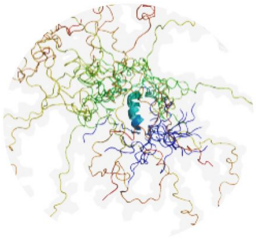
GRK 2467



Self-assembly of intrinsically disordered proteins into amyloid fibrils and liquid condensates

Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der Naturwissenschaftlichen Fakultät I

des Instituts für Biochemie und Biotechnologie

der Martin-Luther-Universität Halle-Wittenberg

MARTIN-LUTHER-UNIVERSITÄT
HALLE-WITTENBERG

vorgelegt von

Luca M. Lauth

Erstgutachter - Prof. Dr. Jochen Balbach

Zweitgutachter - Prof. Dr. Stefan Hüttelmaier

Externer Gutachter - Assistant Prof. Dr. Claudio Iacobucci

Verteidigung erfolgte am 21.03.2024

Contents

I Abbreviations

Abs.	absorption
amp	ampicillin
AUC	analytical ultracentrifugation
b	length of a Kuhn segment
CD	circular dichroism
DMSO	dimethyl sulfoxide
DSBU	disuccinimidyl dibutyric urea
DTT	dithiotreitol
DMEM	Dulbeccos Modified Eagle Medium
E. coli	*Escherichia coli*
EDC	N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride
EDTA	ethylenediaminetetraacetic acid
FUS	fused in sarcoma
GAG	glycosaminoglycan
H3C-protease	human rhinovirus (HRV) 3C protease
His_6	hexahistidine
ICG	inter-chromatin granule
IDP	intrinsically disordered protein
IDR	intrinsically disordered region
IMAC	immobilized metal affinity chromatography
IPTG	isopropyl-ß-D-thiogalactopyranoside
ITC	isothermal titration calorimetry
L	contour length
kan	kanamycin
κ	rate of secondary nucleation and growth
KO	knockout
λ	rate of primary nucleation and growth
LB	lysogeny broth
LC	liquid chromatography
LCD	low complexity domain
LLPS	liquid-liquid phase separation
LSPS	liquid-solid phase separation
MALAT1	metastasis associated lung adenocarcinoma transcript 1
MD	molecular dynamics
MRW	mean residue weight
MS	mass spectrometry
MSA	multiple sequence alignment
MW	molecular weight
MWCO	molecular weight cut-off
Ni^{2+}-NTA	nickel-nitrilotriacetic complex
NMR	nuclear magnetic resonance
NS	nuclear speckle
OD	optical density

OVCAR-3	human ovarian cancer cell line
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PLD	prion-like domain
PTH	parathyroid hormone
PTH1R	PTH receptor 1
PTH2R	PTH receptor 2
RBP	RNA binding protein
R_c	cross-sectional radius
R_g	radius of gyration
rpm	rounds per minute
RFP	red fluorescent protein
RRM	RNA recognition motif
RT	room temperature
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SR proteins	serine/arginine-rich proteins
SRSF1	serine-arginine-rich splicing factor 1
sulfo-NHS	N-Hydroxysulfosuccinimid
sulfo-SDA	sulfosuccinimidyl 4,4'-azipentanoate
TCEP	Tris(2-carboxyethyl)phosphine
TEM	transmission electron microscopy
ThT	Thioflavin T
UV	ultraviolet
WAXS	wide angle X-ray scattering

II Summary

Protein self-assembly plays a vital role in a variety of physiological processes, including stress response, intracellular transport and gene expression regulation. Protein self-assembly gives rise to rich phase behaviors resulting in diverse condensed matter states with varying physical properties. Protein condensates may, for example, have solid properties exhibiting a high degree of structural order, properties that are characteristics of crystals and amyloid fibrils. In contrast, liquid-like protein condensates exist which typically exhibit fast molecular rearrangements. The physical properties of protein condensates determine their physiological functions and are dependent on the protein species involved. In this study, the self-assembly of two largely disordered proteins, the parathyroid hormone (PTH) and the serine/arginine rich splicing factor 1 (SRSF1), was characterized.

PTH is a peptide hormone with an α-helical N-terminus that is sufficient and necessary for full activity at its principal receptor, PTH1R. The disordered C-terminal region of PTH, however, remains largely uncharacterized. Being a peptide hormone, PTH is presumably stored in the form of functional amyloid fibrils. The proposed storage site for these amyloid fibrils in PTH-producing cells are acidic secretory granules. However, mature PTH is considerably positively charged under acidic conditions which has been found to impede unassisted self-assembly into amyloid fibrils before. In this study, the role of the polyanion heparin on promoting fibril formation of PTH was investigated. Using a combination of isothermal titration calorimetry (ITC), circular dichroism (CD) spectroscopy, nuclear magnetic resonance (NMR) spectroscopy, X-ray diffraction techniques and fluorescence-based assays, it was demonstrated that heparin binds PTH with submicromolar affinity and facilitates its conversion into fibrillar seeds, enabling rapid formation of amyloid fibrils under acidic conditions. In absence of heparin, PTH remained in a soluble monomeric state. Heparin-like surfaces may therefore be required *in vivo*

to convert PTH efficiently into fibrillar deposits.

Fibril formation of PTH was also found to be impacted by its disordered C-terminal region. Here, molecular dynamics (MD) simulations and cross-linking mass spectrometry (XL-MS) were used to identify long-range intra-chain contacts between the fibrillation-prone N-terminal region and the disordered C-terminus. The C-terminal region may regulate the propensity of PTH to undergo fibrillation through an effect of surface occlusion.

Liquid-like protein condensation was investigated on SRSF1, a protein that incorporates into nuclear speckles (NS) which are biomolecular condensates in the nucleus. SRSF1 is an RNA binding protein (RBP) with an extended disordered low complexity domain (LCD) that is highly enriched in arginine/serine residues (= RS domain). Extensive serine phosphorylations within the LCD have been shown to affect functional aspects. However, the condensation behavior as a function of differential phosphorylations has not been described, yet. Following a divide-and-conquer approach, several phospho-mimicking mutants of the isolated SRSF1 LCD were designed. A combination of *in cellulo* and *in vitro* experiments was employed to shed light on the phosho-state dependent subcellular localization and condensation behavior of these SRSF1 LCD variants. While all variants incorporated into NS, the condensation behavior was found to significantly differ among the variants suggesting that differential phosphorylation might be an effective means to tune the phase behavior of arginine/serine rich proteins and thus regulate the formation of NS.

III Zusammenfassung

Die Selbstorganisation von Proteinen spielt eine wichtige Rolle bei einer Reihe von physiologischen Prozessen, einschließlich Stressreaktion, intrazellulärem Transport und Regulierung der Genexpression. Die Selbstorganisation von Proteinen resultiert in einem vielfältigen Phasenverhalten, das zu verschiedenen Zuständen kondensierter Materie mit unterschiedlichen physikalischen Eigenschaften führt. Proteinkondensate können bespielsweise feste Eigenschaften mit einem hohen Maß an struktureller Ordnung aufweisen, Eigenschaften, die für Kristalle und Amyloidfibrillen charakteristisch sind. Im Gegensatz dazu gibt es flüssigkeitsähnliche Proteinkondensate, die typischerweise eine schnelle molekulare Umordnung aufweisen. Die physikalischen Eigenschaften von Proteinkondensaten bestimmen ihre physiologischen Funktionen und sind von den beteiligten Proteinspezies abhängig. In dieser Studie wurde die Selbstorganisation zweier weitgehend ungeordneter Proteine, des Parathormons (PTH) und des Serin-Arginin-reichen Spleißfaktors 1 (SRSF1), charakterisiert.

PTH ist ein Peptidhormon mit einem α-helikalen N-Terminus, der für die volle Stimulation seines primären Rezeptors, dem PTH1R, ausreichend und notwendig ist. Der ungeordnete C-terminale Bereich von PTH ist dagegen noch weitgehend uncharakterisiert. Als Peptidhormon wird PTH vermutlich in Form von funktionellen Amyloidfibrillen gespeichert. Der wahrscheinliche Speicherort für diese Amyloidfibrillen in PTH-produzierenden Zellen sind saure sekretorische Vesikel. Unter sauren Bedingungen ist PTH jedoch deutlich positiv geladen, was die Selbstorganisation zu Amyloidfibrillen in früheren *in vitro* Experimenten behindert hat. In dieser Arbeit wurde die Rolle des Polyanions Heparin bei der Bildung von PTH-Fibrillen untersucht. Mithilfe einer Kombination aus isothermer Titrationskalorimetrie (ITC), Zirkulardichroismus(CD)-Spektroskopie, Kernspinresonanz(NMR)-Spektroskopie, Röntgenbeugungstechniken und fluoreszenzbasierten Assays wurde gezeigt, dass Heparin PTH mit submikromolarer Affinität bindet und dessen Umwandlung in

fibrilläre Keime erleichtert, was eine schnelle Bildung von Amyloidfibrillen unter sauren Bedingungen ermöglicht. In Abwesenheit von Heparin verblieb PTH in einem löslichen und monomeren Zustand. In zellulärer Umgebung sind Heparin-ähnliche Oberflächen daher möglicherweise erforderlich, um PTH effizient in fibrilläre Ablagerungen umzuwandeln. Es wurde ferner festgestellt, dass die Fibrillenbildung von PTH auch durch seine intrinsisch ungeordnete C-terminale Region beeinflusst wird. Mittels Molekulardynamik(MD)-Simulationen und Cross-Linking-Massenspektrometrie (XL-MS) wurden intra-Peptid Kontakte zwischen der zur Fibrillierung neigenden N-terminalen Region und dem ungeordneten C-Terminus identifiziert. Die C-terminale Region könnte die Neigung von PTH zur Fibrillierung durch einen Effekt der Oberflächenokklusion beeinflussen.

Flüssige Proteinkondensation wurde an SRSF1 untersucht, einem Protein, das in nukleären Speckle lokalisiert, welche biomolekulare Kondensate im Zellkern darstellen. SRSF1 ist ein RNA-bindendes Protein (RBP) mit einer ungeordneten Domäne niedriger Komplexität (LCD), die stark mit den Aminosäuren Arginin/Serin angereichert ist (= RS-Domäne). Es hat sich gezeigt, dass zahlreiche Serin-Phosphorylierungen innerhalb der LCD funktionelle Aspekte des Proteins beeinflussen. Das Kondensationsverhalten in Abhängigkeit der unterschiedlichen Phosphorylierungs-Zustände ist jedoch noch nicht beschrieben worden. Mehrere Protein-Konstrukte der isolierten SRSF1 LCD wurden generiert, die verschiedene Phosphorylierungs-Zustände des Proteins nachahmen. Zell-basierte Experimente und solche *in vitro* wurde eingesetzt, um die vom Phosphorylierungs-Zustand abhängige subzelluläre Lokalisierung und um das Kondensationsverhalten dieser SRSF1 LCD Varianten zu untersuchen. Während alle Varianten in nukleären Speckle lokalisierten, wurde ein stark unterschiedliches Kondensationsverhalten zwischen den Varianten beobachtet, was darauf hindeutet, dass dynamische Phosphorylierungen bzw. Dephosphorylierungen ein wirksames Mittel sein könnte, um das Phasenverhalten von Arginin/Serin-reichen Proteinen zu beeinflussen und somit die Bildung von nukleären Speckle zu regulieren.

IV Publications

Lauth, L.M., Voigt, B., Bhatia, T., Machner, L., Balbach, J. and Ott, M. (2022), Heparin promotes rapid fibrillation of the basic parathyroid hormone at physiological pH. FEBS Lett. 596: 2928-2939. https://doi.org/10.1002/1873-3468.14455

Piersimoni, L., Abd el Malek, M., Bhatia, T.,..., Lauth, L.M. et al. (2022), Lighting up Nobel Prize-winning studies with protein intrinsic disorder. Cell. Mol. Life Sci. 79, 449. https://doi.org/10.1007/s00018-022-04468-y

V Self-assembly of intrinsically disordered proteins into amyloid fibrils and liquid condensates

1 Introduction

1.1 Intrinsically disordered proteins

Protein disorder has emerged as a feature of proteins that defies the traditional structure-function paradigm governing protein science (see Figure 1) [1]. Unlike proteins that adopt well-defined three-dimensional structures, IDPs lack a stable tertiary structure under physiological conditions [2]. Instead, they exist as highly dynamic ensembles, inter-converting between a range of conformational states. This inherent disorder is now being recognized as a prevalent and functional feature in a wide array of biological processes [3–5].

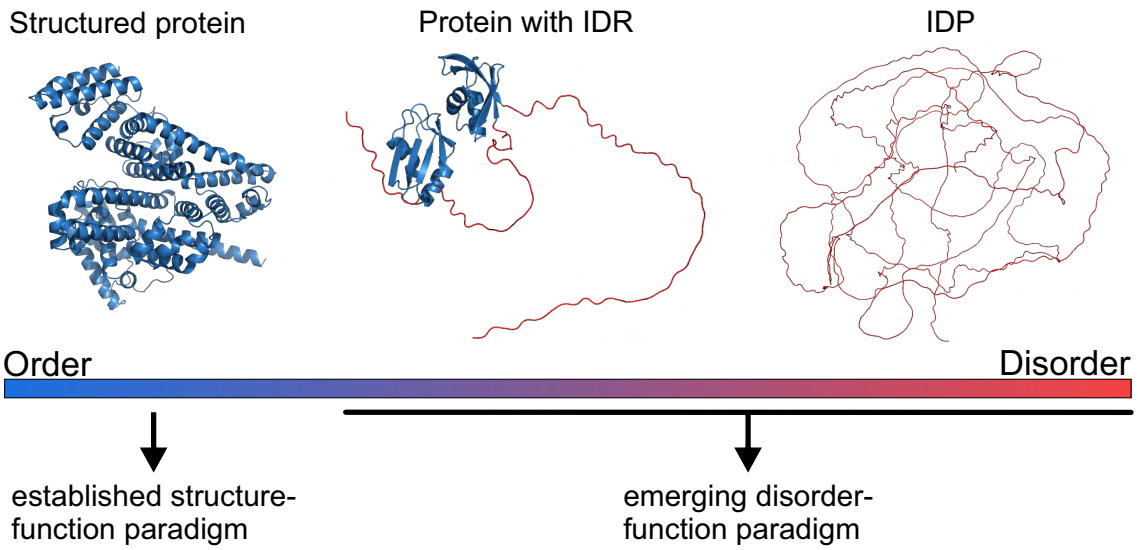


Figure 1: **Order and disorder in proteins.** IDPs and IDRs extend the tranditional structure-function paradigm by the disorder-function paradigm.

It is not long ago that protein disorder was associated with artifacts and

protein misfolding. Globular proteins with well-defined structures were the main focus in early structural biology research. Only since the late 20th century, the understanding of protein disorder as a significant biological feature was propelled by pionneering work by Keith Dunker, among others [6–9].

Since then, IDPs have gained traction in the life sciences: Protein disorder is increasingly appreciated as a crucial feature in the eukaryotic proteome. Over 50% of the human proteome is disordered including fully intrinsically disordered proteins (IDPs) or proteins containing intrinsically disordered regions (IDRs) [10]. This high frequency of protein disorder does not seem to be a random product of evolution. In a study concerning molecular evolution, researchers found that protein disorder is subject to strong selective pressure suggesting that it plays key roles in biological processes [11]. Indeed, there is accumulating evidence that IDPs and proteins containing IDRs are crucial components in a myriad of biological processes ranging from transcriptional regulation and signal transduction [4] to cellular transport [12] and molecular recognition [13]. In fact, protein disorder was recently highlighted by Piersimoni and colleagues to be an essential feature in numerous Nobel Prize-awarded scientific discoveries [14].

Protein disorder enables dynamic and context-dependent interactions with multiple partners [15]. This is ensured by characteristic attributes that provide a remarkable level of control over cellular processes: The presence of molecular recognition features (MoRFs), which undergo disorder-to-order transition upon partner binding [16]; a high flexibility, allowing for promiscuous interactions with various partners [17]; the capability to bind partners with high specificity but moderate affinity, enabling spontaneous and fast dissociation [18]; extremely fast association rates aiding rapid signal transduction [18]; and extensive possibilities for post-translational modifications (PTMs), providing means to tune functions of IDPs or IDRs over space and time through changes in their physicochemical properties [19]. These characteristics render IDPs and proteins containing IDRs especially useful in cellular signaling and for the regulation of various cellular

processes.

Several sequence properties have been observed to be common in protein disorder. Uversky and collegues have proposed that protein disorder is dependent on two sequence features: High net charge and low hydrophobicity [2]. A high net charge at neutral pH corresponds to a non-neutral isoelectric point promoting solubility and preventing aggregation under physiological conditions. Low hydrophobicity again promotes solubility and prevents the hydrophobic collapse which generally is a major driving force in protein folding. As a result, polar and charged residues typically make up a large fraction in disordered proteins. This tendency of IDPs or IDRs to have a bias towards a subset of the 20 canonical amino acids results in ”low complexity”. However, the term ”low complexity” is only loosely defined and the extent of sequence bias can vary greatly between different disordered sequences [20].

Due to the inherent lack of a stable three-dimensional structure, protein disorder can be challenging to characterize with traditional biophysical techniques. X-ray crystallography, for example, typically fails to resolve protein disorder due to their poor crystallization behavior [21]. On the other hand, X-ray diffraction displays a powerful tool to obtain structural information of an ensemble of proteins in solution [22]. While traditional structure elucidation with nuclear magnetic resonance (NMR) spectroscopy can be challenging due to low chemical shift dispersion, advancements in the technique allow to gain valuable insights into the conformational dynamics of protein disorder [23]. Novel fluorescence spectroscopic techniques even allow for investigations on single-molecule level [24].

Computational techniques have advanced, as well. Various prediction tools have been developed, some of which score the probability of peptide sequences to be disordered [25], to undergo phase separation [26] and to be binding sites [27]. Molecular dynamics (MD) simulations running with novel force-fields can provide insights into the spatiotemporal properties of protein disorder [28]. Here, an

especially convincing strategy is to integrate MD simulations with experimentally derived data [29].

Which methods to use for investigations, is also dependent on the state of the disordered protein. Under physiological conditions, IDPs and proteins containing IDRs rarely stay isolated. They typically interact with various binding partners and/or organize into higher-order assemblies. In the following, this collective behavior of proteins will be approached from a physicochemical perspective. Afterwards, the peculiarities of protein disorder will be discussed in this context.

1.2 Organization of proteins into higher-order assemblies

Understanding the complex behavior of cells is a major goal in the life sciences. For a long time, the discussion has been dominated by the famous paradigm of molecular biology: genes encode proteins and each protein performs unique functions. Changes in cell functions occur by modulating the activity of these proteins through either expression control, post-translational modifications (PTMs) or protein degradation [30]. Many individual proteins have been extensively studied so far to reveal their unique molecular functions. However, the mechanistic understanding of individual proteins is not sufficient to describe the complexity of all cellular processes: Many biological macromolecules do not operate in isolation but rather exert functions collectively in higher-order assemblies. There is accumulating evidence on how this collective behavior affects various aspects of living cells, including metabolism, signaling, intracellular organization and stress response.

Phase behavior of colloidal systems

Self-assembly of proteins is a well-recognized phenomenon which can be approached physicochemically by studying their phase behavior. Thermodynamically, phase transitions are driven by the minimization of the global free energy of the

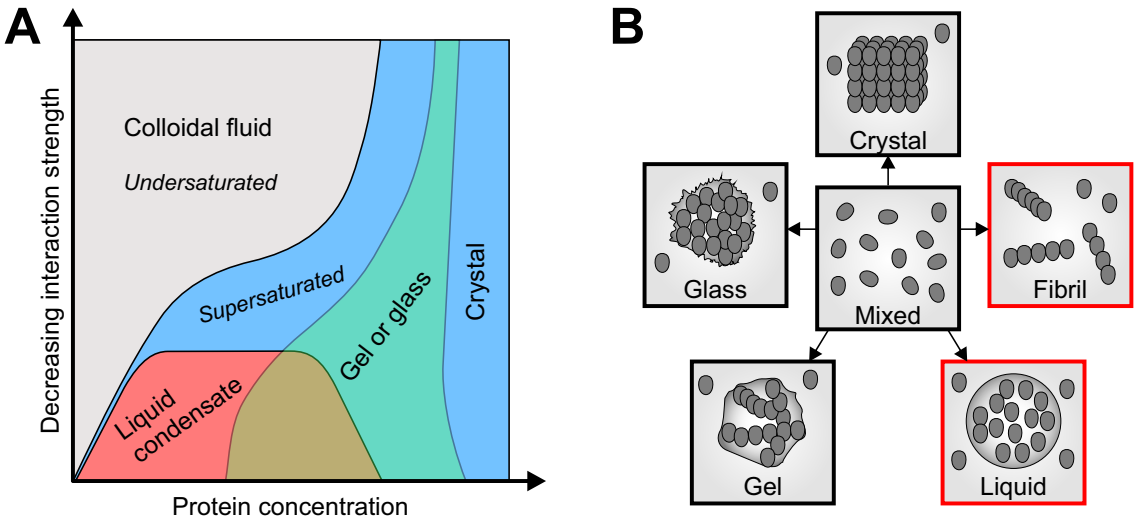


Figure 2: **Physical states of protein assemblies.** A: A protein phase diagram displaying various physical states is depicted. The phase behavior may vary between different proteins and some proteins may not access all states. B: Protein assemblies vary in shape and physical properties. Fibrils and liquid condensates (both highlighted in red) are the main focus in this work. Adapted from [30].

system [31]. In this context, the free energy of mixing, ΔG_m, is a relevant term to describe the solution properties of a given protein. ΔG_m is related to the enthalpy, ΔH_m, and entropy, ΔS_m, of mixing as given in Equation 1:

$$\Delta G_m = \Delta H_m - T\Delta S_m \tag{1}$$

It follows that minimization of the system involves enthalpic contributions which comprise the interaction potentials between the components of the system while entropic contributions account for the degrees of freedom available to the components of the system [32].

As long as the system obeys the thermodynamic rules, the phase behavior of proteins can be very complex with several potentially accessible phase states. Here, the concept of colloids (from the greek ”kolla”, meaning glue) provides a useful framework to describe the phase behavior of particles [33]. Research on the phase behavior of globular model proteins such as lysozyme and γ-crystallins have revealed that they can access multiple phase states (see Figure 2A) [34–37]. In an undersaturated solution, proteins are soluble and a well-mixed colloidal fluid is formed. However, when the concentration exceeds the solubility limit of

the protein (supersaturated solution), the likelihood of protein condensation is drastically increased [30]. Various condensed matter states can be adopted with very different physical properties (see Figure 2B, Table 1) [38].

Table 1: **Physical properties of protein condensates.**

Protein phase state	**Properties**
Liquids	order over short distances, dynamic structures, quick rearrangement of molecules, no positional memory
Solids	order over short and long distances, molecules are spatially confined by neighboring molecules, positional memory
Liquid condensate	forms through demixing of supersaturated solutions, metastable, may form crystals, gels or glasses over time
Gel	percolated system-spanning and cross-linked network, no flow in steady-state
Glass	amorphous, non-equilibrium solid, very slow diffusion
Crystal	highly ordered structure, equilibrium solid
Fibril	one-dimensional pseudo-crystal

For example, the diffusion rates can vary greatly between different phase states with associated biological implications: Liquid condensates allow for free diffusion and may facilitate biochemical reactions while solid assemblies like glasses and crystals restrict diffusion and thus suppress reactions.

Importantly, the formation of protein condensates is not only dependent on protein concentration but also external factors such as solution temperature, pH and ionic strength [30]. Typically, interactions between proteins are relatively weak. For globular proteins such as lysozyme, researchers found a range of possible condensate species that exchange with the surrounding medium dynamically [35, 39]. Here, the interactions within the protein condensates are sufficiently weak and accessible for external factors and thus these assemblies are easily reversed. Hence, protein dilution or changes in pH and ionic strength can lead to the disintegration of these condensates [35].

However, in some cases, proteins associate so strongly that the assembled species can be considered irreversible on practical time scales. These assemblies have solid properties and may be amorphous aggregates with gel- or glass-like

properties or highly ordered fibrillar species (crystal-like). In these cases, the interactions are too strong and not overly sensitive towards external factors. Hence, break up of these assemblies requires harsh conditions including the use of denaturants like urea and sodium dodecyl sulfate (SDS) [30].

Phase behavior of (bio)polymers

In the theoretical colloidal science, the particles under investigation are assumed to be static bodies with mostly spherical or ellipsoidal shape. This approximation holds up reasonably well for globular proteins, for which the early phase diagrams were experimentally determined (see above). IDPs and proteins containing IDRs, however, deviate considerably from the colloidal nature [33]. Instead, they are dynamic structures that fluctuate between multiple conformational states. Hence, the behavior of protein disorder in solution can be more accurately described using a conceptual framework provided by the polymer physicists Paul Flory and Maurice Loyal Huggins [31, 40]. In the Flory-Huggins model, the researchers described the thermodynamics of polymer solutions. To conceptualize, they modelled polymer solutions as an infinite lattice in which each space is occupied by either a polymer bead or a solvent molecule (see Figure 3A). They introduced the Flory parameter, χ, to quantify the ratio of polymer-polymer and polymer-solvent interactions as given in Equation 2:

$$\chi = \frac{2u_{ps} - u_{ss} - u_{pp}}{2k_B T} \tag{2}$$

Here, the terms u_{xy} define the energies for interactions between polymer (subscript p) and solvent (subscript s) species. $k_B T$ quantifies the thermal energy at a given temperature T. According to the equation, the value of χ quantifies the energetic penalty of lattice spaces adjacent to a polymer bead being occupied by a solvent molecule - the larger χ, the higher this energetic penalty.

For $\chi > 0$, the polymer is in a so-called ”poor solvent”, meaning that

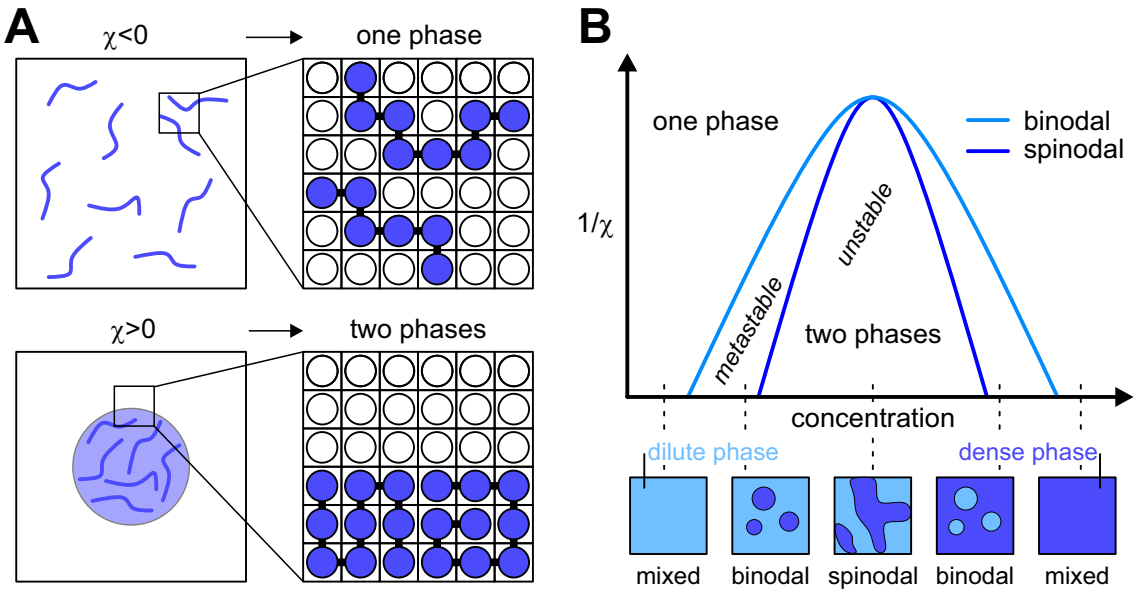


Figure 3: **Flory-Huggins model for polymer solutions.** A: Polymer solutions are modelled as an infinite lattice with spaces occupied by either polymer beads (violet) or solvent molecules (white). χ determines whether polymers mix with the solvent (upper scheme) or phase separate (lower scheme). B: Phase separation is concentration dependent. Below the spinodal curve, phase separation must occur via spinodal decomposition. In the area between the binodal and spinodal curves, phase separation requires nucleation to occur.

polymer-polymer interactions become energetically favorable over polymer-solvent interactions. In this case, de-mixing into a phase separated state with a polymer dense phase and a polymer-depleted phase is favored. According to the Flory-Huggins theory, phase separation can occur via two regimes - metastable binodal phase separation and spinodal decomposition (see Figure 3B). Binodal phase separation requires a nucleation process to occur. Spinodal decomposition, on the other hand, is a near-instantaneous process and must occur spontaneously [41].

The Flory-Huggins model implies that the entropy of mixing is invariably positive ($0 < \Delta S_m$) as indicated by the decreased availability of degrees of freedom for lattice placement in the phase separated state (see Figure 3A). Accordingly, the entropically favored state of mixing needs to be overcome by sufficiently positive enthalpic contributions ($\Delta H_m > T\Delta S_m$) in order for phase separation to occur. While the Flory-Huggins model assumes simple homopolymers, proteins/peptides are typically very complex heteropolymers consisting of up to 20 amino acids.

This allows for the occurence of diverse attractive protein-protein interactions. Apart from covalent bonds such as the disulfide bridge, proteins can interact through relatively weak non-covalent interactions including hydrogen bonds and charge-charge, dipole-dipole, cation-π and π-π interactions (see Figure 4) [42].

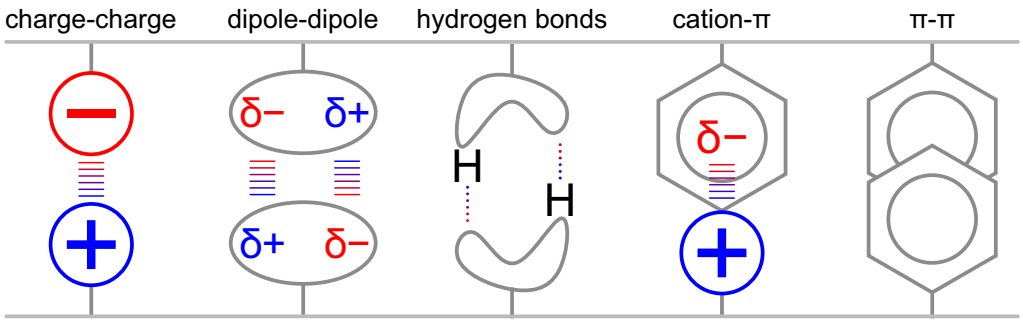


Figure 4: **Possible non-covalent interactions between amino acids.**

Even though sequence-specific interactions are neglected, the basic Flory-Huggins model was successfully used to describe the phase behavior of the germ granule protein DDX4 [43]. Additionally, more recently developed extensions of the model allow to account for electrostatic interactions [44] and charge patterning [45], making the model more versatile in the context of heteropolymers.

Even though life scientists greatly benefit from the strong theoretical background of phase behaviors provided by physicists, the significance of protein condensates in cellular systems remained largely unclear. This has changed more recently with the increasing realization that various membraneless organelles are biomolecular condensates forming through a process termed liquid-liquid phase separation (LLPS). As mentioned above, protein phase separation does not necessarily yield liquid-like phases but can also result in solid phases via a process termed precipitation or liquid-solid phase separation (LSPS). Solid condensed phases in cells include amyloids which are highly ordered structures with several physiological and pathophysiological roles associated to them. They form through nucleated processes and require metastable mixing conditions meaning that the mixed system is globally unstable but locally stable. In the following, both, solid amyloid

fibrils and liquid-like protein condensates will be introduced.

1.3 Liquid condensates

While the basic thermodynamic principles for phase separation to occur were briefly introduced above, the process of LLPS in particular requires further special conditions.

Principles of liquid-liquid phase separation

The term ”liquid” generally refers to a type of phase in which no long-range order exists in the molecular structure. Moreover, liquids do not possess a defined shape nor shape-memory, in other words, they flow [30]. In biological systems, many biomolecular condensates exhibit these characteristics [46–50].

Over the last years, great progress has been made to identify sequence-based and structural determinants in peptides/proteins that drive LLPS. In their seminal paper, Rosen and coworkers have demonstrated that biomolecular LLPS is mediated by multivalent interactions, i.e. interactions with multiple binding sites per macromolecule [51]. More specifically, they observed that pairs of interacting proteins underwent LLPS when their concentrations passed a certain threshold and the stoichiometries were suitable. Importantly, both protein partners encoded a variable number of repeated modular binding domains and molecular recognition motifs, respectively. They found that with more available interaction sites per protein, i.e. with higher multivalence, the propensity for LLPS was increased, as interpreted from a reduced critical concentration for de-mixing.

Why is multivalence so important for the formation of liquid condensates? Thermodynamically, multivalent interactions are advantageous as they increase the avidity of the involved proteins. In other words, the accumulated strength of the individual interactions favors protein-protein over protein-solvent interactions. Moreover, in poor solvent conditions - a situation that applies to many proteins in water - complex formation further reduces solubility and thereby enhances the propensity for phase separation. This is due to a lower entropic penalty

associated with the confinement of a complex into a condensed phase versus the confinement of its individual components [46].

Equally important is the fact that multivalent interactions allow for the development of a three-dimensional network that spans the protein dense phase. The degree of multivalence is correlated with the cross-link density of the network with higher degrees enabling for a higher cross-link density. Notably, there are reported cases of proteins which attain multivalence through self-assembly [52–54].

In biological systems, IDPs and proteins containing IDRs are often main drivers for the formation of biomolecular condensates [55]. Similar to multidomain proteins, protein disorder can provide multiple weak interaction sites which can be single residues or short linear motifs (SLiMs) of usually 3-10 residues in length [27]. Importantly, these multivalent interactions are often transient, allowing for fast structural reconfigurations [56].

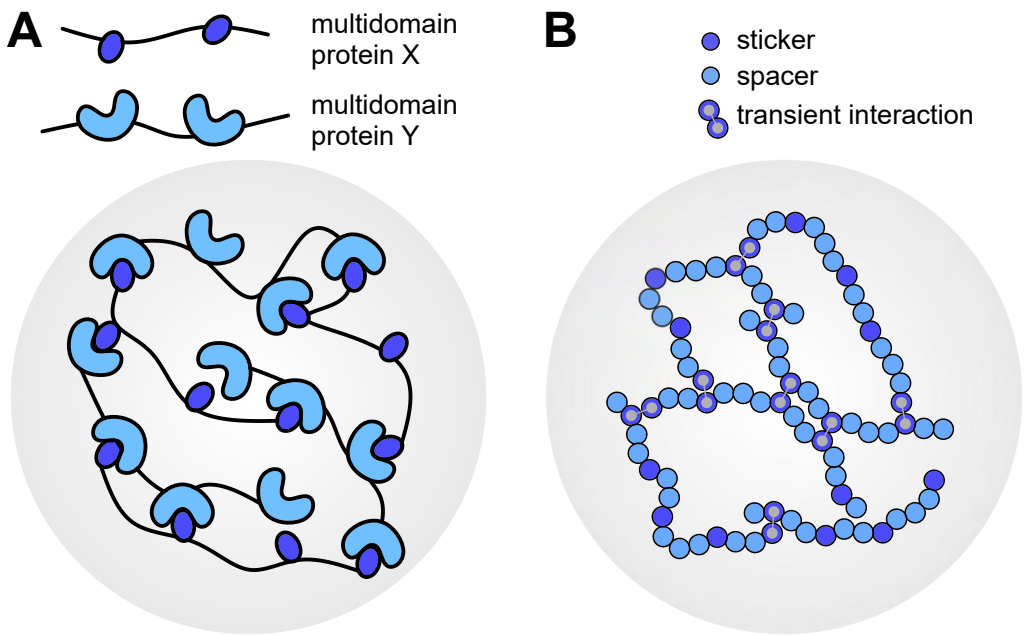


Figure 5: **Stickers and spacers model.** The stickers and spacers model can be applied to multidomain proteins (A) and IDRs (B) to conceptualize the LLPS behavior.

To make multivalence-mediated LLPS of proteins conceptually and computationally tractable, Pappu and coworkers have developed the stickers and spacers

framework based on the theory of associative polymers (see Figure 5) [57]. In this simplistic model, proteins are assumed to consist of two components, stickers and spacers. Stickers primarily contribute to the interaction potential by serving as adhesive elements. Spacers, on the other hand, connect the stickers and determine the interactions with the solvent.

The stickers and spacers framework as a concept describing multidomain proteins (see Figure 5A) can be applied to IDPs/IDRs (see Figure 5B), as well. Like in multidomain proteins, their propensity to undergo LLPS is encoded in the interaction strength between stickers, their valence and patterning, and the physicochemical properties of the spacers [55]. Importantly, in order to employ the stickers and spacers model, both components need to be identified in the disordered sequence. This can be achieved by mutational studies changing putative sticker sites and assessing its effect on the phase behavior. This approach was used to demonstrate the importance of tyrosine (Tyr) and arginine (Arg) residues in the phase separation of FET family proteins such as FUS (fused in sarcoma) [58]. However, it is important to note that spacer sequences also influence the phase behavior of proteins and therefore, this approach may not unambiguously identify stickers. A more accurate approach was recently demonstrated by Mittag and coworkers, who identified aromatic residues in prion-like domains (PLDs) as stickers by monitoring their chain compaction as a function of valence (number of aromatic residues per chain) [59]. They further observed that the patterning of aromatic residues determined whether proteins underwent LLPS or assembled into amorphous aggregates, highlighting the importance of effective spacer regions in mediating LLPS.

In case of polyampholytic proteins, patterning of oppositely charged stickers modulates the phase behavior. Mittal and coworkers showed that the propensity of the P granule protein LAF-1 to undergo LLPS is enhanced with increased charge segregation [60]. Blocks of high net charge therefore appear to exhibit increased stickiness, an interpretation that is supported by the observed collapse of mixed-

charge peptides with a high degree of charge segregation [61]. However, it is important to note that evenly distributed charge patterns do not necessarily abolish LLPS behavior but rather reduce the propensity for this process [60].

Thus far, the given examples of proteins that undergo LLPS *in vitro* can do so in isolation, meaning that no interactions with other macromolecular species are required. This particular behavior is referred to as homotypic LLPS. However, there are protein systems that only undergo LLPS in presence of specific counterparts. In this heterotypic behavior, at least two components condense via a process called complex coacervation [42]. *In vitro*, complex coacervation has been observed for systems of oppositely charged polyelectrolytes. For example, Deniz and coworkers found that Arg-rich polyelectrolytes formed liquid condensates when a charge-neutralizing amount of RNA was added [62]. Similarly, the Arg-rich protamine underwent complex coacervation with the polyanionic hyaluronic acid [63].

In vitro characterization of the phase behavior of a protein can be approached in several ways. Phase separations can be directly observed via microscopy, where liquid condensates usually appear as droplets [64]. Since these droplets scatter light, phase separated solutions appear turbid giving rise to an apparent absorbance [64]. This phenomenon may be used in spectroscopic measurements to distinguish between mono-phasic and bi-phasic solutions [65]. When screening the phase behavior over a range of protein concentrations and external factors such as temperature, pH and ionic strength, a phase diagram can be generated. Moreover, the material properties of the protein dense phase can be monitored using an array of different techniques including fluorescence recovery after bleeching (FRAP), fluorescence correlation spectroscopy (FCS) and atomic force microscopy (AFM) [41].

These *in vitro* studies greatly advance our mechanistic understanding of protein phase behavior. As an important branch of research, this accelerates the emerging field concerning biomolecular condensates which arise intracellularly in the form

of various membraneless organelles.

LLPS underlies the formation of membraneless organelles

The intracellular space is a highly crowded environment. Typically, biological macromolecules occupy 20-30% of the total intracellular volume [66]. Of course, this significant fraction is not occupied by only a few highly concentrated macromolecular species but rather many low concentrated ones. As a result, the intracellular space can be interpreted as a very complex environment. To manage this complexity, cells require means to organize their multi-component space. Compartmentalization is an effective way to spatially confine particular, often functionally related components to certain subcellular locations. In eukaryotes, compartmentalization is achieved through several organelles including mitochondria, the nucleus and the Golgi network. The given examples have in common that they are membrane-bound and thus clearly separated from the cytosol.

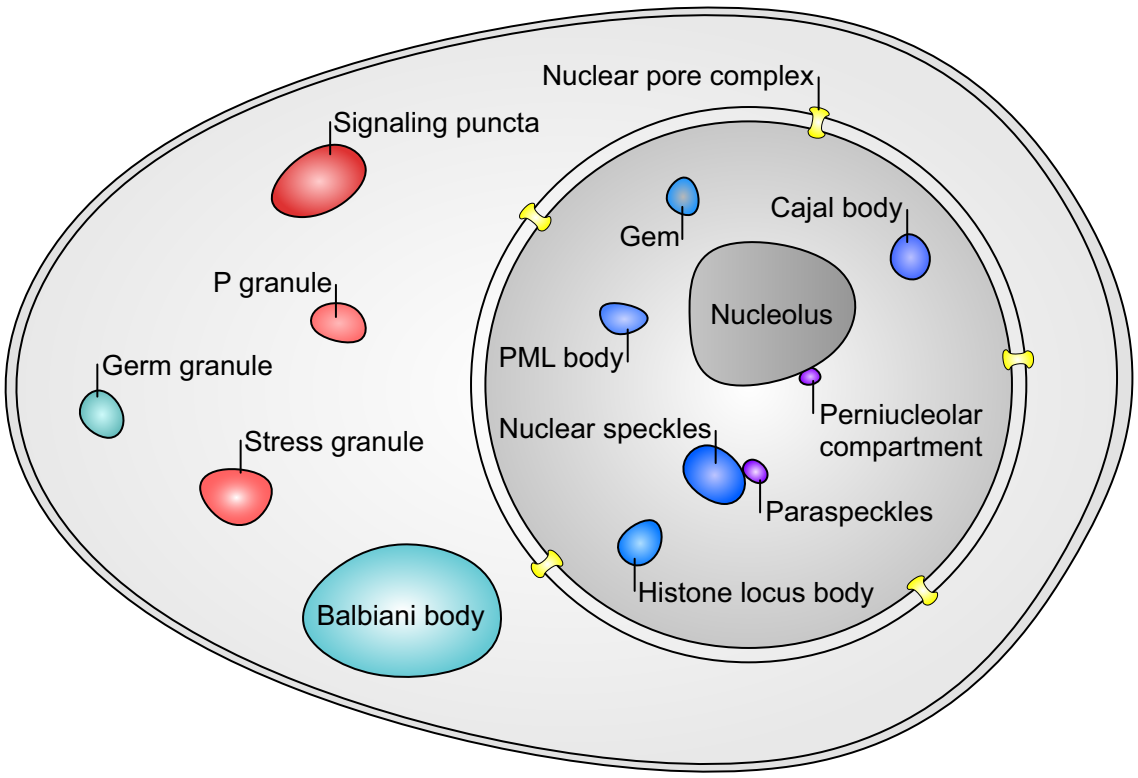


Figure 6: **Membraneless organelles in eukaryotic cells.** Numerous membraneless organelles have been identified in the cytosol and the nucleus. Some are cell type specific.

In more recent years, however, the role of membraneless organelles (MLOs) has come into focus as an additional means to organize the intracellular space. Unlike canonical organelles, MLOs lack a surrounding membrane and yet many distinct MLOs exist in eukaryotic cells and appear as granules in electron microscopy images (see Figure 6). Although the first discovery of MLOs dates back to the 18th century (the nucleolus was discovered in the 1770s) [47], the study of MLOs in context of phase separation has only gained traction more recently with the pioneering work by Hyman and coworkers [48]. They were the first to show that the germline specific P granules in *Caenorhabditis elegans* are liquid droplets that dynamically form through condensation. Since then, many MLOs have been studied with their formation and physicochemical properties in focus. Indeed, accumulated evidence suggests that LLPS underlies the formation of many MLOs which typically retain liquid-like properties [47, 48, 50, 67, 68] but may also transition to gel-like states [69].

In general, MLOs are multi-component assemblies of mesoscopic size (usually 0.2 - 2 μm) exhibiting various shapes [47]. While MLOs typically share a similar mode of formation and dynamics, they differ in subcellular location, function and composition often containing many specific proteins and RNAs and/or DNAs [46, 47]. In accordance with the theoretical background discussed in section 1.3, proteins that localize to MLOs usually form multivalent interactions with other proteins or nucleic acids, often mediated by protein disorder [70]. Indeed, proteins possessing IDRs are frequently identified to be major - and in some cases essential - components of MLOs such as the nucleolus [50], paraspeckles [71] and nuclear speckles [49]. The disordered sequences often exhibit a bias towards charged and aromatic amino acids enhancing their propensity to undergo LLPS [56, 72, 73].

The architecture of MLOs can be considered to consist of LLPS-prone proteins that make up the ”scaffold” and subsequently recruite ”clients” to form the final MLO [74]. Typical MLOs are dynamic, allowing for flux inside and outside of the

condensate. This enables MLOs to reversibly form and dissolve in response to several stimuli such as osmotic and thermal stress. Thus, dynamic MLOs are in constant exchange with their surroundings and therefore avoid thermodynamically stable accumulation which distinguishes them from amorphous protein aggregates [47]. Yet, it is important to note that biomolecular condensates may also serve as intermediate states promoting aberrant LSPS that might be associated with disease [75].

The organization of biological macromolecules in MLOs has various proposed functions [46, 47]. In MLOs, the local concentration of specific proteins and nucleic acids is significantly increased, potentially accelerating biochemical reactions. This behavior was evident in histone mRNA processing which takes place in histone locus bodies and was significantly slowed down when critical components of the reactions failed to condense [76].

MLOs may also regulate the specificity of biochemical reactions by compositional control of the biomolecular condensate. In other words, concentrating a protein with only a subset of its interaction partners may favor certain biological processes to occur. An enzyme may be concentrated with only a few of its possible substrates or, similarly, components acting on a specific pathway may be concentrated while those of alternative pathways may be excluded [77].

In contrast, biomolecular condensates may have the capacity to sequester proteins [78] or act as buffers [79] to maintain certain cellular concentrations in the surrounding medium and thereby controlling their activity.

Importantly, the process of condensate formation in cells can be tightly controlled by alterations in external factors such as temperature [80] and salt conditions [81]. Moreover, post-translational modifications (PTMs) provide a strong means to dynamically regulate the phase behavior of certain proteins. Arg methylation in the P granule protein DDX4, for example, diminishes its propensity to undergo LLPS, probably in part due to the decrease in possible cation-π interactions [43]. Tuning the phase behavior of proteins via PTMs is a fast process (occurs in a matter

of minutes). Mechanisms to regulate the phase behavior that occur on slower time-scales exist as well, including control via alternative splicing and evolutionary processes [46].

As discussed above, biomolecular phase transitions do not necessarily yield liquid condensates. In fact, in biological systems, numerous protein condensates have solid properties. Examples for LSPS include the formation of actin filaments and microtubules. Amyloids represent another interesting form of solid phases as they are linked to both, severe diseases and physiological roles. In the following, a brief overview of amyloids will be given.

1.4 Amyloid fibrils

In the mid-19th century, Schleiden and later Virchow coined the term 'amyloid' (from ancient greek: amylon = starch) to describe the iodine-stained deposits detected in liver during autopsies [82]. As iodine-staining is commonly used to detect starch, it was believed that these deposits primarily contained carbohydrates. The proposed chemical nature was challenged when a high nitrogen content had been discovered in these deposits more aligning with a high protein content [83]. Later, indeed, proteins could be verified to be the primary constituent of these deposits. Nevertheless, the inaccurate term 'amyloid' continued to be used to refer to these proteinaceous deposits.

Essentially, amyloids are a form of LSPS arising through nucleation and growth [84]. In the following, the structural features of amyloids, their physiological and pathophysiological roles and their formation will be described.

Structure of amyloid fibrils

Amyloids are highly insoluble protein assemblies that are composed of ordered linear protein aggregates termed fibrils [85]. Amyloid fibrils are often vizualized by electron microscopy and usually appear as unbranched structures with a length up to several μm. Fiber diffraction experiments confirmed the earlier proposed

structural similarity between different amyloids. In the fibrillar core, amyloids share a cross-β structure, resulting in X-ray diffraction patterns with major reflections at 4.7 Å and 10 Å. These reflections arise on orthogonal axes and result from (i) the stacking of β-strands with an intra-strand distance of 4.7 Å creating an elongated β-sheet along the fiber axis and (ii) the side-chain packing between the sheets perpendicular to the fiber axis with an inter-sheet distance of 10 Å. More detailed structural studies revealed that the β-sheets within protofilaments can either be oriented in a parallel or anti-parallel manner depending on the precursor polypeptide. Yet, they generally exhibit an in-register setup, meaning that each β-strand lies exactly on top of each other in a parallel configuration, while in an anti-parallel configuration one β-strand lies exactly in line with the β-strand two rows below [82, 85].

(Patho)physiology of amyloids

In humans, amyloids are associated with more than 20 diseases, ranging from neurodegenerative disorders such as Alzheimer's and Parkinson's disease to protein aggregate myopathies, type II diabetes and prion disease [86–88]. Although it is established that amyloid deposition is a consequence of these disorders, it remains uncertain whether they are the causative agent for disease progression or rather a byproduct. In case of the well-studied Alzheimer's disease (AD), not the density of amyloid plaques but the concentration of soluble Aβ peptide was found to correlate with the severity of AD, suggesting that amyloids do not progress the disease state and rather represent the end point of the disorder [82, 89]. However, it is important to note that the current data situation is still not conclusive and the toxic species need to be identified for every amyloidosis separately.

Having thus far only been associated with disease, it was surprising when functional amyloids were first identified in the 2000s [90–94]. Functional amyloids fulfill various physiological roles and can be divided into groups according to their functionality, some of which include (i) *structure*; the structural group comprises amyloids with structural functions. A prominent example are curli as the first ever

identified functional amyloid. Curli are extracellular amyloids produced in some enteric bacteria such as *E. coli*. They are a component of the extracellular matrix and involved in biofilm formation [95]. (ii) *Information*; this group contains amyloids involved in protein-based epigenetic memory and are composed of functional prions. A well-documented case in flies is the prion Orb2 which can switch from being a translational repressor to a translational activator by conformational conversion aiding long-term memory [96]. (iii) *Storage*; in this group, amyloid deposits offer an efficient and inert way of protein/peptide storage. In their seminal paper, Riek and coworkers demonstrated that many peptide hormones from the endocrine system are stored in form of amyloids in secretory granules prior to secretion [94].
In light of the existence of both, functional as well as disease-related amyloids, the question arises what the determining features for their (patho)physiological roles are. Three features are being discussed including the hypothesis that (i) disease-related amyloids are toxic and differ structurally from functional amyloids, (ii) the formation of functional amyloids is tightly controlled reducing potential toxic effects and (iii) fibril formation of functional amyloids is fast compared to disease-related amyloids diminishing potential toxicity associated with intermediate species such as small oligomers [85]. Importantly, evidence for each feature can be found and therefore, each system requires individual assessment.

Understanding the kinetics of fibril formation

To understand the microscopic processes that underlie fibril formation, it is essential to disentangle the complex mechanisms and quantify the isolated molecular events.

Assuming the classic nucleation theory [84], fibril formation can be divided into two fundamental steps: nucleation and growth (see Figure 7). While nucleation describes processes that lead to the formation of new amyloid-compatible structures, growth processes comprise fibril elongation and monomer dissociation [97]. Fibrillar growth (including monomer association and dissociation) is typically rapid and therefore the main process contributing to the increase in total fibrillar mass

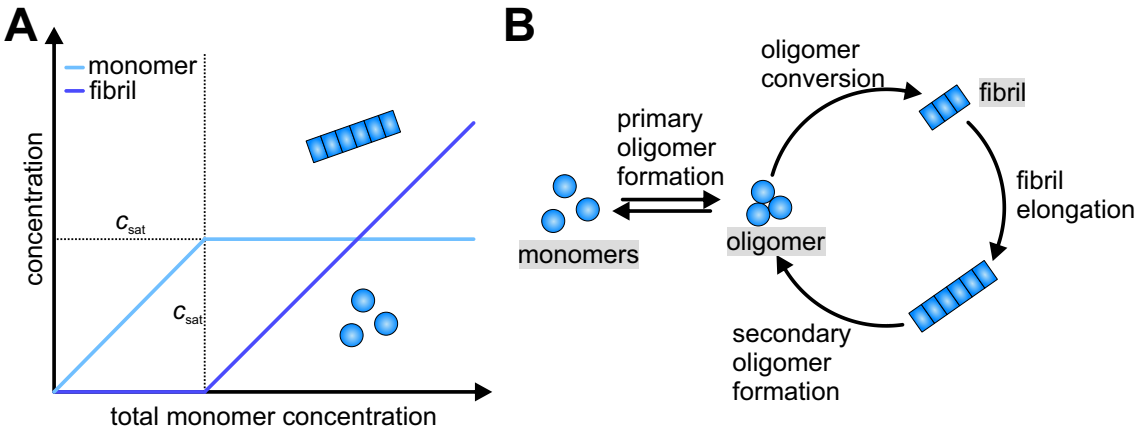


Figure 7: **Processes involved in amyloid formation.** A: At total monomer concentrations c_{tot} below the saturation concentration c_{sat}, protein solutions are stable. When c_{tot} exceeds c_{sat}, protein clusters including fibrils form so that the monomer concentration remains at c_{sat}. B: For amyloid formation from monomeric peptides, various steps including primary oligomer formation, oligomer conversion and fibril elongation need to occur. Secondary oligomer formation through existing fibrils results in an auto-catalytic cycle.

and the concomitant decrease of the free monomer concentration. Nucleation processes are responsible for the increase in the total concentration of aggregates and can be divided into primary and secondary nucleation processes. The spontaneous formation of fibril-compatible oligomers from monomers is described by the primary nucleation. Secondary nucleation, instead, requires existing fibrils and can itself be divided into monomer-dependent and monomer-independent processes. Monomer-dependent processes include surface-induced nucleation events and lateral branching. On the other hand, fragmentation of filaments is a monomer-independent process that leads to an increase in the total number concentration of aggregates. It is important to distinguish between monomer-dependent secondary nucleation and heterogeneous primary nucleation. Secondary nucleation takes place on the surface of existing fibrils while heterogeneous primary nucleation occurs on foreign surfaces. Consequently, these two processes behave very differently. Monomer-dependent secondary nucleation leads to an auto-catalytic amplification of aggregates with the potential for exponential growth, while in heterogeneous primary nucleation, the available catalytic surface does not experience growth over time [84, 97].

1.5 IDPs in this work

In this work, the behavior of the parathyroid hormone (PTH) and the serine/arginine rich splicing factor 1 (SRSF1) was investigated with respect to their self-assembly into condensates. Both are biologically unrelated proteins with extended IDRs and introduced in the following.

Parathyroid hormone

The discovery of the parathyroid glands in the late 19th century marked the beginning of research on a yet unknown and, from the thyroid system, separated entity. Soon after, the parathyroid was established as an endocrine system with the function to secrete the parathyroid hormone (PTH). Even without prior knowledge about the molecular properties of the hormone and its target structures, PTH was identified to regulate serum calcium ion and phosphate levels [98]. Additionally, the pathophysiologic effect of both, PTH excess and deficiency, was investigated. In case of excess parathyroid action (= hyper-parathyroidism), osteoporosis was observed, while insufficient PTH (= hypo-parathyroidism) was found to induce hypocalcemia [99].

More recent advances in structural biology have led to a better understanding of the molecular properties of PTH. PTH is a peptide hormone that is 84 amino acids in length (see Figure 8A). High resolution techniques such as solution nuclear magnetic resonance (NMR) spectroscopy have provided structures of the N-terminal region, predominantly showing two α-helices (amino acids 3-11 & 21-30) joined by a short flexible 'hinge' region [100, 101]. Advances in understanding the cellular and molecular actions of PTH have been made after the identification of its principal receptor, PTH1R (see Figure 8C) [102]. PTH1R is highly expressed in bone and kidney [103] and is equally activated by both, full length and N-terminal PTH peptides [104]. Hence, the N-terminal region of PTH is necessary and sufficient for full activity at PTH1R. Indeed, osteoporosis is effectively treated with an N-terminal fragment of PTH comprising the amino acids 1-34 (Teriparatide is

an approved drug in the US and the EU) [105].

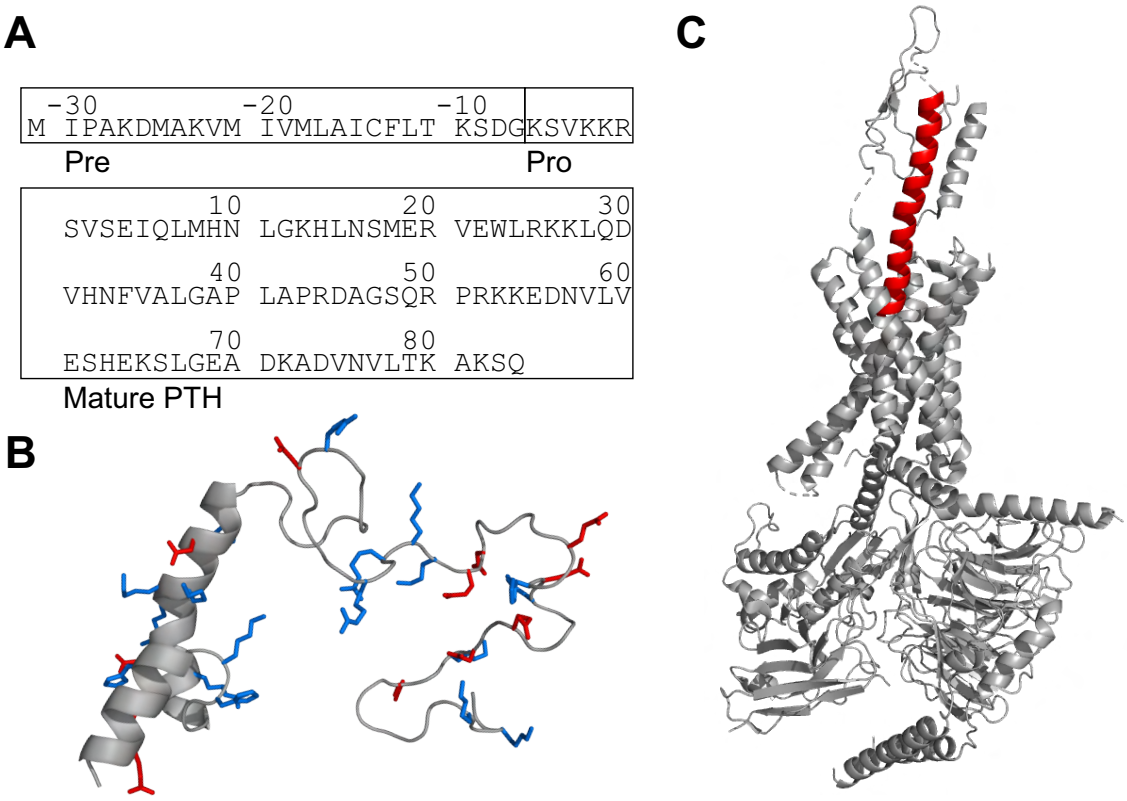


Figure 8: **PTH and its receptor PTH1R.** A: Polypeptide sequence of PTH including the pre- and pro-sequence. The Pre- and Pro-sequence are subsequently cleaved off the peptide during intracellular transport. B: Cartoon representation of PTH_{1-84} based on the PDB entry 1ZWA with a modelled C-terminal disordered region. C: Cartoon representation of PTH1R (gray) in complex with PTH_{1-34} (red) based on PDB entry 7VVK.

The C-terminal region of PTH was shown to have a limited propensity for α-helicity and appears to be largely disordered (see Figure 8B) [106]. Although the greater part of PTH is supposedly disordered, the best understood function lies in the α-helical N-terminus with its receptor-stimulating activity. Little is known about the remaining C-terminal region and the functions associated to it. Hence, for a holistic description of the peptide hormone, a better characterization of the C-terminal part of the protein is required.

The intracellular lifetime of PTH is complex. PTH is synthesized in specialized cells, the parathyroid chief cells [103]. As a secretory protein, PTH passes through the secretory pathway of the producing cell and eventually is secreted into the blood

stream. The intracellular transport of secretory proteins is generally accompanied by a maturation of the proteins. In case of PTH, an N-terminally extended peptide, the so-called prepro-PTH, is initially translated (see Figure 8A). The 'pre'-sequence corresponds to the signal peptide that ensures translocation to the endoplasmic reticulum (ER). After a signal peptidase cleaves off the signal peptide, the generated pro-PTH translocates to the Golgi apparatus. The six amino acids long 'pro'-sequence is cleaved off by the protease furin to yield mature PTH. Mature PTH (hereafter referred to as PTH_{1-84}) is stored in secretory granules until PTH is secreted into the blood stream [103].

In 2009, Riek and coworkers showed that functional amyloids are the natural storage form of many peptide hormones from the endocrine system [94]. They further identified secretory granules to be the sites of amyloid storage. Importantly, the organelles that constitute the secretory pathway experience a gradual acidification as they progress towards the state of secretion [107]. This causes a pH gradient across the pathway which has implications on the sorting of cargo proteins and the regulation of post-translational modifications (PTMs). Furthermore, this gradient suggests that the formation of functional amyloids is adapted to lower pH values [94, 108, 109]. Accordingly, many peptide hormones exhibit an isoelectric point (pI) approaching the granule-relevant pH of ~ 5.5. This renders proteins less soluble at lower pH and simultaneously enhances their propensity to aggregate, for example, into amyloid fibrils [108, 110–112]. Indeed, numerous peptide hormones spontaneously form amyloids under granule-relevant pH conditions *in vitro*, as well [94].

The calculated pI value of PTH_{1-84}, however, is ~ 8.9. Moreover, PTH_{1-84} contains four histidine residues with a pK_A value of ~ 6.2 causing its net charge to significantly increase with decreasing pH: at pH 7.4, the net charge of PTH_{1-84} is ~ +1.4 while it increases to ~ +5.9 at pH 5.5 (calculated using Prot pi Protein Tool https://www.protpi.ch/Calculator/ProteinTool). This increased net charge should result in more extended chain conformations of the largely disordered PTH_{1-84}

and to enhanced inter-chain repulsion [113]. As a consequence, the inherent predisposition for the formation of amyloids should be affected [114]. Indeed, a previously conducted *in vitro* experiment in which different buffer conditions were screened revealed that amyloids form preferably at basic pH, approaching the pI of PTH_{1-84} [113]. Yet, *in vivo*, amyloids of PTH_{1-84} have been observed in acidic secretory granules of parathyroid chief cells [109]. Hence, it is conceivable that under physiological conditions, PTH_{1-84} only efficiently forms amyloids in presence of additional factors. A group of factors potentially facilitating amyloid formation of PTH_{1-84} include polyanionic glycosaminoglycans (GAGs). Being negatively charged, they may favorably interact with the positively charged PTH_{1-84} and induce amyloid formation, for example, via effective heterogeneous primary nucleation. Additionally, GAGs are well known constituents of the secretory pathway and therefore plausible interaction partners of PTH_{1-84}.

Serine/arginine rich splicing factor 1

In eukaryotic systems, pre-mRNA splicing is an essential co-transcriptional process to regulate gene expression and to significantly extend the diversity of gene products [115]. Splicing is a maturation process during which introns are removed and the remaining exons are ligated. Especially in higher eukaryotes, the so-called alternative splicing, i.e. the differential inclusion of introns or exonic regions, can give rise to several transcripts from a single gene [116].

An important group of regulators involved in splicing is the phylogenetically conserved serine/arginine rich (SR) protein family which is present in all plants and metazoans [117]. The SR protein family comprises 12 members in humans (SRSF1 - SRSF12) and while all of them are predominantly nuclear, SRSF1, SRSF3, SRSF7, and SRSF10 can shuttle between the nucleus and the cytosol [118]. SR proteins have been found to be involved in all steps of mRNA metabolism from transcription and mRNP packing in the nucleus to translation and mRNA decay in the cytosol [117–119]. However, SR proteins are best known to be essential components of constitutive splicing and key regulators of alternative splicing [117,

120]. Regulation of splicing is performed by SR proteins in a partially redundant manner. Yet, most SR proteins are essential for normal development as studies on knockout (KO) animal models of mice and *Drosophila* showed embryonic lethality [121]. This emphasizes the important and distinct roles of the different members of this protein family.

SR proteins are RNA-binding proteins (RBPs) exhibiting a modular domain structure that features one or two RNA recognition motifs (RRM) and the eponymous disordered C-terminal RS domain containing numerous Arg/Ser dipeptide repeats [117, 118, 121]. The best understood SR protein is the prototypical serine/arginine rich splicing factor 1 (SRSF1) which will be the focus of the present work. SRSF1 comprises a canonical RRM (RRM1) and a non-canonical pseudo-RRM (RRM2) that is followed by a relatively short RS domain (51 residues) at the C-terminus (see Figure 9). In SRSF1, both RRMs are necessary for efficient RNA binding and pre-mRNA splicing [122], yet, substrate specificity is dominated by the RRM2 *in vivo* [123]. Furthermore, it has been proposed that the RRMs mediate both, interactions with the U1-70K protein component of the U1 small nuclear ribonucleoprotein (snRNP) and the formation of the early spliceosomal complex [124]. Here, the RS domain has been proposed to play a regulatory role.

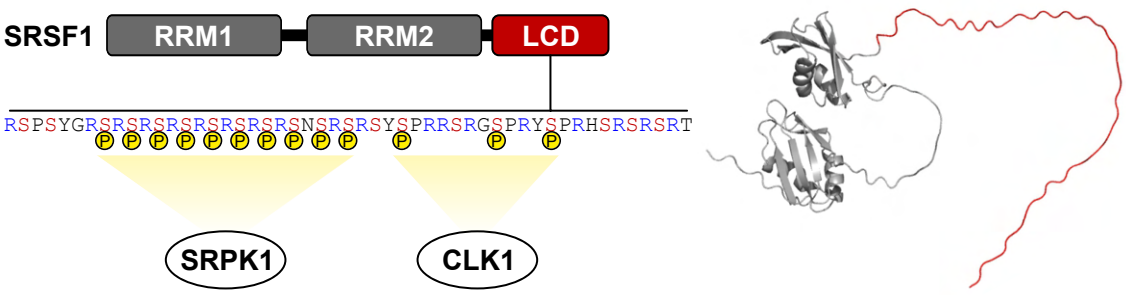


Figure 9: **The LCD of SRSF1 gets extensively phosphorylated under physiological conditions.** The general protein architecture of SRSF1 consists of two ordered RRMs and a C-terminal disordered LCD (top left) and is structurally visualized by Alphafold (right). The LCD exhibits numerous Arg/Ser dipeptide repeats and gets highly phosphorylated by its specific kinases, SRPK1 and CLK1, in a regiospecific manner.

While deletion mutants of SRSF1 that lack the entire RS domain retained splicing activity [122, 123, 125], the RS domain has distinct functions. In the RS domain, long stretches of Arg/Ser dipeptide repeats are sporadically interrupted by other residues resulting in a remarkable degree of low-complexity with Arg/Ser making up more than 75% of the total sequence. The RS domain presents a highly tunable platform for protein-protein interactions and mediates the subcellular localization of SRSF1 in nuclear speckles [117].

Nuclear speckles (NS), or interchromating granule clusters, are nuclear MLOs mainly composed of the long non-coding RNA (lncRNA) MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) and proteins with functions in gene expression regulation including splicing factors [47, 126]. The functions of NS are not unambiguously understood. However, two roles have been proposed: (i) NS may function as reservoirs for RNA processing factors. This potential role is based on a proteomic analysis of isolated NS that revealed the presence of hundreds of RNA processing factors [127]. Nonetheless, neither DNA nor active transcription have been detected in NS suggesting their role in buffering nucleoplasmic concentrations of RNA processing factors [128]. (ii) In contrast, NS may also promote gene expression through facilitating transcription and RNA maturation processes. In this suggested role, NS would provide the necessary factors in a condensed form potentially enabling faster kinetics compared to the dilute nucleoplasm [129]. This role has been proposed as transcriptionally active genes tend to localize in close proximity to NS [130]. Active transcription and local chromatin decondensation have been observed to occur close to NS, as well. Moreover, (pre-)mRNAs were found to accumulate in NS when splicing or nuclear export failed [128].

Recently, NS have been proposed to exhibit a layered architecture with a proteinaceous core that is surrounded by a shell enriched in RNAs including MALAT1 [47, 129]. Interestingly, proteins that reside in NS tend to feature Arg/Ser-rich domains raising the possibility that this specific ”flavor” of low-complexity sequence provides

cohesion in these biomolecular condensates [129]. Indeed, the largely disordered proteins SRRM2 and SON have recently been proposed to function as scaffold proteins in NS [131]. They are suggested to nucleate the formation of NS through their extended Arg/Ser-rich domains by recruiting other proteins that are prone to undergo phase separation.

The role of protein phosphorylation in the process of NS formation and RNA processing activity is subject of many recent scientific efforts. The high abundance of Ser residues in the RS domain of SRSF1 allows for extensive phosphorylation by specific kinases such as the SR protein kinase 1 (SRPK1) and CDC like kinase 1 (CLK1) (see Figure 9) [132]. In the cytosol, SRPK1 phosphorylates about 10 Ser in the long stretch of RS-dipeptide repeats generating the so-called hypophosphorylated state [133]. The hypophosphorylated RS domain is recognized as a nuclear localization signal (NLS) by transportin-SR, an SR protein specific importin [134]. After shuttling to the nucleus, the RS domain mediates subcellular localization in NS [119]. While the role of multisite phosphorylation in governing the phase behavior of SRSF1 is not well understood, the general importance of this relationship is widely accepted. Several lines of evidence suggest that phosphorylations and dephosphorylations regulate the splicing activity and subcellular localization of SRSF1 [117, 119, 135–137]. The hypophosphorylated state was found to inhibit the formation of the early spliceosomal complex while further phosphorylation (= hyperphosphorylated state) promoted it. It was further suggested that the hyperphosphorylated state is released from NS [119]. Yet, conclusive data are still missing to unambiguously define the role of differential phosphorylation on the subcellular localization and the phase behavior of SRSF1 and other RS domain containing proteins.

1.6 Scope of this thesis

The self-assembly of proteins into higher order complexes or condensates is an increasingly recognized process with numerous physiological and disease-related roles associated to it. Therefore, it is of great interest to understand the molecular principles underlying the formation of these assemblies. The physicochemical properties of protein condensates can vary significantly depending on the proteins involved in the assembly.

The peptide hormone PTH_{1-84} was found to assemble into amyloids being the presumed natural storage form inside secretory granules [109, 113]. However, the precise mechanism by which PTH_{1-84} forms amyloids under physiological conditions remained unclear. Previous studies *in vitro* suggested that amyloid formation of PTH_{1-84} is favored when the pH conditions approach the basic pI of the peptide [113], a scenario that is found reversed in the continuously acidifying secretory pathway [107]. As a result, PTH_{1-84} becomes considerably positively charged during its intracellular life time indicating that negatively charged counterparts might be required for efficient amyloid formation. Therefore, it was hypothesized that *in vivo* amyloid formation of PTH_{1-84} may rely on polyanionic glycosaminoglycans (GAGs) or similar structures which are abundant in the late stages of the secretory pathway [138]. Indeed, Riek and coworkers proposed GAGs to be important helper molecules for amyloid formation of several other peptide/protein hormones found in secretory granules [94]. The central goal of this work was to characterize the effect of heparin, a widely used GAG, on the amyloid formation of PTH_{1-84} *in vitro* under granule-relevant pH conditions. Several biophysical techniques were used including isothermal titration calorimetry (ITC) to study the PTH:heparin interaction and circular dichroism (CD) spectroscopy to reveal potential structural changes in PTH_{1-84} induced by heparin. The role of heparin in the amyloid formation of PTH_{1-84} was investigated using small & wide angle X-ray scattering (SAXS & WAXS), thioflavin T (ThT) fluorescence assays, transmission electron microscopy (TEM) imaging and nuclear magnetic resonance (NMR) spectroscopy.

An additional aim of this work was to shed light on potential functions of the supposedly largely disordered C-terminus of PTH_{1-84}. So far, most research on the parathyroid system focussed on the N-terminal part of PTH_{1-84} which had been established to be sufficient for activation of the canonical PTH1R [104]. Although the existence of discrete receptors for the C-terminal region of PTH_{1-84} was suggested, functional data regarding this region of the peptide hormone are scarce [101]. In this work, a structural approach was chosen in an attempt to infer a functional role of the C-terminal region. Based on molecular dynamics (MD) simulations of the PTH_{1-84}, it was speculated that long-range intra-chain interactions between the α-helical N-terminal region and the disordered C-terminal region might develop. Cross-linking MS (XL-MS) was used to further investigate the chain properties of monomeric PTH_{1-84} in solution. Finally, the effect of the disordered C-terminal region on amyloid formation of PTH_{1-84} was monitored and discussed as a potentially physiologically relevant function.

With SRSF1, an unrelated protein system was investigated regarding its ability to self-assemble into liquid-like condensates. This ability appears to be an important driving force for the formation of nuclear speckles (NS) which are biomolecular condensates that SRSF1 and other RNA processing factors localize to [49]. Importantly, SRSF1 and many other NS proteins comprise LCDs that exhibit a high enrichment in Arg and Ser residues [117, 126, 131], a sequence feature that was earlier proposed to provide cohesion in NS [49]. Furthermore, protein phosphorylation is a well-documented PTM in the NS proteome with regulatory functions [124, 136]. Although previous results suggest a correlation between phosphorylation and subcellular localization [135, 139], a detailed relationship between the two remains elusive. In this work, a combined *in cellulo* and *in vitro* approach was followed to investigate the condensation behavior of SRSF1 variants with a focus on the isolated SRSF1 LCD. Several phospho-states of SRSF1 LCD were approximated by phospho-mimicking mutants of which the subcellular localization was tested via confocal laser scanning microscopy. Moreover, the phase behavior

in vitro of these SRSF1 LCD variants was investigated using a combination of microscopy and turbidity assays. Observed differences in the phase behavior were discussed based on the physicochemical properties of the peptides.

2 Material & Methods

2.1 Text editing, graph plotting and figure design

This thesis was written with the LaTeX editor Texmaker. Graphs were plotted with OriginPro 2019 (OriginLab). Figures were generated with the CorelDRAW graphics suite (Corel Corporation). Structures of biological macromolecules were visualized and rendered using PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.)

2.2 Laboratory equipment

Laboratory equipment used for standard procedures is listed in Table 2. Further equipment is specified in the respective sections.

2.3 Chemicals

Unless stated otherwise, chemicals were purchased from Sigma-Aldrich (Munich), New England Biolabs (Frankfurt) and Carl Roth (Karlsruhe). Unfractionated porcine intestinal heparin with an average molecular weight of 20 kDa was sourced from Carl Roth (order ID: 7692.1). The heparin sulphate analog fondaparinux (pentasaccharide) was sourced from Sigma-Aldrich (order ID: SML1240). Poly-U RNA (20mer) was purchased from Eurofins Genomics.

2.4 Peptides

PTH_{1-84} (Uniprot ID: P01270, AA sequence: SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNFVALGAPLAPRDAGSQRPRKKEDNVLVESHEKSLGEADKADVNVLTKAKSQ), PTH_{1-34} (AA sequence: SVSEIQLMHNLGKHLNSMERVEWLRKKLQDVHNF), PTH_{35-84} (AA sequence: VALGAPLAPRDAGSQRPRKKEDNVLVESHEKSLGEADKADVNVLTKAKSQ) and SRSF1 LCD (Uniprot ID: Q07955; AA 199-248, sequence: SPSYGRSRSRSRSRSRSRSRSNSRSRSYSPRRSRGSPRYSPRHSRSRSRT)

Table 2: **Standard laboratory equipment used in this thesis.**

Name	Description	Manufacturer
Avanti J-20	Centrifuge for cell harvesting	Beckman Coulter
L8-60M	Ultracentrifuge for inclusion body isolation	Beckman Coulter
Optima TLX	Ultracentrifuge for separation of protein fibrils	Beckman Coulter
JLA 8.1000	Fixed-angle rotor for Avanti J-20 centrifuge fitting 1 L volumes	Beckman Coulter
JA 25.50	Fixed-angle rotor for Avanti J-20 centrifuge fitting 50 mL volumes	Beckman Coulter
45 Ti	Fixed-angle rotor for L8-60M ultracentrifuge fitting 70 mL volumes	Beckman Coulter
miniSpin plus	tabletop centrifuge for 1-2 mL volumes	Eppendorf
Centrifuge 5418 R	tabletop centrifuge for 1-2 mL volumes	Eppendorf
FiveEasy	benchtop pH meter	Mettler Toledo
LE438	pH electrode for FiveEasy pH meter	Mettler Toledo
Mastercycler personal	PCR thermocycler	Eppendorf
PowerPac Basic	Power supply for agarose gelectrophoresis	Bio-Rad
EPS-301	Power supply for SDS-PAGE	Amersham Biosciences
ThermoMixer C	Thermomixer for 1-2 mL tubes	Eppendorf
Genesys 180	Spectrophotometer	Thermo Fisher Scientific

were synthesized by the Core Unit Peptide Technologies of the University Leipzig using microwave-assisted solid-phase peptide synthesis (CEM GmbH, Germany). Phospho-mimicking mutants of SRSF1 LCD (AA 198-248, sequence (Hypo-D): RSPSYGRDRDRDRDRDRDRDRDNDRDRSYSPRRSRGSPRYSPRHSRSRSRT, (Hyper-D): RSPSYGRDRDRDRDRDRDRDRDNDRDRSYDPRRSRGDPRYDPRHSRSRSRT, (Total-D): RDPDYGRDRDRDRDRDRDRDRDNDRDRDYDPRRDRGDPRYDPRHDRDRDRT) were synthesized by Genosphere Biotechnologies. The successful synthesis of the peptides was verified by mass spectrometry.

2.5 Consumables

Consumables are listed in Table 3. Further consumables are specified in the respective sections.

Table 3: **Standard consumables used in this thesis.**

Name	Description	Manufacturer
Amicon Ultra-0.5, -4 and -15	centrifugal filters with 3 and 10 kDa MWCO	Merck
Pur-A-Lyzer Midi/Maxi	Dialysis tubes for small volumes	Sigma Aldrich
Agarose Gel Extraction Kit	Spin-column bades DNA cleanup from agarose gels	Jena Bioscience
Fast-n-Easy Plasmid Mini-Prep Kit	Column based isolation of plasmid DNA	Jena Bioscience
PCR Cloning kit	Cloning of blunt-end DNA	New England Biolabs
Pierce Protease and Phosphatase Inhibitor	EDTA-free tablets to prevent protein degradation	Thermo Fisher Scientific
Lysozyme (from egg white)	Agent to degrade the bacterial cell wall	Merck
Benzonase	Endonuclease for DNA and RNA degradation	Merck
Quick-Load Purple 1 kb plus DNA ladder	Molecular marker for agarose gel electrophoresis	New England Biolabs
SeeBlue Plus2	Prestained protein ladder for SDS-PAGE	Invitrogen

2.6 Cloning, expression and purification of SRSF1

Cloning of SRSF1 constructs

The plasmid containing the coding sequence of SRSF1 (murine origin, amino acid sequence is 100% identical to human SRSF1) cloned into a pcDNA3.1(+)-derived vector was obtained from Peter Zorn (AG Köhn, MLU Halle). pcDNA3.1(+) is a mammalian expression vector of which the backbone has been modified in-house to place a streptavidin-binding-peptide(SBP)-FLAG-RFP-tag (SFR-tag) N-terminal to the desired protein. This multipurpose tag features high affinity to biotin through SBP, a FLAG octapeptide for protein purification as well as charac-

terization and a red fluorescent protein (RFP) for *in cellulo* fluorescence imaging. This plasmid was used for transfections of human cell lines and subsequent experiments.

In pcDNA3.1(+), the coding sequence of SRSF1 is flanked by the restriction sites EcoRI (5') and XhoI (3') for convenient subcloning into vectors serving different purposes. For subcloning into a pET41a(+)-derived vector, both the SRSF1 (insert) and the vector backbone were cut with EcoRI and XhoI. Ligation was performed using the T4 ligase (New England Biolabs, Frankfurt). pET41a(+) is a bacterial expression vector which was modified in-house (AG Köhn, MLU Halle) to place a hexahistidine(His_6)-tag flanked by a H3C-protease recognition site N-terminal to the desired protein. This tag allows SRSF1 to be purified via Ni^{2+}-NTA immobilized metal affinity chromatography (IMAC), while enabling the proteolytic removal of the tag afterwards.

In addition, SRSF1 mutants mimicking multisite phosphorylation states of the RS domain were generated. Mutants with 10 S/D mutations (termed SRSF1 Hypo-D) and 13 S/D mutations (termed SRSF1 Hyper-D) were obtained from Eurofins Genomics, Ebersberg. More specifically, the purchased DNAs encoded the amino acids 183-248 of SRSF1 and contained the native BssSI (5') and a XhoI (3') restriction site for cloning purposes. The coding sequences were amplified via PCR using Q5 DNA polymerase (New England Biolabs, Frankfurt) and the follwing primers: forward - CACGAGGGAGAAACTGCCTACATCCGG; reverse: GGCTCGAGTTATGTACGAGAGCGAGATCTGCTATGACG. The amplified DNA was cloned into the pMiniT 2.0-vector (New England Biolabs, Frankfurt) according to the manufacturers' protocol. Cutting with BssSI and XhoI generated an insert of the mutated part of SRSF1 (183-248). Additionally, SRSF1 (1-183) was cut out of pcDNA3.1(+) with EcoRI and BssSI. In a DNA ligation approach with two inserts, the mutated DNA (183-248) was ligated with the wildtype SRSF1 fragment (1-183) and inserted into either the pET41a(+)-derived or the pcDNA3.1(+)-derived vectors previously cut with EcoRI and XhoI.

Constructs of the isolated C-terminal RS domain (amino acids 184-248) of SRSF1 (hereafter termed SRSF1 LCD) were generated in the aforementioned mammalian and bacterial expression vectors using the following strategy: The SRSF1 LCD encoding DNA was amplified via PCR and the following primers: forward - GGGAATTCGAGGGAGAAACTGCCTACATCCGGGTTAAA; reverse - GGCTCGAGTTATGTACGAGAGCGAGATCTGCTATGACG. The amplified DNA was cloned into the pCR-Blunt II-TOPO-vector (Invitrogen, Karlsruhe) according to the manufacturers' protocol. For subcloning into pcDNA3.1(+) and pET41a(+) vectors, the aforementioned procedure consisting of restrictions with EcoRI and XhoI and subsequent ligation was performed. Coding sequences of generated constructs are shown in the appendix.

Expression & purification of SRSF1

Several approaches for expression and subsequent purification of SRSF1 variants were followed.

Single-expression of SRSF1 FL

SRSF1 constructs were expressed in *E. coli* BL21DE3+ in LB medium supplemented with 0.1 mg/mL kanamycin (kan). Bacteria were grown at 37 °C while shaking at 130 rpm. At an OD_{600} ~ 0.7, the temperature was decreased to 20 °C for 30 min before expression was induced with 1 mM IPTG. After over-night expression, cells were harvested by centrifugation and frozen at -20 °C until further use.

SRSF1 wildtype constructs form inclusion bodies in the expression host *E. coli*. Therefore, an inclusion body preparation is required for SRSF1 FL purification. 15 g cells were thawed in 50 mL buffer (0.1 M Tris-HCl, 1 mM EDTA; pH 7) supplemented with lysozyme and protease inhibitor (cOmplete EDTA-free tablet, Roche) for 30 min and homogenized using an ultraturrax (10000 rpm). After cell disruption with a french press, benzonase (Merck, Darmstadt) was used in presence of 2 mM $MgCl_2$ to digest nucleic acids for 30 min at 25 °C. 0.5 volumes of 60 mM EDTA, 6% Triton X-100, 1.5 M NaCl, pH 7, were added to the suspension and incubated

for 30 min at 4 °C under stirring. Inclusion bodies were spun down at 96000 g for 60 min at 10 °C. The pellet was resuspended in 100 mL 0.1 M Tris-HCl, 20 mM EDTA, pH 7 buffer, using an ultraturrax. Inclusion bodies were again spun down at 96000 g for 60 min at 10 °C and subsequently resuspended in 100 mL 8 M urea, 50 mM NaP_i, 100 mM NaCl, 10 mM DTT, pH 8, using an ultraturrax and incubated for 2 h at 25 °C under stirring. The pH was decreased to ~ pH 3 by dropwise addition of HCl and the suspension was incubated over night at 4 °C under stirring. Insoluble cell debris including nucleic acids was spun down at 96000 g at 10 °C for 90 min. For renaturation, the collected supernatant fraction was subjected to dialysis over night at 6 °C against refolding buffer (50 mM HEPES, 50 mM L-Arg, 50 mM L-Glu, 0.5 mM TCEP, 1 mM EDTA, pH 7.4). Aggregated protein was removed by centrifugation and the soluble fraction was collected.

Co-expression of SRSF1 FL and SRPK1

Based on a previously published report, a protocol was established for co-expression of SRSF1 FL and its specific kinase SRPK1 [140]. T7 Express lysY/Iq (New England Biolabs, Frankfurt) *E. coli* cells were transformed with the pET41a(+)-derived vector encoding N-terminally His_6-tagged SRSF1 FL and pGEX-6p-1 encoding SRPK1. Cells were grown until OD_{600} ~ 0.8 was reached before protein expression was induced with 1 mM IPTG. Cells were incubated over-night at 20 °C and then harvested by centrifugation.

Cells were resuspended in lysis buffer (50 mM HEPES, 300 mM NaCl, 0.05% Triton X-100, pH 7.4) supplemented with lysozyme (Merck, Darmstadt) and protease/phosphatase (Thermo Fisher Scientific, Dreieich) inhibitor mix. Cell wall disruption was performed via french press. Cell debris was removed by centrifugation (JA 25.50 rotor, 20000 rpm). The supernatant was loaded onto a 5 mL HisTrap column run by an Äkta Exlorer (Cytiva, Freiburg im Breisgau) at 14 °C. Protein was eluted with lysis buffer containing 300 mM imidazol. To remove imidazol, the eluted protein was subjected to dialysis against lysis buffer while Prescission protease (Cytiva, Freiburg im Breisgau) was used to cleave off the N-terminal His_6-tag of

SRSF1 FL over night at 6 °C. The solution was applied to 1 mL of glutathione agarose (Jena Bioscience, Jena) and incubated for 3 h at 6 °C while shaking to remove the GST-tagged Prescission protease. Afterwards, the solution was again applied to a 5 mL HisTrap column to remove any free His_6-tag, uncleaved SRSF1 FL and other unspecifically binding contaminants.

Single-expression of SRSF1 LCD variants

SRSF1 LCD variants were expressed in *E. coli* BL21DE3+ cells as described in section 2.6. The harvested cells were resuspended in 40 mL of lysis buffer (8 M urea, 20 mM NaP_i, pH 8) and lysed via french press. Cell debris was removed by centrifugation (JA 25.50 rotor, 20000 rpm) and the supernatant was loaded onto a 5 mL HisTrap column run by an Äkta Exlorer (Cytiva, Freiburg im Breisgau) at 14 °C. After extensive washing with lysis buffer (until no changes in the UV profile were noticable), the column was extensively washed with washing buffer (20 mM NaP_i, 1 M NaCl, 20 mM imidazole, pH 8). Subsequently, the bound protein was eluted with elution buffer (20 mM NaP_i, 1 M NaCl, 500 mM imidazole, pH 8). After dialysis against SEC buffer (20 mM NaP_i, 0.5 M NaCl, pH 7.4) over night at 6 °C, size exclusion chromatography (SEC) was performed using a HiLoad Superdex 75 pg column (Cytiva, Freiburg im Breisgau) run by an Äkta Exlorer (Cytiva, Freiburg im Breisgau) at 14 °C. Protein fractions were collected and their identity assessed via SDS-PAGE analysis.

2.7 Cloning of SRPK1

The coding sequence of SRPK1 was obtained from Addgene in form of the plasmid pDONR223-SRPK1 (item ID: 23582). The gene was amplified via PCR with the following primers: forward - GGGAATTCATGGAGCGGAAAGTGCTTGCGCTCCAGGCCCGAAA; reverse - GGTCTAGATTAGGAGTTAAGCCAAGGGTGCCGGAGACACTCGG. The further cloning strategy was similar to what is described in section 2.6. In short, after cloning into the pMiniT 2.0 vector, the gene was cut out with EcoRI and XbaI and subcloned into both,

the aforementioned pET-41a(+)-derived vector and a pGEX6p1-derived vector (bacterial expression vector encoding an N-terminal GST-tag linked via a H3C-protease recognition site) previously cut with EcoRI and XbaI. Coding sequences of generated constructs are shown in the appendix.

2.8 Cloning of CLK1

CLK1 was obtained from cDNA (HEK293T cell line) using the following primers: forward - GAATTCATGAGACACTCAAAGAGAACTTACTGTCC; reverse: CTC-GAGCTATATACTTTTCTTCAGAAGGTCAAAGAAAGGA. A truncated construct of CLK1 (hereafter termed CLK1s) comprising amino acids 148-484 was obtained similarly with the following primers: forward - GAATTCCACCTGATCTGTCA-GAGTGGAGACGTACTAAG; reverse - CTCGAGCTATATACTTTTCTTCAGAAGGT-CAAAGAAAGGA. A cloning strategy similar to the one used for SRSF1 constructs was followed. In short, the PCR products were cloned into pMiniT 2.0 vectors. The sequence of the genes were validated via DNA sequencing at Eurofins. For subcloning into the earlier described pET41a(+)-derived vector, both the genes as well as the vector backbone were cut with EcoRI and XhoI before ligation. Coding sequences of generated constructs are shown in the appendix.

2.9 SDS-polyacrylamide gelelectrophoresis and western blot analysis

SDS-polyacryladmide gelelectrophoresis (SDS-PAGE) separates proteins according to their molecular weight. SDS-PAGE was used to assess protein purification, test expression of protein constructs in mammalian cell lines and during sample preparation for cross-linking mass spectrometry.

In case of test expressions, HEK293T cells were transfected with 2 μg plasmid DNA encoding protein constructs. After incubation for 48 h, cells were harvested and lysed in lysis buffer (50 mM Tris-HCl pH 7.4, 50 mM NaCl, 1% SDS, 2 mM $MgCl_2$) containing TurboNuclease (Jena Bioscience, Jena). The BCA Protein

Assay Kit (Bio-Rad Laboratories GmbH, Feldkirchen) was used to determine the protein concentration. Samples were prepared using 4x NuPAGE LDS sample buffer (Invitrogen GmbH, Karlsruhe) and SDS-PAGE was performed with either pre-cast 4-12% Bis-Tris SDS gels (Invitrogen GmbH, Karlsruhe) or self-cast 15% gels.

Proteins were then either visualized using coomassie or blotted on a nitrocellulose membrane (according to the manufacturers' protocol) for subsequent immunodetection. After antibody incubation, a Li-COR Odyssey scanner (Li-COR Biosciences GmbH, Bad Homburg) was used to detect infrared signals from IRDye680RD/800CW-linked secondary antibodies. A list of antibodies used in this study is given in Table 4.

Table 4: **List of antibodies used in this work.**

Name	**Method**	**Company**	**order ID**
GST Tag	Western blot	Sigma	SAB4301139
Phosphoepitope SR proteins	Western blot	Millipore	MABE50
Flagtag	Western blot	Sigma	F3165
SC35	Immunofluorescence	Sigma	S4045
ASF (SRSF1, N-terminus)	Western blot	Bethyl	a302-052a
ASF (SRSF1, C-terminus)	Western blot	Bethyl	a302-053a
His-probe (H-3)	Western blot	Santa Cruz	sc-8036

2.10 Analytic ultracentrifugation

Analytical ultracentrifugation (AUC) is a versatile method to study macromolecular characteristics such as molecular weight and hydrodynamic properties under biologically relevant conditions.

Centrifugation of particulate samples results in sedimentation towards the opposite of the center of rotation. Therefore, in sedimentation velocity experiments, high speed centrifugation depletes the solutes away from the rotational center and a pellet is created at the bottom of the cell. In AUC, the resulting gradient in the distribution of solutes along the cell is continuously monitored through an optical detection system. This gradient over time is used to calculate e.g. the sedimenta-

tion coefficient and the molecular weight of the solutes.
In sedimentation equilibrium experiments, a lower rotational speed is used to allow an equilibrium in the rates of sedimentation and back-diffusion in the cell. In the equilibrium state, the absorbance of the solutes is scanned along the cell. The radial absorption profile provides information about the homogeneity of the sample and potential aggregation states [141].

AUC experiments

AUC was performed in collaboration with PD Dr. Hauke Lilie, MLU Halle, using a ProteomeLab XL-1 analytical ultracentrifuge (Beckman Coulter, Krefeld). 150 μM PTH_{1-84} was solubilized in citrate buffer (20 mM, pH 5.5) and potential aggregates were spun down by centrifugation (17000 g, 30 min). Sedimentation velocity experiments were performed at 40000 rpm at 20 °C for 6 h. Sedimentation equilibrium experiments were conducted at 20000 rpm and 20 °C. The data were recorded at a wavelength of 280 nm and analyzed using the software Sedfit [142].

2.11 Circular dichroism spectroscopy

Circular dichroism (CD) spectroscopy is a non-destructive method to study protein secondary structure. CD spectroscopy provides low resolution information on protein structure compared to NMR spectroscopy and X-ray crystallography but is less demanding in terms of sample preparation and time requirements.
Plane polarized light consists of clockwise (L) and counterclockwise (R) rotating polarized components of equal magnitude. The differential absorption of these components gives rise to the CD effect. Chiral chromophores and chromophores covalently linked to chiral centres of a molecule tend to absorp the two components of plane polarized light differentially and thus generate a CD signal. In proteins, peptide bonds are optically active and absorb in the far UV spectrum (below 240 nm). Various secondary structures of proteins give rise to characteristic CD spectra (see Figure 10) [143, 144].

The CD signal is reported in terms of ellipticity (θ) in degrees (deg) and is com-

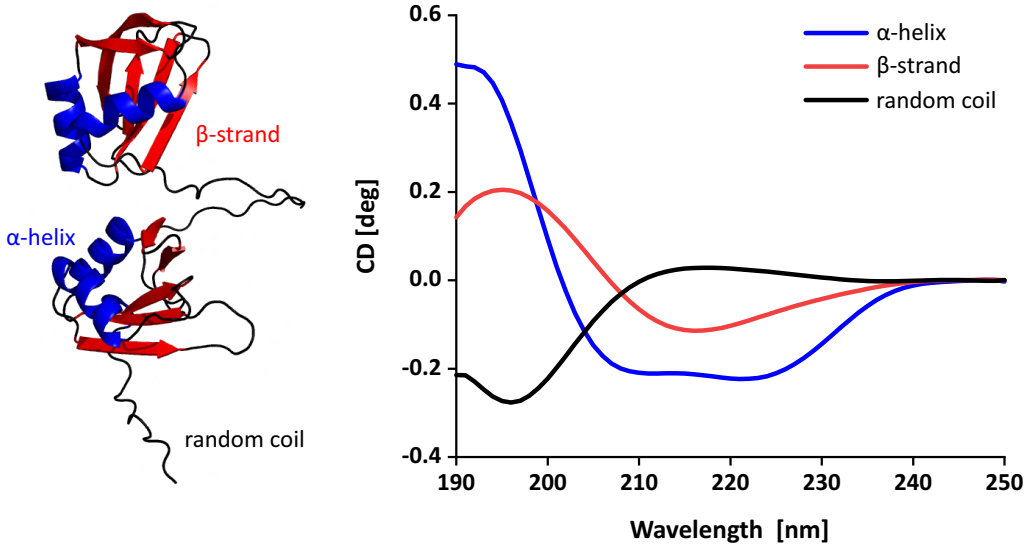


Figure 10: **Basic CD spectra of protein secondary structures.** Protein secondary structures such as α-helices (blue), β-strands (red) and random coil (black) give rise to characteristic CD spectra [144].

monly normalized to obtain the mean residue ellipticity ($\theta_{MRW,\lambda}$) using Equation 3.

$$\theta_{MRW,\lambda} = \frac{MRW \cdot \theta}{10 \cdot c \cdot d} \tag{3}$$

Here, c is the protein concentration in g/mL and d is the cuvette's path length in cm. MRW refers to the mean residue weight, MRW = M/(N-1), where M is the protein's molecular mass and N is the number of residues.

CD measurements

CD measurements were carried out on a Jasco J-810 spectrophotometer (Jasco Deutschland GmbH, Pfungstadt) using a 0.01 cm pathlength quarz cuvette (Hellma GmbH & Co. KG, Müllheim, Germany). PTH_{1-84}, PTH_{1-34} or PTH_{35-84} were concentrated to 100 μM in citrate buffer (20 mM, pH 5.5) or sodium phosphate buffer (50 mM, pH 7.4). Heparin containing samples were set to a final molar ratio of 1:10 (heparin:PTH_{1-84}) and incubated for 1 h at 20 °C. As the amplitude of the CD signal of heparin was much smaller compared to PTH_{1-84}, the CD spectra of the solution containing PTH_{1-84} and heparin were corrected for potential heparin contributions by adding equal heparin concentrations in the reference buffer. The exposure time was set to 50 nm/min and spectra were obtained from

averaging 64 subsequent measurements at 20 °C after subtracting the reference buffer spectrum.

2.12 Infrared spectroscopy

Infrared (IR) spectroscopy is a valuable technique to investigate structural features of proteins [145]. IR spectroscopic measurements were performed by Dr. Maria Ott using an ALPHA 2 Platinum-ATR FT-IR spectrometer (Bruker, Rheinstetten). 2 μL of a 500 μM protein solution were applied on the diamond crystal sample stage and dried for up to 30 min at 37 °C for decreased background contributions arising from water. The increased protein concentration due to drying did not affect intermolecular interactions as verified by comparative measurements pre- and post-drying. Spectra were recorded at 37 °C with 32 scans in the mid-IR region (400-4000 cm^{-1}) and a resolution of 2 cm^{-1}. Background was subtracted.

2.13 Small & wide angle X-ray scattering

Small & wide angle X-ray scattering (SAXS & WAXS) are widely used techniques to gain structural insights into a wide range of particulate matter such as biological macromolecules including proteins. Unlike NMR spectroscopy and cryo electron microscopy, SAXS & WAXS are not limited in the sizes of the proteins under investigation. Furthermore, protein crystallization is not required as measurements typically take place in solution. However, crystalline samples such as protein fibrils are amenable to this method, as well. Labeling of the molecules under investigation is not required [22].

Structures of different length scales can be accessed by analyzing the scattering behavior under different angles. According to Bragg's law, an increase in the length scale of the scattering structure is associated with a decrease in the scattering angle. SAXS & WAXS are therefore complementary techniques accessing different length scales spanning small interatomic distances in the low Å length scale up to large ~ 100 nm structures.

WAXS

WAXS experiments were performed in the lab of Prof. Dr. Thomas Thurn-Albrecht, physics department, MLU Halle. Data analysis was performed by Dr. Maria Ott. For WAXS experiments, PTH_{1-84} fibrils were formed in citrate buffer (20 mM, pH 5.5) in the presence of heparin (1:10 molar ratio). The fibril containing suspension was spun at 200,000 g for 10 min. A fibril pellet was obtained and transferred into a ring-shaped aluminum holder (2 mm thick and with a central hole of 1.5 mm diameter) and left to dry overnight. The scattering measurements were performed at room temperature under vacuum.

SAXS

SAXS experiments were performed in the lab of Prof. Dr. Thomas Thurn-Albrecht, physics department, MLU Halle. Data analysis was performed by Dr. Maria Ott. All X-ray scatterung experiments were performed in transmission mode using a SAXSLAB laboratory setup (Retro-F) equipped with an AXO microfocus X-ray source. The AXO multilayer X-ray optic (AXO Dresden GmbH, Dresden, Gemany) was used as a monochromator for Cu-K_α radiation ($\lambda = 0.154$ nm). A two-dimensional detector (PILATUS3 R 300K; DECTRIS, Baden, Switzerland) was used to record the 2D scattering patterns.

SAXS experiments were conducted using refillable capillaries with an outer diameter of 1 mm (BioSAS JSP stage, SAXSLAB/Xenocs, Grenoble, France). The intensities were angular-averaged and plotted versus the scattering angle q. The measurements were performed in citrate buffer (20 mM, pH 5.5), at room temperature and corrected for background, transmission and sample geometry. Solutions containing heparin and PTH_{1-84}, respectively, were measured for 10 h. A solution of mixed heparin and PTH_{1-84} (1/10 molar ratio) was used for a time dependent series of 12 consecutive measurements for 2 h each. Due to the lower scattering contrast of heparin and, consequently, the need to measure at higher concentra-

tions, we performed experiments at two concentrations (100 μM and 1 mM) with subsequent extrapolation to zero concentration [146].

The data analysis for all measurements included the determination of the radius of gyration, R_g, using the Guinier approximation with $\ln(I(q)) \approx R_g^2/3 \cdot q^2$ for the low q-range, $qR_g < 1.3$. The intermediate q-range was fitted to access the shapes of the scatterers using suitable form factor descriptions, $P(q)$. For the scattering of PTH_{1-84}, a linear polymer model (Debye function) was used[147]:

$$P(q) = \phi(\Delta\rho)^2 2V_c \left(e^{(qR_g)^2} + (qR_g)^2 - 1\right)/(qR_g)^4 \tag{4}$$

In Equation 4, V_c is the chain's volume of occupation, $\Delta\rho$ the scattering contrast and ϕ the volume fraction.

For solutions of heparin and heparin with PTH_{1-84}, a description of a semi-flexible chain, as introduced by Pedersen and Schurtenberger [148, 149], gave the best fitting result. The main fitting parameters are the so called Kuhn length, b, reflecting the stiffness of the chain, the chain's cross-sectional radius, R_c, and the contour length, L, describing the total length of the chain. To strengthen the uniqueness of the description, the parameters were matched with results from other analyses techniques: R_c was additionally derived from the Guinier relation $\ln(I(q)) \cdot q \approx R_{c,g}^2/2 \cdot q^2$ for the q-range $1.5 < qR_{c,g} < 2.1$ with $R_c = \sqrt{2}R_{c,g}$ and found to be concentration independent. The parameter b was determined by the minimum of the plot $I(q)\cdot q^2$ vs q via the relation $b \approx 4.6/q^*$ and subsequent extrapolation to zero concentration [150]. The contour length, L, was also determined by extrapolation, using the concentration dependent scattering intensities $I(q)$ and $I(q)q$ with $q \to 0$ [146].

2.14 Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a versatile technique to study protein structure and dynamics in a residue-specific resolution. Nuclear magnetic resonance (NMR) spectroscopic measurements were carried out by

Bruno Voigt (Prof. Dr. Jochen Balbach, MLU Halle). 2D ^{1}H-^{15}N HSQC spectra of 100 μM ^{15}N labeled PTH_{1-84} were recorded on a Bruker Ascend 500 MHz spectrometer at pH 5.5 and 25 °C in the absence and presence of 100 μM fondaparinux. Chemical shift assignments were performed by standard triple resonance experiments based on the published values [151].

2.15 Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) is a label-free method to study the thermodynamics of biomolecular interactions. ITC allows to determine the dissociation constant, K_D, stoichiometry, n, enthalpy change, ΔH and entropy change, ΔS, of binding in a single experiment. This thermodynamic profile is obtained through direct measurements of the heat that is either released or absorbed during a binding event.

ITC measurements of the PTH:heparin interaction

ITC was performed in collaboration with Lisa Machner (Prof. Dr. Stephan Feller, MLU Halle).

ITC measurements were carried out using a MicroCal iTC200 (Malvern Panalytical GmbH, Kassel) calorimeter at 25 °C and 750 rpm stirring. Titrations of a 125 μM heparin solution into a 150 μM PTH_{1-84} solution (both citrate buffer (20 mM, pH 5.5)) were analyzed. The titration regime included an initial 0.5 μL injection and subsequent 39 $\times$ 1 μL injections with a time interval of 300 s. The reference cell contained citrate buffer (20 mM, pH 5.5). Thermograms were integrated using NITPIC software (version 1.3.0) [152]. Data fitting was performed using CHASM software with the built-in one site model and optional use of multi-sites models [153].

2.16 Transmission electron microscopic imaging

Transmission electron microscopic (TEM) imaging provides a significantly higher resolution than traditional light microscopic imaging, allowing for the detec-

tion of protein aggregates such as amyloid fibrils. TEM imaging was performed by Bruno Voigt (Prof. Dr. Jochen Balbach, MLU Halle).
5 μL of a 20 times dilution of the respective fibrillated sample was transferred to a 200 mesh Formvar/Carbon coated Cu grid (Plano GmbH, Wetzlar, Germany). After 3 min of waiting time, the grids were first cleaned in water for 3 $\times$ 10 s and then negatively stained with 1 percent (w/v) uranyl acetate for a further 60 s. TEM images were taken with an electron microscope (EM 900; Zeiss, Jena) at 80 kV acceleration voltage.

2.17 Thioflavin-T assay

Thioflavin-T (ThT) is a benzothiazole dye that specifically stains amyloid fibrils. Upon incorporation into amyloid fibrils, the ThT fluorescence intensity increases by ~ 160 times. Therefore, the dye is a suitable sensor to monitor fibril formation *in vitro*.

ThT assay experiments of PTH fibrillation

ThT assays were carried out on a FLUOstar Omega well plate reader (BMG Labtech GmbH, Ortenberg, Germany) using Greiner 96 F-bottom (non-binding) wellplates (Greiner Bio-One GmbH, Frickenhausen) in collaboration with Bruno Voigt (Prof. Dr. Jochen Balbach, MLU Halle).
All experiments were performed at 37 °C. Every 300 s, the ThT fluorescence was monitored at 480 nm with excitation at 450 nm. Samples were orbitally shaken with 300 rpm for 150 s prior to each measurement point. The sample volume was 150 μL with concentrations of 150 μM PTH_{1-84} and 50 μM ThT in citrate buffer (20 mM, pH 5.5) or sodium phosphate buffer (50 mM, pH 7.4). A series of heparin concentrations (0 μM, 3 μM, 7.5 μM, 15 μM, 30 μM, 150 μM, 600 μM) was tested to assess its role in PTH_{1-84} fibrillation. All measurements were done in triplicates with subsequent averaging of the ThT intensities.
ThT fluorescence curves were analyzed in collaboration with Dr. Maria Ott. Recent molecular rate kinetic descriptions of fibrillar growth enable to disentangle two main

fibrillation processes [154]. While primary nucleation describes the generation of structured nuclei in the absence of fibrils, secondary pathways address the generation of nuclei on fibrillar surfaces. Both processes affect the time dependence of the fibrillar growth in a different way and hence, their contributions to the total increase of fluorescence, $\Delta F(t)$, can be analyzed using the function

$$\Delta F(t) = \left(1 - \frac{1}{\frac{\lambda^3}{3\kappa^3}(e^{\kappa t} - 1) + 1}\right) \Delta F_{pl} \tag{5}$$

where ΔF_{pl} describes the plateau of the fluorescence assay. λ and κ are the rate constants of nucleation-and-growth for primary and secondary processes, respectively. λ and κ comprise the individual microscopic rates of oligomerization, k_{o1} and k_{o2}, respectively, the rate of conversion of the unstructured oligomers into seeds, k_c, as well as the rate of elongation via monomer addition k_+ [155, 156]:

$$\lambda = (2k_{o1}k_c k_+ m^{n1})^{1/3}, \kappa = (2k_{o2}k_c k_+ m^{n2+1})^{1/3} \tag{6}$$

The parameter m displays the total mass of the monomers and $n1$ and $n2$ the sizes of the nuclei for each process. In this context, λ would also include heterogeneous seeding, as it might be induced by surfaces.

2.18 Determination of the solubility of PTH

PTH_{1-84} fibrils (1 mL of 150 μM PTH_{1-84} monomer) were formed in presence of heparin (1:10 molar ratio) in pH 5.5 (20 mM citrate) or pH 7.4 (50 mM sodium phosphate) buffer at 37 °C and 350 rpm agitation for 3 h. The chosen incubation time ensured completion of fibril formation. Fibrils were pelleted via ultracentrifugation at 200,000 g for 10 min. The PTH_{1-84} concentration of the supernatant fraction was determined using UV spectroscopy and taken as the critical concentration, c^*.

2.19 Quantification of monomer release from fibrils

Fibril pellets were obtained as mentioned in section 2.18, resuspended in 100 μL pH 7.4 buffer (50 mM sodium phosphate, 5% mannitol) and subjected to dialysis through 20k MWCO Slide-A-Lyzer Mini dialysis devices (Fisher Scientific GmbH, Schwerte) against 1 mL pH 7.4 buffer (50 mM sodium phosphate, 5% mannitol). The protein concentration outside of the dialysis device was determined using UV spectroscopy at several time points.

2.20 Immunofluorescence imaging of human cell lines

Immunofluorescence imaging was performed in collaboration with Dr. Marcel Köhn at the Core Facility Imaging (Charles Tanford Centre, MLU Halle).
For immunofluorescence imaging (IF), the human cell line OVCAR-3 was transfected with mammalian expression plasmids encoding SRSF1 constructs (see section 2.6). More specifically, ~ 170,000 cells suspended in DMEM were seeded on a glass cover slip placed in a well plate. A liposome-based transfection approach (lipofection) was followed using Lipofectamine 3000 (Invitrogen, Karlsruhe) according to the manufacturer's protocol. Cells were incubated for 48 h at 37 °C with a medium exchange after 24 h.
To prepare for IF, the cells were washed with PBS and fixated on the cover slips with 4% formaldehyde. Cell membranes were permeabilized with 0.5% Triton X-100 in PBS. To block epitopes, 5% BSA in PBS was used for 30 min.
For co-staining nuclear speckles, cells were incubated with the primary antibody anti-SC-35 (SRSF2; 1:500 in 2% BSA) produced in mouse (ID: S4045; Sigma-Aldrich, Munich) for 1 h. Cells were washed with PBS and incubated with the secondary antibody anti-mouse (1:2000 in 2% PBS) conjugated to an Alexa Fluor 488 dye (DiaNova, Hamburg) for 30 min. Nuclei were stained with Dapi (1:2000 in 5% PBS) for 3 min. Dapi was removed and 5% PBS was added. Cover slips were washed twice in H_2O and ethanol. Dried cover slips were placed on a drop of mounting media on a glass slide with the cells facing downwards.

Images were recorded using a Leica LSM SP5X confocal microscope equipped with a 63x oil objective and controlled via the Leica Application Suite X software. Images were processed using ImageJ software [157].

2.21 Imaging of condensates *in vitro*

Phase contrast microscopy was performed to detect protein condensates of SRSF1 LCD constructs in collaboration with Dr. Nadine Bley (Core facility imaging, MLU Halle). A 40 μM SRSF1 LCD (AA 199-248) solution was prepared in 50 mM HEPES, pH 7.4 buffer, with either no further additions or charge-neutralizing amounts of RNA (poly-U, 20mer) or heparin (average mass: 20 kDa). Samples were pipetted into wells of a transparent Costar 96-well plate (Corning Inc., USA) and images were recorded using a Nikon TE2000E equipped with a 20x magnification objective.

For preliminary results concerning the phase behavior of several SRSF1 LCD mutants, light microscopy was used (Nikon S1, 40x objective). Peptides were dissolved in either 50 mM HEPES, pH 7.4 buffer, or 50 mM sodium phosphate (NaP_i), pH 7.4 buffer, and adjusted to 400 μM. Again, samples were pipetted into wells of a transparent Costar 96-well plate (Corning Inc., USA). For measurements at 4 °C, well plates were stored at 4 °C prior to imaging. In select samples, poly-U RNA (20mer) was added in a molar ratio of 1:40 (RNA:peptide).

2.22 Turbidity assays

Turbidity assays were performed to assess the condensation behavior of SRSF1 variants using a CLARIOstar Plus (BMG Labtech GmbH, Ortenberg) reader. All measurements were performed in Greiner 96 half area wellplates (Greiner Bio-One GmbH, Frickenhausen) with a sample volume of 35 μL. In a concentration/temperature dependent series of measurements, SRSF1 LCD was dissolved in HEPES buffer (pH 7.4) with varying final protein concentrations (1 μM, 5 μM, 10 μM, 25 μM and 50 μM). Poly-U RNA (20mer) was added in a

charge-neutralizing amount, assuming -1.5 charges per phosphate and -1 charge per Arg residue. Turbidity measurements were performed in absorption mode in a series of temperatures (25 °C, 30 °C, 35 °C and 40 °C) with incubation times of 2 min at each temperature step.

NaCl concentration dependent (0 mM, 50 mM, 100 mM, 200 mM, 500 mM and 1000 mM) measurements were performed at a protein concentration of 25 μM at 25 °C. Measurements were conducted in triplicates. Concentration/temperature/Ionic strength dependent measurements were performed with double orbital shaking at 300 rpm for 30 sec prior to each photometric readout (200 flashes).

The effect of an increase in the buffer's ionic strength on preformed protein:RNA condensates (35 μL of 25 μM protein with charge-neutralizing amount of poly-U RNA) was tested via 5 μL injections of a 5 M NaCl solution with subsequent double orbital shaking at 500 rpm for 180 sec.

2.23 Molecular dynamics simulations

Molecular dynamics (MD) simulations are powerful computational techniques to aid structural protein science. They provide detailed insights into the spatiotemporal behavior of proteins and thus can be used to complement datasets derived from experimental approaches.

Modelling of PTH$_{1-84}$

MD simulations on PTH$_{1-84}$ were conducted by Bargeen Turzo (Assoc. Prof. Dr. Steffen Lindert, Ohio state university, USA).

PTH$_{1-84}$ was modeled based on the solution NMR structure of PTH$_{1-34}$ (PDB: 1ZWA). The remaining amino acid sequence (i.e. PTH$_{35-84}$) was first modeled with Rosetta comparative modeling [158] by using the sequence of PTH$_{1-84}$ as the target sequence and the NMR structure of PTH$_{1-34}$ (PDB: 1ZWA) as the only template structure. A total of 100 structures were generated. Following this, the

lowest scoring structure was subjected to the Rosetta FloppyTail [159] application and 250 structures were generated. The representative structures, on which MD simulations were performed, were manually selected based on the criterium to be close to the experimentally derived radius of gyration, R_g, of 23 Å (lab-internal value, SAXS-derived).

MD simulations

The system was prepared by solvating it in a cubic water box with a 15 Å buffer and explicit modeling of water molecules was done using the TIP3 model [158]. After adding water molecules, the net charge of the system was +2 which was neutralized by adding 150 mM NaCl. The fully assembled system was subsequently minimized in the NpT ensemble (isothermal-isobaric ensemble). Short-range, non-bonded interactions were limited to a distance cutoff of 12.0 Å while the smoothing functions had a cutoff of 10.0 Å. For long-range electrostatic interactions, the particle-mesh Ewald (PME) [159] method was employed. The entire system underwent an energy minimization protocol consisting of 2000 steps.

Following the minimization, a temperature ramp was performed, gradually increasing the temperature from 60 K to 300 K over a simulation duration of 0.24 ns. The pressure was maintained at 1 atm using the Nosé-Hoover Langevin piston [160, 161] method. Distance cutoffs and electrostatic interactions were handled as described above. The equations of motion were integrated using the r-RESPA multiple time step scheme [161], updating short-range interactions every 1 step and long-range electrostatic interactions every 2 steps. The time step used for integration was 2 fs in all simulations. During this step, which lasted 0.29 ns, the atoms in the ”backbone” selection were restrained.

In the subsequent equilibration step, the temperature was maintained at 300 K using Langevin dynamics. The remaining inputs and parameters were kept identical except for the duration: The equilibration step lasted 1.00 ns, during which the backbone atoms of the protein were restrained.

Finally, the production run was conducted with all inputs and parameters as above. The production run lasted 25 ns and no atoms were constrained.
All the steps of the MD simulation in this study were performed using the NAMD package [162] and the CHARMM36 force field [163, 164].

2.24 Cross-linking mass spectrometry

Chemical cross-linking in combination with mass spectrometry (XL-MS) is a powerful technique to study protein structures and protein-protein interactions. Protein specific cross-linking reagents covalently connect two amino acids which are in close spatial proximity to each other. Cross-linkers with different spacer lengths (e.g. zero-, short-, medium-length) are able to bridge a multitude of distances between amino acids and thus may be described as molecular rulers. The obtained distance constraints may be integrated with other techniques employed in structural biology, such as cryo-electron microscopy and molecular dynamics simulations, to address biological questions that have been difficult to answer. In this work, XL-MS was used to reveal potential intra-protein interactions in PTH_{1-84}.

Cross-linking

Cross-linking mass spectrometry was performed in collaboration with Alessio Di Ianni (Prof. Dr. Andrea Sinz, MLU Halle).
50 μM PTH_{1-84} dissolved in 50 mM HEPES buffer, pH 7.4, was cross-linked with 5 fold molar excess of three cross-linkers (sulfosuccinimidyl 4,4'-azipentanoate (sulfo-SDA), N-(3-dimethylaminopropyl)-N'-ethylcarbodiimide hydrochloride (EDC)/N-Hydroxysulfosuccinimid (sulfo-NHS) and disuccinimidyl dibutyric urea (DSBU)) independently (see Figure 11). Immediately before incubation with the protein for 1 h at 20 °C, the cross-linking reagents were dissolved in neat DMSO (EDC, sulfo-SDA, DSBU) or H_2O (sulfo-NHS). While cross-linking using EDC/sulfo-NHS and DSBU occurs spontaneously in a one-pot reaction, sulfo-SDA is a photo-inducible cross-linker. Here, after protein labelling for 1 h, cross-linking reactions

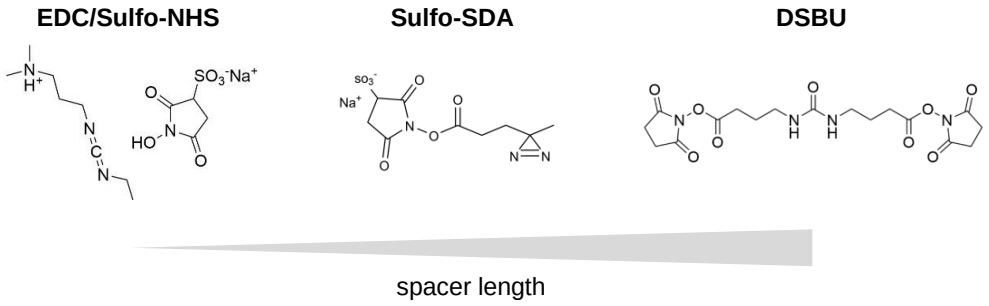


Figure 11: **Chemical cross-linkers used in this work.** Three cross-linkers with different spacer lengths were used in conjunction with mass spectrometry. EDC/sulfo-NHS is a zero-length cross-linker, while sulfo-SDA and DSBU are short (3.9 Å) and medium length (12.5 Å) spacers, respectively [165].

were performed using an ANUJ6186 diode (wavelength: 365 nm) controlled by an Aicure UJ30 (both Panasonic, Hamburg) at max settings for 30 sec irradiation into the opened reaction tube (0.5 mL reaction tube). Cross-linked samples were subjected to SDS-PAGE (15% gel) with a loaded protein mass of 4 μg PTH_{1-84}. Coomassie-stained gels were destained and bands corresponding to monomeric PTH_{1-84} were excised and cut into 1 mm^2 pieces. Gel pieces were destained using 100 μL of a 1:1 solution of 100 mM ammoniumbicarbonate (ABC) and acetonitrile (ACN). After incubation for 30 min while shaking, 500 μL ACN was added and incubated for 10 min to dry gel pieces. After removing all liquid from the tubes, gel pieces were covered with 40 μL of 100 mM ABC and in-gel enzymatic digestion was started with the addition of 20 μL of 12.5 ng/μL Glu-C (dissolved in 100 mM ABC). After digestion at 37 °C for 16 h, trypsin digestion was initiated by addition of 10 μL of 25 ng/μL trypsin (dissolved in 100 mM ABC) and subsequent incubation at 37 °C for 4 h. For peptide extraction, 80 μL of extraction buffer (1:2 solution of 5% TFA and ACN) was added and incubated at 37 °C for 15 min. All liquid was transferred to a new reaction tube and a second extraction was performed by addition of 80 μL of extraction buffer and incubation at 37 °C for 30 min. Again, all liquid was transferred to the new reaction tube. The extracted peptides were concentrated via a Savant SPD1010 SpeedVac concentrator (Thermo Fisher Scientific, Dreieich) to $<$ 50 μL. The solution was topped up to 50 μL with 0.1%

TFA to yield an LC-MS/MS-ready sample.

Mass-spectrometry

Peptide mixtures were subjected to LC-MS/MS. The setup included an Ultimate 3000 RSCL nano HPLC system (Thermo Fisher Scientific, Dreieich) coupled to a timsTOF Pro mass spectrometer equipped with a CaptiveSpray source (Bruker Daltonics, Bremen). Peptides were trapped on a C18 column (precolumn Acclaim PepMap 100, 300 μm x 5 mm, 5 μm, 100 Å; Thermo Fisher Scientific, Dreieich) and separated on a self-packed C18 column. After trapping, peptides were eluted by a linear 90 min water-acetonitrile gradient from 3%-30% (v/v) ACN. For elution, a constant flow rate of 300 nL/min was employed. The column was washed at a flow rate of 300 nL/min with the following gradient: 50% (v/v) to 85% (v/v) ACN (5 min), 85% (v/v) ACN (5 min), 85% (v/v) to 3% (v/v) ACN (5 min). All separations were performed at RT.

For the timsTOF Pro settings, the following parameters were adapted, starting from the PASEF method for standard proteomics. The values for mobility-dependent collision energy ramping were set to 95 eV at an inversed reduced mobility (1/k0) of 1.6 V s/cm^2 and 23 eV at 0.73 V s/cm^2. Collision energies were linearly interpolated between these two 1/k0 values and kept constant above or below. No merging of TIMS scans was performed. Target intensity per individual PASEF precursor was set to 20000. The scan range was set between 0.6 and 1.6 V s/cm^2 with a ramp time of 166 ms. 14 PASEF MS/MS scans were triggered per cycle (2.57 s) with a maximum of seven precursors per mobilogram. Precursor ions in a m/z range between 100 and 1700 with charge states $\geq$ 3+ and $\leq$ 8+ were selected for fragmentation. Active exclusion was enabled for 0.4 min (mass width 0.015 Th, 1/k0 width 0.015 V s/cm^2).

Data analysis was performed by Alessio Di Ianni (Prof. Dr. Andrea Sinz, MLU Halle). Cross-links were identified with MeroX 2.0.1.7 software using the settings described in Table 5.

Table 5: **MeroX settings for cross-link identification.**

Setting	**Description**
Cleavage pattern	C-terminal to Arg/Lys, N-terminal to Glu/Asp (3 missed cleavages allowed in both cases)
Peptide length	5 - 30 amino acids
Cross-linker specificity	
EDC/sulfo-NHS	K,S,T,Y,N-terminus,D,E
Sulfo-SDA	no specificity
DSBU	K,S,T,Y,N-terminus
Search algorithm	quadratic mode
Precursor mass accuracy	10 ppm
Fragment ion mass accuracy	20 ppm
Signal-to-noise ratio	$>$ 1.5
Precursor mass correction	enabled
False discovery rate (FDR) cut-off	1%
Minimum score cut-off	20

3 Results & Discussion

3.1 Heparin promotes rapid fibrillation of the basic PTH at physiological pH

There is evidence that peptide hormones undergo fibrillation as they pass through the acidifying secretory pathway [94]. Fibril deposits in secretory granules are discussed to function as a storage form for peptide hormones. Many of these peptides, including glucagon, urocortin III and corticotropin-releasing factor, spontaneously form fibrils *in vitro* under granule-relevant pH conditions [94]. PTH_{1-84}, however, preferentially forms fibrils at a pH close to its basic pI of 9.1 [113]. It was therefore hypothesized that fibril formation of PTH_{1-84} at granule-relevant pH conditions might be induced by polyanionic structures. To test this hypothesis, the interaction between PTH_{1-84} and the polyanion heparin was characterized. Furthermore, the impact of heparin on fibril formation of PTH_{1-84} was determined.

PTH_{1-84} binds heparin

Given the positively charged nature of PTH_{1-84} at pH 5.5, it was hypothesized that attractive coulomb interactions would drive its binding to the polyanionic heparin. To investigate the thermodynamics of the heparin:PTH_{1-84} binding, isothermal titration calorimetry (ITC) was employed (see section 2.15). The heparin:PTH_{1-84} interaction resulted in a negative isotherm, indicating an exothermic binding event (see Figure 12A). By using a model that assumes independent identical binding sites, the dissociation constant, K_D, was determined to be 213 $\pm$ 29 nM, and the stoichiometry, n, was found to be 0.09 $\pm$ 0.01, indicating an average of 11 PTH_{1-84} molecules bound per heparin molecule. The enthalpy change, ΔH, was calculated to be -72.5 $\pm$ 1.0 kJ mol^{-1}, while the entropy change, ΔS, was determined to be 0.89 $\pm$ 0.01 kJ mol^{-1} K^{-1}. Furthermore, the change in Gibbs free energy, ΔG, was calculated to be -38.0 $\pm$ 0.3 kJ mol^{-1}, which aligns well with a previously reported value of -37 kJ mol^{-1} by Middaugh and coworkers [166].

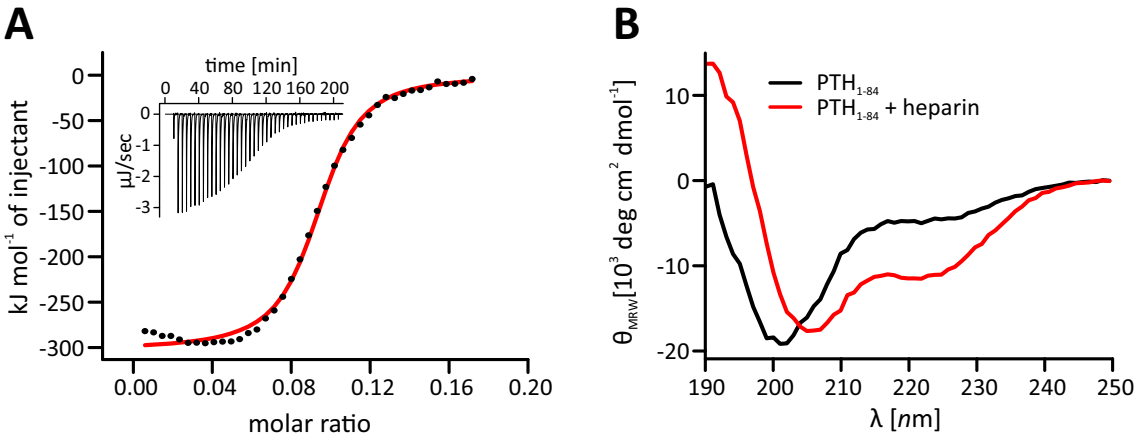


Figure 12: **Heparin binds PTH$_{1-84}$ and induces structural changes.** A: Isothermal titration calorimetric data of PTH$_{1-84}$ with heparin as displayed by the integrated heat plot (dots) together with a single-site ligand-binding model fit (red line). The x-axis gives the molar ratio of heparin vs. PTH$_{1-84}$. Inset: The raw titration data of the heat resulting from each injection of heparin into the PTH$_{1-84}$ solution. B: Circular dichroism spectrum of 100 μM PTH$_{1-84}$ in absence (black) and presence of heparin (1:10 molar ratio heparin:PTH$_{1-84}$, red line). The latter was corrected for heparin contributions. Figure adapted from [167].

Compared to the unfractionated heparin used in this study with an average mass of 20 kDa, Middaugh and coworkers used unfractionated heparin with a lower average mass of 12-16 kDa. Despite this difference, ΔG remained unaffected, while the stoichiometry exhibited a correlation with the size of the heparin molecules. More specifically, the present study determined a stoichiometry of ~ 11 PTH$_{1-84}$ molecules per heparin molecule, whereas Middaugh and coworkers observed a binding ratio of ~ 8 PTH$_{1-84}$ per smaller sized heparin. Considering an average mass of 533 Da per disaccharide unit, it can be inferred that one PTH$_{1-84}$ molecule binds ~ 3 disaccharide units of heparin. This finding supports the beads-on-a-string model previously proposed for other protein-heparin systems [168, 169]. Notably, at low ligand (heparin) concentrations, a slight enthalpic deviation of approximately 20 kJ mol^{-1} was observed, which may be attributed to a second binding process involving higher affinity and smaller exothermic enthalpy [153]. However, attempts to quantify this contribution by using multi-sites models yielded a high fitting error and were therefore not applied.

Heparin induces changes in the secondary structure of PTH$_{1-84}$

Circular dichroism (CD) spectroscopy was employed to assess the relative secondary structure composition of PTH$_{1-84}$ in a citrate buffer (20 mM, pH 5.5) in the absence and presence of heparin (see Figure 12B). In absence of heparin, PTH$_{1-84}$ gave rise to a CD spectrum characterized by a prominent negative ellipticity near 200 nm and a less pronounced minimum around 222 nm. This distinctive CD pattern is typical for IDPs with contributions from α-helices [143] and aligns with structural data obtained from various PTH constructs using X-ray crystallography and NMR spectroscopy [100, 170, 171]. In contrast, the presence of heparin resulted in a markedly different CD spectrum, characterized by a reduction in negative ellipticity around 200 nm and a significant increase in negative ellipticity within the 210-230 nm range. As irregular peptide structures exhibit relatively low optical activity in the latter wavelength range, the CD spectrum suggests an increase of ordered secondary structure elements in PTH$_{1-84}$ as an effect of its interaction with heparin. Interestingly, the CD spectrum analysis tool BeStSel [172] identified an elevated α-helical content in the presence of heparin (12% without heparin, 33% with heparin), accompanied by a decrease in the content of β-strands (25% without heparin, 13% with heparin, see appendix Figure 34). Comparing pH 7.4 to pH 5.5, a similar structural composition was observed for PTH$_{1-84}$, albeit with a less pronounced change in secondary structure upon heparin binding at pH 7.4. Based on these findings, it can be concluded that the polyanionic heparin not only binds to PTH$_{1-84}$ at low pH, potentially shielding its positive charges, but also induces a modification in its ordered secondary structure.

The size and shape of heparin:PTH$_{1-84}$ associates

Small-angle X-ray scattering (SAXS) experiments provide information about the size, shape, and interactions of dissolved particles, thereby enabling the examination of the molecular arrangement of PTH$_{1-84}$ in conjunction with heparin.

The absolute scattering intensities of solutions containing PTH_{1-84}, heparin or heparin:PTH_{1-84} (at a 1:10 molar ratio) are displayed in Figure 13A. PTH_{1-84} gives rise to a scattering intensity profile (represented by the black dataset in Figure 13A) which conforms to a linear polymer model of a flexible chain, as described by Equation 4, indicative of the disordered nature of the peptide. The application of this model yields a radius of gyration, R_g, of 2.9 $\pm$ 0.5 nm. In contrast, the shape of the heparin molecules (represented by the gray dataset) can be described by a semi-flexible chain model, which is defined by a Kuhn length, b, and a contour length, L. Here, L represents the average length of the heparin backbone and was determined to be $\approx$ 30 nm, while b reflects the stiffness of the chain and was determined to be 19.2 $\pm$ 0.7 nm as described in section 2.13. The cross-sectional radius of the chain, R_c, was 0.45 $\pm$ 0.03 nm. These values are consistent with previously reported data [146, 150].

Immediately upon the addition of PTH_{1-84}, the scattering intensity of the heparin solution increased by two orders of magnitude (depicted by the red dataset in Figure 13A). By applying the semi-flexible chain model to the heparin:PTH_{1-84} mixture, a reasonable agreement could be achieved without altering the previously described parameters (red line in Figure 13A). However, the chain's cross-sectional radius substantially increased from R_c = 0.45 $\pm$ 0.03 nm to 5.3 $\pm$ 0.2 nm, supporting the beads-on-a-string model mentioned earlier (illustrated in Figure 13B). Over time, the scattering intensities further increased at low q values, eventually reaching a plateau after 22 h (indicated by the colored datasets in Figure 13A). The exponent of the resulting power-law dependence, $I(q) \sim q^{D}$, corresponds to the dimension, *D*, of a mass fractal. *D* was determined to be 2.04, characteristic of diffusion-limited aggregation with preferred growth at the tips and reduced branching [173]. This slow and predominantly linear growth of heparin:PTH_{1-84} constructs is an initial indication of a fibrillar growth process.

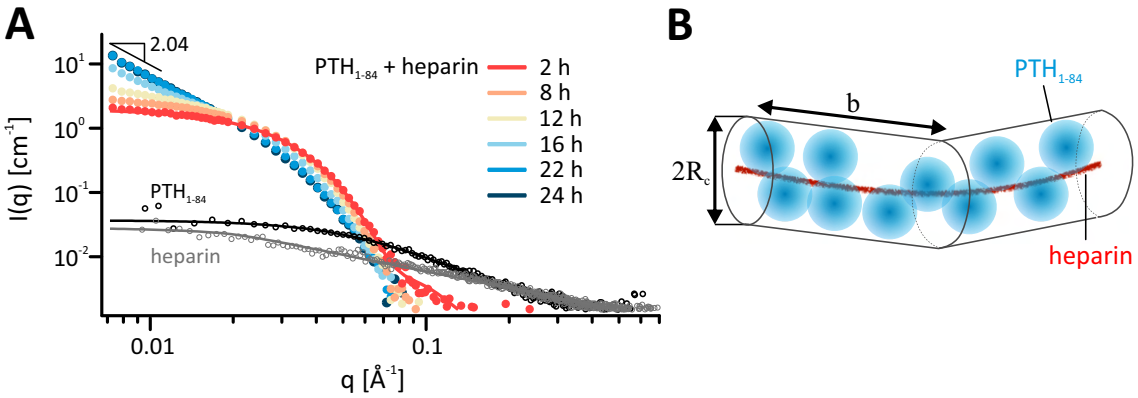


Figure 13: **Heparin induces aggregation of PTH$_{1-84}$.** A: X-ray scattering intensities of heparin (gray symbols), PTH$_{1-84}$ (black symbols) and a 1:10 molar ratio heparin:PTH$_{1-84}$ solution at different timepoints after mixing (red to blue symbols). Lines represent adapted linear chain model functions and are described in the text. All measurements were performed in citrate buffer (20 mM, pH 5.5). B: Scheme of the initial PTH$_{1-84}$ binding to a single heparin molecule, as described by the model of a semiflexible linear chain (red data set in C). Figure adapted from [167].

Heparin induces fibril formation of PTH$_{1-84}$

The predominantly linear growth of heparin-induced PTH$_{1-84}$ aggregates may suggest fibril formation. As determined by analytical ultracentrifugation (AUC), in absence of heparin, PTH$_{1-84}$ is monomeric in solution and a monodispers sample is yielded with no detectable aggregation (see Figure 14A-C). However, after incubation with heparin, aggregates in the form of fibrils were indeed detected using transmission electron microscopy (TEM) (see Figure 14D). In the TEM images, long and slightly curved fibrils of about 16 nm thickness can be seen. Branching was rarely seen, which aligns with the fractal dimension obtained from SAXS experiments. The cross-β sheet structure of amyloid fibrils with characteristic reflections at 4.7 Å and 10.1 Å (see Figure 14E) [87, 113] was confirmed by WAXS experiments. Overall, it can be concluded that heparin induces fibrillation of PTH$_{1-84}$ at granule-relevant pH.

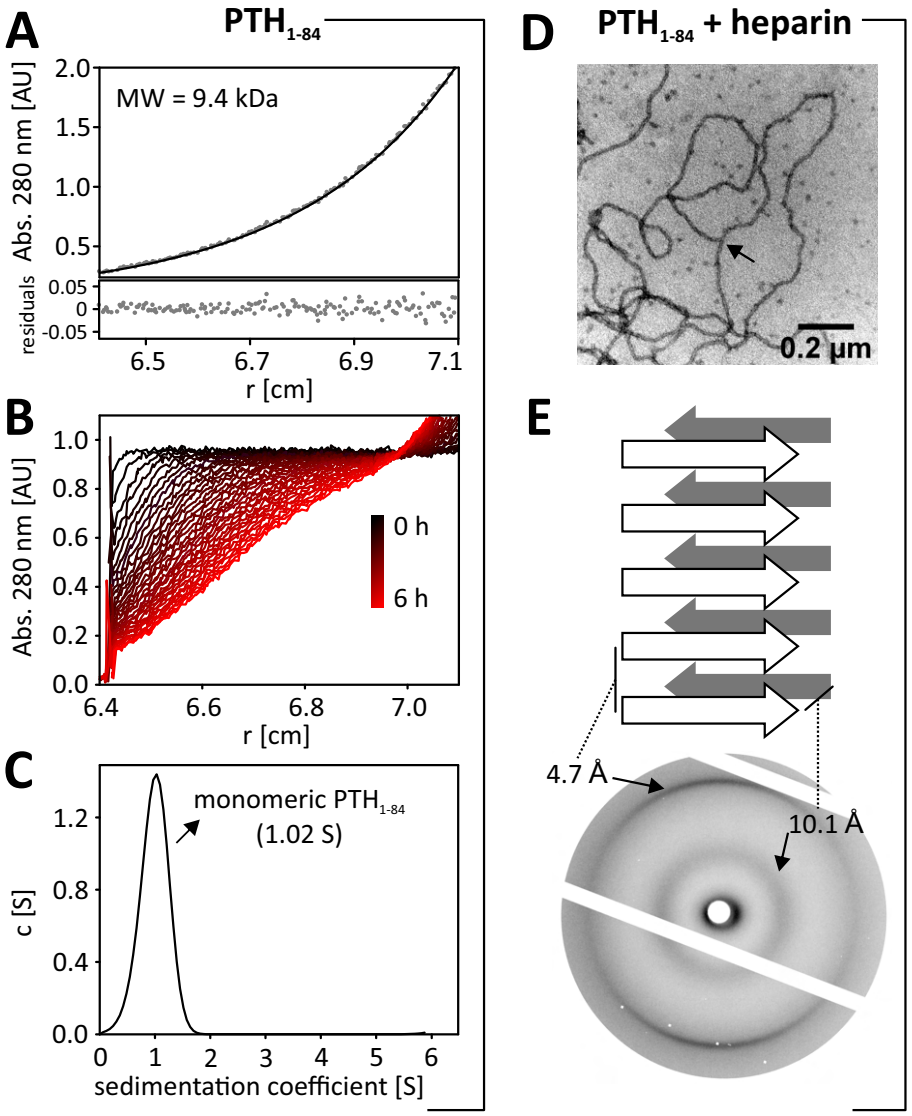


Figure 14: **Heparin induces PTH$_{1-84}$ fibril formation.** A-C: Analytical ultracentrifugation experiments with PTH$_{1-84}$ in absence of heparin at pH 5.5. Data obtained in a sedimentation equilibrium study (A) indicates the presence of monomeric PTH$_{1-84}$ (9424 Da). Homogeneity of the sample was confirmed in a sedimentation velocity experiment (B-C). D: Aggregates formed in presence of heparin were analyzed using transmission electron microscopy. Images display PTH$_{1-84}$ fibrils with a curvilinear structure and sporadic fibrillar branching (arrow). E: X-ray diffraction of PTH$_{1-84}$ fibrils formed in presence of heparin at pH 5.5. Reflections at 4.7 Å and 10.1 Å are consistent with a cross-β structure. Figure adapted from [167].

Fibrillation properties

As PTH$_{1-84}$ requires a partial conformational change from α-helices to β-strands to form amyloid fibrils with a cross-β structure [113], the growth process was monitored using Thioflavin T (ThT), a fluorescent dye that binds to β-sheet surfaces.

In the absence of heparin, there was no detectable fibril buildup from the initially homogeneous and monomeric PTH_{1-84} (see Figure 15A, black dots), even after 70 h of incubation at 37 °C and with orbital shaking. Subsequent centrifugation of the incubated sample did not result in a visible protein precipitate, suggesting that PTH_{1-84} remained in a soluble and monomeric state throughout the incubation time.

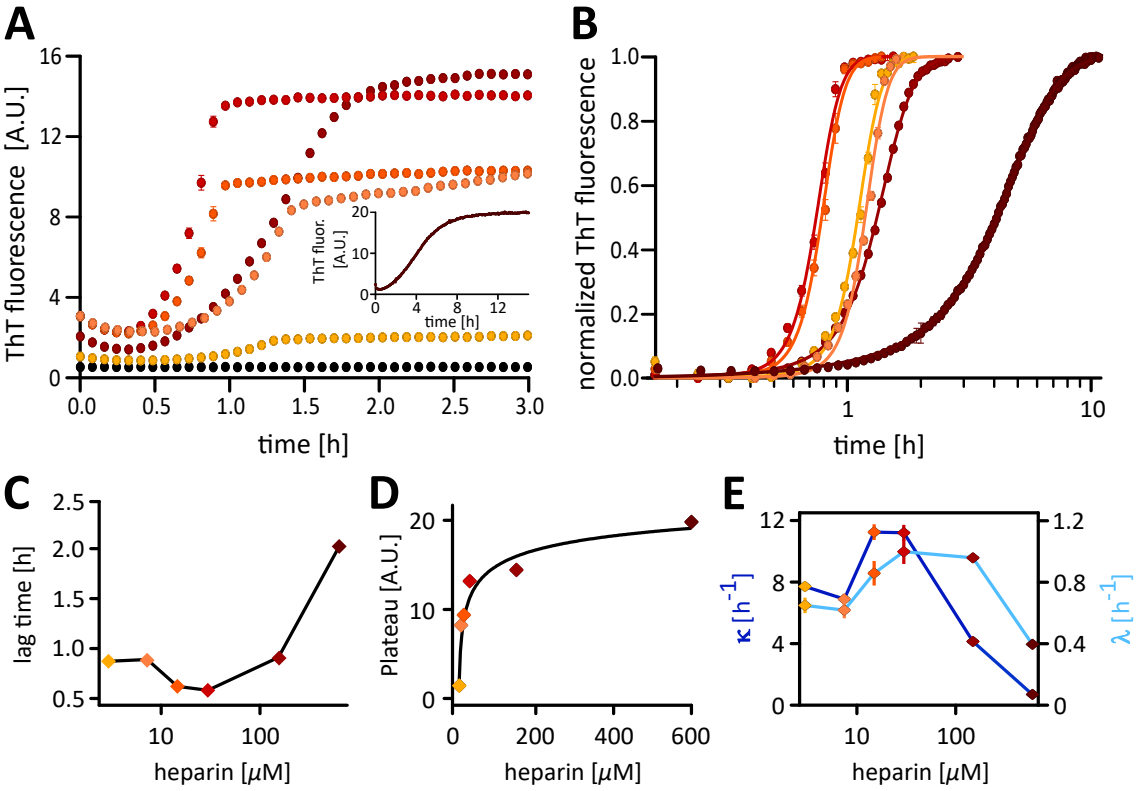


Figure 15: **Kinetics of heparin-induced PTH_{1-84} fibril formation.** A: ThT fluorescence was used to monitor fibrillation of PTH_{1-84} (150 μM, 37 °C, 300 rpm orbital shaking) at several heparin concentrations (black - no heparin; yellow - 3 μM; light orange - 7.5 μM; orange - 15 μM; red - 30 μM; dark red - 150 μM; purple (inset) - 600 μM). Only in presence of heparin, the ThT fluorescence increased significantly with heparin concentration dependent lag times (C) and plateau values (D). Data points were averaged from triplicates and shown with standard deviation error bars. B: Normalization and subsequent fitting using eq. 5 yielded heparin dependent nucleation and growth rates λ (primary fibrillation pathways) and κ (secondary fibrillation pathways) (E). Figure adapted from [167].

In contrast, when heparin was present, PTH_{1-84} rapidly formed fibrils. The ThT fluorescence intensities (plateau value) increased with increasing heparin concentrations until saturation was reached eventually (see Figure 15D). This

suggests that heparin may bind to the fibrillar surfaces as found for fibrils formed by the amyloid beta peptide [174]. At heparin:PTH_{1-84} ratios of 1:10 and 1:5 (15 μM/30 μM heparin), the lag time passed a minimum value of 0.6 h (see Figure 15B+C). A seeding mechanism alone would result in a purely heparin concentration dependent decrease in lag times [175]. Therefore, the heparin-induced fibril formation of PTH_{1-84} appears to follow a more complex aggregation process rather than being solely driven by seeding.

To gain further insights into the predominant mechanisms of fibrillation, the dataset was examined with respect to the overall rates of primary and secondary fibrillation pathways, λ and κ, as given by Equation 5. The fitting results are displayed in the normalized dataset of Figure 15B and the derived values of λ and κ are given in Figure 15E. While λ reflects the generation of structured nuclei in the absence of fibrils, κ addresses the generation of nuclei on the surface of the growing fibrillar phase [154]. It is important to mention that both rates comprise not only their specific microscopic rates of nucleation but also the more general rate of fibrillar growth via monomer addition (see Equation 6). For concentrations below the critical molar ratio of 1:10 (heparin:PTH_{1-84}), one can expect that not all of the PTH_{1-84} molecules are bound to heparin. For low concentrations (up to 7.5 μM, 1:20 ratio) it was found that both rates do not depend on heparin levels. Higher heparin concentrations caused both rates to increase until a plateau of about 150% of their initial values was reached. The significant increase of κ implies that heterogeneous seeding, i.e. the heparin molecules act as seeds on their own, cannot be the leading mechanism by which heparin induces fibrillation. Heterogeneous seeding would affect λ only and, hence, lead to an increase of the nucleation-to-growth factor $\lambda^3/3\kappa^3$ of Equation 5. The factor was not only found to be minor, indicative for reactions in which the nucleation step is rate-limiting [176], but also to be independent from the heparin concentration with, e.g., $(2.0 \pm 0.2) \cdot 10^{-4}$ for the 1:50 molar ratio and $(1.5 \pm 0.3) \cdot 10^{-4}$ for the 1:10 molar ratio (heparin:PTH_{1-84}). Hence, the elevated values need to be either explained

by an increased rate of structural conversion into seeds, or by fibrillar growth (k_c or k_+ of Equation 6). If the latter would be true, λ and κ should be still enhanced at higher heparin concentrations, at which most of the PTH_{1-84} molecules are bound to heparin. However, at heparin concentrations above the critical molar ratio of 1:10, a plateau of λ and κ was observed followed by a decrease (see Figure 15E), meaning that heparin-bound PTH_{1-84} can form primary and secondary seeds but cannot contribute to the fibrillar growth as growth seems to require unbound PTH_{1-84}-molecules. Hence, the leading mechanism of heparin at high and low concentrations is the structural conversion of bound PTH_{1-84} into amyloid-prone seeds.

The heparin-induced fibril formation of PTH_{1-84} is pH dependent

The dominance of electrostatic forces in the interactions between heparin/GAGs and proteins is well-documented [177]. As the charge state of PTH_{1-84} is sensitive to variations within the physiological pH range, it was hypothesized that changes in pH conditions may affect the efficiency of fibrillation. To check for this potential pH dependence, PTH_{1-84} fibrils were formed in presence of heparin in a 3 h incubation at 37 °C in buffers adjusted to pH 5.5 or pH 7.4. Subsequently, the fibrillar content was separated through ultracentrifugation (see section 2.18). Importantly, it was confirmed that immediately after mixing, heparin:PTH_{1-84} complexes were not pelleted by ultracentrifugation. This suggests that the fibrils that formed within the 3 h incubation were the primary species pelleted. Assuming fundamental thermodynamic principles, the remaining protein concentration in the supernatant should approximate the critical concentration for fibril formation [176]. According to this procedure, the critical concentration for fibril formation, c^*, was determined to be ~ 2.6 times higher at pH 7.4 compared to pH 5.5 (see Figure 16A). This observation suggests that the conversion of monomers into fibrils is more efficient at lower pH levels which is in line with pH-dependent plateau intensity values of ThT fluorescence experiments. According to the crystallization model by Finke and Watzky [176], plateau values of ThT experiments, ΔF_{pl}, do not only directly

correlate with the fibrillar mass but also with the supersaturation of the solution, $\sigma = (c - c^*)/c^*$. This leads to the following equation:

$$c^*_{pH7.4}/c^*_{pH5.5} \approx c/c^*_{pH5.5} - (\Delta F_{pl}^{pH7.4}/\Delta F_{pl}^{pH5.5})(c/c^*_{pH5.5} - 1)$$

where c is the total protein concentration in the sample. Using this equation with the experimentally obtained values from $c^*_{pH5.5}$ of 17 μM and the ThT plateau intensities of the PTH_{1-84} fibrillation (see appendix Figure 35), these considerations would lead to a ratio $c^*_{pH7.4}/c^*_{pH5.5}$ of 3.2. Although slightly higher, this value is in reasonable agreement with the above mentioned value obtained through ultracentrifugation experiments. Potential pH dependent changes in the spectral properties of ThT were considered by control measurements and thus cannot explain the differences. However, increased surface charges in PTH_{1-84} caused by low pH may modify the interactions of ThT with the fibrils [178].

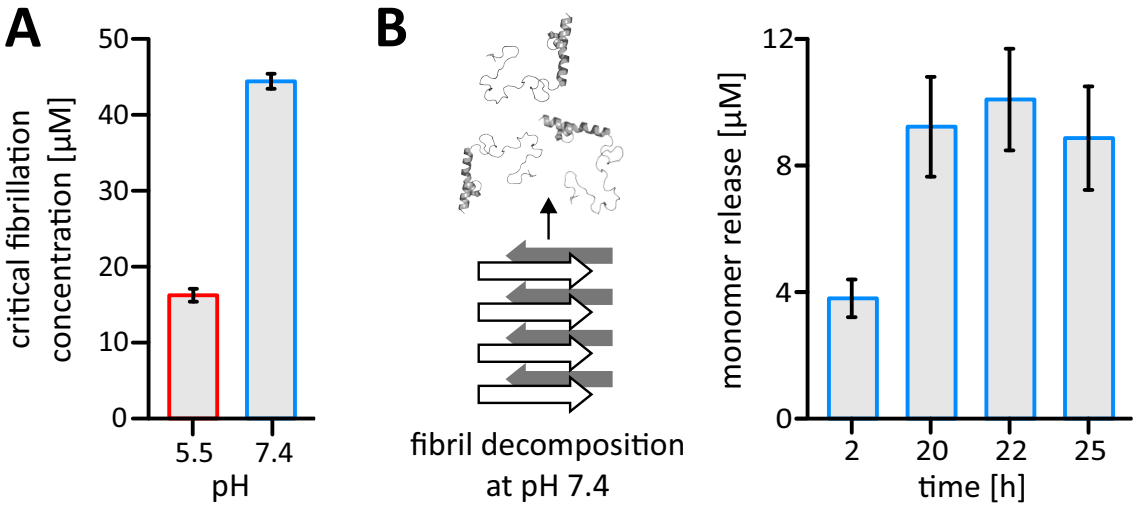


Figure 16: **pH dependent critical concentration for fibril formation and monomer release.** A: At pH 5.5, the critical concentration for fibril formation is reduced compared to pH 7.4. Fibrillation was induced with 1:10 molar ratio heparin:PTH_{1-84}. Data points are averaged from triplicates with standard deviation error bars shown. B: PTH_{1-84} fibrils formed in presence of heparin at pH 5.5 release monomers at pH 7.4 in a time dependent manner. Data points are averaged from triplicates with standard deviation error bars shown. Figure adapted from [167].

According to the relation $\Delta G = -RT \ln(c^*)$, the calculated energetic advantage

of fibril formation is $\Delta G \approx 22.38\ k$J/mol at pH 5.5 and $\Delta G \approx 20.05\ k$J/mol at pH 7.4. These values are lower compared to previously published values of ΔG~$30 - 50$ kJ/mol [179] and may reflect on the functionality and potential reversibility of the fibrils [85]. Surprisingly, according to the relation $\Delta(\Delta G) = -RT \ln(c^*_{7.4}/c^*_{5.5})$, the energetic advantage of fibril formation at pH 5.5 is only $\Delta(\Delta G) \approx 2.3 - 2.8\ k$J/mol. Thus, compared to the total gain of free energy for amyloid fibrillation, lowering the pH leads only to a small, but apparently significant decrease in ΔG.

The observed pH dependence of fibril formation points towards a heparin:peptide interaction that is mainly charge driven and modulated by the protonation state of the four histidine residues within PTH_{1-84}. Due to histidine's unique pK_A value close to the physiological pH (pH ≈ 6.2), their side chains are mainly unprotonated at pH 7.4 while protonated at pH 5.5. This switch-like property of histidines has been shown to be essential in numerous biochemical processes such as pH sensoric functions [180, 181] as well as the modulation of GAG:protein interactions [182].

However, PTH_{1-84} lacks a classical heparin binding consensus sequence (XBBXBX / XBBBXXBX, with X being a hydropathic and B a basic residue) [183], even when accounting for protonated histidines. Yet, these particular amino acid sequences are not a prerequisite for heparin binding as some proteins lacking them retain this capacity [184]. In these cases, heparin-binding is discussed to occur through structurally similar binding motifs, in which the involved residues may be distant in the primary structure but align to a structural binding motif [184]. In case of the largely disordered PTH_{1-84} this modality of binding seems particularly intriguing. PTH_{1-84} might dynamically adopt a high affinity conformational state with particular residues arranged in a structural heparin binding motif. At granule-relevant low pH, protonated histidines might partially form a structural heparin binding motif in PTH_{1-84}, while unprotonated histidines might lose this capacity. Similarly, seemingly minute changes in the sequence of peptides have already been reported to potentially have significant effects on the GAG-mediated

fibril formation [185].

To gain more insights into the binding sites of PTH_{1-84} for GAGs, a residue resolved 2D NMR experiment in the presence and absence of fondaparinux was recorded (see appendix Figure 36). As a pentameric heparin sulphate analog, fondaparinux has a well defined molecular weight of 1728 Da and is therefore suitable in size for NMR spectroscopy [186]. Through chemical shift analysis, heparin binding was found to affect the first 40 residues such as H10, K13, H14, R20, R25, K26, K27, and H32 with the latter four residues contributing to the fibrillar core [113]. Additionally, the residues in the basic stretch R52, K53, and K54 were affected by heparin binding. The involvement of a considerable fraction of basic residues in heparin binding further supports the hypothesis that PTH_{1-84} adopts one or several high affinity states for heparin binding. Overall, through the interaction with heparin, PTH_{1-84} likely overcomes the energetic penalty associated with increased intra-chain repulsion at low pH enabling fibril formation. Compared to neutral pH, fibril formation is even more efficient.

The release of monomers from PTH_{1-84} fibrils

A key feature of typical functional amyloid fibrils is that they are reversible structures, i.e. they are capable of releasing monomers in solution [187–190]. In case of the PTH system, receptor activation involves the binding of the ligand, PTH_{1-84}, to its receptor, e.g. PTH1R. Resolved structures of the ligand in complex with the receptor [191–193] indicate a delicate and tight interplay between the two suggesting that receptor activation requires the ligand in the monomeric form. Therefore, it was hypothesized that fibrils formed at pH 5.5 in presence of heparin might release monomers in a buffer adjusted to neutral pH to approximate blood-like conditions. To test this, a dialysis approach was followed using a cutoff filter of 20k MWCO to let PTH_{1-84} monomers pass the membrane (for more details see section 2.19). Here, the PTH_{1-84} monomer concentration outside of the dialysis membrane increased over time (see Figure 16B) which indicates a time dependent monomer release. This matches reports of similar functional amyloids

[187, 194] and agrees well with an earlier study by Balbach and coworkers, who found a similar release of monomers from PTH_{1-84} fibrils [113]. However, the here used dialysis approach to quantify the monomer release over time might yield underrepresentative values since complex formation of heparin and freshly released monomeric PTH_{1-84} is likely to occur and may interfere with the diffusion across the dialysis membrane.

Summary

A major aim of the present *in vitro* investigation was to approximate two key environmental conditions that may affect the fibril formation of PTH_{1-84} in the cellular context (see Figure 17): The pH was adjusted to a granule-relevant acidic value (pH 5.5) to mimic the conditions in which PTH_{1-84} fibrils are supposed to form *in vivo*. This low pH condition was combined with the addition of the GAG heparin, as GAGs are well-documented constituents of the Golgi apparatus. By using heparin, the effect of a linear polyanionic macromolecule, that resembles potential intracellular interaction partners of PTH_{1-84}, could be studied *in vitro*. Under these conditions, it was found that PTH_{1-84} binds with a sub-micromolar affinity to heparin. Heparin induced ordered secondary structure and aggregation of PTH_{1-84} and was found to be indispensable for rapid PTH_{1-84} fibril formation. This accelerated kinetic is consistent with the potentially short lifetime of PTH_{1-84} in cells which has been determined to be as low as 30 min by using pulse-chase experiments [195, 196]. It further agrees with the proposed property of functional amyloids according to which they need to form more rapidly compared with disease-associated fibrils to avoid accumulation of potentially toxic oligomers [85]. Notably, the mean β-strand content as determined by CD spectroscopy did not correlate with individual seeding capabilities (comparison of PTH_{1-84} free and bound to heparin at low pH) nor the efficiency of fibril formation (comparison of PTH_{1-84} fibril formation with heparin at high and low pH) for heparin:PTH_{1-84} associates. This finding underscores the importance of heparin or equivalent

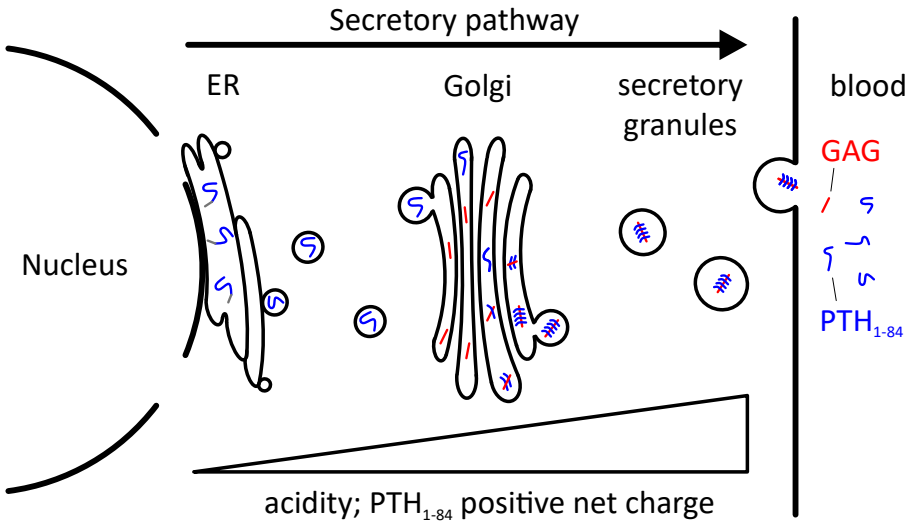


Figure 17: **Model of intracellular fibril formation of PTH$_{1-84}$.** Illustration of PTH (blue) passing through the continuously acidifying secretory pathway. The increasing positive net charge of PTH$_{1-84}$ may drive interactions with glycosaminoglycans (GAGs) that are abundant in the secretory pathway [138, 197]. Based on the here presented *in vitro* data, it is proposed that GAGs interact with PTH$_{1-84}$ in cells and induce its fibrillation. Figure adapted from [167].

polyanionic structures and their interactions during protein assembly.

Structurally, heparin:PTH$_{1-84}$ assemblies can be described as semiflexible heparin chains decorated with several PTH$_{1-84}$ molecules each. Fibril formation is induced by enhancing both, the primary and secondary nucleation pathways, predominantly by structural conversion of PTH$_{1-84}$ into amyloid-compatible seeds. It is conceivable that fibril formation of PTH$_{1-84}$ might follow a similar mechanism *in vivo* possibly relying on endogenous GAGs or glycosylated proteins (see Figure 17). This proposed relationship is supported by previous studies which revealed that PTH$_{1-84}$ is co-stored and co-secreted with proteoglycans [196, 198].

Strikingly, the heparin:PTH$_{1-84}$ interaction enabled more efficient fibril formation at the granule-relevant low pH compared to neutral pH. The high efficiency of fibril formation, as indicated by a reduced critical fibrillation concentration at the lower pH, suggests that the polyanion-driven fibril formation of PTH$_{1-84}$ is enhanced under pH conditions present in later stages of the secretory pathway, e.g. the trans-Golgi network and secretory granules. The pH sensitivity of this process might have a regulating function for sorting and further transport of PTH$_{1-84}$. A

regulatory role of the GAG-assisted formation of functional amyloid fibrils in the secretory pathway was previously proposed [94].

3.2 The intrinsically disordered C-terminus of PTH affects its fibrillation behavior

Experimental data indicate that PTH_{1-84} is a polypeptide without a stable three-dimensional fold but with an α-helical propensity, especially in the N-terminal region [100, 101, 106]. A well-defined structure of the C-terminal region, however, was neither resolved by NMR spectroscopy nor by X-ray crystallography [106], raising the possibility that this region lacks a stable fold and is rather flexible. It is well known that the N-terminus, more specifically PTH_{1-34}, is necessary and sufficient for PTH1R activation [104, 105]. Yet, the function of the C-terminal region (PTH_{35-84}) remains largely unknown. In this work, it is intended to investigate a potential functional role of the PTH_{35-84} regarding PTH_{1-84} fibrillation.

The C-terminus of PTH_{1-84} is highly sequence conserved and structurally disordered

Functional properties of proteins are subject to evolutionary selective pressure. Therefore, the investigation of sequence conservation provides valuable information for protein scientists. Using multiple sequence alignment (MSA), the sequence conservation of PTH_{1-84} was analyzed among higher mammals (see Figure 18). As expected, the mostly structured N-terminal region (PTH_{1-34}) is extraordinarily conserved (79% sequence identity, 97% sequence similarity) among the tested mammals, reflecting the delicate and specific nature of receptor activation [191–193]. Although to a lesser extent, the remaining C-terminal region (PTH_{35-84}) is highly conserved, as well. A sequence identity of 46% and sequence similarity of 70% indicates that the residues in this region are subject to high selective pressure. This suggests that PTH_{35-84} might be critical for proper function of the peptide hormone.

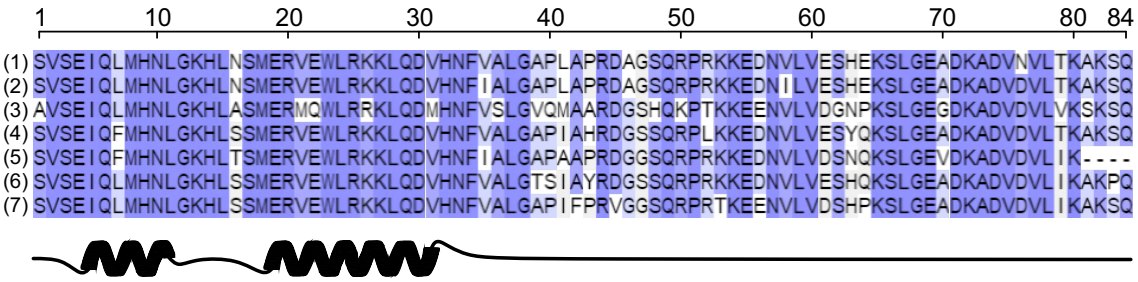


Figure 18: **The sequence of PTH$_{1-84}$ is highly conserved.** A multiple sequence alignment (MSA) of PTH$_{1-84}$ is displayed using peptide sequences from (1) *Homo sapiens*, (2) *Macaca mulatta*, (3) *Mus musculus*, (4) *Canis lupus familiaris*, (5) *Loxodonta africana*, (6) *Orcinus orca* and (7) *Dasypus novemcinctus*. The MSA was performed using the UniProt Align tool (based on CLUSTAL Omega) and colored according to similarity. Secondary structure elements are displayed spanning the sequence.

While the α-helix-rich N-terminal PTH$_{1-34}$ was successfully resolved via NMR spectroscopy and X-ray crystallography [100, 171], the C-terminal PTH$_{35-84}$ seems to be unamenable to these high resolution structure elucidation techniques, possibly because of an inherent lack of a stable structure in that region of the peptide. To test this hypothesis, we compared the secondary structure content of various PTH variants (PTH$_{1-84}$, PTH$_{1-34}$ and PTH$_{35-84}$) using CD spectroscopy (see Figure 19A). As described above (see section 3.1), the full length PTH$_{1-84}$ gave rise to a CD spectrum with a pronounced negative ellipticity close to 200 nm and a more shallow minimum around 222 nm indicative of a largely disordered peptide with α-helical content [143]. The CD spectrum of the N-terminal fragment PTH$_{1-34}$ closely resembles the basic spectrum of an α-helix (see Figure 10 for comparison) with small contributions of random coil, matching NMR derived structures showing two helices joined by a flexible hinge region [100]. In contrast, the CD spectrum of the C-terminal fragment PTH$_{35-84}$ indicates that a major part is disordered with limited α-helical propensity [143]. A complementary dataset obtained from infrared (IR) spectroscopic measurements confirms the finding that PTH$_{1-84}$ is a peptide divided into a mainly α-helical N-terminus and a disordered C-terminus (see Figure 19B). This agrees with the previously proposed structural properties of the peptide hormone [100, 101, 106]. In summary, the C-terminal

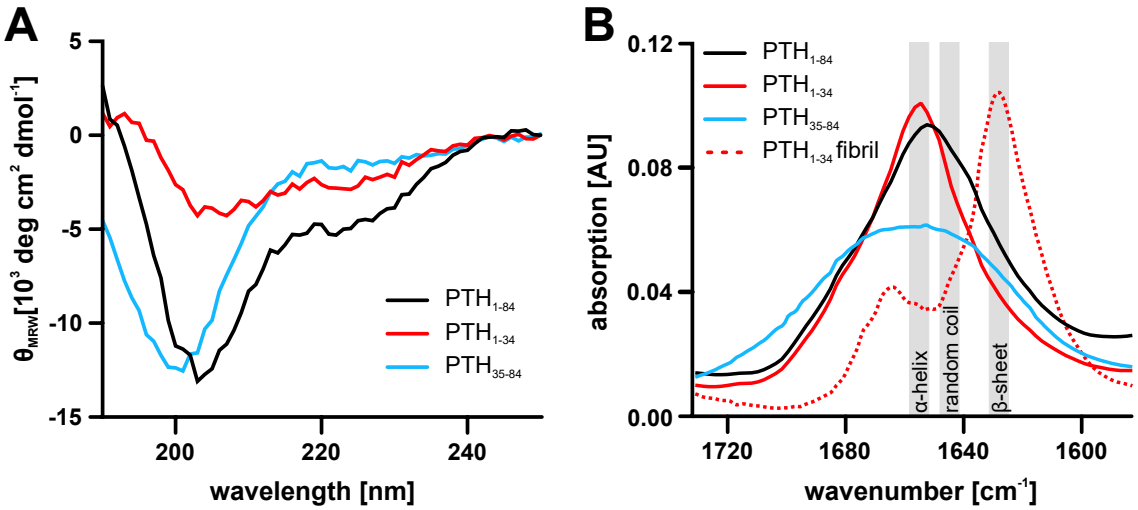


Figure 19: **Secondary structure elucidation of PTH variants.** A: Circular dichroism spectra of PTH variants (PTH_{1-84} - black, PTH_{1-34} - red, PTH_{35-84} - blue) are shown. For all measurements, PTH variants were dissolved in 50 mM HEPES buffer, pH 7.4, and concentrated to 100 μM. B: Infrared spectra of PTH variants in solution (solid line: PTH_{1-84} - black, PTH_{1-34} - red, PTH_{35-84} - blue) and PTH_{1-34} fibrils (red dotted line) are shown. Typical ranges of secondary structure elements are indicated by gray bars [145].

region of PTH_{1-84} displays a highly conserved amino acid sequence alongside with structural disorder as confirmed by CD and IR spectroscopy. This finding contrasts the tendency for a high residue variation typically associated with the evolution of disordered regions [11] and therefore raises the question which yet unknown sequence specific functions PTH_{35-84} may have.

The disordered C-terminal region forms contacts with the α-helical N-terminus

With the secondary structural characteristics of PTH_{1-84} established, the spatiotemporal behavior of the polypeptide chain should be investigated in more detail. An MD simulation on a single PTH_{1-84} molecule was performed. In Figure 20, the time-averaged intra-chain dimensions over a 25 ns simulation are displayed in form of a heatmap displaying time-averaged inter-residue (C_α-C_α) distances. Here, the two N-terminal α-helices in PTH_{1-84} gave rise to a very regular distance pattern, indicating that these secondary structures are stable throughout the simulation. Indeed, sampling through all 3D structures created during the simulation confirmed

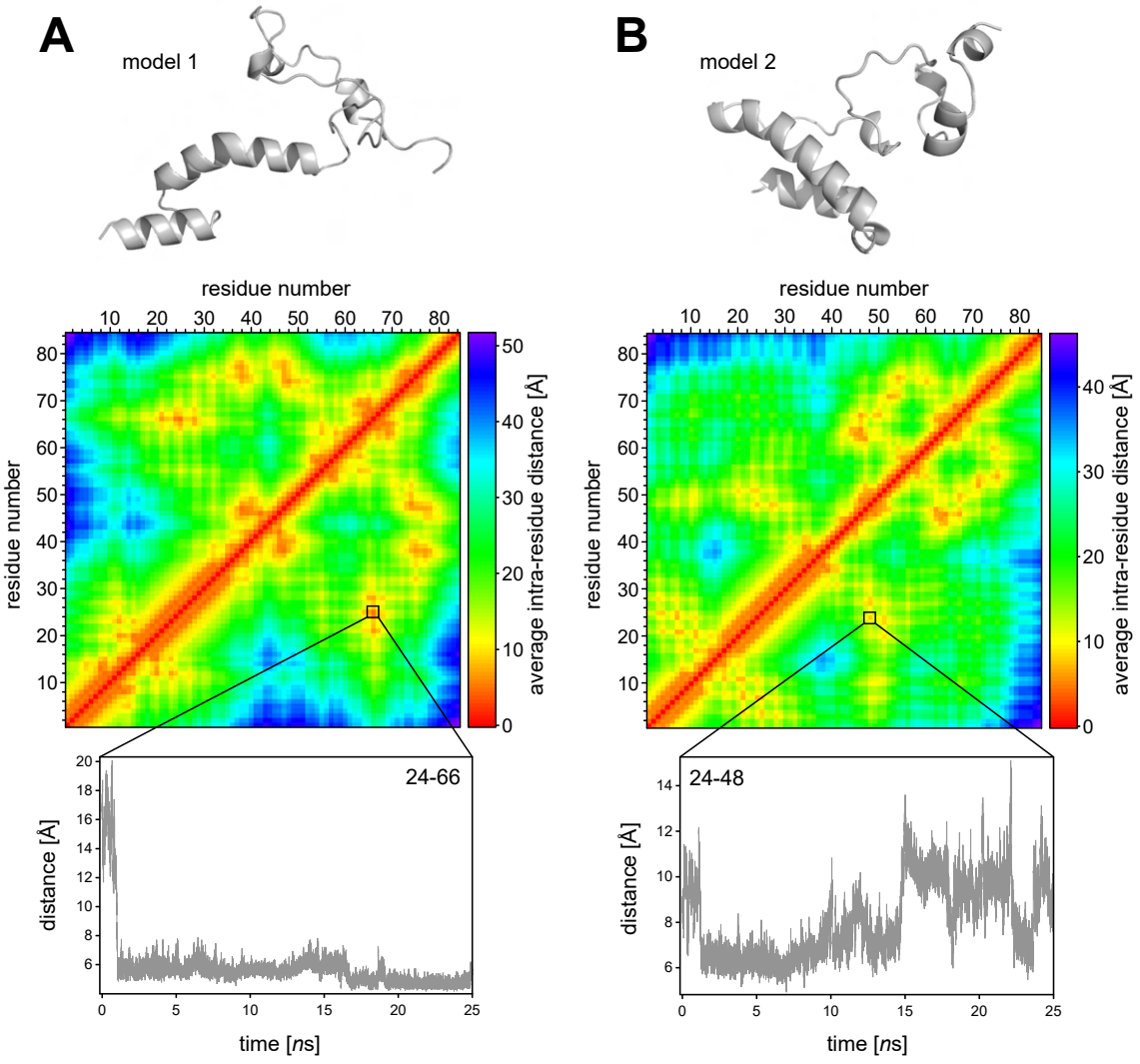


Figure 20: **MD simulations reveal long-range intra-chain contacts in PTH$_{1-84}$.** Two heatmaps displaying the time-averaged inter-residue (C_α-C_α) distances of PTH$_{1-84}$ (center images) were generated from independent 25 ns MD simulations with separate initial models (upper panel, A: model 1; B: model 2). Time traces of selected long-range contacts are displayed (lower panel).

this structural stability which is in agreement with the predominant peptide architecture consisting of two short α-helices as observed in solution NMR structures [100, 199, 200]. The C-terminal part of PTH$_{1-84}$ remained structurally disordered throughout the simulation with sporadic formations of transient α-helices. When focusing on longer ranged interactions within the peptide, it was found that parts of the disordered C-terminal region were averagely close in proximity to the second α-helix (here defined as distances below 12 Å). To gain more insights into these prevalent long range contacts, the temporal behavior was analyzed for the residue-

pair 24-66, with residue 24 being part of the second α-helix and residue 66 being part of the disordered C-terminal region (see Figure 20A). Over the course of the MD simulation, the inter-residue distance (C_α-C_α) between residue 24 and 66 strongly reduced from 16 Å to 6 Å within 2 ns and maintained in close proximity for the remaining 23 ns of the simulation. Considering the additional space occupied by the respective residues, this distance is sufficiently short for both, long- and short-range protein-protein interactions such as salt bridges [201]. This finding suggests the presence of N- to C-terminal intra-chain interactions in PTH_{1-84}. In a second MD simulation using a different initial model, this result could be partly reproduced (see Figure 20B), although with slightly different residues involved in the potential interaction (residue pair 24-48). However, in a third independent simulation, no N- to C-terminal contacts could be discerned from the heatmap (see appendix Figure 37) if the threshold is defined as a distance of 12 Å - yet, relative to ideal chain dimensions, the peptide was still compacted in the relevant regions allowing for potential intra-chain interactions to develop. It has to be noted that any differences in the chain dimensions from separate MD simulations may arise from the system not reaching equilibrium. This is a well-known problem concerning MD simulations on IDPs as their conformational space is large and may be solved by extending the duration of the MD simulations [202]. Yet, overall, the reasonably consistent dataset suggests a prevalent conformation in which the second α-helix in the N-terminal region resides in close proximity to the disordered C-terminal region.

To gain more insights into the potential N- to C-terminal intra-chain interactions of PTH_{1-84}, cross-linking experiments were performed and analyzed using mass spectrometry (XL-MS). Three cross-linkers covering a range of zero- to medium-sized linker lengths were used (see Figure 21, Figure 11). To ensure that obtained cross-links were indeed intra-chain and not inter-chain cross-links, an SDS-PAGE was performed to isolate the monomeric protein fraction (see appendix Figure 38). DSBU, a medium-length cross-linker, yielded the most intra-chain cross-links which

were distributed across the whole sequence. The shorter the linker length was, the less intra-chain cross-links were identified: The short linker sulfo-SDA still yielded numerous cross-links within the N-terminal region of PTH_{1-84} as well as one N- to C-terminal cross-link. The zero-length cross-linker EDC/sulfo-NHS yielded a single cross-link within the N-terminal region and a second one bridging the N- to C-terminal regions. Exemplary fragment ion spectra are shown in Figure 39, Figure 40 and Figure 41.

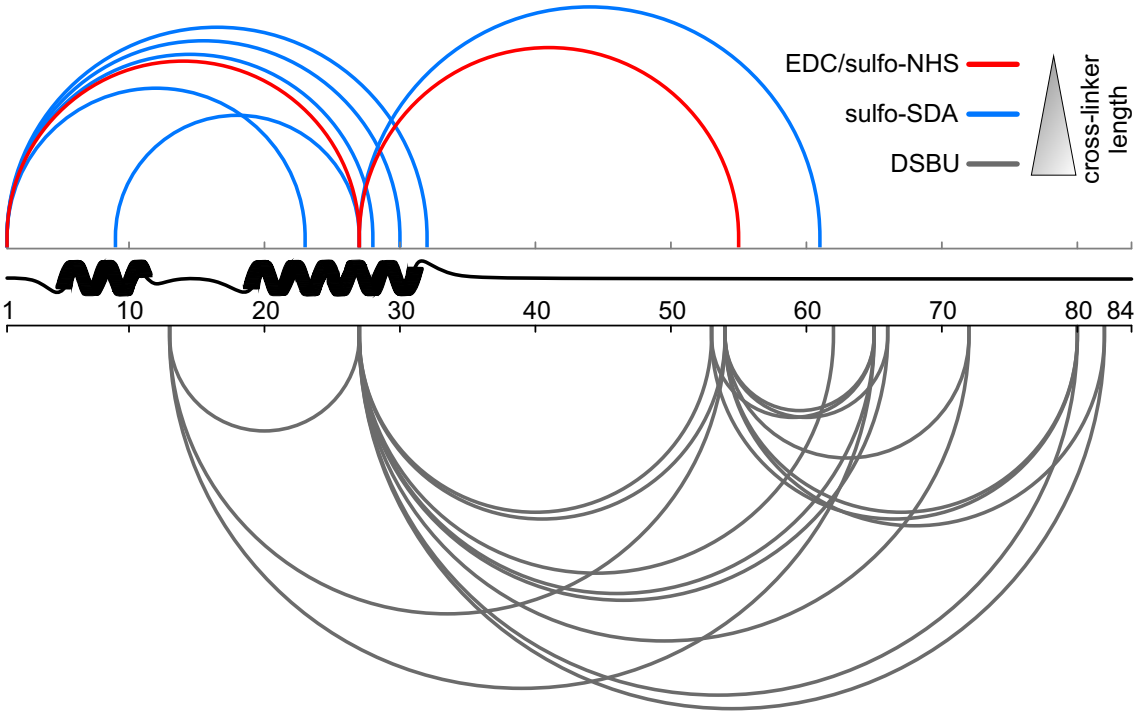


Figure 21: **Intra-chain cross-linking of PTH_{1-84}.** The intra-chain cross-linking pattern across the PTH_{1-84} peptide is displayed. Three cross-linkers with different linker-lengths were used: EDC/sulfo-NHS (zero-length, red), sulfo-SDA (short length, blue) and DSBU (medium length, gray). All mapped cross-links were obtained in three independent measurements.

Identifying similar N- to C-terminal cross-links for different linker lengths highlights the inherent flexibility and dynamic nature of the disordered C-terminal region. Importantly, and in line with the previously shown MD simulations, the identified N- to C-terminal cross-links confirm that PTH_{1-84} can sample through relatively compact states including those that bring residues in very close proximity potentially resulting in stabilizing attractive interactions. Similar self-interacting

states have been observed before in polyampholytic pepides which collapse under low-salt conditions [203]. The existence of such stabilizing interactions could explain why self-interacting conformational states are prevalent in the MD simulations. Since the reaction kinetics of DSBU and EDC/sulfo-NHS are slow compared to typical reconfiguration time scales of IDPs (~ 50-100 ns) [204], the identified cross-links indeed suggest that self-interacting states are long-lived conformations [205]. However, it has to be noted that the XL-MS-derived dataset shown here does not provide quantitative information about the conformational ensemble. Therefore, it cannot be inferred to what extent the self-interacting states are populated. Yet, the reasonable consistency across several MD simulations indicate that these states are energetically favored.

In summary, the data suggest that the disordered C-terminus of PTH$_{1-84}$ favorably interacts with the second α-helix in the N-terminus. This raises the question whether these N- to C-terminal contacts are physiologically important.

The C-terminus affects fibril formation of PTH$_{1-84}$

Interestingly, the above defined contact sites within the second α-helix are positioned within - or close to - the fibrillar core of PTH$_{1-84}$ amyloid fibrils [113]. Thus, it was hypothesized that these potential interactions might have implications on the behavior of fibril formation. To test this hypothesis, the fibril formation of several PTH constructs was monitored via ThT fluoresecence experiments (see Figure 22, performed by Twinkle Bhatia). In this ThT fluorescence experiment, the samples containing PTH$_{1-84}$ and PTH$_{1-34}$ gave rise to a sigmoidal increase in fluorescence intensity after a specific lag time has been passed indicating the formation of fibrils. The lag time varied considerably. PTH$_{1-84}$ exhibited a lag time of 204 $\pm$ 13 h, while the lag time for PTH$_{1-34}$ was shorter (ca. 135 h). This suggests that the C-terminal fragment in PTH$_{1-84}$ acts in a retarding capacity regarding the onset of fibril formation. This retardation may be caused by an IDR-mediated surface occlusion effect hindering efficient nucleation which has been found to impede fibril formation before [206]. Interestingly, while on its own, PTH$_{35-84}$ did not cause

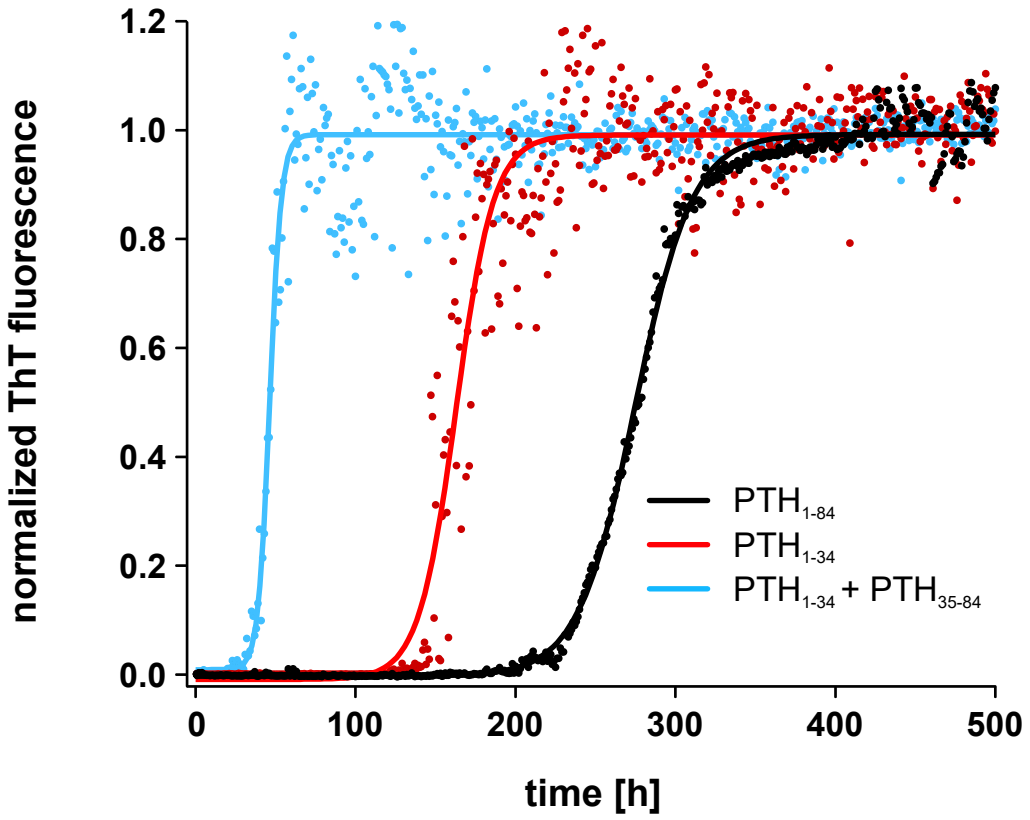


Figure 22: **Kinetics of fibril formation of various PTH variants.** Fibril formation of PTH_{1-84} (black), PTH_{1-34} (red), PTH_{1-34} + PTH_{35-84} (1:1 molar ratio, blue) and PTH_{35-84} (inset, gray) was monitored using ThT fluorescence. Proteins were concentrated to 150 μM in 50 mM NaP_i buffer, pH 7.4, with 150 mM NaCl. Except for the non-fibrillating PTH_{35-84}, data are shown normalized to the fluorescence intensity.

any increase in ThT fluorescence (see appendix Figure 42), the equimolar addition to PTH_{1-34} caused a considerable decrease in the lag time of PTH_{1-34} (35 $\pm$ 4 h). This is surprising considering the above mentioned fibrillation-retarding effect of the C-terminal fragment in PTH_{1-84} and hence suggests a more complex interplay between the disordered C-terminal region and the fibril formation of PTH_{1-84}.

In summary, the presented data strongly suggest that the C-terminal region of PTH_{1-84} affects the fibril formation of the peptide hormone. This observation matches previous reports in which the fibrillation of peptides was found to be affected by their disordered regions [207, 208]. A high sequence conservation within the C-terminal region indicates that functional aspects including a potential modulation of fibril formation are sequence-specific. However, additional experiments need to be performed to challenge the hypothesis of a causality between

specific N- to C-terminal contacts and fibrillation properties of PTH_{1-84}. Moreover, it has to be noted that the present dataset is preliminary and needs to be tested for reproducibility.

3.3 Effects of multisite phosphorylation on the condensation behavior of SRSF1

Serine/arginine-rich splicing factors (SRSFs or SR proteins) are RNA-binding proteins (RBPs) known to be regulators of constitutive and alternative splicing [117]. SR proteins and other splicing components typically locate to particular membraneless organelles, the nuclear speckles (NS) [120, 126]. NS are protein/RNA-condensates in the nucleus that have been proposed to act as reservoirs for RNA processing factors such as splicing factors [47]. However, there is evidence that splicing events do not occur within nuclear speckles but rather at the nucleoplasmic-speckle interface. Following this concept, splicing factors need to be released from the core of NS to exert their function. Post-translational modifications of splicing factors may regulate these dynamic translocations. Indeed, multisite phosphorylation of SR proteins have been shown to affect subcelluar localization and splicing activity [119]. Yet, a detailed description of the impact of these modifications on the condensation behavior is still elusive. In this project, the relationship of multisite phosphorylation and the condensation behavior of the prototypical SR protein, SRSF1, should be investigated.

The LCD of SRSF1 is necessary and sufficient for nuclear speckle targeting

SR proteins incorporate into NS and share a similar architecture consisting of at least one N-terminal RRM and a C-terminal LCD enriched in Arg/Ser residues often referred to as an RS domain [121]. It has previously been shown that LCDs with various amino acid compositions drive condensation and might thus play an important role in the formation and maintenance of certain biomolecular condensates [49, 55, 131]. Since Arg/Ser-rich proteins are prevalent in NS [55],

it is conceivable that a shared ”sequence grammar” within the LCDs provides cohesion in these particular condensates [49]. Therefore, it was hypothesized that isolated Arg/Ser-rich LCDs of NS targeting proteins might be sufficient for proper subcellular localization. To test this, a set of SBP-FLAG-RFP(SFR)-tagged SRSF1 protein constructs was designed comprising a full-length variant (SRSF1 FL) and a deletion mutant only containing the RS domain (SRSF1 LCD) (see Figure 23A). The successful overexpression of both SRSF1 constructs and the isolated SFR-tag was demonstrated via western blot analysis of the cell lysates (see Figure 23B). The subcellular localization of SRSF1 FL and SRSF1 LCD overexpressed in the human cell line OVCAR-3 was monitored (see Figure 23C).

Here, SRSF1 FL was mainly nuclear and colocalized with NS. This could be further validated by a color intensity analysis across a speckle giving rise to superimposing intensity profiles of both, the green (SC35) and the red (RFP constructs) channels (see Figure 23D). This is in agreement with previously conducted localization studies that indicate high enrichment of SRSF1 in NS [123, 126]. Moreover, this suggests that the N-terminal tag which includes the RFP does not interfere with the subcellular localization of the protein. The uniform distribution of the isolated tag across the cytosol and nucleoplasm further substantiates that the tag does not significantly alter the subcellular localization of respective proteins.

Like SRSF1 FL, SRSF1 LCD was mainly localized in the nucleus implying that the active translocation into the nucleus is not impaired by deletion of the RRM domains. This result was expected since the RS domain itself provides a nuclear localization signal (NLS) and is recognized by a specific importin β receptor, the transportin-SR [134]. As transportin-SR specifically recognizes multisite phosphorylated RS domains [210], the nuclear localization of SRSF1 LCD further suggests that the phosphorylation state resembles at least partially the physiological state of the wildtype protein. Inside the nucleus, SRSF1 LCD was localized throughout the nucleoplasm with a clear enrichment in NS. Together

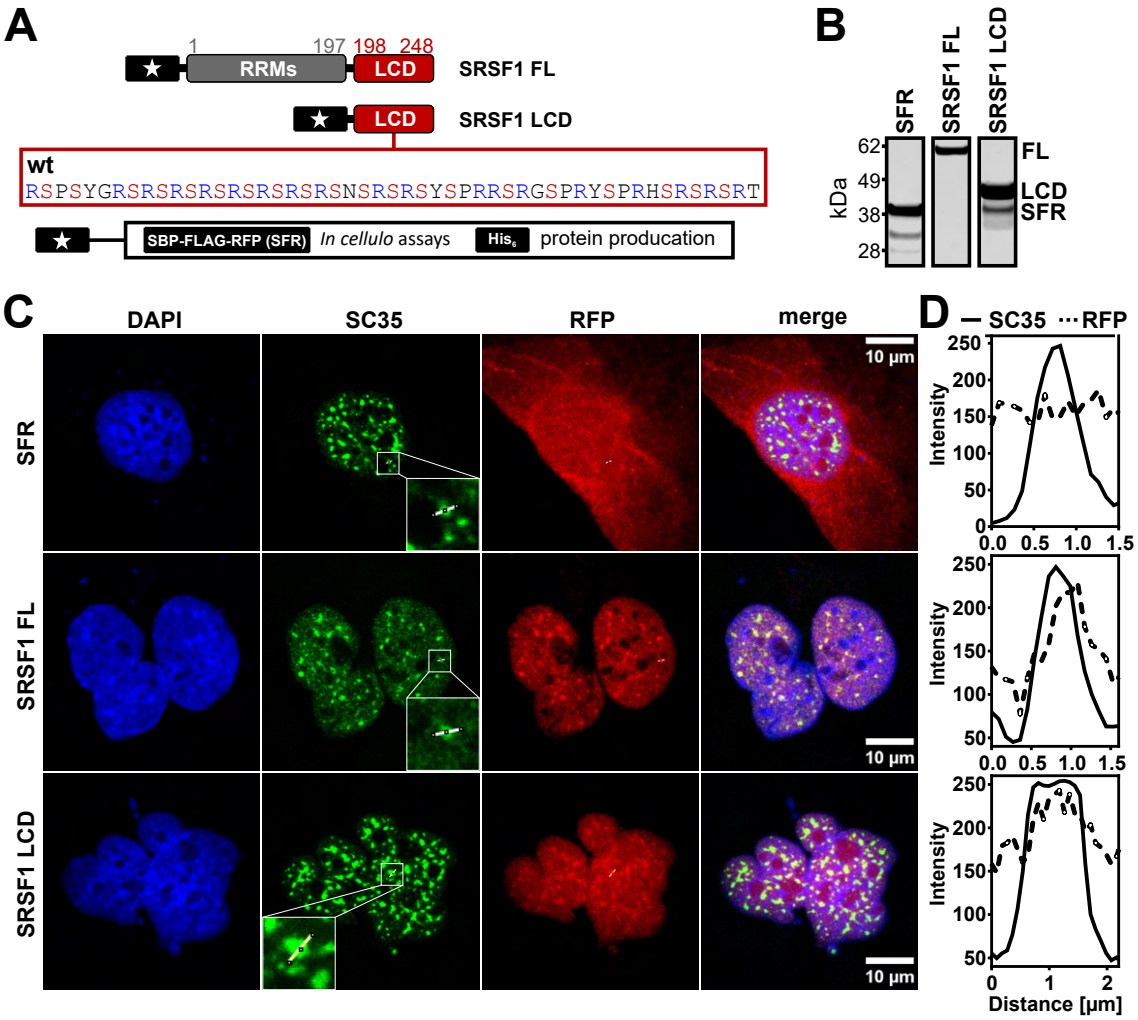


Figure 23: **Subnuclear localization of SRSF1 constructs.** A: Construct design of human SRSF1 FL (AA 1-248, Uniprot-ID: Q07955). In the design of the isolated LCD, in order to ensure proper interaction with SRPK1 the RS domain of SRSF1 was extended by a short fragment of the adjacent RRM to yield SRSF1 LCD (AA 183-248) [209]. For expression in mammalian systems, a multipurpose tag (SFR) consisting of a strepdavidin-binding protein (SBP), FLAG sequence and RFP was fused N-terminally. For protein purification from bacterial expression, a hexahistidine-tag was attached N-terminally. B: Western blot analysis of over-expression of protein constructs in HEK293t cells. Proteins were detected with anti-FLAG antibody. C: Confocal images of the OVCAR-3 cell line expressing indicated RFP-fused constructs. Nuclei were stained with DAPI and nuclear speckles were stained with anti-SC35. Scale bar: 10 μm. D: Intensity profiles across nuclear bodies (indicated by white lines in (C), see insets) confirm enrichment of SRSF1 constructs in nuclear speckles.

with a previously published report demonstrating that deletion of RS dipeptide repeats within the LCD abolished proper subnuclear localization [211], the here presented results indicate that the LCD of SRSF1 is sufficient and necessary for targeting NS. Therefore, it is conceivable that the ”sequence grammar” of this and

similar Arg/Ser-rich domains provides cohesion in NS and might be an important driver in the formation of these biomolecular condensates.

Effects of mixed-charge mutations in SRSF1 LCD on the subcellular localization

Multisite phosphorylation is a frequently described PTM regarding the NS proteome [117, 212] in general as well as SRSF1 specifically [213]. In a number of splicing factors, these PTMs have been shown to regulate subcellular localization and splicing activity [135, 136, 139, 214]. A multitude of specific kinases perform phosphorylations on different NS components. In case of SRSF1, two kinases have been found to phosphorylate multiple Ser in the LCD. The SR protein kinase 1 (SRPK1) performs ~ 10 Ser phosphorylations in the extended Arg/Ser-dipeptide region via a processive mechanism (see Figure 9) [133, 215]. Phosphorylation events associated to SRPK1 have been reported to occur after the translation of SRSF1 in the cytosol yielding a so-called hypophosphorylated state [119]. CDC like kinase 1 (CLK1) is a primarily nuclear kinase [216] that can per se perform phosphorylations at every Ser site, including those that preceede proline residues (see Figure 9) [133, 136]. Hence, further CLK1-mediated phosphorylations of the hypophosphorylated state yield a so-called hyperphosphorylated state. In general, (de)phosphorylations of SRSF1 appear to be highly regulated and to affect splicing activity [137, 217].

Differential multisite phosphorylations in SRSF1 have been proposed to impact the subcellular localization of the protein as well [119, 132, 137]. Yet, a detailed relationship between the two aspects remains to be elucidated. Here, for subcellular localization studies, several phospho-mimicking mutants (Ser to Asp) were generated reflecting the hypo- (= SRSF1 LCD$_{Hypo-D}$) and hyperphosphorylated (= SRSF1 LCD$_{Hyper-D}$) states of SRSF1 (see Figure 24A). To isolate the functionality of the RS domain in that regard, the focus was set on SRSF1 LCD constructs. Again, the correct expression of the constructs was validated via western blot

analysis and for microscopic imaging, the constructs were overexpressed in the OVCAR-3 cell line (see Figure 24B).

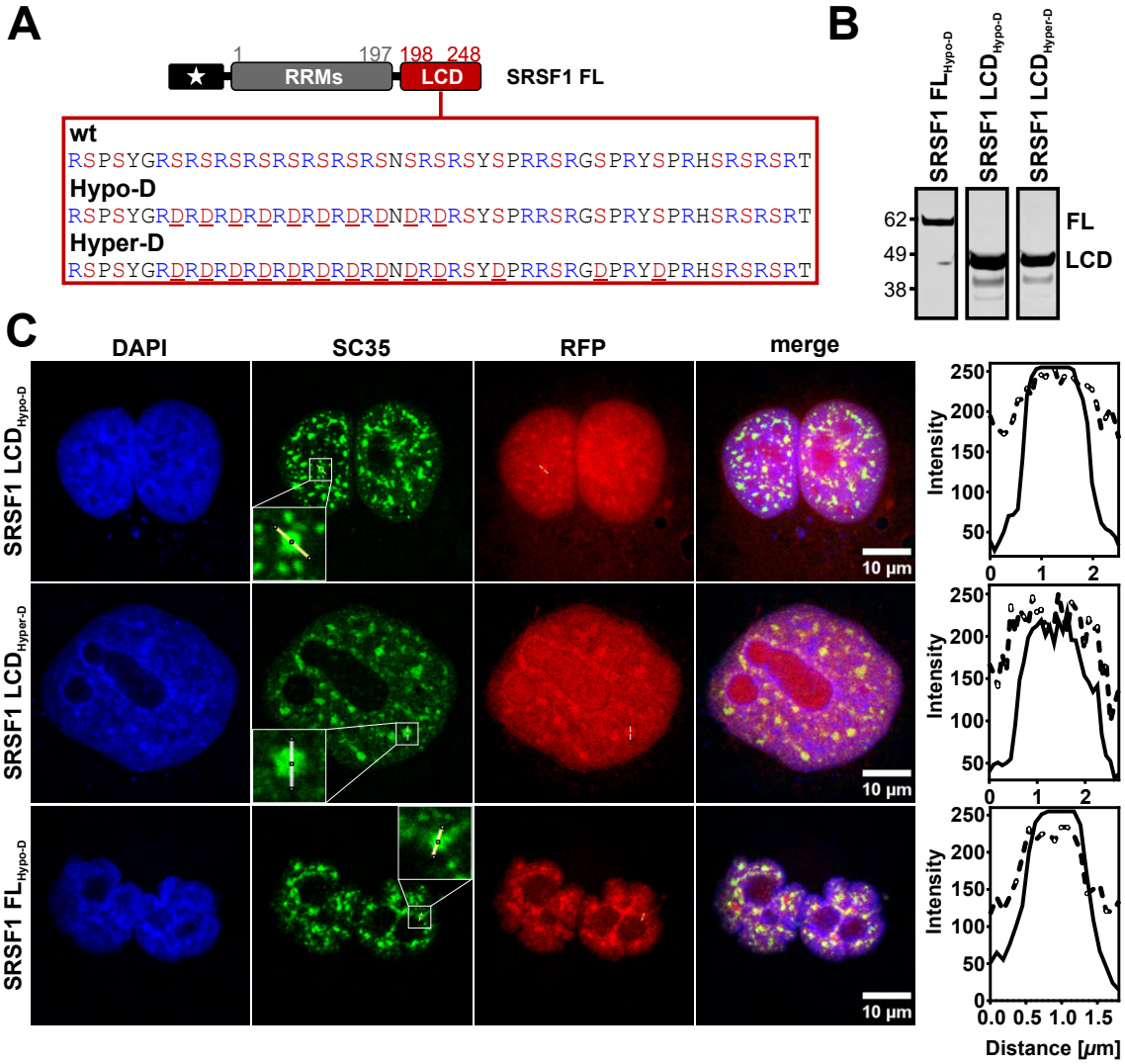


Figure 24: **Subnuclear localization of phospho-mimicking SRSF1 variants.** A: Construct design of phospho-mimicking mutants of SRSF1. The mutated LCD sequence (red box) is shown to specify multiple sites of S/D mutations (sites are underlined) mimicking the hypo- and hyperphosphorylated sates (denoted as Hypo-D and Hyper-D, respectively). B: Western blot analysis of overexpression of protein constructs in HEK293t cells. Proteins were detected with anti-FLAG antibody. C: Confocal images of the OVCAR-3 cell line expressing indicated RFP-fused constructs. Nuclei were stained with DAPI and nuclear speckles were stained with anti-SC35. Scale bar: 10 μm. D: Intensity profiles across nuclear bodies (indicated by white lines in (C)) confirm enrichment of SRSF1 constructs in nuclear speckles.

Similarly to the wildtype SRSF1 LCD both SRSF1 LCD$_{Hypo-D}$ and SRSF1 LCD$_{Hyper-D}$ were predominantly nuclear (see Figure 24C), indicating that the

cytoplasmic-nucleoplasmic transport was not impeded by the mutations. It is therefore likely that transportin-RS successfully recognizes these mutants to mediate their transport. Since the transportin-SR-mediated transport of SRSF1 is phosphorylation dependent, the predominantly nuclear localization of the SRSF1 LCD mutants further suggests that their physicochemical properties resemble those of the phosphorylation states. Hence, these mutants can be used as a proxy to study phosphorylation dependent processes - an approach frequently used to address similar scientific questions [211, 218].

Inside the nucleus, both SRSF1 LCD mutants shared a similar distribution with a clear enrichment in NS akin to the wildtype SRSF1 LCD localization pattern (see Figure 24C,D), Figure 23). To interpret this observation, multiple factors need to be taken into account: (i) there is evidence that NS promote gene expression by providing the necessary components for co-transcriptional pre-mRNA splicing and ensure other RNA metabolic steps to occur [47, 131, 219]. Even though active transcription was not found to occur inside NS, they are frequently found in close proximity to active transcription sites possibly because these processes benefit from the high local concentration of processing factors in the condensate [128, 129]. Accordingly, the RNA processing factors that reside in NS, including SRSF1, may not be required to diffuse a long distance from the NS to fulfill their functions. In confocal microscopy imaging, any differences in subcellular localization as a function of differential phosphorylation might therefore be below the detection limit. The here presented data supports this scenario as no strong changes in subcellular localization could be detected for the different phospho-mimicking mutants of SRSF1 LCD.

(ii) To detect potential phospho-state dependent molecular rearrangements in NS or in their close vicinity, super resolution imaging techniques may be required. Recent super resolution images indeed resolved a layered structure of NS with a mostly proteinaceous core and a surrounding layer composed of mainly RNA that is proposed to be active in co-transcriptional RNA processing [47, 129]. According

to this model, Arg/Ser-rich processing factors such as SRSF1 are recruited to the proteinaceous core of NS by the scaffold proteins SRRM2 and SON and get dynamically released to perform their functions in RNA processing, potentially mediated by differential phosphorylation. Indeed, the kinase CLK1 was shown to incorporate into NS and was proposed to mobilized SRSF1 through hyperphosphorylation [220]. Although it is conceivable that these PTMs are relevant for dynamic rearrangements in NS, this relationship is not clearly understood.

(iii) Cells overexpressing SRSF1 LCD constructs were often found to exhibit an abnormal phenotype (e.g. increased proliferation) suggesting that overexpression of the proteins caused interference with the normal physiology. The abnormal physiology may affect the subnuclear arrangement of biological components and therefore makes clear interpretations difficult.

Overall, these data suggest that the phosphorylation-state dependent phase behavior of SRSF1 LCD cannot be resolved in the cellular environment with the setup used. This problem might be overcome by addressing the questions in a simplified *in vitro* system. In this work, protein purification of several SRSF1 variants was attempted to investigate their condensation behavior in more detail *in vitro*.

Purification of SRSF1 variants

To investigate the phase behavior of SRSF1 variants in more detail *in vitro*, purified protein was needed. An initial western blot analysis was performed to test the heterologous expression in bacteria and protein solubility after lysis of several SRSF1 variants (see Figure 25A). In general, all SRSF1 variants were expressed, yet protein solubility was highly dependent on the nature of the variant. For example, heterologously expressed SRSF1 FL was mainly present in the pellet (P) fraction after lysis which agrees with previous publications describing the formation of SRSF1 containing inclusion bodies in *E. coli* [122, 140, 221]. Thus, protein purification of SRSF1 FL needs to be performed under denaturing conditions. Based on literature, however, refolding of SRSF1 FL seems problematic with very little yield of successfully folded and soluble protein [140, 221].

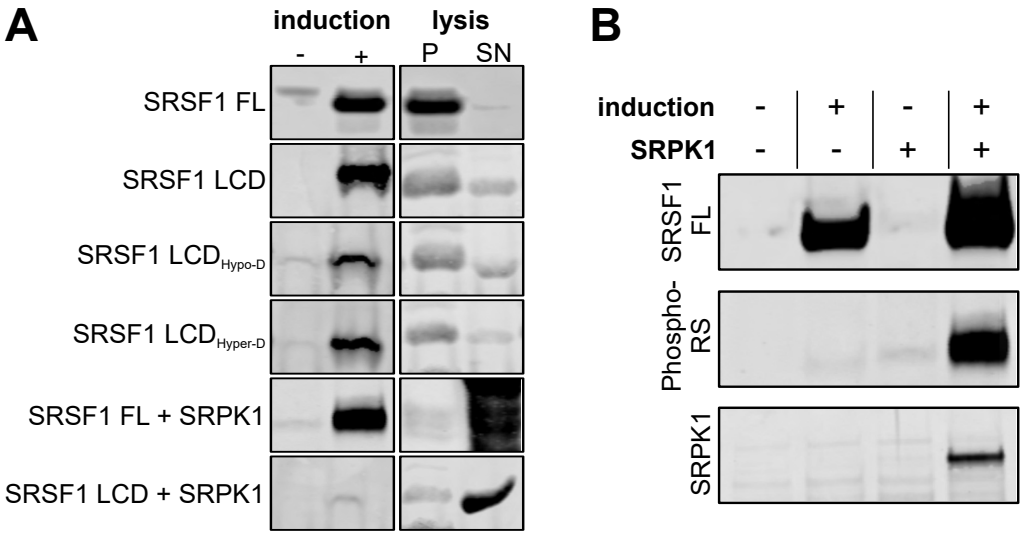


Figure 25: **Expression analysis and solubility test of SRSF1 variants.** A: The indicated SRSF1 variants were successfully expressed after induction with IPTG (+). After cell lysis, SRSF1 FL was almost exclusively found in the pellet (P) fraction. SRSF1 LCD variants were found in both, the pellet and supernatant (SN) fractions. Co-expression with the kinase SRPK1 resulted in SRSF1 variants to be mainly in the SN fraction. Proteins were detected using the His-probe antibody. SRSF1 LCD + SRPK1 was detected using SRSF1 C-terminal antibody (induction) and phospho-RS antibody (lysis). SRSF1 FL + SRPK1 lysis samples were stained with coomassie. B: Co-expression of SRSF1 FL with SRPK1 results in phosphorylations within the RS-rich LCD of SRSF1. SRSF1 FL was detected using SRSF1 N-terminal antibody. Phosphorylated SRSF1 FL and SRPK1 were detected using phospho-RS as well as GST antibodies, respectively.

Therefore, a previously published protocol describing the heterologous co-expression of SRSF1 FL and its specific kinase, SRPK1, should be followed [140]. According to the publication, SRPK1 expressed in *E. coli* performs multisite phosphorylation in the RS domain of SRSF1 FL to a degree that resembles the native state of SRSF1 FL in human cells. Importantly, they found the phosphorylated SRSF1 FL fraction to be soluble under native conditions, enabling the use of a non-denaturing purification procedure.

Initially, for heterologous co-expression, BL21DE3 *E. coli* cells were transformed with pET41a encoding SRSF1 FL (with an N-terminal His_6-tag) and pGEX-6P-1 encoding SRPK1 (with an N-terminal GST-tag) (see section 2.6). However, co-transformed cells were not growing in LB medium which may be attributed to potential toxicity of basaly expressed proteins [222]. To achieve tighter expres-

sion control, the cell system was changed to a T7 Express lysY/Iq (New England Biolabs, Frankfurt) *E. coli* strain. Indeed, this bacterial strain was growing in LB medium after co-transformation.

To check if expressed SRPK1 phosphorylates SRSF1 FL in the expression host, a western blot analysis was performed (see Figure 25B). According to the western blot, SRSF1 FL was expressed at high levels independent of SRPK1 expression. Yet, only in presence of SRPK1, phosphorylated SRSF1 FL was detected indicating that the previously published procedure could be successfully reproduced.

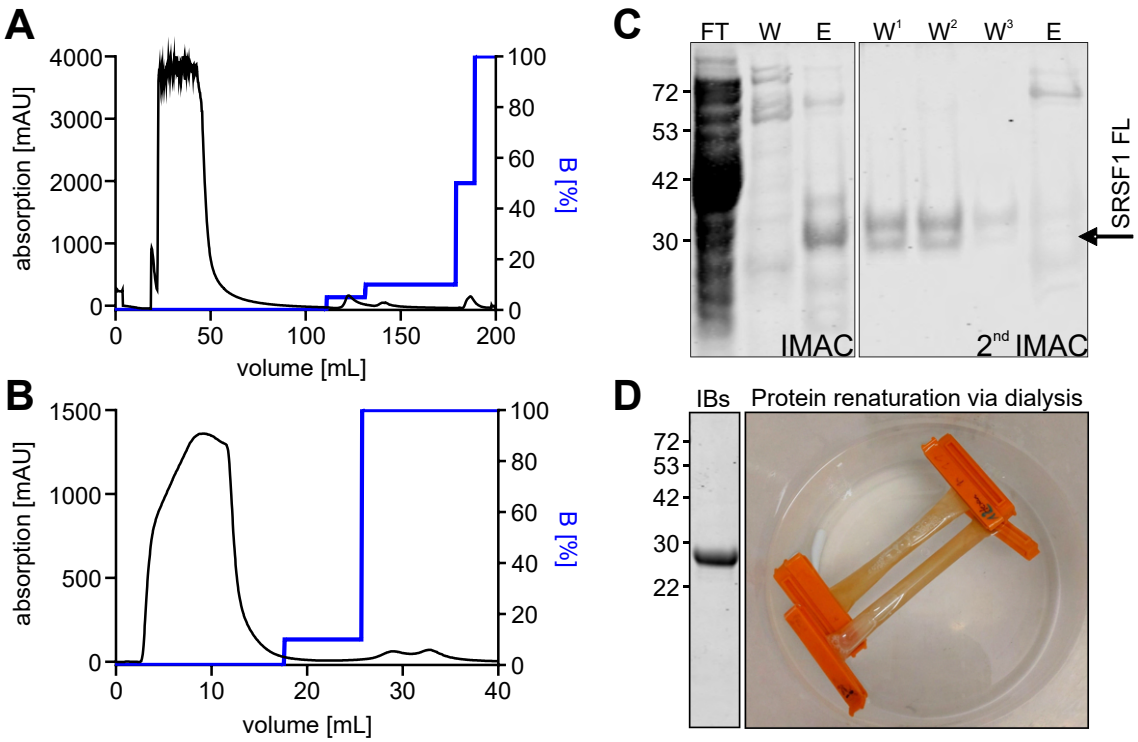


Figure 26: **Protein purification of SRSF1 FL.** SRSF1 FL obtained from co-expression with SRPK1 was purified using an IMAC-based approach (A-C). A: Chromatogram of an IMAC run after lysis. The following fractions were collected and analyzed via SDS-PAGE (C): flowthrough (FT) - 25-30 mL; wash (W) - 120-125 mL; elution (E) - 183-188 mL. B: Chromatogram of a second IMAC run after proteolytic removal of the hexahistidine-tag. For SDS-PAGE analysis (C), wash fractions (W^1-W^3) were collected (0-15 mL). Fraction E was collected (35-40 mL). D: Purification of SRSF1 FL obtained from single-expression was attempted via denaturing inclusion body preparation. Pure SRSF1 FL was obtained under denaturing conditions. Protein renaturation via dialysis against refolding buffer resulted in protein precipitation (see image).

After co-expression, cells were lysed to continue with SRSF1 FL purification.

A major fraction of the expressed SRSF1 FL was found in the supernatant (SN) after centrifugation of the cell lysate suggesting that the protein is soluble. Taking advantage of the N-terminal His_6-tag in SRSF1 FL, an IMAC-based purification approach was followed, which removed contaminants effectively (see Figure 26A,C). To remove the His_6-tag, eluted SRSF1 FL was subjected to proteolysis via H3C protease. In a second IMAC-run, the free His_6-tag was bound to the Ni^{2+}-NTA column and SRSF1 FL was collected in the flow through (see Figure 26B). This procedure seemed to produce relatively pure SRSF1 FL protein according to the SDS-PAGE analysis (see Figure 26C). However, after the second IMAC-run, two bands close to the expected apparent molecular weight of SRSF1 FL were observed. The exact nature of these observed bands has yet to be investigated. Moreover, using this co-expression approach, the achievable homogeneity of the SRSF1 FL protein was a major concern. Multisite phosphorylation of SRSF1 FL in the expression host likely generates a heterogeneous ensemble of phosphorylated states. Further, batch-to-batch variations of different purification attempts would hamper comparability of downstream analyses. To study the phosphorylation dependent condensation behavior of SRSF1 FL, however, the initial protein sample under investigation should be well defined and ideally homogeneous. Therefore, rather than optimizing the above mentioned purification procedure, SRSF1 FL production from inclusion bodies was attempted.

Single-expression of SRSF1 FL in BL21DE3 *E. coli* cells resulted in the formation of inclusion bodies as previously described [122]. SRSF1 FL could be isolated from inclusion bodies effectively using the protocol described in section 2.6. According to an SDS-PAGE analysis, the preparation of SRSF1 FL from inclusion bodies via 8 M urea yielded high amounts of SRSF1 FL with no apparent protein contaminations (see Figure 26D). Renaturation of SRSF1 FL has been reported to be inefficient with only a small fraction of successfully folded and soluble protein [221]. This difficulty might be linked to the physicochemical properties of the Arg/Ser-rich LCD of SRSF1 FL. As the LCD of SRSF1 FL is enriched in Arg,

the isoelectric point of SRSF1 FL is exceptionally high at 10.26 (calculated using Prot Pi Protein Tool, https://www.protpi.ch/Calculator/ProteinTool). Hence, anionic molecules like nucleic acids tend to be attracted via electrostatic forces. Indeed, a significant amount of nucleic acids accompanied the preparation of inclusion bodies (at first apparent in the gel-like cell lysate prior to DNase treatment and later observed in the significant co-sedimentation together with inclusion bodies resulting in a transparent rim surrounding the inclusion bodies). The presence of unspecific nucleic acids may have hindered efficient renaturation of SRSF1 FL in the past. Therefore, the here used protocol includes measures to minimize nucleic acid contaminations. After solubilization of SRSF1 FL containing inclusion bodies in 8 M urea, the solution is acidified to reduce the solubility of nucleic acids which could be removed by centrifugation. According to the UV ratio 260 nm/280 nm = 0.7, nucleic acids were excluded effectively. To test, whether minimizing the nucleic acid content resolved the issue of SRSF1 FL renaturation, refolding was attempted via dialysis against non-denaturing buffer. However, a large fraction of SRSF1 FL precipitated and only a small fraction was soluble (see Figure 26D). An SDS-PAGE analysis of the soluble fraction revealed the presence of both, full length SRSF1 and significant amount of degradation products (data not shown). These results suggest that both, renaturation as well as subsequent handling of SRSF1 needed to be optimized in future purification attempts to yield homogeneous protein samples. However, in light of these challenges and the typically associated time requirements for overcoming them, it was decided to focus on the production of SRSF1 LCD variants. Here, variants that mimic the hypophosphorylated and hyperphosphorylated state were generated. As these variants exhibit physicochemical properties close to the multisite phosphorylated SRSF1, it is conceivable that expression in bacteria yields soluble protein in analogy to the co-expression approach mentioned above. Moreover, since SRSF1 LCD variants lack the ordered RRMs, potentially required renaturation of the proteins may be less complex.

SRSF1 LCD variants were partly soluble but a considerable fraction appeared to be insoluble, possibly forming inclusion bodies (see Figure 25). In the following, the purification of SRSF1 LCD$_{Hyper-D}$ is described exemplarily (see Figure 27). To not lose any of the protein fractions, cell lysis was performed under denaturing conditions (8 M urea containing buffer). The His$_6$-tagged proteins were purified using IMAC (see Figure 27A). However, high molecular weight contaminants were observed in the SDS-PAGE analysis (see Figure 27C). To remove these contaminants, size eclusion chromatography (SEC) was performed (see Figure 27B). Based on SDS-PAGE analysis, contaminants were successfully removed, resulting in relatively pure SRSF1 LCD$_{Hyper-D}$ (see Figure 27D).

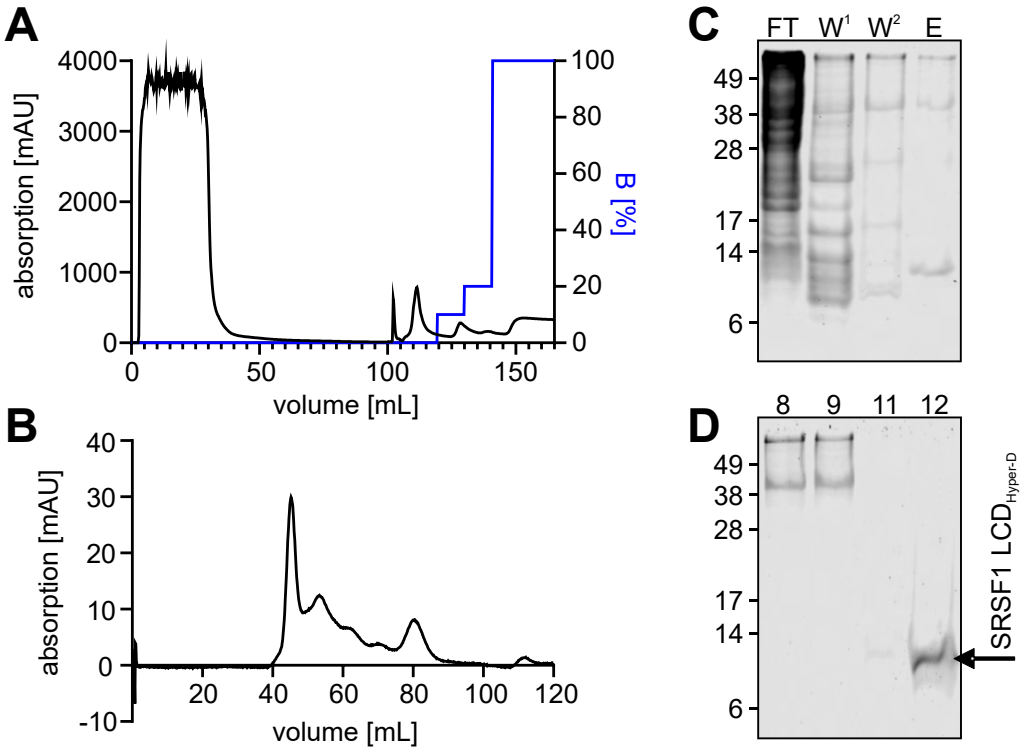


Figure 27: **Protein purification of SRSF1 LCD$_{Hyper-D}$.** A: SRSF1 LCD$_{Hyper-D}$ was purified using an IMAC-based approach. The following fractions were collected and analyzed via SDS-PAGE (C): flowthrough (FT) - 5-20 mL; first wash (W^1) - 100-115 mL; second wash (W^2) - 125-140 mL; elution (E) - 147-157 mL. B: Chromatogram of the subsequent SEC run . For SDS-PAGE analysis (D), the following peak fractions according to 280 nm absorption were collected: fraction 8 - 42-49 mL; fraction 9 - 42-49 mL; fraction 11 - 49-59 mL; fraction 12 - 79-85 mL. Putative SRSF1 LCD$_{Hyper-D}$ was eluted with fraction 12.

However, investigations of protein phase behavior *in vitro* requires large quantities of pure protein. Although, the so far established protocol for protein purification of SRSF1 LCD variants resulted in relatively pure protein batches, the yield is still to be optimized. Due to time limitations, synthetically produced SRSF1 LCD variants were purchased for further characterizations *in vitro*.

SRSF1 LCD forms liquid condensates in presence of polyanions *in vitro*

In cellulo, SRSF1 LCD is incorporated into NS suggesting a major role of the domain in condensate formation. Next, the condensation behavior of SRSF1 LCD should be investigated via *in vitro* experiments with pure peptide in solution. First, a solution of 40 μM SRSF1 LCD was prepared. The solution was translucent and images recorded with a phase contrast microscope showed a homogeneous, mono-phasic solution indicating that SRSF1 LCD does not spontaneously form condensates under the present conditions (see Figure 28A).

The pronounced sequence bias of the peptide is likely a major contributor to the observed lack of condensates. With Arg residues making up almost 40% of the AA sequence, SRSF1 LCD has an extremely high pI value of 13 and a net charge of +18 charges at pH 7.4 (calculated using the Prot pi Protein Tool, https://www.protpi.ch/Calculator/ProteinTool). The high density of positive charges in the peptide may render self-organization into higher order complexes energetically unfavorable through mutual repulsion. Furthermore, a potentially high solubility limit of the peptide may hinder phase separation. Given the polycationic nature of the peptide, it was hypothesized that charge neutralization via a polyanionic macromolecule might induce condensate formation via complex coacervation. Indeed, when poly-U RNA (20mer) was added in a charge-neutralizing amount, the peptide solution turned turbid and condensates were found in the form of droplets (see Figure 28A). To clarify, whether condensate formation was a result of specific interactions with RNA, heparin (~ 20 kDa) as another polyanionic macromolecule was tested. Similar to the effect of RNA addition, a charge-neutralizing amount of heparin induced condensate formation indicating that the process is a result of

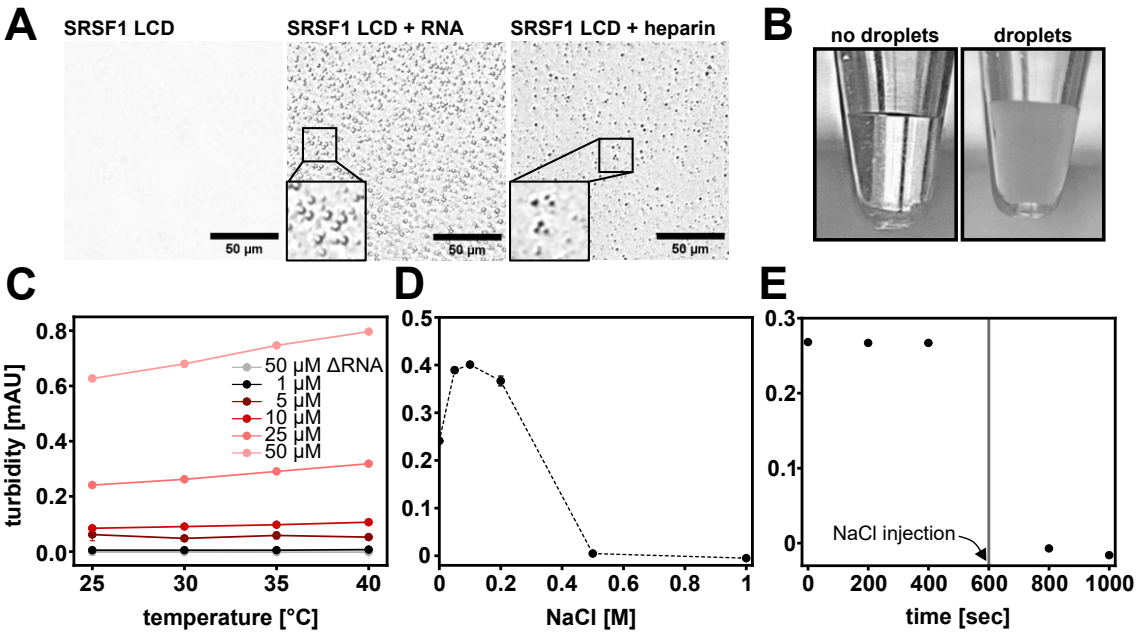


Figure 28: **SRSF1 LCD forms liquid condensates in presence of polyanions *in vitro*.** A: Phase contrast microscope images reveal droplet formation of SRSF1 LCD (40 μM) in presence of charge-neutralizing amounts of poly-U (20mer) RNA and heparin (~ 20 kDa). B: Condensate formation of SRSF1 LCD in presence of polyanions results in increased turbidity compared to a well-mixed sample. Turbidity assays (λ = 450 nm) (C-E) were performed to characterize the condensation behavior of SRSF1 LCD depending on concentration, temperature (C) and ionic strength (D) (error bars indicate standard deviation with a sample size of n = 3). E: Formed droplets dissolve after injecting 5 μL of a 5 M NaCl solution (625 mM final NaCl concentration).

unspecific interactions with polyanions.

To understand the phase behavior in more detail, condensate formation of SRSF1 LCD should be further characterized. Since condensates scatter light, samples in which condensates have formed appear turbid (see Figure 28B). The apparent absorbance of a turbid sample can be used to quantify condensate formation using a photometric approach [41, 64]. A series of turbidity assays was performed where condensate formation was induced with charge-neutralizing amounts of poly-U RNA (20mer). First, turbidity was measured depending on the peptide concentration and temperature (see Figure 28C). With increasing peptide concentrations, the samples got more turbid starting at around 5 μM. Adjusting the temperature had a minor effect on the turbidity of the samples suggesting that complex coacervation of SRSF1 LCD:polyanion is not highly temperature sensitive in the tested range.

This is in line with previously published data on a similar Arg-rich polyelectrolyte system. Here, at low salt concentrations, a lower critical solution temperature (LCST) behavior was found [63] indicating an entropically driven phase separation. This entropic contribution was proposed to arise through a counterion and hydration water release which is a common phenomenon in electrostatic interactions [223]. Interestingly, in the low temperature regime (~ 0 - 20 °C), they found an initial disruption of condensates with increasing temperature implying an additional upper critical solution tempreature (UCST) behavior which is dominated by enthalpic contributions for demixing. This behavior was not observed in this work, possibly because of the limited temperature range (25 - 40 °C) that was tested. In the absence of RNA, no apparent absorbance was detected confirming the absence of droplets in the microscope image presented in Figure 28A.

As condensate formation of the polycationic SRSF1 LCD is induced by a polyanionic macromolecule, the interaction profile was speculated to be charge-dominated. To test this hypothesis, the turbidity was monitored in buffer with varying NaCl concentrations to adjust the ionic strength (see Figure 28D) [62]. With increasing ionic strength, the charge screening effect becomes more prominent and thereby gradually diminishes charge-charge interactions. In low ionic strength conditions up to 200 mM NaCl concentration, the sample was turbid indicating the presence of condensates. Increased NaCl concentrations (500 mM and 1000 mM) resulted in a translucent solution devoid of condensates. Therefore, it can be concluded that condensate formation of SRSF1 LCD is induced by polyanions through mainly electrostatic interactions. These interactions are likely mediated by the Arg residues in the peptide and the phosphate backbone of the RNA as has been proposed earlier for a similar protein:RNA system [62].

To test, whether condensate formation is reversible, a highly concentrated NaCl solution (5 M) was injected into a turbid sample (see Figure 28E). If condensates formed were highly dynamic and liquid-like structures, they should undergo mixing with the surrounding solution due to a disruption of charge-charge interactions

between SRSF1 LCD and the RNA. Indeed, the initially turbid sample turned translucent after NaCl injection suggesting that condensate formation is reversible. The relatively short time-scale on which the condensates dissolved suggests that they are highly dynamic structures and liquid-like in nature.

In conclusion, SRSF1 LCD undergoes complex coacervation with polyanions to yield dynamic and reversible condensates (see Figure 29). Complex coacervation was found to be dominated by electrostatic interactions.

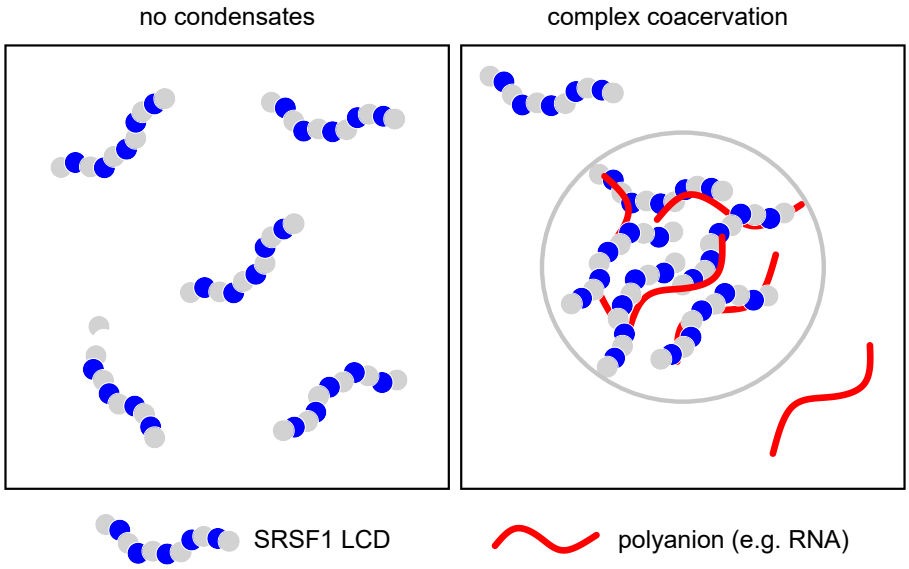


Figure 29: **Model of SRSF1 LCD condensation with polyanions.** Wildtype SRSF1 LCD does not form condensates through homotypic contacts (left box) but through heterotypic contacts (right box) with oppositely charged macromolecules such as RNA.

Phase behavior of mixed-charge variants of SRSF1 LCD

Like other SR proteins, SRSF1 is extensively phosphorylated in cells [119, 135, 224]. Multisite phosphorylations within SRSF1 LCD transforms the initially polycationic peptide into a polyampholytic one. Here, phospho-state mimicking mixed-charge variants of SRSF1 LCD should be studied via microscopy with a focus on whether changes in the peptide's charge properties affect its phase behavior. As described before (see section 3.3), SRSF1 LCD variants exhibited varying degrees of Ser/Asp mutations (see Figure 31A). Asp features a negatively charged residue that is physicochemically comparable to phospho-Ser and thus

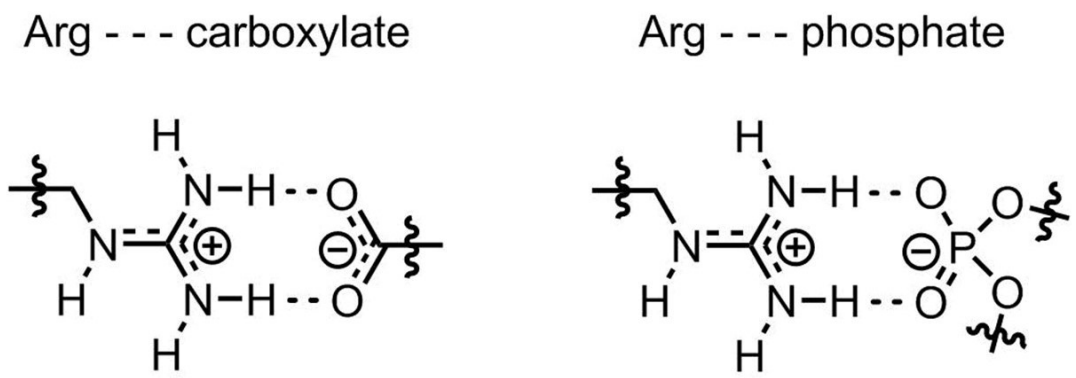


Figure 30: **Interactions of Arg residues with carboxylate and phosphate.** Both, the Arg-carboxalate (left) and Arg-phosphate (right) interactions are driven by charge-charge interactions and hydrogen bonding. Figure adapted from: www.chemsynbio.com/new/output/arginine.html.

may be used as a phospho-mimetic. Importantly, both phospho-Ser and Asp display a similar interaction with Arg residues forming charge-charge interactions and hydrogen bonding (see Figure 30). Therefore, potentially LLPS driving Arg-pospho-Ser interactions likely also exist in Ser/Asp mutated peptides [225].

In microscopy of solutions of various SRSF1 LCD variants, a complex phase behavior was observed with a clear sequence-dependence governing the process of demixing. In a 50 mM HEPES buffer (pH 7.4), SRSF1 LCD$_{Total-D}$ formed condensates in the shape of droplets, while the less extensively mutated variants stayed mixed with the solution. The condensation of SRSF1 LCD$_{Total-D}$ followed a UCST behavior with droplets only occurring at lower temperatures and dynamically dissolving with increasing temperatures. This implies a phase behavior dominated by enthalpic contributions which likely arise from inter-chain charge-charge interactions and hydrogen bonding between Arg and Asp residues [45].

The striking differences in the phase behavior among the variants may be explained with the net charge characteristics. SRSF1 LCD variants that did not form condensates exhibit a high positive net charge at pH 7.4, while SRSF1 LCD$_{Total-D}$ is close to overall charge neutrality (see Figure 32). Therefore, like-charge repulsion should be minimized in SRSF1 LCD$_{Total-D}$ possibly causing a higher propensity to undergo phase separation. Similarly, the polyampholytic protein DDX4 is close

to charge neutrality overall which possibly contributes to its propensity to undergo phase separation [45].

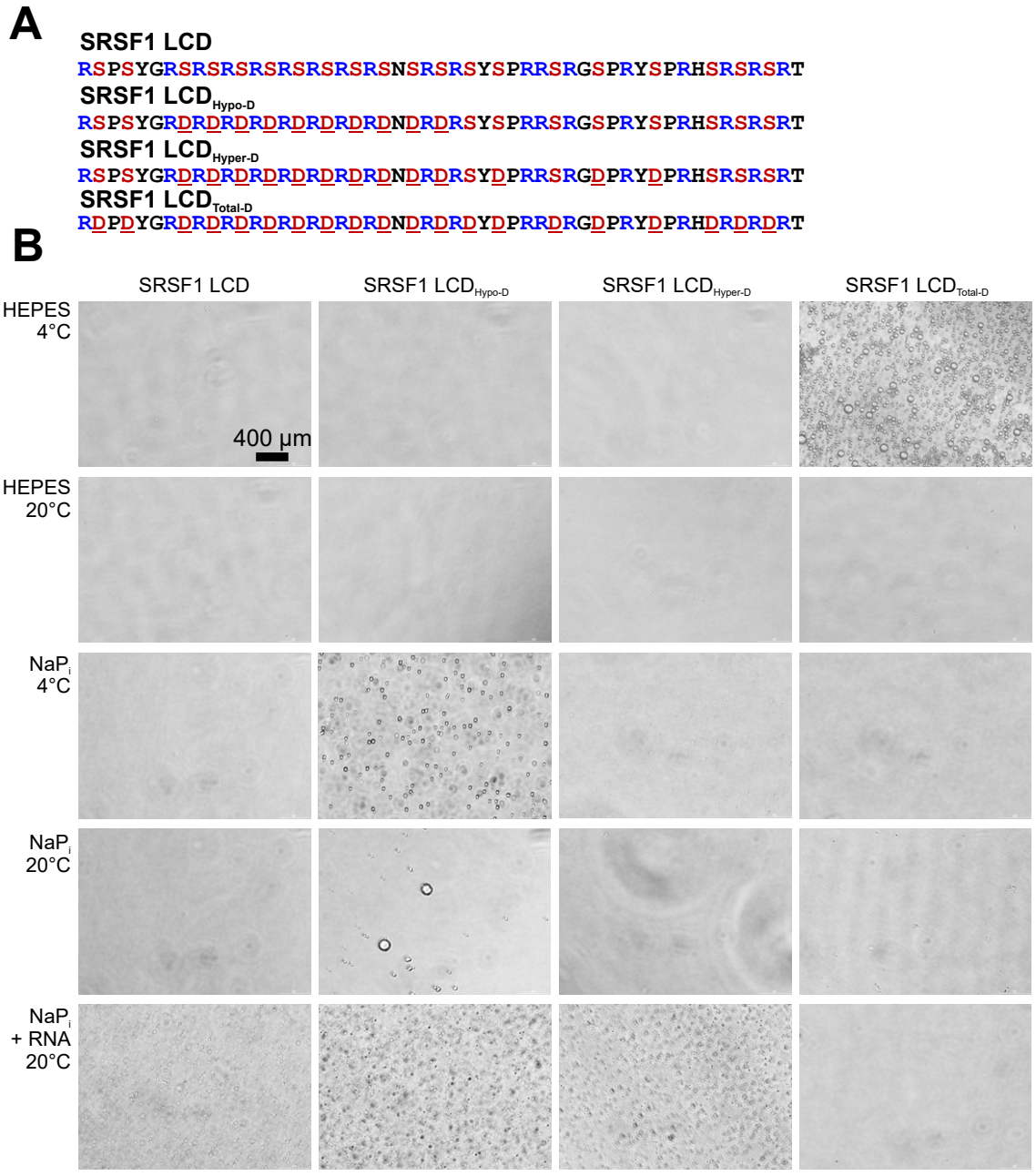


Figure 31: **Phospo-mimicking mutations in SRSF1 LCD affect its phase behavior *in vitro*.** A: AA sequences of SRSF1 LCD variants used in *in vitro* condensation assays. Mutated residues are underlined. B: Microscopy images of several SRSF1 LCD variants, concentrated to 400 μM, are shown. Different buffers (50 mM HEPES or 50 mM NaP$_i$, both pH 7.4) and temperatures (4 ℃ and 20 ℃) were screened as indicated. The effect of added poly-U (20mer) RNA was tested (1:40 RNA:peptide molar ratio), as well.

To test for buffer effects potentially impacting the phase behavior of the

peptides, a NaP_i (pH 7.4) buffer system was used for comparison. Indeed, stark differences were found in the phase behavior of SRSF1 LCD variants compared to the HEPES-based buffer system. In NaP_i buffer, condensates in form of droplets were detected for SRSF1 LCD_{Hypo-D} and also SRSF1 $LCD_{Hyper-D}$, albeit to a much lesser extent. Wildtype SRSF1 LCD and SRSF1 $LCD_{Total-D}$, however, did not form detectable condensates. The observed phase behavior of SRSF1 LCD variants in NaP_i buffer is not explainable by the net charge properties of the peptides alone but rather suggests a strong buffer effect. While incorporation of a certain number of negatively charged residues in the peptide may aid phase separation through direct inter-chain Arg-Asp interactions, peptide-peptide interactions have been shown to be modulated by multivalent counterions as well, giving rise to rich phase behaviors [62, 226, 227]. The here used NaP_i buffer provides counterions in form of phosphate ions which have been shown to interact strongly with positively charged residues such as Arg [228] and thereby to modulate the phase behavior of disordered peptides [226]. Therefore, it is conceivable that phosphate ions bind to Arg residues and allow for ion-bridging peptide-peptide interactions resulting in phase separation. In the slightly negatively charged SRSF1 $LCD_{Total-D}$, Arg-phosphate interactions may result in a peptide charge inversion effect which could impede phase separation as observed in similar systems [62, 226, 227]. In the NaP_i buffer, condensates were only found under cold solution conditions suggesting a UCST behavior like described for the HEPES buffer system. Importantly, an additional LCST behavior cannot be excluded at this point as additional data points at higher temperatures are required for assessment. In fact, Arg-rich peptides have previously been found to exhibit an additional LCST behavior, however in form of complex coacervation in combination with polyanions [63]. Similar observations could be made with wildtype SRSF1 LCD before (see section 3.3). As polyanions - more specifically RNAs - are consitutents of NS, their impact on the condensation behavior of phospho-state mimicking SRSF1 LCD variants

should be tested. Poly-U RNA (20mer) was added to the peptide solutions at a molar ratio of 40:1 peptide:RNA prior to imaging. With the addition of RNA, SRSF1 LCD$_{Hypo-D}$ and SRSF1 LCD$_{Hyper-D}$ formed condensates which could also be detected for wildtype SRSF1 LCD, agreeing with previously shown results obtained in the HEPES buffer system (see section 3.3). For SRSF1 LCD$_{Total-D}$, no condensates were detected which again might be due to protein charge inversion through polyanions binding to the peptide [62, 226, 227]. In contrast, the overall positively charged peptides SRSF1 LCD$_{Hypo-D}$ and SRSF1 LCD$_{Hyper-D}$ likely formed condensates via complex coacervation driven by an RNA binding induced charge neutralization - and notably at elevated temperatures (20 °C). Therefore, the development of heterotypic contacts with polyanions appears to alter the temperature sensitivity of the phase behavior of SRSF1 LCD variants. Importantly, this scenario reflects the complex molecular composition within NS which are known to comprise proteins as well as RNA. Hence, heterotypic contacts *in vivo* may support condensate formation even at elevated body temperature.

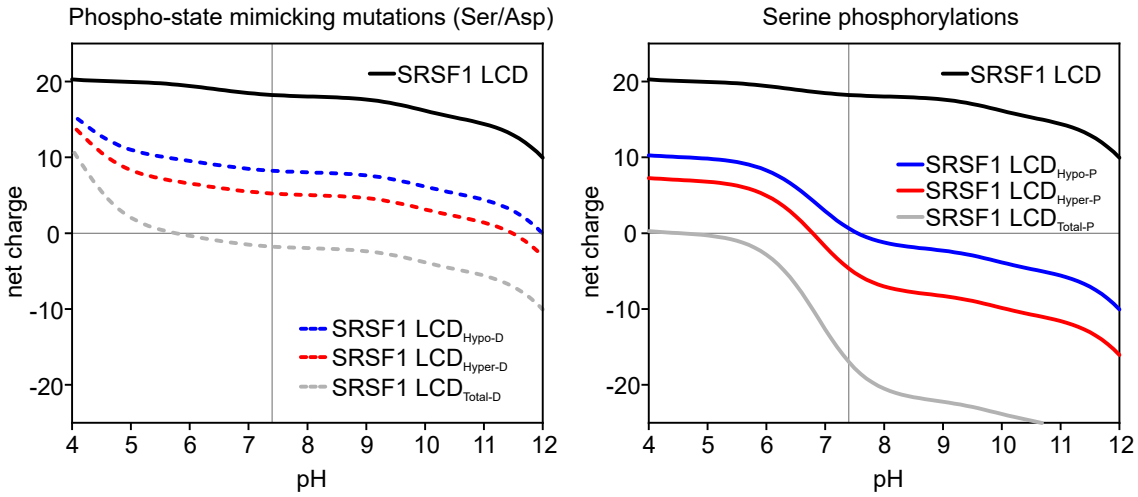


Figure 32: **Net charge characteristics of SRSF1 LCD variants.** The pH dependent net charge properties of wildtype SRSF1 LCD (straight black line), phospho-state mimicking mutants (SRSF1 LCD$_{Hypo-D}$ (broken blue line), SRSF1 LCD$_{Hyper-D}$ (broken red line), SRSF1 LCD$_{Total-D}$ (broken gray line)) (left graph) and differentially phosphorylated SRSF1 LCD variants (SRSF1 LCD$_{Hypo-P}$ (straight blue line), SRSF1 LCD$_{Hyper-P}$ (straight red line), SRSF1 LCD$_{Total-P}$ (straight gray line)) (right graph) are displayed.

Although, it has to be noted that the presented results are preliminary, the currently available data suggest that the condensation behavior of phospho-mimicking SRSF1 LCD variants is altered compared to the wildtype SRSF1 LCD. Among the mutated SRSF1 LCD variants, there appears to be a tendency towards a lower propensity to undergo phase separation the more negative charges are present in the peptide. This matches a previous report by Jedd and coworkers who found that increasing the net-negative charge in Arg-rich mixed-charge peptides abrogated their propensity to form condensates [49]. Moreover, in cells, a similar trend was observed: with increasing net-negative charge, the NS incorporation was abolished indicating that the observed phase behavior of the peptides *in vitro* is a reasonable approximation of the NS formation in cells.

The observed relationship between the charge properties and the phase behavior of Arg-rich peptides further agrees with a previously proposed model of phosphorylation-dependent subcellular localization of SRSF1 [137]. According to this model, hypophosphorylation of SRSF1 is performed by SRPK1 in the cytoplasm which primes the protein for nuclear import and subsequent incorporation into NS. Further phosphorylation of SRSF1 via the nuclear kinase CLK1 yields the hyperphosphorylated state which is released from the NS into the nucleoplasm. Accordingly, the phase behavior of SRSF1 appears to be tightly regulated by differential phosphorylation. Interestingly, the hypophosphorylated state exhibits a neutral net charge at pH 7.4 (see Figure 32) which may explain its high propensity to incorporate into NS. The more extensively phosphorylated variants, however, exhibit a negative net charge, supposedly rendering phase separation unfavorable. This has been previously proposed to be an effect of counteracting the condensation-promoting effect of Arg residues [49].

It is evident - from results shown here and elsewhere - that the Arg/Ser-rich LCD of SRSF1 is an important feature for the incorporation into NS (see Figure 23C) [123]. However, the increased NS incorporation of full-length SRSF1 compared to the isolated SRSF1 LCD (see Figure 23C) indicates that the RRMs contribute

to the protein's phase behavior. Indeed, RRM-mediated specific interactions with the NS residing long non-coding RNA MALAT1 (metastasis-associated lung adenocarcinoma transcript 1) have been proposed to aid NS formation before [212, 229].

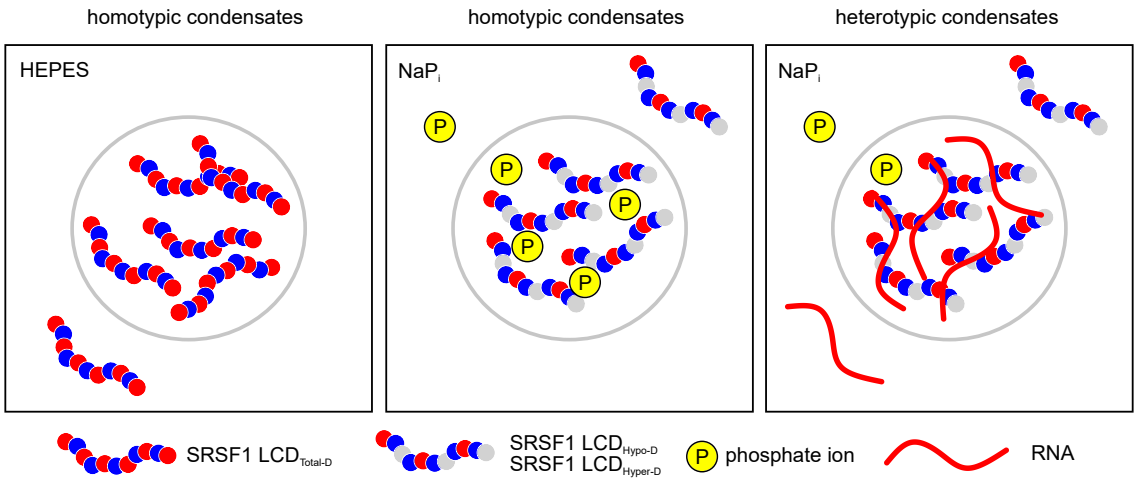


Figure 33: **Model of condensate formation of mixed-charge SRSF1 LCD variants.** SRSF1 LCD$_{Total-D}$ undergoes condensation via homotypic contacts in HEPES buffer (left box). In NaP$_i$ buffer, SRSF1 LCD$_{Hypo-D}$ and SRSF1 LCD$_{Hyper-D}$ undergo homotypic condensation possibly induced by phosphate ions (center box) and heterotypic condensation (right box) via complex coacervation with RNA.

Overall, the data suggest that the phase behavior of SRSF1 is governed by multisite phosphorylation-dependent formation of homotypic and heterotypic contacts and RRM-mediated interactions with RNAs. Therefore, dynamic protein translocations from NS to the nucleoplasm may be regulated tightly through a complex interplay of cellular processes. Efforts in this work were focused on investigating the effect of physiologically relevant phospho-mimicking mutations on the phase behavior of the isolated SRSF1 LCD. Indeed, the phase behavior of SRSF1 LCD was found to vary strongly depending on whether the peptide was mutated at all, and if so, to what extent (see Figure 33). Variations in the negative charge density had an effect on the development of homotypic and heterotypic contacts with lower densities favoring heterotypic contacts with anionic species (phosphate ions/RNA).

Notably, compared to the phospho-mimicking Asp residue which exhibits a single negative charge, actual phosphates exhibit a ~ 1.5 negative charge at pH 7.4. Therefore, the charge properties of phosphorylations are not ideally substituted by the phospho-mimicking mutations which also becomes evident from the net charge characteristics of the various peptides (see Figure 32). Therefore, investigations on phosphorylated SRSF1 LCD variants are required to fully understand the peptide's phase behavior and how it is related to PTMs.

4 Conclusion & Outlook

Protein self-assembly is a critical phenomenon underlying numerous physiological processes. Protein assemblies are a very heterogeneous group with strong variations in their shape and physical properties. Depending on the proteins involved and other factors such as concentration and buffer conditions, assemblies may, for example, display solid or liquid properties. In cells, these various kinds of protein assemblies were found to aid normal physiology.

In this work, the mechanisms of protein self-assembly were studied on two separate systems: the parathyroid system with its amyloid forming peptide hormone PTH and the SR protein system with its nuclear speckle-associated protein SRSF1.

Amyloid formation of PTH_{1-84} is affected by heparin and the disordered C-terminal region

Amyloids were previously proposed to be the natural storage form of numerous peptide hormones from the pituitary system [94]. In case of the peptide hormone PTH_{1-84}, studies concerning its amyloid formation as a potential storage form remained sparse. Although amyloid structures have been identified early on in parathyroid chief cells, a mechanistic understanding of their formation was still elusive, especially under granule-relevant pH conditions (pH 5.5). In this work, the fibril formation of PTH_{1-84} was investigated under near physiological conditions *in vitro*. Under these circumstances, PTH_{1-84} did not undergo spontaneous fibril formation. This might be attributed to the higher positive net charge of the peptide at granule-relevant pH settings, possibly retarding fibril formation through mutual repulsion. Heparin, as a polyanionic glycosaminoglycan (GAG), was found to enable rapid fibril formation of PTH_{1-84} under these conditions, primarily by facilitating the conversion of PTH_{1-84} into amyloid prone structures. As GAGs are well-known components of the Golgi network, PTH_{1-84} may interact with heparin-related structures in cells in a similar way. Therefore, it is conceivable that the heparin-assisted fibrillation of PTH_{1-84}, as characterized *in vitro*, has physiological

relevance. In the future, the role of heparin or chemically similar structures *in vivo* needs to be assessed through cell-based experiments. Intracellular interactors may be revealed by mass spectrometric analysis of granule-preparations of PTH-producing cell lines. This endeavor would greatly benefit from the more recent advances in mass spectrometry for GAG analysis [230].

The α-helical N-terminus of PTH_{1-84} is the focus of most efforts concerning protein structure and activity. This bias in research interest likely stems from the fact that PTH_{1-34} is necessary and sufficient for full receptor stimulation at PTH1R [104]. The C-terminal fragment of PTH_{1-84} is less well characterized. In this work, a high sequence conservation within the disordered C-terminal region of PTH_{1-84} was interpreted to be an indication for a specific function of that part of the peptide. Potential intra-peptide N- to C-terminal contacts were found via MD simulations of PTH_{1-84} and confirmed via cross-linking MS experiments. Interestingly, some of these long range intra-peptide contacts showed the proposed fibrillar core region of PTH_{1-84} to be involved. Therefore, a potential role of the C-terminal region in fibril formation was investigated. Although, data from ThT fluorescence experiments on different PTH variants clearly indicate that fibril formation is affected by the C-terminal region, interpretations of the current dataset proves to be difficult. In order to allow for conclusive statements about the role of the C-terminal region in modulating the fibril formation of PTH_{1-84}, the current dataset needs to be extended. A clear causality between the obtained long range contacts and the modulation of fibril formation remains missing and could be achieved by a mutagenesis-based screening. More specifically, point mutants in which the observed long range contacts are abolished should be examined with a focus on their propensity to form fibrils.

Multisite phosphorylations regulate the condensation behavior of SRSF1

Protein post-translational modifications (PTMs) are vital means to regulate protein functions. Over 400 different types of PTMs have been identified which impact a wide range of cellular processes including enzymatic function

[231], protein degradation [232], protein solubility [233, 234] and folding [235], protein-protein interactions [233, 236] and protein localization [237]. Protein phosphorylation is one of the most common PTMs and, importantly, a reversible process. Thus, proteins can be dynamically phosphorylated and dephosphorylated allowing for a context-dependent tuning of various functional aspects. For example, phosphorylations in TDP-43 have been shown to disrupt RNA splicing activity and condensate formation [218, 238].

Phosphorylations are a well-recognized PTM concerning the nuclear speckle (NS) proteome, as well [117, 119, 126]. Here, protein functions in RNA metabolism have been found to be regulated through phosphorylations before [124, 214]. Yet, the effect of differential phosphorylations on the subcellular localization of these proteins remains less well understood.

In this ongoing study, the condensation behavior of the prototypical SR protein SRSF1 was investigated *in vitro* and *in cellulo*. First, the isolated Arg/Ser-rich SRSF1 LCD was found to be sufficient for NS incorporation suggesting that the specific AA composition - and perhaps patterning - provides cohesion in this type of biomolecular condensate as has been proposed before [49]. The effect of extensive phospho-mimicking mutations in the SRSF1 LCD has been investigated concerning the subcellular localization *in cellulo* and the phase behavior *in vitro*. In cells, a similar speckle pattern was observed for all tested SRSF1 LCD variants. Yet, considerable differences in the observed phase behavior *in vitro* suggest that differential phospohorylations in SRSF1 LCD may regulate the propensity to undergo condensation. More specifically, it was found that the higher the negative charge density was (as a result of varying numbers of S/D mutations), the lower the propensity to undergo condensation through homotypic contacts or through complex coacervation with RNA. This is in line with a previous report [49] and agrees with a current model describing the phosphorylation-dependent subcellular localization of SRSF1 [216].

Phospho-mimicking mutations have been successfully used before to infer

PTM-related changes in protein functions [238]. It has to be noted, though, that they are chemically not fully equivalent to actual phosphorylations as they differ in their charge properties. Since charge-charge interactions contribute to the condensation of SRSF1 LCD, it would be intriguing to investigate the phase behavior of phosphorylated species. Phosphorylations of SRSF1 LCD may be achieved via specific kinases including SRPK1 an CLK1 [133]. Following this, the phase behavior of SRSF1 LCD as a function of PTMs can be more accurately described. Here, an intriguing prospect is to study the phase behavior of complex mixtures, i.e. several phospho-states of SRSF1 LCD in solution. Fluorescent labeling of differentially phosphorylated species may reveal specifically organized subcompartments in dense protein phases. Coexisting liquid phases have been observed to underlie nucleolar subcompartments before [50]. Similarly, the formation of protein dilute phase ”vacuoles” inside the protein dense phase as an effect of excess RNA was previously reported [62].

5 References

(1) Edsall, J. T. Some comments on proteins and protein structure. *Proceedings of the Royal Society of London. Series B, Biological sciences* **1953**, *141*, 97–103.

(2) Uversky, V. N.; Gillespie, J. R.; Fink, A. L. Why are ”natively unfolded” proteins unstructured under physiologic conditions? *Proteins* **2000**, *41*, 415–427.

(3) Dyson, H. J.; Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature reviews. Molecular cell biology* **2005**, *6*, 197–208.

(4) Wright, P. E.; Dyson, H. J. Intrinsically disordered proteins in cellular signalling and regulation. *Nature reviews. Molecular cell biology* **2015**, *16*, 18–29.

(5) Kulkarni, P.; Bhattacharya, S.; Achuthan, S.; Behal, A.; Jolly, M. K.; Kotnala, S.; Mohanty, A.; Rangarajan, G.; Salgia, R.; Uversky, V. Intrinsically Disordered Proteins: Critical Components of the Wetware. *Chemical reviews* **2022**, *122*, 6614–6633.

(6) Romero, P.; Obradovic, Z.; Kissinger, C. R.; Villafranca, J. E.; Garner, E.; Guilliot, S.; Dunker, A. K. Thousands of proteins likely to have long disordered regions. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* **1998**, 437–448.

(7) Dunker, A. K.; Garner, E.; Guilliot, S.; Romero, P.; Albrecht, K.; Hart, J.; Obradovic, Z.; Kissinger, C.; Villafranca, J. E. Protein disorder and the evolution of molecular recognition: theory, predictions and observations. *Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* **1998**, 473–484.

(8) Dunker, A. K. et al. Intrinsically disordered protein. *Journal of molecular graphics & modelling* **2001**, *19*, 26–59.

(9) Dunker, A. K.; Oldfield, C. J.; Meng, J.; Romero, P.; Yang, J. Y.; Chen, J. W.; Vacic, V.; Obradovic, Z.; Uversky, V. N. The unfoldomics decade: an update on intrinsically disordered proteins. *BMC genomics* **2008**, *9 Suppl 2*, S1.

(10) Deiana, A.; Forcelloni, S.; Porrello, A.; Giansanti, A. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PloS one* **2019**, *14*, e0217889.

(11) Liu, J.; Zhang, Y.; Lei, X.; Zhang, Z. Natural selection of protein structural and functional properties: a single nucleotide polymorphism perspective. *Genome biology* **2008**, *9*, R69.

(12) Knockenhauer, K. E.; Schwartz, T. U. The Nuclear Pore Complex as a Flexible and Dynamic Gate. *Cell* **2016**, *164*, 1162–1171.

(13) Yang, J.; Gao, M.; Xiong, J.; Su, Z.; Huang, Y. Features of molecular recognition of intrinsically disordered proteins via coupled folding and binding. *Protein science : a publication of the Protein Society* **2019**, *28*, 1952–1965.

(14) Piersimoni, L. et al. Lighting up Nobel Prize-winning studies with protein intrinsic disorder. *Cellular and molecular life sciences : CMLS* **2022**, *79*, 449.

(15) Bondos, S. E.; Dunker, A. K.; Uversky, V. N. Intrinsically disordered proteins play diverse roles in cell signaling. *Cell communication and signaling : CCS* **2022**, *20*, 20.

(16) van der Lee, R. et al. Classification of intrinsically disordered regions and proteins. *Chemical reviews* **2014**, *114*, 6589–6631.

(17) Song, J.; Ng, S. C.; Tompa, P.; Lee, K. A. W.; Chan, H. S. Polycation-p interactions are a driving force for molecular recognition by an intrinsically disordered oncoprotein family. *PLoS computational biology* **2013**, *9*, e1003239.

(18) Dogan, J.; Gianni, S.; Jemth, P. The binding mechanisms of intrinsically disordered proteins. *Physical chemistry chemical physics : PCCP* **2014**, *16*, 6323–6331.

(19) Bah, A.; Forman-Kay, J. D. Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. *The Journal of biological chemistry* **2016**, *291*, 6696–6705.

(20) Mier, P.; Andrade-Navarro, M. A. Assessing the low complexity of protein sequences via the low complexity triangle. *PloS one* **2020**, *15*, e0239154.

(21) Ota, M.; Koike, R.; Amemiya, T.; Tenno, T.; Romero, P. R.; Hiroaki, H.; Dunker, A. K.; Fukuchi, S. An assignment of intrinsically disordered regions of proteins based on NMR structures. *Journal of structural biology* **2013**, *181*, 29–36.

(22) Kikhney, A. G.; Svergun, D. I. A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Letters* **2015**, *589*, 2570–2577.

(23) Gibbs, E. B.; Cook, E. C.; Showalter, S. A. Application of NMR to studies of intrinsically disordered proteins. *Archives of biochemistry and biophysics* **2017**, *628*, 57–70.

(24) Hofmann, H. Understanding disordered and unfolded proteins using single-molecule FRET and polymer theory. *Methods and applications in fluorescence* **2016**, *4*, 042003.

(25) Erdős, G.; Pajkos, M.; Dosztányi, Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic acids research* **2021**, *49*, W297–W303.

(26) Hardenberg, M.; Horvath, A.; Ambrus, V.; Fuxreiter, M.; Vendruscolo, M. Widespread occurrence of the droplet state of proteins in the human proteome. *Proceedings of the National Academy of Sciences of the United States of America* **2020**, *117*, 33254–33262.

(27) Martín, M.; Brunello, F. G.; Modenutti, C. P.; Nicola, J. P.; Marti, M. A. MotSASi: Functional short linear motifs (SLiMs) prediction based on genomic single nucleotide variants and structural data. *Biochimie* **2022**, *197*, 59–73.

(28) Mu, J.; Liu, H.; Zhang, J.; Luo, R.; Chen, H.-F. Recent Force Field Strategies for Intrinsically Disordered Proteins. *Journal of chemical information and modeling* **2021**, *61*, 1037–1047.

(29) Srivastava, A.; Tiwari, S. P.; Miyashita, O.; Tama, F. Integrative/Hybrid Modeling Approaches for Studying Biomolecules. *Journal of Molecular Biology* **2020**, *432*, 2846–2860.

(30) Alberti, S. The wisdom of crowds: regulating cell function through condensed states of living matter. *Journal of cell science* **2017**, *130*, 2789–2796.

(31) Martin, E. W.; Mittag, T. Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions. *Biochemistry* **2018**, *57*, 2478–2487.

(32) Falahati, H.; Haji-Akbari, A. Thermodynamically driven assemblies and liquid-liquid phase separations in biology. *Soft matter* **2019**, *15*, 1135–1154.

(33) Stradner, A.; Schurtenberger, P. Potential and limits of a colloid approach to protein solutions. *Soft matter* **2020**, *16*, 307–323.

(34) Dumetz, A. C.; Chockla, A. M.; Kaler, E. W.; Lenhoff, A. M. Effects of pH on protein-protein interactions and implications for protein phase behavior. *Biochimica et biophysica acta* **2008**, *1784*, 600–610.

(35) Dumetz, A. C.; Chockla, A. M.; Kaler, E. W.; Lenhoff, A. M. Protein phase behavior in aqueous solutions: crystallization, liquid-liquid phase separation, gels, and aggregates. *Biophysical journal* **2008**, *94*, 570–583.

(36) Galkin, O.; Vekilov, P. G. Control of protein crystal nucleation around the metastable liquid-liquid phase boundary. *Proceedings of the National Academy of Sciences* **2000**, *97*, 6277–6281.

(37) Vekilov, P. G. Dense Liquid Precursor for the Nucleation of Ordered Solid Phases from Solution. *Crystal Growth & Design* **2004**, *4*, 671–685.

(38) McManus, J. J.; Charbonneau, P.; Zaccarelli, E.; Asherie, N. The physics of protein self-assembly. *Current Opinion in Colloid & Interface Science* **2016**, *22*, 73–79.

(39) Vekilov, P. G. Phase transitions of folded proteins. *Soft matter* **2010**, *6*, 5254.

(40) Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer physics of intracellular phase transitions. *Nature Physics* **2015**, *11*, 899–904.

(41) Alberti, S.; Gladfelter, A.; Mittag, T. Considerations and Challenges in Studying Liquid-Liquid Phase Separation and Biomolecular Condensates. *Cell* **2019**, *176*, 419–434.

(42) Abbas, M.; Lipiński, W. P.; Wang, J.; Spruijt, E. Peptide-based coacervates as biomimetic protocells. *Chemical Society reviews* **2021**, *50*, 3690–3705.

(43) Nott, T. J.; Petsalaki, E.; Farber, P.; Jervis, D.; Fussner, E.; Plochowietz, A.; Craggs, T. D.; Bazett-Jones, D. P.; Pawson, T.; Forman-Kay, J. D.; Baldwin, A. J. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Molecular Cell* **2015**, *57*, 936–947.

(44) Overbeek, J. T. G.; Voorn, M. J. Phase separation in polyelectrolyte solutions. Theory of complex coacervation. *Journal of Cellular and Comparative Physiology* **1957**, *49*, 7–26.

(45) Lin, Y.-H.; Forman-Kay, J. D.; Chan, H. S. Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles. *Physical review letters* **2016**, *117*, 178101.

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

(47) Hirose, T.; Ninomiya, K.; Nakagawa, S.; Yamazaki, T. A guide to membraneless organelles and their various roles in gene regulation. *Nature reviews. Molecular cell biology* **2023**, *24*, 288–304.

(48) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P granules are liquid droplets that localize by controlled dissolution/condensation. *Science (New York, N.Y.)* **2009**, *324*, 1729–1732.

(49) Greig, J. A.; Nguyen, T. A.; Lee, M.; Holehouse, A. S.; Posey, A. E.; Pappu, R. V.; Jedd, G. Arginine-Enriched Mixed-Charge Domains Provide Cohesion for Nuclear Speckle Condensation. *Molecular Cell* **2020**, *77*, 1237–1250.e4.

(50) Feric, M.; Vaidya, N.; Harmon, T. S.; Mitrea, D. M.; Zhu, L.; Richardson, T. M.; Kriwacki, R. W.; Pappu, R. V.; Brangwynne, C. P. Coexisting Liquid Phases Underlie Nucleolar Subcompartments. *Cell* **2016**, *165*, 1686–1697.

(51) Li, P.; Banjade, S.; Cheng, H.-C.; Kim, S.; Chen, B.; Guo, L.; Llaguno, M.; Hollingsworth, J. V.; King, D. S.; Banani, S. F.; Russo, P. S.; Jiang, Q.-X.; Nixon, B. T.; Rosen, M. K. Phase transitions in the assembly of multivalent signalling proteins. *Nature* **2012**, *483*, 336–340.

(52) Wang, A.; Conicella, A. E.; Schmidt, H. B.; Martin, E. W.; Rhoads, S. N.; Reeb, A. N.; Nourse, A.; Ramirez Montero, D.; Ryan, V. H.; Rohatgi, R.; Shewmaker, F.; Naik, M. T.; Mittag, T.; Ayala, Y. M.; Fawzi, N. L. A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing. *The EMBO Journal* **2018**, *37*, DOI: 10.15252/embj.201797452.

(53) Mitrea, D. M.; Cika, J. A.; Guy, C. S.; Ban, D.; Banerjee, P. R.; Stanley, C. B.; Nourse, A.; Deniz, A. A.; Kriwacki, R. W. Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA. *eLife* **2016**, *5*, DOI: 10.7554/eLife.13571.

(54) Marzahn, M. R. et al. Higher-order oligomerization promotes localization of SPOP to liquid nuclear speckles. *The EMBO Journal* **2016**, *35*, 1254–1275.

(55) Borcherds, W.; Bremer, A.; Borgia, M. B.; Mittag, T. How do intrinsically disordered protein regions encode a driving force for liquid-liquid phase separation? *Current opinion in structural biology* **2021**, *67*, 41–50.

(56) Mohanty, P.; Shenoy, J.; Rizuan, A.; Mercado-Ortiz, J. F.; Fawzi, N. L.; Mittal, J. A synergy between site-specific and transient interactions drives the phase separation of a disordered, low-complexity domain. *Proceedings of the National Academy of Sciences of the United States of America* **2023**, *120*, e2305625120.

(57) Choi, J.-M.; Holehouse, A. S.; Pappu, R. V. Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions. *Annual review of biophysics* **2020**, *49*, 107–133.

(58) Wang, J.; Choi, J.-M.; Holehouse, A. S.; Lee, H. O.; Zhang, X.; Jahnel, M.; Maharana, S.; Lemaitre, R.; Pozniakovsky, A.; Drechsel, D.; Poser, I.; Pappu, R. V.; Alberti, S.; Hyman, A. A. A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins. *Cell* **2018**, *174*, 688–699.e16.

(59) Martin, E. W.; Holehouse, A. S.; Peran, I.; Farag, M.; Incicco, J. J.; Bremer, A.; Grace, C. R.; Soranno, A.; Pappu, R. V.; Mittag, T. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science (New York, N.Y.)* **2020**, *367*, 694–699.

(60) Schuster, B. S.; Dignon, G. L.; Tang, W. S.; Kelley, F. M.; Ranganath, A. K.; Jahnke, C. N.; Simpkins, A. G.; Regy, R. M.; Hammer, D. A.; Good, M. C.; Mittal, J. Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proceedings of the National Academy of Sciences* **2020**, *117*, 11421–11431.

(61) Das, R. K.; Pappu, R. V. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 13392–13397.

(62) Banerjee, P. R.; Milin, A. N.; Moosa, M. M.; Onuchic, P. L.; Deniz, A. A. Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets. *Angewandte Chemie (International ed. in English)* **2017**, *56*, 11354–11359.

(63) Hong, Y.; Najafi, S.; Casey, T.; Shea, J.-E.; Han, S.-I.; Hwang, D. S. Hydrophobicity of arginine leads to reentrant liquid-liquid phase separation behaviors of arginine-rich proteins. *Nature communications* **2022**, *13*, 7326.

(64) Wang, Z.; Zhang, G.; Zhang, H. Protocol for analyzing protein liquid–liquid phase separation. *Biophysics Reports* **2019**, *5*, 1–9.

(65) 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. *Nature communications* **2019**, *10*, 2909.

(66) Munishkina, L. A.; Cooper, E. M.; Uversky, V. N.; Fink, A. L. The effect of macromolecular crowding on protein aggregation and amyloid fibril formation. *Journal of molecular recognition : JMR* **2004**, *17*, 456–464.

(67) Fox, A. H.; Nakagawa, S.; Hirose, T.; Bond, C. S. Paraspeckles: Where Long Noncoding RNA Meets Phase Separation. *Trends in biochemical sciences* **2018**, *43*, 124–135.

(68) Guillén-Boixet, J. et al. RNA-Induced Conformational Switching and Clustering of G3BP Drive Stress Granule Assembly by Condensation. *Cell* **2020**, *181*, 346–361.e17.

(69) Boeynaems, S.; Alberti, S.; Fawzi, N. L.; Mittag, T.; Polymenidou, M.; Rousseau, F.; Schymkowitz, J.; Shorter, J.; Wolozin, B.; van den Bosch, L.; Tompa, P.; Fuxreiter, M. Protein Phase Separation: A New Phase in Cell Biology. *Trends in cell biology* **2018**, *28*, 420–435.

(70) Uversky, V. N. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Current opinion in structural biology* **2017**, *44*, 18–30.

(71) Yamazaki, T.; Souquere, S.; Chujo, T.; Kobelke, S.; Chong, Y. S.; Fox, A. H.; Bond, C. S.; Nakagawa, S.; Pierron, G.; Hirose, T. Functional Domains of NEAT1 Architectural lncRNA Induce Paraspeckle Assembly through Phase Separation. *Molecular Cell* **2018**, *70*, 1038–1053.e7.

(72) Xue, S.; Gong, R.; He, F.; Li, Y.; Wang, Y.; Tan, T.; Luo, S.-Z. Low-complexity domain of U1-70K modulates phase separation and aggregation through distinctive basic-acidic motifs. *Science advances* **2019**, *5*, eaax5349.

(73) Molliex, A.; Temirov, J.; Lee, J.; Coughlin, M.; Kanagaraj, A. P.; Kim, H. J.; Mittag, T.; Taylor, J. P. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **2015**, *163*, 123–133.

(74) Ditlev, J. A.; Case, L. B.; Rosen, M. K. Who's In and Who's Out-Compositional Control of Biomolecular Condensates. *Journal of Molecular Biology* **2018**, *430*, 4666–4684.

(75) Patel, A. et al. A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation. *Cell* **2015**, *162*, 1066–1077.

(76) Tatomer, D. C.; Terzo, E.; Curry, K. P.; Salzler, H.; Sabath, I.; Zapotoczny, G.; McKay, D. J.; Dominski, Z.; Marzluff, W. F.; Duronio, R. J. Concentrating pre-mRNA processing factors in the histone locus body facilitates efficient histone mRNA biogenesis. *The Journal of cell biology* **2016**, *213*, 557–570.

(77) Castellana, M.; Wilson, M. Z.; Xu, Y.; Joshi, P.; Cristea, I. M.; Rabinowitz, J. D.; Gitai, Z.; Wingreen, N. S. Enzyme clustering accelerates processing of intermediates through metabolic channeling. *Nature biotechnology* **2014**, *32*, 1011–1018.

(78) Li, H.; Leo, C.; Zhu, J.; Wu, X.; O'Neil, J.; Park, E. J.; Chen, J. D. Sequestration and inhibition of Daxx-mediated transcriptional repression by PML. *Molecular and cellular biology* **2000**, *20*, 1784–1796.

(79) Eldar, A.; Elowitz, M. B. Functional roles for noise in genetic circuits. *Nature* **2010**, *467*, 167–173.

(80) Jiang, H.; Wang, S.; Huang, Y.; He, X.; Cui, H.; Zhu, X.; Zheng, Y. Phase transition of spindle-associated protein regulate spindle apparatus assembly. *Cell* **2015**, *163*, 108–122.

(81) Lin, Y.; Protter, D. S. W.; Rosen, M. K.; Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Molecular Cell* **2015**, *60*, 208–219.

(82) Rambaran, R. N.; Serpell, L. C. Amyloid fibrils: abnormal protein assembly. *Prion* **2008**, *2*, 112–117.

(83) Friedreich, N.; Kekulé, A. Zur Amyloidfrage. *Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin* **1859**, *16*, 50–65.

(84) Zhang, Z.; Huang, G.; Song, Z.; Gatch, A. J.; Ding, F. Amyloid Aggregation and Liquid-Liquid Phase Separation from the Perspective of Phase Transitions. *The journal of physical chemistry. B* **2023**, DOI: 10.1021/acs.jpcb.3c01426.

(85) Otzen, D.; Riek, R. Functional Amyloids. *Cold Spring Harbor perspectives in biology* **2019**, *11*, DOI: 10.1101/cshperspect.a033860.

(86) Chiti, F.; Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annual review of biochemistry* **2006**, *75*, 333–366.

(87) Greenwald, J.; Riek, R. Biology of amyloid: structure, function, and regulation. *Structure (London, England : 1993)* **2010**, *18*, 1244–1260.

(88) Riek, R.; Eisenberg, D. S. The activities of amyloids from a structural perspective. *Nature* **2016**, *539*, 227–235.

(89) Lue, L. F.; Kuo, Y. M.; Roher, A. E.; Brachova, L.; Shen, Y.; Sue, L.; Beach, T.; Kurth, J. H.; Rydel, R. E.; Rogers, J. Soluble amyloid beta peptide concentration as a predictor of synaptic change in Alzheimer's disease. *The American journal of pathology* **1999**, *155*, 853–862.

(90) Chapman, M. R.; Robinson, L. S.; Pinkner, J. S.; Roth, R.; Heuser, J.; Hammar, M.; Normark, S.; Hultgren, S. J. Role of Escherichia coli curli operons in directing amyloid fiber formation. *Science (New York, N.Y.)* **2002**, *295*, 851–855.

(91) Maddelein, M.-L.; Dos Reis, S.; Duvezin-Caubet, S.; Coulary-Salin, B.; Saupe, S. J. Amyloid aggregates of the HET-s prion protein are infectious. *Proceedings of the National Academy of Sciences* **2002**, *99*, 7402–7407.

(92) Gebbink, M. F. B. G.; Claessen, D.; Bouma, B.; Dijkhuizen, L.; Wösten, H. A. B. Amyloids–a functional coat for microorganisms. *Nature reviews. Microbiology* **2005**, *3*, 333–341.

(93) Fowler, D. M.; Koulov, A. V.; Alory-Jost, C.; Marks, M. S.; Balch, W. E.; Kelly, J. W. Functional amyloid formation within mammalian tissue. *PLoS biology* **2006**, *4*, e6.

(94) Maji, S. K.; Perrin, M. H.; Sawaya, M. R.; Jessberger, S.; Vadodaria, K.; Rissman, R. A.; Singru, P. S.; Nilsson, K. P. R.; Simon, R.; Schubert, D.; Eisenberg, D.; Rivier, J.; Sawchenko, P.; Vale, W.; Riek, R. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science (New York, N.Y.)* **2009**, *325*, 328–332.

(95) Evans, M. L.; Chapman, M. R. Curli biogenesis: order out of disorder. *Biochimica et biophysica acta* **2014**, *1843*, 1551–1558.

(96) Harvey, Z. H.; Chen, Y.; Jarosz, D. F. Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Molecular Cell* **2018**, *69*, 195–202.

(97) Michaels, T. C. T.; Qian, D.; Šarić, A.; Vendruscolo, M.; Linse, S.; Knowles, T. P. J. Amyloid formation as a protein phase transition. *Nature Reviews Physics* **2023**, *5*, 379–397.

(98) Collip, J. B. The extraction of a parathyroid hormone which will prevent or control parathyroid tetany and which regulates the levels of blood calcium. *Journal of Biological Chemistry* **1925**, *63*, 395–438.

(99) Potts, J. T. Parathyroid hormone: past and present. *The Journal of endocrinology* **2005**, *187*, 311–325.

(100) Marx, U. C.; Adermann, K.; Bayer, P.; Forssmann, W. G.; Rösch, P. Solution structures of human parathyroid hormone fragments hPTH(1-34) and hPTH(1-39) and bovine parathyroid hormone fragment bPTH(1-37). *Biochemical and Biophysical Research Communications* **2000**, *267*, 213–220.

(101) Murray, T. M.; Rao, L. G.; Divieti, P.; Bringhurst, F. R. Parathyroid hormone secretion and action: evidence for discrete receptors for the carboxyl-terminal region and related biological actions of carboxyl- terminal ligands. *Endocrine reviews* **2005**, *26*, 78–113.

(102) Jüppner, H.; Abou-Samra, A. B.; Freeman, M.; Kong, X. F.; Schipani, E.; Richards, J.; Kolakowski, L. F.; Hock, J.; Potts, J. T.; Kronenberg, H. M. A G protein-linked receptor for parathyroid hormone and parathyroid hormone-related peptide. *Science (New York, N.Y.)* **1991**, *254*, 1024–1026.

(103) Habener, J. F.; Rosenblatt, M.; Potts, J. T. Parathyroid hormone: biochemical aspects of biosynthesis, secretion, action, and metabolism. *Physiological reviews* **1984**, *64*, 985–1053.

(104) Potts, J. T.; Tregear, G. W.; Keutmann, H. T.; Niall, H. D.; Sauer, R.; Deftos, L. J.; Dawson, B. F.; Hogan, M. L.; Aurbach, G. D. Synthesis of a biologically active N-terminal tetratriacontapeptide of parathyroid hormone. *Proceedings of the National Academy of Sciences* **1971**, *68*, 63–67.

(105) Bodenner, D.; Redman, C.; Riggs, A. Teriparatide in the management of osteoporosis. *Clinical interventions in aging* **2007**, *2*, 499–507.

(106) Gronwald, W.; Schomburg, D.; Harder, M. P.; Mayer, H.; Paulsen, J.; Wingender, E.; Wray, V. Structure of recombinant human parathyroid hormone in solution using multidimensional NMR spectroscopy. *Biological chemistry Hoppe-Seyler* **1996**, *377*, 175–186.

(107) Paroutis, P.; Touret, N.; Grinstein, S. The pH of the secretory pathway: measurement, determinants, and regulation. *Physiology (Bethesda, Md.)* **2004**, *19*, 207–215.

(108) Dannies, P. S. Prolactin and growth hormone aggregates in secretory granules: the need to understand the structure of the aggregate. *Endocrine reviews* **2012**, *33*, 254–270.

(109) Anderson, T. J.; Ewen, S. W. Amyloid in normal and pathological parathyroid glands. *Journal of clinical pathology* **1974**, *27*, 656–663.

(110) Furukawa, K.; Aguirre, C.; So, M.; Sasahara, K.; Miyanoiri, Y.; Sakurai, K.; Yamaguchi, K.; Ikenaka, K.; Mochizuki, H.; Kardos, J.; Kawata, Y.; Goto, Y. Isoelectric point-amyloid formation of a-synuclein extends the generality of the solubility and supersaturation-limited mechanism. *Current Research in Structural Biology* **2020**, *2*, 35–44.

(111) Nespovitaya, N.; Gath, J.; Barylyuk, K.; Seuring, C.; Meier, B. H.; Riek, R. Dynamic Assembly and Disassembly of Functional b-Endorphin Amyloid Fibrils. *Journal of the American Chemical Society* **2016**, *138*, 846–856.

(112) Marek, P. J.; Patsalo, V.; Green, D. F.; Raleigh, D. P. Ionic strength effects on amyloid formation by amylin are a complicated interplay among Debye screening, ion selectivity, and Hofmeister effects. *Biochemistry* **2012**, *51*, 8478–8490.

(113) Gopalswamy, M.; Kumar, A.; Adler, J.; Baumann, M.; Henze, M.; Kumar, S. T.; Fändrich, M.; Scheidt, H. A.; Huster, D.; Balbach, J. Structural characterization of amyloid fibrils from the human parathyroid hormone. *Biochimica et biophysica acta* **2015**, *1854*, 249–257.

(114) Priyadarshini, M.; Bano, B. Conformational changes during amyloid fibril formation of pancreatic thiol proteinase inhibitor: effect of copper and zinc. *Molecular biology reports* **2012**, *39*, 2945–2955.

(115) Rogalska, M. E.; Vivori, C.; Valcárcel, J. Regulation of pre-mRNA splicing: roles in physiology and disease, and therapeutic prospects. *Nature reviews. Genetics* **2023**, *24*, 251–269.

(116) Baralle, F. E.; Giudice, J. Alternative splicing as a regulator of development and tissue identity. *Nature reviews. Molecular cell biology* **2017**, *18*, 437–451.

(117) Slišković, I.; Eich, H.; Müller-McNicoll, M. Exploring the multifunctionality of SR proteins. *Biochemical Society transactions* **2022**, *50*, 187–198.

(118) Twyffels, L.; Gueydan, C.; Kruys, V. Shuttling SR proteins: more than splicing factors. *The FEBS journal* **2011**, *278*, 3246–3255.

(119) Das, S.; Krainer, A. R. Emerging functions of SRSF1, splicing factor and oncoprotein, in RNA metabolism and cancer. *Molecular cancer research : MCR* **2014**, *12*, 1195–1204.

(120) Galganski, L.; Urbanek, M. O.; Krzyzosiak, W. J. Nuclear speckles: molecular organization, biological function and role in disease. *Nucleic acids research* **2017**, *45*, 10350–10368.

(121) Wagner, R. E.; Frye, M. Noncanonical functions of the serine-arginine-rich splicing factor (SR) family of proteins in development and disease. *BioEssays : news and reviews in molecular, cellular and developmental biology* **2021**, *43*, e2000242.

(122) Cáceres, J. F.; Krainer, A. R. Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains. *The EMBO Journal* **1993**, *12*, 4715–4726.

(123) Cáceres, J. F.; Misteli, T.; Screaton, G. R.; Spector, D. L.; Krainer, A. R. Role of the modular domains of SR proteins in subnuclear localization and alternative splicing specificity. *The Journal of cell biology* **1997**, *138*, 225–238.

(124) Cho, S.; Hoang, A.; Sinha, R.; Zhong, X.-Y.; Fu, X.-D.; Krainer, A. R.; Ghosh, G. Interaction between the RNA binding domains of Ser-Arg splicing factor 1 and U1-70K snRNP protein determines early spliceosome assembly. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 8233–8238.

(125) Zuo, P.; Manley, J. L. Functional domains of the human splicing factor ASF/SF2. *The EMBO Journal* **1993**, *12*, 4727–4737.

(126) Spector, D. L.; Lamond, A. I. Nuclear speckles. *Cold Spring Harbor perspectives in biology* **2011**, *3*, DOI: 10.1101/cshperspect.a000646.

(127) Quinodoz, S. A. et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **2018**, *174*, 744–757.e24.

(128) Chen, Y.; Belmont, A. S. Genome organization around nuclear speckles. *Current opinion in genetics & development* **2019**, *55*, 91–99.

(129) Fei, J.; Jadaliha, M.; Harmon, T. S.; Li, I. T. S.; Hua, B.; Hao, Q.; Holehouse, A. S.; Reyer, M.; Sun, Q.; Freier, S. M.; Pappu, R. V.; Prasanth, K. V.; Ha, T. Quantitative analysis of multilayer organization of proteins and RNA in nuclear speckles at super resolution. *Journal of cell science* **2017**, *130*, 4180–4192.

(130) Smith, K. P.; Moen, P. T.; Wydner, K. L.; Coleman, J. R.; Lawrence, J. B. Processing of endogenous pre-mRNAs in association with SC-35 domains is gene specific. *The Journal of cell biology* **1999**, *144*, 617–629.

(131) Ilik, İ. A.; Malszycki, M.; Lübke, A. K.; Schade, C.; Meierhofer, D.; Aktaş, T. SON and SRRM2 are essential for nuclear speckle formation. *eLife* **2020**, *9*, DOI: 10.7554/eLife.60579.

(132) Ngo, J. C. K.; Chakrabarti, S.; Ding, J.-H.; Velazquez-Dones, A.; Nolen, B.; Aubol, B. E.; Adams, J. A.; Fu, X.-D.; Ghosh, G. Interplay between SRPK and Clk/Sty kinases in phosphorylation of the splicing factor ASF/SF2 is regulated by a docking motif in ASF/SF2. *Molecular Cell* **2005**, *20*, 77–89.

(133) Velazquez-Dones, A.; Hagopian, J. C.; Ma, C.-T.; Zhong, X.-Y.; Zhou, H.; Ghosh, G.; Fu, X.-D.; Adams, J. A. Mass spectrometric and kinetic analysis of ASF/SF2 phosphorylation by SRPK1 and Clk/Sty. *The Journal of biological chemistry* **2005**, *280*, 41761–41768.

(134) Kataoka, N.; Bachorik, J. L.; Dreyfuss, G. Transportin-SR, a nuclear import receptor for SR proteins. *The Journal of cell biology* **1999**, *145*, 1145–1152.

(135) Gui, J. F.; Lane, W. S.; Fu, X. D. A serine kinase regulates intracellular localization of splicing factors in the cell cycle. *Nature* **1994**, *369*, 678–682.

(136) Keshwani, M. M.; Aubol, B. E.; Fattet, L.; Ma, C.-T.; Qiu, J.; Jennings, P. A.; Fu, X.-D.; Adams, J. A. Conserved proline-directed phosphorylation regulates SR protein conformation and splicing function. *Biochemical Journal* **2015**, *466*, 311–322.

(137) Aubol, B. E.; Wu, G.; Keshwani, M. M.; Movassat, M.; Fattet, L.; Hertel, K. J.; Fu, X.-D.; Adams, J. A. Release of SR Proteins from CLK1 by SRPK1: A Symbiotic Kinase System for Phosphorylation Control of Pre-mRNA Splicing. *Molecular Cell* **2016**, *63*, 218–228.

(138) Dick, G.; Akslen-Hoel, L. K.; Grøndahl, F.; Kjos, I.; Prydz, K. Proteoglycan synthesis and Golgi organization in polarized epithelial cells. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **2012**, *60*, 926–935.

(139) Rai, A. K.; Chen, J.-X.; Selbach, M.; Pelkmans, L. Kinase-controlled phase transition of membraneless organelles in mitosis. *Nature* **2018**, *559*, 211–216.

(140) Yue, B. G.; Ajuh, P.; Akusjärvi, G.; Lamond, A. I.; Kreivi, J. P. Functional coexpression of serine protein kinase SRPK1 and its substrate ASF/SF2 in Escherichia coli. *Nucleic acids research* **2000**, *28*, E14.

(141) Lebowitz, J.; Lewis, M. S.; Schuck, P. Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein science : a publication of the Protein Society* **2002**, *11*, 2067–2079.

(142) Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical journal* **2000**, *78*, 1606–1619.

(143) Kelly, S. M.; Jess, T. J.; Price, N. C. How to study proteins by circular dichroism. *Biochimica et biophysica acta* **2005**, *1751*, 119–139.

(144) Abriata, L. A. A Simple Spreadsheet Program To Simulate and Analyze the Far-UV Circular Dichroism Spectra of Proteins. *Journal of Chemical Education* **2011**, *88*, 1268–1273.

(145) Barth, A. Infrared spectroscopy of proteins. *Biochimica et biophysica acta* **2007**, *1767*, 1073–1101.

(146) Pavlov, G.; Finet, S.; Tatarenko, K.; Korneeva, E.; Ebel, C. Conformation of heparin studied with macromolecular hydrodynamic methods and X-ray scattering. *European biophysics journal : EBJ* **2003**, *32*, 437–449.

(147) Roe, R. J., *Methods of X-ray and neutron scattering in polymer science*; Topics in polymer science; Oxford University Press: New York and Oxford, 2000.

(148) Pedersen, J. S.; Schurtenberger, P. Scattering Functions of Semiflexible Polymers with and without Excluded Volume Effects. *Macromolecules* **1996**, *29*, 7602–7612.

(149) Chen, W.-R.; Butler, P. D.; Magid, L. J. Incorporating intermicellar interactions in the fitting of SANS data from cationic wormlike micelles. *Langmuir : the ACS journal of surfaces and colloids* **2006**, *22*, 6539–6548.

(150) Khorramian, B. A.; Stivala, S. S. Small-angle x-ray scattering of high- and low-affinity heparin. *Archives of biochemistry and biophysics* **1986**, *247*, 384–392.

(151) Kumar, A.; Gopalswamy, M.; Wishart, C.; Henze, M.; Eschen-Lippold, L.; Donnelly, D.; Balbach, J. N-terminal phosphorylation of parathyroid hormone (PTH) abolishes its receptor activity. *ACS chemical biology* **2014**, *9*, 2465–2470.

(152) Keller, S.; Vargas, C.; Zhao, H.; Piszczek, G.; Brautigam, C. A.; Schuck, P. High-precision isothermal titration calorimetry with automated peak-shape analysis. *Analytical chemistry* **2012**, *84*, 5066–5073.

(153) Le, V. H.; Buscaglia, R.; Chaires, J. B.; Lewis, E. A. Modeling complex equilibria in isothermal titration calorimetry experiments: thermodynamic parameters estimation for a three-binding-site model. *Analytical Biochemistry* **2013**, *434*, 233–241.

(154) Meisl, G.; Rajah, L.; Cohen, S. A. I.; Pfammatter, M.; Šarić, A.; Hellstrand, E.; Buell, A. K.; Aguzzi, A.; Linse, S.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Scaling behaviour and rate-determining steps in filamentous self-assembly. *Chemical science* **2017**, *8*, 7087–7097.

(155) Michaels, T. C. T.; Šarić, A.; Curk, S.; Bernfur, K.; Arosio, P.; Meisl, G.; Dear, A. J.; Cohen, S. I. A.; Dobson, C. M.; Vendruscolo, M.; Linse, S.; Knowles, T. P. J. Dynamics of oligomer populations formed during the aggregation of Alzheimer's Ab42 peptide. *Nature chemistry* **2020**, *12*, 445–451.

(156) Dear, A. J.; Michaels, T. C. T.; Meisl, G.; Klenerman, D.; Wu, S.; Perrett, S.; Linse, S.; Dobson, C. M.; Knowles, T. P. J. Kinetic diversity of amyloid oligomers. *Proceedings of the National Academy of Sciences of the United States of America* **2020**, *117*, 12087–12094.

(157) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **2012**, *9*, 671–675.

(158) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics* **1983**, *79*, 926–935.

(159) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. *The Journal of chemical physics* **1993**, *98*, 10089–10092.

(160) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *The Journal of chemical physics* **1994**, *101*, 4177–4189.

(161) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of chemical physics* **1995**, *103*, 4613–4621.

(162) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of computational chemistry* **2005**, *26*, 1781–1802.

(163) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; Mackerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone f, y and side-chain q(1) and q(2) dihedral angles. *Journal of chemical theory and computation* **2012**, *8*, 3257–3273.

(164) MacKerell, A. D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The journal of physical chemistry. B* **1998**, *102*, 3586–3616.

(165) Müller, M. Q.; Dreiocker, F.; Ihling, C. H.; Schäfer, M.; Sinz, A. Cleavable cross-linker for protein structure analysis: reliable identification of cross-linking products by tandem MS. *Analytical chemistry* **2010**, *82*, 6958–6968.

(166) Kamerzell, T. J.; Joshi, S. B.; McClean, D.; Peplinskie, L.; Toney, K.; Papac, D.; Li, M.; Middaugh, C. R. Parathyroid hormone is a heparin/polyanion binding protein: binding energetics and structure modification. *Protein science : a publication of the Protein Society* **2007**, *16*, 1193–1203.

(167) Lauth, L. M.; Voigt, B.; Bhatia, T.; Machner, L.; Balbach, J.; Ott, M. Heparin promotes rapid fibrillation of the basic parathyroid hormone at physiological pH. *FEBS Letters* **2022**, *596*, 2928–2939.

(168) Venkataraman, G.; Sasisekharan, V.; Herr, A. B.; Ornitz, D. M.; Waksman, G.; Cooney, C. L.; Langer, R.; Sasisekharan, R. Preferential self-association of basic fibroblast growth factor is stabilized by heparin during receptor dimerization and activation. *Proceedings of the National Academy of Sciences* **1996**, *93*, 845–850.

(169) Kuschert, G. S.; Coulin, F.; Power, C. A.; Proudfoot, A. E.; Hubbard, R. E.; Hoogewerf, A. J.; Wells, T. N. Glycosaminoglycans interact selectively with chemokines and modulate receptor binding and cellular responses. *Biochemistry* **1999**, *38*, 12959–12968.

(170) Chen, Z.; Xu, P.; Barbier, J. R.; Willick, G.; Ni, F. Solution structure of the osteogenic 1-31 fragment of the human parathyroid hormone. *Biochemistry* **2000**, *39*, 12766–12777.

(171) Jin, L.; Briggs, S. L.; Chandrasekhar, S.; Chirgadze, N. Y.; Clawson, D. K.; Schevitz, R. W.; Smiley, D. L.; Tashjian, A. H.; Zhang, F. Crystal Structure of Human Parathyroid Hormone 1–34 at 0.9-Å Resolution. *Journal of Biological Chemistry* **2000**, *275*, 27238–27244.

(172) Micsonai, A.; Moussong, É.; Wien, F.; Boros, E.; Vadászi, H.; Murvai, N.; Lee, Y.-H.; Molnár, T.; Réfrégiers, M.; Goto, Y.; Tantos, Á.; Kardos, J. BeStSel: webserver for secondary structure and fold prediction for protein CD spectroscopy. *Nucleic acids research* **2022**, *50*, W90–W98.

(173) Nicolás-Carlock, J. R.; Carrillo-Estrada, J. L. A universal dimensionality function for the fractal dimensions of Laplacian growth. *Scientific reports* **2019**, *9*, 1120.

(174) Madine, J.; Pandya, M. J.; Hicks, M. R.; Rodger, A.; Yates, E. A.; Radford, S. E.; Middleton, D. A. Site-Specific Identification of an Ab Fibril-Heparin Interaction Site by Using Solid-State NMR Spectroscopy. *Angewandte Chemie* **2012**, *124*, 13317–13320.

(175) Cukalevski, R.; Yang, X.; Meisl, G.; Weininger, U.; Bernfur, K.; Frohm, B.; Knowles, T. P. J.; Linse, S. The Ab40 and Ab42 peptides self-assemble into separate homomolecular fibrils in binary mixtures but cross-react during primary nucleation. *Chemical science* **2015**, *6*, 4215–4233.

(176) Crespo, R.; Rocha, F. A.; Damas, A. M.; Martins, P. M. A generic crystallization-like model that describes the kinetics of amyloid fibril formation. *The Journal of biological chemistry* **2012**, *287*, 30585–30594.

(177) Lander, A. D. Targeting the glycosaminoglycan-binding sites on proteins. *Chemistry & Biology* **1994**, *1*, 73–78.

(178) Arad, E.; Green, H.; Jelinek, R.; Rapaport, H. Revisiting thioflavin T (ThT) fluorescence as a marker of protein fibrillation - The prominent role of electrostatic interactions. *Journal of colloid and interface science* **2020**, *573*, 87–95.

(179) Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M. Metastability of native proteins and the phenomenon of amyloid formation. *Journal of the American Chemical Society* **2011**, *133*, 14160–14163.

(180) Mideros-Mora, C.; Miguel-Romero, L.; Felipe-Ruiz, A.; Casino, P.; Marina, A. Revisiting the pH-gated conformational switch on the activities of HisKA-family histidine kinases. *Nature communications* **2020**, *11*, 769.

(181) Wu, W.; Celma, C. C.; Kerviel, A.; Roy, P. Mapping the pH Sensors Critical for Host Cell Entry by a Complex Nonenveloped Virus. *Journal of virology* **2019**, *93*, DOI: 10.1128/JVI.01897-18.

(182) Elimova, E.; Kisilevsky, R.; Ancsin, J. B. Heparan sulfate promotes the aggregation of HDL-associated serum amyloid A: evidence for a proamyloidogenic histidine molecular switch. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **2009**, *23*, 3436–3448.

(183) Cardin, A. D.; Weintraub, H. J. Molecular modeling of protein-glycosaminoglycan interactions. *Arteriosclerosis (Dallas, Tex.)* **1989**, *9*, 21–32.

(184) Muñoz, E. M.; Linhardt, R. J. Heparin-binding domains in vascular biology. *Arteriosclerosis, thrombosis, and vascular biology* **2004**, *24*, 1549–1557.

(185) Jha, N. N.; Anoop, A.; Ranganathan, S.; Mohite, G. M.; Padinhateeri, R.; Maji, S. K. Characterization of amyloid formation by glucagon-like peptides: role of basic residues in heparin-mediated aggregation. *Biochemistry* **2013**, *52*, 8800–8810.

(186) Gao, Q.; Yang, J.-Y.; Moremen, K. W.; Flanagan, J. G.; Prestegard, J. H. Structural Characterization of a Heparan Sulfate Pentamer Interacting with LAR-Ig1-2. *Biochemistry* **2018**, *57*, 2189–2199.

(187) Jackson, M. P.; Hewitt, E. W. Why are Functional Amyloids Non-Toxic in Humans? *Biomolecules* **2017**, *7*, DOI: 10.3390/biom7040071.

(188) Jacob, R. S.; Das, S.; Ghosh, S.; Anoop, A.; Jha, N. N.; Khan, T.; Singru, P.; Kumar, A.; Maji, S. K. Amyloid formation of growth hormone in presence of zinc: Relevance to its storage in secretory granules. *Scientific reports* **2016**, *6*, 23370.

(189) Anoop, A. et al. Elucidating the role of disulfide bond on amyloid formation and fibril reversibility of somatostatin-14: relevance to its storage and secretion. *The Journal of biological chemistry* **2014**, *289*, 16884–16903.

(190) Maji, S. K.; Schubert, D.; Rivier, C.; Lee, S.; Rivier, J. E.; Riek, R. Amyloid as a depot for the formulation of long-acting drugs. *PLoS biology* **2008**, *6*, e17.

(191) Zhao, L.-H. et al. Structure and dynamics of the active human parathyroid hormone receptor-1. *Science (New York, N.Y.)* **2019**, *364*, 148–153.

(192) Pioszak, A. A.; Xu, H. E. Molecular recognition of parathyroid hormone by its G protein-coupled receptor. *Proceedings of the National Academy of Sciences* **2008**, *105*, 5034–5039.

(193) Pioszak, A. A.; Parker, N. R.; Gardella, T. J.; Xu, H. E. Structural basis for parathyroid hormone-related protein binding to the parathyroid hormone receptor and design of conformation-selective peptides. *The Journal of biological chemistry* **2009**, *284*, 28382–28391.

(194) Chatterjee, D. et al. Co-aggregation and secondary nucleation in the life cycle of human prolactin/galanin functional amyloids. *eLife* **2022**, *11*, DOI: 10.7554/eLife.73835.

(195) Habener, J. F.; Amherdt, M.; Ravazzola, M.; Orci, L. Parathyroid hormone biosynthesis. Correlation of conversion of biosynthetic precursors with intracellular protein migration as determined by electron microscope autoradiography. *The Journal of cell biology* **1979**, *80*, 715–731.

(196) Muresan, Z.; MacGregor, R. R. The release of parathyroid hormone and the exocytosis of a proteoglycan are modulated by extracellular Ca2+ in a similar manner. *Molecular biology of the cell* **1994**, *5*, 725–737.

(197) Prydz, K. Determinants of Glycosaminoglycan (GAG) Structure. *Biomolecules* **2015**, *5*, 2003–2022.

(198) Gorr, S. U.; Hamilton, J. W.; Cohn, D. V. Sulfated secreted forms of bovine and porcine parathyroid chromogranin A (secretory protein-I). *Journal of Biological Chemistry* **1991**, *266*, 5780–5784.

(199) Roesch, P.; Marx, U. C., *STRUCTURE OF BOVINE PARATHYROID HORMONE FRAGMENT 1-37, NMR, 10 STRUCTURES*, 1997.

(200) Weidler, M.; Marx, U. C.; Seidel, G.; Roesch, P. The solution structure of human parathyroid hormone-related protein(1-34) in near physiological solution. **1999**, *444*, 239–244.

(201) Debiec, K. T.; Gronenborn, A. M.; Chong, L. T. Evaluating the strength of salt bridges: a comparison of current biomolecular force fields. *The journal of physical chemistry. B* **2014**, *118*, 6561–6569.

(202) Kasahara, K.; Terazawa, H.; Takahashi, T.; Higo, J. Studies on Molecular Dynamics of Intrinsically Disordered Proteins and Their Fuzzy Complexes: A Mini-Review. *Computational and structural biotechnology journal* **2019**, *17*, 712–720.

(203) Müller-Späth, S.; Soranno, A.; Hirschfeld, V.; Hofmann, H.; Rüegger, S.; Reymond, L.; Nettels, D.; Schuler, B. From the Cover: Charge interactions can dominate the dimensions of intrinsically disordered proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 14609–14614.

(204) Soranno, A.; Buchli, B.; Nettels, D.; Cheng, R. R.; Müller-Späth, S.; Pfeil, S. H.; Hoffmann, A.; Lipman, E. A.; Makarov, D. E.; Schuler, B. Quantifying internal friction in unfolded and intrinsically disordered proteins with single-molecule spectroscopy. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 17800–17806.

(205) Ziemianowicz, D. S.; Ng, D.; Schryvers, A. B.; Schriemer, D. C. Photo-Cross-Linking Mass Spectrometry and Integrative Modeling Enables Rapid Screening of Antigen Interactions Involving Bacterial Transferrin Receptors. *Journal of proteome research* **2019**, *18*, 934–946.

(206) Shillcock, J. C.; Hastings, J.; Riguet, N.; Lashuel, H. A. Non-monotonic fibril surface occlusion by GFP tags from coarse-grained molecular simulations. *Computational and structural biotechnology journal* **2022**, *20*, 309–321.

(207) Gallardo, J.; Escalona-Noguero, C.; Sot, B. Role of a-Synuclein Regions in Nucleation and Elongation of Amyloid Fiber Assembly. *ACS chemical neuroscience* **2020**, *11*, 872–879.

(208) El Mammeri, N.; Duan, P.; Dregni, A. J.; Hong, M. Amyloid fibril structures of tau: Conformational plasticity of the second microtubule-binding repeat. *Science advances* **2023**, *9*, eadh4731.

(209) Ngo, J. C. K.; Giang, K.; Chakrabarti, S.; Ma, C.-T.; Huynh, N.; Hagopian, J. C.; Dorrestein, P. C.; Fu, X.-D.; Adams, J. A.; Ghosh, G. A sliding docking interaction is essential for sequential and processive phosphorylation of an SR protein by SRPK1. *Molecular Cell* **2008**, *29*, 563–576.

(210) Lai, M. C.; Lin, R. I.; Huang, S. Y.; Tsai, C. W.; Tarn, W. Y. A human importin-beta family protein, transportin-SR2, interacts with the phosphorylated RS domain of SR proteins. *Journal of Biological Chemistry* **2000**, *275*, 7950–7957.

(211) Cazalla, D.; Zhu, J.; Manche, L.; Huber, E.; Krainer, A. R.; Cáceres, J. F. Nuclear export and retention signals in the RS domain of SR proteins. *Molecular and cellular biology* **2002**, *22*, 6871–6882.

(212) Tripathi, V.; Ellis, J. D.; Shen, Z.; Song, D. Y.; Pan, Q.; Watt, A. T.; Freier, S. M.; Bennett, C. F.; Sharma, A.; Bubulya, P. A.; Blencowe, B. J.; Prasanth, S. G.; Prasanth, K. V. The nuclear-retained noncoding RNA MALAT1 regulates alternative splicing by modulating SR splicing factor phosphorylation. *Molecular Cell* **2010**, *39*, 925–938.

(213) Aubol, B. E.; Plocinik, R. M.; Hagopian, J. C.; Ma, C.-T.; McGlone, M. L.; Bandyopadhyay, R.; Fu, X.-D.; Adams, J. A. Partitioning RS domain phosphorylation in an SR protein through the CLK and SRPK protein kinases. *Journal of Molecular Biology* **2013**, *425*, 2894–2909.

(214) Aubol, B. E.; Wozniak, J. M.; Fattet, L.; Gonzalez, D. J.; Adams, J. A. CLK1 reorganizes the splicing factor U1-70K for early spliceosomal protein assembly. *Proceedings of the National Academy of Sciences of the United States of America* **2021**, *118*, DOI: 10.1073/pnas.2018251118.

(215) Hagopian, J. C.; Ma, C.-T.; Meade, B. R.; Albuquerque, C. P.; Ngo, J. C. K.; Ghosh, G.; Jennings, P. A.; Fu, X.-D.; Adams, J. A. Adaptable molecular interactions guide phosphorylation of the SR protein ASF/SF2 by SRPK1. *Journal of Molecular Biology* **2008**, *382*, 894–909.

(216) George, A.; Aubol, B. E.; Fattet, L.; Adams, J. A. Disordered protein interactions for an ordered cellular transition: Cdc2-like kinase 1 is transported to the nucleus via its Ser-Arg protein substrate. *The Journal of biological chemistry* **2019**, *294*, 9631–9641.

(217) Liu, H.; Gong, Z.; Li, K.; Zhang, Q.; Xu, Z.; Xu, Y. SRPK1/2 and PP1a exert opposite functions by modulating SRSF1-guided MKNK2 alternative splicing in colon adenocarcinoma. *Journal of experimental & clinical cancer research : CR* **2021**, *40*, 75.

(218) Da Gruijs Silva, L. A.; Simonetti, F.; Hutten, S.; Riemenschneider, H.; Sternburg, E. L.; Pietrek, L. M.; Gebel, J.; Dötsch, V.; Edbauer, D.; Hummer, G.; Stelzl, L. S.; Dormann, D. Disease-linked TDP-43 hyperphosphorylation suppresses TDP-43 condensation and aggregation. *The EMBO Journal* **2022**, *41*, e108443.

(219) Ilık, İ. A.; Aktaş, T. Nuclear speckles: dynamic hubs of gene expression regulation. *The FEBS journal* **2021**, DOI: 10.1111/febs.16117.

(220) Keshwani, M. M.; Hailey, K. L.; Aubol, B. E.; Fattet, L.; McGlone, M. L.; Jennings, P. A.; Adams, J. A. Nuclear protein kinase CLK1 uses a non-traditional docking mechanism to select physiological substrates. *Biochemical Journal* **2015**, *472*, 329–338.

(221) Aubol, B. E.; Chakrabarti, S.; Ngo, J.; Shaffer, J.; Nolen, B.; Fu, X.-D.; Ghosh, G.; Adams, J. A. Processive phosphorylation of alternative splicing factor/splicing factor 2. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 12601–12606.

(222) Dekel, N.; Eisenberg-Domovich, Y.; Karlas, A.; Meyer, T. F.; Bracher, F.; Lebendiker, M.; Danieli, T.; Livnah, O. Expression, purification and crystallization of CLK1 kinase - A potential target for antiviral therapy. *Protein expression and purification* **2020**, *176*, 105742.

(223) Chang, L.-W.; Lytle, T. K.; Radhakrishna, M.; Madinya, J. J.; Vélez, J.; Sing, C. E.; Perry, S. L. Sequence and entropy-based control of complex coacervates. *Nature communications* **2017**, *8*, 1273.

(224) Xiang, S.; Gapsys, V.; Kim, H.-Y.; Bessonov, S.; Hsiao, H.-H.; Möhlmann, S.; Klaukien, V.; Ficner, R.; Becker, S.; Urlaub, H.; Lührmann, R.; de Groot, B.; Zweckstetter, M. Phosphorylation drives a dynamic switch in serine/arginine-rich proteins. *Structure (London, England : 1993)* **2013**, *21*, 2162–2174.

(225) Mandell, D. J.; Chorny, I.; Groban, E. S.; Wong, S. E.; Levine, E.; Rapp, C. S.; Jacobson, M. P. Strengths of hydrogen bonds involving phosphorylated amino acid side chains. *Journal of the American Chemical Society* **2007**, *129*, 820–827.

(226) Lenton, S.; Hervø-Hansen, S.; Popov, A. M.; Tully, M. D.; Lund, M.; Skepö, M. Impact of Arginine-Phosphate Interactions on the Reentrant Condensation of Disordered Proteins. *Biomacromolecules* **2021**, *22*, 1532–1544.

(227) Zhang, F.; Weggler, S.; Ziller, M. J.; Ianeselli, L.; Heck, B. S.; Hildebrandt, A.; Kohlbacher, O.; Skoda, M. W. A.; Jacobs, R. M. J.; Schreiber, F. Universality of protein reentrant condensation in solution induced by multivalent metal ions. *Proteins* **2010**, *78*, 3450–3457.

(228) Woods, A. S.; Ferré, S. Amazing stability of the arginine-phosphate electrostatic interaction. *Journal of proteome research* **2005**, *4*, 1397–1402.

(229) Tripathi, V.; Song, D. Y.; Zong, X.; Shevtsov, S. P.; Hearn, S.; Fu, X.-D.; Dundr, M.; Prasanth, K. V. SRSF1 regulates the assembly of pre-mRNA processing factors in nuclear speckles. *Molecular biology of the cell* **2012**, *23*, 3694–3706.

(230) Pepi, L. E.; Sanderson, P.; Stickney, M.; Amster, I. J. Developments in Mass Spectrometry for Glycosaminoglycan Analysis: A Review. *Molecular & cellular proteomics : MCP* **2021**, *20*, 100025.

(231) Ryšlavá, H.; Doubnerová, V.; Kavan, D.; Vaněk, O. Effect of posttranslational modifications on enzyme function and assembly. *Journal of proteomics* **2013**, *92*, 80–109.

(232) Zeng, L.-R.; Vega-Sánchez, M. E.; Zhu, T.; Wang, G.-L. Ubiquitination-mediated protein degradation and modification: an emerging theme in plant-microbe interactions. *Cell research* **2006**, *16*, 413–426.

(233) Kundinger, S. R.; Dammer, E. B.; Yin, L.; Hurst, C.; Shapley, S.; Ping, L.; Khoshnevis, S.; Ghalei, H.; Duong, D. M.; Seyfried, N. T. Phosphorylation regulates arginine-rich RNA-binding protein solubility and oligomerization. *The Journal of biological chemistry* **2021**, *297*, 101306.

(234) Haltiwanger, R. S.; Lowe, J. B. Role of glycosylation in development. *Annual review of biochemistry* **2004**, *73*, 491–537.

(235) Del Monte, F.; Agnetti, G. Protein post-translational modifications and misfolding: new concepts in heart failure. *Proteomics. Clinical applications* **2014**, *8*, 534–542.

(236) Marshall, C. J. Protein prenylation: a mediator of protein-protein interactions. *Science (New York, N.Y.)* **1993**, *259*, 1865–1866.

(237) Rocks, O.; Peyker, A.; Kahms, M.; Verveer, P. J.; Koerner, C.; Lumbierres, M.; Kuhlmann, J.; Waldmann, H.; Wittinghofer, A.; Bastiaens, P. I. H. An acylation cycle regulates localization and activity of palmitoylated Ras isoforms. *Science (New York, N.Y.)* **2005**, *307*, 1746–1752.

(238) Wang, K.; Wang, L.; Wang, J.; Chen, S.; Shi, M.; Cheng, H. Intronless mRNAs transit through nuclear speckles to gain export competence. *The Journal of cell biology* **2018**, *217*, 3912–3929.

VI Appendix

DNA and amino acid sequences

DNA sequences encoding specific proteins were sequenced by Eurofins Genomics. In the following, the DNA and amino acid sequences of constructs generated in this work are listed.

SRSF1 constructs

SRSF1 FL with an N-terminal SBP-FLAG-RFP-tag (in pcDNA3.1(+)).

```
atg gac gag aag acc acc ggc tgg cgc ggc ggc cac gtg gtg gag ggc ctg gcc ggc gag
 M   D   E   K   T   T   G   W   R   G   G   H   V   V   E   G   L   A   G   E
ctg gag cag ctg cgc gcc cgc ctg gag cac cac ccc cag ggc cag cgc gag ccc aag ctt
 L   E   Q   L   R   A   R   L   E   H   H   P   Q   G   Q   R   E   P   K   L
ggt acc gag ctc gga tct atg gac tac aag gac gac gac gac aag gga tct atg gcc tcc
 G   T   E   L   G   S   M   D   Y   K   D   D   D   D   K   G   S   M   A   S
tcc gag gac gtc atc aag gag ttc atg cgc ttc aag gtg cgc atg gag ggc tcc gtg aac
 S   E   D   V   I   K   E   F   M   R   F   K   V   R   M   E   G   S   V   N
ggc cac gag ttc gag atc gag ggc gag ggc gag ggc cgc ccc tac gag ggc acc cag acc
 G   H   E   F   E   I   E   G   E   G   E   G   R   P   Y   E   G   T   Q   T
gcc aag ctg aag gtg acc aag ggc ggc ccc ctg ccc ttc gcc tgg gac atc ctg tcc cct
 A   K   L   K   V   T   K   G   G   P   L   P   F   A   W   D   I   L   S   P
cag ttc cag tac ggc tcc aag gcc tac gtg aag cac ccc gcc gac atc ccc gac tac ttg
 Q   F   Q   Y   G   S   K   A   Y   V   K   H   P   A   D   I   P   D   Y   L
aag ctg tcc ttc ccc gag ggc ttc aag tgg gag cgc gtg atg aac ttc gag gac ggc ggc
 K   L   S   F   P   E   G   F   K   W   E   R   V   M   N   F   E   D   G   G
gtg gtg acc gtg acc cag gac tcc tcc ctg cag gac ggc gag ttc atc tac aag gtg aag
 V   V   T   V   T   Q   D   S   S   L   Q   D   G   E   F   I   Y   K   V   K
ctg cgc ggc acc aac ttc ccc tcc gac ggc ccc gta atg cag aag aag acc atg ggc tgg
 L   R   G   T   N   F   P   S   D   G   P   V   M   Q   K   K   T   M   G   W
gag gcc tcc acc gag cgg atg tac ccc gag gac ggc gcc ctg aag ggc gag atc aag atg
 E   A   S   T   E   R   M   Y   P   E   D   G   A   L   K   G   E   I   K   M
agg ctg aag ctg aag gac ggc ggc cac tac gac gcc gag gtc aag acc acc tac atg gcc
 R   L   K   L   K   D   G   G   H   Y   D   A   E   V   K   T   T   Y   M   A
aag aag ccc gtg cag ctg ccc ggc gcc tac aag acc gac atc aag ctg gac atc acc tcc
 K   K   P   V   Q   L   P   G   A   Y   K   T   D   I   K   L   D   I   T   S
cac aac gag gac tac acc atc gtg gaa cag tac gag cgc gcc gag ggc cgc cac tcc acc
 H   N   E   D   Y   T   I   V   E   Q   Y   E   R   A   E   G   R   H   S   T
ggc gcc gga tcc gga gaa ttc atg tcg gga ggt ggt gtg atc cgt ggc ccg gcg ggg aac
 G   A   G   S   G   E   F   M   S   G   G   G   V   I   R   G   P   A   G   N
aac gac tgc cgc atc tac gtg ggt aac cta cct ccg gat atc cga acc aag gac atc gag
 N   D   C   R   I   Y   V   G   N   L   P   P   D   I   R   T   K   D   I   E
gac gtg ttt tac aaa tac ggc gcc atc cgc gac atc gac ctg aag aac cgc cgc ggg gga
 D   V   F   Y   K   Y   G   A   I   R   D   I   D   L   K   N   R   R   G   G
ccg ccc ttc gcc ttc gtt gag ttc gag gac ccg cga gac gcg gaa gat gcg gtg tac ggt
 P   P   F   A   F   V   E   F   E   D   P   R   D   A   E   D   A   V   Y   G
cgc gac ggc tac gac tac gac ggc tac cgg ctg cgg gta gag ttt ccc cga agc ggc cgc
 R   D   G   Y   D   Y   D   G   Y   R   L   R   V   E   F   P   R   S   G   R
ggg acc ggc cga ggc ggc ggc ggg ggt gga ggc ggc ggc gcc ccg aga ggc cgc tat ggc
 G   T   G   R   G   G   G   G   G   G   G   G   G   A   P   R   G   R   Y   G
ccg ccg tcc agg cgg tcc gag aac aga gtg gtt gtc tct gga ctg cct ccg agt gga agc
 P   P   S   R   R   S   E   N   R   V   V   V   S   G   L   P   P   S   G   S
tgg cag gac tta aag gat cac atg cgt gag gca ggt gat gta tgt tac gct gat gtt tac
 W   Q   D   L   K   D   H   M   R   E   A   G   D   V   C   Y   A   D   V   Y
cga gat ggc act ggt gtc gtg gag ttt gta cgg aaa gaa gat atg acg tat gca gtt cga
 R   D   G   T   G   V   V   E   F   V   R   K   E   D   M   T   Y   A   V   R
```

```
aaa ctg gat aac act aag ttt aga tct cac gag gga gaa act gcc tac atc cgg gtt aaa
 K   L   D   N   T   K   F   R   S   H   E   G   E   T   A   Y   I   R   V   K
gtt gat ggg ccc aga agt cca agt tat gga aga tct cga tct cga agc cgt agt cgt agc
 V   D   G   P   R   S   P   S   Y   G   R   S   R   S   R   S   R   S   R   S
aga agc cgt agc aga agc aac agc agg agt cgc agt tac tcc cca agg aga agc aga gga
 R   S   R   S   R   S   N   S   R   S   R   S   Y   S   P   R   R   S   R   G
tca cca cgc tat tct ccc cgt cat agc aga tct cgc tct cgt aca taa
 S   P   R   Y   S   P   R   H   S   R   S   R   S   R   T   -
```

SRSF1 FL_{Hypo-D} with an N-terminal SBP-FLAG-RFP-tag (in pcDNA3.1(+)).

```
atg gac gag aag acc acc ggc tgg cgc ggc ggc cac gtg gtg gag ggc ctg gcc ggc gag
 M   D   E   K   T   T   G   W   R   G   G   H   V   V   E   G   L   A   G   E
ctg gag cag ctg cgc gcc cgc ctg gag cac cac ccc cag ggc cag cgc gag ccc aag ctt
 L   E   Q   L   R   A   R   L   E   H   H   P   Q   G   Q   R   E   P   K   L
ggt acc gag ctc gga tct atg gac tac aag gac gac gac gac aag gga tct atg gcc tcc
 G   T   E   L   G   S   M   D   Y   K   D   D   D   D   K   G   S   M   A   S
tcc gag gac gtc atc aag gag ttc atg cgc ttc aag gtg cgc atg gag ggc tcc gtg aac
 S   E   D   V   I   K   E   F   M   R   F   K   V   R   M   E   G   S   V   N
ggc cac gag ttc gag atc gag ggc gag ggc gag ggc cgc ccc tac gag ggc acc cag acc
 G   H   E   F   E   I   E   G   E   G   E   G   R   P   Y   E   G   T   Q   T
gcc aag ctg aag gtg acc aag ggc ggc ccc ctg ccc ttc gcc tgg gac atc ctg tcc cct
 A   K   L   K   V   T   K   G   G   P   L   P   F   A   W   D   I   L   S   P
cag ttc cag tac ggc tcc aag gcc tac gtg aag cac ccc gcc gac atc ccc gac tac ttg
 Q   F   Q   Y   G   S   K   A   Y   V   K   H   P   A   D   I   P   D   Y   L
aag ctg tcc ttc ccc gag ggc ttc aag tgg gag cgc gtg atg aac ttc gag gac ggc ggc
 K   L   S   F   P   E   G   F   K   W   E   R   V   M   N   F   E   D   G   G
gtg gtg acc gtg acc cag gac tcc tcc ctg cag gac ggc gag ttc atc tac aag gtg aag
 V   V   T   V   T   Q   D   S   S   L   Q   D   G   E   F   I   Y   K   V   K
ctg cgc ggc acc aac ttc ccc tcc gac ggc ccc gta atg cag aag aag acc atg ggc tgg
 L   R   G   T   N   F   P   S   D   G   P   V   M   Q   K   K   T   M   G   W
gag gcc tcc acc gag cgg atg tac ccc gag gac ggc gcc ctg aag ggc gag atc aag atg
 E   A   S   T   E   R   M   Y   P   E   D   G   A   L   K   G   E   I   K   M
agg ctg aag ctg aag gac ggc ggc cac tac gac gcc gag gtc aag acc acc tac atg gcc
 R   L   K   L   K   D   G   G   H   Y   D   A   E   V   K   T   T   Y   M   A
aag aag ccc gtg cag ctg ccc ggc gcc tac aag acc gac atc aag ctg gac atc acc tcc
 K   K   P   V   Q   L   P   G   A   Y   K   T   D   I   K   L   D   I   T   S
cac aac gag gac tac acc atc gtg gaa cag tac gag cgc gcc gag ggc cgc cac tcc acc
 H   N   E   D   Y   T   I   V   E   Q   Y   E   R   A   E   G   R   H   S   T
ggc gcc gga tcc gga gaa ttc atg tcg gga ggt ggt gtg atc cgt ggc ccg gcg ggg aac
 G   A   G   S   G   E   F   M   S   G   G   G   V   I   R   G   P   A   G   N
aac gac tgc cgc atc tac gtg ggt aac cta cct ccg gat atc cga acc aag gac atc gag
 N   D   C   R   I   Y   V   G   N   L   P   P   D   I   R   T   K   D   I   E
gac gtg ttt tac aaa tac ggc gcc atc cgc gac atc gac ctg aag aac cgc cgc ggg gga
 D   V   F   Y   K   Y   G   A   I   R   D   I   D   L   K   N   R   R   G   G
ccg ccc ttc gcc ttc gtt gag ttc gag gac ccg cga gac gcg gaa gat gcg gtg tac ggt
 P   P   F   A   F   V   E   F   E   D   P   R   D   A   E   D   A   V   Y   G
cgc gac ggc tac gac tac gac ggc tac cgg ctg cgg gta gag ttt ccc cga agc ggc cgc
 R   D   G   Y   D   Y   D   G   Y   R   L   R   V   E   F   P   R   S   G   R
ggg acc ggc cga ggc ggc ggc ggg ggt gga ggc ggc ggc gcc ccg aga ggc cgc tat ggc
 G   T   G   R   G   G   G   G   G   G   G   G   G   A   P   R   G   R   Y   G
ccg ccg tcc agg cgg tcc gag aac aga gtg gtt gtc tct gga ctg cct ccg agt gga agc
 P   P   S   R   R   S   E   N   R   V   V   V   S   G   L   P   P   S   G   S
tgg cag gac tta aag gat cac atg cgt gag gca ggt gat gta tgt tac gct gat gtt tac
 W   Q   D   L   K   D   H   M   R   E   A   G   D   V   C   Y   A   D   V   Y
cga gat ggc act ggt gtc gtg gag ttt gta cgg aaa gaa gat atg acg tat gca gtt cga
 R   D   G   T   G   V   V   E   F   V   R   K   E   D   M   T   Y   A   V   R
aaa ctg gat aac act aag ttt aga tct cac gag gga gaa act gcc tac atc cgg gtt aaa
```

```
 K   L   D   N   T   K   F   R   S   H   E   G   E   T   A   Y   I   R   V   K
gtt gat ggg ccc aga agt cca agt tat gga aga tct cga tct cga agc cgt agt cgt agc
 V   D   G   P   R   S   P   S   Y   G   R   S   R   S   R   S   R   S   R   S
aga agc cgt agc aga agc aac agc agg agt cgc agt tac tcc cca agg aga agc aga gga
 R   S   R   S   R   S   N   S   R   S   R   S   Y   S   P   R   R   S   R   G
tca cca cgc tat tct ccc cgt cat agc aga tct cgc tct cgt aca taa
 S   P   R   Y   S   P   R   H   S   R   S   R   S   R   T   -
```

SRSF1 LCD with an N-terminal SBP-FLAG-RFP-tag (in pcDNA3.1(+)).

```
atg gac gag aag acc acc ggc tgg cgc ggc ggc cac gtg gtg gag ggc ctg gcc ggc gag
 M   D   E   K   T   T   G   W   R   G   G   H   V   V   E   G   L   A   G   E
ctg gag cag ctg cgc gcc cgc ctg gag cac cac ccc cag ggc cag cgc gag ccc aag ctt
 L   E   Q   L   R   A   R   L   E   H   H   P   Q   G   Q   R   E   P   K   L
ggt acc gag ctc gga tct atg gac tac aag gac gac gac gac aag gga tct atg gcc tcc
 G   T   E   L   G   S   M   D   Y   K   D   D   D   D   K   G   S   M   A   S
tcc gag gac gtc atc aag gag ttc atg cgc ttc aag gtg cgc atg gag ggc tcc gtg aac
 S   E   D   V   I   K   E   F   M   R   F   K   V   R   M   E   G   S   V   N
ggc cac gag ttc gag atc gag ggc gag ggc gag ggc cgc ccc tac gag ggc acc cag acc
 G   H   E   F   E   I   E   G   E   G   E   G   R   P   Y   E   G   T   Q   T
gcc aag ctg aag gtg acc aag ggc ggc ccc ctg ccc ttc gcc tgg gac atc ctg tcc cct
 A   K   L   K   V   T   K   G   G   P   L   P   F   A   W   D   I   L   S   P
cag ttc cag tac ggc tcc aag gcc tac gtg aag cac ccc gcc gac atc ccc gac tac ttg
 Q   F   Q   Y   G   S   K   A   Y   V   K   H   P   A   D   I   P   D   Y   L
aag ctg tcc ttc ccc gag ggc ttc aag tgg gag cgc gtg atg aac ttc gag gac ggc ggc
 K   L   S   F   P   E   G   F   K   W   E   R   V   M   N   F   E   D   G   G
gtg gtg acc gtg acc cag gac tcc tcc ctg cag gac ggc gag ttc atc tac aag gtg aag
 V   V   T   V   T   Q   D   S   S   L   Q   D   G   E   F   I   Y   K   V   K
ctg cgc ggc acc aac ttc ccc tcc gac ggc ccc gta atg cag aag aag acc atg ggc tgg
 L   R   G   T   N   F   P   S   D   G   P   V   M   Q   K   K   T   M   G   W
gag gcc tcc acc gag cgg atg tac ccc gag gac ggc gcc ctg aag ggc gag atc aag atg
 E   A   S   T   E   R   M   Y   P   E   D   G   A   L   K   G   E   I   K   M
agg ctg aag ctg aag gac ggc ggc cac tac gac gcc gag gtc aag acc acc tac atg gcc
 R   L   K   L   K   D   G   G   H   Y   D   A   E   V   K   T   T   Y   M   A
aag aag ccc gtg cag ctg ccc ggc gcc tac aag acc gac atc aag ctg gac atc acc tcc
 K   K   P   V   Q   L   P   G   A   Y   K   T   D   I   K   L   D   I   T   S
cac aac gag gac tac acc atc gtg gaa cag tac gag cgc gcc gag ggc cgc cac tcc acc
 H   N   E   D   Y   T   I   V   E   Q   Y   E   R   A   E   G   R   H   S   T
ggc gcc gga tcc gga gaa ttc gag gga gaa act gcc tac atc cgg gtt aaa gtt gat ggg
 G   A   G   S   G   E   F   E   G   E   T   A   Y   I   R   V   K   V   D   G
ccc aga agt cca agt tat gga aga tct cga tct cga agc cgt agt cgt agc aga agc cgt
 P   R   S   P   S   Y   G   R   S   R   S   R   S   R   S   R   S   R   S   R
agc aga agc aac agc agg agt cgc agt tac tcc cca agg aga agc aga gga tca cca cgc
 S   R   S   N   S   R   S   R   S   Y   S   P   R   R   S   R   G   S   P   R
tat tct ccc cgt cat agc aga tct cgc tct cgt aca taa
 Y   S   P   R   H   S   R   S   R   S   R   T   -
```

SRSF1 LCD$_{Hypo-D}$ with an N-terminal SBP-FLAG-RFP-tag (in pcDNA3.1(+)).

```
atg gac gag aag acc acc ggc tgg cgc ggc ggc cac gtg gtg gag ggc ctg gcc ggc gag
 M   D   E   K   T   T   G   W   R   G   G   H   V   V   E   G   L   A   G   E
ctg gag cag ctg cgc gcc cgc ctg gag cac cac ccc cag ggc cag cgc gag ccc aag ctt
 L   E   Q   L   R   A   R   L   E   H   H   P   Q   G   Q   R   E   P   K   L
ggt acc gag ctc gga tct atg gac tac aag gac gac gac gac aag gga tct atg gcc tcc
 G   T   E   L   G   S   M   D   Y   K   D   D   D   D   K   G   S   M   A   S
tcc gag gac gtc atc aag gag ttc atg cgc ttc aag gtg cgc atg gag ggc tcc gtg aac
 S   E   D   V   I   K   E   F   M   R   F   K   V   R   M   E   G   S   V   N
ggc cac gag ttc gag atc gag ggc gag ggc gag ggc cgc ccc tac gag ggc acc cag acc
 G   H   E   F   E   I   E   G   E   G   E   G   R   P   Y   E   G   T   Q   T
```

```
gcc aag ctg aag gtg acc aag ggc ggc ccc ctg ccc ttc gcc tgg gac atc ctg tcc cct
 A   K   L   K   V   T   K   G   G   P   L   P   F   A   W   D   I   L   S   P
cag ttc cag tac ggc tcc aag gcc tac gtg aag cac ccc gcc gac atc ccc gac tac ttg
 Q   F   Q   Y   G   S   K   A   Y   V   K   H   P   A   D   I   P   D   Y   L
aag ctg tcc ttc ccc gag ggc ttc aag tgg gag cgc gtg atg aac ttc gag gac ggc ggc
 K   L   S   F   P   E   G   F   K   W   E   R   V   M   N   F   E   D   G   G
gtg gtg acc gtg acc cag gac tcc tcc ctg cag gac ggc gag ttc atc tac aag gtg aag
 V   V   T   V   T   Q   D   S   S   L   Q   D   G   E   F   I   Y   K   V   K
ctg cgc ggc acc aac ttc ccc tcc gac ggc ccc gta atg cag aag aag acc atg ggc tgg
 L   R   G   T   N   F   P   S   D   G   P   V   M   Q   K   K   T   M   G   W
gag gcc tcc acc gag cgg atg tac ccc gag gac ggc gcc ctg aag ggc gag atc aag atg
 E   A   S   T   E   R   M   Y   P   E   D   G   A   L   K   G   E   I   K   M
agg ctg aag ctg aag gac ggc ggc cac tac gac gcc gag gtc aag acc acc tac atg gcc
 R   L   K   L   K   D   G   G   H   Y   D   A   E   V   K   T   T   Y   M   A
aag aag ccc gtg cag ctg ccc ggc gcc tac aag acc gac atc aag ctg gac atc acc tcc
 K   K   P   V   Q   L   P   G   A   Y   K   T   D   I   K   L   D   I   T   S
cac aac gag gac tac acc atc gtg gaa cag tac gag cgc gcc gag ggc cgc cac tcc acc
 H   N   E   D   Y   T   I   V   E   Q   Y   E   R   A   E   G   R   H   S   T
ggc gcc gga tcc gga gaa ttc gag gga gaa act gcc tac atc cgg gtt aaa gtt gat ggg
 G   A   G   S   G   E   F   E   G   E   T   A   Y   I   R   V   K   V   D   G
ccc aga agt cca agt tat gga aga gat cga gat cga gat cgt gat cgt gat aga gat cgt
 P   R   S   P   S   Y   G   R   D   R   D   R   D   R   D   R   D   R   D   R
gat aga gat aac gat agg gat cgc agt tac tcc cca agg aga agc aga gga tca cca cgc
 D   R   D   N   D   R   D   R   S   Y   S   P   R   R   S   R   G   S   P   R
tat tct ccc cgt cat agc aga tct cgc tct cgt aca taa
 Y   S   P   R   H   S   R   S   R   S   R   T   -
```

SRSF1 LCD$_{Hyper-D}$ with an N-terminal SBP-FLAG-RFP-tag (in pcDNA3.1(+)).

```
atg gac gag aag acc acc ggc tgg cgc ggc ggc cac gtg gtg gag ggc ctg gcc ggc gag
 M   D   E   K   T   T   G   W   R   G   G   H   V   V   E   G   L   A   G   E
ctg gag cag ctg cgc gcc cgc ctg gag cac cac ccc cag ggc cag cgc gag ccc aag ctt
 L   E   Q   L   R   A   R   L   E   H   H   P   Q   G   Q   R   E   P   K   L
ggt acc gag ctc gga tct atg gac tac aag gac gac gac gac aag gga tct atg gcc tcc
 G   T   E   L   G   S   M   D   Y   K   D   D   D   D   K   G   S   M   A   S
tcc gag gac gtc atc aag gag ttc atg cgc ttc aag gtg cgc atg gag ggc tcc gtg aac
 S   E   D   V   I   K   E   F   M   R   F   K   V   R   M   E   G   S   V   N
ggc cac gag ttc gag atc gag ggc gag ggc gag ggc cgc ccc tac gag ggc acc cag acc
 G   H   E   F   E   I   E   G   E   G   E   G   R   P   Y   E   G   T   Q   T
gcc aag ctg aag gtg acc aag ggc ggc ccc ctg ccc ttc gcc tgg gac atc ctg tcc cct
 A   K   L   K   V   T   K   G   G   P   L   P   F   A   W   D   I   L   S   P
cag ttc cag tac ggc tcc aag gcc tac gtg aag cac ccc gcc gac atc ccc gac tac ttg
 Q   F   Q   Y   G   S   K   A   Y   V   K   H   P   A   D   I   P   D   Y   L
aag ctg tcc ttc ccc gag ggc ttc aag tgg gag cgc gtg atg aac ttc gag gac ggc ggc
 K   L   S   F   P   E   G   F   K   W   E   R   V   M   N   F   E   D   G   G
gtg gtg acc gtg acc cag gac tcc tcc ctg cag gac ggc gag ttc atc tac aag gtg aag
 V   V   T   V   T   Q   D   S   S   L   Q   D   G   E   F   I   Y   K   V   K
ctg cgc ggc acc aac ttc ccc tcc gac ggc ccc gta atg cag aag aag acc atg ggc tgg
 L   R   G   T   N   F   P   S   D   G   P   V   M   Q   K   K   T   M   G   W
gag gcc tcc acc gag cgg atg tac ccc gag gac ggc gcc ctg aag ggc gag atc aag atg
 E   A   S   T   E   R   M   Y   P   E   D   G   A   L   K   G   E   I   K   M
agg ctg aag ctg aag gac ggc ggc cac tac gac gcc gag gtc aag acc acc tac atg gcc
 R   L   K   L   K   D   G   G   H   Y   D   A   E   V   K   T   T   Y   M   A
aag aag ccc gtg cag ctg ccc ggc gcc tac aag acc gac atc aag ctg gac atc acc tcc
 K   K   P   V   Q   L   P   G   A   Y   K   T   D   I   K   L   D   I   T   S
cac aac gag gac tac acc atc gtg gaa cag tac gag cgc gcc gag ggc cgc cac tcc acc
 H   N   E   D   Y   T   I   V   E   Q   Y   E   R   A   E   G   R   H   S   T
ggc gcc gga tcc gga gaa ttc gag gga gaa act gcc tac atc cgg gtt aaa gtt gat ggg
 G   A   G   S   G   E   F   E   G   E   T   A   Y   I   R   V   K   V   D   G
```

```
ccc aga agt cca agt tat gga aga gat cga gat cga gat cgt gat cgt gat aga gat cgt
 P   R   S   P   S   Y   G   R   D   R   D   R   D   R   D   R   D   R   D   R
gat aga gat aac gat agg gat cgc agt tac gat cca agg aga agc aga gga gat cca cgc
 D   R   D   N   D   R   D   R   S   Y   D   P   R   R   S   R   G   D   P   R
tat gat ccc cgt cat agc aga tct cgc tct cgt aca taa
 Y   D   P   R   H   S   R   S   R   S   R   T   -
```

SRSF1 FL with an N-terminal His-H3C-tag (in pET41a(+)).

```
atg ggc cat cat cat cat cat cac ctg gaa gtt ctg ttc cag ggg ccc gga tcc ccg gaa
 M   G   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   G   S   P   E
ttc atg tcg gga ggt ggt gtg atc cgt ggc ccg gcg ggg aac aac gac tgc cgc atc tac
 F   M   S   G   G   G   V   I   R   G   P   A   G   N   N   D   C   R   I   Y
gtg ggt aac cta cct ccg gat atc cga acc aag gac atc gag gac gtg ttt tac aaa tac
 V   G   N   L   P   P   D   I   R   T   K   D   I   E   D   V   F   Y   K   Y
ggc gcc atc cgc gac atc gac ctg aag aac cgc cgc ggg gga ccg ccc ttc gcc ttc gtt
 G   A   I   R   D   I   D   L   K   N   R   R   G   G   P   P   F   A   F   V
gag ttc gag gac ccg cga gac gcg gaa gat gcg gtg tac ggt cgc gac ggc tac gac tac
 E   F   E   D   P   R   D   A   E   D   A   V   Y   G   R   D   G   Y   D   Y
gac ggc tac cgg ctg cgg gta gag ttt ccc cga agc ggc cgc ggg acc ggc cga ggc ggc
 D   G   Y   R   L   R   V   E   F   P   R   S   G   R   G   T   G   R   G   G
ggc ggg ggt gga ggc ggc ggc gcc ccg aga ggc cgc tat ggc ccg ccg tcc agg cgg tcc
 G   G   G   G   G   G   G   A   P   R   G   R   Y   G   P   P   S   R   R   S
gag aac aga gtg gtt gtc tct gga ctg cct ccg agt gga agc tgg cag gac tta aag gat
 E   N   R   V   V   V   S   G   L   P   P   S   G   S   W   Q   D   L   K   D
cac atg cgt gag gca ggt gat gta tgt tac gct gat gtt tac cga gat ggc act ggt gtc
 H   M   R   E   A   G   D   V   C   Y   A   D   V   Y   R   D   G   T   G   V
gtg gag ttt gta cgg aaa gaa gat atg acg tat gca gtt cga aaa ctg gat aac act aag
 V   E   F   V   R   K   E   D   M   T   Y   A   V   R   K   L   D   N   T   K
ttt aga tct cac gag gga gaa act gcc tac atc cgg gtt aaa gtt gat ggg ccc aga agt
 F   R   S   H   E   G   E   T   A   Y   I   R   V   K   V   D   G   P   R   S
cca agt tat gga aga tct cga tct cga agc cgt agt cgt agc aga agc cgt agc aga agc
 P   S   Y   G   R   S   R   S   R   S   R   S   R   S   R   S   R   S   R   S
aac agc agg agt cgc agt tac tcc cca agg aga agc aga gga tca cca cgc tat tct ccc
 N   S   R   S   R   S   Y   S   P   R   R   S   R   G   S   P   R   Y   S   P
cgt cat agc aga tct cgc tct cgt aca taa
 R   H   S   R   S   R   S   R   T   -
```

SRSF1 FL_{Hypo-D} with an N-terminal His-H3C-tag (in pET41a(+)).

```
atg ggc cat cat cat cat cat cac ctg gaa gtt ctg ttc cag ggg ccc gga tcc ccg gaa
 M   G   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   G   S   P   E
ttc atg tcg gga ggt ggt gtg atc cgt ggc ccg gcg ggg aac aac gac tgc cgc atc tac
 F   M   S   G   G   G   V   I   R   G   P   A   G   N   N   D   C   R   I   Y
gtg ggt aac cta cct ccg gat atc cga acc aag gac atc gag gac gtg ttt tac aaa tac
 V   G   N   L   P   P   D   I   R   T   K   D   I   E   D   V   F   Y   K   Y
ggc gcc atc cgc gac atc gac ctg aag aac cgc cgc ggg gga ccg ccc ttc gcc ttc gtt
 G   A   I   R   D   I   D   L   K   N   R   R   G   G   P   P   F   A   F   V
gag ttc gag gac ccg cga gac gcg gaa gat gcg gtg tac ggt cgc gac ggc tac gac tac
 E   F   E   D   P   R   D   A   E   D   A   V   Y   G   R   D   G   Y   D   Y
gac ggc tac cgg ctg cgg gta gag ttt ccc cga agc ggc cgc ggg acc ggc cga ggc ggc
 D   G   Y   R   L   R   V   E   F   P   R   S   G   R   G   T   G   R   G   G
ggc ggg ggt gga ggc ggc ggc gcc ccg aga ggc cgc tat ggc ccg ccg tcc agg cgg tcc
 G   G   G   G   G   G   G   A   P   R   G   R   Y   G   P   P   S   R   R   S
gag aac aga gtg gtt gtc tct gga ctg cct ccg agt gga agc tgg cag gac tta aag gat
 E   N   R   V   V   V   S   G   L   P   P   S   G   S   W   Q   D   L   K   D
cac atg cgt gag gca ggt gat gta tgt tac gct gat gtt tac cga gat ggc act ggt gtc
 H   M   R   E   A   G   D   V   C   Y   A   D   V   Y   R   D   G   T   G   V
gtg gag ttt gta cgg aaa gaa gat atg acg tat gca gtt cga aaa ctg gat aac act aag
```

```
 V   E   F   V   R   K   E   D   M   T   Y   A   V   R   K   L   D   N   T   K
ttt aga tct cac gag gga gaa act gcc tac atc cgg gtt aaa gtt gat ggg ccc aga agt
 F   R   S   H   E   G   E   T   A   Y   I   R   V   K   V   D   G   P   R   S
cca agt tat gga aga gat cga gat cga gat cgt gat cgt gat aga gat cgt gat aga gat
 P   S   Y   G   R   D   R   D   R   D   R   D   R   D   R   D   R   D   R   D
aac gat agg gat cgc agt tac tcc cca agg aga agc aga gga tca cca cgc tat tct ccc
 N   D   R   D   R   S   Y   S   P   R   R   S   R   G   S   P   R   Y   S   P
cgt cat agc aga tct cgc tct cgt aca taa
 R   H   S   R   S   R   S   R   T   -
```

SRSF1 LCD with an N-terminal His-H3C-tag (in pET41a(+)).

```
atg ggc cat cat cat cat cat cac ctg gaa gtt ctg ttc cag ggg ccc gga tcc ccg gaa
 M   G   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   G   S   P   E
ttc gag gga gaa act gcc tac atc cgg gtt aaa gtt gat ggg ccc aga agt cca agt tat
 F   E   G   E   T   A   Y   I   R   V   K   V   D   G   P   R   S   P   S   Y
gga aga tct cga tct cga agc cgt agt cgt agc aga agc cgt agc aga agc aac agc agg
 G   R   S   R   S   R   S   R   S   R   S   R   S   R   S   R   S   N   S   R
agt cgc agt tac tcc cca agg aga agc aga gga tca cca cgc tat tct ccc cgt cat agc
 S   R   S   Y   S   P   R   R   S   R   G   S   P   R   Y   S   P   R   H   S
aga tct cgc tct cgt aca taa
 R   S   R   S   R   T   -
```

SRSF1 LCD$_{Hypo-D}$ with an N-terminal His-H3C-tag (in pET41a(+)).

```
atg ggc cat cat cat cat cat cac ctg gaa gtt ctg ttc cag ggg ccc gga tcc ccg gaa
 M   G   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   G   S   P   E
ttc gag gga gaa act gcc tac atc cgg gtt aaa gtt gat ggg ccc aga agt cca agt tat
 F   E   G   E   T   A   Y   I   R   V   K   V   D   G   P   R   S   P   S   Y
gga aga gat cga gat cga gat cgt gat cgt gat aga gat cgt gat aga gat aac gat agg
 G   R   D   R   D   R   D   R   D   R   D   R   D   R   D   R   D   N   D   R
gat cgc agt tac tcc cca agg aga agc aga gga tca cca cgc tat tct ccc cgt cat agc
 D   R   S   Y   S   P   R   R   S   R   G   S   P   R   Y   S   P   R   H   S
aga tct cgc tct cgt aca taa
 R   S   R   S   R   T   -
```

SRSF1 LCD$_{Hyper-D}$ with an N-terminal His-H3C-tag (in pET41a(+)).

```
atg ggc cat cat cat cat cat cac ctg gaa gtt ctg ttc cag ggg ccc gga tcc ccg gaa
 M   G   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   G   S   P   E
ttc gag gga gaa act gcc tac atc cgg gtt aaa gtt gat ggg ccc aga agt cca agt tat
 F   E   G   E   T   A   Y   I   R   V   K   V   D   G   P   R   S   P   S   Y
gga aga gat cga gat cga gat cgt gat cgt gat aga gat cgt gat aga gat aac gat agg
 G   R   D   R   D   R   D   R   D   R   D   R   D   R   D   R   D   N   D   R
gat cgc agt tac gat cca agg aga agc aga gga gat cca cgc tat gat ccc cgt cat agc
 D   R   S   Y   D   P   R   R   S   R   G   D   P   R   Y   D   P   R   H   S
aga tct cgc tct cgt aca taa
 R   S   R   S   R   T   -
```

SRPK1 constructs

SRPK1 with an N-terminal GST-H3C-tag (in pGEX6p1).

```
atg tcc cct ata cta ggt tat tgg aaa att aag ggc ctt gtg caa ccc act cga ctt ctt
 M   S   P   I   L   G   Y   W   K   I   K   G   L   V   Q   P   T   R   L   L
ttg gaa tat ctt gaa gaa aaa tat gaa gag cat ttg tat gag cgc gat gaa ggt gat aaa
 L   E   Y   L   E   E   K   Y   E   E   H   L   Y   E   R   D   E   G   D   K
tgg cga aac aaa aag ttt gaa ttg ggt ttg gag ttt ccc aat ctt cct tat tat att gat
 W   R   N   K   K   F   E   L   G   L   E   F   P   N   L   P   Y   Y   I   D
ggt gat gtt aaa tta aca cag tct atg gcc atc ata cgt tat ata gct gac aag cac aac
```

```
 G   D   V   K   L   T   Q   S   M   A   I   I   R   Y   I   A   D   K   H   N
atg ttg ggt ggt tgt cca aaa gag cgt gca gag att tca atg ctt gaa gga gcg gtt ttg
 M   L   G   G   C   P   K   E   R   A   E   I   S   M   L   E   G   A   V   L
gat att aga tac ggt gtt tcg aga att gca tat agt aaa gac ttt gaa act ctc aaa gtt
 D   I   R   Y   G   V   S   R   I   A   Y   S   K   D   F   E   T   L   K   V
gat ttt ctt agc aag cta cct gaa atg ctg aaa atg ttc gaa gat cgt tta tgt cat aaa
 D   F   L   S   K   L   P   E   M   L   K   M   F   E   D   R   L   C   H   K
aca tat tta aat ggt gat cat gta acc cat cct gac ttc atg ttg tat gac gct ctt gat
 T   Y   L   N   G   D   H   V   T   H   P   D   F   M   L   Y   D   A   L   D
gtt gtt tta tac atg gac cca atg tgc ctg gat gcg ttc cca aaa tta gtt tgt ttt aaa
 V   V   L   Y   M   D   P   M   C   L   D   A   F   P   K   L   V   C   F   K
aaa cgt att gaa gct atc cca caa att gat aag tac ttg aaa tcc agc aag tat ata gca
 K   R   I   E   A   I   P   Q   I   D   K   Y   L   K   S   S   K   Y   I   A
tgg cct ttg cag ggc tgg caa gcc acg ttt ggt ggt ggc gac cat cct cca aaa tcg gat
 W   P   L   Q   G   W   Q   A   T   F   G   G   G   D   H   P   P   K   S   D
ctg gaa gtt ctg ttc cag ggg ccc ctg gga tcc ccg gaa ttc atg gag cgg aaa gtg ctt
 L   E   V   L   F   Q   G   P   L   G   S   P   E   F   M   E   R   K   V   L
gcg ctc cag gcc cga aag aaa agg acc aag gcc aag aag gac aaa gcc caa agg aaa tct
 A   L   Q   A   R   K   K   R   T   K   A   K   K   D   K   A   Q   R   K   S
gaa act cag cac cga ggc tct gct ccc cac tct gag agt gat cta cca gag cag gaa gag
 E   T   Q   H   R   G   S   A   P   H   S   E   S   D   L   P   E   Q   E   E
gag att ctg gga tct gat gat gat gag caa gaa gat cct aat gat tat tgt aaa gga ggt
 E   I   L   G   S   D   D   D   E   Q   E   D   P   N   D   Y   C   K   G   G
tat cat ctt gtg aaa att gga gat cta ttc aat ggg aga tac cat gtg atc cga aag tta
 Y   H   L   V   K   I   G   D   L   F   N   G   R   Y   H   V   I   R   K   L
ggc tgg gga cac ttt tca aca gta tgg tta tca tgg gat att cag ggg aag aaa ttt gtg
 G   W   G   H   F   S   T   V   W   L   S   W   D   I   Q   G   K   K   F   V
gca atg aaa gta gtt aaa agt gct gaa cat tac act gaa aca gca cta gat gaa atc cgg
 A   M   K   V   V   K   S   A   E   H   Y   T   E   T   A   L   D   E   I   R
ttg ctg aag tca gtt cgc aat tca gac cct aat gat cca aat aga gaa atg gtt gtt caa
 L   L   K   S   V   R   N   S   D   P   N   D   P   N   R   E   M   V   V   Q
cta cta gat gac ttt aaa ata tca gga gtt aat gga aca cat atc tgc atg gta ttt gaa
 L   L   D   D   F   K   I   S   G   V   N   G   T   H   I   C   M   V   F   E
gtt ttg ggg cat cat ctg ctc aag tgg atc atc aaa tcc aat tat cag ggg ctt cca ctg
 V   L   G   H   H   L   L   K   W   I   I   K   S   N   Y   Q   G   L   P   L
cct tgt gtc aaa aaa att att cag caa gtg tta cag ggt ctt gat tat tta cat acc aag
 P   C   V   K   K   I   I   Q   Q   V   L   Q   G   L   D   Y   L   H   T   K
tgc cgt atc atc cac act gac att aaa cca gag aac atc tta ttg tca gtg aat gag cag
 C   R   I   I   H   T   D   I   K   P   E   N   I   L   L   S   V   N   E   Q
tac att cgg agg ctg gct gca gaa gca aca gaa tgg cag cga tct gga gct cct ccg cct
 Y   I   R   R   L   A   A   E   A   T   E   W   Q   R   S   G   A   P   P   P
tcc gga tct gca gtc agt act gct ccc cag cct aaa cca gct gac aaa atg tca aag aat
 S   G   S   A   V   S   T   A   P   Q   P   K   P   A   D   K   M   S   K   N
aag aag aag aaa ttg aag aag aag cag aag cgc cag gca gaa tta cta gag aag cga atg
 K   K   K   K   L   K   K   K   Q   K   R   Q   A   E   L   L   E   K   R   M
cag gaa att gag gaa atg gag aaa gag tcg ggc cct ggg caa aaa aga cca aac aag caa
 Q   E   I   E   E   M   E   K   E   S   G   P   G   Q   K   R   P   N   K   Q
gaa gaa tca gag agt cct gtt gaa aga ccc ttg aaa gag aac cca cct aat aaa atg acc
 E   E   S   E   S   P   V   E   R   P   L   K   E   N   P   P   N   K   M   T
caa gaa aaa ctt gaa gag tca agt acc att ggc cag gat caa acg ctt atg gaa cgt gat
 Q   E   K   L   E   E   S   S   T   I   G   Q   D   Q   T   L   M   E   R   D
aca gag ggt ggt gca gca gaa att aat tgc aat gga gtg att gaa gtc att aat tat act
 T   E   G   G   A   A   E   I   N   C   N   G   V   I   E   V   I   N   Y   T
cag aac agt aat aat gaa aca ttg aga cat aaa gag gat cta cat aat gct aat gac tgt
 Q   N   S   N   N   E   T   L   R   H   K   E   D   L   H   N   A   N   D   C
gat gtc caa aat ttg aat cag gaa tct agt ttc cta agc tcc caa aat gga gac agc agc
 D   V   Q   N   L   N   Q   E   S   S   F   L   S   S   Q   N   G   D   S   S
aca tct caa gaa aca gac tct tgt aca cct ata aca tct gag gtg tca gac acc atg gtg
```

```
 T   S   Q   E   T   D   S   C   T   P   I   T   S   E   V   S   D   T   M   V
tgc cag tct tcc tca act gta ggt cag tca ttc agt gaa caa cac att agc caa ctt caa
 C   Q   S   S   S   T   V   G   Q   S   F   S   E   Q   H   I   S   Q   L   Q
gaa agc att cgg gca gag ata ccc tgt gaa gat gaa caa gag caa gaa cat aac gga cca
 E   S   I   R   A   E   I   P   C   E   D   E   Q   E   Q   E   H   N   G   P
ctg gac aac aaa gga aaa tcc acg gct gga aat ttt ctt gtt aat ccc ctt gag cca aaa
 L   D   N   K   G   K   S   T   A   G   N   F   L   V   N   P   L   E   P   K
aat gca gaa aag ctc aag gtg aag att gct gac ctt gga aat gct tgt tgg gtg cac aaa
 N   A   E   K   L   K   V   K   I   A   D   L   G   N   A   C   W   V   H   K
cat ttc act gaa gat att caa aca agg caa tat cgt tcc ttg gaa gtt cta atc gga tct
 H   F   T   E   D   I   Q   T   R   Q   Y   R   S   L   E   V   L   I   G   S
ggc tat aat acc cct gct gac att tgg agc acg gca tgc atg gcc ttt gaa ctg gcc aca
 G   Y   N   T   P   A   D   I   W   S   T   A   C   M   A   F   E   L   A   T
ggt gac tat ttg ttt gaa cct cat tca ggg gaa gag tac act cga gat gaa gat cac att
 G   D   Y   L   F   E   P   H   S   G   E   E   Y   T   R   D   E   D   H   I
gca ttg atc ata gaa ctt ctg ggg aag gtg cct cgc aag ctc att gtg gca gga aaa tat
 A   L   I   I   E   L   L   G   K   V   P   R   K   L   I   V   A   G   K   Y
tcc aag gaa ttt ttc acc aaa aaa ggt gac ctg aaa cat atc acg aag ctg aaa cct tgg
 S   K   E   F   F   T   K   K   G   D   L   K   H   I   T   K   L   K   P   W
ggc ctt ttt gag gtt cta gtg gag aag tat gag tgg tcg cag gaa gag gca gct ggc ttc
 G   L   F   E   V   L   V   E   K   Y   E   W   S   Q   E   E   A   A   G   F
aca gat ttc tta ctg ccc atg ttg gag ctg atc cct gag aag aga gcc act gcc gcc gag
 T   D   F   L   L   P   M   L   E   L   I   P   E   K   R   A   T   A   A   E
tgt ctc cgg cac cct tgg ctt aac tcc taa
 C   L   R   H   P   W   L   N   S   -
```

SRPK1 with an N-terminal His-H3C-tag (in pET41a(+)).

```
atg ggc cat cat cat cat cat cac ctg gaa gtt ctg ttc cag ggg ccc gga tcc ccg gaa
 M   G   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   G   S   P   E
ttc atg gag cgg aaa gtg ctt gcg ctc cag gcc cga aag aaa agg acc aag gcc aag aag
 F   M   E   R   K   V   L   A   L   Q   A   R   K   K   R   T   K   A   K   K
gac aaa gcc caa agg aaa tct gaa act cag cac cga ggc tct gct ccc cac tct gag agt
 D   K   A   Q   R   K   S   E   T   Q   H   R   G   S   A   P   H   S   E   S
gat cta cca gag cag gaa gag gag att ctg gga tct gat gat gat gag caa gaa gat cct
 D   L   P   E   Q   E   E   E   I   L   G   S   D   D   D   E   Q   E   D   P
aat gat tat tgt aaa gga ggt tat cat ctt gtg aaa att gga gat cta ttc aat ggg aga
 N   D   Y   C   K   G   G   Y   H   L   V   K   I   G   D   L   F   N   G   R
tac cat gtg atc cga aag tta ggc tgg gga cac ttt tca aca gta tgg tta tca tgg gat
 Y   H   V   I   R   K   L   G   W   G   H   F   S   T   V   W   L   S   W   D
att cag ggg aag aaa ttt gtg gca atg aaa gta gtt aaa agt gct gaa cat tac act gaa
 I   Q   G   K   K   F   V   A   M   K   V   V   K   S   A   E   H   Y   T   E
aca gca cta gat gaa atc cgg ttg ctg aag tca gtt cgc aat tca gac cct aat gat cca
 T   A   L   D   E   I   R   L   L   K   S   V   R   N   S   D   P   N   D   P
aat aga gaa atg gtt gtt caa cta cta gat gac ttt aaa ata tca gga gtt aat gga aca
 N   R   E   M   V   V   Q   L   L   D   D   F   K   I   S   G   V   N   G   T
cat atc tgc atg gta ttt gaa gtt ttg ggg cat cat ctg ctc aag tgg atc atc aaa tcc
 H   I   C   M   V   F   E   V   L   G   H   H   L   L   K   W   I   I   K   S
aat tat cag ggg ctt cca ctg cct tgt gtc aaa aaa att att cag caa gtg tta cag ggt
 N   Y   Q   G   L   P   L   P   C   V   K   K   I   I   Q   Q   V   L   Q   G
ctt gat tat tta cat acc aag tgc cgt atc atc cac act gac att aaa cca gag aac atc
 L   D   Y   L   H   T   K   C   R   I   I   H   T   D   I   K   P   E   N   I
tta ttg tca gtg aat gag cag tac att cgg agg ctg gct gca gaa gca aca gaa tgg cag
 L   L   S   V   N   E   Q   Y   I   R   R   L   A   A   E   A   T   E   W   Q
cga tct gga gct cct ccg cct tcc gga tct gca gtc agt act gct ccc cag cct aaa cca
 R   S   G   A   P   P   P   S   G   S   A   V   S   T   A   P   Q   P   K   P
gct gac aaa atg tca aag aat aag aag aag aaa ttg aag aag aag cag aag cgc cag gca
 A   D   K   M   S   K   N   K   K   K   K   L   K   K   K   Q   K   R   Q   A
```

```
gaa tta cta gag aag cga atg cag gaa att gag gaa atg gag aaa gag tcg ggc cct ggg
 E   L   L   E   K   R   M   Q   E   I   E   E   M   E   K   E   S   G   P   G
caa aaa aga cca aac aag caa gaa gaa tca gag agt cct gtt gaa aga ccc ttg aaa gag
 Q   K   R   P   N   K   Q   E   E   S   E   S   P   V   E   R   P   L   K   E
aac cca cct aat aaa atg acc caa gaa aaa ctt gaa gag tca agt acc att ggc cag gat
 N   P   P   N   K   M   T   Q   E   K   L   E   E   S   S   T   I   G   Q   D
caa acg ctt atg gaa cgt gat aca gag ggt ggt gca gca gaa att aat tgc aat gga gtg
 Q   T   L   M   E   R   D   T   E   G   G   A   A   E   I   N   C   N   G   V
att gaa gtc att aat tat act cag aac agt aat aat gaa aca ttg aga cat aaa gag gat
 I   E   V   I   N   Y   T   Q   N   S   N   N   E   T   L   R   H   K   E   D
cta cat aat gct aat gac tgt gat gtc caa aat ttg aat cag gaa tct agt ttc cta agc
 L   H   N   A   N   D   C   D   V   Q   N   L   N   Q   E   S   S   F   L   S
tcc caa aat gga gac agc agc aca tct caa gaa aca gac tct tgt aca cct ata aca tct
 S   Q   N   G   D   S   S   T   S   Q   E   T   D   S   C   T   P   I   T   S
gag gtg tca gac acc atg gtg tgc cag tct tcc tca act gta ggt cag tca ttc agt gaa
 E   V   S   D   T   M   V   C   Q   S   S   S   T   V   G   Q   S   F   S   E
caa cac att agc caa ctt caa gaa agc att cgg gca gag ata ccc tgt gaa gat gaa caa
 Q   H   I   S   Q   L   Q   E   S   I   R   A   E   I   P   C   E   D   E   Q
gag caa gaa cat aac gga cca ctg gac aac aaa gga aaa tcc acg gct gga aat ttt ctt
 E   Q   E   H   N   G   P   L   D   N   K   G   K   S   T   A   G   N   F   L
gtt aat ccc ctt gag cca aaa aat gca gaa aag ctc aag gtg aag att gct gac ctt gga
 V   N   P   L   E   P   K   N   A   E   K   L   K   V   K   I   A   D   L   G
aat gct tgt tgg gtg cac aaa cat ttc act gaa gat att caa aca agg caa tat cgt tcc
 N   A   C   W   V   H   K   H   F   T   E   D   I   Q   T   R   Q   Y   R   S
ttg gaa gtt cta atc gga tct ggc tat aat acc cct gct gac att tgg agc acg gca tgc
 L   E   V   L   I   G   S   G   Y   N   T   P   A   D   I   W   S   T   A   C
atg gcc ttt gaa ctg gcc aca ggt gac tat ttg ttt gaa cct cat tca ggg gaa gag tac
 M   A   F   E   L   A   T   G   D   Y   L   F   E   P   H   S   G   E   E   Y
act cga gat gaa gat cac att gca ttg atc ata gaa ctt ctg ggg aag gtg cct cgc aag
 T   R   D   E   D   H   I   A   L   I   I   E   L   L   G   K   V   P   R   K
ctc att gtg gca gga aaa tat tcc aag gaa ttt ttc acc aaa aaa ggt gac ctg aaa cat
 L   I   V   A   G   K   Y   S   K   E   F   F   T   K   K   G   D   L   K   H
atc acg aag ctg aaa cct tgg ggc ctt ttt gag gtt cta gtg gag aag tat gag tgg tcg
 I   T   K   L   K   P   W   G   L   F   E   V   L   V   E   K   Y   E   W   S
cag gaa gag gca gct ggc ttc aca gat ttc tta ctg ccc atg ttg gag ctg atc cct gag
 Q   E   E   A   A   G   F   T   D   F   L   L   P   M   L   E   L   I   P   E
aag aga gcc act gcc gcc gag tgt ctc cgg cac cct tgg ctt aac tcc taa
 K   R   A   T   A   A   E   C   L   R   H   P   W   L   N   S   -
```

CLK1 constructs

CLK1 FL with an N-terminal His-H3C-tag (in pET41a(+)).

```
atg ggc cat cat cat cat cat cac ctg gaa gtt ctg ttc cag ggg ccc gga tcc ccg gaa
 M   G   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   G   S   P   E
ttc atg aga cac tca aag aga act tac tgt cct gat tgg gat gac aag gat tgg gat tat
 F   M   R   H   S   K   R   T   Y   C   P   D   W   D   D   K   D   W   D   Y
gga aaa tgg agg agc agc agc agt cat aaa aga agg aag aga tca cat agc agt gcc cag
 G   K   W   R   S   S   S   S   H   K   R   R   K   R   S   H   S   S   A   Q
gag aac aag cgc tgc aaa tac aat cac tct aaa atg tgt gat agc cat tat ttg gaa agc
 E   N   K   R   C   K   Y   N   H   S   K   M   C   D   S   H   Y   L   E   S
agg tct ata aat gag aaa gat tat cat agt cga cgc tac att gat gag tac aga aat gac
 R   S   I   N   E   K   D   Y   H   S   R   R   Y   I   D   E   Y   R   N   D
tac act caa gga tgt gaa cct gga cat cgc caa aga gac cat gaa agc cgg tat cag aac
 Y   T   Q   G   C   E   P   G   H   R   Q   R   D   H   E   S   R   Y   Q   N
cat agt agc aag tct tct ggt aga agt gga aga agt agt tat aaa agc aaa cac agg att
 H   S   S   K   S   S   G   R   S   G   R   S   S   Y   K   S   K   H   R   I
cac cac agt act tca cat cgt cgt tca cat ggg aag agt cac cga agg aaa aga acc agg
 H   H   S   T   S   H   R   R   S   H   G   K   S   H   R   R   K   R   T   R
```

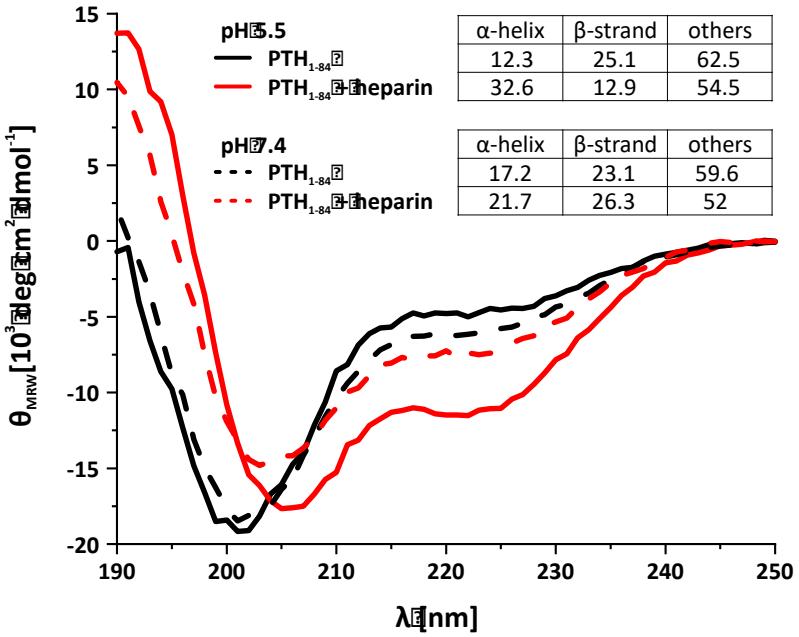
```
agt gta gag gat gat gag gag ggt cac ctg atc tgt cag agt gga gac gta cta agt gca
 S   V   E   D   D   E   E   G   H   L   I   C   Q   S   G   D   V   L   S   A
aga tat gaa att gtt gat act tta ggt gaa gga gct ttt gga aaa gtt gtg gag tgc atc
 R   Y   E   I   V   D   T   L   G   E   G   A   F   G   K   V   V   E   C   I
gat cat aaa gcg gga ggt aga cat gta gca gta aaa ata gtt aaa aat gtg gat aga tac
 D   H   K   A   G   G   R   H   V   A   V   K   I   V   K   N   V   D   R   Y
tgt gaa gct gct cgc tca gaa ata caa gtt ctg gaa cat ctg aat aca aca gac ccc aac
 C   E   A   A   R   S   E   I   Q   V   L   E   H   L   N   T   T   D   P   N
agt act ttc cgc tgt gtc cag atg ttg gaa tgg ttt gag cat cat ggt cac att tgc att
 S   T   F   R   C   V   Q   M   L   E   W   F   E   H   H   G   H   I   C   I
gtt ttt gaa cta ttg gga ctt agt act tac gac ttc att aaa gaa aat ggt ttt cta cca
 V   F   E   L   L   G   L   S   T   Y   D   F   I   K   E   N   G   F   L   P
ttt cga ctg gat cat atc aga aag atg gca tat cag ata tgc aag tct gtg aat ttt ttg
 F   R   L   D   H   I   R   K   M   A   Y   Q   I   C   K   S   V   N   F   L
cac agt aat aag ttg act cac aca gac tta aag cct gaa aac atc tta ttt gtg cag tct
 H   S   N   K   L   T   H   T   D   L   K   P   E   N   I   L   F   V   Q   S
gac tac aca gag gcg tat aat ccc aaa ata aaa cgt gat gaa cgc acc tta ata aat cca
 D   Y   T   E   A   Y   N   P   K   I   K   R   D   E   R   T   L   I   N   P
gat att aaa gtt gta gac ttt ggt agt gca aca tat gat gac gaa cat cac agt aca ttg
 D   I   K   V   V   D   F   G   S   A   T   Y   D   D   E   H   H   S   T   L
gta tct aca aga cat tat aga gca cct gaa gtt att tta gcc cta ggg tgg tcc caa cca
 V   S   T   R   H   Y   R   A   P   E   V   I   L   A   L   G   W   S   Q   P
tgt gat gtc tgg agc ata gga tgc att ctt att gaa tac tat ctt ggg ttt acc gta ttt
 C   D   V   W   S   I   G   C   I   L   I   E   Y   Y   L   G   F   T   V   F
cca aca cac gat agt aag gag cat tta gca atg atg gaa agg att ctt gga cct cta cca
 P   T   H   D   S   K   E   H   L   A   M   M   E   R   I   L   G   P   L   P
aaa cat atg ata cag aaa acc agg aaa cgt aaa tat ttt cac cac gat cga tta gac tgg
 K   H   M   I   Q   K   T   R   K   R   K   Y   F   H   H   D   R   L   D   W
gat gaa cac agt tct gcc ggc aga tat gtt tca aga cgc tgt aaa cct ctg aag gaa ttt
 D   E   H   S   S   A   G   R   Y   V   S   R   R   C   K   P   L   K   E   F
atg ctt tct caa gat gtt gaa cat gag cgt ctc ttt gac ctc att cag aaa atg ttg gag
 M   L   S   Q   D   V   E   H   E   R   L   F   D   L   I   Q   K   M   L   E
tat gat cca gcc aaa aga att act ctc aga gaa gcc tta aag cat cct ttc ttt gac ctt
 Y   D   P   A   K   R   I   T   L   R   E   A   L   K   H   P   F   F   D   L
ctg aag aaa agt ata tag
 L   K   K   S   I   -
```

CLK1 ΔN with an N-terminal His-H3C-tag (in pET41a(+)).

```
atg ggc cat cat cat cat cat cac ctg gaa gtt ctg ttc cag ggg ccc gga tcc ccg gaa
 M   G   H   H   H   H   H   H   L   E   V   L   F   Q   G   P   G   S   P   E
ttc cac ctg atc tgt cag agt gga gac gta cta agt gca aga tat gaa att gtt gat act
 F   H   L   I   C   Q   S   G   D   V   L   S   A   R   Y   E   I   V   D   T
tta ggt gaa gga gct ttt gga aaa gtt gtg gag tgc atc gat cat aaa gcg gga ggt aga
 L   G   E   G   A   F   G   K   V   V   E   C   I   D   H   K   A   G   G   R
cat gta gca gta aaa ata gtt aaa aat gtg gat aga tac tgt gaa gct gct cgc tca gaa
 H   V   A   V   K   I   V   K   N   V   D   R   Y   C   E   A   A   R   S   E
ata caa gtt ctg gaa cat ctg aat aca aca gac ccc aac agt act ttc cgc tgt gtc cag
 I   Q   V   L   E   H   L   N   T   T   D   P   N   S   T   F   R   C   V   Q
atg ttg gaa tgg ttt gag cat cat ggt cac att tgc att gtt ttt gaa cta ttg gga ctt
 M   L   E   W   F   E   H   H   G   H   I   C   I   V   F   E   L   L   G   L
agt act tac gac ttc att aaa gaa aat ggt ttt cta cca ttt cga ctg gat cat atc aga
 S   T   Y   D   F   I   K   E   N   G   F   L   P   F   R   L   D   H   I   R
aag atg gca tat cag ata tgc aag tct gtg aat ttt ttg cac agt aat aag ttg act cac
 K   M   A   Y   Q   I   C   K   S   V   N   F   L   H   S   N   K   L   T   H
aca gac tta aag cct gaa aac atc tta ttt gtg cag tct gac tac aca gag gcg tat aat
 T   D   L   K   P   E   N   I   L   F   V   Q   S   D   Y   T   E   A   Y   N
ccc aaa ata aaa cgt gat gaa cgc acc tta ata aat cca gat att aaa gtt gta gac ttt
 P   K   I   K   R   D   E   R   T   L   I   N   P   D   I   K   V   V   D   F
```

```
ggt agt gca aca tat gat gac gaa cat cac agt aca ttg gta tct aca aga cat tat aga
 G   S   A   T   Y   D   D   E   H   H   S   T   L   V   S   T   R   H   Y   R
gca cct gaa gtt att tta gcc cta ggg tgg tcc caa cca tgt gat gtc tgg agc ata gga
 A   P   E   V   I   L   A   L   G   W   S   Q   P   C   D   V   W   S   I   G
tgc att ctt att gaa tac tat ctt ggg ttt acc gta ttt cca aca cac gat agt aag gag
 C   I   L   I   E   Y   Y   L   G   F   T   V   F   P   T   H   D   S   K   E
cat tta gca atg atg gaa agg att ctt gga cct cta cca aaa cat atg ata cag aaa acc
 H   L   A   M   M   E   R   I   L   G   P   L   P   K   H   M   I   Q   K   T
agg aaa cgt aaa tat ttt cac cac gat cga tta gac tgg gat gaa cac agt tct gcc ggc
 R   K   R   K   Y   F   H   H   D   R   L   D   W   D   E   H   S   S   A   G
aga tat gtt tca aga cgc tgt aaa cct ctg aag gaa ttt atg ctt tct caa gat gtt gaa
 R   Y   V   S   R   R   C   K   P   L   K   E   F   M   L   S   Q   D   V   E
cat gag cgt ctc ttt gac ctc att cag aaa atg ttg gag tat gat cca gcc aaa aga att
 H   E   R   L   F   D   L   I   Q   K   M   L   E   Y   D   P   A   K   R   I
act ctc aga gaa gcc tta aag cat cct ttc ttt gac ctt ctg aag aaa agt ata tag
 T   L   R   E   A   L   K   H   P   F   F   D   L   L   K   K   S   I   -
```

Supporting figures



	α-helix	β-strand	others
pH 5.5			
PTH_{1-84}	12.3	25.1	62.5
PTH_{1-84} + heparin	32.6	12.9	54.5
pH 7.4			
PTH_{1-84}	17.2	23.1	59.6
PTH_{1-84} + heparin	21.7	26.3	52

Figure 34: **pH dependent circular dichroism spectra of PTH_{1-84}.** Buffer effect on circular dichroism spectra of PTH_{1-84} in absence and presence of heparin (1/10 molar ratio heparin:PTH_{1-84}). Secondary structure calculation was performed using the BeStSel online tool [172]. Adapted from [167].

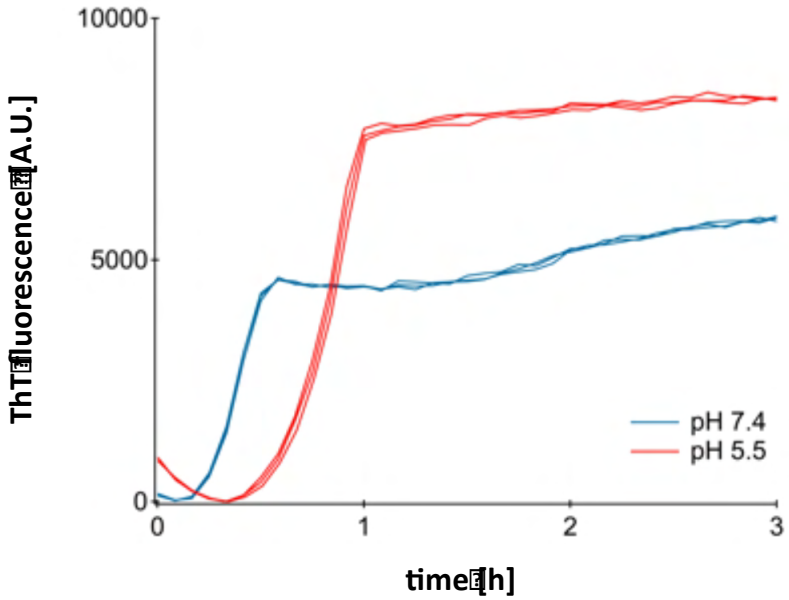


Figure 35: **pH dependent ThT curves of PTH$_{1-84}$.** Buffer effect on ThT curves of PTH$_{1-84}$ in absence and presence of heparin (1:10 molar ratio heparin:PTH$_{1-84}$). Adapted from [167].

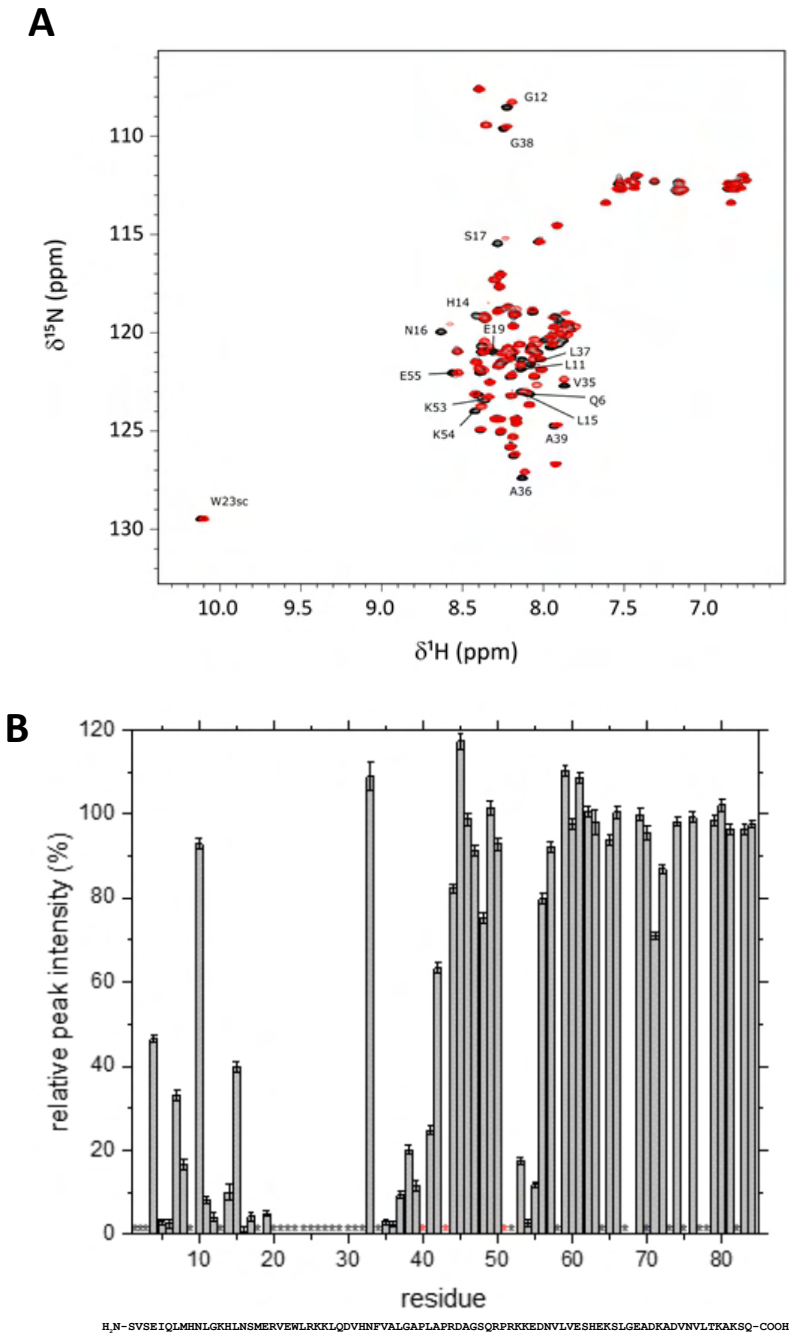


Figure 36: **NMR analysis of the fondaparinux:PTH$_{1-84}$ interaction.** A: 2D ^{1}H-^{15}N HSQC spectrum of 100 μM PTH1-84 at pH 5.5 and 25 ℃ in the absence (black) and presence (red) of fondaparinux at a 1:1 stochiometry. Residues showing major changes in chemical shifts upon addition of the sugar are labeled. B: Relative NMR intensity of the PTH$_{1-84}$ backbone amides in the presence of fondaparinux. Residues with missing or ambiguous assignments are marked by a black asterisk, prolyl residues with a red asterisk. The PTH$_{1-84}$ primary sequence is shown. Adapted from [167].

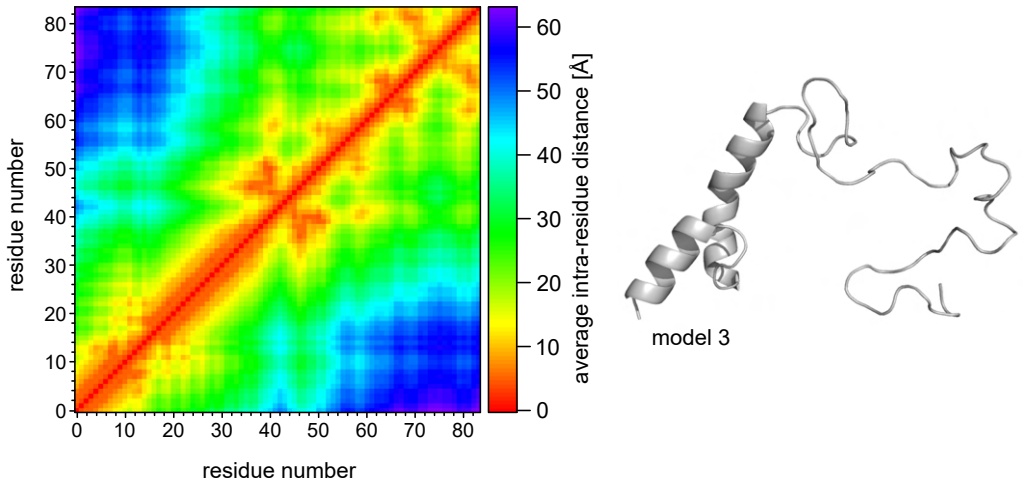


Figure 37: **MD simulation of PTH$_{1-84}$.** A heatmap (left panel) displays the time-averaged inter-residue (C_α-C_α) distances of PTH$_{1-84}$ obtained from a 25 ns MD simulation with the initial model (model 3) shown (right panel).

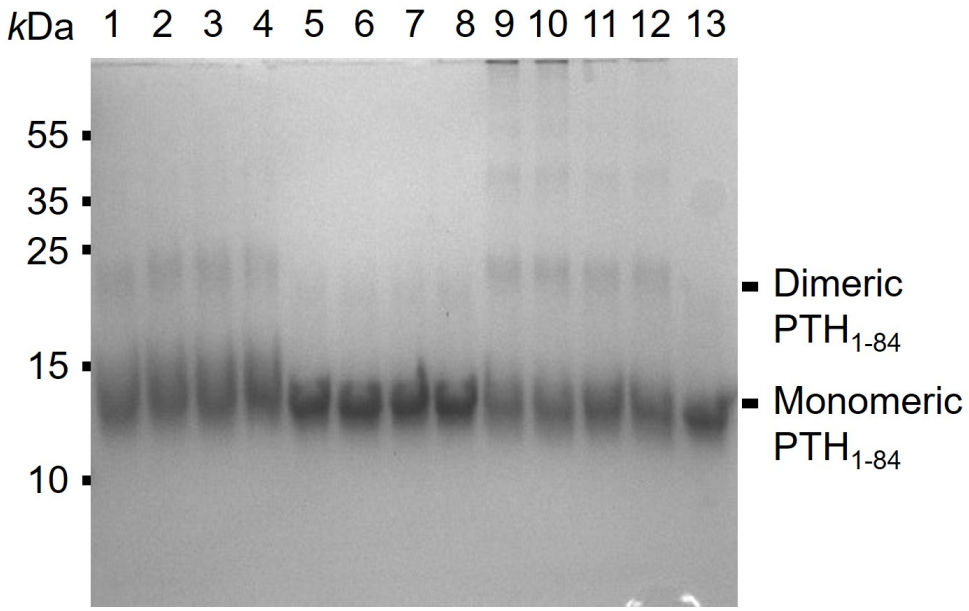


Figure 38: **SDS-PAGE of cross-linked PTH$_{1-84}$.** After cross-linking, protein samples were subjected to SDS-PAGE to isolate monomeric PTH$_{1-84}$. 4 independent experiments were conducted per cross-linking agent (DSBU - lanes 1-4; EDC/Sulfo-NHS - lanes 5-8; Sulfo-SDA - lanes 9-12; control (no cross-linker) - lane 13).

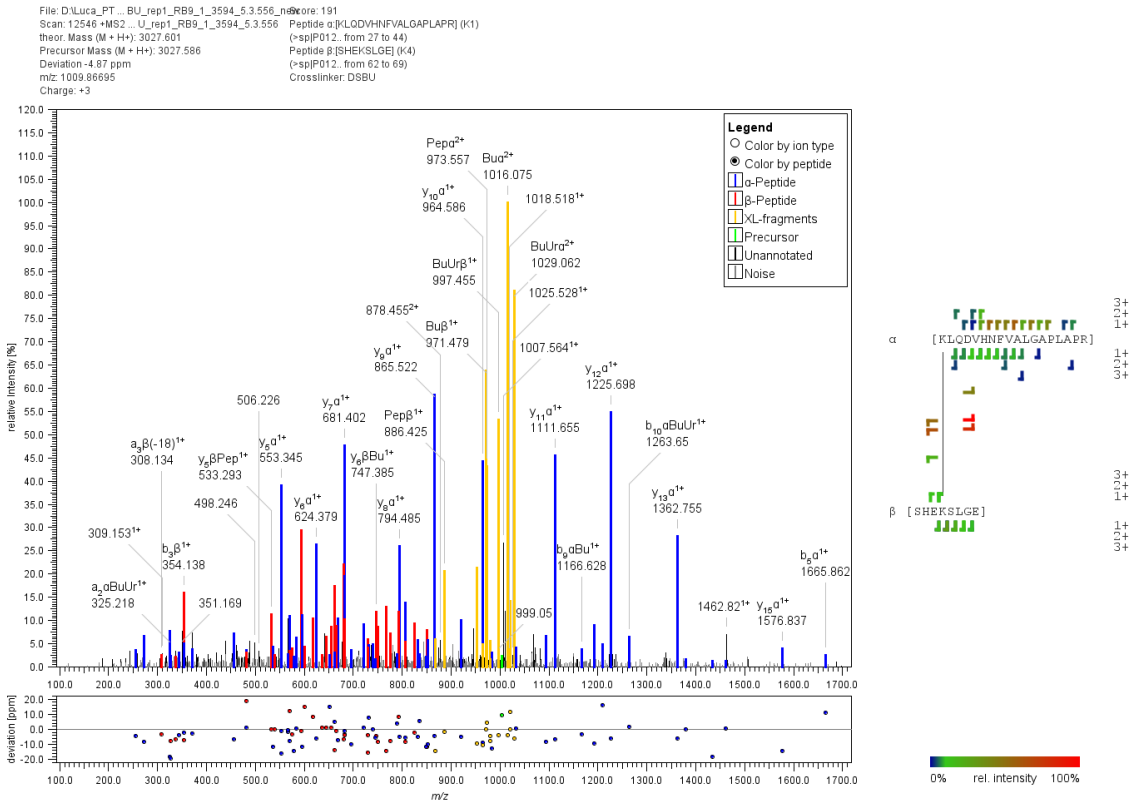


Figure 39: **Fragment ion spectrum of a +n charged DSBU cross-link (Peptide1-Peptide2) at m/z 1009.867, analyzed using MeroX software.** Exemplary deisotoped fragment ion spectrum for a DSBU cross-link. DSBU diagnostic doublets are shown in yellow, b- and y-type ion series for the α and β peptides are presented in blue and red, respectively.

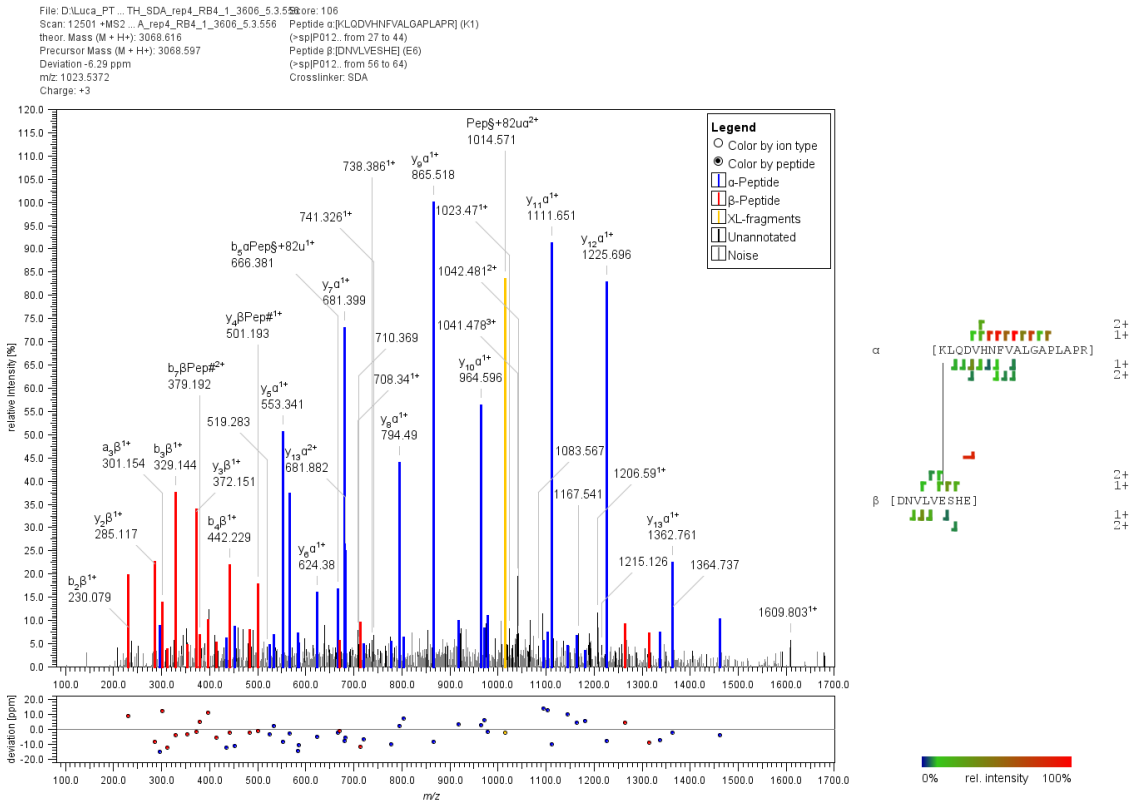


Figure 40: **Fragment ion spectrum of a +n charged sulfo-SDA cross-link (Peptide1-Peptide2) at m/z 1023.537, analyzed using MeroX software.** Exemplary de-isotoped fragment ion spectrum for a sulfo-SDA cross-link. b- and y-type ion series for the α and β peptides are presented in blue and red, respectively.

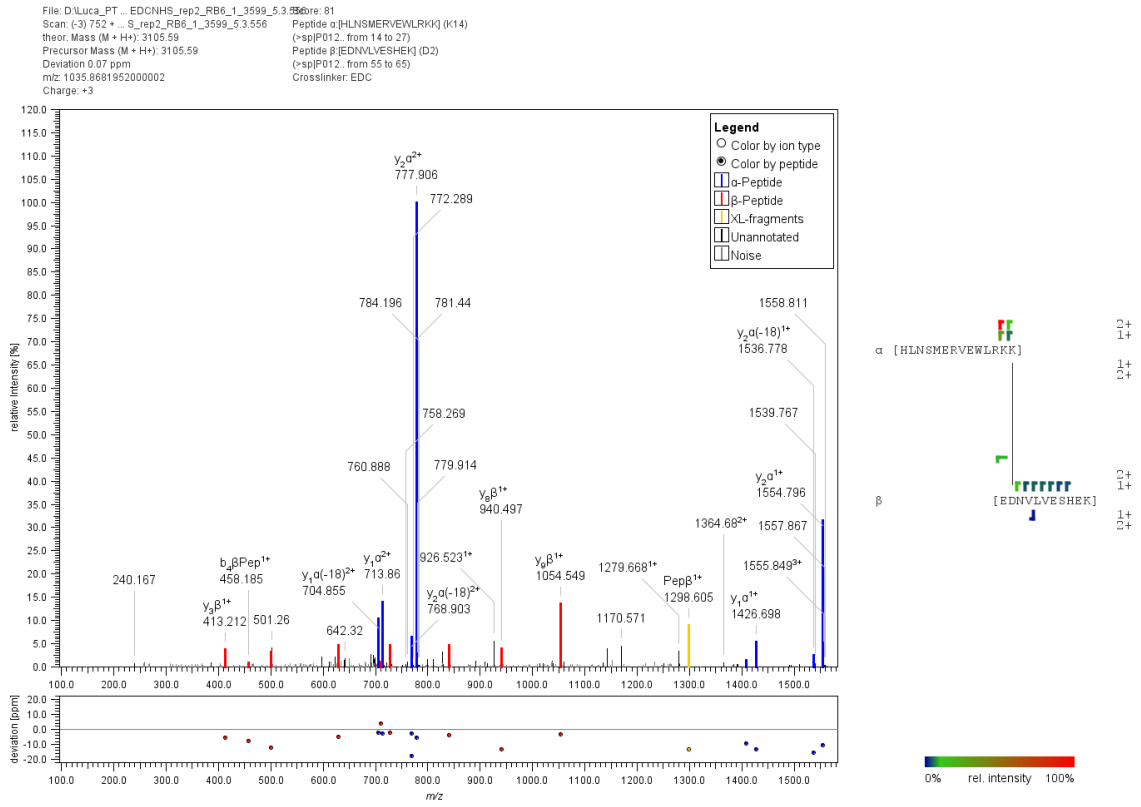


Figure 41: **Fragment ion spectrum of a +n charged EDC/sulfo-NHS cross-link (Peptide1-Peptide2) at m/z 1035.868, analyzed using MeroX software.** Exemplary de-isotoped fragment ion spectrum for an EDC/sulfo-NHS cross-link. b- and y-type ion series for the α and β peptides are presented in blue and red, respectively.

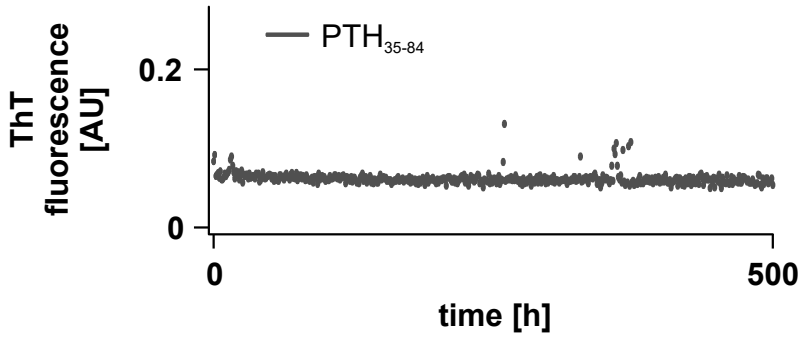


Figure 42: **ThT fluorescence curve of PTH$_{35-84}$.** 150 μM PTH$_{35-84}$ in 50 mM NaP$_i$ buffer, pH 7.4. Over the course of the assay, no increase in ThT fluorescence was detected.

List of Figures

List of Tables

VII Acknowledgements

I am extremely grateful to my supervisor Dr. Maria Ott. Without her trust, her support and her invaluable patience, this endeavor would not have been possible. I also could not have undertaken this journey without my defense committee who generously shared expertise. Additionally, I would like to express my deepest gratitude to the DFG, who funded the project. Likewise, I am very grateful to the members of my thesis advisory committee, Prof. Dr. Jochen Balbach and Prof. Dr. Milton Stubbs for the invaluable scientific discussions.

I am deeply indebted to Dr. Marcel Köhn, with whom a very close collaboration was established during the project. I highly appreciate his tireless support and his generous offer to have me work in his labs. I am incredibly thankful to PD Dr. Hauke Lilie and apl. Prof. Dr. Ralph Golbik who both shared their expertise and supported me in practical work. Special thanks to my peers and good friends Daniele, Alessio, Marina, Jaime, Peter, Tala, Marija and Twinkle. Without their support, both, inside and outside of the lab, this work would have been impossible.

Lastly, I would like to mention my family and friends who have always managed to motivate me and offered moral and emotional support.

VIII Declaration

I hereby declare that I wrote the submitted dissertation without any unauthorized external assistance and used only sources acknowledged in the work. All textual passages which are appropriated verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. In carrying out this research, I complied with the rules of standard scientific practice.

_____

Luca M. Lauth

Halle (Saale), den 03.05.2024