Università degli Studi di Milano-Bicocca

Dipartimento di Biotecnologie e Bioscienze

Dottorato di ricerca in Tecnologie Convergenti per i Sistemi Biomolecolari

XXXV Ciclo

Compaction and condensation properties of intrinsically disordered model proteins

Greta Bianchi

Matr. 791906

Tutor: Dott.ssa Sonia Longhi
Supervisor: Prof.ssa Stefania Brocca

Anno Accademico 2021/2022

Table of Contents

Abbreviations

ALS: Amyotrophic Lateral Sclerosis;

APL: Acute Promyelocytic Leukemia;

CBP: CREB-Binding Protein;

CI: Compaction Index;

CRM: Charge Residue Model;

CSD: Charge State Distribution;

DD: Dimerization Domain;

DH: Debye-Huckel;

DMVs: Double-Membrane Vescicles;

EBV: Epstein-Barr virus;

eIF: Eukaryotic Initiation Factor;

eIF2: Eukaryotic translation Initiation Factor 2;

EOM: Ensemble Optimization Method;

ER: Endoplasmatic Reticulum;

ESI: Electrospray Ionization;

FAK: Focal Adhesion Kinase;

FCR: Fraction of Charged Residues;

FCS: Fluorescence Correlation Spectroscopy;

FH: Flory Huggins;

FRAP: Fluorescence Recovery After Photobleaching;

FTD: Frontotemporal Dementia;

FUS: Fused in Sarcoma;

GFC: Gel-Filtration Chromatography;

GFP: Green Fluorescent Protein;

GPC: Gel-Permeation Chromatography;

HIV: Human Immunodeficiency Virus;

hPIV3: human parainfluenza virus 3;

HSP70: Heat Shock Protein 70 kDa;

IAV: Influenza A Virus;

IBs: Inclusion Bodies;

IDP: Intrinsically Disordered Proteins;

IDR: Intrinsically Disordered Regions;

IM-MS: Ion Mobility Mass Spectrometry;

LARKS: Low-complexity Aromatic-Rich Kinked Segments;

LAT: Activation of T-cell;

LCRs: Low-Complexity Regions;

LLPS: Liquid-Liquid Phase Separation;

MAD: Mitotic spindle Assembly checkpoint protein MAD1;

MALDI: Matrix Assisted Laser Desorption/Ionization;

MAPK: Mitogen-Activated Protein Kinase;

MAX: Myc-associated factor X;

MeV: Measles Virus;

MLO: Membraneless Organelle;

MLX: Max-Like protein X;

MONDOA: MLX interacting protein;

MS: Mass Spectrometry;

MuV: Mumps Virus;

NBs: Negri Bodies;

NC: Nucleocapsid;

NCPR: Net Charge Per Residue;

NFM: human Medium Neurofilament protein;

NICD: Nephrin Intracellular Domain;

NMR: Nuclear Magnetic Resonance;

OGT: *O*-GlcNAc Transferase;

PAMPs: Pathogen Associated Molecular Patterns;

PEG: Polyethylene glycol

PB: Poisson-Boltzmann;

PDB: Protein Data Bank;

PIV5: parainfluenza virus 5;

PKR: dsRNA-dependent Protein Kinase;

PML: Promyelocytic Leukemia

PrLD: Prion-Like Domains;

PTM: Post-Translational Modification;

RABV: Rabies Virus;

RAM: RBP-Jk-Associated-Molecule;

ROI: Region Of Interest;

RBP: RNA Binding Protein;

R_g: gyration radius

R_h: hydrodynamic radius

RPA: Random Phase Approximation;

SASA: Solvent Accessible Surface Area;

SAXS: Small Angle X-ray Scattering;

SCD: Sequence Charge Decoration;

SEC: Size-Exclusion Chromatography;

SG: Stress Granule

SLiMS: Short Linear Motifs;

SOS: Son of Sevenless;

SV5: Simian Virus 5;

TDP-43: TAR DNA binding protein 43;

VSV: Vescicular Stomatitis Virus;

Abstract

Proteins lacking a stable tertiary structure are defined as intrinsically disordered proteins (IDPs). The lack of permanent folding gives rise to an ensemble of conformers whose conformational properties are mainly dictated by the amino acid composition. The resulting conformational plasticity gives IDPs the ability to rapidly and reversibly respond to environmental changes and to interact with multiple partners, acquiring an ordered structure upon binding or maintaining fuzzy conformations. As a consequence, IDPs are largely involved in signal transduction cascades and protein condensation phenomena (designated as liquid-liquid phase separation), which give rise to the so-called "membraneless organelles" (MLOs). MLOs include many cytoplasmatic and nuclear condensates, such as stress granules and the nucleolus, which exert crucial biological functions in the cell. IDPs maintain this interaction network primarily through electrostatic contacts, being their chains particularly enriched in cationic and anionic residues, together with prolines and glycines. Besides the net charge of the proteins, the distribution of opposite charges (or charge patterning) has emerged as a key feature dictating the overall size and condensation propensity of IDPs. A sequence parameter, κ, has been introduced to quantitatively describe charge patterning in polypeptide chains. Investigating the relevance of charge patterning on IDP conformation and phase separation is the main objective of this thesis, which is organized into two sections. In the first section, the effect of charge distribution on the conformational ensemble of model IDPs was studied, in parallel assessing the influence of other sequence features such as proline content and secondary structure elements. Three model intrinsically disordered regions (IDRs), namely N_{TAIL}, PNT4 and NFM, similarly charged yet with different proline fractions, were permutated in order to obtain from each wild-type (wt) sequence two κ-variants with different distributions of charged residues: a "Low-κ" variant, with a more regular alternation of oppositely charged residues than in the wt protein; a "High-κ" variant with a more pronounced

separation of opposite charges. The overall amino acid composition and the coordinates of uncharged residues were not altered. Conformational properties of wt and *κ*-variants were assessed combining size-exclusion chromatography and native mass spectrometry. Experimental data confirm that charge clustering induces the remodeling of the IDP conformational ensemble, promoting chain compaction and/or increasing the spherical shape in a protein-specific manner, depending on the sequence context.

In the second section, charge patterning was analyzed in relation to phase separation phenomena. The disordered N-terminal domain (*h*NTD) of human topoisomerase I was chosen as the model IDR for this study due the high density of its charged residues and the propensity for phase separation predicted by bioinformatic methods. It was observed that the charge pattern of NTDs follows a clear evolutionary trend, with vertebrates showing an extremely regular distribution of opposite charges, while yeasts and fungi present the strongest charge separation. Two *h*NTD charge permutants, named M*κ*-NTD and H*κ*-NTD, with progressive clustering of opposite charges, were recombinantly produced to assess the impact of different charge patterns on phase separation. The pH jump and the addition of RNA proved to be effective stimuli for the condensation of all model proteins. Turbidimetric and confocal microscopy analyses confirmed that *h*NTD undergoes phase separation through electrostatic interactions and that the increased clustering of residues with opposite charges impairs condensate morphology and sensitivity to salts and RNA.

Overall, the results included in this thesis help delineate the multifaceted role of charge patterning in determining both single-chain and multiple-chain properties. Hopefully, this work will contribute to decipher the mechanisms underlying the conformational control of IDPs, which is reflected in their cellular functionality.

Riassunto

Le proteine prive di una struttura terziaria stabile sono definite proteine intrinsecamente disordinate (IDP). La mancanza di un *folding* permanente produce un *ensemble* di conformeri le cui proprietà sono dettate principalmente dalla composizione amminoacidica. La risultante plasticità conformazionale conferisce alle IDP l'abilità di rispondere rapidamente e reversibilmente ai cambiamenti ambientali e di interagire con numerosi *partner*, acquisendo una struttura più ordinata in conseguenza al legame oppure mantenendo una conformazione dinamica. Di conseguenza, le IDP sono ampiamente coinvolte nelle cascate di trasduzione del segnale e nei fenomeni di condensazione proteica (denominati separazione di fase liquido-liquido), i quali originano i cosiddetti "organelli privi di membrana" (MLO). Gli MLO includono svariati condensati citoplasmatici e nucleari, come gli *stress granules* ed il nucleolo, che svolgono funzioni biologiche cruciali nella cellula. Le IDP mantengono il proprio *network* di interazioni principalmente attraverso contatti elettrostatici, essendo le loro catene particolarmente ricche di residui cationici ed anionici, così come di proline e glicine. Assieme alla carica netta, la distribuzione delle cariche opposte (o *patterning* delle cariche) è emersa come caratteristica chiave responsabile delle dimensioni complessive e della propensione alla condensazione delle IDP. Un parametro di sequenza, κ, è stato introdotto per descrivere in maniera quantitativa il *patterning* delle cariche nelle catene polipeptidiche.

Investigare l'importanza del *patterning* delle cariche in relazione alla conformazione delle IDP e alla separazione di fase è l'obiettivo principale di questa tesi, che si organizza in due sezioni. Nella prima sezione, è stato studiato l'effetto della distribuzione delle cariche sull'*ensemble* conformazionale di IDP modello, determinando parallelamente l'influenza di altre proprietà di sequenza come il contenuto di proline e gli elementi di struttura secondaria. Tre regioni

intrinsecamente disordinate modello (IDR), ovvero N_{TAIL}, PNT4 e NFM, similmente cariche ma con differente contenuto di proline, sono state permutate così da ottenere, a partire da ciascuna sequenza *wild-type* (wt), due varianti di κ che presentano una differente distribuzione dei residui carichi: un costrutto "Low κ", che mostra un'alternanza più regolare rispetto alla proteina wt dei residui dotati di carica opposta; una variante "High κ" con una separazione più pronunciata delle cariche elettriche opposte. La composizione amminoacidica complessiva e le coordinate dei residui non carichi non sono state alterate. Le proprietà conformazionali della proteina wt e delle varianti di κ sono state determinate combinando la cromatografia ad esclusione molecolare e la spettrometria di massa nativa. I dati sperimentali suggeriscono che la clusterizzazione delle cariche produce il rimodellamento dell'*ensemble* conformazionale delle IDP, il quale promuove la compattazione della catena polipeptidica e/o l'acquisizione di una forma più sferica in maniera proteino-specifica, secondo il contesto di sequenza. Nella seconda sezione, il *patterning* delle cariche è stato analizzato in relazione ai fenomeni di separazione di fase. Il dominio disordinato N-terminale (*h*NTD) della topoisomerasi I umana è stato scelto come IDR modello per questo studio a causa dell'elevata densità di residui carichi e della sua propensione alla separazione di fase predetta attraverso metodi bioinformatici. E' stato osservato che il *pattern* delle cariche degli NTD segue un chiaro *trend* evolutivo, con i vertebrati che mostrano un'alternanza estremamente regolare di residui con carica opposta, mentre lieviti e funghi possiedono la più forte separazione di carica osservata. Due mutanti di carica di *h*NTD, Mκ-NTD e Hκ-NTD, caratterizzati da una progressiva clusterizzazione dei residui cationici ed anionici, sono stati prodotti per via ricombinante al fine di verificare l'impatto del *patterning* delle cariche sul processo di separazione di fase. Sono stati impiegati il salto di pH e l'aggiunta di RNA come stimoli efficienti per promuovere la condensazione delle proteine modello. Le analisi di turbidimetria e microscopia confocale hanno confermato che *h*NTD va

incontro a separazione di fase attraverso interazioni elettrostatiche e che una più spiccata clusterizzazione dei residui amminoacidici con carica opposta intacca la morfologia dei condensati e la sensibilità a sali e RNA. Complessivamente, i risultati inclusi in questa tesi di dottorato aiutano a delineare il ruolo sfaccettato della distribuzione delle cariche nel determinare sia le proprietà della singola catena che quelle multi-catena. Possibilmente, questo lavoro contribuirà a decifrare i meccanismi che soggiacciono il controllo conformazionale delle IDP, che si riflette inevitabilmente sulla funzionalità cellulare di questa classe di proteine.

1. Introduction

1.1 Finding the order out of the disorder: introducing intrinsically disordered proteins

Imagine the marvel -and skepticism- of scientists when intrinsically disordered proteins (IDPs) were discovered in 90s', in a time when structural biology was already a solid and fruitful field of investigation, dominated since some decades by Anfinsen's structure/function paradigm. It is possibly due to this marvel and skepticism that we owe the perhaps belated recognition of the new structural class of IDPs. Indeed, the concept of protein disorder traces back to 1940 by Pauling [1]. Experimental evidences of IDPs were reported in 1950s and 1960s [2, 3]. From 1970s to 1990s, X-ray crystallography helped determining plenty of protein structures, channeled in the Protein Data Bank (PDB [4, 5]), and the early investigation on protein disorder was sadly forgotten [6]. It was in the mid-1990s that IDPs gained popularity, perhaps due to the application of Nuclear Magnetic Resonance (NMR) and computational modelling to protein structure analysis [3]. Structural disorder was not a simple and random deviation from the hierarchical folding, yet it was finally approached as a different mode of protein organization, relevant in several crucial biological functions. Assigning functionality to the disorder represented the major conquest, not because of the paucity of evidences supporting that, yet – and again - because it represented a groundbreaking concept for the accepted structural biology canons. Luckily, the interest in studies concerning IDPs has exponentially increased (**Figure 1**), partly because of their involvement in numerous diseases, from cancerogenesis to neurodegeneration. Consequently, several databases specifically devoted to IDPs have been developed, with DisProt (https://www.disprot.org) arising as a fundamental repository of disorder-related information [7-9]. This database is manually curated and provides information on the structure and function of fully disordered proteins (*i.e.* those

lacking tertiary structure along the entire protein sequence) and regions (IDRs, that are disordered moieties included within larger, structured proteins). Other large IDP/IDR repositories include Mobi Database (http://protein.bio.unipd.it/mobi2/ [10]), the Database of Disorder Protein Prediction (D^2P^2, http://d2p2.pro [11]) and the DescribeProt Database (http://biomine.cs.vcu.edu/servers/DESCRIBEPROT/ [12]).

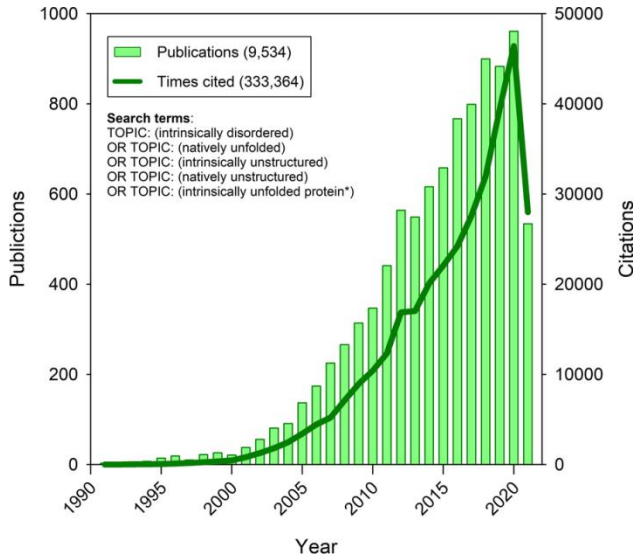


Figure 1. Timeline of the number of publications related to structural disorder (light green bars) and the number of papers citing those articles (dark green line). Data were retrieved from Web of Science on August 01, 2021. Figure from [6].

Yet, what does it mean for a protein chain to be disordered? Structural disorder has more to do with lacking a stable tertiary structure than assuming just purely random-coil conformations. Indeed, in contrast to folded proteins, which acquire a stable tertiary structure as the result of Gibbs free energy minimization, IDPs sample various isoenergetic inter-converting conformers (the so-called *conformational ensemble*), which evolve one into another at the equilibrium, undergoing more dramatic conformational fluctuations than their globular counterparts. This is where the disorder potentiality lies. Since conformational

malleability expands the interaction ability of a protein, and IDPs/IDRs are more abundant in proteomes of higher eukaryotes than in microbes, it has been hypothesized that disorder content correlates with complexity of organisms and is thus evolutionarily favored [13-15]. A case in point is that of viral proteomes, in which the abundance of disordered proteins has been related to the need to minimize the number of proteins. Indeed, a single IDP/IDR may interact with plenty of ligands, conjugating pleiotropy and genetic compaction [16].

The IDP binding mode may imply both the preservation of conformational flexibility, as for "fuzzy complexes" [17] and a disorder-to-order transition induced by the binding to a ligand. In the latter case, an ordered structure can be achieved through at least two distinct mechanisms, namely "conformational selection" and "induced fit". The conformational selection occurs when conformational changes precedes the binding, so that the ligand picks and stabilizes a high-energy binding conformation [18]. In the case of induced fit, the conformational rearrangement follows the binding of the ligand.

Post-translational modifications (PTMs) largely contribute to control these conformational fluctuations. The IDP accessibility to modifying and de-modifying enzymes may explain their high modification rate [19]. These modifications may include the addition of functional chemical groups (phosphoryl-, alkyl-, glycosyl-, acyl-groups, or small proteins as the ubiquitine) or the alteration (e.g. by oxidation) of pre-existing side chain properties, altering the energetics of IDP conformational landscape [20]. A single PTM could impair relevant intra-chain contacts, thus swelling the chain, or conversely stabilize a secondary structure motif producing a conformation collapse. Furthermore, it may modulate the ability to interact of IDPs. To cite an instance, the binding capabilities of 4E-BP2 IDP towards 4E eukaryotic initiation factor (eIF) are suppressed by multiple hierarchical phosphorylation events, determining a disorder-to-order transition responsible for structuring a four-stranded-β-domain incompatible with the binding [21]. An IDP

may receive single or multiple modifications simultaneously, on a same or different amino acid, acting in an additive or competitive manner. This gives a glimpse of the complexity of biological regulation of IDP structure.

Being the protein structure encoded by the amino acid sequence, evidences for the disordered nature have been sought at the residue-level. Interestingly, IDPs display marked sequence degeneracy, namely an impoverishment of the amino acid alphabet in favor of some residue type forming repetitive interaction motifs or low-complexity regions (LCRs). Prolines and glycines are quite represented, acting as secondary structure breakers due to the steric rigidity or enhanced flexibility of the sidechains, respectively. Residues that contribute to core folding, such as the hydrophobic ones, are generally depleted, while amino acids more abundant on the surface of globular proteins, such as charged ones, are enriched. Indeed, theoretical studies have modeled IDPs as "charge decorated" polymers, drawing copiously from polymer physics to rationalize their peculiar behaviour [22-26]. IDPs have been referred to as either polyelectrolytes – bearing a disproportion of multiple charges of the same sign – or, more frequently (approximately the 75% of IDPs) as polyampholytes – characterized by electrostatic neutrality [27]. The ensuing paragraph presents an in-depth investigation of the parameters used to classify and capture the IDP/IDR electrostatics, focusing in particular on the impact of density and distribution of charges on the compactness of disordered proteins.

1.2 Sequence-based conformational classification of IDPs: how protein charge governs IDP conformation.

Sequence-based classification of IDPs began with the description of the charge-hydropathy (*C-H*) correlation, namely the "Uversky's plot", which simply classifies IDPs/IDRs according to their position in the two-dimensional space defined by the mean hydrophobicity "*H*" and the mean net absolute charge "*q*" [28]. This latter is equivalent to the absolute value of the net charge per residue (|NCPR|), a sequence parameter defined as the difference between the fraction of positive (f_+) and negative (f_-) residues [27]. However, the accuracy of the plot is limited by the employment of an absolute value of charge, which prevents considerations on the polarity of the polypeptide chain and assigns the same coordinates to polypeptides with equal but opposite charges. Therefore, the *C-H* plot has been revised, considering as *C* the full range of NCPR values (**Figure 2a**). Nonetheless, the binary order/disorder classification was inadequate to predict the conformational preferences of polyampholyte/polyelectrolyte IDPs. Indeed, same NCPR may correspond to both polyampholytes and polyelectrolytes, which, instead, display distinct dimensions and conformation preferences, as verified by Mao and collaborators on protamines [29], arginine-rich IDPs [30] involved in chromatin condensation [31] during spermatogenesis [32] and packaging of viral genomes [33]. The attempt to overcome the Uversky's plot limitations and provide a further detailing of IDP conformational scenario led to a three dimensional representation, the Mao-Pappu's plot (**Figure 2b**), representing variations of *H* with respect to f_+ and f_- treated separately and not cumulatively through the NCPR. Thus, sequences with low *H*, which previously were indiscriminately collected as disordered, could be distinguished into swollen coils or disordered globules, according to their belonging to polyelectrolytes or polyampholytes respectively. Indeed, polyelectrolytes (characterized by $f_+ >> f_-$ or *viceversa*) display more expanded conformations, due to the electrostatic repulsions of unshielded similar

charges and favorable polymer solvation [34]. Conversely, polyampholytes (with $f_+ \approx f_-$) assume more compact conformations, due to the balancing of opposite charges. Nevertheless, even Mao-Pappu's classification could not tackle the conformational heterogeneity occurring within the class of polyampholytic IDPs. That depends on the entity of the charged fraction (described by the FCR parameter, obtained as the sum of f_+ and f_-), which further distinguishes polyampholytes between "strong" (FCR ≥ 0.3 and NCPR ≤ 0.35) and "weak" (FCR < 0.3 and NCPR < 0.25). The Das-Pappu's phase diagram (**Figure 2c**) implements this classification, helping to further detail general globules into several conformational classes dependently on the FCR value [27, 35]. In this diagram, the 70% of IDPs previously classified as globules were found to be either strong polyampholytes (R3 region, **Figure 2c**) exhibiting coil, hairpin and chimera conformations, or to occupy an intermediate boundary region (R2) between R1 and R3 (**Figure 2c**), with a conformational behavior dictated by several factors [27].

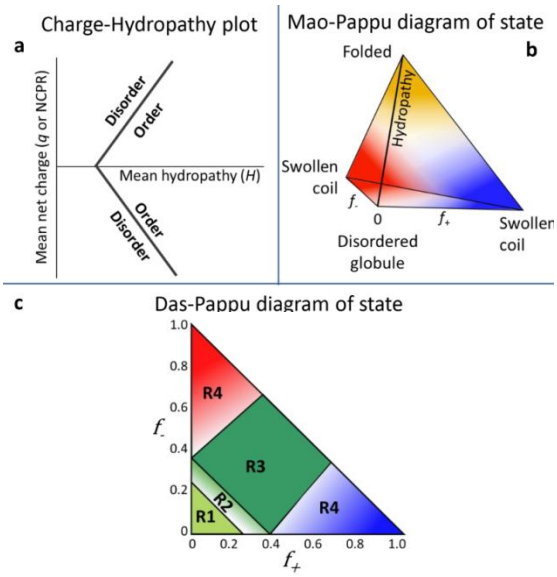


Figure 2. Conformational phase diagrams of IDPs. (**a**) Revised Uversky's plot representing the mean net charge *q* (or NCPR) versus mean hydropathy (H). The two solid lines demark the boundary between ordered and disordered polypeptide

chains [28, 36]. (**b**) Mao-Pappu's three-dimensional plot defined by f_+, f_- and H. Since high hydropathy and large fractions of charged residues are mutually exclusive, the space has a pyramid shape. At the top of the pyramid, the yellow area marks the folded protein area. The blue and red areas correspond to positively-charged and negatively-charged polyelectrolytes respectively, assuming swollen coil conformations. In correspondence to the zero, polyampholyte sequences with disordered globule conformations are found [29]. (**c**) Das-Pappu's phase diagram. It contains four regions (R1-R4) corresponding to distinct conformational classes: R1, weak polyampholytes or weak polyelectrolytes that form globules or tadpole-like conformations; R3, strong polyampholytes with non-globular conformations, such as coil-like, hairpin-like, or a mixture; R2, continuum of conformations between those in R1 and R3; R4, strong polyelectrolytes with FCR > 0.35 and |NCPR| > 0.3, which sample coil-like conformations [35]. Figure adapted from [34].

However, even the FCR alone could not catch the conformational heterogeneity of strong polyampholytes [27]. Indeed, whilst weak polyampholytes are univocally predicted to assume globular conformations by atomistic simulations, strong ones have conformational preferences dependent on the combination of both the FCR and linear pattern of charged residues [27]. Early studies on the key role of charge pattern in polyampholyte conformation date back to the second half of 1990s. The simulated gyration radii (R_g) of twenty randomly charged co-polymers proved that the location of charges decides the interplay between repulsive electrostatic interactions (the "polyelectrolyte effect") between similar charges and the attractive forces between opposite charges ("the polyampholyte effect" [37]). A quantitative parameterization of charge pattern became indispensable to properly investigate its connection with IDP conformation. Rohit Pappu's group introduced the empirical parameter κ [27]. κ quantifies the charge asymmetry partitioning the protein sequence into N overlapping windows (or blobs) of five/six amino acids (for

sequences lacking proline residues). The local charge asymmetry of each *i* segment is calculated as in Eq. 1:

Eq. 1 $$\sigma_i = \frac{(f_+ - f_-)_i^2}{(f_+ - f_-)_i}$$

The squared deviation of asymmetry is obtained as in Eq. 2:

Eq. 2 $$\delta = \frac{\sum_{i=1}^{N_{blob}} (\sigma_i - \sigma)^2}{N_{blob}}$$

κ is defined by the ratio between δ and the maximal value for a given amino acid composition (δ_{max}; Eq. 3)

Eq. 3 $$\kappa = \frac{\delta}{\delta_{max}}$$

κ varies between 0, obtained when opposite charges alternate regularly, and 1, when opposite charges are segregated into two clusters. Both these extremes are accessible to polyampholytes entirely composed of charge residues, and not reachable by natural amino acid sequences, which reflect a more heterogeneous composition and however fall within this interval of κ values. In their seminal paper, Pappu and collaborators designed a set of thirty synthetic sequences $(Glu\text{-}Lys)_{25}$, differing in the permutation of oppositely charged monomers ($0 \leq \kappa \leq 1$, NCPR = 0; *Das et al., 2013*). Employing all-atom Monte Carlo simulations applied to Flory's random coil model [38, 39], the authors confirmed Srivastava-Mutukumar's intuitions, illustrating that the ensemble-averaged R_g inversely correlates with κ [27]. The more "blocky" the sequence, the larger the chain collapse.

Nonetheless, the intuitive computation of κ has limitations, as not accounting for long-range electrostatic interactions. Better performances in this sense are offered by Sawle-Ghosh's analytical Sequence Charge Decoration (SCD) parameter, described by the Eq. 4:

Eq. 4 $$\text{SCD} = \frac{1}{N}\left[\sum_{m=2}^{N}\sum_{n=1}^{m-1} q_m q_n (m-2)^{1/2}\right],$$

where *m* and *n* are the sequence positions in a N-mer chain and q_m and q_n are the charges of the residues in those coordinates [40]. As for κ, SCD is equal to 0 when the distribution of oppositely charged residues is perfectly regular. The SCD absolute value increases with both charge patterning and polymer size and assumes values ≤ 0. Lin and Chan tested SCD performances in predicting R_g, proving a smoother R_g – SCD correlation than the R_g-κ one (**Figure 3**)[41]. This dependence is remarkable, since the simulated R_g has no formal relationship with the analytical theory employed for SCD calculation. That could depend on the SCD accounting for all pairwise interactions, both short-ranged and long-ranged [41]. Nevertheless, Ghosh's formalism is defective in depicting collapsed globules, since the model overestimates the global dimensions, possibly due to the employed dielectric constant or the negligence in accounting for hydrogen-bonding and ionization equilibria [34, 40, 41].

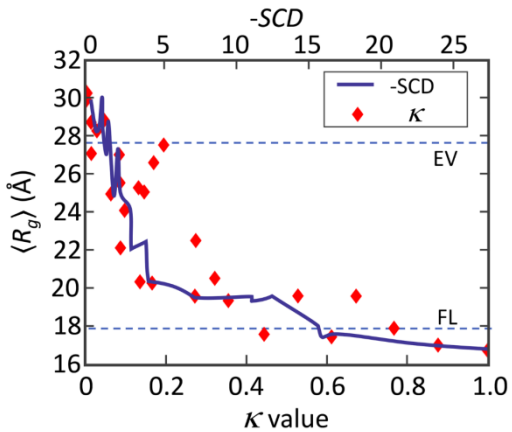


Figure 3. Ensemble-averaged R_g for the $(Glu\text{-}Lys)_{25}$ system versus κ and *–SCD*. The blue dashed lines represent the R_g values expected for the same sequences when modeled as excluded-volume (EV)-limit polymers and as Flory random coils (Flory limit, FL). Figure from [34].

Proofs of the conformational heterogeneity of sequences with same FCR were produced also by Firman and Ghosh that applied the Sawle-Ghosh's heteropolymer theory [40] to the entire DisProt database [42]. To do this, they implemented the model to account for collapsed globules and coil-to-globule transitions [42]. Interestingly, they created a grid defined by f_+ and f_- and computed the average chain expansion parameter x (defined as the ratio between the end-to-end distance of a given protein and that in the Flory random coil limit) for all the protein sequences belonging to the same bin (**Figure 4**). Along the diagonal, the x parameter is low, suggesting that polyampholytes with equal f_+ and f_- assume globule-like conformation (**Figure 4A)**. Increasing the net charge produces more expanded conformations, reaching the coil-like state of polyelectrolytes (high x value)[42]. To determine the "intrabin variation", that is the conformational heterogeneity occurring for proteins with the same FCR, the authors computed the standard deviation in x for all the proteins within the same bin (**Figure 4B**), observing a similar trend. The major standard deviations occur further away from the diagonal [42]. These evidences were supported by computing the fraction of proteins within a bin predicted to be coil-like (**Figure 4C**). Low values of this fraction correspond to globule-like structures (for instance for poorly charged, neutral polyampholytes), whereas the higher the fraction the more expanded the conformations (as for polyelectrolytes). Interestingly, some of the bins already highlighted as highly heterogeneous displayed a fraction of 0.5, indicative of a mixture of both globules and coils. Firman and Ghosh named these bins "sweet spots" to indicate proteins for which sequence specificity, including post-translational modifications as the phosphorylation or mutations, strongly impacts on chain dimensions [42]. The authors reproduced the same analyses for charge-permutated synthetic polymers composed of lysine, aspartic acid and alanine, without any interference of polar and hydrophobic residues present in naturally occurring IDPs. These model polypeptides behave as expected, confirming the hypothesis that strong polyampholytes, which are prevalent components of the

"sweet spots" region, have conformation strongly dictated by the charge content and charge pattern. Conversely, both high and low net charge proteins, namely polyelectrolytes and weak polyampholytes with NCPR ≈ 0, assume coil-like and globular conformations respectively, regardless of charge patterning.

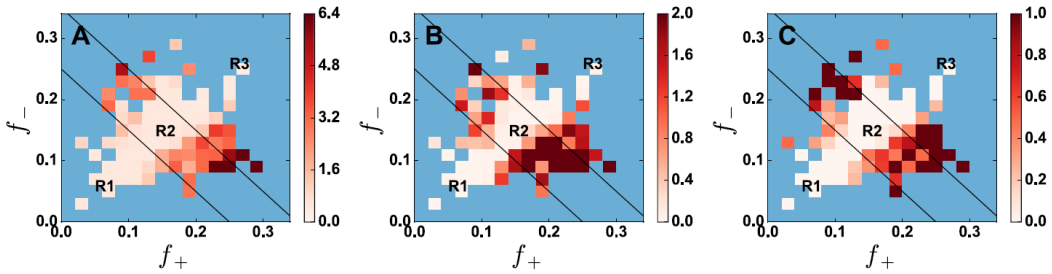


Figure 4. (**A**) Heat map of protein size distribution predicted by the Ghosh's analytical model for the DisProt entries [42]. The bins correspond to f_+ and f_- classes, while the color scale represents the average expansion index x at T = 300 K. The map represents with different color the average x among proteins belonging to a given bin. Solid lines define R1, R2 and R3 regions as in *Das et al., 2015*. Blue bins correspond to classes containing less than two proteins. (**B**) Heat map of the standard deviation in x for a given bin. (**C**) Heat map of the IDP fraction in the coil state within a given bin. Figure from [42].

Sequence specificity helps rationalizing the dependence of IDP conformation on salt exposure as well. Depending on their net charge, IDPs shrink or swell with increasing ionic strength. Polyelectrolytic IDPs undergo compaction, due to the shielding of repulsive electrostatic interactions, as proved for human prothymosin α (Protα) [43]. The stronger the polyelectrolyte, the larger the collapse effect, as predicted by the polyelectrolyte theory [44]. Conversely, the behavior of polyampholytes results less intuitive. In general terms, polyampholytes are expected to expand upon salt addition, since the electrostatic attractive contacts are impaired by charge screening. In some cases, the IDP conformation responds to salt with an initial expansion, followed by a collapse as the ionic strength increases

[45]. This non-monotonic trend has been observed for Myc, Myc-associated factor X (MAX), Mitotic spindle assembly checkpoint protein MAD1 (MAD), Max-like protein X (MLX) and MLX interacting protein (MONDOA). More in detail, evidences collected for Myc suggested that this response reversal is due to the prevalence of hydrophobic interactions over charge shielding. Moreover, different salts produce various compaction effect, echoing the salting-out effect along the Hofmeister series [45].

Huihui and collaborators proposed that the conformational response of polyampholytes to salts could be affected by charge patterning [44]. The authors employed the SCD charge pattern metric and reformulated it in low and high salt regimes ($SCD_{lowsalt}$ and $SCD_{highsalt}$, respectively; for the equations see [44]), in order to predict the salt-induced chain shrinkage or expansion accordingly to the SCD sign. Indeed, Huihui predicted that, near the zero salt limit, sequences with positive $SCD_{lowsalt}$ would collapse upon salt addition, whereas negative $SCD_{lowsalt}$ would result in a chain expansion. In high salt conditions, sequences with $SCD_{highsalt} > 0$ would shrinkage with further increase of the ionic strength [44]. Thus, the possibility of polyampholyte sequences with $SCD_{lowsalt} < 0$ and $SCD_{highsalt} > 0$ may justify the observed non-monotonic behavior.

An experimental assessment of Huihui's theory was provided by measuring the salt-dependent conformations of κ permutants (κ14, with $\kappa = 0.14$; κ56, with $\kappa = 0.56$; FCR = 0.291) of the C-terminus of human cell-cycle inhibitory protein (p27-C, $\kappa = 0.31$, FCR = 0.291)[46]. Ion-mobility mass spectrometry (IM-MS) experiments indicated that the conformational ensemble of each construct is modulated by the ionic strength of the spraying solution [46]. The permutant bearing less clustered charged residues, *i.e.* κ14, turned out to be insensitive to salt addition, adopting extended conformations. This in agreement with previous SAXS measurements [47] indicating that IDPs with FCR ≥ 0.3 and low κ behave as self-avoiding random walks [27]. At high salt, nonetheless, a small fraction of

conformers become more compact, possibly due to the screening of proximal like-charges and the attenuation of their repulsions. The permutant with the most clustered charged residues, κ56, results more compact and responds to salt addition monotonically, through a progressive swelling of the chain, consistently with the shielding of attractive electrostatic contacts between oppositely charged blocks [46]. Conversely, wild-type κ31 revealed a non-monotonic dependence of its conformational transitions on salts. In the medium-salt regime, a shift towards expanded conformers was observed, as expected for strong polyampholytes; a further increase in the ionic strength was accompanied instead by the accumulation of compact structures. Thus, this re-entrant behavior, intrinsically encoded in wild-type protein charge clustering, could be necessary for the biological functionality of the polypeptide. Indeed, for instance, p27-C charge pattern proved to modulate the phosphorylation efficiency at Thr187, that is the primary site of protein phosphorylation [47]. As a matter of fact, it was observed that permutating the charged residues (even without mutating Thr187) greatly affects the PTM efficiency at this residue [47]. Another example of the functional relevance of charge pattern is represented by RBP-Jk-associated-molecule (RAM) region, an IDR belonging to the intracellular domain of Notch receptor (NICD) [48]. Herein, charge distribution mediates RAM binding affinity for CSL, an element of the tertiary complex NICD-CSL-MAML involved in Notch activation. Among RAM charge permutants, both R_g and CSL affinity decreased with increased blockiness. Therefore, charge pattern could be assumed, for all intents and purposes, as an evolutionary-selected trait, able to influence IDP conformational preferences and partner interactions [34].

The predicted impact of charge pattern on IDP conformation, however, may be affected by background compositional features, for instance the fraction of proline residues (P_{PRO}), notoriously enriched in IDPs. Indeed, Marsh and Forman-Kay included this factor – the frequency of Pro residues, P_{PRO} - into the equation of

theoretical hydrodynamic radius (R_h) – which returns the average molecular compactness of an IDP - besides the net charge [49]. An experimental paper suggested that differences in protein responsiveness to charge pattern modifications reflects the proline content of IDPs [50]. In this work, two viral IDRs - N_{TAIL} from Measles virus nucleoprotein N and PNT4 from Hendra virus phosphoprotein P – with similar FCR, NCPR, and κ values yet different proline content (~ 5% and ~11%, respectively) have been scrambled, originating two charge permutants per each, with low and high κ. Merging SAXS, size-exclusion chromatography (SEC) and limited proteolysis, the authors proved the compaction trend with increasing κ, yet the responsiveness in terms of compaction turned out be lower in PNT4, that is the protein with the largest proline content [50]. This suggests that prolines, together with chain length and different secondary structure content, may be synergistically at play to hinder IDP conformational response to charge pattern [51]. Understanding this remains an open challenge, particularly considering that even single residue properties could dramatically influence the overall IDP conformational ensemble. Indeed, a recent publication evidenced that the identity of basic residues can modulate the stability of strong polyampholytes, with arginines favoring globules more than lysines [52]. In parallel, the shorter side chain of aspartate compared to glutamate one proved to favor metastable necklace-like conformations [52]. Reaching the awareness of IDP sequence-conformation relationship will permit to understand the cellular exploitation of disorder and, hopefully, to employ this knowledge in the design of intelligent stimuli-responsive biomaterials [53].

1.3 Liquid-liquid phase separation and the new era of membrane-less organelles

Scientists have long strived for finding order in the apparent chaos of biological systems. Although disorder is the spontaneous drift of thermodynamic systems, cells have needed to evolve levels of supramolecular organisation able to circumvent the stochastic nature of biochemical reactions and Brownian molecular motion [54]. In addition to the compartmentalization of the protoplasm in organelles, based on lipid bilayers, a more dynamic layering of sub-cellular organization has been implemented. Protein "quinary structure", namely the fifth level of protein structure, refers to heterogeneous macromolecular complexes, hold together by weak and transient contacts [55], which further organize cellular biochemical processes and signal transduction cascades [56]. Quinary structures promote multi-enzyme complexes, as the Krebs cycle "metabolon" [57, 58], favoring substrate channeling of reaction intermediates and enhancing metabolic fluxes [59-62]. Also, quinary structure goes beyond the proper positioning of enzymes acting in chain reactions, resulting in the formation of rather complex supramolecular structures that have been termed as membrane-less organelles (MLOs).

MLOs marry the needs for buffering the cellular noise - which produces fluctuations in nucleic acid and protein concentrations - and relying on dynamic, environment-sensitive condensates. A plethora of MLOs have been found in the nucleoplasm – including the nucleolus, nuclear speckles, Cajal bodies, histone–locus bodies, and promyelocytic leukemia (PML) bodies [63, 64] – and in the cytoplasm – as the P bodies, stress granules and germ granules [65, 66]. They arise from liquid-liquid phase separation (LLPS), a physical demixing occurring in supersaturated systems of adhesive macromolecules. Indeed, whenever the solubility limit is reached, namely the solute saturation concentration (C_{sat}), two immiscible phases are produced, differing in the partitioning of the solute yet

displaying equal chemical potential (μ). That occurs since μ varies non-monotonically with concentrations in an interacting system (**Figure 5**). Consequently, no net diffusive fluxes are allowed, particularly at phase boundaries [67, 68], determining the "structural integrity" of condensates, without any energy squander. Nevertheless, this does not imply that any diffusion across interfaces is prevented. Random movements of particles between phases occur, with same number of outgoing molecules as incoming ones, which preserve the equivalence in chemical potential. In thermodynamic terms, the achieved equilibrium corresponds to a free energy minimum [69], in which the entropic cost of order acquisition is counterbalanced by the enthalpic contribution of intermolecular interactions, with long-range contacts and moderate binding energy promoting liquid condensation [70].

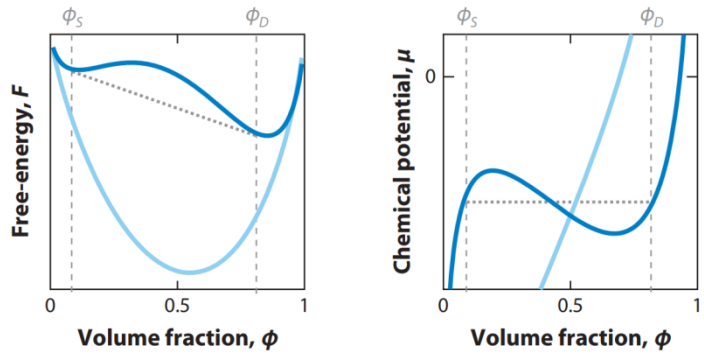


Figure 5. Thermodynamics of phase separating systems. (**Left**) The free energy (F) is a function of volume fraction. In the absence of interactions, free energy is unimodal, since $F = -TS^{mix}$, where T is the temperature and S^{mix} is the "mixing entropy", that is the entropy associated with the mixing (light blue curve). In an interacting system, the F curve becomes multimodal (dark blue curve). The [Φ_S – Φ_D] represents the range of volume fraction where condensation occurs. (**Right**) Without intermolecular interactions, the chemical potential (μ) is monotonic (light blue curve). In case of demixing, the curve becomes non-monotonic and two

different values of volume fractions (Φ_S and Φ_D) display same μ being the system thermodynamically stable. Figure adapted from [68].

Environmental parameters – such as temperature, pH, and ionic strength – influence the networking capabilities of macromolecules. Therefore, the generation of phase diagrams, depicting the set of experimental conditions under which LLPS occurs, normally considers two parameters, such as volume fraction (Φ) and a specific stimulus (as the solution pH; **Figure 6**). The solid line represents the "coexistence curve" or "binodal curve" and delimits the two-phase regime from the solution state. The dotted line instead represents the "spinodal curve", which demarcates a sub-region of the phase separating area in which condensation occurs spontaneously, through spinodal decomposition, without the need for nucleation events. The area between the spinodal and binodal curves, on the other hand, defines a metastable region, in which condensation requires nucleation events. There is, in fact, a nucleation barrier, below which molecular nuclei shrink and above which they keep growing until equilibrium is reached (*Martin et al., 2021*). The nucleation kinetics depends on the probability of reaching such a nucleation barrier and it is strictly protein- and context- dependent. Possibly, conformational fluctuations or oligomerization steps could be involved in reaching critical nuclei [67, 71].

An interesting consideration related to the biological role of two-phase systems requires recalling the definition of "tie line", *i.e.*, a horizontal line, which crosses the phase diagram in the two-phase region, intersecting the binodal and spinodal curves (it is the horizontal yellow line in the diagram in **Figure 6**). Along a given tie line, corresponding to a specific stimulus condition, the concentrations of the dense and dilute phases are conserved, yet the volume fractions of the two phases change relatively to each other [72]. This may represent a cellular "passive filter" to control local concentration of reagents regardless of transient fluctuations in the total concentration [54].

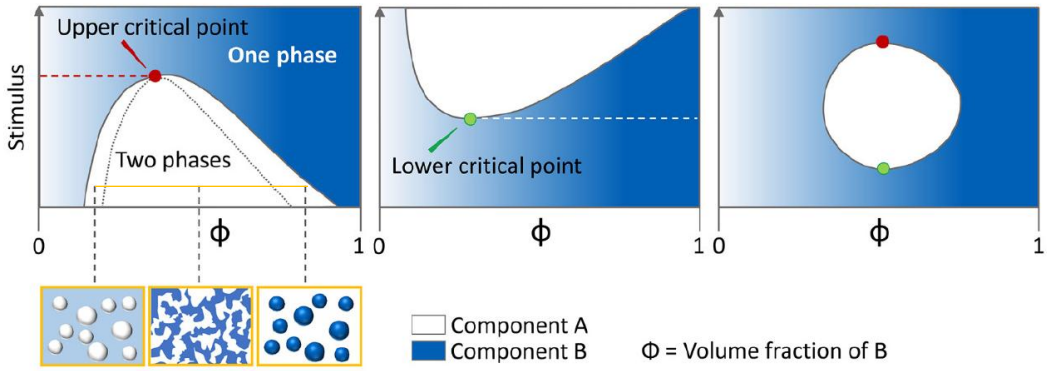


Figure 6. Phase diagram of a two-component mixture. LLPS can occur below or above the critical value of a specific stimulus (temperature, pH, ionic strength etc.). The diagram illustrates the binodal curve (or coexistence curve, solid line) of three reference systems in response to the stimulus. (**Left side**): condensation occurs below a threshold point, namely the upper critical point (UCP). Within the two-phase regime, the spinodal line (dotted line) delimits the conditions in which phase separation occurs spontaneously. The pictorial representations in the yellow boxes show the condensation of components A e B (left and right) and spinodal decomposition (in the middle). (**Middle**): phase separation occurs above the lower critical point (LCP). (**Right side**): the system features both UCP and LCP. Figure adapted from [34].

Therefore, MLOs may act as "protected" bioreactors where catalytic conversion is controllable, guaranteeing continuous influxes of substrates and effluxes of products (without involving carrier proteins), high enzyme concentration and reduction of possible interferences. MLOs can control chemical reactions, by boosting or slowering their kinetics through the material state of the condensate (*i.e.* by lowering or increasing viscosity, respectively), modifying their specificity (*i.e.* by including or excluding a specific substrate) and activity (*i.e.* by including or excluding potential inhibitors) [67, 73]. These general functions are exerted by different and specific MLOs. To cite an instance, stress granules (SGs) are widely

characterized, cytoplasmic assemblies of ribonucleoproteins (mRNPs) containing untraslated mRNAs whose translation is stalled due to environmental stress, drugs or viral infections [74, 75]. Cells can benefit of SGs as "preservation clusters" [76], saving the energy of degrading or re-synthesizing mRNAs after stress removal. Interestingly, SGs contribute to enhance the innate immune response against viruses as well, concentrating and activating numerous anti-viral proteins, such as RIG-1, PKR, RNAseL, and OAS [75]. Not by chance, viruses often hamper SG assembly through the proteolytic cleavage of their principal component G3BP [77]. In addition, SGs sequester components from TOR, RACK1 and TRAF2 signalling pathways, modulating multiple cascades of signal transduction [78-81]. Another example of MLO versatility is offered by the PML-bodies. Discovered because of their dysfunction in acute promyelocytic leukemia (APL), PML-bodies are involved in plenty of functions, ranging from facilitating protein SUMOylation, storaging proteins (regulating nuclear protein availability), enhancing anti-viral response, regulating cellular senescence and stemness [82]. PML-bodies possibly contribute in the regulation of chromatin dynamics, since several chromatin-related proteins are stored within them, as HP1, involved in heterochromatin formation [83], CREB-binding protein (CBP) [84], and DAXX, an histone chaperone associating with chromatin remodeler ATRX [85, 86].

Despite this functional heterogeneity, the MLO assembly is common and displays a hierarchical mechanism [87]. First, the "scaffold" molecules – namely the core components of the condensate – self-associate through heterotypic interactions, increasing the local concentration and undergoing LLPS [88]. Secondly, the "client" species – that are molecules dispensable for the MLO assembly - are recruited and distributed throughout its body. Typically, clients display lower valency, and hence low apparent affinity, implying a reduced competition with high-affinity scaffolds and a more dynamic exchange throughout the compartment [87]. Their recruitment is governed by the scaffold stoichiometric ratios. Any

change in the valency of client molecules results in weaker or stronger recruitment, thus impacting on its degree of partitioning [87, 89]. The binary classification into scaffolds and clients finds a limitation in the interchangeability of roles for some molecular species (*e.g.*, when the valency of the client approaches that of the scaffold [87, 90]) or in the presence of internal sub-compartments that drastically alter the canonical structure of droplet/bulk system [91-94].

1.4 Liquid condensation as a preferred state for polypeptide chains.

Proteins and nucleic acids provide the multivalency/modularity required for the intermolecular contacts underlying their condensation. Multivalency implies the harboring of multiple sites for both intra- and intermolecular interactions and is eminently prominent in IDPs and IDRs [88]. Their sequence degeneracy, favoring low complexity, encodes residue repetitions likely forming short linear motifs (SLiMs) of interaction. For instance, arginine-glycine-glycine repetitions form the RGG motif, a SLiM typical of RNA binding proteins (RBPs). In general, SLiMs promote three-dimensional networking of protein chains, thus behaving as "stickers" [95]. The massive abundance of IDRs participating in LLPS confirms their involvement in shaping the properties of the condensates. Experimental data were obtained on many different IDPs that include – to cite some examples – Laf-1 protein [96], RNA helicase Ddx4 [97], hnRNPA1 [98], TAR DNA binding protein 43 (TDP-43)[99]. Besides the SLiMs, IDPs/IDRs have many features suitable for LLPS. To start with, they lack a stable tertiary structure, preferring a dynamic conformational *ensemble*, which expands their potential interactome and therefore molecular networking [100]. In addition to it, their expanded conformation offers accessibility to post-translational modification enzymes, which act as positive/negative switchers for LLPS [101-103]. Finally, the presence of IDPs/IDRs in condensates guarantees a more porous architecture if compared to globular protein condensates, which results in more dynamic reciprocal exchanges with the surrounding environment [100].

Among the interactions mediated by IDPs/IDRs, cation-π, π- π stacking, electrostatic networking and transient cross-β-contacts are predominant in LLPS condensation. In particular, π-π stacking - due to delocalized π electrons of aromatic residues (tyrosine, tryptophan, and phenylalanine) - proved to be crucial for the majority of to-date known phase separating proteins [104]. That is the case of NICD, whose LLPS capability strongly depends on tyrosine networking through

hydrophobic and electrostatic interactions [105]. Cation-π contacts – formed between positively charged residues and aromatic ones – are also relevant for several proteins to achieve phase separation. Another example is given by Ddx4, whose phenylalanine-glycine (FG) motifs interact with arginine-glycine (RG) ones driving its phase separation both *in vitro* and *in vivo* [97]. However, besides cation-π interactions, π-π stacking and electrostatic networking contribute to Ddx4 LLPS equally [73, 97, 105]. Therefore, rather than the uniqueness and specificity of the interactions involved, it is the transient nature and weakness of those contacts to be crucial in promoting LLPS and discerning it from irreversible aggregation. That is particularly evident in the so-called LARKS (Low-complexity Aromatic-Rich Kinked Segments), kinked cross-β sheets formed by prion-like domains (PrLDs) of RBPs, as fused in sarcoma (FUS) protein [106, 107]. PrLDs are a subset of low complexity regions, characterized by uncharged polar amino acids and glycines, found in many RBPs involved in neurodegenerative disorders associated to pathological aggregation, as amyotrophic lateral sclerosis (ALS) [108]. LARKS proved to be less thermodynamically stable than canonical amyloid fibril β-sheets [106], suggesting that the discriminating factor between cross-β interactions associated either with LLPS or irreversible aggregation resides in the low-energy and transient nature of the former ones [73]. The extent of protein multivalency controls the strength and half-life of these interactions, which also depends on the spatial organization of the "stickers" among the "spacers". The former consist in the adhesive units primarily contributing to the interaction network. The latter connect the stickers, influencing their ability to interact and the chain solvation [100]. Indeed, solvating spacers can suppress the coacervation, at least when protein concentration is medium-low [100]. Identifying the stickers within a protein chain represents a major challenge. The first strategy involves a mutagenesis approach, in which individual residues or short motifs are destroyed or modified and the impact on phase behaviour is evaluated [95, 109]. Of course, the removal of a sticker hinders the LLPS. For instance, this approach led to the identification

of tyrosines and arginines as major stickers in LCRs of proteins belonging to FET family [95]. Nonetheless, being the spacers equally relevant in determining protein condensation, stickers may not be identified unambiguously. Another strategy pinpoints the stickers considering that they are involved – in addition to interchain contacts relevant for the LLPS - in intramolecular interactions responsible for chain compaction in solution [100]. This approach helped the identification of aromatic residues as stickers within the PrLD region of hnRNAP1 [110]. Interestingly, the authors were able to experimentally derive the interaction strength of sticker pairwise aromatic-aromatic interactions. Small interaction energies were assigned to sticker-spacer and spacer-spacer contacts [110]. Currently, the implementation of the stickers-and-spacers model to IDPs is limited to PrLD regions [95, 110, 111] and further work will be needed to include other sticker types.

Considerations on the distribution of charged residues and their role in LLPS are examined in depth in the following paragraph.

1.5 The role of charge pattern in LLPS

Electrostatics represents one of the major energetic component governing the phase separation of IDPs/IDRs [41]. Seminal experimental papers proved LLPS sensitivity to charge pattern, in Ddx4 disordered N-terminus ($Ddx4^{N1}$) [97], NICD protein [105] and LAF-1 [112]. Polymer physics theories aimed at rationalizing this correlation. Unfortunately, explicit-chain models of phase-separated states are computationally expensive, due to the need of tracking the configurations of a relatively large multiple-chain system. For this reason, early efforts have been made to develop analytical theories for IDP phase behaviour, possibly accounting for long-range electrostatic interactions beyond the Flory-Huggins (FH) theory [113], which considers attractive contact-like Coulomb interactions between residue pairs [114, 115]. Analytical theories proved beneficial for modelling polyelectrolytic and polyampholytic systems, which IDPs are approximated to. Beginning with the Debye-Hűckle (DH) theory in 1923 – namely the linearised Poisson-Boltzmann (PB) equation – [116], further implementations of DH slowly emerged – as the Derjaguin, Landau, Verwey, and Overbeek (DLVO) theory [117, 118], accounting for van der Waals forces, or the beyond mean-field PB equation theories, such as Random-Phase Approximation (RPA) approach. RPA has been widely employed for polyelectrolyte and polyampholyte modelling and, despite other theories as DH, treats chain connectivity explicitly, accounting for sequence-specific long-range electrostatic interactions [113]. Therefore, it proved effective in modelling the electrostatic impact over polyampholyte phase separation, particularly with regard to charge pattern implications, as verified by successfully reproducing the experimentally-checked Ddx4 phase behaviour [113]. $Ddx4^{N1}$ owns a blocky alternation of opposite charges associated with *in vitro* and *in vivo* phase behaviour [97]. Its scrambling to a regular alternation of positive and negative residues ($Ddx4^{CS}$, without modifying the amino acid composition and net charge) turned out to suppress the LLPS of this construct [97]. RPA simulations

(carried out in salt-free conditions) of phase diagrams of Ddx4^{N1} and Ddx4^{CS} confirmed the trend, with a three-fold decrease in the simulated critical temperature T^*_{cr} observed for Ddx4^{CS}. Considering that T^* is the highest temperature at which phase separation occurs, simulations of [113] indicated that a weaker LLPS propensity is expected for the scrambled mutant. Such an agreement constituted a promising premise towards the employment of RPA approach in rationalizing the correlation between charge pattern and conditions at which phase separation occurs. Note that this kind of goal would have been previously unfeasible/unattainable through mean-field theories. An attempt to elaborate an all-embracing computational theory for charge distribution impact on LLPS has been performed by Lin and Chan applying RPA to the Das and Pappu's set of 30 KE sequences, designed to sample the κ value interval from 0 to 1 [27], in order to determine their phase diagrams [41]. Interestingly, the T^*_{cr} trend followed largely the κ value one, leading to the conclusion that the increase in charge blockiness ($\kappa \to 1$) associates with a major critical temperature, hence with improved phase separating properties. Since κ is also correlated with the IDP *Rg*, a functional relationship between T^*_{cr} and *Rg* was investigated, leading to the approximate power law $T^*_{cr} \approx 9.8 \times 10^7 (Rg)^{-5.83}$, with *Rg* expressed in Å. This power-law effectively allowed to connect single-chain properties, as the chain conformation (dictated for the IDPs by their charge content and patterning), with properties pertaining multiple-chain systems, as the phase separation propensity. A consistent yet even smoother correlation was observed between T^*_{cr} and the SCD parameter [41], probably due to SCD accounting for long-range residue-interactions [40]. Lin-Chan's model, thus, indicates that the intense electrostatics between sufficiently long stretches of oppositely charged residues ($\kappa \to 1$) may promote both intra-chain and inter-chain interactions, favoring conformational compaction and tendency to phase separate respectively. On the contrary, regularly alternating

opposite charges ($\kappa \rightarrow 0$) mutually neutralize, resulting in expanded conformations and a reduced propensity towards condensation (**Figure 7**).

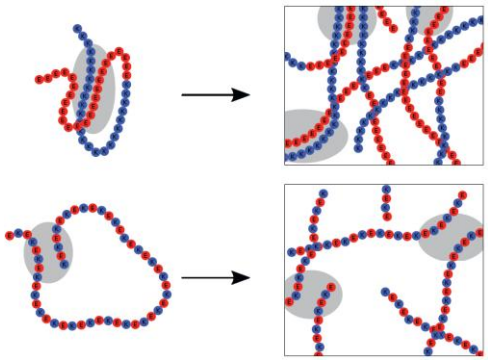


Figure 7. Schematic representation of the correlation between charge pattern and phase separation. (**Top**) Long clusters of like charges promote intra-chain and inter-chain contacts (grey areas). (**Bottom**) Interspersed oppositely charged residues form weak electrostatic networking, resulting in loose conformations and reduced intermolecular contacts. Blue spheres represent lysine (K), red ones glutamic acid (E). Figure from [41].

In an attempt to exceed the limitations inevitably associated with analytical theories, such as the RPA approach, a coarse-grained chain model using a simple lattice grid was applied to further discern the dependence of LLPS on the charge distribution of IDPs [119]. Coarse-grain models assume various levels of simplified polypeptide chain representation [120] – for instance, one or two united atoms collectively are assumed to describe a single residue side-chain– permitting explicit-chain simulations of phase separation to be computationally tractable (*Das et al., 2018*). Moreover, the employment of a lattice model permits a significant computational acceleration with respect to continuum-coarse grained models [121, 122], which however – together with atomistic simulations – have been more and more employed due to their greater structural and energetic resolution [119].

For instance, this simple lattice model was applied to two sequences of the Das and Pappu's data set [27], characterized by significantly different charge patterns ($\kappa =$

0.0009 and κ = 0.1354; *SCD* = -0.41 and *SCD* = -4.35), in order to simulate their phase diagrams. In this case, the simulated T^*_{cr} was found to be the highest for the blocky polymer, consistent with the RPA model for charge-dependent phase behavior [41].

Despite being in qualitative agreement, analytical and explicit-chain simulations diverge significantly in quantitative terms. For instance, critical concentrations simulated by RPA were appreciably lower than those obtained by explicit-chain modelling. Quantitative divergences between the two approaches are especially evident when considering the relationship between T^*_{cr} and κ, and less dramatic when considering the relationship between T^*_{cr} and -*SCD* instead [41]. For instance, the ratio between T^*_{cr} of high κ/low κ sequence is 1.9 for explicit-chain simulations, whereas it is approximately 12 in RPA simulations [119]. This comparison suggests that precautions need to be taken before quantitatively interpreting RPA and analytical theory predictions. Nonetheless, the consistency in the simulated trend strengthens the employment of RPA for reliable qualitative modelling of IDP phase behaviors.

Successive publications employing in continuum explicit-chain modeling have further questioned the κ/-SCD correlation with T^*_{cr}. Indeed, these simulations for sequences belonging to Das-Pappu's dataset confirmed the general trend predicted by RPA and lattice models [119]. However, RPA proved to over-estimate LLPS propensity for low κ (and low –SCD) sequences, with respect to continuum explicit-chain. Similarly, lattice model introduced some limitations, due to the spatial order invoked by the lattice grid [119]. Besides that, Lin and Chan's model qualitatively holds, as endorsed by another recent publication by Hazra and Levy employing explicit-chain coarse-grain simulations [123].

However, does this computational framework properly describe LLPS driven by charge pattern? Could one really suppose that increasing κ (namely charge pattern) to its maximum theorical extent will lead to a properly liquid condensation? Of

course, the aforementioned simulations are run for polyampholytes and not realistic polypeptide chains. Thus, the applicability of Lin and Chan's model to IDPs/IDRs surely requires further considerations. In this respect, the review article included in the *Appendix* section of this thesis proposed the existence of a "compaction threshold", consisting in an upper value of κ beyond which the enrichment of charged patches would rather lead to a disproportion of intramolecular interactions over the interchain ones or to a solid condensate where networking has lost his transience and weakness [34]. This hypothesis was tested on a restricted yet significant dataset of sequences from PhasePro database [124], a curated repository of experimentally validated LLPS drivers. Within PhasePro, 28 entries were selected, corresponding to IDRs for which an electrostatically driven phase separation has been observed. Then a κ-value distribution, employing CIDER webserver [125], was extracted. Each frequency class was compared with that of IDRs from the DisProt database. As shown in **Figure 8**, the frequency of sequences in PhasePro with $0.2 < \kappa < 0.25$ is larger compared to DisProt. On the contrary, the frequency of sequences with $0.25 < \kappa < 0.3$ is less than half in PhasePro with respect to DisProt. These evidences may indicate that IDPs/IDRs undergoing electrostatically driven LLPS display κ values mostly comprised between 0.2 and 0.25 [34]. Such a degree of patterning probably allows better excluding the solvent and favoring interchain interactions. Nevertheless, the limited size of PhasePro entries considered points to the need of addressing this issue in a more systematic way, so as to confirm the existence of an optimal κ value promoting LLPS.

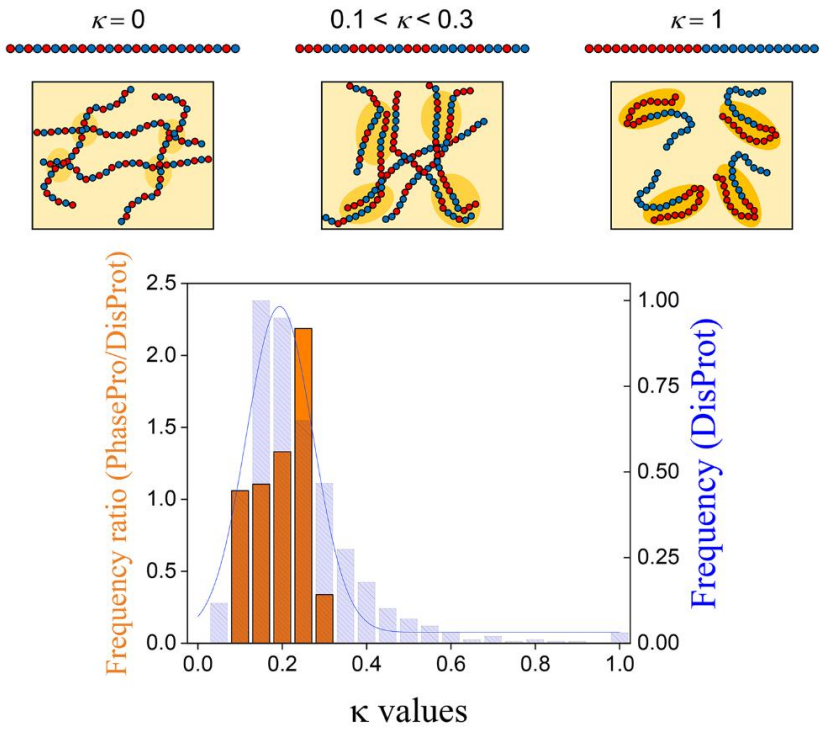


Figure 8. Charge pattern impacts on electrostatically-driven LLPS. **(Top)** Lysine-Glutamic acid (KE) sequences (K in blue, E in red) with different κ value classes. κ = 0: cooperative intermolecular electrostatic interactions between strictly alternating opposite charges are weak (pale yellow areas, inspired by *Lin and Chan, 2017*) and no condensation occurs. 0.1 < κ < 0.3: more blocky charge distribution promotes protein networking (yellow areas). κ = 1: fully blocky patterning favors intramolecular interactions which out-compete interchain ones (dark yellow areas). **(Bottom)** The orange histogram shows the ratio between the κ value frequency classes of 28 PhasePro entries and the ones from DisProt database (left-vertical axis). The shadowed blue bar histogram (right-vertical axis) shows the κ value frequency classes from DisProt. Sequences from PhasePro database were manually retrieved and analyzed for their level of disorder by IUPred (only regions with an overall disorder level of 0.6 in a scale 0-1 were employed to compute κ). κ values were calculated through CIDER webserver. DisProt entries were filtered by

discarding sequences shorter than 20 residues or devoid of charged residues (FCR = 0) and redundant sequences (sequences with same DisProt_ID and if starting residues and ending ones of the two compared regions were respectively comparable). Figure adapted from [34].

This hypothesis is in line with recent computational simulations concerning the impact of charge pattern on the structure and dynamics of polyampholyte condensates [123]. An experimental paper supported these findings as well, showing that the cluster of too many sticky elements may hamper protein condensation [110]. Martin and collaborators showed that a uniform, nonrandom pattern of aromatic stickers is conserved along the PrLD linear sequences of RBPs such as FUS, TAF15, EWSR1, hnRNPA2B1, and hnRNPA3 [110]. Increased charge separation unbalances the LLPS-aggregation interplay toward amorphous aggregation, confirming the relevance of spacers in modulating interchain interactions and therefore phase behaviour.

1.6 The threat of aberrant phase separation

It stands out clearly that MLOs control plenty of physiological functions and that genetic or environmental aberrations on them weight heavily on cellular health. As a matter of fact, aberrant condensates are associated with loss-of-function/gain-of-function diseases, leading to neurodegeneration, inflammation and cancer. The disorders could occur on different levels, (*i)* assembly of the MLO, (*ii*) ectopic localization of the condensate, (*iii*) alteration of the material properties and (*iv*) of protein composition [126] (**Figure 9**). For instance, mutations affecting the valency of scaffolding proteins and their sticker properties could alter the intermolecular networking, affecting the size/number and morphology of the condensates. Besides, aberrations occurring on protein solubility could modify the C_{sat} value possibly reducing it and therefore promoting anomalous condensation. Mutations could concern client proteins as well, determining changes in the partition coefficient [126]. In addition, the mislocalization of a protein in a different compartment (such as a nuclear protein in the cytoplasm or *viceversa*) could result in its abnormal responsiveness to environmental conditions and produce aberrant condensation.

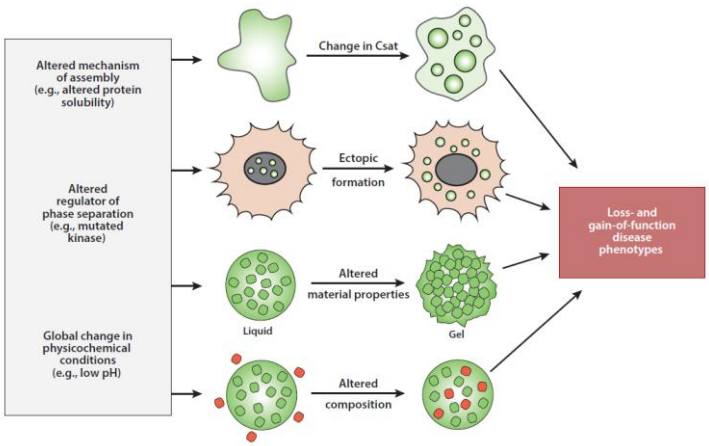


Figure 9. Possible disease phenotypes due to aberrant phase separation. MLOs could be impaired at the assembly-level (*e.g.* changes in scaffold protein C_{sat}), in

the localization (*e.g.* ectopic formation in a different cellular compartment), regulation (*e.g.* alterations in regulatory kinases), material state (*e.g.* liquid-to-solid transitions) and composition (*e.g.* recruitment of an aberrant client). Figure from [126].

Alterations of the material state of MLOs pose a major threat. By their nature, the majority of these condensates forms and dissolves transiently, in a stimulus-sensitive manner. As a matter of fact, their stabilization often compromises MLO homeostasis, leading to detrimental outcomes, such as the formation of amyloid-like fibers [98, 127, 128]. The development of numerous neurodegenerative pathologies, including amyotrophic lateral sclerosis (ALS), Alzheimer's disease, and frontotemporal dementia (FTD), is associated with this type of aberrant maturation. Commonly, proteins undergoing this phenomenon display highly hydrophobic regions and convert easily into β-sheets [129]. Therefore, slight changes in aggregation propensity are sufficient to tip the LLPS/aggregation equilibrium [129]. For instance, the nuclear protein TDP-43 constitutes the main protein that accumulates in the spinal cords of ALS and FTD patients [130, 131]. Mutations in TDP-43 are associated with its depletion from the nucleus, formation of toxic cytoplasmic aggregates, and deregulation of nuclear RNA processing. [132]. Most TDP-43 ALS-related mutations occurs in the C-terminal glycine-rich region of the protein. Mutations Q331K, M337V, Q343R, N345K, R361S, and N390D have been demonstrated to enhance TDP-43 aggregation *in vitro* and cytotoxicity in yeast cells [133]. Moreover, it has been reported that A315T isoform forms amyloid fibrils *in vitro* and causes neuronal death when added to cultured neurons [132, 134]. In general terms, aberrations in the prion-like C-terminal domain – which normally facilitates the phase separation of TDP-43 to form SGs and RNA granules – compromises the dynamics and disassembly of the granules, leading to the aggregation of the protein [135, 136]. Similarly to TDP-43, fibrillization-prone mutations have been identified in other proteins that undergo

phase separation, such as the D262V mutation in *h*nRNPA1 [98, 137]. However, one should keep in mind that irreversible amyloid-like fibers are not by definition the spontaneous evolution of LLPS. Indeed, in addition to fibrillation, numerous amorphous aggregation pathways may be accessible besides fibrillization. For instance, FUS proved to form irregular aggregates [138] and TDP-43 to originate oligomers [139], tufted particles [140] and granulo-filamentous aggregates [99]. It has to be clarified whether intermediate stages are required from matured MLOs to reach fibrillization. It has been suggested that "labile fibrils" may form reversibly, [141] being subsequently replaced by more stable fibrillar aggregates [127]. It remains an open question whether fibrillar aggregation can be considered an off-pathway step of droplet maturation or whether LLPS is a necessary precursor for this pathological state.

Interestingly, droplet maturation could also have a physiological outcome, when covering pathways alternative to amyloid-like fibers. This is the case of *oskar* granules, RNA-protein condensates found in *Drosophila melanogaster* oocytes which behave as solid *in vitro* and *in vivo* [142]. *Oskar* RBP assembly involve a rapid transition from liquid to glassy solid as a result of the entanglement, where chains wrap around each other and cannot cross [67], rather than fibrillization. Interestingly, glass states are easy to fluidize again and therefore remain responsive to changes in the condensate composition and the environment.

Relying intensively on biomolecular condensation and being so dramatic the consequences of its aberration, cells resort to three main control mechanisms to monitor MLOs. To begin with, membrane surfaces have emerged as regulatory platform to control condensate assembly [143]. For instance, evidences suggested that membranes may reduce protein threshold concentration [88, 144] by restricting molecular diffusion to a two-dimensional plane, promoting LLPS when needed. Moreover, membrane surfaces offer a spatio-temporal control over MLO assembly throughout the cell. For instance, it has been shown that the complex composed of the linker for the activation of T-cell (LAT)-Grb2-Son of Sevenless (SOS)

undergoes LLPS on specific districts of lipid-bilayers where the formation of supramolecular condensation nuclei is initiated [145-147]. In the cell, LAT protein is phosphorylated upon T-cell receptor activation, recruiting Grb2 and SOS. The latter undergoes a conformational change promoting Ras activation and the amplification of the signal cascade [148]. Condensation is thus circumscribed in space (T-cell membranes) and time (in response to the activation stimulus), in order to maximize transduction when needed. Endomembrane surfaces exert similar control over LLPS. An example is offered by endoplasmatic reticulum (ER) membranes. TIS11B is a RBP that undergoes LLPS in close proximity to ER membranes, forming a sub-cellular compartment that facilitates protein-protein interactions required for protein trafficking to cellular surface [149]. Similar spatial control mechanisms are put in place to regulate LLPS at specific interfaces of ER-Golgi intermediate compartments [150, 151].

An additional mechanism of LLPS control is offered by PTMs which effectively modulate valency and strength of protein intermolecular interactions [152]. Phosphorylation is a reversible PTM that allows a quick tuning of phase separation properties. Depending on its impact on overall net charge, phosphorylation may positively or negatively modulate condensation in a protein-specific manner. For instance, it disrupts the electrostatic networking required for the LLPS of FUS and CPEB4 [138, 153]. On the contrary, phosphorylation of Tau alters its net charge and charge patterning, enhancing the protein condensation properties and leading to disease-related aggregated forms [154, 155].

Interestingly, on the same residues targeted by phosphorylation (serine and threonine), a concurrent PTM – the O-linked N-acetylglucosamynation (O-Glc-NAc) - contributes to LLPS modulation [143]. It remains to be understood how these PTMs interact when they occur on the same protein target and whether larger sugar modifications are able to sterically hinder intermolecular interactions.

Plenty of other PTMs have been reported to modulate LLPS, such as arginine methylation and citrullination, both inhibiting cation-π interactions. The former has

been reported for FUS [156, 157], Ddx4 [97] and hnRNPA2 [158]. Citrullination has been proposed to counteract pathological aggregation of FUS and several ALS-related proteins [159]. A similar role has been suggested for lysine acetylation of Tau protein which inhibits its phase separation and aggregation [160, 161]. Nonetheless, that seems to be a protein-specific effect, since acetylation of TDP-43 in the RNA-binding domain promotes the formation of aggregates [162]. All things considered, finding an all-embracing rule for PTM regulation appears hard. One may simply infer that the effect of any PTM needs to be evaluated in the specific context of the target protein.

To complete the overview of LLPS control mechanisms, it is necessary to mention the energy-consuming strategies. They mainly include chaperones and helicases, which remodel MLOs by regulating their material state and dynamics. For instance, chaperones localize within SGs and regulate their material properties and disassembly during recovery from stress [143]. A recent publication showed that ALS-related SOD1 variants accumulate in SGs as misfolded proteins, increasing SG viscosity and causing an aberrant liquid-to-solid transition [163]. Chaperones actively prevent aberrant SGs and promote their clearance after stress [163]. Similar to chaperones, DEAD-box helicases can localize to stress granules, regulating their dynamics and dissolution [94, 164].

1.7 The LLPS from the RNA side

RNA molecules can either undergo condensation *per se*, in the absence of protein components, or form complexes with RBPs. Well-known phase-separating proteins such as FUS, TDP-43, and hnRNPA1 are indeed RBPs [98, 165]. Nonetheless, RNA is often required for the proper formation of MLOs even when the proteins involved are not directly participating in RNA metabolism or do not include RBPs. Instead, examples of RNA homotypic condensation is given by RNAs that contain triplet expansion and are associated with neurological disorders (Huntington's and ALS [166]). Long repeats promote formation of foci, which may exhibit liquid-like behavior *in vivo* [167-169].

Why the RNA? Apparently, RNA structure, sequence, and length collectively tune MLO assembly. For instance, the secondary structure of mRNA has been reported to modulate the identity of the droplets formed by Whi3 [170]. Whi3 forms indeed condensates with distinct composition and function depending on whether it includes cyclin CLN3 or actin BNI1 mRNAs, which are structurally different and for this reason mutually exclusive in the binding. A recent publication suggested that highly structured RNAs act as scaffolds for condensate nucleation, particularly in the nuclear environment, which is characterized by high RNA concentrations. This is the case of structured long non-coding RNA Neat1, which promotes condensation of FUS in the presence of a high background concentration of RNA, which normally solubilizes FUS condensates [171]. On the contrary, the complex structure of rRNAs (and not their composition) was demonstrated to attenuate the *in-vitro* condensation of PGL-3 protein [172]. Consistent evidences on RNA structure effect on the LLPS behaviour were also found in other protein-RNA condensates such as those involving poly(proline – arginine) (PR_{30}). This type of protein has been studied *in vitro* with homopolymeric RNA consisting of poly-rG/rU/rA/rC. Interestingly, PR_{30}-RNA condensates take profoundly different forms depending on the type of nucleic acid employed. While poly-rU/rA/rC mixed with

PR_{30} produce spherical assemblies, poly-rG form open, fractal-like network structure reminiscent of kinetically arrested phase separation [173]. This different morphology could depend on the propensity of poly-rG to form stable G-quadruplexes [174]. Having reported conflicting evidences of structured RNA impact on condensate behaviour, one may conclude that there is not a general rule, yet a protein-specific effect.

The same PR_{30} experimental system has also allowed studying the effects of RNA composition, which seem to relate mainly to condensate viscosity and protein dynamics. In particular, PR_{30} condensates with poly-rA were found to be more viscous than those formed with poly-rC and poly-rU [175]. This can be explained by considering that poly-adenine sequences, with their purine bases, promote stiffer conformations than those involving poly-pyrimidine molecules, which are endowed with a single aromatic ring [166].

Finally, RNA length and concentration were found to collectively modulate the LLPS in the cell, defining the total number of potential RNA binding sites [175]. Longer polyribonucleotide sequences trigger multiple interactions, favoring phase separation. This is the case of PGL-3 condensates, which form more efficiently with mRNAs longer than 0.5 kb [172]. The RNA binding footprint for most RBPs is 10-20 nucleotides [176], suggesting that 20-40 nucleotides is the minimal length to guarantee multivalency [175].

Concerning the RNA concentration, Martin's group proved for several RBPs (FUS, TDP43, EWSR1, TAF15, and hnRNPA1) that high concentration ratios of RNA: protein prevent phase separation, while low ratios promote it [171]. Thus, nuclear RNAs that are not-specifically interacting provide a buffering control over RBP condensation, while specific RNA interactors, which bind RBP with high affinity, promote LLPS only when necessary (such as in the case of Neat1 RNA with FUS). Supporting this interpretation is the observation that disease-related, aberrant RBP aggregation occurs in the cytoplasm, where the RNA concentration is not sufficient to efficiently suppress LLPS [171].

The system of PR_{30} and homopolymeric RNA is also useful in studying rheological properties of condensates. Indeed, a "hardening" effect of adding poly-rU RNA on PEG-induced preformed droplets has been observed, showing that increasing the density at interaction sites negatively affects the coacervate dynamics [173]. However, increasing the poly-rU concentration in the PR_{30} solution has no impact on condensate dynamics, since it probably does not produce a preferential recruitment of RNA within the droplets, yet a simple effect on the number of PR_{30}-RNA droplets [173, 177]. A different effect has been observed for LAF-1 condensates [96]. Indeed, low concentrations of $polyU_{50}$ added to the protein solution caused a reduction in viscosity, probably due to the more dynamic protein-RNA interactions replacing the stiffer and more stable IDP-IDP contacts. Overall, the results reported here suggest that RNA differently affects the rheological properties of RNA-protein condensates, dependently on the protein type and the timing by which it is added to the system.

1.8 Phase separation of viral IDPs/IDRs

Due to the transversal distribution of IDRs in proteomes, there is no surprise that LLPS has been largely exploited even by viruses to maximize genome replication or viral particle assembly [178]. Indeed, phase separation underlies the formation of the so-called “viral factories”, such as the inclusion bodies (IBs), that are sites where specific viral and host proteins concentrate, in order to facilitate the replication and shield the viral machinery from the host defenses (**Figure 10**). Actually, membraneless viral factories are one of the possible bodies observable in infected cells, generally more common for negative-strand RNA viruses. Indeed, viral factories can be membrane-delimited as well, including double-membrane vesicles (DMVs) or spherules, namely membrane invaginations [179].

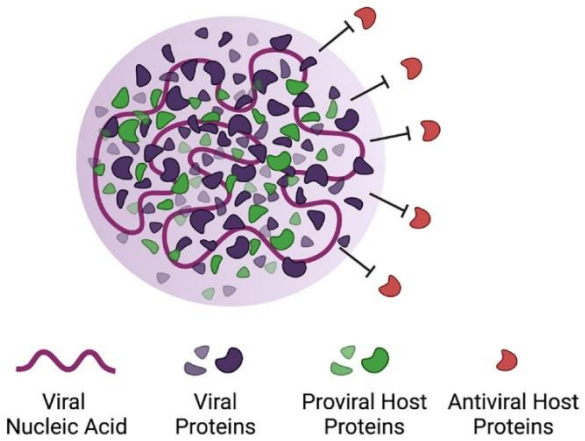


Figure 10. Membraneless viral factories concentrate viral nucleic acids, viral proteins and endogenous proviral proteins, protecting the replication and assembly apparatus from factors with antiviral activity. Figure from [179].

Among the first reported, the liquid-like inclusions known as Negri bodies (NBs) have to be mentioned [180]. They form in neuron cytoplasm upon rabies virus (RABV) infection, serving as platforms for the enrichment of viral transcription or

replication factors. Indeed, the phosphoprotein P (P protein) – one of the components of these inclusion bodies – recruits its cytoplasmatic partners endowed with proviral activity, as the focal adhesion kinase (FAK) and the major inducible heat shock protein 70 kDa (HSP70) [181, 182]. Similar bodies have been reported for other *Mononegalovirales*, as the vescicular stomatitis virus (VSV) [183] and Measles virus (MeV) [184]. Plenty of inclusions characterizing several paramyxoviruses (*Mononegalovirales* order) – as human parainfluenza virus 3 (hPIV3) [185], parainfluenza virus 5 (PIV5)[186], mumps virus (MuV)[186], Nipah virus, and simian virus 5 (SV5) [187] – have not still been univocally associated with LLPS, although the spherical morphology of droplets and the ability to coalescence support this interpretation. Therefore, it is predicted that many of these inclusions, initially identified as "viral aggregates", will be re-interpreted in the light of phase separation. In agreement with this, P and N proteins (main components of the mentioned inclusions) are characterized by structural disorder and multivalency, all desired pre-requites for LLPS.

Other examples have been found among viruses relevant to human health. In the case of influenza A virus (IAV), viral inclusions are formed that display liquid-like properties and are employed for the assembly of the eight-partite genome [188]. Considering human immunodeficiency virus (HIV), the nucleocapsid (NC) protein shows Zn^{2+}-dependent LLPS [189]. Furthermore, it was recently shown that the N protein from SARS-CoV-2 undergoes phase separation, perhaps for the purpose of promoting nucleocapsid formation, as in *Mononegalovirales* inclusion bodies [190]. The capability of SARS-CoV-2 N protein to partition into liquid compartments formed by host proteins as hnRNPA2, FUS, and TDP43 suggests that the viral protein is able to trigger SG assembly, favoring viral replication [191].

Indeed, besides assisting the infection stages, LLPS is exploited by viruses to hijack the host immune response, sequestering cellular sensors that detect pathogen associated molecular patterns (PAMPs) or proteins responsible for the activation of

the innate humoral response [178]. For instance, several viruses have developed strategies to interfere with SGs, which may assemble upon viral infection [179], for instance by sequestering their main components. This is the case of West Nile virus, whose 3'stem-loop structure of RNA genome interacts with two SG proteins, TIA1 and TIAR, thus preventing stress granule assembly. Furthermore, this interaction facilitates the synthesis of viral genomic RNA [192]. Another example is offered by RSV IBs, which proved to sequester *O*-GlcNAc transferase (OGT) and phosphorylated mitogen-activated protein kinase (MAPK)/p38. Because OGT is responsible for attaching *O*-GlcNAc to cellular proteins, which enables their translocation into SGs [193], its sequestration by RSV is sufficient to impair SG assembly [194]. Furthermore, transcriptional activity of NF-κB is of pivotal importance in developing innate immunity, especially when its activity is regulated by the association with MAPK/p38 and p65 subunits. The retention of both these subunits into condensates blocks the translocation of NF-κB to the nucleus, impairing the mechanism that triggers the innate immunity [195]. In addition, the further sequestration of melanoma differentiation-associated protein 5 (MDA5) and mitochondrial antiviral-signaling protein (MAVS) into IBs of RSV impairs the antiviral interferon response [196, 197].

Interestingly, viral condensates are also able to interact with specific host genes, tuning the trascriptome of the host cells in favor of the virus. For instance, Epstein-Barr virus (EBV) – a DNA virus – exploits its transcription regulators EBNA2 and EBNALP to form liquid-like droplets in the nuclei of the infected cells. They recruit other co-activators and transcription factors at the sites of super-enhancers of MYC and Runx3, promoting their transcription.

All things considered, it is not surprising that plenty of studies are focusing on the condensation properties of viral proteins. These studies could help finding innovative antiviral therapies. To cite a few, Risso-Ballester and collaborators identified a compound – namely a steroidal alkaloid cyclopamine and its chemical analogue (A3E) – capable of hardening RSV IBs, where RNA replication occurs.

A3E compound could suppress RSV infection in mice [198], demonstrating that rheological properties of condensates are related to virus infectivity. In another paper, Wang and colleagues tested a peptide, identified by a screening approach, which targeted the dimerization domain (DD) of SARS-CoV-2 N protein [199]. DD is strictly necessary to promote LLPS of N protein. The peptide is able to suppress the condensation, increasing the antiviral response both *in vitro* and in mice. Together, these evidences support the potentialities of antiviral therapies targeting LLPS. This represents a novel route for drug design and hopefully it will gain more and more relevance in treating virus infections.

1.9 Methods

Here, a brief overview of methodologies employed in this thesis is presented. A distinction is made between the techniques used for the structural characterisation of IDPs and methods applied to properly assess the LLPS behaviour. The two subchapters reflect this distinction.

Techniques for structural determination of IDPs

Size-exclusion chromatography (SEC)

SEC, also named as molecular-exclusion chromatography, gel-filtration or gel-permeation chromatography (GFC or GPC), represents a precious laboratory technique employed for the separation of molecules based on their hydrodynamic dimensions (R_h) or Stokes' radius (**Figure 11**)[200].

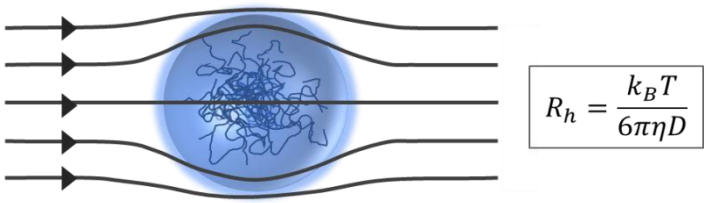


Figure 11. Schematic representation and definition of R_h. Figure from [34].

The chromatographic separation is achieved employing a stationary phase with beads endowed with a defined range of pore dimensions, which is correlated to the column cut-off. If molecules suspended in the mobile phase have smaller hydrodynamic dimensions than pore size, they are able to fit into the pores by force of diffusion and thus they pass through the column with a larger retention time. Conversely, bigger molecules cannot fit into the pores, thus simply follow the solvent flow and exit the column with a smaller retention time. For molecules with a comparable shape, R_h is proportional to molecular mass [200]. Therefore, SEC is largely employed to determine the approximate size and molecular weight of

proteins. The proper column calibration is a necessary pre-requisite for deriving quantitative information on molecular mass from SEC measurements. Indeed, the correlation between the column retention capabilities and the hydrodynamic dimensions of protein standards has to be robustly determined. The retention of a given molecule by a chromatographic column is represented by the column partition coefficient, K_D, calculated through the Eq. 5 [201, 202].

Eq. 5 $$K_D = \frac{V_{EL} - V_O}{V_T - V_o}$$

where V_{EL} is the retention volume of the molecule (namely the volume at which it is eluted by the column), V_O is the void volume (that is the volume of mobile phase between the stationary phase beads) and V_T is the total volume (the volume of all the liquid within the column).

The K_D is then simply correlated to the known hydrodynamic dimensions of the standards, thus creating the calibration curve (R_h *vs.* K_D). By doing this, it is possible to determine the hydrodynamic dimensions and the molecular mass of any protein sample, simply comparing the retention time of the unknown protein with data obtained for the standards whose molecular weight is known, provided that they belong to the same conformational class. Indeed, SEC is able to highlight the different conformational states a protein assumes and R_h has been shown to vary accordingly to the conformational compactness of the protein. As a matter of fact, it has been observed that the hydrodynamic volume of a single-domain globular protein in the molten globule, pre-molten globule and unfolded state increases of 1.5, 3 and 12 times compared to native state, respectively [16, 203, 204]. Consequently, an empirical set of equations has been derived, correlating the Stokes' radius with the apparent molecular mass of globular proteins in different conformational states [205]. Thus, being the molecular mass dependency on Stokes' radius extremely conformation-specific, based on the hydrodynamic

dimensions evaluated for a generic IDP it is possible to assign it to a specific structural class. Indeed, it is sufficient to compare the measured R_h with the theoretical values obtained with these equations for a polypeptide chain of proper molecular mass. This has opened the doors to IDP conformational characterization, helping identifying the presence of elements of residual structure. For instance, this approach was applied to assess the different conformational compaction properties of α-, β-, and ϒ-synuclein, which are predicted as fully disordered proteins. As a matter of fact, the experimentally-determined R_h of the three proteins were compared to the theorical Stokes' radius calculated for a completely unfolded polypeptide of comparable molecular mass. While β-synuclein displayed a similar value, suggesting an extended conformation, α- and ϒ-synuclein revealed more compact. Interestingly, employing the set of equations cited above, it was possible to verify that at acidic pH the three proteins assumed pre-molten globule-like conformations, implying a pH-induced ordering [206].

In parallel with the determination of R_h through SEC, in the attempt to produce reliable R_h predictions, Marsh and Forman-Kay introduced a simple power-law relation (Eq. 6), which proved effective for both folded and chemically denatured proteins [49].

Eq. 6 $$R = R_0 N^{\nu}$$

where R is R_h or R_g, N is the number of residues, R_0 and ν are constants [207, 208]. IDPs, though, display a greater variation in compaction, ascribable to different sequence features, which prevent the use of Eq.6 for reliable R_h predictions. An extension to this equation was thus introduced, accounting for the proline number P_{PRO}, the absolute net charge (|Q|), and the presence of a polyhistidine tag (S_{his}; Eq.7). The employment of this equation further improved the performances of Eq.6 when applied to IDPs.

Eq. 7 $$R_h = (AP_{PRO} + B)(C|Q| + D)S_{his} * R_0 N^{\nu}$$

with A = 1.24, B = 0.904; C = 0.00759; D = 0.963; S_{his} = 0.901; R_0 = 2.49; ν = 0.509 [49].

Overall, it is evident that sequence determinants are particularly decisive for the overall size of IDPs, as emerges from Eq.7. Experimental determination of R_h by SEC and comparison with reliable theoretical references represent an extremely informative approach to determine the conformational properties of IDPs, combined with the advantage of being based on a widespread, inexpensive technique that is accessible to most laboratories.

Native mass spectrometry

With the advent of electrospray ionization (ESI) and matrix assisted laser desorption/ionization (MALDI), mass spectrometry (MS) established as a solid technique for the structural determination of proteins and protein complexes [209]. Subsequent development of gentle ionization sources (such as nano-ESI) made it possible to preserve non-covalent interactions in the vanishing-solvent conditions of the electrospray, resulting in an effective ionization and transfer to the gas phase of the protein in a native-like state (*i.e.* native mass spectrometry, native-MS). This encouraged the use of the technique for conformational characterization studies of IDPs.

Briefly, in a nano-ESI experiment, a dilute solution of the analyte is pumped at low speed through a capillary. A high voltage (positive or negative) is then applied to the capillary outlet, which provides the electrical field required to produce charge separation at the surface of the solution. The liquid flowing out of the needle produces the so-called "Taylor cone". When the surface tension fails to balance the electrostatic repulsions (the "Rayleigh limit") [210], the droplets with an excess of similar charges detach from the cone and move towards the inlet of the mass spectrometer (**Figure 12**) **[**211**]**. The transfer of the analyte to the gas phase can occur through two mechanisms, the "charged residue model" (CRM) or the "ion

evaporation model". In the former case, the evaporation-dissociation cycle proceeds until an ion is released in the gas-phase. Indeed, the gradual evaporation of the solvent leads to an increase in charge density, causing the progressive fission of larger droplets into smaller ones [212]. In the latter, instead, the higher charge density causes coulombic repulsions that overcome surface tension and result in the extrusion of ions from the surface [213]. After desolvation and ionization, the ions are transferred to the mass analyzer (*e.g.* Orbitrap, quadruple, ion trap) and sorted according to mass/charge ratio (*m/z*). Proteins, especially IDPs, produce in nano-ESI experiments a series of bell-shaped peaks, namely the "charge-state distribution" (CSD) [214-217].

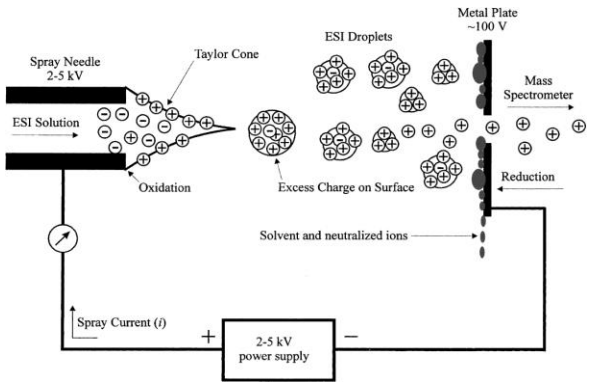


Figure 12. Schematic representation of the working mechanism of ESI. Figure from [211].

The CSD of a given protein analyte is invariably affected by instrumental settings, solvent properties [218], and, above all, by the global protein compactness at the moment of gas-phase transfer [217]. Indeed, conformationally heterogeneous samples, such as the IDP ones, result in multimodal distributions which pinpoint the presence of coexisting conformers or, more properly, of different conformational subpopulations each internally displaying similar compaction properties. Conversely, folded proteins display narrower CSD, due to their

structural homogeneity, reaching higher *m/z* (thus lower charge state). The reason for this is that CSD is governed by the availability of surface-accessible ionizable residues. Proteins characterized by secondary and tertiary structure elements normally exhibit a finite number of protonation sites and therefore assume low-charge states during ionization. The conformational ensemble of IDPs, on the other hand, is composed of a mixed population of conformers with varying numbers of accessible ionization sites, which attain low- and high-charge states. By transforming the abscissa of the MS spectrum from *m/z* to *z*, the Gaussian fitting of the CSD allows these different conformational components to be identified and their compaction characterized. Indeed, from ESI-MS experiments, it is possible to deduce quantitative information on the "solvent accessible surface area" (SASA) [215, 219, 220] which, in turn, reflects the compaction state of the analyzed protein. Initially, the power-law correlation between CSD and SASA has been highlighted for globular proteins, leading to $\log(Z_{av})$ *vs.* log(SASA) plots. The original equation [221] has been updated several times [220], as the available datasets have increased, leading to Eq.8 as the final form [215].

Eq. 8 $y = 0.60x - 3.28$

A similar linear correlation has been also observed for unfolded proteins, using average SASA estimated from simulated ensembles [215, 219, 222]. Eq. 9 has been shown to fit the high-charge states reached by IDPs and proteins undergoing chemical denaturation [215]. In contrast, the charge-to-mass relationship emerged as largely influenced by the conformational state of proteins, thus leading to the conclusion that molecular weight is not sufficient to determine the ionization behaviour of protein analytes [216, 219, 221].

Eq. 9 $y = 0.91x - 6.01$

SASA values derived from the CSDs can be used for the conformational classification of IDPs by means of a compaction index (*CI*). This index had been

previously defined for R_h [223], yet was reformulated for MS-derived SASA to assess the degree of protein compaction estimated by mass spectrometry and compare proteins of different sizes, both at the level of individual conformers and of the entire conformational ensemble [216]. Its value ranges from 0 to 1, with the minimal value corresponding to fully disordered conformations and the maximal one to compact structures, and is calculated as described in [216] and also reported in the *Material and Methods* of Section 2.1, in the Results of this thesis.

The comparison of ESI-MS with in-solution and computational methods for the characterization of conformational ensembles of IDPs revealed both elements of constituency and discrepancy. Considering the former, CSDs of α-synuclein in the absence of lipids revealed the presence of a predominantly extended component [224-226], in agreement with in-solution spectroscopy [227-230], single-molecule spectroscopy [224, 231, 232] and SAXS analyses [233] using the ensemble optimization method (EOM). ESI-MS proved particularly useful for characterizing metal ion-induced conformational transitions, such as in the case of UreG, which compacts upon addition of Zn^{2+}, as assessed by NMR [234], or methallothionein-2A, which compacts upon coordination with 7 Cd^{2+} ions as predicted by molecular dynamics simulations [215]. However, the accuracy of ESI-MS measurements has been recently debated. Besides point discrepancies, such as in the case of acyl-carrier protein ACP from *V.harvey*, whose MS spectra revealed a conformation completely distant from the one in solution, an interesting work questioned the reliability of ESI-MS with a systematic approach, comparing mass spectrometry with SAXS-EOM on a set of 7 IDPs, reporting serious structural differences depending on the used technique. Although the authors attributed the observed discrepancies to artifacts occurring during the ionization phase, it is reasonable to question whether the SAXS-EOM approach fails to resolve all conformational components of the ensemble (whose relative abundance might be low) or native-MS actually produces some aberrations during electrospray [217]. Further investigations are required, especially considering that several other studies

comparing native-MS with SAXS-EOM come to opposite conclusions [224, 225, 233].

All things considered, despite the aforementioned limitations, native-MS emerges as a promising technique for the conformational investigation of IDPs [209, 235-237].

Fourier transform infrared (FTIR) spectroscopy

FTIR spectroscopy proved to be a robust technique for gaining insights into protein secondary structure and aggregation. Indeed, it has been successfully applied in characterizing the residual structure of IDPs, induced folding and aggregation [238]. In infra-red (IR) spectroscopy, IR radiation passes through the sample, with part of the radiation being absorbed and part transmitted. The resulting spectrum represents the absorption and transmission of molecules, defining a unique molecular fingerprint [239]. Modern IR spectrometers are usually FTIR spectrometers, which means that the signal measured by the detector is correlated to the spectrum by a Fourier transform [240]. FTIR spectrometers provide better signal-to-noise performances and are widely preferred in the mid-infrared region (4000-200 cm^{-1}), where the main protein absorption bands are located. These include the amide I (1700-1600 cm^{-1}), the amide II (1600-1500 cm^{-1}) and the amide III (1400-1200 cm^{-1}) bands. The amide I band is due to the stretching vibration of the C=O peptide bond and is particularly sensitive to the C=O environment, thus, the secondary structure of the protein. Due to the strong absorption of water between 1640-1650 cm^{-1}, most of the structure determination by amide I is performed in D_2O solutions [241]. However, the assignment of protein structure to FTIR spectra can be also performed for proteins in H_2O solution. The amide II band is the result of contributions from several different backbone modes (NH in-plain bending and CN stretching, with small contributions from C=O bending, CC and NC stretching vibrations), and shows less conformation sensitivity than the amide I counterpart [242]. Furthermore, the bands of amide II are overlapping with

those originating from side-chain vibrations [243]. Finally, the amide III band, mainly due to the in-phase combination of the in-plane bending of the NH and the stretching vibration of the CN, proved to be highly sensitive to secondary structure folding, without interference from H_2O absorption, yet much weaker than the amide I signal [241].

In the amide I band, each element of the secondary structure produces a different C=O stretching frequency, due to the unique geometry and hydrogen bonding pattern [244]. The large overlap of the underlying component bands makes amide I "featureless" and instrumentally unresolvable [244]. Consequently, mathematical methods are needed to enhance the resolution of the spectrum by narrowing down the large absorbing components in the amide I region [245]. Second-derivative analysis is one of the most frequently used methods employed for this purpose, as is also the case in this thesis work, which provides better spectral resolution since the semi-amplitude of the band is reduced by a factor 2.7 (**Figure 13**)[246]. Interestingly, in the second derivative, the peaks appear negative and their height is proportional to the original height, but inversely proportional to the square of the original half-width [247]. As a result, the sharpest peaks, such as water vapor or noise, are intensified relative to the broad components of the secondary structure elements. Therefore, in order to obtain reliable information from second derivative analysis, a high signal-to-noise ratio and instrument purging are required.

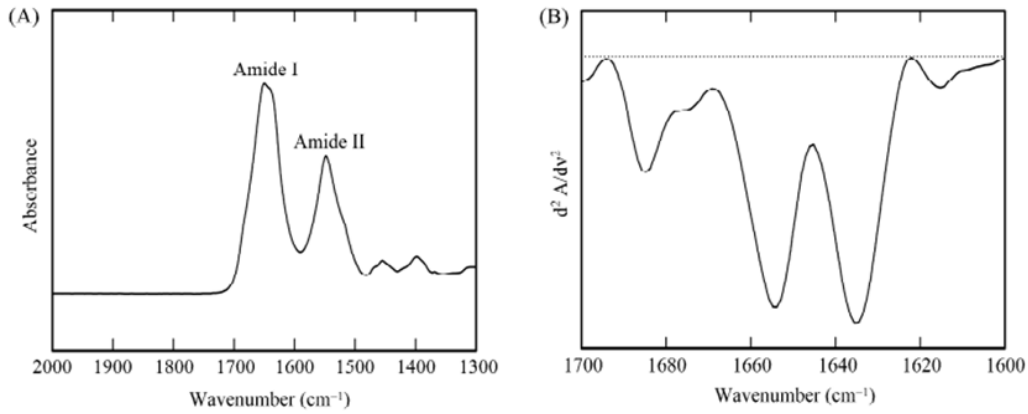


Figure 13. (A) Original FTIR spectrum in the mid-infrared region (2000-1300 cm^{-1}), showing Amide I and II absorption bands. **(B)** Second derivative spectrum. Figure adapted from [244].

After identification of the component bands within the amide I region, it is necessary to perform the complex assignment to the secondary structure elements. Since these partially overlap (as in the case of random coils and α-helices in aqueous solutions) and the positions of the peaks may vary depending on specific interactions, a critical interpretation of the assignment is required [246], also relying on orthogonal techniques as circular dichroism.

Random coil structures are characterized by a broad band centered at 1654 cm^{-1} in H_2O, where α-helices also respond (1660-1648 cm^{-1}). The use of D_2O allows better discrimination of the two elements, as the random coil shifts to 1645 cm^{-1}. For instance, a broad peak at 1645 cm^{-1} has been obtained for α-synuclein in D_2O, consistent with its completely disordered nature in the absence of lipids [246].

As for the β-sheets, the IR response is characterized by two bands, at 1633 cm^{-1} and 1686 cm^{-1} in water, both downshifted in D_2O. The band position and intensity strongly depend on the geometry of the structure, the number of strands per sheet [248], and the strength of H-bonds [249, 250]. However, the positions of the native intramolecular β-sheets are different from those of the intermolecular sheets

associated with protein aggregation [246]. The ability to discern the two has enabled the application of FTIR in protein aggregation and amyloid studies [251, 252].

Once the peaks have been assigned, it is possible to quantitatively assess the abundance of each identified secondary structure element. To this end, it is necessary to perform curve fitting of the measured spectrum, using Lorentzian or mixed Lorentzian-Gaussian functions [246]. The input parameters of the fitting (such as the baseline, the peak position, the band width and intensity) are crucial in the analysis. At least two rounds of fitting are recommended to find the best set of fitting functions [246]. The fractional area of each Lorentzian/Gaussian component compared to the total area (sum of all components) provides the percentage of each element of secondary structure.

Protein samples are often analyzed using FTIR spectroscopy in attenuated total reflection (ATR) mode, due to the ease of sampling. This technique is extremely useful, as it allows proteins to be studied in the form of films, obtained by solvent evaporation of diluted proteins directly on the ATR plate. Indeed, the sample is placed directly on the ATR element, which has a higher refractive index than the sample. The IR beam reaches the ATR element and is totally reflected, reaching the detector after one or more reflections. With each reflection, the evanescent wave of the beam penetrates the sample where it is absorbed [246], producing the absorption spectrum.

In conclusion, FTIR is an extremely powerful tool for characterizing secondary structure elements within protein samples. Although IDPs lack a stable tertiary conformation, residual structure elements are present and define the degree of compaction of these proteins. Therefore, the use of these techniques in IDP studies is extremely valuable and informative.

Techniques to assess in-vitro LLPS behaviour

Microscopy detection of in-vitro LLPS condensates

Visualization of condensates deriving from LLPS is carried out through microscopy techniques. A simple approach for *in-vitro* morphological studies consists in employing light microscopy. However, fluorescence microscopy is widely used as well, especially when imaging is performed within cellular compartments. Fluorescent labeling can be performed by Green Fluorescent Protein (GFP)-tag fusion, at the N- or C-terminus of the protein of interest, or by fluorescent dyes. Considering *in-vitro* studies, it is generally advisable to use an inverted rather than an upright microscope, as condensates tend to settle by gravity [72]. It is mandatory to identify a pre-LLPS condition, in which the protein sample is homogenous and no condensates are observed. This ensures that condensation is controllable and induced by a specific trigger, such as changes in solution pH (pH jump, as performed in this thesis), ionic strength, temperature or nucleic acid addition. Samples may require incubation time (from minutes to hours) before imaging. When comparing different samples, both incubation times and imaging parameters should be kept constant [72]. An issue to consider is the interaction of the sample with the surface of the glass slide, which can affect the material properties of the droplet. Therefore, it is advisable to compare the behaviour of condensates on slides with different coatings, such as polyethylene glycol (PEG) or lipids [72].

Microscopic observation of LLPS samples allows assessing the presence of phase-separation hallmarks, namely the roundness of the droplets, the propensity to undergo fusion/fission events with fast kinetics and the recovery of fluorescence after photobleaching, investigated through FRAP technique. All these features must be carefully verified to confirm the LLPS behaviour of the tested sample.

Fluorescence Recovery After Photobleaching (FRAP)

Fluorescence Recovery After Photobleaching (FRAP) is a microscopy-based technique performed on droplets to assess their liquidity. Depending on the material properties of the condensates, the molecules display different mobility, which has an impact on the recovery kinetics after photobleaching. Indeed, in FRAP, fluorescently-labeled molecules are photobleached with a high-intensity laser pulse within a region of interest (ROI). Photobleaching is the result of photochemical damage of the fluorescent probes, which permanently lose their fluorescence capacity [253]. The recovery of fluorescence within the ROI is monitored over time, resulting in a graph of the recovery kinetics (**Figure 14**). The principle is simple: if fluorescent molecules are able to diffuse freely, the photobleached ones will exchange with the surroundings, completely restoring the initial fluorescent signal within the ROI. Conversely, if the diffusion is slowed down by the material properties of the condensate, the original fluorescence will be only partially, if at all, restored. Therefore, the recovery graph contains information on the mobile fraction (M_f) and the recovery half-time, this latter corresponding to the moment when half of the fluorescence is recovered [254]. To calculate the mobile fraction, it is necessary to compare the fluorescence in the bleached region after full recovery (I_∞) with the fluorescence before bleaching (I_i) and immediately after bleaching (I_0; Eq. 10):

Eq.10 $$M_f = \frac{I_\infty - I_0}{I_{i-I_0}}$$

The immobile fraction, IM_f, corresponding to the population of bleached molecules that do not diffuse, is obtained as (Eq. 11):

Eq. 11 $$IM_f = 1 - M_f,$$

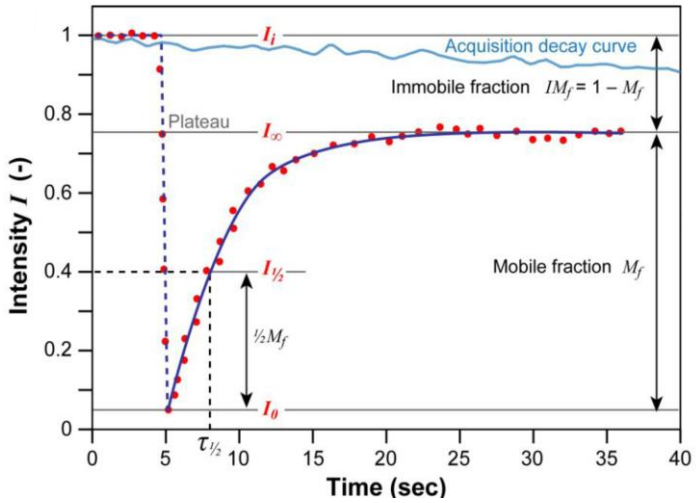


Figure 14. A typical FRAP curve. Starting from the initial maximal fluorescence (I_i), the signal drops to I_0 after photobleaching. Over time, the fluorescence signal recovers from I_0 to the maximum recovery value I_∞. The mobile fraction (M_f) and the immobile fraction (IM_f) are derived from this plot using equations 10 and 11. The light blue line corresponds to a reference photobleaching curve to correct for the loss of fluorescence outside the ROI during data acquisition. Figure adapted from [255].

FRAP is normally implemented in confocal microscopes, along with straightforward experimental protocols, which contribute to the wide application of this method to LLPS studies. As a matter of fact, FRAP has been largely applied to measure the diffusivity of fluorescently labeled protein/RNA molecules within droplets both *in vitro* and *in vivo* [93, 96, 98, 256, 257]. However, understanding FRAP results can be tricky and a critical interpretation is necessarily required. For instance, multi-component condensates may show different diffusivities according to the probe molecule analyzed [258, 259]. This depends on different intermolecular interactions and spatial heterogeneities [260], and possibly reflects the scaffold-client model. Furthermore, since the FRAP recovery curve can be

fitted with appropriate diffusion models to extract the apparent diffusion coefficients, it is important to remind that the chosen model strongly influences the extracted coefficients [261]. Thus, more cautiously, it could be argued that FRAP can provide a solid qualitative idea of the molecular diffusivity within a condensate and give valuable insights into how the material states of phase-separating droplets change over time or under different environmental conditions [93, 262, 263]. For more accurate estimates of diffusion, at the molecule level within liquid states, fluorescence correlation spectroscopy (FCS) can be used [72, 264].

Overall, FRAP proves to be a valuable technique for demonstrating the liquid state of LLPS condensates. Its versatility, both *in-vitro* and *in-vivo* experiments, and ease of application make it a widely used approach to get valuable information on diffusion kinetics within phase separating droplets.

Turbidity Measurements

LLPS condensates scatter visible light, making them detectable by optical density measurements (either at 340 nm or 600 nm). Using a simple spectrophotometer, trigger conditions for LLPS can be identified, *e.g.* by determining the C_{sat} through the onset of the scattering. Furthermore, by monitoring the optical density as a function of time, information on the kinetics course of the condensation phenomenon can be obtained. Nonetheless, these simple turbidity measurements detect a variety of assemblies. Therefore, parallel microscopy analyses are indispensable to properly assess phase separation phenomena.

2. Main results and discussion

As discussed in the introductory paragraph, besides net charge, the distribution of charged residues along the primary structure sensibly affects IDP dimensions and interactability. Nonetheless, few considerations have been made on how the sequence context of an IDP may influence its responsiveness to variations in the patterning of charges. **Section 3.1** debates this topic, verifying how IDPs with similar charge properties yet slightly different sequence backgrounds respond to variations in the pattern of cationic and anionic residues. Three model IDRs – N_{TAIL}, PNT4, and the human medium neurofilament protein (NFM) – were chosen accordingly to their sequence features. Indeed, they share a relatively high FCR (mean FCR ~ 0.33), a rather low absolute value of NCPR (mean absolute value ~ 0.04) and low κ values (mean $\kappa \sim 0.17$), yet different proline content (5.2%, 11.4% and 0.7% respectively). Each IDR has been rationally designed to produce two "κ permutants", which achieve the lowest and highest κ value accessible to the amino acid composition of each natural (wild-type) protein. The variants have been designated as "Low κ" and "High κ". The conformational properties of wild-type and κ permutants were assessed through native-MS and SEC measurements.

Concerning SEC analyses, negligible volume effects (in terms of R_h) were observed for the charge clusterisation of PNT4, namely the IDR with the largest proline content. To better elucidate the effect of charge clusterisation on the IDPs under study, a R_h-based compaction index (CI_R) was introduced. CI_R revealed a simple descriptor, useful for comparing the conformational properties of IDPs of different lengths, as in this case. This comparison revealed that the average compactness of PNT4 is not influenced by the charge distribution, *i.e.* κ value, whereas N_{TAIL} and NFM respond significantly to charge clustering.

Differently from SEC analyses, CSDs resulting from the nano-ESI process captured a general loss of conformational heterogeneity associated with the

progressive clustering of opposite charges for all three sets of κ permutants, suggesting a global effect of compaction as κ increases. The compactness of the ensembles was compared by computing the compaction index based on average SASA ($\overline{CI}_{\text{SASA}}$), which refers to the extent of protein ionization. Interestingly, the high-κ variant of PNT4 displays a unique compaction excursion which, however, is not caused by technical artefacts. Possibly, the observed discrepancy of the SEC/MS analyses lies in the different information provided by the two techniques. Indeed, R_h mainly captures conformational effects on volume, whilst SASA reflects contributions on the shape due to charge redistribution. This emerged applying a mathematical model that approximates the shape of globular proteins to ellipsoids and that was employed herein for the first time to obtain coarse-grained information on IDP ensembles. In general terms, it was observed that the increase in charge blockiness is accompanied by conformational shrinkage and loss of oblateness, yet in a protein specific manner. Indeed, as κ increases, NFM - the protein with the lowest proline content in the experimental set – displays the greatest reduction in volume, while PNT4 - which is the model protein with the highest proline number - undergoes the greatest reshaping, moving from a highly prolate ellipsoid to a more spherical geometry. It is premature to infer that prolines alone can hinder electrostatically-induced compaction. Indeed, an overlapping effect of protein dimensions and secondary structure content could be in place. Overall, one may conclude that, in addition to the expected conformational response to charge clustering, multiple sequence context parameters could collectively impact on the behavior of IDP ensembles.

However, since charge patterning is so decisive in the conformational fate of IDPs, it is reasonable to believe that it may also determine the networking capabilities and condensation properties of this class of proteins. This hypothesis has been extensively investigated computationally and led to Lin and Chan's theory that the phase separation propensity of disordered proteins is positively correlated with the

increase in κ value. In partial contrast, the review article included in the ***Appendix*** of this thesis discusses this issue, arguing for the existence of a compaction threshold above which LLPS is not preferred, being the intrachain contacts too dominant. **Section 3.2** further explores the correlation between LLPS and charge patterning, choosing as a model the N-terminal disordered domain (*h*NTD, residues 1-214) of the human topoisomerase I (*h*TOP1). This domain is required for nuclear and nucleolar import of the protein, which unwinds both positive and negative DNA supercoiling, and is predicted to undergo phase separation, although there is no experimental evidence to support this. Sequence analysis revealed that the *h*NTD is almost devoid of hydrophobic and aromatic residues, yet is rich in charged residues, suggesting that the possible LLPS is electrostatically driven. The charge distribution is fairly regular, especially along the first 116 residues preceding the region containing the nuclear localization signals (NLSs), giving the *h*NTD an overall κ value of 0.131. Interestingly, the charge pattern of *h*NTD seems to correlate with the phylogeny of TOP1. Low κ values are widely conserved in the NTD of vertebrates, while more blocky distributions (high κ values) are spreaded among lower eukaryotes, such as fungi. This could suggest a selective pressure for evenly distributed charged residues in the disordered domain of TOP1, perhaps coevolved with specific mechanisms to control its condensation. Applying the same strategy as described for the design of κ permutants in **Section 3.1**, two synthetic κ variants of *h*NTD with increasing charge separation have been rationally designed, namely one characterized by a κ = 0.224 and designated as "medium κ-NTD" (Mκ-NTD), and one by a κ = 0.298 referred to as "high κ-NTD" (Hκ-NTD). Mutagenesis has involved the 1-116 region (*h*NTD$^{1\text{-}116}$), while keeping the NLS region (*h*NTD$^{117\text{-}214}$) unchanged. To begin with, LLPS was triggered by a rapid reduction in solution pH (pH jump). Turbidity assays have been performed for all three NTD proteins, varying protein and salt concentrations. Protein concentration positively influences demixing, whilst increasing ionic strength

suppresses condensation for *h*NTD and Mκ-NTD, yet not for Hκ-NTD. This confirms that the LLPS of model proteins is governed by electrostatics and suggests that the variant with highly clustered charges needs the presence of salts to attenuate intrachain interactions and permit protein condensation. The morphological characterisation (shape, size, coalescence propensity) performed by fluorescence confocal microscopy on the GFP-tagged NTD variants supports the notion that *h*NTD and Mκ-NTD permutant encounter canonical LLPS, forming round coalescent droplets, while Hκ-NTD forms amorphous, bunch-shaped condensates.

The use of poly(A) as a RNA mimic to trigger phase separation helped to investigate the possible role of the low-κ NTDs observed in the TOP1 of vertebrates. Indeed, while *h*NTD and Mκ-NTD produce biomolecular condensates of similar size and liquid nature, Hκ-NTD does not respond to short-term exposure of any poly(A) concentration, suggesting that it has lost this level of regulation. Interestingly, in the presence of poly(A), *h*NTD shows droplets five times larger than the condensates produced by pH jump. This indicates a stronger response to the RNA stimulus, which may be the main LLPS modulator of the protein in the cell. However, it is premature to infer information on the condensation behaviour of *h*TOP1, as (*i*) this is an *in-vitro* study that does not reproduce the complexity of the cellular environment; (*ii*) only the NTD sequence has been considered and not the full-length protein; (*iii*) *h*TOP1 may need different protein partners or specific RNAs (such as rRNAs that are abundant in the nucleolus) to undergo physiological LLPS. Nevertheless, evidences gathered so far point to *h*NTD as a phase-separating IDR whose low charge clustering may play a key role in responding to LLPS stimuli. Increasing its κ value up to a certain extent (Mκ-NTD) enhances the propensity for phase separation, as supported by Lin and Chan's model. However, beyond this limit, condensation can be counteracted by intrachain interactions, resulting in aberrant phase separation (Hκ-NTD). All this suggests that charge

clustering cannot be increased indefinitely without disturbing the balance of forces promoting LLPS. It is generally true that IDPs with high-κ values are rare in nature. Overall, this work shows that even modest variations in the natural charge distribution ($\Delta\kappa = 0.167$ between *h*NTD and Hκ-NTD) are enough to subvert the phase separation properties.

3. Results

Section 3.1

Distribution of Charged Residues Affects the Average Size and Shape of Intrinsically Disordered

Bianchi, G., Mangiagalli, M., Barbiroli, A., Longhi, S., Grandori, R., Santambrogio, C*., & Brocca, S*. (**2022**). *Biomolecules*, 12(4), 561.
*Corresponding author

Section 3.2

Charge distribution affects the liquid-liquid phase separation of the highly-charged N-terminal domain of human topoisomerase 1.

Bianchi, G., Mangiagalli, M., Ami, D., Longhi, S., Natalello, A., Tompa, P.*, & Brocca, S*.
*Corresponding author

(Manuscript in preparation)

biomolecules

MDPI

Article

Distribution of Charged Residues Affects the Average Size andShape of Intrinsically Disordered Proteins

Greta Bianchi[1], Marco Mangiagalli[1], Alberto Barbiroli[2], Sonia Longhi[3], Rita Grandori[1], Carlo Santambrogio[1,*] and Stefania Brocca[1,*]

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, 20126 Milan, Italy; greta.bianchi@unimib.it (G.B.); marco.mangiagalli@unimib.it (M.M.); rita.grandori@unimib.it (R.G.)

[2]Departement of Food, Environmental and Nutritional Sciences, University of Milan, 20133 Milan, Italy; alberto.barbiroli@unimi.it

[3]Laboratory Architecture et Fonction des Macromolécules Biologiques (AFMB), UMR 7257, Centre National de la Recherche Scientifique (CNRS), Aix Marseille University, 13288 Marseille, France; sonia.longhi@univ-amu.fr

* Correspondence: carlo.santambrogio@unimib.it (C.S.); stefania.brocca@unimib.it (S.B.); Tel.: +39-02-6448-3363 (C.S.); +39-02-6448-3518 (S.B.)

Abstract: Intrinsically disordered proteins (IDPs) are ensembles of interconverting conformers whose conformational properties are governed by several physico-chemical factors, including their amino acid composition and the arrangement of oppositely charged residues within the primary structure. In this work, we investigate the effects of charge patterning on the average compactness and shape of three model IDPs with different proline content. We model IDP ensemble conformations as ellipsoids, whose size and shape are calculated by combining data from size-exclusion chromatography and native mass spectrometry. For each model IDP, we analyzed the wild-type protein and two synthetic variants with permuted positions of charged residues, where positive and negative amino acids are either evenly distributed or segregated. We found that charge clustering induces remodeling of the conformational ensemble, promoting compaction and/or increasing spherical shape. Our data illustrate that the average shape and volume of the ensembles depend on the charge distribution. The potential effect of other factors, such as chain length, number of proline residues, and secondary structure content, is also discussed. This methodological approach is a straightforward way to model IDP average conformation and decipher the salient sequence attributes influencing IDP structural properties.

Keywords: charge clustering; polyelectrolytes; average shape of conformational ensembles; charged- residue patterning; hydrodynamic radius; solvent-accessible surface area; proline content; conformational compactness; ellipsoid model

1. Introduction

Intrinsically disordered proteins (IDPs) and regions have a biased sequence composition compared to folded counterparts, being enriched in disorder-promoting and charged amino acids and depleted in order promoting ones [1–3]. The high number of charged residues (Asp, Glu, Arg, Lys) has enabled modeling IDPs as either polyelectrolytes or polyampholytes, depending on the presence of same- or opposite-sign charges, respectively. The charge state of polyampholytes is often described by the total fraction of charged residues (*FCR*), obtained as the sum of the fractions of positive (f_+) and negative residues (f_-), and by the net charge per residue (*NCPR*), calculated as the difference between f_+ and f_- [4]. In addition to these coarse-grain parameters, the linear distribution of positive and negative charges, described by κ or sequence charge decoration parameters [5,6], is also an important feature in determining protein compactness. More in detail, computational and experimental data show that charge segregation promotes protein compaction [7–10].

IDPs consist of fluctuating and interconverting conformations that constitute "conformational ensembles". Size-exclusion chromatography (SEC), which enables molecule separation based on their hydrodynamic radius (R_h), is one of the most popular and easy to apply techniques to study the compaction of proteins, including IDPs. Experimentally, R_h can be determined from the chromatographic elution volume, using a calibration curve obtained with proteins of known R_h, or known molecular mass belonging to the same structural class [11,12]. Achieving a more quantitative description of IDP ensembles requires methods capable of dealing with heterogeneous molecular systems, such as nuclear magnetic resonance (NMR), small-angle X-ray scattering (SAXS), mass spectrometry (MS) combined with labeling techniques, high-speed atomic force microscopy, and Förster resonance energy transfer and non-denaturing mass spectrometry (native MS) to cite a few [13–17]. Native MS has been extensively employed to characterize the properties

of heterogeneous conformational ensembles, enabling the detection of even poorly populated states [18–21]. Indeed, gentle ionization conditions, such as those obtained by nano-electrospray ionization (nanoESI), preserve non-covalent interactions under the vanishing-solvent conditions of the electrospray, leading to protein ionization and transfer to the gas phase. The final protein net charge is mainly dictated by structural compactness under controlled conditions. Thus, charge state distributions (CSDs) in nanoESI spectra reveal the main components of conformational ensembles [17,20,22]. Unfolded/disordered proteins achieve higher charge states than their globular counterparts. For both folded and unfolded chains, the average charge state correlates with the solvent-accessible surface area (*SASA*), reflecting chain compactness [17,23–25].

In spite of the seminal and breaking-through studies by Pappu and co-workers that illuminated the relationships between charge distribution and conformational properties of IDPs [7,8,10], a full understanding of how the sequence of IDPs encodes their conformation is still lacking, thereby preventing, for instance, the *ex nihilo* design of IDPs with a precise set of desired conformational properties. With the goal of shedding light on these still unsolved issues, here we have studied the effect of charge segregation on three model IDPs that exhibit similar content in overall charged residues, net charge, and hydropathy, but different content of proline residues and secondary structure, and slightly different size. Charged residues within these model IDPs were permutated to obtain different κ- variants (**Figure 1**), and the three sets of proteins were characterized by SEC and ESI-MS. Experimentally derived R_h and *SASA* values were used to obtain coarse-grained structural information on these IDP ensembles using a recently published model, originally developed for globular proteins, that approximates the geometry of a protein to an ellipsoid [26].

Results show how the changes in average volume and shape triggered by the distribution of charged residues are variously affected by the frequency of proline

residues. In addition, we discussed the potential role of other factors such as secondary structure content and amino acid chain length.

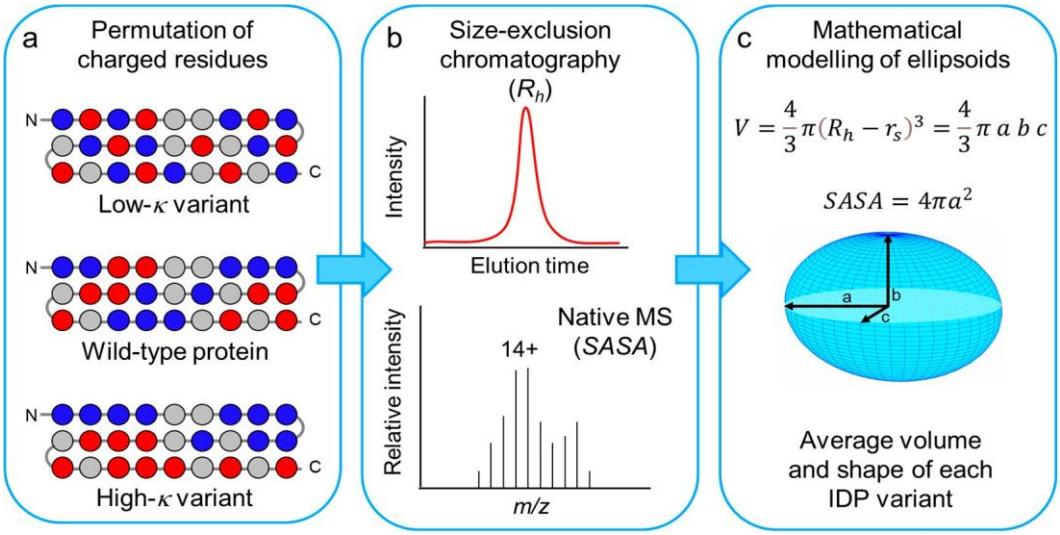


Figure 1. Scheme of the experimental plan used in this work. (**a**) Scheme of the primary structures of a protein set, derived from a generic wild-type IDP by distributing more evenly the oppositely charged residues (low-κ variant) or by clustering them in two blocks at the N- and C-moieties (high-κ variant). Only charged residues were permutated, preserving the original location in the sequence of non-charged residues (see also Figure S1). Blue and red spheres indicate positively and negatively charged residues, respectively. Gray spheres indicate all the other amino acid residues. (**b**) The conformational ensemble of each model IDP was investigated by size-exclusion chromatography (SEC) and native mass spectrometry (MS), to derive experimental values of Rh and *SASA*. (**c**) R_h and *SASA* values were combined to calculate the volume and depict the average shape from the ensemble of each model IDP.

2. Materials and Methods

2.1. Gene Design and Cloning

The model molecules employed in this study are IDPs derived from the measles virus N protein, N_{TAIL} [27], from the Hendra virus P protein, PNT4 [28], and from the human medium neurofilament protein, NFM (UniProtKB ID: P07197) [29,30]. The region used in this work (residues 790–916) belongs to the KE-rich tail of NFM, which is predicted to be intrinsically disordered. The rules followed for NFM gene design are those used for PNT4 and N_{TAIL} [8]. Briefly, we conceived low-*κ* and high-*κ* variants sharing with wild type (wt) the same number of charged residues and the same position of non-charged residues and differing just in the distribution of positively (Lys, Arg) and negatively (Glu, Asp) charged residues along the sequence. In high-*κ* sequences, positively and negatively charged residues are clustered in the N- and C-terminal regions, respectively. On the contrary, in low-κ sequences, positively and negatively charged residues are more evenly distributed than in the wt sequence. Synthetic genes encoding for NFM were optimized for expression in *Escherichia coli* (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). Each synthetic gene encodes a protein with an N-terminal hexa-histidine (6xHis) tag, while a stop codon has been inserted immediately before the XhoI restriction site, thereby excluding from the coding region the vector-encoded 6xHis tag. The amino acid sequences are shown in Figure S1. *Escherichia coli* DH5α™ strain (Invitrogen, Waltham, MA, USA) was used for plasmid DNA propagation.

2.2. Production and Purification of κ Variants

The *E. coli* strain BL21 (DE3) (EMD Millipore, Billerica, MA, USA) was used for protein heterologous production. Cultures were grown in ZYM-5052 medium [31], and recombinant IDPs were extracted and purified as described by Tedeschi and

co-authors [9]. Briefly, recombinant proteins were purified from the soluble fraction of the bacterial lysate by gravity-flow, immobilized-metal affinity chromatography using a nickel-nitrilotriacetic acid agarose resin (ABT, Torrejon de Ardoz, Madrid, Spain). The fractions exhibiting the highest concentration were pooled, and buffers were exchanged for phosphate-buffered saline (PBS, 150 mM NaCl, 50 mM sodium phosphate, pH 7.0) or ultrapure ammonium acetate buffer (ammonium acetate 50 mM, pH 6.95, Merck KGaA, Darmstadt, Germany) by gel filtration on PD-10 columns (GE Healthcare, Little Chalfont, UK). Protein concentration was determined by Bradford protein assay (Bio-Rad, Hercules, CA, USA), using bovine serum albumin as a standard.

2.3. Bioinformatics Analysis

Sequence analysis of model proteins was performed using CIDER [32] and IUPred [33] web servers. IUPred provides a score that characterizes the disordered tendency of each position along the sequence. The score ranges from 0 to 1, with predicted scores above 0.5 indicating disorder. CIDER was used with default parameters to compute κ values and local sequence properties such as *NCPR*, *FCR*, and the mean hydrophobicity in the 0–9 scaled Kyte-Doolittle hydropathy score.

2.4. Far-UV Circular Dichroism (CD) Spectroscopy

Far-UV CD analyses were carried out in PBS using a Jasco J-815 spectropolarimeter (Jasco Europe, Lecco, Italy) in a 1-mm path-length quartz cuvette. Measurements were performed 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). Experiments were performed in triplicate. Mean ellipticity values per residue ([θ]) were calculated as described by

Tedeschi and co-authors [9]. The deconvolution of CD spectra to assess secondary structure content was performed using the BestSel program [34].

2.5. Analytical SEC

Recombinant IDPs produced in this work were analyzed by SEC within the day they were purified. Chromatographic separations were carried out on a Superose 12 10/300 GL column (GE Healthcare, Milan, Italy), in mobile phase PBS, at a flow rate 0.5 mL/min. The chromatographic system was composed of a Waters Delta 600 pump, a 600 Controller, and a 2487 Dual λ Absorbance Detector; all managed through the Empower Pro Software (Waters Corporation, Milford, MA, USA). Chromatograms were recorded at 220 nm. The calibration curve was built using the following standards: Apo-ferritin (horse spleen, 443 kDa, R_h 6.1 nm), Alcohol dehydrogenase (yeast, 150 kDa, R_h 4.6 nm), BSA (bovine serum, 66 kDa, R_h 3.5 nm), Ovalbumin (chicken egg, 43 kDa, R_h 2.8 nm), Carbonic anhydrase (bovine erythrocytes, 29 kDa, R_h 2.1 nm), Cytochrome C (horse heart, 12.4 kDa, R_h 1.7 nm [35].

Firstly, for each standard protein the distribution coefficient (K_d) was calculated:

$$K_d = \frac{V_e - V_0}{V_t - V_o} \quad (1)$$

where V_e is the elution volume, V_0 the void volume, and V_t the total volume. Uracil (0.112 kDa) and Blue dextran (2000 kDa) were used for V_t and V_0 determination.

Finally, the calibration curve Log(R_h) vs. K_d was built and the interpolated linear equation used to calculate IDPs hydrodynamic radii from their K_d values. IDPs were run at least in triplicate.

The theoretical radius (R_t) was calculated according to the empirical Equation (2) [36].

$$R_t = (1.24\ P_{pro} + 0.904)\ (0.00759\ |Q| + 0.963)\ S_{his*} \quad (2)$$

where P_{pro} is the number of proline residues, $|Q|$ the absolute net charge and the S_{his*} is 0.901 or 1 depending on whether a 6xHistag is present or absent, respectively.

R_h values were used to calculate the compaction index (*CI*), which provides a simple and continuous descriptor useful for comparing conformational properties of IDPs of different lengths [23,37]. The *CI* derived from the experimental value of R_h (CI_R) was calculated by applying the following equation [37]:

$$CI_R = \frac{R^D - R_h}{R^D - R^{NF}} \quad (3)$$

where R_h is the experimental value, R^D and R^{NF} are the theoretical values of a chemically denatured or a folded protein, calculated on the basis of power-law Equations (4) and (5), which describe their dependence on the number of residues, N [11].

$$R^{NF} = 4.92 \bullet N^{0.285} \quad (4)$$

$$R^D = 2.49 \bullet N^{0.509} \quad (5)$$

2.6. Native MS Analyses

Protein solutions in 50 mM ammonium acetate, pH 7.0, were brought to a concentration of 10 µM, and samples under non-denaturing conditions were directly injected at room temperature into an Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific, Waltham, MA, USA) equipped with a nano-electrospray ion source. Metal-coated borosilicate capillaries with medium-length emitter tips of 1 µm internal diameter were used to infuse the sample. To assess the effect of electrostatic interactions, protein samples were also analyzed at higher ionic strength (200 mM ammonium acetate pH 7.0) and low pH (no buffer, 1% formic acid, pH 2.5). The following instrumental setting was applied: ion spray voltage, 1.1–1.2 kV; ion-transfer tube temperature, 275 °K; AGC target, $4x10^5$; maximum

injection time, 100 ms. Spectra were averaged over 1-min acquisition. Multi-Gaussian fitting of MS spectra was performed employing the program OriginPro 2020 (OriginLab Corporation, Northampton, MO, USA), and *CI* of single conformers (CI^{i}_{SASA}) and ensembles ($\overline{CI}_{SASA}$) were calculated as follows [16]:

$$CI^{i}_{SASA} = \frac{A^{c} - A^{0}}{A^{c} - A^{f}} \quad (6)$$

$$\overline{CI}_{SASA} = \sum_{i=1}^{n} w_i . \; CI^{i}_{SASA} \quad (7)$$

where A^c and A^f are the solvent-accessible surface areas derived by native MS for reference, random coil (*c*) and folded (*f*) proteins of the same size of the protein under study, A^0 is the solvent-accessible surface areas derived by native MS for the conformer (exploiting the charge state—*SASA* relationship), w_i is the relative amount of the conformer with compaction index CI^{i}_{SASA}.

Statistical significance of experimental differences was estimated by performing a Welch's t-test on three independent datasets.

2.7. Application of Ellipsoid Model

The ellipsoid model assumes that the average conformation of a given protein can be represented by an ellipsoid with semi-axes *a*, *b*, and *c* (a ≥ b ≥ c) [26]. The experimental ellipsoid volume depicting the conformation of the IDP averaged over the ensemble can be estimated by the volume of a sphere given by the following formula:

$$V = \frac{4}{3}\pi(R_h - r_s)^3 \quad (8)$$

where r_s represents the hydration shell (generally assumed to be 5 Å) [38,39], and R_h the hydrodynamic radius obtained by SEC experiments. The geometrical volume of an ellipsoid is expressed as:

$$V = \frac{4}{3}\pi abc, \quad (9)$$

To calculate a, the quadratic relationship with SASA given by the model of Wu and co-authors [26] can be exploited:

$$SASA = 4\pi a^2, \quad (10)$$

Then, *b* and *c* values can be approximated by weighted averages between the extreme conditions of prolate ($a > b = c$) and oblate ($a = b > c$) ellipsoids, according to the equations published by Wu and co-authors [26].

Thomsen's approximation was employed to calculate the ellipsoid surface area (maxi- mal discrepancy to real surface ~1%).

The ellipsoid *flattening* was described through the values of f_b and f_c, calculated according to the formulas:

$$f = \frac{(a-b)}{a}; f = \frac{(a-c)}{a} \quad (11)$$

Both indices report the eccentricity of axial elliptic sections of the ellipsoid, and span in the range [0;1), where 0 corresponds to a circular section.

3. Results

3.1. Design of Model IDPs by Permutation of Charged Residues

The model IDPs used in this work are the viral proteins PNT4 and N_{TAIL} and a C-terminal IDR from the human NFM. These IDPs are similar in length, theoretical hydrodynamic radius (R_t), charge density, and charge segregation, as witnessed by their κ value (Table 1). Values of κ vary between 0 and 1, with 0 indicating evenly mixed positive and negative residues, and 1 referring to the complete segregation of oppositely charged residues along the linear sequence [4]. In our model proteins, the number of positive and negative charges is well balanced, producing a rather low *NCPR* (mean absolute value 0.038 ± 0.017), and opposite charges are evenly distributed along the sequence, thereby resulting in rather low κ values (mean value 0.167 ± 0.041). The three proteins differ in the fraction of proline residues, which is 0.7%, 5.2%, and 11.4% for NFM, N_{TAIL}, and PNT4, respectively. Among disorder-

promoting residues, proline residues are also recognized to disfavor α-helical and β-structures [40], and to promote extended conformations by conferring rigidity to the backbone [36]. For each model IDP, a "high-κ" and a "low-κ" variants were designed by permuting charged residues while keeping the position of all other residues unchanged. Table 1 summarizes, for each model protein and its variants, the κ parameter, *NCPR*, and *FCR* values calculated using the CIDER webserver [28].

Table 1. Features of the three model proteins and their derived κ variants. Sequence features were computed using CIDER [28]; the theoretical radius R_t was calculated according to Marsh and Forman-Kay [35].

Protein	N° residues	N° prolines	Mean Hydropathy	FCR	NCPR	R_t (nm)	κ	Variant
N_{TAIL}	134	7	3.35	0.299	-0.045	2.64	0.078	Low κ
							0.153	wt
							0.431	High κ
NFM	136	1	3.40	0.390	-0.051	2.54	0.037	Low κ
							0.134	wt
							0.516	High κ
PNT4	114	13	3.26	0.298	0.018	2.54	0.044	Low κ
							0.213	wt
							0.421	High κ

In the high-κ variants, positive and negative charged residues are clustered in two distinct blocks at the N- and C-terminal moieties of the sequence, while in low-κ variants, these residues are evenly alternated along the sequence, as highlighted by their *NCPR* profiles (**Figure 2a–c**, upper panels, and **Figure S1**). The degree of disorder predicted by IUPred [33] is conserved within each set of model proteins derived by permutation from the respective wt sequence (**Figure 2a–c**, lower panels). The three sets of proteins were recombinantly produced and purified by immobilized-metal affinity chromatography and experimentally assessed by CD analysis in the far-UV (**Figure S2**). The CD spectra of wt IDPs display the typical

trait of structural disorder with a negative peak at ~200 nm (black line in **Figure S2**). Worthy to note, all the spectra of wt IDPs present a small shoulder at ~220 nm, which indicates the presence of some elements of helical secondary structure. Despite the common high level of disorder predicted by IUPred, deconvolution of CD spectra indicates that in all the three model IDPs the α-helical content tends to increase along with the values of κ (**Figure S2**, inset).

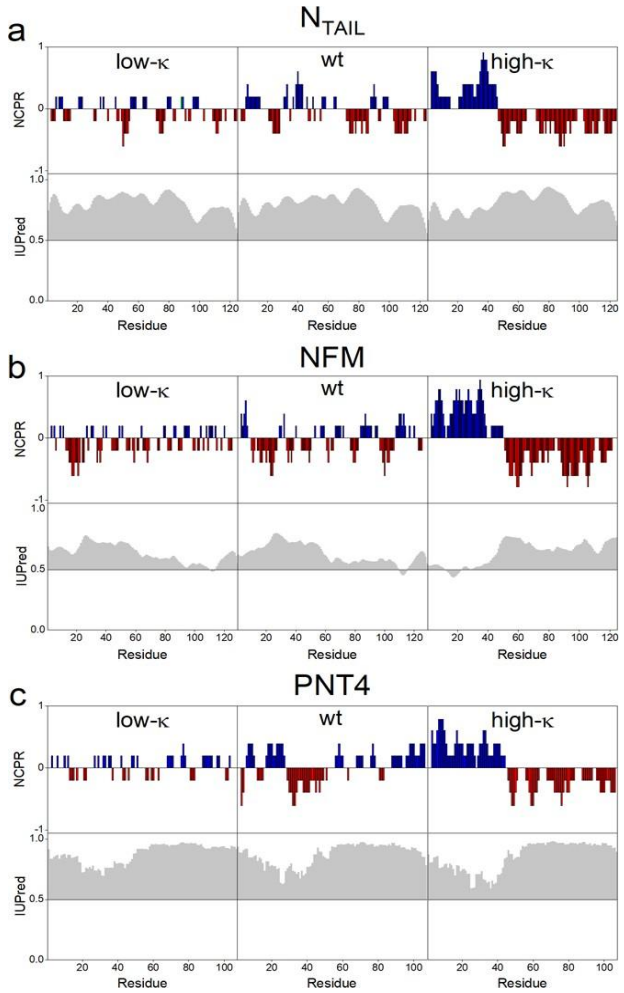


Figure 2. Comparative bioinformatic analyses of N_{TAIL} (**a**), NFM (**b**) and PNT4 (**c**). Upper panels: The *FCR*, fraction of charged residues, was calculated by CIDER [32]. Each model protein contains charged residues at high density, with

red and blue bars indicating negative and positive charges, respectively. The increase in κ value is reflected in the progressively more "blocky" distribution of charged residues. Lower panel: each protein is predicted to be predominantly disordered by IUPred [33]. The discrepancy from the disorder threshold value (0.5) in the IUPred score is shaded in gray. The IUPred and CIDER outputs were generated using the default options of the respective web server.

3.2. Impact of Charge Clustering on the R_h of the Model IDPs

Size-exclusion chromatography was employed to estimate the R_h values of the three sets of model IDPs (Table 2). Experimental R_h values of wt N_{TAIL} and wt PNT4 (2.71 ± 0.09 and 2.34 ± 0.11 nm, respectively) are close to the theoretical ones (Table 1) and similar to the previously determined ones [9]. The R_h of wt NFM (3.31 ± 0.12) is determined here for the first time. We observed that R_h decreases as κ increases for N_{TAIL} and NFM, but not for PNT4 (Table 2). To compare the compaction properties of IDPs with different chain lengths, R_h data were used for the calculation of the R_h-based *CI* (CI_R, defined in Equation (2)). The value of *CI* ranges from 0 to 1, corresponding to minimal and maximal compaction, respectively [37]. Analysis of the CI_R confirms that N_{TAIL} and NFM significantly respond to charge segregation, while PNT4 average compactness is not affected by the κ value (Figure 3a).

Table 2. Hydrodynamic radii (R_h) and average solvent accessible surface area (*SASA*) of the three model proteins and their derived κ variants. Mean values and standard deviations from three independent measurements are reported. Volume, surface area and flattening indices of the ellipsoids were derived from the model proposed by Wu and co-authors [23].

Protein variant		R_h (nm)	SASA (nm^2)	Volume (nm^3)	f_b*	f_c*
N_{TAIL}	Low κ	2.78 ± 0.03	113.2 ± 2.1	49.5 ± 2.2	0.26 ± 0.02	0.41 ± 0.16
	Wt	2.73 ± 0.03	105.6 ± 1.4	46.3 ± 1.8	0.25 ± 0.01	0.39 ± 0.16
	High κ	2.58 ± 0.05	89.3 ± 1.0	37.5 ± 2.3	0.24 ± 0.02	0.38 ± 0.16
NFM	Low κ	3.37 ± 0.05	136.2 ± 2.0	99.1 ± 5.0	0.14 ± 0.02	0.23 ± 0.12
	Wt	3.31 ± 0.04	124.5 ± 2.4	93.2 ± 4.8	0.12 ± 0.02	0.19 ± 0.10
	High κ	3.05 ± 0.10	81.5 ± 0.4	69.7 ± 8.0	-0.03 ± 0.08	0.02 ± 0.09
PNT4	Low k	2.39 ± 0.04	106.8 ± 3.0	28.1 ± 2.1	0.42 ± 0.03	0.53 ± 0.13
	Wt	2.36 ± 0.05	106.8 ± 2.5	26.8 ± 2.1	0.44 ± 0.03	0.54 ± 0.12
	High k	2.43 ± 0.05	69.0 ± 1.0	29.9 ± 2.4	0.19 ± 0.03	0.31 ± 0.15

* flattening indices relative to *b (1-b/a)* and *c (1-c/a)* axis.

3.3. Impact of Charge Clustering on the Conformational Ensemble of the Model IDPs

Native MS was employed to assess the conformational properties of the three sets of IDPs. In this approach, the CSDs resulting from the nanoESI process reflect the overall compactness and relative amounts of the main conformers in the original solution [17,18,22]. Native-MS spectra obtained under non-denaturing conditions for the three variants of N_{TAIL} (**Figure 4a**), NFM, and PNT4 (**Figure S3**) display multimodal CSDs, highlighting the heterogeneous conformational ensemble typical of IDPs. Multi-Gaussian deconvolution of the MS spectra of the wt IDPs (**Figure 4b–d**, central row) indicates that these variants exist in three main conformational components. For each component, the *SASA* and the corresponding *CI* (CI^i_{SASA} defined in Equation (7)) were calculated as recently described [17]. The components were classified as "extended" ($CI^i_{SASA} < 0.25$), "intermediate" ($0.25 < CI^i_{SASA} < 0.75$), and "compact" ($CI^i_{SASA} > 0.75$) (Figure S4). In all the model IDPs, the three main conformational components observed in the wt IDPs also characterize the ensemble of low-*κ* variants, but not that of high-*κ* variants, which includes only the "inter- mediate" and "compact" components (**Figure 4**). These

data indicate that charge clustering induces a loss of heterogeneity of conformational components, in favor of more compact states, in agreement with the increase in secondary structure observed by CD spectroscopy on our model proteins and also with results obtained on p27 by ion-mobility MS [8]. To gain a more comprehensive view of charge clustering effects on IDP conformation, we calculated the *CI* based on the average *SASA* ($\overline{CI}_{SASA}$), which weights the CI^{i}_{SASA} (**Figure S4**) by the relative abundance (**Figure S5**) of the corresponding conformational component. The analysis of $\overline{CI}_{SASA}$ indicates that the protein compactness increases with κ (**Figure 3b**). These results are in good agreement with those obtained by SEC, confirming the general trend of protein compaction at increasing κ values and the peculiar behavior of PNT4.

In this latter case, the $\overline{CI}_{SASA}$ does not vary for low-κ and wt variants, and it strongly increases just for high-κ variants (**Figure 3**). Overall, the largest differences between MS and SEC results are obtained for the high-κ variants. To rule out possible technical artifacts, control MS experiments were carried out, exposing high-κ variants to acidic pH (formic acid 1%, pH 2.5) or higher ionic strength (ammonium acetate 200 mM). Indeed, electrostatic interactions are expected to be attenuated by the extensive protonation of all ionizable groups under very low pH conditions or by the charge shielding by salt ions. The resulting spectra show an increased amount of the components at high charge states, indicating that protein compaction is actually driven by in-solution electrostatic interactions (**Figure S6**).

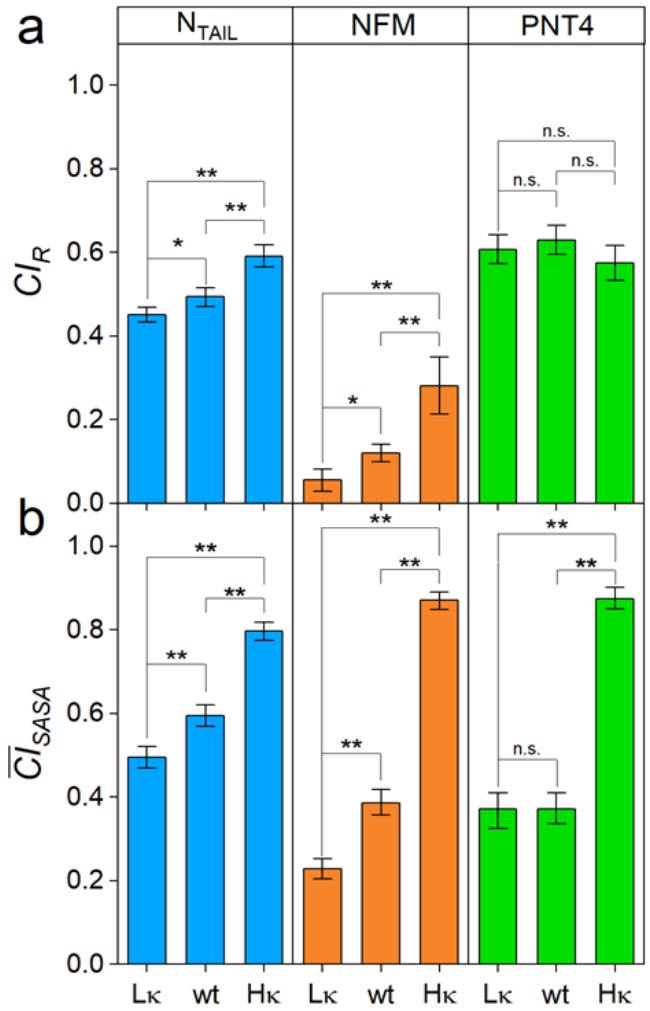


Figure 3. Compactness of the model IDPs. (**a**) *CI* derived from the R_h (CI_R); (**b**) *CI* derived from the average *SASA* of the conformational ensemble ($\overline{CI}_{SASA}$) of N_{TAIL}, NFM and PNT4 variants (Lκ: low-κ; wt: wild type; Hκ: high-κ). Mean values of three independent measurements are shown with error bars indicating standard deviations. Statistical analyses were carried out using Welch's t-test, n.s.: not significant $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$.

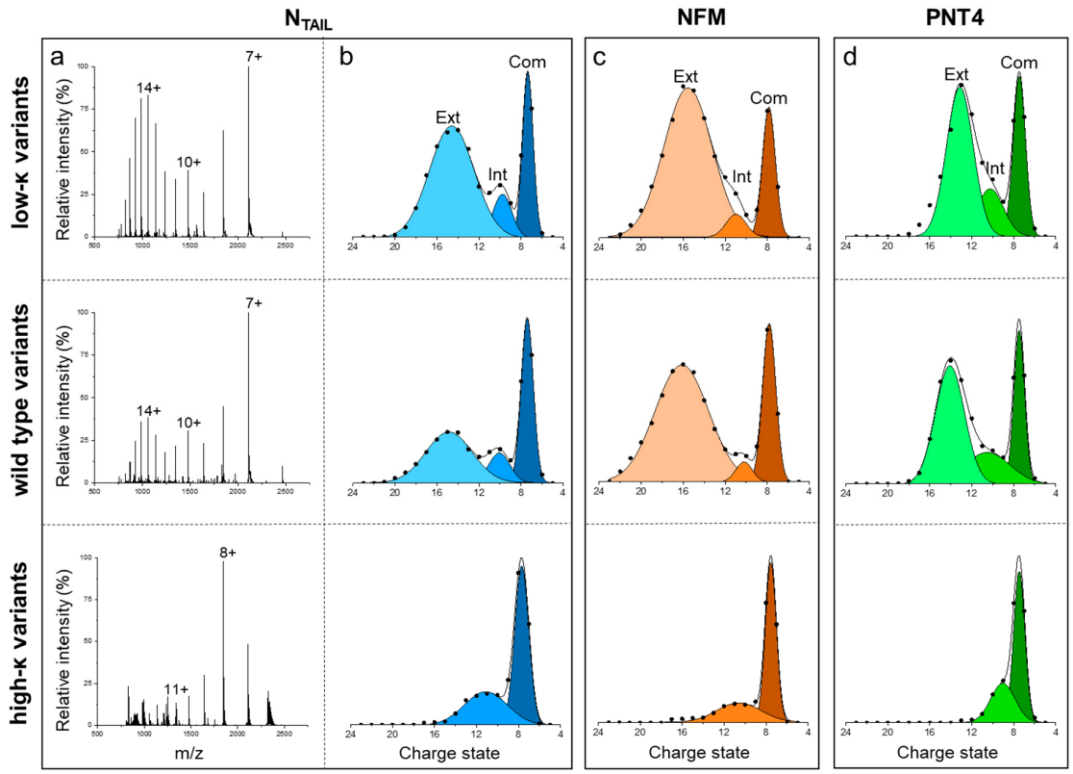


Figure 4. Native-MS analyses. (**a**) NanoESI-MS spectra of N_{TAIL} variants acquired under non- denaturing conditions (50 mM ammonium acetate pH 7.0). The most intense signal of each peak- envelope is labeled by the corresponding charge state. (**b–d**) Multi-Gaussian deconvolution of the MS spectra obtained for N_{TAIL} (**b**), NFM (**c**) and PNT4 (**d**), in the low-*κ* (upper row), wt (central row) and high-*κ* (bottom row) variants. Extended (Ext), intermediate (Int) and compact (Com) species are colored with different shades and labeled in the upper panels. MS spectra of NFM and PNT4 variants are reported in Figure S3.

3.4. Average Shape of the Model IDPs

The geometric, ensemble-averaged shape of each protein under investigation was predicted by combining the results for R_h and *SASA*, as reported by Wu and co-authors [26]. The model was originally applied to approximate the shape of globular proteins to an ellipsoid, whose elongation (prolate-shaped) and/or

flattening (oblate-shaped) describe the protein conformational transitions. The volume of the ellipsoid can be estimated from the experimentally derived R_h, through Equation (8). By collating Equations (8) and (9), one obtains:

$$V = \frac{4}{3}\pi abc = \frac{4}{3}\pi(R_h - r_s)^3 \quad (12)$$

The average length of the *a*-axis was calculated through Equation (10), while the length of the *b* and *c* axes were obtained as described by Wu and co-authors [26]. The application of this model to the nine IDPs under investigation resulted in the values shown in Table 2 and Table S1 and represented in **Figure 5**, in which ellipsoid volumes and shapes are related to κ values. Comparing wt variants, NFM has the largest volume, followed by N_{TAIL} and PNT4. Considering the effects induced by charge clustering, and therefore moving from the lowest towards the highest κ values, a clear linear and negative correlation can be observed in the case of NFM and N_{TAIL} (overall reduction in volume of ~30% and ~25%, respectively) (**Figure 5**, Table 2). On the other hand, a neglectable effect was observed in the case of PNT4, for which the volume remains almost constant among the three variants, reflecting little variation of their $_{Rh}$ value.

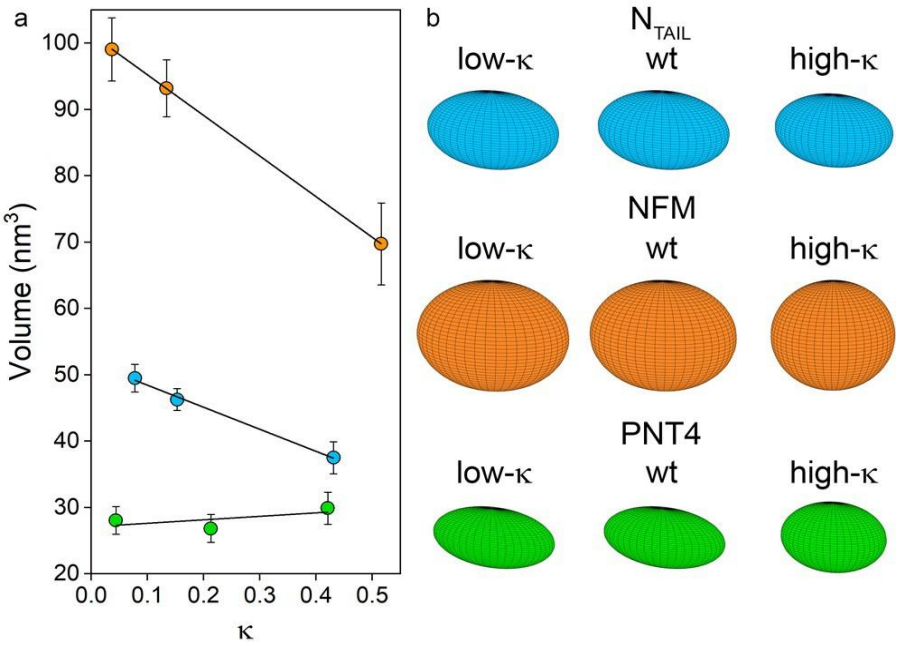


Figure 5. Relationship between ellipsoid volume and κ values. (**a**) Regression of ellipsoid volume and κ for N_{TAIL} (light blue), NFM (orange) and PNT4 (green). The equation of trend lines are: y = −33.4 x + 51.7, R2 = 0.987 for N_{TAIL}, y = −61.2 x + 101.4, R2 = 0.998 for NFM and y = 4.3 x + 27.2, R2 = 0.710 for PNT4. Mean values of three independent measurements are represented, with error bars indicating standard deviations. (**b**) Geometry of the model proteins as obtained by applying the ellipsoid model.

The shape of an ellipsoid depends on the length ratio of the *a*, *b*, and *c* axes, which in turn was derived from the experimental data of *SASA* (Table 2, **Figure 3**). The shape of an ellipsoid can be described by flattening indices (*i.e.*, *fb* and *fc*), which report the eccentricity of axial elliptic sections. These indices vary in the range [0; 1), where 0 corresponds to circle sections, while elliptic sections of increasing eccentricity are obtained as the index approaches 1 (Table 2). Comparing wt variants, NFM has the most spherical conformation, followed by N_{TAIL} and PNT4 (which has the most prolate ensemble). As the κ value increases, the spheroid reshaping reflects the trends observed by native MS and reported in terms of $\overline{CI}$

$_{SASA}$ with N_{TAIL} experiencing the smallest changes, and NFM and PNT4 the greatest ones (Table 2, **Figure 3b**). Indeed, on the basis of the flattening indices, the oblateness of N_{TAIL} is not significantly affected by κ, while NFM and PNT4 tend to approach a spherical shape.

4. Discussion

Computational and experimental works have already shown that charge clustering causes an overall increase in protein conformational compactness [7–10]. However, few data are available in terms of quantitative description of various conformational components within a heterogeneous ensemble. Our work highlights that the conformational ensembles of IDPs can be experimentally dissected by native MS to capture components of different *SASA* and abundance. Our results show that charge segregation triggers a loss of heterogeneity of conformational components, in favor of more compact and intermediate states. At the same time, we used SEC to monitor the average Rh and observed an overall shrinkage resulting from charge clusterisation.

To integrate the two kinds of information resulting from MS and SEC, and to obtain coarse-grained information on the shape of IDP ensembles, we applied a recently published model, which approximates the shape of globular proteins to ellipsoids [26]. The applicability of this "ellipsoid model" to IDPs, herein explored for the first time, is supported by three observations: (*i*) the relationship between CSD and *SASA* was proved to be independent of the folded or disordered nature of the proteins [20,23,25]; (*ii*) the ellipsoid model was successfully applied to depict the conformational changes induced by denaturation [26]. The broad molecular mass range of globular proteins for which the model was shown to hold true (*i.e.*, ~9 kDa to ~70 kDa) [26] argues for the applicability of this model to the three model IDPs herein investigated whose mass falls within this range. This model substantially helped us in translating and rationalizing the

conformational effects induced by charge clustering into the shrinkage and loss of oblateness of each IDP ensemble, while providing evidence of singular, protein-specific compaction behaviors. The observation that each ellipsoid undergoes volume and shape changes in a protein-specific manner argues for a multifactorial response to charge segregation. Although referring to a small set of proteins, and hence likely not directly generalizable to all IDPs, our data suggest that proline content, chain length, and secondary structure content are potentially all involved in the response to charge segregation.

Proline content appears to play a relevant role in modulating the average conformational properties of the ensemble. Indeed, the abundance of proline residues (PNT4 > N_{TAIL} > NFM) promotes the ellipsoid oblateness in wt variants and counteracts the volume shrinking induced by κ. This is in line with the observations that proline disfavors α- and β structures [36,41] because of the conformational constraints imposed by its pyrrolidine ring [42] and the higher stiffness conferred by the preference towards the trans conformation of the Xaa-Pro peptide bonds [36]. Our data show that an increase from 0.7 to 5.2%, and then to 11% in proline content causes a significant reduction in the compaction response associated with charge clustering. Remarkably, the mean frequency of proline residues is 4.57 ± 0.05 and 8.11 ± 0.63 in databases of structured (*i.e.*, PDB Select 25 [43]) and disordered proteins (*i.e.*, DisProt [44,45]), respectively. In this scenario, proline residues would strongly hinder compaction driven by electrostatic interactions and reduce IDP propensity for induced folding. This indirectly supports the hypothesis that a high proline content is a compositional trait typical of "unfoldable IDPs", in contrast to IDPs prone to undergo induced folding, which instead exhibit, at least locally, compositional features nearly overlapping with those of folded proteins [2,36,46]. This hypothesis is corroborated by the analyses of large protein datasets [46].

Polypeptide length may also affect the ellipsoid oblateness in wt variants and counter- act κ-induced volume shrinking. Indeed, PNT4 (the shortest protein under investigation) responds to increasing κ with small volume changes and pronounced shape remodeling (from highly prolate ellipsoid to a more spherical geometry in the high-κ variant), whereas NFM (the longest chain herein studied) shows the greatest volume excursion among variants. Unfortunately, it is difficult to disentangle the contribution of chain length and proline content to charge clustering responsiveness: indeed, the attempt at rationalizing our experimental data and at dissecting the effect of protein length is hampered by the fact that PNT4 has the highest fraction of proline residue and NFM the lowest among our model proteins, thus making the effect of size and proline content overlapping.

Finally, the role of secondary structure content appears controversial. For each of the three proteins, charge clustering triggers an increase in the α-helical content. This could be related to the loss of heterogeneity among conformational components in favor of more compact and intermediate states observed by MS experiments. However, α-helical content does not correlate with compaction in terms of CIR and volume shrinkage (*e.g.*, PNT4). This behavior seems to be consistent with previous studies indicating that the propensity of IDPs for compactness, unlike that of globular proteins, is not correlated with α-helical content [36,47]. Unfortunately, the paucity of data concerning the effects of charge segregation on IDP secondary structure makes it difficult to detail trends and deserves more extensive and systematic study.

Overall, our experimental data, complemented by the ellipsoid model, indicate that the extent of compaction and shape remodeling triggered by charge separation is modulated by multiple parameters that can concur, either individually or collectively, to counteract the expected response. Among the possible sequence features affecting IDP conformational responsiveness to charge clustering, the Lys/Arg and Asp/Glu ratio, recently reported by Zeng and co-authors [48], is a plausible factor that deserves further investigation. Many additional ones are

probably at play and still remain elusive, thereby preventing our ability to fully rationalize and model the conformational behavior of IDPs.

5. Conclusions

In summary, the effect of charge segregation on the conformational properties of IDP ensembles was studied by applying a mathematical model that integrates experimental data from two orthogonal techniques, i.e., SEC and native MS. This original approach was proved to be more informative compared to the single techniques, delineating a distinct and protein-specific compaction behavior in terms of the size and shape of each conformational ensemble. The structural information afforded by this approach relies on techniques that are more accessible compared to more elaborate techniques, such as ion mobility, NMR, or SAXS, usually applied for the study of IDP ensembles. Potentially transposable on a larger scale, *i.e*., by using available experimental datasets of *SASA* and R_h, this approach could also serve as an asset to a more systematic study of the individual factors influencing the compaction behavior of IDPs triggered by charge segregation.

Although we do not pretend to extend our findings to all IDPs, our work identified proline content, protein size, and intrinsic content in ordered secondary structure as factors governing IDP responsiveness. We hope that the present study will stimulate and foster future studies aimed at a systematic analysis of the elements that contribute to the con- formational behavior of IDPs in response to charge clustering. In addition to unraveling the physicochemical rules underlying the response to charge segregation, these elements may account for sequence-specific and biologically relevant properties of proteins, such as the propensity to undergo induced folding or to exhibit partner-mediated conformational polymorphism. The next challenge will be to decipher the hierarchy of elements governing IDP conformation and how they can be modeled to better predict IDP behavior.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/biom12040561/s1: Table S1: Size of semi-axes of ellipsoids; Figure S1: Amino acid sequences of model IDPs; Figure S2: Far-UV CD spectra of model IDPs; Figure S3: Native MS analyses of model IDPs; Figure S4: Compactness of conformational components of model IDPs studied by native MS; Figure S5: Relative abundance of conformational components of model IDPs studied by native MS; Figure S6: NanoESI-MS spectra of model IDPs under conditions affecting electrostatic interactions.

Author Contributions: Conceptualization, G.B., M.M., R.G., C.S. and S.B.; Formal analysis, C.S.; Investigation, G.B., A.B. and C.S.; Methodology, A.B., S.L., R.G., C.S. and S.B.; Supervision, C.S. and S.B; Visualization, M.M., C.S. and S.B.; Writing—original draft, S.B.; Writing—review and editing, G.B., M.M., A.B., S.L., R.G. and S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. This work was partly supported by Fondo di Ateneo of the University Milano-Bicocca (FAR 2020-ATE-0031, SB; FAR 2020-ATE-0319, CS), by the CNRS (S.L.). G.B. acknowledges support by the Italian national PhD program. M.M. benefits from an Assegno di Ricerca of the University Milano-Bicocca.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank Giulia Tedeschi for her contribution in the experimental set up.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Uversky, V.N. Intrinsically Disordered Proteins and Their "Mysterious" (Meta) Physics. Front. Phys. 2019, 7, 10. [CrossRef]

2. Theillet, F.-X.; Kalmar, L.; Tompa, P.; Han, K.-H.; Selenko, P.; Dunker, A.K.; Daughdrill, G.W.; Uversky, V.N. The Alphabet of Intrinsic Disorder: I. Act like a Pro: On the Abundance and Roles of Proline Residues in Intrinsically Disordered Proteins. Intrinsically Disord. Proteins 2013, 1, e24360. [CrossRef] [PubMed]

3. Ruff, K.M. Predicting Conformational Properties of Intrinsically Disordered Proteins from Sequence. In Intrinsically Disordered Proteins; Kragelund, B.B., Skriver, K., Eds.; Methods in Molecular Biology; Springer US: New York, NY, USA, 2020; Volume 2141, pp. 347–389. ISSN 978-1-07-160523-3.

4. Mao, A.H.; Crick, S.L.; Vitalis, A.; Chicoine, C.L.; Pappu, R.V. Net Charge per Residue Modulates Conformational Ensembles of Intrinsically Disordered Proteins. Proc. Natl. Acad. Sci. USA 2010, 107, 8183–8188. [CrossRef] [PubMed]

5. Das, R.K.; Pappu, R.V. Conformations of Intrinsically Disordered Proteins Are Influenced by Linear Sequence Distributions of Oppositely Charged Residues. Proc. Natl. Acad. Sci. USA 2013, 110, 13392–13397. [CrossRef]

6. Sawle, L.; Ghosh, K. A Theoretical Method to Compute Sequence Dependent Configurational Properties in Charged Polymers and Proteins. J. Chem. Phys. 2015, 143, 085101. [CrossRef] [PubMed]

7. Das, R.K.; Huang, Y.; Phillips, A.H.; Kriwacki, R.W.; Pappu, R.V. Cryptic Sequence Features within the Disordered Protein p27

Kip1 Regulate Cell Cycle Signaling. Proc. Natl. Acad. Sci. USA 2016, 113, 5616–5621. [CrossRef]

8. Beveridge, R.; Migas, L.G.; Das, R.K.; Pappu, R.V.; Kriwacki, R.W.; Barran, P.E. Ion Mobility Mass Spectrometry Uncovers the Impact of the Patterning of Oppositely Charged Residues on the Conformational Distributions of Intrinsically Disordered Proteins. J. Am. Chem. Soc. 2019, 141, 4908–4918. [CrossRef]

9. Tedeschi, G.; Salladini, E.; Santambrogio, C.; Grandori, R.; Longhi, S.; Brocca, S. Conformational Response to Charge Clustering in Synthetic Intrinsically Disordered Proteins. Biochim. Et Biophys. Acta Gen. Subj. 2018, 1862, 2204–2214. [CrossRef]

10. Sherry, K.P.; Das, R.K.; Pappu, R.V.; Barrick, D. Control of Transcriptional Activity by Design of Charge Patterning in the Intrinsically Disordered RAM Region of the Notch Receptor. Proc. Natl. Acad. Sci. USA 2017, 114, E9243–E9252. [CrossRef]

11. Uversky, V.N. What Does It Mean to Be Natively Unfolded?: Natively Unfolded Proteins. Eur. J. Biochem. 2002, 269, 2–12. [CrossRef]

12. Schramm, A.; Bignon, C.; Brocca, S.; Grandori, R.; Santambrogio, C.; Longhi, S. An Arsenal of Methods for the Experimental Characterization of

Intrinsically Disordered Proteins—How to Choose and Combine Them? Arch. Biochem. Biophys. 2019, 676, 108055. [CrossRef] [PubMed]

13. Kodera, N.; Noshiro, D.; Dora, S.K.; Mori, T.; Habchi, J.; Blocquel, D.; Gruet, A.; Dosnon, M.; Salladini, E.; Bignon, C.; et al. Structural and Dynamics Analysis of Intrinsically Disordered Proteins by High-Speed Atomic Force Microscopy. Nat. Nanotechnol. 2021, 16, 181–189. [CrossRef] [PubMed]

14. Gomes, G.-N.W.; Krzeminski, M.; Namini, A.; Martin, E.W.; Mittag, T.; Head-Gordon, T.; Forman-Kay, J.D.; Gradinaru, C.C. Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET. J. Am. Chem. Soc. 2020, 142, 15697–15710. [CrossRef] [PubMed]

15. Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge Interactions Can Dominate the Dimensions of Intrinsically Disordered Proteins. Proc. Natl. Acad. Sci. USA 2010, 107, 14609–14614. [CrossRef]

16. Corti, R.; Marrano, C.A.; Salerno, D.; Brocca, S.; Natalello, A.; Santambrogio, C.; Legname, G.; Mantegazza, F.; Grandori, R.; Cassina, V. Depicting Conformational Ensembles of α-Synuclein by Single Molecule Force Spectroscopy and Native Mass Spectroscopy. Int. J. Mol. Sci. 2019, 20, 5181. [CrossRef]

17. Santambrogio, C.; Natalello, A.; Brocca, S.; Ponzini, E.; Grandori, R. Conformational Characterization and Classification of Intrinsically Disordered Proteins by Native Mass Spectrometry and Charge-State Distribution Analysis. Proteomics 2019, 19, 1800060. [CrossRef]

18. Li, J.; Santambrogio, C.; Brocca, S.; Rossetti, G.; Carloni, P.; Grandori, R. Conformational Effects in Protein Electrospray-Ionization Mass Spectrometry: Native Protein Esi-Ms. Mass Spec. Rev. 2016, 35, 111–122. [CrossRef]

19. Konijnenberg, A.; van Dyck, J.F.; Kailing, L.L.; Sobott, F. Extending Native Mass Spectrometry Approaches to Integral Membrane Proteins. Biol. Chem. 2015, 396, 991–1002. [CrossRef]

20. Kaltashov, I.A.; Bobst, C.E.; Abzalimov, R.R. Mass Spectrometry-Based Methods to Study Protein Architecture and Dynamics: MS-Based Methods to Study Protein Architecture and Dynamics. Protein Sci. 2013, 22, 530–544. [CrossRef]

21. Mehmood, S.; Allison, T.M.; Robinson, C.V. Mass Spectrometry of Protein Complexes: From Origins to Applications. Annu. Rev. Phys. Chem. 2015, 66, 453–474. [CrossRef]

22. Natalello, A.; Santambrogio, C.; Grandori, R. Are Charge-State Distributions a Reliable Tool Describing Molecular Ensembles of Intrinsically Disordered Proteins by Native MS? J. Am. Soc. Mass Spectrom. 2017, 28, 21–28. [CrossRef] [PubMed]

23. Testa, L.; Brocca, S.; Grandori, R. Charge-Surface Correlation in Electrospray Ionization of Folded and Unfolded Proteins. Anal. Chem. 2011, 83, 6459–6463. [CrossRef] [PubMed]

24. Hall, Z.; Robinson, C.V. Do Charge State Signatures Guarantee Protein Conformations? J. Am. Soc.

Mass Spectrom. 2012, 23, 1161–1168. [CrossRef] [PubMed]

25. Kaltashov, I.A.; Mohimen, A. Estimates of Protein Surface Areas in Solution by Electrospray Ionization Mass Spectrometry. Anal. Chem. 2005, 77, 5370–5379. [CrossRef] [PubMed]

26. Wu, H.; Zhang, R.; Zhang, W.; Hong, J.; Xiang, Y.; Xu, W. Rapid 3-Dimensional Shape Determination of Globular Proteins by Mobility Capillary Electrophoresis and Native Mass Spectrometry. Chem. Sci. 2020, 11, 4758–4765. [CrossRef]

27. Longhi, S.; Receveur-Bréchot, V.; Karlin, D.; Johansson, K.; Darbon, H.; Bhella, D.; Yeo, R.; Finet, S.; Canard, B. The C-Terminal Domain of the Measles Virus Nucleoprotein Is Intrinsically Disordered and Folds upon Binding to the C-Terminal Moiety of the Phosphoprotein. J. Biol. Chem. 2003, 278, 18638–18648. [CrossRef]

28. Habchi, J.; Mamelli, L.; Darbon, H.; Longhi, S. Structural Disorder within Henipavirus Nucleoprotein and Phosphoprotein: From Predictions to Experimental Assessment. PLoS ONE 2010, 5, e11684. [CrossRef]

29. Yuan, A.; Rao, M.V.; Veeranna; Nixon, R.A. Neurofilaments and Neurofilament Proteins in Health and Disease. Cold Spring Harb. Perspect. Biol. 2017, 9, a018309. [CrossRef]

30. Herrmann, H.; Aebi, U. Intermediate Filaments: Structure and Assembly. Cold Spring Harb. Perspect. Biol. 2016, 8, a018242. [CrossRef]

31. Studier, F.W. Protein Production by Auto-Induction in High-Density Shaking Cultures. Protein Expr. Purif. 2005, 41, 207–234. [CrossRef]

32. Holehouse, A.S.; Das, R.K.; Ahad, J.N.; Richardson, M.O.G.; Pappu, R.V. CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins. Biophys. J. 2017, 112, 16–21. [CrossRef] [PubMed]

33. Dosztanyi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: Web Server for the Prediction of Intrinsically Unstructured Regions of Proteins Based on Estimated Energy Content. Bioinformatics 2005, 21, 3433–3434. [CrossRef] [PubMed]

34. Micsonai, A.; Wien, F.; Bulyáki, É.; Kun, J.; Moussong, É.; Lee, Y.-H.; Goto, Y.; Réfrégiers, M.; Kardos, J. BeStSel: A Web Server for Accurate Protein Secondary Structure Prediction and Fold Recognition from the Circular Dichroism Spectra. Nucleic Acids Res. 2018, 46, W315–W322. [CrossRef] [PubMed]

35. Irvine, G.B. Determination of Molecular Size by Size-Exclusion Chromatography (Gel Filtration). Curr. Protoc. Cell Biol. 2000, 6, 5.5.1–5.5.16. [CrossRef]

36. Marsh, J.A.; Forman-Kay, J.D. Sequence Determinants of Compaction in Intrinsically Disordered Proteins. Biophys. J. 2010, 98, 2383–2390. [CrossRef]

37. Brocca, S.; Testa, L.; Sobott, F.; Šamalikova, M.; Natalello, A.; Papaleo, E.; Lotti, M.; De Gioia, L.; Doglia, S.M.; Alberghina, L.; et al. Compaction Properties of an Intrinsically Disordered Protein: Sic1 and Its Kinase-Inhibitor Domain. Biophys. J. 2011, 100, 2243–2252. [CrossRef]

38. Sinha, S.K.; Chakraborty, S.; Bandyopadhyay, S. Thickness of the

Hydration Layer of a Protein from Molecular Dynamics Simulation. J. Phys. Chem. B 2008, 112, 8203–8209. [CrossRef]

39. Pal, S.; Bandyopadhyay, S. Effects of Protein Conformational Flexibilities and Electrostatic Interactions on the Low-Frequency Vibrational Spectrum of Hydration Water. J. Phys. Chem. B 2013, 117, 5848–5856. [CrossRef]

40. Rath, A.; Davidson, A.R.; Deber, C.M. The Structure of ?Unstructured? Regions in Peptides and Proteins: Role of the Polyproline II Helix in Protein Folding and Recognition. Biopolymers 2005, 80, 179–185. [CrossRef]

41. Dunker, A.K.; Lawson, J.D.; Brown, C.J.; Williams, R.M.; Romero, P.; Oh, J.S.; Oldfield, C.J.; Campen, A.M.; Ratliff, C.M.; Hipps, K.W.; et al. Intrinsically Disordered Protein. J. Mol. Graph. Model. 2001, 19, 26–59. [CrossRef]

42. Adzhubei, A.A.; Sternberg, M.J.E. Left-Handed Polyproline II Helices Commonly Occur in Globular Proteins. J. Mol. Biol. 1993,

229, 472–493. [CrossRef] [PubMed]

43. Berman, H.M. The Protein Data Bank. Nucleic Acids Res. 2000, 28, 235–242. [CrossRef] [PubMed]

44. Sickmeier, M.; Hamilton, J.A.; LeGall, T.; Vacic, V.; Cortese, M.S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V.N.; et al. DisProt: The Database of Disordered Proteins. Nucleic Acids Res. 2007, 35, D786–D793. [CrossRef] [PubMed]

45. Piovesan, D.; Tabaro, F.; Micˇetic´, I.; Necci, M.; Quaglia, F.; Oldfield, C.J.; Aspromonte, M.C.; Davey, N.E.; Davidovic´, R.; Dosztányi, Z.; et al. DisProt 7.0: A Major Update of the Database of Disordered Proteins. Nucleic Acids Res 2017, 45, D219–D227. [CrossRef] [PubMed]

46. Vymeˇtal, J.; Vondrášek, J.; Hlouchová, K. Sequence Versus Composition: What Prescribes IDP Biophysical Properties? Entropy

2019, 21, 654. [CrossRef]

47. Blocquel, D.; Habchi, J.; Gruet, A.; Blangy, S.; Longhi, S. Compaction and Binding Properties of the Intrinsically Disordered C-Terminal Domain of Henipavirus Nucleoprotein as Unveiled by Deletion Studies. Mol. BioSyst. 2012, 8, 392–410. [CrossRef]

48. Zeng, X.; Ruff, K.M.; Pappu, R.V. Competing Interactions Give Rise to Two-State Behavior and Switch-like Transitions in Charge-Rich Intrinsically Disordered Proteins. bioRxiv 2022. [CrossRef]

Supplementary information

Table S1. Size of the semi-axes of ellipsoids assimilated to the model proteins studied in this work

Protein variant		*a (nm)*	*b (nm)*	*c (nm)*	*Surface* (nm^2)*
N_{TAIL}	Low κ	3.00 ± 0.03	2.21 ± 0.03	1.78 ± 0.04	67.5 ± 1.9
	Wt	2.90 ± 0.02	2.17 ± 0.03	1.76 ± 0.03	64.4 ± 1.5
	High κ	2.67 ± 0.01	2.02 ± 0.04	1.66 ± 0.06	55.8 ± 2.1
NFM	Low κ	3.29 ± 0.02	2.83 ± 0.04	2.54 ± 0.07	104.5 ± 3.4
	Wt	3.15 ± 0.03	2.78 ± 0.04	2.54 ± 0.07	100.0 ± 3.3
	High κ	2.55 ± 0.01	2.62 ± 0.02	2.49 ± 0.09	81.9 ± 6.2
PNT4	Low k	2.92 ± 0.04	1.69 ± 0.07	1.36 ± 0.03	48.1 ± 2.3
	Wt	2.91 ± 0.04	1.64 ± 0.08	1.34 ± 0.03	46.8 ± 2.4
	High k	2.34 ± 0.02	1.89 ± 0.05	1.61 ± 0.07	47.4 ± 2.3

* calculated by Thomsen's approximation (maximal discrepancy ~1%)

N_{TAIL}

```
Low-k     MHHHHHHTTERDISRAVGPDQAQVSFLHGKQSENRLPELGGKEKERVDQSRGEARESYED   60
wt        MHHHHHHTTEDKISRAVGPRQAQVSFLHGDQSENELPRLGGKEDRRVKQSRGEARESYRE   60
High-k    MHHHHHHTTRKRISRAVGPRQAQVSFLHGRQSRNRLPRLGGKRKRRVRQSRGRADESYED   60

Low-k     TGPSRASDARAAHLPTGTPLDIDTASRSSQDPQRSERSADALLRLQAMAGISERQGSDTD   120
wt        TGPSRASDARAAHLPTGTPLDIDTASESSQDPQDSRRSADALLRLQAMAGISEEQGSDTD   120
High-k    TGPSEASDAEAAHLPTGTPLDIDTASESSQDPQDSEESADALLELQAMAGISEDQGSDTD   120

Low-k     TPIVYNDRNLLDGS  134
wt        TPIVYNDRNLLDGS  134
High-k    TPIVYNDENLLDGS  134
```

NFM

```
Low-k     MHHHHHHDGAKGSDKEKIAVNGEVEGKEEVKQETKEKGSGDEREKGVVTNGLDLSPAKEK   60
wt        MHHHHHHKGAKGSRKEDIAVNGEVEGKEEVEQETKEKGSGREEEKGVVTNGLDLSPADEK   60
High-k    MHHHHHHKGAKGSRKKKIAVNGKVKGKKKVKQKTKKKGSGRKKKKGVVTNGLKLSPAKED   60

Low-k     EGGDKSEEKVVVTETVEKITSEGGRGATEYITKSVTVTQEVKEHKETFKEKLVSTEKVEK   120
wt        KGGDKSEEKVVVTKTVEKITSEGGDGATKYITKSVTVTQKVEEHEETFEEKLVSTKKVEK   120
High-k    DGGDESEEEVVVTETVEEITSEGGDGATEYITESVTVTQEVEEHEETFEEELVSTEEVED   120

Low-k     VTSHAIVEKVTQSDGS  136
wt        VTSHAIVKEVTQSDGS  136
High-k    VTSHAIVEEVTQSDGS  136
```

PNT4

```
Low-k     MHHHHHHKETPRVDRDKSLMQDSCKEGGVPERLPMLSREFRCSGSDRPIIQELREKGSHP   60
wt        MHHHHHHEETPDVRRKDSLMQDSCKRGGVPKRLPMLSEEFECSGSDDPIIQELEREGSHP   60
High-k    MHHHHHHKRTPRVRRKKSLMQRSCKRGGVPKRLPMLSRKFRCSGSKRPIIQRLEDEGSHP   60

Low-k     GGSLDLREPPQSSGNSRNQPDRQLKTGDAASPGGVQRPGTPMPKSEIMPIKEGS   114
wt        GGSLRLREPPQSSGNSRNQPDRQLKTGDAASPGGVQRPGTPMPKSRIMPIKKGS   114
High-k    GGSLELEEPPQSSGNSENQPDEQLDTGDAASPGGVQEPGTPMPDSEIMPIDDGS   114
```

Figure S1. Amino acid sequences of model IDPs. The wt, low-κ and high-κ variants within each of the three sets have been aligned to highlight the sequence design. Positively charged residues are shown in blue, negatively charged residues in red, yellow highlights proline residues. Sequence alignments were carried out with Clustal Omega (*Sievers et al, 2011*, doi:10.1038/msb.2011.75)

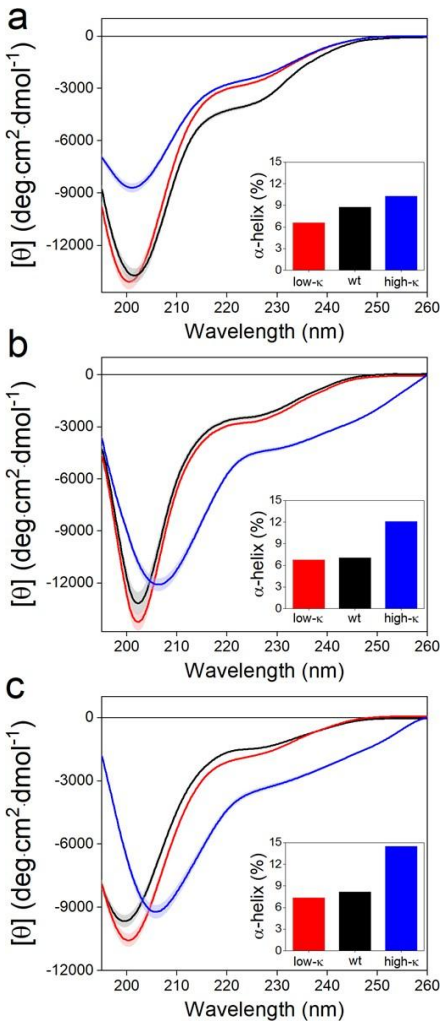


Figure S2. Far-UV CD spectra of model IDPs. CD spectra of NTAIL (a), NFM (b), and PNT4 (c) variants. CD analyses were performed in PBS. The shadowed area refers to the standard deviation of three independent measurements. In black the spectrum of wt proteins, while red and blue are used for the spectra of low-κ and high-κ permutants, respectively. Inset of each panel: percentage of α-helical content as estimated by BeStSel software (*Micsonai et al., 2018*; doi:10.1093/nar/gky497).

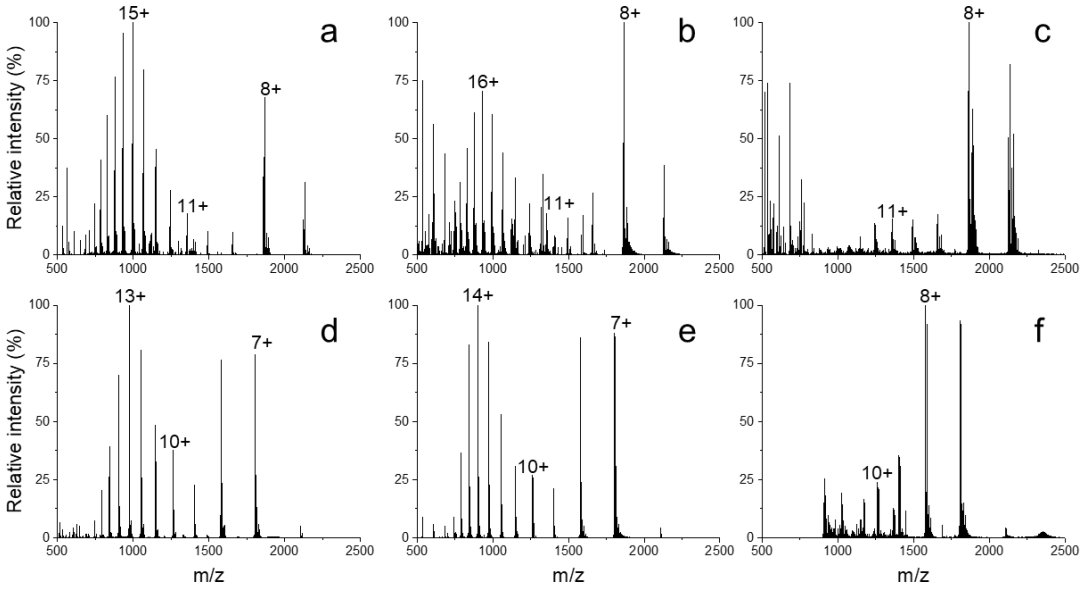


Figure S3. Native MS analyses of model IDPs. NanoESI-MS spectra of NFM (a-c) and PNT4 (d-f) variants acquired under non-denaturing conditions (50 mM ammonium acetate, pH 7.0). Low-κ variants (a, d), wt variants (b, e) and high-κ variants (c, f) are reported.

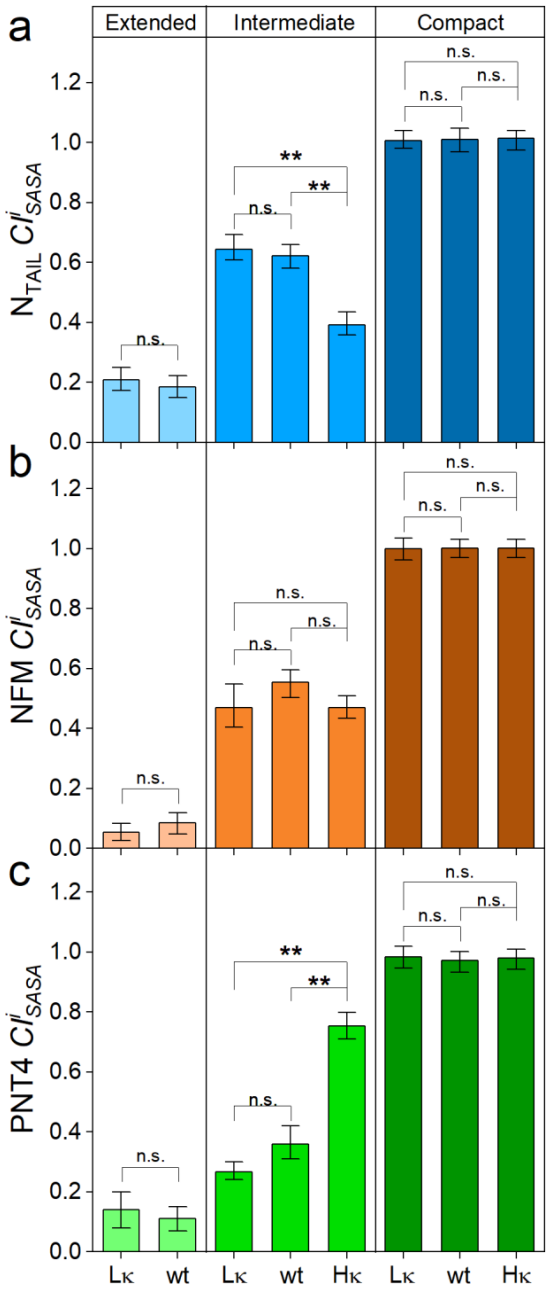


Figure S4. Compactness of conformational components of model IDPs studied by native MS. Bar chart reporting CIi of NTAIL (a), NFM (b) and PNT4 (c). Mean values of three independent measurements are shown with error bars indicating standard deviations. Statistical analyses were carried out using Welch's t-test (n.s.: $p > 0.05$, *: $p < 0.05$, **: $p < 0.01$).

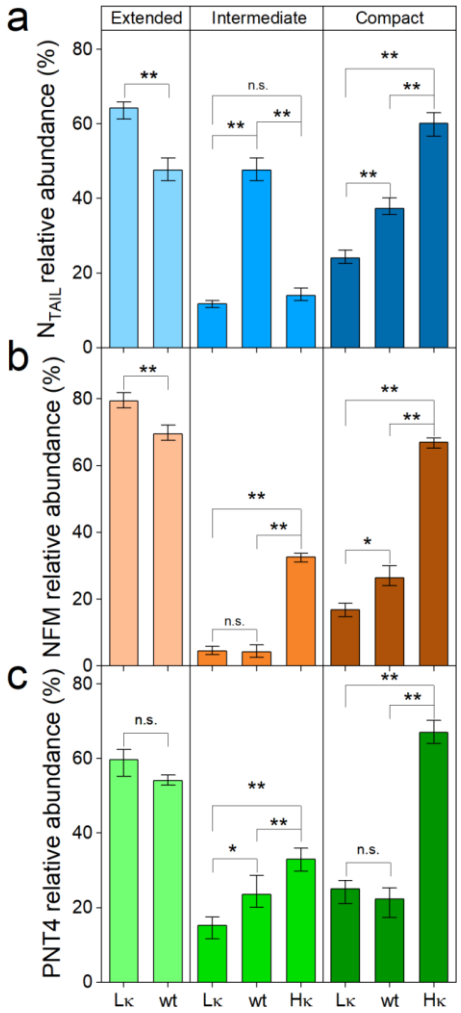


Figure S5. Relative abundance of conformational components of model IDPs studied by native MS. Bar chart reporting the relative abundance of single conformational components (extended; intermediate; compact) of NTAIL (a), NFM (b) and PNT4 (c). Mean values of three independent measurements are represented with error bars indicating standard deviations. Statistical analyses were carried out using Welch's t-test ($p > 0.05$ n.s., * $p < 0.05$, **$p < 0.01$).

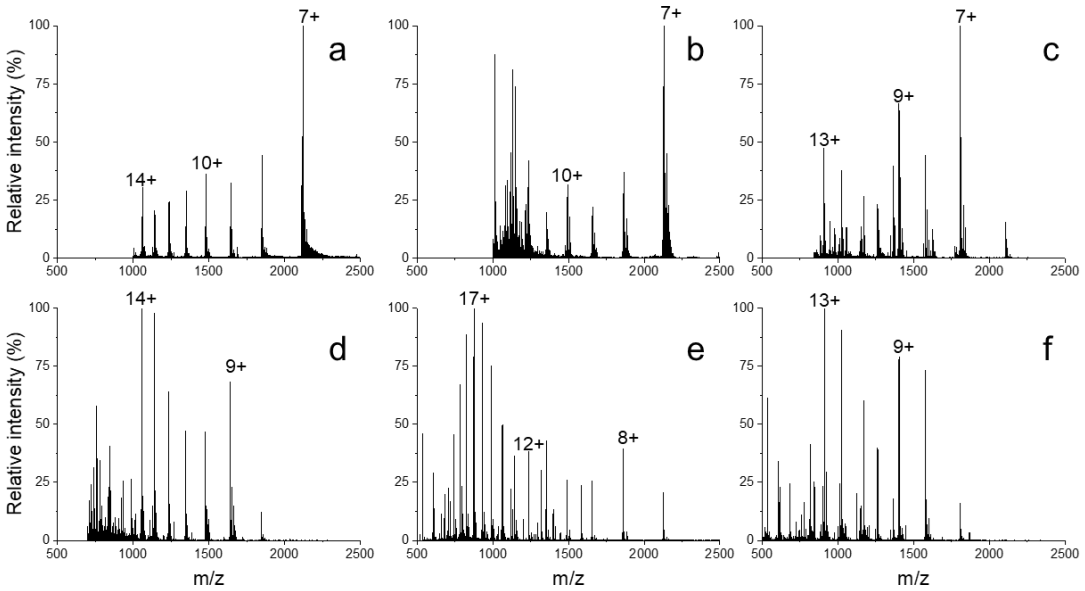


Figure S6. NanoESI-MS spectra of model IDPs under conditions affecting electrostatic interactions. NanoESI-MS spectra of high-κ variants of NTAIL (a, d), NFM (B, E) and PNT4 (C, F) acquired in the presence of 200 mM ammonium acetate pH 7.0 (a-c), and in the presence of 1% formic acid (d-f).

Article in preparation

Charge distribution affects the liquid-liquid phase separation of the highly-charged N-terminal domain of human topoisomerase 1

Greta Bianchi[1], Marco Mangiagalli[1], Diletta Ami[1], Sonia Longhi[2], Antonino Natalello[1], Peter Tompa[3]*, Stefania Brocca[1]*

[1]Department of Biotechnology and Biosciences, University of Milano-Bicocca, Piazza della Scienza 2, 20126, Milan, Italy
[2]Lab. Architecture et Fonction des Macromolécules Biologiques (AFMB), Aix-Marseille University and CNRS, UMR 7257, 13288, Marseille, France; sonia.longhi@univ-amu.fr
[3]VIB-VUB Center for Structural Biology, VIB, Pleinlaan 2, 1050 Brussels, Belgium

*Correspondence: stefania.brocca@unimib.it (S.B.); peter.tompa@vub.be (P.T); +39 02 6448 3518 (S.B.); +32 2 629 1924 (P.T)

ABSTRACT

Liquid-liquid phase separation (LLPS) underlies the formation of biomolecular condensates, which are involved in several cellular processes from the formation of membraneless organelles to chromatin remodeling. Intrinsically disordered regions (IDRs) and nucleic acids, especially RNA, are mainly involved in the formation of biomolecular condensates because of their biased composition and multivalency. Understanding the forces driving the LLPS phenomenon is important considering the key role played by membraneless organelles in many biological functions and pathologies. Electrostatic interactions are known to support protein LLPS, but the role played by the distribution of charged residues remains to be clarified.
The N-terminal domain (NTD) of TOP1, an enzyme able to relax DNA supercoils and hence involved DNA replication and chromatin remodeling, is intrinsically disordered and highly-charged. We found that the charge patterning of NTD follows a clear evolutionary trend, with vertebrates showing the most regular alternation of opposite charges and fungi the most pronounced segregation. Here we document the ability of human TOP1 NTD (*h*NTD) to *in-vitro* undergo LLPS through electrostatic interactions. Using *h*NTD permutants, we show that increasing clustering of charged residues affects the sensitivity of condensates to salt and RNA. In conclusion, our results highlight that an optimal charge distribution promoting in-vitro LLPS exists and that charge segregation impairs phase separation leading to abnormal condensates insensitive to RNA.

Keywords (5):, intrinsically disordered protein, protein coacervate, biomolecular condensate, electrostatic interaction; TOP1.

Abbreviations: **FCR**: fraction of charged residues; **FRAP**: Fluorescence Recovery After Photobleaching; **FTIR**: Fourier transform infrared (spectroscopy); **GFP**: green fluorescent protein; ***h*TOP1**: human DNA topoisomerase I; **IDP/IDR**: intrinsically disordered protein/region; **IMAC**: immobilized-metal affinity chromatography;; **LLPS**: liquid-liquid phase separation; **NCPR**: net charge per residue; **NLS**: nuclear localization sequence; ***h*NTD**: N-terminal domain of *h*TOP1; **PB**: phosphate buffer; **ROI**: region of interest.

1. Introduction

Liquid-liquid phase separation (LLPS) plays a role in the formation of biomolecular condensates, *i.e.* liquid droplets containing proteins and other macromolecules such as nucleic acids [1, 2]. Biomolecular condensates are involved in several biological processes including the organization of the intracellular space through the formation of membraneless organelles (MLOs) such as the nucleolus and the Cajal bodies [3], chromatin reorganization [4], reduction of cellular noise [5], and sensing [6]. Biomolecular condensates can undergo liquid-to-solid state transitions which, in some cases, lead to protein aggregation [7-10] with implications in various pathologies such as cancer [11-13] and neurodegenerative diseases [14-16].

At the molecular level, LLPS is driven by interactions between biomolecules (protein or nucleic acids) with multivalent (un)structured domains. The involvement of intrinsically disordered proteins and regions (IDPs/IDRs) [17-19] in biocondensation phenomenon is related not only to the lack of a well-defined three-dimensional structure, but also to their peculiar composition and sequence, often containing charged residues, and low-complexity regions [20]. Several computational and experimental studies on LLPS processes highlight the importance of both hydrophobic and electrostatic interactions such as π–π stacking, cation–π, and charge–charge interactions [21-26]. Electrostatic interactions depend on several factors, not only the total positive and negative charges or the net charge of a protein, but also the charge distribution along its linear sequence [27]. Charge distribution can be simply described by the κ parameter: when κ value is 0, the opposite charges alternate uniformly one by one along the sequence; when κ value is 1, the opposite charges are segregated into two clusters [27]. Although the role of κ parameter in IDR compactness has been extensively studied [28-31], its impact on LLPS propensity is still largely unknown. A few computational studies with experimental data suggest that electrostatic interchain interactions mostly occur when the κ value is $0.1 < \kappa < 0.3$; in fact, when $\kappa = 0$, interchain interactions are

weak, whereas when κ = 1 the most favored contacts are intrachain interactions [32, 33].

In this work, we investigated the role of charge distribution in LLPS propensity of the N-terminal domain (NTD) of human topoisomerase 1 (*h*TOP1). DNA topoisomerases (EC 5.6.2.1) are ubiquitous nuclear enzymes active in relaxing DNA supercoils introduced during DNA replication and transcription [34-37]. Overall, *h*TOP1 is composed of four domains including a disordered N-terminal domain (hereafter referred to as *h*NTD, residues 1-214), a highly conserved core domain bearing the DNA interaction motifs (residues 215-635), a linker region (residues 636-712) and a highly conserved C-terminal domain (residues 713-765), which contains the active site tyrosine [38]. *h*NTD is highly charged, contains the nuclear localization sequences (NLSs) and plays a role in several physiological functions including relaxation activity and camptothecin sensitivity [39-47]. Long IDRs and LLPS are common features in various proteins involved in DNA replication: in *Drosophila melanogaster* the long IDRs enclosed in ORC, Cdc6 and Cdt1 drive LLPS thereby regulating chromosome recruitment and the initiation of DNA replication [48]. The presence of a long N-terminal IDR in *h*TOP1 might be involved in its condensation, as is the case of the C-terminal disordered domain of yeast and human *h*TOPIIs [49]. To date, the direct involvement of the *h*TOP1 NTD in LLPS phenomena has been just hypothesized. The objective of the present work is dual: first we aim at assessing the ability of *h*NTD to undergo phase separation *in vitro*, in relation to stimuli or "order parameters" such as ionic strength, pH changes, and interaction with RNA, and then at investigating the effects of charge patterning on its LLPS behavior.

Our data indicate that *h*NTD *per se* can undergo phase separation through electrostatic interactions, which can be strongly influenced by charge clustering.

Overall, the differences observed among eukaryotic orthologous of TOP1 suggest that charge distribution within NTD might add a level of regulation to the formation of biomolecular condensates.

2. Material and Methods

2.1. Sequence analyses of *h*TOP1

Sequence analyses of *h*TOP1 (UniProt code: P11387) were carried out using CIDER [50], IUPred2A [51, 52] and FuzDrop [53, 54] softwares. Nuclear localization sites (NLSs) were manually identified based on literature data [40].

*h*TOP1 orthologs were retrieved using OrthoDB v10 (https:// www.orthodb.org), a database of putative orthologous gene sequences from genomes of eukaryotes, prokaryotes and viruses [55], and the sequence of *h*TOP1 gene (NP_003277.1) as a query. OrthoDB identifies 1649 putative orthologs of the *h*TOP1 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 *h*TOP1. The representative sequences were manually selected from a protein BLAST (blastp) analysis [56] using the *h*TOP1 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 [57]. For each NTD sequence, FCR, NCPR and κ value were calculated using CIDER [58]. The phylogenetic tree of full length TOP1s was constructed by the Neighbour Joining algorithm [59] using MEGAX [60].

2.2. Cloning and mutagenesis

The sequence encoding the N-terminal domain of *h*TOP1 (*h*NTD, residues 1-214) was optimized for expression in *Escherichia coli* cells, chemically synthesized (Genscript, Piscataway, NJ, USA) and cloned into the pET-21a vector (EMD, Millipore, Billerica, MA, USA) between the *Nde*I and *Xho*I sites (Jena Biosciences, Jena, Germany). We designed two variants of *h*NTD, namely medium κ (Mκ-NTD) and high κ (Hκ-NTD), by modifying the distribution of charged residues only within the first 116 amino acid residues (*h*NTD$^{1\text{-}116}$), while maintaining unchanged

the NLSs-containing region (*h*NTD$^{117-214}$). The overall strategy of permutation of *h*NTD$^{1-116}$ was previously described in [30, 31]*;* 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 Mκ-NTD and Hκ-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 DH5α™ (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 *Spe*I site, yielding the following constructs: pET-21a[*h*NTD-GFP], pET-21a[Mκ-NTD-GFP] and pET-21a[Hκ-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 [61].

2.3. Production and purification of *h*NTD and its variants

Pre-cultures of BL21(DE3) cells were grown at 37°C in Lennox medium to OD_{600} 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 OD_{600} 0.6-0.8 before the induction with 1% lactose or 0.1 mM isopropryil-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. [30]. 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 (50 mM sodium phosphate, 50-300 mM NaCl, pH 10.5) or a 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.

2.4. 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. The effect of condensation was assayed at different concentrations of protein, i.e. 25, 50 and 100 μM, 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 T*i* 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). Fluorescence Recovery After Photobleaching (FRAP) analyses were carried out on stationary droplets. These were bleached (100% intensity, scan speed 0.125

sec/frame, 121.09 ms bleach, optical zoom 5x) and observed every second for one minute. To study droplet recovery, data were normalized employing the formula:

$$\text{Normalized Recovery} = \frac{(FROI_{BG})}{(FNB_{BG})} * \frac{Mean\,(FNB_{BG}^{BeforeBleaching})}{Mean\,(FROI_{BG}^{BeforeBleaching})}$$

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; $Mean\,(FROI_{BG}^{BeforeBleaching})$ corresponds to the average $FROI_{BG}$ before the bleaching; $Mean\,(FNB_{BG}^{BeforeBleaching})$ represents the average FNB_{BG} before the bleaching.

2.5. RNA-induced LLPS

For RNA experiments, recombinant proteins were suspended in PBS. Confocal microscopy analyses were carried out on GFP-fused proteins as described in 2.4. Poly(A) RNA (P9403, Sigma Aldrich) was added at different concentrations (0, 0.5, 2, 5 mg/mL) to *h*NTD-GFP, M*κ*-NTD-GFP, and Hκ-NTD-GFP (final 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), exciting the GFP with a 488 nm argon laser. Morphological and FRAP analyses were conducted as described in 2.4.

2.6. Fourier transform infrared (FTIR) spectroscopy

The secondary structure of *h*NTD, Mκ-NTD, and Hκ-NTD was monitored by FTIR measurements 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. The formation of protein condensates was induced by either pH jump or RNA addition (0.5 mg/mL) as described in previous paragraphs. pH jump experiments were carried out at 50 mM of NaCl for *h*NTD and M*κ*-NTD proteins (protein concentration: 50 μM) and 150 mM of NaCl for H*κ*-NTD variant (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. For RNA experiments, after the addition of 0.5 mg/mL of poly(A) RNA, samples were incubated for 3 minutes at room temperature and centrifuged at 20,800 g at 4°C for 10 minutes. The supernatant was collected and analyzed by FTIR, while the pellet was washed with PBS before the measurement. Three microliters of each sample were deposited on the single reflection diamond element of the ATR device [62, 63]. The ATR/FTIR spectra were collected after solvent evaporation 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: resolution 2 cm^{-1}, scan speed 25 kHz, 1024 scan co-additions, triangular apodization, and dry-air purging.

ATR-FTIR absorption spectra were corrected for buffer contribution, normalized at the Amide I (1700–1600 cm^{-1}) band area and were smoothed using the Savitsky-Golay method (25 points) before second derivative calculation. Spectral analyses were performed with the Resolutions-Pro software (Varian Australia Pty Ltd., Mulgrave, Victoria, Australia). At least three independent measurements were performed for each condition.

3. Results

3.1 Three evolutionary conserved charge patterns are present in the NTDs of TOP1 sequences

Sequence analysis with predictive tools indicates that *h*NTD (214 residues in length) is disordered and prone to phase separate (**Fig. 1A**). The *h*NTD is depleted in hydrophobic and aromatic residues (Phe: 3, Trp: 3 and Tyr: 3) and enriched in acidic (Glu + Asp: 59) and basic (Lys + Arg: 70) residues, with a κ value of 0.128. To investigate whether this charge distribution is a peculiarity of the *h*NTD or is widespread among eukaryotic organisms, we analyzed the NTD sequences of TOP1 from 55 different eukaryotic taxa, sampling the broad evolutionary pathway

taken by this protein. As expected for IDRs, which are characterized by a biased sequence composition, all the selected NTDs are highly charged, rich in Ser and Ala and poor in Cys and aromatic residues (**Fig. 1C**). The good balance between residues of opposite charge is evidenced by a very low Net Charge Per Residue (*NCPR,* mean value: 0.036 ± 0.035;[27]), which appears as an evolutionarily conserved trait despite the high variability of NTD length and sequence (**Table S1**). The distribution of charged residues appears to be correlated with the phylogeny of TOP1s (**Fig. 1D**) and, on the basis of κ values of NTDs, it is possible to distinguish three different groups: group 1 including vertebrates, with $\kappa < 0.15$ (in blue in **Fig. 1D**); group 2 with $0.15 < \kappa < 0.3$ including protozoa, insects and some plants, (in gray in **Fig. 1D**), and finally group 3 including fungi and the rest of plants and protozoa, with $\kappa > 0.3$ (in red in **Fig. 1D**). Overall, charged residues in NTDs from vertebrates seem to follow a typical pattern, not conserved among lower eukaryotes, suggesting that this distribution could play a physiological role related to functions unique to vertebrates. The wide charge distribution naturally occurring in the NTD of TOP1 sequences could make this IDR a good model to study the effect of charge distribution in LLPS propensity. Thus, we addressed the question as to whether the *h*NTD, characterized by a low κ value, is able to undergo LLPS.

3.2 Electrostatic interactions drive the *in-vitro* LLPS of *h*NTD

To investigate the ability of *h*NTD to undergo LLPS, we recombinantly produced it and we triggered the phase separation through an abrupt change of pH (pH jump), under controlled salt concentration. Dramatic pH changes are an efficient method to trigger protein condensation and hence to study its kinetics *in vitro* [64]. To analyze its propensity to condense, *h*NTD was prepared in an alkaline buffer at pH 10.5 and 50 mM of NaCl and then, by adding a small amount of HCl, the solution pH was lowered to 7.5. Upon the pH jump, phase separation caused turbidity in the samples, which was monitored at 600 nm over time. We observed that the turbidity of the samples increases to a maximum within the first 10 minutes and then

stabilizes or slowly decreases (**Fig. 2A**). The protein concentration affects the maximum turbidimetry value, but not its kinetics (**Fig. 2A**). To test if the protein condensation is driven by electrostatic charges, we also performed the pH jump in the presence of growing salt concentrations. A marked loss of turbidity was observed both at 150 mM and 300 mM of NaCl (**Fig. 2A**), suggesting an important role of electrostatic interactions in *h*NTD condensation.

The morphology and liquid behavior of pH-induced condensates were investigated by confocal fluorescence microscopy and FRAP analysis on *h*NTD-GFP. The fusion of *h*NTD with GFP did not affect its condensation propensity and kinetics (**Fig. S2**). *h*NTD-GFP condensates obtained at low salt concentration , *i.e.* 50 mM NaCl, form droplets with an average area of 0.65 ± 0.35 μm^2 (**Fig. 2B),** a strong tendency to coalesce (**Fig. 2C**) and liquid nature as indicated by recovered fluorescence after photobleaching (~ 80% within 1 min) (**Fig. 2D**). The size of the droplets decreases with increasing salt concentration; their areas are halved from 50 to 150 mM NaCl and reduced ten times with 300 mM NaCl (**Fig. 2B)**. Overall, our results show that ionic strength counteracts *h*NTD condensation, indicating that electrostatic charges play a key role in its LLPS propensity.

The secondary structure content of condensates was monitored by Fourier transform infrared (FTIR) spectroscopy directly applied on protein condensates harvested by centrifugation of samples subjected to pH jump (**Fig. 2E**). Before the pH jump, the second derivative spectrum of *h*NTD shows a main component at ~1644 cm^{-1}, which can be assigned to disordered structures [61, 65]. The low intensity peaks at ~1677 and at ~1688 cm^{-1}, which occurs in the spectral regions of turns and β-sheets (black line in **Fig. 2E**), suggest the existence of some residual secondary structure [63, 65-67]. After the pH jump, the FTIR spectrum of the supernatant (blue line in **Fig 2E**) is similar to that observed before the pH jump with the main peak shifted from ~1644 to ~1652 cm^{-1}, suggesting that the pH jump induces only small changes in the secondary structure of *h*NTD. The FTIR spectrum of the pellet (red line in **Fig. 2E**) collected after the pH jump shows the

main peak at ~1652 cm^{-1} and a new component at low intensity around 1630 cm^{-1} that suggests the formation of intermolecular β-sheets [65, 68, 69]. Overall, our data indicate that pH-induced condensation of *h*NTD does not involve drastic conformational changes in its disordered structure and that intermolecular β-sheet structures contribute only minimally to stabilize protein condensates.

In conclusion, we observed that the pH jump triggers the liquid condensation of the disordered, highly charged *h*NTD. Considering the effect of salts, we hypothesize that, under low ionic strength conditions, *h*NTD behaves as a weak polyampholyte readily involved in a net of intermolecular interactions. These interactions are lost as soon as the ions in solution shield the charges of the polyampholyte, leading to the dissolution of the droplets. After having documented the propensity of *h*NTD to undergo LLPS, we next addressed the question as to whether charge distribution plays a role in this phenomenon, which would point to possible functional differences between NTD regions at different evolutionary distances from *h*NTD.

3.3. Charge segregation strongly influences *h*NTD condensation

A closer look inside *h*NTD revealed that the nuclear and nucleolar localization sequences (NLSs), which play a key role in nuclear transport, are located within the region spanning residues 117-214 [40]. NLSs are rich in basic amino acids and represent sequence constraints that affect the charge distribution of this region (κ=0.224, **Fig. 1B**). Therefore, to assay the role of charge distribution in LLPS, we decided to modify the charge pattern of *h*NTD^{1-116} (κ=0.039) according to the characteristics emerged from our sequence analysis.

Two synthetic sequences of *h*NTD^{1-116} designated as "medium κ" (κ: 0.226) and "high κ" (κ: 0.373) were obtained applying a permutation scheme previously described in [30, 31]. Briefly, the permutants share with *h*NTD^{1-116} the same position of non-charged residues and the number of charged ones, differing only in the distribution of positive and negative residues. Each synthetic sequence was then fused to the wild-type *h*$NTD^{117-214}$ (*i.e.* the NLS-containing region). The new NTD

constructs containing the medium κ or high κ version of *h*NTD$^{1\text{-}116}$ are referred to as Mκ-NTD (κ: 0.224) and Hκ-NTD (κ: 0.298), respectively (**Fig. S1**). Our attempts to go further in opposite charged residues clustering have been unsuccessful, due to either expression or purification and manipulation of the resulting proteins (data not shown). The permutation of charged residues does not affect the degree of disorder and the predicted LLPS propensity of Mκ-NTD and Hκ-NTD (**Fig. S3**). The Mκ-NTD and Hκ-NTD variants were recombinantly produced in *E. coli* cells and their propensity to form protein condensates were assayed as previously described for the wild-type *h*NTD.

The experimental conditions used to trigger the condensation of *h*NTD were also applied to Mκ-NTD and Hκ-NTD. Mκ-NTD reveals a greater propensity to undergo pH-induced LLPS compared to *h*NTD, as indicated by both turbidimetry and morphological analyses (**Fig. 3**). In turbidimetry experiments, Mκ-NTD reaches optical densities 2-3 times higher than *h*NTD at the same protein concentration. Note also the faster kinetics of Mκ-NTD condensation. For example, at 100 μM Mκ-NTD reaches the highest optical density after only 5 minutes, to be compared to the∼30 minutes required for *h*NTD (**Fig. 3A**). Although droplets of Mκ-NTD-GFP exhibit the same liquid behavior in terms of coalescence and FRAP, they are 10 times larger than those observed for *h*NTD-GFP condensates (average droplet area of 6.95 ± 2.67 μm^2 at 50 mM NaCl) (**Fig. 3B-D**). Similarly to *h*NTD-GFP, Mκ-NTD-GFP droplets become smaller by increasing the ionic strength, as inferred from morphological and turbidimetric analyses, especially at 300 mM NaCl (**Fig. 3A, B**). The FTIR spectrum of bulk Mκ-NTD is similar to that of *h*NTD with a main peak at ~1644 cm^{-1} (black line in **Fig. 3E**), suggesting that the charge segregation does not affect *per se* the protein secondary structure.The FTIR spectrum of the supernatant collected after the pH jump is similar to that obtained for *h*NTD (**Fig. 3E** and **2E**, respectively). Interestingly, the pellet spectrum of Mκ-NTD (blue line in **Fig. 3E**) displays an intense component at ~1630 cm^{-1},

ascribable to intermolecular β-sheet structures, which likely underlies the higher LLPS propensity of M*κ*-NTD , compared to *h*NTD (**Fig.2E**). Overall, the charge distribution of M*κ*-NTD seems to optimize the intermolecular electrostatic interactions that drive the LLPS.

The propensity of H*κ*-NTD to undergo LLPS has been very difficult to investigate, due to the low yield of recombinant production in *E. coli* cells and its tendency to spontaneously condense or aggregate at protein concentration and buffer ionic strength at which the other NTD variants remain soluble. For this reason, the highest concentration used in this work for H*κ*-NTD is just 25 μM. Despite the limitation of using this single concentration, we were able to obtain a reliable picture of the pH-induced behavior of LLPS and the effect caused by salts. An atypical scenario emerges from the pH-jump experiments, as the turbidity of samples at low salt concentration is very low compared to *h*NTD and M*κ*-NTD, and, *viceversa*, unexpectedly increases with salt concentrations (**Fig. 4A**). Confocal microscopy analysis indicates that pH-induced condensates of H*κ*-NTD-GFP have a peculiar shape and are unable to coalesce (**Fig. 4B, C**), although FRAP experiments still reveal a liquid behavior (~70% recovery in one minute) (**Fig. 4D**). In the FTIR spectrum of H*κ*-NTD, two main components are present: a peak at ~1654 cm^{-1}, which can be assigned to α-helices and/or random coils, and a peak at ~1639 cm^{-1}, ascribable to β-sheet structures with possible contributions of random coils. Thus, after band narrowing by second derivative calculation, *h*NTD and Mκ-NTD display a featureless band in the 1620-1660 cm^{-1} region while Hκ-NTD exhibits two well resolved components, as observed for folded proteins (*Barth, 2007*). These data suggest that the H*κ*-NTD variant has a more ordered secondary structure content than *h*NTD and M*κ*-NTD (black line in **Fig. 4E**). Therefore, it can be inferred that charge segregation can induce the formation of intra-chain electrostatic interactions, which stabilize more ordered conformations, as already reported for other high-*κ* permutants [30, 31]. After the pH jump at 150 mM NaCl, the FTIR spectrum of the supernatant displays the same components detected in the

bulk solution prior to the pH jump (red line in **Fig. 4E**) with, however, a higher intensity of the 1654 cm^{-1} peak. In the pellet the peak of the intermolecular β-sheets is shifted to ~1626 cm^{-1}, suggesting that the H*κ*-NTD condensates are characterized by a unique β-sheet structure that differs from that observed in *h*NTD and M*κ*-NTD (blue line in **Fig. 4E**).

Overall, our data indicate that charge clustering confers to H*κ*-NTD a stronger polyampholyte trait compared to M*κ*-NTD, with interactions probably less multivalent and less fluctuating. In the case of H*κ*-NTD, intramolecular interactions may be so prevalent that only higher salt concentrations are compatible with condensation, swelling the chain and promoting intermolecular interactions which result, however, in condensates with aberrant characteristics.

3.4 Phase separation of *h*NTD is modulated by addition of RNA

RNA is a highly negatively charged molecule. In the case of condensations driven by electrostatic interactions, the addition of RNA molecules, irrespective of their sequence, may have an effect that is comparable to that of adding salt [25]. It is expected that low concentrations of RNA can improve the network of electrostatic interactions and induce LLPS, while high concentrations disrupt electrostatic and polar interactions, thus reducing the propensity to condense [25, 70]. The interaction with RNA may also depend on the amino acid composition of proteins, whereby proteins rich in lysine or arginine may undergo complex coacervation phenomena that have been well described [71]. To assess how RNA modulates the LLPS propensity of *h*NTD-GFP, experiments were conducted on each NTD variant in PBS and in the presence of different poly(A) RNA concentrations ranging from 0.5 to 5 mg/mL. At low RNA concentration (0.5 mg/mL), both *h*NTD-GFP and M*κ*-NTD-GFP undergo LLPS, generating biomolecular condensates that are similar in terms of size and liquid nature (**Fig. 5** and **S4A, S4C**). By contrast, H*κ*-NTD-GFP is not responsive to RNA (**Fig. 5E**), unless incubated for 90 minutes at room temperature with 0.5 mg/mL of RNA, a condition where we observed the

formation of amorphous condensates (**Fig. 5F**), which recover 60 % of fluorescence after photobleaching (**Fig. S4E**). To investigate the composition and the secondary structure of biomolecular condensates, we performed FTIR analysis on condensates collected by centrifugation. In the Amide I region, the RNA (black dotted line in **Fig. 5B, D, G**) second derivative spectrum displays a broad absorption at around ~1649 cm^{-1} and a well resolved component at ~1602 cm^{-1} that can be employed as a RNA marker band since proteins do not have detectable peaks at this wavenumber. In the case of *h*NTD, the ~1602 cm^{-1} component was undetectable in the supernatant (red line in **Fig. 5B**) while it is well evident in the pellet spectrum (blue line in **Fig. 5B)**, which is characterized also by a broad component at ~1641 cm^{-1}. After subtraction of the RNA absorption (**Fig. S4B**), the pellet displays a main broad component at ~1630 cm^{-1}, assigned to β-sheets while the supernatant shows a main component at ~1652 cm^{-1}, ascribable to α-helical and random coil structures. Similar results were observed for the M*κ*-NTD, where, however, the RNA marker band at ~1602 cm^{-1} was detected both in the supernatant and in the pellet (**Fig. 5D** and **Fig. S4D**).

A different behavior was instead observed for the H*κ*-NTD absorption, where the RNA marker band was found only for the supernatant. In particular, the H*κ*-NTD supernatant spectrum is characterized by two main components at ~1654-1652 cm^{-1} and at ~1639 cm^{-1} (**Fig. 5G** and **Fig. S4F**), assigned as described above. In the pellet, the main component at ~1626 cm^{-1} (**Fig. 5G** and **Fig. S4F**) occurs in the typical spectral region of β-sheet structures in protein aggregates.

Each variant presents a peculiar and dose-dependent RNA response profile. In the case of *h*NTD-GFP, 2 mg/mL of RNA counteracts the formation of biomolecular condensates (**Fig. 5A**). By contrast, at this RNA concentration, M*κ*-NTD undergoes LLPS with condensates very similar to that observed at low RNA concentration (**Fig. 5C**). Overall, these results are in agreement with our previous hypothesis, *i.e.* that the intermolecular interactions are stronger in M*κ*-NTD than in the wild-type

protein. A completely anomalous behavior is observed in the case of Hκ-NTD, since the condensation does not appear to be induced by RNA (**Fig. 5E**).

4. Discussion

The evolutionary trend leads the NTD sequence to progressively attenuate its charge clustering. Understanding the forces governing LLPS is a hot topic because of the numerous biological and pathological processes involving biomolecular condensates [1, 2, 7, 12, 13, 72-74]. This work shows that the distribution of charged residues plays an important role in the LLPS of the highly-charged *h*NTD. More precisely, *h*NTD can be considered as a polyampholyte, since not only 60% of its residues are charged, but also the positively and negatively charged residues are well balanced. Moreover, their linear distribution along the protein sequence yields a value of $\kappa = 0.128$. A closer look in the NTD of *h*TOP1 orthologs indicates that despite their sequence hypervariability among eukaryotes it is still possible to delineate a very clear evolutionary trend in the extent of opposite charge clustering. The most marked charge clustering is found among fungi and unicellular eukaryotes ($\kappa > 0.3$), whose NTD is virtually coincident with the NLS-containing region. The evolutionary pathway of NTD goes progressively towards acquiring an extra N-terminal structure juxtaposed to the NLSs and bearing charges that alternate very regularly, as observed in vertebrates ($\kappa < 0.15$). In between, insects and plants present an intermediate situation, with κ values of their NTDs comprised between 0.15 and 0.3. Our phylogenetic analysis thus enabled us to identify at least three values of the κ parameter, representative of three distinct stages of the NTD evolutionary process.

The next step in the present work consisted in experimentally assessing possible differences in the *in-vitro* LLPS propensity of three synthetic variants of the NTD polypeptide, *i.e. h*NTD, Mκ- and Hκ-NTD, in an attempt of exploring the nature of the driving forces of their condensation.

Electrostatic interactions are the driving force of the hNTD LLPS and moderate charge clustering enhances its condensation.

Our results show that *h*NTD, corresponding to the wild-type domain, has an intrinsic propensity to undergo LLPS through electrostatic interactions, which is tunable by pH, ionic strength and RNA (**Fig.6**). In our system, the linear distribution of electrostatic charges within the NTD^{1-116} also appears to play a non-negligible role, as indicated by the different response to ionic strength and pH of the *h*NTD variants. At low salt concentration, we observed that the M*κ*-NTD droplets are larger than those formed by *h*NTD (**Fig.6**). FTIR spectroscopy analysis of the condensates provides information on the nature and intensity of the forces at play: the interactions stabilizing the M*κ*-NTD condensates are stronger compared to that of *h*NTD droplets, as indicated by the intensity of the peak at ~1630 cm^{-1}. These behaviors are in agreement with the computational simulations carried out by Li and Chan [33], which predicted weak intermolecular electrostatic interactions for peptides with low *κ*-values and progressively more favored intermolecular interactions for peptides with higher *κ*-values. By contrast, the H*κ*-NTD is unable to undergo LLPS at low ionic strength (**Fig.6**). The FTIR analysis shows a more ordered secondary structure for H*κ*-NTD compared to the other two variants ascribable to the formation of intramolecular interactions. Overall, charge clustering may counteract LLPS by favoring intramolecular interactions and reducing intermolecular connectivity [75]. The behavior of H*κ*-NTD corroborates our previous insights derived from the analysis of proteins in the PhasePro database [76] capable of undergoing electrostatics-based LLPS: in spite of the small size of the set, we argued that increasing the value of *κ* cannot favor LLPS indefinitely, with excessive charge clustering rather beginning to hamper the propensity to form intermolecular interactions [32]. Our hypothesis is herein corroborated by the experimentally observed LLPS behavior of the NTD variants in the presence of salts.

Marked charge clustering causes suboptimal electrostatic interactions and an aberrant condensation behavior of NTD.

Although the general effect of salts in inducing LLPS is not yet fully understood, the response of our model proteins at higher salt concentration better highlights the anomalous properties of H*κ*-NTD condensates. Indeed, higher ionic strengths prevent LLPS of *h*NTD and M*κ*-NTD, probably due to the shielding of charges by the ions, which thus counteract intermolecular interactions, as observed in weak polyampholytes (**Fig.6**) [77]. At higher salt concentrations, the Hκ-NTD variant exhibits an anomalous LLPS behavior, characterized by the formation of amorphous condensates forming droplet crumbs that have no propensity to coalescence, despite retaining liquid properties, likely due to the high number of Lys residues involved in electrostatics (**Fig.6**) [71]. This is reminiscent of the PR-polyrG assemblies, whose alike morphology is attributed to a kinetically arrested phase separation originating large networks where molecules are still able to diffuse [78]. Unlike what has been hypothesized to explain the reentrant phase diagram of FUS, TDP-43, Brd4, Sox2, and A11, hydrophobic interactions can poorly explain H*κ*-NTD condensation in the presence of higher salt concentrations. Indeed, the content of *h*NTD is definitely low in hydrophobic residues, and the number and position of non-charged residues is unchanged in all the polypeptides tested. One possible explanation for H*κ*-NTD salt response is that this construct requires the presence of ions to initially break intramolecular interactions thus enabling intermolecular contacts to be established. However, the aberrant condensates triggered at higher ionic strength may be due to suboptimal electrostatic interactions, probably stronger, less transient and fluctuating than in the case of NTD variants carrying less clustered charges.

Charge clustering attenuates the response to *RNA during NTD coacervation*.

Concerning the RNA response, it is known that it critically regulates the phase behavior of proteins, through both non-specific interactions, governed by the

polyelectrolytic nature of its backbone, and specific ones, depending on its sequence. In the context of non-specific interactions, RNA has been described as a destroyer of protein-protein electrostatic networking or as a trigger factor. Low RNA:protein ratios have been reported to promote phase separation in liquid droplets, possibly disrupting inappropriate intramolecular contacts or forming electrostatic interactions with cationic peptides, while high ratios were found to prevent droplet formation *in vitro* [70, 79, 80]. Consistently with these findings, we observed that low poly(A) RNA concentrations were very effective in inducing the formation of *h*NTD biomolecular condensates (**Fig.6**), while high RNA concentrations (≥ 2 mg/mL) prevented it. Similarly, M*κ*-NTD responded to RNA addition (**Fig.6**) in a dose-dependent manner, yet higher poly(A) concentrations were needed to suppress the coacervation (5 mg/mL). Moreover, while M*κ*-NTD/RNA condensates retained approximately the same size of pH-jump induced droplets (6.35 ± 3.21 μm^2; 6.95 ± 2.67 μm^2 respectively), *h*NTD condensates obtained in presence of poly(A) display a five-fold increase in size if compared to pH-induced ones, even at low salt concentrations. This suggests that charge clustering may attenuate the differences related to the condensation stimuli, reflecting on the different sensitivity of the NTD towards environmental fluctuations. The H*κ*-NTD response to RNA further corroborates this hypothesis. Indeed, its condensation is not induced or controlled by poly(A) addition (**Fig.6**), pointing to the loss of this layer of regulation, likely exerted on condensation mechanisms even in a physiological context. As a matter of fact, our results suggest that the local RNA concentration present in different subcellular compartments may play a key role in the LLPS of *h*NTD, as observed for instance for other RNA binding proteins such as FUS and TDP43 [70].

Thus, besides having a major role in determining dimensions of IDP/IDRs [28-31], net charge and patterning also govern their LLPS behavior, as to date just hypothesized [33, 81-83].

Biological relevance of NTD architecture and charge distribution.

Based on the results of our *in-vitro* experiments, it is tempting to hypothesize that the ability of *h*NTD to undergo LLPS may also drive LLPS of *h*TOP1 *in vivo*. Although this hypothesis awaits experimental verification, it is widely known that long IDRs induce LLPS in proteins involved in DNA replication and transcription processes, including yeast TOP2 and human TOP2α and TOP2β [26, 48, 49]. Taken together, our data also suggest the main evolutionary constraints which a disordered, supercharged protein such as NTD is subjected to so as to preserve its ability to undergo LLPS. Among these, the extent of charge separation plays a key, multifaceted role. Indeed, charge patterning seems directly involved in promoting intermolecular interactions over intramolecular and solvent ones, supporting a dynamic network of contacts, possibly maintaining a good responsiveness to salt and RNA concentration and, in general, to the surrounding molecular environment. Altering the degree of clustering can produce a variety of effects ranging from slightly interfering with RNA responsiveness (as in the case of M*κ*-NTD), to completely deregulating LLPS behavior (as in the case of H*κ*-NTD).

The heterogeneity of the NTD charge distribution contrasts with the presence of NLSs, which instead imposes tight sequence constraints including charge clustering. Although the data available on the NLS identification are few, important differences emerge from the comparison of the *h*NTD with their counterparts in budding yeasts. In budding yeasts, the NTD coincides with the NLS region, while in human and in vertebrates an extra sequence, with a low *κ*-value, is juxtaposed upstream of the NLS regions. This observation, together with our results on *h*NTD variants, which mimic the naturally occurring variations in charge distribution, suggests that the evolution of TOP1 has also been shaped by mechanisms controlling its LLPS.

Author contributions

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**; 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.

Acknowledgements

This work was partly supported by University of Milano-Bicocca (Fondo di Ateneo, grant ID 29192 to S.B.). GB acknowledges support for a PhD research fellow from the University of Milano-Bicocca. MM and DA acknowledge support for a post doc research fellow (Assegno di Ricerca) position from the University of Milano-Bicocca. The authors thank Brunno Rocha Levonne, for his assistance in confocal microscopy experiments, Junaid Ahmed and Joris Van Lindt for their support in RNA-mediated LLPS experiments.

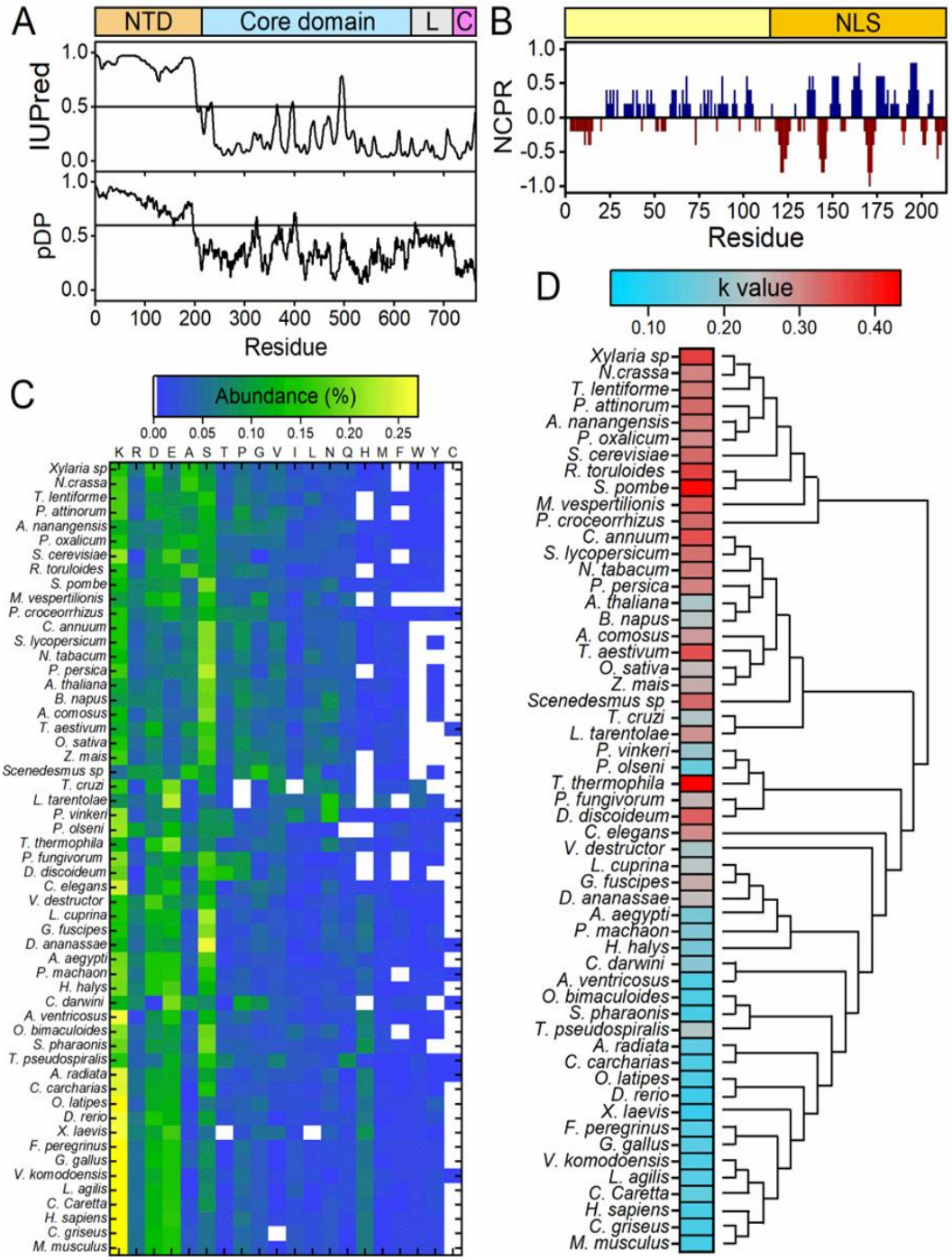


Figure 1. The charge distribution of the N-terminal disordered domain of TOP1 is not conserved. **A**) Architecture, disorder and LLPS predictions of

*h*TOP1. This enzyme is composed of: N-terminal domain (NTD), core domain, linker (L) and catalytic (C) domains. The IUPRED2A disorder prediction score is reported in the middle panel. The black line indicates the cutoff value of 0.5, residues with a score above this cutoff are predicted to be disordered. The LLPS propensity score predicted with *FuzDrop* software is reported in the bottom panel. The black line indicates the cutoff value of 0.6, residues with a score above this cutoff are predicted to undergo LLPS. **B**) Plot of the net charge per residue (NCPR) over a 5-residue sliding window for *h*NTD as obtained using CIDER. **C**) Heat map representation of the amino acid composition of the NTD of TOP1 orthologous sequences. The precise boundaries of the NTDs were obtained from Pfam analysis. **D**) Correlation between the κ values of the NTD and the phylogeny of TOP1 orthologous sequences.

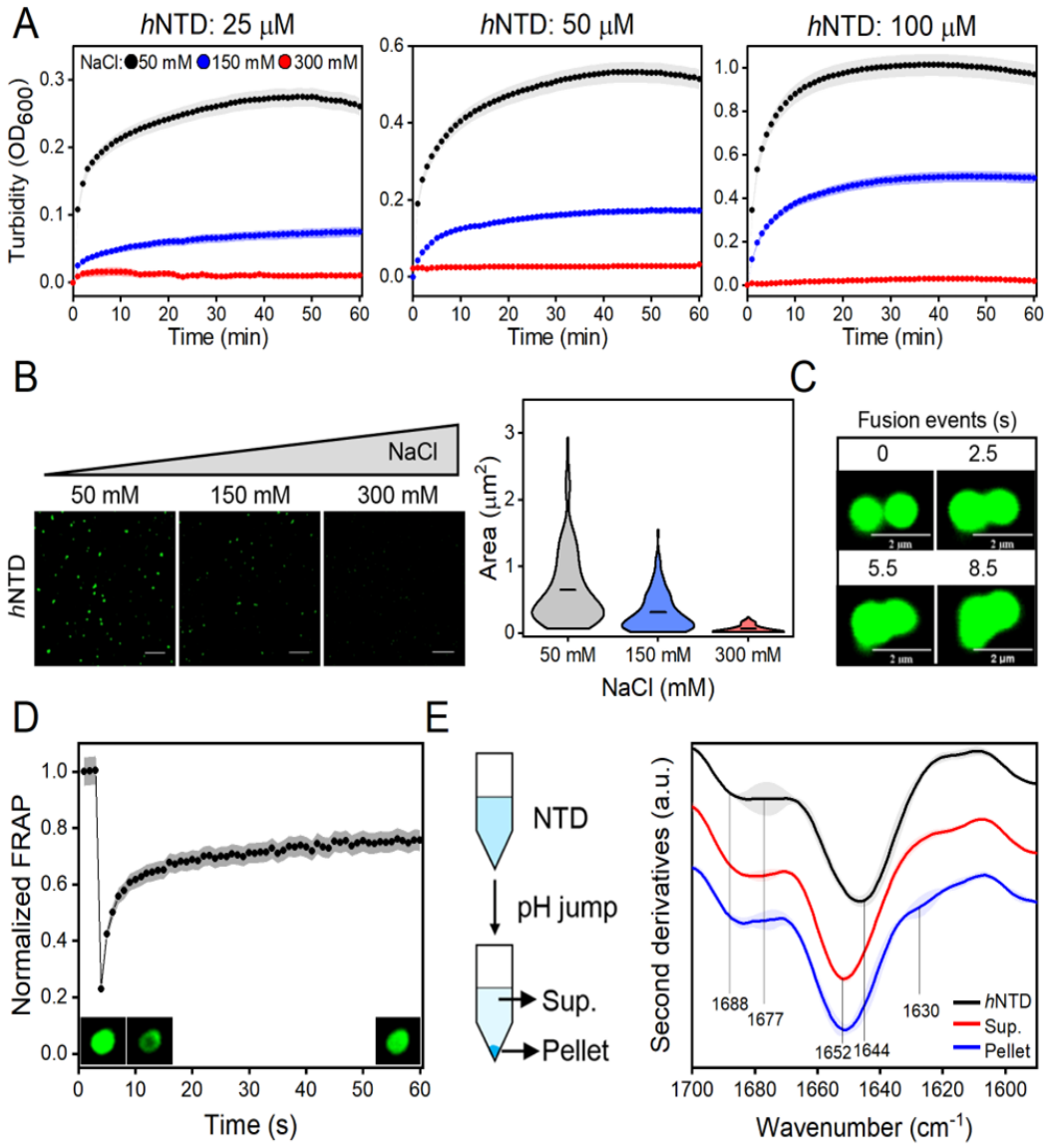


Figure 2. The LLPS of *h*NTD is driven by electrostatic interactions. A) Turbidity plots of *h*NTD after pH jump. Experiments were carried out on three different protein concentrations at three different salt concentrations. **B)** Morphological analysis of *h*NTD-GFP condensates. Confocal microscopy images of *h*NTD-GFP (50 μM) condensates were taken after the pH jump at different salt concentrations. Scale bar: 10 μm. 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 μm. **D)** FRAP analysis of *h*NTD-GFP pH-induced droplets (50 μM, n = 23). Confocal microscopy images show droplets before, after bleaching and at the final phase of recovery (1 min). The fluorescence signal was normalized from 0 to 1. **E)** Secondary structure analysis of *h*NTD condensates. LLPS was triggered by pH jump and protein condensates were separated from supernatant by centrifugation. The FTIR second derivative spectra of *h*NTD were collected before the pH jump and after the jump on the supernatant and the pellet. Mean second derivatives of the absorption spectra from three experiments were reported and the shadowed area refers to the standard deviation of the data (n = 3).

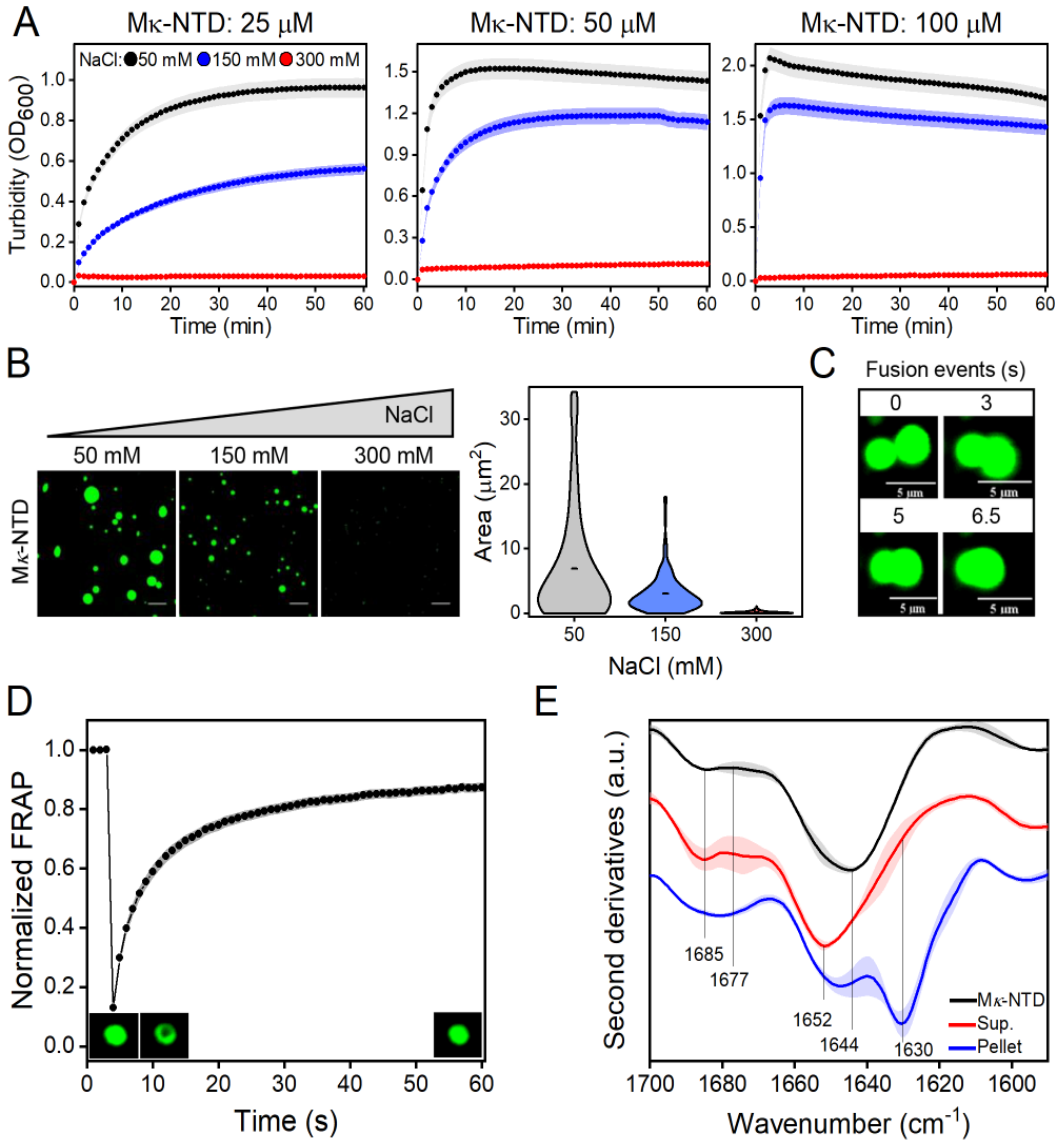


Figure 3. M*κ*-NTD has a strong propensity to condense. A) Turbidity plots of M*κ*-NTD after pH jump. Experiments were carried out on three different protein concentrations at three different salt concentrations. **B)** Morphological analysis of M*κ*-NTD-GFP condensates. Confocal microscopy images of M*κ*-NTD-GFP (50 μM) were taken after the pH jump at different salt concentrations. Scale bar: 10 μm. Each image is representative of three independent measurements. The area of

M*κ*-NTD-GFP droplets was calculated using the FIJI ImageJ software v2.0 (n = 150 droplets). **C)** Fusion events of M*κ*-NTD-GFP droplets induced by pH jump. Scale bar, 5 μm. **D)** FRAP analysis of M*κ*-NTD-GFP pH-induced droplets (50 μM, n=75). Confocal microscopy images show droplets before, after bleaching and at the final phase of recovery (1 min). The fluorescence signal was normalized from 0 to 1. **E)** Secondary structure analysis of M*κ*-NTD condensates. LLPS was triggered by pH jump and protein condensates were separated from supernatant by centrifugation. The FTIR spectra of M*κ*-NTD were collected before the pH jump and after the jump on the supernatant and the pellet. Mean second derivatives of the absorption spectra from three experiments were reported and the shadowed area refers to the standard deviation of the data (n = 3).

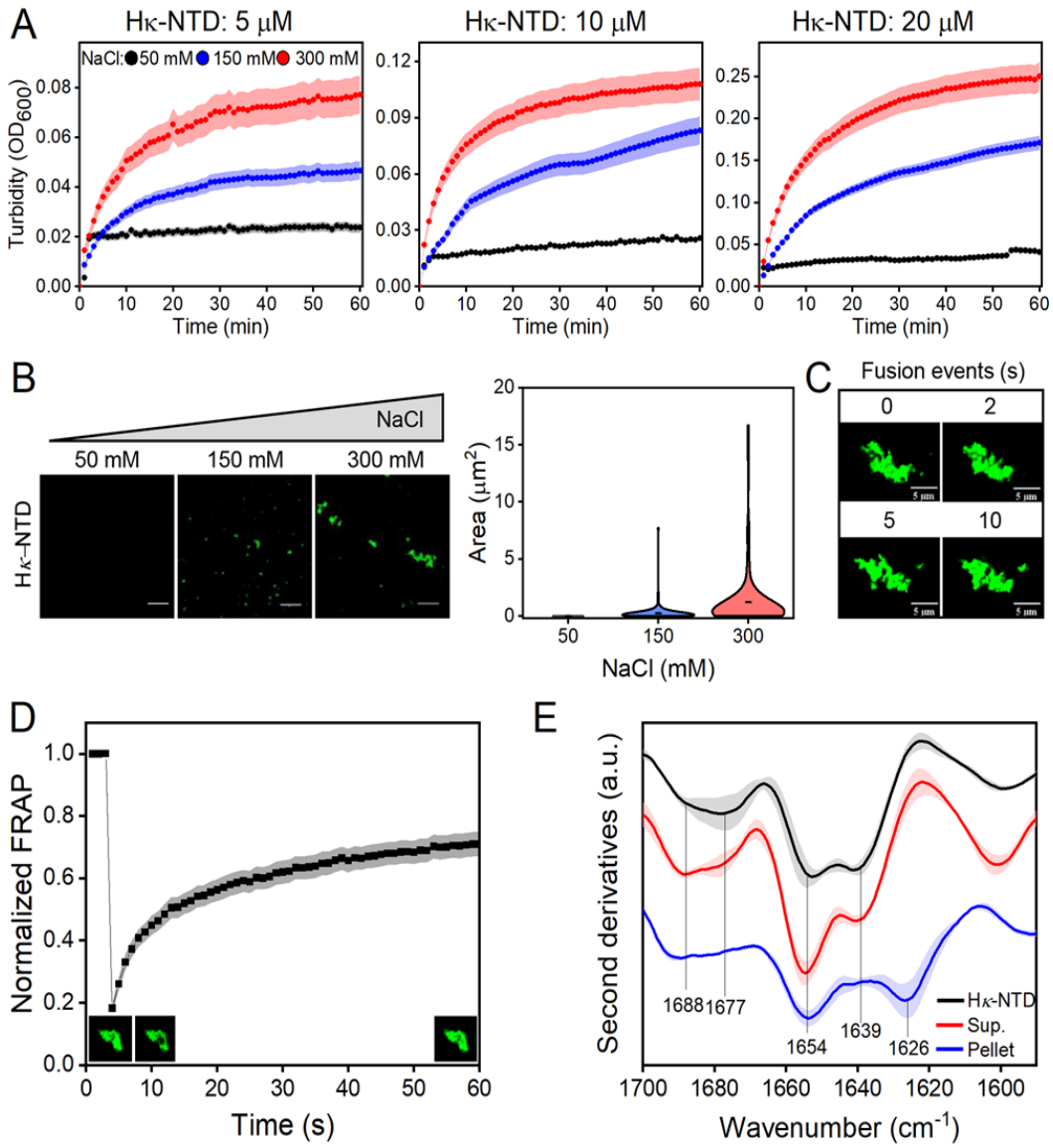


Figure 4. H*κ*-NTD forms anomalous protein condensates. A) Turbidity plots of H*κ*-NTD after pH jump. Experiments were carried out on three protein concentrations, at three different salt concentrations. **B)** Morphological analysis of H*κ*-NTD-GFP condensates. Confocal microscopy images of H*κ*-NTD-GFP (25 μM) were taken after the pH jump at different salt concentrations. Scale bar: 10 μm. Each image is representative of three independent measurements. The area of H*κ*-NTD-GFP condensates was calculated using the FIJI ImageJ software v2.0 (n =

150 droplets). **C)** Absence of fusion events of H*κ*-NTD-GFP condensates. Scale bar, 5 μm. **D)** FRAP analysis of H*κ*-NTD-GFP condensates induced by pH jump (25 μM, n = 28). Confocal microscopy images show droplets before, after bleaching and at the final phase of recovery (1 min). The fluorescence signal was normalized from 0 to 1. **E)** Secondary structure analysis of H*κ*-NTD condensates. LLPS was triggered by pH jump and protein condensates were separated from supernatant by centrifugation. The FTIR spectra of H*κ*-NTD were collected before the pH jump and after the jump on the supernatant and the pellet. Mean second derivatives of the absorption spectra from three experiments were reported and the shadowed area refers to the standard deviation of the data (n = 3).

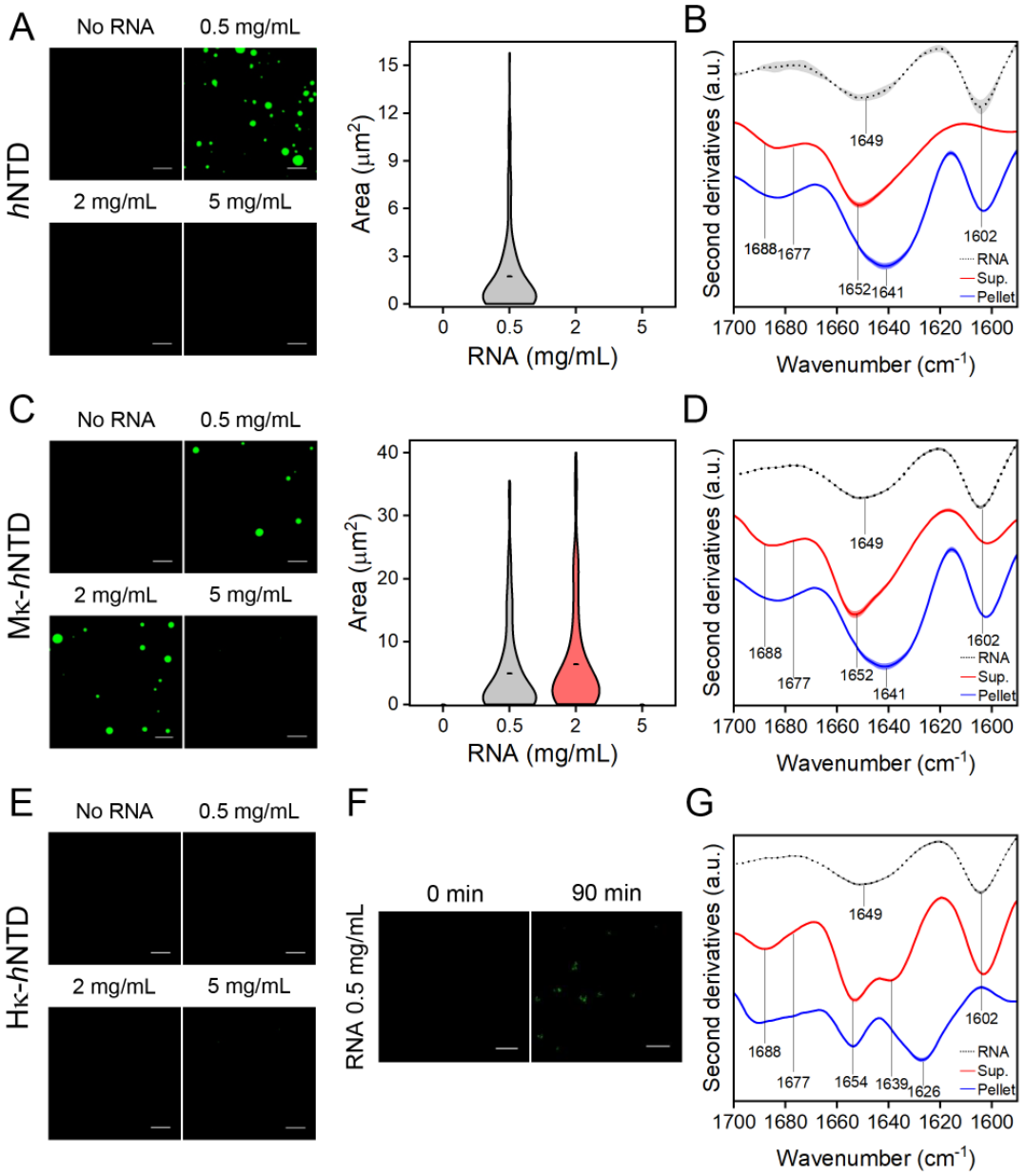


Figure 5. Phase separation of *h*NTD is modulated by addition of RNA. **A, C, E)**

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 μM) were taken after a three-minute incubation with different

poly(A) concentrations (0.5, 2, 5 mg/mL). Scale bar 10 μm. The area 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 triggered by the addition of 0.5 mg/mL poly(A) and protein condensates were separated from supernatant by centrifugation. The FTIR spectra of RNA and after it addition on the supernatant and the pellet. Mean second derivatives of the absorption spectra from two experiments were reported and the shadowed area refers to the standard deviation of the data (n = 2). The black dotted line corresponds to free RNA (0.5 mg/mL). All spectra are corrected for PBS contribution. **F)** Amorphous condensates of H*κ*-NTD-GFP appear 90 minutes after the addition on RNA (0.5 mg/mL). Scale bar 10 μm.

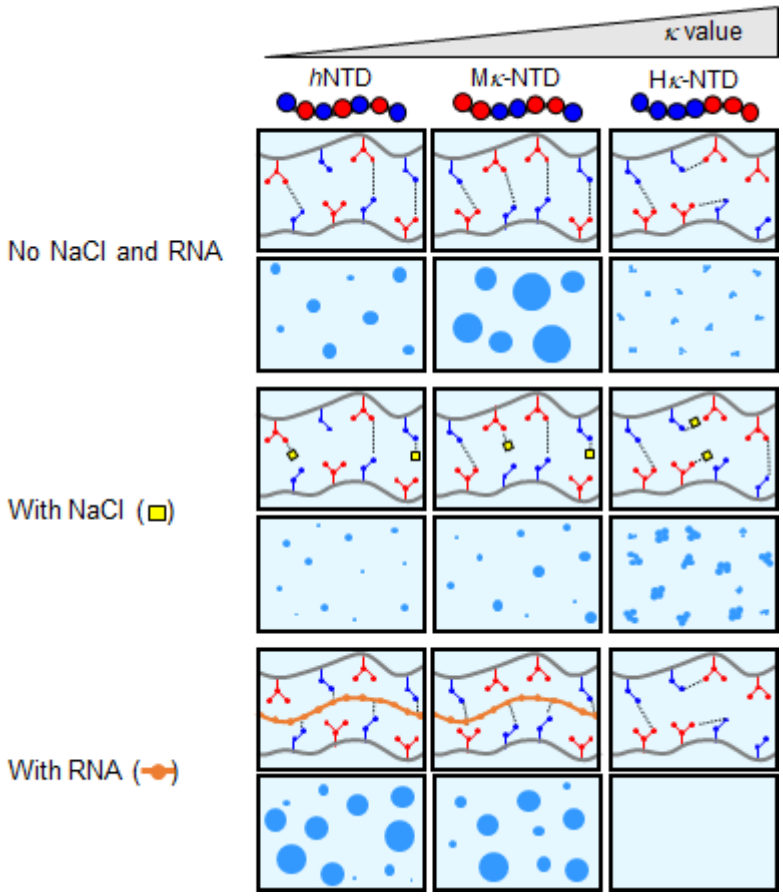


Figure 6. Schematic representation of LLPS behavior of *h*NTD and its charge permutants, in response to salt and RNA addition. *h*NTD is a highly charged polyampholyte with a low κ value as indicated by the alternation of positively (blue spheres) and negatively charged (red spheres) residues. The distribution of charged residues was gradually increased in charge permutants, obtaining the Mκ-NTD and Hκ-NTD variants. Condensation occurs when intermolecular interactions (dash line) prevail over the solvation forces and intramolecular interactions, resulting in liquid droplets, represented as light blue spheres of different size. **Upper panel)** Upon pH-jump, in the low-salt regime and absence of RNA, condensation of *h*NTD leads to the formation of small liquid droplets. Mκ-NTD shows a greater LLPS propensity, as indicated by the larger droplet size, likely motivated by more

favorable intermolecular interactions. The further increase of *κ*-values in H*κ*-NTD intensifies the intramolecular interactions, producing amorphous and non-coalescent condensate "crumbs". **Middle panel)** As expected for condensates driven by electrostatic forces, the addition of salts interferes with pH-induced LLPS due to the shielding of charges by salt ions. The droplets formed by wild-type and M*κ*-NTD are smaller than those obtained at low ionic strength. In contrast, the H*κ*-NTD droplets are larger, which may be due to the loss of intramolecular interaction and the increased propensity for intermolecular interactions. **Lower panel)** RNA is assimilable to a negatively charged polyelectrolyte that can bind positively charged proteins and strongly enhances their propensity for phase separation. Consequently, wild-type and M*κ*-NTD respond to low concentrations of RNA, with morphologically similar and larger droplets than those formed with the pH jump. On the other hand, H*κ*-NTD is unable to form droplets in the presence of RNA, probably due to the persistence of intramolecular interactions.

References

1 Banani, S. F., Lee, H. O., Hyman, A. A. and Rosen, M. K. (2017) Biomolecular condensates: organizers of cellular biochemistry. Nature reviews Molecular cell biology. **18**, 285-298

2 Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B. and Van Den Bosch, L. (2018) Protein phase separation: a new phase in cell biology. Trends in cell biology. **28**, 420-435

3 Shin, Y. and Brangwynne, C. P. (2017) Liquid phase condensation in cell physiology and disease. Science. **357**, eaaf4382

4 Welsh, T. J., Shen, Y., Levin, A. and Knowles, T. P. J. (2018) Mechanobiology of protein droplets: force arises from disorder. Cell. **175**, 1457-1459

5 Klosin, A., Oltsch, F., Harmon, T., Honigmann, A., Jülicher, F., Hyman, A. A. and Zechner, C. (2020) Phase separation provides a mechanism to reduce noise in cells. Science. **367**, 464-468

6 Yoo, H., Triandafillou, C. and Drummond, D. A. (2019) Cellular sensing by phase separation: Using the process, not just the products. Journal of Biological Chemistry. **294**, 7151-7159

7 Babinchak, W. M. and Surewicz, W. K. (2020) Liquid–liquid phase separation and its mechanistic role in pathological protein aggregation. Journal of molecular biology. **432**, 1910-1925

8 Ray, S., Singh, N., Kumar, R., Patel, K., Pandey, S., Datta, D., Mahato, J., Panigrahi, R., Navalkar, A. and Mehra, S. (2020) α-Synuclein aggregation nucleates through liquid–liquid phase separation. Nature chemistry. **12**, 705-716

9 Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K. M., Bennett, R. E., Dujardin, S., Laskowski, P. R., MacKenzie, D., Kamath, T. and Commins, C. (2018) Tau protein liquid–liquid phase separation can initiate tau aggregation. The EMBO journal. **37**, e98049

10 Xing, Y., Nandakumar, A., Kakinen, A., Sun, Y., Davis, T. P., Ke, P. C. and Ding, F. (2020) Amyloid aggregation under the lens of liquid–liquid phase separation. The journal of physical chemistry letters. **12**, 368-378

11 Lu, J., Qian, J., Xu, Z., Yin, S., Zhou, L., Zheng, S. and Zhang, W. (2021) Emerging roles of liquid–liquid phase separation in cancer: from protein aggregation to immune-associated signaling. Frontiers in Cell and Developmental Biology. **9**, 1465

12 Mehta, S. and Zhang, J. (2022) Liquid–liquid phase separation drives cellular function and dysfunction in cancer. Nature Reviews Cancer. **22**, 239-252

13 Taniue, K. and Akimitsu, N. (2022) Aberrant phase separation and cancer. The FEBS journal. **289**, 17-39

14 Zbinden, A., Pérez-Berlanga, M., De Rossi, P. and Polymenidou, M. (2020) Phase separation and neurodegenerative diseases: a disturbance in the force. Developmental cell. **55**, 45-68

15 Ahmad, A., Uversky, V. N. and Khan, R. H. (2022) Aberrant liquid-liquid phase separation and amyloid aggregation of proteins related to neurodegenerative diseases. International Journal of Biological Macromolecules. **220**, 703-720

16 Boyko, S. and Surewicz, W. K. (2022) Tau liquid–liquid phase separation in neurodegenerative diseases. Trends in Cell Biology

17 Uversky, V. N. (2002) What does it mean to be natively unfolded? European Journal of Biochemistry. **269**, 2-12
18 Wright, P. E. and Dyson, H. J. (2015) Intrinsically disordered proteins in cellular signalling and regulation. Nature reviews Molecular cell biology. **16**, 18-29
19 Habchi, J., Tompa, P., Longhi, S. and Uversky, V. N. (2014) Introducing protein intrinsic disorder. Chemical reviews. **114**, 6561-6588
20 Borcherds, W., Bremer, A., Borgia, M. B. and Mittag, T. (2021) How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation? Current opinion in structural biology. **67**, 41-50
21 Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A. and Drechsel, D. (2018) A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. Cell. **174**, 688-699
22 Ruff, K. M., Choi, Y. H., Cox, D., Ormsby, A. R., Myung, Y., Ascher, D. B., Radford, S. E., Pappu, R. V. and Hatters, D. M. (2022) Sequence grammar underlying the unfolding and phase separation of globular proteins. Molecular Cell. **82**, 3193-3208
23 Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B. and Mittal, J. (2018) Sequence determinants of protein phase behavior from a coarse-grained model. PLoS computational biology. **14**, e1005941
24 Saar, K. L., Morgunov, A. S., Qi, R., Arter, W. E., Krainer, G., Lee, A. A. and Knowles, T. P. J. (2021) Learning the molecular grammar of protein condensates from sequence determinants and embeddings. Proceedings of the National Academy of Sciences. **118**, e2019053118
25 Krainer, G., Welsh, T. J., Joseph, J. A., Espinosa, J. R., Wittmann, S., de Csilléry, E., Sridhar, A., Toprakcioglu, Z., Gudiškytė, G. and Czekalska, M. A. (2021) Reentrant liquid condensate phase of proteins is stabilized by hydrophobic and non-ionic interactions. Nature communications. **12**, 1-14
26 Parker, M. W., Kao, J. A., Huang, A., Berger, J. M. and Botchan, M. R. (2021) Molecular determinants of phase separation for Drosophila DNA replication licensing factors. Elife. **10**, e70535
27 Das, R. K. and Pappu, R. V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proceedings of the National Academy of Sciences. **110**, 13392-13397
28 Das, R. K., Huang, Y., Phillips, A. H., Kriwacki, R. W. and Pappu, R. V. (2016) Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. Proceedings of the National Academy of Sciences. **113**, 5616-5621
29 Sherry, K. P., Das, R. K., Pappu, R. V. and Barrick, D. (2017) Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. Proceedings of the National Academy of Sciences. **114**, E9243-E9252
30 Tedeschi, G., Salladini, E., Santambrogio, C., Grandori, R., Longhi, S. and Brocca, S. (2018) Conformational response to charge clustering in synthetic intrinsically disordered proteins. Biochimica et Biophysica Acta (BBA)-General Subjects. **1862**, 2204-2214
31 Bianchi, G., Mangiagalli, M., Barbiroli, A., Longhi, S., Grandori, R.,

Santambrogio, C. and Brocca, S. (2022) Distribution of Charged Residues Affects the Average Size and Shape of Intrinsically Disordered Proteins. Biomolecules. **12**, 561
32 Bianchi, G., Longhi, S., Grandori, R. and Brocca, S. (2020) Relevance of electrostatic charges in compactness, aggregation, and phase separation of intrinsically disordered proteins. International Journal of Molecular Sciences. **21**, 6208
33 Lin, Y.-H. and Chan, H. S. (2017) Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. Biophysical journal. **112**, 2043-2046
34 Corbett, K. D. and Berger, J. M. (2004) Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. Annu. Rev. Biophys. Biomol. Struct. **33**, 95-118
35 Vos, S. M., Tretter, E. M., Schmidt, B. H. and Berger, J. M. (2011) All tangled up: how cells direct, manage and exploit topoisomerase function. Nature reviews Molecular cell biology. **12**, 827-841
36 Roca, J. (1995) The mechanisms of DNA topoisomerases. Trends in biochemical sciences. **20**, 156-160
37 Champoux, J. J. (2001) DNA topoisomerases: structure, function, and mechanism. Annual review of biochemistry. **70**, 369-413
38 Leppard, J. B. and Champoux, J. J. (2005) Human DNA topoisomerase I: relaxation, roles, and damage control. Chromosoma. **114**, 75-85
39 Alsner, J., Svejstrup, J. Q., Kjeldsen, E., Sørensen, B. S. and Westergaard, O. (1992) Identification of an N-terminal domain of eukaryotic DNA topoisomerase I dispensable for catalytic activity but essential for in vivo function. Journal of Biological Chemistry. **267**, 12408-12411
40 Mo, Y.-Y., Wang, C. and Beck, W. T. (2000) A novel nuclear localization signal in human DNA topoisomerase I. Journal of Biological Chemistry. **275**, 41107-41113
41 Bharti, A. K., Olson, M. O. J., Kufe, D. W. and Rubin, E. H. (1996) Identification of a Nucleolin Binding Site in Human Topoisomerase I (∗). Journal of Biological Chemistry. **271**, 1993-1997
42 Shaiu, W.-L. and Hsieh, T.-s. (1998) Targeting to transcriptionally active loci by the hydrophilic N-terminal domain of Drosophila DNA topoisomerase I. Molecular and cellular biology. **18**, 4358-4367
43 Mao, Y., Mehl, I. R. and Muller, M. T. (2002) Subnuclear distribution of topoisomerase I is linked to ongoing transcription and p53 status. Proceedings of the National Academy of Sciences. **99**, 1235-1240
44 Dasari, J. B., Soren, B. C., Ottaviani, A., Tesauro, C., Marino, S., Messina, B. and Fiorani, P. (2020) Swapping of The N-Terminal Domain of Human Topoisomerase 1B with the Corresponding Plasmodium Falciparum Counterpart Strongly Impairs Enzyme Activity. Reports of Biochemistry & Molecular Biology. **8**, 366
45 Lisby, M., Olesen, J. R., Skouboe, C., Krogh, B. O., Straub, T., Boege, F., Velmurugan, S., Martensen, P. M., Andersen, A. H. and Jayaram, M. (2001) Residues within the N-terminal domain of human topoisomerase I play a direct role in relaxation. Journal of Biological Chemistry. **276**, 20220-20227
46 Wright, C. M., Van Der Merwe, M., DeBrot, A. H. and Bjornsti, M.-A. (2015) DNA topoisomerase I domain interactions impact enzyme activity and sensitivity to camptothecin. Journal of Biological Chemistry. **290**, 12068-12078
47 Wang, W., Navarro, S., Azizyan, R. A., Baño-Polo, M., Esperante, S. A.,

Kajava, A. V. and Ventura, S. (2019) Prion soft amyloid core driven self-assembly of globular proteins into bioactive nanofibrils. Nanoscale. **11**, 12680-12694
48 Parker, M. W., Bell, M., Mir, M., Kao, J. A., Darzacq, X., Botchan, M. R. and Berger, J. M. (2019) A new class of disordered elements controls DNA replication through initiator self-assembly. Elife. **8**, e48562
49 Jeong, J., Lee, J. H., Carcamo, C. C., Parker, M. W. and Berger, J. M. (2022) DNA-Stimulated Liquid-Liquid Phase Separation by Eukaryotic Topoisomerase II Modulates Catalytic Function. bioRxiv
50 Holehouse, A. S., Ahad, J., Das, R. K. and Pappu, R. V. (2015) CIDER: classification of intrinsically disordered ensemble regions. Biophysical Journal. **108**, 228a
51 Mészáros, B., Erdős, G. and Dosztányi, Z. (2018) IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic acids research. **46**, W329-W337
52 Erdős, G. and Dosztányi, Z. (2020) Analyzing protein disorder with IUPred2A. Current Protocols in Bioinformatics. **70**, e99
53 Hatos, A., Tosatto, S. C. E., Vendruscolo, M. and Fuxreiter, M. (2022) FuzDrop on AlphaFold: visualizing the sequence-dependent propensity of liquid–liquid phase separation and aggregation of proteins. Nucleic Acids Research
54 Miskei, M., Antal, C. and Fuxreiter, M. (2016) FuzDB: database of fuzzy complexes, a tool to develop stochastic structure-function relationships for protein complexes and higher-order assemblies. Nucleic acids research, gkw1019
55 Kriventseva, E. V., Kuznetsov, D., Tegenfeldt, F., Manni, M., Dias, R., Simão, F. A. and Zdobnov, E. M. (2019) 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**, D807-D811
56 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. Journal of molecular biology. **215**, 403-410
57 Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S. and Richardson, L. J. (2021) Pfam: The protein families database in 2021. Nucleic acids research. **49**, D412-D419
58 Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. and Pappu, R. V. (2017) CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. Biophysical journal. **112**, 16-21
59 Saitou, N. and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular biology and evolution. **4**, 406-425
60 Kumar, S., Stecher, G., Li, M., Knyaz, C. and Tamura, K. (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Molecular biology and evolution. **35**, 1547
61 Tedeschi, G., Mangiagalli, M., Chmielewska, S., Lotti, M., Natalello, A. and Brocca, S. (2017) 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**, 2543-2550
62 Donnarumma, F., Leone, S., Delfi, M., Emendato, A., Ami, D.,

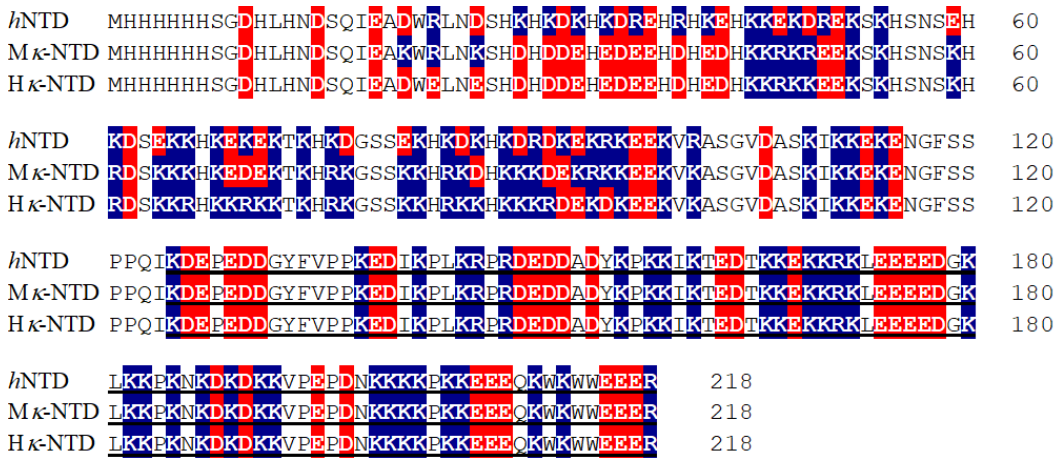
Laurents, D. V., Natalello, A., Spadaccini, R. and Picone, D. (2020) Probing structural changes during amyloid aggregation of the sweet protein MNEI. The FEBS Journal. **287**, 2808-2822
63 Ami, D. and Natalello, A. (2022) Characterization of the Conformational Properties of Soluble and Insoluble Proteins by Fourier Transform Infrared Spectroscopy. In Insoluble Proteins. pp. 439-454, Springer
64 Van Lindt, J., Bratek-Skicki, A., Nguyen, P. N., Pakravan, D., Durán-Armenta, L. F., Tantos, A., Pancsa, R., Van Den Bosch, L., Maes, D. and Tompa, P. (2021) A generic approach to study the kinetics of liquid–liquid phase separation under near-native conditions. Communications biology. **4**, 1-8
65 Natalello, A., Ami, D. and Doglia, S. M. (2012) Fourier transform infrared spectroscopy of intrinsically disordered proteins: measurement procedures and data analyses. In Intrinsically disordered protein analysis. pp. 229-244, Springer
66 Yang, H., Yang, S., Kong, J., Dong, A. and Yu, S. (2015) Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. Nature protocols. **10**, 382-396
67 Goormaghtigh, E., Cabiaux, V. and Ruysschaert, J.-M. (1994) Determination of soluble and membrane protein structure by Fourier transform infrared spectroscopy. Physicochemical methods in the study of biomembranes, 405-450
68 Shivu, B., Seshadri, S., Li, J., Oberg, K. A., Uversky, V. N. and Fink, A. L. (2013) Distinct β-sheet structure in protein aggregates determined by ATR–FTIR spectroscopy. Biochemistry. **52**, 5176-5183
69 Barth, A. (2007) Infrared spectroscopy of proteins. Biochimica et Biophysica Acta (BBA)-Bioenergetics. **1767**, 1073-1101
70 Maharana, S., Wang, J., Papadopoulos, D. K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J. and Franzmann, T. M. (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science. **360**, 918-921
71 Ukmar-Godec, T., Hutten, S., Grieshop, M. P., Rezaei-Ghaleh, N., Cima-Omori, M.-S., Biernat, J., Mandelkow, E., Söding, J., Dormann, D. and Zweckstetter, M. (2019) Lysine/RNA-interactions drive and regulate biomolecular condensation. Nature communications. **10**, 1-15
72 Alberti, S. and Dormann, D. (2019) Liquid-liquid phase separation in disease. Annu. Rev. Genet. **53**, 171-194
73 Mitrea, D. M. and Kriwacki, R. W. (2016) Phase separation in biology; functional organization of a higher order. Cell Communication and Signaling. **14**, 1
74 Forman-Kay, J. D., Kriwacki, R. W. and Seydoux, G. (2018) Phase separation in biology and disease. Journal of molecular biology. **430**, 4603
75 Garaizar, A., Sanchez-Burgos, I., Collepardo-Guevara, R. and Espinosa, J. R. (2020) Expansion of intrinsically disordered proteins increases the range of stability of liquid–liquid phase separation. Molecules. **25**, 4705
76 Mészáros, B., Erdős, G., Szabó, B., Schád, É., Tantos, Á., Abukhairan, R., Horváth, T., Murvai, N., Kovács, O. P. and Kovács, M. (2020) PhaSePro: the database of proteins driving liquid–liquid phase separation. Nucleic acids research. **48**, D360-D367
77 Vancraenenbroeck, R., Harel, Y. S., Zheng, W. and Hofmann, H. (2019) Polymer effects modulate binding

affinities in disordered proteins. Proceedings of the National Academy of Sciences. **116**, 19506-19512
78 Boeynaems, S., Holehouse, A. S., Weinhardt, V., Kovacs, D., Van Lindt, J., Larabell, C., Van Den Bosch, L., Das, R., Tompa, P. S. and Pappu, R. V. (2019) Spontaneous driving forces give rise to protein− RNA condensates with coexisting phases and complex material properties. Proceedings of the National Academy of Sciences. **116**, 7889-7898
79 Aumiller, W. M. and Keating, C. D. (2016) Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. Nature chemistry. **8**, 129-137
80 Polymenidou, M. (2018) The RNA face of phase separation. Science. **360**, 859-860
81 Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T. and Forman-Kay, J. D. (2015) Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Molecular cell. **57**, 936-947
82 Pak, C. W., Kosno, M., Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., Yunus, A. A., Liu, D. R., Pappu, R. V. and Rosen, M. K. (2016) Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. Molecular cell. **63**, 72-85
83 Schuster, B., Dignon, G. L., Tang, W. S., Kelley, F., Ranganath, A. K., Jahnke, C. N., Simpkins, A. G., Regy, R. M., Hammer, D. A. and Good, M. C. (2020) Identifying Sequence Perturbations to an Intrinsically Disordered Protein that Determine Its Phase Separation Behavior. bioRxiv

Supplementary information

Group	Subgroup	Species	NTD		
			κ value	*NCPR*	Length (aa)
Metazoa	Vertebrata	*Homo sapiens* (P11387)	0.128	0.051	214
		Mus musculus (NP_033434.2)	0.126	0.056	216
		Cricetulus griseus (CAA79748.1)	0.126	0.056	216
		Gallus gallus (NP_990441.1)	0.139	0.060	215
		Falco peregrinus (XP_005235944.1)	0.139	0.060	215
		Danio rerio (NP_001037789.1)	0.122	0.058	207
		Oryzias latipes (XP_023810923.1)	0.134	0.070	199
		Carcharodon carcharias (XP_041059975.1)	0.132	0.082	220
		Amblyraja radiata (XP_032897763.1)	0.145	0.091	220
		Lacerta agilis (XP_033009285.1)	0.140	0.060	215
		Varanus komodoensis (KAF7240897.1)	0.139	0.079	214
		Caretta caretta (XP_048675811.1)	0.149	0.056	215
		Xenopus laevis (XP_018089694.1)	0.114	0.063	270
	Lophotrochozoan	*Sepia pharaonis* (CAE1163795.1)	0.126	0.037	374
		Octopus bimaculoides (OX=37653)	0.133	0.074	326
	Nematoda	*Trichinella pseudospiralis* (KRX92265.1)	0.230	0.028	216
		Caenorhabditis elegans (NP_001379240.1)	0.294	0.016	184
	Arthropoda	*Araneus ventricosus* (GBM09797.1)	0.134	0.047	387
		Caerostris darwini (GIY45528.1)	0.190	-0.008	127
		Varroa destructor (XP_022671948.1)	0.220	-0.049	676
		Papilio machaon (KPJ12715.1)	0.180	0.013	374
		Halyomorpha halys (XP_024217857.1)	0.174	-0.015	395
		Aedes aegypti (EAT46253.1)	0.169	-0.012	421
		Drosophila ananassae (XP_001963839.2)	0.248	-0.020	455
		Glossina fuscipes (XP_037889119.1)	0.265	-0.034	475
		Lucilia cuprina (XP_023291221.2)	0.236	-0.029	476
Protista	Alveolata	*Perkinsus olseni* (KAF4667565.1)	0.154	0.131	130
		Tetrahymena thermophila (XP_001025304.2)	0.437	-0.253	133
		Plasmodium vinckei petteri (EUD73215.1)	0.204	0.084	179
	Amoebozoa	*Dictyostelium discoideum* (XP_639222.1)	0.337	-0.019	309
		Planoprotostelium fungivorum (PRP82603.1)	0.256	0.037	217
	Euglenozoa	*Leishmania tarentolae* (GET92410.1)	0.289	-0.163	43
		Trypanosoma cruzi (ESS68442.1)	0.225	-0.133	46
Fungi	Ascomycota	*Penicillium oxalicum* (XP_049968292.1)	0.296	0.009	228
		Phialophora_attinorum (XP_018005181.1)	0.324	0.000	242
		Aspergillus nanangensis (KAF9886811.1)	0.309	0.020	246
		Neurospora crassa (XP_011393242.1)	0.303	0.004	286
		Xylaria sp. FL0933 (KAI1279976.1)	0.366	0.000	240
		Trichoderma lentiforme (KAF3074930.1)	0.315	0.037	273
		Saccharomyces cerevisiae (P04786)	0.324	-0.042	142
		Schizosaccharomyce pombe (NP_596209.1)	0.415	0.010	192
	Basidiomycota	*Pisolithus croceorrhizus* (KAI6131672.1)	0.326	0.014	212
		Rhodotorula toruloides (XP_016275408.1)	0.361	-0.028	287
		Malassezia vespertilionis (PKI83396.1)	0.342	-0.089	101
Viridiplantae	Streptophyta	*Brassica napus* (XP_013726218.2)	0.234	0.056	357
		Arabidopsis thaliana (OAO95951.1)	0.221	0.063	366
		Zea mays (AQK50797.1)	0.263	0.104	192
		Oryza sativa (XP_015650112.1)	0.252	0.070	371
		Triticum aestivum (XP_044326396.1)	0.352	0.011	266
		Ananas comosus (XP_020110486.1)	0.279	0.045	333
		Prunus persica (XP_020425762.1)	0.302	0.070	201
		Nicotiana tabacum (AAD08711.1)	0.312	0.057	332
		Solanum lycopersicum (XP_010323847.2)	0.319	0.065	336
		Capsicum annuum (XP_016565649.1)	0.346	0.061	343
	Chlorophyta	*Scenedesmus sp* (KAF8067313.1)	0.328	0.004	234

Table S1: List of species from which the TOP1 orthologues considered in this study originate. The NCPR and *κ*-value were calculated using CIDER (http://pappulab.wustl.edu/CIDER/analysis/) on the NTD domains identified by analyzing the architecture of each protein with Pfam (https://www.ebi.ac.uk/interpro/search/text/398365647/#table).



S1. Amino acid sequences of NTDs used as model IDPs. The amino acid sequences of *h*NTD, and the Mκ- and Hκ-NTD permutants have been aligned to highlight the sequence design. Positively charged residues are shown in blue, negatively charged residues in red, underlined are the residues belonging to NLS.

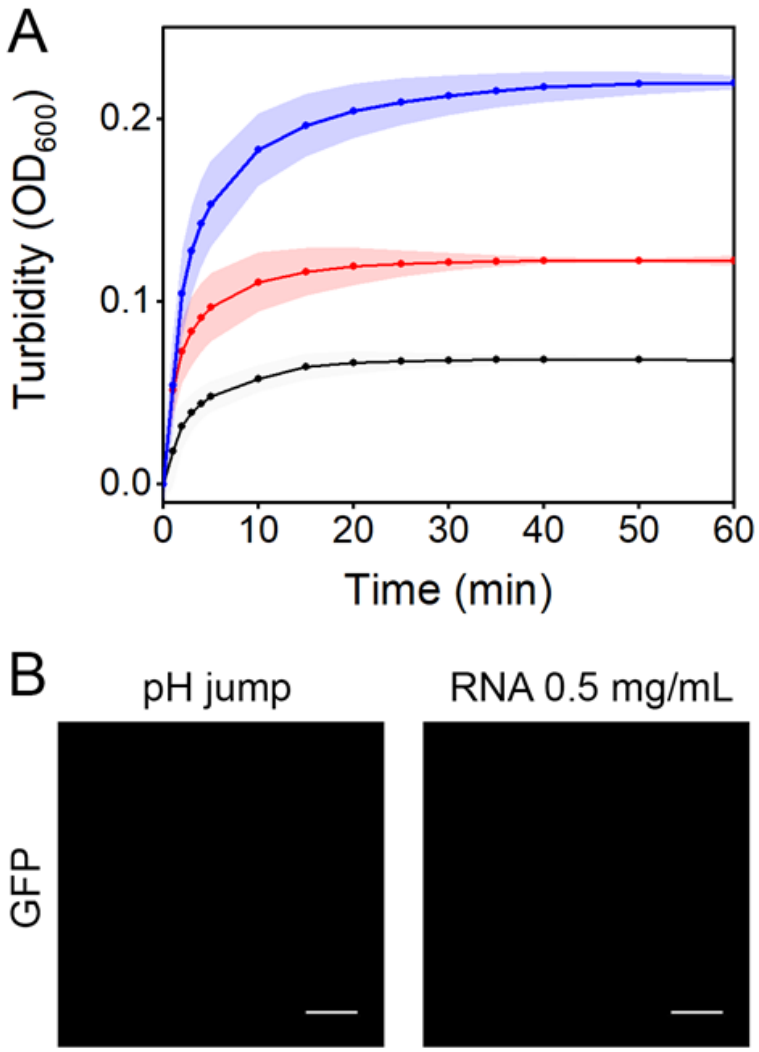


Fig. S2. The fusion of *h*NTD with GFP does not affect its condensation propensity and kinetics. A) Turbidity plots of *h*NTD-GFP after pH jump. Experiments were carried out on three different protein concentrations (25 μM, 50 μM, 100 μM) at 150 mM NaCl. **B)** Confocal microscopy analysis of GFP subjected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

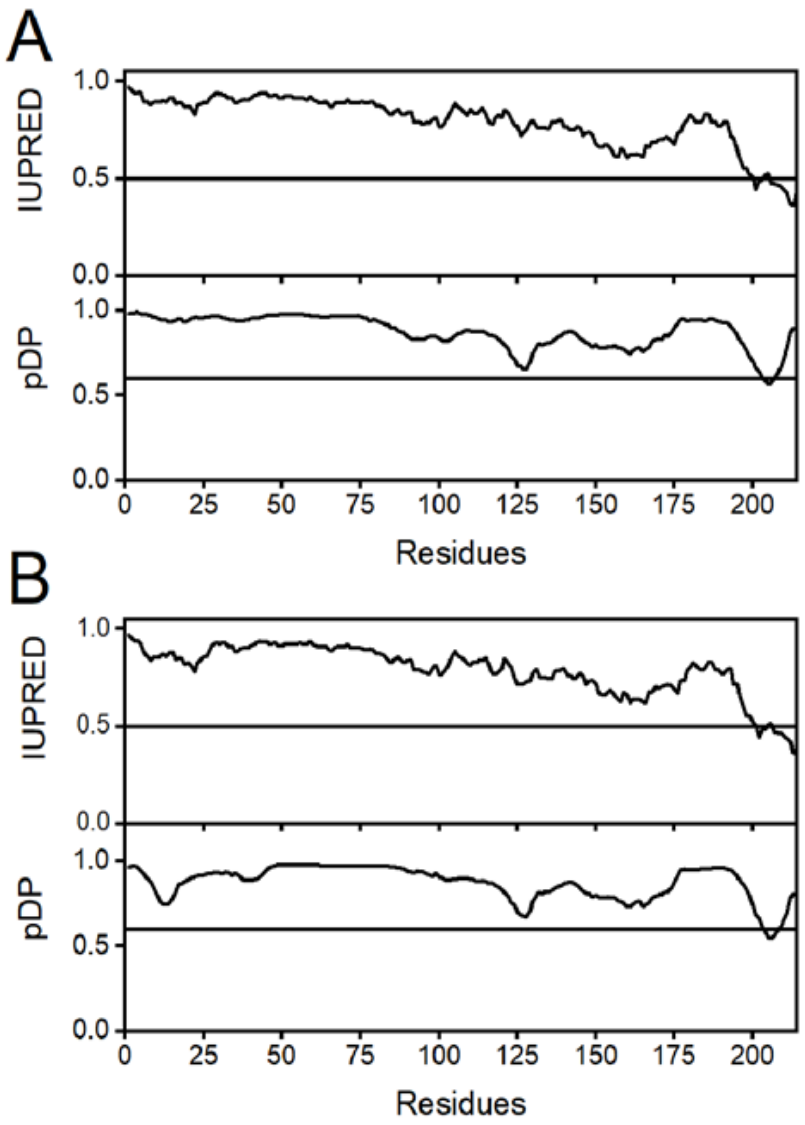


Fig. S3. Predictions of disorder and LLPS propensity for M*κ*-NTD and H*κ*-NTD. The permutation of charged residues did not affect the degree of disorder predicted with IUPred, and propensity to undergo LLPS, measured in terms of droplet-promoting probabilities of residues (pDP) by the FuzDrop server (https://fuzdrop.bio.unipd.it/predictor; [53]) for M*κ*-NTD (**A)** and H*κ*-NTD (**B**).

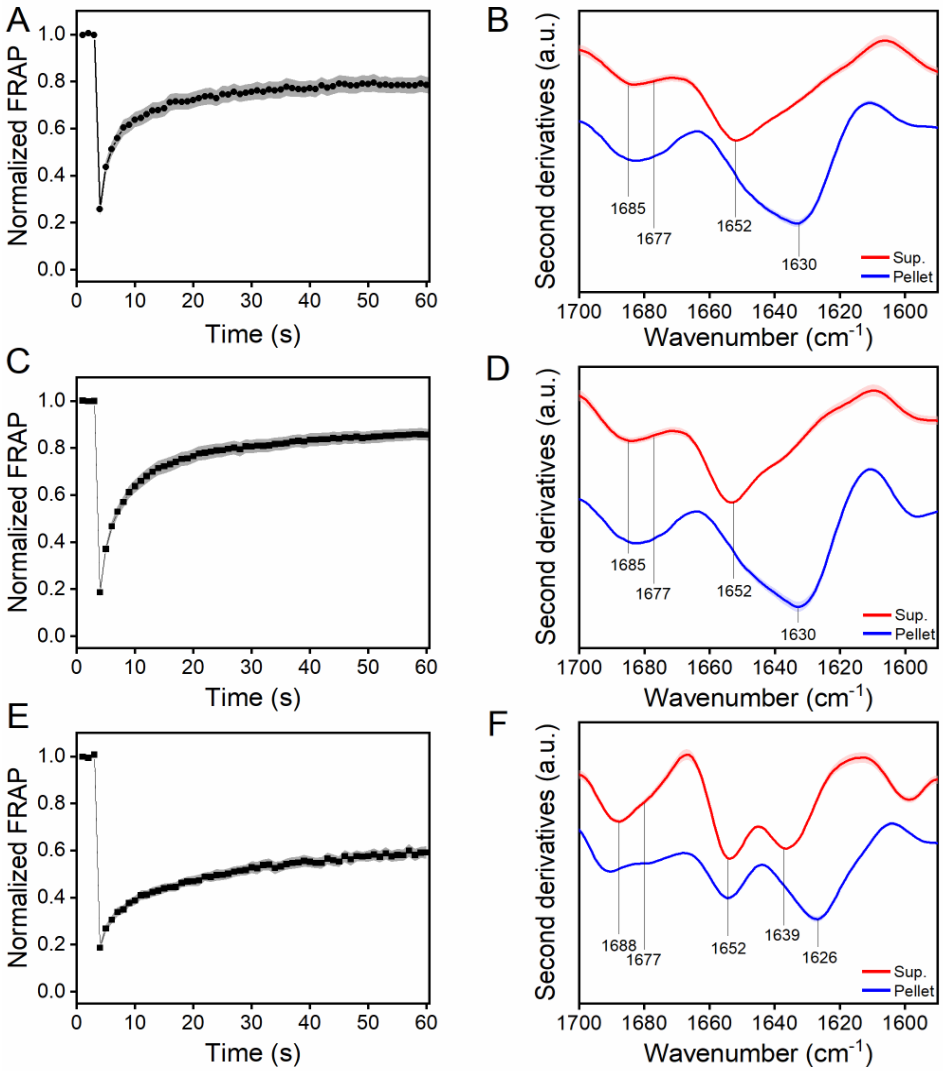


Fig. S4 Fluorescence recovery and secondary structure analysis of condensates obtained in presence of RNA. A, C, E) FRAP analysis of *h*NTD-GFP **(A)**, M*κ*-NTD-GFP **(C)**, and H*κ*-NTD-GFP **(E)** condensates induced by poly(A) addition (0.5 mg/mL; protein concentration 20 μM; n = 25). The fluorescence signal was normalized from 0 to 1. **B, D, F)** Second derivatives of FTIR absorption spectra of *h*NTD, M*κ*-NTD, and H*κ*-NTD condensates obtained in presence of 0.5 mg/mL poly(A), subtracted for RNA and PBS contribution. Red line: supernatant; blue line: pellet.

4. Thesis references

1 Pauling, L. (1940) A theory of the structure and process of formation of antibodies. Journal of the American Chemical Society. **62**, 2643-2657

2 Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M. and Hipps, K. W. (2001) Intrinsically disordered protein. Journal of molecular graphics and modelling. **19**, 26-59

3 Dunker, A. K. and Oldfield, C. J. (2015) Back to the future: nuclear magnetic resonance and bioinformatics studies on intrinsically disordered proteins. Intrinsically Disordered Proteins Studied by NMR Spectroscopy, 1-34

4 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The protein data bank. Nucleic acids research. **28**, 235-242

5 Berman, H. M., Battistuz, T., Bhat, T. N., Bluhm, W. F., Bourne, P. E., Burkhardt, K., Feng, Z., Gilliland, G. L., Iype, L. and Jain, S. (2002) The protein data bank. Acta Crystallographica Section D: Biological Crystallography. **58**, 899-907

6 Bondos, S. E., Dunker, A. K. and Uversky, V. N. (2021) On the roles of intrinsically disordered proteins and regions in cell communication and signaling. ed.)^eds.). pp. 1-9, Springer

7 Vucetic, S., Obradovic, Z., Vacic, V., Radivojac, P., Peng, K., Iakoucheva, L. M., Cortese, M. S., Lawson, J. D., Brown, C. J. and Sikes, J. G. (2005) DisProt: a database of protein disorder. Bioinformatics. **21**, 137-140

8 Sickmeier, M., Hamilton, J. A., LeGall, T., Vacic, V., Cortese, M. S., Tantos, A., Szabo, B., Tompa, P., Chen, J. and Uversky, V. N. (2007) DisProt: the database of disordered proteins. Nucleic acids research. **35**, D786-D793

9 Piovesan, D., Tabaro, F., Mičetić, I., Necci, M., Quaglia, F., Oldfield, C. J., Aspromonte, M. C., Davey, N. E., Davidović, R. and Dosztányi, Z. (2017) DisProt 7.0: a major update of the database of disordered proteins. Nucleic acids research. **45**, D219-D227

10 Piovesan, D. and Tosatto, S. C. E. (2018) Mobi 2.0: an improved method to define intrinsic disorder, mobility and linear binding regions in protein structures. Bioinformatics. **34**, 122-123

11 Oates, M. E., Romero, P., Ishida, T., Ghalwash, M., Mizianty, M. J., Xue, B., Dosztanyi, Z., Uversky, V. N., Obradovic, Z. and Kurgan, L. (2012) D2P2: database of disordered protein predictions. Nucleic acids research. **41**, D508-D516

12 Zhao, B., Katuwawala, A., Oldfield, C. J., Dunker, A. K., Faraggi, E., Gsponer, J., Kloczkowski, A., Malhis, N., Mirdita, M. and Obradovic, Z. (2021) DescribePROT: database of amino acid-level protein structure and function predictions. Nucleic Acids Research. **49**, D298-D308

13 Dunker, A. K. and Obradovic, Z. (2001) The protein trinity—linking function and disorder. Nature biotechnology. **19**, 805-806

14 Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. and Jones, D. T. (2004) Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. Journal of molecular biology. **337**, 635-645

15 Schad, E., Tompa, P. and Hegyi, H. (2011) The relationship between proteome size, structural disorder and organism complexity. Genome biology. **12**, 1-13

16 Habchi, J., Tompa, P., Longhi, S. and Uversky, V. N. (2014) Introducing protein intrinsic disorder. Chemical reviews. **114**, 6561-6588

17 Tompa, P. and Fuxreiter, M. (2008) Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. Trends in biochemical sciences. **33**, 2-8

18 Paul, F. and Weikl, T. R. (2016) How to distinguish conformational selection and induced fit based on chemical relaxation rates. PLOS Computational Biology. **12**, e1005067

19 Theillet, F.-X., Binolfi, A., Frembgen-Kesner, T., Hingorani, K., Sarkar, M., Kyne, C., Li, C., Crowley, P. B., Gierasch, L. and Pielak, G. J. (2014) Physicochemical properties of cells and their effects on intrinsically disordered proteins (IDPs). Chemical reviews. **114**, 6661-6714
20 Bah, A. and Forman-Kay, J. D. (2016) Modulation of intrinsically disordered protein function by post-translational modifications. Journal of Biological Chemistry. **291**, 6696-6705
21 Bah, A., Vernon, R. M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L. E. and Forman-Kay, J. D. (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch. Nature. **519**, 106-109
22 Higgs, P. G. and Joanny, J. F. (1991) Theory of polyampholyte solutions. The Journal of chemical physics. **94**, 1543-1554
23 Ha, B. Y. and Thirumalai, D. (1997) Persistence length of intrinsically stiff polyampholyte chains. Journal de Physique II. **7**, 887-902
24 Gutin, A. M. and Shakhnovich, E. I. (1994) Effect of a net charge on the conformation of polyampholytes. Physical Review E. **50**, R3322
25 Dobrynin, A. V., Colby, R. H. and Rubinstein, M. (2004) Polyampholytes. Journal of Polymer Science Part B: Polymer Physics. **42**, 3513-3538
26 Dobrynin, A. V. and Rubinstein, M. (2005) Theory of polyelectrolytes in solutions and at surfaces. Progress in Polymer Science. **30**, 1049-1118
27 Das, R. K. and Pappu, R. V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proceedings of the National Academy of Sciences. **110**, 13392-13397
28 Uversky, V. N., Gillespie, J. R. and Fink, A. L. (2000) Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins: structure, function, and bioinformatics. **41**, 415-427
29 Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L. and Pappu, R. V. (2010) Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proceedings of the National Academy of Sciences. **107**, 8183-8188
30 Liang, J. F., Yang, V. C. and Vaynshteyn, Y. (2005) The minimal functional sequence of protamine. Biochemical and biophysical research communications. **336**, 653-659
31 Brewer, L. R., Corzett, M. and Balhorn, R. (1999) Protamine-induced condensation and decondensation of the same DNA molecule. Science. **286**, 120-123
32 Ward, W. S. and Coffey, D. S. (1991) DNA packaging and organization in mammalian spermatozoa: comparison with somatic cells. Biol Reprod. **44**, 569-574
33 Hatton, T., Zhou, S. and Standring, D. N. (1992) RNA-and DNA-binding activities in hepatitis B virus capsid protein: a model for their roles in viral replication. Journal of virology. **66**, 5232-5241
34 Bianchi, G., Longhi, S., Grandori, R. and Brocca, S. (2020) Relevance of electrostatic charges in compactness, aggregation, and phase separation of intrinsically disordered proteins. International Journal of Molecular Sciences. **21**, 6208
35 Das, R. K., Ruff, K. M. and Pappu, R. V. (2015) Relating sequence encoded information to form and function of intrinsically disordered proteins. Current opinion in structural biology. **32**, 102-112
36 Uversky, V. N. (2019) Intrinsically disordered proteins and their "mysterious"(meta) physics. Frontiers in Physics. **7**, 10
37 Srivastava, D. and Muthukumar, M. (1996) Sequence dependence of conformations of polyampholytes. Macromolecules. **29**, 2324-2326
38 Flory, P. J. and Volkenstein, M. (1969) Statistical mechanics of chain molecules. Biopolymers: Original Research on Biomolecules. **8**, 699-700
39 Flory, P. J. (1953) Principles of polymer chemistry. Cornell University Press

40 Sawle, L. and Ghosh, K. (2015) A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. The Journal of chemical physics. **143**, 08B615_611
41 Lin, Y.-H. and Chan, H. S. (2017) Phase separation and single-chain compactness of charged disordered proteins are strongly correlated. Biophysical journal. **112**, 2043-2046
42 Firman, T. and Ghosh, K. (2018) Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. The Journal of chemical physics. **148**, 123305
43 Müller-Späth, S., Soranno, A., Hirschfeld, V., Hofmann, H., Rüegger, S., Reymond, L., Nettels, D. and Schuler, B. (2010) Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proceedings of the National Academy of Sciences. **107**, 14609-14614
44 Huihui, J., Firman, T. and Ghosh, K. (2018) Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. The Journal of chemical physics. **149**, 085101
45 Vancraenenbroeck, R., Harel, Y. S., Zheng, W. and Hofmann, H. (2019) Polymer effects modulate binding affinities in disordered proteins. Proceedings of the National Academy of Sciences. **116**, 19506-19512
46 Beveridge, R., Migas, L. G., Das, R. K., Pappu, R. V., Kriwacki, R. W. and Barran, P. E. (2019) Ion mobility mass spectrometry uncovers the impact of the patterning of oppositely charged residues on the conformational distributions of intrinsically disordered proteins. Journal of the American Chemical Society. **141**, 4908-4918
47 Das, R. K., Huang, Y., Phillips, A. H., Kriwacki, R. W. and Pappu, R. V. (2016) Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. Proceedings of the National Academy of Sciences. **113**, 5616-5621
48 Sherry, K. P., Das, R. K., Pappu, R. V. and Barrick, D. (2017) Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. Proceedings of the National Academy of Sciences. **114**, E9243-E9252
49 Marsh, J. A. and Forman-Kay, J. D. (2010) Sequence determinants of compaction in intrinsically disordered proteins. Biophysical journal. **98**, 2383-2390
50 Tedeschi, G., Salladini, E., Santambrogio, C., Grandori, R., Longhi, S. and Brocca, S. (2018) Conformational response to charge clustering in synthetic intrinsically disordered proteins. Biochimica et Biophysica Acta (BBA)-General Subjects. **1862**, 2204-2214
51 Bianchi, G., Mangiagalli, M., Barbiroli, A., Longhi, S., Grandori, R., Santambrogio, C. and Brocca, S. (2022) Distribution of Charged Residues Affects the Average Size and Shape of Intrinsically Disordered Proteins. Biomolecules. **12**, 561
52 Zeng, X., Ruff, K. M. and Pappu, R. V. (2022) Competing interactions give rise to two-state behavior and switch-like transitions in charge-rich intrinsically disordered proteins. bioRxiv
53 Srinivasan, N., Bhagawati, M., Ananthanarayanan, B. and Kumar, S. (2014) Stimuli-sensitive intrinsically disordered protein brushes. Nature communications. **5**, 1-8
54 Riback, J. A. and Brangwynne, C. P. (2020) Can phase separation buffer cellular noise? Science. **367**, 364-365
55 Wirth, A. J. and Gruebele, M. (2013) Quinary protein structure and the consequences of crowding in living cells: Leaving the test-tube behind. BioEssays. **35**, 984-993
56 Cohen, R. D. and Pielak, G. J. (2017) A cell is more than the sum of its (dilute) parts: A brief history of quinary structure. Protein Science. **26**, 403-413
57 Krebs, H. A. and Johnson, W. A. (1937) Acetopyruvic acid (αγ-diketovaleric acid) as an intermediate metabolite in animal tissues. Biochemical Journal. **31**, 772
58 Krebs, H. A. and Johnson, W. A. (1937) Metabolism of ketonic acids in animal tissues. Biochemical Journal. **31**, 645
59 Vélot, C., Mixon, M. B., Teige, M. and Srere, P. A. (1997) Model of a quinary structure between Krebs TCA cycle enzymes: a model for the metabolon. Biochemistry. **36**, 14271-14276

60 Miles, E. W., Rhee, S. and Davies, D. R. (1999) The molecular basis of substrate channeling. Journal of Biological Chemistry. **274**, 12193-12196
61 Robinson Jr, J. B., Inman, L., Sumegi, B. and Srere, P. A. (1987) Further characterization of the Krebs tricarboxylic acid cycle metabolon. Journal of Biological Chemistry. **262**, 1786-1790
62 Ovádi, J. and Sreret, P. A. (1999) Macromolecular compartmentation and channeling. International review of cytology. **192**, 255-280
63 Mao, Y. S., Zhang, B. and Spector, D. L. (2011) Biogenesis and function of nuclear bodies. Trends in Genetics. **27**, 295-306
64 Handwerger, K. E. and Gall, J. G. (2006) Subnuclear organelles: new insights into form and function. Trends in cell biology. **16**, 19-26
65 Decker, C. J. and Parker, R. (2012) P-bodies and stress granules: possible roles in the control of translation and mRNA degradation. Cold Spring Harbor perspectives in biology. **4**, a012286
66 Mitrea, D. M. and Kriwacki, R. W. (2016) Phase separation in biology; functional organization of a higher order. Cell Communication and Signaling. **14**, 1
67 Banani, S. F., Lee, H. O., Hyman, A. A. and Rosen, M. K. (2017) Biomolecular condensates: organizers of cellular biochemistry. Nature reviews Molecular cell biology. **18**, 285-298
68 Hyman, A. A., Weber, C. A. and Jülicher, F. (2014) Liquid-liquid phase separation in biology. Annu. Rev. Cell Dev. Biol. **30**, 39-58
69 Biswas, S., Mukherjee, B. and Chakrabarti, B. (2021) Thermodynamics of droplets undergoing liquid-liquid phase separation. arXiv preprint arXiv:2104.00651
70 Asherie, N., Lomakin, A. and Benedek, G. B. (1996) Phase diagram of colloidal solutions. Physical review letters. **77**, 4832
71 Martin, E. W., Harmon, T. S., Hopkins, J. B., Chakravarthy, S., Incicco, J. J., Schuck, P., Soranno, A. and Mittag, T. (2021) A multi-step nucleation process determines the kinetics of prion-like domain phase separation. Nature communications. **12**, 1-12
72 Alberti, S., Gladfelter, A. and Mittag, T. (2019) Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. Cell. **176**, 419-434
73 Gomes, E. and Shorter, J. (2019) The molecular language of membraneless organelles. Journal of Biological Chemistry. **294**, 7115-7127
74 Buchan, J. R., Muhlrad, D. and Parker, R. (2008) P bodies promote stress granule assembly in Saccharomyces cerevisiae. The Journal of cell biology. **183**, 441-455
75 Protter, D. S. W. and Parker, R. (2016) Principles and properties of stress granules. Trends in cell biology. **26**, 668-679
76 Antifeeva, I. A., Fonin, A. V., Fefilova, A. S., Stepanenko, O. V., Povarova, O. I., Silonov, S. A., Kuznetsova, I. M., Uversky, V. N. and Turoverov, K. K. (2022) Liquid–liquid phase separation as an organizing principle of intracellular space: overview of the evolution of the cell compartmentalization concept. Cellular and Molecular Life Sciences. **79**, 1-32
77 Reineke, L. C. and Lloyd, R. E. (2013) Diversion of stress granules and P-bodies during viral infection. Virology. **436**, 255-267
78 Arimoto, K., Fukuda, H., Imajoh-Ohmi, S., Saito, H. and Takekawa, M. (2008) Formation of stress granules inhibits apoptosis by suppressing stress-responsive MAPK pathways. Nature cell biology. **10**, 1324-1332
79 Takahara, T. and Maeda, T. (2012) Transient sequestration of TORC1 into stress granules during heat stress. Molecular cell. **47**, 242-252
80 Thedieck, K., Holzwarth, B., Prentzell, M. T., Boehlke, C., Kläsener, K., Ruf, S., Sonntag, A. G., Maerz, L., Grellscheid, S.-N. and Kremmer, E. (2013) Inhibition of mTORC1 by astrin and stress granules prevents apoptosis in cancer cells. Cell. **154**, 859-874
81 Kim, Y.-S., Nedospasov, S. A. and Liu, Z.-g. (2005) TRAF2 plays a key, nonredundant role in LIGHT-lymphotoxin β receptor signaling. Molecular and cellular biology. **25**, 2130-2137

82 Corpet, A., Kleijwegt, C., Roubille, S., Juillard, F., Jacquet, K., Texier, P. and Lomonte, P. (2020) PML nuclear bodies and chromatin dynamics: catch me if you can! Nucleic Acids Research. **48**, 11890-11912
83 Seeler, J.-S., Marchio, A., Sitterlin, D., Transy, C. and Dejean, A. (1998) Interaction of SP100 with HP1 proteins: a link between the promyelocytic leukemia-associated nuclear bodies and the chromatin compartment. Proceedings of the National Academy of Sciences. **95**, 7316-7321
84 Doucas, V., Tini, M., Egan, D. A. and Evans, R. M. (1999) Modulation of CREB binding protein function by the promyelocytic (PML) oncoprotein suggests a role for nuclear bodies in hormone signaling. Proceedings of the National Academy of Sciences. **96**, 2627-2632
85 Drané, P., Ouararhni, K., Depaux, A., Shuaib, M. and Hamiche, A. (2010) The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3. 3. Genes & development. **24**, 1253-1265
86 Lewis, P. W., Elsaesser, S. J., Noh, K.-M., Stadler, S. C. and Allis, C. D. (2010) Daxx is an H3. 3-specific histone chaperone and cooperates with ATRX in replication-independent chromatin assembly at telomeres. Proceedings of the National Academy of Sciences. **107**, 14075-14080
87 Banani, S. F., Rice, A. M., Peeples, W. B., Lin, Y., Jain, S., Parker, R. and Rosen, M. K. (2016) Compositional control of phase-separated cellular bodies. Cell. **166**, 651-663
88 Li, P., Banjade, S., Cheng, H.-C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J. V., King, D. S. and Banani, S. F. (2012) Phase transitions in the assembly of multivalent signalling proteins. Nature. **483**, 336-340
89 Kwon, I., Kato, M., Xiang, S., Wu, L., Theodoropoulos, P., Mirzaei, H., Han, T., Xie, S., Corden, J. L. and McKnight, S. L. (2013) Phosphorylation-regulated binding of RNA polymerase II to fibrous polymers of low-complexity domains. Cell. **155**, 1049-1060
90 Schmit, J. D., Feric, M. and Dundr, M. (2021) How hierarchical interactions make membraneless organelles tick like clockwork. Trends in Biochemical Sciences. **46**, 525-534
91 Brangwynne, C. P. (2011) Soft active aggregates: mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies. Soft Matter. **7**, 3052-3059
92 Wang, J. T., Smith, J., Chen, B.-C., Schmidt, H., Rasoloson, D., Paix, A., Lambrus, B. G., Calidas, D., Betzig, E. and Seydoux, G. (2014) Regulation of RNA granule dynamics by phosphorylation of serine-rich, intrinsically disordered proteins in C. elegans. Elife. **3**, e04591
93 Feric, M., Vaidya, N., Harmon, T. S., Mitrea, D. M., Zhu, L., Richardson, T. M., Kriwacki, R. W., Pappu, R. V. and Brangwynne, C. P. (2016) Coexisting liquid phases underlie nucleolar subcompartments. Cell. **165**, 1686-1697
94 Jain, S., Wheeler, J. R., Walters, R. W., Agrawal, A., Barsic, A. and Parker, R. (2016) ATPase-modulated stress granules contain a diverse proteome and substructure. Cell. **164**, 487-498
95 Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A. and Drechsel, D. (2018) A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. Cell. **174**, 688-699
96 Elbaum-Garfinkle, S., Kim, Y., Szczepaniak, K., Chen, C. C.-H., Eckmann, C. R., Myong, S. and Brangwynne, C. P. (2015) The disordered P granule protein LAF-1 drives phase separation into droplets with tunable viscosity and dynamics. Proceedings of the National Academy of Sciences. **112**, 7189-7194
97 Nott, T. J., Petsalaki, E., Farber, P., Jervis, D., Fussner, E., Plochowietz, A., Craggs, T. D., Bazett-Jones, D. P., Pawson, T. and Forman-Kay, J. D. (2015) Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Molecular cell. **57**, 936-947
98 Molliex, A., Temirov, J., Lee, J., Coughlin, M., Kanagaraj, A. P., Kim, H. J., Mittag, T. and Taylor, J. P. (2015) Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. Cell. **163**, 123-133

99 McGurk, L., Gomes, E., Guo, L., Shorter, J. and Bonini, N. M. (2018) Poly (ADP-ribose) engages the TDP-43 nuclear-localization sequence to regulate granulo-filamentous aggregation. Biochemistry. **57**, 6923-6926
100 Borcherds, W., Bremer, A., Borgia, M. B. and Mittag, T. (2021) How do intrinsically disordered protein regions encode a driving force for liquid–liquid phase separation? Current opinion in structural biology. **67**, 41-50
101 Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. and Obradović, Z. (2002) Intrinsic disorder and protein function. Biochemistry. **41**, 6573-6582
102 Dunker, A. K., Cortese, M. S., Romero, P., Iakoucheva, L. M. and Uversky, V. N. (2005) Flexible nets: the roles of intrinsic disorder in protein interaction networks. The FEBS journal. **272**, 5129-5148
103 Xu, B., He, G., Weiner, B. G., Ronceray, P., Meir, Y., Jonikas, M. C. and Wingreen, N. S. (2020) Rigidity enhances a magic-number effect in polymer phase separation. Nature communications. **11**, 1-8
104 Vernon, R. M., Chong, P. A., Tsang, B., Kim, T. H., Bah, A., Farber, P., Lin, H. and Forman-Kay, J. D. (2018) Pi-Pi contacts are an overlooked protein feature relevant to phase separation. elife. **7**
105 Pak, C. W., Kosno, M., Holehouse, A. S., Padrick, S. B., Mittal, A., Ali, R., Yunus, A. A., Liu, D. R., Pappu, R. V. and Rosen, M. K. (2016) Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. Molecular cell. **63**, 72-85
106 Hughes, M. P., Sawaya, M. R., Boyer, D. R., Goldschmidt, L., Rodriguez, J. A., Cascio, D., Chong, L., Gonen, T. and Eisenberg, D. S. (2018) Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. Science. **359**, 698-701
107 Luo, F., Gui, X., Zhou, H., Gu, J., Li, Y., Liu, X., Zhao, M., Li, D., Li, X. and Liu, C. (2018) Atomic structures of FUS LC domain segments reveal bases for reversible amyloid fibril formation. Nature structural & molecular biology. **25**, 341-346
108 King, O. D., Gitler, A. D. and Shorter, J. (2012) The tip of the iceberg: RNA-binding proteins with prion-like domains in neurodegenerative disease. Brain research. **1462**, 61-80
109 Lin, Y., Currie, S. L. and Rosen, M. K. (2017) Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. Journal of Biological Chemistry. **292**, 19110-19120
110 Martin, E. W., Holehouse, A. S., Peran, I., Farag, M., Incicco, J. J., Bremer, A., Grace, C. R., Soranno, A., Pappu, R. V. and Mittag, T. (2020) Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science. **367**, 694-699
111 Murthy, A. C., Dignon, G. L., Kan, Y., Zerze, G. H., Parekh, S. H., Mittal, J. and Fawzi, N. L. (2019) Molecular interactions underlying liquid− liquid phase separation of the FUS low-complexity domain. Nature structural & molecular biology. **26**, 637-648
112 Schuster, B., Dignon, G. L., Tang, W. S., Kelley, F., Ranganath, A. K., Jahnke, C. N., Simpkins, A. G., Regy, R. M., Hammer, D. A. and Good, M. C. (2020) Identifying Sequence Perturbations to an Intrinsically Disordered Protein that Determine Its Phase Separation Behavior. bioRxiv
113 Lin, Y.-H., Forman-Kay, J. D. and Chan, H. S. (2016) Sequence-specific polyampholyte phase separation in membraneless organelles. Physical review letters. **117**, 178101
114 Huggins, M. L. (1941) Solutions of long chain compounds. The Journal of chemical physics. **9**, 440-440
115 Flory, P. J. (1941) Thermodynamics of high polymer solutions. The Journal of Chemical Physics. **9**, 660-660
116 Debye, P. and Hückel, E. (1923) Zur theorie der elektrolyte. II. Das Grenzgesetz für die elektrische Leitfähigkeit. Phys. **305**

117 Deraguin, B. V. and Landau, L. (1941) Theory of the stability of strongly charged lyophobic sols and of the adhesion of strongly charged particles in solution of electrolytes. Acta Physicochim: USSR. **14**, 633-662
118 Verwey, E. J., Overbeek, J. T. G. and Nes, K. v. (1948) Theory of the stability of lyophobic colloids: the interaction of sol particles having an electric double layer.
119 Das, S., Amin, A. N., Lin, Y.-H. and Chan, H. S. (2018) Coarse-grained residue-based models of disordered protein condensates: utility and limitations of simple charge pattern parameters. Physical Chemistry Chemical Physics. **20**, 28558-28574
120 Kmiecik, S., Gront, D., Kolinski, M., Wieteska, L., Dawid, A. E. and Kolinski, A. (2016) Coarse-grained protein models and their applications. Chemical reviews. **116**, 7898-7936
121 Kolinski, A. and Skolnick, J. (1996) Lattice models of protein folding, dynamics, and thermodynamics. RG Landes
122 Mirny, L. and Shakhnovich, E. (2001) Protein folding theory: from lattice to all-atom models. Annual review of biophysics and biomolecular structure. **30**, 361-396
123 Hazra, M. K. and Levy, Y. (2020) Charge pattern affects the structure and dynamics of polyampholyte condensates. Physical Chemistry Chemical Physics. **22**, 19368-19375
124 Mészáros, B., Erdős, G., Szabó, B., Schád, É., Tantos, Á., Abukhairan, R., Horváth, T., Murvai, N., Kovács, O. P. and Kovács, M. (2020) PhaSePro: the database of proteins driving liquid–liquid phase separation. Nucleic acids research. **48**, D360-D367
125 Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O. G. and Pappu, R. V. (2017) CIDER: Resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. Biophysical journal. **112**, 16-21
126 Alberti, S. and Dormann, D. (2019) Liquid-liquid phase separation in disease. Annu. Rev. Genet. **53**, 171-194
127 Babinchak, W. M. and Surewicz, W. K. (2020) Liquid–Liquid Phase Separation and its Mechanistic Role in Pathological Protein Aggregation. Journal of Molecular Biology
128 Verdile, V., De Paola, E. and Paronetto, M. P. (2019) Aberrant phase transitions: side effects and novel therapeutic strategies in human disease. Frontiers in genetics. **10**, 173
129 Webber, C. J., Lei, S. E. and Wolozin, B. (2020) The pathophysiology of neurodegenerative disease: Disturbing the balance between phase separation and irreversible aggregation. Progress in molecular biology and translational science. **174**, 187-223
130 Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T., Grossman, M. and Clark, C. M. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science. **314**, 130-133
131 Cohen, T. J., Lee, V. M. Y. and Trojanowski, J. Q. (2011) TDP-43 functions and pathogenic mechanisms implicated in TDP-43 proteinopathies. Trends in molecular medicine. **17**, 659-667
132 Prasad, A., Bharathi, V., Sivalingam, V., Girdhar, A. and Patel, B. K. (2019) Molecular mechanisms of TDP-43 misfolding and pathology in amyotrophic lateral sclerosis. Frontiers in molecular neuroscience. **12**, 25
133 Johnson, B. S., Snead, D., Lee, J. J., McCaffery, J. M., Shorter, J. and Gitler, A. D. (2009) TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. Journal of Biological Chemistry. **284**, 20329-20339
134 Guo, W., Chen, Y., Zhou, X., Kar, A., Ray, P., Chen, X., Rao, E. J., Yang, M., Ye, H. and Zhu, L. (2011) An ALS-associated mutation affecting TDP-43 enhances protein aggregation, fibril formation and neurotoxicity. Nature structural & molecular biology. **18**, 822-830
135 de Boer, E. M. J., Orie, V. K., Williams, T., Baker, M. R., De Oliveira, H. M., Polvikoski, T., Silsby, M., Menon, P., van den Bos, M. and Halliday, G. M. (2021) TDP-43 proteinopathies: a new wave of neurodegenerative diseases. Journal of Neurology, Neurosurgery & Psychiatry. **92**, 86-95

136 Conicella, A. E., Zerze, G. H., Mittal, J. and Fawzi, N. L. (2016) ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. Structure. **24**, 1537-1549
137 Lin, Y., Protter, D. S. W., Rosen, M. K. and Parker, R. (2015) Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. Molecular cell. **60**, 208-219
138 Monahan, Z., Ryan, V. H., Janke, A. M., Burke, K. A., Rhoads, S. N., Zerze, G. H., O'Meally, R., Dignon, G. L., Conicella, A. E. and Zheng, W. (2017) Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. The EMBO journal. **36**, 2951-2967
139 French, R. L., Grese, Z. R., Aligireddy, H., Dhavale, D. D., Reeb, A. N., Kedia, N., Kotzbauer, P. T., Bieschke, J. and Ayala, Y. M. (2019) Detection of TAR DNA-binding protein 43 (TDP-43) oligomers as initial intermediate species during aggregate formation. Journal of Biological Chemistry. **294**, 6696-6709
140 Sun, Y., Medina Cruz, A., Hadley, K. C., Galant, N. J., Law, R., Vernon, R. M., Morris, V. K., Robertson, J. and Chakrabartty, A. (2018) Physiologically important electrolytes as regulators of TDP-43 aggregation and droplet-phase behavior. Biochemistry. **58**, 590-607
141 Babinchak, W. M., Haider, R., Dumm, B. K., Sarkar, P., Surewicz, K., Choi, J.-K. and Surewicz, W. K. (2019) The role of liquid–liquid phase separation in aggregation of the TDP-43 low-complexity domain. Journal of Biological Chemistry. **294**, 6306-6317
142 Bose, M., Lampe, M., Mahamid, J. and Ephrussi, A. (2022) Liquid-to-solid phase transition of oskar ribonucleoprotein granules is essential for their function in Drosophila embryonic development. Cell. **185**, 1308-1324
143 Snead, W. T. and Gladfelter, A. S. (2019) The control centers of biomolecular phase separation: how membrane surfaces, PTMs, and active processes regulate condensation. Molecular cell. **76**, 295-305
144 Banjade, S., Wu, Q., Mittal, A., Peeples, W. B., Pappu, R. V. and Rosen, M. K. (2015) Conserved interdomain linker promotes phase separation of the multivalent adaptor protein Nck. Proceedings of the National Academy of Sciences. **112**, E6426-E6435
145 Huang, W. Y. C., Yan, Q., Lin, W.-C., Chung, J. K., Hansen, S. D., Christensen, S. M., Tu, H.-L., Kuriyan, J. and Groves, J. T. (2016) Phosphotyrosine-mediated LAT assembly on membranes drives kinetic bifurcation in recruitment dynamics of the Ras activator SOS. Proceedings of the National Academy of Sciences. **113**, 8218-8223
146 Su, X., Ditlev, J. A., Hui, E., Xing, W., Banjade, S., Okrut, J., King, D. S., Taunton, J., Rosen, M. K. and Vale, R. D. (2016) Phase separation of signaling molecules promotes T cell receptor signal transduction. Science. **352**, 595-599
147 Su, X., Ditlev, J. A., Rosen, M. K. and Vale, R. D. (2017) Reconstitution of TCR signaling using supported lipid bilayers. In The Immune Synapse. pp. 65-76, Springer
148 Iversen, L., Tu, H.-L., Lin, W.-C., Christensen, S. M., Abel, S. M., Iwig, J., Wu, H.-J., Gureasko, J., Rhodes, C. and Petit, R. S. (2014) Ras activation by SOS: Allosteric regulation by altered fluctuation dynamics. Science. **345**, 50-54
149 Ma, W. and Mayr, C. (2018) A membraneless organelle associated with the endoplasmic reticulum enables 3′ UTR-mediated protein-protein interactions. Cell. **175**, 1492-1506
150 Hanna Iv, M. G., Block, S., Frankel, E. B., Hou, F., Johnson, A., Yuan, L., Knight, G., Moresco, J. J., Yates Iii, J. R. and Ashton, R. (2017) TFG facilitates outer coat disassembly on COPII transport carriers to promote tethering and fusion with ER–Golgi intermediate compartments. Proceedings of the National Academy of Sciences. **114**, E7707-E7716
151 Johnson, A., Bhattacharya, N., Hanna, M., Pennington, J. G., Schuh, A. L., Wang, L., Otegui, M. S., Stagg, S. M. and Audhya, A. (2015) TFG clusters COPII-coated transport carriers and promotes early secretory pathway organization. The EMBO journal. **34**, 811-827
152 Brangwynne, C. P., Tompa, P. and Pappu, R. V. (2015) Polymer physics of intracellular phase transitions. Nature Physics. **11**, 899-904

153 Guillén-Boixet, J., Buzon, V., Salvatella, X. and Méndez, R. (2016) CPEB4 is regulated during cell cycle by ERK2/Cdk1-mediated phosphorylation and its assembly into liquid-like droplets. Elife. **5**, e19298
154 Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. and Zweckstetter, M. (2017) Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. Nature communications. **8**, 1-13
155 Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K. M., Bennett, R. E., Dujardin, S., Laskowski, P. R., MacKenzie, D., Kamath, T. and Commins, C. (2018) Tau protein liquid–liquid phase separation can initiate tau aggregation. The EMBO journal. **37**
156 Hofweber, M., Hutten, S., Bourgeois, B., Spreitzer, E., Niedner-Boblenz, A., Schifferer, M., Ruepp, M.-D., Simons, M., Niessing, D. and Madl, T. (2018) Phase separation of FUS is suppressed by its nuclear import receptor and arginine methylation. Cell. **173**, 706-719
157 Qamar, S., Wang, G., Randle, S. J., Ruggeri, F. S., Varela, J. A., Lin, J. Q., Phillips, E. C., Miyashita, A., Williams, D. and Ströhl, F. (2018) FUS phase separation is modulated by a molecular chaperone and methylation of arginine cation-π interactions. Cell. **173**, 720-734
158 Ryan, V. H., Dignon, G. L., Zerze, G. H., Chabata, C. V., Silva, R., Conicella, A. E., Amaya, J., Burke, K. A., Mittal, J. and Fawzi, N. L. (2018) Mechanistic view of hnRNPA2 low-complexity domain structure, interactions, and phase separation altered by mutation and arginine methylation. Molecular cell. **69**, 465-479
159 Tanikawa, C., Ueda, K., Suzuki, A., Iida, A., Nakamura, R., Atsuta, N., Tohnai, G., Sobue, G., Saichi, N. and Momozawa, Y. (2018) Citrullination of RGG motifs in FET proteins by PAD4 regulates protein aggregation and ALS susceptibility. Cell reports. **22**, 1473-1483
160 Carlomagno, Y., Chung, D.-e. C., Yue, M., Castanedes-Casey, M., Madden, B. J., Dunmore, J., Tong, J., DeTure, M., Dickson, D. W. and Petrucelli, L. (2017) An acetylation–phosphorylation switch that regulates tau aggregation propensity and function. Journal of Biological Chemistry. **292**, 15277-15286
161 Ferreon, J. C., Jain, A., Choi, K.-J., Tsoi, P. S., MacKenzie, K. R., Jung, S. Y. and Ferreon, A. C. (2018) Acetylation disfavors tau phase separation. International journal of molecular sciences. **19**, 1360
162 Cohen, T. J., Hwang, A. W., Restrepo, C. R., Yuan, C.-X., Trojanowski, J. Q. and Lee, V. M. Y. (2015) An acetylation switch controls TDP-43 function and aggregation propensity. Nature communications. **6**, 1-13
163 Mateju, D., Franzmann, T. M., Patel, A., Kopach, A., Boczek, E. E., Maharana, S., Lee, H. O., Carra, S., Hyman, A. A. and Alberti, S. (2017) An aberrant phase transition of stress granules triggered by misfolded protein and prevented by chaperone function. The EMBO journal. **36**, 1669-1687
164 Hilliker, A., Gao, Z., Jankowsky, E. and Parker, R. (2011) The DEAD-box protein Ded1 modulates translation by the formation and resolution of an eIF4F-mRNA complex. Molecular cell. **43**, 962-972
165 Murray, D. T., Kato, M., Lin, Y., Thurber, K. R., Hung, I., McKnight, S. L. and Tycko, R. (2017) Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. Cell. **171**, 615-627
166 Rhine, K., Vidaurre, V. and Myong, S. (2020) RNA droplets. Annual review of biophysics. **49**, 247
167 Aumiller Jr, W. M., Pir Cakmak, F., Davis, B. W. and Keating, C. D. (2016) RNA-based coacervates as a model for membraneless organelles: formation, properties, and interfacial liposome assembly. Langmuir. **32**, 10042-10053
168 Fay, M. M., Anderson, P. J. and Ivanov, P. (2017) ALS/FTD-associated C9ORF72 repeat RNA promotes phase transitions in vitro and in cells. Cell reports. **21**, 3573-3584
169 Jain, A. and Vale, R. D. (2017) RNA phase transitions in repeat expansion disorders. Nature. **546**, 243-247

170 Langdon, E. M., Qiu, Y., Ghanbari Niaki, A., McLaughlin, G. A., Weidmann, C. A., Gerbich, T. M., Smith, J. A., Crutchley, J. M., Termini, C. M. and Weeks, K. M. (2018) mRNA structure determines specificity of a polyQ-driven phase separation. Science. **360**, 922-927
171 Maharana, S., Wang, J., Papadopoulos, D. K., Richter, D., Pozniakovsky, A., Poser, I., Bickle, M., Rizk, S., Guillén-Boixet, J. and Franzmann, T. M. (2018) RNA buffers the phase separation behavior of prion-like RNA binding proteins. Science. **360**, 918-921
172 Saha, S., Weber, C. A., Nousch, M., Adame-Arana, O., Hoege, C., Hein, M. Y., Osborne-Nishimura, E., Mahamid, J., Jahnel, M. and Jawerth, L. (2016) Polar positioning of phase-separated liquid compartments in cells regulated by an mRNA competition mechanism. Cell. **166**, 1572-1584
173 Boeynaems, S., Holehouse, A. S., Weinhardt, V., Kovacs, D., Van Lindt, J., Larabell, C., Van Den Bosch, L., Das, R., Tompa, P. S. and Pappu, R. V. (2019) Spontaneous driving forces give rise to protein− RNA condensates with coexisting phases and complex material properties. Proceedings of the National Academy of Sciences. **116**, 7889-7898
174 Petrovic, A. G. and Polavarapu, P. L. (2008) The Quadruplex− Duplex Structural Transition of Polyriboguanylic Acid. The Journal of Physical Chemistry B. **112**, 2245-2254
175 Boeynaems, S., Alberti, S., Fawzi, N. L., Mittag, T., Polymenidou, M., Rousseau, F., Schymkowitz, J., Shorter, J., Wolozin, B. and Van Den Bosch, L. (2018) Protein phase separation: a new phase in cell biology. Trends in cell biology. **28**, 420-435
176 Hafner, M., Landthaler, M., Burger, L., Khorshid, M., Hausser, J., Berninger, P., Rothballer, A., Ascano, M., Jungkamp, A.-C. and Munschauer, M. (2010) PAR-CliP-a method to identify transcriptome-wide the binding sites of RNA binding proteins. JoVE (Journal of Visualized Experiments), e2034
177 Boeynaems, S., Bogaert, E., Kovacs, D., Konijnenberg, A., Timmerman, E., Volkov, A., Guharoy, M., De Decker, M., Jaspers, T. and Ryan, V. H. (2017) Phase separation of C9orf72 dipeptide repeats perturbs stress granule dynamics. Molecular cell. **65**, 1044-1055
178 Brocca, S., Grandori, R., Longhi, S. and Uversky, V. (2020) Liquid–liquid phase separation by intrinsically disordered protein regions of viruses: Roles in viral life cycle and control of virus–host interactions. International Journal of Molecular Sciences. **21**, 9045
179 Li, H., Ernst, C., Kolonko-Adamska, M., Greb-Markiewicz, B., Man, J., Parissi, V. and Ng, B. W.-L. (2022) Phase separation in viral infections. Trends in Microbiology
180 Nikolic, J., Le Bars, R., Lama, Z., Scrima, N., Lagaudrière-Gesbert, C., Gaudin, Y. and Blondel, D. (2017) Negri bodies are viral factories with properties of liquid organelles. Nature communications. **8**, 1-13
181 Lahaye, X., Vidy, A., Pomier, C., Obiang, L., Harper, F., Gaudin, Y. and Blondel, D. (2009) Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication. Journal of virology. **83**, 7948-7958
182 Fouquet, B., Nikolic, J., Larrous, F., Bourhy, H., Wirblich, C., Lagaudrière-Gesbert, C. and Blondel, D. (2015) Focal adhesion kinase is involved in rabies virus infection through its interaction with viral phosphoprotein P. Journal of virology. **89**, 1640-1651
183 Heinrich, B. S., Maliga, Z., Stein, D. A., Hyman, A. A. and Whelan, S. P. J. (2018) Phase transitions drive the formation of vesicular stomatitis virus replication compartments. MBio. **9**, e02290-02217
184 Zhou, Y., Su, J. M., Samuel, C. E. and Ma, D. (2019) Measles virus forms inclusion bodies with properties of liquid organelles. Journal of virology. **93**, e00948-00919
185 Zhang, S., Jiang, Y., Cheng, Q., Zhong, Y., Qin, Y. and Chen, M. (2017) Inclusion body fusion of human parainfluenza virus type 3 regulated by acetylated α-tubulin enhances viral replication. Journal of virology. **91**, e01802-01816
186 Katoh, H., Kubota, T., Kita, S., Nakatsu, Y., Aoki, N., Mori, Y., Maenaka, K., Takeda, M. and Kidokoro, M. (2015) Heat shock protein 70 regulates degradation of the mumps virus phosphoprotein via the ubiquitin-proteasome pathway. Journal of Virology. **89**, 3188-3199

187 Ringel, M., Heiner, A., Behner, L., Halwe, S., Sauerhering, L., Becker, N., Dietzel, E., Sawatsky, B., Kolesnikova, L. and Maisner, A. (2019) Nipah virus induces two inclusion body populations: Identification of novel inclusions at the plasma membrane. PLoS pathogens. **15**, e1007733
188 Alenquer, M., Vale-Costa, S., Etibor, T. A., Ferreira, F., Sousa, A. L. and Amorim, M. J. (2019) Influenza A virus ribonucleoproteins form liquid organelles at endoplasmic reticulum exit sites. Nature communications. **10**, 1-19
189 Monette, A., Niu, M., Chen, L., Rao, S., Gorelick, R. J. and Mouland, A. J. (2020) Pan-retroviral nucleocapsid-mediated phase separation regulates genomic RNA positioning and trafficking. Cell reports. **31**, 107520
190 Perdikari, T. M., Murthy, A. C., Ryan, V. H., Watters, S., Naik, M. T. and Fawzi, N. L. (2020) SARS-CoV-2 nucleocapsid protein undergoes liquid-liquid phase separation stimulated by RNA and partitions into phases of human ribonucleoproteins. BioRxiv
191 Espinosa, J. R., Joseph, J. A., Sanchez-Burgos, I., Garaizar, A., Frenkel, D. and Collepardo-Guevara, R. (2020) Liquid network connectivity regulates the stability and composition of biomolecular condensates with many components. Proceedings of the National Academy of Sciences. **117**, 13238-13247
192 Emara, M. M. and Brinton, M. A. (2007) Interaction of TIA-1/TIAR with West Nile and dengue virus products in infected cells interferes with stress granule formation and processing body assembly. Proceedings of the National Academy of Sciences. **104**, 9041-9046
193 Ohn, T., Kedersha, N., Hickman, T., Tisdale, S. and Anderson, P. (2008) A functional RNAi screen links O-GlcNAc modification of ribosomal proteins to stress granule and processing body assembly. Nature cell biology. **10**, 1224-1231
194 Fricke, J., Koo, L. Y., Brown, C. R. and Collins, P. L. (2013) p38 and OGT sequestration into viral inclusion bodies in cells infected with human respiratory syncytial virus suppresses MK2 activities and stress granule assembly. Journal of virology. **87**, 1333-1347
195 Jobe, F., Simpson, J., Hawes, P., Guzman, E. and Bailey, D. (2020) Respiratory syncytial virus sequesters NF-κB subunit p65 to cytoplasmic inclusion bodies to inhibit innate immune signaling. Journal of virology. **94**, e01380-01320
196 Sun, Y. and López, C. B. (2017) The innate immune response to RSV: Advances in our understanding of critical viral and host factors. Vaccine. **35**, 481-488
197 Lifland, A. W., Jung, J., Alonas, E., Zurla, C., Crowe Jr, J. E. and Santangelo, P. J. (2012) Human respiratory syncytial virus nucleoprotein and inclusion bodies antagonize the innate immune response mediated by MDA5 and MAVS. Journal of virology. **86**, 8245-8258
198 Risso-Ballester, J., Galloux, M., Cao, J., Le Goffic, R., Hontonnou, F., Jobart-Malfait, A., Desquesnes, A., Sake, S. M., Haid, S. and Du, M. (2021) A condensate-hardening drug blocks RSV replication in vivo. Nature. **595**, 596-599
199 Wang, S., Dai, T., Qin, Z., Pan, T., Chu, F., Lou, L., Zhang, L., Yang, B., Huang, H. and Lu, H. (2021) Targeting liquid–liquid phase separation of SARS-CoV-2 nucleocapsid protein promotes innate antiviral immunity by elevating MAVS activity. Nature Cell Biology. **23**, 718-732
200 Uversky, V. N. (2012) Size-exclusion chromatography in structural analysis of intrinsically disordered proteins. In Intrinsically disordered protein analysis. pp. 179-194, Springer
201 Le Maire, M., Ghazi, A., Møller, J. V. and Aggerbeck, L. P. (1987) The use of gel chromatography for the determination of sizes and relative molecular masses of proteins. Interpretation of calibration curves in terms of gel-pore-size distribution. Biochemical Journal. **243**, 399-404
202 Ui, N. (1979) Rapid estimation of the molecular weights of protein polypeptide chains using high-pressure liquid chromatography in 6 M guanidine hydrochloride. Analytical Biochemistry. **97**, 65-71
203 Uversky, V. N. and Dunker, A. K. (2010) Understanding protein non-folding. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics. **1804**, 1231-1264

204 Uversky, V. N. and Dunker, A. K. (2012) Multiparametric analysis of intrinsically disordered proteins: looking at intrinsic disorder through compound eyes. ed.)^eds.), ACS Publications
205 Uversky, V. N. (2002) What does it mean to be natively unfolded? European Journal of Biochemistry. **269**, 2-12
206 Uversky, V. N., Li, J., Souillac, P., Millett, I. S., Doniach, S., Jakes, R., Goedert, M. and Fink, A. L. (2002) Biophysical properties of the synucleins and their propensities to fibrillate: inhibition of α-synuclein assembly by β-and γ-synucleins. Journal of Biological Chemistry. **277**, 11970-11978
207 Kohn, J. E., Millett, I. S., Jacob, J., Zagrovic, B., Dillon, T. M., Cingel, N., Dothager, R. S., Seifert, S., Thiyagarajan, P. and Sosnick, T. R. (2004) Random-coil behavior and the dimensions of chemically unfolded proteins. Proceedings of the National Academy of Sciences. **101**, 12491-12496
208 Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A. and Smith, L. J. (1999) Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques. Biochemistry. **38**, 16424-16431
209 Beveridge, R., Chappuis, Q., Macphee, C. and Barran, P. (2013) Mass spectrometry methods for intrinsically disordered proteins. Analyst. **138**, 32-42
210 Taflin, D. C., Ward, T. L. and Davis, E. J. (1989) Electrified droplet fission and the Rayleigh limit. Langmuir. **5**, 376-384
211 Cech, N. B. and Enke, C. G. (2001) Practical implications of some recent studies in electrospray ionization fundamentals. Mass spectrometry reviews. **20**, 362-387
212 Dole, M., Mack, L. L., Hines, R. L., Mobley, R. C., Ferguson, L. D. and Alice, M. B. (1968) Molecular beams of macroions. The Journal of chemical physics. **49**, 2240-2249
213 Iribarne, J. V. and Thomson, B. A. (1976) On the evaporation of charged ions from small droplets. J. Chem. Phys. **64**, 2287-2294
214 Kaltashov, I. A. and Abzalimov, R. R. (2008) Do ionic charges in ESI MS provide useful information on macromolecular structure? Journal of the American Society for Mass Spectrometry. **19**, 1239-1246
215 Li, J., Santambrogio, C., Brocca, S., Rossetti, G., Carloni, P. and Grandori, R. (2016) Conformational effects in protein electrospray-ionization mass spectrometry. Mass spectrometry reviews. **35**, 111-122
216 Santambrogio, C., Natalello, A., Brocca, S., Ponzini, E. and Grandori, R. (2019) Conformational characterization and classification of intrinsically disordered proteins by native mass spectrometry and charge-state distribution analysis. Proteomics. **19**, 1800060
217 Natalello, A., Santambrogio, C. and Grandori, R. (2016) Are charge-state distributions a reliable tool describing molecular ensembles of intrinsically disordered proteins by native MS? Journal of The American Society for Mass Spectrometry. **28**, 21-28
218 Mirza, U. A. and Chait, B. T. (1994) Effects of anions on the positive ion electrospray ionization mass spectra of peptides and proteins. Analytical chemistry. **66**, 2898-2904
219 Testa, L., Brocca, S. and Grandori, R. (2011) Charge-surface correlation in electrospray ionization of folded and unfolded proteins. Analytical chemistry. **83**, 6459-6463
220 Hall, Z. and Robinson, C. V. (2012) Do charge state signatures guarantee protein conformations? Journal of the American Society for Mass Spectrometry. **23**, 1161-1168
221 Kaltashov, I. A. and Mohimen, A. (2005) Estimates of protein surface areas in solution by electrospray ionization mass spectrometry. Analytical chemistry. **77**, 5370-5379
222 Estrada, J., Bernadó, P., Blackledge, M. and Sancho, J. (2009) ProtSA: a web application for calculating sequence specific protein solvent accessibilities in the unfolded ensemble. BMC bioinformatics. **10**, 1-8

223 Brocca, S., Testa, L., Sobott, F., Šamalikova, M., Natalello, A., Papaleo, E., Lotti, M., De Gioia, L., Doglia, S. M. and Alberghina, L. (2011) Compaction properties of an intrinsically disordered protein: Sic1 and its kinase-inhibitor domain. Biophysical journal. **100**, 2243-2252
224 Natalello, A., Benetti, F., Doglia, S. M., Legname, G. and Grandori, R. (2011) Compact conformations of α-synuclein induced by alcohols and copper. Proteins: Structure, Function, and Bioinformatics. **79**, 611-621
225 Frimpong, A. K., Abzalimov, R. R., Uversky, V. N. and Kaltashov, I. A. (2010) Characterization of intrinsically disordered proteins with electrospray ionization mass spectrometry: conformational heterogeneity of α-synuclein. Proteins: Structure, Function, and Bioinformatics. **78**, 714-722
226 Phillips, A. S., Gomes, A. F., Kalapothakis, J. M. D., Gillam, J. E., Gasparavicius, J., Gozzo, F. C., Kunath, T., MacPhee, C. and Barran, P. E. (2015) Conformational dynamics of α-synuclein: insights from mass spectrometry. Analyst. **140**, 3070-3081
227 Nakamura, K., Nemani, V. M., Wallender, E. K., Kaehlcke, K., Ott, M. and Edwards, R. H. (2008) Optical reporters for the conformation of α-synuclein reveal a specific interaction with mitochondria. Journal of Neuroscience. **28**, 12305-12317
228 Bisaglia, M., Mammi, S. and Bubacco, L. (2009) Structural insights on physiological functions and pathological effects of±-synuclein. The FASEB Journal. **23**, 329-340
229 Li, C., Lutz, E. A., Slade, K. M., Ruf, R. A. S., Wang, G.-F. and Pielak, G. J. (2009) 19F NMR studies of α-synuclein conformation and fibrillation. Biochemistry. **48**, 8578-8584
230 Lee, J. C., Langen, R., Hummel, P. A., Gray, H. B. and Winkler, J. R. (2004) α-Synuclein structures from fluorescence energy-transfer kinetics: Implications for the role of the protein in Parkinson's disease. Proceedings of the National Academy of Sciences. **101**, 16466-16471
231 Kaltashov, I. A., Bobst, C. E. and Abzalimov, R. R. (2013) Mass spectrometry-based methods to study protein architecture and dynamics. Protein Science. **22**, 530-544
232 Brucale, M., Tessari, I., Bubacco, L. and Samorì, B. (2012) Single-molecule force spectroscopy of chimeric polyprotein constructs containing intrinsically disordered domains. In Intrinsically Disordered Protein Analysis. pp. 47-56, Springer
233 Curtain, C. C., Kirby, N. M., Mertens, H. D. T., Barnham, K. J., Knott, R. B., Masters, C. L., Cappai, R., Rekas, A., Kenche, V. B. and Ryan, T. (2015) Alpha-synuclein oligomers and fibrils originate in two distinct conformer pools: a small angle X-ray scattering and ensemble optimisation modelling study. Molecular BioSystems. **11**, 190-196
234 D'Urzo, A., Santambrogio, C., Grandori, R., Ciurli, S. and Zambelli, B. (2014) The conformational response to Zn (II) and Ni (II) binding of Sporosarcina pasteurii UreG, an intrinsically disordered GTPase. JBIC Journal of Biological Inorganic Chemistry. **19**, 1341-1354
235 Beveridge, R., Covill, S., Pacholarz, K. J., Kalapothakis, J. M. D., MacPhee, C. E. and Barran, P. E. (2014) A mass-spectrometry-based framework to define the extent of disorder in proteins. Analytical chemistry. **86**, 10979-10991
236 Santambrogio, C., Ponzini, E. and Grandori, R. (2022) Native mass spectrometry for the investigation of protein structural (dis) order. Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics, 140828
237 Testa, L., Brocca, S., Santambrogio, C., D'Urzo, A., Habchi, J., Longhi, S., Uversky, V. N. and Grandori, R. (2013) Extracting structural information from charge-state distributions of intrinsically disordered proteins by non-denaturing electrospray-ionization mass spectrometry. Intrinsically Disordered Proteins. **1**, e25068
238 Natalello, A. and Doglia, S. M. (2010) Intrinsically disordered proteins and induced folding studied by fourier transform infrared spectroscopy. Instrumental analysis of intrinsically disordered proteins, VN Uversky and S. Longhi, Editors. 2010, Wiley, 225-252
239 Dutta, A. (2017) Fourier transform infrared spectroscopy. Spectroscopic methods for nanomaterials characterization, 73-93

240 Barth, A. (2007) Infrared spectroscopy of proteins. Biochimica et Biophysica Acta (BBA)-Bioenergetics. **1767**, 1073-1101
241 Cai, S. and Singh, B. R. (2004) A distinct utility of the amide III infrared band for secondary structure estimation of aqueous protein solutions using partial least squares methods. Biochemistry. **43**, 2541-2549
242 Krimm, S. and Bandekar, J. (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. Advances in protein chemistry. **38**, 181-364
243 Chirgadze, Y. N., Fedorov, O. V. and Trushina, N. P. (1975) Estimation of amino acid residue side-chain absorption in the infrared spectra of protein solutions in heavy water. Biopolymers: Original Research on Biomolecules. **14**, 679-694
244 Kong, J. and Yu, S. (2007) Fourier transform infrared spectroscopic analysis of protein secondary structures. Acta biochimica et biophysica Sinica. **39**, 549-559
245 Smith, L. J., Fiebig, K. M., Schwalbe, H. and Dobson, C. M. (1996) The concept of a random coil: Residual structure in peptides and denatured proteins. Folding and Design. **1**, R95-R106
246 Natalello, A., Ami, D. and Doglia, S. M. (2012) Fourier transform infrared spectroscopy of intrinsically disordered proteins: measurement procedures and data analyses. In Intrinsically disordered protein analysis. pp. 229-244, Springer
247 Susi, H. and Byler, D. M. (1986) [13] Resolution-enhanced fourier transform infrared spectroscopy of enzymes. In Methods in enzymology. pp. 290-311, Elsevier
248 Zandomeneghi, G., Krebs, M. R. H., McCammon, M. G. and Fändrich, M. (2004) FTIR reveals structural differences between native β-sheet proteins and amyloid fibrils. Protein science. **13**, 3314-3321
249 Torii, H., Tatsumi, T. and Tasumi, M. (1998) Effects of hydration on the structure, vibrational wavenumbers, vibrational force field and resonance raman intensities of N-methylacetamide. Journal of Raman Spectroscopy. **29**, 537-546
250 Barth, A. and Zscherp, C. (2002) What vibrations tell about proteins. Quarterly reviews of biophysics. **35**, 369-430
251 Ami, D., Lavatelli, F., Rognoni, P., Palladini, G., Raimondi, S., Giorgetti, S., Monti, L., Doglia, S. M., Natalello, A. and Merlini, G. (2016) In situ characterization of protein aggregates in human tissues affected by light chain amyloidosis: a FTIR microspectroscopy study. Scientific reports. **6**, 1-12
252 Ruysschaert, J.-M. and Raussens, V. (2018) ATR-FTIR analysis of amyloid proteins. In Peptide Self-Assembly. pp. 69-81, Springer
253 Peters, R. (2013) Fluorescence Photobleaching and Photoactivation Techniques. Fluorescence Microscopy, 215
254 Meddens, M. B. M., de Keijzer, S. and Cambi, A. (2014) High Spatiotemporal Bioimaging Techniques to Study the Plasma Membrane Nanoscale Organization. In Fluorescence Microscopy. pp. 49-63, Elsevier
255 Ishikawa-Ankerhold, H. C., Ankerhold, R. and Drummen, G. P. C. (2012) Advanced fluorescence microscopy techniques—Frap, Flip, Flap, Fret and flim. Molecules. **17**, 4047-4132
256 Banerjee, P. R., Milin, A. N., Moosa, M. M., Onuchic, P. L. and Deniz, A. A. (2017) Reentrant phase transition drives dynamic substructure formation in ribonucleoprotein droplets. Angewandte Chemie. **129**, 11512-11517
257 Kroschwald, S., Maharana, S., Mateju, D., Malinovska, L., Nüske, E., Poser, I., Richter, D. and Alberti, S. (2015) Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. elife. **4**, e06807
258 Chalupníková, K., Lattmann, S., Selak, N., Iwamoto, F., Fujiki, Y. and Nagamine, Y. (2008) Recruitment of the RNA helicase RHAU to stress granules via a unique RNA-binding domain. Journal of Biological Chemistry. **283**, 35186-35198

259 Weidtkamp-Peters, S., Lenser, T., Negorev, D., Gerstner, N., Hofmann, T. G., Schwanitz, G., Hoischen, C., Maul, G., Dittrich, P. and Hemmerich, P. (2008) Dynamics of component exchange at PML nuclear bodies. Journal of cell science. **121**, 2731-2743
260 Keating, C. D. (2021) Liquid-liquid Phase Coexistence and Membraneless Organelles. Academic Press
261 Taylor, N. O., Wei, M.-T., Stone, H. A. and Brangwynne, C. P. (2019) Quantifying dynamics in phase-separated condensates using fluorescence recovery after photobleaching. Biophysical journal. **117**, 1285-1300
262 Kaur, T., Alshareedah, I., Wang, W., Ngo, J., Moosa, M. M. and Banerjee, P. R. (2019) Molecular crowding tunes material states of ribonucleoprotein condensates. Biomolecules. **9**, 71
263 Zhang, H., Elbaum-Garfinkle, S., Langdon, E. M., Taylor, N., Occhipinti, P., Bridges, A. A., Brangwynne, C. P. and Gladfelter, A. S. (2015) RNA controls PolyQ protein phase transitions. Molecular cell. **60**, 220-230
264 Wei, M.-T., Elbaum-Garfinkle, S., Holehouse, A. S., Chen, C. C.-H., Feric, M., Arnold, C. B., Priestley, R. D., Pappu, R. V. and Brangwynne, C. P. (2017) Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. Nature chemistry. **9**, 1118-1125

5. Appendix

Section 5.1

Relevance of electrostatic charges in compactness, aggregation, and phase separation of intrinsically disordered proteins.

Bianchi, G., Longhi, S., Grandori, R., & Brocca, S. (2020). *International Journal of Molecular Sciences*, 21(17), 6208.

International Journal of

Molecular Sciences

MDPI

Review

Relevance of Electrostatic Charges in Compactness,Aggregation, and Phase Separation of Intrinsically Disordered Proteins

Greta Bianchi [1], Sonia Longhi [2], Rita Grandori [1] and Stefania Brocca [1,*]

[1] Department of Biotechnology and Biosciences, University of Milano-Bicocca, 20126 Milano, Italy; g.bianchi31@campus.unimib.it (G.B.); rita.grandori@unimib.it (R.G.)

[2] Laboratoire Architecture et Fonction des Macromolécules Biologiques (AFMB),

Aix-Marseille University and CNRS, UMR 7257, 13288 Marseille, France; sonia.longhi@afmb.univ-mrs.fr

[*] Correspondence: stefania.brocca@unimib.it; Tel.: +39-338-666-5111

Received: 31 July 2020; **Accepted**: 23 August 2020; **Published**: 27 August 2020

Abstract: The abundance of intrinsic disorder in the protein realm and its role in a variety of physiological and pathological cellular events have strengthened the interest of the scientific community in understanding the structural and dynamical properties of intrinsically disordered proteins (IDPs) and regions (IDRs). Attempts at rationalizing the general principles underlying both conformational properties and transitions of IDPs/IDRs must consider the abundance of charged residues (Asp, Glu, Lys, and Arg) that typifies these proteins, rendering them assimilable to polyampholytes or polyelectrolytes. Their conformation strongly depends on both the charge density and distribution along the sequence (i.e., charge decoration) as highlighted by recent experimental and theoretical studies that have introduced novel descriptors. Published experimental data are revisited herein in the frame of this formalism, in a new and possibly unitary perspective. The physicochemical properties most directly affected by charge density and distribution are compaction and solubility, which can be described in a relatively simplified way by tools of polymer physics. Dissecting factors controlling such properties could contribute to better understanding complex biological phenomena, such as fibrillation and phase separation. Furthermore, this knowledge is expected to have enormous practical implications for the design, synthesis, and exploitation of bio-derived materials and the control of natural biological processes.

Keywords: charge density; fraction of net charge; net charge per residue; charge decoration; linear pattern of charge distribution; charge segregation; polyampholyte; polyelectrolyte

1.Introduction

Intrinsically disordered proteins (IDPs) and protein regions (IDRs) escape the paradigm of protein folding and must, rather, be described as conformational ensembles of interconverting conformers. Conformers of biological relevance can be poorly populated and cannot be easily isolated, unless bound to a ligand or partner acting as a conformational stabilizer [1]. The cellular environment (pH, temperature, ionic force, concentration of osmolytes) can influence the relative distribution of such metastable conformers, thereby acting as biochemical sensors and signal transducers [2]. Not surprisingly, IDRs are often located at the protein N- or C-terminus and act as interaction hubs in protein–protein networks [3,4]. This feature is instrumental to IDP/IDR involvement in crucial physiological processes, such as transcription, translation, and cell cycle regulation [2,5–7], and underlies the relationships between IDPs/IDRs and diseases (cancer, inflammation, or neurodegeneration) [8,9]. A central challenge for structural biology is understanding how sequence and sequence composition encode structural disorder. Depletion in hydrophobic residues, enrichment in structure-breakers (particularly prolines and glycines), along with polar and charged residues, represent the most common compositional traits and have been employed as diagnostic traits of structural disorder [10,11].

In IDPs/IDRs, the most frequent amino acids with ionizable side chains (i.e., groups that ionize between pH 1 and 14) are Asp and Glu, Lys, and Arg. The ionization behavior is mainly dictated by their equilibrium constant of acid dissociation (the "intrinsic pK_a" value) and their electrostatic environment, including pH. Under physiological conditions of nearly neutral pH, all these residues are charged, although their pK_a can be influenced by several factors, such as dehydration by the Born effect, Coulomb and charge–dipole interactions [12]. Hence, the apparent pK_a can reflect the presence of neighboring peptide bonds, the proximity with other charged groups, hydrogen donors/acceptors, solvent

exposition, or burial inside a protein structure, being very sensitive to local conformation [12,13]. Several experimental and computational techniques can be applied to obtain pK_a values of residues embedded in a protein structure. Among these, nuclear magnetic resonance (NMR) spectroscopy allows measurement of the pH dependence of chemical shifts. Such experiments have highlighted that Asp, Glu, and Lys residues in disordered polypeptides, as well as in solvent-exposed regions of ordered proteins [12], have pK_a values close to those measured in random coil models [14], with short- and medium-range interactions prevailing on long-range electrostatic ones [15,16]. Arg represents a peculiar case. On one side, it is highly basic due to charge delocalization; on the other side, the guanidinium group is a very weakly hydrated cation [17]. This latter property facilitates Arg residue burial in hydrophobic micro-environments [18] and its stacking interactions with aromatic protein residues [17,19]. Theoretical studies have modeled IDPs as ideal "charge-decorated" polymers, drawing copiously from polymer physics to describe their peculiar behavior [20–24]. Indeed, in polymer physics, IDPs have been referred to as either polyelectrolytes, with multiple charges of the same sign, or, more frequently (~75% of IDPs), as polyampholytes, carrying both positive and negative charges [25]. The (un)balancing of opposite charges, i.e., the extent of net charge, affects the IDP conformational fate: Electrical neutrality enables polyampholytes to collapse, whereas unbalanced charges result in structural expansion due to repulsive forces [26]. This review focuses on the parameters introduced so far to capture the peculiar electrostatics of IDPs/IDRs and the involvement of electrostatic properties in their physio-pathological roles. In particular, we will discuss how charge density and distribution affect IDP/IDR compactness, aggregation, solubility, fibrillation, and phase separation. Basic information on compaction parameters, fibrillation, and phase separation are presented in three appendices. Describing the mathematical formalisms underlying polypeptide structure goes beyond the aims of this review, which is meant to

provide an overview of its practical implications of IDP conformational behavior in the field of protein science and cellular biochemistry.

2. From Mean Net Charge to Linear Patterns of Charged Residues

2.1.Compositional Classes of IDPs and Phase Diagrams of Protein Conformation

Protein collapsibility is governed by the interplay of intra-chain and chain–solvent interactions. The early concept of an empirical charge-hydropathy (C-H) correlation has been expressed by the so-called Uversky's plot [27], which classifies IDPs/IDRs according to their position in the two-dimensional space of mean hydrophobicity "H" and mean net absolute charge "*q*". This latter is equivalent to the absolute value of net charge per residue (|NCPR|), a more recently introduced parameter defined as the difference between the fractions of positively (f_+) and negatively (f_-) charged residues. In the original Uversky's plot, the line *"<q> = 2.785 <H> −1.151*" demarcates the boundary between IDPs/IDRs and natively folded proteins [27], assigning the same coordinates to oppositely charged polymers. To account for the polarity of charged proteins, more recent versions of the C-H plot represent the full range of NCPR values (*i.e.*, not only |NCPR|) with the "*H-q*" space crossed by two "mirror" boundaries *("<q> = 2.785 <H> −1.151"* and *"<q> = −2.785 <H> +1.151"*) (Figure 1a) [28]. Still, such plots fail to capture differences that may underlie polymers with similar NCPR values yet endowed with a different number of charged residues. This issue has been illustrated by atomistic simulations and experimental investigations by Pappu's group on a repertoire of protamines, small arginine-rich nuclear proteins [29]. In spite of their identical NCPR values, these polypeptides possess different dimensions and local conformational preferences. Therefore, with the aim of enhancing Uversky's C-H phase diagram, the Mao-Pappu's three-dimensional plot represents the hydrophobicity H on the vertical axis and (f_+) and (f_-) on two horizontal axes, where they vary independently and not cumulatively, as in the case

of NCPR (Figure 1b) [29]. Thereby, sequences with low mean hydrophobicity, which initially were collectively considered as "natively unfolded proteins", are now distinguished into "swollen coils" and "disordered globules", according to their fractions of charged residues. When $f_+ \gg f_-$ and vice versa, which implies large NCPR values, polypeptides can be considered as "polyelectrolytes" and may behave as extended swollen coils. Indeed, the presence of multiple unshielded charges induces chain expansion due to electrostatic repulsions and favorable polymer–solvent interactions, similarly to an ideal polymer in a good solvent [30,31]. On the other hand, when $f_+ \approx f_-$, NCPR is close to zero, and polypeptides are "polyampholytes", which behave as disordered globules governed by attractive interactions. Further, polyampholytes can be classified as "strong" or "weak", depending on whether they possess a large or small fraction of charged residues (FCR, calculated as the sum of f_+ and f_-) and display specific conformational preferences.

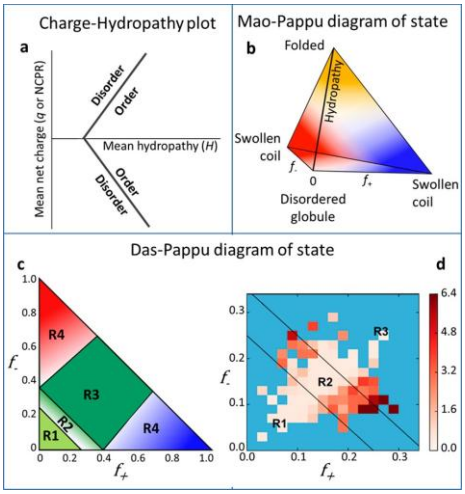


Figure 1. Conformational phase diagrams. (**a**) Uversky's plot of mean net charge (q or NCPR) versus hydropathy (H). The two solid lines mark the boundary between disordered and folded proteins [27,28]. (**b**) Mao-Pappu's phase diagram of conformations for intrinsically disordered proteins (IDPs) and protein regions

(IDRs). The three-dimensional sequence space is defined by f_+, f_-, and mean hydropathy. Since high hydropathy and high fractions of charged residues are mutually exclusive, the space is shaped as a pyramid. The yellow area, at the top of the "pyramid", represents naturally folded proteins, the red and blue regions at the base correspond, respectively, to negatively and positively charged polyelectrolytes (figure inspired by [29]). (**c**) Das-Pappu's phase diagram of IDP/IDR conformations. The diagram contains four regions (R1-R4) representing distinct conformational classes. R1, weak polyampholytes or weak polyelectrolytes that form globules or tadpole-like conformations. R3, strong polyampholytes that form non-globular conformations, such as coil-like, hairpin-like, or a mixture. R2, continuum of conformations between those in R1 and R3. R4, strong polyelectrolytes with FCR > 0.35 and |NCPR| > 0.3, which sample coil-like conformations approaching the excluded-volume limit [32]. (**d**) Heat map of the protein size distribution predicted by the analytical Ghosh's model applied to the DisProt entries. The bins correspond to (f_+, f_-) classes. The color scale represents the values of x, an expansion index calculated for each protein comparing the ensemble-averaged end-to-end distance predicted by Ghosh's model with the ensemble-averaged end-to-end distance in the Flory random coil limit, in the absence of any interaction, at T = 300 K. The map represents with color of different intensity the average of x values among proteins within a given bin. Blue bins correspond to classes containing less than two proteins. Black lines define R1, R2, and R3 regions as reported in panel C. Reproduced from [33] with the permission of AIP Publishing.

An even better correlation between FCR and IDP conformations is offered by a more recent version of the conformational disorder plot, the so-called Das-Pappu's phase diagram [11,25,32] (Figure 1c). Herein, low-NCPR IDPs/IDRs are no longer indiscriminately identified as globules, yet they occupy distinct conformational classes—globules, coils, hairpins, chimeras—according to their FCR values. Table

1 lists some examples of proteins belonging to each of these conformational classes. Nevertheless, this classification, as clearly stated by the authors, is valid for IDPs/IDRs of at least 30 residues, with a low overall hydropathy and low proline content [25,32]. Furthermore, it does not provide any insight into how protein dimension varies within these classes [33]. When comparing experimental data with predictions inspired to FCR, or more complex composition-based heuristics, collapsed globules turn out to be less frequent than predicted [33–36]. Possible reasons for these discrepancies could be searched in the weaknesses of either the experimental or the computational approaches: (i) Collapsed globules have higher aggregation propensity compared to expanded coils, hampering structural characterization at the high protein concentrations required for some biophysical techniques (e.g., NMR, small-angle X-ray scattering (SAXS), etc.); and (ii) the efficiency of prediction algorithms could be hindered by the complexity of the intramolecular interactions' governing compactness, as well as the interplay with the physicochemical environment [37].

Table 1. Examples of intrinsically disordered proteins (IDPs) and protein regions (IDRs) belonging to the distinct regions of the Das-Pappu's conformational phase diagram.

Protein Class	FCR	NCPR	Representative Proteins	References
R1, Globules	<0.25	<0.25	α-Synuclein (residues 1-100)	[11]
R2, Globules and coils	0.25 ≤ FCR ≤ 0.35	≤0.35	Tau repeat domain	[11]
R3, Polyampholyte coils or hairpins	>0.35	≤0.35	LEA_4 proteins[1], NSP12	[38]
R4, Polyelectrolytic semi-flexible rods or coils	>0.35	>0.35	Synthetic polyE and polyK; protamines, NP1[3], RAG2[4]	[29,32]

[1]Late Embryogenesis Abundant proteins from Arabidopsis thaliana; [2]nucleoporin Nsp1 (UniProt ID: P14907) of *S. cerevisiae*; [3]NP1 (UniProt ID: O13030), residues 5–24 of *Cynops pyrrhogaster* protamine 1; [4]RAG2 (UniProt ID: P21784), residues 392–411 ('acidic hinge') of *Cricetulus griseus* V(D)J recombination-activating protein 2.

2.2. The Concept of Linear Patterning of Charges and Its Parametrization

Asymmetry in electrostatic potentials is a recurrent feature in protein structure, found at the level of the protein backbone [39], secondary-structure elements [40], and supersecondary structure motifs [41]. Herein, we will focus on the effect of the charge distribution and polarization within the protein sequence, considering both the backbone structure and sequence-specific features encoded by the R-groups, i.e., its sequence specificity [37]. In this regard, an important aspect to be considered is the linear charge patterning. Indeed, while theoretical and computational works suggest that weak polyampholytes (i.e., low-FCR proteins) preferentially form globules, strong polyampholytes (i.e., high-FCR proteins) behave very differently from one another, according to the linear distribution of oppositely charged residues in their amino acid sequence [25].

The conformation of high-FCR proteins with an identical charge composition but different segregation of cationic and anionic residues was studied by Srivastava and Muthukumar already in the second half of the 1990s [26]. Monte Carlo (MC) simulations showed substantial differences in the radius of gyration (R_g, defined in Appendix A) between two groups of polymers, in which opposite charges are regularly interspersed or clustered at the two extremities, as a result of the interplay between intrachain electrostatic attractions and repulsions [26]. More recently, the same issue was systematically tackled by either computational simulations or scalable analytical theories, offering a coherent envision, yet using different parameters to quantitatively describe charge patterning. The group of Rohit Pappu has introduced the empirical parameter κ as a measure of the overall charge asymmetry [25]. Upon partitioning the protein sequence into N overlapping segments (or blobs, of a size of four to six amino acids, for sequences lacking proline residues), the charge asymmetry of each i segment was calculated as:

$$\sigma_i = \frac{(f_+ - f_-)_i^2}{(f_+ + f_-)_i} \quad (1)$$

The squared deviation of asymmetry was obtained as:

$$\delta = \frac{\sum_{i=1}^{N_{blob}} (\sigma_i - \sigma)^2}{N_{blob}} \quad (2)$$

Finally, κ was defined as the ratio between δ and the maximal value for a given amino acid composition δ_{max} ($\kappa = \delta/\delta_{max}$). The minimum value of κ is 0, obtained when opposite charges alternate one by one. The maximum value of κ is 1, accessible to polyampholytes entirely composed of charged residues, when opposite charges are segregated into two clusters. In the seminal Pappu's work, 30 synthetic variants of a neutral 50-mer (NCPR = 0) were designed to share an identical amino acid composition—$(Glu\text{-}Lys)_{25}$—but different κ values ($0 \leq \kappa \leq 1$), by permutations of oppositely charged residues [25]. The so-called (Glu-Lys)25 system of sequences was analyzed using all-atom MC simulations applied to the

Flory's random-coil model [42,43], showing that their ensemble-averaged Rg values inversely correlate with κ [25,32]. Overall, uniformly distributed charges (κ = 0) cause expanded conformations, whereas maximal segregation of oppositely charged residues (κ = 1) results in more compact structures. Calculation of the κ value can be performed for any protein sequence through the web server CIDER (Classification of Intrinsically Disordered Ensemble Regions) (http://pappulab.wustl.edu/CIDER/), developed by Pappu's lab [11].

Ghosh's work, instead, tackled the charge decoration issue from a more analytical perspective, introducing a general formalism to describe heteropolymer configurational properties, in the light of sequence specificity [44]. Following the coarse-grained approach introduced by Muthukumar [45], Sawle and Ghosh described pairwise, intra-chain, and short- and long-range interaction forces, taking into account charge patterning by the "sequence charge decoration" parameter (*SCD*), defined as:

$$SCD = 1/N[\sum_{m=2}^{N} \sum_{n=1}^{m-1} q_m q_n (m-n)^{1/2}], \qquad (3)$$

where m and n are the sequence positions within a N-mer chain, and q_m and q_n are the residue charges at those coordinates.

Similarly to κ, the *SCD* value tends to 0 in polypeptide sequences with uniformly distributed opposite charges. Unlike κ, the absolute value of *SCD* increases not only with charge segregation but also with polymer size and is ≤ 0 (*SCD* = 0 for perfectly alternated positive and negative charges). The correlation between κ and *SCD*, assessed on the $(Glu\text{-}Lys)_{25}$ system, is linear, with R2 = 0.95 (Figure 2). Plotting the simulated R_g values [25] against κ or *SCD* suggests that the correspondence between R_g and *SCD* is even more effective than between R_g and κ, although there is no relationship between Sawle and Ghosh's formalism and the MC simulations employed for R_g computation. The better performance of *SCD* could depend on the fact that this parameter takes into account all pairwise

interactions, regardless of the residue position, while the κ parameter is computed by averaging over stretches of few consecutive charges (blobs) [46]. Nevertheless, Ghosh's formalism is not well suited to describe collapsed globules (e.g., the model overestimates Rg), possibly due to the employed value of the dielectric constant or to neglecting hydrogen-bonding and ionization equilibria in the model [44,46].

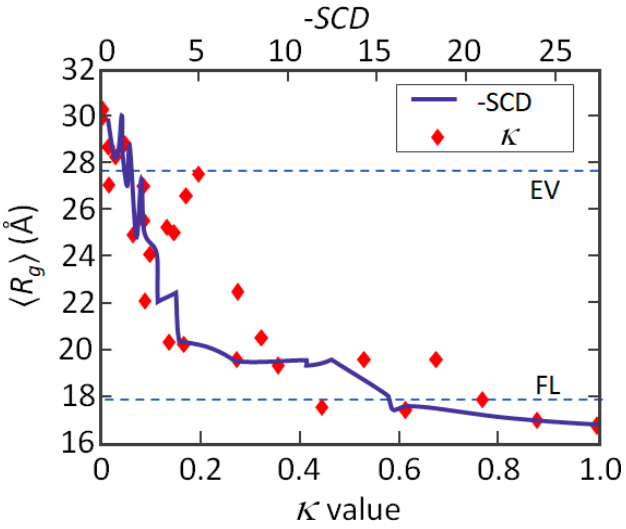


Figure 2. Parametrization of charge patterning. Ensemble-averaged radii of gyration <Rg> for sequence variants of the (Glu-Lys)25 system versus κ [25] and *SCD* [44]. The two dashed lines intersect the ordinate at <Rg> values expected for the sequences of the (Glu-Lys)25 system, modeled as excluded volume (EV)-limit polymers or as Flory random coils (Flory limit, FL), respectively (figure adapted from [46]).

Ghosh's model has been implemented recently, to account for collapsed globules and coil–globule transitions [33]. Herein, the mathematical formalism relies on minimization of Firman and Ghosh's free energy (βF), which allows chain conformational properties to be inferred. Moreover, Ghosh's analytical model was applied to predict the size distribution from the whole DisProt database [47–49], revealing significant size differences, even among IDPs with similar FCR values

(Firman and Ghosh, 2018). This result is illustrated by a phase diagram, showing the average normalized protein size in the $f_+ - f_-$ space (Figure 1d). Here, each bin of the heat map corresponds to an (f_+, f_-) class, for which the average value of the chain expansion parameter x at 300 K is given. The chain expansion parameter is defined as:

$$x = R^2_{ee} / R^2_{ee,\,frc} \qquad (4)$$

where R_{ee} is the end-to-end distance of the protein of interest and $R_{ee,frc}$ is that in the Flory random coil limit (in the absence of any interaction) of the same length. Higher values correspond to darker colors of the scale. The observation that proteins with similar values of f_+ and f_- are predicted to have different degrees of compactness further confirms that FCR is not per se sufficient to predict chain compactness. Charge decoration, instead, as captured by Ghosh's model, seems to account for the main sequence determinants of chain conformation.

Moreover, the same analytical model predicts different responses to salt concentration for proteins with similar FCR, highlighting once again the role of charge patterning as a determinant of the conformational behavior of polypeptides [50] (see also below in the next section).

To conclude, according to Pappu's and Ghosh's models, charge patterning effectively captures sequence specificity among polypeptide chains that are identical in terms of length, composition, and net charge. In this context, the overall number of charged residues (*i.e.*, FCR) and their pattern seems to act synergistically, and independently of NCPR, to determine conformational properties of polyampholytic IDPs.

3. Relevance of Electrostatic Charges in Compaction/Expansion

Charge density and distribution deeply affect conformational states and their transitions, being modulated by pH and salt, as experimentally demonstrated

[29,51,52]. The effects of salt in weakening both attractive and repulsive interactions had been well-established previously from a theoretical point of view. According to Debye-Hückel's theory of charge screening and Higgs and Joanny's polyampholyte theory [20,53,54], salt addition is expected to produce either conformational expansion within polyampholytes (in which attractive forces are prevalent) or increased compactness within polyelectrolytes (in which repulsive forces prevail). This section provides an overview of experimental investigations depicting the role of charge density and charge patterning on IDPs compactness.

3.1. Effects of Charges and Their Screening on Collapse/Expansion Transitions

Charge-mediated conformational transitions rely on both the effective ionization state and solvent exposure of charged residues. The contribution of electrostatics to compactness has been evaluated performing experiments at increasing salt concentrations by single-molecule Förster resonance energy transfer (smFRET), allowing the measurement of molecular distances in the range of 1–10 nm between fluorescence tags in individual proteins. Seminal studies on IDPs have explored the response of polyelectrolytes, such as the C-terminus of ProTα (ProTα, residues 52–111; FCR = 0.700, |NCPR| = 0.533) and the N-terminal domain of HIV-1 integrase (IN, residues 1–56; FCR = 0.267, |NCPR| = 0.067), to 1 M KCl, resulting, respectively, in a 30% and 10% reduction of R_g compared to the absence of salt [51]. The compaction effect, overall ascribable to the attenuation of electrostatic repulsions, depends on the net charge, being more remarkable for stronger polyelectrolytes, according to polyelectrolyte theory [50]. A subsequent systematic study on the N-terminus (residues 1–90) of the *Saccharomyces cerevisiae* cyclin-dependent kinase inhibitor Sic1 (hereafter called "Sic1") has provided insights into this phenomenon, also suggesting its complexity [55]. Sic1 contains 11 positively charged residue (FCR = 0.122, i.e., 12% of charged residues) and is a weak polyelectrolyte. Figure 3a shows the results of its salt titration monitored by smFRET, with increasing KCl concentrations progressively reducing Sic1 size.

Although monotonic, the composite trend of the experimental curve hints to the contribution of several phenomena, which possibly include the different accessibility to the solvent of charged residues and their different response to salt, and the influence of hydrophobic interactions, which prevail upon charge neutralization. Noteworthy, Sic1 undergoes an overall 40% reduction of its R_g in the presence of 1 M KCl. Sic1 compaction is greater than that observed for ProTα (−30%), in spite of a lower |NCPR| (0.122) than that of ProTα (0.533). How to explain the unexpectedly marked compaction of Sic1? Sic1 is a uniformly charged polyelectrolyte (FCR = |NCPR|), while ProTα, as well as IN, are "partial polyampholytes", as indicated by the non-null difference between their FCR and |NCPR| values. Thus, the strong compaction effect elicited by salt in Sic1 can be explained by repulsion screening and a lack of swelling effects, which likely occurs in ProTα and IN because of the presence of annealed charges. Thus, NCPR and FCR cannot individually explain salt dependence, which is better rationalized by taking into account the balance between attractive and repulsive forces.

When attractive forces are predominant, it is foreseeable that salt induces conformational swelling.

Clear examples of this behavior are offered by IDRs belonging to human Myc (residues 353–434), MAX (Myc-associated factor X, residues 22–102), MAD (Mitotic spindle assembly checkpoint protein MAD1, residues 55–136), MLX (Max-like protein X, residues 128–215), and MONDOA (MLX-interacting protein, residues 718–797) [56]. Such highly charged polypeptides (0.3 < FRC < 0.4) behave more markedly as polyampholytes (0.05 < |NCPR| < 0.11), with "annealed" charges conferring compact conformation in the absence of salt. When exposed to low salt concentrations (up to 0.6 M KCl), the screening of attractive interactions causes Rg expansion (Figure 3b) [56]. Above 0.6 M, a chain re-collapse is observed, probably due to hydrophobic interactions, which prevail upon charge screening. Noteworthy, different salts, *e.g.*, LiCl, NaCl, and CsCl, cause

compaction to different extents. This salt specificity is reminiscent of the variable salting-out effect along the Hofmeister series and led to the hypothesis that similar factors come into play [56].

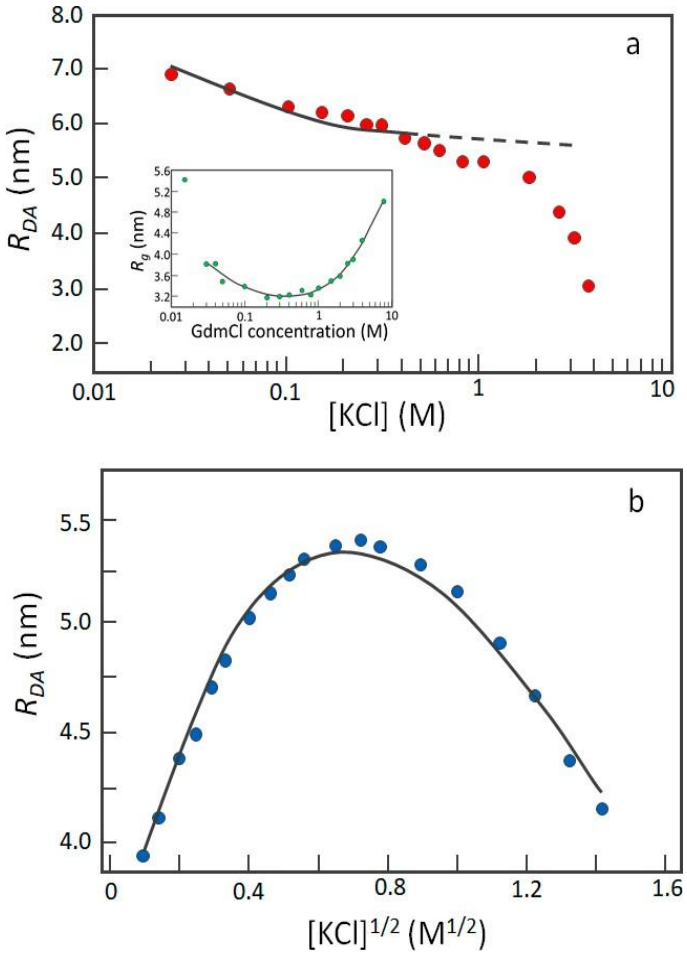


Figure 3. Dependence of IDP compaction on salt concentration. The end-to-end distance obtained from single-molecule Förster resonance energy transfer (smFRET) is expressed as donor-acceptor distances (R_{DA}) *vs.* KCl concentration. (a) The uniformly charged polyelectrolyte Sic1 undergoes collapse, due to shielding of unbalanced electrostatic charges and attenuation of their repulsive forces. The black solid line represents the fitting by the model described in [55]. At higher KCl concentrations, hydrophobic interactions are likely to overlap with the

charge screening effect. The inset shows the behavior of Sic1 R_g exposed to GdmCl. Here, non-chaotropic concentrations (<1 M) cause protein collapse, while higher denaturing concentrations lead to conformational swelling [55], conferring a non-monotonic behavior significantly different from the main plot of panel (a). Figures adapted from [55]. (b) The polyampholyte Myc undergoes expansion due to the weakening of attractive electrostatic forces at a low KCl concentration (<0.6 M). Here the R_{DA} is plotted versus [KCl] 2 to allow fitting of the experimental results by the polyampholyte theory [56]. The dependence of size is modelled on the root square of the ionic strength, which corresponds to the root square of the concentration for a completely dissociated mono-ionic salt. At higher concentrations, the chain collapses again, probably due to the prevalence of hydrophobic interactions over charge shielding. The black solid line represents the fitting by the model described in [56]. Figure adapted from [56].

For the sake of completeness, it is necessary to mention that several studies on polyelectrolytic IDPs have employed the denaturing salt guanidinium chloride (GdmCl), although its behavior likely reflects the overlapping effects of charge screening, preferential solvation, and chaotropic effects, especially at high concentrations [57].

To conclude, NCPR and FCR are useful parameters to roughly predict whether an IDP behaves as a polyelectrolyte or a polyampholyte in its response to salts, within the general frame provided by Debye-Hückel's theory, and the polyelectrolyte and the polyampholyte theories. More recently, it has been proposed that chain expansion or compaction induced by salt depends also on charge patterning [50]. An experimental assessment of this theory is given by the different salt sensitivity of protein permutants obtained by simply varying the position of charge residues ("κ variants") [58] (see Section 3.3). Understanding how environmental conditions affect IDP compactness will contribute to rationalize their function in the cellular

context. This knowledge can be exploited also to better control the performance of IDP-based biomaterials and devices [50,59].

3.2. Exploitation of Charge Patterning in Stimuli-Sensitive Biopolymers

In the field of material sciences, linear polymers designed for the fabrication of solid surfaces are referred to as "polymer brushes" and have been exploited, for instance, to confer anti-biofouling and anti-frictional properties [60]. IDP-inspired polyampholyte brushes reversibly undergo expansion/collapse transitions in response to external stimuli (i.e., pH, ionic strength, temperature), consistently with their FCR and NCPR [61,62]. IDP brushes profit from a large repertoire of building blocks (i.e., canonical and unnatural amino acids) and, as typical for proteins, offer multiple hierarchical levels of structural organization dependent on their primary structure and post-translational modifications [63]. For these reasons, synthetic IDPs combine the advantages of synthetic polymers and polypeptides.

A fine example of stimuli-sensitive protein brush has been developed by Kumar's group [59]. A recombinant IDR, rNFH-SA, derived from the heavy subunit of the rat neurofilament complex [64], was grafted in an oriented manner onto a quartz support to functionalize its surface. rNFH-SA is a highly charged polyampholyte, as inferable from its FCR (0.429) and |NCPR| (0.014) values. In addition, our analysis on charge patterning ($\kappa = 0.074$) suggests that rNFH-SA may display an expanded conformation. This protein behaves as a polymer brush capable of swelling and collapsing in response to changes in solution pH and ionic strength, in a rather wide dynamic range, not yet fully explained in the light of the theory illustrated in the previous paragraph [59]. Overall, rNFH-SA behaves as qualitatively expected for weak polyelectrolytes, which collapse with increasing salt concentration. A deeper knowledge of polyampholyte electrostatics (charge density and patterning) and of polymer physics could further help in developing

"smart biomaterials" with desired properties and capable of complementing the array of already available organic/synthetic polymers.

3.3. Effects of Linear Charge Patterning over Protein Compaction

Theoretical studies on charge-decoration and its impact on polyampholyte conformation have been supported by experimental investigations. A plethora of orthogonal biophysical techniques have been employed to explore this dependence. To cite an instance, Tedeschi and collaborators carried out a systematic comparison between three κ-variants for two ~100-residue viral IDPs, merging evidence from SAXS, size-exclusion chromatography (SEC), and limited proteolysis [65]. For each protein, which displays similar values of FCR (~0.3), |NCPR| (<0.05), and κ (~0.2), the authors designed two sequence permutants, shuffling the charged residue positions in order to achieve the highest and lowest possible κ values (average values over the different proteins: κ_{min}~0.08 and κ_{max}~0.42) compatible with their natural amino acid composition while keeping the location of non-polar residues unchanged. By doing so, the conformational variability of the variants, relative to the wild-type form, could only be imputable to charge-patterning changes. A direct correlation between κ and protein size was observed. In addition, the study provided hints suggesting that differences in protein responsiveness to charge clustering also reflect differences in proline content (which is indeed different in the two model IDRs considered in that study). In particular, proline residues seem to counteract the compaction effect exerted by charge segregation.

A similar approach, yet enriched in further insights into biological implications, was used by Kriwacki and co-workers and applied to the C-terminal domain of the human cell-cycle inhibitory protein $p27^{Kip1}$ (residues 96–198, hereafter called "p27"), integrating computational simulations and biophysical techniques [66]. The authors kept the primary p27 phosphorylation site (Thr187) unmodified, altering the charge distribution around it, to lower (lowest κ value = 0.14) or increase

(highest κ value = 0.78) the κ value relative to the wild-type protein (0.31). Thus, besides the expected κ–Rg inverse correlation, assessed by atomistic simulations and in-bulk conventional techniques (SAXS), the authors could also document differences in phosphorylation efficiency that could be ascribed to sequence-encoded features. Indeed, the efficiency of Thr187 phosphorylation increases with the "local" NCPR of the so-called auxiliary motifs (residues 100–180) [66], highlighting the relevance of linear charge patterns in supporting (or contrasting) a primary physiological function.

An even finer, yet consistent, characterization of p27 sequence permutants (κ values of 0.14 and 0.56) was performed by Barran's group, exploiting native mass spectrometry coupled to ion mobility, a valuable technique to interrogate IDP/IDP ensemble conformational heterogeneity [67]. Collisional cross-sections of the permutants proved that charge patterning dramatically affects IDP/IDR compactness, with the high-κ variant displaying a lower conformational heterogeneity, compared to the wild-type and low-κ variant [58]. The latter turned out to be insensitive to increasing salt concentrations, whereas the high-κ variant displayed a conformational expansion at high ionic strength [58].

Other insights into the functional relevance of charge patterning have been obtained for RAM (RBP-Jk-associated-molecule) region, a 111-residue IDR belonging to the intracellular domain of the Notch receptor (NICD) and involved in a transmembrane cell-to-cell pathway controlling cellular differentiation and stem-cell fate [68]. The limited dispersion of κ values among distantly related RAM orthologues led to lay the hypothesis that its charge patterning responds to a functional requirement, i.e., mediating its binding affinity for CSL, an element of the tertiary complex (NICD-CSL-MAML) involved in Notch activation. Among RAM charge permutants, it was observed that R_g and an affinity for CSL decrease with increasing charge segregation, causing a significant loss in Notch transcriptional activation. Thus, experimental and computational data consistently

suggest that charge decoration influences the conformational preferences of IDPs and can be considered as an evolutionary-selected trait of crucial importance for their functions.

4. Relevance of Electrostatic Charges in Protein Solubility/Aggregation and Fibrillation

Protein solubility corresponds to the ability of a polypeptide chain to dissolve into a solution, notably aqueous, and is governed by the competition among chain–solvent, inter-chain, and chain–chain interactions. Such labile equilibrium is severely impacted by solvent, ionic strength, temperature, and pH. Typically, at pH values higher or lower than the protein pI, protein–protein interactions are disfavored, in favor of chain–solvent ones, therefore increasing its solubility. Recent studies suggest that the correlation between protein pI and the pH of their (sub)cellular environment is simply a neutral "by-product" of the main adaptive selection aimed, instead, at favoring structural metastability [69]. In analogy with this original view, it could be hypothesized that protein solubility is the result of a trade-off between metastability and biological activity.

The issue of protein solubility has been widely addressed by physical statistics, considering the polymer–solvent interaction parameter χ [70,71] and polymer–solvent interaction energy. Briefly, χ can be considered as a measure of solvation enthalpy, namely the enthalpy associated with transferring the polymer from the gas phase into water. Therefore, a distinction can be drawn between "good" ($\chi < 0$) and "poor" ($\chi > 0$) solvents, regarding the ability to solvate a given polymer. Polymers expand and dissolve in a good solvent, while they collapse in a poor one. Referring to polypeptide chains, the propensity to be solvated depends on both backbone and side chains. Although in aqueous media the protein backbone is prone to collapse, it is the interplay among the sidechains, backbone, and solvent that decides the actual solvation fate of a protein and could support or reverse the

intrinsic backbone-compaction propensity [37]. Thus, sequence specificity could account for the divalent nature of water, being a poor solvent for globular folded proteins and a good one for IDPs [37].

A plethora of predictive programs have been developed to infer aggregation propensity from the primary structure [72,73]. The next section will focus on the role of electrostatic charges in determining IDP solubility and aggregation properties.

4.1. Effects of Charge Density on Protein Solubility/Aggregation

The effect of electrostatic charges on protein solubility is controversial. Modulating protein conformation and solvation through the manipulation of pH-sensitive groups represents an exciting challenge [74–76], limited in practice by the difficulty of producing well-folded charge variants of globular proteins. The high designability of IDPs [77], herein meant as the number of sequences encoding conformational ensembles of similar compaction properties, is exploitable to conceive synthetic solubility-enhancing tags [78]. Solubility-enhancing tags can promote solubilization through a dual mechanism: (*i*) By increasing the relative proportion of solubility-enhancing amino acids with respect to the overall residue composition of the fusion construct [79]; and (*ii*) by acting as "entropic bristles" (EBs) through random movements around their point of attachment. EBs entropically exclude the contact with large particles, *i.e.*, other proteins/peptides, thus reducing the probability of the fusion protein to undergo aggregation, without excluding small molecules, such as water, salts, metals, or cofactors, which in fact increase solubility [80].

A first effort to prove IDPs as effective solubility tags was performed by Santner and collaborators, who compared the solubilization performances of four synthetic intrinsically disordered tags to those of several well-established folded tags [81]. In this pivotal work, polyelectrolytes of different lengths (60, 144, 250 residues) and

net charge (−24, −25, −41, and −65) yet similar pI (from 2.5 to 3) were tested. Interestingly, the chain length turned out to be more crucial than the sequence composition, with larger proteins being more effective EBs [81]. It should be emphasized that the proteins selected in this study, although presenting diverse net charges, have a seemingly high charge density, i.e., NCPR values (~−0.40; −0.29; −0.26). Therefore, the chain length, rather than the charge density, represents the most significantly diverse parameter among the analyzed proteins, without ruling out any contribution of charge density itself.

This issue was more directly addressed through a set of synthetic IDPs derived from the N-terminus moiety of measles virus phosphoprotein (PNT, 230 residues) [82], whose sequence naturally possesses an acidic pI (4.88) and an almost balanced set of oppositely charged residues, thus resulting in an NCPR ~0 (−0.071). PNT synthetic variants have the same length, FCR (0.257 ± 0.004), and hydropathy score (3.826 ± 0.067), yet they display different NCPR values (from −0.248 to +0.216) and therefore different pIs (from 3.37–9.61) [83].

As expected, each synthetic protein experiences a solubility loss at its pI. Furthermore, the "aggregation intensity", namely the proportion of insoluble protein, turned out to be directly correlated to |NCPR|, with low-NCPR proteins remaining mostly soluble and almost aggregation free, independently of pH. PNT variants more responsive to pH are able to "transmit" their aggregation propensity to resilient proteins, such as green fluorescent protein, embedded in the same chimeric construct [83]. Overall, these observations may contribute to understanding the behavior of IDPs in response to events affecting protein NCPR (i.e., post-translational modifications, mutations, environmental changes). For instance, it could be argued that high-NCPR proteins (i.e., polyelectrolytic IDPs) are much more sensitive than polyampholytes to even slight pH changes. Results from [83] have led to the development of an empirical equation suitable to predict pH-dependent aggregation of amyloidogenic IDPs and, hence, to promote the

design of synthetic solubility/aggregation tags, as well as reversibly aggregating nanofibrillar materials [84].

4.2. Relevance of Electrostatic Charges on Fibrillation

Amyloid fibrils have been associated with important biological functions [85,86] and a plethora of pathologies, including socially relevant neurodegenerative diseases, such as Alzheimer's and Parkinson's diseases [87]. Aggregation can be triggered in proteins, as the result of either "ordering" of disordered regions or "disordering" of well-folded structures [88]. For a more detailed description of fibril structure, see Appendix B. The research on the "amylome", the ensemble of amyloidogenic proteins in a proteome, has indicated that most proteins can form amyloid fibrils in vivo [89], with the involvement of different sequence features. Most frequently, canonical amyloid fibrils are encoded by stretches of 5–15 residues of aliphatic and aromatic amino acids (Riek, 2018).

Nevertheless, the aggregation of prions and prion-like domains is apparently independent of aliphatic residues and influenced by pH [90–92]. In this subclass of amyloids, which are self-perpetuating and infectious, sequence determinants have been elusive for years, since cryptically encoded by rather long (at least 60 residues) and disordered sequences, containing a few amino acid types (i.e., low-complexity regions) rich of Tyr, Gly, and polar residues (mostly Gln and Asn) [90,93–95], which argue for the involvement of hydrogen bonds, van der Waals, and π–π stacking interactions [96–98]. Short linear sequence motifs rich in polar residues (*e.g.*, SYSGYS from human FUS protein, or GNNQQNY from yeast Sup35) have also been recognized as "nucleation centers" or "soft-amyloid cores", effectively promoting both spontaneous and seeded aggregation in proteins [99,100], and natural and synthetic peptides [101–103].

The structures of human prion protein (PrP) fibrils, recently solved by Cryo-electron microscopy (Cryo-EM), have clarified the role of hydrophobic, polar, and

charged residues [104,105]. The fibrils obtained from residues 23–231 of human PrP display a hydrophobic and compact core stabilized by an intramolecular disulfide bond (between Cys179 and Cys214), while a mostly hydrophilic surface is exposed to the exterior. Indeed, each fibril consists of two protofibrils intertwined in a left-handed helix, with Lys194 and Glu196 from opposing subunits forming salt bridges and creating a hydrophilic cavity at the interface of the two protofibrils [105]. This structure is coherent with previous data suggesting the critical role of pH in promoting prion fibrillization [92].

As concerning canonical amyloid fibrils, polar and charged residues are frequently arranged in disordered protrusions at the fibril edges of the hydrophobic fibril core, being functional to elongation through transient interactions with incoming monomers [106]. Such flexible charged structures have been detected, for instance, in the aggregation products of α-synuclein, Aß peptide, and tau protein [106]. Nonetheless, electrostatic interactions can also play a direct role stabilizing the cross-β-sheet structure. In this respect, we summarize here available data on the impact of charge decoration and electrostatic networks in α-synuclein (α-Syn) fibrillation, generally regarded as the pathological hallmark of Parkinson's disease [107,108] and other synucleinopathies [109,110]. Charge density and, *ante litteram*, charge segregation of α-Syn have been deeply investigated for their effects on fibrillation. α-Syn is an IDP composed of 140 residues (FCR = 0.279; NCPR = −0.064; κ = 0.172) containing an N-terminal domain with a highly conserved α-helical-prone lipid-binding region (residues 1–60), a central hydrophobic region (NAC) essential for α-Syn aggregation (residues 61–95). and a C-terminal tail (residues 96–140), acting as an interaction hub for several proteins [111–115] (Figure 4a). The NCPR profile calculated by CIDER (Figure 4b) suggests a sharper distinction between an amphipathic and amphoteric N-terminal moiety (residues 1–102), in which opposite charges alternate almost regularly (κ = 0.082; FCR = 0.311), and a highly acidic C-terminus (residues 103–140) with

highly clustered negative residues (FCR = −0.368). In vitro fibrillation of α-Syn is markedly pH dependent and occurs at higher rates at acidic pH [116,117]. This acidification effect can be explained considering the α-Syn domain structure. Indeed, acidification increases the net charge of the N-terminal moiety (from +6 at neutral pH to +17 at pH 3) and neutralizes the negative charge of the C-terminal domain (from −15 at neutral pH to 0 at pH 3). Albeit this transition involves a similar number of charged residues in each protein moiety (+17 at the N-terminus and −15 at the C-terminus), the N-terminal region at an acidic pH displays a markedly lower charge density (local $NCPR_{1–102}$ = +0.17) than the C-terminal domain at neutral pH (local $NCPR_{103–140}$ = −0.39). Such a reduction in charge density weakens intra- and inter-molecular electrostatic repulsions, as well as solvent interactions, and ultimately enhances hydrophobic interactions responsible for fibrillation [116,117].

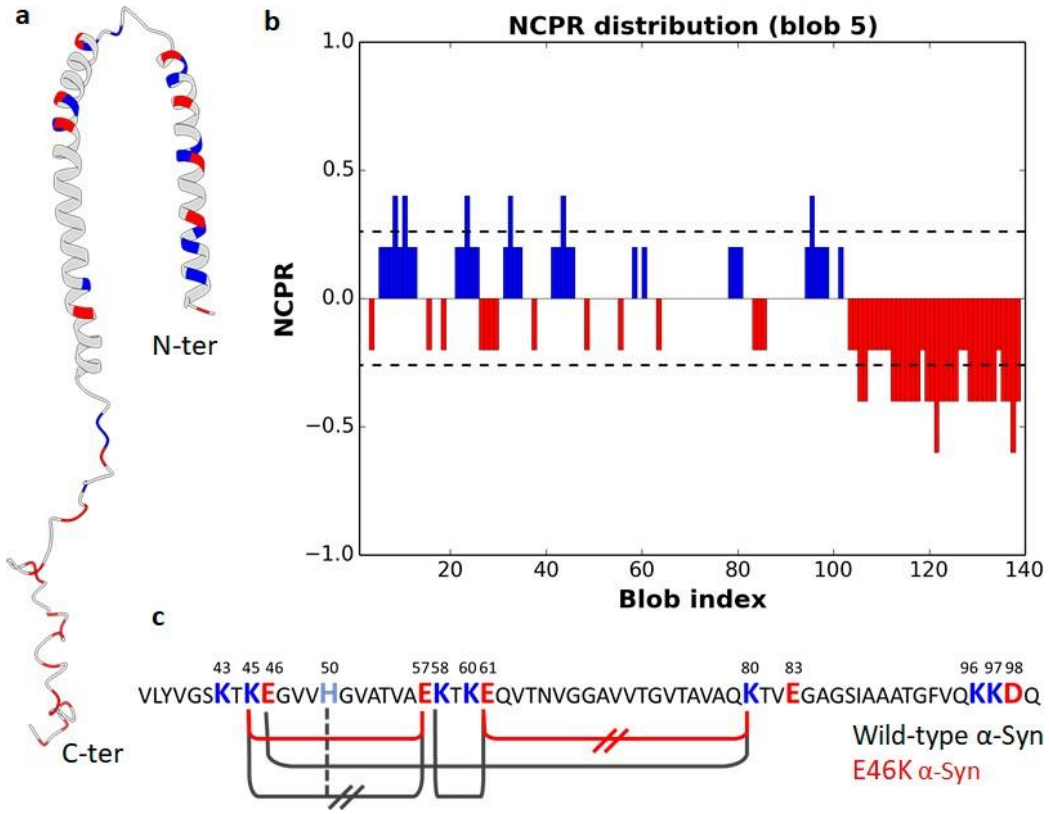


Figure 4. Charge density and distribution of α-Syn. (**a**) Cartoon representation of an NMR structure of micelle-bound human α-Syn (PDB 1XQ8; [118]). Image created with UCSF Chimera [119]. (**b**) NCPR profile along the linear sequence of α-Syn. The blue and red peaks denote positive and negative charges, respectively (plot obtained by CIDER, [11]). (**c**) Primary sequence of the wild-type α-Syn fibril core (aa 37–99). Charged residues are shown in colored bold letters and those interacting in the structures of acetylated wild-type (PDB 6A6B; [120]) and E46K (PDB 6L4S; [121]) α-Syn fibrils are connected by black and red solid lines, respectively. Intermolecular interactions are marked by transverse parallel lines. Figure adapted from [121].

The role of the C-terminal tail in pH-driven fibrillation of α-Syn has been further supported by experimental [122] and computational studies [123]. Data reported so

far depict the C-terminal domain as an effective EB. As long as it is highly charged, expanded. and capable of large conformational fluctuations, it acts as an intramolecular chaperone, counteracting aggregation. The N-terminal region, instead, populates an ensemble of unfolded conformations with some intrinsic helical propensity, in equilibrium with membrane-bound, monomeric and multimeric, and helical structures [118,124,125] (Figure 4a). Cryo-EM studies indicate that the fibril core of α-Syn consists of residues 37–99, while the N- and C-terminus remain flexible and not resolved. The structural detail of fibrils indicates the key role of a network of electrostatic interactions involving intramolecular (i.e., E46-K80, K58-E61) and intermolecular salt bridges (K45-E57) (Figure 4c) [120]. The analysis of the E46K variant, associated with a severe form of familial Parkinson's disease, highlights the importance of electrostatic interactions in defining the fibril morphology, as well. Indeed, with respect to the wild type, the E46K variant reshapes the above-mentioned electrostatic network, and forms a smaller fibril core (residues 45–99) and a distinct fold. This is of utmost relevance with regard to the pathogenic mechanism, as E46K fibrils are less resistant to proteases and mechanical stress and, therefore, more prone to propagation [121].

As expected from a role of electrostatic interactions in fibrillation, salts affect the aggregation kinetics of amyloid proteins. However, they do that in a highly complex way. A systematic analysis of the effect of salts on protein aggregation kinetics has been performed by testing ions from the Hofmeister series in real-time quaking-induced conversion assays [126]. This study reveals different effects of ions, depending on their position in the Hofmeister series, in line with a crucial role of protein hydration in fibril formation [127]. The effect is more remarkable for anions than cations and dependent on the biological matrix. Furthermore, the dependence of aggregation kinetics on Hofmeister ions is highly protein specific. The PrP and tau have opposite trends along the series, while α-Syn displays a

bimodal response, with enhanced kinetics at both ends of the series. These results are in line with a complex interplay of electrostatic, hydrophobic, and hydration effects governing protein fibrillation and with the involvement of specific residues of different nature in these proteins.

5. Relevance of Charge Decoration in Phase Separation

Spatio-temporal control of intracellular reactions is based on a finely regulated molecular trafficking through cellular compartments. Besides membrane-limited structures, membrane-less organelles [128–130] contribute to compartmentalization, separating molecules by liquid demixing and confining them in droplets at higher local concentrations than the surrounding matrix [131,132], thereby creating dynamic proteinaceous microreactors [133] (see Appendix C). LLPS can occur through heterotypic (i.e., interaction and de-mixing of two or more polymers) or homotypic coacervation (i.e., single-polymer self-association) [134,135]. Coacervation usually occurs at concentrations and temperatures thermodynamically favoring self-interaction of polymers. Biomolecular condensates are generally enriched with multivalent molecules prone to establishing multiple intra-chain and interchain interactions [128]. Therefore, proteins with a modular architecture, encompassing low-complexity regions and/or repeats of short-linear motifs, are particularly well suited to nucleate coacervation. It is therefore not surprising that IDPs/IDRs display a strong propensity to undergo LLPS.

Different types of non-covalent interactions have been implicated as triggering factors: Hydrogen bonding, cation–π contacts, electrostatic and dipolar attractions, and π–π interactions between aromatic rings [128,136] (see also Appendix C). Our understanding of sequence determinants of phase behavior is in its infancy and we have just started learning its “grammar” [136,137]. Aromatic residues (Phe, Tyr), along with charged ones (Arg in particular), have often been shown to be key

determinants of in vivo and in vitro LLPS occurring through short-range attractive forces due to π–π or cation–π interactions [134,136,138–142]. Fewer studies have investigated long-range electrostatic interactions between charged residues and their patterning on LLPS [46,139,142], although they might have a prominent role, due to the compositional features of IDPs [143]. To assess the role of electrostatic interactions, experiments are typically carried out in the presence of varying concentrations of salt, most often represented by NaCl. A seminal work on the N-terminal domain of Ddx4 (Ddx4N), a primary constituent of human germ granules, has unveiled that its coacervation is dominated by π–cation interactions involving Phe and Arg residues. The LLPS of Ddx4N is hence extremely sensitive to ionic force, as highlighted through experiments at increasing concentrations of NaCl [134].

Two other illustrative examples are provided by hnRNPDL and NPM1, which are involved in a form of muscular dystrophy and in the spatial organization of the nucleolus, respectively. HnRNPDL (and its isoforms) was found to only form liquid-like droplets at low ionic strength [144], and homotypic LLPS of NPM1 is similarly impaired at high NaCl concentrations [145]. Overall, these behaviors are consistent with the polyampholyte theory and with the hypothesis that LLPS is driven by the annealing of opposite-sign charges or by cation–π interactions.

In contrast with the three examples above, the N-terminal prion-like domain of galectin-3 undergoes LLPS only when the NaCl concentration is increased above 600 mM, with LLPS being driven by π–π interactions between aromatic residues [146].

Examples also exist where salt, namely NaCl, does not exert any significant impact on LLPS, suggesting that the formation of coacervates relies on hydrophobic interactions. An illustrative example is provided by PNT3, a viral protein region that undergoes LLPS with concomitant formation of amyloid-like fibrils [147].

Although this protein is classified as a strong polyampholyte (FCR = 0.364, |NCPR| = 0.164), NaCl does not affect its ability to form aggregates (at least up to 300 mM). In support for the hypothesis of the involvement of π–π interactions, the most amyloidogenic region contains three contiguous tyrosine residues whose replacement with alanine residues abrogates fibrillation [147]. Finally, we can cite the case of γ44-gliadin, a wheat storage protein with an intrinsically disordered domain that undergoes LLPS in a salt-dependent manner in spite of its very weakly charged nature (FCR = 0.04 and NCPR = 0) [148]. In particular, increasing NaCl concentrations were found to lead to a drastic decrease in the number of droplets and to an increase of γ44-gliadin saturation concentrations, corresponding to equilibrium concentrations above which phase separation occurs in in vitro experiments. This behavior argues for a contribution of electrostatic interactions in the formation of γ44-gliadin liquid-like droplets. Since the few charges are mainly located in the C-terminal domain, the authors proposed that this unequal charge distribution along the sequence likely promotes directional interactions: Aromatic residues of the N-terminal domains and positively charged residues of the C-terminal ones could participate in LLPS by π–cation interactions, as already reported [134,149,150].

The relentless tug of war between electrostatics and the hydrophobic effect is also a determinant for the link between aggregation and LLPS. To cite an instance, a connection between aggregation and LLPS of tau protein has for a long time been suggested [151–153] and recently ruled out [154]. For tau protein, LLPS is driven by complex coacervation mechanisms, dominated by long-range multivalent electrostatic attractions [155]. Instead, the amyloidogenic pathway seems driven by hydrophobic interactions [154]. In line with the behavior of tau, in the case of hnRNPDL, LLPS exerts a protective role against fibril formation [144]. On the contrary, when LLPS is triggered in vitro under high-salt concentrations, thereby becoming partially driven by hydrophobic contacts, a direct correlation with

increased amyloid propensity is observed, suggesting that this two phenomena not only could occur under coinciding conditions but could also positively influence one another [154].

Although we can try to rationalize these complex responses, we are still far from acquiring the ability to predict them. The emerging scenario from the available literature data is that the effect of salt on LLPS is poorly predictable and highly protein dependent, analogously to its impact on protein fibrillation (see Section 4). This lack of a clear trend may depend on the double role of charged residues in the so-called architecture of "stickers" and "spacers" [141,156,157]. Stickers are protein motifs or domains reversibly interacting with other protein molecules or nucleic acids, thus creating coacervate networks. Spacers are located between stickers and preferentially interact with solvent molecules, acting as a scaffold [141,156,157]. In contrast with aromatic residues, which are definitely enriched in stickers' elements, charged residues could play a role either in stickers, by electrostatic attractions, or in spacer regions, by conferring solubility and flexibility to the scaffold itself. This ambiguity can also entail the difficulty of designing electrostatically driven LLPS models with predictable and controllable behavior [142].

An even smaller number of studies deal with the role of charge distribution in LLPS. Coacervation was found to strongly depend on the segregation of opposite charges in the $Ddx4^N$ and Nephrin intracellular domain, which also a share similar value of κ (= 0.237 and 0.217, respectively) [134]. For instance, LLPS was suppressed by attenuating charge segregation in $Ddx4^N$ in a permutant with κ = 0.053 [134]. A systematic analysis of the relationship between LLPS propensity and the linear pattern of charge distribution, namely the values of κ, has been addressed by computationally predicting the phase diagrams of $(Glu\text{-}Lys)_{25}$ by a random-phase approximation approach [46]. Overall, the system shows a binodal curve with an upper critical temperature of coacervation (see Figure A2), which

correlates well with κ. Since Rg decreases with increasing κ, a power law linking $_{Rg}$ to the critical temperature (CT) holds as well [46]. Thus, the more compact the conformation, the higher the critical temperature, which is overall consistent with experimental evidence [134,158]. We propose, however, the existence of a "compaction threshold", above which the dependence of CT on κ is inverted. That is, compaction would promote phase separation up to a certain limit, beyond which highly collapsed conformations would instead disfavor inter-chain interactions. In our hypothesis, IDPs/IDRs with low-to-moderate κ values display a higher propensity to undergo LLPS, while those containing interspersed charged residues (κ~0), as well as those with highly segregated charges (κ~1), present a lower coacervation propensity (Figure 5, top panel). To test this hypothesis, IDRs from the PhaSePro database [159] undergoing electrostatically driven phase separation were analyzed herein to extract a κ-value distribution. Each frequency class was compared with that of IDRs from the entire DisProt database. As shown in Figure 5, the frequency of sequences with $0.2 < \kappa < 0.25$ in PhasePro is double compared to DisProt. By contrast, the frequency of sequences with $0.25 < \kappa < 0.3$ is less than half (Figure 5, bottom panel). These data indicate that IDRs undergoing electrostatically driven phase separation have κ values mostly comprised between 0.2 and 0.25. Such κ values, as in the case of Ddx4 [66,160] and NICD [158], are those that probably allow to better exclude the solvent and favor inter-chain interactions. Nonetheless, the still limited size of PhaSePro (28 entries of IDPs/IDRs undergoing electrostatically driven phase separation at the time of this analysis) points to the need of addressing this issue in a more systematic way in the future, so as to be able to confirm the existence of an optimal value of κ favoring phase separation and possibly draw general conclusions. Our hypothesis is in line with a recent work indicating how critically important the patterning of sticking elements is, with the implication that too many sticky elements may hamper coacervation [141].

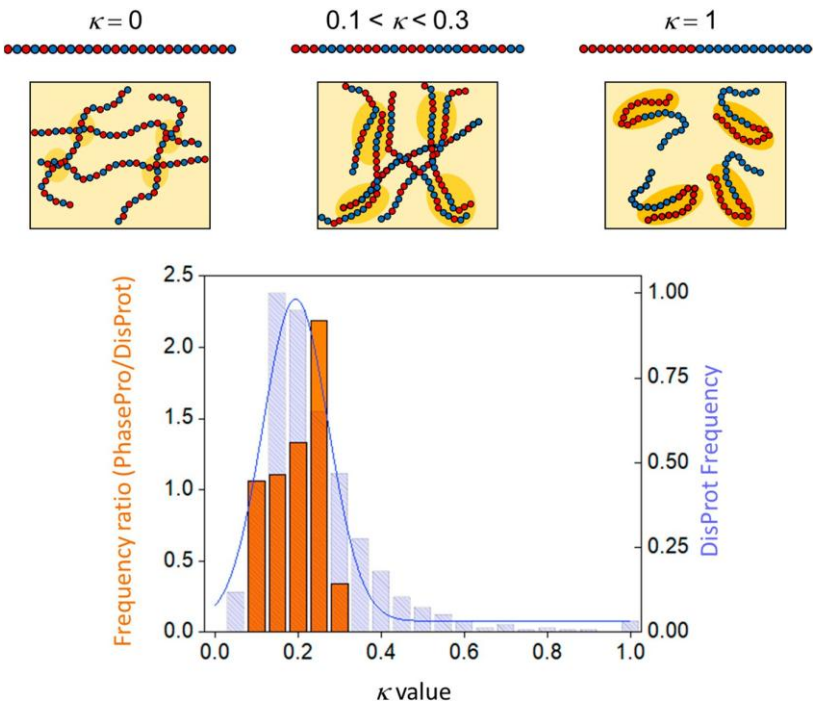


Figure 5. Charge decoration and propensity to undergo electrostatically driven phase separation. (Top) Lys-Glu sequences of different κ values, with K and E residues in red and blue, respectively. Charge distribution is related to protein compactness and phase separation (square boxes). κ = 0: attractions within and among polyampholytes lacking long same-charge clusters are weak. These chains are overall expanded and show weak scarcely cooperative inter-chain interactions, as symbolized by small ovals in pale yellow (scheme inspired by [46]). 0.1< κ < 0.3: most favorable inter-chain interactions among chains presenting blocks of segregated charges. κ = 1: complete, or almost complete, charge segregation favors strong intra-chain interactions (dark yellow areas) that, reflecting monomolecular events, efficiently out compete inter-chain attractions. (Bottom) IDP/IDR propensity to undergo electrostatically driven LLPS as a function of κ values. The orange histogram shows the ratio between PhasePro (subset of electrostatically driven phase separation) and DisProt k-class frequencies [159] (left vertical axis). The κ-class frequencies from DisProt are shown as the shadowed blue bar

histogram (right-hand vertical axis). Sequences from the PhaSePro database were manually retrieved and further analyzed for their level of disorder by IUPred [161]. Only regions with an overall disorder level higher than 0.6 (in a scale 0–1) were used to compute κ values. The latter were calculated through CIDER webserver (http://pappulab.wustl.edu/CIDER/, [11]). DisProt entries were filtered by discarding sequences shorter than 20 amino acids or devoid of charged residues (FCR = 0) and redundant sequences. Two sequences were considered redundant if they were associated to the same DisProt_ID and if the start (residues 1–20) and the end (last 20 residues) of the two compared regions are respectively comparable.

An obvious limitation in LLPS studies is that in vitro experiments cannot reproduce the complexity of the intracellular environment. Therefore, the picture needs to be further detailed, bearing in mind the peculiar features of the cellular context, in addition to simplified molecular models.

6. Conclusions

Today more than ever, knowledge concerning the role of electrostatics on the structure and function of IDPs/IDRs benefits from theoretical, experimental, and heuristic contributions from the fields of polymer physics and computational science. Tapping into such concepts has become imperative albeit challenging, as already pondered more than 20 years ago by one of the fathers of modern molecular biology [162]. Useful descriptors of IDP/IDR charge density and patterning, such as NCPR, κ, and SCD, arise from this very productive crosstalk between polymer theory, biophysics, computational simulations, and protein science, and contribute to deciphering the hidden structural code of IDPs/IDRs. The unfolded states of proteins have long been recognized as crucial models to interpret relevant biological processes, such as protein folding, membrane translocation, and stability, aided by the theoretical framework of statistical and polymer physics [163]. In this regard, IDPs/IDRs represent an interesting

experimental model, as they offer significantly populated unfolded states in the absence of denaturants. One useful aspect resides in their remarkable designability and stimuli responsiveness. The high designability of disordered proteins allows for functional remodeling and modification of entire biological networks [164,165].We have learnt that such a reshaping can be obtained by tuning simple sequence parameters, such as the fraction of positive and negative charges and their clustering. The abundance and distribution of charged residues can confer specific sensitivity to changes in the environment, i.e., pH, ionic strength and ligands [56,83]. This plasticity is likely one of the reasons of their evolutionary success in regulatory networks. An interdisciplinary approach is essential for (i) a deeper understanding of the molecular mechanisms underlying physiological and pathological events, and (ii) translating our knowledge on polymers physics into de novo design of polypeptides with the desired properties of compactness, fibrillation, and phase separation, envisaging expectedly impactful biotechnological applications. Our knowledge of polymer physics does not yet allow to fully understand and especially control these events, particularly concerning phase separation, fibrillation, and their connection. This is partly due to the fact that this research field is still in its infancy. This area of research will benefit from the growth of dedicated databases, such as AmyPro [166], CPAD [167], DrLLPS [168], or PhasePro [159], and from data analysis through advanced data-mining tools, which will also become increasingly available in the future.

Author Contributions: G.B. and S.B.: conceptualization and preparation of original draft; preparation of figures. G.B., S.B., S.L. and R.G. wrote and revised the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors thank Rita Pancsa (Institute of Enzymology, Research Centre for Natural Sciences of the Hungarian Academy of Sciences, Budapest, Hungary) for data mining of PhaSePro database and Antoine Schramm (previously in the AFMB lab in Marseille and now in the Max-Planck Campus Tübingen, Germany) for the analysis of DisProt database; Antonino Natalello

and Marco Mangiagalli (University of Milano-Bicocca, Department of Biotechnology and Biosciences, Milano, Italy) for fruitful discussion and critical reading of the manuscript. The authors gratefully acknowledge Marco Mangiagalli for his support in figure preparation.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

Cryo-EM	Cryo-electron microscopy
CT	Critical temperature
FCR	Fraction of charged residue
GdmCl	Guanidinium chloride
IDP	Intrinsically disordered protein
IDR	Intrinsically disordered region
LCP	Lower critical point
LLPS	Liquid-liquid phase separation
MC	Monte Carlo (simulation)
NCPR	Net charge per residue
NMR	Nuclear magnetic resonance (spectroscopy)
PNT	N-terminus moiety of measles virus phosphoprotein
ProTα	C-terminal domain of Protα
PrP	Human prion protein
RAM	RBP-Jk-associated-molecule (region)
R_{DA}	Donor-acceptor distance
R_g	Radius of gyration;
R_h	Hydrodynamic radius
rNFH-SA	Disordered region from the heavy subunit of the rat neurofilament complex
SAXS	Small angle X-ray scattering (spectroscopy)
SCD	Sequence charge decoration

SEC	Size-exclusion chromatography
smFRET	Single-molecule Förster resonance energy transfer
α-Syn	α-Synuclein
UCP	Upper critical point

Appendix A. Parameters of Polymer Dimension: Hydrodynamic Radius, Radius of Gyration, and Their Relationships

According to the IUPAC recommendation, the dimensions of linear flexible macromolecules are usually expressed as the ensemble averages of the end-to-end distance, S, and radius of gyration R_g [169]. Notably, a pronounced decoupling between R_g and S has been observed, especially for heteropolymers, in both globule and coil-like states, probably due to the chemical heterogeneity of interactions [37]. Along R_g and S, the R_h can also return a reasonably adequate representation of IDP average dimensions and coil-globule transitions. The R_g and R_h parameters are measurable by biochemical/biophysical techniques and suitable to investigate the impact of sequence determinants on IDP/IDR conformation. As physical principles underlying R_g and R_h are distinct, they report on slightly different protein features.

The R_h is the radius of an idealized sphere having the same diffusion coefficient as the molecule of interest (Figure A1a). It can be calculated according to the Stokes–Einstein relation in Equation (A1), where k_B is the Boltzmann constant, T is the temperature, η is the viscosity coefficient of the medium, and D is the translational diffusion coefficient:

$$R_h = \frac{k_B T}{6\pi\eta D} \quad \text{(A1)}$$

R_h can be experimentally measured by SEC, pulsed-field-gradient NMR, dynamic light scattering, and analytical ultracentrifugation.

The R_g is numerically assimilable to R_h, yet conceptually different and specifically used in polymer physics to describe the dimension of a polymer chain (Figure A1b). It can be measured through static light scattering, small angle neutron- and X-ray scattering, or calculated from simulated conformational ensembles. For a polymer composed of N subunits, R_g is defined by Equation (A2):

$$R_g^2 = \frac{1}{N+1}\sum_{i=0}^{N} < (R_i - R_{CM})^2 >, \quad \text{(A2)}$$

where R_i is the vector indicating the position of any subunit and R_{CM} is the vector indicating the position of the polymer mass center [170]. R_g is averaged over the polymer ensemble, as denoted by the angular brackets < . . . > (Figure A1b). Notably, R_g is also proportional to the end-to-end distance S, according to the following equation:

$$R_g^2 = S^2/6. \quad \text{(A3)}$$

For the sake of completeness, S is given by:

$$S^2 = <(R_N - R_0)^2>, \quad \text{(A4)}$$

where S is the average end-to-end distance between the position of the first (R_0) and the latter subunit (R_N) over the ensemble.

For an ideal polymer whose residue units do not interact with each other, *i.e.*, a freely jointed chain, R_g is given by:

$$R_g = \frac{1}{\sqrt{6}}1\sqrt{Na}, \quad \text{(A5)}$$

where *a* is dependent on the polymer stiffness and can vary over orders of magnitude. The scaling law of Equation (A5) could be re-written in more general terms as Equation (A6), to relate both R_g and R_h to chain length, *i.e.*, to *N*:

$$R_{xs} = R_{0xs}N^{\nu}_{xs}, \quad \text{(A6)}$$

where R0 is the value for the compact state, in a given solvent, for structures where the volume scales linearly with the number of subunits; $x = (g, h)$ determines whether the relationship refers to R_g or R_h; and s = (folded, unfolded, IDP) refers to the specific conformational class. It can be observed that the scaling law has not yet been declined for the different categories of IDPs. Empirically determined values for these parameters reveal that the scaling exponents v can vary from 0.33, for folded proteins, to 0.6 for disordered ones [171].

An empirical linear relationship between the R_g of a simulated conformational ensemble and its R_h was derived by Choy and collaborators [172]:

$$R_g / R_h = aR_g + b. \quad \text{(A7)}$$

Although a roughly linear relationship exists for each protein, the values of a and b are different and vary with the protein length. This issue was subsequently addressed by Nygaard and co-workers [171], through an empirical method based on Rg and Rh data for 100 conformations of 30 IDPs or IDP-like peptides:

$$\frac{R_g}{R_n}\left(N, R_g\right) = \frac{\alpha_1(R_g - a_2 N^{0.33})}{N^{0.60} - N^{0.33}} + \alpha_3 \quad \text{(A8)}$$

where α_1 = (0.216 ± 0.001) $Å^{-1}$, α_2 = (4.06 ± 0.02) Å, and α3 = (0.821±0.002), obtained as the best-fit values for empirical data from calculations/simulations.

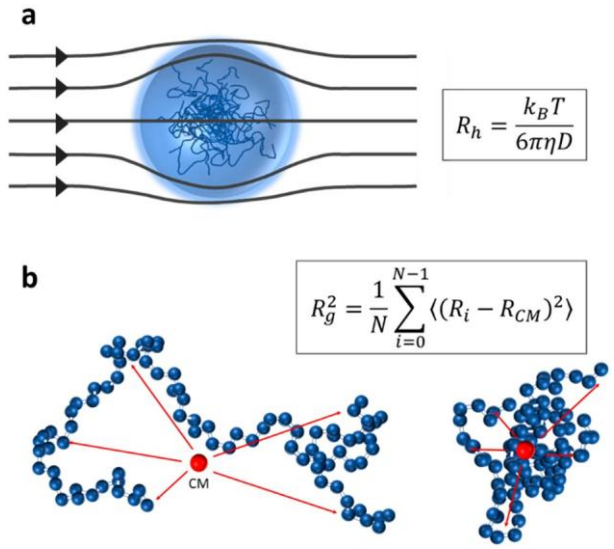


Figure A1. (**a**) Schematic representation and definition of R_h (Figure adapted from [171]). (**b**) Schematic representation and definition of R_g showing two conformations sampled from the ensemble of an ideal polymer, a freely jointed chain of 85 subunits, highlighting the position of their mass center (CM, at position R_{CM}) and its distance from given subunits at generic position R_i. A bias toward compact or extended conformations, reshaping the conformational ensemble, results in different average distances between CM and each subunit and, hence, different $< R_g >$ values.

Appendix B. Protein Fibrillation and Peptide Self-Assembly

The misfolding of proteins and peptides can lead to supramolecular assembly and ordered amyloid-type fibrils [173], as typically observed in neurodegenerative disorders. This transition results in a cross-β (core) structure [173] and can be triggered by different physicochemical stimuli (*e.g.*, temperature, Coulomb

interaction, pH, metal ions, and chemical additives). Chain segments embedded within the cross-β core generally possess hydrophobic clusters, few charged residues, and patterns of alternating hydrophobic and hydrophilic amino acids, as well as an intrinsic β-sheet propensity [173]. Amyloid fibrils maximize the number of hydrogen bonds and hydrophobic interactions along the fibril axis, which generally is achieved through a parallel in-register arrangement of strands. Such an extended hydrogen-bonded β-sheet imparts mechanical strength and stability to amyloid fibrils, regardless of the folded or disordered nature of the native state [174–176]. A continuum of accessible pathways for amyloid conversion has been suggested, dependent on the environmental conditions, sequence, and conformational state of the amyloidogenic monomer [173,177–179]. Such alternative mechanisms could converge into the formation of growth-competent nuclei or aggregates, which support conversion into amyloid-like oligomers and eventually fibrils. This variety of fibril formation pathways could concern even a same polypeptide chain, resulting in polymorphism of the mature fibrils.

A way to tackle the extraordinary complexity of these structural transitions consists in investigating and manipulating self-assembling peptides (16–20 residues). Self-assembly refers to the association of two or more molecules, giving rise to ordered nanostructures (e.g., nanofibers, nanoribbons, nanotubes, or vesicles), from a few nanometer to hundreds of micron size. Such materials attract enormous interest as being bio-compatible and suitable to green technology and biomedical applications. Their formation is due to the synergistic and cooperative effect of various intramolecular and intermolecular non-covalent interactions, including hydrogen bonding, π–π stacking, electrostatic, hydrophobic and van der Waals interactions, as well as metal–ion coordination, which are also relevant for the formation of amyloid fibrils [180]. In line with the focus of this review, some examples of electrostatically driven self-assembly are described here. Peptide nanostructures based on electrostatic interactions have been reviewed [181]. The 20-mer MAX1 is

mainly composed of Val and Lys residues (pI: 10.85) and reversibly responds to pH with gelification [182]. Under basic conditions, due to the neutralization of Lys side chains, the peptide folds into an amphiphilic β-hairpin with one face lined by hydrophobic Val residues and the other face exposing hydrophilic Lys residues. Self-assembly then occurs between hairpins both laterally, via H-bond formation, and facially, by hydrophobic association of the valine-rich faces of the folded peptide [182].

Another example is represented by the group of ionic self-complementary peptides, with a net charge that is almost zero at neutral pH and special amphiphilic features. In particular, RADA (RADA4 or RADA16) has been widely studied as it forms relatively regular nanofibers giving rise to a hydrogel exploitable for cell culturing, encapsulation, and molecule delivery in vivo. One side of the RADA4 monomer is thought to consist predominantly of nonpolar hydrophobic Ala (A), and the other side of alternating oppositely charged amino acids, namely Arg (R) and Asp (D). Recent publications suggest that the surface net charge is a crucial physicochemical parameter in protein aggregation [183,184]. The latter work, based on CD analysis and metadynamics simulations, concludes that RADA16 fibrillation could be easily modulated by pH and ionic strength. In particular, fibril formation is promoted at pH~pI, when a low net charge is achieved. This is due to poor electrostatic repulsions, favoring intermolecular interactions, and to the peptide conformation. Indeed, at this pH, RADA16 is predicted to have a β-hairpin conformation, promoting fibril formation. Ionic strength can promote fibrillation even at pH far from pI and in the presence of a large number of uncompensated charges, by shielding effects. This study highlights the main aspects that have to be taken into consideration when describing protein fibrillation, i.e., the effect of polymer charges and solution charges on conformation, aggregation propensity, and fibril shape. These issues are addressed in more detail in Section 4 of the main text.

Appendix C. Liquid–Liquid Phase Separation

In polymer chemistry, liquid–liquid phase separation (LLPS) occurs whenever polymer self-interactions prevail over polymer–solvent ones (poor solvent conditions), defining two immiscible phases, a low-concentration diluted phase and a high-concentration condensed one [128]. Figure A2 illustrates this concept referring to three kinds of stimuli-sensitive systems. Both phases retain liquid-like properties and the same chemical potential, thus impeding any net diffusive flux between them, yet permitting the rapid exchange of single molecules [128]. In cellular protoplasm, phase separation could account for the formation of proteinaceous, membraneless compartments assimilable to organelles, employed to control chemical reactions (specificity, inhibition, kinetics), store biomolecules, sequester damaging factors, and enhance signal transduction [129]. Such biomolecular condensates are optically resolvable as micron-sized, spherical, and deformable coacervates [129,150] localized either in the cytoplasm or in the nucleus. Among others, we can list, in the nucleus, nucleoli [185,186], Cajal bodies [187], PML (promyelocytic leukemia) nuclear bodies [188,189], and nuclear speckles [190]; and in the cytoplasm, P bodies, stress, and germ granules [191].

Multivalent molecules, such as IDPs and modular multi-domain proteins, elicit the formation of either large homogeneous or heterogeneous complexes, reducing macromolecule solvation preceding phase separation. All kinds of weak and non-specific interactions can contribute to such adhesive contacts. Specific involvement in coacervation has been documented for cation–π contacts, between positively charged residues and π electrons in aromatic residues, π–π interactions between aromatic rings, electrostatic attractions, and dipolar forces [128,136], along with hydrogen bonding [192] and hydrophobic interactions [193]. Poly-ions, as well as RNA or ssDNA, can seed LLPS by favoring the interactions of RNA with the RNA-binding domain [130] or RNA base pairing [129]. The consequent extent of interactions and spatial ordering within the droplet dictates its state of matter,

which could be simply liquid, liquid-crystalline, liquid-gel, semi-crystalline-solid, crystalline-solid, solid-gel, or solid [129]. In particular, IDP-based liquid droplets likely undergo a progressive phase transition towards viscoelastic and then solid aggregates, eventually interrupting any material exchange with the cytoplasm. Several mechanisms could account for this hardening process, including the growth of amyloid fibrils, vitrification of amorphous aggregates, crosslinking, or simply entanglement of disordered chains [128]. Changes in the droplet physical properties are related to several functions [194,195]. For instance, LLPS and other forms of hardening may contribute to reduce the kinetics of reactions, benefiting from the increase in viscosity to slow the diffusion of molecules and thus reaction rates. This has led to envisage a biotechnological exploitation of the LLPS phenomenon in the field of enzymology and heterogeneous-phase catalysis [196]. However, the precise mechanisms underlying phase separation, and most of all their regulation, still remain to be fully elucidated.

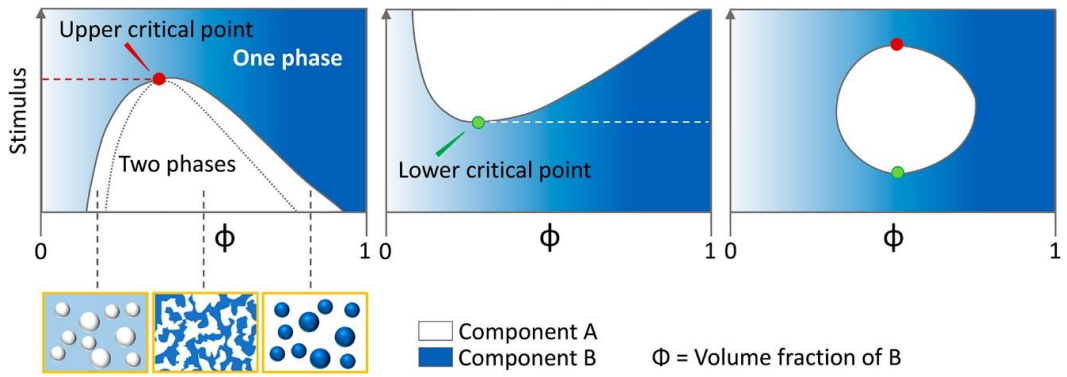


Figure A2. A two-component mixture can give rise to a single-phase well-mixed system, or to a two-phase system, in which the two components are separated. The separation into two phases can occur below or above a critical value of temperature, pH, ionic strength, etc. The diagram illustrates the coexistence curve (binodal) of three reference systems in response to a generic stimulus. (**Left side):** the transition can occur below a threshold point, which is therefore called the upper

critical point (UCP). Phase separation is only observed upon a rate-limiting nucleation event, under conditions lying in between the binodal (indicated as solid line) and spinodal curves (indicated as dotted line). Once the spinodal curve is crossed, nucleation becomes dispensable and phase separation spontaneously occurs by spinodal decomposition. The pictorial schemes in the yellow boxes show spinodal decomposition (in the middle) and de-mixed states of components A and B (left and right); (**Middle**): phase separation occurs beyond the lower critical point (LCP); (**Right side**): this system features both UCP and LCP behavior.

References

1. Boehr, D.D.; Nussinov, R.; Wright, P.E. The role of dynamic conformational ensembles in biomolecular recognition. Nat. Chem. Biol. 2009, 5, 789–796. [CrossRef]

2. Wright, P.E.; Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. Nat. Rev. Mol. Cell Biol. 2015, 16, 18–29. [CrossRef]

3. Uversky, V.N. The most important thing is the tail: Multitudinous functionalities of intrinsically disordered protein termini. FEBS Lett. 2013, 587, 1891–1901. [CrossRef]

4. Hegyi, H.; Schad, E.; Tompa, P. Structural disorder promotes assembly of protein complexes. BMC Struct. Biol.

2007, 7, 65. [CrossRef]

5. Wright, P.E.; Dyson, H.J. Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm. J. Mol. Biol. 1999, 293, 321–331. [CrossRef]

6. Iakoucheva, L.M.; Brown, C.J.; Lawson, J.D.; Obradovic´, Z.; Dunker, A.K. Intrinsic disorder in cell-signaling and cancer-associated proteins. J. Mol. Biol. 2002, 323, 573–584. [CrossRef]

7. Galea, C.A.; Wang, Y.; Sivakolundu, S.G.; Kriwacki, R.W. Regulation of cell division by intrinsically unstructured proteins: Intrinsic flexibility, modularity, and signaling conduits. Biochemistry 2008, 47, 7598–7609. [CrossRef]

8. Uversky, V.N.; Oldfield, C.J.; Dunker, A.K. Intrinsically disordered proteins in human diseases: Introducing the D2 concept. Annu. Rev. Biophys. 2008, 37, 215–246. [CrossRef]

9. Tantos, A.; Kalmar, L.; Tompa, P. The role of structural disorder in cell cycle regulation, related clinical proteomics, disease development and drug targeting. Expert Rev. Proteom. 2015, 12, 221–233. [CrossRef]

10. Uversky, V.N. What does it mean to be natively unfolded? Eur. J. Biochem. 2002, 269, 2–12. [CrossRef]

11. Holehouse, A.S.; Das, R.K.; Ahad, J.N.; Richardson, M.O.G.; Pappu, R.V. CIDER: Resources to analyze sequence-ensemble relationships of intrinsically disordered proteins. Biophys. J. 2017, 112, 16–21. [CrossRef]

12. Pace, C.N.; Grimsley, G.R.; Scholtz, J.M. Protein ionizable groups: PK values and their contribution to protein stability and solubility. J. Biol. Chem. 2009, 284, 13285–13289. [CrossRef]

13. Bashford, D.; Karplus, M. pKa‘s of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. Biochemistry 1990, 29, 10219–10225. [CrossRef]

14. Bundi, A.; Wüthrich, K. 1H-NMR parameters of the common amino acid residues measured in aqueous

solutions of the linear tetrapeptides H-Gly-Gly-X-L-Ala-OH. Biopolym. Orig. Res. Biomol. 1979, 18, 285–297. [CrossRef]

15. Pujato, M.; Bracken, C.; Mancusso, R.; Cataldi, M.; Tasayco, M.L. pH dependence of amide chemical shifts in natively disordered polypeptides detects medium-range interactions with ionizable residues. Biophys. J. 2005, 89, 3293–3302. [CrossRef]

16. Neira, J.L.; Rizzuti, B.; Iovanna, J.L. Determinants of the pKa values of ionizable residues in an intrinsically disordered protein. Arch. Biochem. Biophys. 2016, 598, 18–27. [CrossRef]

17. Mason, P.E.; Neilson, G.W.; Dempsey, C.E.; Barnes, A.C.; Cruickshank, J.M. The hydration structure of guanidinium and thiocyanate ions: Implications for protein stability in aqueous solution. Proc. Natl. Acad. Sci. USA 2003, 100, 4557–4561. [CrossRef]

18. Fitch, C.A.; Platzer, G.; Okon, M.; Garcia-Moreno, E.B.; McIntosh, L.P. Arginine: Its pKa value revisited.

Protein Sci. 2015, 24, 752–761. [CrossRef]

19. Chakrabarti, P.; Bhattacharyya, R. Geometry of nonbonded interactions involving planar groups in proteins.

Prog. Biophys. Mol. Biol. 2007, 95, 83–137. [CrossRef]

20. Higgs, P.G.; Joanny, J.F. Theory of polyampholyte solutions. J. Chem. Phys. 1991, 94, 1543–1554. [CrossRef]

21. Ha, B.Y.; Thirumalai, D. Persistence length of intrinsically stiff polyampholyte chains. J. De Phys. II

1997, 7, 887–902. [CrossRef]

22. Gutin, A.M.; Shakhnovich, E.I. Effect of a net charge on the conformation of polyampholytes. Phys. Rev. E

1994, 50, R3322. [CrossRef] [PubMed]

23. Dobrynin, A.V.; Colby, R.H.; Rubinstein, M. Polyampholates. J. Polym. Sci. Part B Polym. Phys.

2004, 42, 3513–3538. [CrossRef]

24. Dobrynin, A.V.; Rubinstein, M. Theory of polyelectrolytes in solutions and at surfaces. Prog. Polym. Sci.

2005, 30, 1049–1118. [CrossRef]

25. Das, R.K.; Pappu, R.V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. Proc. Natl. Acad. Sci. USA 2013, 110, 13392–13397. [CrossRef]

26. Srivastava, D.; Muthukumar, M. Sequence dependence of conformations of polyampholytes. Macromolecules

1996, 29, 2324–2326. [CrossRef]

27. Uversky, V.N.; Gillespie, J.R.; Fink, A.L. Why are “natively unfolded” proteins unstructured under physiologic conditions? Proteins Struct. Funct. Bioinform. 2000, 41, 415–427. [CrossRef]

28. Uversky, V.N. Intrinsically disordered proteins and their “mysterious”(meta) physics. Front. Phys. 2019, 7, 10. [CrossRef]

29. Mao, A.H.; Crick, S.L.; Vitalis, A.; Chicoine, C.L.; Pappu, R.V. Net charge per residue modulates conformational ensembles of intrinsically disordered proteins. Proc. Natl. Acad. Sci. USA 2010, 107, 8183–8188. [CrossRef]

30. Pappu, R.V.; Wang, X.; Vitalis, A.; Crick, S.L. A polymer physics perspective on driving forces and mechanisms for protein aggregation. Arch. Biochem. Biophys. 2008, 469, 132–141. [CrossRef]

31. Rapaport, D.C. Configurational properties of polymers in a good solvent.

J. Phys. A Math. Gen. 1976, 9, 1521. [CrossRef]

32. Das, R.K.; Ruff, K.M.; Pappu, R.V. Relating sequence encoded information to form and function of intrinsically disordered proteins. Curr. Opin. Struct. Biol. 2015, 32, 102–112. [CrossRef] [PubMed]

33. Firman, T.; Ghosh, K. Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins. J. Chem. Phys. 2018, 148, 123305. [CrossRef] [PubMed]

34. Fuertes, G.; Banterle, N.; Ruff, K.M.; Chowdhury, A.; Mercadante, D.; Koehler, C.; Kachala, M.; Girona, G.E.; Milles, S.; Mishra, A. Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements. Proc. Natl. Acad. Sci. USA 2017, 114, E6342–E6351. [CrossRef] [PubMed]

35. Gibbs, E.B.; Lu, F.; Portz, B.; Fisher, M.J.; Medellin, B.P.; Laremore, T.N.; Zhang, Y.J.; Gilmour, D.S.; Showalter, S.A. Phosphorylation induces sequence-specific conformational switches in the RNA polymerase II C-terminal domain. Nat. Commun. 2017, 8, 1–11. [CrossRef]

36. Martin, E.W.; Holehouse, A.S.; Pappu, R.V.; Mittag, T. Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and Upon Multisite Phosphorylation. Biophys. J. 2017, 112, 510a. [CrossRef]

37. Holehouse, A.S.; Pappu, R.V. Collapse transitions of proteins and the interplay among backbone, sidechain, and solvent interactions. Annu. Rev. Biophys. 2018, 47, 19–39. [CrossRef]

38. Knox-Brown, P.; Rindfleisch, T.; Günther, A.; Balow, K.; Bremer, A.; Walther, D.; Miettinen, M.S.; Hincha, D.K.; Thalhammer, A. Similar Yet Different–Structural and Functional Diversity among Arabidopsis thaliana LEA_4 Proteins. Int. J. Mol. Sci. 2020, 21, 2794. [CrossRef]

39. Gunner, M.R.; Saleh, M.A.; Cross, E.; Wise, M. Backbone dipoles generate positive potentials in all proteins: Origins and implications of the effect. Biophys. J. 2000, 78, 1126–1144. [CrossRef]

40. Hol, W.G.J. Effects of the α-helix dipole upon the functioning and structure of proteins and peptides.

Adv. Biophys. 1985, 19, 133–165. [CrossRef]

41. Xu, S. Cross-β-sheet structure in amyloid fiber formation. J. Phys. Chem. B 2009, 113, 12447–12455. [CrossRef] [PubMed]

42. Flory, P.J.; Volkenstein, M. Statistical mechanics of chain molecules. Biopolym. Orig. Res. Biomol. 1969, 8, 699–700. [CrossRef]

43. Flory, P.J. Principles of Polymer Chemistry; Cornell University Press: Ithaca, NY, USA, 1953.

44. Sawle, L.; Ghosh, K. A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins. J. Chem. Phys. 2015, 143, 08B615_611. [CrossRef] [PubMed]

45. Muthukumar, M. Collapse transition of a stiff chain. J. Chem. Phys. 1984, 81, 6272–6276. [CrossRef]

46. Lin, Y.-H.; Chan, H.S. Phase separation and single-chain compactness

of charged disordered proteins are strongly correlated. Biophys. J. 2017, 112, 2043–2046. [CrossRef] [PubMed]

47. Hatos, A.; Hajdu-Soltész, B.; Monzon, A.M.; Palopoli, N.; Álvarez, L.; Aykac-Fas, B.; Bassot, C.; Benítez, G.I.; Bevilacqua, M.; Chasapi, A. DisProt: Intrinsic protein disorder annotation in 2020. Nucleic Acids Res. 2020, 48, D269–D276. [CrossRef] [PubMed]

48. Piovesan, D.; Tabaro, F.; Micˇetic´, I.; Necci, M.; Quaglia, F.; Oldfield, C.J.; Aspromonte, M.C.; Davey, N.E.; Davidovic´, R.; Dosztányi, Z. DisProt 7.0: A major update of the database of disordered proteins. Nucleic Acids Res. 2017, 45, D219–D227. [CrossRef] [PubMed]

49. Sickmeier, M.; Hamilton, J.A.; LeGall, T.; Vacic, V.; Cortese, M.S.; Tantos, A.; Szabo, B.; Tompa, P.; Chen, J.; Uversky, V.N. DisProt: The database of disordered proteins. Nucleic Acids Res. 2007, 35, D786–D793. [CrossRef] [PubMed]

50. Huihui, J.; Firman, T.; Ghosh, K. Modulating charge patterning and ionic strength as a strategy to induce conformational changes in intrinsically disordered proteins. J. Chem. Phys. 2018, 149, 085101. [CrossRef]

51. Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. Charge interactions can dominate the dimensions of intrinsically disordered proteins. Proc. Natl. Acad. Sci. USA 2010, 107, 14609–14614. [CrossRef]

52. Sizemore, S.M.; Cope, S.M.; Roy, A.; Ghirlanda, G.; Vaiana, S.M. Slow internal dynamics and charge expansion in the disordered protein CGRP: A comparison with amylin. Biophys. J. 2015, 109, 1038–1048. [CrossRef] [PubMed]

53. Wright, M.R. An Introduction to Aqueous Electrolyte Solutions; John Wiley & Sons: Hoboken, NJ, USA, 2007.

54. Debye, P.; Hückel, E. Zur theorie der elektrolyte. II. Das Grenzgesetz Für Die Elektr. Leitfähigkeit. Phys. 1923, 305.

55. Liu, B.; Chia, D.; Csizmok, V.; Farber, P.; Forman-Kay, J.D.; Gradinaru, C.C. The effect of intrachain electrostatic repulsion on conformational disorder and dynamics of the Sic1 protein. J. Phys. Chem. B 2014, 118, 4088–4097. [CrossRef]

56. Vancraenenbroeck, R.; Harel, Y.S.; Zheng, W.; Hofmann, H. Polymer effects modulate binding affinities in disordered proteins. Proc. Natl. Acad. Sci. USA 2019, 116, 19506–19512. [CrossRef] [PubMed]

57. England, J.L.; Haran, G. Role of solvation effects in protein denaturation: From thermodynamics to single molecules and back. Annu. Rev. Phys. Chem. 2011, 62, 257–277. [CrossRef] [PubMed]

58. Beveridge, R.; Migas, L.G.; Das, R.K.; Pappu, R.V.; Kriwacki, R.W.; Barran, P.E. Ion mobility mass spectrometry uncovers the impact of the patterning of oppositely charged residues on the conformational distributions of intrinsically disordered proteins. J. Am. Chem. Soc. 2019, 141, 4908–4918. [CrossRef]

59. Srinivasan, N.; Bhagawati, M.; Ananthanarayanan, B.; Kumar, S. Stimuli-sensitive intrinsically disordered

protein brushes. Nat. Commun. 2014, 5, 1–8. [CrossRef] [PubMed]

60. Stuart, M.A.C.; Huck, W.T.S.; Genzer, J.; Müller, M.; Ober, C.; Stamm, M.; Sukhorukov, G.B.; Szleifer, I.; Tsukruk, V.V.; Urban, M. Emerging applications of stimuli-responsive polymer materials. Nat. Mater. 2010, 9, 101–113. [CrossRef]

61. Biesalski, M.; Johannsmann, D.; Rühe, J. Electrolyte-induced collapse of a polyelectrolyte brush. J. Chem. Phys.

2004, 120, 8807–8814. [CrossRef]

62. Ayres, N.; Cyrus, C.D.; Brittain, W.J. Stimuli-responsive surfaces using polyampholyte polymer brushes prepared via atom transfer radical polymerization. Langmuir 2007, 23, 3744–3749. [CrossRef]

63. Lei, R.; Lee, J.P.; Francis, M.B.; Kumar, S. Structural regulation of a neurofilament-inspired intrinsically disordered protein brush by multisite phosphorylation. Biochemistry 2018, 57, 4019–4028. [CrossRef] [PubMed]

64. Yuan, A.; Rao, M.V.; Nixon, R.A. Neurofilaments and neurofilament proteins in health and disease. Cold Spring Harb. Perspect. Biol. 2017, 9, a018309. [CrossRef] [PubMed]

65. Tedeschi, G.; Salladini, E.; Santambrogio, C.; Grandori, R.; Longhi, S.; Brocca, S. Conformational response to charge clustering in synthetic intrinsically disordered proteins. Biochim. Biophys. Acta (BBA) Gen. Subj. 2018, 1862, 2204–2214. [CrossRef] [PubMed]

66. Das, R.K.; Huang, Y.; Phillips, A.H.; Kriwacki, R.W.; Pappu, R.V. Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling. Proc. Natl. Acad. Sci. USA 2016, 113, 5616–5621. [CrossRef] [PubMed]

67. Beveridge, R.; Chappuis, Q.; Macphee, C.; Barran, P. Mass spectrometry methods for intrinsically disordered proteins. Analyst 2013, 138, 32–42. [CrossRef]

68. Sherry, K.P.; Das, R.K.; Pappu, R.V.; Barrick, D. Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor. Proc. Natl. Acad. Sci. USA 2017, 114, E9243–E9252. [CrossRef]

69. Loell, K.; Nanda, V. Marginal protein stability drives subcellular proteome isoelectric point. Proc. Natl. Acad. Sci. USA 2018, 115, 11778–11783. [CrossRef]

70. Flory, P.J. Thermodynamics of high polymer solutions. J. Chem. Phys. 1941, 9, 660. [CrossRef]

71. Huggins, M.L. Solutions of long chain compounds. J. Chem. Phys. 1941, 9, 440. [CrossRef]

72. Tsiolaki, P.L.; Nastou, K.C.; Hamodrakas, S.J.; Iconomidou, V.A. Mining databases for protein aggregation: A review. Amyloid 2017, 24, 143–152. [CrossRef]

73. Pallarès, I.; Ventura, S. Advances in the prediction of protein aggregation propensity. Curr. Med. Chem.

2019, 26, 3911–3920. [CrossRef] [PubMed]

74. Højgaard, C.; Kofoed, C.; Espersen, R.; Johansson, K.E.; Villa, M.; Willemoës, M.; Lindorff-Larsen, K.; Teilum, K.; Winther, J.R. A soluble,

folded protein without charged amino acid residues. Biochemistry 2016, 55, 3949–3956. [CrossRef] [PubMed]

75. Lawrence, M.S.; Phillips, K.J.; Liu, D.R. Supercharging proteins can impart unusual resilience. J. Am. Chem. Soc. 2007, 129, 10110–10112. [CrossRef] [PubMed]

76. Boyken, S.E.; Benhaim, M.A.; Busch, F.; Jia, M.; Bick, M.J.; Choi, H.; Klima, J.C.; Chen, Z.; Walkey, C.; Mileant, A. De novo design of tunable, pH-driven conformational changes. Science 2019, 364, 658–664. [CrossRef] [PubMed]

77. Liu, Z.; Huang, Y. Advantages of proteins being disordered. Protein Sci. 2014, 23, 539–550. [CrossRef]

78. Minde, D.P.; Halff, E.F.; Tans, S. Designing disorder: Tales of the unexpected tails. Intrinsically Disord. Proteins

2013, 1, e26790. [CrossRef]

79. Davis, G.D.; Elisee, C.; Newham, D.M.; Harrison, R.G. New fusion protein systems designed to give soluble expression in Escherichia coli. Biotechnol. Bioeng. 1999, 65, 382–388. [CrossRef]

80. Hoh, J.H. Functional protein domains from the thermally driven motion of polypeptide chains: A proposal.

Proteins Struct. Funct. Bioinform. 1998, 32, 223–228. [CrossRef]

81. Santner, A.A.; Croy, C.H.; Vasanwala, F.H.; Uversky, V.N.; Van, Y.-Y.J.; Dunker, A.K. Sweeping away protein aggregation with entropic bristles: Intrinsically disordered protein fusions enhance soluble expression. Biochemistry 2012, 51, 7250–7262. [CrossRef]

82. Karlin, D.; Longhi, S.; Receveur, V.; Canard, B. The N-terminal domain of the phosphoprotein of morbilliviruses belongs to the natively unfolded class of proteins. Virology 2002, 296, 251–262. [CrossRef]

83. Tedeschi, G.; Mangiagalli, M.; Chmielewska, S.; Lotti, M.; Natalello, A.; Brocca, S. Aggregation properties of a disordered protein are tunable by pH and depend on its net charge per residue. Biochim. Biophys. Acta (BBA) Gen. Subj. 2017, 1861, 2543–2550. [CrossRef] [PubMed]

84. Santos, J.; Iglesias, V.; Santos-Suárez, J.; Mangiagalli, M.; Brocca, S.; Pallarès, I.; Ventura, S. pH-Dependent Aggregation in Intrinsically Disordered Proteins Is Determined by Charge and Lipophilicity. Cells 2020, 9, 145. [CrossRef] [PubMed]

85. Perrett, S.; Pham, C.L.L.; Kwan, A.H.; Sunde, M. Functional amyloid: Widespread in Nature, diverse in purpose. Essays Biochem. 2014, 56, 207–219. [CrossRef] [PubMed]

86. Otzen, D.; Riek, R. Functional amyloids. Cold Spring Harb. Perspect. Biol. 2019, 11, a033860. [CrossRef]

87. Benson, M.D.; Buxbaum, J.N.; Eisenberg, D.S.; Merlini, G.; Saraiva, M.J.M.; Sekijima, Y.; Sipe, J.D.; Westermark, P. Amyloid nomenclature 2018: Recommendations by the International Society of Amyloidosis (ISA) nomenclature committee. Amyloid 2018, 25, 215–219. [CrossRef]

88. Ebo, J.S.; Guthertz, N.; Radford, S.E.; Brockwell, D.J. Using protein engineering to understand and modulate

aggregation. Curr. Opin. Struct. Biol. 2020, 60, 157–166. [CrossRef]

89. Ciryam, P.; Tartaglia, G.G.; Morimoto, R.I.; Dobson, C.M.; Vendruscolo, M. Neurodegenerative diseases and widespread aggregation are associated with supersaturated proteins. Cell Rep. 2013, 5, 43. [CrossRef]

90. Alberti, S.; Halfmann, R.; King, O.; Kapila, A.; Lindquist, S. A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell 2009, 137, 146–158. [CrossRef]

91. Wickner, R.B.; Shewmaker, F.P.; Bateman, D.A.; Edskes, H.K.; Gorkovskiy, A.; Dayani, Y.; Bezsonov, E.E. Yeast prions: Structure, biology, and prion-handling systems. Microbiol. Mol. Biol. Rev. 2015, 79, 1–17. [CrossRef]

92. Yamaguchi, K.-I.; Kuwata, K. Formation and properties of amyloid fibrils of prion protein. Biophys. Rev.

2018, 10, 517–525. [CrossRef]

93. Uptain, S.M.; Lindquist, S. Prions as protein-based genetic elements. Annu. Rev. Microbiol. 2002, 56, 703–741. [CrossRef] [PubMed]

94. Toombs, J.A.; Petri, M.; Paul, K.R.; Kan, G.Y.; Ben-Hur, A.; Ross, E.D. De novo design of synthetic prion domains. Proc. Natl. Acad. Sci. USA 2012, 109, 6519–6524. [CrossRef] [PubMed]

95. Angarica, V.E.; Ventura, S.; Sancho, J. Discovering putative prion sequences in complete proteomes using probabilistic representations of Q/N-rich domains. BMC Genom. 2013, 14, 316.

96. Tycko, R.; Wickner, R.B. Molecular structures of amyloid and prion fibrils: Consensus versus controversy.

Acc. Chem. Res. 2013, 46, 1487–1496. [CrossRef] [PubMed]

97. Hughes, M.P.; Sawaya, M.R.; Boyer, D.R.; Goldschmidt, L.; Rodriguez, J.A.; Cascio, D.; Chong, L.; Gonen, T.; Eisenberg, D.S. Atomic structures of low-complexity protein segments reveal kinked β sheets that assemble networks. Science 2018, 359, 698–701. [CrossRef]

98. Guenther, E.L.; Cao, Q.; Trinh, H.; Lu, J.; Sawaya, M.R.; Cascio, D.; Boyer, D.R.; Rodriguez, J.A.; Hughes, M.P.; Eisenberg, D.S. Atomic structures of TDP-43 LCD segments and insights into reversible or pathogenic aggregation. Nat. Struct. Mol. Biol. 2018, 25, 463–471. [CrossRef]

99. Batlle, C.; De Groot, N.S.; Iglesias, V.; Navarro, S.; Ventura, S. Characterization of soft amyloid cores in human prion-like proteins. Sci. Rep. 2017, 7, 1–16. [CrossRef]

100. Fernández, M.R.; Batlle, C.; Gil-García, M.; Ventura, S. Amyloid cores in prion domains: Key regulators for prion conformational conversion. Prion 2017, 11, 31–39. [CrossRef]

101. Sant'Anna, R.; Fernández, M.R.; Batlle, C.; Navarro, S.; De Groot, N.S.; Serpell, L.; Ventura, S. Characterization of amyloid cores in prion domains. Sci. Rep. 2016, 6, 1–10. [CrossRef]

102. Díaz-Caballero, M.; Navarro, S.; Fuentes, I.; Teixidor, F.; Ventura, S. Minimalist prion-inspired polar self-assembling peptides. ACS Nano 2018, 12, 5394–5407. [CrossRef]

103. Wang, W.; Navarro, S.; Azizyan, R.A.; Baño-Polo, M.; Esperante, S.A.; Kajava, A.V.; Ventura, S. Prion soft amyloid core driven self-assembly of globular proteins into bioactive nanofibrils. Nanoscale 2019, 11, 12680–12694. [CrossRef] [PubMed]

104. Glynn, C.; Sawaya, M.R.; Ge, P.; Gallagher-Jones, M.; Short, C.W.; Bowman, R.; Apostol, M.; Zhou, Z.H.; Eisenberg, D.S.; Rodriguez, J.A. Cryo-EM structure of a human prion fibril with a hydrophobic, protease-resistant core. Nat. Struct. Mol. Biol. 2020, 27, 417–423. [CrossRef] [PubMed]

105. Wang, L.-Q.; Zhao, K.; Yuan, H.-Y.; Wang, Q.; Guan, Z.; Tao, J.; Li, X.-N.; Sun, Y.; Yi, C.-W.; Chen, J. Cryo-EM structure of an amyloid fibril formed by full-length human prion protein. Nat. Struct. Mol. Biol. 2020, 1–5. [CrossRef] [PubMed]

106. Tompa, P. Structural disorder in amyloid fibrils: Its implication in dynamic interactions of proteins. FEBS J.

2009, 276, 5406–5415. [CrossRef]

107. Spillantini, M.G.; Schmidt, M.L.; Lee, V.M.Y.; Trojanowski, J.Q.; Jakes, R.; Goedert, M. α-Synuclein in Lewy bodies. Nature 1997, 388, 839–840. [CrossRef]

108. Goedert, M. Alpha-synuclein and neurodegenerative diseases. Nat. Rev. Neurosci. 2001, 2, 492–501. [CrossRef]

109. Peng, C.; Gathagan, R.J.; Lee, V.M.Y. Distinct α-Synuclein strains and implications for heterogeneity among

α-Synucleinopathies. Neurobiol. Dis. 2018, 109, 209–218. [CrossRef]

110. Twohig, D.; Nielsen, H.M. α-synuclein in the pathophysiology of Alzheimer's disease. Mol. Neurodegener.

2019, 14, 23. [CrossRef]

111. Uversky, V.N. A protein-chameleon: Conformational plasticity of α-synuclein, a disordered protein involved in neurodegenerative disorders. J. Biomol. Struct. Dyn. 2003, 21, 211–234. [CrossRef]

112. Vamvaca, K.; Volles, M.J.; Lansbury, P.T., Jr. The first N-terminal amino acids of α-synuclein are essential for α-helical structure formation in vitro and membrane binding in yeast. J. Mol. Biol. 2009, 389, 413–424. [CrossRef]

113. Breydo, L.; Wu, J.W.; Uversky, V.N. α-Synuclein misfolding and Parkinson's disease. Biochim. Biophys. Acta (BBA) Mol. Basis Dis. 2012, 1822, 261–285. [CrossRef] [PubMed]

114. Jensen, P.H.; Hager, H.; Nielsen, M.S.; Højrup, P.; Gliemann, J.; Jakes, R. α-Synuclein binds to tau and stimulates the protein kinase A-catalyzed tau phosphorylation of serine residues 262 and 356. J. Biol. Chem. 1999, 274, 25481–25489. [CrossRef]

115. Kawamata, H.; McLean, P.J.; Sharma, N.; Hyman, B.T. Interaction of α-synuclein and synphilin-1: Effect of Parkinson's disease-associated mutations. J. Neurochem. 2001, 77, 929–934. [CrossRef] [PubMed]

116. McClendon, S.; Rospigliosi, C.C.; Eliezer, D. Charge neutralization and collapse of the C-terminal tail of alpha-synuclein at low pH. Protein Sci. 2009, 18, 1531–1540. [CrossRef] [PubMed]

117. Wu, K.-P.; Weinstock, D.S.; Narayanan, C.; Levy, R.M.; Baum, J.

Structural reorganization of α-synuclein at low pH observed by NMR and REMD simulations. J. Mol. Biol. 2009, 391, 784–796. [CrossRef]

118. Ulmer, T.S.; Bax, A.; Cole, N.B.; Nussbaum, R.L. Structure and dynamics of micelle-bound human alpha-synuclein. J. Biol. Chem 2005, 280, 9595. [CrossRef]

119. Pettersen, E.F.; Goddard, T.D.; Huang, C.C.; Couch, G.S.; Greenblatt, D.M.; Meng, E.C.; Ferrin, T.E. UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. 2004, 25, 1605–1612. [CrossRef]

120. Li, Y.; Zhao, C.; Luo, F.; Liu, Z.; Gui, X.; Luo, Z.; Zhang, X.; Li, D.; Liu, C.; Li, X. Amyloid fibril structure of

α-synuclein determined by cryo-electron microscopy. Cell Res. 2018, 28, 897–903. [CrossRef]

121. Zhao, K.; Li, Y.; Liu, Z.; Long, H.; Zhao, C.; Luo, F.; Sun, Y.; Tao, Y.; Su, X.-D.; Li, D. Parkinson's disease associated mutation E46K of α-synuclein triggers the formation of a distinct fibril structure. Nat. Commun. 2020, 11, 1–9. [CrossRef]

122. Izawa, Y.; Tateno, H.; Kameda, H.; Hirakawa, K.; Hato, K.; Yagi, H.; Hongo, K.; Mizobata, T.; Kawata, Y. Role of C-terminal negative charges and tyrosine residues in fibril formation of α-synuclein. Brain Behav. 2012, 2, 595–605. [CrossRef]

123. Park, S.; Yoon, J.; Jang, S.; Lee, K.; Shin, S. The role of the acidic domain of α-synuclein in amyloid fibril formation: A molecular dynamics study. J. Biomol. Struct. Dyn. 2016, 34, 376–383. [CrossRef] [PubMed]

124. Bartels, T.; Choi, J.G.; Selkoe, D.J. α-Synuclein occurs physiologically as a helically folded tetramer that resists aggregation. Nature 2011, 477, 107–110. [CrossRef] [PubMed]

125. Burré, J.; Sharma, M.; Südhof, T.C. α-Synuclein assembles into higher-order multimers upon membrane binding to promote SNARE complex formation. Proc. Natl. Acad. Sci. USA 2014, 111, E4274–E4283. [CrossRef] [PubMed]

126. Metrick, M.A.; do Carmo Ferreira, N.; Saijo, E.; Hughson, A.G.; Kraus, A.; Orrú, C.; Miller, M.W.; Zanusso, G.; Ghetti, B.; Vendruscolo, M. Million-fold sensitivity enhancement in proteopathic seed amplification assays for biospecimens by Hofmeister ion comparisons. Proc. Natl. Acad. Sci. USA 2019, 116, 23029–23039. [CrossRef]

127. Munishkina, L.A.; Henriques, J.; Uversky, V.N.; Fink, A.L. Role of protein–water interactions and electrostatics in α-synuclein fibril formation. Biochemistry 2004, 43, 3289–3300. [CrossRef]

128. Banani, S.F.; Lee, H.O.; Hyman, A.A.; Rosen, M.K. Biomolecular condensates: Organizers of cellular biochemistry. Nat. Rev. Mol. Cell Biol. 2017, 18, 285–298. [CrossRef]

129. Boeynaems, S.; Alberti, S.; Fawzi, N.L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; Van Den Bosch, L. Protein phase separation: A new phase in cell biology. Trends Cell Biol. 2018, 28, 420–435. [CrossRef]

130. Gomes, E.; Shorter, J. The molecular language of membraneless organelles. J. Biol. Chem. 2019, 294, 7115–7127. [CrossRef]

131. Brangwynne, C.P. Soft active aggregates: Mechanics, dynamics and self-assembly of liquid-like intracellular protein bodies. Soft Matter 2011, 7, 3052–3059. [CrossRef]

132. Hyman, A.A.; Brangwynne, C.P. Beyond stereospecificity: Liquids and mesoscale organization of cytoplasm.

Dev. Cell 2011, 21, 14–16. [CrossRef]

133. Reed, E.H.; Schuster, B.S.; Good, M.C.; Hammer, D.A. SPLIT: Stable Protein Coacervation Using a Light Induced Transition. ACS Synth. Biol. 2020, 9, 500–507. [CrossRef] [PubMed]

134. Nott, T.J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T.D.; Bazett-Jones, D.P.; Pawson, T.; Forman-Kay, J.D. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. Mol. Cell 2015, 57, 936–947. [CrossRef] [PubMed]

135. Veis, A. A review of the early development of the thermodynamics of the complex coacervation phase separation. Adv. Colloid Interface Sci. 2011, 167, 2–11. [CrossRef]

136. Wang, J.; Choi, J.-M.; Holehouse, A.S.; Lee, H.O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. Cell 2018, 174, 688–699. [CrossRef] [PubMed]

137. Martin, E.W.; Mittag, T. Relationship of sequence and phase separation in protein low-complexity regions.

Biochemistry 2018, 57, 2478–2487. [CrossRef] [PubMed]

138. McGurk, L.; Gomes, E.; Guo, L.; Mojsilovic-Petrovic, J.; Tran, V.; Kalb, R.G.; Shorter, J.; Bonini, N.M. Poly (ADP-ribose) prevents pathological phase separation of TDP-43 by promoting liquid demixing and stress granule localization. Mol. Cell 2018, 71, 703–717. [CrossRef] [PubMed]

139. Li, H.-R.; Chiang, W.-C.; Chou, P.-C.; Wang, W.-J.; Huang, J.-R. TAR DNA-binding protein 43 (TDP-43) liquid–liquid phase separation is mediated by just a few aromatic residues. J. Biol. Chem. 2018, 293, 6090–6098. [CrossRef]

140. Lin, Y.; Currie, S.L.; Rosen, M.K. Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs. J. Biol. Chem. 2017, 292, 19110–19120. [CrossRef]

141. Martin, E.W.; Holehouse, A.S.; Peran, I.; Farag, M.; Incicco, J.J.; Bremer, A.; Grace, C.R.; Soranno, A.; Pappu, R.V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. Science 2020, 367, 694–699. [CrossRef]

142. Schuster, B.; Dignon, G.L.; Tang, W.S.; Kelley, F.; Ranganath, A.K.; Jahnke, C.N.; Simpkins, A.G.; Regy, R.M.; Hammer, D.A.; Good, M.C. Identifying Sequence Perturbations to an Intrinsically Disordered Protein that Determine Its Phase Separation Behavior. bioRxiv 2020. [CrossRef]

143. Mitrea, D.M.; Kriwacki, R.W. Phase separation in biology; functional organization of a higher order.

Cell Commun. Signal. 2016, 14, 1. [CrossRef] [PubMed]

144. Batlle, C.; Yang, P.; Coughlin, M.; Messing, J.; Pesarrodona, M.; Szulc, E.; Salvatella, X.; Kim, H.J.; Taylor, J.P.; Ventura, S. hnRNPDL Phase Separation Is Regulated by Alternative Splicing and Disease-Causing Mutations Accelerate Its Aggregation. Cell Rep. 2020, 30, 1117–1128. [CrossRef] [PubMed]

145. Mitrea, D.M.; Cika, J.A.; Stanley, C.B.; Nourse, A.; Onuchic, P.L.; Banerjee, P.R.; Phillips, A.H.; Park, C.-G.; Deniz, A.A.; Kriwacki, R.W. Self-interaction of NPM1 modulates multiple mechanisms of liquid–liquid phase separation. Nat. Commun. 2018, 9, 1–13. [CrossRef] [PubMed]

146. Chiu, Y.-P.; Sun, Y.-C.; Qiu, D.-C.; Lin, Y.-H.; Chen, Y.-Q.; Kuo, J.-C.; Huang, J.-R. Liquid-liquid phase separation and extracellular multivalent interactions in the tale of galectin-3. Nat. Commun. 2020, 11, 1–12. [CrossRef]

147. Salladini, E.; Debarnot, C.; Delauzun, V.; Murrali, M.G.; Sutto-Ortiz, P.; Spinelli, S.; Pierattelli, R.; Bignon, C.; Longhi, S. Phase transition and amyloid formation by a viral protein as an additional molecular mechanism of virus-induced cell toxicity. bioRxiv 2018. [CrossRef]

148. Sahli, L.; Renard, D.; Solé-Jamault, V.; Giuliani, A.; Boire, A. Role of protein conformation and weak interactions on γ-gliadin liquid-liquid phase separation. Sci. Rep. 2019, 9, 1–13. [CrossRef]

149. Kim, S.; Huang, J.; Lee, Y.; Dutta, S.; Yoo, H.Y.; Jung, Y.M.; Jho, Y.; Zeng, H.; Hwang, D.S. Complexation and coacervation of like-charged polyelectrolytes inspired by mussels. Proc. Natl. Acad. Sci. USA 2016, 113, E847–E853. [CrossRef]

150. Nott, T.J.; Craggs, T.D.; Baldwin, A.J. Membraneless organelles can melt nucleic acid duplexes and act as biomolecular filters. Nat. Chem. 2016, 8, 569–575. [CrossRef]

151. Ambadipudi, S.; Biernat, J.; Riedel, D.; Mandelkow, E.; Zweckstetter, M. Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. Nat. Commun. 2017, 8, 1–13. [CrossRef]

152. Wegmann, S.; Eftekharzadeh, B.; Tepper, K.; Zoltowska, K.M.; Bennett, R.E.; Dujardin, S.; Laskowski, P.R.; MacKenzie, D.; Kamath, T.; Commins, C. Tau protein liquid–liquid phase separation can initiate tau aggregation. EMBO J. 2018, 37, e98049. [CrossRef]

153. Majumdar, A.; Dogra, P.; Maity, S.; Mukhopadhyay, S. Liquid–Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules.

J. Phys. Chem. Lett. 2019, 10, 3929–3936. [CrossRef] [PubMed]

154. Lin, Y.; Fichou, Y.; Zeng, Z.; Hu, N.Y.; Han, S. Electrostatically driven complex coacervation and amyloid aggregation of tau are independent processes with overlapping conditions. ACS Chem. Neurosci. 2020, 11, 615–627. [CrossRef] [PubMed]

155. Boyko, S.; Qi, X.; Chen, T.-H.; Surewicz, K.; Surewicz, W.K. Liquid–liquid phase separation of tau protein: The crucial role of electrostatic interactions. J. Biol. Chem. 2019, 294, 11054–11059. [CrossRef]

156. Harmon, T.S.; Holehouse, A.S.; Rosen, M.K.; Pappu, R.V. Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins. Elife 2017, 6, e30294. [CrossRef] [PubMed]

157. Choi, J.-M.; Dar, F.; Pappu, R.V. LASSI: A lattice model for simulating phase transitions of multivalent proteins. PLoS Comput. Biol. 2019, 15, e1007028. [CrossRef]

158. Pak, C.W.; Kosno, M.; Holehouse, A.S.; Padrick, S.B.; Mittal, A.; Ali, R.; Yunus, A.A.; Liu, D.R.; Pappu, R.V.; Rosen, M.K. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. Mol. Cell 2016, 63, 72–85. [CrossRef] [PubMed]

159. Mészáros, B.; Erdo˝s, G.; Szabó, B.; Schád, É.; Tantos, Á.; Abukhairan, R.; Horváth, T.; Murvai, N.; Kovács, O.P.; Kovács, M. PhaSePro: The database of proteins driving liquid–liquid phase separation. Nucleic Acids Res. 2020, 48, D360–D367. [CrossRef] [PubMed]

160. Lin, Y.-H.; Forman-Kay, J.D.; Chan, H.S. Sequence-specific polyampholyte phase separation in membraneless organelles. Phys. Rev. Lett. 2016, 117, 178101. [CrossRef]

161. Dosztányi, Z.; Csizmok, V.; Tompa, P.; Simon, I. IUPred: Web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. Bioinformatics 2005, 21, 3433–3434. [CrossRef]

162. Alberts, B. The cell as a collection of protein machines: Preparing the next generation of molecular biologists. Cell 1998, 92, 291–294. [CrossRef]

163. Smith, L.J.; Fiebig, K.M.; Schwalbe, H.; Dobson, C.M. The concept of a random coil: Residual structure in peptides and denatured proteins. Fold. Des. 1996, 1, R95–R106. [CrossRef]

164. Huang, H.; Sarai, A. Analysis of the relationships between evolvability, thermodynamics, and the functions of intrinsically disordered proteins/regions. Comput. Biol. Chem. 2012, 41, 51–57. [CrossRef] [PubMed]

165. Tompa, P.; Fersht, A. Structure and Function of Intrinsically Disordered Proteins; CRC Press: Boca Raton, FL, USA, 2009.

166. Varadi, M.; De Baets, G.; Vranken, W.F.; Tompa, P.; Pancsa, R. AmyPro: A database of proteins with validated amyloidogenic regions. Nucleic Acids Res. 2018, 46, D387–D392. [CrossRef] [PubMed]

167. Thangakani, A.M.; Nagarajan, R.; Kumar, S.; Sakthivel, R.; Velmurugan, D.; Gromiha, M.M. CPAD, curated protein aggregation database: A repository of manually curated experimental data on protein and peptide aggregation. PLoS ONE 2016, 11. [CrossRef]

168. Ning, W.; Guo, Y.; Lin, S.; Mei, B.; Wu, Y.; Jiang, P.; Tan, X.; Zhang, W.; Chen, G.; Peng, D. DrLLPS: A data resource of liquid–liquid phase separation in eukaryotes. Nucleic Acids Res. 2020, 48, D288–D295. [CrossRef]

169. Stepto, R.; Chang, T.; Kratochvíl, P.; Hess, M.; Horie, K.; Sato, T.; Vohlídal, J. Definitions of terms relating to individual macromolecules, macromolecular assemblies, polymer

solutions, and amorphous bulk polymers (IUPAC Recommendations 2014). Pure Appl. Chem. 2015, 87, 71–120. [CrossRef]

170. Fixman, M. Radius of gyration of polymer chains. J. Chem. Phys. 1962, 36, 306–310. [CrossRef]

171. Nygaard, M.; Kragelund, B.B.; Papaleo, E.; Lindorff-Larsen, K. An efficient method for estimating the hydrodynamic radius of disordered protein conformations. Biophys. J. 2017, 113, 550–557. [CrossRef]

172. Choy, W.-Y.; Mulder, F.A.A.; Crowhurst, K.A.; Muhandiram, D.R.; Millett, I.S.; Doniach, S.; Forman-Kay, J.D.; Kay, L.E. Distribution of molecular size within an unfolded state ensemble using small-angle X-ray scattering and pulse field gradient NMR techniques. J. Mol. Biol. 2002, 316, 101–112. [CrossRef]

173. Chiti, F.; Dobson, C.M. Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. Annu. Rev. Biochem. 2017, 86, 27–68. [CrossRef]

174. L Almeida, Z.; Mm Brito, R. Structure and Aggregation Mechanisms in Amyloids. Molecules 2020, 25, 1195. [CrossRef] [PubMed]

175. Gazit, E. The “correctly folded” state of proteins: Is it a metastable state? Angew. Chem. Int. Ed.

2002, 41, 257–259. [CrossRef]

176. Baldwin, A.J.; Knowles, T.P.J.; Tartaglia, G.G.; Fitzpatrick, A.W.; Devlin, G.L.; Shammas, S.L.; Waudby, C.A.; Mossuto, M.F.; Meehan, S.; Gras, S.L. Metastability of native proteins and the phenomenon of amyloid formation. J. Am. Chem. Soc. 2011, 133, 14160–14163. [CrossRef] [PubMed]

177. Kumar, S.; Udgaonkar, J.B. Conformational conversion may precede or follow aggregate elongation on alternative pathways of amyloid protofibril formation. J. Mol. Biol. 2009, 385, 1266–1276. [CrossRef] [PubMed]

178. Wetzel, R.; Shivaprasad, S.; Williams, A.D. Plasticity of amyloid fibrils. Biochemistry 2007, 46, 1–10. [CrossRef] [PubMed]

179. Kumar, S.; Udgaonkar, J.B. Structurally distinct amyloid protofibrils form on separate pathways of aggregation of a small protein. Biochemistry 2009, 48, 6441–6449. [CrossRef] [PubMed]

180. Wang, J.; Liu, K.; Xing, R.; Yan, X. Peptide self-assembly: Thermodynamics and kinetics. Chem. Soc. Rev.

2016, 45, 5589–5604. [CrossRef] [PubMed]

181. Rehm, T.H.; Schmuck, C. Ion-pair induced self-assembly in aqueous solvents. Chem. Soc. Rev. 2010, 39, 3597–3611. [CrossRef]

182. Schneider, J.P.; Pochan, D.J.; Ozbas, B.; Rajagopal, K.; Pakstis, L.; Kretsinger, J. Responsive hydrogels from the intramolecular folding and self-assembly of a designed peptide. J. Am. Chem. Soc. 2002, 124, 15030–15037. [CrossRef]

183. Sun, Y.; Zhang, Y.; Tian, L.; Zhao, Y.; Wu, D.; Xue, W.; Ramakrishna, S.; Wu, W.; He, L. Self-assembly behaviors of molecular designer functional RADA16-I peptides: Influence of motifs, pH, and assembly

time. Biomed. Mater. 2016, 12, 015007. [CrossRef]

184. Owczarz, M.; Casalini, T.; Motta, A.C.; Morbidelli, M.; Arosio, P. Contribution of electrostatics in the fibril stability of a model ionic-complementary peptide. Biomacromolecules 2015, 16, 3792–3801. [CrossRef] [PubMed]

185. Montgomery, T.H. Comparative Cytological Studies with Especial Regard to the Morphology of the Nucleolus; Ginn: Cambridge, UK, 1900; Volume 5.

186. Pederson, T. The nucleolus. Cold Spring Harb. Perspect. Biol. 2011, 3, a000638. [CrossRef] [PubMed]

187. Ramon y Cajal, S.R. Un sencillo metodo de coloracion seletiva del reticulo protoplasmatico y sus efectos en los diversos organos nerviosos de vertebrados e invertebrados. Trab. Lab. Invest. Biol. (Madr.) 1903, 2, 129–221.

188. de Thé, H.; Chomienne, C.; Lanotte, M.; Degos, L.; Dejean, A. The t (15; 17) translocation of acute promyelocytic leukaemia fuses the retinoic acid receptor α gene to a novel transcribed locus. Nature 1990, 347, 558–561. [CrossRef] [PubMed]

189. Melnick, A.; Licht, J.D. Deconstructing a Disease: RAR, Its Fusion Partners, and Their Roles in the Pathogenesis of Acute Promyelocytic Leukemia. BloodJ. Am. Soc. Hematol. 1999, 93, 3167–3215. [CrossRef]

190. Ramon y Cajal, S.R. El núcleo de las células piramidales del cerebro humano y de algunos mamíferos.

Trab. Lab. Invest. Biol. (Madr.) 1910, 8, 27–62.

191. Decker, C.J.; Parker, R. P-bodies and stress granules: Possible roles in the control of translation and mRNA degradation. Cold Spring Harb. Perspect. Biol. 2012, 4, a012286. [CrossRef]

192. Gabryelczyk, B.; Cai, H.; Shi, X.; Sun, Y.; Swinkels, P.J.M.; Salentinig, S.; Pervushin, K.; Miserez, A. Hydrogen bond guidance and aromatic stacking drive liquid-liquid phase separation of intrinsically disordered histidine-rich peptides. Nat. Commun. 2019, 10, 1–12. [CrossRef]

193. Murthy, A.C.; Dignon, G.L.; Kan, Y.; Zerze, G.H.; Parekh, S.H.; Mittal, J.; Fawzi, N.L. Molecular interactions underlying liquid− liquid phase separation of the FUS low-complexity domain. Nat. Struct. Mol. Biol. 2019, 26, 637–648. [CrossRef]

194. Pancsa, R.; Schad, E.; Tantos, A.; Tompa, P. Emergent functions of proteins in non-stoichiometric supramolecular assemblies. Biochim. Biophys. Acta (BBA) Proteins Proteom. 2019, 1867, 970–979. [CrossRef]

195. Bratek-Skicki, A.; Pancsa, R.; Meszaros, B.; Van Lindt, J.; Tompa, P. A guide to regulation of the formation of biomolecular condensates. FEBS J. 2020, 287, 1924–1935. [CrossRef] [PubMed]

196. Küffner, A.M.; Prodan, M.; Zuccarini, R.; Capasso Palmiero, U.; Faltova, L.; Arosio, P. Acceleration of an Enzymatic Reaction in Liquid Phase Separated Compartments Based on Intrinsically Disordered Protein Domains. ChemSystemsChem 2020, 2. [CrossRef]