UNIVERSITY OF CALIFORNIA SAN DIEGO

Biophysical investigation of protein liquid-to-solid phase transitions and their modulation by small heat shock proteins

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Chemistry

by

Raymond F. Berkeley

Committee in charge:

Professor Galia T. Debelouchina, Chair
Professor Rommie E. Amaro
Professor Michael D. Burkart
Professor Lalit Deshmukh
Professor Eugene W. Yeo

2023

The dissertation of Raymond F. Berkeley is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2023

DEDICATION

For Jan, Steve, and Alan.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ACD	α-crystallin domain
ALS	amyotrophic lateral sclerosis
ATP	adenosine triphosphate
AUC	analytical ultracentrifugation
Bis-ANS	4,4'-dianilino-1,1'-binaphthyl-5,5'-disulfonic acid
BMRB	Biological Magnetic Resonance Data Bank
CAPS	*N*-cyclohexyl-3-aminopropanesulfonic acid
CG-MD	coarse-grained molecular dynamics
CP	cross polarization
Cryo-EM	cryogenic electron microscopy
CTD	C-terminal domain
DARR	dipolar assisted rotational resonance
DEER	double electron-electron resonance
DNA	deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
EMDB	Electron Microscopy Data Bank
EPL	expressed protein ligation
EPR	electron paramagnetic resonance
FID	free induction decay
FPLC	fast protein liquid chromatography
FRAP	fluorescence recovery after photobleaching
FRET	fluorescence resonance energy transfer

FTLD	frontotemporal lobar degeneration
FUS	Fused in sarcoma
hnRNPA1	heterogeneous nuclear ribonucleoprotein A1
HP1α	heterochromatin protein 1^{α}
HPLC	high-performance liquid chromatography
HSP/sHsp	heat shock protein/small heat shock protein
HSQC	heteronuclear single quantum coherence
IDP	intrinsically disordered protein
INEPT	insensitive nuclei enhanced by polarization transfer
IPTG	isopropyl β-D-1-thiogalactopyranoside
LC	low complexity
LLPS	liquid-liquid phase separation
MAS	magic-angle spinning
MBP	maltose binding protein
MD	molecular dynamics
MTSL	*S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate
NCC	normalize cross-correlation
NCL	native chemical ligation
NMR	nuclear magnetic resonance
NTD	N-terminal domain
PACSY	protein structure and chemical shift spectroscopy
PDB	Protein Data Bank

PRE	paramagnetic relaxation enhancement
PTM	post-translational modification
RCF	relative centrifugal force
RF	radiofrequency
RNA	ribonucleic acid
STRIDE	structural identification
TCEP	tris(2-carboxyethyl)phosphine
TDP-43	TAR DNA-binding protein 43
TEV	tobacco etch virus
ThT	thioflavin T
TMAO	trimethylamine *N*-oxide
TOBSY	total through bond correlation spectroscopy

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I'm deeply grateful to the many people who have supported, motivated, and inspired me during my scientific training. In particular, my mom, dad, and brother have always had my back, both as my core support system during difficult times and sources of joy and inspiration the rest of the time. This thesis would not have been possible without Steve's composure and dry sense of humor, Jan's openness, enthusiasm, and interest in even the most trivial topics, and Alan's dry wit and stoicism.

The same goes for my grandparents Marie, John, Joanie, and David, who have been my great supporters. I'm also grateful for my extended family, especially Sara, whose prose I envy, and Pete, whose zeal for life is difficult to match.

I'm also lucky to have a great group of friends, too large to list here. They all know who they are, and I appreciate them.

Finally, I'm thankful for the supportive and engaging lab that I am lucky to be a part of. The Debelouchina Lab has been a fantastic place to train as a scientist, and I'm grateful for the friendship of all our lab members. I'm especially grateful to Galia, who has mentored, supported, and championed me through graduate school.

Chapter 2, in full, is a reprint of the material as it appears in *Biophysical Journal*: Berkeley, R. F., Kashefi, M. & Debelouchina, G. T. Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC. *Biophys J* **120**, 1276–1287 (2021). The dissertation author was the primary investigator and author of this material.

Chapter 3, in part, is currently being prepared for submission for publication. Berkeley, R. F., Plonski, A. P., Cook, B. D., McGuire, K. L., Herzik, M. A., & Debelouchina, G. T. The dissertation author was the primary investigator and author of this material.

Chapter 4, in part, is currently being prepared for submission for publication. Berkeley, R. F., Plonski, A. P., Cook, B. D., McGuire, K. L., Herzik, M. A., & Debelouchina, G. T. The dissertation author was the primary investigator and author of this material.

Chapter 5, in full, is a reprint of the material as it appears in *Chemical Science*: Berkeley, R. F. & Debelouchina, G. T. Chemical tools for study and modulation of biomolecular phase transitions. *Chem Sci* **13**, 14226–14245 (2022). The dissertation author was the primary investigator and author of this material.

VITA

RESEARCH EXPERIENCE

University of California San Diego, La Jolla, CA September 2017 - Present
Advisor: Galia Debelouchina
Characterization of the physical and biochemical basis of protein phase transitions using chemical modulators of protein phase separation and integrative structural and chemical biology techniques.

University of California Santa Cruz, Santa Cruz, CA August 2012 - November 2015
Advisor: Scott Lokey
Investigation of the chemical basis for the passive permeability of non-Lipinskian cyclic peptide macrocycles; determination of the mechanism of action of a bioactive small molecule by whole genome sequencing of resistant mutants.

University of Oxford, Oxfordshire, UK January 2015 - March 2015
Advisor: Stuart Conway
Synthesis of analogs of a dimethylisoxazole-based bromodomain inhibitor.

PROFESSIONAL EXPERIENCE

Mendel Biological Solutions, Hayward, CA November 2015 - August 2017
Advisors: Peter Repetti & Joshua Armstrong
Discovery and development of bioactive natural products for the enhancement of crop performance in agriculture.

EDUCATION

University of California San Diego
Ph.D., Chemistry, 2023

University of California Santa Cruz
B.S., Molecular, Cell, & Developmental Biology with a minor in Bioinformatics, 2015

AWARDS AND HONORS

2022 ACS Graduate Student and Postdoctoral Scholar Recognition Program Award for Leadership in Mentoring

2021 Bruno Zimm Award

2019–2020 San Diego Fellowship

2018–2020 NIH Chemistry-Biology Interface Predoctoral Training Program

SERVICE

2022 ACSSA Undergraduate Research Symposium Judge

2021 Debelouchina Lab UCSD Mentor-Mentee Workshop

2021 Mira Mesa High School ScienceBridge Program Speaker

2021 ACSSA Undergraduate Research Symposium Moderator

2020 UCSD Grad Pals Mentorship Program

2018–2020 ChemPAL Mentorship Program

2012 Juvenile Diabetes Research Foundation Counselor

INVITED AND CONTRIBUTED TALKS

2023 Biophysical Society Annual Meeting (IDP subgroup early career speaker), San Diego, CA

2022 UT Southwestern Biophysics Seminar Series (lightning talk), Dallas, TX

2022 International Council on Magnetic Resonance in Biological Systems (lightning talk), Boston, MA

2021 UCSD Graduate Student Seminar Series, Virtual

2021 San Diego NMR Supergroup, Virtual

2021 San Diego Python Users Group, Virtual

2019 Chemistry Biology Interfaces Symposium, UC San Diego, La Jolla, CA

2019 CBI-CRIN Industry Interaction Day (Lightning Talk), UC San Diego, La Jolla, CA

2018 Chemistry Biology Interfaces Symposium, UC San Diego, La Jolla, CA

OTHER CONFERENCES AND POSTERS

2023 Biophysical Society Annual Meeting (poster), San Diego, CA

2022 International Council on Magnetic Resonance in Biological Systems (poster), Boston, MA

2022 Southern California Users of Magnets (poster), Santa Barbara, CA

2021 Protein Society Annual Symposium (poster), Virtual

2021 Experimental Nuclear Magnetic Resonance Conference (poster), Virtual

2021 UCSD Graduate Student Seminar Series, Virtual

2021 Biophysical Society Annual Meeting (poster), Virtual

2020 Biomolecular Solid-State NMR Winter School, Stowe, VT

2016 PyData San Francisco, San Francisco, CA

2016 HPLC 2016, San Francisco, CA

2015 William S. Johnson Symposium (poster), Stanford University, Palo Alto, CA

PUBLICATIONS

1. Berkeley, R. F., & Debelouchina, G. T. (2022). Chemical tools for study and modulation of biomolecular phase transitions. *Chemical Science*.

2. Berkeley, R. F., Kashefi, M., & Debelouchina, G. T. (2021). Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC. *Biophysical Journal*.

3. Lim, B. J., Berkeley, R. F., & Debelouchina, G. T. (2019). Fused split inteins: Tools for introducing multiple protein modifications. In *Methods mol. biol.* (Vol. 2133, pp. 161–181). New York, NY: Humana.

4. Schwochert, J., Turner, R., Thang, M., Berkeley, R. F., Ponkey, A. R., Rodriguez, K. M., ... Lokey, R. S. (2015). Peptide to Peptoid Substitutions Increase Cell Permeability in Cyclic Hexapeptides. *Org. Lett.*, *17*, 2928–2931.

5. Wride, D. A., Pourmand, N., Bray, W. M., Kosarchuk, J., Nisam, S., Quan, T., ... Lokey, R. S. (2014). Confirmation of the cellular targets of benomyl and rapamycin using next-generation sequencing of resistant mutants in S. cerevisiae. *Mol. BioSyst.*, *10*(12), 3179–3187.

ABSTRACT OF THE DISSERTATION

Biophysical investigation of protein liquid-to-solid phase transitions and their modulation by small heat shock proteins

by

Raymond Berkeley

Doctor of Philosophy in Chemistry

University of California San Diego, 2023

Professor Galia Debelouchina, Chair

Many of the proteins found in the pathological protein fibrils and aggregates that are associated with neurodegenerative disease also exhibit tendencies for liquid-liquid phase separation (LLPS) both *in vitro* and in cells. The transition between the LLPS state and the aggregate state can be modulated by protein chaperones that can block the liquid-to-solid transition and help maintain the LLPS state. In this dissertation, we characterize structural features exhibited by FUS LC, a protein that has been demonstrated to facilitate LLPS and form fibrils, as it undergoes a liquid-to-solid transition from the LLPS state to the fibril state. We subsequently examine the α-crystallin chaperone HSPB1 as it interacts with FUS LC in the LLPS state, and compare the structural features of HSPB1 without a client protein to its

structural features in the presence of a phase-separated and an aggregated client protein. The heterogeneous nature of the chaperone and client system and the wide range of dynamics adopted by the component proteins present unique challenges for structural interrogation. We therefore integrate solid-state magic angle spinning NMR spectroscopy, coarse-grained molecular dynamics simulations, and cryo-EM to build a model for the structures and dynamics adopted by FUS LC during the liquid-to-solid transition and HSPB1 as it interacts with client proteins, with a specific focus on the role of the N-terminal domain of HSPB1 in client recognition and chaperone function.

An introduction to protein self-assembly and structural biology

Introduction and synopsis

Life is the result of a series of complex interactions between four major classes of biomolecules. Each class has historically been assigned a specific role in the cell: Proteins are molecular machines that do work, nucleic acids are information carriers, glycans are platforms for energy storage and load-bearing, and lipids facilitate compartmentalization. However, like all processes in the real world of biology, the biomolecular division of labor is not so clear-cut. The ribosome is perhaps the most prominent example of this–one of the most important cellular machines, responsible for manufacturing all of the protein in the mammalian cell, is constructed from RNA. It is just one of many RNA enzymes[1,2]. Similarly non-canonical functions exist for other nucleic acids, as well as proteins, glycans, and lipids.

Proteins, which are typically functional molecular machines, can also act as signaling molecules[3], structural elements in organelles or in the cell[4,5], carriers of heritable information[6,7], or as scaffolds that facilitate the organization of other biomolecules[8-12]. Many of these non-canonical protein functions are the result of protein self-assembly. Self-assembly enables protein function on length- and timescales that are inaccessible to monomers. Dynamic assembly and disassembly allows for an additional layer of regulation for the behavior of protein self-assemblies that does not exist for individual protein molecules. These features in turn facilitate the non-canonical roles for proteins listed above, which require proteins to act on large length scales and short timescales.

This dissertation is concerned with the structure, dynamics, and function of proteins that sit in-between the traditional definition of proteins as molecular machines, and a burgeoning role of protein assemblies as scaffolds for the organization and concentration of other biomolecules in space and in time. Many of the proteins that exist at this intersection are partially or entirely intrinsically disordered, and their bulk chemistry plays as much of a role in their behavior as the folded structures that they may–or may not–adopt. These intrinsically disordered proteins self-assemble through liquid-liquid phase separation (LLPS), a process that generates a dynamic protein compartment with a discrete phase boundary[10-12]. The unique chemical environment inside an LLPS droplet enables LLPS droplets of different compositions to sequester and concentrate other biomolecules whose properties favor the LLPS environment over the environment outside of the LLPS droplet. LLPS droplets provide a mechanism for fine spatiotemporal control over the location and concentration of other biomolecules in the cell, and thereby act as crucibles for biochemical activity. The mobile and dynamic nature of LLPS droplets and the intrinsically disordered proteins that comprise them provides an environment within which sequestered proteins and other biomolecules can engage in their function.

Proteins that undergo LLPS also often have a tendency to form rigid structures such as aggregates and amyloid fibrils. Although amyloid-like assemblies are sometimes beneficial and can have distinct biological roles[13], amyloid and protein aggregation are more often associated with neurodegenerative diseases including Alzheimer's, amyotrophic lateral sclerosis, and Parkinson's disease[14]. The connection between LLPS of intrinsically disordered proteins and pathological phase transitions into more solid structures has been demonstrated in vitro and is a

possible etiological factor in neurodegenerative disease[15]. The question of whether pathological protein aggregation proceeds through an intermediate LLPS state, occurs from monomeric disordered protein, or is the result of some combination of these two processes is unclear. The impact of understanding the mechanisms that link LLPS and pathological aggregation therefore extend beyond fundamental science, as these processes are implicated in diseases that impact public health.

The goal of this dissertation is to characterize the relationship between protein LLPS and pathological protein aggregation. This includes the study of chaperone proteins that can modulate LLPS and protein phase transitions. By chapter, the major areas of focus are:

- Interactions that drive pathological liquid-to-solid transitions of intrinsically disordered proteins (Chapter 2);
- Changes in structure and dynamics that occur within populations of proteins during liquid-to-solid transitions (Chapter 2);
- The structural features of chaperone proteins that can modulate phase transitions, and the changes in structure and dynamics that occur within these multifaceted proteins as they engage clients in LLPS or aggregated states (Chapters 3 and 4);
- A review of small molecules and other chemical tools that can modulate protein liquid-liquid phase separation and protein liquid-to-solid transitions (Chapter 5).

All of the proteins under study in this dissertation undergo self-assembly, either into LLPS droplets, gels, aggregates, or oligomers that regulate protein function. To face the challenges

inherent in studying systems that form high-order assemblies, we employ a range of chemical tools and structural biology methodologies. Magic angle spinning solid-state nuclear magnetic resonance (MAS NMR) spectroscopy is our method of choice for extracting molecular level information from our samples, which exhibit a high degree of heterogeneity in both structure and dynamics. MAS NMR enables the characterization of these heterogeneous samples at room temperature, and it allows us to capture the changes in dynamics and structure that they exhibit over time. To supplement our NMR studies, we employ coarse-grained molecular dynamics (CG-MD) simulations to capture the behavior of individual protein monomers in LLPS droplets, as well as cryogenic electron microscopy (cryo-EM) to characterize the overall architecture of protein oligomers. The remainder of this introduction will focus on a discussion of proteins, protein folding, and protein self-assembly, along with a brief introduction to each of the structural biology methods employed in this work.

Protein folding in real time

Proteins are polymers of amino acids that are translated into linear chains by the ribosome using an mRNA template. Each amino acid is added to the growing chain via an amide coupling reaction that occurs between the C-terminus of the nascent polypeptide chain, which is linked to the previous tRNA in the P site of the ribosome, and an aminoacyl-tRNA in the A site of the ribosome. Each amino acid is linked via an amide bond that is the result of a dehydration reaction between the amine and carboxylic acid moieties on each amino acid. Resonance in the amide bond confers partial double bond character across the peptide bond. The torsion angles about the N-Cα and Cα-C' bonds are referred to as the φ and ψ angles, respectively (**Fig. 1.1**

). As the chain of amino acids is constructed from within the ribosome, the linear polypeptide extends from the ribosome exit site[16].

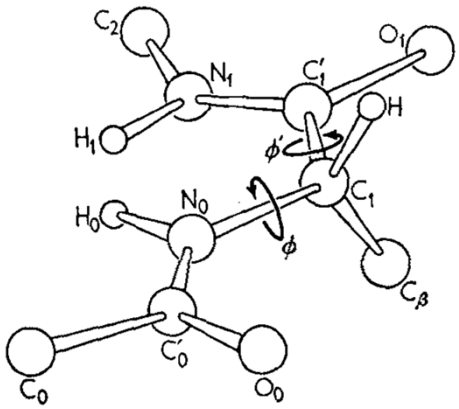


Figure 1.1: The protein backbone as illustrated by G.N. Ramachandran. Note the φ and ψ (here denoted φ') angles. Reprinted from Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. J Mol Biol 7, 95-99 (1963) with permission from Elsevier.

Even before clearing the ribosomal exit site, the polypeptide begins to explore conformations towards the adoption of a final, low-energy conformational state. The final structure of the folded polypeptide is determined by a myriad of processes. Most fundamentally, the folded structure is encoded in the chemistry embedded within the protein's sequence, with the final folded state being that which results in the most thermodynamically favorable arrangement of polar and nonpolar chemistries in the sidechains and peptide backbone of the protein in the water solvent[17]. The planar character of the amide bond and steric hinderance restricts rotation about the φ and ψ angles, limiting the final protein conformation[18].The kinetics of translation can influence the final fold of the protein[19,20], and promote the adoption of a fold that sits at a local or global thermodynamic energy minimum on the protein's folding landscape[21]. The final structure of newly synthesized proteins is also influenced by third parties–usually other proteins that either

install post-translational modifications or engage in protein-protein interactions to bias the nascent peptide chain towards a particular final conformation. The combination of the many processes that contribute to protein folding means that protein folding is not the result of a true random search of conformational space[22], but rather a directed process that proceeds towards a final low-energy folded state[23].

The most basic elements of protein structure are α-helices and β-sheets, collectively referred to as secondary structure. Secondary structural elements are formed via a repeated network of hydrogen bonds between amide protons and carbonyl oxygens in the protein's amide backbone. In an α-helix, an amino acid will form hydrogen bonds with the amino acid four positions up or down the sequence, forming a compact coil like structure in which the peptide backbone forms a spiral (**Fig. 1.2A**). In a β-sheet, the peptide backbone is relatively flat and each amino acid is engaged in hydrogen bonding interactions with an adjacent strand of the protein that can be positioned anywhere in the sequence (**Fig. 1.2B**). Stable β-sheets can form between parallel (i.e. both protein backbones running from the N to C terminus) or antiparallel peptide sequences. The local nature of secondary structure, especially α-helices, means that these structures often form early in the protein folding process, sometimes even in the exit tunnel of the ribosome[24].

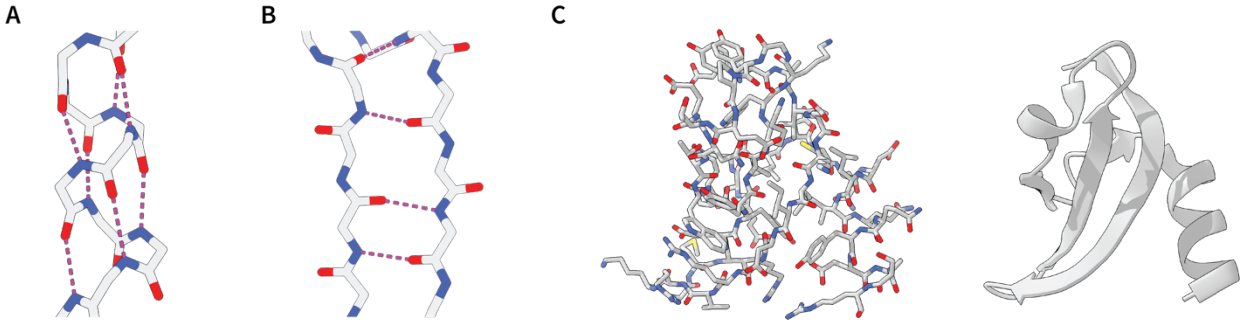


Figure 1.2: Protein Secondary Structure. (A) Backbone hydrogen bonding in an α-helix. **(B)** Backbone hydrogen bonding in a β-sheet. **(C)** Stick and ribbon representation of PDB 1KNA, which contains both β-sheet and α-helical elements.

Tertiary protein structure refers to larger structures that consist of combinations of secondary structure elements (**Fig. 1.2C**[25]). Many tertiary folds have been characterized and common tertiary folds can be observed in proteins with different primary sequences. As proteins begin to adopt secondary structure, it is likely that tertiary structure is accessed through small, partially folded subunits that are referred to as "foldons"[26,27].

The entire process of protein folding generally occurs on millisecond timescales, although some proteins can take much longer to fold into their final state[28,29]. Once folded, proteins are highly dynamic molecules that engage in motion and conformational changes throughout their lifetime. Conformational plasticity is often required for proteins to engage in their desired function. The initial folding process merely marks the beginning of a protein's dynamic existence.

The rise and fall of the structure-function relationship in protein folding

The idea that protein function follows form is central to our understanding of biology. The structure-function paradigm was probably first put into writing by Emil Fisher at the turn of the 20th century, when he compared α-glucoside hydrolysis in beer yeast extracts as requiring an interaction between an enzyme and sugar like a "lock and key"[30]. In the 1920s, the reversibility of protein denaturation and the relationship between protein function and denaturation were demonstrated experimentally[31-34]. This work culminated in a basic theory of protein structure and a structure-function relationship, initially posited by Hsien Wu in 1931[35,36] and subsequently developed in an influential paper by Alfred Mirsky and Linus Pauling in 1936[37]. Further work on globular protein structure followed, most notably on the roles of hydrogen bonding in the

maintenance of α-helices and β-sheets by Linus Pauling[38,39] and the role of hydrophobic interactions in protein structure by Walter Kauzmann[40]. The first generally accepted solution of a protein's structure was performed for myoglobin by John Kendrew and coworkers in 1958 and 1960[41,42].

Most natural proteins have a free energy landscape with pronounced minima that represent folded, low energy states. The transition from the unfolded nascent polypeptide to the final folded state is thought to be bridged by an intermediate "molten globule" state that is rich in secondary structure but still searching for the most favorable tertiary structure represented in the final folded state[43-48]. Until the late 1980s, the field of structural biology focused on these folded states, and generally adhered to Anfinsen's dogma, which states that protein structure is determined by the amino acid sequence alone, and each sequence corresponds to a single, low-energy functional structure that is kinetically accessible and reasonably stable[17]. The idea that folded protein structure is necessary for protein function is a logical one. The widespread use of crystallography for characterizing biomolecular structure amplified the focus on folded protein structure.

As the field of structural biology further came into its own, however, it became clear that not all proteins respected Anfinsen's conclusions about protein folding–many proteins did not have a low-energy folded state at all, and instead retained their lack of order after translation while also carrying out biological functions in this disordered state[49]. Functional intrinsically disordered proteins (IDPs) go against the paradigm that protein structure confers function, and instead

suggest that conformational ensembles, or the chemical properties of populations of proteins in the bulk phase, can be an essential factor in protein function and behavior.

IDPs have turned out to be ubiquitous in all areas of biology. Rather than classifying proteins as IDPs or folded proteins, it is more accurate to say that all proteins exist along a continuum of structure, with entirely disordered IDPs on one end and highly structured proteins on the other. Most proteins contain domains that exist at various locations along this continuum[50]. Predictions of disorder in eukaryotic proteomes suggest that more than 30% of proteins contain disordered regions of 50 or more residues[51]. Measured by expression level, proteins containing disordered regions are highly represented in some of the most abundant classes of protein in the cell, including transcription factors, proteins involved in packaging DNA, and subunits of the ribosome[52].

Without a structural component to satisfy the structure-function paradigm, IDP behavior must be characterized through the ensemble of conformational states sampled by the IDP. Often, when IDPs participate in protein-protein interactions, they undergo a transition from disorder to order, also known as "folding upon binding"[53-56]. Disorder-to-order transitions enable behaviors that go beyond the one-to-one lock-and-key binding paradigm, such as allowing a single protein to bind to multiple targets[57], or facilitating protein trafficking behavior that would not be accessible to the folded state[58,59].

Other IDPs engage in protein-protein interactions while retaining their disorder[60]. There are numerous examples of biologically significant interactions between folded proteins and

disordered binding partners[61-64]. In extreme cases, high-affinity protein-protein interactions occur between two IDPs which retain their disorder during binding[65]. The fact that IDPs engage in biologically meaningful interactions while retaining their disorder has forced a reevaluation of the structure-function paradigm and broadened our understanding of how proteins can adopt their role as molecular machines in the cell.

Mesoscale protein assemblies and liquid-liquid phase separation

Higher-order assemblies of IDPs can take on the role of scaffolding and compartmentalization platforms for other biomolecules. In cells, biomolecular organization is usually achieved through lipid bilayer membranes, which form boundaries that can separate groups of biomolecules into distinct areas. Compartments created by lipid bilayers often contain groups of biomolecules that enact a particular cellular function as organelles. To enter or exit an organelle, a molecule must cross the lipid bilayer that encloses it.

There are some organelles, like nucleoli, that are not encapsulated by membranes. The structural integrity of these membraneless organelles is maintained by networks of interactions between their constituent biomolecules, rather than a lipid bilayer. In 2009, it was demonstrated that P granules, a class of membraneless organelle, have liquid-like properties[66]. For example, P granules can wet surfaces, flow, fuse, and drip. Additionally, the formation of P granules was shown to occur through condensation and dissolution. Further investigation of similar membraneless organelles has revealed that many of these structures form as the result of protein liquid-liquid phase separation (LLPS)[10-12]. The recent recognition of the fact that membraneless

organelles form through the dynamic process of LLPS, rather than a more static means of assembly, has brought on a deep shift in our understanding of biomolecular organization and regulation in the cell.

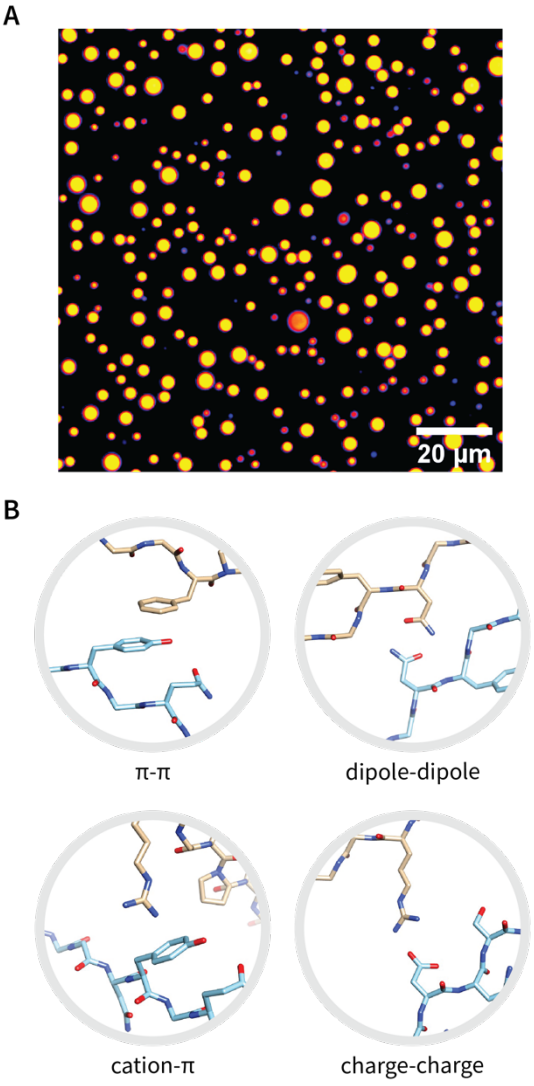


Figure 1.3: Liquid-liquid phase separation. (A) Fluorescently labeled LLPS droplets *in vitro*. **(B)** The four categories of intermolecular interactions that drive protein LLPS.

In protein LLPS droplets (**Fig. 1.3A**), transient multivalent interactions between proteins maintain the integrity of the droplet rather than an encapsulating lipid bilayer. The nature of these interactions produces a liquid-like, mobile environment within the LLPS droplet. The low density and high permeability of LLPS droplets promotes the inclusion of other biomolecules that may prefer the chemical environment within the LLPS droplet[67]. IDPs and disordered regions within proteins are often drivers of LLPS. IDPs are enriched in hydrophilic residues that engage in π-π, dipole-dipole, cation-π, and charge-charge interactions that support LLPS (**Fig. 1.3B**)[68,69]. Environmental changes or post-translational modifications on IDPs can induce phase transitions that enable a layer of organizational regulation that would not be achievable with a membrane-bound compartment. The ability of LLPS to rapidly compartmentalize biomolecules means that it is often associated with processes that require finely-tuned spatiotemporal control, like DNA damage repair, regulation of transcription and gene expression, and mRNA processing.

In the case of LLPS, the role of proteins that facilitate droplet formation extends beyond that of molecular machines. IDPs that drive LLPS behave as scaffolds for the recruitment and concentration of third-party biomolecules, forming dynamic, mesoscale protein assemblies that act as crucibles for biochemical reactions[10-12]. The dynamic nature of LLPS droplets, and their ability to condense and dissolve in response to stimuli, provides cells with a system that can quickly react to environmental demands.

Pathological protein aggregation and liquid-to-solid phase transitions

The conformational plasticity of IDPs is driven by the fact that interactions of IDPs with the solvent are favorable–IDPs are often enriched in hydrophilic residues, or they do not contain hydrophobic residues with sequence distributions that promote the formation of the hydrophobic cores present in globular proteins. This does not mean that it is impossible for IDPs to fold or to form rigid structures–the conformational landscape simply contains fewer or less stable low-energy conformations. If the solvent environment is not favorable, or if a population of IDPs samples a low-energy folded state to a sufficient degree, conformational collapse to a rigid, folded state can occur.

Since the sequence of an IDP does not predispose it towards a folded state, the most stable static conformation for most IDPs is typically a β-sheet mediated by backbone hydrogen bonds rather than structure encoded in amino acid sidechain chemistry[70]. Rigid β-sheet-rich aggregates can form self-templating fibrillar assemblies that promote fibril formation by other copies of the IDP. If a stable β-sheet is not sufficiently favored, or if an IDP experiences unfavorable conditions but does not have time to sample the fibril state, the protein can collapse into a conformationally amorphous but dynamically rigid state. This process is referred to as protein aggregation. If conditions disfavor the soluble disordered state for an IDP, the progression towards a thermodynamically favorable β-sheet-rich structure like an amyloid fibril or a kinetically favorable amorphous aggregate will depend on whether or not the system can overcome the free-energy barrier inherent in the adoption of a low-energy fibril state[71]. For both amorphous aggregates and β-sheet fibrils, the dense network of hydrogen bonds and hydrophobic

interactions that form the foundation of these self-assemblies produces very stable structures that are difficult to reverse or disassemble. The presence of irreversible protein aggregates and fibrils in the brain are hallmarks of numerous neurodegenerative diseases[72].

The set of sequences in IDPs that promote LLPS overlaps significantly with the set that can lead to protein aggregation. Indeed, most of the proteins found in pathological aggregates in neurodegenerative diseases, like tau[73], α-synuclein[74], TDP-43[75], and FUS[15,76] undergo LLPS both in vitro and in cells. A significant body of evidence that LLPS droplets may act as nucleation points for pathological aggregation, or that LLPS and aggregation are linked in some other way, has emerged over the last few years[15,74-83]. This relationship is not only significant in that it represents a novel and biologically relevant route to pathological fibril formation, but also in that it represents a disconnect between traditional approaches to fibril preparation in vitro and biological modes of fibril formation. Understanding how the formation of fibrils proceeds in the biologically relevant context of LLPS, coupled with new approaches for studying patient-derived protein fibril samples[84-87], could provide insights into fibril formation and neurodegenerative disease etiology that were inaccessible with fibrils generated *in vitro* from seeds. Characterizing the changes in IDP structure and dynamics that occur during the liquid-to-solid transition described here is the primary focus of Chapter 2 of this dissertation.

Molecular Chaperones

The pathological nature of protein aggregation and the error-prone nature of protein folding in general is addressed by a class of proteins called molecular chaperones. Some molecular

chaperones actively influence protein folding by hydrolyzing ATP and kinetically partitioning partially folded proteins towards their folded states[88-91]. This class of molecular chaperones is usually represented by structurally complex oligomeric protein machines such as HSP10/HSP60, HSP70, and HSP90. Although these molecular chaperones are powerful facilitators of protein folding, they require significant energy investment to perform their function. They also cannot engage a structurally diverse set of clients–they often interact with nascent proteins at or just downstream of the ribosome, or they require cooperation with other proteins to engage specific misfolded clients in the cytosol. Examples of this cooperative behavior can be found with HSP70 and the wide array of HSP40 facilitator proteins in eukaryotes[92-94], or the priming of aggregated proteins by small heat shock proteins for disaggregation by HSP70[95,96]. In both cases, however, HSP70 requires a second protein to facilitate its function.

ATP-independent molecular chaperones have simpler structures and functions compared to their ATP-dependent counterparts. While ATP-dependent chaperones can actively disaggregate or facilitate the folding of client proteins, ATP-independent chaperones primarily bind to unfolded or misfolded proteins in a passive manner. In this aspect they can be considered to act as buffers that physically block or modulate aberrant folding while priming misfolded proteins for active refolding by ATP-dependent molecular chaperones. For this reason, ATP-independent chaperones are often referred to as “holdases”, a designation that perhaps does a better job of capturing their role in the cell.

The most common class of ATP-independent molecular chaperones are the small heat shock proteins (sHsps). The simplicity of sHsp function is mirrored in the simple structure and domain

architecture that they exhibit: all sHsps have a tripartite domain architecture with an N-terminal domain (NTD) of variable length that is generally rigid but disordered, a folded α-crystallin domain (ACD) that is highly conserved across sHsps, and a disordered and dynamic C-terminal domain (CTD). In the absence of client proteins, sHsps self-assemble into pseudosymmetric oligomers that form polydisperse populations. Although sHsps have a simple structure and domain organization, the complexity of these polydisperse oligomers presents challenges for structural studies. A full discussion of the structure and organization of sHsps in the absence of client proteins can be found in the introduction to Chapter 3 of this dissertation.

The ubiquitous nature of sHsps and their passive, ATP-independent mechanism of action positions them in a central role in the protein quality control machinery. As "first-responders" in the unfolded protein response, sHsps can engage a diverse array of unfolded or misfolded clients. Emerging evidence suggests that different sHsp domains are responsible for engaging clients in different classes of self-assembly, with the disordered but rigid NTD being responsible for engaging misfolded, aggregated clients and the β-sheet rich ACD engaging fibrillar, β-sheet-rich clients[97-100]. Additionally, there is functional evidence that sHsps interact with aggregation-prone proteins in their disordered, monomeric state[101], and can directly modulate the LLPS state[97,102-105]. The role of sHsps as protective agents in models of neurodegenerative disease suggests that their ability to prevent against aberrant liquid-to-solid transitions of disease-associated proteins is linked to their role in preventing neurodegenerative disease[106-109]. The changes in dynamics that sHsps exhibit when interacting with different types of clients and the contributions of different domains to its chaperone activity are the subject of Chapter 4 of this dissertation.

The structural biology toolbox

Fundamental discoveries in structural biology are often propelled by advances in technology and the development of new methods for examining biomolecular structure and dynamics. The protein systems examined in this dissertation exhibit a wide range of sizes and dynamics–from small, mobile, disordered proteins in LLPS droplets to massive, rigid protein assemblies like those found in aggregates and amyloid fibrils. We employ a range of structural biology methodologies to face the challenges inherent in the biological systems under study, most notably nuclear magnetic resonance spectroscopy, cryo-electron microscopy, and molecular dynamics simulations. In this section of the introduction, we provide a brief primer for each of these methods, and an overview of how each method contributes to the characterization of the biomolecular systems in this dissertation.

Nuclear magnetic resonance spectroscopy

The primary method of investigation in this dissertation is magic angle spinning solid-state nuclear magnetic resonance spectroscopy (MAS NMR). Although there is no one-size-fits-all methodology in structural biology, the versatility of MAS NMR makes it the ideal choice for characterizing the structurally and dynamically diverse biomolecular systems under study here. For example, MAS NMR is the only technique capable of delivering information at atomic resolution for biomolecules occupying the broad ranges of dynamics exhibited by samples undergoing liquid-to-solid transitions[110-112]. In addition to the ranges of dynamics that can be characterized by NMR, the ability to dissect heterogeneous samples using experiments that filter

on dynamics, and to do so at room temperature, is unique to MAS NMR and extremely valuable for real-world biological samples. The flexibility of NMR underscores its value as a probe for biologically relevant samples that recapitulate the wide range of structure and dynamics found in the milieu of a living cell.

The NMR experiment is enabled by the fact that atomic nuclei possess an intrinsic quantized angular momentum, or spin, as well as an intrinsic magnetic moment[113,114]. The relative magnitude of the magnetic moment of a nucleus is described by a property called the gyromagnetic ratio, or γ, which is parallel (for nuclei with a positive γ) or antiparallel (for nuclei with a negative γ) to the spin orientation[115]. In the presence of an external magnetic field, nuclear spins will align with or against the magnetic field and precess around it at a frequency proportional to γ, a process referred to as Larmor precession. The degree of alignment and the population of spins aligned with or against the field is dependent on the gyromagnetic ratio and is proportional to the strength of the external magnetic field (**Fig. 1.4**). Small differences in the interaction energy between nuclear magnetic moments, whose orientation is determined by their spin state, and the external magnetic field exist due to differences in the incidence angle of this interaction. Since spin states are quantized, these energy differences are discrete and therefore manifest as splitting in observed energy levels, a phenomenon known as the Zeeman effect[116].

For a spin-½ nucleus, the NMR signal represents the difference between two spin states, “spin up” and “spin down” along the alignment vector. The energy difference between these two spin states is described by the following equation[112]:

$$\Delta E = \gamma\hbar(1 - \sigma)B_0 \quad (1)$$

Here, ΔE is the energy difference; γ is the gyromagnetic ratio, a fundamental property of every isotope describing its precession frequency at a given magnetic field strength; $\hbar$ is the reduced Planck's constant; σ is the chemical shielding constant, a value that quantifies the effect on the magnetic field around a given nucleus by its surrounding electrons; and B_0 is the strength of the external magnetic field[112,113]. Differences in the gyromagnetic ratio of different isotopes can dictate their sensitivity in NMR experiments. A number of spin-½ isotopes are available for nuclei abundant in biomolecules, with the most useful being ^{1}H ($\gamma = 267.5 \cdot 10^6$ rad $\cdot s^{-1} \cdot T^{-1}$, or 500 MHz at 11.74 T), ^{13}C ($\gamma = 67.3 \cdot 10^6$ rad $\cdot s^{-1} \cdot T^{-1}$, or 125.7 MHz at 11.74 T), and ^{15}N ($\gamma = -27.1 \cdot 10^6$ rad $\cdot s^{-1} \cdot T^{-1}$, or 50.7 MHz at 11.74 T)[113]. Although the natural abundance of ^{1}H is 99.9%, ^{13}C is 1.1%, and ^{15}N is 0.3%[117], the low receptivity of ^{13}C and ^{15}N can be circumvented by recombinantly expressing proteins in ^{13}C- and ^{15}N-enriched media and thereby uniformly labeling protein with NMR-active nuclei. Approaches to selective[118] or segmental labeling[119] also exist that can be used as a tool to provide more detailed insight into the behavior of complex biomolecules[120].

In practical terms, Equation 1 tells us that the measurable energy difference in an NMR experiment scales with the magnetic field strength and that changes in the electronic environment around a given nucleus can induce small changes in the behavior of that nucleus. The measurable change in behavior is perhaps more intuitive if we reframe the equation in frequency terms by removing the reduced Planck's constant:

$$\omega_0 \nu = \gamma(1-\sigma)B_0 \quad (2)$$

Here, ω_0 ν represents the precession frequency of a nucleus under the influence of an external magnetic field B_0 and chemical shielding σ. To decouple measured ω_0 values from the external magnetic field strength, NMR spectra are usually presented as chemical shift values in parts per million (ppm) away from a reference compound (usually tetramethyl silane for solution NMR and adamantane for MAS NMR) at the external magnetic field strength. This facilitates the comparison of NMR spectra acquired on different magnets operating at different magnetic field strengths.

The spin polarization anisotropy (the actual ratio of nuclei in each spin state) for a given ΔE, which determines the amount of measurable signal in an NMR experiment, can be calculated using the Boltzmann distribution[112,113]:

$$\frac{N_+}{N_-} = e^{-\frac{\Delta E}{kT}} = e^{-\frac{\gamma\hbar(1-\sigma)B_0}{kT}} \quad (3)$$

Equation 3 tells us that magnitude of the spin polarization anisotropy (the ratio of "spin up" and "spin down" nuclei for a spin-½ nucleus) is proportional to the ratio of magnetic (via ΔE) and thermal (T) energies. The magnitude can therefore be increased by increasing the external magnetic field B_0 or reducing the temperature T. Unfortunately, the small energy difference between the spin up and down states means spin polarization anisotropy at commercially available external magnetic field strengths is still very low, especially when compared to electron spin anisotropy[112,113]. For example, most of the experiments in this dissertation were performed

at a magnetic field strength of 750 MHz ^{1}H Larmor frequency, or approximately 17.6 Tesla, at 275 K. Evaluating the equation for the Boltzmann distribution with these parameters reveals that the ^{1}H spin polarization anisotropy corresponds to only about 1 in 7,700. The situation is even worse for ^{13}C and ^{15}N, with spin polarization anisotropies of approximately 1 in 30,100 and 1 in 75,400, respectively.

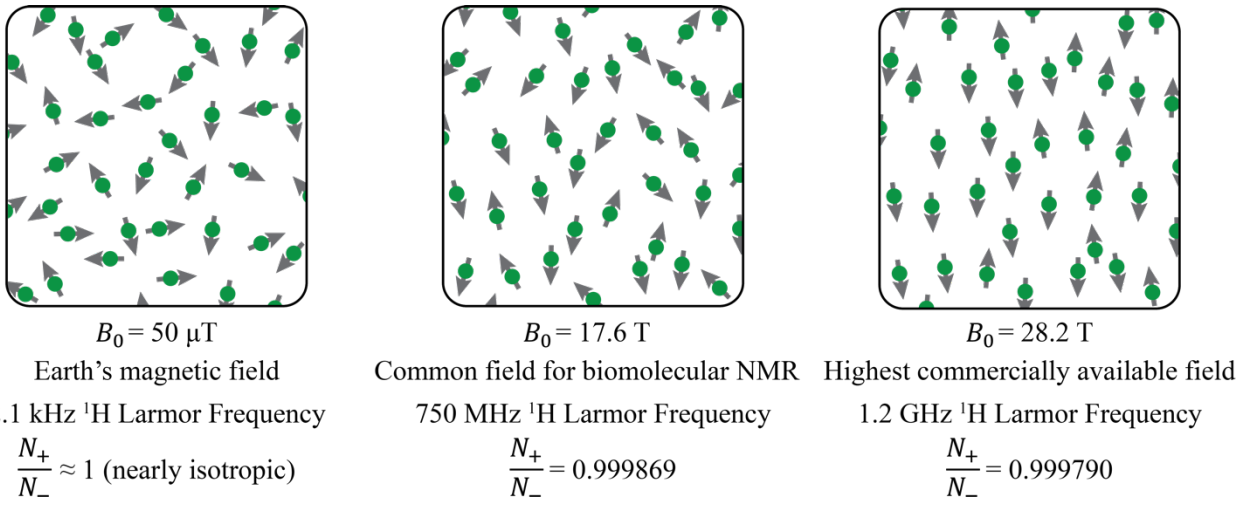


Figure 1.4: The influence of magnetic fields on nuclear spin polarization anisotropy. The degree of anisotropy is amplified for the purpose of illustration.

Even in the presence of a strong external magnetic field, magnetization from nuclear spin polarization anisotropy along the axis of B_0 is very weak–so weak that it is difficult to detect and impractical to use for routine experiments. To circumvent this, NMR experiments are performed by detecting spin magnetization in the transverse plane.

In a basic NMR experiment, nuclear spins precessing at ω_0 are manipulated by irradiation with electromagnetic energy at a resonant frequency. For most biomolecular NMR experiments, the external magnetic field strengths that are used place this frequency in the radiofrequency (RF)

regime. Starting from a population of spins at thermal equilibrium, with a magnetization vector along the axis of B_0, an RF pulse is applied whose frequency is set to match the Larmor frequency of the nuclei being manipulated. The amplitude and duration of the pulse is tailored to provide the energy required to move the magnetization 90° from the B_0 axis, a process called excitation. After excitation, spin magnetization returns to the equilibrium state over time through a process referred to as relaxation. The precession of the nuclear magnetization vector that occurs in the transverse plane as spins relax produces an electromagnetic field that produces current in a coil surrounding the sample. This current is the signal that is measured in an NMR experiment.

Relaxation is straightforward to understand when presented on the Bloch sphere, a polar coordinate set that can be used to describe the magnetization vector generated by a population of spins (**Fig. 1.5A**). In the context of our basic NMR experiment, the axis of the external magnetic field B_0 lies along Z, and our signal is measured in the transverse plane that spans X and Y. Figure 1.5A uses the Bloch sphere to illustrate the basic 90° pulse NMR experiment described above, where an RF pulse rotates spin magnetization from Z to Y, and the magnetization gradually relaxes back to thermal equilibrium along Z. Two distinct processes are responsible for spin relaxation. The first is longitudinal relaxation, described by the time constant T_1 (**Fig. 1.5B**). This process represents the return of spin magnetization to thermodynamic equilibrium along Z after excitation. The second process is transverse relaxation, described by the time constant T_2 (**Fig. 1.5C**). T_2 relaxation represents the decoherence, or dispersion, of spin magnetization in the transverse plane. This type of relaxation is the consequence of variation between the magnetic microenvironments experienced by the individual nuclei that make up the population of spins in

our sample. T_1 and T_2 times are intimately linked to the dynamics of the system at hand. Because T_1 describes the return of magnetization to the B_0 axis, it is affected by molecular rotational correlation times. Nuclei that are tumbling or moving at frequencies near their Larmor frequency at a given magnetic field strength will have fast T_1 times, and T_1 will decrease as the rotational frequency diverges from the Larmor frequency.

Molecular motion also has a large effect on T_2 relaxation, which is the direct result of the influence of the local environment of a spin. Because the influence of the local interactions that impact T_2 (which are discussed later in this chapter) tends to be more pronounced in rigid systems, T_2 relaxation is more rapid in large or rigid protein assemblies. The relationship between molecular motion and T_2 relaxation can be useful for interpreting macromolecular dynamics, both for global motion (like tumbling) and local motion within macromolecules.

The combined effect of T_1 and T_2 relaxation on a population of excited spins is a signal whose oscillation frequency and rate of decay encode information about both the chemical environment of the precessing nuclei in this population of spins and the dynamics of the molecules that contain them (**Fig. 1.5D**). This signal is detected in the transverse plane as a "free induction decay" (FID)–a decaying signal that represents the sum of the precession frequencies of the spins in our sample. This signal can be transformed from the time to frequency domain (usually using a Fourier transform) to provide an interpretable picture of the spin precession–and, by proxy, the chemical environment and dynamics–of the spins in our sample.

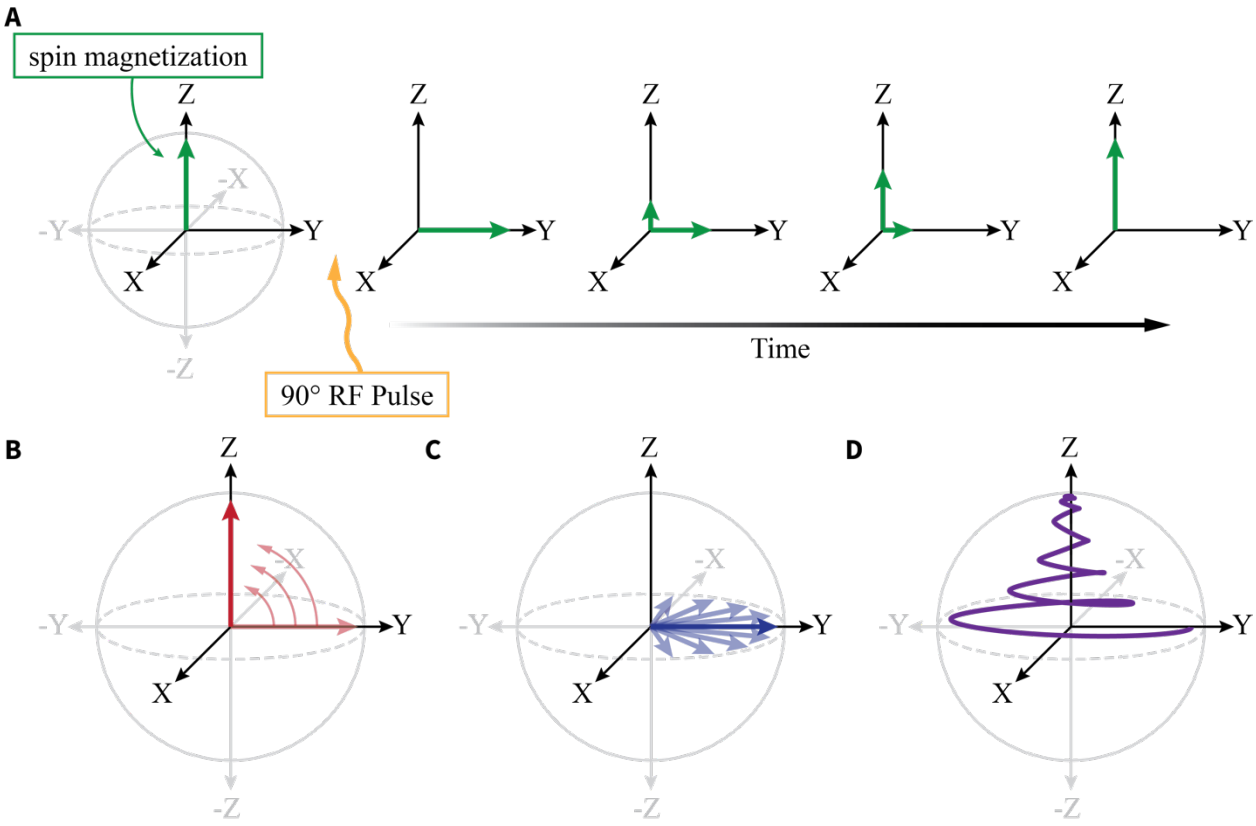


Figure 1.5: Relaxation on the Bloch sphere. (A) Magnetization vector over time after the application of a 90° pulse. **(B)** Longitudinal relaxation. **(C)** Transverse relaxation. **(D)** Magnetization trajectory under the influence of both longitudinal and transverse relaxation.

Multidimensional nuclear magnetic resonance spectroscopy

The first NMR experiments characterized the chemical environment within samples through NMR absorption by delivering RF pulses at fixed frequencies and sweeping the external magnetic field strength[121,122]. Soon after these experiments, improvements in hardware delivered better magnetic field homogeneity which in turn enhanced resolution[123]. This improved resolution enabled finer detection of chemical shielding effects (σ) and the ability to detect and manipulate spin-spin couplings between nuclei[124]. This advance provided the opportunity to perform NMR experiments that report upon connectivity and distance between individual nuclei

by selectively detecting spin-spin couplings with different frequencies and constructing an additional spectral dimension through the information encoded therein.

Spin-spin couplings are particularly useful in that they can provide information about distances between spins and connectivity within molecules. Multidimensional NMR experiments can be performed to extract this information. Multidimensional NMR experiments are composed from a series of one-dimensional experiments that iterate across a second frequency parameter (the evolution time) in order to capture interactions between the coupled nuclei. A Fourier transform across the first (detected) dimension produces what is effectively an absorption spectrum in the second dimension (**Fig. 1.6**). A 2D Fourier transform produces a contour plot with signal corresponding to coupled nuclei along each frequency axis (**Fig. 1.6**).

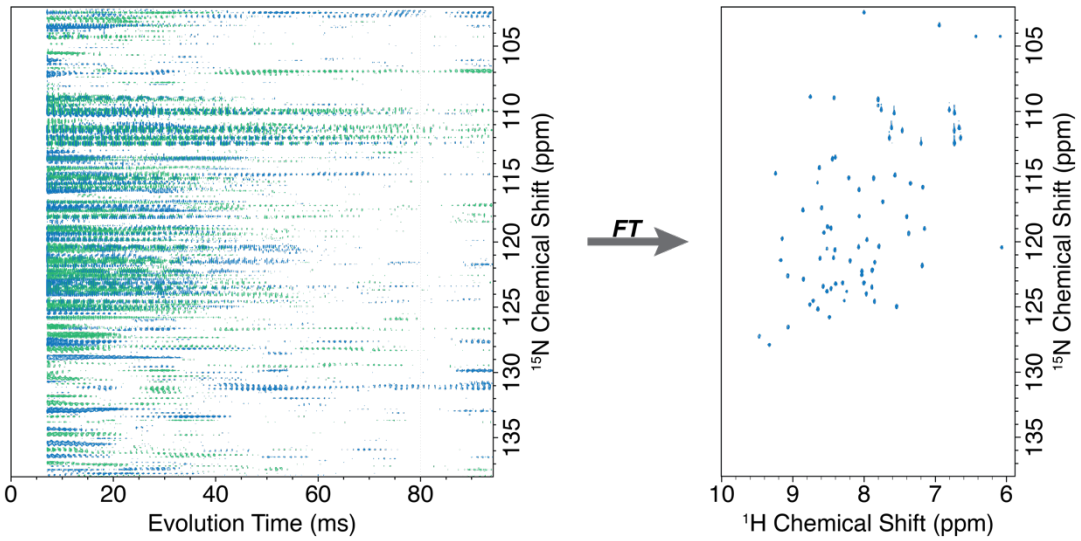


Figure 1.6: A ^{1}H-^{15}N HSQC 2D NMR spectrum after 1D and 2D Fourier transformation. Note the evolution time axis on the 1D FT spectrum (left).

Solid-state nuclear magnetic resonance spectroscopy and magic angle spinning

In Equation 2, we showed that precession frequency ω_0 is determined by the gyromagnetic ratio γ, the external magnetic field strength B_0, and the chemical shielding constant σ. Since the gyromagnetic ratio is a fixed property of an isotope, and the magnetic field is generally homogeneous in biomolecular NMR experiments, the chemical shielding constant is often the primary source of information about the chemical environment of a given population of spins in an NMR experiment. For proteins, chemical shift values derived from slight changes in chemical shielding correlate strongly enough with unique chemistries in amino acids to make confident assignments of amino acid types. Supplementing these assignments with the connectivity information provided by higher dimensional NMR experiments often allows for the unambiguous assignment of atoms in proteins.

Although Equations 1 and 2 are sufficient to describe the energy of an isolated spin in an NMR experiment, other interactions can influence the local magnetic environment around an atom. In the complex world of biomolecular NMR, it is rarely the case that an isolated spin is being considered. For spin-½ isotopes, one of the most useful interactions is the dipolar coupling interaction, which occurs through space between the magnetic field of two nuclear spins. Its strength is proportional to the gyromagnetic ratio and scales as the cube root of the internuclear distance, making it a valuable source of distance information between spins. In isotopes with more complex magnetic moments, through-space quadrupolar interactions also exist. Lastly, *J*-coupling interactions, mediated by the electrons in covalent bonds between nuclei, also exist and can be leveraged to extract information about through-bond connectivity in molecules. The sum

of these interactions can be used to describe the total energy experienced by spins in an NMR experiment, which can be expressed as the NMR Hamiltonian:

$$\widehat{H}_{NMR} = \widehat{H}_{Zeeman} + \widehat{H}_J + \widehat{H}_{CS} + \widehat{H}_{DD} + \widehat{H}_Q \quad (4)$$

The relative orders of magnitude of each interaction are important for understanding the role of each in the NMR experiment:

Table 1.1: Properties of NMR Hamiltonains

Hamiltonian	Name	Magnitude (Hz)	Isotropy
$\widehat{H}_{Zeeman}$	Zeeman	10^8	Isotropic
$\widehat{H}_Q$	Quadrupolar	10^6	Anisotropic
$\widehat{H}_{CS}$	Chemical Shift	10^3	Isotropic and anisotropic
$\widehat{H}_{DD}$	Dipole-Dipole	10^3	Anisotropic
$\widehat{H}_J$	J	10	Isotropic

It is important to note that many of these interactions are anisotropic, which means that their magnitude is dependent on the orientation of the interaction with respect to the external magnetic field. For a population of spins situated at random orientations with respect to the external magnetic field, it follows that the effect of these interactions would span a range of magnitudes. The result of this is an NMR spectrum with broad lines that are the result of the wide range of precession frequencies exhibited by spins in the sample.

Small soluble proteins often tumble in solution at frequencies higher than the magnitude of anisotropic chemical shift and dipole-dipole interactions. This means that the line-broadening effects of these interactions do not manifest in an NMR spectrum as the broadening induced by them is effectively "averaged out". The result are peaks in the NMR spectrum that are centered at the chemical shift (**Fig. 1.6**). As rotational correlation time increases, however, anisotropic interactions are no longer averaged and spectra become difficult or impossible to interpret. As a rule of thumb, this limitation precludes the use of most standard solution NMR experiments for proteins with a molecular weight higher than approximately 50 kDa.

To circumvent this issue, we can employ magic angle spinning (MAS) to remove the undesirable effects of anisotropic interactions[112,125,126]. This is achieved by taking advantage of the fact that the magnitude of the relevant anisotropic interactions have a $3\,cos^2\,\theta - 1$ relationship with the external magnetic field, where θ is the angle between the internuclear axis (for dipolar couplings) or the axis with the largest deshielding value (for the anisotropic component of the chemical shift) and the external magnetic field. Setting θ to an angle that sets this relationship to zero reduces the anisotropy from three to two dimensions, allowing simple spinning along the θ axis to completely average the anisotropy in three dimensions. This angle is 54.74°, often referred to as the "magic angle" (**Fig. 1.7A**). Just like isotropic tumbling in solution NMR, the MAS frequency must be greater than the magnitude of the anisotropic interactions for effective averaging to occur. The line broadening effects of anisotropic NMR interactions, and their resolution by MAS, can be seen in the first reported double-resonance MAS experiment depicted in the figure below (**Fig. 1.7B**)[126].

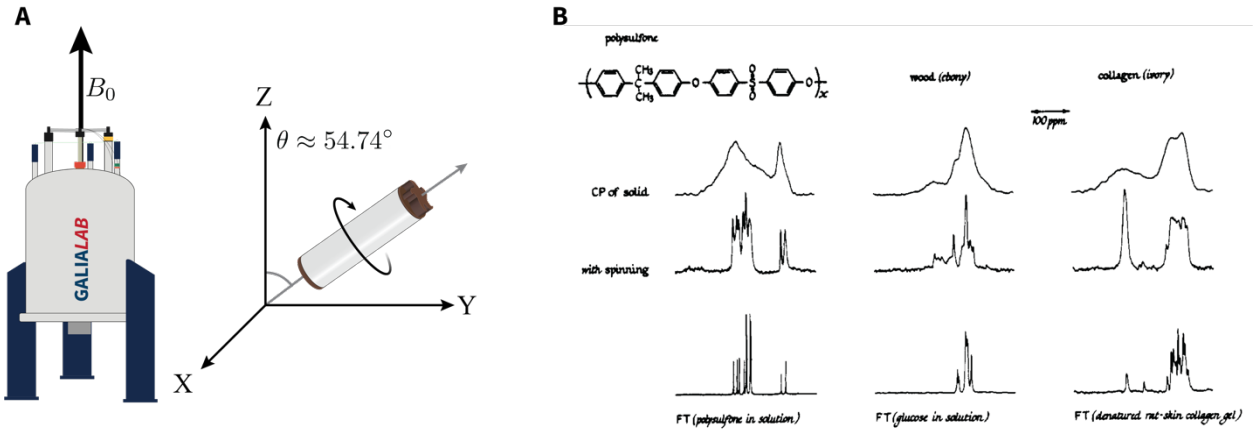


Figure 1.7: Magic-angle spinning. A) The magic angle and its orientation with respect to the external magnetic field B_0. The angle is depicted using an MAS rotor. **B)** An early example of solid-state NMR spectra of polymers with and without MAS compared to solution NMR spectra of the component monomers. Adapted with permission from Schaefer, J. & Stejskal, E. Carbon-13 nuclear magnetic resonance of polymers spinning at the magic angle. *Journal of the American Chemical Society* 98, 1031-1032 (1976). Copyright 1976 American Chemical Society.

The control over anisotropic dipolar and chemical shift interactions afforded by MAS offers a few key benefits over solution NMR. The first is the obvious fact that it enables the analysis of large rigid systems by NMR. Even though MAS NMR experiments are performed in a different format (i.e. in a MAS rotor), solution NMR experiments can be performed under MAS conditions with only minor modifications. This key advantage means that systems that exhibit heterogeneous dynamics can be dissected using a single MAS NMR sample. In this dissertation we take full advantage of this by examining systems that undergo time-dependent liquid-to-solid transitions from LLPS droplets to aggregates and fibrils, a process that involves broad transitions across regimes of structure and dynamics that would be impossible to characterize by any other method.

To characterize systems that exhibit heterogeneous dynamics, we can employ MAS NMR experiments that leverage polarization transfer steps that are only efficient for spins adopting a

particular regime of dynamics. To examine dynamic components within a heterogeneous system, such as those that are moving on timescales that could be associated with a liquid, we can use experiments based on the Insensitive Nuclei Enhanced by Polarization Transfer (INEPT) pulse sequence[127]. INEPT-based experiments leverage *J*-coupling interactions to establish correlations between coupled spins and to provide polarization enhancement for dilute spins such as ^{13}C or ^{15}N by transferring polarization from high-γ spins such as ^{1}H. This means that correlations are observed through bonds, and 1D, 2D or 3D spectra will only have peaks that correspond to spins coupled through a covalent bond. Additionally, the evolution time necessary for INEPT polarization transfer to occur is longer than the T_2 relaxation time for rigid systems–this means that INEPT-based experiments are selective for dynamic components of our sample, while rigid spins are invisible in the NMR spectrum.

The complement to INEPT-based pulse sequences are cross-polarization (CP)-based experiments[128,129]. CP leverages polarization transfer through the dipolar coupling interaction to provide polarization enhancement and correlation information between spins. Because the dipolar coupling interaction is only measurable for systems engaged in dynamics that are slower than the magnitude of the dipolar interaction (**Table 1.1**), the CP experiment is not efficient for dynamic systems and therefore acts as an effective filter for rigid components of our NMR sample. In this dissertation, we use CP-based experiments extensively to characterize the rigid subpopulations of proteins undergoing liquid-to-solid transitions (Chapter 2) and the rigid components of high-order oligomers of chaperone proteins (Chapter 4).

Molecular dynamics and coarse graining

Molecular dynamics (MD) simulations are a valuable complement to empirical structural biology studies. With MAS NMR, distances between atoms and general regimes of dynamics can be calculated which provide a population-level description of the behavior of biomolecular systems. Simulations are often used to produce snapshots of the behavior of individual molecules based on the bulk measurements performed by other empirical methods in order to yield a comprehensive picture of the structure and dynamics of biomolecules.

Molecular dynamics simulations use forces derived from first principles or from experimental observation to calculate the motion of a molecule in discrete timesteps. Generally, this involves calculating the forces acting on each individual atom in a system and modeling the changes in position and motion that those forces import. More modern approaches involve the integration of force fields that describe the sum of a number of different forces acting on every atom in a biomolecular system. Due to the sheer number of calculations required to simulate the complex array of forces that govern biomolecular motion, MD simulations are often computationally intensive. For example, a single NVIDIA RTX 3090 can compute about 13 ns of simulation time per day for the ribosome, a system with about 2 million atoms, using GROMACS 2022[130-132]. This limitation makes it difficult to simulate large biomolecular systems such as protein self-assemblies, or to capture rare events.

Strategies exist to counter the computational restraints imposed on MD. To capture rare events in a reasonable amount of time, enhanced sampling methods can be employed to introduce bias

against frequently sampled states in a simulation[133]. Enhanced sampling approaches generally either lower potential energy barriers between rare states, therefore increasing the probability that a structure will transition between low energy states (a notable example of this approach is the umbrella sampling approach[134]) or increasing the free energy in the system and selecting rare conformations as starting points for continued simulation (a notable example of this approach is replica exchange molecular dynamics[135]). To simulate large systems, tradeoffs in accuracy can be made. The most common approach here is coarse graining, which involves treating groups of atoms–usually every amino acid for proteins–in a biomolecule as a single entity and establishing a collective set of interaction parameters for that entity. This allows for the simulation of larger systems by reducing the interactions experienced by all atoms in an amino acid in a protein down to a single parameter set.

In this dissertation, we simulate proteins in the LLPS state using a coarse-grained model that has been shown to be capable of recapitulating phase separation *in silico*[136]. Modeling of LLPS is achieved using only three potentials. Bonds between the beads that represent each amino acid in the protein are modeled with a harmonic potential:

$$\mathrm{V}(r) = \frac{1}{2}\mathrm{k}(r - r_0)^2 \tag{5}$$

Where V is the potential energy at some displacement r from the resting bond length r_0 and k is the spring constant. Long-range electrostatic interactions are modeled using Coulomb's law with Debye-Hückel electrostatic screening to scale the interaction to simulate a salt concentration of 100 mM in aqueous media[136,137]:

$$E_{ij}(r) = \left(\frac{q_i q_j}{4\pi D r}\right)^{-\frac{r}{\kappa}} \tag{6}$$

Where E is the energy between two charges at some distance r, q_i and q_j are the particle charges, D is the dielectric constant, and κ is the Deby screening length. Finally, all short-range interactions, including Pauli repulsion and London dispersion forces, are captured in the Lennard-Jones potential:

$$\Phi_{LJ} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad (7)$$

Where r is the interaction distance, ϵ is the dispersion energy, and σ is the bead size. The Lennard-Jones potential is scaled using a hydrophobicity score λ[136,138] as $\Phi_{LJ} + (1-\lambda)\epsilon$ if $r \leq 2^{(1/6)}\sigma$ (the repulsive part of the Lennard-Jones potential) or as $\lambda\Phi_{LJ}$ if $r \geq 2^{(1/6)}\sigma$ (the attractive part of the Lennard-Jones potential). These potentials, combined with the efficiency afforded by coarse-graining, enable the study of large protein LLPS systems by MD.

Cryo-electron microscopy

The final method in the structural biology toolbox utilized in this dissertation is cryo-electron microscopy (cryo-EM). Although the resolution revolution is in full swing[139], we leverage cryo-EM for its ability to provide insight into very large structures rather than very small ones– namely, the architecture of oligomeric structures adopted by molecular chaperones. Solid-state NMR has been used to characterize chaperone oligomers in the past[140], but cryo-EM offers a more straightforward route towards the determination of the ensemble of oligomers likely formed by many sHsps in the absence of client proteins[141].

Cryo-EM uses a focused, homogeneous electron beam to image biomolecules that have been frozen in vitreous ice. Electrons that interact with biomolecules in the ice are diffracted and refocused to form a two-dimensional image. A large number of images of individual proteins are collected and images that share the same orientation in the ice sample are aggregated into 2-dimensional "views" of the protein. An ideal cryo-EM experiment will collect views of the target protein from every possible angle around the protein of interest, providing a 360° perspective of the protein. This collection of two-dimensional views can be reconstructed into a three-dimensional structure of the protein.

With MAS NMR, interactions between atoms are discovered via coupling between spins. In monomeric proteins, each spin has a unique chemical shift that can be determined during the assignment process. In a protein self-assembly, however, multiple copies of the same protein exist in an oligomeric structure. This adds a significant challenge to the determination of structure within monomers in the oligomer, and of the oligomer itself, because we need to determine whether a through-space coupling is being made between a spin within the same protein, or the same spin in another copy of the protein that is nearby due to the oligomer architecture. In cryo-EM, this problem doesn't exist–the architecture of the oligomer is its most obvious feature. This positions cryo-EM as a complement to MAS NMR as it provides information that would be more difficult to access with an NMR-based approach. Our use of cryo-EM as a complement to MAS NMR is discussed in more detail in Chapter 3.

References

1 Scott, W. G., Finch, J. T. & Klug, A. The crystal structure of an AII-RNAhammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991-1002 (1995).

2 Doudna, J. A. & Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **418**, 222-228 (2002).

3 Blundell, T., Cutfield, J., Dodson, G., Dodson, E., Hodgkin, D. & Mercola, D. The structure and biology of insulin. *Biochemical Journal* **125**, 50P (1971).
4 Bielajew, B. J., Hu, J. C. & Athanasiou, K. A. Collagen: quantification, biomechanics and role of minor subtypes in cartilage. *Nature Reviews Materials* **5**, 730-747 (2020).
5 Dominguez, R. & Holmes, K. C. Actin structure and function. *Annual review of biophysics* **40**, 169-186 (2011).
6 Weissmann, C. The state of the prion. *Nature Reviews Microbiology* **2**, 861-871 (2004).
7 Saad, S. & Jarosz, D. F. Protein self-assembly: a new frontier in cell signaling. *Current opinion in cell biology* **69**, 62-69 (2021).
8 Petroski, M. D. & Deshaies, R. J. Function and regulation of cullin–RING ubiquitin ligases. *Nature reviews Molecular cell biology* **6**, 9-20 (2005).
9 Dhanasekaran, D., Kashef, K., Lee, C., Xu, H. & Reddy, E. Scaffold proteins of MAP-kinase modules. *Oncogene* **26**, 3185-3202 (2007).
10 Alberti, S. Phase separation in biology. *Current Biology* **27**, R1097-R1102 (2017).
11 Alberti, S., Gladfelter, A. & Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419-434 (2019).
12 Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annual review of cell and developmental biology* **30**, 39-58 (2014).
13 Fowler, D. M., Koulov, A. V., Balch, W. E. & Kelly, J. W. Functional amyloid–from bacteria to humans. *Trends in biochemical sciences* **32**, 217-224 (2007).
14 Chiti, F. & Dobson, C. M. Protein misfolding, functional amyloid, and human disease. *Annu. Rev. Biochem.* **75**, 333-366 (2006).
15 Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S. & Franzmann, T. M. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066-1077 (2015).
16 Nissen, P., Hansen, J., Ban, N., Moore, P. B. & Steitz, T. A. The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**, 920-930 (2000).
17 Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223-230 (1973).
18 Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. Stereochemistry of polypeptide chain configurations. *J Mol Biol* **7**, 95-99 (1963).
19 Purvis, I. J., Bettany, A. J., Santiago, T. C., Coggins, J. R., Duncan, K., Eason, R. & Brown, A. J. The efficiency of folding of some proteins is increased by controlled rates of translation in vivo: A hypothesis. *Journal of molecular biology* **193**, 413-417 (1987).
20 Zhang, G., Hubalewska, M. & Ignatova, Z. Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature structural & molecular biology* **16**, 274-280 (2009).
21 Dill, K. A. Polymer principles and protein folding. *Protein Science* **8**, 1166-1180 (1999).
22 Levinthal, C. Are there pathways for protein folding? *Journal de chimie physique* **65**, 44-45 (1968).
23 Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Bioinformatics* **21**, 167-195 (1995).
24 Gilbert, R. J., Fucini, P., Connell, S., Fuller, S. D., Nierhaus, K. H., Robinson, C. V., Dobson, C. M. & Stuart, D. I. Three-dimensional structures of translating ribosomes by Cryo-EM. *Molecular cell* **14**, 57-66 (2004).

25 Jacobs, S. A. & Khorasanizadeh, S. Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **295**, 2080-2083 (2002).
26 Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. Protein folding intermediates: native-state hydrogen exchange. *Science* **269**, 192-197 (1995).
27 Englander, S. W. & Mayne, L. The case for defined protein folding pathways. *Proceedings of the National Academy of Sciences* **114**, 8253-8258 (2017).
28 Kim, P. S. & Baldwin, R. L. Intermediates in the folding reactions of small proteins. *Annual review of biochemistry* **59**, 631-660 (1990).
29 Kubelka, J., Hofrichter, J. & Eaton, W. A. The protein folding 'speed limit'. *Current opinion in structural biology* **14**, 76-88 (2004).
30 Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **27**, 2985-2993 (1894).
31 Anson, M. & Mirsky, A. On some general properties of proteins. *The Journal of general physiology* **9**, 169 (1925).
32 Spiegel-Adolf, M. Hitzeveränderungen des albumins. *Biochem. Ztschr.* **170**, 126 (1926).
33 Wu, H. & Lin, K.-H. Denaturation of Hemoglobin. *Proceedings of the Society for Experimental Biology and Medicine* **24**, 705-707 (1927).
34 Mirsky, A. & Anson, M. Protein coagulation and its reversal: the reversal of the coagulation of hemoglobin. *The Journal of General Physiology* **13**, 133 (1929).
35 Wu, H. Studies on denaturation of proteins XIII. A theory of denaturation. *Chinese Journal of Physiology* **5**, 321–344 (1931).
36 Edsall, J. T. in *Advances in protein chemistry* Vol. 46 1-5 (Elsevier, 1995).
37 Mirsky, A. E. & Pauling, L. On the structure of native, denatured, and coagulated proteins. *Proceedings of the National Academy of sciences* **22**, 439-447 (1936).
38 Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences* **37**, 205-211 (1951).
39 Pauling, L. & Corey, R. B. Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proceedings of the National Academy of Sciences* **37**, 729-740 (1951).
40 Kauzmann, W. in *Advances in protein chemistry* Vol. 14 1-63 (Elsevier, 1959).
41 Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R., Wyckoff, H. & Phillips, D. C. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181**, 662-666 (1958).
42 Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å. resolution. *Nature* **185**, 422-427 (1960).
43 Matouschek, A., Kellis Jr, J. T., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122-126 (1989).
44 Hughson, F. M., Wright, P. E. & Baldwin, R. L. Structural characterization of a partly folded apomyoglobin intermediate. *Science* **249**, 1544-1548 (1990).
45 Jennings, P. A. & Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **262**, 892-896 (1993).
46 Arai, M. & Kuwajima, K. Rapid formation of a molten globule intermediate in refolding of α-lactalbumin. *Folding and Design* **1**, 275-287 (1996).

47 Raschke, T. M. & Marqusee, S. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature structural biology* **4**, 298-304 (1997).
48 Fujiwara, K., Arai, M., Shimizu, A., Ikeguchi, M., Kuwajima, K. & Sugai, S. Folding− Unfolding Equilibrium and Kinetics of Equine β-Lactoglobulin: Equivalence between the Equilibrium Molten Globule State and a Burst-Phase Folding Intermediate. *Biochemistry* **38**, 4455-4463 (1999).
49 Sigler, P. B. Acid blobs and negative noodles. *Nature* **333**, 210-212 (1988).
50 Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nature reviews Molecular cell biology* **6**, 197-208 (2005).
51 Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M. & Hipps, K. W. Intrinsically disordered protein. *Journal of molecular graphics and modelling* **19**, 26-59 (2001).
52 Van Der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J. & Jones, D. T. Classification of intrinsically disordered regions and proteins. *Chemical reviews* **114**, 6589-6631 (2014).
53 Lee, H., Mok, K. H., Muhandiram, R., Park, K.-H., Suk, J.-E., Kim, D.-H., Chang, J., Sung, Y. C., Choi, K. Y. & Han, K.-H. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. *Journal of Biological Chemistry* **275**, 29426-29432 (2000).
54 Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N. & Dunker, A. K. Coupled folding and binding with α-helix-forming molecular recognition elements. *Biochemistry* **44**, 12454-12470 (2005).
55 Receveur-Bréchot, V., Bourhis, J. M., Uversky, V. N., Canard, B. & Longhi, S. Assessing protein disorder and induced folding. *Proteins: Structure, Function, and Bioinformatics* **62**, 24-45 (2006).
56 Wright, P. E. & Dyson, H. J. Linking folding and binding. *Current opinion in structural biology* **19**, 31-38 (2009).
57 Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proceedings of the National Academy of Sciences* **93**, 11504-11509 (1996).
58 Daughdrill, G. W., Chadsey, M. S., Karlinsey, J. E., Hughes, K. T. & Dahlquist, F. W. The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, σ28. *Nature structural biology* **4**, 285-291 (1997).
59 Plaxco, K. W. & Groβ, M. The importance of being unfolded. *Nature* **386**, 657-659 (1997).
60 Tompa, P. & Fuxreiter, M. Fuzzy complexes: polymorphism and structural disorder in protein–protein interactions. *Trends in biochemical sciences* **33**, 2-8 (2008).
61 Baker, J. M., Hudson, R. P., Kanelis, V., Choy, W.-Y., Thibodeau, P. H., Thomas, P. J. & Forman-Kay, J. D. CFTR regulatory region interacts with NBD1 predominantly via multiple transient helices. *Nature structural & molecular biology* **14**, 738-745 (2007).
62 Mittag, T., Orlicky, S., Choy, W.-Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M. & Forman-Kay, J. D. Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor. *Proceedings of the National Academy of Sciences* **105**, 17772-17777 (2008).

63 Hendus-Altenburger, R., Pedraz-Cuesta, E., Olesen, C. W., Papaleo, E., Schnell, J. A., Hopper, J. T., Robinson, C. V., Pedersen, S. F. & Kragelund, B. B. The human Na+/H+ exchanger 1 is a membrane scaffold protein for extracellular signal-regulated kinase 2. *BMC biology* **14**, 1-17 (2016).
64 Milles, S., Mercadante, D., Aramburu, I. V., Jensen, M. R., Banterle, N., Koehler, C., Tyagi, S., Clarke, J., Shammas, S. L. & Blackledge, M. Plasticity of an ultrafast interaction between nucleoporins and nuclear transport receptors. *Cell* **163**, 734-745 (2015).
65 Borgia, A., Borgia, M. B., Bugge, K., Kissling, V. M., Heidarsson, P. O., Fernandes, C. B., Sottini, A., Soranno, A., Buholzer, K. J. & Nettels, D. Extreme disorder in an ultrahigh-affinity protein complex. *Nature* **555**, 61-66 (2018).
66 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* **324**, 1729-1732 (2009).
67 Wei, M.-T., Elbaum-Garfinkle, S., Holehouse, A. S., Chen, C. C.-H., Feric, M., Arnold, C. B., Priestley, R. D., Pappu, R. V. & Brangwynne, C. P. Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles. *Nature chemistry* **9**, 1118-1125 (2017).
68 Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A. & Drechsel, D. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688-699. e616 (2018).
69 Choi, J.-M., Holehouse, A. S. & Pappu, R. V. Physical principles underlying the complex biology of intracellular phase transitions. *Annual review of biophysics* **49**, 107-133 (2020).
70 Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nature reviews Molecular cell biology* **15**, 384-396 (2014).
71 Yoshimura, Y., Lin, Y., Yagi, H., Lee, Y.-H., Kitayama, H., Sakurai, K., So, M., Ogi, H., Naiki, H. & Goto, Y. Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. *Proceedings of the National Academy of Sciences* **109**, 14446-14451 (2012).
72 Ross, C. A. & Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nature medicine* **10**, S10-S17 (2004).
73 Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. & Zweckstetter, M. Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature communications* **8**, 275 (2017).
74 Ray, S., Singh, N., Kumar, R., Patel, K., Pandey, S., Datta, D., Mahato, J., Panigrahi, R., Navalkar, A. & Mehra, S. α-Synuclein aggregation nucleates through liquid–liquid phase separation. *Nature chemistry* **12**, 705-716 (2020).
75 Conicella, A. E., Zerze, G. H., Mittal, J. & Fawzi, N. L. ALS mutations disrupt phase separation mediated by α-helical structure in the TDP-43 low-complexity C-terminal domain. *Structure* **24**, 1537-1549 (2016).
76 Altmeyer, M., Neelsen, K. J., Teloni, F., Pozdnyakova, I., Pellegrino, S., Grøfte, M., Rask, M.-B. D., Streicher, W., Jungmichel, S. & Nielsen, M. L. Liquid demixing of

intrinsically disordered proteins is seeded by poly (ADP-ribose). *Nature communications* **6**, 8088 (2015).
77 Murakami, T., Qamar, S., Lin, J. Q., Schierle, G. S. K., Rees, E., Miyashita, A., Costa, A. R., Dodd, R. B., Chan, F. T. & Michel, C. H. ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron* **88**, 678-690 (2015).
78 Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Molecular cell* **60**, 231-241 (2015).
79 Mukherjee, S., Sakunthala, A., Gadhe, L., Poudyal, M., Sawner, A. S., Kadu, P. & Maji, S. K. Liquid-liquid phase separation of α-synuclein: a new mechanistic insight for α-synuclein aggregation associated with Parkinson's disease pathogenesis. *Journal of Molecular Biology* **435**, 167713 (2023).
80 Huang, S., Mo, X., Wang, J., Ye, X., Yu, H. & Liu, Y. α-Synuclein phase separation and amyloid aggregation are modulated by C-terminal truncations. *FEBS letters* **596**, 1388-1400 (2022).
81 Wegmann, S., Eftekharzadeh, B., Tepper, K., Zoltowska, K. M., Bennett, R. E., Dujardin, S., Laskowski, P. R., MacKenzie, D., Kamath, T. & Commins, C. Tau protein liquid–liquid phase separation can initiate tau aggregation. *The EMBO journal* **37**, e98049 (2018).
82 Kanaan, N. M., Hamel, C., Grabinski, T. & Combs, B. Liquid-liquid phase separation induces pathogenic tau conformations in vitro. *Nature communications* **11**, 2809 (2020).
83 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* **163**, 123-133 (2015).
84 Fitzpatrick, A. W., Falcon, B., He, S., Murzin, A. G., Murshudov, G., Garringer, H. J., Crowther, R. A., Ghetti, B., Goedert, M. & Scheres, S. H. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547**, 185-190 (2017).
85 Zhang, W., Tarutani, A., Newell, K. L., Murzin, A. G., Matsubara, T., Falcon, B., Vidal, R., Garringer, H. J., Shi, Y. & Ikeuchi, T. Novel tau filament fold in corticobasal degeneration. *Nature* **580**, 283-287 (2020).
86 Radamaker, L., Baur, J., Huhn, S., Haupt, C., Hegenbart, U., Schönland, S., Bansal, A., Schmidt, M. & Fändrich, M. Cryo-EM reveals structural breaks in a patient-derived amyloid fibril from systemic AL amyloidosis. *Nature Communications* **12**, 875 (2021).
87 Kollmer, M., Close, W., Funk, L., Rasmussen, J., Bsoul, A., Schierhorn, A., Schmidt, M., Sigurdson, C. J., Jucker, M. & Fändrich, M. Cryo-EM structure and polymorphism of Aβ amyloid fibrils purified from Alzheimer's brain tissue. *Nature communications* **10**, 4760 (2019).
88 Sharma, S. K., De Los Rios, P., Christen, P., Lustig, A. & Goloubinoff, P. The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nature chemical biology* **6**, 914-920 (2010).
89 Nunes, J. M., Mayer-Hartl, M., Hartl, F. U. & Müller, D. J. Action of the Hsp70 chaperone system observed with single proteins. *Nature communications* **6**, 6307 (2015).
90 Sekhar, A., Rosenzweig, R., Bouvignies, G. & Kay, L. E. Mapping the conformation of a client protein through the Hsp70 functional cycle. *Proceedings of the National Academy of Sciences* **112**, 10395-10400 (2015).

91 Sekhar, A., Rosenzweig, R., Bouvignies, G. & Kay, L. E. Hsp70 biases the folding pathways of client proteins. *Proceedings of the National Academy of Sciences* **113**, E2794-E2801 (2016).
92 Fan, C.-Y., Lee, S. & Cyr, D. M. Mechanisms for regulation of Hsp70 function by Hsp40. *Cell stress & chaperones* **8**, 309 (2003).
93 Liu, Q., Liang, C. & Zhou, L. Structural and functional analysis of the Hsp70/Hsp40 chaperone system. *Protein Science* **29**, 378-390 (2020).
94 Faust, O., Abayev-Avraham, M., Wentink, A. S., Maurer, M., Nillegoda, N. B., London, N., Bukau, B. & Rosenzweig, R. HSP40 proteins use class-specific regulation to drive HSP70 functional diversity. *Nature* **587**, 489-494 (2020).
95 Żwirowski, S., Kłosowska, A., Obuchowski, I., Nillegoda, N. B., Piróg, A., Ziętkiewicz, S., Bukau, B., Mogk, A. & Liberek, K. Hsp70 displaces small heat shock proteins from aggregates to initiate protein refolding. *The EMBO journal* **36**, 783-796 (2017).
96 Gonçalves, C. C., Sharon, I., Schmeing, T. M., Ramos, C. H. & Young, J. C. The chaperone HSPB1 prepares protein aggregates for resolubilization by HSP70. *Scientific Reports* **11**, 17139 (2021).
97 Liu, Z., Zhang, S., Gu, J., Tong, Y., Li, Y., Gui, X., Long, H., Wang, C., Zhao, C. & Lu, J. Hsp27 chaperones FUS phase separation under the modulation of stress-induced phosphorylation. *Nature structural & molecular biology* **27**, 363-372 (2020).
98 Freilich, R., Betegon, M., Tse, E., Mok, S.-A., Julien, O., Agard, D. A., Southworth, D. R., Takeuchi, K. & Gestwicki, J. E. Competing protein-protein interactions regulate binding of Hsp27 to its client protein tau. *Nature communications* **9**, 4563 (2018).
99 Mainz, A., Peschek, J., Stavropoulou, M., Back, K. C., Bardiaux, B., Asami, S., Prade, E., Peters, C., Weinkauf, S. & Buchner, J. The chaperone αB-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nature structural & molecular biology* **22**, 898-905 (2015).
100 Baughman, H. E., Pham, T.-H. T., Adams, C. S., Nath, A. & Klevit, R. E. Release of a disordered domain enhances HspB1 chaperone activity toward tau. *Proceedings of the National Academy of Sciences* **117**, 2923-2929 (2020).
101 Baughman, H. E., Clouser, A. F., Klevit, R. E. & Nath, A. HspB1 and Hsc70 chaperones engage distinct tau species and have different inhibitory effects on amyloid formation. *Journal of Biological Chemistry* **293**, 2687-2700 (2018).
102 Yoo, H., Bard, J. A., Pilipenko, E. V. & Drummond, D. A. Chaperones directly and efficiently disperse stress-triggered biomolecular condensates. *Molecular Cell* **82**, 741-755. e711 (2022).
103 Lu, S., Hu, J., Arogundade, O. A., Goginashvili, A., Vazquez-Sanchez, S., Diedrich, J. K., Gu, J., Blum, J., Oung, S. & Ye, Q. Heat-shock chaperone HSPB1 regulates cytoplasmic TDP-43 phase separation and liquid-to-gel transition. *Nature cell biology* **24**, 1378-1393 (2022).
104 Darling, A. L., Dahrendorff, J., Creodore, S. G., Dickey, C. A., Blair, L. J. & Uversky, V. N. Small heat shock protein 22 kDa can modulate the aggregation and liquid–liquid phase separation behavior of tau. *Protein Science* **30**, 1350-1359 (2021).
105 Gallagher, E. R. & Holzbaur, E. L. The selective autophagy adaptor p62/SQSTM1 forms phase condensates regulated by HSP27 that facilitate the clearance of damaged lysosomes via lysophagy. *Cell Reports* **42** (2023).

106 Mateju, D., Franzmann, T. M., Patel, A., Kopach, A., Boczek, E. E., Maharana, S., Lee, H. O., Carra, S., Hyman, A. A. & Alberti, S. An aberrant phase transition of stress granules triggered by misfolded protein and prevented by chaperone function. *The EMBO journal* **36**, 1669-1687 (2017).
107 Sharp, P. S., Akbar, M. T., Bouri, S., Senda, A., Joshi, K., Chen, H.-J., Latchman, D. S., Wells, D. J. & de Belleroche, J. Protective effects of heat shock protein 27 in a model of ALS occur in the early stages of disease progression. *Neurobiology of disease* **30**, 42-55 (2008).
108 Jesse, C. M., Bushuven, E., Tripathi, P., Chandrasekar, A., Simon, C. M., Drepper, C., Yamoah, A., Dreser, A., Katona, I. & Johann, S. ALS-associated endoplasmic reticulum proteins in denervated skeletal muscle: implications for motor neuron disease pathology. *Brain pathology* **27**, 781-794 (2017).
109 Dierick, I., Irobi, J., Janssens, S., Theuns, J., Lemmens, R., Jacobs, A., Corsmit, E., Hersmus, N., Van Den Bosch, L. & Robberecht, W. Genetic variant in the HSPB1 promoter region impairs the HSP27 stress response. *Human mutation* **28**, 830-830 (2007).
110 Matlahov, I. & van der Wel, P. C. Hidden motions and motion-induced invisibility: dynamics-based spectral editing in solid-state NMR. *Methods* **148**, 123-135 (2018).
111 Mandala, V. S. & Hong, M. High-sensitivity protein solid-state NMR spectroscopy. *Current opinion in structural biology* **58**, 183-190 (2019).
112 Reif, B., Ashbrook, S. E., Emsley, L. & Hong, M. Solid-state NMR spectroscopy. *Nature Reviews Methods Primers* **1**, 2 (2021).
113 Levitt, M. H. *Spin dynamics: basics of nuclear magnetic resonance*. (John Wiley & Sons, 2013).
114 Bloch, F. The principle of nuclear induction. *Science* **118**, 425-430 (1953).
115 Pauli, W. Zur Frage der theoretischen Deutung der Satelliten einiger Spektrallinien und ihrer Beeinflussung durch magnetische Felder. *Naturwissenschaften* **12**, 741-743 (1924).
116 Zeeman, P. Over de invloed eener magnetisatie op den aard van het door een stof uitgezonden licht. *Verslagen van de Gewone Vergaderingen der Wisen Natuurkundige Afdeeling (Koninklijk Akademie van Wetenschappen te Amsterdam)* **5**, 181-184 (1896).
117 Rosman, R., Schultz, L., Shima, M. & Taylor, P. Atomic weights of the elements. *Pure and Applied Chemistry*, 1593-1607 (1999).
118 Lacabanne, D., Meier, B. H. & Böckmann, A. Selective labeling and unlabeling strategies in protein solid-state NMR spectroscopy. *Journal of biomolecular NMR* **71**, 141-150 (2018).
119 Ghosh, R., Madrid, C. L. & Frederick, K. K. Segmental Labeling: Applications to Protein NMR and DNP. *eMagRes* **9** (2020).
120 Clark, E. T., Sievers, E. E. & Debelouchina, G. T. A Chemical Biology Primer for NMR Spectroscopists. *Journal of magnetic resonance open* **10**, 100044 (2022).
121 Bloch, F., Hansen, W. W. & Packard, M. Nuclear Induction. *Physical Review* **69**, 127-127 (1946).
122 Purcell, E. M., Torrey, H. C. & Pound, R. V. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review* **69**, 37-38 (1946).
123 Duer, M. J. *Solid state NMR spectroscopy: principles and applications*. (John Wiley & Sons, 2008).
124 Bax, A. *Two Dimensional Nuclear Magnetic Resonance in Liquids*, Colorado State University, (1982).

125 Andrew, E., Bradbury, A. & Eades, R. Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature* **182**, 1659-1659 (1958).
126 Schaefer, J. & Stejskal, E. Carbon-13 nuclear magnetic resonance of polymers spinning at the magic angle. *Journal of the American Chemical Society* **98**, 1031-1032 (1976).
127 Morris, G. A. & Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society* **101**, 760-762 (1979).
128 Pines, A., Gibby, M. & Waugh, J. Proton-enhanced nuclear induction spectroscopy. A method for high resolution NMR of dilute spins in solids. *The Journal of Chemical Physics* **56**, 1776-1777 (1972).
129 Pines, A., Gibby, M. G. & Waugh, J. Proton-enhanced NMR of dilute spins in solids. *The Journal of chemical physics* **59**, 569-590 (1973).
130 Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E. & Berendsen, H. J. GROMACS: fast, flexible, and free. *Journal of computational chemistry* **26**, 1701-1718 (2005).
131 Bock, L. V., Blau, C., Schröder, G. F., Davydov, I. I., Fischer, N., Stark, H., Rodnina, M. V., Vaiana, A. C. & Grubmüller, H. Energy barriers and driving forces in tRNA translocation through the ribosome. *Nature structural & molecular biology* **20**, 1390-1396 (2013).
132 Kinghorn, D. Molecular Dynamics Benchmarks GPU Roundup GROMACS NAMD2 NAMD 3alpha on 12 GPUs. *Puget Systems* (2022).
133 Yang, Y. I., Shao, Q., Zhang, J., Yang, L. & Gao, Y. Q. Enhanced sampling in molecular dynamics. *The Journal of chemical physics* **151**, 070902 (2019).
134 Torrie, G. M. & Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187-199 (1977).
135 Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical physics letters* **314**, 141-151 (1999).
136 Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B. & Mittal, J. Sequence determinants of protein phase behavior from a coarse-grained model. *PLoS computational biology* **14**, e1005941 (2018).
137 Debye, P. & Hückel, E. De la theorie des electrolytes. I. abaissement du point de congelation et phenomenes associes. *Physikalische Zeitschrift* **24**, 185-206 (1923).
138 Kapcha, L. H. & Rossky, P. J. A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *Journal of molecular biology* **426**, 484-498 (2014).
139 Kühlbrandt, W. The resolution revolution. *Science* **343**, 1443-1444 (2014).
140 Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S., Kühne, R., Stout, J. R., Higman, V. A., Klevit, R. E., van Rossum, B.-J. & Oschkinat, H. Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers. *Nature structural & molecular biology* **17**, 1037-1042 (2010).
141 Kaiser, C. J., Peters, C., Schmid, P. W., Stavropoulou, M., Zou, J., Dahiya, V., Mymrikov, E. V., Rockel, B., Asami, S. & Haslbeck, M. The structure and oxidation of the eye lens chaperone αA-crystallin. *Nature structural & molecular biology* **26**, 1141-1150 (2019).

Biophysical Journal
Article

Biophysical Society

Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC

Raymond F. Berkeley,[1] Maryam Kashefi,[1] and Galia T. Debelouchina[1,*]
[1]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California

ABSTRACT A subset of the proteins found in pathological protein fibrils also exhibit tendencies for liquid-liquid phase separation (LLPS) both in vitro and in cells. The mechanisms underlying the connection between these phase transitions have been challenging to study due to the heterogeneous and dynamic nature of the states formed during the maturation of LLPS protein droplets into gels and solid aggregates. Here, we interrogate the liquid-to-solid transition of the low-complexity domain of the RNA-binding protein FUS (FUS LC), which has been shown to adopt LLPS, gel-like, and amyloid states. We employ magic-angle-spinning NMR spectroscopy, which has allowed us to follow these transitions in real time and with residue-specific resolution. We observe the development of β-sheet structure through the maturation process and show that the final state of FUS LC fibrils produced after LLPS is distinct from that grown from fibrillar seeds. We also apply our methodology to FUS LC G156E, a clinically relevant FUS mutant that exhibits accelerated fibrillization rates. We observe significant changes in dynamics during the transformation of the FUS LC G156E construct and begin to unravel the sequence specific contributions to this phenomenon with computational studies of the phase-separated state of FUS LC and FUS LC G156E.

SIGNIFICANCE The presence of protein aggregates in the brain is a common pathological sign of neurodegenerative disease. Recent work has revealed that a subset of the proteins found in these aggregates can also form liquid-liquid droplets and gels. Although the interconversion from one state to another can have vast implications for cell function and disease, the molecular mechanisms that underlie these processes are not well understood. Here, we combine magic-angle-spinning NMR spectroscopy with other biophysical and computational tools to follow the transitions of the RNA-binding protein FUS. This approach has allowed us to observe real-time changes in structure and dynamics as the protein undergoes these transitions, and to reveal the intricate effects of disease-relevant mutations on the transformation process.

INTRODUCTION

Neuronal protein aggregates are hallmarks of neurodegenerative disease whose biological genesis is still poorly understood (1,2). Although their cellular functions may vary, proteins found in these aggregates can contain intrinsically disordered domains with low-complexity (LC) charge-patterned sequences. These sequences not only promote the adoption of fibrillar or aggregated states but are also important drivers for liquid-liquid phase separation (LLPS) and the formation of membraneless organelles (3,4). As such, many proteins that are found in pathological neuronal aggregates can exist in different biophysical states that span a broad range of dynamic regimes, including liquid-liquid droplets, hydrogels, and amyloid fibrils. Some examples include proteins associated with Alzheimer disease (tau) (5–10), Parkinson's disease (α-synuclein) (11–14), and frontotemporal lobar degeneration (TDP-43 and FUS) (15,16).

Recent work has started to establish connections between these biophysical states in vitro. For example, the repeated coacervation and dissolution of liquid droplets of RNA-binding proteins such as hnRNPA1 has been shown to accelerate the formation of protein aggregates (17). Imaging experiments often show fibrils that grow from the center of TDP-43, FUS, and hnRNPA1 droplets, suggesting that phase-separated protein environments may serve as centers for the nucleation and growth of amyloid in the presence or absence of RNA (18–20). Reversible amyloid-like fibers of

Submitted October 19, 2020, and accepted for publication February 8, 2021.

*Correspondence: gdebelouchina@ucsd.edu

Editor: Jason Kahn.
https://doi.org/10.1016/j.bpj.2021.02.008

FUS segments have also been found embedded within hydrogels (21). Although it is known that phase separation is not a requirement for the growth of fibrils in vitro (6), the growing body of evidence linking mature LLPS protein droplets and hydrogels to aggregates and fibrils suggests that the aberrant transition between these states could be a physiologically relevant factor in disease pathogenesis (7,16,19).

The current biophysical toolbox contains well-established methodologies that can characterize each of those states separately. For example, liquid droplets have been analyzed with a variety of imaging, spectroscopic, and computational approaches that have provided valuable information regarding the sequence requirements and the nature of protein-protein interactions that drive LLPS (20,22–34). On the other hand, magic-angle-spinning (MAS) NMR spectroscopy and cryoelectron microscopy (cryo-EM) have revealed the common structural principles that underlie the formation of stable, β-sheet-rich amyloid fibrils (5,16,35–37). Although more challenging because of their viscous, dynamic, and heterogeneous nature, hydrogels have also been amenable to characterization by MAS NMR approaches (38,39). Yet, to establish a comprehensive view of how these states are connected on the molecular level and how these connections may break in disease requires a strategy that can ideally observe these transformations in real time and in the same sample at atomic resolution. Here, we explore the capability of MAS NMR spectroscopy to achieve this goal.

MAS NMR spectroscopy is a versatile structural technique that allows the investigation of biological samples of different sizes, complexity, and material state (40,41). Although the NMR resonances arising from these proteins may show line broadening because of slow molecular tumbling, strong dipolar interactions, and chemical shift anisotropy, the line broadening can be removed by spinning the sample at 54.7° (the magic angle) (42,43). During MAS, tailored pulse sequences can reintroduce the magnetic spin interactions in a controlled manner that allows the extraction of structural information such as the chemical shift, the protonation state, the distance between atoms, or their relative orientation. Spin-spin interactions can be reintroduced based on the scalar couplings between covalently bonded atoms, similar to solution NMR, enabling the description of the mobile components in the sample (44). On the other hand, dipolar-based pulse sequences can be used to identify protein regions or sample components that experience slow motions and detectable dipolar interactions through space (45). The combination of the two approaches provides the opportunity to dynamically edit the NMR spectra and to describe both the mobile and rigid components of the sample (46).

In previous work, we have used MAS NMR spectroscopy to follow the transition from the liquid droplet to the gel state of the chromatin related protein HP1α (47). This approach allowed us to detect specific serine residues that experience large changes in mobility and that appear to be important for the formation of cross-linking interactions in the gel state. Here, we extend this approach to the LC domain of the RNA-binding protein Fused in sarcoma (FUS), and we follow the transformation of individual samples of FUS LC from the liquid droplet to the gel and amyloid states in real time. We also compare the wild-type sequence and a sequence that harbors a pathogenic G156E mutation known to increase the likelihood of amyotrophic lateral sclerosis-spectrum diseases in patients (19). This mutation is particularly interesting from the biophysical point of view as it introduces additional negative charge in the sequence of FUS LC. However, unlike other FUS modifications such as phosphorylation or phosphomimetic substitutions that introduce negative charge and impede LLPS, the G156E mutation not only supports LLPS but also greatly accelerates the onset of FUS fibrils in vitro (16,19,48). To gain further insight into the molecular origins of this puzzling behavior, we complement our experimental studies with coarse-grained simulations of wild-type and G156E FUS LC and uncover subtle differences in the protein-protein interaction landscape of their liquid-droplet states. These differences may be amplified through the subsequent steps of the transformation process leading to the increased aggregation propensity of the G156E mutation.

MATERIALS AND METHODS

Expression of FUS LC and related constructs

FUS LC was prepared from recombinant Rosetta (DE3) competent *Escherichia coli* cells (MilliporeSigma, Burlington, MA) that had been transformed with a plasmid encoding the 6xHis-MBP-TEV-FUS (1–163) sequence. This plasmid was a gift from Nicolas Fawzi (RRID: Addgene_98653; Addgene plasmid no. 98653; http://n2t.net/addgene:98653, Addgene, Watertown, MA) (48). Seed cultures were grown to saturation from freshly transformed colonies and inoculated at a 1% v/v ratio into either Luria-Bertani or $^{15}N/^{13}C$-M9 medium supplemented with kanamycin (50 μg/mL). The cultures were grown at 37°C to an OD600 of ~0.7 and protein expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside. The cultures were allowed to express protein for 4 h at 37°C before being harvested by centrifugation at 10,000 × *g* and 4°C for 30 min. After decanting the supernatant, cell pellets were stored at −80°C for later use. In addition to the wild-type FUS LC fusion protein, the pathogenic mutant FUS LC G156E protein was prepared. To generate the mutant, the requisite G to E mutation was introduced to the wild-type FUS LC plasmid using an NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, Ipswich, MA). Expression and purification conditions of the G156E mutant protein were identical to those of the wild-type.

Purification of FUS LC and related constructs

Here, we followed the published protocol by Burke et al. with some modifications (31). The frozen cell pellet was thawed, resuspended in lysis buffer (20 mM sodium phosphate, 300 mM sodium chloride, Roche cOmplete EDTA-free Protease Inhibitor Cocktail (MilliporeSigma, Burlington, MA) (pH 7.4, 4°C), and lysed by pulsed sonication for 30 min at 4°C using a Qsonica sonicator (Qsonica, Newton, CT) with a 1/8″ diameter probe tip at 12 kHz

(60%) output. The lysate was cleared by centrifugation at 20,000 × *g* for 30 min and the supernatant was incubated with Thermo Scientific HisPur Ni-NTA Resin (Thermo Fisher Scientific, Waltham, MA) for 30 min at 4°C. The suspension of beads was washed with 10 column volumes of lysis buffer containing 10 mM imidazole, and protein was eluted with two column volumes of lysis buffer containing 250 mM imidazole. The eluted protein was incubated with 6xHis-tagged Tobacco Etch Virus (TEV) protease at a ratio of 1:150 TEV to MBP-FUS LC for 5 h and 25°C to cleave MBP from FUS LC. FUS LC crashes out upon cleavage to yield a cloudy suspension with white clumps of aggregated protein. After TEV cleavage, 8 M urea was added to the reaction mixture to solubilize the aggregated FUS LC. The TEV reaction was monitored by SDS-PAGE (note that FUS LC does not bind Coomassie, so cleavage was verified by the gel shift of the MBP band). The mixture containing solubilized FUS LC was diluted to 2 mg/ml with 20 mM CAPS and 150 mM sodium chloride (pH 11), and subjected to size-exclusion chromatography over a GE HiLoad 16/600 Superdex 75-pg column (GE Healthcare, Chicago, IL). Purity was verified by high-performance liquid chromatography (HPLC), mass spectrometry, and by comparing the A260 and A280 absorbance to verify that no nucleic acid contaminants were present in the sample. This protocol generally yields 10–15 mg of FUS LC per liter of culture in both Luria-Bertani and M9 media.

Protein labeling with small molecule fluorophores

To produce Cy3-labeled FUS LC, an FUS LC fusion protein with a Cys-Ser-Gly C-terminal tag (effectively FUS LC(1–166) S164C) was generated by introducing the requisite sequence to the wild-type FUS LC plasmid using an NEBuilder HiFi DNA Assembly Cloning Kit. The 6xHis-MBP-FUS LC(1–166) S164C construct was expressed and subjected to Ni-NTA purification as described above. The eluent containing 6xHis-MBP-FUS LC (1–166) S164C was diluted to a final concentration of 20 μM with reaction buffer (20 mM sodium phosphate, 300 mM sodium chloride (pH 7.4)), and 500 μM TCEP and 80 μM Cy3-maleamide (4 M eq.; APExBIO, Houston, TX) were added. The reaction was allowed to proceed for 60 s at 25°C in the dark before quenching with excess β-mercaptoethanol (>200 M eq.). The reaction mixture was then transferred to dialysis tubing with 10 kDa molecular weight cutoff and dialyzed twice into 20 mM sodium phosphate, 300 mM sodium chloride, and 100 μM TCEP (pH 7.4) at 4°C. Once the majority of the residual Cy3 was removed by dialysis, the labeled protein was removed from the dialysis tubing and subjected to TEV cleavage and size-exclusion chromatography as described above. The labeling efficiency was ~30% as determined by analytical HPLC and mass spectrometry.

Microscopy of liquid-liquid phase-separated droplets

LLPS was induced by the dilution of 1.8 mM stock solutions of FUS LC (with 5% FUS LC(1–166) S164C-Cy3) in 20 mM CAPS, 150 mM sodium chloride (pH 11) with 20 mM sodium phosphate, and 150 mM sodium chloride (pH 7.4) at 25°C. To mitigate the risk of aggregation and to ensure quick mixing of the components, the phosphate buffer was added to the protein stock solutions. Phase separation was apparent by the sample rapidly becoming cloudy and opaque. The phase-separated protein was added to a microscope slide and allowed to incubate quiescently over the course of the imaging experiment. To prevent both the evaporation of the buffer and the mechanical perturbation of the droplets, the borders of the microscope coverslips were coated with a small amount of vacuum grease that served to raise the slide and to hermetically seal the sample.

Fluorescence recovery after photobleaching

Droplet samples were prepared as described above and imaged on an Olympus FV1000 Confocal microscope (Olympus Scientific Solutions Americas Corp., Waltham, MA). Six droplets with 5- to 10-μm diameter were subjected to photobleaching. A circular region with a diameter of one-half of the droplet diameter was bleached for each droplet. The diameter of the bleach region spanned from the center of the droplet to the droplet edge. Photobleaching was performed for 2 s using an FV3000 Hybrid Scan Unit (Olympus Scientific Solutions Americas Corp.) in Tornado Scanning mode. Fluorescence intensity was recorded within the bleached region every 2 s. Data were normalized to the pre- and postbleach fluorescence. All microscopy images were processed with Fiji/ImageJ (49,50) and data were analyzed and visualized with SciPy tools (51–54).

Thioflavin T assays

Spectra were recorded at 25°C on a Molecular Devices SpectraMax i3x Fluorometer (Molecular Devices, San Jose, CA). The excitation wavelength was 440 nm, and emission was recorded from 465 to 520 nm at a scan speed of 1 nm/s. LLPS was induced by the dilution of 1.8 mM stock solutions of FUS LC (stored in 20 mM CAPS and 150 mM sodium chloride (pH 11)) with 20 mM sodium phosphate and 150 mM sodium chloride (pH 7.4) to a final concentration of 300 μM FUS LC. A phase-separated stock solution was kept, and an aliquot was drawn at each time point for analysis. Each aliquot contained 30 μM FUS LC and 20 μM thioflavin T (ThT) diluted with 20 mM sodium phosphate and 150 mM sodium chloride (pH 7.4).

Solid-state nuclear magnetic resonance experiments

All experiments were performed using 3.2-mm thin-walled zirconia MAS rotors with 50 μL sample volume. LLPS of 30 mg ^{15}N, ^{13}C-labeled FUS LC or ^{15}N, ^{13}C-labeled FUS LC G165E was induced as described above at a concentration of 300 μM and the droplets were transferred into the rotor by gentle centrifugation at 3000 × *g* using a device built in-house. This condensed the sample into a single proteinaceous phase with final concentration of ~400 mg/ml (23 mM). This concentration was determined based on the difference between the initial A280 absorbance of the protein sample and the final absorbance of the supernatant after rotor packing. This resulted in ~20 mg of protein inside the rotor. Spectra were acquired on a 750-MHz (17.6 T) NMR spectrometer equipped with a 3.2-mm E^{free} triple resonance HCN MAS probe (Bruker Biospin, Billerica, MA). All experiments were performed at MAS frequency of 11.11 kHz. The sample was cooled with a stream of dry nitrogen gas maintained at 285 K while we estimate that the sample temperature during the MAS experiments is 10–15° higher. More details regarding the MAS NMR experimental settings are given in the Supporting materials and methods. Data were visualized and analyzed with NMRFAM-Sparky (55). The Sparky files have been deposited at https://doi.org/10.5281/zenodo.4521765.

Statistical analysis of NMR chemical shifts

Chemical shift and protein coordinate data were acquired from the Protein Data Bank (PDB) and Biological Magnetic Resonance Bank (BMRB) with PACSY Maker (56). For each amino acid in FUS LC, the PACSY database was filtered using secondary structure classifications generated by STRIDE (57) to produce a data set of chemical shifts associated with residues in PDB structures that are in either β-sheet or random coil conformations. Only chemical shifts with unambiguous assignments were considered. Because the BMRB only contains one-dimensional (1D) assignments, chemical shifts were projected into two dimensions by considering atom connectivity for each amino acid and plotting theoretical correlations using the chemical shifts for directly bonded atoms on a per-protein basis. Data were analyzed and results were visualized using

SciPy tools (51–54). Bruker NMR data were parsed with Nmrglue (58). All codes are available upon request. It should be noted that this approach is similar to that taken by PLUQ (59).

Coarse-grained molecular dynamics simulations

Molecular dynamics simulations were performed with HOOMD-blue (60) using a set of hydrophobicity-scaled pair potentials initially described by Dignon, et al. (27). This model has been shown to be effective for simulating LLPS of FUS LC and other proteins (32,48,61,62). Briefly, the model defines three interaction potentials: a potential for bonded interactions, a single potential representing short-range nonbonded interactions, and a potential representing electrostatic interactions. Bonds are represented by a harmonic potential with a bond length of 3.8 Å and a spring constant of 10 kJ/Å^2. Short range nonbonded interactions are represented by a standard Lennard-Jones potential that has been scaled to the hydrophobicity of each interaction pair. This scaled Lennard-Jones potential was utilized as implemented in the azplugins package for HOOMD-blue (63), and the hydrophobicity scaling parameters were identical to those described in Dignon, et al. (27,64). Electrostatic interactions were represented by the Yukawa potential with a Debye screening length of 1 nm and a dielectric constant of 80 to mimic an aqueous solvent containing 100 mM salt. For each simulation, 100 FUS LC or FUS LC G156E monomers were prepared in a linear configuration using mBuild (65). The coordinates of each particle in each monomer were then randomized such that each bond length was fixed at 3.8 Å and no particles overlapped. Monomers were then packed into a 50 × 50 × 50 nm simulation cell using PACKMOL (66), resulting in a protein concentration of 22 mg/mL (1.3 mM) before phase separation. Charges and masses were assigned to each particle within HOOMD-blue. Production simulations were performed on a single NVIDIA Tesla K80 GPU. Each simulation was run for a total of 1.1 μs with 10 fs timesteps using a Langevin integrator at 300 K. Simulations were analyzed and contact maps were visualized using tools from the SciPy software stack (51–54). Simulation snapshots were visualized using Ovito Pro (67). Two 300-ns replicates of each simulation were performed and analyzed to confirm that contact map results were not dependent of the initial configuration of the simulation.

RESULTS AND DISCUSSION

Liquid droplets of FUS LC undergo transition to gels and fibers at neutral pH

The full-length FUS protein can undergo a liquid-to-solid phase transition that is accelerated by the presence of a clinically relevant G156E mutation (19). We chose to work with the N-terminal LC domain comprising residues 1–163 of the protein (FUS LC) to investigate the differential behavior of FUS LC and the pathological FUS LC G156E mutant in vitro. The LC domain of FUS has been extensively studied in the LLPS state (31,32,48) and includes segments that are known to form rigid fibril cores (16,68). We prepared recombinantly purified protein (Fig. S1 and S2) mixed with 5% Cy3-labeled FUS LC and initiated phase separation under physiological pH and low-salt conditions. We observed the immediate formation of liquid-liquid phase-separated droplets that were subsequently placed in hermetically sealed microscope slides and monitored for several weeks (Fig. 1 *A*). The well-defined droplets coarsened into gel-like structures over the course of several days. As the droplets continued to age, fibrillar structures developed that appeared to protrude from the gel cores in a manner consistent with previous reports for FUS and other proteins with LC sequences capable of undergoing liquid-to-solid phase transitions (12,21). After incubation for a full month, these fibrillar structures had grown into a dense network covering the microscope slide. Two-month-old samples were also subjected to negative staining and transmission electron microscopy (TEM) in which dense clumps of fibers were also observed (Fig. S3). In contrast to wild-type FUS LC, the FUS LC G156E mutant formed fibrillar species much more rapidly, with clear fibrillar protrusions appearing after only a day of incubation (Fig. 1 *B*). Thus, the FUS LC and FUS LC G156E constructs recapitulate the transformation behavior of the full-length protein at neutral pH as described previously (19,69,70).

The maturation of FUS LC droplets was further interrogated by fluorescence recovery after photobleaching (FRAP) experiments (Fig. S4 A). Although the recovery kinetics of FUS LC droplets decreased over the course of several weeks, fluorescence recovery was not completely abrogated even after 8 weeks, suggesting that mobile components remained in the phase-separated structures even after this long period of incubation. The FUS LC G156E sample also exhibited fast recovery at the beginning (Fig. S4 B) but transitioned quickly to a fibrillar state with limited mobility.

The observation of fibrillar structures in our FUS LC and FUS LC G156E samples by microscopy prompted us to determine whether these structures have the characteristics of β-sheet-rich amyloid fibers. ThT is a switchable small molecule fluorophore that is often used as a test for the presence of amyloid (71). We initiated LLPS of FUS LC and FUS LC G156E samples and performed a binding assay with ThT (Fig. 1 *C*). Consistent with the trends observed in the microscopy experiments, FUS LC G156E formed fibers within the first 2 days and with similar kinetics to a positive control sample containing the amyloidogenic Aβ-peptide. On the other hand, ThT fluorescence of the wild-type FUS LC sample started to increase after approximately a week, whereas a negative control sample containing solubilized FUS LC in high pH CAPS buffer did not form amyloid over the course of the experiment. Although ThT fluorescence is also sensitive to the viscosity of the solvent (72) and may increase in the liquid droplet and gel environments, the changes in fluorescence intensity are consistent with the timeline of fibrillar protrusion formation in the microscopy experiments. Therefore, we interpret the increase in ThT fluorescence for the FUS LC and the FUS LC G156E samples as an indication of β-sheet amyloid formation. We also performed the experiment in the presence of 1,6-hexanediol, a small molecule known to disrupt protein droplets (Fig. S5; (73)). In this case, no ThT fluorescence was observed over the course of this experiment indicating that under our experimental conditions LLPS and amyloid formation are coupled.

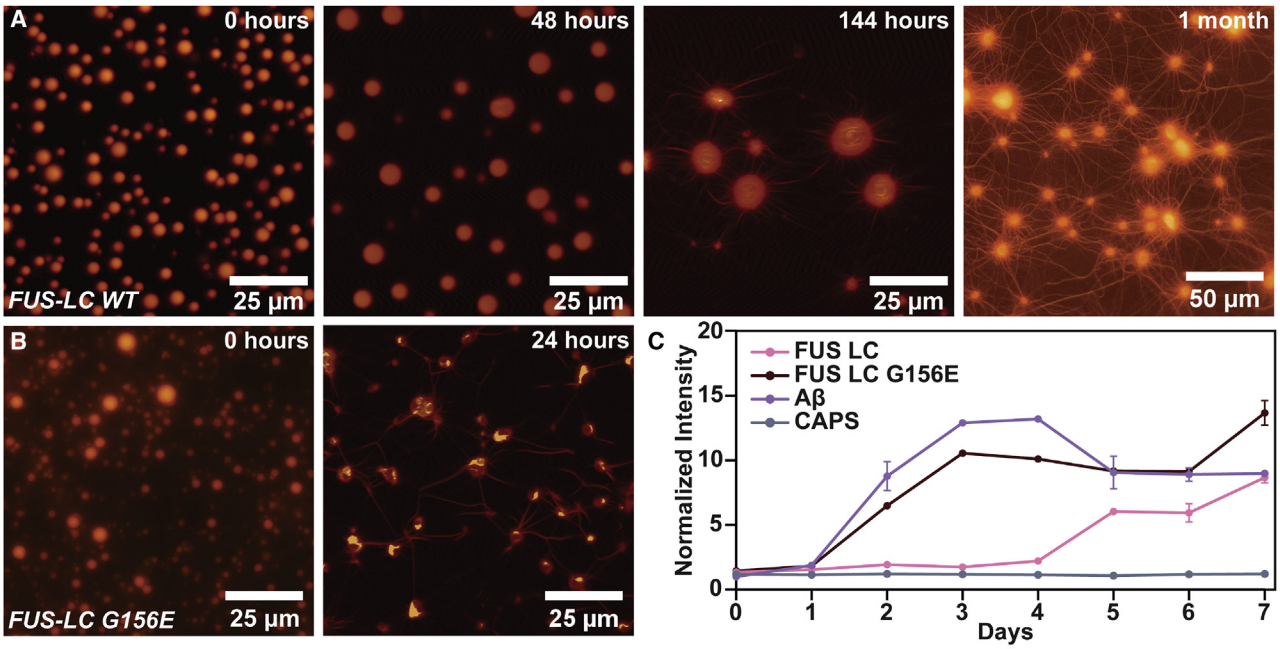


FIGURE 1 Microscopy and ThT analysis of FUS LC and FUS LC G156E maturation. (*A*) Maturation of wild-type FUS LC droplets. Droplets coarsen and loosen their spherical character after 48 h. Fibrils begin to appear after 144 h. After a full month, a dense fibril network is visible. (*B*) Maturation of FUS LC G156E droplets. The progression of the G156E mutant is accelerated, with aggregates and fibrils appearing after 24 h. (*C*) Fibrillization rates of FUS LC and FUS LC G156E characterized by ThT. The Aβ peptide was used as positive control whereas FUS LC in CAPS buffer (pH 11) served as negative control. The error bars represent standard deviation from three independent measurements.

Taken together, the imaging, FRAP, and ThT data indicate that wild-type FUS LC droplets can coarsen into gel-like states and mature into amyloid fibrils over time under near-physiological conditions. For the FUS LC G156E mutant, the transition from droplets to gels is more abrupt and the formation of the amyloid state is significantly accelerated.

The structure and dynamics of FUS LC change during the transformation process

The interactions and dynamics of FUS LC in the liquid-droplet state have been characterized extensively by solution NMR (31,32,48). On the other hand, solid-state MAS NMR spectroscopy and cryo-EM have revealed several possible amyloid states that can be formed by FUS LC (16,35,36,68). We sought to establish a connection between these observations by following the transformation from droplet to amyloid on the molecular level in real time using MAS NMR. The maturation of FUS LC droplets into gels and fibrils is ideally suited to this approach as the transformation of the wild-type protein takes several weeks and, thus, allows sufficient time for the collection of multidimensional NMR experiments at different time points.

To capture the range of dynamic regimes within the sample, we performed two different types of MAS NMR experiments. First, the INEPT-based pulse sequence was used to capture mobile components in the sample. This can include the mobile segments of an otherwise slow tumbling protein system or the mobile subpopulation of proteins in a heterogeneous sample (16,47,74,75). Second, dipolar-based experiments such as cross-polarization (CP) were used to describe the slow tumbling (rigid) components of the sample like those subpopulations in the gel and amyloid states (76,77). As control, we also used a direct polarization (DP) ^{13}C experiment which detects all carbon atoms in the sample irrespective of their mobility. We prepared two samples, one of ^{15}N,^{13}C FUS LC and one of ^{15}N,^{13}C FUS LC G156E, and subjected each to LLPS. The LLPS droplets were then gently collapsed into a single condensed phase and into an MAS NMR rotor. The final concentration of both FUS LC constructs was ~23 mM; within the range of the expected concentration of FUS LC in LLPS droplets (32). We used INEPT, CP, and DP experiments to follow the transformation of the samples over the course of 30 days for FUS LC and 12 days for FUS LC G156E (Fig. 2).

The two samples exhibited different behavior over the course of these experiments. For the wild-type FUS LC sample, the CP signal started to increase after a week, eventually reaching a plateau at 3.5 times the original integrated signal (Fig. 2, *A* and *C*). This indicates the emergence of rigid components in the sample with a timeline that is consistent with the formation of the fibrillar species detected by microscopy and ThT fluorescence. Furthermore, a comparison of the initial and final CP spectra shows the appearance of several new peaks (Fig. 2 *A*; Fig. S6, *A* and *B*). This includes peaks

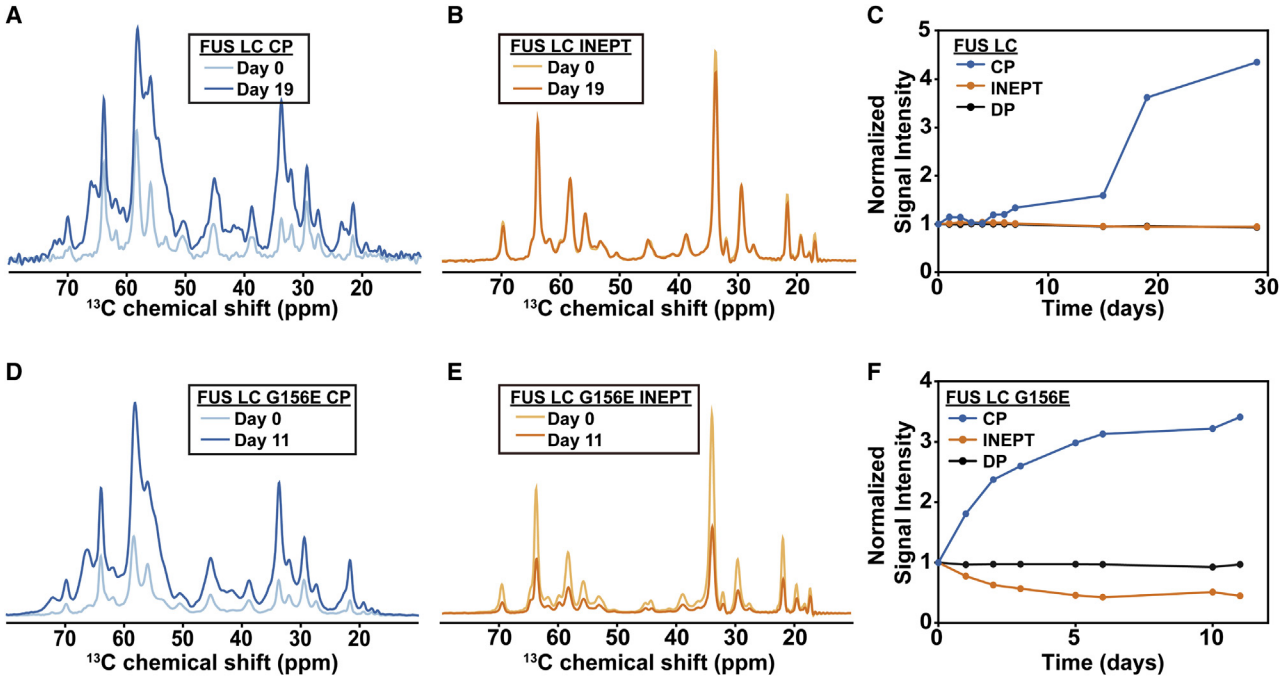


FIGURE 2 Characterization of FUS LC and FUS LC G156E maturation by MAS NMR spectroscopy. (*A*) Early and late-stage FUS LC CP spectra are shown. (*B*) Early and late-stage FUS LC INEPT spectra are shown. (*C*) Integrated signal intensity of the aliphatic region of FUS LC over time for CP, INEPT, and DP experiments is shown. (*D*) Early and late-stage FUS LC G156E CP spectra are shown. (*E*) Early and late-stage FUS LC G156E INEPT spectra are shown. (*F*) Integrated signal intensity of the aliphatic region of FUS LC G156E over time for CP, INEPT and DP experiments is shown. The error bars are based on the integrated noise level and are too small to visualize.

consistent with the chemical shifts of threonine, serine, and potentially glycine, asparagine, aspartate, and tyrosine. These peaks represent new chemical environments and suggest that the rigid components in the sample undergo a structural change over time, perhaps from an unstructured or oligomeric state to a β-sheet-rich amyloid fold. In contrast, the FUS LC G156E sample changed much more rapidly over the course of a week (Fig. 2, *D* and *F*). A comparison of the CP spectra shows a 3-fold increase in signal and the appearance of new structural environments during this time (Fig. S6, *C* and *D*). Overall, the initial and final CP spectra have similar shapes between the two samples although there are differences in intensity and line width for some peaks (Fig. S7, *A* and *C*).

The time course of the INEPT experiments is also noteworthy. Although the INEPT spectra of the wild-type protein changed very little over the course of three weeks (Fig. 2, *B* and *C*), the signals of the G156E sample decreased significantly during the first week, and then stabilized at ~50% of the initial intensity (Fig. 2, *E* and *F*). In these heterogenous samples, the INEPT signals are consistent with random coil confirmation and potentially arise from three different processes. First, they may reflect mobile monomers or low-molecular-weight oligomers that experience fast diffusion and rotational correlation times. These mobile components should also be detectable in FRAP experiments and, as our data indicate, their contribution diminishes over time (Fig. S4). Second, the INEPT signals may result from the gel phase. The gel state forms early on and involves the formation of noncovalent cross-linking interactions between monomers. These interactions may involve only certain regions of the monomer while other segments remain mobile enough to be detected by INEPT. We have observed this scenario in other protein systems that undergo gelation (47). And finally, the INEPT signals may arise from mobile regions of the amyloid state. For example, the published structure of seeded amyloid fibers of FUS LC (1–214) has a rigid β-sheet core that spans 57 residues (39–95), while the C-terminus is dynamic and unstructured and clearly visible in INEPT experiments (16). In addition, the relative ratio of gel to amyloid might determine the INEPT trajectory. For example, in the wild-type sample the buildup of β-sheet structure is slow, and the majority of the protein most likely remains in the gel state throughout the time course of the experiment. In a situation where the gel state dominates and there is a large mobile component, the INEPT signal should remain relatively constant over time, which is what we observe for the FUS LC sample. On the other hand, the G156E mutant sample quickly converts to amyloid, thus reducing the contribution of the gel state. Even if the final fibril structure contains mobile regions, the overall INEPT signal may still decrease. A final note regarding dynamics is that in heterogeneous and viscous samples such as this one, there are often motions on an intermediate timescale that are too slow for INEPT-based experiments and too fast for CP-based experiments (46). Thus,

it is possible that our experimental strategy is "blind" to some of the cross-linking interactions that contribute to the gel state or to some of the oligomeric species that lead to amyloid formation.

It should be noted that the slow transformation from a liquid-like to a more rigid β-sheet-rich state described here is distinct from the temperature dependent formation of reversible gels. Previous work has indicated that, upon cooling, FUS LC liquid condensed phases rapidly become a rigid, opaque gel, resulting in line broadening and loss of signal in the ^{1}H–^{15}N HSQC spectra acquired under solution NMR conditions (31). This reversible gel state forms quickly (within seconds), melts quickly (with seconds of heating), and is distinct from the "hydrogel" state (78), which forms by transition (rather than phase separation) of a solution of concentrated protein into amyloid fibrils. Using MAS NMR spectroscopy, we observe a similar reversible phenomenon with our initial stage FUS LC droplet samples (Fig. S8). When the sample temperature was decreased by 15°C, the ^{13}C INEPT signals almost disappeared. At the same time, the CP signals arising from rigid components increased in intensity but also in line width. The process was reversible, as an increase in temperature brought back the signals to their initial intensity and linewidth (although some small changes in the shape of the CP spectrum were noted). Similar to the solution NMR case, this behavior is consistent with the formation of reversible gels. Unlike the slow transformation described above, however, the CP signals at low temperatures did not indicate the formation of new β-sheet structures.

LLPS results in amyloid fibers with distinct structures

Intrigued by the structural changes detected in the 1D MAS NMR experiments described above, we recorded two-dimensional (2D) correlation spectra of the end-states of the two samples (30 days for FUS LC and 12 days for FUS LC G156E). In particular, we extended the 1D INEPT experiment into a 2D ^{1}H–^{13}C INEPT correlation spectrum that provides a more detailed picture of the residues that remain mobile at the end of the time course (Fig. 3 *A*). The 2D INEPT spectra of the wild-type and G156E samples contain similar amino acid types (including glycine, threonine, serine, glutamine, alanine, proline, and methionine) with chemical shifts that are consistent with random coil. The INEPT spectrum of the G156E construct has lower intensity overall despite the comparable number of scans and sample amount in the rotor, consistent with the 50% reduction of INEPT signal observed over the course of the 1D experiments. Despite the general overlap between the two spectra, a careful comparison reveals differences in the glycine region where one set of glycine residues remains the same and another set experiences a different chemical environment. This is also supported by the 1D INEPT spectra (Fig. S7, *B* and *D*) in which the G156E Gly Cα-region clearly shows two peaks. Although we were not able to assign these two different glycine groups, there appears to be subtle differences between the dynamic environments in the wild-type and mutant FUS LC samples.

To complement the residue-specific studies of mobile sample components, we also recorded 2D ^{13}C–^{13}C DARR spectra of the final state (Fig. 3 *B*). This experiment relies on a dipolar-assisted rotational resonance condition to recouple ^{13}C atoms in protein regions that experience reduced motions and rotational correlation times (79). Given the composition of the two samples, this experiment is expected to report on the rigid core of the final amyloid state. Both samples gave rise to relatively well-resolved DARR spectra where many individual crosspeaks could be identified and analyzed. To determine whether the spectra indeed report on a β-sheet-rich amyloid fold, we compared the position of the crosspeaks against a statistical analysis of chemical shifts deposited in the BMRB and the PDB (Fig. S9). Plotting these distributions with our DARR spectrum reveals that many of the correlations lie nearer to the center of the β-sheet distribution than the coil-like distribution. These observations suggest that the rigid components appearing in our DARR spectrum contain β-sheet structure.

Unlike previous studies of FUS LC fibrils (16,35,68), which used seeded fibers in isolation, our samples are a heterogeneous mixture of gel-like and fibrillar states. Therefore, the signal-to-noise in the dipolar experiments is relatively low, and this precluded the collection of other multidimensional MAS NMR correlations (e.g., three-dimensional NCACX and three-dimensional NCOCX). While this prevented us from performing site-specific assignments of the DARR spectra, we were able to compare our data with the DARR spectra of the published FUS LC (1–214) structure of seeded amyloid fibers (Fig. S10; (16)). This comparison revealed significant differences between the two spectra. First, our spectrum is relatively broader, indicating the presence of heterogeneity and/or intermediate dynamics that can interfere with the timescale of the NMR experiment and cause line broadening. This is expected as our sample is more heterogenous and dynamic by design. Second, a significant number of crosspeaks appear to be shifted or missing from our spectrum. This includes crosspeaks for the unique residues Asp46 and Pro72 that appear to be substantially shifted as well as crosspeaks for residues such as Gln93 that are not present altogether. Other residues that potentially experience different environments or are not part of the core include Thr45, Thr47, Ser70, Thr78, and Ser90.

Recent structural studies of FUS LC fibrils have indicated the presence of two possible β-sheet cores (16,68). In constructs that comprise residues 1–214 or 1–110, the β-sheet core spans residues 39–95 and has the same structure in both cases. In a construct that lacks the N-terminal portion of the sequence and covers residues 111–214, a different β-sheet core can be formed by residues

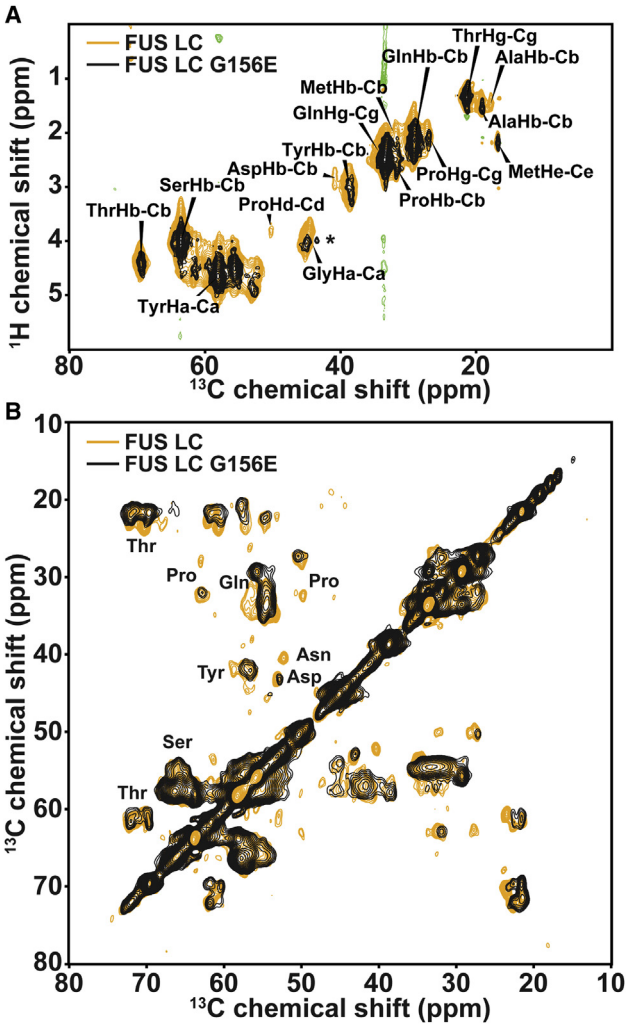


FIGURE 3 2D correlation spectra of FUS LC and FUS LC G156E. (*A*) End-state ^{1}H–^{13}C INEPT spectra of FUS LC (*orange*) and FUS LC G156E (*black*) are shown. The asterisks denotes a new glycine crosspeak that appears in the FUS LC G156E sample. (*B*) End-state ^{13}C–^{13}C DARR correlation spectra of FUS LC (*orange*) and FUS LC G156E (*black*) are shown.

112–150. Because our spectrum clearly contains aspartate and multiple threonine residues, amino acids that are only present or enriched in the N-terminal portion of the sequence, we conclude that the core of the LLPS fibrils investigated here is generally located in a similar region as the 39–95 β-sheet core. However, the structure of the core is different for the LLPS-derived fibrils as manifested by the significant chemical shift changes for some crosspeaks. The missing signals for some residues (e.g., Gln93) might also indicate that the β-sheet core of the LLPS-derived fibrils is smaller compared to the 39–95 core of the seeded FUS LC (1–214) fibers. It should also be noted that our DARR spectrum contains a much more intense tyrosine Cα-Cβ-region, potentially implying a more prominent role for these residues in the structure and dynamics of the final state.

Finally, we compared the DARR spectra of the wild-type FUS LC and the G156E mutant (Fig. 3 *B*). The two spectra are much more similar to each other than to the DARR spectrum of the seeded FUS LC (1–214) amyloids. Common features among the LLSP-derived samples include the overlapping signals for Pro72 and Asp46, the prominently missing Gln93 signal, and the overlapping shapes of the threonine and serine regions of the spectra. Therefore, it appears that the LLPS-derived fibers of FUS LC and FUS LC G156E share a similar β-sheet core located in the N-terminal segment of the protein sequence. The structure and dynamics of the two folds, however, are not completely identical, as indicated by the loss of intensity and small chemical shift changes for some crosspeaks in the G156E sample.

Therefore, the combined analyses of the INEPT and DARR spectra discussed above reveal several important observations. LLPS-derived FUS LC (1–163) amyloid fibrils have a disordered C-terminus similar to the seeded fibers of FUS LC (1–214). Similarly, the LLPS-derived fibers favor an N-terminal β-sheet core approximately centered around residues 39–95 of the protein sequence. This core, however, is potentially smaller and distinctly folded compared to the N-terminal core of fibers obtained through seeding. Therefore, the liquid droplet environment appears to influence the structure of the coupled amyloid state. The crowded conditions during LLPS may favor different types of protein-protein interactions that lead to the observed differences. For example, the valence and patterning of tyrosine residues is an important determinant of FUS LLPS (4), and amyloid formation involves the formation of steric zippers with good side-chain complementarity (16,68,80). The interactions of the tyrosine residues therefore may fine tune the exact position of the steric zippers in the FUS LC fibers. FUS is an important component of cellular stress granules which have LLPS properties and are often misregulated in amyotrophic lateral sclerosis (81,82). Thus, LLPS conditions may also influence the structure of pathogenic FUS fibers in neurodegenerative disease.

The G156E mutation exposes the amyloidogenic core of FUS LC

Previous work has shown that phosphorylation of serine and threonine residues in the FUS LC domain can inhibit LLPS (16,48). Incubation at high pH also disrupts LLPS likely because of the deprotonation and subsequent negative charge of tyrosine side chains in FUS LC under these conditions (31). It is therefore noteworthy that the disease-relevant G156E mutant, which introduces a negatively charged residue into FUS LC, can not only undergo LLPS but also exhibits faster aggregation kinetics than the wild-type. Furthermore, based on the published structures of seeded fibrils (16,68) and our MAS NMR analysis, this residue does

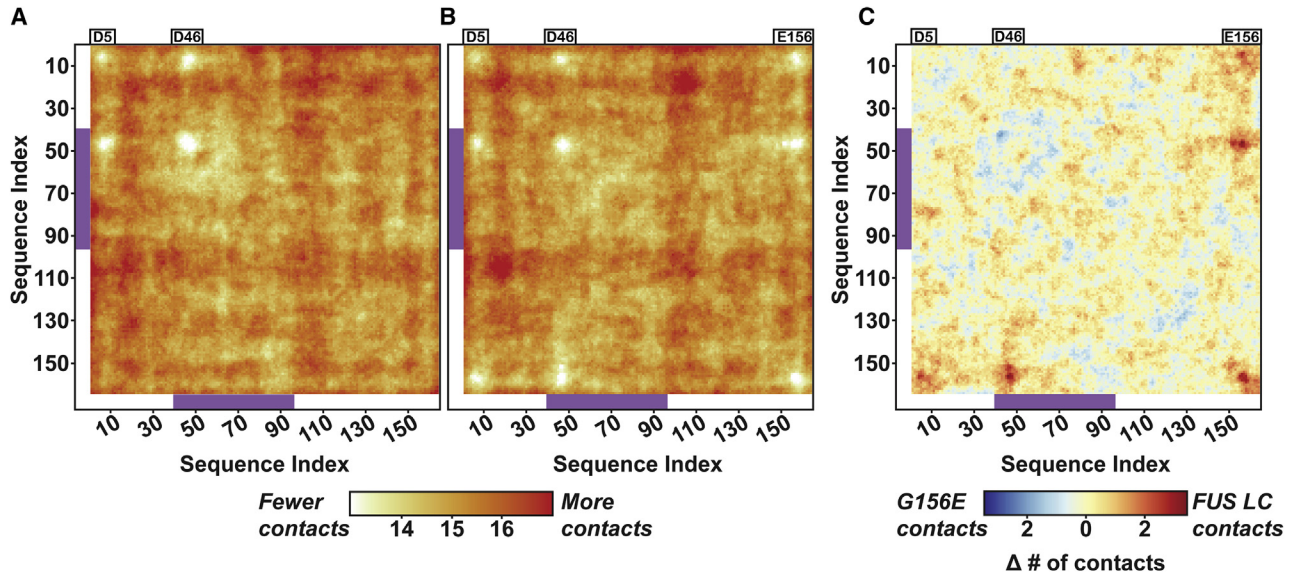


FIGURE 4 Coarse-grained simulations of FUS LC and FUS LC G156E droplets. (*A*) Intermolecular contact map for FUS LC is shown. Regions with a large number of contacts appear in orange, whereas regions with a relatively low number of contacts appear in white. (*B*) Intermolecular contact map for FUS LC G156E is shown. The color coding for both maps is the same and both maps represent contacts within a 15-Å cutoff. Distance is measured from the bead centers. (*C*) Difference map in which red indicates regions where FUS LC makes more contacts, whereas blue indicates regions where FUS LC G156E makes more contacts. The purple rectangles denote the fibril core of FUS LC (1–214) as determined in Ref. 16.

not appear to participate in the ordered fibril core. Therefore, it is still unclear how this mutation promotes amyloidogenesis from an LLPS state. To gain insight into the early events that may drive this behavior, we used coarse-grained molecular dynamics simulations to model interactions in liquid droplets formed by wild-type FUS LC and the FUS LC G156E mutant. We applied a hydrophobicity scale model (64), initially described by Dignon et al., which is capable of recapitulating LLPS in silico and has been used to analyze interactions in wild-type FUS, FUS LC, and phosphomimetic substitutions of FUS LC (27,32).

Separate simulations of 100 dispersed FUS LC or FUS LC G156E monomers in random configurations were prepared and allowed to proceed for 1.1 μs. In both simulations, dispersed monomers rapidly coalesced into a dynamic droplet-like state within 30 ns. The droplet persisted throughout the simulation (Fig. S11), with some monomers occasionally breaking free from the LLPS droplet and rejoining after a few ns. We analyzed the coalesced state by constructing intermolecular contact maps that represent the number of pair-wise amino acid contacts within a specified radius. For each simulation run, 100 frames from the droplet state were extracted and interparticle distances were measured for all particles in the simulation. Particles that were within the specified distance radius were counted as a contact. Intramolecular (i.e., intramonomer) interactions were not counted. The mean number of contacts across all frames and monomers was used to construct the contact maps.

In alignment with previous simulations of FUS LC, there are no notable regions of high contact propensity within a contact radius of up to 7 Å (Fig. S12; (32)). This is also consistent with experimental literature that suggests that LLPS of FUS LC is not driven by specific regions in the sequence (32). At longer contact distances (8–15 Å), however, some features start to appear in the contact map for FUS LC (Fig. 4 *A*; Fig. S12). This includes 1) fewer contacts between the regions surrounding the two native negatively charged residues, Asp5 and Asp46, and 2) increased contacts between a segment encompassing residues 15–25 and another segment comprising the more aliphatic region of FUS LC between residues 99–110. It is noteworthy that those interacting regions flank the rigid fibril core of the seeded fibers (39–95) but are not part of it.

The 15 Å contact map for FUS LC G156E appears similar to the contact map of the wild-type construct (Fig. 4, *B* and *C*) with one exception. The introduction of a negative charge at position 156 decreases the number of contacts between the C-terminus and the N-terminal Asp5 and Asp 46 so that two additional repulsive regions are observed in the map. Overall, the repulsion between these negatively charged residues results in reduced interactions of the C-terminus with the N-terminal part of the protein. On the other hand, the higher contact propensities of regions 15–25 and 99–110 do not appear to be affected by the presence of the additional negative charge. These conclusions hold across three different simulations with independent starting conditions for each sample and are not caused by surface effects (Fig. S13 and S14). In light of these observations, we propose that the introduction of the G156E mutation into FUS LC reduces the number of

nonproductive interactions between the termini and leads to greater exposure of the inner segments of the protein. This increases the propensity for protein-protein interactions that, in turn, leads to amyloid formation on an accelerated timescale.

CONCLUSION

The LC domain of FUS can undergo phase transitions that lead to the formation of liquid droplets, gels, and amyloid. Our work demonstrates that that FUS LC droplets formed at neutral pH mature into amyloid-like states over the course of several weeks, a transition that can be significantly accelerated by the disease-relevant mutation G156E. The slow kinetics of the process allowed us to follow this transition by MAS NMR spectroscopy and to observe the formation of characteristic β-sheet signatures at residue-specific resolution in real time. Whereas the resulting β-sheet core of these LLPS-derived fibrils is located in the same region of the sequence as the rigid core formed by seeded fibers, the core structure of the LLPS-derived fibrils appears to be different. Furthermore, when the LLPS environment was eliminated by the addition of 1,6-hexanediol, no amyloid fibrils were observed in the course of the experiment. These observations imply that liquid droplet environments can influence both the efficiency of fibril formation and their structural signatures. As FUS is often located in stress granules in the cell (81), the crowded conditions promoted by LLPS may play an important role in its pathogenic amyloid formation.

In contrast to the differences observed between the fibril core structure of seeded FUS LC fibrils and LLPS-derived fibrils, the fibril core structures of the LLPS-derived FUS LC and FUS LC G156E constructs appear relatively similar to each other. The most striking differences between the wild-type protein and the pathological mutant were maturation rate rather than the final structure, with the FUS LC G156E protein forming fibrils much earlier than its wild-type counterpart under the same conditions. Although the wild-type FUS LC sample has a significant mobile component even after several weeks, the G156E mutant matures much more rapidly and loses mobility over the course of several days. Thus, a single negatively charged mutation at a region of the sequence that does not participate in fibril core interactions can have dramatic consequences on the fibrillization process. Our coarse-grained simulations suggest that the presence of this negatively charged residue near the C-terminus increases the repulsion between the termini and promotes enhanced intermolecular contacts between inner regions of the protein. This mechanism may also explain the increased aggregation propensity of the full-length G156E mutant (19).

Our approach presented here combines imaging, MAS NMR spectroscopy, and computational methods to provide new molecular insights into the elusive transformation of a protein from the liquid droplet to the gel and amyloid states. Our strategy is versatile and provides a platform for the analysis of other proteins that exhibit similar behavior in viscous and heterogeneous environments. This approach can also be extended to explore the role of relevant biological components and small molecules on phase separation, gelation and amyloid formation.

SUPPORTING MATERIAL

Supporting Material can be found online at https://doi.org/10.1016/j.bpj.2021.02.008.

AUTHOR CONTRIBUTIONS

R.F.B. and G.T.D. conceived the project and wrote the manuscript. R.F.B. prepared samples, performed microscopy, ThT assays, NMR experiments, and computational studies. M.K. performed ThT assays and NMR experiments. G.T.D. performed NMR experiments and supervised the study. All authors analyzed the data and commented on the manuscript.

ACKNOWLEDGMENTS

We are grateful to A. de Angelis for help with the NMR spectrometers, to J. Yang's lab at UCSD for help with the ThT assays, to T. Meerloo for help with TEM, to J. Mittal for generously sharing simulations code, and to N. Fawzi for helpful discussions and comments.

This work was supported by a Research Education Component associated with NIH Grant P30 AG062429, an R21 AG069064 award to G.T.D., a T32 GM112584 fellowship to R.F.B., the BRC for NMR Molecular Imaging of Proteins at UCSD (P41 EB002031), and the UCSD Microscopy Core (NINDS NS047101).

REFERENCES

1. Soto, C., and S. Pritzkow. 2018. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nat. Neurosci.* 21:1332–1340.
2. Mathieu, C., R. V. Pappu, and J. P. Taylor. 2020. Beyond aggregation: pathological phase transitions in neurodegenerative disease. *Science.* 370:56–60.
3. Wang, J., J.-M. Choi, …, A. A. Hyman. 2018. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell.* 174:688–699.e16.
4. Martin, E. W., A. S. Holehouse, …, T. Mittag. 2020. Valence and patterning of aromatic residues determine the phase behavior of prion-like domains. *Science.* 367:694–699.
5. Fitzpatrick, A. W. P., B. Falcon, …, S. H. W. Scheres. 2017. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature.* 547:185–190.
6. Lin, Y., Y. Fichou, …, S. Han. 2020. Electrostatically driven complex coacervation and amyloid aggregation of tau are independent processes with overlapping conditions. *ACS Chem. Neurosci.* 11:615–627.
7. Wegmann, S., B. Eftekharzadeh, …, B. T. Hyman. 2018. Tau protein liquid-liquid phase separation can initiate tau aggregation. *EMBO J.* 37:e98049.
8. Dregni, A. J., V. S. Mandala, …, M. Hong. 2019. In vitro 0N4R tau fibrils contain a monomorphic β-sheet core enclosed by dynamically heterogeneous fuzzy coat segments. *Proc. Natl. Acad. Sci. USA.* 116:16357–16366.

9. Ambadipudi, S., J. Biernat, …, M. Zweckstetter. 2017. Liquid-liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nat. Commun.* 8:275.

10. Ambadipudi, S., J. G. Reddy, …, M. Zweckstetter. 2019. Residue-specific identification of phase separation hot spots of Alzheimer's-related protein tau. *Chem. Sci. (Camb.).* 10:6503–6507.

11. Schweighauser, M., Y. Shi, …, M. Goedert. 2020. Structures of α-synuclein filaments from multiple system atrophy. *Nature.* 585:464–469.

12. Ray, S., N. Singh, …, S. K. Maji. 2020. α-Synuclein aggregation nucleates through liquid-liquid phase separation. *Nat. Chem.* 12:705–716.

13. Araki, K., N. Yagi, …, H. Mochizuki. 2019. Parkinson's disease is a type of amyloidosis featuring accumulation of amyloid fibrils of α-synuclein. *Proc. Natl. Acad. Sci. USA.* 116:17963–17969.

14. Tuttle, M. D., G. Comellas, …, C. M. Rienstra. 2016. Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nat. Struct. Mol. Biol.* 23:409–415.

15. Li, Y. R., O. D. King, …, A. D. Gitler. 2013. Stress granules as crucibles of ALS pathogenesis. *J. Cell Biol.* 201:361–372.

16. Murray, D. T., M. Kato, …, R. Tycko. 2017. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell.* 171:615–627.e16.

17. Gui, X., F. Luo, …, D. Li. 2019. Structural basis for reversible amyloids of hnRNPA1 elucidates their role in stress granule assembly. *Nat. Commun.* 10:2006.

18. Babinchak, W. M., R. Haider, …, W. K. Surewicz. 2019. The role of liquid-liquid phase separation in aggregation of the TDP-43 low-complexity domain. *J. Biol. Chem.* 294:6306–6317.

19. Patel, A., H. O. Lee, …, S. Alberti. 2015. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell.* 162:1066–1077.

20. Lin, Y., D. S. Protter, …, R. Parker. 2015. formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell.* 60:208–219.

21. Liu, Z., S. Zhang, …, C. Liu. 2020. Hsp27 chaperones FUS phase separation under the modulation of stress-induced phosphorylation. *Nat. Struct. Mol. Biol.* 27:363–372.

22. Mitrea, D. M., B. Chandra, …, R. W. Kriwacki. 2018. Methods for physical characterization of phase-separated bodies and membraneless organelles. *J. Mol. Biol.* 430:4773–4805.

23. Mittag, T., and J. D. Forman-Kay. 2007. Atomic-level characterization of disordered protein ensembles. *Curr. Opin. Struct. Biol.* 17:3–14.

24. Martin, E. W., and T. Mittag. 2018. Relationship of sequence and phase separation in protein low-complexity regions. *Biochemistry.* 57:2478–2487.

25. van der Lee, R., M. Buljan, …, M. M. Babu. 2014. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* 114:6589–6631.

26. Brangwynne, C. P., P. Tompa, and R. V. Pappu. 2015. Polymer physics of intracellular phase transitions. *Nat. Phys.* 11:899–904.

27. Dignon, G. L., W. Zheng, …, J. Mittal. 2018. Sequence determinants of protein phase behavior from a coarse-grained model. *PLOS Comput. Biol.* 14:e1005941.

28. Pak, C. W., M. Kosno, …, M. K. Rosen. 2016. Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein. *Mol. Cell.* 63:72–85.

29. Feric, M., N. Vaidya, …, C. P. Brangwynne. 2016. Coexisting liquid phases underlie nucleolar subcompartments. *Cell.* 165:1686–1697.

30. Dignon, G. L., R. B. Best, and J. Mittal. 2020. Biomolecular phase separation: from molecular driving forces to macroscopic properties. *Annu. Rev. Phys. Chem.* 71:53–75.

31. Burke, K. A., A. M. Janke, …, N. L. Fawzi. 2015. Residue-by-Residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Mol. Cell.* 60:231–241.

32. Murthy, A. C., G. L. Dignon, …, N. L. Fawzi. 2019. Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain. *Nat. Struct. Mol. Biol.* 26:637–648.

33. Molliex, A., J. Temirov, …, J. P. Taylor. 2015. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell.* 163:123–133.

34. Banani, S. F., A. M. Rice, …, M. K. Rosen. 2016. Compositional control of phase-separated cellular bodies. *Cell.* 166:651–663.

35. Murray, D. T., and R. Tycko. 2020. Side chain hydrogen-bonding interactions within amyloid-like fibrils formed by the low-complexity domain of FUS: evidence from solid state nuclear magnetic resonance spectroscopy. *Biochemistry.* 59:364–378.

36. Ding, X., F. Sun, …, S.-Z. Luo. 2020. Amyloid-forming segment induces aggregation of FUS-LC domain from phase separation modulated by site-specific phosphorylation. *J. Mol. Biol.* 432:467–483.

37. Iadanza, M. G., R. Silvers, …, S. E. Radford. 2018. The structure of a β_2-microglobulin fibril suggests a molecular basis for its amyloid polymorphism. *Nat. Commun.* 9:4517.

38. Ader, C., S. Frey, …, M. Baldus. 2010. Amyloid-like interactions within nucleoporin FG hydrogels. *Proc. Natl. Acad. Sci. USA.* 107:6281–6285.

39. Kennedy, S. B., E. R. deAzevedo, …, M. Hong. 2001. Dynamic structure of a protein hydrogel: a solid-state NMR study. *Macromolecules.* 34:8675–8685.

40. Mandala, V. S., and M. Hong. 2019. High-sensitivity protein solid-state NMR spectroscopy. *Curr. Opin. Struct. Biol.* 58:183–190.

41. Marchanka, A., B. Simon, …, T. Carlomagno. 2015. RNA structure determination by solid-state NMR spectroscopy. *Nat. Commun.* 6:7024.

42. Andrew, E. R., A. Bradbury, and R. G. Eades. 1958. Nuclear magnetic resonance spectra from a crystal rotated at high speed. *Nature.* 182:1659.

43. Maricq, M. M., and J. S. Waugh. 1979. NMR in rotating solids. *J. Chem. Phys.* 70:3300–3316.

44. Baldus, M., and B. H. Meier. 1996. Total correlation spectroscopy in the solid state. The use of scalar couplings to determine the through-bond connectivity. *J. Magn. Reson. A.* 121:65–69.

45. Takegoshi, K., S. Nakamura, and T. Terao. 2001. 13C–1H dipolar-assisted rotational resonance in magic-angle spinning NMR. *Chem. Phys. Lett.* 344:631–637.

46. Matlahov, I., and P. C. A. van der Wel. 2018. Hidden motions and motion-induced invisibility: dynamics-based spectral editing in solid-state NMR. *Methods.* 148:123–135.

47. Ackermann, B. E., and G. T. Debelouchina. 2019. Heterochromatin protein HP1α gelation dynamics revealed by solid-state NMR spectroscopy. *Angew. Chem. Int.Ed.* 58:6300–6305.

48. Monahan, Z., V. H. Ryan, …, N. L. Fawzi. 2017. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *EMBO J.* 36:2951–2967.

49. Schindelin, J., I. Arganda-Carreras, …, A. Cardona. 2012. Fiji: an open-source platform for biological-image analysis. *Nat. Methods.* 9:676–682.

50. Rueden, C. T., J. Schindelin, …, K. W. Eliceiri. 2017. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics.* 18:529.

51. Virtanen, P., R. Gommers, …, P. van Mulbregt; SciPy 1.0 Contributors. 2020. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods.* 17:261–272.

52. McKinney, W. 2010. Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference.* 445:56–61.

53. Hunter, J. D. 2007. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* 9:90–95.

54. Harris, C. R., K. J. Millman, …, T. E. Oliphant. 2020. Array programming with NumPy. *Nature.* 585:357–362.

55. Lee, W., W. M. Westler, …, J. L. Markley. 2009. PINE-SPARKY: graphical interface for evaluating automated probabilistic peak assignments in protein NMR spectroscopy. *Bioinformatics.* 25:2085–2087.

56. Lee, W., W. Yu, …, J. L. Markley. 2012. PACSY, a relational database management system for protein structure and chemical shift analysis. *J. Biomol. NMR.* 54:169–179.

57. Heinig, M., and D. Frishman. 2004. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* 32 (*Suppl 2*):W500–W502.

58. Helmus, J. J., and C. P. Jaroniec. 2013. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *J. Biomol. NMR.* 55:355–367.

59. Fritzsching, K. J., Y. Yang, …, M. Hong. 2013. Practical use of chemical shift databases for protein solid-state NMR: 2D chemical shift maps and amino-acid assignment with secondary-structure information. *J. Biomol. NMR.* 56:155–167.

60. Anderson, J. A., J. Glaser, and S. C. Glotzer. 2020. HOOMD-blue: a Python package for high-performance molecular dynamics and hard particle Monte Carlo simulations. *Comput. Mater. Sci.* 173:109363.

61. Schuster, B. S., G. L. Dignon, …, J. Mittal. 2020. Identifying sequence perturbations to an intrinsically disordered protein that determine its phase-separation behavior. *Proc. Natl. Acad. Sci. USA.* 117:11421–11431.

62. Conicella, A. E., G. L. Dignon, …, N. L. Fawzi. 2020. TDP-43 α-helical structure tunes liquid-liquid phase separation and function. *Proc. Natl. Acad. Sci. USA.* 117:5883–5894.

63. mphowardlab/azplugins, 2020. mphowardlab.

64. Kapcha, L. H., and P. J. Rossky. 2014. A simple atomic-level hydrophobicity scale reveals protein interfacial structure. *J. Mol. Biol.* 426:484–498.

65. Klein, C., J. Sallai, …, P. T. Cummings. 2016. A hierarchical, component based approach to screening properties of soft matter. *In* Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015. R. Q. Snurr, C. S. Adjiman, D. A. Kofke, R. Q. Snurr, C. S. Adjiman, and D. A. Kofke, eds, pp. 79–92.

66. Martínez, L., R. Andrade, …, J. M. Martínez. 2009. PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* 30:2157–2164.

67. Stukowski, A. 2009. Visualization and analysis of atomistic simulation data with OVITO–the Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.* 18:015012.

68. Lee, M., U. Ghosh, …, R. Tycko. 2020. Molecular structure and interactions within amyloid-like fibrils formed by a low-complexity protein sequence from FUS. *Nat. Commun.* 11:5735.

69. Murakami, T., S. Qamar, …, P. St George-Hyslop. 2015. ALS/FTD mutation-induced phase transition of FUS liquid droplets and reversible hydrogels into irreversible hydrogels impairs RNP granule function. *Neuron.* 88:678–690.

70. Nomura, T., S. Watanabe, …, Y. Furukawa. 2014. Intranuclear aggregation of mutant FUS/TLS as a molecular pathomechanism of amyotrophic lateral sclerosis. *J. Biol. Chem.* 289:1192–1202.

71. Naiki, H., K. Higuchi, …, T. Takeda. 1989. Fluorometric determination of amyloid fibrils in vitro using the fluorescent dye, thioflavin T1. *Anal. Biochem.* 177:244–249.

72. Stsiapura, V. I., A. A. Maskevich, …, K. K. Turoverov. 2008. Thioflavin T as a molecular rotor: fluorescent properties of thioflavin T in solvents with different viscosity. *J. Phys. Chem. B.* 112:15893–15902.

73. Kroschwald, S., S. Maharana, and A. Simon. 2017. Hexanediol: a chemical probe to investigate the material properties of membraneless compartments. *Matters (Zur.).* 3:e201702000010.

74. Andronesi, O. C., S. Becker, …, M. Baldus. 2005. Determination of membrane protein structure and dynamics by magic-angle-spinning solid-state NMR spectroscopy. *J. Am. Chem. Soc.* 127:12965–12974.

75. Debelouchina, G. T., G. W. Platt, …, R. G. Griffin. 2010. Magic angle spinning NMR analysis of β2-microglobulin amyloid fibrils in two distinct morphologies. *J. Am. Chem. Soc.* 132:10414–10423.

76. Hartmann, S. R., and E. L. Hahn. 1962. Nuclear double resonance in the rotating frame. *Phys. Rev.* 128:2042–2053.

77. Pines, A., M. G. Gibby, and J. S. Waugh. 1972. Proton-enhanced nuclear induction spectroscopy. A method for high resolution NMR of dilute spins in solids. *J. Chem. Phys.* 56:1776–1777.

78. Kato, M., T. W. Han, …, S. L. McKnight. 2012. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell.* 149:753–767.

79. Takegoshi, K., S. Nakamura, and T. Terao. 2003. 13C–1H dipolar-driven 13C–13C recoupling without 13C rf irradiation in nuclear magnetic resonance of rotating solids. *J. Chem. Phys.* 118:2325–2341.

80. Sawaya, M. R., S. Sambashivan, …, D. Eisenberg. 2007. Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature.* 447:453–457.

81. Wheeler, J. R., T. Matheny, …, R. Parker. 2016. Distinct stages in stress granule assembly and disassembly. *eLife.* 5:e18413.

82. Wolozin, B., and P. Ivanov. 2019. Stress granules and neurodegeneration. *Nat. Rev. Neurosci.* 20:649–666.

Biophysical Journal, Volume *120*

Supplemental information

Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC

Raymond F. Berkeley, Maryam Kashefi, and Galia T. Debelouchina

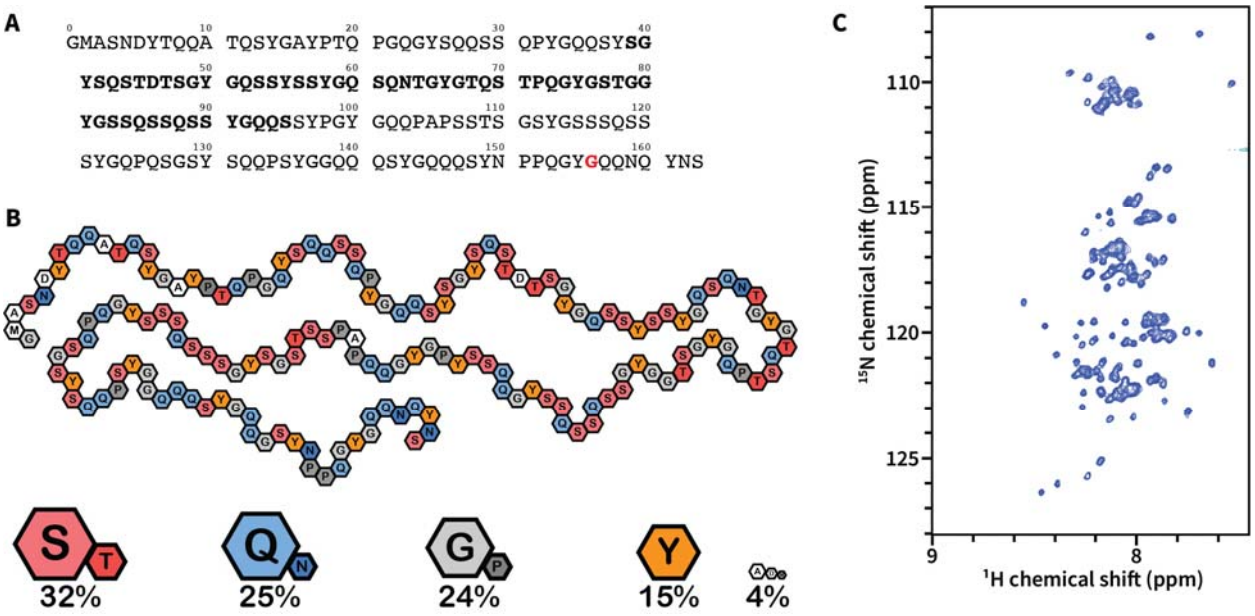


Figure S1. A) The sequence of FUS LC (1 -163) used in this work. The N-terminal glycine is left after cleavage with TEV protease. The seeded fibril core determined by Murray *et al.*[1] is highlighted in bold, and the site of the G156E mutation is marked in red. **B)** Amino acid composition analysis of the FUS LC construct. **C)** Solution NMR spectrum of 150 µM FUS LC obtained at 800 MHz in 20 mM MES, pH 5.5 at 25 °C similar to the conditions used by Burke *et al.*[2] where FUS LC can form droplets but not amyloids.

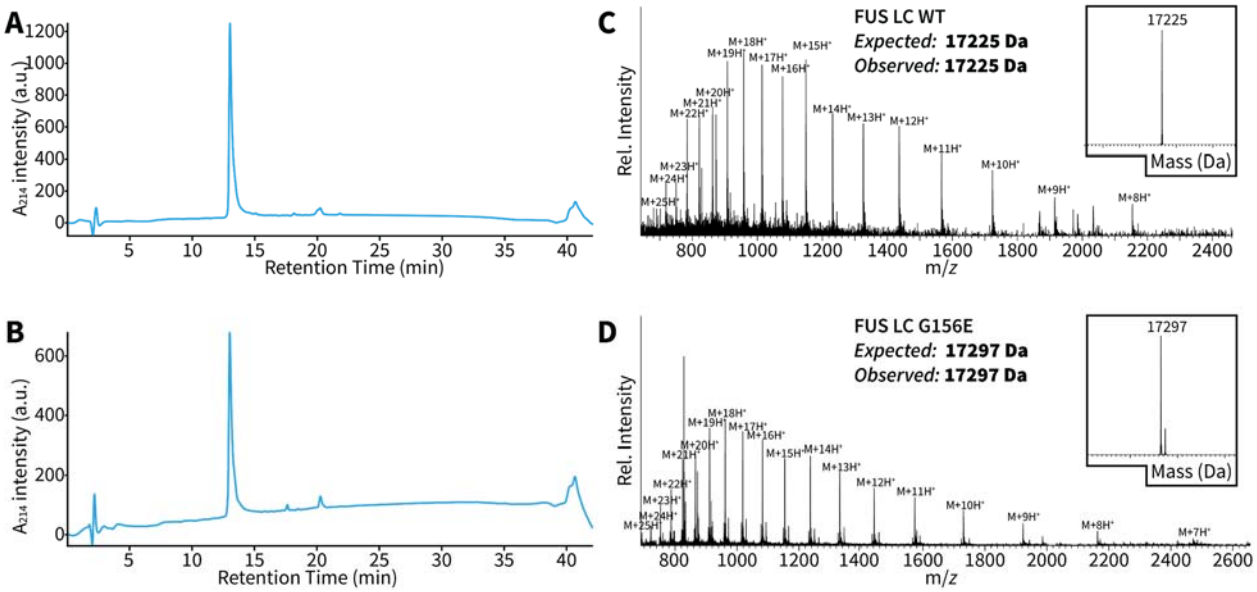


Figure S2. (A-B) Representative analytical C18 reverse-phase HPLC chromatograms of purified FUS LC and FUS LC G156E, respectively. **(C-D)** ESI-MS of the purified proteins.

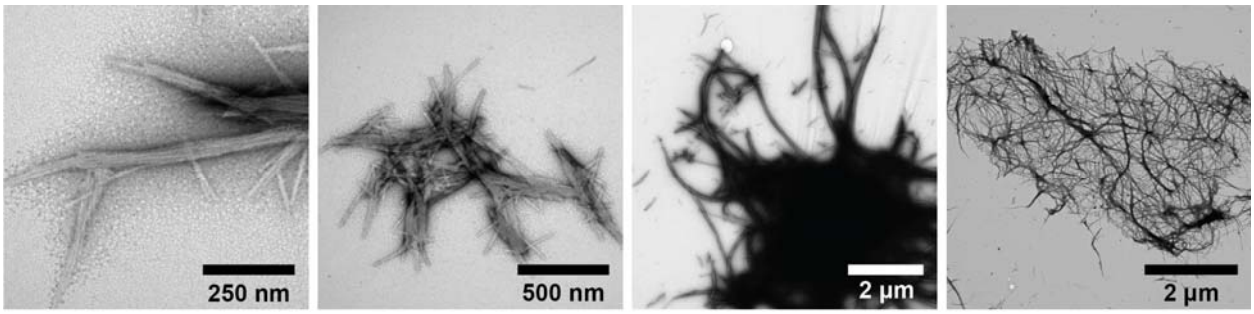


Figure S3. Representative transmission electron microscopy (TEM) images of a wild-type FUS LC sample after two months of maturation.

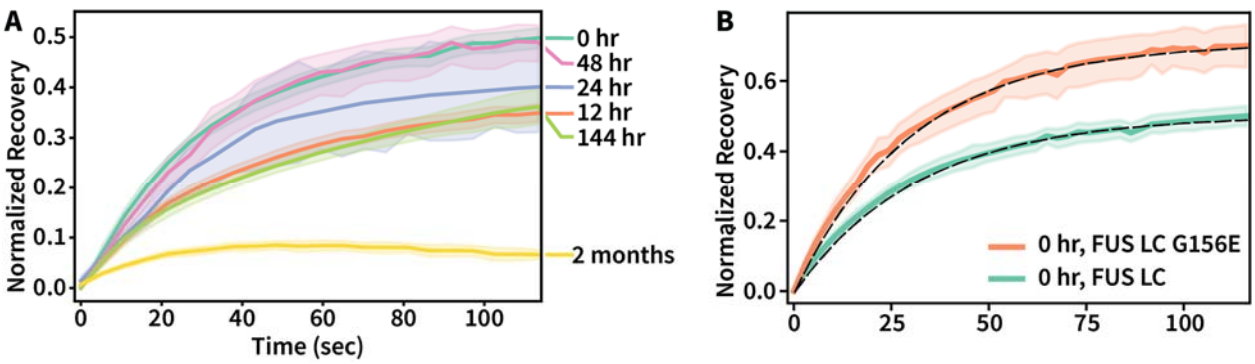


Figure S4. A) FRAP recovery curves of Cy3-labeled FUS LC droplets (5% Cy3-labeled protein) followed over the course of several weeks as they mature into gels and fibrillar species. During the first few days, recovery is highly heterogeneous as not all droplets have reached the same gel-like state. Six unique droplets were photobleached for each time point and data are plotted as the mean recovery with a 95% confidence interval. **B)** FRAP recovery of FUS LC G156E droplets recorded immediately after LLPS initiation indicating a high degree of mobility at the initial stages of the transformation process. Six unique droplets were photobleached for each time point and data are plotted as the mean recovery with a 95% confidence interval. The data were fit with the one phase association model in GraphPad Prism using the equation $Y = Y_0 + (Plateau - Y_0) * (1 - e^{-t/\tau})$ where Y_0 represents the beginning of the FRAP curve and is set to zero. The $Plateau$ represents the mobile fraction which is 0.50 ± 0.01 for the WT FUS LC sample and 0.71 ± 0.01 for the G156E FUS LC mutant. τ represents the characteristic recovery time which is 32 ± 2 s for the WT FUS LC sample, and 30 ± 2 for the G156E FUS LC mutant.

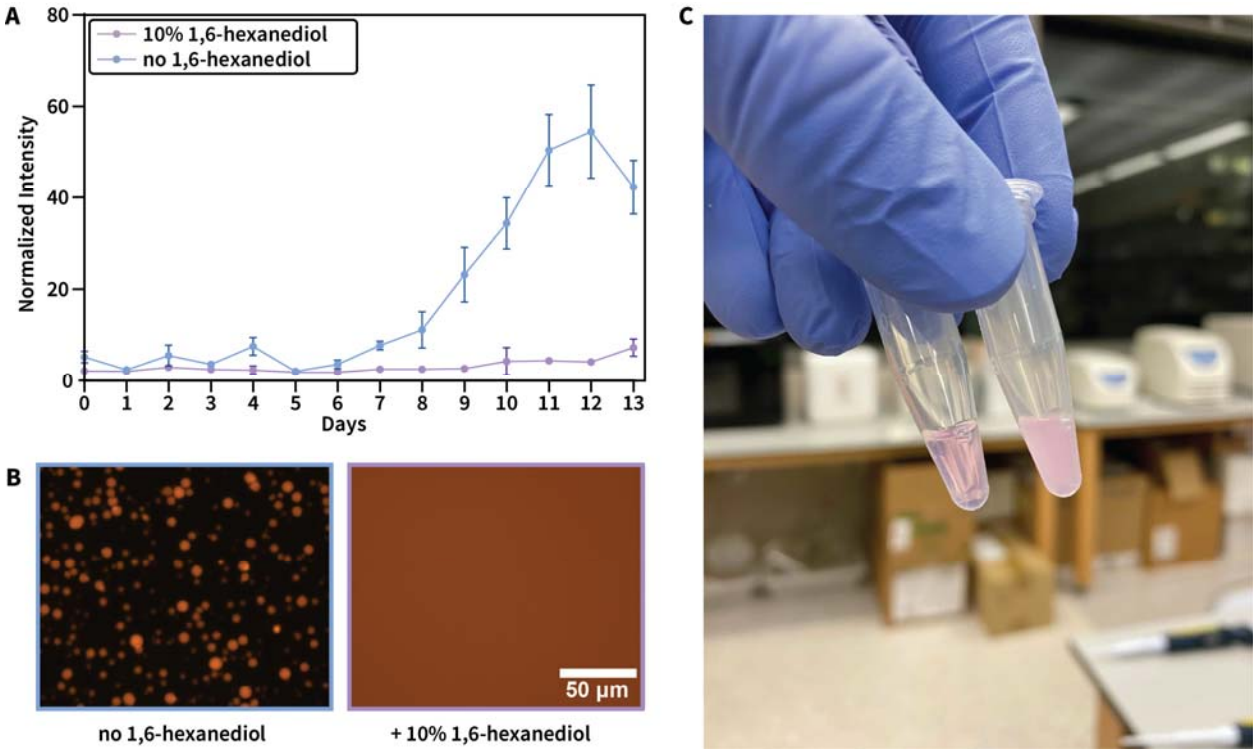


Figure S5. A) Thioflavin T fluorescence assay of FUS LC amyloid formation in the presence and absence of 10% 1,6-hexanediol. Experiments were done in triplicate. **B)** Imaging of Cy3 labeled FUS LC samples in the presence and absence of 10% 1,6-hexanediol. **C)** Cy3 labeled FUS LC samples with and without 10% 1,6-hexanediol. Inhibition of LLPS is apparent in the hexanediol sample (left).

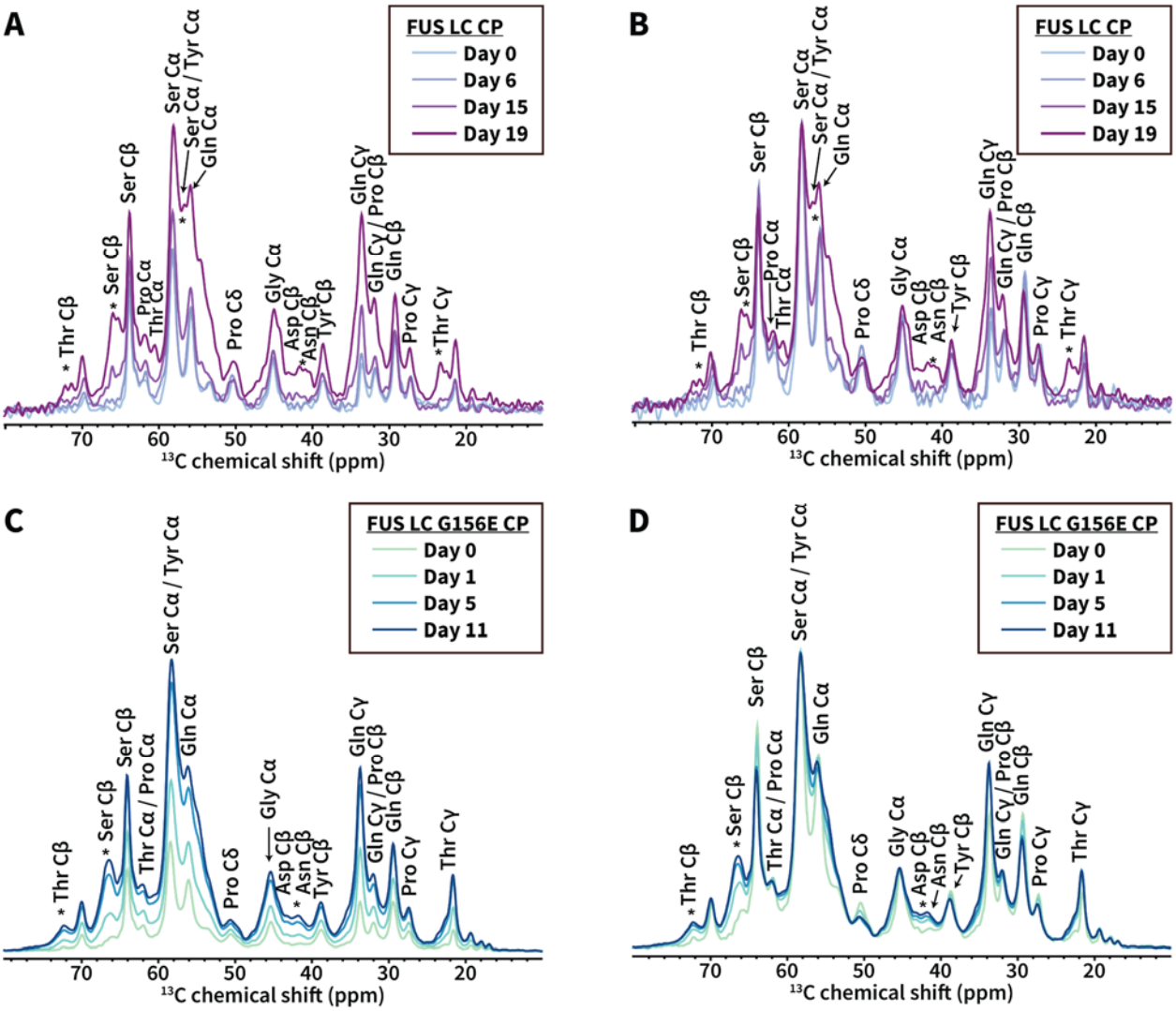


Figure S6. A) Representative CP spectra of the FUS LC sample acquired on different days along the transformation process. **B)** The same spectra as in (A) but normalized to the highest intensity peak of the day 19 spectrum to highlight changes in chemical shifts throughout the transformation process**. C)** Representative CP spectra of the FUS LC G165E sample acquired at different days along the transformation process. **D)** The same spectra as in (C) but normalized to the highest intensity peak of the day 11 spectrum to highlight changes in chemical shifts throughout the transformation process. The asterisks highlight peaks where significant changes are observed. Note that changes in the 40 – 45 ppm region may arise from Gly Cα, Asp Cβ, Asn Cβ or Tyr Cβ resonances.

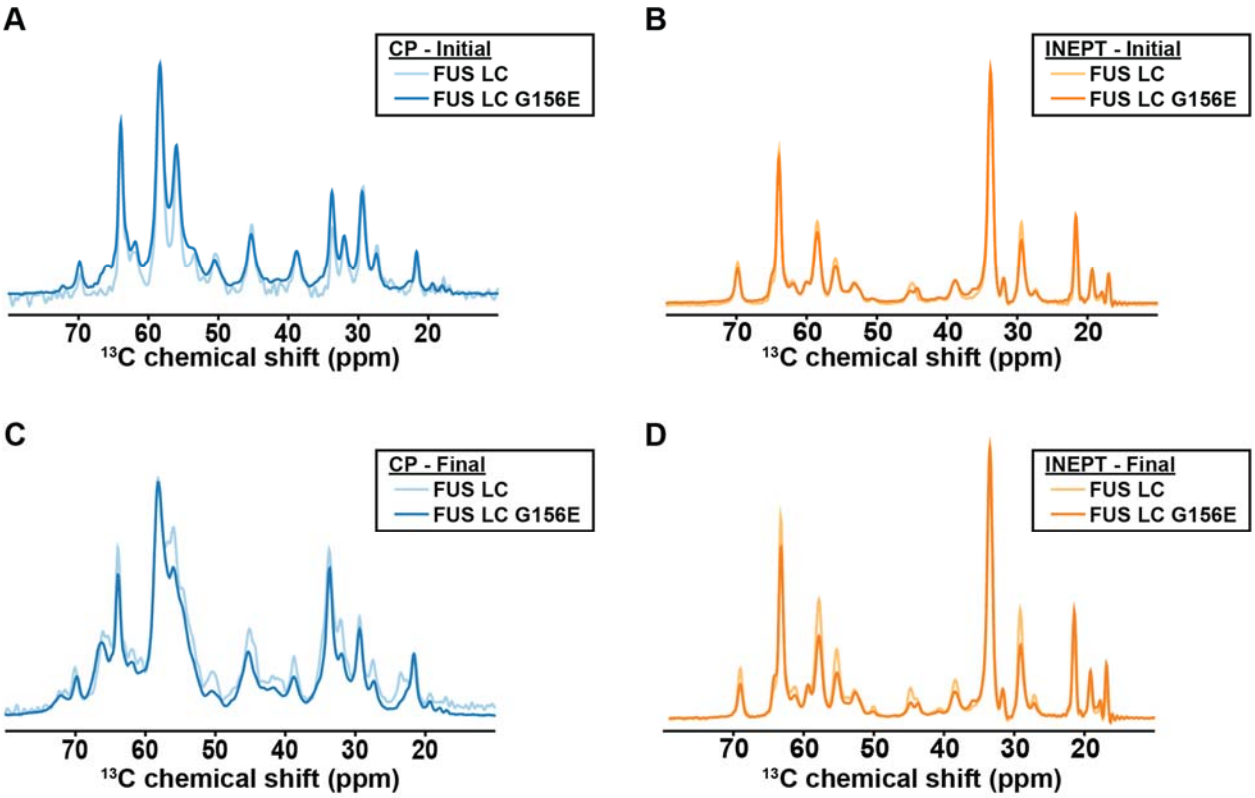


Figure S7. Comparison of the initial and final states of the FUS LC and the FUS LC G156E samples. **A)** CP spectra acquired on day 0, **B)** INEPT spectra acquired on day 0, **C)** CP spectra acquired on day 19 for FUS LC and day 11 for FUS LC G156E, **D)** INEPT spectra acquired on day 19 for FUS LC and day 11 for FUS LC G156E. Each set of spectra is normalized to the highest intensity peak. The FUS LC data were acquired with 128 scans, while the FUS G156E spectra were acquired with 1600 scans (see also **Table S1**).

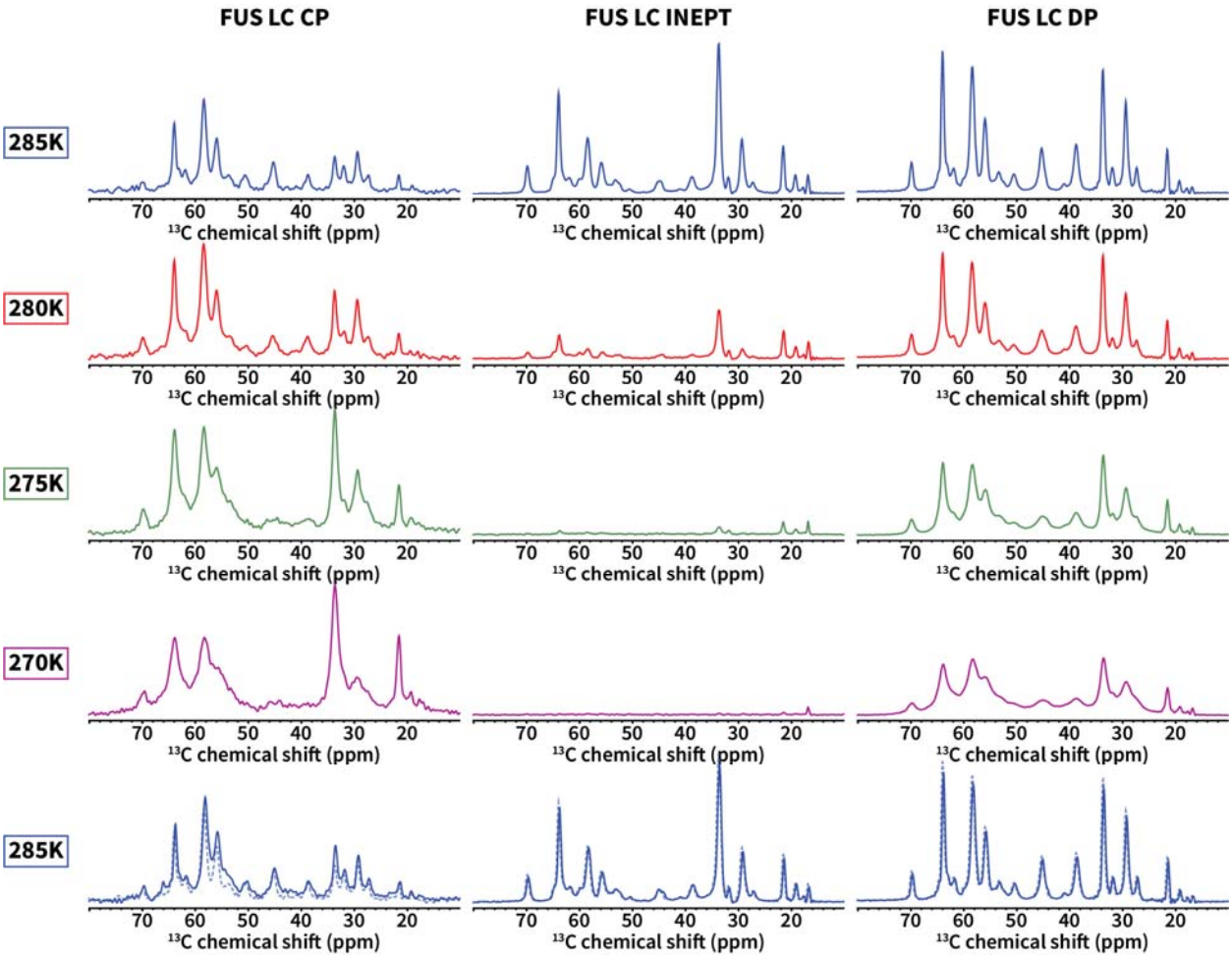


Figure S8. Temperature dependence of the FUS LC CP, INEPT and DP MAS NMR spectra. The reported temperature values reflect the temperature of the cooling gas, while we estimate that the actual sample temperature is 10 - 15 °C higher. The bottom spectra compare data obtained at 285 K before (dashed line, same as top spectrum) and after (solid line) cooling the sample.

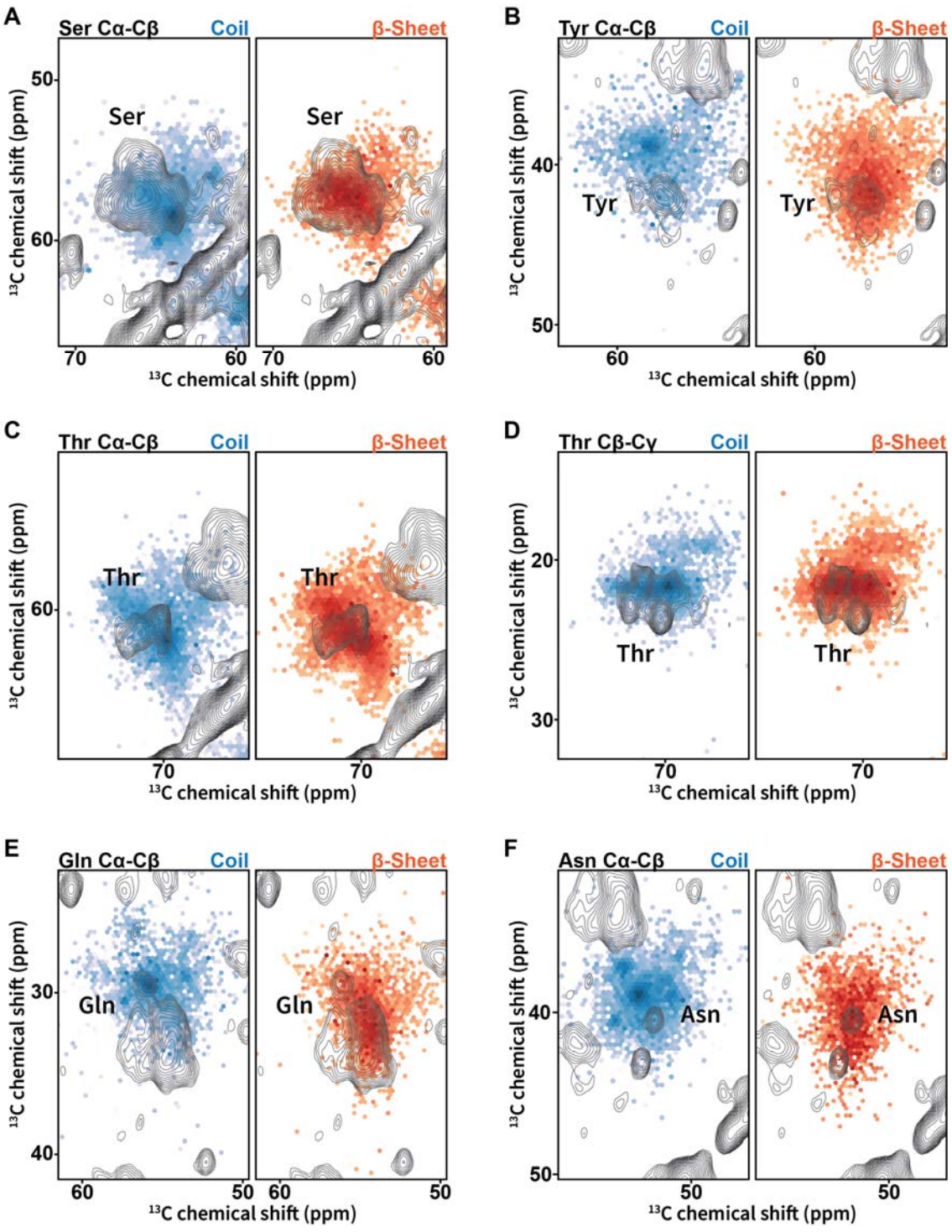


Figure S9. Analysis of the β-sheet content of the 2D DARR correlation spectra of FUS LC based on chemical shift statistics from the BMRB. For each residue type, red densities denote the statistical distribution of β-sheet chemical shifts, while blue densities represent random coil chemical shifts. Although there is some overlap between β-sheet and coil densities, the correlations observed in our DARR experiment fall well within the β-sheet statistical distributions, suggesting that the final state of FUS LC contains β-sheet secondary structure.

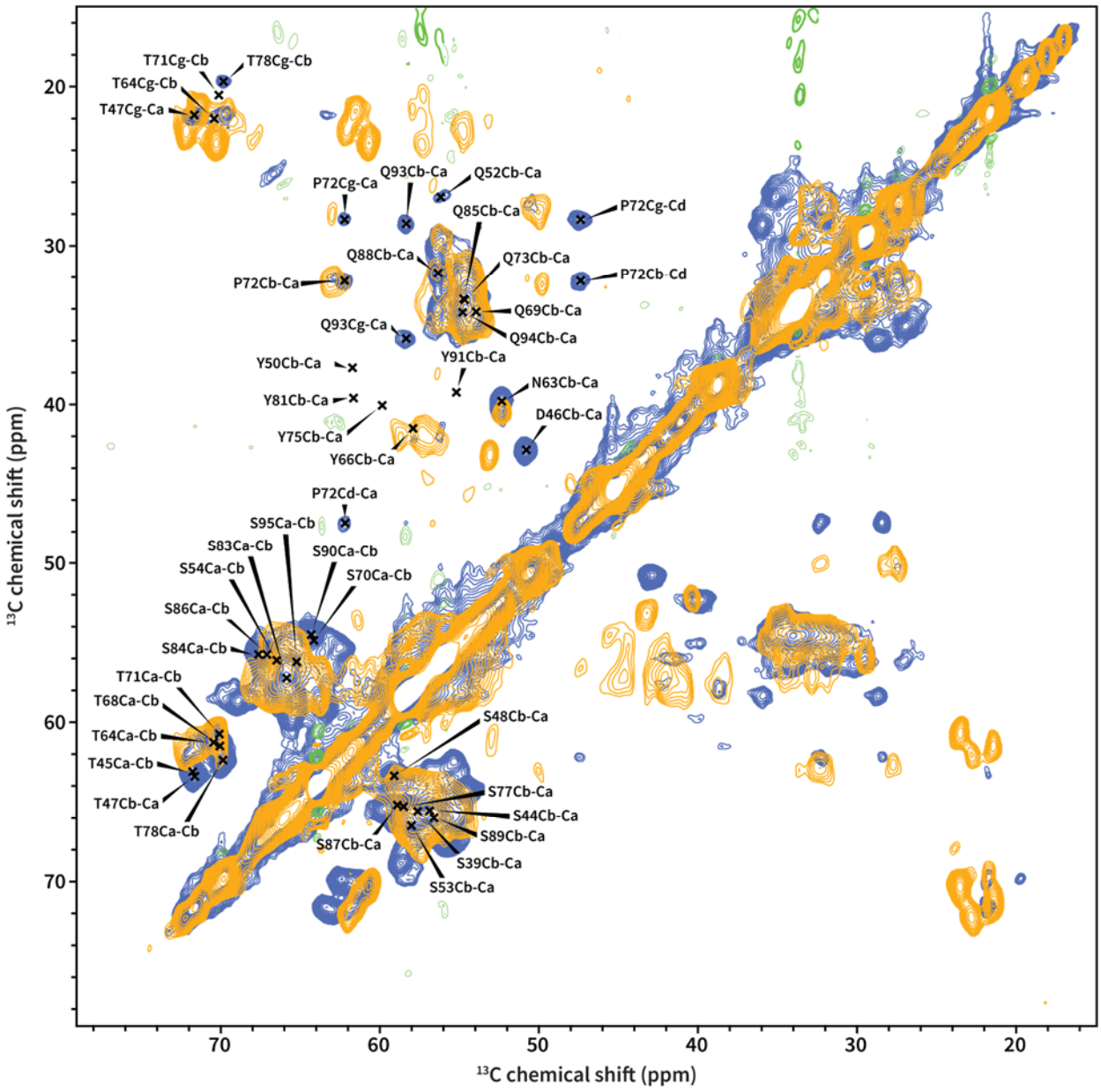


Figure S10. Comparison of the DARR spectra of the FUS LC(1-163) sample used in this work (orange) and the DARR spectra of FUS LC(1-214) from Murray *et al*. (blue)[1].The assignments correspond to the FUS LC(1-214) sample as determined by Murray *et al*.

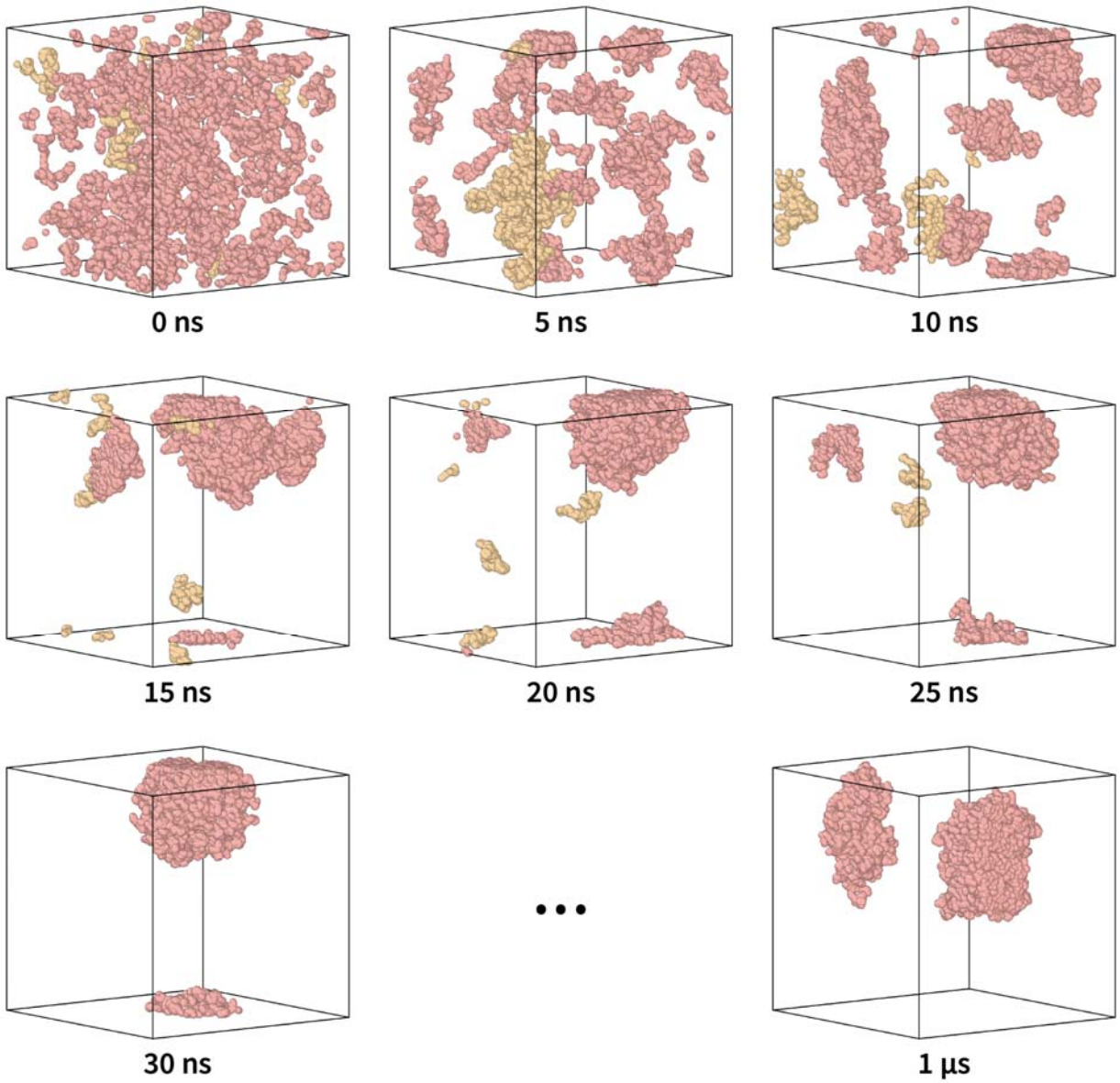


Figure S11. Snapshots of the production coarse-grained molecular dynamics run of FUS LC with FUS LC monomers colored by cluster. Clusters are separated by at least 3.2 nm. Particles representing 100 FUS LC monomers were placed into a 50 nm x 50 nm x 50 nm simulation cell at the beginning of the run. As the simulation progressed, the protein rapidly condensed into a discrete droplet. The droplet integrity was maintained throughout the entire run, with few monomers occasionally breaking free and then rejoining the droplet.

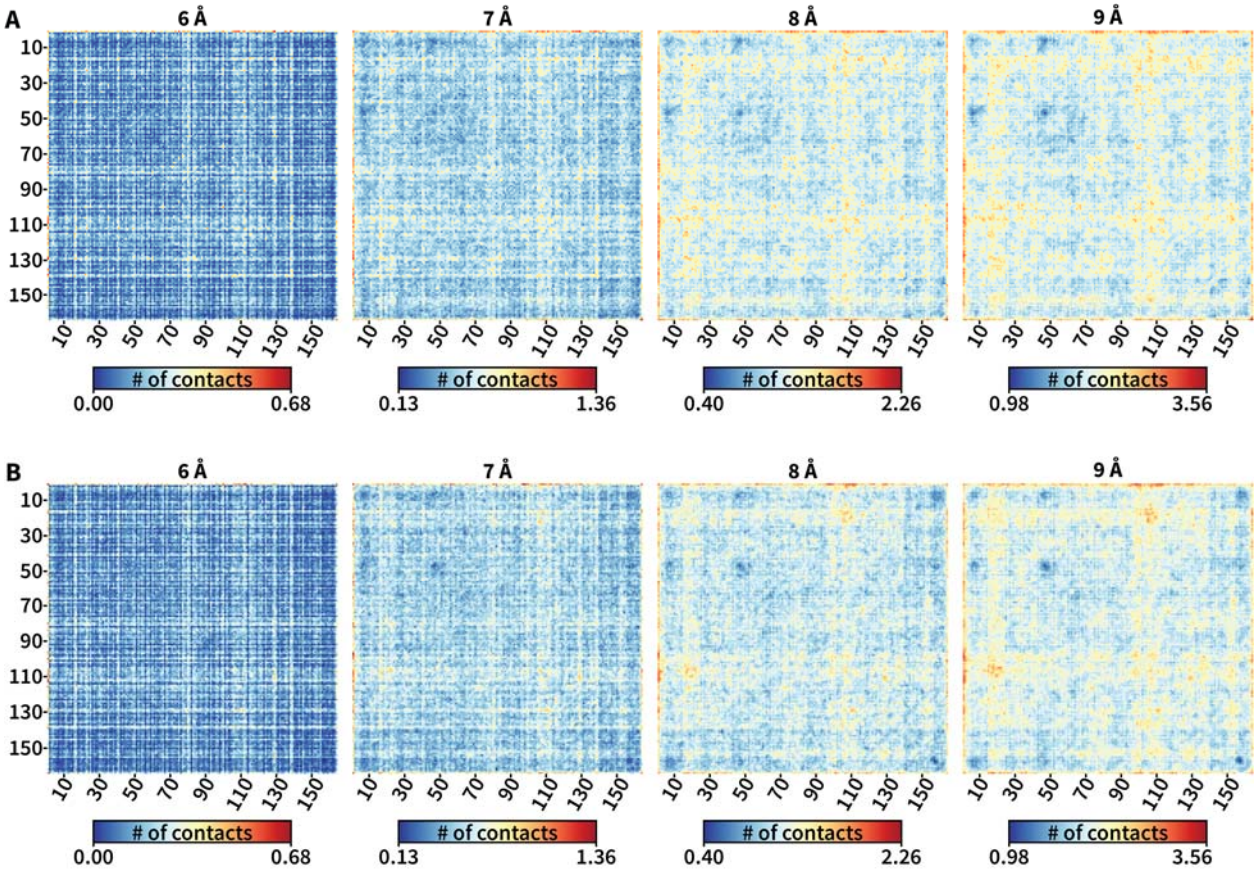


Figure S12. Intermolecular contact maps for FUS LC and FUS LC G156E with varying cutoff radii. Distances are measured from bead centers. The contacts maps for each cutoff radius use the same color coding for the number of contacts. **A)** Maps for FUS LC. **B)** Maps for FUS LC G156E.

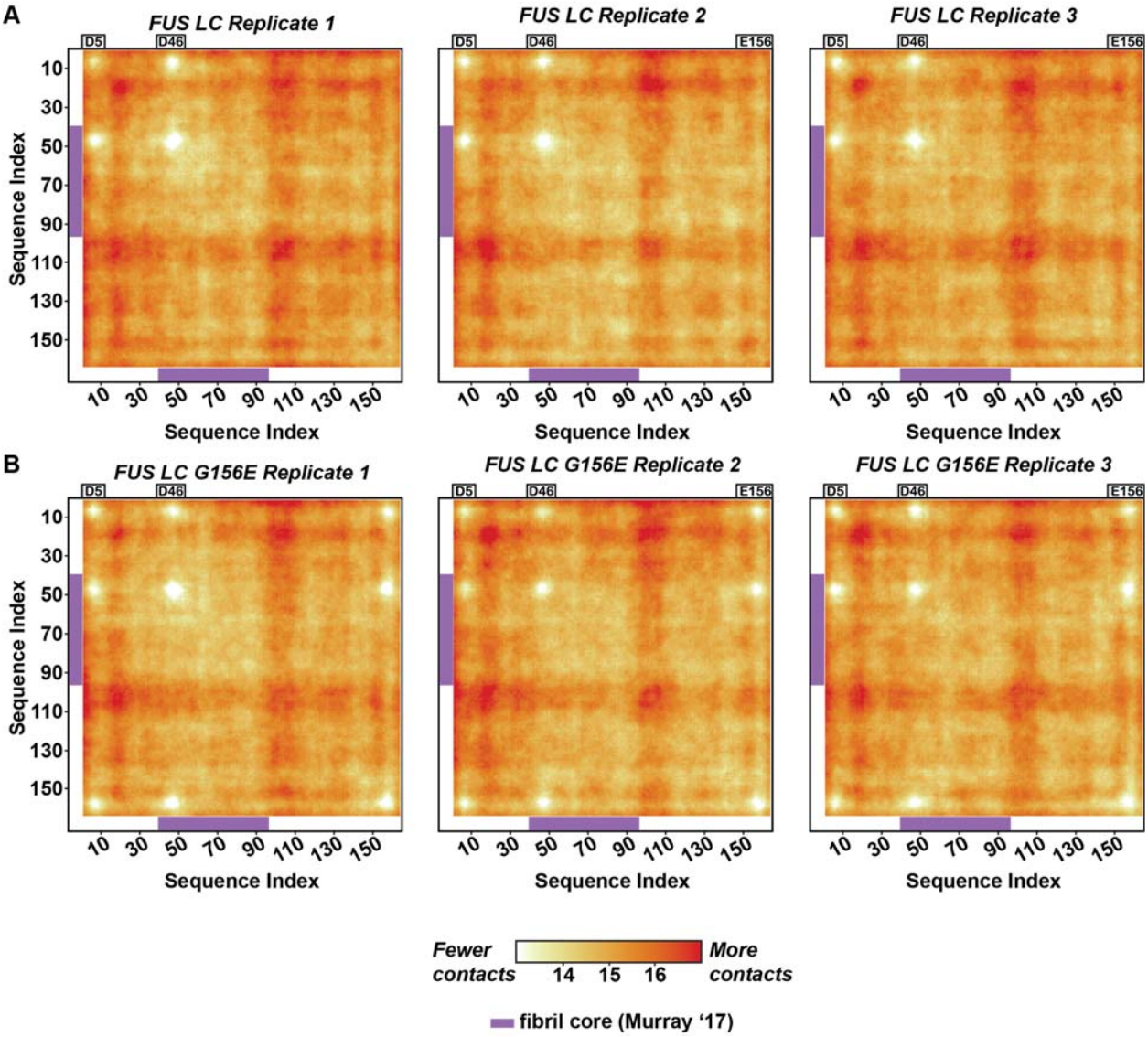


Figure S13. Contact maps generated for three independent replicates of the coarse-grained simulations of **A)** FUS LC, and **B)** FUS LC G156E. Each replicate was run for 300 ns starting from an independent initial condition and each contact map averages data over 300 frames in the droplet state. The contact maps represent contacts within 15 Å and use the same color coding for the number of contacts.

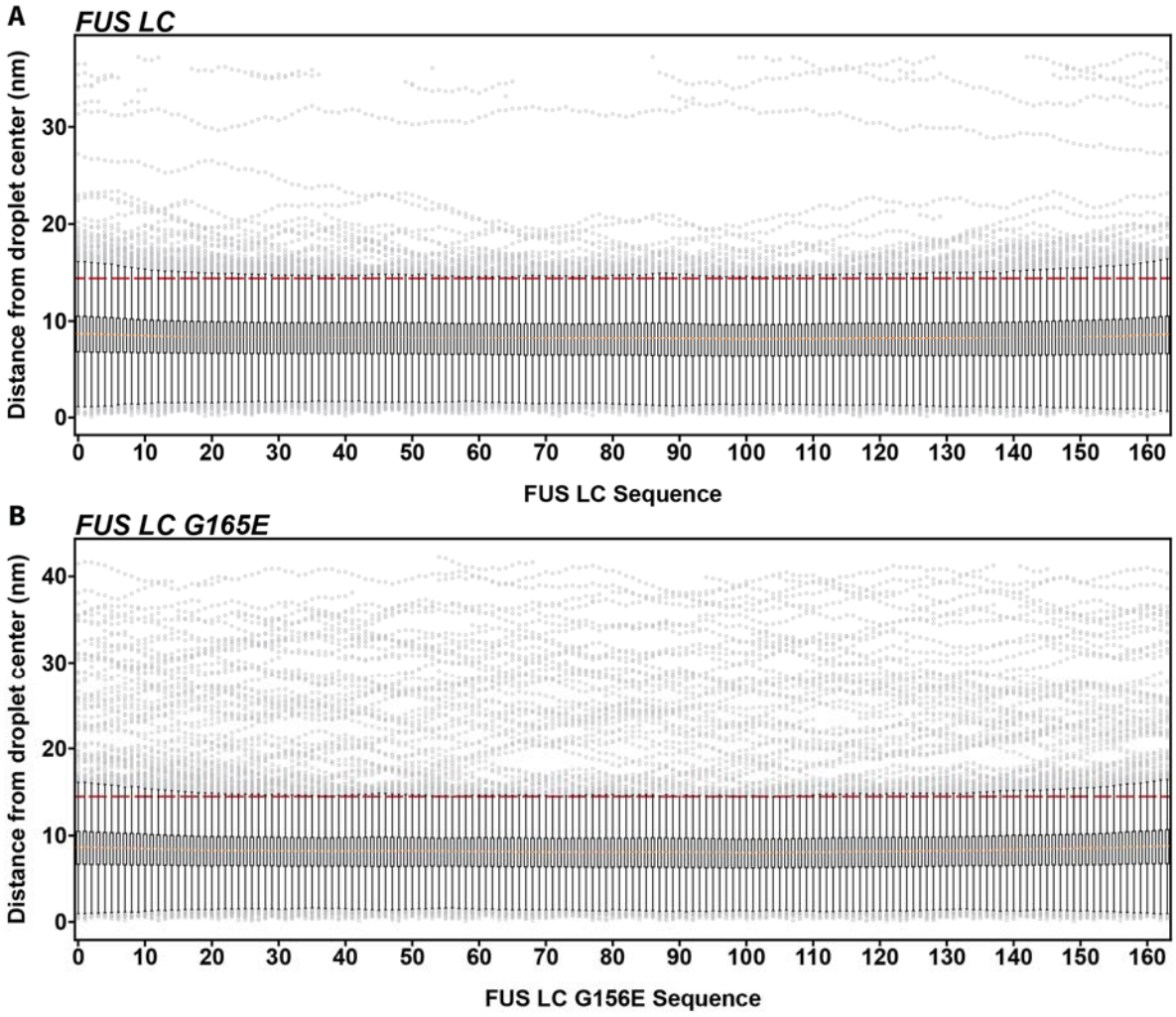


Figure S14. A boxplot depicting the distribution of distances from the center of the LLPS droplet for each residue in **A)** FUS LC, and **B)** FUS LC G156E. The data are based on 100 simulation frames. With 100 monomers in each simulation, the boxplot for each residue represents 100 x 100 = 10,000 observations. Median distances for each residue are presented in orange and the approximate droplet radius is represented by a red dashed line. The median distance from the droplet center is 8.0 – 8.7 nm for each residue and the medians appear similar within each protein and between the wild-type and the mutant protein. Since the droplets contain 100 monomers and are small (the average droplet radius is ~ 14 nm), the surface to volume ratio is large. However, based on these data we can conclude that there are no significant differences between the two samples regarding how often a residue is on the surface or in the interior of the droplet. Therefore, the observed differences in the contact maps presented in Figure 4 are not due to surface effects. We note that G156E FUS LC monomers appear more likely to break off the droplet as represented by the larger number of outliers in the boxplot for that sample. The presence of these outliers does not have a significant effect the median distance from the droplet center as the number of outliers is still small when compared to the total number of monomers represented in the data.

Table S1. Parameters for NMR experiments. All experiments were performed at MAS frequency of 11.11 kHz, 285 K, 750 MHz ^{1}H Larmor frequency with a 3.2 mm E^{free} HCN probe.

Experiment	**FUS LC**		**FUS LC G156E**	
Sample amount	**20**	**mg**	**20**	**mg**
CP				
Number of scans	128	scans	1600	scans
Recycle delay	4	s	4	s
CP contact time	1.75	ms	1.75	ms
Decoupling	SPINAL64		SPINAL64	
Decoupling power	90	kHz	90	kHz
Acquisition time	9	ms	9	ms
INEPT				
Number of scans	128	scans	1600	scans
Recycle delay	4	s	4	s
^{13}C π/2	5	µs	5	µs
^{1}H π/2	2.75	µs	2.75	µs
Refocusing parameter	140	Hz	140	Hz
Decoupling	SPINAL64		SPINAL64	
Decoupling power	90	kHz	90	kHz
Acquisition time	9	ms	9	ms
DP				
Number of scans	128	scans	1600	scans
Recycle delay	4	s	4	s
^{13}C π/2	5	µs	5	µs
Decoupling	SPINAL64		SPINAL64	
Decoupling power	90	kHz	90	kHz
Acquisition time	9	ms	9	ms
2D-INEPT				
Number of scans	32	scans	32	scans
Recycle delay	3	s	3	s
^{13}C π/2	5	µs	5	µs
^{1}H π/2	2.75	µs	2.75	µs
Refocusing parameter	140	Hz	140	Hz
Decoupling	swftppm		swftppm	
Decoupling power	90	kHz	90	kHz
t1 acquisition	33.3	ms	33.3	ms
t1 points	500		500	
CP-DARR				
Number of scans	160	scans	64	scans
Recycle delay	4	s	4	s
CP contact time	1.75	ms	1.75	ms
Decoupling	SPINAL64		SPINAL64	
Decoupling power	90	kHz	90	kHz
^{1}H power during DARR	11.1	kHz	11.1	kHz
DARR mixing	20	ms	20	ms
t1 acquisition	3.2	ms	3.2	ms
t1 points	256		256	

Supporting References

1. Murray, D. T. *et al.* Structure of FUS Protein Fibrils and Its Relevance to Self-Assembly and Phase Separation of Low-Complexity Domains. *Cell* **171**, 615-627.e16 (2017).

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

ACKNOWLEDGEMENTS

Chapter 2, in full, is a reprint of the material as it appears in *Biophysical Journal*: Berkeley, R. F., Kashefi, M. & Debelouchina, G. T. Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC. *Biophys J* **120**, 1276–1287 (2021). The dissertation author was the primary investigator and author of this material.

Chapter 3: The small heat shock protein HSPB1 forms cage-like oligomers in the absence of a client

Introduction

Protein folding, like many processes in biology, is error-prone[1,2]. Proteins are flexible and dynamic molecules, and their structural plasticity is often the result of marginal thermodynamic stability that can lead to misfolding. Misfolded proteins are generally considered to be undesirable and in many cases are implicated in aging and disease[2-5]. In mammalian cells, thousands of proteins are involved in preventing or resolving protein misfolding and maintaining homeostasis in the folded protein population, an equilibrium referred to as proteostasis[6,7]. Since all proteins are capable of misfolding, the mechanisms that facilitate proteostasis must be universal enough to prevent the pathological misfolding of all proteins. This one-to-many relationship is compelling from a structural perspective–how can a protein-protein interaction be general enough that it can address all of the possible routes of protein misfolding that are present in the proteome?

One of the central classes of proteins involved in proteostasis are molecular chaperones, which engage misfolded or partially folded proteins[8]. Molecular chaperones generally act either during translation as trans-ribosomal facilitators of protein folding (providing a potential solution to Levinthal's Paradox[9]) or as modulators of a terminal unfolded or misfolded state where they bind to protein aggregates and fibrils[10,11]. The former category is mostly populated with ATP-dependent chaperone complexes that bind to exposed hydrophobic residues in nascent

polypeptides and transiently block aggregation in these regions, facilitating protein folding through kinetic partitioning[11]. Notable members of this class of molecular chaperones in mammals are HSP10/HSP60[12], HSP70[13], and HSP90[14,15]. The latter category of molecular chaperones is more diverse and includes both ATP-dependent active molecular chaperones and ATP-independent molecular chaperones that bind unfolded or misfolded proteins passively without actively facilitating refolding. These passive molecular chaperones are often referred to as "holdases", a functional nomenclature that describes their role as binders that sustain unfolded proteins in a folding-competent state and buffer the more complex functionality of active chaperones such as HSP70 or HSP90. In many cases, protein folding or protein refolding involves interplay between ATP-dependent active molecular chaperones and ATP-independent holdases.

Small heat shock proteins (sHsps) are a superfamily of holdases that are some of the most diverse and widespread molecular chaperones[16-20]. They are also some of the best studied– small heat shock proteins were identified as a distinct class of proteins in the 1980s[21,22], were first recognized to have chaperone activity in the early 1990s[23,24], and have been implicated in a wide range of biological processes and disease pathologies since[25]. All sHsps share the same tripartite domain architecture despite having a poorly conserved primary sequence[26]. The core component is the α-crystallin domain (ACD), a 90 – 100 residue domain that adopts a characteristic β-sandwich fold containing two layers of five and three antiparallel β-sheet strands each. The N-terminal domain (NTD) is of variable length and sequence, and it is generally considered to be partially or totally disordered yet largely immobile. The evidence for this unusual conformational heterogeneity without dynamic motion is chiefly provided by difficulties resolving the NTD with

most structural biology methods, including solution NMR spectroscopy[27], cryo-EM[28-30], and crystallography[31,32]. Some studies have resolved small regions of consistent secondary structure within the NTD using solid-state NMR[33] or by capturing sHsps complexed with short client peptides[34]. The C-terminal domain (CTD) is generally shorter than the NTD and ACD and is disordered and dynamic enough to be observed by solution NMR. All three domains play important roles in sHsp function.

Although the domain architecture of sHsps is simple, sHsps have been refractory to structural characterization due to their formation of intricate polydisperse and heterogeneous oligomers[28,35,36]. In addition to the different dynamics observed in the NTD, ACD, and CTD of individual sHsp monomers, oligomers themselves are highly dynamic and engage in subunit exchange which can further complicate structural characterization[37,38]. Nevertheless, several structures or structural models for full-length sHsps exist. Crystal structures of multiple monodisperse sHsp oligomers have been determined, consistently revealing cage-like symmetrical or pseudosymmetrical assemblies constructed from a dimer base unit[32,39,40]. Negative stain microscopy and cryo-EM have been used to show that polydisperse sHsps form similar cage-like oligomer structures, with examples from αA-crystallin[28], αB-crystallin[41,42], Hsp16.5[43,44], Hsp21[29], Hsp26[30,45], and Sip1[46]. Electron microscopy methods have the advantage of being able to identify multiple oligomer states within a single sample, which can lead to powerful analysis of both the architecture of oligomers and the relative distribution of each oligomer[28]. Lastly, a solid-state NMR structure of an αB-crystallin 24mer, the result of a tour-de-force of MAS NMR spectroscopy, is one of the most comprehensive and detailed descriptions of the structure of a sHsp oligomer in the literature[33,47]. As expected from functional data, almost all

sHsp oligomers are built from dimers that are formed across two ACD domains, with interactions between the CTD (usually via a conserved IXI motif) and the β4/β8 groove of an adjacent ACD as well as interactions between the NTDs of adjacent dimers contributing to the integrity of the cage-like structure in many cases.

The size distribution and degree of polydispersity exhibited by sHsp oligomers varies, and is tied to the primary sequence of each sHsp[44]. In addition to primary sequence, abiotic factors such as temperature, pH, concentration, oxidation state, and the presence of post-translational modifications can also affect oligomer distribution[20,28,48-52]. For a given sHsp, the oligomer profile is directly linked to the capacity to act as a holdase, but the relationship is inconsistent– both low[46,53,54] and high[46,55] molecular weight oligomers have been implicated in molecular chaperone activity. For example, in the case of αB-crystallin and HSPB1, the two most abundant mammalian sHsps, phosphorylation at three specific residues in the NTD shifts the distribution of oligomer sizes towards smaller oligomers and increases chaperone activity[35,37,56-62]. As a counterpoint, heat shock of HSP18.1 from pea induces a transition from inactive monodisperse dodecamers to a population of larger polydisperse oligomers that bind to clients[54].

It is possible that different oligomer sizes expose different parts of the same sHsp in order to bind different clients. For example, in a recent report by Liu et al, wild-type HSPB1, which forms large oligomers, was shown to engage the intrinsically disordered protein FUS LC in the monomeric state and prevent condensate formation while a phosphomimetic 3D-HSPB1 construct which forms smaller oligomers was shown to allow condensate formation but protect against β-sheet formation of FUS LC[62]. Delbecq and Klevit recently demonstrated that the

chaperone activity of HSPB5 towards αLac at early monomeric, and late, aggregated stages of its aggregation pathway were determined by the ability of HSPB5 to alter its oligomeric distribution[63]. In a similar vein, there is growing evidence that individual domains of sHsps are responsible for engaging different types of clients, with the NTD and ACD domains often being implicated in binding different types of clients, or clients in different states[62,64,65].

To investigate the role of sHsp domains in client chaperoning, we set out to explore the relationship between HSPB1 and a series of intrinsically disordered proteins in multiple states. There is already mounting evidence that HSPB1 is capable of interacting with FUS LC and affecting the phase transitions of FUS LC between the monomeric, liquid-liquid phase separated, and fibrillar state[62]. The status of HSPB1 as the most highly expressed sHsp in the human proteome and the functional evidence of the relationship between HSPB1 and FUS LC phase transitions positions this system as an ideal one for investigating chaperone function through a structural lens. In order to provide context for our investigation of HSPB1-client interactions described in Chapter 4 of this dissertation, we first sought to characterize the structure of HSPB1 and HSPB1 oligomers in the absence of a client. Characterization of the oligomer structure of the client-free state of HSPB1 provides context for the changes in HSPB1 structure and dynamics observed when challenged with client proteins in phase separated, aggregated, and fibrillar states.

Materials and methods

Expression of HSPB1 and related constructs

HSPB1 was prepared from recombinant Rosetta (DE3) competent *Escherichia coli* cells (MilliporeSigma, Burlington, MA) that had been transformed with a plasmid encoding the full-length HSPB1 sequence (**Fig. 3.S1**). The HSPB1 sequence was sourced from a template plasmid encoding the 6xHis-TEV-HSPB1 construct, a gift from Don Cleveland, from which the 6xHis-TEV sequenced had been excised using an NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, Ipswitch, MA) and custom primers (Integrated DNA Technologies, Inc., Coralville, IA). Seed cultures were grown to saturation from freshly transformed colonies and inoculated at a 1% v/v ratio into Luria-Bertani medium supplemented with kanamycin (50 μg/mL). Cultures were grown in LB at 37°C to an OD600 of ~ 0.6, at which time the temperature was reduced to 18 °C and protein expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside. Cultures were allowed to express protein for 16 hours at 18 °C before being harvested by centrifugation at 5,000 RCF and 4°C for 30 minutes. After decanting the supernatant, cell pellets were either immediately used for purification of HSPB1 or stored at -80°C for later use.

Purification of HSPB1 and related constructs

We found that TEV cleavage of the His-TEV-HSPB1 was unreliable, with some preparations containing small amounts of uncleaved His-TEV-HSPB1 in addition to the tag-free HSPB1

cleavage product (possibly due to steric occlusion of the N-terminal TEV cleavage sequence in HSPB1 oligomers). To circumvent this, we turned to a tagless ammonium sulfate-mediated purification inspired by approaches previously described in the literature[47,66-68].

Briefly, the HSPB1 cell pellet was thawed if necessary, resuspended in lysis buffer (50 mM sodium phosphate, 50 mM sodium chloride, Roche cOmplete™ EDTA-free Protease Inhibitor Cocktail, pH 7.4) and lysed by pulsed sonication for 30 minutes using a Qsonica sonicator with a 1/8" diameter probe tip at 12 kHz (60%) output and 4°C. The lysate was cleared by centrifugation at 30,000 RCF for 30 minutes and the supernatant was decanted into an Erlenmeyer flask with a stir bar. To this was added saturated ammonium sulfate (4.1 M), dropwise, to a final concentration of 40% saturation (1.64 M) at room temperature. The solution was allowed to stir for 30 minutes, after which the precipitated protein was pelleted by centrifugation at 30,000 RCF for 30 minutes at 4 °C. The supernatant was discarded and the pellet was redissolved in 20 mM tris(hydroxymethyl)aminomethane, 10 mM magnesium chloride, 30 mM ammonium chloride, pH 7.6 at room temperature. The resulting solution was desalted by running over PD-10 desalting columns packed with Sephadex G-25 resin (Cytiva, Marlborough, MA) equilibrated in the same buffer. The desalted HSPB1 solution was further purified by ion exchange chromatography over a Mono Q anion exchange column (Cytiva, Marlborough, MA) equilibrated in the same buffer. Protein was eluted from the column using a 0-250 mM NaCl gradient. The fractions containing HSPB1 were concentrated and subjected to reverse-phase HPLC purification over a Waters XBridge Peptide BEH C18 OBD prep column (130 Å pore size, 10 μm particle size, 19 mm X 250 mm) using a Waters 2535 Binary Gradient Module equipped with a 2484 UV/Vis detector (Waters Corporation, Milford, MA). Sample

purity was confirmed using a Waters XBridge Peptide BEH C18 column (300 Å pore size, 5 µm particle size, 2.1 mm X 100 mm) on the same HPLC system. Sample identity was confirmed by intact mass QTOFMS using an Agilent 1260 Infinity Binary LC coupled to a 6230 Accurate-Mass TOFMS system (Agilent Technologies, Santa Clara, CA). All HPLC purifications utilized a gradient method of water and acetonitrile with 0.1% trifluoroacetic acid as the mobile phases. After HPLC purification, samples were lyophilized and refolded by dialysis at 4 °C from 6 M guanidine hydrochloride, 50 mM sodium phosphate, 50 mM sodium chloride, pH 7.4, into the same buffer without denaturant. If samples were not used immediately after preparation, they were concentrated to ~1 mM and flash frozen at -80 °C for storage.

Analytical Ultracentrifugation

The analytical ultracentrifugation sedimentation velocity experiment was performed using a ProteomeLab XL-I (Beckman Coulter, Brea, CA) analytical ultracentrifuge, with absorbance detection at 280 nm. Samples of HSPB1 were prepared in 50 mM sodium phosphate, 50 mM sodium chloride, +/- 2 mM tris(2-carboxyethyl)phosphine, pH 7.4 at 0.3 mg/ml (A280 = 0.5) and loaded into 2-channel cells equipped with sapphire windows and spun using an An-50 Ti 8-place rotor at 40,000 rpm and 20°C for 20 hours. Data were analyzed using Sedfit[69].

Cryo-electron microscopy sample preparation and data collection

Samples of HPSB1 were prepared in 50 mM sodium phosphate, 50 mM sodium chloride, 5 mM tris(2-carboxyethyl)phosphine, pH 7.4 at concentrations ranging from 0.5 mg/ml to 4 mg/ml.

Samples were allowed to equilibrate at room temperature for at least 1 hour before application to the grid. Freshly glow-discharged Quantifoil R 2/1 holey carbon-on-copper grids (Quantifoil Micro Tools GmbH, Jena, Germany) were loaded into a Vitrobot Mark II System (Thermo Fisher Scientific, Waltham, MA) and 4 µl of HSPB1 sample were applied at 4°C and 100% relative humidity to prevent sample evaporation. The blotting parameters were set to blot for 4 seconds with a blot force of -10. After blotting, the grid was plunge-frozen into liquid ethane. Following plunge freezing, the grids were clipped and stored under liquid nitrogen until data collection.

For data collection, grids were loaded into an autoloader cassette and transferred into a Titan Krios G3 transmission electron microscope equipped with a K2 Summit direct electron detector with a Bioquantum energy filter. Movies were acquired at a 300 kV accelerating voltage with a calibrated pixel size of 1.1 Å/pixel. The total electron dose was 64.5 *e*-/Å^2, fractionated over multiple frames to allow for dose-weighted motion correction during data processing.

Cryo-electron microscopy image processing

All cryo-EM data were processed using CryoSPARC v4.2[70]. Image preprocessing, including patch motion correction, contrast transfer function estimation, blob picking, and initial 2D classification was performed during data acquisition with CryoSPARC Live.

To prepare volumes for particle sorting via heterogeneous refinement, a dataset of ~3.48 M particles was picked from 6,204 curated movies using an elliptical blob picker with a 100

ångström minimum radius and 300 ångström maximum radius. Particles were extracted in 360 pixel boxes and Fourier cropped to 90 pixels. The resulting particle stack was pared down to ~1.34 M particles based on normalized cross correlation (NCC) score and power score cutoffs, and further curated to ~377 k particles via 2D classification. The final 2D classes were used to generate templates from which a subsequent round of template picking yielded ~4.09 M particles. This particle stack was pared to ~499 k particles using NCC score and power score cutoffs and further curated to ~424 k particles via 2D classification. These particles were used to train a Topaz model[71] using a learning rate of 3×10^{-4} over 20 epochs (**Fig. 3.S2**). This model picked ~276 k particles. The resulting particle stack was curated to ~239 k particles via 2D classification.

The particle stack provided by the Topaz model was subjected to multi-class ab-initio reconstruction with C1 (no symmetry) and D3 symmetry. The volumes generated by these runs were subsequently subjected to 3D classification. Three dominant classes emerged from the symmetric dataset. These classes, along with unique classes from the asymmetric 3D classification, were used as templates for heterogeneous refinement.

In order to circumvent potential issues with sampling, a fresh set of ~3.48 M particles were extracted from the original set of curated movies. Particles were extracted using a box size of 360 pixels and Fourier cropped to 90 pixels. The resulting particle stack was pared down to ~2.93 M particles based on normalized cross correlation (NCC) score and power score cutoffs and further curated to ~2.83 M particles via 2D classification. The resulting particle stack was subjected to heterogeneous refinement using 4 structures from the asymmetric 3D classification,

4 structures from the D3 symmetric 3D classification, and a T20S volume acting as a junk class[72]. All classes resolved to between 10 and 11 ångström at a Fourier shell correlation cutoff of 0.143[73].

Results

HSPB1 forms dodecameric oligomers after low-temperature refolding

It is well known that the oligomer state and polydispersity of small heat shock proteins are dependent on concentration and temperature[20,28,48-52]. To prepare pure samples of HSPB1, we employed preparative HPLC, which circumvents challenges presented by the formation of noncovalent assemblies of protein by denaturing the protein at high pressures in organic solvent and separating monomeric protein over a reverse-phase C18 column. After refolding at 4 °C, we found that HSPB1 exhibited an FPLC elution and AUC sedimentation profile that is most consistent with a dominant dodecamer species (**Fig. 3.1A, 3.1B**). The wide FPLC and AUC profiles suggest that the oligomer distribution is still polydisperse. This oligomer distribution was invariant to thermal equilibration (**Fig. 3.1A**). The oligomer distribution was also unaffected by reduction with 2 mM TCEP (**Fig. 3.1B**), which is consistent with previous reports that show that reduction of HSPB1 enhances chaperone activity through the release of monomers rather than modulation of the oligomer distribution[53]. Indeed, transmission electron microscopy data with HSPB1 were more consistent with previous reports for reduced αA-crystallin, which predominantly forms 12-, 16-, and 20mers that appear as largely uniform particles by negative stain TEM and cryo-EM (**Fig. 3.1C, 3.1D**). In contrast to the primarily dodecameric species

obtained via HPLC purification, protein purified by size-exclusion chromatography alone exhibited more

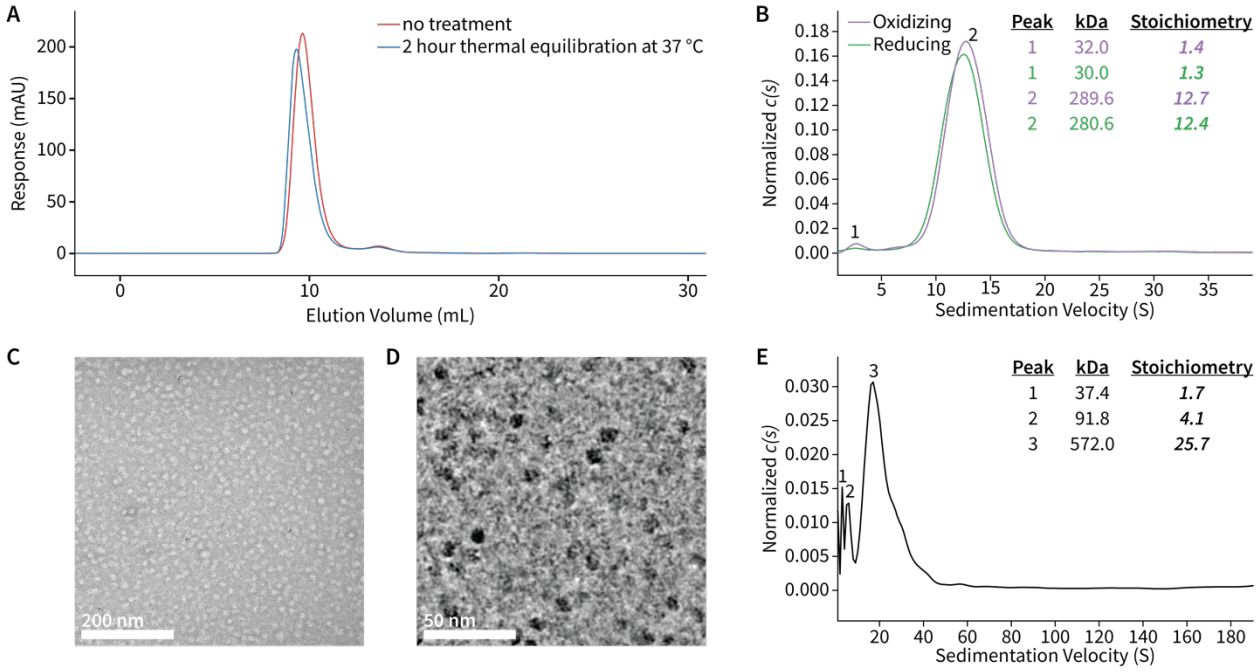


Figure 3.1: Characterization of HSPB1 oligomer size and polydispersity. A) FPLC elution profiles of HPLC-purified and refolded HSPB1 with and without thermal equilibration at 37 °C. **B)** Analytical ultracentrifugation profiles of HPLC-purified and refolded HSPB1 with and without TCEP. **C)** A negative stain micrograph of HPLC-purified and refolded HSPB1. **D)** A cryo-electron micrograph of HPLC-purified and refolded HSPB1. **E)** Analytical ultracentrifugation profile of FPLC-purified HSPB1.

pronounced polydispersity and a trend towards much larger oligomers, with the elution profile centered around the 24mer (**Fig. 3.1E**).

The architecture of HSPB1 dodecamers is heterogeneous with at least two dominant cage-like subtypes

In order to determine the 3-dimensional architecture of HSPB1 oligomers, we turned to cryo-electron microscopy (cryo-EM). The dynamic nature and architectural heterogeneity of HSPB1 oligomers pose challenges for particle classification. Initial attempts at 2D classification of

particles yielded many classes that appeared to represent symmetric or pseudo-symmetric structures with partially blurry densities, likely representative of partial unfolding, domain swapping, monomer or dimer exchange, or interconversion between oligomer states. Despite the obvious heterogeneity in the 2D classes, many of the most highly represented classes look cage-like and appear to have one or more symmetry axes (**Fig. 3.2A**). Many of the classes are remarkably similar to those observed for αA-crystallin[28].

Clear indications of cage-like oligomer formation by HSPB1 can be seen in both raw micrographs and in 2D classes (**Fig. 3.1C, 3.1D, 3.2A**). Despite this, attempts to generate 3D representations of these cage-like structures using symmetry-free ab-initio reconstruction approaches invariably lead to volumes that are poorly resolved on one side (**Fig. 3.2B**). While it is possible that symmetry inherent in HSPB1 oligomers is causing asymmetric assignment of particles onto the volume during the ab-initio reconstruction process, it is more likely that actual heterogeneity in the oligomer integrity, dispersity, and architecture is leading to classification problems that confound the ab-initio reconstruction algorithm. Attempts to resolve this heterogeneity at the sample level by crosslinking with BS3 were unsuccessful, as were attempts to use a detergent cocktail to improve oligomer homogeneity or view distribution. Micrographs from grids prepared using manual plunge techniques[74], a Vitrobot, or an SPT Labtech chameleon[75] all produced similar 2D classes and ab-initio reconstructions.

The heterogeneity observed in the data presents a significant classification challenge. In the literature, this challenge has been addressed for similar sHsps by providing synthetic oligomer models as templates for refinement[28]. In an attempt to reduce the amount of bias in our

classification and refinement approach, we instead carried out 3D classification using template volumes derived from multi-class ab-initio reconstructions with and without imposed D3 symmetry under the expectation that many of the particles in the dodecamer state would adopt this symmetry. This approach generated a number of volumes from which three dominant symmetrical structural classes emerged. The first is a trefoil-like structure, which looks similar to the structure of the Hsp21 dodecamer from *Arabidopsis thaliana* reported by Yu, et al. (EMD-30261)[29]. The second is an elongated barrel-like structure that looks similar to the αA-crystallin dodecamer reported by Kaiser, et al. (EMD-4895)[28]. The third is a flattened barrel-like structure with distinct densities for the ACD domains in HSPB1 and a hollow center (**Fig. 3.2C**). These structures can be fit with six copies of the structure predicted for the HSPB1 dimer by AlphaFold-multimer (**Fig. 3.S3B**) in a pseudosymmetrical manner (**Fig. 3.2C**). Model fitting and literature models suggest that the elongated and barrel-like volumes represent "stretched" and "flattened" versions of the same HSPB1 architecture. In all cases, it is likely that oligomer

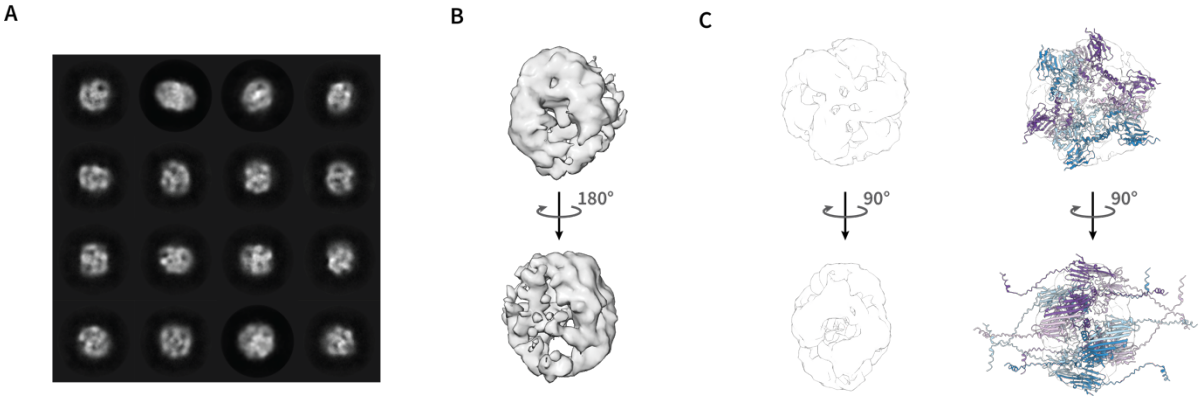


Figure 3.2: HSPB1 oligomers by cryo-EM. A) 2D classes of HSPB1 show structures consistent with pseudosymmetrical cage-like oligomers. **B)** Ab-initio volumes without enforced symmetry are only partially resolved. **C)** Forcing D3 symmetry generates volumes that can be fit with predicted structures for HSPB1 dimers in reasonable oligomer geometries. Note that the CTDs in the model remain in the disordered configuration predicted by AlphaFold-multimer.

integrity is maintained through NTD-NTD and ACD-CTD interactions. For the barrel and trefoil-like volumes, NTDs are oriented towards the center of the oligomer. Scattered densities are present inside the center of the oligomer. For the elongated volume, NTD-NTD interactions occur along the edge of the barrel. In all three cases, insufficient space exists within the oligomer to allow for a significant degree of NTD dynamics. While our cryo-EM data does not resolve the CTD, the orientation of HSPB1 dimers in all oligomer models indicates the possibility of dynamic interactions occurring between the ACD and CTD–an observation that is consistent with the NMR results presented in Chapter 4, and with the HSPB1 literature[27,51,76].

Heterogeneity in oligomer architecture and dispersity preclude the solution of the structure of an HSPB1 oligomer to high resolution

Having obtained plausible structural models for HSPB1 oligomers, we conducted a heterogeneous refinement to sort particles into sets representing each oligomer architecture. In general, refinement reduced the resolution of features in volumes derived from ab-initio reconstruction or 3D classification (**Fig. 3.3**). Heterogeneous refinement can help quantify the distribution of particles in each oligomeric state. For HSPB1, the majority of the particles in symmetrical classes were assigned to the flattened (**Fig. 3.3C**) or elongated (**Fig. 3.3B**) barrel-like architectures, and the minority fell into the trefoil architecture (**Fig. 3.3A**). However, in a heterogeneous refinement with 4 D3-derived classes, 4 asymmetric classes, and one T20S junk class[72], 26% of the particles were assigned to the T20S class, and the remaining particles were divided almost evenly across the symmetric and asymmetric classes, suggesting that the refinement algorithm had difficulty assigning particles to the correct volume. It is therefore likely

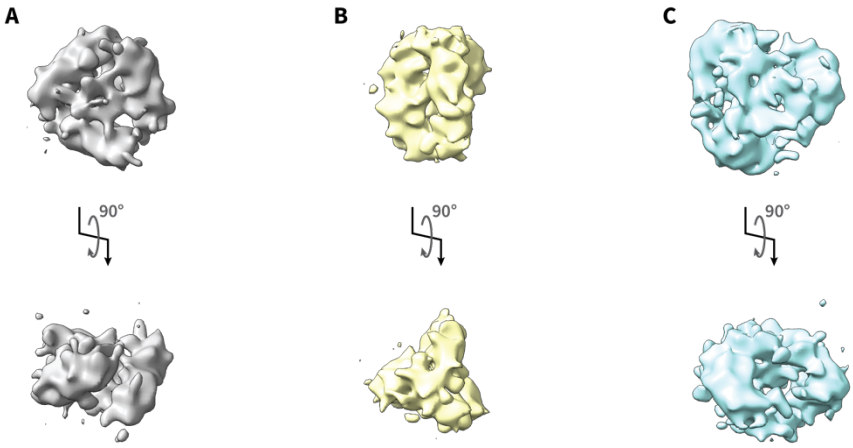


Figure 3.3: Symmetric HSPB1 oligomer classes after heterogeneous refinement with C1 volumes and a T20S junk volume. A) The trefoil oligomer class. **B)** The elongated oligomer class. **C)** The barrel oligomer class.

that classification challenges are preventing the refinement of HSPB1 oligomers, and that further particle curation is required to push characterization of HSPB1 oligomers into a high-resolution regime.

A number of avenues are open for further curation of the HSPB1 particle stack. First, 2D classes of HSPB1 oligomers suggest that oligomers with D4 and D5 symmetry exist in the data in addition to the D3 symmetrical structures that have been the focus of the HSPB1 data analysis so far. Providing D4- or D5-symmetrical templates may provide more appropriate targets for particle sorting during heterogeneous refinement. This can be achieved by either performing ab-initio reconstructions with enforced symmetry, or by generating synthetic volumes in the mold of Kaiser, et al[28]. Sorted particles could be resubmitted to the standard cryo-EM data processing pipeline for an unbiased analysis of each symmetry class.

A second possibility is that the trefoil state lies on a trajectory of continuous stretching motion apparently occupied by the flattened and elongated barrel-like state identified in the current

dataset. If a continuous transition between the three dominant D3 oligomer classes is represented in the particle stack, this heterogeneity would hinder the performance of the standard cryo-EM data processing approach described in this chapter despite the presence of a consistent, albiet dynamic, structure within the data. Methods for capturing continuous motion within cryo-EM data that leverage variational autoencoders trained on cryo-EM data are under active development[77-80]. Although capturing global motion like that exhibited in the stretching of a cage-like HSPB1 oligomer is currently not possible, future advances in approaches for capturing continuous motion in cryo-EM data may facilitate the analysis of this dataset.

Finally, although a large amount of data has been collected so far, comprehensive sampling of all oligomer views could still be an issue if many different oligomer states and many different oligomer architectures for each state are present. The sampling problem is exacerbated by the fact that it is clear in the raw micrographs that many of the HSPB1 oligomers are fragmented, missing dimers, or aggregated into much larger high-order oligomer structures. This is unsurprising given that the molecular mechanism of HSPB1 chaperoning likely involves disruption of the oligomer, and the presence of these partially formed or high-order oligomers may be unavoidable even in a pure sample of HSPB1. It is possible that the acquisition of more data or the merging of the multiple existing HSPB1 datasets could be necessary to provide enough data to generate structures for all of the oligomer states present in an HSPB1 sample.

Conclusion

HSPB1 and other small heat shock proteins are challenging subjects of structural study due to their propensity for forming heterogeneous, polydisperse oligomers. The central role of HSPB1 as a modulator of pathological protein phase transitions and protein misfolding makes it an important structural biology target nonetheless. In the past, the NTD and ACD of HSPB1 have both been implicated in its ability to act as a molecular chaperone. These regions of the protein, along with the CTD, are also important for maintaining the integrity of oligomers of the inactive, client free state. In this work, we set out to understand the high-order architecture of HSPB1 oligomers in order to build models for oligomer structure that will contextualize studies of HSPB1 in the client-bound state. We find that the 12mer is the predominant oligomer state of HSPB1 after low-temperature refolding, and that there are likely multiple 12mer architectures present in the HSPB1 oligomer population. Despite the heterogeneity in the structure of HSPB1 oligomers, all of the models derived from our cryo-EM data analysis suggest that NTD-NTD and ACD-CTD interactions are responsible for maintaining the integrity of HSPB1 oligomers. Additionally, the orientation of the NTD in our models implies that there is insufficient space to allow for movement of the NTD. These results are consistent with our observations regarding the structure and dynamics of the NTD and ACD in the next chapter, and provide insight into the changes that may be occurring the in structure of HSPB1 as it engages client proteins.

In the next chapter of this dissertation, we investigate the dynamics of HSPB1's NTD, ACD, and CTD domains in the presence of client proteins in different states of self-assembly using magic-angle spinning solid-state NMR. Integrating the architectural information regarding HSPB1 oligomers in the absence of client with dynamics information about HSPB1's domains in the

presence of both liquid-liquid phase separated and fibrillar client proteins yields key information into the structure-function relationship of HSPB1's tripartite domain architecture.

References

1 Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884-890 (2003).
2 Chiti, F. & Dobson, C. M. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annual review of biochemistry* **86**, 27-68 (2017).
3 Soto, C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nature Reviews Neuroscience* **4**, 49-60 (2003).
4 Soto, C. & Pritzkow, S. Protein misfolding, aggregation, and conformational strains in neurodegenerative diseases. *Nature neuroscience* **21**, 1332-1340 (2018).
5 Hipp, M. S., Kasturi, P. & Hartl, F. U. The proteostasis network and its decline in ageing. *Nature reviews Molecular cell biology* **20**, 421-435 (2019).
6 Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *science* **319**, 916-919 (2008).
7 Klaips, C. L., Jayaraj, G. G. & Hartl, F. U. Pathways of cellular proteostasis in aging and disease. *Journal of Cell Biology* **217**, 51-63 (2018).
8 Hartl, F. U., Bracher, A. & Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **475**, 324-332 (2011).
9 Levinthal, C. Are there pathways for protein folding? *Journal de chimie physique* **65**, 44-45 (1968).
10 Hartl, F. U. Molecular chaperones in cellular protein folding. *Nature* **381**, 571-580 (1996).
11 Balchin, D., Hayer-Hartl, M. & Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **353**, aac4354 (2016).
12 Levy-Rimler, G., Bell, R. E., Ben-Tal, N. & Azem, A. Type I chaperonins: not all are created equal. *FEBS letters* **529**, 1-5 (2002).
13 Mayer, M. P. Hsp70 chaperone dynamics and molecular mechanism. *Trends in biochemical sciences* **38**, 507-514 (2013).
14 Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature reviews Molecular cell biology* **11**, 515-528 (2010).
15 Schopf, F. H., Biebl, M. M. & Buchner, J. The HSP90 chaperone machinery. *Nature reviews Molecular cell biology* **18**, 345-360 (2017).
16 Kappé, G., Leunissen, J. A. & de Jong, W. W. Evolution and diversity of prokaryotic small heat shock proteins. *Small stress proteins*, 1-17 (2002).
17 Franck, E., Madsen, O., van Rheede, T., Ricard, G., Huynen, M. A. & de Jong, W. W. Evolutionary diversity of vertebrate small heat shock proteins. *Journal of molecular evolution* **59**, 792-805 (2004).
18 Waters, E. R. & Vierling, E. Plant small heat shock proteins–evolutionary and functional diversity. *New Phytologist* **227**, 24-37 (2020).

19 Carra, S., Alberti, S., Arrigo, P. A., Benesch, J. L., Benjamin, I. J., Boelens, W., Bartelt-Kirbach, B., Brundel, B. J., Buchner, J. & Bukau, B. The growing world of small heat shock proteins: from structure to functions. *Cell Stress and Chaperones* **22**, 601-611 (2017).
20 Janowska, M. K., Baughman, H. E., Woods, C. N. & Klevit, R. E. Mechanisms of small heat shock proteins. *Cold Spring Harbor Perspectives in Biology* **11**, a034025 (2019).
21 Ingolia, T. D. & Craig, E. A. Four small Drosophila heat shock proteins are related to each other and to mammalian alpha-crystallin. *Proceedings of the National Academy of Sciences* **79**, 2360-2364 (1982).
22 Petko, L. & Lindquist, S. Hsp26 is not required for growth at high temperatures, nor for thermotolerance, spore development, or germination. *Cell* **45**, 885-894 (1986).
23 Horwitz, J. Alpha-crystallin can function as a molecular chaperone. *Proceedings of the National Academy of Sciences* **89**, 10449-10453 (1992).
24 Jakob, U., Gaestel, M., Engel, K. & Buchner, J. Small heat shock proteins are molecular chaperones. *Journal of Biological Chemistry* **268**, 1517-1520 (1993).
25 Tanguay, R. M. & Hightower, L. E. *The big book on small heat shock proteins*. Vol. 8 (Springer, 2015).
26 Kriehuber, T., Rattei, T., Weinmaier, T., Bepperling, A., Haslbeck, M. & Buchner, J. Independent evolution of the core domain and its flanking sequences in small heat shock proteins. *The FASEB Journal* **24**, 3633-3642 (2010).
27 Alderson, T. R., Benesch, J. L. & Baldwin, A. J. Proline isomerization in the C-terminal region of HSP27. *Cell Stress and Chaperones* **22**, 639-651 (2017).
28 Kaiser, C. J., Peters, C., Schmid, P. W., Stavropoulou, M., Zou, J., Dahiya, V., Mymrikov, E. V., Rockel, B., Asami, S. & Haslbeck, M. The structure and oxidation of the eye lens chaperone αA-crystallin. *Nature structural & molecular biology* **26**, 1141-1150 (2019).
29 Yu, C., Leung, S. K. P., Zhang, W., Lai, L. T. F., Chan, Y. K., Wong, M. C., Benlekbir, S., Cui, Y., Jiang, L. & Lau, W. C. Y. Structural basis of substrate recognition and thermal protection by a small heat shock protein. *Nature communications* **12**, 3007 (2021).
30 White, H. E., Orlova, E. V., Chen, S., Wang, L., Ignatiou, A., Gowen, B., Stromer, T., Franzmann, T. M., Haslbeck, M. & Buchner, J. Multiple distinct assemblies reveal conformational flexibility in the small heat shock protein Hsp26. *Structure* **14**, 1197-1204 (2006).
31 Mchaourab, H. S., Lin, Y.-L. & Spiller, B. W. Crystal structure of an activated variant of small heat shock protein Hsp16. 5. *Biochemistry* **51**, 5105-5112 (2012).
32 Kim, K. K., Kim, R. & Kim, S.-H. Crystal structure of a small heat-shock protein. *Nature* **394**, 595-599 (1998).
33 Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S., Kühne, R., Stout, J. R., Higman, V. A., Klevit, R. E., van Rossum, B.-J. & Oschkinat, H. Solid-state NMR and SAXS studies provide a structural basis for the activation of αB-crystallin oligomers. *Nature structural & molecular biology* **17**, 1037-1042 (2010).
34 Sluchanko, N. N., Beelen, S., Kulikova, A. A., Weeks, S. D., Antson, A. A., Gusev, N. B. & Strelkov, S. V. Structural basis for the interaction of a human small heat shock protein with the 14-3-3 universal signaling regulator. *Structure* **25**, 305-316 (2017).

35 Jovcevski, B., Kelly, M. A., Rote, A. P., Berg, T., Gastall, H. Y., Benesch, J. L., Aquilina, J. A. & Ecroyd, H. Phosphomimics destabilize Hsp27 oligomeric assemblies and enhance chaperone activity. *Chemistry & biology* **22**, 186-195 (2015).
36 Aquilina, J. A., Benesch, J. L., Bateman, O. A., Slingsby, C. & Robinson, C. V. Polydispersity of a mammalian chaperone: mass spectrometry reveals the population of oligomers in αB-crystallin. *Proceedings of the National Academy of Sciences* **100**, 10611-10616 (2003).
37 Peschek, J., Braun, N., Rohrberg, J., Back, K. C., Kriehuber, T., Kastenmüller, A., Weinkauf, S. & Buchner, J. Regulated structural transitions unleash the chaperone activity of αB-crystallin. *Proceedings of the National Academy of Sciences* **110**, E3780-E3789 (2013).
38 Inoue, R., Sakamaki, Y., Takata, T., Wood, K., Morishima, K., Sato, N., Okuda, A., Shimizu, M., Urade, R. & Fujii, N. Elucidation of the mechanism of subunit exchange in αB crystallin oligomers. *Scientific Reports* **11**, 1-9 (2021).
39 Kennaway, C. K., Benesch, J. L., Gohlke, U., Wang, L., Robinson, C. V., Orlova, E. V., Saibi, H. R. & Keep, N. H. Dodecameric structure of the small heat shock protein Acr1 from Mycobacterium tuberculosis. *Journal of Biological Chemistry* **280**, 33419-33425 (2005).
40 Van Montfort, R. L., Basha, E., Friedrich, K. L., Slingsby, C. & Vierling, E. Crystal structure and assembly of a eukaryotic small heat shock protein. *Nature structural biology* **8**, 1025-1030 (2001).
41 Peschek, J., Braun, N., Franzmann, T. M., Georgalis, Y., Haslbeck, M., Weinkauf, S. & Buchner, J. The eye lens chaperone α-crystallin forms defined globular assemblies. *Proceedings of the National Academy of Sciences* **106**, 13272-13277 (2009).
42 Braun, N., Zacharias, M., Peschek, J., Kastenmüller, A., Zou, J., Hanzlik, M., Haslbeck, M., Rappsilber, J., Buchner, J. & Weinkauf, S. Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach. *Proceedings of the National Academy of Sciences* **108**, 20491-20496 (2011).
43 Koteiche, H. A., Chiu, S., Majdoch, R. L., Stewart, P. L. & Mchaourab, H. S. Atomic models by cryo-EM and site-directed spin labeling: application to the N-terminal region of Hsp16. 5. *Structure* **13**, 1165-1171 (2005).
44 Mishra, S., Chandler, S. A., Williams, D., Claxton, D. P., Koteiche, H. A., Stewart, P. L., Benesch, J. L. & Mchaourab, H. S. Engineering of a polydisperse small heat-shock protein reveals conserved motifs of oligomer plasticity. *Structure* **26**, 1116-1126. e1114 (2018).
45 Haslbeck, M., Walke, S., Stromer, T., Ehrnsperger, M., White, H. E., Chen, S., Saibil, H. R. & Buchner, J. Hsp26: a temperature-regulated chaperone. *The EMBO journal* **18**, 6744-6751 (1999).
46 Fleckenstein, T., Kastenmüller, A., Stein, M. L., Peters, C., Daake, M., Krause, M., Weinfurtner, D., Haslbeck, M., Weinkauf, S. & Groll, M. The chaperone activity of the developmental small heat shock protein Sip1 is regulated by pH-dependent conformational changes. *Molecular cell* **58**, 1067-1078 (2015).
47 Jehle, S., van Rossum, B., Stout, J. R., Noguchi, S. M., Falber, K., Rehbein, K., Oschkinat, H., Klevit, R. E. & Rajagopal, P. αB-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer. *Journal of molecular biology* **385**, 1481-1497 (2009).

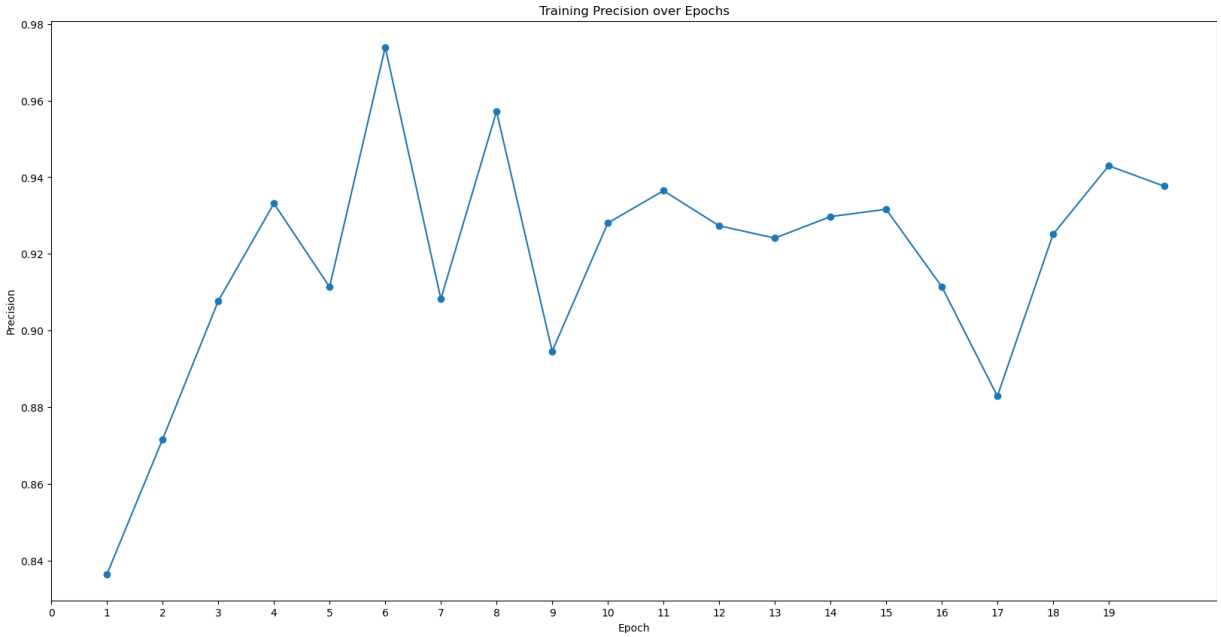
48 Baldwin, A. J., Lioe, H., Robinson, C. V., Kay, L. E. & Benesch, J. L. αB-crystallin polydispersity is a consequence of unbiased quaternary dynamics. *Journal of molecular biology* **413**, 297-309 (2011).
49 Lelj-Garolla, B. & Mauk, A. G. Self-association and chaperone activity of Hsp27 are thermally activated. *Journal of Biological Chemistry* **281**, 8169-8174 (2006).
50 Sha, E., Nakamura, M., Ankai, K., Yamamoto, Y. Y., Oka, T. & Yohda, M. Functional and structural characterization of HspB1/Hsp27 from Chinese hamster ovary cells. *FEBS open bio* **9**, 1826-1834 (2019).
51 Lelj-Garolla, B. & Mauk, A. G. Roles of the N-and C-terminal sequences in Hsp27 self-association and chaperone activity. *Protein Science* **21**, 122-133 (2012).
52 Shashidharamurthy, R., Koteiche, H. A., Dong, J. & Mchaourab, H. S. Mechanism of chaperone function in small heat shock proteins: dissociation of the HSP27 oligomer is required for recognition and binding of destabilized T4 lysozyme. *Journal of Biological Chemistry* **280**, 5281-5289 (2005).
53 Alderson, T. R., Roche, J., Gastall, H. Y., Dias, D. M., Pritišanac, I., Ying, J., Bax, A., Benesch, J. L. & Baldwin, A. J. Local unfolding of the HSP27 monomer regulates chaperone activity. *Nature communications* **10**, 1068 (2019).
54 Stengel, F., Baldwin, A. J., Painter, A. J., Jaya, N., Basha, E., Kay, L. E., Vierling, E., Robinson, C. V. & Benesch, J. L. Quaternary dynamics and plasticity underlie small heat shock protein chaperone function. *Proceedings of the National Academy of Sciences* **107**, 2007-2012 (2010).
55 Bepperling, A., Alte, F., Kriehuber, T., Braun, N., Weinkauf, S., Groll, M., Haslbeck, M. & Buchner, J. Alternative bacterial two-component small heat shock protein systems. *Proceedings of the National Academy of Sciences* **109**, 20407-20412 (2012).
56 Mühlhofer, M., Peters, C., Kriehuber, T., Kreuzeder, M., Kazman, P., Rodina, N., Reif, B., Haslbeck, M., Weinkauf, S. & Buchner, J. Phosphorylation activates the yeast small heat shock protein Hsp26 by weakening domain contacts in the oligomer ensemble. *Nature Communications* **12**, 6697 (2021).
57 Ahmad, M. F., Raman, B., Ramakrishna, T. & Rao, C. M. Effect of phosphorylation on αB-crystallin: differences in stability, subunit exchange and chaperone activity of homo and mixed oligomers of αB-crystallin and its phosphorylation-mimicking mutant. *Journal of molecular biology* **375**, 1040-1051 (2008).
58 Ito, H., Kamei, K., Iwamoto, I., Inaguma, Y., Nohara, D. & Kato, K. Phosphorylation-induced change of the oligomerization state of αB-crystallin. *Journal of Biological Chemistry* **276**, 5346-5352 (2001).
59 Koteiche, H. A. & Mchaourab, H. S. Mechanism of chaperone function in small heat-shock proteins: phosphorylation-induced activation of two-mode binding in αB-crystallin. *Journal of Biological Chemistry* **278**, 10361-10367 (2003).
60 Ecroyd, H., Meehan, S., Horwitz, J., Aquilina, J. A., Benesch, J. L., Robinson, C. V., Macphee, C. E. & Carver, J. A. Mimicking phosphorylation of αB-crystallin affects its chaperone activity. *Biochemical Journal* **401**, 129-141 (2007).
61 Weeks, S. D., Muranova, L. K., Heirbaut, M., Beelen, S., Strelkov, S. V. & Gusev, N. B. Characterization of human small heat shock protein HSPB1 α-crystallin domain localized mutants associated with hereditary motor neuron diseases. *Scientific Reports* **8**, 1-15 (2018).

62 Liu, Z., Zhang, S., Gu, J., Tong, Y., Li, Y., Gui, X., Long, H., Wang, C., Zhao, C. & Lu, J. Hsp27 chaperones FUS phase separation under the modulation of stress-induced phosphorylation. *Nature structural & molecular biology* **27**, 363-372 (2020).
63 Delbecq, S. P. & Klevit, R. E. HSPB5 engages multiple states of a destabilized client to enhance chaperone activity in a stress-dependent manner. *Journal of Biological Chemistry* **294**, 3261-3270 (2019).
64 Freilich, R., Betegon, M., Tse, E., Mok, S.-A., Julien, O., Agard, D. A., Southworth, D. R., Takeuchi, K. & Gestwicki, J. E. Competing protein-protein interactions regulate binding of Hsp27 to its client protein tau. *Nature communications* **9**, 4563 (2018).
65 Mainz, A., Peschek, J., Stavropoulou, M., Back, K. C., Bardiaux, B., Asami, S., Prade, E., Peters, C., Weinkauf, S. & Buchner, J. The chaperone αB-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nature structural & molecular biology* **22**, 898-905 (2015).
66 Rajagopal, P., Liu, Y., Shi, L., Clouser, A. F. & Klevit, R. E. Structure of the α-crystallin domain from the redox-sensitive chaperone, HSPB1. *Journal of biomolecular NMR* **63**, 223-228 (2015).
67 Clouser, A. F. & Klevit, R. E. pH-dependent structural modulation is conserved in the human small heat shock protein HSBP1. *Cell Stress and Chaperones* **22**, 569-575 (2017).
68 Baughman, H. E., Pham, T.-H. T., Adams, C. S., Nath, A. & Klevit, R. E. Release of a disordered domain enhances HspB1 chaperone activity toward tau. *Proceedings of the National Academy of Sciences* **117**, 2923-2929 (2020).
69 Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophysical journal* **78**, 1606-1619 (2000).
70 Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nature methods* **14**, 290-296 (2017).
71 Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A. J. & Berger, B. Positive-unlabeled convolutional neural networks for particle picking in cryo-electron micrographs. *Nature methods* **16**, 1153-1160 (2019).
72 Danev, R. & Baumeister, W. Cryo-EM single particle analysis with the Volta phase plate. *elife* **5**, e13046 (2016).
73 Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *Journal of molecular biology* **333**, 721-745 (2003).
74 Nguyen, H. P., McGuire, K. L., Cook, B. D. & Herzik Jr, M. A. Manual blot-and-plunge freezing of biological specimens for single-particle cryogenic electron microscopy. *JoVE (Journal of Visualized Experiments)*, e62765 (2022).
75 Levitz, T. S., Weckener, M., Fong, I., Naismith, J. H., Drennan, C. L., Brignole, E. J., Clare, D. K. & Darrow, M. C. Approaches to using the chameleon: robust, automated, fast-plunge cryoEM specimen preparation. *Frontiers in Molecular Biosciences* **9** (2022).
76 Carver, J. A., Grosas, A. B., Ecroyd, H. & Quinlan, R. A. The functional roles of the unstructured N-and C-terminal regions in αB-crystallin and other mammalian small heat-shock proteins. *Cell Stress and Chaperones* **22**, 627-638 (2017).

77 Punjani, A. & Fleet, D. J. 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *Journal of structural biology* **213**, 107702 (2021).
78 Zhong, E. D., Bepler, T., Berger, B. & Davis, J. H. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. *Nature methods* **18**, 176-185 (2021).
79 Kinman, L. F., Powell, B. M., Zhong, E. D., Berger, B. & Davis, J. H. Uncovering structural ensembles from single-particle cryo-EM data using cryoDRGN. *Nature Protocols* **18**, 319-339 (2023).
80 Rosenbaum, D., Garnelo, M., Zielinski, M., Beattie, C., Clancy, E., Huber, A., Kohli, P., Senior, A. W., Jumper, J. & Doersch, C. Inferring a continuous distribution of atom coordinates from cryo-EM images using VAEs. *arXiv preprint arXiv:2106.14108* (2021).

```
1   MTERR VPFSL LRGPS WDPFR DWYPH SRLFD QAFGL PRLPE EWSQW LGGSS
51  WPGYV RPLPP AAIES PAVAA PAYSR ALSRQ LSSGV SEIRH TADRW RVSLD
101 VNHFA PDELT VKTKD GVVEI TGKHE ERQDE HGYIS RCFSR KYTLP PGVDP
151 TQVSS SLSPE GTLTV EAPMP KLATQ SNEIT IPVTF ESRAQ LGGPE AAKSD
201 ETAAK *
```

Figure 3.S1: The sequence of HSPB1. The NTD is green, the ACD is orange, and the CTD is purple.



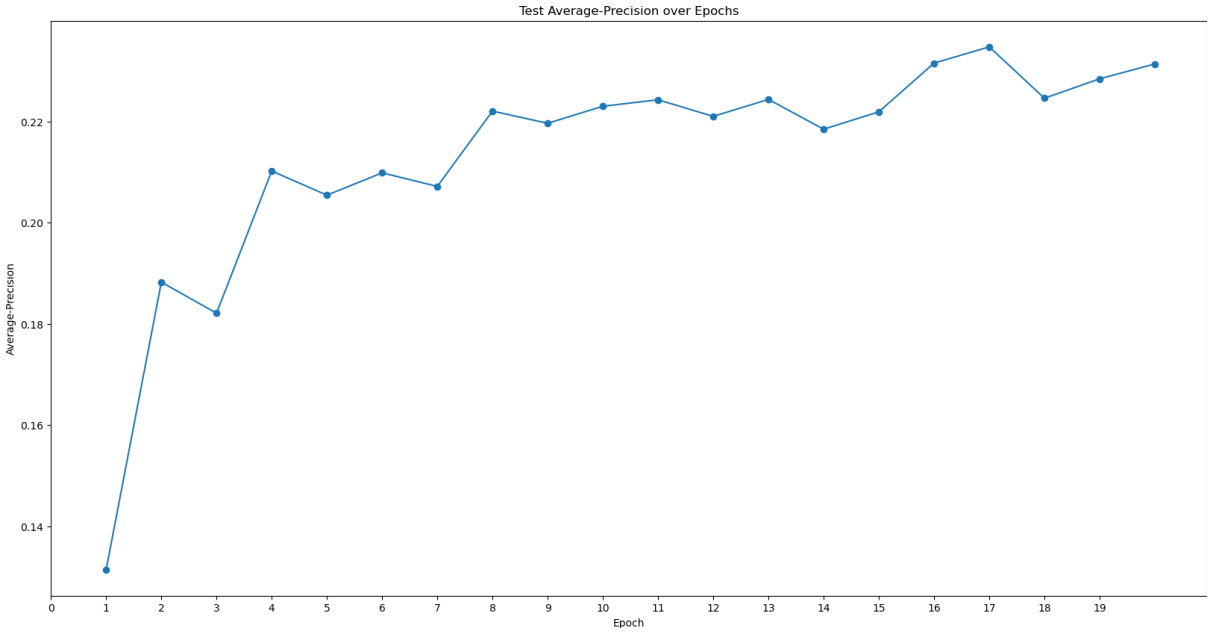


Figure 3.S2: **Training Precision and Test Average-Precision for Topaz**

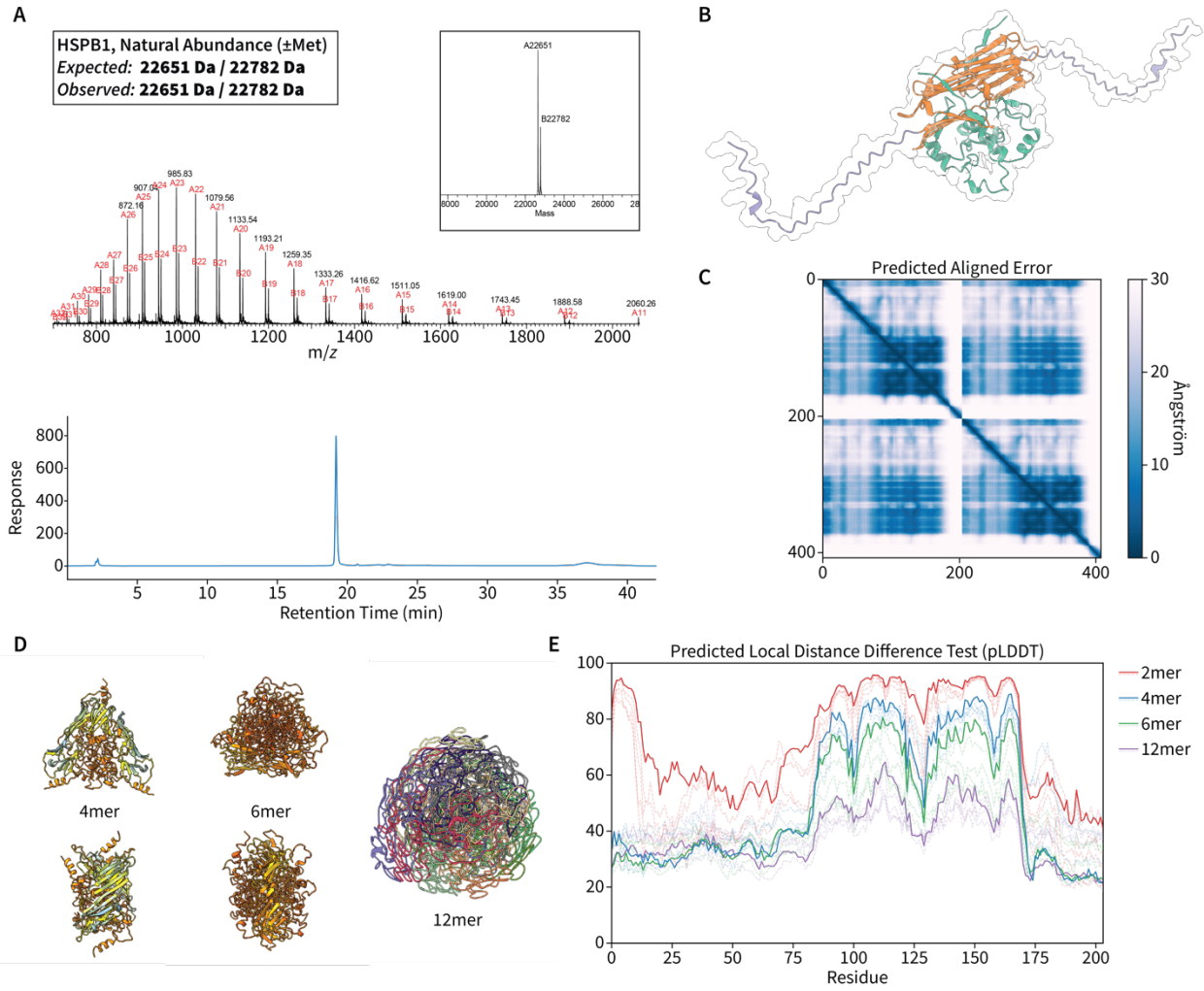


Figure 3.S3: Characterization of HSPB1 *in vitro* and *in silico*. A) QTOF-MS data and a representative analytical HPLC chromatogram for purified HSPB1. **B)** AlphaFold-multimer predicted structure for the HSPB1 dimer. The NTD is presented in green, the ACD in orange, and the CTD in purple. **C)** Predicted Aligned Error (PAE) for the AlphaFold-multimer predicted structure for the HSPB1 dimer. **D)** AlphaFold-multimer predicted structures for 4mer, 6mer, and 12mer of HSPB1.The 4mer and 6mer predictions are colored by pLDDT. The 12mer prediction is colored by chain. PLDDT scores for all predictions are provided in Figure 3.S3E. **E)** Predicted Local Distance Difference Test (pLDDT) scores for all AlphaFold-multimer predictions. Scores for the top 5 predicted structures for each oligomer class are presented, with the best score represented by a solid opaque line and the remaining scores presented as dashed semi-transparent lines.

ACKNOWLEDGEMENTS

Chapter 3, in part, is currently being prepared for submission for publication. Berkeley, R. F., Plonski, A. P., Cook, B. D., McGuire, K. L., Herzik, M. A., & Debelouchina, G. T. The dissertation author was the primary investigator and author of this material.

Characterization of the small heat shock protein HSPB1 in the presence of a liquid-liquid phase separated client

Introduction

The discovery of protein liquid-liquid phase separation (LLPS) and its role in biomolecular organization in the cell has introduced a new dimension to the study of protein self-assembly[1-4]. Unlike the interactions that underpin the formation of fibrils, aggregates, and other static protein assemblies, LLPS is driven primarily by transient, nonspecific interactions between residues in participating proteins[6]. This process yields mesoscale structures in which proteins retain their intrinsic disorder and exhibit a range of dynamics, which extend from rapid motions comparable to those observed for monomeric proteins[7] to very slow and rigid dynamics[8].

At length scales beyond those of the component proteins, LLPS droplets can adopt material states ranging from liquids to gels and solids[9-12]. Liquid-like LLPS assemblies exhibit emergent properties of liquids such as wetting and fusion[13], phenomena that result from surface tension at the phase boundary of the droplet[14,15].

Many proteins known to undergo LLPS are also prone to forming aggregates and fibrils. LLPS droplets have been observed to undergo phase transitions between material states *in vitro*, maturing from liquid-like droplets to gels and aggregates with time[9-12]. The relationship between the liquid droplet state and the solid aggregate state and the factors that contribute to phase transitions between these two states are not fully understood. It is probable that the protein-protein interactions that contribute to the maturation of LLPS droplets into gels *in vitro* play a

role in the formation of pathological protein aggregates and fibrils *in vivo*. Studies have shown that mutations in proteins implicated in neurodegenerative disease etiology can accelerate the maturation of LLPS droplets formed by these proteins. For example, the FUS G156E mutation discussed in Chapter 2, which has been shown to accelerate the liquid-to-solid transition of FUS[16] and FUS' low-complexity domain (FUS LC)[12], is associated with familial amyotrophic lateral sclerosis[17]. Similar irreversible liquid-to-solid transitions have been identified in proteins that are found in pathological aggregates in neurodegenerative disease, including tau[18], α-synuclein[19], and TDP-43[20], and it is likely that mutations and dysregulation in post-translational modifications contribute to the onset of pathological aggregation in these proteins as well.

Pathological aggregates are targets for association by molecular chaperones, particularly small heat shock proteins (sHsps). Unlike other molecular chaperones, sHsps are striking in their lack of substrate specificity and ability to engage client proteins in different states[21]. Similar to pathological liquid-to-solid transitions, the exact mechanism of action of sHsps is not completely understood, with the only common denominator being that they participate in interactions with proteins as part of the unfolded protein response, particularly as a reaction to cellular stress. The versatility of sHsps in client binding stands in contrast to the relatively simple tripartite domain architecture and structure of sHsps, which are discussed at length in Chapter 3 of this dissertation. Emerging evidence suggests that the disordered NTD plays an outsized role in engaging amorphous clients and that the β-sheet-rich ACD is responsible for binding β-sheet-rich fibrillar clients[22,23]. Regardless of the binding mode, sHsps have been shown to interact with intrinsically disordered clients, aggregates, and fibrils, which suggests that they are capable of engaging client proteins at every stage of the LLPS droplet maturation process. Indeed, a number

of recent studies have shown that sHsps are present during and capable of modulating liquid-to-solid transitions of proteins both *in vitro* and *in vivo*[23-25].

In this chapter, our aim is to characterize the structure and dynamics of the sHsp HSPB1 during interactions with client proteins at each step of a liquid-to-solid transition. This work builds directly on Chapter 3, where we investigated the architecture of HSPB1 oligomers in the absence of any client protein using cryo-electron microscopy. To model LLPS systems, we use phase-separated FUS LC, which HSPB1 has been shown to interact with in both the LLPS and fibril state[26]. As a point of comparison, we use α-synuclein fibrils, which are also known to be clients of HSPB1[27]. These models of HSPB1 in the client-free, LLPS-bound, and fibril-bound state allow us to dissect the structure and dynamics of the three domains of HSPB1 in order to better understand the role of each domain in binding clients in different states.

To address the challenges associated with the structural plasticity and tendency towards self-assembly of each of the proteins involved in this study, we employ magic angle spinning (MAS) NMR spectroscopy. MAS NMR is uniquely capable of extracting information about both structure and dynamics from the heterogeneous systems under study at physiological temperatures. MAS NMR has been used to investigate the structural characteristics and dynamics of sHsp oligomers in the past[28-30]. Our approach uses a suite of experiments that leverage insensitive nuclei enhancement by polarization transfer (INEPT)[31] and cross polarization (CP)[32] to enhance the NMR signal. INEPT uses J-coupling interactions to transfer polarization between nuclei that are connected by a covalent bond. Although the J-coupling interaction is isotropic and is therefore not subject to line broadening induced by anisotropy, the

faster T_2 relaxation times associated with rigid molecules can eclipse the time needed for coherence transfer between two nuclei. The practical consequence of this is that INEPT-based polarization transfer is only efficient for molecules engaging in rapid motions with longer T_2 times, and that INEPT-based experiments will filter for mobile components of an NMR sample. Unlike INEPT, CP uses anisotropic dipolar coupling for polarization transfer between nuclei. This interaction occurs through space and is efficient only for nuclei that are engaging in slow motions. Taken in conjunction, we can apply INEPT-based and CP-based experiments to interrogate protein systems with heterogeneous dynamics by capturing dynamic, mobile parts of the protein with INEPT-based experiments and rigid parts of the protein with CP-based experiments. In this work, we use this approach to dissect the structure and dynamics of the NTD, ACD, and CTD of HSPB1 in the presence of clients such as FUS LC and α-synuclein, and in the context of LLPS and amyloids.

Materials and methods

Expression and purification of HSPB1 and related constructs

HSPB1 was purified via ammonium sulfate precipitation and preparative HPLC as described in Chapter 3 of this dissertation. Briefly, the HSPB1 cell pellet was thawed if necessary, resuspended in lysis buffer (50 mM sodium phosphate, 50 mM sodium chloride, Roche cOmplete™ EDTA-free Protease Inhibitor Cocktail, pH 7.4) and lysed by pulsed sonication for 30 minutes using a Qsonica sonicator with a 1/8” diameter probe tip at 12 kHz (60%) output and 4°C. The lysate was cleared by centrifugation at 30,000 RCF for 30 minutes and the supernatant

was decanted into an Erlenmeyer flask with a stir bar. To this was added saturated ammonium sulfate (4.1 M), dropwise, to a final concentration of 40% saturation (1.64 M) at room temperature. The solution was allowed to stir for 30 minutes, after which the precipitated protein was pelleted by centrifugation at 30,000 RCF for 30 minutes at 4 °C. The supernatant was discarded and the pellet was redissolved in 20 mM tris(hydroxymethyl)aminomethane, 10 mM magnesium chloride, 30 mM ammonium chloride, pH 7.6 at room temperature. The resulting solution was desalted by running over PD-10 desalting columns packed with Sephadex G-25 resin (Cytiva, Marlborough, MA) equilibrated in the same buffer. The desalted HSPB1 solution was further purified by ion exchange chromatography over a Mono Q anion exchange column (Cytiva, Marlborough, MA) equilibrated in the same buffer. Protein was eluted from the column using a 0-250 mM NaCl gradient. The fractions containing HSPB1 were concentrated and subjected to reverse-phase HPLC purification over a Waters XBridge Peptide BEH C18 OBD prep column (130 Å pore size, 10 µm particle size, 19 mm X 250 mm) using a Waters 2535 Binary Gradient Module equipped with a 2484 UV/Vis detector (Waters Corporation, Milford, MA). Sample purity was confirmed using an analytical Waters XBridge Peptide BEH C18 column (300 Å pore size, 5 µm particle size, 2.1 mm X 100 mm) on the same HPLC system. Sample identity was confirmed by intact mass QTOF-MS using an Agilent 1260 Infinity Binary LC coupled to a 6230 Accurate-Mass TOF-MS system (Agilent Technologies, Santa Clara, CA). All HPLC purifications utilized a gradient method of water and acetonitrile with 0.1% trifluoroacetic acid as the mobile phases. After HPLC purification, samples were lyophilized and refolded by dialysis at 4 °C from 6 M guanidine hydrochloride, 50 mM sodium phosphate, 50 mM sodium chloride, pH 7.4, into the same buffer without denaturant. If samples were not used

immediately after preparation, they were concentrated to ~1 mM and flash frozen at -80 °C for storage.

All HSPB1 intein constructs were prepared using HiFi Assembly (New England Biolabs, Ipswich, MA). Constructs bearing a 6x-Histidine tag were purified by Ni-NTA affinity chromatography. Briefly, the cell pellet was thawed if previously frozen, resuspended in lysis buffer (50 mM sodium phosphate, 50 mM sodium chloride, Roche cOmplete™ EDTA-free Protease Inhibitor Cocktail, pH 7.4) and lysed by pulsed sonication for 30 minutes using a Qsonica sonicator with a 1/8" diameter probe tip at 12 kHz (60%) output and 4°C. The lysate was cleared by centrifugation at 30,000 RCF for 30 minutes and the supernatant was incubated with 3 mL Thermo Scientific HisPur Ni-NTA Resin per liter of culture (Thermo Fisher Scientific, Waltham, MA) for 30 min at 4°C. The suspension of beads was washed with 10 column volumes of lysis buffer containing 10 mM imidazole, and protein was eluted with two column volumes of lysis buffer containing 250 mM imidazole. Ni-NTA binding was confirmed using an SDS-PAGE gel. For non-intein constructs, the eluted protein was incubated with 6xHis-tagged Tobacco Etch Virus (TEV) protease at a ratio of 1:100 TEV to protein for 4 h and 25°C to remove the 6x-Histidine tag. For intein constructs, this step was performed after intein splicing followed by a round of preparative HPLC to purify the spliced 6xHis-MBP-TEV-HSPB1 construct. The TEV splicing reaction mixture was subjected to reverse-phase HPLC purification over a Waters XBridge Peptide BEH C18 OBD prep column (130 Å pore size, 10 µm particle size, 19 mm X 250 mm) using a Waters 2545 Binary Gradient Module equipped with a 2484 UV/Vis detector (Waters Corporation, Milford, MA). Sample purity was confirmed using an analytical Waters XBridge Peptide BEH C18 column (300 Å pore size, 5 µm particle size, 2.1 mm X 100 mm) on

the same HPLC system. Sample identity was confirmed by intact mass QTOFMS using an Agilent 1260 Infinity Binary LC coupled to a 6230 Accurate-Mass TOFMS system (Agilent Technologies, Santa Clara, CA). All HPLC purifications utilized a gradient method of water and acetonitrile with 0.1% trifluoroacetic acid as the mobile phases. After HPLC purification, samples were lyophilized and refolded by dialysis from 6 M guanidine hydrochloride, 50 mM sodium phosphate, 50 mM sodium chloride, pH 7.4, into the same buffer without denaturant. If samples were not used immediately after preparation, they were concentrated to 1-2 mM and flash frozen at -80 °C for storage.

Intein Splicing and Purification

Initial attempts to perform splicing reactions with HSPB1$_{NTD}$-Cfa^{N}-6xHis and 6xHis-Cfa^{C}-HSPB1$_{ACD-CTD}$ (**Fig. 4.S1A**) were confounded by poor solubility of the Cfa^{N} construct. Although splicing did proceed under these conditions, we sought to improve splicing performance by introducing an MBP solubility tag to the Cfa^{N} construct to produce a final 6xHis-MBP-TEV-HSPB1$_{NTD}$-Cfa^{N}-6xHis fusion protein (**Fig. 4.S1B**). This construct was more soluble.

To prepare segmentally labeled HSPB1, ^{15}N,^{13}C-6xHis-MBP-TEV-HSPB1$_{NTD}$-Cfa^{N}-6xHis and 6xHis-Cfa^{C}-HSPB1$_{ACD-CTD}$ were subjected to Ni-NTA chromatography as described in the previous section. Fractions containing the intein products were diluted in 50 mM sodium phosphate, 200 mM sodium chloride, 1 mM EDTA, 2 mM TCEP, pH 8.0 and incubated for 30 minutes at 30 °C before mixing to a final concentration of 15 μM ^{15}N,^{13}C-6xHis-MBP-TEV-HSPB1$_{NTD}$-Cfa^{N}-6xHis and 30 μM 6xHis-Cfa^{C}-HSPB1$_{ACD-CTD}$. The intein splicing reaction was

allowed to proceed overnight. The splicing reaction was quenched by adding 8 M guanidine HCl and 4% trifluoroacetic acid at a 3:1 v/v ratio of reaction to quenching solution and purified by preparative HPLC as described in the previous section.

Expression and purification of FUS LC

FUS LC was purified as described previously[12]. FUS LC was prepared from recombinant Rosetta (DE3) competent *Escherichia coli* cells (MilliporeSigma, Burlington, MA) that had been transformed with a plasmid encoding the 6xHis-MBP-TEV-FUS (1–163) sequence. This plasmid was a gift from Nicolas Fawzi (Addgene plasmid no. 98653; Addgene, Watertown, MA)[33]. Seed cultures were grown to saturation from freshly transformed colonies and inoculated at a 1% v/v ratio into either Luria-Bertani or $^{15}N/^{13}C$-M9 medium supplemented with kanamycin (50 μg/mL). The cultures were grown at 37 °C to an OD600 of ~0.7 and protein expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside. The cultures were allowed to express protein for 4 h at 37°C before being harvested by centrifugation at 10,000 × g and 4 °C for 30 min. After decanting the supernatant, cell pellets were stored at −80 °C for later use. In addition to the wild-type FUS LC fusion protein, the pathogenic mutant FUS LC G156E protein was prepared. To generate the mutant, the requisite G to E mutation was introduced to the wild-type FUS LC plasmid using an NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs, Ipswich, MA). Expression and purification conditions of the G156E mutant protein were identical to those of the wild-type.

To purify recombinantly expressed FUS LC, we followed the published protocol by Burke et al. with some modifications[7]. The frozen cell pellet was thawed, resuspended in lysis buffer (20 mM sodium phosphate, 300 mM sodium chloride, Roche cOmplete EDTA-free Protease Inhibitor Cocktail (MilliporeSigma, Burlington, MA) (pH 7.4, 4°C), and lysed by pulsed sonication for 30 min at 4°C using a Qsonica sonicator (Qsonica, Newton, CT) with a 1/8″ diameter probe tip at 12 kHz (60%) output. The lysate was cleared by centrifugation at 20,000 × *g* for 30 min and the supernatant was incubated with Thermo Scientific HisPur Ni-NTA Resin (Thermo Fisher Scientific, Waltham, MA) for 30 min at 4°C. The suspension of beads was washed with 10 column volumes of lysis buffer containing 10 mM imidazole, and protein was eluted with two column volumes of lysis buffer containing 250 mM imidazole. The eluted protein was incubated with 6xHis-tagged Tobacco Etch Virus (TEV) protease at a ratio of 1:150 TEV to MBP-FUS LC for 5 h and 25°C to cleave MBP from FUS LC. FUS LC crashes out upon cleavage to yield a cloudy suspension with white clumps of aggregated protein. After TEV cleavage, 8 M urea was added to the reaction mixture to solubilize the aggregated FUS LC. The TEV reaction was monitored by SDS-PAGE (note that FUS LC does not bind Coomassie, so cleavage was verified by the gel shift of the MBP band). The mixture containing solubilized FUS LC was diluted to 2 mg/ml with 20 mM CAPS and 150 mM sodium chloride (pH 11), and subjected to size-exclusion chromatography over a GE HiLoad 16/600 Superdex 75-pg column (GE Healthcare, Chicago, IL). Purity was verified by high-performance liquid chromatography (HPLC), mass spectrometry, and by comparing the A260 and A280 absorbance to verify that no nucleic acid contaminants were present in the sample. This protocol generally yields 10–15 mg of FUS LC per liter of culture in both Luria-Bertani and M9 media.

Expression and purification of α-synuclein

Samples of α-synuclein were prepared via an ammonium sulfate-based approach similar to the one described for HSPB1 in Chapter 3 of this dissertation. Recombinant α-synuclein was prepared from Rosetta (DE3) competent *Escherichia coli* cells (MilliporeSigma, Burlington, MA) that had been transformed with a plasmid encoding the wild-type α-synuclein sequence. Seed cultures were grown to saturation from freshly transformed colonies and inoculated at a 1% v/v ratio into either Luria-Bertani or $^{15}N/^{13}C$-M9 medium supplemented with ampicillin (100 μg/mL). The cultures were grown at 37 °C to an OD600 of ~0.6 and protein expression was induced by the addition of 1 mM isopropyl-β-D-thiogalactoside. The cultures were allowed to express protein overnight at 18 °C before being harvested by centrifugation at 10,000 × g and 4 °C for 30 min. After decanting the supernatant, cell pellets were stored at −80 °C for later use.

To purify α-synuclein, the cell pellet was thawed if previously frozen, resuspended in lysis buffer (50 mM sodium phosphate, 50 mM sodium chloride, Roche cOmplete™ EDTA-free Protease Inhibitor Cocktail, pH 7.4) and lysed by pulsed sonication for 30 minutes using a Qsonica sonicator with a 1/8” diameter probe tip at 12 kHz (60%) output and 4°C. The lysate was cleared by centrifugation at 30,000 RCF for 30 minutes and the supernatant was decanted into an Erlenmeyer flask with a stir bar. To this was added saturated ammonium sulfate (4.1 M), dropwise, to a final concentration of 40% saturation (1.64 M) at room temperature. The solution was allowed to stir for 30 minutes, after which the precipitated protein was pelleted by centrifugation at 30,000 RCF for 30 minutes at 4 °C. The supernatant was discarded and the pellet was redissolved in 20 mM tris(hydroxymethyl)aminomethane, 10 mM magnesium

chloride, 30 mM ammonium chloride, pH 7.6 at room temperature. The resulting solution was desalted by running over PD-10 desalting columns packed with Sephadex G-25 resin (Cytiva, Marlborough, MA) equilibrated in the same buffer. The desalted α-synuclein solution was concentrated and subjected to reverse-phase HPLC purification over a Waters XBridge Peptide BEH C18 OBD prep column (130 Å pore size, 10 µm particle size, 19 mm X 250 mm) using a Waters 2535 Binary Gradient Module equipped with a 2484 UV/Vis detector (Waters Corporation, Milford, MA). Sample purity was confirmed using a Waters XBridge Peptide BEH C18 column (300 Å pore size, 5 µm particle size, 2.1 mm X 100 mm) on the same HPLC system. Sample identity was confirmed by intact mass QTOFMS using an Agilent 1260 Infinity Binary LC coupled to a 6230 Accurate-Mass TOFMS system (Agilent Technologies, Santa Clara, CA). All HPLC purifications utilized a gradient method of water and acetonitrile with 0.1% trifluoroacetic acid as the mobile phases. After HPLC purification, samples were lyophilized and refolded by dialysis at 4 °C from 6 M guanidine hydrochloride, 50 mM sodium phosphate, 50 mM sodium chloride, pH 7.4, into the same buffer without denaturant. If samples were not used immediately after preparation, they were concentrated to ~1 mM and flash frozen at -80 °C for storage.

Fibril formation of α-synuclein

Our approach to preparing α-synuclein fibrils was based on protocols described in the literature[34]. Monomeric α-synuclein was prepared in 500 µl aliquots at 5 mg/ml in 50 mM sodium phosphate, 50 mM sodium chloride, 2 mM TCEP, pH 7.4. Aliquots were incubated for 7

days at 37 °C with constant agitation at 1,000 RPM on an orbital shaker. Fibril formation was confirmed by negative-stain TEM.

Negative Staining and TEM

Freshly glow-discharged Formvar 300 mesh carbon-on-copper grids (Electron Microscopy Sciences, Hatfield, PA) were placed onto 10 µl drops of sample arrayed on a Parafilm sheet. After allowing protein to adsorb onto the grids for 5 minutes, the grids were washed with buffer 3 times and applied to drops of 1% w/v uranyl acetate (Ladd Research Industries, Williston, VT) arrayed on a Parafilm sheet and allowed to stain for 1 minute. After staining, excess uranyl acetate solution was wicked from each grid and the grids were allowed to dry for at least an hour before insertion into the microscope. Grids were viewed using a JEOL JEM-1400Plus transmission electron microscope operating at 80 kV and micrographs were recorded using a Gatan OneView digital camera.

Solid-state nuclear magnetic resonance experiments

For samples containing HSPB1 only, protein was concentrated in 10 kDa MWCO centrifugal concentrators and packed into the rotor either by pipetting or scooping the material (in the case of the hydrogel observed at extremely high HSPB1 concentrations). For experiments with LLPS FUS LC, stocks of 1.2 mM monomeric FUS LC in 20 mM CAPS, 150 mM sodium chloride, pH 11 were subjected to LLPS by rapid dilution into 50 mM sodium phosphate, 300 mM sodium chloride, 5 mM TCEP, 0.1% sodium azide, pH 7.4 with HSPB1 to a final concentration of 300

μM FUS LC and 150 μM HSPB1. For experiments with α-synuclein fibrils, fibril stocks were pelleted by ultracentrifugation and resuspended in 50 mM sodium phosphate, 50 mM sodium chloride, 5 mM TCEP, 0.1% sodium azide, pH 7.4 with HSPB1. Samples were packed at a final concentration of 3 mg, or 2.51 mM ^{15}N, ^{13}C-HSPB1, and 20 mg, or 27.66 mM α-synuclein.

All experiments were performed using 3.2 mm zirconia MAS rotors with either 30 μL or 50 μL (thin-walled) sample volumes. Spectra were acquired on a 17.6 T (750 MHz ^{1}H Larmor frequency) NMR spectrometer equipped with a 3.2 mm E^{free} H/C/N MAS probe (Bruker Biospin, Billerica, MA). All reported experiments were performed at an MAS frequency of 11.11 kHz. The temperature was set to 275 K, although control CP, INEPT, DP, and CP-DARR experiments were acquired for HSPB1 at 15 kHz MAS and 275 K as well as 11.11 kHz MAS and 298 K with no notable differences in the spectra. We estimate that the true sample temperature during all experiments is 10-15K higher than the set temperature based on temperature shift experiments with water. Pulse parameters and other details are provided in the supporting materials.

All NMR data were analyzed and visualized with POKY[35], nmrglue[36], and matplotlib[37]. Statistical analyses of chemical shift data were performed with the help of PACSY[38]. For 2D distributions of proline chemical shifts, BMRB assignments for proline were collected and a tuple of theoretical assignments was generated for every pair of carbons in each assigned proline. This approach results in a dataset of pairs of chemical shifts representing every assigned pair of proline carbons in the BMRB, and can be taken as a proxy for expected cross-peaks in CP-DARR experiments.

Results

HSPB1 forms dodecamers in which the NTD is sequestered and rigid in the absence of a client protein

In our hands, HPLC-purified and refolded HSPB1 forms dodecameric oligomers as determined by size-exclusion chromatography and analytical ultracentrifugation (**Fig. 3.1**). We have also shown that refolded HSPB1 forms cage-like oligomer structures by cryo-EM (**Fig. 3.2, 3.3**). Model fitting to these densities suggests that the oligomers are maintained via NTD-NTD interactions that occur either inside of the cage or around the edge of the cage depending on the oligomer architecture.

Studies on the structure and dynamics of sHsp domains typically agree that the NTD of sHsps are disordered and rigid, the ACD is rich in β-sheet secondary structure and rigid, and the CTD is disordered and dynamic[39]. To characterize the local structure and dynamics of HSPB1 in the absence of a client protein, we performed MAS NMR experiments on samples of full-length HSPB1. As demonstrated in previous studies, we found that the high molecular weight of HSPB1 oligomers and the sedimentation effects of MAS precluded the need for PEG precipitation or any other treatment in order to achieve efficient cross-polarization (CP) for oligomeric HSPB1 samples[40,41]. We also prepared samples at a range of concentrations (400 μg, 15 mg, and 35 mg; corresponding to approximately 340 μM, 13 mM, and 30 mM, respectively, in a 3.2 mM thin-walled MAS NMR rotor), and found that although HSPB1 forms a gel at concentrations above 30 mM, the shape of the NMR spectra and the relative intensity of correlations in the spectra are

similar. This suggests that the range of HSPB1 concentrations used in this study does not impact oligomer integrity, and that changes in the structure and dynamics observed in the presence of client proteins are the result of interactions with client proteins rather than a concentration effect (**Fig. 4.S3, 4.S4**).

We recorded a suite of CP- and INEPT-based experiments to capture the different dynamics exhibited by the three domains in HSPB1 (**Fig. 4.1**). Consistent with the high mobility reported for the CTD, resonances in 1D and 2D HN-INEPT, HC-INEPT, and INEPT-TOBSY experiments performed with HSPB1 samples are consistent with the residue distribution in the CTD of HSPB1 (**Fig. 4.1A, 4.1B**). The identity of correlations in our HN-INEPT experiments matches assigned ^{1}H-^{15}N HSQC data present in the literature (**Fig. 4.1B**)[5].

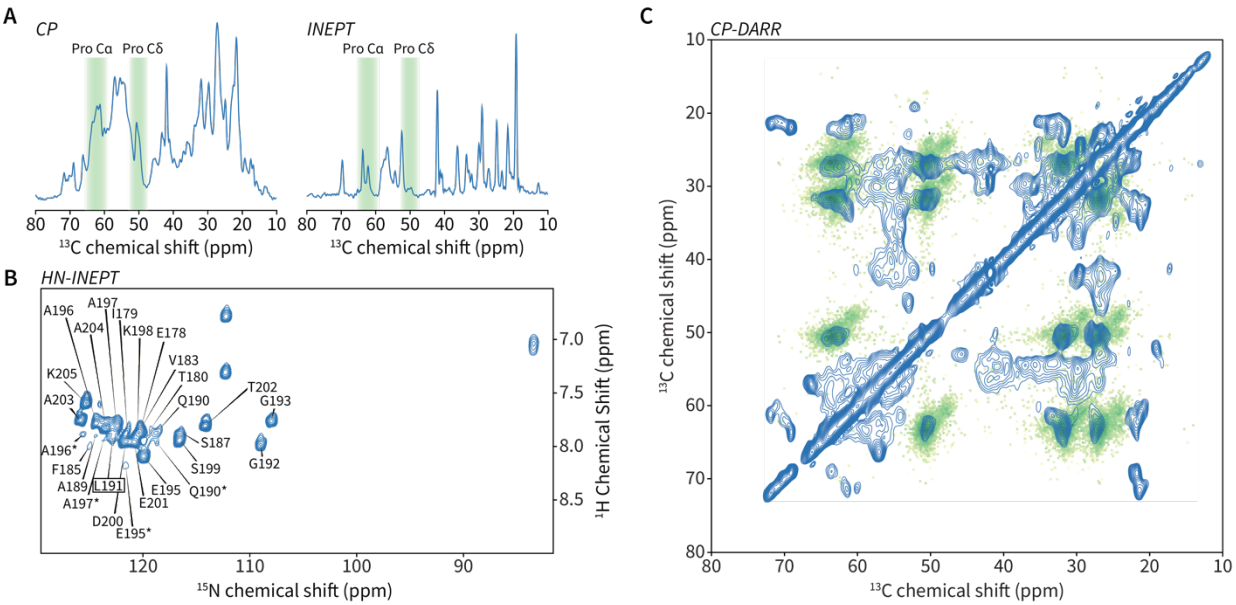


Figure 4.1. Characterization of HSPB1 by MAS NMR. (A) 1D ^{13}C CP and INEPT spectra of the aliphatic region of HSPB1. Green bars represent probability distributions for diagnostic proline chemical shifts in the BMRB. **(B)** ^{1}H,^{15}N-INEPT spectrum of HSPB1. Assignments for residues in the CTD are taken from Alderson et al., 2017[5]. **(C)** ^{13}C-^{13}C CP-DARR of HSPB1. Green hexbin plot represents the distribution of chemical shifts for all carbon-carbon pairs in proline.

Based on our cryo-EM data with HSPB1, as well as MAS NMR data and structural models for other sHsps[29,42], we expected that both the NTD and ACD regions would be rigid enough to generate signal in CP-based MAS NMR experiments. Even if the NTD is rigid enough to be observed by CP-based MAS NMR, it is possible that the disordered nature of the NTD would induce line broadening severe enough to preclude the interpretation of signal from the NTD. Indeed, we observed significant line broadening in all of our CP-based experiments, likely due to structural heterogeneity in the NTD and the overall structural and architectural heterogeneity of HSPB1 oligomers. These effects are particularly noticeable in the DARR, with obvious line-broadening between 52 and 60 ppm and 25 and 45 ppm, the general area where Cα-Cβ correlations would be expected, respectively (**Fig. 4.1C**).

Despite the line-broadening effects of heterogeneity in HSPB1's structure, a number of diagnostic residues enable interpretation of CP-based spectra of HSPB1. Analysis of the amino acid composition of the NTD and ACD reveals that there are three unique residues in the ACD (lysine, asparagine, and cysteine) and no unique residues in the NTD (**Fig. 4.2A**). There are also a handful of residues whose relative abundance in either domain makes them good representatives of their respective domain, such as threonine or valine in the ACD and proline, alanine, and leucine in the NTD (**Fig. 4.2A**). To leverage a unique residue in an analysis by MAS NMR without resonance assignments, the residue must contain atoms with non-degenerate chemical shifts.

Given the requirements for diagnostic residues, the only signal capable of disambiguating the ACD and NTD contributions is that of lysine Cε, which is both unique to the ACD and exhibits a

characteristic chemical shift that does not overlap with any of the other residues in HSPB1 (**Fig. 4.2B**). In our HSPB1 experiments, lysine Cε manifests as a clear, sharp peak at 42.1 ppm (**Fig. 4.2B**). The narrow linewidth of this peak suggests that the chemical environment of the part of HSPB1 containing lysine is more or less uniform, an observation that is consistent with the fact that all of the lysines in the ACD are present in rigid, folded regions of the HSPB1 ACD and they point out towards the solvent in the dodecamer models of HSPB1 presented in Chapter 3 of this dissertation (**Fig. 4.3A**). In addition to lysine Cε, threonine and valine Cγ are also reasonable indicators for the ACD that exhibit signals between 21.5 and 22.5 ppm (**Fig. 4.2B**).

Although there are no residues that are totally unique to the NTD, proline serves as a good indicator due to its relative abundance in the NTD as compared to the ACD (**Fig. 4.2A**) and non-degenerate Cα and Cδ chemical shifts. Proline correlations appear in the DARR spectrum as five distinctive peaks on each side of the diagonal that correspond to each combination of through-space correlations between carbons around the proline ring. The identity of these peaks can be confirmed using chemical shift statistics for proline in the BMRB, overlaid on our spectra in a green hexbin plot for the DARR or green bars for the 1D CP and INEPT spectra (**Fig. 4.1A, 4.1C, 4.3B**).

To confirm that the diagnostic proline residues in our CP spectra do indeed correspond to prolines in the NTD, we developed an intein-mediated segmental labeling approach for HSPB1. (**Fig. 4.2C**). Although there is a single cysteine residue in the middle of the ACD (C136), we elected to introduce an S83C mutation at the juncture of the NTD and ACD to unequivocally separate the NTD and ACD in our NMR spectra. We chose to use the Cfa_{GEP} engineered split

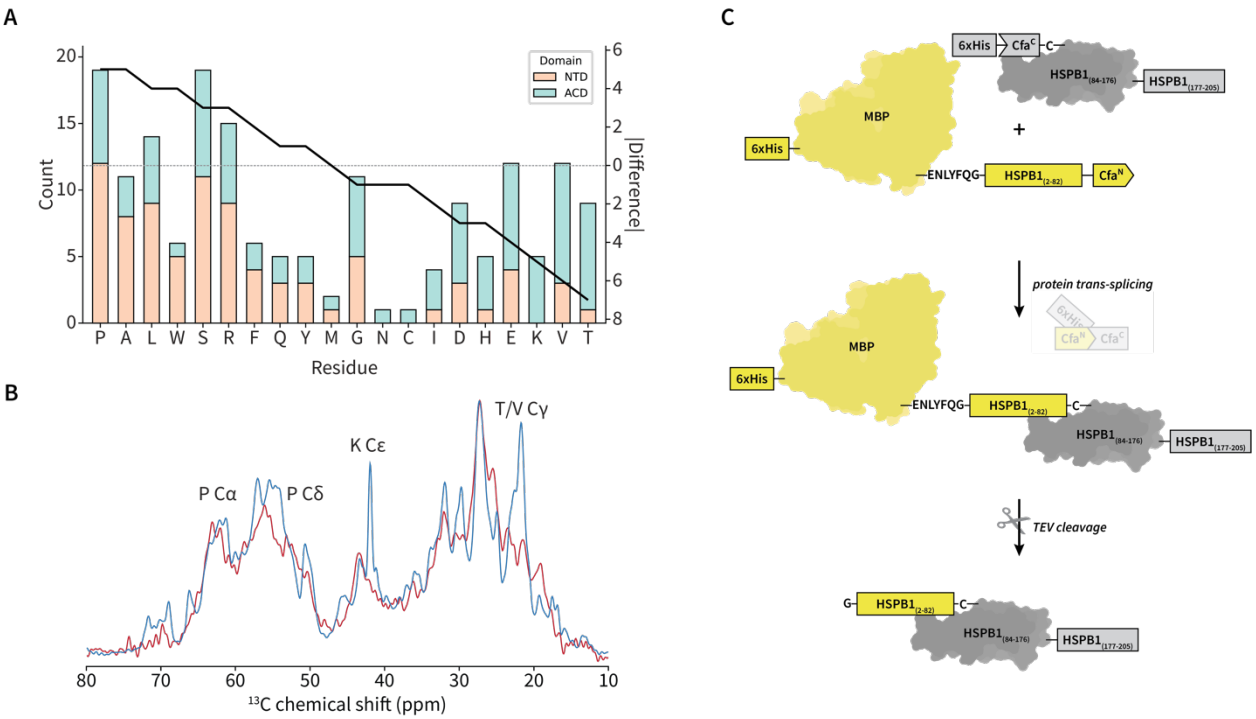


Figure 4.2. Dissecting NMR signals from the NTD and ACD with split inteins. (A) A stacked bar plot representing the residue distribution of the NTD and ACD of HSPB1. Bars representing the NTD are orange and bars representing the ACD are green. Bar height represents the number of residues in each domain, indexed on the left of the plot. The line represents the absolute difference in residue abundance between the NTD and ACD, indexed on the right of the plot. **(B)** 1D ^{13}C CP spectra of the segmentally labeled HSPB1 construct in red and the uniformly labeled HSPB1 construct in blue. Notable residues are labeled. **(C)** An overview of the protein splicing approach, with the ^{15}N,^{13}C-labeled construct in yellow and the natural abundance construct in gray.

intein system to catalyze splicing due to its rapid kinetics, insensitivity to extein sequence, and demonstrated performance under a range of splicing conditions[43,44]. We also introduced an MBP fusion tag to enhance the otherwise poor solubility of the NTD-Cfa^N construct. This strategy enabled the preparation of enough segmentally labeled HSPB1 for MAS NMR studies.

CP spectra of segmentally labeled HSPB1 S83C containing a ^{15}N,^{13}C-labeled NTD show that a significant amount of resolvable signal is derived from the NTD despite its largely disordered nature (**Fig. 4.2C**). Signals associated with residues that are overrepresented in the ACD, such as

those for lysine Cε and threonine and valine Cγ, are absent from the spectrum (**Fig. 4.2B**). Perhaps most importantly, the shape of the spectrum suggests that a significant portion of the proline CP signal comes from prolines in the NTD. This result indicates that the characteristic proline Cα and Cδ signals are effective indicators for NTD behavior in CP-derived MAS NMR spectra of HSPB1.

The NTD of HSPB1 becomes more dynamic in the presence of phase-separated FUS LC

HSPB1 has an established role in modulating LLPS and blocking pathological liquid-to-solid phase transitions of FUS LC and similar RNA-binding proteins[23,25]. The NTD and ACD domains of sHsps have been functionally implicated in engaging disordered and fibrillar clients, respectively[22,23], but there is little information regarding the structure of HSPB1 in the presence of either class of client, nor any information regarding the structure of any sHsp in the presence of an LLPS client.

To characterize the interactions between HSPB1 and an LLPS client, we prepared mixed samples containing FUS LC and ^{15}N, ^{13}C-labeled HSPB1. Spectra acquired with a sample containing 15 mg of FUS LC and 15 mg ^{15}N, ^{13}C-HSPB1 appeared similar to spectra of ^{15}N, ^{13}C-HSPB1 without FUS LC, suggesting that the majority of HSPB1 remained in the oligomeric state under these conditions (**Fig. 4.S3**). To capture HSPB1 in a highly activated, client-engaged state, we instead leveraged the fact that HSPB1 partitions into FUS LC LLPS droplets by gently packing LLPS droplets of FUS LC in the presence of ^{15}N, ^{13}C-HSPB1 into an MAS NMR rotor by gentle centrifugation[7,12]. To prepare this sample, we induced LLPS of monomeric FUS LC in CAPS buffer at pH 11 by dilution into a phosphate buffer at pH 7.4 containing ^{15}N, ^{13}C-HSPB1 for a

final concentration of 300 μM FUS LC and a 2:1 molar ratio of FUS LC:^{15}N, ^{13}C-HSPB1. The sample was incubated for 30 minutes on ice to further promote LLPS and the resulting sample was packed into an MAS rotor. FUS LC LLPS was weaker in the presence of HSPB1, with only 42% of FUS LC demixing into droplets (**Fig. 4.S4A**) as compared to ~66% under similar buffer conditions in the absence of HSPB1[12]. Notably, very little HSPB1 partitions into FUS LC droplets (**Fig. 4.S4A**) in general, with the majority of the labeled HSPB1 remaining in the supernatant after centrifugation of the LLPS droplets into the MAS rotor. Comparisons to dilute HSPB1 samples in the absence of client protein suggest that 250-500 μg of HSPB1 are present in the copartitioned sample (**Fig 4.S4B**).

Spectra for CP-based experiments of the copartitioned HSPB1 sample reveal several key differences when compared to spectra of HSPB1 in the absence of client (**Fig. 4.3B, 4.3C**). The most striking feature is that the distinctive proline Cα and Cδ peaks are much less prominent in both the CP and DARR spectra. In the CP spectrum (**Fig. 4.3B**), the loss of proline signal is particularly striking for Cδ. Despite a significant change in the proline signal, the lysine Cε signal is much less affected, aside from a possible small reduction in intensity. Thus, when the two key diagnostic residues in HSPB1 are considered, and a substantial reduction of signal is observed for proline Cα and Cδ without a concurrent loss of signal for lysine Cε, differential behavior of the NTD and ACD is evident. Given the poor efficiency of CP-based experiments for highly mobile regions of protein, it is likely that the NTD of HSPB1 is becoming more dynamic in the presence of FUS LC without a concurrent increase in dynamics for the ACD.

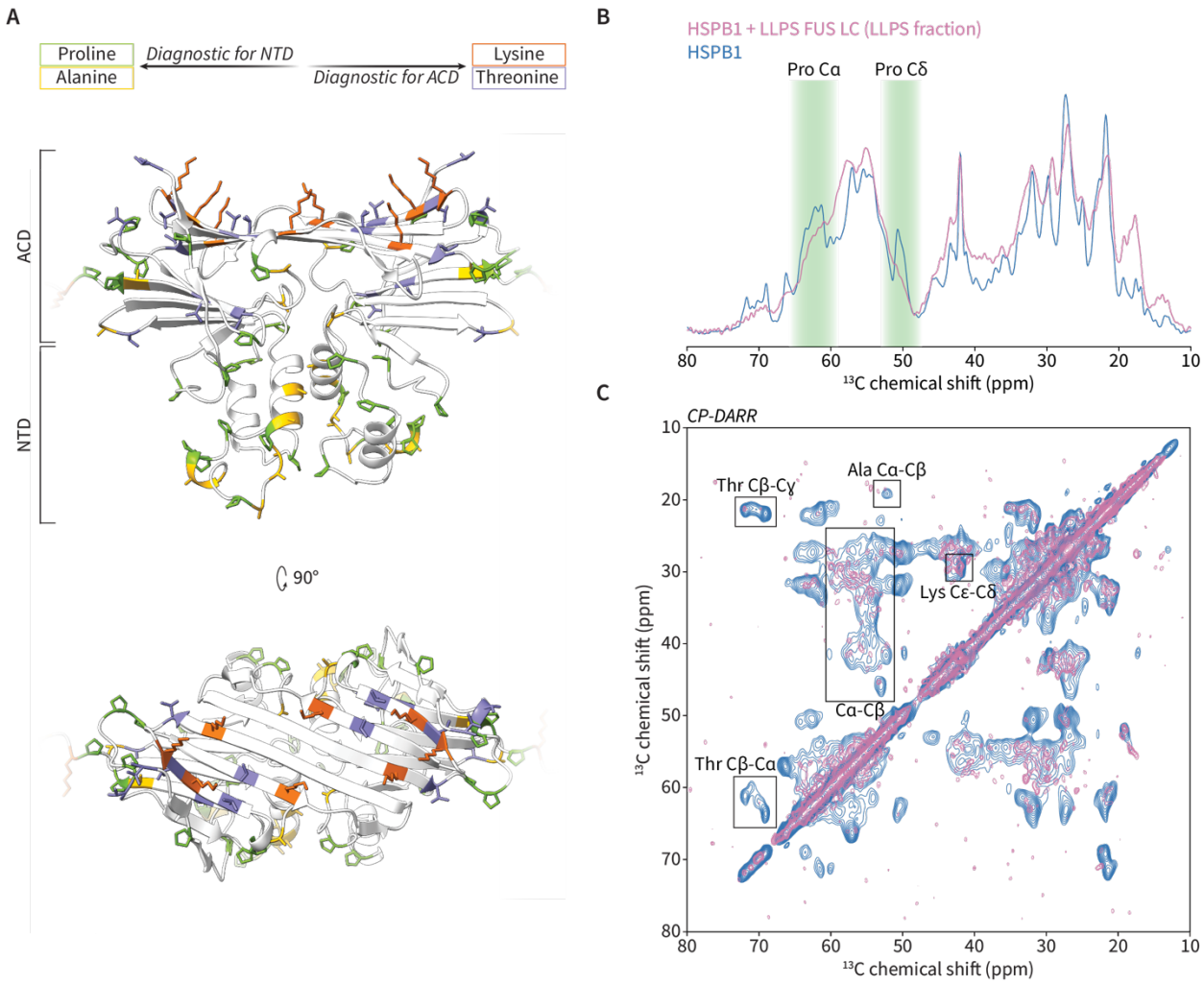


Figure 4.3. MAS NMR data for HSPB1 copartitioned with FUS LC LLPS droplets (A) An overview of the structural distribution of diagnostic residues in the NTD and ACD of HSPB1 as predicted by AlphaFold2. Prolines are rendered in green, lysines in orange, threonines in purple, and alanines in yellow. The diagnostic power of the labeled residues for the NTD or ACD is noted. **(B)** 1D ^{13}C CP spectra for HSPB1 in blue overlaid with the spectrum for the HSPB1 and FUS LC LLPS copartitioned sample in pink. Chemical shift probability distributions for proline residues from the BMRB are denoted in green. **(C)** ^{13}C-^{13}C CP-DARR spectra of HSPB1 only (blue) and the HSPB1 and FUS LC LLPS copartitioned sample (pink). Correlations representing diagnostic residues other than proline are labeled.

In addition to the changes observed in signal intensity of diagnostic proline and lysine residues, there is a significant increase in signal upfield of 20 ppm. The only carbon chemical shifts in that region for proteins are alanine Cβ and isoleucine Cδ and Cγ2. Alanine is enriched the NTD, although there are a number of alanines present in the ACD β-sheet formed by strands β3, β8,

and β9 (**Fig. 4.3A**). Isoleucine, on the other hand, is enriched in the ACD. The fact that every isoleucine residue contributes two Cβ signals increases its relevance in a comparative analysis of the NTD and ACD. Further work is required to determine which residues are responsible for this change.

Although poor signal-to-noise hinders detailed interpretation of the DARR spectrum, it is apparent that densities for proline, as well as threonine and alanine–unambiguous residues with diagnostic potential for either the NTD or ACD–have all become indistinguishable from the noise (**Fig. 4.3C**). Despite this, the broad signal in the region corresponding to Cα-Cβ correlations from 25 to 45 ppm in the first dimension and 52 to 60 ppm in the second dimension remains visible despite being less intense than the proline, threonine, or alanine correlations in the client-free DARR. Finally, peaks corresponding to lysine Cε remain strong in in the DARR.

In addition to changes in features in CP-based spectra, differences in signal intensity reveal changes in the overall mobility of HSPB1 in the presence of client. To establish a more accurate comparison for the copartitioned sample and to determine whether the observed changes were not merely a result of concentration effects, we prepared a dilute sample containing 400 μg HSPB1. Noteworthy changes in the copartitioned sample, such as the decrease in proline signal intensity, were absent in the HSPB1 dilute sample (**Fig. 4.S2B**). Interestingly, the copartitioned sample exhibited higher signal intensity in the CP spectrum but lower intensity in the INEPT spectrum compared to the dilute sample despite identical experimental parameters for CP and INEPT experiments between the two samples (**Fig. 4.S2B**). This suggests that the overall mobility of HSPB1 is reduced when LLPS FUS LC is present, either through protein-protein interactions with the client, or due to the viscous nature of the LLPS environment, or both.

Taken together, changes in CP-based spectra induced by the LLPS FUS LC client imply that the degree of dynamics in HSPB1 have increased, especially in the NTD as reported by a significant reduction in proline signal. The fact that the only truly diagnostic residue in HSPB1, lysine, remains largely unaffected by the introduction of LLPS FUS LC indicates that the ACD remains relatively rigid and immobile even when LLPS FUS LC is present. A structural model of the HSPB1 dimer (**Fig. 4.3A**), along with our cryo-EM models for HSPB1 oligomers, can help contextualize the differences observed between HSPB1 spectra acquired in the presence and absence of LLPS FUS LC. The NTD, rich in proline and alanine residues, is tucked beneath the folded ACD, likely engaging in interactions with the NTD of adjacent HSPB1 dimers to maintain oligomer integrity in the absence of a client. Note that although transient secondary structure is thought to exist in HSPB1 and is predicted by AlphaFold2 in the structural model, the NTD is thought to be largely disordered and the AlphaFold2 confidence in the NTD structure prediction is relatively low (**Fig. 3.S3C, 3.S3D, 3.S3E**). This rigidity manifests as strong proline signal in the HSPB1 CP spectrum. Threonines are spread throughout the ACD and are absent from the NTD aside from T2, which is unlikely to be observable in our CP-based experiments. Lysine, the diagnostic residue least affected by LLPS FUS LC, is only present in the upper β-sheet consisting of the β4, β5, and β6+7 strands. In our models of HSPB1 dodecamers, the upper β-sheet faces the outside of the oligomer. Upon introduction of a dynamic and disordered LLPS client, the NTD is exposed and free to engage in transient, dynamic interactions with FUS LC. It is possible that this change does not necessarily lead to oligomer decomposition, as ACD-CTD interactions are also involved in maintaining oligomer integrity and oligomer architectures that partially or completely expose the NTD have been established for other sHsps[45]. Indeed, the

slower bulk dynamics of HSPB1 in the copartitioned sample is consistent with previous reports for sHsps demonstrating that amorphous clients promote the formation of larger oligomers[46] and that the large oligomer state is predominant in the presence of LLPS droplets[23].

Changes in the dynamics of the NTD are distinct in the presence of α-synuclein fibrils

The ACD of sHsps is most often implicated in interactions with fibrillar clients[22]. To determine whether the distinction between the NTD and ACD's role in engaging disordered or fibrillar clients holds true for our system, we prepared samples of α-synuclein fibrils as model fibrillar clients for HSPB1. HSPB1 appears to interact with α-synuclein fibrils at various ratios, with obvious client engagement occurring at all tested ratios of HSPB1 and α-synuclein (**Fig. 4.4A**). Using a similar rationale to the copartitioned FUS LC samples, we attempted to bias the population of HSPB1 in our sample away from HSPB1 oligomers and towards α-synuclein-bound HSPB1 by maximizing the ratio of HSPB1:α-synuclein. To this end, we resuspended 20 mg α-synuclein fibrils with 3 mg ^{15}N, ^{13}C-HSPB1 (a 1:10 molar ratio) and packed the resulting slurry into a MAS NMR rotor. To reinforce our conclusions from this sample, we also prepared a second rotor with a similar amount of natural abundance α-synuclein. Data acquired on the α-synuclein sample were adjusted to account for differences in signal-to-noise resulting from differences in the number of scans (the CP spectrum the natural abundance sample was divided by $\sqrt{2}$ to account for its acquisition with twice the number of scans) and subtracted from the mixed α-synuclein and ^{15}N, ^{13}C-HSPB1 sample.

The CP spectrum of the copacked HSPB1 and α-synuclein sample shows broader linewidths overall, and the profile of the spectrum does not change significantly after subtracting the α-synuclein spectrum (**Fig. 4.4B**). Other than the broader peaks in the copacked sample, there are fewer striking differences between the client-free HSPB1 sample and the copacked α-synuclein sample than there are when comparing the client-free HSPB1 sample and the copartitioned LLPS FUS LC sample. The diagnostic proline signal absent in the copartitioned LLPS FUS LC and HSPB1 sample are present in the α-synuclein sample, especially the Cα. The intensity changes in the aliphatic region of the spectrum observed for the copartitioned LLPS sample also do not seem to occur for the copacked α-synuclein sample. Overall, this suggests that the NTD does not experience the changes in dynamics seen for the LLPS sample when presented with a fibrillar client. Further work is required to fully evaluate the effect of fibrillar client binding by HSPB1,

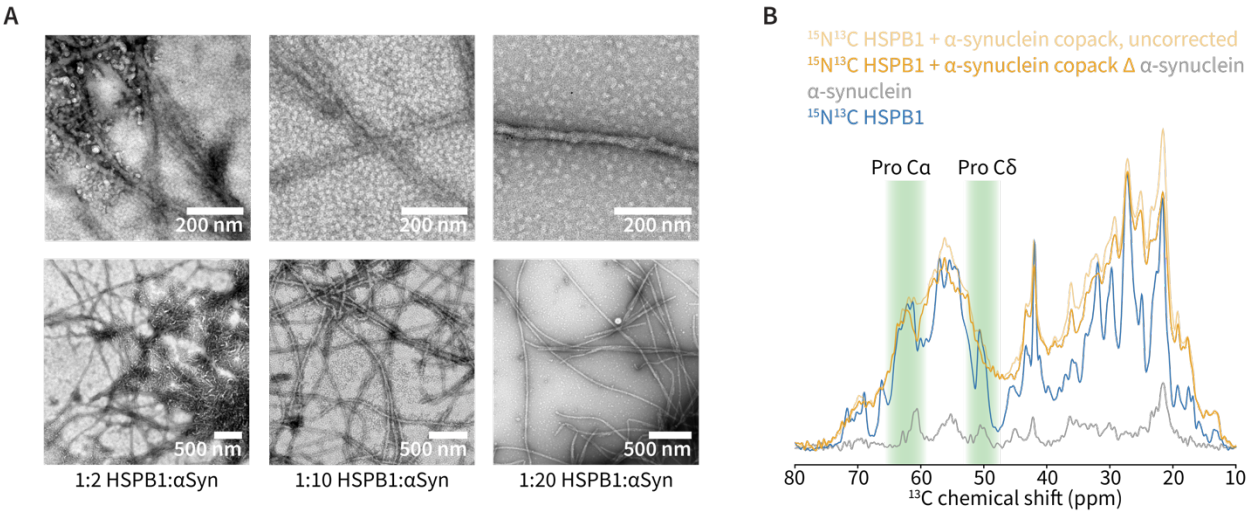


Figure 4.4. Characterization of HSPB1 interactions with α-synuclein fibrils (A) Negative stain TEM images of HSPB1 and α-synuclein fibrils at low and high ratios. **(B)** 1D ^{13}C CP spectra of HSPB1 in the absence and presence of α-synuclein at a 1:10 molar ratio. HSPB1 in the absence of α-synuclein is in blue, a spectrum of natural abundance α-synuclein is in gray, and a spectrum of the HSPB1 sample copacked with α-synuclein fibrils is in dashed orange. A difference spectrum of the natural abundance α-synuclein data subtracted from the copacked data is in orange.

but these results suggests that the mode of binding for fibrillar clients is distinct from that of LLPS clients.

Conclusion

HSPB1 and related small heat shock proteins have the ability to interact with client proteins across various states of pathological self-assembly. This versatility is a key component of the unfolded protein response and is vital for maintaining proteostasis. Attempts to unveil the structural mechanism of the promiscuity of sHsps towards clients have been confounded by the propensity of sHsps to themselves self-assemble into dynamic, polydisperse oligomers. In this chapter, we attempt to dissect the tripartite domain architecture of HSPB1 using MAS NMR in order to understand to behavior of HSPB1 domains in the presence of clients in different states. In particular, we focus on as-yet uncharacterized interactions between HSPB1 and an LLPS client. Using information about the architecture of HSPB1 oligomers presented in Chapter 3 of this dissertation to contextualize our results, we characterize the differences in MAS NMR spectra acquired in the absence and presence of FUS LC in the LLPS state. Based on the observed signals in CP and DARR spectra of HSPB1, we find that the NTD domain of HSPB1 becomes more dynamic in the presence of an LLPS client, and that changes observed in the presence of an LLPS client are distinct from those observed in the presence of a fibrillar client. These results not only build upon previous findings related to the structural differences in sHsps when bound to fibrillar or aggregated cleints[22], but also begin to provide a structural rationale for the ability of HSPB1 to modulate the liquid-to-solid phase transitions of FUS and other RNA-binding proteins[23,25].

References

1 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* **324**, 1729-1732 (2009).

2 Brangwynne, C. P., Mitchison, T. J. & Hyman, A. A. Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes. *Proceedings of the National Academy of Sciences* **108**, 4334-4339 (2011).

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

4 Gomes, E. & Shorter, J. The molecular language of membraneless organelles. *Journal of Biological Chemistry* **294**, 7115-7127 (2019).

5 Alderson, T. R., Benesch, J. L. & Baldwin, A. J. Proline isomerization in the C-terminal region of HSP27. *Cell Stress and Chaperones* **22**, 639-651 (2017).

6 Wang, J., Choi, J.-M., Holehouse, A. S., Lee, H. O., Zhang, X., Jahnel, M., Maharana, S., Lemaitre, R., Pozniakovsky, A. & Drechsel, D. A molecular grammar governing the driving forces for phase separation of prion-like RNA binding proteins. *Cell* **174**, 688-699. e616 (2018).

7 Burke, K. A., Janke, A. M., Rhine, C. L. & Fawzi, N. L. Residue-by-residue view of in vitro FUS granules that bind the C-terminal domain of RNA polymerase II. *Molecular cell* **60**, 231-241 (2015).

8 Riback, J. A., Katanski, C. D., Kear-Scott, J. L., Pilipenko, E. V., Rojek, A. E., Sosnick, T. R. & Drummond, D. A. Stress-triggered phase separation is an adaptive, evolutionarily tuned response. *Cell* **168**, 1028-1040. e1019 (2017).

9 Kato, M., Han, T. W., Xie, S., Shi, K., Du, X., Wu, L. C., Mirzaei, H., Goldsmith, E. J., Longgood, J. & Pei, J. Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**, 753-767 (2012).

10 Murray, D. T., Kato, M., Lin, Y., Thurber, K. R., Hung, I., McKnight, S. L. & Tycko, R. Structure of FUS protein fibrils and its relevance to self-assembly and phase separation of low-complexity domains. *Cell* **171**, 615-627. e616 (2017).

11 Ackermann, B. E. & Debelouchina, G. T. Heterochromatin protein HP1α gelation dynamics revealed by solid-state NMR spectroscopy. *Angewandte Chemie International Edition* **58**, 6300-6305 (2019).

12 Berkeley, R. F., Kashefi, M. & Debelouchina, G. T. Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC. *Biophysical Journal* **120**, 1276-1287 (2021).

13 Hyman, A. A., Weber, C. A. & Jülicher, F. Liquid-liquid phase separation in biology. *Annual review of cell and developmental biology* **30**, 39-58 (2014).

14 Alberti, S., Gladfelter, A. & Mittag, T. Considerations and challenges in studying liquid-liquid phase separation and biomolecular condensates. *Cell* **176**, 419-434 (2019).

15 Widom, B. Note on the interfacial tension of phase-separated polymer solutions. *Journal of statistical physics* **52**, 1343-1351 (1988).

16 Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S. & Franzmann, T. M. A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162**, 1066-1077 (2015).
17 Ticozzi, N., Silani, V., LeClerc, A. L., Keagle, P., Gellera, C., Ratti, A., Taroni, F., Kwiatkowski, T. J., McKenna-Yasek, D. M. & Sapp, P. C. Analysis of FUS gene mutation in familial amyotrophic lateral sclerosis within an Italian cohort. *Neurology* **73**, 1180-1185 (2009).
18 Ambadipudi, S., Biernat, J., Riedel, D., Mandelkow, E. & Zweckstetter, M. Liquid–liquid phase separation of the microtubule-binding repeats of the Alzheimer-related protein Tau. *Nature communications* **8**, 275 (2017).
19 Ray, S., Singh, N., Kumar, R., Patel, K., Pandey, S., Datta, D., Mahato, J., Panigrahi, R., Navalkar, A. & Mehra, S. α-Synuclein aggregation nucleates through liquid–liquid phase separation. *Nature chemistry* **12**, 705-716 (2020).
20 Schmidt, H. B., Barreau, A. & Rohatgi, R. Phase separation-deficient TDP43 remains functional in splicing. *Nature communications* **10**, 4890 (2019).
21 Mymrikov, E. V., Daake, M., Richter, B., Haslbeck, M. & Buchner, J. The chaperone activity and substrate spectrum of human small heat shock proteins. *Journal of Biological Chemistry* **292**, 672-684 (2017).
22 Mainz, A., Peschek, J., Stavropoulou, M., Back, K. C., Bardiaux, B., Asami, S., Prade, E., Peters, C., Weinkauf, S. & Buchner, J. The chaperone αB-crystallin uses different interfaces to capture an amorphous and an amyloid client. *Nature structural & molecular biology* **22**, 898-905 (2015).
23 Liu, Z., Zhang, S., Gu, J., Tong, Y., Li, Y., Gui, X., Long, H., Wang, C., Zhao, C. & Lu, J. Hsp27 chaperones FUS phase separation under the modulation of stress-induced phosphorylation. *Nature structural & molecular biology* **27**, 363-372 (2020).
24 Boczek, E. E., Fürsch, J., Niedermeier, M. L., Jawerth, L., Jahnel, M., Ruer-Gruß, M., Kammer, K.-M., Heid, P., Mediani, L. & Wang, J. HspB8 prevents aberrant phase transitions of FUS by chaperoning its folded RNA-binding domain. *Elife* **10**, e69377 (2021).
25 Lu, S., Hu, J., Arogundade, O. A., Goginashvili, A., Vazquez-Sanchez, S., Diedrich, J. K., Gu, J., Blum, J., Oung, S. & Ye, Q. Heat-shock chaperone HSPB1 regulates cytoplasmic TDP-43 phase separation and liquid-to-gel transition. *Nature cell biology* **24**, 1378-1393 (2022).
26 Liu, Q., Liang, C. & Zhou, L. Structural and functional analysis of the Hsp70/Hsp40 chaperone system. *Protein Science* **29**, 378-390 (2020).
27 Cox, D., Whiten, D. R., Brown, J. W., Horrocks, M. H., San Gil, R., Dobson, C. M., Klenerman, D., Van Oijen, A. M. & Ecroyd, H. The small heat shock protein Hsp27 binds α-synuclein fibrils, preventing elongation and cytotoxicity. *Journal of Biological Chemistry* **293**, 4486-4497 (2018).
28 Jehle, S., van Rossum, B., Stout, J. R., Noguchi, S. M., Falber, K., Rehbein, K., Oschkinat, H., Klevit, R. E. & Rajagopal, P. αB-crystallin: a hybrid solid-state/solution-state NMR investigation reveals structural aspects of the heterogeneous oligomer. *Journal of molecular biology* **385**, 1481-1497 (2009).
29 Jehle, S., Rajagopal, P., Bardiaux, B., Markovic, S., Kühne, R., Stout, J. R., Higman, V. A., Klevit, R. E., van Rossum, B.-J. & Oschkinat, H. Solid-state NMR and SAXS studies

provide a structural basis for the activation of αB-crystallin oligomers. *Nature structural & molecular biology* **17**, 1037-1042 (2010).

30 Jehle, S., Vollmar, B. S., Bardiaux, B., Dove, K. K., Rajagopal, P., Gonen, T., Oschkinat, H. & Klevit, R. E. N-terminal domain of αB-crystallin provides a conformational switch for multimerization and structural heterogeneity. *Proceedings of the National Academy of Sciences* **108**, 6409-6414 (2011).

31 Morris, G. A. & Freeman, R. Enhancement of nuclear magnetic resonance signals by polarization transfer. *Journal of the American Chemical Society* **101**, 760-762 (1979).

32 Pines, A., Gibby, M. G. & Waugh, J. Proton-enhanced NMR of dilute spins in solids. *The Journal of chemical physics* **59**, 569-590 (1973).

33 Monahan, Z., Ryan, V. H., Janke, A. M., Burke, K. A., Rhoads, S. N., Zerze, G. H., O'Meally, R., Dignon, G. L., Conicella, A. E. & Zheng, W. Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity. *The EMBO journal* **36**, 2951-2967 (2017).

34 Guerrero-Ferreira, R., Taylor, N. M., Mona, D., Ringler, P., Lauer, M. E., Riek, R., Britschgi, M. & Stahlberg, H. Cryo-EM structure of alpha-synuclein fibrils. *elife* **7**, e36402 (2018).

35 Lee, W., Rahimi, M., Lee, Y. & Chiu, A. POKY: a software suite for multidimensional NMR and 3D structure calculation of biomolecules. *Bioinformatics* **37**, 3041-3042 (2021).

36 Helmus, J. J. & Jaroniec, C. P. Nmrglue: an open source Python package for the analysis of multidimensional NMR data. *Journal of biomolecular NMR* **55**, 355-367 (2013).

37 Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in science & engineering* **9**, 90-95 (2007).

38 Lee, W., Yu, W., Kim, S., Chang, I., Lee, W. & Markley, J. L. PACSY, a relational database management system for protein structure and chemical shift analysis. *Journal of biomolecular NMR* **54**, 169-179 (2012).

39 Haslbeck, M., Franzmann, T., Weinfurtner, D. & Buchner, J. Some like it hot: the structure and function of small heat-shock proteins. *Nature structural & molecular biology* **12**, 842-846 (2005).

40 Mainz, A., Jehle, S., van Rossum, B. J., Oschkinat, H. & Reif, B. Large protein complexes with extreme rotational correlation times investigated in solution by magic-angle-spinning NMR spectroscopy. *Journal of the American Chemical Society* **131**, 15968-15969 (2009).

41 Mainz, A., Religa, T. L., Sprangers, R., Linser, R., Kay, L. E. & Reif, B. NMR spectroscopy of soluble protein complexes at one mega-dalton and beyond. *Angewandte Chemie International Edition* **52**, 8746-8751 (2013).

42 Kaiser, C. J., Peters, C., Schmid, P. W., Stavropoulou, M., Zou, J., Dahiya, V., Mymrikov, E. V., Rockel, B., Asami, S. & Haslbeck, M. The structure and oxidation of the eye lens chaperone αA-crystallin. *Nature structural & molecular biology* **26**, 1141-1150 (2019).

43 Stevens, A. J., Brown, Z. Z., Shah, N. H., Sekar, G., Cowburn, D. & Muir, T. W. Design of a split intein with exceptional protein splicing activity. *Journal of the American Chemical Society* **138**, 2162-2165 (2016).

44 Stevens, A. J., Sekar, G., Shah, N. H., Mostafavi, A. Z., Cowburn, D. & Muir, T. W. A promiscuous split intein with expanded protein engineering applications. *Proceedings of the National Academy of Sciences* **114**, 8538-8543 (2017).

45 Yu, C., Leung, S. K. P., Zhang, W., Lai, L. T. F., Chan, Y. K., Wong, M. C., Benlekbir, S., Cui, Y., Jiang, L. & Lau, W. C. Y. Structural basis of substrate recognition and thermal protection by a small heat shock protein. *Nature communications* **12**, 3007 (2021).
46 Kulig, M. & Ecroyd, H. The small heat-shock protein αB-crystallin uses different mechanisms of chaperone action to prevent the amorphous versus fibrillar aggregation of α-lactalbumin. *Biochemical Journal* **448**, 343-352 (2012).

Table 4.S1. Parameters for MAS NMR experiments. Experiments were performed at a MAS frequency of 11.11 kHz at 275 K unless otherwise indicated. All experiments were performed at a magnetic field strength corresponding to a 750 MHz ^{1}H Larmor frequency using a 3.2 mm E^{free} HCN probe. Note that multiple experiments were acquired for HSPB1, and the HN-INEPT, HC-INEPT, NCA, and INEPT-TOBSY were performed on a sample containing 20 mg HSPB1.

Experiment	HSPB1		HSPB1 + FUS LC		HSPB1 + FUS LC LLPS	
Sample Amount	30	mg	15 / 15	mg	<1 / 10	mg
CP						
Number of scans	8	scans	64	scans	10240	
Recycle delay	5	s	5	s	5	s
^{13}C π/2	50	kHz	50	kHz	50	kHz
^{1}H π/2	80.6	kHz	80.6	kHz	80.6	kHz
CP contact time	1000	µs	1000	µs	1000	µs
Decoupling program	SPINAL64		SPINAL64		SPINAL64	
	74.6	kHz	74.6	kHz	74.6	kHz
Decoupling power	160	W	160	W	160	W
Acquisition time	9.9	ms	9.9	ms	9.9	ms
INEPT						
Number of scans	64	scans	64	scans	10240	
Recycle delay	4	s	4	s	4	s
^{13}C π/2	50	kHz	50	kHz	50	kHz
^{1}H π/2	80.6	kHz	80.6	kHz	80.6	kHz
Refocusing parameter	140	Hz	140	Hz	140	Hz
Decoupling program	SPINAL64		SPINAL64		SPINAL64	
	74.6	kHz	74.6	kHz	74.6	kHz
Decoupling power	160	W	160	W	160	W
Acquisition time	9	ms	9	ms	9	ms
DP						
Number of scans	4	scans	64	scans	10240	
Recycle delay	5		5		5	
^{13}C π/2	50	kHz	50	kHz	50	kHz
Decoupling program	SPINAL64		SPINAL64		SPINAL64	
	87.9	kHz	87.9	kHz	87.8	kHz
Decoupling power	160	W	160	W	160	W
Acquisition time	10.2	ms	10.2	ms	10.2	ms
CP-DARR						
Number of scans	64	scans	64	scans	384	
Recycle delay	5	s	5	s	5	s
^{13}C π/2	50	kHz	50	kHz	50	kHz
^{1}H π/2	80.6	kHz	80.6	kHz	80.6	kHz
CP contact time	400	µs	400	µs	400	µs
Decoupling program	SPINAL64		SPINAL64		SPINAL64	

Table 4.S1. Parameters for MAS NMR experiments. Experiments were performed at a MAS frequency of 11.11 kHz at 275 K unless otherwise indicated. All experiments were performed at a magnetic field strength corresponding to a 750 MHz ^{1}H Larmor frequency using a 3.2 mm E^{free} HCN probe. Note that multiple experiments were acquired for HSPB1, and the HN-INEPT, HC-INEPT, NCA, and INEPT-TOBSY were performed on a sample containing 20 mg HSPB1.

	74.6	kHz	74.6	kHz	74.6	kHz
Decoupling power	160	W	160	W	160	W
DARR mixing time	20	ms	20	ms	20	ms
^{1}H power during DARR sequence	11.11	kHz	11.11	kHz	11.11	kHz
t1 acquisition	3.2	ms	3.2	ms	3.2	ms
t1 points	256		256		256	
HN-INEPT						
Number of scans	160	scans				
Recycle delay	3	s				
^{15}N π/2	42	kHz				
^{1}H π/2	79.4	kHz				
Refocusing parameter	92	Hz				
Decoupling program	SWf-TPPM					
	100	kHz				
Decoupling power	160	W				
t1 acquisition	13	ms				
t1 points	320					
HC-INEPT	227		42			
Number of scans	64	scans	64			
Recycle delay	5	s	5	s		
^{13}C π/2	50	kHz	50	kHz		
^{1}H π/2	80.6	kHz	80.6	kHz		
Refocusing parameter	140	Hz	140	Hz		
Decoupling program	SWf-TPPM		SWf-TPPM			
	74.6	kHz	74.6	kHz		
Decoupling power	160	W	160	W		
t1 acquisition	33.3	ms	33.3			
t1 points	500		500			
INEPT-TOBSY	331					
Number of scans	64	scans				
Recycle delay	4					
^{13}C π/2	50	kHz				
^{1}H π/2	83.3	kHz				

Table 4.S1. Parameters for MAS NMR experiments. Experiments were performed at a MAS frequency of 11.11 kHz at 275 K unless otherwise indicated. All experiments were performed at a magnetic field strength corresponding to a 750 MHz ^{1}H Larmor frequency using a 3.2 mm E^{free} HCN probe. Note that multiple experiments were acquired for HSPB1, and the HN-INEPT, HC-INEPT, NCA, and INEPT-TOBSY were performed on a sample containing 20 mg HSPB1.

Refocusing parameter	140	Hz
Decoupling program	SWf-TPPM	
	98	kHz
Decoupling power	160	W
t1 acquisition	9	ms
t1 points	720	

Table 4.S1. Parameters for MAS NMR experiments. Experiments were performed at a MAS frequency of 11.11 kHz at 275 K unless otherwise indicated. All experiments were performed at a magnetic field strength corresponding to a 750 MHz ^{1}H Larmor frequency using a 3.2 mm E^{free} HCN probe. Note that multiple experiments were acquired for HSPB1, and the HN-INEPT, HC-INEPT, NCA, and INEPT-TOBSY were performed on a sample containing 20 mg HSPB1.

Experiment	HSPB1 + α-Synuclein fibrils		α-Synuclein		
Sample Amount	3 / 20	mg	20	mg	
CP	21				
Number of scans	10240		20480		
Recycle delay	5	s	5	s	
^{13}C π/2	50	kHz	50	kHz	50
^{1}H π/2	80.6	kHz	80.6	kHz	80.6
CP contact time	1000	µs	1000	µs	
Decoupling program	SPINAL64		SPINAL64		
	80.6	kHz	80.6	kHz	
Decoupling power	160	W	160	W	
Acquisition time	9.9	ms	9.9	ms	
INEPT	33				
Number of scans	10240		20480		
Recycle delay	4	s	4	S	
^{13}C π/2	50	kHz	50	kHz	
^{1}H π/2	80.6	kHz	80.6	kHz	
Refocusing parameter	140	Hz	140	Hz	
Decoupling program	SPINAL64		SPINAL64		
	80.6	kHz	80.6	kHz	
Decoupling power	160	W	160	W	
Acquisition time	9	ms	9	ms	
DP	20				
Number of scans	10240		20480		
Recycle delay	5		5		
^{13}C π/2	50	kHz	50	kHz	
Decoupling program	SPINAL64		SPINAL64		
	87.8	kHz	87.8	kHz	
Decoupling power	160	W	160	W	
Acquisition time	10.2	ms	10.2	ms	
CP-DARR	30				
Number of scans	384				
Recycle delay	5	s			
^{13}C π/2	50	kHz			
^{1}H π/2	80.6	kHz			
CP contact time	400	µs			

Table 4.S1. Parameters for MAS NMR experiments. Experiments were performed at a MAS frequency of 11.11 kHz at 275 K unless otherwise indicated. All experiments were performed at a magnetic field strength corresponding to a 750 MHz ^{1}H Larmor frequency using a 3.2 mm E^{free} HCN probe. Note that multiple experiments were acquired for HSPB1, and the HN-INEPT, HC-INEPT, NCA, and INEPT-TOBSY were performed on a sample containing 20 mg HSPB1.

Decoupling program	SPINAL64	
Decoupling power	74.6	kHz
	160	W
DARR mixing time	20	ms
^{1}H power during DARR sequence	11.11	kHz
t1 acquisition	3.2	ms
t1 points	256	

A

```
1   MTERR VPFSL LRGPS WDPFR DWYPH SRLFD QAFGL PRLPE EWSQW LGGSS
51  WPGYV RPLPP AAIES PAVAA PAYSR ALSRQ LSSGV SEIRH TADRW RVSLD
101 VNHFA PDELT VKTKD GVVEI TGKHE ERQDE HGYIS RCFTR KYTLP PGVDP
151 TQVSS SLSPE GTLTV EAPMP KLATQ SNEIT IPVTF ESRAQ LGGPE AAKSD
201 ETAAK *
```

B

```
1   MKSSH HHHHH VKIIS RKSLG TQNVY DIGVG EPHNF LLKNG LVASN CGVSE
51  IRHTA DRWRV SLDVN HFAPD ELTVK TKDGV VEITG KHEER QDEHG YISRC
101 FSRKY TLPPG VDPTQ VSSSL SPEGT LTVEA PMPKL ATQSN EITIP VTFES
151 RAQLG GPEAA KSDET AAK*
```

C

```
1   MGSDK IHHHH HHSSG TKIEE GKLVI WINGD KGYNG LAEVG KKFEK DTGIK
51  VTVEH PDKLE EKFPQ VAATG DGPDI IFWAH DRFGG YAQSG LLAEI TPDKA
101 FQDKL YPFTW DAVRY NGKLI AYPIA VEALS LIYNK DLLPN PPKTW EEIPA
151 LDKEL KAKGK SALMF NLQEP YFTWP LIAAD GGYAF KYENG KYDIK DVGVD
201 NAGAK AGLTF LVDLI KNKHM NADTD YSIAE AAFNK GETAM TINGP WAWSN
251 IDTSK VNYGV TVLPT FKGQP SKPFV GVLSA GINAA SPNKE LAKEF LENYL
301 LTDEG LEAVN KDKPL GAVAL KSYEE ELAKD PRIAA TMENA QKGEI MPNIP
351 QMSAF WYAVR TAVIN AASGR QTVDE ALKDA QTNSG SDITS LYKKA EGGTE
401 NLYFQ GHTER RVPFS LLRGP SWDPF RDWYP HSRLF DQAFG LPRLP EEWSQ
451 WLGGS SWPGY VRPLP PAAIE SPAVA APAYS RALSR QLSCL SYDTE ILTVE
501 YGFLP IGKIV EERIE CTVYT VDKNG FVYTQ PIAQW HNRGE QEVFE YCLED
551 GSIIR ATKDH KFMTT DGQML PIDEI FERGL DLKQV DGLPH HHHHH *
```

Figure 4.S1: Amino acid sequences of the constructs used in this work. A) Sequence for HSPB1. B) Sequence for 6xHis-Cfa^{C}-HSPB1$_{ACD-CTD}$. Cfa^{N} is pink, the ACD is orange, and the CTD is purple. His tag and linker sequences are gray. **C) The amino acid sequence of 6xHis-MBP-TEV-HSPB1$_{NTD}$-Cfa^{N}-6xHis.** MBP is gold, the TEV cleavage sequence is blue, the NTD is green, and Cfa^{C} is brown. His tag and linker sequences are gray. **D) Sequence for the segmentally labeled construct of HSPB1.** The TEV cleavage scar is blue, the NTD is green, the residual Cfa cysteine is pink, the ACD is orange, and the CTD is purple. The ^{15}N, ^{13}C-labeled residues are highlighted in yellow. **E) Sequence for α-synuclein.**

D

```
1   GHTER RVPFS LLRGP SWDPF RDWYP HSRLF DQAFG LPRLP EEWSQ WLGGS
51  SWPGY VRPLP PAAIE SPAVA APAYS RALSR QLSCG VSEIR HTADR WRVSL
101 DVNHF APDEL TVKTK DGVVE ITGKH EERQD EHGYI SRCFT RKYTL PPGVD
151 PTQVS SSLSP EGTLT VEAPM PKLAT QSNEI TIPVT FESRA QLGGP EAAKS
201 DETAA K*
```

E

```
1   MDVFM KGLSK AKEGV VAAAE KTKQG VAEAA GKTKE GVLYV GSKTK EGVVH
51  GVATV AEKTK EQVTN VGGAV VTGVT AVAQK TVEGA GSIAA ATGFV KKDQL
101 GKNEE GAPQE GILED MPVDP DNEAY EMPSE EGYQD YEPEA *
```

Figure 4.S1: Amino acid sequences of the constructs used in this work. A) Sequence for HSPB1. B) Sequence for 6xHis-Cfa^{C}-HSPB1$_{ACD-CTD}$. Cfa^{N} is pink, the ACD is orange, and the CTD is purple. His tag and linker sequences are gray. **C) The amino acid sequence of 6xHis-MBP-TEV-HSPB1$_{NTD}$-Cfa^{N}-6xHis.** MBP is gold, the TEV cleavage sequence is blue, the NTD is green, and Cfa^{C} is brown. His tag and linker sequences are gray. **D) Sequence for the segmentally labeled construct of HSPB1.** The TEV cleavage scar is blue, the NTD is green, the residual Cfa cysteine is pink, the ACD is orange, and the CTD is purple. The ^{15}N, ^{13}C-labeled residues are highlighted in yellow. **E) Sequence for α-synuclein.**

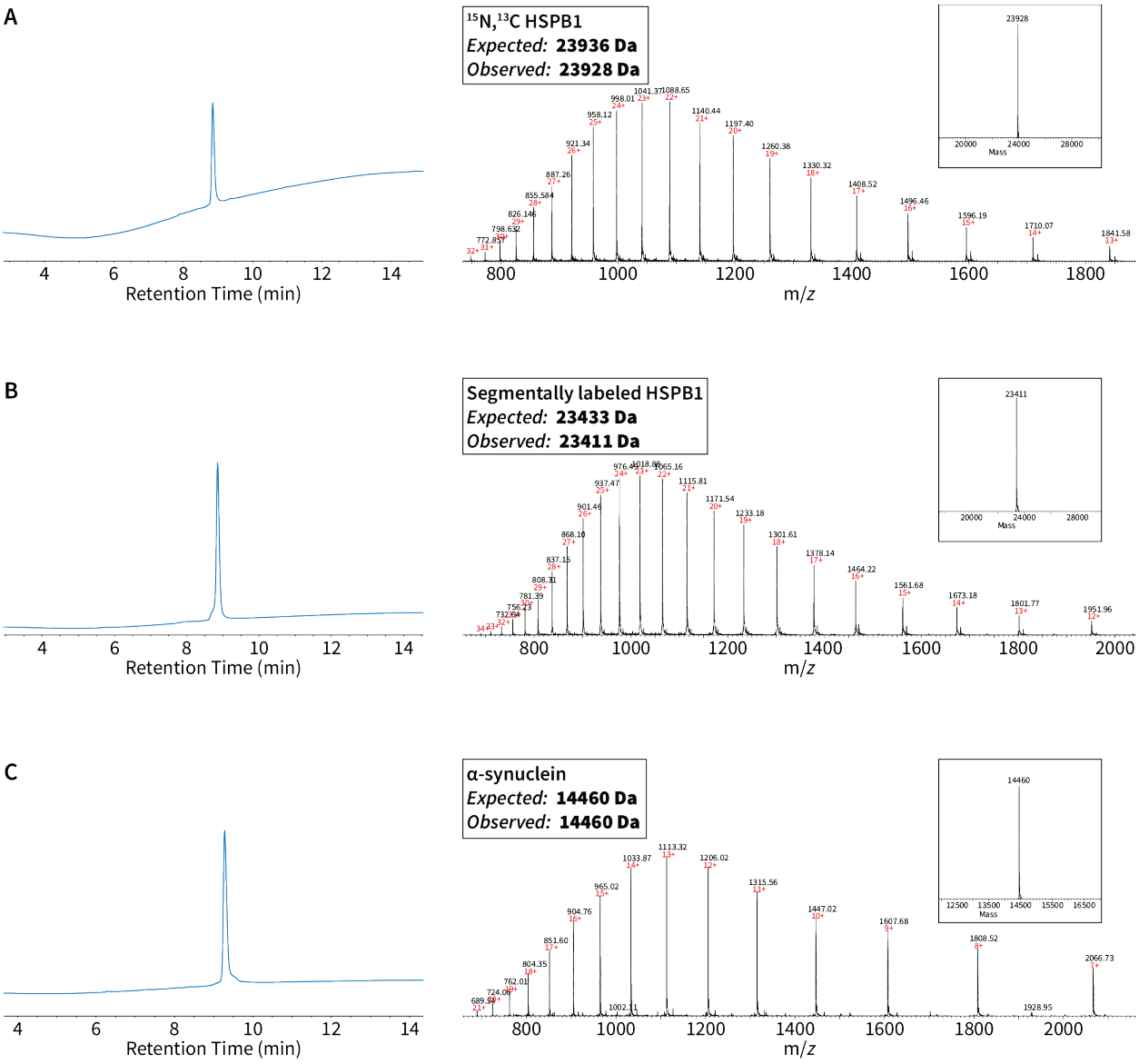


Figure 4.S2. HPLC-MS data for constructs used in this work. A) HPLC-MS data for uniformly ^{15}N, ^{13}C-labeled HSPB1. B) HPLC-MS data for segmentally labeled HSPB1. Note that this construct contains a ^{15}N, ^{13}C-labeled NTD. **C) HPLC-MS data for α-synuclein.**

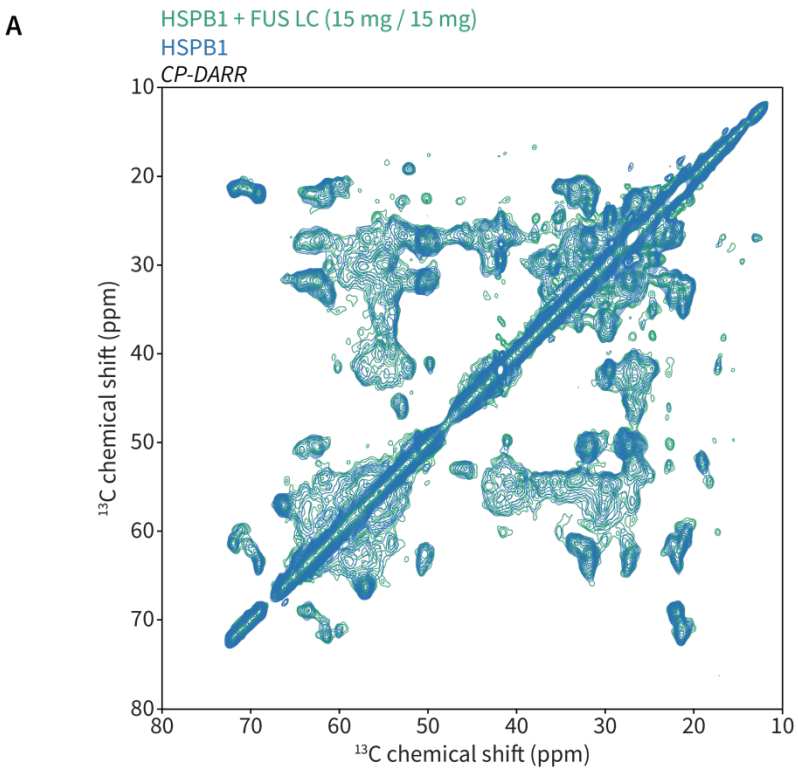


Figure 4.S3. Characterization of a low ratio sample of HSPB1 and FUS LC. A) A ^{13}C-^{13}C CP-DARR experiment of HSPB1 in blue overlaid with the same experiment performed on a sample containing approximately 15 mg HSPB1 and 15 mg FUS LC in green.

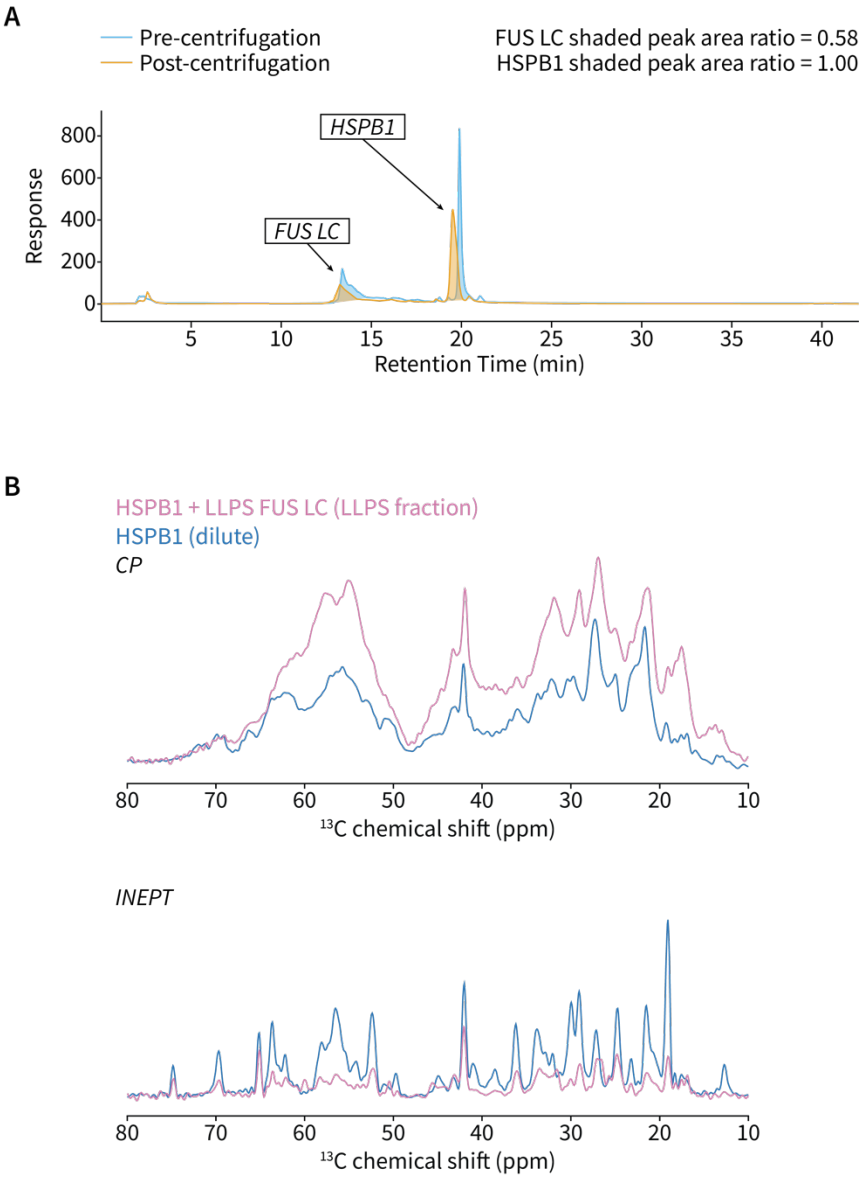


Figure 4.S4. Quantitation of HSPB1 in FUS LC LLPS droplets. A) Analytical HPLC traces of HSPB1 and LLPS FUS LC before rotor packing and the supernatant of the same sample after rotor packing by centrifugation. Peaks representing FUS LC and HSPB1 are shaded and integrated. **B)** 1D ^{13}C CP and INEPT spectra of a dilute sample containing 400 µg of HSPB1 in blue overlaid with the spectra from the HSPB1 and FUS LC LLPS copartitioned sample. Experimental parameters for the CP and INEPT experiments are identical between the two samples.

ACKNOWLEDGEMENTS

Chapter 4, in part, is currently being prepared for submission for publication. Berkeley, R. F., Plonski, A. P., Cook, B. D., McGuire, K. L., Herzik, M. A., & Debelouchina, G. T. The dissertation author was the primary investigator and author of this material.

Chemical Science

ROYAL SOCIETY OF CHEMISTRY

PERSPECTIVE

Cite this: *Chem. Sci.*, 2022, **13**, 14226

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 3rd September 2022
Accepted 21st November 2022

DOI: 10.1039/d2sc04907d

rsc.li/chemical-science

Chemical tools for study and modulation of biomolecular phase transitions

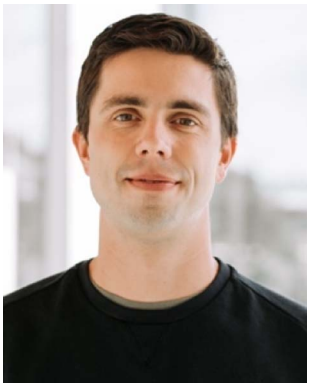
Raymond F. Berkeley and Galia T. Debelouchina *

Biomolecular phase transitions play an important role in organizing cellular processes in space and time. Methods and tools for studying these transitions, and the intrinsically disordered proteins (IDPs) that often drive them, are typically less developed than tools for studying their folded protein counterparts. In this perspective, we assess the current landscape of chemical tools for studying IDPs, with a specific focus on protein liquid–liquid phase separation (LLPS). We highlight methodologies that enable imaging and spectroscopic studies of these systems, including site-specific labeling with small molecules and the diverse range of capabilities offered by inteins and protein semisynthesis. We discuss strategies for introducing post-translational modifications that are central to IDP and LLPS function and regulation. We also investigate the nascent field of noncovalent small-molecule modulators of LLPS. We hope that this review of the state-of-the-art in chemical tools for interrogating IDPs and LLPS, along with an associated perspective on areas of unmet need, can serve as a valuable and timely resource for these rapidly expanding fields of study.

Introduction

Intrinsically disordered proteins and intrinsically disordered protein regions (henceforth collectively referred to as IDPs) are common in the proteome.[1,2] All IDPs share a relatively flat folding energy landscape, and often completely lack a detectable folded low-energy state.[3] Unlike proteins constructed from random amino acid sequences, which are also often intrinsically disordered, proteomic IDPs typically exhibit low sequence complexity with primary sequences enriched in a small number of amino acids.[4–6] Despite their unfolded or conditionally folded nature, IDPs can act as important regulators of cellular functions, often through low-affinity interactions that are driven by the bulk chemical properties of residues in the IDP.[2] Many IDPs can also undergo a process called liquid–liquid phase separation (LLPS) that allows them to concentrate into dynamic and spatially resolved condensates.[7] This capability allows IDPs to

Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA, USA. E-mail: gdebelouchina@ucsd.edu



Ray Berkeley is a PhD candidate in the Chemical Biology program at UC San Diego working under the supervision of Professor Galia Debelouchina. Ray's work focuses on using solid-state NMR to understand the molecular interactions and structural features that drive protein phase transitions. He is also interested in chemical tools that modulate intrinsically disordered proteins and protein liquid–liquid phase separation. Prior to UCSD, Ray earned his B.S. in Molecular & Cell Developmental Biology with a minor in Bioinformatics from UC Santa Cruz, where he worked in Scott Lokey's laboratory studying yeast chemical genetics and the properties of cyclic peptide macrocycles.

Dr Galia Debelouchina obtained a PhD in Physical Chemistry in the lab of Prof. Robert Griffin at the Massachusetts Institute of Technology. There, she used solid-state NMR spectroscopy to study amyloid fibril structure. She then moved to Princeton University and the lab of Prof. Tom Muir where she used chemical biology and biophysics tools to interrogate chromatin structure and dynamics. She is currently an Assistant Professor in the Department of Chemistry and Biochemistry at UC San Diego. Her group combines NMR spectroscopy with chemical biology approaches to investigate the molecular interactions that drive protein phase transitions.

rapidly and selectively recruit other biomolecules and entrenches their role as regulators of intracellular organization, with many common IDPs such as p53, FUS, and α-synuclein exhibiting pleiotropic regulatory behaviors that are essential for cell viability.[8,9]

IDPs present unique challenges to structural biologists – their conformational plasticity and the range of dynamics that they can exhibit are difficult to characterize with tools that have been optimized for folded proteins. While early structural studies of IDPs often focused on proteins that fold upon binding, this behavior is not universal. For example, some IDPs engage in high affinity interactions while retaining their disorder, while others engage in nonspecific low affinity interactions that drive biomolecular LLPS and are dependent on their intrinsically disordered nature.[10–12] In addition to the conceptual challenges brought on by the unstructured nature of these proteins, the sensitivity of IDPs to environmental conditions requires special consideration when working with them *in vitro*. Characterizing the structure–function relationship of IDPs, therefore, necessitates a shift in the way that we think about protein structure and a coincident shift in the methods and tools that we use to interrogate it, especially in the context of LLPS.[13,14]

In this perspective, we discuss how chemical biology can aid structural studies of IDPs, with an emphasis on chemical tools that are compatible with LLPS. This will include strategies for the recombinant preparation of IDPs, tools that enable the efficient and site-selective introduction of chemical probes and isotopic labels, and chemical modulators of IDPs and LLPS. We take an application-centric approach to highlight real use cases that are enabled by the current state-of-the-art in IDP chemical biology. We hope to provide insight into best practices for handling and studying these systems, along with a call for the development of minimally perturbative small-molecule chemical tools to aid their analysis and functional manipulation.

IDPs present distinct challenges for biophysical investigations *in vitro*

When working with IDPs *in vitro*, their intrinsic disorder brings considerations that are not present for folded proteins. IDPs are often less soluble than their globular counterparts near their isoelectric points due to the lack of a distinct fold, and they are often prone to LLPS or aggregation under these conditions due to nonspecific interactions between hydrophobic residues.[15] On the other hand, the lack of a need to maintain a fold can facilitate IDP survival at extreme pH levels and temperatures that would lead to denaturation and aggregation of folded proteins, which can enable unique purification and manipulation strategies. The conditions under which IDPs are handled must therefore be chosen with these considerations in mind. In this section, we will discuss challenges associated with working with IDPs *in vitro*, along with approaches for purification that are distinct from those of typically folded proteins. We will highlight examples of IDP preparations that strategically employ pH, temperature, and other abiotic factors in ways that would not be applicable for folded proteins. This section will also provide some context for the challenges that need to be overcome when developing chemical tools to aid the structural and functional analysis of IDPs. For a more comprehensive overview of strategies and recommendations for purifying IDPs, the interested reader is referred to two excellent resources by Graether and Alberti *et al.*[16,17]

Whether expressed in *E. coli* or other recombinant expression systems, one of the challenges in handling IDPs is avoiding undesirable aggregation or LLPS during the purification process. Fusing solubility tags such as MBP or GST to an IDP of interest can be a useful strategy for maintaining its solubility.[18,19] As these solubility tags are large and may interfere with downstream functional and biophysical assays, they are typically removed at the end of the purification process through a suitably engineered cleavable site. This requires an additional purification step to remove the cleaved tag. Cleavage of a solubility tag after purification can be a facile approach for performing LLPS studies.[20] In a compelling recent example of this approach, Morin *et al.* use an MBP-Klf4 fusion protein to construct a model describing the role of prewetting in the sequence-specific surface condensation of the transcription factor Klf4, which forms small LLPS condensates on DNA. In this work, the MBP-Klf4 fusion is capable of adsorbing onto DNA, but does not undergo LLPS. After adsorption, the MBP tag is removed and Klf4 condenses into droplets around sequences known to promote Klf4 binding.[21] This strategy allows for the disambiguation of Klf4 adsorption and LLPS. In this case, a solubility tag is used to enable an LLPS study in an environment that would be sensitive to other means of LLPS initiation, such as a pH jump or changing the IDP or salt concentration.

Since IDPs are often enriched in hydrophilic and charged residues, tuning the pH can be a straightforward way to improve the performance of a purification protocol. Although there are reports that leverage pH in order to purposefully precipitate the IDP of interest,[22] in most cases the selection of a buffer that optimizes solubility is preferred. Buffers that maintain unusually high or low pH conditions are often advantageous, and the disordered nature of the IDP means that pH-induced denaturation is not an issue. As an example, some purification approaches for the low-complexity domain of Fused in Sarcoma (FUS LC), which undergoes LLPS and/or aggregation at neutral pH, involve the extensive use of *N*-cyclohexyl-3-aminopropanesulfonic acid (CAPS), a buffer that maintains a pH of 10–11. At such high pH, the tyrosine residues on FUS are deprotonated and the protein is highly charged, which promotes solubility and enables the subsequent purification of the LC domain by size-exclusion chromatography without the need for a denaturant.[18,23–25] On the other hand, low-pH conditions are often used to purify α-synuclein.[26] This first step in the purification process precipitates many undesired cellular proteins, thereby leveraging α-synuclein's resistance to pH-dependent denaturation. A potential disadvantage of pH-based purification protocols is that they may not be compatible with chemical biology approaches for protein labeling or modification, as the reactions are typically sensitive to pH.

Temperature is another practical consideration that may require a different approach than that taken for a typical globular protein. Since there is no fold to maintain, some IDPs are resistant to high temperatures, a unique property that can be exploited with high-temperature protein purification approaches. Tau and α-synuclein, for example, are often purified by boiling crude cell lysate, which denatures and precipitates most cellular proteins and leaves a soluble fraction that is highly enriched in the desired protein.[26] On the other hand, many IDPs undergo thermoresponsive phase transitions which may require the use of mild temperatures around 25 °C throughout the protein preparation,[7] a characteristic that is counter-intuitive to those who are used to maintaining temperatures closer to 4 °C throughout the purification of a globular protein. Tolerance to higher temperatures may be beneficial for chemical labeling approaches of IDPs as it can speed up the relevant reactions.

For IDPs that are especially aggregation-prone, chemical denaturants are often necessary to achieve reasonable yields from a recombinant protein preparation. In many IDP preparations, chaotropes such as urea or guanidinium hydrochloride can be used to redissolve aggregated protein or to maintain solubility at a pH or temperature that would otherwise induce aggregation or LLPS.[27] Chemical denaturants may also be required to keep IDPs soluble during size-exclusion or ion-exchange chromatography purification steps. In some cases, the IDP will be sequestered into inclusion bodies during expression, and urea or guanidinium hydrochloride may be required for extracting proteins from the inclusion body and for mitigating the risk of aggregation in subsequent steps.[28] Chemical denaturants are often compatible with cysteine chemistry and even intein-based segmental labeling approaches, and can therefore be useful in the preparation of modified and labeled IDPs.[29]

Introduction of chemical probes for imaging and spectroscopic studies

Understanding the structure, dynamics, interactions, and functions of IDPs often relies on fluorescence-based approaches such as fluorescence recovery after photobleaching (FRAP) and fluorescence resonance energy transfer (FRET), and/or spectroscopic studies by nuclear magnetic resonance (NMR) and electron paramagnetic resonance (EPR).[30–32] A common requirement for these techniques is the site-specific installation of chemical probes that report on the properties of the IDP or its environment. A wide range of technologies for site-specific labeling have been demonstrated, and many of these approaches are applicable to IDPs.[33–35] Here we review the applications of cysteine chemistry, unnatural amino acid incorporation through genetic means (amber suppression), inteins, and sortase, with a particular emphasis on the unique challenges presented by IDPs and LLPS. We focus primarily on the installation of small chemical probes such as fluorescent labels, rather than the use of large fusion fluorescent proteins, as these approaches may be less familiar to the reader and may have distinct advantages in certain applications.

Cysteine is the chemical handle of choice for facile labeling of IDPs

A wide range of small-molecule chemical tools are available for the site-selective modification of proteins under aqueous conditions.[36,37] Of these, the most robust and practical reactions take place at nucleophilic cysteine residues. Cysteines are particularly rare in IDPs and are therefore valuable reactive handles that can be targeted selectively if present or added into the recombinantly-produced proteins if needed. The potent nucleophilicity of the cysteine thiol allows for chemoselectivity even in the presence of other nucleophiles such as the primary amines on lysines and protein N-termini.

Covalent labeling at cysteine residues is a popular strategy for introducing small-molecule fluorescent probes in proteins including IDPs (Fig. 1A). This is most often achieved through maleimide or iodoacetamide functionalized dyes and there are many commercially available options covering a wide range of absorbance and emission properties, including cyanine-based probes, the Alexa Fluor® series, and the BODIPY family of dyes.[38] Labelling at a cysteine residue with a maleimide-based probe is often as simple as incubating the dye with the protein for less than an hour and subsequently removing the unreacted dye *via* gel filtration or reverse-phase chromatography.[39–41] In the case of IDPs, it may be beneficial to perform the labeling step while the protein is still fused to the solubility tag or under denaturing conditions to avoid issues with aggregation or premature LLPS. The straightforward and robust nature of cysteine chemistry make this the preferred method of choice for labeling IDPs, especially if the purification protocol is challenging and delivers relatively low yields. Sometimes, it may also be possible to use the primary amines of lysine residues to attach fluorescent probes through *N*-hydroxysuccinimide ester chemistry.[42] However, there are typically many more lysine residues in a protein compared to cysteine and it is difficult to control the specificity of these reactions, especially in the context of IDPs where all side-chains are solvent exposed and accessible. The unique reactivity of the α-amine of a protein's N-terminus can also be exploited to attach imaging or spectroscopic probes through chemical or enzymatic means. Interested readers are referred to the comprehensive review of these methods by Rosen & Francis.[43]

In addition to fluorescent labeling, cysteine is a useful handle for introducing spin labels for NMR- and EPR-based methods. A common NMR approach for IDP studies is the paramagnetic relaxation enhancement (PRE) experiment.[44] PREs are particularly powerful in detecting weak intra- and inter-molecular interactions at residue-specific resolution in dynamic biological systems and are often applied to characterize the molecular basis of LLPS.[18,23,24,45,46] In a PRE experiment, the protein of interest is labeled with a paramagnetic relaxation enhancement probe, typically a nitroxide-based stable radical moiety such as *S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1*H*-pyrrol-3-yl)methyl methanesulfonothioate

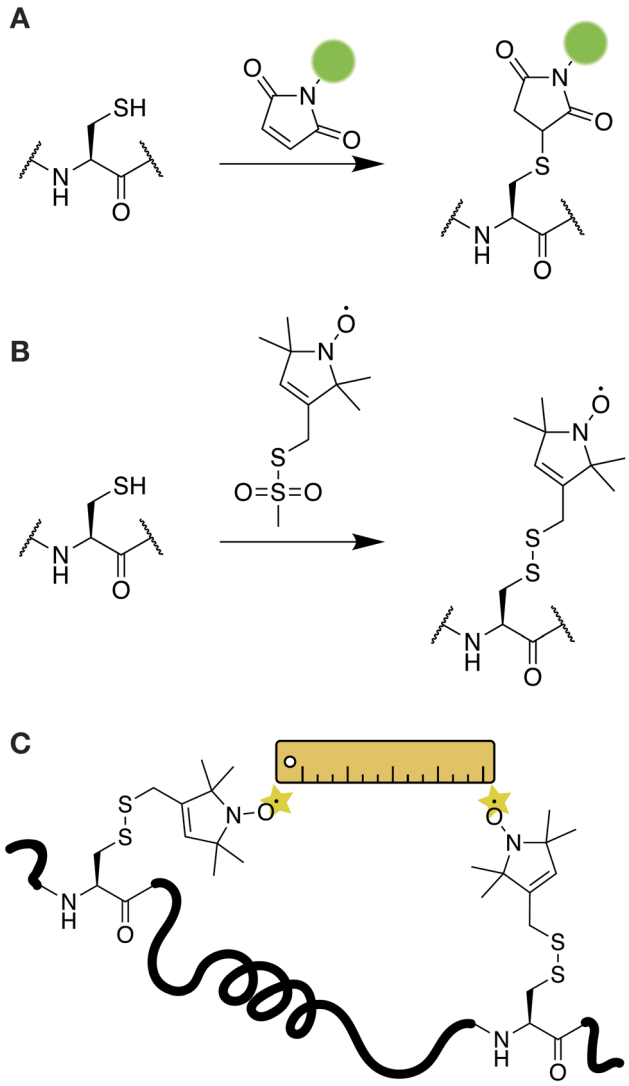


Fig. 1 Cysteine approaches for introducing imaging and spectroscopic probes in IDPs. (A) Alkylation reactions with maleimides are often used to attach fluorescent probes. (B) The EPR probe MTSL can be introduced through a disulfide oxidation reaction. (C) The distance between two MTSL probes can be measured through an experiment called double electron–electron resonance (DEER).

(MTSL).[44] The addition of MTSL results in an oxidation reaction and the formation of a disulfide bond with the targeted cysteine residue (Fig. 1B). The nitroxide moiety on the probe can induce distance dependent relaxation effects that reduce the peak intensity for residues within a 10–25 Å radius.[13] This information can be used to construct a map of the residues that participate in intra- or intermolecular interactions for IDPs. For example, PRE-based NMR experiments have been used to describe the transient interactions formed by the low complexity domains of FUS, TDP-43, hnRNPA1, and hnRNPA2 in LLPS environments.[23,24,45,47,48]

The MTSL probe can also enable the characterization of IDP behavior by EPR. For example, MTSL-labeled tau was used to characterize tau dynamics within liquid–liquid droplets and to report on tau–water interactions.[42] In this case, the native cysteine residues of tau were mutated to serine, and a new cysteine was introduced at a different position in the protein sequence for labeling so that the relevant interactions could be captured. A commonly used EPR experiment is double electron–electron resonance (DEER), which measures up to 10 nm distances between two electron spin probes.[49] DEER is conceptually similar to FRET experiments and can be used to build a structural model of the protein of interest and to characterize protein–protein interactions (Fig. 1C).[50,51] If intramolecular DEER is performed, then two spin labels need to be introduced in the protein. As DEER can probe distances between two identical spin probes, labeling can be easily achieved by introducing two cysteine residues at the relevant positions in the IDP. For example, in a recent study, DEER of a doubly-labeled construct of the FUS LC domain was used to interrogate the dynamics and conformational distributions of the protein within a phase separated sample.[52] EPR and DEER can also be performed with Gd^{3+}-based spin probes attached through cysteine chemistry.[51,53]

Despite their popularity, some important considerations need to be taken into account when working with cysteine-based labeling approaches. When oxidation-based reactions are used for labeling (*e.g.* with MTSL), the protein cysteines need to be reduced and available before the labeling reaction takes place.[54] Once the disulfide bond between the protein and the probe is formed, care must be taken to avoid reducing conditions or agents in the sample buffer. Even the more chemically resilient maleimide labeling reactions can be susceptible to hydrolysis and thiol exchange, processes that can be exacerbated by higher pH and long storage in aqueous solutions.[55] Cysteine-based reactions are also not bioorthogonal and are therefore not suitable for in-cell applications. Finally, using cysteine labeling approaches, it is difficult to introduce two different probes on the same IDP as may be required for intramolecular FRET experiments. In this case, cysteine chemistry may be combined with unnatural amino acid incorporation by genetic means and bioorthogonal labeling reactions, as discussed below.

Unnatural amino acids provide flexibility for specific and multiple labeling

In cases where labeling at a cysteine is not an option, or a second distinct chemical probe needs to be site-specifically introduced to an IDP of interest, the incorporation of an unnatural amino acid (UAA) can help expand the reactive scope of the target protein.[56,57] The introduction of an entirely unique chemistry into the IDP with a UAA precludes any cross-reactivity with other nucleophilic residues or issues with multiple labeling that come along with the classical thiol-reactive chemical probes described in the previous section. Unnatural amino acids can be introduced by genetic means through a technique often referred to as amber suppression (Fig. 2A).[56,58,59] In this case, the amber stop codon, UAG/TAG, is assigned to the UAA and cloned at the desired position in the protein sequence. At the same time, the cells are transformed or transfected with a second plasmid that encodes an engineered

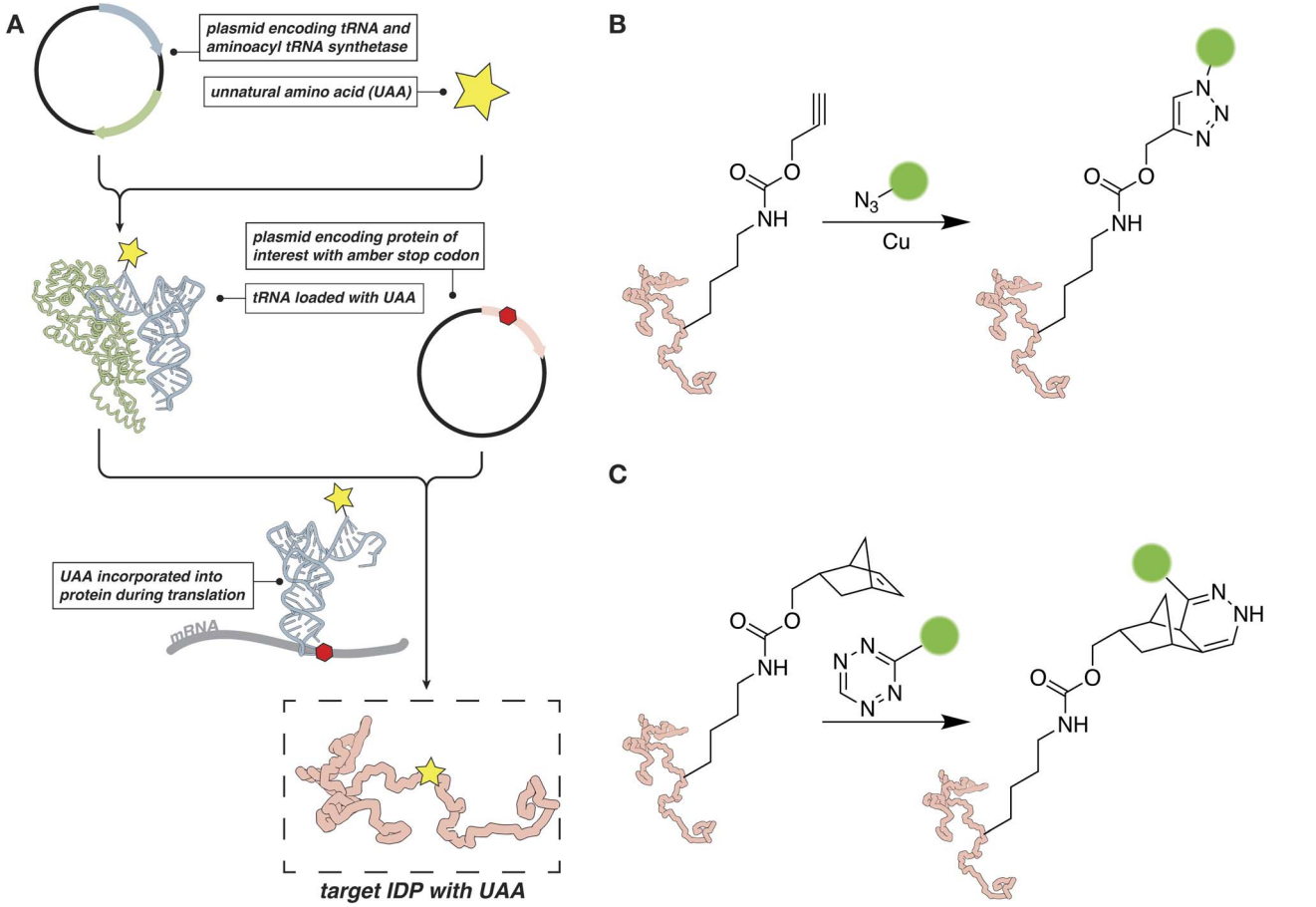


Fig. 2 Amber suppression as a tool for studying IDPs. (A) An overview of the amber suppression strategy. (B) Bioorthogonal CuAAC reaction for protein labeling. (C) Bioorthogonal labeling reaction based on tetrazine–norbornene chemistry.

tRNA/tRNA-synthetase pair. The tRNA synthetase can recognize the UAA and load it onto the cognate tRNA. The loaded tRNA, in turn, recognizes the amber stop codon and delivers the UAA to the ribosome for incorporation into the growing protein sequence.

Amber suppression is often used to introduce a chemical handle for bioorthogonal chemical reactions, such as an azide, tetrazine, strained alkene or alkyne.[60–65] These chemical moieties can then be targeted with a suitable "warhead" carrying the fluorescent or spectroscopic probe of interest. The bioorthogonal reactions can be performed both *in vitro* on the purified protein and in cells. For example, copper-catalyzed [3 + 2] azide–alkyne cycloaddition (CuAAC) reactions have been used to attach EPR probes to proteins in mammalian cells, while tetrazine-based reagents have been used to tether fluorescent probes (Fig. 2B and C).[62,65] If a suitable engineered tRNA/tRNA-synthetase pair has been developed, fluorescent probes, spin labels, or other suitable moieties, can be introduced directly as the UAA.[56–58,66] For example, several reports from Schmidt *et al.*, describe the evolution and subsequent application of a $tRNA^{Pyl}$/pyrrolysyl-tRNA-synthetase pair capable of installing a spin-labeled nitroxide-based amino acid.[67,68] This allows for the direct introduction of a probe without the need for performing chemistry on the target protein after expression, a capability that could prove especially useful for IDPs that are sensitive to abiotic conditions.

From a conceptual point of view, amber suppression is relatively straightforward to implement as it requires only the addition of two plasmids and the UAA before protein expression. In practice, it can often severely reduce the yield of the desired protein as truncation products are very common. This problem is even more pronounced if two or more UAAs need to be installed.[69] Despite these shortcomings, amber suppression has great potential for the biophysical studies of IDPs. For example, it can be used in combination with cysteine chemistry or alone to install two distinct fluorescent and spectroscopic probes on the same protein. More importantly, as these reactions can be performed in cells, bioorthogonal chemistry and amber suppression can provide an alternative to fluorescent proteins for LLPS studies in the cellular milieu. In the context of spectroscopic probes, these strategies can enable the structural characterization of IDPs by EPR or PRE NMR in a native environment.[65,70] While there are currently few examples in the literature of successful applications to IDP and LLPS studies,[71] we expect that continuing developments to improve efficiency of UAA incorporation and speed of bioorthogonal labeling will

make these approaches a reliable and useful option for biophysical analysis both *in vitro* and in cells.[72,73]

Inteins and sortase are versatile tools for segmental labeling of IDPs

While IDPs are often studied by NMR, their repetitive sequences and low chemical shift dispersion can make resonance assignments challenging.[74] In such cases, it is helpful to segmentally label the protein, so that only a portion of the sequence is visible by NMR while the properties of the full-length polypeptide are preserved.[75,76] Segmental labeling is often performed with split inteins, protein engineering tools that can connect two separate protein segments through a native peptide bond in a process called *trans*-splicing (Fig. 3A).[77] To perform segmental labeling, the protein of interest (*i.e.* the extein) is divided into two fragments called the N- and C-exteins, respectively. Each fragment is fused with the corresponding N- or C-intein and the fusions are expressed and purified separately, so that each fusion construct can be labeled as desired (*e.g.* ^{15}N and natural abundance, or ^{15}N and ^{13}C respectively). After purification, the two constructs are mixed and the N- and C-inteins come together through non-covalent electrostatic interactions and adopt the functional intein horseshoe-like fold.[78] Upon folding, the assembled intein performs *trans*-splicing of the extein fragments and releases the newly ligated native protein of interest. Since most optimized split inteins use cysteine chemistry to carry out the *trans*-splicing reaction, segmental labeling requires that the extein is split at a native cysteine residue or that a cysteine is introduced at the desired location. Some inteins may also require a few additional residues beyond the junction cysteine, although the most efficient engineered intein so far, Cfa GEP, is quite tolerant to variations in the extein sequence.[29,79,80] Some inteins can also carry out splicing reactions at serine or threonine junctions, although their full extein dependency is less known.[81,82]

While inteins are gaining traction as segmental labeling tools for a variety of proteins, the biggest challenge for their application to most IDPs is the requirement that the *trans*-splicing reaction is performed under denaturing conditions to keep the reactants and final product soluble. The presence of urea or guanidinium can interfere with the folding of the intein and severely reduce the efficiency of splicing. The best intein to use in such cases is the Cfa GEP intein, which was engineered to withstand harsh conditions and can carry out splicing in buffers that contain up to 4 M guanidinium hydrochloride and 6 M urea.[29,79] There are also inteins that are tolerant to very high salt concentrations and can in principle be used for splicing in such conditions that often prevent phase separation of the protein of interest.[83] Extein–intein fusions may also have increased expression levels and higher solubility when compared to the full-length IDP construct, although this may be highly protein dependent. In cases where the solubility and purification of the extein–intein needs to be improved, solubility or affinity tags such as MBP or His_6 can be fused on the intein-side of the construct.[29,84] At this position, the tags do not interfere with the splicing reaction and are automatically removed from the extein when splicing takes place.

In addition to segmental labeling, inteins have other useful applications. For example, they can be adapted to attach C-terminal small molecule fluorophores to proteins of interest.[77,79,85] In this case, an engineered or native contiguous intein is fused to the C-terminal side of the extein. The fused intein can be cleaved in the presence of a thiol which generates a C-terminal thioester on the extein. The thioester can then be reacted with a small peptide bearing the desired fluorophore, thus generating the labeled protein. This strategy may be useful when fluorescent proteins such as GFP are not compatible with LLPS studies (*i.e.* if they interfere with LLPS) and when cysteine labeling is not an option. In a different adaptation of this strategy, the intein can be hydrolyzed from the extein at slightly basic pH.[86–89] If a suitable affinity tag is attached on the intein side, *e.g.* a His_6 or chitin, then the intein can be hydrolyzed directly during affinity purification. In this case, the intein and the tag remain on the column while the pure protein of interest

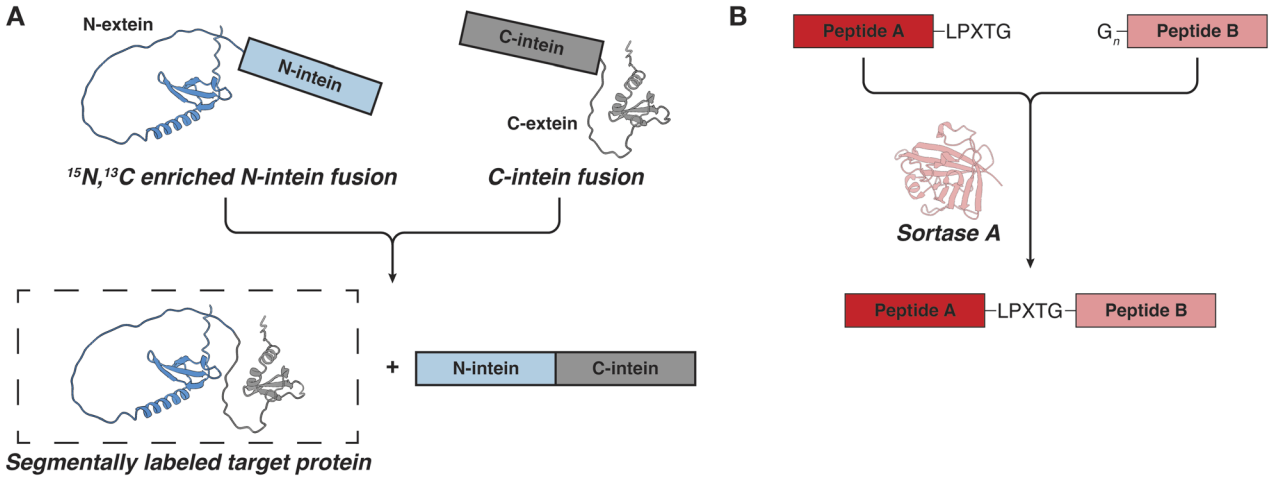


Fig. 3 Approaches for segmental labeling of IDPs for NMR studies. (A) Intein mediated protein *trans*-splicing. (B) Ligation of recombinant or synthetic polypeptides by sortase.

is released in solution. This strategy circumvents the need for a protease cleavage step to remove the affinity tag and may be a useful option when such steps are problematic in the IDP purification protocols.

An alternative tool for segmental labeling and protein modification is the transpeptidase sortase (Fig. 3B). This protein can stitch together two polypeptides in *trans*, provided that they carry an appropriate signal peptide.[90–92] The recognition signal typically consists of the LPXTG sequence (where X is any amino acid) on one peptide, and one to five glycine residues on the other peptide.[90] The two peptides are mixed and sortase is added in *trans* to carry out the ligation reaction, resulting in a minimum six residue "scar" in the newly formed protein. If one of the peptides is prepared in minimal isotopically labeled media, the resulting full-length protein will be segmentally labeled. In the context of IDPs and LLPS, there are several important considerations that need to be kept in mind.[93,94] First, to carry out the ligation reaction, sortase needs to be folded, which precludes the use of high concentrations of denaturants. Second, the reaction requires the presence of Ca^{2+} ions in the buffer, which may be incompatible with LLPS studies for some proteins.[92] And third, as the product polypeptide also contains the LPXTG recognition motif, the sortase reaction may be reversible and care must be taken to remove the product quickly and to minimize the generation of undesired products.[93] Nevertheless, the sortase-based ligations are conceptually elegant and may provide a useful alternative when the protein of interest is incompatible with intein splicing (*e.g.* if the protein charge interferes with the binding and folding of the intein fragments which proceeds through electrostatic interactions). Sortase has also been used extensively to attach proteins to surfaces and to add IDRs to the folded regions of proteins.[95,96] Interestingly, sortase has an IDR which undergoes a disorder-to-order transition upon binding of the signal peptide and Ca^{2+} ions.[97]

Introduction of post-translational modifications

Innate biological LLPS is often controlled by post-translational modifications (PTMs) including phosphorylation, acetylation, methylation, and ubiquitination.[98,99] Faithfully recapitulating PTMs *in vitro* is therefore crucial for constructing experiments that interrogate LLPS. While there are a variety of chemical and biochemical tools for installing or modelling PTMs after a protein has been recombinantly purified, IDPs present unique challenges for the applications of these tools.[98] Here, we review enzymatic approaches, bioisostere substitutions, cysteine chemistry and native chemical ligation, with the goal to give a range of options to biophysicists and biochemists interested in understanding the role of PTMs in IDPs and LLPS mechanisms and interactions.

The most common strategy to introduce PTMs in recombinantly produced IDPs is enzymatic modification. This approach is flexible as it can generally be applied to any substrate protein if the appropriate enzyme can be purchased or expressed and purified in house. Enzymes also introduce the same chemistries that are seen in cells, avoiding issues that may arise with PTM isosteres or PTM mimetics that result from other methods for PTM installation. The chemical accuracy and ease of use of this approach has made it the method of choice for many studies involving a variety of combinations of IDPs and PTMs, including phosphorylation of tau,[100] acetylation of FUS,[101] and mono-, di-, and trimethylation of histone tails[102] (among many others). The major drawback with enzymatic PTMs is that the activity of the enzyme can be difficult to control, and the resulting protein can be either overmodified (with the PTM present at multiple sites on the protein) or undermodified where the desired PTM is not installed efficiently. Controlling the installation site is also an issue, especially in unfolded IDPs whose residues are entirely solvent exposed. The proteins that result from mixtures of sites and distributions of PTMs that are generated by enzymes are often hard to separate from each other. These factors limit the enzymatic approach to cases where an appropriately active enzyme is available and there is either only one specific substrate residue in the protein or a distribution of modifications is desired, similar to the cases illustrated above.

A second straightforward strategy to incorporate PTMs is to genetically encode a bioisostere into the protein of interest. Bioisosteres, commonly used in medicinal chemistry and chemical biology, are atoms or functional groups with similar chemical and physical properties.[103] In the context of PTMs, perhaps the best known example is the substitution of phosphorylated serine or threonine residues with glutamic acid. Similar to enzymatic PTM incorporation, ease of application is a major benefit with this approach: bioisosteric residues can be encoded in the protein with cloning, and no further modification is required after protein purification. The drawback is that few good bioisosteres for PTMs exist among the canonical amino acids. Despite this limitation, this strategy has been commonly used in the literature, especially in the context of phosphorylation.[104] This includes a number of studies by the Fawzi group that reveal the influence of site-specific phosphorylation on the structural distributions and LLPS propensities of both FUS and TDP-43.[23,46] As more tRNA/tRNA synthetase pairs that encode pre-modified amino acids are developed, we expect that the incorporation of chemically accurate PTMs *via* genetic encoding by amber suppression will become a valuable strategy for incorporating PTMs into IDPs. It is already possible to encode phosphotyrosine and acetyllysine, for example, and new tRNA/tRNA synthetases are actively being developed.[105,106]

Cysteine chemistry presents a convenient way to introduce PTM mimics that are inaccessible with genetically encoded

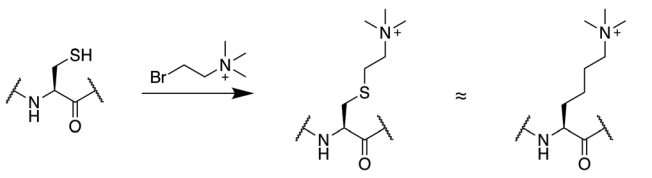


Fig. 4 Cysteine alkylation can be used to prepare methyl lysine mimics.

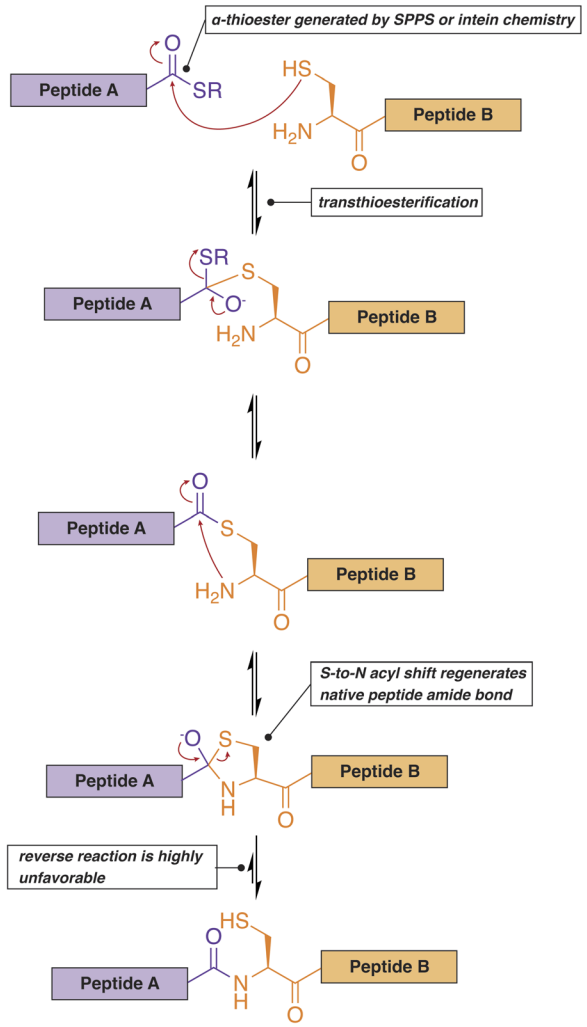


Fig. 5 Native chemical ligation can be used to ligate two synthetic or recombinant polypeptides. One of the peptides ends with a C-terminal thioester while the other peptide contains an N-terminal cysteine residue.

bioisosteres. Mimics of lysine methylation are common targets for this approach (Fig. 4). To use this strategy, a lysine residue is substituted with a cysteine, which is then alkylated with appropriate reagents to generate a mono-, di-, or trimethyllysine as desired.[107] This produces a sidechain that is the same length as lysine but bearing a sulfur instead of a carbon atom at the γ position. The reaction is compatible with denaturing conditions and is very popular in chromatin studies as it presents a relatively straightforward way to generate methylated histones.[107–110] It is important to note that in binding studies, alkylated cysteines display slightly higher K_d values compared to native methylated lysine residues.[109] They are, however, a good option when large amounts of methylated protein are needed, and are especially useful in the context of isotopic labeling for NMR studies.[108] Similarly, cysteine-based chemistries have been leveraged to introduce acetyl-lysine and methyl-arginine mimics into histone tails after cysteine mutations at endogenous lysine or arginine residues.[111–113] Asymmetric disulfide linkages can also be used to attach ubiquitin at well-defined positions in a protein sequence.[114,115]

Native chemical ligation (NCL) represents another versatile option for the site-selective introduction of PTMs in IDPs (Fig. 5).[116,117] Unlike most of the methods discussed above, this technique can generate proteins with multiple well-defined PTMs, enabling the construction of highly accurate models of post-translationally modified systems *in vitro*. In this case, a synthetic peptide bearing the necessary modifications is prepared through Fmoc- or Boc-mediated solid-phase peptide synthesis. The peptide is then ligated to another synthetic or recombinant fragment to build the full-length protein of interest. To perform NCL, the N-terminal fragment must contain a C-terminal α-thioester, while the C-terminal fragment needs to start with a nucleophilic amino acid such as cysteine. The thioester can be generated synthetically (native chemical ligation) or *via* the use of an intein that is fused to a recombinant protein or peptide fragment (expressed protein ligation, EPL).[116,118] The α-thioester serves as a reactive handle for the formation of a native amide protein backbone *via* a *trans*-thioesterification reaction with the cysteine sidechain of the second peptide fragment. Since the development of NCL, however, the range of chemistries that can be formed at this ligation site has extended far beyond cysteine to include many other natural and unnatural amino acids,[119] with recent methodologies expanding the scope to extremely challenging residues such as proline.[120] Another notable improvement in the NCL methodology is the development of the C-terminal hydrazide as a more stable and flexible replacement for the C-terminal α-thioester.[121] It should also be noted that the ligation reaction can proceed in the presence of urea or guanidinium hydrochloride and is therefore compatible with the production of IDPs that are prone to aggregation or premature LLPS.[122]

The primary advantage of using NCL or EPL to introduce PTMs into IDPs is that these techniques offer the capability to site-specifically introduce multiple distinct PTMs if desired. An excellent illustration of the value that this capability can offer to IDP research is provided by recent work by Ge *et al.*, in which a semisynthetic construct of the partially disordered JARID2 protein was produced using NCL.[123] In this report, multiple NCL reactions were utilized to generate a protein construct containing two distinct post translational modifications on individual residues in a site-specific manner. NCL and EPL have also been used extensively to produce α-synuclein and tau bearing a wide variety of PTMs, sometimes in combination with segmental isotopic labeling for NMR studies.[124–128]

Noncovalent small-molecule chemical tools for modulating LLPS and IDP behavior

One of the most exciting frontiers for IDP chemical biology is the development of noncovalent modulators of IDPs and LLPS. Disordered proteins do not offer hydrophobic pockets or other defined structural features that can serve as targets for small molecules. IDPs are therefore often avoided in high-throughput

screens for small-molecule effectors due to the high potential for failure. Despite this, successful screens for high-value protein targets such as Aβ, Myc, and α-synuclein are present in the literature,[129–134] suggesting that screening is a viable strategy for the discovery of chemical tools and drug leads that target IDPs directly.[134] Rational structure-based design has also been challenging although there is now at least one example of a successful structure–activity relationship (SAR)-based campaign against IDPs in the literature.[135] Given the central roles of IDPs and LLPS in cell biology and disease, the dearth of small-molecules capable of selectively engaging IDPs for use as chemical probes or therapeutic leads is an important problem in the field and makes this area of research especially exciting.[136] In this section, we will outline the precedent and explore the prospects for the discovery and development of such small-molecule modulators of IDPs and LLPS.

The current chemical landscape of LLPS modulators is sparse and lacks specificity

The chemical tool most often used to study LLPS is 1,6-hexanediol, a general disruptor of this process. Akyl alcohols including 1,6-hexanediol were first used to probe interactions between the FG-nucleoporins that gate the nuclear pore complex, which are IDPs that exhibit behavior consistent with LLPS.[137–139] Since its demonstration as a modulator of FG-nucleoporins, 1,6-hexanediol has been used as a convenient probe for assaying the properties of droplet-like structures in cells, for the rapid characterization of LLPS *in vitro*, or for differentiating between LLPS droplets and solids.[140–143] Despite its widespread use, 1,6-hexanediol is typically added at high concentrations of 5–10% w/v. Such high concentrations can interfere with the analysis of *in vitro* LLPS systems that are sensitive to abiotic conditions as well as with studies done in cells where 1,6-hexanediol disrupts a broad spectrum of cellular processes and can be cytotoxic.[140,144,145] Additionally, 1,6-hexanediol works primarily by disrupting hydrophobic interactions, a mechanism that does not address electrostatic, dipolar, and cation-π interactions that are also known to drive LLPS.[5] This suggests that the efficacy of 1,6-hexanediol in disrupting LLPS droplets may vary.[5,146,147] Conclusions drawn from experiments that rely on 1,6-hexanediol must therefore be considered with care and with the limitations of this tool in mind.

In addition to 1,6-hexanediol, a number of alkyl alcohols have been used to characterize LLPS. As a general trend, less hydrophobic alkyl alcohols (2,5-hexanediol and 1,2,3-hexanetriol, for example) are less effective at disrupting LLPS droplets, which is consistent with the proposed mechanism of action.[141] This difference in efficacy can be used as a tool for differentiating between phase-separated structures with different susceptibilities to disruption by alkyl alcohols.[143]

On the other end of the spectrum from 1,6-hexanediol and related disruptors lie small-molecule promoters of LLPS. Promoters include physiologically-relevant small molecule hydrotopes such as ATP, which can enhance the propensity of FUS to undergo LLPS in a concentration-dependent manner.[148,149] A synthetic example of a hydrotope capable of enhancing LLPS is 4,4′-dianilino-1,1′-binaphthyl-5,5′-disulfonic acid (bis-ANS). This molecule, along with a handful of similar but less effective naphthalene sulfonate derivatives, has been shown to promote LLPS for a number of common LLPS-prone

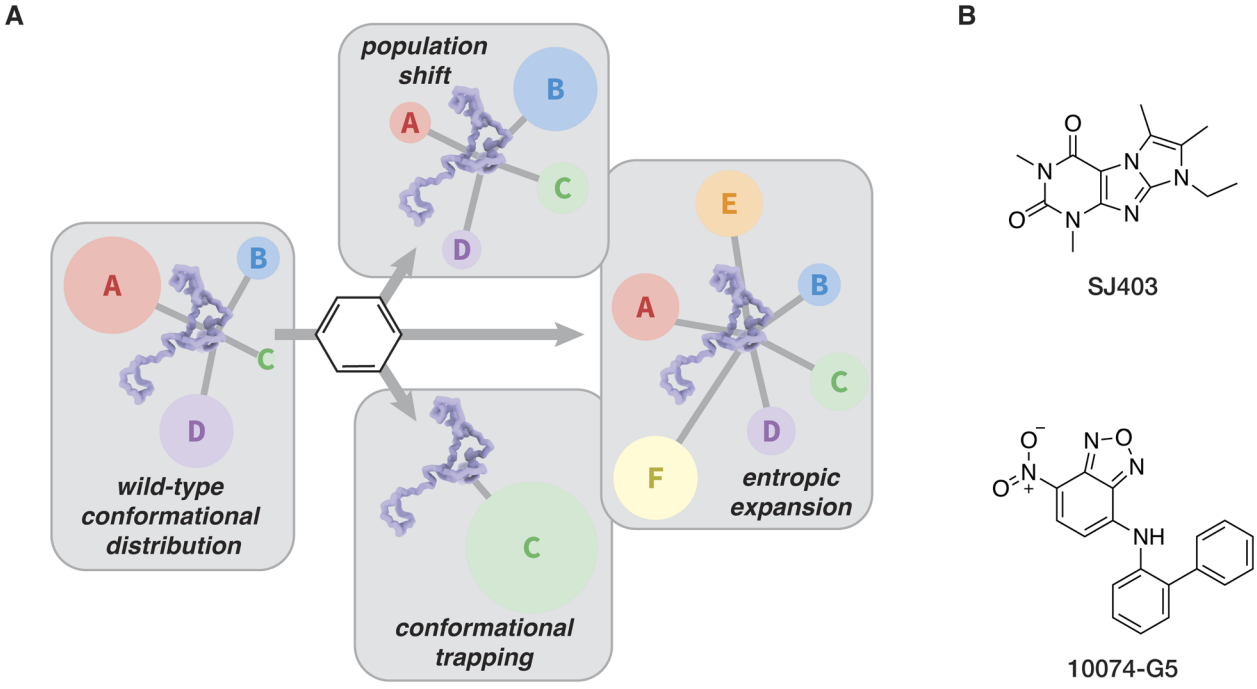


Fig. 6 Noncovalent small molecule modulators of IDPs. (A) Unique modes of action for small molecule modulators of IDPs, with A–F representing arbitrary populations of conformers. (B) Two small molecule modulators discussed in the text.

model proteins such as TDP-43 low complexity domain, tau, and FUS LC.[150] The study that introduces bis-ANS further demonstrates that Congo Red, a small molecule probe that is used as a reporter for amyloid, is capable of promoting LLPS in a similar manner to bis-ANS. This work suggests that bivalent, negatively charged compounds with hydrophobic cores can serve as hydrotopic drivers for LLPS. Some biologically important small molecules have also been shown to be capable of maintaining LLPS and modulating the onset of the liquid-to-solid transition of LLPS droplets. In one of the few systematic studies of its kind, Jonchhe *et al.* found that hydrophobic moieties in small molecules delay the onset of a liquid-to-solid phase transition in tau LLPS droplets.[151] In the same work, TMAO is shown to be an especially potent inhibitor of the liquid-to-solid transition, which is consistent with the observation that amphiphilic compounds like bis-ANS and ATP can drive LLPS.

Although they do not necessarily modulate LLPS directly, LLPS-sensitive probes that fluoresce or otherwise report on LLPS, are valuable tools for characterizing this process and may prove superior to disruptive reporters such as 1,6-hexanediol in many experimental contexts. Molecular rotors such as thioflavin T are sensitive to viscosity and can be used to probe LLPS while also serving as a tool for identifying and visualizing protein aggregation and fibrilization.[152] Recently, a novel aggregation-induced emission fluorogen, sodium 1,2-bis[4-(3-sulfonatopropoxyl)phenyl]-1,2-diphenylethene (BSPOTPE), was demonstrated to be capable of reporting on LLPS by partitioning into droplets *in vitro* and fluorescing in a viscosity-dependent manner.[153] The continued development of chemical tools capable of reporting on LLPS without the need for toxic or otherwise disruptive concentrations of 1,6-hexanediol will allow for less perturbative *in vitro* and in-cell assays. Molecules designed to non-perturbatively probe LLPS characteristics may also be useful for providing SAR data for chemistries capable of selectively partitioning into LLPS droplets. Just like biological LLPS condensates, whose composition is tuned by the chemistry of the client and scaffold proteins, it is feasible that small molecules could be tuned to partition into LLPS droplets with a particular set of chemical properties.[5,154] For example, recently, a handful of cancer therapeutics were shown to selectively partition into biomolecular LLPS condensates both *in vitro* and in cells.[155] It is our hope that further development of selective small-molecule LLPS modulators will lead to better chemical tools for studying LLPS and better therapeutics for addressing clinically-relevant LLPS dysregulation.

Rethinking the physicochemical basis of binding to IDPs

In order to develop more effective small molecule modulators of IDPs and LLPS, it is important to first consider the properties that would make a modulator effective in the first place. In general, three outcomes are available for small molecule binding to an IDP (Fig. 6A). First, the conformational distribution of an IDP may be changed or the conformational space available to an IDP may be reduced, a mechanism referred to as a population shift or conformational restriction.[156,157] Alternatively, the conformational states available to an IDP may increase, a mechanism referred to as entropic expansion.[158,159] Lastly, a small-molecule IDP modulator may bind to and induce a single conformation of an IDP, which we refer to as conformational trapping.

The population shift mechanism relies on transient enthalpy-driven binding of a small molecule to an IDP. Many IDPs adopt transient secondary structure, either due to a predominant low-energy conformation or as the result of a transient interaction with another protein (in the case of LLPS, this interaction is often with another copy of the same protein). As an example, adoption of transient secondary structure by an IDP has been demonstrated by Conicella *et al.*, who showed that transient α-helices in the C-terminal domain of TDP-43 exist both in the dispersed and LLPS phase, and that mutations that enhance the propensity for the formation of α-helices also enhance the propensity for LLPS.[47,160] Molecular dynamics and NMR are particularly powerful for providing insight into the role of transient secondary structure in the behavior of IDPs and IDP interactors. In a recent study, Zhu *et al.* integrated solution NMR data with all-atom molecular dynamics to provide a structure-based explanation for SAR differences between a family of bisphenol A-based modulators of the intrinsically disordered transactivation domain of the androgen receptor.[161] Rational design of IDPs that target transient structure has been achieved by clustering snapshots from molecular dynamics simulations and using favored structural ensembles as targets for docking experiments.[162]

The discovery and optimization of a molecule that confers a population shift in an IDP can also be performed in the absence of a full structural ensemble, as illustrated in work described by the Shelat, Zuo, and Kriwacki groups that shows the discovery and optimization of a small molecule capable of engaging the IDP $p27^{Kip1}$. An initial fragment-based NMR screen against $p27^{Kip1}$ yielded SJ403 (Fig. 6B), which can bind transient clusters of hydrophobic residues.[163] SJ403 is shown to induce a population shift through this binding mode, and further SAR work on this molecule generated a compound capable of sequestering $p27^{Kip1}$ into small soluble oligomers *via* the hydrophobic binding mechanism.[135,156] This effort provides an elegant example of the implementation of well-established drug discovery and optimization approaches to discover a small-molecule modulator of an IDP.

The entropic expansion mechanism introduces a much stronger focus on the entropic contributions to interactions between a small-molecule modulator and IDPs.[158] In the conformational trapping mechanism, and to a lesser degree the population shift mechanism, binding can be optimized through rational design principles against a favored IDP conformation. The strategy mirrors that of lead optimization for a folded protein, where the main goal is to introduce changes to a small molecule binder in order to decrease the enthalpy of binding.[164] In the entropic expansion approach, on the other hand, the entropic benefit of introducing more diversity into the conformational ensemble of a population of IDPs is considered. Enthalpic contributions provide a degree of target specificity through transient, weak interactions with

structural or sequence motifs in IDPs, while the entropic contribution favors interactions with the IDPs and effectively reduces unwanted LLPS or aggregation by expanding the range of conformations that the proteins can adopt.[158] The only example to date of a molecule that works through the entropic expansion mechanism is 10074-G5 (Fig. 6B), a small molecule inhibitor of amyloid-β_{42} aggregation discovered by the Knowles, Dobson, and Vendruscolo groups.[165] Since many IDPs undergo pathological conformational collapse into aggregates and fibrils, it is clear that molecules that exploit the entropic expansion mechanism could be viable therapeutic leads. It is possible that other small molecules that protect against aggregation of IDPs while preserving their wild-type function work through this mechanism.[166,167] Further efforts exploring the entropic expansion mechanism will be valuable for the development of better chemical tools for studies of IDPs and LLPS.

For both the population shift and entropic expansion mechanisms, a framework for initial small-molecule binding and selectivity must exist. For enthalpically driven binding, selectivity can be explained in some cases by the existence of transient binding pockets. Recently, Robustelli *et al.* provided an alternative mechanism for selectivity in small molecule binders of IDPs. Using all-atom molecular dynamics simulations and solution NMR studies of a fragment of α-synuclein and fasudil, a known ligand for α-synuclein, Robustelli and coworkers demonstrated that the spacing of low-affinity binding sites for fasudil along the α-synuclein sequence can compensate for the lack of a single well-conserved or high-affinity binding site. This mechanism for selectivity, called "dynamic shuttling" provides a compelling explanation for small-molecule selectivity towards specific IDP sequences that is compatible with both the population shift and entropic expansion paradigms for small-molecule-IDP interactions.[168]

A final mechanism for IDP interactions with small molecules is conformational trapping. Due to the large entropic cost associated with restricting an IDP to a single conformation, this mode of interaction is challenging to achieve with a small molecule. An example of conformational trapping can be found in the biotinylated 5-aryl-isoxazole-3-carboxyamide molecule (b-isox) described by Kato *et al.* In this case, the microcrystalline form of b-isox is the active modality. Grooves on the crystal surface provide binding sites capable of overcoming the entropic penalty associated with conformational trapping of IDPs, and multiple IDPs associated with RNA granules can be trapped and subsequently isolated using b-isox as a chemical tool.[169]

Future directions

A quarter of a century of IDP research has revealed that unstructured polypeptides carry out a vast number of important and precise functions in the cell.[10,170,171] Despite these advances, there is still much to learn regarding how cells translate a multitude of non-specific interactions into specific biological outcomes, and how the cellular machinery perceives functional and pathological states of IDPs. Further developments in both structural biology and protein chemistry will be crucial to these efforts.

Here, we have reviewed the current state of the art in chemical and biochemical approaches that allow the preparation of IDPs for imaging and structural studies and enable the investigation of PTMs in physiologically relevant contexts (summarized in Table 1). While these tools have become much more efficient and versatile over time, the handling requirements and aggregation propensity of many IDPs still present a tremendous challenge. The development of the Cfa GEP intein, for example, enabled *trans*-splicing and segmental labeling of many new proteins, including IDPs such as FUS.[29,79,172] Yet, more can be done to improve the efficiency of inteins in high concentrations of guanidinium hydrochloride and urea, conditions that are essential for the handling of many aggregation and LLPS prone IDPs. The development of cysteine alkylation approaches to mimic lysine methylation, on the other hand, has allowed NMR spectroscopists to study the role of this modification using specifically methylated, biologically relevant samples.[107,108,110] The design of similarly robust and easy to implement methodologies to create acetylation and phosphorylation mimics or modifications on recombinantly produced templates, would be highly beneficial.

Going forward, we expect that protein engineering efforts on IDPs will shift more and more to the cellular environment. This reflects a recent increase in efforts from imaging, NMR and EPR spectroscopy to understand how cells shape the conformational ensembles, interactions, and dynamics of IDPs and LLPS-based compartments.[51,173,174] While most imaging efforts so far have relied on fusion fluorescent proteins such as GFP, the significant bulk of these tags may not be compatible with all IDPs or LLPS studies.[41] In these situations, alternative labeling approaches with small molecule fluorescent probes are needed. Here, we expect that unnatural amino acid incorporation through amber suppression and bioorthogonal labeling methodologies would be particularly valuable. In addition to the delivery of fluorescent probes in cells, this technology is also gaining traction for EPR and sensitivity-enhanced NMR spectroscopy in cells.[65,70] Recent developments have focused on improving the efficiency of UAA incorporation, and in particular, the installation of two or three UAAs in mammalian cells.[66,72,73] It should also be mentioned that the intein approach can also be adapted to studies in cells, where it can be used to control IDP function and for the installation of PTMs.[175,176]

Perhaps the greatest challenge facing the IDP chemical biology field is the development of specific small molecule modulators that allow regulation of IDP and LLPS functions in cells. These efforts invariably tie back to more detailed understanding of the structural ensemble, dynamics, and interactions of the IDP of interest. These developments will no doubt require close collaborations between synthetic chemists, structural biologists, and computational chemists. A number of NMR studies, often complemented with all atom molecular dynamics simulations, have already provided valuable information regarding the interactions between small molecules and

Table 1 Summary of chemical tools suitable for IDP labeling and modulation, as discussed in the text

Goal	Strategy	Advantages	Disadvantages	References
Purification of IDPs	Purification under denaturing conditions	• In principle, compatible with all IDP/IDRs	• Denaturant must be removed • Refolding required for constructs that contain both folded and IDR domains	27, 28 and 172
	Purification at extreme pH	• Straightforward recovery of target protein	• Applicable to only a small subset of proteins	18 and 22–26
	Purification at extreme temperature	• Straightforward recovery of target protein	• Applicable to only a small subset of proteins	26
Labeling with spectroscopic or imaging probes	Maleimide or iodoacetamide-based probes	• Site-specific at cysteine residues • Compatible with isotopic labeling for NMR	• Not applicable in cells • Potential labeling at multiple sites • Probes may perturb IDP function and LLPS	36–41
	N-Hydroxysuccininimide ester-based chemistry	• Site-specific at lysine residues • Compatible with isotopic labeling for NMR	• Not applicable in cells • Potential labeling at multiple sites • Probes may perturb IDP function and LLPS	42
	Fusion protein-based approaches	• Genetically encoded • Applicable in cells	• Large fusion proteins may perturb IDP function and LLPS	See ref. 41 and 183 for comparison of the effects on LLPS of small molecule labeling *vs.* fusion protein labeling
	Bioorthogonal chemical approaches and amber suppression	• Genetically encoded • Applicable in cells • Site-specific labeling • May be compatible with isotopic labeling for NMR	• UAA incorporation may be inefficient and lead to low protein yields • Probes may perturb IDP function and LLPS	56, 61–67, 70, 184 and 185
	Native chemical ligation and expressed protein ligation	• Site-specific labeling • Multiple probes can be introduced at the same time • Access to a wide range of chemical probes • May be compatible with isotopic labeling for NMR	• Segments must be accessible synthetically • Difficult to label sites away from the N- or C-terminus • Final protein yields may be low • Not applicable in cells • Probes may perturb IDP function and LLPS	77, 79, 116 and 118 Also see relevant references for introducing PTMs through NCL and EPL below
Segmental labeling for NMR spectroscopy	Cfa GEP	• *Trans*-splicing can be achieved under denaturing conditions • High efficiency • Robust to a range of extein sequences • May be applied in cells	• Splicing reaction must take place at a cysteine residue	29, 79 and 172
	MCM2	• Salt-inducible splicing can be used to control intein activity • Robust reactivity under high-salt conditions • Serine-based reaction mechanism provides versatility	• Splicing cannot be performed under denaturing conditions • May not be applicable in cells • Splicing rate and extein sequence compatibility are poor compared to Cfa	83
	Sortase	• A possible alternative to intein-based methods • May be used to attach proteins to cell surfaces	• Requires the insertion of a five/six amino acid scar into the target protein • Continued reaction between reactants and	90–96 and 186

Table 1 (*Contd.*)

Goal	Strategy	Advantages	Disadvantages	References
			desired product may reduce yield	
Introduction of post-translational modifications	Enzymatic methods	• Straightforward introduction of native PTMs	• Lack of control over stoichiometry and location of PTM installation	100–102
	Genetic encoding of a bioisostere	• Straightforward introduction of PTM mimics • Applicable in cells	• Scope of PTMs is limited to those that can be mimicked effectively by a bioisostere	23, 46 and 104
	Amber suppression	• Introduction of native PTMs or PTM mimics • Applicable in cells	• Scope of PTMs is limited to those that have a tRNA synthetase available • UAA incorporation may be inefficient and lead to low protein yields	105 and 106
	Cysteine alkylation	• Efficient and specific introduction of methyl-lysine mimics • Reactions are compatible with denaturing conditions	• Methyl-lysine mimic may not faithfully reproduce the function of the native PTM • Not applicable in cells	41, 107, 109 and 110
	Native chemical ligation and expressed protein ligation	• Site-specific introduction of PTMs • Multiple PTMs can be introduced at the same time • Access to a wide range of modifications	• Segments must be accessible synthetically • Difficult to introduce PTMs away from the N- or C-terminus • Final protein yields may be low • Not applicable in cells	125–128, 187 and 188

IDPs.[135,156,158,163,165,168] There have also been exciting developments in the coarse-grained simulations of LLPS[177–180] and we are looking forward to the extension of these studies to all atom simulations and the incorporation of small molecule modulators.[181,182]

While IDPs present significant challenges to chemical and structural biologists, they also put forward valuable opportunities to hone existing chemical and biophysical methodologies and to develop more efficient and precise tools. Such developments will have far reaching implications for the chemical and biological fields and will undoubtedly enrich our understanding of protein function in health and disease.

Author contributions

R. F. B. and G. T. D. conceived the idea and wrote the manuscript. R. F. B. prepared the figures.

Conflicts of interest

The authors declare no conflicts of interest.

Acknowledgements

This work was supported by a Research Education Component associated with NIH Grant P30 AG062429, an R21 AG069064 award to G. T. D., and a T32 GM112584 fellowship to R. F. B.

References

1 R. van der Lee, M. Buljan, B. Lang, R. J. Weatheritt, G. W. Daughdrill, A. K. Dunker, M. Fuxreiter, J. Gough, J. Gsponer, D. T. Jones, P. M. Kim, R. W. Kriwacki, C. J. Oldfield, R. V. Pappu, P. Tompa, V. N. Uversky, P. E. Wright and M. M. Babu, Classification of intrinsically disordered regions and proteins, *Chem. Rev.*, 2014, **114**, 6589–6631.

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

3 J. N. Onuchic, Z. Luthey-Schulten and P. G. Wolynes, Theory of protein folding: the energy landscape perspective, *Annu. Rev. Phys. Chem.*, 1997, **48**, 545–600.

4 R. Neme, C. Amador, B. Yildirim, E. McConnell and D. Tautz, Random sequences are an abundant source of bioactive RNAs or peptides, *Nat. Ecol. Evol.*, 2017, **1**, 0217.

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

6 I. Pritisanac, R. M. Vernon, A. M. Moses and J. D. Forman Kay, *Entropy and Information within Intrinsically Disordered Protein Regions*, Entropy (Basel), 2019, p. 21.

7 S. F. Banani, H. O. Lee, A. A. Hyman and M. K. Rosen, Biomolecular condensates: organizers of cellular biochemistry, *Nat. Rev. Mol. Cell Biol.*, 2017, **18**, 285–298.

8 M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, *Nat. Genet.*, 2000, **25**, 25–29.

9 Gene Ontology Consortium, The Gene Ontology resource: enriching a GOld mine, *Nucleic Acids Res.*, 2021, **49**, D325–D334.

10 H. J. Dyson and P. E. Wright, Coupling of folding and binding for unstructured proteins, *Curr. Opin. Struct. Biol.*, 2002, **12**, 54–60.

11 A. Borgia, M. B. Borgia, K. Bugge, V. M. Kissling, P. O. Heidarsson, C. B. Fernandes, A. Sottini, A. Soranno, K. J. Buholzer, D. Nettels, B. B. Kragelund, R. B. Best and B. Schuler, Extreme disorder in an ultrahigh-affinity protein complex, *Nature*, 2018, **555**, 61–66.

12 M. K. Hazra and Y. Levy, Affinity of disordered protein complexes is modulated by entropy-energy reinforcement, *Proc. Natl. Acad. Sci. U. S. A.*, 2022, **119**, e2120456119.

13 A. C. Murthy and N. L. Fawzi, The (un)structural biology of biomolecular liquid-liquid phase separation using NMR spectroscopy, *J. Biol. Chem.*, 2020, **295**, 2375–2384.

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

15 G. Tedeschi, M. Mangiagalli, S. Chmielewska, M. Lotti, A. Natalello and S. Brocca, Aggregation properties of a disordered protein are tunable by pH and depend on its net charge per residue, *Biochim. Biophys. Acta, Gen. Subj.*, 2017, **1861**, 2543–2550.

16 S. P. Graether, Troubleshooting Guide to Expressing Intrinsically Disordered Proteins for Use in NMR Experiments, *Front. Mol. Biosci.*, 2018, **5**, 118.

17 S. Alberti, S. Saha, J. B. Woodruff, T. M. Franzmann, J. Wang and A. A. Hyman, A User's Guide for Phase Separation Assays with Purified Proteins, *J. Mol. Biol.*, 2018, **430**, 4806–4820.

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

19 M. Vivoli Vega, A. Nigro, S. Luti, C. Capitini, G. Fani, L. Gonnelli, F. Boscaro and F. Chiti, Isolation and characterization of soluble human full-length TDP-43 associated with neurodegeneration, *FASEB J.*, 2019, **33**, 10780–10793.

20 J. Van Lindt, A. Bratek-Skicki, P. N. Nguyen, D. Pakravan, L. F. Duran-Armenta, A. Tantos, R. Pancsa, L. Van Den Bosch, D. Maes and P. Tompa, A generic approach to study the kinetics of liquid-liquid phase separation under near-native conditions, *Commun. Biol.*, 2021, **4**, 77.

21 J. A. Morin, S. Wittmann, S. Choubey, A. Klosin, S. Golfier, A. A. Hyman, F. Julicher and S. W. Grill, Sequence-dependent surface condensation of a pioneer transcription factor on DNA, *Nat. Phys.*, 2022, **18**, 271–276.

22 C. P. Pedersen, P. Seiffert, I. Brakti and K. Bugge, Production of Intrinsically Disordered Proteins for Biophysical Studies: Tips and Tricks, *Methods Mol. Biol.*, 2020, **2141**, 195–209.

23 Z. Monahan, V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerze, R. O'Meally, G. L. Dignon, A. E. Conicella, W. Zheng, R. B. Best, R. N. Cole, J. Mittal, F. Shewmaker and N. L. Fawzi, Phosphorylation of the FUS low-complexity domain disrupts phase separation, aggregation, and toxicity, *EMBO J.*, 2017, **36**, 2951–2967.

24 A. C. Murthy, G. L. Dignon, Y. Kan, G. H. Zerze, S. H. Parekh, J. Mittal and N. L. Fawzi, Molecular interactions underlying liquid-liquid phase separation of the FUS low-complexity domain, *Nat. Struct. Mol. Biol.*, 2019, **26**, 637–648.

25 R. F. Berkeley, M. Kashefi and G. T. Debelouchina, Real-time observation of structure and dynamics during the liquid-to-solid transition of FUS LC, *Biophys. J.*, 2021, **120**, 1276–1287.

26 A. D. Stephens, D. Matak-Vinkovic, A. Fernandez-Villegas and G. S. Kaminski Schierle, Purification of Recombinant alpha-synuclein: A Comparison of Commonly Used Protocols, *Biochemistry*, 2020, **59**, 4563–4572.

27 M. Pesarrodona, I. Latorre and X. Salvatella, Intrinsically Disordered Proteins (IDP): Purification Under Denaturing Conditions, *Methods Mol. Biol.*, 2022, **2406**, 359–370.

28 B. D. Fonda, K. M. Jami, N. R. Boulos and D. T. Murray, Identification of the Rigid Core for Aged Liquid Droplets of an RNA-Binding Protein Low Complexity Domain, *J. Am. Chem. Soc.*, 2021, **143**, 6657–6668.

29 A. J. Stevens, Z. Z. Brown, N. H. Shah, G. Sekar, D. Cowburn and T. W. Muir, Design of a Split Intein with Exceptional Protein Splicing Activity, *J. Am. Chem. Soc.*, 2016, **138**, 2162–2165.

30 N. O. Taylor, M. T. Wei, H. A. Stone and C. P. Brangwynne, Quantifying Dynamics in Phase-Separated Condensates Using Fluorescence Recovery after Photobleaching, *Biophys. J.*, 2019, **117**, 1285–1300.

31 L. A. Metskas and E. Rhoades, Single-Molecule FRET of Intrinsically Disordered Proteins, *Annu. Rev. Phys. Chem.*, 2020, **71**, 391–414.

32 S. Weickert, J. Cattani and M. Drescher, in *Electron Paramagnetic Resonance*, ed. V. Chechik and D. M. Murphy, 2018, vol. 26, pp. 1–37.

33 E. A. Hoyt, P. M. S. D. Cal, B. L. Oliveira and G. J. L. Bernardes, Contemporary approaches to site-selective protein modification, *Nat. Rev. Chem.*, 2019, **3**, 147–171.

34 G. T. Debelouchina and T. W. Muir, A molecular engineering toolbox for the structural biologist, *Q. Rev. Biophys.*, 2017, **50**, e7.

35 E. T. Clark, E. E. Sievers and G. T. Debelouchina, A Chemical Biology Primer for NMR Spectroscopists, *J. Magn. Reson. Open*, 2022, 10–11.

36 J. N. deGruyter, L. R. Malins and P. S. Baran, Residue-Specific Peptide Modification: A Chemist's Guide, *Biochemistry*, 2017, **56**, 3863–3873.

37 O. Boutureira and G. J. Bernardes, Advances in chemical protein modification, *Chem. Rev.*, 2015, **115**, 2174–2195.

38 Y. Fu and N. S. Finney, Small-molecule fluorescent probes and their design, *RSC Adv.*, 2018, **8**, 29051–29061.

39 K. Rhine, M. A. Makurath, J. Liu, S. Skanchy, C. Lopez, K. F. Catalan, Y. Ma, C. M. Fare, J. Shorter, T. Ha, Y. R. Chemla and S. Myong, ALS/FTLD-Linked Mutations in FUS Glycine Residues Cause Accelerated Gelation and Reduced Interactions with Wild-Type FUS, *Mol. Cell*, 2020, **80**, 1139.

40 Z. R. Grese, A. C. Bastos, L. D. Mamede, R. L. French, T. M. Miller and Y. M. Ayala, Specific RNA interactions promote TDP-43 multivalent phase separation and maintain liquid properties, *EMBO Rep.*, 2021, **22**, e53632.

41 A. G. Larson, D. Elnatan, M. M. Keenen, M. J. Trnka, J. B. Johnston, A. L. Burlingame, D. A. Agard, S. Redding and G. J. Narlikar, Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin, *Nature*, 2017, **547**, 236–240.

42 Y. Lin, Y. Fichou, A. P. Longhini, L. C. Llanes, P. Yin, G. C. Bazan, K. S. Kosik and S. Han, Liquid-Liquid Phase Separation of Tau Driven by Hydrophobic Interaction Facilitates Fibrillization of Tau, *J. Mol. Biol.*, 2021, **433**, 166731.

43 C. B. Rosen and M. B. Francis, Targeting the N terminus for site-selective protein modification, *Nat. Chem. Biol.*, 2017, **13**, 697–705.

44 G. M. Clore, Practical Aspects of Paramagnetic Relaxation Enhancement in Biological Macromolecules, *Methods Enzymol.*, 2015, **564**, 485–497.

45 V. H. Ryan, G. L. Dignon, G. H. Zerze, C. V. Chabata, R. Silva, A. E. Conicella, J. Amaya, K. A. Burke, J. Mittal and N. L. Fawzi, Mechanistic View of hnRNPA2 Low-Complexity Domain Structure, Interactions, and Phase Separation Altered by Mutation and Arginine Methylation, *Mol. Cell*, 2018, **69**, 465–479.

46 A. Wang, A. E. Conicella, H. B. Schmidt, E. W. Martin, S. N. Rhoads, A. N. Reeb, A. Nourse, D. Ramirez Montero, V. H. Ryan, R. Rohatgi, F. Shewmaker, M. T. Naik, T. Mittag, Y. M. Ayala and N. L. Fawzi, A single N-terminal phosphomimic disrupts TDP-43 polymerization, phase separation, and RNA splicing, *EMBO J.*, 2018, **37**.

47 A. E. Conicella, G. H. Zerze, J. Mittal and N. L. Fawzi, ALS Mutations Disrupt Phase Separation Mediated by alpha-Helical Structure in the TDP-43 Low-Complexity C-Terminal Domain, *Structure*, 2016, **24**, 1537–1549.

48 I. Ritsch, E. Lehmann, L. Emmanouilidis, M. Yulikov, F. Allain and G. Jeschke, Phase Separation of Heterogeneous Nuclear Ribonucleoprotein A1 upon Specific RNA-Binding Observed by Magnetic Resonance, *Angew. Chem., Int. Ed.*, 2022, **61**, e202204311.

49 G. Jeschke, DEER distance measurements on proteins, *Annu. Rev. Phys. Chem.*, 2012, **63**, 419–446.

50 N. A. Eschmann, E. R. Georgieva, P. Ganguly, P. P. Borbat, M. D. Rappaport, Y. Akdogan, J. H. Freed, J. E. Shea and S. Han, Signature of an aggregation-prone conformation of tau, *Sci. Rep.*, 2017, **7**, 44739.

51 F. X. Theillet, A. Binolfi, B. Bekei, A. Martorana, H. M. Rose, M. Stuiver, S. Verzini, D. Lorenz, M. van Rossum, D. Goldfarb and P. Selenko, Structural disorder of monomeric alpha-synuclein persists in mammalian cells, *Nature*, 2016, **530**, 45–50.

52 L. Emmanouilidis, L. Esteban-Hofer, F. F. Damberger, T. de Vries, C. K. X. Nguyen, L. F. Ibanez, S. Mergenthal, E. Klotzsch, M. Yulikov, G. Jeschke and F. H. Allain, NMR and EPR reveal a compaction of the RNA-binding protein FUS upon droplet formation, *Nat. Chem. Biol.*, 2021, **17**, 608–614.

53 E. H. Yardeni, T. Bahrenberg, R. A. Stein, S. Mishra, E. Zomot, B. Graham, K. L. Tuck, T. Huber, E. Bibi, H. S. McHaourab and D. Goldfarb, Probing the solution structure of the E. coli multidrug transporter MdfA using DEER distance measurements with nitroxide and Gd(III) spin labels, *Sci. Rep.*, 2019, **9**, 12528.

54 V. Venditti and N. L. Fawzi, Probing the Atomic Structure of Transient Protein Contacts by Paramagnetic Relaxation Enhancement Solution NMR, *Methods Mol. Biol.*, 2018, **1688**, 243–255.

55 S. D. Fontaine, R. Reid, L. Robinson, G. W. Ashley and D. V. Santi, Long-term stabilization of maleimide-thiol conjugates, *Bioconjugate Chem.*, 2015, **26**, 145–152.

56 L. Wang, A. Brock, B. Herberich and P. G. Schultz, Expanding the genetic code of Escherichia coli, *Science*, 2001, **292**, 498–500.

57 L. Wang and P. G. Schultz, Expanding the genetic code, *Angew. Chem., Int. Ed.*, 2004, **44**, 34–66.

58 C. C. Liu and P. G. Schultz, Adding new chemistries to the genetic code, *Annu. Rev. Biochem.*, 2010, **79**, 413–444.

59 J. W. Chin, Expanding and reprogramming the genetic code, *Nature*, 2017, **550**, 53–60.

60 K. Lang and J. W. Chin, Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins, *Chem. Rev.*, 2014, **114**, 4764–4806.

61 K. Lang, L. Davis, J. Torres-Kolbus, C. J. Chou, A. Deiters and J. W. Chin, Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction, *Nat. Chem.*, 2012, **4**, 298–304.

62 K. Lang, L. Davis, S. Wallace, M. Mahesh, D. J. Cox, M. L. Blackman, J. M. Fox and J. W. Chin, Genetic Encoding of Bicyclononynes and trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels-Alder Reactions, *J. Am. Chem. Soc.*, 2012, **134**, 10317–10320.

63 T. Plass, S. Milles, C. Koehler, J. Szymanski, R. Mueller, M. Wiessler, C. Schultz and E. A. Lemke, Amino acids for Diels-Alder reactions in living cells, *Angew. Chem., Int. Ed.*, 2012, **51**, 4166–4170.

64 J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang and P. G. Schultz, Addition of p-azido-L-phenylalanine to the genetic code of Escherichia coli, *J. Am. Chem. Soc.*, 2002, **124**, 9026–9027.
65 P. Widder, J. Schuck, D. Summerer and M. Drescher, Combining site-directed spin labeling in vivo and in-cell EPR distance determination, *Phys. Chem. Chem. Phys.*, 2020, **22**, 4875–4879.
66 D. de la Torre and J. W. Chin, Reprogramming the genetic code, *Nat. Rev. Genet.*, 2021, **22**, 169–184.
67 M. J. Schmidt, J. Borbas, M. Drescher and D. Summerer, A genetically encoded spin label for electron paramagnetic resonance distance measurements, *J. Am. Chem. Soc.*, 2014, **136**, 1238–1241.
68 M. J. Schmidt, A. Fedoseev, D. Bucker, J. Borbas, C. Peter, M. Drescher and D. Summerer, EPR Distance Measurements in Native Proteins with Genetically Encoded Spin Labels, *ACS Chem. Biol.*, 2015, **10**, 2764–2771.
69 A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao and P. G. Schultz, A versatile platform for single- and multiple-unnatural amino acid mutagenesis in Escherichia coli, *Biochemistry*, 2013, **52**, 1828–1837.
70 B. J. Lim, B. E. Ackermann and G. T. Debelouchina, Targetable Tetrazine-Based Dynamic Nuclear Polarization Agents for Biological Systems, *Chembiochem*, 2020, **21**, 1315–1319.
71 C. M. Haney, R. F. Wissner, J. B. Warner, Y. X. J. Wang, J. J. Ferrie, D. J. Covell, R. J. Karpowicz, V. M. Y. Lee and E. J. Petersson, Comparison of strategies for non-perturbing labeling of alpha-synuclein to study amyloidogenesis, *Org. Biomol. Chem.*, 2016, **14**, 1584–1592.
72 W. E. Robertson, L. F. H. Funke, D. de la Torre, J. Fredens, T. S. Elliott, M. Spinck, Y. Christova, D. Cervettini, F. L. Boge, K. C. Liu, S. Buse, S. Maslen, G. P. C. Salmond and J. W. Chin, Sense codon reassignment enables viral resistance and encoded polymer synthesis, *Science*, 2021, **372**, 1057–1062.
73 E. D. Ficaretta, C. J. J. Wrobel, S. J. S. Roy, S. B. Erickson, J. S. Italia and A. Chatterjee, A Robust Platform for Unnatural Amino Acid Mutagenesis in E. coli Using the Bacterial Tryptophanyl-tRNA synthetase/tRNA pair, *J. Mol. Biol.*, 2022, **434**, 167304.
74 J. Yao, H. J. Dyson and P. E. Wright, Chemical shift dispersion and secondary structure prediction in unfolded and partly folded proteins, *FEBS Lett.*, 1997, **419**, 285–289.
75 T. Yamazaki, T. Otomo, N. Oda, Y. Kyogoku, K. Uegaki, N. Ito, Y. Ishino and H. Nakamura, Segmental isotope labeling for protein NMR using peptide splicing, *J. Am. Chem. Soc.*, 1998, **120**, 5591–5592.
76 R. Xu, B. Ayers, D. Cowburn and T. W. Muir, Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 388–393.
77 N. H. Shah, G. P. Dann, M. Vila-Perello, Z. H. Liu and T. W. Muir, Ultrafast Protein Splicing is Common among Cyanobacterial Split Inteins: Implications for Protein Engineering, *J. Am. Chem. Soc.*, 2012, **134**, 11338–11341.
78 N. H. Shah, E. Eryilmaz, D. Cowburn and T. W. Muir, Naturally split inteins assemble through a "capture and collapse" mechanism, *J. Am. Chem. Soc.*, 2013, **135**, 18673–18681.
79 A. J. Stevens, G. Sekar, N. H. Shah, A. Z. Mostafavi, D. Cowburn and T. W. Muir, A promiscuous split intein with expanded protein engineering applications, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 8538–8543.
80 N. H. Shah, E. Eryilmaz, D. Cowburn and T. W. Muir, Extein residues play an intimate role in the rate-limiting step of protein trans-splicing, *J. Am. Chem. Soc.*, 2013, **135**, 5839–5847.
81 P. Carvajal-Vallejos, R. Pallisse, H. D. Mootz and S. R. Schmidt, Unprecedented Rates and Efficiencies Revealed for New Natural Split Inteins from Metagenomic Sources, *J. Biol. Chem.*, 2012, **287**, 28686–28696.
82 M. Bhagawati, T. M. E. Terhorst, F. Fusser, S. Hoffmann, T. Pasch, S. Pietrokovski and H. D. Mootz, A mesophilic cysteine-less split intein for protein trans-splicing applications under oxidizing conditions, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 22164–22172.
83 A. Ciragan, S. M. Backlund, K. M. Mikula, H. M. Beyer, O. H. Samuli Ollila and H. Iwai, NMR Structure and Dynamics of TonB Investigated by Scar-Less Segmental Isotopic Labeling Using a Salt-Inducible Split Intein, *Front. Chem.*, 2020, **8**, 136.
84 K. K. Frederick, V. K. Michaelis, M. A. Caporini, L. B. Andreas, G. T. Debelouchina, R. G. Griffin and S. Lindquist, Combining DNP NMR with segmental and specific labeling to study a yeast prion protein strain that is not parallel in-register, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 3642–3647.
85 N. H. Shah and T. W. Muir, Inteins: nature's gift to protein chemists, *Chem. Sci.*, 2014, **5**, 446–461.
86 B. J. Lim, R. F. Berkeley and G. T. Debelouchina, Fused Split Inteins: Tools for Introducing Multiple Protein Modifications, *Methods Mol. Biol.*, 2020, **2133**, 163–181.
87 S. Batjargal, C. R. Walters and E. J. Petersson, Inteins as Traceless Purification Tags for Unnatural Amino Acid Proteins, *J. Am. Chem. Soc.*, 2015, **137**, 1734–1737.
88 S. R. Chong, F. B. Mersha, D. G. Comb, M. E. Scott, D. Landry, L. M. Vence, F. B. Perler, J. Benner, R. B. Kucera, C. A. Hirvonen, J. J. Pelletier, H. Paulus and M. Q. Xu, Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element, *Gene*, 1997, **192**, 271–281.
89 M. W. Southworth, K. Amaya, T. C. Evans, M. Q. Xu and F. B. Perler, Purification of proteins fused to either the amino or carboxy terminus of the Mycobacterium xenopi gyrase A intein, *Biotechniques*, 1999, **27**, 110–120.
90 H. Ton-That, G. Liu, S. K. Mazmanian, K. F. Faull and O. Schneewind, Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of Staphylococcus aureus at the LPXTG motif, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 12424–12429.

91 S. K. Mazmanian, G. Liu, H. Ton-That and O. Schneewind, Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall, *Science*, 1999, **285**, 760–763.

92 A. W. Jacobitz, M. D. Kattke, J. Wereszczynski and R. T. Clubb, Sortase Transpeptidases: Structural Biology and Catalytic Mechanism, *Adv. Protein Chem. Struct. Biol.*, 2017, **109**, 223–264.

93 H. Morgan, W. BruceTurnbull and M. Webb, Challenges in the use of sortase and other peptide ligases for site-specific protein modification, *Chem. Soc. Rev.*, 2022, **51**, 4121–4145.

94 K. V. Boyko, E. A. Rosenkranz, D. M. Smith, H. L. Miears, M. Oueld Es Cheikh, M. Z. Lund, J. C. Young, P. N. Reardon, M. Okon, S. L. Smirnov and J. M. Antos, Sortase-mediated segmental labeling: a method for segmental assignment of intrinsically disordered regions in proteins, *PLoS One*, 2021, **16**, e0258531.

95 L. Freiburger, M. Sonntag, J. Hennig, J. Li, P. J. Zou and M. Sattler, Efficient segmental isotope labeling of multi-domain proteins using Sortase A, *J. Biomol. NMR*, 2015, **63**, 1–8.

96 C. P. Guimaraes, M. D. Witte, C. S. Theile, G. Bozkurt, L. Kundrat, A. E. Blom and H. L. Ploegh, Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions, *Nat. Protoc.*, 2013, **8**, 1787–1799.

97 K. Moritsugu, T. Terada and A. Kidera, Disorder-to-order transition of an intrinsically disordered region of sortase revealed by multiscale enhanced sampling, *J. Am. Chem. Soc.*, 2012, **134**, 7094–7101.

98 A. Bah and J. D. Forman-Kay, Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications, *J. Biol. Chem.*, 2016, **291**, 6696–6705.

99 M. Hofweber and D. Dormann, Friend or foe-Post-translational modifications as regulators of phase separation and RNP granule dynamics, *J. Biol. Chem.*, 2019, **294**, 7137–7150.

100 A. Savastano, D. Flores, H. Kadavath, J. Biernat, E. Mandelkow and M. Zweckstetter, Disease-Associated Tau Phosphorylation Hinders Tubulin Assembly within Tau Condensates, *Angew. Chem., Int. Ed.*, 2021, **60**, 726–730.

101 M. Saito, D. Hess, J. Eglinger, A. W. Fritsch, M. Kreysing, B. T. Weinert, C. Choudhary and P. Matthias, Acetylation of intrinsically disordered regions regulates phase separation, *Nat. Chem. Biol.*, 2019, **15**, 51–61.

102 E. T. Usher, K. E. W. Namitz, M. S. Cosgrove and S. A. Showalter, Probing multiple enzymatic methylation events in real time with NMR spectroscopy, *Biophys. J.*, 2021, **120**, 4710–4721.

103 A. I. Green and G. M. Burslem, Focused Libraries for Epigenetic Drug Discovery: The Importance of Isosteres, *J. Med. Chem.*, 2021, **64**, 7231–7240.

104 L. A. Gruijs da Silva, F. Simonetti, S. Hutten, H. Riemenschneider, E. L. Sternburg, L. M. Pietrek, J. Gebel, V. Dotsch, D. Edbauer, G. Hummer, L. S. Stelzl and D. Dormann, Disease-linked TDP-43 hyperphosphorylation suppresses TDP-43 condensation and aggregation, *EMBO J.*, 2022, **41**, e108443.

105 X. Luo, G. Fu, R. E. Wang, X. Zhu, C. Zambaldo, R. Liu, T. Liu, X. Lyu, J. Du, W. Xuan, A. Yao, S. A. Reed, M. Kang, Y. Zhang, H. Guo, C. Huang, P. Y. Yang, I. A. Wilson, P. G. Schultz and F. Wang, Genetically encoding phosphotyrosine and its nonhydrolyzable analog in bacteria, *Nat. Chem. Biol.*, 2017, **13**, 845–849.

106 S. J. Elsasser, R. J. Ernst, O. S. Walker and J. W. Chin, Genetic code expansion in stable cell lines enables encoded chromatin modification, *Nat. Methods*, 2016, **13**, 158–164.

107 M. D. Simon, F. Chu, L. R. Racki, C. C. de la Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar and K. M. Shokat, The site-specific installation of methyl-lysine analogs into recombinant histones, *Cell*, 2007, **128**, 1003–1012.

108 M. Shoaib, Q. Chen, X. Shi, N. Nair, C. Prasanna, R. Yang, D. Walter, K. S. Frederiksen, H. Einarsson, J. P. Svensson, C. F. Liu, K. Ekwall, M. Lerdrup, L. Nordenskiold and C. S. Sorensen, Histone H4 lysine 20 mono-methylation directly facilitates chromatin openness and promotes transcription of housekeeping genes, *Nat. Commun.*, 2021, **12**, 4800.

109 K. Hiragami-Hamada, S. Soeroes, M. Nikolov, B. Wilkins, S. Kreuz, C. Chen, I. A. De La Rosa-Velazquez, H. M. Zenn, N. Kost, W. Pohl, A. Chernev, D. Schwarzer, T. Jenuwein, M. Lorincz, B. Zimmermann, P. J. Walla, H. Neumann, T. Baubec, H. Urlaub and W. Fischle, Dynamic and flexible H3K9me3 bridging via HP1beta dimerization establishes a plastic state of condensed chromatin, *Nat. Commun.*, 2016, **7**, 11310.

110 B. E. Ackermann and G. T. Debelouchina, Heterochromatin Protein HP1alpha Gelation Dynamics Revealed by Solid-State NMR Spectroscopy, *Angew. Chem., Int. Ed.*, 2019, **58**, 6300–6305.

111 R. Huang, M. A. Holbert, M. K. Tarrant, S. Curtet, D. R. Colquhoun, B. M. Dancy, B. C. Dancy, Y. Hwang, Y. Tang, K. Meeth, R. Marmorstein, R. N. Cole, S. Khochbin and P. A. Cole, Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation, *J. Am. Chem. Soc.*, 2010, **132**, 9986–9987.

112 F. Li, A. Allahverdi, R. Yang, G. B. Lua, X. Zhang, Y. Cao, N. Korolev, L. Nordenskiold and C. F. Liu, A direct method for site-specific protein acetylation, *Angew. Chem., Int. Ed.*, 2011, **50**, 9611–9614.

113 D. D. Le, A. T. Cortesi, S. A. Myers, A. L. Burlingame and D. G. Fujimori, Site-specific and regiospecific installation of methylarginine analogues into recombinant histones and insights into effector protein binding, *J. Am. Chem. Soc.*, 2013, **135**, 2879–2882.

114 C. Chatterjee, R. K. McGinty, B. Fierz and T. W. Muir, Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation, *Nat. Chem. Biol.*, 2010, **6**, 267–269.

115 G. T. Debelouchina, K. Gerecht and T. W. Muir, Ubiquitin utilizes an acidic surface patch to alter chromatin structure, *Nat. Chem. Biol.*, 2017, **13**, 105–110.

116 P. E. Dawson, T. W. Muir, I. Clark-Lewis and S. B. Kent, Synthesis of proteins by native chemical ligation, *Science*, 1994, **266**, 776–779.
117 R. E. Thompson and T. W. Muir, Chemoenzymatic Semisynthesis of Proteins, *Chem. Rev.*, 2020, **120**, 3051–3126.
118 T. W. Muir, D. Sondhi and P. A. Cole, Expressed protein ligation: a general method for protein engineering, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 6705–6710.
119 V. Agouridas, O. El Mahdi, V. Diemer, M. Cargoet, J. M. Monbaliu and O. Melnyk, Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations, *Chem. Rev.*, 2019, **119**, 7328–7443.
120 O. Fuchs, S. Trunschke, H. Hanebrink, M. Reimann and O. Seitz, Enabling Cysteine-Free Native Chemical Ligation at Challenging Junctions with a Ligation Auxiliary Capable of Base Catalysis, *Angew. Chem., Int. Ed.*, 2021, **60**, 19483–19490.
121 G. M. Fang, Y. M. Li, F. Shen, Y. C. Huang, J. B. Li, Y. Lin, H. K. Cui and L. Liu, Protein chemical synthesis by ligation of peptide hydrazides, *Angew. Chem., Int. Ed.*, 2011, **50**, 7645–7649.
122 J. C. Shimko, C. J. Howard, M. G. Poirier and J. J. Ottesen, Preparing semisynthetic and fully synthetic histones h3 and h4 to modify the nucleosome core, *Methods Mol. Biol.*, 2013, **981**, 177–192.
123 E. J. Ge, K. S. Jani, K. L. Diehl, M. M. Muller and T. W. Muir, Nucleation and Propagation of Heterochromatin by the Histone Methyltransferase PRC2: Geometric Constraints and Impact of the Regulatory Subunit JARID2, *J. Am. Chem. Soc.*, 2019, **141**, 15029–15039.
124 M. R. Pratt, T. Abeywardana and N. P. Marotta, Synthetic Proteins and Peptides for the Direct Interrogation of alpha-Synuclein Posttranslational Modifications, *Biomolecules*, 2015, **5**, 1210–1227.
125 M. Hejjaoui, S. Butterfield, B. Fauvet, F. Vercruysse, J. Cui, I. Dikiy, M. Prudent, D. Olschewski, Y. Zhang, D. Eliezer and H. A. Lashuel, Elucidating the role of C-terminal post-translational modifications using protein semisynthesis strategies: alpha-synuclein phosphorylation at tyrosine 125, *J. Am. Chem. Soc.*, 2012, **134**, 5196–5210.
126 D. Ellmer, M. Brehs, M. Haj-Yahya, H. A. Lashuel and C. F. W. Becker, Single Posttranslational Modifications in the Central Repeat Domains of Tau4 Impact its Aggregation and Tubulin Binding, *Angew. Chem., Int. Ed.*, 2019, **58**, 1616–1620.
127 M. Haj-Yahya and H. A. Lashuel, Protein Semisynthesis Provides Access to Tau Disease-Associated Post-translational Modifications (PTMs) and Paves the Way to Deciphering the Tau PTM Code in Health and Diseased States, *J. Am. Chem. Soc.*, 2018, **140**, 6611–6621.
128 B. Pan, J. H. Park, T. Ramlall, D. Eliezer, E. Rhoades and E. J. Petersson, Chemoenzymatic Semi-synthesis Enables Efficient Production of Isotopically Labeled alpha-Synuclein with Site-Specific Tyrosine Phosphorylation, *ChemBioChem*, 2021, **22**, 1440–1447.
129 H. V. Erkizan, Y. Kong, M. Merchant, S. Schlottmann, J. S. Barber-Rotenberg, L. Yuan, O. D. Abaan, T. H. Chou, S. Dakshanamurthy, M. L. Brown, A. Uren and J. A. Toretsky, A small molecule blocking oncogenic protein EWS-FLI1 interaction with RNA helicase A inhibits growth of Ewing's sarcoma, *Nat. Med.*, 2009, **15**, 750–756.
130 N. Krishnan, D. Koveal, D. H. Miller, B. Xue, S. D. Akshinthala, J. Kragelj, M. R. Jensen, C. M. Gauss, R. Page, M. Blackledge, S. K. Muthuswamy, W. Peti and N. K. Tonks, Targeting the disordered C terminus of PTP1B with an allosteric inhibitor, *Nat. Chem. Biol.*, 2014, **10**, 558–566.
131 X. Yin, C. Giap, J. S. Lazo and E. V. Prochownik, Low molecular weight inhibitors of Myc-Max interaction and function, *Oncogene*, 2003, **22**, 6151–6159.
132 C. Yu, X. Niu, F. Jin, Z. Liu, C. Jin and L. Lai, Structure-based Inhibitor Design for the Intrinsically Disordered Protein c-Myc, *Sci. Rep.*, 2016, **6**, 22298.
133 M. Convertino, A. Vitalis and A. Caflisch, Disordered binding of small molecules to Abeta(12-28), *J. Biol. Chem.*, 2011, **286**, 41578–41588.
134 G. Toth, T. Neumann, A. Berthet, E. Masliah, B. Spencer, J. Tao, M. F. Jobling, S. J. Gardai, C. W. Bertoncini, N. Cremades, M. Bova, S. Ballaron, X. H. Chen, W. Mao, P. Nguyen, M. C. Tabios, M. A. Tambe, J. C. Rochet, H. D. Junker, D. Schwizer, R. Sekul, I. Ott, J. P. Anderson, B. Szoke, W. Hoffman, J. Christodoulou, T. Yednock, D. A. Agard, D. Schenk and L. McConlogue, Novel Small Molecules Targeting the Intrinsically Disordered Structural Ensemble of alpha-Synuclein Protect Against Diverse alpha-Synuclein Mediated Dysfunctions, *Sci. Rep.*, 2019, **9**, 16947.
135 L. I. Iconaru, S. Das, A. Nourse, A. A. Shelat, J. Zuo and R. W. Kriwacki, Small Molecule Sequestration of the Intrinsically Disordered Protein, p27(Kip1), Within Soluble Oligomers, *J. Mol. Biol.*, 2021, **433**, 167120.
136 Y. Cheng, T. LeGall, C. J. Oldfield, J. P. Mueller, Y. Y. Van, P. Romero, M. S. Cortese, V. N. Uversky and A. K. Dunker, Rational drug design via intrinsically disordered protein, *Trends Biotechnol.*, 2006, **24**, 435–442.
137 K. Ribbeck and D. Gorlich, The permeability barrier of nuclear pore complexes appears to operate via hydrophobic exclusion, *EMBO J.*, 2002, **21**, 2664–2671.
138 N. Nag, S. Sasidharan, V. N. Uversky, P. Saudagar and T. Tripathi, Phase separation of FG-nucleoporins in nuclear pore complexes, *Biochim. Biophys. Acta, Mol. Cell Res.*, 2022, **1869**, 119205.
139 N. Shulga and D. S. Goldfarb, Binding dynamics of structural nucleoporins govern nuclear pore complex permeability and may mediate channel gating, *Mol. Cell. Biol.*, 2003, **23**, 534–542.
140 S. Kroschwald, S. Maharana and A. Simon, Hexanediol: a chemical probe to investigate the material properties of membrane-less compartments, *Matters*, 2017, **3**, e201702000010.

141 Y. Lin, E. Mori, M. Kato, S. Xiang, L. Wu, I. Kwon and S. L. McKnight, Toxic PR Poly-Dipeptides Encoded by the C9orf72 Repeat Expansion Target LC Domain Polymers, *Cell*, 2016, **167**, 789–802.

142 J. R. Wheeler, T. Matheny, S. Jain, R. Abrisch and R. Parker, Distinct stages in stress granule assembly and disassembly, *Elife*, 2016, **5**, e18413.

143 M. Shi, K. You, T. Chen, C. Hou, Z. Liang, M. Liu, J. Wang, T. Wei, J. Qin, Y. Chen, M. Q. Zhang and T. Li, Quantifying the phase separation property of chromatin-associated proteins under physiological conditions using an anti-1,6-hexanediol index, *Genome Biol.*, 2021, **22**, 229.

144 R. Duster, I. H. Kaltheuner, M. Schmitz and M. Geyer, 1,6-Hexanediol, commonly used to dissolve liquid-liquid phase separated condensates, directly impairs kinase and phosphatase activities, *J. Biol. Chem.*, 2021, **296**, 100260.

145 R. Meduri, L. S. Rubio, S. Mohajan and D. S. Gross, Phase-separation antagonists potently inhibit transcription and broadly increase nucleosome density, *J. Biol. Chem.*, 2022, 102365, DOI: **10.1016/j.jbc.2022.102365**.

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

147 S. S. Patel, B. J. Belmont, J. M. Sante and M. F. Rexach, Natively unfolded nucleoporins gate protein diffusion across the nuclear pore complex, *Cell*, 2007, **129**, 83–96.

148 A. Patel, L. Malinovska, S. Saha, J. Wang, S. Alberti, Y. Krishnan and A. A. Hyman, ATP as a biological hydrotrope, *Science*, 2017, **356**, 753–756.

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

150 W. M. Babinchak, B. K. Dumm, S. Venus, S. Boyko, A. A. Putnam, E. Jankowsky and W. K. Surewicz, Small molecules as potent biphasic modulators of protein liquid-liquid phase separation, *Nat. Commun.*, 2020, **11**, 5574.

151 S. Jonchhe, W. Pan, P. Pokhrel and H. Mao, Small Molecules Modulate Liquid-to-Solid Transitions in Phase-Separated Tau Condensates, *Angew. Chem., Int. Ed.*, 2022, **61**, e202113156.

152 V. I. Stsiapura, A. A. Maskevich, V. A. Kuzmitsky, V. N. Uversky, I. M. Kuznetsova and K. K. Turoverov, Thioflavin T as a molecular rotor: fluorescent properties of thioflavin T in solvents with different viscosity, *J. Phys. Chem. B*, 2008, **112**, 15893–15902.

153 C. Q. Liang, L. Wang, Y. Y. Luo, Q. Q. Li and Y. M. Li, Capturing protein droplets: label-free visualization and detection of protein liquid-liquid phase separation with an aggregation-induced emission fluorogen, *Chem. Commun.*, 2021, **57**, 3805–3808.

154 S. F. Banani, A. M. Rice, W. B. Peeples, Y. Lin, S. Jain, R. Parker and M. K. Rosen, Compositional Control of Phase-Separated Cellular Bodies, *Cell*, 2016, **166**, 651–663.

155 I. A. Klein, A. Boija, L. K. Afeyan, S. W. Hawken, M. Fan, A. Dall'Agnese, O. Oksuz, J. E. Henninger, K. Shrinivas, B. R. Sabari, I. Sagi, V. E. Clark, J. M. Platt, M. Kar, P. M. McCall, A. V. Zamudio, J. C. Manteiga, E. L. Coffey, C. H. Li, N. M. Hannett, Y. E. Guo, T. M. Decker, T. I. Lee, T. Zhang, J. K. Weng, D. J. Taatjes, A. Chakraborty, P. A. Sharp, Y. T. Chang, A. A. Hyman, N. S. Gray and R. A. Young, Partitioning of cancer therapeutics in nuclear condensates, *Science*, 2020, **368**, 1386–1392.

156 D. Ban, L. I. Iconaru, A. Ramanathan, J. Zuo and R. W. Kriwacki, A Small Molecule Causes a Population Shift in the Conformational Landscape of an Intrinsically Disordered Protein, *J. Am. Chem. Soc.*, 2017, **139**, 13692–13700.

157 P. Herrera-Nieto, A. Perez and G. De Fabritiis, Small Molecule Modulation of Intrinsically Disordered Proteins Using Molecular Dynamics Simulations, *J. Chem. Inf. Model.*, 2020, **60**, 5003–5010.

158 G. T. Heller, P. Sormanni and M. Vendruscolo, Targeting disordered proteins with small molecules using entropy, *Trends Biochem. Sci.*, 2015, **40**, 491–496.

159 G. T. Heller, M. Bonomi and M. Vendruscolo, Structural Ensemble Modulation upon Small-Molecule Binding to Disordered Proteins, *J. Mol. Biol.*, 2018, **430**, 2288–2292.

160 A. E. Conicella, G. L. Dignon, G. H. Zerze, H. B. Schmidt, A. M. D'Ordine, Y. C. Kim, R. Rohatgi, Y. M. Ayala, J. Mittal and N. L. Fawzi, TDP-43 alpha-helical structure tunes liquid-liquid phase separation and function, *Proc. Natl. Acad. Sci. U. S. A.*, 2020, **117**, 5883–5894.

161 J. Zhu, X. Salvatella and P. Robustelli, Small molecules targeting the disordered transactivation domain of the androgen receptor induce the formation of collapsed helical states, *Nat. Commun.*, 2022, **13**, 6390.

162 G. Toth, S. J. Gardai, W. Zago, C. W. Bertoncini, N. Cremades, S. L. Roy, M. A. Tambe, J. C. Rochet, C. Galvagnion, G. Skibinski, S. Finkbeiner, M. Bova, K. Regnstrom, S. S. Chiou, J. Johnston, K. Callaway, J. P. Anderson, M. F. Jobling, A. K. Buell, T. A. Yednock, T. P. Knowles, M. Vendruscolo, J. Christodoulou, C. M. Dobson, D. Schenk and L. McConlogue, Targeting the intrinsically disordered structural ensemble of alpha-synuclein by small molecules as a potential therapeutic strategy for Parkinson's disease, *PLoS One*, 2014, **9**, e87133.

163 L. I. Iconaru, D. Ban, K. Bharatham, A. Ramanathan, W. Zhang, A. A. Shelat, J. Zuo and R. W. Kriwacki, Discovery of Small Molecules that Inhibit the Disordered Protein, p27(Kip1), *Sci. Rep.*, 2015, **5**, 15686.

164 G. Klebe, Applying thermodynamic profiling in lead finding and optimization, *Nat. Rev. Drug Discovery*, 2015, **14**, 95–110.

165 G. T. Heller, F. A. Aprile, T. C. T. Michaels, R. Limbocker, M. Perni, F. S. Ruggeri, B. Mannini, T. Lohr, M. Bonomi, C. Camilloni, A. De Simone, I. C. Felli, R. Pierattelli, T. P. J. Knowles, C. M. Dobson and M. Vendruscolo, Small-molecule sequestration of amyloid-beta as a drug discovery strategy for Alzheimer's disease, *Sci. Adv.*, 2020, **6**(45), eabb5924.

166 J. Pujols, S. Pena-Diaz, D. F. Lazaro, F. Peccati, F. Pinheiro, D. Gonzalez, A. Carija, S. Navarro, M. Conde-Gimenez,

J. Garcia, S. Guardiola, E. Giralt, X. Salvatella, J. Sancho, M. Sodupe, T. F. Outeiro, E. Dalfo and S. Ventura, Small molecule inhibits alpha-synuclein aggregation, disrupts amyloid fibrils, and prevents degeneration of dopaminergic neurons, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 10481–10486.

167 B. Xu, X. Mo, J. Chen, H. Yu and Y. Liu, Myricetin Inhibits alpha-Synuclein Amyloid Aggregation by Delaying the Liquid-to-Solid Phase Transition, *Chembiochem*, 2022, **23**, e202200216.

168 P. Robustelli, A. Ibanez-de-Opakua, C. Campbell-Bezat, F. Giordanetto, S. Becker, M. Zweckstetter, A. C. Pan and D. E. Shaw, Molecular Basis of Small-Molecule Binding to alpha-Synuclein, *J. Am. Chem. Soc.*, 2022, **144**, 2501–2510.

169 M. Kato, T. W. Han, S. Xie, K. Shi, X. Du, L. C. Wu, H. Mirzaei, E. J. Goldsmith, J. Longgood, J. Pei, N. V. Grishin, D. E. Frantz, J. W. Schneider, S. Chen, L. Li, M. R. Sawaya, D. Eisenberg, R. Tycko and S. L. McKnight, Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels, *Cell*, 2012, **149**, 753–767.

170 P. Romero, Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guilliot and A. K. Dunker, Thousands of proteins likely to have long disordered regions, *Pac. Symp. Biocomput.*, 1998, 437–448.

171 H. J. Dyson and P. E. Wright, NMR illuminates intrinsic disorder, *Curr. Opin. Struct. Biol.*, 2021, **70**, 44–52.

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

173 J. M. Plitzko, B. Schuler and P. Selenko, Structural Biology outside the box-inside the cell, *Curr. Opin. Struct. Biol.*, 2017, **46**, 110–121.

174 E. Luchinat and L. Banci, In-cell NMR: From target structure and dynamics to drug screening, *Curr. Opin. Struct. Biol.*, 2022, **74**, 102374.

175 Y. David, M. Vila-Perello, S. Verma and T. W. Muir, Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins, *Nat. Chem.*, 2015, **7**, 394–402.

176 J. A. Gramespacher, A. J. Stevens, D. P. Nguyen, J. W. Chin and T. W. Muir, Intein Zymogens: Conditional Assembly and Splicing of Split Inteins via Targeted Proteolysis, *J. Am. Chem. Soc.*, 2017, **139**, 8074–8077.

177 G. L. Dignon, W. Zheng, Y. C. Kim, R. B. Best and J. Mittal, Sequence determinants of protein phase behavior from a coarse-grained model, *PLoS Comput. Biol.*, 2018, **14**, e1005941.

178 Z. Benayad, S. von Bulow, L. S. Stelzl and G. Hummer, Simulation of FUS Protein Condensates with an Adapted Coarse-Grained Model, *J. Chem. Theory Comput.*, 2021, **17**, 525–537.

179 A. P. Latham and B. Zhang, Unifying coarse-grained force fields for folded and disordered proteins, *Curr. Opin. Struct. Biol.*, 2022, **72**, 63–70.

180 J. A. Joseph, A. Reinhardt, A. Aguirre, P. Y. Chew, K. O. Russell, J. R. Espinosa, A. Garaizar and R. Collepardo-Guevara, Physics-driven coarse-grained model for biomolecular phase separation with near-quantitative accuracy, *Nat. Comput. Sci.*, 2021, **1**, 732–743.

181 M. Paloni, R. Bailly, L. Ciandrini and A. Barducci, Unraveling Molecular Interactions in Liquid-Liquid Phase Separation of Disordered Proteins by Atomistic Simulations, *J. Phys. Chem. B*, 2020, **124**, 9009–9016.

182 W. Zheng, G. L. Dignon, N. Jovic, X. Xu, R. M. Regy, N. L. Fawzi, Y. C. Kim, R. B. Best and J. Mittal, Molecular Details of Protein Condensates Probed by Microsecond Long Atomistic Simulations, *J. Phys. Chem. B*, 2020, **124**, 11671–11679.

183 N. M. Kanaan, C. Hamel, T. Grabinski and B. Combs, Liquid-liquid phase separation induces pathogenic tau conformations in vitro, *Nat. Commun.*, 2020, **11**, 2809.

184 J. E. Hoffmann, T. Plass, I. Nikic, I. V. Aramburu, C. Koehler, H. Gillandt, E. A. Lemke and C. Schultz, Highly Stable trans-Cyclooctene Amino Acids for Live-Cell Labeling, *Chem.–Eur. J.*, 2015, **21**, 12266–12270.

185 P. Widder, F. Berner, D. Summerer and M. Drescher, Double Nitroxide Labeling by Copper-Catalyzed Azide-Alkyne Cycloadditions with Noncanonical Amino Acids for Electron Paramagnetic Resonance Spectroscopy, *ACS Chem. Biol.*, 2019, **14**, 839–844.

186 Y. Ge, L. Chen, S. Liu, J. Zhao, H. Zhang and P. R. Chen, Enzyme-Mediated Intercellular Proximity Labeling for Detecting Cell-Cell Interactions, *J. Am. Chem. Soc.*, 2019, **141**, 1833–1837.

187 B. Fierz, C. Chatterjee, R. K. McGinty, M. Bar-Dagan, D. P. Raleigh and T. W. Muir, Histone H2B ubiquitylation disrupts local and higher-order chromatin compaction, *Nat. Chem. Biol.*, 2011, **7**, 113–119.

188 B. Pan, E. Rhoades and E. J. Petersson, Chemoenzymatic Semisynthesis of Phosphorylated alpha-Synuclein Enables Identification of a Bidirectional Effect on Fibril Formation, *ACS Chem. Biol.*, 2020, **15**, 640–645.

ACKNOWLEDGEMENTS

Chapter 5, in full, is a reprint of the material as it appears in *Chemical Science*: Berkeley, R. F. & Debelouchina, G. T. Chemical tools for study and modulation of biomolecular phase transitions. *Chem Sci* **13**, 14226–14245 (2022). The dissertation author was the primary investigator and author of this material.