AN ABSTRACT OF THE DISSERTATION OF

Sarah A. Clark for the degree of Doctor of Philosophy in Molecular and Cellular Biology presented on February 13, 2018.

Title: Three Sites are Better than One: Exploring Multivalency in LC8 Binding Partners.

Abstract approved: _____
Elisar J. Barbar

Hub proteins bind a large number of partners to facilitate structural changes and downstream protein interactions. LC8, a highly conserved protein homodimer, is a unique hub that regulates the activity of proteins in a wide range of cellular processes by binding to intrinsically disordered regions. With many of these systems, LC8 is known to function as a dimerization engine, bringing together two disordered chains to facilitate structural change. But for the significant and increasingly large class of multivalent LC8 binding partners, the role of LC8 binding is less clear.

This thesis reports on work aimed at analyzing the structure and function of complexes formed between LC8 and three multivalent protein partners: ASCIZ, Chica, and Nucleoporin 159. Four chapters of original work include one review, two primary research reports, and one unpublished study. The review (chapter 2) classifies different types of intrinsically disordered multivalent protein assemblies and highlights the unique features of complexes formed between LC8 and its multivalent protein partners. Chapter 3 is an in-depth study of the factors that influence partner protein binding to LC8 using the multivalent protein Chica as a model system. The 'anchored flexibility model' of LC8 motif recognition is introduced, which proposes that a few residues are essential for binding, while others modulate binding affinity. This model and the structural data presented in this work enhance our understanding of how LC8 recognizes and binds to its protein partners. Chapter 4 focuses on the structural and functional characterization of the protein ASCIZ, a transcription factor for LC8. This work demonstrates how the concerted action of multiple binding sites in an intrinsically disordered region enable a gradient of ASCIZ activity and finely tune the transcriptional level of LC8. Chapter 5 is a biophysical examination of the structure and assembly process of the complex formed between Nucleoporin 159 and LC8.

Finally, chapter 6 discusses the impacts and highlights of the reported work, and compares the properties of the multivalent protein complexes presented in chapters 3-5. Individually, these results provide a detailed biophysical description of three unique macromolecular protein assemblies. Together, they emphasize the structural and functional flexibility of LC8 as hub protein.

Three Sites are Better than One: Exploring Multivalency in LC8 Binding Partners

by
Sarah A. Clark

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented February 13, 2018
Commencement June 2018

Doctor of Philosophy dissertation of Sarah A. Clark presented on February 13, 2018.

APPROVED:

Major Professor, representing Molecular and Cellular Biology

Director of the Molecular and Cellular Biology Program

Dean of the Graduate School

I understand that my dissertation will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my dissertation to any reader upon request.

Sarah A. Clark, Author

ACKNOWLEDGEMENTS

First, I would like to thank my mentor, Dr. Elisar Barbar. Elisar, thank you for granting me the freedom and independence to pursue my own ideas in the lab. Even when my experiments failed (*Bim and Beclin…*) you never micromanaged me and continually encouraged the development of new projects. Your confidence in me made a huge difference and allowed me to grow as a scientist. I feel very lucky to have had an advisor who encouraged me to attend conferences, travel to Australia, establish connections, and pursue my own interests.

I would like to thank the Biochemistry and Biophysics Department for welcoming me wholeheartedly even though I'm a Molecular and Cellular Biology student. The effort to include me in everything from the very beginning was so important to me. I thank my graduate committee members, especially Dr. Barbara Taylor and Dr. Andy Karplus. Barb, thank you for all of your help and guidance over the years and for teaching me how to be a good instructor. Andy, thank you for your enthusiasm, advice, and scientific insights. Your guidance has made me a far better writer, presenter, and scientist. Also, I thank my graduate mentors, Dr. Afua Nyarko and Dr. Jing Jie, for their infinite patience, amazing organizational skills, and willingness to share their knowledge.

I thank my collaborators - Dr. Jörg Heirhorst, Dr. Radovan Fiala, Dr. Grant Pierce, and Dr. Steve Reichow - for their help and contributions to my thesis project. Jörg, thank you for welcoming me into your lab and teaching a biophysics student how to perform cell culture work. Radovan, thank you for your thorough, patient email replies to all of my questions, and for the hours and hours of time that you devoted to NMR experiments on my very challenging protein. Grant, thank you for driving many hours to Melbourne to teach me how to run analytical ultracentrifugation experiments. Your kindness left an impression on me. Steve, thank you for showing me how perseverance can pay off and for teaching me how to perform electron microscopy experiments.

I have made some terrific friends at OSU - Nathan Jespersen, Dr. Andrew Brereton, Andrew Popchock, and Michelle Wiley, to name a few. Thank you all for making this experience so positive and full of fun moments. Nathan, my partner in crime, thank you for always listening and being willing to bat around new ideas. Your joyfulness and genuine enthusiasm is infectious, and made the lab such an entertaining space.

Finally, I would like to thank my family, especially my mom, dad, and Aunt Kathy, for their encouragement and support. I have you all to thank for where I am today. Ryan, my other

half, thank you for your endless support on everything and for helping me keep things in perspective.

CONTRIBUTION OF AUTHORS

Nathan Jespersen contributed to the introduction and chapter 2. Afua Nyarko contributed to experiments and text in chapter 3. Frank Löhr contributed to NMR experiments in chapter 3. P. Andrew Karplus contributed to crystallographic analysis and text in chapter 3. Steve Reichow and Janette Myers performed electron microscopy experiments in chapter 4. Radovan Fiala and Jirka Novacek contributed to NMR experiments in chapter 4. Grant Pearce performed analytical ultracentrifugation experiments in chapters 4 and 5. Ashleigh King and Jörg Heierhorst contributed to cellular transcription assays in chapter 4. Elisar Barbar was involved in the design, writing, and analysis of all experiments and chapters.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF FIGURES

Figure Page

LIST OF FIGURES (continued)

Figure Page

LIST OF TABLES

Table Page

Three Sites are Better Than One: Exploring Multivalency in LC8 Binding Partners

Chapter 1

Introduction

Dynein light chain 8

Overview

Dynein light chain (Dynll1, also known as LC8) is a ubiquitously expressed, essential hub protein with nearly 60 confirmed binding partners. LC8 was initially described as the seventh light chain component of the outer arm axonemal dynein in *Chlamydomonas reinhardtii* in 1982[1], and was later identified as a subunit of the cytoplasmic motor proteins dynein 1 and 2[2, 3] (Figure 1.1, left). In the following years, LC8 was found to interact with many proteins that are unrelated to dynein or microtubule-based transport, leading researchers to conclude that LC8 acts as a cargo adaptor protein for dynein[4, 5]. This hypothesis was discredited by the discovery that a large portion of cytoplasmic LC8 is not associated with dynein[3] and the observation that LC8 binds putative cargo and dynein at the same site[6, 7]. Crystal structures of LC8 bound to peptides clearly illustrate that LC8's binding groove can only accommodate one peptide at a time[8-10] (Figure 1.1, right). As LC8 is a homodimer with two binding grooves, it binds two chains simultaneously, promoting dimerization of the partner protein [6, 7, 11]. The hub hypothesis, proposed in 2008[12], describes LC8 as a hub protein that promotes dimerization and folding of partially disordered proteins in a diverse array of cellular processes. The wide range of partners suggests a role for LC8 as a general biochemical regulator akin to the well-described regulatory proteins 14-3-3[13], Hsp90[14], and calmodulin[15]. However, unlike the aforementioned proteins, LC8 acts as both a regulatory chaperone and a scaffolding protein.

LC8 exists as a 20.6 kDa homodimer in solution (10.3 kDa monomer) with a dissociation constant of 83 nM at neutral pH[16, 17]. LC8 is only found as a monomer in highly acidic conditions below pH 4.8[16] or when phosphorylated at Ser88[18]. There are two dynein light chain isoforms: Dynll1 (LC8) and Dynll2. This text refers to dynein light chain as LC8 because it is the predominant isoform and has been used in the majority of functional studies. LC8 and Dynll2 are 93% identical at an amino acid level, differing by only 6 out of 89 amino acids (Figure 1.2a). While the two isoforms are expressed at different levels depending on the tissue type, it is difficult to draw conclusions as to where LC8 and Dynll2 preferentially localize due to discrepancies in mRNA and protein distribution from different studies [19, 20]. The authors of one study noted that these differences might be due to LC8's role as a hub protein[19] and therefore reflect a difference in LC8 partner protein distribution.

From an evolutionary perspective, LC8 can be divided into three phylogenetic groups: animals, Schistomsoma parasites, and fungi and plants[19]. Mammals, fruit flies, and zebrafish contain both light chain paralogs, while all other species only contain the gene for one light chain

protein, which exhibits the greatest sequence similarity to Dynll2[19]. Mammalian LC8 sequences are identical, and human, *Xenopus*, and *Drosophila* sequences are ~94% identical. Human LC8 is 81% identical to the unicellular protozoan parasite *Trypanosoma brucei* and, interestingly, was recently characterized as a component of the parasite's basal transcription machinery[21, 22]. Importantly, although the plant kingdom lacks a dynein motor protein[23], LC8 is found in plants with 40% sequence identity to the human protein, emphasizing its broader role as a hub protein. In the fava bean plant (*Vicia faba*), LC8 interacts with phototropin, a protein that regulates signaling in response to blue light[24]. However, very little is known about the role of LC8 in plants or its plant-specific binding partners.

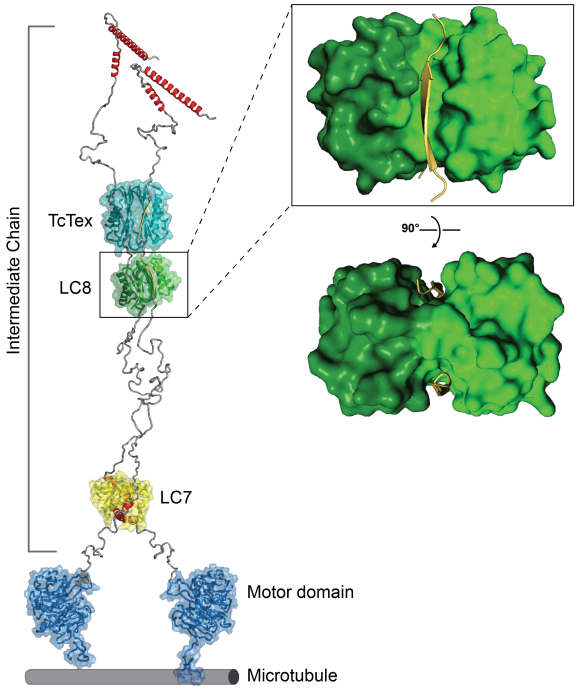


Figure 1.1: Dynein light chain 8.
LC8 (green) binds to two copies of the dynein intermediate chain (gray), which aids in assembly of the dynein motor protein. As the LC8 binding groove can only accommodate one ligand at a time (box), LC8 cannot serve as a cargo adaptor protein when bound to the intermediate chain.

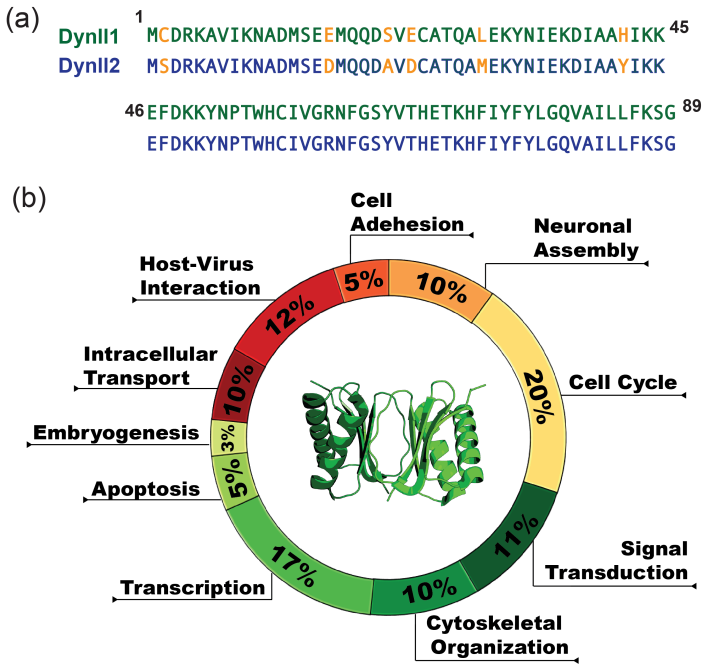


Figure 1.2. LC8 interacts with binding partners from a wide range of cellular processes. (a) An amino acid sequence alignment of the two isoforms of LC8, Dynll1 and Dynll2, show that they differ by only six amino acids (orange). (b) The 58 known LC8 binding partners are sorted into ten categories based on their biological process listed in Uniprot[25]: cell adhesion, neuronal assembly, cell cycle, signal transduction, cytoskeletal organization, transcription, apoptosis, embryogenesis, intracellular transport, and host-virus interaction. Proteins that perform multiple functions are placed in more than one category.

The LC8 Recognition Motif

LC8 has 58 confirmed partner proteins from a diverse selection of cellular pathways (Figure 1.2b). The largest functional category is cell cycle regulation, which contains 20% of the known partner proteins. Remaining binding partners are dispersed throughout nine additional categories that were assigned based on their 'biological process' listed in Uniprot[25]. Multiple methods have been used to identify new LC8 binding partners including phage display[26], proteomic analysis[27], and pepscan[28]. These proteins share a short linear motif that mediates their binding to LC8[29]. The 10 amino acid motif typically includes the residues TQT at positions 7-9[26, 29] and is found in segments of the protein partners that are intrinsically disordered[12]. Analysis of 92 recognition motifs from all 58 currently known partner proteins (some of which contain multiple recognition motifs), demonstrates that the residues TQT are indeed the most common amino acids at these

positions, while all other motif positions are highly variable (Figure 1.3a). This observation is the basis of the '*anchored flexibility model*' of LC8 motif recognition, which states that the TQT residues are "essential" for motif binding, while the other 7 motif residues modulate affinity[30]. The range of natural motif affinity for LC8 ranges from 0.16 μM [31] to 42.7 μM[32], as determined by binding of 24 synthetic peptides to LC8 (Figure 1.3b). A phage display study by Rapali et al. (2011) identified the ideal consensus motif for a tight binder as VSRGTQTE and confirmed this result by synthesizing artificial peptides containing variations of this motif and measuring their binding to LC8[26] (Figure 1.3b, top).

The prevalence of the TQT residues gave rise to the moniker "TQT motif" for the LC8 recognition sequence. However, other variations exist, such as TMT in the motor protein Myosin 5A[33] and TQC in the spindle assembly protein anastral spindle 2[34]. While these variations bind to LC8 because the side chains of methionine or cysteine fit in the binding groove, their interactions are not optimal and result in weaker binding. A recent study analyzed LC8 binding to 26 putative TQT motif peptides that were identified by phage display experiments and found that the TQT residues are necessary, but not sufficient, for LC8 binding. The authors examine the molecular determinants of LC8 binding using information from all known LC8 binding motifs and conclude that predicting binding partners based on motif sequence and structure is difficult due to the remarkable plasticity of the LC8 binding groove (Jespersen and Barbar, unpublished data).

Investigation of the TQT motif from two well-studied LC8 partner proteins, the intermediate chain of dynein (IC)[8] and brain nitric oxide synthase (nNos)[35], demonstrates that the motif sequence is conserved between species and further emphasizes which residues are important for binding (Figure 1.3c,d). IC is very highly conserved in all eukaryotes, with only minimal amino acid changes from humans to yeast. nNos is less conserved, but the importance of residue positions 3, 4, 7-10 is evident.

Additionally, accumulating evidence suggests that motif sequence is not the only determinant of binding affinity. Partner protein oligomerization state[26], the presence of transient secondary structure at the motif location[36], and cooperativity between multiple motifs[36] also affect binding affinity.

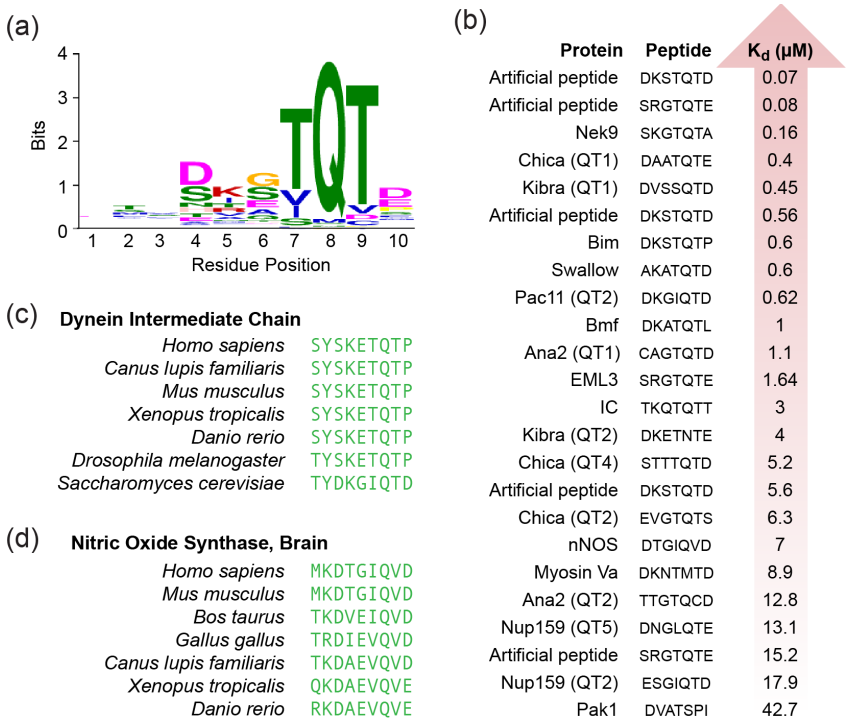


Figure 1.3. The LC8 recognition motif.
(a) A sequence logo of the LC8 binding motif derived from sequences of 92 motifs, showing the prevalence of TQT residues at positions 7-9. Height of amino acids indicates their relative frequency at that position. (b) The sequence and dissociation constant of LC8 binding to 24 synthetic peptides are shown. Peptides are ranked by their K_d value. Sequence alignment of the LC8 binding motif from (c) dynein intermediate chain and (d) nitric oxide synthase demonstrates a high degree of sequence conservation.

Structural Basis of Binding

LC8 interacts with its binding partners via a highly conserved hydrophobic groove at the dimer interface[8, 9] (Figure 1.4a). Sequence-based alignment of LC8 from 58 species ranging from human to the parasitic worm *Echinococcus granulosus* clearly indicates that the dimer interface and binding groove are the best conserved portions (Figure 1.4b, magenta), while the exterior of the protein is less conserved (cyan, white). This result implies that both the binding interaction and the recognition motif are evolutionarily conserved.

Two grooves, one on each side of the dimer interface, allow LC8 to bind two ligands simultaneously and dimerize its partner protein. The two grooves bind cooperatively; when the first ligand binds, affinity for the second ligand is enhanced, increasing the likelihood that a

second ligand will bind[37]. As LC8 dimerizes all of its known partner proteins, it is proposed to serve as a cellular 'dimerization engine' that alters the oligomeric state of its partner proteins in order to modify their function[12]. This idea is discussed further in the next section.

Thirteen structures of human or *Drosophila* LC8 bound to peptides have been solved[5, 8, 9, 26, 30-34, 37-40], providing atomic-level insight into LC8-ligand interactions and explaining why the TQT residues are so highly conserved. The LC8 dimer interface is composed of a 12-strand β-sandwich, where five β strands are contributed by each LC8 monomer (Figure 1.4a, β1-β5, green) and the recognition motif is the sixth (yellow). As the binding groove is formed by the dimer interface, monomeric LC8 cannot interact with binding partners[18]. The recognition motif is always in an intrinsically disordered region of the partner protein and folds into a β-strand once bound to LC8. For this reason, motifs that have high helical propensity have lower affinities[36], while those with polyproline II or beta-strand propensity tend to bind more tightly[36, 41]. A network of hydrogen bonds between side chain and backbone atoms anchors the ligand in the binding groove (Figure 1.4c). The importance of the TQT residues is clear from their hydrogen bonding interactions; the Q is involved in interactions with both LC8 monomers, while both T's are completely buried within the binding pocket. The remainder of the motif residues participate in hydrogen bonding interactions with various β3 residues which vary depending on motif sequence.

LC8 is able to accommodate a wide variety of peptide sequences in the binding groove due to its structural plasticity. Peptide binding causes a shift in tertiary structure that exposes additional hydrophobic surface for binding and results in ordering of LC8 around the binding groove[37]. An in-depth ensemble analysis of all apo-LC8 and peptide bound-LC8 structures in the protein data bank illustrates this structural change. The binding groove is very dynamic in the apo-LC8 structure (Figure 1.4d, β3), but considerably more ordered in the bound protein (Figure 1.4e, β3). This demonstrates that peptide binding orders the binding groove, locking it into a single conformation.

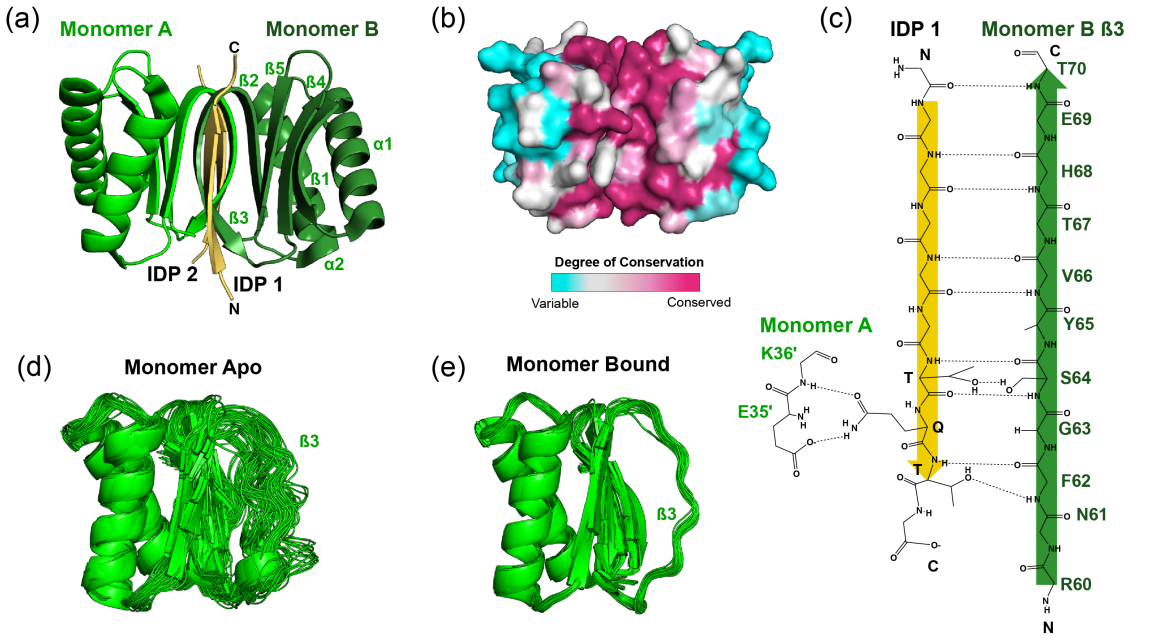


Figure 1.4. Structural basis of ligand binding to LC8.
(a) A ribbon diagram of the LC8 dimer (green and dark green) bound to two chains of the Chica QT1 peptide (yellow, PDB 5E0M). Structures of all LC8/peptide complexes look essentially the same. LC8 secondary structure elements for one monomer are labeled. (b) The degree of amino acid conservation of LC8 homologs is mapped onto the ribbon structure of LC8 (color-coded gradient from blue-variable to magenta-conserved). Sequences of LC8 were analyzed with ConSurf, a bioinformatics tool for estimating the evolutionary conservation of amino acid positions in proteins. (c) Hydrogen bonding interactions for backbone residues and conserved TQT residues are shown for a peptide (yellow) bound to LC8 (green). These interactions are very highly conserved regardless of the sequence of non-TQT residues. (d) 93 apo-LC8 structures and (e) 61 ligand bound-LC8 structures are overlaid using the program the Ensemblator[42]. Structures are derived from the PDB and some are NMR structures that contain multiple models. Only one LC8 monomer is shown and the region with the highest variability, β3, is labeled.

Structure of LC8:Partner Complexes

LC8 is referred to as a cellular dimerization engine because it alters the function of its binding partners by changing their oligomeric state. LC8 dimerizes all of its known partner proteins and often aids in the formation of **(1)** a coiled-coil or **(2)** a self-association domain distal from the binding site[12]. Dimerization of the partner protein by LC8 can also **(3)** promote binding of other proteins[43]. These three types of LC8:partner protein complexes are discussed in greater detail below.

(1) Coiled-coil formation. LC8 binding to the *Drosophila* protein Swallow results in the formation of a coiled-coil 17 residues from the TQT motif[6] (Figure 1.5a). The coiled-coil is weak as a monomer in the absence of LC8, but about 7-fold stronger when it is dimerized as a result of LC8 binding[44]. Swallow binding to LC8 is necessary for bicoid RNA localization during *Drosophila* oogenesis[45]; however, it is unknown why Swallow must be dimerized to perform this function. LC8 similarly dimerizes the centriole duplication factor anastral spindle 2 (Ana2) by binding at two sites, one on either side of a coiled-coil. The coiled-coil then oligomerizes to form a tetramer, which likely increases Ana2 avidity for centriole binding factors[34]. The interaction between Ana2 and LC8 is necessary for localization and stabilization of proteins involved in regulating mitotic spindle assembly[46].

(2) Self-association. In the case of dynein intermediate chain (IC), LC8 binding results in the self-association of two alpha helices 69 amino acids away from the TQT motif[47] (Figure 1.5b). This alpha helical self-association domain gains helicity upon LC8 binding[47] and increases the affinity of LC8 for IC 6-fold[48]. Interestingly, the long linker separating the TQT motif and the self-association site remains disordered after LC8 binding[47], allowing the intermediate chain to retain the flexibility that is likely needed for binding of other regulatory proteins[49].

(3) Protein partner binding. A third consequence of LC8 dimerization is the binding of additional proteins to the bivalent complex. In dynein, LC8 dimerization of the intermediate chain results in 50-fold binding enhancement of a second light chain, Tctex[43] (Figure 1.5c, top), and binding enhancement of yeast dynactin (Nip100) [50]. The second binding event occurs with a higher affinity after a bivalent scaffold is formed because the entropic cost associated with linking two disordered chains is reduced. Additionally, the complex formed by Bim and LC8 is capable of binding to the autophagic regulatory protein Beclin-1. The three proteins form a functional macromolecular complex *in vivo*, but only when LC8 is complexed with Bim[51]. LC8 binding to Bim inhibits the induction of apoptosis[52] and the tertiary Bim/LC8/Beclin-1 complex inhibits autophagy[51]. Interaction with LC8 is essential for both of these functions.

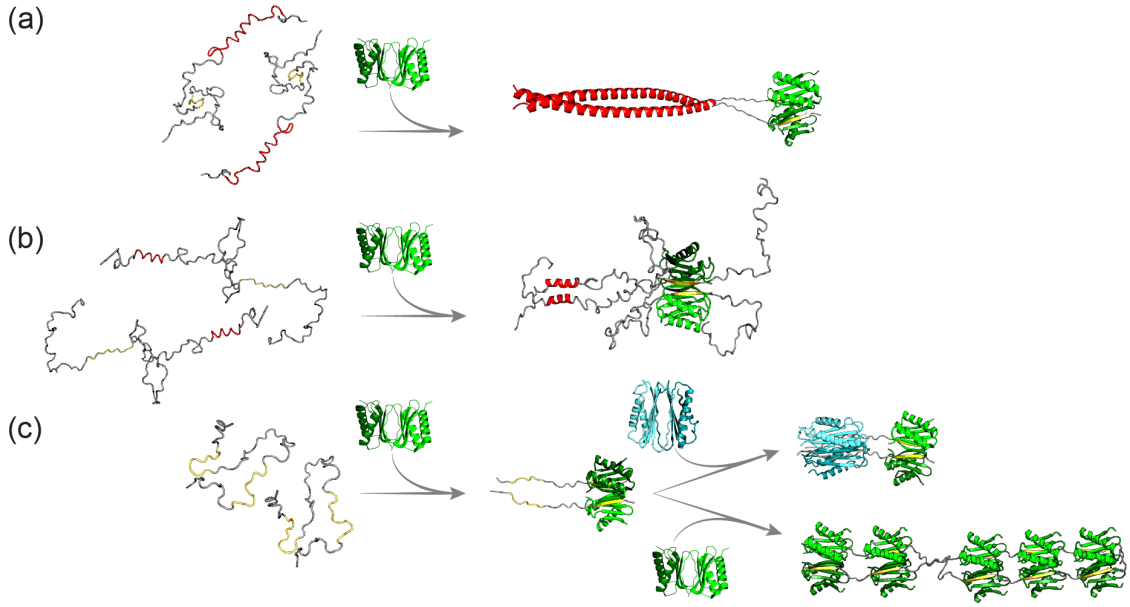


Figure 1.5. Three types of LC8: partner protein assemblies.
Structural models depicting the three main types of complexes formed by LC8 and its protein partners are shown. LC8 (green) binds to two disordered chains, resulting in (a) the formation of a coiled-coil, (b) formation of a self-association domain, or (c) an additional binding event. Binding of an additional protein partner is shown for dynein light chain TcTex[37] (blue) and the multivalent LC8 binding partner Nucleoporin 159[36].

Regulation of LC8

Although LC8 plays a significant role in dozens of cellular processes, little is known about how LC8 itself is regulated. Both phosphorylation and pH have been proposed as potential regulators of LC8 activity by inducing the inactive monomeric state[16, 18]. Phosphorylation of serine 88, located proximal to the binding groove, shifts the LC8 monomer-dimer equilibrium towards a monomeric state[18]. Ser88 phosphorylation occurs *in vivo* by Pak1 (p21-activated kinase-1) and promotes the tumorigenic potential of breast cancer cells[53], presumably by inhibiting LC8 from interacting with its binding partners.

Additionally, LC8 redox state alters its interaction with IκBα and thus the activation of NF-κB, a transcription factor that regulates immune and inflammatory responses[54]. Reactive oxygen species oxidize a pair of cystines in the LC8 binding groove, causing them to form a disulfide bond that prevent LC8 from binding IκBα. The downstream consequence of this redox event is activation of NF-κB, which is normally maintained in a repressed state[54].

Alternatively, some partner proteins are post-translationally modified to alter their interaction with LC8. Bim and Bmf are proapoptitic proteins whose association with LC8 is mediated by phosphorylation of the recognition motif. The c-Jun NH2-terminal kinase (JNK) phosphorylates Bim and Bmf, preventing them from binding to LC8, and causing induction of apoptosis[55]. The protein kinase Nek9 similarly autophosphorylates a serine near the recognition motif, reducing LC8 binding, and activation of Nek6. Activated Nek6 can then initiate a signaling cascade that leads to mitotic spindle formation[31]. While all of these processes play a role in regulating LC8 activity, none suggest a general mechanism that could govern LC8 binding or selection of protein partners.

A third, recently discovered method for regulating the cellular level of LC8 is negative autoregulation by its transcription factor, ASCIZ (ATM-substrate Chk-Interacting Zn^{2+} finger). Human ASCIZ contains 11 TQT motifs in an intrinsically disordered C-terminal domain. Cell-based studies suggest that these multiple sites function as a sensor for cellular LC8 concentrations; high concentrations of LC8 turn transcription off, while low LC8 levels increase transcription[56]. This 'sensor hypothesis' of LC8 regulation could explain how the cellular concentration of LC8 is tuned to match the needs of the cell.

Regulation of Cellular Processes

LC8 is essential for a wide variety of cellular processes from cell cycle regulation to viral infection. When LC8 levels are indirectly reduced via ASCIZ knockdown, various developmental and mitotic detects are observed. Mice with mutations in ASCIZ that prevent LC8 transcription die in late embryogenesis and exhibit serious developmental defects in the kidneys and lungs[57-59]. *Drosophila* ASCIZ knockouts die in early embryogenesis and localized wing knockdowns using RNAi show mitotic defects[60]. Many *in vivo*[61] and cell culture[62, 63] studies have been performed in which the LC8 binding site on a partner protein is mutated, impairing its ability to interact with LC8. In every instance, the impaired TQT motif has a profound impact on the system. While it is clear that LC8 is essential for these cellular processes, its specific role in the process often remains elusive.

LC8 binding partners can be grouped into 10 broad categories based on their main cellular function: neuronal assembly, cell cycle, signal transduction, cytoskeletal organization, transcription, apoptosis, embryogenesis, intracellular transport, host-virus interaction, and cell adhesion (Figure 1.2b). The most populated category is cell cycle regulation, in which LC8 interacts with 12 proteins, most of which are involved in mitotic spindle organization. In the other

9 categories, LC8 mainly alters the subcellular localization of its targets. The neuronal assembly protein Bassoon, for example, relies on LC8 for proper trafficking in neurons and correct distribution among synapses[64]. Similarly, the Ras-like small GTPase exchange factor, RasGRP3, displays altered subcellular location when it cannot bind to LC8, which affects diacylglycerol signaling[65]. A notable third category of LC8 function is structural support, for which cytoplasmic dynein-1 is the best-studied example. LC8, together with two other light chains, dimerizes the intermediate chain of dynein, creating a stable assembly and indirectly facilitating cargo binding[7]. LC8 also binds to the yeast nuclear pore protein nucleoporin 159 and stabilizes the nuclear pore complex[38].

Multivalent Partner Proteins

A feature of some LC8 partner proteins is the prevalence of multiple TQT motifs in one protein, often located within the same domain (Figure 1.6). Approximately 21% of known LC8 binding partners contain more than one functional TQT motif and thus simultaneously bind multiple LC8 dimers[4, 28, 34, 38, 56, 62, 64, 66, 67]. In comparison to monovalent partners, multivalent proteins have not been well studied *in vitro*, in part due to the distance between LC8 motifs for some partners (Figure 1.6, ex: Bassoon), necessitating the purification of a very large intrinsically disordered protein. Only two multivalent proteins have been characterized *in vitro*: Nucleoporin 159 (Nup159) and ASCIZ.

Nup159 is an essential component of the nuclear pore complex in yeast. The structure and function of the Nup159:LC8 complex has been well-studied both *in vitro* and *in vivo*. Nup159 contains 5 tandem TQT motifs in an intrinsically disordered region that bind cooperatively to LC8[36], forming a rigid, ladder-like structure that aids in nuclear pore formation[68].

ASCIZ, by comparison, is markedly less understood. *In vivo* studies demonstrate that ASCIZ is a transcription factor for LC8[56, 58] and a 2011 study identified 11 LC8 binding motifs in ASCIZ's C-terminal domain using a pepscan experiment[69]. ASCIZ is proposed to function as a sensor for the cellular concentration of LC8 on the basis of cell culture studies, which show that when LC8 levels are high, the TQT motifs on ASCIZ are occupied and transcription is turned off[56]. When LC8 levels are low or the TQT sites are mutated to prevent binding, transcription is very high[56]. However, the structural and functional role of 11 binding sites, the most for any known LC8 partner protein, is unclear. Further examination of ASCIZ and additional multivalent partners would provide valuable insight into the function LC8:multivalent protein complexes.

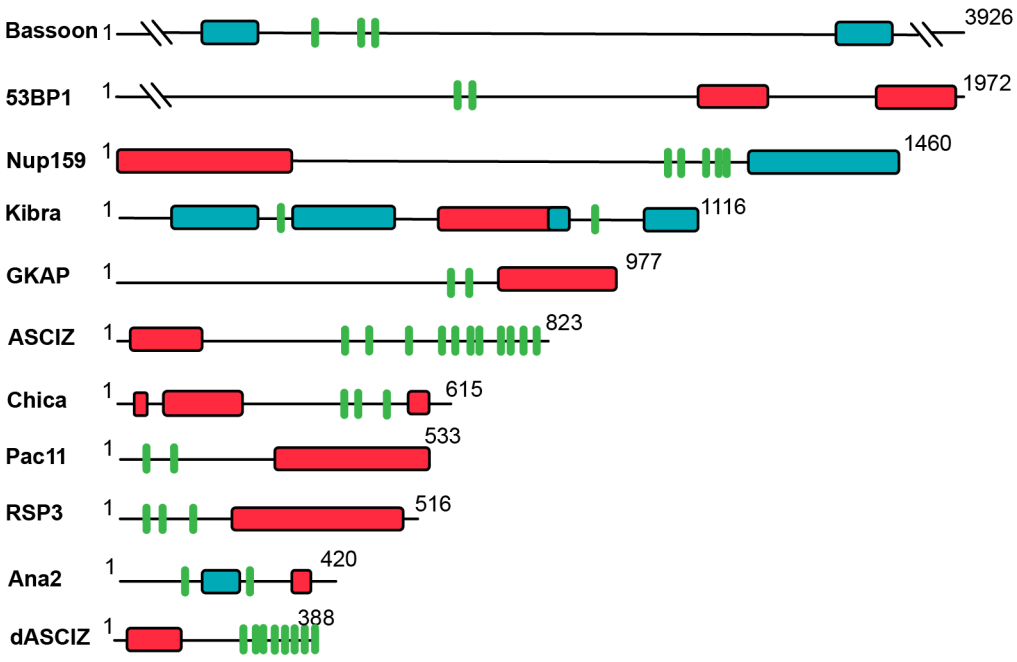


Figure 1.6. LC8 binding partners have multiple recognition motifs in intrinsically disordered regions.

Sequence-based predictions of order (red boxes), disorder (black lines), coiled-coil (blue boxes), and LC8 binding motifs (green bars) are shown for residues 1000−2500 of Bassoon, a protein involved in neuronal trafficking[64]; residues 500−1972 of p53 binding protein 1 (53BP1), a protein which activates p53-dependent gene transcription[4]; yeast nucleoporin, Nup159, a protein in the nuclear pore complex[36]; Kibra, a protein involved in tissue homeostasis and regulation of organ size[66]; guanylate kinase-associated protein (GKAP), a protein that orchestrates protein remodeling at synapses[28]; DNA damage repair protein and transcription factor ATM-Substrate Chk-interacting zinc finger (ASCIZ)[69]; the intermediate chain (Pac11) subunit of the yeast cytoplasmic dynein molecular motor[43]; Chica, a spindle-associated adaptor protein[62]; the flagellar radial spoke protein 3, RSP3[67]; anastral spindle-2 (Ana2), a protein involved in centriole duplication[34]; and the *Drosophila* homologue of ASCIZ (dASCIZ)[60]. Sequence predictions of order and disorder were obtained using the program PSIPRED[70], where our criteria for structure was based on >10% probability of predicted structure in a 50+ amino acid stretch, and coiled-coils were predicted using the program Paracoil2[71], where predicted coiled-coils with a p-value <0.025 were considered significant. Figure adapted from Clark et al. Protein Sci. (2015)[30].

Biophysical Techniques

To probe the interactions between LC8 and its multivalent binding partners, I have isolated and characterized many protein constructs using a variety of biochemical and biophysical techniques. A total of 23 different constructs were used in this work, the majority of which were designed by me. Proteins were expressed in *Escherichia coli* host cell lines and purified using either a 6-His tag or a 6-His tag in conjunction with an MBP-tag or Z-tag for enhanced solubility. TALON affinity resin was used for the first purification step from cell lysate, and was commonly followed by ion exchange chromatography and size exclusion chromatography to increase protein purity.

The three main biophysical techniques used in these studies are: nuclear magnetic resonance spectroscopy (NMR), analytical ultra centrifugation (AUC), and isothermal titration calorimetry (ITC). These techniques provide highly complementary information on protein structure, dynamics, size, and thermodynamic parameters that together paint a picture of a protein or macromolecular complex.

NMR Spectroscopy

NMR exploits the magnetic properties of atomic nuclei to investigate the structure and dynamics of molecules in solution. An atom's chemical environment alters its presentation in NMR spectra, allowing one to examine protein structure and protein-protein interactions. For protein residue assignments, I utilized conventional 3D experiments as well as 5D experiments for proteins that were challenging to assign by 3D methods. Protein residue assignment enables calculation of secondary structure, including areas of transient or weak structure, and facilitates additional NMR experiments. A variety of NMR pulse sequences were used to investigate protein backbone dynamics on a ns-ps timescale, which measures bond vibrations. Dynamics experiments provide a wealth of information concerning regions of flexibility or restricted motion that are a fantastic complement to structural studies.

Analytical Ultra Centrifugation

AUC is a powerful method for analyzing the hydrodynamic properties of a particle by measuring the evolution of the sample concentration as a result of the applied centrifugal field. AUC can be used for many applications, such as calculating association or dissociation constants, but in our case, it was applied to the determination of protein molecular weight using sedimentation velocity experiments. Ultraviolet light absorption at 280 nm was used to monitor sample sedimentation over time, which directly correlates to the protein hydrodynamic radius. The AUC results

revealed the presence (or absence) of multiple species and allowed for accurate determination of their molecular weights.

Isothermal Titration Calorimetry

ITC is a technique used to determine the thermodynamic parameters of biomolecular interactions in solution. While there are many methods that can evaluate binding thermodynamics, ITC is unique in that it directly measures the enthalpy (ΔH) and indirectly measures the entropy (-TΔS) contributions to binding. In this work, ITC was used to compare the thermodynamic parameters associated with LC8 binding to different partners.

Dissertation Contents

My work on protein biophysics began with research that is not included in this dissertation. I was introduced to protein NMR during my rotation in Dr. Elisar Barbar's lab, where I solved the soltuion structure of the N-terminal domain of the Nsa-2 protein, a component of ribosome biogenesis. I was a co-author on the publication that includes this work[72], which illuminates the relay network of assembly factors that contribute to ribosome maturation. Additionally, as a part of my rotation in Dr. Andy Karplus' lab, I helped develop the program the Ensemblator, a highly useful tool that enables global and local comparison of protein structures. This work is published in Clark et al. *Protein Science* 2015[73]. These two projects, although unrelated to my thesis project, seeded my interest in protein NMR and established a solid theoretical foundation in protein biophysics that proved vital for my future studies.

This dissertation includes four chapters of original work, three of which are either published or represent a manuscript in preparation. Chapter 2 is a review published in *FEBS Letters* that examines different types of multivalent intrinsically disordered protein (IDP) assemblies[74]. The unique characteristics of LC8-associated IDP complexes are discussed in depth, which is a useful introduction to chapters 3-5, as each of these chapters focuses on a different LC8:IDP assembly.

Chapter 3 is a research article published in *Biochemistry* that examines the factors that influence LC8 binding to the TQT motif. The article uses the multivalent spindle assembly protein Chica as a model system to investigate why some motifs bind tightly and others bind weakly or not at all. We present the '*anchored flexibility model'* of LC8 motif recognition, which proposes that the TQT residues are essential for binding, while the other residues modulate

binding affinity. This model and the structural data presented in this work enhance our understanding of how LC8 recognizes and binds to its protein partners.

Chapter 4 is a research article in preparation for submission to *eLife* that illuminates the mechanism behind a novel method of negative feedback regulation that controls LC8 transcription. The article examines the interaction between LC8 and it's transcription factor ASCIZ and demonstrates how a combination of multivalency and intrinsic disorder allow ASCIZ to tune LC8 transcription. This research uses a combination of biophysical techniques to examine the structure of the ASCIZ:LC8 complex and *in vivo* transcription assays to test our *in vitro* hypotheses.

Chapter 5 is a partially complete study of the multivalent nucleoporin protein Nup159 and its interaction with LC8. In this study, I examine the structure and assembly process of the Nup159:LC8 complex. While this work is incomplete, is a useful addition to this dissertation because valuable insight can be gained from comparing the Nup159:LC8 complex to the multivalent proteins ASCIZ and Chica.

Chapter 6 summarizes the important findings from these studies and highlights their impact on the field. A detailed comparison of the three multivalent LC8 partner proteins - ASCIZ, Chica, and Nup159 - is presented. I conclude by discussing two ongoing experiments as well as suggestions for future work.

Chapter 2

Multivalent IDP assemblies: Unique Properties of LC8-Associated, IDP Duplex Scaffolds

Sarah A. Clark, Nathan Jespersen, Clare Woodward, and Elisar Barbar

Published in *FEBS Letters* (2015) 589, 2543-2551

Abstract

A wide variety of subcellular complexes are composed of one or more intrinsically disordered proteins (IDPs) that are multivalent, flexible, and characterized by dynamic binding of diverse partner proteins. These multivalent IDP assemblies, of broad functional diversity, are classified here into five categories distinguished by the number of IDP chains and the arrangement of partner proteins in the functional complex. Examples of each category are summarized in the context of the exceptional molecular and biological properties of IDPs. One type – IDP duplex scaffolds – is considered in detail. Its unique features include parallel alignment of two IDP chains, formation of new self-associated domains, enhanced affinity for additional bivalent ligands, and ubiquitous binding of the hub protein LC8. For two IDP duplex scaffolds, dynein intermediate chain IC and nucleoporin Nup159, these duplex features, together with the inherent flexibility of IDPs, are central to their assembly and function. A new type of IDP-LC8 interaction, distributed binding of LC8 among multiple IDP recognition sites, is described for Nup159 assembly.

Abbreviations

IC: dynein intermediate chain, LC8: dynein light chain 8, LC7: dynein light chain 7, Tctex1: An LC8-like dynein light chain, Dyn2: LC8 in yeast, DID: Dyn2 interaction domain, PP1: protein phosphatase 1, I-2: inhibitor 2, Keap1: Kelch ECH associating protein 1, Nrf2: nuclear erythroid 2-related factor 2

Introduction

'Intrinsic disorder' is a collective term that embodies the ensemble nature and inherent flexibility of the structure of intrinsically disordered proteins (IDPs). The remarkable pleiotropy of IDP function arises from an essential feature of ensemble structure, namely, that each IDP chain samples numerous interconverting conformations among which the specific local structure, and the degree to which it is favored, is coded in the amino acid sequence [75]. The simplest conceptualization of an IDP must include a dynamic equilibrium among a multiplicity of conformations with varying degrees of order, and is irreducible to an 'average conformation that lacks order' [76-78]. Widely recognized characteristics of ensemble structure that are crucial to IDP function are: flexibility that confers versatility [79, 80], reversibility that confers ready responsiveness to local cellular changes [81, 82], and assortment of ligand binding sites along the IDP sequence that optimizes concerted pathways and cascades [83-85]. The very broad range of essential biological functions to which IDP ensemble attributes are adapted [77, 86-92] are illustrated in numerous regulatory pathways in which an IDP complex is a critical constituent, e.g., the Wnt signaling pathway [93], mitochondrial initiated cell death [94], regulation of eukaryotic cell division [95], and a DNA damage repair pathway [79].

Multivalent IDPs simultaneously bind multiple partner proteins and are well suited to large macromolecular assemblies [88, 96, 97]. Multivalent IDP assemblies encompass supramolecular complexes of one or more multivalent IDP chains with one or more partner ligands that tend to be folded protein domains. The partners may be different proteins and/or multiple copies of the same protein. A growing recognition of the ubiquity and functional significance of multivalent IDP assemblies is reflected in their burgeoning interest among researchers in protein-related fields. Functional adaptations of the physical attributes of IDPs, in general, are eloquently described and reviewed [87-89, 96, 98, 99]. LC8/IDP duplex scaffolds constitute one category of multivalent IDP assemblies. To place IDP duplexes in context within the larger field, we developed a general classification scheme for multivalent IDP assemblies based on a few simple structural criteria, i.e., the number of IDP chains and the arrangement of partner proteins involved in forming a functional complex (Figure 2.1). In this review we discuss the unique structure/function properties of IDP duplex scaffolds, that is, features duplexes possess in addition to the many other remarkable features they have in common with single chain scaffolds. LC8/IDP duplex scaffolds function as core components in a broad array of cellular assemblies, each employing a different IDP and coordinating the collective activity of bound IDP partner proteins. Example IDPs in duplex scaffolds include IC in the dynein cargo domain [43, 100], Nup159 in the nuclear pore [36], zinc

finger protein ASCIZ [56], signal transduction protein Kibra [66], RNA-binding protein Swallow [44], mitotic protein Chica [62, 101], and rabies virus phosphoprotein.

The effort in our lab to characterize one category of multivalent IDP assemblies – polybivalent IDP duplex scaffolds – has led us to seek a general classification scheme based on a few simple criteria, i.e., the number of IDP chains and the arrangement of partner proteins involved in forming a functional multivalent IDP assembly, and to explicate IDP duplex scaffolds vis-à-vis other categories of multivalent IDP assemblies.

In Figure 2.1 we offer such a scheme. The discussion includes a brief summary of representative examples of each category along with a more extended analysis of features unique to polybivalent IDP duplex scaffolds, the category in which a duplex –formed by association of two IDP chains with the protein LC8– presents multiple sites for binding bivalent partners.

Categories of Multivalent IDP assemblies

We group multivalent IDP assemblies into A) Binary complexes, B) IDP single chain scaffolds, C) IDP duplex scaffolds, D) Higher order IDP associations, and E) IDP multi-site collective binding ligands. These five types are illustrated in Figure 2.1, and discussed in order below.

Binary Complexes

Binary complex assemblies (Figure 2.1A) consist of an IDP chain with several recognition motifs of variable length along its sequence that specifically interact with different sites on the surface of one folded partner. This group represents coupled binding and folding, a well described process in which an IDP binds several sites on a folded domain and in the process acquires a three-dimensional structure [102-106]. In folding-upon-binding, multiple IDP binding sequences interact specifically with as many folded domain sites, often distributed distally on the folded surface. This interaction mode is highly specific, as well as easily reversible in response to changes in cellular requirements. Two well known examples are illustrated in Figure 2.1A (upper frame), which depicts the interactions between protein phosphatase 1 (PP1) and intrinsically disordered regions (IDRs) of spinophilin (top) and protein inhibitor-2 (bottom). Each directs substrate specificity by blocking one of the three PP1 substrate binding sites. The intrinsically disordered nature of these two proteins allows them to "wrap around" PP1 and binds sites on different faces of the protein and cover one or more substrate binding sites [107, 108].

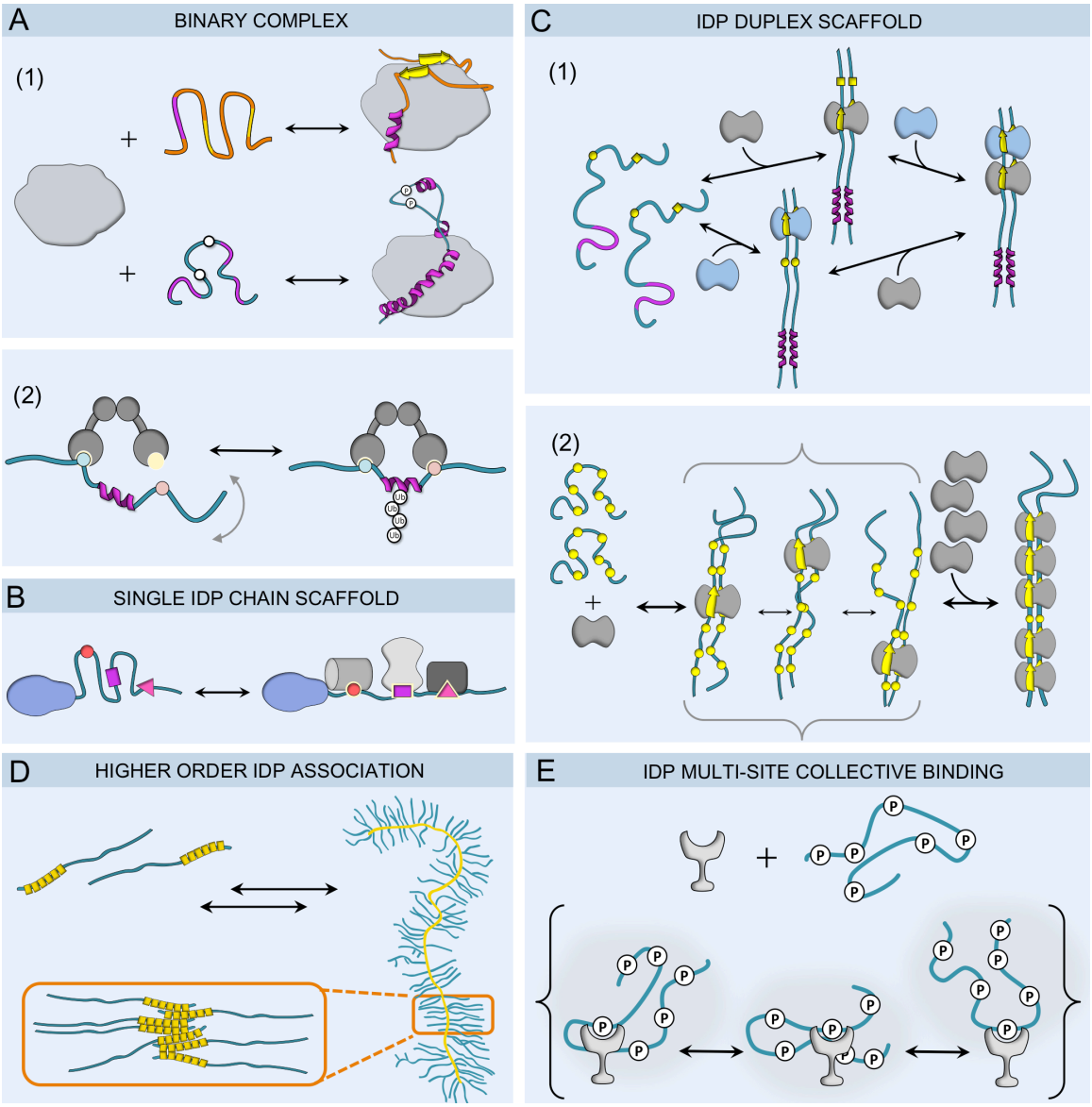


Figure 2.1. Multivalent IDP assemblies. Five categories are grouped by the number of IDP chains and arrangement of partner proteins.

(**A**) Binary complexes composed of one multivalent IDP and one binding partner. (1) Folding-upon-binding. PP1 (grey shape, left) binds either of two IDPs, spinophilin (upper chain) or protein inhibitor-2, I-2, (lower chain). Upon binding to PP1, spinophilin undergoes a "disorder to order" transition and forms two β-strands (yellow) and an α-helix (pink) at three PP1 contact points. In the coupled folding and binding of I-2, the IDP wraps around PP1 to form an extended α-helix (pink), and incorporates a loop structure that brings exposed phosphorylation sites Ser86 and Thr72 (white spheres) into close proximity, and subsequently turns on PP1 phosphatase activity. (2) The disordered domain of Nrf2 (aqua chain) has two binding motifs (blue, pink) for Keap1 (grey). In a proposed "hinge and latch" process, Keap1 binding to the first Nrf2 site

exposes six lysines in an α-helix (pink) located between the two motifs. Keap1 binding to both sites results in polyubiquitination (white spheres) of the helix (pink), and proteolysis of Nrf2. (**B**) Single IDP chain scaffolds. In this illustration a generalized multivalent IDR (aqua) with three linear motifs (sphere, rod, triangle) binds three partner proteins (cylinder, bi-lobed shape, rectangle) and facilitates interactions between the proteins in a concerted series of reactions. The intrinsically disordered region is part of a larger protein that also has a folded domain (blue pear shape). (**C**) Duplex IDP scaffolds consist of two IDP chains having at least one LC8 cross-link, and sites for other bivalent ligands and/or self-association domains. (1) Dynein IC binds LC8 and Tctex1 to form a scaffold (rightmost structure) with an ordered self-association domain (pink) and additional bivalent binding sites (not shown). Binding of one dimer, LC8, to the intrinsically disordered intermediate chain enhances the binding of the second dimer, Tctex1, and vice versa. (2) Nup159 DID chains (blue, at left) have six apparent LC8 motifs (yellow spheres) but binds five copies of yeast LC8 (Dyn2) to form a rigid duplex (rightmost structure). At lower molar ratios of Dyn2:Nup159, there is distributed binding of LC8 among the motifs, as depicted in the center brackets by the three flexible duplex species. (**D**) Higher order association of polyvalent IDPs can induce a phase change to form a large network of chains bound together by multivalent interactions. Shown is the formation of a protein hydrogel, as in wheat germ cell RNP granules, when repeats in the C-terminal domain of the protein self-associate. (**E**) In collective binding, a multivalent IDP interacts with one binding partner in a dynamic process (depicted in brackets) in which multiple IDP sites interact transiently and repeatedly with a single site on the partner. In the example shown, a single site on the receptor Cdc4 (crescent shape) interacts with multiple phosphorylated sites on the disordered cyclin-dependent kinase inhibitor (aqua chain). References and further discussion of the five categories are given in the text.

Another example in this category (Figure 2.1A (2)) is Keap1, a hub protein that modifies its intrinsically disordered binding partner Nrf2 through a proposed "hinge and latch" process [109, 110], leading to its polyubiquitination and subsequent degradation. The high-affinity site (blue) retains the ligand and acts as a hinge, while the weak-binding site (yellow) serves as a latch that operates only when Keap1 is in a specific conformation that is dependent on the cellular redox environment [110, 111]. Disorder and flexibility allow the weak-binding site frequent contact with the second, unoccupied binding site on Keap1 (Figure 2.1A(2), grey arrow).

IDP Scaffolds

IDP scaffolds, (Figures 2.1B,C), present a series of recognition motifs distributed along their sequence to a suite of partner proteins. Assortment of bound partners along the IDP facilitates the concerted operation of multiple components in a common function, along with efficient integration of regulatory proteins. Each motif is short, about 8-10 residues, and when

bound, is incorporated into the existing fold of the partner as an element of secondary structure flanked by disordered linkers.

IDP scaffolds attached at intervals to partner proteins are dynamic and flexible assemblies, in flux with numerous cell components in a tightly packed cellular environment, poised for specific and reversible interactions with an ever-changing pool of reactive partners and their regulators. Pleotropic function, ubiquitous distribution in intracellular systems, concerted action, fast response to changes in cellular needs, modulation by reversible binding of regulators, and participation in alternative pathways to the same end, are all recognized aspects of IDP scaffold structure/function relationships [89]. As these remarkable properties are shared by both single chain and duplex IDP scaffolds, we ask what *additional* features are offered or displayed only by duplex scaffolds. We answer this question in the context of a general overview of IDP duplex scaffolds with specific examples from our lab, illustrated in Figure 2.1C.

IDP Single Chain Scaffolds

The most numerous and well described group of IDP scaffolds contain one IDP chain bound to several or many partner proteins, Figure 2.1B. The intrinsic disorder of the IDP assists catalytic or signaling pathways by spatially and temporally integrating essential functional components. The limited space allotted here to single chain IDP scaffolds does not reflect their relative importance, but rather that excellent reviews of this broadly distributed and functionally essential category are readily available in the literature [89, 91, 112, 113].

A representative example of a disordered scaffold punctuated with short linear motifs is RNase E, an essential endoribonuclease conserved across many bacterial phyla [114]. Each RNase E chain binds three partner proteins whose collective action is important in RNA degradation and turnover and in RNA precursor processing [115]. The endoribonuclease activity of RNAse E resides in the structured N-terminal half of the protein, while the intrinsically disordered C-terminal region serves as a scaffold for the prokaryotic RNA degradosome consisting of three proteins: the DEAD-box RNA helicase, the glycolytic enzyme enolase, and the exoribonuclease (different grey shapes) [116-118]. RNAse E is an unusual scaffold in that its binding partners do not directly interact with each other [116], but rather, they function in sequence on a flexible scaffold that brings the three proteins into close proximity.

IDP Duplex Scaffolds

IDP duplex scaffolds, Figure 2.1C, are composed of two IDP chains in parallel alignment, connected by one or many bivalent partner proteins and/or by inter-chain interactions of identical sequence segments. Specific cross-linking sites are distributed along each IDP chain, and are either motif sequences or self-association sequences. Each motif (yellow sphere) recognizes a bivalent ligand (bi-lobed shape). Self-association sequences (pink) interact in the duplex to form inter-chain, dimeric domains.

While IDP duplex scaffolds share many extraordinary properties with single chain scaffolds, as outlined above, they also display unique features, namely: (1) parallel alignment of two IDP chains cross-linked by reversible, non-covalent protein-protein interactions; (2) formation of new ordered, self-associated domains resulting from interactions of identical sequences within each chain; (3) enhancement, relative to the same IDP in monomeric form, of binding affinity for additional bivalent ligands and of self-association tendency, and, (4) ubiquitous binding of LC8, a folded, bivalent protein first described in dynein. All four features are observed in all currently reported IDP duplexes, and new examples in the literature are steadily accruing. The growing range of biological functions served by IDP duplex scaffolds is impressive, from dynein assembly and regulation [43, 47-49], to nuclear pores [36, 38], to virus maturation [63, 119]. Additional examples are associated with mRNA localization and lung development [44, 45, 58]. While the full functional implications of IDP duplexes are still under examination in our lab and others, these four intriguing and unique characteristics of IDP duplex structure are well documented.

(1) Parallel alignment – in-register cross-linking – is the combined result of symmetrical binding of a bivalent ligand(s) and of dimerization of self-association sequences. Since there are typically multiple binding sites for binary partners along the IDP sequence [36, 43, 69], it is conceivable that cross-linked isoforms could occur if, for example, a partner binds one chain at motif number 1 and the second chain at motif number 2; but such out-of-register species have not been observed.

(2) Self-association domains of IDP duplexes cross-link the duplex along with binary partner proteins, and importantly, they *mediate the specificity of an IDP sequence for duplex formation* [47]. Once the duplex is formed, self-association domains also *extend* the specificity of an IDP sequence for regulatory and/or accessory ligands to other ligand types. This is accomplished by

formation of new ordered surfaces that may bind additional ligands, including those that are monovalent.

In the absence of partner proteins, apo ensembles of IDPs involved in duplex scaffolds are equilibrating populations of monomers and dimers. For some, the equilibrium favors monomers while for others the ensemble favors dimers; the relative monomer/dimer population of self-association domains presumably reflects the tendency of a given self-association sequence to engage inter-chain IDP interactions. In one well-characterized example, the dynein intermediate chain (IC), the apo ensemble favors monomers and also displays nascent order within the self-association sequences [11, 43, 47, 49]. In another case where the apo ensemble favors dimers, such as in Swallow and Myosin V [6, 44, 120], the self-association domains interact but retain conformational heterogeneity and flexibility. In a third example such as rabies phosphoprotein P [121], the self association sequences interact strongly and fold into highly ordered domains amenable to X-ray crystallography.

The ensemble nature of IDP structure means that tight binding of a new ligand to the self-association domain does not require that the major conformational population have high affinity self-association sequences. An ensemble that weakly favors self-dimerization can still bind specific dimer-targeted ligands by conformational selection. A notable consequence is that self-associating domains may constitute new targets for drugs directed against pathological IDP duplexes (e.g. RNA viruses [63, 119]) whether or not the self-association domain favors stable, folded dimers – any dimeric conformation in the ensemble could be 'selected' by the drug, and so shift the equilibrium to that conformation.

(3) Bivalency refers to the entropic enhancement of an additional binding event; it may be conceptualized as the increase in local effective concentration of motifs or self-association sequences or as an entropic 'pre-payment' of numerous degrees of freedom lost when two chains are constrained in a cross-linked duplex [43, 47, 48, 122]. Bivalency in the dynein IC duplex is discussed below.

(4) Cross-linking by at least one LC8 is *always* observed in the known IDP duplex scaffolds. Two IDP chains bind LC8, one in each of two grooves near the LC8 dimer interface. The crystal structure of the complex of LC8 and two copies of a peptide corresponding to the LC8 recognition motif in the IDP sequence is shown in Figure 2.2. The motif peptides are aligned in parallel, and each is incorporated as a new strand in the existing β-sheet of LC8. In one IDP

duplex, dynein intermediate chain discussed below, additional cross-linking proteins are observed and in other duplexes, multiple LC8s bind along the IDP chain [36, 64, 67, 69]. While an IDP duplex scaffold may bind other bivalent partners, e.g., Tctex1 and LC7 with dynein IC [43, 48], only LC8 occurs in every IDP duplex reported thus far.

In summary, the unique characteristics of IDP duplex scaffold structure – parallel alignment, self-association domains, bivalency, universal cross-linking by at least one LC8 – are mutually reinforcing and interdependent. For example, cross-linking at two segments (self-association domain *and* bivalent partner motif) promotes in-register alignment, parallel alignment promotes bivalency, and bivalency promotes binding of additional bivalent ligands, and so on. The unique characteristics of IDP duplex scaffolds, while ultimately a manifestation of the inherent properties of IDPs, originate in their unique duplex organization.

Higher Order IDP Associations

IDP assemblies composed of higher order self-associated complexes, Figure 2.1D, can reversibly incorporate and/or release other functional proteins. Often the assemblies undergo a phase transition and form a hydrogel, and act as a subcellular organelle unbounded by a membrane [123, 124]. A characteristic feature of these IDPs is low complexity sequences of phenylalanine-glycine, polyglutamine, or [G/S]Y[G/S] repeats [125-128], that often interact with homotypic and heterotypic proteins containing the same repeat [129, 130]. Intermolecular affinities result in networks of multivalent interactions forming an aggregate that is phase-separated from the surrounding solution (e.g. liquid droplets or hydrogels). The yeast germ cell RNP granules are one such example where these membrane-less organelles provide a means for cellular polarization of RNA during cell division [131].

Multi-site Collective Binding

Figure 2.1E illustrates an IDP in which multiple sites of a polyvalent ligand collectively engage a single site of a partner protein. This unusual binding mode involves interactions of the folded partner with various binding sites along the IDP chain. This category is represented at present by a polyphosphorylated IDP, in which several phosphate sites interact collectively with one site on a receptor [132].

LC8 Cross-linking of IDP Duplex Scaffolds

In duplex scaffolds, the IDP is specific to the system, but in all known examples, LC8 is a cross-linking partner protein. As such, LC8 has been termed an IDP dimerization 'hub' protein [12], and its broad occurrence indicates its central cellular functions. The name LC8 was coined in dynein literature to abbreviate 'light chain 8' of the dynein cargo attachment domain. Before the LC8 hub hypothesis it was common to assume that LC8 serves primarily as an adaptor molecule connecting dynein to dynein cargo. Thus, proteins that are LC8-associated were commonly assumed to function in conjunction with dynein transport of cargo. Since then, the hub idea has had numerous corroborating examples. In all LC8-IDP complexes for which structural information is available, around a dozen thus far, LC8 cross-links an IDP duplex scaffold in diverse functional systems with no proven dynein association.

The number of LC8 dimers in an IDP duplex spans the gamut, varying from one in Swallow and rabies phosphoprotein [6, 63], to several in Chica and nucleoporin Nup159 [36, 62], to many in ASCIZ (ATM-Substrate Chk Interacting Zn^{2+} Finger) [69]. Interestingly, ASCIZ functions as an LC8 transcription factor, and it has been proposed that LC8 binding to ASCIZ mediates the level of LC8 expression in the cell [56].

In principle, IDP duplexes cross-linked by bivalent proteins other than LC8 are possible but not yet identified, since many new IDP duplexes are initially recognized as LC8-binding proteins. However, the universality of LC8 as the IDP duplex crosslink is consistent with the example of rabies phosphoprotein, an IDP and one of only five proteins coded by the virus genome, which recruits host LC8 rather than any other cross-linking protein [63], and with the proposed LC8 self-regulation of its own expression.

In crystal structures of LC8 bound to short linear motifs from several different IDPs [8, 9, 34], e.g. Figure 2.2, all motifs interact with residues in identical symmetric grooves of the LC8 dimer. This binding versatility is ascribed to the flexibility of LC8 in that region [8, 133, 134]. The core of the LC8 homodimer is composed of a 12-stranded β-sandwich where each monomer contributes 5 β-strands and the sixth β-strand is formed by the bound peptide. The antiparallel β-sheet forces antiparallel binding of the IDP partner and therefore parallel orientation of the two IDP chains [8, 9]. While the LC8 consensus motif is identified [26] as an 8- to 10-residue sequence called a TQT, SQT or QT motif, not all such motifs bind LC8 [36, 38], and some motifs known to bind LC8 do not contain glutamine or threonine [61]. The variation of LC8 motif sequences suggests that involvement of LC8 in IDP duplexes has a long evolutionary history. If so, LC8 itself is expected to show some measure of evolutionary sequence divergence, and this is observed to a modest

extent in LC8 sequences from yeast and higher organisms, and in the structural relatedness of LC8 to another dynein subunit, Tctex1 [43], a structural homolog of LC8 that binds a second near-by motif on dynein intermediate chain IC. In vertebrate IC, the TQ motif is present in both Tctex1 and LC8 recognition motifs (SKVTQV and SKETQT in *Danio rerio*) suggesting that the ancestral IC was bound to two LC8 dimers. Indeed in yeast, two LC8 dimers (Dyn2 in yeast) bind near-by motifs on IC [135].

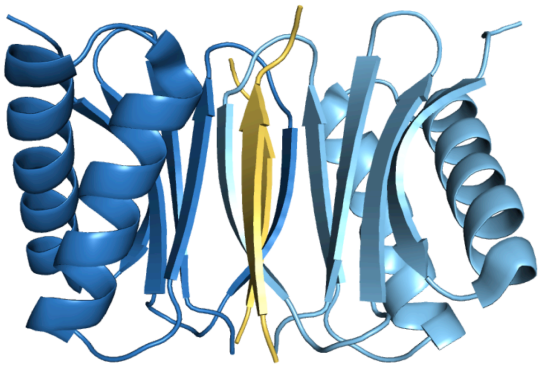


Figure 2.2. Crystal structure of LC8 bound to two recognition motif peptides.
Dimeric LC8 (light and dark blue monomers) forms a complex with two copies of a peptide (yellow) corresponding to an LC8 recognition sequence. The peptides are aligned in parallel, and each is incorporated as an additional β-strand in the β-sheet of an LC8 monomer subunit. In the LC8/IDP complex, one recognition motif on each IDP chain binds LC8, as in Figures 2.1C and 2.3.

Given the flexibility and disorder of IDPs, study of the molecular bases of duplex function requires a multidisciplinary approach. For this, we integrate NMR spectroscopy for study of residue level interactions of flexible domains, X-ray crystallography for atomic level structure of stable complexes, and isothermal titration calorimetry for determination of binding energetics. 13 crystal structures of LC8 bound to peptides have been solved [8, 9, 26, 31, 33, 34, 37, 39, 40, 61], revealing features of the LC8 recognition motif that confer affinity and specificity. NMR studies of the full-length LC8 binding domain have been performed for multiple binding partners [36, 83], shedding light on the transient structure and dynamics of these regions. A cryo-EM image of Nup159, showing 5 LC8 homodimers bound to two Nup159 chains in parallel has also been published and, remarkably, the rigid scaffold is so stable that is has been used as a marker in

cryo-EM experiments [136]. Additionally, *in vivo* studies of LC8-IDP duplexes are complicated by the ubiquitous nature of LC8 and the presence of multiple LC8 binding sites. Mice deficient in LC8 exhibit developmental defects [58, 137] and *Drosophila* with reduced LC8 levels die in early embryonic stages of development [60, 138]. Many LC8 binding partners contain multiple binding sites that must be fully removed in order to study the functional impact of LC8 binding *in vivo*, which becomes challenging when there is ambiguity in identifying all binding motifs. Below, we summarize the results of our structure/function studies of two LC8-IDP duplexes that integrate many of these approaches: dynein intermediate chain IC and nucleoporin Nup159.

Dynein IC Duplex Scaffold

Dynein intermediate chain IC is a core component of the cytoplasmic dynein cargo attachment domain. LC8 binding to apo IC dimerizes IC to form an IDP duplex scaffold (illustrated in Figure 2.1C, frame (1), and Figure 2.3) [43]. IC contains adjacent recognition motifs for LC8 and for Tctex1. The ternary complex is a stable polybivalent scaffold in which an LC8 dimer and a Tctex1 dimer each bind the same two IC chains. In the absence of binding partners, apo IC constructs are predominantly monomeric, with nascent order in the self-association domain [7, 11, 139]. When either LC8 or Tctex1 is bound to apo IC, affinity for the other is enhanced 50-fold relative to apo IC, due to a bivalency effect, namely, a reduced entropic penalty for the second binding event [43]. Similarly, binding of LC8 to apo IC enhances self-association interactions in a region of IC that is 84 residues away from the LC8 recognition motif, in the C-terminal helix self-association domain (pink) on the rightmost duplex of Figure 2.1C, frame (1). A bivalency effect is also observed when an artificial cross-link (inserted disulfide) is engineered at this site; in disulfide cross-linked IC, LC8 binding is enhanced 6-fold [47]. Not every duplex cross-link results in binding enhancement of other ligands. Dimeric light chain 7 (LC7) does not enhance overall LC8 binding affinity relative to monovalent IC, and *vise versa*, [47], possibly because LC7 binds to residues in the self-association domain and so disrupts self-association interactions.

The coupling of LC8 binding and IC self-association suggests that in the apo IC ensemble a small population of self-associated dimer has bivalency-enhanced LC8 affinity and so LC8 selectively binds this conformation, and initiates the assembly process. Then Tctex1 binding further stabilizes the duplex, which subsequently binds LC7. Later binding of LC7 to an already stabilized duplex is expected because LC7 disrupts the self-association domain and folds those residues into an LC7-bound helix, as shown by a crystal structure [47]. In the LC7-bound duplex, disruption of IC self-association interactions are apparently compensated by the combined effects

of bivalency and of formation of new transient IC-IC contacts detected by fluorescence energy transfer when both LC8 and LC7 are bound but not when only one is bound [47]. With LC8 bound, IC-IC self-association involves one helix from each chain, presumably packed against the other as in Figure 2.1C. With LC7 bound, these residues are folded into the LC7 structure, as described above. With both LC8 and LC7 bound, transient IC-IC interactions stabilize the linker between LC8 and LC7 (Figure 2.3, and legend). This conformational versatility of IC segments resides in the inherent flexibility and adaptability of IDP complex ensembles.

Thus, the assembly of two IC chains into a flexible polybivalent scaffold is modulated by long range coupling between IC self-association and LC8 binding and by subsequent binding of multiple additional bivalent partners. In its fully bound state, this polybivalent scaffold remains partially disordered and therefore structurally pliable [47, 83]. In addition to providing a duplex scaffold for the three dimeric dynein light chains, IC also interacts directly with components of several protein complexes (Dynactin, NudE, and RZZ) that regulate the function and activity of the dynein motor complex in the cell [140], and with the protein Huntingtin, implicated in Huntington's Disease [141].

Among the regulatory proteins that interact with IC, we are particularly interested in dynactin, which is critical in normal cellular functions [142], and in NudE, a 'nuclear distribution' protein that is essential in diverse processes including kinetochore and centrosome migration [143, 144]. Dynactin and NudE are localized with dynein in cellular compartments and they both bind to proximate regions on IC and with similar *in vitro* affinity, with NudE binding to region 1 but dynactin binding both regions 1 and 2 (Figure 2.3) [49, 83]. The coordination of their binding to IC is therefore a pertinent question. We have proposed that events that modify region 2 (brown), but do not significantly affect region 1 (orange), could regulate dynactin binding, but have limited effect on NudE binding [49, 145]. Region 2 is close to a disordered linker that has phosphorylation sites, suggesting that phosphorylation events may modulate binding of NudE *versus* dynactin (Figure 2.3). In summary, disorder in apo IC is an integral part of its assembly, and disorder retained in the IDP duplex is an integral part of its function and regulation of cargo binding and transport.

Considering its many interactions with light chains, regulatory molecules, and putative dynein cargo, IC appears to be a key modulator of dynein assembly and attachment to cargoes. Since the majority of these interactions are localized to the N-terminal 300 amino acid segment which is largely disordered [139] and is rich in phosphorylation and alternative splicing sites [146, 147], the disordered regions of IC therefore figure importantly in the functional versatility and binding-partner diversity of the entire complex.

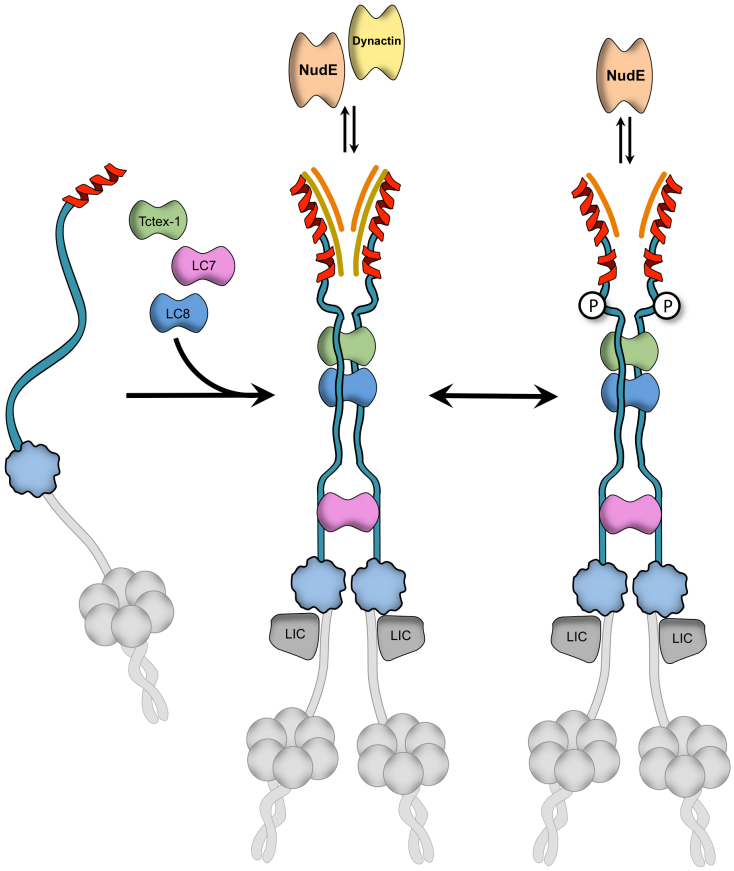


Figure 2.3. Schematic representation of cytoplasmic dynein IC duplex and two regulatory proteins.

(*left*) The dynein complex of intermediate chain, IC (blue), and heavy chain (grey). The 300 amino acid N-terminal domain of IC (aqua chain) is intrinsically disordered and monomeric except for a short single α-helix (red). The C-terminal domain of IC (light blue shape), predicted to be ordered, binds the tail of the heavy chain (light grey). In the motor domain of the heavy chain, six spheres represent the sites for ATP hydrolysis. (*middle*) Three homodimeric light chains Tctex1 (green), LC8 (blue), and LC7 (purple) bind IC and form a duplex. We hypothesize that LC8 binding initiates the duplex formation. Transient IC-IC contacts, present when both LC8 and LC7 are bound, are indicted by short segments of close, parallel blue chains. The IC duplex is essential for formation of the dimeric motor domain that is necessary for dynein activity [39, 148]. Light intermediate chains (LIC) are dark grey. The N-terminal domain of IC also binds dynein regulator proteins NudE (light orange, bi-lobed) and dynactin (light yellow, bi-lobed); both are dimeric coiled-coils that bind to residues in the helical region (red). NudE binds only to Region 1 (orange bar), while dynactin binds to Regions 1 and 2 (brown bar), although the overall IC affinity is similar for both. Regions 1 and 2 are separated by a short flexible linker. The resulting IC duplex contains other disordered linkers, one separating Region 2 from the Tctex1 binding site, and another separating the LC8 binding from the LC7 site. (*right*) Phosphorylation of the disordered linker between Region 2 and the Tctex1 binding motif prevents dynactin binding [149].

Nup159 Duplex Scaffold

The LC8-Nup159 duplex scaffold illustrates IDP duplexes that are incorporated into multicomponent subcellular structures, in this case the yeast nuclear pore complex (NPC), a 60 MDa complex that directs nucleocytoplasmic transport. One essential module of the NPC is the Nup82 subcomplex, which is located on the cytoplasmic side of the NPC and is involved in nuclear mRNA export. The Nup82 module is comprised of Dyn2 (the yeast LC8 ortholog), Nup159, Nup82 and NSP proteins [68]. Nup159 is a 159 kDa protein whose sequence is predicted to be a β-propeller at the N-terminal end, followed by a long segment of FG repeats lacking secondary structure, then by the Dyn2 interacting domain (DID) sequence bearing multiple recognition motifs for Dyn2, and at the C-terminus a predicted helical segment that includes coiled-coil motifs. The DID between the FG repeats and the predicted coiled-coil forms a duplex of 2 Nup chains cross-linked by 5 Dyn2 dimers, as diagrammed in Figure 2.1C, frame (2).

In a series of constructs of the DID containing increasing numbers of Dyn2 recognition motifs, isothermal titration calorimetry (ITC) shows that when one motif is present, Dyn2 binding is very weak, but when two motifs are present, the average affinity is significantly increased, indicating cooperative binding due to bivalency [36]. Similarly, when three recognition motifs are present, fits to the single ITC binding curve give an average Dyn2 affinity that is significantly higher than when only two motifs are present. When the fourth and fifth recognition motifs are included in the constructs, however, the trend is ended, and the average affinity is slightly decreased.

In full length DID, with all motifs present, ITC data show a stoichiometry of 5 Dyn2 per duplex in the presence of excess Dyn2. But at substoichiometric ratios of Dyn2:Nup159, there is no Dyn2 binding preference for one motif *versus* another, resulting in a mixture of partially bound duplexes [36]. This is inferred from NMR titration methods that monitor involvement of specific residues in Nup159-Dyn2 binding interactions; in the titration experiments, loss of intensity is measured for samples with increasing Dyn2:Nup159 ratios. Diminished intensity of backbone NH peaks indicates direct involvement of that residue in Dyn2 binding and/or restricted local motion due to proximate binding. We observe lower peak intensity along the entire DID sequence even at Dyn2 concentrations equivalent to one Dyn2 dimer per DID duplex [Clark & Barbar, unpublished data], meaning that binding of the first Dyn2 molecules occurs at multiple sites.

This distributed binding of Dyn2 among the five Nup159 motifs is reminiscent of the multisite collective binding of Figure 2.1E, where multiple phosphorylation sites on an IDP

apparently interact transiently with the binding pocket of the partner protein, resulting in a dynamic ensemble of IDP/ligand complexes. While this system has some similarity to the Nup-Dyn2 complexes at substoichiometric ratios of Dyn2, there is a significant difference: in functional Dyn2-Nup159 duplexes, all IDP sites are *fully* bound to ligand. Thus, the multi-site binding of Dyn2 is more a distributed binding process, rather than a collective binding mode. As more Dyn2 dimers are added, their binding is distributed all along the duplex until, ultimately, the assembly attains the final ratio of 5 Dyn2 per duplex.

Distributed binding of Dyn2 retains Nup159 flexibility in ensembles of partially cross-linked duplexes, and we expect that this is essential to the biological activity of Nup159. A possible function of distributed Dyn2 binding is to maintain flexibility at lower Dyn2 concentrations before the final assembly of five Dyn2 in the Nup159 duplex that is apparently quite rigid and thus unfavorable energetically, i.e. of very low entropy. Distributed binding could optimize the system-wide energetics as increasing numbers of Dyn2 are added by averaging the entropic penalty over numerous binding reactions, and maximizing the bivalent affinity enhancement of various Nup159 motifs for Dyn2.

The expectation that the fully bound Dyn2-DID complex is rigid comes from electron microscopic studies of nuclear pore formation[38]. In the Dyn2-DID complex, a ~20 nm long rod-like structure is observed in electron micrographs of the DID domain [38]. In these structures five Dyn2 dimers are stacked like beads on two Nup159 strands and the rigidity of this parallel arrangement is suggested to help protrude Nup159 FG repeats and the N-terminal β-propeller domain toward the central transport channel [36, 38]. Thus it appears that the Dyn2-Nup159 duplex is a key organizing component of the NPC complex, directing FG repeats towards the central transport channel to form a cargo-accessible domain. Consistent with the proposal that the functional adaptation of multiple LC8 binding sites is distributed binding, our ITC data show that DID constructs having either the last two motifs or the first three motifs form a maximally stable complex [36], and therefore the evolutionary selection of five Dyn2 cross-links is apparently not for overall stability.

In addition to distributed binding, Nup159 has a sequence near the DID that is crucially involved in pore formation, and that appears to be a self-association domain; this is significant because self-association, alignment and duplex stability are mutually reinforcing. Recent structural data for a reconstituted Nup82 module suggest that the C-terminal portion of Nup159, which includes the DID and the C-terminal helices, forms the structural backbone of the complex, along which the other subunits – Nup82, Nsp1, and Dyn2 – become organized, and that this

segment of Nup159 is required for organization of the whole complex [68]. Critical to this process is a short, weakly predicted coiled-coil sequence separated from the DID domain by a flexible linker, and shown *in vivo* to be essential for NPC assembly. Since IDP self-association and LC8 binding are mutually reinforcing, the weakly predicted coiled-coil sequence in Nup159 likely forms a self-association domain in the duplex. The *in vivo* demonstration that the weakly predicted coiled-coil sequence adjacent to the DID is essential for NPC assembly suggests that in the duplex these residues form a self-association domain which plays a critical role in nuclear pore assembly.

In summary, the thermodynamics of Dyn2 binding to Nup159, taken together with NMR titration experiments of Nup159 with Dyn2, implies that the essential involvement of the Nup159-Dyn2 complex in nuclear pore formation arises from unique properties of the IDP duplex scaffold. These include self-association and distributed binding, the latter a new process in multivalent IDP assemblies. We expect that, in general, for IDP duplexes having multiple LC8 motifs, distributed binding underlies the function of multiple LC8 cross-links. In the Nup159-Dyn2 duplex, distributed binding is apparently a novel adaptation of IDP flexibility to produce duplex rigidity.

Future Directions for IDP duplex scaffolds

Current studies of multivalent IDP duplex assemblies tend to focus on IDP-partner protein interactions. It is clear that duplex scaffolds align in parallel, form self-associated domains, display bivalent enhancement of affinity for additional ligands, and are always cross-linked by LC8. Other aspects of growing interest in IDP duplexes are the processes associated with flexible sequence segments, or linkers, between IDP binding sites for partner proteins. Flexible linkers often connect essential functional domains, and as such are well placed to host novel regulatory processes such as competition of regulatory molecules for overlapping binding sites[83, 112, 150], alternative splicing, and posttranslational modification. An example of a regulatory effect from posttranslational modification, is in the dynein IC duplex where phosphorylation in a disordered linker abolishes IC binding to the regulator dynactin [145]. An example in the Nup159 duplex is the linker between the DID and a self-association domain, which when deleted, abolishes binding of Dyn2 to DID *in vivo* even though the DID domain is left intact [68]. Flexible linkers also contain dynamic self-association domains that acquire alternative ordered conformations, as in IC segments that make up the binding site of LC7[112]. Thus, although only a handful of IDP duplex scaffolds are well characterized thus far, upcoming molecular-level studies promise new

structure/function advances since, remarkably, IDP duplexes participate in essential cellular functions in over 100 diverse systems [12, 26, 151].

Funding

This work was supported by the National Institutes of Health Grant GM 084276 and the National Science Foundation Grant MCB 0818896.

Chapter 3

The Anchored Flexibility Model in LC8 Motif Recognition: Insights from the Chica Complex

Sarah Clark, Afua Nyarko, Frank Löhr, P. Andrew Karplus, and Elisar Barbar

Published in *Biochemistry* (2015) 55, 199-209

Abstract

LC8 is a dimeric hub protein involved in a large number of interactions central to cell function. It binds short linear motifs – usually containing a Thr-Gln-Thr (TQT) triplet – in intrinsically disordered regions of its binding partners, some of which have several LC8 recognition motifs in tandem. Hallmarks of the 7-10 amino acid motif are a high variability of LC8 binding affinity, and extensive sequence permutation outside the TQT triplet. To elucidate the molecular basis of motif recognition, we use a 69-residue segment of the human Chica spindle adaptor protein that contains four putative TQT recognition motifs in tandem. NMR-derived secondary chemical shifts and relaxation properties show that the Chica LC8 binding domain is essentially disordered with a dynamically restricted segment in one linker between motifs. Calorimetry of LC8 binding to synthetic motif-mimicking peptides shows that the first motif dominates LC8 recruitment. Crystal structures of the complexes of LC8 bound to each of two motif peptides show highly ordered and invariant TQT-LC8 interactions and more flexible and conformationally variable non-TQT-LC8 interactions. These data highlight rigidity in both LC8 residues that bind TQT and in the TQT portion of the motif as an important new characteristic of LC8 recognition. Based on these data and others in the literature, we propose that LC8 recognition is based on rigidly fixed interactions between LC8 and TQT residues that act as an anchor, coupled with inherently flexible interactions between LC8 and non-TQT residues. The ‘anchored flexibility’ model explains the requirement for the TQT triplet and the ability of LC8 to accommodate a large variety of motif sequences and affinities.

Abbreviations

Ana2, anastral spindle-2; ASCIZ, ATM-Substrate Chk-Interacting Zn^{2+} Finger; QT1p, QT2p, QT3p, and QT4p, peptides corresponding to recognition motifs 1-4, respectively; ITC, isothermal titration calorimetry;

Introduction

LC8, also known as DLC1 or DYNLL, is a highly conserved, essential protein that interacts with numerous protein partners involved in widely diverse functions. Originally identified as a subunit of the motor complex dynein [152], LC8 was later shown to be a component of many large multi-protein complexes [12, 26, 151] involved in broadly varying processes ranging from tumor suppression to viral replication and DNA damage repair [4, 56, 66, 101, 153]. LC8, a homodimer at physiological conditions [154], invariably binds pairs of the same partner, always a disordered protein or a disordered region of a partially folded protein [12, 155]. Crystal structures show that all recognition motifs presented by LC8 partners bind in the same manner to two symmetrical grooves on the LC8 dimer despite their highly variable sequences. The ~10-amino acid recognition motifs are incorporated as anti-parallel β-strands [8, 31-34, 39, 40, 156]. In its partner-bound form (Figure 3.1a), the LC8 dimer is a 12-stranded β-sandwich where each monomer contributes 5 β-strands (β1-β5, blue) and the recognition motif is the sixth (orange). The antiparallel β-sheet is flanked on each side by pairs of helices (blue).

The 10-amino-acid motif recognized by LC8 includes the signature sequence Thr-Gln-Thr (TQT) at the C-terminus. In a motif logo generated from recognition sequences of 11 crystal structures of LC8/peptide complexes (Figure 3.1b) residues are numbered from the conserved glutamine at position zero and adjacent residues -1 and 1 [8]. While the TQ_0T is an almost universal feature of known motif sequences, other positions show notable conservation, such as aspartate at position 2 and to a smaller extent, residues at -2, -3 and -4. A few partners have different triplet sequences, such as GTQ_0CD in Ana2 and ATS_0PI in Pak1 [32, 34], and Q_0 replacement by methionine in Myosin 5A [120] or asparagine in Kibra [100]. The variability in motif sequence outside TQ_0T is accompanied by a similarly wide range of binding affinities, from 160 nM to 42 μM [31, 32]. In summary, while the TQT triplet is very highly conserved, a diverse array of peptide sequences outside positions TQ_0T are accommodated in the binding groove.

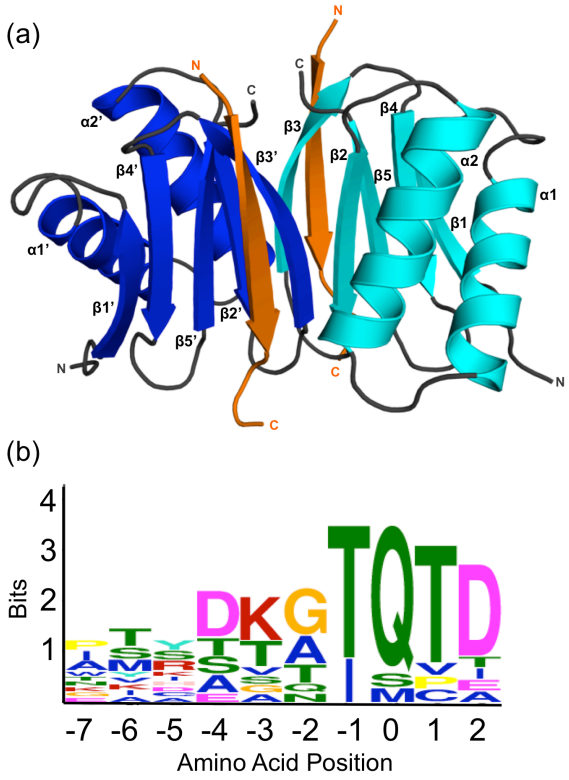


Figure 3.1. LC8 binds many partners in symmetrical grooves at its dimer interface. (a) A ribbon diagram of the LC8 dimer (blue and cyan) bound to two chains of the Chica QT1 peptide determined in this study (orange). Structures of all LC8/peptide complexes look essentially the same. LC8 secondary structure elements are labeled. (b) A sequence logo of LC8 binding motifs derived from sequences of the motifs in the 11 crystal structures reported for LC8/peptide complexes. Height of amino acids indicates their relative frequency at that position.

Chica, a protein with multiple LC8 recognition motifs, is a 64-kDa mitotic spindle-associated protein that binds LC8 to form a complex that promotes asymmetrical cortical localization of dynein and correct spindle orientation. Experiments that interfere with LC8-Chica association, including LC8 siRNA knockdown or specific mutations of the TQT residues, demonstrate that Chica interaction with LC8 is necessary to target dynein to the cell cortex and to orient mitotic spindles for the proper onset of mitosis [62]. Other proteins in this intriguing group of LC8 partners with multiple recognition motifs include nucleoporin Nup159, with six TQT motifs in tandem of which five bind LC8 [36], and the transcription factor ASCIZ with 17 motifs of which

11 bind LC8 [69]. Chica has four conserved TQT motifs (QT1-4) within the intrinsically disordered segment of its C-terminal half (Figure 3.2). Mutation of the TQT sequences within QT1, QT2, and QT4 completely abolish binding to LC8 [62], a strong indication that LC8 does not bind QT3. The one-residue linker separating QT2 and QT3 implies that only one of these can bind LC8.

Here, using nuclear magnetic resonance spectroscopy (NMR), isothermal titration calorimetry (ITC), and X-ray crystallography, we have determined structural and dynamic properties of the LC8 binding domain of Chica and the thermodynamic and structural characteristics of its interactions with LC8. These studies were carried out on a 69-amino acid Chica construct containing the four TQT recognition motifs and on short synthetic peptides corresponding to each of the recognition motifs. We show which TQT motif does not bind LC8, and of the three that do bind, which dominates LC8 recruitment by Chica. Our findings with Chica support a model of LC8-motif binding that is anchored by a set of favorable and conserved interactions with the TQT residues, and that is modulated in a difficult-to-predict manner by the variable residues at other motif positions that serve to decrease or enhance the overall affinity.

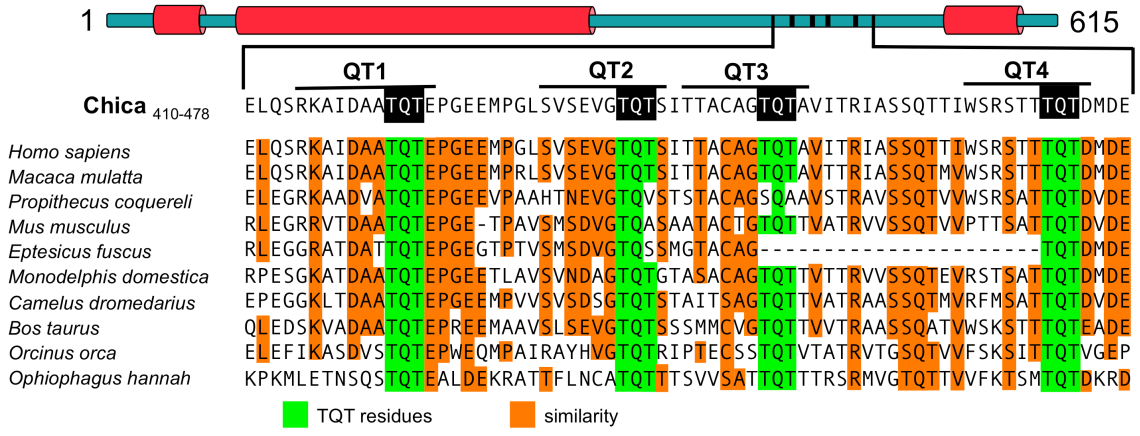


Figure 3.2. The LC8 binding region of Chica is disordered and conserved.
Four putative LC8 binding sites, QT1, QT2, QT3, and QT4 (black highlight) are nestled within a segment of Chica (410-478) predicted to be disordered (teal). Predicted structured regions are shown as red cylinders. Conserved TQT residues in each motif are highlighted in black and a black bar indicates the full 10-amino acid sequence. A sequence alignment across 10 species is shown below with 100% identity shown in green and similarity shown in orange. Species were chosen to cover a wide genus range; percent sequence identity relative to the *Homo sapiens* sequence ranges from 67-96%.

Results

Solution Properties of Chica$_{410-478}$

Sequence-based secondary structure prediction of Chica using the PSIPRED algorithm [157] predicts three helical regions (Figure 3.2, red cylinders) while the rest of the protein, primarily the C-terminal half, is predicted to be unstructured (Figure 3.2, teal). Within the disordered region, the segment corresponding to residues 410-478, Chica$_{410-478}$, contains four putative LC8 recognition sites with the canonical TQT recognition motif. Based on the position of the TQT triplet in each motif, we assign the LC8 interacting motifs to residues 415-424 (QT1), 433-442 (QT2), 444-453 (QT3) and 466-475 (QT4) (Figure 3.2, black bars). Sequence alignment of Chica$_{410-478}$ from 9 mammals and 1 reptile shows high but not perfect conservation of the TQT sites (Figure 3.2, green).

Recombinant N-terminally His-tagged Chica$_{410-478}$ was produced, and in agreement with the predicted disorder, its ^{15}N HSQC spectrum (Figure 3.3a) has a pronounced lack of amide proton chemical shift dispersion, and the CD spectrum shows limited ellipticity at 220 nm and a large negative ellipticity at 200 nm (Figure 3.3b). Resonance assignments were obtained for all residues (Figure 3.3a). Secondary chemical shifts indicate some weak β-strand character for residues within QT1 and the linker separating QT3 and QT4, but no apparent secondary structure preference in the rest of the sequence (Figure 3.3c).

To provide insight into the local dynamics of specific segments within Chica$_{410-478}$, ^{15}N longitudinal (T_1), transverse (T_2), and ^{1}H-^{15}N heteronuclear NOE relaxation parameters were measured. *A plot of T_2:T_1 ratios versus residue number suggests motional heterogeneity (Figure 3.3d). T_2/T_1 values* range from 0.53 to 0.96, with an average value of 0.62 (dotted line), with the lowest for residues in the linker separating QT3 and QT4, suggesting motional restriction in this region associated with a tendency to form transient structure. Measured heteronuclear NOE values at 15 °C are all zero, as no peaks were detected in the NOE subspectrum, except for the C-terminal residue 478, which has a value of -0.3 (data not shown). We conclude that in this highly disordered segment, there are regions of transient structure in the linker between QT3 and QT4 inferred from the combination of β-sheet propensity from secondary chemical shifts and restricted dynamics.

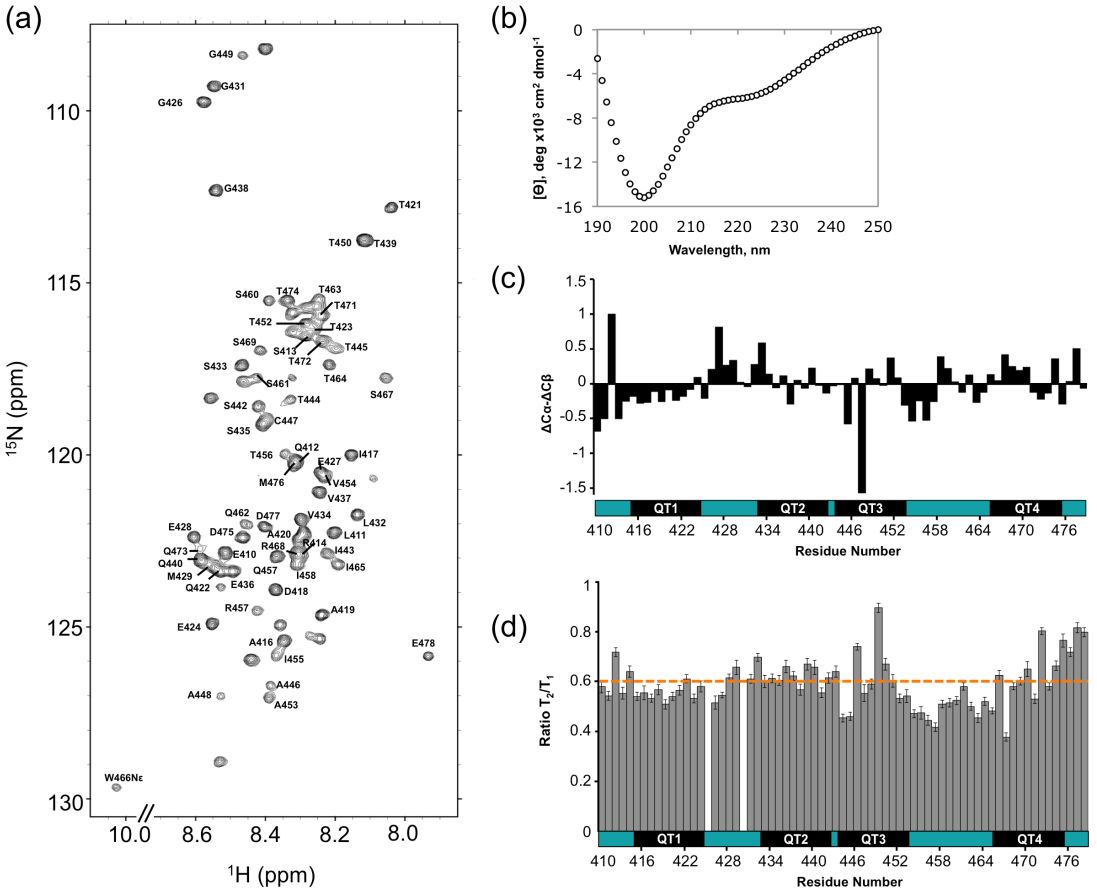


Figure 3.3. Spectroscopic analyses of Chica$_{410-478}$.
(a) [^{15}N-^{1}H]-BEST-TROSY spectrum showing backbone assignments for all 67 non-proline residues. Unassigned residues correspond to the additional N-terminal residues from the expression vector. The spectrum was recorded at 15 ^{o}C. (b) Far UV CD spectrum collected at 25 ^{o}C. (c) A plot of chemical shift differences *versus* residue number. ΔCα and ΔCβ values were calculated by subtracting the random coil chemical shift from the experimentally determined chemical shift value [158]. ΔCα – ΔCβ values > +/- 1.0 ppm were considered significant. (d) Plot of T_2/T_1 relaxation per residue. Segments corresponding to recognition motifs, QT1, QT2, QT3, and QT4 are shown below the plot.

Binding of LC8 to Chica$_{410-478}$

ITC experiments (Figure 3.4a and Table 3.1) show that LC8 binds Chica$_{410-478}$ with a K_d of 0.4 μM and a stoichiometry of 3:1 indicating that, consistent with earlier mutagenesis studies [62], only three of the TQT recognition motifs bind LC8. NMR titration studies to identify which motifs bind were unsuccessful due to limited solubility of the complex at increasing LC8 concentrations. To determine the relative affinities of the recognition motifs, we synthesized corresponding short peptides QT1p, QT2p, QT3p, and QT4p and measured their binding to LC8 by ITC (Figure 3.4b-d and Table 4.1). LC8 interacts with QT1p with a dissociation constant of 0.4 μM, identical to the affinity of the full-length segment, while QT2p and QT4p bind to LC8 with over 10-fold weaker affinity with K_d values of 6.3 and 5.2 μM, respectively (Figure 3.4c,e). The interaction with QT3p is too weak to be measured accurately by ITC (Figure 3.4d). Interestingly, the thermodynamic parameters of QT1p and Chica$_{410-478}$ are very similar; both show favorable binding entropy compared to QT2p and QT4p (Table 3.1), implying that the interaction of LC8 with QT1 is the dominant binding reaction in the full-length construct. The TΔS terms are of opposite sign for QT1p versus QT2p and QT4p, resulting in more than an order of magnitude higher affinity constant for QT1p, even though the ΔH term is more favorable for QT2p and QT4p binding.

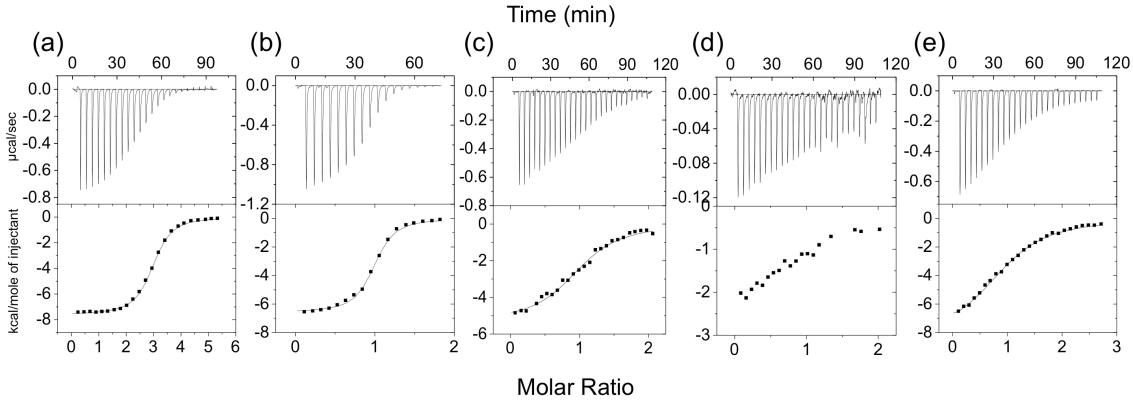


Figure 3.4. LC8-Chica interactions monitored by isothermal titration calorimetry. Representative thermograms are shown for the interaction of LC8 with (a) Chica$_{410-478}$, and peptides corresponding to (b) QT1p (c) QT2p, (d) QT3p, and (e) QT4p recognition sequences.

Construct	Construct Sequence[a]	N	K_d (μM)	ΔH (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
Chica $_{410-478}$		3	0.4 ± 0.1	-7.6 ± 0.2	1.2 ± 0.4	-8.8± 0.2
QT1p (414-424)	SYRKAIDAATQTEE	1	0.4 ± 0.1	-6.8 ± 0.2	1.9 ± 0.2	-8.7 ± 0.2
QT2p (433-443)	SYSVSEVGTQTSI	1	6.3 ± 0.4	-8.5 ± 1.1	-1.4± 1.1	-7.1 ± 0.04
QT3p (444-454)	SYTTACAGTQTAV	---	ND	---	---	---
QT4p (466-476)	SYWSRSTTTQTDM	1	5.2 ± 0.1	-8.2 ± 0.1	-1.2 ± 0.3	-7.0 ± 0.2

Table 3.1. Thermodynamic parameters of the LC8-Chica interactions.
[a]Non-native residues added to the N-terminus of each peptide to increase solubility and improve concentration determination are underlined.

Crystal Structures of LC8-Chica Motif Complexes

To elucidate the molecular determinants of the LC8-Chica interaction, crystal structures of QT1p and QT4p in complex with LC8 were solved. QT1p and QT4p were chosen because they bracket the range of binding thermodynamics and affinities (Table 3.1). Although the complexes were crystallized under different conditions (see "Experimental Procedures"), the crystal forms are isomorphous and contain one LC8 monomer and peptide in the unit cell. After facile solution by molecular replacement, refinements led to a LC8-QT1p complex at 1.3 Å resolution with R/R_{free} values of 19.6%/20.2%, and a LC8-QT4p complex at 1.6 Å with R/R_{free} values of 20.7%/23.2% (Table 3.2).

Both peptides bind to LC8 in a similar, extended conformation (Figure 3.5) in keeping with previously reported crystal structures of LC8 with bound partner peptides. The conserved TQT portions of QT1p and QT4p show identical hydrogen bonding interactions with LC8 and identical buried water molecules (Figure 3.5b,c). Both structures also highlight similar electrostatic interactions between a side chain carboxylate at position 2 (Glu424 in QT1p or Asp475 in QT4p) and Lys9 of LC8.

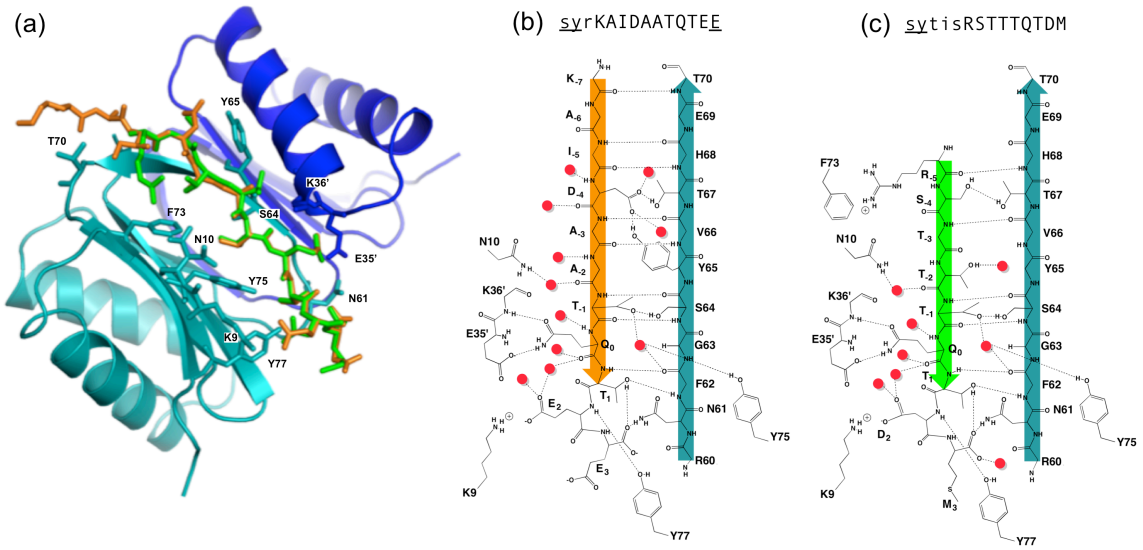


Figure 3.5. Comparisons of crystal structures of the LC8-QT1p and LC8-QT4p complexes. (a) Chica peptides QT1p(orange) and QT4p (green) overlaid after alignment of the mainchain residues of their respective LC8 homodimers. The side chains of LC8 residues that form hydrogen bonds or a cation-π interaction (F73) with QT1p or QT4p are shown as sticks and labeled. For the sake of clarity, only one peptide bound to LC8 is shown. A schematic of (b) QT1p (orange) bound to LC8 compared to (c) QT4p (green) bound to LC8 highlights similarities and differences in LC8 residues involved in the hydrogen bonding or stabilizing interactions based upon the peptide sequence. The sequence of each peptide is shown above each diagram with capital letters corresponding to residues that are observed in the crystal structure, while those in small letters are part of the peptide but are not observed and underlined residues are nonnative.

	PDB code: 5E0L	PDB-code: 5E0M
Data statistics		
structure	LC8-QT1p	LC8-QT4p
space group	$P6_122$	$P6_122$
unit cell *a, b, c* (Å)	44.67, 44.67, 204.17	44.18, 44.18, 204.71
resolution (Å)	51.01-1.31 (1.33-1.31)[a]	68.34-1.64 (1.67-1.64)
completeness (%)	100.0 (99.1)	100.0 (100.0)
unique reflections	30438 (1438)	15900 (754)
multiplicity	30.8 (8.0)	33.4 (24.7)
R_{meas} (%)	8.3[b] (47.9)	37.8[c] (931)
$\langle I/\sigma \rangle$	20.6[d] (0.4)	9.8[e] (0.5)
$CC_{1/2}$	1.0 (0.26)	1.0 (0.17)
Refinement statistics		
Resolution range (Å)	38.7-1.31 (1.35-1.31)	38.6-1.65 (1.78-1.65)
R-factor (%)	19.6 (47.2)	20.7 (44.8)
R_{free} (%)	20.2 (48.2)	23.2 (47.3)
protein residues	97	95
water/SO_4 molecules	59/1	60/2
sulfate molecules	1	2
total atoms	1704	1688
rmsd[f] lengths (Å)	0.014	0.007
rmsd angles (°)	1.3	1.1
Ramachandran plot[g]		
φ,ψ-preferred (%)	97	97
φ,ψ-allowed (%)	3	3
φ,ψ-outliers (%)	0.0	0.0
B-factors		
<main chain> (Å^2)	33	34
<side chains and waters> (Å^2)	46	45

Table 3.2. Data collection and refinement statistics of X-ray crystal structures.
[a]Values in parentheses are for the highest-resolution shell. [b]R_{meas} in the inner shell (51.02-7.18 Å) is 4.6%. [c]R_{meas} in the inner shell (68.34-8.98 Å) is 15.8%. [d]$\langle I/\sigma \rangle$ in the inner shell is 75.2 and falls to ~2 at 1.55 Å. [e]$\langle I/\sigma \rangle$ in the inner shell is 35.4 and falls to ~2 at 2.1 Å. [f]rmsd = root mean square deviation. [g]Preferred, allowed, and outlier regions as assigned by Molprobity.

Structural Differences Between LC8-QT1p and LC8-QT4p

The most obvious difference between LC8-QT1p and LC8-QT4p complexes is that the well-defined electron density in QT1p extends the ordered β-strand by two residues compared to QT4p (Figure 3.6a), implying a high degree of disorder in the N-terminal residues of LC8-QT4p. The last five residues overlay well in the LC8 binding groove with a backbone RMSD of 0.19 Å (Figure 3.5a). Two notable differences localized to the N-terminal half of the peptide are consistent with differences in QT1p and QT4p binding affinities. The first is at position -3, where a threonine (Thr470) in QT4p substitutes for an alanine in QT1p. Thr470 does not fit the LC8 binding pocket as well and, relative to Ala419, the peptide backbone of Thr470 is pushed out ~0.9 Å at the Cα atoms (Figure 3.6b). This distorts the hydrogen bonding geometry at three locations: between Thr470 O and Val66 N, Ser469 O and a water molecule, and Thr471 N and a water molecule (Figure 3.6b, pink arrows). The perturbed hydrogen bond geometry and altered packing of Thr470 backbone atoms could contribute to the lower binding affinity of QT4p. At position -4, both the QT4p Ser469 side chain oxygen and the QT1p Asp418 side chain oxygen hydrogen bond with the side chain oxygen of Thr67 (Figure 3.5b).

The second difference is the presence of Arg468 at position -5 in QT4p. In QT1p, a buried Ile occupies this position. A Val or Ile in the -5 position has been proposed to be an important determinant of binding affinity both by its observed packing against the imidazole group of LC8 His68 [26] and by a pepscan analysis showing that Val or Ile at the -5 position enhances binding [159]. Although a buried Arg at position -5 is presumably destabilizing compared to an Ile or Val, it appears to be acceptable at this position due to a stabilizing cation- π interaction that it makes with Phe73.

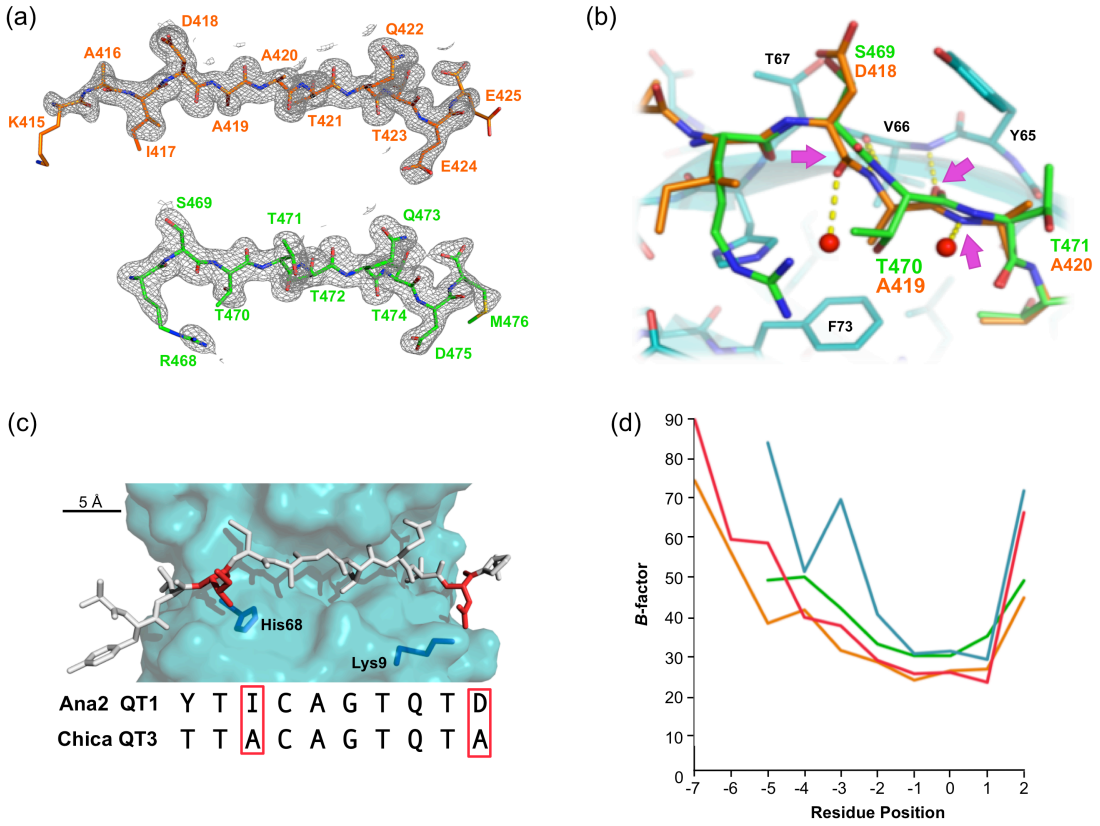


Figure 3.6. Crystal structures of motif peptides bound to LC8.
(a) The $2F_O$-F_C electron density for Chica QT1p (top) and Chica QT4p (bottom) is shown as grey mesh, contoured at 1.0σ. (b) The main chain conformation of the QT4p (green) is different from the QT1p (orange) at position -3 (T_{-3}) due to a buried T470 in LC8. QT4p H-bonding geometry is unfavorable at three locations (pink arrows). (c) Ana2 QT1 peptide (PDB 4QH7) is shown in the LC8 binding groove with positions of lysine 9 and histidine 68 labeled. Ana2 QT1 and Chica QT3p peptides have highly similar motifs but very different affinities; the primary sequence differences are shown in red rectangles. (d) The *B*-factors of backbone QT1p atoms (orange), side chain QT1p atoms (red), backbone QT4p atoms (green), and side chain QT4p atoms (blue) are plotted at each residue position. *B*-factors for backbone or side chain atoms were averaged using the program 'baverage' available through the CCP4 suite [160].

Discussion

Chica Motifs Span the Range of LC8 Affinity

Although LC8 sequences are highly conserved (94% identity between *Drosophila* and human, and 100% between rat/mice and human), the LC8 recognition motifs carried on LC8 partners are rather variable in sequence except for the TQT consensus triplet. Equally variable is motif affinity for LC8. Full answers to pressing questions of LC8 biological function remain elusive, and the adaptive function of LC8 motif diversity is unknown. One approach is to ask how, at the molecular level, LC8 accommodates so many motif sequences and apply the answers to development of predictive principles for identifying new LC8 partners. In experiments directed toward these general goals, we analyze the structural biology of LC8-Chica interactions to elucidate features that guide LC8-motif recognition.

Among multiple LC8-binding partners (Figure 3.7), some have TQT motifs of varying LC8 affinity, and a few have TQT motifs that do not bind LC8 at all. The protein studied here, human Chica, contains four potential LC8 recognition motifs, out of which three bind LC8. Spectroscopic measurements confirm that $Chica_{410-478}$ is intrinsically disordered as predicted from sequence (Figure 3.3) and calorimetric studies show that it binds three LC8 molecules (Figure 3.4 and Table 3.1), as expected from previous reports [62]. Affinities of the TQT motifs, determined by calorimetric experiments with synthetic 14-amino acid peptides corresponding to each of the four motifs, vary from 0.4 mM for QT1p to immeasurably weak for QT3p (Table 3.1). In this single protein, the affinities of LC8-binding sequences span the range of LC8 recognition motifs, from 0.16 mM for Nek9 [31] to 42.7 mM for Pak1 [32]. Thus Chica is an ideal model system for studying the molecular bases of TQT motif recognition by LC8.

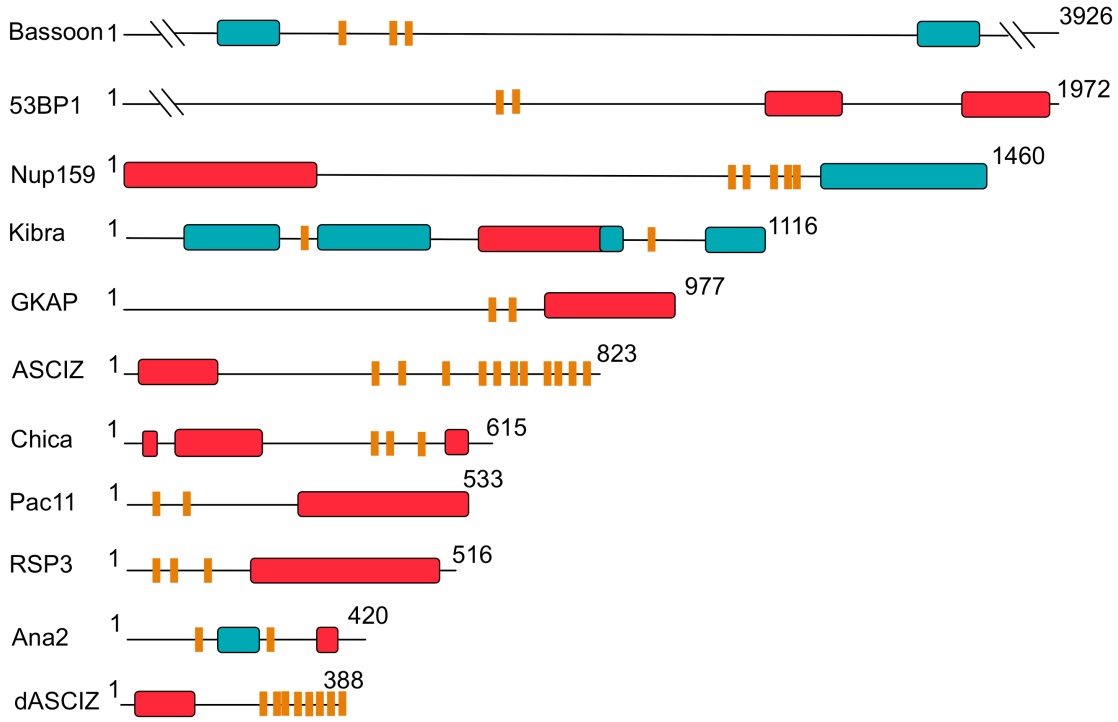


Figure 3.7. LC8 binding partners have multiple recognition motifs in intrinsically disordered regions.

Sequence-based predictions of order (red boxes), disorder (black lines), coiled-coil (blue boxes) and LC8 binding motifs (orange bars) are shown for residues 1000–2500 of Bassoon, a protein involved in the organization of the cytomatrix at the nerve terminals active zone; residues 500-1972 of p53 binding protein 1 (53BP1), a protein which activates p53-dependent gene transcription; yeast nucleoporin, Nup159, a protein in the nuclear pore complex; Kibra, a protein involved in tissue homeostatis and regulation of organ size; guanylate kinase-associated protein (GKAP), a protein that orchestrates protein remodeling at synapses; DNA damage repair protein and transcription factor ATM-Substrate Chk-Interacting Zinc-Finger (ASCIZ); the intermediate chain (Pac11) subunit of the yeast cytoplasmic dynein molecular motor; Chica, a spindle-associated adaptor protein; the flagellar radial spoke protein 3, RSP3; anastral spindle-2 (Ana2), a protein involved in centriole duplication, and the Drosophila homologue of ASCIZ (dASCIZ). Sequence predictions of order and disorder were obtained using the program PSIPRED [157], where our criteria for structure was based on >10% probability of predicted structure in a 50+ amino acid stretch, and coiled-coils were predicted using the program Paracoil2 [71], where predicted coiled-coils with a p-value < 0.025 were considered significant. Putative LC8 binding motifs are based on the following references: bassoon [64], 53BP1 [4], Nup159 [36], Kibra [66], GKAP [28], ASCIZ [69], IC [161], Chica [62], RSP3 [67], Ana2 [34], and dASCIZ[60].

Why Multiple Motifs?

While it is not clear at this stage why some LC8 partner proteins are multivalent, it is clear that they are numerous and biologically diverse, and that their LC8-recognition sites are present exclusively in intrinsically disordered segments of the partner (Figure 3.7). For some LC8 partners with multiple binding sites, various functions of the multiplicity have been proposed. In Nup159, the multiple sites act cooperatively so that binding of the first LC8 enhances binding of the next[36], and apparently contributes to the stability of the complex. Electron microscopic studies of nuclear pore assembly suggest that incorporation of Nup159 into the nuclear pore does not occur until all LC8 sites are filled [68, 162]. In ASCIZ, the multiple motifs could act as a sensor for LC8 cellular concentration [56]. ASCIZ is an LC8 transcription factor, and it is possible that the level of bound LC8 could modulate its LC8-dependent transcription activity. In Ana2, binding of two LC8 dimers promotes self-aggregation of Ana2 into a tetramer, possibly by aiding formation of a coiled-coil region located between the two binding sites [34]. In Pac11, two LC8 sites flank a nascent helix which forms a self-associated helix/helix interface upon LC8 binding [163].

In Chica, the lack of LC8 binding to QT3 is consistent with its lower conservation. Out of the four TQT motifs, QT3 is least conserved: in one species (*Propithecus coquereli)*, it has a mutant TQT sequence and in another (*Eptesicus fuscus)* it is missing entirely (Figure 3.2). The stability of the LC8/Chica$_{410\text{-}478}$ complex containing three LC8 molecules (average binding affinity) is the same as for LC8/QT1p, suggesting that the role of the other motifs is not for forming a stable complex. One possibility for the Chica/LC8 assembly is that QT1/LC8 binding initiates the process and later QT2/LC8 and QT4/LC8 binding promote self-association, perhaps in QT3 and the linker separating QT2 and QT4 similar to that observed with Ana2 and Pac11. This proposal is consistent with secondary structure propensity in the linker separating QT3 and QT4 in free Chica indicated by secondary chemical shifts and restricted dynamics (Figure 3.3c,d).

The Crystal Structures of LC8/Chica Motifs Distinguish Two Regions of Interaction

In LC8-QT1p and LC8-QT4p, the TQT residues of the peptide interact with the same LC8 atoms in a network of polar interactions involving buried atoms in both backbone and sidechains, as shown in Figures 3.5b and 3.5c. LC8 residues in the TQT network are 62-64 at the N-terminus of strand b3 and 35’ and 36’ in the a2 helix of the other LC8 monomeric unit. Three-dimensional positions of every atom in this region of the LC8/motif binding site are highly overlapping in both complexes, with an RMS deviation of 0.25 Å.

In non-TQT regions of LC8/motif binding sites, the correlation of motif position with affinity is highly complex. For example, the main chain conformation of QT4p (Figure 3.6b, green) is different from QT1p (orange) at T_{-3}, resulting in QT4p H-bonding geometry that appears to be less favorable (pink arrows). This is expected to lower affinity, but the actual affinity of QT4 relative to QT1 peptides cannot be ascribed to this effect alone, or even primarily, because there are many other amino acid differences at non-TQT positions, some of which lead to the degree of disorder of positions -6 and -7 in QT4. As a second example of the complexity of correlating affinity with sequence, the sequences of QT2 and QT4 are different at every position outside of the TQT (Figure 3.2), but they bind LC8 with a similar affinity (Table 3.1).

Similarly, it is tempting to expect that QT2 should have lower affinity than QT1, as QT2 lacks the stabilizing electrostatic interactions at position 2 (Glu–>Ser) and the long hydrophobic sidechain at position -5 (Ile–>Ser). However, many other non-TQT positions are different as well, and how the contributions of amino acids at individual positions sum to give the overall affinity is not clear. That is, an *a priori* sequence-based prediction of relative affinity of two motifs, or even prediction of whether a motif binds at all, is not feasible. A further illustration of this complexity comes from a comparison of Chica QT3 which essentially does not bind LC8, to Ana2 QT1 which has a binding affinity of 1.1 mM [34]. Between Chica QT3 and Ana2 QT1, the most notable sequence differences are at position 2 (Asp–>Ala), and at position -5 (Ile–>Ala) (Figure 3.6c). An Asp at position 2 makes a favorable interaction with Lys9 in LC8 and a long hydrophobic sidechain is more stabilizing than an Ala at position -5. The combined effect of the two replacements is reasonably assumed to diminish binding. So it is tempting to reason that these two substitutions together could suggest a position 'rule' for predicting whether TQT motifs in fact bind LC8, especially since Chica QT1 and Ana2 QT1 both have Ile at position -5 and a carboxyl side chains at position 2. But, even this simple correlation does not hold up, because QT2 has Ser at both -5 and 2, and it still binds better than QT3.

In summary, analyses of Chica motif complexes with LC8, and of other LC8-peptide crystal structures (Figure 3.8), support the generalization that there are two functionally distinct parts of the LC8-motif binding site: the TQT residues are the main determinant of motif sequence specificity and the non-TQT residues modulate overall LC8-motif affinity and provide finely tuned association/dissociation constants for selection by LC8. In the TQT region, LC8/TQT interactions are invariant and highly constrained. In contrast, the LC8/non-TQT interactions are highly variable and their contributions to binding stability are complex and interdependent. We suggest the term "anchored flexibility" to capture the manner in which LC8/motif sequence

specificity resides primarily in LC8/TQT interactions and affinity modulation resides in LC8/non-TQT interactions.

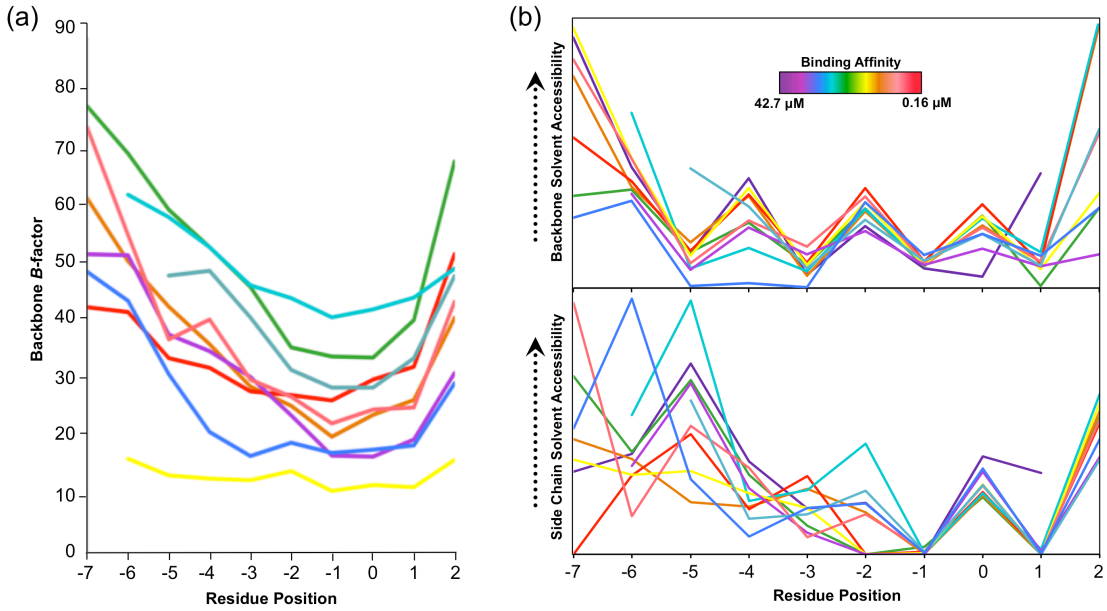


Figure 3.8. Peptides bound to LC8 exhibit B-factors and solvent accessibility patterns that suggest that the TQT anchor is a common recognition property among LC8 partners. (a) The *B*-factors for backbone atoms of peptides bound to the LC8 homodimer are plotted against residue position. *B*-factors for backbone or side chain atoms were averaged using the program 'baverage' available through the CCP4 suite [160]. Each plotted motif is colored according to its binding affinity where red is the tightest and purple is the weakest. They correspond to Nek9[31] (red), Chica QT1p (pink), Swallow[8] (orange), Ana2 QT1p[34] (yellow), nNOS[156] (green), IC[8] (cyan), Chica QT4p (teal), Myosin 5a[33] (blue), Ana2 QT2p[34] (purple), and Pak1[32] (dark purple). (b) Solvent accessibility of motif residues bound to the LC8 homodimer are plotted against residue position for both backbone atoms (top) and side chains atoms (bottom). Solvent accessibility was calculated using the webserver GetArea[164] and a Van der Waals radius of 1.4 Å. Colors are the same as in (a).

Molecular Determinants of LC8 Recognition: The "Anchored Flexibility" Model.
Flexibility of LC8 partners is inherent to the intrinsic disorder first noted as their distinguishing feature [12], and flexibility and shear movement of the LC8 binding groove were proposed as a basis for the wide variability in motif sequence and binding affinity [8, 37, 133, 134]. To the concept of groove flexibility we add the feature of a TQT 'anchor' within one binding site that in other regions is dynamic.

As the LC8/Chica motif complexes illustrate (Figure 3.5), the 'anchor' region of the LC8 binding site engages peptide TQT residues in a network of polar interactions involving buried H-bond donors and acceptors in both LC8 subunits. Such a buried network requires optimal H-bonding geometry, the distortion of which is energetically costly, especially in an apolar milieu. The TQT region of the LC8-motif complex is highly ordered and rigidly packed (anchored) compared to the other more flexible motif positions (2, -2 to -7). The plasticity of the non-TQT regions of the complex is reflected in the conformational accommodation of numerous sequence permutations with similarly variable affinities. The flexibility resides in non-TQT residues and LC8 residues that bind them, while the anchor is composed of TQT residues and LC8 residues 35', 36', and 62-64. It is likely that these features underlie the evolutionary fitness of LC8-motif interactions.

Structural parameters of the two complexes of Chica motif peptides with LC8 (Figure 3.5 and 4.6a) and other LC8-peptide complexes (Figure 3.8) support the anchored flexibility model. If TQT-LC8 interacting residues are more rigid and tightly packed than other motif residues, then they should have smaller *B*-factors and low solvent accessibility in crystal structures of LC8-peptide complexes. A comparison of the *B*-factors of LC8-QT1p and LC8-QT4p (Figure 3.6d) upholds the expected correlation. For QT1p backbone (orange) and sidechains (red), TQT atoms have the lowest *B*-factors in the peptide. Further, in LC8-QT1p and LC8-QT4p, LC8 residues 62-64 bound to TQT have the lowest *B*-factors in the complex (data not shown). In apo LC8, *B*-factors of residues 62-64 are among the lowest [37], suggesting that a high degree of order in unbound LC8 minimizes the unfavorable entropy loss associated with rigid binding of the TQT anchor. The same trend is observed in 7 other LC8-peptide complexes. TQT residues exhibit considerably lower B-factors than residues at other positions, indicating that they are more tightly packed and inaccessible to solvent (Figure 3.8a).

An analysis of the solvent accessibility profiles of 9 LC8-peptide structures reveals the expected trend and shows a similar backbone solvent accessibility profile among all peptides for residues at positions -6 to 2 (Figure 3.8b, top). The alternating high/low solvent accessibility

pattern is consistent with the elongated β-strand structure of the peptide and shows that all peptides, regardless of their various binding affinities, form a strand at these positions with the ends being more flexible. The TQT "anchor" is clearly seen in the complete burial of side chains in the LC8 binding pocket at positions -1 to 1 (Figure 3.8b, bottom). Solvent accessibility for side chains at other positions varies considerably, exemplifying the plasticity of the non-TQT regions of the peptide binding groove.

The model does not require that all non-TQT positions are equally variable, and the motif logo in Figure 3.1b demonstrates amino acid preferences at some non-TQT positions. Because the non-TQT peptide residues and associated LC8 residues are flexible in the complex, destabilizing energetic effects of an unfavorable contact introduced by a specific mutation may be readily compensated by numerous small, stabilizing conformational adjustments. Strain and unfavorable contacts are eased by relaxation processes of equilibrating conformations so that numerous sequence permutations are accommodated in the non-TQT regions of the binding site. Still, the positions adjacent to TQT (-2 and 2) are more constrained than other positions due to their proximity to TQT, and residue preference at these positions should be higher than for others, as observed especially for position 2 (logo, Figure 3.1) where the Asp side chain interacts with Lys 9 of LC8.

In summary, the anchored flexibility model explains the requirement for the TQT triplet and the ability of LC8 grooves to accommodate numerous sequence permutations outside the TQT. Simple rules of position-dependent affinity effects are, at our present level of knowledge, ineffective in predicting new additions to the LC8 interactome from motif sequence. Given this, we suggest that predictive algorithms for identifying new LC8 binding partners should be based on statistical position preferences which awaits identification of binding affinities of a large number of binding motifs.

Materials and Methods

Cloning, Protein Expression and Purification

A gene encoding mammalian Chica (FAM83D, accession number NP112181) residues 410-478 was synthesized (GenScript, Piscataway, NJ), subcloned into pET15 vector (Novagen) which added a a 17 non-native residue extension containing a hexahistidine tag, and the construct was transformed into *E. coli* BL21-DE3 cell line. *E. coli* strains were grown at 37 °C in LB or in minimal media with ^{12}C or ^{13}C glucose, and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively. Recombinant protein over-expression was induced with 0.2 mM IPTG and growth

continued at 18 °C for 12 h. Cells were harvested and purified under native conditions using Qiagen's Ni-NTA purification protocol (Qiagen). Recombinant proteins were further purified on a SuperdexTM 75 gel filtration column (GE Health). The purity of the recombinant His-tagged protein, as assessed by SDS polyacrylamide gels, was > 95%. The pure protein was stored at 4 °C, and used within 2 weeks. LC8 was overexpressed and purified as previously described [154].

Peptide Design and Synthesis

Peptides corresponding to the four putative recognition sequences, SYRKAIDAATQTEE (QT1p), SYSVSEVGTQTSI (QT2p), SYTTACAGTQTAV (QT3p), SYWSRSTTTQTDM (QT4p) were synthesized (GenScript, Piscataway, NJ). Non-native residues (Table 4.1, underlined) were added to the N-terminus of each peptide to facilitate solubility and peptide concentration determination by absorbance at 280 nm. In QT1p, the C-terminal Glu replaces a Pro residue in the wild type sequence to increase solubility. This replacement is justified as in the crystal structure the side chain of E425 in QT1p does not form a hydrogen bond with LC8, suggesting minimal contribution to binding affinity from the replacement.

Isothermal Titration Calorimetry

Binding thermodynamics of the $Chica_{410-478}$/peptide-LC8 interactions were obtained at 25°C with a VP-ITC microcalorimeter (Microcal). Chica peptides were dissolved in the reaction buffer composed of 50 mM sodium phosphate, 50 mM sodium chloride, 0.5 mM sodium azide, pH 7.5 to final concentrations of 280 - 300 μM and then added to 30 - 35 μM of LC8 in the reaction cell. For binding to $Chica_{410-478}$, LC8 at a concentration of 280 μM was titrated into 10 μM $Chica_{410-478}$ in the reaction cell. Peak areas were integrated and fitted to a single-site binding model in Origin 7.0 from which the stoichiometry (N), dissociation constant (K_d), the change in enthalpy (ΔH) and entropy (ΔS) were obtained. Reported data are the average of three independent experiments.

Circular Dichroism

CD data were collected on a Jasco720 spectropolarimeter. Three scans collected at 1 nm step resolution, and 1.0 nm bandwidth were averaged for the spectrum of $Chica_{410-478}$ at a concentration of 10 μM in buffer composed of 10 mM sodium phosphate, pH 7.3. The temperature of the experiment was maintained at 25 °C by connecting the 1 mm water-jacketed sample cell to a water bath.

NMR Experiments

NMR measurements were collected at 15°C, using 150 - 250 μM isotopically (^{15}N or ^{13}C/^{15}N) labeled Chica$_{410-478}$ in buffer composed of 10 mM sodium phosphate, 10 mM sodium chloride, 1 mM sodium azide, pH 7.0, with 50 mM arginine/glutamate (to improve solubility [165]), a protease inhibitor mixture (Roche Applied Science), and 2,2-dimethylsilapentane-5-sulfonic acid for ^{1}H chemical shifts referencing. All experiments were collected on a Bruker Avance 600 spectrometer equipped with a cryoprobe. The CA, CB, N, CO and HA chemical shifts were assigned from 2D [^{15}N,^{1}H] BEST-TROSY [165, 166], 3D- BEST-TROSY- HNCACB and HN(CO)CACB, and 3D-BEST-TROSY-HNCO, and HA(CACON)H. T_1 relaxation measurements experiments were recorded with relaxation delay times ranging from 10 to 900 ms, and the T_2 relaxation data were acquired using relaxation delays ranging from 22.72 to 318.08 ms. Curve fitting was done with the rate analysis interface of NMRview [167]. Steady-state ^{1}H-^{15}N heteronuclear NOE acquired with TROSY detection were based on described methods using 6 seconds total saturation time [168, 169]. Unless otherwise stated, all spectra were processed with Topspin™ (Bruker Biosciences) and analyzed with Sparky [170]. NMR titration experiments with LC8 were attempted, but data collection was hampered by the propensity of Chica$_{410-478}$ to aggregate during the titration.

Protein Crystallization and Structure Determination

Crystals grew within two weeks at room temperature using the hanging-drop, vapor diffusion method. For the QT1p complex, hexagonal rods were grown by mixing a solution of 0.8 mM LC8 and 2 mM QT1p in 20 mM Tris (pH 8.0) in a 1:1 (v/v) drop with a reservoir solution of 0.1 M calcium acetate, 0.1 M sodium cacodylate, 15% polyethylene glycol 8000 (pH 5.5). For the QT4p complex, rounded barrel-shaped crystals were grown using the same LC8 stock but a large excess of QT4p and a reservoir solution of 0.2 M lithium sulfate, 0.1 M Bis-Tris, 25% polyethylene glycol 3350 (pH 5.5). Crystals were transferred to a cryoprotectant consisting of reservoir solution plus 20% (v/v) glycerol, and then flash-frozen on nylon loops in liquid nitrogen.

Diffraction data were collected on beam line 5.0.2 at the Advanced Light Source, Berkeley National Labs with 180 1° oscillations for all crystals. Data were processed using Mosflm [171] and for both complexes data from two crystals were merged to improve the completeness (Table 4.2). Both crystals have one molecule in the asymmetric unit and a solvent content of 57%.

The LC8-QT1p and LC8-QT4p peptide structures were solved at resolutions of 1.3 and 1.6 Å, respectively, by molecular replacement using PHENIX AutoMR [172] with apo-LC8 (PDB 3BRI) as the search model. The resulting F_o-F_c map showed clear density for bound peptide, confirming the molecular replacement solution. Models for bound QT1p and QT4p were constructed using the LC8-Swa complex (PDB 3E2B) as a template. With 10% of the data set aside for cross-validation, the models were subjected to iterative refinement using Coot [173] and phenix.refine [172], including TLS refinement [174]; ordered water molecules were added using the criteria of having reasonable hydrogen bonding partners and a peak in the $2F_o$-F_c electron density map of at least 1σ. For the bound peptides, side chain atoms were included for all of the residues that were modeled, even if there was minimal or no density for some of the side chain atoms.

PDB and BMRB Accession Codes

The coordinates of LC8-QT1p and LC8-QT4p have been deposited in the RCSB protein data bank under accession codes 5E0L and 5E0M, respectively. The chemical shifts for $Chica_{410-478}$ have been deposited in the Biological Magnetic Resonance Data Bank under accession code 26684.

Acknowledgements

This work is supported by National Institutes of Health Grant GM 084276 to EB. We acknowledge the support of the protein core facility in the OSU Environmental Health Sciences Center (NIH/NIEHS 00210), and access to the Research Infrastructure activity in the 7th Framework Programme of the EC (Project number: 261863, Bio-NMR) (Frankfurt, Germany). The authors also wish to thank Prof. Clare Woodward and the Barbar lab for valuable discussions.

Chapter 4

Multivalency Regulates Activity in an Intrinsically Disordered Transcription Factor

Sarah A. Clark, Janette B. Myers, Radovan Fiala, Jirka Novacek, Grant Pearce, Ashleigh King, Jörg Heierhorst, Steve L. Reichow, and Elisar Barbar

In preparation for submission to *eLife*, February 2018

Abstract

The transcription factor ASCIZ (ATMIN, ZNF822) has an unusually high number of recognition motifs specific to its gene product, the hub protein LC8 (DYNLL1). Using a combination of biophysical methods, structural analysis by NMR and electron microscopy, and cellular transcription assays, we developed a model that describes how the concerted action of intrinsic disorder and multiple LC8 binding events tune LC8 transcription. We demonstrate that the primarily disordered ASCIZ binds multiple copies of the LC8 dimer to form a dynamic ensemble of low occupancy complexes with a gradient of transcriptional activity that is inversely proportional to LC8 occupancy. A preference for low occupancy complexes at saturating concentrations of LC8 is observed with both human and *Drosophila* ASCIZ, suggesting that a dynamic equilibrium and negative cooperativity are conserved features of ASCIZ-LC8 interactions. The prevalence of intrinsic disorder and multivalency among transcription factors suggests that formation of multiple complexes with variable occupancies could be a widespread mechanism for tuning transcriptional regulation.

Introduction

Common mechanisms for regulating the activity of transcription factors include post-translational modifications or cooperativity among multiple activators and repressors[175]. Some transcription factors contain multiple regulatory sites for either post-translational modifications[176] or binding partners[177], and their activity is thus tuned by the combined action of these components. Recent studies have revealed a high degree of intrinsic disorder in transcription factors, indicating that the inherent dynamical behavior harbored by these structures is critical for these regulatory events to take place [87, 178]. Our developing understanding suggests that the intrinsically disordered domains in transcription factors may provide a multivalent platform for the recruitment of regulatory binding partners [179, 180]. While the functional consequence of multivalent binding to an intrinsically disordered region has been described for a few transcription systems [181, 182], it remains unclear for the vast majority of cases.

ASCIZ (ATMIN-Substrate Chk-Interacting Zn^{2+} finger) is an 88 kDa protein that has recently been identified as a transcription factor for the hub protein, LC8 (dynein light chain 8)[56]. Mice with mutations in ASCIZ that prevent LC8 transcription die in late embryogenesis and exhibit serious developmental defects in kidneys and lungs [57-59]. *Drosophila* ASCIZ knockouts die in early embryogenesis and localized knockdowns using RNAi show mitotic defects [60]. Mutant phenotypes in *Drosophila*, developing mouse B lymphocytes, and cultured cells are rescued by ectopic overexpression of LC8, demonstrating that the observed defects of ASCIZ knockouts are due to ASCIZ regulation of LC8 expression [58, 60, 183]. In addition, it has recently been shown that conditional knockout of LC8 almost perfectly copies the corresponding phenotypes of ASCIZ knockouts in mouse B cell development and B cell lymphomagenesis [184, 185].

LC8 is a highly conserved 20.6 kDa protein homodimer (10.3 kDa monomer) that acts as a chaperone by facilitating self-association of its primarily disordered partners [12, 74, 151] (Figure 4.1a). LC8 binding is associated with a range of cellular processes, from cell division to apoptosis, underscoring LC8's essential role as a regulatory hub [52, 62]. LC8 preferentially binds to a 10-amino acid motif in intrinsically disordered protein segments (IDPs) containing highly conserved TQT residues at positions 7-9 [12, 26] (Figure 4.1b). In complex with LC8, the otherwise intrinsically disordered motif adopts a beta strand conformation [8, 9] (Figure 4.1a,b). Analysis of the 11 crystal structures of LC8 bound to short peptides containing the motif explains why the TQT residues are essential for binding; the Q is involved in interactions with both LC8 monomers, while both T's are fully buried and thus evolutionarily constrained. In these

interactions, TQT acts as the motif anchor while the other seven highly variable motif residues modulate affinity, as described in the *anchored flexibility model* of LC8 motif recognition [30]. LC8 binds one motif in each of its two symmetrical binding grooves (Figure 4.1a), creating an IDP duplex that serves as a bivalent scaffold. This scaffold aids in higher order complex assembly by promoting binding of other proteins, including additional LC8 dimers (Figure 4.1c), and enhancing self-association and oligomerization processes that often involve coiled-coil formation [43, 44]. In recent years, the number of experimentally characterized LC8 partners has risen to more than 40, and prediction methods indicate that dozens more may specifically bind LC8 [26]. Gaining insight into how LC8 interacts with partner proteins, and how LC8 levels in the cell are balanced, is therefore paramount to understanding the regulation of many cellular processes.

A distinctive feature of ASCIZ is the high number of LC8 recognition motifs within its C-terminal domain. Although some LC8 partners have multiple recognition motifs[38, 62, 64], human ASCIZ contains 11 functional LC8 binding sites [69], the most of any partner protein identified to date. This enrichment in LC8-binding sites is found to be conserved throughout the animal kingdom, underscoring the importance of multiple motifs in ASCIZ function (Figure 4.1d). *In vivo* transcription assays demonstrate that ASCIZ regulates LC8 transcription via a system of negative autoregulation, for which the mechanism is not well understood. Disruption of the ASCIZ/LC8 interaction via mutation of the TQT sites results in an increased level of LC8 transcription, while overexpression of LC8 decreases transcription [56]. This observation led to the hypothesis that ASCIZ acts as a sensor for cellular LC8 and regulates LC8 transcription levels according to cellular needs [56]. As LC8 expression levels vary among tissue types [186] and LC8 overexpression enhances the survival and proliferation of breast cancer cells in culture [53], regulation of LC8 levels is critical for cellular health and homeostasis. However, while high levels of LC8 inhibit ASCIZ transcriptional activity[56], it is not known how this activity is controlled at the molecular level nor the reason for multiple binding sites.

In this work, we use a combination of biophysical, structural, and molecular biology tools to explore the relationship between ASCIZ multivalency and LC8 transcription. We show that human and *Drosophila* ASCIZ (dASCIZ) bind to multiple LC8 dimers simultaneously to form a dynamic mixture of complexes, of which a low occupancy intermediate is most stable. Transcriptional assays with human ASCIZ demonstrate that the number of bound LC8 dimers modulates ASCIZ transcription, creating a gradient of activity. These observations support a novel model of autoregulation, whereby ASCIZ engages in a dynamic equilibrium of multivalent interactions that tune the level of ASCIZ transcriptional activity. The prevalence of intrinsic

disorder and multivalency among transcription factors suggests that formation of multiple complexes with variable occupancies could be a wide-spread mechanism for tuning transcriptional regulation.

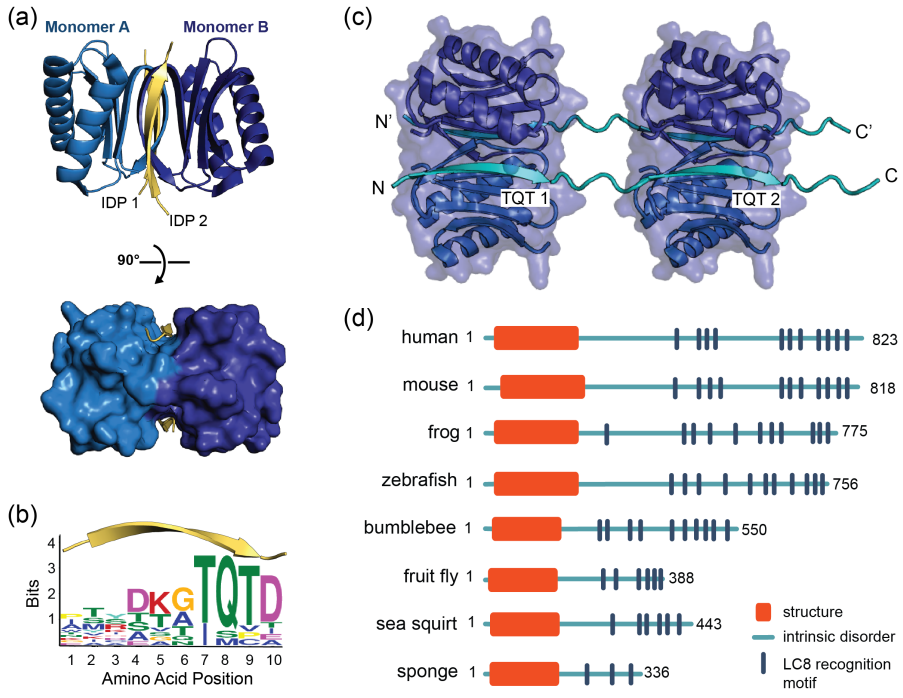


Figure 4.1. LC8 dimerizes its protein partners.
(a) Ribbon diagram of dimeric LC8, where each monomer is colored a different shade of blue and bound to a representative peptide in yellow (PDB 5E0L). Sidechain residues corresponding to the conserved TQT motif are displayed. A single LC8 dimer binds two peptides, one on each side of the dimerization interface and arranged in a parallel fashion. (b) A sequence logo of LC8 binding motifs derived from sequences of the motifs in the 11 crystal structures reported for LC8/peptide complexes. Height of amino acids indicates their relative frequency at that position. (c) A crystal structure of two LC8 dimers bound to two copies of the same intrinsically disordered chains illustrates how LC8 can bind to multivalent partner proteins (PDB 3GLW). (d) Sequence-based predictions of order (red boxes), disorder (blue lines), and LC8 binding motifs (dark blue bars) are shown for 10 ASCIZ proteins from the animal kingdom. ASCIZ proteins for different species were identified from a BLAST search [187] against the human protein. Sequence predictions of order and disorder were obtained with PSIPRED [70], where a criteria for order based on >10% probability of predicted structure in a 50+ amino acid stretch. LC8 binding sites for human ASCIZ and *Drosophila* ASCIZ were obtained from the literature [60, 69]. Putative LC8 binding sites for the other species were identified based on the presence of TQT residues.

Results

Unbound ASCIZ Is a Primarily Disordered Monomer

The 45 kDa *Drosophila* ASCIZ protein, dASCIZ, has a sequence predicted to contain four Zn-finger motifs at the N-terminus followed by a 243 amino acid long region of intrinsic disorder. The disordered region has six putative LC8 recognition sites identified by a canonical TQT motif (Figure 4.2a, dark blue bars): QT1 (residues 251-262), QT2 (274-285), QT4 (323-334), QT5 (340-351), QT6 (354-365), and QT7 (374-385). QT3 (285-296) lacks the TQT residues but is identified experimentally as an LC8 recognition site in this work (below). Purification of full-length dASCIZ is impeded by poor expression levels and insolubility, and therefore we designed and produced constructs corresponding to the zinc finger domain, ZnF (residues 1-156, red bar Figure 4.2a) and the LC8-binding domain, LBD (named dLBD in *Drosophila*, residues 241-388, Figure 4.2a).

Sedimentation velocity analysis of the ZnF domain and dLBD (Figure 4.2b) indicates that each is a monomer in solution with molecular weights of 17.4 kDa and 18 kDa, respectively (theoretical MW 17.6 kDa and 17 kDa). The CD spectrum of ZnF shows a large negative ellipticity at 208 nm and a small negative ellipticity at 222 nm (Figure 3.2c, red) indicative of a mix of alpha helix and loop structure, consistent with CD spectra of other ZnF proteins [188]. The dLBD CD spectrum has a large negative ellipticity at 200 nm, indicating that it is primarily disordered (Figure 4.2c, blue). From 5D NMR experiments, backbone assignments for 90% of the 148 residues in dLBD were obtained (Figure 4.2d). A high level of disorder in dLBD is revealed by the limited amide proton chemical shift dispersion in ^{15}N HSQC spectra (Figure 4.2d), and a lack of secondary structure preference is further supported by small ΔCα-ΔCβ chemical shift differences from random coil values (Figure 4.2e). Together these data demonstrate that dASCIZ contains an N-terminal structured domain as well as a long intrinsically disordered domain, and constructs of each domain are monomeric in solution.

Local dynamics of the LC8 binding domain were assessed by NMR measurement of the ^{15}N longitudinal (R_1), transverse (R_2) relaxation, and ^{1}H-^{15}N heteronuclear NOEs. R_2/R_1 values range from 0.4 to 2.4 with an average of 1.1 (Figure 4.2f). Relatively higher R_2/R_1 values for residues 321-363 suggest motional restriction in this region, indicating a tendency to form transient structure. Heteronuclear NOE values measured at 10°C are very low overall, with values ranging from -0.1 to 0.3, but are also slightly higher for residues 321-363 (Figure 4.2g). Together, the R_2/R_1 and heteronuclear NOE values indicate that dLBD is highly flexible with transient ordered structure in its C-terminal half.

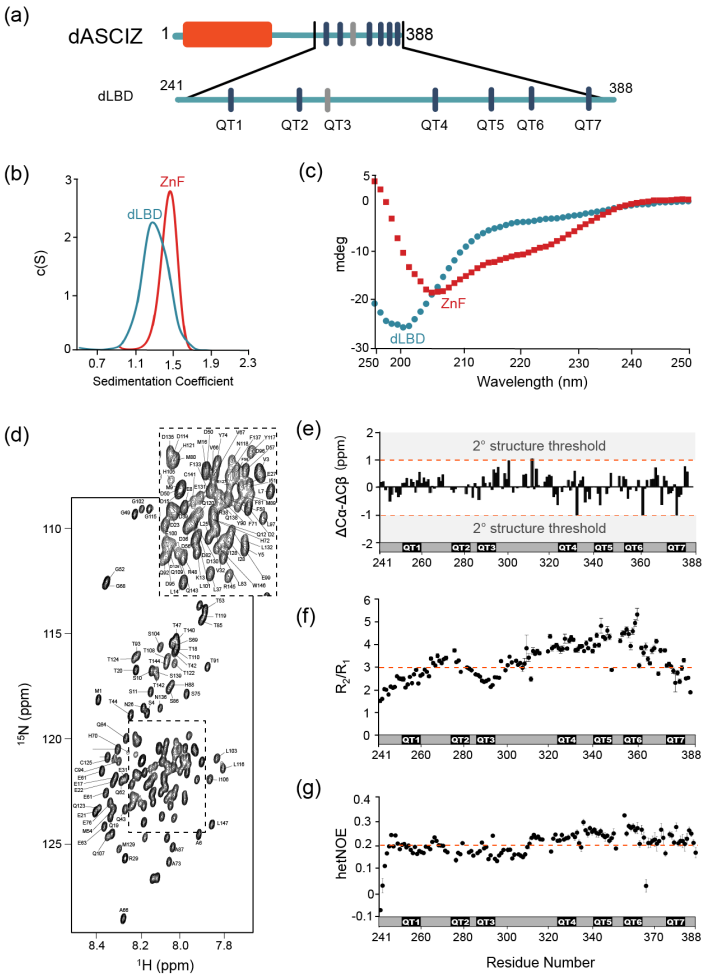


Figure 4.2. The domain structure of ASCIZ.
(a) Domain structure of dASCIZ, showing the ZnF domain (red) and 7 LC8 binding motifs in its C-terminal domain (blue). Dark blue bars indicate predicted TQT motifs and gray bars indicate the TMT motif (QT3) identified in this study. (b) Sedimentation velocity analysis of the ZnF domain (red), collected at 10 °C, and the dLBD (blue), collected at 25 °C. (c) Far UV CD spectrum of the ZnF domain (red squares) and the dLBD (blue circles), both collected at 10 °C. (d) [^{15}N–^{1}H]-BEST-TROSY spectrum at 850 MHz showing backbone assignments for 133 non-proline residues. Unassigned peaks correspond to the additional N-terminal residues from the expression vector. The spectrum was recorded at 10 °C. (e) A plot of secondary chemical shift differences versus residue number. ΔCα and ΔCβ values were calculated by subtracting the random coil chemical shift from the experimentally determined chemical shift value [189]. ΔCα − ΔCβ values > ± 1.0 ppm are considered to be significant. (f) Plots of R_2/R_1 and (g) heteronuclear NOE values measured at 10 °C per residue. A dotted line is placed at the average value of both plots to aid in visualization. (b-c) Segments corresponding to LC8 recognition motifs, QT1, QT2, QT3, QT4, QT5, QT6, and QT7 are shown.

The presence of transient structure at the C-terminus was validated by generating shorter constructs of the dLBD that include the first three (QT1-3), middle three (QT2-4 and QT4-6), and last four binding sites (QT4-7) (Figure 4.3a). All constructs are of a similar size, varying from 68-84 residues in length. Circular dichroism demonstrates that the two C-terminal constructs, QT4-6 and QT4-7, are slightly more ordered than N-terminal constructs, QT1-3 and QT2-4 (SI Figure 4.1). Size-exclusion chromatography supports this result, as the QT4-6 and QT4-7 constructs elute at a later time point, indicating that they are more compact than the N-terminal constructs (SI Figure 4.1).

Identification and Binding Affinities of LC8 Recognition Motifs in ASCIZ

Isothermal titration calorimetry (ITC) experiments were performed to verify that all six putative recognition motifs in dASCIZ bind LC8 and to estimate the overall binding affinity of dLBD for LC8. Surprisingly, we found that full length dLBD binds LC8 with a dLBD:LC8 stoichiometry of 1:7 (two chains of dLBD for 7 LC8 dimers) and an overall K_d of 1.4 μM (Figure 4.3a, Table 1), indicating the presence of one non-TQT LC8-binding site. A plausible candidate was identified by examination of the dASCIZ sequence, namely a TMT motif corresponding to residues 285-296 (designated QT3 in Figure 4.3a). To confirm the functionality of this motif, ITC binding of LC8 was measured for constructs QT1-3, QT2-4, QT4-6, and QT4-7 (Figure 4.3b). QT1-3 and QT2-4 contain the TMT binding motif, and both bind LC8 with a stoichiometry of 3, demonstrating that this TMT motif is the seventh LC8 recognition site (Figure 4.3b, Table 4.1). Each of the other two constructs bind LC8 with the stoichiometry expected from the number of TQT binding motifs. Interestingly, the construct containing recognition sites 4-6 (QT4-6) binds LC8 with a K_d of 1.0 μM, a slightly higher affinity than the full dLBD construct. QT4-7, on the other hand, binds LC8 with a 1.6 μM affinity and is slightly more entropically disfavored than either QT4-6 or dLBD (Table 4.1). The K_d values of QT1-3, QT2-4, and QT4-7, are 2.4 μM, 4.1 μM, and 1.6 μM, respectively.

Given the small differences in binding affinity among the shorter dLBD constructs, we sought to determine whether any individual site binds LC8 with higher affinity than all others. A series of 14-amino acid peptides were synthesized, each corresponding to one of the 7 recognition motifs, and their LC8 affinity was measured (Figure 4.3c, Table 4.2). The K_d values for all peptides are surprisingly weak. A QT3p-LC8 interaction is not detected at a 30 μM LC8 concentration. Intriguingly, the three sites that bind most tightly when combined in the QT4-6

construct are among the three weakest as individual peptides. These results indicate that, for all 7 sites, *in vitro* LC8 binding affinity is enhanced by the presence of neighboring motifs.

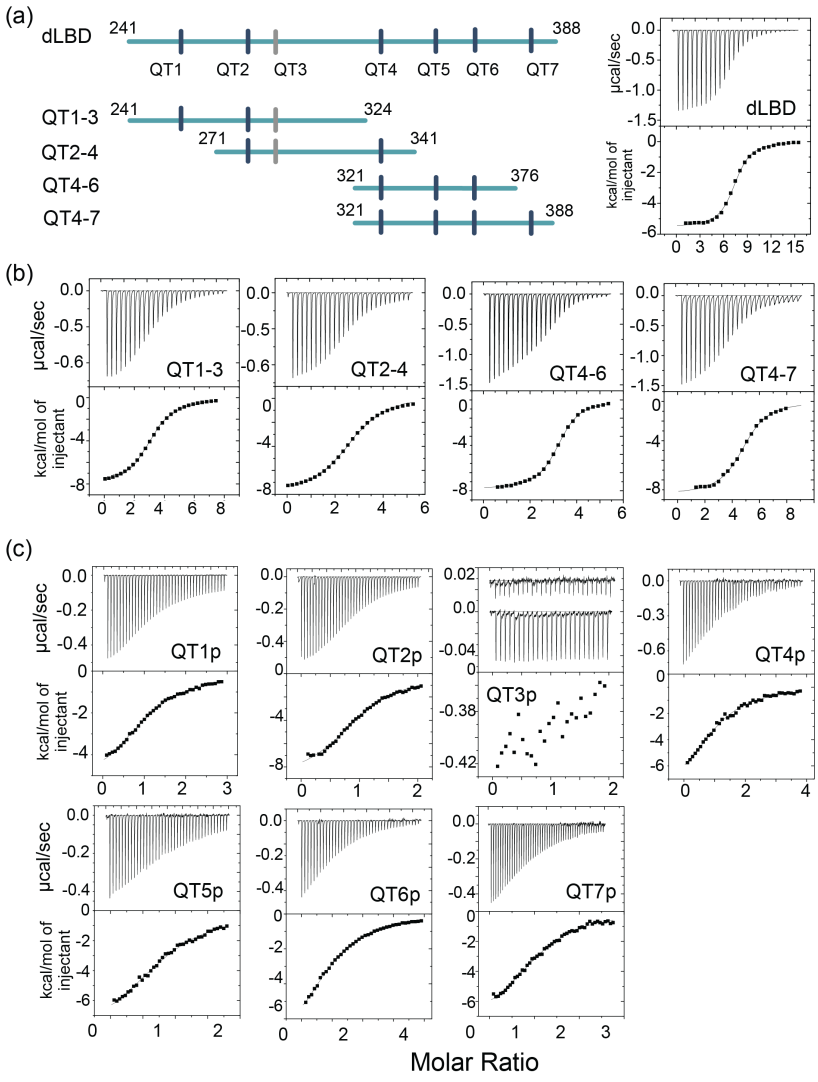


Figure 4.3. LC8-dLBD interactions monitored by ITC.
(a) Construct schematics of the dLBD, QT1-3, QT2-4, QT4-6, and QT4-7 are shown, along with the locations of each TQT motif. A representative isothermal titration plot of LC8 with the dLBD is shown right. (b) Representative isothermal titration plots of LC8 with constructs corresponding to QT1-3, QT2-4, QT4-6, and QT4-7. (b) Representative isothermal titration plots of LC8 with peptides corresponding to QT1p, QT2p, QT3p, QT4p, QT5p, QT6p, and QT7p. The interaction with QT3p is too weak to fit to the binding model, so interactions with buffer are plotted on top to show the difference between QT3p interaction with LC8 and buffer. Higher concentrations were not possible due to poor solubility of the QT3p. Data were fit to a single site binding model using Origin software.

Construct	N	Overall K_d (μM)	Overall ΔH (kcal/mol)	Overall TΔS (kcal/mol)	Overall ΔG (kcal/mol)
dLBD	7.3	1.4 ± 0.1	-5.3 ± 0.2	2.7 ± 0.4	-8.0 ± 0.4
QT1-3 (241-324)	3.2	2.4 ± 0.1	-8.1 ± 0.4	-0.4 ± 0.6	-7.7 ± 0.4
QT2-4 (271-341)	2.7	4.1 ± 0.2	-7.9 ± 0.4	-1.6 ± 0.5	-6.3 ± 0.3
QT4-6 (321-376)	3.0	1.0 ± 0.1	-10.0 ± 0.5	-1.8 ± 0.6	-8.2 ± 0.4
QT4-7 (321-388)	4.0	1.6 ± 0.4	-10.0 ± 0.5	-2.1 ± 0.6	-7.9 ± 0.4

Table 4.1. Thermodynamic parameters of dASCIZ-LC8 interactions.

Peptide	Peptide Sequence[a,b]	N	K_d (μM)	ΔH (kcal/mol)	TΔS (kcal/mol)	ΔG (kcal/mol)
QT1p	ymssQKLDMETQTEe	1.1	14 ± 3.5	-5.9 ± 0.3	0.7 ± 0.4	-6.6 ± 0.3
QT2p	ylapLLRDIETQTPd	1.0	7 ± 0.4	-9.2 ± 0.5	-2.2 ± 0.6	-7.0 ± 0.4
QT3p	ytpdTRGDIGTMTDd	---	weak	-----	-----	-----
QT4p	dlqTSAHMYTQTCd	1.1	15 ± 0.8	-8.7 ± 0.4	-2.1 ± 0.5	-6.6 ± 0.3
QT5p	eelGLSHIQTQTHw	0.9	11 ± 0.6	-8.8 ± 0.4	-2.0 ± 0.5	-6.8 ± 0.3
QT6p	wpdgLYNTQHTQTCd	1.1	20 ± 1.0	-8.6 ± 0.4	-2.2 ± 0.5	-6.4 ± 0.3
QT7p	epdNFQSTCTQTRw	1.1	10 ± 0.5	-7.8 ± 0.4	-0.9 ± 0.5	-6.9 ± 0.3

Table 4.2. Thermodynamic parameters of peptide-LC8 interactions.

[a] the 10-amino acid LC8 binding motif is capitalized

[b] non-native residues added to the N-terminus of each peptide to increase solubility or improve concentration determination are underlined

dLBD and LC8 Form a Dynamic, Low-Occupancy Complex

Since ITC experiments show that full length dLBD binds 7 LC8 dimers, we sought to establish the size of the dLBD:LC8 complex at varying LC8 concentrations by analytical ultracentrifugation. The molecular masses of unbound LC8 and dLBD determined from sedimentation velocity are 23 and 18 kDa, respectively, closely matching their theoretical masses of 24 kDa for LC8 dimer and 17 kDa for dLBD monomer (Figure 4.4a). ASCIZ dLBD titrated with increasing concentrations of LC8 forms complexes that were assessed for dLBD:LC8 molar ratios of 1:1, 1:3, 1:6, and 1:10 (Figure 4.4a). A plot of the continuous size distribution, c(S), vs. sedimentation coefficient shows that titration at sub-saturating concentrations of LC8 results in a broad peak that is likely an equilibrium mixture of complexes with varying LC8 occupancy in exchange with each other and with free dLBD. At a saturating concentration of LC8 (1:10 ratio), a high occupancy complex (197 kDa) is clearly evident, but contrary to expectations, an intermediate complex (approx. 114 kDa) is even more highly populated. The molecular weight of this intermediate corresponds roughly to a 1:3 complex and is approximately twice the intensity of the high occupancy 1:7 complex, a preponderance that persists even at a molar ratio of 1:25 (data not shown), implying that the low occupancy complex is quite stable.

In summary, while the high occupancy complex evident in AUC profiles is consistent with the ITC results, this complex is in equilibrium with many smaller sub-saturated species, the most highly populated of which is an approximately 1:3 complex of dLBD:LC8. The 1:3 complex is favored, relative to the 1:7 complex even in samples having a large excess of LC8.

The presence of stable, low occupancy intermediate sized complexes is supported by small angle X-ray scattering (SAXS) data collected on the dLBD:LC8 complex. A sample composed of dLBD and a large excess of LC8 was injected into an in-line size-exclusion chromatography system and X-ray scattering data were collected for the largest peak. Guinier analysis of the data indicates a monodisperse sample suitable for further analysis (SI Figure 4.2a). Analysis of the distance distribution function suggests a moderately compact structure for dLBD:LC8 complexes, with D_{max} = 240 Å, and a molecular weight of roughly 110 kDa, consistent with a dLBD:LC8 dimer molecular ratio of 2:3 (SI Figure 4.2b), and with the 114 kDa complex observed by AUC. Additionally, a Kratky plot of the scattering data indicates that a dLBD :LC8 sample is a mix of globular domains and intrinsic disorder, consistent with our expectations of low occupancy complexes (SI Figure 4.2c).

Native gel titration of dLBD with LC8 corroborates the presence of a mixture of complexes (Figure 4.4b). When unbound, dLBD and LC8 each run as a single band. When LC8 is

added to dLBD at a 1:1 molar ratio, free LC8 disappears, and the complexes formed migrate above LC8, as a diffuse band most readily explained as duplexes of 2 dLBD chains and one or two LC8 dimers (some complexes have 2 LC8 since some free dLBD persists). As the molar ratio is increased, the diffuse upper band becomes a dark smear, free LC8 accumulates in a pronounced dark band, and the free dLBD band disappears. We think the most likely explanation is that the decreasing mobility of the upper smear indicates increasing sizes of the complexes formed. Unbound LC8 is clearly visible at ratios ≥1:4, indicating the presence of a pool of free LC8 well below LC8 saturation of dLBD.

We similarly assessed the gel mobility of complexes formed by LC8 and the LC8-binding domain of human ASCIZ (LBD), which contains 11 TQT motifs [69](Figure 4.1d). Very similar gel behavior is observed for human LBD and dLBD, although the gel mobility of LBD bands is much lower than the mobility of dLBD bands due to molecular sieving of the much larger human LBD (53 kDa) compared to dLBD (17 kDa); both are disordered and highly extended. Molecular sieving is the dominate effect on LBD mobility as both LBD and dLBD have a pI ~4. As LC8 is added to LBD, free LC8 first disappears then returns as a very dark band at high molar ratios. At the same time, the free LBD band disappears and a lower mobility smear becomes increasingly evident. We conclude that the most likely explanation of LBD titration behavior is the same as for dLBD gels. Decreasing mobility of the upper smear indicates increasing sizes of the complexes formed; heavy bands migrating the same as free LC8 at ratios ≥1:4 indicate the presence of a pool of free LC8 well below LC8 saturation of dLBD. Together the gel titration data for dLBD and LBD suggest that LBD:LC8 complexes form a dynamic ensemble with varying levels of LC8 occupancy in which lower occupancy forms are favored.

Titration of the LBD with LC8 by size-exclusion chromatography supports our native gel interpretation, by demonstrating the formation of a low molecular weight (LMW) and high molecular weight (HMW) complexes (Figure 4.4d). The amount of LBD is held constant, and increases in peak intensity are therefore due to the addition of LC8. Both LMW and HMW complexes form at the first titration point and increase in population as more LC8 is added. A discernable peak corresponding to excess LC8 appears at a ratio of 1:7 and a stable, low occupancy intermediate is in slow exchange with a high occupancy complex, in a manner similar to the dLBD:LC8 complex. As both *Drosophila* and human LBD exhibit this dynamic behavior, it may be a conserved feature of the ASCIZ:LC8 interaction.

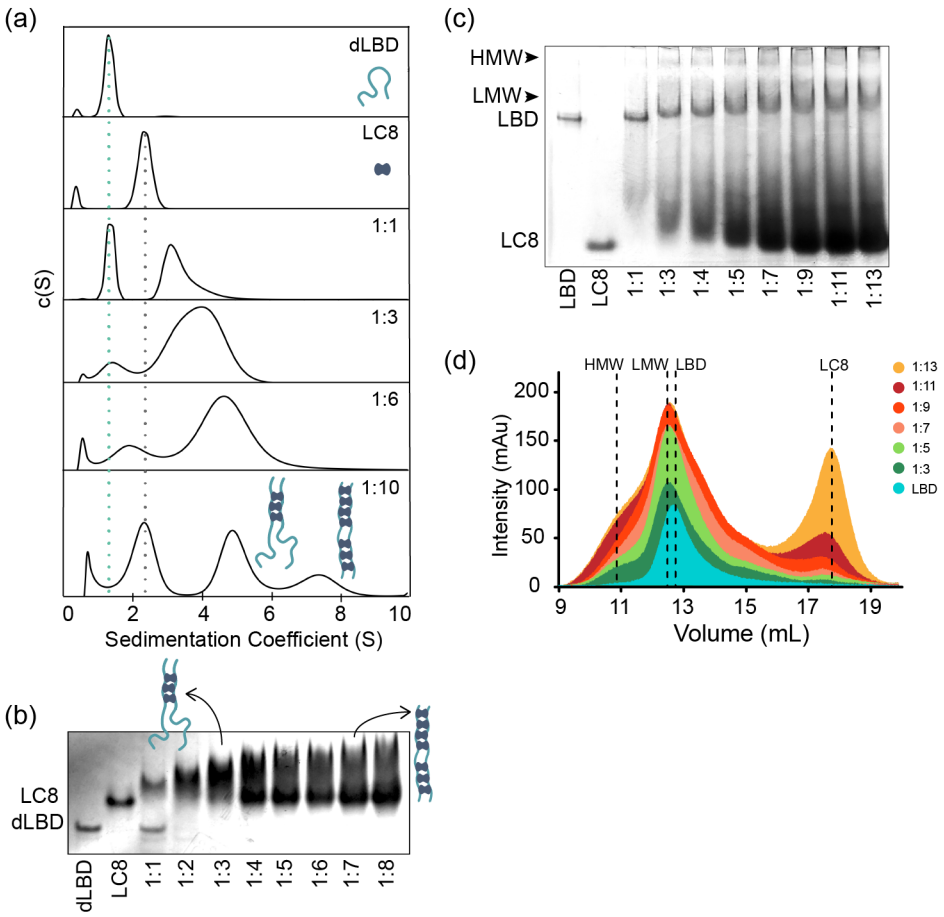


Figure 4.4. ASCIZ and LC8 form a dynamic complex with a low occupancy intermediate. (a) Representative *c(S)* distributions obtained by sedimentation velocity are shown for the dLBD, LC8, and increasing molar ratios of the dLBD: 1 LC8; 1:1, 1:3, 1:6, and 1:10. The molecular masses of unbound LC8 and dLBD determined from sedimentation velocity are 23 and 18 kDa, respectively, closely matching their theoretical masses of 24 kDa for LC8 dimer and 17 kDa for dLBD monomer. Cartoon depictions of the ASCIZ dLBD (light blue) and LC8 (dark blue) are shown to aid in visualization of the complexes formed at a 1:10 ratio. (b) Native gel titration of dLBD with LC8. An increasing concentration of LC8 was added to a constant amount of dLBD, from a molar ratio of (dLBD:LC8) 1:1 to 1:8. As more LC8 is added, the complex migrates more slowly and excess LC8 appears. (c) Native gel titration of human LBD with LC8. As an increasing concentration of LC8 is added to a constant amount of LBD. Arrows indicate the locations of the low molecular weight (LMW) and high molecular weight (HMW) complex. (d) Titration of the LBD with LC8, monitored by size exclusion chromatography on a Superdex 200 gel filtration column. The concentration of LBD is held constant and an increasing amount of LC8 is added, from a molar ratio of 1:3 to 1:13. Peaks corresponding to free LBD, free LC8, low molecular weight complex (LMW) and high molecular weight (HMW) complex are labeled.

Structure and Distribution of the LBD:LC8 Complexes Visualized by Single Particle EM

In order to visualize the various oligomeric states of ASICZ-LC8, we analyzed electron microscopy data of dLBD and human LBD peptides under saturating concentrations of LC8. As a positive control and validation of EM conditions, similar experiments were carried out with complexes of another intrinsically disordered protein with multiple LC8 binding sites, Nucleoporin159 (Nup159), in complex with LC8[38]. Nup:LC8 complexes were clearly visualized as a linear array of 5 stacked densities of the LC8, as previously reported[38] (data not shown), and consistent with the conclusions of Nup:LC8 biophysical solution experiments[36]. In contrast, despite the similar overall affinity toward LC8 reported for Nup159 (K_d = 2.9 μM)[36], in negative stain images of dLBD:LC8 and human LBD:LC8 complexes the vast majority of complex species appear dissociated on the grid (Figure 4.5a, *arrow heads*). However, a few observable complexes could be clearly resolved from raw micrographs, identified as linear stacks of punctate densities, like beads on a string (Figure 4.5a, *squares*), similar to images of Nup159 bound to LC8[38]. Furthermore, although dLBD contains 7 LC8 binding sites and the human LBD peptide has 11 LC8 binding sites, the vast majority of complexes observed by EM appeared to be of low LC8 occupancy.

Reference-free two-dimensional (2D) classification routines were carried out on datasets of ~2000 single particle images of dLBD:LC8 complexes and ~1000 particles of human LBD:LC8 complexes extracted from ~300 and 200 micrographs, respectively. These produced 2D projection averages for dLBD:LC8 and LBD:LC8 oligomers displaying complexes formed with 2 – 4 stacked densities, corresponding to LC8 dimers, which could be deduced from the dimensions of the averaged bead-like densities (~4 nm diameter)[8, 38]. Complexes with 3 or more LC8 dimers displayed significant conformational flexibility in 2D class averages (Figure 5b) and in the single-particle images (Figure 4.5a and SI Figure 4.3). The extent of conformational variability is consistent with ~10 – 20 Å spacing measured between LC8 densities, and the intrinsic flexibility of the IDP duplex chain separating the neighboring LC8 TQT recognition motifs.

The formation of higher-order oligomers appeared relatively rare, and the fully formed complexes were almost completely absent under saturating LC8 conditions. This low population, coupled with the intrinsic conformational heterogeneity, precluded our ability to obtain 2D class averages of the high-occupancy complexes. To overcome this limitation, statistical analysis describing the distribution of oligomeric states was obtained by hand-selection and classification from single-molecule images (SI Figure 4.3). Both dLBD:LC8 and LBD:LC8 complexes form an

ensemble of structures, displaying an exponential distribution with low-occupancy states (*i.e.* 2 – 4 stacked LC8 dimers) being most abundant (Figure 4.5c,d). Density corresponding to the IDP duplex peptide cannot be resolved by negative stain EM, therefore complexes formed with a single LC8 dimer were not included in this analysis, as they could not be distinguished from unbound LC8 dimers.

Together, this analysis shows dLBD and LBD peptides support dynamic assemblies with LC8 that favor low-occupancy states (Figure 4.5d). Although uncommon, high-occupancy and fully-formed complexes of dLBD:LC8 (1:7 ratio) could be identified from the raw single particle images (SI Figure 4.3), further confirming the stoichiometry obtained by our ITC studies. For human LBD:LC8 dataset, complexes containing as many as 7 – 9 LC8 dimers could be distinguished from the single particle image data, while higher-order complexes containing 10 – 11 LC8 dimers were either not distinguishable or were simply absent under the limiting concentrations required for negative stain EM specimen preparation. Never-the-less, the remarkable similarity in distribution of oligomeric species formed by these two peptides, obtained under similar binding conditions, is consistent with the nearly equivalent overall LC8 affinity determined by ITC, and suggests a conserved mechanism of negative cooperativity is used by ASICZ to regulate the formation and distribution of higher-order LC8 assemblies.

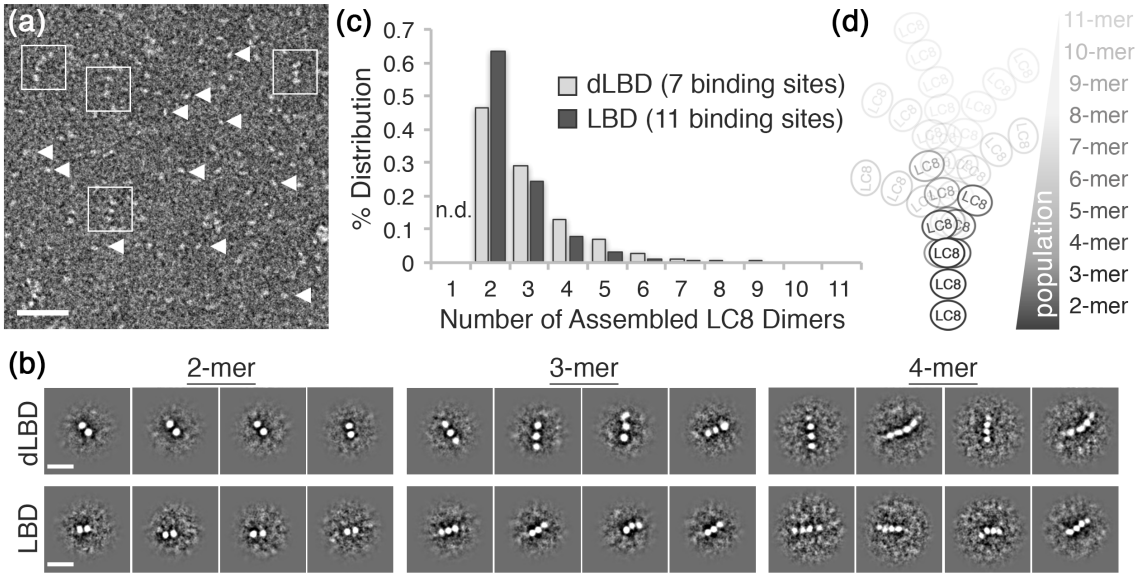


Figure 4.5. dLBD:LC8 and LBD:LC8 complexes visualized by negative stain electron microscopy.

(a) Representative micrograph of negatively stained LBD:LC8 complexes. Identified oligomeric complexes are boxed. Non-oligomeric LC8 dimers are indicated by arrowheads. Scale bar = 100 nm. dLBD:LC8 micrographs had the same appearance *(not shown)*. (b) Representative 2D projection averages of *(top)* dLBD:LC8 oligomers and *(bottom)* LBD:LC8 oligomers. Only low-occupancy oligomers with 2 – 4 LC8 dimers (2mer – 4mers) were successfully averaged. Higher-occupancy oligomers were identified in raw micrographs, but were not averaged due to low population and/or high degree of conformational flexibility. Scale bar = 20 nm. (c) Histogram showing population distribution of LC8 occupancy identified in complexes formed with dLBD *(grey)* and LBD *(black)* peptides, identified from raw micrographs. The population of complexes formed with a single LC8 dimer were not determined *(n.d.)* (d) Illustration representing the distribution of LC8 occupancy and conformational flexibility observed in LBD:LC8 complexes.

NMR Titration of dLBD with LC8 Indicates Preferential Binding of Motifs QT4-6

If dLBD and LC8 form stable intermediate complexes in the presence of excess LC8, which of the seven recognition sites are preferentially bound? To examine interactions between individual motifs and LC8 in the context of the full dLBD sequence, we turned to NMR. Unlabeled LC8 was titrated into solutions of ^{15}N-^{13}C- labeled dLBD, and changes in NH peak intensities were measured in 3D HNCO spectra recorded for dLBD:LC8 molar ratios of 1:0.25, 1:1, 1:2, 1:5, 1:8. Since the dLBD:LC8 complex is unstable at pH <7.5, the titration experiment was carried out at pH 7.5 which results in loss of several low intensity peaks. As LC8 concentration increases, a corresponding decrease in dLBD peak intensity is observed (Figure 4.6a). At a molar ratio of 1:5, less than 10% of the original peak intensity remains at all seven LC8 binding sites. At a ratio of 1:8, all dLBD peaks completely disappear except for peaks corresponding to eight N-terminal residues (241-248), indicating that dLBD is fully bound. The absence of peaks for bound dLBD is attributed to line broadening associated with intermediate exchange processes and/or faster transverse relaxation as a result of increased complex size. Therefore, we consider a decrease in peak intensity as a measure of increased complex formation.

Notably, peaks at the C-terminal half of the protein, QT4-7, decrease more quickly than peaks at the N-terminal half, implying that LC8 preferentially occupies these motifs. This finding prompted us to perform NMR titrations of the QT2-4 and QT4-6 because QT2-4 could further validate LC8 binding to the QT3 motif as observed by ITC, and QT4-6 has the highest LC8 binding affinity by ITC (Figure 4.3b). Further, the two constructs share the QT4 motif, allowing us to assess motif affinity in two sequence contexts. Unlabeled LC8 was titrated into ^{15}N-labeled QT2-4 or QT4-6 and changes in peak intensity were analyzed in 2D HSQC spectra (Figure 4.6b,c). Due to the smaller number of peaks for shorter constructs, HSQC spectra have a sufficiently high resolution to render an HNCO-based titration unnecessary. As with dLBD, there is a gradual decrease in peak intensity as more LC8 is added to QT2-4 or to QT4-6 at molar ratios: 1:0.25, 1:1, 1:2, 1:3, and 1:4. Significantly, peak intensity in the QT4-6 construct decreases at lower LC8 ratio than in QT2-4, confirming the trend we observe in full-length dLBD. In QT4-6, nearly all peaks in the motif region disappear at a ratio of 1:2, while ~30% of peak intensity remains in the QT2-4 construct.

To better visualize the interactions of each LC8 recognition motif with LC8, the average peak intensity (I/I_0) versus LC8 concentration is plotted for each 10-amino acid motif in dLBD, QT2-4, and QT4-6 (Figure 4.6d-g). In dLBD, a clear dichotomy in the pattern of peak disappearance is observed between the first three motifs (Figure 4.6d) and the last four motifs

(Figure 4.6e). This dichotomy is replicated by QT2,3,4 motifs *versus* QT4,5,6 motifs; e.g., the average I/I_0 of QT2-4 (Figure 4.6f) drops to 0.2 at LC8 molar ratio 1:2, twice that observed for QT4-6 (Figure 4.6g) which reaches the same I/I_0 at molar ratio 1:1. The apparent higher LC8 affinity of motifs 4-6 (Figures 4.6e, g), relative to motifs 2-4 (Figures 4.6d, f) is as expected from ITC experiments (Figure 4.4a). The QT4 motif, which is common to both constructs, has a different rate of peak disappearance in each construct, suggesting that motif location, not local sequence, determines its affinity. We conclude from the data in Figure 4.6 that the recognition motifs QT4, QT5 and QT6 are likely the sites favored in stable low occupancy complexes.

In summary, the data in Figure 4.6 support the conclusion that motifs QT4, QT5 and QT6 behave as a unit distinguishable from other motifs by their high affinity for LC8, and that these sites preferentially bind LC8 to form stable low occupancy ASCIZ:LC8 complexes even in the presence of excess LC8. Further, titration with construct QT2-4 confirms an LC8-binding function of the QT3 motif, identified here for the first time by ITC (Figure 4.3b,c Table 4.1); peaks corresponding to QT3 disappear at a similar rate as those of QT2, but at quite different rates from those of the adjacent linker residues.

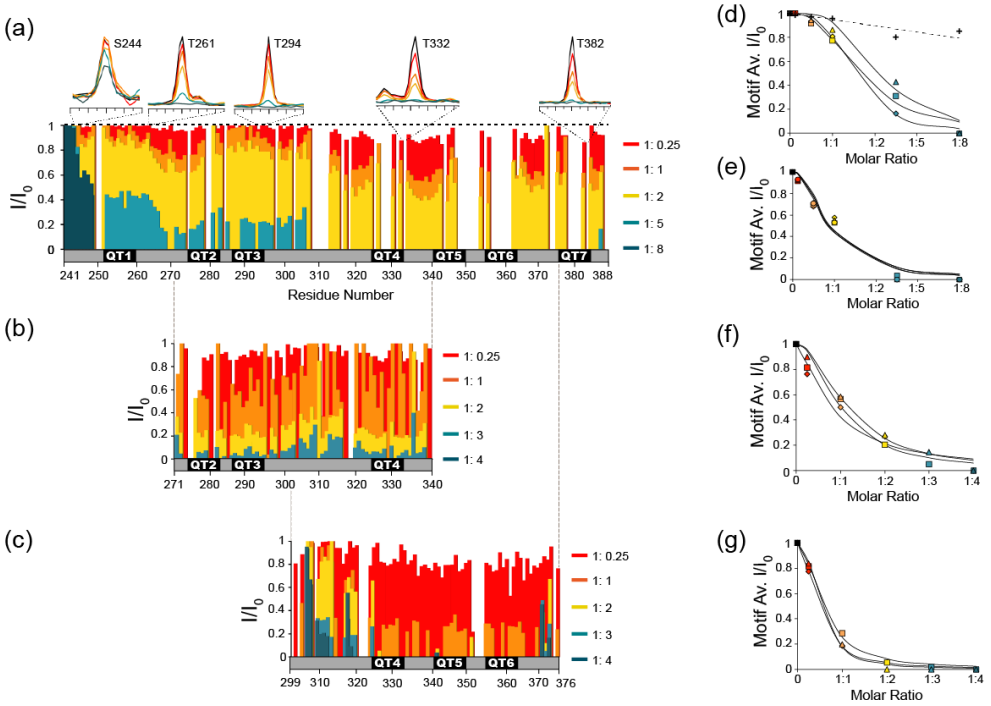


Figure 4.6. NMR titration of the dLBD with LC8.
Relative intensities of non-proline NH peaks in ^{15}N-^{13}C-HNCO spectra were measured for the dLBD:LC8 complex formed at varying molar ratios 1:0.25, 1:1, 1:2, 1:5, and 1:8. Spectra were obtained for LC8 complexes with the full-length dLBD and two shorter constructs, and peak intensities (I) are given relative to the intensity of the same peak in free peptide (I_0). Titration results are shown for dLBD (a), QT2-4 (b) and QT4-6 (c). Above (a) are 1D NMR slices of representative amino acids from several TQT motifs (T261, T294, T332, T382) and a control (S244). The data in (a-c) are alternatively plotted as the average intensity ratio for each 10-amino acid motif versus the molar ratio of dLBD:LC8. For the dLBD complex, titration curves of individual motifs cluster in two groups having higher (d) or lower (e) average intensity ratios at the same dLBD:LC8 molar ratio. Similarly for the two shorter constructs, plots of average I/I_0 for individual motifs cluster in two groups, motifs in QT2-4:LC8 having higher intensity ratios (f), and motifs in QT4-6:LC8 having lower intensity ratios (g). Motif designations in (d) and in (f) are: QT1 (triangle), QT2 (square), QT3 (diamond). Motif designations in (e) and (g) are: QT4 (triangles), QT5 (squares), QT6 (diamond), and QT7 (circle). In (d) values are shown for a negative control group (crosses, dotted line) comprised of those measurable at saturating LC8 concentration in the first 8 amino acids, 241-249. In panels (d-g), black curves are to guide the eye.

LC8 binding to ASCIZ results in a gradient of transcriptional activity.

To investigate how the number of bound LC8 molecules affects the transcriptional activity of ASCIZ, we turned our attention to the human protein whose transcriptional activity can be assayed in cell culture using an ASCIZ knockout mouse embryonic fibroblast cell line[56]. Human ASCIZ has eleven LC8 recognition motifs (Figure 4.7a). To prevent LC8 binding to specific ASCIZ motifs, TQT recognition motifs were mutated to AAA. We developed five human ASCIZ constructs having specified TQT to AAA replacements corresponding to motif clusters in WT ASCIZ. Mutant construct AAA1-4 has an AAA replacement at each TQT motif numbered 1-4 in Figure 4.7a; similarly named mutant constructs are AAA8-11; AAA5-11; AAA1-4, 8-11; and AAA-all.

To assess the impact of these mutations on ASCIZ transcriptional activity, luciferase reporter assays were carried out using immortalized ASCIZ knockout mouse embryonic fibroblasts transiently transfected with each full-length WT or mutant ASCIZ protein and a plasmid containing the LC8 promoter. The measured luciferase activity was normalized against *Renilla* luciferase (Figure 4.7c). Empty vector and the ΔZnF construct showed limited transcriptional activity compared to ASCIZ constructs. Most significantly, transcriptional activity of the AAA mutants can be easily ranked to form an activity gradient (Figure 4.7c) notable for a clear inverse relationship of activity and number of LC8 motifs. The transcriptional activity of a construct with 7 sites, AAA1-4, is 1.5x higher than WT ASCIZ, while a construct with 3 sites, AAA1-4,8-11, has 2.5x the activity of WT ASCIZ.

There are two exceptions to this trend: (1) The construct with zero functional LC8 binding sites (AAA-all) shows lower activity than the construct with 3 sites (AAA1-4,8-11). This could be due to a weaker binding of ASCIZ ZnF domain to the DNA promoter when not complexed to LC8. The dASCIZ ZnF DNA binding domain is monomeric (Figure 4.2c) and this domain is well conserved between the two species. It is also possible that transcription activators of ASCIZ are dimers and thus have tighter binding to ASCIZ cross-linked by LC8. (2) The AAA8-11 construct with 7 sites shows a similar, or even slightly lower transcriptional activity compared to WT ASCIZ. This could be due to its slightly tighter binding to LC8 (lower K_d) relative to WT (Figure 4.7b, Table 4.3), compared to all of the other constructs, which exhibit a higher K_d than WT. For each construct, the luciferase activity vs. K_d plot shows a clear increase in transcriptional activity as ASCIZ affinity for LC8 decreases due to a decreasing number of binding sites (Figure 4.7d).

To confirm LC8 binding *in vitro* for each of the TQT to AAA mutants, identical mutations were made in ASCIZ LBD constructs, named in the same fashion. To completely eliminate LC8 binding in the AAA-all construct, it was necessary to mutate three SQT/VQT motifs in addition to the 11 TQT motifs identified in our ITC experiments and in pepscan experiments[69]. The affinity and stoichiometry of binding of LC8 to WT ASCIZ LBD and each mutant LBD peptide was determined (Figures 4.7b,d and Table 4.3). WT LBD binds LC8 with an LBD:LC8 ratio of 1:11(two chains of ASCIZ to 11 LC8 dimers), as expected based from pepscan data [69], and an average K_d value of 0.9 μM (Table 4.3). The four mutant LBD peptides each bound LC8 with the expected stoichiometry for the number of remaining LC8 recognition motifs and with affinities in the range of 0.7-4.4 μM.

In summary, the data in Figure 4.7 indicate that, in general, ASCIZ transcriptional activity varies inversely with the number of LC8 recognition motifs, and that fine tuning within this trend depends on which motifs are occupied and their specific dissociation constants.

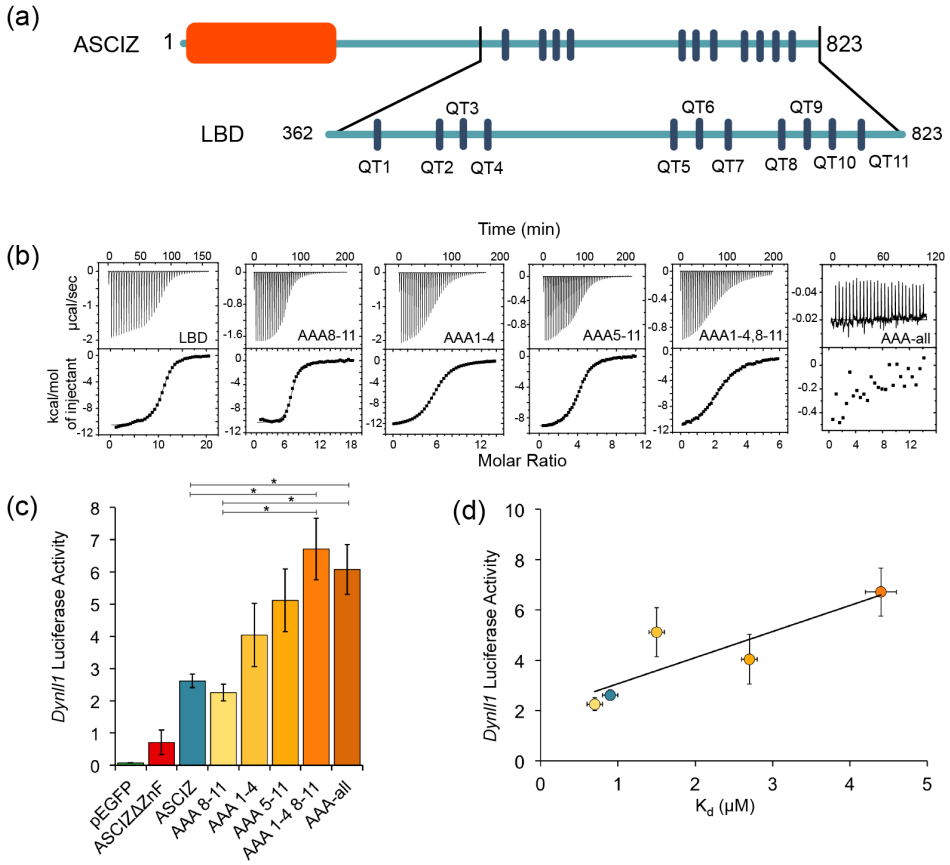


Figure 4.7: The number of LC8 recognition motifs correlates with ASCIZ transcriptional activity.

(a) Domain structure of human ASCIZ, showing 11 LC8 binding motifs as blue bars. (b) Shown are representative isothermal titration plots of LC8 with the LBD and constructs with four (AAA1-4 and AAA8-11), seven (AAA5-11), eight (AAA1-4, 8-11), or fourteen (AAA-all) mutant LC8 binding sites. Data were fit to a single site binding model using Origin software. (c) Firefly luciferase reporter assays of ASCIZ knockout mouse embryo fibroblast cells transiently transfected with WT human ASCIZ, a zinc finger deletion construct, and ASCIZ mutant constructs AAA1-4, AAA8-11, AAA5-11, AAA1-4, 8-11, or AAA-all, along with the *Dynll1* luciferase and *Renilla* luciferase vectors. Error bars are ± S.E. relative to *Renilla* luciferase as a control. Asterisks (*) indicate p-values less than 0.01. (d) Binding affinity of each construct for LC8 is plotted against luciferase activity. Data points are colored according to (c). The AAA-all construct is excluded from the graph because it does not bind to LC8.

Construct	N	Overall K_d (μM)	Overall ΔH (kcal/mol)	Overall TΔS (kcal/mol)	Overall ΔG (kcal/mol)
LBD	11.2	0.9 ± 0.1	-10.6 ± 0.5	-2.4 ± 0.6	-8.2 ± 0.4
AAA8-11	6.7	0.7 ± 0.1	-10.4 ± 0.5	-2.0 ± 0.6	-8.4 ± 0.4
AAA1-4	6.6	2.7 ± 0.1	-12.6 ± 0.6	-5.0 ± 0.7	-7.6 ± 0.4
AAA5-11	4.2	1.5 ± 0.1	-9.2 ± 0.5	-1.3 ± 0.6	-7.9 ± 0.4
AAA1-4, 8-11	2.5	4.4 ± 0.2	-12.2 ± 0.6	-4.9 ± 0.7	-7.3 ± 0.4

Table 4.3. Thermodynamic parameters of ASCIZ-LC8 interactions.

Discussion

Extensive *in vitro* biophysical experiments and *in vivo* transcriptional activation assays illuminate the molecular mechanism of ASCIZ transcriptional activity. We find that the level of ASCIZ transcription is controlled by the number of LC8 molecules bound, such that increasing LC8 concentrations signal decreasing LC8 transcriptional levels. To achieve this negative feedback function, ASCIZ and LC8 engage in a dynamic equilibrium of multiple ASCIZ:LC8 complexes, the most prevalent of which have low LC8 occupancy and a moderate level of transcriptional activity. We propose that this dynamic ensemble of complexes is important for fine-tuning ASCIZ transcriptional activity, and that stable, low occupancy complexes function to maintain a basal transcription rate for LC8.

A low occupancy, dynamic complex is a conserved feature of the ASCIZ:LC8 interaction

Evidence for a dynamic ASCIZ:LC8 ensemble comes from a combination of AUC, gel filtration, native gel, and negative stain electron microscopy data (Figures 4.4-4.5). For both *Drosophila* and human LBD constructs, addition of excess LC8 results in formation of stable low molecular weight (LMW) complexes and a minor population of high molecular weight complexes (Figure 4.4). The ensemble is in dynamic conformational equilibrium, as the ASCIZ duplex is inherently disordered and flexible. The ensemble is also in dynamic binding equilibrium between complexes and free LC8. This exchange is clearly seen by native gel electrophoresis (Figure 4.4b-c), and by AUC and gel column filtration (Figure 4.4a,d). SAXS and AUC analysis demonstrate that low

occupancy complexes are the most stable (SI Figure 4.2). Negative stain electron microscopy experiments visualize 3-4 clearly stacked copies of LC8 and suggest flexibility and fast off-rates in the rest of the complex (Figure 4.5). Complexes with 3- or 4- bound LC8 copies are the most highly populated for both the human and *Drosophila* proteins, possibly due to the clustering of TQT motifs in groups of 3 or 4 separated by linkers (Figure 4.1d).

Additionally, ITC results demonstrate that LC8 binds more tightly when a few specific ASCIZ sites are occupied, rather than all 7 (dLBD) or 11 (LBD) sites. For the dLBD, a construct containing only sites 4-6 (QT4-6) binds LC8 with higher affinity than the full-length. NMR titration of the dLBD and shorter constructs with LC8 suggests that these C-terminal sites are the first to bind LC8, and the sites bound in the low occupancy complexes. Residues in the C-terminal motifs are slightly more ordered in comparison to the N-terminal QT1-3 motifs (Figure 4.2f-g, SI Figure 4.1), which may explain why LC8 binding saturates this region first. The different binding behavior between QT1-3 and QT4-7 offers insight into how regions of differing flexibility can influence complex assembly and binding affinity. A similar dichotomy for the LBD is inferred from the tighter complex formed upon removal of the C-terminal 4 sites (AAA8-11).

ASCIZ is a New Type of Multivalent LC8 Binding Partner

The majority of LC8 binding partner proteins are intrinsically disordered in all or some of their primary sequence and contain a single TQT recognition motif. As a result of LC8 binding, they form a dimeric, linear, bivalent scaffold [12, 151] onto which other proteins assemble. Frequently, the two IDP chains in the bivalent scaffold are self-associated [43], or form a coiled-coil [44] at a site proximal to the TQT motif. The remarkably broad occurrence of LC8/IDP complexes underscores the importance of LC8 in macromolecular assemblies.

Although most LC8-binding partners have only one LC8 binding site, there are some that, like ASCIZ, bind multiple LC8 molecules. Among these, ASCIZ is the only one in which the most stable IDP:LC8 complex has fewer than the maximum number of motifs occupied; for all others, more LC8 copies in the complex results in higher overall binding affinity. For example, Nucleoporin 159 (Nup159) binds five LC8 dimers upstream of a coiled-coil motif to form a rigid assembly readily visible by electron microscopy [38].

As a positive control for negative staining experiments with LBD:LC8, we replicated the experiments in Figure 4.5 using the LC8-binding domain of Nup159 in complex with LC8, and resultant electron micrographs revealed a substantial number of 5-LC8 rod-like structures (data

not shown). As Nup159 and ASCIZ share a similar overall affinity for LC8 (2.9 μM vs. 0.9 μM, respectively[36], with ASCIZ considerably tighter), the stark difference between the Nup159 grids with ubiquitous 5-mer structures and the ASCIZ grids with sparse, heterogeneous, 3-mer and 4-mer structures highlights an intriguing difference between the two complexes. This observation is consistent with the idea that the dynamic properties of the ensemble of LC8-ASCIZ complexes are important for autoregulation of LC8 transcription.

Another multivalent LC8 binding partner, the spindle assembly protein Chica, binds three LC8 dimers to form a bivalent complex[30]. Unlike Nup159 and ASCIZ, one TQT motif binds much more tightly than the rest [30], making it likely that the Chica:LC8 complex assembles through a two-step process where the first motif binds to LC8 and forms a scaffold onto which the next two dimers can assemble. Two other multivalent LC8 binding partners with >2 TQT motifs exist [64, 67], but assembly processes of the LC8 complex have not been studied *in vitro*.

ASCIZ Binding of Multiple LC8 Copies: A New Molecular Mechanism for Negative Feedback Regulation

Previous *in vivo* transcription assays demonstrate that ASCIZ controls LC8 transcription through a system of negative feedback regulation[56]. ASCIZ binding to the LC8 promoter turns on LC8 transcription, while ASCIZ binding to LC8 turns down LC8 transcription[56]. Here we elucidate the structural mechanism by which LC8 binding regulates ASCIZ transcriptional activity. Our major findings are the following: 1) At high levels of LC8, there is a dynamic distribution of ASCIZ:LC8 complexes with variable LC8 occupancies that are in exchange with each other and with free LC8. 2) LC8 preferentially binds to the C-terminal half of dASCIZ, favoring the formation of a stable, low occupancy 3-mer complex. A low occupancy 3-4 mer complex is also favored for human ASCIZ, emphasizing the importance, and functional relevance, of stable intermediate complexes for ASCIZ transcriptional activity. 3) The transcriptional activity of ASCIZ is inversely proportional to its LC8 occupancy and so, like occupancy, is regulated by local LC8 levels. We hypothesize that the gradient of LC8 copies per ASCIZ chain, and the related gradient of ASCIZ transcriptional activity, are finely tuned by the internal flexibility of the complex which allows dynamic access of free LC8 to recognition motifs, and of bound LC8 to the solvent

A model that summarizes this regulation is shown in Figure 4.8, in which the two domains of a free ASCIZ chain (right) are shown as a red helical ribbon representing the N-terminal ZnF domain and a blue disordered ribbon representing the LC8 binding domain.

Flexible ensembles of ASCIZ:LC8 complexes are simplified down to three diagrammatic low occupancy structures. The complex with 3 LC8-bound motifs (middle) has higher transcriptional activity and is more stable; the complex with all motifs LC8-bound (left) has lower transcriptional activity and is less stable. The ASCIZ:LC8 complexes bind and dissociate LC8 dimers, thereby creating a gradient of LC8 occupancies, and a related gradient of transcriptional activity rather than an on/off switch. Low occupancy complexes are arguably important for enabling a fast response to changing LC8 concentrations and for maintaining a basal level of LC8 transcription. Since low occupancy complexes persist at higher LC8 concentrations, they could help maintain a moderate level of transcriptional activity essential for activity of LC8 in numerous cellular process.

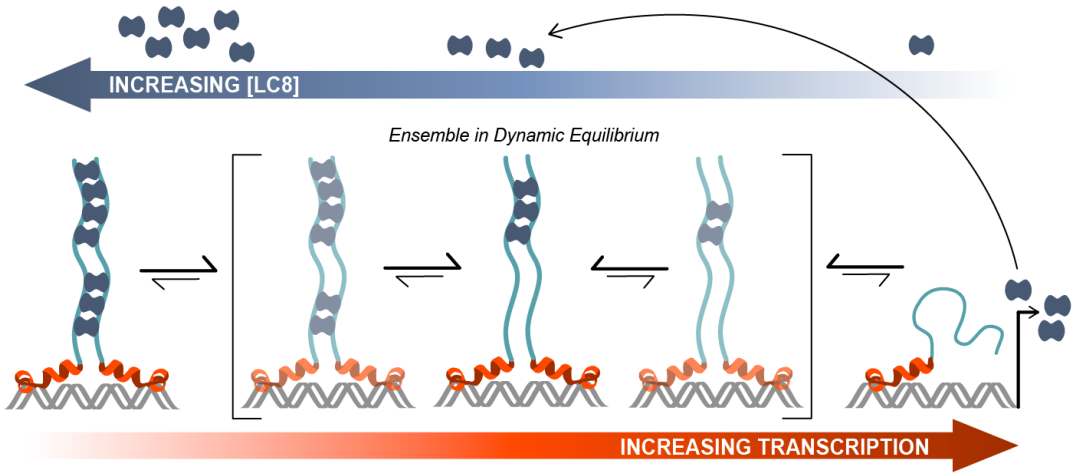


Figure 4.8. Model of ASCIZ regulation of LC8 transcription.
A proposed model of LC8 transcriptional regulation is shown for dASCIZ, but the model also applies to the human protein. Free LC8 dimers (dark blue) bind to ASCIZ and modulate transcriptional activity. A dynamic, low-occupancy complex (center) is the predominant species. LC8 that is produced from ASCIZ transcription returns to the pool of free LC8.

This negative feedback mechanism of LC8 transcriptional regulation by ASCIZ-LC8 binding is a new example of how multivalency and intrinsic disorder can be harnessed to modulate activity. The prevalence of intrinsic disorder in other transcription systems[178] suggests that this type of process may be utilized by other transcription systems. While no examples of transcription factors regulated by binding multiple copies of their own gene product could be found in the literature, there are many diverse examples of multi-site recognition in transcriptional regulation. Many multi-site transcription systems involve multiple phosphorylation events, such as the E26 transformation-specific transcription factor (Ets-1) and p53. Ets-1 is autoinhibited by an intrinsically disordered serine rich domain and a helical inhibitory module. Multi-site phosphorylation of the serine rich region occludes the DNA-binding interface and stabilizes the helical inhibitory module, inhibiting Ets-1 DNA binding ~20-fold[190-192]. The function of the transcription factor p53, a tumor suppressor protein, is modulated by over 50 posttranslational modifications that are proposed to be interdependent[176]. Phosphorylation of specific p53 residues prevents binding to the inhibitory protein HDM2, while increasing binding to the activating proteins CREB-binding protein (CBP) and p300[193]. p53 affinity for CBP/p300 depends on the extent of p53 phosphorylation; successive phosphorylation events increase p53 affinity for the TAZ1, TAZ2, and KIX domains of CBP/p300[194, 195].

Conclusions

Here we integrate structure, dynamics, thermodynamics and hydrodynamics approaches to characterize large, dynamic, disordered complexes that collectively present a molecular model by which the transcription factor ASCIZ maintains stable pools of the hub protein LC8. ASCIZ uses multiple short LC8 recognition motifs in its large disordered domain to form a dynamic ensemble that creates a gradient of transcriptional activity. We verified the main aspects of this model in cells using transcription activity assays in which mutant ASCIZ constructs with lower LC8 occupancy display higher transcriptional activity, while constructs with higher LC8 occupancy have lower activity. We propose that the dynamic nature of the ASCIZ:LC8 complex enables a tunable gradient of transcription activity. The inverse correlation between transcriptional activity and affinity for LC8 supports this hypothesis. Although many other transcription factors are regulated by multisite phosphorylation[196] or multiple binding events to different proteins[177, 197], we find no examples of activity tuned by multivalent binding to the gene product in a negative autoregulatory role, underscoring the novelty and potential impact of this study.

Materials and Methods

Cloning, Protein Expression, and Purification

Studies were carried out using constructs from human ASCIZ (Uniprot O43313) as well as *Drosophila* ASCIZ (dASCIZ) (Uniprot Q9VZU1) which, with its fewer recognition motifs, smaller size, and available mutant phenotypes, is a tractable model of the human ASCIZ. Constructs of the dASCIZ zinc finger domain (ZnF) and the LC8 binding domain (dLBD) were generated by cloning residues 1-156 or 241-388, respectively, of *Drosophila* ASCIZ into the pET2Zt2-1a vector. The constructs were expressed in frame with a hexahistidine tag, Protein A solubility tag, and cleavage site for the tobacco etch virus (TEV) enzyme. Shorter constructs of the dLBD were generated by cloning residues 241-324 (QT1-3), 271-341 (QT2-4), 299-376 (QT4-6), and 321-388 (QT4-7) into the pET2Zt2-1a vector. The human LC8 binding domain (LBD) construct was generated by cloning human ASCIZ (Uniprot O43313) residues 362-823 into the pET24d vector (Novagen) and expressing the construct in frame with a hexahistidine tag and TEV cleavage site.

For the five human ASCIZ mutants, ASCIZ AAA1-4, AAA8-11, AAA5-11, AAA1-4,8-11, and AAA-all, residues 7-9 of the LC8 binding motif (usually the residues TQT), were mutated to AAA to prevent binding. Mutations were performed using either the QuikChange Lightening Mutagenesis Kit (Agilent) or by synthesizing short constructs (300-350 bp) containing the desired mutations and using Gibson Assembly (New England Biosciences) to insert them into the LC8-binding domain (LBD) gene (residues 362-823 of human ASCIZ). All constructs were transformed into *Escherichia coli* Rosetta DE3 cells and expressed at 37°C in LB or minimal autoinduction media with ^{12}C or ^{13}C glycerol and $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively. Recombinant protein expression was induced with 0.4 mM IPTG (for LB cultures) and growth continued at 25°C for 16 hours. Cells were harvested and purified under denaturing conditions using Clontech's TALON His-Tag Purification protocol (Clontech). The solubility tag and/or hexahistidine tag were cleaved by TEV protease and the protein was further purified using strong anion exchange chromatography (Bio-Rad) followed by gel filtration on a Superdex^{TM} 75 gel filtration column (GE Health). The purity of the recombinant proteins, as assessed by SDS-polyacrylamide gels, was > 95%. The pure proteins were stored at 4°C and used within 1 week. LC8 was prepared as previously described [16].

Peptide Design and Synthesis

Peptides corresponding to the seven putative recognition sequences from dASCIZ were commercially synthesized: YMSSQKLDMETQTEE (QT1p), YLAPLLRDIETQTPD (QT2p), YTPDTRGDIGTMTDD (QT3p), DLQTSAHMYTQTCD (QT4p), EELGLSHIQTQTHW (QT5p), WPDGLYNTQHTQTCD (QT6p), and EPDNFQSTCTQTRW (QT7p) (Genescript, Piscataway, NJ). Non-native amino acids (underlined in Table 2) were added to the N-terminus to enhance solubility or concentration determination by UV absorbance.

Isothermal Titration Calorimetry

Binding thermodynamics of the ASCIZ and dASCIZ construct/peptide-LC8 interactions were obtained at 25°C with a VP-ITC microcalorimeter (Microcal). The reaction buffer was composed of 50 mM sodium phosphate, 50 mM sodium chloride, 1 mM sodium azide, 5 mM β-mercaptoethanol, pH 7.5. dLBD was placed in the reaction cell at a concentration of 8 μM and titrated with LC8 at a concentration of 800 μM. For binding of dASCIZ constructs QT1-3, QT2-4, QT4-6, and QT4-7, 10 μM of construct was titrated with 400 μM LC8. For interactions with synthetic peptide, peptides were dissolved in reaction buffer to a final concentration of 300 μM and then added to LC8 at a concentration of 30 μM in the reaction cell. ASCIZ LBD and mutant LBD constructs were placed in the reaction cell at a concentration of 9-16 μM and titrated with 900 μM LC8. Peak areas were integrated and data were fit to a single-site binding model in Origin 7.0 from which the stoichiometry (N), dissociation constant (K_d), and the change in enthalpy (ΔH), and entropy (ΔS) were obtained. Reported data are the average of two or more independent experiments. As the binding-model fit was very good and data were reproducible, error was determined based on a 5% uncertainty in protein concentration calculations.

Circular Dichroism

CD experiments were conducted on a Jasco720 spectropolarimeter in a 1 mm cell. For the spectrum of dLBD, ten scans were averaged at a concentration of 30 μM in a buffer composed of 10 mM sodium phosphate, pH 7.5, at 25°C and 10°C. For the ZnF, ten scans were averaged at a concentration of 25 μM in a buffer composed of 10 mM sodium phosphate, 200 mM sodium sulfate, 50 μM zinc sulfate, pH 7.5, at 10°C, 25°C, and 35°C.

Analytical Ultracentrifugation

Sedimentation velocity experiments for the titration of dLBD and LC8 were performed in a Beckman Coulter Model XL-I analytical ultracentrifuge equipped with UV/Vis scanning optics. Reference (400 μL; 50 mM sodium phosphate, 50 mM sodium chloride, 1 mM sodium azide, 5 mM TCEP, pH 7.5) and sample (380 μL) solutions were loaded into 12 mm double-sector cells with quartz windows and the cells were then mounted in an An-50 Ti 8-hole rotor. LC8 was prepared at a concentration of 15 μM while the concentration of dLBD was varied from 15 – 1.5 μM. Proteins were centrifuged at 50,000 rpm at 20 °C, and radial absorbance data were collected at appropriate wavelengths in continuous mode every 5 minutes without averaging. Data were fit to a continuous size-distribution [c(S)] model using the program SEDFIT [198]. The partial specific volume of the proteins, buffer density, and buffer viscosity were computed using the program SEDNTERP [199].

Sedimentation velocity experiments for the ZnF domain were performed on a Beckman ProteomeLab™ XL-A/XL-I analytical ultracentrifuge in a buffer composed of 50 mM sodium phosphate, 200 mM sodium chloride, 0.4 mM zinc sulfate, 1 mM sodium azide, 2 mM TCEP, pH 7.0. The sample was centrifuged at 40,000 rpm at 20 °C for 7 hours and absorbance data were collected at 286 nm. Data were fit to a continuous size-distribution [c(s)] model using the program SEDPHAT [200].

NMR Experiments

NMR measurements were collected at 10°C, using 300-350 μM isotopically ($^{13}C/^{15}N$ or ^{15}N) labeled dLBD in a buffer at pH 6.5 composed of 10 mM sodium phosphate, 10 mM sodium chloride, 1 mM sodium azide, 10 mM β-mercaptoethanol, a protease inhibitor mixture (Roche Applied Science), and 2-2 dimethylsilapentane-5-sulfonic acid for ^{1}H chemical shifts referencing. Data for backbone assignments were collected on a Bruker Avance 850 MHz spectrometer equipped with a cryoprobe. Five-dimensional HN(CA)CONH and HabCabCONH experiments [201, 202] and a three-dimensional HNCO experiment were acquired with non-uniform sampling of the indirectly detected dimensions and used for sequential assignment of ^{13}C-^{15}N- dLBD (dLBD for LC8 binding domain).

Interaction of unlabeled LC8 and ^{13}C-^{15}N labeled dLBD (residues 241-388) was characterized by collecting three-dimensional BEST-TROSY-HNCO spectra at multiple molar ratios of LC8, 1: 0.25 (dLBD: LC8), 1:1, 1:2, 1:5, and 1:8. For the interaction of unlabeled LC8 with ^{15}N-labeled QT2-4 (residues 271-341) or QT4-6 (residues 321-376), two-dimensional

BEST-TROSY-HSQC spectra were collected at the molar ratios (QT2-4/QT4-6:LC8) 1:0.25, 1:1, 1:2, 1:3, and 1:4. NMR titration data were analyzed and plotted by measuring peak volumes using Sparky and averaging over each 10 amino acid QT motif.

HNCO-based R_1 relaxation measurements experiments were recorded with relaxation delay times ranging from 3 to 630 ms, and the R_2 relaxation data were acquired using relaxation delays ranging from 14.4 to 259 ms. Curve fitting was performed using the rate analysis script Sparky2Rate and the program Curvefit (A. G. Palmer, Columbia University). Steady-state $^{1}H-^{15}N$ heteronuclear NOEs were acquired using 6 s total saturation time. Error bars were determined from the intensities of the baseline noise using the formula $\sigma/(NOE) = [(\sigma I_{sat}/I_{sat})^2 + (\sigma I_{unsat}/I_{unsat})^2]^{1/2}$, where I_{sat} and σI_{sat} correspond to the intensity of the peak and its baseline noise.

All two-dimensional spectra and the three-dimensional HNCO spectra were processed using TopSpin (Bruker Biosciences), and the non- uniformly sampled five-dimensional HN(CA)CONH and HabCabCONH spectra were processed with Sparse Multidimensional Fourier Transform (the software for data processing is available online at the Warsaw University Laboratory (nmr.cent3.uw.edu.pl/software)). All spectra were analyzed with the graphical NMR assignment and integration software Sparky 3.115.

Analytical Size Exclusion Chromatography and Native Gel Titration

LBD at a concentration of 30 μM was incubated with 600 μM LC8 at various molar ratios: (LBD:LC8) 1:3, 1:5, 1:7, 1:9, 1:11, and 1:13. The complex was run on a Superdex 200 analytical column (GE healthcare) in a buffer containing 50 mM phosphate, 50 mM NaCl, 5 mM beta-mercaptoethanol, 1 mM sodium azide, pH 7.8. 100 or 200 μl of protein samples were injected at a flow rate of 0.5 ml/min at room temperature and samples were monitored by UV absorption at 280 nm.

For native gel titrations, dLBD or LBD and LC8 were incubated at the molar ratios listed above and run on a 10% polyacrylamide gel at a constant 10 mAmps for 5-7 hours.

Small angle X-ray scattering

Small-angle X-ray scattering experiments were conducted at the ESRF BioSAXS beamline BM29[203] in Grenoble, France. dLBD and LC8 samples were purified as described above and dialyzed into binding buffer (50 mM sodium phosphate, 50 mM sodium chloride, 10 mM beta-mercaptoethanol, 1 mM sodium azide, pH 7.5) before SAXS measurements. 30 μl of dLBD:LC8 complex (1:8 molar ratio) at five different concentrations for each sample (and buffer) were

exposed to X-rays and scattering data collected using the robotic sample handling available at the beamline. 10 individual frames were collected for every exposure, each 2 seconds in duration using the Pilatus 1M detector (Dectris). Individual frames were processed automatically and independently within the EDNA framework, yielding individual radially averaged curves of normalized intensity versus scattering angle $s=4\pi SIN\theta/\lambda$. Additional data reduction within EDNA utilizes the automatic data processing tools of EMBL-Hamburg ATSAS package[204], to combine timeframes, excluding any data points affected by aggregation induced by radiation damage, yielding the average scattering curve for each exposure series. Matched buffer measurements taken before and after every sample were averaged and used for background subtraction. Merging of separate concentrations and further analysis steps were performed manually using the tools of the ATSAS package[204]. The forward scattering I(0) radius of gyration, Rg were calculated from the guinier approximation[205], the hydrated particle volume was computed using the Porod invariant[206] and the maximum particle size $D_{max,}$ was determined from the pair distribution function computed by GNOM[207] using PRIMUS.

Electron Microscopy

Electron microscopy (EM) studies were conducted using dLBD and LBD peptides incubated with a molar excess of LC8, and the formed complexes were negatively stained for contrast enhancement using established protocols[208]. Briefly, dLBD (50 nM) was mixed with LC8 at a molar ratio 1:8, and human LBD peptide (50 nm) was mixed with LC8 at a molar ratio of 1:13) in EM buffer containing 20 mM Tris, pH 7.5, 50 mM NaCl, 10 mM BME and 1 mM NaN_3. A 3 µl drop of sample was applied to a glow-discharged continuous carbon coated EM specimen grid (400 mesh Cu grid, Ted Pella). Excess protein was removed by blotting with filter paper and washing the grid two times with EM buffer. The specimen was then stained with freshly prepared 0.75% (wt vol^{-1}) uranyl formate (SPI-Chem).

Negatively stained specimens were visualized on a 120 kV TEM (iCorr, FEI) at a nominal magnification of 49,000x at the specimen level. Digital micrographs were recorded on a 2K x 2K CCD camera (FEI Eagle) with a calibrated pixel size of 4.37 Å $pixel^{-1}$ and a defocus of 2.0 – 3.5 um. For the dLBD-LC8 specimen, a total of 2,574 single particle images were extracted from ~300 micrographs, and for human dLBD-LC8 1,234 particles were extracted from ~200 micrographs. Complexes with clear oligomeric structure could be identified and were hand-selected using EMAN2[209]. Single particle images were extracted with a box size of 160 x 160 pixels and CTF-corrected (phase-flipped) in EMAN2. Reference-free 2D class averages were

generated in EMAN2 and RELION 2.0[210] using CTF-corrected and high-pass filtered image datasets. Statistical analysis of oligomeric composition for both datasets was performed by counting the number of subunits identified from single particle images and classifying them manually as 2 – 7mers (dLBD:LC8 complexes) or 2 – 11mers (LBD:LC8 complexes) (SI Figure 4.3). Particles that could not be confidently assigned were discarded, leaving 2,334 oligomers assigned for the dLBD:LC8 dataset and 967 for human LBD:LC8 dataset. Complexes containing only a single LC8 dimer could not be distinguished from unbound LC8 particles, and were not included in our analysis.

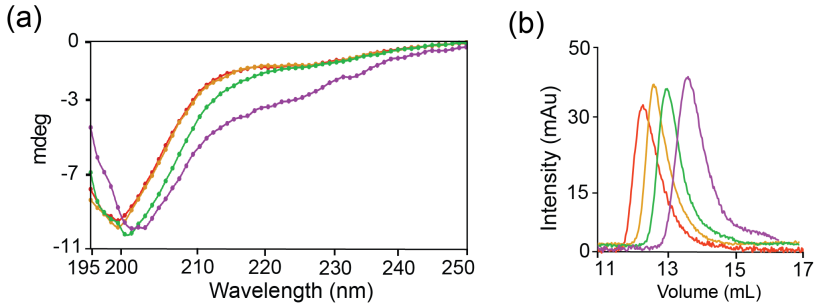
As a positive control, Nucleoporin159 (Nup159) in complex with LC8 was also prepared for negative stain EM under similar conditions to the dLBD/LBD samples, and as previously described[38] (not shown). As a negative control, we prepared EM grids with LC8 alone and dLBD/LBD peptides alone. No oligomeric structures (*i.e.* beads on a string) were observed in these images (not shown).

Transcription Reporter Assays

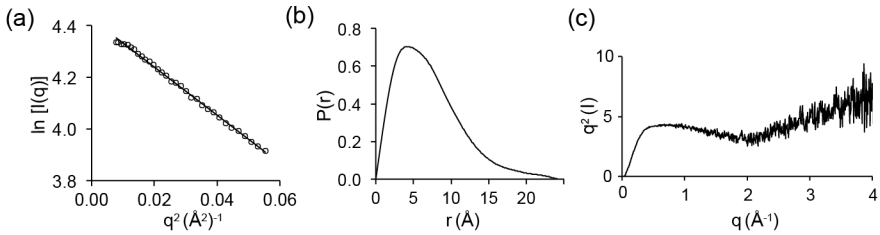
To measure transcriptional activity of ASCIZ mutants, six ASCIZ constructs were cloned into the pEGFP vector (Clontech): WT ASCIZ (1-823), ΔZnF (230-823), ASCIZ AAA1-4, ASCIZ AAA8-11, ASCIZ AAA5-11, and ASCIZ AAA1-4, 8-11. Approximately 2 kbp of the Dynll1 promoter was cloned into the pGL3 vector (Promega) upstream of the firefly luciferase gene as previously described [56]. Using FuGENE 6 (Promega), immortalized ASCIZ KO MEFs [59] were co-transfected with ASCIZ constructs, the Dynll1 promoter, and a pRL-CMV vector containing *Renilla* luciferase for normalization of firefly/luciferase ratios. 24 hours after transfection, cells were transferred to 96-well plates and incubated overnight before determining reporter gene activities using the dual-luciferase reporter assay kit (Promega) and a Polarstar Optima (BMG Labtechnologies) instrument. For assessment of protein expression levels, human U2OS cells were transfected with ASCIZ constructs using FuGENE 6 and were probed with ASCIZ antibody [211].

Acknowledgements

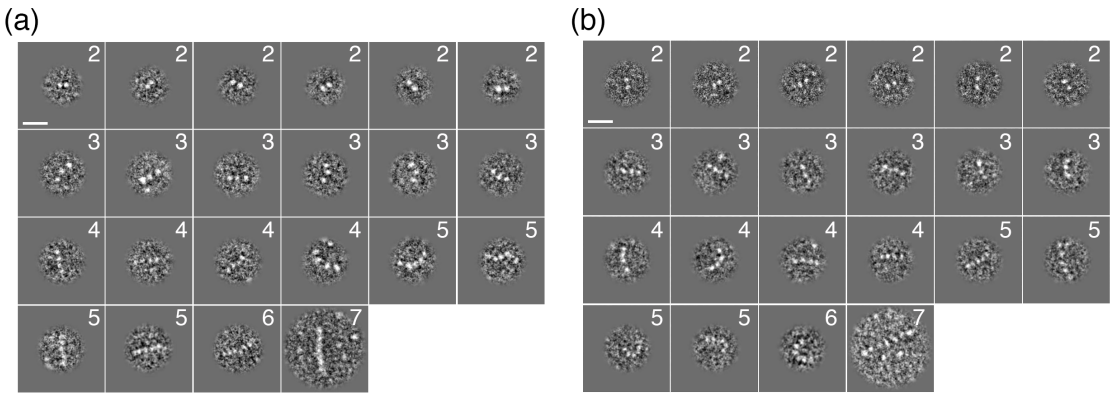
This work was supported by National Institutes of Health Grant GM 084276 to E.B. Support to facilities includes the Oregon State University NMR Facility funded in part by the National Institutes of Health, HEI Grant 1S10OD018518, and by the M. J. Murdock Charitable Trust grant # 2014162. Access to the NMR facility of CEITEC Masaryk University was provided by iNEXT, project number 653706, funded by the Horizon 2020 programme of the European Union. This article reflects only the author's view and the European Commission is not responsible for any use that may be made of the information it contains. JH was supported by the National Health and Medical Research Council of Australia (Senior Research Fellowship APP1022469 and Project Grant APP1026125) and Victorian State Government Operational Infrastructure Support. Electron microscopy was conducted with support from the Multiscale Microscopy Core (OHSU), Advanced Computing Center (OHSU) and the National Institutes of Health Grant R35GM124779 to SLR. Small angle X-ray scattering data was collected at the European Molecular Biology Institute in Grenoble, France.



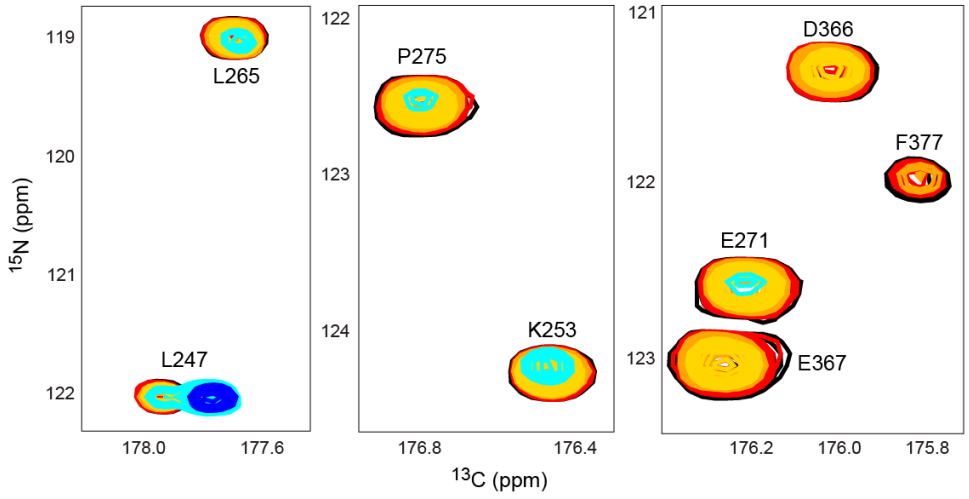
SI Figure 4.1. dLBD C-terminus is transiently structured.
(a) Far UV CD spectrum of dLBD constructs collected at 10 °C: QT1-3 (red), QT2-4 (orange), QT4-6 (green), and QT4-7 (magenta). (b) Size exclusion chromatography of each construct from (a) depicts differences in retention time, using the same color scheme.



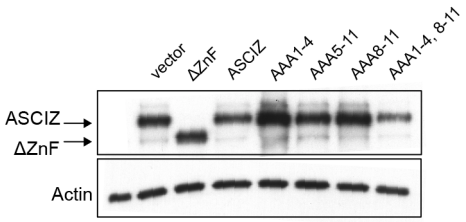
SI Figure 4.2. Primary SAXS data.
(a) A Guinier plot of the experimental scattering data is shown. The solid line represents the Guinier fit, where the linear fit was extended to $q < 1.3/R_g$ [212]. (b) The real-space pairwise distribution function for the dLBD:LC8 complex. (c) Kratky plot of the dLBD:LC8 complex.



SI Figure 4.3. Single particle images of dLBD:LC8 and LBD:LC8 complexes.
(a) Representative images of negatively stained (a) dLBD:LC8 complexes and (b) LBD:LC8 complexes extracted from raw micrographs. Each particle is annotated to indicate the assigned occupancy of LC8 dimers used for statistical analysis in Figure 4.5. Scale bars = 20 nm.



SI Figure 4.4. Representative HNCO slices of dSQTQ titration with LC8.
Shown are representative peaks from HNCO spectra depicting a loss of intensity with an increasing LC8 concentration. Colors are the same as Figure 4.6a: free dSQTQ (black), 1 dSQTQ: 0.25 LC8 (red), 1:1 (orange), 1:2 (gold), 1:5 (blue), and 1:8 (dark blue).



SI Figure 4.5. Western blot of ASCIZ constructs.
Western blot analysis of human U2OS cells transiently transfected with ASCIZ constructs from (Figure 4.8c), probed with ASCIZ antibody.

Chapter 5

Nucleoporin 159 Forms a Stable, Rigid Assembly with LC8

Sarah A. Clark, Grant Pearce, and Elisar Barbar

Introduction

Nucleoporin 159 (Nup159) is a core structural component of the yeast nuclear pore complex and a LC8 binding partner. The interaction between Nup159 and LC8 (Dyn2 in yeast) is necessary for stabilization of the Nup82-Nsp1-Nup159 complex in yeast[38]. Previous work showed that Nup159 contains six LC8 recognition motifs in an intrinsically disordered region, five of which bind to LC8 (Figure 5.1a)[40]. The Nup159:LC8 complex forms a rigid, rod-like structure, where five LC8 dimers are assembled like beads on a string between two disordered Nup159 chains (Figure 5.1b). The bivalent complex can clearly be seen by negative stain electron microscopy[38], and is so stable that it has been used as a tag for subunit identification in electron microscopy[136].

Here we seek to understand the process of complex assembly and characterize the structure of the fully-bound 5-mer complex. Thermodynamic experiments show that the Nup159:LC8 complex is most stable when only three of the five recognitions motifs are bound by LC8[36]. This result hints at the presence of low occupancy intermediates that contribute to the complex assembly process. In this work, we use a combination of biophysical techniques to assess whether these intermediates exist and to determine the order of binding among the five sites. Our results indicate that assembly occurs through random sampling of all five sites, resulting in formation of a low stoichiometry complex in addition to a highly populated 5-mer structure.

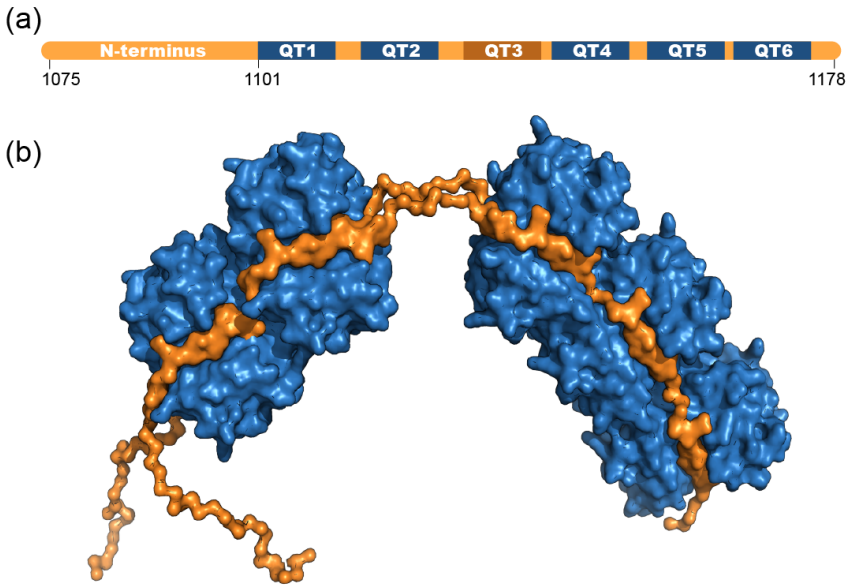


Figure 5.1. Model of the Nup6:LC8 complex.
(a) Construct used in this work. The intrinsically disordered domain of Nup6 is shown in orange, with six LC8 recognition motifs highlighted in blue. (b) A model of the fully-bound Nup6:LC8 complex showing 5 LC8 molecules (blue), bound to two chains of Nup6 (orange).

Results

Nup6 and LC8 form a rigid, 5-mer complex

All experiments were performed using Nup6, a construct which corresponds to Nup159 residues 1075-1178 and contains all 5 LC8 recognition motifs (Figure 5.1a). To assess the process of complex formation, we titrated an increasing amount of LC8 into Nup6 and monitored complex size using analytical ultracentrifugation (AUC). The molar mass of free LC8 determined from sedimentation velocity was 23 kDa, closely matching the theoretical molecular weight of 24 kDa. Free Nup6, with its theoretical molecular weight of 12 kDa, was too small to be measured by AUC. We titrated an increasing amount of LC8 into Nup6 and measured complex size at 5 molar ratios of Nup6 to LC8: 1:0.5, 1:1, 1:2, 1:4, and 1:8 (Figure 5.2a). A broad peak is seen at sub-saturating concentrations of LC8 that is likely a mixture of different complex stoichiometries. At a ratio of 1:8, where LC8 is present at a saturating concentration, two peaks can be seen. The

larger peak roughly corresponds to a molar weight of 142 kDa, which matches the expected size of a 1:5 complex. The smaller peak likely corresponds to a low occupancy 3-mer complex, where 3 of the 5 LC8 binding sites are occupied. Previous ITC experiments demonstrated that a 3-mer complex is more thermodynamically stable than a 5-mer complex due to the entropic cost associated with binding 5 tandem LC8 sites[36].

Native gel titration at the same molar ratios provides additional insight into complex formation (Figure 5.2b). A discreet, high molecular weight band that is likely the fully-bound complex can be seen at the first titration point and it increases in intensity as more LC8 is added. A lower molecular weight smear indicates the presence of multiple complexes with varying stoichiometries, as is seen by AUC. Interestingly, excess Nup6 persists until a saturating 1:5 ratio is reached, suggesting that the low stoichiometry complexes are less stable than the full-bound complex, and thus pull away from the other complexes by mass action. Likewise, free LC8 appears at a ratio of 1:2 by AUC and can also be seen by native gel. Overall, these results demonstrate that the complex forms by sampling multiple stoichiometries instead of through discreet, stepwise process. Both a fully-bound 5-mer and low occupancy 3-mer complex are stable at saturating LC8 concentrations, although the 5-mer complex is the dominant species.

Small angle X-ray scattering data collected on the Nup6:LC8 complex demonstrates that the proteins form an elongated, rigid structure. Guinier analysis of the data reveals that the sample is monodisperse and suitable for further analysis (Figure 5.2c). Analysis of the distance distribution function indicates that the proteins form an elongated structure with D_{max} = 139 Å and a molecular weight of roughly 160 kDa, which corresponds to a 5-mer complex (Figure 5.2d). A Kratky plot of the scattering data shows that the complex is a mix of globular domains and intrinsic disorder, although it is most similar to the profile for a globular protein. This suggests that the Nup6:LC8 complex is a rigid, bivalent scaffold with little flexibility in the linker regions.

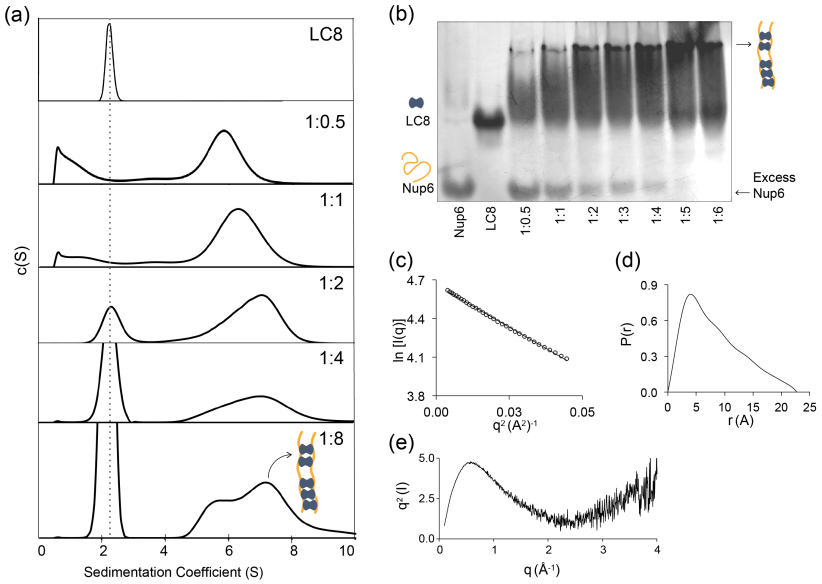


Figure 5.2. Nup6:LC8 complex assembly.
(a) Representative *c(S)* distributions obtained by sedimentation velocity are shown for LC8, and increasing molar ratios of 1 Nup6: 0.5 LC8; 1:1, 1:2, 1:4, and 1:8. Cartoon depictions of Nup6 (orange) and LC8 (dark blue) are shown to aid in visualization of the complexes formed at a 1:8 ratio. (b) Native gel titration of Nup6 with LC8. An increasing concentration of LC8 was added to a constant amount of Nup6, from a molar ratio of (Nup6:LC8) 1:0.5 to 1:6. As more LC8 is added, the high molecular weight complex increases in intensity and the amount of free Nup6 decreases. Primary SAXS data. (c) A Guinier plot of the experimental scattering data is shown. The solid line represents the Guinier fit, where the linear fit was extended to $q < 1.3/R_g$ [212]. (d) The real-space pairwise distribution function for the Nup6:LC8 complex. (d) Kratky plot of the Nup6:LC8 complex.

NMR analysis of Nup6:LC8 complex formation

Previous ITC experiments showed that all 5 recognition motifs bind to LC8 weakly with K_d values >13.1 μM[40]. The first three motifs bind to LC8 more tightly than all five (K_d 0.8 μM vs 2.9 μM, respectively)[36], suggesting that one or more of these motifs bind to LC8 first, creating a bivalent scaffold that nucleates complex formation. To assess the order of binding among the five LC8 recognition motifs, we turned to NMR. Unlabeled LC8 was titrated into ^{1}H-^{15}N labeled Nup6 and 2D HSQC spectra were recorded for molar ratios of (Nup6:LC8) 1:0.1, 1:0.3, 1:0.5, 1:1, 1:3, 1:4, and 1:6 (Figure 5.3a). At the first titration point we saw a uniform decrease in intensity across all five recognition motifs. Addition of more LC8 resulted in a further loss of

intensity, likely due to line broadening associated with intermediate exchange processes and/or faster transverse relaxation as a result of increased complex size (Figure 5.3b). At a ratio of 1:3, 50% of the peak intensity across the five motifs had disappeared and at a ratio of 1:4, only the N-terminal and C-terminal peaks remained. Interestingly, the N-terminal peaks increased in intensity in the fully-bound 5-mer complex, implying an increase in flexibility upon complex formation[213]. The results of this titration imply that there is no binding preference among the five recognition motifs.

NMR dynamics experiments corroborate the titration results, demonstrating that there is no motional heterogeneity in Nup6 that could contribute to preferential recognition of one site over another. A plot of ^{15}N transverse (R_2)/ longitudinal (R_1) values vs. residue position shows a 1.0-2.1 range with an average value of 1.5 that is consistent across all 103 residues (Figure 5.3c). ^{1}H-^{15}N heteronuclear NOE values measured at 25°C are negative overall, with an average value of -0.5, revealing a high degree of flexibility (Figure 5.3d). Together, the R_2/R_1 and heteronuclear NOE values indicate that Nup6 is motionally homogeneous and highly flexible.

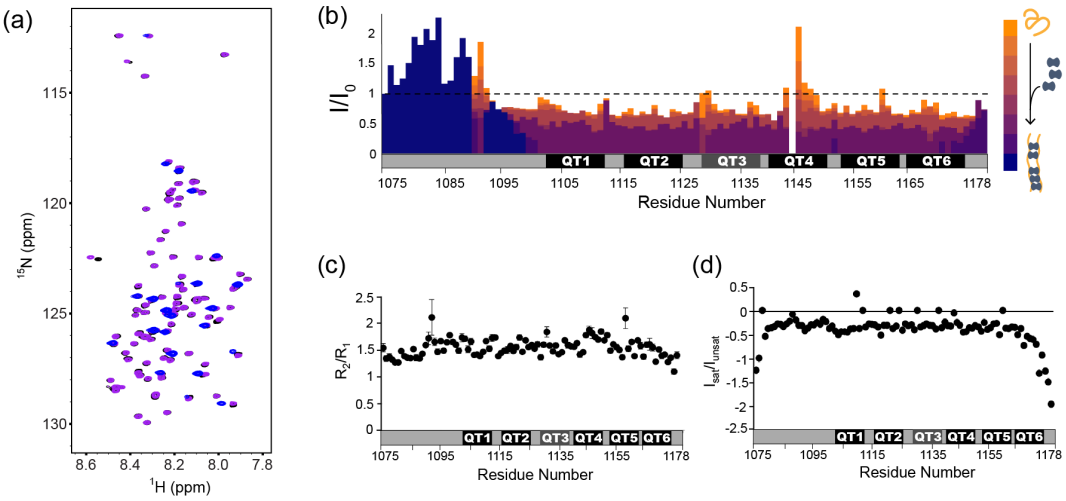


Figure 5.3. NMR titration of Nup6 with LC8.
(a) ^{1}H-^{15}N HSQC spectrum of Nup6 at 700 mHz. Free Nup6 (black) is shown overlaid with spectra from two titration points: 1 Nup6: 3 LC8 (purple) and 1:6 (blue). (b) Relative intensities of non-proline NH peaks in ^{15}N-^{1}H HSQC spectra of Nup6 were obtained at different molar ratios of Nup6: LC8; 1:0.1, 1:0.3, 1:0.5, 1:1, 1:3, 1:4, and 1:6 and plotted against residue number. (c) Plots of R_2/R_1 and (d) heteronuclear NOE values measured at 25 °C per residue. Segments corresponding to recognition motifs, QT1, QT2, QT3, QT4, QT5, and QT6 are shown below the plot. QT3 is shown in a shade of gray to indicate that it does not bind LC8.

Discussion

Here we characterize the structure and assembly process of the Nup6:LC8 complex. AUC and native gel data indicate that the fully-bound 5-mer complex is the most stable, populated species in solution, although a lower stoichiometry 3-mer complex also exists. SAXS demonstrates that this 5-mer is both elongated and rigid, as has also been shown by negative stain electron microscopy[38].

Our results suggest a model of complex assembly where Nup6 and LC8 sample multiple stoichiometries before forming the stable 5-mer complex. The low occupancy intermediate likely contributes to the assembly process by stabilizing the bivalent scaffold so that subsequent LC8 molecules can bind. NMR titration of Nup6 with LC8 demonstrate that LC8 does not preferentially bind to any of the five recognition motifs and instead randomly samples different sites until the final complex is formed.

Curiously, the complex appears to maintain register during the assembly process instead of forming aggregate species or phase-separated granules, as one might expect from studies of similarly intrinsically disordered multivalent proteins[214]. Future work should explore how register is maintained when there is no preference among all five equally weak binding sites. It is possible that this is a unique characteristic of LC8 binding and is thus important for the formation of other multivalent partner:LC8 complexes.

Materials and Methods

Protein Expression and Purification

The Nup6 construct contains 6 TQT motifs and includes residues 1075-1178 (QT1-6). Recombinant LC8 and Nup6 were expressed and purified as previously described[36].

Analytical Ultracentrifugation

Sedimentation velocity experiments for the titration of Nup6 and LC8 were performed in a Beckman Coulter Model XL-I analytical ultracentrifuge equipped with UV/Vis scanning optics. Reference (400 μL; 50 mM sodium phosphate, 50 mM sodium chloride, 1 mM sodium azide, 5 mM TCEP, pH 7.5) and sample (380 μL) solutions were loaded into 12 mm double-sector cells with quartz windows and the cells were then mounted in an An-50 Ti 8-hole rotor. LC8 was prepared at a concentration of 15 μM while the concentration of Nup6 was varied from 15 – 1.5 μM. Proteins were centrifuged at 50,000 rpm at 20 °C, and radial absorbance data were collected at appropriate wavelengths in continuous mode every 5 minutes without averaging. Data were fit

to a continuous size-distribution [c(S)] model using the program SEDFIT[198]. The partial specific volume of the proteins, buffer density, and buffer viscosity were computed using the program SEDNTERP[199].

NMR Data Collection and Analysis

NMR samples were prepared in 10 mM sodium phosphate, 10 mM NaCl, 5mM TCEP, 1 mM sodium azide, pH 6.0 buffer containing 1mM DSS and 10% D_2O. For titration experiments, unlabeled LC8 was mixed with isotopically labeled Nup6 at specific molar ratios and allowed to equilibrate before spectra collection. Data were acquired on a Bruker Avance 700 MHz spectrometer equipped with a $^{1}H/^{15}N$ BBI probehead with the z-axis gradient coil at 25°C. Two-dimensional ^{1}H-^{15}N HSQC spectra were collected for each titration point and processed using TopSpin software (version 3.1.1). Assignments of ^{1}H-^{15}N Nup6 titration spectra were performed using spectra from Nyarko et al. 2013[36]. All spectra were analyzed using graphical NMR assignment and integration software Sparky 3.115[170].

NMR dynamics experiments were acquired on a Bruker Avance 900 MHz spectrometer equipped with a cryoprobe. R_1 relaxation measurements experiments were recorded with relaxation delay times ranging from 10 to 1300 ms, and the R_2 relaxation data were acquired using relaxation delays ranging from 29.4 to 382.7 ms. Curve fitting was performed using the rate analysis script Sparky2Rate and the program Curvefit (A. G. Palmer, Columbia University). Steady-state $^{1}H-^{15}N$ heteronuclear NOEs were acquired using 6 s total saturation time. Error bars were determined from the intensities of the baseline noise using the formula $\sigma/(NOE) = [(\sigma I_{sat}/I_{sat})^2 + (\sigma I_{unsat}/I_{unsat})^2]^{1/2}$, where I_{sat} and σI_{sat} correspond to the intensity of the peak and its baseline noise.

SAXS Data Collection and Analysis

Small-angle X-ray scattering experiments were conducted at the ESRF BioSAXS beamline BM29[203] in Grenoble, France. All samples were purified by His-affinity and amylose affinity chromatography followed by size exclusion chromatography before SAXS measurements. 30 µl of protein solution at three different concentrations for each sample (and buffer) were exposed to X-rays and scattering data collected using the robotic sample handling available at the beamline. 10 individual frames were collected for every exposure, each 2 seconds in duration, using the Pilatus 1M detector (Dectris). Individual frames were processed automatically and independently within the EDNA framework, yielding individual radially averaged curves of normalized

intensity versus scattering angle $s=4\pi SIN\theta/\lambda$. Additional data reduction within EDNA utilizes the automatic data processing tools of EMBL-Hamburg ATSAS package[204] to combine timeframes and exclude any data points affected by aggregation induced by radiation damage, yielding the average scattering curve for each exposure series. Matched buffer measurements taken before and after every sample were averaged and used for background subtraction. Merging of separate concentrations and further analysis steps were performed manually using the tools of the ATSAS package[215]. The forward scattering I(0) and radius of gyration Rg were calculated from the Guinier approximation[205], the hydrated particle volume was computed using the Porod invariant[206] and the maximum particle size D_{max} was determined from the pair distribution function computed by GNOM[207] using PRIMUS.

Chapter 6

Conclusion

Impact

The studies presented here have answered key questions regarding the function and assembly of LC8-partner complexes that contain multiple recognition motifs. These complexes, termed IDP duplex scaffolds, contain many unique structural traits that enable diverse functions in a variety of cellular processes. This thesis work, together with one review, focuses largely on the biophysical analysis of LC8 partner proteins using a variety of techniques including NMR, electron microscopy, protein crystallography, molecular biology, and isothermal titration calorimetry. These results have expanded our understanding of the types of assemblies that LC8 forms with its partner proteins, and have illuminated the novel method of negative feedback regulation that aids in maintaining the cellular concentration of LC8. The important features of each chapter are highlighted below. As each chapter contains information pertaining to a different LC8 binding partner, the similarities and differences between these complexes is also discussed. The remainder of the chapter is dedicated to outlining ongoing work and suggestions for future studies.

Highlights of Reported Work

The purpose of the review (Chapter 2) is to place intrinsically disordered protein (IDP) duplexes in context of the larger field of multivalent IDP assemblies. LC8 binds to 58 known intrinsically disordered partner proteins and is predicted to bind to dozens more, emphasizing the importance of this type of complex. Highlighting the unique features of LC8-associated IDP complexes is important to better understand the function of the complex in these diverse systems.

In Chapter 3, we explore the importance of the motif sequence. There are two important features of this work. First, we introduce the '*anchored flexibility model*' of LC8 motif recognition wherein the TQT residues of the motif serve as the anchor for binding, and the remaining seven motif residues modulate affinity. We find that the TQT residues are highly conserved because the T's are completely buried in the LC8 binding pocket and the Q engages in hydrogen bonding with both LC8 monomers. Second, by analyzing LC8 binding to the multivalent protein Chica, we pinpoint the elements that lead to increased or decreased motif affinity in Chica's three recognition motifs. We conclude that the factors which influence motif affinity are highly complex and difficult to predict with our current level of understanding. Development of a bioinformatics tool that allows one to determine whether a given motif sequence binds to LC8 would be extremely valuable.

The work in Chapter 4 contributed two important advances to our understanding of LC8-partner protein complexes. First, we discovered that LC8 and its multivalent partner ASCIZ form a highly dynamic ensemble of complexes, instead of a single discreet complex as had been observed with other partner proteins. LC8 cooperatively binds to the eleven recognition motifs in human ASCIZ, or the seven motifs in *Drosophila* ASCIZ, forming a mixture of complexes with different stoichiometries. Interestingly, a low occupancy complex is most stable and most highly populated in both species. We speculate that this low-occupancy complex is important for dimerizing ASCIZ and maintaining a basal level of LC8 transcription. Second, we found that the activity level of ASCIZ is determined by the number of LC8 dimers bound. When many LC8 motifs are occupied, transcription is reduced, and when fewer sites are occupied, transcription is increased. This tunable gradient of transcriptional activity is likely critical for maintaining a precise balance of LC8 in the cell. While many transcription systems rely on multiple sites for transcription regulation, we could find no examples where a transcription factor is negatively regulated by binding to many copies of its gene product. As multivalency and intrinsic disorder are common among eukaryotic transcription factors [177, 178], we expect this mechanism is utilized in other systems.

Chapter 5 is centered on the biophysical characterization of a third multivalent binding partner, Nup159. Nup159 and LC8 have previously been shown to form a rigid, rod-like structure where LC8 binds to five tandem repeats in an intrinsically disordered region[36, 38]. We discovered that LC8 randomly samples all five sites during complex assembly instead of preferentially binding to one site, and showed that a low-occupancy complex is populated in addition to the more stable fully-bound complex. This work answered several outstanding questions that were prompted by thermodynamic characterization of the Nup159:LC8 complex[36].

Comparative Analysis of LC8-Multivalent Partner Protein Assemblies

The four multivalent proteins presented in this thesis, *Drosophila* ASCIZ (dASCIZ), human ASCIZ, Chica, and Nup159, all form IDP duplex scaffolds when bound to LC8. The ASCIZ and Nup159 duplex structures can clearly be seen by negative stain electron microscopy[38], while the Chica:LC8 structure is inferred from X-ray crystallography of LC8 bound to Chica peptides[30]. Constructs corresponding to the LC8 binding domain of each of these proteins (*Drosophila* ASCIZ residues 241-388, human ASCIZ residues 362-823, Chica residues 410-478, and Nup159 residues 1075-1178), bind to LC8 with a similar affinity: 1.4 μM for dASCIZ, 1.0 μM for human ASCIZ, 0.4 μM for Chica[30], and 2.9 μM for Nup159[36]. This is an interesting result due to the

disparate number of recognition motifs in each protein, which ranges from three in Chica to eleven in human ASCIZ. Nup159 and both ASCIZ proteins bind to LC8 cooperatively, meaning that the individual motifs are weak and bind more tightly in the context of the full-length protein[36]. This binding enhancement is limited to a few motifs, however. For Nup159 and dASCIZ, binding affinity increases for the first 3 motifs, then decreases as subsequent motifs are bound. Chica differs from these two proteins in that the first motif binds to LC8 most tightly and the remaining two motifs bind 10x weaker[30].

The process of complex assembly with LC8 varies for each of these proteins. Chica binds LC8 most tightly at its first motif (QT1), suggesting that this motif binds first. Once the bivalent scaffold is formed, the two other weak motifs can bind more easily to form the fully-bound 3-mer complex[30]. The intrinsically disordered Nup159 chain is motionally homogeneous (Figure 5.3) and all five motifs bind LC8 with an equally weak affinity[38]. These results, together with NMR titration experiments (Figure 5.3), support a model of complex assembly wherein LC8 randomly samples all five sites until the fully-bound 5-mer complex is formed. This final complex is an elongated, rigid assembly as observed by electron microscopy[38] and SAXS experiments (Figure 5.2). dASCIZ, on the other hand, is motionally heterogeneous in that the C-terminal half of the protein is more restricted then the N-terminal half. Additionally, the C-terminal sites (QT4-7) bind to LC8 more tightly than then N-terminal three sites (QT1-3), and NMR titration demonstrates that these sites bind to LC8 first (Figure 4.6). Interestingly, neither dASCIZ nor ASCIZ readily forms a stable, fully-bound complex with LC8. Both ASCIZ proteins form a dynamic mixture of complex stoichiometries that is not observed for either Nup159 or Chica. For ASCIZ, a low-occupancy complex dominates in solution, suggesting that it is more stable than the fully-bound.

Intriguingly, the biophysical differences between ASCIZ, Nup159 and Chica are reflective of their cellular roles. The Nup159:LC8 complex is a structural component of the yeast nuclear pore complex that is necessary for nuclear pore stability[38]. The rigidity of the Nup159:LC8 complex is likely important for stabilizing the pore and is hypothesized to aid in projecting the FG repeats out of the pore to enable protein transport[38]. In contrast, ASCIZ is a transcription factor whose activity is modulated by the number of LC8 dimers bound. In order to enable fast response to changing cellular conditions, LC8 must bind and dissociate quickly from ASCIZ's disordered domain. This function would require a high degree of plasticity in ASCIZ, which is seen in its dynamic nature and reluctance to form a stable, fully-bound complex.

Although the functional consequences of LC8 binding to Chica are unknown, its role in stabilizing mitotic spindles suggests the formation of a more rigid complex.

The remarkable functional and structural diversity of LC8-partner complexes highlights the flexibility of LC8 as a hub protein, as well as underscores the need for more studies aimed at elucidating the overall cellular role of LC8. For many binding partners, such as Chica, it is known that interaction with LC8 is necessary for proper cellular function, but the cellular-level changes that result from LC8 binding are unclear.

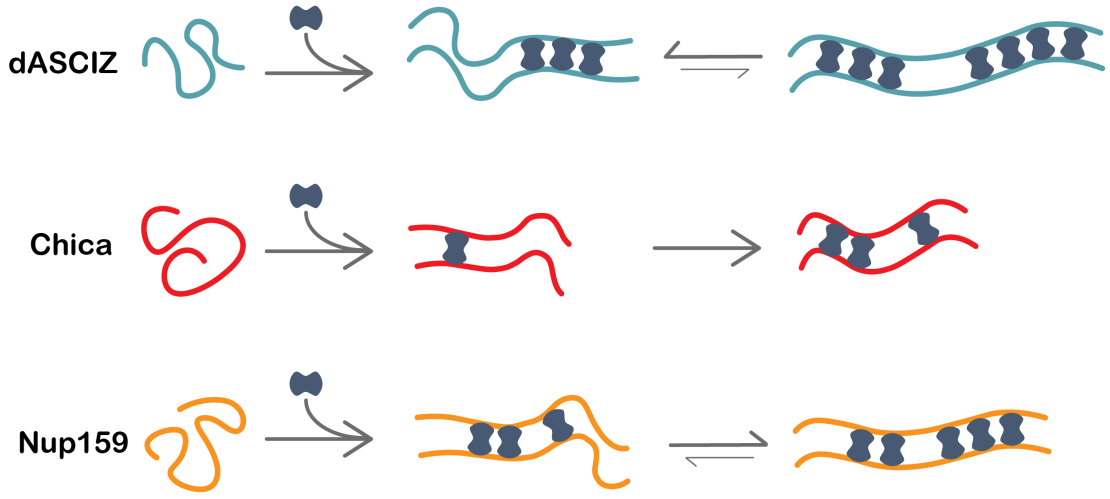


Figure 6.1. Assembly process of three multivalent LC8 partner proteins.
Cartoon schematics are shown for dASCIZ, Chica, and Nup159. LC8 binding to each of these proteins results in the formation of an IDP duplex (right). dASCIZ (blue) forms a stable, partially bound complex with LC8, in addition to a small population of fully-bound complex. Chica binds to LC8 tightly at the first motif and forms a stable fully-bound complex. Nup159 preferentially forms a fully-bound complex with LC8, although a small fraction of low-occupancy complex is also observed. For Nup159 and dASCIZ, the low occupancy complex (center) depicts one of many possible complexes (i.e. sites QT4,5,6 in dASCIZ are shown as occupied, but sites QT4,5,7 are equally as likely to be occupied).

Ongoing Work

ASCIZ Zebrafish Studies

ASCIZ knockout phenotypes have been studied in mouse[57-59] and *Drosophila* models[60]. In Summer 2017 I began work assessing the phenotype of zebrafish in which ASCIZ has been knocked down using a morpholino. Initially, I confirmed that the zebrafish ASCIZ (zASCIZ) gene and LC8 gene products could be quantified by real time PCR (rt-PCR) using mRNA collected from wild-type zebrafish embryos. Both gene products were readily detectable and I found that the zASCIZ mRNA transcript increases steadily over the course of development by comparing mRNA from 24 hr, 48 hr, and 5-day-old zebrafish.

Two morpholinos were tested for their ability to knockdown ASCIZ: e1i1, which targets the first intron/exon splice junction in the zASCIZ gene and e3i3, which targets the third splice junction. Binding of the morpholino to zASCIZ mRNA should result in the incorporation of intron 1 or 3, and thus an incorrect protein product. Both morpholinos were injected into zebrafish embryos at three concentrations: 1.5 mM (stock concentration), 1.0 mM, and 0.75 mM. E3i3 did not produce notable phenotypic effects even at the highest concentration, while injection of the e1i1 morpholino resulted in highly deformed embryos in which the abnormalities were apparent 48 hours post injection. The 1.0 mM concentration was chosen for further studies as the deformities were more obvious than the 0.75 mM concentration, yet not lethal in 48 hours as with the 1.5 mM concentration. Additionally, PCR was used to test both morpholinos for their effect on the zASCIZ mRNA (i.e. whether intron 1 or 3 was incorporated). mRNA was extracted from 24 and 48 hour zebrafish embryos injected with the e1i1 and e3i3 morpholinos and converted to cDNA. The presence or absence of intron 1 or intron 3 was assessed by performing PCR with primers that span the intron. Morpholino e3i3 effectively incorporated intron 3, while morpholino e1i1 incorporated only part of intron 1. Sequencing analysis of intron piece that was incorporated indicates that it is sufficient to frameshift the rest of the sequence, resulting in a non-functional zASCIZ protein.

Rescue of the zASCIZ knockdown phenotype requires co-injection with LC8 or zASCIZ mRNA with the 1.0 mM e1i1 morpholino. I began by injecting LC8 in two amounts: 50 pg and 100 pg (2-4 nL injection of 500 ng/μL mRNA per embryo). The 50 pg injection did not alter the knockdown phenotype. The 100 pg injection showed promise in that more of the zebrafish appeared to be phenotypically normal than the 1.0 mM morpholino control, but it did not fully rescue the phenotype.

Further work on this project will require determining the optimal concentration of e1i1 morpholino for injection and the optimal amount of LC8 or zASCIZ mRNA to include for phenotypic rescue. It is possible that an intermediate concentration of morpholino (0.8-0.9 mM) would still produce visible phenotypic defects, but would be easier to rescue with lower concentrations of mRNA. I had difficultly making pure zASCIZ mRNA, which is why rescue experiments with zASCIZ was not attempted. It is possible that it would be more successful than LC8 rescue, however, as zASCIZ performs cellular roles in addition to acting as an LC8 transcription factor.

An interesting avenue for this research project would be to assess whether ASCIZ chimeras are capable of rescuing the knockdown phenotype. One could make a construct with the zASCIZ DNA binding domain and the human ASCIZ LC8 binding domain (or a more distantly related species) and assess its function. Additionally, attempted rescue with only the zASCIZ DNA binding domain or a construct containing part of the LC8 binding domain would provide insight into the importance of the intrinsically disordered domain in transcriptional regulation.

Identification of ASCIZ Interacting Partners using Mass Spectrometry

In collaboration with Dr. Larry David at the OHSU Mass Spectrometry Center, I have been working to identify binding partners of human ASCIZ. Previous work identified acidic/hydrophobic motifs in ASCIZ's intrinsically disordered domain that likely bind to general transcription factors (Figure 6.2a). This type of motif interacts with activating transcription factors in other systems, driving transcription forward[216-220]. These motifs overlap with LC8 binding motifs in the human and *Drosophila* ASCIZ proteins (Figure 6.2a,b), suggesting that competition between transactivation factors and LC8 modulate ASCIZ activity. The presence of hydrophobic patches is conserved throughout ASCIZ proteins in the animal kingdom (Figure 6.2c), indicating that they play an important functional role. Identification of these interacting partners will provide insight into ASCIZ's system of LC8 transcription regulation and will enable *in vitro* characterization of these transcription factors with ASCIZ.

Initially, I transfected HEK293 cells with MYC-tagged GFP-ASCIZ in the pEGFP vector previously used for transcription assays (Figure 4.7). Transfected cells were harvested after 48 hours and cells were lysed by three rounds of freeze-thaw. Clarified cell lysate was incubated with anti-MYC sepharose overnight at 4°C. The next day, beads were washed with lysis buffer, protein was eluted by heating the resin at 100°C in gel loading dye, and the samples were run into a stacking gel for ~10 minutes. The gel was Coomassie stained and the samples were excised and

sent for mass spectrometry analysis. Unfortunately, this procedure repeatedly yielded a high background and no useable information. It seems that the binding of GFP-ASCIZ to the anti-MYC sepharose was very low, despite a high amount of GFP-ASCIZ in the transfected cells.

In light of these negative results, I turned to the BioID2 method of identifying interaction partners. This method is designed to reveal weak interactions that would normally be difficult to see by conventional pull-down methods[221]. In this method, ASCIZ is expressed as a fusion protein with the BioID2 enzyme, a modified biotin ligase. The cell culture media is supplemented with 50 μM biotin so that, in theory, any protein that contacts ASCIZ will be biotinylated. The biotinylated proteins can then be pulled down using streptavidin-coated beads. This method has multiple advantages over a conventional pull-down method: (1) the ability to identify weak interactions, (2) direct pulldown of interacting partners, enabling use of harsher wash and lysis conditions, and (3) increased pulldown efficiency due to the very strong biotin-streptavidin binding interaction.

Future work on this project will require designing a new GFP-ASCIZ-BioID2 construct, as pulldown experiments with the first construct (BioID2-ASCIZ) was not successful.

(a) dASCIZ$_{241-388}$

MDVSYALEMSSQKLDMETQTEEDDLNEIRNEVLAPLLRDIETQTPDTRGDIGTMTD
DFPEEQEPVAVGSHFHAYSETEPMFDLQTSAHMYTQTCDDLFEELGLSHIQTQTHW
PDGLYNTQHTQTCDEIMDELFPDNFQSTCTQTRWLD

ASCIZ$_{362-823}$

LKESLPLFKIANPIAGEPISTGVQVNFGKSPSNPLQELGNTCQKNSISSINVQTDL
SYASQNFIPSAQWATADSSVSSCSQTDLSFDSQVSLPISVHTQTFLPSSKVTSSIA
AQTDAFMDTCFQSGGVSRETQTSGIESPTDDHVQMDQAGMCGDIFESVHSSYNVAT
GNIISNSLVAETVTHSLLPQNEPKTLNQDIEKSAPIINFSAQNSMLPSQNMTDNQT
QTIDLLSDLENILSSNLPAQTLDHRSLLSDTNPGPDTQLPSGPAQNPGIDFDIEEF
FSASNIQTQTEESELSTMTTEPVLESLDIETQTDFLLADTSAQSYGCRGNSNFLGL
EMFDTQTQTDLNFFLDSSPHLPLGSILKHSSFSVSTDSSDTETQTEGVSTAKNIPA
LESKVQLNSTETQTMSSGFETLGSLFFTSNETQTAMDDFLLADLAWNTMESQFSSV
ETQTSAEPHTVSNF

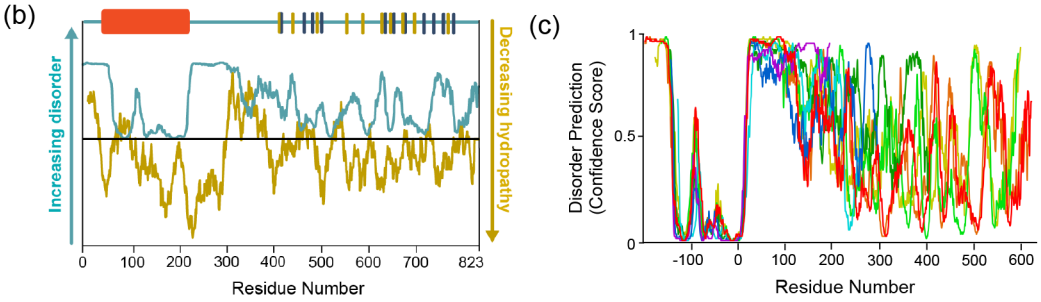


Figure 6.2. Hydrophobic patches are a conserved feature of the LC8 binding domains of ASCIZ proteins.

(a) Putative transactivation motifs were identified by visually inspecting the protein sequence and looking for clusters of 8 or more acidic or hydrophobic amino acids. For each putative transactivation motif, hydrophobic amino acids are highlighted in yellow and acidic amino acids are highlighted in red. LC8 binding sites, identified by pepscan [69], are indicated by black boxes. (b) The disorder prediction for human ASCIZ (blue), obtained using the program Disopred[222], is overlaid with a Kyte-Doolittle hydropathy plot for the same residues (yellow)[223]. A secondary structure schematic of human ASCIZ is shown on top, where dark blue lines indicate the locations of LC8 recognition motifs and yellow lines indicate putative transactivation motifs. (c) Disorder is conserved among multiple ASCIZ proteins from the animal kingdom. Sequence predictions of disorder obtained using the program Disopred [222] are plotted against residue number for 10 ASCIZ proteins. Residue numbers were normalized to human ASCIZ to account for variability in the length of the N-terminal domain. Oscillating behavior of the C-terminal domain is due to the presence of hydrophobic patches.

Future Work

An interesting feature of the complexes formed by LC8 and its multivalent interacting partners that has long been a mystery is how they stay 'on-register' during complex formation. With multiple weak binding sites along a low complexity disordered chain, one might expect that LC8 binding would result in fibers, aggregates, or even phase transitions as a result of LC8 binding to different sites in each groove (e.g. QT1 in binding groove 1, QT5 in binding groove 2). In other systems, phase transitions have been observed for binding of an intrinsically disordered multivalent binding partner to a dimeric ligand[214]. The resulting hydrogels and granules can play a role in increasing the local concentration of biological components[124] and are associated with disease states[224, 225]. These types of structures have not been observed for LC8 partners, despite their biochemical and biophysical similarities. Investigating possible reasons for maintenance of register during complex assembly, such as linker length and linker self-association, would illuminate the functional differences between LC8-partner complexes and other types of multivalent assemblies.

References

[1] Pfister, K. K., Fay, R. B., and Witman, G. B. (1982) Purification and polypeptide composition of dynein ATPases from Chlamydomonas flagella, *Cell motility 2*, 525-547.

[2] Asante, D., Stevenson, N. L., and Stephens, D. J. (2014) Subunit composition of the human cytoplasmic dynein-2 complex, *Journal of cell science 127*, 4774-4787.

[3] King, S. M., Barbarese, E., Dillman, J. F., 3rd, Patel-King, R. S., Carson, J. H., and Pfister, K. K. (1996) Brain cytoplasmic and flagellar outer arm dyneins share a highly conserved Mr 8,000 light chain, *The Journal of biological chemistry 271*, 19358-19366.

[4] Lo, K. W., Kan, H. M., Chan, L. N., Xu, W. G., Wang, K. P., Wu, Z., Sheng, M., and Zhang, M. (2005) The 8-kDa dynein light chain binds to p53-binding protein 1 and mediates DNA damage-induced p53 nuclear accumulation, *The Journal of biological chemistry 280*, 8172-8179.

[5] Fan, J., Zhang, Q., Tochio, H., Li, M., and Zhang, M. (2001) Structural basis of diverse sequence-dependent target recognition by the 8 kDa dynein light chain, *Journal of molecular biology 306*, 97-108.

[6] Wang, L., Hare, M., Hays, T. S., and Barbar, E. (2004) Dynein light chain LC8 promotes assembly of the coiled-coil domain of swallow protein, *Biochemistry 43*, 4611-4620.

[7] Nyarko, A., Hare, M., Hays, T. S., and Barbar, E. (2004) The intermediate chain of cytoplasmic dynein is partially disordered and gains structure upon binding to light-chain LC8, *Biochemistry 43*, 15595-15603.

[8] Benison, G., Karplus, P. A., and Barbar, E. (2007) Structure and dynamics of LC8 complexes with KXTQT-motif peptides: swallow and dynein intermediate chain compete for a common site, *Journal of molecular biology 371*, 457-468.

[9] Liang, J., Jaffrey, S. R., Guo, W., Snyder, S. H., and Clardy, J. (1999) Structure of the PIN/LC8 dimer with a bound peptide, *Nature structural biology 6*, 735-740.

[10] Williams, J. C., Roulhac, P. L., Roy, A. G., Vallee, R. B., Fitzgerald, M. C., and Hendrickson, W. A. (2007) Structural and thermodynamic characterization of a cytoplasmic dynein light chain-intermediate chain complex, *Proceedings of the National Academy of Sciences of the United States of America 104*, 10028-10033.

[11] Benison, G., Nyarko, A., and Barbar, E. (2006) Heteronuclear NMR identifies a nascent helix in intrinsically disordered dynein intermediate chain: implications for folding and dimerization, *Journal of molecular biology 362*, 1082-1093.

[12] Barbar, E. (2008) Dynein light chain LC8 is a dimerization hub essential in diverse protein networks, *Biochemistry 47*, 503-508.

[13] Fu, H., Subramanian, R. R., and Masters, S. C. (2000) 14-3-3 proteins: structure, function, and regulation, *Annual review of pharmacology and toxicology 40*, 617-647.

[14] Taipale, M., Jarosz, D. F., and Lindquist, S. (2010) HSP90 at the hub of protein homeostasis: emerging mechanistic insights, *Nature reviews. Molecular cell biology 11*, 515-528.

[15] Means, A. R., and Dedman, J. R. (1980) Calmodulin--an intracellular calcium receptor, *Nature 285*, 73-77.

[16] Barbar, E., Kleinman, B., Imhoff, D., Li, M., Hays, T. S., and Hare, M. (2001) Dimerization and folding of LC8, a highly conserved light chain of cytoplasmic dynein, *Biochemistry 40*, 1596-1605.

[17] Benison, G., and Barbar, E. (2009) NMR analysis of dynein light chain dimerization and interactions with diverse ligands, *Methods in enzymology 455*, 237-258.

[18] Song, Y., Benison, G., Nyarko, A., Hays, T. S., and Barbar, E. (2007) Potential role for phosphorylation in differential regulation of the assembly of dynein light chains, *The Journal of biological chemistry 282*, 17272-17279.

[19] Wilson, M. J., Salata, M. W., Susalka, S. J., and Pfister, K. K. (2001) Light chains of mammalian cytoplasmic dynein: identification and characterization of a family of LC8 light chains, *Cell motility and the cytoskeleton 49*, 229-240.

[20] Naisbitt, S., Valtschanoff, J., Allison, D. W., Sala, C., Kim, E., Craig, A. M., Weinberg, R. J., and Sheng, M. (2000) Interaction of the postsynaptic density-95/guanylate kinase domain-associated protein complex with a light chain of myosin-V and dynein, *The Journal of neuroscience : the official journal of the Society for Neuroscience 20*, 4524-4534.

[21] Brandenburg, J., Schimanski, B., Nogoceke, E., Nguyen, T. N., Padovan, J. C., Chait, B. T., Cross, G. A., and Gunzl, A. (2007) Multifunctional class I transcription in Trypanosoma brucei depends on a novel protein complex, *The EMBO journal 26*, 4856-4866.

[22] Kirkham, J. K., Park, S. H., Nguyen, T. N., Lee, J. H., and Gunzl, A. (2016) Dynein Light Chain LC8 Is Required for RNA Polymerase I-Mediated Transcription in Trypanosoma brucei, Facilitating Assembly and Promoter Binding of Class I Transcription Factor A, *Molecular and cellular biology 36*, 95-107.

[23] Wickstead, B., and Gull, K. (2007) Dyneins across eukaryotes: a comparative genomic analysis, *Traffic 8*, 1708-1721.

[24] Emi, T., Kinoshita, T., Sakamoto, K., Mineyuki, Y., and Shimazaki, K. (2005) Isolation of a protein interacting with Vfphot1a in guard cells of Vicia faba, *Plant physiology 138*, 1615-1626.

[25] The UniProt, C. (2017) UniProt: the universal protein knowledgebase, *Nucleic acids research 45*, D158-D169.

[26] Rapali, P., Radnai, L., Suveges, D., Harmat, V., Tolgyesi, F., Wahlgren, W. Y., Katona, G., Nyitray, L., and Pal, G. (2011) Directed evolution reveals the binding motif preference of the LC8/DYNLL hub protein and predicts large numbers of novel binders in the human proteome, *PloS one 6*, e18818.

[27] Navarro-Lerida, I., Martinez Moreno, M., Roncal, F., Gavilanes, F., Albar, J. P., and Rodriguez-Crespo, I. (2004) Proteomic identification of brain proteins that interact with dynein light chain LC8, *Proteomics 4*, 339-346.

[28] Rodriguez-Crespo, I., Yelamos, B., Roncal, F., Albar, J. P., Ortiz de Montellano, P. R., and Gavilanes, F. (2001) Identification of novel cellular proteins that bind to the LC8 dynein light chain using a pepscan technique, *FEBS letters 503*, 135-141.

[29] Lo, K. W., Naisbitt, S., Fan, J. S., Sheng, M., and Zhang, M. (2001) The 8-kDa dynein light chain binds to its targets via a conserved (K/R)XTQT motif, *The Journal of biological chemistry 276*, 14059-14066.

[30] Clark, S., Nyarko, A., Lohr, F., Karplus, P. A., and Barbar, E. (2016) The Anchored Flexibility Model in LC8 Motif Recognition: Insights from the Chica Complex, *Biochemistry 55*, 199-209.

[31] Gallego, P., Velazquez-Campoy, A., Regue, L., Roig, J., and Reverter, D. (2013) Structural analysis of the regulation of the DYNLL/LC8 binding to Nek9 by phosphorylation, *The Journal of biological chemistry 288*, 12283-12294.

[32] Lightcap, C. M., Sun, S., Lear, J. D., Rodeck, U., Polenova, T., and Williams, J. C. (2008) Biochemical and structural characterization of the Pak1-LC8 interaction, *The Journal of biological chemistry 283*, 27314-27324.

[33] Bodor, A., Radnai, L., Hetenyi, C., Rapali, P., Lang, A., Kover, K. E., Perczel, A., Wahlgren, W. Y., Katona, G., and Nyitray, L. (2014) DYNLL2 dynein light chain binds to an extended linear motif of myosin 5a tail that has structural plasticity, *Biochemistry 53*, 7107-7122.

[34] Slevin, L. K., Romes, E. M., Dandulakis, M. G., and Slep, K. C. (2014) The mechanism of dynein light chain LC8-mediated oligomerization of the Ana2 centriole duplication factor, *The Journal of biological chemistry 289*, 20727-20739.

[35] Fan, J. S., Zhang, Q., Li, M., Tochio, H., Yamazaki, T., Shimizu, M., and Zhang, M. (1998) Protein inhibitor of neuronal nitric-oxide synthase, PIN, binds to a 17-amino acid residue fragment of the enzyme, *The Journal of biological chemistry 273*, 33472-33481.

[36] Nyarko, A., Song, Y., Novacek, J., Zidek, L., and Barbar, E. (2013) Multiple recognition motifs in nucleoporin Nup159 provide a stable and rigid Nup159-Dyn2 assembly, *The Journal of biological chemistry 288*, 2614-2622.

[37] Benison, G., Karplus, P. A., and Barbar, E. (2008) The interplay of ligand binding and quaternary structure in the diverse interactions of dynein light chain LC8, *Journal of molecular biology 384*, 954-966.

[38] Stelter, P., Kunze, R., Flemming, D., Hopfner, D., Diepholz, M., Philippsen, P., Bottcher, B., and Hurt, E. (2007) Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex, *Nature cell biology 9*, 788-796.

[39] Rao, L., Romes, E. M., Nicholas, M. P., Brenner, S., Tripathy, A., Gennerich, A., and Slep, K. C. (2013) The yeast dynein Dyn2-Pac11 complex is a dynein dimerization/processivity factor: structural and single-molecule characterization, *Molecular biology of the cell 24*, 2362-2377.

[40] Romes, E. M., Tripathy, A., and Slep, K. C. (2012) Structure of a yeast Dyn2-Nup159 complex and molecular basis for dynein light chain-nuclear pore interaction, *The Journal of biological chemistry 287*, 15862-15873.

[41] Morgan, J. L., Jensen, M. R., Ozenne, V., Blackledge, M., and Barbar, E. (2017) The LC8 Recognition Motif Preferentially Samples Polyproline II Structure in Its Free State, *Biochemistry 56*, 4656-4666.

[42] Brereton, A. E., and Karplus, P. A. (2018) Ensemblator v3: Robust atom-level comparative analyses and classification of protein structure ensembles, *Protein science : a publication of the Protein Society 27*, 41-50.

[43] Hall, J., Karplus, P. A., and Barbar, E. (2009) Multivalency in the assembly of intrinsically disordered Dynein intermediate chain, *The Journal of biological chemistry 284*, 33115-33121.

[44] Kidane, A. I., Song, Y., Nyarko, A., Hall, J., Hare, M., Lohr, F., and Barbar, E. (2013) Structural features of LC8-induced self-association of swallow, *Biochemistry 52*, 6011-6020.

[45] Schnorrer, F., Bohmann, K., and Nusslein-Volhard, C. (2000) The molecular motor dynein is involved in targeting swallow and bicoid RNA to the anterior pole of Drosophila oocytes, *Nature cell biology 2*, 185-190.

[46] Wang, C., Li, S., Januschke, J., Rossi, F., Izumi, Y., Garcia-Alvarez, G., Gwee, S. S., Soon, S. B., Sidhu, H. K., Yu, F., Matsuzaki, F., Gonzalez, C., and Wang, H. (2011) An ana2/ctp/mud complex regulates spindle orientation in Drosophila neuroblasts, *Developmental cell 21*, 520-533.

[47] Nyarko, A., and Barbar, E. (2011) Light chain-dependent self-association of dynein intermediate chain, *The Journal of biological chemistry 286*, 1556-1566.

[48] Hall, J., Song, Y., Karplus, P. A., and Barbar, E. (2010) The crystal structure of dynein intermediate chain-light chain roadblock complex gives new insights into dynein assembly, *The Journal of biological chemistry 285*, 22566-22575.

[49] Nyarko, A., Song, Y., and Barbar, E. (2012) Intrinsic disorder in dynein intermediate chain modulates its interactions with NudE and dynactin, *The Journal of biological chemistry 287*, 24884-24893.

[50] Jie, J., Lohr, F., and Barbar, E. (2015) Interactions of Yeast Dynein with Dynein Light Chain and Dynactin: GENERAL IMPLICATIONS FOR INTRINSICALLY DISORDERED DUPLEX SCAFFOLDS IN MULTIPROTEIN ASSEMBLIES, *The Journal of biological chemistry 290*, 23863-23874.

[51] Luo, S., Garcia-Arencibia, M., Zhao, R., Puri, C., Toh, P. P., Sadiq, O., and Rubinsztein, D. C. (2012) Bim inhibits autophagy by recruiting Beclin 1 to microtubules, *Molecular cell 47*, 359-370.

[52] Puthalakath, H., Huang, D. C., O'Reilly, L. A., King, S. M., and Strasser, A. (1999) The proapoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex, *Molecular cell 3*, 287-296.

[53] Vadlamudi, R. K., Bagheri-Yarmand, R., Yang, Z., Balasenthil, S., Nguyen, D., Sahin, A. A., den Hollander, P., and Kumar, R. (2004) Dynein light chain 1, a p21-activated kinase 1-interacting substrate, promotes cancerous phenotypes, *Cancer cell 5*, 575-585.

[54] Jung, Y., Kim, H., Min, S. H., Rhee, S. G., and Jeong, W. (2008) Dynein light chain LC8 negatively regulates NF-kappaB through the redox-dependent interaction with IkappaBalpha, *The Journal of biological chemistry 283*, 23863-23871.

[55] Lei, K., and Davis, R. J. (2003) JNK phosphorylation of Bim-related members of the Bcl2 family induces Bax-dependent apoptosis, *Proceedings of the National Academy of Sciences of the United States of America 100*, 2432-2437.

[56] Jurado, S., Conlan, L. A., Baker, E. K., Ng, J. L., Tenis, N., Hoch, N. C., Gleeson, K., Smeets, M., Izon, D., and Heierhorst, J. (2012) ATM substrate Chk2-interacting Zn2+ finger (ASCIZ) Is a bi-functional transcriptional activator and feedback sensor in the regulation of dynein light chain (DYNLL1) expression, *The Journal of biological chemistry 287*, 3156-3164.

[57] Goggolidou, P., Hadjirin, N. F., Bak, A., Papakrivopoulou, E., Hilton, H., Norris, D. P., and Dean, C. H. (2014) Atmin mediates kidney morphogenesis by modulating Wnt signaling, *Human molecular genetics 23*, 5303-5316.

[58] Goggolidou, P., Stevens, J. L., Agueci, F., Keynton, J., Wheway, G., Grimes, D. T., Patel, S. H., Hilton, H., Morthorst, S. K., DiPaolo, A., Williams, D. J., Sanderson, J., Khoronenkova, S. V., Powles-Glover, N., Ermakov, A., Esapa, C. T., Romero, R., Dianov, G. L., Briscoe, J., Johnson, C. A., Pedersen, L. B., and Norris, D. P. (2014) ATMIN is a transcriptional regulator of both lung morphogenesis and ciliogenesis, *Development 141*, 3966-3977.

[59] Jurado, S., Smyth, I., van Denderen, B., Tenis, N., Hammet, A., Hewitt, K., Ng, J. L., McNees, C. J., Kozlov, S. V., Oka, H., Kobayashi, M., Conlan, L. A., Cole, T. J., Yamamoto, K., Taniguchi, Y., Takeda, S., Lavin, M. F., and Heierhorst, J. (2010) Dual functions of ASCIZ in the DNA base damage response and pulmonary organogenesis, *PLoS genetics 6*, e1001170.

[60] Zaytseva, O., Tenis, N., Mitchell, N., Kanno, S., Yasui, A., Heierhorst, J., and Quinn, L. M. (2014) The novel zinc finger protein dASCIZ regulates mitosis in Drosophila via an essential role in dynein light-chain expression, *Genetics 196*, 443-453.

[61] Lightcap, C. M., Kari, G., Arias-Romero, L. E., Chernoff, J., Rodeck, U., and Williams, J. C. (2009) Interaction with LC8 is required for Pak1 nuclear import and is indispensable for zebrafish development, *PloS one 4*, e6025.

[62] Dunsch, A. K., Hammond, D., Lloyd, J., Schermelleh, L., Gruneberg, U., and Barr, F. A. (2012) Dynein light chain 1 and a spindle-associated adaptor promote dynein asymmetry and spindle orientation, *The Journal of cell biology 198*, 1039-1054.

[63] Tan, G. S., Preuss, M. A., Williams, J. C., and Schnell, M. J. (2007) The dynein light chain 8 binding motif of rabies virus phosphoprotein promotes efficient viral transcription, *Proceedings of the National Academy of Sciences of the United States of America 104*, 7229-7234.

[64] Fejtova, A., Davydova, D., Bischof, F., Lazarevic, V., Altrock, W. D., Romorini, S., Schone, C., Zuschratter, W., Kreutz, M. R., Garner, C. C., Ziv, N. E., and Gundelfinger, E. D. (2009) Dynein light chain regulates axonal trafficking and synaptic levels of Bassoon, *The Journal of cell biology 185*, 341-355.

[65] Okamura, S. M., Oki-Idouchi, C. E., and Lorenzo, P. S. (2006) The exchange factor and diacylglycerol receptor RasGRP3 interacts with dynein light chain 1 through its C-terminal domain, *The Journal of biological chemistry 281*, 36132-36139.

[66] Rayala, S. K., den Hollander, P., Manavathi, B., Talukder, A. H., Song, C., Peng, S., Barnekow, A., Kremerskothen, J., and Kumar, R. (2006) Essential role of KIBRA in co-activator function of dynein light chain 1 in mammalian cells, *The Journal of biological chemistry 281*, 19092-19099.

[67] Gupta, A., Diener, D. R., Sivadas, P., Rosenbaum, J. L., and Yang, P. (2012) The versatile molecular complex component LC8 promotes several distinct steps of flagellar assembly, *The Journal of cell biology 198*, 115-126.

[68] Gaik, M., Flemming, D., von Appen, A., Kastritis, P., Mucke, N., Fischer, J., Stelter, P., Ori, A., Bui, K. H., Bassler, J., Barbar, E., Beck, M., and Hurt, E. (2015) Structural basis for assembly and function of the Nup82 complex in the nuclear pore scaffold, *The Journal of cell biology 208*, 283-297.

[69] Rapali, P., Garcia-Mayoral, M. F., Martinez-Moreno, M., Tarnok, K., Schlett, K., Albar, J. P., Bruix, M., Nyitray, L., and Rodriguez-Crespo, I. (2011) LC8 dynein light chain (DYNLL1) binds to the C-terminal domain of ATM-interacting protein (ATMIN/ASCIZ) and regulates its subcellular localization, *Biochemical and biophysical research communications 414*, 493-498.

[70] McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) The PSIPRED protein structure prediction server, *Bioinformatics 16*, 404-405.

[71] McDonnell, A. V., Jiang, T., Keating, A. E., and Berger, B. (2006) Paircoil2: improved prediction of coiled coils from sequence, *Bioinformatics 22*, 356-358.

[72] Bassler, J., Paternoga, H., Holdermann, I., Thoms, M., Granneman, S., Barrio-Garcia, C., Nyarko, A., Lee, W., Stier, G., Clark, S. A., Schraivogel, D., Kallas, M., Beckmann, R., Tollervey, D., Barbar, E., Sinning, I., and Hurt, E. (2014) A network of assembly factors is involved in remodeling rRNA elements during preribosome maturation, *The Journal of cell biology 207*, 481-498.

[73] Clark, S. A., Tronrud, D. E., and Karplus, P. A. (2015) Residue-level global and local ensemble-ensemble comparisons of protein domains, *Protein science : a publication of the Protein Society 24*, 1528-1542.

[74] Clark, S. A., Jespersen, N., Woodward, C., and Barbar, E. (2015) Multivalent IDP assemblies: Unique properties of LC8-associated, IDP duplex scaffolds, *FEBS letters 589*, 2543-2551.

[75] Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins 42*, 38-48.

[76] Namba, K. (2001) Roles of partly unfolded conformations in macromolecular self-assembly, *Genes to cells : devoted to molecular & cellular mechanisms 6*, 1-12.

[77] Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M., and Obradovic, Z. (2002) Intrinsic disorder and protein function, *Biochemistry 41*, 6573-6582.

[78] Varadi, M., Kosol, S., Lebrun, P., Valentini, E., Blackledge, M., Dunker, A. K., Felli, I. C., Forman-Kay, J. D., Kriwacki, R. W., Pierattelli, R., Sussman, J., Svergun, D. I., Uversky, V. N., Vendruscolo, M., Wishart, D., Wright, P. E., and Tompa, P. (2014) pE-DB: a database of structural ensembles of intrinsically disordered and of unfolded proteins, *Nucleic acids research 42*, D326-335.

[79] Wells, M., Tidow, H., Rutherford, T. J., Markwick, P., Jensen, M. R., Mylonas, E., Svergun, D. I., Blackledge, M., and Fersht, A. R. (2008) Structure of tumor suppressor p53 and its intrinsically disordered N-terminal transactivation domain, *Proceedings of the National Academy of Sciences of the United States of America 105*, 5762-5767.

[80] Qin, B. Y., Liu, C., Srinath, H., Lam, S. S., Correia, J. J., Derynck, R., and Lin, K. (2005) Crystal structure of IRF-3 in complex with CBP, *Structure 13*, 1269-1277.

[81] Van Roey, K., Dinkel, H., Weatheritt, R. J., Gibson, T. J., and Davey, N. E. (2013) The switches.ELM resource: a compendium of conditional regulatory interaction interfaces, *Science signaling 6*, rs7.

[82] Weatheritt, R. J., Gibson, T. J., and Babu, M. M. (2014) Asymmetric mRNA localization contributes to fidelity and sensitivity of spatially localized systems, *Nature structural & molecular biology 21*, 833-839.

[83] Morgan, J. L., Song, Y., and Barbar, E. (2011) Structural dynamics and multiregion interactions in Dynein-dynactin recognition, *The Journal of biological chemistry 286*, 39349-39359.

[84] Marsh, J. A., Dancheck, B., Ragusa, M. J., Allaire, M., Forman-Kay, J. D., and Peti, W. (2010) Structural diversity in free and bound states of intrinsically disordered protein phosphatase 1 regulators, *Structure 18*, 1094-1103.

[85] Oldfield, C. J., Meng, J., Yang, J. Y., Yang, M. Q., Uversky, V. N., and Dunker, A. K. (2008) Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners, *BMC genomics 9 Suppl 1*, S1.

[86] Dunker, A. K., Silman, I., Uversky, V. N., and Sussman, J. L. (2008) Function and structure of inherently disordered proteins, *Current opinion in structural biology 18*, 756-764.

[87] Liu, J., Perumal, N. B., Oldfield, C. J., Su, E. W., Uversky, V. N., and Dunker, A. K. (2006) Intrinsic disorder in transcription factors, *Biochemistry 45*, 6873-6888.

[88] Wright, P. E., and Dyson, H. J. (2014) Intrinsically disordered proteins in cellular signalling and regulation, *Nature reviews. Molecular cell biology 16*, 18-29.

[89] Cortese, M. S., Uversky, V. N., and Dunker, A. K. (2008) Intrinsic disorder in scaffold proteins: getting more from less, *Prog Biophys Mol Biol 98*, 85-106.

[90] Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nature reviews. Molecular cell biology 6*, 197-208.

[91] Babu, M. M., van der Lee, R., de Groot, N. S., and Gsponer, J. (2011) Intrinsically disordered proteins: regulation and disease, *Current opinion in structural biology 21*, 432-440.

[92] Haynes, C., Oldfield, C. J., Ji, F., Klitgord, N., Cusick, M. E., Radivojac, P., Uversky, V. N., Vidal, M., and Iakoucheva, L. M. (2006) Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes, *PLoS computational biology 2*, e100.

[93] Xue, B., Dunker, A. K., and Uversky, V. N. (2012) The roles of intrinsic disorder in orchestrating the Wnt-pathway, *Journal of biomolecular structure & dynamics 29*, 843-861.

[94] Hinds, M. G., Smits, C., Fredericks-Short, R., Risk, J. M., Bailey, M., Huang, D. C., and Day, C. L. (2007) Bim, Bad and Bmf: intrinsically unstructured BH3-only proteins that undergo a localized conformational change upon binding to prosurvival Bcl-2 targets, *Cell death and differentiation 14*, 128-136.

[95] Wang, Y., Fisher, J. C., Mathew, R., Ou, L., Otieno, S., Sublet, J., Xiao, L., Chen, J., Roussel, M. F., and Kriwacki, R. W. (2011) Intrinsic disorder mediates the diverse regulatory functions of the Cdk inhibitor p21, *Nature chemical biology 7*, 214-221.

[96] Stein, A., Pache, R. A., Bernado, P., Pons, M., and Aloy, P. (2009) Dynamic interactions of proteins in complex networks: a more structured view, *The FEBS journal 276*, 5390-5405.

[97] Mittag, T., Kay, L. E., and Forman-Kay, J. D. (2010) Protein dynamics and conformational disorder in molecular recognition, *J Mol Recognit 23*, 105-116.

[98] Smock, R. G., and Gierasch, L. M. (2009) Sending signals dynamically, *Science 324*, 198-203.

[99] Malinovska, L., Kroschwald, S., and Alberti, S. (2013) Protein disorder, prion propensities, and self-organizing macromolecular collectives, *Biochimica et biophysica acta 1834*, 918-931.

[100] Nyarko, A., Hall, J., Hall, A., Hare, M., Kremerskothen, J., and Barbar, E. (2011) Conformational dynamics promote binding diversity of dynein light chain LC8, *Biophysical chemistry 159*, 41-47.

[101] Poisson, N., Real, E., Gaudin, Y., Vaney, M. C., King, S., Jacob, Y., Tordo, N., and Blondel, D. (2001) Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: dissociation of dynein-binding properties and transcriptional functionality of P, *The Journal of general virology 82*, 2691-2696.

[102] Dyson, H. J., and Wright, P. E. (2002) Coupling of folding and binding for unstructured proteins, *Current opinion in structural biology 12*, 54-60.

[103] Rautureau, G. J., Day, C. L., and Hinds, M. G. (2010) Intrinsically disordered proteins in bcl-2 regulated apoptosis, *International journal of molecular sciences 11*, 1808-1824.

[104] Shammas, S. L., Travis, A. J., and Clarke, J. (2013) Remarkably fast coupled folding and binding of the intrinsically disordered transactivation domain of cMyb to CBP KIX, *The journal of physical chemistry. B 117*, 13346-13356.

[105] Sugase, K., Dyson, H. J., and Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature 447*, 1021-1025.

[106] Oldfield, C. J., Cheng, Y., Cortese, M. S., Romero, P., Uversky, V. N., and Dunker, A. K. (2005) Coupled folding and binding with alpha-helix-forming molecular recognition elements, *Biochemistry 44*, 12454-12470.

[107] Cannon, J. F. (2013) How phosphorylation activates the protein phosphatase-1 * inhibitor-2 complex, *Biochimica et biophysica acta 1834*, 71-86.

[108] Peti, W., Nairn, A. C., and Page, R. (2012) Folding of Intrinsically Disordered Protein Phosphatase 1 Regulatory Proteins, *Current physical chemistry 2*, 107-114.

[109] Tong, K. I., Kobayashi, A., Katsuoka, F., and Yamamoto, M. (2006) Two-site substrate recognition model for the Keap1-Nrf2 system: a hinge and latch mechanism, *Biological chemistry 387*, 1311-1320.

[110] Tong, K. I., Katoh, Y., Kusunoki, H., Itoh, K., Tanaka, T., and Yamamoto, M. (2006) Keap1 recruits Neh2 through binding to ETGE and DLG motifs: characterization of the two-site molecular recognition model, *Molecular and cellular biology 26*, 2887-2900.

[111] Taguchi, K., Motohashi, H., and Yamamoto, M. (2011) Molecular mechanisms of the Keap1-Nrf2 pathway in stress response and cancer evolution, *Genes to cells : devoted to molecular & cellular mechanisms 16*, 123-140.

[112] Balazs, A., Csizmok, V., Buday, L., Rakacs, M., Kiss, R., Bokor, M., Udupa, R., Tompa, K., and Tompa, P. (2009) High levels of structural disorder in scaffold proteins as exemplified by a novel neuronal protein, CASK-interactive protein1, *The FEBS journal 276*, 3744-3756.

[113] Gunasekaran, K., Tsai, C. J., Kumar, S., Zanuy, D., and Nussinov, R. (2003) Extended disordered proteins: targeting function with less scaffold, *Trends in biochemical sciences 28*, 81-85.

[114] Kaberdin, V. R., Singh, D., and Lin-Chao, S. (2011) Composition and conservation of the mRNA-degrading machinery in bacteria, *Journal of biomedical science 18*, 23.

[115] Callaghan, A. J., Marcaida, M. J., Stead, J. A., McDowall, K. J., Scott, W. G., and Luisi, B. F. (2005) Structure of Escherichia coli RNase E catalytic domain and implications for RNA turnover, *Nature 437*, 1187-1191.

[116] Vanzo, N. F., Li, Y. S., Py, B., Blum, E., Higgins, C. F., Raynal, L. C., Krisch, H. M., and Carpousis, A. J. (1998) Ribonuclease E organizes the protein interactions in the Escherichia coli RNA degradosome, *Genes & development 12*, 2770-2781.

[117] Tsai, Y. C., Du, D., Dominguez-Malfavon, L., Dimastrogiovanni, D., Cross, J., Callaghan, A. J., Garcia-Mena, J., and Luisi, B. F. (2012) Recognition of the 70S ribosome and polysome by the RNA degradosome in Escherichia coli, *Nucleic acids research 40*, 10417-10431.

[118] Nurmohamed, S., Vaidialingam, B., Callaghan, A. J., and Luisi, B. F. (2009) Crystal structure of Escherichia coli polynucleotide phosphorylase core bound to RNase E, RNA and manganese: implications for catalytic mechanism and RNA degradosome assembly, *Journal of molecular biology 389*, 17-33.

[119] Luthra, P., Jordan, D. S., Leung, D. W., Amarasinghe, G. K., and Basler, C. F. (2015) Ebola virus VP35 interaction with dynein LC8 regulates viral RNA synthesis, *Journal of virology 89*, 5148-5153.

[120] Wagner, W., Fodor, E., Ginsburg, A., and Hammer, J. A., 3rd. (2006) The binding of DYNLL2 to myosin Va requires alternatively spliced exon B and stabilizes a portion of the myosin's coiled-coil domain, *Biochemistry 45*, 11564-11577.

[121] Ivanov, I., Crepin, T., Jamin, M., and Ruigrok, R. W. (2010) Structure of the dimerization domain of the rabies virus phosphoprotein, *Journal of virology 84*, 3707-3710.

[122] Barbar, E., and Nyarko, A. (2014) Polybivalency and disordered proteins in ordering macromolecular assemblies, *Semin Cell Dev Biol*.

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

[124] Weber, S. C., and Brangwynne, C. P. (2012) Getting RNA and protein in phase, *Cell 149*, 1188-1191.

[125] Allen, N. P., Patel, S. S., Huang, L., Chalkley, R. J., Burlingame, A., Lutzmann, M., Hurt, E. C., and Rexach, M. (2002) Deciphering networks of protein interactions at the nuclear pore complex, *Molecular & cellular proteomics : MCP 1*, 930-946.

[126] Brangwynne, C. P., Eckmann, C. R., Courson, D. S., Rybarska, A., Hoege, C., Gharakhani, J., Julicher, F., and Hyman, A. A. (2009) Germline P granules are liquid droplets that localize by controlled dissolution/condensation, *Science 324*, 1729-1732.

[127] Burke, M. G., Woscholski, R., and Yaliraki, S. N. (2003) Differential hydrophobicity drives self-assembly in Huntington's disease, *Proceedings of the National Academy of Sciences of the United States of America 100*, 13928-13933.

[128] Han, T. W., Kato, M., Xie, S., Wu, L. C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., and McKnight, S. L. (2012) Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies, *Cell 149*, 768-779.

[129] Radu, A., Moore, M. S., and Blobel, G. (1995) The peptide repeat domain of nucleoporin Nup98 functions as a docking site in transport across the nuclear pore complex, *Cell 81*, 215-222.

[130] Wetzel, R. (2012) Physical chemistry of polyglutamine: intriguing tales of a monotonous sequence, *Journal of molecular biology 421*, 466-490.

[131] Gsponer, J., and Babu, M. M. (2012) Cellular strategies for regulating functional and nonfunctional protein aggregation, *Cell reports 2*, 1425-1437.

[132] Mittag, T., Orlicky, S., Choy, W. Y., Tang, X., Lin, H., Sicheri, F., Kay, L. E., Tyers, M., and Forman-Kay, J. D. (2008) Dynamic equilibrium engagement of a polyvalent ligand with a single-site receptor, *Proceedings of the National Academy of Sciences of the United States of America 105*, 17772-17777.

[133] Hall, J., Hall, A., Pursifull, N., and Barbar, E. (2008) Differences in dynamic structure of LC8 monomer, dimer, and dimer-peptide complexes, *Biochemistry 47*, 11940-11952.

[134] Fan, J. S., Zhang, Q., Tochio, H., and Zhang, M. (2002) Backbone dynamics of the 8 kDa dynein light chain dimer reveals molecular basis of the protein's functional diversity, *Journal of biomolecular NMR 23*, 103-114.

[135] Stuchell-Brereton, M. D., Siglin, A., Li, J., Moore, J. K., Ahmed, S., Williams, J. C., and Cooper, J. A. (2011) Functional interaction between dynein light chain and intermediate chain is required for mitotic spindle positioning, *Molecular biology of the cell 22*, 2690-2701.

[136] Flemming, D., Thierbach, K., Stelter, P., Bottcher, B., and Hurt, E. (2010) Precise mapping of subunits in multiprotein complexes by a versatile electron microscopy label, *Nature structural & molecular biology 17*, 775-778.

[137] Kim, H., Hyeon, S., Kim, H., Yang, Y., Huh, J. Y., Park, D. R., Lee, H., Seo, D. H., Kim, H. S., Lee, S. Y., and Jeong, W. (2013) Dynein light chain LC8 inhibits osteoclast differentiation and prevents bone loss in mice, *J Immunol 190*, 1312-1318.

[138] Asthana, J., Kuchibhatla, A., Jana, S. C., Ray, K., and Panda, D. (2012) Dynein light chain 1 (LC8) association enhances microtubule stability and promotes microtubule bundling, *The Journal of biological chemistry 287*, 40793-40805.

[139] Makokha, M., Hare, M., Li, M., Hays, T., and Barbar, E. (2002) Interactions of cytoplasmic dynein light chains Tctex-1 and LC8 with the intermediate chain IC74, *Biochemistry 41*, 4302-4311.

[140] Kardon, J. R., and Vale, R. D. (2009) Regulators of the cytoplasmic dynein motor, *Nature reviews. Molecular cell biology 10*, 854-865.

[141] Caviston, J. P., Ross, J. L., Antony, S. M., Tokito, M., and Holzbaur, E. L. (2007) Huntingtin facilitates dynein/dynactin-mediated vesicle transport, *Proceedings of the National Academy of Sciences of the United States of America 104*, 10045-10050.

[142] Gill, S. R., Schroer, T. A., Szilak, I., Steuer, E. R., Sheetz, M. P., and Cleveland, D. W. (1991) Dynactin, a conserved, ubiquitously expressed component of an activator of vesicle motility mediated by cytoplasmic dynein, *The Journal of cell biology 115*, 1639-1650.

[143] Yan, X., Li, F., Liang, Y., Shen, Y., Zhao, X., Huang, Q., and Zhu, X. (2003) Human Nudel and NudE as regulators of cytoplasmic dynein in poleward protein transport along the mitotic spindle, *Molecular and cellular biology 23*, 1239-1250.

[144] Stehman, S. A., Chen, Y., McKenney, R. J., and Vallee, R. B. (2007) NudE and NudEL are required for mitotic progression and are involved in dynein recruitment to kinetochores, *The Journal of cell biology 178*, 583-594.

[145] Barbar, E. (2012) Native disorder mediates binding of dynein to NudE and dynactin, *Biochemical Society transactions 40*, 1009-1013.

[146] Pfister, K. K., Salata, M. W., Dillman, J. F., 3rd, Vaughan, K. T., Vallee, R. B., Torre, E., and Lye, R. J. (1996) Differential expression and phosphorylation of the 74-kDa intermediate chains of cytoplasmic dynein in cultured neurons and glia, *The Journal of biological chemistry 271*, 1687-1694.

[147] Pfister, K. K., Shah, P. R., Hummerich, H., Russ, A., Cotton, J., Annuar, A. A., King, S. M., and Fisher, E. M. (2006) Genetic analysis of the cytoplasmic dynein subunit families, *PLoS genetics 2*, e1.

[148] Trokter, M., Mucke, N., and Surrey, T. (2012) Reconstitution of the human cytoplasmic dynein complex, *Proceedings of the National Academy of Sciences of the United States of America 109*, 20895-20900.

[149] Vaughan, P. S., Leszyk, J. D., and Vaughan, K. T. (2001) Cytoplasmic dynein intermediate chain phosphorylation regulates binding to dynactin, *The Journal of biological chemistry 276*, 26171-26179.

[150] Kim, H., Hyeon, S., Yang, Y., Huh, J. Y., Park, D. R., Lee, H., Seo, D. H., Kim, H. S., Lee, S. Y., and Jeong, W. (2013) Dynein light chain LC8 inhibits osteoclast differentiation and prevents bone loss in mice, *J Immunol 190*, 1312-1318.

[151] Barbar, E., and Nyarko, A. (2014) NMR Characterization of Self-Association Domains Promoted by Interactions with LC8 Hub Protein, *Computational and structural biotechnology journal 9*, e201402003.

[152] King, S. M., Barbarese, E., Dillman, J. F., PatelKing, R. S., Carson, J. H., and Pfister, K. K. (1996) Brain cytoplasmic and flagellar outer arm dyneins share a highly conserved M(r) 8,000 light chain, *Journal of Biological Chemistry 271*, 19358-19366.

[153] Kubota, T., Matsuoka, M., Chang, T. H., Bray, M., Jones, S., Tashiro, M., Kato, A., and Ozato, K. (2009) Ebolavirus VP35 interacts with the cytoplasmic dynein light chain 8, *Journal of virology 83*, 6952-6956.

[154] Barbar, E., Kleinman, B., Imhoff, D., Li, M. G., Hays, T. S., and Hare, M. (2001) Dimerization and folding of LC8, a highly conserved light chain of cytoplasmic dynein, *Biochemistry 40*, 1596-1605.

[155] Clark, S. A., Jespersen, N., Woodward, C., and Barbar, E. (2015) Multivalent IDP assemblies: Unique properties of LC8-associated, IDP duplex scaffolds, *FEBS Lett*.

[156] Liang, J., Jaffrey, S. R., Guo, W., Snyder, S. H., and Clardy, J. (1999) Structure of the PIN/LC8 dimer with a bound peptide, *Nat Struct Mol Biol 272*, 20929 - 20935.

[157] McGuffin, L. J., Bryson, K., and Jones, D. T. (2000) The PSIPRED protein structure prediction server., *Bioinformatics 16*, 404-405.

[158] Kjaergaard, M., Brander, S., and Poulsen, F. M. (2011) Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH, *Journal of biomolecular NMR 49*, 139-149.

[159] Lajoix, A. D., Gross, R., Aknin, C., Dietz, S., Granier, C., and Laune, D. (2004) Cellulose membrane supported peptide arrays for deciphering protein-protein interaction sites: the case of PIN, a protein with multiple natural partners, *Molecular diversity 8*, 281-290.

[160] Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments, *Acta crystallographica. Section D, Biological crystallography 67*, 235-242.

[161] Hall, J., Karplus, P. A., and Barbar, E. (2009) Multivalency in the assembly of intrinsically disordered dynein intermediate chain, *J Biol Chem 284*, 33115-33121.

[162] Stelter, P., Kunze, R., Radwan, M., Thomson, E., Thierbach, K., Thoms, M., and Hurt, E. (2012) Monitoring spatiotemporal biogenesis of macromolecular assemblies by pulse-chase epitope labeling, *Molecular cell 47*, 788-796.

[163] Jie, J., Lohr, F., and Barbar, E. (2015) Interactions of yeast dynein with dynein light chain and dynactin: General implications for intrinsically disordered duplex scaffolds in multi-protein assemblies, *J Biol Chem*.

[164] Fraczkiewicz, R. a. B., W. (1998) Exact and Efficient Analytical Calculation of the Accessible Surface Areas and Their Gradients for Macromolecules, *J. Comp. Chem. 19*, 319-333.

[165] Favier, A., and Brutscher, B. (2011) Recovering lost magnetization: polarization enhancement in biomolecular NMR, *Journal of biomolecular NMR 49*, 9-15.

[166] Lescop, E., Kern, T., and Brutscher, B. (2010) Guidelines for the use of band-selective radiofrequency pulses in hetero-nuclear NMR: example of longitudinal-relaxation-enhanced BEST-type 1H-15N correlation experiments, *J Magn Reson 203*, 190-198.

[167] Johnson, B. A., and Blevins, R. A. (1994) NMRView: A computer program for visualization and analysis of NMR data, *Journal of biomolecular NMR 4*, 603-614.

[168] Ferrage, F., Piserchio, A., Cowburn, D., and Ghose, R. (2008) On the measurement of 15N-{1H} nuclear Overhauser effects, *J Magn Reson 192*, 302-313.

[169] Lakomek, N. A., Ying, J., and Bax, A. (2012) Measurement of (1)(5)N relaxation rates in perdeuterated proteins by TROSY-based methods, *Journal of biomolecular NMR 53*, 209-221.

[170] Goddard, T. D., and Kneller, D. G. SPARKY 3, San Francisco, California.

[171] Leslie, A. G. W. a. P., H.R. (2007) Processing Diffraction Data with Mosflm, In *Evolving Methods for Macromolecular Crystallography*, pp 41-51.

[172] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta crystallographica. Section D, Biological crystallography 66*, 213-221.

[173] Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta crystallographica. Section D, Biological crystallography 66*, 486-501.

[174] Winn, M. D., Isupov, M. N., and Murshudov, G. N. (2001) Use of TLS parameters to model anisotropic displacements in macromolecular refinement, *Acta crystallographica. Section D, Biological crystallography 57*, 122-133.

[175] Banerjee, N. a. Z. M. (2003) Identifying cooperativity among transcription factors controlling the cell cycle in yeast, *Nucleic Acids Res. 31*, 7024-70311.

[176] Meek, D. W., and Anderson, C. W. (2009) Posttranslational modification of p53: cooperative integrators of function, *Cold Spring Harbor perspectives in biology 1*, a000950.

[177] Cantor, A. B., and Orkin, S. H. (2002) Transcriptional regulation of erythropoiesis: an affair involving multiple partners, *Oncogene 21*, 3368-3376.

[178] Minezaki, Y., Homma, K., Kinjo, A. R., and Nishikawa, K. (2006) Human transcription factors contain a high fraction of intrinsically disordered regions essential for transcriptional regulation, *Journal of molecular biology 359*, 1137-1149.

[179] Currie, S. L., Doane, J. J., Evans, K. S., Bhachech, N., Madison, B. J., Lau, D. K. W., McIntosh, L. P., Skalicky, J. J., Clark, K. A., and Graves, B. J. (2017) ETV4 and AP1 Transcription Factors Form Multivalent Interactions with three Sites on the MED25 Activator-Interacting Domain, *Journal of molecular biology 429*, 2975-2995.

[180] Shammas, S. L. (2017) Mechanistic roles of protein disorder within transcription, *Current opinion in structural biology 42*, 155-161.

[181] Dyson, H. J., and Wright, P. E. (2016) Role of Intrinsic Protein Disorder in the Function and Interactions of the Transcriptional Coactivators CREB-binding Protein (CBP) and p300, *The Journal of biological chemistry 291*, 6714-6722.

[182] Uversky, V. N., Oldfield, C. J., Midic, U., Xie, H., Xue, B., Vucetic, S., Iakoucheva, L. M., Obradovic, Z., and Dunker, A. K. (2009) Unfoldomics of human diseases: linking protein intrinsic disorder with diseases, *BMC genomics 10 Suppl 1*, S7.

[183] Jurado, S., Gleeson, K., O'Donnell, K., Izon, D. J., Walkley, C. R., Strasser, A., Tarlinton, D. M., and Heierhorst, J. (2012) The Zinc-finger protein ASCIZ regulates B cell development via DYNLL1 and Bim, *The Journal of experimental medicine 209*, 1629-1639.

[184] King, A., Li, L., Wong, D. M., Liu, R., Bamford, R., Strasser, A., Tarlinton, D. M., and Heierhorst, J. (2017) Dynein light chain regulates adaptive and innate B cell development by distinctive genetic mechanisms, *PLoS genetics 13*, e1007010.

[185] Wong, D. M., Li, L., Jurado, S., King, A., Bamford, R., Wall, M., Walia, M. K., Kelly, G. L., Walkley, C. R., Tarlinton, D. M., Strasser, A., and Heierhorst, J. (2016) The Transcription Factor ASCIZ and Its Target DYNLL1 Are Essential for the Development and Expansion of MYC-Driven B Cell Lymphoma, *Cell reports 14*, 1488-1499.

[186] Chintapalli, V. R., Wang, J., and Dow, J. A. (2007) Using FlyAtlas to identify better Drosophila melanogaster models of human disease, *Nature genetics 39*, 715-720.

[187] Boratyn, G. M., Camacho, C., Cooper, P. S., Coulouris, G., Fong, A., Ma, N., Madden, T. L., Matten, W. T., McGinnis, S. D., Merezhuk, Y., Raytselis, Y., Sayers, E. W., Tao, T., Ye, J., and Zaretskaya, I. (2013) BLAST: a more efficient report with usability improvements, *Nucleic acids research 41*, W29-33.

[188] Ezomo, O. F., Kazuya Takahashi, Yuki Horie, Mohammed S. Mustak and Shunsuke Meshitsuka. (2010) Circular Dichroism Studies on C-terminal Zinc Finger Domain of Transcription Factor GATA-2, *Yonago Acta medica 53*, 25-28.

[189] Tamiola, K., Acar, B., and Mulder, F. A. (2010) Sequence-specific random coil chemical shifts of intrinsically disordered proteins, *Journal of the American Chemical Society 132*, 18000-18003.

[190] Pufall, M. A., Lee, G. M., Nelson, M. L., Kang, H. S., Velyvis, A., Kay, L. E., McIntosh, L. P., and Graves, B. J. (2005) Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region, *Science 309*, 142-145.

[191] Desjardins, G., Meeker, C. A., Bhachech, N., Currie, S. L., Okon, M., Graves, B. J., and McIntosh, L. P. (2014) Synergy of aromatic residues and phosphoserines within the intrinsically disordered DNA-binding inhibitory elements of the Ets-1 transcription factor, *Proceedings of the National Academy of Sciences of the United States of America 111*, 11019-11024.

[192] Lee, G. M., Pufall, M. A., Meeker, C. A., Kang, H. S., Graves, B. J., and McIntosh, L. P. (2008) The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region, *Journal of molecular biology 382*, 1014-1030.

[193] Ferreon, J. C., Lee, C. W., Arai, M., Martinez-Yamout, M. A., Dyson, H. J., and Wright, P. E. (2009) Cooperative regulation of p53 by modulation of ternary complex formation with CBP/p300 and HDM2, *Proceedings of the National Academy of Sciences of the United States of America 106*, 6591-6596.

[194] Lee, C. W., Ferreon, J. C., Ferreon, A. C., Arai, M., and Wright, P. E. (2010) Graded enhancement of p53 binding to CREB-binding protein (CBP) by multisite phosphorylation, *Proceedings of the National Academy of Sciences of the United States of America 107*, 19290-19295.

[195] Teufel, D. P., Bycroft, M., and Fersht, A. R. (2009) Regulation by phosphorylation of the relative affinities of the N-terminal transactivation domains of p53 for p300 domains and Mdm2, *Oncogene 28*, 2112-2118.

[196] Holmberg, C. I., Tran, S. E., Eriksson, J. E., and Sistonen, L. (2002) Multisite phosphorylation provides sophisticated regulation of transcription factors, *Trends in biochemical sciences 27*, 619-627.

[197] Remenyi, A., Scholer, H. R., and Wilmanns, M. (2004) Combinatorial control of gene expression, *Nature structural & molecular biology 11*, 812-815.

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

[199] Hayes, D. L., T.; Philo, J. (1995) Program Sednterp: Sedimentation Interpretation Program, *Alliance Protein Laboratories*.

[200] Vistica, J., Dam, J., Balbo, A., Yikilmaz, E., Mariuzza, R. A., Rouault, T. A., and Schuck, P. (2004) Sedimentation equilibrium analysis of protein interactions with global implicit mass conservation constraints and systematic noise decomposition, *Analytical biochemistry 326*, 234-256.

[201] Motackova, V., Novacek, J., Zawadzka-Kazimierczuk, A., Kazimierczuk, K., Zidek, L., Sanderova, H., Krasny, L., Kozminski, W., and Sklenar, V. (2010) Strategy for complete NMR assignment of disordered proteins with highly repetitive sequences based on resolution-enhanced 5D experiments, *Journal of biomolecular NMR 48*, 169-177.

[202] Novacek, J., Zawadzka-Kazimierczuk, A., Papouskova, V., Zidek, L., Sanderova, H., Krasny, L., Kozminski, W., and Sklenar, V. (2011) 5D 13C-detected experiments for backbone assignment of unstructured proteins with a very low signal dispersion, *Journal of biomolecular NMR 50*, 1-11.

[203] Pernot, P., Round, A., Barrett, R., De Maria Antolinos, A., Gobbo, A., Gordon, E., Huet, J., Kieffer, J., Lentini, M., Mattenet, M., Morawe, C., Mueller-Dieckmann, C., Ohlsson, S., Schmid, W., Surr, J., Theveneau, P., Zerrad, L., and McSweeney, S. (2013) Upgraded ESRF BM29 beamline for SAXS on macromolecules in solution, *Journal of synchrotron radiation 20*, 660-664.

[204] Petoukhov, M. V. K., P. V.; Kikhney A. G. ; Svergun D. I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis, *J. Appl. Cryst. 40*, s223-s228.

[205] A., G. (1938) The diffusion of X-rays under the extremely weak angles applied to the study of fine particles and colloidal suspension, *Comptes Rendus Hebdomadaires Des Seances De L Acad Des Sci. 206*, 1374-1376.

[206] Porod, G. (1982) General Theory, In *Small angel X-ray scattering* (Glatter O, K. O., Ed.), pp 17-51, Academic Press, London.

[207] Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria, *J. Appl. Cryst. 25*, 495-503.

[208] Myers, J. B., Zaegel, V., Coultrap, S. J., Miller, A. P., Bayer, K. U., and Reichow, S. L. (2017) The CaMKII holoenzyme structure in activation-competent conformations, *Nature communications 8*, 15742.

[209] Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007) EMAN2: an extensible image processing suite for electron microscopy, *Journal of structural biology 157*, 38-46.

[210] Scheres, S. H. (2012) RELION: implementation of a Bayesian approach to cryo-EM structure determination, *Journal of structural biology 180*, 519-530.

[211] McNees, C. J., Conlan, L. A., Tenis, N., and Heierhorst, J. (2005) ASCIZ regulates lesion-specific Rad51 focus formation and apoptosis after methylating DNA damage, *The EMBO journal 24*, 2447-2457.

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

[213] Bondarenko, V., Mowrey, D., Liu, L. T., Xu, Y., and Tang, P. (2013) NMR resolved multiple anesthetic binding sites in the TM domains of the alpha4beta2 nAChR, *Biochimica et biophysica acta 1828*, 398-404.

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

[215] Konarev, P. V. V., V. V. ; Sokolova, A. V.; Koch M. H. J. ; Svergun D. I. . (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis, *J. Appl. Cryst. 36*, 1277-1282.

[216] Gill, G., Pascal, E., Tseng, Z. H., and Tjian, R. (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAFII110 component of the Drosophila TFIID complex and mediates transcriptional activation, *Proceedings of the National Academy of Sciences of the United States of America 91*, 192-196.

[217] Almlof, T., Gustafsson, J. A., and Wright, A. P. (1997) Role of hydrophobic amino acid clusters in the transactivation activity of the human glucocorticoid receptor, *Molecular and cellular biology 17*, 934-945.

[218] Cress, W. D., and Triezenberg, S. J. (1991) Critical structural elements of the VP16 transcriptional activation domain, *Science 251*, 87-90.

[219] Ma, J., and Ptashne, M. (1987) A new class of yeast transcriptional activators, *Cell 51*, 113-119.

[220] Drysdale, C. M., Duenas, E., Jackson, B. M., Reusser, U., Braus, G. H., and Hinnebusch, A. G. (1995) The transcriptional activator GCN4 contains multiple activation domains that are critically dependent on hydrophobic amino acids, *Molecular and cellular biology 15*, 1220-1233.

[221] Kim, D. I., Jensen, S. C., Noble, K. A., Kc, B., Roux, K. H., Motamedchaboki, K., and Roux, K. J. (2016) An improved smaller biotin ligase for BioID proximity labeling, *Molecular biology of the cell 27*, 1188-1196.

[222] Ward, J. J., McGuffin, L. J., Bryson, K., Buxton, B. F., and Jones, D. T. (2004) The DISOPRED server for the prediction of protein disorder, *Bioinformatics 20*, 2138-2139.

[223] Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein, *Journal of molecular biology 157*, 105-132.

[224] Patel, A., Lee, H. O., Jawerth, L., Maharana, S., Jahnel, M., Hein, M. Y., Stoynov, S., Mahamid, J., Saha, S., Franzmann, T. M., Pozniakovski, A., Poser, I., Maghelli, N., Royer, L. A., Weigert, M., Myers, E. W., Grill, S., Drechsel, D., Hyman, A. A., and Alberti, S. (2015) A Liquid-to-Solid Phase Transition of the ALS Protein FUS Accelerated by Disease Mutation, *Cell 162*, 1066-1077.

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