Review

Adenosine Triphosphate: The Primordial Molecule That Controls Protein Homeostasis and Shapes the Genome–Proteome Interface

Jianxing Song

Department of Biological Sciences, Faculty of Science, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260, Singapore; dbssjx@nus.edu.sg or songjianxing1000@gmail.com; Tel.: +65-65161013; Fax: +65-67792486

Abstract: Adenosine triphosphate (ATP) acts as the universal energy currency that drives various biological processes, while nucleic acids function to store and transmit genetic information for all living organisms. Liquid–liquid phase separation (LLPS) represents the common principle for the formation of membrane-less organelles (MLOs) composed of proteins rich in intrinsically disordered regions (IDRs) and nucleic acids. Currently, while IDRs are well recognized to facilitate LLPS through dynamic and multivalent interactions, the precise mechanisms by which ATP and nucleic acids affect LLPS still remain elusive. This review summarizes recent NMR results on the LLPS of human FUS, TDP-43, and the viral nucleocapsid (N) protein of SARS-CoV-2, as modulated by ATP and nucleic acids, revealing the following: (1) ATP binds to folded domains overlapping with nucleic-acid-binding interfaces; (2) ATP and nucleic acids interplay to biphasically modulate LLPS by competitively binding to overlapping pockets of folded domains and Arg/Lys within IDRs; (3) ATP energy-independently induces protein folding with the highest efficiency known so far. As ATP likely emerged in the prebiotic monomeric world, while LLPS represents a pivotal mechanism to concentrate and compartmentalize rare molecules for forming primordial cells, ATP appears to control protein homeostasis and shape genome–proteome interfaces throughout the evolutionary trajectory, from prebiotic origins to modern cells.

Keywords: ATP; nucleic acids; liquid–liquid phase separation (LLPS); membrane-less organelles (MLOs); intrinsically disordered regions (IDRs); protein folding; prebiotic evolution

1. Introduction

For life to emerge, two key components are essential: biological fuel providing energy for various biological processes and molecules encoding/storing genetic information. Nature solved these two problems by selecting phosphorus as a key building element for all living organisms [1–4]. Adenosine triphosphate (ATP) emerged as the universal energy currency through the hydrolysis of its high-energy bonds, which are thermodynamically favorable but kinetically controlled (Figure 1a). On the other hand, phosphate esters provide an extremely stable backbone for RNA and DNA, which function to store and transmit genetic information [5]. ATP and nucleic acids are negatively charged molecules that share similar unit compositions. They are composed of nucleotides with phosphate groups, sugars, and nitrogenous bases. However, there are key differences: ATP has three phosphate groups and solely adenine as its base, whereas nucleic acids are the polymers of nucleotides which have a single phosphate group but five distinct bases (Figure 1b). ATP has been proposed to have emerged from unique chemistry in the water of the prebiotic and monomeric world, preceding the polymerization of RNA, DNA, and proteins, as well as the rise of genetically encoded macromolecular engines [6]. Mysteriously, in modern cells, although all known ATP-dependent proteins/enzymes require only micromolar

concentrations, the cellular concentrations of ATP are very high, ranging from 2 to 12 mM depending on cell type [5].
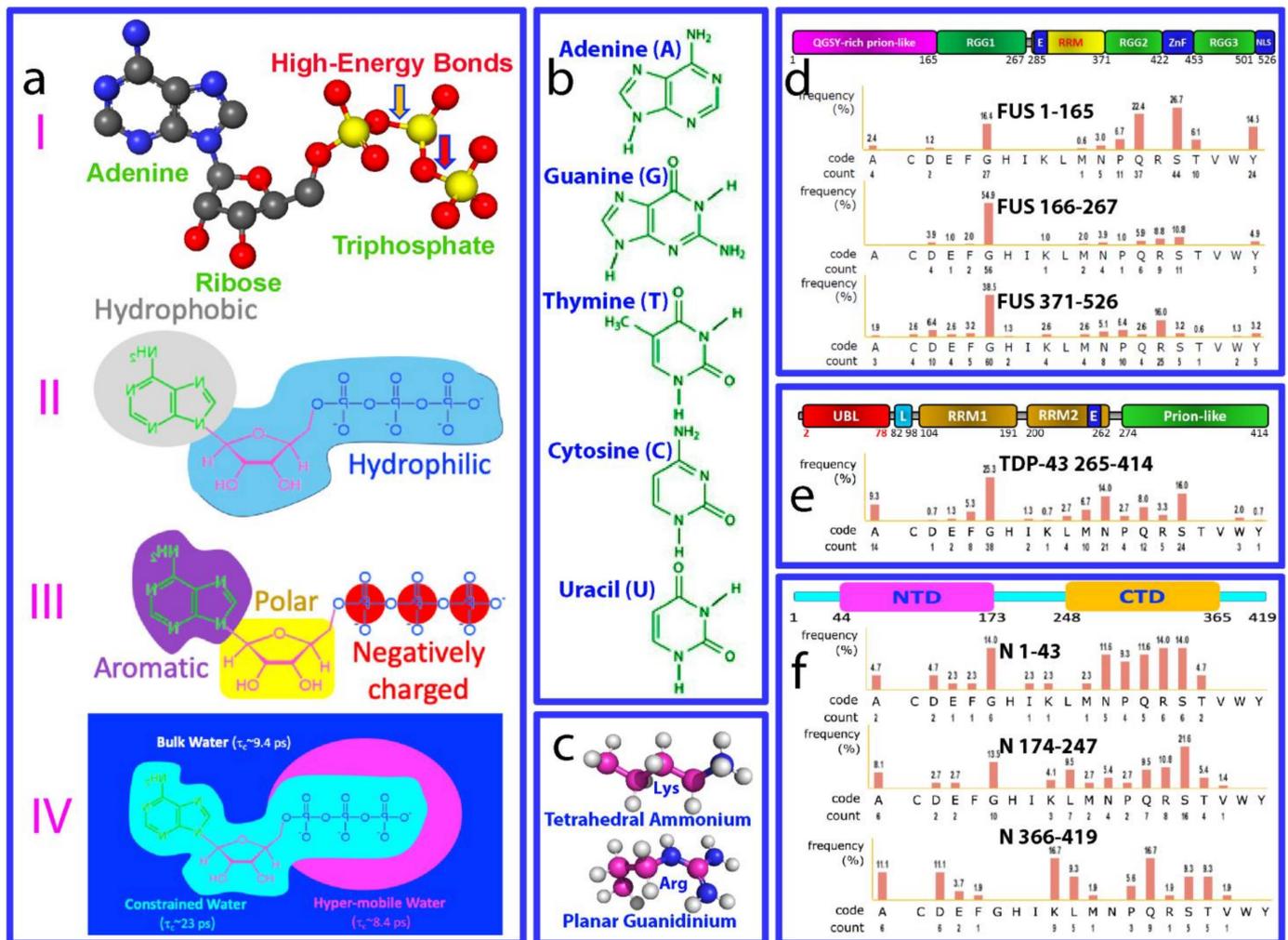
Figure 1. Key properties of ATP and nucleic acids for modulating LLPS of FUS, TDP-43, and SARS-CoV-2 nucleocapsid proteins. (**a**) ATP has unique structural properties and thus may act as (I) energy currency by hydrolysis of high-energy bonds; (II) a biological hydrotrope with the presence of hydrophobic adenine and hydrophilic ribose and triphosphate; (III) a bivalent binder via the aromatic purine ring and highly negatively charged triphosphate chain; and (IV) a hydration mediator resulting from its unique hydration structure previously derived from the results of microwave dielectric spectroscopy, modeled to contain "constrained water" with a dielectric relaxation time (τc) of ~23 ps, as well as "hyper-mobile water" with a τc of ~8.4 ps, even smaller than that of bulk water (9.4 ps). Chemical structures of five nitrogenous bases in DNA and RNA (**b**), and of Arg and Lys side chains (**c**). Amino acid compositions of intrinsically disordered regions (IDRs) of FUS (**d**), TDP-43 (**e**), and SARS-CoV-2 nucleocapsid (N) proteins (**f**).

Recently, liquid–liquid phase separation (LLPS) has been established as the common principle for forming membrane-less organelles (MLOs) and cellular compartments, including nucleoli, Cajal bodies, nuclear speckles, paraspeckles, histone locus bodies, nuclear gems, and promyelocytic leukemia (PML) bodies in the nucleus, as well as P-bodies, stress granules (SGs), and germ granules in the cytoplasm [7–11]. Noticeably, most, if not all, MLOs consist of the proteins abundant in intrinsically disordered regions (IDRs) and nucleic acids. Presently, the unique characteristics of IDRs are widely recognized as vital in facilitating LLPS through the establishment of dynamic and multivalent interactions.

Unlike folded proteins such as lysozyme, which phase-separate only at high concentrations (>mM) [12–14], IDR-rich proteins undergo phase separation at considerably low concentrations (~µM) [14–19]. As such, although MLOs also contain globular client proteins, these proteins do not appear to drive the formation of MLOs.

IDRs exist as dynamic ensembles of conformers due to low energy barriers, making them highly sensitive to external changes. The multivalency of binding sites within IDR-rich proteins enables simultaneous interactions with multiple partners, facilitating simultaneous interactions with multiple copies of themselves (homotypic phase separation) or other biomolecules such as nucleic acids (heterotypic phase separation). LLPS involves the reversible separation of a homogeneous solution into coexisting dense and dilute phases, with dynamic exchange constantly occurring between two phases [3–7,14–20]. However, dynamic liquid droplets may transform into solid-like structures, leading to the formation of cytoplasmic inclusions or amyloid fibrils associated with various diseases [7–22]. LLPS has also been suggested to have been a pivotal mechanism during the primordial evolution to concentrate and compartmentalize various rare molecules, which were presumably scarce in the aqueous pools of the early Earth, thus enabling the acquisition of the complexity essential for forming primordial cells [23].

By contrast, the exact effects and high-resolution mechanisms of ATP and nucleic acids on LLPS are only just starting to be understood. Previously, both RNA [24–27] and DNA [27–29] have been found to exert biphasic effects on the LLPS of IDR-rich proteins/peptides, namely, induction/enhancement at low concentrations but inhibition/dissolution at high concentrations. Based on polymer chemistry principles, this heterotypic LLPS was predicted to be driven by non-specific electrostatic interactions between the negatively charged phosphate backbone of RNA/DNA and the positively charged arginine/lysine residues of proteins [25]. ATP has been found to function as a hydrotrope, preventing/dissolving the LLPS, aggregation, and amyloid fibrillation of hydrophobic proteins [30,31], and later was suggested to serve more as a cosolute [32,33]. Beyond this, ATP has also been shown to exert biphasic effects on the LLPS of FUS [27,28], implying the existence of additional effects/mechanisms by which ATP modulates LLPS.

On a physiochemical level, ATP and nucleic acids can establish extremely intricate interactions with proteins. For example, the aromatic base rings of nucleotides can engage in at least three types of interactions with proteins: (1) π–π stacking interactions with the side chains of Arg and aromatic amino acids; (2) π–cation electrostatic interactions with the side chains of Arg and Lys (Figure 1c); and (3) hydrophobic interactions with hydrophobic patches on protein. Phosphate groups of nucleotides can also exert three electrostatic effects: (1) charge–charge interactions attracting and neutralizing positively charged groups such as the side chains of Arg and Lys residues, as well as repelling negatively charged groups; (2) hydrogen bonds forming particularly through oxygen atoms in the backbone phosphate; (3) screening effects, diminishing the strength of electrostatic interactions between charged molecules and groups [34–36]. Intriguingly, the triphosphate group of ATP appears to possess a unique hydration structure [37]. Indeed, recently it was found that, likely by interacting with protein hydration, ATP can antagonize the crowding-induced destabilization of the human eye-lens protein γS-crystalline at a ratio of 1:1 (ATP:crystalline) [38]. Most unexpectedly, ATP and triphosphate have the capacity to energy-independently induce protein folding with the highest efficiency known so far [39]. On the other hand, almost all cellular functions of nucleic acids manifest by interaction with proteins, including their involvement in LLPS [40–42]. So, a question of fundamental interest arises: is ATP able to mediate the interfaces between proteins and nucleic acids?

This review therefore aims to summarize recent studies on the mechanisms of ATP and nucleic acids and their interplay in modulating the LLPS of FUS, TDP-43, and SARS-CoV-2 nucleocapsid (N) proteins by NMR, which is powerful in elucidating high-resolution mechanisms of LLPS [43]. Human FUS and TDP-43 proteins, as well as SARS-CoV-2 viral nucleocapsid (N) proteins, all contain both folded nucleic-acid-binding domains and large IDRs (Figure 1d–f). FUS and TDP-43 are RNA-binding proteins (RBPs) with a unique

domain structure featuring intrinsically disordered regions (IDRs), including prion-like domains (PLDs) prone to aggregation and cellular toxicity [44–46]. They play crucial roles in transcription, splicing, RNA processing, DNA repair, and stress granule (SG) formation. Mutated FUS and TDP-43 can mislocalize and aggregate, contributing to neurodegenerative diseases like ALS, FTD, Alzheimer's, Parkinson's, and Huntington's. Therefore, understanding their LLPS and modulation is key to unraveling these diseases and has clinical implications for new diagnostics, prognostics, and drug development. On the other hand, the SARS-CoV-2 nucleocapsid (N) protein is one of four viral structural proteins which not only packages genomic RNA (gRNA), but also suppresses the immune system and manipulates host cell machinery to enhance viral infection. It has low mutation rates, with 91% identity to SARS-CoV-1, making it a key candidate for drug development compared to the spike protein [47–49]. LLPS has recently been identified as a mechanism underlying its diverse functions, which are dependent on binding to viral/host-cell nucleic acids. Despite limited understanding, its interaction with gRNA is crucial for RNA genome packaging, a challenge given SARS-CoV-2's large genome (~30 kb). Therefore, any small molecules modulating N protein–nucleic acid interactions could disrupt viral life cycle steps, potentially becoming anti-SARS-CoV-2 drugs [50–56]. Furthermore, accumulating evidence indicates that, via LLPS, proteins and nucleic acids form condensates/compartments in infected cells. This is also seen in other viruses, including measles [57], HIV [58], influenza [59], Hendra [60], vesicular stomatitis [61], Borna disease [62], and rabies [63]. These structures are known as viral factories, viroplasms, or viral replication centers [64]. They play crucial roles in virus–host interactions and the viral life cycle, encompassing genome replication, gene expression, and antiviral immune response [55–64]. Understanding the pivotal roles of viral LLPS not only reveals previously unknown principles of host–virus interactions and the viral life cycle, but also offers a new direction for the development of antiviral strategies and drugs, an area that remains largely unexplored.

This review primarily focuses on illustrating the biophysical principles and forces that govern how ATP and nucleic acids interplay to modulate LLPS. Briefly, ATP binds to pockets within nucleic-acid-binding interfaces of folded domains adopting different structural folds. Strikingly, both ATP and nucleic acids have been shown to mainly interact with Arg/Lys residues within IDRs. In this context, in addition to exerting extremely complex effects on proteins, ATP also interplays with nucleic acids to interact with proteins, including to modulate LLPS by competitively binding to overlapping pockets of folded nucleic-acid-binding domains and Arg/Lys residues within IDRs of human and viral proteins. Therefore, ATP appears to not only act as a primordial molecule to control protein homeostasis—including inducing folding, preventing aggregation, and mediating LLPS, essential for forming primordial cells—but has also shaped protein–nucleic acid interfaces since the dawn of life, the probiotic evolution.

2. ATP Binds Folded Nucleic-Acid-Binding Domains

2.1. ATP Binds the Nucleic-Acid-Binding Pocket of the FUS RRM Domain

FUS, a 526-residue protein, is composed of an N-terminal low-sequence complexity (LC) domain (residues 1–267), followed by an RNA-recognition motif (RRM) domain (residues 285–370) and a C-terminal LC domain enriched in RG/RGG motifs (residues 371–526) (Figure 1d). FUS is prone to aggregation, which is extensively linked to various neurodegenerative diseases, including ALS and FTD [19,21,22,44–46]. The FUS RRM exhibits versatile nucleic-acid-binding capacities in both sequence-specific and sequence-independent manners [65–70]. It stands out as one of the most prevalent domains in eukaryotes, commonly found in heterogeneous nuclear ribonucleo-proteins (hnRNPs) that engage in direct DNA and RNA interactions through one or several RRM domains. Structurally, the FUS RRM shares an overall fold common to other RRMs, evident in structures of its free state [65] or when bound to RNA [66,67]. This fold comprises a four-stranded β-sheet and two perpendicular α-helices. Notably, the FUS RRM possesses a unique and extra-long positively charged "KK" loop crucial for nucleic acid binding (I of

Figure 2a). Intriguingly, the FUS RRM domain displays heightened backbone dynamics and relatively low thermodynamic stability with a melting temperature (Tm) of only 52 °C [69]. In particular, the FUS RRM is characteristic of irreversible unfolding and capable of spontaneous self-assembly into amyloid fibrils, which is essential for cytotoxicity [68–70].
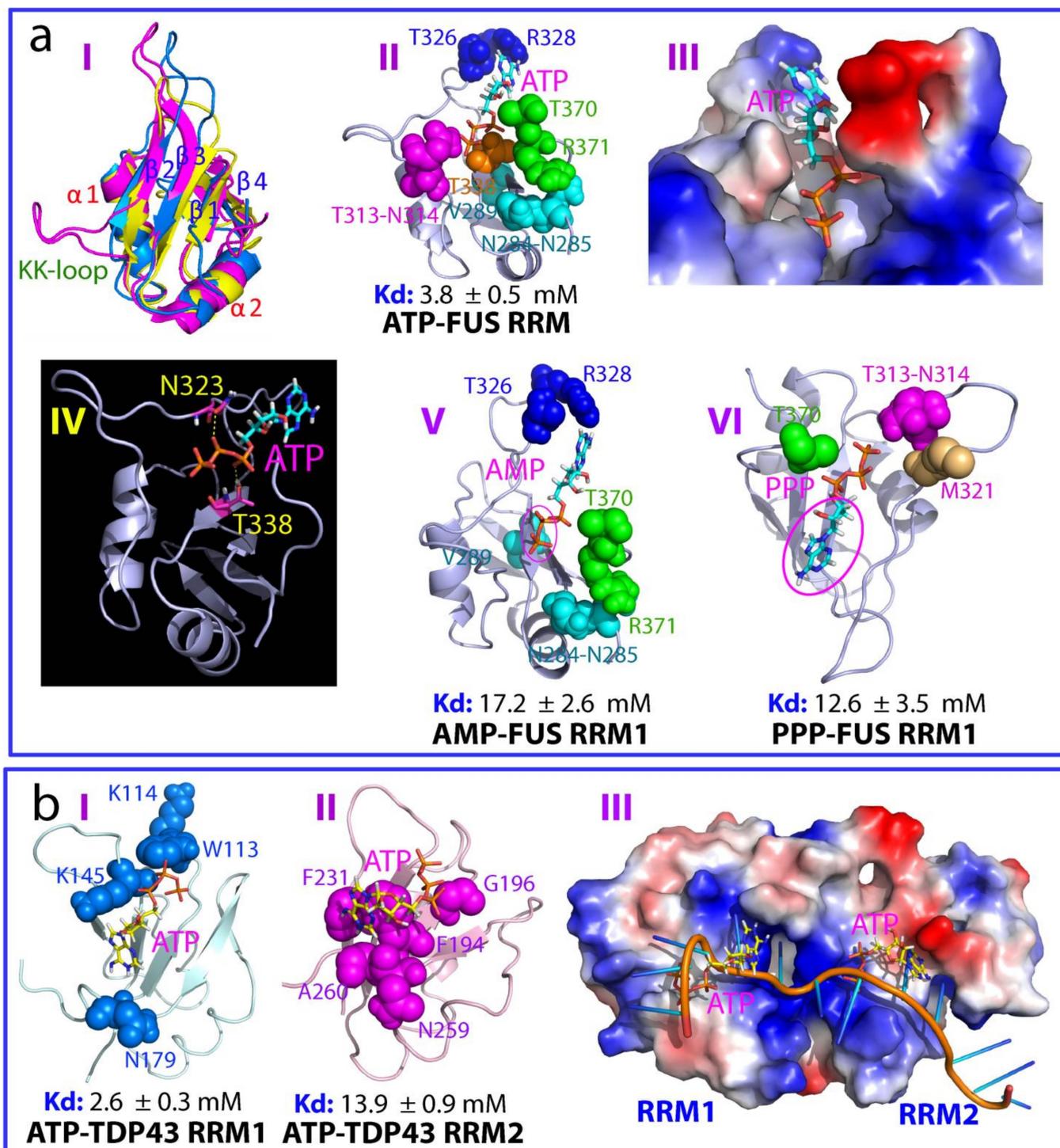
Figure 2. ATP binds nucleic-acid-binding pockets of FUS and TDP-43 RRM domains. (**a**) (I) Superimposition of FUS RRM (purple), TDP-43 RRM1 (blue), and RRM2 (yellow) with secondary structures labeled. (II)–(VI) ATP/AMP/PPP complexes of FUS RRM domain. (**b**) ATP complexes of TDP-43 RRM1 and RRM2 domains.

In a recent investigation [70], ATP was found to bind the FUS RRM domain with a dissociation constant (Kd) of 3.8 mM (II of Figure 2a). This binding occurs at a pocket overlapping with the physiological RNA/ssDNA binding interface. Structural analysis of the ATP-RRM complex revealed that both the adenine aromatic ring and triphosphate chain engage in crucial interactions with specific residues of the FUS RRM. Specifically, ATP occupies a pocket formed by 10 residues distributed across the entire sequence (II of Figure 2a). The aromatic ring of ATP resides in a relatively hydrophobic surface pocket within the FUS RRM domain, making direct contact with the side chain of Arg328, presumably through a π–cation interaction (III of Figure 2a). Meanwhile, the triphosphate chain of ATP is nestled within a pocket featuring a positively charged surface. Notably, the oxygen atoms within the triphosphate chain form two hydrogen bonds with the side chain protons of Asn323 and Thr338, respectively (VI of Figure 2a).

Subsequent NMR titrations with AMP and triphosphate (PPP) further validated the essential roles of both the adenine and triphosphate components of ATP. AMP and PPP were observed to bind to similar pockets, but their binding affinities decreased notably, resulting in Kd values of 17.2 and 12.6 mM, respectively for AMP (V of Figure 2a) and PPP (VI of Figure 2a). On the other hand, ATP, even up to 20 mM, did not influence the thermodynamic stability of the FUS RRM domain with the same Tm of 52 °C. By contrast, under the same buffer conditions, ATP at 3 mM was sufficient to inhibit RRM fibrillization, a pivotal process associated with the acquisition of cytotoxicity linked to ALS and FTD [68]. As shown in II of Figure 2a, ATP binds the pocket which is formed by residues over the entire sequence. As such, ATP binding has been proposed to reduce the structural opening of the FUS RRM, which was shown to initiate fibrillation, thus increasing the kinetic barrier for amyloid fibrillation [69].

2.2. ATP Binds the Nucleic-Acid-Binding Pocket of TDP-43 RRM1 and RRM2

TAR-DNA-binding protein-43 (TDP-43) was initially identified to bind TAR DNA in HIV to inhibit its transcription [71]. TDP-43 is a 414-residue multifunctional protein within the hnRNP family. It exhibits a high affinity for single-stranded DNA and RNA, engaging with over 6000 RNA species, and holds pivotal roles in governing RNA metabolism, encompassing transcription, splicing, and mRNA stability [14,21,22,45]. TDP-43 comprises several distinct domains: the N-terminal domain (NTD) spanning residues 1–80 [72,73], two RNA recognition motif (RRM) domains spanning residues 104–262 [74–76], and a C-terminal PLD domain abundant in Gln/Asn/Ser/Gly over residues 265–414 (Figure 1e). Unlike the FUS RRM, TDP-43's tandem RRM domains lack the "KK-loop" (I of Figure 2a).

Recently, it was demonstrated that ATP also binds to the pockets within the conserved nucleic-acid-binding interfaces of both RRM1 and RRM2 of TDP-43 [77–79], displaying notably distinct affinities (Kd of 2.6 mM for RRM1 and 13.9 mM for RRM2). Within the ATP-RRM1 complex, the triphosphate chain of ATP engages in close interactions with the side chains of K145 and K114 residues, while the ATP aromatic ring fits into a positively charged pocket formed by K136, K137, K176, and K181. Specifically, the aromatic ring of ATP establishes close contacts with the side chains of K176 and K181 via π–cation interactions. Moreover, the three phosphate oxyanions of ATP create five hydrogen bonds with the backbone atoms of S144 and K145 [77]. In contrast, within the ATP-RRM2 complex (II of Figure 2b), the aromatic ring of ATP closely interacts with the aromatic rings of F194 and F231 through π–π interactions (II of Figure 2b). Interestingly, unlike the ATP-RRM1 complex, the ATP-binding pocket on RRM2 appears to exhibit a slightly negative charge. Only one hydrogen bond forms between the β-phosphate oxyanions of ATP and the backbone atom of V195. Strikingly, the structures of TDP-43 RRM1 and RRM2 in complex with ATP closely resemble those observed in their RNA complexes, underscoring ATP's occupation of the conserved pockets within RRM1 and RRM2 for nucleic acid binding (III of Figure 2b). Indeed, further studies have shown that the mutation of crucial nucleic-acid-binding residues disrupts ATP binding. Interestingly, the ALS-linked D169G mutation in TDP-43 RRM1, which is located on the back side of both the ATP and nucleic-acid-binding

interfaces, also largely reduces ATP binding affinity to a Kd of 9.3 mM [78]. However, the ATP-binding pockets of the FUS and TDP-43 RRM domains are much smaller compared to their nucleic-acid-binding interfaces. For instance, while the "KK-loop" is crucial for the nucleic acid binding of the FUS RRM, it does not directly participate in ATP binding, thus resulting in a much lower ATP-binding affinity than for nucleic acids.

ATP exhibits varying effects on the thermal stability of the TDP-43 RRM1 and RRM2 domains and impedes amyloid fibrillation [77–79]. Briefly, TDP-43 RRM1 and RRM2 have Tm values of 57 and 59 °C, respectively. Interestingly, ATP at 15 mM increased the Tm value of RRM1 to 60 °C, but showed no alteration in the Tm of RRM2. On the other hand, under the same conditions, ATP at 3 mM was sufficient to inhibit the formation of amyloid fibril. As such, ATP appears to inhibit the amyloid fibrillation of TDP-43 RRM domains both thermodynamically and kinetically.

Moreover, ATP has also been characterized to bind the nucleic acid binding interface of the RRM domain of human cold-inducible RNA-binding protein (CIRBP) with a Kd of 2.9 $\pm$ 1.4 mM [80]. So, an intriguing question emerges: can ATP bind to nucleic-acid-binding domains lacking the RRM fold? Recent findings reveal that ATP indeed binds the N-terminal acidic domain (AcD) of the 623-residue SYNCRIP, exhibiting a Kd of 3.1 mM (Figure 3a). The AcD lacks any sequence and structural homology for RRMs, but functions as a cryptic RNA-binding domain adopting an all-helix fold [81]. In the complex, ATP binds to an AcD pocket formed by residues situated on helices 3 and 5 (I of Figure 3a), which constitute a significant interface for miRNA binding [82,83]. Notably, the aromatic purine ring of ATP finds placement between the side chains of Arg38 and Lys78, establishing two π–cation interactions, while the triphosphate chain directly interacts with the highly positively charged AcD surface (II of Figure 3a). Additionally, an oxyanion from the β-phosphate of ATP forms two hydrogen bonds, one with the backbone atom of Lys78 and the other with the side chain atom of Gln 82. Despite the overall negatively charged electrostatic surface of AcD, ATP can still fit into the highly positively charged cavity within AcD, which coincidentally serves as a binding site for miRNA (II of Figure 3a).

2.3. ATP Binds NTD and CTD of SARS-CoV-2 Nucleocapsid Protein

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) belongs to the coronavirus family, possessing a genomic RNA (gRNA) of approximately 30 kb, encapsulated by the nucleocapsid (N) protein within a membrane-enveloped virion. This virus has instigated the ongoing pandemic, resulting in >772 million infections and >6.99 million deaths [47]. Among its four structural proteins, the nucleocapsid (N) protein stands out as the sole protein responsible not only for packaging gRNA, but also modulating the host cell's immune response and cellular mechanisms to facilitate viral infection and replication [48,49]. Notably, LLPS has recently been identified as a key process governing the diverse functions of the SARS-CoV-2 N protein [50–55]. Crucially, most of the N protein's functions, including LLPS, are reliant on its interaction with a wide array of viral and host-cell nucleic acids, encompassing various single- and double-stranded RNA/DNA sequences of diverse origins in both sequence-dependent and sequence-independent manners.

The SARS-CoV-2 N protein spans 419 residues and comprises distinct domains (folded N-terminal domain (NTD) and C-terminal domain (CTD)) alongside three large intrinsically disordered regions (IDRs) enriched with Arg/Lys residues (Figure 1f). Studies have affirmed the NTD as a nucleic-acid-binding domain (RBD), crucial for engaging with diverse RNA and DNA in both specific and nonspecific manners [52–55]. Previously, CTD has been recognized for driving dimerization/oligomerization to facilitate the formation of high-order structures [49–55]. Very recently, CTD was also shown to be a nucleic-acid-binding domain [55,84]. Surprisingly, despite the absence of ATP in viruses [5], recent findings have revealed that ATP not only binds to NTD with a Kd of 3.3 mM [52], but also to CTD with a Kd of 1.5 mM [84].
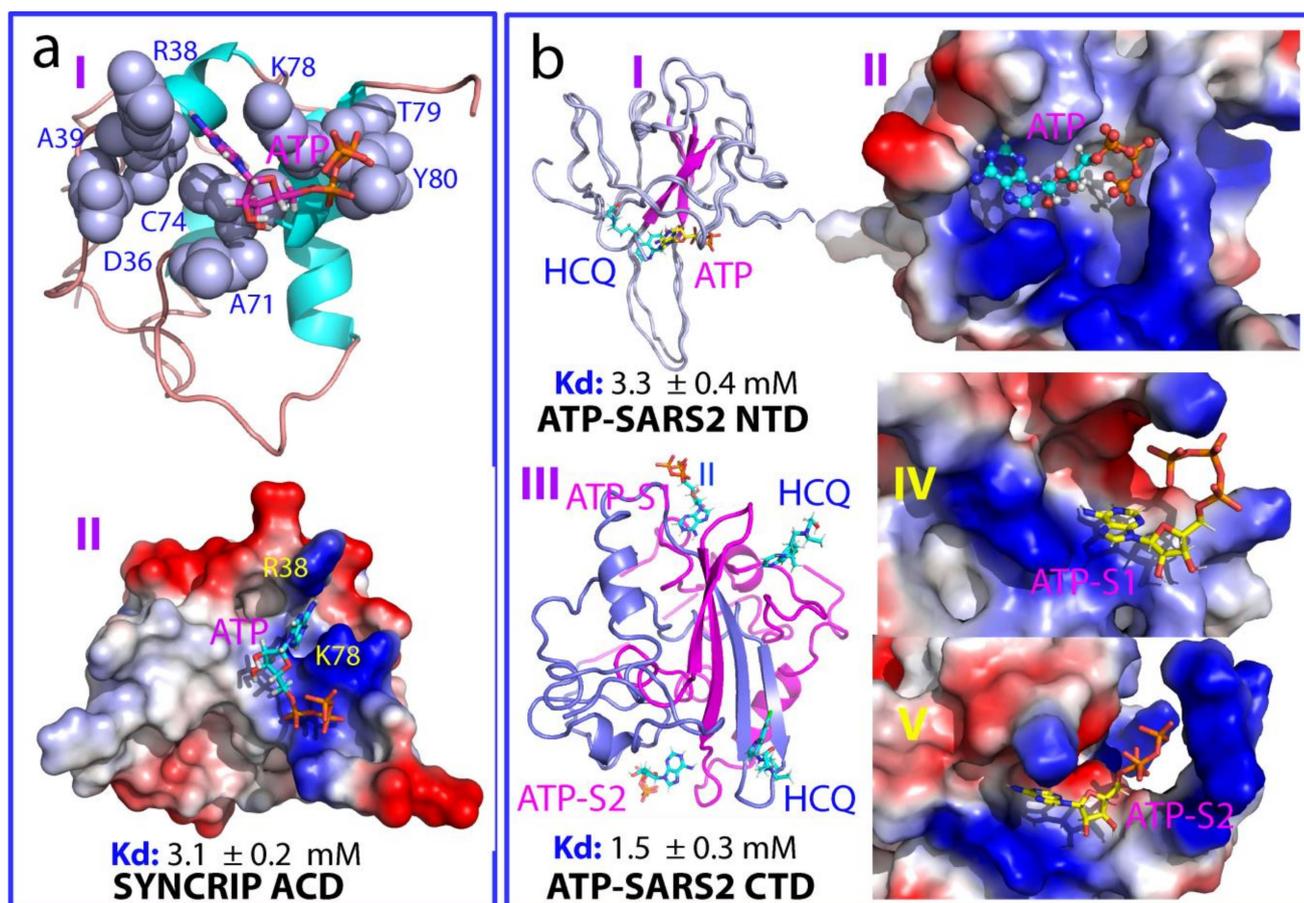
Figure 3. ATP binds nucleic-acid-binding pockets of folded domains beyond RRM. (**a**) ATP complex of the acidic domain (AcD) of Syncrip. (**b**) ATP complexes of N-terminal domain (NTD) and C-terminal domain (CTD) of SARS-CoV-2 nucleocapsid protein, both of which are bound with hydroxychloroquine (HCQ). CTD is a homo-dimer with two monomers colored in blue and purple, respectively.

As shown in Figure 3b, the ATP-NTD complex illustrates ATP occupying a pocket within the extensively positively charged surface (I of Figure 3b). The purine ring of ATP engages in π–cation interactions with multiple Arg residues, while the oxyanions of the β-phosphate from the triphosphate chain form three hydrogen bonds with NTD residues: two with Asn8 and one with Thr9. Recent NMR studies have revealed the structures of the NTD of the SARS-CoV-2 N protein in complex with both single-stranded RNA (ssRNA) containing the sequence UCUCUAAACG and double-stranded RNA (dsRNA) with the sequence CACUGAC [53]. Intriguingly, the comparison of these structures with the ATP-NTD complex highlights that ATP indeed occupies a pocket within the large positively charged surface used by NTD to bind various ssRNA and dsRNA (II of Figure 3b). Consequently, the NTD of the SARS-CoV-2 N protein emerges as the first viral fold capable of binding ATP with a Kd of approximately mM, in addition to the RRM and AcD folds [55].

Unexpectedly, as illustrated in III of Figure 3b, ATP can also bind two distinct binding pockets of CTD comprising 12 residues [55,84]. These pockets span the dimerization interface, consisting of residues from both monomers, notably displaying a highly positive charge (IV and V of Figure 3b). Within the complex, the purine rings of ATP closely interact with the highly positive surface formed by Lys256, Arg259, and Arg262 from one monomer, along with Lys338 from another monomer, establishing π–cation and π–π interactions between the ATP purine ring and the Arg/Lys side chains. Additionally, the triphosphate groups engage in electrostatic interactions with the predominantly positive surface, constituted by Lys342. Compared to previously identified ATP-binding pockets in

other domains capable of binding ATP at mM levels, the CTD pocket features a notably high density of positive charges, characterized by a cluster of Arg/Lys residues. This unique feature potentially contributes to its high binding affinity for ATP. Interestingly, the high conservation of the 12 residues across SARS-CoV-2 variants strongly suggests that ATP binding to CTD might play an indispensable role in completing its life cycle within the host cell, thus representing a key target for the design of anti-SARS-CoV-2 drugs [55,84].

In summary, unlike the weak and non-specific effects found in folded proteins lacking nucleic-acid-binding ability, such as human γ-crystalline [38], profilin-1 [70], and others [85,86], ATP generally binds to the pockets within the nucleic-acid-binding interfaces of folded domains of diverse folds with a Kd of ~mM. Structural analysis of the ATP-binding pockets in these domains elucidates the general determinants governing their interactions: (1) Both the adenine aromatic ring and triphosphate group are essential for binding specificity and affinity. (2) π–cation/π–π interactions of the adenine aromatic ring with Arg/Lys appear to hold greater strength than the π–π interactions of the adenine aromatic ring with aromatic residues. (3) ATP exhibits a higher affinity for Arg compared to Lys, as the adenine aromatic ring of ATP is capable of forming both π–cation/π–π interactions with Arg, whereas it can only engage in a π–cation interactions with Lys. (4) The number of hydrogen bonds formed between the triphosphate chain and protein residues emerges as a critical factor influencing binding affinity. (5) ATP appears to be capable of universally binding to folded nucleic-acid-binding domains regardless of their origin, as evident from its binding to the NTD and CTD of the viral N protein of SARS-CoV-2 that lacks the ability to produce ATP itself.

Intriguingly, unlike ATP binding to kinases, which may have Kd values of ~nM [87,88], ATP only binds folded nucleic-acid-binding domains with Kd values of ~mM. Theoretically, binding affinity and specificity are two crucial factors that govern how proteins interact with other molecules. Binding affinity refers to the strength of the interaction, which signifies how tightly the two molecules bind and the degree of their complex formation. On the other hand, binding specificity describes the selectivity of the protein towards its ligand, which reflects how well the protein distinguishes its specific ligand from structurally similar molecules. Although binding affinity and specificity are often interrelated, high affinity does not always guarantee high specificity, while a protein with a lower affinity can still exhibit high specificity if its binding site is highly selective. In this context, despite its very low affinity, ATP binding to folded nucleic-acid-binding domains is generally expected to be specific for several reasons: (1) ATP-binding pockets are located within nucleic-acid-binding interfaces formed by many residues distributed over entire sequences, which can only be formed upon successful folding; (2) ATP-binding curves are all saturable; (3) the replacement of ATP with AMP or triphosphate would dramatically reduce the binding affinity; (4) the mutations of binding residues such as those of the TDP-43 RRM domain disrupt the binding events.

3. ATP and Nucleic Acids Interplay in Modulating LLPS

3.1. ATP and Nucleic Acids Interplay in Modulating the LLPS of FUS IDRs

Notably, over 70% of the FUS sequence is composed of IDRs, characterized by the absence of bulky hydrophobic residues Leu, Ile, and Val (Figure 1d). These FUS IDRs can be categorized into two distinct types: the polar/aromatic residue-rich PLD (1–165), and the RG-/RGG-rich regions spanning from 166 to 267 and the C-terminal domain (371–526), which are enriched in Gly and Arg/Lys. Within the sequence range of 1 to 165, the QGSY-rich prion-like domain (PLD) exhibits significant sequence identity with the N-terminal yeast prion domain of Sup35 [89], comprising 24 Tyr residues but no Arg/Lys. In vitro systematic dissection studies have revealed that among the FUS domains, only the FUS NTD (1–267) demonstrates autonomous phase separation at low concentrations. Conversely, the isolated RRM or C-terminal domain (CTD) lack this inherent capacity [27,69].

The FUS NTD (1–267), comprising a QGSY-rich PLD and RGG1 (Figure 3a), plays a key role in promoting self-assembly into liquid-like granules, with the potential to transition

into hydrogels and solid aggregates [90,91]. By contrast, the FUS PLD was observed to undergo phase separation only at higher concentrations or in the presence of dextran (I of Figure 3a), acting as a molecular crowding agent [16,27,92]. Notably, NMR investigations revealed a fascinating discovery: the FUS PLD, which lacks autonomous phase separation, unexpectedly acquires this capability at remarkably low concentrations (~1 μM) upon linkage to RGG1 to form an NTD [27,28]. Additionally, it was observed that the isolated FUS PLD could also undergo phase separation at low concentrations upon adding the isolated RGG3 region [93]. These findings strongly support the role of π–π and/or π–cation interactions involving aromatic amino acids and Arg/Lys residues as significant contributors to driving the LLPS of FUS.

A comprehensive investigation aiming to delineate the molecular interactions governing the LLPS of FUS and its domains, alongside the impact of ATP, RNA, and specific/nonspecific ssDNA sequences, was conducted [27]. The examined FUS domains encompassed PLD (1–165), NTD (1–267), and CTD (371–526). The RNA sequence used was UAGUUUG-GUGAU, while the ssDNAs employed were telomeric ssDNA (TssDNA) with a sequence of $(TTAGGG)_4$, forming a G-quadruplex, and $(TTTTTT)_4$ (T24), lacking defined secondary or tertiary structures. Notably, the FUS PLD, incapable of phase separation at low concentrations, did not exhibit induced phase separation in the presence of ATP, RNA, or ssDNA. Corresponding NMR studies highlight the absence of substantial interactions between these molecules and the FUS PLD [27].

Conversely, as detailed previously [27,28], the FUS NTD (1–267), displaying inherent phase separation at very low concentrations, experienced a monotonic dissolution of LLPS upon the gradual addition of ATP, RNA, and two ssDNA molecules (Figure 4a). Further NMR investigations unveiled the mechanism behind this dissolution, illustrating that ATP and nucleic acids achieve this effect by binding to the same set of residues within the RGG1 region. Briefly, the LLPS of the FUS NTD predominantly hinges on π–π and/or π–cation interactions between the aromatic amino acids in the PLD and the Arg/Lys residues in RGG1. In this framework, nucleic acids with multiple base aromatic rings engage in analogous interactions with the Arg/Lys residues in RGG1, and their binding affinity is governed by binding multivalency [94]. Consequently, the base aromatic rings of nucleic acids competitively displace the aromatic rings of the PLD residues from binding to Arg/Lys residues in RGG1, leading to the dissolution of the LLPS of the FUS NTD (III of Figure 4a). Through a similar mechanism, ATP utilizes its adenine aromatic ring to establish π–π and/or π–cation interactions with the Arg/Lys residues in RGG1, albeit with notably lower affinity compared to nucleic acids. At sufficiently high concentrations, the adenine aromatic ring of ATP clusters around the Arg/Lys residues in RGG1, competitively disrupting their interactions with aromatic residues within PLD, thereby causing the dissolution of the FUS NTD's LLPS (III′ of Figure 4a). However, in contrast to nucleic acids, the triphosphate group of ATP exhibits strong interactions with water molecules. Consequently, NTD molecules bound with ATP self-assemble into large and dynamic oligomers, involving the aromatic/hydrophobic PLD residues despite the weak ATP binding (III′ of Figure 4a). Upon introducing ssDNA molecules, which possess a higher affinity for Arg/Lys residues, they competitively displace ATP molecules clustered around these residues, resulting in the disassembly of ATP-NTD oligomers. Consequently, the NTD molecules shift to binding with ssDNA (IV of Figure 4a).

The FUS CTD (371–526) does not inherently undergo phase separation under varied conditions (I of Figure 4b). However, the introduction of ATP, RNA, TssDNA, or T24 induces phase separation, followed by subsequent dissolution. For RNA/DNA, at lower concentrations, they behave as multivalent binders, utilizing their base aromatic rings and phosphate groups to bind the side chains of Arg/Lys. This binding initiates the formation of large and dynamic RNA/DNA-CTD complexes, manifesting as liquid droplets (II of Figure 4b). Yet, at higher concentrations, excessive RNA/DNA binding disrupts these large complexes, leading to the dissolution of the droplets (III of Figure 4b). Similarly, at lower concentrations, ATP acts as a bivalent binder, using the aromatic ring of its adenine

group to bind to the side chains of Arg/Lys as well as its triphosphate chain to engage with Arg/Lys. Consequently, the formation of large and dynamic ATP-CTD complexes will manifest as liquid droplets (II′ of Figure 4b). However, at high concentrations, excessive ATP binding disrupts these large complexes, leading to the dissolution of the droplets (III′ of Figure 4b).
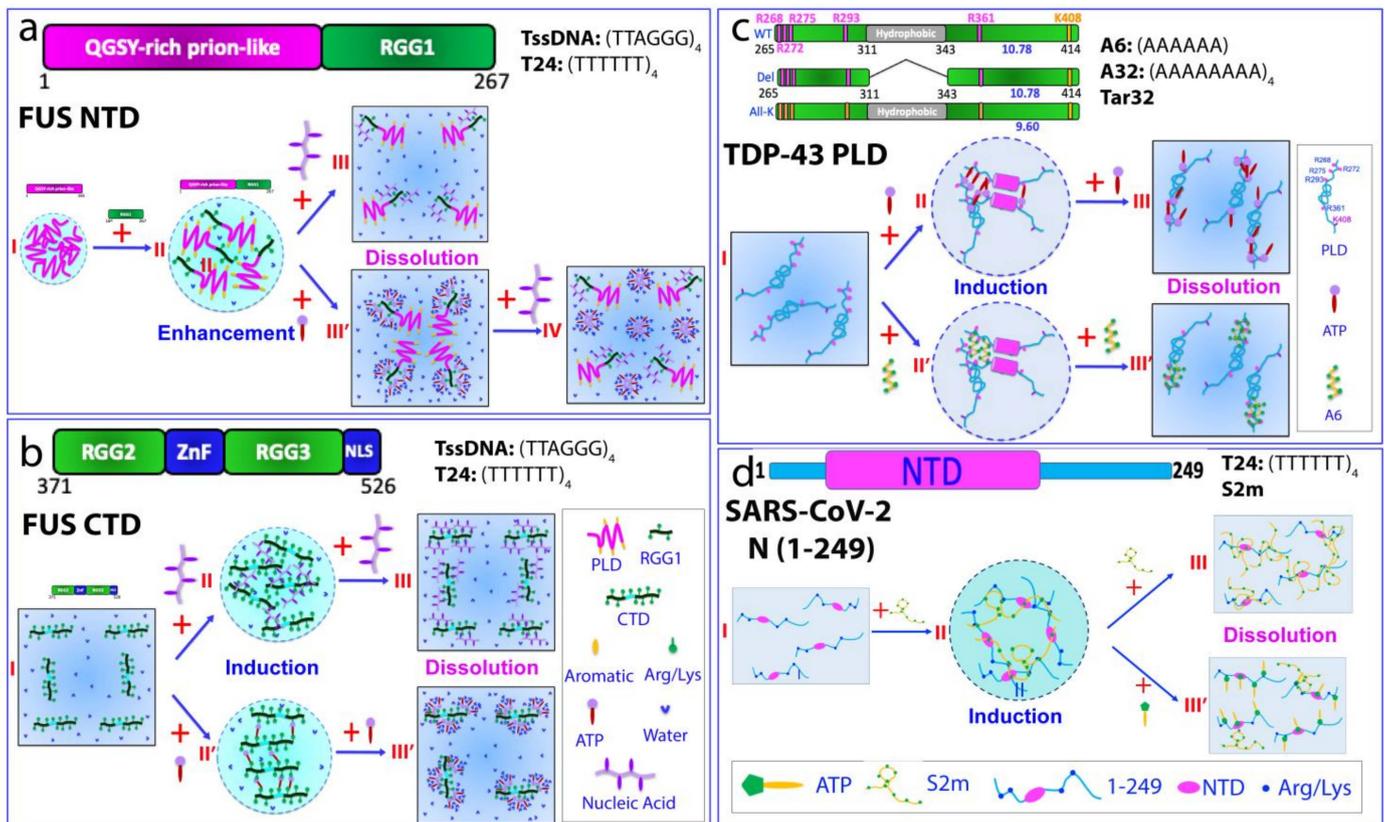
Figure 4. Mechanisms of interplay between ATP and nucleic acids to modulate LLPS. Speculative models to illustrate how ATP and nucleic acids interplay to modulate LLPS: (**a**) FUS NTD; (**b**) FUS CTD; (**c**) TDP-43 PLD and its mutants including the ones with residues 312–342 deleted and with all five Arg mutated to Lys; (**d**) SARS-CoV-2 N (1–249).

Indeed, a comprehensive analysis of NMR data on the CTD revealed strikingly similar patterns of chemical shift differences (CSD) induced by ATP, TssDNA, and T24 [27]. Notably, a significant number of residues, including 25 Arg and 4 Lys residues, displayed pronounced shifts in their HSQC peaks. This observation strongly suggests the specific binding of ATP and ssDNA to these Arg/Lys residues. Intriguingly, the binding of ATP and ssDNA to Arg/Lys residues is not overly reliant on the specific sequence housing these Arg/Lys residues, as long as the region remains disordered. These findings provide compelling evidence that Arg/Lys residues within IDRs play a pivotal role in the induction and dissolution of phase separation by ATP and nucleic acids in the FUS CTD [27].

3.2. How ATP and Nucleic Acids Interplay to Modulate the LLPS of the TDP-43 PLD

TDP-43 encompasses a C-terminal PLD that is rich in Gln/Asn/Ser/Gly residues spanning residues 265–414 (Figure 1e), houses nearly all known ALS-related mutations, and is implicated in the prion-like propagation associated with ALS. Notably, in contrast to prototypical PLDs like the FUS PLD (1–165), which is abundant in polar and aromatic residues (Figure 1d), the TDP43 PLD stands out for its inclusion of 10 evolutionarily conserved methionine residues, as well as a hydrophobic region spanning residues 311–343, which adopt

a nascent helical conformation capable of interacting with membranes and transforming into β-like conformations and amyloid fibrils under neutral pH conditions [95–98].

Extensive investigations have delineated the pivotal role of the evolutionarily conserved hydrophobic region in driving the LLPS of the PLD, as evidenced by the observation that the deletion of residues 311–343 abolishes its ability to undergo phase separation under various in vitro conditions [99,100]. A study at pH 6.1 proposed that the LLPS of the TDP-43 PLD primarily occurs via the intermolecular self-interaction of residues 321–340, promoting a helical propensity [95]. Moreover, the arrangements of hydrophobic 'sticker' residues separated by flexible linkers, electrostatic interactions, and aromatic residues adjacent to Gly or Ser have also been suggested as crucial factors [101,102]. However, other studies have indicated that at pH 6.8, the nascent helical conformation of residues 321–340 transforms into a cross-β oligomer, facilitating biologically relevant LLPS and further amyloidosis [96,97]. Evidently, the oxidation of methionine residues, enhancing helical conformation, impedes cross-β assembly and LLPS [97]. Indeed, pH was shown to be a key factor for α-to-β transformation, LLPS, and the subsequent amyloidosis of the TDP-43 PLD [96–98].

Although the LLPS of the FUS CTD was shown to be biphasically modulated by the binding of ATP and nucleic acids to its Arg/Lys residues (Figure 4b), the high abundance of Arg/Lys residues within the FUS CTD (25 Arg and 4 Lys) makes it extremely challenging to determine the high-resolution mechanism for ATP and RNA/DNA specifically to bind these Arg/Lys residues and modulate LLPS through NMR and site-directed mutagenesis. In contrast, the TDP-43 PLD (265–414) harbors a limited number of Arg and Lys residues— Arg268, Arg272, Arg275, Arg293, Arg361, and Lys408 (Figure 4c)—offering an excellent model for investigating potential LLPS modulation by ATP and nucleic acids, as well as whether such modulation occurs via non-specific electrostatic/salt effects of ATP and nucleic acids or through specific binding to Arg/Lys residues.

Recently, two studies comprehensively examined the impact of ATP and ssDNA on the LLPS of the TDP-43 PLD and probed residue-specific interactions using NMR [99,100]. These investigations scrutinized the effects of ATP and ssDNA at different sequences and lengths (Tar32, A32, and A6) on LLPS dynamics, assessing their interactions with both wild-type and mutated PLDs. These mutations included PLD Δ(311–343), with residues 311–343 deleted but an unaltered pI, and All-K PLD, with all five Arg residues substituted by Lys, resulting in a pI of 9.6 (Figure 4c).

Notably, ATP exhibited a biphasic modulation of the TDP-43 PLD's LLPS, inducing phase separation at low concentrations followed by dissolution at higher concentrations [99]. By contrast, PLD Δ(311–343), which is unable to phase separate under various conditions, failed to undergo LLPS or aggregation upon the addition of ATP, even at ratios as high as 1:1500 (PLD:ATP). However, ATP induced significant NMR HSQC peak shifts in the N-terminal residues of PLD Δ(311–343), resembling patterns observed in the WT PLD. In contrast, the All-K PLD showed a minimal induction of LLPS upon ATP addition, and further ATP additions led to gradual droplet dissolution. Additionally, the ATP-triggered HSQC peak shifts in the All-K PLD resembled WT PLD patterns. Nevertheless, while the WT PLD exhibited peak saturation at a ratio of 1:100 (PLD:ATP), the All-K PLD's shifts remained unsaturated, even at a ratio of 1:500 [99]. These findings unequivocally demonstrate the "residue-specific/selective" binding of ATP to Arg or Lys residues across different TDP-43 PLD mutants, showcasing notably higher affinity to Arg than Lys. Furthermore, unlike the FUS CTD, which relies solely on ATP for LLPS modulation due to its abundance of Arg/Lys residues, the TDP-43 PLD's response to ATP-induced LLPS is contingent on the unique presence of the 311–343 region, despite ATP binding the same residues in both WT and Δ(311–343) PLD variants.

Hence, owing to its low binding affinity, the mechanisms underlying ATP's role in modulating LLPS seem highly contingent on the specific context. In the case of the 156-residue FUS CTD, featuring 25 Arg and 4 Lys residues, ATP's bivalent binding to these residues directly induces LLPS by generating sizable and dynamic complexes at low ATP

concentrations, followed by dissolution at higher ATP concentrations due to excessive binding (Figure 4b). Conversely, for the 150-residue TDP-43 PLD with only five Arg residues and one Lys residue, ATP binding alone seems insufficient to drive LLPS. Thus, it necessitates the coordination of other contributing factors, particularly the oligomerization of the distinctive hydrophobic region, to induce LLPS (Figure 4c).

In a follow-up study [100], three distinct ssDNA sequences were demonstrated to biphasically modulate the LLPS of a TDP-43 WT PLD, but with varying effectiveness. A6 induced maximum LLPS at a ratio of 1:3, with complete dissolution observed at 1:5. Conversely, the stepwise addition of Tar32 led to peak LLPS at a ratio of 1:0.25, followed by dissolution at 1:1 (PLD:Tar32). A32 also triggered and dissolved LLPS in a pattern similar to Tar32. NMR analysis revealed that A6 caused shifts in a limited set of HSQC peaks, resembling the pattern induced by ATP. In contrast, Tar32 induced shifts across a broad spectrum of HSQC peaks throughout the PLD sequence, and A32 exhibited a shift pattern similar to that of Tar32. Notably, in PLD Δ(311–343), A6 caused shifts of a limited set of HSQC peaks, akin to those induced by ATP. However, even at a ratio of 1:5 (PLD:A6), A6 failed to induce LLPS. Conversely, Tar32 triggered LLPS, reaching its peak at 1:0.25, with complete dissolution observed at 1:1 (PLD:Tar32). Intriguingly, A32 displayed LLPS modulation patterns highly reminiscent of Tar32. Subsequent NMR analysis revealed that the residues with significant perturbations by both Tar32 and A32 are very similar, indicating a striking similarity between their effects.

These results suggest that the presence of the hydrophobic region 311–343 does not notably influence the binding affinity of A6, A32, or Tar32 to Arg/Lys residues. However, it plays a significant role in bolstering the force driving LLPS. As such, although ATP and A6 bind to similar residues in WT PLDs and PLD Δ(311–343), they fail to induce the LLPS of PLD Δ(311–343) due to the low binding affinity and absence of an inherent driving force from the oligomerization of the residues 311–343. In contrast, Tar32/A32, which exhibit multivalent binding with stronger binding affinity, are capable of instigating and resolving the LLPS of PLD Δ(311–343), even without the intrinsic driving force.

3.3. How ATP and Nucleic Acids Interplay to Modulate the LLPS of the SARS-CoV-2 N Protein

The N protein of SARS-CoV-2 features three lengthy IDRs abundant in Arg/Lys residues, but lacks Tyr and Trp, having only two Phe residues within these regions. Consequently, the N protein exhibited minimal phase separation capacity in the absence of ATP and nucleic acids. Furthermore, the C-terminal folded domain (CTD) not only facilitates dimerization, but also enhances oligomerization, potentially reinforcing LLPS but also exacerbating aggregation, akin to observations in the hydrophobic region of the TDP-43 PLD. While turbidity measurement and DIC imaging revealed the biphasic effects of ATP and the nucleic acids of both specific and non-specific sequences on the LLPS of the entire N protein, the aggregation-prone complexity hinders detailed residue-specific investigations using NMR techniques [103]. However, a recent NMR characterization of a dissected fragment, N (1–249) composed of the folded NTD and two IDRs, elucidated the modulating mechanism of LLPS by ATP and a 32-mer S2m ssDNA derived from the stem-loop II motif of SARS-CoV-2 [103].

Comprehensive NMR studies reveal that S2m drives both the formation and dissolution of LLPS in N (1–249) primarily through specific interactions with the folded NTD and Arg/Lys residues within the IDRs. A speculative model was formulated to depict this process (Figure 4d). Initially, when S2m is added to the homogeneous N (1–249) solution at lower ratios (I of Figure 4d), a dynamic and multivalent interaction occurs between N (1–249) and S2m, engaging both the NTD and IDR Arg/Lys residues. This interaction results in the formation of large and dynamic complexes, visually observed as liquid droplets (II of Figure 4d). However, as the excess S2m is introduced, multiple S2m molecules bind to a single N (1–249) molecule, causing the disruption of these large and dynamic complexes and leading to the dissolution of liquid droplets into a homogeneous solution (III of Figure 4d). The residue-specific NMR data also indicate a notable interplay

between ATP and S2m in regulating the LLPS of N (1–249) by binding to overlapping sites, i.e., the folded NTD and Arg/Lys regions within the IDRs, despite exhibiting significantly different affinities. Remarkably, ATP at substantially higher concentrations than S2m was capable of disrupting the S2m-induced LLPS of N (1–249) by competitively displacing S2m from binding to the proteins (III' of Figure 4d). Detailed NMR investigations, including ATP/S2m competition experiments, elucidated this mechanism [85]. Notably, due to S2m binding to N (1–249) with a Kd of ~μM, considerably higher than ATP (with a Kd of ~mM), ATP at 20 mM remained insufficient to completely displace S2m from binding to N (1–249) [103].

In summary, experimental investigations into the modulation of the LLPS of FUS, TDP-43 PLD, and N (1–249) composed of both folded nucleic-acid-binding domains and IDRs uncover a common mechanism by which ATP and nucleic acids interplay to modulate LLPS by competitively binding to the nucleic-acid-binding interface of folded domains, as well as arginine/lysine residues within the IDRs. Previously, ATP and RNA have also been shown to interact with the RGG-rich IDRs of human cold-inducible RNA-binding protein (CIRBP) to modulate its LLPS [80]. Furthermore, a detailed NMR study also revealed that ATP can bind to enhance or reduce the phase separation of the C-terminal IDR of an RNA-binding protein, CAPRIN1 [104].

Strikingly, by combining the semiempirical quantum mechanical (SQM) method, mean-field theory, and coarse-grained molecular dynamics (CGMD) simulations, ATP has been shown to act as a bivalent or trivalent binder to inhibit and enhance the phase separation of FUS [33]. Very recently, a comprehensive study using both experimental and simulation methods revealed the unique properties and mechanisms underlying the ATP-induced phase separation of basic intrinsically disordered proteins (bIDPs). Briefly, ATP serves as a bridging agent, cross-linking bIDP chains to create mesh-like networks that manifest as liquid-like droplets. These droplets exhibit unusual physicochemical characteristics, such as very high ATP concentrations within the droplets, rapid fusion rates, low interfacial tension, and high viscosities, thus resulting in extreme shear-thinning behavior [105].

Taken together, these results suggest the following: (1) Unlike the classic high-affinity complexes formed between some folded proteins and ATP/nucleic acid, in which various residue types contribute significantly to binding, and consequently, a minor atomic alteration in ATP/nucleic acids or proteins could dramatically alter binding affinity, ATP/nucleic acids form dynamic complexes with IDRs through bivalent/multivalent binding to Arg/Lys residues within IDRs, which lack defined conformations and are highly accessible to bulk solvent and small molecules. The formation of such complexes is mainly driven by electrostatic interactions between phosphate groups of ATP/nucleic acids and side-chain cations of Arg/Lys, as well as by π–π/π–cation interactions between base aromatic rings and Arg/Lys side chains. As a result, different bases exhibit highly similar affinities to Arg/Lys residues, while RNA and ssDNA are anticipated to bind Arg/Lys residues of IDRs using the same mechanism due to their minor chemical structural differences, suggesting that these bindings might be considered to be "residue-specific/selective". (2) On the other hand, nucleic acids differ from ATP by having multiple covalently linked nucleotides, allowing for the establishment of multivalent but rather independent binding to Arg/Lys, resulting in length-dependent affinity. Two crucial factors governing ATP-induced LLPS are the quantity of Arg/Lys residues within IDRs and the presence of additional driving forces for LLPS. Therefore, when nucleic acids possess adequate lengths and IDRs contain multiple Arg/Lys residues, they can instigate the LLPS of these IDRs, suggesting that the magnitude of nucleic-acid-induced LLPS may exceed the level of current recognition. (3) In particular, ATP, distinguished by its unique triphosphate group, demonstrates a unique capacity for interacting with water molecules and the Arg/Lys side chain. These distinctions may impact the competition between ATP and nucleic acids in binding Arg/Lys residues, as well as their abilities to modulate LLPS. (4) Notably, although nucleic acids may attain a high binding affinity to IDRs by forming multivalent but independent bindings to multiple Arg/Lys residues, such high-affinity binding appears relatively susceptible to displace-

ment by ATP. This is due to the fact that these dynamic complexes are highly accessible, and therefore, ATP can simultaneously displace each independent binding event via the nucleotides of nucleic acids. This phenomenon may rationalize the observation that, on the one hand, a substantial proportion of IDRs possess multiple Arg/Lys residues, rendering them prone to phase separation upon induction by nucleic acids with typical lengths. On the other hand, the occurrence of MLOs is notably limited in cells. It is therefore most likely that high cellular concentrations of ATP may function to inhibit/dissolve the majority of nucleic-acid-induced phase separations of IDRs. Consequently, proteins within MLOs consistently feature folded domains with high nucleic-acid-binding affinities, enabling the establishment of interactions with nucleic acids even in the presence of high cellular concentrations of ATP. (5) Given the highly negative charges of both ATP and nucleic acids, they possess the capability to exert strong electrostatic effects on LLPS and the aggregation of IDRs. However, these effects appear to be highly context-dependent. For instance, in the presence of excessive amounts, if ATP or nucleic acids are bound to Arg/Lys residues within IDRs such as WT TDP-43 PLD, the IDR molecules acquire site-/conformation-specific associations with multiple negative charges, resulting in repulsive electrostatic interactions disrupting LLPS and preventing irreversible aggregation. In contrast, if ATP and/or nucleic acids are unable to bind to IDRs such as the TDP-43 PLD mutant with all Arg/Lys residues mutated, their non-specific screening effect dominates, acting to induce aggregation driven by the oligomerization of the unique hydrophobic region [100].
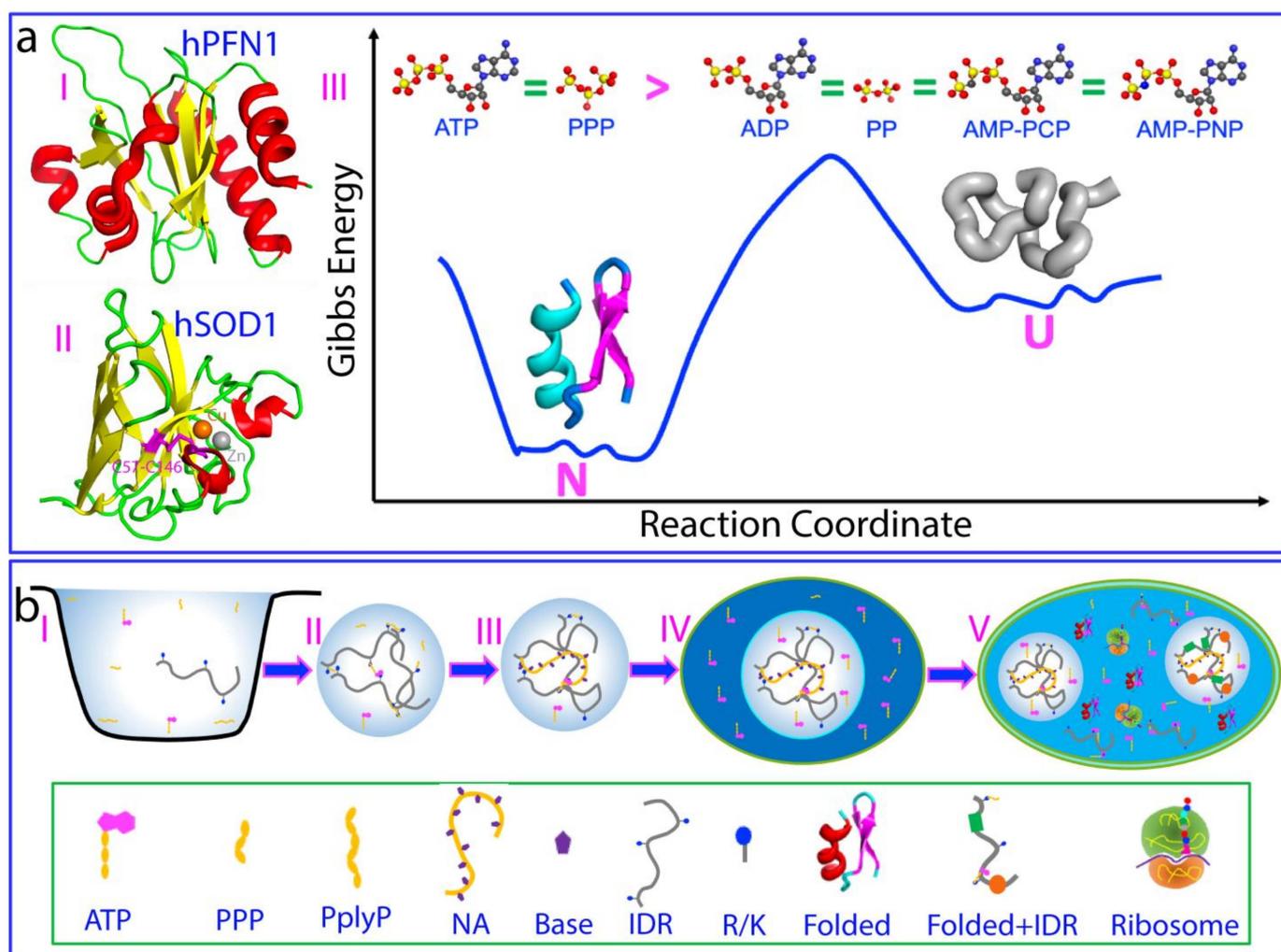
4. ATP Induces Protein Folding with the Highest Efficiency

Approximately 50% of the proteome needs to fold from the unfolded state (U) to the folded state (F) to fulfill its functional role [106–108]. Given the marginal stability of the folded state, genetic mutations may induce destabilization, resulting in misfolding and aggregation within cellular environments—a prevalent pathological characteristic associated with aging and neurodegenerative diseases [109–111]. Modern cells employ supramolecular machinery, energetically fueled by ATP, as the primary mechanism to handle challenges related to protein folding and misfolding/aggregation [111]. However, the modulation of protein folding during primordial evolution, predating the emergence of complex machineries, remains an unresolved question. Notably, previous studies have demonstrated that ATP can exert proteome-wide control over the solubility and aggregation of proteins both in vitro and in vivo [112,113]. Nevertheless, the direct influence of ATP on the equilibrium of protein folding, a fundamental event in protein homeostasis, hitherto remains unknown.

Recently, it was discovered that ATP can induce the folding of ALS-causing mutants of human profilin 1 (hPFN1) (I of Figure 5a) and CuZn-superoxide dismutase 1 (hSOD1) (II of Figure 5a) [39]. C71G-hPFN1 coexists in both folded and unfolded states, with respective populations of 55.2% and 44.8%, undergoing dynamic exchange at a rate of 11.7 Hz. Nascent hSOD1, lacking metal cofactors and disulfide bridges, remains entirely unfolded, devoid of any stable secondary and tertiary structures. Through the NMR visualization of ATP and 11 related compounds, it has been demonstrated that ATP fully converts C71G-hPFN1 into the folded state at a ratio of 1:2 (C71G:ATP), while inducing nascent hSOD1 into two co-existing states at a ratio of 1:8 (hSOD1:ATP) (III of Figure 5a), the highest efficiency known so far.

Unexpectedly, the remarkable inducing capacity of ATP comes from its triphosphate moiety, previously proposed as a pivotal intermediate in prebiotic chemical evolutions, contributing to the generation of building units for constructing primitive cells and potentially influencing the origin of life [1–4]. However, it is noteworthy that triphosphate exhibits a strong ability to induce aggregation. The inducing capacity can be ranked as follows: ATP = ATPP = PPP > ADP = AMP−PNP = AMP−PCP = PP. Meanwhile, AMP, adenosine, P, and NaCl demonstrate no conversion. Surprisingly, trimethylamine N-oxide (TMAO), a well-known molecule with a general capacity to induce protein folding [108],

fails to exhibit any detectable induction of the folding of C71G-hPFN1, even at a ratio of 1:2000 (C71G:TMAO), where protein aggregation occurs.

Figure 5. ATP enhances protein folding and is proposed to have contributed to the origin of life. (**a**) ATP energy-independently induces protein folding with the highest efficiency known so far. (**b**) ATP plays key roles in controlling protein homeostasis and interactions between nucleic acids and proteins throughout the evolutionary trajectory from prebiotic origins to modern cells.

For nascent hSOD1, 12 cations (Na^+, K^+, Ca^{2+}, Zn^{2+}, Mg^{2+}, Mn^{2+}, Cu^{2+}, Fe^{2+}, Ni^{2+}, Cd^{2+}, Co^{2+}, and Al^{3+}) have been evaluated, but only Zn^{2+} and Fe^{2+} demonstrate the ability to induce folding with a coexistence of unfolded and folded states at a ratio of 1:20 (hSOD1:cation) [114]. Notably, ATP exhibits a distinct capacity to induce folding, leading to the coexistence of unfolded and folded states at a much smaller ratio of 1:8 (hSOD1:ATP) [39]. In contrast, TMAO fails to induce the folding of hSOD1, even at a ratio of 1:1000 (hSOD1:TMAO), where protein aggregation occurs. Mechanistically, a comparison of the conformations and dynamics of ATP- and Zn^{2+}-induced hSOD1 folded states suggests that ATP and triphosphate induce folding by enhancing the intrinsic folding capacity encoded in the protein sequences. Specifically, ATP and triphosphate have been proposed to induce folding mainly by interacting with and displacing water molecules that hydrogen-bond with the unfolded protein [39].

Therefore, nature appears to select triphosphate as a central intermediate in prebiotic evolution, possibly also due to its notably high efficiency in inducing protein folding. Intriguingly, polyphosphates may employ a similar mechanism to triphosphate to generally

facilitate protein folding. Consequently, inorganic polyphosphates persist as primordial chaperones in some single-cell organisms [115,116]. Nevertheless, owing to the propensity of high triphosphate concentrations to induce aggregation in unfolded or partially unfolded protein states, modern cells maintain extremely low concentrations of triphosphate. Nature marvelously creates ATP that possesses three integrated abilities: to effectively induce protein folding, inhibit aggregation, and enhance thermodynamic stability by joining adenosine with triphosphate [39].

5. ATP Action from Prebiotic Evolution to Modern Cells

A recent study suggests that ATP emerged through the phosphorylation of ADP via a universally conserved intermediate, acetyl phosphate (AcP), during the early stages of biochemical evolution [6]. Intriguingly, the highest ATP yield was observed in highly unsalted water, such as HPLC-grade water, at an optimal pH of approximately 5.5–6, while the yield significantly diminished upon the introduction of salts and alterations in pH [6]. Remarkably, the aqueous condition observed in this study [6] coincides with the solution condition in which all insoluble proteins could be solubilized [117], also akin to the conditions present in prebiotic oceans or water bodies [6,110]. Notably, AcP phosphorylated only ADP but not other nucleoside diphosphates, contributing to the universal conservation of ATP across diverse life forms.

Therefore, as illustrated in part I of Figure 5b, at the early stages of biochemical evolution, water was highly unsalted with a slightly acidic pH in which all primordial peptides/proteins were soluble [110,117]. The presence of ATP and/or triphosphate might drive the phase separation of primordial peptides/proteins with cationic residues, thus serving to concentrate these rare peptides/proteins (II of Figure 5b). Furthermore, the binding of ATP to these primordial peptides/proteins might also have protected them from aggregation upon the coming increase in salt concentration in the water [6,110]. Later, as nucleic acids emerged, primordial peptides/proteins may have become mainly bound to nucleic acids, leading to phase separation and the formation of dynamic protein–nucleic acid complexes (III of Figure 5b). Subsequently, the phase-separated droplets may have been separated from their surroundings through encapsulation within a membrane system, thereby giving rise to the formation of primordial cells (IV of Figure 5b). In primordial cells, mechanisms for the self-reproduction of proteins and nucleic acids might have evolved and established. As such, with the extensive sampling of protein sequences, some proteins become folded with the induction of ATP, while some remain intrinsically disordered. Furthermore, under modulation by ATP, protein–nucleic acid complexes also evolved into two categories: well-folded ones, as exemplified by ribosomes, and dynamic droplets for forming membrane-less organelles (V of Figure 5b).

In modern cells, genetic information is typically stored in genomes (DNA in most organisms, but RNA in some viruses). Protein–nucleic acid interactions play a central role in the expression of genetic information into proteins (proteomes). Consequently, the interfaces between genomes and proteomes are crucial for regulating, maintaining, and expressing genetic information. In eukaryotic cells, most nucleic-acid-binding proteins contain both folded and intrinsically disordered domains (Figure 6a). The binding affinity and specificity of these domains for nucleic acids vary widely, depending on the proteins and their target DNA/RNA sequences. Folded domains can have binding affinities (Kd) for nucleic acids ranging from picomolar (pM) to millimolar (mM). On the other hand, RG/RGG-rich motifs within intrinsically disordered regions (IDRs) of over 1700 human proteins were previously identified as nucleic acid binders [118,119], while recent research indicates that Arg/Lys residues within IDRs are sufficient to bind nucleic acids, with affinity highly dependent on binding valency [100].

As ATP closely resembles nucleotides, the building blocks of nucleic acids (Figure 6b), it can bind to pockets within the nucleic-acid-binding interfaces of folded domains. Consequently, ATP competes with nucleic acids for binding to these proteins. However, whether ATP can displace nucleic acids from these proteins depends heavily on the ratios of binding

affinity and concentrations among ATP, nucleic acids, and proteins. Additionally, ATP seems capable of readily displacing nucleic acids from binding to Arg/Lys within IDRs because nucleic acids bind Arg/Lys to form dynamic and accessible complexes in a multivalent manner, but with each binding event occurring rather independently. Strikingly, it was demonstrated that ATP could even directly interact with and destabilize the CAG-RNA hairpin structure, thus modulating RNA–protein interactions [120]. Therefore, ATP appears to play a central role in shaping the genome–proteome interface (Figure 6c) by modulating the formation of folded protein–nucleic acid complexes and IDR nucleic acid membrane-less organelles.

Figure 6. ATP shapes the genome–proteome interface. (**a**) Major interactions between nucleic acids and proteins composed of intrinsically disordered regions (IDRs) and folded domains including the nucleic-acid-binding domain. (**b**) Major effects that ATP exerts on proteins composed of intrinsically disordered regions (IDRs) and well-folded domains including the nucleic-acid-binding domain. (**c**) ATP sits in between the genome and proteome in modern cells to regulate various biological functions.

ATP is composed of both adenosine and a triphosphate group with a unique hydration property [37], and is thus capable of exerting extremely complex effects on proteins and nucleic acids, including hydrotropic, cosolute, and/or screening effects at high concentrations in a context-dependent manner (Figure 6b). The triphosphate group in ATP can directly alter protein hydration even at very low molar ratios, counteracting the crowding-induced destabilization of human eye-lens γ-crystalline at 1:1 (crystalline:ATP) [38] and enhancing the protein folding of C71G-hPFN1 at 1:2 (C71G:ATP) and hSOD1 at 1:8 (hSOD1:ATP) [39]. In modern cells, ATP not only serves as the energy currency for numerous biological processes, but also energy-independently regulates protein homeostasis. The findings reviewed here suggest that ATP also sits at the genome–proteome interface, energy-independently maintaining and controlling the expression of genetic information (Figure 6c).

In the future, two areas of great interest need to be addressed, but they come with fundamental challenges. The first is that understanding the multifaceted capacities of ATP requires decoding the core mechanisms by which ATP mediates protein hydration. However, achieving this requires a significant breakthrough in experimentally and computationally deciphering the abnormality of water [121–124]. The second relates to revealing the biological consequences of ATP's effects. This is central but also fundamentally challeng-

ing. This is due to several factors: (1) ATP's involvement in a plethora of energy-dependent and energy-independent processes that are interconnected or coupled; (2) ATP has emerged in prebiotic chemistry, implying that its energy-independent effects might have been extensively integrated into various cellular processes during evolution; and (3), the primary challenge, the fact that biological function represents an "emergence" from complex systems/networks of biochemical/biophysical interactions that cannot be understood or predicted solely by analyzing the individual components of the systems/networks, largely due to non-linearity [125].

References

1. Todd, A.R. *Where There's Life There's Phosphorus*; Nakamura, K., Kageyama, M., Oshima, T., Eds.; Science and Society Press: Tokyo, Japan, 1981; p. 275.
2. Westheimer, F.H. Why nature chose phosphates. *Science* **1987**, *235*, 1173–1178. [CrossRef] [PubMed]
3. Bowler, M.W.; Cliff, M.; Walthoand, J.P.; Blackburn, G.M. Why did nature select phosphate for its dominant roles in biology? *New J. Chem.* **2010**, *34*, 784–794. [CrossRef]
4. Kamerlin, S.C.L.; Sharma, P.K.; Prasad, R.B.; Warshel, A. Why nature really chose phosphate. *Q. Rev. Biophys.* **2013**, *46*, 1–132. [CrossRef] [PubMed]
5. Leningher, A. *Principles of Biochemistry*; W. H. Freeman and Company: New York, NY, USA, 2005.
6. Pinna, S.; Kunz, C.; Halpern, A.; Harrison, S.A.; Jordan, S.F.; Ward, J.; Werner, F.; Lane, N. A prebiotic basis for ATP as the universal energy currency. *PLoS Biol.* **2022**, *20*, e3001437. [CrossRef] [PubMed]
7. Brangwynne, C.P.; Eckmann, C.R.; Courson, D.S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A.A. Germline Pgranules are liquid droplets that localize by controlled dissolution/condensation. *Science* **2009**, *324*, 1729–1732. [CrossRef] [PubMed]
8. Hyman, A.A.; Weber, C.A.; Jülicher, F. Liquid-liquid phase separation in biology. *Annu. Rev. Cell Dev. Biol.* **2014**, *30*, 39–58. [CrossRef] [PubMed]
9. Shin, Y.; Brangwynne, C.P. Liquid phase condensation in cell physiology and disease. *Science* **2017**, *357*, 6357. [CrossRef] [PubMed]
10. Mitrea, D.M.; Kriwacki, R.W. Phase separation in biology; functional organization of a higher order. *Cell Commun. Signal.* **2016**, *14*, 1. [CrossRef] [PubMed]
11. Handwerger, K.E.; Gall, J.G. Subnuclear organelles: New insights into form and function. *Trends Cell Biol.* **2006**, *16*, 19–26. [CrossRef]
12. Overbeek, J.T.G.; Voorn, M.J. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *J. Cell. Comp. Physiol.* **1957**, *49*, 7–26. [CrossRef]
13. Taratuta, V.G.; Holschbach, A.; Thurston, G.M.; Blankschtein, D.; Benedek, G.B. Liquid-liquid phase separation of aqueous lysozyme solutions: Effects of pH and salt identity. *J. Phys. Chem.* **1990**, *94*, 2140–2472. [CrossRef]
14. Song, J. Molecular mechanisms of phase separation and amyloidosis of ALS/FTD-linked FUS and TDP-43. *Aging Dis.* **2023**. [CrossRef] [PubMed]
15. Toretsky, J.A.; Wright, P.E. Assemblages: Functional units formed by cellular phase separation. *J. Cell Biol.* **2014**, *206*, 579–588. [CrossRef] [PubMed]
16. Wang, J.; Choi, J.M.; Holehouse, A.S.; Lee, H.O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; et al. Molecular grammar underlying the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **2018**, *174*, 688–699.e16. [CrossRef] [PubMed]
17. Brangwynne, C.P.; Tompa, P.; Pappu, R.V. Polymer physics of intracellular phase transitions. *Nat. Phys.* **2015**, *11*, 899–904. [CrossRef]
18. Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **2019**, *176*, 419–434. [CrossRef] [PubMed]
19. Dang, M.; Song, J. A review of the effects of ATP and hydroxychloroquine on the phase separation of the SARS-CoV-2 nucleocapsid protein. *Biophys. Rev.* **2022**, *14*, 709–715. [CrossRef] [PubMed]
20. Brady, J.P.; Farber, P.J.; Sekhar, A.; Lin, Y.-H.; Huang, R.; Bah, A.; Nott, T.J.; Chan, H.S.; Baldwin, A.J.; Forman-Kay, J.D.; et al. Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E8194–E8203. [CrossRef] [PubMed]
21. Da Cruz, S.; Cleveland, D.W. Understanding the role of TDP-43 and FUS/TLS in ALS and beyond. *Curr. Opin. Neurobiol.* **2011**, *21*, 904–919. [CrossRef]

22. Portz, B.; Lee, B.L.; Shorter, J. FUS and TDP-43 Phases in Health and Disease. *Trends Biochem. Sci.* **2021**, *46*, 550–563. [CrossRef]

23. Keating, C.D. Aqueous phase separation as a possible route to compartmentalization of biological molecules. *Acc. Chem. Res.* **2012**, *45*, 2114–2124. [CrossRef] [PubMed]

24. Maharana, S.; Wang, J.; Papadopoulos, D.K.; Richter, D.; Pozniakovsky, A.; Poser, I.; Bickle, M.; Rizk, S.; Guillén-Boixet, J.; Franzmann, T.M.; et al. RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **2018**, *360*, 918–921. [CrossRef] [PubMed]

25. Milin, A.N.; Deniz, A.A. Reentrant phase transitions and non- equilibrium dynamics in membraneless organelles. *Biochemistry* **2018**, *57*, 2470–2477. [CrossRef] [PubMed]

26. Ukmar-Godec, T.; Hutten, S.; Grieshop, M.P.; Rezaei-Ghaleh, N.; Cima-Omori, M.-S.; Biernat, J.; Mandelkow, E.; Söding, J.; Dormann, D.; Zweckstetter, M. Lysine/RNA-interactions drive and regulate biomolecular condensation. *Nat. Commun.* **2019**, *10*, 2909. [CrossRef] [PubMed]

27. Kang, J.; Lim, L.; Lu, Y.; Song, J. A unified mechanism for LLPS of ALS/FTLD-causing FUS as well as its modulation by ATP and oligonucleic acids. *PLoS Biol.* **2019**, *17*, e3000327. [CrossRef] [PubMed]

28. Kang, J.; Lim, L.; Song, J. ATP enhances at low concentrations but dissolves at high concentrations liquid-liquid phase separation (LLPS) of ALS/FTD-causing FUS. *Biochem. Biophys. Res. Commun.* **2018**, *504*, 545–551. [CrossRef]

29. Wang, L.; Kang, J.; Lim, L.; Wei, Y.; Song, J. TDP-43 NTD can be induced while CTD is significantly enhanced by ssDNA to undergo liquid-liquid phase separation. *Biochem. Biophys. Res. Commun.* **2018**, *499*, 189–195. [CrossRef]

30. Patel, A.; Malinovska, L.; Saha, S.; Wang, J.; Alberti, S.; Krishnan, Y.; Hyman, A.A. ATP as a biological hydrotrope. *Science* **2017**, *356*, 753–756. [CrossRef]

31. Rice, A.M.; Rosen, M.K. ATP controls the crowd. *Science* **2017**, *356*, 701–702. [CrossRef]

32. Mehringer, J.; Do, T.-M.; Touraud, D.; Hohenschutz, M.; Khoshsima, A.; Horinek, D.; Kunz, W. Hofmeister versus Neuberg: Is ATP really a biological hydrotrope? Cell Rep. *Phys. Sci.* **2021**, *2*, 100343. [CrossRef]

33. Hautke, A.; Ebbinghaus, S. The emerging role of ATP as a cosolute for biomolecular processes. *Biol. Chem.* **2023**, *404*, 897–908. [CrossRef] [PubMed]

34. Salonen, L.M.; Ellermann, M.; Diederich, F. Aromatic rings in chemical and biological recognition: Energetics and structures. *Angew. Chem. Int. Ed. Engl.* **2011**, *50*, 4808–4842. [CrossRef] [PubMed]

35. Ren, C.-L.; Shan, Y.; Zhang, P.; Ding, H.-M.; Ma, Y.-Q. Uncovering the molecular mechanism for dual effect of ATP on phase separation in FUS solution. *Sci. Adv.* **2022**, *8*, eabo7885. [CrossRef] [PubMed]

36. Song, J. ATP energy-independently controls protein homeostasis with unique structure and diverse mechanisms. *Protein Sci.* **2021**, *30*, 1277–1293. [CrossRef] [PubMed]

37. Mogami, G.; Wazawa, T.; Morimoto, N.; Kodama, T.; Suzuki, M. Hydration properties of adenosine phosphate series as studied by microwave dielectric spectroscopy. *Biophys. Chem.* **2011**, *154*, 1–7. [CrossRef] [PubMed]

38. He, Y.; Kang, J.; Song, J. ATP antagonizes the crowding-induced destabilization of the human eye-lens protein γS-crystallin. *Biochem. Biophys. Res. Commun.* **2020**, *526*, 1112–1117. [CrossRef] [PubMed]

39. Kang, J.; Lim, L.; Song, J. ATP induces folding of ALS-causing C71G-hPFN1 and nascent hSOD1. *Commun. Chem.* **2023**, *2*, 223. [CrossRef] [PubMed]

40. Roden, C.; Gladfelter, A.S. RNA contributions to the form and function of biomolecular condensates. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 183–195. [CrossRef] [PubMed]

41. Lafontaine, D.L.J.; Riback, J.A.; Bascetin, R.; Brangwynne, C.P. The nucleolus as a multiphase liquid condensate. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 165–182. [CrossRef] [PubMed]

42. Bhat, P.; Honson, D.; Guttman, M. Nuclear compartmentalization as a mechanism of quantitative control of gene expression. *Nat. Rev. Mol. Cell Biol.* **2021**, *22*, 653–670. [CrossRef]

43. Emmanouilidis, L.; Esteban-Hofer, L.; Jeschke, G.; Allain, F.H.T. Structural biology of RNA-binding proteins in the context of phase separation: What NMR and EPR can bring? *Curr. Opin. Struct. Biol.* **2021**, *70*, 132–138. [CrossRef] [PubMed]

44. Schwartz, J.C.; Cech, T.R.; Parker, R.R. Biochemical properties and biological functions of FET proteins. *Annu. Rev. Biochem.* **2015**, *84*, 355–379. [CrossRef] [PubMed]

45. Ling, S.C.; Polymenidou, M.; Cleveland, D.W. Converging mechanisms in ALS and FTD: Disrupted RNA and protein homeostasis. *Neuron* **2013**, *79*, 416–438. [CrossRef] [PubMed]

46. Cléry, A.; Blatter, M.; Allain, F.H. RNA recognition motifs: Boring? Not quite. *Curr. Opin. Struct. Biol.* **2008**, *18*, 290–298. [CrossRef] [PubMed]

47. World Health Organization. *WHO Coronavirus Disease (COVID-19) Dashboard*; World Health Organization: Geneva, Switzerland, 2020.

48. Masters, P.S. Coronavirus genomic RNA packaging. *Virology* **2019**, *537*, 198–207. [CrossRef] [PubMed]

49. Cong, Y.; Ulasli, M.; Schepers, H.; Mauthe, M.; V'Kovski, P.; Kriegenburg, F.; Thiel, V.; de Haan, C.A.M.; Reggiori, F. Nucleocapsid protein recruitment to replication-transcription complexes plays a crucial role in coronaviral life cycle. *J. Virol.* **2020**, *94*, e01925-19. [CrossRef] [PubMed]

50. Iserman, C.; Roden, C.A.; Boerneke, M.A.; Sealfon, R.S.; McLaughlin, G.A.; Jungreis, I.; Fritch, E.J.; Hou, Y.J.; Ekena, J.; Weidmann, C.A.; et al. Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell* **2020**, *80*, 1078–1091. [CrossRef] [PubMed]

51. Lu, S.; Ye, Q.; Singh, D.; Cao, Y.; Diedrich, J.K.; Yates, J.R., 3rd; Villa, E.; Cleveland, D.W.; Corbett, K.D. The SARS-CoV-2 nucleocapsid phosphoprotein forms mutually exclusive condensates with RNA and the membrane-associated M protein. *Nat. Commun.* **2021**, *12*, 502. [CrossRef]

52. Dang, M.; Li, Y.; Song, J. ATP biphasically modulates LLPS of SARS-CoV-2 nucleocapsid protein and specifically binds its RNA-binding domain. Biochem Biophys. *Res. Commun.* **2021**, *19*, 50–55.

53. Dinesh, D.C.; Chalupska, D.; Silhan, J.; Koutna, E.; Nencka, R.; Veverka, V.; Boura, E. Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* **2020**, *16*, e1009100. [CrossRef]

54. Peng, Y.; Du, N.; Lei, Y.; Dorje, S.; Qi, J.; Luo, T.; Gao, G.F.; Song, H. Structures of the SARS-CoV-2 nucleocapsid and their perspectives for drug design. *EMBO J.* **2020**, *39*, e105938. [CrossRef] [PubMed]

55. Dang, M.; Song, J. Structural basis of anti-SARS-CoV-2 activity of HCQ: Specific binding to N protein to disrupt its interaction with nucleic acids and LLPS. *QRB Discov.* **2021**, *2*, e13. [CrossRef] [PubMed]

56. Brocca, S.; Grandori, R.; Longhi, S.; Uversky, V. Liquid–liquid phase separation by intrinsically disordered protein regions of viruses: Roles in viral life cycle and control of virus–host interactions. *Int. J. Mol. Sci.* **2020**, *21*, 9045. [CrossRef] [PubMed]

57. Guseva, S.; Milles, S.; Jensen, M.R.; Schoehn, G.; Ruigrok, R.W.; Blackledge, M. Structure, dynamics and phase separation of measles virus RNA replication machinery. *Curr. Opin. Virol.* **2020**, *41*, 59–67. [CrossRef] [PubMed]

58. Scoca, V.; Di Nunzio, F. Membraneless organelles restructured and built by pandemic viruses: HIV-1 and SARS-CoV-2. *J. Mol. Cell Biol.* **2021**, *13*, 259–268. [CrossRef] [PubMed]

59. Monette, A.; Niu, M.; Chen, L.; Rao, S.; Gorelick, R.J.; Mouland, A.J. Pan-retroviral nucleocapsid-mediated phase separation regulates genomic RNA positioning and trafficking. *Cell Rep.* **2020**, *31*, 107520. [CrossRef] [PubMed]

60. Salladini, E.; Gondelaud, F.; Nilsson, J.F.; Pesce, G.; Bignon, C.; Murrali, M.G.; Fabre, R.; Pierattelli, R.; Kajava, A.V.; Horvat, B.; et al. Identification of a region in the common amino-terminal domain of Hendra Virus P, V, andWproteins responsible for phase transition and amyloid formation. *Biomolecules* **2021**, *11*, 1324. [CrossRef] [PubMed]

61. Heinrich, B.S.; Maliga, Z.; Stein, D.A.; Hyman, A.A.; Whelan, S.P. Phase transitions drive the formation of vesicular stomatitis virus replication compartments. *mBio* **2018**, *9*, e02290-17. [CrossRef] [PubMed]

62. Hirai, Y.; Tomonaga, K.; Horie, M. Borna disease virus phosphoprotein triggers the organization of viral inclusion bodies by liquid-liquid phase separation. *Int. J. Biol. Macromol.* **2021**, *192*, 55–63. [CrossRef]

63. Nikolic, J.; Le Bars, R.; Lama, Z.; Scrima, N.; Lagaudrière-Gesbert, C.; Gaudin, Y.; Blondel, D. Negri bodies are viral factories with properties of liquid organelles. *Nat. Commun.* **2017**, *8*, 58. [CrossRef]

64. Bianchi, G.; Brocca, S.; Longhi, S.; Uversky, V. Liaisons dangereuses: Intrinsic Disorder in Cellular Proteins Recruited to Viral Infection-Related Biocondensates. *Int. J. Mol. Sci.* **2023**, *24*, 2151. [CrossRef] [PubMed]

65. Liu, X.; Niu, C.; Ren, J.; Zhang, J.; Xie, X.; Zhu, H.; Feng, W.; Gong, W. The RRM domain of human fused in sarcoma protein reveals a non-canonical nucleic acid binding site. *Biochim. Biophys. Acta* **2013**, *1832*, 375–385. [CrossRef] [PubMed]

66. Loughlin, F.E.; Lukavsky, P.J.; Kazeeva, T.; Reber, S.; Hock, E.-M.; Colombo, M.; Von Schroetter, C.; Pauli, P.; Cléry, A.; Mühlemann, O.; et al. The Solution Structure of FUS Bound to RNA Reveals a Bipartite Mode of RNA Recognition with Both Sequence and Shape Specificity. *Mol. Cell* **2019**, *73*, 490–504.e6. [CrossRef] [PubMed]

67. Jutzi, D.; Campagne, S.; Schmidt, R.; Reber, S.; Mechtersheimer, J.; Gypas, F.; Schweingruber, C.; Colombo, M.; von Schroetter, C.; Loughlin, F.E.; et al. Aberrant interaction of FUS with the U1 snRNA provides a molecular mechanism of FUS induced amyotrophic lateral sclerosis. *Nat. Commun.* **2020**, *11*, 6341. [CrossRef] [PubMed]

68. Sun, Z.; Diaz, Z.; Fang, X.; Hart, M.P.; Chesi, A.; Shorter, J.; Gitler, A.D. Molecular determinants and genetic modifiers of aggregation and toxicity for the ALS disease protein FUS/TLS. *PLoS Biol.* **2011**, *9*, e1000614. [CrossRef] [PubMed]

69. Lu, Y.; Lim, L.; Song, J. RRM domain of ALS/FTD-causing FUS characteristic of irreversible unfolding spontaneously self-assembles into amyloid fibrils. *Sci. Rep.* **2017**, *7*, 1043. [CrossRef] [PubMed]

70. Kang, J.; Lim, L.Z.; Song, J. ATP binds and inhibits the neurodegeneration-associated fibrillization of the FUS RRM domain. *Commun. Biol.* **2019**, *2*, 223. [CrossRef] [PubMed]

71. Ou, S.H.; Wu, F.; Harrich, D.; García-Martínez, L.F.; Gaynor, R.B. Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. *J. Virol.* **1995**, *69*, 3584–3596. [CrossRef] [PubMed]

72. Qin, H.; Lim, L.; Wei, Y.; Song, J. TDP-43 N terminus encodes a novel ubiquitin-like fold and its unfolded form in equilibrium that can be shifted by binding to ssDNA. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 18619–18624. [CrossRef]

73. Afroz, T.; Hock, E.-M.; Ernst, P.; Foglieni, C.; Jambeau, M.; Gilhespy, L.A.B.; Laferriere, F.; Maniecka, Z.; Plückthun, A.; Mittl, P.; et al. Functional and dynamic polymerization of the ALS-linked protein TDP-43 antagonizes its pathologic aggregation. *Nat. Commun.* **2017**, *8*, 45. [CrossRef]

74. Kuo, P.H.; Doudeva, L.G.; Wang, Y.T.; Shen, C.K.; Yuan, H.S. Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.* **2009**, *37*, 1799–1808. [CrossRef] [PubMed]

75. Kuo, P.H.; Chiang, C.H.; Wang, Y.T.; Doudeva, L.G.; Yuan, H.S. The crystal structure of TDP-43 RRM1-DNA complex reveals the specific recognition for UG- and TG-rich nucleic acids. *Nucleic Acids Res.* **2014**, *42*, 4712–4722. [CrossRef] [PubMed]

76. Lukavsky, P.J.; Daujotyte, D.; Tollervey, J.R.; Ule, J.; Stuani, C.; Buratti, E.; Baralle, F.E.; Damberger, F.F.; Allain, F.H. Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. *Nat. Struct. Mol. Biol.* **2013**, *20*, 1443–1449. [CrossRef] [PubMed]

77. Dang, M.; Kang, J.; Lim, L.; Song, J. ATP is a cryptic binder of TDP-43 RRM domains to enhance stability and inhibit ALS/AD-associated fibrillation. *Biochem. Biophys. Res. Commun.* **2020**, *522*, 247–253. [CrossRef] [PubMed]

78. Dang, M.; Song, J. ALS-causing D169G mutation disrupts the ATP-binding capacity of TDP-43 RRM1 domain. *Biochem. Biophys. Res. Commun.* **2020**, *524*, 459–464. [CrossRef]

79. Dang, M.; Li, Y.; Song, J. Tethering-induced destabilization and ATP-binding for tandem RRM domains of ALS-causing TDP-43 and hnRNPA1. *Sci. Rep.* **2021**, *11*, 1034. [CrossRef] [PubMed]

80. Zhou, Q.; Usluer, S.; Zhang, F.; Lenard, A.J.; Bourgeois, B.M.; Madl, T. ATP regulates RNA-driven cold inducible RNA binding protein phase separation. *Protein Sci.* **2021**, *30*, 1438–1453. [CrossRef] [PubMed]

81. He, Y.; Kang, J.; Lim, L.; Song, J. ATP binds nucleic-acid-binding domains beyond RRM fold. *Biochem. Biophys. Res. Commun.* **2020**, *522*, 826–831. [CrossRef]

82. Beuck, C.; Williamson, J.R.; Wüthrich, K.; Serrano, P. The acidic domain is a unique structural feature of the splicing factor SYNCRIP. *Protein Sci.* **2016**, *25*, 1545–1550. [CrossRef]

83. Hobor, F.F.; Dallmann, A.; Ball, N.J.; Cicchini, C.; Battistelli, C. A cryptic RNAbinding domain mediates Syncrip recognition and exosomal partitioning of miRNA targets. *Nat. Commun.* **2018**, *9*, 831. [CrossRef]

84. Dang, M.; Song, J. CTD of SARS-CoV-2 N protein is a cryptic domain for binding ATP and nucleic acid that interplay in modulating phase separation. *Protein Sci.* **2021**, *31*, 345–356. [CrossRef] [PubMed]

85. Nishizawa, M.; Walinda, E.; Morimoto, D.; Kohn, B.; Scheler, U.; Shirakawa, M.; Sugase, K. Effects of weak nonspecific interactions with ATP on proteins. *J. Am. Chem. Soc.* **2021**, *143*, 11982–11993. [CrossRef] [PubMed]

86. Hu, G.; Ou, X.; Li, J. Mechanistic Insight on General Protein-Binding Ability of ATP and the Impacts of Arginine Residues. *J. Phys. Chem. B* **2022**, *126*, 4647–4658. [CrossRef] [PubMed]

87. Scheeff, E.D.; Bourne, P.E. Structural evolution of the protein kinase-like superfamily. *PLoS Comput. Biol.* **2005**, *1*, e49. [CrossRef] [PubMed]

88. Kanev, G.K.; de Graaf, C.; de Esch, I.J.; Leurs, R.; Würdinger, T.; Westerman, B.A.; Kooistra, A.J. The landscape of atypical and eukaryotic protein kinases. *Trends Pharmacol. Sci.* **2019**, *40*, 818–832. [CrossRef] [PubMed]

89. Michelitsch, M.D.; Weissman, J.S. A census of glutamine/asparagine-rich regions: Implications for their conserved function and the prediction of novel prions. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 11910–11915. [CrossRef] [PubMed]

90. Murray, D.T.; Kato, M.; Lin, Y.; Thurber, K.R.; Hung, I.; McKnight, S.L.; Tycko, R. Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low Complexity Domains. *Cell* **2017**, *171*, 615–627.e16. [CrossRef] [PubMed]

91. Kato, M.; McKnight, S.L. The low-complexity domain of the FUS RNA binding protein self-assembles via the mutually exclusive use of two distinct cross-β cores. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2114412118. [CrossRef]

92. Burke, K.A.; Janke, A.M.; Rhine, C.L.; Fawzi, N.L. Residue-by-Residue View of In Vitro FUS Granules that Bind the C-Terminal Domain of RNA Polymerase II. *Mol. Cell.* **2015**, *60*, 231–241. [CrossRef]

93. Murthy, A.C.; Tang, W.S.; Jovic, N.; Janke, A.M.; Seo, D.H.; Perdikari, T.M.; Mittal, J.; Fawzi, N.L. Molecular interactions contributing to FUS SYGQ LC-RGG phase separation and co-partitioning with RNA polymerase II heptads. *Nat. Struct. Mol. Biol.* **2021**, *28*, 923–935. [CrossRef]

94. Song, J.; Ni, F. NMR for the design of functional mimetics of protein-protein interactions: One key is in the building of bridges. *Biochem. Cell Biol.* **1998**, *76*, 177–188. [CrossRef] [PubMed]

95. Conicella, A.E.; Zerze, G.H.; Mittal, J.; Fawzi, N.L. ALS mutations disrupt phase separation mediated by a-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **2016**, *24*, 1537–1549. [CrossRef] [PubMed]

96. Lin, Y.; Zhou, X.; Kato, M.; Liu, D.; Ghaemmaghami, S.; Tu, B.P.; McKnight, S.L. Redox mediated regulation of an evolutionarily conserved cross-β structure formed by the TDP43 low complexity domain. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 28727–28734. [CrossRef] [PubMed]

97. Gu, J.; Zhou, X.; Sutherland, L.; Kato, M.; Jaczynska, K.; Rizo, J.; McKnight, S.L. Oxidative regulation of TDP-43 self-association by a β-to-α conformational switch. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2311416120. [CrossRef] [PubMed]

98. Lim, L.; Wei, Y.; Lu, Y.; Song, J. ALS-causing mutations significantly perturb the self-assembly and interaction with nucleic acid of the intrinsically disordered prion-like domain of TDP-43. *PLoS Biol.* **2016**, *14*, e1002338. [CrossRef] [PubMed]

99. Dang, M.; Lim, L.; Kang, J.; Song, J. ATP biphasically modulates LLPS of TDP-43 PLD by specifically binding arginine residues. *Commun. Biol.* **2021**, *4*, 714. [CrossRef]

100. Dang, M.; Li, T.; Zhou, S.; Song, J. Arg/Lys-containing IDRs are cryptic binding domains for ATP and nucleic acids that interplay to modulate LLPS. *Commun. Biol.* **2022**, *5*, 1315. [CrossRef] [PubMed]

101. Schmidt, H.B.; Barreau, A.; Rohatgi, R. Phase separation-deficient TDP43 remains functional in splicing. *Nat. Commun.* **2019**, *10*, 4890. [CrossRef] [PubMed]

102. 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]

103. Dang, M.; Li, T.; Song, J. ATP and nucleic acids competitively modulate LLPS of the SARS-CoV2 nucleocapsid protein. *Commun. Biol.* **2023**, *6*, 80. [CrossRef]

104. Kim, T.H.; Payliss, B.J.; Nosella, M.L.; Lee, I.T.; Toyama, Y.; Forman-Kay, J.D.; Kay, L.E. Interaction hot spots for phase separation revealed by NMR studies of a CAPRIN1 condensed phase. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2104897118. [CrossRef] [PubMed]

105. Kota, D.; Prasad, R.; Zhou, H.X. Adenosine Triphosphate Mediates Phase Separation of Disordered Basic Proteins by Bridging Intermolecular Interaction Networks. *J. Am. Chem. Soc.* **2024**, *146*, 1326–1336. [CrossRef] [PubMed]
106. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230. [CrossRef] [PubMed]
107. Baldwin, R.L. New directions in the study of Peptide h-bonds and Peptide solvation. *Adv. Protein Chem.* **2005**, *72*, ix–xi.
108. Rose, G.D.; Fleming, P.J.; Banavar, J.R.; Maritan, A. A backbone-based theory of protein folding. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 16623–16633. [CrossRef] [PubMed]
109. Chiti, F.; Dobson, C.M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **2006**, *75*, 333–366. [CrossRef] [PubMed]
110. Song, J. Why do proteins aggregate? "Intrinsically insoluble proteins" and "dark mediators" revealed by studies on "insoluble proteins" solubilized in pure water. *F1000Research* **2013**, *2*, 94. [CrossRef] [PubMed]
111. Bukau, B.; Weissman, J.; Horwich, A. Molecular chaperones and protein quality control. *Cell* **2006**, *125*, 443–451. [CrossRef] [PubMed]
112. Sridharan, S.; Kurzawa, N.; Werner, T.; Günthner, I.; Helm, D.; Huber, W.; Bantscheff, M.; Savitski, M.M. Proteome-wide solubility and thermal stability profiling reveals distinct regulatory roles for ATP. *Nat. Commun.* **2019**, *10*, 1155. [CrossRef]
113. Hayes, M.H.; Peuchen, E.H.; Dovichi, N.J.; Weeks, D.L. Dual roles for ATP in the regulation of phase separated protein aggregates in Xenopus oocyte nucleoli. *eLife* **2018**, *7*, e35224. [CrossRef]
114. Lim, L.; Kang, J.; Song, J. Extreme diversity of 12 cations in folding ALS-linked hSOD1 unveils novel hSOD1-dependent mechanisms for Fe^{2+}/Cu^{2+}-induced cytotoxicity. *Sci. Rep.* **2023**, *13*, 19868. [CrossRef] [PubMed]
115. Gray, M.J.; Wholey, W.-Y.; Wagner, N.O.; Cremers, C.M.; Mueller-Schickert, A.; Hock, N.T.; Krieger, A.G.; Smith, E.M.; Bender, R.A.; Bardwell, J.C.; et al. Polyphosphate is a primordial chaperone. *Mol. Cell* **2014**, *53*, 689–699. [CrossRef] [PubMed]
116. Rao, N.N.; Gómez-García, M.R.; Kornberg, A. Inorganic polyphosphate: Essential for growth and survival. *Annu. Rev. Biochem.* **2009**, *78*, 605–647. [CrossRef] [PubMed]
117. Li, M.; Liu, J.; Ran, X.; Fang, M.; Shi, J.; Qin, H.; Goh, J.M.; Song, J. Resurrecting abandoned proteins with pure water: CD and NMR studies of protein fragments solubilized in salt-free water. *Biophys. J.* **2006**, *91*, 4201–4209. [CrossRef] [PubMed]
118. Thandapani, P.; O'Connor, T.R.; Bailey, T.L.; Richard, S. Defining the RGG/RG motif. *Mol. Cell* **2013**, *50*, 613–623. [CrossRef]
119. Rajyaguru, P.; Parker, R. RGG motif proteins: Modulators of mRNA functional states. *Cell Cycle* **2012**, *11*, 2594–2599. [CrossRef]
120. Hautke, A.; Voronin, A.; Idiris, F.; Riel, A.; Lindner, F.; Lelièvre, A.; Zhu, J.; Appel, B.; Fatti, E.; Weis, K.; et al. CAG-repeat RNA hairpin folding and recruitment into nuclear speckles with a pivotal role of ATP as a cosolute. *J. Am. Chem. Soc.* **2023**, *145*, 9571–9583. [CrossRef]
121. Baldwin, R.L. Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 13052–13056. [CrossRef] [PubMed]
122. Levy, Y.; Onuchic, J.N. Water mediation in protein folding and molecular recognition. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 389–415. [CrossRef]
123. Chong, S.H.; Ham, S. Distinct role of hydration water in protein misfolding and aggregation revealed by fluctuating thermodynamics analysis. *Acc. Chem. Res.* **2015**, *48*, 956–965. [CrossRef]
124. Laage, D.; Elsaesser, T.; Hynes, J.T. Water dynamics in the hydration shells of biomolecules. *Chem Rev.* **2017**, *117*, 10694–10725. [CrossRef] [PubMed]
125. Holland, J.H. *Complexity: A Very Short Introduction, Very Short Introductions*; Oxford Academic: Oxford, UK, 2014.