

Washington University in St. Louis

## Washington University Open Scholarship

---

Engineering and Applied Science Theses &  
Dissertations

McKelvey School of Engineering

---

Winter 1-15-2021

# Uncovering the Roles and Evolved Sequence Grammar of Hypervariable Intrinsically Disordered Proteins in Bacterial Cell Division

Megan Cohan

*Washington University in St. Louis*

Follow this and additional works at: [https://openscholarship.wustl.edu/eng\\_etds](https://openscholarship.wustl.edu/eng_etds)



Part of the [Biology Commons](#), [Biomedical Engineering and Bioengineering Commons](#), and the [Biophysics Commons](#)

---

### Recommended Citation

Cohan, Megan, "Uncovering the Roles and Evolved Sequence Grammar of Hypervariable Intrinsically Disordered Proteins in Bacterial Cell Division" (2021). *Engineering and Applied Science Theses & Dissertations*. 606.

[https://openscholarship.wustl.edu/eng\\_etds/606](https://openscholarship.wustl.edu/eng_etds/606)

This Dissertation is brought to you for free and open access by the McKelvey School of Engineering at Washington University Open Scholarship. It has been accepted for inclusion in Engineering and Applied Science Theses & Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact [digital@wumail.wustl.edu](mailto:digital@wumail.wustl.edu).

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering

Department of Biomedical Engineering

Dissertation Examination Committee:

Rohit V. Pappu, Chair

Alex S. Holehouse

Petra A. Levin

Timothy M. Lohman

Andrea Soranno

Michael D. Vahey

Uncovering the Roles and Evolved Sequence Grammar of Hypervariable Intrinsically Disordered  
Proteins in Bacterial Cell Division

by

Megan Claire Cohan

A dissertation presented to  
The Graduate School  
of Washington University in  
partial fulfillment of the  
requirements for the degree  
of Doctor of Philosophy

January 2021  
St. Louis, Missouri

© 2021, Megan Claire Cohan

# Table of Contents

List of Figures.....	vi
List of Tables .....	x
List of Abbreviations .....	xi
Acknowledgments .....	xiii
Abstract of the Dissertation .....	xvii
<b>1 Introduction.....</b>	<b>1</b>
1.1 Preamble .....	1
1.2 Features of bacterial cell division protein FtsZ.....	3
1.2.1 Bacterial cell division and FtsZ.....	3
1.2.2 Dynamics and assembly of FtsZ in the Z-ring .....	7
1.2.3 Assembly of FtsZ <i>in vitro</i> .....	9
1.2.4 Architecture of FtsZ.....	13
1.3 Intrinsically Disordered Proteins/Regions .....	20
1.3.1 Definition of Intrinsically Disordered Regions .....	20
1.3.2 Hypervariability in intrinsically disordered regions.....	21
1.3.3 Decoding the sequence-to-function relationship of IDRs .....	22
1.3.4 IDPs in bacteria.....	24
1.3.5 Disorder in bacterial cell division .....	25
1.4 Scope of the Thesis .....	27
1.5 References.....	31
1.6 Appendix.....	49
<b>2 Dissecting the functional contributions of the intrinsically disordered C-terminal tail of <i>B. subtilis</i> FtsZ.....</b>	<b>50</b>

2.1	Preamble .....	50
2.2	Introduction.....	52
2.3	Material and Methods .....	56
2.3.1	Protein expression and purification.....	56
2.3.2	Dynamic Light Scattering (DLS) assay .....	57
2.3.3	Transmission Electron Microscopy (TEM).....	57
2.3.4	GTPase assay .....	58
2.3.5	Determination of $c_A$ .....	58
2.3.6	Michaelis-Menten fits .....	59
2.4	Results.....	60
2.4.1	The CTL and the CTP modules have different effects on Bs-FtsZ assembly .....	60
2.4.2	Increased salt concentration weakens the formation of higher-order assemblies for WT Bs-FtsZ.....	70
2.4.3	The CTL weakens the driving forces for forming active polymers .....	73
2.4.4	The CTL and CTP have opposing effects on the enzymatic activity .....	76
2.5	Discussion .....	80
2.6	References.....	88
2.7	Appendix.....	99
<b>3</b>	<b>Information theoretic measures for quantifying sequence-ensemble relationships of intrinsically disordered proteins.....</b>	<b>101</b>
3.1	Preamble .....	101
3.2	Introduction.....	103
3.3	Material and Methods .....	108
3.3.1	Simulations of sequences of FtsZ-CTTs .....	108
3.4	Results.....	109
3.4.1	Global parameters that define conformational ensembles.....	109
3.4.2	The ensemble entropy matrix approach .....	112
3.4.3	Generating comparative assessments of SERs .....	116

3.4.4	Comparative assessments of SERs across a set of sequences of identical lengths and amino acid compositions.....	117
3.4.5	Comparison of the difference ensemble entropy matrices $\Delta AB$ for the PT8s-WT and PT5s-WT pairs	120
3.4.6	Quantitative SERs for IDRs derived from the same functional family across orthologs .....	122
3.5	Discussion .....	128
3.6	References .....	132
3.7	Appendix.....	143
<b>4</b>	<b>Uncovering non-random sequence patterns encoded in sequences of hypervariable IDPs / IDRs.....</b>	<b>145</b>
4.1	Preamble .....	145
4.2	Introduction.....	147
4.3	The z-score method.....	152
4.3.1	Null model generation.....	152
4.3.2	Calculation of patterning parameters .....	153
4.3.3	Development of the z-score matrix .....	154
4.4	Application of the z-score method.....	157
4.4.1	Application of the z-score matrix method to delineate random and non-random patterning features	157
4.4.2	Using the z-score matrix method to connect sequence-encoded information to IDR function and bacterial phenotype .....	160
4.5	Discussion .....	174
4.6	References .....	180
4.7	Appendix.....	187
<b>5</b>	<b>Impact of the molecular grammar of the <i>B. subtilis</i> FtsZ C-terminal linker on function and bacterial cell division.....</b>	<b>199</b>
5.1	Preamble .....	199
5.2	Introduction.....	201

5.3	Material and Methods .....	205
5.3.1	General methods .....	205
5.3.2	Cloning CTL variants.....	205
5.3.3	Immunoblotting.....	206
5.3.4	Growth Curves & Immunofluorescence microscopy .....	207
5.3.5	All-Atom Simulations of CTT Sequence Variants.....	208
5.3.6	Protein Purification .....	208
5.3.7	90° Light Scattering Assay .....	209
5.3.8	Transmission Electron Microscopy (TEM).....	209
5.3.9	GTPase Activity Assay .....	210
5.3.10	Fluorescence Correlation Spectroscopy (FCS).....	210
5.3.11	Shannon Entropy Calculation.....	211
5.4	Results.....	212
5.4.1	Design of C-terminal linker variants .....	212
5.4.2	CTT- $\kappa$ influences the conformational properties of CTT sequences .....	214
5.4.3	Designed set of CTL variants includes random and non-random sequence features .....	218
5.4.4	Increasing the CTT $\kappa$ value converts the CTL into a sticker.....	220
5.4.5	CTL variants that do not meet the criteria suggested for a functional CTL disrupt cell division .....	222
5.4.6	CTL variants that have sticker-like properties also disrupt function <i>in vitro</i> .....	228
5.4.7	Linear sequence patterning can explain discrepancies between spacer CTLs.....	231
5.5	Discussion .....	235
5.6	References.....	242
5.7	Appendix.....	251
<b>6</b>	<b>Making the case for disordered proteins and biomolecular condensates in bacteria....</b>	<b>262</b>
6.1	Preamble .....	262
6.2	Introduction.....	264
6.2.1	Eukaryotic condensates appear to form via phase transitions .....	266
6.2.2	Scaffolds versus clients.....	267
6.2.3	Descriptors of the organization of molecular matter within condensates.....	268

6.3	Making a case for condensates in bacteria.....	270
6.3.1	Spatiotemporal control of bacterial cell division .....	270
6.3.2	Spatiotemporal control of polarity .....	272
6.3.3	Spatiotemporal control of transcription and post-transcriptional regulation.....	274
6.3.4	Management of phosphate levels and synthesis of multivalent phosphates.....	277
6.4	Discussion.....	279
6.5	References.....	282
<b>7</b>	<b>Concluding Remarks and Future Directions .....</b>	<b>293</b>
7.1	Preamble .....	293
7.2	Concluding Remarks.....	295
7.3	References.....	304
	<b>Curriculum Vitae.....</b>	<b>310</b>

## List of Figures

Figure 1.1:	Architecture of FtsZ includes three distinct domains.....	14
Figure 1.2:	Normalized conservation scores for different regions of FtsZ.....	15
Figure 1.3:	Disorder in the <i>B. subtilis</i> FtsZ interactome.....	26
Figure 2.1:	Design of C-terminal tail truncation variants .....	55
Figure 2.2:	Concentration-dependent assembly of FtsZ.....	61
Figure 2.3:	TEM image of WT FtsZ.....	63
Figure 2.4:	Concentration-dependent assembly of $\Delta$ CTT .....	64
Figure 2.5:	TEM image of long single-stranded protofilaments of $\Delta$ CTT .....	65
Figure 2.6:	Concentration-dependent assembly of $\Delta$ CTP.....	66
Figure 2.7:	$\Delta$ CTP TEM image .....	67
Figure 2.8:	Concentration-dependent assembly of $\Delta$ CTL .....	68
Figure 2.9:	TEM image of $\Delta$ CTL .....	69
Figure 2.10:	Concentration-dependent assembly of WT with an increasing concentration of KCl .....	71
Figure 2.11:	TEM images of WT in high salt.....	72



Figure 2.12: Measurement of GTPase activity (millimoles of GTP hydrolyzed per minute) as a function of WT concentration.....	74
Figure 2.13: Measurement of GTPase activity (millimoles of GTP hydrolyzed per minute) as a function of CTT variant concentration .....	75
Figure 2.14: Bar plots of the apparent Michaelis-Menten constants .....	77
Figure 2.15: Bar plots of the catalytic rates ( $k_{cat}$ ) for each variant at 4 and 6 $\mu$ M.....	78
Figure 2.16: Bar plots of the catalytic efficiency ( $k_{eff}$ ) for each variant at 4 and 6 $\mu$ M .....	79
Figure 2A.1: Parameters that quantify enzyme activity were extracted using fits to a FtsZ Michaelis-Menten model .....	99
Figure 2A.2: Assessment of the fits of GTPase Activity versus GTP concentration obtained using a Hill model .....	100
Figure 3.1: Adaptation of a communication channel to describe protein design, focusing on IDP design.....	105
Figure 3.2: Illustration of the relevant global conformational parameters .....	110
Figure 3.3: Illustration of conformational features of IDPs/IDRs extracted from all-atom simulations.....	111
Figure 3.5: Example of a two-dimensional probability distribution of conformational properties used to quantify the information theoretic entropy.....	114
Figure 3.6: Example of an entropy matrix .....	116
Figure 3.7: Comparative assessments of SERs for the RAM regions of NICD variants.....	119
Figure 3.8: Comparison of the difference ensemble entropy matrices $\Delta AB$ for the PT8s-WT and PT5s-WT pairs.....	121
Figure 3.9: Scatter plot of CTT sequence parameters summarized in terms of CTT-length and the Fraction of Charged Residues (FCR) .....	123
Figure 3.10: Pairwise distances of sequences versus sequence ensemble entropy matrices .....	126
Figure 3.11: Histogram of SERs and sequence similarities.....	127
Figure 4.1: Histograms of the cosine similarities of the vectorized compositions .....	149
Figure 4.3: Depiction of a z-score matrix with eight residues / residue types.....	156
Figure 4.4: Percentage of CTL sequences that have at least one non-random feature versus those that are fully random.....	159
Figure 4.5: Z-score matrices of the FtsZ CTLs from <i>B. subtilis</i> , <i>E. coli</i> , and <i>C. crescentus</i> .....	161
Figure 4.6: Frequency of observing z-scores above 1.5 for each assessed feature.....	163
Figure 4.7: Direct comparison of z-score matrices from <i>C. crescentus</i> RNase E and <i>E. coli</i> RNase E.....	165
Figure 4.8: The feature-specific frequency of observing z-scores above 2 for the system of RNase E orthologs .....	167
Figure 4.9: <i>E. coli</i> SSB CTD z-score matrix .....	169
Figure 4.10: The feature-specific frequency of observing non-random features for the system of SSB orthologs .....	170

Figure 4.11: The frequency of observing a non-random value for each assessed feature for the C-terminal SSB IDLs.....	173
Figure 4.12: Z-score matrices of the disordered domains within the FtsZ interactome .....	179
Figure 4A.1: Fitting to a gamma distribution versus a normal distribution.....	187
Figure 4A.2: Goodness of fit of the gamma distribution as a function of the fraction of the residues of interest for $\kappa$ -based patterning parameters. ....	189
Figure 4A.3: The expected value for a given parameter depends on the sequence content .....	190
Figure 4A.4: The expected value of SCD depends on the sequence charge content and length	191
Figure 4A.5: Goodness of fit of the gamma distribution as a function of the fraction of the residues of interest for $\Omega$ -based patterning parameters .....	192
Figure 4A.6: The expected value for the $\Omega$ parameter depends on the sequence content .....	193
Figure 4A.7: Cumulative probability of observing a maximum z-score of a certain magnitude for each FtsZ CTL from the system of orthologs .....	194
Figure 4A.8: Average amino acid frequencies within the set of CTLs. ....	195
Figure 4A.9: Sequences of the CTDs from RNases E .....	196
Figure 4A.10: Cumulative probability of observing a maximum z-score of a certain magnitude for each RNase E CTD from the system of orthologs .....	197
Figure 4A.11: Cumulative probability of observing a maximum z-score of a certain magnitude for each SSB IDL from the system of orthologs .....	198
Figure 5.1: The hydrodynamic radius ( $R_H$ ) of TMR-labeled peptides of the WT and variant CTT sequences .....	215
Figure 5.2: Plots of length normalized values of the mean radii of gyration for each of the designed CTT variants .....	216
Figure 5.3: Plot of the relative Shannon entropy ( $s'$ ) versus $\kappa$ for the WT CTT and each of the designed CTT variants .....	218
Figure 5.4: Z-score matrices for the CTT variants and WT FtsZ .....	219
Figure 5.5: Apparent hydrodynamic diameter ( $D_H$ ) of the assemblies formed by WT and the CTT variants in the absence of GTP .....	221
Figure 5.6: Immunofluorescence micrographs of <i>B. subtilis</i> expressing the WT FtsZ (left) versus six different variants of FtsZ with redesigned CTL sequences .....	224
Figure 5.7: Length-to-Z-ring (L/R) ratio for <i>B. subtilis</i> expressing different FtsZ variants.....	225
Figure 5.8: Data for growth phenotypes are shown in terms of the normalized optical density at 600 nm (OD600).....	226
Figure 5.9: Growth profiles of the CTT- $\kappa$ variants co-expressed with native WT FtsZ .....	228
Figure 5.10: TEM images of morphologies obtained for k46 and k72.....	230
Figure 5.11: Relative GTPase Activities of the CTT variants that have phenotypic impacts ....	230
Figure 5.12: Relative SLS intensities of the CTT variants that supported bacterial cell division .....	232
Figure 5.13: Negative stain EM images of k40. ....	233

Figure 5.14: Relative GTPase Activities of the CTT variants that supported bacterial cell division .....	234
Figure 5A.1: Distribution of z-scores from the $\kappa_+$ - parameter values observed from a system of 1208 FtsZ orthologs .....	251
Figure 5A.2: Design space of CTL variants .....	252
Figure 5A.3: Two-dimensional conformational distributions from atomistic simulations.....	253
Figure 5A.4: DLS data in the absence of GTP including new variants .....	255
Figure 5A.5: Additional variants also follow the trend of compaction with increasing k value as measured by FCS .....	256
Figure 5A.6: Time dependent Western blot quantifying the protein levels of each variant and the native FtsZ in <i>B. subtilis</i> .....	257
Figure 5A.7: Relative SLS intensities values of CTT variants that behave as stickers .....	258
Figure 5A.8: Negative stain EM images of wild type and k40 in high salt conditions .....	259
Figure 5A.9: Impact of salt on wild type and k40 assembly as measured by SLS .....	260
Figure 5A.10: Z-score matrix for the randomly scrambled CTL sequence assessed in the study conducted by Buske and Levin.....	261
Figure 6.1: Conceptual depiction of intrinsic versus emergent sticker multivalence .....	265
Figure 6.2: Schematics to distinguish bond percolation without phase separation .....	267
Figure 6.3: Summary depiction of the exemplars of bacterial condensates discussed in this work .....	279

# List of Tables

Table 3.1: Summary of $\kappa$ values and parameters extracted from all atom simulations for the RAM region extracted from the WT and designed NICD variants .....	118
Table 3A.1: Sequence ID numbers and the UniProt ID.....	144
Table 5.1: Sequences of the FtsZ CTL variants with the conserved CTP motif .....	214
Table 5.2: Summary of the results of specific IDR parameters hypothesized to impact the function of <i>B. subtilis</i> FtsZ .....	222
Table 5.3: Summary of results from all experiments on FtsZ CTL variants compared to WT ..	236
Table 5A.1 Table of stains and plasmids referenced in this study.....	254
Table 5A.2: Additional variants considered for future experiments.....	255

# List of Abbreviations

<b>AFD</b> .....	autonomously foldable domain
<b>AFP</b> .....	autonomously foldable protein
<b>Bs-FtsZ</b> .....	<i>Bacillus subtilis</i> FtsZ
<b>Cc-FtsZ</b> .....	<i>Caulobacter crescentus</i> FtsZ
<b>Ec-FtsZ</b> .....	<i>Escherichia coli</i> FtsZ
<b>CDF</b> .....	cumulative distribution function
<b>Cryo-EM</b> .....	cryogenic electron microscopy
<b>CTL</b> .....	C-terminal linker
<b>CTP</b> .....	C-terminal peptide
<b>CTT</b> .....	C-terminal tail
<b>CTV</b> .....	C-terminal variable region
<b>DLS</b> .....	dynamic light scattering
<b>FCR</b> .....	fraction of charged residues
<b>FCS</b> .....	fluorescence correlation spectroscopy
<b>FtsZ</b> .....	Filamentous temperature sensitive mutant Z
<b>FtsZ-GDP</b> .....	FtsZ bound to GDP
<b>FtsZ-GTP</b> .....	FtsZ bound to GTP
<b>GDP</b> .....	guanosine diphosphate

**GTP** ..... guanosine triphosphate

**IDP**..... intrinsically disordered protein

**IDR** ..... intrinsically disordered region

**MES** ..... 2-(N-morpholino) ethanesulfonic acid

**NCPR** ..... net charge per residue

**NO** ..... nucleoid occlusion

**PDF** ..... probability distribution function

**PSBP**..... phase separated aided bond percolation

**SAXS** ..... small-angle X-ray scattering

**SCD**..... sequence charge decoration

**SER**..... sequence-ensemble relationship

**SLiM**..... short linear motif

**SLS** ..... static light scattering

**smFRET** ..... single molecule Förster resonance transfer

**SSB** ..... single-stranded DNA binding protein

**TEM** ..... transmission electron microscopy

**TIRF** ..... total internal reflection fluorescence

**WT**..... wild type (*Bacillus subtilis* FtsZ)

# Acknowledgments

First, I would like to thank my thesis committee, Alex S. Holehouse, Petra A. Levin, Timothy M. Lohman, Andrea Soranno, and Michael D. Vahey, for their help and guidance throughout my thesis work. Significantly, Tim Lohman's several insightful suggestions helped shape the body of this thesis, and Petra Levin inspired and cultivated my interest in and knowledge of FtsZ through our key collaboration. As also a past member of the Pappu Lab, Alex Holehouse has been an incredible mentor to me, and I am grateful that I was able to learn from and work with him throughout my entire graduate career.

I would also like to thank the Pappu Lab, both past and present members, for their collective support. The fellowship, feedback, and discussions were instrumental to my thesis work. The lab was also an incredible source of mentorship for me, and I would like to thank Drs. Tyler Harmon, Ammon Posey, and Kiersten Ruff for their help in shaping who I am today as a scientist. Ammon served as my first mentor in the lab; I am grateful for the time he took to teach me all about IDPs and for all of his help at the bench. To my conference room-buddy Kiersten, I am grateful to her for her friendship and for teaching me to challenge myself to think about the data in a different and often better way. I know we will climb many more mountains together. I also appreciate the significant assistance from Anna Eddelbuettel (now Hancock), who was an undergraduate in the lab, from Kasia Kornacki, our lab tech, and from individuals from Petra Levin's lab, with whom I closely collaborated. Of course, my work in this lab would not have been possible without the generous support of our funding sources, specifically the Jean and Sidney Grossman fellowship

and the National Science Foundation that has continued to support the Pappu lab's efforts on disordered proteins.

To those that I have met through the BALSAs (Biotechnology and Life Science Advising) Group, I am thankful for their support and encouragement through our shared community. BALSAs was a creative outlet for me during graduate school, and I am so grateful to have such enriching and meaningful friendships and experiences from my involvement as a director. I will miss starting every Wednesday with this group, and I cannot wait to see how the organization continues to evolve and grow. I would also like to thank my broader support group, including my local and far away friends. My workout buddies kept me sane through some of the most challenging times in my studies, allowing me space to decompress. I am so lucky to have a group of friends that are supportive and fun, with a diverse set of interests.

It is difficult to sum in a few words the gratitude that I feel to my PI, Rohit Pappu. The lab culture of support and scientific rigor is his creation, and I owe my growth and success in my graduate career to this environment. Rohit's friendship and mentorship have made a lasting impact on me, as he invested not only in my development as a scientist but also as a person. He constantly challenged me to advocate for myself and to aim high, practices that have ultimately helped me land my dream job. His dedication to his work and his family (both inside and outside of the lab) is nothing short of inspiring, and I am honored to have worked so closely with him. I won the lottery on mentors, and to say I will miss him is an incredible understatement.



To my family, I am overwhelmed with gratitude. The Whites have offered me a home away from home, and I am genuinely thankful for their constant encouragement. My parents are an endless source of inspiration, support, and love, and to them, I owe everything, especially for encouraging / allowing me to come to St. Louis in the first place to follow this path. From my father, I learned that everything worth doing deserves my best effort, and from my mother, I gained a true passion for learning and an eager spirit that always inspires me to take on the next challenge. Though, I am most grateful to them both for showing me that it is always the best to go through life with gentleness, humility, and humor. Wholeheartedly, I am at my best when I am my parents' daughter. As for my younger brother, I hope that when I grow up, I can be just like him – curious, intelligent, creative, and passionate. He will always give the best hugs, and I know that no matter what, he is there for me.

And last, but certainly not least, to the love of my life, Michael, for being my number one fan, advocate, supporter, and best friend. He always manages to know exactly what I need in a given moment, whether it is a sounding board, a devil's advocate, a motivational speaker, a coffee provider, or, most often, a source of a good laugh. It did not take long after I met him for me to realize that he is the real reason I ended up in St. Louis. Michael, you – and the life we have created together – are what truly matters, and I cannot wait for our next adventure.

Megan Claire Cohan

*Washington University in St. Louis*

*January 2021*

*Dedicated to my parents*

# **Abstract of the Dissertation**

Uncovering the Roles and Evolved Sequence Grammar of Hypervariable Intrinsically Disordered  
Proteins in Bacterial Cell Division

by

Megan Claire Cohan

Doctor of Philosophy in Biomedical Engineering

Washington University in St. Louis, 2020

Professor Rohit V. Pappu, Ph.D. Chair

Across all domains of life, a defining hallmark of the onset of cell division is the formation of a cytokinetic ring at the center of the cell. Cell division is a tightly controlled process that involves various regulatory factors that modulate the assembly of the cytokinetic ring. In rod-shaped bacteria, the ring is termed the Z-ring after the protein FtsZ, which is foundational to ring formation and is the bacterial homolog of tubulin. Like tubulin, FtsZ is an assembling GTPase, where GTP binding promotes the cooperative assembly into FtsZ polymers that laterally associate to form bundles. While the GTPase domain drives FtsZ polymerization, the formation of these higher-order structures requires domains outside the folded core. FtsZ has a bristled architecture, where a disordered tail, called the C-terminal tail (CTT), flanks the folded domain. The essentiality of the CTT was established through deletion experiments; however, the exact role that the CTT

plays within the context of FtsZ function remained unclear. Here, we establish that the CTT, containing an intra-and intermolecular interaction motif (CTP) and a disordered linker (CTL), has a sticker-and-spacer architecture, where the CTL modulates the interactions of the CTP. We find that the modules of the CTT not only influence FtsZ assembly but also impact the catalytic efficiency of the GTPase domain. These findings add to recent findings that implicate disordered regions tethered to enzymes in auto-regulatory activities.

The findings summarized above were obtained by focusing our investigations on the CTT of the FtsZ protein from *B. subtilis* (Bs-FtsZ). Is the stickers-and-spacers model applicable to understanding the functions of CTTs from other bacterial FtsZs? We analyzed the sequences of 1208 orthologous FtsZs, and the results show that while the CTPs and the core domains are reasonably well conserved, the CTLs are hypervariable across orthologs. The results of the sequence analysis have several implications: It might reflect a form of convergent evolution whereby different CTL sequences are interoperable with one other because different sequences serve the functionality of being spacers. Alternatively, the variation could be an example of divergent evolution, whereby changes to the CTLs engender different functionalities in different bacteria. Answering these questions will require methods to identify common sequence patterns across orthologous CTLs, and this cannot be achieved using traditional multiple sequence alignment approaches. Accordingly, we introduce computational methods that enable the quantitative analysis of conserved / distinct sequence-ensemble relationships across a family of IDRs. Additionally, we introduce a new method to uncover cryptic sequence patterns that define disordered regions as random versus non-random. These methods are shown to be applicable for high-throughput analysis of CTLs derived from different FtsZs. They are also effective in

uncovering sequence patterns that are cryptic but conserved in intrinsically disordered regions (IDRs) from other bacterial proteins.

Given the role of sequence-ensemble relationships and non-random motifs in IDP/IDR function, we hypothesized that these features might influence function and, therefore, might be encoded for within the amino acid sequence of the FtsZ CTL. This implies that designed CTL sequence variants that result in significant changes to these sequence features and to the ability of the CTL to function as a spacer could perturb function. To test this hypothesis, we developed scrambled sequence variants of the *B. subtilis* FtsZ CTL using the patterning of oppositely charged residues as a design parameter. Leveraging new methodologies, we found that the designed variants caused changes to the sequence-ensemble relationships, the non-random sequence patterning, and / or the spacer properties. Each variant was tested for complementary functions to wild type *in vitro* and *in vivo*. Indeed, deviations from wild type features had phenotypic impacts and / or influenced FtsZ assembly and activity, showing that the CTL is not a random disordered sequence but instead has specifically encoded sequence features that dictate function. As the global need to combat antibiotic-resistant infections continues to mount, studies that further understand the functions that IDRs contribute to essential bacterial processes such as cell division can be leveraged to create next-generation antibiotics.

# Chapter 1

## Introduction

### 1.1 Preamble

The focus of this thesis is on understanding how the sequence-encoded features of the intrinsically disordered C-terminal tail (CTT) of the FtsZ protein contribute to its functions in cytokinetic ring assembly and bacterial cell division. We develop and deploy experimental and computational methodologies that can apply to other disordered proteins. We focus specifically on intrinsically disordered regions (IDRs) that are not conserved in terms of both the linear amino acid sequence and the overall composition within their functional families. This opening chapter introduces the relevant concepts, broader scope, and the overall motivations underlying this dissertation. It is organized as follows: First, we present an introduction to bacterial cell division and the components involved, demonstrating the essential role of FtsZ throughout the process. Next, we describe the known molecular features of this protein and discuss recent work establishing the vital role of the disordered CTT. We then connect what is known regarding the CTT to recently discovered principles regarding sequence-ensemble-function relationships of intrinsically disordered proteins (IDPs) and illustrate how the CTT presents a unique and useful

model system for future assessments to generate progress in the IDP field. Finally, we outline the scope of the thesis.

## 1.2 Features of bacterial cell division protein FtsZ

### 1.2.1 Bacterial cell division and FtsZ

Cell division across all domains of life is initiated by the formation of a cytokinetic ring [1-9]. A group of proteins, termed the divisome, collectively functions to constrict the ring in the center of the cell, synthesize new material, and pinch the cell in two [10]. In most bacteria and some archaea, the foundation for the divisome is FtsZ (filamentous temperature sensitive mutant Z), which is a homolog of eukaryotic tubulin [11-18]. At the nascent division site between the two segregated nucleoids / chromosomes, FtsZ assembles into the cytokinetic ring, termed the Z-ring [11-17]. This positioning is highly precise and results in the occupation of only ~2% of the total cell volume [19, 20]. The precise control over spatial localization involves both positive and negative regulators of cell division and FtsZ assembly [21-23]. Recent evidence has suggested that transient membrane-anchored assemblies of FtsZ precede the formation of the cytokinetic ring and that the aforementioned regulators play an imperative role in properly ensuring that these temporary structures become more stable at the requisite location [24]. The names and specific functions of the regulatory proteins involved are bacterium-specific, and many of the features, roles, and components of divisomes have yet to be fully elucidated [10]. However, general rules, such as the involvement of negative and positive FtsZ regulators, have been uncovered that have broad relevance across different bacterial species.

Negative regulation prevents aberrant and delocalized assembly of FtsZ. The MinCDE and the nucleoid occlusion (NO) systems have been identified as negative regulators that exert control over precision in spatial localization. In the Min system of *Escherichia coli* (*E. coli*), MinC



gradients allow for the Min proteins to form a spatiotemporal oscillator that inhibits FtsZ assembly, as MinC is a direct antagonist of FtsZ [25]. Therefore, at the cell poles where MinC concentration is high, FtsZ is prevented from assembling [26]. *Bacillus subtilis* (*B. subtilis*) also uses a similar mechanism but uses DivIVA as a replacement for MinE [27]. DivIVA is unrelated to MinE and instead has homology to Wag31, a protein in bacteria such as *Mycobacterium tuberculosis* involved in defining the cell shape [28].

Negative regulation is also enabled by the NO system, which prevents Z-ring assembly over or in the nucleoid [29]. In the *E. coli* NO system, the relevant players are the protein SlmA, DNA, and FtsZ [30, 31]. SlmA binds to both the DNA and FtsZ [32]; however, SlmA can only bind FtsZ when SlmA itself is bound to DNA at a site away from the center of the cell. Therefore, this mechanism serves to sequester FtsZ and prevent aberrant assembly. In *B. subtilis*, the NO system is similar but instead uses the protein Noc [29].

Interestingly, cell division in *B. subtilis* and *E. coli* is not dependent upon either the MinCDE or NO systems for division or septum localization. Deletions of both of these systems in *B. subtilis* and *E. coli* result in division and growth phenotypes that are mostly indistinguishable from wild type [33, 34]. Recent evidence suggests that the MinCDE and NO systems enable efficient resource utilization rather than enabling positional control [35]. Other systems mechanisms might instead be essential for regulating cell division in *B. subtilis* and *E. coli* [10]. One hypothesized mechanism is that the chromosome is used to position the divisome, as with the MatP protein in *E. coli* and *B. subtilis*. MatP binds to DNA at the DNA replication terminus binding site, otherwise known as the Ter macrodomain of the nucleoid, which is localized to the center of the cell during division and connects to the Z-ring [36]. In the case where the DNA is damaged,

the Z-ring location is impaired, and *E. coli* cells divide asymmetrically [37]. Thus, the Ter-linkage ensures that the Z-ring is correctly located with respect to the genetic material and could be how, in the absence of other negative regulatory systems, the Z-ring still finds its way to the middle of the dividing cell.

Once the location is determined, the cell cycle progresses through two distinct steps: an early stage and a late stage. The early stage involves the assembly of the cytokinetic ring and includes proteins such as FtsZ and its membrane anchoring proteins; the late stage is the cell invagination and separation, including proteins required for functions such as cell wall synthesis [10]. In this first step, membrane anchoring proteins such as FtsA, an actin homolog, connects the Z-ring to the membrane via an amphipathic helix, and this protein is conserved across *B. subtilis*, *Caulobacter crescentus* (*C. crescentus*), and *E. coli* [38, 39]. Interestingly, there has been debate about the role of FtsA in the transition from early to late stages of cytokinesis. This debate centers around the ability of the amphipathic helix of FtsA to invaginate the cell wall, which is thought to trigger the recruitment of other late-stage proteins. Indeed, bottom-up reconstitutions of the Z-ring in systems of FtsZ, FtsA, energetic material, and liposomal membranes have shown that this minimal system can drive liposomal constriction [40]. However, this study does not consider membrane potential, which plays an essential role in FtsA function [41], therefore bringing to question whether these results are relevant *in vivo*. Indeed, recent work on the divisome with cryo-electron microscopy (cryo-EM) did not find evidence for membrane invagination driven by the Z-ring structure [42]. These results indicate that even though it is an essential step, membrane anchoring is likely not the direct transition from one stage of the cell cycle to the next. It is important to note that FtsA itself is not essential as other proteins are also involved with membrane

anchoring, such as ZipA in *E. coli* [43-45], EzrA and SepF in *B. subtilis* [46-48], and FzIC in *C. crescentus* [47].

Positive regulators of FtsZ driven cell division also include a family of proteins involved in the early stage that contribute to the stability and the integrity of the Z-ring structure called the Zap proteins [46]. The Zap proteins are implicated in filling gaps between FtsZ clusters, enabling FtsZ higher-order assembly, holding the structure together, and changing the Z-ring dynamics [36, 49, 50]. However, the apparent redundancy of this family of proteins with ZapA, B, C, and D all having similar functions in *E. coli* has yet to be fully explained, though recent crystallographic studies suggest that these proteins bind at different sites on FtsZ [36, 46, 51, 52]. The most commonly studied member of this family is ZapA, as it is widely conserved across many bacterial systems [46]. While ZapA is nonessential in both *E. coli* and *B. subtilis*, its overexpression has been shown to overcome lethality induced by over-expressions or knockouts, as was shown in *B. subtilis* with MinD and DivIVA, respectively [46]. ZapA is also known to compete with MinC for binding to FtsZ, indicating that these two proteins regulate the assembly state of FtsZ, which depends on each protein's local concentration [53].

With a stable and anchored Z-ring, the late-stage components of the divisome undergo a division cascade that, in *E. coli*, involves the following steps: (1) connection of the Z-ring to the group of proteins responsible for completing the cell cycle. These proteins include the lipoproteins, the amidases that break up the peptidoglycan, which is the polymer that constitutes the bacterial cell wall, and the enzymes that synthesize the new peptidoglycan in the septal cell wall [54]. (2) Transmembrane proteins FtsQ (DivIB), FtsL, and FtsB (DivIC) connect the cytoplasmic divisome to the periplasmic components. This complex first assembles and then is stabilized through

interactions with FtsW, FtsI, and last arriving FtsN [38]. (3) Once FtsN arrives, constriction begins, and peptidoglycan synthesis proceeds [55]. While the arrival of FtsN is an essential first step, the actual trigger for the division cascade is thought to be ABC-transporter like complex FtsEX, which enables an interaction between FtsA and FtsN and results in a conformational change in FtsA [56, 57]. This change subsequently facilitates the interactions that activate peptidoglycan synthesis [56]. Importantly, the ATPase activity of FtsEX is essential to start constriction at mid-cell, further implicating this functionality as the start of the division cascade [58].

Under normal conditions, once this cascade occurs, the cell progresses through division resulting in two equivalent daughter cells [59]. While there are many proteins and factors involved, this entire process relies on the tight regulation of FtsZ assembly. It follows then that understanding the mechanism by which FtsZ assembles and disassembles, as an autonomous entity, in reconstitutions with components of the divisome and in the context of live cells, will provide both a foundational and integrative understanding of the bacterial cell division process. Here, we pursue a bottom-up approach focused on dissecting the molecular functions associated with different modules within FtsZ and assessing how they vary across orthologs.

### **1.2.2 Dynamics and assembly of FtsZ in the Z-ring**

Like tubulin, FtsZ is a GTPase that can bind and hydrolyze GTP [60]. While monomeric FtsZ can bind GTP, the catalytic site is formed at the interface between two FtsZ subunits – see Appendix for example (**Figure 1A.1**) [61]. This is because the N-terminal side of FtsZ contains a Rossman fold, common to nucleotide-hydrolyzing enzymes, and the amino acids for GTP binding. In contrast, the C-terminal end contains the highly conserved T7 loop that contains the amino acid

sequence required for GTP hydrolysis [14]. Mutations to this site diminish the overall GTPase activity of FtsZ [62]. Therefore, the minimal unit of a GTPase active FtsZ is a dimer.

In cell division, FtsZ polymers serve as a treadmilling platform for the division machinery, especially for the enzymes required to synthesize peptidoglycans [63, 64]. The treadmilling observed in FtsZ is reminiscent of the association / dissociation that occurs in actin filaments [65]. This discovery settled a longtime debate in the FtsZ field as to whether FtsZ polymers underwent dynamic instability like tubulin [66, 67], treadmilled like actin [62], or severed stochastically and annealed in the middle of filaments [68]. While a case for each mechanism was independently made, all previous assessments used indirect measures to interrogate the assembly dynamics, such as a tryptophan reporter assay [68], site-directed mutational analyses [67], and non-hydrolyzable analogs of FtsZ [62].

In two papers published in *Science* in 2017, direct assessments of FtsZ dynamics were made by visualizing filaments with total internal reflection fluorescence (TIRF) microscopy [63, 69]. Both groups observed that as GTP hydrolysis occurs within the polymers, GDP-bound monomers are released from the minus end of the filament, while GTP-bound subunits are added to the positive end. This highly dynamic treadmilling (20 – 50 nm/s) and the resulting subunit turnover occurs on the order of seconds [70, 71]. Mutations to the GTP hydrolysis site in FtsZ reduces Z-ring dynamics, directly coupling subunit turnover and Z-ring dynamics to the GTP hydrolysis rate [70]. While some studies have shown that FtsZ polymers can exchange nucleotides in the middle of the protofilament, GDP dissociation occurs at a rate estimated to be 30 – 50 times slower than disassembly, providing further evidence that the primary method for subunit exchange is at the filament ends [72]. Recent work showed that disruptions to FtsZ filament assembly

observed *in vitro* have deleterious effects on the dynamics of treadmilling [73]. While this finding built upon several previous observations that *in vitro* FtsZ assembly is related to *in vivo* Z-ring assembly, stability, and integrity [74], it provided evidence that FtsZ assembly mechanisms *in vitro* might also connect directly to Z-ring dynamics.

### 1.2.3 Assembly of FtsZ *in vitro*

FtsZ forms single-stranded polymers *in vitro*, and GTP-binding promotes this polymerization [17]. Single-stranded polymers, like those formed by FtsZ, can often form via an isodesmic process where an identical association constant characterizes each monomer addition step. In an isodesmic process, there is no critical concentration, and within a pool of filaments, a diverse range of lengths will be observed [75, 76]. *In vitro* experiments, however, indicate that FtsZ assembly is highly cooperative, marked by distinct transitions [14, 77]. The hallmark of a cooperative assembly reaction such as polymerization is the presence of at least one threshold concentration, referred to often as a critical concentration [78]. In this context, it is worth noting that the formation of FtsZ assemblies appears to involve the crossing of one or more critical concentration thresholds, as shown with fluorescence reporter assays [68, 79]. This is, in part, the evidence for cooperative assembly, a topic we will return to in detail in **Chapter 2** of this thesis. FtsZs from *E. coli* and *B. subtilis* assemble into single-stranded protofilaments when the concentrations of FtsZ monomers cross a threshold concentration of 1  $\mu\text{M}$  (buffer conditions for *E. coli* FtsZ studies: 50 mM MES, 100 mM KAc, 5 mM MgAc, 1 mM EGTA, pH 6.5; for *B. subtilis* FtsZ studies: 50 mM MES, 2.5 mM MgCl<sub>2</sub>, 1 mM EGTA, 50 mM KCl, pH 6.5) [68, 80]. Numerous models have been developed to describe how FtsZ assembles cooperatively [75, 81-83]. Some of these models suggest that once part of the polymer, FtsZ subunits undergo

conformational changes to facilitate the population of a high-affinity state that favors polymerization [81, 84]. Recent crystallographic studies refer to this high-affinity state as an “open” conformation within the protofilament [84]. Given that this open conformation exists in an assembly with more than one subunit, it further supports the inference that the nucleus for FtsZ cooperative assembly is a dimer [68].

While the cooperative assembly of the single-stranded FtsZ protofilament is now widely accepted, further studies have aimed to understand the observation of a second critical concentration within FtsZ assembly [68, 85]. In some bacterial systems, FtsZ protofilament bundling can occur *in vitro*, and this is especially evident in *B. subtilis* FtsZ [70]; therefore, it has been hypothesized that the presence of the second critical concentration represents the concentration threshold that has to be crossed for the lateral association of protofilaments / polymers of FtsZ [86, 87]. Indeed, above, but not below, concentrations of 2-2.5  $\mu\text{M}$  FtsZ, bundled protofilaments are observed [82, 88]. The extent of bundling is concentration-dependent, which, coupled with the presence of a threshold concentration, could be an indication of a condensation aided bond percolation networking transition [89, 90]. This implies that the number of protofilaments involved in the largest network of protofilaments (bundle) will increase as the concentration increases.

Recent mutational studies have proposed binding site locations on the GTPase domain that could be associated with lateral protofilament assembly [91]. These mutational analyses have included R174D [92] and in a separate study, K121L and D304L [91] both in *E. coli*. The R174D mutant can facilitate linear polymerization but does not enable lateral bundling, even in the presence of bundling agents such as  $\text{Ca}^{2+}$ , [92] the effects of which are discussed in detail below.

Interestingly, simultaneously mutating another site, L169R, rescued the ability to bundle protofilaments [93]. Data from photo-induced cross-links showed that the two sites K121 and D304 might be involved in lateral interactions [91]. However, in this model, the protofilaments were assumed to be anti-parallel, an orientation that appears to be at odds with the widely accepted protofilament treadmilling model during constriction [91]. In these collective studies, the site mutations resulted in phenotypic consequences, though it is unclear whether this results from the loss of bundling or some downstream consequence. Furthermore, the interrogations of how mutations impact the second critical concentration have not yet been conducted. In this context, it is worth noting that the models discussed to this point primarily implicate / involve the structures adopted by the conserved and autonomously foldable GTPase domain. The body of this thesis work implicates more than just the folded GTPase domain as a determinant of the threshold concentration for lateral associations of FtsZ filaments.

When studying FtsZ assembly, as with any system that exhibits threshold behavior, it is imperative to consider solution conditions. Starting with the presumed most essential component, GTP, it should be noted that while GTP-binding facilitates FtsZ polymerization, the presence of GTP is not required to generate polymers. Polymers of FtsZ bound to GDP (FtsZ-GDP) are observable in the absence of GTP, although this is only observed at concentrations of FtsZ that are at least an order of magnitude higher when compared to the threshold concentration required in the presence of GTP [94, 95]. Without the divalent cation magnesium, FtsZ-GTP polymers behave like FtsZ-GDP, as magnesium is required for FtsZ to hydrolyze GTP [96]. Therefore, to generate active FtsZ polymers *in vitro*, near-physiological concentrations of 1-3 mM of  $Mg^{2+}$  should be used. A tight balancing act is needed for maintaining single-stranded polymers because concentrations of  $Mg^{2+}$  that are higher than 10 mM will induce lateral associations of



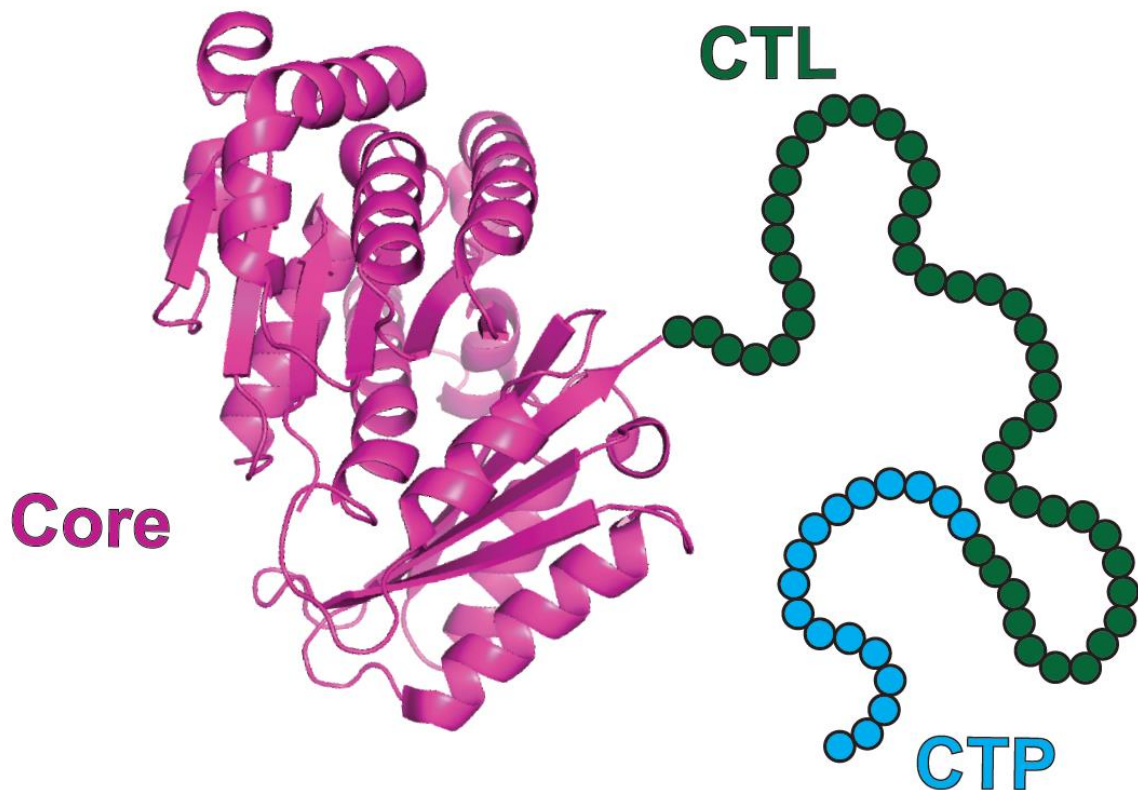
protofilaments, giving rise to thick bundles (5 – 15 nm) [71, 97, 98]. This assembly size increase results in an approximately 3-fold increase in the observed light scattering intensity [99, 100].  $\text{Ca}^{2+}$  is another divalent cation that induces bundles. These bundles are thicker and longer than those formed in the presence of  $\text{Mg}^{2+}$ , as observed qualitatively and shown quantitatively via a 10-fold increase in light scattering intensity in [99]. These bundles formed in the presence of  $\text{Ca}^{2+}$  are reversible and are assumed to consist only of active FtsZ. As a result, calcium-cycling has been used to purify FtsZ; in this process, high concentrations (~20 mM) of  $\text{Ca}^{2+}$  in the presence of GTP can be used to induce bundling of FtsZ protofilaments; the bundles are then sedimented, the pellet is resuspended in a buffer with a calcium-chelator, and the process is repeated [101].

Monovalent salts and pH are also important factors in FtsZ polymerization. FtsZ polymers that are laterally bundled have been primarily observed in buffer conditions with pH ranges from 6.0-7.0 and low salt (50 mM KCl) [102, 103]. Potassium is essential for forming single-stranded protofilaments [104, 105]. In contrast, the addition of sodium causes an upward shift to the critical concentration, as it directly destabilizes the FtsZ dimer [104, 105]. In conditions intended to mimic solution conditions *in vivo*, namely, pH of 7.4-7.7 and a salt concentration of 350 mM KCl, primarily single-stranded protofilaments are observed, with little to no bundles [14]. Similar observations have been made even at a pH of 6.5 and KCl concentrations of 100 mM [102]. These results could be interpreted to mean that the bundles observed *in vitro* have little to no relevance to those seen *in vivo*. However, mutations that impact FtsZ bundling *in vitro* are known to have deleterious impacts on the phenotype [15, 91, 102]. Further, bundling *in vitro* is recovered in the presence of crowding agents, such as PEG-Dextran and Ficoll, which are intended to mimic the crowded cellular environment [106]. Other relevant ions and solutes may also modulate the phase boundaries of FtsZ assembly; these likely include phosphate and glutamate, given that the

intracellular concentrations of these ions are estimated to be in the millimolar range [107-110]. Indeed, one study has reported glutamate-induced bundling of FtsZ, though this study was conducted at a high 1 M glutamate concentration [107]. These collective results show that modulating the concentration of one or more solute may reveal important information about each component's impact on FtsZ assembly but does not necessarily relate to the *in vivo* behaviors or mimic the cellular environment. Further studies that consider the entirety of the cellular environment are required to make direct inferences.

#### 1.2.4 Architecture of FtsZ

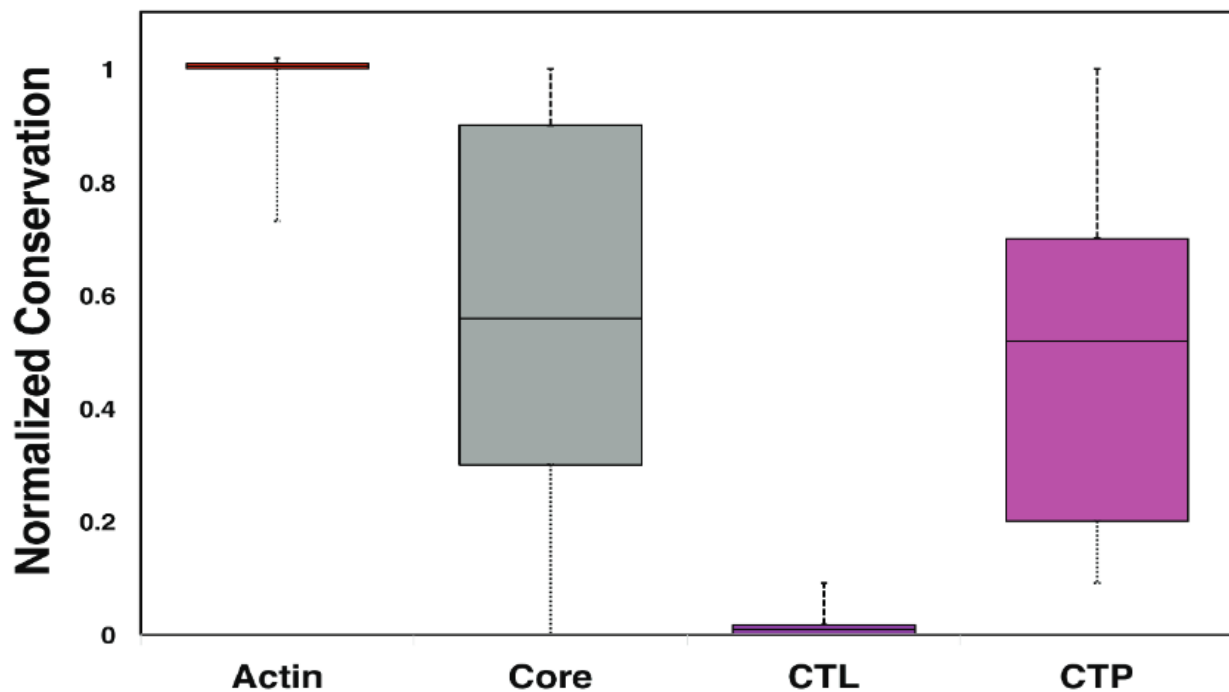
Based on recent studies, [102, 111, 112] clues to the observed complexities and regulation of FtsZ assembly come from the distinctive protein architecture, which includes modules other than the GTPase domain (**Figure 1.1**). The modular architecture of FtsZ encompasses three essential domains: a folded N-terminal core domain that forms an active GTP hydrolysis site upon dimerization, an intrinsically disordered linker (CTL), and a C-terminal peptide (CTP) [102, 111, 112]. Collectively, the CTL and the CTP are known as the C-terminal tail or CTT. While often considered a part of the core, the N-terminus of FtsZ contains a short region that is predicted to be disordered, and it is not part of the Rossmann fold of the GTPase domain [113]. Interestingly, this short tail is not present in tubulin [113]. Ascribing functions to this 5-30 residue tail has been difficult as mutational studies in *E. coli* have also removed part of the GTPase domain. To date, the only identified function of this region is in chloroplasts, where it has been implicated in organelle targeting [114, 115]. Given that the localization sequence is not present in or relevant for prokaryotic orthologs of FtsZ, there may be another unknown function for this region.



**Figure 1.1: Architecture of FtsZ includes three distinct domains.** The *Bacillus subtilis* FtsZ (Bs-FtsZ) core (residues 1-315) is represented using a pink ribbon structure rendering of the model deposited in the protein data bank (PDB id: 2vxy) and displayed here using PyMOL (The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC). The 68-residue CTT of Bs-FtsZ spans residues 315 – 383 and is depicted using 50 green beads for the CTL and 18 cyan beads for the CTP.

At the extreme C-terminus of the CTT is the CTP, which can be further broken down into two pieces, a conserved molecular recognition element [116] that coordinates heterotypic protein-protein interactions involving FtsZ and a C-terminal variable region (CTV) that, as its name suggests, is not well conserved [102]. The CTV can be 0-30 amino acids long and consists mostly of polar and basic residues [113]. In the CTP, most FtsZs contain a conserved motif involved with facilitating intermolecular interactions [39, 43, 52]. This motif has the consensus sequence of **DxxDxPxFL** [113]. Cryptic versions of this motif can be identified by searching for a Pro residue

in the sixth position and a Phe residue in the eighth position, as these are the most highly conserved features of the motif (> 95%) [113]. The normalized conservations of each of FtsZ domains across a system of 1208 orthologs identified in [112] are shown in **Figure 1.2**. Here, we use the globular actin (G-actin) sequence as a standard for calibrating a high degree of conservation [117]. The hypervariable CTV lowers the conservation score of the CTP, which in our analysis encompasses the CTV.



**Figure 1.2: Normalized conservation scores for different regions of FtsZ.** The conservation scores, in terms of percent sequence identities, were computed using CLUSTAL-W. In each box and whisker plot, the solid horizontal line represents the mean conservation score. The box outlines the middle 50% of the values, and the whiskers represent the entire distribution. All conservation scores were normalized to the scores for sequences of G-actin. By definition, the mean normalized score is 1 for G-actin. The mean normalized conservation scores are ~0.57 and ~0.55 for the core domain and CTP motifs, respectively, in contrast to the low score of the CTL.

The interaction partners that engage with FtsZ through the CTP include many of the interactome proteins described above, including FtsA, SepF, MinC, and ZipA [118]. The CTP was

also shown to be involved with FtsZ *in vitro* assembly in the absence of modulatory proteins [102]. *E. coli* FtsZ (Ec-FtsZ), which has a noticeably diminished ability to associate laterally *in vitro*, has a different CTP from *B. subtilis* FtsZ (Bs-FtsZ), and these differences can mostly be attributed to the CTV region [102, 113]. This points to the potential involvement of the CTP in lateral associations of FtsZ protofilaments *in vitro*. Indeed, in the study conducted by Buske et al. [102], swapping the 6-residue CTV of the Bs-FtsZ with the neutral 4-residue CTV of Ec-FtsZ led to a reduction of Bs-FtsZ bundling to levels like Ec-FtsZ. Deleting the domain also significantly diminished lateral assembly, indicating that the CTV is essential for lateral interactions [102]. Since the *E. coli* and *B. subtilis* cores are both negatively charged, it follows then that the basic residues in the CTV could be acting as a polycation helping to prevent electrostatic repulsions between filaments. This model was supported by the further weakening of bundling when the basic residues within the CTV were swapped with acidic residues [102]. Cryo-EM experiments also showed that only single-stranded protofilaments are observed with CTP deletions even at concentrations of Bs-FtsZ that are 10-fold higher [95]. Adding a stabilizing agent (FtsZ-binding anti-bacterial agent PC190723) [95] or  $Mg^{2+}$  [98] could induce bundling of the CTP deletion construct that is indistinguishable from wild type. However, all mutations that disrupted bundling also resulted in phenotypic impacts when expressed *in vivo* [102]. Taken together, these results suggest that the presence and charge of the CTP are imperative not only for protein networking but also for enabling the lateral bundling of protofilaments.

In contrast to the CTP and GTPase core, the CTL varies substantially across the system of orthologs, as shown in **Figure 1.2**. These sequences not only vary by an order of magnitude in length (~30-300 residues) but also vary substantially in composition and sequence patterns [119]. Despite this variability, multiple studies over the past few years have shown that the deletion of

the disordered CTL impairs cell division in different bacterial systems, primarily *E. coli*, *C. crescentus*, and *B. subtilis* [98, 111, 112, 120, 121]. The studies of each system revealed two significant findings: (1) the CTL is essential for FtsZ-mediated cell division, and (2) there appear to be different, bacterium-specific requirements that define a functional CTL. The following section describes each of these studies in detail and lays out the unanswered questions regarding the functionality of the CTL, some of which we focused on answering in this thesis.

The first indication of the presence of a disordered domain C-terminal to the core came from its lack of resolution in early crystallography studies of FtsZ [60]. Approximately fifteen years later, as the interest in intrinsically disordered proteins / regions (IDPs / IDRs) grew, the first studies detailing the essentiality of the CTL were published. In the study conducted by Buske and Levin in *B. subtilis*, deletion of the CTL caused a mis-localization and filamenting phenotype whereby FtsZ molecules made aberrant assemblies away from the mid-section in cells that also became aberrantly long [111, 120]. Replacing the CTL with a rigid alpha-helical domain from human beta-catenin yielded diffuse FtsZ puncta that compromised Z-ring formation and cell division [111]. Both variations impacted the ability of FtsZ to form protofilaments, and neither supported protofilament bundling. The same phenotypic impacts of a CTL deletion were shown in *E. coli* [120]. Two years later, work in *C. crescentus* also revealed that a *C. crescentus* FtsZ (Cc-FtsZ) CTL deletion caused the cells to lyse and bulge, and this had further downstream effects on cell wall synthesis, [121] as it impairs the ability to coordinate interactions with FtsA [122]. All deletion variants also resulted in significant reductions in GTPase activity [111, 120, 121]. Together, these results suggested that a disordered CTL was essential for *in vitro* assembly and GTPase activity as well as *in vivo* functions.

What then are the sequence requirements of a functional CTL? The CTLs of Ec-FtsZ and Bs-FtsZ are similar in length (~50 residues), and swapping one for the other did not impact cell growth [111, 120]. However, the Ec-FtsZ CTL did not rescue Bs-FtsZ *in vitro* function; specifically, bundling was significantly reduced [111]. Further interrogations of CTL sequence and length variants showed that amino acid sequences that were double the length did not complement wild type *in vivo* behavior in both Bs-FtsZ and Ec-FtsZ [111, 120]. Cryo-electron microscopy studies showed that this was likely because the CTL plays the role of a spacer in Bs-FtsZ, keeping the protofilaments apart that are otherwise closer together in CTL deletions. Accordingly, longer CTL sequences increased the distance between bundled protofilaments, presumably impacting the structural integrity [95]. These observations implied that there might be a length requirement for the CTL of FtsZ.

However, not all CTL sequences of equal approximate length were complimentary to wild type function, and these discrepancies were most apparent *in vitro* [95, 111, 120]. This behavior was also shown in Cc-FtsZ, where CTLs of various lengths did and did not complement function; for instance, designed CTL sequences that were approximately 20% of the original length were able to complement WT, and some designs that were closer to the wild type length did not [121]. Swapping the CTL for one from a different ortholog also did not necessarily complement the wild type function in a non-length dependent way, indicating that there are system-specific requirements for the CTL of FtsZ [98, 111, 120, 121, 123]. Importantly, titrations of the CTL length cannot be achieved in a generic, sequence agnostic manner. Instead, changes to the length are achieved while simultaneously changing other variables, such as the fractional composition of different amino acids and the context patterning of these residues. These concomitant changes make it challenging to attribute functional discrepancies to a single variable such as CTL length.

In contrast to the Bs-FtsZ, deletion of the CTL in Cc-FtsZ was thought to enhance “bundling” *in vitro* [98]. Notably, however, the morphologies of the  $\Delta$ CTL assemblies were different than those formed by wild type, and the presence of bundles can most likely be ascribed to the increase in magnesium concentration and the use of nonhydrolyzable analogs, both of which will enable bundling [96]. Indeed, in conditions that mimicked the experiments conducted with Bs-FtsZ, bundles were not observed with the  $\Delta$ CTL Cc-FtsZ [98]. When reconstituted on supported lipid bilayers,  $\Delta$ CTL Cc-FtsZ had a significantly reduced turnover rate and formed structures that were large and distinct from the dynamic, short clusters formed by Cc-FtsZ [123]. Taken together with previous evidence connecting turnover with hydrolysis, this reduced subunit turnover was thought to have a direct impact on GTPase functionality, which resulted in the decreased activity observed for  $\Delta$ CTL Cc-FtsZ with respect to wild type [123].

The collection of studies focused on uncovering the roles of the disordered CTT have underscored the importance of this disordered region in FtsZ activity, assembly, Z-ring formation, dynamics, and even cell size. They have also raised several unanswered questions regarding the contributions made by each module within the CTT to the assortment of functions coordinated by FtsZ. **Chapter 2** of this thesis dissects the contributions of each component of the CTT and relates them to the overall protein behavior. However, to further understand the evolved sequence-encoded functional information of this non-conserved disordered CTT, we first explore its role in the context of the known principles governing disordered domains.



## 1.3 Intrinsically Disordered Proteins/Regions

### 1.3.1 Definition of Intrinsically Disordered Regions

Anfinsen noted that “... the three-dimensional structure of a native protein in its normal milieu (solvent, pH, ionic strength, temperature, presence of other components such as metal ions or prosthetic groups, and others) is the one in which the Gibbs free energy of the whole system is lowest; that is the native conformation is determined by the totality of inter-atomic interactions and hence by the amino acid sequence” [124]. The defining feature of Anfinsen’s thermodynamic hypothesis is that the physiologically relevant conformational ensemble is one that minimizes the Gibbs free energy of the protein plus milieu system. Accordingly, for a given sequence, the conformational ensemble, defined as the collection of conformations accessible via spontaneous fluctuations, will be determined by the totality of sequence-encoded interactions and their modulation by the surrounding solvent. The resultant ensembles will be characterized by different degrees of conformational heterogeneity, and this depends on the interplay between sequence-encoded and solvent-mediated interactions.

Intrinsically disordered proteins or regions (IDPs / IDRs) and stable autonomously foldable proteins / domains (AFPs / AFDs) are two extreme instantiations of conformational heterogeneity. The ensembles of AFPs can be described in terms of either one or a small number of structures and fluctuations about these structures. In contrast, a high degree of conformational heterogeneity is the defining hallmark of IDPs / IDRs [125]. Accordingly, no single structure can be used as being representative of conformations within the ensemble. Instead, the totality of the ensemble must be utilized to model IDPs / IDRs [126]. **Chapter 3** of this thesis provides a new methodology

to enable holistic comparisons of IDP / IDR sequence-encoded conformational ensembles, focusing on the impact of conformational heterogeneity and its variation with the sequence.

More often than not, IDRs are found tethered to AFDs, either as a linker between more than one AFD or as a tail attached to an AFD [126]. The presence of one or more AFD will influence the conformational ensembles of the IDRs [127]. IDRs are known to serve many functions that influence the AFD(s) to which they are tethered, including acting as scaffolds for displaying short linear motifs, binding sites or interfaces, and modulators of emergent phase behavior [128-131]. Variations to IDR sequences, including complete deletions, have shown that the IDRs tethered to AFDs are imperative to the molecular functions and the resulting cellular phenotypes [132-137]. These observations, as well as analysis of compositional biases, suggest that IDRs are not randomly inserted but have sequence-dependent, context-specific functions [138]. Given the interplay between AFDs and IDRs, it follows that there are likely co-evolutionary relationships between AFDs and IDRs. However, identifying these relationships has unique challenges given that IDRs and AFDs appear to follow different constraints in terms of sequence evolution and sequence-structure-function versus sequence-ensemble-function relationships [138-141].

### **1.3.2 Hypervariability in intrinsically disordered regions**

AFPs / AFDs can be classified into distinct sequence families based on the combination of sequence-structure, structure-function, and sequence-function relationships [142]. Indeed, it is often true that AFDs with similar sequences have similar structural preferences. This paves the way for identifying sequence families and constructing sequence phylogenies based on multiple

sequence alignments that identify regions of sequence identity, similarity, and dissimilarity due to insertions or deletions.

In contrast to AFDs, IDPs / IDRs are not constrained to adopt specific folds, although many IDPs / IDRs fold when bound to their binding partners [143]. Instead, many (not all) IDPs / IDRs are often characterized by hypervariability within sequence families [119, 138, 144-146]. This refers to IDPs / IDRs that are involved in similar functions across orthologs but are also characterized by considerable differences in sequence features such as amino acid compositions, numbers of amino acids, numbers of interaction motifs, and the sequence contexts of these motifs [138, 140, 141]. As a result, multiple sequence alignments must incorporate numerous gaps to maximize sequence similarity and / or identity within an alignment. This lowers the statistical significance of an alignment because the likelihood of realizing an alignment produced by hypervariable sequences can become equivalent to the likelihoods produced by alignments of random sequences. Therefore, multiple sequence alignments are not useful for uncovering the selection principles that underlie the evolution of hypervariable intrinsically disordered sequences. Instead, new tools are required to uncover the extent to which sequence compositions [138], conformational ensembles [125, 147], and sequence patterns [141, 148] are conserved.

### **1.3.3 Decoding the sequence-to-function relationship of IDRs**

IDRs feature prominently in eukaryotic proteins, estimated to be involved in approximately 30-40% of proteins [126, 149]. These regions are associated with a range of molecular functions that contribute to an assortment of cellular processes [150-153]. As discussed above, IDR/IDPs defy the well-established sequence-structure-function paradigm that provides a coherent

framework for connecting sequence-encoded information to molecular functions for proteins that fold into well-defined three-dimensional structures [130]. This is because conformational heterogeneity is a defining hallmark of IDRs / IDPs [153-156]. Despite this sequence-intrinsic conformational heterogeneity, it has become clear that IDRs and IDPs have definable sequence-to-ensemble relationships [156-158]. The overall sizes, shapes, and amplitudes of conformational fluctuations of IDRs / IDPs are governed by compositional parameters such as the chain length, fraction of charged residues, the overall proline contents, and the net charge per residue [147, 159-161]. Additional contributions to sequence-to-ensemble relationships include the linear patterning of residues such as oppositely charged residues and proline and charged residues [162, 163]. In hypervariable sequences, however, it becomes challenging to dissect the features that influence, or in some cases, are imperative to function, as many compositions and even sequence patterns can encode for a similar ensemble [138]. Therefore, different sequences within the same sequence family can have vastly different compositions and / or sequence parameters. This makes it challenging to infer coherent sequence-ensemble-function relationships for hypervariable IDRs.

Progresses in advancing our understanding of sequence-ensemble relationships of IDRs / IDPs [164, 165] have opened the door to using sequence design as a tool to investigate the functional and phenotypic effects of modulating sequence-to-conformation relationships of IDPs / IDRs [132, 133, 137]. These studies have been directed at yeast proteins that control cell cycle arrest [132], mating-type switching [163], and transcriptional activation [137]. Studies have also been directed at mammalian systems that regulate cell differentiation [133], tumor suppression [166], and transcription [167]. There is also growing interest in the roles of IDRs / IDPs as modulators of the driving forces for the formation, regulation, and dissolution of biomolecular condensates [129, 168-175]. As these condensates gain recognition for their roles in different

facets of the biology of life [90, 144, 176-181], more has been learned about the sequence architectures that govern the behavior of IDRs as drivers and regulators of condensate formation and dissolution.

### **1.3.4 IDPs in bacteria**

In contrast to eukaryotic systems and viruses, where the functional importance of IDRs / IDPs is well established [153, 164, 182, 183], the prevailing dogma is that bacterial proteins largely conform to the classical sequence-structure-function paradigm [184, 185]. This view has emerged from bioinformatics analyses, which show that only a small percentage (~4%) of bacterial proteins include long IDRs [186-188]. The apparent bias against IDRs / IDPs in bacterial proteomes has been ascribed to a variety of factors. These include shorter protein lengths, the importance of well-defined structure in metabolic enzymes, and the supposedly focused range of functions associated with bacterial proteins [186, 189, 190].

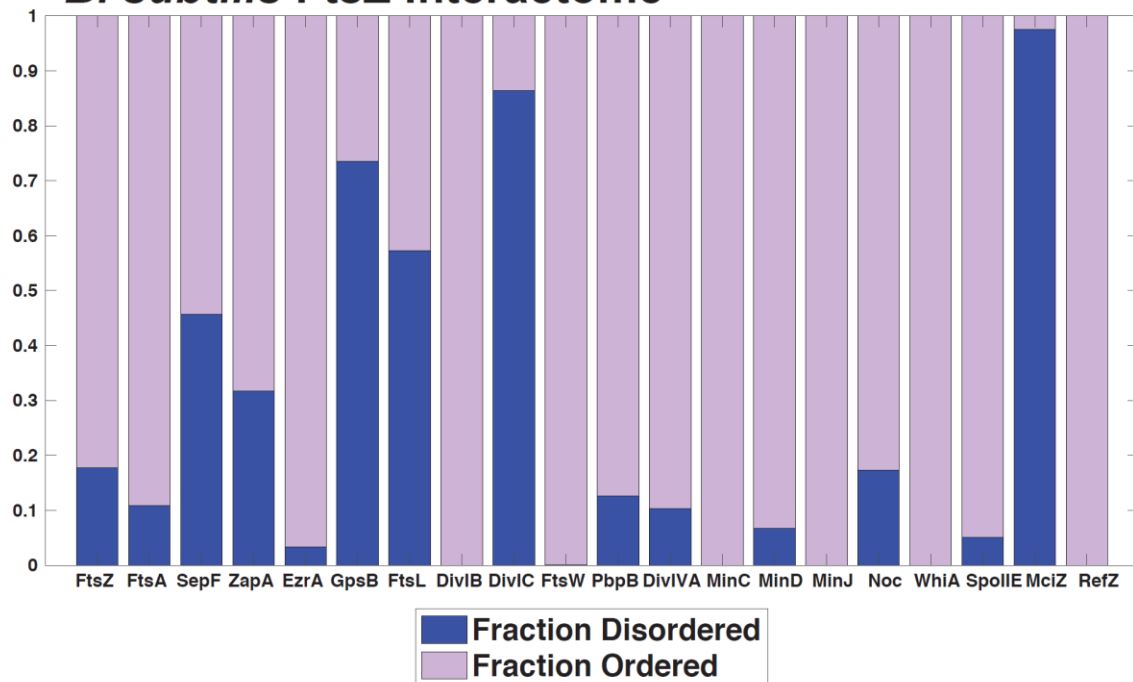
However, recent studies have demonstrated that while IDRs / IDPs may make up only a small fraction of proteins / protein regions in bacterial proteomes compared to eukaryotic ones, they are involved in a number of essential processes. Prominent bacterial IDRs / IDPs are featured in proteins involved in regulating cell division [111, 112], DNA replication [191, 192], protein and RNA quality control [193-199], bacterial warfare [200-202], biofilm formation [203-205], and chemotaxis [206-208]. These findings raise a key question: Do the sequence-ensemble-function relationships uncovered for eukaryotic IDRs / IDPs have any bearing on the functions and phenotypes that are influenced by IDRs / IDPs in bacteria? And if so, can we leverage the diversity of bacterial systems to begin to decode sequence-function relationships within hypervariable IDRs? To start to answer these questions, we focused on the C-terminal disordered tail of FtsZ,

which, as noted in previous sections, is essential in regulating cell division in rod-shaped bacteria [111, 112].

### **1.3.5 Disorder in bacterial cell division**

It is noteworthy that FtsZ serves as an interaction hub for many of the modulatory proteins of the divisome [23, 209, 210]. As discussed above, these proteins assist in anchoring, stabilizing, and providing spatiotemporal control of the cytokinetic ring [211]. Interestingly, despite the paucity of IDRs in bacteria, the interactome of FtsZ is enriched for proteins with IDRs (as shown with *B. subtilis* in **Figure 1.3**). While this dissertation focuses on FtsZ, it is imperative also to uncover the functional contributions of IDRs in the divisome to understand how these modules contribute to networks of interactions within and beyond the divisome, thereby enabling spatiotemporal regulation of bacterial cell division [98, 111, 112, 120, 122].

## *B. subtilis* FtsZ Interactome



**Figure 1.3: Disorder in the *B. subtilis* FtsZ interactome.** Of the 19 proteins listed to interact with FtsZ that are also members of the divisome, 13, or 68%, are predicted to contain IDRs [22]. The fraction of disorder is calculated based on the number of amino acids that are predicted to be disordered divided by the total protein length. The remaining fraction is the fraction that is ordered.

## 1.4 Scope of the Thesis

The goal of this thesis is to understand the molecular sequence architecture and role of the hypervariable C-terminal tail (CTT) of bacterial cell division protein, FtsZ. While the essentiality of the CTT has been established in previous work, questions regarding the molecular functions of the distinct modules and the sequence features that encode these functions have remained unanswered. Using mutational studies with deletion constructs, we quantify the impacts of the different CTT modules on the folded GTPase domain of FtsZ in **Chapter 2**. We find that the C-terminal linker (CTL) and peptide (CTP) play critical roles in FtsZ assembly and activity, predominantly because the CTL modulates the intra- and intermolecular interactions that are enabled through the CTP. Because of these observations, we propose to describe the modular features of the CTT using a *sticker-and-spacer* architecture. Therefore, a functional CTL must act as a spacer, not engaging in intermolecular interactions of its own but rather coordinating the interactions of the CTP sticker.

Leveraging recent discoveries and concepts from the field of intrinsically disordered proteins (IDPs), we sought to decode the evolved sequence features hidden within the set of hypervariable CTT sequences. Like folded domains, IDPs can be defined by their sequence-to-ensemble relationships; however, unlike folded domains, which often have one globally stable conformation, IDPs are characterized by conformational heterogeneity. Therefore, different methodologies are needed for comparative analyses of sequence-ensemble relationships. In **Chapter 3**, we develop and deploy information theory-based methods to compare the conformational ensembles of various IDPs to one another. This method allows us to demonstrate that despite having different sequences and compositions, sequences can encode for similar



conformational ensembles. We use this methodology to introduce the concept of *functional length*, referred to previously as *thermodynamic length*, which is an important feature for defining the functionality of CTTs from different orthologs, meaning that CTTs with similar *functional* lengths will have similar functionalities. Despite significant deviations in their primary sequence, we find that most CTTs encode for similar sequence-to-ensemble relationships. This points to convergent evolution through the conservation or maintenance of sequence-ensemble relationships.

To further determine the criteria of a functional CTT, we next focused on uncovering potentially conserved sequence features. Unlike other IDR sequence families that might preserve composition, hypervariable sequences deviate significantly both in linear sequence features and sequence compositions. To detect non-random features that are therefore ostensibly selected for or against, we compared the observed patterning-based parameters, known to impact both conformational ensemble and molecular function, to the values statistically expected from the most appropriate null models, namely the random ensemble of sequences generated by scrambling sequences while keeping the amino acid composition fixed. Using this methodology, we discuss in **Chapter 4** the observed system-specific requirements of a functional CTL. Importantly, we uncover significant cryptic sequence features that require further investigations to elucidate the potential functional implications. We propose that a functional CTT must then also preserve random / non-random sequence patterns. We also showcase the application of our analysis, referred to as uncovering non-random patterns encoded in hypervariable IDR sequences, to IDRs from two different systems.

In **Chapter 5**, we use *de novo* sequence design to test the following hypotheses: a designed CTL must (1) act as a spacer, (2) maintain the sequence-ensemble relationships, and (3) preserve non-random sequence patterns. For these designs, we used the patterning parameter  $\kappa$  defined by

the degree of oppositely charged residue segregation [212]. This design approach is justified because segregated blocks of oppositely charged residues can cause a gain-of-function through charge-based interactions. Accordingly, we reasoned and then demonstrate that high degrees of charge segregation convert the spacer CTL to a sticker. It is noteworthy that it is impossible to alter a single patterning parameter while keeping all other parameters fixed. Our multi-pronged strategy of generating designed sequence variants that maintain overall amino acid composition allows us to assess how unconstrained sequence parameters change and hence alter molecular functions, self-assembly, and cellular phenotypes when we make focused changes to the Bs-FtsZ CTL by repatterning the mixing versus segregation of oppositely charged residues. We find that the preservation of sequence-ensemble relationships and maintenance and spacer-like behavior are important for maintaining cellular phenotypes. While changes to CTL sequence patterns generate apparently binary phenotypic outcomes, the situation is much more complex *in vitro*. A significant challenge remains with regard to understanding how the changes observed *in vitro* are compensated for *in vivo*. Another significant finding is that the truly non-random sequence features with FtsZ CTLs are highly bacterium specific. Might this point to the selection of sequence features being driven by different adaptation and evolutionary pressures? The work in this thesis sets the stage for being able to answer these questions using a combination of methods developed and deployed here and designing large-scale libraries that leverage the grammars uncovered in this work.

Building on the insights gained from the first five chapters, in **Chapter 6**, we associate the bacterial disordered proteome with spatiotemporal control of bacteria by including a review of recent work in this space and relating concepts learned about IDPs in eukaryotic systems to these systems in bacteria. Finally, in **Chapter 7**, we summarize the main findings and contributions of

this dissertation and conclude with implications and future directions. In doing so, we place our findings within the broader context, offering speculations on how next-generation antibiotics might be developed by targeting IDRs in bacterial proteins.

## 1.5 References

1. Straight, A.F. and C.M. Field, *Microtubules, membranes and cytokinesis*. *Curr Biol*, 2000. **10**(20): p. R760-70.
2. Margolin, W., *Spatial regulation of cytokinesis in bacteria*. *Curr Opin Microbiol*, 2001. **4**(6): p. 647-52.
3. Tolliday, N., N. Bouquin, and R. Li, *Assembly and regulation of the cytokinetic apparatus in budding yeast*. *Curr Opin Microbiol*, 2001. **4**(6): p. 690-5.
4. Huckaba, T.M. and L.A. Pon, *Cytokinesis: rho and formins are the ringleaders*. *Curr Biol*, 2002. **12**(23): p. R813-4.
5. Baluska, F., D. Menzel, and P.W. Barlow, *Cytokinesis in plant and animal cells: endosomes 'shut the door'*. *Dev Biol*, 2006. **294**(1): p. 1-10.
6. Guizetti, J. and D.W. Gerlich, *Cytokinetic abscission in animal cells*. *Semin Cell Dev Biol*, 2010. **21**(9): p. 909-16.
7. Rincon, S.A. and A. Paoletti, *Molecular control of fission yeast cytokinesis*. *Semin Cell Dev Biol*, 2016. **53**: p. 28-38.
8. Thieleke-Matos, C., et al., *Emerging Mechanisms and Roles for Asymmetric Cytokinesis*. *Int Rev Cell Mol Biol*, 2017. **332**: p. 297-345.
9. Addi, C., J. Bai, and A. Echard, *Actin, microtubule, septin and ESCRT filament remodeling during late steps of cytokinesis*. *Curr Opin Cell Biol*, 2018. **50**: p. 27-34.
10. den Blaauwen, T., L.W. Hamoen, and P.A. Levin, *The divisome at 25: the road ahead*. *Curr Opin Microbiol*, 2017. **36**: p. 85-94.
11. Lutkenhaus, J., *FtsZ ring in bacterial cytokinesis*. *Mol Microbiol*, 1993. **9**(3): p. 403-9.
12. Rothfield, L., S. Justice, and J. Garcia-Lara, *Bacterial cell division*. *Annu Rev Genet*, 1999. **33**: p. 423-48.
13. Janakiraman, A. and M.B. Goldberg, *Recent advances on the development of bacterial poles*. *Trends Microbiol*, 2004. **12**(11): p. 518-25.

14. Erickson, H.P., D.E. Anderson, and M. Osawa, *FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one*. *Microbiol Mol Biol Rev*, 2010. **74**(4): p. 504-28.
15. Huang, K.H., J. Durand-Heredia, and A. Janakiraman, *FtsZ ring stability: of bundles, tubules, crosslinks, and curves*. *J Bacteriol*, 2013. **195**(9): p. 1859-68.
16. Meier, E.L. and E.D. Goley, *Form and function of the bacterial cytokinetic ring*. *Curr Opin Cell Biol*, 2014. **26**: p. 19-27.
17. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division*. *Seminars in Cell and Developmental Biology*, 2015. **37**: p. 3-10.
18. Erickson, H.P. and M. Osawa, *Cell division without FtsZ—a variety of redundant mechanisms*. *Molecular microbiology*, 2010. **78**(2): p. 267.
19. Yu, X.C. and W. Margolin, *FtsZ ring clusters in min and partition mutants: role of both the Min system and the nucleoid in regulating FtsZ ring localization*. *Mol Microbiol*, 1999. **32**(2): p. 315-26.
20. Harry, E., J. Rodwell, and R. Wake, *Co-ordinating DNA replication with cell division in bacteria: a link between the early stages of a round of replication and mid-cell Z ring assembly*. *Molecular microbiology*, 1999. **33**(1): p. 33-40.
21. Adams, D.W., L.J. Wu, and J. Errington, *Nucleoid occlusion protein Noc recruits DNA to the bacterial cell membrane*. *The EMBO journal*, 2015. **34**(4): p. 491-501.
22. Errington, J. and L.J. Wu, *Cell cycle machinery in Bacillus subtilis*, in *Prokaryotic Cytoskeletons*. 2017, Springer. p. 67-101.
23. Wagstaff, J. and J. Löwe, *Prokaryotic cytoskeletons: protein filaments organizing small cells*. *Nature Reviews Microbiology*, 2018. **16**(4): p. 187-201.
24. Walker, B.E., J. Mannik, and J. Mannik, *Transient Membrane-Linked FtsZ Assemblies Precede Z-Ring Formation in Escherichia coli*. *Curr Biol*, 2020. **30**(3): p. 499-508 e6.
25. Lutkenhaus, J. and S. Du, *E. coli cell cycle machinery*, in *Prokaryotic Cytoskeletons*. 2017, Springer. p. 27-65.
26. Pichoff, S. and J. Lutkenhaus, *Escherichia coli division inhibitor MinCD blocks septation by preventing Z-ring formation*. *J Bacteriol*, 2001. **183**(22): p. 6630-5.

27. Edwards, D.H. and J. Errington, *The Bacillus subtilis DivIVA protein targets to the division septum and controls the site specificity of cell division*. Mol Microbiol, 1997. **24**(5): p. 905-15.
28. Kang, C.M., et al., *Wag31, a homologue of the cell division protein DivIVA, regulates growth, morphology and polar cell wall synthesis in mycobacteria*. Microbiology (Reading), 2008. **154**(Pt 3): p. 725-735.
29. Adams, D.W., L.J. Wu, and J. Errington, *Nucleoid occlusion protein Noc recruits DNA to the bacterial cell membrane*. EMBO J, 2015. **34**(4): p. 491-501.
30. Abbondanzieri, E.A. and A.S. Meyer, *More than just a phase: the search for membraneless organelles in the bacterial cytoplasm*. Current Genetics, 2019. **65**(3): p. 691-694.
31. Monterroso, B., et al., *Bacterial FtsZ protein forms phase-separated condensates with its nucleoid-associated inhibitor SlmA*. EMBO Rep, 2019. **20**(1).
32. Bernhardt, T.G. and P.A. De Boer, *SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over chromosomes in E. coli*. Molecular cell, 2005. **18**(5): p. 555-564.
33. Bailey, M.W., et al., *Evidence for divisome localization mechanisms independent of the Min system and SlmA in Escherichia coli*. PLoS Genet, 2014. **10**(8): p. e1004504.
34. Rodrigues, C.D. and E.J. Harry, *The Min system and nucleoid occlusion are not required for identifying the division site in Bacillus subtilis but ensure its efficient utilization*. PLoS Genet, 2012. **8**(3): p. e1002561.
35. Yu, Y., et al., *The Min System Disassembles FtsZ Foci and Inhibits Polar Peptidoglycan Remodeling in Bacillus subtilis*. mBio, 2020. **11**(2).
36. Buss, J., et al., *A multi-layered protein network stabilizes the Escherichia coli FtsZ-ring and modulates constriction dynamics*. PLoS Genet, 2015. **11**(4): p. e1005128.
37. Raghunathan, S., et al., *Asymmetric chromosome segregation and cell division in DNA damage-induced bacterial filaments*. Mol Biol Cell, 2020: p. mbcE20080547.
38. Busiek, K.K., et al., *The early divisome protein FtsA interacts directly through its 1c subdomain with the cytoplasmic domain of the late divisome protein FtsN*. J Bacteriol, 2012. **194**(8): p. 1989-2000.

39. Krupka, M., et al., *Role of the FtsA C terminus as a switch for polymerization and membrane association*. mBio, 2014. **5**(6): p. e02221.
40. Osawa, M. and H.P. Erickson, *Liposome division by a simple bacterial division machinery*. Proc Natl Acad Sci U S A, 2013. **110**(27): p. 11000-4.
41. Strahl, H. and L.W. Hamoen, *Membrane potential is important for bacterial cell division*. Proc Natl Acad Sci U S A, 2010. **107**(27): p. 12281-6.
42. Daley, D.O., U. Skoglund, and B. Soderstrom, *FtsZ does not initiate membrane constriction at the onset of division*. Sci Rep, 2016. **6**(1): p. 33138.
43. Hale, C.A., A.C. Rhee, and P.A. de Boer, *ZipA-induced bundling of FtsZ polymers mediated by an interaction between C-terminal domains*. J Bacteriol, 2000. **182**(18): p. 5153-66.
44. Lopez-Montero, I., et al., *Intrinsic disorder of the bacterial cell division protein ZipA: coil-to-brush conformational transition*. FASEB J, 2013. **27**(8): p. 3363-75.
45. Schoenemann, K.M., D.E. Vega, and W. Margolin, *Peptide Linkers within the Essential FtsZ Membrane Tethers ZipA and FtsA Are Nonessential for Cell Division*. J Bacteriol, 2020. **202**(6): p. e00720-19.
46. Gueiros-Filho, F.J. and R. Losick, *A widely conserved bacterial cell division protein that promotes assembly of the tubulin-like protein FtsZ*. Genes & Development, 2002. **16**(19): p. 2544-2556.
47. Haeusser, D.P., et al., *The division inhibitor EzrA contains a seven-residue patch required for maintaining the dynamic nature of the medial FtsZ ring*. J Bacteriol, 2007. **189**(24): p. 9001-10.
48. Duman, R., et al., *Structural and genetic analyses reveal the protein SepF as a new membrane anchor for the Z ring*. Proceedings of the National Academy of Sciences, 2013. **110**(48): p. E4601-E4610.
49. Fu, G., et al., *In vivo structure of the E. coli FtsZ-ring revealed by photoactivated localization microscopy (PALM)*. PLoS One, 2010. **5**(9): p. e12682.
50. Caldas, P., et al., *Cooperative ordering of treadmilling filaments in cytoskeletal networks of FtsZ and its crosslinker ZapA*. Nat Commun, 2019. **10**(1): p. 5744.

51. Roach, E.J., et al., *Structure and Mutational Analyses of Escherichia coli ZapD Reveal Charged Residues Involved in FtsZ Filament Bundling*. J Bacteriol, 2016. **198**(11): p. 1683-1693.
52. Schumacher, M.A., et al., *Structural and Functional Analyses Reveal Insights into the Molecular Properties of the Escherichia coli Z Ring Stabilizing Protein, ZapC*. J Biol Chem, 2016. **291**(5): p. 2485-98.
53. Scheffers, D.J., *The effect of MinC on FtsZ polymerization is pH dependent and can be counteracted by ZapA*. FEBS Lett, 2008. **582**(17): p. 2601-8.
54. Pazos, M., et al., *Z-ring membrane anchors associate with cell wall synthases to initiate bacterial cell division*. Nat Commun, 2018. **9**(1): p. 5090.
55. Weiss, D.S., *Last but not least: new insights into how FtsN triggers constriction during Escherichia coli cell division*. Mol Microbiol, 2015. **95**(6): p. 903-9.
56. Du, S., S. Pichoff, and J. Lutkenhaus, *FtsEX acts on FtsA to regulate divisome assembly and activity*. Proc Natl Acad Sci U S A, 2016. **113**(34): p. E5052-61.
57. Yang, D.C., et al., *An ATP-binding cassette transporter-like complex governs cell-wall hydrolysis at the bacterial cytokinetic ring*. Proceedings of the National Academy of Sciences, 2011. **108**(45): p. E1052-E1060.
58. Arends, S.J., R.J. Kustus, and D.S. Weiss, *ATP-binding site lesions in FtsE impair cell division*. J Bacteriol, 2009. **191**(12): p. 3772-84.
59. Weart, R.B. and P.A. Levin, *Growth rate-dependent regulation of medial FtsZ ring formation*. Journal of Bacteriology, 2003. **185**(9): p. 2826-34.
60. Lowe, J. and L.A. Amos, *Crystal structure of the bacterial cell-division protein FtsZ*. Nature, 1998. **391**(6663): p. 203-6.
61. Oliva, M.A., D. Trambaiolo, and J. Lowe, *Structural insights into the conformational variability of FtsZ*. Journal of Molecular Biology, 2007. **373**(5): p. 1229-42.
62. Redick, S.D., et al., *Mutants of FtsZ targeting the protofilament interface: effects on cell division and GTPase activity*. J Bacteriol, 2005. **187**(8): p. 2727-36.
63. Bisson-Filho, A.W., et al., *Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division*. Science, 2017. **355**(6326): p. 739-743.



64. Monteiro, J.M., et al., *Peptidoglycan synthesis drives an FtsZ-treadmilling-independent step of cytokinesis*. Nature, 2018. **554**(7693): p. 528-532.
65. Wanger, M., et al., *The actin treadmill*. Can J Biochem Cell Biol, 1985. **63**(6): p. 414-21.
66. Scheffers, D. and A.J. Driessen, *The polymerization mechanism of the bacterial cell division protein FtsZ*. FEBS Lett, 2001. **506**(1): p. 6-10.
67. Scheffers, D.J., T. den Blaauwen, and A.J. Driessen, *Non-hydrolysable GTP-gamma-S stabilizes the FtsZ polymer in a GDP-bound state*. Mol Microbiol, 2000. **35**(5): p. 1211-9.
68. Chen, Y., et al., *A rapid fluorescence assay for FtsZ assembly indicates cooperative assembly with a dimer nucleus*. Biophys J, 2005. **88**(1): p. 505-14.
69. Yang, X., et al., *GTPase activity-coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis*. Science, 2017. **355**(6326): p. 744-747.
70. Anderson, D.E., F.J. Gueiros-Filho, and H.P. Erickson, *Assembly dynamics of FtsZ rings in Bacillus subtilis and Escherichia coli and effects of FtsZ-regulating proteins*. J Bacteriol, 2004. **186**(17): p. 5775-81.
71. Barrows, J.M. and E.D. Goley, *FtsZ dynamics in bacterial division: What, how, and why?* Curr Opin Cell Biol, 2020. **68**: p. 163-172.
72. Huecas, S., et al., *The interactions of cell division protein FtsZ with guanine nucleotides*. J Biol Chem, 2007. **282**(52): p. 37515-28.
73. García-Soriano, D.A., et al., *The speed of FtsZ treadmilling is tightly regulated by membrane binding*. Scientific reports, 2020. **10**(1): p. 1-9.
74. Monahan, L.G., A. Robinson, and E.J. Harry, *Lateral FtsZ association and the assembly of the cytokinetic Z ring in bacteria*. Mol Microbiol, 2009. **74**(4): p. 1004-17.
75. Romberg, L., M. Simon, and H.P. Erickson, *Polymerization of FtsZ, a Bacterial Homolog of Tubulin: IS ASSEMBLY COOPERATIVE?* Journal of Biological Chemistry, 2001. **276**(15): p. 11743-11753.
76. Smulders, M.M., et al., *How to distinguish isodesmic from cooperative supramolecular polymerisation*. Chemistry—A European Journal, 2010. **16**(1): p. 362-367.

77. Bi, E.F. and J. Lutkenhaus, *FtsZ ring structure associated with division in Escherichia coli*. Nature, 1991. **354**(6349): p. 161-4.
78. Crick, S.L. and R.V. Pappu, *Thermodynamic and Kinetic Models for Aggregation of Intrinsically Disordered Proteins*, in *Protein and Peptide Folding, Misfolding, and Non-Folding*. 2012. p. 413-440.
79. Caplan, M.R. and H.P. Erickson, *Apparent cooperative assembly of the bacterial cell division protein FtsZ demonstrated by isothermal titration calorimetry*. Journal of Biological Chemistry, 2003. **278**(16): p. 13784-8.
80. Hill, N.S., et al., *A moonlighting enzyme links Escherichia coli cell size with central metabolism*. PLoS Genet, 2013. **9**(7): p. e1003663.
81. Miraldi, E.R., P.J. Thomas, and L. Romberg, *Allosteric models for cooperative polymerization of linear polymers*. Biophys J, 2008. **95**(5): p. 2470-86.
82. Ruiz-Martinez, A., et al., *Efficient Multiscale Models of Polymer Assembly*. Biophys J, 2016. **111**(1): p. 185-96.
83. Ruiz-Martinez, A., et al., *Efficient models of polymerization applied to FtsZ ring assembly in Escherichia coli*. Proc Natl Acad Sci U S A, 2018. **115**(19): p. 4933-4938.
84. Wagstaff, J.M., et al., *A polymerization-associated structural switch in FtsZ that enables treadmilling of model filaments*. MBio, 2017. **8**(3).
85. Chen, Y. and H.P. Erickson, *Rapid in vitro assembly dynamics and subunit turnover of FtsZ demonstrated by fluorescence resonance energy transfer*. J Biol Chem, 2005. **280**(23): p. 22549-54.
86. Gonzalez, J.M., et al., *Cooperative behavior of Escherichia coli cell-division protein FtsZ assembly involves the preferential cyclization of long single-stranded fibrils*. Proceedings of the National Academy of Sciences of the United States of America 2005. **102**(6): p. 1895-900.
87. Romberg, L. and P.A. Levin, *Assembly dynamics of the bacterial cell division protein FTSZ: poised at the edge of stability*. Annual Review of Microbiology, 2003. **57**: p. 125-54.
88. Huecas, S., et al., *Energetics and geometry of FtsZ polymers: nucleated self-assembly of single protofilaments*. Biophys J, 2008. **94**(5): p. 1796-806.

89. Choi, J.M., F. Dar, and R.V. Pappu, *LASSI: A lattice model for simulating phase transitions of multivalent proteins*. PLoS Comput Biol, 2019. **15**(10): p. e1007028.
90. Choi, J.M., A.S. Holehouse, and R.V. Pappu, *Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions*. Annu Rev Biophys, 2020. **49**: p. 107-133.
91. Guan, F., et al., *Lateral interactions between protofilaments of the bacterial tubulin homolog FtsZ are essential for cell division*. Elife, 2018. **7**: p. e35578.
92. Koppelman, C.M., et al., *R174 of Escherichia coli FtsZ is involved in membrane interaction and protofilament bundling, and is essential for cell division*. Mol Microbiol, 2004. **51**(3): p. 645-57.
93. Haeusser, D.P., V.W. Rowlett, and W. Margolin, *A mutation in Escherichia coli ftsZ bypasses the requirement for the essential division gene zipA and confers resistance to FtsZ assembly inhibitors by stabilizing protofilament bundling*. Mol Microbiol, 2015. **97**(5): p. 988-1005.
94. Huecas, S. and J.M. Andreu, *Polymerization of nucleotide-free, GDP- and GTP-bound cell division protein FtsZ: GDP makes the difference*. FEBS Lett, 2004. **569**(1-3): p. 43-8.
95. Huecas, S., et al., *Self-Organization of FtsZ Polymers in Solution Reveals Spacer Role of the Disordered C-Terminal Tail*. Biophys J, 2017. **113**(8): p. 1831-1844.
96. Monterroso, B., G. Rivas, and A.P. Minton, *An equilibrium model for the Mg(2+)-linked self-assembly of FtsZ in the presence of GTP or a GTP analogue*. Biochemistry, 2012. **51**(31): p. 6108-13.
97. Froschauer, E.M., et al., *Fluorescence measurements of free [Mg<sup>2+</sup>] by use of mag-fura 2 in Salmonella enterica*. FEMS Microbiology Letters, 2004. **237**(1): p. 49-55.
98. Sundararajan, K. and E.D. Goley, *The intrinsically disordered C-terminal linker of FtsZ regulates protofilament dynamics and superstructure in vitro*. J Biol Chem, 2017. **292**(50): p. 20509-20527.
99. Mukherjee, A. and J. Lutkenhaus, *Analysis of FtsZ assembly by light scattering and determination of the role of divalent metal cations*. J Bacteriol, 1999. **181**(3): p. 823-32.
100. Sundararajan, K. and E.D. Goley, *Cytoskeletal Proteins in Caulobacter crescentus: Spatial Orchestrators of Cell Cycle Progression, Development, and Cell Shape*, in *Sub-cellular biochemistry*. 2017. p. 103-137.

101. Rivas, G., et al., *Magnesium-induced linear self-association of the FtsZ bacterial cell division protein monomer. The primary steps for FtsZ assembly.* J Biol Chem, 2000. **275**(16): p. 11740-9.
102. Buske, P.J. and P.A. Levin, *Extreme C terminus of bacterial cytoskeletal protein FtsZ plays fundamental role in assembly independent of modulatory proteins.* J Biol Chem, 2012. **287**(14): p. 10945-57.
103. Pacheco-Gomez, R., et al., *The pH dependence of polymerization and bundling by the essential bacterial cytoskeletal protein FtsZ.* PLoS One, 2011. **6**(6): p. e19369.
104. Ahijado-Guzmán, R., et al., *Control by Potassium of the Size Distribution of Escherichia coli FtsZ Polymers Is Independent of GTPase Activity.* Journal of Biological Chemistry, 2013. **288**(38): p. 27358-27365.
105. Mendieta, J., et al., *Structural and functional model for ionic (K(+)/Na(+)) and pH dependence of GTPase activity and polymerization of FtsZ, the prokaryotic ortholog of tubulin.* J Mol Biol, 2009. **390**(1): p. 17-25.
106. Monterroso, B., et al., *Charged Molecules Modulate the Volume Exclusion Effects Exerted by Crowders on FtsZ Polymerization.* PLoS One, 2016. **11**(2): p. e0149060.
107. Beuria, T.K., et al., *Glutamate-induced assembly of bacterial cell division protein FtsZ.* J Biol Chem, 2003. **278**(6): p. 3735-41.
108. Commichau, F.M., et al., *Glutamate metabolism in Bacillus subtilis: gene expression and enzyme activities evolved to avoid futile cycles and to allow rapid responses to perturbations of the system.* J Bacteriol, 2008. **190**(10): p. 3557-64.
109. Whatmore, A.M., J.A. Chudek, and R.H. Reed, *The effects of osmotic upshock on the intracellular solute pools of Bacillus subtilis.* J Gen Microbiol, 1990. **136**(12): p. 2527-35.
110. McCleary, W.R., *Molecular mechanisms of phosphate homeostasis in Escherichia coli.* Escherichia coli—Recent Advances on Physiology, Pathogenesis and Biotechnological Applications. A. Samie, ed. InTechOpen, 2017: p. 333-357.
111. Buske, P.J. and P.A. Levin, *A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo.* Mol Microbiol, 2013. **89**(2): p. 249-63.
112. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division.* Semin Cell Dev Biol, 2015. **37**: p. 3-10.

113. Vaughan, S., et al., *Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota*. Journal of molecular evolution, 2004. **58**(1): p. 19-29.
114. Beech, P.L., et al., *Mitochondrial FtsZ in a chromophyte alga*. Science, 2000. **287**(5456): p. 1276-9.
115. Fujiwara, M. and S. Yoshida, *Chloroplast targeting of chloroplast division FtsZ2 proteins in Arabidopsis*. Biochem Biophys Res Commun, 2001. **287**(2): p. 462-7.
116. Oldfield, C.J., et al., *Coupled folding and binding with alpha-helix-forming molecular recognition elements*. Biochemistry, 2005. **44**(37): p. 12454-70.
117. Holm, L. and C. Sander, *Protein structure comparison by alignment of distance matrices*. J Mol Biol, 1993. **233**(1): p. 123-38.
118. Adams, D.W. and J. Errington, *Bacterial cell division: assembly, maintenance and disassembly of the Z ring*. Nature Reviews Microbiology, 2009. **7**(9): p. 642-53.
119. Cohan, M.C., K.M. Ruff, and R.V. Pappu, *Information theoretic measures for quantifying sequence-ensemble relationships of intrinsically disordered proteins*. Protein Eng Des Sel, 2019. **32**(4): p. 191-202.
120. Gardner, K.A., D.A. Moore, and H.P. Erickson, *The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide*. Mol Microbiol, 2013. **89**(2): p. 264-75.
121. Sundararajan, K., et al., *The bacterial tubulin FtsZ requires its intrinsically disordered linker to direct robust cell wall construction*. Nat Commun, 2015. **6**: p. 7281.
122. Barrows, J.M., et al., *FtsA Regulates Z-Ring Morphology and Cell Wall Metabolism in an FtsZ C-Terminal Linker-Dependent Manner in Caulobacter crescentus*. J Bacteriol, 2020. **202**(7): p. e00693-19.
123. Sundararajan, K., et al., *Species- and C-terminal linker-dependent variations in the dynamic behavior of FtsZ on membranes in vitro*. Mol Microbiol, 2018. **110**(1): p. 47-63.
124. Anfinsen, C.B., *Principles that govern the folding of protein chains*. Science, 1973. **181**(4096): p. 223-230.

125. Lyle, N., R.K. Das, and R.V. Pappu, *A quantitative measure for protein conformational heterogeneity*. J Chem Phys, 2013. **139**(12): p. 121907.
126. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins*. Chem Rev, 2014. **114**(13): p. 6589-631.
127. Mittal, A., et al., *Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins*. J Mol Biol, 2018. **430**(16): p. 2403-2421.
128. Harmon, T.S., A.S. Holehouse, and R.V. Pappu, *Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins*. New Journal of Physics, 2018. **20**(4): p. 045002.
129. Harmon, T.S., et al., *Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins*. eLife, 2017. **6**: p. e30294.
130. Babu, M.M., *The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease*. Biochemical Society Transactions, 2016. **44**(5): p. 1185-1200.
131. Di Fiore, B., et al., *The Mitotic Checkpoint Complex Requires an Evolutionary Conserved Cassette to Bind and Inhibit Active APC/C*. Mol Cell, 2016. **64**(6): p. 1144-1153.
132. Das, R.K., et al., *Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling*. Proc Natl Acad Sci U S A, 2016. **113**(20): p. 5616-21.
133. Sherry, K.P., et al., *Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor*. Proc Natl Acad Sci U S A, 2017. **114**(44): p. E9243-e9252.
134. Banjade, S. and M.K. Rosen, *Phase transitions of multivalent proteins can promote clustering of membrane receptors*. Elife, 2014. **3**.
135. Nott, T.J., et al., *Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles*. Mol Cell, 2015. **57**(5): p. 936-947.
136. Pak, C.W., et al., *Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein*. Mol Cell, 2016. **63**(1): p. 72-85.
137. Staller, M.V., et al., *A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain*. Cell Syst, 2018. **6**(4): p. 444-455 e6.

138. Holehouse, A.S., et al., *CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins*. Biophys J, 2017. **112**(1): p. 16-21.
139. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Curr Opin Struct Biol, 2015. **32**: p. 102-12.
140. Zarin, T., et al., *Selection maintains signaling function of a highly diverged intrinsically disordered region*. Proc Natl Acad Sci U S A, 2017. **114**(8): p. E1450-E1459.
141. Zarin, T., et al., *Proteome-wide signatures of function in highly diverged intrinsically disordered regions*. Elife, 2019. **8**: p. e46883.
142. Orengo, C.A., A.E. Todd, and J.M. Thornton, *From protein structure to function*. Current opinion in structural biology, 1999. **9**(3): p. 374-382.
143. Dyson, H.J. and P.E. Wright, *Intrinsically unstructured proteins and their functions*. Nature reviews Molecular cell biology, 2005. **6**(3): p. 197-208.
144. Franzmann, T.M., et al., *Phase separation of a yeast prion protein promotes cellular fitness*. Science, 2018. **359**(6371): p. eaao5654.
145. Gruszka, D.T., et al., *Disorder drives cooperative folding in a multidomain protein*. Proc Natl Acad Sci U S A, 2016. **113**(42): p. 11841-11846.
146. Riback, J.A., et al., *Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response*. Cell, 2017. **168**(6): p. 1028-1040 e19.
147. Mao, A.H., N. Lyle, and R.V. Pappu, *Describing sequence-ensemble relationships for intrinsically disordered proteins*. Biochemical Journal, 2013. **449**: p. 307-318.
148. Forman-Kay, J.D. and T. Mittag, *From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins*. Structure, 2013. **21**(9): p. 1492-1499.
149. Dunker, A.K., et al., *Intrinsic disorder and protein function*. Biochemistry, 2002. **41**: p. 6573-6582.
150. Dunker, A.K., C.J. Brown, and Z. Obradovic, *Identification and functions of usefully disordered proteins*. . Advances in Protein Chemistry, 2002. **62**: p. 25-49.

151. Iakoucheva, L.M., et al., *Intrinsic disorder in cell-signaling and cancer-associated proteins*. Journal of Molecular Biology, 2002. **323**(3): p. 573-84.
152. Iakoucheva, L.M., et al., *The importance of intrinsic disorder for protein phosphorylation*. Nucleic Acids Research, 2004. **32**(3): p. 1037-1049.
153. Wright, P.E. and H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation*. Nat Rev Mol Cell Biol, 2015. **16**(1): p. 18-29.
154. Babu, M.M., R.W. Kriwacki, and R.V. Pappu, *Versatility from protein disorder*. Science, 2012. **337**: p. 1460-1461.
155. Lyle, N., R.K. Das, and R.V. Pappu, *A quantitative measure for protein conformational heterogeneity*. Journal of Chemical Physics, 2013. **139**(12): p. 121907(1-12).
156. Forman-Kay, J.D. and T. Mittag, *From sequence and forces to structure, function, and evolution of intrinsically disordered proteins*. Structure, 2013. **21**(9): p. 1492-9.
157. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Current Opinion in Structural Biology, 2015. **32**: p. 102-12.
158. Gibbs, E.B. and S.A. Showalter, *Quantification of Compactness and Local Order in the Ensemble of the Intrinsically Disordered Protein FCPI*. Journal of Physical Chemistry B, 2016. **120**(34): p. 8960-9.
159. Holehouse, A.S., et al., *CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins*. Biophysical Journal, 2017. **112**(1): p. 16-21.
160. Hofmann, H., et al., *Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy*. Proc Natl Acad Sci U S A, 2012. **109**(40): p. 16155-60.
161. Muller-Spath, S., et al., *Charge interactions can dominate the dimensions of intrinsically disordered proteins*. Proceedings of the National Academy of Sciences of the United States of America, 2010. **107**(33): p. 14609-14614.
162. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(33): p. 13392-13397.



163. Martin, E.W., et al., *Sequence Determinants of the Conformational Properties of an Intrinsically Disordered Protein Prior to and upon Multisite Phosphorylation*. Journal of the American Chemical Society, 2016. **138**(47): p. 15323-15335.
164. Mollica, L., et al., *Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment*. Front Mol Biosci, 2016. **3**: p. 52.
165. Schneider, R., et al., *Visualizing the molecular recognition trajectory of an intrinsically disordered protein using multinuclear relaxation dispersion NMR*. J Am Chem Soc, 2015. **137**(3): p. 1220-9.
166. Borchers, W., et al., *Disorder and residual helicity alter p53-Mdm2 binding affinity and signaling in cells*. Nature Chemical Biology, 2014. **10**(12): p. 1000-2.
167. Portz, B., et al., *Structural heterogeneity in the intrinsically disordered RNA polymerase II C-terminal domain*. Nature Communications, 2017. **8**: p. 15231.
168. Mitrea, D.M., et al., *Self-interaction of NPM1 modulates multiple mechanisms of liquid-liquid phase separation*. Nature Communications, 2018. **9**(1): p. 842.
169. Lin, Y.H., J.D. Forman-Kay, and H.S. Chan, *Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates*. Biochemistry, 2018.
170. Dao, T.P., et al., *Ubiquitin Modulates Liquid-Liquid Phase Separation of UBQLN2 via Disruption of Multivalent Interactions*. Molecular Cell, 2018. **69**(6): p. 965-978.e6.
171. Wei, M.T., et al., *Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles*. Nature Chemistry, 2017. **9**(11): p. 1118-1125.
172. Lin, Y.H., J.D. Forman-Kay, and H.S. Chan, *Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates*. Biochemistry, 2018. **57**(17): p. 2499-2508.
173. Hernandez-Vega, A., et al., *Local Nucleation of Microtubule Bundles through Tubulin Concentration into a Condensed Tau Phase*. Cell Reports, 2017. **20**(10): p. 2304-2312.
174. Mitrea, D.M., et al., *Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA*. Elife, 2016. **5**.
175. Al-Husini, N., et al., *alpha-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies*. Molecular Cell, 2018. **71**(6): p. 1027-1039 e14.

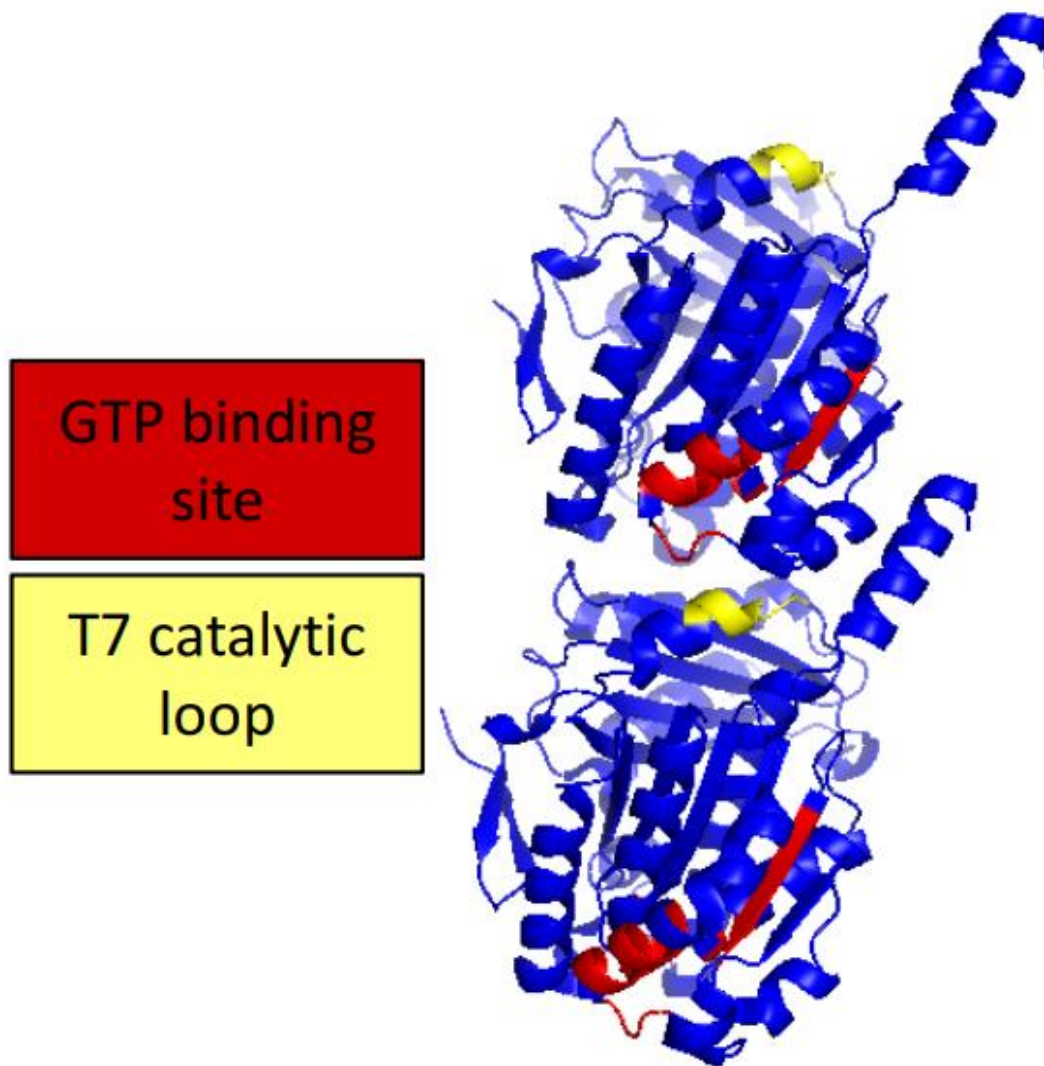
176. Banani, S.F., et al., *Biomolecular condensates: organizers of cellular biochemistry*. Nat Rev Mol Cell Biol, 2017. **18**(5): p. 285-298.
177. Hyman, A.A., C.A. Weber, and F. Julicher, *Liquid-liquid phase separation in biology*. Annu Rev Cell Dev Biol, 2014. **30**: p. 39-58.
178. Wang, J., et al., *A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins*. Cell, 2018. **174**(3): p. 688-699.e16.
179. Woodruff, J.B., et al., *The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin*. Cell, 2017. **169**(6): p. 1066-1077 e10.
180. Emenecker, R.J., A.S. Holehouse, and L.C. Strader, *Emerging Roles for Phase Separation in Plants*. Dev Cell, 2020. **55**(1): p. 69-83.
181. Holehouse, A.S. and R.V. Pappu, *Functional Implications of Intracellular Phase Transitions*. Biochemistry, 2018. **57**(17): p. 2415-2423.
182. Berlow, R.B., H.J. Dyson, and P.E. Wright, *Hypersensitive termination of the hypoxic response by a disordered protein switch*. Nature, 2017. **543**(7645): p. 447-451.
183. Jensen, M.R., et al., *Intrinsic disorder in measles virus nucleocapsids*. Proc Natl Acad Sci U S A, 2011. **108**(24): p. 9839-44.
184. Dunker, A.K., et al., *Intrinsically disordered proteins and multicellular organisms*. Semin Cell Dev Biol, 2015. **37**: p. 44-55.
185. Pavlovic-Lazetic, G.M., et al., *Bioinformatics analysis of disordered proteins in prokaryotes*. BMC Bioinformatics, 2011. **12**: p. 66.
186. Yruela, I., et al., *Evidence for a Strong Correlation Between Transcription Factor Protein Disorder and Organismic Complexity*. Genome Biology and Evolution, 2017. **9**(5): p. 1248-1265.
187. Chen, J.W., et al., *Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder*. J Proteome Res, 2006. **5**(4): p. 888-98.
188. Brown, C.J., et al., *Evolution and disorder*. Current Opinion in Structural Biology, 2011. **21**(3): p. 441-6.

189. Dunker, A.K., et al., *Intrinsically disordered protein*. J Mol Graph Model, 2001. **19**(1): p. 26-59.
190. Brown, C.J., A.K. Johnson, and G.W. Daughdrill, *Comparing models of evolution for ordered and disordered proteins*. Molecular Biology of Evolution, 2010. **27**(3): p. 609-21.
191. Shereda, R.D., et al., *SSB as an organizer/mobilizer of genome maintenance complexes*. Crit Rev Biochem Mol Biol, 2008. **43**(5): p. 289-318.
192. Kozlov, A.G., et al., *Intrinsically disordered C-terminal tails of E. coli single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA*. J Mol Biol, 2015. **427**(4): p. 763-774.
193. Ait-Bara, S. and A.J. Carpousis, *RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs*. Mol Microbiol, 2015. **97**(6): p. 1021-135.
194. Ait-Bara, S., A.J. Carpousis, and Y. Quentin, *RNase E in the gamma-Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains*. Mol Genet Genomics, 2015. **290**(3): p. 847-62.
195. Jakob, U., R. Kriwacki, and V.N. Uversky, *Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function*. Chem Rev, 2014. **114**(13): p. 6779-805.
196. Reichmann, D., et al., *Order out of disorder: working cycle of an intrinsically unfolded chaperone*. Cell, 2012. **148**(5): p. 947-57.
197. Updegrove, T.B., A. Zhang, and G. Storz, *Hfq: the flexible RNA matchmaker*. Current Opinion in Microbiology, 2016. **30**: p. 133-138.
198. Quan, S., et al., *Super Spy variants implicate flexibility in chaperone action*. Elife, 2014. **3**: p. e01584.
199. Foit, L., et al., *Chaperone activation by unfolding*. Proceedings of the National Academy of Sciences of the United States of America 2013. **110**(14): p. E1254-62.
200. Hecht, O., et al., *A common interaction for the entry of colicin N and filamentous phage into Escherichia coli*. Journal of Molecular Biology, 2009. **388**(4): p. 880-93.

201. Papadakos, G., et al., *Consequences of inducing intrinsic disorder in a high-affinity protein-protein interaction*. Journal of the American Chemical Society, 2015. **137**(16): p. 5252-5.
202. Bonsor, D.A., N.A. Meenan, and C. Kleanthous, *Colicins exploit native disorder to gain cell entry: a hitchhiker's guide to translocation*. Biochemical Society Transactions, 2008. **36**(Pt 6): p. 1409-13.
203. Gu, S., et al., *The role of intrinsic disorder and dynamics in the assembly and function of the type II secretion system*. Biochimica et Biophysica Acta, 2017. **1865**(10): p. 1255-1266.
204. Gruszka, D.T., et al., *Disorder drives cooperative folding in a multidomain protein*. Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(42): p. 11841-11846.
205. Whelan, F. and J.R. Potts, *Two repetitive, biofilm-forming proteins from Staphylococci: from disorder to extension*. Biochemical Society Transactions, 2015. **43**(5): p. 861-6.
206. Daughdrill, G.W., et al., *The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28*. Nature Structural Biology, 1997. **4**(4): p. 285-91.
207. Dedmon, M.M., et al., *FlgM gains structure in living cells*. Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(20): p. 12681-4.
208. Ding, L., et al., *Functional characterization of FlgM in the regulation of flagellar synthesis and motility in Yersinia pseudotuberculosis*. Microbiology, 2009. **155**(Pt 6): p. 1890-900.
209. Yang, X., et al., *GTPase activity-coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis*. Science, 2017. **355**(6326): p. 744.
210. Anderson, D.E., F.J. Gueiros-Filho, and H.P. Erickson, *Assembly dynamics of FtsZ rings in Bacillus subtilis and Escherichia coli and effects of FtsZ-regulating proteins*. Journal of Bacteriology, 2004. **186**(17): p. 5775-81.
211. den Blaauwen, T., L.W. Hamoen, and P.A. Levin, *The divisome at 25: the road ahead*. Current Opinion in Microbiology, 2017. **36**: p. 85-94.
212. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*. Proc Natl Acad Sci U S A, 2013. **110**(33): p. 13392-7.



## 1.6 Appendix



**Figure 1A.1:** FtsZ is a GTPase whose active site is formed by the interface between two **FtsZ monomers**. The cartoon representation, generated using the VMD software package, is based on the coordinates for *M. jannaschii* FtsZ dimers obtained from the protein data bank (PDB id: 1W5B). In each subunit, residues that make up the GTP binding site are colored in red, whereas residues that make up the T7 activation loop are colored in yellow.

## Chapter 2

# Dissecting the functional contributions of the intrinsically disordered C-terminal tail of *B. subtilis* FtsZ

### 2.1 Preamble

This chapter is based on the published work, Cohan, M.C., Eddelbuettel, A.M.P., Levin P.A., Pappu R.V. (2020). Dissecting the functional contributions of the intrinsically disordered C-terminal tail of *B. subtilis* FtsZ. *Journal of Molecular Biology*, **432**: 3205-3221. All experiments and data were collected and analyzed by MCC. AMPE helped with troubleshooting key experiments and protein purification. MCC and RVP organized the results, analyzed them, and wrote / revised the manuscript. PAL provided the deletion constructs and essential feedback to the manuscript.

This work focuses on FtsZ, which is a bacterial GTPase that is central to the spatial and temporal control of cell division. It is a filament-forming enzyme that encompasses a well-folded core domain and a disordered C-terminal tail (CTT). The CTT is essential for ensuring proper assembly of the cytokinetic ring, and its deletion leads to mislocalization of FtsZ, aberrant assembly, and cell death. In this chapter, we dissect the contributions of modules within the

disordered CTT to the assembly and enzymatic activity of *Bacillus subtilis* FtsZ (Bs-FtsZ). The CTT features a hypervariable C-terminal linker (CTL) and a conserved C-terminal peptide (CTP). Our *in vitro* studies show that the CTL weakens the driving forces for forming single-stranded active polymers and suppresses lateral associations of these polymers, whereas the CTP promotes the formation of alternative assemblies. Accordingly, in full-length Bs-FtsZ, the CTL acts as a spacer that spatially separates the CTP sticker from the core, thus ensuring filament formation through core-driven polymerization and lateral associations through CTP-mediated interactions. We also find that the CTL weakens GTP binding while enhancing the catalytic rate, whereas the CTP has opposite effects. The joint contributions of the CTL and CTP make Bs-FtsZ an enzyme that is only half as efficient when compared to the truncated version that lacks the CTT. Overall, our data suggest that the CTT acts as an auto-regulator of Bs-FtsZ assembly and an auto-inhibitor of enzymatic activity. Based on our results, we propose hypotheses regarding the hypervariability of CTLs and compare FtsZs to other bacterial proteins with tethered IDRs.



## 2.2 Introduction

Cell division in rod-shaped bacteria is a complex process that begins with the assembly of the tubulin-like cell division protein FtsZ at the site of cell division [1]. *In vivo* studies focused on *Bacillus subtilis* and *Escherichia coli* have shown that FtsZ, which is a GTPase, is responsible for recruiting ~20 additional proteins that together orchestrate the assembly of the division septum [2-9]. FtsZ serves as the foundation on which the rest of the division machinery is assembled, and the machinery itself is frequently referred to as the “Z-ring” [1]. As the first protein at the nascent division site, FtsZ is a target for factors controlling the timing, formation, and location of the Z-ring via direct or indirect interactions with FtsZ [3, 10]. Recent efforts have uncovered modulatory proteins that: (1) prevent Z-ring assembly at aberrant locations (via the functions of nucleoid occlusion proteins such as EzrA and Noc in *B. subtilis*, SlmA in *E. coli*, and the Min proteins in both) [11-14]; (2) tether the Z-ring to the membrane (*E. coli* ZipA and FtsA in both *E. coli* and *B. subtilis*) [15, 16]; and (3) facilitate the assembly of FtsZ into higher-order structures (the Zap proteins in *E. coli* and *B. subtilis*) [10, 17]. Although the key roles of many of the essential proteins are yet to be fully elucidated, it is clear that FtsZ is a central player in the divisome and that FtsZ-mediated assembly of the cytokinetic ring is a tightly regulated process [18].

The formation of Z-rings is governed, in part, by the intrinsic ability of FtsZ to hydrolyze GTP and self-assemble into linear, single-stranded polymers that laterally associate to form higher-order structures [19-23]. Key aspects of the assembly process have been uncovered through systematic structural and mechanistic studies performed *in vitro* and *in vivo* focusing predominantly on FtsZs from *E. coli* and *B. subtilis*, and more recently, *Caulobacter crescentus*

[3, 5, 6, 24-28]. Monomers of FtsZ bind to GTP, and this facilitates the formation of single-stranded linear polymers that are also referred to as protofilaments [29, 30]. The active site of FtsZ is formed at the interface between a pair of FtsZ monomers, which implies that the dimer is the minimal active form of the enzyme [29, 31]. GTP hydrolysis drives the treadmilling of the polymers, thereby controlling the length distribution and dynamics of FtsZ polymers [9, 32-35]. GTP-dependent assembly of FtsZ *in vitro* depends on a variety of factors, including solution conditions and FtsZ concentration [36-39].

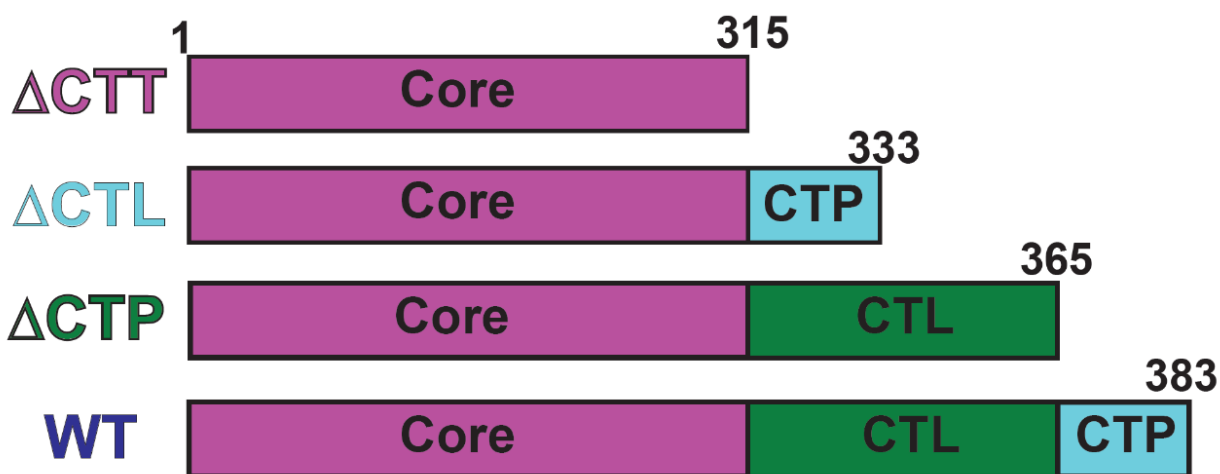
FtsZ is a prominent archetype of *filament-forming enzymes* [40], which are also referred to as *living polymers* in the physical literature [41, 42]. The assembly of filament-forming enzymes often requires the crossing of a threshold concentration to facilitate the formation of active filaments. Previous studies identified the presence of at least two threshold concentrations for GTP-dependent assembly of FtsZ, referred to hereafter as  $c_A$  and  $c_B$ , that are in the micromolar range [20, 21, 43-46]. Increasing the concentration above the threshold concentration  $c_A$  is required for the formation of active protofilaments, which are single-stranded polymers. The crossing of the threshold concentration  $c_B > c_A$  leads to the formation of higher-order assemblies that appear to be characterized by the lateral association of linear polymers [47]. From a thermodynamic standpoint, the existence of threshold concentrations implies that both the formation of single-stranded polymers and the lateral association of these polymers are cooperative processes [41].

Considerable effort has been invested into uncovering the determinants of cooperativity in FtsZ assembly [20, 21, 43-46, 48, 49]. These efforts have been motivated in part by the observation that dimers of FtsZ are thought to be the cooperative units that control the assembly of single-stranded linear polymers [45]. Yet, the nucleus for polymerization appears to be independent of

the binding of GTP and GTP hydrolysis [49]. Additionally, there is the notion that the formation of single-stranded polymers cannot be governed by a threshold concentration [50]. However, this view ignores several nuances whereby conformational transitions and / or anisotropic interactions among subunits, especially in ligand-dependent systems such as FtsZ, can give rise to a threshold concentration for the formation of single-stranded protofilaments [40-42, 51-55].

FtsZ has a tripartite architecture consisting of a globular catalytic domain and an intrinsically disordered C-terminal tail (CTT) that encompasses two modules *viz.*, a 40-300 residue hypervariable linker (CTL) and a highly conserved 10-20 amino acid peptide at the extreme C-terminus called the C-terminal peptide (CTP) [50]. **Figure 1.1** shows a schematic of this architecture using FtsZ from *B. subtilis* (Bs-FtsZ) as an archetype. Recent studies have established the *in vivo* importance of CTTs in *B. subtilis*, *E. coli*, and *C. crescentus* [3, 47, 50, 56-62]. In all FtsZs studied to date, the CTT is not required for forming linear polymers or hydrolyzing GTP [56, 58, 62]. However, there are clear morphological differences in the assemblies that are formed and the bacterial phenotypes that are observed for cells expressing wild type FtsZs versus variants where the CTT has been deleted [56, 58, 62]. Deletion of the CTL has several consequences [50, 57, 58, 61]. These include reducing GTP hydrolysis and subunit exchange, inhibiting cell division, and eventually causing cell death via a filamenting phenotype. Systematic *in vitro* studies have also begun to uncover the mechanisms by which CTTs influence FtsZ assembly [3, 47, 50, 56-62]. Specifically, *in vitro* studies show that the CTP influences the lateral associations among protofilaments and serves as a binding motif for modulatory proteins that regulate FtsZ assembly *in vivo* [47, 63-65].

To understand why the disordered CTT is essential for FtsZs *in vivo*, we focus here on *in vitro* biophysical and biochemical investigations to uncover the functions of each of the FtsZ modules / domains using FtsZ from *B. subtilis* (Bs-FtsZ) as an archetypal system. Our studies, which build on previous foundational contributions [3, 47, 50, 56-60, 62], are directed at four different constructs *viz.*, full-length Bs-FtsZ (designated as WT),  $\Delta$ CTT which lacks the C-terminal tail,  $\Delta$ CTL which lacks the C-terminal linker, and  $\Delta$ CTP which lacks the C-terminal peptide (**Figure 2.1**). For each of these constructs, we quantify the following quantities: (i) The concentration dependence and morphologies of different assemblies; (ii) the apparent affinity for GTP; and (iii) the efficiency of GTP hydrolysis. Taken together, the results from our systematic experiments provide clear insights regarding the contributions made by each of the modules of Bs-FtsZ in controlling assembly, cooperativity, and regulating GTPase activity.



**Figure 2.1: Design of C-terminal tail truncation variants.** Schematics showing the block architectures of the four constructs *viz.*, WT,  $\Delta$ CTT,  $\Delta$ CTL, and  $\Delta$ CTP

## 2.3 Material and Methods

### 2.3.1 Protein expression and purification

Plasmids for Bs-FtsZ and the tail variants were acquired from the Levin Lab. These were cloned into the pET-21b(+) expression vector through *E. coli* strain AG1111. The resulting plasmids were mini-prepped and freshly transformed into C41(DE3) cells and made into glycerol stocks. With minimal adjustments, the growths and purifications were performed using the protocol deployed in Buske et al., [57]. Here, 500 mL of Lysogeny broth (LB) medium was inoculated 1:100 with an overnight culture. Cells were grown at 37°C until A600 ~0.6-0.8, and then cells were induced with 1 mM isopropyl  $\beta$ -d-1-thiogalactopyranoside. Cells were then grown for an additional 4h at 37°C. The cells were then harvested by centrifugation, and the cell pellets were stored at -80°C. Purification was performed on an AKTA FLPC system using a Superdex 75 size exclusion column. Peak fractions were analyzed by SDS-PAGE, pooled together, and dialyzed overnight in 1 L of FtsZ dialysis buffer (50 mM 2-Morpholinoethanesulfonic acid (MES), 50 mM KCl, 2.5 mM Mg<sub>2</sub>Cl<sub>2</sub>, 1 mM EGTA, pH 6.5). Protein concentration was quantified using UV-Vis via phenylalanine absorbance at 280 nm ( $\epsilon = 2980 \text{ M}^{-1} \text{ cm}^{-1}$ ) and confirmed using a Pierce 660 nm assay with BSA or tubulin as a standard (Thermo Fisher Scientific). Purified protein stocks were concentrated, separated into aliquots, flash frozen on liquid N<sub>2</sub>, and stored at -80°C. Prior to use, FtsZ aliquots were thawed on ice, well mixed, and the concentration was reconfirmed using UV-Vis absorbance.

### **2.3.2 Dynamic Light Scattering (DLS) assay**

FtsZ was diluted into an MES reaction buffer (50 mM MES, 2.5 mM Mg<sub>2</sub>Cl<sub>2</sub>, 1 mM EGTA, 1 mM GTP, pH 6.5) with KCl concentrations varying from 50-200 mM KCl as specified) to a total volume of 50  $\mu$ L. All samples were incubated at 30°C with 1 mM GTP for 5 minutes prior to measurement on a Malvern ZEN3600 Zetasizer instrument (Nano Research Facility, Washington University in St. Louis). For the data collection and analysis, a standard program for aggregating proteins was used that was developed by the instrument manufacturer. The inherent output of dynamic light scattering measurements is an autocorrelation function of scattering intensity, which is a deconvolution of the contributions of the size of scatterers and the number density of scatterers. The intensity in the Rayleigh regime scales as the sixth power of the size of the scatterer. This intrinsic weighting of the scattering intensity is accounted for when analyzing autocorrelation function to extract percent number distributions, which are the preferred quantities for comparative analyses of DLS data across different constructs.

### **2.3.3 Transmission Electron Microscopy (TEM)**

Samples were prepared in MES reaction buffer with 6  $\mu$ M of FtsZ and the specified KCl concentration. Prior to preparing the glow-discharged copper grids, each sample was incubated for 5 minutes in the presence of 1 mM GTP to allow for adequate assembly. Each sample was quickly rinsed with deionized (DI) water, then stained three times with 1% uranyl acetate for 5-seconds each, with the solution wicked away, and followed by a 10 second wait time in between stains. Samples were visualized using an FEI Company Transmission Electron Microscope (Nano Research Facility, Washington University in St. Louis).

### 2.3.4 GTPase assay

GTP hydrolysis activity was monitored using a coupled GTPase assay [66]. For this, we used a 96-well TECAN plate reader, and measurements were conducted in an MES reaction buffer that included 1 mM phosphoenolpyruvate, 250  $\mu$ M NADH, and 40 units/ml of both lactose dehydrogenase and pyruvate kinase. From the following equation, the linear decline of NADH absorbance at 340 nm was monitored over 30 minutes. The steepest decline rate for a 5-minute consecutive stretch was related to the GTPase activity by the following manipulation of Beer's law [67], which yields:  $\left( \frac{\text{moles of GTP hydrolyzed}}{\text{min}} \right) = \left( \frac{\Delta A_{340}}{\epsilon_{\text{NADH}} L V_a} \right)$ . Here,  $\Delta A_{340}$  is the slope of the decline,  $\epsilon_{\text{NADH}}$  is the extinction coefficient for NADH at 340 nm ( $6220 \text{ M}^{-1}\text{cm}^{-1}$ ),  $L$  is the path length of the cuvette (0.401 cm), and  $V_a$  is the observation volume (150  $\mu$ L). Each trial was performed at least three times.

### 2.3.5 Determination of $c_A$

The threshold concentration ( $c_A$ ) and its variance were determined by randomly selecting a trial at each considered concentration and prescribing a linear fit [68]. This process was repeated  $10^3$  times. We only considered points that correspond to non-zero values along the ordinate. The mean and variance were then calculated from the resulting vector of intercepts along the abscissae. All analyses were performed using MATLAB.

### 2.3.6 Michaelis-Menten fits

Data for enzyme kinetics were fit to:  $a = \left( \frac{v_{\max} S}{K_M + S} \right)$ . Here,  $a$  is the measured GTPase

activity,  $S$  is the substrate concentration, which is GTP, and  $K_M$  is the apparent Michaelis constant.

A random trial was selected from each tested GTP concentration, and the data were fit to the equation for enzymatic activity. Values of  $K_M$  and  $v_{\max}$  were determined using a nonlinear regression model based on the *fitnlm* function in MATLAB. This process was repeated  $10^3$  times using bootstrapping with replacement to extract a mean and variance. The data were also fit using

a Hill model, where  $a = \left( \frac{v_{\max} S^n}{K_M^{1/n} + S^n} \right)$ . The two models were compared using the residuals from

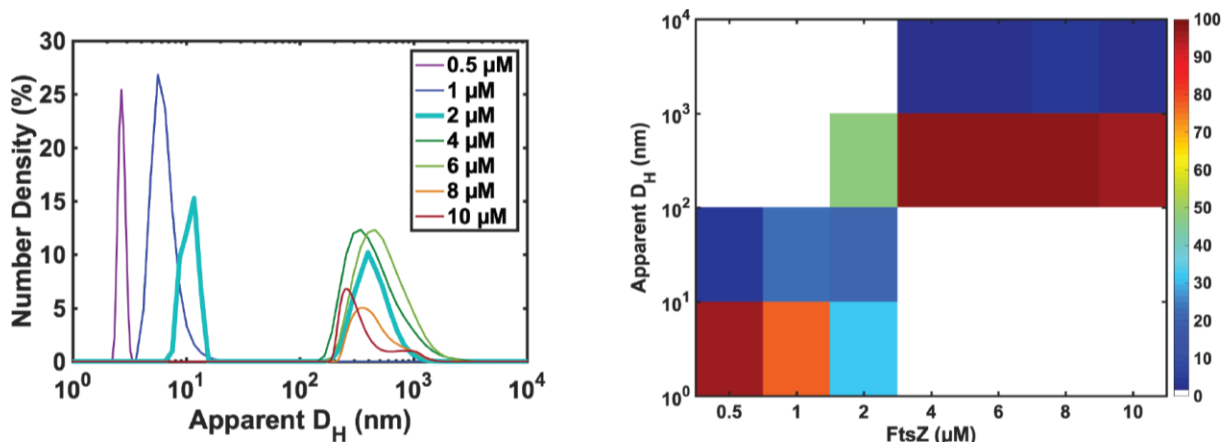
fitting the data to determine the best fit. In all cases, the Michaelis-Menten model provided the best fit to the activity data (see Appendix).



## 2.4 Results

### 2.4.1 The CTL and the CTP modules have different effects on Bs-FtsZ assembly

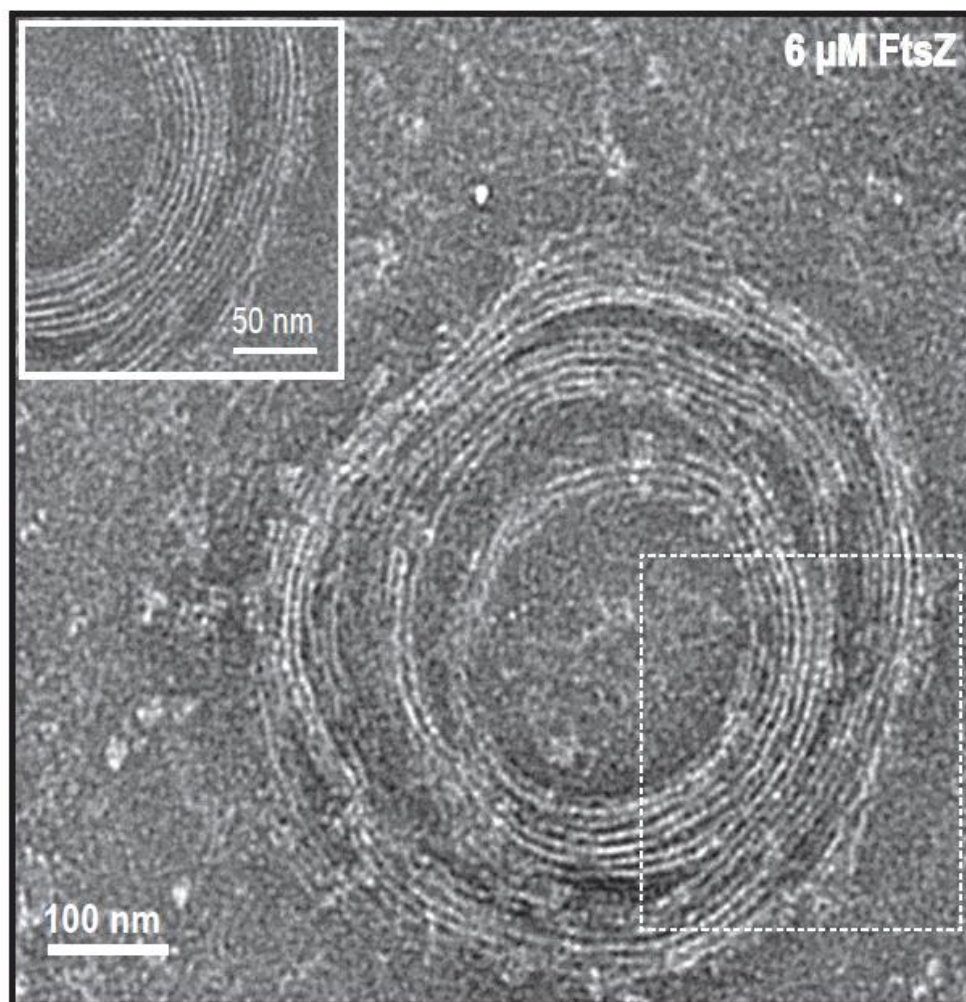
In the presence of GTP, FtsZ can form single-stranded linear polymers that associate laterally to form higher-order assemblies. The sizes and morphologies of assemblies are likely to be heterogeneous and dependent on FtsZ concentration. We used dynamic light scattering (DLS), a correlation spectroscopy method [69, 70], to uncover the size distributions of assemblies formed by FtsZ as a function of protein concentration in an MES reaction buffer (50 mM MES, 50 mM KCl, 2.5 mM MgCl<sub>2</sub>, 1 mM EGTA, 1 mM GTP, and pH 6.5). The time correlation functions from raw scattering data yielded distributions of scattering intensities that were converted to number density distributions. These distributions quantify the number densities of hydrodynamic diameters ( $D_H$ ) that are observed in solution. The conversion between intensity and number distributions is obtained using an instrument-specific algorithm based on Mie's theory for scattering [70].



**Figure 2.2: Concentration-dependent assembly of FtsZ.** (Left) Dynamic light scattering (DLS) measurements are used to estimate the fraction of the total number of scatterers that are of a specific apparent diameter ( $D_H$ ). The percent number density is plotted as a function of  $D_H$  for Bs-FtsZ concentrations of 0.5, 1, 2, 4, 6, 8, and 10  $\mu$ M. This analysis shows that the presence of a threshold concentration of 2  $\mu$ M is evident in the bimodality of the population distribution (bolded cyan curve). (Right) The number density function inferred from DLS data is integrated and represented using a checkerboard heat map, where each box shows the percent occupancy for species that lie within specific intervals of  $D_H$  values (ordinate) for a specific protein concentration (abscissa). Boxes shown in white correspond to intervals with zero occupancies.

First, we analyzed the size distributions of assemblies formed by the WT construct as a function of protein concentration. Data for the number density distributions are shown in **Figure 2.2**. At a protein concentration of 2  $\mu$ M, we observe a bimodal distribution. Following Kozlov et al. [71], we interpret this bimodality to imply cooperativity in assembly, which for FtsZ implies the presence of a concentration threshold of  $\sim 2\mu$ M where we observe the coexistence of two different types of assemblies *viz.*, 10 nm-sized assemblies that are most likely protofilaments that coexist with higher-order assemblies that are likely to be bundles of filaments and / or other types of structures. The DLS data are further summarized by integrating the number distribution to quantify the populations associated with assemblies that correspond to specific size intervals. Although number density distributions are complete representations of the DLS data, they are probability density functions (PDFs). To enable comparative assessments across different solution

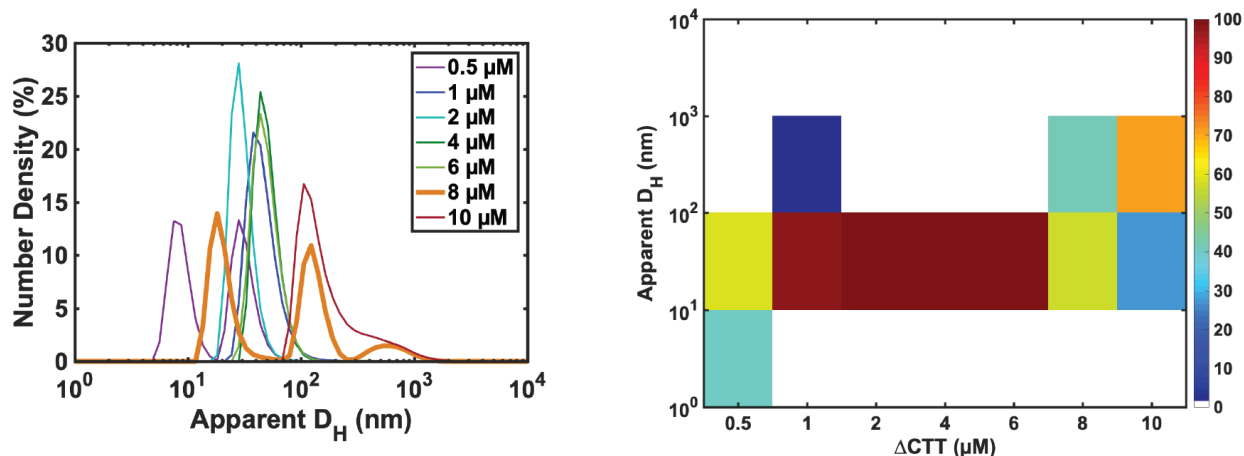
conditions and constructs, we summarize the information contained in the DLS number density distributions as interval-specific cumulative distribution functions (CDFs), which are integrals of PDFs, i.e., areas under the number density distributions calculated across specific size intervals. CDFs are optimal choices because they quantify the total probability rather than the probability density. The CDFs, calculated as integrals over the PDFs across specific intervals, are displayed in heatmaps. The color within each box quantifies the total percent probability of observing scatterers within the size interval corresponding to the box in question (**Figure 2.2**). The advantages of CDFs over PDFs have been well established [72], especially for comparative assessments across different distributions. We use these as devices that enable comparative assessments of differences in assembly sizes at various protein concentrations.



**Figure 2.3: TEM image of WT FtsZ.** Transmission electron microscopy (TEM) indicates the formation of higher-order assemblies such as multi-filament bundles and wreath-like structures at 6  $\mu\text{M}$  FtsZ.

**Figure 2.2** shows evidence for a threshold protein concentration of 2  $\mu\text{M}$  beyond which the assemblies that form in solution are larger than 100 nm. Transmission electron microscopy (TEM) images obtained at a concentration of 6  $\mu\text{M}$ , a representative of which is shown in **Figure 2.3**, show the bundling of filaments into wreath-like structures that are larger than 100 nm in diameter. In contrast, the assemblies that form at concentrations below 2  $\mu\text{M}$  are smaller than 100 nm. The concentration-dependent DLS data for WT Bs-FtsZ point to the existence of a threshold

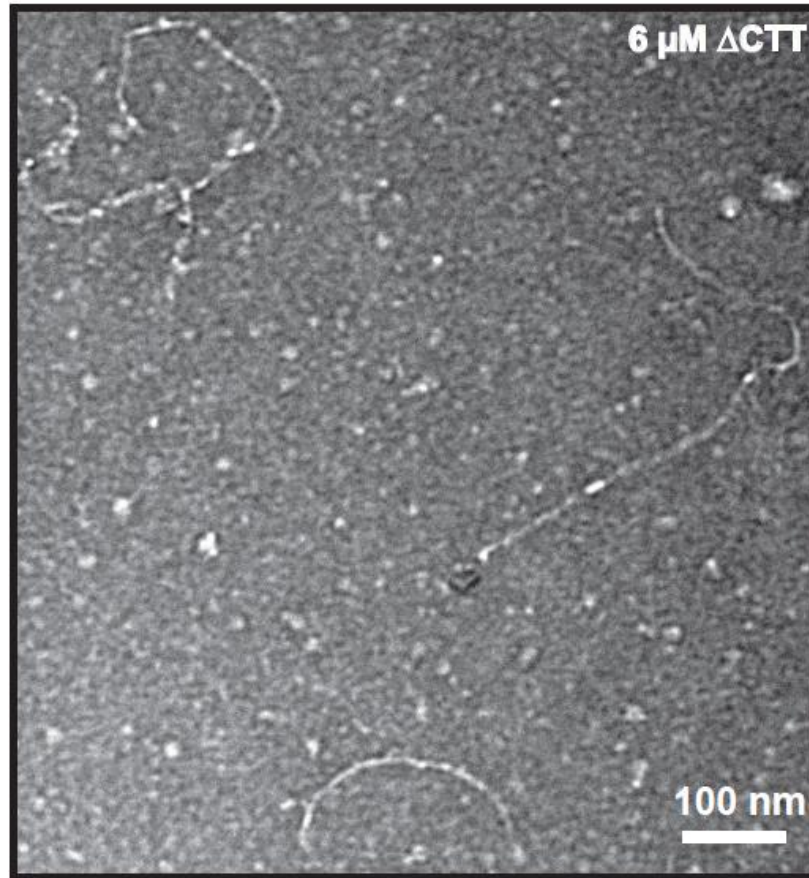
concentration,  $c_B \approx 2 \mu\text{M}$  for the formation of higher-order assemblies of filaments in the solution conditions deployed in our experiments.



**Figure 2.4: Concentration-dependent assembly of  $\Delta\text{CTT}$ .** (Left) For  $\Delta\text{CTT}$ , the number density probabilities as a function of  $D_H$  shows that most structures fall between 101 and 102 nm. There is a bimodality in the distribution at 8  $\mu\text{M}$  (bolded orange curve) (Right) Checkerboard heatmap of the  $\Delta\text{CTT}$   $D_H$  distribution as measured by DLS. The particle size is enhanced at low concentrations compared to WT and largely maintains size with an increase in concentration.

Next, we asked how the deletion of the CTT, the CTP, and the CTL impact the assemblies formed under conditions that were identical to those used for the study of the WT. **Figure 2.4** summarizes the DLS data for  $\Delta\text{CTT}$ . At low protein concentrations, 0.5  $\mu\text{M}$  – 2  $\mu\text{M}$ ,  $\Delta\text{CTT}$  forms assemblies that are larger than those observed for WT. The concentration at which we observe a bimodal distribution for the number density of scatterers shifts upward to 8  $\mu\text{M}$  vis-à-vis the threshold of 2  $\mu\text{M}$  for the WT. This is similar to previous observations of the condensation of  $\Delta\text{CTT}$  filaments in the presence of crowders and stabilizers [23, 56, 58]. Our data suggest that while  $\Delta\text{CTT}$  forms larger filaments when compared to WT, the absence of the CTT weakens its ability to form higher-order assemblies. Support for this inference comes from direct visualization using TEM, which shows the formation of long, single-stranded filaments and the absence of

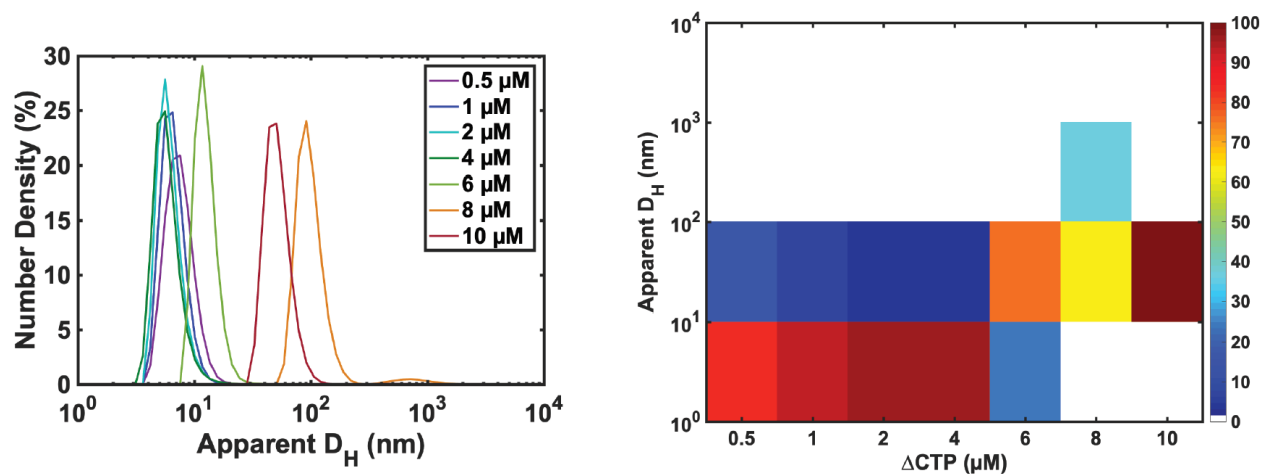
bundles or wreaths at a protein concentration of 6  $\mu\text{M}$  (**Figure 2.5**). These morphologies are distinct from those observed for WT under identical solution conditions (**Figure 2.3**).



**Figure 2.5: TEM image of long single-stranded protofilaments of  $\Delta\text{CTT}$ .** TEM reveals that these sizes correspond to long, linear, and curved protofilaments.

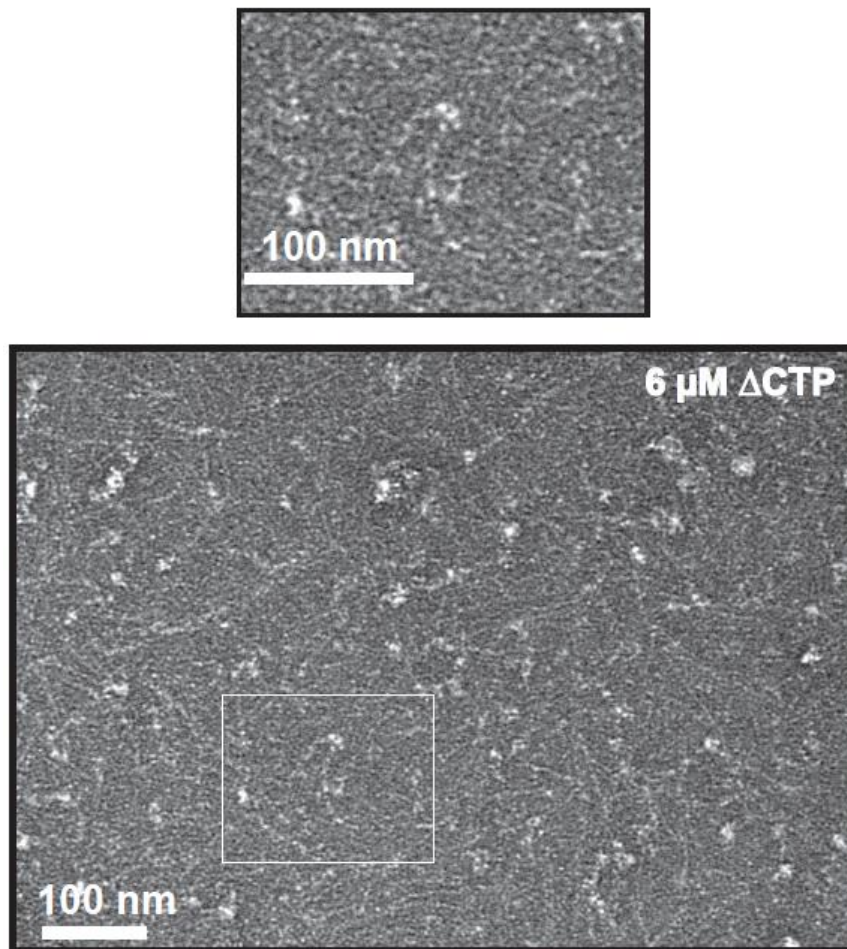
In order to understand why  $\Delta\text{CTT}$  is able to form long, single-stranded polymers more robustly than WT while showing a weakened ability to form higher-order assemblies, we analyzed the DLS and TEM data for  $\Delta\text{CTP}$  and  $\Delta\text{CTL}$ . These constructs help us separate contributions made by the CTL and the CTP to the overall behavior of WT. Surprisingly, we observe a significant diminution in the sizes of assemblies formed by  $\Delta\text{CTP}$ . It appears that the disordered CTL suppresses assembly on all length scales, leading to the formation of assemblies that are at least an

order of magnitude smaller than  $\Delta\text{CTT}$  across protein concentrations ranging from  $0.5\ \mu\text{M}$  –  $8\ \mu\text{M}$  (Figure 2.6).



**Figure 2.6: Concentration-dependent assembly of  $\Delta\text{CTP}$ .** (Left) Number density (%) versus  $D_H$  for various  $\Delta\text{CTP}$  concentrations. No significant bimodalities are observed. (Right) Heatmap of the  $\Delta\text{CTP}$  DLS data shows a low instance of large assemblies and mostly small structures at least an order of magnitude smaller than the structures formed by the  $\Delta\text{CTT}$  variant.

Unlike the WT and  $\Delta\text{CTT}$  constructs, we do not observe the onset of bimodality in any of the number density distributions for  $\Delta\text{CTP}$  (Figure 2.6). The DLS data suggest that, in the concentration range studied here,  $\Delta\text{CTP}$  likely forms short linear polymers that do not form bundles or other higher-order assemblies. This conjecture is confirmed by TEM analysis, performed at a  $\Delta\text{CTP}$  concentration of  $6\ \mu\text{M}$ , which is identical to the concentrations used for WT and  $\Delta\text{CTT}$ . The TEM image shows the formation of significantly shortened linear polymers and an absence of higher-order assemblies (Figure 2.7). We conclude, in accord with previous findings of Huecas et al. [56], that the CTL acts as an excluded volume *spacer* that substantially weakens the formation of single-stranded polymers and inhibits lateral associations.

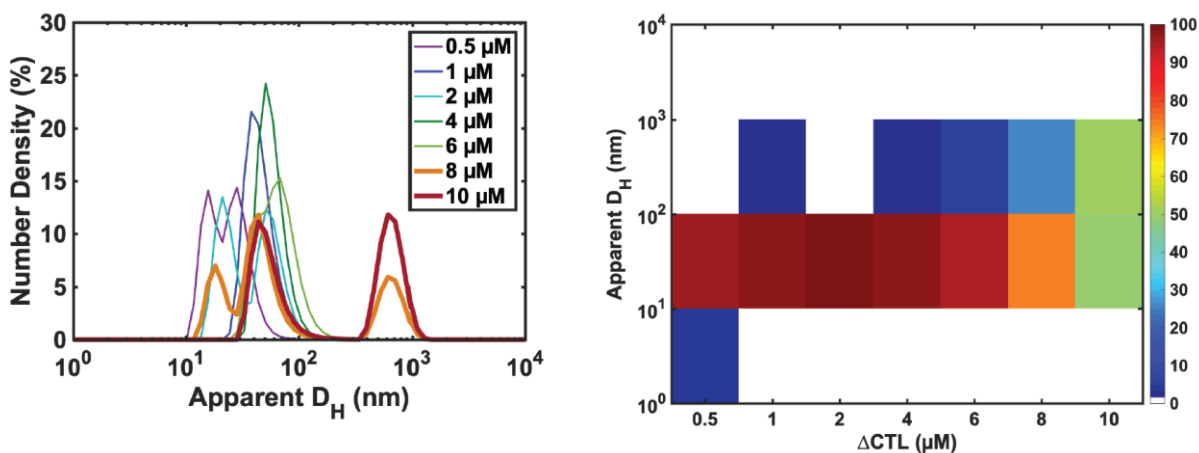


**Figure 2.7:  $\Delta$ CTP TEM image.** This assembly is predominantly short single-stranded protofilaments, as observed with TEM.

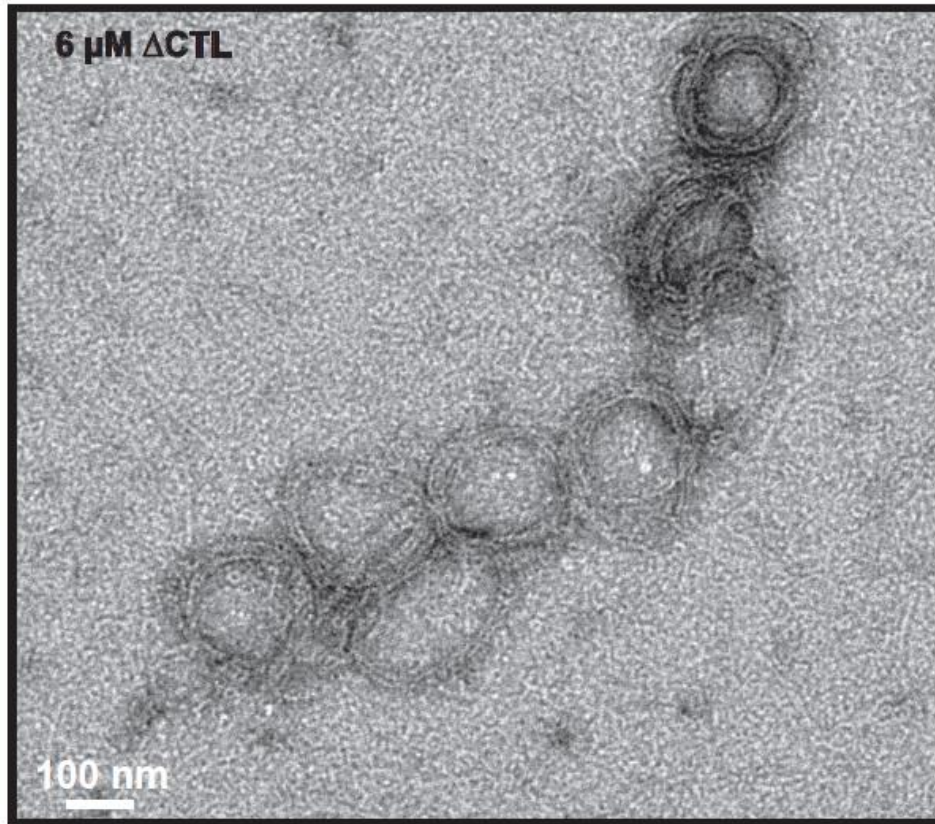
Finally, we studied the assemblies and morphologies formed by  $\Delta$ CTL, the construct lacking the C-terminal linker (**Figures 2.8**). In the concentration range between 0.5  $\mu$ M and 6  $\mu$ M, we observe a clear upward shift in the sizes of assemblies formed by  $\Delta$ CTL when compared to those of  $\Delta$ CTT and  $\Delta$ CTP. TEM analysis, performed at a  $\Delta$ CTL concentration of 6  $\mu$ M, shows the formation of alternate morphologies (**Figure 2.9**). We refer to these morphologies as *mini rings* because they appear to be miniature versions of the wreath / ring-like structures that we observe for the WT under similar conditions. Direct comparisons of the DLS data for WT and  $\Delta$ CTL show



that the mini rings are at least an order of magnitude smaller in size than the structures formed by WT. The number density distributions become bimodal and even multi-modal above protein concentrations of 8  $\mu\text{M}$ , which points to association or condensation of the mini-ring structures. The DLS data and TEM analysis suggest that the CTP module is akin to a *sticker* that engages in cohesive intramolecular interactions to drive the formation of alternative assemblies. The implication is that in the WT, the CTL provides spatial separation between the core and the CTP that enables the formation of larger assemblies by facilitating polymer formation through core-mediated interactions and lateral associations via CTP-mediated interactions.



**Figure 2.8: Concentration-dependent assembly of  $\Delta\text{CTL}$ .** (Left) A plot of the number density against  $D_H$  for various  $\Delta\text{CTL}$  concentrations. The populations are more heterogeneous than  $\Delta\text{CTT}$ , and the clearest bimodality is observed at 10  $\mu\text{M}$ . (Right) DLS measurements of  $\Delta\text{CTL}$  show that the distributions at each protein concentration more closely match those of  $\Delta\text{CTT}$  than that of WT. However, enhanced assemblies are shown at lower concentrations. At high concentrations, these assemblies are still smaller than those observed by WT



**Figure 2.9: TEM image of  $\Delta$ CTL.** TEM shows that the higher-order assemblies formed by the  $\Delta$ CTL variant are rings that are visibly smaller than WT rings. These are referred to as mini rings.

Taken together, the data presented in **Figures 2.2-9** lead to a coherent picture for how the different modules / domains of FtsZ contribute to the assembly of Bs-FtsZ. Of the four constructs we studied, the  $\Delta$ CTT forms the longest single-stranded polymers. Deletion of the CTT weakens the ability of filaments to form higher-order structures. Our data suggest that the CTP and CTL modules may be thought of as *stickers* and *spacers*, respectively [56, 73-75]. The CTP sticker provides cohesive interactions and appears to engage in interactions with the core and with itself, thereby giving rise to mini-ring structures that are fundamentally different from the single-stranded polymers formed by  $\Delta$ CTT (compare **Figures 2.5** and **2.9**). The observation of bi- and multimodality in the DLS data suggests that the CTP can also promote lateral associations and

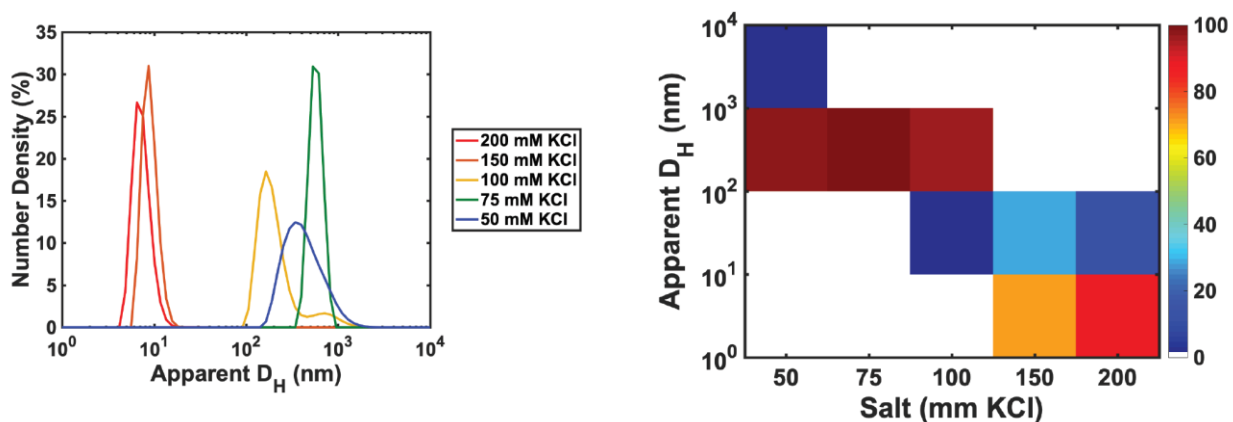
lead to the formation of higher-order assemblies. However, without the CTL providing the necessary spacing between the core and the CTP, the higher-order assemblies formed by  $\Delta$ CTL are smaller than those formed by WT. In direct contrast, the CTL acts as a spacer that carves out an excluded volume, thus ensuring spatial separation of the CTP sticker from the core and suppressing assembly on all length scales [56].

#### **2.4.2 Increased salt concentration weakens the formation of higher-order assemblies for WT Bs-FtsZ**

Previous work showed that a positively charged CTP is necessary to drive the bundling of FtsZ protofilaments [47]. Studies have also shown that increases in the concentration of monovalent salts can reduce the extent of FtsZ assembly [36, 47]. Therefore, we hypothesize that CTP-mediated interactions that are spatially separated from the core are likely to be electrostatic in nature and that bundling must be influenced by alterations to these interactions. Here, we use the term electrostatic in a very generic sense to refer to Coulomb interactions that can be screened due to the presence of dissociated solution ions. This is a zeroth-order assumption, although this does not rule out ion-specific effects, entropic effects due to ion release, or condensation due to ion-mediated associations among like-charged molecules. To test the zeroth-order hypothesis, we analyzed the salt (KCl) dependence of higher-order assemblies formed by WT Bs-FtsZ. We compared these results to the variant that lacks the CTP.

We measured the size distributions of FtsZ assemblies as a function of increased KCl concentration for WT Bs-FtsZ. These measurements were performed in an otherwise consistent MES reaction buffer at 6  $\mu$ M WT Bs-FtsZ. Under these conditions, WT Bs-FtsZ forms large

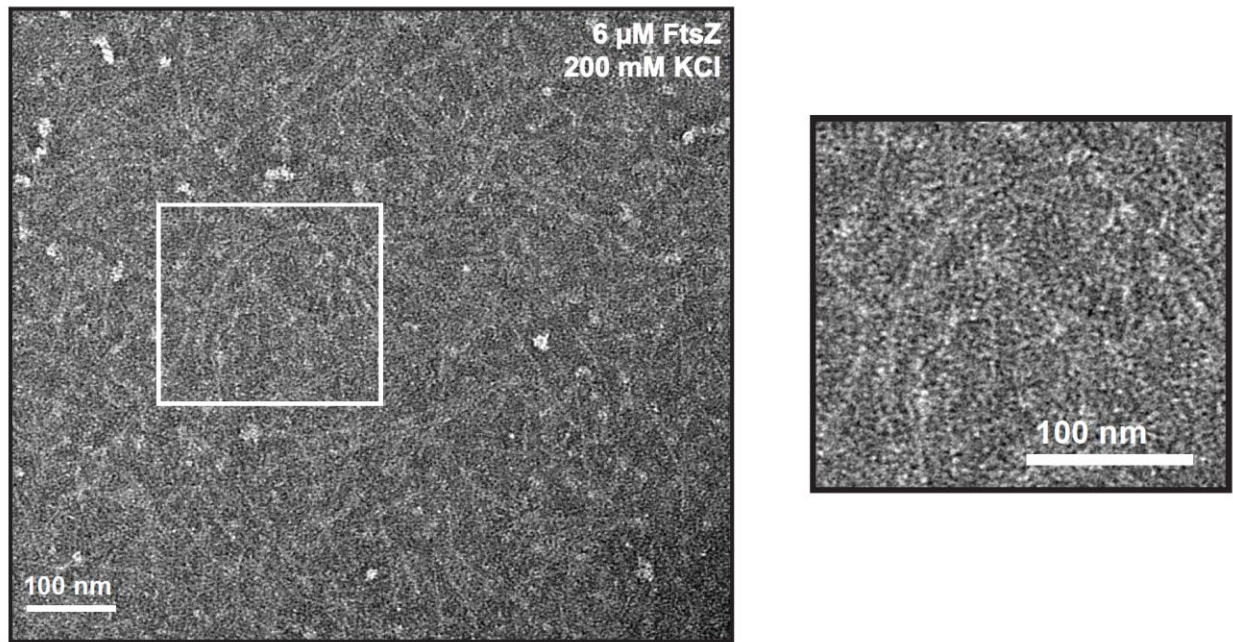
wreath-like structures (**Figure 2.3**). As the concentration of KCl in the buffer increases, we observe a concomitant reduction in the sizes of assemblies formed by WT Bs-FtsZ. This is summarized in the number density distributions extracted from DLS measurements and the integrals of these distributions presented as checkerboard heatmaps (**Figure 2.10**). Sizes of the dominant assemblies in solution decrease by two orders of magnitude as the salt concentration increases from 50 mM KCl to 200 mM KCl (**Figure 2.10**). The size distributions at relatively high salt are similar to those formed by  $\Delta$ CTP at similar protein concentrations.



**Figure 2.10: Concentration-dependent assembly of WT with an increasing concentration of KCl.** (Left) DLS measurements are used to estimate the percentage of the total number of particles that are of a specific apparent diameter ( $D_H$ ) for KCl concentrations of 50, 75, 100, 150, and 200 mM. (Right) Heatmap of the DLS data for different salt concentrations shows that the apparent size of the assemblies decreases with increasing salt concentration. FtsZ assembly size decreases by over an order of magnitude as the salt concentration is increased to 200 mM. The size of the assemblies formed in 150 and 200 mM KCl is similar to the size formed by  $\Delta$ CTP at the same concentration of 6  $\mu$ M.

Comparison of TEM images collected using material extracted from solutions with 50 mM KCl (**Figure 2.3**) versus 200 mM KCl (**Figure 2.11**) shows that higher-order structures are disassembled at high salt, and the structures that persist are short single-stranded polymers. In fact, the structures observed in the presence of 200 mM KCl are most similar in morphology to those observed for  $\Delta$ CTP (**Figure 2.7**). The most straightforward interpretation of data shown in **Figure**

**2.10-11** is that lateral associations among protofilaments can give rise to bundled higher-order assemblies, and these are promoted primarily by electrostatic interactions, mediated mainly by the CTP sticker. These interactions are screened at high salt, thereby unmasking the assembly inhibiting effects of the CTL spacer.



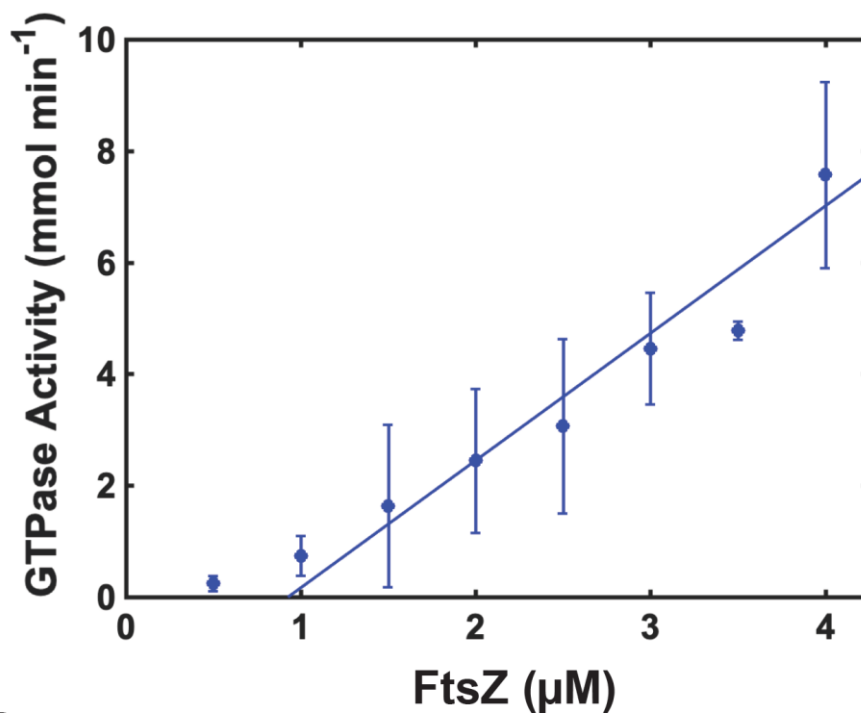
**Figure 2.11: TEM images of WT in high salt.** In 200 mM KCl, TEM images show the formation of single-stranded protofilaments.

Our data thus far suggest that within the WT Bs-FtsZ, core-driven interactions give rise to single-stranded linear polymers. These interactions are weakened by the CTL, which also spatially separates the core from CTP-mediated electrostatic interactions. This spatial separation appears to help drive interactions that lead to the formation of wreath- and ring-like structures. We next asked how distinct domains in the tail contribute to the formation of active polymers.

### 2.4.3 The CTL weakens the driving forces for forming active polymers

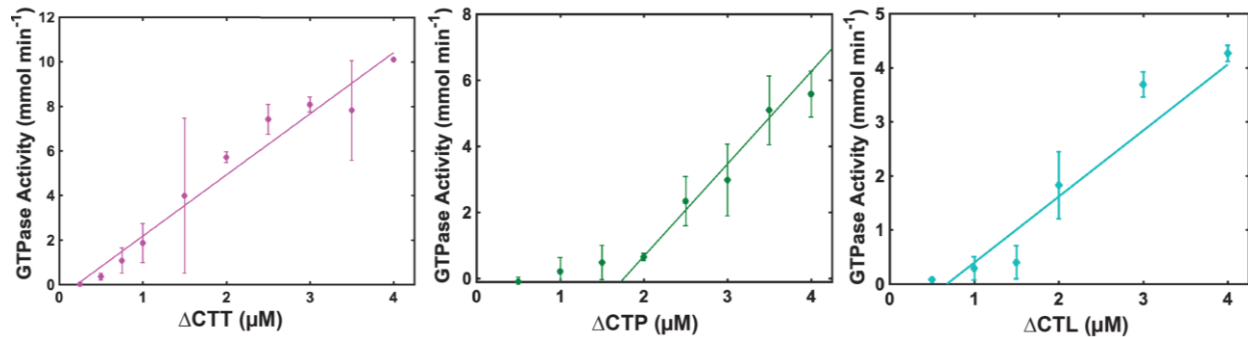
FtsZ is a filament-forming enzyme, and the formation of active polymers requires the crossing of an apparent concentration threshold ( $c_A$ ). We refrain from referring to this as a critical concentration because this parlance has precise implications for concentration fluctuations at critical points in systems that undergo first-order and continuous phase transitions [41, 76]. The presence of an apparent threshold concentration is the hallmark of a cooperative transition [76]. Lower apparent threshold concentrations for forming active polymers imply stronger driving forces and / or higher cooperativity for forming these polymers. By estimating the values of  $c_A$  for each of the four constructs, we investigated the effects of the CTT and modules within the CTT on the strength of driving forces for cooperative assembly into active polymers. It is difficult to estimate the value of  $c_A$  using DLS. This is because the scattering intensity decreases as the concentration is lowered, leading to a significant reduction of signal versus noise.

Additionally, the use of purely spectrophotometric methods based on spinning samples down and separating them into supernatant and pellet are confounded by the lack of tryptophan residues in Bs-FtsZ. Therefore, such an approach requires either amino acid substitutions or the incorporation of fluorescent dyes at suitable positions that do not disrupt WT Bs-FtsZ behavior. These are non-trivial modifications that can confound interpretations, especially given the subtle interplay between CTT modules and the core domain. Therefore, we leveraged extant methods in the FtsZ literature and used the onset of activity as a function of FtsZ concentration as a proxy for quantifying the apparent concentration threshold for the onset of active polymers [38]. Specifically, we measured the GTPase activity (millimoles of GTP hydrolyzed per minute) as a function of FtsZ concentration.



**Figure 2.12: Measurement of GTPase activity (millimoles of GTP hydrolyzed per minute) as a function of WT concentration.** GTPase activity is measured using a regenerative-coupled GTPase assay. The onset of activity pinpoints the location of the threshold concentration for the formation of active polymers. The FtsZ concentration versus GTPase activity data were fit using a linear regression model, and the intercept along the abscissa or the onset of inflection points to the location of  $c_A$ .

The intercept along the abscissa of a linear fit to the activity data is used as an estimate of  $c_A$ . This is valid because GTP binding and dimer formation are obligatory for polymer formation and GTPase activity. In our analysis, we only fit points that have a GTPase activity value greater than zero, taking into account the error in our measurements. Our analysis yielded an estimate for  $c_A$  to be  $0.92 \pm 0.12 \mu\text{M}$  for Bs-FtsZ in 50 mM KCl, consistent with estimates from previous studies showing that  $c_A$  lies between 0.5 and 1  $\mu\text{M}$  [20, 21, 43-46] (**Figure 2.12**).



**Figure 2.13: Measurement of GTPase activity (millimoles of GTP hydrolyzed per minute) as a function of CTT variant concentration.** (Left) GTPase activity as a function of  $\Delta$ CTT concentration. (Middle) GTPase activity as a function of  $\Delta$ CTL concentration. (Right) GTPase activity as a function of  $\Delta$ CTP concentration shows the most apparent evidence for a critical concentration threshold.

**Figure 2.13** shows our data for the GTPase activity of the  $\Delta$ CTT construct. Our analysis of these data yields an estimate of  $0.21 \pm 0.18 \mu\text{M}$  for the  $c_A$  of  $\Delta$ CTT. The lower value of  $c_A$  for  $\Delta$ CTT vis-à-vis WT and its ability to assemble into long single-stranded polymers suggests that active polymers form via a cooperative process that is driven mainly by interactions between Bs-FtsZ cores. Our analysis of the GTPase data for the  $\Delta$ CTP construct yields an estimate of  $\sim 1.76 \pm 0.11 \mu\text{M}$  for  $c_A$  (**Figure 2.13**). This is almost twice the value of  $c_A$  for WT Bs-FtsZ and an order of magnitude higher than  $\Delta$ CTT. These data indicate that the CTL spacer weakens the driving forces for the assembly of single-stranded active polymers while maintaining the cooperative nature of this assembly process. Finally, analysis of the activity data for  $\Delta$ CTL (**Figure 2.13**) yields an estimate of  $0.67 \pm 0.07 \mu\text{M}$  for  $c_A$ , and this value is slightly less than that of the WT construct. Overall, our estimates for  $c_A$  follow the trend whereby  $c_A(\Delta\text{CTT}) < c_A(\Delta\text{CTL}) < c_A(\text{WT}) < c_A(\Delta\text{CTP})$ . These data show that as a whole, the CTT weakens the driving forces for forming active polymers. This weakening of the driving forces for forming active polymers derives

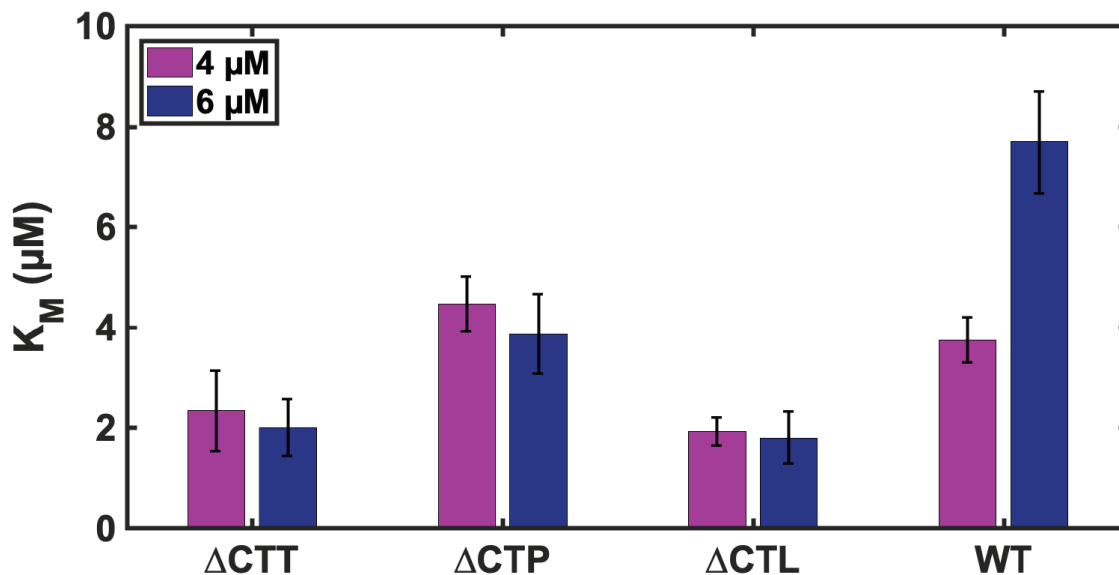


exclusively from the CTL spacer and is overcome by the CTP sticker, although not enough to lower the value of  $c_A$  to be equivalent to that of  $\Delta CTT$ .

#### 2.4.4 The CTL and CTP have opposing effects on the enzymatic activity

The Michaelis-Menten formalism [77] has been used as a minimal model for analyzing the enzyme kinetics of FtsZs [38, 78]. However, FtsZ is a filament-forming enzyme, and its overall activity will be a convolution of contributions from the full range of polymers and higher-order assemblies. Extant methods do not allow us to parse the species-specific contributions to enzyme kinetics. Accordingly, we analyzed measurements of enzyme kinetics for each of the four constructs using two minimal models *viz.*, the Michaelis-Menten model [77] and the Hill model [79], respectively. In general, the data for enzyme kinetics measured for all constructs at two different concentrations (4  $\mu\text{M}$  and 6  $\mu\text{M}$ ) are optimally fit using the Michaelis-Menten model (**Figure 2A.1**) when compared to the Hill equation (**Figures 2A.2**). It is worth noting that while the data from measurements of GTPase activity versus GTP concentration are shown to obey Michaelis-Menten kinetics for all four constructs (**Figures 2A.1**), the inferred value of the apparent Michaelis constant  $K_M$  *viz.*, the concentration of GTP at which 50% of the sites from of all active species in the ensemble are occupied increases with FtsZ concentration for WT (**Figure 2.14**). This reflects the complexities associated with analyzing enzyme kinetics using a simple model that does not account for species-specific contributions from different types of assemblies formed by the WT enzyme. Our analysis in **Figure 2.2** shows that the WT enzyme forms a mixture of single-stranded polymers and higher order structures at concentrations of 4  $\mu\text{M}$  and 6  $\mu\text{M}$ . However, at these concentrations, higher-order assembly is either minimal or non-existent for  $\Delta CTT$ ,  $\Delta CTP$ , and  $\Delta CTL$ , respectively (**Figures 2.2 – 10**). Accordingly, for each  $\Delta CTT$ ,  $\Delta CTP$ , and  $\Delta CTL$  the

apparent  $K_M$  values we obtain within error do not show a dependence on enzyme concentration (Figure 2.14).

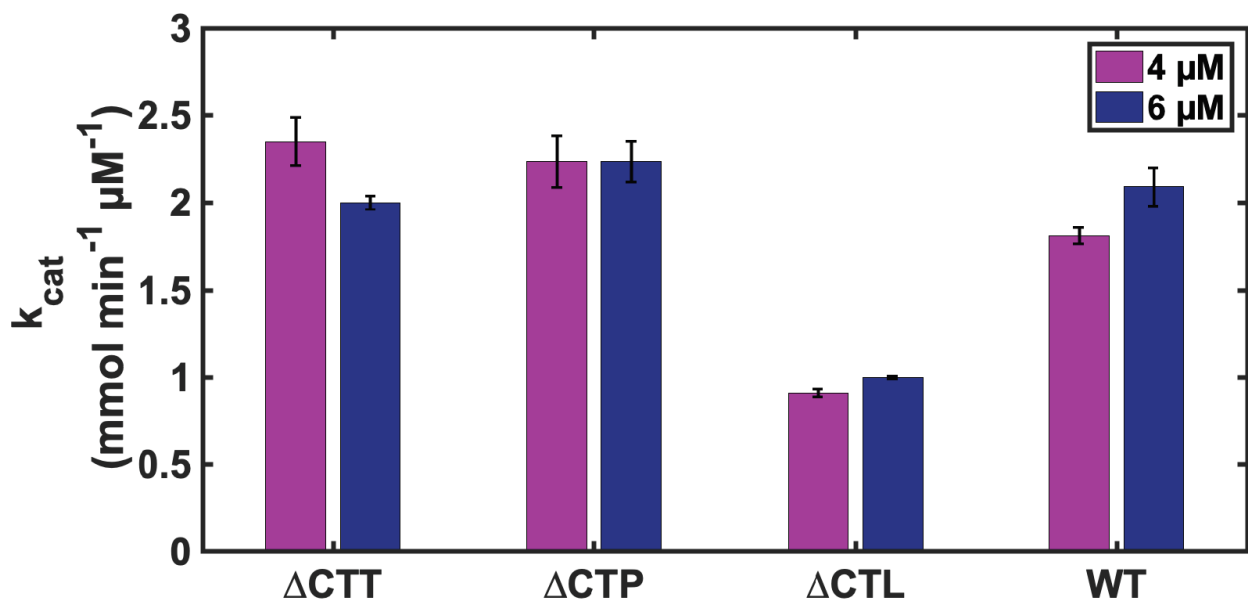


**Figure 2.14: Bar plots of the apparent Michaelis-Menten constants ( $K_M$ ).** The  $K_M$  for each variant at 4 and 6  $\mu\text{M}$  enzyme shows that the  $\Delta\text{CTP}$  enzyme has the highest apparent  $K_M$ .

With the caveat that the enzyme kinetics for WT likely reflect a convolution of multiple species, we compared the construct-specific values for  $K_M$  and  $v_{\text{max}}$  and used these to quantify enzyme efficiencies. We reiterate that this exercise is most meaningful for the different deletion constructs and less so for WT.

Overall, our measurements show that FtsZs that lack the CTL, as in  $\Delta\text{CTL}$  and  $\Delta\text{CTT}$ , have the lowest apparent  $K_M$  values (Figure 2.14). The addition of the CTL, which suppresses polymerization and inhibits higher-order assembly, leads to an increase in apparent  $K_M$  (see data for  $\Delta\text{CTP}$ ). For the WT, increased bundling increases the apparent  $K_M$ . A parsimonious interpretation of the inferred values for the apparent  $K_M$  is that the CTL weakens GTP binding, whereas the CTP enhances GTP binding. At this juncture, this interpretation is highly qualitative,

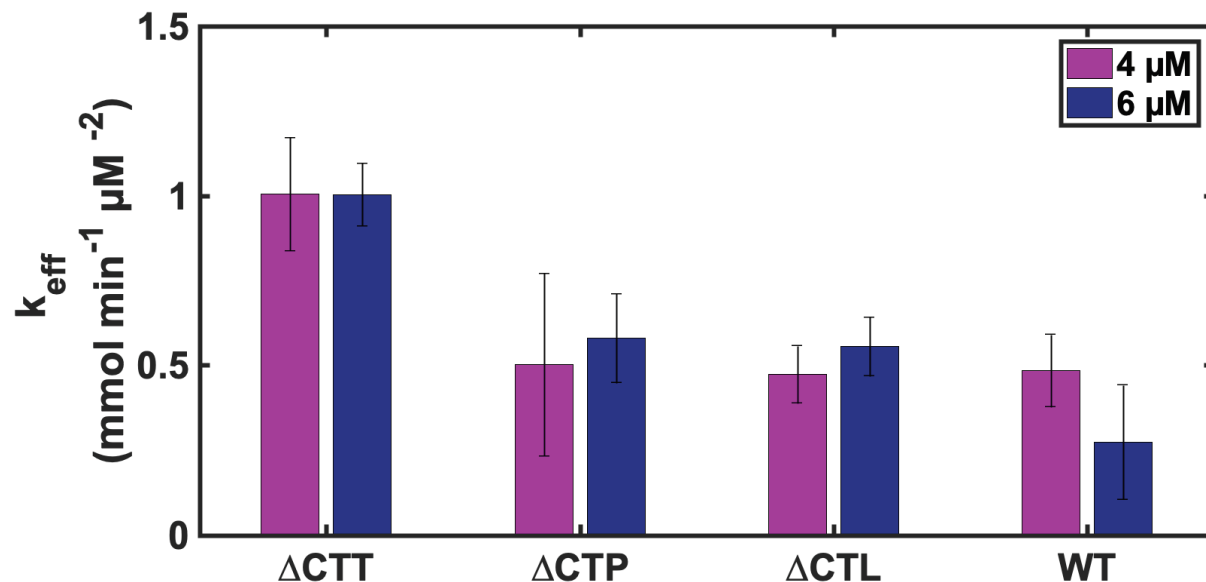
and further work is needed to dissect the species-specific contributions to enzymatic activity. Unfortunately, this remains a serious challenge, and we envisage the need for an elaborate joint effort that combines measurements of single filament dynamics and mesoscopic simulations that are suitably parameterized to enable testable predictions.



**Figure 2.15:** Bar plots of the catalytic rates ( $k_{\text{cat}}$ ) for each variant at 4 and 6  $\mu\text{M}$ . These parameters show that the CTP slows the catalytic rate ( $\Delta\text{CTL}$  data) in comparison to the  $\Delta\text{CTT}$  enzyme.

The catalytic rate constant ( $k_{\text{cat}}$ ) or turnover number calculated as the ratio of  $v_{\text{max}}$  to the total concentration of the enzyme is higher for  $\Delta\text{CTT}$  when compared to WT (**Figure 2.15**). Since the rate of GTP hydrolysis is coupled to the rate of subunit turnover, this likely corresponds to an enhanced rate of subunit exchange [32, 33]. We find that the CTP reduces the catalytic rate of the core, as shown by data obtained for  $\Delta\text{CTL}$ , indicating that the subunit exchange is slowed by the CTP. Data for the  $\Delta\text{CTP}$  construct suggest that having only the CTL and not the CTP restores, and even slightly enhances, the catalytic rate of the core alone. The WT construct falls in the middle

of these two extremes, indicating that the reduced subunit exchange due to the presence of the CTP sticker can be offset by the presence of the CTL spacer.



**Figure 2.16:** Bar plots of the catalytic efficiency ( $k_{\text{eff}}$ ) for each variant at 4 and 6  $\mu\text{M}$ . The variant without the CTT is the most efficient enzyme.

Our data suggest that the modules of the CTT contribute to an auto-inhibitory function as far as overall enzyme activity is concerned. Indeed, the  $\Delta\text{CTT}$  is the most efficient enzyme, where efficiency ( $k_{\text{eff}}$ ) is quantified by the ratio  $k_{\text{cat}}/K_{\text{M}}$  (**Figure 2.16**) at both concentrations (4  $\mu\text{M}$  and 6  $\mu\text{M}$ ). While the  $k_{\text{cat}}$  of  $\Delta\text{CTL}$  is less than that of  $\Delta\text{CTP}$ , the opposite is true for the apparent  $K_{\text{M}}$ . This leads to an approximately equivalent  $k_{\text{eff}}$  for  $\Delta\text{CTP}$  and  $\Delta\text{CTL}$ . WT Bs-FtsZ is the least efficient enzyme, emphasizing auto-inhibition by the tail through a combination of the contributions of the  $\Delta\text{CTL}$  and  $\Delta\text{CTP}$  modules and the ability of WT to form bundles that lead to a reduction of overall activity.

## 2.5 Discussion

The intrinsically disordered CTT of Bs-FtsZ has two distinct modules, *viz.*, a hypervariable CTL and a well-conserved CTP [80]. In this work, we focused on dissecting the functional contributions of each of these modules within the CTT and the overall contribution of the CTT. Taken together, our data suggest that the CTT acts as an *auto-regulator* of assembly and an *auto-inhibitor* of enzymatic activity. We find that the CTL weakens the driving forces for forming single-stranded polymers, thereby leading to shorter polymers when compared to  $\Delta$ CTT. The CTL also inhibits the formation of higher-order assemblies. From a functional standpoint, the CTL acts as a *spacer* with a finite excluded volume and helps in spatially separating the CTP from the core of Bs-FtsZ – a feature that has been established by Huecas et al. for different FtsZs [56]. The ability of the CTL to function as a spacer also helps alleviate interactions involving the CTP, which, if left unregulated, will give rise to alternative mini-ring structures (**Figure 2.9**). The assembly suppressing activities of the CTL appear to ensure that the CTP engages primarily in interactions that drive higher-order assemblies. Additionally, *in vivo*, the spatial delocalization of the CTP from the core should help ensure that the CTP is free to coordinate the network of interactions involving FtsZ and modulatory proteins that help regulate the Z-ring assembly, localization, anchoring, and dynamics [3]. It stands to reason that there will be a competition between homotypic CTP-mediated lateral associations that promote higher-order assemblies of FtsZ and heterotypic interactions that involve the CTP and modulatory proteins. There could also be a synergy between homotypic associations giving rise to lateral associations and vacant CTP sites engaging in heterotypic interactions. Experiments designed to test the interplay between

homotypic and heterotypic CTP-mediated interactions for FtsZs extracted from different bacteria would be of immense utility.

Our data suggest that while the CTL may be thought of as a spacer, the CTP plays the role of a *sticker*. In  $\Delta$ CTL, the CTP enables the formation of alternative mini-ring structures (**Figure 2.9**), whereas, in the WT, the CTP enables the formation of higher-order assemblies. The latter likely arises through CTP-mediated electrostatic interactions (**Figure 2.10**). Our data are in accord with previous findings showing that the CTP promotes lateral bundling of protofilaments in *B. subtilis* [47, 57] and in *E. coli* [62]. We interpret the weakening of higher-order assemblies at high salt concentrations to imply that CTP-dependent electrostatic interactions are among the main drivers of lateral associations that promote higher-order assemblies. Taken together, our data suggest that the spacer activities of the CTL and the sticker nature of the CTP make the CTT an auto-regulator of FtsZ assembly.

Previous studies of Ahijado-Guzmán et al., have quantified the effects of KCl concentrations on inferred size distributions of *E. coli* FtsZ (Ec-FtsZ) polymers formed in the presence of GTP [36]. Comparisons between our data and those of Ahijado-Guzmán et al., would appear to point to discrepancies, but closer scrutiny reveals mutual consistency. Our DLS data for the size distributions show that Bs-FtsZ assemblies become narrow and shift to lower values in 200 mM KCl versus 50 mM KCl (**Figures 2.10** and **2.11**). This implies a weakening of higher-order assemblies at higher concentrations of KCl. The data of Ahijado-Guzmán et al., [36] obtained for Ec-FtsZ in KCl concentrations between 100 mM and 500 mM suggests the opposite behavior where the narrow distributions shift toward smaller sizes in 100 mM KCl versus 500 mM KCl. This is a reflection of the shorter linear polymers seen at 100 mM KCl in comparison to 500

mM KCl, as Ec-FtsZ forms minimal bundles under these conditions. Therefore, the experiments of Ahijado-Guzmán et al. are probing a different phase boundary when compared to our assay. However, in conditions where bundling is enabled in the presence of a non-hydrolyzable analog, the behavior we observe wherein higher salt decreases assembly size is recapitulated for Ec-FtsZ as well. It is worth emphasizing that the impact of salt concentration, ion valence, and ion hydrophobicity will likely contribute differently to the driving forces for and the assemblies of FtsZs derived from different orthologs. The totality of our data and those of Ahijado-Guzmán et al., [36] point to subtle interplays between the conformations of the FtsZ cores (which are very similar between Ec-FtsZ and Bs-FtsZ) and the CTTs, which are very different, especially across the CTLs, between Ec-FtsZ and Bs-FtsZ [80]. Clearly, a comparative analysis of the salt and osmolyte dependence of assemblies and driving forces for assembly across FtsZ orthologs will be imperative for uncovering differences in synergies among GTPase cores and disordered CTTs.

As far as enzyme activity is concerned, the presence of the CTL in  $\Delta$ CTP weakens GTP binding vis-à-vis  $\Delta$ CTT whereas the presence of the CTP in  $\Delta$ CTL enhances GTP binding. Inferences regarding GTP binding are drawn from values for the apparent  $K_M$  estimated for different constructs (**Figure 2.14**). In contrast to the apparent  $K_M$  values, the inferred  $k_{cat}$  values suggest near  $\Delta$ CTT-like values for  $\Delta$ CTP and a significantly diminished value for  $\Delta$ CTL (**Figure 2.15**). It is possible that the effect on  $k_{cat}$  derives from the formation of shorter polymers when compared to  $\Delta$ CTT. These data suggest that the CTP might serve as an auxiliary binding site for GTP, especially when it is not spatially separated from the core via the CTL. Support for this inference comes from the higher  $k_{cat}$  values we infer for WT when compared to  $\Delta$ CTL. The differing effects of the CTL and CTP lead to a WT enzyme that is roughly half as efficient as the

core domain alone. This is seen in comparisons of inferred values for  $k_{\text{eff}}$  of  $\Delta\text{CTT}$  vs. WT (**Figure 2.16**). Indeed, the CTT appears to act as an auto-inhibitor of enzymatic activity, a feature that is realized via convolution of the different contributions of the CTL and CTP modules. Further, it is known that the hydrolysis state of the bound nucleotide modulates the curvature of the protofilament [81]. Therefore, the presence of the CTL and CTP modules not only regulates activity but could also impact the ability of the filament to curve and likewise impact the ability to form a ring.

In many proteins, IDRs tethered to folded domains are known to contribute as auto-inhibitors of the functions controlled by folded domains [82-87]. Along these lines, Trudeau et al., showed that the activities of auto-inhibitory IDRs are more frequently modified by other modulatory proteins than the folded domains to which they are tethered [88]. Indeed, it is noteworthy that the auto-regulatory and auto-inhibitory functions of the CTT, due to differing contributions from the CTL and CTP, are congruent (from a phenomenological standpoint) with data reported by Li et al., for a series of variants of the glucocorticoid receptor [89]. The impacts of different types of spacers were quantified on the effects of hormone binding on allosteric coupling and transcriptional activation. These results were analyzed using a quantitative framework for ensemble allostery developed by Mothlagh and Hilser [90]. Li et al., substituted linkers that they deemed to be inert and used this to uncover positive coupling between one of the tethered IDRs and the structured domain while also uncovering evidence for autoregulation through negative coupling between the two tethered IDRs [89]. The studies of Li et al., provide a template for probing the effects of different CTLs on the nature of the coupling between the CTP and the core domain. In this context, it is worth emphasizing that the CTL is a hypervariable



module characterized by an assortment of lengths and compositional biases [80]. Accordingly, designing an inert linker becomes non-trivial because there are likely to be cryptic interactions involving the CTL that can only be dissected using a combination of simulations [91] and phenotypic characterization [92]. Specifically, we will need to titrate sequence features within the CTL of different FtsZ orthologs to probe the impact of changes to sequence patterns within CTLs. This approach, inspired by recent efforts of Das et al., [93] and Sherry et al., [94] is being used for the CTL of Bs-FtsZ [92]. Specifically, since the CTL of Bs-FtsZ is a polyampholyte [95], we are investigating the effects of altering the linear patterning of oppositely charged residues away from the patterning observed for the WT CTL. These designs, which maintain the amino acid composition, show clear cellular phenotypes [92]. Our findings in the current study provide a clear basis for interpreting the emerging data.

The active role of the CTT as a modulator of Bs-FtsZ assembly and enzymatic activity also mirrors recent discoveries regarding the distinct roles of disordered C-terminal linkers and conserved C-terminal tips in single-stranded DNA binding proteins referred to as SSBs [96]. The SSBs share coarse-grained architectural similarities with FtsZs. Their oligomerizing N-terminal folded oligonucleotide binding (OB) domains are connected via hypervariable C-terminal linkers (CTLs) to conserved C-terminal tips that encompass acidic residues and a short linear motif that helps coordinate the interactions of SSB interaction proteins (SIPs) [97]. Recent molecular dissection studies on SSBs from *E. coli* (Ec-SSB) [96, 98, 99] show that the CTL in Ec-SSB is essential for cooperative binding of single-stranded DNA. Notably, the CTL of Ec-SSB helps with spatial delocalization of the tip from the OB fold – a feature that is shared with the CTL of Bs-

FtsZ, suggesting a converging theme of disordered linkers serving as spacers that separate stickers such as short linear motifs from folded domains.

As noted above, the CTLs of FtsZs are hypervariable, and this is akin to that of the CTLs of SSBs as well as other oligomerizing / self-assembling systems such as the bacterial RNA chaperone Hfq [84]. In the context of FtsZs, the lengths, compositions, and sequence features of CTLs vary considerably across FtsZ orthologs [50, 80]. In direct contrast, the sequences of CTPs and core domains remain largely conserved [80]. The hypervariability of CTLs in FtsZs might be an example of convergent evolution. In this scenario, a large number of disparate sequences, providing they are intrinsically disordered, might be able to serve as modulators of cooperative assembly of FtsZ while spatially separating the conserved core and CTP, weakening GTP binding, and enhancing the catalytic rate.

On the other hand, the hypervariability of CTLs might also be an example of divergent evolution. Previous studies have shown that CTLs from different bacteria are not always interoperable with one another [59, 60, 62]. Sundararajan and Goley also showed that the  $\Delta$ CTT of *C. crescentus* FtsZ (Cc-FtsZ) is less efficient than the WT. This implies that the mechanism that is operative in Cc-FtsZ might be the opposite of what we have uncovered here for Bs-FtsZ. In Cc-FtsZ, the catalytic inefficiency of the core, possibly due to interactions among core domains that drive alternative inactive assemblies, appears to be alleviated by alternative CTL and CTP modules. Indeed, analysis using CIDER [100] shows that the CTLs of Bs-FtsZ versus Cc-FtsZ are very different from one another in terms of their sequence lengths ( $N = 50$  versus 176), the fraction of charged residues (0.41 versus 0.24), the net charge per residue (0.06 versus  $-0.05$ ), and Kyte-Doolittle hydrophathy (3.12 versus 3.72). These results suggest that the sequences of CTLs and

cores might co-vary in order to achieve a requisite level of enzymatic efficiency and assembly, giving rise to distinct rules / features that underlie the sequence features of CTLs in distinct classes of bacteria.

CTL length is another source of variability among FtsZs, and it is thought to be the key determinant of differences among FtsZs [56, 60, 62]. However, informatics analysis has established an inverse correlation between CTL length and the fraction of charged residues within CTLs [50, 80]. This feature, which is evident even in the comparisons of the CTLs from Bs-FtsZ and Cc-FtsZ, is important because recent studies have established a distinction between sequence length (also referred to as *apparent length* [91]) and *functional length* (also referred to as *effective length* [80] or *thermodynamic length* [91]). While the apparent length is simply the number of amino acids in a sequence, the functional length reflects the fact that intrinsically disordered regions (IDRs) of different sequence lengths can have similar end-to-end distance and shape distributions because amino acid compositions as well as sequence patterning effects, and not just apparent length, will control the conformational properties of IDRs [101].

In an ongoing analysis, we have discovered that thirteen out of the ~20 essential proteins that are involved in *B. subtilis* bacterial cell division encompass IDRs that are longer than 30 residues (**Figure 1.3**). This points to the potential importance of these regions as regulators of cell division. Furthermore, despite having considerably smaller proteomes when compared to eukaryotes, bacteria exert tight control over all processes that are vital to their life cycles. In contrast to eukaryotic systems and viruses, where the functional importance of IDRs is well established [102-105], the prevailing view is that bacterial proteins conform to the classical sequence-structure-function paradigm [106, 107]. This view has emerged mainly from

bioinformatics analysis, which shows that only a small percentage of bacterial proteins encompass long IDRs [108-110]. However, several recent studies have demonstrated that while IDRs make up a small fraction of proteins / protein regions in bacterial proteomes, the synergies between IDRs and folded domains contribute directly to an assortment of functions. Prominent bacterial IDRs include regions within proteins that are involved in regulating cell division [50, 57], single-stranded DNA binding [96, 97], protein and RNA quality control [111-117], bacterial warfare [118-120], biofilm formation [121-123], and chemotaxis [124-126]. In each of these systems, it seems important and essential to undertake molecular dissections of IDR functions and uncover the coevolution between folded domains and IDRs.

## 2.6 References

1. Bi, E.F. and J. Lutkenhaus, *FtsZ ring structure associated with division in Escherichia coli*. Nature, 1991. **354**(6349): p. 161-4.
2. Aarsman, M.E.G., et al., *Maturation of the Escherichia coli divisome occurs in two steps*. Molecular Microbiology, 2005. **55**(6): p. 1631-1645.
3. den Blaauwen, T., L.W. Hamoen, and P.A. Levin, *The divisome at 25: the road ahead*. Current Opinion in Microbiology, 2017. **36**: p. 85-94.
4. Commichau, F.M., N. Pietack, and J. Stülke, *Essential genes in Bacillus subtilis: a re-evaluation after ten years*. Molecular BioSystems, 2013. **9**(6): p. 1068-1075.
5. Wagstaff, J. and J. Löwe, *Prokaryotic cytoskeletons: protein filaments organizing small cells*. Nature Reviews Microbiology, 2018. **16**(4): p. 187-201.
6. Löwe, J. and L.A. Amos, *Crystal structure of the bacterial cell-division protein FtsZ*. Nature, 1998. **391**(6663): p. 203-206.
7. Fu, G., et al., *In Vivo Structure of the E. coli FtsZ-ring Revealed by Photoactivated Localization Microscopy (PALM)*. PLOS ONE, 2010. **5**(9): p. e12680.
8. Bisson-Filho, A.W., et al., *Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division*. Science, 2017. **355**(6326): p. 739-743.
9. Anderson, D.E., F.J. Gueiros-Filho, and H.P. Erickson, *Assembly dynamics of FtsZ rings in Bacillus subtilis and Escherichia coli and effects of FtsZ-regulating proteins*. Journal of Bacteriology, 2004. **186**(17): p. 5775-81.
10. Huang, K.H., J. Durand-Heredia, and A. Janakiraman, *FtsZ ring stability: of bundles, tubules, crosslinks, and curves*. J Bacteriol, 2013. **195**(9): p. 1859-68.
11. Levin, P.A., I.G. Kurtser, and A.D. Grossman, *Identification and characterization of a negative regulator of FtsZ ring formation in Bacillus subtilis*. Proceedings of the National Academy of Sciences, 1999. **96**(17): p. 9642.
12. Monterroso, B., et al., *Bacterial division FtsZ forms liquid condensates with nucleoid-associated Z-ring inhibitor SlmA*. EMBO Reports, 2018. **20**(1): p. e45946.

13. Yu, X.-C. and W. Margolin, *FtsZ ring clusters in min and partition mutants: role of both the Min system and the nucleoid in regulating FtsZ ring localization*. *Molecular Microbiology*, 1999. **32**(2): p. 315-326.
14. Haeusser, D.P., et al., *The Division Inhibitor EzrA Contains a Seven-Residue Patch Required for Maintaining the Dynamic Nature of the Medial FtsZ Ring*. *Journal of Bacteriology*, 2007. **189**(24): p. 9001.
15. Busiek, K.K., et al., *The Early Divisome Protein FtsA Interacts Directly through Its 1c Subdomain with the Cytoplasmic Domain of the Late Divisome Protein FtsN*. *Journal of Bacteriology*, 2012. **194**(8): p. 1989.
16. Rowlett, Veronica W. and W. Margolin, *3D-SIM Super-resolution of FtsZ and Its Membrane Tethers in Escherichia coli Cells*. *Biophysical Journal*, 2014. **107**(8): p. L17-L20.
17. Gueiros-Filho, F.J. and R. Losick, *A widely conserved bacterial cell division protein that promotes assembly of the tubulin-like protein FtsZ*. *Genes & Development*, 2002. **16**(19): p. 2544-2556.
18. Sundararajan, K. and E.D. Goley, *Cytoskeletal Proteins in Caulobacter crescentus: Spatial Orchestrators of Cell Cycle Progression, Development, and Cell Shape*, in *Sub-cellular biochemistry*. 2017. p. 103-137.
19. Anderson, D.E., F.J. Gueiros-Filho, and H.P. Erickson, *Assembly dynamics of FtsZ rings in Bacillus subtilis and Escherichia coli and effects of FtsZ-regulating proteins*. *J Bacteriol*, 2004. **186**(17): p. 5775-81.
20. Chen, Y. and H.P. Erickson, *Rapid in Vitro Assembly Dynamics and Subunit Turnover of FtsZ Demonstrated by Fluorescence Resonance Energy Transfer*. *Journal of Biological Chemistry*, 2005. **280**(23): p. 22549-22554.
21. Lan, G., et al., *Polymerization and Bundling Kinetics of FtsZ Filaments*. *Biophysical Journal*, 2008. **95**(8): p. 4045-4056.
22. Lan, G., et al., *Condensation of FtsZ filaments can drive bacterial cell division*. *Proceedings of the National Academy of Sciences*, 2009. **106**(1): p. 121.
23. Guan, F., et al., *Lateral interactions between protofilaments of the bacterial tubulin homolog FtsZ are essential for cell division*. *eLife*, 2018. **7**: p. e35578.

24. Oliva, M.A., D. Trambaiolo, and J. Lowe, *Structural insights into the conformational variability of FtsZ*. *Journal of Molecular Biology*, 2007. **373**(5): p. 1229-42.
25. González, J.M., et al., *Essential Cell Division Protein FtsZ Assembles into One Monomer-thick Ribbons under Conditions Resembling the Crowded Intracellular Environment*. *Journal of Biological Chemistry*, 2003. **278**(39): p. 37664-37671.
26. Erickson, H.P., *Modeling the physics of FtsZ assembly and force generation*. *Proceedings of the National Academy of Sciences*, 2009. **106**(23): p. 9238.
27. Moore, D.A., et al., *Probing for Binding Regions of the FtsZ Protein Surface through Site-Directed Insertions: Discovery of Fully Functional FtsZ-Fluorescent Proteins*. *Journal of Bacteriology*, 2017. **199**(1): p. e00553-16.
28. Meier, E.L. and E.D. Goley, *Form and function of the bacterial cytokinetic ring*. *Current Opinion in Cell Biology*, 2014. **26**: p. 19-27.
29. Bramhill, D. and C.M. Thompson, *GTP-dependent polymerization of Escherichia coli FtsZ protein to form tubules*. *Proceedings of the National Academy of Sciences*, 1994. **91**(13): p. 5813.
30. Scheffers, D.-J. and A.J.M. Driessen, *The polymerization mechanism of the bacterial cell division protein FtsZ*. *FEBS Letters*, 2001. **506**(1): p. 6-10.
31. Scheffers, D.-J., et al., *GTP Hydrolysis of Cell Division Protein FtsZ: Evidence that the Active Site Is Formed by the Association of Monomers*. *Biochemistry*, 2002. **41**(2): p. 521-529.
32. Stricker, J., et al., *Rapid assembly dynamics of the Escherichia coli FtsZ-ring demonstrated by fluorescence recovery after photobleaching*. *Proceedings of the National Academy of Sciences*, 2002. **99**(5): p. 3171.
33. Yang, X., et al., *GTPase activity-coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis*. *Science*, 2017. **355**(6326): p. 744.
34. Mateos-Gil, P., et al., *Depolymerization dynamics of individual filaments of bacterial cytoskeletal protein FtsZ*. *Proceedings of the National Academy of Sciences*, 2012. **109**(21): p. 8133.
35. Mingorance, J., et al., *Visualization of Single Escherichia coli FtsZ Filament Dynamics with Atomic Force Microscopy*. *Journal of Biological Chemistry*, 2005. **280**(21): p. 20909-20914.

36. Ahijado-Guzmán, R., et al., *Control by Potassium of the Size Distribution of Escherichia coli FtsZ Polymers Is Independent of GTPase Activity*. Journal of Biological Chemistry, 2013. **288**(38): p. 27358-27365.
37. Mukherjee, A. and J. Lutkenhaus, *Analysis of FtsZ Assembly by Light Scattering and Determination of the Role of Divalent Metal Cations*. Journal of Bacteriology, 1999. **181**(3): p. 823.
38. Hill, N.S., et al., *A Moonlighting Enzyme Links Escherichia coli Cell Size with Central Metabolism*. PLOS Genetics, 2013. **9**(7): p. e1003663.
39. Dhaked, H.P.S., et al., *Regulation of Streptococcus pneumoniae FtsZ assembly by divalent cations: paradoxical effects of Ca<sup>2+</sup> on the nucleation and bundling of FtsZ polymers*. The FEBS Journal, 2019. **286**(18): p. 3629-3646.
40. Park, C.K. and N.C. Horton, *Novel insights into filament-forming enzymes*. Nat Rev Mol Cell Biol, 2020. **21**(1): p. 1-2.
41. Greer, S.C., *Reversible polymerizations and aggregations*. Annual Review of Physical Chemistry, 2002. **53**(1): p. 173-200.
42. Dudowicz, J., K.F. Freed, and J.F. Douglas, *Lattice model of living polymerization. I. Basic thermodynamic properties*. The Journal of Chemical Physics, 1999. **111**(15): p. 7116-7130.
43. Mukherjee, A. and J. Lutkenhaus, *Dynamic assembly of FtsZ regulated by GTP hydrolysis*. EMBO Journal, 1998. **17**(2): p. 462-9.
44. Ruiz-Martinez, A., et al., *Efficient Multiscale Models of Polymer Assembly*. Biophysical Journal, 2016. **111**(1): p. 185-196.
45. Chen, Y., et al., *A rapid fluorescence assay for FtsZ assembly indicates cooperative assembly with a dimer nucleus*. Biophysical Journal, 2005. **88**(1): p. 505-14.
46. Huecas, S., et al., *Energetics and Geometry of FtsZ Polymers: Nucleated Self-Assembly of Single Protofilaments*. Biophysical Journal, 2008. **94**(5): p. 1796-1806.
47. Buske, P.J. and P.A. Levin, *Extreme C terminus of bacterial cytoskeletal protein FtsZ plays fundamental role in assembly independent of modulatory proteins*. Journal of Biological Chemistry, 2012. **287**(14): p. 10945-57.



48. Miraldi, E.R., P.J. Thomas, and L. Romberg, *Allosteric models for cooperative polymerization of linear polymers*. Biophysical Journal, 2008. **95**(5): p. 2470-86.
49. Romberg, L., M. Simon, and H.P. Erickson, *Polymerization of FtsZ, a Bacterial Homolog of Tubulin: IS ASSEMBLY COOPERATIVE?* Journal of Biological Chemistry, 2001. **276**(15): p. 11743-11753.
50. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division*. Seminars in Cell and Developmental Biology, 2015. **37**: p. 3-10.
51. Crick, S.L. and R.V. Pappu, *Thermodynamic and Kinetic Models for Aggregation of Intrinsically Disordered Proteins*, in *Peptide Folding, Misfolding, and Nonfolding*. 2010, John Wiley & Sons: Hoboken, NJ. p. 413-440.
52. Kasai, M., S. Asakura, and F. Oosawa, *The cooperative nature of G-F transformation of actin*. Biochimica et Biophysica Acta, 1962. **57**(1): p. 22-31.
53. Meisl, G., et al., *Scaling behaviour and rate-determining steps in filamentous self-assembly*. Chemical Science, 2017. **8**(10): p. 7087-7097.
54. Vitalis, A. and R.V. Pappu, *Assessing the contribution of heterogeneous distributions of oligomers to aggregation mechanisms of polyglutamine peptides*. Biophysical Chemistry, 2011. **159**(1): p. 14-23.
55. Wagstaff, J.M., et al., *A Polymerization-Associated Structural Switch in FtsZ That Enables Treadmilling of Model Filaments*. mBio, 2017. **8**(3): p. e00254-17.
56. Huecas, S., et al., *Self-Organization of FtsZ Polymers in Solution Reveals Spacer Role of the Disordered C-Terminal Tail*. Biophysical Journal, 2017. **113**(8): p. 1831-1844.
57. Buske, P.J. and P.A. Levin, *A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo*. Molecular Microbiology, 2013. **89**(2): p. 249-63.
58. Sundararajan, K. and E.D. Goley, *The intrinsically disordered C-terminal linker of FtsZ regulates protofilament dynamics and superstructure in vitro*. The Journal of Biological Chemistry, 2017. **292**(50): p. 20509-20527.
59. Sundararajan, K., et al., *The bacterial tubulin FtsZ requires its intrinsically disordered linker to direct robust cell wall construction*. Nature Communications, 2015. **6**: p. 7281.

60. Sundararajan, K., et al., *Species- and C-terminal linker-dependent variations in the dynamic behavior of FtsZ on membranes in vitro*. *Molecular Microbiology*, 2018. **110**(1): p. 47-63.
61. Barrows, J.M., et al., *FtsA Regulates Z-Ring Morphology and Cell Wall Metabolism in an FtsZ C-Terminal Linker-Dependent Manner in Caulobacter crescentus*. *J Bacteriol*, 2020. **202**(7): p. e00693-19.
62. Gardner, K.A.J.A., D.A. Moore, and H.P. Erickson, *The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide*. *Molecular Microbiology*, 2013. **89**(2): p. 264-275.
63. Sogues, A., et al., *Essential dynamic interdependence of FtsZ and SepF for Z-ring and septum formation in Corynebacterium glutamicum*. *Nature Communications*, 2020. **11**(1): p. 1641.
64. Hale, C.A., A.C. Rhee, and P.A.J. de Boer, *ZipA-Induced Bundling of FtsZ Polymers Mediated by an Interaction between C-Terminal Domains*. *Journal of Bacteriology*, 2000. **182**(18): p. 5153.
65. Ma, X. and W. Margolin, *Genetic and Functional Analyses of the Conserved C-Terminal Core Domain of Escherichia coli FtsZ*. *Journal of Bacteriology*, 1999. **181**(24): p. 7531.
66. Ingberman, E. and J. Nunnari, *A continuous, regenerative coupled GTPase assay for dynamin-related proteins*. *Methods in Enzymology*, 2005. **404**: p. 611-9.
67. Swinehart, D.F., *The beer-lambert law*. *Journal of chemical education*, 1962. **39**(7): p. 333.
68. Hill, N.S., et al., *A moonlighting enzyme links Escherichia coli cell size with central metabolism*. *PLoS Genet*, 2013. **9**(7): p. e1003663.
69. Hou, S., et al., *Characterization of Caulobacter crescentus FtsZ Protein Using Dynamic Light Scattering*. *Journal of Biological Chemistry*, 2012. **287**(28): p. 23878-23886.
70. Stetefeld, J., S.A. McKenna, and T.R. Patel, *Dynamic light scattering: a practical guide and applications in biomedical sciences*. *Biophysical Reviews*, 2016. **8**(4): p. 409-427.
71. Kozlov, A.G., M.K. Shinn, and T.M. Lohman, *Regulation of Nearest-Neighbor Cooperative Binding of E. coli SSB Protein to DNA*. *Biophys J*, 2019. **117**(11): p. 2120-2140.

72. Arora, J.S., *Chapter 20 - Additional Topics on Optimum Design*, in *Introduction to Optimum Design (Third Edition)*, J.S. Arora, Editor. 2012, Academic Press: Boston. p. 731-784.
73. Choi, J.-M., F. Dar, and R.V. Pappu, *LASSI: A lattice model for simulating phase transitions of multivalent proteins*. PLOS Computational Biology, 2019. **15**(10): p. e1007028.
74. Choi, J.-M., A.S. Holehouse, and R.V. Pappu, *Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions*. Annual Review of Biophysics, 2020. **49**(1): p. <https://doi.org/10.1146/annurev-biophys-121219-081629>.
75. Martin, E.W., et al., *Valence and patterning of aromatic residues determine the phase behavior of prion-like domains*. Science, 2020. **367**(6478): p. 694-699.
76. Yeomans, J.M., *Statistical Mechanics of Phase Transitions*. 1992, Oxford, UK: Oxford University Press.
77. Cárdenas, M.L., *Michaelis and Menten and the long road to the discovery of cooperativity*. FEBS Letters, 2013. **587**(17): p. 2767-2771.
78. Arjes, H.A., et al., *Mutations in the bacterial cell division protein FtsZ highlight the role of GTP binding and longitudinal subunit interactions in assembly and function*. BMC Microbiology, 2015. **15**(1): p. 209.
79. Weiss, J.N., *The Hill equation revisited: uses and misuses*. The FASEB Journal, 1997. **11**(11): p. 835-841.
80. Cohan, M.C., K.M. Ruff, and R.V. Pappu, *Information theoretic measures for quantifying sequence–ensemble relationships of intrinsically disordered proteins*. Protein Engineering, Design and Selection, 2019. **32**(4): p. 191-202.
81. Lu, C., M. Reedy, and H.P. Erickson, *Straight and curved conformations of FtsZ are regulated by GTP hydrolysis*. J Bacteriol, 2000. **182**(1): p. 164-70.
82. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins*. Chem Rev, 2014. **114**(13): p. 6589-631.
83. Kim, Y., et al., *Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLa mismatch repair complex*. Nucleic Acids Research, 2018. **47**(4): p. 1823-1835.

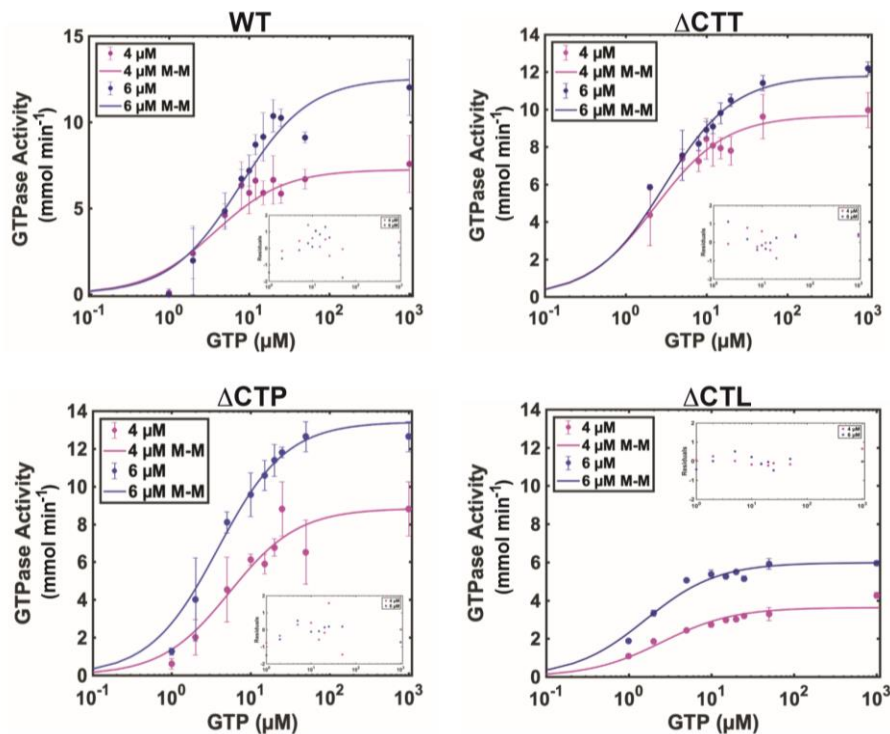
84. Santiago-Frangos, A., et al., *Acidic C-terminal domains autoregulate the RNA chaperone Hfq*. eLife, 2017. **6**: p. e27049.
85. He, F., et al., *Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding*. Proceedings of the National Academy of Sciences, 2019. **116**(18): p. 8859.
86. Bista, M., M. Petrovich, and A.R. Fersht, *MDMX contains an autoinhibitory sequence element*. Proceedings of the National Academy of Sciences, 2013. **110**(44): p. 17814.
87. Weems, A. and M. McMurray, *The step-wise pathway of septin hetero-octamer assembly in budding yeast*. eLife, 2017. **6**: p. e23689.
88. Trudeau, T., et al., *Structure and intrinsic disorder in protein autoinhibition*. Structure, 2013. **21**(3): p. 332-41.
89. Li, J., et al., *Genetically tunable frustration controls allostery in an intrinsically disordered transcription factor*. eLife, 2017. **6**: p. e30688.
90. Motlagh, H.N. and V.J. Hilser, *Agonism/antagonism switching in allosteric ensembles*. Proceedings of the National Academy of Sciences, 2012. **109**(11): p. 4134.
91. Mittal, A., et al., *Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins*. Journal of Molecular Biology, 2018. **430**(16): p. 2403-2421.
92. Cohan, M.C., et al., *Evolved sequence features within the intrinsically disordered tail influence FtsZ assembly and bacterial cell division*. bioRxiv, 2018: p. 301622.
93. Das, R.K., et al., *Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling*. Proc Natl Acad Sci U S A, 2016. **113**(20): p. 5616-21.
94. Sherry, K.P., et al., *Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor*. Proc Natl Acad Sci U S A, 2017. **114**(44): p. E9243-e9252.
95. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(33): p. 13392-13397.

96. Kozlov, A.G., et al., *Intrinsically disordered C-terminal tails of E. coli single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA*. J Mol Biol, 2015. **427**(4): p. 763-74.
97. Shereda, R.D., et al., *SSB as an Organizer/Mobilizer of Genome Maintenance Complexes*. Critical Reviews In Biochemistry And Molecular Biology, 2008. **43**(5): p. 289-318.
98. Shinn, M.K., et al., *Are the intrinsically disordered linkers involved in SSB binding to accessory proteins?* Nucleic Acids Research, 2019. **47**(16): p. 8581-8594.
99. Kozlov, A.G., et al., *Glutamate promotes SSB protein–protein Interactions via intrinsically disordered regions*. Journal of Molecular Biology, 2017. **429**(18): p. 2790-2801.
100. Holehouse, A.S., et al., *CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins*. Biophysical Journal, 2017. **112**(1): p. 16-21.
101. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Current Opinion in Structural Biology, 2015. **32**: p. 102-12.
102. Berlow, R.B., H.J. Dyson, and P.E. Wright, *Hypersensitive termination of the hypoxic response by a disordered protein switch*. Nature, 2017. **543**(7645): p. 447-451.
103. Wright, P.E. and H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation*. Nat Rev Mol Cell Biol, 2015. **16**(1): p. 18-29.
104. Jensen, M.R., et al., *Intrinsic disorder in measles virus nucleocapsids*. Proc Natl Acad Sci U S A, 2011. **108**(24): p. 9839-44.
105. Mollica, L., et al., *Binding Mechanisms of Intrinsically Disordered Proteins: Theory, Simulation, and Experiment*. Front Mol Biosci, 2016. **3**: p. 52.
106. Dunker, A.K., et al., *Intrinsically disordered proteins and multicellular organisms*. Semin Cell Dev Biol, 2015. **37**: p. 44-55.
107. Pavlovic-Lazetic, G.M., et al., *Bioinformatics analysis of disordered proteins in prokaryotes*. BMC Bioinformatics, 2011. **12**: p. 66.
108. Yruela, I., et al., *Evidence for a Strong Correlation Between Transcription Factor Protein Disorder and Organismic Complexity*. Genome Biology and Evolution, 2017. **9**(5): p. 1248-1265.

109. Chen, J.W., et al., *Conservation of intrinsic disorder in protein domains and families: II. functions of conserved disorder*. Journal of Proteome Research, 2006. **5**(4): p. 888-98.
110. Brown, C.J., et al., *Evolution and disorder*. Current Opinion in Structural Biology, 2011. **21**(3): p. 441-6.
111. Ait-Bara, S., A.J. Carpousis, and Y. Quentin, *RNase E in the gamma-Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains*. Molecular Genetics and Genomics, 2015. **290**(3): p. 847-62.
112. Ait-Bara, S. and A.J. Carpousis, *RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs*. Molecular Microbiology, 2015. **97**(6): p. 1021-135.
113. Jakob, U., R. Kriwacki, and V.N. Uversky, *Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function*. Chemical Reviews, 2014. **114**(13): p. 6779-805.
114. Reichmann, D., et al., *Order out of disorder: working cycle of an intrinsically unfolded chaperone*. Cell, 2012. **148**(5): p. 947-57.
115. Updegrave, T.B., A. Zhang, and G. Storz, *Hfq: the flexible RNA matchmaker*. Current Opinion in Microbiology, 2016. **30**: p. 133-138.
116. Quan, S., et al., *Super Spy variants implicate flexibility in chaperone action*. Elife, 2014. **3**: p. e01584.
117. Foit, L., et al., *Chaperone activation by unfolding*. Proceedings of the National Academy of Sciences of the United States of America 2013. **110**(14): p. E1254-62.
118. Hecht, O., et al., *A common interaction for the entry of colicin N and filamentous phage into Escherichia coli*. Journal of Molecular Biology, 2009. **388**(4): p. 880-93.
119. Papadakos, G., et al., *Consequences of inducing intrinsic disorder in a high-affinity protein-protein interaction*. Journal of the American Chemical Society, 2015. **137**(16): p. 5252-5.
120. Bonsor, D.A., N.A. Meenan, and C. Kleanthous, *Colicins exploit native disorder to gain cell entry: a hitchhiker's guide to translocation*. Biochemical Society Transactions, 2008. **36**(Pt 6): p. 1409-13.

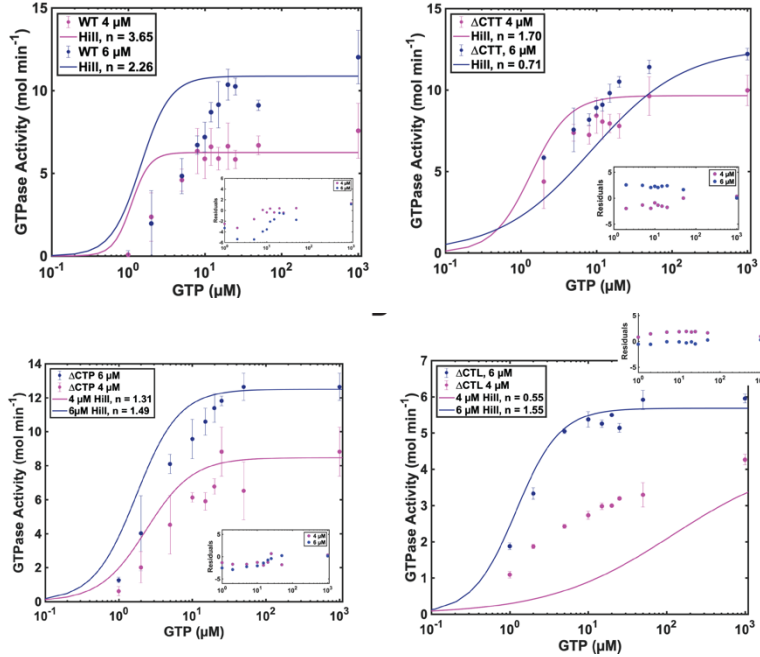
121. Gu, S., et al., *The role of intrinsic disorder and dynamics in the assembly and function of the type II secretion system*. *Biochimica et Biophysica Acta*, 2017. **1865**(10): p. 1255-1266.
122. Gruszka, D.T., et al., *Disorder drives cooperative folding in a multidomain protein*. *Proceedings of the National Academy of Sciences of the United States of America*, 2016. **113**(42): p. 11841-11846.
123. Whelan, F. and J.R. Potts, *Two repetitive, biofilm-forming proteins from Staphylococci: from disorder to extension*. *Biochemical Society Transactions*, 2015. **43**(5): p. 861-6.
124. Daughdrill, G.W., et al., *The C-terminal half of the anti-sigma factor, FlgM, becomes structured when bound to its target, sigma 28*. *Nature Structural Biology*, 1997. **4**(4): p. 285-91.
125. Dedmon, M.M., et al., *FlgM gains structure in living cells*. *Proceedings of the National Academy of Sciences of the United States of America*, 2002. **99**(20): p. 12681-4.
126. Ding, L., et al., *Functional characterization of FlgM in the regulation of flagellar synthesis and motility in Yersinia pseudotuberculosis*. *Microbiology*, 2009. **155**(Pt 6): p. 1890-900.
127. Weiss, J.N., *The Hill equation revisited: uses and misuses*. *FASEB J*, 1997. **11**(11): p. 835-41.

## 2.7 Appendix



**Figure 2A.1: Parameters that quantify enzyme activity were extracted using fits to a Michaelis-Menten model.** GTP is the substrate of the GTPase FtsZ. Each subplot shows the GTPase activity as a function of GTP concentration for the specified variant of the enzyme, Bs-FtsZ. Data at 4 and 6 μM enzyme concentration are fit to a Michaelis-Menten model as described in Materials and Methods. The inset shows the residuals between the model and the data at each concentration. To estimate the goodness of fit, we use the normalized sum of the square of the residuals ( $s^2$ ) where the normalization accounts for the number of parameters that are used in the fitting process. (Top Left) Data for WT Bs-FtsZ at 4 and 6 μM. The  $s^2$  values are 0.72 and 1.09, respectively. (Top Right) Data for ΔCTT. The  $s^2$  values are 0.31 and 0.27 for 4 and 6 μM, respectively. (Bottom Left) Data for ΔCTP at 4 μM ( $s^2 = 0.84$ ) and 6 μM ( $s^2 = 0.49$ ). (Bottom Right) Data for ΔCTL at 4 μM ( $s^2 = 0.09$ ) and 6 μM ( $s^2 = 0.11$ ). The fits shown here are to be compared to those shown in **Figure 2A.2** where the data were fit using a Hill equation. In all cases, the fits based on the Michaelis-Menten model are superior to those obtained using the Hill model.





**Figure 2A.2: Assessment of the fits of GTPase Activity versus GTP concentration obtained using a Hill model [127].** Raw data for subplots 2A.2 are the same as shown in **Figure 2A.1**. Here, the data is fit using a Hill enzyme model, which allows for negative or positive cooperativity in activity as described in the Materials and Methods. The inset shows the residuals between the model and the data at each concentration. To estimate the goodness of fit, we use the normalized sum of the square of the residuals ( $s^2$ ). (A) Data for WT Bs-FtsZ at 4 (pink) and 6 (blue)  $\mu\text{M}$ . The  $s^2$ -values are 2.78 and 13.65, respectively. (B) Data for  $\Delta\text{CTT}$ . The  $S^2$  values are 2.98 and 6.59 for 4 and 6  $\mu\text{M}$ , respectively. (C) Data for  $\Delta\text{CTP}$  at 4  $\mu\text{M}$  ( $s^2= 3.13$ ) and 6  $\mu\text{M}$  ( $s^2= 4.53$ ). (D) Data for  $\Delta\text{CTL}$  at 4  $\mu\text{M}$  ( $s^2= 3.89$ ) and 6  $\mu\text{M}$  ( $s^2= 0.19$ ). For every enzyme and concentration tested, fits to the Hill equation resulted in a larger  $s^2$ -value.

## Chapter 3

# Information theoretic measures for quantifying sequence-ensemble relationships of intrinsically disordered proteins

### 3.1 Preamble

This chapter is based on the published work, Cohan, M.C., Ruff, K.M., Pappu, R.V. (2019). Information theoretic measures for quantifying sequence–ensemble relationships of intrinsically disordered proteins, *Protein Engineering, Design and Selection*, **32**: 191–202. All work and analyses were conducted by MCC. KMR provided critical insights that helped to shape the narrative. RVP and MCC compiled and wrote the manuscript. Rahul Das allowed for the use of his simulation results on the Notch Intracellular Domain (NICD) system.

As discussed in the preceding chapters, intrinsically disordered proteins (IDPs) contribute to a multitude of functions. It follows that *de novo* design of IDPs should open the door to modulating functions and phenotypes controlled by these systems. Recent design efforts have focused on compositional biases and specific sequence patterns as the design features. Analysis of the impact of these designs on sequence-function relationships indicates that individual sequence/compositional parameters are insufficient for describing sequence-function relationships

in IDPs. To remedy this problem, we have developed information theoretic measures for the sequence–ensemble relationships (SERs) of IDPs. These measures rely on the prior availability of statistically robust conformational ensembles derived from all-atom simulations. In this chapter, we show that the measures we have developed are useful for comparing sequence–ensemble relationships even when the sequence is poorly conserved, as is the case for the hypervariable C-terminal tail of FtsZ. Based on our results, we propose that *de novo* designs of IDPs, guided by knowledge of their SERs, should provide improved insights into their sequence–ensemble–function relationships. Therefore, when designing a functional C-terminal tail of FtsZ, we suggest that SERs must be considered.

## 3.2 Introduction

Advances in *de novo* design [1] have given rise to proteins with new folds [2-4], novel functions [5], controllable dynamics [6, 7], and unnaturally high stabilities [8, 9]. Historically, protein design was cast as an inverse protein-folding problem [10-13]. One prescribes a structure or a fold and uses design principles to uncover the family of sequences that are compatible with the fold. In this approach, one seeks the set of sequences for which the free energy of folding, defined as the difference between standard state free energies of the folded and unfolded states, is negative. In addition to being able to design sequences that are compatible with a prescribed fold, modern tools in protein design are yielding novel folds with bespoke functions leading to a revolution in synthetic biology [14, 15].

Advances in protein design may be traced to improvements in our understanding of sequence-structure-function relationships of proteins [16-19]. These improvements are manifest in being able to codify relationships between sequence and structure. Improvements in *de novo* structure prediction, which essentially represent the ability to relate sequence to structure, have also gone a long way toward enabling rapid advances in protein design [20]. When fluctuations around well-ordered structures have to be incorporated into the designs, one can enhance computational design strategies using novel Monte Carlo sampling [21, 22].

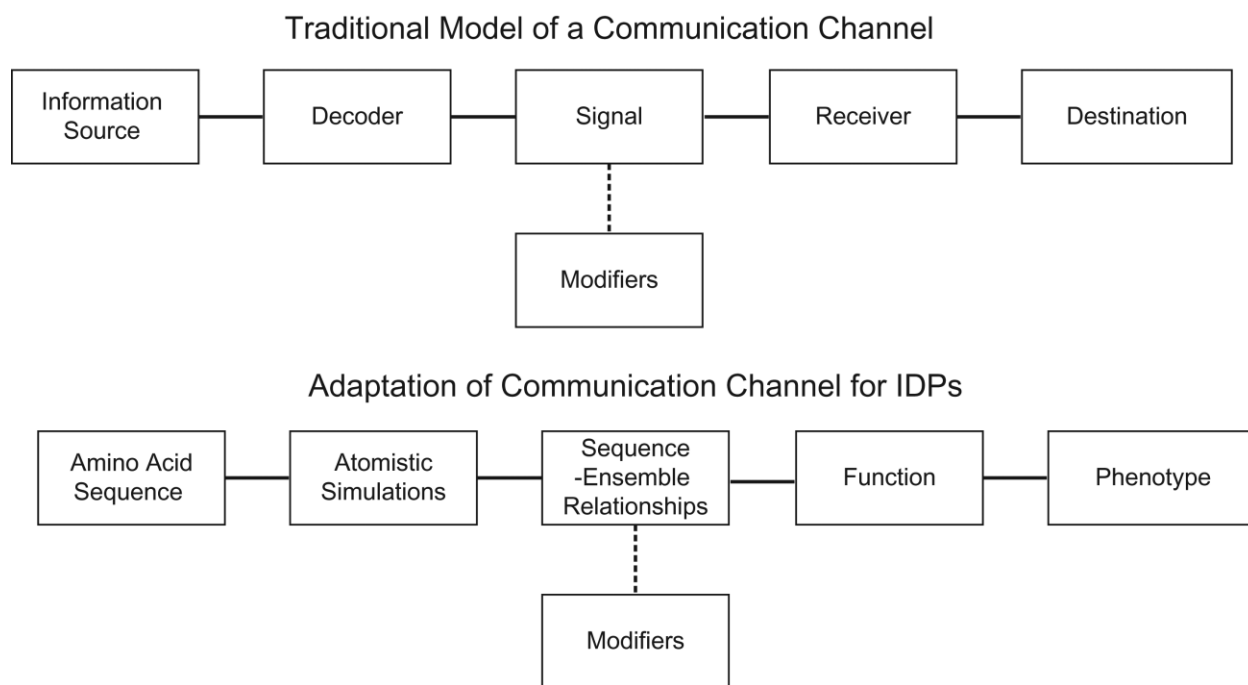
The preceding narrative summarizes the state-of-the-art for the design of proteins that spontaneously adopt well-defined folds under typical folding conditions. We refer to these proteins as autonomously folded proteins (AFPs). However, over a third of eukaryotic proteomes feature proteins or regions that are defined by significant conformational heterogeneity and are referred to

as intrinsically disordered proteins / regions (IDPs / IDRs) [23-27]. It is becoming increasingly clear that IDPs / IDRs have important functional roles, especially in the context of controlling the outcomes of decision-making and influencing circuits in cells. Therefore, it stands to reason that IDPs / IDRs provide a prime target for protein design. However, it is often the case that sequences of IDPs / IDRs are poorly conserved, even when they belong to the same functional family across orthologs [28-30]. This stands in direct contrast to AFPs, where proteins that contribute to similar functions often have similar sequences [31] and hence similar structures, although exceptions to this rule do exist [32, 33]. Despite these exceptions, multiple sequence alignments (MSAs) are highly informative for inferring sequence-structure relationships of AFPs, but they are not very useful in classifying IDPs / IDRs unless considerable prior knowledge is brought to bear on designing substitution matrices that are used in sequence alignments [28].

It has been observed, however, that IDPs / IDRs retain similarities in terms of amino acid compositions, even when sequence similarities are minimal [29, 30]. This has led to the development and deployment of various tools that enable the computation of compositional parameters for IDPs / IDRs [34]. These include parameters such as average hydrophathy, the fraction of charged residues (FCR), the net charge per residue (NCPR) [35], the patterning of oppositely charged residues along the linear sequence [36-38], and the patterning of proline and charged residues vis-à-vis other residues [39].

Sequences of IDPs / IDRs can be compared to one another in terms of coarse-grained compositional parameters or by comparing compositional profiles across adjacent or non-adjacent sequence windows [34]. Recent efforts have focused on *de novo* redesigns of specific IDRs by altering compositional biases and patterning parameters to influence overall dimensions, specific

molecular functions, phase behavior, and cellular phenotypes [39-48]. Analysis of changes to specific sequence parameters on sequence-function / sequence-phenotype relationships has revealed the fact that no single compositional parameter can serve as an adequate design feature that connects IDP / IDR sequences to their functions [41, 43, 44]. What we require are quantitative measures that account for the totality of ensemble features encoded by IDP sequences [49].



**Figure 3.1: Adaptation of a communication channel to describe protein design, focusing on IDP design.** In a traditional communication channel, the information source produces the message, which is then decoded and converted into a signal for transmission over the channel by the decoder. The actual signal that is transmitted is a convolution of the intrinsic signal and extrinsic modifications introduced in the form of encryption, noise, or ancillary signals. The transmission is processed by a receiver and relayed to its intended destination. We propose that the model of a communication channel can be adapted to describe proteins, such that the amino acid sequence (information source) encodes protein function (receiver) and resulting cellular phenotype (destination). IDPs exhibit conformational heterogeneity. Therefore, analysis of all-atom simulations that considers the entire ensemble of conformations needs to be used to decode the information contained in the IDP sequence.

An analogy to the design of communication channels (**Figure 3.1**) helps make the point about the importance of sequence-ensemble relationships for IDPs / IDRs [50]. In a traditional

communication channel [51], the information source produces the message, which is then decoded and converted into a signal for transmission over the channel. The actual transmission is a convolution of the intrinsic signal and extrinsic modifications introduced in the form of encryption, noise, or ancillary signals. The transmission is processed by a receiver and relayed to its intended destination. In our conceptualization of the analogy to communication channels, the information source is the protein sequence (**Figure 3.1**). The key decoding unit that facilitates protein design and formalizes analogies between communication channels and sequence-structure-function channels are the sequence-encoded and decodable sequence-structure or sequence-ensemble relationships. For AFPs, these can be gleaned by combining MSAs [31] and structural comparisons [52-54]. The situation is quite different for IDPs / IDRs because no singular structure provides a suitable representation or abstraction for the types of conformations that these sequences can adopt. Efforts over the past decade have uncovered a series of rules and heuristics that connect the sequences of IDPs / IDRs to conformational ensembles that they adopt [35, 50, 55-61]. We propose that SERs serve as quantitative proxies for sequence-structure relationships and pave the way to understanding and modulating how the information encoded in IDP / IDR sequences contributes to protein function and cellular phenotypes.

How are SERs quantified? Recent advances have enabled all-atom simulations with sufficient throughput for a variety of IDPs / IDRs [35, 62-87]. The use of implicit solvation models combined with advances in Monte Carlo sampling enables the efficiency required for being able to simulate a large number of sequences derived from similar functional families [35, 62, 81, 88-91]. Comparisons to experiments suggest that the conformational ensembles that result from the use of efficient simulations based on implicit solvation models have the accuracy that should

enable the construction of quantitative SERs [61, 73-75, 79, 92]. Further, these simulations can be combined with genetic algorithms to design IDPs / IDRs that fit specified criteria for SERs [93]. What we require is a formal set of measures to quantify SERs for IDPs / IDRs. This, we propose, will allow us to uncover the design principles that connect information encoded in IDP / IDR sequences to their functions and the cellular phenotypes they influence.

Given our analogy between protein design and the design of communication channels, we use methods from information theory to quantify SERs. We show that these measures enable large-scale comparisons of SERs across designed and naturally occurring sequence families. These measures reveal the inadequacies of using compositional parameters as the only parameters to be modulated for tuning sequence-encoded information in IDPs / IDRs. We find that sequences with similar compositional biases can have different SERs. We also find that highly dissimilar sequences can have similar SERs, quite possibly leading to disparate sequences being part of similar functional families across orthologs. These findings appear to highlight one of the reasons for the considerable sequence dissimilarities that have been observed for IDPs / IDRs that belong to similar functional families across orthologs.



## 3.3 Material and Methods

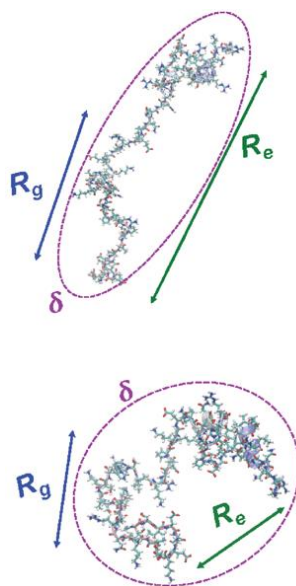
### 3.3.1 Simulations of sequences of FtsZ-CTTs

All-atom Monte Carlo simulations were performed using the ABSINTH implicit solvent model and forcefield paradigm as made available in version 2.0 CAMPARI simulation package (<http://campari.sourceforge.net>) [81, 89, 94]. Simulations were based on the abs\_3.2\_opls.prm parameter set in conjunction with optimized parameters for neutralizing and excess Na<sup>+</sup> and Cl<sup>-</sup> ions [95]. Simulations were performed using a spherical droplet with a diameter of 285 Å with explicit ions to mimic a concentration of 10 mM NaCl. Temperature replica-exchange Monte Carlo (T-REMC) [96] was utilized to enhance conformational sampling. The temperature schedule ranged from 280 K to 400 K. Ensembles corresponding to a temperature of 310 K were used in the analysis reported in this work. Three independent sets of T-REMC simulations were performed for each CTT sequence. In all, the ensembles for each CTT sequence were extracted from a collection of T-REMC simulations, where each simulation deploys  $4.6 \times 10^7$  Monte Carlo steps. In each simulation, the first  $10^6$  steps were discarded as equilibration. Simulation results were analyzed using the CAMPARItraj routines that are available at <http://pappulab.wustl.edu/CTraj.html>. The results for the RAM regions of the WT and designed variants of NICD were those of Sherry et al. [97].

## 3.4 Results

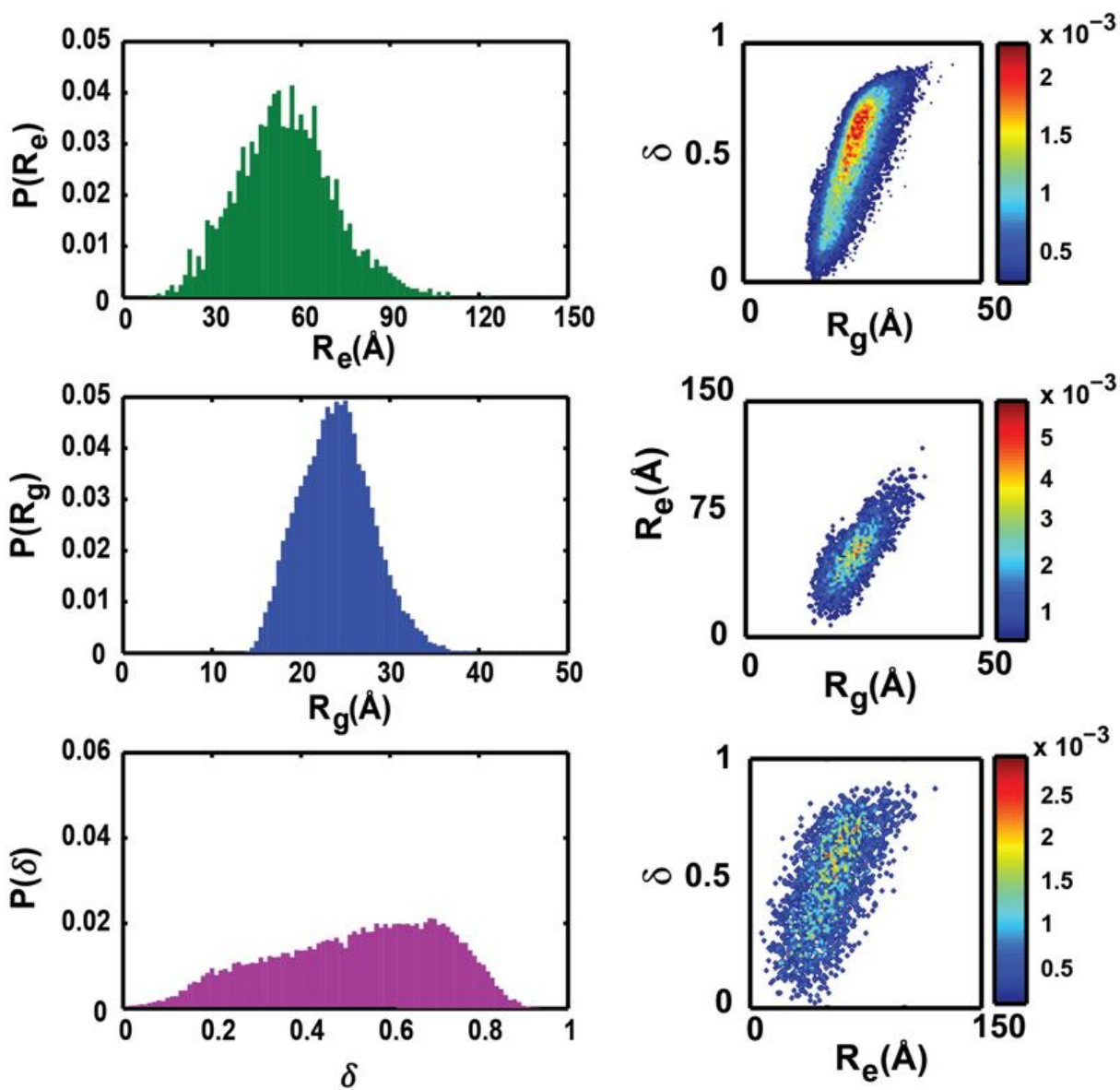
### 3.4.1 Global parameters that define conformational ensembles

All-atom simulations of disordered systems may be performed in one of two modes: IDRs tethered to ordered domains can be simulated in their full protein contexts; alternatively, IDRs can be treated as autonomous units, and the sequence-intrinsic conformational preferences of IDRs are then extracted from simulations. The latter mode is the more pervasive approach. However, recent developments in enhanced sampling [98] have enabled the simulations of IDRs tethered as disordered tails to ordered domains or as linkers between ordered domains. A typical simulation will yield an ensemble of conformations that can be analyzed using a series of global and local parameters. We focus here on global parameters that are central to polymeric descriptions of heterogeneous systems, namely, radii of gyration ( $R_g$ ), asphericity ( $\delta$ ), and end-to-end distance ( $R_e$ ) – see **Figure 3.2**.



**Figure 3.2: Illustration of the relevant global conformational parameters.** Two distinct conformations are shown from the disordered C-terminal tail of *B. subtilis* FtsZ. Each conformation within the ensemble has a set of properties. Here, we focus on three distinct properties:  $R_g$ ,  $R_e$ , and  $\delta$  (see text).

Each of the three parameters can be gleaned from small-angle X-ray scattering (SAXS) measurements [99], although the estimate's reliability will depend on the parameter itself.  $R_g$  quantifies the average distance, for a given conformation, of all of the atoms with respect to its centroid; accordingly, it is a measure of the overall size, primarily in terms of the density of the polymer in question. Analysis of a SAXS profile in the Guinier regime enables direct estimates of  $\langle R_g \rangle$  averaged over the thermodynamically relevant ensemble of the system. For a given conformation,  $\delta$  quantifies the shape of the polymer, and it is calculated using the eigenvalues of the gyration tensor [99]. Ensemble-averaged values of  $\delta$ , denoted as  $\langle \delta \rangle$ , can be extracted by inferring parameters for the average inertial ellipsoid that describes all of the SAXS data for a given system [75]. Values of  $\langle \delta \rangle$  can also be extracted from measurements of rotational diffusion, although care must be taken when connecting quantities derived from hydrodynamic measurements to parameters that are derived from the inertial ellipsoid. Similar concerns apply to conversions of hydrodynamic radii ( $\langle R_h \rangle$ ) from translational diffusion measurements to estimates of  $\langle R_g \rangle$  [80]. Finally, ensemble-averaged values of  $R_e$ , which refer to the ensemble-averaged distance between the ends of a chain, can be inferred from SAXS measurements but are more readily obtained from single-molecule Förster resonance transfer (smFRET) measurements. For homopolymers in the ideal solvent limit,  $\langle R_g \rangle$  and  $\langle R_e \rangle$  differ from one another by a multiplicative factor. However, away from the ideal solvent limit and specifically for heteropolymers such as IDPs / IDRs, the conformation-specific and ensemble averaged values of  $R_g$  and  $R_e$  can become decoupled from one another [100, 101].

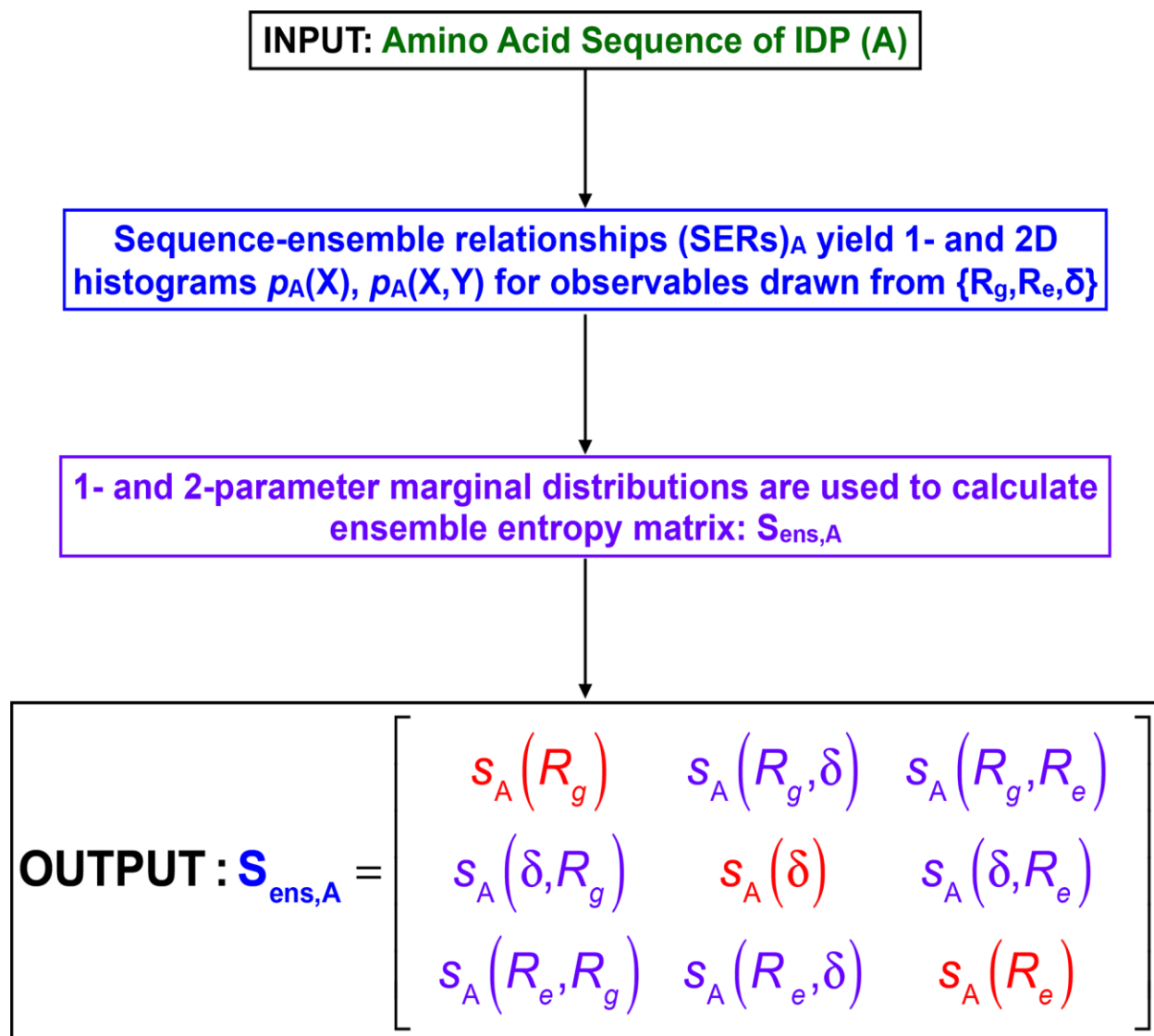


**Figure 3.3: Illustration of conformational features of IDPs/IDRs extracted from all-atom simulations.** (Left) One-parameter marginal distributions  $p(R_g)$ ,  $p(R_e)$ , and  $p(\delta)$  for the conformational ensemble of the disordered C-terminal tail (CTT) of *B. subtilis* FtsZ. The histograms each have 100 bins of equivalent width. (Right) Two-dimensional histograms of the resulting two-parameter marginal distributions of  $p(R_g, \delta)$ ,  $p(R_g, R_e)$ , and  $p(\delta, R_e)$  shown with distribution-specific color bars.

For a given sequence, the values of  $R_g$ ,  $\delta$ , and  $R_e$  are conformation-specific, and for a heterogeneous ensemble of conformations, converged, statistically robust simulations will yield distributions of values for these parameters. Accordingly, to first order, a complete description of conformational ensembles in terms of global features can be cast as a three-parameter probability density function *viz.*,  $p(R_g, \delta, R_e)$ . Features of this three-parameter distribution can be gleaned from three different marginal one-parameter distributions,  $p(R_g)$ ,  $p(\delta)$ , and  $p(R_e)$  – see left panel of **Figure 3.3** – or three different marginal two-parameter distributions,  $p(R_g, \delta)$ ,  $p(R_g, R_e)$ , and  $p(\delta, R_e)$  – see right panel of **Figure 3.3**. For a given sequence, the one- and two-parameter marginal distributions provide a visual and quantitative description of conformational heterogeneity. We use these distributions to compute quantitative SERs as described next.

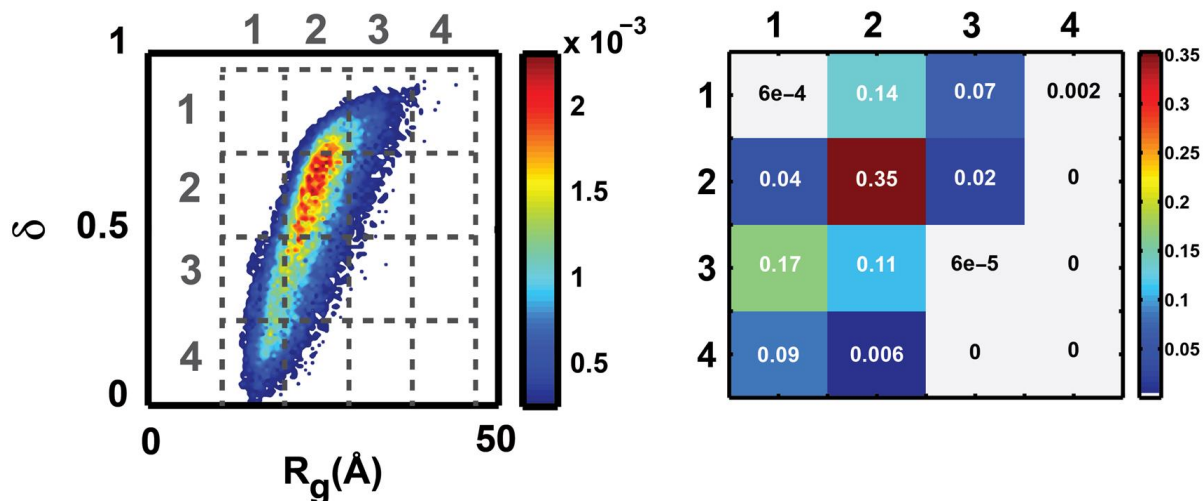
### 3.4.2 The ensemble entropy matrix approach

**Figure 3.4** summarizes the overall approach that we follow to arriving at a concise, quantitative, information theoretic description of the conformational ensemble for a given IDP sequence that is based on analysis of simulation results for one- and two-parameter marginal distributions. Consider the case of a two-parameter distribution  $p(R_g, \delta)$  shown in the left panel of **Figure 3.5**.



**Figure 3.4: Summary of the workflow used to generate the ensemble entropy matrix.** The workflow is depicted for a sequence of interest, designated as A.

The two-parameter space is tiled into an  $n \times n$  grid, and the integral of  $p(R_g, \delta)$  is computed for each of the grids, leading to a grid of probabilities as shown in the righthand side of **Figure 3.5** where  $n = 4$ .



**Figure 3.5: Example of a two-dimensional probability distribution of conformational properties used to quantify the information theoretic entropy.** (Left) The two-parameter space is tiled into  $n \times n$  grids. (Right) Grid of probabilities derived from the 2-parameter distribution shown left.

In general, if  $(X,Y)$  are the parameters of interest, shown for  $(X,Y) \equiv (R_g, \delta)$ , then the information theoretic entropy  $s(X,Y)$  using the grid of probabilities is computed as:

$$s(X,Y) = - \prod_{i=1}^n \prod_{j=1}^n p(X_i, Y_j) \ln p(X_i, Y_j) \quad \text{Equation 3.1}$$

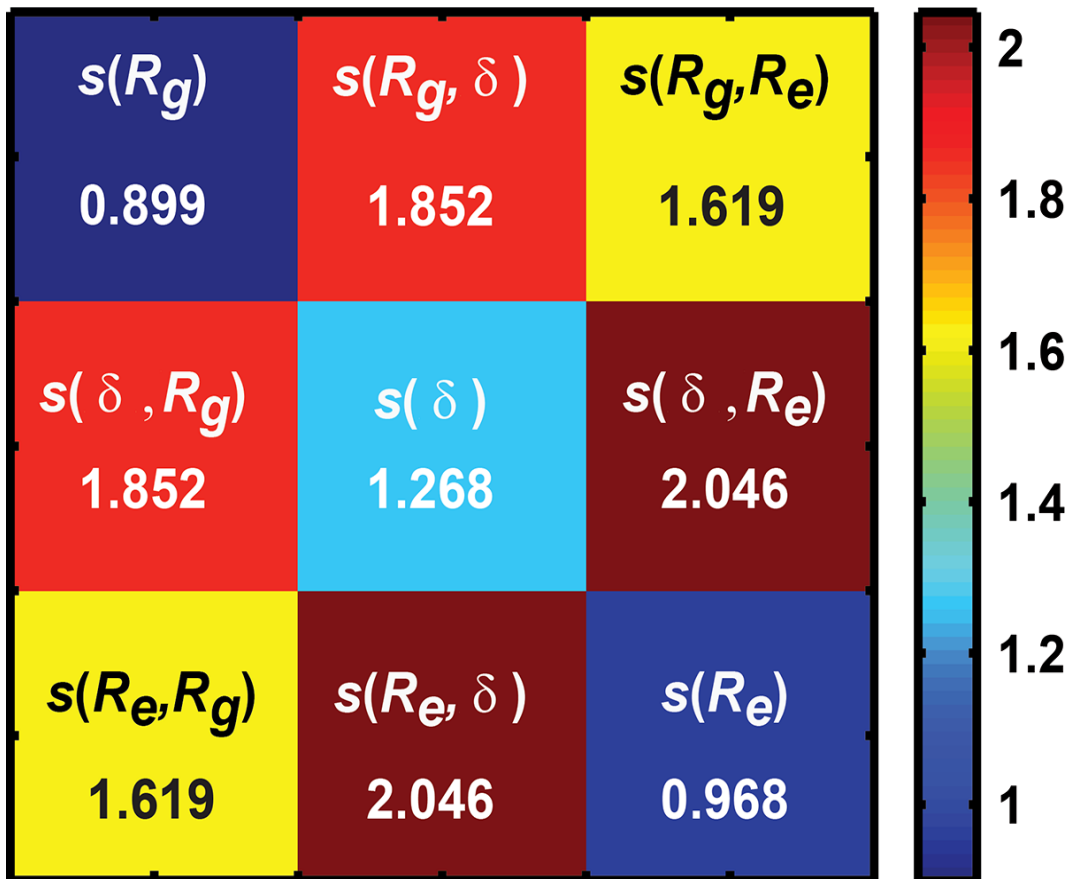
For a one-parameter distribution tiled into  $n$  bins, the corresponding information theoretic entropy is computed as:

$$s(X) = - \prod_{i=1}^n p(X_i) \ln p(X_i) \quad \text{Equation 3.2}$$

The information theoretic entropies computed using a total of three one-parameter and three two-parameter marginal distributions become elements of the symmetric square matrix that we refer to as the *ensemble entropy matrix*, denoted as  $\mathbf{S}_{\text{ens}}$ , and computed as:

$$\mathbf{S}_{\text{ens}} = \begin{bmatrix} s(R_g) & s(R_g, d) & s(R_g, R_e) \\ s(d, R_g) & s(d) & s(d, R_e) \\ s(R_e, R_g) & s(R_e, d) & s(R_e) \end{bmatrix} \quad \text{Equation 3.3}$$

The ensemble entropy matrix provides a concise description of the SERs for a specific sequence. For an idealized, maximally heterogeneous system with a flat distribution, setting  $n = 4$  equal-sized intervals will lead to upper limits of 1.386 for the diagonal elements and 2.773 for the off-diagonal elements of  $\mathbf{S}_{\text{ens}}$ . **Figure 3.6** shows the ensemble entropy matrix that we compute from all-atom simulations for an archetypal polyampholytic IDP *viz.*, the 67-residue disordered C-terminal tail (CTT) from the bacterial cell division protein FtsZ of *B. subtilis*.





**Figure 3.6: Example of an entropy matrix.** From the grid of probabilities for parameters X and Y (see **Figure 3.5**) for  $R_g$  &  $\delta$ , the information theoretic entropy  $s(X,Y)$  of the two-dimensional distribution is computed, and the values are shown in the matrix.

### 3.4.3 Generating comparative assessments of SERs

The ensemble entropy matrix can be calculated using simulation results for a sequence of interest. If we have a family of sequences, then the simulated ensembles for each sequence can be reduced to a sequence-specific ensemble entropy matrix. For a pair of sequences A and B, we shall denote the corresponding ensemble entropy matrices as  $\mathbf{S}_{\text{ens,A}}$  and  $\mathbf{S}_{\text{ens,B}}$ , respectively. For this pair of sequences, we define a difference ensemble entropy matrix as:  $\Delta_{AB} = (\mathbf{S}_{\text{ens,A}} - \mathbf{S}_{\text{ens,B}})$ . The dissimilarity between SERs for sequences A and B is calculated as the Frobenius norm of the difference ensemble entropy matrix according to:

$$D_{AB} = \|\Delta_{AB}\|_F = \sqrt{\sum_{i=1}^3 \sum_{j=i}^3 \left(D_{ij}^{(AB)}\right)^2} \quad \text{Equation 3.4}$$

In **Equation 3.4**,  $\|\Delta_{AB}\|_F$  denotes the Frobenius norm of  $\Delta_{AB}$  and  $D_{ij}^{(AB)}$  are the elements of  $\Delta_{AB}$ . If the SERs, described quantitatively in terms of  $\mathbf{S}_{\text{ens}}$  matrices, are essentially identical for a pair of sequences A and B, then  $D_{AB} \approx 0$ ; conversely, for a pair of sequences whose SERs are maximally dissimilar,  $D_{AB} = 5.369$ . In reality, the constraints imposed by chain connectivity and excluded volume considerations imply that the upper bound will be considerably lower than the theoretical maximum set by assuming an ensemble defined by flat one- and two-parameter marginal distributions. However, the theoretical lower and upper bounds provide a useful rubric for comparing SERs across pairs of sequences. This is first illustrated for a set of sequences of identical length and amino acid composition. We then follow this up by using  $\mathbf{S}_{\text{ens}}$  derived

dissimilarity measures to compare SERs for sequences derived from the same functional family across orthologs.

### **3.4.4 Comparative assessments of SERs across a set of sequences of identical lengths and amino acid compositions**

A significant amount of IDP sequences are classified as strong polyampholytes. In these systems, the FCR values are greater than 0.3, whereas the NCPR values are close to zero [102]. The calculated and measured values for ensemble-averaged radii of gyration ( $\langle R_g \rangle$ ) of strong polyampholytic IDPs are governed by the mixing versus segregation of oppositely charged residues within the linear sequence [36]. This feature, referred to as charge patterning, can be quantified by a parameter known as  $\kappa$  or other variants of this parameter [38]. For a fixed amino acid composition, one can design a series of sequence variants characterized by  $\kappa$  values that range from 0 to 1, where 0 implies sequences where the oppositely charged residues are evenly distributed along the linear sequence, and 1 implies that the oppositely charged residues are segregated into distinct blocks along the sequence.

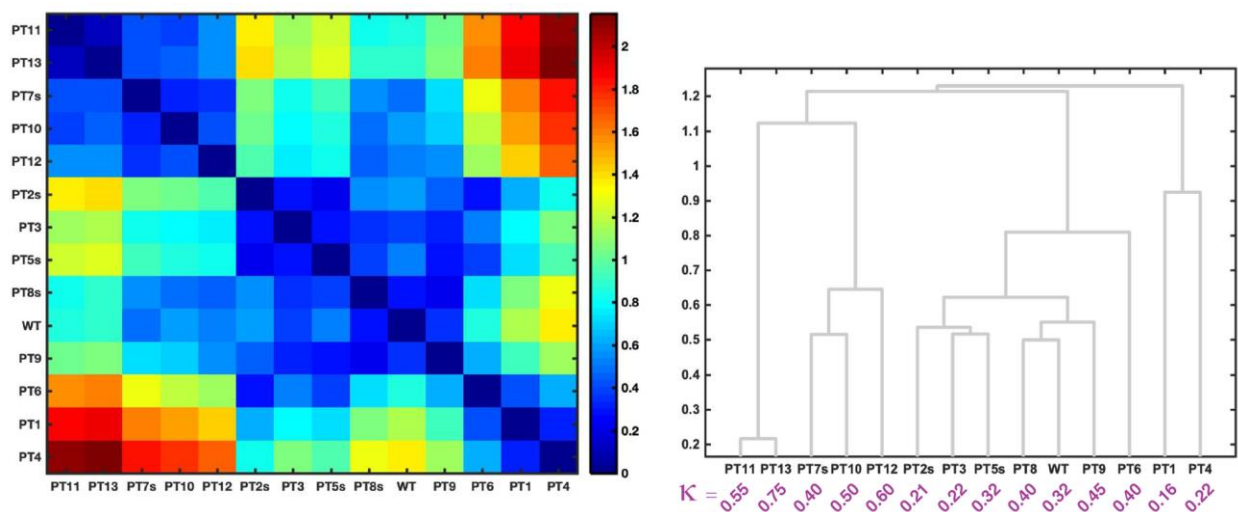
Recent investigations have quantified the impact of changes to  $\kappa$  on the global dimensions of different IDPs and the functions controlled by these IDPs [43, 44]. One such example is the Notch intracellular domain (NICD), which controls the transcription of Notch genes that contribute to the determination of cell fate in metazoans [103, 104]. NICD has a bipartite RAM-ANK architecture comprising of an IDR in the form of a 120-residue RAM region that is fused to an Ankyrin (ANK) repeat domain. Sherry et al. [43] recently designed a set of NICD variants with identical ANK domains but differed in the charge patterning of their RAM regions. The amino

acid compositions of the RAM regions and the sequence of the motif that binds to the transcription factor CSL were also identical across the designed variants. In all, thirteen distinct NICD variants were designed, and the  $\kappa$  values of the RAM regions of these sequences are shown in **Table 3.1**. Sherry et al. found that the computed  $\langle R_g \rangle$  values and measured hydrodynamic radii of RAM variants show an inverse correlation with  $\kappa$  [43]. The Pearson  $r$ -values quantifying the inverse correlations were 0.895 and 0.858, respectively. Interestingly, the average asphericity values showed a considerably weaker inverse correlation with  $\kappa$  with a Pearson  $r$ -value of 0.55.

<b>Variant</b>	$\kappa$	$\langle R_g \rangle \text{ \AA}$	$\langle R_e \rangle \text{ \AA}$	$\langle \delta \rangle$
<b>WT</b>	0.32	26.93	42.00	0.33
<b>PT1</b>	0.16	29.80	58.77	0.35
<b>PT2s</b>	0.21	28.19	46.66	0.31
<b>PT3s</b>	0.22	27.55	52.53	0.32
<b>PT4</b>	0.22	32.43	64.24	0.42
<b>PT5s</b>	0.32	27.43	46.21	0.32
<b>PT6</b>	0.40	28.14	53.46	0.33
<b>PT7s</b>	0.40	25.48	40.43	0.30
<b>PT8s</b>	0.44	26.23	49.55	0.31
<b>PT9</b>	0.45	26.40	48.67	0.36
<b>PT10</b>	0.50	25.04	45.41	0.28
<b>PT11</b>	0.55	24.85	47.05	0.26
<b>PT12</b>	0.60	24.85	49.19	0.37
<b>PT13</b>	0.75	24.65	42.88	0.28

**Table 3.1: Summary of  $\kappa$  values and parameters extracted from all-atom simulations for the RAM region extracted from the WT and designed NICD variants. Data are from the work of Sherry et al., [97].**

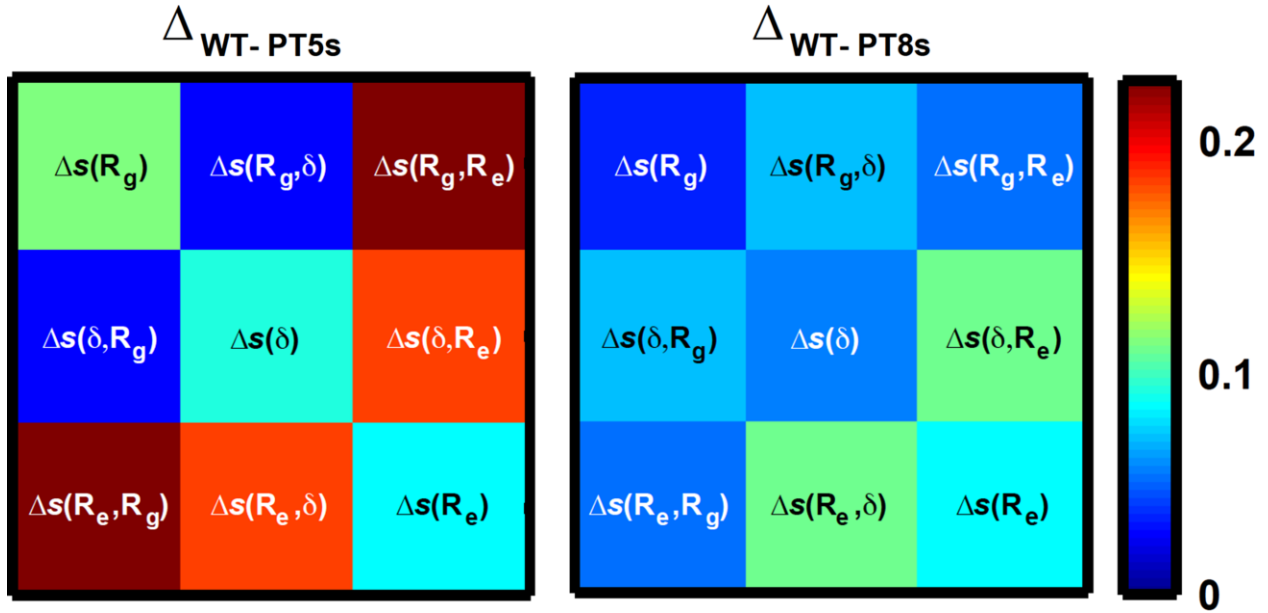
We computed ensemble entropy matrices using results from all-atom simulations [43] for each of RAM regions from each of 13 NICD variants as well as the WT sequence. These matrices were used to compute pairwise dissimilarities ( $D_{AB}$ ) between SERs, and the results are shown in the form of a checkerboard plot on the left side of **Figure 3.7**. The RAM variants whose SERs are most similar to those of the WT sequence are PT8s ( $\kappa=0.44$ ), PT9 ( $\kappa=0.45$ ), PT3 ( $\kappa=0.22$ ), PT5s ( $\kappa=0.32$ ), and PT7s ( $\kappa=0.40$ ). The pairwise dissimilarities derived from ensemble entropy matrices were used to construct a dendrogram that arranges sequences in terms of the extent of similarity of their SERs. This is shown on the right side of **Figure 3.7** for the RAM variants. This dendrogram provides a visual summary of the detailed information presented in the checkerboard plot of panel A. It highlights the fact that statistically robust conformational distributions generated from all-atom simulations can be used to quantify similarities / dissimilarities between pairs of IDPs.



**Figure 3.7: Comparative assessments of SERs for the RAM regions of NICD variants. (Left)** Checkerboard plot of the similarity between ensemble entropy matrices of the system of Notch-RAM variants. The dissimilarity between ensemble entropy matrices  $S_{ens,A}$  and  $S_{ens,B}$  is calculated as the Frobenius norm of the difference ensemble entropy matrix according to **Equation 3.4**. **(Right)** Dendrogram of sequences grouped based on the similarities of their  $S_{ens}$  matrices.

### 3.4.5 Comparison of the difference ensemble entropy matrices $\Delta_{AB}$ for the PT8s-WT and PT5s-WT pairs

The results for the designed RAM variants illustrate the importance of using multi-parameter conformational distributions for quantifying and comparing SERs. To make this point, we consider pairwise comparisons of the WT RAM with RAM variants of PT5s and PT8s, respectively. The dissimilarity measure  $D_{AB}$  is smallest for the PT8s-WT pair even though PT5s has a  $\kappa$ -value that is identical to WT RAM. Comparison of the difference ensemble entropy matrices  $\Delta_{AB}$  for the PT8s-WT and PT5s-WT pairs, shown in **Figure 3.8**, reveals the following: Despite having identical  $\kappa$ -values, identical ensemble-averaged  $R_g$  values, and very similar ensemble-averaged asphericity values, the two-parameter marginal distributions  $p(R_g, R_e)$  and  $p(\delta, R_e)$  are considerably different across the WT and PT5s sequences. This is underscored by the magnitudes of the differences between  $s(R_g, R_e)$  and  $s(\delta, R_e)$  for WT RAM versus the RAM region from PT5s. These differences arise because of the sequence-specificity of size and shape distributions as well as sequence-specificity in the decoupling of  $R_g$  and  $R_e$  distributions. We also computed difference ensemble entropy matrices for the RAM regions of WT ( $\kappa=0.32$ ) and PT8s ( $\kappa=0.44$ ). Interestingly, while the ensemble-averaged  $R_g$ ,  $R_e$ , and  $\delta$  values of PT8s are more different from those of WT RAM when compared to PT5s versus WT RAM (**Table 3.1**), the overall dissimilarity measure  $D_{AB}$  is lowest for PT8s vis-à-vis the WT RAM.

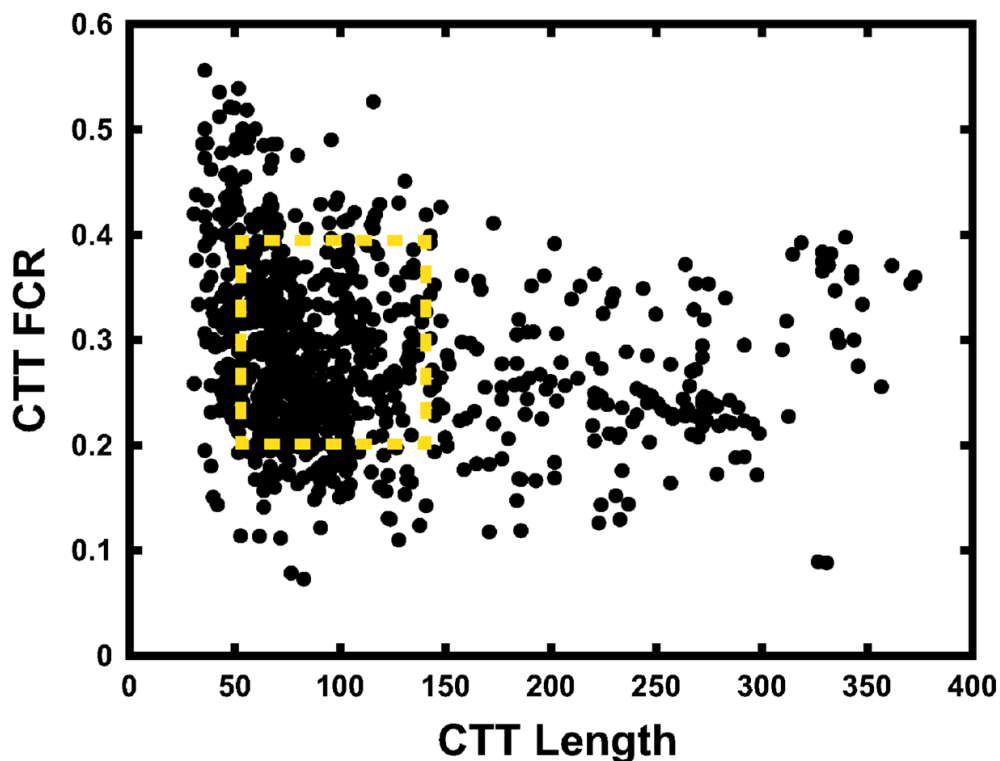


**Figure 3.8: Comparison of the difference ensemble entropy matrices  $\Delta_{AB}$  for the PT8s-WT and PT5s-WT pairs.** The color bar for both of the difference matrices is shown on the right-hand side.

The comparisons illustrated above highlight two crucial features of IDPs: similarities in the values of first moments of one-parameter marginal distributions such as  $\langle R_g \rangle$ ,  $\langle R_e \rangle$ , and  $\langle \delta \rangle$  can mask significant dissimilarities in the details of conformational ensembles. Secondly, dissimilarities in values for the first moments need not necessarily mean that the overall conformational ensembles have equivalent dissimilarities. Instead, conformational fluctuations can give rise to similar distributions, thereby offsetting differences in first moments. Conformational fluctuations are the defining hallmark of systems such as IDPs / IDRs. It is therefore important to account for conformational distributions to account for quantitative similarities / dissimilarities between sequence-specific ensembles. Our results emphasize the need for computing SERs using entire distributions, preferably computed in terms of multiple parameters, rather than over-interpreting changes observed from the scaling of first moments such as  $\langle R_g \rangle$  or  $\langle R_e \rangle$  with chain length,  $\kappa$ , or proxies for these parameters [36, 37, 105].

### 3.4.6 Quantitative SERs for IDRs derived from the same functional family across orthologs

Covariation in sequence helps unmask cryptic sequence-structure relationships, and this can be used to improve structure prediction, uncover the determinants of protein function, and enable novelties in protein design [52, 106, 107]. Although covariation analysis has been adapted to predict the presence of specific structural motifs within IDPs / IDRs, this type of analysis requires large-scale MSAs and a high degree of sequence conservation as well. Most IDPs / IDRs are characterized by very poor sequence conservation. Examples include the RAM region of NICD [108], discussed in the preceding section, and the disordered C-terminal tail of the bacterial protein FtsZ [109]. Cell division in rod-shaped bacteria involves the polymerization and assembly of FtsZ into so-called Z-rings that form at the midsection of the dividing cell [110, 111]. FtsZ, which is a bacterial homolog of tubulin, is also a GTPase, and it has a sequence architecture that is modular (**Figure 1.1**). The GTPase, referred to as the core, is an ordered domain with a C-terminal tail (CTT) attached to it. The CTT is essential for Z-ring formation *in vivo* in bacteria [112].



**Figure 3.9: Scatter plot of CTT sequence parameters summarized in terms of CTT-length and the Fraction of Charged Residues (FCR).** To compare the conformational distributions of CTTs, we performed all-atom simulations for 58 distinct CTTs drawn from the yellow region (UniProt IDs are listed in the Appendix in **Table 3A-1**).

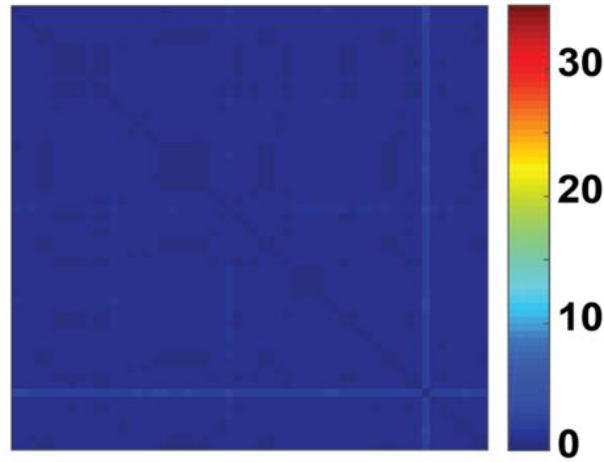
A recent analysis quantified various sequence features for CTT sequences derived from 1208 different FtsZ proteins across bacterial orthologs [109]. A summary of this analysis in **Figure 3.9** shows a scatter plot in a two-parameter space defined by the apparent lengths of CTTs and CTT FCR. This analysis suggests a confounding level of variation for the CTT sequences. The average CTT sequence is 60-70 residues long with an average FCR value of 0.3. However, there are significant deviations from the average values for CTT length and FCR. In fact, the distribution of points in **Figure 3.9** has an overall L-shape, suggesting that longer CTTs tend to have lower FCR values, whereas shorter sequences have higher FCR values. A recent study showed that the



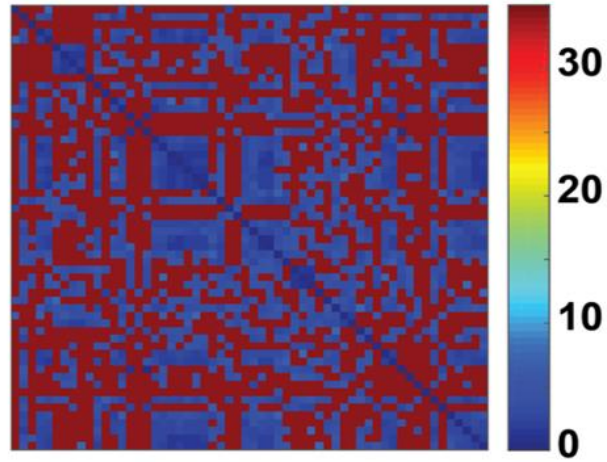
relevant parameter for disordered linkers and tails is the so-called thermodynamic or *functional length* [83] and not the number of residues within an IDR sequence, which we refer to as the apparent length. Instead, the functional length is governed by the thermodynamically preferred end-to-end distance realized by the disordered linker or tail that is encoded by at least two parameters, namely, the apparent length and the FCR, and is also influenced by extrinsic parameters such as solution conditions.

Covariation in the values of apparent lengths and FCR suggest that there are likely to be preferred conformational distributions encoded by functional CTTs. To quantify and compare these conformational distributions, we performed all-atom simulations for 58 distinct CTTs. The sequences of these CTTs span a range of length and FCR values, and they are drawn from the yellow bounded region within **Figure 3.9**. For our analysis, we first computed the distances between pairs of sequences for FtsZ cores derived from MSAs of the cores alone. As expected, the sequences of the core GTPase domains of FtsZ proteins show minimal variation across orthologs (**Figure 3.10**). In contrast, from a similar analysis performed on the CTT sequences, we find that the pairwise sequence similarities have an apparent bimodality (**Figure 3.10**); pairs of CTT sequences are either highly similar or highly dissimilar. While the former is expected, the latter is surprising because the CTTs are essential for FtsZ function, and yet there appears to be a range of very different sequences that can function as CTTs.

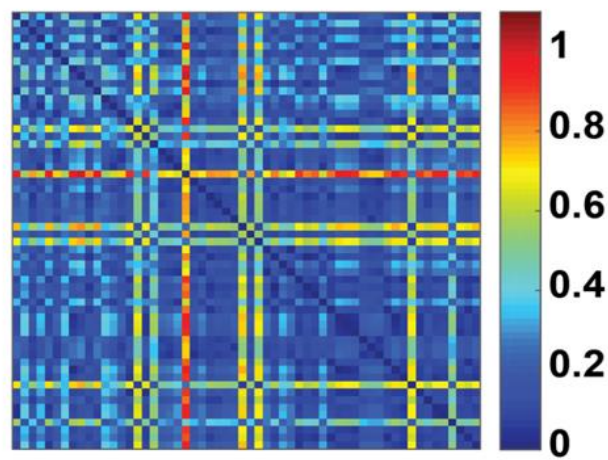
**Pairwise Distances between  
FtsZ Cores from MSA**



**Pairwise Distances between  
FtsZ CTTs from MSAs**



**Pairwise Frobenius norms of differences  
in ensemble entropy matrices for FtsZ CTTs**

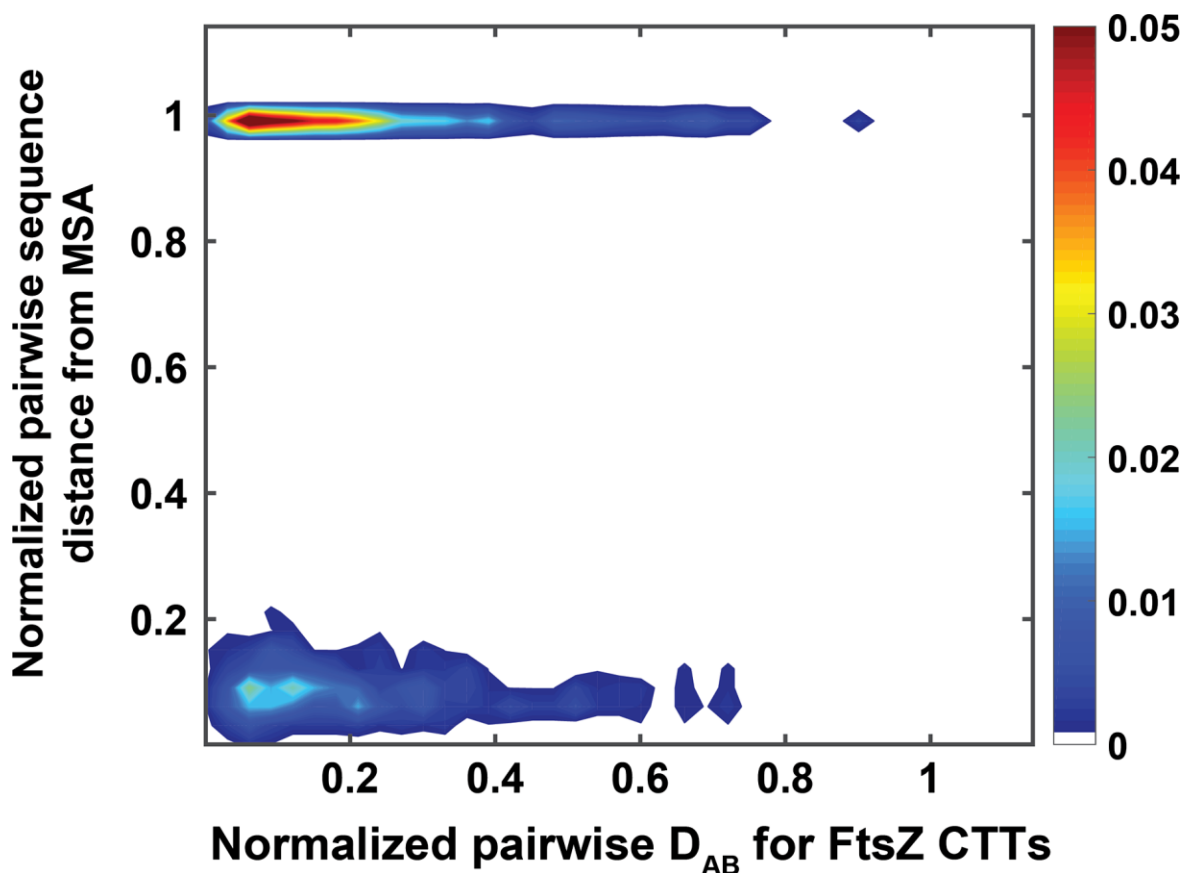


**Figure 3.10: Pairwise distances of sequences versus sequence ensemble entropy matrices** (Top) Checkerboard plot of the pairwise distances extracted from a MSA shows minimal variation between pairs of cores. (Middle) Data from MSAs of CTT sequences exhibit a bimodality whereby the sequences are either very similar or very different from one another. The color bar used to annotate all panels are the same. (Bottom) Checkerboard plot of pairwise differences between ensemble entropy matrices for FTsZ CTTs.

We used simulation results for 58 different CTTs and computed pairwise dissimilarity measures ( $D_{AB}$ ) using the SERs that we obtained for each of the CTTs. The results are shown as a checkerboard plot in **Figure 3.11**. Since the IDR lengths differ for this analysis, we normalized the length-dependent parameters ( $R_g$  and  $R_e$ ) by the square root of the length before computing the one- and two-parameter marginal distributions required to construct the ensemble entropy matrices. The bimodality that is apparent in the comparison of CTT sequences is not manifest in the  $D_{AB}$  used to quantify similarities / dissimilarities in SERs.

Next, we computed the Frobenius norms of pairwise difference ensemble entropy matrices and normalized these values by the maximal norms. We also computed normalized pairwise distances between CTT sequences, where the normalization was performed using the maximal difference between sequences. **Figure 3.11** shows a 2-dimensional histogram of the two normalized values computed for all 58 FtsZ-CTT sequences. If the normalized Frobenius norms of pairwise difference ensemble entropy matrices *and* normalized pairwise sequence differences are less than 0.5, the implication is that similar sequences will have similar SERs. This region, which corresponds to the lower-left corner of the 2-dimensional histogram, encompasses 40% of the CTTs studied here. If normalized Frobenius norms of pairwise difference ensemble entropy matrices and normalized pairwise sequence distance differences are both above 0.5, then the differences in CTT sequences translate to differences in SERs – corresponds to 7% of CTTs studied here.

Interestingly, 46% of the CTTs studied here fall into the top left region of the 2-dimensional histogram. In this region, the normalized Frobenius norms of pairwise difference ensemble entropy matrices are below 0.5, whereas the normalized differences between sequences exceed 0.5. These CTT sequences have very similar SERs despite having very different sequences. Overall, the results highlight the value of analyzing SERs across sequences derived from orthologs. The implication is that for over 86% of the FtsZ CTTs studied here, the SERs, quantified in terms of multi-parameter conformational distributions, are largely preserved, even though in 46% of these sequences, the pairwise sequence comparisons show considerable divergence.



**Figure 3.11: Histogram of SERs and sequence similarities.** Two-dimensional histogram of the normalized pairwise CTT  $D_{AB}$  values and normalized pairwise distances between aligned CTT sequences. Both distributions are normalized by their maximum values.

## 3.5 Discussion

IDPs / IDRs feature prominently in eukaryotic proteomes [113]. As semantic descriptions for systems exhibiting different degrees of conformational heterogeneity were being developed, terms such as intrinsically unstructured proteins were introduced [25] and even used in the literature [114-116] [117]. This phraseology can erroneously be taken to imply that IDPs are unstructured. However, being disordered and being unstructured are not the same [118], and this has become clear with intense efforts yielding formal descriptions of sequence-ensemble relationships for IDPs / IDRs. Indeed, IDPs / IDRs come in distinct sequence flavors [34, 35, 119], and the types of ensembles that are accessible to IDPs / IDRs are governed by the information encoded within their sequence. They can also be influenced by a variety of extrinsic factors, including solution conditions [120], concentrations of ligands, and epigenetic modifications [39]. Disorder refers to measures that quantify the multi-parameter, sequence-specific conformational distributions that reflect the interplay of chain-solvent and intra-chain interactions [49]. In this context, it is worth noting that numerous bioinformatics servers predict disordered regions within protein sequences. These predictors are valuable because they identify, with reasonable accuracy, the regions that are likely to be IDPs / IDRs as opposed to AFDs or autonomously folded domains. These types of disorder predictions serve as the starting point for quantitative studies of SERs. Having identified a disordered region, one can perform suitable all-atom simulations of the required throughput to generate thermodynamically accurate and statistically robust descriptions of conformational ensembles. Unlike an AFD that can often be described in terms of one or a small set of structures, with atomic coordinates in models for the structures, one needs statistical descriptions of SERs. Here, we introduced a three-parameter distribution function that we

decompose into three one-parameter and three two-parameter marginal distributions to then construct an ensemble entropy matrix ( $\mathbf{S}_{\text{ens}}$ ) for a given IDP / IDR sequence. These  $\mathbf{S}_{\text{ens}}$  matrices serve to synthesize the SERs by quantifying the information contained in distribution functions. Construction of pairwise difference ensemble entropy matrices and the calculation of Frobenius norms of these difference matrices allow us to compare sequences to one another through the lens of their SERs.

Our results reveal interesting insights regarding the SERs of IDPs / IDRs. Analysis of the RAM regions of NICD variants shows that while a single sequence encoded parameter such as  $\kappa$  is useful for generating novel variants, it does not fully describe SERs. Interestingly, considerable attention has focused on the calculation / measurement of first moments of conformational distributions such as ensemble-averaged values for  $R_g$ ,  $R_e$ , and asphericity. These quantities show coherent variations with parameters such as chain length and sequence patterning for homopolymers and low complexity systems such as block copolymers. These simple systems are defined by the homogeneity of interactions on all length scales. However, most IDPs / IDRs are finite-sized heteropolymers. The sidechains are of different sizes, feature different functional groups, and encode very different balances between sidechain-solvent and sidechain-sidechain interactions. These chemical details cannot be readily captured using coarse-grained approximations for heteropolymers [101]. Recent studies have highlighted the importance of chemical heterogeneity on decoupling size and shape fluctuations and also the fluctuations of  $R_g$  and  $R_e$  [75, 100, 121]. This type of decoupling raises caution about inferring SERs purely from the scaling of ensemble-averaged values of  $R_g$ ,  $R_e$ , or asphericity. Instead, a complete description of SERs requires measurements of sizes, shapes, and conformational distributions and / or dynamics

[122, 123]. Alternatively, one can pursue all-atom simulations providing they are efficient enough to enable sufficient throughput while also ensuring that they are accurate.

Of course, one cannot be certain of the accuracy of forcefields that are used for simulations of IDPs / IDRs. Considerable efforts have gone into the refinement of forcefields that are interoperable with explicit representations of solvent molecules [68]. We have primarily focused on the development, testing, refinement, and deployment of the ABSINTH implicit solvation model and forcefield paradigm for the simulations of IDPs / IDRs. To date, this model, aided by a variety of enhanced sampling methods, has been used to make predictions of SERs for well over  $3 \times 10^4$  IDPs / IDRs of different lengths and sequence complexities. Recent efforts have also focused on simulations of IDRs as tails and linkers tethered to ordered domains [83, 98]. Where possible, these simulations have been used to make testable predictions for scrutiny by experiment or reanalyzed using reweighting approaches based on experimental data. A new version of ABSINTH, known as ABSINTH-C [124], was developed to remedy inaccuracies in the descriptions of local conformational equilibria. At the level of the conformational distributions studied here, it appears that the two versions generate roughly equivalent results. We propose that it should be possible to deploy ABSINTH / ABSINTH-C based simulations in high throughput investigations of SERs for IDPs / IDRs and combine this with a modified version of a recently developed genetic algorithm for designing sequences with bespoke SERs [93].

Our results for FtsZ CTTs underscore the importance of using SERs as a signal that can be modulated through sequence design. For example, one could conceive a design strategy that generates a diverse library of CTTs that support bacterial cell division by ensuring the preservation of SERs. These designs can be guided by taxonomic inferences gleaned from a dendrogram that

synthesizes all of the data in the matrix of all pairwise dissimilarity values. This dendrogram, shown in **Figure 3.12**, is a similarity tree that groups CTT sequences based on the similarities of their SERs. We propose that to a first approximation, the CTTs with similar or identical SERs could be functionally interoperable with one another. We anticipate that the use of SERs, fueled by advancements in computational efficiency and accuracy, will enable the emergence of new design paradigms that target the sequences of IDRs / IDPs for impacting cellular processes and circuits controlled by proteins with disordered regions.



**Figure 3.12: Dendrogram that arranges FtsZ CTT sequences based on similarities of SERs.** The sequence IDs are derived from **Table 3A.1**.



## 3.6 References

1. Baker, D., *What has de novo protein design taught us about protein folding and biophysics?* Protein Sci, 2019. **28**(4): p. 678-683.
2. Smith, B.A. and M.H. Hecht, *Novel proteins: from fold to function.* Curr Opin Chem Biol, 2011. **15**(3): p. 421-6.
3. ElGamacy, M., et al., *An Interface-Driven Design Strategy Yields a Novel, Corrugated Protein Architecture.* ACS Synthetic Biology, 2018. **7**(9): p. 2226-2235.
4. Pessi, A., et al., *A designed metal-binding protein with a novel fold.* Nature, 1993. **362**(6418): p. 367-369.
5. Khare, S.D. and S.J. Fleishman, *Emerging themes in the computational design of novel enzymes and protein-protein interfaces.* Febs Letters, 2013. **587**(8): p. 1147-54.
6. Johansson, K.E. and K. Lindorff-Larsen, *Structural heterogeneity and dynamics in protein evolution and design.* Current Opinion In Structural Biology, 2018. **48**: p. 157-163.
7. Kundert, K. and T. Kortemme, *Computational design of structured loops for new protein functions.* Biol Chem, 2019. **400**(3): p. 275-288.
8. Marcos, E., et al., *De novo design of a non-local beta-sheet protein with high stability and accuracy.* Nature Structural & Molecular Biology, 2018. **25**(11): p. 1028-1034.
9. Geiger-Schuller, K., et al., *Extreme stability in de novo-designed repeat arrays is determined by unusually stable short-range interactions.* Proceedings of the National Academy of Sciences of the United States of America, 2018. **115**(29): p. 7539-7544.
10. Kuhlman, B. and D. Baker, *Exploring folding free energy landscapes using computational protein design.* Current Opinion In Structural Biology, 2004. **14**(1): p. 89-95.
11. Ponder, J.W. and F.M. Richards, *Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes.* Journal of Molecular Biology, 1987. **193**(4): p. 775-91.
12. Yue, K. and K.A. Dill, *Inverse protein folding problem: designing polymer sequences.* Proc Natl Acad Sci U S A, 1992. **89**(9): p. 4163-7.

13. Bowie, J.U., R. Luthy, and D. Eisenberg, *A method to identify protein sequences that fold into a known three-dimensional structure*. Science, 1991. **253**(5016): p. 164-70.
14. Gainza-Cirauqui, P. and B.E. Correia, *Computational protein design-the next generation tool to expand synthetic biology applications*. Current Opinion In Biotechnology, 2018. **52**: p. 145-152.
15. Pleiss, J., *Protein design in metabolic engineering and synthetic biology*. Current Opinion In Biotechnology, 2011. **22**(5): p. 611-7.
16. Fischer, M., et al., *MarkUs: a server to navigate sequence-structure-function space*. Nucleic Acids Research, 2011. **39**(Web Server issue): p. W357-61.
17. Marchler-Bauer, A., et al., *CDD: a Conserved Domain Database for the functional annotation of proteins*. Nucleic Acids Res, 2011. **39**(Database issue): p. D225-9.
18. Qi, Y. and N.V. Grishin, *Structural classification of thioredoxin-like fold proteins*. Proteins, 2005. **58**(2): p. 376-88.
19. Shakhnovich, B.E., et al., *ELISA: structure-function inferences based on statistically significant and evolutionarily inspired observations*. BMC Bioinformatics, 2003. **4**: p. 34.
20. Heinke, F., et al., *eProS--a database and toolbox for investigating protein sequence-structure-function relationships through energy profiles*. Nucleic Acids Research, 2013. **41**(Database issue): p. D320-6.
21. Kundert, K. and T. Kortemme, *Computational design of structured loops for new protein functions*. Biological Chemistry, 2019. **400**(3): p. 275-288.
22. Mignon, D. and T. Simonson, *Comparing three stochastic search algorithms for computational protein design: Monte Carlo, replica exchange Monte Carlo, and a multistart, steepest-descent heuristic*. Journal of Computational Chemistry, 2016. **37**(19): p. 1781-93.
23. Wright, P.E. and H.J. Dyson, *Intrinsically disordered proteins in cellular signalling and regulation*. Nat Rev Mol Cell Biol, 2015. **16**(1): p. 18-29.
24. van der Lee, R., et al., *Classification of Intrinsically Disordered Regions and Proteins*. Chemical Reviews, 2014. **114**(13): p. 6589-6631.

25. Wright, P.E. and H.J. Dyson, *Intrinsically unstructured proteins: Re-assessing the protein structure-function paradigm*. Journal of Molecular Biology, 1999. **293**(2): p. 321-331.
26. Dunker, A.K., et al., *Intrinsic disorder and protein function*. Biochemistry, 2002. **41**(21): p. 6573-6582.
27. Forman-Kay, J.D. and T. Mittag, *From Sequence and Forces to Structure, Function, and Evolution of Intrinsically Disordered Proteins*. Structure, 2013. **21**(9): p. 1492-1499.
28. Lange, J., L.S. Wyrwicz, and G. Vriend, *KMAD: knowledge-based multiple sequence alignment for intrinsically disordered proteins*. Bioinformatics, 2016. **32**(6): p. 932-6.
29. Moesa, H.A., et al., *Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification*. Molecular BioSystems, 2012. **8**(12): p. 3262-3273.
30. Brown, C.J., et al., *Evolution and disorder*. Current Opinion in Structural Biology, 2011. **21**(3): p. 441-446.
31. Pirovano, W. and J. Heringa, *Multiple sequence alignment*. Methods in Molecular Biology, 2008. **452**: p. 143-61.
32. Bryan, P.N. and J. Orban, *Proteins that switch folds*. Current Opinion In Structural Biology, 2010. **20**(4): p. 482-8.
33. Wasserman, H. and E.O. Saphire, *More than Meets the Eye: Hidden Structures in the Proteome*. Annual Reviews in Virology, 2016. **3**(1): p. 373-386.
34. Holehouse, A.S., et al., *CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins*. Biophysical Journal, 2017. **112**(1): p. 16-21.
35. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Current Opinion In Structural Biology, 2015. **32**: p. 102-12.
36. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(33): p. 13392-13397.

37. Firman, T. and K. Ghosh, *Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins*. Journal of Chemical Physics, 2018. **148**(12): p. 123305.
38. Sawle, L. and K. Ghosh, *A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins*. Journal of Chemical Physics, 2015. **143**(8): p. 085101.
39. Martin, E.W., et al., *Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation*. Journal of the American Chemical Society, 2016. **138**(47): p. 15323-15335.
40. Beh, L.Y., L.J. Colwell, and N.J. Francis, *A core subunit of Polycomb repressive complex 1 is broadly conserved in function but not primary sequence*. Proceedings of the National Academy of Sciences USA, 2012. **109**(18): p. E1063-E1071.
41. Staller, M.V., et al., *A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain*. Cell Systems, 2018. **6**(4): p. 444-455.e6.
42. Beveridge, R., et al., *Ion Mobility Mass Spectrometry Uncovers the Impact of the Patterning of Oppositely Charged Residues on the Conformational Distributions of Intrinsically Disordered Proteins*. Journal of the American Chemical Society, 2019. **141**(12): p. 4908-4918.
43. Sherry, K.P., et al., *Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor*. Proceedings of the National Academy of Sciences of the United States of America, 2017. **114**(44): p. E9243-e9252.
44. Das, R.K., et al., *Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling*. Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(20): p. 5616-21.
45. Pak, C.W., et al., *Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein*. Molecular Cell, 2016. **63**(1): p. 72-85.
46. Nott, T.J., et al., *Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles*. Mol Cell, 2015. **57**(5): p. 936-947.
47. Franzmann, T.M., et al., *Phase separation of a yeast prion protein promotes cellular fitness*. Science, 2018. **359**(6371).
48. Wang, J., et al., *A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins*. Cell, 2018. **174**(3): p. 688-699.e16.

49. Lyle, N., R.K. Das, and R.V. Pappu, *A quantitative measure for protein conformational heterogeneity*. Journal of Chemical Physics, 2013. **139**(12).
50. Csizmok, V., et al., *Dynamic Protein Interaction Networks and New Structural Paradigms in Signaling*. Chemical Reviews, 2016. **116**(11): p. 6424-62.
51. Shannon, C.E., *A mathematical theory of communication*. The Bell System Technical Journal, 1948. **27**(3): p. 379-423.
52. Reynolds, K.A., et al., *Evolution-based design of proteins*. Methods in Enzymology, 2013. **523**: p. 213-235.
53. Hopf, T.A., et al., *Sequence co-evolution gives 3D contacts and structures of protein complexes*. Elife, 2014. **3**: p. 03430.
54. Marks, D.S., et al., *Protein 3D structure computed from evolutionary sequence variation*. PLoS One, 2011. **6**(12): p. e28766.
55. Brucale, M., B. Schuler, and B. Samori, *Single-Molecule Studies of Intrinsically Disordered Proteins*. Chemical Reviews, 2014. **114**(6): p. 3281-3317.
56. Jensen, M.R. and M. Blackledge, *Testing the validity of ensemble descriptions of intrinsically disordered proteins*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(16): p. E1557-E1558.
57. Kragelj, J., et al., *Conformational Propensities of Intrinsically Disordered Proteins from NMR Chemical Shifts*. Chemphyschem, 2013. **14**(13): p. 3034-3045.
58. Marsh, J.A. and J.D. Forman-Kay, *Sequence Determinants of Compaction in Intrinsically Disordered Proteins*. Biophysical Journal, 2010. **98**(10): p. 2383-2390.
59. Mittag, T. and J.D. Forman-Kay, *Atomic-level characterization of disordered protein ensembles*. Current Opinion in Structural Biology, 2007. **17**(1): p. 3-14.
60. Liu, B.X., et al., *The Effect of Intrachain Electrostatic Repulsion on Conformational Disorder and Dynamics of the Sic1 Protein*. Journal of Physical Chemistry B, 2014. **118**(15): p. 4088-4097.
61. Gibbs, E.B., E.C. Cook, and S.A. Showalter, *Application of NMR to studies of intrinsically disordered proteins*. Archives of Biochemistry and Biophysics, 2017. **628**: p. 57-70.

62. Ruff, K.M., R.V. Pappu, and A.S. Holehouse, *Conformational preferences and phase behavior of intrinsically disordered low complexity sequences: insights from multiscale simulations*. Current Opinion In Structural Biology, 2018. **56**: p. 1-10.
63. Zerze, G.H., et al., *Evolution of All-Atom Protein Force Fields to Improve Local and Global Properties*. Journal of Physical Chemistry Letters, 2019. **10**(9): p. 2227-2234.
64. Zheng, W., et al., *Inferring properties of disordered chains from FRET transfer efficiencies*. Journal of Chemical Physics, 2018. **148**(12): p. 123329.
65. Dignon, G.L., et al., *Relation between single-molecule properties and phase behavior of intrinsically disordered proteins*. Proceedings of the National Academy of Sciences USA, 2018. **115**(40): p. 9929-9934.
66. Zerze, G.H., R.B. Best, and J. Mittal, *Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations*. Journal of Physical Chemistry B, 2015. **119**(46): p. 14622-30.
67. Wuttke, R., et al., *Temperature-dependent solvation modulates the dimensions of disordered proteins*. Proceedings of the National Academy of Sciences of the United States of America, 2014. **111**(14): p. 5213-5218.
68. Best, R.B., W. Zheng, and J. Mittal, *Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association*. Journal of Chemical Theory and Computation, 2014. **10**(11): p. 5113-5124.
69. Ganguly, D., W.H. Zhang, and J.H. Chen, *Synergistic folding of two intrinsically disordered proteins: searching for conformational selection*. Molecular Biosystems, 2012. **8**(1): p. 198-209.
70. De Sancho, D. and R.B. Best, *Modulation of an IDP binding mechanism and rates by helix propensity and non-native interactions: association of HIF1alpha with CBP*. Molecular Biosystems, 2012. **8**(1): p. 256-67.
71. Turjanski, A.G., et al., *Binding-induced folding of a natively unstructured transcription factor*. PLoS Computational Biology, 2008. **4**(4): p. e1000060.
72. Vitalis, A., X.L. Wang, and R.V. Pappu, *Quantitative characterization of intrinsic disorder in polyglutamine: Insights from analysis based on polymer theories*. Biophysical Journal, 2007. **93**(6): p. 1923-1937.

73. Newcombe, E.A., et al., *Tadpole-like Conformations of Huntingtin Exon 1 Are Characterized by Conformational Heterogeneity that Persists regardless of Polyglutamine Length*. *Journal of Molecular Biology*, 2018. **430**(10): p. 1442-1458.
74. Warner, J.B.t., et al., *Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and Pathological Polyglutamine Lengths*. *Journal of the American Chemical Society*, 2017. **139**(41): p. 14456-14469.
75. Fuertes, G., et al., *Decoupling of size and shape fluctuations in heteropolymeric sequences reconciles discrepancies in SAXS vs. FRET measurements*. *Proceedings of the National Academy of Sciences USA*, 2017. **114**(31): p. E6342-e6351.
76. Ruff, K.M., T.S. Harmon, and R.V. Pappu, *CAMELOT: A machine learning approach for coarse-grained simulations of aggregation of block-copolymeric protein sequences*. *Journal of Chemical Physics*, 2015. **143**(24): p. 243123.
77. Vitalis, A. and A. Caflisch, *50 Years of Lifson-Roig Models: Application to Molecular Simulation Data*. *Journal of Chemical Theory and Computation*, 2012. **8**(1): p. 363-73.
78. Vitalis, A. and A. Caflisch, *Micelle-like architecture of the monomer ensemble of Alzheimer's amyloid-beta peptide in aqueous solution and its implications for Abeta aggregation*. *Journal of Molecular Biology*, 2010. **403**(1): p. 148-165.
79. Das, R.K., S.L. Crick, and R.V. Pappu, *N-Terminal Segments Modulate the alpha-Helical Propensities of the Intrinsically Disordered Basic Regions of bZIP Proteins*. *Journal of Molecular Biology*, 2012. **416**(2): p. 287-299.
80. Mao, A.H., et al., *Net charge per residue modulates conformational ensembles of intrinsically disordered proteins*. *Proceedings of the National Academy of Sciences of the United States of America*, 2010. **107**(18): p. 8183-8188.
81. Radhakrishnan, A., et al., *Improved Atomistic Monte Carlo Simulations Demonstrate That Poly-L-Proline Adopts Heterogeneous Ensembles of Conformations of Semi-Rigid Segments Interrupted by Kinks*. *Journal of Physical Chemistry B*, 2012. **116**(23): p. 6862-6871.
82. Harmon, T.S., et al., *Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins*. *Elife*, 2017. **6**: p. e30294.
83. Mittal, A., et al., *Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins*. *Journal of Molecular Biology*, 2018. **430**(16): p. 2403-2421.

84. Robustelli, P., S. Piana, and D.E. Shaw, *Developing a molecular dynamics force field for both folded and disordered protein states*. Proceedings of the National Academy of Sciences of the United States of America, 2018. **115**(21): p. E4758-e4766.
85. Piana, S., et al., *Water dispersion interactions strongly influence simulated structural properties of disordered protein states*. Journal of Physical Chemistry B, 2015. **119**(16): p. 5113-23.
86. Metskas, L.A. and E. Rhoades, *Conformation and Dynamics of the Troponin I C-Terminal Domain: Combining Single-Molecule and Computational Approaches for a Disordered Protein Region*. Journal of the American Chemical Society, 2015. **137**(37): p. 11962-9.
87. Gibbs, E.B. and S.A. Showalter, *Quantification of Compactness and Local Order in the Ensemble of the Intrinsically Disordered Protein FCPI*. Journal of Physical Chemistry B, 2016. **120**(34): p. 8960-9.
88. Vitalis, A. and R.V. Pappu, *ABSINTH: A New Continuum Solvation Model for Simulations of Polypeptides in Aqueous Solutions*. Journal of Computational Chemistry, 2009. **30**(5): p. 673-699.
89. Vitalis, A. and R.V. Pappu, *Methods for Monte Carlo simulations of biomacromolecules*. Annual Reports in Computational Chemistry, 2009. **5**: p. 49-76.
90. Vitalis, A. and R.V. Pappu, *A simple molecular mechanics integrator in mixed rigid body and dihedral angle space*. Journal of Chemical Physics, 2014. **141**(3).
91. Mittal, A., et al., *The ABSINTH implicit solvation model and forcefield paradigm for use in simulations of intrinsically disordered proteins*, in *Computational Approaches to Protein Dynamics: From Quantum to Coarse-Grained Methods*, M. Fuxreiter, Editor. 2015, CRC Press: Taylor & Francis Group. p. 181-204.
92. Munshi, S., et al., *Protein plasticity driven by disorder and collapse governs the heterogeneous binding of CytR to DNA*. Nucleic Acids Research, 2018. **46**(8): p. 4044-4053.
93. Harmon, T.S., et al., *GADIS: Algorithm for designing sequences to achieve target secondary structure profiles of intrinsically disordered proteins*. Protein Engineering, Design and Selection, 2016. **29**(9): p. 339-346.
94. Vitalis, A. and R. Pappu, *ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions*. Journal of Computational Chemistry, 2009. **30**(5): p. 673-699.



95. Mao, A.H. and R.V. Pappu, *Crystal lattice properties fully determine short-range interaction parameters for alkali and halide ions*. Journal of Chemical Physics, 2012. **137**(6): p. 064104(1-9).
96. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*. Chemical Physics Letters, 1999. **314**(1-2): p. 141-151.
97. Sherry, K.P., et al., *Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor*. Proc Natl Acad Sci U S A, 2017. **114**(44): p. E9243-E9252.
98. Mittal, A., et al., *Hamiltonian Switch Metropolis Monte Carlo Simulations for Improved Conformational Sampling of Intrinsically Disordered Regions Tethered to Ordered Domains of Proteins*. Journal of Chemical Theory and Computation, 2014. **10**(8): p. 3550-3562.
99. Steinhauser, M.O., *A molecular dynamics study on universal properties of polymer chains in different solvent qualities. Part I. A review of linear chain properties*. Journal of Chemical Physics, 2005. **122**(9): p. 094901.
100. Song, J., et al., *Conformational Heterogeneity and FRET Data Interpretation for Dimensions of Unfolded Proteins*. Biophysical Journal, 2017. **113**(5): p. 1012-1024.
101. Ruff, K.M. and A.S. Holehouse, *SAXS versus FRET: A Matter of Heterogeneity?* Biophysical Journal, 2017. **113**(5): p. 971-973.
102. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Curr Opin Struct Biol, 2015. **32**: p. 102-12.
103. Johnson, S.E., et al., *Thermodynamic analysis of the CSL x Notch interaction: distribution of binding energy of the Notch RAM region to the CSL beta-trefoil domain and the mode of competition with the viral transactivator EBNA2*. Journal of Biological Chemistry, 2010. **285**(9): p. 6681-92.
104. Kopan, R., *Notch Signaling*. Cold Spring Harbor Perspectives in Biology, 2012. **4**(10): p. a011213.
105. Riback, J.A., et al., *Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water*. Science, 2017. **358**(6360): p. 238-241.

106. Raman, A.S., K.I. White, and R. Ranganathan, *Origins of Allostery and Evolvability in Proteins: A Case Study*. Cell, 2016. **166**(2): p. 468-480.
107. Salinas, V.H. and R. Ranganathan, *Coevolution-based inference of amino acid interactions underlying protein function*. Elife, 2018. **7**: p. e34300.
108. Bertagna, A., et al., *The effects of conformational heterogeneity on the binding of the Notch intracellular domain to effector proteins: a case of biologically tuned disorder*. Biochemical Society Transactions, 2008. **36**: p. 157-166.
109. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division*. Seminars in Cell & Developmental Biology, 2015. **37**: p. 3-10.
110. den Blaauwen, T., L.W. Hamoen, and P.A. Levin, *The divisome at 25: the road ahead*. Current Opinion in Microbiology, 2017. **36**: p. 85-94.
111. Wehrens, M., et al., *Size Laws and Division Ring Dynamics in Filamentous Escherichia coli cells*. Current Biology, 2018. **28**(6): p. 972-979.e5.
112. Buske, P.J. and P.A. Levin, *A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo*. Molecular Microbiology, 2013. **89**(2): p. 249-63.
113. Liu, J.G., et al., *Intrinsic disorder in transcription factors*. Biochemistry, 2006. **45**(22): p. 6873-6888.
114. Harauz, G., et al., *Myelin basic protein-diverse conformational states of an intrinsically unstructured protein and its roles in myelin assembly and multiple sclerosis*. Micron, 2004. **35**(7): p. 503-42.
115. Huhn, A.J., et al., *The high-molecular-weight kininogen domain 5 is an intrinsically unstructured protein and its interaction with ferritin is metal mediated*. Protein Science, 2014. **23**(8): p. 1013-22.
116. Takahashi, M., et al., *Polyglutamine tract binding protein-1 is an intrinsically unstructured protein*. Biochimica Biophysica Acta, 2009. **1794**(6): p. 936-43.
117. Bencivenga, D., et al., *p27(Kip1) and human cancers: A reappraisal of a still enigmatic protein*. Cancer Letters, 2017. **403**: p. 354-365.

118. Smyth, E., et al., *Solution structure of native proteins with irregular folds from Raman optical activity*. Biopolymers, 2001. **58**(2): p. 138-51.
119. Hofmann, H., et al., *Polymer scaling laws of unfolded and intrinsically disordered proteins quantified with single-molecule spectroscopy*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(40): p. 16155-16160.
120. Moses, D., et al., *Revealing the Hidden Sensitivity of Intrinsically Disordered Proteins to their Chemical Environment*. The Journal of Physical Chemistry Letters, 2020. **11**: p. 10131-10136.
121. Peran, I., et al., *Unfolded states under folding conditions accommodate sequence-specific conformational preferences with random coil-like dimensions*. Proceedings of the National Academy of Sciences, 2019: p. 201818206.
122. Jensen, M.R., R.W. Ruigrok, and M. Blackledge, *Describing intrinsically disordered proteins at atomic resolution by NMR*. Current Opinion In Structural Biology, 2013. **23**(3): p. 426-35.
123. Jensen, M.R., et al., *Exploring free-energy landscapes of intrinsically disordered proteins at atomic resolution using NMR spectroscopy*. Chemical Reviews, 2014. **114**(13): p. 6632-60.
124. Choi, J.M. and R.V. Pappu, *Improvements to the ABSINTH Force Field for Proteins Based on Experimentally Derived Amino Acid Specific Backbone Conformational Statistics*. Journal of Chemical Theory and Computation, 2019. **15**(2): p. 1367-1382.
125. Consortium, U., *UniProt: a worldwide hub of protein knowledge*. Nucleic acids research, 2019. **47**(D1): p. D506-D515.

### 3.7 Appendix

Sequence ID	Uniprot ID
1	E4Q020
2	Q7VI30
3	G2RZ44
4	D8K1P8
5	E4TFX0
6	B3CLB8
7	D8JZV8
8	A6LSH0
9	C4L5U8
10	Q5GS34
11	I0I9A4
12	D3FTC8
13	Q6F708
14	D7AEK5
15	Q73H51
16	B8GMN5
17	A1U3F3
18	B3E3X6
19	D5CP17
20	Q2S9Z7
21	B3DVV6
22	E4ZBJ0
23	Q07WI9
24	Q1GZ07
25	Q6MT25
26	Q661W5
27	Q7VQI4
28	A8ZXV8
29	H6MRY6
30	D4G8Q5
31	A5UIR7
32	F5Y8Y2
33	Q50318

34	Q6AE67
35	K7SIK3
36	P45500
37	I4CCP3
38	F0RYV6
39	B2FPA8
40	Q2IY32
41	D0M7Z9
42	D4GAI3
43	F8HZP5
44	G7UWC1
45	Q6AJ58
46	Q3K2P0
47	B3EIK4
48	B4SH45
49	E4T4J7
50	Q8KGD8
51	G0HEC9
52	E4RT28
53	G8PCU7
54	C6SRV4
55	Q313P8
56	D1C0C9
57	F5ZJY1
58	H8KHW3

**Table 3A.1: Sequence ID numbers and the UniProt ID.** Identifiers of FtsZ proteins that were used for this analysis. The disordered C-terminal linkers were extracted, and the analysis was performed on the core and CTT domain separately [125].

## Chapter 4

# Uncovering non-random sequence patterns encoded in sequences of hypervariable IDPs / IDRs

### 4.1 Preamble

Intrinsically disordered proteins / regions (IDPs / IDRs) pose unique challenges for efforts aimed at connecting linear amino acid sequence information to protein function. Since a single globally stable fold is by definition not realized for IDPs / IDRs, a specific linear amino acid sequence is not a prerequisite for structure and function. As a result, multiple sequences can potentially encode for similar functions, resulting in hypervariability. In **Chapter 3**, we demonstrated how disparate sequences can encode for similar sequence-ensemble relationships (SERs). Therefore, sequences that are very different can have similar conformational ensembles and potentially participate in similar functions. Extracting SERs across large numbers of orthologs is non-trivial. For example, in **Chapter 3**, we analyzed SERs for 58 out of ~1200 orthologs of the FtsZ CTTs. There are 20-fold more orthologs whose SERs need evaluation. While simulations have become increasingly more tractable, there also remains the issue of not being able to rely solely on SERs alone to identify non-random sequence patterns that might have functional

significance. This is especially likely in hypervariable IDRs, whereby new, bacterium-specific patterns might be the norm rather than the exception. Focusing on identifying shared relationships or features across orthologs can gloss over cryptic, ortholog-specific features that might have a direct bearing on function. To bridge the gap, we next asked if we could infer the presence of non-random sequence features that are likely to be ortholog-specific signatures. If such features exist, then it would follow that the combination of non-random features and conserved SERs together contribute to maintenance of functions and to adaptations that enable IDRs to function in different bacteria. Non-random sequence patterns might be a way to impart modular functional regulation onto IDRs whereby a motif becomes the site for epigenetic control or coordinating a network of protein-protein interactions.

In this chapter, we develop methods to extract non-random features within sequences of IDPs / IDRs. To do so, we impose constraints that enable the comparison of sequence patterns across large numbers of randomly generated sequence variants with fixed amino acid compositions. Each naturally occurring IDP / IDR has a z-score associated with a specific patterning parameter. This helps us assess if the patterning parameter of interest is likely non-random. We suggest that this methodology can be used to uncover cryptic sequence features and to guide *de novo* sequence design of IDP / IDRs. In addition to the CTTs of FtsZs, we also deploy our method for identifying non-random sequence patterns in IDRs from two other families of bacterial proteins.

## 4.2 Introduction

Recent estimates suggest that approximately 40% of eukaryotic proteomes contain proteins that are either entirely disordered or include disordered domains [1, 2]. Conformational heterogeneity is the defining hallmark of intrinsically disordered proteins/regions (IDPs / IDRs). Under typical approximations of physiological solution conditions (pH 7.4, 37°C, and 100-300 mOsm of salts and solutes), IDPs / IDRs are best described as adopting heterogeneous ensembles of conformations, although some IDPs / IDRs can be coaxed into differently stable folds when bound to their binding partners [3]. Accordingly, we distinguish IDPs / IDRs from autonomously foldable proteins or domains (AFPs / AFDs), which adopt essentially singular and stable folds.

AFDs can be classified into distinct sequence families based on sequence-structure-function relationships [4, 5]. It is often true that folded proteins with similar sequences have similar structural preferences [6]. This paves the way for identifying sequence families and constructing sequence phylogenies based on multiple sequence alignments [4]. In contrast to AFDs, many IDPs / IDRs are characterized by hypervariability within sequence families [7-9]. While often used to describe genetic regions with high-degrees of variability [10], here we co-opt the term *hypervariability* to describe certain types of IDPs / IDRs, and we focus on those that feature prominently in bacterial proteomes. Hypervariable IDPs / IDRs are involved in similar functions across orthologs but are also characterized by significant differences in sequence features. As a result, multiple sequence alignments must incorporate numerous gaps to maximize sequence similarity and identity within an alignment [11, 12]. These gaps lower the statistical significance of an alignment because the likelihood of realizing an alignment produced by hypervariable sequences can become equivalent to the likelihoods produced by alignments of random sequences.



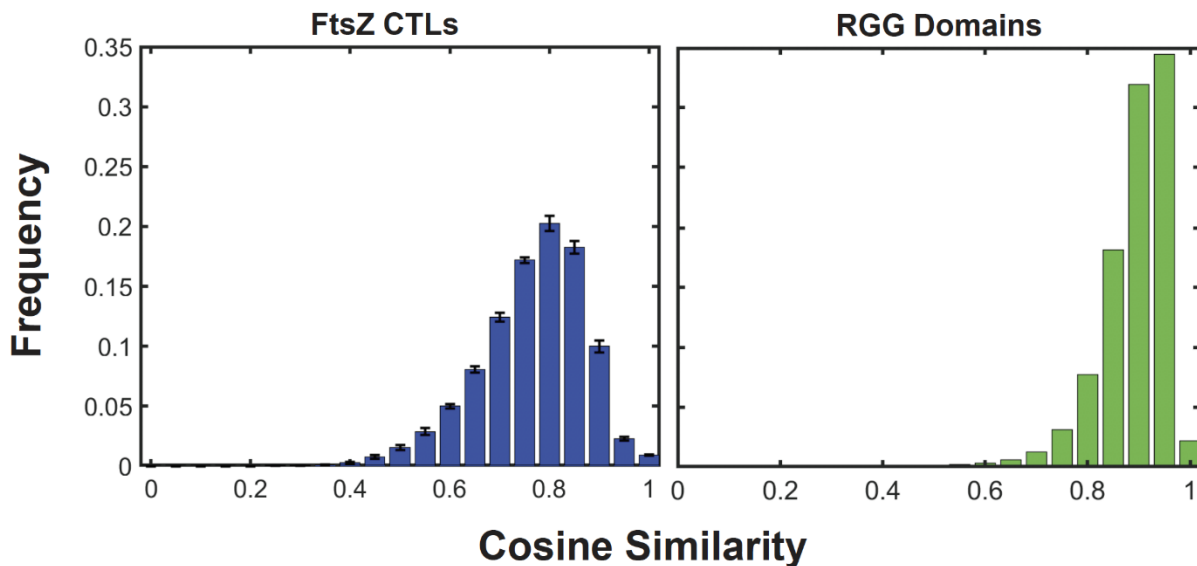
Rather than performing multiple sequence alignments, the compositions of a set of non-conserved IDPs / IDRs have been compared to identify conserved compositional biases and further assess the potential implication of that conserved feature on the function [13]. Conserved parameters that impact function include compositional biases of amino acids [14], the lengths of sequences [15], the net charge per residue [16, 17], and the number of occurrences of specific short linear motifs (SLiMs) [18-23]. One example of a family of IDRs with a well-studied composition-to-function relationship is the disordered RGG domain found in proteins that drive phase separation [24-26]. While the actual sequences cannot be aligned without numerous gaps, the compositional profiles are fairly similar across this family of sequences [25].

To quantify compositional similarities across a set of sequences, one can reduce each protein sequence to a 20×1 compositional vector, where each component is the fractional composition of each amino acid. The similarity between a pair of compositional vectors  $\mathbf{C}_i$  and  $\mathbf{C}_j$

is calculated as:  $\cos\theta_{ij} = \frac{\mathbf{C}_i \times \mathbf{C}_j}{|\mathbf{C}_i| |\mathbf{C}_j|}$  (method adapted from [27]). If  $\cos\theta_{ij} = 1$ , then sequences  $i$  and  $j$

have identical amino acid compositions. For pairs of sequences with totally different compositions,  $\cos\theta_{ij}$  will be close to zero. Given a set of sequences for IDPs / IDRs, we can compute the distribution of pairwise compositional (dis)similarities by quantifying the distribution of  $\cos\theta_{ij}$  values, calculated for each pair of sequences  $i$  and  $j$  in the set. We applied this method to interrogate (dis)similarities in amino acid compositions across the set of C-terminal linkers (CTLs) of bacterial cell division protein FtsZ. Previous studies suggest that the CTLs derived from orthologous FtsZs are hypervariable [9, 11, 28-32].

Using the 1208 C-terminal tail (CTT) sequences from FtsZ identified previously [31], we focused on the sequences of CTLs and ignored the CTP / sticker motifs [32, 33]. The distribution of pairwise cosine similarities of FtsZ CTLs is shown alongside the pairwise similarities calculated for a set of 407 RGG domains extracted from the human nucleolar proteome (**Figure 4.1**). For RGG domains, the pairwise similarities are typically greater than 0.8, whereas the CTLs from FtsZs have a wider distribution spanning from 0.4 to 1. This discrepancy is not a result of the disparate number of sequences between the RGGs and the CTLs; in this analysis, we extracted 407 sequences randomly from the complete set of FtsZ CTL sequences, computed the pairwise similarities, and repeated the process  $10^2$  times. Accordingly, IDRs such as the CTLs from FtsZs are hypervariable because they belong to the same functional family but diverge in linear order and composition profiles.



**Figure 4.1: Histograms of the cosine similarities of the vectorized compositions of the FtsZ CTLs (left) in comparison to human nucleolar RGG domains (right).** The height of each bar represents the frequency of observing a cosine similarity between two sequences within the binned interval. Each interval encompasses all values  $\pm 0.05$  of its mid-point. The FtsZ CTL bar plot (left) contains error bars, showing the deviations realized from the bootstrapping analysis.

In contrast to composition-based parameters, context-dependent parameters change with the linear sequence. These parameters include binary patterning parameters that quantify either the segregation or clustering of a specific residue / residue type with respect to another residue / residue type or all other residues in the sequence. Previous studies have demonstrated the functional importance of context-dependent parameters through mutational studies where sequence compositions were fixed, and variations to the linear sequence patterns altered the context-dependent parameters and resulted in perturbed functions [11, 23, 34-37]. These findings have prompted comparisons of context-dependent features within a set of hypervariable sequences as a route to inferring evolutionarily conserved functions. For instance, Buske et al., observed that despite having different compositions, the patterning of oppositely charged residues, quantified in terms of the parameter ( $\kappa$ ), was bounded between 0.15 and 0.4 for the CTLs of FtsZs [31]. This observation led to the conjecture that the bounds on  $\kappa$  might be indicative of functional significance.

Here, we first assessed the validity of comparing sequence patterning parameters across sequences with different compositions and lengths. Details of this analysis are presented in the Appendix to this chapter. Most importantly, we find that the likelihood of observing a specific context-dependent parameter is dependent upon the composition and length of the sequence. Therefore, accurate comparisons of context-dependent sequence parameters require a robust approach that accounts for differences in sequence composition and length. This requires two advances, which we have developed here: The first is the use of a suitable null model that allows us to assess the significance, i.e., extent of non-randomness of a parameter of interest. Second, by ensuring that biases due to composition and length are resolved through the use of an appropriate

null model, we use the method of z-scores to annotate each sequence in terms of patterning specific z-scores. This allows us to uncover cryptic, non-random sequence patterns for each ortholog and group orthologs by shared non-random sequence patterns. We propose that the presence of one or more non-random patterns within an IDR could highlight the potential functional importance of these patterns. This work represents a generalization of the approaches brought to bear for identifying SLiMs [7, 19, 20, 38]. In addition to generating hypotheses regarding functionally important sequence patterns within hypervariable IDRs, our work, if it stands up to the scrutiny of system-specific assays, could imply that there are considerably more than a million motifs across “IDRomes” [39].

## 4.3 The z-score method

### 4.3.1 Null model generation



**Figure 4.2: Z-score method workflow.** The process includes generating the null-scramble model (“null model”) and calculating the deviation of the observed value from the null model.

The workflow to generate the null model is summarized in **Figure 4.2** and is as follows. For each sequence of interest, we generate  $10^5$  distinct scrambled sequences wherein the composition is fixed, and the linear sequence is varied at random. The patterning parameter of interest is calculated for each randomly generated sequence in the set of scrambles. Irrespective of the parameter of interest, the random manner of generating sequence scrambles leads to maximum entropy-based gamma distributions, as shown in **Figure 4A.1** of the Appendix. The mean ( $\mu$ ) of the gamma distribution is the null expectation for the patterning parameter of interest. For the relevant parameter of the sequence of interest, the z-score ( $z$ ) represents the deviation of the observed value from  $\mu$  (**Equation 4.1**). The z-score quantifies the direction of the deviation

(positive or negative) and the number of standard deviations ( $\sigma$ ) that the observed patterning parameter ( $x$ ) is away from the expected value ( $\mu$ ).

$$z(x) = \frac{\mu - x}{\sigma} \quad \text{Equation 4.1}$$

For each set of hypervariable sequences, we define a z-score threshold value to delineate random from non-random sequence features. If the magnitude of the z-score is less than the threshold value, the patterning feature is labeled as being random. Conversely, if the magnitude of the z-score exceeds the threshold value, then the patterning feature of interest is labeled as being non-random.

### 4.3.2 Calculation of patterning parameters

*Kappa*: The kappa ( $\kappa$ ) value for a given sequence quantifies the degree of linear segregation within a sequence between two specific residues or two types of residues, denoted as r1 and r2 (**Equations 4.2 and 4.3**) [40]. In a sequence with where the fraction of residues of types r1 and r2 are  $f_{r1}$  and  $f_{r2}$ , respectively, we start by computing an asymmetry parameter defined as:

$$\sigma_{r1r2} = \frac{(f_{r1} - f_{r2})^2}{(f_{r1} + f_{r2})} \quad \text{Equation 4.2}$$

As previously described [40, 41],  $\kappa$  is essentially the mean squared deviation of local asymmetry values, calculated over sequence windows labeled  $i$  ( $\sigma_{r1r2,i}$ ), and the global asymmetry ( $\sigma_{r1r2}$ ) (**Equation 4.3**). A local window corresponds to the size of a thermal blob, which spans 5 - 6 amino acids. Previously the  $\kappa$  value for a particular sequence was normalized by the maximum possible mean squared deviation. However, this becomes moot here since  $\kappa$  turns out not to be directly

comparable across different compositions. Accordingly, we repurpose  $\kappa$  to refer to unnormalized values.

$$\kappa = \frac{\sum_{i=1}^{N_{windows}} (\sigma_{r1r2,i} - \sigma_{r1r2})^2}{N_{windows}} \quad \text{Equation 4.3}$$

*Omega:* Instead of quantifying the patterning of a pair of residue types with respect to one another, one can quantify the linear patterning of one class of residues or one residue type with respect to all other residues or residue types. This one-against-all patterning parameter is denoted as omega ( $\Omega$ ). As defined below, it quantifies how a specific residue or type of residue ( $r$ ) within a sequence is positioned (mixed or segregated) with respect to all other residues / residue types. Accordingly, the relevant asymmetry is calculated as follows (**Equation 4.4**) [37]:

$$\sigma_{rX} = \frac{(f_r - f_{X \neq r})^2}{(f_r + f_{X \neq r})} \quad \text{Equation 4.4}$$

The  $\Omega$  value is also calculated by comparing local asymmetry to global asymmetry to determine the overall patterning (**Equation 4.5**).

$$\Omega = \frac{\sum_{i=1}^{N_{windows}} (\sigma_{rX,i} - \sigma_{rX})^2}{N_{windows}} \quad \text{Equation 4.5}$$

### 4.3.3 Development of the z-score matrix

The z-score matrix is a set of z-scores that includes quantifications of a set of pairwise patterns and one-against-all patterning parameters. These are computed for specific combinations of residues / residue types. The eight residues / residue types used in this work were chosen based on unique stereochemical properties and / or distinctive functional groups. They include: polar ( $\mu$ )

(S,T,N,Q,C,H), hydrophobic (h) (I,L,M,V), positively charged or basic residues (+) (R,K), negatively charged or acidic residues (-) (E,D), aromatic ( $\pi$ ) (F,W,Y), alanine (A), proline (P), and glycine (G).

The organization of a typical z-score matrix is shown in **Figure 4.3**, where the residues or residue types are labeled along the axes. We use an 8×8 matrix where only the upper triangular values are considered, though the matrix size can change as we change the numbers of residues / residue types of interest. The off-diagonal elements in the matrix are  $\kappa$ -values that quantify the linear mixing versus segregation of pairs of residues / residue types. When sequences contain fewer than 12% of a residue or residue type, all z-scores that involve that residue type are set to 0. This is further explained in the Appendix (**Figure 4A.2 and Figure 4A.5**).



$\mu$	$\Omega_{\mu}$	$\kappa_{\mu h}$	$\kappa_{\mu +}$	$\kappa_{\mu -}$	$\kappa_{\mu \pi}$	$\kappa_{\mu A}$	$\kappa_{\mu P}$	$\kappa_{\mu G}$
$h$		$\Omega_h$	$\kappa_{h +}$	$\kappa_{h -}$	$\kappa_{h \pi}$	$\kappa_{h A}$	$\kappa_{h P}$	$\kappa_{h G}$
$+$			$\Omega_+$	$\kappa_{+-}$	$\kappa_{+ \pi}$	$\kappa_{+ A}$	$\kappa_{+ P}$	$\kappa_{+ G}$
$-$				$\Omega_-$	$\kappa_{- \pi}$	$\kappa_{- A}$	$\kappa_{- P}$	$\kappa_{- G}$
$\pi$					$\Omega_{\pi}$	$\kappa_{\pi A}$	$\kappa_{\pi P}$	$\kappa_{\pi G}$
$A$						$\Omega_A$	$\kappa_{AP}$	$\kappa_{AG}$
$P$							$\Omega_P$	$\kappa_{PG}$
$G$								$\Omega_G$
	$\mu$	$h$	$+$	$-$	$\pi$	$A$	$P$	$G$

**Figure 4.3: Depiction of a z-score matrix with eight residues / residue types.** Here, polar ( $\mu$ ) (S,T,N,Q,C,H), hydrophobic ( $h$ ) (I,L,M,V), positive ( $+$ ) (R,K), negative ( $-$ ) (E,D), aromatic ( $\pi$ ) (F,W,Y), alanine (A), proline (P), and glycine (G) are considered for unique properties.

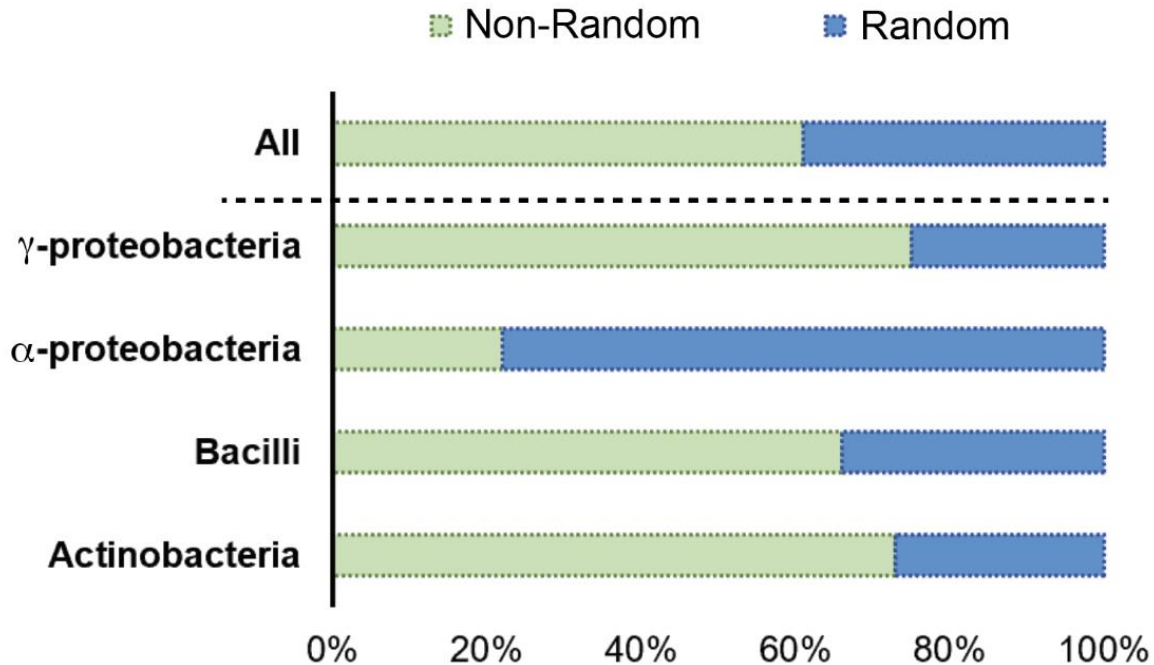
## 4.4 Application of the z-score method

### 4.4.1 Application of the z-score matrix method to delineate random and non-random patterning features

Patterning parameters cannot be compared amongst sequences with varying compositions and lengths (see Appendix, **Figures 4A.2 – 6**) [41]. Therefore, we used the z-score method (**Section 4.3**) to quantify deviations of a series of patterning parameters from the expected values based on a null model to assess if the parameter should be classified as being random or non-random. Sequences that do not have non-random values for any of the patterning parameters are considered entirely random. For fully random sequences, the entire z-score matrix (**Section 4.3.3**) will have z-score values less than the absolute value of the threshold value.

Using the 1208 CTL sequences from FtsZs, we calculated the maximum z-score magnitude (absolute value) for each CTL. We used this information to quantify the dependence of the fraction of random sequences on the z-score threshold value. The resulting cumulative distribution function (CDF) is shown in the Appendix (**Figure 4A.7**). For each z-score magnitude, the CDF quantifies the fraction of sequences that will be considered random if that value was set as the z-score threshold. For example, a threshold value of 1 will classify 90% of sequences as having non-random patterning for at least one feature of interest. In contrast, a threshold value of 2 will classify 35% of sequences as having at least one non-random patterning parameter. We set 1.5 as the threshold such that approximately two-thirds of the CTL sequences have at least one non-random feature and are therefore considered non-random (65.6% versus 33.4% random) (**Figure 4.4**).

We previously showed that the CTL functions as a spacer in *B. subtilis*, separating the C-terminal peptide (CTP) sticker from the GTPase core domain. In *B. subtilis* FtsZ, the core promotes linear assembly and enzymatic activity, while the sticker promotes higher-order interactions that are imperative to forming the cytokinetic ring in bacteria [30, 32, 42]. The CTL spacer exhibits auto-regulatory behavior by impacting the core and the sticker [32]. Despite having disparate sequences, simulations of the CTL sequences showed that most CTLs had similar degrees of conformational heterogeneity [11]. And yet, mutational studies have shown that not all CTLs can be swapped for one another and function the same [29, 30, 43-45]. In this set of orthologous CTL sequences, 27 distinct phyla and 51 phylum classes were represented. Four phylum classes were enriched within the data set, each independently comprising over 9% of the total sequences. These four phyla were  $\gamma$ -proteobacteria (n = 192),  $\alpha$ -proteobacteria (n = 147), bacilli (n = 105), and actinobacteria (n = 122). **Figure 4.4** shows the class-specific deviations from the collective (All) split between random and non-random sequences. The bacilli class most closely mimics the breakdown of the collective set (~60/40). Strikingly,  $\alpha$ - and  $\gamma$ -proteobacteria have approximately opposite percentage splits of random and non-random CTL sequences. 78% of  $\alpha$ -proteobacteria CTL sequences are random, whereas 75% of  $\gamma$ -proteobacteria are non-random.



**Figure 4.4: Percentage of CTL sequences that have at least one non-random feature versus those that are fully random.** Two-thirds (~ 65.6%) of the CTL sequences assessed (n = 1208) were considered non-random. Phylum classes that contained at least ~9% of the sequences (105 /1208) were further analyzed to uncover any class-specific trends. These included  $\gamma$ -proteobacteria (75% non-random (nr) / 27% random (r); n = 192),  $\alpha$ -proteobacteria (22% nr / 78% r; n = 147), bacilli (66% nr / 34% r; n = 105), and actinobacteria (73% nr / 27% r; n = 122).

In addition to the absolute value of the z-scores, the sign of the z-score also contains pertinent information. If a sequence has a patterning parameter that deviates from the expected value in the negative direction, the residues of interest are more well-mixed with respect to one another or the rest of the sequence than expected. Conversely, if a sequence has a patterning parameter that deviates in the positive direction, the residues are more segregated from one another than expected. Therefore, positive deviations indicate a blockier sequence architecture than expected, and negative deviations indicate a more well-mixed sequence as far as the patterning parameter is concerned.

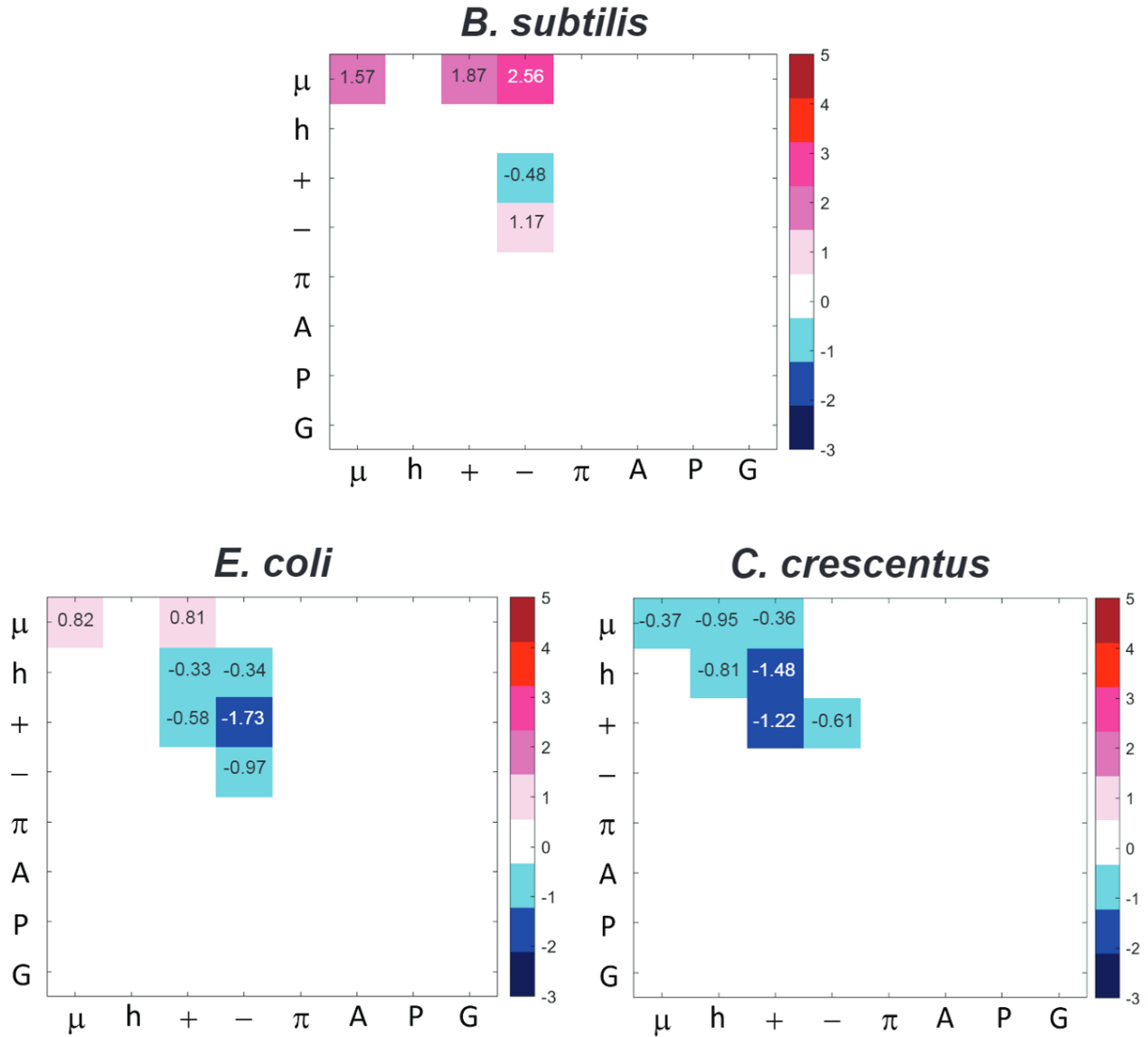
#### 4.4.2 Using the z-score matrix method to connect sequence-encoded information to IDR function and bacterial phenotype

Further analysis of the CTL sequences from three well-studied systems, bacilli *B. subtilis* [30-32],  $\gamma$ -proteobacteria *E. coli* [29], and  $\alpha$ -proteobacteria *C. crescentus* [28, 44-46], showed that the sequences not only diverge in terms of patterns that are random versus non-random but also in the significance of different sequence patterns within each of the CTLs (**Figure 4.5**). In the *B. subtilis* CTL, the relevant non-random patterning corresponds to the segregation of polar and charged residues, specifically negatively charged residues, with respect to one another. The parameter  $\kappa_{\mu-}$  shows a significant positive deviation from the random expectation value.

The  $\kappa$  value ( $\kappa_{+-}$ ) for the *E. coli* CTL is much lower than the randomly expected value, implying that these residues are more well-mixed than would be expected at random. With the threshold z-score value set at 1.5 to quantify non-randomness, all of the patterns we analyzed in the *C. crescentus* CTL show insignificant deviations from their values that are expected at random. This is seen from the fact that the absolute values of the z-score matrix elements do not exceed 1.5. However, there is a tendency toward well-mixed patterns of basic and hydrophobic residues, as evidenced by the nearly non-random value for  $\kappa_{+h}$ . The implication is that linear clusters of hydrophobic residues are avoided by interspersing basic residues between these groups.

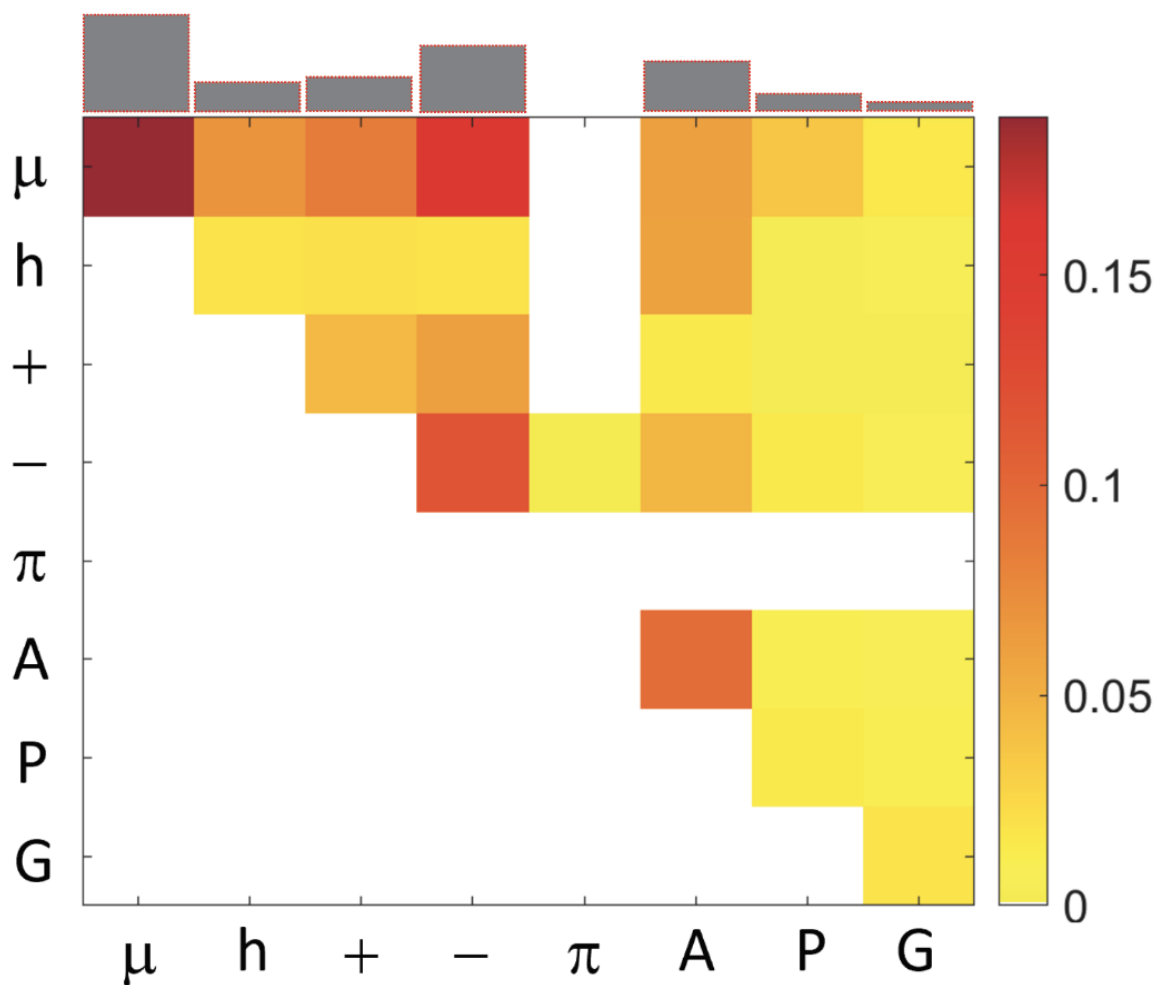
Overall, the patterning analysis applied to CTLs and summarized above suggest that each system appears to follow its own set of rules for sequence patterning features that are random versus non-random. In this context, it is worth noting that mutational studies that swap CTL domains have shown evidence for apparent specificity that is concordant with the findings from

our analysis that identifies random versus non-random patterns in CTLs of FtsZs. For example, the CTLs of *C. crescentus* and *E. coli* are not functionally interoperable with one another [29] [28]. Likewise, replacing the CTL of *B. subtilis* FtsZ CTL with that from *E. coli* FtsZ alters the functions of FtsZ *in vitro* [30]).



**Figure 4.5: Z-score matrices of the FtsZ CTLs from *B. subtilis*, *E. coli*, and *C. crescentus*.** A color bar is used to illustrate the z-score values. The values are calculated as described in the development of the z-score matrix section (Chapter 4.3.3). Features were considered random if the z-scores were less than 1.5 and greater than -1.5 ( $-1.5 \leq z \leq 1.5$ ).

We assessed which, if any, of the CTLs might be representative of the full set of FtsZ orthologs by analyzing the conservation of non-random features. Counting each feature that deviates significantly ( $z < -1.5$  or  $z > 1.5$ ), we find that over 80% of the deviations are in the positive direction. Given the significant representation of this category, we focused our subsequent analysis on CTL sequences with non-random features that have positive deviations from the expected values. Across the complete set of z-score matrices, we tallied the number of positive deviations ( $z > 1.5$ ) observed for each parameter. To compute a positive deviation frequency for each feature, we divided the number of positive deviations by the total number of interrogated sequences. **Figure 4.6** shows the frequency of non-randomness for each feature. The bar graph on top demonstrates the relative frequency of observing a positive deviation involving each residue / residue type. Aromatic residues are underrepresented in FtsZ CTL sequences (< 5% of total residues) and are not considered in this analysis. The total frequencies of each amino acid for the CTLs are shown in the Appendix (**Figure 4A.8**). The prominent non-random feature that emerges is that polar and negatively charged residues tend to be more segregated from one another in the linear sequence than would be expected at random. This is a feature we identified for the CTL of Bs-FtsZ as well. This analysis summarized in **Figure 4.6** helps identify features that may have functional implications. Mutational studies based on the design of chimeras or the design of consensus sequences will likely help in assessing if the ortholog-specific non-random sequence patterns within FtsZ CTLs have any functional significance.



**Figure 4.6: Frequency of observing z-scores above 1.5 for each assessed feature.** Features that are light yellow are rare ( $< 0.05$ ), and features that are deep red are relatively common ( $> 0.15$ ). A bar graph is placed on top of the matrix to represent the relative frequency of observing z-score values greater than 1.5 for features involving each residue / residue type. The height order of the relative bar graph and, therefore, the order of significance is polar, negative, alanine, positive, hydrophobic, proline, glycine, then aromatic residues.

### Analysis of IDRs in RNases E

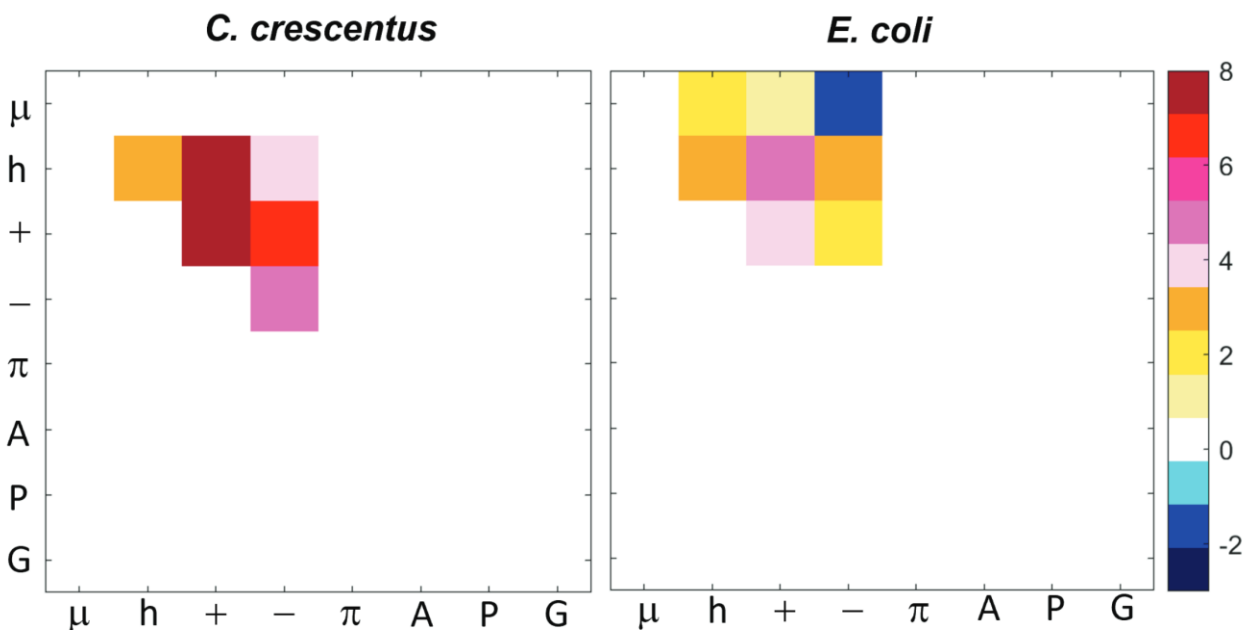
Next, we applied the z-score methodology to a system that contains a disordered domain with a patterning feature that is known to have functional implications. In bacteria, the protein RNase E is a critical driver of the formation of the RNA degradosome [47-49]. The architecture of this protein includes a conserved DEAD-box RNA helicase and a disordered C-terminal domain



(CTD). In *C. crescentus*, the RNase E CTD is necessary and sufficient to drive phase separation, similar to the previously mentioned RGG domain. *In vivo*, RNase E drives the formation of cytoplasmic foci that colocalize with other exonucleases. This degradation body has been termed Bacterial Ribonucleoprotein body or BR-body [49]. The CTD of the *C. crescentus* RNase E has a blocky patterning of oppositely charged residues [48]. The blocky nature leads to encoded multivalence of opposite charges [50], and this architecture is essential for the formation of BR bodies [49]. In contrast, the *E. coli* RNase E, which lacks the blocky patterning of oppositely charged residues, does not form cytoplasmic condensates; instead, it forms membrane-tethered puncta and does not exhibit liquid-like behavior *in vitro* under the solution conditions that have been investigated to date [47-49]. Sequences of each CTD are shown in the Appendix (**Figure 4A.9**).

Are there distinct sequence patterns that distinguish the CTDs of RNases E from *C. crescentus* and *E. coli*? To answer this question, we performed the z-score analysis on the CTDs from *C. crescentus* and *E. coli* to test whether our method captures the previously noted CTD sequence architectures that result in different phenotypes (**Figure 4.7**). To set the z-score threshold, we used the same rationale used in our analysis of patterning features within the CTLs of FtsZs. Specifically, we set the threshold such that approximately 60% of sequences would be considered non-random. The resulting cumulative distribution is shown in the Appendix (**Figure 4A.10**), and the threshold z-value was set to 2.0. Like the CTL from FtsZ, most z-scores that fall outside the threshold value bounds are positive (90% of 2041 non-random z-scores). We find that 62% of RNase E CTDs contains at least one feature that has a z-score  $> 2$ . Basic residues are most frequently involved in the non-random patterning parameters. The preference for separation of

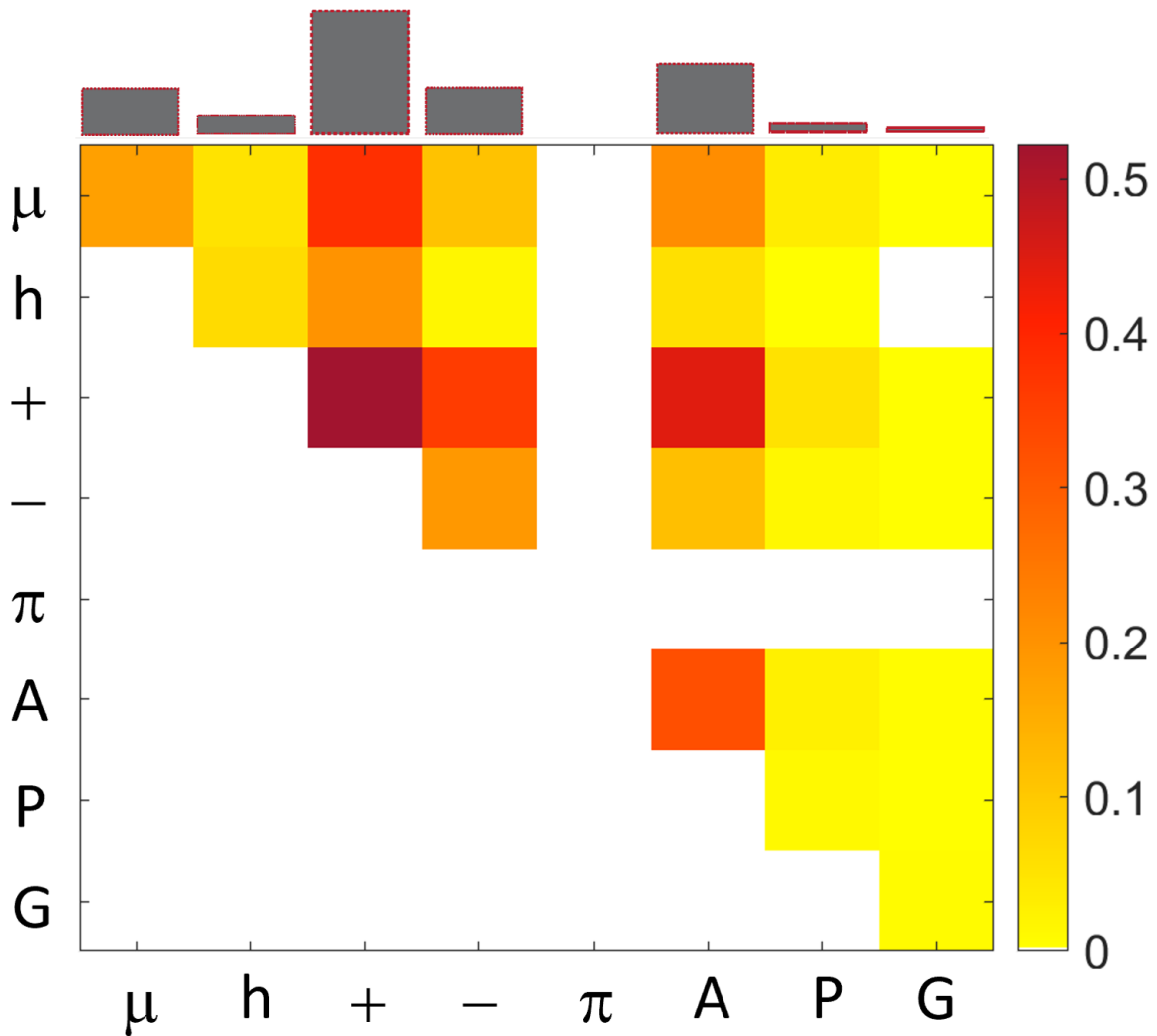
positive charges from the rest of the residues ( $\Omega_+$ ) is one of the dominant features across the system of RNase E orthologs (**Figure 4.8**).



**Figure 4.7: Direct comparison of z-score matrices from *C. crescentus* RNase E and *E. coli* RNase E.** Patterns associated with charges in *C. crescentus* RNase E (left) are  $> 4$  standard deviations away from the null-scramble model in the positive direction. *E. coli* RNase E (right) shows non-random segregation of positive residues from other residues, and hydrophobic residues also contribute to non-random patterns. Unlike the *C. crescentus* RNase E, patterns involving negative residues do not significantly deviate from the null-scramble expectation.

To gain specific insights regarding patterning correlates of phenotypic differences across orthologs, we focused on quantifying the distinctions in the patterning features within CTDs of RNases E from *C. crescentus* and *E. coli*. We find that acidic and basic residues are significantly segregated within the sequence of the CTD from *C. crescentus* RNase E. The z-scores of  $\Omega_-$ ,  $\Omega_+$ , and  $\kappa_{+-}$  are above 4.5, and this indicates that acidic and basic residues are segregated from each other and the rest of the sequence. In descending order, the most significant z-scores are  $\kappa_{+-}$ ,  $\Omega_+$ ,  $\kappa_{h+}$ ,  $\Omega_-$ ,  $\kappa_{h-}$ , and  $\Omega_h$ , respectively. These features point to a clear preference for the linear

segregation of acidic, basic, and hydrophobic residues into distinct linear clusters along the CTD (**Appendix Figure 4A.9**). Accordingly, the CTD of the *C. crescentus* RNase E may be classified as a blocky polyampholyte that also includes clusters of hydrophobic residues. Comparative analysis shows that the extent of non-randomness of the distinct patterns listed above is diminished for the CTD of *E. coli* RNase E. Lowering of the extent of linear segregation of acidic, basic, and hydrophobic residues is evident from the increased importance of the segregation of basic and hydrophobic residues with respect to polar residues (see values of  $\kappa_{\mu+}$  and  $\kappa_{\mu h}$  in the panel for *E. coli* in **Figure 4.7**). The overall weakening of the mutual linear segregation of charged and hydrophobic residues appears to correlate with the observation of distinct phenotypes and driving forces for CTD mediated phase transitions of RNases E that control the formation of BR bodies. Whether our conjectures have merit will have to be resolved through sequence design experiments geared toward understanding the phenotypic consequences of altering the extents of linear segregation / mixing of charged and hydrophobic residues within CTDs of RNases E.



**Figure 4.8: The feature-specific frequency of observing z-scores above 2 for the system of RNase E orthologs.** 1149 CTDs from a system of RNase E were scrambled to generate the z-score matrices. All z-scores greater than 2 were tallied and divided by the total number of CTD sequences to calculate a frequency of observing a positive non-random deviation for each sequence feature. The bar graph displays the relative frequency of a non-random feature involving each residue / residue type. Z-scores involving positive residues are most frequently above 2.

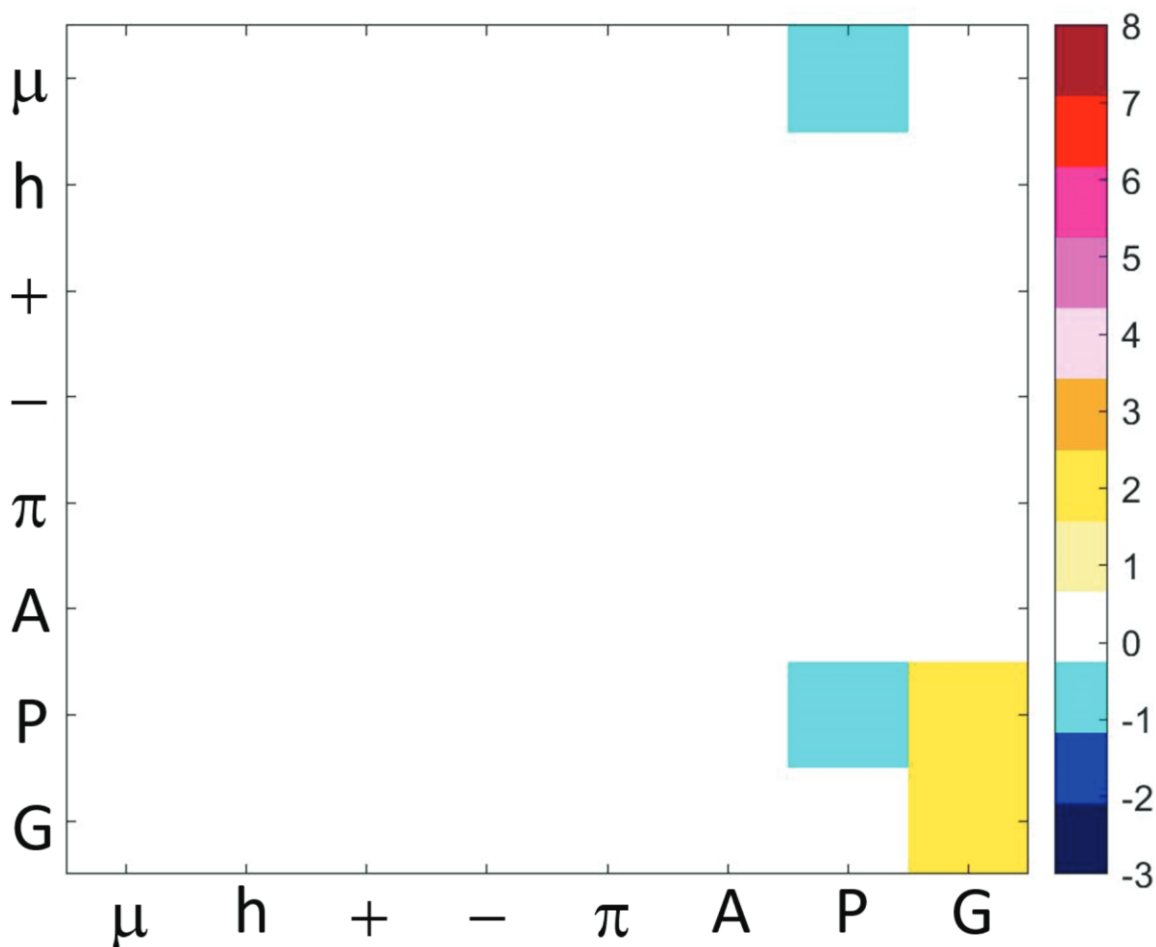
## Analysis of IDRs in bacterial single-stranded DNA binding proteins

Another example of an essential IDR in bacteria is the intrinsically disordered linker (IDL) in single-stranded DNA binding proteins (SSBs) that play critical roles in bacterial DNA replication and repair. Their modular architecture, which is reminiscent of FtsZ, includes an ordered DNA-binding domain (OB-fold), followed by a hypervariable intrinsically disordered linker (IDL / spacer) that is connected to a conserved C-terminal tip (sticker) [51-53]. Recent work has also shown that the SSB from *E. coli* can mediate phase separation with DNA [54]; however, unlike RNase E, the sequence requirements for this function have yet to be elucidated. Using the z-score methodology, we sought to generate testable hypotheses regarding sequence features that are likely to distinguish functional IDLs from random ones.

We find that for *E. coli* SSB IDL, the distinctly non-random feature pertains to the linear segregation of Gly from all other residues. Gly residues form a series of short linear clusters, segregated from all other residues in the SSB IDL sequences. The blocks of Gly residues are often interspersed by Pro, giving rise to positive deviations of  $\kappa_{PG}$  from the null model (**Figure 4.9**). These features are highlighted in the sequence of the IDL from *E. coli* SSB: TMQML**GG**RQ**SGG**APAG**GG**NI**GGG**QPQ**GGW**GPQPQPQ**GG**NQFS**GG**AQSRPQQSAPAAP**SNEPP**.

Cursory analysis suggests that the *E. coli* SSB IDL has many of the features that are reminiscent of elastomeric IDRs that are known to be drivers of responsive phase transitions that control elastic responses of materials such as extracellular matrices [55-58]. Examples of such elastomeric IDRs include Gly-rich regions within resilin, as seen with repeats of PSSSYGAP**GGGN**GG**R** that confer elastic properties to resilin. Other examples include stretches such as P**GGG**QQ from Q-rich proteins such as gluten [59], Gly, and Ser rich motifs in silk[60],

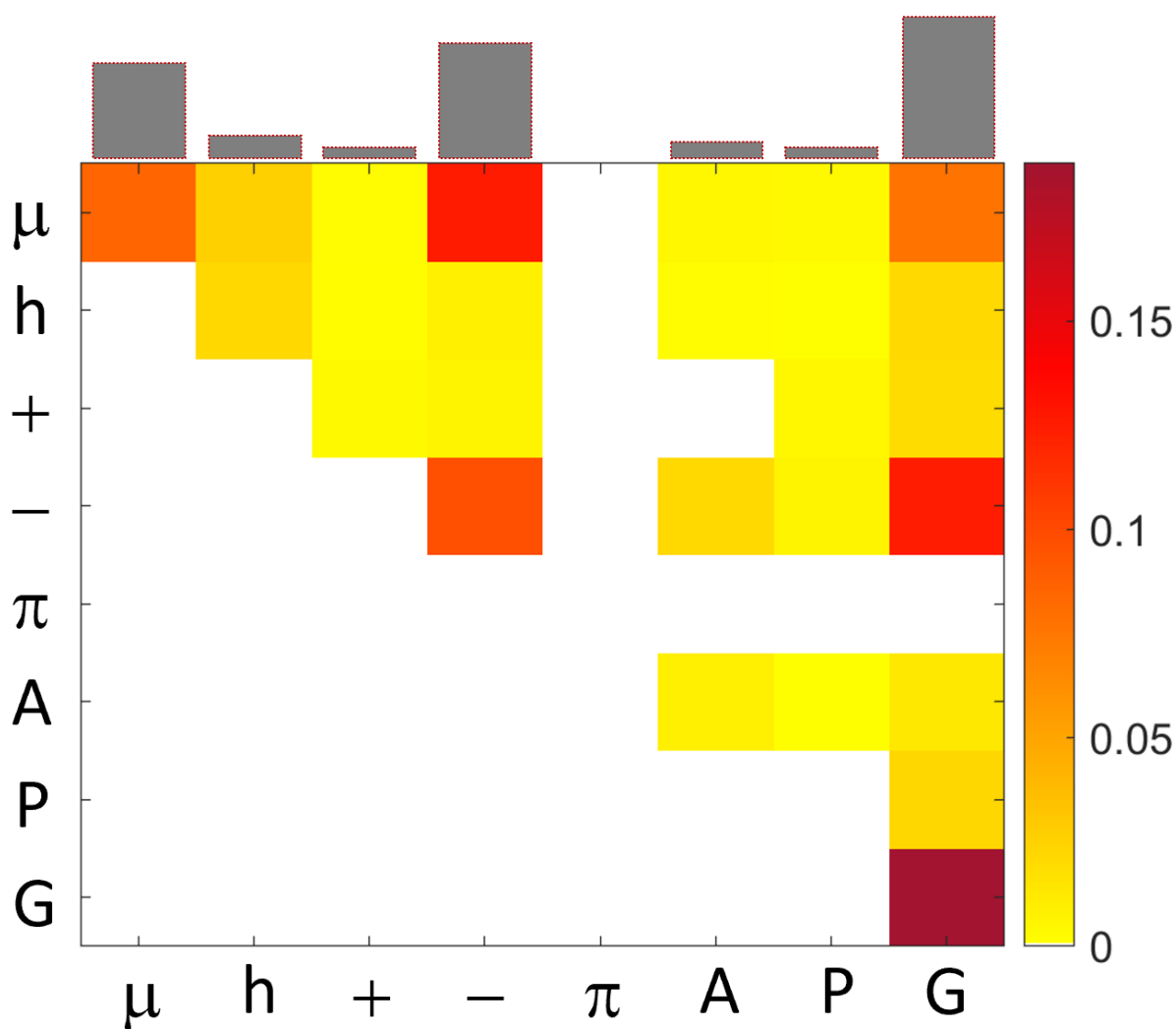
motifs such as YGHGGN/G in cell wall proteins of higher plants [61], and repetitive motifs such as FGGMGGGKGG from abductin, a protein that makes up hinge ligaments that control the swimming behaviors of mollusks [62].



**Figure 4.9: *E. coli* SSB CTD z-score matrix shows glycine and proline positioning is significant.** Matrix elements are colored corresponding to the z-score value.

We assessed whether non-random segregation of Gly residues is a conserved feature across a system of 1523 orthologs. Using the 60% cut off for identifying sequences with non-random patterning features, we set the z-score threshold value to 2.6. The cumulative distribution of the magnitude of maximum z-score values is shown in the Appendix (**Figure 4A.11**). Only positive

deviations were considered as 99% of the observed deviations were in the positive direction. As with the *E. coli* SSB IDL, we find a recurring theme (see **Figure 4.10**) of non-random linear segregation of Gly and other polar residues from all other residues ( $\Omega_G$  and  $\Omega_\mu$ ) as well from one another (see values of  $\kappa_{\mu G}$ ). It also appears that the interplay between Gly residues and acidic or polar residues (specifically Gln) could be essential for SSB functions.

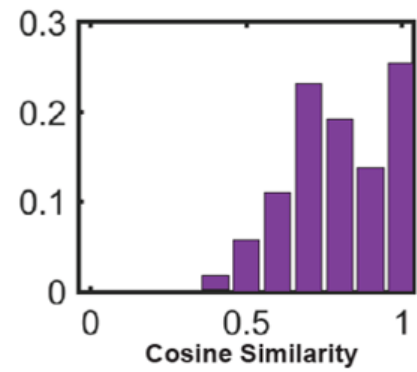
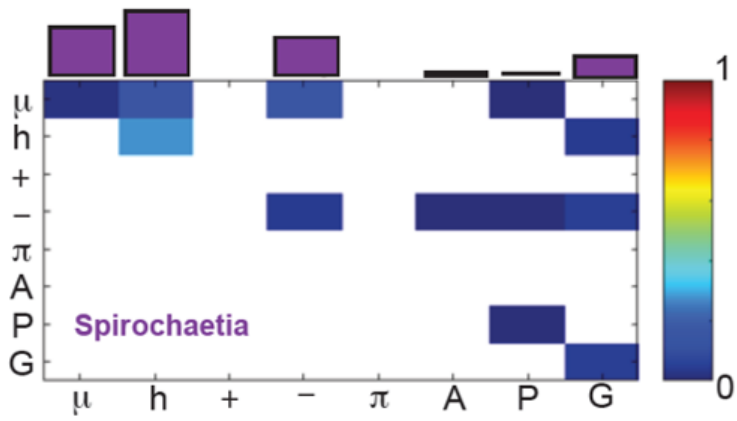
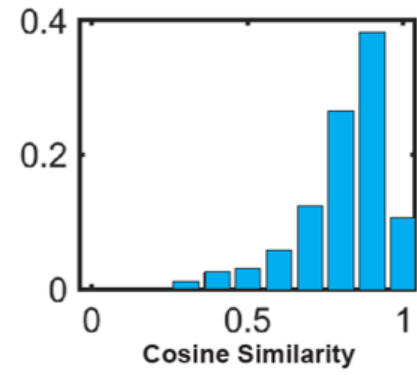
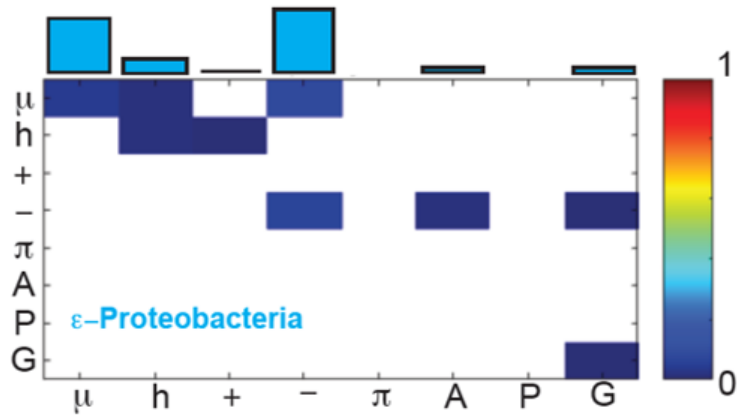
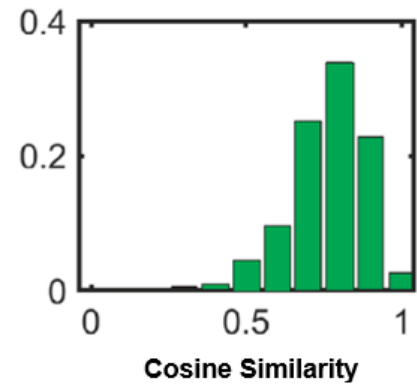
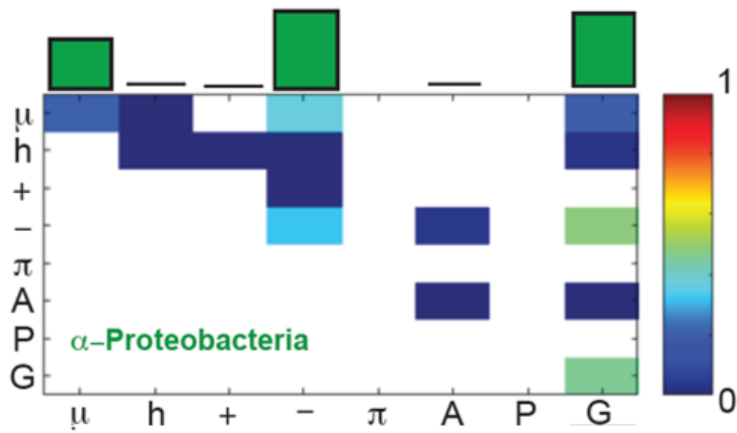


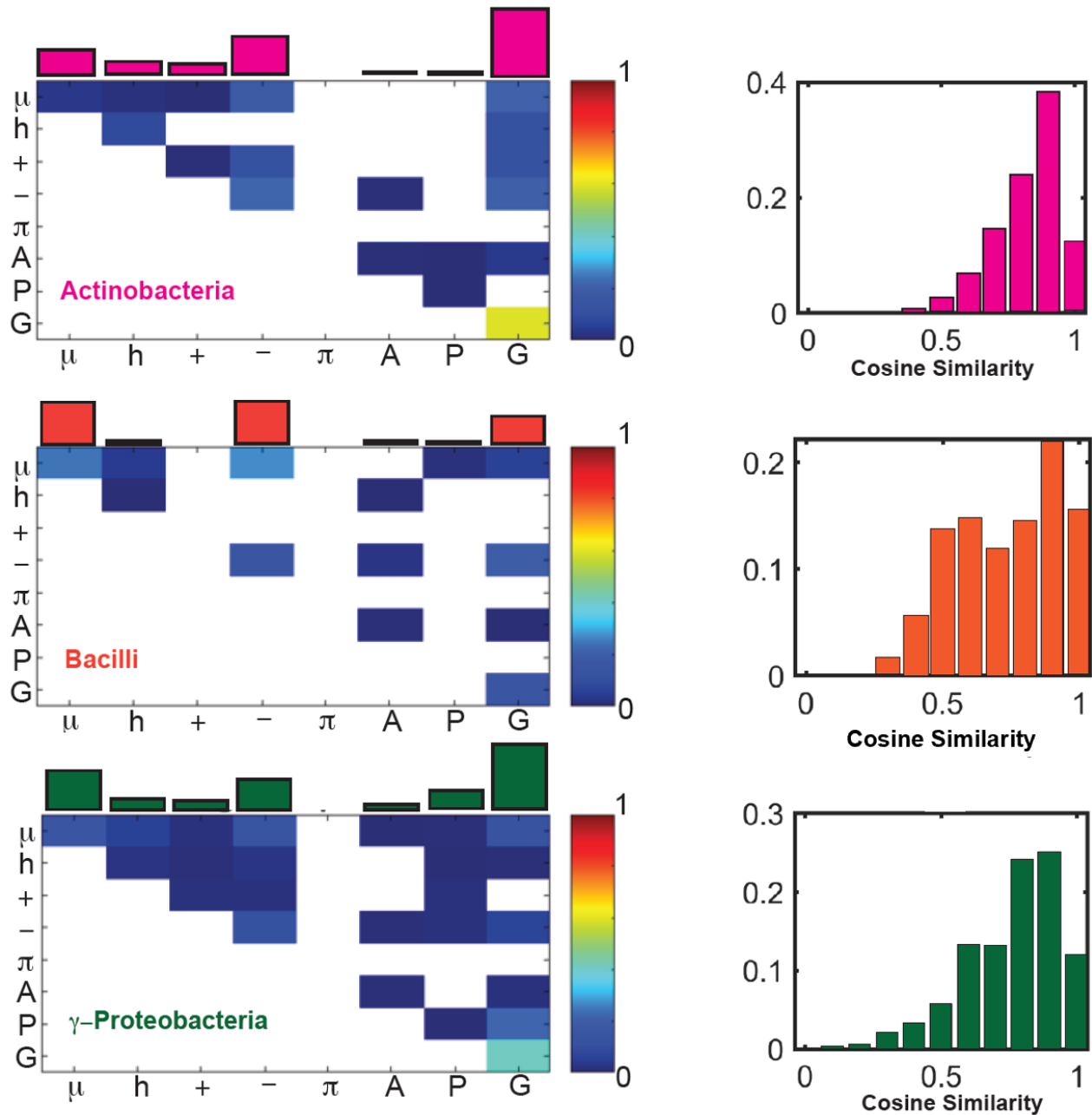
**Figure 4.10: The feature-specific frequency of observing non-random features for the system of SSB orthologs.** All z-scores greater than 2.6 were tallied and divided by the total number of IDL sequences to calculate a frequency of observing a positive non-random deviation for each sequence feature. The bar graph displays the relative frequency of a non-random feature involving

each residue / residue type. Z-scores involving glycine residues are most frequently above 2.6 and is dominated by three features: glycine, negatively charged residues, and polar residues.

Finally, we asked if the linear segregation of Gly and / or polar residues was a phylum- / class-specific feature. Classes that contained more than ~5% of the total number of sequences were analyzed separately (**Figure 4.11**). We confirmed that the sequences were still hypervariable by computing the pairwise cosine distances. Non-random patterning of Gly residues was conserved across four out of the seven evaluated classes, indicating that it is mostly conserved across different species, suggesting a positive selection pressure for this feature. The elastomeric features of SSB IDLs are noteworthy because they present testable hypotheses regarding the IDL-mediated phase behavior, which has been reported recently for *E. coli* SSBs [54]. How SSBs use such elastomeric features in the context of heterotypic networks of SSB-interacting proteins [53] remains entirely unclear and suggests a plausible route for investigation pending key preliminary experiments.







**Figure 4.11: The frequency of observing a non-random value for each assessed feature for the C-terminal SSB IDLs.** The data are organized by phylum class. The binned frequencies of the pairwise cosine similarities are included to the right of each non-random z-score frequency matrix. Eight phylum classes were analyzed: actinobacteria (n = 192), bacilli (n = 334),  $\gamma$ -proteobacteria (n = 362),  $\alpha$ -proteobacteria (n = 137),  $\epsilon$ -proteobacteria (n = 143), and spirochaetia (n = 107). As with the previous analysis, the bars on the top of the matrix represent the relative frequencies of observing a non-random feature ( $z > 1.5$ ) involving each residue / residue type.

## 4.5 Discussion

Here, we present a method to analyze a set of orthologous yet hypervariable intrinsically disordered domains. A conserved linear sequence is not required in IDPs / IDRs. This might be taken to mean that random IDRs can be functionally equivalent to the cognate IDRs. However, such conclusions could be erroneous if selected sequence features, specifically patterning features that are important for determining sequence-function relationships of IDRs, are masked by hypervariability. Here, we developed a method that is tailormade to uncover random versus non-random sequence patterns in hypervariable IDRs.

Binary patterning parameters measure how a residue, residue type, or motif is positioned within the sequence with respect to itself and other residues. An example of this type of parameter is  $\kappa_{+-}$  that quantifies the degree to which oppositely charged residues are segregated or mixed in the linear sequence. The conventional versions of  $\kappa$  and  $\Omega$ , generated to measure how segregated charged and proline residues were from the rest of the residues can be repurposed to interrogate any residue or residue type of interest. However, as previously discussed [41] and built upon in the Appendix, the values of these parameters cannot be compared across sequences with very different compositions because the value of the patterning parameter that would be expected at random is inherently dependent upon the composition of the sequence. To overcome this challenge, we use the deviation (z-score) of the observed patterning parameter from a null model. The random scrambling we perform is unbiased. It does not consider the weighted probability of observing such a sequence based upon factors such as codon usage and evolutionary deviation, which recent work aims to consider [8]. Instead, our method is a zeroth-order approximation of a

uniform background that allows for comparisons of hypervariable sequences in a statistically meaningful way.

Before applying this method, we first set a threshold z-score value to distinguish random and non-random features. This allows us to use the nomenclature of “random” versus “non-random” features to emphasize potentially significant patterns within the sequence. For the set of sequences analyzed in this work, we set the z-score threshold value such that approximately 60% of the hypervariable sequences have at least one non-random feature.

We used the z-score method to identify non-random sequence patterns within the hypervariable CTLs of FtsZs. We find that the fraction of CTL sequences presenting non-random features depends on the phylum / class (**Figure 4.4**). This implies that, while bacteria can be drastically different from one another, some features encoded within the CTL are likely to be evolutionarily conserved. For instance, the differences we observe between the  $\gamma$ - (mostly non-random) and  $\alpha$ - (mostly random) proteobacteria could imply that FtsZs from  $\alpha$ -proteobacteria are more likely to tolerate a swap for an entirely random CTL than FtsZs from  $\gamma$ -proteobacteria.

Analysis of three CTLs from well-studied FtsZs, *B. subtilis*, *C. crescentus*, and *E.coli*, reveals that these CTL sequences have specific features that diverge from the null model (**Figure 4.5**). In the CTL of Bs-FtsZ, the patterning of polar residues and charged residues is non-random. Analyzing the entire set of hypervariable CTL sequences shows that polar residues are most prominent as generators of non-random sequence patterns (**Figure 4.6**). This finding becomes particularly interesting in the context of observations of the involvement of CTL phosphorylation in function (conversations with Petra Levin and Sathya Nagarajan). This leads to questions about the potential roles for Ser / Thr phosphorylation and other post-translational modifications in CTLs

and assessments of how the context of the potentially phosphorylated residues impacts FtsZ function.

In the z-score matrix of the *E. coli* FtsZ CTL, we observe a non-random patterning of oppositely charged residues ( $\kappa_{+-}$ ). The negative z-score implies that these residues are more well-mixed in the linear sequence than expected based on the null model. While the exact role of the CTL in *E. coli* has yet to be elucidated, well-mixed oppositely charged residues within a disordered linker could ensure that there is an adequate excluded volume keeping the core and CTP apart. In the absence of other dominant features, we reason then that the *E. coli* FtsZ CTL, like the *B. subtilis* CTL, might also behave like a spacer.

Next, we used our z-score analysis method to examine the C-terminal domains (CTDs) from RNases E in *E. coli* and *C. crescentus* (**Figure 4.7**). When swapped with one another, these domains are known to impact function and the *in vivo* phenotype by affecting the nature of the RNA degradosome. In *C. crescentus*, this body is a biomolecular condensate that exhibits liquid-like properties, whereas, in *E. coli*, it is a membrane-bound punctum [49]. This discrepancy has been shown to be dictated by differences in the CTD sequence. In the z-score matrix of the *C. crescentus* CTD, the blocky charge architecture of this sequence is a feature that significantly diverges from the null expectation. While the positively charged residues of the *E. coli* CTD are non-randomly positioned, the negative residues are essentially random with regard to their position in relation to positively charged residues as well as in relation to other groups of residues. The non-random positioning of positive charges is a feature that is consistent with over 50% of RNase E CTDs (**Figure 4.8**). This could be a feature that is relevant for RNA binding within the degradosome. Comparing these sequences, it appears that the RNases E provide an example of

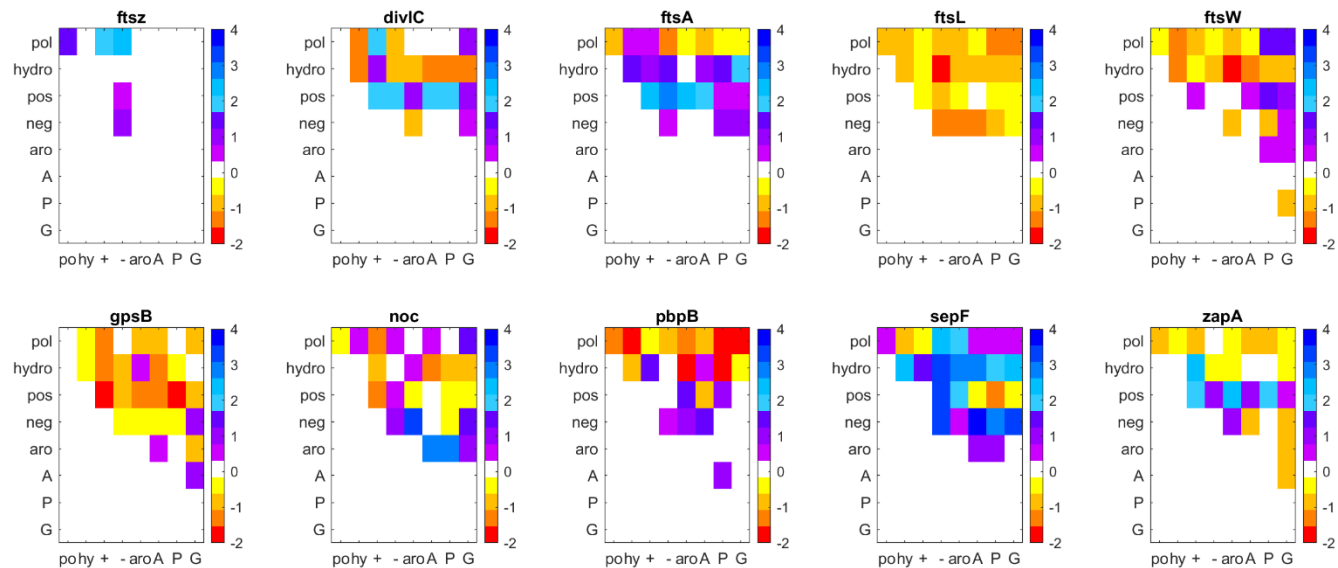
adaptation, whereby the charge patterning and other features within the CTD are the target of evolutionary shuffling to provide control over the formation of BR bodies.

For the SSB system, we observed a conserved non-random patterning of Gly residues whereby these residues tend to be found in short clusters along the linear sequence (**Figures 4.9 – 11**). We hypothesize that this feature could have implications for the cooperative binding of SSBs to single-stranded DNA [51-53] and for the driving forces for phase-separation of SSBs [54, 63-67]. SSBs form homo-tetramers that generate a tetra-valence of stickers to coordinate interactions with SSB interacting proteins [53, 66, 68]. The *E. coli* SSB tetramer binds cooperatively to single-stranded DNA, and the binding modes are classified by the number of nucleotides that are occluded by individual tetramers [63, 64, 66]. Cooperativity of single-stranded DNA binding is governed by sequence features of the IDL [51-53]. Specifically, cooperativity is enhanced when the IDL has features that are akin to low complexity domains enriched in polar amino acids, primarily proline, glutamine, and glycine [68, 69]. Conversely, cooperativity is diminished for long IDLs enriched in charged residues [51, 52]. The features that appear to govern cooperativity are also those that are non-random in the z-score analysis.

A recent study has shown that in response to DNA damage, membrane-associated SSBs form condensates at the sites of DNA damage [54]. These condensates are multicomponent bodies and concentrate other factors that contribute to DNA processing and metabolism. *In vitro* studies showed that the IDL is essential for driving the formation of liquid-like condensates. It could be that the patterning of Gly residues plays an important role in condensate formation – a feature that would be true of elastomeric sequences [58, 70].

Our analysis goes beyond the pair of binary patterning parameters considered to this point and introduces a general approach for identifying truly non-random binary patterns. The method allows the identification of non-random patterns within a sequence and comparative analyses of non-random patterns across sequences. This is aided by the use of sequence-specific z-score matrices. However, the picture that emerges may not be a complete description of the patterns that are either selected for or against within IDRs. Our analysis does not account for the fact that the IDRs invariably integrate the proteins of interest into more extensive protein-protein interaction networks. The selection pressures exerted by interaction networks exert and the covariation of sequence patterns among IDRs that make up nodes in networks remains unclear. An example of such an interactome is the *B. subtilis* FtsZ divisome [71] (**Figure 4.12**). Here, we report the z-score matrices for IDRs drawn from the known interactors of *B. subtilis* FtsZ (**Figure 1.3**). Further analysis will be required to uncover the potential co-evolution of patterning features across orthologous divisomes.

The work presented here offers a new methodology to compare hypervariable sequences and drive inferences about conserved sequence parameters to function relationships. We look forward to its usage as a driver of IDP / IDR design for uncovering sequence-ensemble-function relationships and for identifying features that are likely selected for among functional orthologs. It is worth noting that we have focused exclusively on binary patterns. This analysis masks higher-order correlations that may or may not be fully captured by considering the full set of binary patterning parameters. Generalizations that go beyond analyses of binary patterning parameters might become essential as we try to uncover the sequence-function relationships encoded by hypervariable IDRs.



**Figure 4.12: Z-score matrices of the disordered domains within the FtsZ interactome.** Each node has features that diverge from the null model and could be significant to intra- and inter-molecular functions.



## 4.6 References

1. Ward, J.J., et al., *Prediction and functional analysis of native disorder in proteins from the three kingdoms of life*. J Mol Biol, 2004. **337**(3): p. 635-45.
2. Peng, Z., M.J. Mizianty, and L. Kurgan, *Genome-scale prediction of proteins with long intrinsically disordered regions*. Proteins: Structure, Function, and Bioinformatics, 2014. **82**(1): p. 145-158.
3. Wright, P.E. and H.J. Dyson, *Linking folding and binding*. Current Opinion in Structural Biology, 2009. **19**(1): p. 31-38.
4. Wiley, E.O. and B.S. Lieberman, *Phylogenetics: theory and practice of phylogenetic systematics*. 2011: John Wiley & Sons.
5. Higgins, D.G. and P.M. Sharp, *CLUSTAL: a package for performing multiple sequence alignment on a microcomputer*. Gene, 1988. **73**(1): p. 237-244.
6. Russell, R.B., et al., *Recognition of analogous and homologous protein folds: analysis of sequence and structure conservation*. Journal of molecular biology, 1997. **269**(3): p. 423-439.
7. Zarin, T., et al., *Selection maintains signaling function of a highly diverged intrinsically disordered region*. Proc Natl Acad Sci U S A, 2017. **114**(8): p. E1450-E1459.
8. Zarin, T., et al., *Proteome-wide signatures of function in highly diverged intrinsically disordered regions*. Elife, 2019. **8**: p. e46883.
9. Cohan, M.C. and R.V. Pappu, *Making the Case for Disordered Proteins and Biomolecular Condensates in Bacteria*. Trends Biochem Sci, 2020. **45**(8): p. 668-680.
10. Stoneking, M., *Hypervariable sites in the mtDNA control region are mutational hotspots*. The American Journal of Human Genetics, 2000. **67**(4): p. 1029-1032.
11. Cohan, M.C., K.M. Ruff, and R.V. Pappu, *Information theoretic measures for quantifying sequence-ensemble relationships of intrinsically disordered proteins*. Protein Eng Des Sel, 2019. **32**(4): p. 191-202.
12. Lange, J., L.S. Wyrwicz, and G. Vriend, *KMAD: knowledge-based multiple sequence alignment for intrinsically disordered proteins*. Bioinformatics, 2016. **32**(6): p. 932-936.

13. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Current Opinion in Structural Biology, 2015. **32**: p. 102-12.
14. Moesa, H.A., et al., *Chemical composition is maintained in poorly conserved intrinsically disordered regions and suggests a means for their classification*. Molecular BioSystems, 2012. **8**(12): p. 3262-3273.
15. Schlessinger, A., et al., *Protein disorder—a breakthrough invention of evolution?* Current opinion in structural biology, 2011. **21**(3): p. 412-418.
16. Mao, A.H., et al., *Net charge per residue modulates conformational ensembles of intrinsically disordered proteins*. Proceedings of the National Academy of Sciences, 2010. **107**(18): p. 8183-8188.
17. Strickfaden, S.C., et al., *A mechanism for cell-cycle regulation of MAP kinase signaling in a yeast differentiation pathway*. Cell, 2007. **128**(3): p. 519-531.
18. Ba, A.N.N., et al., *Proteome-wide discovery of evolutionary conserved sequences in disordered regions*. Science signaling, 2012. **5**(215): p. rs1-rs1.
19. Davey, N.E., M.S. Cyert, and A.M. Moses, *Short linear motifs - ex nihilo evolution of protein regulation*. Cell Commun Signal, 2015. **13**: p. 43.
20. Davey, N.E., et al., *Attributes of short linear motifs*. Mol Biosyst, 2012. **8**(1): p. 268-81.
21. Gomes, E. and J. Shorter, *The molecular language of membraneless organelles*. Journal of Biological Chemistry 2019. **294**(18): p. 7115-7127.
22. Martin, E.W. and T. Mittag, *Relationship of Sequence and Phase Separation in Protein Low-Complexity Regions*. Biochemistry, 2018. **57**(17): p. 2478-2487.
23. Martin, E.W., et al., *Valence and patterning of aromatic residues determine the phase behavior of prion-like domains*. Science, 2020. **367**(6478): p. 694-699.
24. Chong, P.A., R.M. Vernon, and J.D. Forman-Kay, *RGG/RG motif regions in RNA binding and phase separation*. Journal of molecular biology, 2018. **430**(23): p. 4650-4665.
25. Varadi, M., et al., *Functional advantages of conserved intrinsic disorder in RNA-binding proteins*. PloS one, 2015. **10**(10): p. e0139731.

26. Boeynaems, S., et al., *Phase Separation of C9orf72 Dipeptide Repeats Perturbs Stress Granule Dynamics*. *Molecular Cell* 2017. **65**(6): p. 1044-1055 e5.
27. Lyle, N., R.K. Das, and R.V. Pappu, *A quantitative measure for protein conformational heterogeneity*. *J Chem Phys*, 2013. **139**(12): p. 121907.
28. Sundararajan, K., et al., *The bacterial tubulin FtsZ requires its intrinsically disordered linker to direct robust cell wall construction*. *Nat Commun*, 2015. **6**: p. 7281.
29. Gardner, K.A., D.A. Moore, and H.P. Erickson, *The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide*. *Mol Microbiol*, 2013. **89**(2): p. 264-75.
30. Buske, P.J. and P.A. Levin, *A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo*. *Mol Microbiol*, 2013. **89**(2): p. 249-63.
31. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division*. *Semin Cell Dev Biol*, 2015. **37**: p. 3-10.
32. Cohan, M.C., et al., *Dissecting the Functional Contributions of the Intrinsically Disordered C-terminal Tail of Bacillus subtilis FtsZ*. *J Mol Biol*, 2020. **432**(10): p. 3205-3221.
33. Vaughan, S., et al., *Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota*. *Journal of molecular evolution*, 2004. **58**(1): p. 19-29.
34. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. *Curr Opin Struct Biol*, 2015. **32**: p. 102-12.
35. Beveridge, R., et al., *Ion mobility mass spectrometry uncovers the impact of the patterning of oppositely charged residues on the conformational distributions of intrinsically disordered proteins*. *Journal of the American Chemical Society*, 2019. **141**(12): p. 4908-4918.
36. Sherry, K.P., et al., *Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor*. *Proceedings of the National Academy of Sciences*, 2017. **114**(44): p. E9243-E9252.

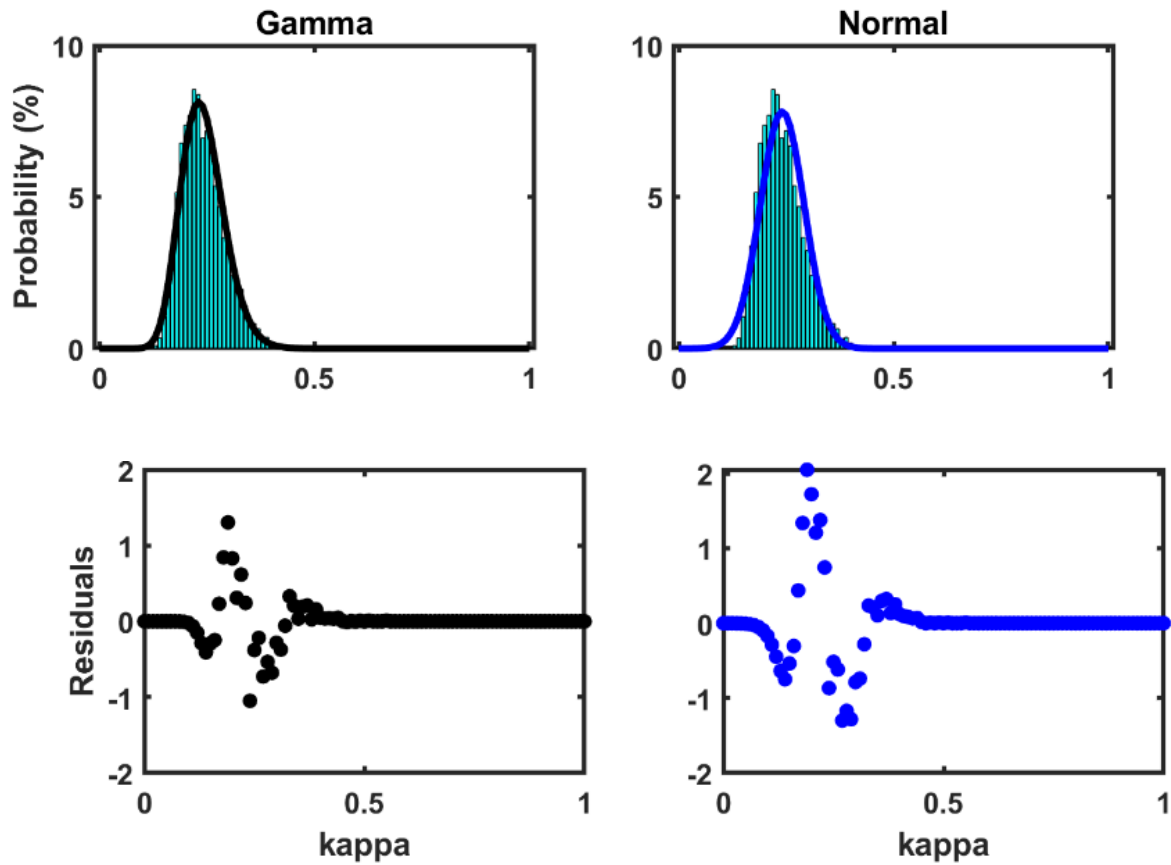
37. Martin, E.W., et al., *Sequence determinants of the conformational properties of an intrinsically disordered protein prior to and upon multisite phosphorylation*. Journal of the American Chemical Society, 2016. **138**(47): p. 15323-15335.
38. Van Roey, K., et al., *Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation*. Chem Rev, 2014. **114**(13): p. 6733-78.
39. Tompa, P., et al., *A million peptide motifs for the molecular biologist*. Mol Cell, 2014. **55**(2): p. 161-9.
40. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*. Proc Natl Acad Sci U S A, 2013. **110**(33): p. 13392-7.
41. Holehouse, A.S., et al., *CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins*. Biophys J, 2017. **112**(1): p. 16-21.
42. Buske, P.J. and P.A. Levin, *Extreme C terminus of bacterial cytoskeletal protein FtsZ plays fundamental role in assembly independent of modulatory proteins*. J Biol Chem, 2012. **287**(14): p. 10945-57.
43. Barrows, J.M., et al., *FtsA Regulates Z-Ring Morphology and Cell Wall Metabolism in an FtsZ C-Terminal Linker-Dependent Manner in Caulobacter crescentus*. J Bacteriol, 2020. **202**(7): p. e00693-19.
44. Sundararajan, K. and E.D. Goley, *The intrinsically disordered C-terminal linker of FtsZ regulates protofilament dynamics and superstructure in vitro*. J Biol Chem, 2017. **292**(50): p. 20509-20527.
45. Sundararajan, K., et al., *Species- and C-terminal linker-dependent variations in the dynamic behavior of FtsZ on membranes in vitro*. Mol Microbiol, 2018. **110**(1): p. 47-63.
46. Sundararajan, K., J.M. Barrows, and E.D. Goley, *Determinants of FtsZ C-terminal linker-dependent regulation of cell wall metabolism in <em>Caulobacter crescentus</em>*. bioRxiv, 2019: p. 632075.
47. Ait-Bara, S. and A.J. Carpousis, *RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs*. Mol Microbiol, 2015. **97**(6): p. 1021-135.

48. Ait-Bara, S., A.J. Carpousis, and Y. Quentin, *RNase E in the gamma-Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains*. Mol Genet Genomics, 2015. **290**(3): p. 847-62.
49. Al-Husini, N., et al., *alpha-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies*. Molecular Cell, 2018. **71**(6): p. 1027-1039 e14.
50. Lin, Y.H., J.D. Forman-Kay, and H.S. Chan, *Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles*. Phys Rev Lett, 2016. **117**(17): p. 178101.
51. Kozlov, A.G., et al., *Glutamate promotes SSB protein-protein Interactions via intrinsically disordered regions*. J Mol Biol, 2017. **429**(18): p. 2790-2801.
52. Kozlov, A.G., et al., *Intrinsically disordered C-terminal tails of E. coli single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA*. J Mol Biol, 2015. **427**(4): p. 763-774.
53. Shinn, M.K., et al., *Are the intrinsically disordered linkers involved in SSB binding to accessory proteins?* Nucleic Acids Res, 2019. **47**(16): p. 8581-8594.
54. Harami, G.M., et al., *Phase separation by ssDNA binding protein controlled via protein-protein and protein-DNA interactions*. Proc Natl Acad Sci U S A, 2020. **117**(42): p. 26206-26217.
55. Li, L., M.B. Charati, and K.L. Kiick, *Elastomeric polypeptide-based biomaterials*. J Polym Sci A Polym Chem, 2010. **1**(8): p. 1160-1170.
56. Dzuricky, M., S. Roberts, and A. Chilkoti, *Convergence of Artificial Protein Polymers and Intrinsically Disordered Proteins*. Biochemistry, 2018. **57**(17): p. 2405-2414.
57. Quiroz, F.G., et al., *Intrinsically disordered proteins access a range of hysteretic phase separation behaviors*. Science advances, 2019. **5**(10): p. eaax5177.
58. Roberts, S., M. Dzuricky, and A. Chilkoti, *Elastin-like polypeptides as models of intrinsically disordered proteins*. FEBS letters, 2015. **589**(19): p. 2477-2486.
59. Athamneh, A.I. and J.R. Barone, *Enzyme-mediated self-assembly of highly ordered structures from disordered proteins*. Smart materials and structures, 2009. **18**(10): p. 104024.

60. Holland, G.P., et al., *Determining secondary structure in spider dragline silk by carbon-carbon correlation solid-state NMR spectroscopy*. J Am Chem Soc, 2008. **130**(30): p. 9871-7.
61. Ringli, C., B. Keller, and U. Ryser, *Glycine-rich proteins as structural components of plant cell walls*. Cellular and Molecular Life Sciences CMLS, 2001. **58**(10): p. 1430-1441.
62. Cao, Q., Y. Wang, and H. Bayley, *Sequence of abductin, the molluscan 'rubber' protein*. Current biology, 1997. **7**(11): p. R677-R678.
63. Dubiel, K., et al., *Structural Mechanisms of Cooperative DNA Binding by Bacterial Single-Stranded DNA-Binding Proteins*. J Mol Biol, 2019. **431**(2): p. 178-195.
64. Kozlov, A.G., R. Galletto, and T.M. Lohman, *SSB–DNA binding monitored by fluorescence intensity and anisotropy*, in *Single-Stranded DNA Binding Proteins*. 2012, Springer. p. 55-83.
65. Shereda, R.D., et al., *SSB as an organizer/mobilizer of genome maintenance complexes*. Crit Rev Biochem Mol Biol, 2008. **43**(5): p. 289-318.
66. Antony, E. and T.M. Lohman. *Dynamics of E. coli single stranded DNA binding (SSB) protein-DNA complexes*. in *Seminars in cell & developmental biology*. 2019. Elsevier.
67. Kozlov, A.G., M.K. Shinn, and T.M. Lohman, *Regulation of Nearest-Neighbor Cooperative Binding of E. coli SSB Protein to DNA*. Biophys J, 2019. **117**(11): p. 2120-2140.
68. Bianco, P.R., *The tale of SSB*. Prog Biophys Mol Biol, 2017. **127**: p. 111-118.
69. Bianco, P.R., et al., *The IDL of E. coli SSB links ssDNA and protein binding by mediating protein–protein interactions*. Protein Science, 2017. **26**(2): p. 227-241.
70. MacEwan, S.R. and A. Chilkoti, *Elastin-like polypeptides: Biomedical applications of tunable biopolymers*. Peptide Science, 2010. **94**(1): p. 60-77.
71. Wagstaff, J. and J. Löwe, *Prokaryotic cytoskeletons: protein filaments organizing small cells*. Nature Reviews Microbiology, 2018. **16**(4): p. 187-201.
72. Firman, T. and K. Ghosh, *Sequence charge decoration dictates coil-globule transition in intrinsically disordered proteins*. The Journal of chemical physics, 2018. **148**(12): p. 123305.



## 4.7 Appendix

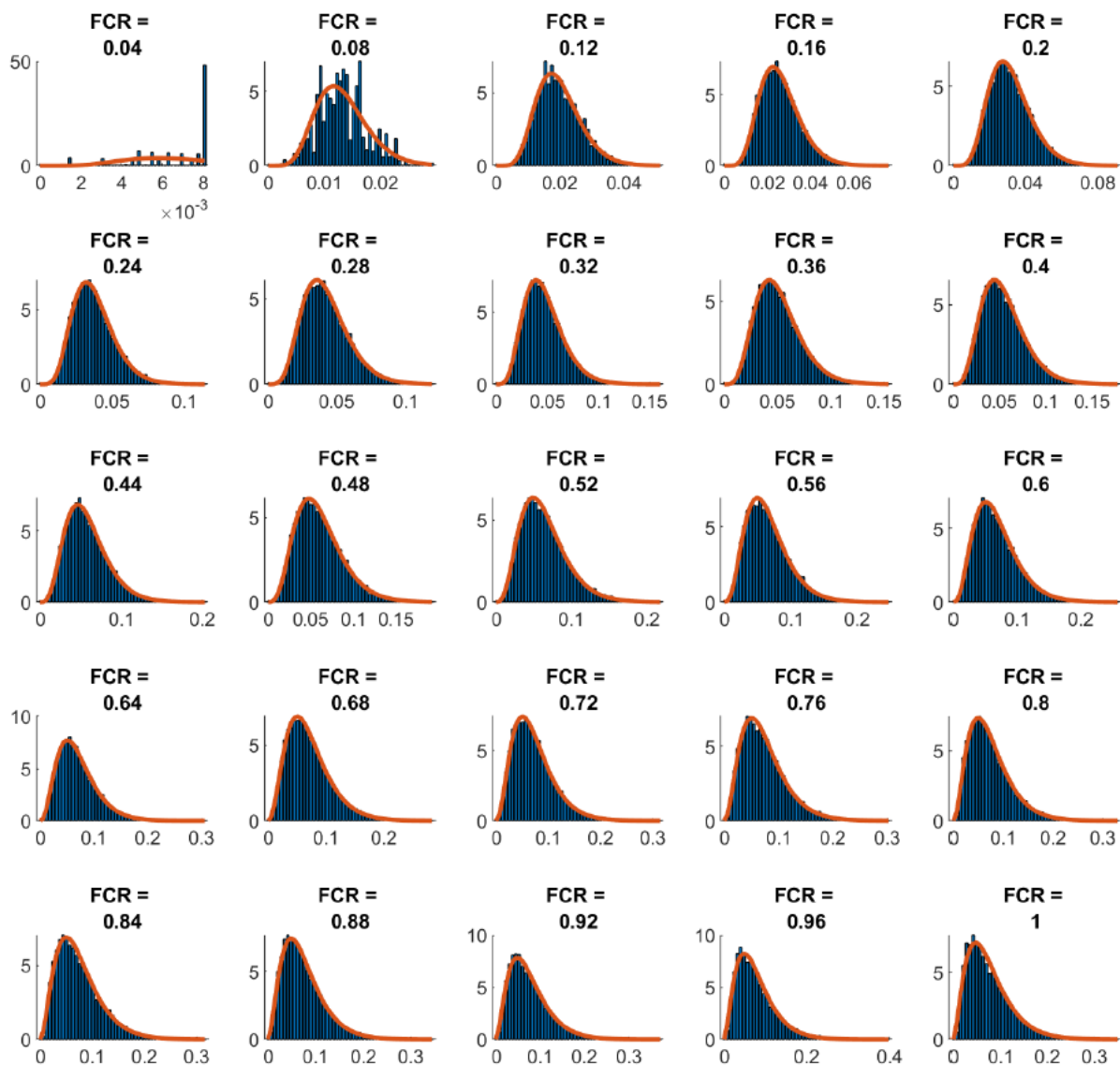


**Figure 4A.1: Fitting to a gamma distribution versus a normal distribution.** Sample of patterning parameter ( $\kappa$ ) values from scrambled sequences fit to a gamma distribution (top left) compared to a normal distribution fit (top right). Residuals of each fit are shown on the bottom left and right quadrants. The  $R^2$  values are 0.99 (left) and 0.95 (right).

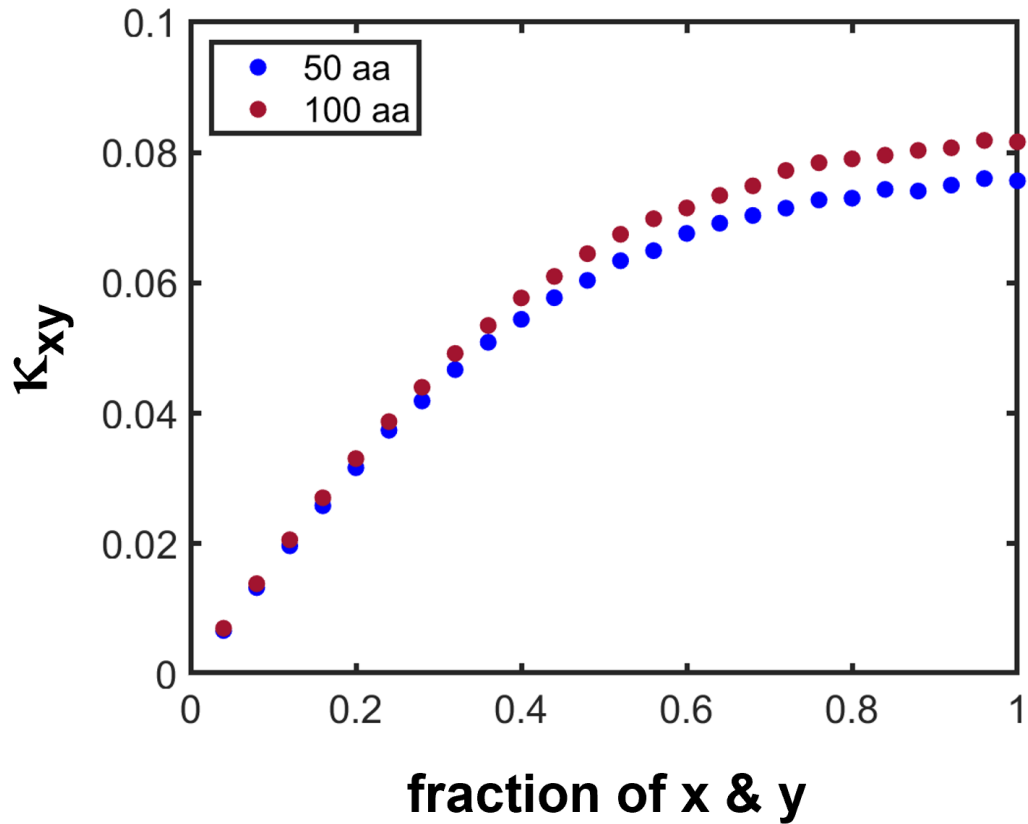


### *Demonstrating composition-dependence of patterning parameters*

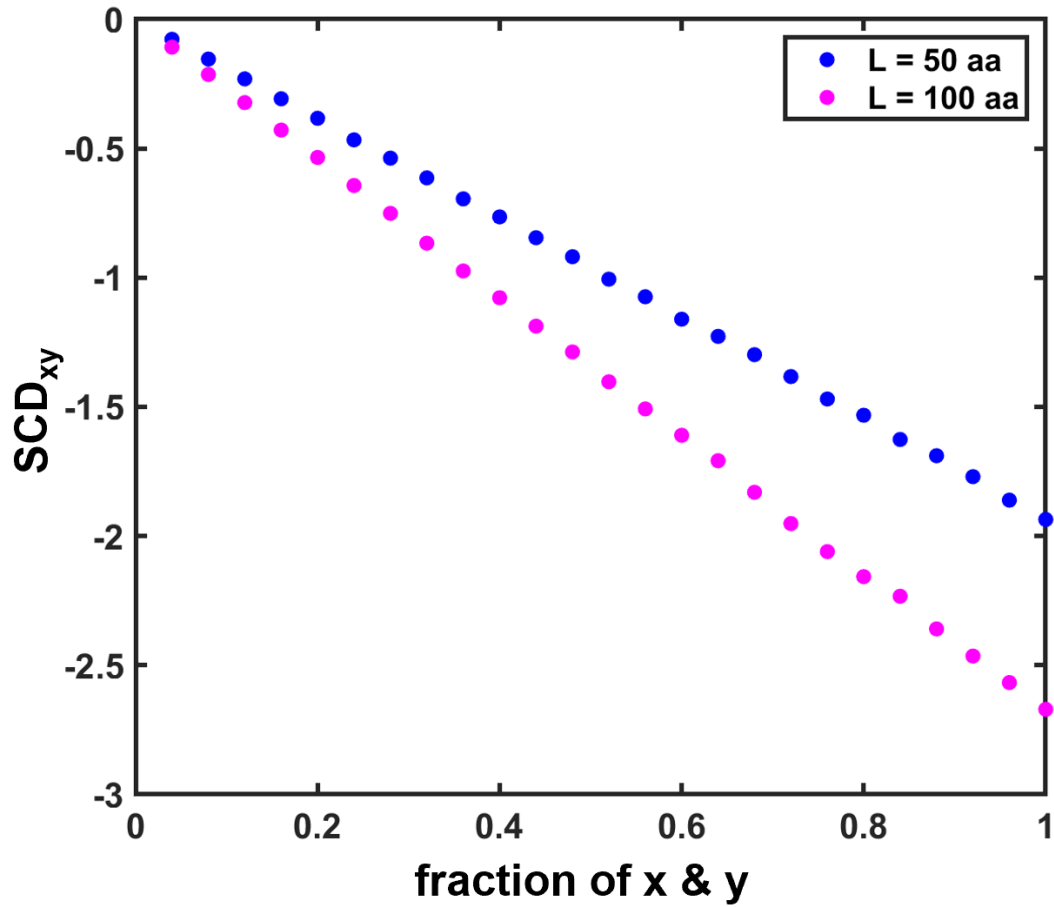
In this section, we built upon the work of Holehouse et al. [41] that demonstrated that sequences of varying compositions have different underlying expected values for a patterning parameter of interest. While this analysis was performed using the conventional version of  $\kappa$  ( $\kappa_{+-}$ ), the approach can be applied to other residues and residue types, as demonstrated in **Section 4.3.2**. Test sequences of three amino acids (A, E, and K) were generated that titrated the content of the sequence such that the fraction of charged residues (FCR) was varied. The sequences were randomly scrambled  $10^5$  times, and the  $\kappa_{+-}$  values were calculated. At sequence compositions above an FCR of 0.1, a gamma distribution could fit the data (**Figure 4A.2**). The mean of the resulting distribution was considered the expected value ( $\mu$ ) for the given sequence. The analysis was repeated for a sequence of twice the length (50 versus 100 amino acids). In **Figure 4A.3**, we present the residue-type agnostic version of the data ( $\kappa_{xy}$  versus  $\kappa_{+-}$ ) and show that the expected value changes with varying sequence composition and length. This observation is not unique to  $\kappa$ -based patterning parameters. Using Sequence Charge Decoration (SCD) [72] instead to assess the patterning of oppositely charged residues revealed a similar vulnerability to comparing sequences across compositions and lengths (**Figure 4A.4**). A similar assessment was performed with  $\Omega$ -based calculations using a set of sequences that titrated the content of the residue of interest ( $x$ ) and A. The goodness of fit for these data constrains the usage to a tighter window where only fractional compositions above 0.12 and below 0.92 were applicable (**Figure 4A.5**). Again, these data showed that the expected value of  $\Omega_x$  varies with the fraction of residue  $x$  (**Figure 4A.6**). This analysis shows that we need a different approach, based on the comparison of z-scores, to compare patterning features across sequences of different compositions and lengths.



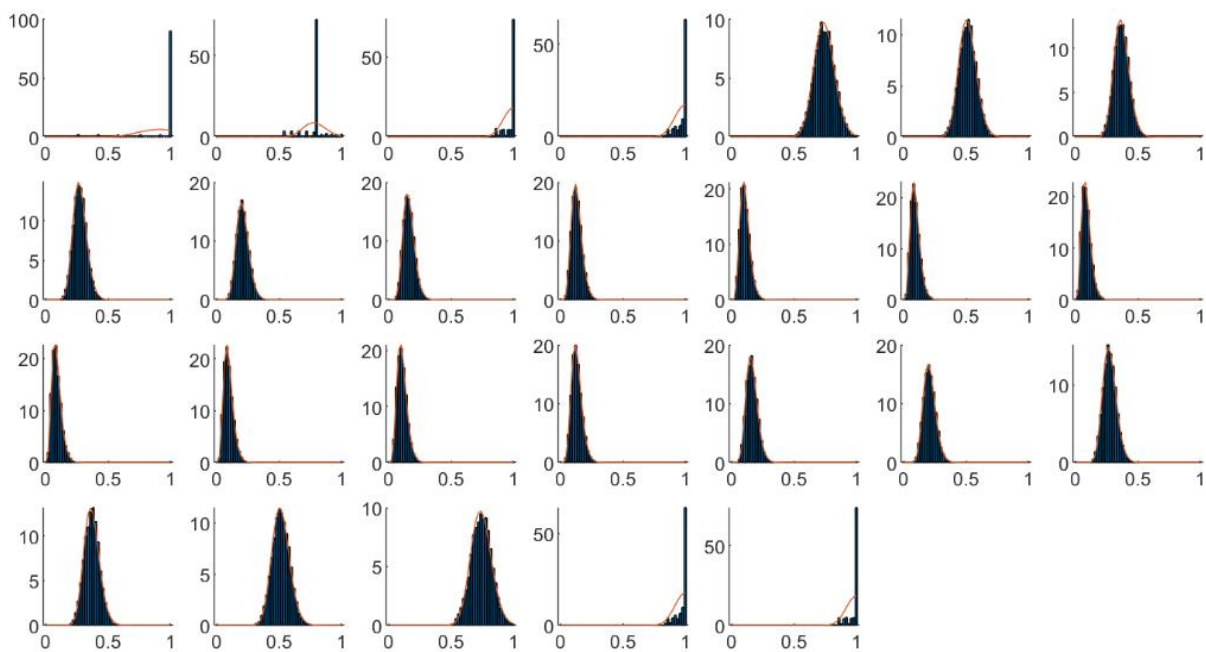
**Figure 4A.2: Goodness of fit of the gamma distribution as a function of the fraction of the residues of interest for  $\kappa$ -based patterning parameters.** Sequences of 50 amino acids and three different residues, A, E, and K, were used. The net charge per residue for each sequence was fixed at 0, while the fraction of charged residues (FCR) was varied. The unnormalized  $\kappa$  value for the positive and negative residues ( $\kappa_{+-}$ ) was calculated.



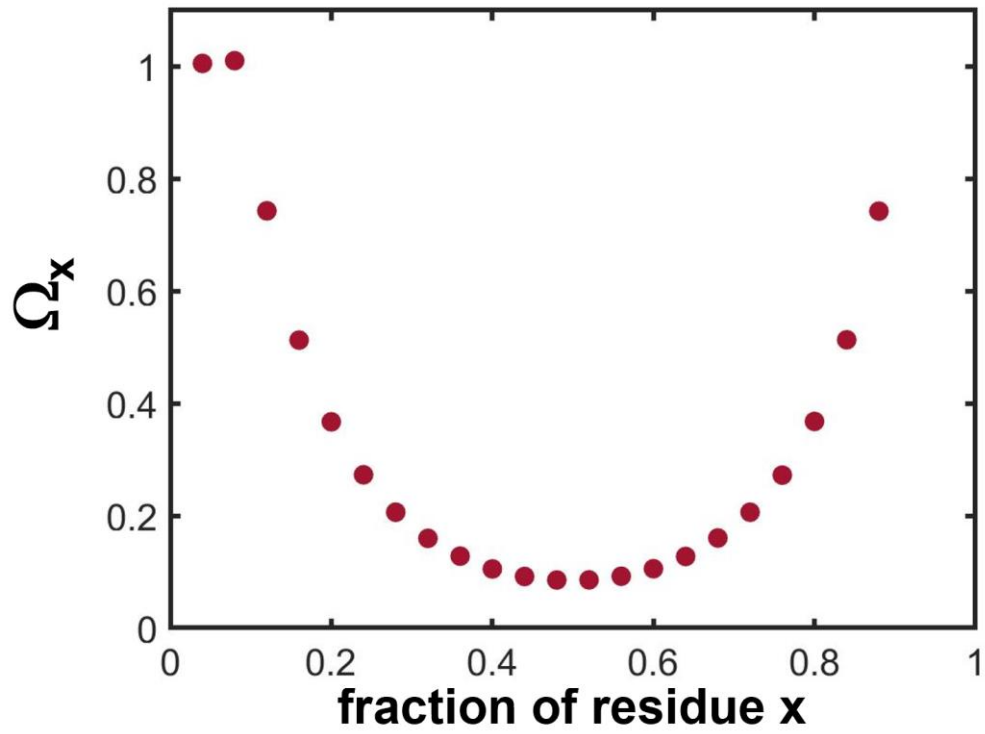
**Figure 4A.3:** The expected value for a given parameter depends on the sequence content. The null-scramble expectations of  $\kappa$  for a 50 and 100 amino acid sequence are shown as a function of residue x and y ( $\kappa_{xy}$ ). Length and composition alter the underlying expected value.



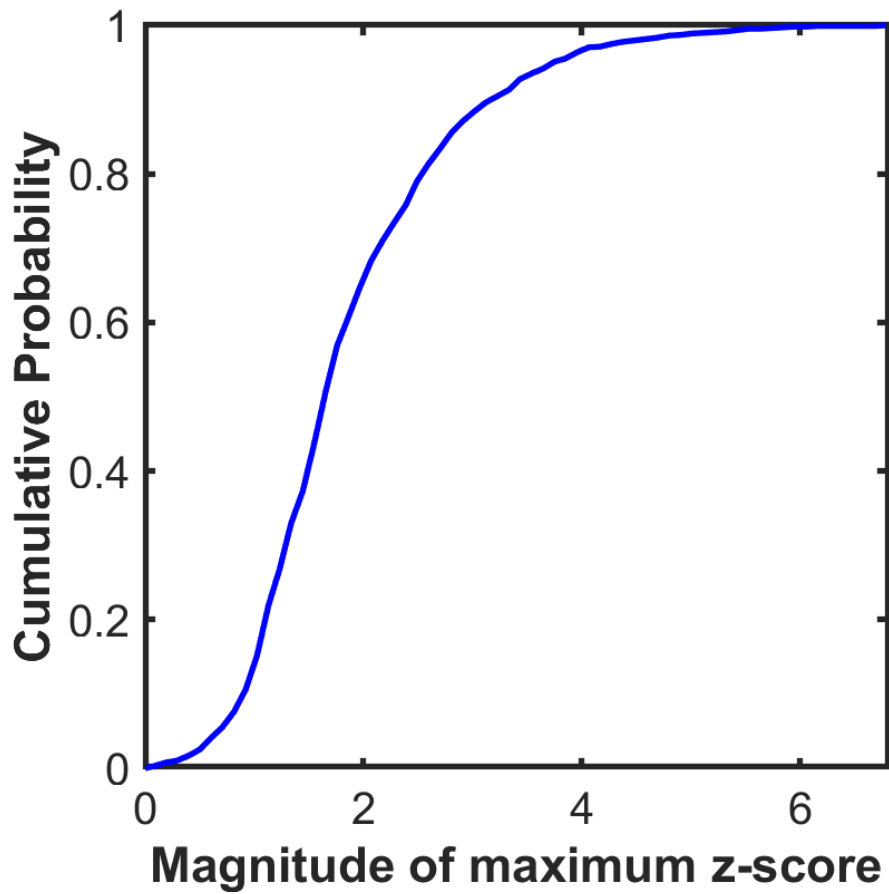
**Figure 4A.4: The expected value of SCD depends on the sequence charge content and length.** The null-scramble expectations of SCD are plotted for a 50 and 100 amino acid sequence as a function of residue x that is positively charged and y that is negatively charged (SCD<sub>xy</sub>). The expectation of SCD is length-dependent.



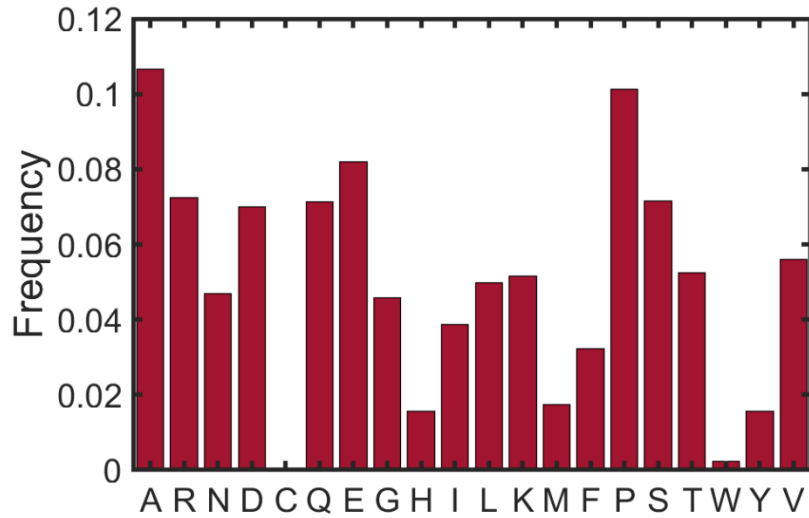
**Figure 4A.5: Goodness of fit of the gamma distribution as a function of the fraction of the residues of interest for  $\Omega$ -based patterning parameters.** Each subplot is a distribution of  $\Omega$  values for scrambles of 50 amino acid sequences that increase in fractional residue content from 0 (upper left) to 1 bottom right by 0.04 for each figure (row value increases first).



**Figure 4A.6:** The expected value for the  $\Omega$  parameter depends on the sequence content. The plot shows the null-scramble expectations of  $\Omega$  for a 100 amino acid as a function of the fraction of residue x ( $\Omega_x$ ).



**Figure 4A.7: Cumulative probability of observing a maximum z-score of a certain magnitude for each FtsZ CTL from the system of orthologs.** At each magnitude, the cumulative probability represents the fraction of sequences that will be considered random if that value was set as the z-score threshold.



**Figure 4A.8: Average amino acid frequencies within the set of CTLs.** The frequencies of each amino acid for each sequence are averaged across the system of orthologs to avoid overrepresentation of specific amino acids due to potential enrichments within sequences of certain lengths.



*E. coli* RNase E

FCR = 0.346 | NCPR = 0.025

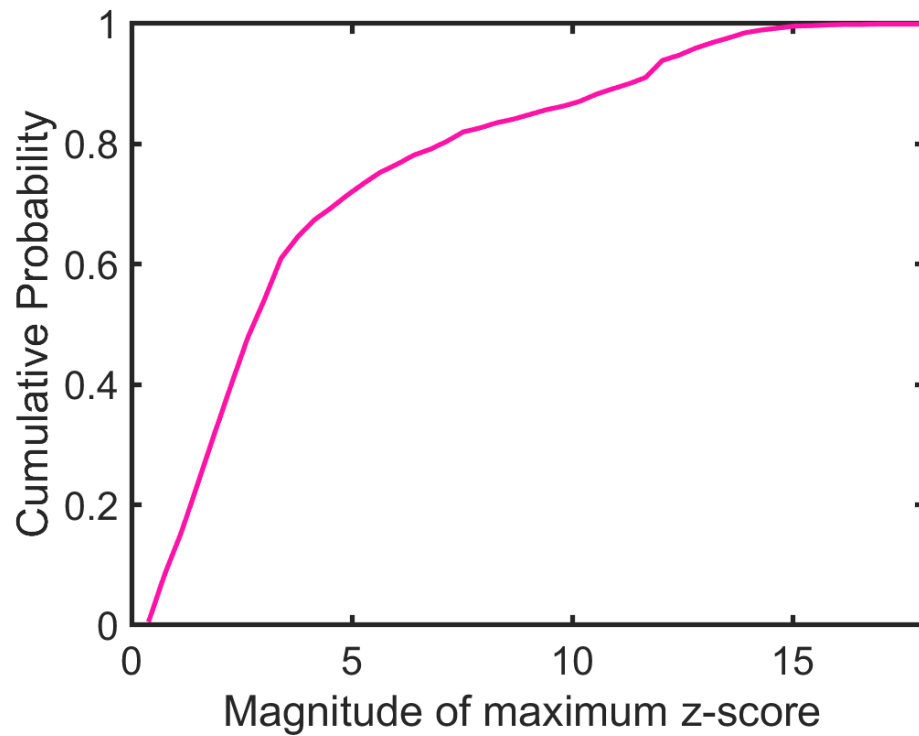
RQDGVRCVIV PNDQMETPHY HVLRVRKGEETPTLSYMLPK LHEEAMALPS  
EEEF AERKRP EQPALATFAM PDVPPAPTPA EPAA PVVAPA PKAAPATPAA  
PAQPGLLSRF FGALKALFSG GEETKPTEQP APKAEAKPER QQDRRKPRQN  
NRRDRNERRD TRSERTEGSD NREENRRNRR QAQQQTAETR ESRQQAEVTE  
KARTADEQQA PRRERSRRRN DDKRQAQQA KALNVEEQSV QETEQEERVR  
PVQPRRKQRQ LNQKVRYEQS VAE EAVVAPV VEETVAAEPI VQEAPAPRTE  
LVKVPLPVVA QTAPEQQEEN NADNRDNGGM PRRSRRSPRH LRVSGQRRRR  
YRDERYPTQS PMPLTVA

*C. crescentus* RNase E

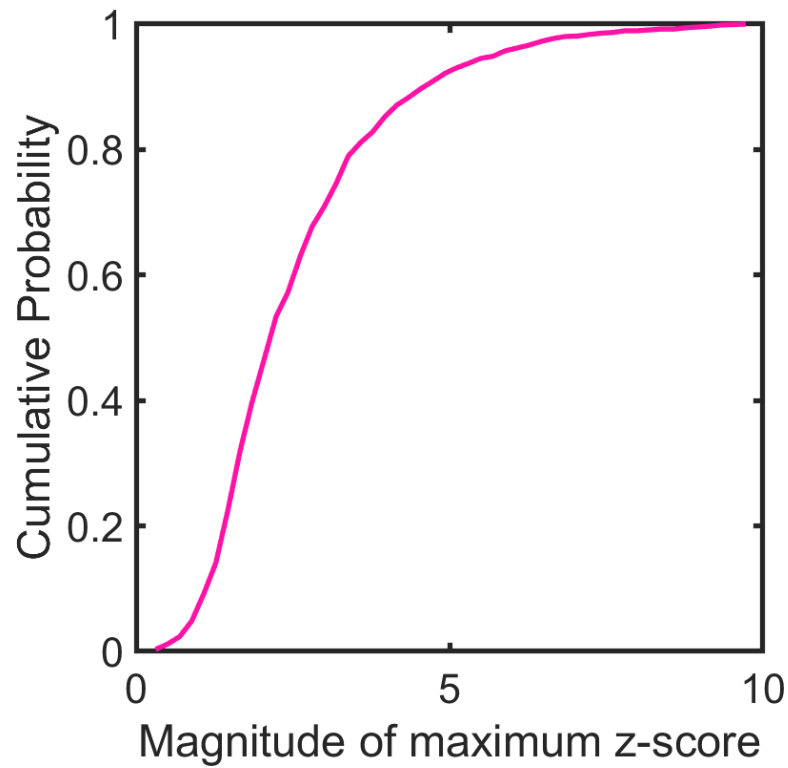
FCR = 0.406 | NCPR = -0.085

TGVLEGTTHV CEHCEGTGRV RSVESSALAA LRAVEAEALK GSGSVILKVS  
RSVGLYILNE KR DYLQRLLT THGLFVSVVV DDSLHAGDQE IERTELGERI  
AVAPPPFVEE DDDFDPNAYD DEEEEDDVIL DDEDDTDRED TDDDDATTRK  
SARDDERGDR KGRRGRRDRN RGRGRRDERD GETESEDEDV VAEGADEDRG  
EFGDDDEGGR RRRRRGRRGG RRGREDGDR PTDAFVWIRP RVPFGENVFT  
WHDP AALVGG GESRRQ APEP RVDAATEAAP RPERAEREER PGRERGRGR  
DRGRRQRDEA PVAEMTSVES ATVEAAEPFE APILAPPVIA GPPADVWVEL  
PEVEEAPKKP KR SRARGKKA TETSVEAIDT VTEVAAEAPA PETAEPEAVE  
VAPPAPTVEA APEPGPVVEA VEEAQP AEPD PNEITAPPEK PRRGWWRR

**Figure 4A.9: Sequences of the CTDs from RNases E.** Negative residues are colored in red, and positively charged residues are colored in blue. The blocky architecture of *C. crescentus* RNase E is shown by the enrichment of blue or red stretches within the sequence. FCR stands for the fraction of charged residues, and NCPR stands for the net charge per residue. Both sequences are highly charged and polyampholytic, which means the net charge per residue is approaching 0 and indicates an equal balance of positively and negatively charged residues.



**Figure 4A.10: Cumulative probability of observing a maximum z-score of a certain magnitude for each RNase E CTD from the system of orthologs.** At each magnitude, the cumulative probability represents the fraction of sequences that will be considered random if that value was set as the z-score threshold.



**Figure 4A.11: Cumulative probability of observing a maximum z-score of a certain magnitude for each SSB IDL from the system of orthologs.** At each magnitude, the cumulative probability represents the fraction of sequences that will be considered random if that value was set as the z-score threshold.

## Chapter 5

# Impact of the molecular grammar of the *B. subtilis* FtsZ C-terminal linker on function and bacterial cell division

### 5.1 Preamble

This chapter builds upon the observations made in Chapters 2-4 and proposes a design parameter to test the working hypothesis that a functional C-terminal linker (CTL) of FtsZ must act as a spacer, must preserve sequence-ensemble relationships, and must maintain encoded non-random sequence features. Here, we use the patterning of oppositely charged residues, which refers to the extent of linear mixing / segregation of oppositely charged residues, as the design parameter and direct a combination of biophysical and cellular studies to further our understanding of how the sequence of the CTL of FtsZ contributes to Z-ring assembly.

To assess the impact of systemically titrating the features of the CTL sequence, we performed experiments, aided by sequence design, to quantify the impact of sequence on the

functions of FtsZ from *B. subtilis*. Z-ring formation is robust if and only if the extent of linear mixing / segregation of oppositely charged residues within the CTT sequences does not significantly perturb sequence-ensemble relationships and the ability of the CTL to function as a spacer. Otherwise, aberrant, CTT-mediated, FtsZ assemblies impair Z-ring formation. Furthermore, the non-random sequence features of CTLs also had functional significance because FtsZ variants with fully random CTLs behaved differently than wild type *in vitro*. Taken together, our results suggest that CTT sequences vary according to the *Goldilocks precept* of being “just right.” This appears to be achieved through a combination of maintaining sequence-ensemble relationships as well as sequence-encoded intra- and intermolecular interactions.

## 5.2 Introduction

Cell division is initiated by the formation of a cytokinetic ring involving key proteins at the nascent division site [1-9]. In rod-shaped bacteria, the main building block of the cytokinetic ring, also known as the Z-ring, is the essential GTPase FtsZ, a homolog of eukaryotic tubulin [10-16]. FtsZ forms single-stranded polymers *in vitro*, and GTP-binding promotes FtsZ polymerization [16]. FtsZ polymers serve as a treadmilling platform for the division machinery, particularly enzymes required to synthesize peptidoglycans, the massive polymers that make up the bacterial cell wall [17, 18].

In bacteria, the FtsZ ring is dynamic, with subunit turnover occurring on the order of seconds [19]. Experiments *in vitro* indicate that FtsZ polymerization and Z-ring formation involve phase transitions, in which FtsZ monomers associate to form linear polymers referred to as filaments or protofilaments. These filaments associate laterally to form bundles [13, 20]. In the presence of GTP, FtsZ forms single-stranded polymers or filaments [21] at concentrations above 1  $\mu\text{M}$  [22-24]. The folded N-terminal core domain forms a complete GTPase upon dimerization. The interface between the nucleotide-binding site and the T7 loop (in dimers and higher-order polymers) is the active site for GTP hydrolysis [25]. FtsZ subunits undergo conformational changes to facilitate the population of a high-affinity state that favors polymerization [26]. This cooperativity contributes to the requirement of a threshold concentration for FtsZ assembly, which has been described by numerous models [27-31]. However, in some FtsZs like the one from *B. subtilis*, the formation of assemblies, i.e., bundles of filaments, requires the crossing of two concentration thresholds [32, 33].

In previous work, we showed that the mechanism of FtsZ assembly is governed by its modular architecture (**Figure 1.1**) [34]. FtsZ encompasses two main domains, the GTPase core and the C-terminal tail (CTT). The CTT can be further subdivided into an intrinsically disordered linker (CTL) and a C-terminal peptide (CTP) [16, 35, 36]. Recent work showed that the CTT modules impact the overall efficiency of the GTPase domain, imparting an auto-regulatory functionality of the core [34]. Based on this work, which focused on the *B. subtilis* FtsZ, we proposed showed that the CTP and CTL modules play the role of a *sticker* and *spacer*, respectively, in FtsZ assembly (Chapter 2, [34, 37]). The designation of a sticker implies that the CTP enables a precise network of homotypic and heterotypic protein-protein interactions.

In contrast, the spacer keeps the CTP physically separate from the GTPase core, thereby ensuring modularity of core versus sticker functionalities. Random sequences are unlikely to fulfill the role of a spacer since a *bona fide* spacer is one that should be inert in that its interactions with itself, with the core, with other spacers, the sticker, and the surrounding solvent should counterbalance one another. In the absence of the CTT, the second critical concentration threshold is not observed. The core requires a sticker (CTP) properly separated from itself by a spacer (CTL) to enable protofilament bundling.

Within this *stickers-and-spacers* framework for proteins that drive phase transitions, stickers are the interaction motifs that engage in physical crosslinks. Spacers provide the linear and / or spatial scaffolds for stickers. To first order, the driving forces for phase transitions are governed by the valence and linear / spatial patterning of stickers, whereas the cooperativity and material properties of phase transitions are governed by excluded volumes of spacers [38-41]. This is relevant because FtsZs undergo GTP-assisted polymerization that leads to the formation of

single-stranded polymers, each comprising multiple FtsZ subunits [33, 42] giving rise to emergent multivalence of CTP stickers. Polymers of FtsZ, defined by a multivalence of CTP stickers, can engage in a condensation transition that leads to bundling via homotypic interactions among single-stranded polymers [56]. Alternatively, single-stranded polymers and / or bundles can engage in heterotypic interactions with FtsZ interacting proteins. The heterotypic networking transition involves modulatory proteins that assist in anchoring, stabilizing, and providing spatiotemporal control of the cytokinetic ring [48, 58].

Since the CTP is a molecular recognition element [43] that coordinates heterotypic protein-protein interactions involving FtsZ [44], it follows that the sequence of this motif is conserved across all orthologs, with some variations likely because of the differing interactomes [30, 45-48]. Conversely, as illustrated in the previous chapter of this thesis, the CTL is hypervariable (**Figure 1.2**). Yet, the presence of a disordered CTL is essential to the functions of FtsZs from *B. subtilis* [36], *E. coli* [49], and *C. crescentus* [50]. Given the emerging insights about sequence-ensemble relationships of IDPs / IDRs, the observations regarding the importance and role of the CTL / CTT of FtsZ raise several unanswered questions regarding the sequence grammar of functional CTLs / CTTs. In mutational studies of these CTLs, none of the variants tested to date have yielded fully equivalent results. While some CTLs have worked in FtsZs from other orthologs, most reveal that there are system-specific requirements. Furthermore, it is difficult to decode the rules of a functional CTL using information gathered from these studies such as truncations, elongations, or other variants of the sequence that seemingly change a single parameter (the length, net charge, etc.) will actually alter multiple parameters, including amino acid composition and the patterning of residues [51, 52]. These changes will in turn affect the balance of sequence-encoded interactions



that contribute to sequence-ensemble relationships of the disordered CTL [53, 54] and will impact the valence and position of sequence motifs (**Chapter 4**).

In the previous chapters of this thesis, we showed that each CTL sequence within a system of orthologs appears to encode for system-specific and unique features. Here, we use *de novo* sequence design to test the working hypothesis that there are requirements for a functional CTL in *B. subtilis* FtsZ: (1) the CTL must act as a spacer; (2) the CTL must preserve the sequence-to-ensemble relationships; and (3) the CTL must maintain the molecular grammar in the form of non-random and random sequence patterns. By keeping length and amino acid composition fixed, we explore the functional and phenotypic impacts of sequence variations to the CTL of FtsZ. We leverage the polyampholytic nature of the *B. subtilis* FtsZ CTL to design re-patterned variants whereby we change  $\kappa_{+-}$ , i.e., the linear patterning of oppositely charged residues. Based on previous studies, [55] [56, 57] and in **Chapter 4**, we propose that increasing  $\kappa_{+-}$  will convert spacers into stickers by increasing the likelihood of realizing complementary electrostatic interactions between blocks of opposite charges. In accord with recent studies [57-59], we show that there is an underlying sequence grammar governing functional CTL sequences that distinguish functional FtsZs and robust Z-ring formers from those that do not support FtsZ assembly, GTPase activity, Z-ring formation, and ultimately, cell division.

## 5.3 Material and Methods

### 5.3.1 General methods

*B. subtilis* strains expressing CTL variants of FtsZ were derived from PAL 2084 and were grown in 0.5% xylose to induce wild-type expression or 0.1 mM IPTG to induce CTL variant expression. Vent DNA polymerase was used for PCR (New England Biolabs). All restriction enzymes were purchased from New England Biolabs. All genomic DNA extractions were performed using the Wizard Genomic DNA Purification Kit (Promega). Plasmid preparations were made using the NucleoSpin Plasmid Kit (Macherey-Negel). Gel/PCR purifications were performed using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Negel). T4 DNA ligase was used for ligations (New England Biolabs).

### 5.3.2 Cloning CTL variants

The CTL variant strains were constructed as described previously [36]. The bacterial strains and plasmids used in this study are listed in **Table 5A.1** of the Appendix. Synthetic double-stranded oligonucleotides of the CTL variants were ordered from Integrated DNA Technologies, digested using restriction enzymes, and ligated into pPJ19, which contains FtsZ under the control of the Pspac promoter that is inducible with 0.1 mM IPTG, with restriction sites flanking the CTL. A BamHI site after amino acid 315 and an XmaI site before residue 366 result in the insertion of amino acid pairs GS and PG N- and C-terminal to the CTL, respectively. The plasmid was transformed into PAL 644, a strain of *E. coli* derived from PAL930. PAL 644 contains the low copy plasmid pBS58 expressing *E. coli* ftsQAZ, which allows the sub-cloning of *B. subtilis* FtsZ.

FtsZ was amplified in pPJ19, the product of which was restriction digested, purified, and ligated into pDR67. The multiple cloning site in the vector pDR67 contains the 5'- and 3'-ends of the amyE gene on either side, which allows the insertion of FtsZ into the amyE locus by homologous recombination. The purified plasmid was transformed into PAL 522, a derivative of the JH642 wild-type strain of *B. subtilis*. Genomic DNA was purified and transformed into MEO 1, a derivative of PAL 2084 containing a copy of wild-type FtsZ under the control of the Pxyl promoter, inducible with 0.5% xylose. The cells were made competent and transformed with purified genomic DNA from PAL 2084, which knocks out the chromosomal WT copy of FtsZ. Plasmids were verified by restriction digests and sequencing.

### **5.3.3 Immunoblotting**

Immunoblotting was performed as described previously [60]. Cells were grown overnight in Luria-Bertani (LB) medium at 37°C with 100 µg/mL ampicillin, 100 µg/mL spectinomycin, 5 µg/mL chloramphenicol and 0.5% xylose. They were then back-diluted 1:100 and grown in 0.5% xylose until the cells reached mid-log phase. The cells were then washed twice with LB, diluted 1:100, and grown to mid-log phase in 0.1 mM IPTG. The cells were lysed with lysozyme and detergent. Loading was normalized to the OD<sub>600</sub> at sampling. The blot was probed using affinity-purified polyclonal rabbit anti-FtsZ antibodies and goat anti-rabbit antibodies conjugated to horseradish peroxidase (Jackson ImmunoResearch Laboratories). Immunoblots were developed using the ECL Western Blotting detection reagents (GE Healthcare) and visualized with the luminescent image analyzer ImageQuant LAS 4000 mini (GE Healthcare).

### 5.3.4 Growth Curves & Immunofluorescence microscopy

Cells were grown overnight under the same media conditions as the immunoblots in 0.5% xylose, back-diluted 1:100, and grown in 0.5% xylose until the cells reached the mid-log phase. The cells were then washed twice with LB, diluted to OD<sub>600</sub> 0.004, and grown in 0.1 mM IPTG. Starting 1 hour after induction, the OD<sub>600</sub> was measured every 30 minutes for 4 hours. Immunofluorescence microscopy was performed as described previously [35]. Cells were grown using the same media conditions overnight in 0.5% xylose, back-diluted 1:100, and grown in 0.5% xylose until the cells reached the mid-log phase. The cells were then washed twice with LB, diluted 1:100, and grown in 0.1 mM IPTG for five generations (~2.5 hours). The cells were harvested and fixed with 16% paraformaldehyde/0.5% glutaraldehyde. The cells were lysed with 2 mg/mL lysozyme. FtsZ was detected with affinity-purified polyclonal rabbit anti-FtsZ serum combined with goat anti-rabbit serum conjugated to Alexa488 (Life Technologies). Cell walls were stained with wheatgerm agglutinin conjugated to tetramethylrhodamine, and DNA was stained with DAPI. Slides were visualized with an Olympus BX51 microscope with Chroma filters and a Hamamatsu OrcaERG camera and processed using Openlab version 5.2.2 (Improvision) and Adobe Photoshop CS version 8.0 (Adobe Systems). The cell length/Z-ring (L/R) ratio was calculated as described by Weart et al. 2007. The L/R ratio was calculated as the sum of the total cell length of a population of cells divided by the total number of Z-rings in that population.

### 5.3.5 All-Atom Simulations of CTT Sequence Variants

All-atom Monte Carlo simulations were performed using the ABSINTH implicit solvent model and forcefield paradigm as made available in the CAMPARI simulation package (<http://campari.sourceforge.net>) [61-63]. Simulations used were performed by Dr. Kiersten Ruff (Pappu Lab) and utilized the `abs_3.2_opls.prm` parameter set in conjunction with optimized parameters for neutralizing and excess  $\text{Na}^+$  and  $\text{Cl}^-$  ions [64]. Simulations were performed using a spherical droplet with a diameter of 285 Å with explicit ions to mimic a concentration of 10 mM NaCl. Temperature replica-exchange Monte Carlo (T-REMC) [65] was utilized to improve conformational sampling. The temperature schedule ranged from 280 K to 400 K. Ensembles corresponding to a temperature of 310 K were used in the analysis reported in this work. Three independent sets of T-REX simulations were performed for each CTT sequence. In all, the ensembles for each CTT sequence were extracted from simulations, where each T-REX simulation deploys  $4.6 \times 10^7$  Monte Carlo steps. In each simulation, the first  $10^6$  steps were discarded as equilibration. Simulation results were analyzed using the MDTraj [66] and CAMPARITraj routines that are available at <http://pappulab.wustl.edu/CTraj.html>.

### 5.3.6 Protein Purification

FtsZ variants were cloned into the pET-21b(+) expression vector through *E. coli* strain AG1111. The resulting plasmids were mini-prepped and freshly transformed into C41(DE3) cells and made into glycerol stocks. 500 mL of LB medium was inoculated 1:100 with an overnight culture. Cells were grown at 37°C until  $A_{600} \sim 0.6-0.8$ , and then the cells were induced with isopropyl IPTG to a final concentration of 1 mM. Cells were grown for an additional 4h at 37°C,

and then the cells were harvested by centrifugation, and the cell pellets were stored at  $-80^{\circ}\text{C}$ . Purification was carried out as previously described [35]. Peak fractions were analyzed by SDS-PAGE, pooled together, and dialyzed overnight in 1 L of FtsZ dialysis buffer (50 mM MES 50 mM KCl, 2.5 mM  $\text{MgCl}_2$ , 1 mM EGTA, pH 6.5). Protein preparations were concentrated, separated into aliquots, flash frozen on liquid  $\text{N}_2$ , and stored at  $-80^{\circ}\text{C}$ . Before use, FtsZ aliquots were thawed on ice, well mixed, and the concentration was quantified using Pierce 660 nm assay with tubulin as a standard (Thermo Fisher Scientific).

### **5.3.7 $90^{\circ}$ Light Scattering Assay**

Assembly reactions contained 5  $\mu\text{M}$  FtsZ in an MES buffer solution (50 mM MES, 2.5 mM  $\text{MgCl}_2$ , 1 mM EGTA, pH 6.5, with salt concentrations varying from 50-200 mM KCl as specified). Measurements were recorded every quarter of a second at  $30^{\circ}\text{C}$ . A 1-minute baseline was established before adding 1 mM GTP to the reaction. At least three trials were conducted for each salt concentration and each variant. All data were collected and exported into Microsoft Excel, and the subsequent analysis was performed using MATLAB. The average baseline was subtracted from each data point.

### **5.3.8 Transmission Electron Microscopy (TEM)**

Samples were prepared in conditions mimicking the light scattering assays, with a lower concentration of FtsZ (2.5  $\mu\text{M}$ ). Before preparing the copper grids, each sample was incubated for 10 minutes in the presence of 1 mM GTP to allow for adequate assembly. Each sample was stained three times with 2% uranyl acetate for 20-seconds each. Each staining involved wicking the solution away and waiting 10 seconds in between stains. Samples were visualized using an FEI

Company Transmission Electron Microscope (TEM). FtsZ TEM images were analyzed using ImageJ software [67].

### 5.3.9 GTPase Activity Assay

GTP hydrolysis activity was monitored using a coupled GTPase assay [68]. The assay was conducted using a 96-well plate reader in the same buffer conditions as the light scattering assay (5  $\mu$ M FtsZ WT / variant, and 1 mM GTP) and included 1 mM phosphoenolpyruvate, 250  $\mu$ M NADH, and 40 units/ml of both lactose dehydrogenase and pyruvate kinase. The linear decline of NADH absorbance at 340 nm was monitored over 30 minutes. The steepest decline rate for a 5-minute consecutive stretch was related to the GTPase activity by manipulating Beer's law [69],

which yields:  $\left( \frac{\text{moles of GTP hydrolyzed}}{\text{min}} \right) = \left( \frac{\Delta A_{340} V_a}{\epsilon_{\text{NADH}} L} \right)$ . Here,  $\Delta A_{340}$  is the slope of the decline,

$\epsilon_{\text{NADH}}$  is the extinction coefficient for NADH at 340 nm (6220  $\text{M}^{-1}\text{cm}^{-1}$ ),  $L$  is the path length of the cuvette (0.401 cm), and  $V_a$  is the observation volume (150  $\mu$ L). Each trial was performed in triplicate.

### 5.3.10 Fluorescence Correlation Spectroscopy (FCS)

TMR-labeled peptides of CTT variants were synthesized by Watson Bio. The peptides were thawed in containers with desiccant while the surfaces of the 8-well plates (0.17 +/- 0.005 mm thickness) were passivated by adding 150-200 $\mu$ L of BSA (2 mg/ml) to the wells and letting it sit for 10-15 minutes. After 15 minutes, the surfaces were washed 3-5x with 1 mL of DI water. Three stocks were generated by dilution in MES buffer (above), assumed to have the same viscosity as water: (1) 50 nM peptide, (2) 5 nM peptide, (3) 5 nM free dye. The free dye was used

to determine the hydrodynamic radius of the beam and used to calculate the hydrodynamic radius ( $R_H$ ) of the peptides per the Stokes-Einstein Equation [70]. The diffusion coefficient of TMR with the instrument setup used here is  $4.30 \times 10^{-10} \text{ m}^2 \text{ s}^{-1}$ . All data were collected on a Confocor II LSM system (Carl Zeiss-Evotec, Jena, Germany) with a  $\times 40$  water-immersion objective, similarly to the methods described previously [71]. The sample was excited at 500 nm, and emission was collected at 620 nm. Data for fluorescence intensity autocorrelation functions were analyzed with Zeiss Confocor II FCS software.

### 5.3.11 Shannon Entropy Calculation

The Shannon entropies were quantified, as described in **Chapter 3** [53]. Here, only two-parameter distributions were used with  $R_g$  and  $\delta^*$ . For each sequence-specific distribution, we tiled the shape- and size-axes into four evenly sized regions, giving rise to a total of 16 bins. The boundaries for each of the bins were computed using the maximum and minimum observed values

for the  $\frac{R_g}{\sqrt{n}}$  and  $\delta^*$  across all seven variants and segmenting each axis into four equivalent regions:

$$\frac{R_g}{\sqrt{n}} \leq 1; 1 < \frac{R_g}{\sqrt{n}} \leq 3; 3 < \frac{R_g}{\sqrt{n}} \leq 5; 5 < \frac{R_g}{\sqrt{n}} \leq 7; \text{ and } \delta^* \leq 0.25; 0.25 < \delta^* \leq 0.5; 0.5 < \delta^* \leq 0.75; 0.75 <$$

$\delta^* \leq 1$ . The Shannon entropy for each sequence variant using the following equation:

$$S = - \prod_{i=1}^b \prod_{j=1}^b p_{ij} \ln p_{ij}. \text{ Here, } b = 4 \text{ is the total number of tiles along each axis, and } p_{ij} \text{ is the probability}$$

density associated with bin  $(i,j)$ . If we assume that each bin is equally populated, then the maximum Shannon entropy is  $S_{\max} = -\ln(1/16) = 2.77$ . To compare the sequence-specific conformational

heterogeneities, we calculated relative Shannon entropies as:  $s' = \left( \frac{S}{S_{\max}} \right)$ .



## 5.4 Results

### 5.4.1 Design of C-terminal linker variants

To test the established criteria for a functional CTL (**Chapters 2-4**), we required a design parameter for *de novo* sequence scramble variants. This parameter must influence the spacer properties of the CTL, the conformational ensemble, and the non-random and random sequence patterns. The last two items will be related as the overall sizes, shapes, and amplitudes of conformational fluctuations of intrinsically disordered sequences are governed by a combination of sequence composition, the length of the IDR, and the linear patterning of residues [52]. Therefore, for fixed composition and length, systematic alteration of a linear patterning parameter is the best approach to modulate the conformational ensemble and assess the impact of these alterations on FtsZ functions and cellular phenotypes.

Despite the variability of FtsZ CTL sequences, most CTT sequences are best described as nearly symmetric polyampholytes [16]. This means that the fractions of acidic and basic residues within each sequence are similar, although not identical. In polyampholytic sequences, the linear patterning of oppositely charged residues, also known as *charge patterning* [72], can generate distinct blocks of charges or can be uniformly mixed along the sequence. This patterning, in turn, impacts the conformational properties of the domain [73]. The linear mixing / segregation of oppositely charged residues can be quantified using different parameters [73-75]. The one we use here is designated as  $\kappa$  (identical to  $\kappa_{+,-}$  from **Chapter 4**), where  $0 \leq \kappa \leq 1$  [73].

If the oppositely charged residues are well mixed, then the sequence will have low  $\kappa$  values. In contrast, the separation of oppositely charged residues into discrete blocks within the linear sequence leads to higher values of  $\kappa$ . In well-mixed sequences, preferential solvation of charged residues combined with a counterbalancing of intra-chain electrostatic repulsions and attractions will promote chain expansion. Increased linear segregation of oppositely charged residues causes systematic compaction of IDRs, leading to smaller  $R_g$  values, more spherical shapes for individual molecules, and lower amplitudes of conformational fluctuations.

Based on previous analyses in **Chapter 4**, it is clear that the  $\kappa$  values across CTT sequences drawn from 1208 orthologous FtsZ proteins [16] are mostly random. The  $\kappa$  values from over 70% of sequences fall within one standard deviation of the statistically expected value derived from the null-scramble model, shown in the Appendix (**Figure 5A.1**). This analysis also showed a clear selection against sequences with significant positive deviations from the null model. Significant positive deviations from the null-scramble model are a result of blocky charge architectures in polyampholytic sequences. IDRs with this architecture have been shown to have sticker-like properties, such as with the condensate-forming bacterial protein RNase E, described in **Chapter 4**, and also with other coacervate forming systems (i.e., DDX4 [76-78], and NICD [79]). Together, these insights suggest that alterations to the  $\kappa$  parameter will influence the conformational ensemble; specifically, we propose that high  $\kappa$  values within the CTL should compromise its ability to function as a spacer through a gain-of-function sticker mechanism.

To test the impact of changing CTT- $\kappa$  values on FtsZ function, we designed a set of sequence variants of *B. subtilis* FtsZ. In these variants, we fixed the GTPase core domain, the amino acid composition of the CTL, and the sequence of the CTP. We then shuffled the positions

of charged and neutral residues within the *B. subtilis* CTL to generate sequence variants of the CTT that span the spectrum of  $\kappa$  values between 0.1 and 0.8, keeping the CTP constant with the intent to preserve the interactome of *B. subtilis* FtsZ. There are roughly  $10^{40}$  sequences that satisfy the design criteria (**Appendix Figure 5A.2**). Of these, we selected six variants of *B. subtilis* FtsZ. Each variant is distinguished by the  $\kappa$  value for the CTT, which spans from 0.15 to 0.72 (**Appendix Figure 5A.2** and **Table 5.1**). Each of the designed *B. subtilis* FtsZ variants is designated as  $k_x$ , where  $x$  is the CTT  $\kappa$  value multiplied by 100.

FtsZ Variant	CTT-Sequence	CTT- $\kappa$
<i>B. Subtilis</i> WT	IEQEKDVTKPQRPSLNQSIKTHNQSVPKREPKREEPQQQNTVSRHTSQPADDDTLDIPTFLRNRNKRG	0.19
k15	GSHQPKPEQKSEANQSRQVTRQLHSRNVPEIQKDKVPPQSTNTSTPQRIPGDDTLDIPTFLRNRNKRG	0.15
k18	GSVQNKQIEKQKEPERRRQRTHVESSPHQPSSQPVDRNPLPKTNQTQEKASITPGDDTLDIPTFLRNRNKRG	0.18
k34	GSHQKPRQKVNKRQSEIRVPQSELSRSPTQENEEQSQPPAKTKNVQDTIHTPPGDDTLDIPTFLRNRNKRG	0.34
k40	GSHPPEQQISKTRKRHVTSQQEDEPPAKQKSLRRQNQEINVVNSSKPEETTPPGDDTLDIPTFLRNRNKRG	0.40
k46	GSPQIHQPKPQKKRSSNPTDQHKSAVTTPRKRVLIRQQQTSEVEVEESNQNPAGDDTLDIPTFLRNRNKRG	0.46
k72	GSNSQTQIRKRRKKKKRSSHQIVNLPNAPPDEEEESEPHQSQTQPTVTQVQPGDDTLDIPTFLRNRNKRG	0.72

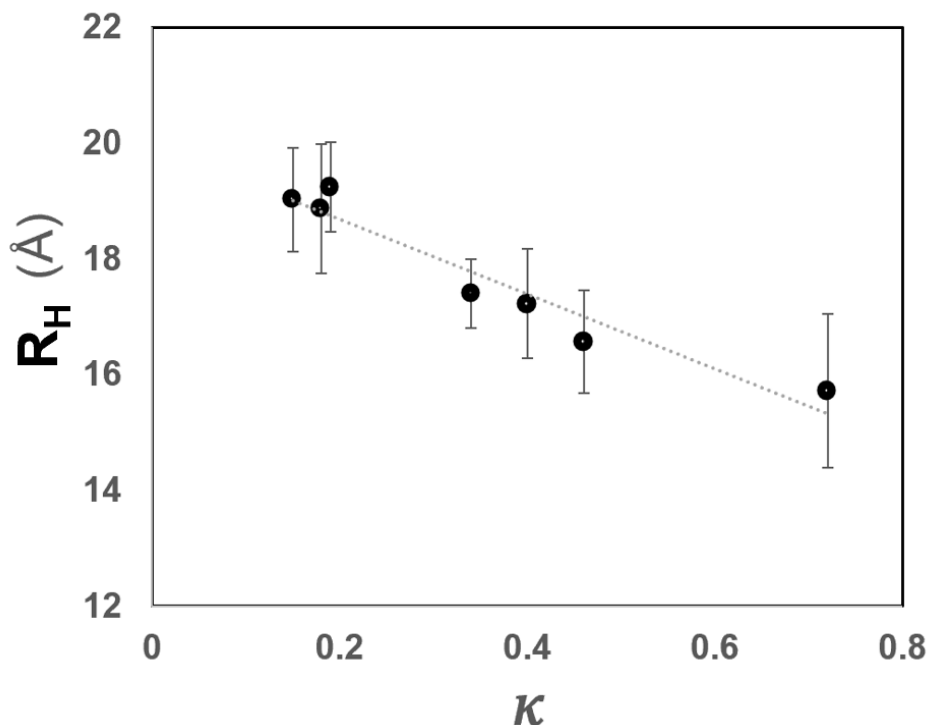
**Table 5.1: Sequences of the FtsZ CTL variants with the conserved CTP motif.** We chose six CTT sequences with  $\kappa_{+}$ -values spanning from 0.15 to 0.72 (see **Figure 5.A-3** for the design space). Column 1 shows the name that was assigned to each *B. subtilis* FtsZ variant; columns 2 and 3 show the redesigned sequence of the CTL and the  $\kappa_{+}$  value for the variant in question, respectively. Values of  $\kappa_{+}$  were calculated over the entire CTT, and this includes the CTL and CTP sequences. Variant sequences are 4 residues longer because of the cloning artifacts.

In the following three sections, we report results from our investigations of how the designed sequences influence the essential properties of *B. subtilis* FtsZs *in vitro* and *in vivo*.

#### 5.4.2 CTT- $\kappa$ influences the conformational properties of CTT sequences

To illustrate the influence of CTT- $\kappa$  on the conformations of CTT sequences, we performed Fluorescence Correlation Spectroscopy (FCS) experiments on the CTT variants. Each CTT variant was synthesized as a TMR-labeled peptide. Artifacts that are later introduced from expressing these variants recombinantly were also included in the sequence. The inferred hydrodynamic radii

( $R_h$ ) are shown in **Figure 5.1**. As  $\kappa$  increases, there is a clear trend toward compaction that follows linearly with CTT- $\kappa$  ( $R^2 = 0.94$ ).

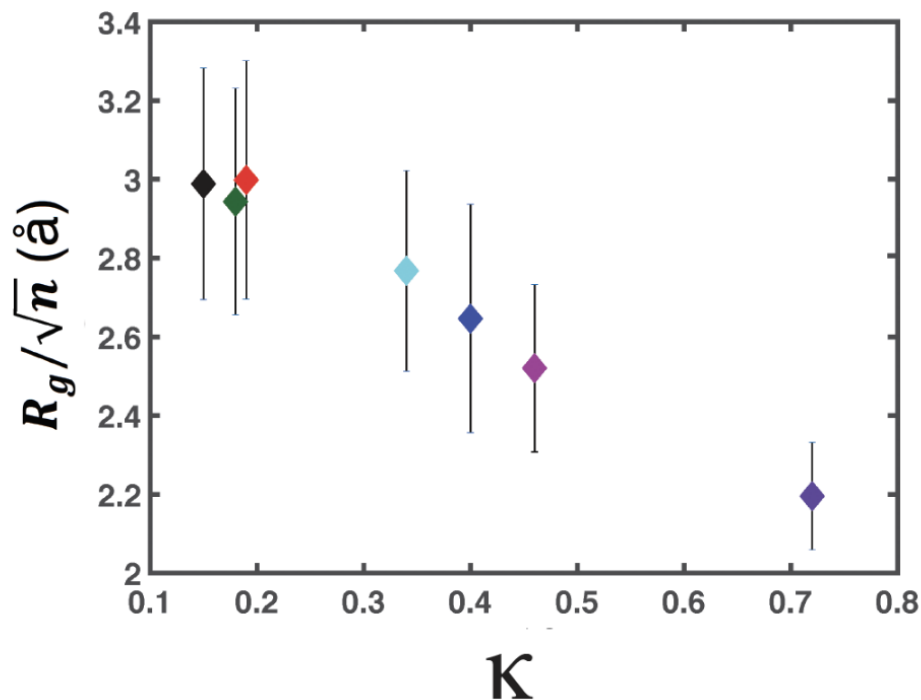


**Figure 5.1: The hydrodynamic radius ( $R_H$ ) of TMR-labeled peptides of the WT and variant CTT sequences.** These data show that the  $R_H$  is high for CTT sequences with  $\kappa$  values that are low, and we observe a monotonic decrease as  $\kappa$  values increase. These results are fit to a linear slope  $R_H(\kappa) = -6.45\kappa + 19.97$  with an  $R^2$  value of 0.94. This slope implies  $R_H$  decreases by 0.65 Å as  $\kappa$  increases by 0.1.

Using data from atomistic simulations of the WT CTT from *B. subtilis* FtsZ and the six designed variants, we quantified the preference for compact structures in terms of the length-normalized mean radii of gyration  $\frac{R_g}{\sqrt{n}}$  that are plotted for the WT CTT and designed variants

(**Figure 5.2**). Variants of the *B. subtilis* CTT sequence with  $\kappa$  values that are higher than the WT again become increasingly more compact with increasing  $\kappa$ . It follows, in accord with previous

observations for other systems [57, 59, 73], that increasing the linear segregation of oppositely charged residues engenders compaction of the IDRs.



**Figure 5.2: Plots of length-normalized values of the mean radii of gyration for each of the designed CTT variants.** These values were obtained by dividing the mean  $R_g$  values for each variant by the square root of the length ( $n$ ) of the corresponding sequence. The extra residues present in each CTL variant were a product of cloning. These data show that the length normalized  $R_g$  is high for CTT sequences with  $\kappa$  values that are low. In contrast, we observe a monotonic decrease as  $\kappa$  values increase beyond that of the WT.

The sequence-intrinsic conformational heterogeneity of the WT *B. subtilis* FtsZ CTT sequence is borne out in terms of the broad spectrum of both the sizes and shapes that this sequence

adopts. Therefore, average values of  $\frac{R_g}{\sqrt{n}}$  do not describe the influence of CTT- $\kappa$  on the overall

conformational heterogeneity, which is quantifiable in terms of the distributions of overall sizes (radii of gyration), shapes (asphericities), and amplitudes of fluctuations about different canonical and non-canonical structural motifs [59, 80-82]. As shown in **Chapter 3**, the diversity of sizes and

shapes can be quantified using the  $(\frac{R_g}{\sqrt{n}}, \delta^*)$  distributions, shown in the Appendix (**Figure 5A.3**)

and summarized in terms of the relative Shannon entropies ( $s'$ ) (**Figure 5.3**). These results show that despite doubling the  $\kappa$  value, the values of  $s'$  for CTT sequences with  $\kappa$  between 0.15 and 0.4 are bounded between 0.55 and 0.6. These bounds suggest that despite the monotonic decrease of

$\frac{R_g}{\sqrt{n}}$  with  $\kappa$ , the conformational heterogeneity is not impacted because enhancements in shape

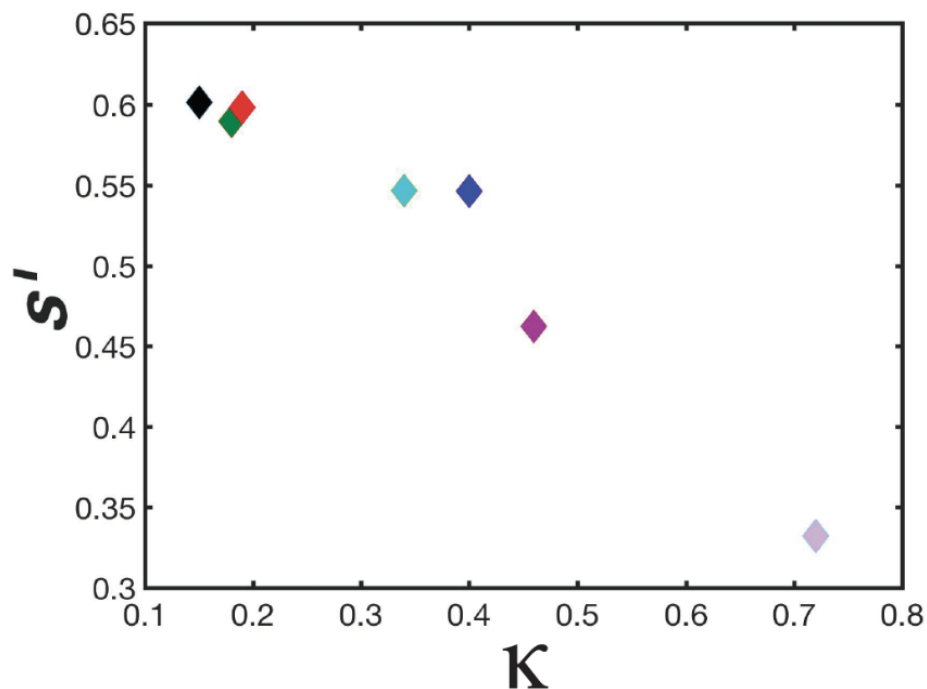
fluctuations offset any diminution of fluctuations in size. However, there is an apparent reduction

in  $s'$  as  $\kappa$  increases beyond 0.4. Thus, CTT- $\kappa$  indeed influences the sizes, shapes, and fluctuations

of the CTTs. In the absence of all other considerations, we hypothesize that sequences that

maintain WT-like conformational ensembles (k15, k18, k34, and k40) should behave the most like

WT.

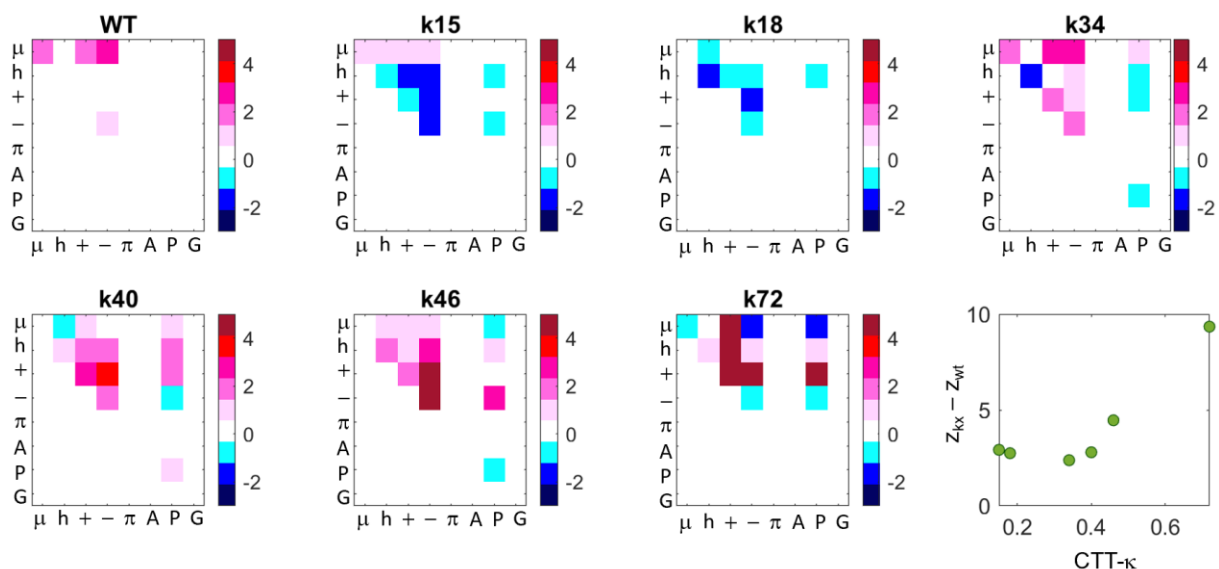


**Figure 5.3: Plot of the relative Shannon entropy ( $s'$ ) versus  $\kappa$  for the WT CTT and each of the designed CTT variants.** The sequence-specific two-dimensional histograms are found in the Appendix (**Figure 5A.3**).

### **5.4.3 Designed set of CTL variants includes random and non-random sequence features**

Since CTT- $\kappa$  impacts the sequence-ensemble relationships (SERs), we next sought to glean the effect of sequence scrambling on context-dependent sequence properties. As described in **Chapter 4**, solely interrogating the linear patterning of two residues will not paint the complete picture of the random/non-random features encoded within a sequence. Instead, there are likely multiple features at play that contribute to the conformational ensemble, function, and interactome of an IDR and the protein to which it belongs. We quantify the significance of features in terms of z-scores, the number of standard deviations, the observed parameter is away from the statically expected value for a null-scramble model. In the context of WT *B. subtilis* FtsZ, the positioning of its polar and negatively charged residues is non-random (populated colors in the matrix), implying a potential functional significance (**Chapter 4, Figure 5.4**). Using a z-score of 1.5 as the threshold set in **Chapter 4** for CTL sequences for random / non-random values, we find that k18 most closely represents a random CTL sequence. It does not have a binary patterning parameter that is considered non-random. The variant k15 has non-random features that are more scrambled than expected. As the design parameter ( $\kappa$ ) increases, the parameters involving charged residues become more non-random. The z-scores for the design parameter specifically increase monotonically with  $\kappa$ , as  $z = 2.83, 3.83, 4.83, 9.17$  for k34, k40, k46, and k72, respectively. This monotonic increase follows with the design space shown in **Figure 5A.2** as sequences with these  $\kappa$  values are less probable. Ongoing investigations will include sequences that fix the non-charged

residues and scramble only the charged residues (**Table 5A.2**). This will maintain the z-scores for non-charge related parameters and therefore help to separate charge-related impacts from non-charge related impacts.



**Figure 5.4: Z-score matrices for the CTT variants and WT FtsZ.** Despite seemingly changing one variable ( $\kappa_{+-}$ ), each variant has a unique set of random and non-random parameters that are also impacted. k18 is the closest approximation to a fully random CTL sequence, with k15 also showing mostly random sequence patterns. The bottom right figure shows the Euclidian difference between each variant z-score matrix and WT. The CTT variant k34 has the lowest difference, while k72 is a significant deviation from WT.

In the bottom right panel of **Figure 5.4**, we show the Euclidian distance between each z-score matrix for the variants and WT. The distances are calculated as  $\sqrt{\sum_{j=i}^8 \sum_{i=1}^8 (z_{wtij} - z_{xij})^2}$  and are plotted as a function of CTT- $\kappa$ . Even though k18 is the closest variant in terms of the  $\kappa$  value, when considering the entire z-score matrix, the variant k34 is the most similar to the WT CTT sequence. We also find that k34 conserves the non-random patterning of polar and negatively charged residues observed with the WT CTT. Therefore, with this design parameter, we create

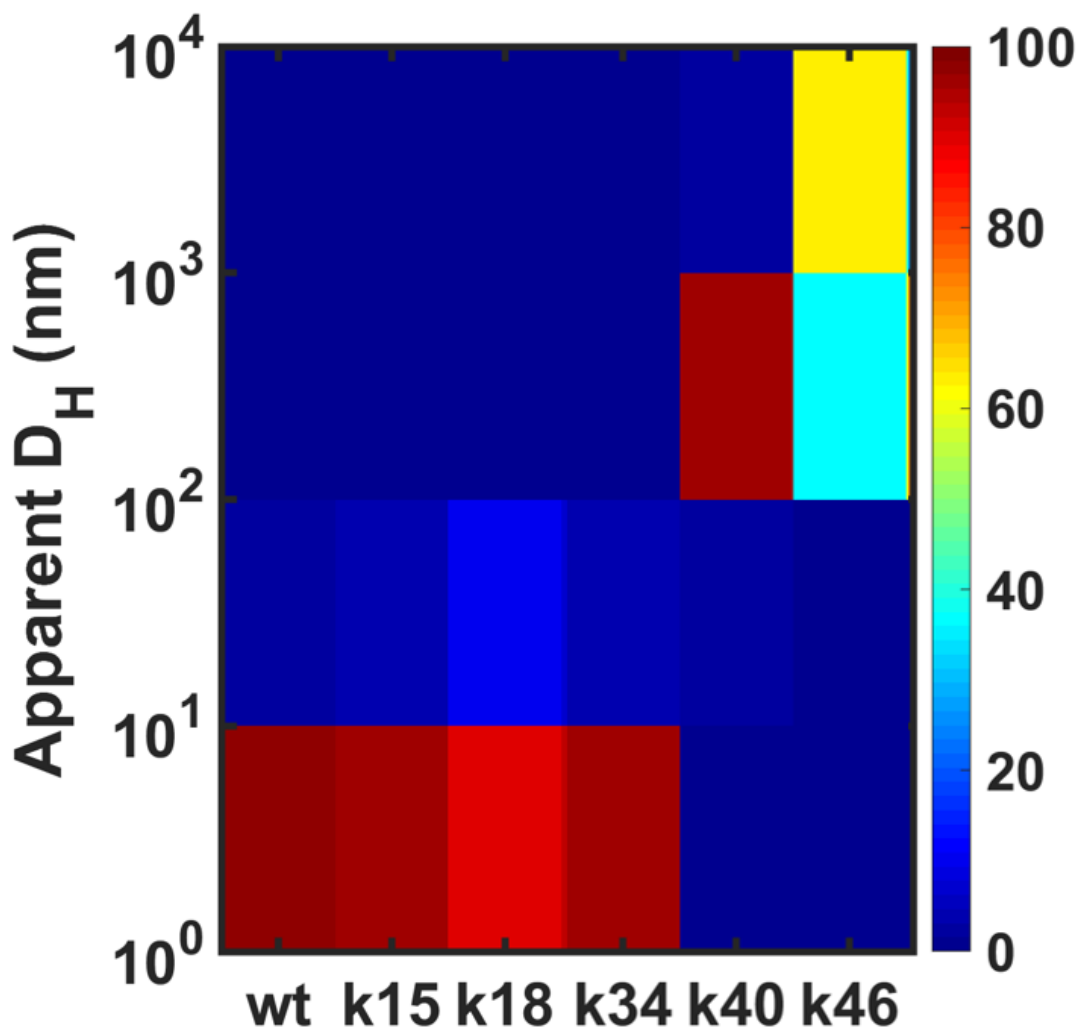


three types of sequences: (1) random (k18, k15), (2) WT-like (k34), and (3) distinct in addition to displaying non-random sequence patterns (k40, k46, and k72). From the hypothesis generated in **Chapter 4**, the WT-like sequence (k34) should then function the most like WT.

#### **5.4.4 Increasing the CTT $\kappa$ value converts the CTL into a sticker**

Next, we tested our hypothesis that segregation of oppositely charged residues into a charge-blocky architecture could turn the spacer CTL into a sticker. In previous work, we established that in the presence of GTP, the threshold concentration for forming higher-order assemblies is 2  $\mu\text{M}$  [34]. In the absence of GTP, this threshold concentration shifts up at least three-fold, where at 6  $\mu\text{M}$ , no higher-order assemblies are observed (**Figure 5.5**). Over 90% of the assemblies (% of the total number) have an apparent diameter ( $D_H$ ) between 1 and 10 nm, as measured by dynamic light scattering (DLS). This observation holds for variants k15, k18, and k34. Variants k40 and k46, however, can form assemblies that have apparent  $D_H$  values of  $10^2$  to  $10^4$  nm. This increase in assembly size indicates a gain-of-function that, in the absence of all other variables, can only be attributed to the changed features of the CTL. We propose that this gain-of-function in CTTs with high- $\kappa$  values is due to a conversion of the CTL from a spacer to a sticker, which refers to the gain of non-native interactions through the repatterned CTL. The yield of k72 was too low to perform the DLS experiments; however, experiments with the previously described new variants with similarly high  $\kappa$  values (0.53 and 0.59) also demonstrated the ability to assemble into presumed higher-order structures at 6  $\mu\text{M}$  in the absence of GTP (**Figure 5A.4**). This, along with subsequent experiments, indicates that k72 is also likely a sticker. While these variants do show a linear increase in compaction as measured by FCS (**Figures 5.1 & 5A.5**), which is a proxy

for the strength of the homotypic interactions involving the sticker, it is also likely that the repatterned sticker-like CTL variants engage in non-native, gain-of-function interactions with the core domain [54].



**Figure 5.5: Apparent hydrodynamic diameter ( $D_H$ ) of the assemblies formed by WT and the CTT variants in the absence of GTP.** Number densities of particles of each diameter were quantified from intensity distributions in DLS measurements. The results are reported as a heatmap to show the frequency of observing a particle of a certain magnitude of size for each variant. WT does not form large assemblies in the absence of GTP; however, k40 and k46 are able to form assemblies that are apparently at least two orders of magnitude larger. All experiments are performed at an FtsZ concentration of 6  $\mu$ M.

In **Table 5.2**, we summarize the results of the preceding experiments on the designed sequence variants. Features colored in red are hypothesized to have deleterious functional implications, and green features are hypothesized to impact function.

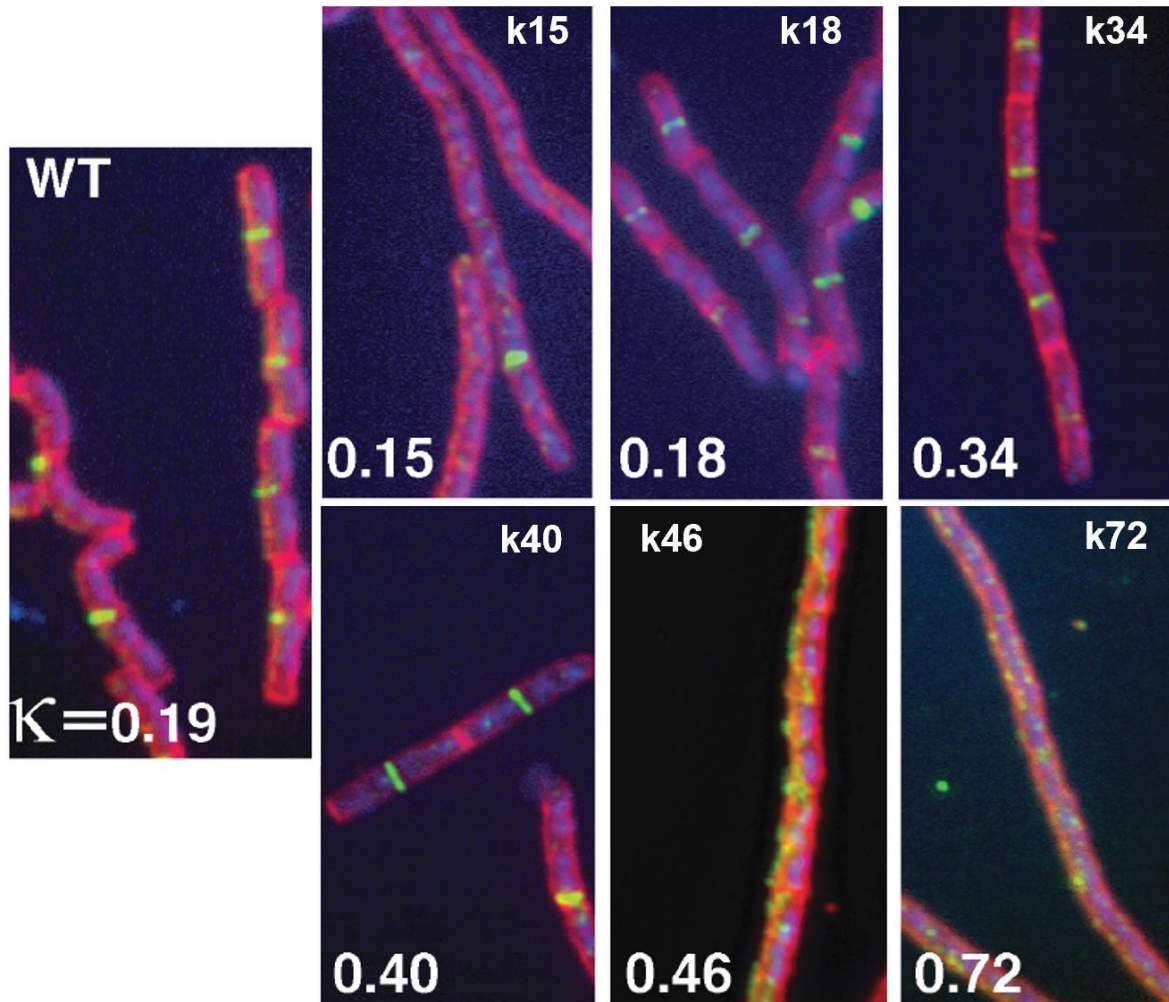
Variant	Conformational Ensemble	Non-random sequence features	Spacer/Sticker
k15	Like WT	Random	Spacer
k18	Like WT	Random	Spacer
k34	Like WT	WT-like	Spacer
k40	Like WT	Somewhat distinct	Sticker
k46	Unlike WT	Distinct	Sticker
k72	Unlike WT	Distinct	Sticker

**Table 5.2: Summary of the results of specific IDR parameters hypothesized to impact the function of *B. subtilis* FtsZ**

#### 5.4.5 CTL variants that do not meet the criteria suggested for a functional CTL disrupt cell division

To assess the impact of designed FtsZ variants on function, we used in-cell investigations to probe the impact of designed changes to CTL sequences on cell division and cell growth in *B. subtilis* (**Figure 5.6**). All variants, including WT, were expressed using the same plasmid (**Table 5A.1**). However, the protein levels of all CTT- $\kappa$  variants were less than that of WT, especially those of k15, k46, and k72 (**Figure 5A.6**). Given that all variants are expressed using the same

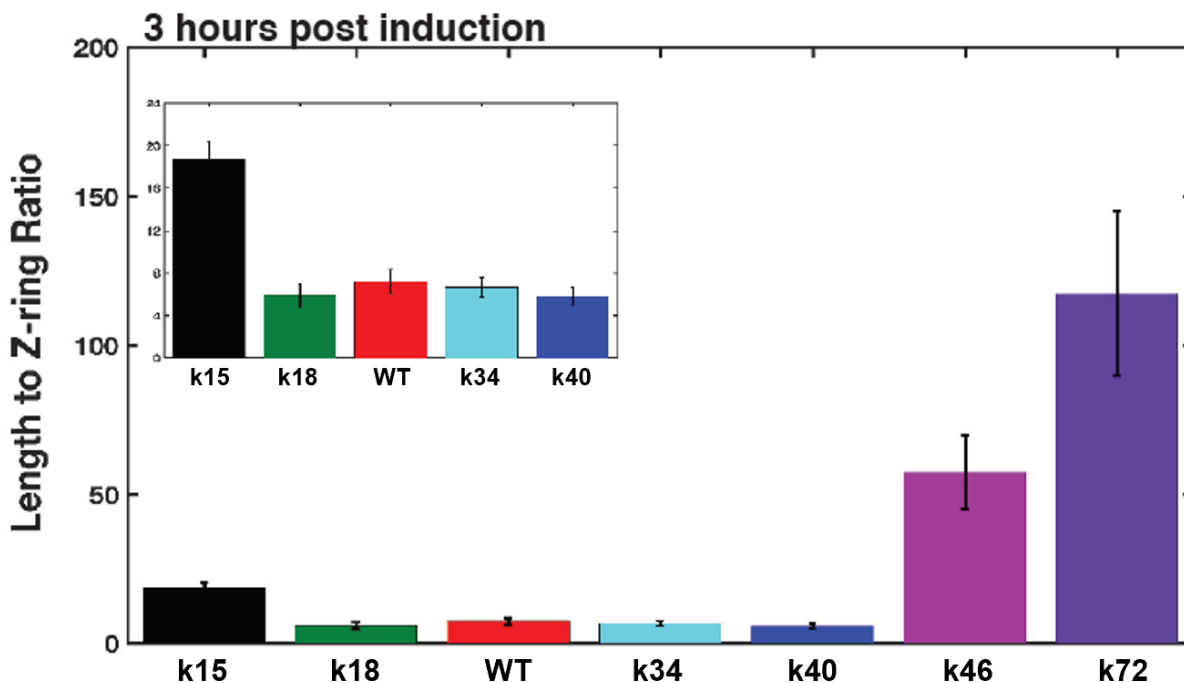
plasmid, this suggests that designed variants are prone to degradation. In cells where k15 reaches a sufficient concentration, well-defined Z-rings are formed (**Figure 5.6**). The cellular levels of k18, k34, and k40 are only slightly compromised by the designed changes to the CTL sequences (**Figure 5A.6**). Cells expressing these three FtsZ variants show robust Z-ring formation. This is true even though the  $\kappa$  values for their CTT sequences span a range from 0.18 – 0.4. However, Z-ring formation is compromised for cells expressing variants k46 and k72 (**Figure 5.6**). These results suggest that increased segregation of oppositely charged residues within the CTT sequence leads to a disruption of Z-ring formation and haphazard localization of FtsZ puncta that inhibit cell division.



**Figure 5.6: Immunofluorescence micrographs of *B. subtilis* expressing the WT FtsZ (left) versus six different variants of FtsZ with redesigned CTL sequences (shown on the right as a 2×3 grid of micrographs). Each micrograph includes the label of the variant (top) and the  $\kappa$ -value of the CTT sequence (bottom). Z-ring formation is robust for variants k18-k40. In contrast, for variants with CTT  $\kappa$ -values that are larger than 0.40 (see k46 and k72), we observe a haphazard distribution of puncta instead of robust Z-ring formation. For k15, we observed both properly formed Z-rings and diffuse distributions of FtsZ. Images courtesy of Stephen Grigsby, Levin Lab.**

To further assess the robustness / disruption of Z-ring formation, we quantified the L/R ratio, which describes the ratio of the average length of the cell to the number of Z-rings observed [83]. This ratio is a measure of the fitness and robustness of cell division in rod-shaped bacteria. The L/R ratio is  $7.2 \pm 1.2$  for cells expressing WT FtsZ. Compromised fitness is indicated by L/R

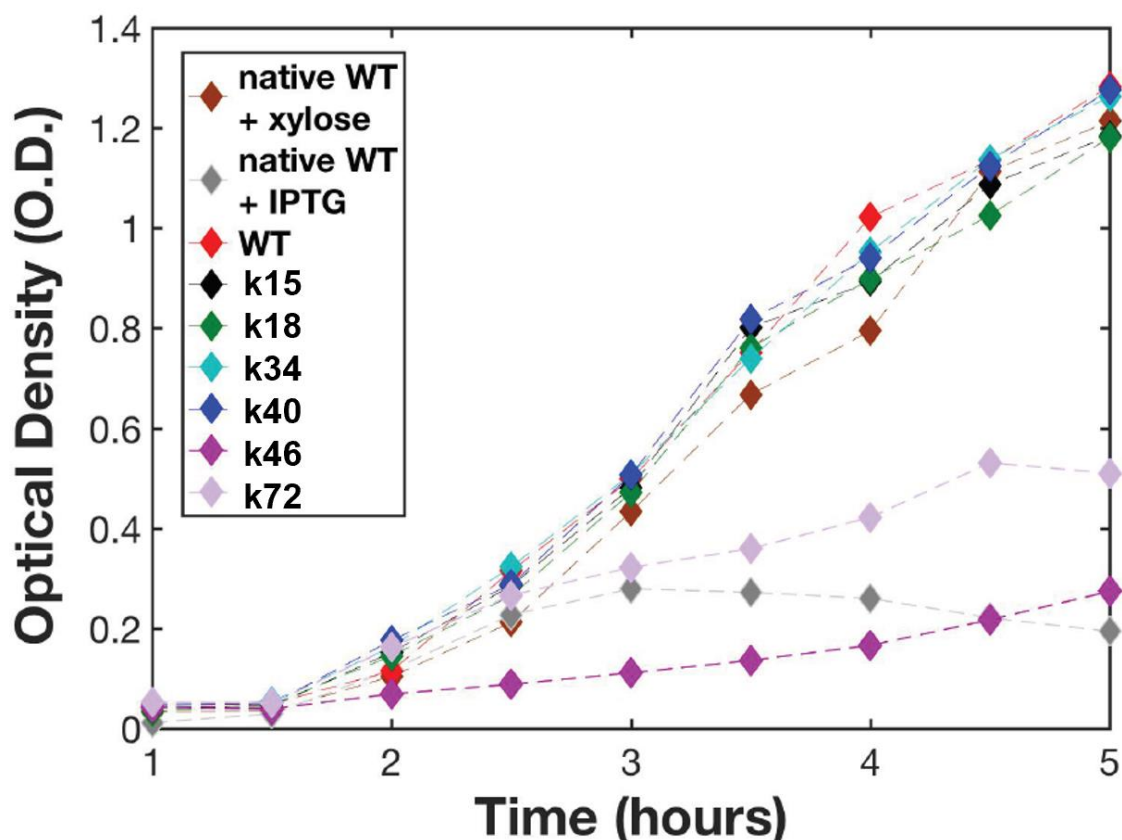
ratios that are larger than the WT. The L/R ratio is  $5.9 \pm 1.1$ ,  $6.6 \pm 1.0$ , and  $5.8 \pm 0.9$  for cells expressing variants k18, k34, and k40, respectively (**Figure 5.7**). These results suggest robust Z-ring formation and division for cells expressing FtsZ variants k18, k34, and k40. In contrast, the L/R ratio is  $18.8 \pm 1.6$ ,  $57.5 \pm 12.5$ , and  $117.3 \pm 27.6$  for cells expressing variants k15, k46, and k72, respectively.



**Figure 5.7: Length-to-Z-ring (L/R) ratio for *B. subtilis* expressing different FtsZ variants.** The value of the L/R is similar to that of WT for cells expressing variants k18, k34, and k40. The error bars quantify the standard error in our estimate of the mean L/R values — *analysis courtesy of Stephen Grigsby, Levin Lab.*

The higher L/R ratios for k15 and k72 are likely due to lowered cellular levels of these proteins (**Appendix Figure 5A.6**). Relating this to sequence patterns for k15, this could mean that non-random segregation of charges and other sequence features signal for degradation. Clues from western blots indicate that for k72, a cryptic proteolytic motif was created. However, the *in vivo*

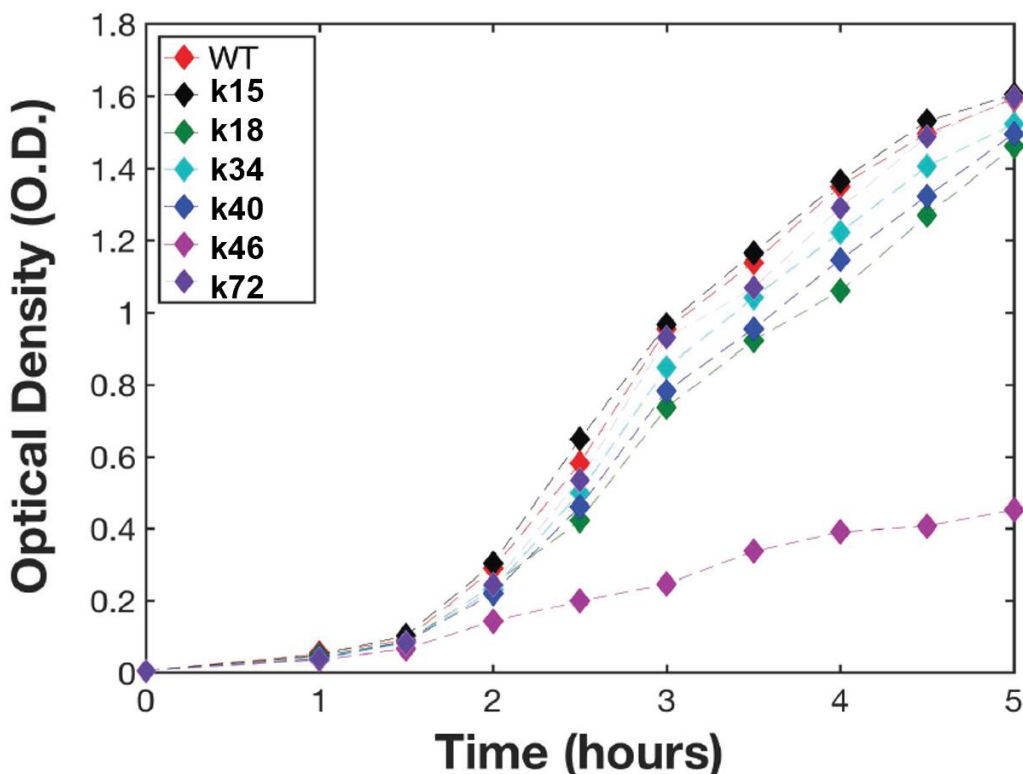
behavior we observe for cells expressing k46 cannot be explained by reduced protein levels alone. As an additional measure of cell division efficiency, we monitored the rate of change of optical density at 600 nm ( $OD_{600}$ ), which is related to the size and concentration of cells in the mid-log phase of growth (**Figure 5.8**). A reduction in  $OD_{600}$  would arise from the combination of a decrease in the efficiency of cell division, compromised metabolism, and increased cell death.



**Figure 5.8:** Data for growth phenotypes are shown in terms of the normalized optical density at 600 nm ( $OD_{600}$ ). This quantifies the concentration of viable bacteria in a suspension. *B. subtilis* expressing FtsZ variants k46 and k72 ( $\kappa > 0.4$ ) show a weak growth phenotype in contrast to all other variants that have WT profiles for growth. Bacteria expressing k72 displays a similar growth profile as the strain lacking FtsZ (native + IPTG), whereas the bacteria expressing k46 show little to no growth.

As shown in **Figure 5.8**, cells expressing the variants k15, k18, k34, and k40 showed growth rates similar to those of cells expressing WT FtsZ. In contrast, k46 and k72 show a

reduction in the increase of  $OD_{600}$ , albeit with different growth patterns. k72 follows a similar pattern of the knockdown strain (grey diamonds), in which the  $OD_{600}$  values increase at first and then decrease. This trend indicates that k72 is nonfunctional and is likely degraded. Unlike k72, k46 exhibits an almost immediate arrest in growth, suggesting that k46 might be directly toxic to the cells. Co-expression of the native WT and the CTL variant rescues the growth of k72 (**Figure 5.9**). This rescue provides additional evidence that k72 is degraded. The variant k46, however, is a dominant-negative mutation because co-expression with WT is unable to rescue growth fully. This result supports the model that k46 is directly toxic to the cells (**Figure 5.9**). Taken together, variants that significantly alter the SERs or z-score matrices with sticker-like properties did not support growth.

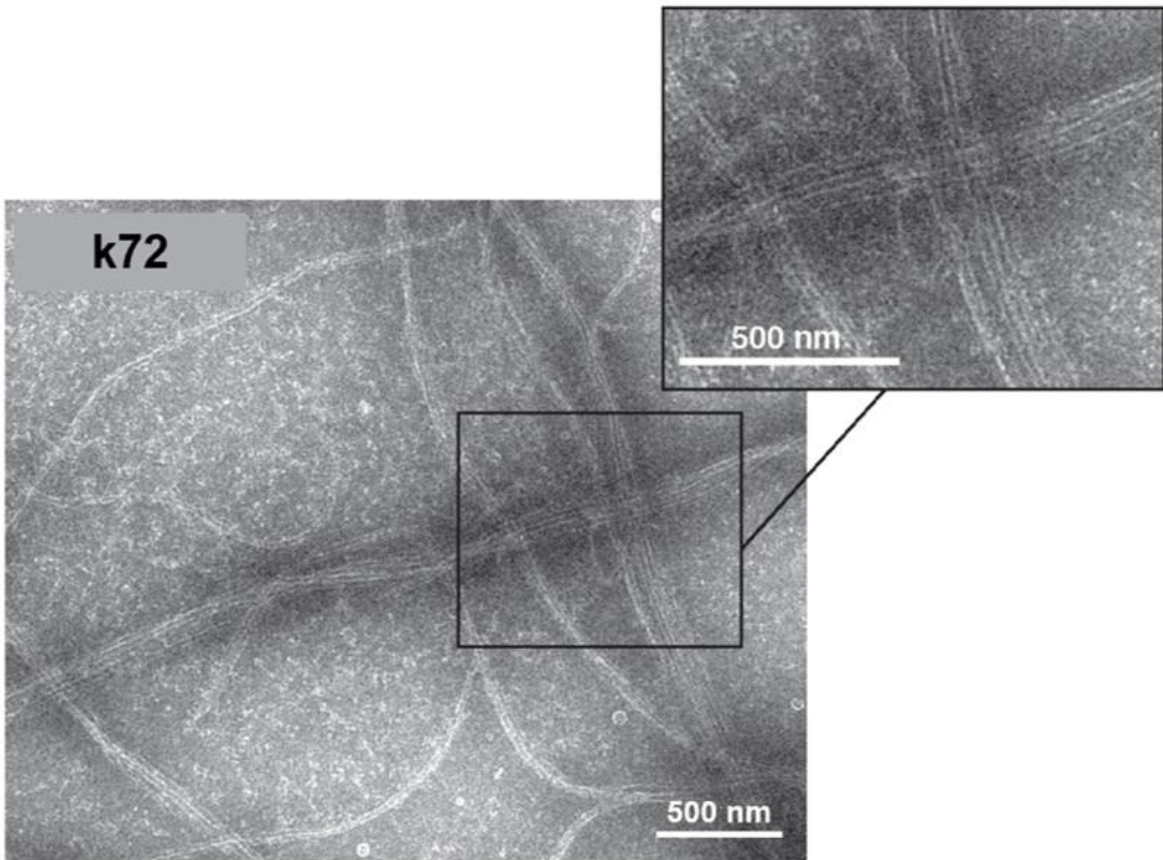
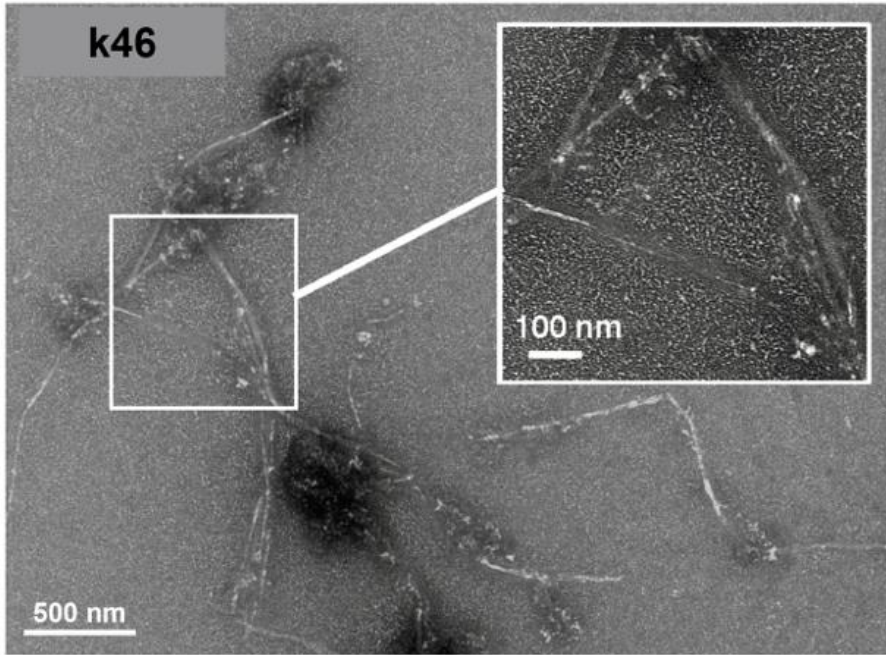




**Figure 5.9: Growth profiles of the CTT- $\kappa$  variants co-expressed with native WT FtsZ.** The variant k46 is a dominant-negative mutation and cannot be rescued by WT FtsZ, whereas k72 is rescued.

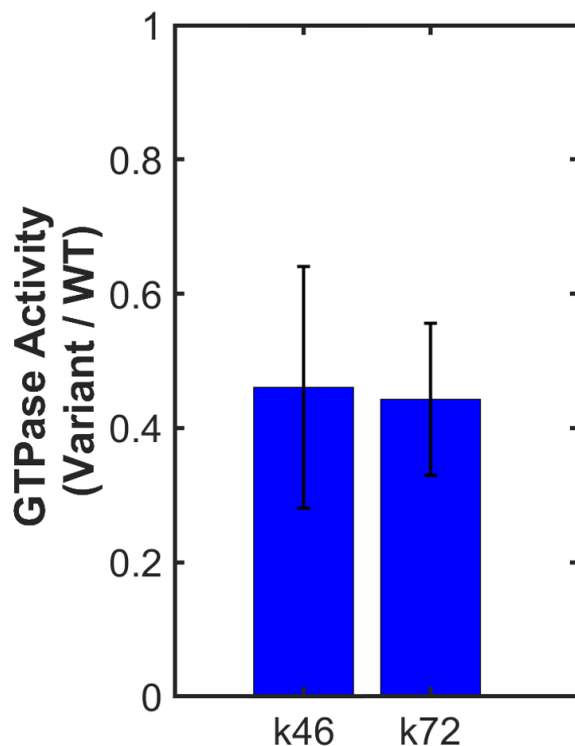
#### **5.4.6 CTL variants that have sticker-like properties also disrupt function *in vitro***

Given the deleterious phenotypic impacts of k46 and k72, we asked if these variants impacted FtsZ activity and polymerization *in vitro*. From negative-stain transmission electron microscopy (TEM) images, we see that in the presence of GTP, both k46 and k72 form large linear filaments that are laterally associated in a manner unlike WT FtsZ (**Figure 5.10**). Of the designed variants, these variants have the most extreme degree of segregation of oppositely charged residues. Increased segregation of oppositely charged residues within the CTT creates a sticker that drives increased self-association of FtsZ via inter-tail interactions among filaments. These interactions generate profoundly different assemblies than the morphologies observed for wild-type FtsZ (**Chapter 2**, [13, 14, 34, 84-87]).



**Figure 5.10: TEM images of morphologies obtained for k46 and k72.** Increased segregation of oppositely charged residues leads to the formation of linear filamentous “tracks” characterized by inter-filament interactions likely involving the sticker CTTs.

Additionally, the GTPase activities of these variants at 6  $\mu\text{M}$  FtsZ were significantly reduced to approximately half of the WT level (**Figure 5.11**). This decrease is similar to the reduction of catalytic rate ( $k_{cat}$ ) observed for the  $\Delta\text{CTL}$  variant in **Chapter 2**, in which the CTP sticker influences the catalytic domain. A reduced catalytic rate is indicative of reduced rates of subunit turnover for assemblies formed by FtsZ variants with high  $\kappa$ /sticker-like CTT sequences. However, while a GTP concentration of 1 mM is assumed to be significantly above twice the  $K_M$ , the Michaelis-Menten parameters of the variants have yet to be quantified.

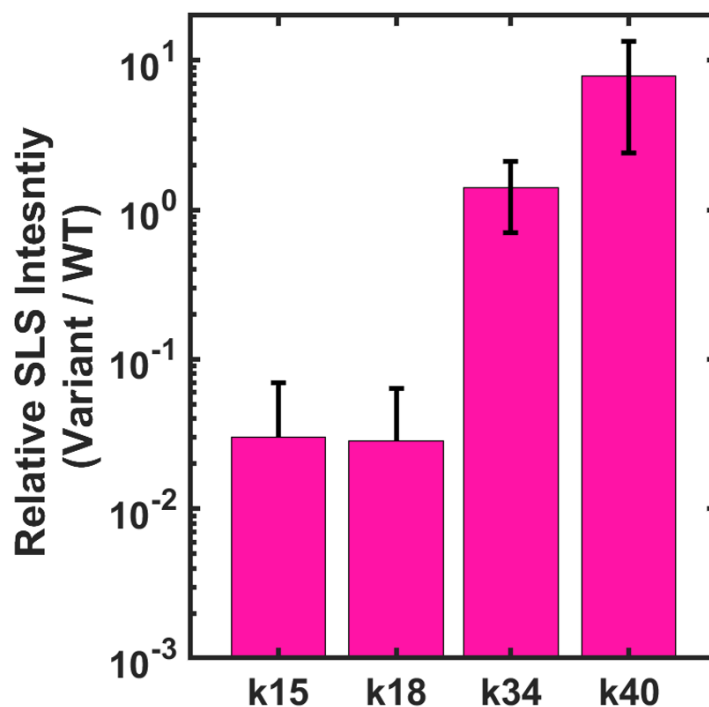


**Figure 5.11: Relative GTPase Activities of the CTT variants that have phenotypic impacts.** k46 and k72 have significantly reduced activities (< 50%) with respect to WT.

Right-angle light scattering is a sensitive method for studying FtsZ polymerization and bundling of FtsZ polymers in the presence of GTP [83]. Despite the apparent larger morphological sizes, the light scattering intensity of k72 and k46 is slightly lower than WT (**Appendix Figure 5A.7**). One possible explanation for this is that the assemblies become insoluble and, therefore, undetected by light scattering.

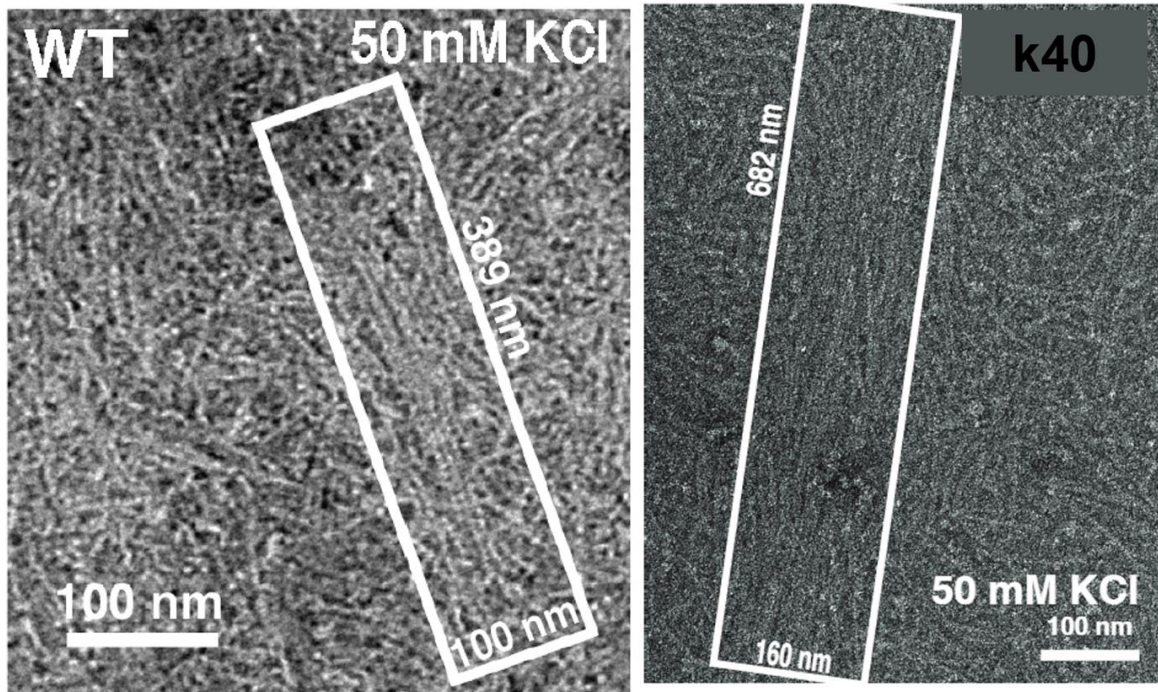
#### **5.4.7 Linear sequence patterning can explain discrepancies between spacer CTLs**

In the variants that supported growth, we asked whether they showed any deviations from WT *in vivo*. We repeated the SLS experiment performed with k46 and k72 to quantify the relative SLS intensities with respect to WT in the presence of GTP (**Figure 5.12**). Relative SLS intensities showed two different types of assemblies that corresponded with the random / non-random feature assessment of the variants. Random CTL sequences (k15 and k18) had significantly reduced assembly with respect to WT. The WT-like sequence (k34) had a similar assembly to WT.



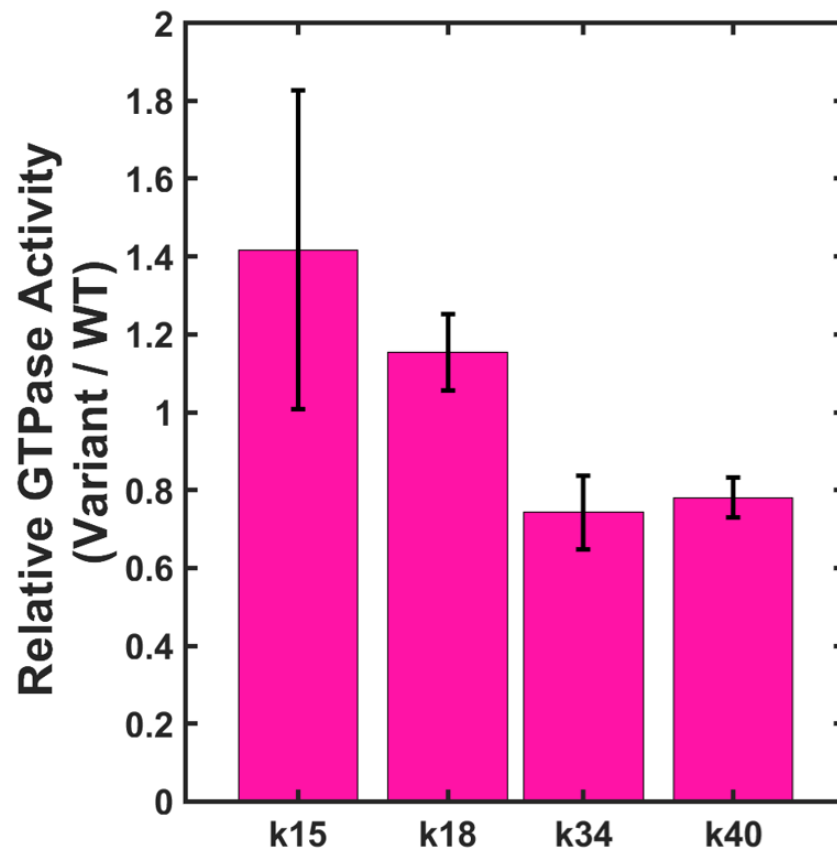
**Figure 5.12: Relative SLS intensities of the CTT variants that supported bacterial cell division.** The fully random sequences k15 and k18 have higher GTPase activities than WT, whereas k34 and k40 have reduced activities with respect to WT.

The variant k40, which showed sticker-like properties and divergent/distinct patterning, had a significantly higher SLS intensity. Simulations predict the CTT of k40 is 10% smaller than the WT CTT (**Figure 5.2**). Compaction is mediated by increased electrostatic interactions between charged blocks on the CTT sequence. While the simulations illustrate charge-dependent intramolecular interactions, these interactions could also occur in an intermolecular fashion and relate to morphological observations. Compared to the WT FtsZ, k40 forms longer polymers that become part of thicker bundles (**Figure 5.13**). At higher salt concentrations, the morphological differences between k40 and WT FtsZ are minimal. Both k40 and WT FtsZ form short single-stranded protofilaments, indicating that these enhancing interactions are screened with increased salt concentrations (**Figure 5A.8**). This results in a similar SLS intensity (**Figures 5A.9**).



**Figure 5.13: Negative stain EM images of k40.** These images show that the structures formed by k40 (right) are similar to those formed as a precursor to the Z-ring for WT (left). The white box with dimensions 389 nm by 100 nm for WT and 682 nm by 160 nm for k40 outlines the bundled protofilaments. The observed bundles are longer and wider for k40.

The relative GTPase activity of these variants showed the opposite effect of the SLS data (Figure 5.14). For k15 and k18, the GTPase activity was higher than WT. For k34 and k40, the GTPase activity was lower than WT. This result is a convolution of many potential variables, including assembly state and CTL-dependent variations to Michaelis-Menten parameters. However, this indicates that at the same FtsZ concentration, sequence-encoded features within the CTL alter the function of the GTPase core domain. Overall, these results suggest that the linear sequence patterning, the conformational ensemble, and the overall behavior of the domain (i.e., sticker versus spacer) impact function. Each variable must be considered when designing a functional CTL.



**Figure 5.14: Relative GTPase Activities of the CTT variants that supported bacterial cell division.** The fully random sequences k15 and k18 have higher GTPase activities than WT, whereas k34 and k40 have reduced activities with respect to WT.

## 5.5 Discussion

FtsZ plays an essential role in scaffolding the cell division machinery, and this is accomplished in part by the contributions of its disordered C-terminal tail (CTT). The CTT contains a sticker-spacer architecture whereby the disordered C-terminal linker (CTL) spacer connects the C-terminal peptide (CTP) sticker to the GTPase domain. Together, these domains impart an auto-regulatory function on the enzyme [34]. While the CTP is well conserved across a system of FtsZ orthologs, the CTL is hypervariable, meaning the length, sequence composition, and linear order of the sequence varies significantly. As the functional significance of the CTL continues to be highlighted, efforts to understand the sequence grammar of the hypervariable domain have grown. Previous work converged upon a hypothesis that a functional CTL sequence must act as a spacer, must preserve its conformational ensemble, and incorporates random as well as non-random features within the sequence (**Chapters 2 – 4**).

To test these criteria for the CTL of *B. subtilis* FtsZ, we utilized an approach driven by sequence design and guided by physical principles underlying SERs, sticker-spacer architectures, and the molecular sequence patterns of IDRs [73]. Altering sequence parameters, such as length, composition, and the patterning of residues, will affect the conformational ensembles and, therefore, the functions of IDPs / IDRs. Considerable attention has focused on the lengths of CTLs and their impact on FtsZ functions [50, 83, 86, 88-91]. However, titrating length creates a multivariable problem where the composition and patterning of residues are also varied, therefore clouding the definitive relationship between length and function. Here, we pursued a different approach. By maintaining composition and length, we focused instead on a single sequence



parameter  $\kappa$  that directly modulates the interplay between local and non-local interactions and hence intra- as well as intermolecular interactions involving the CTT. We summarize the results from these experiments in **Table 5.3** below, which is an expansion of **Table 5.2**.

Variant	Conformational Ensemble	Sequence features	Spacer/ Sticker	Growth	Expression	GTPase Activity	Assembly
k15	Like WT	Random	Spacer	Like WT	Reduced	Higher	Smaller
k18	Like WT	Random	Spacer	Like WT	Similar	Higher	Smaller
k34	Like WT	Like WT	Spacer	Like WT	Similar	Slightly lower	Like WT
k40	Like WT	Partially distinct	Sticker	Like WT	Similar	Slightly lower	Larger
k46	Unlike WT	Distinct	Sticker	Compromised	Similar	Lower	Distinct
k72	Unlike WT	Distinct	Sticker	Compromised	Reduced	Lower	Distinct

**Table 5.3: Summary of results from all experiments on FtsZ CTL variants compared to WT.**

From our simulations, we find a clear diminution of conformational heterogeneity as quantified by the reduction in Shannon entropy values for CTT sequences with  $\kappa > 0.4$  (**Figure 5.3**). Our experiments show that this has biological and biophysical consequences. Sequences with higher- $\kappa$  values have intermolecular interactions that are so strong we observe FtsZs that can assemble into higher-order assemblies in the absence of GTP (**Figure 5.4**). We consider this gain-of-function to be the creation of stickers within the CTT. In the presence of GTP, large linear polymers are observed with inter-polymeric tracks that are consistent with potential stabilization via tail-to-tail associations (**Figure 5.10**). The GTPase activity of the core is also significantly

reduced with the swap of the k46 and k72 CTLs for the WT CTL, which indicates a potential reduction of subunit turnover within these structures (**Figure 5.11**). Borgia et al., have recently reported these types of strong, charge-driven interactions for polyelectrolytic IDPs, where oppositely charged IDPs form binary, albeit disordered, high-affinity complexes [92].

Charge-based interactions also drive phase transitions via complex coacervation demonstrated for the intrinsically disordered Nephlin intracellular domain [93]. The driving forces for phase separation increase multi-fold for variants of NICD with long blocks of charged residues that are considered stickers. Similar results were observed by altering the patterning of oppositely charged residues within the IDR of DDX4, a protein that forms Nuage bodies [94]. Theories show that the stickers, or driving forces for phase separation, become stronger with increased separation of charged residues [77, 95, 96]. Taken together, our results suggest that the lack of CTT sequences that have a non-random segregation of charged residues comes from a functional drive to prevent the conversion of the spacer CTL to a sticker. This helps minimize associations that lead to strong albeit aberrant multivalent interactions amongst segregated blocks of charge within CTTs. Design and investigations of additional sequence variants that impact the conformational ensemble via alternative intrachain interactions that are not dominantly charge-driven (i.e., polar or polyglutamine tracts [97, 98]) will separate the impact of the change in the conformational ensemble from the creation of a sticker to dissect the contribution of each variable to function.

For FtsZ variants with lower CTT- $\kappa$  values, we find that the conformational ensembles are similar to WT, and these sequences are mostly complementary to WT function *in vivo* (k15, k18, k34, and k40). However, some discrepancies both *in vivo* and *in vitro* cannot be explained by conformational ensembles alone. Instead, we turn to the molecular sequence grammar and the

corresponding random / non-random patterning to further explain the differences (**Figure 5.4**). The variant that behaves most like WT in all of the experiments is k34. This variant has the closest z-score matrix to WT, where the non-random and random features are largely conserved. Particularly, the patterns of polar and negative residues within the CTL of k34 are, like WT, also non-random. The ability of k34 to complement WT function points to the potential contribution of polar and negative residues in the multitude of interactions between the CTP, the CTL, the GTPase core, and the FtsZ interactome. Ongoing simulations of the CTL variants in the context of the FtsZ monomer (CTL tethered to the core), the FtsZ dimer (CTLs tethered to two conjoined monomers), and the FtsZ protofilament (CTLs in a three-mer) could shed light on the contributions of cryptic sequence patterns to the SERs of CTLs that are influenced by the excluded volume and surface potential of one or more FtsZ cores. These results could further explain the functional significance of the non-random patterning of polar and negatively charged residues.

The variants k15 and k18 are also interesting as they represent negative deviations from the random expectation (more scrambled) and fully random CTL sequences, respectively. In experiments conducted in *B.subtilis*, the presence of k15 is sparse (**Figure 5A.6**). This scarcity leads to an overall compromised ability to form Z-rings due to a lack of availability. Still, the k15 FtsZ maintains the ability to divide correctly when present, as shown by the growth curves (**Figures 5.6 – 10**). Considering the sequence patterns, one could hypothesize that randomly mixed sequences lack complexity, and such sequences are targeted for degradation. We note that similar proteolytic biases have been observed in eukaryotic sequences [99]. These variants also behave distinctly from WT *in vitro*, as the relative assembly size is reduced by over an order of magnitude (**Figure 5.12**). These relative assembly sizes are reminiscent of those formed by the  $\Delta$ CTP

construct [34], implying that random CTLs might create too strong of a spacer and impact the ability of the CTP to function as a sticker. Despite the small size, GTPase activity is observed for these variants, and we, therefore, deduce that these variants are still able to form active protofilaments, given the requirement of a dimer for hydrolysis (**Figure 5.14**). Like the  $\Delta$ CTP construct, the relative GTPase activity at a single concentration is higher than WT. This increase could mean that random CTLs that act as strong spacers weaken the ability of the CTT to function as an auto-inhibitor. However, this result is a convolution of many potential variables associated with the enzyme activity, including assembly state, ligand binding affinity, concentration-dependent catalytic rate, and more. Further experiments will seek to deconvolve these contributions and determine the impact of random CTLs on the catalytic domain of FtsZ.

With variant k40, we observe multiple different features at play that can be considered to create a “tug-of-war” effect. While the conformational ensemble is similar to WT, there is a slight increase in the overall compaction that corresponds to an increase in the intramolecular interactions within the domain. We observe a gain-of-function whereby k40 can assemble into higher-order structures in the absence of GTP exhibiting sticker-like behavior, as seen with k46 and k72 (**Figure 5.5**). However, when GTP is added, core-driven polymerization wins out, essentially pulling the “rope” towards conventional linear and GTP-driven polymerization (**Figure 5.10**). This translates to a complementary CTL variant that supports Z-ring formation and bacterial growth *in vivo* (**Figure 5.6 – 10**). However, the increased interaction strength still contributes to the *in vitro* assembly mechanisms. We find that structures formed by k40 are significantly larger than WT, as it forms long and highly bundled protofilaments. Bundled protofilaments are thought to be a precursor of the Z-ring, and similar morphologies are also observed with WT [14, 83] (**Figure**

**5.13).** Increasing the KCl concentration in the buffer reduces the contribution of CTT-mediated interactions in k40, and WT and k40 similarly show single-stranded protofilament morphologies (**Figure 5A.8** and **5A.9**). Therefore, while the CTT of k40 can be involved in enhanced tail-mediated interactions, especially at low salt, this may not impact polymerization or bundling of k40 in the high physiological salt conditions of the *B. subtilis* bacterial cell [100]. These data provide a potential explanation for the observed robustness of Z-ring formation and cell division for cells expressing k40.

In the work that established the essentiality of a disordered CTL, Buske and Levin included a randomly scrambled CTL that was intended to preserve variables such as the length, the composition, and the apparent flexibility of the disordered domain [83]. While Buske and Levin also included other sequences in this mutational study, these sequences varied in composition and, therefore, do not meet the requirements for assessment here. The CTT- $\kappa$  value of the scrambled sequence (called Scr) was near that of wild type, as expected for a randomly generated sequence. Using the data presented in this study, we find that the relative activity and apparent assembly size of this variant were approximately 0.70 and  $5 \times 10^{-1}$ , respectively [83]. Bundles, but not rings, are also observed *in vitro*. We infer from these results that the variant is similar, but not identical, to the wild type. Indeed, the distance between the Scr patterning matrix and the wild type patterning matrix is the lowest ( $z_{\text{scr}} - z_{\text{wt}} = 1.95$ ). However, the non-random patterning of polar and negative is not preserved in this scramble (**Figure 5A.10**). Whether or not this contributes to the *in vitro* discrepancies, especially those observed in morphology and GTPase activity, could be the subject of further investigations. Together with other analyses in this chapter, it also brings to bear the

question of the involvement of the CTL properties in protofilament stiffness and subunit exchange dynamics.

Our findings substantiate the hypothesis that not all disordered sequence scrambles of CTL of the same length can be interoperable within *B. subtilis* FtsZ. Along with composition and length, the design of CTL variants must consider the strength of the spacer, the non-random / random sequence patterns, and the conformational ensemble. Our findings suggest that the CTT / CTL encoded effects must be just right because the CTT plays multiple regulatory roles. Similar results have been uncovered from a deep mutational scan of the transactivation domain of Gcn4, which is an essential transcription factor in yeast [58]. Our investigations open the door for systematic high-throughput experiments guided by recent computational advances [101] that enable the design of CTT sequences with different lengths, amino acid compositions, and patterning. These methods can be applied to other enzymes with N- and C-terminal IDRs to uncover the combination of sequence features that contribute to the functionalities influenced by IDPs / IDRs.

## 5.6 References

1. Straight, A.F. and C.M. Field, *Microtubules, membranes and cytokinesis*. Curr Biol, 2000. **10**(20): p. R760-70.
2. Margolin, W., *Spatial regulation of cytokinesis in bacteria*. Curr Opin Microbiol, 2001. **4**(6): p. 647-52.
3. Tolliday, N., N. Bouquin, and R. Li, *Assembly and regulation of the cytokinetic apparatus in budding yeast*. Curr Opin Microbiol, 2001. **4**(6): p. 690-5.
4. Huckaba, T.M. and L.A. Pon, *Cytokinesis: rho and formins are the ringleaders*. Current Biology, 2002. **12**(23): p. R813-4.
5. Baluska, F., D. Menzel, and P.W. Barlow, *Cytokinesis in plant and animal cells: endosomes 'shut the door'*. Dev Biol, 2006. **294**(1): p. 1-10.
6. Guizetti, J. and D.W. Gerlich, *Cytokinetic abscission in animal cells*. Semin Cell Dev Biol, 2010. **21**(9): p. 909-16.
7. Rincon, S.A. and A. Paoletti, *Molecular control of fission yeast cytokinesis*. Semin Cell Dev Biol, 2016. **53**: p. 28-38.
8. Thieleke-Matos, C., et al., *Emerging Mechanisms and Roles for Asymmetric Cytokinesis*. Int Rev Cell Mol Biol, 2017. **332**: p. 297-345.
9. Addi, C., J. Bai, and A. Echard, *Actin, microtubule, septin and ESCRT filament remodeling during late steps of cytokinesis*. Current Opinion in Cell Biology, 2018. **50**: p. 27-34.
10. Lutkenhaus, J., *FtsZ ring in bacterial cytokinesis*. Mol Microbiol, 1993. **9**(3): p. 403-9.
11. Rothfield, L., S. Justice, and J. Garcia-Lara, *Bacterial cell division*. Annu Rev Genet, 1999. **33**: p. 423-48.
12. Janakiraman, A. and M.B. Goldberg, *Recent advances on the development of bacterial poles*. Trends Microbiol, 2004. **12**(11): p. 518-25.
13. Erickson, H.P., D.E. Anderson, and M. Osawa, *FtsZ in bacterial cytokinesis: cytoskeleton and force generator all in one*. Microbiol Mol Biol Rev, 2010. **74**(4): p. 504-28.

14. Huang, K.H., J. Durand-Heredia, and A. Janakiraman, *FtsZ ring stability: of bundles, tubules, crosslinks, and curves*. J Bacteriol, 2013. **195**(9): p. 1859-68.
15. Meier, E.L. and E.D. Goley, *Form and function of the bacterial cytokinetic ring*. Current Opinion in Cell Biology, 2014. **26**: p. 19-27.
16. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division*. Seminars in Cell and Developmental Biology, 2015. **37**: p. 3-10.
17. Bisson-Filho, A.W., et al., *Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division*. Science, 2017. **355**(6326): p. 739-743.
18. Monteiro, J.M., et al., *Peptidoglycan synthesis drives an FtsZ-treadmilling-independent step of cytokinesis*. Nature, 2018. **554**(7693): p. 528-532.
19. Anderson, D.E., F.J. Gueiros-Filho, and H.P. Erickson, *Assembly dynamics of FtsZ rings in Bacillus subtilis and Escherichia coli and effects of FtsZ-regulating proteins*. Journal of Bacteriology, 2004. **186**(17): p. 5775-81.
20. Bi, E.F. and J. Lutkenhaus, *FtsZ ring structure associated with division in Escherichia coli*. Nature, 1991. **354**(6349): p. 161-4.
21. Mukherjee, A. and J. Lutkenhaus, *Guanine nucleotide-dependent assembly of FtsZ into filaments*. Journal of Bacteriology, 1994. **176**(9): p. 2754-8.
22. Ruiz-Martinez, A., et al., *Efficient Multiscale Models of Polymer Assembly*. Biophys J, 2016. **111**(1): p. 185-96.
23. Chen, Y., et al., *A rapid fluorescence assay for FtsZ assembly indicates cooperative assembly with a dimer nucleus*. Biophys J, 2005. **88**(1): p. 505-14.
24. Chen, Y. and H.P. Erickson, *Rapid in vitro assembly dynamics and subunit turnover of FtsZ demonstrated by fluorescence resonance energy transfer*. J Biol Chem, 2005. **280**(23): p. 22549-54.
25. Oliva, M.A., D. Trambaiolo, and J. Lowe, *Structural insights into the conformational variability of FtsZ*. Journal of Molecular Biology, 2007. **373**(5): p. 1229-42.
26. Miraldi, E.R., P.J. Thomas, and L. Romberg, *Allosteric models for cooperative polymerization of linear polymers*. Biophysical Journal, 2008. **95**(5): p. 2470-86.



27. Huecas, S., et al., *Energetics and geometry of FtsZ polymers: nucleated self-assembly of single protofilaments*. Biophys J, 2008. **94**(5): p. 1796-806.
28. Miraldi, E.R., P.J. Thomas, and L. Romberg, *Allosteric models for cooperative polymerization of linear polymers*. Biophys J, 2008. **95**(5): p. 2470-86.
29. Romberg, L., M. Simon, and H.P. Erickson, *Polymerization of FtsZ, a Bacterial Homolog of Tubulin: IS ASSEMBLY COOPERATIVE?* Journal of Biological Chemistry, 2001. **276**(15): p. 11743-11753.
30. Wagstaff, J. and J. Lowe, *Prokaryotic cytoskeletons: protein filaments organizing small cells*. Nat Rev Microbiol, 2018. **16**(4): p. 187-201.
31. Wagstaff, J.M., et al., *A polymerization-associated structural switch in FtsZ that enables treadmilling of model filaments*. MBio, 2017. **8**(3).
32. Caplan, M.R. and H.P. Erickson, *Apparent cooperative assembly of the bacterial cell division protein FtsZ demonstrated by isothermal titration calorimetry*. Journal of Biological Chemistry, 2003. **278**(16): p. 13784-8.
33. Chen, Y., et al., *A rapid fluorescence assay for FtsZ assembly indicates cooperative assembly with a dimer nucleus*. Biophysical Journal, 2005. **88**(1): p. 505-14.
34. Cohan, M.C., et al., *Dissecting the Functional Contributions of the Intrinsically Disordered C-terminal Tail of Bacillus subtilis FtsZ*. J Mol Biol, 2020. **432**(10): p. 3205-3221.
35. Buske, P.J. and P.A. Levin, *Extreme C terminus of bacterial cytoskeletal protein FtsZ plays fundamental role in assembly independent of modulatory proteins*. Journal of Biological Chemistry, 2012. **287**(14): p. 10945-57.
36. Buske, P.J. and P.A. Levin, *A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo*. Molecular Microbiology, 2013. **89**(2): p. 249-63.
37. Cohan, M.C. and R.V. Pappu, *Making the Case for Disordered Proteins and Biomolecular Condensates in Bacteria*. Trends Biochem Sci, 2020. **45**(8): p. 668-680.
38. Martin, E.W., et al., *Valence and patterning of aromatic residues determine the phase behavior of prion-like domains*. Science, 2020. **367**(6478): p. <http://dx.doi.org/10.1126/science.aaw8653>.

39. Choi, J.-M., F. Dar, and R.V. Pappu, *LASSI: A lattice model for simulating phase transitions of multivalent proteins*. PLoS computational biology, 2019. **15**(10).
40. Harmon, T.S., et al., *Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins*. eLife, 2017. **6**: p. e30294.
41. Harmon, T.S., A.S. Holehouse, and R.V. Pappu, *Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins*. New Journal of Physics, 2018. **20**(4): p. 045002.
42. Erickson, H.P., *Modeling the physics of FtsZ assembly and force generation*. Proceedings of the National Academy of Sciences, 2009. **106**(23): p. 9238.
43. Oldfield, C.J., et al., *Coupled folding and binding with alpha-helix-forming molecular recognition elements*. Biochemistry, 2005. **44**(37): p. 12454-70.
44. Adams, D.W. and J. Errington, *Bacterial cell division: assembly, maintenance and disassembly of the Z ring*. Nature Reviews Microbiology, 2009. **7**(9): p. 642-53.
45. Amos, L.A. and J. Löwe, *Overview of the diverse roles of bacterial and archaeal cytoskeletons*, in *Prokaryotic Cytoskeletons*. 2017, Springer. p. 1-26.
46. Errington, J. and L.J. Wu, *Cell cycle machinery in Bacillus subtilis*, in *Prokaryotic Cytoskeletons*. 2017, Springer. p. 67-101.
47. Lutkenhaus, J. and S. Du, *E. coli cell cycle machinery*, in *Prokaryotic Cytoskeletons*. 2017, Springer. p. 27-65.
48. Vaughan, S., et al., *Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota*. Journal of molecular evolution, 2004. **58**(1): p. 19-29.
49. Gardner, K.A.J.A., D.A. Moore, and H.P. Erickson, *The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide*. Molecular Microbiology, 2013. **89**(2): p. 264-275.
50. Sundararajan, K., et al., *The bacterial tubulin FtsZ requires its intrinsically disordered linker to direct robust cell wall construction*. Nat Commun, 2015. **6**: p. 7281.

51. Kozlov, A.G., et al., *Intrinsically disordered C-terminal tails of E. coli single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA*. J Mol Biol, 2015. **427**(4): p. 763-74.
52. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Current Opinion in Structural Biology, 2015. **32**: p. 102-12.
53. Cohan, M.C., K.M. Ruff, and R.V. Pappu, *Information theoretic measures for quantifying sequence-ensemble relationships of intrinsically disordered proteins*. Protein Eng Des Sel, 2019. **32**(4): p. 191-202.
54. Mittal, A., et al., *Sequence-to-Conformation Relationships of Disordered Regions Tethered to Folded Domains of Proteins*. J Mol Biol, 2018. **430**(16): p. 2403-2421.
55. Al-Husini, N., et al., *alpha-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies*. Molecular Cell, 2018. **71**(6): p. 1027-1039 e14.
56. Beveridge, R., et al., *Ion Mobility Mass Spectrometry Uncovers the Impact of the Patterning of Oppositely Charged Residues on the Conformational Distributions of Intrinsically Disordered Proteins*. J Am Chem Soc, 2019. **141**(12): p. 4908-4918.
57. Sherry, K.P., et al., *Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor*. Proc Natl Acad Sci U S A, 2017. **114**(44): p. E9243-e9252.
58. Staller, M.V., et al., *A High-Throughput Mutational Scan of an Intrinsically Disordered Acidic Transcriptional Activation Domain*. Cell Syst, 2018. **6**(4): p. 444-455 e6.
59. Das, R.K., et al., *Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling*. Proc Natl Acad Sci U S A, 2016. **113**(20): p. 5616-21.
60. Weart, R.B. and P.A. Levin, *Growth rate-dependent regulation of medial FtsZ ring formation*. Journal of Bacteriology, 2003. **185**(9): p. 2826-34.
61. Vitalis, A. and R.V. Pappu, *Methods for Monte Carlo simulations of biomacromolecules*. Annual Reports in Computational Chemistry, 2009. **5**: p. 49-76.
62. Vitalis, A. and R. Pappu, *ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions*. Journal of Computational Chemistry, 2009. **30**(5): p. 673-699.

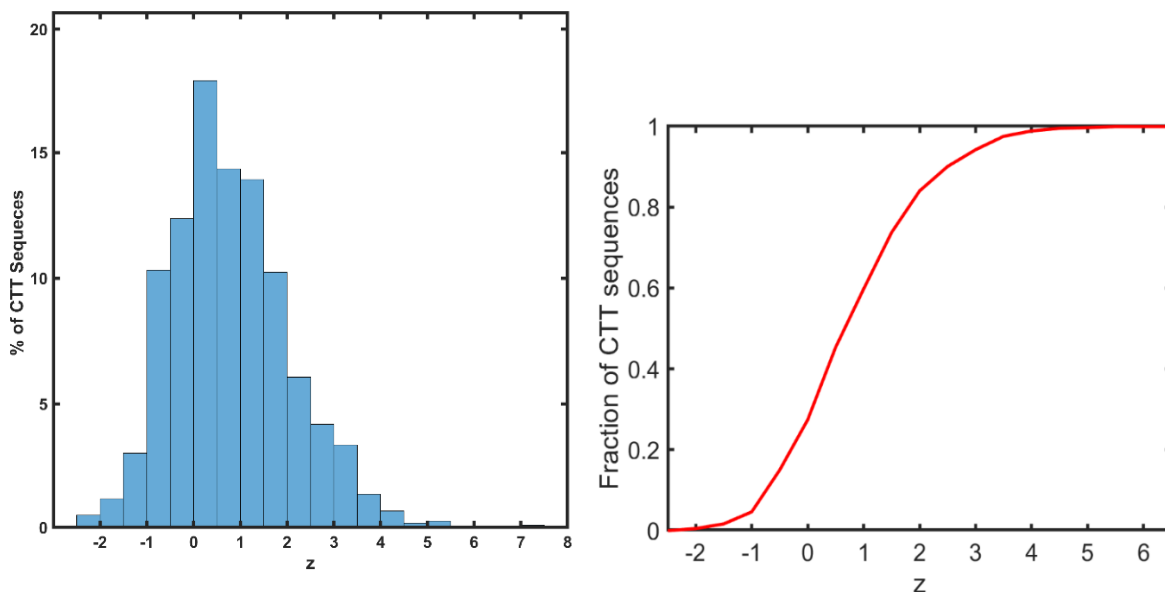
63. Radhakrishnan, A., et al., *Improved atomistic Monte Carlo simulations demonstrate that poly-L-proline adopts heterogeneous ensembles of conformations of semi-rigid segments interrupted by kinks*. Journal of Physical Chemistry B, 2012. **116**(23): p. 6862-6871.
64. Mao, A.H. and R.V. Pappu, *Crystal lattice properties fully determine short-range interaction parameters for alkali and halide ions*. Journal of Chemical Physics, 2012. **137**(6): p. 064104(1-9).
65. Sugita, Y. and Y. Okamoto, *Replica-exchange molecular dynamics method for protein folding*. Chemical Physics Letters, 1999. **314**(1-2): p. 141-151.
66. McGibbon, R.T., et al., *MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories*. Biophysical Journal, 2015. **109**(8): p. 1528-32.
67. Schindelin, J., et al., *Fiji: an open-source platform for biological-image analysis*. Nat Methods, 2012. **9**(7): p. 676-82.
68. Ingberman, E. and J. Nunnari, *A continuous, regenerative coupled GTPase assay for dynamin-related proteins*. Methods in Enzymology, 2005. **404**: p. 611-9.
69. Swinehart, D.F., *The beer-lambert law*. Journal of chemical education, 1962. **39**(7): p. 333.
70. Elson, E.L. and D. Magde, *Fluorescence correlation spectroscopy. I. Conceptual basis and theory*. Biopolymers: Original Research on Biomolecules, 1974. **13**(1): p. 1-27.
71. Crick, S.L., et al., *Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions*. Proceedings of the National Academy of Sciences, 2006. **103**(45): p. 16764-16769.
72. Holehouse, A.S., et al., *CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins*. Biophysical Journal, 2017. **112**(1): p. 16-21.
73. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*. Proceedings of the National Academy of Sciences of the United States of America, 2013. **110**(33): p. 13392-13397.
74. Sawle, L., J. Huihui, and K. Ghosh, *All-Atom Simulations Reveal Protein Charge Decoration in the Folded and Unfolded Ensemble Is Key in Thermophilic Adaptation*. J Chem Theory Comput, 2017. **13**(10): p. 5065-5075.

75. Sawle, L. and K. Ghosh, *A theoretical method to compute sequence dependent configurational properties in charged polymers and proteins*. J Chem Phys, 2015. **143**(8): p. 085101.
76. Lin, Y.H. and H.S. Chan, *Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated*. Biophys J, 2017. **112**(10): p. 2043-2046.
77. Lin, Y.H., J.D. Forman-Kay, and H.S. Chan, *Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles*. Phys Rev Lett, 2016. **117**(17): p. 178101.
78. Lin, Y.H., J.D. Forman-Kay, and H.S. Chan, *Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates*. Biochemistry, 2018. **57**(17): p. 2499-2508.
79. Pak, C.W., et al., *Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein*. Mol Cell, 2016. **63**(1): p. 72-85.
80. Mao, A.H., N. Lyle, and R.V. Pappu, *Describing sequence-ensemble relationships for intrinsically disordered proteins*. Biochemical Journal, 2013. **449**: p. 307-318.
81. Das, R.K., S.L. Crick, and R.V. Pappu, *N-Terminal Segments Modulate the alpha-Helical Propensities of the Intrinsically Disordered Basic Regions of bZIP Proteins*. Journal of Molecular Biology, 2012. **416**(2): p. 287-299.
82. Lyle, N., R.K. Das, and R.V. Pappu, *A quantitative measure for protein conformational heterogeneity*. Journal of Chemical Physics, 2013. **139**(12): p. 121907(1-12).
83. Buske, P.J. and P.A. Levin, *A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo*. Mol Microbiol, 2013. **89**(2): p. 249-63.
84. Fu, G., et al., *In vivo structure of the E. coli FtsZ-ring revealed by photoactivated localization microscopy (PALM)*. PLoS One, 2010. **5**(9): p. e12682.
85. Guan, F., et al., *Lateral interactions between protofilaments of the bacterial tubulin homolog FtsZ are essential for cell division*. Elife, 2018. **7**: p. e35578.
86. Huecas, S., et al., *Self-Organization of FtsZ Polymers in Solution Reveals Spacer Role of the Disordered C-Terminal Tail*. Biophys J, 2017. **113**(8): p. 1831-1844.

87. Sundararajan, K. and E.D. Goley, *The intrinsically disordered C-terminal linker of FtsZ regulates protofilament dynamics and superstructure in vitro*. J Biol Chem, 2017. **292**(50): p. 20509-20527.
88. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division*. Semin Cell Dev Biol, 2015. **37**: p. 3-10.
89. Gardner, K.A., D.A. Moore, and H.P. Erickson, *The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide*. Mol Microbiol, 2013. **89**(2): p. 264-75.
90. Barrows, J.M., et al., *FtsA Regulates Z-Ring Morphology and Cell Wall Metabolism in an FtsZ C-Terminal Linker-Dependent Manner in Caulobacter crescentus*. J Bacteriol, 2020. **202**(7): p. e00693-19.
91. Sundararajan, K., et al., *Species- and C-terminal linker-dependent variations in the dynamic behavior of FtsZ on membranes in vitro*. Mol Microbiol, 2018. **110**(1): p. 47-63.
92. Borgia, A., et al., *Extreme disorder in an ultrahigh-affinity protein complex*. Nature, 2018. **555**(7694): p. 61-66.
93. Pak, C.W., et al., *Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein*. Molecular Cell, 2016. **63**(1): p. 72-85.
94. Nott, T.J., et al., *Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles*. Mol Cell, 2015. **57**(5): p. 936-47.
95. Lin, Y.H., J.D. Forman-Kay, and H.S. Chan, *Theories for Sequence-Dependent Phase Behaviors of Biomolecular Condensates*. Biochemistry, 2018.
96. Chang, L.W., et al., *Sequence and entropy-based control of complex coacervates*. Nature Communications, 2017. **8**(1): p. 1273.
97. Williamson, T.E., et al., *Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin*. J Mol Biol, 2010. **396**(5): p. 1295-309.
98. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Curr Opin Struct Biol, 2015. **32**: p. 102-12.

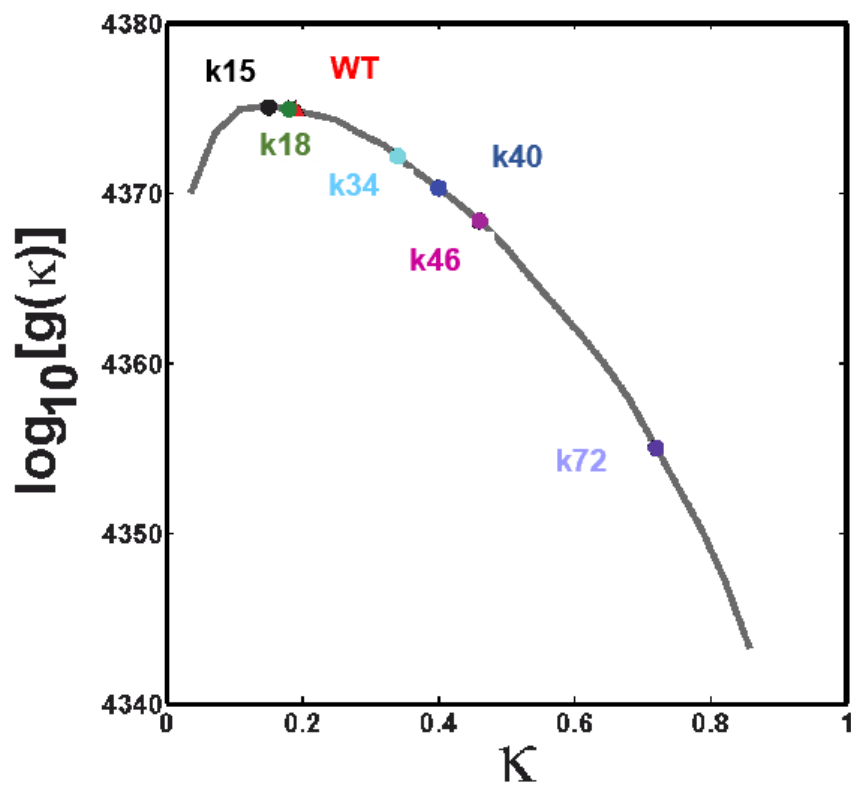
99. Fishbain, S., et al., *Sequence composition of disordered regions fine-tunes protein half-life*. Nat Struct Mol Biol, 2015. **22**(3): p. 214-21.
100. Whatmore, A.M., J.A. Chudek, and R.H. Reed, *The effects of osmotic upshock on the intracellular solute pools of Bacillus subtilis*. J Gen Microbiol, 1990. **136**(12): p. 2527-35.
101. Harmon, T.S., et al., *GADIS: Algorithm for designing sequences to achieve target secondary structure profiles of intrinsically disordered proteins*. Protein Engineering, Design & Selection, 2016. **29**(9): p. 339-46.

## 5.7 Appendix

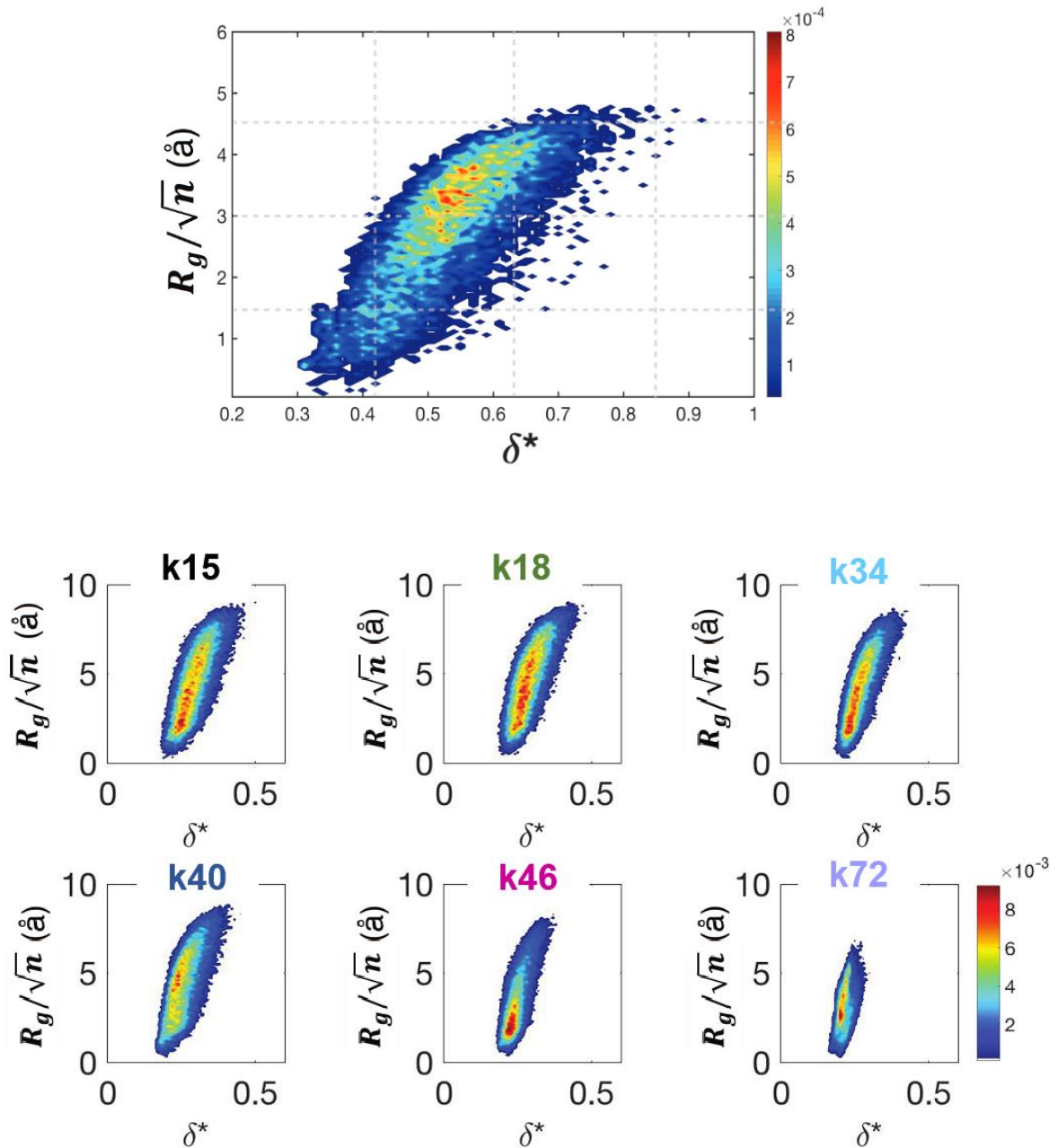


**Figure 5A.1: Distribution of z-scores from the  $\kappa_{+-}$  parameter values observed from a system of 1208 FtsZ orthologs.** Most sampled FtsZ orthologs have oppositely charged residue patterning parameter values near the null-scramble expectation. (Left) Histogram of z-values for 1208 C-terminal tail (CTT) FtsZ sequences. As shown in the cumulative distribution function (CDF; right), approximately 92% of sequences fall within 2 standard deviations from the null model, and ~ 72% of sequences fall within 1.





**Figure 5A.2: Design space of CTL variants.** Keeping the CTP sequence fixed, we designed sequence variants of the *B. subtilis* FtsZ by fixing the amino acid compositions and shuffling residues within the CTL to generate proteins with CTT sequences that have altered linear patterning of oppositely charged residues. There are roughly  $10^{40}$  sequences that meet the design criteria.



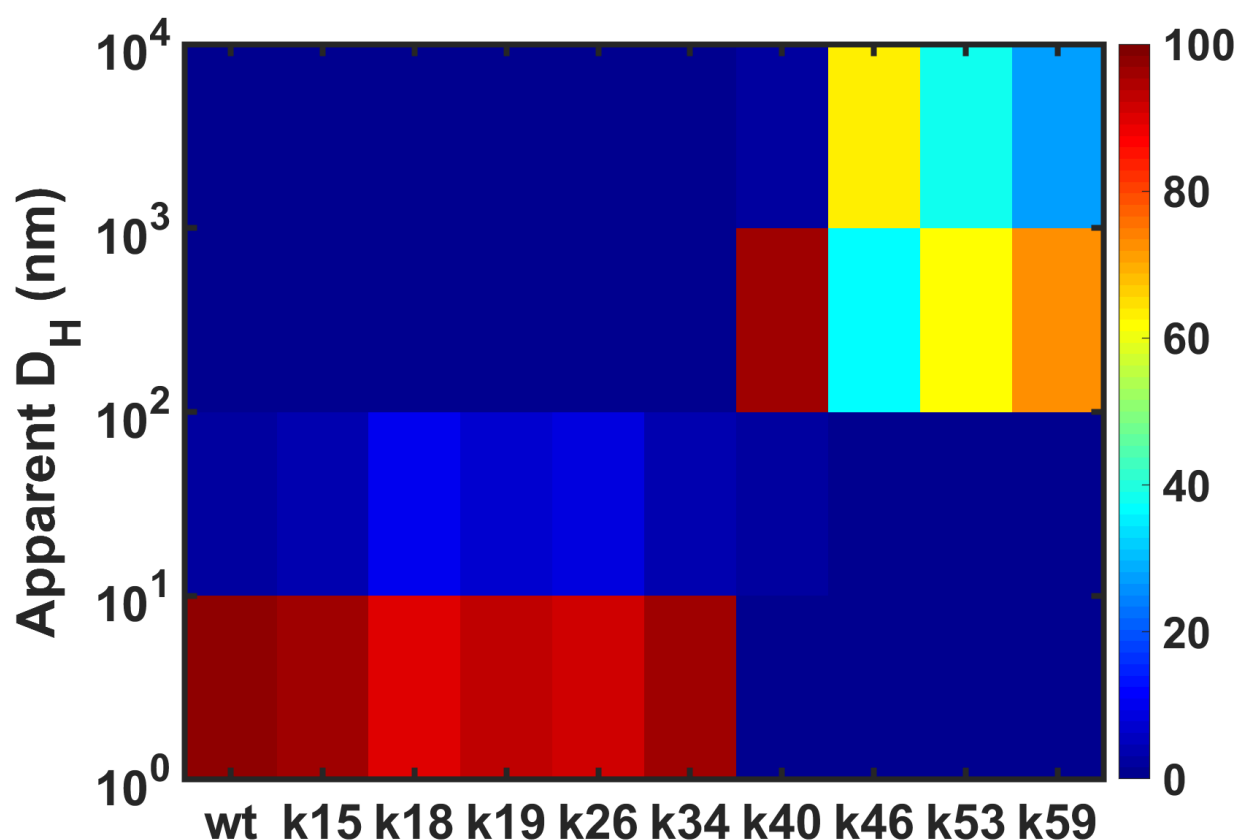
**Figure 5A.3: Two-dimensional conformational distributions from atomistic simulations of *B. subtilis* FtsZ (top) and all CTT variants (bottom).** These distributions quantify the joint probability density  $\rho(R_g/\sqrt{n}, \delta^*)$  of normalized sizes ( $R_g/\sqrt{n}$ ) and shapes ( $\delta^*$ ). In the simulations, each CTT variant was modeled as an autonomous unit.

Strain or Plasmid	Genotype/features	Source
<b><i>B. subtilis</i></b>		
JH642	<i>trpC2 pheA1</i>	Perego et al 1988
PAL 2084	JHC642 <i>thrC::P<sub>xyI</sub>-ftsZ</i>	Wearat Levin 2003
PAL 3171	PAL 2084 <i>amyE::P<sub>spac</sub>-ftsZ</i>	Buske Levin 2012
PAL 3491	PAL 2084 <i>amyE::P<sub>spac</sub>-ftsZCTLV1</i>	This study
PAL 3495	PAL 2084 <i>amyE::P<sub>spac</sub>-ftsZCTLV2</i>	This study
PAL 3499	PAL 2084 <i>amyE::P<sub>spac</sub>-ftsZCTLV3</i>	This study
PAL 3503	PAL 2084 <i>amyE::P<sub>spac</sub>-ftsZCTLV4</i>	This study
PAL 3507	PAL 2084 <i>amyE::P<sub>spac</sub>-ftsZCTLV5</i>	This study
PAL 3511	PAL 2084 <i>amyE::P<sub>spac</sub>-ftsZCTLV6</i>	This study
<b><i>E. coli</i></b>		
AG1111	<i>DZR200- MC1061 F'lacIQ lacZM15 Tn10 (tet)</i>	Ireton et al 1993
PAL 930	AG1111 + pBS58	Wearat Levin 2003
<b>Plasmids</b>		
pPJ19	pET21b(+)- <i>ftsZ946-51 Bam HI 1090-95 Xma I stop</i>	Buske Levin 2013
pPJ53	pET21b(+)- <i>ftsZCTLV1 stop</i>	This study
pPJ54	pDR67- <i>ftsZCTLV1 stop</i>	This study
pPJ55	pET21b(+)- <i>ftsZCTLV2 stop</i>	This study
pPJ56	pDR67- <i>ftsZCTLV2 stop</i>	This study
pPJ57	pET21b(+)- <i>ftsZCTLV3 stop</i>	This study
pPJ58	pDR67- <i>ftsZCTLV3 stop</i>	This study
pPJ59	pET21b(+)- <i>ftsZCTLV4 stop</i>	This study
pPJ60	pDR67- <i>ftsZCTLV4 stop</i>	This study
pPJ61	pET21b(+)- <i>ftsZCTLV5 stop</i>	This study
pPJ62	pDR67- <i>ftsZCTLV5 stop</i>	This study
pPJ63	pET21b(+)- <i>ftsZCTLV6 stop</i>	This study
pPJ64	pDR67- <i>ftsZCTLV6 stop</i>	This study

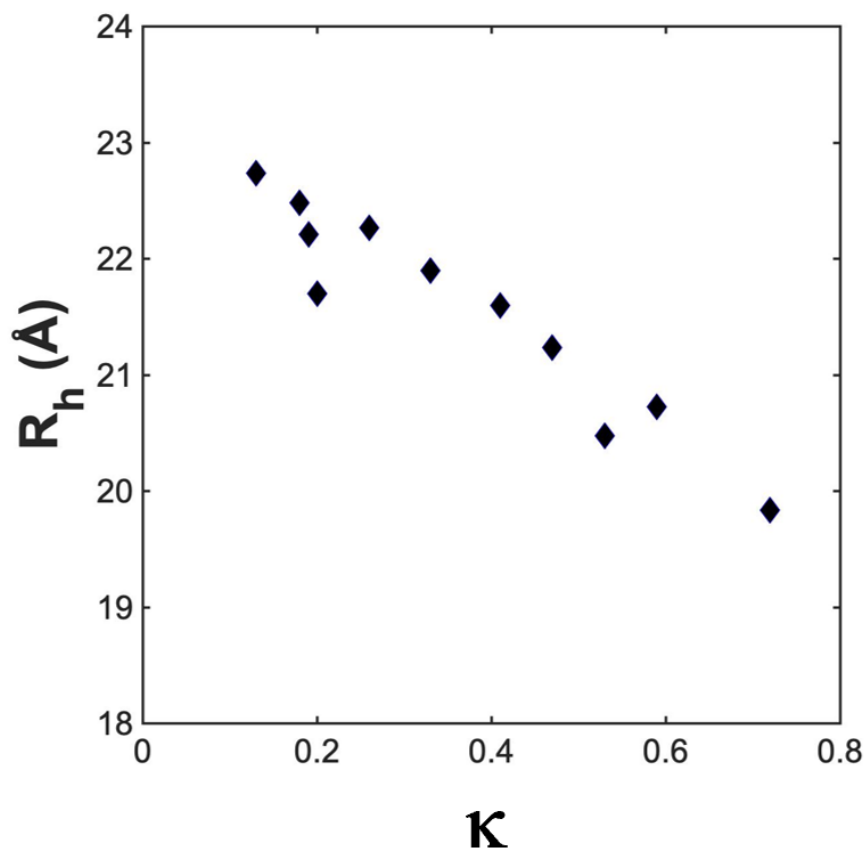
**Table 5A.1 Table of stains and plasmids referenced in this study.** Relevant sources are cited in the third column. Here, the nomenclature used in the study and the strain tables generated by the Levin Lab (collaborator) vary. k15 is k15, V2 is k18, V3 is k34, V4 is k40, V5 is k46, and V6 is k72.

k19	GSIKQERKVTKPQEPSLNQSI <b>R</b> THNQSVPEKEPDERRPQQQNTVSKHTSQPA[CTP]	0.19
k26	GSIEQKEKVTKPQRPSLNQSIETHNQSVPRRKPREKDPQQQNTVSEHTSQPA[CTP]	0.26
k53	GSIEQEEDVTEPQEPSLNQSI <b>R</b> THNQSVPKKKPRKKRPQQQNTVSRHTSQPA[CTP]	0.53
k59	GSPIVLEEDEETSEVSTNQKTQQV <b>K</b> PPPINSSQPQQHHPQRAQRKRRKKTSN[CTP]	0.59

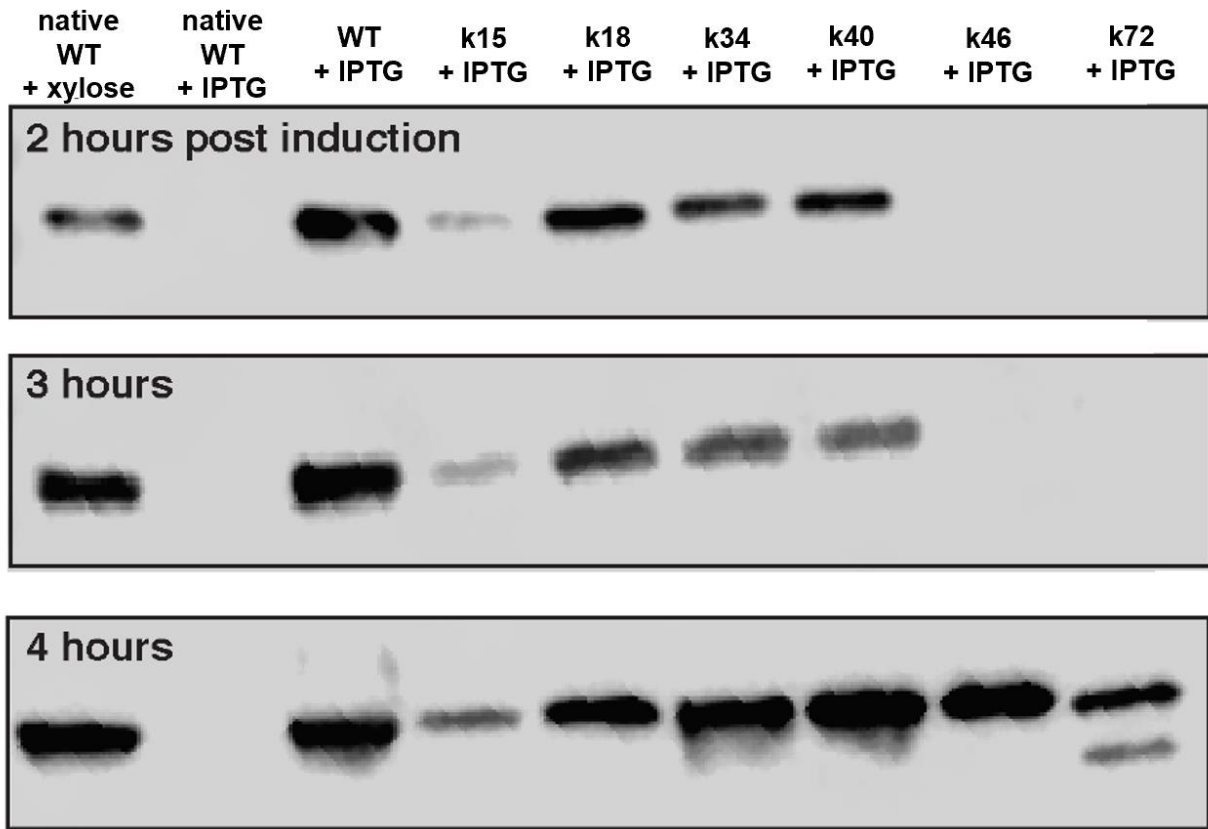
**Table 5A.2: Additional variants considered for future experiments.** The name of the new variant is in the first column, the CTL sequence follows in the second column, and the CTT- $\kappa$  values are in the last column.



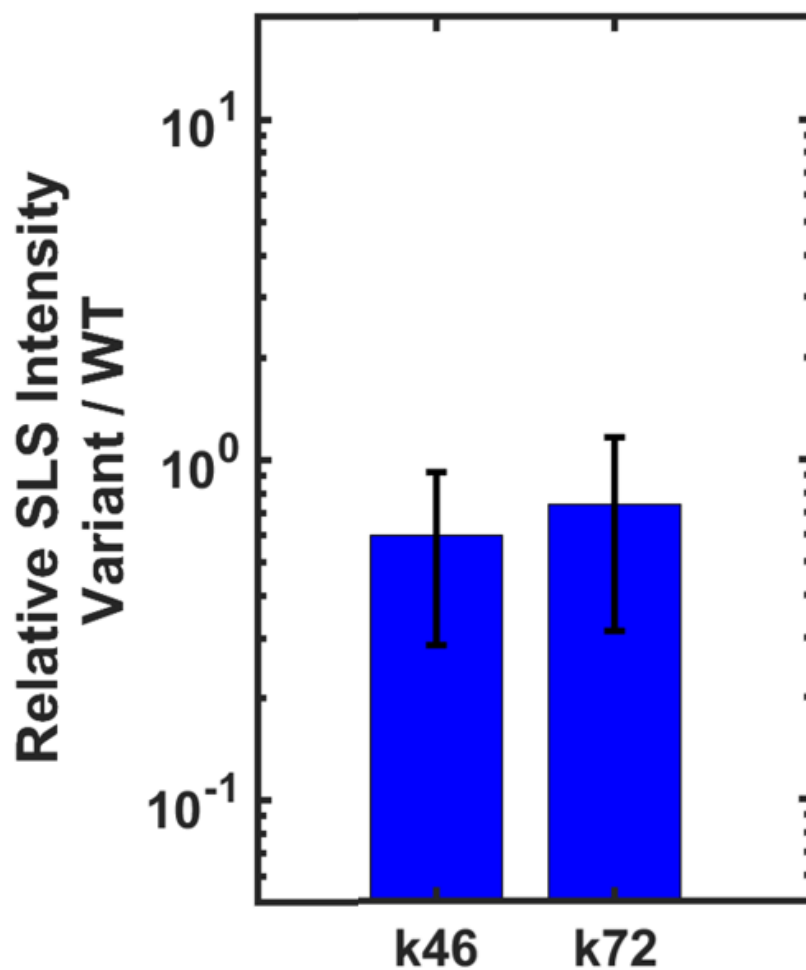
**Figure 5A.4: DLS data in the absence of GTP, including new variants.** High CTT- $\kappa$  variants also show sticker-like behavior as higher-order assemblies are observed in the absence of GTP. Data are identical to **Figure 5.5** with new variants added to the matrix.



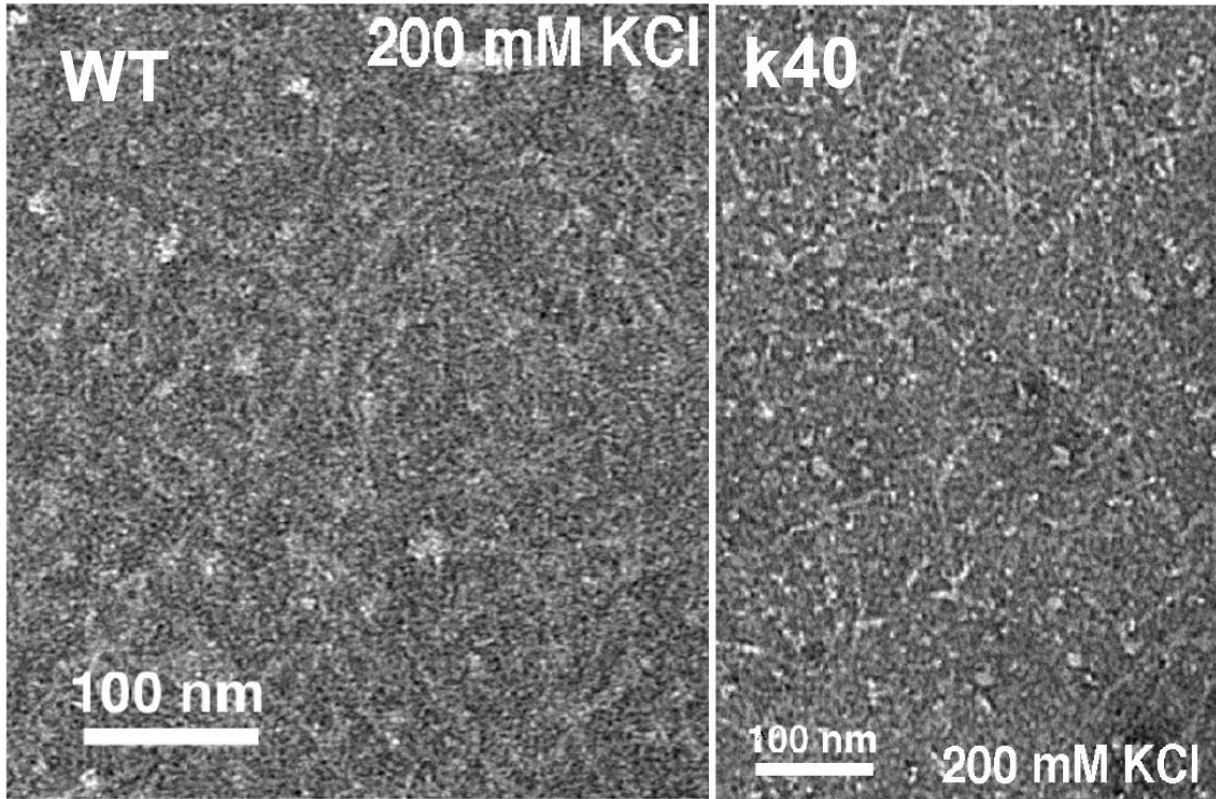
**Figure 5A.5: Additional variants also follow the trend of compaction with increasing  $\kappa$  value as measured by FCS.** Hydrodynamic radius values are plotted as a function of CTT- $\kappa$ . Data are identical to **Figure 5.1**, with new variants added to the plot.



**Figure 5A.6: Time-dependent Western blot quantifying the protein levels of each variant and the native FtsZ in *B. subtilis*** (see Table 5.S-1 for stain and plasmid information). The native FtsZ is inducible by xylose, and all FtsZ variants, including FtsZ WT, are inducible by IPTG. The time points were taken 2, 3, and 4 hours after induction. At each time point, the FtsZ CTT- $\kappa$  variants are present at lower concentrations than that of WT and native FtsZ. When available, k72 shows two distinct bands. Without xylose, the native FtsZ is not visible, indicating it is likely not present in the bacteria expressing FtsZ variants

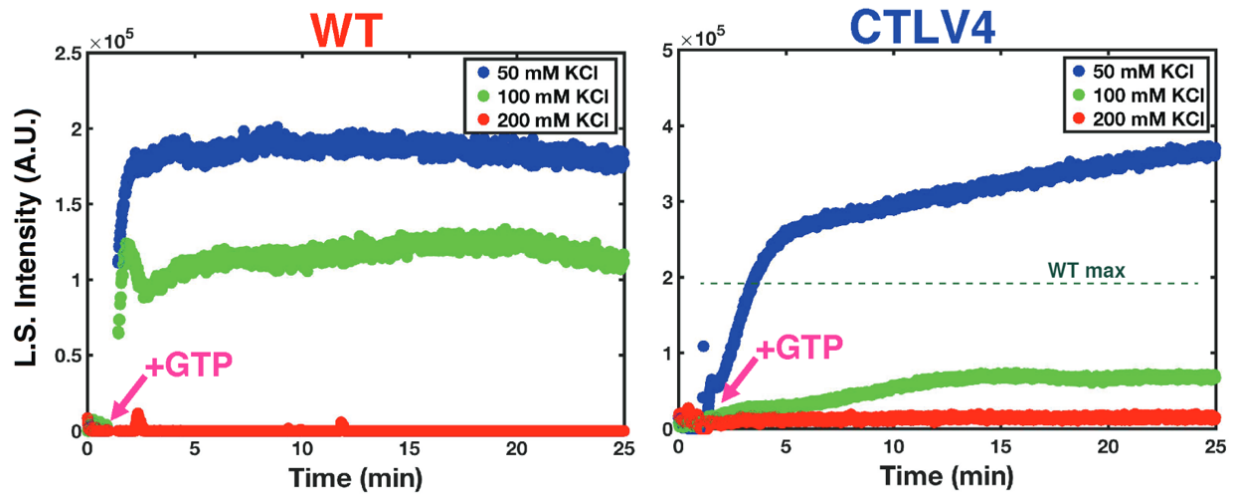


**Figure 5A.7: Relative SLS intensities values of CTT variants that behave as stickers.** The variants k46 and k72 have approximately the same intensity magnitude of WT. However, this data is apparently at odds with the TEM images showing that these variants assemble into much larger sized structures, indicating that they might be insoluble and, therefore, not fully detected by SLS.

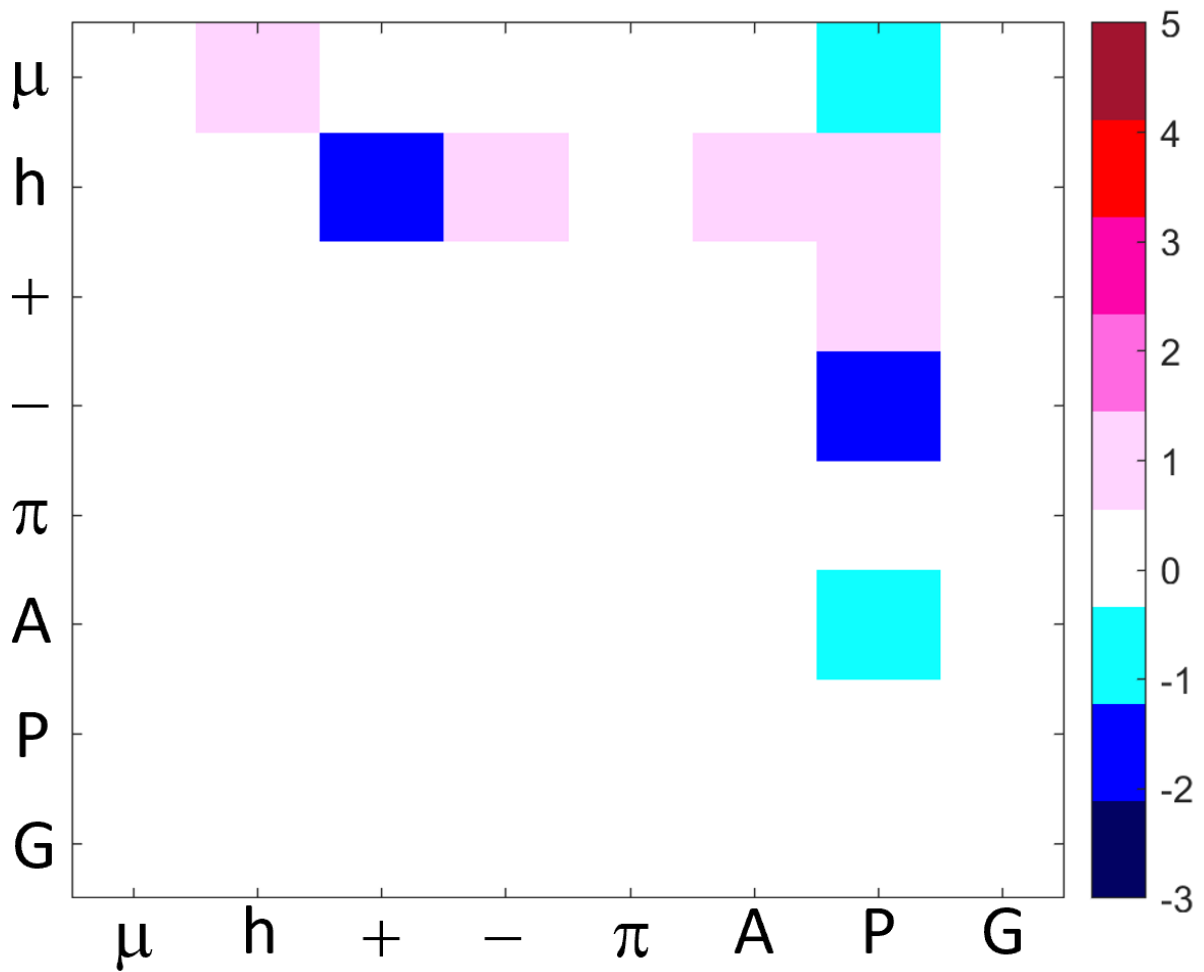


**Figure 5A.8: Negative stain EM images of wild type and k40 in high salt conditions.** Increasing the KCl concentration 4-fold resulted in structures formed by k40 are identical to those formed by WT. Both versions of FtsZ form single-stranded protofilaments. Interestingly, the relative GTPase activity of k40 with respect to wild type in identical conditions increases almost 2-fold to 1.38. This indicates that the dynamics within the protofilaments may be CTL-mediated.





**Figure 5A.9: Impact of salt on wild type and k40 assembly as measured by SLS.** Representative data traces from light scattering for WT FtsZ in the presence of differing amounts of KCl. The arrow marks the time point for the introduction of GTP into the solvent mixture. At low salt, linear FtsZ filaments form bundled assemblies. The bundling is weakened at higher salt concentrations for both wild type (left) and k40 (right).



**Figure 5A.10: Z-score matrix for the randomly scrambled CTL sequence assessed in the study conducted by Buske and Levin [83].** While this matrix is closest to that of wild type in terms of the calculated distance, the sequence is considered to be generally random.

## Chapter 6

# Making the case for disordered proteins and biomolecular condensates in bacteria

### 6.1 Preamble

This chapter is based on the following article: Cohan, M.C., Pappu, R.V. (2020). Making the Case for Disordered Proteins and Biomolecular Condensates in Bacteria. *Trends in Biochemical Sciences*, **45**: 668 – 680. Intrinsically disordered proteins / regions (IDPs / IDRs) contribute to a diverse array of molecular functions in eukaryotic systems. There is also growing recognition that membraneless biomolecular condensates, many of which are organized or regulated by IDPs / IDRs, can enable spatial and temporal regulation of complex biochemical reactions in eukaryotes. Taken together with the findings from this thesis, we assess if (and how) membraneless biomolecular condensates and IDPs / IDRs are functionally involved in key cellular processes and molecular functions in bacteria. In this chapter, we summarize conceptual underpinnings of condensate assembly and leverage these concepts by connecting them to recent findings that implicate specific

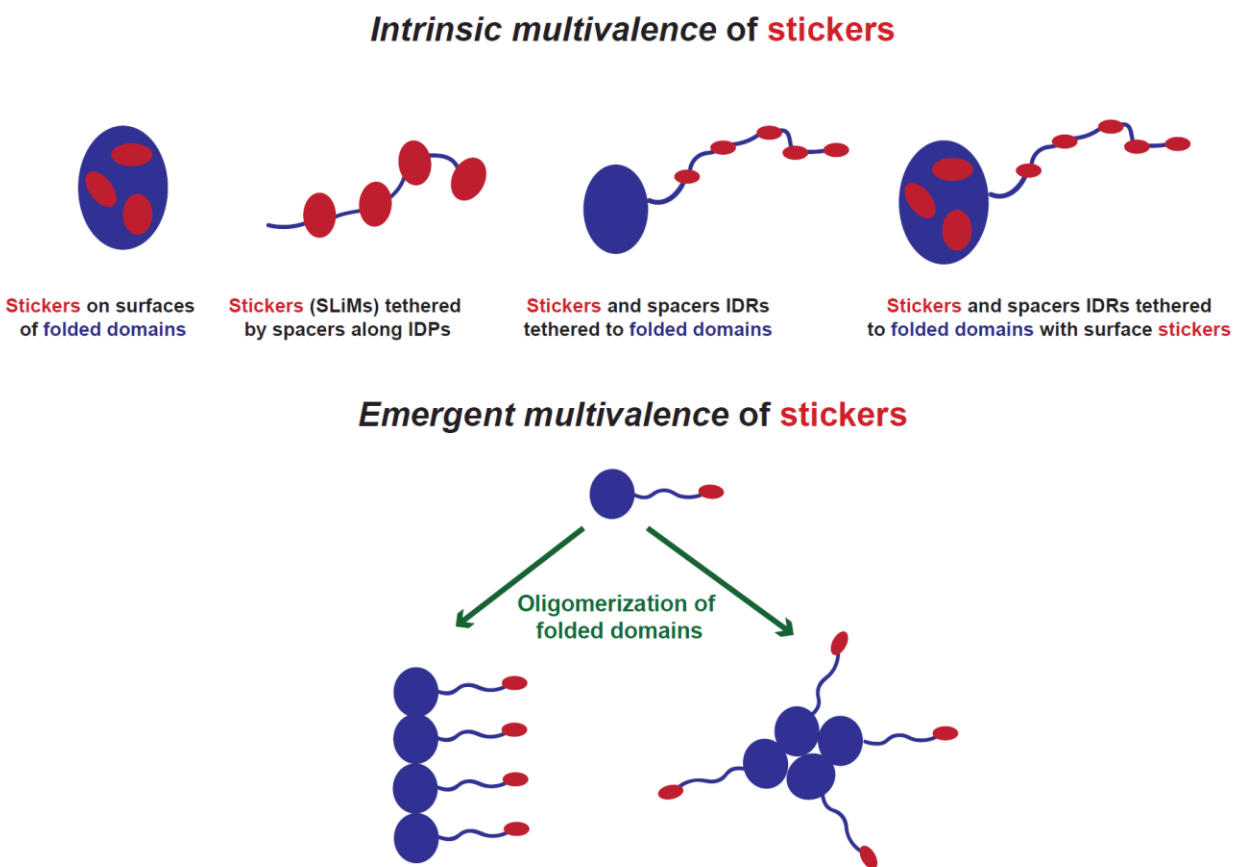
types of condensates and IDPs / IDRs in important cellular level processes and molecular functions in bacterial systems.

## 6.2 Introduction

Spatial and temporal regulation of cellular matter is necessary for the control of transcription, protein quality control, cell signaling, and responses to stimuli [1]. In many eukaryotic systems, the reversible formation of membraneless biomolecular condensates – referred to hereafter as condensates – provides spatiotemporal control by organizing biochemical reactions, enabling quality control, and concentrating cellular matter [2-5]. Condensates are defined as non-stoichiometric assemblies of multivalent macromolecules [6] that concentrate biomolecules [2]. *Multivalence* refers to the numbers of specific interaction sites [2, 5, 7, 8]. Specific interaction sites can be short linear motifs within intrinsically disordered regions (IDRs) (see **Chapter 1.3**) or hot spots / sectors on the surfaces of folded domains of proteins [9, 10] (**Figure 6.1**).

Associative polymers are “macromolecules with attractive groups” [11]. These systems can be described using a *stickers-and-spacers* framework. *Stickers* are attractive groups that form reversible physical crosslinks with one other. These are non-covalent interactions such as hydrogen bonds, cation-pi or pi-pi interactions, or a hierarchy of electrostatic interactions [9, 12, 13]. An associative polymer that has more than one sticker is a *multivalent* polymer. To zeroth order, we distinguish associative polymers from one another by their *intrinsic valence* (**Figure 6.1**), defined as the *maximum number of crosslinks* ( $n_c$ ) that a chain can form. Intrinsic valence is quantified using knowledge of the number of sequence-encoded stickers in the chain and the number of bonds (coordination number) that an individual sticker can make. Some systems can self-assemble to form higher-order oligomers. These higher-order assemblies can enable novel crosslinks that cannot be formed by the constituent protomers. Accordingly, higher-order oligomers are characterized by an *emergent valence* (**Figure 6.1**), which we define to be the *maximum number of crosslinks* ( $n_c$ ) that

the oligomer can form through stickers that are not involved in forming higher-order oligomers. Spacers, on the other hand, provide the linear and / or spatial separation between stickers, and their interactions with the solvent contribute to the driving forces for spontaneous phase transitions [14-19] – a defining feature of multivalent macromolecules [2, 4, 8, 11, 20-24].

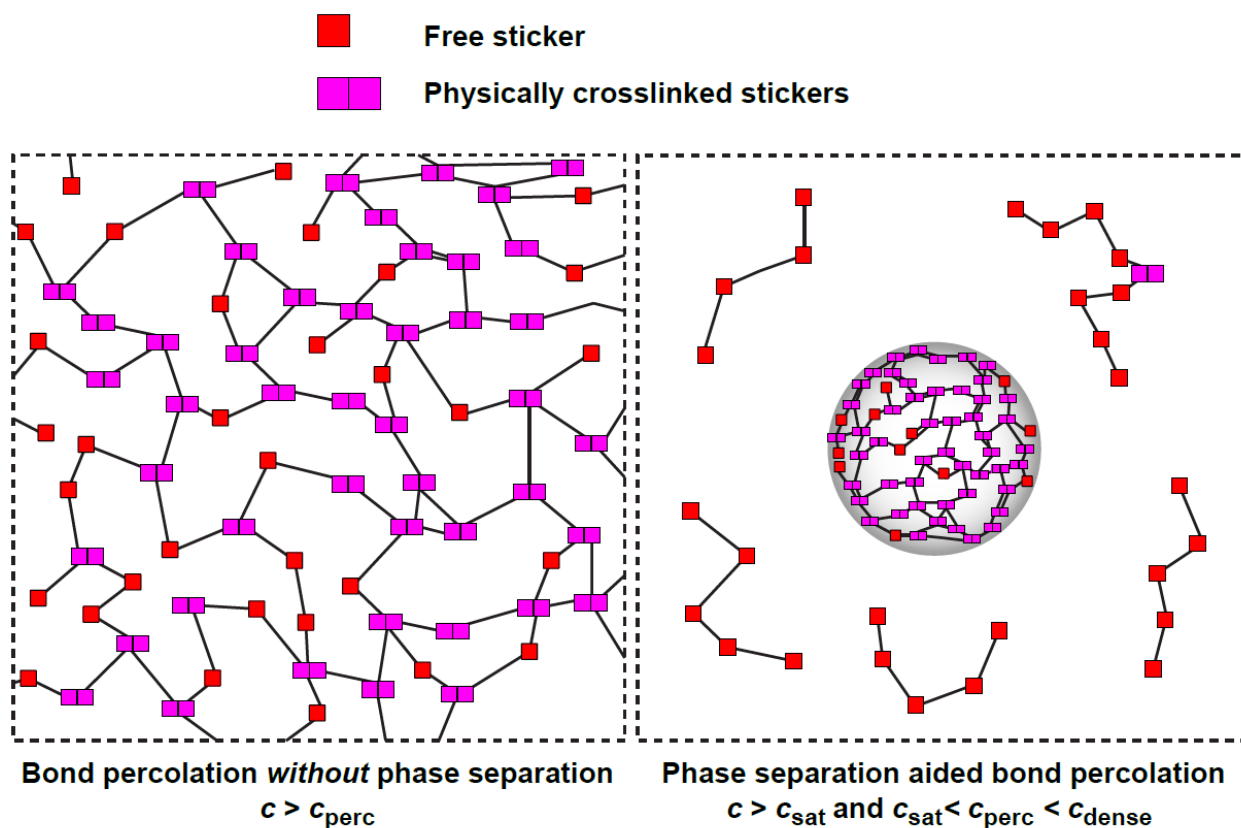


**Figure 6.1: Conceptual depiction of intrinsic versus emergent sticker multivalence.** (Top) The intrinsic multivalence of stickers can be encoded via different architectures, as shown in the four panels. (Bottom) The emergent multivalence of stickers occurs when a protein with a single sticker or interaction motif assembles into a larger complex by forming oligomers via a linear process (left-hand side) or oligomerizing (right-hand side).

### 6.2.1 Eukaryotic condensates appear to form via phase transitions

Multivalent macromolecules can drive condensate formation via spontaneous phase transitions [14, 15, 17-19, 25, 26]. Inter-chain physical crosslinks among stickers can enable a *networking transition* known as *bond percolation* [10, 11, 14-16, 20]. This happens above a system-specific threshold concentration known as the percolation threshold and designated as  $c_{\text{perc}}$ ; above  $c_{\text{perc}}$ , the collection of multivalent macromolecules in solution will form a system-spanning network [14, 15, 18] (**Figure 6.2**). These transitions are also known as sol-gel transitions [27], but we prefer the term bond percolation to avoid confusion with *ad hoc* definitions of gels [28].

Multivalent macromolecules can also undergo density transitions known as *phase separation*. Here, a binary mixture comprising associative polymers and a solvent separates into two distinct phases, namely, a macromolecule-rich phase that coexists with a macromolecule-deficient phase [11, 14, 20]. Phase separation in a binary mixture is defined by the presence of a threshold concentration known as the saturation concentration  $c_{\text{sat}}$  [16], which corresponds to the concentration of macromolecules in the coexisting dilute phase; conversely, the concentration of macromolecules in the coexisting dense phase is designated by  $c_{\text{dense}}$ . If  $c_{\text{perc}} < c_{\text{sat}}$ , then bond percolation occurs without phase separation [18]. However, it is often the case that  $c_{\text{sat}} < c_{\text{perc}}$  and  $c_{\text{sat}} < c_{\text{perc}} < c_{\text{dense}}$ . In this scenario, multivalent macromolecules will undergo *phase separation aided bond percolation* (PSBP) transitions (**Figure 6.2**) [18].



**Figure 6.2: Schematics to distinguish bond percolation without phase separation (left) versus phase separation aided bond percolation (right).** The black segments are spacers that are interspersed between stickers, shown as red squares. A pair of magenta squares signify a pair of stickers that are connected via a physical crosslink.

## 6.2.2 Scaffolds versus clients

The driving forces for PSBP transitions are governed by the strengths of sticker-sticker interactions, the valence of stickers, the coordination number per sticker, the solvation properties of spacers, and the weaker, isotropic spacer-spacer, and sticker-spacer attractions [14, 15]. The lower the values of  $c_{sat}$  and  $c_{perc}$ , the stronger the driving forces for PSBP transitions. Accordingly, system-specific driving forces for PSBP transitions can be quantified in terms of the values of  $c_{sat}$  and  $c_{perc}$  for a given set of solution conditions. This quantification can be achieved via measurements [16, 29], analytical or quasi-analytical calculations [16, 20, 30-32], and computer



simulations [14, 18, 29, 33, 34]. Multivalent macromolecules that drive PSBP transitions under physiologically relevant conditions (low-to-sub-micromolar protein concentrations, 150 mM monovalent salt, pH 7.4, 25 – 37°C) [35] are referred to as *scaffolds* [2, 7]. Conversely, macromolecules that preferentially accumulate into condensates formed by scaffolds are referred to as *clients* [2, 7, 36]. To zeroth order, scaffolds and clients can be distinguished by the valence of stickers (higher in scaffolds versus clients) and the solvation profiles of spacers (clients are likely to have spacers that are preferentially solvated) [2, 7, 18, 36].

### 6.2.3 Descriptors of the organization of molecular matter within condensates

PSBP transitions lead to condensate-spanning networks [18]; this is in contrast to the system-spanning networks that form via bond percolation *without* phase separation [18]. The multiscale structures of condensates can be described by topological features (inferred from rheological measurements [17, 37-41]), such as the degree of connectedness of multivalent macromolecules and the extent of crosslinking. Structural descriptors also include the mean and variance of protein density as well as more detailed descriptors such as distribution functions, derivable from scattering measurements [42], that describe the distributions of distances and orientations of molecules with respect to one another.

Physical crosslinks among stickers are reversible, although the timescales for making and breaking crosslinks will be governed by different factors that influence the material properties of condensates [11]. They will have fluid-like properties if physical crosslinks make and break on timescales that are commensurate with macromolecule diffusion. Such condensates are best described as *viscoelastic network fluids* where the delineation between elastic and viscous regimes is governed by timescales associated with the making and breaking of crosslinks as well as the

extent of crosslinking and the density within condensates [43]. We emphasize that PSBP transitions can lead to a range of condensate types that include viscoelastic network fluids, viscous liquids, liquid crystals, or different types of solids [4, 25, 44]. The most appropriate designation for the state of molecular matter will be governed by the interplay between the long-range and short-range translational and orientational ordering of molecules and the types of symmetries that are made or broken in conjunction with PSBP transitions [45].

In the extant literature, many condensates are referred to as liquids, and the relevant phase transition that gives rise to condensates is referred to as liquid-liquid phase separation (LLPS) [3, 46, 47]. This comes from the observation that many (although not all) condensates have spherical shapes *in vitro*, *in vivo*, or both [4]. Spherical shapes point to the minimization of the interfacial tension between condensates and the surrounding milieu [47]. Condensates can also wet surfaces and fuse with one another [4]. Scaffolds and clients can have rich dynamical features within condensates, and these pertain to the making and breaking of physical crosslinks, diffusive motions within condensates, and the exchange of molecules across the interface between condensates and the surrounding milieu. Accordingly, condensates with spherical shapes and dynamical features that are concordant with those of dense polymeric liquids are referred to as liquids, although by rigorous physical classifications, these are not simple liquids [48] (a term that is reserved for liquids formed by spherical particles that interact via short-range van der Waals interactions [49]). Instead, condensates are better described as complex fluids or network fluids [43, 50]. Given that scaffold molecules can be mapped onto stickers-and-spacers architectures, we choose the rigor associated with the terms PSBP transitions and viscoelastic network fluids [10] over the lack of clarity and precision engendered by terms such as LLPS and liquids.

## **6.3 Making a case for condensates in bacteria**

Given the lack of membrane-bound organelles in bacteria and the importance of spatial and temporal regulation for various biochemical reactions, it would seem that condensates, regulated by reversible PSBP transitions, might play crucial roles in enabling spatiotemporal regulation within bacteria. However, when compared to eukaryotes, there are fewer reports of condensates or of the role of PSBP transitions in prokaryotes [51]. This is probably because IDRs make up less than 2-5% of bacterial proteomes [52], a relevant issue given that IDRs with specific sequence features can be prominent scaffolds for PSBP transitions in eukaryotic systems [2-4, 9, 10, 16, 25, 35, 38, 44, 46, 53-56]. The paucity of IDRs in prokaryotes might be taken to mean that they are less important or relevant for bacteria. However, studies over the past decade have demonstrated that IDRs are, in fact, key players within proteins that drive or regulate cell division, bacterial warfare, community organization, DNA and RNA metabolism, RNA folding, and nucleoid organization [51, 57-68]. In what follows, we present an overview of cellular processes and molecular functions in bacterial systems influenced by IDRs. In doing so, we make a case for the importance of disorder and for condensates in processes such as division, transcription, post-transcriptional regulation, and stress responses.

### **6.3.1 Spatiotemporal control of bacterial cell division**

The initiation of cell division in rod-shaped bacteria is marked by the formation of a cytokinetic ring at the cell center [69]. A family of proteins, termed the divisome, collectively functions to constrict the ring in the center of the cell, synthesize new material, and pinch the cell in two [70]. The bacterial cytokinetic ring is scaffolded by the tubulin homolog FtsZ, which serves

as a treadmilling interaction hub for many of the adaptor proteins of the divisome [71-73]. FtsZ encompasses a highly conserved tubulin-like GTPase domain that is connected via a disordered C-terminal linker (CTL) to a conserved C-terminal peptide (CTP) motif that coordinates a network of FtsZ interacting proteins [74, 75].

CTL sequences can vary in length by an order of magnitude (~30-300 residues); they also vary substantially in composition and sequence patterns [76]. Despite this *hypervariability* (**Chapter 1.3.2**), deletion of the disordered CTL impairs cell division, causing a mislocalization and filamenting phenotype whereby FtsZ molecules make aberrant assemblies away from the mid-section in cells that also become aberrantly long [58, 77]. **Chapter 2**, focused on the *B. subtilis* FtsZ, showed that the CTP and CTL modules play the role of a sticker and spacer, respectively [78]. This is relevant because FtsZs undergo GTP-assisted polymerization that leads to the formation of single-stranded polymers, each comprising multiple FtsZ subunits [75, 79] giving rise to emergent multivalence of CTP stickers. Polymers of FtsZ, defined by a multivalence of CTP stickers, can condense to form wreaths and bundles via homotypic interactions among single-stranded polymers [78]. Alternatively, single-stranded polymers and / or bundles can engage in heterotypic interactions with FtsZ interacting proteins. The heterotypic networking transition involves adaptor proteins that assist in anchoring, stabilizing, and providing spatiotemporal control of the cytokinetic ring [70, 80]. Additional examples of IDRs in the FtsZ interactome have been uncovered (**Figure 1.3**), and distinct roles have been identified in positioning the Z-ring and regulating enzymatic activity [81-86]. These findings further highlight the overall role of protein disorder in the formation of the bacterial cytokinetic ring.

While many proteins within the FtsZ interactome are involved in coordinating cell division, others are involved in ensuring that the cytokinetic ring has an appropriate spatial location that does not overlap with the nucleoid [70]. This is a manifestation of intricate spatial regulation, and multiple proteins are involved in the formation of compartments that aid in the nucleoid occlusion process. Some of these inhibit FtsZ activity or spatially sequester FtsZ [72, 87, 88], and others aid in nucleoid occlusion. Recent studies have identified specific roles for condensates (as defined by Banani et al., [2]) as regulators of FtsZ inhibition and nucleoid occlusion [51, 89]. Among the relevant scaffolds for these condensates are SlmA and FtsZ, as well as DNA molecules that house SlmA binding sites. In *E. coli*, SlmA binds to bacterial DNA and to the disordered CTT of FtsZ [90]. Importantly, SlmA can only bind the CTT of FtsZ when SlmA itself has bound to DNA away from the center of the cell. Fluid-like condensates form in cell-like media that include SlmA, FtsZ, DNA that contains SlmA binding sites, polyethylene glycol, and Dextran [89, 91]. These condensates dissolve in the presence of GTP, suggesting an interplay among FtsZ sequestration into condensates, GTP hydrolysis, GTP-mediated FtsZ polymerization, and dissolution of condensates [89]. This implicates condensate formation in the

### **6.3.2 Spatiotemporal control of polarity**

Another area of spatial and temporal organization pertains to the regulation of asymmetric division and polarity that has been reported in *C. crescentus*. Here, a differentiated stalked cell can enter the cell division cycle, and in doing so, it ensures that the cytoplasmic space has two sides: a future swarmer and the old stalk cell [68, 92, 93]. Polar organizing protein Z (PopZ), a major player in this differentiation, is 57% disordered [92, 94-96]. PopZ oligomerizes and forms an excluding network at the dividing cell poles, referred to as a microdomain. The resulting pole organization

gives an orientation to the cell. Side-specific PopZ IDR-interacting proteins can penetrate the selective 200 nm PopZ microdomain, which was recently shown to have many of the defining features [2] of condensates [97].

After cell division, the PopZ microdomain interacts with membrane-anchored proteins PleC and PodJ at the site of the newly formed swarmer cell pole. Emerging data suggest that PodJ also encompasses a long intracellular IDR [98]. In order to undergo subsequent cycles of cell division, the swarmer cell must first differentiate into a stalk cell. This remodeling of the cell pole and the swarmer-to-stalk cell differentiation coincides with the displacement of PleC and PodJ by membrane-anchored SpmX and DivJ, both of which encompass long, intracellular IDRs [98-100]. The emergent multivalence of stickers (whose identities are yet to be determined) converts the pole into a signaling hub that promotes differentiation into stalk cells [93, 101]. New evidence suggests that interactions between SpmX and the PodJ IDR are essential for preventing aberrant assembly of PodJ in the cytoplasmic space during cell division [98]. Each step of the swarmer-to-stalk cell differentiation appears to involve the formation / dissolution of condensates that concentrate scaffold-like proteins with IDRs. Whether the mechanisms of condensate formation involve PSBP transitions remains to be determined, although preliminary results hint toward such a mechanism.

Another important function of PopZ lies in controlling the orientation of the chromosomal centromere (*parS*) through direct and indirect interactions with the ParA/ParB chromosome segregation machinery [93]. Prior to replication, *parS* that is bound to ParB is located at the top of the original PopZ microdomain. Once replicated, one of the ParB-*parS* complexes is trafficked by the ParA ATPase to the other new pole site by following a ParA concentration gradient. Recent studies suggest that this system is also regulated by the formation of a biomolecular condensate

where, in the presence of *parS*, ParB accumulates locally to form dynamic, liquid-like nanoscale condensates [102]. This discovery stemmed from the observation that ParB weakly oligomerizes and encompasses a low complexity IDR that drives the emergent multivalence.

### **6.3.3 Spatiotemporal control of transcription and post-transcriptional regulation**

In bacterial DNA replication and repair, single-stranded DNA binding proteins (SSBs) play an essential role. Their modular architecture, which is reminiscent of FtsZ, includes an ordered DNA-binding domain (OB fold), followed by a hypervariable intrinsically disordered linker (IDL / spacer) that is connected to a conserved C-terminal tip (sticker) [62, 63, 103]. SSBs oligomerize to form homo-tetramers that generate a tetra-valence of stickers that coordinate interactions with SSB interacting proteins. The *E. coli* SSB tetramer binds cooperatively to single-stranded DNA, and the binding modes are classified by the number of nucleotides that are occluded by individual tetramers. Cooperativity of single-stranded DNA binding is governed mainly by sequence features of the IDL. Cooperativity is enhanced when the IDL has features that are akin to low complexity domains enriched in polar amino acids. Conversely, cooperativity is diminished for long IDLs that are enriched in charged residues [62, 103].

The IDL mediated cooperativity of binding to single-stranded DNA derives from the ability of SSBs to form beads-on-strings architectures along single-stranded DNA. Phase transitions are infinitely cooperative transitions [104]. Accordingly, it stands to reason that proteins capable of one form of cooperative transition might also be able to drive highly cooperative phase transitions as well. Indeed, a recent study has shown that in response to DNA damage, membrane-associated

SSBs form puncta at the site of the DNA damage [105]. These condensates are multicomponent bodies and concentrate other critical factors that contribute to DNA processing and metabolism. *In vitro* studies showed that the IDL is essential for driving the formation of liquid-like condensates, even though binding to DNA is independent of the presence of the IDL. Interestingly, while the IDL is not required for cell viability, the lack of the IDL impacts bacterial sensitivity to UV radiation [62]. The inability to form condensates might affect the responsiveness of bacteria to stress, a hypothesis that we formulate based on observations in yeast [25, 106]. To test this hypothesis, we need a comprehensive understanding of how different IDLs across different bacterial systems contribute to PSBP transitions *in vitro* and *in vivo*.

Like SSBs, DNA-binding protein from starved cells (Dps) also binds DNA via a cooperative process. The resulting structures are membraneless protein and DNA complexes [107, 108]. Recently, it was observed that these complexes exclude restriction enzymes but selectively allow the partitioning of RNA polymerase (RNAP) [51]. Further characterizations of this complex are needed in order to identify the material properties and understand their functions vis-à-vis the emerging biophysics and biology of condensates. Similarly, “transcriptional foci” in *E. coli* also involve RNAP and bear the hallmarks of condensates. In slow-growing conditions, RNAP is uniformly dispersed throughout the cell, whereas in fast-growing conditions, RNAP becomes localized to distinct puncta, potentially optimizing ribosomal production [109-111]. Puncta that were previously thought to form via DNA-binding were shown to dissolve with the addition of 1’6 hexanediol to the media, indicating that the puncta might involve noncovalent protein-protein interactions. To identify potential candidates involved in these interactions, a proteome-wide screen for disordered proteins that are involved in rRNA transcription was performed. Among the identified candidates were antitermination factors such as NusA. The modular architecture of NusA



includes two C-terminal acid-rich repeat domains connected by flexible linkers, analogous to a conventional sticker-spacer architecture. Indeed experiments showed that NusA forms liquid-like droplets that fuse *in vitro* and facilitates the formation of foci *in vivo* [112]. Further, the NusA components of the foci diffuse more rapidly than their DNA-bound LacI counterparts. Conclusions from this work lead to the working model that NusA facilitates the formation of dynamic RNAP clusters via PSBP transitions and that this process is controlled by nutrient availability. The functions of these puncta, as they relate to transcriptional efficiency and the role of the IDRs, are yet to be described. However, the extant data provide evidence for the potential existence of nucleolus-like condensates in bacteria.

Another process controlled by condensates in bacteria is that of RNA degradation [113-115]. In this system, RNase E is a critical driver of the formation of the RNA degradosome. The architecture of RNase E includes a conserved DEAD-box RNA helicase and a disordered C-terminal domain (CTD). In *C. crescentus*, the RNase E CTD is necessary and sufficient to drive phase separation. RNase E drives the formation of cytoplasmic foci *in vivo* that colocalize with other exonucleases. This degradation body has been termed **Bacterial Ribonucleoprotein body** or BR-body. It is noteworthy that RNA dependent DEAD-box ATPases (DDXs) form liquid-like condensates in both prokaryotes and eukaryotes that provide spatiotemporal control over RNA processing and synergies with RNA-protein condensates [116]. Indeed, it would be interesting and important to explore the synergies, if any, between BR bodies and bacterial condensates formed by DDXs in general.

The CTD of the *C. crescentus* RNase E has a blocky patterning of oppositely charged residues. The di-block nature of this CTD leads to encoded multivalence of opposite charges [30],

and this architecture appears to be essential for the formation of BR bodies. In theoretical work, Lin and Chan [30] have emphasized the importance of blocky charge patterns for generating may be referred to as “super stickers” that enhance the driving forces for PSBP transitions. Similar findings were obtained by Pak et al., [117] who showed that sequence clustering of like-charged residues drives *de novo* condensate formation by the intracellular domain of Nephhrin.

Interestingly, the *E. coli* RNase E lacks the blocky patterning of oppositely charged residues. Extant data suggest that this protein does not form cytoplasmic condensates; instead, it forms membrane-tethered puncta in *E. coli* [113, 115, 118, 119]. Rigorous biophysical characterizations via requisite *in vitro* comparisons of the driving forces and mechanisms of PSBP transitions of the *C. crescentus* and *E. coli* RNases E could be highly informative and shed light on whether this protein and others like it are exemplars of how sequence hypervariability of IDRs in bacteria provide an evolutionary strategy for tuning condensate formation and the functions regulated by IDRs.

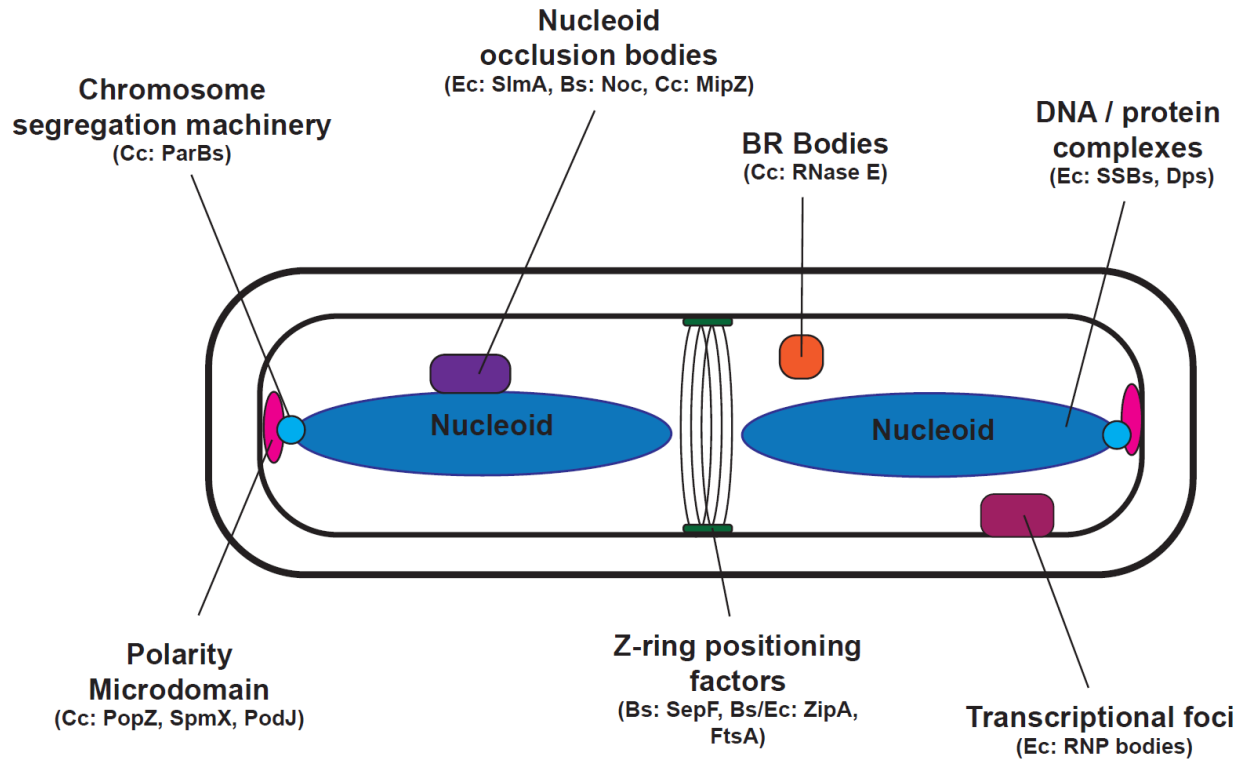
#### **6.3.4 Management of phosphate levels and synthesis of multivalent phosphates**

Bacteria utilize novel mechanisms to manage intracellular phosphate levels, particularly at times of nutrient deprivation. Under these conditions, bacteria will burn ATP to synthesize long polyphosphate chains that coalesce into large granules required for survival. These PolyP granules sequester certain proteins, including the polyphosphate synthesizing enzyme Ppk1 [120]. Polyphosphates are reminiscent of nucleic acid backbones, and it appears that the physical principles relevant to the formation protein-RNA/DNA condensates [117, 121] is transferable for understanding the formation with and regulation of PolyP granules by their clients. Bacteria also utilize the controlled synthesis of other multivalent phosphates, such as (p)ppGpp, to regulate ATP

/ GTP dependent processes. It is known that (p)ppGpp directly impacts RNAP activity [122], and this negative feedback could benefit from the efficiency a condensate provides. However, further investigations are required to substantiate this hypothesis.

Multivalent phosphates can drive PSBP transitions through a combination of interactions that include complimentary electrostatic interactions with multivalent proteins that encompass cationic and aromatic stickers [123]. We reason that multivalent phosphates in bacteria might be positive regulators that drive PSBP transitions via heterotypic interactions with positively charged multivalent proteins; they could also be negative regulators that destabilize PSBP transitions driven by heterotypic interactions involving negatively charged multivalent proteins. While such insights have emerged from studies of synthetic systems *in vitro* [124, 125], whether or not they prevail in living cells remains to be ascertained.

## 6.4 Discussion



**Figure 6.3: Summary depiction of the exemplars of bacterial condensates discussed in this work.** Key, condensate-specific scaffold proteins are included as labels associated with condensates and the species in which they have been reported (Ec = *E. coli*, Cc = *C. crescentus*, and Bs = *B. subtilis*).

We have focused on specific examples (**Figure 6.3**) to make the case that IDRs and biomolecular condensates have relevant and important roles to play in bacteria. In doing so, we have relied on the minimalist definition provided by Banani et al. [2] for designating membraneless bodies as condensates. Whether these bodies form and dissolve via spontaneous (passive) or driven (active) PSBP transitions [53] will need to be resolved via systematic assessments *in vitro* and *in vivo*. There is a debate regarding criteria that should be used for invoking PSBP transitions as the

mechanisms for condensate formation and dissolution in living cells [126]. While this debate is interesting, it is also premature because settling on truly objective criteria will require rigorous assessments of the complexities of PSBP transitions in multicomponent systems, i.e., systems with hundreds of different types of molecules [14, 125, 127], context dependencies of scaffolds versus clients [8, 36], and recognizing the distinctions between clients and other types of non-scaffold molecules that do not just partition into condensates but actually alter the overall phase behavior via thermodynamic linkage [128-130]. For now, what we can assert is that condensates (as per the Banani et al. definition [2]) do form in bacteria and that growing evidence implicates IDRs as central players in crucial activities that control bacterial life cycles.

As we gain deeper insights into the molecular grammar [16] that drives condensate formation via PSBP transitions as well as the complexities of multicomponent systems, we should be able to transfer these insights unto equivalent systems in bacteria. This will require systematic and synergistic multipronged investigations *in vitro* and *in vivo* that take on the array of systems discussed here and new systems that emerge from further analyses of the fascinating complexities in bacteria. Indeed, work in bacteria has already pointed us toward novel cell-wide phase transitions that might be relevant within sub-cellular regions in eukaryotes. Specifically, metabolic activities within bacteria directly influence the material properties of the cytoplasm. In the absence of metabolic activity, bacterial cytoplasms can undergo transitions into glassy states, and they do so without changes to the temperature of the surroundings [131]. This is interesting because glass transitions are temperature-dependent mobility transitions, whereby molecules become immobilized below the glass transition temperature [132]. This can come about through a combination of temperature-dependent increases in the barriers to breaking inter-sticker crosslinks, random freezing-in of crosslinked multivalent macromolecules, and reduced spacer-driven

reptation (slithering motions) and overall diffusion [133]. It appears that chemical work, in the form of metabolic activity, acts as an intracellular energy source that controls the effective temperature within cells to control the cell-wide glassy behavior. Hence, it is reasonable to expect that local energy gradients provide directed ways of controlling the fluidity of condensates. The relative simplicity of bacterial systems makes them fertile as investigative playgrounds for exploring the synergies among PSBP transitions, glass transitions, and active processes [134].

## 6.5 References

1. Surovtsev, I.V. and C. Jacobs-Wagner, *Subcellular Organization: A Critical Feature of Bacterial Cell Replication*. Cell, 2018. **172**(6): p. 1271-1293.
2. Banani, S.F., et al., *Biomolecular condensates: organizers of cellular biochemistry*. Nat Rev Mol Cell Biol, 2017. **18**(5): p. 285-298.
3. Alberti, S., *Phase separation in biology*. Current Biology, 2017. **27**(20): p. R1097-r1102.
4. Shin, Y. and C.P. Brangwynne, *Liquid phase condensation in cell physiology and disease*. Science, 2017. **357**(6357): p. eaaf4382.
5. Holehouse, A.S. and R.V. Pappu, *Functional Implications of Intracellular Phase Transitions*. Biochemistry, 2018. **57**(17): p. 2415-2423.
6. Powers, S.K., et al., *Nucleo-cytoplasmic Partitioning of ARF Proteins Controls Auxin Responses in Arabidopsis thaliana*. Molecular Cell, 2019. **76**(1): p. 177-190.e5.
7. Banani, S.F., et al., *Compositional Control of Phase-Separated Cellular Bodies*. Cell, 2016. **166**(3): p. 651-663.
8. Su, X., et al., *Phase separation of signaling molecules promotes T cell receptor signal transduction*. Science, 2016. **352**(6285): p. 595-9.
9. Brangwynne, C.P., P. Tompa, and R.V. Pappu, *Polymer physics of intracellular phase transitions*. Nature Physics, 2015. **11**(11): p. 899-904.
10. Posey, A.E., A.S. Holehouse, and R.V. Pappu, *Phase separation of intrinsically disordered proteins*, in *Methods in enzymology*. 2018, Elsevier. p. 1-30.
11. Rubinstein, M. and A.V. Dobrynin, *Solutions of Associative Polymers*. Trends in Polymer Science, 1997. **5**(6): p. 181-186.
12. Nott, T.J., et al., *Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles*. Mol. Cell, 2015. **57**(5): p. 936-947.
13. Vernon, R.M., et al., *Pi-Pi contacts are an overlooked protein feature relevant to phase separation*. eLife 2018. **7**: p. e31486.

14. Choi, J.-M., F. Dar, and R.V. Pappu, *LASSI: A lattice model for simulating phase transitions of multivalent proteins*. PLoS computational biology, 2019. **15**(10).
15. Choi, J.-M., A.S. Holehouse, and R.V. Pappu, *Physical Principles Underlying the Complex Biology of Intracellular Phase Transitions*. Annual Review of Biophysics, 2020. **49**: p. 107-133.
16. Wang, J., et al., *A Molecular Grammar Governing the Driving Forces for Phase Separation of Prion-like RNA Binding Proteins*. Cell, 2018. **174**(3): p. 688-699. e16.
17. Feric, M., et al., *Coexisting liquid phases underlie nucleolar subcompartments*. Cell, 2016. **165**(7): p. 1686-1697.
18. Harmon, T.S., et al., *Intrinsically disordered linkers determine the interplay between phase separation and gelation in multivalent proteins*. eLife, 2017. **6**: p. e30294.
19. Harmon, T.S., A.S. Holehouse, and R.V. Pappu, *Differential solvation of intrinsically disordered linkers drives the formation of spatially organized droplets in ternary systems of linear multivalent proteins*. New Journal of Physics, 2018. **20**(4): p. 045002.
20. Semenov, A.N. and M. Rubinstein, *Thermoreversible Gelation in Solutions of Associative Polymers. I. Statics*. Macromolecules, 1998. **31**(4): p. 1373-1385.
21. Flory, P.J., *Molecular Size Distribution in Three Dimensional Polymers. I. Gelation*. Journal of the American Chemical Society, 1941. **63**(11): p. 3083-3090.
22. Stockmayer, W.H., *Theory of Molecular Size Distribution and Gel Formation in Branched-Chain Polymers*. The Journal of Chemical Physics, 1943. **11**(2): p. 45-55.
23. Bracha, D., et al., *Mapping Local and Global Liquid Phase Behavior in Living Cells Using Photo-Oligomerizable Seeds*. Cell, 2018. **175**(6): p. 1467-1480.e13.
24. Ghosh, A., K. Mazarakos, and H.-X. Zhou, *Three archetypical classes of macromolecular regulators of protein liquid–liquid phase separation*. Proceedings of the National Academy of Sciences, 2019. **116**(39): p. 19474.
25. Franzmann, T.M., et al., *Phase separation of a yeast prion protein promotes cellular fitness*. Science, 2018. **359**(6371): p. eaao5654.
26. Maharana, S., et al., *RNA buffers the phase separation behavior of prion-like RNA binding proteins*. Science, 2018. **360**(6391): p. 918-921.



27. Innocenzi, P., *Understanding sol–gel transition through a picture. A short tutorial.* Journal of Sol-Gel Science and Technology, 2020: p. 1-7.
28. Almdal, K., et al., *Towards a phenomenological definition of the term ‘gel’.* Polymer Gels and Networks, 1993. **1**(1): p. 5-17.
29. Martin, E.W., et al., *Valence and patterning of aromatic residues determine the phase behavior of prion-like domains.* Science, 2020. **367**(6478): p. 694-699.
30. Lin, Y.-H., J.D. Forman-Kay, and H.S. Chan, *Sequence-Specific Polyampholyte Phase Separation in Membraneless Organelles.* Physical Review Letters, 2016. **117**(17): p. 178101.
31. Lin, Y.H. and H.S. Chan, *Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated.* Biophys J, 2017. **112**(10): p. 2043-2046.
32. Choi, J.-M., A.A. Hyman, and R.V. Pappu, *Generalized models for bond percolation transitions of associative polymers.* Physical Review E, 2020. **102**(4): p. 042403.
33. Dignon, G.L., et al., *Sequence determinants of protein phase behavior from a coarse-grained model.* PLoS Computational Biology, 2018. **14**(1): p. e1005941.
34. McCarty, J., et al., *Complete Phase Diagram for Liquid–Liquid Phase Separation of Intrinsically Disordered Proteins.* The Journal of Physical Chemistry Letters, 2019. **10**(8): p. 1644-1652.
35. Alberti, S., et al., *A User's Guide for Phase Separation Assays with Purified Proteins.* Journal of Molecular Biology 2018. **430**(23): p. 4806-4820.
36. Ditlev, J.A., L.B. Case, and M.K. Rosen, *Who's In and Who's Out-Compositional Control of Biomolecular Condensates.* Journal of Molecular Biology 2018. **430**(23): p. 4666-4684.
37. Roberts, S., et al., *Injectable tissue integrating networks from recombinant polypeptides with tunable order.* Nature Materials, 2018. **17**(12): p. 1154-1163.
38. Wei, M.T., et al., *Phase behaviour of disordered proteins underlying low density and high permeability of liquid organelles.* Nature Chemistry, 2017. **9**(11): p. 1118-1125.
39. Brangwynne, C.P., et al., *Germline P granules are liquid droplets that localize by controlled dissolution/condensation.* Science, 2009. **324**(5935): p. 1729-1732.

40. Brangwynne, C.P., T.J. Mitchison, and A.A. Hyman, *Active liquid-like behavior of nucleoli determines their size and shape in Xenopus laevis oocytes*. Proceedings of the National Academy of Sciences USA, 2011. **108**(11): p. 4334-4339.
41. Jawerth, L.M., et al., *Salt-Dependent Rheology and Surface Tension of Protein Condensates Using Optical Traps*. Physical Review Letters, 2018. **121**(25): p. 258101.
42. Mitrea, D.M., et al., *Nucleophosmin integrates within the nucleolus via multi-modal interactions with proteins displaying R-rich linear motifs and rRNA*. eLife, 2016. **5**.
43. Dias, C.S., N.A.M. Araújo, and M.M. Telo da Gama, *Dynamics of network fluids*. Advances in Colloid and Interface Science, 2017. **247**: p. 258-263.
44. Shin, Y., et al., *Spatiotemporal Control of Intracellular Phase Transitions Using Light-Activated optoDroplets*. Cell, 2017. **168**(1-2): p. 159-171.e14.
45. Woodruff, J.B., et al., *The Centrosome Is a Selective Condensate that Nucleates Microtubules by Concentrating Tubulin*. Cell, 2017. **169**(6): p. 1066-1077 e10.
46. Mitrea, D.M. and R.W. Kriwacki, *Phase separation in biology; functional organization of a higher order*. Cell Commun. Signal., 2016. **14**(1): p. 1-20.
47. Hyman, A.A., C.A. Weber, and F. Jülicher, *Liquid-liquid phase separation in biology*. Annu. Rev. Cell Dev. Biol., 2014. **30**: p. 39-58.
48. Taylor, N., et al., *Biophysical characterization of organelle-based RNA/protein liquid phases using microfluidics*. Soft Matter, 2016. **12**(45): p. 9142-9150.
49. Chandler, D., J.D. Weeks, and H.C. Andersen, *Van der Waals Picture of Liquids, Solids, and Phase Transformations*. Science, 1983. **220**(4599): p. 787.
50. Witten, T., A., and P. Pincus, A., *Structured fluids: Polymers, colloids, surfactants*. 2004, New York: Oxford University Press. 216.
51. Abbondanzieri, E.A. and A.S. Meyer, *More than just a phase: the search for membraneless organelles in the bacterial cytoplasm*. Current Genetics, 2019. **65**(3): p. 691-694.
52. van der Lee, R., et al., *Classification of Intrinsically Disordered Regions and Proteins*. Chemical Reviews, 2014. **114**(13): p. 6589-6631.

53. Berry, J., C.P. Brangwynne, and M. Haataja, *Physical principles of intracellular organization via active and passive phase transitions*. Reports on Progress in Physics, 2018. **81**(4): p. 046601.
54. Lin, Y., S.L. Currie, and M.K. Rosen, *Intrinsically disordered sequences enable modulation of protein phase separation through distributed tyrosine motifs*. Journal of Biological Chemistry, 2017. **292**(46): p. 19110-19120.
55. Putnam, A., et al., *A gel phase promotes condensation of liquid P granules in Caenorhabditis elegans embryos*. Nature Structural and Molecular Biology, 2019. **26**(3): p. 220-226.
56. Mittag, T. and R. Parker, *Multiple modes of protein-protein interactions promote RNP granule assembly*. Journal of Molecular Biology, 2018. **430**(23): p. 4636-4649.
57. Buske, P.J. and P.A. Levin, *Extreme C terminus of bacterial cytoskeletal protein FtsZ plays fundamental role in assembly independent of modulatory proteins*. Journal of Biological Chemistry, 2012. **287**(14): p. 10945-57.
58. Buske, P.J. and P.A. Levin, *A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo*. Molecular Microbiology, 2013. **89**(2): p. 249-63.
59. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division*. Seminars in Cell and Developmental Biology, 2015. **37**: p. 3-10.
60. Sundararajan, K. and E.D. Goley, *Cytoskeletal Proteins in Caulobacter crescentus: Spatial Orchestrators of Cell Cycle Progression, Development, and Cell Shape*, in *Sub-cellular biochemistry*. 2017. p. 103-137.
61. Sundararajan, K. and E.D. Goley, *The intrinsically disordered C-terminal linker of FtsZ regulates protofilament dynamics and superstructure in vitro*. The Journal of Biological Chemistry, 2017. **292**(50): p. 20509-20527.
62. Kozlov, A.G., et al., *Intrinsically disordered C-terminal tails of E. coli single-stranded DNA binding protein regulate cooperative binding to single-stranded DNA*. J Mol Biol, 2015. **427**(4): p. 763-74.
63. Shinn, M.K., et al., *Are the intrinsically disordered linkers involved in SSB binding to accessory proteins?* Nucleic Acids Research, 2019. **47**(16): p. 8581-8594.

64. Jakob, U., R. Kriwacki, and V.N. Uversky, *Conditionally and transiently disordered proteins: awakening cryptic disorder to regulate protein function*. Chemical Reviews, 2014. **114**(13): p. 6779-805.
65. Gu, S., et al., *The role of intrinsic disorder and dynamics in the assembly and function of the type II secretion system*. Biochimica et Biophysica Acta, 2017. **1865**(10): p. 1255-1266.
66. Gruszka, D.T., et al., *Disorder drives cooperative folding in a multidomain protein*. Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(42): p. 11841-11846.
67. Ding, L., et al., *Functional characterization of FlgM in the regulation of flagellar synthesis and motility in Yersinia pseudotuberculosis*. Microbiology, 2009. **155**(Pt 6): p. 1890-900.
68. Bowman, G.R., et al., *Oligomerization and higher-order assembly contribute to sub-cellular localization of a bacterial scaffold*. Molecular microbiology, 2013. **90**(4): p. 776-795.
69. Bi, E.F. and J. Lutkenhaus, *FtsZ ring structure associated with division in Escherichia coli*. Nature, 1991. **354**(6349): p. 161-4.
70. den Blaauwen, T., L.W. Hamoen, and P.A. Levin, *The divisome at 25: the road ahead*. Current Opinion in Microbiology, 2017. **36**: p. 85-94.
71. Yang, X., et al., *GTPase activity-coupled treadmilling of the bacterial tubulin FtsZ organizes septal cell wall synthesis*. Science, 2017. **355**(6326): p. 744.
72. Wagstaff, J. and J. Löwe, *Prokaryotic cytoskeletons: protein filaments organizing small cells*. Nature Reviews Microbiology, 2018. **16**(4): p. 187-201.
73. Anderson, D.E., F.J. Gueiros-Filho, and H.P. Erickson, *Assembly dynamics of FtsZ rings in Bacillus subtilis and Escherichia coli and effects of FtsZ-regulating proteins*. Journal of Bacteriology, 2004. **186**(17): p. 5775-81.
74. Löwe, J. and L.A. Amos, *Crystal structure of the bacterial cell-division protein FtsZ*. Nature, 1998. **391**(6663): p. 203-206.
75. Erickson, H.P., *Modeling the physics of FtsZ assembly and force generation*. Proceedings of the National Academy of Sciences, 2009. **106**(23): p. 9238.

76. Cohan, M.C., K.M. Ruff, and R.V. Pappu, *Information theoretic measures for quantifying sequence-ensemble relationships of intrinsically disordered proteins*. *Protein Eng Des Sel*, 2019. **32**(4): p. 191-202.
77. Gardner, K.A.J.A., D.A. Moore, and H.P. Erickson, *The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide*. *Molecular Microbiology*, 2013. **89**(2): p. 264-275.
78. Cohan, M.C., et al., *Dissecting the Functional Contributions of the Intrinsically Disordered C-terminal Tail of Bacillus subtilis FtsZ*. *J Mol Biol*, 2020. **432**(10): p. 3205-3221.
79. Chen, Y., et al., *A rapid fluorescence assay for FtsZ assembly indicates cooperative assembly with a dimer nucleus*. *Biophysical Journal*, 2005. **88**(1): p. 505-14.
80. Aarsman, M.E.G., et al., *Maturation of the Escherichia coli divisome occurs in two steps*. *Molecular Microbiology*, 2005. **55**(6): p. 1631-1645.
81. den Blaauwen, T., L.W. Hamoen, and P.A. Levin, *The divisome at 25: the road ahead*. *Curr Opin Microbiol*, 2017. **36**: p. 85-94.
82. Lopez-Montero, I., et al., *Intrinsic disorder of the bacterial cell division protein ZipA: coil-to-brush conformational transition*. *FASEB J*, 2013. **27**(8): p. 3363-75.
83. Yan, K., K.H. Pearce, and D.J. Payne, *A conserved residue at the extreme C-terminus of FtsZ is critical for the FtsA-FtsZ interaction in Staphylococcus aureus*. *Biochemical and biophysical research communications*, 2000. **270**(2): p. 387-392.
84. Krupka, M., et al., *Role of the FtsA C terminus as a switch for polymerization and membrane association*. *mBio*, 2014. **5**(6): p. e02221.
85. Schoenemann, K.M., D.E. Vega, and W. Margolin, *Peptide Linkers within the Essential FtsZ Membrane Tethers ZipA and FtsA Are Nonessential for Cell Division*. *J Bacteriol*, 2020. **202**(6): p. e00720-19.
86. Sogues, A., et al., *Essential dynamic interdependence of FtsZ and SepF for Z-ring and septum formation in Corynebacterium glutamicum*. *Nature Communications*, 2020. **11**(1): p. 1641.
87. Adams, D.W., L.J. Wu, and J. Errington, *Nucleoid occlusion protein Noc recruits DNA to the bacterial cell membrane*. *The EMBO journal*, 2015. **34**(4): p. 491-501.

88. Errington, J. and L.J. Wu, *Cell cycle machinery in Bacillus subtilis*, in *Prokaryotic Cytoskeletons*. 2017, Springer. p. 67-101.
89. Monterroso, B., et al., *Bacterial FtsZ protein forms phase-separated condensates with its nucleoid-associated inhibitor SlmA*. EMBO Reports, 2019. **20**(1).
90. Bernhardt, T.G. and P.A. De Boer, *SlmA, a nucleoid-associated, FtsZ binding protein required for blocking septal ring assembly over chromosomes in E. coli*. Molecular cell, 2005. **18**(5): p. 555-564.
91. Monterroso, B., et al., *Microenvironments created by liquid-liquid phase transition control the dynamic distribution of bacterial division FtsZ protein*. Scientific Reports, 2016. **6**: p. 35140.
92. Holmes, J.A., et al., *Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles*. Proceedings of the National Academy of Sciences, 2016. **113**(44): p. 12490-12495.
93. Ptacin, J.L., et al., *Bacterial scaffold directs pole-specific centromere segregation*. Proceedings of the National Academy of Sciences, 2014. **111**(19): p. E2046-E2055.
94. Bowman, G.R., et al., *A polymeric protein anchors the chromosomal origin/ParB complex at a bacterial cell pole*. Cell, 2008. **134**(6): p. 945-955.
95. Ebersbach, G., et al., *A self-associating protein critical for chromosome attachment, division, and polar organization in caulobacter*. Cell, 2008. **134**(6): p. 956-68.
96. Holmes, J.A., et al., *Caulobacter PopZ forms an intrinsically disordered hub in organizing bacterial cell poles*. Proc Natl Acad Sci U S A, 2016. **113**(44): p. 12490-12495.
97. Lasker, K., et al., *Selective sequestration of signalling proteins in a membraneless organelle reinforces the spatial regulation of asymmetry in Caulobacter crescentus*. Nature Microbiology, 2020.
98. Zhao, W., et al., *A circuit of protein-protein regulatory interactions enables polarity establishment in a bacterium*. bioRxiv, 2018: p. 503250.
99. Curtis, P.D. and Y.V. Brun, *Getting in the loop: regulation of development in Caulobacter crescentus*. Microbiol Mol Biol Rev, 2010. **74**(1): p. 13-41.

100. Curtis, P.D., et al., *The scaffolding and signalling functions of a localization factor impact polar development*. Mol Microbiol, 2012. **84**(4): p. 712-35.
101. Perez, A.M., et al., *A localized complex of two protein oligomers controls the orientation of cell polarity*. MBio, 2017. **8**(1): p. e02238-16.
102. Guilhas, B., et al., *ATP-Driven Separation of Liquid Phase Condensates in Bacteria*. Molecular Cell, 2020. **79**(2): p. 293-303.e4.
103. Kozlov, A.G., et al., *Glutamate promotes SSB protein–protein Interactions via intrinsically disordered regions*. Journal of Molecular Biology, 2017. **429**(18): p. 2790-2801.
104. Sugar, I.P., *Cooperativity and classification of phase transitions. Application to one- and two-component phospholipid membranes*. The Journal of Physical Chemistry, 1987. **91**(1): p. 95-101.
105. Harami, G.M., et al., *Phase separation by ssDNA binding protein controlled via protein-protein and protein-DNA interactions*. Proc Natl Acad Sci U S A, 2020. **117**(42): p. 26206-26217.
106. Riback, J.A., et al., *Stress-Triggered Phase Separation Is an Adaptive, Evolutionarily Tuned Response*. Cell, 2017. **168**(6): p. 1028-1040 e19.
107. Karas, V.O., I. Westerlaken, and A.S. Meyer, *The DNA-Binding Protein from Starved Cells (Dps) Utilizes Dual Functions To Defend Cells against Multiple Stresses*. Journal of Bacteriology, 2015. **197**(19): p. 3206-3215.
108. Antipov, S.S., et al., *The nucleoid protein Dps binds genomic DNA of Escherichia coli in a non-random manner*. PLoS One, 2017. **12**(8): p. e0182800.
109. Weber, S.C., A.J. Spakowitz, and J.A. Theriot, *Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm*. Physical review letters, 2010. **104**(23): p. 238102.
110. Weber, S.C., *Sequence-encoded material properties dictate the structure and function of nuclear bodies*. Curr Opin Cell Biol, 2017. **46**: p. 62-71.
111. Weng, X., et al., *Spatial organization of RNA polymerase and its relationship with transcription in Escherichia coli*. Proc Natl Acad Sci U S A, 2019. **116**(40): p. 20115-20123.

112. Ladouceur, A.-M., et al., *Clusters of bacterial RNA polymerase are biomolecular condensates that assemble through liquid–liquid phase separation*. Proceedings of the National Academy of Sciences, 2020. **117**(31): p. 18540-18549.
113. Ait-Bara, S. and A.J. Carpousis, *RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs*. Molecular Microbiology, 2015. **97**(6): p. 1021-135.
114. Ait-Bara, S., A.J. Carpousis, and Y. Quentin, *RNase E in the gamma-Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains*. Molecular Genetics and Genomics, 2015. **290**(3): p. 847-62.
115. Al-Husini, N., et al., *alpha-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies*. Molecular Cell 2018. **71**(6): p. 1027-1039.e14.
116. Hondele, M., et al., *DEAD-box ATPases are global regulators of phase-separated organelles*. Nature, 2019. **573**(7772): p. 144-148.
117. Pak, C.W., et al., *Sequence determinants of intracellular phase separation by complex coacervation of a disordered protein*. Molecular Cell 2016. **63**(1): p. 72–85.
118. Ait-Bara, S. and A.J. Carpousis, *RNA degradosomes in bacteria and chloroplasts: classification, distribution and evolution of RNase E homologs*. Mol Microbiol, 2015. **97**(6): p. 1021-135.
119. Ait-Bara, S., A.J. Carpousis, and Y. Quentin, *RNase E in the gamma-Proteobacteria: conservation of intrinsically disordered noncatalytic region and molecular evolution of microdomains*. Mol Genet Genomics, 2015. **290**(3): p. 847-62.
120. Racki, L.R., et al., *Polyphosphate granule biogenesis is temporally and functionally tied to cell cycle exit during starvation in Pseudomonas aeruginosa*. Proceedings of the National Academy of Sciences, 2017. **114**(12): p. E2440-E2449.
121. Sing, C.E., *Development of the modern theory of polymeric complex coacervation*. Advances in Colloid and Interface Science, 2017. **239**: p. 2-16.
122. Hauryliuk, V., et al., *Recent functional insights into the role of (p)ppGpp in bacterial physiology*. Nature reviews. Microbiology, 2015. **13**(5): p. 298-309.
123. Aumiller, W.M. and C.D. Keating, *Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles*. Nature Chemistry, 2016. **8**(2): p. 129-137.



124. Fanalista, F., et al., *FtsZ-Induced Shape Transformation of Coacervates*. *Advanced Biosystems*, 2018. **2**(9): p. 1800136.
125. Banerjee, P.R., et al., *Reentrant Phase Transition Drives Dynamic Substructure Formation in Ribonucleoprotein Droplets*. *Angewandte Chemie International Edition*, 2017. **56**(38): p. 11354-11359.
126. McSwiggen, D.T., et al., *Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences*. *Genes & Development*, 2019. **33**(23-24): p. 1619-1634.
127. Riback, J.A., et al., *Composition-dependent thermodynamics of intracellular phase separation*. *Nature*, 2020. **581**(7807): p. 209-214.
128. Yoshizawa, T., et al., *Nuclear Import Receptor Inhibits Phase Separation of FUS through Binding to Multiple Sites*. *Cell*, 2018. **173**(3): p. 693-705.e22.
129. Posey, A.E., et al., *Profilin reduces aggregation and phase separation of huntingtin N-terminal fragments by preferentially binding to soluble monomers and oligomers*. *Journal of Biological Chemistry*, 2018. **293**: p. 3734-3746.
130. Wyman, J. and S.J. Gill, *Ligand-linked phase changes in a biological system: applications to sickle cell hemoglobin*. *Proceedings of the National Academy of Sciences*, 1980. **77**(9): p. 5239.
131. Parry, Bradley R., et al., *The Bacterial Cytoplasm Has Glass-like Properties and Is Fluidized by Metabolic Activity*. *Cell*, 2014. **156**(1): p. 183-194.
132. Debenedetti, P.G. and F.H. Stillinger, *Supercooled liquids and the glass transition*. *Nature*, 2001. **410**(6825): p. 259-267.
133. Munder, M.C., et al., *A pH-driven transition of the cytoplasm from a fluid- to a solid-like state promotes entry into dormancy*. *eLife*, 2016. **5**: p. e09347.
134. Halfmann, R., *A glass menagerie of low complexity sequences*. *Current Opinion in Structural Biology*, 2016. **38**: p. 18-25.

# Chapter 7

## Concluding Remarks and Future Directions

### 7.1 Preamble

This dissertation establishes that the hypervariable disordered C-terminal tail of FtsZ is not only essential for function but contains system-specific sequence encoded features. Previous studies identified the essentiality of each of the modular components of *B. subtilis* FtsZ. Here, we establish that the CTT, containing a C-terminal intra-and intermolecular interaction motif (CTP) and a disordered linker (CTL), has a sticker-and-spacer architecture, where the CTL modulates the interactions of the CTP (**Chapter 2**). We find that in this capacity, the modules of the CTT not only influence FtsZ assembly but also impart an auto-regulatory function on the active GTPase domain. While the sequence of the CTP is mostly conserved, the linear sequence of the CTL is hypervariable, having significant variations in both length and composition across a system of orthologs. This clouds the ability to utilize any evolutionarily conserved features to understand sequence-to-function relationships. **Chapters 3-4** were dedicated to method development to parse

out potentially conserved features of hypervariable sequences. These methods involved comparing conformational ensembles of the disordered domains and non-random sequence encoded features. In **Chapter 5**, we show that the preservation of spacer-like behavior, conformational ensemble, and sequence features is essential for a functional CTL. We additionally highlight recent discoveries in the disordered proteome of bacteria. Leveraging concepts learned from eukaryotic systems, we make a case for the involvement of disordered proteins in bacterial spatiotemporal control (**Chapter 6**). In the following chapter, we summarize the advancements made through this body of work, acknowledge unanswered questions, and suggest potential next steps in building on these results to investigate further the properties of other bristled and assembling enzymes and other disordered domains in bacteria. Last, we comment on how these results might be applied in the formulation of next-generation antibiotics.

## 7.2 Concluding Remarks

Across all domains of life, the onset of cell division is hallmarked by the formation of a cytokinetic ring at the center of the cell [1]. Bacteria are no exception; cell division in bacteria is a tightly controlled process and involves regulatory factors that participate in several roles [2, 3]. One such imperative role involves modulating the assembly of the cytokinetic ring, termed the Z-ring after the protein that makes up its foundation, FtsZ, the bacterial homolog of tubulin [2, 4-7] [8].

Like tubulin, FtsZ is an assembling GTPase, where GTP binding promotes FtsZ monomer assembly into polymers [7, 9-11]. This process is cooperative, and as such, there is a critical concentration of FtsZ ( $\sim 1 \mu\text{M}$ ) required to enable the assembly of protofilaments (FtsZ polymers) [12-15]. In this process, the formation of the FtsZ dimer is considered to be the nucleation step for subsequent polymerization [12, 16-18]. In the context of the Z-ring, longitudinal FtsZ protofilaments are laterally associated into bundles to generate the apparent wreath-like structure [5, 19-29]. While the GTPase domain drives FtsZ polymerization, the modulation and formation of higher-order structures involve domains outside of the folded core [30-36].

FtsZ has a bristled architecture, meaning that its folded domain is flanked by a disordered tail [31, 35, 37, 38]. This C-terminal tail (CTT) contains two essential domains: a conserved interaction peptide (CTP) and a non-conserved linker (CTL) that connects the CTP to the core [8]. Deletion of the CTP results in a loss of bundling *in vitro* [31], and it is the known interaction module with many factors that promote bundling *in vivo* [18, 19, 23, 39]. Whether the bundling is driven by modulatory proteins involved in FtsZ crosslinking by interacting directly with the CTP [18, 19, 23, 39], driven by FtsZ condensation through interactions mediated by the CTP [24, 30, 32, 34, 37,

40], or a result of some combination of these intra- and intermolecular interactions is a topic of debate in the FtsZ field [2]. Importantly, recent work has revealed that one size does not fit all to describe FtsZ and Z-ring assembly, and there are species-specific mechanisms that relate to different *in vivo* Z-ring morphologies, dynamics, and functions [2]. In all studied systems, however, the CTT plays an imperative role in coordinating higher-order assemblies, and its deletion results in cell death [32, 36, 40].

In this dissertation, we focus on the FtsZ from *B. subtilis* (Bs-FtsZ), an ortholog known to bundle on its own through interactions mediated by the CTP [4, 20, 30-32]. With Bs-FtsZ, the appearance of bundles *in vitro* has interestingly corresponded with an observed second critical concentration in FtsZ assembly [14, 30, 41]. Therefore, it was hypothesized that bundling causes this threshold behavior [14, 41]. While the CTP is involved in bundling [31], the molecular function of the CTL in the context of FtsZ assembly had not yet been elucidated. Furthermore, deletions and substitutions of the CTL resulted in an impacted GTPase [32, 34-36, 40]; however, these observations were convoluted by several factors, including the differing assembly states of the FtsZs.

In **Chapter 2** [42], we focused on parsing out the contributions of each module of FtsZ to the assembly and activity of FtsZ. Our data collectively suggest that the CTT acts as an *auto-regulator* of assembly and an *auto-inhibitor* of enzymatic activity. In the construct that lacks the CTL, the CTP enables the formation of alternative mini-ring structures (**Figure 2.9**), whereas, in the WT, the CTP enables higher-order assemblies. In accord with previous findings [31, 32], these assemblies likely arise through CTP-mediated electrostatic interactions (**Figure 2.10**). Conversely, we find that the CTL weakens the driving forces for forming single-stranded polymers and higher-

order assemblies, leading to shorter, single-stranded polymers than the FtsZ that lacks the CTT and the FtsZ that lacks the CTL. From a functional standpoint, the CTL acts as a *spacer* with a finite excluded volume and helps in spatially separating the CTP from the core of Bs-FtsZ – a feature that has been established by Huecas et al. for different FtsZs [30]. The ability of the CTL to function as a spacer also helps alleviate interactions involving the CTP, and these assembly suppressing activities appear to ensure that the CTP engages primarily in interactions that drive the requisite higher-order assemblies. Taken together, our data suggest that the CTT has a *sticker and spacer* architecture and that the CTT is an auto-regulator of FtsZ assembly.

In the context of FtsZ activity, the CTT acts as an auto-inhibitor. In the absence of the tail, FtsZ is a more efficient GTPase. The reduction of efficiency in the presence of the CTT stems from the CTL and CTP having distinct modulatory functions. Using Michaelis-Menten kinetics, we describe the impacts of each module on the active domain and find that while the CTL weakens the apparent ability of GTP to bind to the core, the CTP reduces the catalytic rate. These modules together have a compensatory effect, and the overall reduction of efficiency is non-additive [42]. It appears then that the GTPase activity rate is “fine-tuned” by the CTT. The auto-regulatory and inhibitory behaviors of a disordered domain are not unique to Bs-FtsZ, as there are several examples where an IDR imparts an auto-regulatory function [43-48]. Given the reduced proteome size of bacteria and yet the tight regulation of critical processes, one could reason that bacteria might frequently take advantage of this type of enzyme architecture, as it imparts control without requiring additional proteins. The disordered domain of bacterial chaperone Hfq is one such example [45, 49]. Additionally, the divisome, a network of proteins that imparts incredible control on the cell cycle, includes several enzymes with disordered domains [3, 20]. Therefore, we reason that other FtsZ CTT-like architectures and functions may exist within the divisome.

While the results described here are specific to Bs-FtsZ, they present a universal sticker-spacer architecture for the CTT of FtsZ that is in accordance with other previously reported findings [30-36]. We propose that while the sequence of the CTP may dictate the type of interactions in which the FtsZ can participate (i.e., *in vivo* modulatory proteins and *in vitro* protofilament crosslinking), the non-conserved CTL regulates the strength of the interactions, the apparent valence of interaction sites, and the material properties of the resulting structures. Therefore, on the zeroth order, discrepancies amongst FtsZ structures and dynamics could then be attributed to the sequence-encoded differences of the CTT.

The motif architecture of the CTP can be mostly compared using the conventional tools deployed for folded domains. The sequence features of the CTL, however, are more difficult to parse out. As common with intrinsically disordered proteins / regions (IDPs / IDRs), the CTL is hypervariable, meaning that despite contributing to similar functions, the linear amino acid sequence and the overall sequence composition is not conserved. This is because unlike folded domains, IDRs/IDPs are characterized by having no defining and globally stable secondary structure; instead, IDRs/IDPs sample a wide range of conformations [50, 51]. Often, the heterogeneity encoded within the ensemble of conformations is imperative to IDR/IDP function, and linear sequences that result in changes to the conformational ensemble have been shown to impact cellular phenotype [46, 52-57]. Importantly, IDPs/IDRs are also common sites of interaction / binding motifs; therefore, within an apparently hypervariable domain, the context and valence of these motifs can also be imperative to function [43, 58-61]. The totality of these observations led to the need for new methodologies that considered the conservation of sequence-ensemble relationships (SERs) (**Chapter 3**) and the presence and positioning of non-random sequence-encoded motifs to compare hypervariable IDPs/IDRs (**Chapter 4**).

In **Chapters 3** and **4**, we focused on building tools that enable the analysis of conserved sequence-encoded properties that might have relevance to the observed functions. To briefly summarize **Chapter 3**, the information theoretic entropy of the conformational ensemble was calculated by repurposing a traditional communication channel to apply to SERs. From all-atom simulations, a distribution of sizes and shapes was generated, representing the totality of the sampled conformations. The probabilities of observing a conformation with a specific range of sizes and shapes were used to quantify the information theoretic entropy of a sequence. These values were compared across CTTs from the system of FtsZ orthologs. Interestingly, despite having disparate sequences and sequence compositions, most FtsZ CTT sequences encode for similar SERs. This might reflect a form of convergent evolution whereby different CTL sequences are interoperable because different sequences serve the functionality of being spacers.

**Chapter 4** focused on identifying non-random sequence features within hypervariable sequences. The method involved generating random scrambles of the sequence of interest, calculating the values of the sequence parameter of interest for each scramble, fitting these values to a distribution, and comparing the observed value to the expected value. The more significant the deviation from the expected value, quantified by the z-score, the more non-random the observed value is. This analysis was performed on CTL sequences and showed that each CTL sequence encoded for a specific combination of sequence features that were random / non-random. This variation could be an example of divergent evolution, whereby changes to the CTL sequences engender different functionalities in different bacteria.

Are the CTLs of FtsZ then an example of convergent evolution (spacer behavior / similar SERs), divergent evolution (system-specific non-random sequence patterning), or a combination of



both? Based on the preceding analysis, we hypothesized that the CTL of FtsZ might be a combination of both and that these features are encoded within the amino acid sequence. If this were true, designed CTL sequence scrambles that result in significant changes to these features would perturb function. To answer this question, we used the patterning of oppositely charged residues within the CTL sequence as the design parameter (quantified by the parameter  $\kappa$  [55]) to generate scramble variants of the CTL (**Chapter 5**). This parameter was chosen for multiple reasons: (1) the fraction of charged residues is significantly high (> 30%) with an approximately equal balance of basic and acidic residues, (2) the segregation of oppositely charged residues is known to impact the SERs [55, 62], (3) the random alteration of one patterning parameter will result in variations to other patterning parameters, and (4) large blocks of charge residues are known to engage in electrostatic-mediated interactions that can promote sticker-like behavior [63-67]. Lastly, there appeared to be some evolutionary selection against  $\kappa$  values that represent the segregation of opposite charges, though as illustrated in **Chapter 4**, this observation was not properly normalized for composition [37].

Using these sequence designs, we found that the CTL sequences that encoded for sticker-like behavior and caused significant perturbations to SERs resulted in FtsZs that did not support Z-ring formation or cell division. Some discrepancies were only manifested *in vitro*, showing deviations from wild type behavior in terms of FtsZ assembly and activity. These differences could only be explained by the divergences of the residue patterning as these sequences largely preserved SERs and spacer behavior. The Bs-FtsZ CTL sequence contains non-random patterning of its polar and acidic residues; this implies that the positioning of these residue types could be functionally significant in the context of a sequence that is otherwise random. Interestingly, the CTL sequence that most closely preserved the sequence patterns of the wild type behaved the most similarly to

wild type. These results imply that, along with composition and length, the design of CTL variants must consider the strength of the spacer, the non-random / random sequence patterns, and the conformational ensemble.

Compared to wild type, CTL sequences that most closely approximated fully random sequences did not support equivalent levels of higher-order assembly and had higher GTPase activities. Based on **Chapter 2**, these results indicate that these CTLs can be thought of as strong spacers. While this behavior largely did not impact *in vivo* function in optimal growth conditions, CTLs could contribute to robustness. Therefore, the functional discrepancies between strong and optimal spacers may be revealed in non-ideal conditions. Taken together, our findings substantiate the hypothesis that not all disordered sequence scrambles of CTL of the same length can be interoperable within *B. subtilis* FtsZ, implying that CTL sequences are likely an example of both convergent and divergent evolution. Further experiments involving the FtsZs from *B. subtilis*, *E. coli*, *C. crescentus*, and other species can help to elucidate what sequence features are purposefully encoded within each unique CTL sequence.

Despite making up a small percentage (~5%) of the proteomes, bacterial proteins with IDRs, like the CTT of FtsZ, are becoming more implicated in critical bacterial processes, such as those outlined in **Chapter 6**. In both **Chapters 3 and 4**, we show that the developed methodologies for the hypervariable CTL can be and are intended to be used on systems outside of FtsZ. Indeed, the use of these methodologies revealed features that could further explain the observed functions and phenotypes, such as the ability to form biomolecular condensates in bacteria (**Chapter 6**), as was shown with RNases E [67]. Using the z-score method, we found that the charge blocky architecture in the disordered C-terminal domain (CTD) of *C. crescentus* RNase E was decidedly non-random,

and this sequence architecture is also imperative to form the liquid-like condensates in the bacterial cytoplasm that were involved with RNA processing. On the other hand, in the *E. coli* RNase E, which does not form cytoplasmic biomolecular condensates, the oppositely charged residues within the CTD were more randomly dispersed with respect to one another. In this case, the identified non-random features had a functional significance. It follows then that other RNase E CTDs that have similar non-random charge patterns to the *C. crescentus* RNase E might also drive phase separation. The z-score method could enable the identification of these orthologs. Furthermore, in the RNA degradosome system, the exact proteins involved varies from system to system [68]. Therefore, IDRs with this architecture can be searched for across all proteins implicated in this process to identify the relevant player that may help form this complex in each given system.

Unlike with RNase E, it is unclear what sequence features within the intrinsically disordered linker (IDL) of SSBs contribute to its ability to phase separate. The z-score method identified proline and glycine residues as non-randomly situated within the *E. coli* SSB sequence, and this feature is largely conserved across all orthologs of SSB. Such features are reminiscent of elastomeric sequences that are intrinsically disordered and drivers of responsive phase transitions and controllers of elastic responses [69-72]. While this hypothesis requires testing, it provides the basis for a design principle to determine the sequence requirements of a functional IDL in the context of biomolecular condensate formation.

Importantly, further uncovering the functions of bacterial IDRs both in and outside of the context of bacterial biomolecular condensates will help to understand more about the organisms that are becoming an increasing societal threat. According to the CDC and the WHO, we face a growing public health challenge with antibiotic resistance [73, 74]. In 2019, 2.8 million individuals

in the United States contracted an antibiotic-resistant infection, with a mortality rate of 1.25%. Since 2013, the prevalence has increased by 40%, and the overall mortality rate has increased by 13% [75]. Accordingly, the WHO currently lists 13 different bacteria as priority pathogens that pose a potential threat due to gained resistance [74]. Of the 60 new antimicrobial agents currently in clinical development, only approximately half of these agents target these priority pathogens, and as with all clinical trials, their successes are not guaranteed. Further, the WHO has indicated that these clinical-stage therapeutics have limited benefit over existing treatments [74]. Current antibiotics target the bacterial processes identified when bacteria were still thought to be “bags of enzymes” [76], such as disrupting cell membrane function, protein synthesis, DNA / RNA synthesis, and cell wall synthesis. To address this growing public health crisis, we suggest leveraging the new insights brought to bear in this dissertation. Specifically, the bacterial processes tightly controlled by IDPs and the crucial functions performed via the formation of biomolecular condensates represent new mechanisms for which novel antibiotics can be developed.

## 7.3 References

1. Straight, A.F. and C.M. Field, *Microtubules, membranes and cytokinesis*. Curr Biol, 2000. **10**(20): p. R760-70.
2. Barrows, J.M. and E.D. Goley, *FtsZ dynamics in bacterial division: What, how, and why?* Curr Opin Cell Biol, 2020. **68**: p. 163-172.
3. den Blaauwen, T., L.W. Hamoen, and P.A. Levin, *The divisome at 25: the road ahead*. Curr Opin Microbiol, 2017. **36**: p. 85-94.
4. Anderson, D.E., F.J. Gueiros-Filho, and H.P. Erickson, *Assembly dynamics of FtsZ rings in Bacillus subtilis and Escherichia coli and effects of FtsZ-regulating proteins*. J Bacteriol, 2004. **186**(17): p. 5775-81.
5. Bi, E.F. and J. Lutkenhaus, *FtsZ ring structure associated with division in Escherichia coli*. Nature, 1991. **354**(6349): p. 161-4.
6. Bisson-Filho, A.W., et al., *Treadmilling by FtsZ filaments drives peptidoglycan synthesis and bacterial cell division*. Science, 2017. **355**(6326): p. 739-743.
7. Bramhill, D. and C.M. Thompson, *GTP-dependent polymerization of Escherichia coli FtsZ protein to form tubules*. Proc Natl Acad Sci U S A, 1994. **91**(13): p. 5813-7.
8. Vaughan, S., et al., *Molecular evolution of FtsZ protein sequences encoded within the genomes of archaea, bacteria, and eukaryota*. Journal of molecular evolution, 2004. **58**(1): p. 19-29.
9. Huecas, S. and J.M. Andreu, *Polymerization of nucleotide-free, GDP- and GTP-bound cell division protein FtsZ: GDP makes the difference*. FEBS Lett, 2004. **569**(1-3): p. 43-8.
10. Mukherjee, A. and J. Lutkenhaus, *Dynamic assembly of FtsZ regulated by GTP hydrolysis*. EMBO J, 1998. **17**(2): p. 462-9.
11. Mukherjee, A. and J. Lutkenhaus, *Guanine nucleotide-dependent assembly of FtsZ into filaments*. Journal of Bacteriology, 1994. **176**(9): p. 2754-8.
12. Chen, Y., et al., *A rapid fluorescence assay for FtsZ assembly indicates cooperative assembly with a dimer nucleus*. Biophys J, 2005. **88**(1): p. 505-14.

13. Hill, N.S., et al., *A moonlighting enzyme links Escherichia coli cell size with central metabolism*. PLoS Genet, 2013. **9**(7): p. e1003663.
14. Ruiz-Martinez, A., et al., *Efficient Multiscale Models of Polymer Assembly*. Biophys J, 2016. **111**(1): p. 185-96.
15. Ruiz-Martinez, A., et al., *Efficient models of polymerization applied to FtsZ ring assembly in Escherichia coli*. Proc Natl Acad Sci U S A, 2018. **115**(19): p. 4933-4938.
16. Caplan, M.R. and H.P. Erickson, *Apparent cooperative assembly of the bacterial cell division protein FtsZ demonstrated by isothermal titration calorimetry*. Journal of Biological Chemistry, 2003. **278**(16): p. 13784-8.
17. Chen, Y. and H.P. Erickson, *Rapid in vitro assembly dynamics and subunit turnover of FtsZ demonstrated by fluorescence resonance energy transfer*. J Biol Chem, 2005. **280**(23): p. 22549-54.
18. Roach, E.J., et al., *Structure and Mutational Analyses of Escherichia coli ZapD Reveal Charged Residues Involved in FtsZ Filament Bundling*. J Bacteriol, 2016. **198**(11): p. 1683-1693.
19. Buss, J., et al., *A multi-layered protein network stabilizes the Escherichia coli FtsZ-ring and modulates constriction dynamics*. PLoS Genet, 2015. **11**(4): p. e1005128.
20. Errington, J. and L.J. Wu, *Cell cycle machinery in Bacillus subtilis*, in *Prokaryotic Cytoskeletons*. 2017, Springer. p. 67-101.
21. Fu, G., et al., *In vivo structure of the E. coli FtsZ-ring revealed by photoactivated localization microscopy (PALM)*. PLoS One, 2010. **5**(9): p. e12682.
22. Guan, F., et al., *Lateral interactions between protofilaments of the bacterial tubulin homolog FtsZ are essential for cell division*. Elife, 2018. **7**: p. e35578.
23. Gueiros-Filho, F.J. and R. Losick, *A widely conserved bacterial cell division protein that promotes assembly of the tubulin-like protein FtsZ*. Genes & Development, 2002. **16**(19): p. 2544-2556.
24. Huang, K.H., J. Durand-Heredia, and A. Janakiraman, *FtsZ ring stability: of bundles, tubules, crosslinks, and curves*. J Bacteriol, 2013. **195**(9): p. 1859-68.
25. Lutkenhaus, J., *FtsZ ring in bacterial cytokinesis*. Mol Microbiol, 1993. **9**(3): p. 403-9.

26. Lutkenhaus, J. and S. Du, *E. coli cell cycle machinery*, in *Prokaryotic Cytoskeletons*. 2017, Springer. p. 27-65.
27. Marrington, R., et al., *FtsZ fiber bundling is triggered by a conformational change in bound GTP*. *Journal of Biological Chemistry*, 2004. **279**(47): p. 48821-48829.
28. Meier, E.L. and E.D. Goley, *Form and function of the bacterial cytokinetic ring*. *Curr Opin Cell Biol*, 2014. **26**: p. 19-27.
29. Pazos, M., et al., *Z-ring membrane anchors associate with cell wall synthases to initiate bacterial cell division*. *Nat Commun*, 2018. **9**(1): p. 5090.
30. Huecas, S., et al., *Self-Organization of FtsZ Polymers in Solution Reveals Spacer Role of the Disordered C-Terminal Tail*. *Biophys J*, 2017. **113**(8): p. 1831-1844.
31. Buske, P.J. and P.A. Levin, *Extreme C terminus of bacterial cytoskeletal protein FtsZ plays fundamental role in assembly independent of modulatory proteins*. *J Biol Chem*, 2012. **287**(14): p. 10945-57.
32. Buske, P.J. and P.A. Levin, *A flexible C-terminal linker is required for proper FtsZ assembly in vitro and cytokinetic ring formation in vivo*. *Mol Microbiol*, 2013. **89**(2): p. 249-63.
33. Barrows, J.M., et al., *FtsA Regulates Z-Ring Morphology and Cell Wall Metabolism in an FtsZ C-Terminal Linker-Dependent Manner in *Caulobacter crescentus**. *J Bacteriol*, 2020. **202**(7): p. e00693-19.
34. Sundararajan, K. and E.D. Goley, *The intrinsically disordered C-terminal linker of FtsZ regulates protofilament dynamics and superstructure in vitro*. *J Biol Chem*, 2017. **292**(50): p. 20509-20527.
35. Sundararajan, K., et al., *The bacterial tubulin FtsZ requires its intrinsically disordered linker to direct robust cell wall construction*. *Nat Commun*, 2015. **6**: p. 7281.
36. Sundararajan, K., et al., *Species- and C-terminal linker-dependent variations in the dynamic behavior of FtsZ on membranes in vitro*. *Mol Microbiol*, 2018. **110**(1): p. 47-63.
37. Buske, P.J., et al., *An intrinsically disordered linker plays a critical role in bacterial cell division*. *Semin Cell Dev Biol*, 2015. **37**: p. 3-10.
38. Lowe, J. and L.A. Amos, *Crystal structure of the bacterial cell-division protein FtsZ*. *Nature*, 1998. **391**(6663): p. 203-6.

39. Caldas, P., et al., *Cooperative ordering of treadmilling filaments in cytoskeletal networks of FtsZ and its crosslinker ZapA*. Nat Commun, 2019. **10**(1): p. 5744.
40. Gardner, K.A., D.A. Moore, and H.P. Erickson, *The C-terminal linker of Escherichia coli FtsZ functions as an intrinsically disordered peptide*. Mol Microbiol, 2013. **89**(2): p. 264-75.
41. Huecas, S., et al., *Energetics and geometry of FtsZ polymers: nucleated self-assembly of single protofilaments*. Biophys J, 2008. **94**(5): p. 1796-806.
42. Cohan, M.C., et al., *Dissecting the Functional Contributions of the Intrinsically Disordered C-terminal Tail of Bacillus subtilis FtsZ*. J Mol Biol, 2020. **432**(10): p. 3205-3221.
43. van der Lee, R., et al., *Classification of intrinsically disordered regions and proteins*. Chem Rev, 2014. **114**(13): p. 6589-631.
44. Kim, Y., et al., *Intrinsically disordered regions regulate both catalytic and non-catalytic activities of the MutLalpha mismatch repair complex*. Nucleic Acids Res, 2019. **47**(4): p. 1823-1835.
45. Santiago-Frangos, A., et al., *Acidic C-terminal domains autoregulate the RNA chaperone Hfq*. Elife, 2017. **6**: p. e27049.
46. He, F., et al., *Interaction between p53 N terminus and core domain regulates specific and nonspecific DNA binding*. Proc Natl Acad Sci U S A, 2019. **116**(18): p. 8859-8868.
47. Bista, M., M. Petrovich, and A.R. Fersht, *MDMX contains an autoinhibitory sequence element*. Proc Natl Acad Sci U S A, 2013. **110**(44): p. 17814-9.
48. Weems, A. and M. McMurray, *The step-wise pathway of septin hetero-octamer assembly in budding yeast*. Elife, 2017. **6**: p. e23689.
49. Kozlov, A.G., et al., *Glutamate promotes SSB protein-protein Interactions via intrinsically disordered regions*. J Mol Biol, 2017. **429**(18): p. 2790-2801.
50. Lyle, N., R.K. Das, and R.V. Pappu, *A quantitative measure for protein conformational heterogeneity*. J Chem Phys, 2013. **139**(12): p. 121907.
51. Mao, A.H., N. Lyle, and R.V. Pappu, *Describing sequence-ensemble relationships for intrinsically disordered proteins*. Biochemical Journal, 2013. **449**: p. 307-318.



52. Das, R.K., S.L. Crick, and R.V. Pappu, *N-Terminal Segments Modulate the alpha-Helical Propensities of the Intrinsically Disordered Basic Regions of bZIP Proteins*. Journal of Molecular Biology, 2012. **416**(2): p. 287-299.
53. Das, R.K., S.L. Crick, and R.V. Pappu, *N-terminal segments modulate the  $\alpha$ -helical propensities of the intrinsically disordered basic regions of bZIP proteins*. Journal of Molecular Biology, 2012. **416**: p. 287-299.
54. Das, R.K., et al., *Cryptic sequence features within the disordered protein p27Kip1 regulate cell cycle signaling*. Proc Natl Acad Sci U S A, 2016. **113**(20): p. 5616-21.
55. Das, R.K. and R.V. Pappu, *Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues*. Proc Natl Acad Sci U S A, 2013. **110**(33): p. 13392-7.
56. Das, R.K., K.M. Ruff, and R.V. Pappu, *Relating sequence encoded information to form and function of intrinsically disordered proteins*. Curr Opin Struct Biol, 2015. **32**: p. 102-12.
57. Sherry, K.P., et al., *Control of transcriptional activity by design of charge patterning in the intrinsically disordered RAM region of the Notch receptor*. Proc Natl Acad Sci U S A, 2017. **114**(44): p. E9243-E9252.
58. Davey, N.E., M.S. Cyert, and A.M. Moses, *Short linear motifs - ex nihilo evolution of protein regulation*. Cell Commun Signal, 2015. **13**: p. 43.
59. Davey, N.E., et al., *Attributes of short linear motifs*. Mol Biosyst, 2012. **8**(1): p. 268-81.
60. Van Roey, K., et al., *Short linear motifs: ubiquitous and functionally diverse protein interaction modules directing cell regulation*. Chem Rev, 2014. **114**(13): p. 6733-78.
61. Tompa, P., et al., *Intrinsically disordered proteins: emerging interaction specialists*. Curr Opin Struct Biol, 2015. **35**: p. 49-59.
62. Holehouse, A.S., et al., *CIDER: Resources to Analyze Sequence-Ensemble Relationships of Intrinsically Disordered Proteins*. Biophys J, 2017. **112**(1): p. 16-21.
63. Lin, Y.H. and H.S. Chan, *Phase Separation and Single-Chain Compactness of Charged Disordered Proteins Are Strongly Correlated*. Biophys J, 2017. **112**(10): p. 2043-2046.
64. Pak, C.W., et al., *Sequence Determinants of Intracellular Phase Separation by Complex Coacervation of a Disordered Protein*. Mol Cell, 2016. **63**(1): p. 72-85.

65. Sing, C.E., *Development of the modern theory of polymeric complex coacervation*. *Advances in Colloid and Interface Science*, 2017. **239**: p. 2-16.
66. Brady, J.P., et al., *Structural and hydrodynamic properties of an intrinsically disordered region of a germ cell-specific protein on phase separation*. *Proc Natl Acad Sci U S A*, 2017. **114**(39): p. E8194-E8203.
67. Al-Husini, N., et al., *alpha-Proteobacterial RNA Degradosomes Assemble Liquid-Liquid Phase-Separated RNP Bodies*. *Molecular Cell*, 2018. **71**(6): p. 1027-1039 e14.
68. Arranz, A.T., V. de Crecy-Lagard, and H. de Reuse, *Bacterial RNA Degradosomes: Molecular Machines under Tight Control*. 2019.
69. Li, L., M.B. Charati, and K.L. Kiick, *Elastomeric polypeptide-based biomaterials*. *J Polym Sci A Polym Chem*, 2010. **1**(8): p. 1160-1170.
70. Dzuricky, M., S. Roberts, and A. Chilkoti, *Convergence of Artificial Protein Polymers and Intrinsically Disordered Proteins*. *Biochemistry*, 2018. **57**(17): p. 2405-2414.
71. Quiroz, F.G., et al., *Intrinsically disordered proteins access a range of hysteretic phase separation behaviors*. *Science advances*, 2019. **5**(10): p. eaax5177.
72. Roberts, S., M. Dzuricky, and A. Chilkoti, *Elastin-like polypeptides as models of intrinsically disordered proteins*. *FEBS letters*, 2015. **589**(19): p. 2477-2486.
73. Kadri, S.S., *Key Takeaways From the US CDC's 2019 Antibiotic Resistance Threats Report for Frontline Providers*. *Critical Care Medicine*, 2020.
74. Organization, W.H., *Lack of New Antibiotics Threatens Global Efforts to Contain Drug-Resistant Infections*.
75. *Antibiotic resistance threats in the United States, 2019*. 2019.
76. Shapiro, L. and R. Losick, *Cell Biology of Bacteria: A Subject Collection from Cold Spring Harbor Perspectives in Biology*. 2011: Cold Spring Harbor Laboratory Press.

# MEGAN CLAIRE COHAN

Ph.D. Candidate in the Lab of Rohit V. Pappu, Ph.D.  
Department of Biomedical Engineering  
Washington University in St. Louis  
Saint Louis, MO 63130

Email: mcohan15@gmail.com  
Phone: (401) 575-9757

---

## Professional Profile

---

- Biomedical Engineering Ph.D. Candidate at Washington University in St. Louis utilizing a cross-disciplinary approach to uncover the molecular function of protein disorder in bacterial cell division
- Acting Vice President and project advisor for a student-run nonprofit consulting group
- Current Biotechnology Equity Research Fellow for health care portfolio at investment firm with over \$4B in assets under management
- Incoming Associate at McKinsey & Company (Class of 2021)

---

## Education

---

**Ph.D., Biomedical Engineering** **August 2015 – December 2020**

Cellular and Molecular Bioengineering and Biophysics  
Department of Biomedical Engineering, Washington University in St. Louis  
Thesis Advisor: Rohit V. Pappu, Ph.D.

**B.S., Chemical and Biomolecular Engineering** **September 2011 – May 2015**

Molecular and Cellular Engineering Concentration  
Department of Chemical and Biomolecular Engineering, Johns Hopkins University

---

## Research Experience

---

**Graduate Researcher**, Lab of Rohit V. Pappu, Ph.D. July 2015 – Present

Cellular and Molecular Biophysics & Bioengineering  
Department of Biomedical Engineering, McKelvey School of Engineering  
Washington University in St. Louis, St. Louis, MO  
Project: Uncovering the roles and evolved sequence grammar of hypervariable intrinsically disordered proteins in bacterial cell division

**Graduate Researcher**, Lab of Oliver Daumke, Ph.D. June 2016 – August 2016

Structural Biology of Membrane-Associated Processes  
Max-Delbrück-Centrum für Molekulare Medizin (MDC), Berlin, Germany

Project: Characterizing stabilization mutations of dynein orthologs for crystallography studies

**Undergraduate Research Assistant**, Lab of John Cooper, M.D. Ph.D. June 2014 – August 2014  
Department of Biochemistry and Biophysics, Division of Biology and Biomedical Sciences  
Washington University in St. Louis, St. Louis, MO  
Project: Improving a MATLAB-based cell motility analysis program to quantify neutrophil migration

**Undergraduate Research Assistant**, Lab of Rebecca Schulman, Ph.D. September 2013 – May 2014  
Dynamic and Adaptive Biomolecular Materials  
Department of Chemical and Biomolecular Engineering, Whiting School of Engineering  
Johns Hopkins University, Baltimore, MD  
Project: Programming the controlled nucleation and growth of DNA nanotubes

---

### **Appointments & Awards**

---

*Awardee*, Ceil M. DeGutis Prize in Chemical Biology, Washington University in St. Louis May 2020  
*Vice President*, The Biotechnology and Life Sciences Advising (BALSA) Group March 2020 – Present  
*Biotechnology Equity Research Fellow*, Kennedy Capital Management, Inc. October 2019 – Present  
*Co-Chair*, Intrinsically Disordered Proteins Gordon Research Seminars (postponed) June 2020  
*Director of Human Resources*, The BALSA Group March 2019 – March 2020  
*Travel Award*, 3<sup>rd</sup> Annual Bacterial Cell Biology, Fusion Conferences February 2020  
*Subgroup Representative*, Intrinsically Disordered Proteins, Biophysical Society March 2019  
*Consultant*, The BALSA Group June 2018 – March 2019  
*Travel Award*, Intrinsically Disordered Proteins Gordon Research Conference June 2018  
*Travel Award*, Cellular Mechanisms Driven by Phase Separation, EMBO May 2018  
*Travel Award*, Annual BMES Meeting, Washington University in S October 2017  
*Organizing Committee Member*, Protein Folding Consortium June 2017  
*Graduate Student Representative*, Diversifying Graduate Students Spring 2017  
*Travel Award*, Intrinsically Disordered Proteins Gordon Research Conference June 2016  
*Research Abroad Award*, Washington University in St. Louis Summer 2016  
*Jean & Sydney Grossman Fellowship*, Washington University in St. Louis August 2015  
*1<sup>st</sup> Place Product Design*, Chemical and Biomolecular Engineering, Johns Hopkins University May 2015  
*Instrument Sales Intern*, Agilent Technologies May 2013 – May 2015  
*Dean's List*, Johns Hopkins University December 2014 – May 2015

---

### **Teaching & Mentorship Experience**

---

*Lecturer*, Remote Teamwork, The BALSA Group August 2020  
*Lecturer*, Networking & Pitch Decks, The BALSA Foundation May 2020

<i>Lecturer</i> , Colligative Properties, Bioengineering Thermodynamics (BME320)	November 2019
<i>Project Advisor</i> , The BALSAs Group	August 2019 – Present
<i>Lecturer</i> , Business Communication, The BALSAs Group	July 2019
<i>Project Manager</i> , The BALSAs Group	March 2019 – August 2019
<i>Lecturer</i> , Introduction to Phase Transitions, Bioengineering Thermodynamics (BME320)	September 2016
<i>Teaching Assistant</i> , Bioengineering Thermodynamics (BME320)	September 2016 – May 2018
<i>Mentor</i> , Young Scientists Program	August 2015 – May 2018
<i>Mentor</i> , Bringing Ladies and Scientists Together	September 2014 – May 2015
<i>Athlete Mentor</i> , Johns Hopkins University Mentorship Program (JUMP)	August 2011 – May 2014

---

## Publications

---

### Published

- Cohan, M.C.**, Pappu, R.V. (2020). Making the Case for Disordered Proteins and Biomolecular Condensates in Bacteria. *Trends in Biochemical Sciences*, **45**: 668 - 680
- Cohan, M.C.**, Eddelbuettel, A.M.P., Levin P.A., Pappu R.V. (2020). Dissecting the functional contributions of the intrinsically disordered C-terminal tail of *B. subtilis* FtsZ. *Journal of Molecular Biology*, **432**: 3205-3221
- Cohan, M.C.**, Ruff, K.M., Pappu, R.V. (2019). Information theoretic measures for quantifying sequence–ensemble relationships of intrinsically disordered proteins, *Protein Engineering, Design and Selection*, **32**: 191–202.
- Mittal, A., Holehouse, A. S., **Cohan, M. C.**, Pappu, R. V. (2018). Sequence-to-conformation relationships of disordered regions tethered to folded domains of proteins. *Journal of Molecular Biology*, **430**: 2403-2421.
- Bridgen, D. T., Fearing, B. V., Jing, L., Sanchez-Adams, J., **Cohan, M. C.**, Guilak, F., Chen J., Setton, L. A. (2017). Regulation of human nucleus pulposus cells by peptide-coupled substrates. *Acta Biomaterialia*, **55**: 100-108.
- Mukherjee, S., Kim, J., Mooren, O. L., Shahan, S. T., **Cohan, M.**, & Cooper, J. A. (2015). Role of cortactin homolog HS1 in transendothelial migration of natural killer cells. *PloS One*, **10**: e0118153.

### In preparation

- Cohan, M.C.**, Bullock, J, Ruff, K.M., Levin P.A., Pappu, R.V. Patterning of oppositely charged residues within the intrinsically disordered tail of FtsZ influences *in vitro* assembly and cell division in *B. subtilis*. In preparation for *Molecular Cell* (2020). Published on bioRxiv (2018)
- Cohan, M.C.**, Pappu, R.V. Decoding the conserved sequence features within a family of hypervariable sequences.

---

## Invited Research Talks & Poster Presentations

---

### Invited Talks

Biochemistry and Molecular Biophysics Presentation of the Ceil M. DeGuitis Prize – virtual via Washington University in St. Louis (May 2020). “Small but mighty! Intrinsically disordered domains as key regulators of bacterial processes.”

Intrinsically Disordered Proteins Scientific Interest Group – virtual via Stanford University (April 2020). “Decoding the molecular grammar of the intrinsically disordered C-terminal tail of FtsZ.”

Center for Science & Engineering of Living Systems – Washington University in St. Louis (March 2020). Canceled due to COVID-19.

Biophysical Society 63rd Annual Meeting – Baltimore, MD (March 2019). “Measures for quantifying sequence-ensemble relationships of intrinsically disordered regions.”

Washington University in St. Louis Plant and Microbial Biosciences Annual Retreat – St. Louis Zoo (October 2018). “Impact of conserved sequence parameters of the disordered tail of FtsZ on function & phenotype”

IDP Gordon Research Symposium – Les Diablerets, Switzerland (June 2018). “Impact of conserved sequence features within the disordered C-terminal Tail of FtsZ on protein function and cellular phenotype.”

Protein Folding Consortium Annual Meeting – Berkeley, California (June 2017). “Impact of conserved charge patterning within disordered domains on protein function and cellular phenotype”

### Poster Presentations

EMBO Bacterial Cell Division – Lund, Sweden (June 2019). Decoding the molecular grammar of the intrinsically disordered divisome in *B. subtilis*.

Biophysical Society 63<sup>rd</sup> Annual Meeting – Baltimore, MD (February 2019). Measures for quantifying sequence-ensemble relationships of intrinsically disordered regions.

ASCB & EMBO Annual Meeting – San Diego, CA (December 2018). The disordered C-terminal tail of FtsZ modulates the phase boundaries of FtsZ assembly and regulates bacterial cell division.

Gordon Research Conference: IDPs – Les Diablerets, Switzerland (July 2018). Measures for quantifying the evolution of sequence-to-conformation relationships of intrinsically disordered regions.

EMBO Cellular Mechanisms Driven by Phase Separation– Heidelberg, Germany (May 2018). Charge patterning within the C-terminal disordered tail modulates the phase boundaries of FtsZ assembly and regulates cell division in rod-shaped bacteria

Biophysical Society 62<sup>nd</sup> Annual Meeting – San Francisco, California (February 2018). Sequence-encoded charge patterning of the intrinsically disordered tail of FtsZ impacts polymerization and bacterial cell division

ASCB & EMBO Annual Meeting – Philadelphia, PA (December 2017). Sequence grammar of the C-terminal disordered tail determines FtsZ self-assembly and cell division in rod-shaped bacteria

CBSE Advances Symposium – Saint Louis, MO (November 2017). Sequence grammar of the C-terminal disordered tail determines FtsZ self-assembly and cell division in rod-shaped bacteria

BMES 49<sup>th</sup> Annual Meeting – Phoenix, Arizona (October 2017). The intrinsically disordered tail of FtsZ impacts polymerization and bacterial cell division through sequence-encoded charge patterning.

Biophysical Society 61st Annual Meeting – New Orleans, LA (February 2017). The intrinsically disordered tail of FtsZ impacts polymerization and bacterial cell division through sequence-encoded charge patterning.

Gordon Research Conference: IDPs – Les Diabrelets, Switzerland (June 2016). Impact of conserved charge patterning in the intrinsically disordered tail of FtsZ protofilament assembly and cell division in rod shaped bacteria.

Biophysical Society 60th Annual Meeting – Los Angeles, CA (February 2016). Impact of conserved charge patterning in the intrinsically disordered tail of FtsZ protofilament assembly and cell division in rod shaped bacteria.

Johns Hopkins University ChemBE Senior Design – Baltimore, MD (April 2015). Nic-Balm: A Smoking Cessation Device

---

## Certifications

---

**Financial Modeling, Training the Street** Fall 2020  
Olin School of Business, Washington University in St. Louis

**Entrepreneurship for Bioscientists (E4B)** Fall 2019  
Washington University in St. Louis

**Business Concepts for Life Scientists** Summer 2019  
The Career Center, Washington University in St. Louis

---

## Professional Societies

---

Biophysical Society & Intrinsically Disordered Proteins Subgroup 2015 – Present  
BMES 2017-2018  
American Society Cell Biology 2017-2019

---

## Languages

---

**Foreign:** German (basic); **Native:** English

**Computational:** MATLAB (advanced), Python (intermediate)