AN ABSTRACT OF THE DISSERTATION OF

Kayla A. Jara for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on May 31, 2022.

Title: Context Matters: Assembly and Disorder in the Regulation of Multivalent Complexes.

Abstract approved:

_____

Elisar J. Barbar

Intrinsically disordered proteins (IDPs), protein regions (IDRs), and protein complexes continue to emerge at the forefront of protein science. Proteins and protein regions lacking specific structure are found in all organisms, and often have vital roles in numerous biological processes. Breaking the well-known structure-function paradigm, the understanding of disorder-based functionality is constantly expanding. In addition to their structural plasticity and dynamic conformational flexibility, IDPs/IDRs often interact with their binding partners multivalently to serve regulatory purposes.

This thesis reports on two example systems analyzing the impact of disorder and complex assembly on regulation while also showcasing why the context in which we study these systems really matters. Four chapters of original work include one protocol book chapter and three primary research reports. The book chapter, chapter 2, serves as a guide in using NMR to probe interactions of IDPs, using dynein intermediate chain (IC) as a model example. Chapters 3 and 4 report on a domain of the transcription factor, ASCIZ, and interactions with its own product, dynein light chain (LC8). Chapter 3 classifies the domain of study as a predominantly in-register binder of LC8 while chapter 4 further explores how binding motif specificity and the linker length between binding sites impact complex assembly. This work serves as a foundation and model system for more intricate cases, such as the focus of chapter 5,

dynein IC subcomplexes. In chapter 5, a regulatory mechanism involving autoinhibition via long range intramolecular interactions that is relieved by multivalent complex assembly is presented. IC is studied and biophysically characterized, for the first time, in context of its entire N-terminal domain, in the reconstitution of subcomplexes, and in the full-length protein. Finally, chapter 6 summarizes the impact and highlights of the reported work and presents suggestions for work moving forward. These studies provide detailed description of two various multivalent and disordered protein assemblies that together serve as a blueprint for studying these systems in context for both the IDP and dynein fields.

Context Matters: Assembly and Disorder in the Regulation of Multivalent Complexes

by
Kayla A. Jara

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented May 31, 2022
Commencement June 2022

Doctor of Philosophy dissertation of Kayla A. Jara presented on May 31, 2022.

APPROVED:

Major Professor, representing Biochemistry and Biophysics

Head of the Department of Biochemistry and Biophysics

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.

Kayla A. Jara, Author

ACKNOWLEDGEMENTS

I would like to start by thanking my mentor, Dr. Elisar Barbar. The lab motto has always been *Insist, Persist, Take Risks* and I can't think of anyone that is a better embodiment of those qualities. Elisar, thank you for seeing the best in me and always believing that I was capable of whatever I was trying to accomplish. Your unwavering confidence in me has pushed me to grow tremendously both as a person and as a scientist. Thank you for modeling strength and compassion in your leadership roles and for being such a fierce advocate for women and for science.

To the Barbar Besties – I'm so thankful to have crossed paths with you all, of which there are *many*. Aidan, joining the lab together has always given us an advantage. I've enjoyed our countless conversations and your countless interesting takes on anything debatable. Thank you for being the lab's resident 'computer guy' and thank you for teaching me to be steadfast in what I believe in. Jesse, you came into the lab as a bit of an outsider, thank you for being so persistent in winning us over with your upbeat attitude and willingness to help-out wherever you can. Hannah, watching you mature from undergrad student to lab assistant to grad student has been amazing. Your infectious laughter is exactly what the lab needed as we tried to emerge from the pandemic. Douglas, thank goodness you began your Postdoc in time for us to overlap. Where were you when nobody wanted to take my side in office arguments?? Getting to know you and Erika has been a blessing, thank you for your friendship, comradery, and predisposition to avoid conflict. Zhen, we are beyond lucky to have you as our lab manager. The day you joined the lab was the day everything became better for all of us. Thank you for all the work you do to keep the lab and our projects running. To the numerous undergrads that have been a part of the besties during my time in the lab, thank you for your curiosity, determination, and eagerness to learn. In particular, thank you Sanjana, Seth, Cat, Coban, and Gretchen for helping me become a better teacher and mentor. It takes something special to be a Barbar Bestie and I'm so lucky to have shared the lab with all of you.

To the Biochemistry and Biophysics department, thank you for feeling like family. To my graduate committee members, Dr. Afua Nyarko, Dr. Colin Johnson,

and Dr. Maria Franco thank you for your guidance and endless support. I've enjoyed sitting in on faculty meetings in my capacity as GSA President where I always felt welcome and heard. Whether through short interactions in the halls or as classroom instructors, it's clear that BB faculty care immensely about our department's students and cultivating an environment to foster student success. I want to especially thank Kari Van Zee for all she does for the department on behalf of students and for always being willing to lend an ear when needed. To my cohort – Aidan, Amber, Phil, and Ruben – it was a pleasure to act as our 'group mom'. There's no better group to commiserate with and you all have pushed me to be a better researcher. Ally, Shauna, and Elise, thank you for being so willing to share your knowledge with me and for your care for the cohorts below you. Thank you, my fellow BB grads, for entrusting me with roles such as GSA president and BPS Student Chapter president, I hope I've done you all proud.

Finally, I would like to thank my family and close friends for their incredible encouragement and support. Mom and Dad, thank you for believing in me so fiercely and helping me through my tough transition into grad school. You two will always be my biggest role models and I'm so lucky to have you as my #1 fans. Matt and Lisa, my Super Sibs, thank you for always wanting to protect me and for never letting me forget that even though I may be considered the smartest, I'll also always be the youngest. Amber and Laken, thank you for being Winston's favorite aunties and for pushing Kayla and I to be more adventurous. There are no better people to serve as our chosen family and our time in Oregon wouldn't be the same without you two by our sides. Winston, thanks for adding so much joy and love to our home. You're currently lying next to me as I type and although I know you'd rather me not be, you do a great job keeping me company while I work. Last but certainly not least, Kayla, my better half with the same name, there are no words to sufficiently express my gratitude for your support, patience, and love. Thank you for taking the leap with me and always reminding me that things will be ok. Thank you for everything you do for me and Winston, for being 'the fun one', and for always keeping me grounded.

CONTRIBUTION OF AUTHORS

Patrick Reardon contributed to experiments and text in chapters 3 and 4 and assisted with AUC and NMR experiments in chapter 5. Amber Rolland contributed to MS experiments and analysis in chapters 3 and 4. Andrew Swansiger contributed to the MS experiments in chapter 4. James Prell contributed to the MS analysis in chapters 3 and 4. Nikolaus Loening contributed to experiments, text, and analysis in chapter 5. Zhen Yu and Prajna Woonnimani contributed to protein production via baculovirus methods in chapter 5. Coban Brooks and Cat Vesely contributed experiments in chapter 5. Elisar Barbar was involved in the design, writing, and analysis of all experiments and chapters.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

LIST OF APPENDIX FIGURES

Figure Page

Context Matters: Assembly and Disorder in the Regulation of Multivalent Complexes

Chapter 1

Introduction

Intrinsically Disordered Proteins

Overview

Proteins lacking any specific fold and structure have been termed intrinsically disordered proteins (IDPs). The abundance and functional importance of these IDPs and proteins with disordered regions (IDRs), is revolutionizing protein science fields[1]. IDPs/IDRs are found in all proteomes, across organisms and kingdoms[2–5]. In eukaryotes, it's estimated that 30% of all proteins contain IDRs >30 amino acids in length[6]. Prevalent even in enzymes, the protein class at the height of structure-function relationships, IDRs are truly everywhere[7]. Recognition of the prevalence of IDPs/IDRs and their roles as biologically active proteins rapidly grew, to the point now where they are no longer considered exceptions but rather a fact of existence[1].

IDPs and IDRs are characterized by low sequence diversity, few hydrophobic residues, an abundance of charged residues, and areas of sequence repeats. These features culminate in a plethora of benefits over their structured protein counterparts. For example, a lack of folding and structure make IDRs highly assessable and thus ideal locations for post-translational modifications and numerous binding interactions[6]. Due in part to their high charge state, as well as their abundance of short linear binding motifs, disordered regions are promiscuous in their binding interactions, and have been found to facilitate the formation of many large, complex protein assemblies[8]. While ordered proteins are constrained by the requirement to conserve their 3-D structure, disordered proteins are under no such evolutionary pressures and therefore display much higher mutation rates, providing efficient means for adaptability[9].

In addition to their many benefits, IDPs/IDRs are also extremely functionally diverse. IDPs/IDRs are most commonly found in signaling pathways[10] and regulatory processes[6] partially due to their high prevalence in kinases that conduct protein phosphorylation and play important roles in the regulation of cell proliferation, growth, and apoptosis[11]. IDPs/IDRs are also commonly found in scaffolding assemblies that bring together a specific set of proteins to participate in molecular

pathways[12]. Due to their diverse set of vital cellular functions, IDP misregulation and aggregation result in a number of diseases such as Alzheimer's disease[13] and breast cancer[14]. Surprisingly, despite their prevalence and importance to biology, IDPs remain mechanistically poorly understood due to inherent difficulties with their structural characterization.

While some IDPs adopt structure upon binding to a partner protein, others retain their disorder adding flexibly and "fuzziness" to protein complexes[15]. Such characteristics often interfere with traditional structural methods such as x-ray crystallography and cryo-electron microscopy (cryoEM). Luckily, other biophysical methods such as circular dichroism (CD), analytical ultracentrifugation (AUC), isothermal titration calorimetry (ITC), nuclear magnetic resonance (NMR) spectroscopy, and native mass spectrometry (MS) can be used in combination to effectively study IDPs. All such methods were employed in the work presented in this thesis and are covered more thoroughly later and within individual chapters.

Multivalency

In addition to their structural plasticity and dynamic confirmational flexibility, IDPs/IDRs often interact with their binding partners multivalently, a property crucial to the work presented in this thesis. Compared to monovalent interactions, in which ligands bind a single location of a partner, multivalent interactions involve linked associations of ligands binding to multiple sites and often confer increased binding affinity[16–18].

Multivalent IDP assemblies can be categorized as binary complexes, IDP single chain scaffolds, IDP duplex scaffolds, higher order IDP associations, and IDP multi-site collective binding ligands[17]. Binary complexes contain an IDP chain with multiple recognition motifs that specifically interact with different sites on one folded partner and upon binding, adopt 3D structure[17,19–23]. IDP scaffolds have a series of recognition motifs along their sequence for multiple partners and when bound, are incorporated into the secondary structure fold of the partner while linker regions between sites remain disordered. Single chain scaffolds are incredibly abundant and

well described in the literature[12,24–26]. Duplex scaffolds are composed of two IDP chains connected by one or more dimeric partners with two symmetrical binding sites and/or by self-association interactions within the chain[16,17]. Higher order IDP associations are composed of self-associated complexes that can reversibly incorporate other functional proteins[27–29]. Finally, IDP multi-site collective binding ligands occur when multiple sites of a polyvalent ligand engage a single site on a partner[30].

Duplex scaffolds represent a particularly unique category of IDP assemblies as the biological functions they serve are vast and diverse. Cases in which many of the same or multiple dimeric ligands bind multiple sites across disordered chains is common for partners of the hub protein, LC8, with notable examples: 1) ASCIZ, the transcription factor for LC8 that binds up to 11 LC8 dimers in its human form and 2) dynein intermediate chain, that binds LC8 and many additional dimeric proteins. Further discussion of these assemblies is presented below.

Dynein Light Chain 1

Overview

Dynein light chain 1 (LC8) was first described as a component of axonemal dynein[31], and was later identified as a subunit of the cytoplasmic motor proteins, dynein 1 and 2[32,33]. LC8 forms a stable 21 kDa dimer under normal cellular conditions with an association constant of ~80 nM[34]. LC8's dimeric structure contains two, five-stranded anti-parallel β-sheets and opposite of the dimer inface, a pair of α-helices[35–43] (Figure 1.1). Upon phosphorylation at reside Ser88[44] or when exposed to a pH below 4.8[34], LC8 is known to exist in its monomeric form, however because the binding groove is localized to the dimer interface, monomeric versions are inactive[44]. Within the LC8 homodimer, two symmetrical binding grooves are housed that allow for LC8's primary function as a dimerization engine of its intrinsically disordered (or regions thereof) binding partners[17,18,45] (Figure 1.1). LC8 is a highly conserved protein across species[46] and is present in all examined eukaryotes[47], including plants

lacking dynein[47,48]. Together these properties pointed at expanded functional roles for LC8 and it is now known as a ubiquitously expressed and essential hub protein with numerous binding partners that represent a wide variety of cellular functions[49,50].

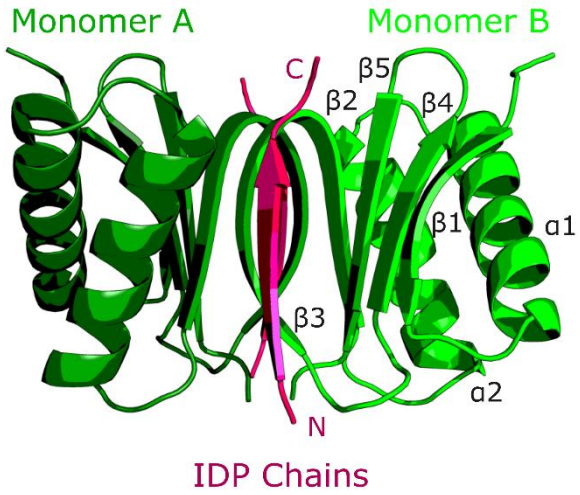

Figure 1.1 LC8 dimer structure and binding groove. A ribbon diagram of the LC8 dimer (monomer units in dark and light green) bound to two chains of an IDP peptide (pink) (PDB code 2P2T). The IDP peptide adopts β-strand structure upon binding in LC8's central binding groove. Secondary structural elements are labeled on LC8 monomer B (light green). This LC8/IDP bound structure is representative of all LC8 partner complexes.

Out of LC8's over 100 confirmed partners, IDPs preforming functions such as intracellular transport[51,52], nuclear pore formation[53], viral interactions[54–56], and transcription[57–60] are represented. Overall, LC8 binders fall into 9 generalized categories: neurotransmission, viral infection, transcription, tumor suppression, cell death, development, cell division, cell transport, and cell signaling. LC8 and its partners are broadly expressed and distributed throughout the cell[49], however LC8's precise role in many cellular functions is unknown due to the fact that LC8 knockouts are lethal in mice[61]. More studies using the newly generated limb bud mesoderm-restricted conditional knock-out of LC8, the first available animal and cell-line model of its kind[61], are needed to continue to expand the understanding of LC8 functionality.

LC8 is a non-enzymatic protein, meaning its main role in its many assemblies is structural in nature. LC8-facilitated duplex assemblies can be functionally and structurally categorized as those that include 1) self-association formation[50,52], 2) coiled-coil formation[62–64], 3) conformational ensemble restriction[55], and 4) protein regulation via multivalency[16,53,59,65,66]. Examples of all such assemblies are illustrated in Figure 1.2. Some of these categories are straight forward and have been well described, as is the interaction of LC8 with the *Drosophila* protein, Swallow. Upon binding LC8, Swallow forms a coiled-coil 17 residues away that is 7-fold stronger than in the absence of LC8[63,64]. However, other categories remain more ambiguous, namely multivalent assemblies including LC8. In some instances, one or more other bivalent partners will bind the LC8 IDP duplex[65,66], while in others the multiple binding sites are all occupied by LC8[43,53,59]. Regardless of their type, the specific downstream functions of these multivalent complexes remain variable and, in some cases, unknown.

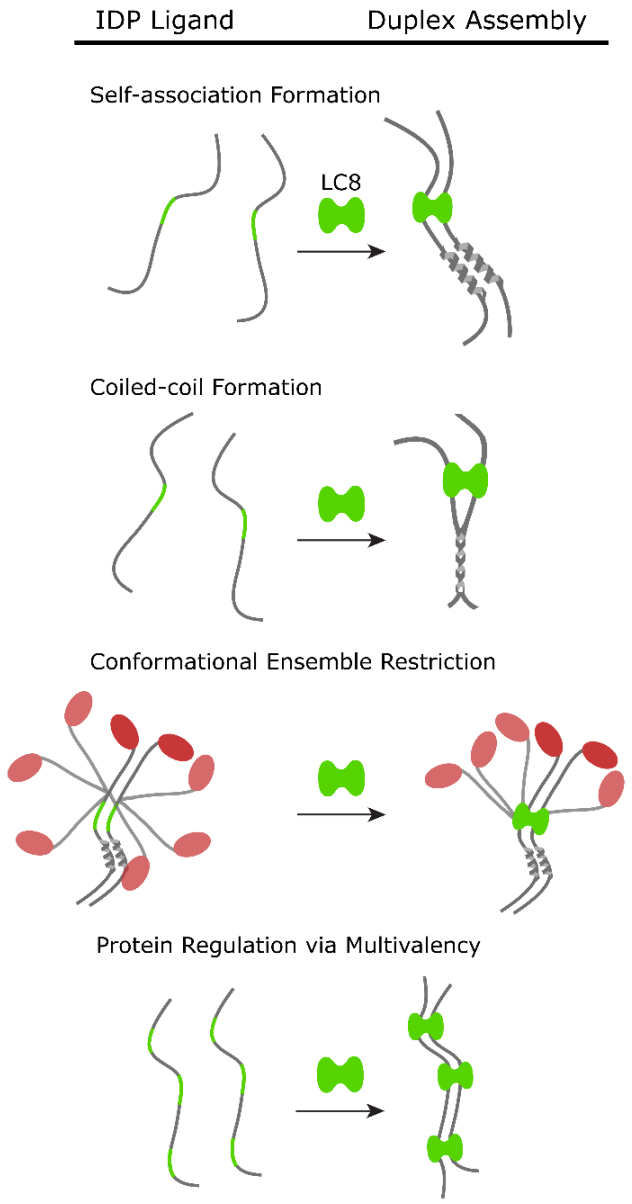

Figure 1.2 LC8-facilitated IDP duplex assemblies. Binding interactions between IDP ligands (grey) and LC8 dimers (green) lead to a variety of assemblies. Some ligands have a self-association or coiled-coil region near the LC8 motif, and upon binding form such structure (top and top middle). As seen with RavP[55], LC8 binding restricts the conformational ensemble (faded) of disordered linkers and ordered domains (red) instead (bottom middle). Finally, multivalent ligands with more than one LC8 motif bind to multiple LC8 dimers by means of regulation (bottom). Across these categories IDP/LC8 duplex result in stabilization of a complex, often facilitating downstream interactions with additional protein partners (not shown).

Binding Recognition Motif

LC8 partner proteins share a short (8 amino acid) linear recognition motif that mediate their binding to LC8[55] (Figure 1.3). The motif lies within intrinsically disordered regions of proteins, however upon binding to LC8 form a single β-strand structure that is integrated into an LC8 anti-parallel β-sheet[35]. The binding motif contains some level of variation, however is typically anchored by a threonine-glutamine-threonine (TQT) sequence[43]. This TQT anchor is highly enriched upon known LC8 binders and as such, these residues are known within the larger motif as positions -1, 0, and +1, respectively[43,62] (Figure 1.3). The LC8 binding interface is dynamic[62,67,68], thus allowing for sequence variation within binders, although at the same time steric and enthalpic restrictions do apply. One such restriction includes the fact that the side chains of residues in the -1 and +1 positions are entirely buried and thus propose a strong preference to branched side chains that are either hydrophobic or participate in hydrogen bonding. Although past motif prediction analyses provided initial information of the LC8 binding motif[38,69],via the recent creation of an LC8 binders database and subsequent open access LC8 binding prediction algorithm, the 8 amino acid motif has been more thoroughly characterized[49].

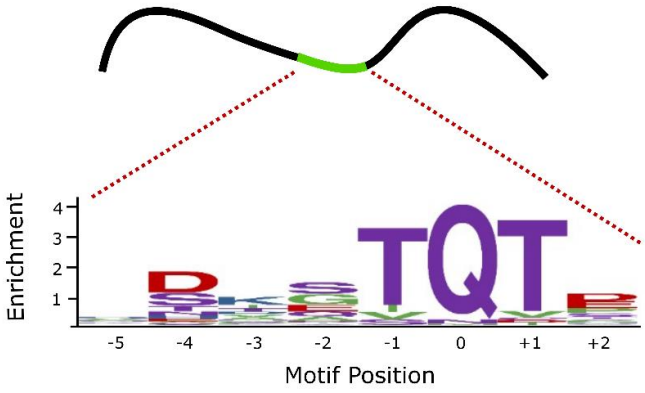

Figure 1.3. LC8 recognition motif. LC8 partner proteins (black) share an 8 amino acid linear recognition motif (green) that mediate LC8 binding. The motif is anchored by a highly enriched TQT sequence. Figure adapted from Jespersen et al. Life Sci. Alliance (2019)[49].

Distinct preferences and exclusions for positions within the LC8 binding motif outside of the anchor include -4 positions capable of hydrogen binding, -3 positions with larger positive side chains, and strand breaking +2 positions. Also of important consideration, pre-anchor prolines, a high concentration of charges, or bulky hydrophobic groups at the -2 position all negatively impact the likelihood that a sequence will bind to LC8[49]. In line with the dynamic nature of LC8's binding grooves, thermodynamic analyses of tight binders indicate that a wide range of entropy and enthalpy compensations are sampled among binding sequences and that entropic factors modulate LC8 binding to accommodate the variation seen in partner sequences[49,68].

LC8's Transcription Factor: ASCIZ

One extremely unique LC8 partner is the transcription factor, ASCIZ (ATMIN-Substrate Chk-Interacting Zn^{2+} finger)[38,58], that not only has an 11 LC8 recognition motifs in its human version, but also interacts with LC8 via a dynamic ensemble of low occupancy LC8 bound complexes[59]. Human ASCIZ is 88 kDa and contains an N-terminal zinc finger domain followed by a 600 amino acid disordered tail (Figure 1.3). *In vivo* knockouts of ASCIZ impose consequences similar to LC8 knockouts, although with slight attenuation, indication that a vital ASCIZ function is to regulate the cellular concentrations of its own gene product, LC8[57,61,70]. *In vivo* experiments also demonstrated that overexpression of LC8 causes a decrease in LC8 transcription via an ASCIZ-mediated pathway[58]. A current model suggests ASCIZ acts as a sensor of LC8 concentrations and is able to fine-tune LC8 transcription accordingly[59].

Biophysical characterization of *Drosophila* (7 LC8 sites) and human ASCIZ, show that ASCIZ/LC8 interactions display both positive and negative cooperativity. Also, unlike Nup159 from yeast with 5 LC8 binding sites, ASCIZ does not form rigid stacked complexes but instead more heterogenous complexes by negative-stain EM analysis[59] (Figure 1.3). This may be due to the disordered linkers between LC8 binding sites in ASCIZ that vary considerably in length (3 to 30 residues). Unique to

Drosophila ASCIZ, one of the binding motifs contains a TMT rather than the canonical TQT anchor. Cooperativity, linker length, and motif specificity are all important ASCIZ/LC8 regulation properties further explored in chapters 3 and 4.

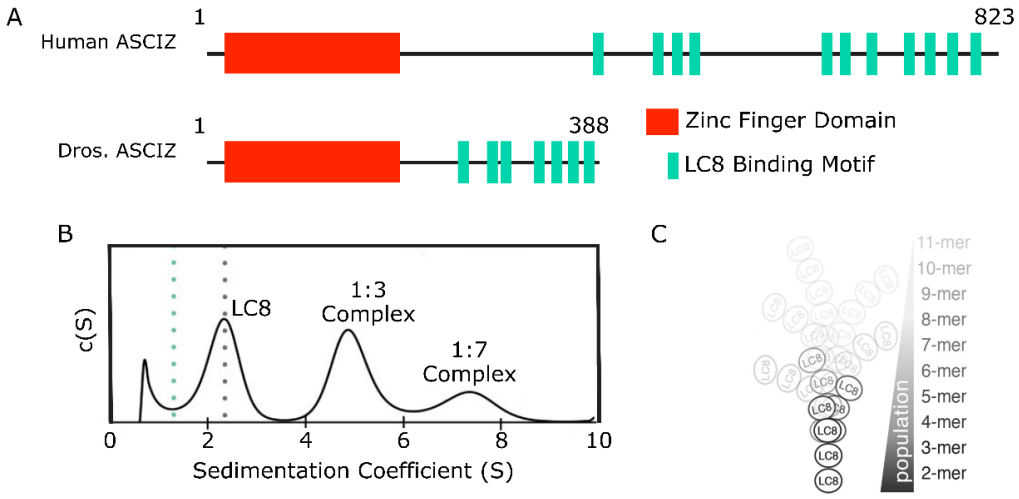

Figure 1.4. ASCIZ forms heterogenous complexes with LC8. A) Domain architecture diagrams of human and *Drosophila* ASCIZ. ASCIZ contains an N-terminal zinc finger domain (orange box), and a disordered C-terminal domain housing multiple LC8 binding motifs (teal boxes). B) An SV-AUC experiment of the LC8 binding domain of Dros. ASCIZ with LC8 (1:10 molar ratio ASCIZ:LC8). Even at this oversaturating amount of LC8, the fully bound 1:7 complex is not favored, but rather a low occupancy filled 1:3 complex. C) Illustration representing the complexes formed for human ASCIZ and LC8. Only low occupied complexes were seen by negative stain EM. This figure is adapted from Clark et al. eLife (2018) [59].

Dynein Intermediate Chain

Cytoplasmic Dynein Subunits

Cytoplasmic dynein 1 is a 1.4-MDa microtubule-based motor protein complex responsible for all minus-end-directed transport of cargos such as membranes, RNAs, proteins, and viruses[71]. Dynein has been proven as an essential gene, as mutations in the dynein transport machinery have been linked to neurological and neurodevelopmental diseases such as Perry syndrome, spinal muscular atrophy-lower

extremity predominant (SMA-LED), Charcot-Marie-Tooth disease, and lissencephaly[72]. Furthermore, a single dynein performs functions that it takes ~40 kinesins (plus-end directed motors) to do[71], underscoring dynein's significance and also suggesting a high level of intricate regulation processes and interactions within and between dynein subunits.

Dynein is composed of heavy, intermediate, light intermediate, and light chain subunits (Figure 1.5). The heavy chain (HC) of dynein contains a C-terminal motor domain and an N-terminal tail domain. The motor domain houses ATP binding sites and connect to microtubules via a microtubule binding domain (MTBD)[73–76]. The intermediate chain (IC) of dynein can also be thought of as having a distinct N- and C-terminus with the C-terminal WD40 domain serving as the attachment point to HC and the extended N-terminal disordered domain housing binding sites for the three dynein light chains: Tctex, LC8, and LC7. Two copies of IC bind HC, and the light chains exist as homodimers, binding across both copies of IC[52,77–81]. The light intermediate chain (LIC) is connected to HC via its RAS-like domain as well as N- and C-terminal helices. Again, two copies of LIC bind to HC and the LIC also contains an extended disordered chain, however C-terminally[73,82,83]. Work presented in this thesis will focus on the interactions of IC and the light chains, as well as how the IC/light chain subcomplex interacts with other non-dynein partner proteins.

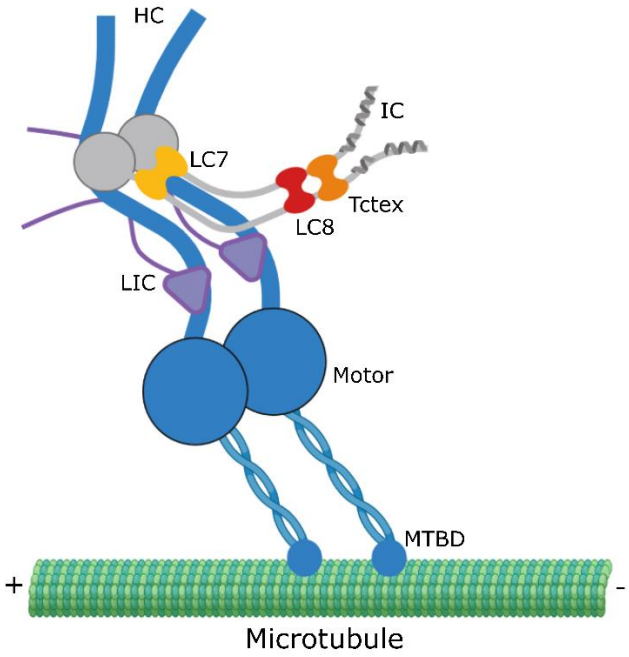

Figure 1.5. Cytoplasmic dynein subunits and configuration. Cartoon model of cytoplasmic dynein. The two dynein heavy chains (HC) contain a C-terminal motor domain and an N-terminal tail domain (blue). The motor domain connects to microtubules (green) via a microtubule binding domain (MTBD)[73–76]. The intermediate chain (IC) of dynein has a C-terminal WD40 domain that attaches to HC and a disordered N-terminal domain that is extended (grey). N-IC binds the three dynein light chains: Tctex (orange), LC8 (red), and LC7 (yellow). The light chains are homodimers, bringing together two copies of IC[52,77–81]. The light intermediate chain (LIC) is connected to HC via its RAS-like domain and contains an extended C-terminal tail (purple)[73,82,83].

IC Subcomplex: A Polybivalent Scaffold

Much of dynein has been well characterized thanks to advancements in techniques such as cryo-electron microscopy (cryoEM), however IC contains almost 300 amino acids of disorder that remain missing from resolved dynein structures[73,82,84]. This disorder makes up almost the entire N-terminal half of IC and is the location of binding interactions with dynein light chain subunits (Figure 1.5). The subcomplex formed by IC, Tctex, LC8, and LC7 is best described as a polybivalent

scaffold, where the first binding event will pay the entropic cost for subsequent bivalent binding events[16,17,52,81,85].

Some IDPs fold upon being incorporated into complexes[86], while others retain their disorder. The IC/light chains assembly represents a new class of IDP complexes, identified in the Barbar lab as IDP duplex complexes[16], where tight 3-D packing occurs only at the protein interfaces formed by linear IC motifs that adopt structure when bound to the dimeric light chains which bring together two chains of IC to form a duplex. The linkers connecting these motifs remain completely disordered[16,17,52,81,85]. Each homodimeric light chain has a corresponding recognition motif on IC that alone has propensity to form β-strands (motifs for Tctex and LC8) or α-helices (motif for LC7), yet only forms such secondary structure when bound and incorporated into the fold of their respective ligand. The IC/light chain subcomplex is best described as a polybivalent scaffold as the linking of two IC chains results in a duplex containing multiple other bivalent sites. The first binding event pays the entropic cost thus reducing the entropic penalty of subsequent ligand binding[16,81]. This type of binding enhancement is thought to be modulated by the length of disordered linkers separating binding sites. Such a mechanism has already been observed for the IC/Tctex/LC8 ternary complex in *Drosophila*, in which a 50-fold binding enhancement was measured for the second light chain, separated by a 3-reside linker[81]. Binding enhancement is expected to be much less so with a longer linker, such as is seen in both the linker between LC8 and LC7 binding sites as well as the linker between the biding site for Tctex and more N-terminal helices involved in the binding of non-dynein partner proteins[85,87].

Interactions with p150^{Glued} and NudE

To date, the best characterized regulator of dynein is dynactin, another large, multiprotein complex. Dynactin contains 11 different subunits[88] and connects to IC via its largest subunit, p150^{Glued} [89] (Figure 1.6A). p150^{Glued} contains an N-terminal cytoskeleton-associated protein, glycine-rich (CAP-Gly) domain that is involved with the binding of microtubules and two coiled-coil domains (CC1 and CC2)[88] (Figure

1.6B). A section of the second coiled-coil region, CCB1, of p150^{Glued} binds to the single α-helix (SAH) domain of IC, a domain well conserved across various species[85,90–93]. This interaction is essential for processivity along microtubules, correct spindle formation, and proper cellular division[88,89,94–96]. Studies using proteins from *Drosophila* have identified a short, second helix in IC (following the SAH domain) to be involved in binding p150^{Glued} in addition to the SAH[90,91]. This is further supported by work done on small fragments of IC from a thermophilic fungus that show that binding of p150^{Glued} is stronger than that for mammalian counterparts, and binding is weakened when the short helix (H2) following the SAH region is absent[97].

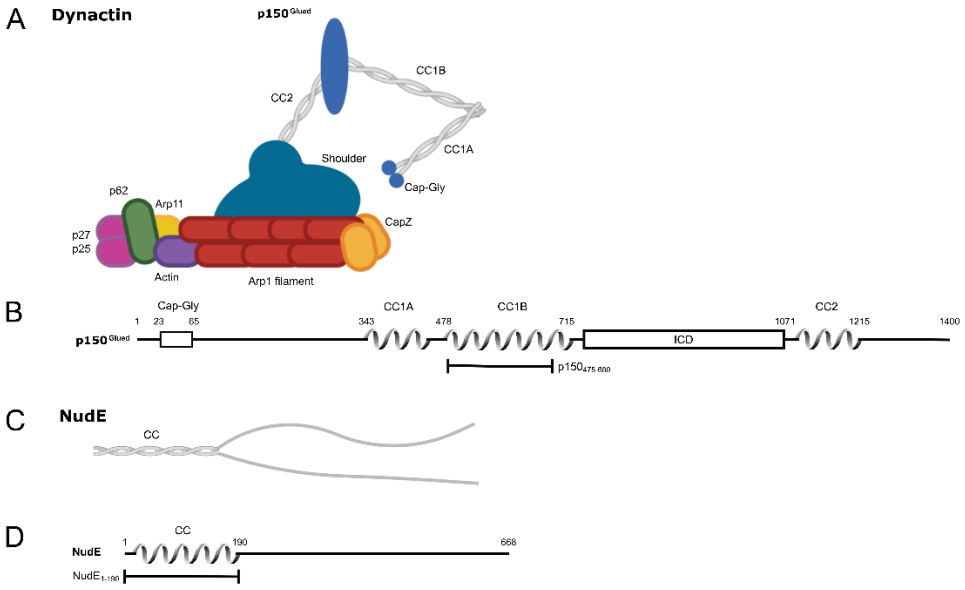

Figure 1.6. Dynactin p150^{Glued} and NudE structure and domain architecture. A) Cartoon model of dynactin. Dynactin is built around an eight-actin filament (Arp1, red) with proteins p62 (green), p27 (pink), p25 (pink), actin monomer (purple), and actin related protein (Arp11, yellow) at one end and capping proteins (CapZ, orange) at the other. Extending from the shoulder domain (teal) is the p150^{Glued} subunit, with two coiled-coil regions (grey), an intercoil domain (blue), and Cap-Gly domains. B) p150^{Glued} from *Chaetomium thermophilum* is predicted to have a Cap-Gly domain near the N-terminus, and two coiled-coil domains, CC1 and CC2, that are separated by an intercoil domain (ICD). CC1 is further divided into two regions called CC1A and CC1B. p150^{Glued} binds to dynein IC via the CC1B domain. p150$_{478\text{-}680}$ (p150$_{CC1B}$)

is the construct used in this work. C) Cartoon model of full-length NudE. The N-terminus is a coiled-coil (CC) domain while the C-terminus is disordered. D) NudE from *Chaetomium thermophilum* is predicted to have an N-terminal coil-coiled (CC) region followed by a long tail of disorder. $NudE_{1-190}$ ($NudE_{CC}$) is the construct used in this work.

Also binding IC at its SAH domain is the nuclear distribution protein (NudE), a protein required for a diverse set of functions. NudE has been shown to aid in kinetochore and chromosome migration, Golgi complex organization, centrosome duplication, mitotic spindle positioning, and membrane transport[98–102]. NudE contains a long, N-terminal coiled-coil domain followed by a disordered C-terminal domain[103] (Figure 1.6C-D). The binding site of NudE has been previously mapped to the SAH domain of mammalian and *Drosophila* ICs with a binding affinity similar to that exhibited by $p150^{Glued}$ [65,92].

IC partnering with either $p150^{Glued}$ or NudE impacts IC regulation and the dynein complex as they dictate interactions with activating adaptors, impact the selection of cargo attachment, and aid in determining how the dynein complex orients for microtubule binding[71]. Despite the importance of these interactions, the details of the molecular basis for which regulator is bound at any given time is still unclear and remains a challenge in the field. Due to the disorder/flexibility of N-IC, a recurring theme in dynein studies is a lack of detail about IC structure and binding interactions. Although residue specific studies have been performed on IC, they are thus far limited to small regions of IC and to conditions far removed from native biological systems.

We recently introduced *Chaetomium thermophilum* (CT), a thermophilic filamentous fungus, as a new system for dynein studies[97]. The CT IC_{1-260} (residues 1-260) construct predominantly used in the work presented in chapter 5, is far longer than previously studied CT, *Drosophila*, and rat constructs [65,77,81,85,90,92,93,99,100]. Utilizing almost the entire N-terminus domain of IC allows, for the first time, characterization of IC interactions in the context of its assembly with the light chains, as has been shown vital for other disorder-driven systems [104].

Biophysical Techniques

To probe the interactions of the disordered, multivalent proteins and assemblies presented in this thesis, I have employed numerous biochemical and biophysical techniques. As one would expect when characterizing such complex systems, many different proteins and constructs were a part of this work. I was fortunate to inherit some of these constructs from fellow lab members and collaborators, while others I engineered myself. Most of the proteins were expressed in *Escherichia coli* cell lines, except for full-length CT IC that was instead expressed using baculovirus methods. In all cases, proteins were purified via 6xHis tag affinity purification followed by anion exchange and/or size exclusion chromatography.

As IDPs, proteins with a mix of order and disorder, and macromolecular assemblies often require, data was obtained through a combination of complementary biophysical techniques. The main techniques used in this work are described below.

Circular Dichroism (CD)

CD is a technique that relies on the principle that molecules differentially absorb right vs. left-handed circularly polarized light. This absorption manifests in either positive or negative CD signal and for proteins, the signal within the wavelength range of ~180-250 nm has implications of specific secondary structure (α-helices, β-strands, and random coils). CD is also a simple technique that requires very little sample, is fast and inexpensive, and implements no size restrictions. CD was useful when determining the helical content of different protein constructs, confirming protein disorder, and measuring protein stability across a temperature range (measured in loss or retainment of secondary structure). My first step in protein characterization was often CD.

Analytical Ultracentrifugation (AUC)

AUC is a method used to analyze hydrodynamic properties of molecules in solution via measurement of optical density/refractive index across a centrifuge cell over time.

Relying on mass and fundamental laws of gravity, quantitative analysis of biomolecules and their interactions in their native states is possible. In an AUC experiment, centrifugal force is applied at either a high (sedimentation velocity) or low (sedimentation equilibrium) speed for applications such as the determination of molecular weight, association state, and association or dissociation constants of protein samples. AUC is also a relatively inexpensive method that requires moderate amount of sample. I was able to utilize both equilibrium (SE) and velocity (SV) type experiments to confirm expected protein association states and monitor binding interactions of many different samples.

Isothermal Titration Calorimetry (ITC)

ITC is the go-to technique used to determine the thermodynamic parameters of biomolecular interactions. Although ITC requires much higher sample concentrations than CD or AUC, the information provided is tenfold. By measuring the heat change associated with reactions in solution, ITC provides binding affinity (K_a or k_d), reaction stoichiometry (N), enthalpy (ΔH), entropy ($-T\Delta S$), and Gibbs free energy (ΔG) values for a binding interaction. Proven vital for comparing the binding interactions between QT mutants and LC8, ITC was also useful in characterizing IC interactions, although it quickly became clear that they contained a level of complexity better suited to other methods.

NMR Spectroscopy

From the moment I had my photo taken with our 800 MHz NMR during my recruitment visit, I hoped that it would be an instrument I used often. Joining the Barbar lab practically ensured this, as there is no better technique for the characterization of IDPs and to specifically map binding sites on proteins. Using radio waves in the presence of a strong magnetic field to excite NMR-active nuclei, signals arise that are sensitive to chemical environment. The 2D HSQC type experiment provides a ‘protein fingerprint’ in which every peak corresponds to each protein residue. Following assignment via 3D experiments, protein structure,

dynamics, and binding interactions can be investigated. A wide variety of NMR experiments were utilized in the work presented within this thesis. Of special note, perdeuteration methods proved essential for study of QT/LC8 complexes and study of the IC subcomplexes served as an example of the power of NMR as disordered regions were still visible in complexes up to 200 kDa.

Dissertation Contents

This dissertation includes four chapters of original work, all of which are either published or represent a manuscript in preparation for submission. Chapter 2 is a book chapter accepted for publication in a Dynein specific edition of the lab protocol series, Methods in Molecular Biology, published by *Springer Nature*. The chapter serves as a guide in using NMR to probe interactions of intrinsically disordered proteins (IDPs) with IC as a model example. I present our routinely used methodology in a simple, easy-to-follow format that could be of use to those interested in implementing NMR into their own research. The chapter concludes with an overview of how to couple NMR results with various other biophysical techniques depending on the desired research outcome, thus setting up chapters 3-5 which all incorporate data from ITC, AUC, and MALS in addition to NMR.

Chapter 3 is a research article published in the *Journal of Biological Chemistry* that shows that LC8 forms a predominantly in-register complex when bound to an IDP domain of the multivalent transcription factor, ASCIZ. The domain used, $QT_{2\text{-}4}$, houses three LC8 binding sites that vary in their motif specificity as well as in the length of the linkers between sites, thus serving as a model system for studying multivalency and disorder in complex assembly.

Chapter 4 is a research article that follows up on the work presented in chapter 3 and is in preparation for submission. In this work, mutant constructs were created, abolishing either one or two LC8 binding motifs at a time to better understand the contribution each site has on complex formation. Analyzing binding by ITC and AUC shows that in every case, independent of motif specificity or linker length, two LC8 binding sites are better than one and behave cooperatively while the cooperativity of a

third site is location dependent. Furthermore, longer linker lengths between binding sites disrupts LC8/IDP duplex assembly formation.

Chapter 5 is a research article in preparation for submission that illuminates a regulatory mechanism for dynein IC. We identify autoinhibition within IC caused by strong, long range intramolecular interactions that cover IC's N-terminal single α-helix. Reconstitution of IC subcomplexes demonstrate that this autoinhibition is regulated by both multivalency and disorder and underlies selection between multiple binding partners at the same site. Utilizing almost the entire N-terminus domain of IC allowed, for the first time, characterization of IC interactions in the context of its assembly with the light chains. We describe how assembly with the light chains relieves IC autoinhibition and regulates binding of IC to $p150^{Glued}$ and NudE and show that autoinhibition, and the regulatory model we propose persist in the full-length CT IC protein.

Chapter 6 summarizes the important findings from these studies and their impact on both the dynein and multivalent/disordered protein fields. The importance of context is highlighted, and I conclude by discussing my suggestions for future directions for the work presented.

Chapter 2

NMR analysis of the interactions and conformational plasticity of the dynein intermediate chain

Kayla A. Jara and Elisar J. Barbar

Accepted for publication in *Methods in Mol. Biol*, Jan 2022

Abstract

Cytoplasmic dynein complexes play crucial roles in intracellular transport of cellular organelles. While the motor domain of dynein is well characterized by techniques such as X-ray crystallography and cryo-electron microscopy (Cryo-EM), structural representations of dynein usually include only the more packed and easily resolved regions, and omit the long flexible and poorly structured regions. One such flexible region is the N-terminal half of the intermediate chain (IC), which contains almost 300 amino acids that are predicted to be disordered. This level of disorder makes IC very challenging to study by X-ray crystallography and Cryo-EM, but amenable to study by solution nuclear magnetic resonance (NMR), a powerful technique that can elucidate residue-specific information in dynamic structure, and transient binding interactions of associated proteins. Here we describe the methods we use to understand flexible and disordered proteins including protein expression, purification, sample preparation, and NMR data acquisition.

Keywords

Dynein intermediate chain, NMR spectroscopy, intrinsically disordered proteins, protein-protein interactions, transient structure, transient interactions.

Introduction

Dynein is a large complex composed of numerous protein subunits that function in the cellular processes of motility and cargo attachment. The dynein motor domain is situated in the heavy chains (HCs), and houses sites for ATP binding and hydrolysis. Each tail of the HC binds to one copy of the intermediate chain (IC), a subunit that plays key roles in modulating dynein interactions and activity [51,71,85,90,105]. The IC has two separate domains: a mostly disordered N-terminal domain (N-IC), and a well-ordered C-terminal domain (C-IC). The degree of disorder in N-IC is conserved amongst a variety of species [97] and is consistently predicted by structure algorithms such as PSIPRED and AlphaFold [106,107], whereas C-IC folds into a β-propeller and contains the binding site for the HC [73,82] (Figure 2.1A-B).

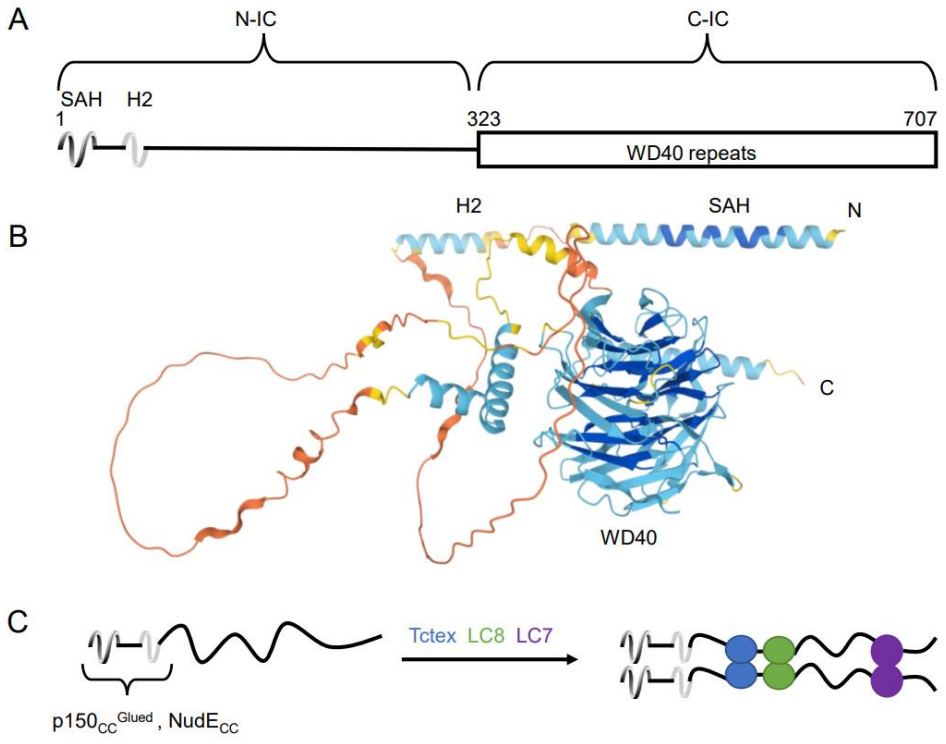

Figure 2.1. Dynein intermediate chain architecture, disorder, and assembly. A) A generalized domain architecture diagram for IC. Notable features of N-IC are a single α-helix (SAH), second helix (H2), and linker regions of disorder. C-IC contains the well folded WD40 domain. B) AlphaFold structure prediction for Cytoplasmic

dynein 1 intermediate chain 2 from Homo sapiens. Varying coloration depicts model confidence, dark blue (very high), light blue (confident), yellow (low), and orange (very low). Orange regions are indicative of unstructured regions. C) Apo, monomeric N-IC (left) with binding sites for coiled-coil regions of p150^{Glued} and NudE noted. The addition of dynein light chains, Tctex (blue), LC8 (green), and LC7 (purple) cause the formation of a dimeric IC complex with disorder in-between binding sites remaining.

A small amount of structure within N-IC is attributed to the N-terminal single alpha helix (SAH), and a short helix (H2; Figure 2.1B) [106,108,109]. Although the N-terminal residues of dynein from yeast were originally thought to form a coiled-coil, they lack the characteristic hydrophobic seam found in coiled coils. Instead, the first ~30 residues are rich in highly charged residues (~65% Arg, Lys, and Glu), suggesting helix stabilizing electrostatic interactions between residues *i* and *i* +4 and *i* and *i* + 3, which are characteristic of a single α-helix (SAH) domain [85,110]. Over the last decade we have characterized the SAH and contiguous H2 regions by NMR in yeast, Drosophila, fungal, and mammalian variations of IC [65,85,90–92,97]. Of note, H2 forms a fully ordered helix in some species but is only a nascent helix in others, and this structural variation is correlated with its importance in contributing to its binding affinity with dynactin p150^{Glued}; necessary for binding when a nascent helix, and does not contribute to binding when folded [97].

Apo-IC (unbound) from metazoa is monomeric and interacts with the three dynein light chain subunits and with many non-dynein partner proteins. N-IC binds to the coiled-coil regions of non-dynein regulatory proteins such as dynactin p150^{Glued} and nuclear distribution protein (NudE) with its SAH and H2 domains [65,85,89,90,92,97,111]. Downstream from this binding site, metazoan IC interacts with homodimeric dynein light chain subunits (LCs; Tctex, LC8, and LC7) to form a dimeric subcomplex that is best described as a polybivalent intrinsically disordered duplex scaffold (8-13). The structure of this scaffold is such that, when assembled, tight packing occurs only at the protein-protein interfaces [77,81] leaving linker regions completely disordered [52,87,91] (Figure 2.1C). These attributes of N-IC make it extremely challenging to characterize by X-ray crystallography and cryo-EM but an excellent candidate for study by NMR. Although these other techniques are extremely

useful in determining structures of the many well-ordered dynein subunits, and for providing an overall picture of the organization of the dynein complex, disordered regions and subunits often evade detection in both techniques due to their inherent flexibility.

NMR has long been used to characterize proteins and protein-protein interactions. Every protein will generate a unique, 2D ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) "fingerprint" spectrum, as the chemical shift for each HSQC peak is dependent on the structural and chemical environment that each protein residue experiences [112]. Disordered proteins exhibit a characteristic narrow chemical shift dispersion, while well-structured proteins show much more dispersion (Figure 2.2A-B). 2D NMR experiments are an improvement upon 1D ^{1}H-NMR, as they produce protein spectra that are incredibly complex and crowded (Figure 2.2C). The same drawback can be a feature of 2D spectra, especially for large proteins and proteins containing significant disorder, as residues with similar environments generate overlapping peaks. Thus, acquiring 3D spectra is often necessary to accurately assign peaks in a 2D experiment. In experiments such as ^{1}H-^{15}N-^{13}C HNCO, HN(CA)CO, HNCACB and HN(CO)CACB, each peak can be described by three frequencies and thus distinguished from the others in order to assign a protein backbone. Additional experiments such as a TOCSY-HSQC or HCCH-TOCSY are used to effectively assign side chains. Data collected from a suite of 3D experiments can be pieced together like a puzzle as the information obtained from each is complementary.

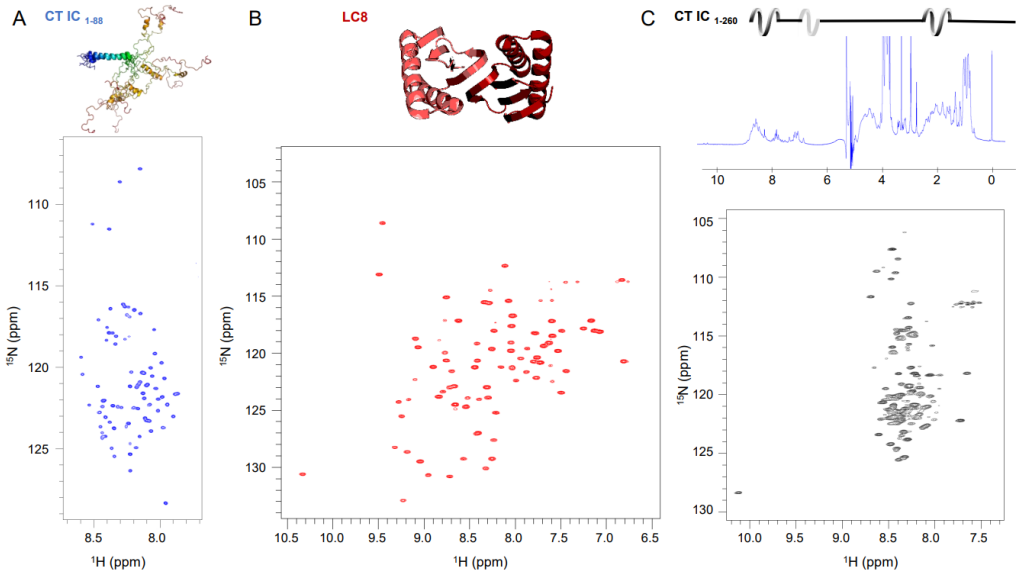

Figure 2.2. NMR spectra comparison. A) ^{15}N-HSQC of IC$_{1-88}$ (residues 1-88) from *Chaetomium thermophilum* (CT). The narrow chemical shift dispersion seen is indicative of a disordered protein, matching the structural ensemble above of 10 overlaid solved structures, aligned by the N-term (blue). B) ^{15}N-HSQC of full-length LC8 from *Drosophila*. The wide chemical shift dispersion is indicative of a well folded protein, matching the structure above. C) Domain architecture of CT IC$_{1-260}$ (residues 1-260) with corresponding 1D (^{1}H) and 2D (^{15}N-TROSY) NMR experiments below.

Perhaps the primary limitations of solution NMR are those exacerbated by increasing protein size. With the introduction of transverse relaxation-optimized spectroscopy (TROSY), studies on proteins as large as 900 kDa have been reported [113,114]. TROSY based experiments select slowly relaxing components to significantly enhance signal sensitivity and thus have narrower lines than HSQC spectra. TROSY experiments are crucial to our study of IC complexes due to both the numerous binding sites that can be captured by a single IC construct, and the dimeric nature of IC complexes with masses ranging from ~60-200 kDa [52,65,85,87,90–92,97].

Once a 2D fingerprint of a protein is assigned, residue specific information can be obtained of both unbound and bound protein samples. For example, in our studies of multiple IC constructs from multiple species, NMR has permitted us to map binding sites on IC, determine structural propensity (including transient helices),

analyze dynamics of unbound and bound IC, map a self-association domain, determine effects of mutations on protein dynamics and binding interactions, create a model representing the structural ensemble of IC bound to LCs, and characterize transient tertiary structures of IC [52,65,85,87,90–92,97]. In many cases, the results listed above were jointly analyzed with data from other biophysical methodologies such as circular dichroism (CD), isothermal titration calorimetry (ITC), and analytical ultracentrifugation (AUC). Although it is important to confirm results from disparate techniques, it is also notable that NMR alone is able to provide atom- and residue-level information about the structure and dynamics of biomolecules in solution. Examples include residue specific average structure, transient intermediate structure, and intermolecular interactions.

Protein-protein interactions can be monitored by adding non-isotopically labeled protein into the sample of an appropriately labeled target protein. Upon binding, peaks for residues that are involved in the interaction shift and/or lose intensity as their chemical environment and/or structure and size are impacted. Therefore, binding sites may be effectively mapped by tracking chemical shifts and decreases in intensities. Protein dynamics are measured using (^{1}H)-^{15}N heteronuclear nuclear Overhauser effect enhancements (NOEs), T1 and T2 relaxation, and CLEANEX experiments. The heteronuclear NOE experiment, measuring movement of N-H vectors, provides information concerning which parts of the protein are undergoing movement the most rapidly when compared to the overall tumbling of the molecule. Measurements of T1 and T2 relaxation provide information about the overall tumbling rate, and this relaxation is rapid for a free, disordered protein, and decreases if bound in a larger and more structured complex. T2 relaxation also gives information about conformational heterogeneity where multiple different conformations interconvert on microsecond to millisecond time scales giving rise to broad peaks. The CLEANEX experiment, a way of measuring rapid amide hydrogen exchange with water, confirms which residues are disordered and exposed, as their amide backbones are available for rapid exchange as opposed to those that are well-folded or buried. Comparison of free and bound dynamics will reveal what structural

changes are occurring upon complex formation and will pinpoint exact regions of retained disorder post assembly. Information about a protein's secondary structure can be achieved through secondary chemical shift analysis, pinpointing regions of α-helicity (positive values), β-strand (negative values), and disorder (values close to zero). Additionally, three-bond H_N-H_α scalar coupling constants can be used to validate secondary chemical shift analysis as coupling constants in the range of 4-6 Hz imply α-helical secondary structure, 3 Hz or less imply β-strands, and 6-8 Hz imply regions of disorder.

Numerous experiments, each collecting specific and complementary data, are necessary for structure determination. For example, NOE restraints, chemical shift-based backbone dihedral angle restraints, residual dipolar coupling (RDC) restraints, and scalar coupling restrains were combined to create a model of the structural ensemble for an N-terminal construct of IC from *Chaetomium thermophilum* [97]. Additionally, paramagnetic relaxation enhancement (PRE) experiments can be employed, especially when transient tertiary interactions are suspected [91]. Traditionally used to probe long-range contacts, PRE-experiments involve introducing a single cysteine mutation to attach a paramagnetic spin-label (MTSL) [115]. The spin label causes relaxation enhancement to residues nearby in sequence as well as residues that are otherwise spatially proximate [116,117]. Residues that participate in binding or are otherwise spatially close to the spin label will be identified by diminished peak intensities in 2D spectra. Once the necessary distance restraints have been collected, they are entered into structure calculation software to find satisfactory structures that are void of energy violations.

Our lab has made tremendous progress in studying and characterizing N-IC structure and dynamics as well as its interactions with dynein light chains and non-dynein proteins from multiple species (rat, *Drosophila*, yeast, fungus). Our findings have provided a new appreciation of the importance of bivalency in dynein assembly, and of how NudE and $p150^{Glued}$ can compete for the same site of binding on IC [52,65,85,87,90–92,97]. Our methods, described herein, can be applied universally to any disordered protein or region within a larger complex.

Materials

Protein Expression

1. *E. coli* Rosetta (DE3) cells containing expression plasmid. When designing the plasmid there are general guidelines and recommendations to keep in mind (see Note 1).
2. Luria-Broth (LB) Media: 10 g NaCl, 10 g Bacto-Tryptone, 5 g yeast extract in 1 L dIH_2O. Autoclave to sterilize and store at 25°C.
3. 100 ml, 0.1 g/ml $MgSO_4$ stock. Autoclave to sterilize and store at 25˚C.
4. 1,000× trace metal mix: 2 mM H_3BO_3 , 2 mM $CuSO_4$, 2 mM $CoCl_2$, 10 mM $MnCl_2$, 2 mM $NiSO_4$, 2 mM $(NH_4)_6Mo_7O_{24}$. Filter sterilize and store at 25°C.
5. MJ9 minimal media: 1.0 g $^{15}NH_4Cl$, 9.0 g KH_2PO_4, 6.0 g K_2HPO_4, 0.5 g $NaC_6H_7O_7$ in 1 L dIH_2O. pH adjust to 7.2-7.4 and autoclave to sterilize. Post autoclave add in 10.0 ml of $MgSO_4$ stock, 400 µL of trace metal mix, 650 µL MEM Vitamins, and a sterile filtered 0.3% glucose (w/v) and 0.0005% (w/v) thiamine solution. Media should be made with $^{15}NH_4Cl$ as the sole nitrogen source in order to obtain uniformly ^{15}N labeled sample. However, it is recommended that preliminary test expressions with $^{14}NH_4Cl$ are performed to ensure optimal protein expression and solubility (see Note 2). ^{13}C glucose may also be added post-autoclave depending on desired experiments
6. 1 M stock solution of Isopropyl β-D-1-thiogalactopyranoside (IPTG) in dIH_2O, sterile filtered, stored in 1 ml aliquots at -20˚C.
7. 50 mg/ml stock solution of Kanamycin in H_2O, sterile filtered, stored in 1 ml aliquots at -20˚C. Substitution of a different antibiotic may be necessary, depending on the plasmid used.

Protein Purification

1. 1,000× Phenylmethanesulfonylfluoride (PMSF) (0.17 mg/ml) and Pepstatin (0.30 mg/ml) stock solution in 100% ethanol. Store at -20°C.

2. Lysis/Equilibration Buffer: 50 mM phosphate, 300 mM NaCl, 5 mM $C_3H_4N_2$, 1 mM NaN_3, pH 8, store at 25˚C.
3. Wash Buffer: Same as Lysis/Equilibration buffer but with 10 mM imidazole ($C_3H_4N_2$).
4. Elution Buffer: Same as Lysis/Equilibration buffer but with 350 mM imidazole.
5. Final Wash Buffer: Same as Lysis/Equilibration buffer but with 500 mM imidazole.
6. TALON Metal Affinity Resin
7. 2× Loading dye: 40% (w/v) glycerol, 125 mM Tris–HCl, pH 6.8, 100 mM DTT, 2% (w/v) sodium dodecyl sulfate (SDS), 0.025% (w/v) bromophenol blue.
8. ~100 µM TEV protease stock, stored in 1 ml aliquots at -20˚C.

NMR Sample Preparation

1. NMR buffer: A stock of working NMR buffer (see Note 3) such as 500 ml of 50 mM NaCl, 50 mM phosphate, 1mM NaN_3, pH 6.5. Stored at 25˚C.
2. NMR tube: Use a clean, dry NMR tube. Conventional, shigemi, or shaped tubes are all acceptable (See Note 4).
3. D_2O
4. 1 ml, 100 mM stock of 2,2-dimethylsilapentane-5-sulfonic acid (DSS) in dIH_2O. Store at 4˚C.
5. cOmplete protease inhibitor cocktail, stored in 20 µL aliquots at -20˚C.

NMR Data Acquisition/Processing

1. Pulse Sequences: dependent on desired information. ^{15}N-TROSY or ^{15}N-HSQC experiments are suggested for initial validation, whereas triple resonance sequences (HNCO, HN(CA)CO, HNCACB, HN(CO)CACB) should be used for backbone assignment. CLEANEX, T1, T2, and HetNOE experiments are also often commonly used to detect the most flexible residues

and conformational dynamics information. NOESY, HCCH TOCSY, and HNHA are additional possible experiments (see Note 5).

2. High field NMR spectrometer, 700-MHz or higher. We utilize an 800-MHz Bruker Avance III HD NMR spectrometer equipped with a 5-mm triple resonance (HCN) cryogenic probe.
3. Data processing: any processing software can be used for post-acquisition processing of the NMR spectra collected. Examples include TopSpin 4.0.7 (Bruker) and NMRPipe [118]. Peak assignment and relaxation analysis can be performed using CCPN Analysis 2.5.0 [119] and/or nmrviewJ [120]. For 3D experiments that employ nonuniform sampling, spectra should be reconstructed using SCRUB [121]. Backbone dihedral angle restraints may be generated using TALOS-N [122]. Structure calculations may be done using software such as ARIA 2.3 [123].

Methods

In our experience with the methods below, "target protein" refers to the dynein IC while "binding proteins" are proteins such as $p150^{Glued}$, NudE, Tctex, LC8, and LC7. We've chosen to be more general in our descriptions so that they may easily be applied to other protein subunits and regions within dynein, as well as the numerous dynein binding proteins in existence.

Protein Expression

Prior to following the steps below, prepare and autoclave all necessary media (LB and isotopically labeled MJ9) and the proper containers (test tube and flasks). The container used in each growth step should be three to five times the volume of the media for proper aeration to take place. All post-autoclave additives to the MJ9 media should also be added. See Figure 2.3 for a simplified procedural diagram.

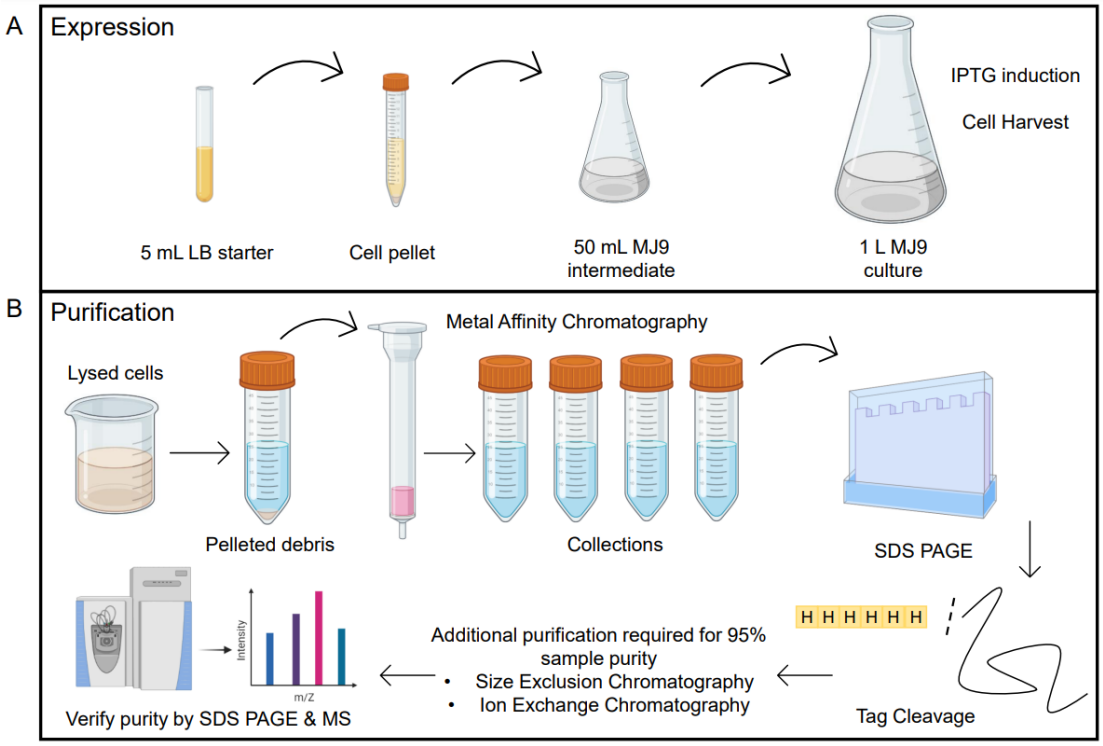

Figure 2.3. Methods for protein expression and purification. A) Schematic of protein expression beginning with inoculation and growth of *E.coli* cells in a 5 ml LB culture. Cells are then pelleted after ~7-12 hrs and resuspended in the 50 ml, isotopically labeled MJ9 intermediate culture. After ~10-12 hrs, cells are transferred into the final, full-sized culture and growth continues until OD_{600}= 0.6-0.8. Cells are then induced with IPTG and harvested 6-16 hrs later. B) Harvested cells are resuspended in lysis buffer, lysed, and cell debris is pelleted. Metal affinity chromatography serves as the first round of purification and collected fractions are ran on SDS-PAGE. Further purification includes tag cleavage, gel filtration and/or ion exchange chromatography in order to reach ~95% sample purity. Sample purity is verified by SDS-PAGE and mass spectrometry to ensure there is no aggregation and/or degradation. Portions of this figure were created with Biorender.com.

1. Using a frozen glycerol stock or single colony from an LB plate with appropriate antibiotic, inoculate 5 ml of LB media containing the appropriate antibiotic. Incubate the 5 ml culture at 37˚C in a shaking incubator at 200 rpm until cloudy (~7-12 hrs). We find it is most convenient to start this growth in the morning/early afternoon, and let the growth proceed throughout the day.
2. Once 5 ml starter is grown (significantly cloudy), pellet the cells at 1,000 x g for 10 min at 20˚C. Decant off the LB media into the waste and resuspend the

cells in 50-100 ml of the MJ9 media. Grow the 50-100 ml starter culture at 37˚C in a shaking incubator at 200 rpm for 10-12 hrs. We find it convenient to let this starter grow overnight.

3. Following the growth of this starter culture, either place the starter at 4˚C to pause growth, or transfer the starter culture directly into the full-sized MJ9 growth media (typically 1 L; see Note 6). When transferring the starter culture to the larger volume, make sure to take an initial OD_{600} measurement to track cell growth.
4. The full-sized MJ9 growth (typically 1 L) is achieved at 37˚C in a shaking incubator at 200 rpm until OD_{600} = 0.6 – 0.8 is reached. At this point, 0.4 mM IPTG is added. Expression may be continued at 37˚C for 4-6 hrs, or the incubator temperature may be dropped to 26˚C for overnight expression (12-16 hrs).
5. Harvest cells at 3,000 x g for 20 min at 4˚C. Cell pellets may be used directly for protein purification, or stored at -20˚C or -80˚C.

Protein Purification

There are many useful methods of protein purification. Briefly described below is our workflow using metal affinity chromatography as an initial step, followed by His_6 tag cleavage. See Figure 2.3 for a simplified procedural diagram.

1. Resuspend the cell pellet in 20-25 ml of lysis buffer. At this time, also add the appropriate amount of protease inhibitor stock solution. Once cells are thoroughly resuspended, sonicate the slurry until cells are properly lysed. Clarify the lysate via centrifugation at 17,500 x g for 30 min at 4˚C.
2. Prepare and equilibrate the affinity chromatography gravity column. A bed volume of 3-5 ml is appropriate and should be equilibrated with 5x the bed volume of lysis buffer. Add the clarified lysate supernatant onto the column and collect the flow through.

3. Proceed to wash the column with wash, elution, and then final wash buffer. Visualize each collection via SDS-PAGE to verify protein expression and purification. The target protein should predominantly be in the elution collection.
4. Cleave off the His_6 tag using TEV protease. Purification of the cleaved target protein from TEV and the His_6-tag may be done in a variety of ways (see Note 7)

Often, further purification steps will be needed in order to reach the desired 95% purity. We suggest methods such as ion exchange chromatography and/or size exclusion chromatography (SEC). Sample purity should always be assessed via SDS-PAGE gel and mass spectrometry (the latter for first time samples) at the end of the purification prior to final NMR sample preparation. Avoiding issues such as sample aggregation and degradation are imperative to the collection of high-quality data.

NMR Sample Preparation

1. Protein sample concentration can be variable; however, a more concentrated sample is often desired. We suggest a concentration in the range of 200-500 µM.
2. Following complete protein purification, prepare a sample of appropriate volume in the chosen NMR buffer, for the type of NMR tube that will be used. The sample should also contain appropriate protease inhibitors, 5-10% D_2O (for the NMR lock), 0.2 mM DSS (for calibration), and 1-5 mM of reducing agent if necessary (see Note 8).

NMR Data Collection

Data collection is highly dependent on the quality of the sample and the desired information. See our suggested flow chart for an order of experiments and possible paths of data collection (Figure 2.4).

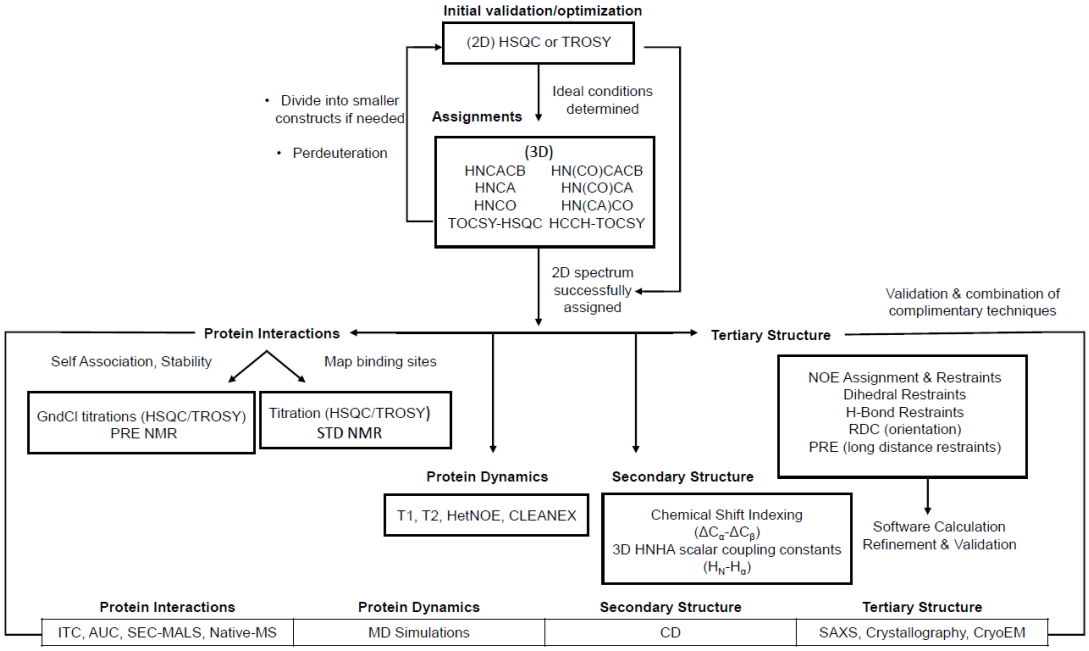

Figure 2.4. NMR data collection flow chart. Different NMR experiments organized by desired information. Where appropriate, alternatives and troubleshooting measures are listed. Complimentary biophysical techniques are listed at the bottom.

A brief outline and special considerations:

1. Begin by collecting a 2D ^{15}N-TROSY or ^{15}N-HSQC spectrum (see Note 5), being sure to test a variety of temperature conditions in order to find an optimal temperature for future experiments. The optimal temperature will be that in which the most peaks possible are visible and of similar intensity (see Note 9).
2. Unless assignments for the target protein have already been obtained, 3D assignment spectra will be required. Since these experiments vary in sensitivity, it's important to process the spectra along the way to ensure parameters have been set properly, so that the data are useable. We recommend continually taking ^{15}N-TROSY or ^{15}N-HSQC experiments to access sample stability and check for degradation (appearance of new and usually less intense peaks). If complications arise preventing adequate and

accurate assignment of the 2D spectrum, a “divide and conquer” method is suggested (see Note 10) and/or protein perdeuteration (see Note 11).

3. After assignments are obtained, additional experiments will depend on desired information (Figure 2.4).
4. To assess protein interactions, perform a titration of non-isotopically labeled binding protein into an appropriately labeled target protein solution, with the goal of achieving maximum binding (~100% bound; see Note 12). If many peaks shift dramatically, as would be the case for a disordered to structured transition, assignment spectra may need to be repeated on a bound sample. For an alternative method see Note 13.
5. To measure protein dynamics, collect data from heteronuclear NOE, T1 and T2 relaxation, and CLEANEX experiments.
6. Secondary structure determination is achieved from chemical shift analysis of the backbone (HNCO, HN(CA)CO, HNCACB, HN(CO)CACB experiments) and side-chain resonance assignments (^{15}N-TOCSY-HSQC and HCCH-TOCSY experiments). In addition, three-bond H_N-H_α scalar coupling constants can be found by measuring the intensity ratio between the cross-peaks and the diagonal peaks in a 3D HNHA experiment and also imply the type of secondary structure present.
7. Numerous experiments are necessary for structure determination. Nuclear Overhauser effect spectroscopy (NOESY), chemical shift-based backbone dihedral angle restraints, residual dipolar coupling (RDC) restraints, and scalar coupling restrains are combined to create a model of the structural ensemble.
8. Paramagnetic relaxation enhancement (PRE) experiments can be employed to probe long-range contacts.

Notes

1. A variety of expression cell lines are appropriate, and we recommend examining available strains and selecting whichever provides the best expression/yield of soluble protein. We design our plasmids using pET type vectors that include an N-terminal His_6 tag followed by a tobacco etch virus (TEV) protease cleavage site. The His_6 tag allows purification via metal affinity chromatography, and TEV cleavage allows tag removal, to minimize the concern of additional residues in NMR spectra. At the plasmid design stage, it may also be best to mutate cysteine residues, especially if the target protein contains many. Cysteine residues should be mutated to either alanine or serine depending on the predicted structural environment of the native cysteine. Finally, if the target protein has a low extinction coefficient (low A_{280}), adding a tryptophan will aid in successful protein concentration determination by A_{280}. If a low A_{280} is unavoidable, we suggest using A_{205} as an alternative [124].
2. We recommend carrying out test expressions in non-isotopically labeled MJ9 prior to a ^{15}N- MJ9 expression. Test expressions can be scaled down to any size, however we recommend a 50 ml minimum. Test expressions can also be used to determine the best expression conditions for induction (temperature and timing).
3. A variety of NMR buffers are appropriate and will depend on the target protein and protein binding partners to be used. Finding appropriate buffer conditions in which the target protein and any binders are soluble is critical. We recommend screening multiple buffers to achieve this [125]. Once a buffer is chosen, making a 500 ml or larger stock is recommended to continually use for final buffer exchange prior to NMR, as doing so allows for greater consistency between samples.
4. Clean, dry, 5-mm NMR tubes should be used. Regular NMR tubes require more volume (~400-600 µL) but often are easier to shim once in the

instrument. To aid in concentration concerns, we often use Shigemi tubes as they only require ~300 µL of volume. For higher salt containing samples, shaped tubes should be considered.

5. Band-selective excitation short transient (BEST) and TROSY variants of 2D and 3D experiments can be advantageous for large, disordered protein samples [126,127]. If the unbound protein is better suited for HSQC type experiments, but there is a desire to study binding with large or multiple partners, it is advantageous to use TROSY experiments instead, as they will likely become necessary as total mass increases.
6. Pausing the intermediate MJ9 starter in the 4˚C is recommended if post-induction expression is to be carried out overnight.
7. One quick and effective way to purify the cleaved target protein away from TEV protease and the His tag is by the use of "reverse" metal affinity chromatography. Ensure that the sample is in a low imidazole buffer prior to following the same protocol steps for the initial affinity purification. During this "reverse" purification the target protein should be present in the flow-through while undesired products (*i.e.*, the His tag) remain in the bound fraction. Other purification methods include those mentioned (SEC or ion exchange). These methods are only appropriate if target protein characteristics are amenable (size, charge, etc.).
8. Any reducing agent may be chosen (*e.g.*, TCEP, DTT, βME). To ensure that a high enough concentration is being used to adequately reduce disulfides, run "reduced" vs. "non-reduced" samples on SDS-PAGE. The non-reduced sample should be made with loading dye void of reducing agent.
9. In general, matching the temperature of other techniques/experiments (ITC) is advantageous for cohesive data analysis. In some cases, one ideal temperature does not exist or is not feasible. When this occurs, it is likely due to having a mixture of disorder and structure, in which peaks for the disordered residues are best observed at a lower temperature, and the structured residues are only

visible at higher temperatures. In such cases, experiments at multiple temperatures become necessary.

10. A “divide and conquer method” refers to the process of breaking up one long protein construct into two or more smaller constructs. This method often works well with long, disordered protein fragments as the 2D spectra for fragments will overlay well with that of the full-length protein as the residues experience the same structure and chemical environment independent of fragment length. Smaller proteins exhibit relaxation rates that are more suitable to 3D experiments and thus can circumvent issues that are encountered with larger protein fragments.
11. Perdeuteration requires protein expression be done in MJ9 media made in D_2O, and the use of deuterated glucose as the ^{13}C source. Expect cultures to grow much slower and take care to include gradual introduction of D_2O; start with the regular 5 ml LB starter, progress to a 25 ml LB D_2O starter, then continue on to the 100 ml and full-sized culture of D_2O MJ9.
12. The number of titration points will depend on many factors, including binding affinity. The goal is to have enough data points to accurately track gradual peak change over the course of binding. It’s common to collect five or more titration points along the way to achieving maximum binding.
13. Saturation-transfer difference (STD) experiments. STD experiments work best when ligands are in fast exchange with binding partners. In an STD experiment, peak intensity is caused by the transfer of magnetization from the ligand to the binding partner [128,129]. Significant STD effects indicate direct transfer of saturated magnetization.

Chapter 3

The dynein light chain 8 (LC8) binds predominantly "in-register" to a multivalent intrinsically disordered partner

Kayla A. Jara[1], Patrick N. Reardon[1], Amber D. Rolland, Delaney A. Smith, Hanh T. M. Hoang, James S. Prell, and Elisar J. Barbar

[1]Authors contributed equally

Published in *J. Biol. Chem.* (2020) 295(15) 4912-4922

Keywords

Intrinsically disordered protein, protein dynamics, LC8, multivalent interactions, analytical ultracentrifugation, nuclear magnetic resonance (NMR), native mass spectrometry.

Abstract

Dynein Light Chain 8 (LC8) interacts with intrinsically disordered proteins (IDPs) and influences a wide range of biological processes. It is becoming apparent that, of the numerous IDPs that interact with LC8, many contain multiple LC8 binding sites. While it is established that LC8 forms parallel IDP duplexes with some partners, such as nucleoporin Nup159 and dynein intermediate chain, the molecular details of these interactions, and LC8's interactions with other diverse partners, remain largely uncharacterized. LC8 dimers could bind in either a paired "in-register" or in a heterogeneous "off-register" manner to any of the available sites on a multivalent partner. Here, we show using NMR chemical shift perturbation, analytical ultracentrifugation, and native electrospray ionization mass spectrometry, that LC8 forms a stable "in-register" complex when bound to an intrinsically disordered domain of the multivalent regulatory protein ASCIZ. Using saturation transfer difference NMR, we demonstrate that at sub-stoichiometric LC8 concentrations there is preferential binding to one of the three LC8 recognition motifs, showing for the first time that the binding process is "in-register". Further, the dynamic behavior for the three sites and the size of the fully bound complex confirm an "in-register" complex. Dynamics measurements also reveal that coupling between sites is dependent on the linker length separating these sites. These results identify linker length and motif specificity as drivers of "in-register" binding in multivalent complex assembly.

Introduction

LC8 is a highly conserved protein found across nearly all eukaryotes. First identified as a component of cytoplasmic dynein[48], LC8 is now recognized as a hub protein that interacts with >100 different proteins involved in numerous cellular processes as diverse as synaptic signaling, virus replication, and apoptosis[50,130,131]. LC8 is a homodimer of 10kDa subunits, with symmetrical binding grooves for binding an eight amino acid motif in intrinsically disordered protein (IDP) clients[132,62,35]. Thus LC8 promotes higher order complex formation by dimerizing client proteins binding on each LC8 protomer.

The short linear recognition sequence includes highly conserved TQT amino acids in the C-terminal half of the motif that are essential for binding[62,78,133]. High resolution structures show that this characteristic TQT forms critical interactions that anchor the peptides to LC8 and support high affinity binding[43]. The residues flanking the TQT anchor are less conserved, but contribute significantly to the overall binding affinity and correct orientation of the motifs in the binding grooves, supporting the formation of parallel IDP dimers[49]. While the majority of identified LC8 binding motifs have the TQT anchor, some variations in these core residues occur in natural sequences[36,49].

A number of LC8 client proteins display multiple LC8 binding motifs that are thought to play an important role in their biological functions[17,59,134,135]. Nup159 from *Saccharomyces cerevisiae* exhibits five such motifs[53,136], separated by short linker regions that are generally only two or three amino acids in length. Characterization by negative stain electron microscopy (EM) of the LC8/Nup159 complex shows that the protein forms a rigid stacked structure. In contrast, another multivalent LC8 binding partner, ASCIZ, has eleven LC8 binding motifs spaced throughout a ~460 amino acid intrinsically disordered region[58,59,137]. The disordered linkers between LC8 binding sites in ASCIZ vary considerably in length, from 3 to 27 amino acids, and could contribute to the heterogeneous population of the LC8/ASCIZ complexes observed by negative stain EM analysis[59].

To date, ASCIZ is the only known regulator of LC8 transcription[138]. ASCIZ has the highest number of LC8 binding sites of any protein identified in either humans or *D. melanogaster*[57,137]. ASCIZ from *D. melanogaster* has six predicted binding sites which were confirmed using short synthetic peptides[59]. A seventh site (called QT3), which contains a TMT instead of the canonical TQT, does not bind as a synthetic peptide, but larger constructs containing QT3 support binding. Interestingly, ASCIZ binding to LC8 forms heterogenous complexes of varying LC8 occupancy, suggesting that the multiple interaction motifs may act as a sensor for LC8 concentration inside the cell[58,59].

The presence of multiple LC8 binding sites on a given IDP raises important mechanistic questions about LC8 binding. For proteins that present a single IDP binding site, interaction with a multivalent client results in decoration of the IDP with the binding protein (Figure 3.1A). The situation for LC8 is more complicated, because LC8 is a homodimer and presents two parallel binding groves. In IDP clients with only one binding site, LC8 dimer interacts with each of two copies, resulting in formation of an IDP duplex (Figure 3.1B). It is tempting to extrapolate this mechanism to IDPs with multiple binding sites, where the LC8 dimers interact with the sites on the client IDPs in a paired or "in-register" fashion, as shown in Figure 1B. However, this is not the only possible model for LC8 binding to multivalent IDPs. LC8 could bind to any of the available sites resulting in "off-register" binding, as shown in Figure 3.1B. A second factor that could influence the formation of LC8/IDP complexes is the differences in the linker length between LC8 binding sites. Indeed, the closely spaced LC8 binding sites in Nup159 appear to yield a relatively "in-register" and rigid structure, while the more varied spacings in ASCIZ yield a more flexible and dynamic structure. These differences in overall flexibility could underlie the functional differences in the respective complexes, where one functions as an assembly scaffold, while the other as a molecular sensor.

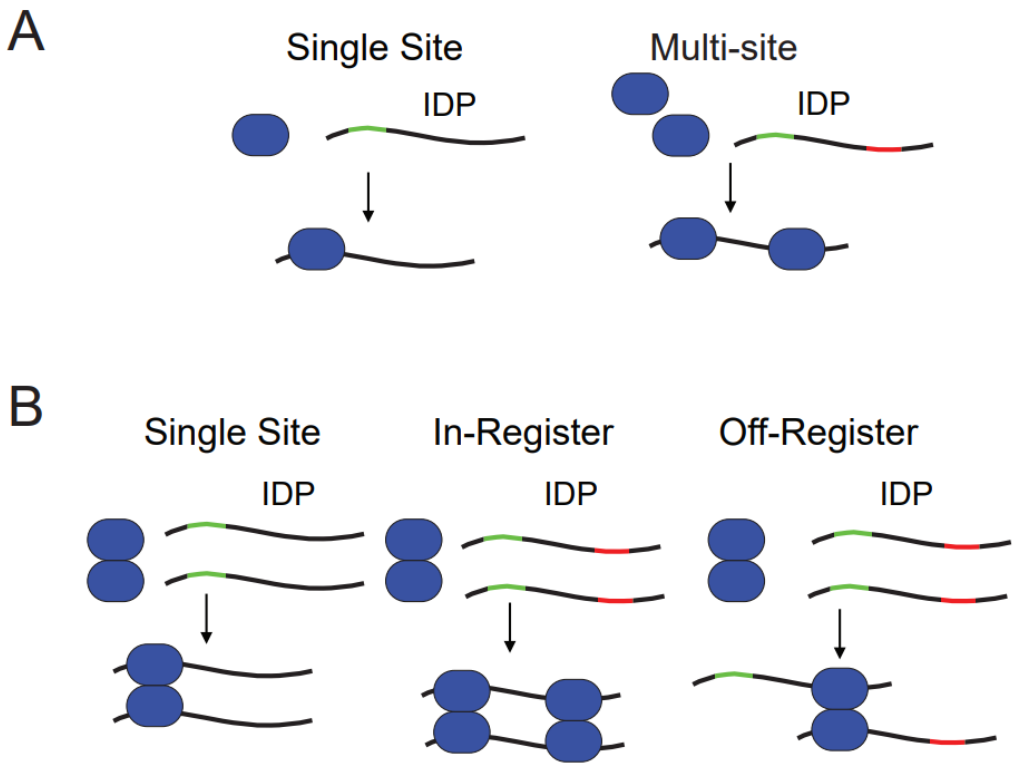

Figure 3.1. Diagram of LC8 interaction models. A) Model of a protein with one recognition motif interacting with an IDP client. Interaction with a monovalent IDP results in a 1:1 complex. Interaction with a multivalent IDP results in decoration of the IDP with the binding partner. B) Models of homodimeric LC8 interacting with an IDP client. Interaction of LC8 with a monovalent (single site) IDP results in dimerization of the IDP. "In-Register" binding model for LC8 interaction with multivalent IDP clients. A given LC8 dimer interacts with the equivalent binding site on each IDP client, resulting in dimerization of the IDP client with multiple bound LC8's. "Off-Register" binding model for LC8 interaction with multivalent IDP clients. A given LC8 dimer interacts with non-equivalent binding sites resulting in the formation of dimers and higher order oligomers. For B, only two LC8 binding sites are shown for clarity.

In this study, we address the mechanism of assembly using a multivalent domain of ASCIZ containing the QT2, QT3 and QT4 recognition motifs, referred to as QT2-4 (Figure 3.2)[59]. This construct was designed to contain 1) multiple LC8 binding sites, 2) varying linker length between motifs, and 3) the QT3 site whose binding to LC8 has not been directly observed. Our study provides the first evidence of "in-register" binding during complex assembly, and suggests a role for linker

lengths in modulating flexibility and LC8 occupancy in multivalent LC8/IDP complexes in general.

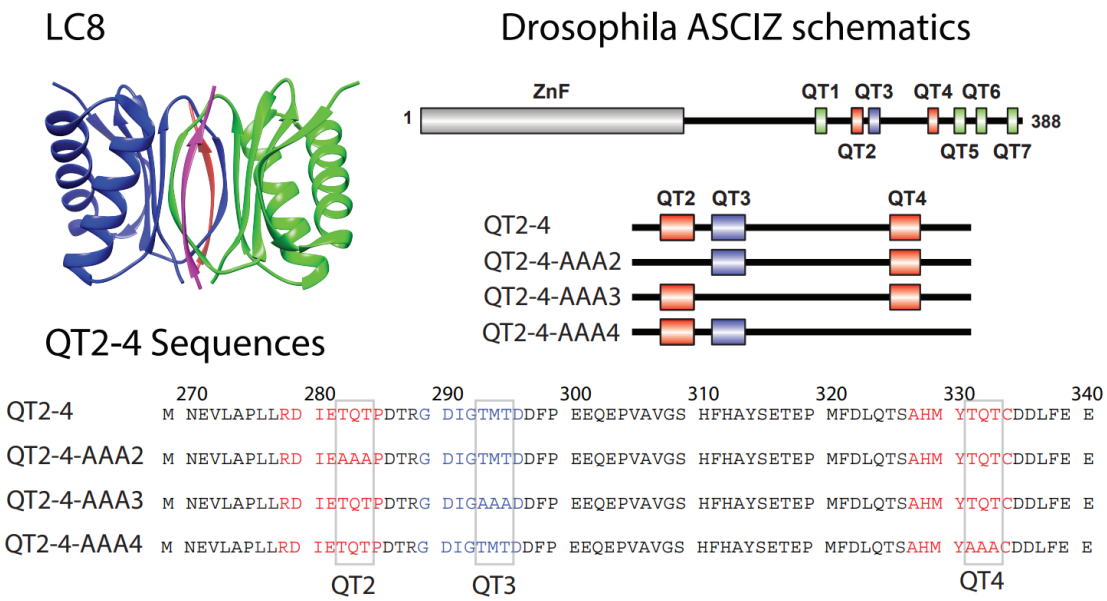

Figure 3.2. Proteins used in this study. Top left panel, Ribbon diagram of LC8 showing each monomer in the LC8 homodimer (blue and green). Bound peptides are shown in red and pink. The coordinates were obtained from PDB 2P2T. LC8 graphic was generated in Chimera[139]. Top right panel, Schematic diagram of ASCIZ showing the location of the LC8 binding sites and the constructs used in this study, including AAA variants that abolish each recognition motif one at a time. Bottom panel, Amino acid sequence of QT2-4. QT2 and QT4 are shown in red. QT3 is shown in blue. Core binding site residues (TQT or TMT) are indicated by grey boxes.

Results

Complex Formation Monitored by Sedimentation Velocity Analytical Ultracentrifugation (SV-AUC) and Native ESI-MS

SV-AUC analysis of the complex shows that the proteins are in a dynamic equilibrium at ratios of QT2-4:LC8 up to 1:3 (Figure 3.3). At a 1:1 molar ratio of QT2-4 to LC8, the sedimentation coefficient is shifted to ~2.9 S, larger than the ~2.5 S observed for LC8 alone and consistent with binding to QT2-4. Higher LC8 concentrations further shift the sedimentation coefficient towards the fully bound state, consistent with incorporation of additional LC8 dimers. We also note that at ratios of 1:2 and greater, we observe a distinct population of unbound LC8. Titrating with a ratio of 1:4 QT2-4:LC8 results in the formation of a fully occupied complex having a sedimentation coefficient of ~4.5 S. Importantly, we found that the complex is stable, exhibiting the same sedimentation coefficient after purification with size exclusion chromatography. We did not detect larger order aggregates or oligomers.

To isolate the contribution of each binding motif to complex formation, we performed SV-AUC on mutant QT2-4 proteins bound to LC8 (see Figure 3.2 for the mutant constructs). QT2-4-AAA4 bound to LC8 exhibited a relatively narrow c(S) distribution with a sedimentation coefficient of ~3.2 S, lower than the LC8 saturated wild type QT2-4 and consistent with a lower molecular weight complex. QT2-4-AAA3 and QT2-4-AAA2 bound to LC8 exhibited a broader c(S) profile, with maximum S values at ~2.8 S and ~2.7 S, respectively. Increased broadness in the c(S) plot suggests that the complexes formed by QT2-4-AAA3 and QT2-4-AAA2 are more heterogeneous than the complex formed by QT2-4-AAA4. The lower sedimentation coefficients are also consistent with dynamic complexes that average to lower molecular weight complexes when compared to wild type or QT2-4-AAA4.

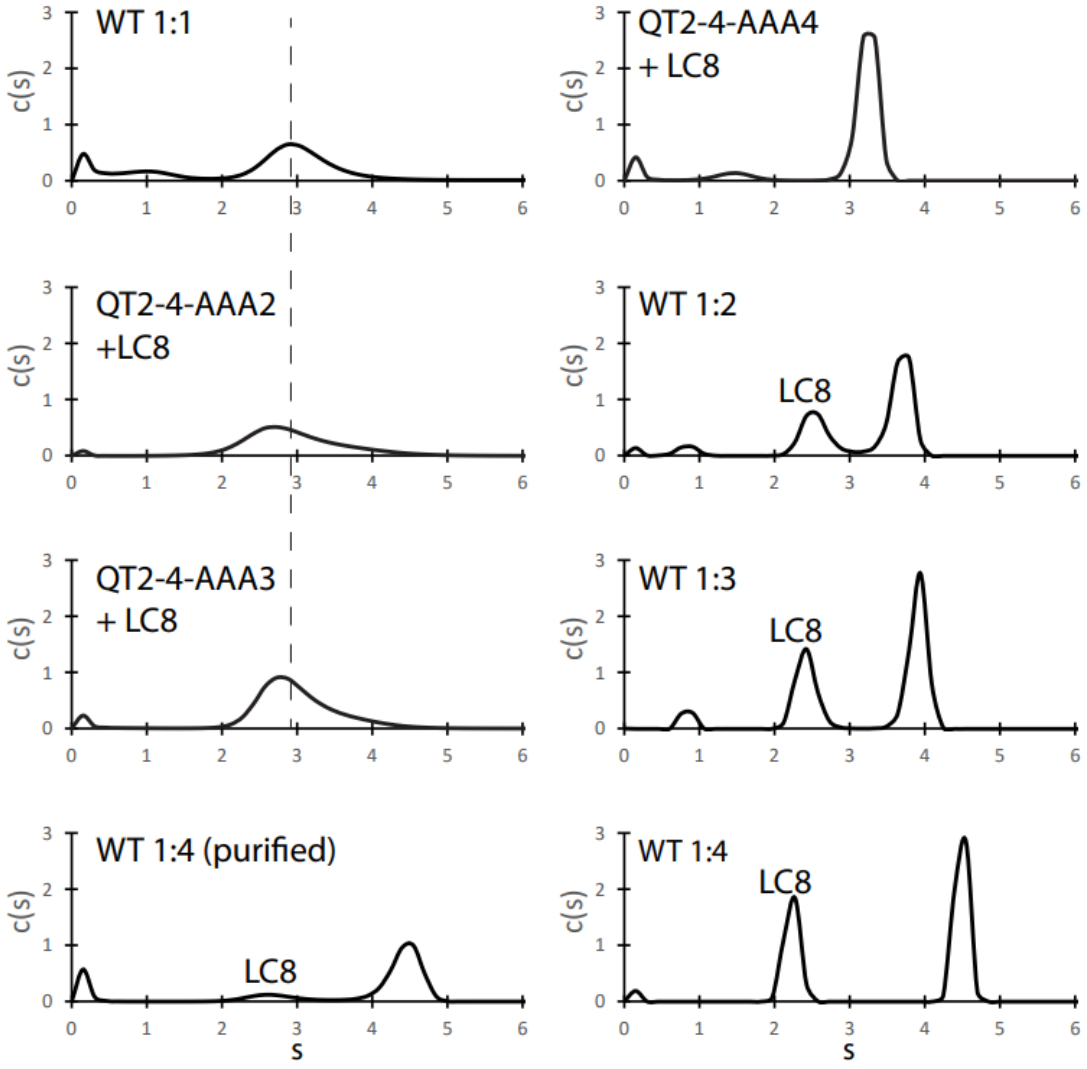

Figure 3.3. Sedimentation velocity analytical ultracentrifugation of ASCIZ QT2-4 bound to LC8. Populations corresponding to free LC8 are labeled. Experiments were performed at four molar ratios of WT QT2-4 to LC8. For the highest ratio (1:4) data are shown with and without purification using size exclusion chromatography to remove excess LC8. Results for mutant QT2-4 are shown after mixing with excess LC8 and purification with size exclusion chromatography. The dashed line is centered on the main peak for the 1:1 sample.

Using native electrospray ionization mass spectrometry (native ESI-MS) measurements of the individual protein subunits along with the mass of the complex(es) allows identification of complex stoichiometries for those formed by QT2-4 bound to LC8. Masses of monomeric QT2-4 (8645.84 ± 0.05 Da) and

monomeric LC8 (10638.5 ± 0.2 Da) were determined first and closely matched the expected masses calculated from each protein sequence (Table 3.1). Further native mass spectra of both QT2-4 and LC8 acquired at a series of diluted concentrations, (Figures A1.1 and A1.2, respectively) confirmed that QT2-4 remains monomeric while LC8 can form dimers in solution.

Table 3.1. Expected and measured masses for all proteins and complexes identified by native mass spectrometry.

ACCURATE MONOMER MASSES		
Protein, z+	**Sequence mass (Da)**	**Measured mass (average ± standard deviation, Da)**
QT2-4, 4-7+	8592.37	8645.84 ± 0.05
LC8, 3-7+	10770.31	10638.52 ± 0.20
MASSES OF QT2-4/LC8 SPECIES		
Species/Complex, z+	**Expected mass (Da) based on measured masses (above)**	**Measured average mass (Da)**
LC8 monomer, 3-8+	10638.52 ± 0.20	10753 ± 2
LC8 dimer, 8-10+	21277.0 ± 0.4	21503 ± 1
QT2-4 monomer + 1 LC8 dimer	29922.9 ± 0.4	30154.2 ± 0.9
QT2-4 monomer + 2 LC8 dimers	51200 ± 1	51765 ± 8
QT2-4 dimer + 3 LC8 dimers	81123 ± 1	81925 ± 21
QT2-4 dimer + 4 LC8 dimers	102400 ± 2	103497 ± 15

Native ESI-MS of the QT2-4/LC8 complex identified six different species, the associated peaks of which are labeled in Figure 3.4. Two of these six correspond to the LC8 monomer and dimer, which are expected to be present upon dilution of the reconstituted complex. The others include QT2-4/LC8 complexes with masses that correspond to QT2-4 monomer bound by one LC8 dimer, QT2-4 monomer bound by two LC8 dimers, QT2-4 dimers bound by three LC8 dimers and QT2-4 dimer bound by four LC8 dimers (Table 3.1). Detected complex abundances are shown alongside calculated Poisson probability distributions in Figure 3.4. The two smaller complexes of QT2-4 monomer bound by one or two LC8 dimers are detected at higher abundances than predicted by Poisson distributions to the lowest concentration studied and thus may represent complex assembly intermediates. The stoichiometry of the largest complexes identified were confirmed via collision-induced dissociation (CID). Of the two larger complexes, $(QT2\text{-}4)_2(LC8\ dimer)_3$ was consistently detected

at higher abundance than $(QT2\text{-}4)_2(LC8\ dimer)_4$. Dilutions of the purified complex (1 µM to 100 nM) show that $(QT2\text{-}4)_2(LC8\ dimer)_3$ is more stable than $(QT2\text{-}4)_2(LC8\ dimer)_4$, with the larger complex not observed at the lowest concentration (Figure 3.4). Taken together, we consider the fully-formed complex to primarily be the more stable $(QT2\text{-}4)_2(LC8\ dimer)_3$ complex.

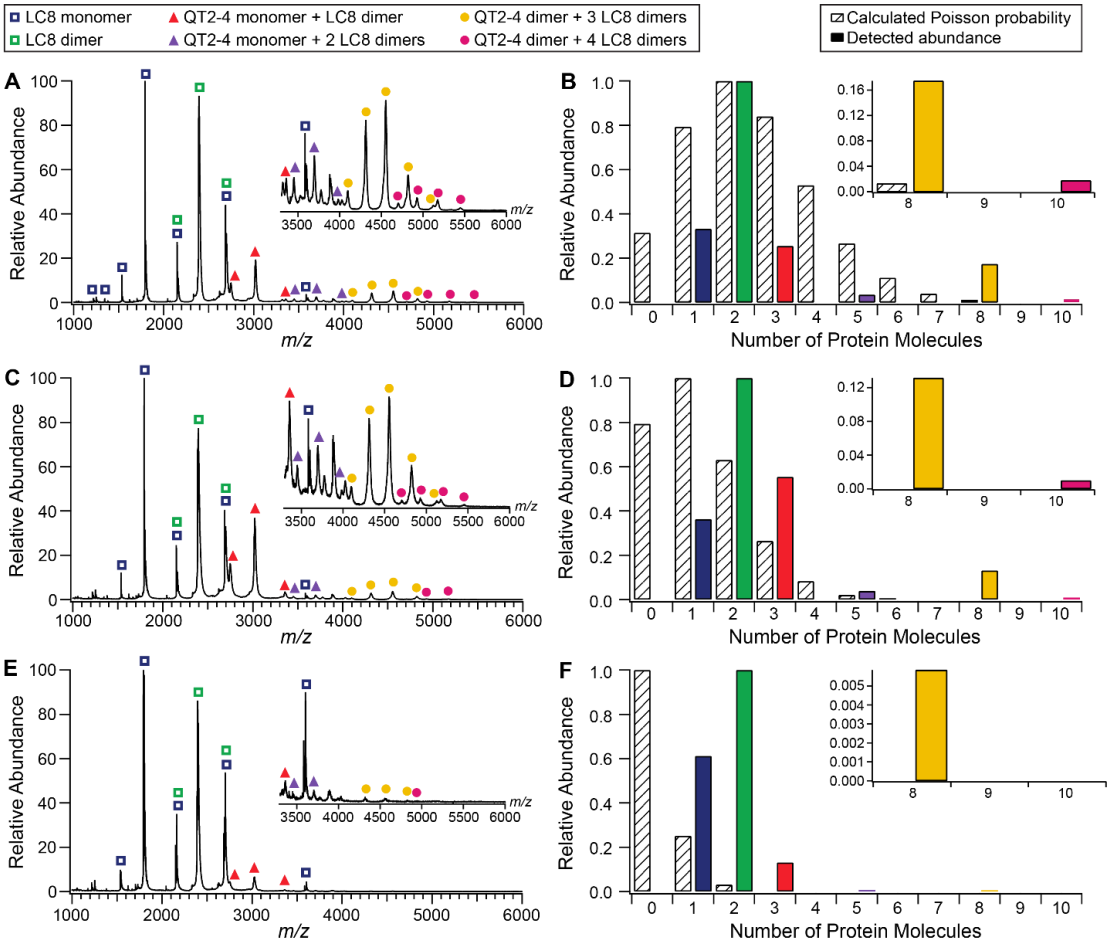

Figure 3.4. Native mass spectra and abundance distributions of QT2-4/LC8 species at a concentration of 1 µM (A, B), 500 nM (C, D), and 100 nM (E, F). Insets in mass spectra panels (A, C, E) show the *m/z* 3300-6000 region of the corresponding mass spectrum within each panel. Peaks associated with each of the six species identified are labeled in the legend. Distribution panels (B, D, F) display the calculated Poisson probability for 0-10 protein molecules in a droplet of 100 nm radius (striped bars) and the experimentally detected abundances of the six QT2-4/LC8 species (solid bars with colors matching the legend on the left). Probabilities and abundances in each main distribution panel were normalized to the most probable

or abundant state. Insets in distribution panels provide clarity for comparison of the relative calculated Poisson probabilities and abundances of the two largest species detected. For experimentally detected abundances, the six identified species are represented by the number of protein molecules as follows: LC8 monomer (1), LC8 dimer (2), QT2-4 monomer + LC8 dimer (3), QT2-4 monomer + 2 LC8 dimers (5), QT2-4 dimer + 3 LC8 dimers (8), QT2-4 dimer + 4 LC8 dimers (10).

Structural Characterization of QT2-4/LC8 Complex

Having determined that the complex composed of two chains of QT2-4 bound by three LC8 dimers (mass of ~80 kDa) is the most populated and stable complex, we assigned its backbone resonances using triple labeled (^{2}H, ^{13}C, ^{15}N) protein and a suite of TROSY-based 3D NMR experiments. Structural characterization of LC8 bound to IDP with multiple LC8 binding sites using NMR spectroscopy has been hampered by the disappearance of resonances corresponding to residues associated with LC8 binding[53,59,87]. Overall peak attenuation is caused by a reduction in global tumbling time, resulting in increased relaxation. To observe these peaks, we used perdeuteration due to its favorable impact on T_2 relaxation[140]. The spectrum of deuterated QT2-4 bound to LC8 shows that many of the QT2-4 resonances remain observable and the bound spectrum exhibits a large number of chemical shift perturbations when compared to the spectrum of unbound QT2-4 (Figure 3.5 and Figure A1.3). Assignments for ~78% of the backbone amide resonances of QT2-4 bound to LC8 are shown in Figure 3.5.

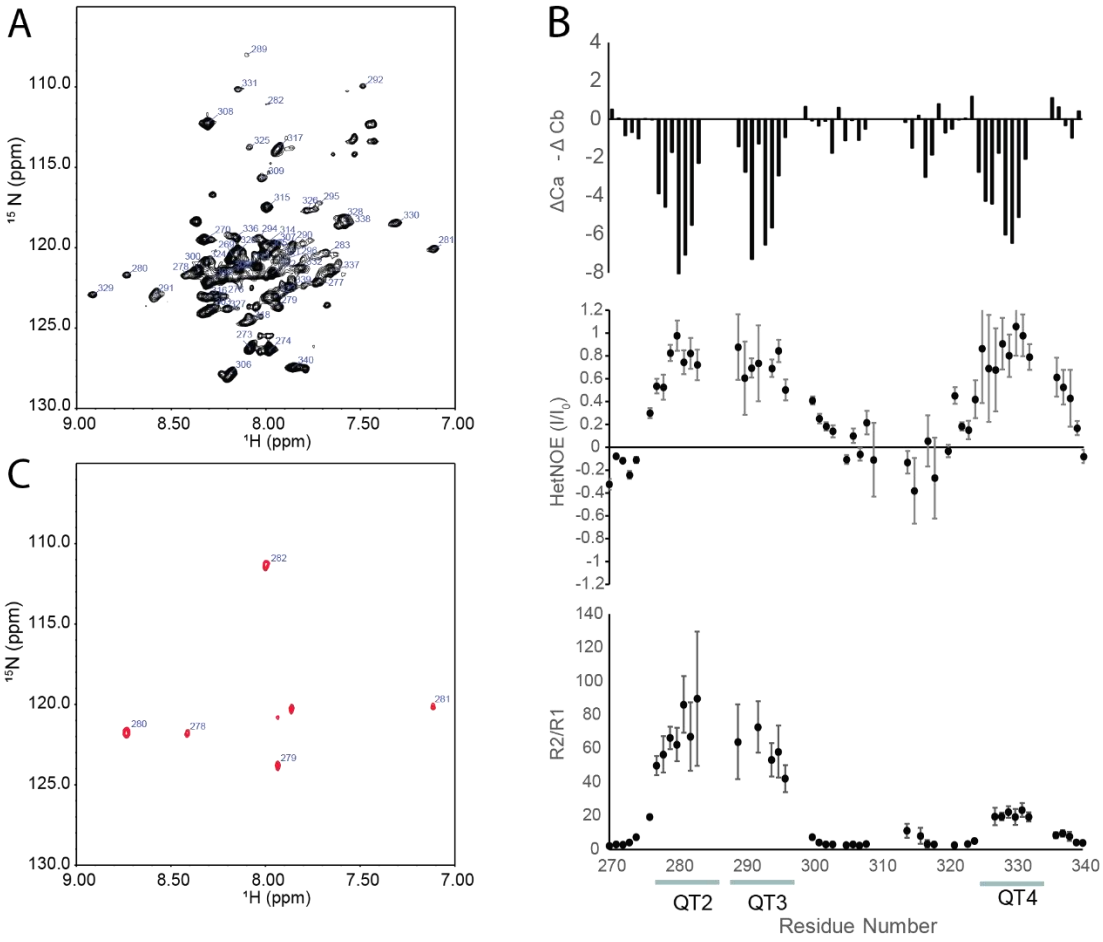

Figure 3.5. Characterization of LC8 bound to QT2-4. A) Assigned TROSY-HSQC spectrum of ASCIZ QT2-4 bound to LC8. B) (top) ΔCa-ΔCb chemical shift indexing for ASCIZ QT2-4 bound to LC8. Regions of contiguous negative values indicate beta-strand, contiguous positive values indicate alpha helix and values near zero indicate disorder. NMR based dynamics analysis of QT2-4 bound to LC8. Heteronuclear NOE (middle) and R_2/R_1 (bottom) are shown. LC8 binding sites are indicated by the green lines. Gaps are due to unassigned residues or residues that could not be accurately fit due to overlap. C) Cross saturation transfer difference spectrum of ASCIZ QT2-4 bound to LC8 at 1:1 molar ratio of QT2-4 to LC8. Resonances that are saturated by irradiation on the LC8 appear as negative peaks (red). Amino acid labels correspond to residues in QT2.

Chemical shift indexing of QT2-4 bound to LC8 reveals that the chemical shifts for the resonances corresponding to the three LC8 binding sites are characteristic of beta-strand structure (Figure 3.5). This observation is consistent with crystal structures of LC8 bound to short peptides containing an LC8 binding site,

where the peptide also adopts beta-strand like structure[37,43,62]. Residues located between QT3 and QT4, such as G308 and T317, exhibit chemical shifts consistent with a random coil and show modest or no chemical shift perturbations upon binding, indicating that the region between QT3 and QT4 remains unstructured and that their local chemical environment remains largely unchanged by LC8 binding.

It is interesting to note that binding of the QT3 peptide to LC8 was not detected using isothermal titration calorimetry (ITC)[59], but the QT2-4 construct binds LC8 with a stoichiometry of 3 and a mass consistent with binding at all three sites. These somewhat contradictory results raise the open question: is QT3 an LC8 binding site? Here we show beyond doubt that QT3 binds LC8 in the context of QT2-4 and adopts a beta-strand in the bound complex. We identified two glycine residues (G289 and G292) in QT3 that undergo significant chemical shift perturbations upon LC8 binding when compared to unbound QT2-4 (See Figure A1.3), while G308, which is not located in an LC8 binding site, does not undergo any chemical shift perturbation upon addition of LC8, confirming that changes in the spectra are at or near the LC8 binding regions. QT3 residues T287, T293 and T295 shift significantly in the LC8 bound spectrum, further demonstrating that LC8 interacts with QT3 in context of the QT2-4 construct. While QT3 lacks the canonical TQT, having TMT instead, binding of TMT has been previously reported[36].

Early Binding Events Identified by Cross Saturation Transfer Difference

Cross saturation transfer difference NMR[141] utilized perdeuterated and ^{15}N labeled QT2-4 mixed with unlabeled LC8 to probe the interface between LC8 and QT2-4. At a molar ratio of 1:1 QT2-4 to LC8, resonances corresponding to residues R278, D279, I280 and E281 showed reduced intensity by saturation on the methyl resonances of LC8, indicating that these residues are in close proximity to LC8. In the absence of LC8, the on-resonance saturation pulse did not result in saturation of any QT2-4 resonances, indicating that any residual protonated QT2-4 that may be present does not significantly influence the spectrum. Interestingly, at this molar ratio, the majority of residues with reduced intensity from the saturation pulse are from QT2,

demonstrating that, of the three binding sites in QT2-4, QT2 is most likely to be occupied at sub-stoichiometric LC8 concentrations (see Figure 3.5). This suggests that the LC8 binding sites in ASCIZ are not equivalent and that LC8 preferentially occupies QT2 over the other sites in QT2-4, even though the affinity for the QT2 linear peptide is only modestly (~2-fold) tighter than the QT4 linear peptide[59]. Bound state peak intensities are generally highest in the QT2 motif at the lower LC8 ratios, consistent with more stable interaction at this site (Figure A1.4).

Dynamics Analysis of QT2-4 Bound to LC8

At a saturating concentration of LC8, where all three sites are fully occupied, R_2 increases at each recognition motif while R_1 decreases, consistent with increased apparent rotational correlation time and LC8 binding (see Figure A1.5). A plot of R_2/R_1 conveys this trend, with a dramatic increase in R_2/R_1 at each of the LC8 binding regions relative to the linker (Figure 3.5). In contrast to the binding regions, the linker exhibits R_2/R_1 ratios consistent with unstructured protein. Heteronuclear NOE analysis confirms that all three binding regions become structured upon interaction with LC8, with peak intensity ratios approaching 1 (Figure 3.5). The linker region remains unstructured with peak intensity ratios near or below zero. Interestingly, the R_2/R_1 ratio of residues in QT2 and QT3 are similar in magnitude and both much larger than the R_2/R_1 ratio of residues in QT4. This observation indicates that QT2 and QT3, and the associated LC8 dimers, exhibit similar restricted global correlation times and likely tumble as a single unit. In contrast, the lower R_2/R_1 ratio observed in QT4 suggests that the unstructured linker between QT3 and QT4 decouples motion between these binding sites and allows QT4 to tumble independently of QT2 and QT3.

Discussion

Macromolecular complexes involving IDPs have been implicated in a wide range of biological processes. Key to the function of these complexes is the

mechanism by which they form and the overall architecture of the resulting complexes. Our study examined the mechanism and architecture of the complex formed between dimeric LC8 and multivalent QT2-4 with 3 sites of varying motif specificity and linker lengths separating them. We show the first evidence for preferential binding to one motif that drives “in-register” complex formation and identify both compositional and conformational heterogeneity as common features of LC8/ASCIZ complexes.

“In-register” vs. “Off-register” Binding

As illustrated in Figure 3.1, LC8 interacting with a multivalent binding partner could do so either “in-register” or “off-register”. To date, there has been little evidence that differentiates between these two models. Our data support “in-register” binding and lead to an overall binding model that is illustrated in Figure 3.6A. “In-register” binding is consistent with our observation of a single fully bound complex in sedimentation velocity and a single set of resonances in the ^{15}N TROSY NMR spectrum. The saturation transfer difference experiment showed that the majority of saturation transfer is localized to a single binding site at a low ratio of LC8 to QT2-4, an observation that is consistent with “in-register” binding, as “off-register” binding would result in saturation transfer at all of the binding sites due to lack of client site differentiation by LC8. Finally, native ESI-MS shows that the predominant species in the fully bound complex consists of two monomeric QT2-4 chains and three LC8 homodimers, as expected from the “in-register” binding model. While other complexes are observed by native ESI-MS, the detected populations indicate that they are minor populations when compared to the “in-register” complex. Furthermore, we did not observe evidence for these complexes by SV-AUC, consistent with them being lower population states in solution when compared to the “in-register” complex.

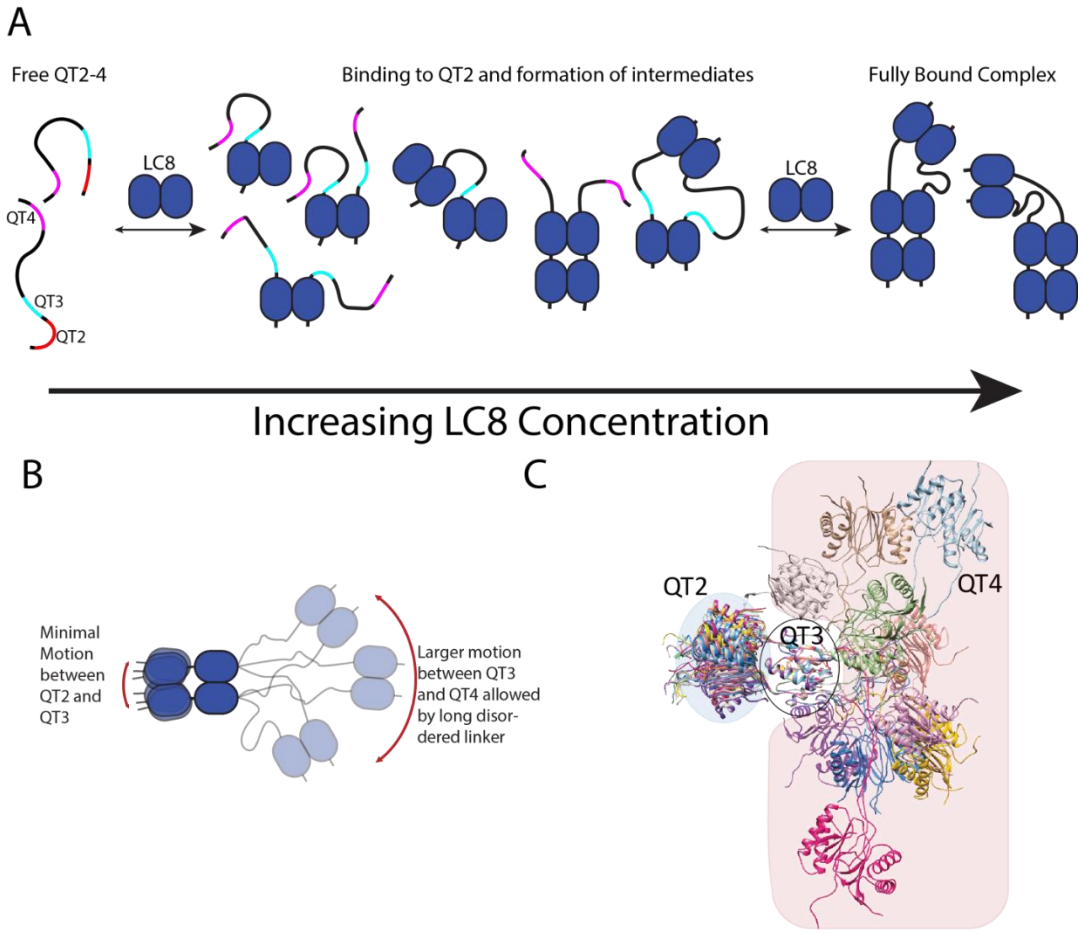

Figure 3.6. LC8/QT2-4 complex assembly. A) Model of LC8 binding to QT2-4 showing compositional heterogeneity. The blue ovoids represent LC8, while the black lines represent QT2-4. Binding sites on QT2-4 are shown in red (QT2), cyan (QT3) and magenta (QT4). Binding is initiated by interaction of LC8 with QT2, binding to either one chain of QT2-4 or to two chains forming a duplex of QT2-4. Subsequent LC8 homodimers interact with the newly formed duplex. The close proximity of the dimerized QT2-4 chains promote "in-register" binding. At high concentrations of LC8, a stable complex is formed, with all LC8 binding sites occupied. B) Model of QT2-4 bound to LC8 showing conformational heterogeneity. The black lines represent ASCIZ QT2-4. The red arrows indicate the extent of motion imparted by the disordered linkers. Motion between QT2 and QT3 is relatively modest due to the short linker and increase steric constraint between LC8 dimers. Motion between QT3 and QT4 is larger due to the conformational freedom provided by the long disordered linker. C) Atomistic model of the relative motional freedom of QT2-4. Models aligned on LC8 bound to QT3. Blue shaded area emphasizes the freedom of QT2, red shaded area emphasizes the freedom of QT4. The model was generated by fixing the interaction motif and LC8 coordinates to those observed in pdb file 2P2T and performing molecular dynamics in XPLOR-NIH. The model is an illustration of the possible variation in motion between binding sites, not an explicit ensemble structure.

The R_2/R_1 data also support "in-register" binding. If binding were "off-register" each LC8 monomer would bind a different LC8 site on QT2-4, resulting in linkers of differing length between each LC8 dimer. In turn, this is expected to give rise to similar R_2/R_1 values for each binding site, as motion between each binding site would be the result of a mixture of the different possible binding modes. Instead, we see different R_2/R_1 behavior for QT2 and QT3, vs QT4, as predicted for "in-register" binding.

"In-register" binding of stable complexes is observed with electron microscopy (EM). Negative stain EM on the cytoplasmic pore filaments of the nuclear pore complex from yeast shows a rigid, ladder like assembly of the light chain, Dyn2 (LC8 in yeast), supporting "in-register" binding[142]. Another recent cryo-EM structure of the dynein-2 complex suggests that LC8 forms a complex that is architecturally similar to an "in-register" complex, having the same ladder-like structure, despite significant sequence differences in the binding motifs[143]. With negative stain EM, it is only possible to image LC8 and not the IDP partner and therefore "in-register" binding is inferred and not directly observed. Our work presents direct evidence that LC8 preferentially binds at one site and forms "in-register" complexes with multivalent ASCIZ QT2-4.

Step-wise Dynamic Binding

We propose a step-wise model for LC8 interaction with multivalent ASCIZ QT2-4 that is initiated at the QT2 site of highest affinity, as determined from saturation transfer, dimerizing the client protein and orienting the client for "in-register" binding (Figure 3.6A). Subsequent addition of LC8 is supported by the close proximity of the dimerized chains, with weaker binding sites, such as QT3, experiencing enhanced affinity from nearby LC8 interaction. Previous studies on model systems have shown that tethering recognition motifs to binding proteins enhances the interaction by increasing the effective concentration[144,145]. LC8 affinity enhancement due to bivalency has been fully characterized when other light chains

interact with a nearby site[81]. Additional intermediate complexes having only a single QT2-4 chain are also possible, as indicated by our native ESI-MS data.

Interestingly, while LC8 initiates binding to QT2-4 by interacting with QT2, subsequent interaction with the other sites appears to be more complicated. We note that the fully bound complex between QT2-4 and LC8 is stable and does not rapidly dissociate, indicating that the overall off-rate must be slow. Importantly, if a stable fully bound complex is formed, we would expect to detect the complex in SV-AUC, as the overall off-rate would be slow on the sedimentation timescale. However, we do not observe the fully bound complex in the 1:3 or lower ratio complexes, suggesting that the fully bound complex is not formed until high levels of LC8 are present.

The exact nature of the intermediates remains to be fully determined, but our results provide additional insight. SV-AUC analysis of the binding motif mutants show that each site does not contribute equally to overall complex formation. In QT2-4-AAA4, the presence of QT2 and QT3 supports formation of a stable complex, consistent with increased occupancy of the binding sites. This result indicates that close proximity between QT2 and QT3 enhances binding at QT3. In contrast, mutation of either QT2 or QT3 results in a dynamic mixture of complexes with average lower molecular weight. Taken together, these results provide evidence that linker length plays an important role in multivalent interactions with a longer linker resulting in complexes with more heterogeneous dynamic equilibrium.

The sedimentation coefficients for the mutant QT2-4 proteins bracket the wild type QT2-4 sedimentation coefficient at the 1:1 ratio, suggesting that wild type QT2-4 bound to LC8 at the 1:1 ratio is a mixture of complexes that averages between one and two LC8 homodimers bound to QT2-4. The complexes could consist of a mixture of LC8 bound at any of the binding sites, but QT2 is likely more stably occupied based on the saturation transfer results. In the context of the full length ASCIZ, we also observe a complex dynamic set of intermediates, which tune ASCIZ transcription activity, rather than behaving like a simple on/off switch[59]. A set of partially occupied stable complexes also observed in the shorter QT2-4 confirms that a dynamic

complex of multiple partially occupied LC8/ASCIZ complexes is a feature of ASCIZ and possibly other multivalent partners.

Complex Flexibility

In addition to the compositional heterogeneity that is observed during the titration, the fully occupied LC8/QT2-4 complex has significant conformational heterogeneity, as it is a mixture of flexible and rigid regions, illustrated in Figure 3.6B and C. The short linker between QT2 and QT3 greatly constricts the overall motion between LC8 homodimers bound at these locations, as inferred from the large R_2/R_1 values, and forces them to behave as a unit. The long disordered linker between QT3 and QT4 enables QT4 to move more freely with respect to QT3 and QT2. Thus, differences in disordered linker length dictate the relative conformational freedom of the complexes formed by multivalent LC8 binding proteins. Finally, the increased rigidity at QT2 and QT3 may further support "in-register" binding by limiting interactions between sites flanking the QT2 and QT3 unit. Results with the Rabies virus protein, RavP, also show that LC8 binding can alter the conformational space sampled by a protein, and limit interactions between sites separated by LC8[55]. With Nup159, LC8 binding forms a rigid structure that is proposed to enhance the cytoplasmic accessibility of Phe-Gly repeats to nuclear transport proteins. Nup159 has short disordered linkers separating the LC8 binding sites, similar to the linker between QT2 and QT3 in QT2-4, and thus would have similar restricted mobility. In contrast to Nup159, QT2-4 retains IDP-like behavior in the region between QT3 and QT4, suggesting that LC8 bound ASCIZ would not adopt a fully rigid structure. Therefore, the mechanism of action for ASCIZ does not rely on formation of a fully rigid structure, as is the case with Nup159, and this added flexibility could offer another layer of regulatory control that contributes to the buffered transcriptional activity observed for ASCIZ.

Conclusions

Here we show that binding of LC8 to a multivalent partner results in formation of an "in-register" complex. Such "in-register" binding guides the formation, structure and function of higher order complexes. The dynamic behavior of multivalent LC8 complexes is linked to the binding affinity of the individual sites and the length of the intervening disordered linkers. Longer disordered linkers introduce increased flexibility between regions with bound LC8, resulting in both compositional and conformational heterogeneity, while shorter linkers can lead to enhanced binding of nearby weaker sites and reduced flexibility in the complex. The combination of motif affinity and specificity, along with variation in disordered linker length in a multivalent partner, is a promising emerging mechanism for an exquisitely tunable system of binding and regulation. We anticipate this mechanism to be general across many biological processes, given the growing number of multivalent LC8 client proteins and the essential role of LC8 in nearly all cellular functions.

Experimental Procedures

Cloning, Protein Expression and Purification

Cloning of *Drosophila* ASCIZ QT2-4 (ASCIZ residues 271-341) (Figure 3.2) with various mutations of recognition motifs was performed using QuickChange Lightening Mutagenesis kit (Agilent). Resulting constructs verified by sequencing are QT2-4-AAA2, QT2-4-AAA3, and QT2-4-AAA4 where the number indicates the LC8 recognition motif whose TQT was replaced with AAA and thus has lost binding at this particular site. Proteins were expressed and purified according to previously published procedures[59]. For perdeuteration, *Escherichia coli* Rosetta DE3 cells, transformed with a pET2Zt2-1a vector with a sequence encoding QT2-4, were grown in Luria Broth prepared in 99.9% D_2O overnight. Modified M9 minimal media prepared with 99.9% D_2O was inoculated from the overnight culture. ^{13}C and ^{15}N were supplied with uniformly labeled ^{2}H-^{13}C-glucose (0.01 M) and ^{15}N ammonium chloride (0.02 M), respectively. Cultures were grown at 37° C to an optical density of

0.6 at 600 nm, induced by adding IPTG to a final concentration of 0.4 mM, and harvested after 6 hours of induction. Proteins were purified under denaturing conditions using TALON His-Tag Purification protocol (Clontech), then dialyzed into 20 mM Tris-HCl, 5 mM β-mercaptoethanol, 1 mM sodium azide, pH 8.0 affinity tag cleavage buffer. Complete cleavage of the His-Tag required incubation with in-house produced His-tagged TEV protease (1:100) for 4 hours at room temperature and confirmed on SDS-PAGE. Further purification using anion exchange Macro-Prep High Q Support resin (Bio-Rad, Hercules, California) was followed by size exclusion chromatography on a Superdex 75 column (GE Healthcare) in a buffer composed of 20 mM Tris-HCl, 50 mM NaCl, 5 mM β-mercaptoethanol, 1 mM sodium azide, pH 7.5 yielded protein with >95% purity as determined by SDS-PAGE. Protein concentrations were determined by absorbance measurements at 280 nm with extinction coefficients for LC8 of 14,440 M^{-1} cm^{-1} and QT2-4 of 2,980 M^{-1} cm^{-1}. When estimating a concentration for SEC purified complexes, absorbance at 280 nm was still used along with the assumption that the majority of formed complex in solution followed the expected stoichiometry of 1:3 (QT2-4 monomer: LC8 dimer) and that very little excess of either free protein would be present.

Analytical Ultracentrifugation

Sedimentation velocity analytical ultracentrifugation (SV-AUC) was performed using a Beckman Coulter Optima XL-A analytical ultracentrifuge, equipped with absorbance optics. LC8 was mixed with QT2-4 at ratios of 1:1, 1:2, 1:3 and 1:4 (molar ratio of QT2-4:LC8). An additional sample of QT2-4 was mixed with LC8 at a 1:4 molar ratio then re-purified using gel filtration chromatography on a Superdex 200 column (GE Healthcare). SV-AUC of the QT2-4 AAA complexes was performed on reconstituted complexes purified by gel filtration chromatography. The concentration of protein complex in the final samples was estimated to be 10 µM. Buffer conditions for SV-AUC analysis were 20 mM Tris-HCl, 50 mM NaCl, 5 mM Tris(2-carboxyethyl)phosphine, 1 mM sodium azide, pH 7.5. The complexes were loaded into standard, 12 mm pathlength, 2-channel sectored centerpieces and

centrifuged at 42,000 rpm and 20° C. 300 scans were acquired at 280 nm with no interscan delay. Data were fit to a c(S) distribution using the software SEDFIT[146]. Buffer density was calculated to be 1.0009 g/ml using Sednterp[147].

Native Electrospray Ionization Mass Spectrometry (Native ESI-MS)

Purified samples of the individual proteins (QT2-4 and LC8) and of the QT2-4/LC8 complex were flash-frozen in liquid nitrogen and stored at -80 °C until further use. A 50 µL aliquot of each sample was buffer exchanged into 200 mM ammonium acetate, pH 7.50, using Micro Bio-Spin™ columns (Bio-Rad Laboratories, Inc.). All native mass spectra were collected using a Waters Synapt G2-S*i* time-of-flight mass spectrometer with a nanoelectrospray ionization (nanoESI) source. A volume of 3-5 µL of sample was loaded into borosilicate capillaries (i.d. 0.78 mm) pulled to a fine tip using a Flaming-Brown P-97 micropipette puller (Sutter Instruments). A platinum wire was placed in electrical contact with the solution, and a voltage of +0.5-0.7 kV was applied to the wire to initiate electrospray. Data were acquired with the source at ambient temperature, Trap collision energy at 10 V, Transfer collision energy at 5 V, and Trap gas flow at 10 mL/min. A sampling cone with a small aperture and backing pressure of 1.37 mbar was used in all experiments, and the sampling cone was operated at 25 V (for accurate mass determination), and 50 V (for complex dilution series). Spectra shown were generated by summing data scans collected over 1 min (accurate mass) or 7 min (complex dilution series). A mass calibration profile was generated using cesium iodide clusters prior to acquiring data for accurate mass determination. Complex stoichiometry of the two largest complexes identified was confirmed via collision-induced dissociation (CID) and detected abundance distributions of QT2-4/LC8 complexes were assessed in the context of Poisson probability distributions[148]. In addition to acquiring data of individual proteins at an initial concentration of 25 µM and of the mixed sample (QT2-4/LC8) at 1 µM, spectra were also acquired for a dilution series of each sample at 10 µM and 1 µM (individual proteins), 500 nM, 100 nM, and 10 nM (individuals and complex). Poisson probabilities of observing 0 to 10 protein molecules as non-specific

oligomers were calculated for each of these concentrations using a droplet diameter of 200 nm. Mass spectral peaks were fitted to Gaussian distributions using IGOR Pro, and the resulting areas of each species' charge state peaks were summed to determine abundances for each different species observed. Oligomers and complexes detected with relative abundances well above those expected for non-specific, Poisson-like association occurring during the nanoESI process were determined instead to originate in solution.

NMR Experiments

NMR experiments were carried out on an 800 MHz Bruker Avance III HD NMR spectrometer equipped with a 5 mm triple resonance (HCN) cryogenic probe. All NMR data on the complex were collected at 40° C, as this led to the best overall spectrum of the complex. NMR samples were prepared in 10 mM sodium phosphate, 10 mM NaCl, 1mM sodium azide, 10 mM Tris(2-carboxyethyl)phosphine, pH 6.5 buffer at a final QT2-4 concentration of 0.5 mM. Samples also contained a protease inhibitor mixture (Roche Applied Science, Madison, WI), 10% D_2O, and 0.2 mM 2–2 dimethylsilapentane-5-sulfonic acid for ^1H chemical shift referencing. Backbone assignments were obtained using a suite of three-dimensional NMR experiments, including HNCA, HNCACB, HNCOCACB, HNCO, and HNCACO. All three-dimensional experiments used TROSY, deuterium decoupling, and non-uniform sampling (NUS). Recycle delays were set to 2.5 s for all experiments except HetNOE, which used a recycle delay of 8 s. For the saturation transfer difference experiments[141], saturation was applied during the relaxation delay of a 2D ^{15}N-TROSY-HSQC for 4 s using a 50 ms gaussian pulse with a B_1 field strength of 50 Hz. On resonance and off resonance saturation was applied at 0.7 ppm and -40 ppm respectively. Subtraction of the on and off resonance spectra is incorporated into the phase cycle. T_1 and T_2 experiments incorporated temperature compensation. Relaxation delay times for the T_1 experiments were 0.02, 0.06, 0.1, 0.2, 0.4, 0.6, 0.8, and 1.2 s. Relaxation delays times for the T_2 experiments were 0.0169, 0.0339, 0.0509, 0.0678, 0.0848, 0.1357, 0.1696, 0.2374 s. Triplicate time points were

collected in T_1 and T_2 experiments for error estimation. NMR data were apodized, zero-filled, Fourier transformed, phased, and baseline corrected using nmrPipe[118]. Data were apodized with a shifted sine squared function and zero filled to twice the original size. Artifacts from NUS data collection were removed using SCRUB[121]. Data were visualized and analyzed in nmrviewJ or CARA[120]. Chemical shift indexing was performed using the ΔCα-ΔCβ method[149,150].

Acknowledgements

The authors thank Wendy Hare for help with protein preparations. This work is funded by the National Science Foundation (Award 1617019). The Oregon State University NMR Facility is funded in part by the National Institutes of Health, HEI Grant 1S10OD018518, and by the M. J. Murdock Charitable Trust grant #2014162. Assignments for QT2-4 bound to LC8 have been deposited in the BMRB under accession number 28033. This work was also supported by a National Institutes of Health grant number T32 GM007759 (to A.D.R.). A.D.R. is an ARCS Scholar supported by the ARCS Oregon Chapter.

Chapter 4

Interplay of linker length and motif specificity in regulating assembly of dynamic LC8 multivalent complexes

Kayla A. Jara, Amber D. Rolland, Andrew K. Swansiger, Patrick N. Reardon, James S. Prell, and Elisar J. Barbar

In preparation for submission to *J. Biol. Chem.*, May 2022

Abstract

The ubiquitous hub protein, LC8, is known to bind over 100 different intrinsically disordered protein (IDP) partners involved in a wide variety of cellular functions. LC8/IDP complexes commonly form as duplex scaffold assemblies and cases continue to emerge in which the IDP ligand contains multiple binding sites for LC8. Although the LC8 binding motif has been well characterized, properties surrounding the context of these multivalent binding sites are not fully understood. Variability in both motif specificity and linker lengths separating motifs are well represented in *Drosophila* ASCIZ, especially within a subdomain termed QT2-4. Serving as a model system, past studies utilizing QT2-4 provided the first evidence of in-register binding during LC8/IDP complex assembly and showcased the role that linker length plays in modulating the flexibility of such complexes. Work presented here expands on these results determining how the interplay of linker length and motif specificity regulate the propensity and compositional heterogeneity of dynamic, multivalent LC8 duplexes. We show that a short linker between sites results in stable LC8/IDP duplex assemblies that are readily formed in solution compared to constructs with longer linkers that have a greater propensity for the formation of single chain complexes. Comparison of constructs with similar linker lengths, but variability in motif specificity emphasize that both properties regulate IDP/LC8 complex assembly.

Introduction

Recognition of the prevalence of intrinsically disordered proteins and protein regions (IDPs/IDRs) and their roles as biologically active proteins has rapidly grown[1]. IDPs and IDRs are characterized by low sequence diversity, a lack of hydrophobic residues, an abundance of charged residues, and areas of sequence repeats. Due in part to their high charge state, as well as their abundance of short linear binding motifs, disordered regions are promiscuous in their binding interactions, and facilitate the formation of many large, complex protein assemblies[8]. IDPs/IDRs are also extremely functionally diverse, and in addition to their structural plasticity and dynamic conformational flexibility, they often interact with their binding partners multivalently.

Compared to monovalent interactions, in which ligands bind a single location of a partner, multivalent interactions involve linked associations of ligands binding to multiple sites[16–18]. Multivalent IDP assemblies are considered to belong to one of the following groups; binary complexes, IDP single chain scaffolds, IDP duplex scaffolds, higher order IDP associations, and IDP multi-site collective binding ligands[17]. Of focus in this work, assemblies termed “duplex scaffolds” are composed of two IDP chains connected by one or more bivalent partners with two symmetrical binding sites and/or by self-association interactions within the chain[16,17]. Cases in which the same dimeric ligand binds multiple sites across disordered chains are incredibly common for partners of the hub protein, LC8 (Figure 4.1)[17,49,131].

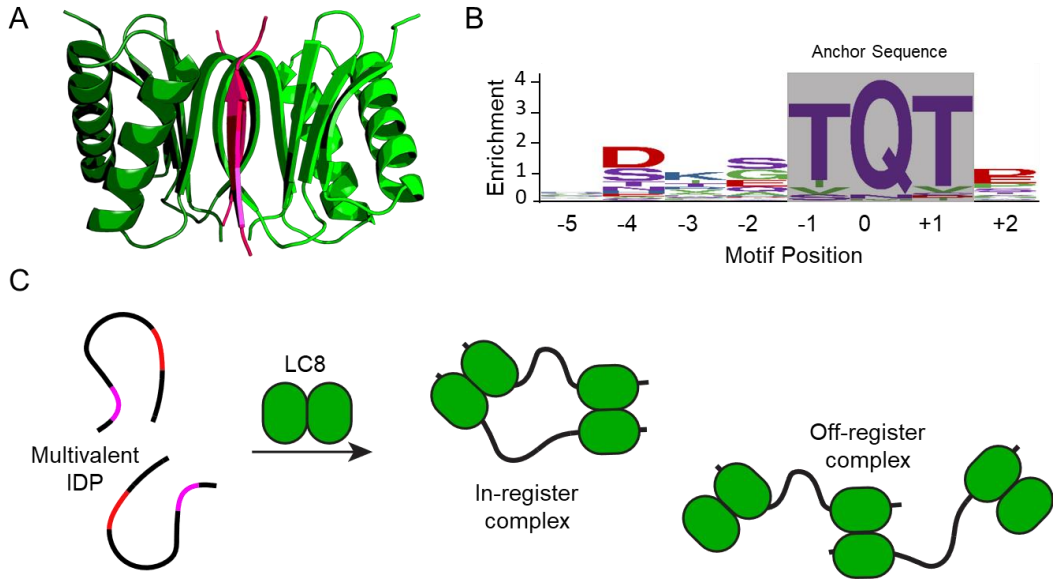

Figure 4.1. LC8 dimer, binding motif, and multivalent assembly. A) Ribbon diagram of the LC8 dimer showing each monomer (light and dark green) bound to disordered peptides (pink and red) that adopt β-strand structure upon binding in LC8's binding groove (Protein Data Bank code 2P2T). B) amino acid enrichment for each position within the LC8 binding motif. The TQT anchor is boxes in gray. C) A model depicting duplex complex assembly of a multivalent IDP and LC8. Both in-register and off-register conformations are shown.

Within the LC8 homodimer, two symmetrical binding grooves are housed that allow for LC8's primary function as a dimerization engine of its intrinsically disordered (or regions thereof) binding partners (Figure 4.1A)[17,18,45]. Out of LC8's over 100 confirmed partners, IDPs preforming functions such as intracellular transport[51,52], nuclear pore formation[53], viral interactions[54–56], and transcription[57–60] are represented. LC8 partner proteins share a short (8 amino acid) linear recognition motif that mediate their binding to LC8[55]. The binding motif contains some level of variation, however is typically anchored by a threonine-glutamine-threonine (TQT) sequence (Figure 4.1B)[43]. Although it is common for LC8 partners to contain multiple LC8 binding motifs, one extremely unique example is the transcription factor, ASCIZ (ATMIN-Substrate Chk-Interacting Zn^{2+} finger)[38,58], that has an astonishing 11 LC8 recognition motifs in its human version[59].

Prior *in vivo* and biophysical studies have characterized ASCIZ functionally as a regulator of the cellular concentrations of its own gene product, LC8[57,61,70]. ASCIZ is thought to act as a sensor that fine-tunes LC8 transcription by interacting with LC8 via a dynamic ensemble of low occupancy bound complexes[59]. *Drosophila* (7 LC8 sites) and human ASCIZ studies, show that ASCIZ/LC8 interactions display both positive and negative cooperativity. Also, unlike Nup159 from yeast, (5 LC8 binding sites) ASCIZ does not form rigid stacked complexes but instead more heterogenous complexes by negative-stain EM analysis[59]. Such heterogeneity may be due to the disordered linkers between LC8 binding sites in ASCIZ that vary considerably in length (3 to 30 residues). Unique to *Drosophila* ASCIZ, one of the LC8 binding motifs contains a TMT rather than the canonical TQT anchor (Figure 4.2). A multivalent domain of *Drosophila* ASCIZ, termed QT2-4 for containing the second, third, and forth LC8 binding sites, serves as a model system of study as this construct contains both the shortest and longest linker lengths between sites and variability in LC8 motifs (Figure 4.2). Our recent studies utilizing QT2-4 provided the first evidence of in-register binding during LC8/IDP complex assembly and suggested that linker length plays a role in modulating the flexibility and LC8 occupancy in multivalent LC8/IDP complexes in general[60]. Work presented here expands on these results with special attention to the interplay of linker length and motif specificity in the regulation of dynamic, multivalent LC8 complexes.

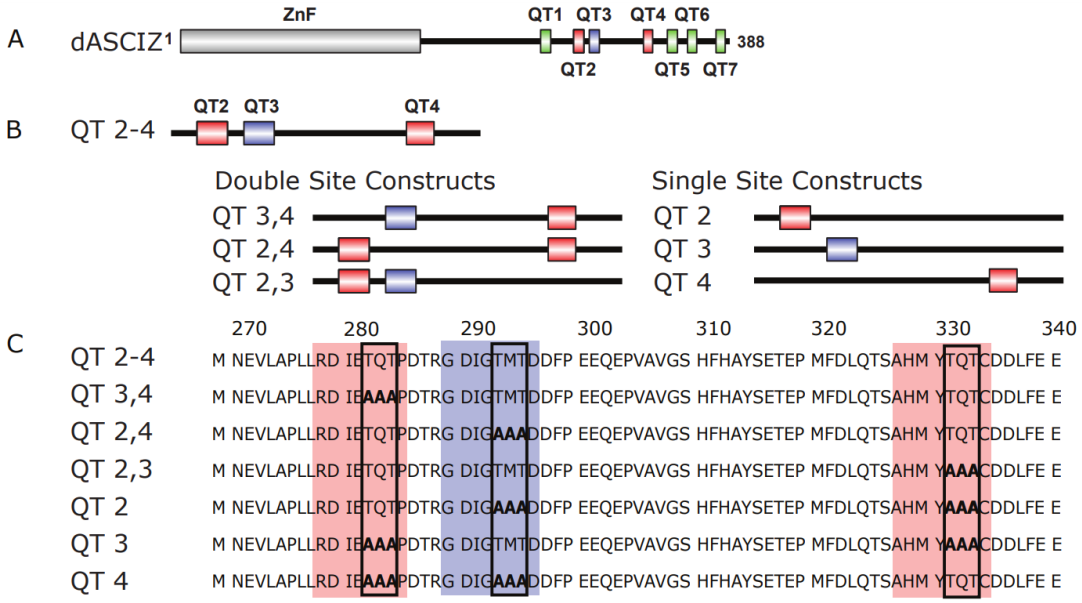

Figure 4.2. ASCIZ domain architecture and constructs. A) Schematic diagram of full-length Drosophila ASCIZ with LC8-binding sites denoted. B) The constructs used in this study including the wild type QT 2-4 (top) and AAA variants that systematically abolish either one (left) or two (right) LC8 recognition motifs from QT 2-4 at a time. Construct nomenclature denotes the binding sites present in each construct. C) Amino acid sequence of each construct is shown with the QT2 and QT4 binding shown in red and QT3 shown in blue. Anchor binding site residues (TQT or TMT) are boxed and are listed as AAA when mutated.

Results

Thermodynamics of QT2-4 Single Site Constructs Interactions with LC8

Previous experiments have documented the binding affinity of QT2-4 and LC8 with a K_d ~ 5 μM (reproduced here in Figure 4.3A) and explored the binding affinities of each LC8 binding motif in QT2-4 individually using short peptide sequences[151]. Here, we created three mutant constructs (QT2, QT3, and QT4) in which two out of three native LC8 recognition motifs in the QT2-4 construct were abolished by replacing the three TQT anchor residues with AAA, so that each binding site could be studied individually, but also in the context of the longer, disordered chain (Figure 4.2B).

ITC experiments of QT2 (Figure 4.3B) and QT4 (Figure 4.3D) with LC8 yield modest dissociation constants (K_d) of 9.3 μM and 15.4 μM respectively. The similarity of these dissociation constants is expected as the QT2 and QT4 LC8 binding sites share the canonical TQT motif anchor, and the slight affinity preference for the QT2 site supports previous results that indicate the site as the first to bind in the context of QT2-4[152]. As expected, the interaction of QT3 with LC8 (Figure 4.3C) yields a much weaker binding affinity (K_d of 35.6 μM) supporting previous data[151]. Interestingly, the ΔH and TΔS values for QT4 (-16.1 and -9.5 kcal/mol) vary from those recorded for QT2 (-10.7 and -3.9 kcal/mol) and QT3 (-8.4 and -2.3 kcal/mol). Especially when compared to QT2, the ΔH and TΔS variability seen for QT4 suggests that the composition of the motif outside of the TQT anchor and/or the distance the center of the anchor lies from the closest terminus, 9 versus 15 residues for QT4 and QT2, respectively, impact the thermodynamics of LC8 binding. In all cases, the ΔG values are between -6 and -7 kcal/mol (Table 4.1).

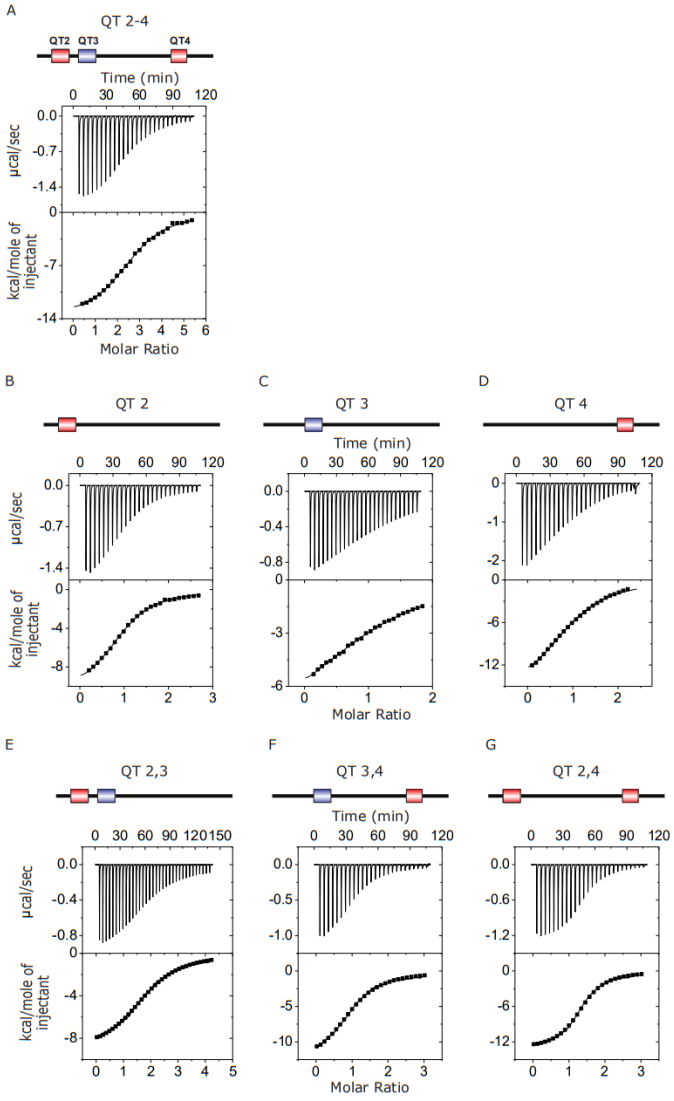

Figure 4.3. LC8-ASCIZ interactions monitored by ITC. A-G) Representative thermograms of the titration of LC8 into QT constructs corresponding to WT QT2-4 (A), the single-site constructs QT2 (B), QT3 (C), and QT4 (D), and the double site constructs QT2,3 (E), QT3,4 (F), and QT2,4 (G). Data were fit to a single site bonding model using Origin software.

Table 4.1. Thermodynamic parameters of the QT-LC8 interactions.

Construct	N	Overall Kd (μM)	Overall ΔH (kcal/mol)	Overall TΔS (kcal/mol)	Overall ΔG (kcal/mol)
QT 2-4	2.7	4.8 ± 0.2	-13.9 ± 0.7	-6.6 ± 0.3	-7.3 ± 0.4
QT 2	1	9.3 ± 0.5	-10.7 ± 0.5	-3.9 ± 0.2	-6.9 ± 0.3
QT 3	1	35.6 ± 2	-8.4 ± 0.4	-2.3 ± 0.1	-6.1 ± 0.3
QT 4	1	15.4 ± 0.8	-16.1 ± 0.8	-9.5 ± 0.5	-6.6 ± 0.3
QT 2,3	1.9	6.3 ± 0.3	-9.2 ± 0.5	-2.1 ± 0.1	-7.1 ± 0.4
QT 3,4	1	5.0 ± 0.3	-13.2 ± 0.7	-6.0 ± 0.3	-7.2 ± 0.4
QT 2,4	1.4	1.8 ± 0.1	-13.2 ± 0.7	-5.4 ± 0.3	-7.8 ± 0.4

Thermodynamics of QT2-4 Double Site Constructs Interactions with LC8

ITC experiments using the double site QT2-4 constructs (QT2,3, QT3,4, and QT2,4) (Figure 4.2B) were applied to identify how abolishing each individual LC8 binding motif, one at a time, impacts binding of the overall construct (Figure 4.3E-G). Since each construct has at least two available LC8 binding motifs, and all display a single binding step, the reported K_d and stoichiometry values are effective values representing overall binding of assembly formation (Table 4.1).

When abolishing any of the three LC8 recognition motifs the binding stoichiometry expectedly decreases compared to that of the wild-type QT2-4 construct, however differences between the mutants were observed (Figure 4.3, Table 4.1). When abolishing the QT4 site (QT2,3) the N value drops to 1.9, indicating that binding is still occurring at the QT2 and QT3 sites (Figure 4.3E). Interestingly, when abolishing either QT2 or QT3 (QT3,4 and QT2,4) the N value decreases even further. In the case of QT3,4, the overall stoichiometry decreases to 1.0, suggesting that only a single LC8 motif is binding (Figure 4.3F). This aligns well with previous data that indicated QT3 binding to be weak and cooperative with QT2[151,152]. It is likely that without the binding of QT2, QT3 is simply unable to bind under these conditions. QT2,4/LC8 binding does not support that the reverse is true however, as an overall stoichiometry of 1.4 was seen, indicating that on average between 1 and 2 LC8s are binding per QT2,4 monomer (Figure 4.3G).

Overall binding affinities of the double site constructs are all stronger than those measured for the single sites, indicating that each combination of double sites is binding cooperatively (Table 4.1). However, dissociation constants do vary between the constructs, and most notably, for the QT2,4/LC8 interaction. It appears that by removing the weakest binding site, the construct binds with the greatest affinity (K_d = 1.8 μM) (Fig 4.3G, Table 4.1).

Complex formation monitored by sedimentation velocity analytical ultracentrifugation (SV-AUC)

To further identify the possible heterogeneity of formed complexes between the QT2-4 double site constructs and LC8, we used SV-AUC to track QT/LC8 complex assembly. SV-AUC analysis of the double site constructs in complex with LC8 show that the proteins are in dynamic equilibrium at molar ratios of QT:LC8 up to 1:3 (Figure 4.4) and that the final complexes formed at saturating LC8 conditions vary in their sedimentation coefficients. Complexes formed with QT3,4 have sedimentation coefficients of ~2.5, ~3.75, and ~4 S at ratios of 1:1, 1:2, and 1:3 respectively (Figure 4.4A). Complexes formed with QT2,4 have sedimentation coefficients of ~2.75, ~4, and ~4 S across the increasing ratios of LC8 (Figure 4.4B). Finally, complexes formed with QT2,3 have sedimentation coefficients of ~3.25, ~3.5, and ~3.75 S (Figure 4.4C).

The sizeable shift in sedimentation coefficient for QT3,4 complexes between the ratios of 1:1 and 1:2 is indicative of the difference between binding 1 versus 2 LC8 dimers, while the much smaller shift between the ratios of 1:2 and 1:3, coupled with the large excess LC8 peak (~2.5 S), likely correspond to shifting the equilibrium such that a fully bound complex with 2 LC8 dimers is favored (Figure 4.4A). Complex formation appears accelerated for QT2,3 as at the lowest ratio of LC8, the largest sedimentation coefficient is seen. Furthermore, even though the complex peak for QT2,3/LC8 at the 1:2 ratio has a slightly smaller sedimentation coefficient than those seen for QT3,4 and QT2,4, the peak is more well defined from the free LC8 peak (Figure 4.4C). Together, these results indicate that QT2,3 and LC8 form less heterogeneous complexes that are in less of a dynamic equilibrium. Conversely, the

broadest peak with the worst definition between complex and LC8 peaks is seen for QT2,4 at the 1:2 ratio, indicating more complex heterogeneity (Figure 4.4B). These results compare well to the difference in linker lengths in each construct, with QT2,3 having a very short linker between the available binding sites and QT2,4 having the longest linker between binding sites. We did not detect larger order aggregates or oligomers in any of the experiments for any construct.

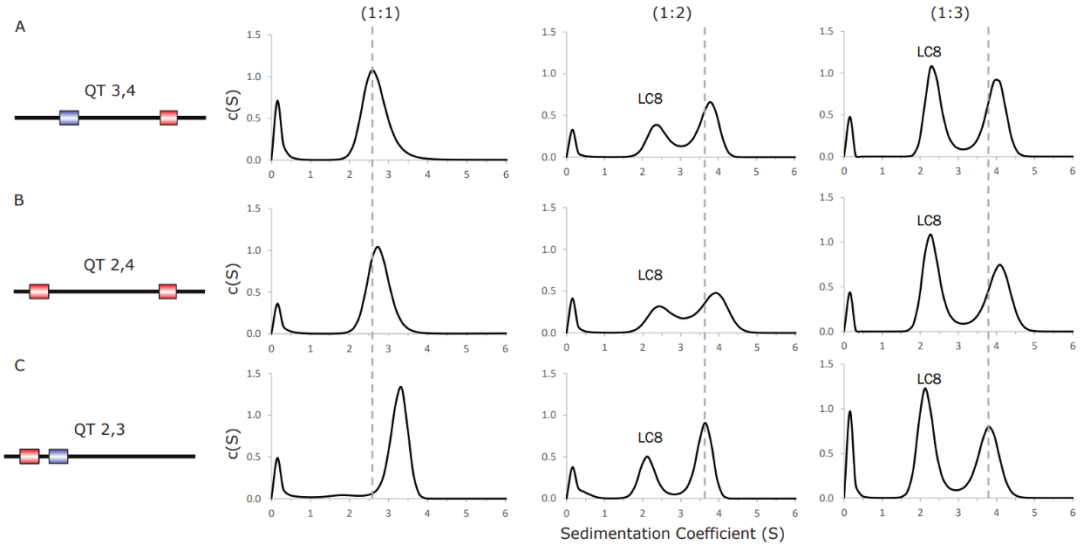

Figure 4.4. Sedimentation velocity analytical ultracentrifugation of double site ASCIZ constructs bound to LC8. SV-AUC titrations of QT3,4 (A), QT2,4 (B), and QT2,3 (C) with LC8 at three separate molar ratios of QT:LC8 (1:1, 1:2, 1:3). When applicable, populations corresponding to free LC8 are labeled. The dashed line is centered on the complex peak with the lowest sedimentation coefficient for each molar ratio.

Complex formation monitored by native ESI-MS

Using native electrospray ionization (ESI)-MS, measurements of the individual protein subunits along with the masses of coordinating QT/LC8 complexes allows for identification of complex stoichiometries for those formed by the double site QT constructs. These experiments are very similar to those used previously to study the complex formation of QT2-4 with LC8[152]. As was observed previously, accurate mass determination for LC8 and the QT constructs match closely with the

expected masses calculated from each protein sequence (Table 4.2) and upon conducting a dilution series, QT3,4, QT2,4, and QT2,3 remain monomeric while LC8 is dimeric in solution.

Table 4.2. Expected and measured masses for all proteins and complexes identified by native ESI-MS

Protein	**Charge states**	**Sequence mass (Da)**	**Measured mass (average ± s.d., Da)**
LC8	6-10+	10639.11	10638.75 ± 0.04
QT2,3	3-7+	8475.26	8452.77 ± 0.07
QT2,4	4-7+	8472.2	8510 ± 1
QT3,4	3-7+	8475.26	8452.47 ± 0.27
Species/complex	**Charge states**	**Expected mass (Da)**	**Measured mass (average ± s.d., Da)**
LC8 monomer	6-9+	10638.75 ± 0.04	10638.82 ± 0.04
LC8 dimer	8-11+	21277.51 ± 0.07	21278.3 ± 0.5
LC8 tetramer	11-14+	42555.01 ± 0.15	42558 ± 2
QT3,4 monomer	5-7+	8452.5 ± 0.3	8452.4 ± 0.5
QT3,4 monomer + LC8 dimer	9-13+	29729.98 ± 0.28	29732.52 ± 0.21
QT3,4 monomer + 2 LC8 dimers	13-16+	51007.49 ± 0.31	51013 ± 7
QT3,4 dimer + 2 LC8 dimers	13-20+	59460.0 ± 0.6	59464 ± 16
QT3,4 dimer + 3 LC8 dimers	16-20+	80737.5 ± 0.6	80766 ± 20
QT2,4 monomer	5-7+	8510 ± 1	8508 ± 6
QT2,4 monomer + LC8 dimer	9-13+	29788 ± 1	29788.1 ± 0.2
QT2,4 monomer + 2 LC8 dimers	15-18+	51065 ± 1	51073 ± 4
QT2,4 dimer + 2 LC8 dimers	14-21+	59576 ± 2	59576 ± 21
QT2,4 dimer + 3 LC8 dimers	16-20+	80853 ± 2	80854 ± 11
QT2,3 monomer	5-7+	8452.77 ± 0.07	8452.63 ± 0.3
QT2,3 monomer + LC8 dimer	10-12+	29730.28 ± 0.10	29732.01 ± 0.33
QT2,3 monomer + 2 LC8 dimers	12-16+	51007.78 ± 0.16	51036 ± 22
QT2,3 dimer + 2 LC8 dimers	13-20+	59460.56 ± 0.20	59510 ± 31
QT2,3 dimer + 3 LC8 dimers	15-20+	80738.06 ± 0.26	80804 ± 54

Further native mass spectra acquired for samples of LC8 mixed in a 2:1 molar ratio with each QT2-4 double site construct identified the same complex stoichiometries with LC8 (Table 4.2, Figure 4.5). The four detected complex stoichiometries correspond to one LC8 dimer bound to a single QT subunit (1:2 QT:LC8), two LC8 dimers bound to a single QT subunit (1:4), two LC8 dimers bound to two QT subunits (2:4), and three LC8 dimers bound to two QT subunits

(2:6). These results mimic those determined for wildtype QT2-4/LC8 complexes, as both the expected “in-register” complex (2:4) and an “off-register” complex (2:6) are present. Of note, in-register complexes are always detected at a greater abundance than the off-register complex, for all QT double site constructs, and the persistence of the in- and off-register complexes at low concentrations indicates they are both native, rather than spurious, complexes (Figure 4.5).

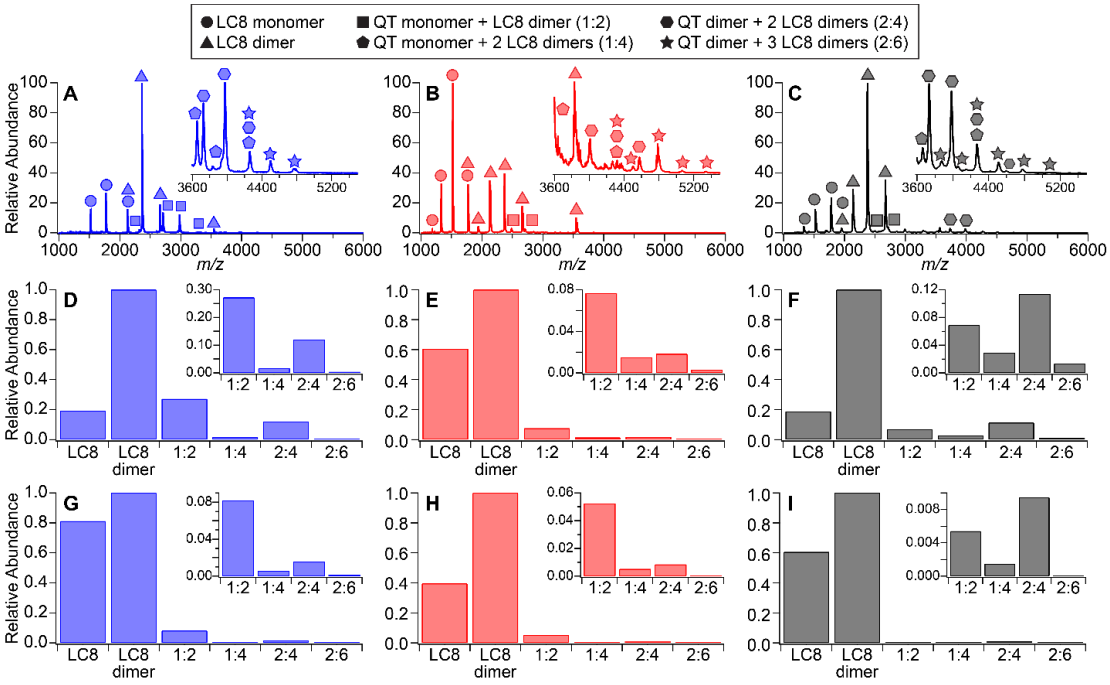

Figure 4.5. QT/LC8 complex species and abundance distributions determined by ESI-MS. Native mass spectra of A) QT3,4 (blue), B) QT2,4 (red), and C) QT2,3 (gray) at 25 µM are shown with individual and complex species labeled. The abundancy distributions of species detected at 25 µM (D-F) and 5 µM (G-I) are shown.

While QT3,4, QT2,4, and QT2,3 all form the same complex stoichiometries with LC8, there are slight differences with respect to their relative detected abundances and abundances upon dilution (Figure 4.5D-I). Considering only the four hetero-oligomeric complexes detected, the 2:4 in-register complex is the most abundant species formed by QT2,3 with LC8 at nearly all concentrations studied (Figure 4.5F and I). By contrast, for both QT2,4 and QT3,4, the intermediate 1:2

complex is the most abundant species detected across all total protein concentrations tested (Figure 4.5D-E and G-H). In general, the off-register 2:6 complex is detected at a low abundance for all three double site constructs. Together, these results suggest that the complexes formed by QT2,3 are more stable than those formed by QT2,4 or QT3,4. This is in alignment with QT2,3 containing the shortest linker between binding sites and echoes the ITC and SV-AUC results presented above.

The QT/LC8 species identified with native ESI-MS also provide evidence for the general mechanism of complex formation. First, an LC8 dimer binds to a single QT subunit, forming the 1:2 complex. This is followed by binding a second LC8 dimer, resulting in a 1:4 complex. Finally, a second QT strand is recruited to form the expected in-register 2:4 complex. Misalignment of the second QT strand would allow a third LC8 dimer to bind, resulting in an off-register 2:6 complex. These data would also support a model whereby the second QT strand is already bound by an LC8 dimer before binding to the intermediate complex in an off-set manner. It is also possible that some proportion of the 2:4 complex population could also represent the binding of two 1:2 complexes, rather than being formed by sequential binding of LC8 dimers to a single QT subunit. This mechanism of assembly and the ensemble of complexes formed by the QT double site constructs are depicted in Figure 4.6.

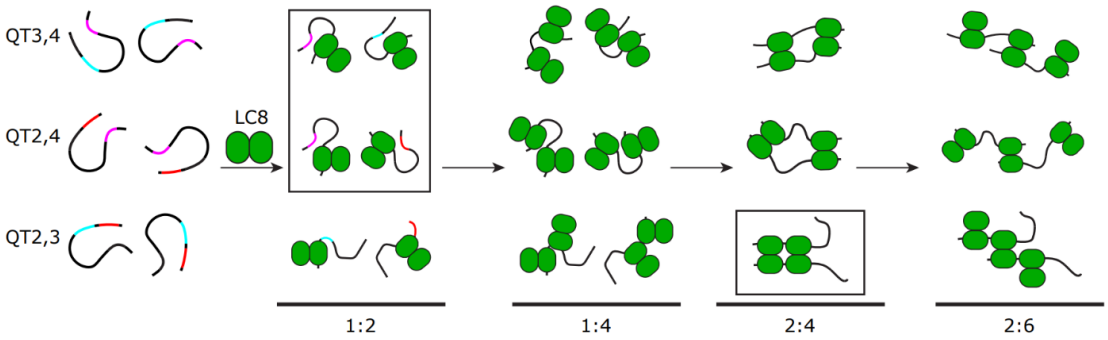

Figure 4.6. Ensemble models of complex species detected for LC8/QT double site constructs. Two monomeric chains of each QT double site construct are shown (far left) with QT2 (red), QT3 (blue), and QT4 (pink) sites color coded. Upon addition of LC8 (green) complexes with 1:2, 1:4, 2:4, and 2:6 stoichiometry form. Models for each construct at each complex stoichiometry are shown, with the most abundant complex species for each QT construct boxed.

Discussion

LC8 commonly forms duplex scaffold assemblies with its many multivalent IDP partners[17,49,131] and cases in which the IDP ligand contains multiple binding sites for LC8 continue to emerge. However, the properties surrounding the context of multivalent LC8 binding sites are not fully understood. Variability in both motif specificity and linker lengths between motifs are well represented in *Drosophila* ASCIZ, especially within the QT2-4 subdomain. Serving as a model system of study, this construct contains both the shortest and longest linker lengths between sites as well as TQT and TMT anchored LC8 motifs. Recent studies utilizing QT2-4 provided the first evidence of in-register binding during LC8/IDP complex assembly and showcased the role that linker length plays in modulating the flexibility of such complexes[60]. Work presented here expands on these results determining how the interplay of linker length and motif specificity regulate the propensity and compositional heterogeneity of dynamic, multivalent LC8 duplexes.

Two LC8 binding sites are better than one and the cooperativity of a 3rd site is location and motif dependent

ITC experiments of single site QT constructs provide motif specific binding affinities in the context of the QT2-4 disordered chain for QT2, QT3, and QT4 (Figure 4.3B-D). Expectedly, QT2 and QT4 show similar solo binding affinities (9.3 and 15.4 μM, respectively) while QT3 is considerably weaker (35.6 μM). This is consistent with the TMT anchor of QT3 compared to the TQT anchors for QT2 and QT4, and the slight favorability for QT2 supports prior evidence of it as the initial sight of LC8 binding within QT2-4. Upon reintroduction of a binding site into each single site construct, forming the double site constructs, ITC experiments indicate variability in LC8 binding cooperation (Figure 4.3E-G). When a third binding site containing a TQT motif is introduced in a flanking position, outside the pair of double-sites, affinity is increased, and positive cooperativity occurs. This positive cooperativity is seen in the binding thermograms and dissociation constants for QT2-4 compared to those of QT3,4 and QT2,3 (Figure 4.3). In contrast, when introducing

a third binding site containing a TMT motif that interrupts two existing sites, as in the difference between the QT2,4 and QT2-4 constructs, negative cooperativity occurs. In the QT2-4 system we are unable to distinguish if location or motif specificity plays a larger role in imparting negative cooperativity in a triple site, multivalent IDP compared to a double site. However, because of the sizeable decrease in affinity shown here, we conclude that it's likely that both properties play a role.

QT2,3 forms stable complexes with LC8 more readily than QT3,4 and QT2,4

Although the double site QT constructs have the same occupancy for two LC8 dimers, it's clear that the constructs vary in their complex assembly. ITC experiments report stoichiometries that vary between 1 and 2, indicating a discrepancy in the number of LC8 dimers binding each construct. With stoichiometries below the expected 2, both QT3,4 and QT2,4 display a decreased ability for complex assembly with LC8 (Figure 4.3, Table 4.1). We attribute this to the increased linker length of 30 and 41 residues, respectively, compared to the very short linker in QT2,3. These results are supported by ESI-MA experiments that detected much higher abundance of the expected 2:4 complex (2 QT monomers, 2 LC8 dimers) for QT2,3 compared to QT3,4 and QT2,4 that remains upon complex dilution (Figure 4.5).

SV-AUC titration experiments tracking QT/LC8 complex formation (Figure 4.4) also show that QT2,3 forms a complex with LC8 more readily compared to QT3,4 and QT2,4, evident by a larger sedimentation coefficient for the complex at a sub-saturating molar ratio of LC8. Furthermore, at a 1:2 molar ratio of LC8, QT2,3 forms a complex with LC8 that is more homogeneous than those for QT3,4 and QT2,4, evident by better definition between the QT2,3/LC8 complex and free LC8 peaks that is lacking for the other constructs (Figure 4.4). The sedimentation coefficient for the QT2,3/LC8 complex doesn't increase between a 1:2 and 1:3 molar ratio of LC8, indicating that QT2,3 had already formed a fully bound complex at the lower ratio. This contrasts with QT3,4 and QT2,4 complexes that increase in sedimentation coefficient, suggesting that both require over-saturating amounts of LC8 to shift the equilibrium towards a fully bound complex. Interestingly, the

complexes formed at over-saturating amounts of LC8 have larger sedimentation coefficients than that of the QT2,3/LC8 complex. This can be explained by either a difference in mass or overall shape of the fully bound complexes, as sedimentation coefficient is determined by both. Lower N values for QT3,4 and QT2,4 ITC binding experiments make it unlikely that a mass difference exists and is further supported by complex species determined for the double site constructs by native ESI-MS (Tables 4.1-4.2). It is therefore likely that the extended, disordered tail following the QT2,3 LC8 binding sites causes a more extended duplex confirmation thus increasing the frictional coefficient of the QT2,3/LC8 complex resulting in a smaller sedimentation coefficient.

Linker length and motif specificity determine the propensity for duplex formation

Although the double site QT constructs all form the same complex species in solution determined by ESI-MS; one LC8 dimer bound to a single QT subunit (1:2 QT:LC8), two LC8 dimers bound to a single QT subunit (1:4), two LC8 dimers bound to two QT subunits (2:4), and three LC8 dimers bound to two QT subunits (2:6), they vary drastically in their detected abundance (Table 4.2, Figure 4.5). Single chain complexes (1:2 and 1:4) are more abundant for long linker containing constructs, QT3,4 and QT2,4, than the short linker containing QT2,3. These single chain complexes are also more abundant than the duplex species, even at the highest concentration tested (Figure 4.5). This indicates that an increased linker length between LC8 binding sites disrupts duplex formation of IDP multivalent complexes. Further, the higher propensity for single chain complexes explains decreased N values for QT3,4 and QT2,4 binding interactions with LC8 by ITC and the increased complex compositional heterogeneity seen in SV-AUC experiments (Figure 4.3-4.4). Notably, motif specificity remains important to complex formation the weak-binding LC8 motif in QT3,4 causes a lower overall LC8 binding stoichiometry by ITC compared to QT2,4 with a similar length linker (Figure 4.2C, Table 4.1). Together, these results highlight the importance of both linker length and motif specificity and

determine their interplay as a regulation mechanism for IDP/LC8 multivalent complex assembly.

Conclusions

Here we show that binding of LC8 to multivalent QT2-4 constructs is complex and governed by factors such as LC8 motif specificity and the length of disordered linkers between LC8 binding sites. Cooperativity between multivalent sites can be positive or negative depending on site location and individual LC8 binding affinity. QT constructs with only two binding sites are all cooperative and upon addition of a flanking third site with strong individual affinity remain so. However, interruption of two binding sites by a third, weaker site results in negative cooperativity instead. Additionally, the multivalent construct with the shortest linker between sites resulted in stable LC8/IDP duplex assemblies that are readily formed in solution compared to constructs with longer linkers that showed a greater propensity for the formation of single chain complexes. Comparison of constructs with similar linker lengths, but variability in motif specificity emphasize that both properties regulate IDP/LC8 complex assembly.

Experimental Procedures

Cloning, Protein Expression, and Purification

Cloning of *Drosophila* ASCIZ QT2–4 (residues 271–341) with various mutations of recognition motifs was performed using either the QuikChange Lightening mutagenesis kit (Agilent) or the Q5 site-directed mutagenesis kit (New England Biolabs). The resulting constructs verified by sequencing are QT2, QT3, QT4, QT2,3, QT2,4, and QT3,4, where the number(s) following ‘QT’ indicate which LC8 recognition motif(s) remain and have not been mutated to AAA. Proteins were expressed and purified according to previously published protocols[151,152].

Isothermal Titration Calorimetry

Binding thermodynamics of the QT/LC8 interactions were obtained with a MicroCal VP-ITC microcalorimeter (Malvern Panalytical, Malvern, United Kingdom). All experiments were obtained at 25˚C and with protein samples in a buffer composed of 50 mM sodium phosphate, 50 mM sodium chloride, 1 mM sodium azide, 5 mM β-mercaptoethanol, pH 7.5. Each experiment was started with a 2 μL injection, followed by 27 to 33 injections of 10 μL. Experiments were conducted with QT in the sample cell at concentrations ranging from 20-50 μM and LC8 in the syringe at concentrations ranging from 400-500 μM. Experiments for the single site constructs resulted in calculated Brandt parameter values (c values) of 5.4, 1.4, and 3.2 for QT2, QT3, and QT4, respectively, indicating that the thermodynamic parameters for each interaction are almost out of an acceptable range for reliability. The data were processed using Origin 7.0 and fit to a simple, single-site binding model; however, these systems are indeed more complicated because LC8 is a dimer binding two IDP chains. Reported data are from two independent experiments. In all cases, the data was reproducible, and the model fits were very good, therefore the reported error is based on a 5% uncertainty in protein concentrations that were determined by absorbance measurement at 280 nm.

Analytical Ultracentrifugation

SV-AUC was performed using a Beckman Coulter Optima XL-A analytical ultracentrifuge, equipped with absorbance optics. LC8 was mixed with each double site QT construct at ratios of 1:1, 1:2, and 1:3 (molar ratio of QT:LC8). Buffer conditions for SV-AUC analysis were 20 mM Tris-HCl, 50 mM NaCl, 5 mM Tris(2-carboxyethyl)phosphine, 1 mM sodium azide, pH 7.5. The complexes were loaded into standard, 12-mm pathlength, two-channel sectored centerpieces and centrifuged at 42,000 rpm and 20 °C. 300 scans were acquired at 280 nm with no interscan delay. The data were fit to a $c(S)$ distribution using the software SEDFIT(33). Buffer density was calculated to be 1.0009 g/ml using Sednterp(34).

Native Electrospray Ionization Mass Spectrometry (Native ESI-MS)

Purified individual protein samples were first diluted to a concentration of 25 µM with 200 mM ammonium acetate, pH 7.4, and a 50 µL aliquot of each was exchanged into this same buffer using Micro Bio-Spin™ 6 columns (Bio-Rad Laboratories, Inc.). A small volume of sample (~3-5 µL) was loaded into a borosilicate glass capillary (i.d. 0.78 mm) pulled to a fine tip (~1-2 µm in diameter) with a Flaming-Brown P-97 micropipette puller (Sutter Instrument), and a platinum wire was placed in electrical contact with the sample solution. Electrospray was initiated by applying a voltage of +1.0 kV to this wire, after which this voltage was lowered to a threshold value, typically +0.6-0.8 kV, to maintain electrospray. All native mass spectra were acquired on a Waters Synapt G2-S*i* time-of-flight mass spectrometer equipped with a nano-electrospray ionization (nano-ESI) source. Prior to native mass spectrometric analysis of these protein samples, a mass calibration profile was created using cesium iodide cluster ions produced by nano-ESI from aqueous cesium iodide solution.

The instrumental settings used were as follows: source at ambient temperature, sample cone collision energy of 25 V, trap collision energy of 25 V, transfer collision energy of 5 V, and trap gas flow rate of 7-7.5 mL/min. Spectra shown represent the summation of data scans acquired over a period of 5 minutes. A native mass spectrum was acquired for each 25 µM buffer-exchanged individual protein sample and used to determine accurate monomer masses. Complexes were formed by mixing LC8 with each QT2-4 mutant at a 2:1 LC8:QT molar ratio to achieve a final total protein concentration of 25 µM. After allowing complex formation to occur overnight at +4°C, native mass spectra were acquired for each 25 µM sample, as well as for a dilution series of each at total protein concentrations of 15 µM, 10 µM, 5 µM, 1 µM, and 500 nM.

After peaks in the native mass spectra of LC8/QT complexes were assigned, the area of each peak was determined by integration with IGOR Pro 9. Detected abundances were calculated by summing the area of each species/complex's charge

state peaks, and these were normalized to the detected abundance of LC8 dimer in each spectrum.

Acknowledgements

The authors acknowledge Delaney Smith and Seth Pickney for their work in cloning the mutant constructs. This work is funded by the National Institutes of Health (R01GM141733) and in part by the University of Oregon Doctoral Dissertation Fellowship and the Peter O'Day Fellowship in Biological Sciences and Office of the Vice President for Research and Innovation at the University of Oregon received by A.D.R. A.D.R. has also been supported by the National Institute of General Medical Sciences (Award 2T32GM00759) and by the ARCS Foundation Oregon Chapter.

Chapter 5

Multivalency, autoinhibition, and protein disorder in the regulation of interactions of dynein intermediate chain with dynactin and the nuclear distribution protein.

Kayla A. Jara, Nikolaus M. Loening, Patrick N. Reardon, Zhen Yu, Prajna Woonnimani, Coban Brooks, Cat H. Vesely, and Elisar J. Barbar.

Submitted to *eLife,* May 2022

Abstract

Cytoplasmic dynein plays crucial roles in the intracellular transport of organelles and other cargoes. Central to dynein function is the intrinsically disordered N-terminal domain of dynein intermediate chain (IC), which binds the three dimeric dynein light chains at multivalent sites, and dynactin p150^{Glued} and nuclear distribution protein (NudE) at overlapping sites. The disorder in IC has hindered cryo-electron microscopy and X-ray crystallography studies of its structure and interactions. Here we use a suite of biophysical methods to reveal how multivalent binding of the three light chains regulate IC interactions with p150^{Glued} and NudE. Using the N-terminal domain or the full-length IC from *Chaetomium thermophilum*, a tractable species to interrogate IC interactions, we identify a significant reduction in IC's binding affinity for p150^{Glued} and a loss of binding to NudE in contrast to the tight binding observed with small IC constructs. We attribute this difference to autoinhibition caused by strong long-range intramolecular interactions that cover IC's N-terminal single α-helix, the site for p150^{Glued} and NudE binding. Reconstitution of IC subcomplexes demonstrate that autoinhibition is differentially regulated by light chains binding underscoring their importance both in assembly and organization of IC, and in selection between multiple binding partners at the same site.

Introduction

Dynein intermediate chain (IC) plays key roles in modulating dynein interactions and activity[51,71,85,90,105]. For example, IC connects the three dynein light chains (Tctex, LC8, and LC7) to the heavy chain, and also serves as the primary binding site of multiple non-dynein proteins essential for dynein regulation, such as the $p150^{Glued}$ subunit of dynactin. Despite this level of importance, high resolution structural information of IC interactions is limited due to its highly disordered N-terminal domain, which hinders studies by methods such as X-ray crystallography and cryo-electron microscopy (cryo-EM).

The primarily disordered N-terminus of IC (N-IC)[51] contains a single α-helix (SAH) and a short helix (H2), each of which is followed by disordered linker regions[85], whereas the C-terminus of IC (C-IC) folds into a β-propeller and contains the binding site for the dynein heavy chain[73,82] (Figure 5.1A). Apo-IC is monomeric but, upon binding the homodimeric dynein light chain subunits (Tctex, LC8, and LC7), it dimerizes to form a subcomplex that is best described as a polybivalent scaffold[16,77–80,105]. In the formation of this subcomplex, the first binding event pays the entropic cost for subsequent bivalent binding events[16,52,81] and the enhancement of subsequent binding events is modulated by the length of the disordered linkers separating the binding sites[81,152]. Such a mechanism has been well-described for the IC/Tctex/LC8 complex, for which a three-residue linker separates the Tctex and LC8 binding sites. In this complex, binding of one light chain to IC results in a 50-fold enhancement in the affinity for binding the other light chain[81]. Each homodimeric light chain has a corresponding binding site on IC that is initially disordered but forms β-strands (for Tctex and LC8) or an α-helix (for LC7) when bound and incorporated into the fold of their respective ligand. The assembly of monomeric IC and the homodimeric light chains is such that, when bound, folding occurs only at the protein-protein interfaces[77,81] while the remaining linker regions stay completely disordered[52,87,91]. In this work, we demonstrate the importance of the flexibility of

these disordered linker regions separating the three dimeric light chains in regulating interactions of IC with non-dynein binding partners.

N-IC binds non-dynein regulatory proteins including dynactin and nuclear distribution protein (NudE)[65,85,89,90,92,97,111]. Dynactin is a multisubunit complex that binds dynein with its largest subunit, p150^{Glued}, and this interaction is required for the recruitment of cargo, dynein processivity, and correct spindle formation in cell division[74,88,89,94,95,153,154]. NudE, on the other hand, regulates dynein recruitment to kinetochores and membranes, centrosome migration, mitotic spindle orientation, and binds LIS1[95,98–100,155–158]. It is well recognized that IC partnering with either p150^{Glued} or NudE impacts the regulation of the dynein complex as this dictates its interactions with adaptors and cargo[71]. However, the molecular processes underlying which regulator is bound at any given time are still unclear. Previously, we showed that in multiple species (rat, Drosophila, yeast) the coiled-coil domains of p150^{Glued} and NudE from the respective species each bind IC at the same site (the SAH region)[65,90,92] but that in a filamentous fungus *(Chaetomium thermophilum)* binding of IC and p150^{Glued} also involves binding of the H2 region[97].

With advancements in cryo-EM, the overall picture of dynein structure and activity is beginning to emerge. For example, cryo-EM images show that dynactin causes the motor domains of dynein to reorient to become parallel to microtubules prior to binding[82,84]. Additionally, adaptors can recruit a second dynein to dynactin for faster movement[73]. Dynein is seemingly a perfect candidate for structural characterization by cryo-EM, as the motor domains are large and symmetric. However, a recurring theme in these studies is that the flexibility of N-IC limits details of its structure and binding interactions with multiple regulators. Although residue specific studies have been performed on IC, they are thus far limited to short fragments and to conditions far removed from native biological systems. Using the combined data from EM structures and *in vitro* studies, we have made a model to aid in visualizing the assembled N-IC subcomplex bound to either p150^{Glued} or NudE thus providing context for the work presented here (Figure 5.1B).

The interactions of short fragments of N-IC with dynein light chains and non-dynein proteins from multiple different species (rat, Drosophila, yeast) demonstrate that NudE and p150^{Glued} compete for the same binding site, however a mechanism for IC partner selection and the importance of bivalency in IC subcomplex assembly have remained elusive [65,77,81,85,90,92,93,99,100]. We recently introduced *Chaetomium thermophilum* (CT), a thermophilic filamentous fungus, as a new system for dynein studies[97]. The CT IC$_{1-260}$ (residues 1-260) construct used in this work is far longer than our previously studied CT, Drosophila, and rat constructs (Figure 5.1, A2.1). Utilizing the entire N-terminal domain of IC allows, for the first time, characterization of IC interactions in the context of its assembly with the light chains, as has been shown vital for other disorder-driven systems[17,59,104,159,160]. However, none of the previously studied systems have the level of complexity (an assembly of five unique proteins) as the system explored here. From our studies of both IC$_{1-260}$ as well as full-length IC, we 1) describe the first recombinant expression and reconstitution of the polybivalent scaffold formed from IC$_{1-260}$ bound by all three light chains (Tctex, LC8, and LC7), as well as by coiled-coil domains of p150^{Glued} or NudE (Figure 5.1), 2) we identify long range tertiary contacts between residues in the C-terminal region (the LC7 binding site) and residues in the N-terminal region (SAH) that inhibit binding to non-dynein proteins 3) we show how assembly with the light chains relieves this autoinhibition and regulates binding of IC to p150^{Glued} and NudE and 4) we demonstrate for the first time the essential role of light chains in both assembly and regulation of the full-length IC.

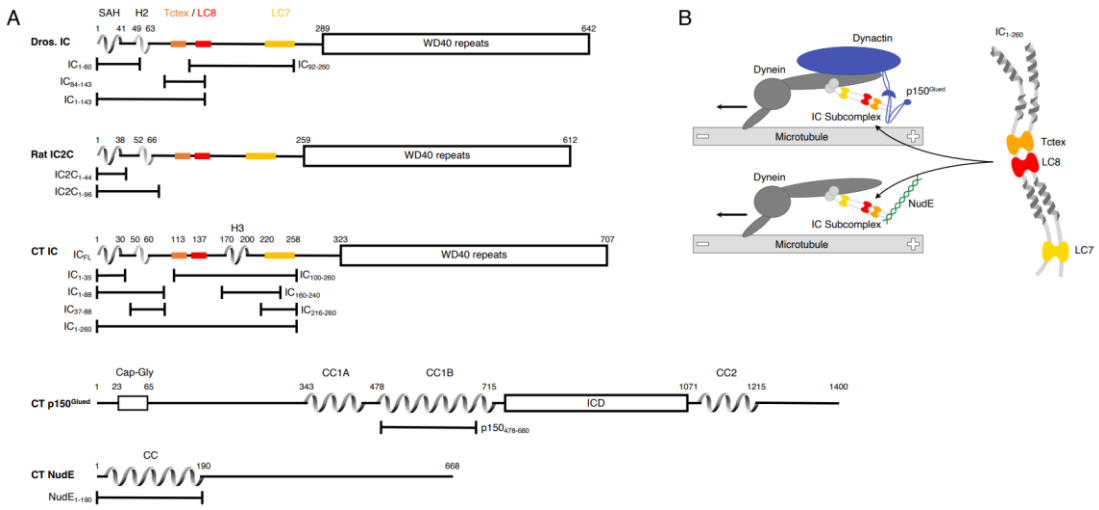

Figure 5.1. Domain architecture for dynein intermediate chain (IC), dynactin p150^{Glued}, and nuclear distribution protein (NudE). A) Domain architecture diagrams for IC from *Drosophila melanogaster* (Dros. IC) and *Rattus norvegicus* (Rat IC2C) and constructs used in earlier work are provided for comparison. Proteins and constructs used in this work are from *Chaetomium thermophilum* (CT). All IC's have an N-terminal single α-helix (SAH), followed by either a transient/nascent or folded second helix (H2). In CT, there is an additional helix (H3). The Tctex (orange), LC8 (red), and LC7 (yellow) binding sites are well characterized in Dros. IC and Rat IC2C, and their position in CT was predicted based on sequence and structure comparison. The C-terminal domain is predicted to contain seven WD40 repeats. The CT constructs IC_{FL}, IC_{1-88}, IC_{37-88}, IC_{1-260}, $IC_{100-260}$, $IC_{140-260}$, $IC_{216-260}$, and $IC_{216-260}$ are used in this paper; the IC_{1-35} construct was used in prior work. CT p150^{Glued} is predicted to have a Cap-Gly domain near the N-terminus, and two coiled-coil domains, CC1 and CC2, that are separated by an intercoil domain (ICD). CC1 is further divided into two regions called CC1A and CC1B. $p150_{478-680}$ ($p150_{CC1B}$) is the construct used in this work. CT NudE is predicted to have an N-terminal coil-coiled (CC) region followed by disorder. $NudE_{1-190}$ ($NudE_{CC}$) is the construct used in this work. B) Contextual models of dynein (dark grey) with the IC subcomplex highlighted (light grey). The top model depicts the interaction between N-IC and the p150^{Glued} subunit of dynactin (blue) while the bottom model depicts the interaction between N-IC and NudE (green). In both models, dynein is a processive motor traveling towards the minus end of a microtubule. For clarity, shown on a larger scale to the right is the IC subcomplex consisting of CT IC_{1-260} with the homodimeric dynein light chains: Tctex (orange), LC8 (red), and LC7 (yellow).

Results

CT p150$_{CC1B}$ and NudE$_{CC}$ are dimeric, while CT IC$_{1-260}$ is monomeric

Studies of interactions of IC constructs with light chains show that two monomeric IC chains are brought together by the dimeric light chains to create a 'ladder-like' polybivalent scaffold[16,52,81,91]. Here we use a construct of IC that includes all the binding sites for the light chains and non-dynein proteins, the full-length light chains, and coiled-coil domains of non-dynein proteins and employ multiple techniques to determine their association states. Sedimentation velocity analytical ultracentrifugation (SV-AUC), which gives information about a protein's mass and shape in solution, was used to determine each protein's heterogeneity and size. A larger sedimentation coefficient (S) indicates a protein with a larger mass and/or a protein with a smaller, shape dependent, frictional ratio[161]. SV-AUC reveals that IC$_{1-260}$, the three light chains, and non-dynein proteins p150$_{CC1B}$ and NudE$_{CC}$ all have similar sedimentation coefficients (in the 2-3 S range, Figure 5.2A). Further, each subunit shows a single sharp peak in the c(S) distribution, indicating that the proteins are homogeneous in solution. In comparison to the sedimentation coefficients for IC$_{1-260}$ and the light chains, both p150$_{CC1B}$ and NudE$_{CC}$ have smaller sedimentation coefficients than would typically be expected for globular proteins with their respective dimeric masses. However, as sedimentation coefficients also depend on shape, the smaller values observed for p150$_{CC1B}$ and NudE$_{CC}$ are consistent with their predicted coiled-coil structures (which result in elongated, less-compact, rod-like shapes) causing slower sedimentation than if they were globular.

Sedimentation equilibrium analytical ultracentrifugation (SE-AUC) confirmed the dimeric coiled-coil state of p150$_{CC1B}$ and NudE$_{CC}$. SE-AUC data fit to a monomer-dimer binding model results in dimerization dissociation constants of 0.03 μM and 0.20 μM for p150$_{CC1B}$ and NudE$_{CC}$, respectively, indicating that both are strong dimers and that, of the two, p150$_{CC1B}$ is the tighter dimer (Figure 5.2B). This result is underscored by circular dichroism (CD) spectra acquired at temperatures ranging from 5-60ºC (Figure 5.2C), which show that while both proteins have α-

helical secondary structure[162], they have different mechanisms of unfolding. Changes in the secondary structure for $p150_{CC1B}$ are most dramatic between 30 and 35ºC, whereas $NudE_{CC}$ shows gradual unfolding across the 5-40ºC temperature range. Also of note, the coiled-coil structures of $p150_{CC1B}$ and $NudE_{CC}$ are confirmed by their CD spectra which show helical structures and values for 222/208 ratios larger than 1 (Figure 5.2C).

To confirm our expectation that IC_{1-260} is a monomer in solution, we used size exclusion chromatography (SEC) with multi-angle light scattering (MALS) detection (Figure 5.2D). The measured mass of approximately 29.5 kDa based on the MALS data is consistent with the expected mass of 29.2 kDa for an IC_{1-260} monomer.

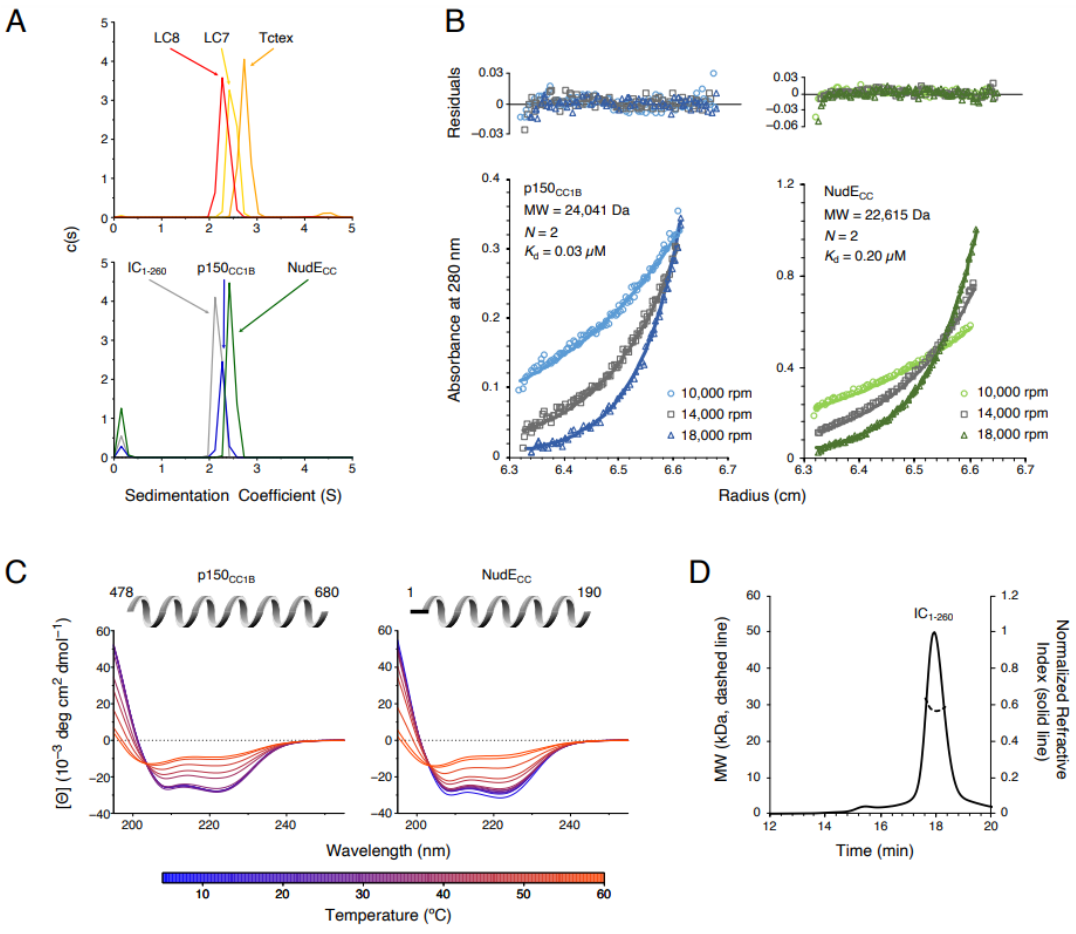

Figure 5.2. CT $p150_{CC1B}$, $NudE_{CC}$, and dynein light chains are dimeric, whereas CT IC_{1-260} is monomeric. A) SV-AUC profiles for LC8 (red), LC7 (yellow), and Tctex (orange) (top), and, IC_{1-260} (grey), $p150_{CC1B}$ (blue), and $NudE_{CC}$ (green)

(bottom). All samples were at protein concentration of 30 μM. B) SE-AUC data for $p150_{CC1B}$ (blue) and $NudE_{CC}$ (green) at three speeds (10,000, 14,000, and 18,000 rpm). Data were fit to a monomer-dimer binding model. The quality of the fits to this model is reflected by the plots of the residuals on top. The monomeric masses determined by fitting this data compare very well to the masses expected based on the sequences for the constructs. The stoichiometry (*N*) values of 2 indicate that both $p150_{CC1B}$ and $NudE_{CC}$ are dimers in solution. C) CD spectra of $p150_{CC1B}$ and $NudE_{CC}$ acquired at temperatures in the 5 to 60ºC range. The shape of the spectra for both $p150_{CC1B}$ and $NudE_{CC}$ indicate α-helical secondary structure, and the 222/208 ratios (1.04 and 1.00 for $p150_{CC1B}$ and $NudE_{CC}$, respectively) are consistent with coil-coiled structures. D) MALS of IC_{1-260} gives an estimated mass of 29.5 kDa, which indicates that on its own, IC_{1-260} exists as a monomer in solution (calculated mass of monomer is 29.2 kDa).

CT IC_{1-260} is stabilized by long range contacts

The N-IC from CT has a domain architecture with structural elements that are similar in all experimentally characterized ICs: an N-terminal single α-helix (SAH), a nascent helix 2 (H2), and long disordered linkers (Figure 5.1A). CT IC is unique, however, in also including a strongly predicted third helix (H3) corresponding to residues 170-200[106,108,109,163–166] (Figure 5.3A, A2.2). To validate this predicted secondary structure, and determine its impact on global stability, we acquired CD spectra of various IC constructs: IC_{1-260}, IC_{1-88} (containing SAH and H2), $IC_{100-260}$ (containing linker, H3, and the LC7 binding site), and $IC_{160-240}$ (containing H3 and shorter linker). All the CD spectra show two minima around 208 and 222 nm, which is indicative of the presence of α-helical secondary structure[162] (Figure 5.3B-E). The estimated fractional helicity[167] values for these constructs at 5ºC is in the range of 20%-35%, which matches the fraction of either NMR-determined or predicted helical residues in each construct (Figure 5.3A). Notably, the longest IC construct (IC_{1-260}) is the most thermostable of the four, and resists unfolding until above 50ºC. Comparatively, IC_{1-88}, $IC_{160-240}$, and $IC_{100-260}$ exhibit some loss in secondary structure significantly below 50ºC. In particular, $IC_{100-260}$ appears to be the least stable of the three, with a gradual loss of secondary structure that begins around 25ºC and a complete loss of helical structure around 40ºC. From this, it is reasonable to hypothesize that some degree of tertiary contacts may exist only within IC_{1-260} and underlie the increase in its structural stability and cooperative unfolding.

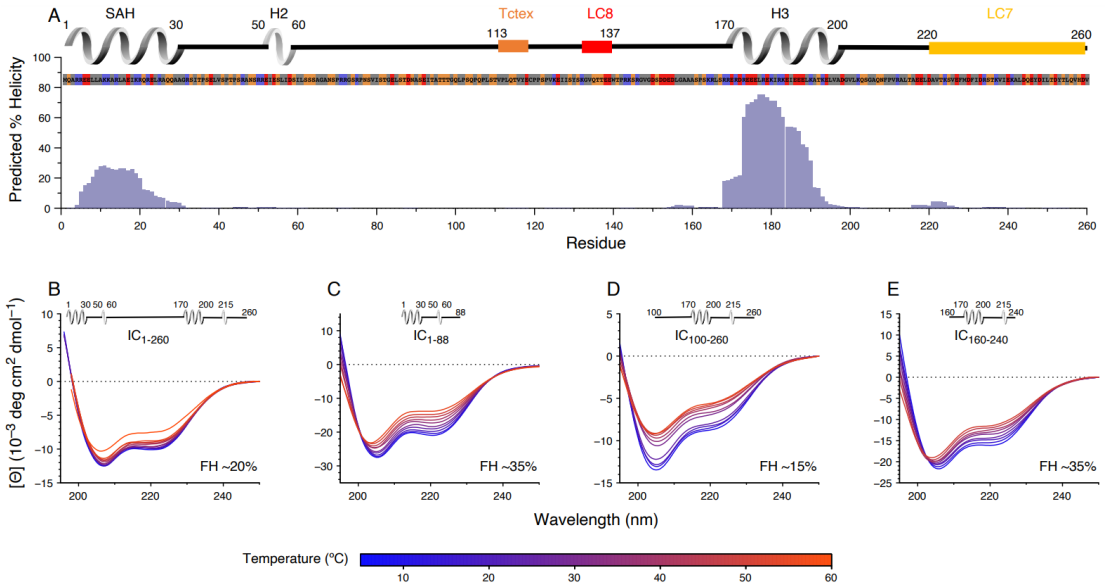

Figure 5.3. Secondary structure and thermal stability of CT IC. A) Agadir prediction for IC_{1-260} showing the percent helicity by residue (purple). Shown above the plot is a schematic structure for IC_{1-260} with labels for SAH, H2 and H3 above the helical structure. The sites for lights chains binding are also indicated. The amino acid sequence under the schematic is colored by amino acid type: hydrophobic (grey), positive (red), negative (blue), neutral (orange). Variable temperature CD spectra of B) IC_{1-260}, C) IC_{1-88}, D) $IC_{100-260}$, and E) $IC_{160-240}$. The shapes of the spectra for all constructs indicates a mixture of α-helical secondary structure and regions of intrinsic disorder. Loss in structure, or lack thereof, over a temperature range of 5-50ºC (blue for lowest, red for highest) indicates how each construct varies in stability and indicates that IC_{1-260} is the most thermally stable. The fractional helicity (FH) of each construct at 5ºC was calculated based on the experimentally observed mean residue ellipticity at 222 nm as explained in the methods section.

To further probe possible tertiary contacts within IC_{1-260}, SV-AUC was employed to determine if there is binding between the IC_{1-88} and $IC_{100-260}$ constructs (Figure A2.3 A). IC_{1-260} has a sedimentation coefficient of approximately 2.2 S, whereas $IC_{100-260}$ has a smaller sedimentation coefficient of 1.5 S due to its lower mass and its expected elongation compared to IC_{1-260} (Figure A2.3 B). Upon addition of IC_{1-88} at a 1:2 molar ratio ($IC_{100-260}$:IC_{1-88}), a complex with a sedimentation coefficient of 2 S is formed, indicating a strong interaction between the two constructs; a peak for the excess of IC_{88} is not observed because IC_{1-88} does not absorb at 280 nm. The IC_{1-88}/$IC_{100-260}$ complex has a slightly smaller sedimentation

coefficient than that of IC_{1-260}, which can be explained by a greater degree of elongation for this complex compared to IC_{1-260}. SEC-MALS determined a mass of 30.3 kDa for the $IC_{1-88}/IC_{100-260}$ complex which matches the expected mass of 30.6 kDa for a 1:1 complex (Figure A2.3C). Together, these data confirm the presence of tertiary contacts within IC_{1-260} and explain the increase in its stability compared to smaller IC constructs (Figure 5.3).

Identifying disordered domains of CT IC in the context of IC_{1-260}

To identify the disordered regions of IC_{1-260}, we used nuclear magnetic resonance (NMR) spectroscopy to study isotopically-labeled protein samples. The limited chemical shift dispersion in the ^{1}H-^{15}N TROSY spectrum for IC_{1-260} at 10ºC (Figure 5.4B), along with appearance of the majority of the peaks in the CLEANEX experiment at this temperature (Figure 5.4C), indicate that the peaks observed in the spectra are for the disordered regions of the protein. Using triple resonance experiments on a $^{2}H/^{13}C/^{15}N$ labeled sample, we assigned almost all of the observable peaks, corresponding to 37% (90 of 245) of the non-proline residues (Figure 5.4B). The handful of "unassigned" peaks in Figure 5.4B mainly correspond to side-chain amides or to peaks from minor conformers that arise due to the slow cis/trans isomerization of the peptide bond between prolines and the amino acid preceding them[168]. The assigned peaks all correspond to residues in disordered linker regions of IC_{1-260} and these peaks vary considerably in intensity (Figure 5.4A).

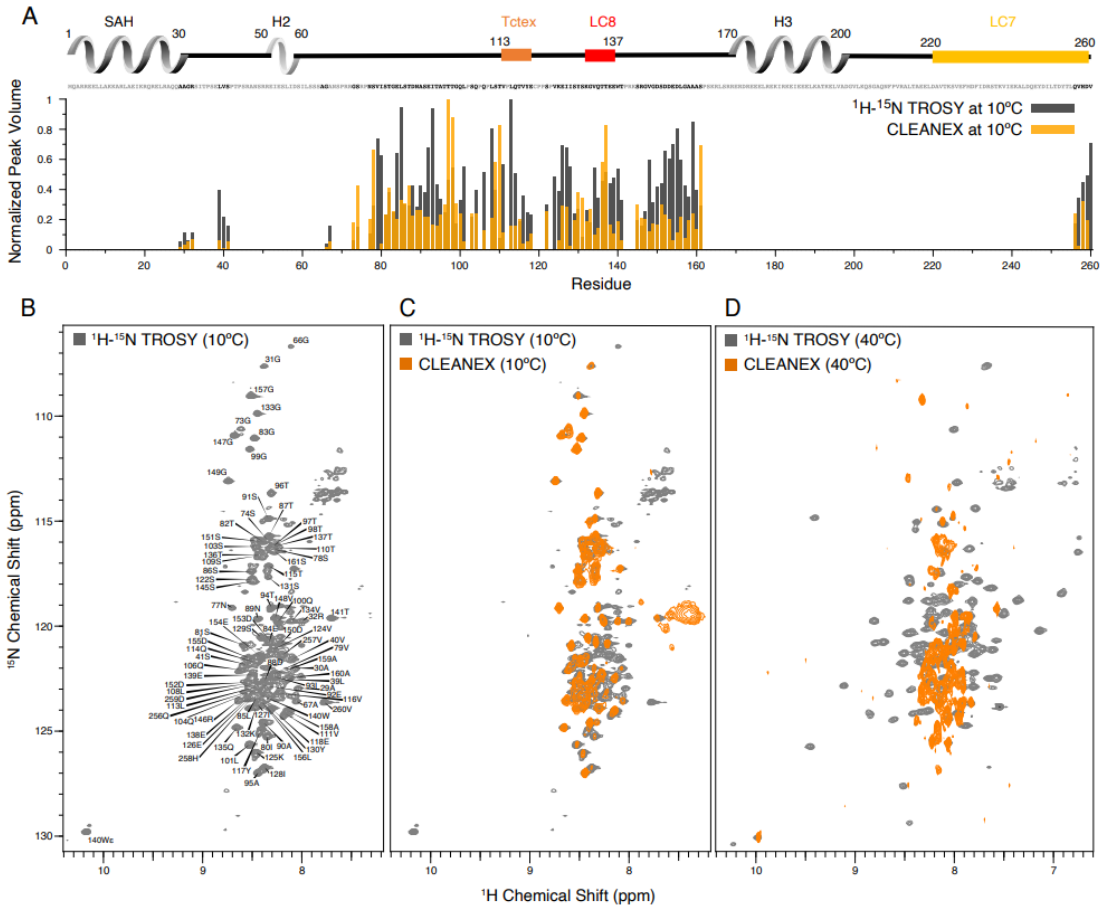

Figure 5.4. Identification of disordered linkers of CT IC$_{1\text{-}260}$ using NMR spectroscopy. A) Plot showing the normalized peak volumes at 10ºC in the ^{1}H-^{15}N TROSY spectrum (grey) and in the CLEANEX spectrum (orange) of the amides that could be assigned. Assigned residues are in black in the sequence above the plot (and unassigned residues in grey); all assigned residues are from disordered regions of IC$_{1\text{-}260}$. B) ^{1}H-^{15}N TROSY spectrum of IC$_{1\text{-}260}$ acquired at 800 MHz at 10ºC showing amide assignments. C) Overlay of a CLEANEX spectrum (orange) with the ^{1}H-^{15}N TROSY spectrum (grey) at 10ºC, shows that most of the assignable residues are in exchange with the solvent on the timescale of the CLEANEX experiment. D) At 40ºC, ^{1}H-^{15}N TROSY spectrum (grey) shows new peaks appearing with greater chemical shift dispersion for IC$_{1\text{-}260}$ in the 800 MHz. Overlaying a CLEANEX spectrum (orange) at this temperature reveals that most of the new peaks in the ^{1}H-^{15}N TROSY spectrum are from amides that are slow to exchange with the solvent and therefore are not observed in the CLEANEX spectrum.

By increasing the temperature to 40ºC, peaks with much greater chemical shift dispersion become visible in the ^{1}H-^{15}N TROSY spectrum (Figure 5.4D), while peaks corresponding to disordered regions of the protein disappear. Based on CD data

(Figure 5.3B), IC_{1-260} does not undergo significant secondary structural changes in the 10 to 40ºC temperature range. Therefore, the newfound peak dispersion at higher temperature is not due to an increase in secondary structure. Rather, the appearance of peaks at elevated temperatures is most likely due to an increased rate of molecular tumbling causing peaks from more structured parts of the protein that were too broad to observe at lower temperatures to narrow and become visible. The conclusion that these peaks belong to residues in ordered regions is supported by their chemical shift dispersion and their absence in the CLEANEX spectrum at 40ºC (Figure 5.4D). Some additional peaks are observed in the CLEANEX spectrum at 40ºC that do not appear in the TROSY spectrum at this temperature; these correspond to amides in disordered regions of the protein that are in exchange with the solvent with exchange rates that makes them detectable in the CLEANEX experiment, but invisible in the TROSY experiment. Together, these spectra support the conclusion that IC_{1-260} contains both structured and disordered regions, with the most disordered regions resulting in peaks that are the most easily assigned by NMR at low temperature. Even with the use of deuterated samples and TROSY-based experiments, the peaks at 40ºC corresponding to more structured regions were, for the most part, not assignable as their rapid relaxation resulted in extremely low signal intensities in backbone assignment experiments. IC_{1-260} samples with salt concentrations of 20 mM and 250 mM were also explored at both 10ºC and 40ºC, to ensure that electrostatic interactions were not the cause for missing peaks (Figure A2.4).

The compact structure of CT IC_{1-260}

To identify the residues involved in the tertiary contacts within IC_{1-260}, we collected NMR data on smaller fragments of IC; IC_{1-88}, $IC_{100-260}$, and $IC_{160-240}$ (Figures 5.1, 5.5A). Based on the long-range interactions observed between IC_{1-88} and $IC_{100-260}$ (Figure 5.3B, S5.3), we hypothesized that an intramolecular interaction between the SAH region (residues 1-30), and the H3 region (residues 170-200) could be contributing to the stability of IC_{1-260}. To test this, we added unlabeled $IC_{160-240}$ to ^{15}N-labeled IC_{1-88}. Peak disappearances and peak shifts for the spectrum of IC_{1-88}

upon addition of $IC_{160-240}$ indicate some degree of interaction (Figure 5.5B-C). However, when conducting the same experiment with unlabeled $IC_{100-260}$ instead of unlabeled $IC_{160-240}$, the effects were much more pronounced (Figure 5.5B-C). In both cases, peaks corresponding to residues in the SAH and H2 regions of IC_{1-88} either lost intensity or disappeared completely, while peaks from linker region residues remained largely unaffected, especially those that are near the C-terminus of IC_{1-88}. To continue to narrow down the location of the interaction, an additional IC construct, $IC_{216-260}$, that does not include the H3 region, was ^{15}N-labeled and completely assigned. Upon titration with ^{15}N-labeled IC_{1-88}, many of the $IC_{216-260}$ peaks disappear or shift (Figure 5.5D-E). From this, we conclude that the tertiary contacts within IC_{1-260} are largely between the SAH/H2 regions and a region near the C-terminus (residues 220-250) that overlaps with the LC7 binding site. We note that these interactions appear to be even stronger in the context of the longer disordered chain encompassing the Tctex and LC8 binding sites based on the more substantial disappearances of peaks from the SAH/H2 regions when ^{15}N-labeled IC_{1-88} was titrated with unlabeled $IC_{100-260}$ (Figure 5.5C).

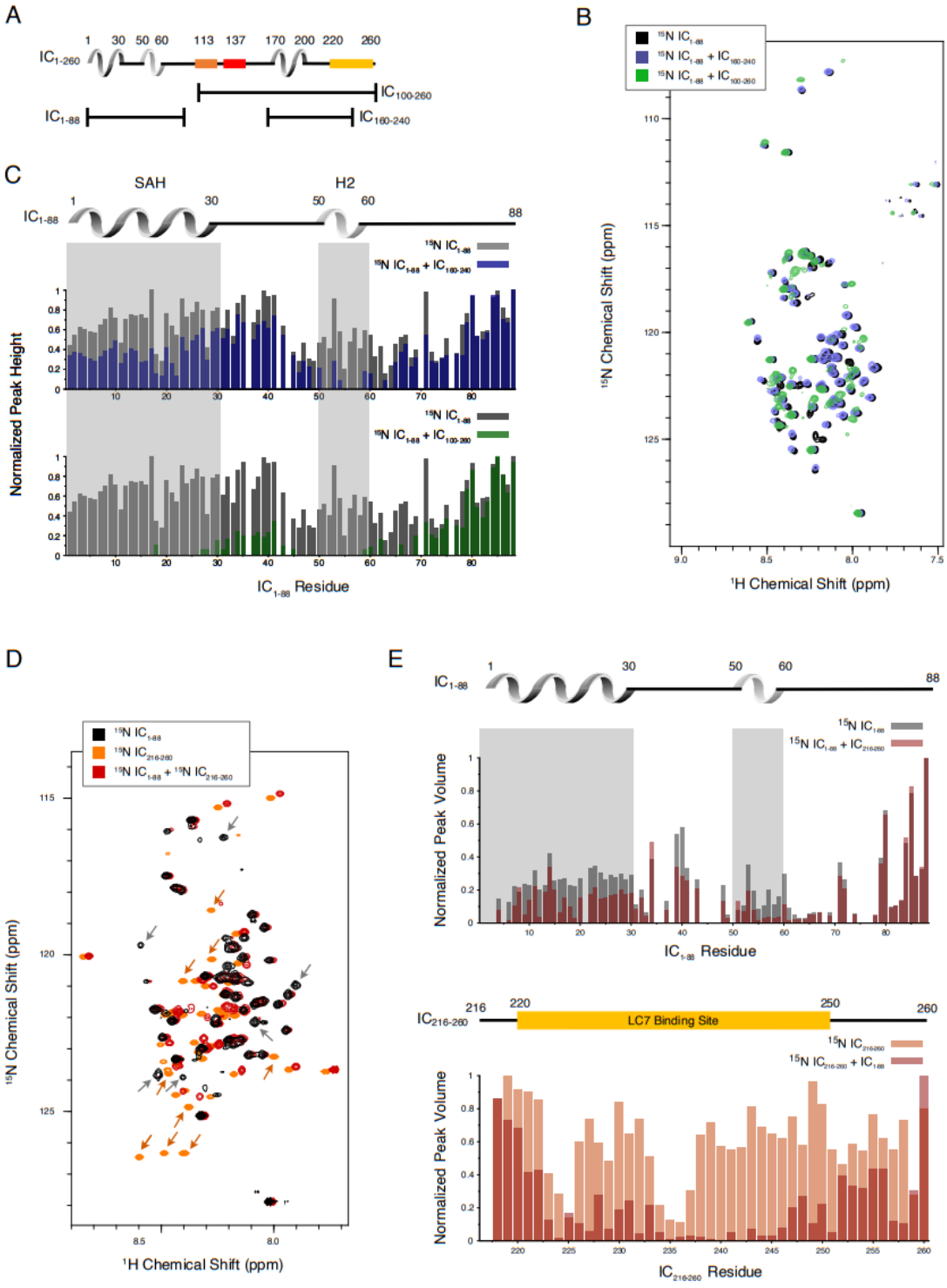

Figure 5.5. Evidence of tertiary contacts between the N and C-termini within CT IC$_{1-260}$. A) Domain architecture diagram for CT IC$_{1-260}$ with bars shown below corresponding to the IC$_{100-260}$, IC$_{1-88}$, and IC$_{160-240}$ constructs. B) ^{1}H-^{15}N TROSY overlays of free ^{15}N-labeled IC$_{1-88}$ (black) and ^{15}N-labeled IC$_{1-260}$ bound to unlabeled

$IC_{160-240}$ (purple) and $IC_{100-260}$ (green). Note, spectra are deliberately offset in the ^{1}H dimension to help visualize overlapping peaks. C) Normalized peak heights in the ^{1}H-^{15}N TROSY spectra for (top) ^{15}N-labeled $IC_{1–88}$ (grey) and ^{15}N-labeled $IC_{1–88}$ + $IC_{160-240}$ (purple) and (bottom) ^{15}N-labeled $IC_{1–88}$ (grey) and ^{15}N-labeled $IC_{1–88}$ + $IC_{100-260}$ (green). D) ^{1}H-^{15}N HSQC overlay of ^{15}N-labeled IC_{1-88} (black), ^{15}N-labeled $IC_{216-260}$ (orange), and ^{15}N-labeled $IC_{216-260}$ bound to ^{15}N-labeled IC_{1-88} (red). Arrows highlight some of the more significant peak disappearances for IC_{1-88} (grey arrows) and $IC_{216-260}$ (orange arrows). Note, spectra are deliberately offset by 0.03 ppm in the ^{1}H dimension to help visualize overlapping peaks. E) Normalized peak volumes in the ^{1}H-^{15}N HSQC spectra for ^{15}N-labeled $IC_{216-260}$ (top, grey columns) and ^{15}N-labeled $IC_{216-260}$ (bottom, orange columns) when free and when in the presence of the other protein (red columns).

Interactions of $p150_{CC1B}$ and $NudE_{CC}$ with CT IC

The interaction between CT IC_{1-88} (a smaller IC construct IC_{1-88} that includes only the p150 and NudE binding domains) and $p150_{CC1B}$ has been previously reported[97] and is shown here only for comparative purposes (Figure 5.6C, top left). Here, we explore the interaction between IC_{1-88} and $NudE_{CC}$ under similar conditions. By SV-AUC, IC_{1-88} bound to $p150_{CC1B}$ shows a sedimentation coefficient of 3.7, whereas for IC_{1-88} bound to $NudE_{CC}$ the coefficient is only 2.7 (Figure 5.6A). This is surprising considering that the two complexes have similar overall masses and binding stoichiometries, and suggests that the complex with $NudE_{CC}$ is less compact than the complex with $p150_{CC1B}$. ITC indicates that, like IC_{1-88} and $p150_{CC1B}$ (Figure 5.6C, top left), IC_{1-88} and $NudE_{CC}$ bind with a 1:1 molar ratio (Figure 5.6C, bottom left), which corresponds to one $NudE_{CC}$ dimer binding two IC_{1-88} monomeric chains. IC_{1-88} binds to $NudE_{CC}$ with a dissociation constant (K_d) of 0.3 μM, which is weaker than the nM affinity estimated for IC_{1-88} binding to $p150_{CC1B}$[97]. Previously published NMR titrations of unlabeled $p150_{CC1B}$ with ^{15}N-labeled IC constructs that contain (IC_{1-88}) or do not contain (IC_{37-88}) the SAH region demonstrate that both the SAH and H2 regions bind to $p150_{CC1B}$[97] (Figure 5.6B). IC_{1-88} binding to $p150_{CC1B}$ exhibits an unusual two-step ITC thermogram (Figure 5.7C, top left) whereas a one-step thermogram is seen when $NudE_{CC}$ is titrated with IC_{1-88} (Figure 5.7C, bottom left), suggesting a difference in the mode of binding. We propose that the two step thermogram is due to binding to both the SAH and H2 regions of IC_{1-88} with

p150$_{CC1B}$, while in contrast the single step thermogram is due to NudE binding only to the SAH region.

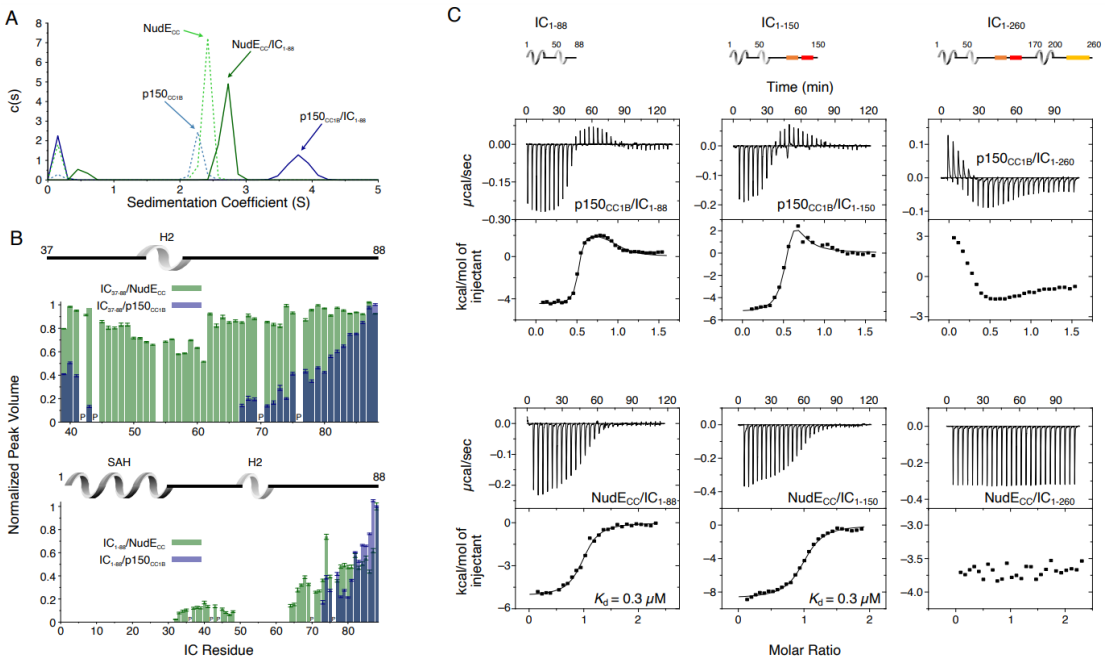

Figure 5.6. Binding interactions of CT IC to p150$_{CC1B}$ and NudE$_{CC}$. A) SV-AUC profiles for samples containing p150$_{CC1B}$ (blue dashed line), NudE$_{CC}$ (green dashed line), IC$_{1-88}$/p150$_{CC1B}$ complex (blue solid line) and IC$_{1-88}$/NudE$_{CC}$ complex (green solid line) show that IC$_{1-88}$ complexes have a larger sedimentation coefficient with p150$_{CC1B}$ than with NudE$_{CC}$. No data were collected for free IC$_{1-88}$ because it has no absorbance at 280 nm. B) Normalized peak volumes in the ^{1}H-^{15}N HSQC spectra for ^{15}N-labeled IC$_{37–88}$ (top) or ^{15}N-labeled IC$_{1–88}$ (bottom) when titrated with unlabeled p150$_{CC1B}$ (blue) and NudE$_{CC}$ (green). "P" indicates proline residues. No peak disappearance for IC$_{37-88}$ was observed when NudE$_{CC}$ was added. C) ITC thermograms for p150$_{CC1B}$ titrated with IC$_{1-88}$ (top left), IC$_{1-150}$ (top middle), and IC$_{1-260}$ (top right), and for NudE$_{CC}$ titrated with IC$_{1-88}$ (bottom left), IC$_{1-150}$ (bottom middle), and IC$_{1-260}$ (bottom right), collected at 25ºC (pH 7.5). For IC$_{1-260}$, reduced and endothermic binding is observed with p150$_{CC1B}$ whereas no binding is observed with NudE$_{CC}$. All results indicate that the H2 region of CT IC binds weakly to p150$_{CC1B}$ but does not bind NudE$_{CC}$.

To identify if NudE$_{CC}$ does indeed only bind to the SAH region or to both the SAH and the H2 regions, NMR titrations with unlabeled NudE$_{CC}$ and ^{15}N-labeled IC$_{1-88}$ or IC$_{37-88}$ constructs were carried out. In contrast to the NMR titration with

$p150_{CC1B}$, peaks for residues corresponding to the H2 region did not disappear upon titration of ^{15}N IC_{37-88} with $NudE_{CC}$ (Figure 6.6B, top), confirming our proposal that the IC H2 region does not directly bind to $NudE_{CC}$ whereas this region does directly bind to $p150_{CC1B}$. For both $p150_{CC1B}$ and $NudE_{CC}$, titration into ^{15}N IC_{1-88} resulted in IC_{1-88} peaks disappearing for both the SAH and H2 regions (Figure 6.6B, bottom). The disappearance of IC_{1-88} peaks for both the SAH and H2 regions with $NudE_{CC}$ can be explained by an interaction between the SAH and H2 regions[91] that relays the change in correlation time of the SAH region upon binding to $NudE_{CC}$ to the H2 region. The difference in how IC binds $p150_{CC1B}$ and $NudE_{CC}$ observed in the ITC experiment is conserved when a somewhat longer IC construct is used (IC_{1-150}, Figure 6.6C, center), whereas binding is severely diminished when using a construct containing the entire N-IC region (IC_{1-260}, Figure 6.6C, right). Further data and discussion of these observations are presented below.

Interactions using CT IC_{1-260}

SV-AUC was initially used to characterize complex formation between IC_{1-260} and each of the other binding partners. The largest sedimentation coefficient was observed for the IC_{1-260}/LC8 complex; for all other complexes, a less dramatic peak shift was seen, a possible indicator of weaker binding, a dynamic equilibrium between free and bound states, and/or a shift to a more elongated conformation (Figure 5.7A). For the IC_{1-260}/$NudE_{CC}$ complex, no significant change was observed in the SV-AUC data relative to unbound $NudE_{CC}$. However, the absence of a peak corresponding to free IC for the sample containing IC_{1-260} and $NudE_{CC}$ indicates that some degree of binding takes place. This result along with our ITC (Figure 5.6C, bottom right) and NMR (Figure 5.7B, right) data for this complex indicates that binding of IC_{1-260} to $NudE_{CC}$ is very weak and is thus only observed at high protein concentrations.

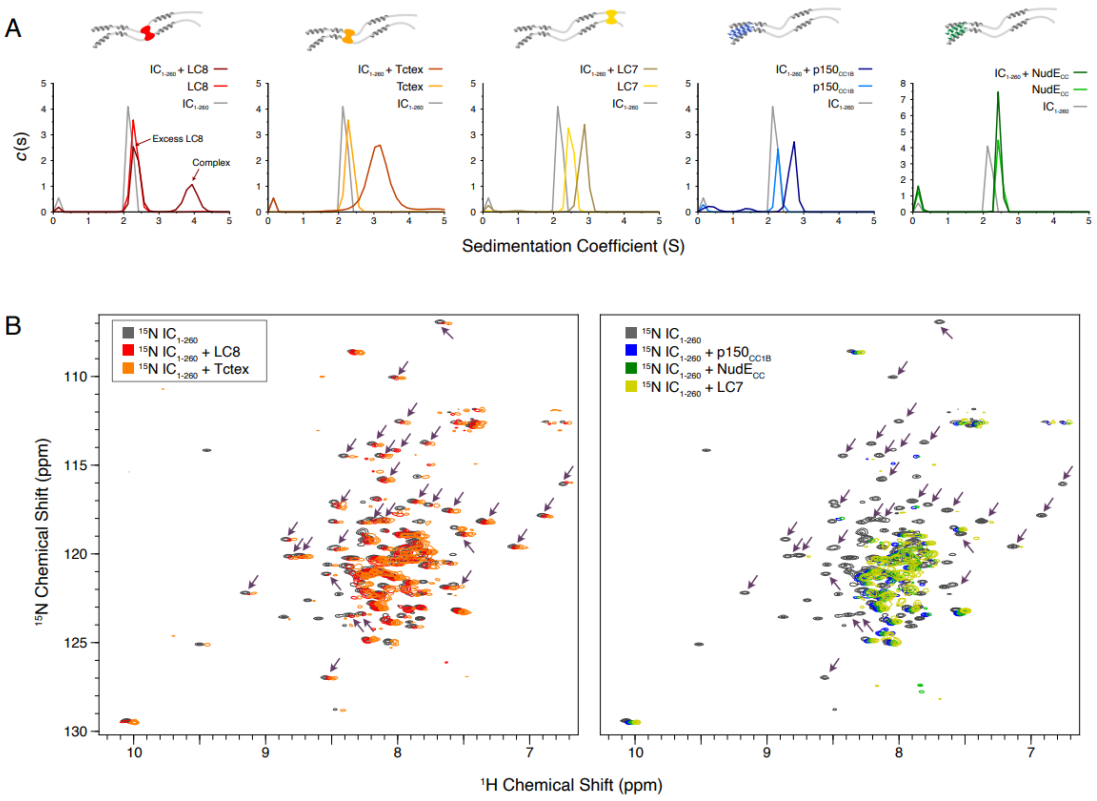

Figure 5.7. Binding characterization of binary complexes of IC_{1-260}. A) SV-AUC of IC_{1-260}/LC8, IC_{1-260}/Tctex, IC_{1-260}/LC7, IC_{1-260}/$p150_{CC1B}$, and IC_{1-260}/$NudE_{CC}$. Data for the binary complexes is overlayed with data for each protein individually to better see shifts in the sedimentation coefficient of the binary complexes. B) ^{1}H-^{15}N TROSY overlays of free ^{15}N-labeled IC_{1-260} (black) and ^{15}N-labeled IC_{1-260} bound to unlabeled binding partners in a 1:1.5 molar ratio. The spectra were offset by 0.03 ppm in the ^{1}H dimension to help illustrate changes in peak intensities. Changes in peak appearances/shifts/disappearances seem to be similar for LC8 (red) and Tctex (orange) versus changes seen for $p150_{CC1B}$ (blue), $NudE_{CC}$ (green), and LC7 (yellow). Arrows indicate peaks that remain when LC8 and Tctex are added to IC_{1-260}, but disappear when $p150_{CC1B}$, $NudE_{CC}$, and LC7 are added.

Although we were unable to assign the NMR spectrum of IC_{1-260} at 40ºC because of poor sensitivity and unfavorable relaxation times, overlays of ^{1}H-^{15}N TROSY spectra for free ^{15}N-labeled IC_{1-260} and binary complexes of ^{15}N-labeled IC_{1-260} with unlabeled binding partners (Figure 5.7B) show distinct patterns of peak disappearances. Upon addition of Tctex or LC8 (Figure 5.7B, left), only a handful of IC_{1-260} peaks disappear and the patterns of disappearances are similar, as expected based on the proximity of the Tctex and LC8 binding sites. When either LC7,

$p150_{CC1B}$, or $NudE_{CC}$ is added, considerably more peaks disappear in the spectra for ^{15}N-labeled IC_{1-260} (Figure 5.7B, right) and, surprisingly, similar patterns of disappearances occur even though the LC7 binding site is at the C-terminus of IC_{1-260} whereas the $p150_{CC1B}$ and $NudE_{CC}$ binding sites are at the N-terminus. The similar patterns of peak disappearances suggests that regions of the N and C-termini of IC_{1-260} interact in such a way that when one end of IC_{1-260} is bound it affects the peak intensities of the other end and vice versa.

Multivalency relieves IC autoinhibition

Following characterization of individual binding events, we sought to reconstitute full N-IC subcomplexes. Each individual subunit was first expressed and purified individually prior to dynein subcomplex formation (IC_{1-260}/Tctex/LC8/LC7), achieved by mixing IC_{1-260} with the dynein light chains in a 1:1.5 (IC to LC) molar ratio. To this dynein subcomplex, $p150_{CC1B}$ or $NudE_{CC}$ was added to create two larger subcomplexes. Each subcomplex was re-purified by size exclusion chromatography (SEC) to remove any excess of the binding partners (which elute at ~215 mL) and to assess their overall stability (Figure 5.8A). The shape and symmetry of the eluting SEC peak for the dynein subcomplex (~140 mL) indicates a weaker binding affinity than when either $p150_{CC1B}$ (~120 mL) or $NudE_{CC}$ (~130 mL) was added. In all cases however, each expected subunit was detected by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) of collected fractions (Figure 5.8B), thus validating successful assembly.

Using SV-AUC, we show that when all dynein light and intermediate chains are present, the fully bound dynein subcomplex sediments as a single peak with a sedimentation coefficient of approximately 5 S (Figure 5.8C). Surprisingly, addition of $NudE_{CC}$, which adds approximately 45 kDa of mass, only slightly increases the sedimentation coefficient. In contrast, addition of $p150_{CC1B}$ increases the sedimentation coefficient to approximately 7 S. The difference in the SV-AUC data between the addition of $p150_{CC1B}$ and $NudE_{CC}$ is surprising based on the similar expected masses of the complexes (201 and 198 kDa, respectively). This difference

may be explained by the overall shape of the two bound complexes, details of which need to be further examined. Also, important to note is that the shift in sedimentation coefficient for $p150_{CC1B}$ when added to the IC/light chains complex is significantly more pronounced than that with IC_{1-260} alone, suggesting that binding to IC_{1-260} is significantly enhanced in the presence of the light chains.

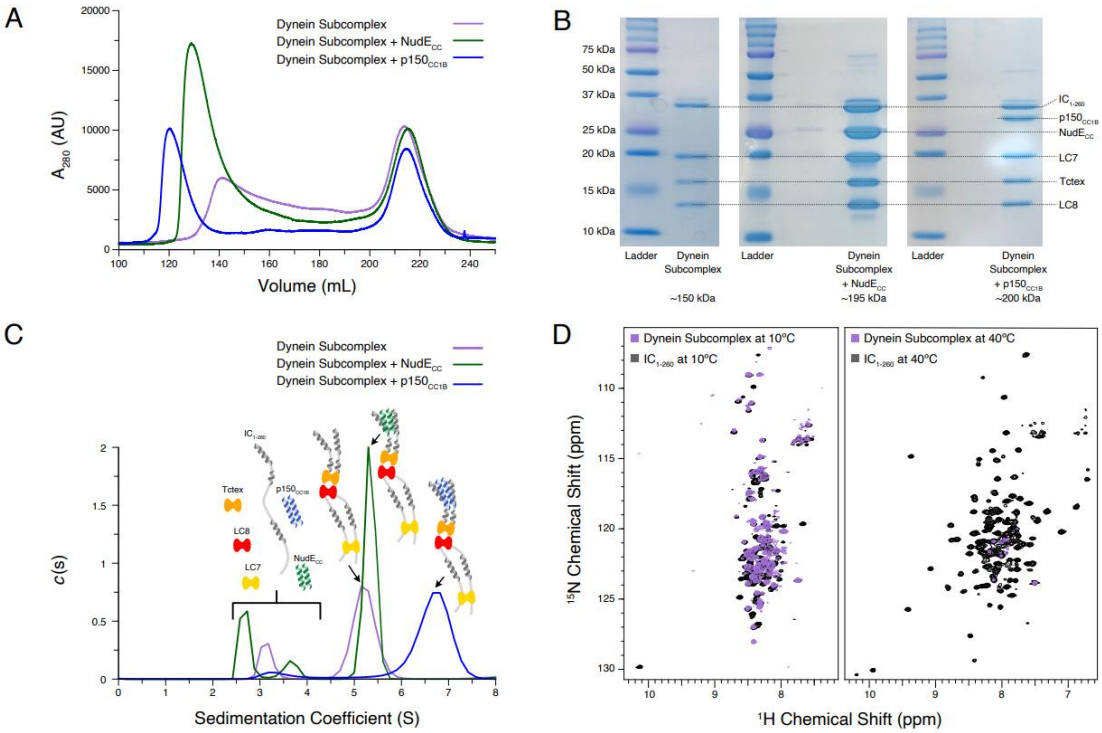

Figure 5.8. Reconstitution and characterization of dynein subcomplexes. A) SEC traces of the dynein subcomplex (IC/light chains) (purple) and the dynein subcomplex with the addition of either $p150_{CC1B}$ (blue) or $NudE_{CC}$ (green). B) SDS-PAGE gels of fractions collected from SEC for all complexes showing all expected proteins. C) SV-AUC profiles of the dynein subcomplex (purple) bound to $p150_{CC1B}$ (blue) or $NudE_{CC}$ (green). D) ^{1}H-^{15}N TROSY overlays of free IC_{1-260} (black) and the dynein subcomplex (purple). At 10ºC many peaks are still of high intensity in the 153 kDa complex, indicating that some regions remain disordered. The very few peaks at 40ºC of the bound is most likely due to the size and tumbling of the subcomplex and consistent with the fact that the majority of the peaks at this temperature are from ordered regions.

NMR spectroscopy indicates that most of the peaks in the disordered regions observed at 10ºC for free IC_{1-260} (Figure 5.8D) remain in the spectrum for the dynein subcomplex (IC_{1-260}/Tctex/LC8/LC7), indicating that there is still significant disorder

in the fully bound complex (153 kDa). In contrast, comparison of the spectra of free IC_{1-260} and the dynein subcomplex at 40ºC shows a drastic disappearance of peaks, indicating that the residues corresponding to these peaks have much longer correlation times in the bound state either due to their involvement in binding or due to the overall increase in weight of the entire subcomplex. Peaks from linker regions not involved in binding light chains that are observed at 10ºC but are not observed at 40ºC likely disappear due to rapid exchange with the solvent at the higher temperature.

Autoinhibition is retained in full-length CT IC

To determine if the autoinhibition seen in IC_{1-260} is also present in the full-length construct, IC_{FL} (residues 1-642) was produced using a baculovirus expression system. The estimated mass of 75.4 kDa for IC_{FL} from SEC-MALS matches closely to the expected monomeric mass of 79 kDa (Figure 5.9A-B). Further, SV-AUC shows a single, homogenous peak with a sedimentation coefficient of 4.0 S, as would be expected for monomeric IC_{FL} (Figure 5.9B top). This is the first study that shows full-length IC is a monomer in solution and requires the light chains for its dimerization. Adding either $p150_{CC1B}$ or $NudE_{CC}$ to IC_{FL} results in a negligible shift of the IC_{FL} peak and, in both cases, a peak corresponding to unbound $p150_{CC1B}$ or $NudE_{CC}$ was observed (Figure 5.9B). This lack of binding between IC_{FL} and $p150_{CC1B}$/$NudE_{CC}$ shows that autoinhibition occurs in IC_{FL} in a similar manner to what we have already observed in the IC_{1-260} construct. SV-AUC experiments on the dynein subcomplex interactions confirm that the autoinhibited state is released by the addition of the light chains. Full length IC bound to the three light chains (IC_{FL}/Tctex/LC8/LC7) has a sedimentation coefficient of 7.0 S but shifts to values of 7.5 and 8 S upon addition of $NudE_{CC}$ and $p150_{CC1B}$, respectively (Figure 5.9B middle). These results mimic those seen for the IC_{1-260} construct and indicate that the addition of the light chains allows $p150_{CC1B}$ or $NudE_{CC}$ to bind by relieving IC autoinhibition.

To further explore the role of each of the light chains in releasing IC autoinhibition, SV-AUC experiments with IC_{FL}/Tctex/LC8 and IC_{FL}/LC7 were conducted. The IC_{FL}/Tctex/LC8 and IC_{FL}/LC7 complexes show peaks with sedimentation coefficients of approximately 5.2 and 4.2 S, respectively (Figure 5.9B). Interestingly, the IC_{FL}/Tctex/LC8 complex exhibits no shift upon the addition of $NudE_{CC}$ but does shift to 6.2 S upon the addition of $p150_{CC1B}$ (Figure 5.9B). The IC_{FL}/LC7 complex, on the other hand, exhibits a shift upon addition of either $NudE_{CC}$ or $p150_{CC1B}$ to 4.5 or 5 S, respectively (Figure 5.9B). These data suggest that the addition of LC7 is sufficient to allow binding of $p150_{CC1B}$ and $NudE_{CC}$ to IC_{FL}, whereas the binding of Tctex and LC8 to IC_{FL} only promotes $p150_{CC1B}$ binding to IC_{FL}.

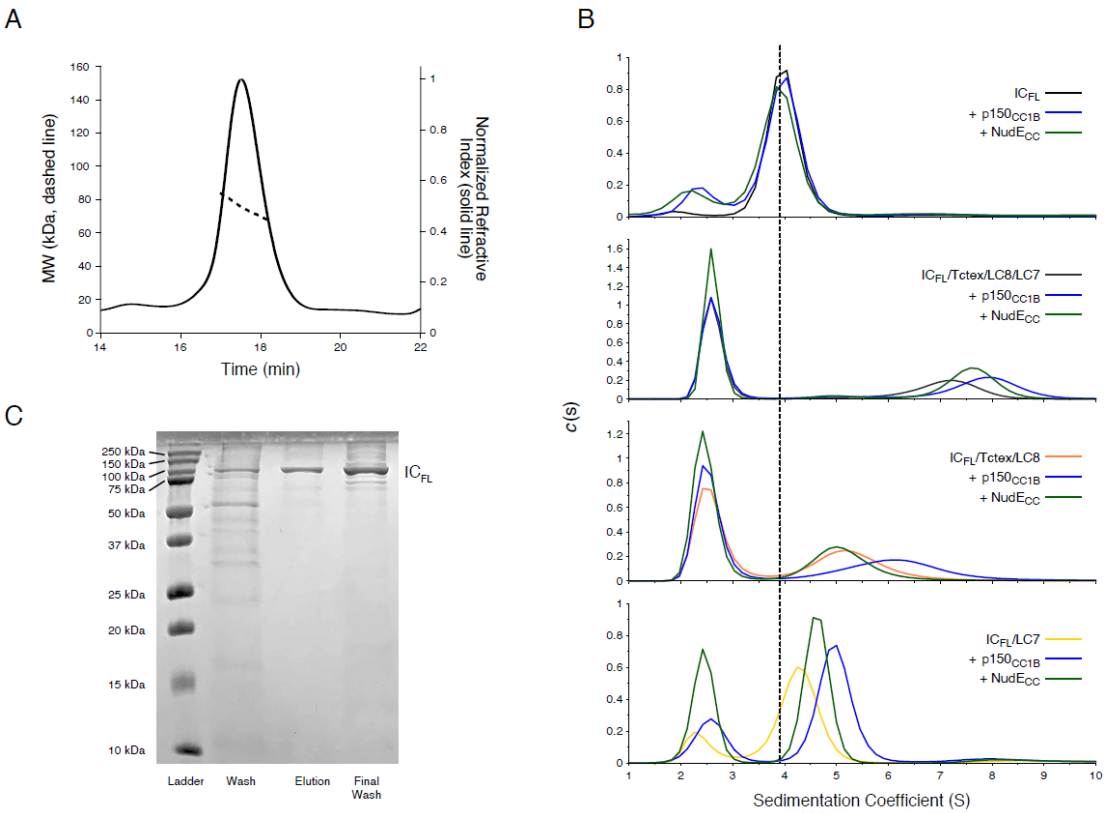

Figure 5.9. Binding characterization of CT IC_{FL} subcomplexes. A) The estimated mass of IC_{FL} from MALS is 75.4 kDa, which indicates that IC_{FL} is a monomer in the absence of binding partners. B) SV-AUC profiles of IC_{FL} (black), IC_{FL} mixed with $p150_{CC1B}$ (blue) or $NudE_{CC}$ (green), and the subcomplexes: IC_{FL}/Tctex/LC8/LC7

(gray), IC_{FL}/Tctex/LC8/LC7/$p150_{CC1B}$ (blue), IC_{FL}/Tctex/LC8/LC7/$NudE_{CC}$ (green), IC_{FL}/Tctex/LC8 (orange), IC_{FL}/Tctex/LC8/$p150_{CC1B}$ (blue), IC_{FL}/Tctex/LC8/$NudE_{CC}$ (green), IC_{FL}/LC7 (yellow), IC_{FL}/LC7/$p150_{CC1B}$ (blue), and IC_{FL}/LC7/$NudE_{CC}$ (green). The black, dashed line is centered on unbound IC_{FL} to help guide the eye. C) SDS-PAGE gel of IMAC fractions (left to right: wash, elution, and final wash) with a band for IC_{FL} migrating in accordance with the expected mass of ~79 kDa.

Discussion

The N-terminus of IC from a variety of species contains a stretch of about 300 amino acids that are primarily disordered, except for a few short α-helices. Within the first 40 residues is a fully ordered helix (the SAH region), followed by a short disordered linker and another region (H2) that forms a fully ordered helix in some species but is only a nascent helix in others[85,92,93,97]. Prior work suggests that a more disordered H2 is correlated with tighter IC/$p150^{Glued}$ binding and has shown that, for CT IC, the H2 region binds directly to $p150^{Glued}$ (although mostly in a nonspecific manner)[97]. In this work, we use a construct of IC that encompasses almost its entire 300-amino acid N-terminus to probe the interactions of CT IC with both $p150_{CC1B}$ and $NudE_{CC}$. This construct allows us to study the assembly of IC into a multivalent subcomplex with three dimeric dynein light chains and with its binding partners, $p150^{Glued}$ and NudE. Further, we demonstrate, using full-length IC, that the mechanisms at place in IC_{1-260} remain in the context of the entire IC protein. A model that illustrates the importance of autoinhibition in dynein regulation and how the assembly of the multivalent IC subcomplex relieves this autoinhibition is presented (Figure 5.10).

CT IC_{1-260} is a partially disordered compact monomer

The primary CT IC construct (IC_{1-260}) used in this work is, to date, the longest IC construct from any species made recombinantly and extensively studied by NMR, ITC, and SV-AUC[65,77,81,85,90,92,93,99,100]. IC_{1-260} far exceeds previously studied constructs from CT (res. 1-35, 37-88, and 1-88) as well as constructs from Drosophila (res. 1-60, 30-143, 84-143, 1-143, 92-260, and 114-260) and rat (res. 1-44, 1-96, and 1-112). Similar to IC from Drosophila, apo CT IC_{1-260} is monomeric (Figure 5.2A, D) and contains significant disorder (Figure 5.4). However, unlike ICs from other

species, CT IC_{1-260} has a sequence-predicted third helix, H3, corresponding to residues 170-200 (Figure 5.3A, A2.2) that is highly consistent with fractional helicity values of IC_{1-260}, $IC_{160-240}$, and $IC_{100-260}$ determined by CD spectroscopy (Figure 5.3B-E). IC_{1-260} is the most thermostable construct, resisting unfolding at temperatures of up to 50ºC (Figure 5.3B), a feature that we attribute to long-range tertiary contacts between the N- and C-termini (Figure 5.5, A2.3). SV-AUC and SEC-MALS experiments confirm the presence of a binding interaction between IC_{1-88} and $IC_{100-260}$ (Figure A2.3). Further, NMR titrations with N- and C-terminal IC constructs identify a self-association interaction between residues in the SAH and the LC7 binding site of IC_{1-260} (Figure 5.5). These tertiary contacts are also seen in comparisons of NMR spectra of ^{15}N-labeled IC_{1-260} upon addition of $p150_{CC1B}$, $NudE_{CC}$, or LC7, which reveal a similar pattern of peak disappearances and chemical shifts that are not the same as those observed upon addition of Tctex or LC8 (Figure 5.7B). Finally, long distance tertiary interactions, and their impact on IC binding interactions, are suggested by the dramatic reduction in binding of $p150_{CC1B}$ or $NudE_{CC}$ with IC_{1-260} compared to IC_{1-88} and IC_{1-150} (Figure 5.6C).

Different modes of binding of IC to $p150^{Glued}$ and NudE

The coiled-coil domains of $p150^{Glued}$ and NudE ($p150_{CC1B}$ residues 478-680, and $NudE_{CC}$ residues 1-190) show similar structure propensity and stability by SV-AUC and CD and have stable dimeric structures with dissociation constants of 0.03 and 0.20 μM, respectively (Figure 5.2). However, despite their many similarities, $p150_{CC1B}$ and $NudE_{CC}$ have different modes of binding to CT IC. First, IC interactions with $p150_{CC1B}$ are multistep, which we attribute to $p150_{CC1B}$ binding to both the SAH and H2 regions of IC[97]. In contrast, $NudE_{CC}$ binds IC in a single step (Figure 5.6C) and only to the SAH region and not to the H2 region (Figure 5.6B). Second, the $p150_{CC1B}/IC_{1-88}$ complex has a larger sedimentation coefficient when compared to the $NudE_{CC}/IC_{1-88}$ complex (3.8 S vs 2.7 S, Figure 5.6A) even though the complexes have similar expected masses and ITC shows that they have the same binding stoichiometry. This suggests that the $p150_{CC1B}/IC_{1-88}$ complex has a more compact

structure. Third, although binding is much weaker with IC_{1-260} (Figure 5.6C) compared to the smaller IC constructs for both $p150_{CC1B}$ and $NudE_{CC}$, binding to $NudE_{CC}$ is significantly weaker and is only detected at the higher concentrations used in NMR experiments (Figure 5.7B). Finally, and most importantly, the light chain/IC assembled complex shows a different sedimentation coefficient when bound to $p150_{CC1B}$ or $NudE_{CC}$. The subcomplexes both have masses of almost 200 kDa, but the $NudE_{CC}$ subcomplex has a sedimentation coefficient of ~5 S compared to ~7 S for the $p150_{CC1B}$ subcomplex (Figure 5.8C). The larger sedimentation coefficient for the $p150_{CC1B}$ subcomplex is unexplained by the small mass difference, and thus indicates a more compact and stable conformation for the subcomplex when $p150_{CC1B}$ is bound. While adding $p150_{CC1B}$ makes the light chain/IC assembled complex heavier and increases the sedimentation coefficient, adding $NudE_{CC}$ does not result in a significant increase, suggesting both weaker binding and increased friction compared to the light chain/IC/ $p150_{CC1B}$ assembled complex. A difference in overall complex mass due to a binding stoichiometry greater than 1:1 for $p150_{CC1B}$ and IC could explain the complex's larger sedimentation coefficient, however this is very unlikely. As shown in previous studies across multiple species, $p150_{CC1B}$ consistently binds IC in a 1:1 stoichiometry[65,90–92,97]. Furthermore, visualization of the light chain/IC/$p150_{CC1B}$ complex and the light chain/IC/$NudE_{CC}$ complex by SDS-PAGE (Figure 5.8B) shows similar intensity ratios between IC_{1-260}/$NudE_{CC}$ and IC_{1-260}/$p150_{CC1B}$, indicating that they exist in the same stoichiometry. Finally, MALS data of the IC_{1-88}/$p150_{CC1B}$ complex gives a mass of 66 kDa, consistent with the mass expected for a $p150_{CC1B}$ dimer and two monomeric IC_{1-88} chains (data not shown). SV-AUC experiments with IC_{FL} further confirm that $p150_{CC1B}$ and $NudE_{CC}$ have different binding modes as, once again, complexes with assembled IC_{FL}/light chain subcomplexes show a larger sedimentation coefficient when bound to $p150_{CC1B}$ (~8 S) than when bound to $NudE_{CC}$ (~7.5 S) (Figure 5.9).

Autoinhibition in IC selects for binding of p150^{Glued} over NudE

Interactions of IC_{1-260} with its binding partners observed by ITC and AUC show weak endothermic binding to $p150_{CC1B}$ but no binding to $NudE_{CC}$. Our data suggest a process, which we refer to as autoinhibition, in which IC_{1-260} adopts a compact structure that covers the SAH region necessary for binding to $NudE_{CC}$ and for strong binding to $p150_{CC1B}$, but still leaves the H2 region partly accessible so that weak binding to $p150_{CC1B}$ can still occur. The SAH region is made inaccessible by tertiary interactions within IC, as seen in an NMR titration of ^{15}N-labeled IC_{1-88} with $IC_{100-260}$ (Figure 5.5B-C). We assign these tertiary interactions to those within the LC7 binding site at the C-terminus via NMR titrations of ^{15}N-labeled $IC_{216-260}$ with IC_{1-88} (Figure 5.5D-E) but note that the binding affinity does appear stronger in the longer construct ($IC_{100-260}$) (Figures A2.3, 5.5B-C) indicating that the full context of the disordered chain encompassing the binding sites for Tctex, LC8, and LC7 is needed for strong, intramolecular interactions. These interactions likely have an autoinhibitory effect, which would explain the reduced binding affinity for $p150_{CC1B}$ and $NudE_{CC}$ for IC_{1-260} in comparison to the shorter constructs. Since this effect is more pronounced for IC_{1-260} binding with $NudE_{CC}$, we propose autoinhibition as a mechanism for partial binding of IC_{1-260} with $p150_{CC1B}$ and thus selection of $p150_{CC1B}$ over $NudE_{CC}$. Autoinhibition as a selection mechanism is underscored by our results with full-length IC (Figure 5.9). Although binding to both $p150_{CC1B}$ and $NudE_{CC}$ are inhibited with IC_{FL} alone, the addition of Tctex and LC8 rescue binding to $p150_{CC1B}$, likely by helping make the nearby H2 region more available. In contrast, binding to both $p150_{CC1B}$ and $NudE_{CC}$ is rescued via LC7 binding to IC_{FL} (Figure 5.9).

Regulation of binding to p150^{Glued} across species and the role of dynein light chains

Based on current data, it is highly likely that autoinhibition in IC is a conserved process for modulating binding to p150^{Glued} and NudE. In Drosophila IC, paramagnetic relaxation enhancement NMR experiments have shown that binding of NudE to the SAH region shifts IC equilibrium toward more open states[91]. Such a preference for more open states when NudE is bound is in line with our data for CT

showing that $NudE_{CC}$ requires an open SAH region to bind, whereas $p150_{CC1B}$ can more easily overcome the autoinhibited (closed) conformation of IC. Similarly, with IC from mammalian species, autoinhibition by interaction between the SAH and H2 regions appears to be modulated by phosphorylation[92,169]. Our data collected on the full N-terminal domain of IC and full-length IC protein from CT demonstrate subcomplex assembly via light chain binding as the primary modulating mechanism for $p150^{Glued}$ and NudE binding. This is an effect that can only be detected with larger constructs of IC, such as CT $IC_{1\text{-}260}$, that contains long disordered linkers separating short helices and binding sites. As summarized by the model shown in Figure 10, we propose that these properties combine to create a multifaceted system of regulation for IC. The disordered linkers allow for the needed flexibility to bring together the SAH region and the C-terminus (200+ amino acids away) and to extend it into an open conformation when LC7 is bound (either on its own, or with the other two light chains), which suggests a unique role for LC7 in regulating IC interactions.

In conclusion, this work shows, for the first time, the recombinant expression of multiple dynein and non-dynein subunits and their reconstitution into three assembled subcomplexes. The dynein subcomplex (IC/Tctex/LC8/LC7) as well as the dynein subcomplex with $p150_{CC1B}$ or with $NudE_{CC}$ were each successfully re-purified by SEC (Figure 5.8A) and verified by SDS-PAGE to show each expected subunit present in the complex (Figure 5.8B). Both the shape and the symmetry of the eluting SEC peak for the dynein subcomplex indicate that the complex is less stable than when either $p150_{CC1B}$ or $NudE_{CC}$ is added. Taken together, these data demonstrate that not only is the stability of the dynein subcomplex enhanced by addition of $p150_{CC1B}$ and $NudE_{CC}$, but also that the ability of $p150_{CC1B}$ and $NudE_{CC}$ to bind to IC is no longer inhibited when the light chains are present. The same conclusion is supported by data with full-length IC, as IC_{FL} alone is unable to bind to either $p150_{CC1B}$ or $NudE_{CC}$ until light chains are bound (Figure 5.9). Interestingly, the binding of just LC7 is enough to rescue IC_{FL} binding to $p150_{CC1B}$ and $NudE_{CC}$, while binding of Tctex and LC8 only relieves autoinhibition enough for $p150_{CC1B}$ to bind

(Figure 5.9B). These results speak to the complexity required in IC regulation and emphasize the roles of autoinhibition and multivalency in this process.

As illustrated in the model in Figure 10, the C-terminal end of IC_{1-260}, which contains the LC7 binding site, interacts with the SAH region and results in a closed conformation. The closed conformation of IC is autoinhibitory, preventing binding of the SAH region by $p150_{CC1B}$ or $NudE_{CC}$. However, because $p150_{CC1B}$ also binds to the H2 region, $p150_{CC1B}$ binding to IC_{1-260} is not completely abolished. Our data suggest that it is the segment containing residues 220-250 (which forms an α-helix upon binding LC7) that makes contacts with the SAH, leaving downstream residues able to initiate binding to LC7 (Figures 5.5, 5.9). From the limited changes in NMR spectra upon Tctex and LC8 binding, we conclude that the overall closed conformation of IC is likely unrelieved by binding of LC8 and/or Tctex since they bind in the center of a long linker, whereas LC7 binding is required to release the SAH region, thereby allowing assembled IC to fully bind to $p150_{CC1B}$ and $NudE_{CC}$.

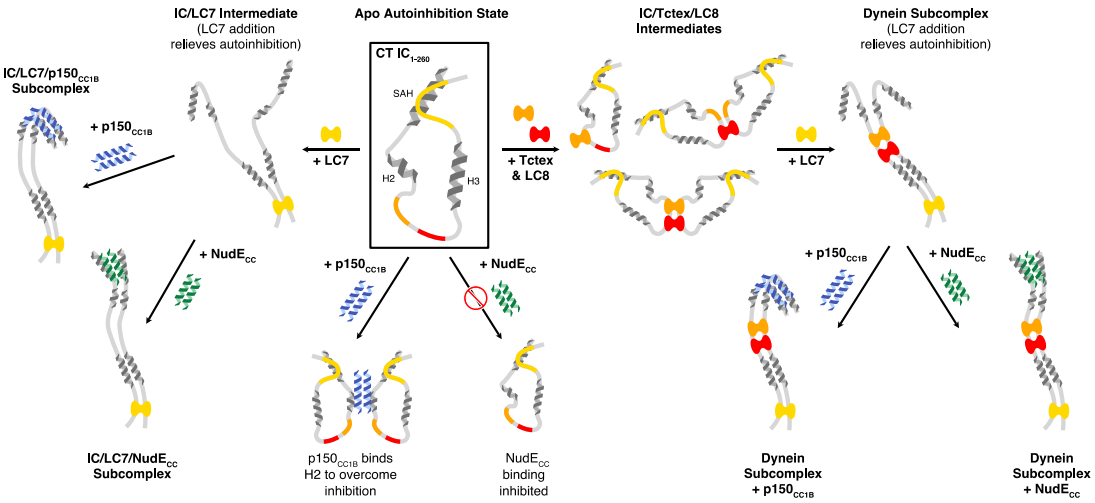

Figure 5.10. A model of CT IC_{1-260} binding interactions and subcomplex assemblies. Apo CT IC_{1-260} is compact and in autoinhibited state (Boxed in black), with the SAH, H2, H3 regions depicted as helices and with colors indicating the LC8 (red), Tctex (orange), and LC7 (yellow) binding sites. When LC7 is added (left arrow) LC7 outcompetes autoinhibition to bind IC_{1-260}, exposing the SAH domain for $p150_{CC1B}$ and NudE binding. When $p150_{CC1B}$ or $NudE_{CC}$ are added (down arrows) to apo IC_{1-260}, autoinhibition prevents $NudE_{CC}$ from binding and reduces the binding affinity of $p150_{CC1B}$. However, as $p150_{CC1B}$ is able to bind to the H2 region of IC_{1-260}, binding is not completely prevented. Addition of Tctex and/or LC8 (right arrow)

leads to a number of possible binary and ternary intermediates. LC8's role of driving IC dimerization is depicted and, in all intermediates, we predict that IC autoinhibition remains based on very limited changes in NMR spectra. Continuing to the right, the addition of LC7 leads to formation of the dynein subcomplex and the release of SAH autoinhibition. The free SAH is now able to resume transient interactions with H2 prior to binding with either $p150_{CC1B}$ or $NudE_{CC}$. Finally, addition of $p150_{CC1B}$ or $NudE_{CC}$ leads to the formation of the p150 and NudE subcomplexes (bottom right). Our SV-AUC data suggest that the NudE subcomplex adopts a more elongated conformation.

Dynein autoinhibition has been a relevant topic for years as dynein is only weakly processive when isolated in vitro[170], and requires the binding of dynactin and/or cargo adaptor proteins for activation[96,171–173]. In a recent study, cryo-EM structures revealed that the autoinhibited form of human cytoplasmic dynein (phi-particle) is stabilized by motor domain self-dimerization and contact between the tails of the heavy chains, whereas the less inhibited form (open-dynein) formed after binding dynactin has the proper motor domain orientation for highly processive motion along the microtubules[82]. Due to the binding relationship between the N-terminus of IC and the CC1B region of the $p150^{Glued}$ subunit of dynactin, there is the possibility that the IC/$p150^{Glued}$ interaction provides the first step in dynein activation. If so, it would stand to reason that IC/$p150^{Glued}$ autoinhibition regulation we identify here in CT is a critical step in dynein function. In future work it will be important to demonstrate that autoinhibition and its subsequent reversal by dynein light chain binding is a conserved process of IC across species and whether this plays a role in the autoinhibition of the overall dynein protein complex.

Materials and Methods

Cloning, Protein Expression and Purification

All studies were carried out using constructs from *Chaetomium thermophilum* (thermophilic fungus, G0SCF1-1). IC_{1-260} (res. 1-260), $p150_{CC1B}$ (res. 478-680), and $NudE_{CC}$ (res. 1-190) constructs were prepared by PCR and cloned into a pET-24d vector with an N-terminal 6×His tag using the Gibson Assembly protocol[174,175]. In addition, a fragment of the IC_{1-260} construct with N-terminal residues removed (IC_{100-}

$_{260}$) was generated by the same method. $IC_{160-240}$ was ordered from GenScript (Piscataway, NJ) and $IC_{216-260}$ was ordered from Azenta Life Sciences (Chelmsford, MA). Full length CT Tctex, LC8, and LC7 were amplified out of a CT cDNA library and cloned into a pET-15b vector with an uncleavable C-terminal 6×His tag. These light chain constructs contain a single, non-native GS linker prior to the tag sequence. For IC, p150, and NudE constructs, an N-terminal tobacco etch virus (TEV) protease cleavage site was included to allow the removal of the 6×His tag, leaving a non-native GAH sequence post cleavage. DNA sequences were verified by Sanger sequencing. IC_{1-88} and IC_{37-88} constructs were prepared previously[97].

Recombinant plasmids were transformed into Rosetta (DE3) *E. coli* cells (Merck KGaA, Darmstadt, Germany) for protein expression. Bacterial cultures for expression of unlabeled proteins were grown in ZYM-5052 autoinduction media at 37ºC for 24 hrs[176], whereas cultures for expression of isotopically-labeled (^{15}N or $^{15}N/^{13}C$) proteins were grown in MJ9 minimal media[177] at 37ºC to an OD_{600} of 0.8 before being induced with 0.4 mM isopropyl-β-D-1-thiogalactopyranoside (IPTG), and continuing expression overnight at 26ºC. Proteins were purified from the cell cultures by immobilized metal affinity chromatography using previously published methods[97]. Complete cleavage of the tag by tobacco etch virus protease was verified by SDS-PAGE analysis.

Proteins were further purified using a Superdex 75 (Cytiva Life Sciences, Pittsburgh, PA) SEC column and then, for IC_{1-260} samples, followed by anion exchange using Macro-Prep High Q Support resin (Bio-Rad, Hercules, California) with elution in 0.1 to 0.2 M sodium chloride. Protein concentrations were determined from absorbance at 205 and 280 nm [124]. Molar extinction coefficients for the constructs used are as follows (ε_{205} & ε_{280}): IC_{1-260} = 853,230 & 11,460 $M^{-1}\,cm^{-1}$, IC_{1-150} = 488,290 & 8,480 $M^{-1}\,cm^{-1}$, $IC_{100-260}$ = 665,620 & 14,440 $M^{-1}\,cm^{-1}$, $IC_{160-240}$ = 387,110 & 2,980 $M^{-1}\,cm^{-1}$, $p150_{CC1B}$ = 685,820 & 9,970 $M^{-1}\,cm^{-1}$, $NudE_{CC}$ = 665,800 & 16,960 $M^{-1}\,cm^{-1}$, Tctex = 649,850 & 31,970 $M^{-1}\,cm^{-1}$, LC8 = 495,640 & 8,480 $M^{-1}\,cm^{-1}$, and LC7 = 533,510 & 6,990 $M^{-1}\,cm^{-1}$. All purified proteins were

stored at 4ºC with a protease inhibitor mixture of pepstatin A and phenylmethanesulfonyl fluoride and used within one week.

Full Length IC Cloning and Expression

Full length CT IC (IC_{FL}) was expressed in Insect Sf9 cells. The sequence of IC_{FL} was codon optimized and cloned into pFastbac1 vector by Genscript (Piscataway, NJ) with an N-terminal 6×His tag followed by a TEV protease cleavage site. IC_{FL} was expressed in Sf9 cells using the multiBAC system following previously published protocols with slight modifications[82,172]. The plasmid was transformed into EmBacY cells (Multibac) and single, white colonies were selected after 2 days and inoculated into 2xTY media supplemented with antibiotics and grown overnight. The bacteria pellet was harvested at 4000 rpm for 10 min and resuspended in 0.3 mL QIAGEN miniprep buffer P1, followed by 0.3 mL P2 buffer. After 5 min incubation, 0.4 mL P3 buffer was added and the mixture was centrifuged at 13,000 rpm for 10 min. The supernatant was added to 0.8 mL ice cold isopropanol and bacmid DNA was pelleted for 10 min at 13,000 rpm. The pellet was washed twice with 70% ethanol, resuspended in H_2O, and stored at 4°C.

Sf9 cells were cultured in SF-900 III SFM (Thermo Fisher) at a shaking speed of 125 rpm at 27°C. 2 μg fresh bacmid DNA was transfected into 2 mL Sf9 cells at 0.9×10^6 cells/2 mL with 8 μL Cellfection II (Thermo Fisher) following manufacturer's protocol. Five days later, 0.5 mL of the transfected culture medium was added to a 50 mL culture of Sf9 cells (0.5×10^6 cells/mL) for P2 infection. Four days later, cells were spun down at 3000 rpm for 15 min at 4°C and the supernatant of P2 virus was collected and stored at 4°C in the dark. Protein expression was induced by adding P2 virus to Sf9 cells (1:100 ratio, 2×10^6 cells/mL). After four days, cells were harvested by centrifugation at 4000 rpm for 20 min at 4°C. The pellet was flash frozen in liquid nitrogen and stored at −80°C for further protein purification. Affinity purification was carried out as described above for other constructs.

IC Structure Prediction

Sequences for IC from a range of species were obtained from the UniProt protein database[178]. *Rattus norvegicus* (rat, UniProt: Q62871-3), *Drosophila melanogaster* (fruit fly, UniProt: Q24246-11), *Saccharomyces cerevisiae* (yeast, UniProt: P40960-1), *Homo sapiens* (human, UniProt: O14576-2), *Danio rerio* (zebrafish,UniProt: A1A5Y4-1), *Callorhinchus milii* (Australian ghost shark, UniProt:V9KAN3-1), *Octopus bimaculoides* (Californian two-spot octopus, UniProt:A0A0L8HM30-1), *Caenorhabditis elegans* (nematode, UniProt: O45087-1), and *Chaetomium thermophilum* (CT, thermophilic fungus, UniProt: G0SCF1-1). The first 260 amino acids of the IC from each species were scored using the Agadir algorithm, which outputs a prediction for percent helicity per residue[108,109,164–166]. Agadir is not developed for prediction of long protein sequences, therefore results were compared to predictions from PSIPRED[106,163] and in all cases agreed well.

Circular Dichroism

CD measurements were made using a JASCO (Easton, Maryland) J-720 circular dichroism spectropolarimeter. Samples consisted of proteins at concentrations of 5-10 μM in 10 mM sodium phosphate buffer (pH 7.5). All experiments were done using a 400 μL cuvette with a path length of 0.1 cm. The data shown are the average of three scans. Thermal unfolding data were collected in increments of 5ºC over a temperature range of 5 to 60ºC. CD measurements were used to estimate the fractional helicity of the samples using the equation below:

$$\text{Fractional Helicity} = \frac{(\theta_{222}^{\text{exp}} + 3000 \text{ deg cm}^2 \text{ dmol}^{-1})}{-36{,}500 \text{ deg cm}^2 \text{ dmol}^{-1}}$$

where $\theta_{222}^{\text{exp}}$ is the experimentally observed residue ellipticity (MRE) at 222 nm[167].

Isothermal Titration Calorimetry

ITC experiments were conducted using a MicroCal VP-ITC microcalorimeter (Malvern Panalytical, United Kingdom). All experiments were performed at 25ºC and with protein samples in a buffer containing 50 mM sodium phosphate (pH 7.5), 50

mM sodium chloride, and 1 mM sodium azide. Samples were degassed at 25ºC prior to loading. Each experiment was started with a 2 µL injection, followed by 27 to 33 injections of 10 µL. Protein concentrations in the cell ranged from 20-40 µM and concentrations in the syringe ranged from 200-400 µM. The data were processed using Origin 7.0 (Malvern Panalytical, Malvern, UK) and fit to a single-site binding model. The recorded data are the averages of 2-3 independent experiments.

NMR measurements and analysis

Samples for NMR were prepared in a 20 mM sodium chloride, 50 mM phosphate (pH 7.4) buffer that included 5% D_2O, 1 mM sodium azide, 0.2 mM 2,2-dimethylsilapentane-5-sulfonic acid (DSS), and 1× cOmplete™ protease inhibitor cocktail (Roche, Basel, Switzerland). NMR spectra were collected over a temperature range of 10-40 ºC using a Avance III HD 800 MHz spectrometer (Bruker Biospin, Billerica, Massachusetts) with a TCI cryoprobe and a Avance NEO 600 MHz spectrometer (Bruker Biospin) equipped with a room-temperature TXI probe. Band-selective excitation short transient (BEST) variants of TROSY-based triple resonance sequences (HNCO, HNCA, HN(CO)CA, HN(CA)CO, HNCACB, HN(CO)CACB were used for backbone assignment of IC_{1-260} at 10ºC [126]. Assignments of ^{15}N-labeled $IC_{216-260}$ was carried out using ^{15}N-seperated TOCSY and NOESY experiments. The accessibility of IC_{1-260} amide protons to exchange with the solvent was determined by measuring peak volumes in a Fast HSQC spectrum with a 20 ms CLEANEX-PM mixing period. The IC_{1-260} and $IC_{216-260}$ concentrations were 350-500 µM for these samples.

NMR experiments of binary complexes (IC with one binding partner) were performed by combining ^{15}N-labeled IC_{1-260} with unlabeled $p150_{CC1B}$, $NudE_{CC}$, Tctex, LC8, or LC7 at a molar ratio of 1:1.5. For the $IC_{216-260}$/$IC_{216-260}$ binary complex, the ^{15}N-labeled components were mixed in a 1:1 molar ratio and each was at a concentration of 350 µM. Spectra for each binary complex were collected at both 10 and 40ºC. NMR experiments of the dynein subcomplex (IC_{1-260} with all light chains) were performed much in the same way. The IC_{1-260} concentration was 250 µM for

these samples. NMR data were processed using TopSpin 3.6 (Bruker) and NMRPipe[118]. For 3D experiments that employed non-uniform sampling the spectra were reconstructed using SCRUB[121]. Peak assignment was performed using CCPN Analysis 2.5.2[119].

Analytical ultracentrifugation

Sedimentation velocity analytical ultracentrifugation (SV-AUC) experiments were performed using a Beckman Coulter Optima XL-A analytical ultracentrifuge, equipped with absorbance optics. For individual proteins, the concentration used was 15-30 μM. For binary complexes, $IC_{1\text{-}260}$ or IC_{FL} were mixed with each binding partner at ratios of 1:1.5 or 1:2 (molar ratio of IC to binding partner). The buffer condition used for all SV-AUC experiments was 25 mM tris(hydroxymethyl)aminomethane hydrochloride (pH 7.4), 150 mM KCl, 5 mM tris (2-carboxyethyl) phosphine, and 1 mM sodium azide. Samples were loaded into standard, 12 mm pathlength, 2-channel sectored centerpieces and centrifuged at 42,000 rpm and 20°C. 300 scans were acquired at 280-297 nm with no interscan delay. Data were fit to a c(S) distribution using SEDFIT[146].

Sedimentation equilibrium analytical ultracentrifugation (SE-AUC) experiments were performed on the same instrument. $NudE_{CC}$ and $p150_{CC1B}$ were each loaded with three concentrations in the range of 15-60 μM in 6 channel centerpieces, centrifuged at three speeds (10,000, 14,000, and 18,000 rpm) and scanned at 280 nm. Samples were scanned every 3 hours until they were at equilibrium (i.e., when the final sequential scans were superimposable), which occurred after 30 to 36 hours of centrifugation. Data were acquired as averages of five measurements of absorbance at each radial position, with a nominal spacing of 0.003 cm between each position. The data from the three speeds and three concentrations were globally fit to a monomer-dimer self-association model and resulted in random residuals. Other models tested did not give adequate variances and random residuals. All experiments were done at 20°C. Data were fit using HETEROANALYSIS[179].

SEC multiangle light scattering

SEC coupled to multiangle light scattering (SEC-MALS) was carried out using a Superdex 200 gel filtration column on an AKTA fast liquid chromatography system (Cytiva Life Sciences) coupled with a DAWN multiple-angle light scattering detector and an Optilab refractive index detector (Wyatt Technology, Santa Barbara, CA). Data for IC_{1-260} was collected for protein samples at a concentration of 200 µM protein in a buffer composed of 50 mM sodium phosphate (pH 7.5), 50 mM sodium chloride, and 1 mM sodium azide. Data for IC_{FL} was collected for protein samples at a concentration of 30 µM in a buffer composed of 25 mM tris(hydroxymethyl)aminomethane hydrochloride (pH 7.4), 150 mM KCl, 5 mM β-mercaptoethanol, and 1 mM sodium azide. Molar mass and error analysis were determined using ASTRA v9, employing a Zimm light scattering model (Wyatt Technology).

Subcomplex reconstitution

Post purification of individual protein, IC_{1-260} was combined with dynein light chains (Tctex, LC8, and LC7). To the assembled subcomplex, the non-dynein proteins $p150_{CC1B}$ or $NudE_{CC}$ were added in a 1:1.5 molar ratio prior to SEC using a Superdex 200 column (Cytiva Life Sciences). Subcomplex reconstitution was verified by SDS-PAGE of SEC fractions, as each protein clearly resolves. When estimating concentrations for complexes purified by SEC, the absorbance at 280 nm was used with the assumption that the majority of formed complex in solution followed the expected stoichiometry of 1:1 (IC monomer: partner monomers) and that very little excess of free protein would be present.

Acknowledgements

We acknowledge support from the National Science Foundation (Award 1617019 for E.J.B. and Award 2003557 for N.M.L.). The Oregon State University NMR Facility is funded in part by the National Institutes of Health (HEI Grant 1S10OD018518) and by the M. J. Murdock Charitable Trust (Grant 2014162). The Lewis & Clark College Bruker Avance NEO 600 NMR spectrometer was purchased with support from the National Science Foundation (Award 1917696) and the M. J. Murdock Charitable Trust (Grant 201811283).

Chapter 6

Conclusion

Impact

The studies presented here provide detailed description of two multivalent and disordered protein assemblies involving the hub protein, LC8. In combination, the work of this thesis serves as a blueprint for studying these systems in context. Beginning with a guide in using NMR to probe interactions of intrinsically disordered proteins, I present our routinely used methodology in a simple, easy-to-follow format that could be of great use to those interested in implementing NMR into their own research. As IDPs continue to prove important in the regulation of numerous cellular pathways and functions, continuing to make NMR approachable for novice users has the possibility for broad and significant impact across biological fields. The chapter concludes with an overview of how to couple NMR results with various other biophysical techniques depending on the desired research outcome, an approach that is proven vital in our work with the QT2-4 subdomain of ASCIZ and intermediate chain subcomplexes of dynein. By eloquently incorporating data from ITC, AUC, SEC-MALS, and ESI-MS in addition to NMR, the IDP interactions in the systems mentioned are well characterized and gaps in the respective fields are addressed. Central especially to the regulation mechanisms presented are not only the methods used but also the context encompassed in our studies. Our work with QT2-4 is the first in-depth look at the interplay of linker length, motif affinity, and motif specificity in LC8/IDP complex assembly that laid the groundwork for our success with the N-terminal domain and full-length constructs of dynein IC. The remainder of this chapter recounts the breakthroughs of my work and proposes future studies to continue much-needed progress in the IDP and dynein fields.

Highlights of Reported Work

Chapter 2 provides a procedural overview for studying the interactions of protein complexes that contain flexibility and disorder. The basics of protein expression, purification, sample preparation, and NMR data acquisition are presented

while acknowledging the need to use multiple biophysical techniques to properly probe such complex systems. This chapter provides an overview of the complications involved when using traditional structural techniques to study disordered proteins, while highlighting the advancements made by using NMR, specifically by our group for dynein IC.

In Chapters 3 and 4, we explore important mechanistic questions about LC8 and multivalent IDP binding. Using QT2-4 as a model multivalent LC8 ligand we provide the first evidence of "in-register" binding during complex assembly and identify linker length as having a modulating role in the flexibility and LC8 occupancy in multivalent LC8/IDP complexes. Longer disordered linkers introduce increased flexibility between regions when bound to LC8, resulting in both compositional and conformational heterogeneity, that deters proper LC8/IDP duplex scaffold formation. Conversely, we show that shorter linkers can lead to enhanced binding of nearby weaker sites, reducing flexibility in the complex and promoting the formation of stable duplexes. Additionally, we show that binding of LC8 to multivalent QT2-4 constructs is complex and governed by factors such as LC8 motif specificity and that cooperativity between multivalent sites can be positive or negative depending on site location and individual LC8 binding affinity. The interplay of linker length and motif affinity and specificity, is a promising emerging mechanism for a tunable system of binding and regulation that we anticipate being shared across many biological processes, given the growing number of multivalent LC8 client proteins and the essential role of LC8 in nearly all cellular functions.

In Chapter 5, we preformed, for the first time, the recombinant expression of multiple dynein and non-dynein subunits and their reconstitution into assembled subcomplexes. Via successful re-purification of the dynein subcomplex (IC/Tctex/LC8/LC7) and subcomplex variations including either $p150_{CC1B}$ or $NudE_{CC}$, we determined that the addition of non-dynein partners increase the subcomplex stability. We also show that $p150_{CC1B}$ and $NudE_{CC}$ experience inhibited binding to both the N-terminal domain (IC_{1-260}) and full-length IC (IC_{FL}) due to long range intramolecular interactions within IC between the N-terminal SAH domain and

the LC7 binding site that impart a closed conformation of IC. IC autoinhibition is released by the binding of the three dynein light chains and interestingly, the binding of just LC7 is also enough to rescue IC_{FL} binding to $p150_{CC1B}$ and $NudE_{CC}$.

Further intricacies regarding the IC autoinhibition mechanism described include a selection process for $p150_{CC1B}$ over $NudE_{CC}$ via the H2 domain. Because $p150_{CC1B}$ binds H2 in addition to the SAH on IC, its binding is less inhibited. Fittingly so, IC_{FL} binding to Tctex and LC8 alone relieves autoinhibition enough for $p150_{CC1B}$ to fully bind, while inhibition is maintained to $NudE_{CC}$. These results not only speak to the complexity required in IC regulation but further emphasize the roles of autoinhibition and multivalency in the process.

These important findings were made possible by the context of our study, the only of its kind. Prior work on IC binding interactions has been limited to only short fragments of N-IC, leaving a mechanism for IC partner selection and the importance of bivalency in IC subcomplex assembly as major gaps in the field. Our studies utilizing IC_{1-260} and IC_{FL} allowed for crucial insight into filling these gaps and has paved the way for future studies on IC and dynein regulation.

Future Work

LC8/IDP Complexes

Many studies have already begun within the Barbar lab that will serve as future work to my studies with QT2-4 and LC8. The positive and negative allostery exhibited by QT2-4 will be further characterized with NMR and negative stain EM to quantitatively determine how each binding site is affected by the others and eventually, whether the effects are retained in a cellular environment. Sources of binding heterogeneity in *Drosophila* ASCIZ will also be investigated with other constructs (QT1-3, QT4-6, and QT4-7) and mutants thereof as well and the full-length LC8 binding domain to better create ‘rules’ around LC8/IDP assembly formation. Because it is such an important transcription factor and multivalent partner of LC8, ASCIZ serves as an ideal system for investigating LC8 multivalent binding.

However, even beyond LC8 and ASCIZ, this work would provide insights that extend to other transcription factors, hub proteins, and their partners.

Outside the context of ASCIZ, other lab projects include the tumor suppressor P53-binding protein (53BP1) that forms a unique assembly with LC8 and may expand our known functional and structural categories for multivalent LC8/IDP complexes. Also, we continue to make progress in our efforts to mathematically model LC8 binding interactions that would provide important kinetic and thermodynamic insights for LC8's numerous ligands and the many systems in which they involve.

Dynein IC

Future work regarding dynein IC can be further divided into the aims below:

1) **Further characterizing autoinhibition as a mechanism for regulation of dynein interactions.**

As we have already shown that autoinhibition effects $p150^{Glued}$ and NudE binding in CT IC_{1-260} and IC_{FL}, a next step would be to determine if the same mechanism of autoinhibition is seen across species of IC and if other regulatory mechanisms, such as phosphorylation, impact this autoinhibition. Prior data on mammalian versions of IC suggest a role for phosphorylation in regulating binding to $p150^{Glued}$, however these studies were done on smaller IC segments. Future studies should use the full-length N-domain and genetic code expansion (GCE) to incorporate phosphoSer at previously identified sites to determine if autoinhibition is observed in mammalian IC and to compare the effects on binding to $p150^{Glued}$ to the shorter domain.

2) **Determining the structure of the dynein IC subcomplexes.**

Our current NMR studies of CT IC_{1-260} could possibly be expanded by incorporating paramagnetic relaxation enhancement, and sortase-mediated segmental labeling methods to further determine the structure of free and light chain bound IC. Additionally, negative stain and cryoEM methods should be explored, with CT IC_{1-260} and IC_{FL}, with the addition of p150 and NudE if tighter, more stable complexes are needed. These studies could result in the first structures of IC subcomplexes and determine conformational differences between p150 and NudE subcomplexes.

References

(1) Uversky, V. N. Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. *Frontiers in Physics* **2019**, *7*.

(2) Xue, B.; Dunker, A. K.; Uversky, V. N. Orderly Order in Protein Intrinsic Disorder Distribution: Disorder in 3500 Proteomes from Viruses and the Three Domains of Life. *Journal of Biomolecular Structure and Dynamics* **2012**, *30* (2), 137–149. https://doi.org/10.1080/07391102.2012.675145.

(3) Peng, Z.; Yan, J.; Fan, X.; Mizianty, M. J.; Xue, B.; Wang, K.; Hu, G.; Uversky, V. N.; Kurgan, L. Exceptionally Abundant Exceptions: Comprehensive Characterization of Intrinsic Disorder in All Domains of Life. *Cell. Mol. Life Sci.* **2015**, *72* (1), 137–151. https://doi.org/10.1007/s00018-014-1661-9.

(4) Ward, J. J.; Sodhi, J. S.; McGuffin, L. J.; Buxton, B. F.; Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *Journal of Molecular Biology* **2004**, *337* (3), 635–645. https://doi.org/10.1016/j.jmb.2004.02.002.

(5) Uversky, V. N. The Mysterious Unfoldome: Structureless, Underappreciated, Yet Vital Part of Any Given Proteome. *Journal of Biomedicine and Biotechnology* **2009**, *2010*, e568068. https://doi.org/10.1155/2010/568068.

(6) Dunker, A. K.; Lawson, J. D.; Brown, C. J.; Williams, R. M.; Romero, P.; Oh, J. S.; Oldfield, C. J.; Campen, A. M.; Ratliff, C. M.; Hipps, K. W.; Ausio, J.; Nissen, M. S.; Reeves, R.; Kang, C.; Kissinger, C. R.; Bailey, R. W.; Griswold, M. D.; Chiu, W.; Garner, E. C.; Obradovic, Z. Intrinsically Disordered Protein. *J Mol Graph Model* **2001**, *19* (1), 26–59. https://doi.org/10.1016/s1093-3263(00)00138-8.

(7) DeForte, S.; Uversky, V. N. Not an Exception to the Rule: The Functional Significance of Intrinsically Disordered Protein Regions in Enzymes. *Mol. BioSyst.* **2017**, *13* (3), 463–469. https://doi.org/10.1039/C6MB00741D.

(8) van der Lee, R.; Buljan, M.; Lang, B.; Weatheritt, R. J.; Daughdrill, G. W.; Dunker, A. K.; Fuxreiter, M.; Gough, J.; Gsponer, J.; Jones, D. T.; Kim, P. M.; Kriwacki, R. W.; Oldfield, C. J.; Pappu, R. V.; Tompa, P.; Uversky, V. N.; Wright, P. E.; Babu, M. M. Classification of Intrinsically Disordered Regions and Proteins. *Chem Rev* **2014**, *114* (13), 6589–6631. https://doi.org/10.1021/cr400525m.

(9) Brown, C. J.; Johnson, A. K.; Dunker, A. K.; Daughdrill, G. W. Evolution and Disorder. *Curr Opin Struct Biol* **2011**, *21* (3), 441–446. https://doi.org/10.1016/j.sbi.2011.02.005.

(10) Tantos, A.; Han, K.-H.; Tompa, P. Intrinsic Disorder in Cell Signaling and Gene Transcription. *Mol Cell Endocrinol* **2012**, *348* (2), 457–465. https://doi.org/10.1016/j.mce.2011.07.015.

(11) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298* (5600), 1912–1934. https://doi.org/10.1126/science.1075762.

(12) Cortese, M. S.; Uversky, V. N.; Dunker, A. K. Intrinsic Disorder in Scaffold Proteins: Getting More from Less. *Prog Biophys Mol Biol* **2008**, *98* (1), 85–106. https://doi.org/10.1016/j.pbiomolbio.2008.05.007.

(13) Alonso, A. C.; Grundke-Iqbal, I.; Iqbal, K. Alzheimer's Disease Hyperphosphorylated Tau Sequesters Normal Tau into Tangles of Filaments and Disassembles Microtubules. *Nat Med* **1996**, *2* (7), 783–787. https://doi.org/10.1038/nm0796-783.

(14) Mark, W.-Y.; Liao, J. C. C.; Lu, Y.; Ayed, A.; Laister, R.; Szymczyna, B.; Chakrabartty, A.; Arrowsmith, C. H. Characterization of Segments from the Central Region of BRCA1: An Intrinsically Disordered Scaffold for Multiple Protein-Protein and Protein-DNA Interactions? *J Mol Biol* **2005**, *345* (2), 275–287. https://doi.org/10.1016/j.jmb.2004.10.045.

(15) Tompa, P.; Fuxreiter, M. Fuzzy Complexes: Polymorphism and Structural Disorder in Protein-Protein Interactions. *Trends Biochem Sci* **2008**, *33* (1), 2–8. https://doi.org/10.1016/j.tibs.2007.10.003.

(16) Barbar, E.; Nyarko, A. Polybivalency and Disordered Proteins in Ordering Macromolecular Assemblies. *Semin Cell Dev Biol* **2015**, *37*, 20–25. https://doi.org/10.1016/j.semcdb.2014.09.016.

(17) Clark, S. A.; Jespersen, N.; Woodward, C.; Barbar, E. Multivalent IDP Assemblies: Unique Properties of LC8-Associated, IDP Duplex Scaffolds. *FEBS Lett* **2015**, *589* (19 Pt A), 2543–2551. https://doi.org/10.1016/j.febslet.2015.07.032.

(18) Forsythe, H. M.; Barbar, E. Chapter Seven - The Role of Dancing Duplexes in Biology and Disease. In *Progress in Molecular Biology and Translational Science*; Uversky, V. N., Ed.; Dancing Protein Clouds: Intrinsically Disordered Proteins in the Norm and Pathology, Part C; Academic Press, 2021; Vol. 183, pp 249–270. https://doi.org/10.1016/bs.pmbts.2021.06.004.

(19) Oldfield, C. J.; Cheng, Y.; Cortese, M. S.; Romero, P.; Uversky, V. N.; Dunker, A. K. Coupled Folding and Binding with Alpha-Helix-Forming Molecular Recognition Elements. *Biochemistry* **2005**, *44* (37), 12454–12470. https://doi.org/10.1021/bi050736e.

(20) Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein. *Nature* **2007**, *447* (7147), 1021–1025. https://doi.org/10.1038/nature05858.

(21) Shammas, S. L.; Travis, A. J.; Clarke, J. Remarkably Fast Coupled Folding and Binding of the Intrinsically Disordered Transactivation Domain of CMyb to CBP KIX. *J. Phys. Chem. B* **2013**, *117* (42), 13346–13356. https://doi.org/10.1021/jp404267e.

(22) Rautureau, G. J. P.; Day, C. L.; Hinds, M. G. Intrinsically Disordered Proteins in Bcl-2 Regulated Apoptosis. *Int J Mol Sci* **2010**, *11* (4), 1808–1824. https://doi.org/10.3390/ijms11041808.

(23) Dyson, H. J.; Wright, P. E. Coupling of Folding and Binding for Unstructured Proteins. *Curr Opin Struct Biol* **2002**, *12* (1), 54–60. https://doi.org/10.1016/s0959-440x(02)00289-0.

(24) Gunasekaran, K.; Tsai, C.-J.; Kumar, S.; Zanuy, D.; Nussinov, R. Extended Disordered Proteins: Targeting Function with Less Scaffold. *Trends Biochem Sci* **2003**, *28* (2), 81–85. https://doi.org/10.1016/S0968-0004(03)00003-3.

(25) Balázs, A.; Csizmok, V.; Buday, L.; Rakács, M.; Kiss, R.; Bokor, M.; Udupa, R.; Tompa, K.; Tompa, P. High Levels of Structural Disorder in Scaffold Proteins as Exemplified by a Novel Neuronal Protein, CASK-Interactive Protein1. *FEBS J* **2009**, *276* (14), 3744–3756. https://doi.org/10.1111/j.1742-4658.2009.07090.x.

(26) Babu, M. M.; van der Lee, R.; de Groot, N. S.; Gsponer, J. Intrinsically Disordered Proteins: Regulation and Disease. *Curr Opin Struct Biol* **2011**, *21* (3), 432–440. https://doi.org/10.1016/j.sbi.2011.03.011.

(27) Allen, N. P. C.; Patel, S. S.; Huang, L.; Chalkley, R. J.; Burlingame, A.; Lutzmann, M.; Hurt, E. C.; Rexach, M. Deciphering Networks of Protein Interactions at the Nuclear Pore Complex. *Mol Cell Proteomics* **2002**, *1* (12), 930–946. https://doi.org/10.1074/mcp.t200012-mcp200.

(28) Burke, M. G.; Woscholski, R.; Yaliraki, S. N. Differential Hydrophobicity Drives Self-Assembly in Huntington's Disease. *Proc Natl Acad Sci U S A* **2003**, *100* (24), 13928–13933. https://doi.org/10.1073/pnas.1936025100.

(29) Brangwynne, C. P.; Eckmann, C. R.; Courson, D. S.; Rybarska, A.; Hoege, C.; Gharakhani, J.; Jülicher, F.; Hyman, A. A. Germline P Granules Are Liquid Droplets That Localize by Controlled Dissolution/Condensation. *Science* **2009**, *324* (5935), 1729–1732. https://doi.org/10.1126/science.1172046.

(30) Mittag, T.; Orlicky, S.; Choy, W.-Y.; Tang, X.; Lin, H.; Sicheri, F.; Kay, L. E.; Tyers, M.; Forman-Kay, J. D. Dynamic Equilibrium Engagement of a Polyvalent Ligand with a Single-Site Receptor. *Proc Natl Acad Sci U S A* **2008**, *105* (46), 17772–17777. https://doi.org/10.1073/pnas.0809222105.

(31) Pfister, K. K.; Fay, R. B.; Witman, G. B. Purification and Polypeptide Composition of Dynein ATPases from Chlamydomonas Flagella. *Cell Motil* **1982**, *2* (6), 525–547. https://doi.org/10.1002/cm.970020604.

(32) Asante, D.; Stevenson, N. L.; Stephens, D. J. Subunit Composition of the Human Cytoplasmic Dynein-2 Complex. *J Cell Sci* **2014**, *127* (Pt 21), 4774–4787. https://doi.org/10.1242/jcs.159038.

(33) King, S. M.; Barbarese, E.; Dillman, J. F.; Patel-King, R. S.; Carson, J. H.; Pfister, K. K. Brain Cytoplasmic and Flagellar Outer Arm Dyneins Share a Highly Conserved Mr 8,000 Light Chain. *J Biol Chem* **1996**, *271* (32), 19358–19366. https://doi.org/10.1074/jbc.271.32.19358.

(34) Barbar, E.; Kleinman, B.; Imhoff, D.; Li, M.; Hays, T. S.; Hare, M. Dimerization and Folding of LC8, a Highly Conserved Light Chain of Cytoplasmic Dynein. *Biochemistry* **2001**, *40* (6), 1596–1605. https://doi.org/10.1021/bi002278+.

(35) Liang, J.; Jaffrey, S. R.; Guo, W.; Snyder, S. H.; Clardy, J. Structure of the PIN/LC8 Dimer with a Bound Peptide. *Nat Struct Biol* **1999**, *6* (8), 735–740. https://doi.org/10.1038/11501.

(36) Bodor, A.; Radnai, L.; Hetényi, C.; Rapali, P.; Láng, A.; Kövér, K. E.; Perczel, A.; Wahlgren, W. Y.; Katona, G.; Nyitray, L. *DYNLL2 Dynein Light Chain Binds to an Extended Linear Motif of Myosin 5a Tail That Has Structural Plasticity*. ACS Publications. https://pubs.acs.org/doi/pdf/10.1021/bi500574z (accessed 2022-05-05). https://doi.org/10.1021/bi500574z.

(37) Slevin, L. K.; Romes, E. M.; Dandulakis, M. G.; Slep, K. C. The Mechanism of Dynein Light Chain LC8-Mediated Oligomerization of the Ana2 Centriole Duplication Factor. *J Biol Chem* **2014**, *289* (30), 20727–20739. https://doi.org/10.1074/jbc.M114.576041.

(38) Rapali, P.; Radnai, L.; Süveges, D.; Harmat, V.; Tölgyesi, F.; Wahlgren, W. Y.; Katona, G.; Nyitray, L.; Pál, G. 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* **2011**, *6* (4), e18818. https://doi.org/10.1371/journal.pone.0018818.

(39) Regué, L.; Sdelci, S.; Bertran, M. T.; Caelles, C.; Reverter, D.; Roig, J. DYNLL/LC8 Protein Controls Signal Transduction through the Nek9/Nek6 Signaling Module by Regulating Nek6 Binding to Nek9. *J Biol Chem* **2011**, *286* (20), 18118–18129. https://doi.org/10.1074/jbc.M110.209080.

(40) Lightcap, C. M.; Kari, G.; Arias-Romero, L. E.; Chernoff, J.; Rodeck, U.; Williams, J. C. Interaction with LC8 Is Required for Pak1 Nuclear Import and Is Indispensable for Zebrafish Development. *PLOS ONE* **2009**, *4* (6), e6025. https://doi.org/10.1371/journal.pone.0006025.

(41) Rao, L.; Romes, E. M.; Nicholas, M. P.; Brenner, S.; Tripathy, A.; Gennerich, A.; Slep, K. C. The Yeast Dynein Dyn2-Pac11 Complex Is a Dynein Dimerization/Processivity Factor: Structural and Single-Molecule Characterization. *Mol Biol Cell* **2013**, *24* (15), 2362–2377. https://doi.org/10.1091/mbc.E13-03-0166.

(42) Benison, G.; Karplus, P. A.; Barbar, E. The Interplay of Ligand Binding and Quaternary Structure in the Diverse Interactions of Dynein Light Chain LC8. *J Mol Biol* **2008**, *384* (4), 954–966. https://doi.org/10.1016/j.jmb.2008.09.083.

(43) Clark, S.; Nyarko, A.; Löhr, F.; Karplus, P. A.; Barbar, E. The Anchored Flexibility Model in LC8 Motif Recognition: Insights from the Chica Complex. *Biochemistry* **2016**, *55* (1), 199–209. https://doi.org/10.1021/acs.biochem.5b01099.

(44) Song, Y.; Benison, G.; Nyarko, A.; Hays, T. S.; Barbar, E. Potential Role for Phosphorylation in Differential Regulation of the Assembly of Dynein Light Chains. *Journal of Biological Chemistry* **2007**, *282* (23), 17272–17279. https://doi.org/10.1074/jbc.M610445200.

(45) Nyarko, A.; Hare, M.; Hays, T. S.; Barbar, E. The Intermediate Chain of Cytoplasmic Dynein Is Partially Disordered and Gains Structure upon Binding to Light-Chain LC8. *Biochemistry* **2004**, *43* (49), 15595–15603. https://doi.org/10.1021/bi048451+.

(46) Wickstead, B.; Gull, K. Dyneins across Eukaryotes: A Comparative Genomic Analysis. *Traffic* **2007**, *8* (12), 1708–1721. https://doi.org/10.1111/j.1600-0854.2007.00646.x.

(47) Pfister, K. K.; Shah, P. R.; Hummerich, H.; Russ, A.; Cotton, J.; Annuar, A. A.; King, S. M.; Fisher, E. M. C. Genetic Analysis of the Cytoplasmic Dynein Subunit Families. *PLOS Genetics* **2006**, *2* (1), e1. https://doi.org/10.1371/journal.pgen.0020001.

(48) King, S. M.; Patel-King, R. S. The M(r) = 8,000 and 11,000 Outer Arm Dynein Light Chains from Chlamydomonas Flagella Have Cytoplasmic Homologues. *J Biol Chem* **1995**, *270* (19), 11445–11452. https://doi.org/10.1074/jbc.270.19.11445.

(49) Jespersen, N.; Estelle, A.; Waugh, N.; Davey, N. E.; Blikstad, C.; Ammon, Y.-C.; Akhmanova, A.; Ivarsson, Y.; Hendrix, D. A.; Barbar, E. Systematic Identification of Recognition Motifs for the Hub Protein LC8. *Life Sci Alliance* **2019**, *2* (4), e201900366. https://doi.org/10.26508/lsa.201900366.

(50) Barbar, E. Dynein Light Chain LC8 Is a Dimerization Hub Essential in Diverse Protein Networks. *Biochemistry* **2008**, *47* (2), 503–508. https://doi.org/10.1021/bi701995m.

(51) Makokha, M.; Hare, M.; Li, M.; Hays, T.; Barbar, E. Interactions of Cytoplasmic Dynein Light Chains Tctex-1 and LC8 with the Intermediate Chain IC74. *Biochemistry* **2002**, *41* (13), 4302–4311. https://doi.org/10.1021/bi011970h.

(52) Nyarko, A.; Barbar, E. Light Chain-Dependent Self-Association of Dynein Intermediate Chain. *J Biol Chem* **2011**, *286* (2), 1556–1566. https://doi.org/10.1074/jbc.M110.171686.

(53) Nyarko, A.; Song, Y.; Nováček, J.; Žídek, L.; Barbar, E. Multiple Recognition Motifs in Nucleoporin Nup159 Provide a Stable and Rigid Nup159-Dyn2 Assembly. *J Biol Chem* **2013**, *288* (4), 2614–2622. https://doi.org/10.1074/jbc.M112.432831.

(54) Raux, H.; Flamand, A.; Blondel, D. Interaction of the Rabies Virus P Protein with the LC8 Dynein Light Chain. *J Virol* **2000**, *74* (21), 10212–10216. https://doi.org/10.1128/jvi.74.21.10212-10216.2000.

(55) Jespersen, N. E.; Leyrat, C.; Gérard, F. C.; Bourhis, J.-M.; Blondel, D.; Jamin, M.; Barbar, E. The LC8-RavP Ensemble Structure Evinces A Role for LC8 in Regulating Lyssavirus Polymerase Functionality. *J Mol Biol* **2019**, *431* (24), 4959–4977. https://doi.org/10.1016/j.jmb.2019.10.011.

(56) Rodriguez Galvan, J.; Donner, B.; Veseley, C. H.; Reardon, P.; Forsythe, H. M.; Howe, J.; Fujimura, G.; Barbar, E. Human Parainfluenza Virus 3 Phosphoprotein Is a Tetramer and Shares Structural and Interaction Features with Ebola Phosphoprotein VP35. *Biomolecules* **2021**, *11* (11), 1603. https://doi.org/10.3390/biom11111603.

(57) Zaytseva, O.; Tenis, N.; Mitchell, N.; Kanno, S. I.; Yasui, A.; Heierhorst, J.; Quinn, L. M. The Novel Zinc Finger Protein DASCIZ Regulates Mitosis in Drosophila via an Essential Role in Dynein Light-Chain Expression. *Genetics* **2014**, *196* (2), 443–453. https://doi.org/10.1534/genetics.113.159541.

(58) Jurado, S.; Conlan, L. A.; Baker, E. K.; Ng, J.-L.; Tenis, N.; Hoch, N. C.; Gleeson, K.; Smeets, M.; Izon, D.; Heierhorst, J. 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. *Journal of Biological Chemistry* **2012**, *287* (5), 3156–3164. https://doi.org/10.1074/jbc.M111.306019.

(59) Clark, S.; Myers, J. B.; King, A.; Fiala, R.; Novacek, J.; Pearce, G.; Heierhorst, J.; Reichow, S. L.; Barbar, E. J. Multivalency Regulates Activity in an Intrinsically Disordered Transcription Factor. *eLife* **2018**, *7*, e40684. https://doi.org/10.7554/eLife.40684.

(60) Reardon, P. N.; Jara, K. A.; Rolland, A. D.; Smith, D. A.; Hoang, H. T. M.; Prell, J. S.; Barbar, E. J. The Dynein Light Chain 8 (LC8) Binds Predominantly "in-Register" to a Multivalent Intrinsically Disordered Partner. *J Biol Chem* **2020**, *295* (15), 4912–4922. https://doi.org/10.1074/jbc.RA119.011653.

(61) King, A.; Hoch, N. C.; McGregor, N. E.; Sims, N. A.; Smyth, I. M.; Heierhorst, J. Dynll1 Is Essential for Development and Promotes Endochondral Bone Formation by Regulating Intraflagellar Dynein Function in Primary Cilia. *Hum Mol Genet* **2019**, *28* (15), 2573–2588. https://doi.org/10.1093/hmg/ddz083.

(62) Benison, G.; Karplus, P. A.; Barbar, E. Structure and Dynamics of LC8 Complexes with KXTQT-Motif Peptides: Swallow and Dynein Intermediate Chain Compete for a Common Site. *J Mol Biol* **2007**, *371* (2), 457–468. https://doi.org/10.1016/j.jmb.2007.05.046.

(63) Wang, L.; Hare, M.; Hays, T. S.; Barbar, E. Dynein Light Chain LC8 Promotes Assembly of the Coiled-Coil Domain of Swallow Protein. *Biochemistry* **2004**, *43* (15), 4611–4620. https://doi.org/10.1021/bi036328x.

(64) Kidane, A. I.; Song, Y.; Nyarko, A.; Hall, J.; Hare, M.; Löhr, F.; Barbar, E. Structural Features of LC8-Induced Self Association of Swallow†. *Biochemistry* **2013**, *52* (35), 10.1021/bi400642u. https://doi.org/10.1021/bi400642u.

(65) Nyarko, A.; Song, Y.; Barbar, E. Intrinsic Disorder in Dynein Intermediate Chain Modulates Its Interactions with NudE and Dynactin. *J Biol Chem* **2012**, *287* (30), 24884–24893. https://doi.org/10.1074/jbc.M112.376038.

(66) Luo, S.; Garcia-Arencibia, M.; Zhao, R.; Puri, C.; Toh, P. P. C.; Sadiq, O.; Rubinsztein, D. C. Bim Inhibits Autophagy by Recruiting Beclin 1 to Microtubules. *Mol Cell* **2012**, *47* (3), 359–370. https://doi.org/10.1016/j.molcel.2012.05.040.

(67) Fan, J.-S.; Zhang, Q.; Tochio, H.; Zhang, M. Backbone Dynamics of the 8 KDa Dynein Light Chain Dimer Reveals Molecular Basis of the Protein's Functional Diversity. *J Biomol NMR* **2002**, *23* (2), 103–114. https://doi.org/10.1023/a:1016332918178.

(68) Hall, J.; Hall, A.; Pursifull, N.; Barbar, E. Differences in Dynamic Structure of LC8 Monomer, Dimer, and Dimer-Peptide Complexes. *Biochemistry* **2008**, *47* (46), 11940–11952. https://doi.org/10.1021/bi801093k.

(69) Erdős, G.; Szaniszló, T.; Pajkos, M.; Hajdu-Soltész, B.; Kiss, B.; Pál, G.; Nyitray, L.; Dosztányi, Z. Novel Linear Motif Filtering Protocol Reveals the Role of the LC8 Dynein Light Chain in the Hippo Pathway. *PLOS Computational Biology* **2017**, *13* (12), e1005885. https://doi.org/10.1371/journal.pcbi.1005885.

(70) Jurado, S.; Smyth, I.; Denderen, B. van; 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.; Heierhorst, J. Dual Functions of ASCIZ in the DNA Base Damage Response and Pulmonary Organogenesis. *PLOS Genetics* **2010**, *6* (10), e1001170. https://doi.org/10.1371/journal.pgen.1001170.

(71) Reck-Peterson, S. L.; Redwine, W. B.; Vale, R. D.; Carter, A. P. The Cytoplasmic Dynein Transport Machinery and Its Many Cargoes. *Nat Rev Mol Cell Biol* **2018**, *19* (6), 382–398. https://doi.org/10.1038/s41580-018-0004-3.

(72) Lipka, J.; Kuijpers, M.; Jaworski, J.; Hoogenraad, C. C. Mutations in Cytoplasmic Dynein and Its Regulators Cause Malformations of Cortical Development and Neurodegenerative Diseases. *Biochem Soc Trans* **2013**, *41* (6), 1605–1612. https://doi.org/10.1042/BST20130188.

(73) Urnavicius, L.; Lau, C. K.; Elshenawy, M. M.; Morales-Rios, E.; Motz, C.; Yildiz, A.; Carter, A. P. Cryo-EM Shows How Dynactin Recruits Two Dyneins for Faster Movement. *Nature* **2018**, *554* (7691), 202–206. https://doi.org/10.1038/nature25462.

(74) Urnavicius, L.; Zhang, K.; Diamant, A. G.; Motz, C.; Schlager, M. A.; Yu, M.; Patel, N. A.; Robinson, C. V.; Carter, A. P. The Structure of the Dynactin Complex and Its Interaction with Dynein. *Science* **2015**, *347* (6229), 1441–1446. https://doi.org/10.1126/science.aaa4080.

(75) Kon, T.; Oyama, T.; Shimo-Kon, R.; Imamula, K.; Shima, T.; Sutoh, K.; Kurisu, G. The 2.8 Å Crystal Structure of the Dynein Motor Domain. *Nature* **2012**, *484* (7394), 345–350. https://doi.org/10.1038/nature10955.

(76) Schmidt, H.; Gleave, E. S.; Carter, A. P. Insights into Dynein Motor Domain Function from a 3.3-Å Crystal Structure. *Nat Struct Mol Biol* **2012**, *19* (5), 492–497, S1. https://doi.org/10.1038/nsmb.2272.

(77) Hall, J.; Song, Y.; Karplus, P. A.; Barbar, E. The Crystal Structure of Dynein Intermediate Chain-Light Chain Roadblock Complex Gives New Insights into Dynein Assembly. *J Biol Chem* **2010**, *285* (29), 22566–22575. https://doi.org/10.1074/jbc.M110.103861.

(78) Lo, K. W.; Naisbitt, S.; Fan, J. S.; Sheng, M.; Zhang, M. The 8-KDa Dynein Light Chain Binds to Its Targets via a Conserved (K/R)XTQT Motif. *J Biol Chem* **2001**, *276* (17), 14059–14066. https://doi.org/10.1074/jbc.M010320200.

(79) Mok, Y. K.; Lo, K. W.; Zhang, M. Structure of Tctex-1 and Its Interaction with Cytoplasmic Dynein Intermediate Chain. *J Biol Chem* **2001**, *276* (17), 14067–14074. https://doi.org/10.1074/jbc.M011358200.

(80) Susalka, S. J.; Nikulina, K.; Salata, M. W.; Vaughan, P. S.; King, S. M.; Vaughan, K. T.; Pfister, K. K. The Roadblock Light Chain Binds a Novel Region of the Cytoplasmic Dynein Intermediate Chain. *J Biol Chem* **2002**, *277* (36), 32939–32946. https://doi.org/10.1074/jbc.M205510200.

(81) Hall, J.; Karplus, P. A.; Barbar, E. Multivalency in the Assembly of Intrinsically Disordered Dynein Intermediate Chain. *J Biol Chem* **2009**, *284* (48), 33115–33121. https://doi.org/10.1074/jbc.M109.048587.

(82) Zhang, K.; Foster, H. E.; Rondelet, A.; Lacey, S. E.; Bahi-Buisson, N.; Bird, A. W.; Carter, A. P. Cryo-EM Reveals How Human Cytoplasmic Dynein Is Auto-Inhibited and Activated. *Cell* **2017**, *169* (7), 1303-1314.e18. https://doi.org/10.1016/j.cell.2017.05.025.

(83) Lee, I.-G.; Olenick, M. A.; Boczkowska, M.; Franzini-Armstrong, C.; Holzbaur, E. L. F.; Dominguez, R. A Conserved Interaction of the Dynein Light Intermediate Chain with Dynein-Dynactin Effectors Necessary for Processivity. *Nat Commun* **2018**, *9* (1), 986. https://doi.org/10.1038/s41467-018-03412-8.

(84) Chowdhury, S.; Ketcham, S. A.; Schroer, T. A.; Lander, G. C. Structural Organization of the Dynein–Dynactin Complex Bound to Microtubules. *Nat Struct Mol Biol* **2015**, *22* (4), 345–347. https://doi.org/10.1038/nsmb.2996.

(85) Jie, J.; Löhr, F.; Barbar, E. Interactions of Yeast Dynein with Dynein Light Chain and Dynactin: GENERAL IMPLICATIONS FOR INTRINSICALLY DISORDERED DUPLEX SCAFFOLDS IN MULTIPROTEIN ASSEMBLIES. *J Biol Chem* **2015**, *290* (39), 23863–23874. https://doi.org/10.1074/jbc.M115.649715.

(86) Wright, P. E.; Dyson, H. J. Linking Folding and Binding. *Current Opinion in Structural Biology* **2009**, *19* (1), 31–38. https://doi.org/10.1016/j.sbi.2008.12.003.

(87) Benison, G.; Nyarko, A.; Barbar, E. Heteronuclear NMR Identifies a Nascent Helix in Intrinsically Disordered Dynein Intermediate Chain: Implications for Folding and Dimerization. *J Mol Biol* **2006**, *362* (5), 1082–1093. https://doi.org/10.1016/j.jmb.2006.08.006.

(88) Schroer, T. A. Dynactin. *Annu Rev Cell Dev Biol* **2004**, *20*, 759–779. https://doi.org/10.1146/annurev.cellbio.20.012103.094623.

(89) King, S. J.; Brown, C. L.; Maier, K. C.; Quintyne, N. J.; Schroer, T. A. Analysis of the Dynein-Dynactin Interaction In Vitro and In Vivo. *Mol Biol Cell* **2003**, *14* (12), 5089–5097. https://doi.org/10.1091/mbc.E03-01-0025.

(90) Morgan, J. L.; Song, Y.; Barbar, E. Structural Dynamics and Multiregion Interactions in Dynein-Dynactin Recognition. *J Biol Chem* **2011**, *286* (45), 39349–39359. https://doi.org/10.1074/jbc.M111.296277.

(91) Morgan, J. L.; Yeager, A.; Estelle, A. B.; Gsponer, J.; Barbar, E. Transient Tertiary Structures of Disordered Dynein Intermediate Chain Regulate Its Interactions with Multiple Partners. *J Mol Biol* **2021**, *433* (18), 167152. https://doi.org/10.1016/j.jmb.2021.167152.

(92) Jie, J.; Löhr, F.; Barbar, E. Dynein Binding of Competitive Regulators Dynactin and NudE Involves Novel Interplay between Phosphorylation Site and Disordered Spliced Linkers. *Structure* **2017**, *25* (3), 421–433. https://doi.org/10.1016/j.str.2017.01.003.

(93) Siglin, A. E.; Sun, S.; Moore, J. K.; Tan, S.; Poenie, M.; Lear, J. D.; Polenova, T.; Cooper, J. A.; Williams, J. C. Dynein and Dynactin Leverage Their Bivalent Character to Form a High-Affinity Interaction. *PLOS ONE* **2013**, *8* (4), e59453. https://doi.org/10.1371/journal.pone.0059453.

(94) Gill, S. R.; Schroer, T. A.; Szilak, I.; Steuer, E. R.; Sheetz, M. P.; Cleveland, D. W. Dynactin, a Conserved, Ubiquitously Expressed Component of an Activator of Vesicle Motility Mediated by Cytoplasmic Dynein. *J Cell Biol* **1991**, *115* (6), 1639–1650. https://doi.org/10.1083/jcb.115.6.1639.

(95) Kardon, J. R.; Vale, R. D. Regulators of the Cytoplasmic Dynein Motor. *Nat Rev Mol Cell Biol* **2009**, *10* (12), 854–865. https://doi.org/10.1038/nrm2804.

(96) Schroeder, C. M.; Vale, R. D. Assembly and Activation of Dynein-Dynactin by the Cargo Adaptor Protein Hook3. *J Cell Biol* **2016**, *214* (3), 309–318. https://doi.org/10.1083/jcb.201604002.

(97) Loening, N. M.; Saravanan, S.; Jespersen, N. E.; Jara, K.; Barbar, E. Interplay of Disorder and Sequence Specificity in the Formation of Stable Dynein-Dynactin Complexes. *Biophys J* **2020**, *119* (5), 950–965. https://doi.org/10.1016/j.bpj.2020.07.023.

(98) Feng, Y.; Walsh, C. A. Mitotic Spindle Regulation by Nde1 Controls Cerebral Cortical Size. *Neuron* **2004**, *44* (2), 279–293. https://doi.org/10.1016/j.neuron.2004.09.023.

(99) Wainman, A.; Creque, J.; Williams, B.; Williams, E. V.; Bonaccorsi, S.; Gatti, M.; Goldberg, M. L. Roles of the Drosophila NudE Protein in Kinetochore Function and Centrosome Migration. *J Cell Sci* **2009**, *122* (Pt 11), 1747–1758. https://doi.org/10.1242/jcs.041798.

(100) Liang, Y.; Yu, W.; Li, Y.; Yu, L.; Zhang, Q.; Wang, F.; Yang, Z.; Du, J.; Huang, Q.; Yao, X.; Zhu, X. Nudel Modulates Kinetochore Association and Function of Cytoplasmic Dynein in M Phase. *MBoC* **2007**, *18* (7), 2656–2666. https://doi.org/10.1091/mbc.e06-04-0345.

(101) Efimov, V. P.; Morris, N. R. The LIS1-Related NUDF Protein of Aspergillus Nidulans Interacts with the Coiled-Coil Domain of the NUDE/RO11 Protein. *J Cell Biol* **2000**, *150* (3), 681–688. https://doi.org/10.1083/jcb.150.3.681.

(102) Lam, C.; Vergnolle, M. A. S.; Thorpe, L.; Woodman, P. G.; Allan, V. J. Functional Interplay between LIS1, NDE1 and NDEL1 in Dynein-Dependent Organelle Positioning. *J Cell Sci* **2010**, *123* (Pt 2), 202–212. https://doi.org/10.1242/jcs.059337.

(103) Derewenda, U.; Tarricone, C.; Choi, W. C.; Cooper, D. R.; Lukasik, S.; Perrina, F.; Tripathy, A.; Kim, M. H.; Cafiso, D. S.; Musacchio, A.; Derewenda, Z. S. The Structure of the Coiled-Coil Domain of Ndel1 and the Basis of Its Interaction with Lis1, the Causal Protein of Miller-Dieker Lissencephaly. *Structure* **2007**, *15* (11), 1467–1481. https://doi.org/10.1016/j.str.2007.09.015.

(104) Bugge, K.; Brakti, I.; Fernandes, C. B.; Dreier, J. E.; Lundsgaard, J. E.; Olsen, J. G.; Skriver, K.; Kragelund, B. B. Interactions by Disorder – A Matter of Context. *Frontiers in Molecular Biosciences* **2020**, *7*, 110. https://doi.org/10.3389/fmolb.2020.00110.

(105) Nyarko, A.; Barbar, E. Light Chain-Dependent Self-Association of Dynein Intermediate Chain. *J Biol Chem* **2011**, *286* (2), 1556–1566. https://doi.org/10.1074/jbc.M110.171686.

(106) Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Research* **2019**, *47* (W1), W402–W407. https://doi.org/10.1093/nar/gkz297.

(107) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(108) Muñoz, V.; Serrano, L. Development of the Multiple Sequence Approximation within the AGADIR Model of Alpha-Helix Formation: Comparison with Zimm-Bragg and Lifson-Roig Formalisms. *Biopolymers* **1997**, *41* (5), 495–509. https://doi.org/10.1002/(SICI)1097-0282(19970415)41:5<495::AID-BIP2>3.0.CO;2-H.

(109) Muñoz, V.; Serrano, L. Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. *Nat Struct Mol Biol* **1994**, *1* (6), 399–409. https://doi.org/10.1038/nsb0694-399.

(110) Peckham, M.; Knight, P. J. When a Predicted Coiled Coil Is Really a Single α-Helix, in Myosins and Other Proteins. *Soft Matter* **2009**, *5* (13), 2493–2503. https://doi.org/10.1039/B822339D.

(111) Vallee, R. B.; McKenney, R. J.; Ori-McKenney, K. M. Multiple Modes of Cytoplasmic Dynein Regulation. *Nat Cell Biol* **2012**, *14* (3), 224–230. https://doi.org/10.1038/ncb2420.

(112) Poulsen, F. A Brief Introduction to NMR Spectroscopy of Proteins. *undefined* **2003**.

(113) Pervushin, K.; Riek, R.; Wider, G.; Wüthrich, K. Transverse Relaxation-Optimized Spectroscopy (TROSY) for NMR Studies of Aromatic Spin Systems in 13C-Labeled Proteins. *J. Am. Chem. Soc.* **1998**, *120* (25), 6394–6400. https://doi.org/10.1021/ja980742g.

(114) Fernández, C.; Wider, G. TROSY in NMR Studies of the Structure and Function of Large Biological Macromolecules. *Current Opinion in Structural Biology* **2003**, *13* (5), 570–580. https://doi.org/10.1016/j.sbi.2003.09.009.

(115) Venditti, V.; Fawzi, N. L. Probing the Atomic Structure of Transient Protein Contacts by Paramagnetic Relaxation Enhancement Solution NMR. *Methods Mol Biol* **2018**, *1688*, 243–255. https://doi.org/10.1007/978-1-4939-7386-6_12.

(116) Clore, G. M. Visualizing Lowly-Populated Regions of the Free Energy Landscape of Macromolecular Complexes by Paramagnetic Relaxation Enhancement. *Mol Biosyst* **2008**, *4* (11), 1058–1069. https://doi.org/10.1039/b810232e.

(117) Ganguly, D.; Chen, J. Structural Interpretation of Paramagnetic Relaxation Enhancement-Derived Distances for Disordered Protein States. *J Mol Biol* **2009**, *390* (3), 467–477. https://doi.org/10.1016/j.jmb.2009.05.019.

(118) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. NMRPipe: A Multidimensional Spectral Processing System Based on UNIX Pipes. *J Biomol NMR* **1995**, *6* (3), 277–293. https://doi.org/10.1007/BF00197809.

(119) Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, M.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins* **2005**, *59* (4), 687–696. https://doi.org/10.1002/prot.20449.

(120) Johnson, B. A. Using NMRView to Visualize and Analyze the NMR Spectra of Macromolecules. In *Protein NMR Techniques*; Downing, A. K., Ed.; Methods in Molecular Biology™; Humana Press: Totowa, NJ, 2004; pp 313–352. https://doi.org/10.1385/1-59259-809-9:313.

(121) Coggins, B. E.; Werner-Allen, J. W.; Yan, A.; Zhou, P. Rapid Protein Global Fold Determination Using Ultrasparse Sampling, High-Dynamic Range Artifact Suppression, and Time-Shared NOESY. *J Am Chem Soc* **2012**, *134* (45), 18619–18630. https://doi.org/10.1021/ja307445y.

(122) Shen, Y.; Bax, A. Protein Backbone and Sidechain Torsion Angles Predicted from NMR Chemical Shifts Using Artificial Neural Networks. *J Biomol NMR* **2013**, *56* (3), 227–241. https://doi.org/10.1007/s10858-013-9741-y.

(123) Rieping, W.; Habeck, M.; Bardiaux, B.; Bernard, A.; Malliavin, T. E.; Nilges, M. ARIA2: Automated NOE Assignment and Data Integration in NMR Structure Calculation. *Bioinformatics* **2007**, *23* (3), 381–382. https://doi.org/10.1093/bioinformatics/btl589.

(124) Anthis, N. J.; Clore, G. M. Sequence-Specific Determination of Protein and Peptide Concentrations by Absorbance at 205 Nm. *Protein Sci* **2013**, *22* (6), 851–858. https://doi.org/10.1002/pro.2253.

(125) Lepre, C. A.; Moore, J. M. Microdrop Screening: A Rapid Method to Optimize Solvent Conditions for NMR Spectroscopy of Proteins. *J Biomol NMR* **1998**, *12* (4), 493–499. https://doi.org/10.1023/a:1008353000679.

(126) Solyom, Z.; Schwarten, M.; Geist, L.; Konrat, R.; Willbold, D.; Brutscher, B. BEST-TROSY Experiments for Time-Efficient Sequential Resonance Assignment of Large Disordered Proteins. *J Biomol NMR* **2013**, *55* (4), 311–321. https://doi.org/10.1007/s10858-013-9715-0.

(127) Zhu, G.; Xia, Y.; Nicholson, L. K.; Sze, K. H. Protein Dynamics Measurements by TROSY-Based NMR Experiments. *Journal of Magnetic Resonance* **2000**, *143* (2), 423–426. https://doi.org/10.1006/jmre.2000.2022.

(128) Wagstaff, J. L.; Taylor, S. L.; Howard, M. J. Recent Developments and Applications of Saturation Transfer Difference Nuclear Magnetic Resonance (STD NMR) Spectroscopy. *Mol Biosyst* **2013**, *9* (4), 571–577. https://doi.org/10.1039/c2mb25395j.

(129) Becker, W.; Bhattiprolu, K. C.; Gubensäk, N.; Zangger, K. Investigating Protein-Ligand Interactions by Solution Nuclear Magnetic Resonance Spectroscopy. *Chemphyschem* **2018**, *19* (8), 895–906. https://doi.org/10.1002/cphc.201701253.

(130) Rapali, P.; Szenes, Á.; Radnai, L.; Bakos, A.; Pál, G.; Nyitray, L. DYNLL/LC8: A Light Chain Subunit of the Dynein Motor Complex and Beyond. *The FEBS Journal* **2011**, *278* (17), 2980–2996. https://doi.org/10.1111/j.1742-4658.2011.08254.x.

(131) Jespersen, N.; Barbar, E. Emerging Features of Linear Motif-Binding Hub Proteins. *Trends in Biochemical Sciences* **2020**, *45* (5), 375–384. https://doi.org/10.1016/j.tibs.2020.01.004.

(132) Fan, J.; Zhang, Q.; Tochio, H.; Li, M.; Zhang, M. Structural Basis of Diverse Sequence-Dependent Target Recognition by the 8 KDa Dynein Light Chain. *J Mol Biol* **2001**, *306* (1), 97–108. https://doi.org/10.1006/jmbi.2000.4374.

(133) Benison, G.; Barbar, E. Chapter 9 NMR Analysis of Dynein Light Chain Dimerization and Interactions with Diverse Ligands. In *Methods in Enzymology*; Biothermodynamics, Part A; Academic Press, 2009; Vol. 455, pp 237–258. https://doi.org/10.1016/S0076-6879(08)04209-2.

(134) Dunsch, A. K.; Hammond, D.; Lloyd, J.; Schermelleh, L.; Gruneberg, U.; Barr, F. A. Dynein Light Chain 1 and a Spindle-Associated Adaptor Promote Dynein Asymmetry and Spindle Orientation. *Journal of Cell Biology* **2012**, *198* (6), 1039–1054. https://doi.org/10.1083/jcb.201202112.

(135) Fejtova, A.; Davydova, D.; Bischof, F.; Lazarevic, V.; Altrock, W. D.; Romorini, S.; Schöne, C.; Zuschratter, W.; Kreutz, M. R.; Garner, C. C.; Ziv, N. E.; Gundelfinger, E. D. Dynein Light Chain Regulates Axonal Trafficking and Synaptic Levels of Bassoon. *J Cell Biol* **2009**, *185* (2), 341–355. https://doi.org/10.1083/jcb.200807155.

(136) Stelter, P.; Kunze, R.; Flemming, D.; Höpfner, D.; Diepholz, M.; Philippsen, P.; Böttcher, B.; Hurt, E. Molecular Basis for the Functional Interaction of Dynein Light Chain with the Nuclear-Pore Complex. *Nat Cell Biol* **2007**, *9* (7), 788–796. https://doi.org/10.1038/ncb1604.

(137) Rapali, P.; García-Mayoral, M. F.; Martínez-Moreno, M.; Tárnok, K.; Schlett, K.; Albar, J. P.; Bruix, M.; Nyitray, L.; Rodriguez-Crespo, I. LC8 Dynein Light Chain (DYNLL1) Binds to the C-Terminal Domain of ATM-Interacting Protein (ATMIN/ASCIZ) and Regulates Its Subcellular Localization. *Biochem Biophys Res Commun* **2011**, *414* (3), 493–498. https://doi.org/10.1016/j.bbrc.2011.09.093.

(138) Jurado, S.; Gleeson, K.; O'Donnell, K.; Izon, D. J.; Walkley, C. R.; Strasser, A.; Tarlinton, D. M.; Heierhorst, J. The Zinc-Finger Protein ASCIZ Regulates B Cell Development via

DYNLL1 and Bim. *J Exp Med* **2012**, *209* (9), 1629–1639. https://doi.org/10.1084/jem.20120785.

(139) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera—A Visualization System for Exploratory Research and Analysis. *Journal of Computational Chemistry* **2004**, *25* (13), 1605–1612. https://doi.org/10.1002/jcc.20084.

(140) Venters, R. A.; Farmer II, B. T.; Fierke, C. A.; Spicer, L. D. Characterizing the Use of Perdeuteration in NMR Studies of Large Proteins:13C,15N And1H Assignments of Human Carbonic Anhydrase II. *Journal of Molecular Biology* **1996**, *264* (5), 1101–1116. https://doi.org/10.1006/jmbi.1996.0699.

(141) Shimada, I.; Ueda, T.; Matsumoto, M.; Sakakura, M.; Osawa, M.; Takeuchi, K.; Nishida, N.; Takahashi, H. Cross-Saturation and Transferred Cross-Saturation Experiments. *Progress in nuclear magnetic resonance spectroscopy* **2009**, *54* (2), 123–140. https://doi.org/10.1016/j.pnmrs.2008.07.001.

(142) Gaik, M.; Flemming, D.; von Appen, A.; Kastritis, P.; Mücke, N.; Fischer, J.; Stelter, P.; Ori, A.; Bui, K. H.; Baßler, J.; Barbar, E.; Beck, M.; Hurt, E. Structural Basis for Assembly and Function of the Nup82 Complex in the Nuclear Pore Scaffold. *J Cell Biol* **2015**, *208* (3), 283–297. https://doi.org/10.1083/jcb.201411003.

(143) Toropova, K.; Zalyte, R.; Mukhopadhyay, A. G.; Mladenov, M.; Carter, A. P.; Roberts, A. J. Structure of the Dynein-2 Complex and Its Assembly with Intraflagellar Transport Trains. *Nat Struct Mol Biol* **2019**, *26* (9), 823–829. https://doi.org/10.1038/s41594-019-0286-y.

(144) Krishnamurthy, V. M.; Semetey, V.; Bracher, P. J.; Shen, N.; Whitesides, G. M. Dependence of Effective Molarity on Linker Length for an Intramolecular Protein−Ligand System. *J. Am. Chem. Soc.* **2007**, *129* (5), 1312–1320. https://doi.org/10.1021/ja066780e.

(145) Sørensen, C. S.; Kjaergaard, M. Effective Concentrations Enforced by Intrinsically Disordered Linkers Are Governed by Polymer Physics. *Proceedings of the National Academy of Sciences* **2019**, *116* (46), 23124–23131. https://doi.org/10.1073/pnas.1904813116.

(146) Schuck, P. Size-Distribution Analysis of Macromolecules by Sedimentation Velocity Ultracentrifugation and Lamm Equation Modeling. *Biophysical Journal* **2000**, *78* (3), 1606–1619. https://doi.org/10.1016/S0006-3495(00)76713-0.

(147) Laue, T.; Shah, B.; Ridgeway, T.; Pelletier, S. Computer-Aided Interpretation of Sedimentation Data for Proteins. *Analytical ultracentrifugation in biochemistry and polymer science* **1992**.

(148) Wheeler, L. C.; Donor, M. T.; Prell, J. S.; Harms, M. J. Multiple Evolutionary Origins of Ubiquitous Cu2+ and Zn2+ Binding in the S100 Protein Family. *PLOS ONE* **2016**, *11* (10), e0164740. https://doi.org/10.1371/journal.pone.0164740.

(149) Wishart, D. S.; Sykes, B. D. The 13C Chemical-Shift Index: A Simple Method for the Identification of Protein Secondary Structure Using 13C Chemical-Shift Data. *J Biomol NMR* **1994**, *4* (2), 171–180. https://doi.org/10.1007/BF00175245.

(150) Wishart, D. S.; Case, D. A. Use of Chemical Shifts in Macromolecular Structure Determination. *Methods Enzymol* **2001**, *338*, 3–34. https://doi.org/10.1016/s0076-6879(02)38214-4.

(151) Clark, S.; Myers, J. B.; King, A.; Fiala, R.; Novacek, J.; Pearce, G.; Heierhorst, J.; Reichow, S. L.; Barbar, E. J. Multivalency Regulates Activity in an Intrinsically Disordered Transcription Factor. *eLife 7*, e36258. https://doi.org/10.7554/eLife.36258.

(152) Reardon, P. N.; Jara, K. A.; Rolland, A. D.; Smith, D. A.; Hoang, H. T. M.; Prell, J. S.; Barbar, E. J. The Dynein Light Chain 8 (LC8) Binds Predominantly "in-Register" to a Multivalent Intrinsically Disordered Partner. *J Biol Chem* **2020**, *295* (15), 4912–4922. https://doi.org/10.1074/jbc.RA119.011653.

(153) Splinter, D.; Razafsky, D. S.; Schlager, M. A.; Serra-Marques, A.; Grigoriev, I.; Demmers, J.; Keijzer, N.; Jiang, K.; Poser, I.; Hyman, A. A.; Hoogenraad, C. C.; King, S. J.; Akhmanova, A. BICD2, Dynactin, and LIS1 Cooperate in Regulating Dynein Recruitment to Cellular Structures. *Mol Biol Cell* **2012**, *23* (21), 4226–4241. https://doi.org/10.1091/mbc.E12-03-0210.

(154) Carter, A. P.; Diamant, A. G.; Urnavicius, L. How Dynein and Dynactin Transport Cargos: A Structural Perspective. *Curr Opin Struct Biol* **2016**, *37*, 62–70. https://doi.org/10.1016/j.sbi.2015.12.003.

(155) McKenney, R. J.; Weil, S. J.; Scherer, J.; Vallee, R. B. Mutually Exclusive Cytoplasmic Dynein Regulation by NudE-Lis1 and Dynactin *. *Journal of Biological Chemistry* **2011**, *286* (45), 39615–39622. https://doi.org/10.1074/jbc.M111.289017.

(156) Wang, S.; Zheng, Y. Identification of a Novel Dynein Binding Domain in Nudel Essential for Spindle Pole Organization in Xenopus Egg Extract. *J Biol Chem* **2011**, *286* (1), 587–593. https://doi.org/10.1074/jbc.M110.181578.

(157) Zyłkiewicz, E.; Kijańska, M.; Choi, W.-C.; Derewenda, U.; Derewenda, Z. S.; Stukenberg, P. T. The N-Terminal Coiled-Coil of Ndel1 Is a Regulated Scaffold That Recruits LIS1 to Dynein. *J Cell Biol* **2011**, *192* (3), 433–445. https://doi.org/10.1083/jcb.201011142.

(158) Yamada, M.; Toba, S.; Yoshida, Y.; Haratani, K.; Mori, D.; Yano, Y.; Mimori-Kiyosue, Y.; Nakamura, T.; Itoh, K.; Fushiki, S.; Setou, M.; Wynshaw-Boris, A.; Torisawa, T.; Toyoshima, Y. Y.; Hirotsune, S. LIS1 and NDEL1 Coordinate the Plus-End-Directed Transport of Cytoplasmic Dynein. *EMBO J* **2008**, *27* (19), 2471–2483. https://doi.org/10.1038/emboj.2008.182.

(159) Berlow, R. B.; Dyson, H. J.; Wright, P. E. Hypersensitive Termination of the Hypoxic Response by a Disordered Protein Switch. *Nature* **2017**, *543* (7645), 447–451. https://doi.org/10.1038/nature21705.

(160) Chen, Y.; Wang, P.; Slep, K. C. Mapping Multivalency in the CLIP-170–EB1 Microtubule plus-End Complex. *Journal of Biological Chemistry* **2019**, *294* (3), 918–931. https://doi.org/10.1074/jbc.RA118.006125.

(161) Cole, J. L.; Lary, J. W.; P Moody, T.; Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods Cell Biol* **2008**, *84*, 143–179. https://doi.org/10.1016/S0091-679X(07)84006-4.

(162) Greenfield, N. J. Using Circular Dichroism Spectra to Estimate Protein Secondary Structure. *Nat Protoc* **2006**, *1* (6), 2876–2890. https://doi.org/10.1038/nprot.2006.202.

(163) *PSIPRED Protein Analysis Workbench: 20 years on | Nucleic Acids Research | Oxford Academic*. https://academic.oup.com/nar/article/47/W1/W402/5480136 (accessed 2021-08-17).

(164) Muñoz, V.; Serrano, L. Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. II. Helix Macrodipole Effects and Rational Modification of the Helical Content of Natural Peptides. *J Mol Biol* **1995**, *245* (3), 275–296. https://doi.org/10.1006/jmbi.1994.0023.

(165) Muñoz, V.; Serrano, L. Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. III. Temperature and PH Dependence. *J Mol Biol* **1995**, *245* (3), 297–308. https://doi.org/10.1006/jmbi.1994.0024.

(166) Muñoz, V.; Serrano, L. Elucidating the Folding Problem of Helical Peptides Using Empirical Parameters. *Nat Struct Biol* **1994**, *1* (6), 399–409. https://doi.org/10.1038/nsb0694-399.

(167) Wei, Y.; Thyparambil, A. A.; Latour, R. A. Protein Helical Structure Determination Using CD Spectroscopy for Solutions with Strong Background Absorbance from 190-230 Nm. *Biochim Biophys Acta* **2014**, *1844* (12), 2331–2337. https://doi.org/10.1016/j.bbapap.2014.10.001.

(168) Alderson, T. R.; Lee, J. H.; Charlier, C.; Ying, J.; Bax, A. Propensity for Cis-Proline Formation in Unfolded Proteins. *Chembiochem* **2018**, *19* (1), 37–42. https://doi.org/10.1002/cbic.201700548.

(169) Vaughan, P. S.; Leszyk, J. D.; Vaughan, K. T. Cytoplasmic Dynein Intermediate Chain Phosphorylation Regulates Binding to Dynactin. *J Biol Chem* **2001**, *276* (28), 26171–26179. https://doi.org/10.1074/jbc.M102649200.

(170) Trokter, M.; Mücke, N.; Surrey, T. Reconstitution of the Human Cytoplasmic Dynein Complex. *PNAS* **2012**, *109* (51), 20895–20900. https://doi.org/10.1073/pnas.1210573110.

(171) McKenney, R. J.; Huynh, W.; Tanenbaum, M. E.; Bhabha, G.; Vale, R. D. Activation of Cytoplasmic Dynein Motility by Dynactin-Cargo Adapter Complexes. *Science* **2014**, *345* (6194), 337–341. https://doi.org/10.1126/science.1254198.

(172) Schlager, M. A.; Hoang, H. T.; Urnavicius, L.; Bullock, S. L.; Carter, A. P. In Vitro Reconstitution of a Highly Processive Recombinant Human Dynein Complex. *EMBO J* **2014**, *33* (17), 1855–1868. https://doi.org/10.15252/embj.201488792.

(173) Olenick, M. A.; Tokito, M.; Boczkowska, M.; Dominguez, R.; Holzbaur, E. L. F. Hook Adaptors Induce Unidirectional Processive Motility by Enhancing the Dynein-Dynactin Interaction *. *Journal of Biological Chemistry* **2016**, *291* (35), 18239–18251. https://doi.org/10.1074/jbc.M116.738211.

(174) Gibson, D. G.; Young, L.; Chuang, R.-Y.; Venter, J. C.; Hutchison, C. A.; Smith, H. O. Enzymatic Assembly of DNA Molecules up to Several Hundred Kilobases. *Nat Methods* **2009**, *6* (5), 343–345. https://doi.org/10.1038/nmeth.1318.

(175) Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R.-Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. A.; Young, L.; Qi, Z.-Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. A.; Smith, H. O.;

Venter, J. C. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science* **2010**, *329* (5987), 52–56. https://doi.org/10.1126/science.1190719.

(176) Studier, F. W. Protein Production by Auto-Induction in High Density Shaking Cultures. *Protein Expr Purif* **2005**, *41* (1), 207–234. https://doi.org/10.1016/j.pep.2005.01.016.

(177) Jansson, M.; Li, Y. C.; Jendeberg, L.; Anderson, S.; Montelione, G. T.; Nilsson, B. High-Level Production of Uniformly ^{15}N- and ^{13}C-Enriched Fusion Proteins in Escherichia Coli. *J Biomol NMR* **1996**, *7* (2), 131–141. https://doi.org/10.1007/BF00203823.

(178) UniProt Consortium. UniProt: A Worldwide Hub of Protein Knowledge. *Nucleic Acids Res* **2019**, *47* (D1), D506–D515. https://doi.org/10.1093/nar/gky1049.

(179) *HeteroAnalysis Contents*. https://core.uconn.edu/files/auf/ha-help/HA-Help.htm (accessed 2021-08-17).

(180) Katoh, K.; Standley, D. M. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* **2013**, *30* (4), 772–780. https://doi.org/10.1093/molbev/mst010.

Appendices

Appendix 1

The dynein light chain 8 (LC8) binds predominantly "in-register" to a multivalent intrinsically disordered partner – supplemental material

Patrick N. Reardon[1], Kayla A. Jara[1], Amber D. Rolland, Delaney A. Smith, Hanh T. M. Hoang, James S. Prell, and Elisar J. Barbar

[1]Authors contributed equally

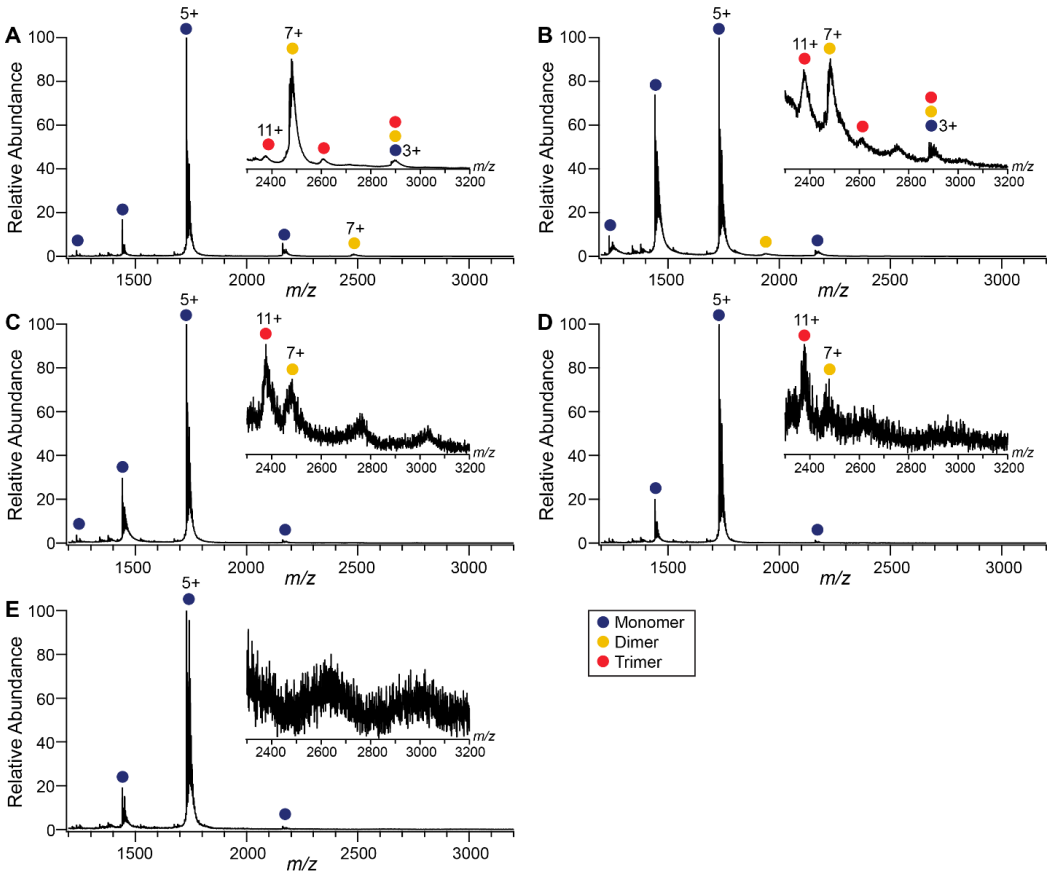

Figure A1.1. Native mass spectra of QT2-4 acquired over a range of concentrations. 25 µM (A), 10 µM (B), 1 µM (C), 500 nM (D), and 100 nM (E). Insets show a zoomed-in version of the *m/z* 2300-3200 region of the corresponding mass spectrum within each panel. The peaks associated with each QT2-4 homo-oligomer are labeled as follows: monomer, blue circles; dimer, yellow circles; trimer, red circles. One charge state per oligomer has been labeled in each panel and inset for clarity.

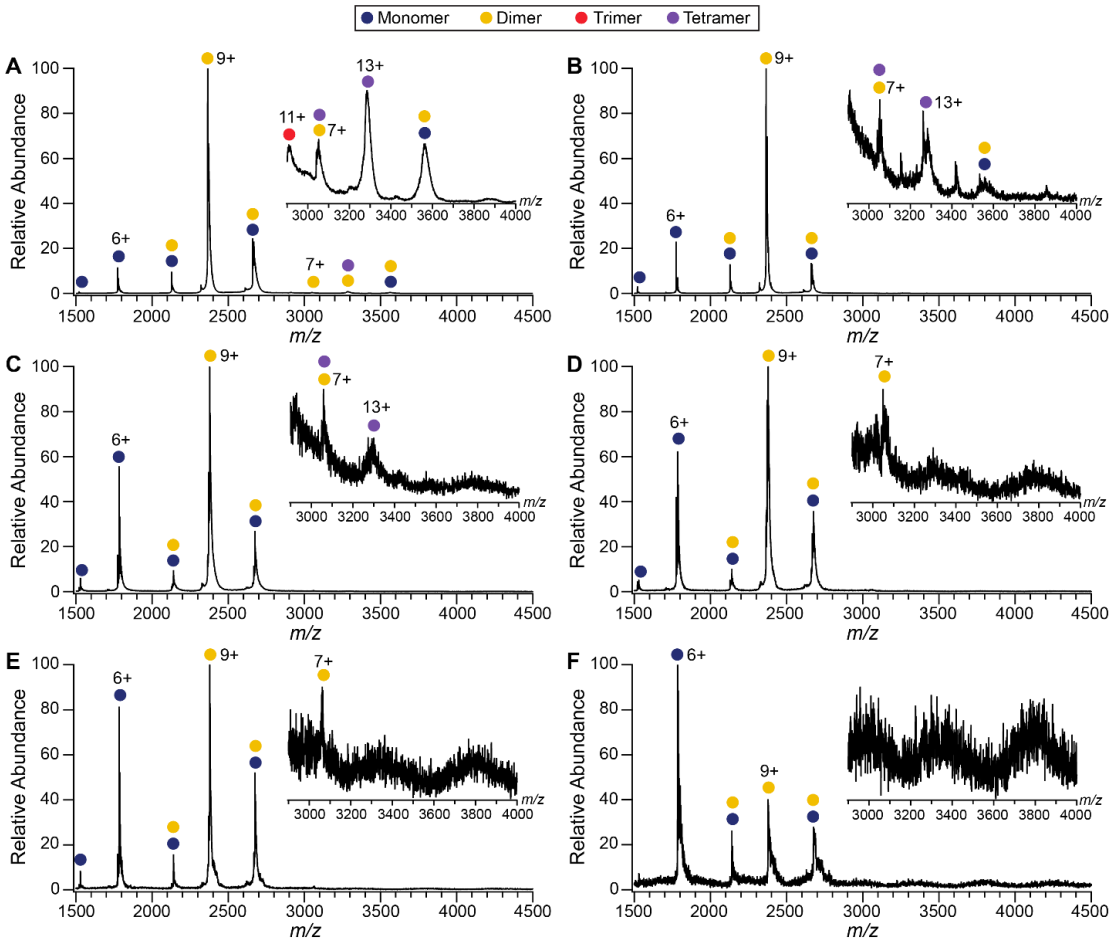

Figure A1.2. Native mass spectra of LC8 acquired over a range of concentrations. 25 µM (A), 10 µM (B), 1 µM (C), 500 nM (D), 100 nM (E), and 10 nM (F). Insets show a zoomed-in version of the *m/z* 2900-4000 region of the corresponding mass spectrum within each panel. The peaks associated with each LC8 homo-oligomer are labeled as follows: monomer, blue circles; dimer, yellow circles; trimer, red circles; tetramer, purple circles. One charge state per oligomer has been labeled in each main panel and inset for clarity.

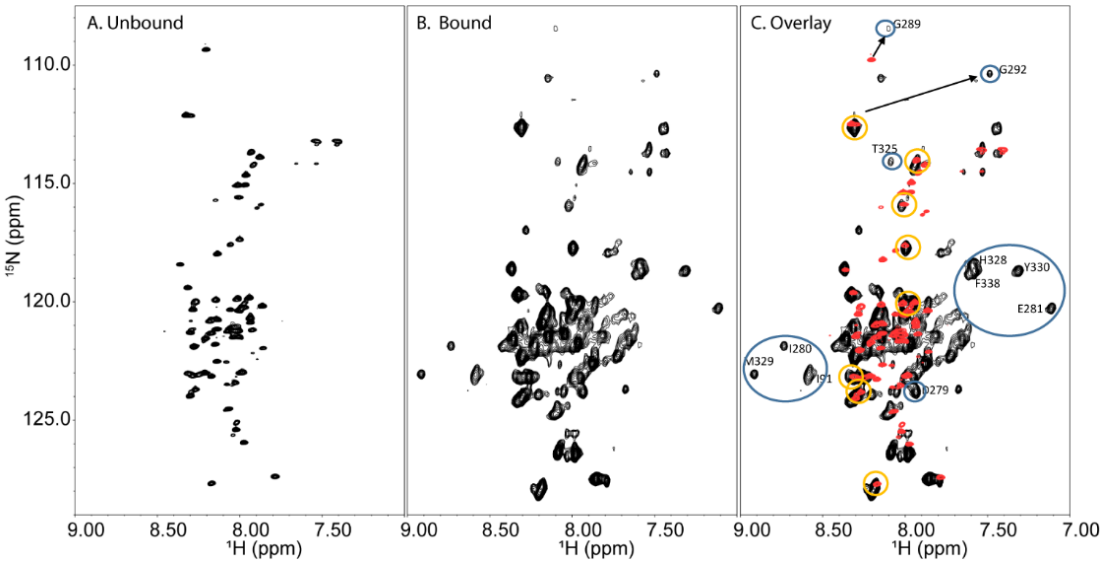

Figure A1.3. NMR spectra of bound and unbound QT2-4 at 40° C. A) ^{15}N TROSY spectrum of unbound QT2-4. B) ^{15}N TROSY spectrum of QT2-4 bound to LC8 (1:4). C) Overlay of A (red) and B (black). Blue circles highlight example resonances that have moved in the bound spectrum. Orange circles highlight example resonance that are not perturbed in the bound spectrum. Arrows specifically show the perturbation of glycines 289 and 292 in QT3 that are perturbed by LC8 binding.

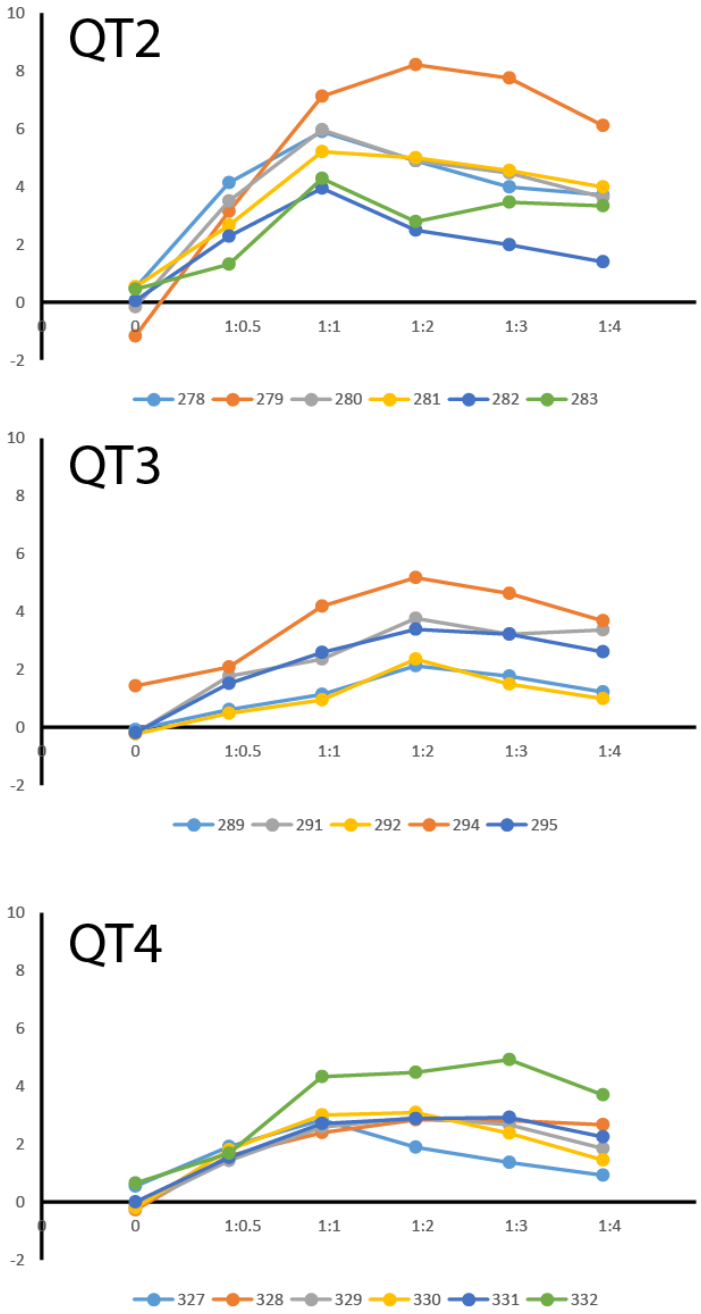

Figure A1.4. NMR titration of QT2-4 with increasing concentrations of LC8. Intensities of 15N-HSQC resonances are shown as a function of LC8 ratio. Results are grouped into separate plots for each LC8 binding motif as indicated in the figure.

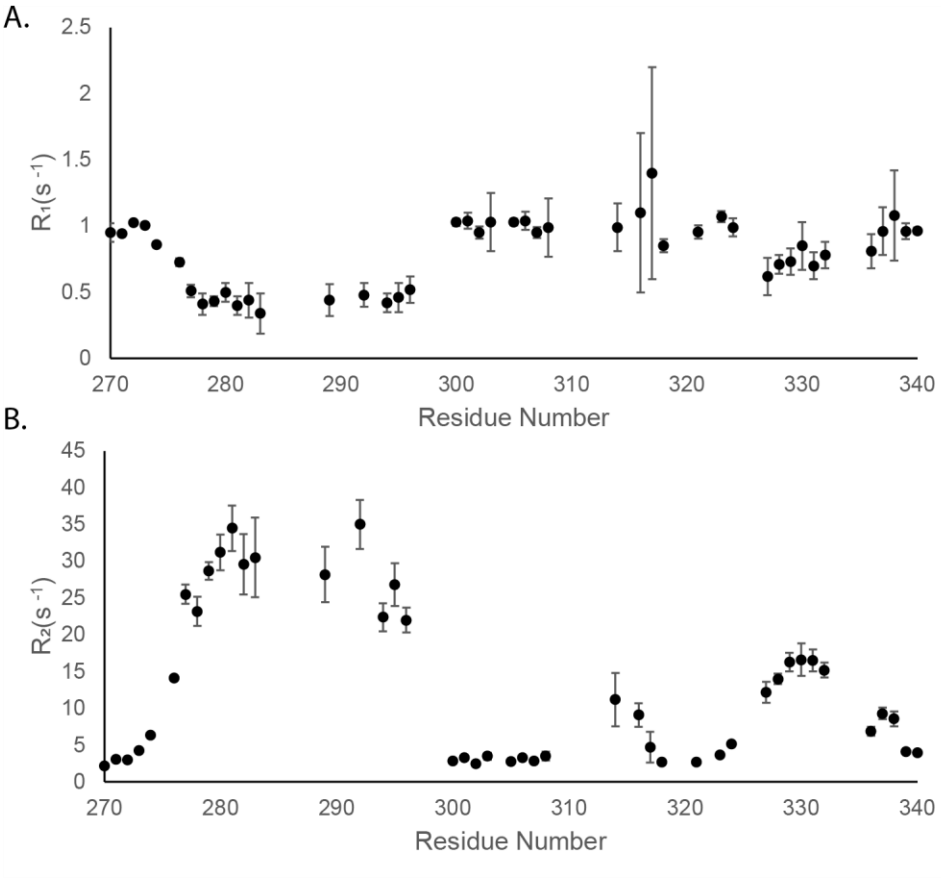

Figure A1.5. R_1 and R_2 relaxation rates for QT2-4 bound to LC8. A) ^{15}N R_1 B) ^{15}N R_2.

Appendix 2

Multivalency, autoinhibition, and protein disorder in the regulation of interactions of dynein intermediate chain with dynactin and the nuclear distribution protein – supplemental material

Kayla A. Jara, Nikolaus M. Loening, Patrick N. Reardon, Zhen Yu, Prajna Woonnimani, Coban Brooks, Cat H. Vesely, and Elisar J. Barbar.

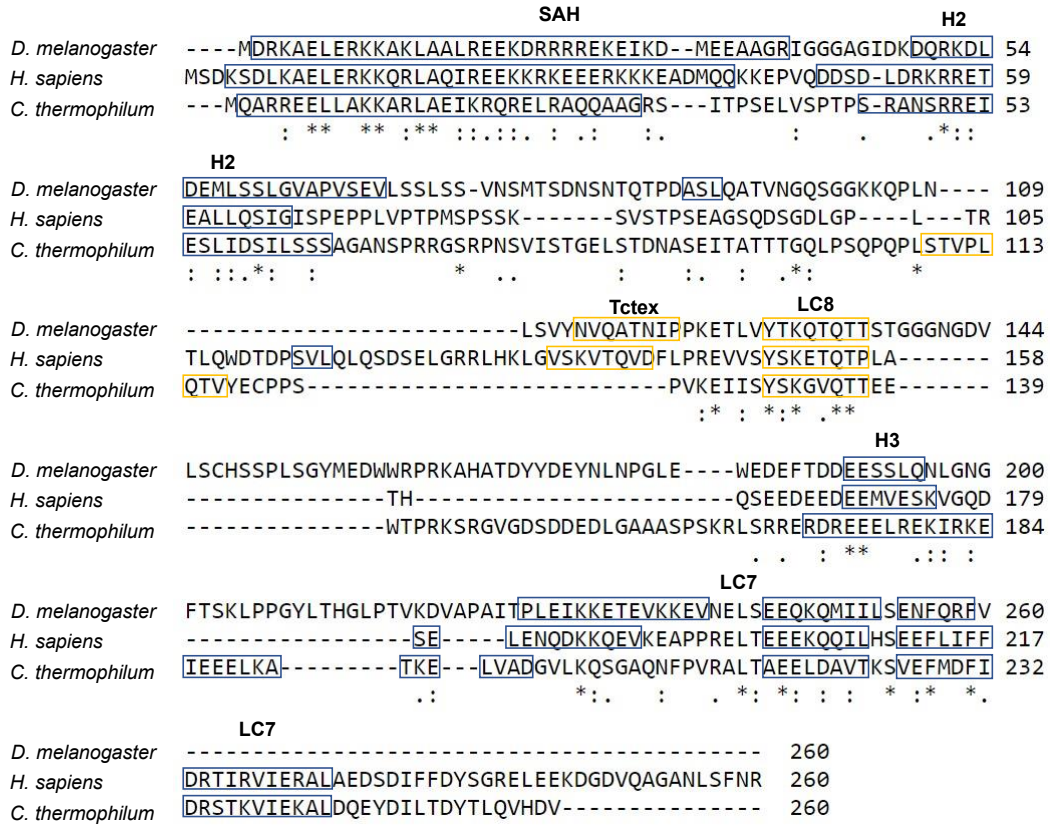

Figure A2.1. IC sequence alignment. Alignment of the first 260 residues of IC from Drosophila, human, and CT using the MAFFT alignment program [180]. Identical (asterisk), strongly similar (colon), and weakly similar (period) residues are shown at the bottom of each alignment. Known or predicted α-helical secondary structure (SAH, H2, H3, and the LC7 binding domain) is highlighted in blue, with the darker shade of blue indicating stronger prediction. Known or predicted binding sites for Tctex and LC8 are highlighted in yellow [49].

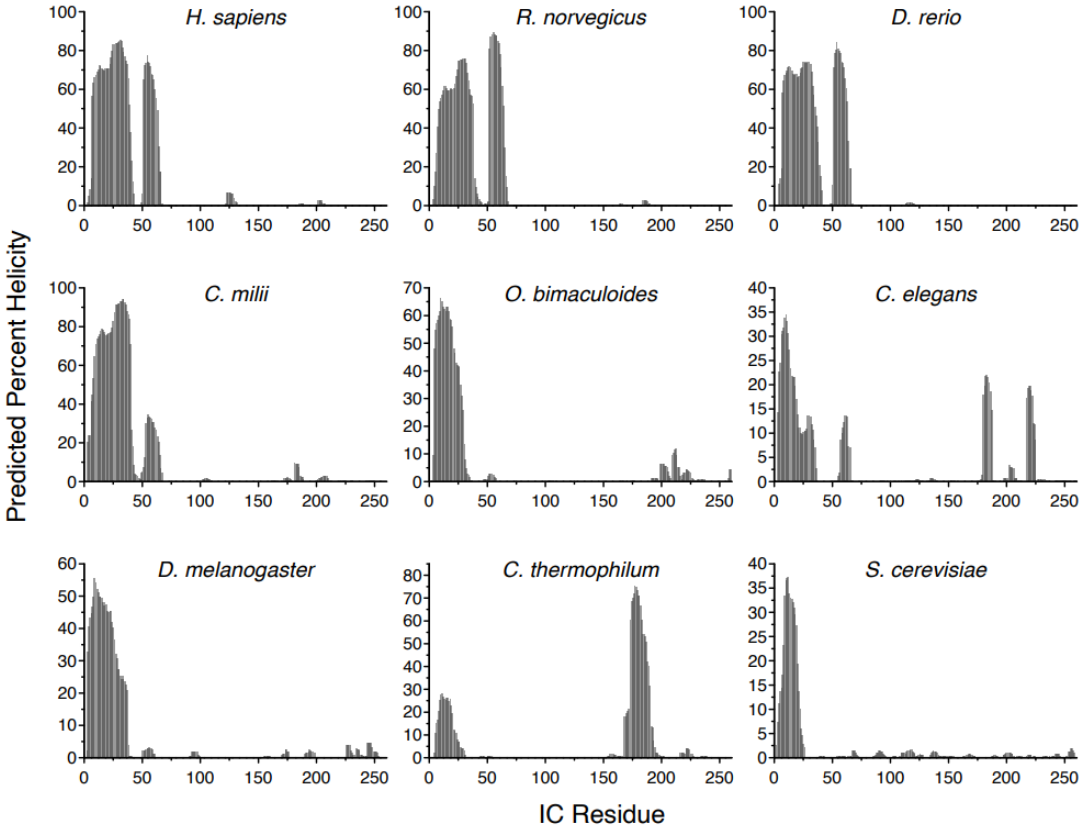

Figure A2.2. Predicted percent Helicity in IC across species. Residue-level percent helicity predictions generated using the Agadir algorithm for the first 260 amino acids of IC from *H. sapiens* (human), *R. norvegicus* (rat), *D. rerio* (zebrafish), *C. milii* (Australian ghostshark), *O. bimaculoides* (Californian two-spot octopus), *C. elegans* (nematode), *D. melanogaster* (fruit fly), *C. thermophilum* (CT, thermophilic fungus), and *S. cerevisiae* (yeast) [108,109,164–166].

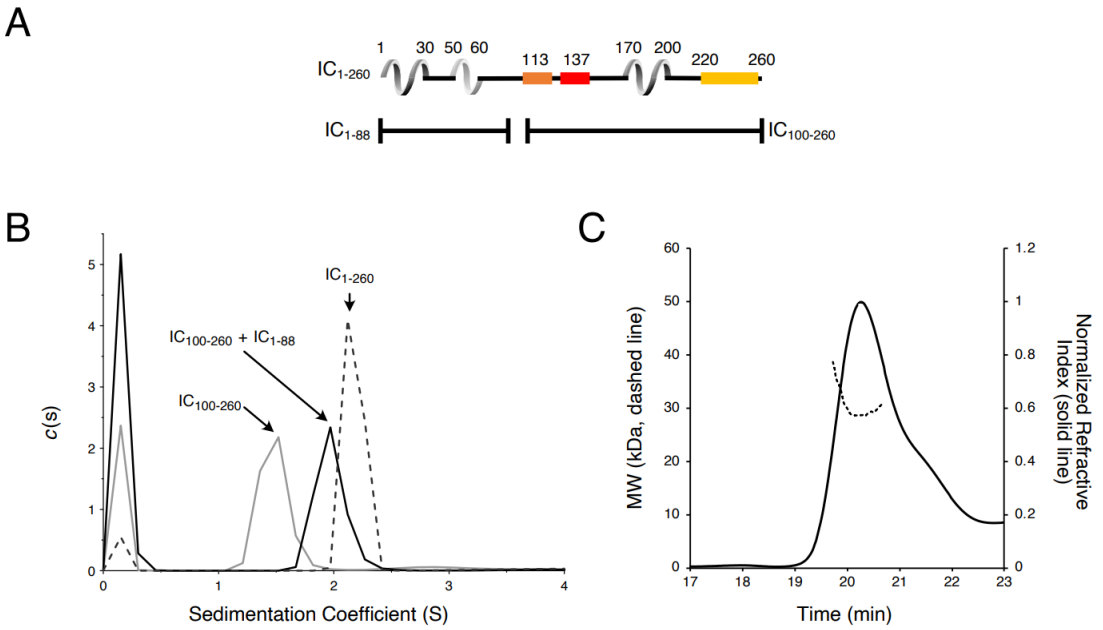

Figure A2.3 IC_{1-88} and $IC_{100-260}$ binding by SV-AUC. A) Domain architecture diagrams of IC_{1-260}, IC_{1-88}, and $IC_{100-260}$. B) SV-AUC experiments of $IC_{100-260}$ (grey), $IC_{100-260}$ mixed with IC_{1-88} at a 1:2 molar ratio (black), and IC_{1-260} (grey dashes). C) The estimated mass of $IC_{100-260}$/IC_{1-88} complex from MALS is 30.3 kDa, which indicates a 1:1 binding stoichiometry.

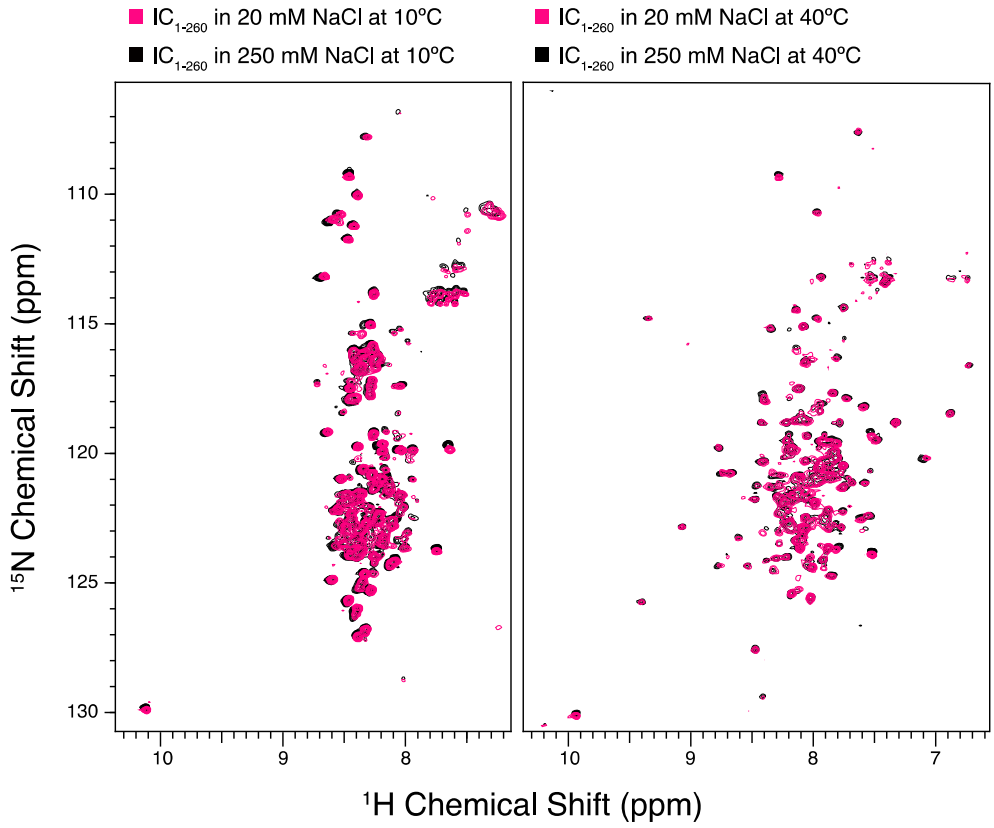

Figure A2.4. NMR spectra of IC are unaffected by salt concentration. ^{1}H-^{15}N TROSY spectra of IC_{1-260} in 20 mM NaCl (black) overlaid with IC_{1-260} in 250 mM NaCl (pink) at both 10ºC (left) and at 40ºC (right).

Appendix 3

Phosphomimetic mutations and Saturation Transfer NMR in the study of mammalian IC binding interactions

Kayla A. Jara and Nikolaus Loening

Overview

Previously published work (Jie et. al. 2017) identified that a phosphomimetic mutation in the mammalian IC-2C isoform at serine residue 84 (S84D) caused reduced binding to a coil-coiled region of dynactin ($p150^{Glued}$), while binding to the Nuclear Distribution protein (NudE) remained unaffected. The combination of ITC, NMR, and CD data lead to a proposed a model in which IC-2C becomes compact upon phosphorylation due to electrostatic interactions of the N-terminal p150/NudE binding site (SAH) and the further C-terminal serine-rich region. The aims of this work were to study the effects of similar phosphomimetic mutations (T89E and S91D) in the serine-rich region to determine if the effect on $p150^{Glued}$ binding is phosphorylation site specific. Further, to determine any differences in $p150^{Glued}$ and NudE binding to IC-2C, saturation transfer NMR was employed to better map the exact binding regions within IC-2C.

Summary and Outlook

Summary

Mammalian IC-$2C_{1-96}$ and subsequent phosphomimetic mutations at residues S84, T89, and S91 were successfully expressed and characterized by CD, ITC, and NMR, including binding experiments with $p150^{Glued}$ and NudE.

CD analysis shows similar α-helical secondary structure among the wild-type (WT) IC-$2C_{1-96}$ and the mutants. Over a temperature range of 5-35˚C, the S84D mutant maintains secondary structure in a similar manner as WT, while the T89E mutant appears to be slightly more temperature stable, and the S91D mutant appears to be slightly less temperature stable, beginning to lose secondary structure between 25-30˚C (Figure A3.1). This data would suggest that introducing phosphomimetic mutations into the serine-rich region of IC-$2C_{1-96}$ does not induce significant structural changes.

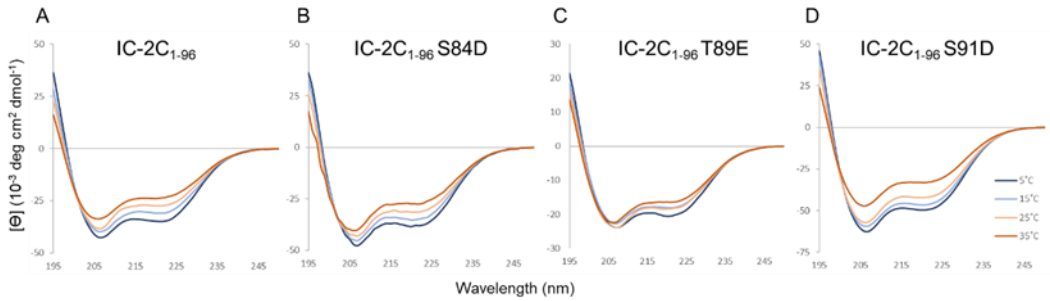

Figure A3.1. Secondary structure and stability analysis of IC-2C$_{1-96}$ and phosphomimetic mutants by CD. Variable temperature CD spectra of (A) IC-2C$_{1-96}$, (B) IC-2C$_{1-96}$ S84D, (C) IC-2C$_{1-96}$ T89E, and (D) IC-2C$_{1-96}$ S91D. The shapes of the spectra for all constructs indicates a mixture of α-helical secondary structure and regions of intrinsic disorder. Loss in structure, or lack thereof, over a temperature range of 5-35ºC (blue for lowest, red for highest) indicates how each construct varies in stability.

The binding interactions of IC-2C$_{1-96}$ and p150^{Glued} were first monitored by ITC (Figure A3.2). The binding thermogram of IC-2C$_{1-96}$ with p150^{Glued} compares well to that reported by Jie et al. (Figure A3.2 A), however, when reproduced here, the interaction of the S84D mutant with p150^{Glued} matches that of WT (Figure A3.2 B). Further, thermograms for the S91D and T89E mutants also indicate binding to p150^{Glued} that is similar to WT. Together, this data indicates no loss of binding to p150^{Glued} caused by phosphomimetic mutations.

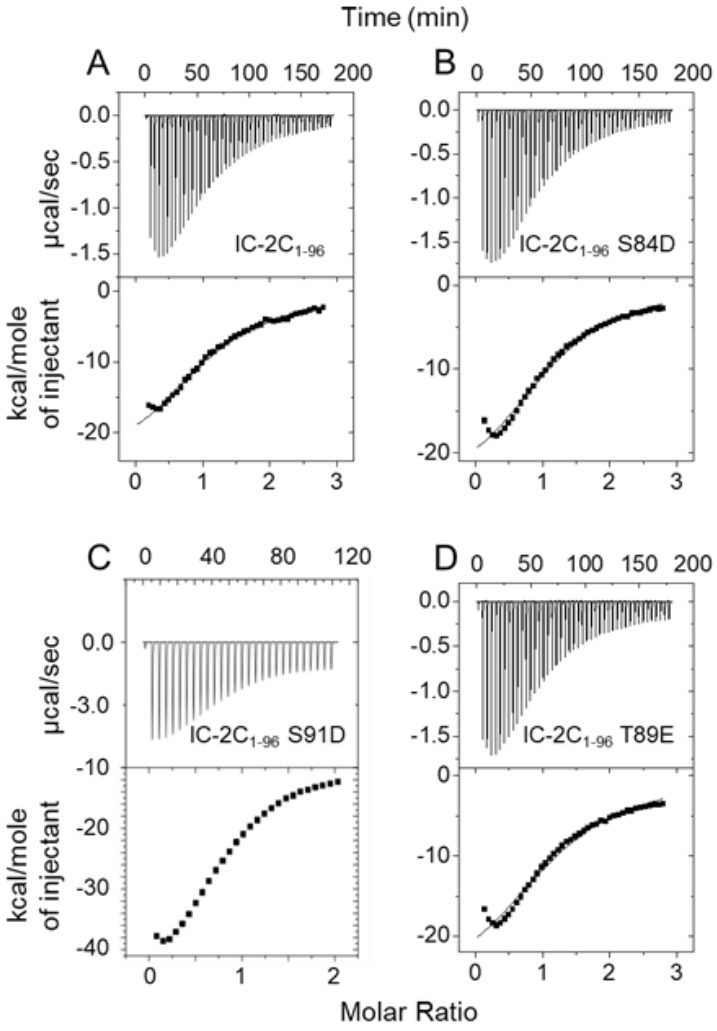

Figure A3.2. Binding of IC-2C$_{1-96}$ and phosphomimetic mutants to p150^{Glued} by ITC. ITC thermograms for p150^{Glued} titrated with (A) IC-2C$_{1-96}$, (B) IC-2C$_{1-96}$ S84D, (C) IC-2C$_{1-96}$ S91D, and (D) IC-2C$_{1-96}$ T89E collected at 25ºC (pH 7.5). All results indicate binding between the IC-2C constructs and p150^{Glued}.

Finally, to better map the binding sites of p150^{Glued} and NudE on IC-2C$_{1-96}$, STD NMR experiments were used. In an STD experiment, peak intensity is caused by the transfer of magnetization from the ligand to the binding partner. In this case, IC-2C$_{1-96}$ was in excess and thus existed predominantly in the free form upon titration of unlabeled p150^{Glued} or NudE. Significant STD effects indicate direct transfer of saturated magnetization. Data collected here at ratios of 50:1 and 30:1 (IC-2C$_{1-96}$: binding partner) indicate that the SAH of IC-2C$_{1-96}$ is directly binding to p150^{Glued}

and NudE, while in both cases the disordered linker regions are not, evident by the high STD intensities for residues 1-20 that are not seen elsewhere. Further, the H2 region does not appear to be directly binding to NudE and only a few weak peaks appear for $p150^{Glued}$ (Figure A3.3 A). This is an important finding as in regular NMR titration experiments, peak attenuation was seen for the SAH and H2 regions upon addition of both $p150^{Glued}$ and NudE. Weak peaks in the linker regions are also seen in the free IC-$2C_{1-96}$ spectrum and are most likely due to the residual fraction of protonated methyls in IC-2C. To further support these findings, CLEANEX experiments were ran at increasing ratios of 5:1 and 1:1 (IC-$2C_{1-96}$: binding partner) and provide more evidence that the SAH region is binding while the linker 2 region is remaining flexible, evident by the high peak intensities seen for resides 80-100 that are not seen elsewhere, except in the free IC-$2C_{1-96}$ spectrum (Figure A3.3 B). Together, these experiments indicate that the serine-rich region home to the phosphomimetic mutations doesn't interact with either binding partner in the WT form.

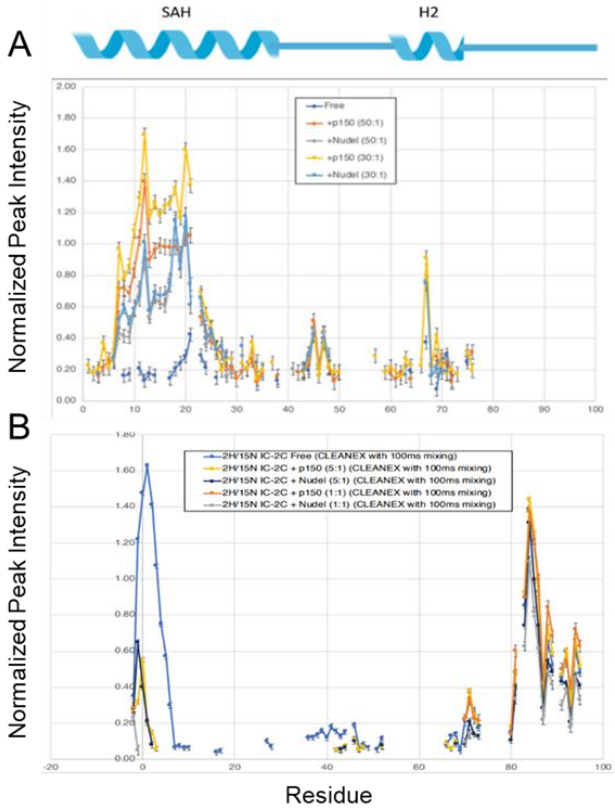

Figure A3.3. Saturation Transfer Difference NMR of IC2C$_{1\text{-}96}$ binding p150^{Glued} and NudE. (A) Peak intensities are plotted for each IC2C residue. Free IC2C (dark blue) represents the basal peak intensity in the free spectrum prior to titration with p150 and NudE at a molar ratio of 50:1 (red and grey respectively) and a molar ratio of 30:1 (yellow and light blue) ratios of ligand to IC2C. These data indicate that the SAH of IC2C is directly binding to p150 and NudE, while in both cases the disordered linker regions are not. (B) CLEANEX peak intensities for free IC2C (light blue), and titration with p150 and NudE at a molar ratio of 5:1 (yellow and dark blue respectively), and a molar ratio of 1:1 (orange and gray). This data supports the STD NMR and the conclusion that the linker 2 region of IC2C is not involved in binding interactions.

Outlook

The data presented above does not support the previously proposed model in which IC-2C becomes compact upon phosphorylation due to electrostatic interactions of the N-terminal p150/NudE binding site (SAH) and the further C-terminal serine-rich region. The aims of this work were to study the effects of similar phosphomimetic mutations (T89E and S91D) in the serine-rich region to determine if

the effect on $p150^{Glued}$ binding is phosphorylation site specific. In these experiments none of the phosphomimetic mutations impacted the binding of IC-2C to $p150^{Glued}$.

The STD NMR experiments presented here serve as a good example of successfully mapping a binding region on an IDP that appears more widely encompassing in traditional NMR titrations. Although similar to the work presented in Chapter 3, the ratios of IC: binding partner vary immensely, providing a different example of successful conditions depending on the type of complex being studied. In future experiments, the same type of characterization but with the incorporation of phosphoserine by genetic code expansion should be explored as although still widely accepted, phosphomimic mutations are often not representative of true phosphorylation. Further, from my work with CTIC (presented in Chapter 5), I would also suggest the use of a longer IC-2C construct to determine if any effects seen for $IC\text{-}2C_{1\text{-}96}$ remain in the context of the entire disordered N-terminal domain or full-length protein.

Experimental Procedures

Mutagenesis and Protein Production

Mammalian $IC\text{-}2C_{1\text{-}96}$ (residues 1-96), $p150^{Glued}$ (residues 382-531), and NudE (residues 1-189) were cloned into pET vectors with an His_6 tag. An internal tobacco etch virus (TEV) protease cleavage site was included in the IC-2C and $p150^{Glued}$ constructs to allow the removal of the His_6 tag. All IC-2C mutations were made using the Quikchange Mutagenesis kit (Agilent). All mutations were verified via Sanger sequencing and then transformed into competent Rosetta (DE3) *E.coli* cells (Merck KGaA, Darmstadt, Germany) for protein expression.

Bacterial cultures for protein expression were grown in Luria-Bertani medium at 37˚C to an optical density (A_{600}) of 0.6-0.7 followed by induction with 0.4 mM isopropyl-β-D-thiogalactopyranoside for 16-18 hr at 26˚C. Cells were harvested, lysed by sonication, and centrifuged to clarify the lysate. Soluble fractions were purified with TALON's Metal Affinity Resin (Takara Bio USA, Mountain View, CA). His_6 tags were cleaved with TEV protease and all proteins were either further

purified using a Superdex 75 (Cytiva, Marlborough, MA) size-exclusion chromatography (SEC) column. Concentrations for IC constructs were determined using absorbance at 205 nm and the concentration of $p150^{Glued}$ and NudE were determined using the absorbance at 280 nm. All proteins were stored at 4˚C and used within one week.

Circular Dichroism

CD measurements were made using a JASCO (Easton, Maryland) J-720 spectropolarimeter. Spectra were recorded on a JASCO 720 spectropolarimeter. Samples consisted of proteins at concentrations of 5-10 µM in 10 mM sodium phosphate buffer (pH 6). All experiments were done using a 400 µL cuvette with a path length of 0.1 cm. Data was taken at 5, 15, 25, and 35˚C and the data shown are the average of three scans.

Isothermal Titration Calorimetry

ITC experiments were conducted using a MicroCal VP-ITC microcalorimeter (Malvern Panalytical, Malvern, United Kingdom). All experiments were performed at 25˚C in buffer composed of 50 mM sodium phosphate, 50 mM sodium chloride, and 1 mM sodium azide (pH 7.2-7.5). Each experiment involved an initial 2µL injection, followed by 26 (10µL) injections. The titrants (IC) were in concentrations of the range 0.25-0.30mM, while the cell protein ($p150^{Glued}$) was in the range of 20-30µM. Data was processed using Origin 7.0 and fit to a single-site binding model.

NMR measurements and analysis

For perdeuteration, *E. coli* Rosetta DE3 cells, transformed with IC-2C, were grown in Luria broth prepared in 99.9% D2O overnight. Modified M9 minimal medium prepared with 99.9% D2O was inoculated from the overnight culture. ^{13}C and ^{15}N were supplied with uniformly labeled [^{2}H-^{13}C] glucose (0.01 M) and [^{15}N] ammonium chloride (0.02 M), respectively. The cultures were grown and purified as mentioned above.

NMR experiments were carried out on an 800-MHz Bruker Avance III HD NMR spectrometer equipped with a 5-mm triple resonance (HCN) cryogenic probe. NMR samples were prepared in 10 mM sodium phosphate, 50 mM NaCl, 1 mM sodium azide, pH 6 buffer at a final IC-2C concentration of 0.6 mM. The samples also contained a protease inhibitor mix (Roche Applied Science), 10% D_2O, and 0.2 mM 2-2 dimethylsilapentane-5-sulfonic acid (DSS) for chemical shift referencing. Assignments of IC-2$C_{1\text{-}96}$ had been obtained previously. For the saturation transfer difference experiments, saturation was applied during the relaxation delay of a 2D ^{15}N TROSY. Samples were made with both $p150^{Glued}$ and NudE at molar ratios of 50:1, 30:1, and 15:1 (IC:binder). The accessibility of IC2$C_{1\text{-}96}$ amide protons to exchange with the solvent was determined by measuring peak volumes in a Fast HSQC spectrum with a 100 ms CLEANEX-PM mixing period. CLEANEX experiments were also ran upon the addition of $p150^{Glued}$ and NudE at molar ratios of 5:1 and 1:1. NMR data were processed using TopSpin 3.6 (Bruker) and NMRPipe. Peak assignment was performed using CCPN Analysis 2.5.2.