AN ABSTRACT OF THE DISSERTATION OF

Jessica Lee Morgan for the degree of Doctor of Philosophy in Biochemistry & Biophysics presented on June 4, 2015.

Title: Characterization of the Structure and Dynamics of the Intrinsically Disordered Dynein Intermediate Chain and its Interactions with Regulatory Binding Partners.

Abstract approved: ______________________________________________________

Elisar J. Barbar

Cells represent microcosms of spatial and temporal structural organization, with the achievement of internal spatial organization relying upon a collection of macromolecular motor complexes to transport and localize components throughout the cell. Cytoplasmic dynein is one such motor complex, and is the principal ATP-dependent motor for retrograde transport along microtubules in the cell. The large (~1.2 MDa) cytoplasmic dynein complex is comprised of multiple protein subunits, including two copies of the intermediate chain (IC), the N-terminal half (‘N–IC’) of which, is central to the dynein cargo attachment sub-domain. N–IC is a prototypical example of the intrinsically disordered protein (IDP) class, serving as a primarily disordered polybivalent molecular scaffold for its numerous binding partners (including regulators of the dynein motor complex), itself often becoming more ordered upon binding interaction. This dissertation presents studies aimed at the biophysical characterization of N–IC itself and also its interactions with several of its binding partners to elucidate structure–dynamics–function relationships, and to gain insights into how these binding interactions might be regulated, as these protein–protein interactions can ultimately determine the sub-cellular targeting and function of dynein within the cell.

Chapter 1 opens with a brief introduction to IDPs, as they are a relatively new class of proteins whose recognition and presence in the reported scientific literature have grown exponentially in the past decade-and-a-half. The prevalence of intrinsically disordered proteins and protein regions in the proteome, the peculiarities in their binding interactions with partners, and their functionality in the absence of fixed, three-dimensional structure are outlined. This sub-section is followed by a thorough review of cytoplasmic dynein motor functions in the cell and its protein subunit composition, with particular emphasis placed upon the intermediate chain—the central protein of this thesis. A thorough review is also
given for regulatory complexes of the dynein motor including dynactin, ZW10/RZZ, and NudE/EL. Chapter 2 presents a review of the premier biophysical technique for the study and characterization of IDPs—solution-state protein NMR spectroscopy—and enumeration of particular considerations that must be taken into account (stemming largely from the conformational dynamics and motional freedom of these polypeptide chains) in the interpretation of data garnered from these techniques when applied to IDPs or unfolded proteins.

Chapters 3 and 4 present original research work on the characterization of the N-terminal 143 residues (IC:1–143) of the Drosophila melanogaster dynein intermediate chain and its interactions with binding partners. The work presented in Chapter 3 demonstrates that, although predominately disordered, IC:1–143 deviates from random coil behavior, particularly in the form of two regions of α-helical structure near the N-terminus. Furthermore, these helical segments were determined to exist in two non-contiguous segments of IC that interact with the regulator dynactin p150Glued protein, and the results of this study provided insights into the biophysical basis by which the IC–p150Glued interaction (and thus the association between dynein and dynactin) might be regulated. In Chapter 4 a more detailed conformational and dynamical examination was performed on IC:1–143 using NMR residual dipolar couplings (RDCs) and paramagnetic relaxation enhancement (PRE) experiments, revealing unprecedented detail concerning further deviations of this protein from random coil behavior. The Tctex1 and LC8 light chains binding regions in IC exhibit enhanced polyproline II conformational sampling (relative to a random coil description), and the IC:1–143 protein exhibits further deviations from random coil behavior in the form of significant transient tertiary structure, shedding light on how the association state of IC with dynactin p150Glued vs. NudE might be controlled in Drosophila melanogaster when both regulatory proteins are simultaneously present.

Chapter 5 presents a summary of the key findings of the work presented in this dissertation, as well as an assessment of outstanding questions in this field and proposed work to help fill these gaps in knowledge. Overall, the results presented in this dissertation provide detailed descriptions of the structure and dynamics of the N-terminal half of N–IC (which constitutes a ‘hotbed’ for binding activity), revealing both subtle and pronounced deviations from random coil behavior in the form of secondary structure of varying degrees, as well as transient tertiary structure, all of which underlie interactions of IC with its binding partners, and also provide insights into the biophysical bases for regulation of these binding interactions.
Characterization of the Structure and Dynamics of the Intrinsically Disordered Dynein Intermediate Chain and its Interactions with Regulatory Binding Partners

by
Jessica Lee Morgan

A DISSERTATION
submitted to
Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 4, 2015
Commencement June 2015

APPROVED:

Major Professor, representing Biochemistry & Biophysics

Chair of the Department of Biochemistry & 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.

Jessica Lee Morgan, Author
ACKNOWLEDGEMENTS

I thank my research supervisor, Elisar Barbar, particularly for her help with my application to attend the 2014 National Graduate Student Symposium at St. Jude Children’s Research Hospital in Memphis, Tennessee, and also for the exposure and access to the terrific contacts she made while she was on sabbatical in western Europe during my third year of graduate school. I also want to heartily thank all the other members of my graduate committee, including: Dr. Victor L. Hsu and Dr. P. Andrew Karplus (both of the Dept. of Biochemistry & Biophysics), Dr. Barbara Taylor (Zoology, and Dept. of Integrative Biology), and my Graduate Council Representative (GCR), Dr. Kevin P. Gable (Dept. of Chemistry). I thank Victor Hsu for not allowing me to slip into complacency when it came to the interpretation of my own NMR data, and for causing me to have to think more deeply about the physical phenomena underlying the experiments and data I collected. I thank Andy Karplus for introducing me at my 3rd year departmental seminar and for his generosity in helping me to prepare for that talk; I also thank Andy and the members of his lab for their kind and generous assistance with multiple attempted protein crystallizations and projects over the years. I thank Barb Taylor for challenging me to see and recognize the ‘bigger picture’ and how my work fits within this greater scheme, and for helping me to develop the mentality of now, always asking myself the questions: “So, what? Why is this important? Why is this particular project worth pursuing?” I thank Kevin Gable for his scientific input and for helping to make sure that all the requirements and documentation were squared-away with the Graduate School. The members of my graduate committee have had a profound impact on my development as a Scientist and as a ‘scientific thinker.’ I thank you all.

I also want to thank Dr. Kerry McPhail (OSU College of Pharmacy) for her incredible generosity and her unending willingness to meet me over at the new 500 MHz instrument in the Chemistry building, time-after-time (even on weekends and at sometimes odd hours), as we tried-out so many different equipment setups, all in order that I should be able to collect data I needed for my research. I also want to thank Rodger Kohnert (Dept. of Chemistry), firstly for entertaining my inquiry as to whether or not the 700 MHz instrument could be set up for protein work, and then bearing with me through the trial-and-error of getting things configured and calibrated for the type of work I wanted to do. After the loss of our own 600 MHz NMR instrument, I could not have collected the NMR data and finished the work presented in Chapter 4 of this dissertation without Kerry and Rodger’s help in using
the remaining instrumentation on campus. I want to thank our various collaborators in western Europe for their eagerness to help, to teach and share their knowledge and expertise, and to engage in exciting collaborative works with us. I also want to thank all of the Karplus Lab folks who participated with our lab in a weekly Structure Journal Club; I truly appreciated the opportunities I had to present my (NMR-heavy) research work to these “X-ray/Structural Biology people,” and appreciated even more, all of the helpful and meaningful feedback furnished over the years.

I thank my former labmates, Yujuan Song and Jing Jie. I thank Yujuan for being both a friend and a mentor to me in the lab; I appreciated the sharing of her wisdom and experience both in life and in all things of a technical nature in the lab. I thank Jing for her camaraderie (and commiseration over the numerous different challenges presented by some of the peskier proteins we each encountered over the years!); I also appreciated our numerous afternoon scientific discussions across the bench. I also want to thank Monsieur Antoine Gruet who, though only with us for a year, brought much-needed joy, laughter, and sunshine into our lab. Last, but not least, I want to thank all of my friends and family—both immediate and extended. Without your love and constant encouragement, I never could have survived all of the trials and tribulations. Thank-you for believing in me, even when I didn’t believe in myself. I love you all.
CONTRIBUTION OF AUTHORS

Yujuan Song contributed molecular biology work to create many of the constructs for proteins described in Chapters 3 and 4. Jonathan Yih and Afua Nyarko contributed some of the IC\textsubscript{TL} protein samples used in NMR experiments described in Chapter 4. Gregory Benison collected the NMR dynamics data for IC\textsubscript{TL} as described in Chapter 4. Malene Ringkjøbing Jensen (Institut de Biologie Structurale (IBS) at Grenoble, France) collected the RDC NMR data as described in Chapter 4. Valéry Ozenne (IBS at Grenoble, France) performed the flexible-meccano/ASTEROIDS ensemble calculations for IC\textsubscript{TL} as described in Chapter 4. Malene Ringkjøbing Jensen and Martin Blackledge (IBS at Grenoble, France) were involved in the design and analysis of experiments pertaining to the ensemble calculations for IC\textsubscript{TL} reported in Chapter 4. Initial protein crystallization screens of the CT-LIC protein were set up by the staff of the High-Throughput Crystallization (HTX) Lab facility (European Molecular Biology Lab (EMBL) at Grenoble, France); Dr. Andrew McCarthy (EMBL at Grenoble, France) collected the synchrotron data on CT-LIC protein crystals from these screens and performed initial analysis of this data reported in Appendix 3. Elisar Barbar was involved in the design and analysis of experiments presented in this dissertation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter 1: Introduction, Literature Review, and Thesis Overview</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intrinsically Disordered Proteins—A Brief Introduction</td>
</tr>
<tr>
<td>Biological Roles and Functions of Protein Intrinsic Disorder</td>
</tr>
<tr>
<td>Examples and Modes of Disordered Proteins Binding to Partners</td>
</tr>
<tr>
<td>Benefits and Functional Advantages of Protein Disorder</td>
</tr>
<tr>
<td>The Cytoplasmic Dynein Motor Complex</td>
</tr>
<tr>
<td>Functions of Cytoplasmic Dynein</td>
</tr>
<tr>
<td>Multi-subunit Composition of the Cytoplasmic Dynein Motor</td>
</tr>
<tr>
<td>The Dynein Heavy Chain (HC)</td>
</tr>
<tr>
<td>The Dynein Light Intermediate Chain (LIC)</td>
</tr>
<tr>
<td>The Dynein Intermediate Chain (IC)</td>
</tr>
<tr>
<td>The Dynein Light Chain Tetex1</td>
</tr>
<tr>
<td>The Dynein Light Chain LC7</td>
</tr>
<tr>
<td>The Dynein Light Chain LC8</td>
</tr>
<tr>
<td>Disorder in Dynein—the Intermediate Chain</td>
</tr>
<tr>
<td>Isoform Diversity of the Intermediate Chain</td>
</tr>
<tr>
<td>Regulatory Phosphorylation of the Intermediate Chain</td>
</tr>
<tr>
<td>Regulators of the Cytoplasmic Dynein Motor Complex</td>
</tr>
<tr>
<td>Dynactin</td>
</tr>
<tr>
<td>Rod-ZW10-Zwilch (RZZ)</td>
</tr>
<tr>
<td>Nuclear distribution protein E (NudE) and NudE-like (NudEL)</td>
</tr>
<tr>
<td>Work Presented in this Dissertation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 2: The Use of Solution-State Protein NMR Spectroscopy in the Study and Characterization of Intrinsically Disordered Proteins</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secondary Chemical Shifts</td>
</tr>
<tr>
<td>Scalar Couplings</td>
</tr>
<tr>
<td>Residual Dipolar Couplings (RDCs)</td>
</tr>
<tr>
<td>Nuclear Overhauser effects (NOEs) and Paramagnetic Relaxation Enhancements (PREs)</td>
</tr>
</tbody>
</table>
## TABLE OF CONTENTS (Continued)

### Chapter 3: Structural Dynamics and Multiregion Interactions in Dynein–Dynactin Recognition

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>89</td>
</tr>
<tr>
<td>Introduction</td>
<td>90</td>
</tr>
<tr>
<td>Experimental Procedures</td>
<td>93</td>
</tr>
<tr>
<td>Results</td>
<td>97</td>
</tr>
<tr>
<td>Resonance Assignments and Secondary Structure of IC:1–143</td>
<td>97</td>
</tr>
<tr>
<td>Dynamics of IC:1–143</td>
<td>100</td>
</tr>
<tr>
<td>Mapping the IC Binding Interface with p150\textsubscript{Glued}</td>
<td>103</td>
</tr>
<tr>
<td>Dynamics of IC:1–143 Bound to p150\textsubscript{Glued}</td>
<td>106</td>
</tr>
<tr>
<td>Thermodynamics of p150\textsubscript{Glued}–IC Interactions</td>
<td>106</td>
</tr>
<tr>
<td>Dynamics of IC:1–143 Bound to p150\textsubscript{Glued} and Light Chains</td>
<td>108</td>
</tr>
<tr>
<td>Discussion</td>
<td>108</td>
</tr>
<tr>
<td>Residual Structure in IC:1–143</td>
<td>111</td>
</tr>
<tr>
<td>Multiregion Binding Footprint for p150\textsubscript{Glued}\textsubscript{221-509} on IC:1–143 of the Intermediate Chain</td>
<td>111</td>
</tr>
<tr>
<td>Structure and Dynamics of Assembled IC</td>
<td>113</td>
</tr>
<tr>
<td>Functional Implications of Disorder in Assembled IC</td>
<td>114</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>115</td>
</tr>
</tbody>
</table>

### Chapter 4: The Role of Residual Structure in Dynein Intermediate Chain in Complex Assembly and Regulation

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>116</td>
</tr>
<tr>
<td>Introduction</td>
<td>117</td>
</tr>
<tr>
<td>Experimental Procedures</td>
<td>118</td>
</tr>
<tr>
<td>Results</td>
<td>121</td>
</tr>
<tr>
<td>Experimental Characterization of Residual Secondary Structure in IC\textsubscript{TL}</td>
<td>126</td>
</tr>
<tr>
<td>Ensemble of Structures Populated by IC\textsubscript{TL} in Solution</td>
<td>130</td>
</tr>
<tr>
<td>Dynamical Characterization of IC\textsubscript{TL}</td>
<td>131</td>
</tr>
<tr>
<td>Design and Characterization of Spin-Labeled IC\textsubscript{1LL} Variants</td>
<td>135</td>
</tr>
<tr>
<td>Detection of Long-Range Interactions within IC\textsubscript{1LL}</td>
<td>139</td>
</tr>
<tr>
<td>Characterization of Long-Range Interactions in Assembled IC\textsubscript{1LL}</td>
<td>141</td>
</tr>
<tr>
<td>Discussion</td>
<td>144</td>
</tr>
<tr>
<td>Residual Secondary Structure in IC\textsubscript{TL}—Enhanced Polyproline II Content</td>
<td>145</td>
</tr>
<tr>
<td>Functional Implications of Enhanced Polyproline II Content in IC\textsubscript{TL}</td>
<td>148</td>
</tr>
</tbody>
</table>
**TABLE OF CONTENTS (Continued)**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Significant Transient Tertiary Structure in apo IC1$_{LL}$</td>
<td>151</td>
</tr>
<tr>
<td>Altered Sampling of Transient Tertiary Structure in Assembled IC1$_{LL}$</td>
<td>153</td>
</tr>
<tr>
<td>Functional Aspects and Implications of Transient Tertiary Structure</td>
<td>153</td>
</tr>
<tr>
<td>in IC1$_{LL}$</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td>156</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>156</td>
</tr>
</tbody>
</table>

**Chapter 5: Concluding Discussion and Outlook**

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Summary</td>
<td>158</td>
</tr>
<tr>
<td>Main Conclusions</td>
<td>158</td>
</tr>
<tr>
<td>Future Work</td>
<td>161</td>
</tr>
</tbody>
</table>

**Bibliography**

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>166</td>
</tr>
</tbody>
</table>

**Appendices**

| Appendix 1: The Role of Residual Structure in Dynein Intermediate Chain in Complex Assembly and Regulation—supplemental material | 193  |
| Appendix 2: Cloning and preparation of the full-length *Drosophila melanogaster* Zeste-white 10 (ZW10) protein and its interactions with the Dynein Intermediate Chain                                      | 204  |
| Appendix 3: Preparation, characterization, and crystallization of the dynein light intermediate chain (LIC) protein from *Chaetomium thermophilum*                                    | 215  |
| Appendix 4: NMR backbone chemical shift assignments and dynamics of the *Drosophila melanogaster* dynein LC7 protein—free and IC7-bound state NMR data                                         | 221  |
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>7</td>
</tr>
<tr>
<td>1.2</td>
<td>15</td>
</tr>
<tr>
<td>1.3</td>
<td>23</td>
</tr>
<tr>
<td>1.4</td>
<td>28</td>
</tr>
<tr>
<td>1.5</td>
<td>34</td>
</tr>
<tr>
<td>1.6</td>
<td>36</td>
</tr>
<tr>
<td>1.7</td>
<td>40</td>
</tr>
<tr>
<td>1.8</td>
<td>45</td>
</tr>
<tr>
<td>1.9</td>
<td>48</td>
</tr>
<tr>
<td>2.1</td>
<td>57</td>
</tr>
<tr>
<td>2.2</td>
<td>67</td>
</tr>
<tr>
<td>2.3</td>
<td>76</td>
</tr>
<tr>
<td>2.4</td>
<td>85</td>
</tr>
<tr>
<td>3.1</td>
<td>94</td>
</tr>
<tr>
<td>3.2</td>
<td>98</td>
</tr>
<tr>
<td>3.3</td>
<td>99</td>
</tr>
<tr>
<td>3.4</td>
<td>101</td>
</tr>
<tr>
<td>3.5</td>
<td>104</td>
</tr>
<tr>
<td>3.6</td>
<td>107</td>
</tr>
<tr>
<td>3.7</td>
<td>109</td>
</tr>
<tr>
<td>3.8</td>
<td>112</td>
</tr>
<tr>
<td>4.1</td>
<td>119</td>
</tr>
<tr>
<td>4.2</td>
<td>128</td>
</tr>
<tr>
<td>4.3</td>
<td>132</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>4.4 Dynamics of IC_{TL}</td>
<td>133</td>
</tr>
<tr>
<td>4.5 Characterization of transient tertiary structure sampled by IC_{1L}-Cys mutants</td>
<td>137</td>
</tr>
<tr>
<td>4.6 Characterization of transient tertiary structure sampled by IC_{1L} in its bound states</td>
<td>142</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>A1.1</td>
<td>194</td>
</tr>
<tr>
<td>A1.2</td>
<td>195</td>
</tr>
<tr>
<td>A1.3</td>
<td>196</td>
</tr>
<tr>
<td>A1.4</td>
<td>197</td>
</tr>
<tr>
<td>A1.5</td>
<td>198</td>
</tr>
<tr>
<td>A1.6</td>
<td>200</td>
</tr>
<tr>
<td>A1.7</td>
<td>201</td>
</tr>
<tr>
<td>A1.8</td>
<td>202</td>
</tr>
<tr>
<td>A2.1</td>
<td>207</td>
</tr>
<tr>
<td>A2.2</td>
<td>208</td>
</tr>
<tr>
<td>A2.3</td>
<td>209</td>
</tr>
<tr>
<td>A2.4</td>
<td>210</td>
</tr>
<tr>
<td>A3.1</td>
<td>216</td>
</tr>
<tr>
<td>A3.2</td>
<td>217</td>
</tr>
<tr>
<td>A3.3</td>
<td>218</td>
</tr>
<tr>
<td>A3.4</td>
<td>219</td>
</tr>
<tr>
<td>A3.5</td>
<td>220</td>
</tr>
<tr>
<td>A4.1</td>
<td>222</td>
</tr>
<tr>
<td>A4.2</td>
<td>223</td>
</tr>
<tr>
<td>A4.3</td>
<td>224</td>
</tr>
<tr>
<td>A4.4</td>
<td>225</td>
</tr>
</tbody>
</table>
Dedicated to Grandpa Joe,

The one who first helped to set me upon this path
Chapter 1

Characterization of the Structure and Dynamics of the Intrinsically Disordered Dynein Intermediate Chain and its Interactions with Regulatory Binding Partners

Introduction, Literature Review, and Thesis Overview
Intrinsically Disordered Proteins—A Brief Introduction

Intrinsically disordered proteins (IDPs) function in the absence of a unique, stable, three-dimensional structure and appear to challenge (Wright et. al. 1999) the more than half-century-old classic structure–function paradigm, wherein it is postulated that the amino acid sequence of a protein determines its three-dimensional structure, and that the prior formation of this 3-D structure is required for the function of the protein. Rather than a static three-dimensional structure, IDPs are best described and represented as an ensemble of dynamic and rapidly interconverting conformers. Once thought to be an anomalous and rare occurrence (aside from loops and linker regions observed in crystal structures of globular proteins), it is now known that intrinsic protein disorder pervades the proteomes of all three major kingdoms of life (Dunker et. al. 2000; Ward et. al. 2004a), being the most prevalent in eukaryotes (Pancsa et. al. 2012). In fact, it is conservatively estimated that greater than 30% of eukaryotic proteins contain long disordered regions (excluding loops or short disordered N- and C-termini found in folded globular proteins) at least 30 residues in length; and nearly 20% of eukaryotic proteins contain disordered regions of length greater than 50 residues (Romero et. al. 1998; Ward et. al. 2004a; Pancsa et. al. 2012).

It should be noted that disorder in protein structure can be either global or local: there are those proteins that are completely or almost completely disordered (e.g. disorder predominates throughout the protein), as well as those that consist of a mixture of stably-structured domains and disordered domains. Disordered protein segments can often be punctuated by single elements of secondary structure, with the degree of secondary structure ranging from transient (or, ‘nascent’) to fully-formed, and transient tertiary structure can arise from short-lived intra-chain contacts, sometimes made even between very distant parts of the disordered protein. For proteins in which disorder is in the minority and confined to one or more distinct segments (usually > 30–40 residues) in the primary sequence, these segments are often referred to as Intrinsically Disordered Regions (‘IDRs’) rather than ‘IDPs.’ It is also of note that IDPs and IDRs have an amino acid compositional bias relative to globular proteins: they tend to be enriched in “disorder-promoting” (Williams et. al. 2001) residues such as aspartic acid, lysine, arginine, serine, glutamine, glutamic acid, and proline, and on average, are depleted in such “order-promoting” residues as cysteine, tryptophan, tyrosine, isoleucine, phenylalanine, valine, leucine, histidine, threonine, and asparagine (Theillet et. al. 2013). In other words, IDPs and IDRs tend to exhibit a combination of low overall hydrophobicity and significant charge content (Uversky et. al. 2000).
Contrary to the traditional view that protein function requires and derives from a folded and stable three-dimensional structure, intrinsically disordered proteins and protein regions are indeed quite functional (Dunker et. al. 2002; Fink 2005; Dyson et. al. 2005). Analysis of the correlation between protein disorder and associated functions demonstrates that protein disorder is associated with a broad repertoire of biological functions and processes, but is most highly represented in such categories as: 1) transcription and transcriptional regulation, 2) cell signaling and signal transduction, and 3) cell cycle regulation (Ward et. al. 2004a; Xie et. al. 2007a,b; Vucetic et. al. 2007).

Transcription is tightly regulated by a number of proteins in the cell, especially transcription factors that include activators, repressors, and enhancer-binding factors; in this context, one implication of protein disorder is in the recognition and binding of specific nucleic acid sequences, an example of which, is the bZIP (basic region leucine zipper) DNA-binding domain of the CREB (cyclic AMP response element-binding protein) transcription factor (Minezaki et. al. 2006). The DNA-interacting protein segments of the bZIP homodimer are unstructured and flexible in the free state, but undergo a transition to a more folded state (increase in $\alpha$-helical structure) upon recognition and binding of the CRE (cAMP response element) target DNA sequence (Santiago-Rivera et. al. 1993); it is believed that the flexible and malleable nature seen in the free state of these DNA-binding segments facilitates the scanning of hundreds of thousands of base pairs by the transcription factor, while the local, DNA-induced protein folding upon recognition of a particular sequence helps to confer specificity in binding (Wright et. al. 1999). In addition to protein–DNA interactions, another implication of disorder in the context of transcription and regulation thereof, is in the protein–protein interactions that facilitate the recruitment and assembly of other transcription-related proteins and machinery by transcription factors. Following phosphorylation at serine-133 in the KID (kinase-inducible domain) of CREB, the transcription co-activator CBP (CREB-binding protein) is recruited through binding interaction between its KIX domain and pKID of CREB; the pKID is largely unstructured in its free state but undergoes a coil-to-helix folding transition upon binding to the KIX domain of CBP (Radhakrishnan et. al. 1997, 1998).

In the category of cell signaling and signal transduction, protein disorder again functions in molecular recognition and binding (Iakoucheva et. al. 2002). For instance, programmed cell death (apoptosis) is mediated by interactions between pro- and anti-apoptotic proteins of the Bcl-2 (B-cell lymphoma-2) family, members of which contain either intrinsically disordered regions...
or whole disordered protein domains that are critical to their function (Rautureau et al. 2010). As an example, upon receipt of a death stimulus signal, the intrinsically disordered pro-apoptotic cytosolic Bim protein is activated and then undergoes a coil-to-helix folding transition upon binding to the anti-apoptotic Bcl-2 protein (Hinds et al. 2007) anchored on the mitochondrial outer membrane surface; Bim binding to Bcl-2 liberates previously-sequestered (due to binding to Bcl-2) pro-apoptotic Bak and Bax proteins, which can then homo-oligomerize to form transmembrane channels, leading to the release of factors such as cytochrome c from the mitochondrial inter-membrane space into the cytosol, which leads to activation of caspases and a proteolytic cascade that ultimately destroys the cell from within (Rautureau et al. 2010). An example of the role that protein disorder plays in signal transduction across the cell membrane would be the case of the cytoplasmic signaling (effector) domains of multi-chain immune recognition receptors (MIRRs) (Sigalov et al. 2011). Antigen receptors on T cells have cytoplasmic domains (on separate subunits from the external ligand recognition domains) that are intrinsically disordered and homo-oligomerize; oligomerization of these cytoplasmic domains does not change the random-coil-like conformations observed (Sigalov et al. 2004, 2007). Upon engagement of extracellular receptor domains, phosphorylation of certain Tyrosine residues occurs in the flexible cytoplasmic domains, triggering intracellular signaling cascades. In the case of the Tyrosine-phosphorylated T cell ζ cytoplasmic domain (ζ_cyt), binding interaction is known to occur both with acidic POPG lipids in vesicles and with the well-structured SIV (Simian Immunodeficiency Virus) Nef protein; in both instances, the disorder of ζ_cyt is retained in its bound states (Sigalov et al. 2006, 2008).

In the category of cell cycle regulation: progression through the cell cycle is controlled in part by the activity of cyclin-dependent kinases (Cdns) that are regulated by their cyclin partners, with which they form active heterodimeric complexes; eight Cdk family members (Cdk1–Cdk8) and nine cyclins (A–I) have been identified thus far. Different pairings of the Cdk and cyclin proteins exist and different heterodimeric complexes are associated with progression through various specific transitions in the cell cycle (Nigg 2001). The activities of Cdk–cyclins throughout the cell cycle are regulated by a combination of mechanisms, including binding by Cdk inhibitor proteins (such as p21^{Waf1/Cip1/Sdi1} and p27^{Kip1}) that function in the deactivation (inhibition of the kinase activity) of the Cdk–cyclin complexes (Yoon et al. 2012). The intrinsically disordered p21 and p27 proteins undergo coupled folding and binding upon interaction with their multiple different Cdk–cyclin targets (Kriwacki et al. 1996; Lacy et al. 2004), and it is the structural plasticity of the p21 and p27 proteins that underlies their ability to
accommodate binding to similar, but topologically distinct, Cdk–cyclin heterodimers (Wang et al. 2011a).

Thus, we have seen several examples demonstrating that protein disorder can indeed be quite functional in biological processes within the cell. In addition to the three major biological function categories discussed above, another interesting function ascribed to protein disorder and exhibited by IDPs is that of a molecular scaffold (Cortese et al. 2008). Scaffold proteins can contain a multitude of IDRs that facilitate molecular recognition and binding interaction with several other proteins. In the context of proteins in signaling networks, scaffold proteins selectively assemble and congregate specific proteins within signaling pathways in order to promote and facilitate interactions between them. Further, signaling scaffold proteins can bind proteins that are elements in a multitude of other signaling pathways, thus providing sensitive coordination between alternate signaling pathways; in this sense, such intrinsically disordered scaffold proteins are considered to be a special subclass of ‘hub’ proteins (Dunker et al. 2005; Haynes et al. 2006). A protein that is likely familiar to the reader (due to its implication in human disease such as cancer) that is also an example of an intrinsically disordered signaling scaffold, is the product of the **BRCA1** gene; between the folded, structured RING and BRCT domains at the N- and C-termini, respectively, the central ~1500 residues of BRCA1 are largely disordered and contain the binding regions for DNA and also a plethora of proteins including Rb and p53 (Mark et al. 2005). In addition to signaling pathways, intrinsically disordered scaffold proteins can also function in macromolecular assembly of complexes and machinery--a prime example of which, is the intermediate chain protein of the dynein motor complex (as we shall see below, Figure 1.2).

**Examples and Modes of Disordered Proteins Binding to Partners.**

In general, the functions of IDPs and IDR s involve molecular recognition and binding to partner ligands such as other proteins, nucleic acids, membranes, etc. Several binding modes exist for these interactions, and though folding-coupled-to-binding is a commonly observed mode of interaction (several examples were given above), it is by no means the exclusive mode exhibited by disordered proteins (Uversky 2011a).

In the category of coupled folding and binding, a disordered protein region can transition from a flexible and dynamic state into one with definite structure upon its binding interaction with a folded, structured partner, forming complexes that can often be relatively static and potentially amenable to structure determination by X-ray crystallography. In its bound and folded
state, the disordered protein region can assume regular secondary structures such as α-helices and β-strands (both types of secondary structure are observed in the context of the dynein intermediate chain binding to its light chains, see below and Figure 3.8), but also irregular structures as well, without any regular pattern of backbone hydrogen bonds or any type of recognizable secondary structure (Uversky 2011a). In addition, it is noteworthy that some disordered protein regions exhibit ‘chameleon-like’ behavior in that, the same basic set of residues can bind to multiple different partners, gaining very different structures upon binding in each case. A prime example of this occurs for the tumor suppressor protein p53 via residues located in its C-terminal regulatory region. The span of residues 374–388 in p53 contains the binding interaction sites for: S100B(ββ), with which p53 forms an α-helix upon binding (Rustandi et al. 2000); sirtuin Sir2, with which p53 (Lysine382-acetylated) forms a β-strand upon binding (Avalos et al. 2002); transcription co-activator CBP (CREB-binding protein, mentioned earlier), with which p53 (Lysine382-acetylated) forms an irregular structure in the bound state (Mujtaba et al. 2004); and Cyclin A of Cdk2/Cyclin A, with which p53 forms another but different irregular structure in the bound state (Lowe et al. 2002) (Figure 1.1a).

It should also be noted that a mutual folding-upon-binding mechanism can occur, wherein the interacting regions of both partners are disordered in the free state but each undergo a synergistic folding transition upon binding one another (Wright et al. 2009); an example of this phenomenon is observed for the binding interaction between IDRs in the Arf tumor suppressor and the ubiquitin E3 ligase Hdm2 proteins (along with p53, these proteins regulate a tumor-suppressor pathway that is disrupted in many types of human cancer) that join to form β-strand-rich oligomeric structures with amyloid resemblance (Sivakolundu et al. 2008).

In addition to some of the relatively static complexes involving IDP/IDR binding discussed above, disordered protein segments can also engage in binding interactions that produce rather dynamic (or, “fuzzy”) complexes in which the disordered segment continues to sample an ensemble of rapidly inter-converting conformations that may be largely devoid of structure even in the bound state (Tompa et al. 2008). Such dynamic complexes are not so amenable to structure determination by classic means such as X-ray crystallography; further, even if such a dynamic protein complex could be made to crystallize and one were able to determine the three-dimensional structure of a single conformational substate for such a dynamic complex, this static structural ‘snapshot’ would represent only one out of a vast continuum of significantly populated states in the biologically-relevant ensemble of different conformations that are transiently sampled by the IDP/IDR in the bound state. Different categories of dynamic, “fuzzy”
**Figure 1.1.** Illustrative portraits of some disorder-based protein complexes. (A) The span of residues 374–388 in the C-terminal regulatory region of the tumor suppressor protein p53 gains different types of secondary structure when in complex with four different binding partners (clockwise, starting from top left): S100B(ββ) (orange surface), p53 forms an α-helix (red ribbon); sirtuin Sir2 (light pink surface), p53 forms a β-strand (magenta ribbon); CBP bromodomain (teal surface), p53 forms an irregular structure (orange thread); and Cyclin A (light blue surface) of Cdk2/Cyclin A, p53 forms another, but different, irregular structure (green thread). (B) In contrast to the relatively static complexes depicted in (A) where a disordered region can adopt a well-defined conformation in the complex, in dynamic “fuzzy” complexes the disordered protein region can fluctuate between various conformations of its bound-state ensemble; a prime example is the dynamic complex of the disordered phosphorylated yeast Cdk inhibitor Sic1 and Cdc4. Multiple phosphorylated motifs (red circles marked with ‘P’ in the center) in pSic1 engage the core Cdc4 binding site (the highlighted blue region at the top of the gray polygon) in a dynamic equilibrium. pSic1 sites not directly bound in the core Cdc4 binding site at any given instant can contribute to the binding energy via long-range electrostatic interactions. A more detailed view of one possible hypothetical interaction is shown (left middle) with the pThr-45 site of pSic1 (shown in stick representation) binding to Cdc4; this hypothetical binding interaction is modeled based upon coordinates in the structure of a phosphorylated cyclin E peptide in complex with Cdc4 (Orlicky et. al. 2003). The parts of pSic1 remaining disordered are depicted as black lines. Figure images adapted with permission (A) from Oldfield et. al., 2008 (Oldfield et. al. 2008), and (B) from Mittag et. al., 2008 (Mittag et. al. 2008).
protein complexes have been enumerated, and much like protein disorder itself—once thought to be an anomalous and rare occurrence (aside from disordered loops and termini in folded, globular proteins or short linkers between folded protein domains)—examples of retained disorder for IDPs/IDRs in the bound state continue to present themselves and steadily accumulate in the literature (Fuxreiter et al. 2012).

One particularly notable example of a “fuzzy” and dynamic protein complex occurs in the case of the disordered Cdk inhibitor Sic1 in yeast (S. cerevisiae inhibitor Sic1 is the functional homologue of mammalian inhibitor protein p27Kip1 mentioned above). In S. cerevisiae the cyclin-dependent kinase Cdc4 binds to other proteins in the cell, targeting and bringing them to the SCF complex containing a ubiquitin ligase; the targeted proteins are ubiquitinated and then degraded by the proteasome. The WD40 β-propeller domain of Cdc4 recognizes short serine- or threonine-phosphorylated sequences in Sic1 and other substrates; Cdc4 contains a single deep pSer/pThr–Pro-binding pocket to which phosphorylated Sic1 (pSic1) and other substrates bind (Orlicky et al. 2003). High affinity binding of Sic1 to Cdc4 appears to require the phosphorylation of about a half dozen Ser/Thr residues, all of which engage the single receptor site of Cdc4 in a dynamic equilibrium, wherein pSic1 remains globally disordered and dynamic, with each single site gaining only transient local ordering upon interaction with the single binding site in Cdc4 (Mittag et al. 2008) (Figure 1.1b). The fast inter-conversion of a multitude of flexible conformers of pSic1 is believed to create a mean electrostatic field that allows even the unbound phosphates to contribute to the binding affinity via long-range electrostatic interactions with the positively-charged surface of Cdc4; in this way, the degree of phosphorylation in pSic1 (related to cell cycle status) can be “sensed” by Cdc4 in an ultrasensitive and thresholded manner (Borg et al. 2007).

**Benefits and Functional Advantages of Protein Disorder.**

Not only can intrinsic protein disorder be quite functional (discussed above), but there are also several functional advantages and benefits associated with this disorder. For instance, one hallmark of IDP/IDR binding to partners is the combination of relatively weak binding affinity, coupled with high specificity. For example, the folding of an IDP/IDR upon binding its target partner is accompanied by a substantial decrease in entropy that offsets the energetically favorable enthalpic contribution of the binding interaction, often leading to a relatively weak binding affinity; in some cases, disorder acts to prevent spontaneous assembly of the IDP/IDR at inappropriate times or locations or with the ‘wrong’ binding partner in that, the spatial coordination of atoms that occurs in the folded state of the disordered segment (e.g. the specific
binding interface that gets presented to a potential partner) must be sufficiently ‘matched’ by the
target binding partner such that the enthalpy of the binding interaction is sufficient to overcome
the entropic penalty due to the folding of the disordered segment (Wright et al. 1999). In this
way, high specificity of binding is achieved by the disordered protein segment, coupled with a
modest binding affinity that allows for reversibility of the binding interaction.

In addition to readily reversible binding interactions, protein disorder has also been
suggested to facilitate enhanced rates of interaction with a binding partner due to possession of a
larger (relative to a folded state) ‘capture radius’ for a specific partner binding site (Shoemaker et.
al. 2000), making IDPs/IDRs ideal for the rapid and reversible interactions involved in cell
signaling and regulation, facilitating rapid regulatory control of the cell’s response to changing
environmental conditions (Wright et al. 1999; Uversky 2011a). Moreover, the protein disorder
can often be used (by nature) to fine-tune the equilibrium and thermodynamics of binding
interactions via the amount of structural disorder present in either the apo or bound states of the
IDP. Among disordered proteins that exhibit folding-upon-binding, the interacting segments
often contain some degree of preexisting (or, ‘preformed’ or ‘nascent’) localized structure in
regions termed “MoRFs” (‘Molecular Recognition Features’; Mohan et. al. 2006) that bear
resemblance to the structure adopted by the IDP in the bound state; in this way, the effective
entropic penalty due to folding-upon-binding is reduced due to the ‘pre-payment’ of some of the
entropic cost via the existence of some structure in the disordered segment in its apo state prior to
binding interaction. In cases where the disordered protein segment retains some degree of
disorder in the bound state (e.g. a dynamic or “fuzzy” complex), the entropic penalty of
association is reduced on the ‘back end’ of the equation, by virtue of the fact that there is less
ordering or structure of the IDP in the bound state.

Structural plasticity present in the unbound state of an IDP can also confer functional
advantage in terms of being able to accommodate binding to several different partner ligands; this
feature underlies the ability of IDPs to act as molecular scaffolds and ‘hub’ proteins (mentioned
above). Intrinsic disorder can endow proteins with a high density of sites for protein–protein
interactions, particularly since disordered interaction sites can be comprised of a few residues
within a short stretch of sequence. A fundamental physical basis for the ability of relatively short
linear stretches in IDRs to participate in more interactions than an ordered region comprised of
the same number of residues is that, by being conformationally extended, unstructured regions
expose more surface area per residue; it was calculated that, in order for a folded globular protein
to be monomeric and present the same amount of interface area as a disordered protein, the folded
protein would need to be 2–3 times larger (that is, have 2–3 times as many residues) as the disordered protein, which would in-turn require larger genome and cell sizes (Gunasekaran et. al. 2003). Thus, intrinsically disordered protein regions have more surface area available for molecular recognition and binding interactions.

Further economy can be realized when a single disordered region can recognize and bind several different partners, or when binding sites overlap in the sequence of the IDR. An example of multiple partners binding to a single disordered region (e.g. ‘binding promiscuity’) was described for the case of the C-terminal regulatory region of p53 in the previous sub-section. Another example of how protein intrinsic disorder facilitates binding-partner promiscuity occurs in the case of the central protein of this thesis—i.e. the N-terminal region of the dynein intermediate chain contains overlapping binding sites for multiple regulatory partners (see below, and Figure 1.4). Aside from just simple economy, overlapping binding sites in an IDP can also facilitate pathway selection (as in the case of signaling scaffolds, for instance) by virtue of mutual exclusivity in the association of the disordered region with one particular partner versus another. Or, in the case of macromolecular complex assembly scaffolds (such as the dynein intermediate chain), mutual exclusivity of binding partners with overlapping binding sites can serve as a ‘switch point’ in that, association with only one type of partner (to the exclusion of others) can select for particular functional activities of the overall macromolecular complex (e.g. dynein) within the cell.

Post-translational modifications (such as phosphorylation, acetylation, methylation, lipidation, etc.) of a polypeptide chain can modulate and extend the functional range of a protein in the cell (Uversky 2013). Several of these post-translational modifications have been frequently observed to occur in, and are associated with, regions of intrinsic protein disorder (Xie et. al. 2007b). Chemical modification of amino acid side chains requires close association between a target protein and the modifying enzyme; the conformational flexibility afforded by disorder facilitates binding of the substrate IDR to the enzyme active site, with the increased accessibility of IDRs providing for efficient modification, whereas for a side chain within a structured region, steric factors might be expected to slow or impede the association and chemical modification of target protein by enzyme. Reversible protein phosphorylation, a major regulatory mechanism in eukaryotic cells, occurs predominately in regions of protein disorder (Iakoucheva et. al. 2004); multiple examples of intrinsically disordered protein phosphorylation appeared in this brief section on IDPs. Phosphorylation/de-phosphorylation allows precise control over the
thermodynamics of binding interactions, thus providing a mechanism for inducibility of said interactions and rapid initiation of regulatory changes in cell signaling and other pathways.

In addition to post-translational modifications, the functional range of proteins can be extended through pre-translational alternative splicing of mRNA transcripts. Alternative splicing of proteins enables the temporal and tissue-specific modulation of protein function needed for processes such as cell development and differentiation (Buljan et. al. 2012). This type of splicing occurs preferentially within regions of protein disorder (Romero et. al. 2006), which can be rationalized by consideration of the effects that insertion/deletion of residues and modules can have on protein structure, stability, and thus function. In the case of folded or globular proteins, splicing within a structured protein region or domain would likely have catastrophic effects on the structure of the rest of the protein (unless, perhaps, if the spliced segment were to be small and physically located on the relatively less-structurally-constrained surface of the protein, for instance), likely leading to its misfolding and aggregation. In contrast, disordered protein segments are not constrained by rigid, fixed secondary or tertiary structure, thus allowing alterations to be made without incurring significant complications and penalties. Protein disorder allows for the insertion/removal of functional binding recognition sequence elements, thus providing for modulation of signaling and regulation pathways (Buljan et. al. 2012), or in the case of macromolecular assembly scaffold IDPs, differential composition and stoichiometries of subunits within the complex; further, removal of some but not all key residues from a disordered binding motif could have the effect of ‘tuning’ the affinity and kinetics of the binding interaction with that site, rather than completely eliminating it.

Having given this brief introduction and overview of intrinsically disordered proteins, their functionalities and participation in biological processes, binding modes, and associated functional advantages, attention is now turned to the dynein motor complex, which contains several different subunits including the intrinsically disordered intermediate chain (IC)—the central protein of this thesis. This brief introduction to IDPs will hopefully have provided sufficient background and context to afford appreciation of the physical features and functionality of the dynein intermediate chain IDP.
The Cytoplasmic Dynein Motor Complex

Cells represent microcosms of spatial and temporal structural organization; their ability to function and to assemble into higher-order structural systems (such as tissues, organs, and entire organisms) depends ultimately upon their achievement of internal spatial organization, compartmentalization, and sequestration of components. To this end, cells employ a system of macromolecular motor complexes (such as kinesins and dyneins) and microtubules to achieve transport of various components throughout the cell (Hirokawa 1998; Ross et al. 2008). The motor complexes convert the chemical energy from ATP hydrolysis into mechanical motion for directed transport of cargoes along microtubule tracks throughout the cell.

Functions of Cytoplasmic Dynein.

Cytoplasmic dynein, whose discovery was first reported in 1987 (Paschal et al. 1987), is the principle microtubule motor responsible for minus-end-directed (retrograde) transport in most eukaryotic cells (Schnapp et al. 1989; Schroer et al. 1989). Cytoplasmic dynein (hereafter referred to simply as ‘dynein’) is involved in myriad critical cellular processes, including: retrograde axonal transport of membranous cargoes in neurons (Schnapp et al. 1989; Waterman-Storer et al. 1997), nuclear migration during mitosis and positioning within cells (Eshel et al. 1993; Salina et al. 2002), positioning and orientation of organelles such as the Golgi apparatus and lysosomes (Corthesy-Theulaz et al. 1992; Lin et al. 1992; Burkhardt et al. 1997; Harada et al. 1998), intermediate compartment transport between the endoplasmic reticulum (ER) and Golgi complex (Presley et al. 1997), and facilitation of endocytic traffic with vesicular transport between early and late endosomes, as well as movement of endocytic cargo from the endosome to the lysosome (Gruenberg et al. 1989; Aniento et al. 1993).

Dynein is also heavily implicated in the mitotic spindle, being involved in centrosome migration and positioning, as well as interphase and mitotic microtubule organization in the cell (Ma et al. 1999; Robinson et al. 1999; Vaisberg et al. 1993). Mitotic dynein is localized to astral microtubules and at sites of the cell cortex (to which the plus ends of astral microtubules may be attached) and is thought to function in anchoring and positioning the spindle poles, as well as in helping to generate the force to separate the poles in anaphase B of mitosis (Carminati et al. 1997; Busson et al. 1998; Sharp et al. 2000a). Dynein also contributes to organization of the spindle and poles by its transport of pericentrin to centrosomes; pericentrin is a conserved protein of the centrosome involved in microtubule nucleation and organization (Purohit et al. 1992).
Additionally, in the progression toward, and through the early stages of mitosis, the interphase microtubule array of the cell is broken down and many of these microtubules become bundled at their minus ends; dynein further contributes to the organization of the spindle via translocation of the minus ends of these bundles along and toward the minus ends of centrosomal microtubules, toward the spindle poles, whence these ‘recycled’ microtubules can be incorporated into the forming mitotic spindle (Heald et al. 1996; Rusan et al. 2002). For microtubules whose plus ends are already attached at kinetochores and thus connected to chromosomes (‘Kinetochore fibers’), dynein helps to transport unattached minus ends of these microtubules (a.k.a. ‘K-fibers’) poleward toward centrosomes (Khodjakov et al. 2003), further contributing to spindle formation and bipolarity.

Dynein is also localized to kinetochores during mitosis (Pfarr et al. 1990; Steuer et al. 1990), where it has been implicated in initial microtubule capture (i.e. attachment of polar microtubules to the kinetochore) and in generation of the initial rapid poleward chromosome movements in prometaphase that facilitate additional microtubule capture, leading to chromosome congression at the spindle equator (metaphase plate) (Echeverri et al. 1996; Sharp et al. 2000b; Yang et al. 2007; Vorozhko et al. 2008; Varma et al. 2008). In addition, upon stable microtubule attachment at kinetochores during prometaphase, dynein is released from kinetochores and a mass egress of dynein toward the spindle poles occurs with coincident significant decrease in the amount of dynein at kinetochores (King et al. 2000a). In this migration along microtubules toward the spindle poles (‘poleward streaming’), dynein functions in the transport of mitotic spindle assembly checkpoint (SAC) proteins [such as Mad1, Mad2, and BubR1, as well as the RZZ complex (see below, ‘Regulators of the Cytoplasmic Dynein Motor Complex’)] away from the kinetochore, contributing to inactivation of the SAC (Howell et al. 2001; Wojcik et al. 2001); inactivation of the SAC leads to activation of the anaphase-promoting complex (APC) and progression of the cell into anaphase (Musacchio et al. 2007). The small kinetochore dynein population remaining at metaphase (following the mass poleward egress that contributes to SAC protein stripping from kinetochores) has also been shown to contribute to the motion of kinetochores (and chromosomes) toward spindle poles during anaphase A (Sharp et al. 2000b; Banks et al. 2001; Yang et al. 2007).

Consistent with the numerous and diverse cellular roles of dynein, defects within the motor and its components, as well as in its associated non-dynein regulatory complexes (see below), are implicated in numerous human diseases, particularly neurological disorders (Reiner et al. 1993; Farrer et al. 2009; Dupuis et al. 2009) and in motor neuron disease and
degeneration (Hafezparast et al. 2003; Puls et al. 2003; Eschbach et al. 2011; Lipka et al. 2013). In addition to the myriad functions of dynein enumerated above, it would seem that dynein has also been co-opted (or, ‘hijacked’) by viruses for transport within cells (Dohner et al. 2005). Upon entry and release into the cytoplasm, numerous viruses—such as Herpes Simplex Virus type 1 (HSV1) (Dohner et al. 2002) and Adenovirus (Kelkar et al. 2004)—rely upon dynein for retrograde transport along microtubules in order to travel from the cell periphery to the more centrally-located nucleus, where the viral DNA can enter into the nucleus via nuclear pores (Scherer et al. 2011). HSV1 typically enters peripheral nerve terminals and travels back along the axon (a lengthy cell projection in neurons) toward the neuronal cell body where the nucleus is located; in the absence of directed transport by a motor such as dynein, it was calculated that, relying solely on random diffusion, it would take 231 years for a particle the size of a herpes virus capsid to travel a distance of 10 mm in the axonal cytoplasm (Dohner et al. 2002). In the case of HSV1 and Adenovirus, docking of viral capsids on dynein occurs via direct interaction of viral capsid proteins with the dynein intermediate chain (IC) and Tctex1 subunits (for HSV1; Ye et al. 2000; Douglas et al. 2004), and with the dynein Intermediate and Light Intermediate Chains (for Adenovirus; Bremner et al. 2009).

**Multi-subunit Composition of the Cytoplasmic Dynein Motor.**

Dynein is a large (~1.2 MDa) multi-subunit protein complex that may be conceptually divided into two functional domains. The first domain is comprised of the dynein heavy chains (HCs), which contain the sites for microtubule binding and are responsible for the ATPase activity to which dynein motion along the microtubule is coupled (Vallee et al. 2004) (Figure 1.2a). The second functional domain is referred to as the ‘cargo attachment complex’ whose constituents include the intermediate chains (ICs) (Paschal et al. 1992), light intermediate chains (LICs), and the light chain proteins Tctex1, LC8, and LC7 (King et al. 1996a,b; Bowman et al. 1999) (Figure 1.2a). Thus, cytoplasmic dynein is comprised of six subunit families (Pfister et al. 2006).

*The Dynein Heavy Chain (HC).*

Cytoplasmic dynein is an assemblage of dimers of its multiple subunits. The heavy chains are ~530 kDa and each contains a ring of six AAA+ ATPases associated with diverse cellular activities ATPase modules, of which four can bind or hydrolyze ATP; in addition to the ring of six AAA+ units, the C-terminal portion of HC also contains an anti-parallel coiled-coil
Figure 1.2. Schematic representation of the cytoplasmic dynein motor complex and configuration of its components; major components are labeled. (A) Cytoplasmic dynein contains two heavy chains (light blue), shown with microtubule-binding stalks attached to a microtubule (orange cylinder). The light intermediate chains (LICs; purple) are attached within the N-terminal portion of the heavy chain chains. Two copies of the intermediate chain (IC; grey) are present in every motor complex; the N-terminal half of IC (‘N-IC’) is predominately disordered, while the C-terminal half (‘C-IC’; grey spheres) is predicted from protein sequence to have a toroidal β-propeller structure. The three dimeric light chains Tctex1 (yellow spheres), LC8 (green spheres), and LC7 (small blue spheres) are depicted in complex with the IC chains, forming a polybivalent scaffold. Also diagrammed is a model for (vesicular) cargo attachment: a generic membranous cargo is depicted (large blue sphere), attached to dynein via the heteromultimeric dynactin complex (teal), containing a dimer of its largest subunit, p150Glued.  

(B) Top–Schematic representation of the dynein intermediate chain (IC) showing its division into domains ‘N–IC’ and ‘C–IC.’ The structurally-determined binding sites (sequence numbering is for D. melanogaster) for the three light chains in N–IC are indicated with the same coloring scheme as in Figure 1.2a. Also shown for N–IC are the predicted (JPred3; Cole et. al. 2008) secondary structural elements and predicted coiled-coil regions (PairCoil2; McDonnell et. al. 2006) [red double lines, ‘ICC’ (aa 1-31) and ‘PrCC’ (aa 207-237)]. The segments of IC outside of these regions are predicted to be disordered (straight line). The ‘serine-rich’ region (Nurminsky et. al. 1998) is indicated with a red bracket (residues 71–86 in D. melanogaster).  

Bottom–A graphic representation of the central protein of my thesis, D. melanogaster dynein intermediate chain residues 1–143 (“IC:1–143”), depicting the primarily disordered nature of the molecule, as well as certain regions that I determined to have structural deviation from random coil behavior (Morgan et. al. 2011).
Figure 1.2. (Continued)
stalk (between the fourth and fifth ATPase domains) with a globular microtubule binding domain (MTBD) located at the tip of the ~15 nm stalk (Cho et al. 2012). ATP hydrolysis within a AAA+ domain produces conformational change within the unit that is propagated through the heavy chain protein to the anti-parallel coiled-coil stalk domain, resulting in a change in the register (e.g. sliding of one coil strand relative to the other), which in-turn results in a conformational change within the attached MTBD, thus effecting a change in its microtubule binding affinity (Nishikawa et al. 2014); in this way, ATP hydrolysis can be coupled to the association state of dynein’s “feet” with the microtubule. The homo-dimerization domain of the heavy chain resides within the N-terminal portion of the protein and this dimerization domain is also overlapped by binding sites for the C-terminal half of IC chains (‘C-IC’) and for the Light Intermediate Chains (LICs) as well (Tynan et al. 2000b) (Figure 1.2). I like to conceptualize this particular junction—the heavy chain homodimerization domain, plus the C-ICs and LICs—as constituting the “pelvis” of cytoplasmic dynein.

The Dynein Light Intermediate Chain (LIC).

The Light Intermediate Chains (LICs) range from ~50–60 kDa and in mammalian species, are encoded by more than one gene [(DYNC1LI1, ‘LIC1’) and (DYNC1LI2, ‘LIC2’)] (Pfister et al. 2006). The LIC1 and LIC2 proteins form homooligomers (but not heterooligomers) and exhibit mutually exclusive binding to the heavy chain (Tynan et al. 2000a); the presence of LIC has been demonstrated to stabilize and strengthen the association between the heavy chain and C-IC (Zhang et al. 2009). LIC1 (but not LIC2) is able to bind the protein pericentrin and is one of the dynein subunits known to function directly in cargo attachment to the motor complex (Tynan et al. 2000a). As mentioned above, in addition to the Intermediate Chains, Adenovirus particles also attach directly to the LIC1 subunits, demonstrating another example of LIC participation in direct cargo attachment (albeit a pathogenic form of cargo) to the cytoplasmic dynein motor (Bremner et al. 2009). In human (HeLa) cells, LIC1 and LIC2 were observed to function redundantly (e.g. the two forms were interchangeable) in dynein for positioning the centrosome close the nuclear envelope in interphase (Raaijmakers et al. 2013). In mammalian cells, LIC1 (much more so than LIC2) has been observed to function in recruitment of dynein to late endosomes and lysosomes, independently of any additional, dynactin-based (see below, ‘Regulators of the Cytoplasmic Dynein Motor Complex’) recruitment mechanism (Tan et al. 2011); in the case of endosomes, dynein/LIC1 recruitment to these compartments is mediated through direct interaction of LIC1 with Rab4-GTPase or Rab11-GTPase/FIP3 complex, for instance (Bielli et al. 2001; Horgan et al. 2010).
In addition, in HeLa cells, the LICs were demonstrated to be essential for all dynein functions in mitosis (with LIC1 and LIC2 often functioning redundantly) (Raaijmakers et. al. 2013). For example, depletion of LIC1 in human cells resulted in metaphase delay, with accumulation of dynein and signal assembly checkpoint proteins (Mad1/2 and ZW10) at kinetochores (Sivaram et. al. 2009); depletion of LIC in Drosophila produced the same mitotic cellular defects (Mische et. al. 2008). Also, various LIC phosphoisoforms (the result of post-translational modification) have been isolated and observed in mammalian species and in Drosophila as well (Hughes et. al. 1995; Mische et. al. 2008). The LIC1 subunit contains several consensus sequence sites for phosphorylation by a number of cellular kinases (Bielli et. al. 2001). In the study of LIC1 depletion and metaphase delay in human cells, it was determined that a Cdk1(cyclin-dependent kinase-1)-phosphorylated form of LIC1 is required for removal of SAC proteins Mad1/2 and ZW10 from the kinetochore and release of metaphase arrest (Sivaram et. al. 2009). Phosphorylation is not a general prerequisite for LIC incorporation into dynein, as the non-phosphorylated form of LIC1 is bound within dynein in interphase cells (Sivaram et. al. 2009), suggesting that the LICs may contribute an additional layer of regulation in dynein function, by virtue of their phosphorylation states at various points during the cell cycle. At present, there are no protein crystal structures available for LIC [however, note work presented in Appendix 3]. In metazoan species, the LICs contain a P-loop motif (a nucleotide-binding domain found in numerous proteins such as ATPases and kinases) that allows interaction with ATP (Hughes et. al. 1995); the purpose or role of this ATP interaction in the function of LIC is not currently known.

The Dynein Intermediate Chain (IC).

The cytoplasmic dynein Intermediate Chain (IC, also known as ‘IC74’) is a ~74 kDa protein (642 amino acid residues in D. melanogaster) encoded by a single gene in Drosophila (Cdic; Nurminsky et. al. 1998) and by two different genes in mammalian species [(DYNC1I1, ‘IC-1’) and (DYNC1I2, ‘IC-2’)] (Pfister et. al. 2006); IC-1 and IC-2 share ~69% sequence identity (Kuta et. al. 2010). Molecular genetic analyses in Drosophila show that IC mutations result in larval lethality, demonstrating that IC serves an essential function in vivo (Boylan et. al. 2002). The C-terminal half of IC (‘C-IC’, residues 290–642 in D. melanogaster) contains seven WD40 repeats and is predicted on the basis of sequence to fold into a toroidal β-propeller structure (Wilkerson et. al. 1995; Nurminsky et. al. 1998); the C-IC subdomain is the site of IC interaction with the dynein heavy chain (Ma et. al. 1999), with this interaction presumably
facilitated through the β-propeller structure of C-IC. The N-terminal half of IC (‘N-IC’, residues 1–289 in *D. melanogaster*) is predominately disordered (see below, ‘Disorder in Dynein—the Intermediate Chain’) under physiological conditions (Makokha *et. al.* 2002; Nyarko *et. al.* 2004). Two IC molecules are present in each cytoplasmic dynein motor complex, and N-IC monomers are joined in a parallel, polybivalent molecular scaffold, containing the binding sites for the three dimeric dynein light chains Tctex1, LC8, and LC7 (Lo *et. al.* 2001; Makokha *et. al.* 2002; Susalka *et. al.* 2002) (Figure 1.2). I like to conceptualize this particular assemblage—two N-ICs, plus the three dimeric dynein light chains—as constituting the “torso” of cytoplasmic dynein.

*The Dynein Light Chain Tctex1.*

The cytoplasmic dynein light chain Tctex1 is ~14 kDa and is encoded by a single gene in *D. melanogaster* and by two genes in mammalian species [(*DYNLT1*, ‘Tctex1’) and (*DYNLT3*, ‘rp3’ a.k.a. ‘Tctex1L’)] (Pfister *et. al.* 2006). Tctex1 and rp3 (which shares ~55% amino acid identity with Tctex1) are homodimers that bind directly to the dynein intermediate chain (King *et. al.* 1998; Mok *et. al.* 2001) and have been demonstrated to compete with one another for binding to IC in *vitro* (Tai *et. al.* 2001). This may help to explain, in part, why Tctex1 and rp3 expression are regulated differentially in both a developmental (e.g. temporal) and in a tissue-specific (e.g. spatial) manner; for instance, rp3 is most the prevalent form present in liver and brain tissue in rat, while Tctex1 is the least represented in those cell types (King *et. al.* 1998). Tctex1/rp3 have a limited function in cargo attachment to the dynein motor; both variants can participate in binding HSV1 to dynein (Douglas *et. al.* 2004) (see above). Tctex1, but not rp3, can bind to the C-terminal cytoplasmic tail of rhodopsin, allowing dynein to translocate rhodopsin-bearing vesicles along microtubules in rod photoreceptor cells of the retina (Tai *et. al.* 1999), thus demonstrating an example of the modulation of cytoplasmic dynein function via differential subunit composition.

*The Dynein Light Chain LC7.*

Cytoplasmic dynein light chain LC7 (a.k.a. ‘Roadblock’) is ~11 kDa and is encoded by a single gene in *Drosophila* (*roadblock, rob1*), where mutations in *rob1* were seen to result in accumulation and other defects in the transport of axonal cargoes (*i.e.* mutations produced a “roadblock” in neuronal axons), and either larval or pupal lethality (Bowman *et. al.* 1999). Mammalian species contain two genes for LC7/Roadblock [(*DYNLRB1*, ‘LC7-1’, a.k.a. ‘Roadblock-1’ or ‘km23-1’) and (*DYNLRB2*, ‘LC7-2’ a.k.a. ‘Roadblock-2’ or ‘km23-2’)] (Pfister *et. al.* 2006). LC7-1 and LC7-2 bind directly to all known protein isoforms of IC-1 and IC-2,
with the LC7 binding site located toward the C-terminal end of N-IC (Susalka et al. 2002) (Figure 1.2). In homodimeric Drosophila LC7, IC was seen to contribute two long α-helical segments when bound—one helix from each of the two IC chains (Hall et al. 2010). Mammalian LC7s can form both hetero- and homodimers (Nikulina et al. 2004); LC7-2 shares 74% identity (83% similarity) with LC7-1. The LC7 variants exhibit differential tissue-dependent expression, and changes in LC7 isoform expression (up-regulation of LC7-1 and down-regulation of LC7-2) have been observed in hepatocellular carcinoma (Jiang et al. 2001).

Like the Tctex1/rp3 light chains, the LC7 subunit can also interact with non-dynein partners. Independent of their association with dynein, LC7-1 and LC7-2 undergo phosphorylation of certain serine residues upon their engagement with the cytoplasmic (receptor tyrosine kinase) domain of a ligand-activated transforming growth factor-β (TGF-β) complex; this phosphorylation of mammalian LC7s is necessary for binding to dynein intermediate chains in the motor complex (Tang et al. 2002; Jin et al. 2013). Activated TGF-β receptors activate (phosphorylate) Smad2 and Smad3, thereby allowing these proteins to associate with LC7-1 and LC7-2, respectively (Jin et al. 2007; Jin et al. 2009). Phosphorylated LC7/Smad complexes are recruited to the rest of the dynein motor complex, which facilitates retrograde transport of the Smads toward the cell interior (where the nucleus is located) (Jin et al. 2013). Smad2/3 are intracellular proteins that transduce extracellular signals from ligand-activation of TGF-β receptors, to the nucleus where they can activate gene transcription; this may help to explain how changes in LC7 isoform levels can, in-turn, lead to a cancerous cellular state.

The Dynein Light Chain LC8.

Like Tctex1 (above), LC8 (whose name derives from the observation that this protein component migrated at ~8 kDa in SDS-PAGE gels, though the protein is actually ~10 kDa) is also homodimeric with a strand-swapped β-sheet interface (with crossover β3 strand at the dimer interface); dynein IC contributes two β-strand segments (one from each of the two IC chains) at the homodimeric interface (Benison et al. 2007b; Williams et al. 2007; Hall et al. 2009) (see Figure 3.8). LC8 is encoded by two genes in Drosophila (Cdlc1 and Cdlc2) and Cdlc1 null mutants caused Drosophila embryonic lethality with excessive apoptosis (Dick et al. 1996). LC8 is also encoded by two genes in mammalian species [(DYNLL1, ‘LC8-1’ a.k.a. ‘LC8a’ or ‘DLC1’) and (DYNLL2, ‘LC8-2’ a.k.a. ‘LC8b’ or ‘DLC2’)] (Pfister et al. 2006). Mammalian LC8a and LC8b proteins differ from each other by only six amino acid residues out of 89, with most residue substitutions being conservative, and comparison of LC8a and LC8b sequences
among different mammalian species revealed that each protein variant was identical across species; though quite similar in sequence, the two protein forms exhibited differential tissue expression patterns in rat (Wilson et. al. 2001). Given their high sequence similarity and the fact that both protein forms are components of cytoplasmic dynein, most studies on the cellular roles of LC8 have not made a distinction between the two different gene products—they are often considered and referred to together under the inclusive term ‘LC8.’

Aside from the dynein intermediate chain, LC8 is well-known to engage in binding interactions with a plethora of non-dynein partners in the cell, and is considered to be a prime example of a “hub” protein (reviewed in Barbar 2008; Rapali et. al. 2011; Barbar et. al. 2014). Observation of its interaction with a multitude of proteins involved in diverse biological functions led to the supposition that LC8 might function as a prolific cargo attachment subunit within the dynein motor complex, and that its diverse partners might presumably be dynein cargo. However, many of these non-dynein binding partners were observed to associate with LC8 in dimeric fashion via their own disordered segments (containing a similar type of binding recognition sequence as that in IC), and bind within the same grooves at the LC8 dimer interface that are occupied by the dynein intermediate chains (Liang et. al. 1999; Fan et. al. 2001; Benison et. al. 2007b). Structural determination and dynamical analyses indicated that dynein IC and Swallow (a non-dynein LC8 binding partner) could not be simultaneously bound to the same LC8 dimer (Benison et. al. 2007b)—something that would be requisite in order for LC8 to function as an adaptor to bind this type of cargo to the dynein motor. Rather, a major functional role that has emerged for LC8 in the cell is that of a dimerization engine, with the ability to promote dimerization in its diverse partners by binding to certain disordered segments within these proteins (Barbar 2008). Consistent with the idea that LC8 might have functions within the cell that are completely dynein-independent, is the observation that a significant soluble pool of LC8 exists within cells that is not bound to or associated with the dynein complex (King et. al. 1996b). In the case of dynein, rather than serving primarily as a cargo adaptor, the emerging view of LC8 function within the motor complex is that it serves an important function in dynein assembly (Benison et. al. 2006).

Indeed, though some of the dynein light chains may exhibit limited participation in cargo attachment to dynein (see above), the three different dynein light chains have increasingly been regarded for their important structural role in assembly and stabilization of the dynein cargo attachment sub-complex (Nyarko et. al. 2011; Trotker et. al. 2012). Support for the regulatory
structural roles served by the light chains derives from several observations. The light chains have highly conserved (across species) binding recognition sequences in IC, and these sequences are located outside of alternative splice regions in IC (Figure 1.3), suggesting that the binding of the LCs to IC is likely to be constitutive and fundamentally required, and is not modulated by isoform diversity in IC (Mok et al. 2001). In addition, there is a reciprocal enhancement of binding affinity for Tctex1 and LC8 to IC via the multivalency effect (Hall et al. 2009), and the binding of LC8 to IC has been observed to effect changes in the structure and order of IC at a site distal to the LC8 binding region (Nyarko et al. 2004, 2011). Further, abrogation of the LC8-IC interaction in Saccharomyces interfered with recruitment of the dynactin complex (see below, ‘Regulators of the Cytoplasmic Dynein Motor Complex’) to dynein in vivo (Stuchell-Brereton et al. 2011); this effect occurs in spite of the fact that LC8 and dynactin have no known direct interaction with one another, suggesting that LC8 occupancy affects the binding of dynactin p150Glued through direct modulation of IC (to which p150Glued binds).

Having reviewed the basic components of the dynein motor complex, focus is now turned to the central component of the dynein cargo attachment complex (and the central protein of my dissertation work)—the dynein intermediate chain.

**Disorder in Dynein—the Intermediate Chain.**

In addition to providing a binding platform for the three dimeric dynein light chains (Figure 1.2), the dynein intermediate chain has also been implicated in the direct binding of various cargoes to dynein, including Adenovirus and Herpes Simplex Virus 1 (discussed above), the CIC-2 chloride channel (Dhani et al. 2003), and also β-catenin (Ligon et al. 2001) and PLAC-24 (Karki et al. 2002)—both proteins that are localized to intercellular adherens junctions (Harris et al. 2010). IC also interacts directly with the protein Huntingtin, implicated in Huntington’s Disease (Caviston et al. 2007), and with components of several protein complexes (Dynactin, NudE/EL, and RZZ) that regulate the function and activity of the dynein motor complex in the cell (discussed further below, and also see ‘Regulators of the Cytoplasmic Dynein Motor Complex’).

As mentioned previously, N–IC is primarily disordered in its apo state (Makokha et al. 2002; Nyarko et al. 2004) and it is aptly categorized as an intrinsically disordered protein (‘IDP’, see section above, ‘Intrinsically Disordered Proteins—A Brief Introduction’). Like many IDPs, though being predominately disordered, N–IC is interspersed with elements of secondary
**Figure 1.3.** The bi-segmental dynactin p150\textsuperscript{Glued} binding footprint on *D. melanogaster* IC and sequence comparisons with mammalian IC isoforms. (A) *Top* sequence is *D. melanogaster*, *middle* and *bottom* sequences are *Rattus norvegicus* IC-1A and IC-2C isoforms, respectively. Identical residues (*), conserved residues (:), and semi-conservative substitutions (.) are designated below the aligned sequences. The N-terminal (sequence-based) predicted coiled-coil regions for all sequences are shown in **boldface**. Regions of alternative splicing for the *R. norvegicus* sequences are outlined in **orange**; *R. norvegicus* residues corresponding to Serine-84 (IC-2C isoform numbering) are outlined in **red**. The crystallographically-determined binding sites for dynein light chains Tctex1 and LC8 in *Drosophila* (Hall *et al.* 2010) are highlighted in **yellow** and **green**, respectively. IC Regions 1 and 2 of the bi-segmental p150\textsuperscript{Glued} binding interface in *Drosophila* are designated above the sequence, as are Linker 1 (designated ‘L1’) and Linker 2. **Bottom—** Regions 1 and 2 (same color coding scheme as *top* panel) and linker segments are highlighted on a proposed model for molecular structure of assembled IC; molecular model image adapted with permission from Figure 8 of Morgan *et al.*, 2011 (Morgan *et al.* 2011). (B) Sequence alignment and comparison of the six total isoforms produced from the two dynein IC genes [(DYNC1I1, ‘IC-1’) and (DYNC1I2, ‘IC-2’)] present in the model mammalian species *Rattus norvegicus* (Vaughan *et al.* 1995). The N-terminal predicted (sequence-based, PairCoil2) coiled-coil regions for all sequences are shown in **boldface**. The two variable (alternative splice) regions for the *R. norvegicus* sequences are underlined in **red**, while the crystallographically-determined binding sites for the mammalian dynein light chains Tctex1 and LC8 (Williams *et al.* 2007) are highlighted in **yellow** and **green**, respectively. The ‘serine-rich region’ for all isoform sequences is highlighted in **pink**. *R. norvegicus* IC residues corresponding to Serine-84 and Threonine-89 (IC-2C isoform numbering) are shown in **red** and **blue boldface**, respectively.
Figure 1.3. (Continued)

A

<table>
<thead>
<tr>
<th>Organism</th>
<th>Region 1</th>
<th>L1</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>M-DR----KAELRKKKKAALREKDRR----RREKEIK--DNEEAAGRIGGAGD1DKQKDR-53</td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-1A</td>
<td>MSDKSELKAELRKKKQAQIREEKKEEKREEKKEAKQKEKEF--VDQDSLDR--KRELT-59</td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-2C</td>
<td>MSDKSELKAELRKKKQAQIREEKKEEKREEKKEAKQKQQEDLEK--KRRTE-60</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organism</th>
<th>Region 2</th>
<th>Linker 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>---LDEMLSSLGVAP---------------VSEVLSLSSYNSMTSNSNQPTPASLQAT---96</td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-1A</td>
<td>EALLQSGISPEPLVQPLHFLTWDQFCYFYLVPITPSPKPLVSSTPSE-AGQDO---DLGFLT-119</td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-2C</td>
<td>EALLQSMGILTSDEP---VPPMPSPKPLVSSTPSE-AGQDQSDGAGVGS-105</td>
<td></td>
</tr>
</tbody>
</table>

'Serine-rich region'

Flexible linker, 'L1'

Flexible linker, 'L2'

B

<table>
<thead>
<tr>
<th>Organism</th>
<th>IC-1A</th>
<th>IC-1B</th>
<th>IC-1C</th>
<th>IC-1D</th>
<th>IC-2A</th>
<th>IC-2B</th>
<th>IC-2C</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>MSDKSELKAELRKKKQAQIREEKKEEKREEKKEAKQKEKEF--VDQDSLDR--KRELT-53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-1A</td>
<td>MSDKSELKAELRKKKQAQIREEKKEEKREEKKEAKQKEKEF--VDQDSLDR--KRELT-57</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-1B</td>
<td>MSDKSELKAELRKKKQAQIREEKKEEKREEKKEAKQKEKEF--VDQDSLDR--KRELT-77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-2A</td>
<td>MSDKSELKAELRKKKQAQIREEKKEEKREEKKEAKQKEKEF--VDQDSLDR--KRELT-77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-2B</td>
<td>MSDKSELKAELRKKKQAQIREEKKEEKREEKKEAKQKEKEF--VDQDSLDR--KRELT-77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-2C</td>
<td>MSDKSELKAELRKKKQAQIREEKKEEKREEKKEAKQKEKEF--VDQDSLDR--KRELT-77</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Organism</th>
<th>IC-1A</th>
<th>IC-1B</th>
<th>IC-1C</th>
<th>IC-1D</th>
<th>IC-2A</th>
<th>IC-2B</th>
<th>IC-2C</th>
</tr>
</thead>
<tbody>
<tr>
<td>D. melanogaster</td>
<td>T9MP5SVEKVPSSAGQSD---DLGFLTRLQMDSDPSVQQLQDSL2LRLKLVSKVTPSPS----180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-1A</td>
<td>T9MP5SVEKVPSSAGQSD---DLGFLTRLQMDSDPSVQQLQDSL2LRLKLVSKVTPSPS----180</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-1B</td>
<td>T9MP5SVEKVPSSAGQSD---DLGFLTRLQMDSDPSVQQLQDSL2LRLKLVSKVTPSPS----143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-2A</td>
<td>T9MP5SVEKVPSSAGQSD---DLGFLTRLQMDSDPSVQQLQDSL2LRLKLVSKVTPSPS----143</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-2B</td>
<td>T9MP5SVEKVPSSAGQSD---DLGFLTRLQMDSDPSVQQLQDSL2LRLKLVSKVTPSPS----172</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R. norvegicus IC-2C</td>
<td>T9MP5SVEKVPSSAGQSD---DLGFLTRLQMDSDPSVQQLQDSL2LRLKLVSKVTPSPS----172</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
structure—either well-formed or nascent. Biophysical characterization (primarily NMR spectroscopy) of various constructs of *D. melanogaster* IC has indicated nascent helicity in apo IC for the region that binds LC7 (Benison *et al.* 2006; Nyarko *et al.* 2011). When bound to LC7, *D. melanogaster* IC is known to form two consecutive α-helical segments (largely consistent with sequence-based secondary structural prediction, Figure 1.2b), as seen in the X-ray crystal structure of the IC/LC7 complex (Hall *et al.* 2010); given the lack of fixed, definite secondary structure in apo IC as compared to the well-defined helical structure seen in the bound form, this region of IC is believed to undergo a disorder-to-order transition (see section above, ‘Intrinsically Disordered Proteins—A Brief Introduction’) upon binding to LC7. There is also a predicted coiled-coil region in IC (‘PrCC’, Figure 1.2b) that partially overlaps with the LC7 recognition site. NMR spectroscopy permitted mapping of a self-association domain to residues 222–231 (numbering in the Cdic2b isoform) within *D. melanogaster* IC, with this interaction being disrupted upon the binding of LC7 (Nyarko *et al.* 2011). Biophysical evidence collected for mammalian IC also indicates a self-associative interaction in a region that includes the LC7 binding site (King *et al.* 2003; Lo *et al.* 2006). Whether a true coiled-coil exists in this region is uncertain, but given the relatively short length of the self-association region in *D. melanogaster* IC, it is believed that the interaction is due to simple helix-helix packing between two chains of IC in this region (Nyarko *et al.* 2011).

Sequence-based secondary structure prediction also suggests α-helical content in the N-terminal portion of N–IC, as well as a predicted coiled-coil coincident with the N-terminal α-helix in both *Drosophila* and mammalian ICs (Figure 1.2b, Figure 1.3). My NMR spectroscopic work on *Drosophila* IC lead to identification of a stable and well-formed α-helix within the first ~40 residues, as well as a more nascent helical segment within residues 48–60 (see Chapter 3), with these two experimentally-determined helical segments being in good agreement with sequence-based prediction of secondary structure. However, various experiments and biophysical measurements on apo IC have given no indication of self-association or coiled-coil formation within this N-terminal region of IC in any species. Rather, we believe that the N-terminal α-helix of IC most probably constitutes a stable single α-helix (SAH) (Peckham *et al.* 2009) that does not, by itself, dimerize through coiled-coil formation as previously assumed. It is possible that, when bound to a dimeric partner (such as dynactin p150<sub>GluD</sub>, NudE, *etc.*), this N-terminal helical segment might pack against a copy of itself in some manner, but there is no evidence to indicate that this occurs in the apo state.
In addition to helical content, N–IC is also predicted from sequence to have β-strand structural propensity in a region that overlaps with the Tctex1 and LC8 light chains binding sites (Figure 1.2b). Structural and dynamical NMR data for this apo IC segment indicate flexibility and a lack of any fixed secondary structure (Benison et al. 2006; Morgan et al. 2011). However, the binding recognition sequences in IC for the Tctex1 and LC8 light chains are seen in crystal structures to form definite β-strands at the dimeric interfaces of the light chains when bound (Williams et al. 2007; Hall et al. 2009). As in the case of binding LC7, IC is believed to also undergo a localized disorder-to-order transition upon binding to Tctex1 and LC8. While there is no evidence of β-strand or fixed secondary structure in this region of apo IC, a recent and more detailed NMR analysis has revealed significant Polyproline II structural content (greater than that in a random coil of the same amino acid sequence) in the regions interacting with Tctex1 and LC8; it is believed that this increased Polyproline II structural content may facilitate initial recognition and docking of this region of an IC chain on the surface of the light chain molecule, just prior to the final binding interaction and assumption of β-strand structure by IC (see Chapter 4).

When bound by the dynein light chains, NMR data demonstrate that IC remains largely disordered and flexible in regions outside the light chains binding sites (Benison et al. 2006; Nyarko et al. 2011; Morgan et al. 2011). Thus, in its assembled state, dynein IC exists as a flexible polyvalent molecular scaffold (see Figure 3.8), with the dimeric light chains helping to assemble and stabilize the IC scaffold. The retained disorder in assembled IC is an important property of the complex in that, it provides a flexible platform for binding other cargo in the remaining exposed sections of IC, and it may also potentially facilitate regulatory post-translational modification (such as phosphorylation; see below) via this molecular flexibility and accessibility.

In addition to the three dynein light chains, N–IC also mediates interaction between dynein and the dynein-modulating dynactin complex (see below, ‘Regulators of the Cytoplasmic Dynein Motor Complex’) by binding directly with dynactin’s p150Glued subunit (Karki et al. 1995; Vaughan et al. 1995). Earlier coarse-grained methods mapped the p150Glued interaction region to residues 1–123 in the rat IC–1A isoform (see below, and Figure 1.3), which corresponds to the region of IC that is N-terminal to the Tctex1 light chain binding site (Vaughan et al. 1995). More recently, the p150Glued binding site was mapped in D. melanogaster IC at residue-level resolution using NMR spectroscopy; a bi-segmental binding footprint for p150Glued on IC was determined (Figure 1.3a), comprising residues 1–41 (‘Region 1’) and 46–75 (‘Region 2’) in
Thus, regions of IC containing helical content participate in the binding interaction with dynactin p150Glued; in that, IC residues 1–40 (within ‘Region 1’) constitute a stable and well-formed \( \alpha \)-helix, while a nascent helical segment was identified within residues 48–60 (part of ‘Region 2’) in apo IC. It is believed that additional ordering of the IC chains may occur in these two helical segments upon binding interaction with p150Glued, while the intervening IC residues 42–45 (‘Linker 1’) remain disordered, as do the residues 76–109 (‘Linker 2’) when IC is assembled with p150Glued and the dynein light chains (see Chapter 3).

Subsequently, the binding site of dynein regulator NudE (see below, ‘Regulators of the Cytoplasmic Dynein Motor Complex’) was also mapped in *D. melanogaster* IC at residue-level resolution using NMR spectroscopy. NudE was demonstrated to bind IC residues 1–40 (e.g. ‘Region 1’, Figure 1.3a, Figure 1.4b), and binding of NudE and dynactin p150Glued were seen to be mutually exclusive, with displacement of NudE from IC by p150Glued (Nyarko et al. 2012). Using more coarse-grained methods (*i.e.* chromatographic analysis), the same conclusions were drawn in the case of mammalian dynein (*R. norvegicus* IC–2C isoform)—namely, that NudE and dynactin p150Glued bind to overlapping, but not identical sites within the N-terminal region of IC, and that the two compete with one another for binding to this region, with p150Glued seen to displace NudE from IC (McKenney et al. 2011).

**Isoform Diversity of the Intermediate Chain.**

From the single gene (*Cdic*) encoding the dynein intermediate chain in *Drosophila*, at least 10 alternative splice isoforms of the protein are produced, with the splicing pattern exhibiting tissue-specificity; in addition to a constitutive set, tissue-specific IC isoforms were observed in *Drosophila* ovaries and neural tissue, where tissue-specific kinds of dynein-dependent transport take place (Nurminsky et al. 1998). However, the alternative splice regions in *Drosophila* only occur in a region of IC that is C-terminal to where light chain LC8 binds, and N-terminal to the LC7 binding site (Nurminsky et al. 1998)—and not in the N-terminal-most region of N–IC where the regulatory dynactin p150Glued and NudE proteins bind.

As mentioned earlier, the dynein intermediate chain is encoded by two different genes in mammalian species [*DYNC1I1*, ‘IC-1’) and (*DYNC1I2*, ‘IC-2’) (Pfister et al. 2006). In contrast to *Drosophila*, the IC genes in mammalian species (human, rat, mouse, etc.) each possess two variable (alternative splice) regions that are N-terminal to the binding site of light chain Tctex1 in IC (Figure 1.3b) (Vaughan et al. 1995; Crackower et al. 1999; Kuta et al. 2010). As in the case of *Drosophila*, mammalian IC isoforms (the product of alternative splicing in the two variable
**Figure 1.4.** Models for dynein intermediate chain assembled with the light chains and with dynactin p150\(^{\text{Glued}}\) or NudE, and a speculative model for the binding of ZW10 to IC. Segments of IC for which NMR spectral characteristics are altered upon partner binding are indicated in **red** (‘Region 1’, residues 1–41) and/or **purple** (‘Region 2’, residues 46–75); the dynein light chains Tctex1 and LC8 are shown in their bound states (**yellow** and **green**, respectively), as depicted in Figure 1.3a, *bottom*. (A) The bi-segmental binding footprint for a coiled-coil segment of dynactin p150\(^{\text{Glued}}\) on *Drosophila* IC (involving both Regions 1 and 2), as determined by NMR spectroscopy (Morgan *et. al.* 2011). (B) The NMR-mapped binding site (‘Region 1’ only) for a coiled-coil segment of NudE on *Drosophila* IC (Nyarko *et. al.* 2012). (C) A speculative model for the binding site of ZW10 (whose structure is unknown) on IC, based upon some preliminary NMR spectroscopy data collected for *Drosophila* IC (see Appendix 2), and based also upon biochemical assays in mammalian (rat) species, of ZW10 binding activity with certain phosphoisoforms of mammalian IC (Whyte *et. al.* 2008).
Figure 1.4. (Continued)

A

\( p_{150}^{\text{Glued}} \)

Flexible linker, 'L1'

Flexible linker, 'L2'

B

NudE

C

ZW10

??
N-terminal regions) also exhibit tissue-specific and temporal (developmental stage-specific) expression patterns. For example, whereas IC-2 isoforms are expressed ubiquitously throughout all rat tissue types (with the IC–2C isoform being present in all tissues and cells examined), IC-1 isoform expression is restricted to rat brain and nervous tissues (Vaughan et al. 1995; Pfister et al. 1996a; Crackower et al. 1999). Further, IC-1 expression is also cell-specific; within the adult rat brain (a mixture of cell types), cortical neurons and glia exhibit different populations of IC isoforms. Glial cells contain only the IC–2C splice isoform, whereas neurons possess four IC protein isoforms, including IC–1A (found only in neurons), IC–1B, IC–2B, and IC–2C (Pfister et al. 1996a). Incidentally, in vitro solution-state binding assays performed on IC isoforms from R. norvegicus have demonstrated that the IC–1A isoform (Figure 1.3b) has a significantly higher binding affinity for dynactin p150Glued than the other isoforms, while IC–2B and IC–2C have the lowest binding affinities of all (E. L. F. Holzbaur, unpublished data). IC–1A is only found in adult neurons, suggesting that this particular isoform has neuron-specific function.

In addition, during rat brain development, the IC–2C isoform is the only one found in neurons prior to E14 (embryonic day 14); between E15 and the fifth day after birth (P5), the relative expression of the neuronal IC protein isoforms changes, demonstrating that expression of IC isoforms is also developmentally (e.g. temporally) regulated in the brain (Pfister et al. 1996b; Susalka et al. 2000). The E15–P5 time period is one of rapid process (axon) extension and initial pattern formation in the rat brain; nerve growth factor-induced neurite differentiation triggers changes in the composition of the IC isoform pool, suggesting that specific IC isoforms play a role in establishment or regulation of retrograde axonal transport during neurite extension (Salata et al. 2001).

In light of the observed IC isoform diversity and differential tissue-type expression of isoforms, one early hypothesis was that the particular versions of IC chains present within dynein complexes might confer cargo binding specificity to individual motor complexes or otherwise adapt the motor complex pool to the types of transport needed by a particular cell type (Pfister et al. 1996a). Over time, evidence to support this hypothesis has gradually accumulated. For example, in rat optic nerve cells it was demonstrated that the IC–2C isoform is the only IC protein associated with membranous cargoes in the fast component of anterograde axonal transport, whereas the slow component was enriched in the IC–1A, IC–1B, and IC–2B isoforms (Dillman et al. 1996; Susalka et al. 2000). In addition, signaling endosomes containing neurotrophin receptor tyrosine kinase B (TrkB) were observed to be transported down the axon terminus toward the rat neuron cell body by cytoplasmic dynein containing the neuron-specific IC–1B
isoform and not by dynein complexes containing the ubiquitous IC–2C isoform; conversely, nerve growth factor TrkA-containing endosomes were transported specifically by dynein complexes containing the IC–2C isoform (Ha et al. 2008). Thus, there is evidence demonstrating that alternatively spliced intermediate chain isoforms provide a basis for dynein cargo specificity and complex function in mammalian cells.

Residue-resolution mapping (as by protein NMR spectroscopy) of the binding interface of IC with dynactin p150\textsuperscript{Glued} in Drosophila suggests an explanation of how dynein-dynactin cargo-binding specificity might be achieved through IC isoform diversity in mammalian species. Through inter-species sequence alignment and extrapolation of the results obtained in Drosophila to mammalian IC (rat, for example), there are small sequence differences in ‘Region 1’ of mammalian IC-1 vs. IC-2 isoforms (Figure 1.3b) that could impact the binding affinity of this region for dynactin p150\textsuperscript{Glued}. More significantly, ‘Region 2’ of the p150\textsuperscript{Glued} binding ‘footprint’ overlaps with the first of two alternative splice regions in this N-terminal region of mammalian IC, generating significant differences among rat IC-1 isoforms (Figure 1.3) and also differences among the IC-2 isoforms. The R. norvegicus IC–1A isoform (as mentioned above, preliminary results indicate that this particular isoform has the strongest binding affinity for dynactin p150\textsuperscript{Glued}) has a significantly extended sequence in this region, relative to the IC–1B and IC–1C isoforms (Figure 1.3b).

In addition, there is further isoform diversity in mammalian ICs due to the second of two alternative splice regions in this N-terminal region of mammalian IC; this second splice region overlaps with ‘Linker 2’ that spans from the C-terminal end of the p150\textsuperscript{Glued} binding region to the N-terminus of the Tctex1 binding site (Figure 1.3). Alternative splicing produces variability in the length of this linker that could modulate IC binding affinity for dynactin p150\textsuperscript{Glued} or other regulatory binding partners via the ‘multivalency effect’ (Hall et al. 2009). Pre-association of IC with a bivalent binding partner such as the dynein light chains Tctex1 or LC8 essentially increases the effective local concentrations of proximate regions in the two IC chains, thus potentially enhancing the binding interaction with another bivalent partner at this proximate site. The extent of binding enhancement for another protein (such as dynactin p150\textsuperscript{Glued}) is expected to depend upon the length of the flexible linker, with a shorter linker segment expected to produce greater enhancement of binding affinity than a longer one. The IC-LC8 (‘Pac11-Dyn2’ in yeast) interaction has been demonstrated in Saccharomyces to enhance the binding of p150\textsuperscript{Glued} (‘Nip100’ in yeast) \textit{in vitro} and recruitment of dynactin to dynein \textit{in vivo}; however, it should be noted that the Nip100 and nearest Dyn2 binding sites in Pac11 are separated by ~20 residues.
(Stuchell-Brereton et al. 2011; Siglin et. al. 2013). In contrast, ITC experiments with *Drosophila* IC:1–143 showed no enhancement of binding affinity for p150\textsuperscript{Glued} with light chains pre-bound to IC (Barbar Lab, *unpublished results*); the length of ‘Linker 2’ in this case is 34 residues—a bit longer than the 20 residues in yeast. In the case of mammalian IC isoforms (assuming that the binding site of p150\textsuperscript{Glued} is roughly analogous in mammalian species to that determined in *Drosophila*) the rat IC–2C isoform would have a ‘Linker 2’ length of 26 residues—shorter than that seen in *Drosophila* (Figure 1.3).

*Regulatory Phosphorylation of the Intermediate Chain.*

In addition to the IC isoform diversity generated by alternative splicing, early work also helped to demonstrate the existence of different IC phosphoisoforms in brain tissue (Pfister et. al. 1996a,b), as well as differential phosphorylation of subunits in dynein pools associated with particular functions in neural tissues (Dillman et. al. 1994). In cells that undergo mitosis (HeLa cells, for example), a temporal aspect of IC phosphorylation was demonstrated in that, only non-phosphorylated IC was detected in interphase cells whereas most of the dynein IC was (serine) hyperphosphorylated in M phase (mitosis); this further suggested that phosphorylation of the IC subunit might serve as a mechanism to alter the function of dynein at particular cell stages (Huang et. al. 1999).

In the dynein intermediate chains of all metazoans characterized thus far, there exists a conserved ‘serine-rich’ domain that is located C-terminal to the predicted coiled-coil region at the N-terminus, and N-terminal to the Tctex1/LC8 light chains binding domain (see Figure 1.3a). To-date, there is still no evidence demonstrating that phosphorylation of IC residues in the ‘serine-rich’ region occurs in *Drosophila*, let alone the impact that any such phosphorylation might have on dynein function in this organism; though, the conservation of this ‘serine-rich’ region across species and throughout evolution suggests an important function. There is, however, accumulating evidence of specific phosphorylation sites in the ‘serine-rich’ region in mammalian (rat and human) IC, as well as *in vivo* demonstration of the impact(s) that these specific IC phosphorylations have on dynein function. Note that the ‘serine-rich’ domain contains several potentially phosphorylatable serines as well as one threonine residue, and is highly conserved (essentially identical) among the rat IC isoforms (Figure 1.3b), with the exception of one additional serine residue being present at the C-terminus of the ‘serine-rich’ domain in rat IC-2 isoforms as compared to IC-1 isoforms. Also, the ‘serine-rich’ domain is flanked by adjoining sequences that differ between the rat IC-1 and IC-2 isoforms; for the serine
residues at the termini of the ‘serine-rich’ domain, this could potentially generate alternative consensus recognition sequences for different cellular kinases.

The first mapping of an IC phosphorylation site was Serine-84 in the rat IC-2C isoform (Figure 1.3) from liver cells; phosphorylation of this specific residue abrogated association of IC with dynactin p150^Glued in vitro and the association of dynein with dynactin in vivo, producing defects in late endosome and lysosome organelle transport and localization (Vaughan et. al. 2001; Towns et. al. 2009). Residue-level resolution mapping by NMR indicated a bi-segmental binding interface of IC with dynactin p150^Glued in Drosophila and isothermal titration calorimetry (ITC) demonstrated that both Regions 1 and 2 of IC contribute energetically to the binding interaction with p150^Glued, suggesting that there might possibly be direct interaction of both Regions 1 and 2 of IC with p150^Glued (see Chapter 3). Through inter-species sequence alignment and extrapolation of the results obtained in Drosophila to mammalian (rat) IC, the C-terminal end of ‘Region 2’ overlaps with Serine-84 (rat IC-2C isoform numbering) (Figure 1.3a); this suggests (as a possible mechanism) that phosphorylation of Serine-84 might potentially block any direct interaction between p150^Glued and at least part of ‘Region 2’ in IC, thereby decreasing the overall binding affinity between IC and p150^Glued, and thus dynein and dynactin. In light of the bi-segmental IC binding footprint of dynactin p150^Glued, the overlapping (but not identical) IC binding footprint of NudE (see above), and the competition observed between p150^Glued and NudE for binding to IC, a regulatory model has been proposed (Figure 1.5) wherein phosphorylation of sites (such as Serine-84 in rat IC-2C isoform) within the ‘serine-rich’ domain of IC could block the binding of p150^Glued, allowing NudE to be ‘selected’ for association with IC and thus dynein in the cell.

In addition to Serine-84, another specific IC phosphorylation site was identified as Threonine-89 in the rat IC-2C isoform (Figure 1.3b); comparison of kidney cell extracts from interphase and mitotic cells revealed phosphorylation of mitotic (but not interphase) dynein IC (Whyte et. al. 2008). Phosphorylation of the Threonine-89 IC residue was seen to significantly reduce its binding to dynactin p150^Glued (much like phosphorylation of the Serine-84 residue discovered in earlier work). However, phosphorylation of Threonine-89 was required for the interaction between IC and ZW10 (of the RZZ complex, see section below ‘Regulators of the Cytoplasmic Dynein Motor Complex’) and this specific form of phosphodynein was seen to accumulate at kinetochores from the time of nuclear envelope breakdown up until metaphase in cells (Whyte et. al. 2008). Results in vivo implicated a direct interaction between this form of phosphodynein and ZW10 as the basis for initial dynein recruitment to kinetochores, and the IC phosphorylation state at Threonine-89 as a basis for its differential interactions with ZW10 and
**Figure 1.5.** Possible regulatory model for selective IC binding to dynactin p150Glued or NudE when both regulatory partners are present in the cell. Dynactin p150Glued binds dynein IC Regions 1 and 2, making dynactin-bound dynein the major product (*left*). Phosphorylation at sites within the unstructured serine-rich domain (overlaps with the C-terminus of Region 2 and the N-terminus of Linker ‘L2’) of IC decreases the binding affinity for p150Glued (Vaughan *et. al.* 2001; Whyte *et. al.* 2008), and could produce selective binding to NudE (*right*). All three proteins are dimeric in the assembled state, but for simplicity are drawn as half-dimers. Figure adapted with permission from Nyarko et al., 2012 (Nyarko *et. al.* 2012).
dynactin $p15^{Glued}$ (Whyte et al. 2008). Some very preliminary NMR results (see Appendix 2, Figure A2.4) suggest that *Drosophila* ZW10 may exhibit a bi-segmental binding footprint on IC, much like that observed for dynactin $p15^{Glued}$ (Figure 1.4). Inter-species IC sequence alignment and extrapolation of the binding interface determined for $p15^{Glued}$ in *Drosophila* to mammalian (rat) IC, suggests that Threonine-89 (rat IC-2C isoform numbering) may lie outside of ‘Region 2’ as defined in *Drosophila* (Figure 1.3a). However, the precise IC binding interfaces for regulatory proteins (such as dynactin $p15^{Glued}$, NudE, ZW10, etc.) remain to be determined in mammalian species (see Chapter 5, ‘Future Work’ section). It may be the case that IC Threonine-89 (rat IC-2C isoform numbering) does overlap with the IC binding sites for $p15^{Glued}$ and ZW10 in mammalian species, in which case, phosphorylation of this residue may act to decrease binding affinity for $p15^{Glued}$ and increase affinity for binding ZW10, through the direct interactions that these regulatory proteins make with this particular region of IC.

In the work of Whyte et al., 2008, phosphorynem (containing IC phosphorylated at Threonine-89, ‘ICpT89’) was seen to disappear from kinetochores by the time of metaphase chromosome alignment; the reduction in kinetochore ICpT89-dynein that had been present through early prometaphase up until the time of metaphase, was coupled with an increase in total dephosphodynein (containing IC not phosphorylated at Threonine-89) along kinetochore microtubules and at the spindle poles—the result of ‘poleward streaming’ that occurs upon bioriented microtubule attachment and chromosome alignment during mitosis (Whyte et al. 2008). Reduction of dynactin levels at the kinetochore did not alter the time progression or levels of ICpT89-dynein at kinetochores (suggesting that this form of phosphorynem is recruited to kinetochores independently of dynactin), but did result in metaphase arrest/delay in the cells. Inhibition of a particular kinetochore-localized protein phosphatase (PP1-$\gamma$) in cells prevented the dephosphorylation of ICpT89-dynein and resulted in its accumulation at the kinetochore and metaphase arrest/delay; inhibition of ICpT89 dephosphorylation also inhibited the usual poleward streaming of dynein–dynactin and resulted in sequestration of dynactin and spindle assembly checkpoint proteins at the kinetochore (Whyte et al. 2008). On the collective basis of these observations, a model was proposed for the regulation of cytoplasmic dynein at kinetochores (Figure 1.6) wherein ICpT89-dynein is recruited to kinetochores by direct interaction with ZW10, and then ICpT89 becomes dephosphorylated upon bioriented microtubule attachment and chromosome alignment, thus allowing the dephospho-IC to bind to $p15^{Glued}$ of dynactin (instead of to ZW10), with spindle assembly checkpoint (SAC) proteins (and the ZW10/RZZ complex, see below) being transported toward the spindle poles through their association with dynein-
Figure 1.6. Proposed sequential model for recruitment and regulation of cytoplasmic dynein at kinetochores. Dynactin (red) can be independently recruited to the kinetochore through direct binding of its p50 (dynamitin) subunit to ZW10 of RZZ (Starr et. al. 1998). Polo-like kinase 1 (Plk1) is a mitotic dynein kinase demonstrated to phosphorylate a specific threonine in IC (upper left-hand corner) (time point $t = 0$). During prometaphase, dynein is targeted to and loaded onto kinetochores via binding of phosphodynein (containing IC phosphorylated at Threonine-89; residue numbering is for the rat IC-2C isoform) to ZW10 (time point $t = 1$). As microtubules attach to kinetochores and chromosomes congress to the metaphase plate (time point $t = 2$), dynein undergoes IC dephosphorylation (by PP1-γ protein phosphatase), shifting its binding from ZW10 to dynactin. This transition stimulates dynein translocation away from kinetochores and toward the spindle poles (‘poleward streaming’), allowing spindle assembly checkpoint (SAC) proteins to be removed from kinetochores (time point $t = 3$), thus allowing progression of the cell into anaphase. Figure adapted with permission from Bader et. al., 2011, and Whyte et. al., 2008 (Bader et. al. 2011; Whyte et. al. 2008).
dynactin (Whyte et al. 2008).

Implicit in this model, is the supposition that the phosphorylation state of certain residues in the ‘serine-rich’ region of IC can undergo post-translational chemical modification when IC is already associated with a binding partner—one whose binding recognition site may overlap with these phosphorylatable IC residues. The precise IC binding site of ZW10 in mammalian species is currently unknown, and whether or not the IC ‘serine-rich’ region remains sufficiently flexible and accessible to kinases or phosphatases in the ZW10 (or p150\textsuperscript{Glued})-bound state remains to be directly demonstrated biochemically, \textit{in vitro}. However, at least in the case of the \textit{Drosophila} IC–p150\textsuperscript{Glued} complex, NMR dynamics experiments demonstrated that there was a degree of retained disorder and flexibility in IC residues at the C-terminal end of ‘Region 2’ (see Figure 3.4), which includes residues in the ‘serine-rich’ region. It is worth noting that, in earlier work with rat dynein IC–1A, Casein kinase II (CKII) was discovered to be a dynein kinase that could bind directly to and phosphorylate dynein IC (the first 120 residues of IC were sufficient for CKII binding) (Karki et al. 1997). Intriguingly, when the rat IC was pre-bound with excess dynactin p150\textsuperscript{Glued}, CKII was still able to bind to IC; however, no biochemical assay was performed in this study to determine whether IC phosphorylation by CKII occurred when p150\textsuperscript{Glued} was bound to the IC (Karki et al. 1997).

In summation of this subsection on the dynein Intermediate Chain, I would classify and highlight the IC protein as an IDP exemplar, exhibiting so many of the features (or, ‘bells and whistles’) typically found in IDPs (discussed above, ‘Intrinsically Disordered Proteins—A Brief Introduction’ section). N–IC is primarily disordered but interspersed with different elements and degrees of secondary (and even tertiary, see Chapter 4) structure. It has been demonstrated to undergo localized folding-coupled-to-binding transitions upon association with dynein light chains, forming either \(\alpha\)-helices or \(\beta\)-strands in its bound states. Its disorder and flexibility underlie its ability to interact with myriad binding partners—several of which (dynactin p150\textsuperscript{Glued}, NudE, ZW10, reviewed below in ‘Regulators of the Cytoplasmic Dynein Motor Complex’) have overlapping binding sites on IC—and to function as a flexible polybivalent scaffold for molecular assembly in the dynein motor complex. Not only does its flexibility allow it to associate with multiple non-dynein regulatory partners (several using the same basic region at the N-terminus of IC), but in its disorder, it is also amenable to modulation of its composition in the form of alternative splice protein isoforms and post-translational modification (which are also properties facilitated by protein intrinsic disorder, discussed earlier). Thus, the disorder in IC facilitates its
versatile interaction with multiple different non-dynein regulatory proteins, and also presents a basis for regulation of its own interaction with those binding partners via its composition (in terms of alternative splice isoforms and differential phosphorylation).

**Regulators of the Cytoplasmic Dynein Motor Complex**

Like dynein, kinesins (the other microtubule-based motor in cells) function in a wide array of cellular activities such as organelle transport and positioning and in cell division (Hirokawa 1998). These functions are performed by greater than 40 distinct kinesin motors, many of which have a particular functional specialization (Kim et. al. 2000). In contrast, there is one basic cytoplasmic dynein motor performing a similar diversity of functions. Dynein’s functional specialization and association with a wide range of cargoes are believed to be governed by several mechanisms: 1) heterogeneity in dynein composition via isoform diversity of its comprising subunits (discussed above, ‘Multi-subunit Composition of the Cytoplasmic Dynein Motor’), 2) post-translational modification—such as phosphorylation—of dynein subunits (also discussed above), and 3) association of the dynein motor complex with a variety of non-dynein regulators and functional adaptors (reviewed in Kardon et. al. 2009b). In this section, I provide a brief overview of three key non-dynein protein complexes that provide regulation and/or expansion of the functional repertoire of the dynein motor complex. Each of these three regulatory complexes (Dynactin, ZW10/RZZ, and NudE/EL) also contain protein components that have been demonstrated to mediate interaction with dynein through direct binding to the dynein intermediate chain—the central protein of this thesis.

**Dynactin.**

Dynactin (*Dynein activator*, 1.2 MDa, 20S) is a heteromultimeric macromolecular complex, first identified through its role in activation of vesicular transport mediated by cytoplasmic dynein (Schroer et. al. 1991; Gill et. al. 1991), and is known also to increase the processivity of transport by dynein along the microtubule (King et. al. 2000b; Ross et. al. 2006; Kardon et. al. 2009a). Dynactin is implicated as an activator or cofactor in nearly all functions of dynein (Karki et. al. 1999) and also serves as a cargo adaptor (Figure 1.2a), allowing dynein to associate with diverse cargoes, especially those that are membranous (Waterman-Storer et. al. 1997; Steffen et. al. 1997; Habermann et. al. 2000). Though dynein is capable of binding to a limited number of cargoes directly through some of its subunits (discussed above), association
with dynactin significantly expands the repertoire of cargo that can be transported by dynein and also helps target dynein to specific cellular locations (Vallee et al. 2012; Raaijmakers et al. 2013). Due to its role as a cargo adaptor, as well as the manner and location in which it associates with dynein (e.g. IC–p150Glued connection), I have often conceptualized dynactin as supplying a pair of prosthetic “arms” that are attached at the upper “torso” of dynein.

Dynactin is the most prominent and extensively characterized regulatory complex of the dynein motor. The 1.2 MDa dynactin complex is comprised of 11 different subunit types, ranging in size from 22 to 150 kDa (Schroer 2004; Hammesfahr et al. 2012); several components are present as dimers or oligomers in the complex, with a total of ~20 polypeptide chains per dynactin molecule (Schafer et al. 1994). The structure of dynactin can be divided into two functionally and morphologically distinct domains: the Arp1 rod and the projecting arm (Figure 1.7a). Among other components, the Arp1 rod domain contains about eight Arp1 (Actin-related protein 1, a.k.a. ‘Centractin’) protein molecules that form a short actin-like filament to which the other subunits of dynactin bind (Schafer et al. 1994; Bingham et al. 1998). Association of Arp1 with spectrin-family proteins is believed to provide a general mechanism by which dynactin binds a variety of membranous cargoes (Holleran et al. 1996); it has been demonstrated that dynactin Arp1 binds directly to βIII spectrin, a specialized spectrin isoform found on the cytosolic surface of the Golgi and other cellular membranes, thus allowing dynactin to provide a link between the dynein motor complex and membranous cargo (Holleran et al. 2001; Muresan et al. 2001). Disruption of dynactin’s association with dynein in interphase cells disrupts intracellular membrane organelle distribution and produces defects in intracellular transport between organelles, such as ER-to-Golgi transport (Burkhardt et al. 1997; Presley et al. 1997).

The projecting arm domain of dynactin (Figure 1.7a) is primarily composed of 2 copies of the p150Glued protein (gene DCTN1 in mammalian species), but also contains 2 copies of p24/22 (gene DCTN3 in mammalian species) and 4 copies of p50 (a.k.a. dynamitin; gene DCTN2 in mammalian species) that help to attach the C-terminal half of p150Glued to the Arp1 rod (Bingham et al. 1998; Schroer 2004; Jacquot et al. 2010). Dynactin’s largest subunit, p150Glued, was first discovered as a 150 kDa polypeptide that co-purified with cytoplasmic dynein (Collins et al. 1989), and was subsequently determined to have extensive sequence and functional similarity with the product of the Drosophila gene Glued that encodes a polypeptide of very similar size (Swaroop et al. 1987; Holzbaur et al. 1991; Waterman-Storer et al. 1996).

The N-terminal ~150 residues of p150Glued contain the microtubule-binding and dynein processivity-enhancing functionalities of the dynactin complex (cytoskeleton-associated protein,
Figure 1.7. Schematic diagrams of the dynactin complex and its components, and its association with the dynein motor. (A) Schematic illustrating the components and structural features of the dynactin complex. The rod-like Arp1 domain is ~10 nm wide and ~40 nm long and is responsible for dynactin binding to membranous cargo; the projecting arm domain (primarily composed of p150\textsuperscript{Glued}) including its globular heads (N-terminal p150\textsuperscript{Glued} CAP-Gly domains) extends ~24 nm from the base of the Arp1 domain to the microtubule (Schafer et al. 1994). Dynein binds along or near the base of this projecting arm (dynein intermediate chain binds to dynactin p150\textsuperscript{Glued}) as depicted in (B). (C) Schematic illustration of the domain organization and associated functions of dynactin p150\textsuperscript{Glued}, containing an N-terminal globular microtubule-binding domain and multiple coiled-coil domains. The ‘CC1A’ sub-domain of the first coiled-coil corresponds to the long rod-like segment of the projecting arm observed in imaging studies, while the ‘CC1B’ sub-domain (to which dynein binds; McKenney et al. 2011) is believed to be more spatially proximate to the Arp1 rod domain. (D) Domain organization and associated residue numbering for the D. melanogaster dynactin p150\textsuperscript{Glued} subunit. Also shown (below) are the sequence-based predicted (JPred3; Cole et al. 2008) \(\alpha\)-helical segments (magenta cylinders) and coiled-coil regions (PairCoil2; McDonnell et al. 2006) (red double lines) for the ‘CC1’ domain of Drosophila p150\textsuperscript{Glued}. Images adapted with permission: (A and C) from Schroer 2004 (Schroer 2004), and (B) from Siglin et al., 2013 (Siglin et al. 2013).
Figure 1.7. (Continued)
glycine rich) domain at the very tip of the dynactin projecting arm (Figure 1.7c). This N-terminal region is proceeded by two segments (‘CC1’ and ‘CC2’, Figure 1.7c,d) that are predicted on the basis of primary sequence to contain extensive α-helical regions including long, extended heptad-repeats characteristic of an elongated coiled-coil. Given the dimeric nature of p150Gluα, coupled with images from deep-etch electron microscopy and other techniques on the dynactin complex (Schafer et al. 1994; Imai et al. 2014), these observations have all led to the widely-accepted model in which p150Gluα forms a parallel homodimer with the ‘CC1’ coiled-coil structure largely constituting the projecting arm of dynactin (Figure 1.7). The ‘CC1’ domain is confirmed by biophysical measurements to be an α-helical coiled-coil and has been conceptually subdivided into two segments, ‘CC1A’ and CC1B’ with a short intervening flexible linker (Figure 1.7c,d); in rat p150Gluα, the ‘CC1A’ fragment was demonstrated to be a stronger dimer and more stable coiled-coil than ‘CC1B’, with ‘CC1A’ aiding in the dimerization and stabilization of the coiled-coil in the ‘CC1B’ sub-domain (Siglin et al. 2013). Binding to the dynein intermediate chain is mediated through ‘CC1B’ of dynactin p150Gluα (McKenney et al. 2011), and it has been suggested that the interaction is mediated more specifically by the C-terminal half of ‘CC1B’ (Siglin et al. 2013); thus, dynein (via its intermediate chain) is believed to bind dynactin p150Gluα in a region along or near the base (extending from the Arp1 rod domain) of the projecting arm (Figure 1.7b).

Functionally, the importance of dynactin, as well as the interaction of dynein with dynactin, is particularly pronounced and has been well-characterized in neurons, where the attachment of vesicular cargo to dynein occurs through its association with dynactin, and where the processivity enhancement due to dynactin is believed to aid in the retrograde transport of such cargoes by dynein over the long distances found in nerve axons (Waterman-Storer et al. 1997; Puls et al. 2003; Levy et al. 2006). Mutant dynactin p150Gluα mice displayed defects in vesicular transport in motors neurons, as well as axonal swelling and axo-terminal degeneration (Laird et al. 2008). Disruption of the dynein-dynactin interaction itself in mice inhibited axonal transport and caused motor neuron degeneration resembling that observed in amyotrophic lateral sclerosis (ALS, a.k.a. ‘Lou Gehrig’s Disease’) (LaMonte et al. 2002).

In addition to roles in neuronal transport, dynein and dynactin become associated with the outer nuclear membrane prior to the onset of mitosis (Salina et al. 2002), and dynactin has a demonstrated requirement as a dynein recruitment factor to the nuclear envelope; dynactin also serves an essential function in centrosomal anchoring to the nuclear envelope up until the end of prophase in cells (Raaijmakers et al. 2013). Dynactin and dynein also localize to astral
microtubules and sites of the mitotic cell cortex, where they are believed to participate in the positioning and anchoring of spindle poles to the cell cortex (Busson et al. 1998); it is proposed that dynactin binds directly to cortical network filaments (such as spectrin proteins, as through its Arp1 rod domain), and binds also to dynein, whose anterograde motion along astral microtubules toward the spindle pole generates tension, thus helping to separate the spindle poles in anaphase B of mitosis (Sharp et al. 2000a). The association of dynein with dynactin is also particularly important for the stripping of SAC proteins and ZW10/RZZ (see below) from kinetochores during prometaphase-to-metaphase (Figure 1.6) (Wojcik et al. 2001; Howell et al. 2001; Whyte et al. 2008; Bader et al. 2010). In addition to linking the two main structural domains of dynactin (Arp1 rod and projecting arm) to one another, p50 (dynamitin) is also prominent for its role in directly linking dynactin to the ZW10 protein of the RZZ complex (Starr et al. 1998).

**Rod-ZW10-Zwilch (RZZ).**

The RZZ complex (also referred to as the ‘ZW10 complex’ or ‘ZW10/RZZ’) is composed of three different protein subunits: Rough deal (‘Rod’, ~240 kDa), Zeste-White 10 (‘ZW10’, ~85 kDa), and Zwilch (~75 kDa), which are unique to metazoans (Scaerou et al. 2001; Williams et al. 2003). Given that the RZZ complex exhibits of mass of ~800 kDa (based upon size exclusion chromatography), and given the masses of the individual protein subunits themselves, it has been hypothesized that the complex might potentially consist of 2 copies of each of the three protein subunits (Williams et al. 2003; Karess 2005). A crystal structure has been determined for the RZZ component protein Zwilch (Civril et al. 2010) however, structural and biophysical information is lacking in the literature for the other two components, including ZW10. The Drosophila version of ZW10 (see Appendix 2) consists of 721 amino acid residues and is predicted on the basis of primary sequence to contain numerous α-helical segments (greater than 50% α-helical content) and only a few short β-strand segments; ZW10 does not appear (from sequence-based prediction) to possess any structural domains known in other proteins (Karess 2005).

Components of the RZZ complex were first discovered in *Drosophila*, where mutations in the zwilch, rod, or zw10 genes produced severe mitotic defects including lagging chromosomes at anaphase and disruption of chromosome segregation (including nondisjunction) in both mitotic and meiotic divisions, leading to significant levels of aneuploidy and consequent organismal lethality; the same phenotypes have also been observed for mutants of the homologous genes in vertebrate species (Smith et al. 1985; Karess et al. 1989; Savoian et al. 2000; Williams et al.
Given their role in chromosome segregation and the chromosomal instability associated with aneuploidy, it is not surprising that gene mutations of RZZ components are relatively frequent in certain types of cancer (Wang et al. 2004a). The critical role of RZZ and its components in mitosis and chromosome segregation is explained by the specific functions that have been attributed to the complex, including its requirement and function in the initial recruitment of spindle assembly checkpoint (SAC) proteins (particularly the Mad1–Mad2 complex; Mad = ‘Mitotic-arrest deficient’) to unattached kinetochores during late prophase (Kops et al. 2005; Buffin et al. 2005). In the absence of critical SAC proteins, the mitotic checkpoint is compromised, allowing premature anaphase to occur despite the presence of chromosomes that are unattached to the mitotic spindle.

In addition to Mad1–Mad2 complex recruitment, ZW10/RZZ is also heavily implicated in the recruitment of both dynein and dynactin to the kinetochore in mitosis; the ZW10 protein is able to bind directly to a phosphorylated form of the dynein intermediate chain protein (discussed earlier), and has also been demonstrated to directly bind the p50 (dynamitin) subunit of dynactin (Starr et al. 1998; Whyte et al. 2008). Recruitment of dynein to ZW10/RZZ at the kinetochore occurs largely through direct interaction of phosphorylated dynein intermediate chain with ZW10 protein (Whyte et al. 2008) though, some dynein recruitment may potentially occur indirectly via dephosphodynein (containing dephosphorylated IC) binding to dynactin p150Glued with dynactin p50 (dynamitin) in-turn binding to ZW10 (Figure 1.8). ZW10/RZZ has been demonstrated to be required for roles that dynein plays during prometaphase that are independent of dynactin, including initial microtubule capture (designated by ‘initial chromosome microtubule attachment’ in Figure 1.8), initial chromosome movement during congression toward the midplane, and efficient progression to metaphase (Yang et al. 2007; Bader et al. 2011).

As individual chromosomes become stably attached to microtubules and approach the metaphase plate during prometaphase, the dynein–dynactin complex (containing dephosphorylated IC; refer to ‘Regulatory Phosphorylation of the Intermediate Chain’ above for a more detailed discussion) transports SAC proteins and the RZZ complex away from kinetochores and toward the spindle poles (Figure 1.6), which helps to silence the mitotic checkpoint and allow progression to anaphase (Wojcik et al. 2001; Howell et al. 2001; Whyte et al. 2008); this process corresponds to the mass egress of dynein away from kinetochores that occurs prior to anaphase. In addition to the classic checkpoint proteins (such as Mad1, Mad2, BubR1, etc., originally identified in yeast), ZW10 and Rod have also been classified as mitotic checkpoint proteins in metazoan species (Chan et al. 2000; Basto et al. 2000) that are stripped from
**Figure 1.8.** A model in which multiple different populations of dynein coexist at kinetochores. In addition to the sequential mechanism of phosphodynein (containing IC phosphorylated at Threonine-89) recruitment to the kinetochore by ZW10/RZZ (followed by IC dephosphorylation and association with dynactin p150Glued, as depicted in Figure 1.6), results from other studies in the literature suggest that multiple populations of dynein may coexist at the kinetochore and are recruited to this location by distinct receptors, with the different dynein–regulator/adaptor complexes performing distinct functions (Stehman et. al. 2007; Whyte et. al. 2008; Bader et. al. 2011; McKenney et. al. 2011; Raaijmakers et. al. 2013). The direct association of phosphodynein with ZW10/RZZ (independent of dynactin) is believed to perform the function of initial microtubule capture at kinetochores and initial chromosome movement during congression to the midplane during prometaphase (Howell et. al. 2001; Yang et. al. 2007; Bader et. al. 2011). The complex of dephospho-dynein with dynactin is implicated in the removal of SAC proteins and ZW10/RZZ from microtubule-attached kinetochores and poleward streaming of these components prior to the commencement of anaphase (Wojcik et. al. 2001; Howell et. al. 2001; Whyte et. al. 2008). Though phosphodynein (through direct binding to ZW10) and dephosphodynein (through indirect association via binding interaction with dynactin p150Glued) are schematically shown to be simultaneously bound to the RZZ complex, it is not known whether this occurs in reality. The sequential model of dynein interaction with ZW10/RZZ at kinetochores (Figure 1.6) proposes that either phosphodynein is directly associated with ZW10 or, upon stable microtubule attachment to kinetochores, the dynein intermediate chain becomes dephosphorylated and dephospho-dynein then shifts its binding to dynactin. However, in the work of Whyte et. al., 2008, while depletion of dynactin did not affect the levels of phosphodynein recruitment to kinetochores during prometaphase, it did decrease the total dynein (phosphodynein + dephosphodynein), suggesting that some population of dynein-bound dephosphodynein may exist at the kinetochore, perhaps simultaneous to the population of phosphodynein that is bound directly to ZW10/RZZ (Whyte et. al. 2008). The kinetochore recruitment and association of dynein with NudE/EL-Lis1 (see also Figure 1.9) is proposed to function in generation of interkinetochore tension of microtubule-attached sister chromatids at the metaphase plate. Figure adapted with permission from Bader et. al., 2010 (Bader et. al. 2010).
Figure 1.8. (Continued)

- Initial Chromosome Microtubule Attachment
- SAC Silencing
- Kinetochore tension (?)

**Diagram Components:**
- ZW10/RZZ Complex
- Dynactin
- NudE/EL-Lis1
- Phospho-Dynein
- Dephospho-Dynein
- Dynein (Additional population)
microtubule-attached kinetochores by dynein in mitosis. Thus, ZW10/RZZ contributes to checkpoint activation by promoting SAC protein (Mad1–Mad2) recruitment to unattached kinetochores during prophase, and to checkpoint inactivation by recruiting dynein/dynactin that subsequently strips the Mad1–Mad2 complex (and RZZ itself) from microtubule-attached kinetochores throughout prometaphase.

In addition to binding interactions with p50 (dynamitin) of dynactin, phosphorylated IC of dynein, and the Rod and Zwilch proteins of the RZZ complex, ZW10 also binds directly (through its N-terminal region) and is recruited to the kinetochore by outer kinetochore protein Zwint-1 (Starr et. al. 2000; Wang et. al. 2004b; Kops et. al. 2005). Thus, it would seem that ZW10 has many different binding partners in the cell. In addition to its more pronounced and prominent role as part of the RZZ complex at the kinetochore in mitosis, ZW10 also appears to have an additional role in dynein targeting and regulation in interphase ER-to-Golgi trafficking and lysosome motility (Hirose et. al. 2004; Varma et. al. 2006). In this context, ZW10 does not exist as a part of the RZZ complex, but rather associates with two other proteins (NAG and RINT-1) in the interphase ‘NRZ’ complex (Civril et. al. 2010) that interacts with SNARE protein Syntaxin-18 (Hirose et. al. 2004) on the ER membrane outer surface. Molecular details of the specific arrangement of dynein with ZW10 in this context are not known, but it has been proposed that the NRZ complex docks at the membrane surface and its association with dynein is mediated via interaction of ZW10 with dynactin p50 (dynamitin), with dynein IC bound to dynactin p150Glued, and dynactin possibly acting via its Arp1 subunits (which bind to spectrin, see above) to co-anchor dynein to the membrane surface (Varma et. al. 2006); in this instance, dynein IC would likely be unphosphorylated in its ‘serine-rich’ region.

**Nuclear distribution protein E (NudE) and NudE-like (NudEL).**

Nuclear distribution protein E (‘NudE’ a.k.a. ‘Nde1’) and the highly homologous NudE-like (‘NudEL’ a.k.a. ‘Ndel1’) proteins (collectively referred to as ‘NudE/EL’) share ~55% sequence identity and are highly similar in both size and predicted secondary structure, each forming homodimers composed of a highly conserved parallel coiled-coil in the N-terminal approximate half of the protein, and a C-terminal half that is predominately unstructured (Figure 1.9a) (Efimov et. al. 2000; Derewenda et. al. 2007; Nyarko et. al. 2012). The unstructured C-terminal regions of NudE/EL bind to the centromere protein F (CENP-F) at kinetochores (Stehman et. al. 2007; Vergnolle et. al. 2007), and in the case of the NudEL form, the dynein
Figure 1.9. NudE/EL-Lis1 structure and association with the dynein motor complex. (A) Bar diagrams of mammalian NudE (similar to NudEL) and Lis1 domain structures, showing coiled-coil (‘CC’, blue), β-propeller (green), and unstructured domains, as well as relevant regions known to mediate interaction with various binding partners; the scale (below) indicates amino acid number. (B) Schematic depicting proposed modes of interaction in the tripartite complex of NudE/EL (blue and yellow), Lis1 (green and black), and dynein (grey); domain coloring is like that in Figure 1.9a. NudE/EL helps to recruit and enhance the binding of Lis1 to dynein (McKenney et. al. 2010). Lis1 interacts with the dynein motor domain when the latter is occupied by ADP and phosphate (P) (i.e. the ‘prepower stroke’ stage of the dynein crossbridge cycle); in this state, Lis1 binding enhances the affinity of dynein for microtubules (dark green), resulting in prolonged interactions with microtubules when dynein is under load. (C) Recent work indicates that NudE and dynactin p150Glued compete for overlapping sites within the dynein intermediate chain (McKenney et. al. 2011; Nyarko et. al. 2012), ensuring that individual dynein motor complexes cannot be simultaneously occupied by both regulators; in the case of NudE, the primary dynein binding site lies within the N-terminal region of the dynein IC subunit. Given its stimulatory effect on dynein processivity, dynactin may be required for fast, long-range transport of smaller vesicular and macromolecular cargoes, while association with NudE/EL-Lis1 transforms the dynein motor to a persistent force-producing state with enhanced multiple motor transport under heavier loads (McKenney et. al. 2010). Dynactin IC is identified as an important nexus for dynein regulation in this context, as phosphorylation of IC within its ‘serine-rich’ region is proposed to modulate selective binding of dynein to one regulator (NudE/EL) versus another (dynactin p150Glued), as described in Figure 1.5. Figure images adapted with permission: (A and B) from Vallee et. al., 2012 (Vallee et. al. 2012), and (C) from McKenney et. al., 2011 (McKenney et. al. 2011).
Figure 1.9. (Continued)

A

<table>
<thead>
<tr>
<th>NudE/EL</th>
<th>1</th>
<th>CC</th>
<th>LIS1</th>
<th>Unstructured domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dynein IC</td>
<td></td>
<td></td>
<td></td>
<td>345</td>
</tr>
<tr>
<td>LIS1</td>
<td>1</td>
<td>LH</td>
<td>CC</td>
<td>β-Propeller (7 WD repeats)</td>
</tr>
<tr>
<td>NudE/EL, Dynein</td>
<td></td>
<td></td>
<td></td>
<td>410</td>
</tr>
</tbody>
</table>

B

Microtubule

C

Cytoplasmic Dynein

(+ ) Dynactin
Fast, Low Load

Competitive

Dynein vs. NudE – Lis1

(+ ) NudE – Lis1
Slow, High Load
heavy chain (Sasaki et al. 2000; Liang et al. 2004), though there is some question as to whether this particular interaction occurs for the NudE form (McKenney et al. 2010). The parallel coiled-coil N-terminal domains of NudE/EL bind to the dynein intermediate chain (the N-terminal region of IC, discussed earlier) (Wang et al. 2011b; Zylkiewicz et al. 2011; McKenney et al. 2011; Nyarko et al. 2012), and to Lis1 (Efimov et al. 2000; Sasaki et al. 2000; Feng et al. 2000; Derewenda et al. 2007; Wang et al. 2011b).

NudE and NudEL were originally identified as binding partners of the protein Lis1 (Efimov et al. 2000; Sasaki et al. 2000). The Lis1 protein (a.k.a. ‘NudF’ in Aspergillus nidulans) is the product of the gene LIS-1—mutations of which are causal in the severe brain developmental disease lissencephaly (from Greek, meaning “smooth brain”), characterized by mislocalization of neurons in the cerebral cortex (due to defects in neuronal migration), producing profound mental retardation, seizures, and other neurological abnormalities (Dobyns et al. 1993).

Lis1 is a ~45 kDa protein with seven WD40 repeats at its C-terminus that form a β-propeller structure (Tarricone et al. 2004), while the N-terminal portion is largely α-helical, consisting of the Lis-homogoy (LH) motif that mediates dimerization of Lis1, followed by a parallel coiled-coil region (Figure 1.9a) (Kim et al. 2004). The C-terminal β-propellers mediate Lis1’s interactions with NudE/EL and with the first AAA domain of the dynein heavy chain (Sasaki et al. 2000; Derewenda et al. 2007). NudE/EL and Lis1 homodimers associate to form a heterotetramer (Tarricone et al. 2004) that associates with cytoplasmic dynein in a tripartite complex (Figure 1.9b) (Sasaki et al. 2000; Niethammer et al. 2000).

Among other functions (reviewed in Vallee et al. 2012), NudE/EL-Lis1 are implicated in regulation of dynein during neuronal migration and migration-coupled movements of nuclei and centrosomes (Shu et al. 2004; Tsai et al. 2007), and in mitosis (Faulkner et al. 2000; Liang et al. 2007; Raaijmakers et al. 2013). Experimental evidence demonstrates that NudE/EL greatly enhance the inherently weak binding affinity of Lis1 for dynein, such that NudE/EL essentially function in recruitment of Lis1 to the dynein molecule (McKenney et al. 2010; Wang et al. 2011b). Lis1 increases the effective strength of the dynein-microtubule interaction in the post-ATP-cleavage (‘prepower stroke’) state of the dynein motor domain, allowing dynein to remain microtubule-bound longer under load and function in sustained force production (McKenney et al. 2010); this may explain the requirement of Lis1 (and NudE/EL) for activities of cytoplasmic dynein that involve very high loads, including transport of cytoskeletal microtubules and nuclei, and transport of, or tension across, chromosomes (Tsai et al. 2007; Grabham et al. 2007) (Figure 1.9c).
Mammalian NudE/EL-Lis1 were found to be essential for all dynein functions in mitosis, with the tripartite complex localizing to several mitotic sites including the kinetochore (Raaijmakers et al. 2013). NudE/EL (recruited through their unstructured C-terminal regions that bind to CENP-F at the kinetochore, discussed above) arrive at the kinetochore during prophase before either dynein or Lis1, suggesting that NudE/EL may help to recruit dynein to the kinetochore and Lis1 to dynein (Stehman et al. 2007). Note that NudEL has been shown to interact with both the dynein motor domain and intermediate chain (though, the interaction with the intermediate chain has been shown to be necessary and sufficient for NudEL regulation of some dynein functions; Zylkiewicz et al. 2011), while NudE’s primary dynein binding site is at the N-terminus of IC (McKenney et al. 2011). Salient points from the work of Stehman et al., 2007, include: in mammalian (LLC-PK1, pig kidney) cells, antibody-interference of NudE/EL produced a decrease in the total amount of dynein present at kinetochores [also note, in HeLa cells, siRNA-depletion of NudE/EL resulted in a ~20% decrease in the total amount of dynein recruited to mitotic kinetochores (Raaijmakers et al. 2013)]; antibody-interference of NudE/EL reduced the degree of interkinetochore tension across microtubule-attached sister chromatids; and loss of NudE/EL activity prevented removal of checkpoint (SAC) proteins (including Mad2) even on microtubule-attached kinetochores that had congressed to the metaphase plate (Stehman et al. 2007).

Given these and other observations, a model has emerged in which NudE/EL helps to recruit Lis1 and dynein to the kinetochore, creating a kinetochore dynein sub-population (of the tripartite complex) (Figure 1.8) that functions in helping to create strong microtubule attachments for bioriented sister chromatids at the metaphase plate, with sufficient tension across the pair of chromatids—for which the sustained force could be suitably generated by the NudE/EL-Lis1/dynein tripartite complex (see above). It should be noted that, in the work of Whyte et al., 2008, both microtubule attachment and kinetochore tension were required for the dephosphorylation of ICpT89-dynein that switched IC binding from ZW10 to dynactin p150Glued, which then facilitated stripping of SAC proteins (including Mad2) from kinetochores by dynein/dynactin (Figure 1.6) (Whyte et al. 2008). The defect in removal of SAC proteins from kinetochores of microtubule-attached chromatids and metaphase arrest produced by NudE/EL depletion (Stehman et al. 2007) could potentially be explained (this is my inference) by the failure to recruit Lis1 and dynein to the kinetochore to form the tripartite NudE/EL-Lis1/dynein complex—which would normally function in generating strong microtubule attachments and interkinetochore tension—the absence of which interferes with the usual conversion of ICpT89-
dynein to its dephosphorylated form that would normally bind dynactin and participate in checkpoint inactivation and progression to anaphase. It is also worth noting that, given our current model for the interaction between dynein IC and NudE (Figure 1.4b, Figure 1.5), a dynein–NudE/EL complex at the kinetochore would be expected to be invariant to changes in IC phosphorylation/de-phosphorylation in its ‘serine-rich’ region that normally affect dynein association with other regulators at the kinetochore such as ZW10/RZZ and dynactin p150Glued.

Work Presented in this Dissertation

This dissertation contains four remaining chapters. In Chapter 2, I give a brief introduction and overview of the solution-state biophysical technique protein NMR spectroscopy (parts of which will be included in an NMR methods review article being prepared for submission to the journal *Intrinsically Disordered Proteins*) that is commonly applied to the characterization and study of IDPs, and is considered to be the most powerful tool available for studying this class of proteins. Given the substantial role that protein NMR spectroscopy played in the characterization of the dynein intermediate chain itself and in its interactions with regulatory proteins, particular emphasis in this chapter is placed upon this biophysical technique and special considerations that must be made when applying these NMR techniques to structural determination in IDPs (as opposed to folded, globular proteins) and in the handling and interpretation of data garnered from these experiments. In a utilitarian sense, information provided in this chapter helps to explain the interpretations of NMR data appearing in the original research work presented in Chapters 3 and 4.

Activities of the cytoplasmic dynein motor complex are indispensable for cell function and survival. A great deal of knowledge about the composition of dynein and the structures and assembly of its individual components and sub-complexes have been garnered through the work of multiple different labs (including ours) throughout the world in the 28 years since cytoplasmic dynein’s discovery was first reported. However many unanswered questions still remain regarding the control of its physical targeting and regulation of its functional behavior within the cell. To put it plainly, we know what ‘the beast’ looks like (X-ray crystal structures are available for nearly all constituent subunits—in-part or whole—in their apo and/or various bound states); and we have learned much about the process and mechanics of subunit assembly to form the functional motor complex, and cryo-electron microscopy images of sub-complexes or whole dynein motors have yielded low-resolution information to complement the higher-resolution X-ray protein crystal structures of the subunits; and we have even learned much about the manner
and control of dynein locomotion along microtubules, with proposed structure-based models for the mechanochemical coupling to ATP hydrolysis. But once assembled, how does dynein “know” where to go in the cell—or, how does the cell give dynein its ‘marching orders’?

Years of research have led to the discovery of protein complexes that can bind to and regulate the sub-cellular targeting and function of the dynein motor complex; a number of these complexes have been determined to possess subunits that engage in direct binding interaction with the intermediate chain (IC) protein of dynein, thus mediating the interaction between the regulatory and dynein complexes. But then, how is the binding association of dynein with these non-dynein regulatory complexes controlled?

Previously published work in the field had suggested (on the basis of results from more coarse-grained methods such as chromatography applied to different protein fragments) that the binding interactions between proteins of non-dynein regulatory complexes (such as dynactin p150\(^\text{Glued}\) and ZW10 of RZZ) and the dynein intermediate chain were mediated by the N-terminal approximate half of N–IC. To gain an understanding of the biophysical bases for these protein–protein interactions, as well as insight into how these interactions might be regulated, I first sought to characterize the structure and dynamics of a fragment of dynein IC (IC:1–143) that also contains the binding sites for dynein light chains Tctex1 and LC8. I performed solution-state protein NMR spectroscopy experiments to yield residue-level-resolution information demonstrating that, although predominately disordered along its length, IC:1–143 exhibits 2 regions of helical structure near the N-terminus. Furthermore, NMR titration of IC with p150\(^\text{Glued}\) evidenced binding interaction with 2 non-contiguous segments (named ‘Region 1’ and ‘Region 2’) of IC that overlapped with the regions of identified helical structure. Isothermal Titration Calorimetry (ITC) measurements confirmed that both regions of IC contribute energetically to the binding interaction with p150\(^\text{Glued}\). Furthermore, when IC:1–143 was bound with both p150\(^\text{Glued}\) and the light chains (an ∼150 kDa complex), the intervening residues between the p150\(^\text{Glued}\) and light chains binding sites remained disordered in the complex. The results of this study provided insights into the biophysical bases by which post-translational modification (phosphorylation of residues in a ‘serine-rich’ region) and alternative splicing in the primarily disordered IC could be used to modulate the interaction between dynein IC and dynactin p150\(^\text{Glued}\). This published work is presented as Chapter 3.

While the work presented in Chapter 3 began to shed light on some deviations of IC:1–143 from random coil behavior, and how the interplay of order and disorder in the IC–p150\(^\text{Glued}\) complex could function in regulation of dynein–dynactin interactions and cargo
recognition, a more detailed analysis of IC:1–143 revealed further deviations of the IDP from random statistical coil behavior. In collaboration with a research group in France (Dr. Martin Blackledge at the EMBL in Grenoble, France), experimentally measured NMR chemical shifts and a collection of residual dipolar coupling (RDC) values were combined with ensemble generation and selection algorithms to create a representative ensemble of the region of IC that binds light chains Tctex1 and LC8. These ensemble calculations evidenced deviation from random coil behavior with elevated populations of Polyproline II conformation being sampled by this region of IC, particularly in the case of the LC8 binding site in IC; in the light chains bound state, these same segments of IC adopt β-strand structure. Further deviation of IC:1–143 from random coil behavior is demonstrated by NMR paramagnetic relaxation enhancement (PRE) experiments, showing transient medium- and long-range contacts within apo IC itself, including interactions between the two non-contiguous helical regions that were previously identified to provide binding sites for dynein regulatory proteins dynactin p150Glued and NudE. NMR-PRE experiments on IC:1–143 in the NudE-bound state demonstrated that IC samples more extended conformations, with decreased (relative to the apo IC state) contacts made between the helical segments present in IC ‘Region 1’ and ‘Region 2’, suggesting a physical basis and mechanism for the observed phenomenon whereby dynactin p150Glued is able to displace NudE for binding to IC when both regulatory proteins are present in solution. This work is in preparation for publication submission and is presented as Chapter 4.

The results of these studies provide evidence that, while the flexibility and intrinsic protein disorder in N–IC underlie its ability to serve as a binding platform for numerous partners (some of which have overlapping binding sites on IC), deviations of this protein from random coil behavior (in the form of secondary structures of varying degrees) seem to occur in segments of the protein that engage in the binding interactions. N–IC exhibits further deviations from pure random coil behavior in the form of transient tertiary structure that, along with the retained disorder in assembled IC, provide insight into potential mechanisms for regulating these binding interactions and thus, the sub-cellular targeting and functionality of the dynein motor complex in the cell. In Chapter 5, I discuss the impacts of these findings, and outline areas for future research in terms of questions that remain to be addressed and that would expand upon insights gained from the research presented in this dissertation.
Chapter 2

The Use of Solution-State Protein NMR Spectroscopy in the Study and Characterization of Intrinsically Disordered Proteins

Jessica L. Morgan
In this chapter, I provide a brief review of the biophysical technique that has been at the core of my dissertation work in the characterization of the structure and dynamics of proteins—Nuclear Magnetic Resonance (NMR) spectroscopy. This technique allows proteins to be studied in the native solution state and provides time- and ensemble-averaged values for the physical properties measured. In the study of intrinsically disordered and unfolded proteins, solution-state NMR spectroscopy is regarded as the most powerful tool available due to the residue-level resolution it affords in both structural and dynamical analyses. Topics covered include chemical shifts, scalar couplings, residual dipolar couplings (RDCs), Nuclear Overhauser effects (NOEs), and Paramagnetic Relaxation Enhancements (PREs).

In the approximate decade-and-a-half since publication of the seminal report “Intrinsically Unstructured Proteins: Re-assessing the Protein Structure-Function Paradigm” by Dyson and Wright (Wright et. al. 1999)—which helped launch the new field of IDP research—rapid advances have been made in the application of NMR spectroscopy to the study of IDPs and unfolded proteins. Much of the theoretical background information detailed below for the different NMR observables can be broadly applied to proteins in general—folded or not. However, additional considerations (stemming largely from the unique conformational dynamics and motional freedom of these polypeptide chains) must be taken into account in the application these particular techniques and resultant data interpretation in the study of IDPs or unfolded proteins.

**Secondary Chemical Shifts**

Resonance (chemical shift) assignments (Figure 2.1) are the first step in analysis of protein structure and dynamics by NMR spectroscopy. NMR resonance assignments correlate frequency signals observed in the NMR spectra of a protein with specific amino acid residues and the nuclei (atoms) contained therein. The frequencies measured are generally normalized relative to a standard compound (such as DSS, TMS, etc.) and then converted into units of parts per million (e.g. the difference between the frequencies for the signal of interest and the signal of the standard reference compound (in units of Hz), divided by the Larmor frequency of the particular type of nucleus being observed) to yield the chemical shift. Nuclei whose chemical shift values are often assigned in the case of a protein include: $^1$H$^N$ (the amide proton in the amino acid residue), $^1$H$^\alpha$ (the alpha proton), $^{15}$N (the amide nitrogen), $^{13}$C$^\alpha$ (the alpha carbon atom), $^{13}$C$^\beta$ (the beta carbon), and $^{13}$C$^\gamma$ (the carbonyl carbon, a.k.a. $^{13}$C$'$). Since the advent of strategies for production of proteins uniformly labeled with stable isotopes such as $^{13}$C and $^{15}$N, as well as the
Figure 2.1. Example of NMR chemical shift assignments for a protein—*Drosophila melanogaster* dynein LC7. (A) $^{13}$C (C$_{\alpha}$, C$_{\beta}$) strip plots for LC7 residues Gln42 – Gln49 taken from the HNCACB experiment collected, with the strip for each residue centered at its amide ($^{1}$H$_{N}$ and $^{15}$N$_{H}$) frequencies. This is an example of the type of data yielded from one particular 3–D triple-resonance NMR experiment. (B) The HNCACB experiment correlates the amide proton and nitrogen of a residue (e.g. residue ‘$i$’ with the C$_{\alpha}$ (red contours) and C$_{\beta}$ (blue contours) of the same residue, and the C$_{\alpha}$ and C$_{\beta}$ of the preceding amino acid residue (e.g. residue ‘$i$ – 1’) in the protein chain; the scalar couplings that underlie these correlations are indicated with blue curved arrows. In the course of performing backbone chemical shift assignments for a protein, the data from an HNCACB experiment is often paired with that from another triple-resonance experiment that contains overlapping information, such as a CBCA(CO)NH experiment which correlates the amide proton and nitrogen of a residue (e.g. residue ‘$i$’) with the C$_{\alpha}$ and C$_{\beta}$ of the preceding amino acid residue (e.g. residue ‘$i$ – 1’), allowing unambiguous assignment of resonances to particular amino acid residues. A $^{1}$H–$^{15}$N Heteronuclear Single Quantum Coherence (HSQC) NMR spectrum shows the correlations between the amide proton and amide nitrogen atoms (highlighted with orange circles) in each amino acid residue (except for Proline, which has no amide proton). Each peak in an HSQC corresponds to one amino acid residue in the protein chain, and the HSQC spectrum is essentially like a “fingerprint” for each protein. (C) $^{1}$H–$^{15}$N HSQC spectrum of *Drosophila melanogaster* LC7 in 12.5 mM Tris (pH 7.5) at 30 °C collected on a 600 MHz Bruker DRX spectrometer; each peak in the spectrum is labeled with the amino acid residue to which it corresponds. Panels ‘A’ and ‘C’ are adapted from Appendix figures A4.1 and A4.2, respectively.
Figure 2.1. (Continued)
development of multi-dimensional $^{13}\text{C}/^{15}\text{N}$-resolved triple-resonance NMR experiments that capitalized on this uniform isotopic labeling (in the early 1990s), backbone chemical shifts have routinely been assigned for proteins using a collection of multi-dimensional triple-resonance NMR experiments, including: HNCACB (Wittekind et al. 1993), CBCA(CO)NH (Grzesiek et al. 1992), (H)CC(CO)NH (Grzesiek et al. 1993), HNCO and HNCA (Kay et al. 1990), etc.

The chemical shift values measured for each nucleus in a protein are exquisitely sensitive to the local physicochemical environment (including the presence of secondary structure) in which the nucleus resides (Wishart et al. 1994a). Early researchers in the field noted strong correlations between $^1\text{H}$, $^{13}\text{C}$, and $^{15}\text{N}$ chemical shifts and protein secondary structure (Wishart et al. 1991). More specifically, the NMR secondary chemical shifts (deviation of experimental chemical shifts from their expected random coil values, $\Delta\delta = [\delta_{\text{experimental}} - \delta_{\text{random coil}}]$) of certain nuclei are known to be highly correlated with $\phi$ and $\psi$ backbone dihedral angles and thus with secondary structure in proteins (Spera et al. 1991). For instance, protein residues in $\alpha$-helical conformations exhibit small negative secondary chemical shifts (SCSs) for $^1\text{H}^{\alpha}$ and $^1\text{H}^{\text{N}}$ nuclei (averaged values of -0.38 and -0.19, respectively), while these same nuclei exhibit small positive SCS values (averaged values of +0.38 and +0.29, respectively) in the context of $\beta$-strand structure (Wishart et al. 1994a). No significant deviations from zero are exhibited in the secondary chemical shift values for $^1\text{H}^{\alpha}$ and $^1\text{H}^{\text{N}}$ nuclei in Polyproline II structure (Lam et al. 2003).

Positive deviations in the secondary chemical shifts of $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}^{\text{O}}$ nuclei are correlated with $\alpha$-helical conformations (averaged values of +2.6 and +1.7 ppm, respectively), while negative deviations (averaged values of -1.4 ppm for each nucleus type) are associated with extended $\beta$-strand-like structure (Wishart et al. 1994a). For $^{13}\text{C}^{\beta}$ nuclei an opposing trend is observed, wherein negative secondary chemical shifts are exhibited by residues in $\alpha$-helical structure (averaged value of -0.38 ppm) and positive deviations (averaged value of +2.16 ppm) are seen for residues with $\beta$-strand structure (Spera et al. 1991). Polyproline II structure is correlated with much smaller magnitude (than for $\beta$-strand) negative secondary chemical shifts in $^{13}\text{C}^{\alpha}$ and $^{13}\text{C}^{\text{O}}$ nuclei (Wishart et al. 1998; Mukrasch et al. 2009). Negative SCS values are exhibited for $^{15}\text{N}$ nuclei in $\alpha$-helical structure (averaged value of -1.7 ppm), while positive values are observed for residues in $\beta$-strand conformations (averaged value of +1.2 ppm; Wishart et al. 1994a) and slightly smaller (than for $\beta$-strand) positive values are seen for Polyproline II
While the averaged (over all amino acid residue types) secondary chemical shift values cited above are for fully-formed secondary structural elements, many IDPs and unfolded proteins exhibit only fractional or transient secondary structure with SCS values that are much smaller in magnitude and closer to zero (e.g. closer to random coil values). These IDPs that lack regions of stable secondary structure exist as rapidly interconverting ensembles of numerous conformers; for inter-conversion rates that are fast on the NMR timescale [e.g. $\geq 10^4$ s$^{-1}$ (Cavanagh et al. 1996)]—as evidenced by single peaks for individual residues in a $^1$H–$^{15}$N HSQC spectrum—the measured chemical shift of a given nucleus represents a population-weighted average of the conformations that compose the ensemble. For a system in which a given protein region is presumed to exist as a mixture of random coil and fully-formed secondary structure (either $\alpha$-helix or $\beta$-strand), quantitative estimates of the secondary structure content (percentages) have sometimes been made by normalizing the observed secondary chemical shift of a nucleus to the empirically determined SCS value expected for the given residue type in regular, fully-formed secondary structure (Weinstock et al. 2008). Alternatively, by combining multiple types of $^{13}$C (Wishart et al. 1994b) or multiple types of $^1$H, $^{13}$C, and $^{15}$N SCS values for a given residue, with weights assigned to the individual types of chemical shifts based upon their reliability for detecting different types of secondary structure (e.g. $\alpha$-helix or $\beta$-strand), a consensus estimate of the secondary structure content or propensities can be calculated (Wang et al. 2002a; Marsh et al. 2006).

Regardless of whether secondary chemical shift data are to be used quantitatively or qualitatively to assess secondary structure content and propensities in proteins, the choice of random coil chemical shift values (e.g. the ‘reference state’ chosen to represent the random coil) is of crucial importance, particularly in the case of IDPs where secondary structures can often be fractional or transient, producing secondary chemical shift values that are of only small magnitude. A prime example of the dependence of calculated SCS values on choice of random coil chemical shift scale was recently highlighted (Tamiola et al. 2010) for the case of the K19 construct of the intrinsically disordered human Tau protein, in which the use of one commonly-employed set of random coil chemical shift values indicated significant $\alpha$-helical propensity throughout the protein, while use of a different commonly-used set of random coil chemical shift values produced completely different SCS values that suggested a predominance of weakly-populated $\beta$-sheet-like regions. Depending on the type of nucleus and amino acid residue type, differences between random coil chemical shift values from different scales can vary.
considerably, even by as much as several ppm. To reiterate, NMR chemical shifts are highly sensitive to local environment, and this includes numerous factors aside from the presence of secondary or tertiary structure. Several influential factors on the chemical shift measured for a particular nucleus in a given type of amino acid residue include: temperature, pH, the presence of denaturants or organic solvents, and the identity of neighboring residues in the primary sequence, etc., and different combinations of values for these factors influence the set of tabulated random coil chemical shift values in a given scale.

Random coil chemical shift scales are grouped into two categories: (1) those derived from statistical analyses of ‘coil libraries’ culled from chemical shifts in an NMR databank (“statistical analysis/database” approach), and (2) those measured from libraries of small model peptides (“model peptide library” approach).

In statistical analysis/database approaches, proteins of known structure (either X-ray crystallography or NMR-derived structures) whose chemical shifts have been compiled or otherwise collected in a databank (such as the Biological Magnetic Resonance Data Bank (BMRB), http://www.bmrb.wisc.edu; Ulrich et. al. 2008), are parsed according to secondary structure (either α-helix or β-strand), with all other residues outside of these two designations being lumped together under the category named ‘coil’ which is taken to represent a generic reference state for the random coil behavior of amino acid residues; average chemical shift values for each type of nucleus (e.g. $^1$H$^N$, $^{13}$C$^\alpha$, $^{15}$N, etc.) in each amino acid residue type have been compiled for each of the three categories: ‘α-helix’, ‘β-strand’, and ‘coil’ (Wishart et. al. 1991; Wang et. al. 2002b; Zhang et. al. 2003). One obvious drawback to this approach is that, the classification of protein structure by three categories (helix, sheet, and “other”) may not be a suitable simplifying assumption; there several types of protein structure (Polyproline II and various types of turn motifs, for example) that are not α-helix or β-strand, but nor are they appropriately classified as ‘random coil’ either. More recently, this type of approach has been refined and updated by exclusion of fragments classified by STRIDE (Heinig et. al. 2004) as $\pi$- or $3_{10}$-helix, and turns or bends, from the ‘coil’ category (De Simone et. al. 2009).

One major shortcoming of statistical analysis/database approaches is that they take averages over a wide array of datasets collected for proteins under widely varying temperature and solvent experimental NMR conditions. It should be noted however, that statistical analysis/database approaches do have the benefit (over model peptide library approaches, *vide infra*) of a large database size with tens or even hundreds of thousands of chemical shift entries.
for the different nucleus types ($^{1}H^{N}, ^{13}C^{\alpha}, ^{15}N$, etc.), over which good statistical analyses can be conducted due to the large sample size.

In model peptide library approaches, small peptides of the form: Ac-GG-X-GG-NH$_2$ (Plaxco et. al. 1997; Schwarzinger et. al. 2000; Kjaergaard et. al. 2011a), Ac-QQ-X-QQ-NH$_2$ (Kjaergaard et. al. 2011b), Ac-AA-X-AA-NH$_2$ (Prestegard et. al. 2013), or Ac-GG-X-(A/P)-GG-NH$_2$ (Wishart et. al. 1995; Lam et. al. 2003), etc., have been used to model and represent the random coil behavior of individual amino acids, where ‘X’ = any of the 20 proteogenic amino acids. One key advantage of the model peptide library approach (over statistical analysis/database approaches, vide supra) is the ability to specify and control the exact experimental conditions under which chemical shift values are measured. Chemical shifts for the central and surrounding residues in these model peptides have been measured at various temperatures and pH values, with different solution compositions including nondenaturing (Kjaergaard et. al. 2011b), mildly denaturing (Wishart et. al. 1995; Lam et. al. 2003; Kjaergaard et. al. 2011a), and denaturing (Schwarzinger et. al. 2000, 2001; Prestegard et. al. 2013) conditions.

While the effects of temperature on $^{13}C$ random coil chemical shifts are thought to be small and relatively negligible, the effect of temperature on $^{1}H^{N}$ and $^{15}N$ chemical shifts is considered to be more significant and for this reason, temperature coefficients have been calculated and reported for these nuclei (and others) in the case of certain model peptide library-derived chemical shift scales (Merutka et. al. 1995; Lam et. al. 2003; Kjaergaard et. al. 2011a); through interpolation, these temperature coefficients allow one to calculate correction factors and thus random coil chemical shift values appropriate for a particular temperature of interest. Though many IDPs are known to undergo changes in structure with changing temperature (they usually have higher secondary structural content at lower temperature), the temperature-dependent changes exhibited by their observed chemical shifts stem from changes in both the actual structural content of the protein, as well as from the intrinsic temperature-dependence of random coil chemical shifts of the nuclei examined. By subtracting the inherent temperature-dependence of the random coil chemical shifts from the exhibited changes that occur with temperature, one can more accurately detect transient or lowly populated states of secondary structure in the protein and monitor how these change with temperature.

In addition to temperature, pH is another important consideration, particularly for residues with titratable side chain groups such as aspartate (pKa = 3.9), glutamate (pKa = 4.1),
and histidine (pKa = 6.1); recall from Chapter 1, that IDPs tend to be enriched (relative to folded proteins) in certain amino acid types, including aspartate and glutamate. For random coil values determined at low pH (as in the experimental conditions: 8M urea, 20°C, pH 2.3 of one popular random coil chemical shift scale; Schwarzinger et al. 2000), the side chains of these three residues are expected to be fully protonated. At pH 5 (as in experimental conditions: 1.0 M urea, 25°C, pH 5.0 of another random coil chemical shift scale; Wishart et al. 1995) the aspartate and glutamate side chains would be partially (~10%) protonated and fully deprotonated at pH 6.5 (as in experimental conditions: 1.0 M urea, 5°C, pH 6.5 of another random coil chemical shift scale; Kjaergaard et al. 2011a). An examination of the pH dependence of NMR chemical shifts of aspartate, glutamate, and histidine residues revealed striking changes in the chemical shifts of the $^1$H, $^{15}$N, and $^{13}$C nuclei in these residues, with the $^{13}$C and $^{15}$N chemical shifts becoming larger with increasing pH (Kjaergaard et al. 2011a,b); over the pH-course of titration, some $^{13}$C chemical shifts were observed to change by as much as several ppm. Not surprisingly, the $^{13}$C and $^{15}$N random coil chemical shift values determined for these residues at pH 6.5 (Kjaergaard et al. 2011a) were seen to be slightly higher than those reported for pH 5.0 (Wishart et al. 1995) and significantly higher than those reported for pH 2.3 (Schwarzinger et al. 2000, 2001). While there is potential utility in having available random coil chemical shift values determined at extremes of pH, for protein NMR spectroscopy, experimental solution conditions are generally kept slightly below neutral pH in order to minimize proton exchange with the aqueous solvent. With the passage of time and repeated, widespread usage, certain random coil chemical shift scales have come to be regarded as the set of “standard” random coil values to be used; this practice has likely been facilitated, for instance, by incorporation of a particular random coil dataset (Schwarzinger et al. 2000, 2001) into modules for secondary chemical shift calculation within free and widely-used NMR data analysis software packages such as NMRView (Johnson et al. 1994). Convenience aside, it is imperative—particularly when trying to detect transient secondary structure in IDPs or unfolded proteins via secondary chemical shift calculations—to consider the experimental conditions under which the subtracted random coil chemical shift values were determined.

In addition to temperature and pH, the presence of denaturants is another important factor to consider. Even small peptides (such as the ones used in model peptide library approaches) have been known to adopt defined structures in aqueous solution (Dyson et al. 1991; Shi et al. 2002, 2005). Additionally, in host peptides where the buttressing residues contain side chains capable of hydrogen-bonding (such as glutamine in Ac-QQ-X-QQ-NH$_2$; Kjaergaard et al. 2011a)
intra-chain hydrogen-bonding interactions between the host (e.g. glutamine) and guest (“X”) residues could potentially alter measured chemical shift values; intra-chain hydrogen-bonding interactions could potentially occur with the peptide terminal groups as well. In an effort to eliminate traces of residual structure or conformational bias (as well as to prevent peptide aggregation) and to achieve an ostensibly better representation of random coil behavior, most model peptide library-derived random coil chemical shift scales include some amount of denaturant in the experimental solution conditions, usually either 1.0 M (Wishart et. al. 1995; Lam et. al. 2003; Kjaergaard et. al. 2011a) or a more significant amount such as 8.0 M urea (Schwarzinger et. al. 2000, 2001; Prestegard et. al. 2013). It should be noted, however, that the presence of denaturants can cause perturbations in observed chemical shifts that are independent of any structural or conformational changes in proteins and peptides due to the denaturant. For instance, in one study examining the effects of the denaturant guanidine hydrochloride (GuHCl) on model peptides of the form Ac-GG-X-GG-NH₂, substantial changes in chemical shifts of ¹H, ¹³C, and ¹⁵N nuclei were observed to occur in linear fashion with increasing GuHCl concentrations, despite the fact that other NMR data (such as ³J_HNΗα scalar couplings) showed no dependence on denaturant concentration, indicating that GuHCl was not perturbing the intrinsic conformational preferences of the peptides (Plaxco et. al. 1997). In the same study, the chemical shifts of both water and even GuHCl itself were seen to be substantially altered in linear fashion with GuHCl concentration. As a side note, correction factors have been calculated for all nuclei of the 20 proteogenic ‘guest’ amino acid residues, to allow (through interpolation) prediction of random coil chemical shifts at any arbitrary GuHCl concentration (Plaxco et. al. 1997).

When using model peptide library-derived random coil chemical shift values, particularly if quantitative assessments are to be made about secondary structure content, the recommendation is to select a set of values that were collected under conditions that match your experimental conditions as closely as possible. There is no single best ‘gold standard’ set of random coil chemical shift values that will cover all experiments. While some experimental conditions that have been used in random coil chemical shift scales might seem rather extreme (such as 8.0 M urea, pH 2.3; Schwarzinger et. al. 2000) and inappropriate to the less- or non-denaturing and less-acidic conditions used in the bulk of modern protein NMR studies, they may be highly applicable and appropriate to another set of experimental circumstances. For instance, the aforementioned set of conditions and associated random coil chemical shift values (Schwarzinger et. al. 2000) were likely developed to serve as a starting-point reference state for NMR studies of folding-initiation processes in denatured model proteins, such as apomyoglobin, which exists in an acid-
denatured (unfolded) state at pH 2.3 (Yao et al. 2001).

Lastly, the Ramachandran-space ($\phi, \psi$) conformational distribution of an amino acid residue is known to be affected by residues adjacent to it in the primary sequence (Ting et al. 2010) and thus, it is not surprising that random coil chemical shifts exhibit nearest-neighbor effects. For instance, $^1$H$^N$ and $^{15}$N amide chemical shifts are correlated with both $\phi_i$ and $\psi_{i-1}$ (that is, $\psi$ of the preceding residue in the primary sequence) main-chain torsion angles (Le et al. 1994; Wishart et al. 1998) and $^1$H$^N$ and $^{15}$N amide random coil chemical shifts for a given residue vary widely, depending upon the identity of the preceding residue; $^{15}$N amide chemical shifts can vary by several ppm in this regard (Braun et al. 1994; Kjaergaard et al. 2011a). In contrast, the random coil chemical shifts of other nuclei, such as $^{13}$C', are affected more strongly by the identity of the proceeding (e.g. ‘$i + 1$’) residue in the primary sequence (Wishart et al. 1998). In light of nearest-neighbor effects on random coil chemical shifts, the creators of several random coil chemical shift scales have measured and incorporated sequence correction factors for $^1$H, $^{13}$C, and $^{15}$N nuclei into their scales, both statistical analysis/database (Wang et al. 2002b; De Simone et al. 2009; Tamiola et al. 2010) and model peptide library (Braun et al. 1994; Wishart et al. 1995; Schwarzinger et al. 2001; Kjaergaard et al. 2011a,b; Prestegard et al. 2013) approaches; these sequence correction factors include both nearest-neighbor residues in the primary sequence (e.g. ‘$i - 1$’ and ‘$i + 1$’), and in some cases, next-nearest-neighbor residues (e.g. ‘$i - 2$’ and ‘$i + 2$’) as well.

In the case of model peptide library approaches, one of the earlier studies (Wishart et al. 1995) addressed the issue of sequence-dependence of random coil chemical shifts through a limited but direct approach, using model peptides of the form: Ac-GG-X-Y-GG-NH$_2$, where ‘Y’ = Ala or Pro and ‘X’ = any of the 20 proteogenic amino acids. Thus, of the 400 possible X-Y dipeptide combinations, 40 were examined directly. In subsequent studies involving Ac-GG-X-GG-NH$_2$ model peptides, simplifying assumptions were made such that sequence correction factors were calculated based upon the effect that the central residue ‘X’ had on neighboring Glycine residues in the peptide (Schwarzinger et al. 2001; Kjaergaard et al. 2011a,b). However, the assumption that all other residue types would be affected by the guest residue ‘X’ in the same way that Glycine was affected, has been called into question; Glycine (with its single hydrogen atom for a side-chain) is known to be “conformationally abnormal” and to be an outlier (relative to other amino acids) in terms of its distribution in ($\phi, \psi$) Ramachandran space. Indeed, Glycine-based sequence correction factors were found to differ significantly from those determined by the same approach with Ac-AA-X-AA-NH$_2$ (Prestegard et al. 2013) and Ac-QQ-X-QQ-NH$_2$.
(Kjaergaard et al. 2011b) model peptide systems. The use of Alanine was considered to be an improvement upon the Glycine system, since Alanine (with its single methyl group side-chain) is the next-highest amino acid in terms of its mass and complexity. For the approach using the Ac-QQ-X-QQ-NH₂ model peptide system, Glutamine was chosen as the ‘host’ residue type based upon a study of the neighbor-dependence of Ramachandran distributions (Ting et al. 2010), where Glutamine was found to be the most representative residue type (Kjaergaard et al. 2011b).

In the model peptide library approach, it would be ideal to determine sequence-dependent corrections to random coil chemical shifts directly from the exact sequence of interest; however, for a relatively simple system in which only nearest-neighbors (e.g. ‘i – 1’ and ‘i + 1’) are taken into consideration, a tripeptide block of ‘X’-‘Y’-‘Z’ within some sort of host peptide system would amount to \((20)^3 = 8,000\) different possible combinations, which would be prohibitive and intractable experimentally. In this particular aspect, statistical analysis/database approaches may have the advantage (over model peptide library approaches) of containing a greater diversity of sequence combinations. Plus, with time, the number of depositions in the relevant databanks will only increase, providing an even greater diversity of sequence combinations.

**Scalar Couplings**

Two types of coupling interactions between spins (nuclei) are known: direct dipolar coupling (see section below, ‘Residual Dipolar Couplings’) and scalar coupling (also known as ‘J-coupling’). Nuclear spin-spin scalar coupling is mediated and relayed via the electrons surrounding the nuclei (Ramsey et al. 1952); thus, scalar coupling occurs ‘through-bonds’ in contrast to the ‘through-space’ interactions in direct dipolar coupling, and the magnitude of the scalar coupling (Figure 2.2a) significantly decreases as the number of intervening bonds increases (the scalar coupling constant is usually given the symbol \(^n J\), where \(n\) is the number of intervening bonds). To the non-practitioner of protein NMR spectroscopy, this type of scalar coupling is probably most familiar in the context of the peak-splittings seen in 1-D (proton) NMR spectra of small model compounds, as might be encountered in an undergraduate Organic Chemistry course. Much like chemical shifts (described above), scalar couplings are sensitive to the conformational properties (including bond and torsional/dihedral angles) of the molecule in which they reside and can thus be used to determine structural information.

In terms of structural characterization, the three-bond scalar (‘vicinal’) couplings are
**Figure 2.2.** NMR scalar coupling constants and secondary structure in proteins.  (A) Examples of typical NMR scalar coupling values (in Hz) measured for various pairs of nuclei (Bystrov 1976). One-bond direct couplings are shown in *black*, two-bond geminal couplings are depicted in *blue*, and three-bond vicinal couplings are in *pink*. Note that it is the sequential correlations via one- and two-bond scalar couplings that are used to establish connectivities between amino acids in the multi-dimensional triple-resonance experiments used for chemical shift assignments. (B) A peptide fragment (L-amino acid residues) depicting the $\phi$ and $\psi$ backbone dihedral angles. At *right*, Newman projections depicting the $\phi$ and $\psi$ peptide backbone torsional angles. (C) A plot of the empirical Karplus curve (*red*) for $^3J_{HNH\alpha}$, calculated using the parameters of Pardi et. al. (Pardi *et. al.* 1984). Various types of secondary structure are indicated on the plot according to their idealized, regular $\phi$ values: $\beta$-sheet, anti-parallel ($\phi = -139^\circ$, *cyan line*); $\beta$-sheet, parallel ($\phi = -119^\circ$, *blue line*); $\beta_1$ left-handed helix (*e.g.* Polyproline II (‘PPII’), $\phi = -75^\circ$, *green line*); and right-handed alpha helix (‘$\alpha_R$', $\phi = -57^\circ$, *pink line*).
Figure 2.2. (Continued)
probably the most familiar, being modeled and calculated from torsion angles \( \theta \) using the Karplus relationship (Karplus 1963; Bystrov 1976):

\[
3J = A \cdot \cos^2(\theta) + B \cdot \cos(\theta) + C \quad \text{[eq. 1]}
\]

In the analysis of conformations in peptides and proteins, several different types of homo- and heteronuclear three-bond scalar couplings have been measured. Among those which are sensitive to and report on the peptide backbone dihedral angle \( \psi \) (Figure 2.2b) are \( ^3J_{HH(\alpha)} \) (Wang et al. 1995; Permi et al. 2000; Löhr et al. 2001), \( ^3J_{NN} \) (Löhr et al. 1998), and \( ^3J_{CP(\alpha)} \) (Seip et al. 1994) scalar couplings, for which Karplus equation [eq. 1] parameters (e.g. A, B, C, and the form of ‘\( \theta \’) have been derived. However, these \( ^3J \) scalar couplings that report on the angle \( \psi \) are generally smaller in magnitude (and considered to be less reliable) than those reporting on the \( \phi \) angle. Some of the three-bond scalar couplings that report on the peptide backbone dihedral angle \( \phi \) include: \( ^3J_{HN(\alpha)} \) (Wang et al. 1995), \( ^3J_{IN\alpha} \) (Pardi et al. 1984), \( ^3J_{CC} \) (Hu et al. 1996), \( ^3J_{CP(\alpha)} \) (Hu et al. 1997), \( ^3J_{IN\beta} \) and \( ^3J_{HC(\alpha)} \) (Bystrov 1976). Parameterizations of the Karplus relation [eq. 1] for the different types of \( ^3J \) scalar couplings have typically been performed through correlations of experimentally measured \( ^3J \) values with known dihedral angles measured from structures of folded proteins. One of the most frequently measured and reported vicinal scalar couplings in protein structure determination, \( ^3J_{IN\alpha} \) (coupling between the amide proton, \( ^1H_N \), and the alpha proton, \( ^1H_\alpha \), within a single amino acid residue), has received much attention in terms of parameterization and refinement of its Karplus equation (Bystrov et al. 1973; Pardi et al. 1984; Ludvigsen et al. 1991; Vuister et al. 1993a; Wang et al. 1996; Schmidt et al. 1999; Case et al. 2000; Markwick et al. 2009):

\[
^3J_{IN\alpha} = A \cdot \cos^2(\phi - 60^\circ) + B \cdot \cos(\phi - 60^\circ) + C \quad \text{[eq. 2]}
\]

Presently, one of the most widely-employed sets of empirically-derived parameters for [eq. 2] are: A = 6.4, B = -1.4, C = 1.9, and ‘\( \theta \’) in [eq.1] is equal to (\( \phi - 60^\circ \)) (Pardi et al. 1984), producing the characteristic Karplus curve shown in Figure 2.2c. Values of \( ^3J_{IN\alpha} \) are generally in the range of 4–5 Hz (< 6 Hz) for a residue in an \( \alpha \)-helix (\( \alpha_\beta \)) and > 8 Hz for a residue in a \( \beta \)-strand (Smith et al. 1996); residues in left-handed \( 3_1 \) helix (e.g. Polyproline II) structure generally have values in the range of ~6–7 Hz (Lam et al. 2003).
Like NMR chemical shifts (*vide supra*), measured NMR scalar coupling constants also represent population-weighted ensemble averages over numerous rapidly interconverting conformers in a protein sample. And in analogy with secondary chemical shifts, secondary scalar couplings can be calculated by subtracting residue-specific random coil values from measured values \[ \Delta^{3}J_{\text{HNH}} = {}^{3}J_{\text{HNH}}(\text{experimental}) - {}^{3}J_{\text{HNH}}(\text{random coil}) \], to identify propensities for local secondary structure. It is important to use residue-specific random coil values, as they do vary significantly among amino acid residue types: ~6 Hz for residues like Glycine and Alanine vs. values of ~7.7 Hz for such residues as Valine and Threonine, for instance in one set of random coil values (Smith *et al.* 1996). Positive deviations of \(^{3}J_{\text{HNH}}\) values from residue-specific random coil values (*e.g.* positive \(\Delta^{3}J_{\text{HNH}}\)) indicate a propensity toward extended β-strand structure, with an average deviation of ~+1.5 Hz corresponding to full β-structure for most residue types, while negative values of \(\Delta^{3}J_{\text{HNH}}\) are indicative of turn or helical propensities, with an average deviation of ~−2.0 Hz corresponding to full α-helical structure for most residue types (Smith *et al.* 1996). Negative values of \(\Delta^{3}J_{\text{HNH}}\) are also consistent with polyproline II structure, and the three-bond coupling values for this type of structure are generally ~0.6 Hz lower than those for random coil (Lam *et al.* 2003).

As was the case for secondary chemical shifts, sets of random coil values for \(^{3}J_{\text{HNH}}\) scalar couplings come in two basic varieties: (1) values predicted using a parameterized Karplus relationship, applied to known \(\phi\) angle values in a ‘coil’ library culled from protein structures in the Protein Data Bank (PDB) (Berman *et al.* 2000) [“prediction/database” approach], and (2) values measured directly from libraries of small model peptides [“model peptide” approach].

In prediction/database approaches, a ‘coil’ library (similar to that described above for the “statistical analysis/database” approach in the ‘Secondary Chemical Shifts’ section) is constructed from \(\phi,\psi\) angles seen in regions devoid of secondary structure in protein crystal structures. \(^{3}J_{\text{HNH}}\) values are predicted (calculated) from X-ray structure-derived \(\phi\) values using a parameterized Karplus relationship [eq. 2]; the calculated \(^{3}J_{\text{HNH}}\) values are then averaged for each residue type. A set of residue-specific \(^{3}J_{\text{HNH}}\) random coil values for 19 amino acid types (Proline is excluded) has been compiled and reported (Smith *et al.* 1996), with values derived from the structures of 85 proteins, using the Karplus parameters of Pardi et. al. (Pardi *et al.* 1984). However, as mentioned above, several different parameterizations of the Karplus relation for \(^{3}J_{\text{HNH}}\) have been performed, in several cases on the basis of a single protein (or maybe 2 proteins).
for which there exist both a crystal structure and measured $^3J_{\text{HNH}_\alpha}$ scalar coupling constants, such as: basic pancreatic trypsin inhibitor (BPTI) at 36°C and pH 3.5 (Pardi et. al. 1984); barley serine proteinase inhibitor (CI-2) and bacterial ribonuclease (Ludvigsen et. al. 1991); staphylococcal nuclease at 35°C and pH 7.0 (Vuister et. al. 1993a); human ubiquitin at 30°C and pH 4.7 (Wang et. al. 1996); D. vulgaris flavodoxin at 27°C and pH 7.0 (Schmidt et. al. 1999). These various sets of Karplus equation parameters do differ from one another (usually more so for $\phi > -60^\circ$).

In a prediction/database approach, different Karplus parameters could yield different predicted $^3J_{\text{HNH}_\alpha}$ values for the same angle $\phi$, and thus yield different residue-specific random coil values.

Concerning the different parameterizations of the Karplus equation for $^3J_{\text{HNH}_\alpha}$, one possible explanation for the diversity of parameter values obtained, could be that there are simply fundamental differences between the different proteins used in each study (e.g. perhaps one protein in one study has a preponderance of residues in the $\alpha$-helical region of Ramachandran space, while another may not, and this would tend to skew the distribution of data points available for the parameter fitting process, in the case of the former). Alternatively, perhaps the experimental conditions (such as temperature and pH) in which the $^3J_{\text{HNH}_\alpha}$ coupling constants for the model protein were collected are contributing to the diversity. The protonation state of a side chain group (sensitive to pH) could conceivably alter the conformational sampling of backbone dihedral angles in a given residue. Temperature is known to affect the values of measured scalar coupling constants (just as temperature can affect the structural content of a protein); for the example of BPTI (above), Pardi et. al. measured $^3J_{\text{HNH}_\alpha}$ coupling constants for their model protein at both 36°C and 68°C, with some residues showing a difference of up to 1.0 Hz between the two temperatures (Pardi et. al. 1984). Incidentally, they selected, “...somewhat arbitrarily...”, to use the data collected at 36°C in their Karplus parameterization (Pardi et. al. 1984). With the exception of proteins derived from (hyper)thermophilic organisms (and also excepting those rare proteins known to undergo cold-denaturation), proteins in the solution state will generally resemble their crystal structures more closely (and make better models for this type of Karplus parameterization), when studied at room temperature or below.

This last point provides segue into what is likely the most important consideration in X-ray structure-based parameterizations of the Karplus relation. There is a fundamental flaw in approaches that try to correlate measured experimental solution-state NMR parameters with conformations measured from static structures (such as protein X-ray crystal structures): the particular ($\phi, \psi$) angles seen for residues in crystal structures are usually singular and yet, when
the same protein is in solution, those same residues are able to sample a multitude of different conformations. The NMR scalar couplings that get measured are averages over these ensembles of conformations, and a particular ensemble-averaged scalar coupling value may not necessarily be consonant with the particular single conformation exhibited by that residue in the static crystal structure. The discrepancy between experimentally measured $^{3}J_{\text{HNHa}}$ values and those predicted (using a parameterized Karplus relation) from crystal structures was demonstrated by Smith et al. for the case of 9 different proteins (Smith et al. 1996); the distributions of $^{3}J_{\text{HNHa}}$ values for the experimentally measured data set and the predicted (based upon the static X-ray crystal structures of the same proteins) set of data exhibited noticeable differences. While the presence of a residue within a particular secondary structure element may help to constrain the range of $\phi$ values it samples in the solution state—thus helping to decrease the magnitude of the discrepancy between experimentally measured and predicted scalar coupling values—this discrepancy could potentially be more significant in the case of residues that fall into the ‘coil’ category. For example, in the case of the protein calbindin D$_{9k}$, while data for residues present in the 4 $\alpha$-helices agreed rather well with values calculated from the static crystal structure, residues belonging to termini, linker, and loop segments of the protein exhibited experimentally measured $^{3}J_{\text{HNHa}}$ values of between 6 and 8 Hz (in the range of random coil values for many residue types), the majority of which deviated by more than 1 Hz from the values predicted (calculated using a parameterized Karplus equation) from the static structure (Otten et al. 2009).

Though there exists a fundamental flaw in traditional approaches for the parameterization of the Karplus equation parameters (e.g. empirical calibrations of measured scalar coupling values against dihedral angles known from static structures), there is potential promise in more recent approaches to Karplus equation parameterization, in which backbone dynamics and motional averaging effects in protein residues have been taken into account (Case et al. 2000; Vogeli et al. 2007; Markwick et al. 2009). Some of these approaches have been able to demonstrate an improvement in self-consistency of the parameterization with improved agreement between experimental couplings and those predicted (calculated) from the generated dynamic ensemble. Karplus parameterizations aside, prediction/database approaches do have the advantage (over model peptide approaches, vide infra) of a large database size with a greater diversity of combinations in the primary sequence surrounding the central residue of interest.

In model peptide approaches, small peptides of the form Ac-GG-X-GG-NH$_{2}$ (Plaxco et al. 1997; Schwarzinger et al. 2000; Shi et al. 2005) have been used to model and represent the random coil behavior of individual amino acids (again, ‘X’ represents potentially any of the 20
proteogenic amino acids). The residue-specific random coil $^3J_{\text{HNH}}$ values reported in these sets were all recorded at 20°C and in two of the three cases, in the presence of denaturant (either 8 M urea or 6 M GuHCl); values from the different sets are in reasonably good agreement with one another. One shortcoming of model peptide approaches, however, is that sequence-dependent effects (other than those due specifically to the residue Glycine) are not taken into account with the use of this single model peptide. As mentioned in the last section (‘Secondary Chemical Shifts’), the conformational distribution of an amino acid residue is affected by neighboring residues (Ting et al. 2010). Evidence of this sequence-dependence comes from experimental measurement of $^3J_{\text{HNH}}$ scalar coupling constants in a 130-residue IDP (a fragment of S. aureus fibronectin-binding protein, devoid of any stable secondary structure), in which coupling constants were seen to be significantly increased for a given residue type when the preceding residue (‘$i-1$’) was one with a β-branched or aromatic sidechain (Penkett et al. 1997). The neighboring residue effect and this particular pattern have also been observed in the ‘coil’ libraries constructed from unstructured residues in PDB protein structures; when a neighboring residue (‘$i-1$’ or ‘$i+1$’) has either an aromatic or β-branched sidechain, the backbone angle $\phi$ of central residue ‘$i$’ is more negative for essentially all amino acids examined (Avbelj et al. 2004). For backbone angles $\phi > -120^\circ$, a more negative $\phi$ value would correlate with an increased $^3J_{\text{HNH}}$ scalar coupling constant value. To-date, though, no sequence-dependent correction factors have been reported for the residue-specific random coil values of $^3J_{\text{HNH}}$.

In addition to three-bond (‘vicinal’) scalar couplings, two-bond (‘geminal’) and one-bond (direct) scalar couplings can also yield structural information in peptides and proteins (Schmidt et al. 2011), though their conformational dependences have not yet been as heavily studied and characterized as for vicinal couplings. $^2J$ couplings related to $^{13}\text{C}'$ and $^{15}\text{N}$ nuclei depend significantly on $\phi, \psi$ torsion angle combinations. Among the several types of $^2J$ couplings that can be measured—such as $^2J_{\text{C$_\gamma$N(+1)}}$, $^2J_{\text{CHN(+1)}}$, $^2J_{\text{HN$_\alpha$C}}$, $^2J_{\text{H$_\alpha$C}}$, $^2J_{\text{H$_\alpha$C$_\beta$}}$, etc.—the $^2J_{\text{C$_\alpha$N(+1)}}$ (coupling between the alpha-Carbon of one residue and the amide Nitrogen of the proceeding residue ‘$i+1$’, Figure 2.2a) exhibits the strongest dependence on molecular conformation and a range of scalar coupling values whose magnitudes are comparable to those seen for $^3J_{\text{HNH}}$ (Schmidt et al. 2010). $^2J_{\text{C$_\alpha$N(+1)}}$ correlates most strongly with $\psi_i$ and to a lesser extent with $\phi_i$; a five-parameter combined Kaplus-type equation (a function of both $\psi_i$ and $\phi_i$) has preliminarily been parameterized on the basis of data from folded ubiquitin and staphylococcal nuclease (Wirmer et
al. 2002). $^2J_{CaN(+1)}$ coupling constants collected for unfolded (denatured) ubiquitin were rather uniform (suggesting that this coupling constant is independent of amino acid sequence and residue type), with mean value $8.0 \pm 0.3$ Hz; application of the parameterized combined Karplus-type equation for $^2J_{CaN(+1)}$ to observed $\phi, \psi$ values for residues in the ‘coil’ library derived from X-ray crystal structures yielded essentially the same value, suggesting that 8.0 Hz might serve as a general (non-residue-specific) random coil coupling constant value (Wirmer et al. 2002). For residues in an $\alpha$-helix, $^2J_{CaN(+1)}$ is smaller (averaged value of 6.1 Hz), and larger (averaged value of 8.5 Hz) for residues in a $\beta$-sheet (Wirmer et al. 2002; Schmidt et al. 2010).

In contrast with the two-bond (‘geminal’) couplings described above, one-bond scalar couplings have found more widespread use in the study of peptides and proteins, including IDPs (Noval et al. 2013; Xiang et al. 2013). One-bond (direct) scalar couplings tend to be significantly larger (Figure 2.2a) than the values observed for two- or three-bond scalar couplings (Bystrov 1976). In the case of $^1J_{CaH\alpha}$, $^1J_{CaC\beta}$, $^1J_{CaC'-}$, and $^1J_{CaN}$ couplings, all depend upon the $\phi$ angle and $\psi$ as well (Schmidt et al. 2009), thus aiding in determination of the angle $\psi$ which is difficult to reliably determine via $^3J$ coupling approaches. Combined Karplus-type equations (functions of both $\phi$ and $\psi$) have been parameterized for $^1J_{CaN}$ (Wirmer et al. 2002), $^1J_{CaC\beta}$ (Cornilescu et al. 2000), and $^1J_{CaH\alpha}$ (Vuister et al. 1992) on the basis of empirical correlation of measured coupling values with known protein structures, with the relations between the scalar couplings and backbone torsion angles being confirmed by theoretical calculations (such as density functional theory) on small peptide mimetics. As an example, in the case of $^1J_{CaH\alpha}$, the relation between scalar coupling values and conformation is described by (Vuister et al. 1992):

$$^1J_{CaH\alpha} = 140.3 + 1.4 \cdot \sin(\psi + 138^\circ) - 4.1 \cdot \cos(2(\psi + 138^\circ)) + 2.0 \cdot \cos(2(\phi + 30^\circ)) \quad [\text{eq. 3}]$$

which was parameterized on the basis of data from the proteins basic pancreatic trypsin inhibitor (BPTI), staphylococcal nuclease, and calmodulin (Vuister et al. 1992). Although one-bond scalar coupling constants, such as $^1J_{CaH\alpha}$, have often mistakenly been thought to have uniform values in proteins, significant and conformationally-dependent variations in their magnitudes occur. For example, in a collection of 3 different proteins, $^1J_{CaH\alpha}$ values exhibited a range of $131.6 - 150.3$ Hz (Vuister et al. 1992), which constitutes a broader range of absolute values (in Hz) than that spanned by $^3J_{HNH\alpha}$ couplings in proteins. As with $^3J_{HNH\alpha}$, secondary couplings can
also be calculated for $1J_{\text{CaH}_\alpha} [\Delta^1J_{\text{CaH}_\alpha} = 1J_{\text{CaH}_\alpha(\text{experimental})} - 1J_{\text{CaH}_\alpha(\text{random coil})}]$ to identify propensities for local secondary structure. Residue-specific random coil values (derived from measurements on amino acids within random coil peptide segments at 35°C) have been reported for 18 amino acids residues (Vuister et al. 1993b). Positive values of $\Delta^1J_{\text{CaH}_\alpha}$ on the order of ~ + 4–5 Hz are indicative of full $\alpha$-helical structure (Vuister et al. 1993b; Schmidt et al. 2009), while small negative values (on the order of ~ −0.6 Hz, for instance) are expected for extended $\beta$-strand structure. Positive deviations of $1J_{\text{CaH}_\alpha}$ values from random coil on the order of ~1.1 Hz are expected for polyproline II structure (Lam et al. 2003). No investigations into whether or not there are nearest-neighbor effects on residue-specific $1J_{\text{CaH}_\alpha}$ random coil values have been reported.

Residual Dipolar Couplings (RDCs)

Direct dipolar couplings originate from through-space magnetic dipole-dipole interaction between two nuclei possessing magnetic moments; since the interaction is ‘through-space’ rather than ‘through-bonds’ (as in the case of $J$-coupling, see section above, ‘Scalar Couplings’), the interacting nuclei need not be joined either directly or indirectly by covalent bonds. However, for intra-molecular dipolar couplings the symbol $^nD$ is often used, where $n$ is the number of covalent bonds joining the two nuclei.

The magnitude of the dipolar coupling between two spins (nuclei) $i$ and $j$ depends upon their inter-nuclear distance, $r$, as well as the relative orientation (angle $\theta$ with respect to the applied static magnetic field, $B_0$) of the internuclear vector (Figure 2.3a). Dipolar couplings between covalently bound nuclei can be intrinsically very strong; for example, for an amide $^1\text{H}^N–^{15}\text{N}$ spin pair with internuclear vector statically oriented (as in the solid state) in the direction of the applied magnetic field (e.g. $\theta = 0^\circ$), the maximum dipolar coupling of ~23 kHz could result [as per eq. 4, below] (Prestegard et al. 2004). However, in an isotropic solution state, rapid molecular tumbling causes the internuclear vector to sample all possible orientations with equal probability such that dipolar couplings average to zero. But, if a small degree of alignment or orientational bias could be introduced into the sample, such that the orientational sampling of the internuclear vector were no longer isotropic with respect to the applied magnetic field $B_0$, then small (“residual”) dipolar couplings can be observed. For conformational exchange occurring on timescales faster than the millisecond, the magnitude of the residual dipolar coupling (RDC)
**Figure 2.3.** Residual dipolar couplings (RDCs) in protein secondary structure elements. (A) Orientation (angle $\theta$) of the internuclear vector (orange) between two nuclei $i$ and $j$ relative to the applied static magnetic field, $B_0$. (B) Plot of the second-order Legendre polynomial of $\cos(\theta)$ [$P_2(\cos(\theta))$]. For positive $\theta < 180^\circ$, $P_2(\cos(\theta))$ changes sign (e.g. is equal to zero) for angles $\theta = 54.74^\circ$ and $125.26^\circ$. (C) Figurative representation of the amide $^{15}\text{N} - ^1\text{H}$ internuclear vector angles in an unfolded protein dissolved in weakly aligning medium with orientation of the long molecular axis parallel to the applied magnetic field, $B_0$. Dipolar couplings measured for $^{15}\text{N} - ^1\text{H}$ internuclear vectors (e.g. $^1\text{D}_{\text{IN}}$) in more extended conformations (i.e. $\theta \approx 90^\circ$) such as Polyproline II and $\beta$-strand will have negative values (see plot in part b), whereas those in helical conformations align rather parallel with the direction of the chain (i.e. $\theta \approx 0^\circ$) and will have positive values. Figure in panel C adapted from Jensen et. al., 2009 with permission (Jensen et. al. 2009).
Figure 2.3. (Continued)

A

B

\[ \frac{3 \cos^2(\theta) - 1}{2} \]

\[ -180 -150 -120 -90 -60 -30 0 30 60 90 120 150 180 \]

\[ \theta \text{ (degrees)} \]

\[ 54.74^\circ \]

\[ 125.26^\circ \]

C

Polyproline II (PPII, \( \beta_p \))

\( \beta \)-strand (\( \beta_\alpha \))

\( \alpha \)-helix (\( \alpha_h \))
between nuclei \(i\) and \(j\) is proportional to the inverse third-power of the internuclear distance \(r\), and to the second-order Legendre polynomial of \(\cos(\theta)\) (Figure 2.3b):

\[
D_{ij} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{8 \pi^3 r^3} \left( \frac{3 \cos^2(\theta) - 1}{2} \right) = D_{\text{max}} \langle P_2(\cos(\theta)) \rangle \quad \text{[eq. 4]}
\]

with

\[
D_{\text{max}} = -\frac{\gamma_i \gamma_j \hbar \mu_0}{8 \pi^3 r^3} \quad \text{[eq. 5]}
\]

where \(\theta\) is the instantaneous angle of the internuclear vector with respect to the applied magnetic field, \(r\) is the internuclear distance, \(\hbar\) is Planck’s constant, \(\mu_0\) is the permeability of space, and \(\gamma_i\), \(\gamma_j\) are the gyromagnetic ratios of nuclei \(i\) and \(j\), respectively, and the average within the angular brackets represents an ensemble average over all molecules in the sample (Meier et al. 2008).

Many molecules, including some proteins (mainly those with a paramagnetic site), possess significant nonzero anisotropic (para)magnetic susceptibility tensors such that they attain orientation in the applied strong magnetic field, sufficient to produce \(^1\)D\(_{\text{HN}}\) RDCs of several hertz, for instance (Tolman et al. 1995). However, the advent of media that could produce weak alignment in a strong magnetic field created the general possibility of measuring RDCs of the desired magnitude for essentially any protein of interest (Tjandra et al. 1997a). Up to that point in time, protein structure determination via NMR relied heavily upon NOEs (see ‘NOEs and PREs’ section below) for the determination of interproton distances; NOEs have an inverse sixth-power dependence \(r^{-6}\) upon the interproton distance, thus limiting their detection to distances \(< 5 \text{ Å}\) (Clore et al. 1989). The newfound accessibility of RDC data proved to be a boon to NMR protein structure determination; RDCs, with their inverse third-power dependence \(r^{-3}\) upon internuclear distance allow greater distances (than with NOEs) and more long-range order to be measured in proteins, leading to improvements in the quality and accuracy of determined protein structures (Tjandra et al. 1997b). In the case of folded proteins, an overall molecular alignment tensor is first determined for the protein by measuring RDCs for spin-pairs with essentially constant internuclear distance \(r\) (such as those joined by a single covalent bond), such that the angles of these internuclear bond vectors (\(i.e.\) the other variable, \(\theta\), in eq. 4) relative to the applied magnetic field (and relative to each other) can be calculated from the measured RDCs (Prestegard et al. 2004). Within this local frame of reference, the \(r\) and \(\theta\) values for various other internuclear vectors can be determined, thus aiding structure determination.
To achieve weak macromolecular orientation in an applied magnetic field, a variety of aligning media have been developed, such as the lipid bicelles that were the first medium reportedly used to collect RDCs in biomolecules (Tjandra et al. 1997a). Bicelles are lipid bilayer disks composed of mixtures of either ester- or ether-linked phospholipids and the composing phospholipids are often of net neutral charge but the lipid bilayer disks can also be doped with charged lipids (Glover et al. 2001); these disks orient with the bilayer normal perpendicular to the applied magnetic field (Forrest et al. 1981; Vold et al. 1996). Similar to bicelles—in terms of being discoid and containing lipids—patches of purple membrane (PM) can also be used as aligning media, though their orientation is with the bilayer normal parallel to the static magnetic field (Sass et al. 1999). Other aligning media include: filamentous phage and rod-like virus particles (Hansen et al. 1998; Clore et al. 1998); polyacrylamide gels that have been compressed or strained either laterally or longitudinally to create anisotropic cavities in which the protein molecules are dissolved (Tycko et al. 2000; Sass et al. 2000); and lyotropic ethylene glycol/alcohol phases, which form uncharged planar bilayers with the hydrophilic polyethylene glycol (PEG) headgroups directed toward the aqueous phase, and with lamellar bilayers forming into concentric tubes of different radii with their long axes aligned parallel to the direction of the applied magnetic field (Rückert et al. 2000; Hoffmann 1994). In the case of most of these aligning media, the alignment results from steric repulsion between the protein and medium, whereas in the case of bacteriophage particles, purple membrane disks, or charged forms of the other media, alignment results from a combination of both steric and electrostatic interactions.

Sterically aligning media are the most commonly used for the measurement of RDCs in IDPs, and in the case of lyotropic PEG/alcohol phases or for acrylamide gel that is radially compressed (i.e. cavities are longitudinally elongated), the proteins align essentially as depicted in Figure 2.3c, with the long molecular axis oriented parallel to the static magnetic field (Obolensky et al. 2007; Jensen et al. 2009). The degree of alignment (and thus the magnitude of the measured RDCs) can often be tuned by altering the concentration of the aligning media and/or its composition for multi-component media; this degree of alignment must often be tuned for a given protein to yield desirable RDC values. If the degree of alignment is too small, the magnitude (and signal-to-noise) of the RDCs may be too small for reliable measurement and not yield useful information. If the degree of alignment is too large, this can potentially lead to increased dipolar broadening of peaks (and thus decreased resolution, which is one of the obstacles encountered in solid-state NMR), as well as increased complexity and crowding of the spectra in that, more correlations (in addition to the shorter-range ones) will be seen from nuclei.
that are more distant from one another. Of course, techniques have been developed that can effectively ‘mask’ the shorter-range proton-proton correlations, thus highlighting the longer-range dipolar correlations observed from higher degrees of alignment, yielding desirable and very useful long-range structural information (Wu et. al. 2002; Boisbouvier et. al. 2003). But still, the general adage concerning the degree of alignment for collection of RDCs in proteins is: Not too little, but not too much.

A large array of different RDCs (both homo- and heteronuclear) can be measured in a protein, stemming from the variety of NMR pulse sequence programs designed to probe them (Prestegard et. al. 2004). As an example, one straightforward method for determining one-bond heteronuclear RDCs in proteins involves the collection of $^1$H–$^{15}$N or $^1$H–$^{13}$C HSQC spectra without heteronuclear decoupling during detection; the resulting spectrum contains peaks that are bifurcated (or, ‘split’) by the coupling between the two nuclei. For the case of nuclei exhibiting scalar coupling to one another, the measured splitting between peaks for the aligned state is the sum of both the scalar and dipolar couplings ($J + D$) between the two nuclei. Dipolar couplings are obtained from the difference between peak splittings measured in the oriented phase ($J + D$) and in the isotropic phase (just $J$). In other words, the same basic NMR experiment is run twice—once with the protein sample in regular solution (e.g. the isotropic phase), and again with the protein sample in aligning medium; mathematically, $D = (J + D) - J$.

Though the measurement and application of RDCs in proteins was initially focused more on structural determination in proteins of the folded variety, the measurement and use of RDCs in structural characterizations of unfolded proteins and IDPs has grown (Shortle et. al. 2001; Mohana-Borges et. al. 2004; Mukrasch et. al. 2007; Jensen et. al. 2011). The RDCs collected on these extended types of molecules are primarily aimed at the determination of localized order along the chain—that is, the RDCs collected are generally more short-range and between nuclei that have scalar coupling interactions with one another (such as $^1$D$_{HN}$, $^1$D$_{C\alpha-H\alpha}$, $^1$D$_{C\beta-C\alpha}$, $^2$D$_{C'-HN}$, $^3$D$_{HN-C\alpha}$, $^3$D$_{IN-H\alpha}$, etc.). But, more long-range interproton ($^1$H$_N$–$^1$H$_\alpha$) and interamide proton ($^1$H$_N$–$^1$H$^N$) RDCs, such as $^1$H$_{N_1}$–$^1$H$_{N_{i+1}}$, $^1$H$_{N_1}$–$^1$H$^N_{i+2}$, and even $^3$H$_{N_1}$–$^1$H$^N_{i+3}$ RDCs (Meier et. al. 2007a,b) have also been collected and utilized in the structural analysis of unfolded proteins.

As a simple illustration of how RDCs can yield structural information in IDPs and unfolded proteins, the simplest case of $^1$D$_{HN}$ will be considered here. As mentioned above, for the case of alignment in lyotropic PEG/alcohol phase or for acrylamide gel that is radially compressed (with longitudinally-oriented cavities), IDPs and unfolded proteins align as depicted in Figure 2.3c with long molecular axis parallel to the static magnetic field. In this instance, the
amide proton-nitrogen internuclear distance, \( r \), is taken to be a constant (vibrationally-averaged bond length of \(~1.02\, \text{Å};\) Yao et al. 2008), so that remaining variable \( \theta \) in [eq. 4] above—that is, the orientational information about the \(^1\text{H}–^{15}\text{N}\) internuclear vector relative to the molecular alignment axis—can be determined from the measured RDCs. In the case of \( \beta \)-strand or polyproline II structure (Figure 2.3c, left), the \(^1\text{H}–^{15}\text{N}\) internuclear vectors have nearly orthogonal orientation with respect to the static magnetic field (\( \theta \) approaches 90°), such that the value of \( P_2(\cos(\theta)) \) is negative (Figure 2.3b) and the measured RDCs have negative sign. In contrast, for \( \alpha \)-helical structure (Figure 2.3c, right), the \(^1\text{H}–^{15}\text{N}\) internuclear vectors have nearly parallel orientation with respect to the static magnetic field (\( \theta \) is closer to 0°), such that the value of \( P_2(\cos(\theta)) \) is positive and the measured RDCs exhibit positive sign. Until more recently, the majority of the RDCs measured and reported for IDPs and unfolded proteins were \(^1\text{D}_{\text{HN}}\). While still providing a powerful and useful source of conformational information, it is recognized that some ambiguities can result from \(^1\text{D}_{\text{HN}}\) taken alone; for example, based upon the analysis just presented here, \(^1\text{D}_{\text{HN}}\) RDCs alone cannot distinguish between a right-handed \( \alpha \)-helix and one that is left-handed. In light of such considerations, the modern trend is towards measurement and inclusion of a greater number and diversity of RDCs in IDPs, so as to raise potential degeneracies and ambiguities in conformation that may have been present in earlier, more simplistic approaches, and to just generally improve the detail and accuracy of molecular conformations determined from RDCs measured in IDPs (Jensen et al. 2009).

**Nuclear Overhauser effects (NOEs) and Paramagnetic Relaxation Enhancements (PREs)**

Like the secondary chemical shifts, scalar couplings, and RDCs reviewed above, nuclear Overhauser enhancements (NOEs) can also provide information on secondary structure in proteins. The nuclear Overhauser effect arises from through-space dipolar coupling between two spins (nuclei), with exchange of magnetization occurring between the two nuclei (usually protons) in a process termed cross-relaxation (Clore et al. 1989). The rate of cross-relaxation \( (\sigma_{ij}^{\text{NOE}}) \) between two nuclei (protons) \( i \) and \( j \) is proportional to the inverse sixth-power of the internuclear (interproton) distance \( r \) :

\[
\sigma_{ij}^{\text{NOE}} = \frac{(\gamma_i)^2(\gamma_j)^2h^2(\mu_0)^2}{160\pi^4(r_{ij})^6}\left(-1 + \frac{6}{1+4(\omega_0)^2(\tau_c)^2}\right) \quad [\text{eq. 6}]
\]
where \( h \) is Planck’s constant, \( \mu_0 \) is the permeability of space, \( r_{ij} \) is the interproton separation distance, \( \gamma_i \) and \( \gamma_j \) are the gyromagnetic ratios of nuclei (protons) \( i \) and \( j \), respectively, \( \omega_0 \) is the Larmor frequency (the \(^1\)H frequency of the NMR instrument used), and \( \tau_c \) is the rotational correlation time for the interproton vector (Cavanagh et al. 1996). The enhancement of a nucleus’ magnetization is in-turn proportional to the cross-relaxation rate (\( \sigma_{ij}^{\text{NOE}} \)); observed intensities of the peaks corresponding to the correlation between the two nuclei (e.g. ‘off-diagonal’ or ‘cross-peaks’ that, in a 2-D spectral plane, are located at the coordinates of the resonance frequency of the first nucleus in one dimension and the resonance frequency of the second nucleus in the other dimension) are proportional to the cross-relaxation rate and thus to the inverse sixth-power of the interproton separation distance. Due to this distance dependence, in order for an NOE cross-peak for two nuclei \( i \) and \( j \) to appear in a spectrum, the two nuclei must be located within \( \sim 5 \) Å or less of each other; thus, NOEs can provide information about the arrangement of different protons within a given residue and also between neighboring residues in the primary sequence (this yields secondary structure information), as well as the arrangement of protons that are distant in the primary sequence, but may be spatially proximate to one another due to packing within a protein structure (e.g. tertiary structural information).

The indication of secondary structure from NOEs for such elements as \( \alpha \)-helices and \( \beta \)-strands derives from the fact that these regular structures each have characteristic intra-molecular distances for specific pairs of protons. For example, in a regular \( \alpha \)-helix the distance between amide protons in neighboring (in the primary sequence) residues [sometimes denoted as \( d_{\alpha N}(i, i+1) \)] is 2.8 Å, while the same \( d_{\alpha N}(i, i+1) \) distance in a regular \( \beta \)-strand is 4.3 Å; the interproton distance between the \( \alpha \)-proton of residue \( i \) and the amide proton of residue ‘\( i + 1 \)’ \( [d_{\alpha N}(i, i+1)] \) is equal to 3.5 Å in an \( \alpha \)-helix, but equal to 2.2 Å in a \( \beta \)-strand (Wuthrich et al. 1984). Also, additional medium-range backbone NOEs can be seen for such proton pairs as \( d_{\alpha N}(i, i+2) \), \( d_{\alpha N}(i, i+3) \), and \( d_{\alpha N}(i, i+4) \) [with distances 4.4 Å, 3.4 Å, and 4.2 Å, respectively], and \( d_{\alpha N}(i, i+2) \) and \( d_{\alpha N}(i, i+3) \) [distances 4.2 Å and 4.9 Å] in an \( \alpha \)-helix, while these same correlations will be absent from the spectrum for a \( \beta \)-strand (Wuthrich et al. 1984). As an example, for the predominately disordered \( D. \ melanogaster \) dynein IC:1–143 protein, \( d_{\alpha N}(i, i+1) \) and \( d_{\alpha N}(i, i+2) \) NOEs observed from an \(^{15}\)N-edited NOESY (Nuclear Overhauser Effect Spectroscopy) experiment were seen for various residues within the N-terminal 40 residues (see Chapter 3, Figure 3.3), which was determined to be an \( \alpha \)-helix; these inter-amide proton...
correlations were not observed for the remainder of residues in the predominately disordered protein (Morgan et. al. 2011).

It should be noted at this point that, in addition to the inverse sixth-power dependence on interproton distance, the cross-relaxation rate between two spins is also dependent upon dynamics (i.e. $\tau_c$ in eq. 6, which corresponds to the effective rotational correlation time of the interproton vector); the intensities of NOEs are sensitive to overall molecular motion of the protein molecule as well as internal motions within the protein, which can change the distance between two nuclei with time (Wuthrich 2003). For mobile and flexible segments of a protein, the internal motion generally results in shorter effective correlation times $\tau_c$ in those regions, which results in a decreased dipolar cross-relaxation rate; this can manifest itself in the form of NOEs that become small or even disappear (Williamson 2009). This phenomenon might possibly have contributed to the absence of observed $d_{NN}(i, i+1)$ and $d_{NN}(i, i+2)$ correlations for residues 48–60 in the aforementioned example of D. melanogaster dynein IC:1–143; on the basis of various NMR data collected, this small segment of the protein was interpreted to be a nascent $\alpha$-helix (Morgan et. al. 2011)—that is, a helix that is transient, conformationally flexible (with greater internal mobility and probable smaller effective $\tau_c$ values, than for the $\alpha$-helix assigned in residues 1–40), and is only partially populated in time, interconverting between random coil and helical conformations. Of course, this dynamical effect would be in addition to the fact that helical conformations in this region are lowly-populated, which could also have lead to weak or non-visible $d_{NN}$ NOEs in this case.

It should also be noted that there exists an additional complication arising from conformational flexibility in terms of the interpretation of NOE data collected in IDPs. For interproton NOEs that can be observed in spectra recorded for such proteins, the rapid interconversion between multiple distinct and different conformations, coupled with the inverse sixth-power interproton distance dependence, creates complexity and ambiguity in the interpretation of NOE peak intensities in terms of interproton distances and the conformations to which they might correspond (Jensen et. al. 2009). As an example, for different polypeptide conformers in which a pair of protons might be separated by distances of 2.5, 3.5, or 4.5 Å, similar NOE intensities would be detected for each case if the conformations were populated at 2, 12, or 53% of the time, respectively (Fiebig et. al. 1996). In other words, the effective distance indicated by NOE intensity is disproportionately weighted toward that of a conformer with shortest interproton distance, even if that state is only very lowly-populated, while conformers with longer interproton distance might be highly-populated and yet make very little contribution
to the observed intensity. For the reasons presented, NOEs are probably most applicable to the detection of more static and well-defined, regular secondary structure elements that may be dispersed throughout and punctuate the flexible and disordered regions within an IDP.

Similar to NOEs in principle, Paramagnetic Relaxation Enhancements (PREs) also have an inverse sixth-power dependence upon the separation distance between two spins. In this case, the two spins are a proton (such as an amide proton present in an amino acid residue in the protein) and an unpaired electron present in an extrinsic radical spin label that has been covalently attached to the protein (Figure 2.4a). The application of spin labels in NMR spectroscopy is based upon the increase in transverse relaxation rates ($R_2$ values) caused in the protons spatially proximate to the paramagnetic center (e.g. the unpaired electron in a nitroxide spin label), with the extent of dipolar peak broadening (faster relaxation, larger effective $R_2$ value) dependent upon the inverse sixth-power of the proton–unpaired electron separation distance. The increase (enhancement) of the proton transverse relaxation rates due to the contribution from the paramagnetic species is given by the Solomon-Bloembergen equation (Solomon 1955; Clore et. al. 2009):

$$\Gamma_2 = \frac{1}{15} \left( \frac{\mu_0}{4\pi} \right)^2 \left( \frac{\gamma_H^2 \gamma_e^2 h^2 s_e(s_e + 1)}{4\pi^2 (r_{H-e})^6} \right) \left( 4\tau_c + \frac{3\tau_c}{1 + (\omega_H^2)(\tau_c)^2} \right)$$

where $h$ is Planck’s constant, $\mu_0$ is the permeability of space, $s_e$ is the electron spin quantum number, $r_{H-e}$ is the distance between the proton and unpaired electron, $\gamma_H$ and $\gamma_e$ are the gyromagnetic ratios of a proton and an electron, respectively, $\omega_H$ is the proton Larmor frequency, and here $\tau_c$ is the correlation time for the electron-proton vector; this equation is based upon assumptions that the vector between the proton and electron is free to undergo isotropic rotational diffusion and that their separation distance $r_{H-e}$ is essentially fixed (Gillespie et. al. 1997a).

While equations for NOEs [eq. 6] and PREs [eq. 7] bear similarity to one another, it should be noted that, in the case of PREs, the gyromagnetic ratio for an electron ($\gamma_e$) is 658 times larger than that of a proton, and the value of this parameter enters quadratically into the equation [eq. 7] (Jahnke 2002). Thus, while the measurable effect of an NOE is limited to (inter-proton) distances of ~5 Å or less, the action underlying the PRE effect can extend to distances of up to ~25 Å, thereby making it suitable for the detection of long-range interactions and as an aid to tertiary structure determination.
Figure 2.4. Paramagnetic Relaxation Enhancements (PREs) in disordered proteins.
(A) Schematic of the covalent modification reaction by which a nitroxide radical spin label, such as MTSSL [(1-oxyl-2,2,5,5-tetramethyl-Δ3-pyrroline-3-yl)methyl methanethiosulfate] (Berliner et al. 1982), is attached to the sidechain of a cysteine residue in a protein (blue). (B) NMR-PRE intensity ratio plots ($I_{ox}/I_{red}$) for α-synuclein with MTSSL spin label attached to individually engineered cysteine residues at positions Q24, S42, Q62, S87, and N103 along the protein sequence. The red dotted lines in each plot represent PRE ratios simulated for spin-labeled random coil structures for each variant. (C) Two-dimensional residual contact plot for native α-synuclein with schematic of domain structure shown below. Darker contour colors (brown) correspond to intra-molecular contacts that are more probable in the PRE-restrained molecular dynamics ensemble as compared to those seen in the simulated random coil ensemble for the same protein. In particular, contacts between residues ~120–140 of the C-terminus and residues ~30–100 in the central region (‘fibril core region’) are indicated. Release of these long-range tertiary interactions are proposed to potentiate, via self-association of exposed fibril core regions, the formation of inter-molecular aggregates in cells of the brainstem that are associated with Parkinson’s disease (Bertoncini et al. 2005). Figures in panels B and C adapted from Dedmon et al., 2005 with permission (Dedmon et al. 2005).
Quantitative measurement of the PRE $\Gamma_2$ rates can be made by collecting an identical set of $T_2$–type experiments (one on the paramagnetic sample, and another on the conjugate diamagnetic version of the sample); $\Gamma_2$ is given by the difference in rates measured for the paramagnetic vs. diamagnetic states (Iwahara et. al. 2007). Alternatively (and more frequently seen in the literature), a more simplistic approach involves the collection of two regular HSQC spectra—one of the sample in its paramagnetic (‘oxidized’) state and another of the same, exact sample with the paramagnetic center in its diamagnetic (‘reduced’) state (Battiste et. al. 2000; Iwahara et. al. 2007). The values of $\Gamma_2$ can be determined from the ratio of the HSQC peak intensities of the sample in its oxidized (‘paramagnetic’) and reduced (‘diamagnetic’) states thusly:

$$\frac{I_{OX}}{I_{red}} = \frac{R_2 * e^{(-\Gamma_2 \cdot \tau_m)}}{R_2 + \Gamma_2} \quad \text{[eq. 8]}$$

where $R_2$ represents the intrinsic transverse relaxation rate of the proton (measured in the absence of any spin-label, or with the spin-label in its diamagnetic state), $\Gamma_2$ is the effective transverse relaxation rate in the presence of the paramagnetic center (and is the value to be determined), and $\tau_m$ corresponds to the INEPT coherence transfer evolution time in a standard HSQC experiment, which is usually on the order of ~10 ms (Gillespie et. al. 1997a; Battiste et. al. 2000). The value of $\Gamma_2$ determined from application of [Eq. 8] can then be used in conjunction with [Eq. 7] to derive the value of the separation distance between a proton and the unpaired electron in the paramagnetic moiety ($r_{H-e}$), and thus structural information.

Some of the earliest applications of the PRE effect in determination of structural information in proteins were for the case of metalloproteins, containing an inherent metallic paramagnetic center or else some (diamagnetic) metal ion that could be substituted with a paramagnetic metal ion such as low-spin Fe$^{3+}$, Cu$^{2+}$, Co$^{2+}$, or certain lanthanides (Bertini et. al. 1999). In such instances, where the paramagnetic moiety has fixed position within the protein, and where the protein is folded with more-or-less fixed structural elements, the metal-proton distances can be expected to exhibit only small fluctuations and can be approximated as vectors of essentially fixed length; in such situations, back-calculation of dipole-dipole separation distances from the measured PREs via the Solomon-Bloembergen equation [Eq. 7] is relatively straightforward (Clore et. al. 2009).
But then consider the case of an extended, heterogeneous ensemble of conformations undergoing dynamic inter-conversion. Similar considerations to those mentioned for the interpretation of NOEs (vide supra) apply to the interpretation of measured PREs in these dynamic and flexible systems. To reiterate, different conformations of the same molecule will possess different proton-electron separation distances and, given the inverse sixth-power distance-dependence of the interaction, the signal contributions made by these different conformers will be biased toward those conformations with shorter proton-electron distances. A given measured $\Gamma_2$ value could correspond equally well to a weakly-populated conformer with shorter proton-electron separation distance, or to a highly-populated conformation with greater separation distance. In light of such considerations, early researchers who measured PREs in unfolded and intrinsically disordered proteins proposed and advocated for a “semi-quantitative” approach to the interpretation of PREs measured in these extended, dynamic systems (Gillespie et. al. 1997b). Indeed, earlier (as well as more recent) examples of studies reported in the literature for PREs measured in unfolded proteins simply presented the PRE data as plots of $I_{\text{ox}}/I_{\text{red}}$ (a.k.a. “$I_{\text{paramagnetic}}/I_{\text{diamagnetic}}$”) ratios and interpretation of said data was highly qualitative in nature, with no ostensible attempt made at calculation of any proton-electron distances (Teilum et. al. 2002; Lietzow et. al. 2002; Platzer et. al. 2011).

In NMR-PRE experiments performed on unfolded or intrinsically disordered proteins, a collection of protein variants is usually created, with spin label attachment sites distributed along the length of the protein, one spin label site per protein variant (Figure 2.4b). Ideally, a spin label attached at one site of the protein (call it ‘site A’) should affect the sites of modification in the other protein variants in the same way that spin labels at these sites affect the protein at ‘site A.’ In other words, reciprocity should ideally be observed. With its ~25 Å range of action, the enhanced sensitivity of PREs permit detection of intra-protein interactions (e.g. transient tertiary structure) that exist at relatively low populations and/or longer distances than could be obtained from standard NOEs. In the case of a spin label covalently attached to an extended polypeptide chain, the unpaired electron in the label is expected to exert significant localized paramagnetic broadening effects (resulting in low values of the $I_{\text{paramagnetic}}/I_{\text{diamagnetic}}$ ratio) for residues that are adjacent in the primary sequence to the site of spin label attachment; for a polypeptide chain behaving as a random coil polymer, the paramagnetic broadening decreases with increasing distance in the primary sequence from the site of spin label attachment (as seen in the example of $\alpha$-synuclein presented in Figure 2.4b). In addition to localized effects of the paramagnetic spin label, deviations of the $I_{\text{paramagnetic}}/I_{\text{diamagnetic}}$ ratios from unity (i.e. values below 1.0) seen in
regions of the protein distant in the primary sequence from the site of spin label attachment give indication of long-range contacts between these parts of the protein.

Though it is acknowledged that singular quantitative distance values cannot be directly calculated from experimental PREs measured in these types of dynamic, extended proteins, computational approaches have been used to generate ensembles of spin-labeled random coil conformations of proteins, with simulated PRE values being calculated over the ensemble of conformers (Teilum et al. 2002; Dedmon et al. 2005; Ozenne et al. 2012b). Deviations of measured PRE values for a protein from those expected for a random coil description of the same protein can further highlight and aid in the identification of regions in the protein that exhibit long-range contacts. Some studies have incorporated experimentally-measured PREs as constraints in restrained molecular dynamics simulations to generate representative protein ensembles (Dedmon et al. 2005; Kristjansdottir et al. 2005); comparisons of the PRE-restrained ensemble with that simulated for a random coil model have sometimes been used to generate 2-dimensional contact probability maps for different sites within a protein (as seen in the example for α-synuclein presented Figure 2.4c). Alternatively, experimental PRE data have been used to select representative ensembles of a protein (from a large pool of possible conformers) that, when taken as an ensemble, are able to reproduce the experimental PRE data (Salmon et al. 2010). However, in the absence of additional long-range conformational information (such as from Small Angle X-Ray Scattering (SAXS) or from diffusion determination techniques that yield information on hydrodynamic radii), such computational approaches mentioned here routinely yield ensembles that are overly-compact, despite their ability to reproduce measured PRE values (Schwalbe et al. 2014).
Chapter 3

Structural Dynamics and Multiregion Interactions in Dynein-Dynactin Recognition

Jessica L. Morgan, Yujuan Song, and Elisar Barbar

Published in *The Journal of Biological Chemistry*, 2011, volume 286(11), pages 39349 - 39359
Abstract

Cytoplasmic dynein is a 1.2-MDa multisubunit motor protein complex that, together with its activator dynactin, is responsible for the majority of minus end microtubule-based motility. Dynactin targets dynein to specific cellular locations, links dynein to cargo, and increases dynein processivity. These two macromolecular complexes are connected by a direct interaction between dynactin’s largest subunit, p150Glued, and dynein intermediate chain (IC) subunit. Here, we demonstrate using NMR spectroscopy and isothermal titration calorimetry that the binding footprint of p150Glued on IC involves two noncontiguous recognition regions, and both are required for full binding affinity. In apo-IC, the helical structure of Region 1, the nascent helix of Region 2, and the disorder in the rest of the chain are determined from coupling constants, amide-amide sequential NOEs, secondary chemical shifts, and various dynamics measurements. When bound to p150Glued, different patterns of spectral exchange broadening suggest that Region 1 forms a coiled-coil and Region 2 a packed stable helix, with the intervening residues remaining disordered. In the 150-kDa complex of p150Glued, IC, and two light chains, the non-interface segments remain disordered. The multiregion IC binding interface, the partial disorder of Region 2 and its potential for post-translational modification, and the modulation of the length of the longer linker by alternative splicing may provide a basis for elegant and multifaceted regulation of binding between IC and p150Glued. The long disordered linker between the p150Glued binding segments and the dynein light chain consensus sequences could also provide an attractive recognition platform for diverse cargoes.

Introduction

Cytoplasmic dynein, the principal motor responsible for retrograde transport along microtubules, is involved in a variety of critical cellular processes, including orientation and assembly of the mitotic spindle and subsequent chromosome segregation, orientation and transport of organelles, and establishment of cell polarity (reviewed in Vallee et al. 2004). Dynein, a large ~1.2-MDa multisubunit protein complex, is conceptually divided into two major functional domains. The motor domain is comprised of the dynein heavy chains and includes the

1The abbreviations used are: IC, 74-kDa dynein intermediate chain corresponding to gene Cdic2b; N-IC, IC residues 1-289; TcTex1, the 12-kDa dynein light chain corresponding to gene Dlc90F; LC8, the 10-kDa dynein light chain corresponding to gene Cdle2; LC7, the 11-kDa dynein light chain corresponding to gene rob1; HSQC, heteronuclear single-quantum coherence; ITC, isothermal titration calorimetry; IDP, intrinsically disordered protein; SUMO, small ubiquitin-related modifier; PDB, Protein Data Bank.
sites for microtubule binding and the ATPase activity; the cargo attachment domain is comprised of the light intermediate chains, the intermediate chains (ICs), and the light chains Tctex1, LC8, and LC7 (Pfister et al. 2006). In Drosophila melanogaster, IC (a 642-amino acid protein) consists of two structurally and functionally distinct subdomains as follows: a predominantly disordered N-terminal region (N-IC), which includes binding sites for the three light chains (Makokha et al. 2002; Lo et al. 2001; Mok et al. 2001; Susalka et al. 2002; Hall et al. 2010), and a C-terminal region, which includes seven WD40 repeats predicted to fold into a compact toroidal β-propeller structure. Molecular genetics analyses show that in D. melanogaster, IC mutations result in larval lethality, demonstrating the essential function of IC in vivo (Boylan et al. 2002).

Commensurate with the numerous functions and activities of dynein within the cell, a multitude of adaptors and regulators have been identified (Kardon et al. 2009b), including dynactin, which is required for most types of cytoplasmic dynein activities (Schroer 2004). Dynactin, a heteromultimeric macromolecular complex, first identified through its role in activation of vesicular transport mediated by cytoplasmic dynein (Gill et al. 1991), also increases the processivity of transport by dynein along microtubules (Kardon et al. 2009a; King et al. 2000b). Dynactin’s largest subunit, p150\textsuperscript{Glued}, was discovered as a 150-kDa polypeptide co-purified with cytoplasmic dynein (Collins et al. 1989), and it was subsequently determined to have extensive sequence and functional similarity to the D. melanogaster gene product Glued, a polypeptide of similar size (Holzbaur et al. 1991).

Early in vitro studies show a direct interaction between dynactin p150\textsuperscript{Glued} and dynein N-IC, implying that this interaction mediates mutual recognition between the two macromolecular complexes (Gill et al. 1991; Kardon et al. 2009b). Subsequent molecular genetics studies in D. melanogaster demonstrate that additional copies of the dynein IC gene suppress the dominant rough eye phenotype of Glued (a mutation in the p150\textsuperscript{Glued} gene) (Boylan et al. 2000). In addition, overexpression of full-length p150\textsuperscript{Glued} in COS-7 cells disrupts dynein-based motility (Quintyne et al. 1999), and conversely, overexpression of small IC fragments that bind p150\textsuperscript{Glued} perturbs endomembrane and microtubule organization in COS-7 and HeLa cells. Both effects are proposed to arise from competitive inhibition (King et al. 2003).

Elucidating the structural basis for dynein-dynactin interactions and how their functional activities are regulated requires characterization of the constituent interacting subunits. N-IC is an intrinsically disordered protein (IDP) (Makokha et al. 2002) and is predicted to form a coiled-coil in its N-terminal ~31 residues. The term “IDP” collectively refers both to proteins that are
completely disordered and to those consisting of a mixture of ordered and disordered residues, with “disordered” structure referring to a flexible ensemble of conformations that are, on average, aperiodic, extended, and not well packed by other protein atoms. IDPs play diverse roles in the promotion of supramolecular assembly and regulation of function in various binding partners and are themselves highly amenable to regulation through post-translational modification. Although the majority of IDPs in complex with binding partners have been structurally characterized in forms resulting from ligand binding-induced disorder-to-order transitions, a new class of IDPs has begun to emerge, in which these proteins do not fully fold even in the ligand-bound state, where they form partially disordered dynamic complexes (Uversky 2011a, and references therein).

In contrast to the predominance of disorder in N-IC, p150Glued is predicted to contain extensive α-helical structures with two major coiled-coil regions (residues 214–567 and 982–1086 in D. melanogaster). Stoichiometric analysis of dynactin indicates two copies of p150Glued per dynactin complex (Gill et al. 1991), and coupled with images from deep-etch electron microscopy, these observations led to the widely accepted model of a homodimeric coiled-coil structure that constitutes the projecting arm of dynactin (Schroer 2004). Analyses of truncation mutants of both IC and p150Glued in various species have produced a coarse-grained approximation of the regions necessary for their binding interaction (King et al. 2003; Vaughan et al. 1995; Karki et al. 1995; Vaughan et al. 2001). Although one study (Vaughan et al. 2002) suggests that the full-length IC includes a region that can interact with a segment (amino acids 600–811 in rat) of p150Glued, another study demonstrates binding between N-terminal residues of IC (residues 1–106 in Rattus norvegicus IC-2C isoform) and a region of p150Glued corresponding to the first predicted coiled-coil (p150Glued amino acids 214–567 in D. melanogaster) (King et al. 2003). To date, there is no report of a comparison of pre- versus post-complex structures of IC and p150Glued.

Although dynein is coupled with dynactin for many of its activities, dynein and dynactin are not constitutively colocalized in cells, which raises the question of how dynein-dynactin binding is regulated. Here, we report the multiregion recognition motif of the p150Glued binding interface in IC, and demonstrate that, in the assembled state, whether in binary combination with p150Glued or in a biologically relevant 150-kDa subcomplex with p150Glued and light chains, IC remains disordered in regions outside the interfaces with its binding partners. These results explain how alternative splicing and phosphorylation of IC might modulate its interaction with p150Glued.
Experimental Procedures

Protein Preparation.

Constructs of the *D. melanogaster* IC (Figure 3.1b) were generated by PCRs using the cDNA of the *Cdic2b* gene (accession number AF 263371.1) as the template. An IC construct corresponding to residues 1–87 (IC:1–87) was generated with an N-terminal hexahistidine (His₆) tag and a protease Factor Xa recognition sequence engineered immediately 5’ to the start codon. The PCR fragment was cloned into pCR2.1 TOPO (Invitrogen), followed by subcloning into a pET15d (Novagen) expression vector. For NMR studies, an IC construct containing two sequential LC8-binding sites (hereafter designated IC:1–143̂LL) was generated by PCR with a synthetic nucleotide template using the sequence LVYTKQTQTT in place of residues 111–120 (described in Hall *et al.* 2009). The IC:1–143̂LL, as well as constructs corresponding to residues 1–143 (IC:1–143) and 1–40 (IC:1–40), were cloned into pET SUMO (Invitrogen). All sequences were verified by automated sequencing prior to transformation into *Escherichia coli* BL21(DE3) host cell lines for protein expression. The p150Glued construct containing residues 221–509 (p150Glued̂221–509) from *D. melanogaster* (accession number AAF49788, *Dctn1* gene) was cloned into pET15d expression vector, containing an N-terminal His₆ tag.

Cells were grown in LB or minimal media at 37 °C to an A₆₀₀ nm of ~0.6. Isotopically labeled ¹⁵N and ¹³C,¹⁵N proteins for NMR studies were prepared using published protocols (Nyarko *et al.* 2011, Makokha *et al.* 2004). Protein expression and purification under native conditions were performed as described previously (Nyarko *et al.* 2004, Barbar *et al.* 2001). The His₆ tag was cleaved from all IC proteins prepared for NMR and isothermal titration calorimetry (ITC) studies using Factor Xa protease (Novagen), and the SUMO tag was cleaved from all SUMO constructs using SUMO protease (Cornell University). Final purification of all proteins was performed via size-exclusion chromatography on a Superdex™ 75 (or Superdex™ 200 in the case of p150Glued̂221–509) (16:60) gel filtration column (GE Healthcare) with a running buffer of 50 mM sodium phosphate (pH 7.3) with 0.2 M sodium sulfate and 1 mM NaN₃.

For NMR experiments, protein concentrations were determined from sequence-based calculated molar extinction coefficients at 280 nm (IC:1–143 and IC:1–143̂LL, 2,980 M⁻¹ cm⁻¹; LC8, 14,565 M⁻¹ cm⁻¹; and p150Glued̂221–509, 1,490 M⁻¹ cm⁻¹). For ITC experiments, protein concentrations were determined using reversed-phase HPLC with detection at 214 nm and calculation of peak intensity; sequence-based calculated molar extinction coefficients at 214 nm are: IC:1–143, 166,840 M⁻¹ cm⁻¹; IC:1–87, 90,667 M⁻¹ cm⁻¹; IC:1–40, 40,296 M⁻¹ cm⁻¹; and p150Glued̂221–509, 392,888 M⁻¹ cm⁻¹ (Kuipers *et al.* 2007).
Figure 3.1. IC constructs and isoforms used in this study. (A) Sequence alignment between fruit fly and rat IC. Top sequence is D. melanogaster (Dm, AF 263371.1), middle and bottom sequences are R. norvegicus IC-1A (IC-1A, NP 062107.1) and IC-2C (IC-2C, AAA 89165.1) isoforms, respectively. Identical residues (*), conserved residues (:), and semi-conservative substitutions (.) are designated below the aligned sequences. Predicted coiled-coil regions for all sequences are shown in boldface. The binding sites for Tctex1 and LC8 are highlighted in light and dark gray, respectively. Regions 1 and 2 of the p150 Glued binding interface are designated above the sequence, as are Linker 1 (abbreviated ‘L1’) and Linker 2. (B) IC constructs used in this study. Also highlighted are regions of predicted secondary structure (helices as cylinders and β-sheet as arrow) (Cole et al. 2008), coiled-coils (McDonnell et al. 2006) (double black bars, ‘ICC’), regions predicted by DISOPRED to be disordered (Ward et al. 2004b) (raised black bars), and binding sites of light chains Tctex1 (light gray) and LC8 (dark gray).
NMR Spectroscopy.

Protein concentrations were 0.5–1 mM for IC:1–143, IC:1–40, IC:1–87, and IC:1–143 LL and up to 2 mM for p150^Gliaed_{221-509} and LC8 in 10 mM sodium phosphate (pH 6.5) with 50 mM NaCl, 1 mM NaN_3, 10% H2O, a mixture of protease inhibitors (Roche Applied Science), and 1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid. Chemical shifts were referenced with internal 2,2-dimethyl-2-silapentane-5-sulfonic acid (Markley et. al. 1998). NMR spectra were collected on a 600 MHz Bruker DRX spectrometer. Triple resonance experiments CBCA(CO)NH (Grzesiek et. al. 1992), HNCA (Wittekind et. al. 1993), and (H)CC(CO)NH (Grzesiek et. al. 1993) were recorded with 1024(H) and at least 48(C) and 40(N) points. An HNN (Panchal et. al. 2001) experiment optimized for disordered proteins was also collected.

The three-dimensional ^1H–^15N NOESY–HSQC experiment was acquired with a mixing time of 0.20 s for IC:1–40 and IC:1–143 at 5 °C. ^3J(HN–Hα) coupling constants for IC:1–40 and IC:1–143 at 5 °C were obtained from a three-dimensional HNHA (Vuister et. al. 1993a) experiment with a dephasing/rephasing delay of 13.05 ms. T2 relaxation data were determined from experiments (Farrow et. al. 1994) recorded for IC:1–143 at 5 and 20 °C and on IC:1–143/p150^Gliaed at 5 °C and IC:1–143 LL/15N-p150^Gliaed/LC8 at 20 °C with a 1.7-s recycle delay and relaxation delay times of 15.44, 30.88, 46.32, 61.76, 77.20, 92.64, and 123.52 ms, and with at least one redundant data point to aid estimation of experimental error. Steady-state ^1H–^15N heteronuclear NOEs (Farrow et. al. 1994) were recorded at 20 °C for both apo- and p150^Gliaed–bound IC:1–143, using a 3-s period of saturation and an additional delay of 1.5 s. Hydrogen/hydrogen exchange spectra were collected for apo- and p150^Gliaed–bound IC:1–143 at 5 °C using the CLEANEX-PM-FHSQC pulse sequence (Hwang et. al. 1998) with a 20-ms mixing time.

Titration between ^15N-labeled IC:1–143 and unlabeled p150^Gliaed_{221-509} was conducted at 5 °C with increasing IC:p150^Gliaed molar ratios up to a 1:2.8 ratio and recording a series of ^1H–^15N HSQC spectra using echo-anti-echo phase discrimination of 256 increments with 1024 points. The spectrum collected for the 1:2.8 ratio evidenced no significant change from that of the 1:2 ratio, suggesting saturation of binding. ^1H–^15N TROSY–HSQC spectra were also collected in tandem during titration but did not improve the spectra. Sample quality was monitored with SDS-PAGE before, during, and after the titration process. The quaternary IC:1–143 LL•p150^Gliaed•LC8 complex was prepared with molar ratios of 2 (p150^Gliaed) and 4 (LC8)
relative to IC:1–143LL, and NMR data were collected at 20 °C.

**NMR Data Analysis.**

NMR spectra were processed with NMRPipe (Delaglio *et al.* 1995) and analyzed using NMRView (Johnson 2004). Sequential assignments for IC:1–143 were based on HNCA, HNCA,CB, and CBCA(CO)NH spectra using Burrow-Owl (Benison *et al.* 2007a) with additional analysis of the (H)CC(CO)NH-TOCSY and HNN spectra. Secondary chemical shifts (ΔδCα and ΔδCβ) were calculated relative to random coil shifts corrected for temperature, pH, and for primary sequence (Kjaergaard *et al.* 2011a). For all dynamics experiments, peak intensities were measured as the peak height at the highest point and the associated error taken to be the spectral baseline noise. T2 and NOE values were calculated as described previously (Benison *et al.* 2006). For titration experiments, peak intensities were measured as peak volumes. To account for differences in concentration across the titration series, a normalization factor was determined from the peak volume of residue 143, which is presumably not altered by p150Glued binding. Changes in peak intensities Ibound/Ifree were calculated as the ratio between the peak volumes in spectra of the complex and apo-IC:1–143. For CLEANEX-PM data, peak intensities were normalized as the ratio to the peak intensity of residue 137.

**Isothermal Titration Calorimetry.**

IC:1–40, IC:1–87, IC:1–143, and p150Glued221-509 were prepared in a buffer containing 50 mM sodium phosphate (pH 7.5) with 50 mM sodium chloride and 0.5 mM NaN3. Binding thermodynamics were determined using a VP-ITC isothermal titration calorimeter (MicroCal, Northampton, MA) at 25 °C with p150Glued in the cell and the IC constructs in the syringe, using cell/syringe concentrations of 23.6 μM (31.7 μM for the titration with IC:1–40), 0.40 mM for p150Glued and each IC construct. Data were processed using Origin 7.0 (OriginLab Corp., Northampton, MA). The average of the last 2–4 enthalpies of injection were subtracted from the binding data prior to fitting. The IC:1–87 and IC:1–143 constructs bind p150Glued with a stoichiometry (n) of 0.99 ± 0.02. Data were fit to a single-site binding model, with small deviations from the best fit. Data reported are for experiments performed in duplicates. Error estimates are based upon deviations from the theoretical best fit, with duplicates yielding similar parameter values and associated uncertainties.
**Sequence Analysis.**

Sequence alignment between the *D. melanogaster* IC isoform 2 (accession number AF263371.1) and *R. norvegicus* IC-1A and IC-2C isoforms (accession numbers NP 062107.1 and AAA89165.1, respectively) was performed with manual adjustment to achieve the alignment in the “serine-rich” region (Nurminsky *et. al.* 1998). The Jpred 3 (Cole *et. al.* 2008) and Paircoil2 (McDonnell *et. al.* 2006) programs were used to predict secondary structure and coiled-coil propensities, respectively. The DISOPRED (Ward *et. al.* 2004b) program was used for disorder prediction.

**Results**

**Resonance Assignments and Secondary Structure of IC:1–143.**

*D. melanogaster* dynein intermediate chain IC:1–143 containing residues 1–143 includes the LC8 (residues 126–138) and Tctex1 (residues 110–122) recognition sequences (Figure 3.1). IC:1–143 is predicted to contain two segments of α-helical secondary structure, encompassing residues 3–36 and 49–59 (the second segment corresponding to residues 53–64 and 54–65 in *R. norvegicus* IC-1A and IC-2C isoforms, respectively), a β-strand in the light chains binding site, and disorder in the remainder of the sequence (Figure 3.1b). IC:1–143 also includes a serine-rich region (Nurminsky *et. al.* 1998), corresponding to residues 71–89 and 97–114 in *D. melanogaster* and the *R. norvegicus* IC-1A isoform, respectively. The ICs in both species contain a predicted coiled-coil at the N terminus, spanning residues 2–35 in *D. melanogaster* and 1–44 and 1–52 in *R. norvegicus* IC-1A and IC-2C, respectively (Figure 3.1a).

The $^1$H–$^{15}$N HSQC spectrum of IC:1–143 at 5 °C (Figure 3.2) exhibits a narrow 7.7 to 8.9 ppm amide chemical shift range, suggesting a predominantly disordered or helical structure. Because of the high sequence redundancy, a “divide-and-conquer” approach to chemical shift assignments was necessary, whereby smaller (and sequentially overlapping) constructs, IC:1–40 and IC:1–87, were assigned to identify and verify residue assignments in IC:1–143. Assignments were further aided by comparison with chemical shifts for IC:84–143 collected at 20 °C (Benison *et. al.* 2006). Backbone amide assignments at 5 °C were completed for 131 of the 138 nonproline residues.

Structural characterization of IC:1–143 by NMR spectroscopy included determination of $^3$J(HN–Hα) scalar coupling constants, analysis of sequential amide-amide NOEs, and measurement of secondary chemical shifts (Figure 3.3). Excluded residues correspond to
Figure 3.2. Resonance assignments of IC:1–143. Two-dimensional $^1$H–$^{15}$N HSQC spectrum of 0.6 mM $^{15}$N IC:1–143 in 10 mM sodium phosphate buffer (pH 6.5), 50 mM NaCl, and 1 mM NaN$_3$ recorded at 5 °C. Peaks are labeled according to their backbone resonance assignments. Several peaks corresponding to the N-terminal 40 residues are indicated; the rest are demonstrated through an overlay of the IC:1–40 HSQC (red) on that of IC:1–143 (black).
Figure 3.3. Secondary structure of IC:1–143. $^3J$-coupling constants are grouped as < 5 Hz (open circles) for helical structure, between 6 and 8 Hz (filled black circles) for random coil, and > 8 Hz (filled gray circles). Sequential amide-amide connectivities are shown short- and medium-range NOEs. Secondary chemical shifts for $C_\alpha$ and $C_\beta$ atoms are presented in bar plot format below the inter-residue NOE connectivities. The secondary structure estimated from $^3J$-coupling values and secondary chemical shifts is plotted above the sequence and includes the following: helices (solid cylinders), nascent helix (outlined cylinder), and random coil conformation (dotted line).
overlapped peaks or to peaks too weak for reliable measurement. The majority of residues included in the analysis exhibit $^3J$-coupling values in the 6–8 Hz range, indicative of disordered conformations; however, there also exist contiguous stretches of residues (3–9, 11–23, and 27–37) near the N terminus with $^3J$ scalar coupling values of less than 5 Hz, indicative of helical structure. Further suggestive of helical structure in this region is the presence of several $d_{NN}(i+1)$ NOEs spanning residues 2–4, 31–36, 37–40, and $d_{NN}(i+2)$ NOEs between residues 32 and 34 (Figure 3.3). Several additional $d_{NN}$ NOEs were present but not unambiguously assigned due to the extreme spectral overlap. No amide-amide NOEs were observed for resolved peaks outside of residues 1–40.

Secondary chemical shifts, sensitive indicators of protein secondary structure, were calculated for the $^{13}C^\alpha$ and $^{13}C^\beta$ nuclei from the difference between the observed and sequence-corrected random coil chemical shifts (Kjaergaard et al. 2011a). Their difference ($\Delta C^\alpha - \Delta C^\beta$) is presented for all assigned resonances (Figure 3.3). The majority of residues show only small deviations from zero, suggestive of random coil conformations. A large positive difference, indicative of helical conformations (Wishart et al. 1991), was observed in stretches spanning residues 3–23 and 26–37 (the apparent punctuation arises from a lack of chemical shift assignments for residues Arg-24 and Arg-25). An additional stretch of lower average positive values was observed for residues 47–60. The predominance of disordered segments with regions of helicity localized to the N-terminus matches the sequence-based structure prediction (Figure 3.1) and is consistent with the CD-detected (Nyarko et al. 2004) structural content.

**Dynamics of IC:1–143.**

Significant deviation from random coil behavior in the nanosecond-picosecond time scale is reflected in the $^1H-^{15}N$ steady-state heteronuclear NOEs recorded at 20 °C (Figure 3.4a); this temperature affords the best contrast in dynamic behavior between different regions of the protein. As expected for a primarily disordered protein, the majority of residues exhibit negative heteronuclear NOE values with larger negative values at the C-terminus. Residues 1–41, however, have an average heteronuclear NOE value of 0.42, whereas residues 48–61 have an average value of 0.16 (Figure 3.4a). The sign and magnitude of the NOE values within these two regions are indicative of some ordered structure.

Backbone dynamics, also inferred from $T_2$ experiments collected at 5 °C (Figure 3.4b), allow motions on the microsecond–millisecond time scale to be probed. Observed $T_2$ values
Figure 3.4. Dynamics of $^{15}$N-labeled IC:1–143 in its apo (A–C) and p150$^{\text{Glued}}$-bound states (D–F). Steady-state $^1$H–$^{15}$N heteronuclear NOE values shown as $I_{\text{sat}}/I_{\text{unsat}}$ for $^{15}$N IC:1–143 in its apo (A) and p150$^{\text{Glued}}_{221-509}$-bound (D) states at pH 6.5 and 20 °C. Average steady-state heteronuclear NOE values calculated for the assigned N-terminal 41 residues and for residues 48–61 are plotted as dotted lines for those regions (A). NOE values lower than negative 1.5 corresponding to the C-terminal residues were truncated in these plots. $T_2$ relaxation rates and CLEANEX-PM relative cross-peak intensities were measured at pH 6.5 and 5 °C for apo (B and C) and p150$^{\text{Glued}}_{221-509}$-bound $^{15}$N IC:1–143 (E and F), respectively. $T_2$ values greater than 0.4 s corresponding to the C-termini were truncated in these plots (B and E). CLEANEX-PM peak intensities (C and F) were recorded with a 20-ms mixing time. Dashed lines are drawn through the plots (A–F) to facilitate visual comparison.
Figure 3.4. (Continued)
range from 0.013 to 0.57 s with an overall average value of 0.114 s. The broad distribution of $T_2$ values is indicative of nonrandom structure in intermediate exchange on the NMR time scale, particularly so for residues evincing the lowest $T_2$ values, such as residues 1–40 (average $T_2$ value of 0.046 s). The lower $T_2$ values in this region are consistent with the increased peak broadness and nebulosity apparent for residues 1–40 in relation to the other peaks (Figure 3.2 and right insets therein). Overall broader peaks are observed for NMR spectra of the IC:1–40 construct alone (data not shown), indicating that exchange broadening is characteristic of this region, independent of the size of the protein. Further information on the conformational ensemble of IC:1–143 was garnered from amide hydrogen/hydrogen exchange experiments (Figure 3.4c). For a 20-ms mixing time, numerous residues in IC:1–143 exhibit strong peak intensities, indicating efficient exchange with the solvent and a lack of ordered structure on this time scale. There are also numerous residues with peaks apparently missing from the CLEANEX-PM spectrum, including residues 4–40 and 48–58, 61–66, and 69–70, indicating a relatively more protected local environment within these regions.

Mapping the IC Binding Interface with p150Glued.

To identify the IC:1–143 interface with p150Glued, an NMR titration experiment was performed wherein $^1$H–$^{15}$N HSQC spectra of $^{15}$N-labeled IC:1–143 were recorded with stepwise addition of unlabeled p150Glued$_{221-509}$ (Figure 3.5). A superimposition of $^1$H–$^{15}$N HSQC spectra of $^{15}$N-labeled IC:1–143, collected in the absence and presence of a saturating amount of unlabeled p150Glued$_{221-509}$ (Figure 3.5a), reveals no novel peaks for IC:1–143 residues in its p150Glued-bound conformation. Rather than moving in chemical shift space, numerous peaks for IC:1–143 were significantly attenuated, presumably due to intermediate chemical exchange effected by the binding association with p150Glued, leading to significant peak broadening and apparent disappearance of the peaks from the spectrum. Of particular interest is the pattern of peak disappearances over the course of the titration. The first set of peaks to disappear are for residues 1–41 (black peaks in Figure 3.5,b-d), followed by a gradual disappearance of peaks for residues 46–75 with increasing p150Glued concentrations (Figure 3.5,b-d, e.g. red, green, and attenuated purple peaks). Collectively, the residue 1–41 peaks (Region 1, Figure 3.5e) are the most strongly attenuated (an average of greater than 90%). Overall, the residue 46–75 peaks (Region 2, Figure 3.5e) are attenuated by an average of 72%, with the residue 48–66 peaks more strongly attenuated (80% on average) than the remaining peaks (residues 67–75, 57% on average), suggesting that the binding behavior is different for these two segments of Region 2.
Figure 3.5. Titration of $^{15}$N-labeled IC:1–143 with p150$^{Glued}_{221-509}$. (A) Overlay of $^1$H–$^{15}$N HSQC spectra of IC:1–143 (black) and p150$^{Glued}_{221-509}$-bound IC:1–143 (purple). Several peaks are labeled whose relative peak volumes are attenuated by more than 70% in the bound state; peaks corresponding to residues 84, 134, and 143 (undiminished relative peak volumes) are also indicated for comparison. The spectra were recorded at pH 6.5 and 5 °C with ~0.6 mM $^{15}$N IC:1–143 in 10 mM sodium phosphate buffer and greater than 2-fold excess of p150$^{Glued}_{221-509}$ in the case of the fully bound state (purple). Portions of HSQC spectra (B–D) show selected sets of peaks during the titration process. Peaks in these overlaid spectra are colored corresponding to $^{15}$N IC:1–143 with 0 equivalents of p150$^{Glued}_{221-509}$ (black), two subsequent additions of p150$^{Glued}_{221-509}$ (red and green, in that order), and greater than 2 equivalents (purple) of p150$^{Glued}_{221-509}$. (E) Numerical plot of relative integrated peak intensity ($I_{\text{bound}}/I_{\text{free}}$) versus residue number. Relative peak intensity is defined as the ratio of the integrated peak volume in the spectrum of the complex to the integrated peak volume in the spectrum of the free IC:1–143 protein. Two distinct series of peaks (labeled Region 1 and Region 2) in IC:1–143 exhibited significant attenuation upon interaction with p150$^{Glued}_{221-509}$. 
Figure 3.5. (Continued)
Dynamics of IC:1–143 Bound to p150Glaed.

To determine the effect of p150Glaed binding on the dynamics of IC:1–143, the same NMR dynamics experiments collected for the apo-state were performed for the bound state (Figure 3.4,d-f). Peak disappearance essentially rendered all residues of Region 1 and a major portion of Region 2 spectroscopically invisible, thus preventing determination of their dynamic properties in the bound state. In $^1$H–$^{15}$N steady-state heteronuclear NOE and $T_2$ spectra, peaks are visible for residues 43–47, 50, 67–69, 71–75, and the span of residues 76–143. Peaks for residues 67–69 and 71–75 (the latter part of Region 2) are somewhat weaker but of sufficient intensity for dynamics characterization. For residues 76–143, the NOE and $T_2$ values are essentially unchanged from their values in the unbound state, indicating little change in the degree of their disorder. For residues 43–47 (linking Region 1 and Region 2), as well as residues 67–69 and 71–75, the NOE and $T_2$ values are also largely unchanged and are indicative of conformational disorder on the picosecond–nanosecond time scale in the bound complex. For a 20-ms mixing time, CLEANEX-PM experiments show that the majority of residues exhibiting strong, visible peak intensities in the apo-state do so for the bound state as well. Notable exceptions include the following: residues 41–47 (excluding residue 43), which roughly correspond to termini of Regions 1 and 2; residues 59, 60, 68, and 72–75, all included within Region 2; and residues 78–81, which are located just C-terminal to Region 2. The diminished intensities are indicative of slower exchange with the solvent, suggesting that the termini and residues surrounding the binding interface become more ordered in the bound state.

Thermodynamics of p150Glaed–IC Interactions.

Isothermal titration calorimetry measurements confirm the newly determined multiregion binding footprint of p150Glaed upon IC. Data were collected for titration of p150Glaed$_{221-509}$ with three IC constructs (Figure 3.1) as follows: IC:1–143, containing both regions of the p150Glaed recognition motif determined here, as well as the smallest previously reported (King et. al. 2003) p150Glaed-binding domain of IC (residues 1–106 of isoform IC-2C in rat, corresponding to residues 1–97 in D. melanogaster); IC:1–87, containing both regions of the recognition motif for p150Glaed; and IC:1–40, containing only Region 1 of the p150Glaed binding footprint. Representative data from each titration are shown in Figure 3.6. p150Glaed$_{221-509}$ binds to both IC:1–143 and IC:1–87 with moderate affinity and similar $K_a$ values of $3.5 \pm 0.3 \mu M$ (with $\Delta H^0 = -5.0 \pm 0.2$ kcal/mol) and $3.6 \pm 0.3 \mu M$ (with $\Delta H^0 = -5.5 \pm 0.2$ kcal/mol), respectively. The titration between IC:1–40 and p150Glaed$_{221-509}$ does not reach saturation, and the data were not fit.
Figure 3.6. Representative ITC data for IC constructs binding to p150Glued. Isothermal titration calorimetry data with thermograms (top panels) and isotherms (bottom panels) shown for titration of p150Glued_{221-509} with IC:1–143 (A), IC:1–87 (B), and IC:1–40 (C). Data were collected at 25 °C in 50 mM sodium phosphate, 50 mM sodium chloride (pH 7.5). For the data collected with titrants IC:1–143 (A) and IC:1–87 (B), solid lines correspond to the nonlinear least squares fit for an A + B → AB binding model.
The weaker binding affinity for IC:1–40, inferred from the failure to reach saturation, suggests that the first 40 residues of IC are involved in only a fraction of the interactions with p150Glued, consistent with the smaller binding enthalpies of the initial injections (Figure 3.6, lower panels). The initial normalized enthalpies of injection for titrations of p150Glued with IC:1–87 and IC:1–143 are similar, although the value for IC:1–40 is approximately half that.

**Dynamics of IC:1–143 Bound to p150Glued and Light Chains.**

To investigate the dynamics of IC:1–143 in a more biologically relevant context, we conducted NMR studies on IC:1–143 bound to p150Glued221-509 and to two light chains. To simplify NMR spectra, we used a construct of IC:1–143 in which the Tctex1 binding region is replaced with an LC8 recognition sequence (e.g. IC:1–143LL). IC:1–143LL binds 2 equivalents of LC8 with higher affinity than IC:1–143 binds Tctex1 and LC8, and the structure of IC at the light chains interface is similar in both cases (Hall et. al. 2009). Noticeably, peaks corresponding to residues in the LC8-binding sites (residues 108–120 and 124–138) are absent from the spectrum of the quaternary complex (Figure 3.7a) as observed with titration of IC:84–143 with Tctex1 and LC8 (Benison et. al. 2006). As in the case of the binary complex between IC:1–143 and p150Glued221-509, a similar disappearance of peaks due to binding with p150Glued also occurs in the spectra of the quaternary complex. Although peaks corresponding to IC residues in the binding interfaces with p150Glued and the light chains are absent, numerous peaks remain with similar intensity as in the apo-form (Figure 3.7a). The mere presence of these intense peaks in the spectrum of a 150-kilodalton complex suggests that the segment connecting the p150Glued–binding site to the LC8 sites retains a considerable degree of conformational flexibility. The T2 values of the remaining peaks (Figure 3.7b) are largely unchanged in the quaternary complex, indicating that residues that are not directly at the binding interfaces with p150Glued and the light chains remain disordered on the nanosecond–picosecond time scale.

**Discussion**

In our ongoing efforts to characterize the structure, function, and regulation of the cargo-attachment domain of dynein, we have mapped the shortest IC segment necessary for binding dynactin p150Glued to IC residues 1–75 (1–101 in rat DIC1A). Mapping is based on NMR analyses, which also reveal three important structural features of IC residues involved in dynein-dynactin binding. First, the binding footprint on IC is multiregional, *i.e.* it involves two noncontiguous IC recognition sequences. Second, the IC regions bound to p150Glued are helical in
Figure 3.7. Dynamics of $^{15}$N-labeled IC:1–143 in a quaternary complex with p150$^{Glued}$ and light chains. (A) Overlay of $^1$H–$^{15}$N HSQC spectra of apo-IC:1–143 (black) and IC:1–143$_{LL}$ in complex with excess p150$^{Glued}_{221-509}$ and LC8 (green); spectra were recorded at pH 6.5 and 20 °C in 10 mM sodium phosphate buffer. Labeled peaks correspond to those whose relative peak volumes remain at 60% or greater intensity for IC:1–143$_{LL}$ in the quaternary complex. (B) T$_2$ relaxation times collected at 20 °C for apo-IC:1–143 (bars) and the remaining visible peaks of IC:1–143$_{LL}$ in complex with excess p150$^{Glued}$ and LC8 (green circles). A dashed line is drawn through the plot to facilitate visual comparison. (C) Strips from SDS-polyacrylamide gel analysis of components and of the assembled IC quaternary complex used for NMR experiments; IC:1–143$_{LL}$ (lane 2), p150$^{Glued}_{221-509}$ (lane 3), a binary complex of IC:1–143$_{LL}$ saturated with p150$^{Glued}_{221-509}$ (lane 4), LC8 (lane 5), and a quaternary complex of IC:1–143$_{LL}$ combined with saturating amounts of both p150$^{Glued}_{221-509}$ and LC8 (lane 6).
Figure 3.7. (Continued)
nature; the more N-terminal Region 1 is likely coiled-coil as inferred from spectral exchange broadening and sequence-based structure prediction of the apoprotein. Third, the intervening residues between IC regions interacting with dynactin remain disordered in the complex. Furthermore, the IC linker residues connecting the p150\textsuperscript{Glued} binding regions and the light chains recognition sequences remain disordered in the quaternary complex. We propose that the multiregion binding footprint of p150\textsuperscript{Glued} on IC, along with the interplay of order and disorder in the resulting complex, function in regulation of dynein-dynactin interactions and in cargo recognition.

Residual Structure in IC:1–143.

NMR experiments show that predominantly disordered apo-IC:1–143 contains two segments of \(\alpha\)-helical structure, namely residues 1–40 and 48–60, largely in agreement with sequence-based prediction (Figure 3.1b). Stretches of helical structure are inferred from \(^3J\)-coupling constants, \(d_{\text{NN}}\) sequential NOEs, and deviation of \(^{13}\text{C}\alpha\) and \(^{13}\text{C}\beta\) secondary chemical shifts from random coil values. These structure-based analyses are further bolstered by complementary dynamics measurements (\(^{1}H—^{15}N\) steady-state heteronuclear NOE, \(T_2\), and CLEANEX-PM) that indicate greater order in the regions of inferred \(\alpha\)-helical structure relative to the rest of the protein. Although structural and dynamics data from NMR suggest a well-formed helix for residues 1–40, residues 48–60 apparently form a nascent helix as indicated by the small positive values of secondary chemical shift differences and of steady-state heteronuclear NOEs. A model depicting these features of IC in its unbound state is presented in Figure 3.8, top.

Multiregion Binding Footprint for p150\textsuperscript{Glued}\textsubscript{221-509} on IC:1–143.

NMR titration of \(^{15}N\)-labeled IC:1–143 with unlabeled p150\textsuperscript{Glued} identifies a new binding motif, namely a two-region binding footprint involving IC residues 1–41 (Region 1) and residues 46–75 (Region 2), with no binding to the intervening linker residues (Figure 3.8, bottom). Intriguingly, both regions correspond to or contain the \(\alpha\)-helical structures identified in apo-IC:1–143, although their structure in the p150\textsuperscript{Glued}–bound state cannot be directly ascertained through NMR spectroscopy as the corresponding peaks essentially disappear from the spectra. ITC analysis demonstrates that both regions are required for full binding affinity. The requirement of both regions for significant IC–p150\textsuperscript{Glued} interaction explains the results of Vaughan and Vallee (Vaughan et. al. 1995), in which binding was observed between a fragment containing the first
Figure 3.8. Model for assembly of dynein intermediate chain with $p_{150}^{\text{Glued}}$ and light chains. The model depicts apo-IC:1–143 (top) as a primarily disordered and monomeric ensemble of conformations with one defined helical region encompassing the N-terminal 40 residues and another region of nascent helicity encompassing residues ~48–60, as determined from NMR measurements. Segments of IC for which NMR spectral characteristics are altered by $p_{150}^{\text{Glued}}$ binding are indicated in red (Region 1, residues 1–41) and purple (Region 2, residues 46–75). The helical nature of Regions 1 and 2 in bound IC (bottom) is inferred as described in the text, whereas all other regions in the bound complex are directly measured by NMR spectroscopy or X-ray crystallography. In accordance with sequence-based prediction of coiled-coil conformations in both the N-terminal $\alpha$-helical region of IC and the $p_{150}^{\text{Glued}}$ construct used in this study, the N-terminal 40 residues of IC in the bound state are depicted in a generic coiled-coil configuration. The $p_{150}^{\text{Glued}}$•IC complex is a tetramer that could be formed either from two $p_{150}^{\text{Glued}}$•IC heterodimeric coiled-coils (i.e. one chain from IC and one chain of $p_{150}^{\text{Glued}}$) or from two homodimeric coiled-coils (i.e. two chains of IC in an IC/IC coiled-coil packed against two chains of $p_{150}^{\text{Glued}}$ coiled-coil). The $p_{150}^{\text{Glued}}$ construct is represented as a box with a striped border. In the IC-bound state, dotted lines in the box indicate that $p_{150}^{\text{Glued}}$ does not interact with IC residues 42–45 (‘Linker 1’). Residues in Linker 1 and Linker 2 (residues 76–106) are disordered, as indicated by NMR (Figure 3.7). The IC•Tctex1•LC8 subcomplex portion of the model is from a crystal structure (Protein Data Bank entry 3FM7; Hall et. al. 2010), with Tctex1 (yellow) and LC8 dimers (green), and the corresponding IC segments in gray (one subunit light gray and the other dark gray). All structures, including apo-Tctex1 (Protein Data Bank entry 1YGT) and apo-LC8 (Protein Data Bank entry 3BRI) were generated using PyMOL (DeLano 2002).
123 residues of rat IC-1A and rat p150Glued, whereas binding was not observed between p150Glued and a fragment of rat IC-1A containing only the first 61 residues (e.g. lacking Region 2), nor with a truncation mutant of IC in which the first 61 residues had been deleted (e.g. lacking Region 1).

**Structure and Dynamics of Assembled IC.**

Our interpretive model for IC in its assembled state with p150Glued and light chains is based upon a combination of x-ray crystallography data, structural and dynamic information from NMR spectroscopy, as well as sequence-based prediction of structural propensity. Both helices of unbound IC (Figure 3.8, top) are included in the multiregion binding footprint of p150Glued, with the first consisting of the entirety of Region 1 whose associated peaks are first to disappear during NMR titration. The significant peak broadening upon binding to p150Glued is attributed to multiple exchange processes possibly including intermediate exchange between bound and free states of IC, as well as exchange between helical and coiled-coil conformations in an oligomerization process that results in a tetrameric p150Glued•IC complex. Evidence from NMR spectroscopy (analysis of identical HSQC spectra for increasingly dilute samples of IC:1–40) as well as analytical ultracentrifugation studies on IC:1–289 (Makokha et. al. 2002) indicate a predominantly monomeric state for apo-IC. The proposed coiled-coil conformation for Region 1 in the bound state derives from the complex exchange processes and sequence-based prediction of a coiled-coil in both IC (residues 1–35, Figure 3.1b) and the p150Glued construct used in this study; the coiled-coil assemblage depicted in the model (Figure 3.8) could represent either an IC/IC coiled-coil packed on a p150Glued/p150Glued coiled-coil (the latter not shown in the model) or two p150Glued/IC hetero-coiled-coils, each formed from one chain of IC and one chain of p150Glued.

The second, nascent helix in apo-IC is contained within the p150Glued recognition motif in Region 2, with associated peaks diminishing more gradually and exhibiting less attenuation of intensity than those of Region 1. The less dramatic disappearance of Region 2 peaks is attributed to less complex exchange broadening processes, likely chemical exchange between the free and p150Glued–bound states of IC as well as structural fluctuation between nascent and fully formed helix within IC. Thus, the nascent helical structure depicted in the model for apo-IC is proposed to persist and perhaps stabilize in the bound state. For residues ~67–75 of Region 2, less attenuation of peak intensity suggests even simpler exchange processes, likely between free and bound IC with minimal structural changes. The retained disorder in this part of Region 2 in
bound p150\textsuperscript{Glued}–IC is confirmed by dynamics experiments. The attribution of peak disappearance primarily to exchange broadening (rather than to the larger size and increased rotational correlation time of the complex with p150\textsuperscript{Glued}) is supported by the observation that titration of IC with \textsuperscript{2}H-labeled p150\textsuperscript{Glued}\textsubscript{221-509} (data not shown) does not improve spectra of p150\textsuperscript{Glued}–bound IC.

Between Regions 1 and 2 of the p150\textsuperscript{Glued}–binding motif in IC is Linker 1, a short intervening segment (residues 42–45) that remains disordered in the bound state. It is possible that Linker 1 turns to allow Region 2 to pack against Region 1, a structure that is an alternative to the model presented in Figure 3.8. In this scenario, the disappearance of peaks in Region 2 could be attributed to its packing against Region 1 and not to direct interaction with p150\textsuperscript{Glued}; our data cannot differentiate between the two possibilities. Linker 2 (residues 76–107), connecting the C-terminal end of the p150\textsuperscript{Glued} recognition sequences with the N-terminal end of the light chains binding domain, also remains disordered in assembled IC. IC residues at the light chains binding interfaces are absent from spectra of the quaternary complex (Figure 3.7a) and so cannot be probed in this study, but the structure of IC bound to the light chains is known to assume β-strand structure (Williams et al. 2007, Benison et al. 2007b, Hall et al. 2009), depicted accordingly in the model (Figure 3.8).

**Functional Implications of Disorder in Assembled IC.**

Assembled IC is a dynamic complex, integrating both order and disorder, with IC remaining largely disordered and flexible in regions outside the segments that form an interface with any of the binding partners. At the IC–p150\textsuperscript{Glued} interface, a multiregion motif suggests distinct regional functions with three-dimensionally packed and conserved Region 1 primarily for recognition and partially disordered and variable Region 2 primarily for regulation by post-translational modification and/or alternative splicing (Figure 3.1a). In this vein, phosphorylation of residues in the serine-rich cluster at the C-terminal end of Region 2 (Vaughan et al. 2001) has been proposed as a potential regulator of IC activity, including its interaction with p150\textsuperscript{Glued} (Vaughan et al. 1995, Nurminsky et al. 1998). Partial disorder in Region 2 would allow access for regulatory chemical modifications of residues in the serine-rich region.

The disorder in Linker 2 may present accessible binding interfaces for other IC-binding partners, possibly including NudC/CL (Zhou et al. 2006), NudE/EL, Spindly, or ZW10 (Kardon et al. 2009b). Linker 2 is also subject to alternative splicing (Figure 3.1a), which produces variability in linker length that could modulate IC binding affinity for any of these binding
partners. Furthermore, the IC–p150Glued interaction may be enhanced by light chains binding through the “multivalency effect” (Hall et al. 2009). Association with a bivalent ligand such as Tctex1 or LC8 essentially increases the local effective concentrations of proximate IC regions, with the extent of binding enhancement for another ligand such as p150Glued depending upon the length of Linker 2; a shorter linker is expected to result in greater binding enhancement than a longer one. The potential modulation of IC–p150Glued interaction by processes related to IC Linker 2 disorder is consistent with the emerging theme of assembled IC as a flexible polybivalent scaffold whose affinities for its multitude of binding partners are finely tuned to provide versatility and reversibility in response to changing conditions in the cellular environment (Hall et al. 2009, Nyarko et al. 2004, Benison et al. 2006).

In summary we propose that a multiregion IC binding interface and disorder of the linkers between binding interfaces together provide a basis for elegant and multifaceted regulation of binding between IC and p150Glued, and thus of the association of dynein with one of its key regulators—Dynactin. The linker disorder retained throughout assembled polybivalent IC may also maintain accessibility of segments involved in cargo recognition.

Acknowledgements

The authors thank Dr. Gregory C. Benison for assistance with NMR experiment setup and Professor Clare Woodward for critical reading of the manuscript, as well as members of the Barbar lab for valuable discussion.
Chapter 4

The Role of Residual Structure in Dynein Intermediate Chain in Complex Assembly and Regulation*

Jessica L. Morgan, Valéry Ozenne, Malene Ringkjøbing Jensen, Martin Blackledge, and Elisar Barbar
Abstract

Cytoplasmic dynein is a microtubule-associated motor protein complex and the primary motor responsible for retrograde transport in the cell. The 1.2 MDa multi-subunit dynein complex is comprised of multiple protein subunits, including two heavy chains, two light intermediate chains, two intermediate chains (ICs), and the three homodimeric light chains Tctex1, LC8, and LC7. The N-terminal half of dynein IC (‘N–IC’) is central to the ‘cargo attachment complex’ and is a prototypical example of the intrinsically disordered protein (IDP) class, serving as a primarily disordered polybivalent molecular scaffold for its numerous binding partners, itself becoming more ordered upon binding interaction. In this study, experimentally measured NMR chemical shifts and a collection of residual dipolar coupling (RDC) values were combined with ensemble generation and selection algorithms to create a representative ensemble of the region of IC that contains the binding sites for light chains Tctex1 and LC8. Ensemble calculations evidence deviation from random statistical coil behavior with higher populations of polyproline II conformation being sampled by the light chains binding segments of IC; in the light chains bound state, these same segments of IC adopt β-strand structure. Further deviation of N–IC from random coil behavior is demonstrated by NMR paramagnetic relaxation enhancement (PRE) experiments, showing transient medium- and long-range contacts within IC itself, including interactions between two non-contiguous helical regions previously identified in IC to contain the binding sites for dynein regulators dynactin and NudE. Intra-molecular contact between these two helical regions is significantly attenuated when IC is bound to NudE. While the flexibility in N-IC may underlie its ability to serve as a binding platform for numerous partners, deviations of this protein from random-coil behavior provide insight into potential mechanisms for regulating these binding interactions and thus the overall behavior of the dynein motor.

* Data to be published as supplemental material in this manuscript are included in Appendix 1.

1The abbreviations used are: IC, 74-kDa dynein intermediate chain corresponding to gene Cdic2b; N–IC, IC residues 1–289; IC111, IC residues 84–143; IC114, IC residues 1–143 with residues 111–120 (SVYNVQATNI) replaced with residues 126–135 (LVYTKQTQTT); p150Glued, the 150-kDa polypeptide corresponding to the Glued gene; NudE, isoform A of the DnudE gene; LC8, the 10-kDa dynein light chain corresponding to gene Cdlc2; Tctex1, the 12-kDa dynein light chain corresponding to gene Dtc90F; SUMO, small ubiquitin-related modifier; IDP, intrinsically disordered protein; HSQC, heteronuclear single-quantum coherence; RDC, residual dipolar coupling; CS, chemical shift; PRE, paramagnetic relaxation enhancement; FM, flexible-mecanico; PPII, polyproline II or ‘poly-L-proline type II’ a.k.a. ‘left-handed 31-helix’; MTSL, (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-methanethiosulfonate; DSS, 2,2-dimethyl-2-silapentane-5 sulfonyl acid.
Introduction

Cytoplasmic dynein is a microtubule-associated macromolecular motor complex and the primary motor responsible for retrograde (minus-end directed) transport in the cell, being involved in such critical cellular processes as assembly and positioning of the mitotic spindle, chromosome segregation, nuclear migration, orientation and transport of organelles, as well as establishment of cell polarity (reviewed in Vallee et al. 2004). The ~1.2 MDa dynein complex is comprised of multiple protein subunits, including two heavy chains (which include the sites for microtubule binding and ATPase activity), two light intermediate chains, two intermediate chains (ICs), and the three homodimeric light chains Tctex1, LC8, and LC7 (Pfister et al. 2006).

Dynein IC (642 residues in Drosophila melanogaster) consists of two structurally and functionally distinct subdomains (Figure 4.1): a long, primarily disordered domain (‘N–IC’, residues 1–289) that includes the binding sites for the three dynein light chains (Makokha et al. 2002; Lo et al. 2001; Mok et al. 2001; Susalka et al. 2002; Hall et al. 2010), and the C-terminal domain (‘C–IC’, residues 290–642) which includes seven WD40 repeats predicted to fold into a toroidal β-propeller structure, providing sites for the binding and assembly of the heavy chains (Ma et al. 1999). N–IC is a prototypical example of the intrinsically disordered protein (IDP) class, whose members are involved in a vast array of cellular processes including molecular recognition, signal transduction, transcriptional regulation, cell cycle control, and assembly of multi-protein complexes (Dyson et al. 2005). An intrinsically disordered protein or protein region is defined by a lack of stable, well-defined three-dimensional structure, and is best described as a flexible and rapidly-interconverting ensemble of conformations lacking regular packing of protein atoms. The flexible unstructured character of IDPs facilitates their incorporation into multi-protein macromolecular assemblies (Cortese et al. 2008), as well as promiscuous binding with myriad protein partners. Indeed, dynein N–IC serves as a flexible polybivalent scaffold for the three homodimeric dynein light chains, wherein two N–IC chains bind to each homodimeric light chain Tctex1, LC8, and LC7 with the crystallographically-determined binding sites corresponding to IC residues 110–122 (Tctex1; Hall et al. 2009), 126–135 (LC8; Hall et al. 2009; Benison et al. 2007b), and 221–258 (LC7; Hall et al. 2010) in Drosophila melanogaster (Figure 4.1). In addition to these components of dynein, N–IC provides a binding platform for numerous non-dynein partners, including the p150Glued subunit of the regulatory complex and dynein cargo adaptor dynactin, the Zeste-white 10 (ZW10) subunit of the RZZ complex, and NudE/EL (reviewed in Kardon et al. 2009b), as well as several putative cargoes of the dynein motor including the protein Huntingtin (Caviston et al. 2007), Herpes
Figure 4.1. Domain organization of *D. melanogaster* IC and constructs used in this study. IC74 (*top*), the full-length 74 kDa *Drosophila melanogaster* intermediate chain containing subdomains ‘Region 1 (R1)’ and ‘Region 2 (R2)’ (residues 1–41 and 46–75, respectively), the binding sites for dynein light chains Tctex1 (110–122, *dots*), LC8 (126–135, *gray shading*), and LC7 (221–258, *diagonal stripes*), and the ‘C-IC’ domain (residues 290–642, *horizontal stripes*) containing the seven WD40 repeats predicted to form a toroidal β-propeller. The IC_TL (T for Tctex1, L for LC8) construct used in this study (IC residues 84–143, *middle*) contains the binding sites for the Tctex1 and LC8 light chains. The IC1_LL construct (IC residues 1–143, *bottom*) has residues 111–120 replaced by a second copy of the LC8 binding sequence. Numbered residues indicated below the IC1_LL schematic are those individually replaced by cysteine to create the single-mutant IC1_LL-K11C, IC1_LL-R25C, IC1_LL-S60C, IC1_LL-S84C, and IC1_LL-L108C constructs to which nitroxide spin labels were attached in this study.
Simplex Virus 1 (Ye et al. 2000), and Adenovirus (Bremner et al. 2009).

The IDP class of proteins includes members that are completely disordered, as well as those consisting of a mixture of ordered (structured) and disordered residues; regions of order or secondary structure are typically confined to relatively short regions, with the degree of secondary structure ranging from transient (or, nascent) to fully-formed. Within the predominantly disordered (Makokha et al. 2002) dynein N–IC domain, we have previously determined through protein NMR spectroscopy, that residues 1–40 exist as a well-formed α-helix and residues 48–60 exhibit nascent helicity (Morgan et al. 2011), and residues 222–231 constitute a helical IC self-association domain (Nyarko et al. 2011). Regions containing preformed or transient elements of secondary structure within an IDP frequently mediate its interactions with binding partners (Fuxreiter et al. 2004). For IDP regions with apparent complete lack of structure, as well as for those with more readily-detected residual secondary structure, the binding interaction between an IDP and target molecule is frequently concomitant with its conformational (‘disorder-to-order’) transition to a folded form within its interacting region(s) (Wright et al. 2009). In the case of dynein N–IC binding to homodimeric light chains Tctex1 and LC8, two chains of N–IC bind to both Tctex1 and LC8 simultaneously, and the 10-(LC8) or 13- (Tctex1) residue recognition sequences in each IC chain undergo a disorder-to-order transition to form extended β-strands that are each incorporated into a β-sheet at the light chain dimer interface (Benison et al. 2006; Benison et al. 2007b; Williams et al. 2007; Hall et al. 2009).

Previous solution-state spectroscopic characterization of the IC chain peptide comprising residues 84–143 (IC_{TL})—containing the binding sites for the dynein light chains Tctex1 and LC8 (Figure 4.1)—evidenced a lack of secondary structure or rigidity in this segment of IC (Benison et al. 2006). In order to determine whether there existed some small amount of residual β-strand structure in the light chains binding segments of apo IC (that might presage the definite β-strand structure that these segments adopt in the bound state), we performed a detailed analysis including application of the flexible-mecanno/ASTEROIDS approach to develop a molecular ensemble description of IC_{TL} on the basis of several different chemical shifts (CSs) and a collection of 4 different types of residual dipolar couplings (RDCs) per residue. This analysis revealed decreased conformational sampling of β-strand structure but increased polyproline II conformational sampling (relative to the random statistical coil description) in the light chains binding segments of IC.
In addition, we report the first evidence of significant transient tertiary structure within the unbound form of IC residues 1–143, garnered from highly sensitive NMR Paramagnetic Relaxation Enhancement (PRE) experiments. The N-terminal 143 residues of D. melanogaster IC (IC:1–143) constitute a ‘hotbed’ for binding activity (Figure 4.1, top), with this fragment containing the binding sites for dynein light chains Tctex1 and LC8, as well as the overlapping binding recognition sites for dynein regulator proteins dynactin p150Glued and NudE in the N-terminal half of IC:1–143 (Morgan et al. 2011; Nyarko et al. 2012). Numerous intra-chain contacts were determined in IC:1–143, the most prominent being the transient intra-molecular packing between two non-contiguous regions (‘Region 1’ and ‘Region 2’ denoted ‘R1’ and ‘R2’, respectively, in Figure 4.1) that constitute the binding recognition ‘footprint’ of dynactin p150Glued upon IC. Intriguingly, binding of NudE to its recognition site (labeled ‘R1’ in Figure 4.1) within IC leads to substantial decrease in the transient intra-molecular contact between Regions 1 and 2 and IC sampling more extended conformations within its N-terminal ~100 residues, suggesting a molecular basis for the selection of one binding partner vs. another (i.e. NudE or dynactin p150Glued) that have overlapping binding sites in IC, when both are simultaneously present in the cell.

Full understanding of how an IDP is able to execute its biological functions in the absence of a stable three-dimensional structure requires a detailed understanding and description of specific local conformational sampling and preferences of residues, as well as a description of the long-range transient tertiary structure of the molecule, and also any potential coupling between these different levels of structure. The unprecedented molecular detail revealed for the N-terminal 143 residues of the dynein intermediate chain in this study provides further insights into the functionality of this protein (i.e. molecular recognition and scaffolding assembly of binding partners) as well as details of how these binding interactions might be regulated.

Experimental Procedures

Protein Preparation.

Site-directed mutageneses of Lys-11, Arg-25, Ser-60, Ser-84, and Leu-108 to generate five IC1LL (IC residues 1–143 with the sequence LVYTQKTQTT in place of residues 111–120; Morgan et al. 2011) single cysteine mutants were performed using the Stratagene QuikChange™ kit (Agilent Technologies) following the manufacturer’s protocol. The IC1LL single cysteine mutants were cloned into pET SUMO (Invitrogen), and all sequences were verified by automated
sequencing prior to transformation into *Escherichia coli* BL21(DE3) host cells for protein expression. The SUMO tag was cleaved from all SUMO constructs using SUMO protease (Cornell University) prior to final purification. Isotopically labeled $^{15}$N and $^{13}$C, $^{15}$N proteins were prepared using published protocols (Makokha *et al.* 2004; Benison *et al.* 2006; Morgan *et al.* 2011). *D. melanogaster* LC8, nNudE (N-terminal domain of NudE, residues 1–174), and p150$^{\text{Glued}}_{221-509}$ (residues 221–509 of *Drosophila* dynactin p150$^{\text{Glued}}$, constituting most of the first coiled-coil, ‘CC1’) proteins were prepared as described previously (Morgan *et al.* 2011; Nyarko *et al.* 2012). Protein concentrations were determined using sequence-based calculated molar extinction coefficients at 280 nm (ProtParam, http://expasy.org).

**NMR Sample Preparation and Experiments.**

For the measurement of RDCs in IC$_{TL}$, $^{13}$C, $^{15}$N-labeled protein was prepared at a concentration of 0.5 mM in 10 mM sodium phosphate (pH 6.5) with 50 mM NaCl, 1 mM NaN$_3$, 10% $^2$H$_2$O, a mixture of protease inhibitors (Roche Applied Science), and 1 mM 2,2-dimethyl-2-silapentane-5 sulfonyl acid (DSS) (Markley *et al.* 1998). The protein sample was aligned in a liquid crystalline phase composed of poly-ethylene glycol (PEG) and n-octanol (Rückert *et al.* 2000). 1D$_{\text{HN-N}}$, 1D$_{\text{Ca-C}}$, 3D$_{\text{HN-Ca}}$, 1D$_{\text{Ca-Ha}}$, and 4D$_{\text{HN-Ha(-1)}}$ RDCs were obtained using 3D BEST-type HNCO and CT-HN(CO)CA experiments (Rasia *et al.* 2011) collected on a Varian 800 MHz instrument equipped with a cryogenic TXI triple-resonance probe at 20°C. The isotropic reference sample of $^{13}$C, $^{15}$N IC$_{TL}$ was prepared identically and the same 3D BEST-type HNCO and CT-HN(CO)CA experiments were performed on a Varian 600 MHz instrument equipped with a cryogenic TXI triple-resonance probe at 20°C.

$T_1$ and $T_2$ relaxation data were determined from experiments (Farrow *et al.* 1994) recorded for $^{15}$N-labeled IC$_{TL}$ at pH 6.5 and 20°C. $T_1$ experiments used a 1.4-second recycle delay, with relaxation delay times of 15.44, 30.88, 46.32, 77.2, 92.64, 108.08, 123.52, and 138.96 milliseconds. $T_2$ experiments used a 1.5-second recycle delay and relaxation delay times of 50, 100, 150, 200, 300, 500, and 1,000 milliseconds. Both $T_1$ and $T_2$ experiments were recorded with at least one redundant data point to aid estimation of experimental error. Steady-state $^1$H–$^{15}$N NOEs (Farrow *et al.* 1994) were recorded at 20°C; spectra recorded in the presence and absence of amide proton saturation were collected in an interleaved manner, and experiments with proton saturation utilized a 3 second period of saturation and additional delay of 1.5 seconds. All dynamics experiments were collected on a 600 MHz Bruker DRX spectrometer. $^3$J($^1$H$_{\text{N}}$–$^1$H$_{\alpha}$) coupling constants for apo IC$_{TL}$ at 20°C were obtained from a three-dimensional HNHA (Vuister
et. al. 1993a) experiment with a dephasing/rephasing delay of 13.05 milliseconds, collected on a 500 MHz Bruker Ascend spectrometer equipped with a room-temperature TXI probe.

To add nitroxide spin label MTSL [(1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-methanethiosulphonate, Toronto Research Chemicals] to the $^{15}$N IC$_{1LL}$ single cysteine-containing mutants, dithiothreitol (DTT) was first removed by dialysis of the protein sample into 10 mM sodium phosphate (pH 7.0), 50 mM NaCl, 1 mM NaN$_3$. Free sulfhydryl groups were then reacted with a 5-fold molar excess of MTSL at 4°C in the dark for 12–16 h. After incubation, unreacted MTSL was removed by extensive dialysis at 4°C in the dark into 10 mM sodium phosphate (pH 6.5), 50 mM NaCl, 1 mM NaN$_3$. Spin-labeled protein samples were concentrated to 0.5–0.7 mM and also contained 10% $^2$H$_2$O, a mixture of protease inhibitors (Roche Applied Science), and 1 mM DSS. Spin-labeled proteins were stored in the dark at 4°C and used within 24 h. The completeness and uniformity of spin-labeling was confirmed using MALDI-TOF mass spectrometry, and no significant mass of disulfide-linked dimeric or non-spin-labeled protein was detected. $^1$H–$^{15}$N HSQC spectra were collected at 5 or 20°C using echo-anti-echo discrimination of 256 increments with 1024 points and a 3.0-second recycle delay on a Bruker 700 MHz spectrometer with a room-temperature BBI probe. After HSQC acquisition on a sample with the spin-label in its oxidized (paramagnetic) state, a 5-fold molar excess of ascorbic acid was added from a concentrated stock (0.5 M ascorbic acid prepared in the same NMR buffer used for the protein samples, pH adjusted to 6.5) to reduce the spin-label to its diamagnetic state, without significantly altering the sample volume or pH; a second HSQC was then collected on the diamagnetic sample with all acquisition parameters remaining unchanged.

NMR-PRE experiments were collected for all five individual nitroxide spin-labeled $^{15}$N IC$_{1LL}$ single cysteine mutants at 5 and 20°C. PRE experiments were also collected for spin-labeled 0.15 mM $^{15}$N IC$_{1LL}$-R25C with 0.9 mM non-isotopically-labeled nNudE at 5°C; spin-labeled $^{15}$N IC$_{1LL}$-R25C with excess non-isotopically-labeled LC8 at 20°C; and spin-labeled 0.15 mM $^{15}$N IC$_{1LL}$-R25C with 0.9 mM and excess non-isotopically-labeled nNudE and LC8, respectively, at 20°C. NMR-PRE experiments recorded for samples of IC with nNudE were recorded at a relatively lower overall protein concentration (IC ~ 0.15 mM and nNudE ~ 0.9 mM) to avoid any contributions from nonspecific interactions due to high protein concentrations that had previously been observed for the IC–nNudE system (Nyarko et. al. 2012). As part of the assay for retained binding function of IC single-cysteine mutants with spin-label attached, $^1$H–$^{15}$N HSQC spectra were also collected for diamagnetic MTSL spin-labeled $^{15}$N IC$_{1LL}$-K11C and $^{15}$N IC$_{1LL}$-R25C each with non-isotopically-labeled p150$^{\text{Ghed}}$$^{221-509}$, and also for 0.5 mM diamagnetic
MTSL spin-labeled $^{15}$N IC1$_{LL}$-R25C with 1.5 mM non-isotopically-labeled nNudE, all collected at 5°C. The sample quality of proteins used in NMR experiments was monitored with SDS-PAGE (both reducing and non-reducing, where appropriate) before and immediately following the conduction of NMR experiments.

**NMR Data Analysis.**

NMR spectra were processed with NMRPipe (Delaglio et al. 1995) and quantitatively analyzed using NMRView (Johnson 2004). Chemical shift assignments for apo IC$_{TL}$ were as reported previously (Benison et al. 2006), with corrections in peak assignments (Morgan et al. 2011). For all dynamics experiments, peak intensities were measured as the peak height at the highest point and the associated error taken to be the spectral baseline noise. T$_1$, T$_2$, and NOE values were calculated as described previously (Benison et al. 2006). Reduced spectral density mapping was performed on the NMR dynamics data using ‘Method 2’ as described by Farrow et al., 1995 (Farrow et al. 1995) to yield values for spectral densities $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$; the micro- to millisecond chemical exchange contribution ($R_{ex}$) to the transverse relaxation rate is omitted in this analysis. To yield indication of secondary structural element content, secondary $^3J(H^N-H^\alpha)$ scalar coupling values were calculated as the difference between experimental $^3J(H^N-H^\alpha)$ scalar couplings and residue-specific random coil values (Plaxco et al. 1997). One-bond $^1J_{\text{Cala}}$ scalar couplings were obtained from the 3D BEST CT-HN(CO)CA-JCH (Rasia et al. 2011) experiment collected on the isotropic apo IC$_{TL}$ sample (above), and to yield further characterization of conformational preferences in IC$_{TL}$, secondary $^1J_{\text{Cala}}$ scalar couplings were calculated from the difference between the observed and residue-specific random coil $^1J_{\text{Cala}}$ values (Vuister et al. 1993b). Secondary chemical shifts ($\Delta\delta^{13}C^\alpha$, $\Delta\delta^{13}C^\beta$, and $\Delta\delta^1H^N$) were calculated relative to random coil chemical shifts corrected for temperature, pH, and for primary sequence (Kjaergaard et al. 2011a,b). Amino acid-specific correction factors to the random coil values for residues preceding Proline were taken from Wishart et al., 1995 (Wishart et al. 1995).

For the 3D BEST HNCO-JNH experiment (Rasia et al. 2011) used to measure one-bond amide N–H splittings, the apparent line splittings were adjusted by the scaling factor $(1 + \lambda)$ with values: $\lambda = 0.345542$ and 0.403530 for isotropic and aligned apo IC$_{TL}$, respectively. RDC values were calculated as the difference between splittings measured for the aligned and isotropic phases of IC$_{TL}$. Protein NMR backbone assignments for the IC1$_{LL}$ single-cysteine variants were based
upon previous assignments for the IC1\textsubscript{1L} construct at 5 and 20\textdegree C (Morgan \textit{et al.} 2011). NMR-PRE effects for the \textsuperscript{15}N IC1\textsubscript{1L} single-cysteine mutants were determined from the ratio of peak intensities ($I_{\text{paramagnetic}}/I_{\text{diamagnetic}}$) measured in the \textsuperscript{1}H–\textsuperscript{15}N HSQC spectra collected for the protein with the MTSL spin-label in the paramagnetic (oxidized) and diamagnetic (reduced) states. PRE ratios are reported as the average over a three-residue window. Simulated PRE profiles for pure random coil ensembles were generated for the specific amino acid sequence of each of the five single-cysteine mutants of IC1\textsubscript{1L} using \textit{flexible-meccano} with explicit modeling of MTSL spin-label mobility (Salmon \textit{et al.} 2010); parameters used in these calculations were: $\tau_C = 5$ ns, $\tau_{\text{mix}} = 10$ ms, $\omega_H = 700$ MHz, 4.0 Hz intrinsic proton linewidth, and an ensemble size of 50,000 conformers.

\textit{Generation of Descriptive Conformational Ensembles from NMR Data.}

A conformational ensemble description of the IC\textsubscript{T1} protein was generated using a recently described protocol combining Monte-Carlo-based statistical coil sampling with ensemble selection on the basis of agreement with experimental NMR data (RDCs and chemical shifts)—the \textit{flexible-meccano}/ASTEROIDS CS-RDC approach (Ozenne \textit{et al.} 2012a). A large initial pool of 60,000 conformers with the amino acid sequence of IC\textsubscript{T1} was generated using the statistical coil model proposed by \textit{flexible-meccano} (Bernado \textit{et al.} 2005; Ozenne \textit{et al.} 2012b), in which individual amino acids broadly (randomly) sample amino acid-specific backbone dihedral angle $\{\phi,\psi\}$ potential wells. The population-weighted amino acid-specific potentials are derived from a compilation of non-secondary structural elements (e.g., segments of amino acid residues that are outside of $\alpha$-helix and $\beta$-sheet structures, such as loop regions) in high-resolution X-ray crystallographic protein structures (Lovell \textit{et al.} 2003); a total of 23 different potential energy wells are employed—one for each of the 20 different amino acid types, plus 3 additional specific potentials that account for the particular backbone conformational propensities of residues that precede Proline, Prolines that precede other Prolines, and Glycines that precede Proline. The peptide chain is then constructed by using the selected $\{\phi,\psi\}$ pairs for each amino acid residue to sequentially connect peptide planes, with residue-specific exclusion volumes used to avoid steric clash between different amino acids within the same conformer chain. Side chains were added to each conformer in the initial large ensemble from a backbone-dependent rotamer library using the SCCOMP program (Eyal \textit{et al.} 2004) as described previously (Huang \textit{et al.} 2012).
Chemical shifts were calculated for each member of the ensemble using the program SPARTA (Shen et al. 2007) as previously described (Jensen et al. 2010), and secondary chemical shifts were calculated using random coil values from RefDB (Zhang et al. 2003). Theoretical RDCs were calculated for the residues of each conformer in the ensemble using PALES (Zweckstetter et al. 2000) to calculate a global alignment tensor for each conformer. A uniform scaling factor (to account for the degree of alignment) was applied to the entire predicted set of RDCs to best reproduce the experimental data. The genetic algorithm ASTEROIDS (Nodet et al. 2009) was used to select conformational sub-ensembles in agreement with the experimental data (\(1^3\)C\(\alpha\), \(1^3\)C\(\beta\), \(1^5\)N, and \(1^5\)H\(\alpha\) chemical shifts and \(1^3\)D\(\alpha\)HN, \(1^3\)D\(\alpha\)HC\(\alpha\), \(1^3\)D\(\alpha\)C\(\alpha\)-HN, \(1^3\)D\(\alpha\)C\(\alpha\)-H\(\alpha\), \(1^3\)D\(\alpha\)C\(\alpha\)-C\(\gamma\), and \(4^3\)D\(\alpha\)HN-H\(\alpha\)) RDCs). The final ensemble was obtained from generations of 100 sub-ensembles of size N=200 that underwent evolution and selection based on agreement with the experimental data, with iterations continued until convergence. Chemical shift and RDC values for each residue were calculated as the average of values over each conformer in the final ensemble. Ensemble-averaged chemical shifts and RDCs were also calculated as described above for a statistical coil ensemble of IC\(\alpha\) generated using flexible-mecano. To quantify the sampling of conformational space and \(\{\phi,\psi\}\) angle distribution for a given ensemble, Ramachandran space is divided into four quadrants, defined as follows: \(\alpha_L\), \(\{\phi > 0^\circ\}\); \(\alpha_R\), \(\{\phi < 0^\circ, -120^\circ < \psi < 50^\circ\}\); \(\beta_p\), \(\{-100^\circ < \phi < 0^\circ, \psi > 50^\circ\ \text{or} \psi < -120^\circ\}\); \(\beta_s\), \(\{-180^\circ < \phi < -100^\circ, \psi > 50^\circ \ \text{or} \psi < -120^\circ\}\). The populations of these quadrants are denoted as \(p(\alpha_L)\), \(p(\alpha_R)\), \(p(\beta_p)\), and \(p(\beta_s)\), respectively.

**Sequence Analysis.**

For the IC\(\alpha\) construct (D. melanogaster IC isoform 2, accession number AF 263371.1) a sequenced-based “bulkiness” function was calculated as defined by Cho et al., 2007 (Cho et al. 2007), with bulkiness values averaged over a five-residue window size, and the resultant multiplied by the bell-shaped function (Schwalbe et al. 1997) with persistence length = 7, taking into account the fact that Serine84 was not the true N-terminal residue in the IC\(\alpha\) construct used for NMR dynamics experiments.

**Results**

**Experimental Characterization of Residual Secondary Structure in IC\(\alpha\).**

Structural characterization via NMR spectroscopy was performed on D. melanogaster dynein intermediate chain, residues 84–143 (Figure 4.1, middle). This construct (IC\(\alpha\)) is
primarily disordered in its apo state (Benison et al. 2006) and includes the LC8 (IC residues 126–135) and Tctex1 (IC residues 110–122) recognition sequences. Characterization of the residual structure and deviations from pure random coil behavior in IC\textsubscript{TTL} included measurement of residual dipolar couplings (RDCs), secondary chemical shifts of selected nuclei, and $^3J(N-H^{\alpha})$ and $^1J(C^{\alpha}-H^{\alpha})$ scalar coupling constants.

For unfolded proteins partially oriented in aligning media, residual dipolar couplings (RDCs) measured between different pairs of nuclei are sensitive probes of local conformational propensity and residual structure (Jensen et al. 2009). $^1D_{HN}$ (a.k.a. $^1D_{N-HN}$), $^1D_{C=H}$, $^1D_{C=C}$, and $^4D_{HN-H(-1)}$ RDCs were measured for IC\textsubscript{TTL} aligned in liquid crystalline polyethylene glycol/alcohol lyotropic phase. $^1D_{HN}$ RDCs from IC\textsubscript{TTL} (Figure 4.2a, left ordinate) globally follow the expected distribution of values for an unfolded protein or a random coil in sterically-aligning media, with predominately negative values throughout and a tapering of values toward zero at the chain termini, consistent with the bell-like shape and distribution predicted from random flight chain models (Obolensky et al. 2007). However, a sign inversion (e.g. positive RDC values) occurs for certain residues in IC\textsubscript{TTL}, such as the span of residues 91–93; such positive RDC values can be qualitatively attributed to or, interpreted as, the presence of local helical or turn motifs (Jensen et al. 2009). In addition, relatively larger negative amplitude RDCs occur in the span of residues 120–134 than in the rest of the protein; larger negative RDCs are associated with locally more extended conformations, such as $\beta$-strand or polyproline II structure. Deviations in this region from values predicted for a random coil (from FM) are found in the other RDCs measured ($^1D_{C=H}$, $^1D_{C=C}$, and $^4D_{HN-H(-1)}$, Figure A1.1).

The NMR secondary chemical shifts (deviation of experimental chemical shifts from their expected random coil values, $\Delta\delta = \delta_{\text{experimental}} - \delta_{\text{random coil}}$) of certain nuclei are known to be highly correlated with $\phi$ and $\psi$ backbone dihedral angles and thus with secondary structure in proteins. Positive deviations in the secondary chemical shifts of $^{13}C^{\alpha}$ nuclei are correlated with a propensity for alpha-helical conformations, while negative deviations are associated with extended $\beta$-strand-like structure (Wishart et al. 1991); positive deviations in the secondary chemical shifts of $^{13}C^{\beta}$ nuclei are correlated with extended $\beta$-strand-like structure, while negative deviations are associated with a propensity for alpha-helical conformations (discussed in...
Figure 4.2. Residual secondary structure in IC<sub>TL</sub>. Structural information from experimental data collected for apo IC<sub>TL</sub> at pH 6.5 and 20°C. (A) $^{1}$D$_{\text{NH}}$ RDCs measured in liquid crystalline aligning media (grey bars, left-hand ordinate) compared to secondary $^{3}$J(H$^{N}$–H$^{\alpha}$)-coupling values (thin black bars, right-hand ordinate). Residues excluded from this presentation correspond to significantly overlapped peaks or to Prolines. (B) Deviations of the $\delta$C$^{\alpha}$ (grey bars, left ordinate) and ($\delta$C$^{\alpha}$ – $\delta$C$^{\beta}$) (thin black bars, right ordinate) chemical shifts from values expected for a random coil ensemble of conformations. (C) Plot of the differences between the observed and residue-specific random coil values for the $^{1}$J$_{\text{C\textalpha}H_{\text{\alpha}}}$ scalar coupling. Above, the crystallographically-determined regions of IC<sub>TL</sub> that bind to dynein light chains Tctex1 (residues 110 – 122) and LC8 (residues 126 – 135) are demarcated by vertical dashed lines; these segments of IC transition to $\beta$-strand structure in the bound state (Hall et al. 2009).
Among secondary NMR chemical shifts, those in the $^{13}$C$^\alpha$ nuclei are the most reliable in distinguishing alpha-helical structure from random coil, while secondary chemical shifts in $^{13}$C$^\beta$ nuclei provide a more reliable means to distinguish $\beta$-strand structure from random coil (Wang et al. 2002a); secondary chemical shifts were calculated in this study for both the $^{13}$C$^\alpha$ and $^{13}$C$^\beta$ nuclei from the difference between the observed and sequence-corrected random coil chemical shifts (Kjaergaard et al. 2011a, b). The values of $\Delta \delta C^\alpha$ and ($\Delta \delta C^\alpha - \Delta \delta C^\beta$) are presented in Figure 4.2b (left and right ordinates, respectively) and have magnitude of less than ±1.5 ppm for all residues in IC$_{TL}$, indicating a lack of rigid, fully-formed secondary structure. Rather, the majority of residues show only small deviations from zero, indicative of random coil conformations. Small positive values over a few consecutive residues (such as the small positive values of $\Delta \delta C^\alpha - \Delta \delta C^\beta$ for residues 90–92 in IC$_{TL}$) are indicative of transiently populated helical or turn conformations in the ensemble, while small negative values over a few consecutive residues are indicative of polyproline II structure or of transient conformational sampling of $\beta$-strand structure.

NMR scalar coupling constants can be directly related to bond torsion angles and measured experimental values can be compared with random coil values to identify residual local secondary structure. The $^{3}J_{HNH^\alpha}$ scalar coupling constant of a residue depends on its $\phi$ backbone torsion angle, providing for discrimination between different regions of secondary structure in Ramachandran space. Positive deviations of $^{3}J_{HNH^\alpha}$ values from residue-specific random coil values [$\Delta^{3}J_{HNH^\alpha} = ^{3}J_{HNH^\alpha(experimental)} - ^{3}J_{HNH^\alpha(random~coil)}$] indicate a propensity toward extended $\beta$-strand structure, with an average deviation of ~+1.5 Hz corresponding to full $\beta$-structure for most residue types (Smith et al. 1996). Negative values of $\Delta^{3}J_{HNH^\alpha}$ are indicative of turn or helical conformational propensities, with an average deviation of ~−2.0 Hz corresponding to full $\alpha$-helical structure for most residue types (Smith et al. 1996). Negative values of $\Delta^{3}J_{HNH^\alpha}$ are also consistent with polyproline II structure; the three-bond coupling values for this type of structure are generally ~0.6 Hz lower than those for random coil (Lam et al. 2003). Secondary $^{3}J_{HNH^\alpha}$ values for IC$_{TL}$ (Figure 4.2a, right ordinate) are primarily of magnitude ±1 Hz or less, suggesting that any $\alpha$-helical or $\beta$-strand structure in this protein is only partial or transiently populated within the ensemble. $^{1}D_{HN}$ RDCs (above) identified a region (residues 120–134) with relatively
larger negative values, suggestive of locally more extended conformations such as β-strand or polyproline II; the predominance of small negative secondary chemical shift values measured for this same region are also consistent with polyproline II structure or with transient β-strand structure. While the $^{1}D_{\text{HN}}$ RDCs and the two $^{13}C$ secondary chemical shifts taken alone cannot reliably distinguish between these two types of secondary structure in this case, the small negative secondary $^{1}J_{\text{INH}}$ values in this region provide the distinction of polyproline II rather than β-strand structure (for which positive values of $^{3}J_{\text{INH}}$ would be expected).

The one-bond heteronuclear coupling constant $^{1}J_{\text{C}_\alpha \text{H}_\alpha}$ for a given amino acid residue is primarily determined by the $\psi$ angle, and by $\phi$ as well, to a lesser extent (Vuister et al. 1992). The $^{1}J_{\text{C}_\alpha \text{H}_\alpha}$ values of residues in IC$_{\text{TL}}$ were measured from the splittings of $\text{H}_\alpha$-$\text{C}_\alpha$ correlations measured along the $^{13}C$ dimension from a 3D BEST CT-HN(CO)CA-JCH (Rasia et al. 2011) experiment collected on the isotropic sample of IC$_{\text{TL}}$. Given the nature and properties of the experiment, neither the C-terminal residue nor residues preceding prolines can be observed and values are not reported for these residues; also, due to complications from additional splittings due to the presence of geminal H$^\text{a}$ protons in glycine residues, values are not reported here for glycine residues. Figure 4.2c shows the difference ($^{\Delta}^{1}J_{\text{C}_\alpha \text{H}_\alpha}$) between observed and residue-specific random coil values (Vuister et al. 1993b) for residues of IC$_{\text{TL}}$ whose constants could be obtained. Positive values of $^{\Delta}^{1}J_{\text{C}_\alpha \text{H}_\alpha}$ on the order of ~4–5 Hz are indicative of full α-helical structure (Vuister et al. 1993b), while small negative values (on the order of ~−0.6 Hz, for instance) are expected for extended β-structure. Positive deviations of $^{1}J_{\text{C}_\alpha \text{H}_\alpha}$ values from random coil on the order of ~1.1 Hz are expected for polyproline II structure (Lam et al. 2003). Few negative values of $^{\Delta}^{1}J_{\text{C}_\alpha \text{H}_\alpha}$ are seen for IC$_{\text{TL}}$ (and those values that are negative are not sequential), with the majority of values being positive and < 2 Hz in magnitude. The majority of values seen for $^{\Delta}^{1}J_{\text{C}_\alpha \text{H}_\alpha}$ in IC$_{\text{TL}}$ are intermediate between the values corresponding to a random coil and the values for polyproline II structure.

**Ensemble of Structures Populated by IC$_{\text{TL}}$ in Solution.**

To obtain more direct insight into the residual secondary structure and ensemble distribution of structures populated by IC$_{\text{TL}}$ in solution, we applied the flexible-mecanno/ASTEROIDS approach (Ozene et al. 2012a) with ensemble selection on the basis of agreement with the experimental RDCs [$^{1}D_{\text{HN}}$, $^{1}D_{\text{C}_\alpha \text{H}_\text{a}}$, $^{1}D_{\text{C}_\alpha \text{C}_\alpha}$, and $^{4}D_{\text{HN-H}_\alpha(-1)}$] and chemical shifts
(CSs) \([^{13}\text{C}^\alpha, ^{13}\text{C}^\beta, ^{13}\text{C}', ^{15}\text{N}, \text{and } ^1\text{H}^\text{N}]\) data determined for IC\textsubscript{TL}. The back-calculated CS and RDC values from the final ASTEROIDS-selected ensemble are in good agreement with the experimentally determined values (Figures A1.1 and A1.2). Populations of the different regions of Ramachandran space (defined in \textit{Methods}) determined from the ensemble analysis of IC\textsubscript{TL} are presented in Figure 4.3. Most notably, the results indicate a general decreased sampling of backbone dihedral angles in the \(\beta_S\) (beta-sheet) quadrant of Ramachandran space, compared to the random statistical coil description. This reduction in beta-sheet (\(\beta_S\)) conformational sampling is largely coincident with increased \(\beta_P\) (polyproline II) population, compared to the statistical coil description. The population of the \(\beta_P\) conformational region is elevated in numerous segments, being most significantly and continuously elevated in residues 123–136 (on average \(\sim 20\%\) higher than expected for random coil), which contains the crystallographically-determined (Hall \textit{et al.} 2009) binding site (IC residues 126–135, sequence highlighted in \textit{green} in Figure 4.3) for dynein light chain LC8. In addition, small increased (above statistical coil) \(\alpha_R\) (right-handed \(\alpha\)-helix) populations are seen in local regions such as residues 91–94 and 110–113. In contrast to the other three quadrants, conformational sampling of the left-handed \(\alpha\)-helix (\(\alpha_L\)) quadrant of Ramachandran space by IC\textsubscript{TL} does not exhibit deviation from that expected for a random statistical coil.

\textbf{Dynamical Characterization of IC\textsubscript{TL}.} 
To gain insights into the motional properties of IC\textsubscript{TL} and to probe for differences in conformational dynamics along the polypeptide chain, NMR \(^{15}\text{N}\) relaxation rates and steady-state \(^1\text{H}–^{15}\text{N}\) heteronuclear NOEs were determined for apo IC\textsubscript{TL}. \(^{15}\text{N} \ T_1\) (Figure 4.4a) and \(T_2\) (Figure 4.4b) relaxation values offer information on backbone motions on the pico- to nanosecond time scale, with \(T_2\) values also containing contributions from motions on the micro- to millisecond time scale. The \(^{15}\text{N} \ T_1\) values (Figure 4.4a) range from 0.41 to 1.53 s (with an overall average value of 0.51 s) with values being predominately homogeneous throughout that majority of the chain and increasingly elevated at the C-terminus; note that residue 84 in the IC\textsubscript{TL} construct used for NMR dynamics experiments is preceded by non-native residues in the purification tag and is thus not the N-terminal residue in this particular construct. The \(^{15}\text{N} \ T_2\) values (Figure 4.4b) range from 0.15 to 0.81 s (with overall average value of 0.19 s) and are predominately homogeneous throughout the majority of the chain but significantly higher for the C-terminal residue. The \(^1\text{H}–^{15}\text{N}\) steady-state heteronuclear NOE values offer information on backbone motions on the
Figure 4.3. Ensemble conformational population distribution of IC_{TL}. Ensemble conformational population distribution of IC_{TL} among the four quadrants of Ramachandran space as defined in Experimental Procedures: β-sheet (β_S), polyproline II (β_P), right-handed α-helix (α_R), and left-handed α-helix (α_L); the population of these quadrants is denoted as p(β_S), p(β_P), p(α_R), and p(α_L), respectively. Population values were generated by application of the ASTEROIDS approach using experimental chemical shifts and RDCs collected for the protein. The population distributions for IC_{TL} are shown in color for each quadrant of conformational space, while calculated distributions for a random statistical coil model (from flexible-mecanno) with the same amino acid sequence as IC_{TL} are shown in black. The amino acid sequence of the protein is shown, with the crystallographically-determined binding sites for dynein light chains Tetex1 and LC8 highlighted in yellow and green, respectively.
Figure 4.4. Dynamics of IC<sub>TL</sub>. Plots of $^{15}$N longitudinal $T_1$ (A) and transverse $T_2$ (B) relaxation data recorded for $^{15}$N IC<sub>TL</sub> at pH 6.5 and 20°C; $T_1$ and $T_2$ values greater than 0.9 and 0.4 seconds, respectively, corresponding to the C-terminal residues were truncated in these plots. (C) Steady state $^1$H-$^{15}$N heteronuclear NOE values are shown as $I_{\text{sat}}/I_{\text{unsat}}$; NOE values lower than negative 3 corresponding to the C-terminal residues were truncated in the plot. (D) The ratio of the spectral density functions $J(\omega_N)/J(0.87\omega_H)$ for IC<sub>TL</sub> at 600 MHz are plotted for each residue (gray circles, left ordinate); the horizontal dashed line indicates the average value of the ratio $J(\omega_N)/J(0.87\omega_H)$ (7.17 ns•rad⁻¹) over all residues. The “bulkiness” function as defined by Zweckstetter and Blackledge (Cho et. al. 2007), multiplied by the bell-shaped function (Schwalbe et. al. 1997), is also shown in this plot (red line, right-hand ordinate). A diagram is shown (top) indicating the crystallographically-determined binding regions of the dynein light chains Tctex1 and LC8 in IC<sub>TL</sub>.
Figure 4.4. (Continued)
pico- to nanosecond time scale and range from +0.032 (residue Valine112) to –4.680 (near the C-terminus), exhibiting greater heterogeneity along the protein sequence (Figure 4.4c) than seen for the T1 and T2 values. The 1H–15N NOE values are primarily negative throughout the sequence (as expected for a disordered or unfolded protein), consistent with high flexibility and a lack of fixed, rigid structure.

The relaxation data collected for apo IC_TL were further analyzed using the reduced spectral density mapping approach of Farrow et al., 1995 (Farrow et al. 1995) to obtain insights into motions on various timescales; this approach estimates the magnitudes of the spectral density function at three frequencies (the zero frequency, the 15N frequency, and 0.87× the 1H frequency) and makes no assumptions about the manner of molecular tumbling or the uniformity of correlation times $\tau_m$ among residues. Spectral density function values $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$ were calculated (Figure A1.3), and the ratios of the spectral density functions $J(\omega_N)/J(0.87\omega_H)$ are shown for each residue of IC_TL in Figure 4.4d. Differences in $J(\omega_N)/J(0.87\omega_H)$ ratios are observed along the protein sequence, with values ranging from 0.62 (near the C-terminus) to 11.29 (occurring in the interior of the protein sequence). Lower values of the $J(\omega_N)/J(0.87\omega_H)$ ratio (reflecting lower $J(\omega_N)$ values and/or higher $J(0.87\omega_H)$) are correlated with increased flexibility and decreased order, while higher values of the ratio are correlated with decreased flexibility and increased order (e.g. increased restriction of motions on the pico- to nanosecond timescale).

Decreasing below-average $J(\omega_N)/J(0.87\omega_H)$ values are seen in the C-terminal portion of the protein, consistent with the increased flexibility expected for this terminal region; a value significantly below the average is also seen for residue Glycine99, reflecting the relatively lower $J(\omega_N)$ and higher $J(0.87\omega_H)$ values seen for this particular residue (Figure A1.3b,c). Regions of the protein with higher-than-average $J(\omega_N)/J(0.87\omega_H)$ values are seen in the interior of the sequence, with localized maxima seen in regions containing Serine92; Leucine110, Valine112, and Tyrosine113; Lysine123, Leucine126, and Glutamine131. The relatively higher values of the $J(\omega_N)/J(0.87\omega_H)$ ratio in these regions indicates increased order (on the pico- to nanosecond timescale) in these residues relative to others in the IC_TL polypeptide chain.

Design and Characterization of Spin-Labeled IC144L Variants.

To obtain further insight into deviations of dynein IC from random-coil behavior, we sought to probe for evidence of any (transient) tertiary structure in a larger fragment containing the first 143 residues of IC. Given the previously-determined dynamic and flexible nature of the N-terminal 143 residues of the Drosophila dynein intermediate chain (IC) (Morgan et al. 2011),
detection of any transient or weakly-populated long-range intra-chain contacts was pursued through highly-sensitive NMR measurement of paramagnetic relaxation enhancement (PRE) effects. While the dipolar interaction between protons (as in $^1$H–$^1$H NOEs) is limited to distances of up to ~5.5 Å, in the case of the dipolar interaction between a free electron and a proton, due to the fact that the gyromagnetic ratio of the electron is ~660 times larger than that of a proton, the range of this interaction underlying the PRE effect extends to distances on the order of up to ~25 Å (Gillespie et. al. 1997b), with this enhanced sensitivity allowing detection of interactions existing at much lower populations and/or at longer distances than permitted by standard NOEs.

To simplify spectral analysis for those experiments in which we wanted to examine PRE effects in a more biologically relevant assembled state of IC (i.e. with the light chains bound, vide infra), we used a base construct of the first 143 residues of Drosophila IC in which the Tctex1 binding region is replaced with an LC8 recognition sequence (e.g. ‘IC1 LL’, Figure 4.1, bottom). IC1 LL binds 2 equivalents of LC8 with higher affinity than the wild-type IC binds Tctex1 and LC8, and the structure of IC in complex with the light chains is highly similar in both cases (Hall et. al. 2009).

Since the primary sequence of this IC construct lacks cysteine, five different single cysteine-containing mutants (IC1 LL-K11C, IC1 LL-R25C, IC1 LL-S60C, IC1 LL-S84C, and IC1 LL-L108C) were constructed to provide attachment sites for the thiol reactive methanethiosulfonate (MTSL) nitroxide radical spin label (see Figure 4.1, bottom). Both of the K11C and R25C sites are located within ‘Region 1’ (labeled ‘R1’ in Figure 4.5, top), corresponding to residues 1–41, which constitutes the binding site for the NuDE dynein regulator (Nyarko et. al. 2012), and is also the first of two regions that comprise the dynactin p150 Glued binding site in D. melanogaster IC (Morgan et. al. 2011). The S60C site is located within ‘Region 2’ (labeled ‘R2’ in Figure 4.5, top), corresponding to residues 46–75, and constituting the second of two regions comprising the p150 Glued binding site on IC (Morgan et. al. 2011). The S84C and L108C sites are located in the C-terminal half of the IC1 LL construct, with the L108C site being just N-terminal to the first light chains binding site (residues 111–120). Previous NMR structural and dynamical characterization of the first 143 residues of Drosophila IC revealed the presence of a well-formed α-helix within residues 1–40 (essentially ‘Region 1’) and a second, more nascent and weakly-populated helix comprised of residues 48–60 (the N-terminal part of ‘Region 2’, Figure 4.5, top) (Morgan et. al. 2011).

The HSQC spectra of the five mutants with MTSL labels in the diamagnetic (reduced) state were highly similar and largely superimposable with each other and the non-mutant, non-
Figure 4.5. Characterization of transient tertiary structure sampled by IC1_{LL}-Cys mutants. Intensity ratio plots for MTSL spin-labeled single cysteine mutants of IC1_{LL} collected at pH 6.5 and 5°C: (A) IC1_{LL}-K11C, (B) IC1_{LL}-R25C, (C) IC1_{LL}-S60C, (D) IC1_{LL}-S84C, and (E) IC1_{LL}-L108C. The ratio of peak intensities in the $^1$H–$^{15}$N HSQC spectra of the oxidized (paramagnetic) and reduced (diamagnetic) states of the MTSL spin-label are plotted as a function of residue number for each of the five constructs. Vertical arrows in the plots indicate the site of MTSL spin label attachment for each mutant construct. Blank spaces in the plots correspond either to the presence of proline residues (positions 65, 89, 107, 121, and 122) or to non-assigned residues. The dotted lines (black) represent the PRE intensity ratios simulated for each of the 5 constructs by the flexible-meccano program for a purely random coil ensemble of conformations. A diagram is shown (top) depicting the sub-domains ‘Region 1 (R1)’ and ‘Region 2 (R2)’ (light gray shading) and the crystallographically-determined (Hall et. al. 2009) dynein light chain LC8 binding sites (dark gray shading) in the IC1_{LL} construct. Corresponding secondary structure elements are indicated below the IC1_{LL} diagram: a well-defined α-helix (gray cylinder) that comprises ‘Region 1’ and a more nascent short helical segment (light gray outlined cylinder) at the N-terminus of ‘Region 2’ (both determined previously for IC:1–143 by NMR spectroscopy; Morgan et. al. 2011), and black rectangles denote the regions of elevated polyproline II conformational sampling identified in this study for the LC8 binding region of IC_{TL} (Figure 4.3). Note that this IC1_{LL} construct contains two LC8 binding recognition sequences, and the enhanced polyproline II conformation sampling seen within the wild-type residues 123–136 in IC_{TL} are conceptually transferred to the identical sequence of amino acid residues present within the first (more N-terminal) LC8 recognition sequence in IC1_{LL}—hence the depiction of two black rectangles in this schematic.
Figure 4.5. (Continued)
spin-labeled (IC1LL) protein (data not shown), with perturbations in peak positions being localized to the site of spin-label attachment and residues adjacent in the primary sequence, suggesting that no significant structural change is introduced by the presence of the spin-labels. Three of the five mutants contained spin-label attachment sites in regions of the protein previously determined to have significant flexibility and a lack of fixed secondary structure (Morgan et. al. 2011), while the remaining two (IC1LL-K11C and IC1LL-R25C) had spin-label attachment sites that fell within a region of the protein previously determined to have well-defined α-helical structure; particular attention was paid to any possible structural or functional consequences of spin-label attachment in this region of the protein. Analysis of chemical shift data for residues 1–41 (‘Region 1’) of IC1LL and the diamagnetic state MTSL-labeled IC1LL-K11C and IC1LL-R25C proteins indicated conservation of the helical structure in this region (Figure A1.4). In addition to conservation of structural properties, the MTSL-labeled IC1LL-K11C and IC1LL-R25C proteins were observed to retain their function with respect to their ability to bind the nNudE and p150Glu221-509 proteins, as judged from the pattern and relative magnitudes of peak attenuation seen in the NMR spectra of isotopically-labeled IC when in complex with unlabeled binding partners p150Glu and NudE (Figure A1.5).

Detection of Long-Range Interactions within IC1LL.

Paramagnetic nitroxide spin labels cause substantial broadening of the resonances (peaks) of nearby nuclei via an increase in transverse relaxation (R2) rate; the manifestation of this dipolar interaction between unpaired electron and nucleus (1H in this study) is a decrease in spectral peak intensity. NMR-PRE profiles for the five single-cysteine IC1LL constructs are presented in Figure 4.5. Intensity ratios (Iparamagnetic / Idiamagnetic) of amide proton peaks were obtained from 1H–15N HSQC spectra acquired for each cysteine mutant with the MTSL nitroxide spin label in either its oxidized (paramagnetic) or reduced (diamagnetic) state. The intensity ratio Iparamagnetic / Idiamagnetic is a convenient indicator of the dipolar interaction between an unpaired electron and a nucleus of interest (amide protons in this study), with a ratio of unity indicating no relaxation enhancement and thus an average distance between the electron and nucleus (amide proton) larger than ~20–25 Å; a ratio of zero indicates complete elimination of the amide proton signal, indicating a very strong and short-range interaction between the two. The localized effects of the nitroxide radical on the NMR spectral peak intensity ratios can be seen in the vicinity of spin-label attachment for each of the IC1LL single-cysteine variants; for the site of spin-label attachment and residues adjacent in the primary sequence, the paramagnetic broadening is significant and results
in low values of the $I_{\text{paramagnetic}} / I_{\text{diamagnetic}}$ ratio. For polypeptides behaving as statistical random coil polymers, the paramagnetic broadening decreases with increasing distance in the primary sequence from the site of spin-label attachment, and simulated PRE profiles are shown (dotted lines, Figure 4.5a–e) for each IC$_{1LL}$ variant based upon a statistical coil model for each protein sequence. In addition to the local effects of the paramagnetic spin-label, each of the five IC$_{1LL}$ variants exhibits additional regions (distant from the site of spin-label attachment) of decreased $I_{\text{paramagnetic}} / I_{\text{diamagnetic}}$ ratios, indicative of long-range contacts between different parts of the protein. While the specific appearances of the PRE profiles differ among the five different sites of spin-label attachment in the protein, the profiles qualitatively mirror and complement one another in terms of the long-range interactions that they indicate.

For spin-label attachment to residue 11 (IC$_{1LL}$-K11C variant, Figure 4.5a), significant paramagnetic broadening (beyond that expected for a statistical coil model) was observed within residues 49–79 (intensity ratios of 0.6 or less) with the greatest decrease in ratios occurring within residues 56–66 (intensity ratios of 0.1 or less); the range of residues 49–79 is largely coincident with the ‘Region 2’ domain of IC, indicating significant transient contact between the N-terminus of ‘Region 1’ and ‘Region 2.’ Additional smaller depressions in peak intensity ratios for IC$_{1LL}$-K11C are seen within residues 88–98 and within residues 106–116 (ratios as low as ~0.47) and residues 123–132 (ratios as low as ~0.55); the latter two residue segments are largely coincident with the light chains binding segments (dark gray shading, Figure 4.5, top). The profile of intensity ratios for spin-label attachment at residue 25 (IC$_{1LL}$-R25C, Figure 4.5b) shows significant paramagnetic broadening within residues 50–76 with the greatest decrease in ratios occurring within residues 54–64 (ratios as low as ~0.2); as with spin-label attachment at residue 11, the spin-label at residue 25 exhibits long-range PRE effects within the light chains binding segments of IC as well. For the IC$_{1LL}$-S60C variant (Figure 4.5c, and also Figure A1.6), two significant depressions in the peak intensity ratios occur within ‘Region 1’, centered about residues 14 and 24 (with intensity ratios as low as ~0.24). Thus, attachment of spin labels at two different sites within ‘Region 1’ indicate transient long-range interaction with residues within ‘Region 2’, and attachment of a spin label at residue 60 within ‘Region 2’ evidenced interactions of this region with residues in ‘Region 1.’ This reciprocity of PRE effects for different spin-labeled variants suggests that the observed effects are not induced by the modifications made in the IC protein but are due to inherent to physical properties (intra-molecular interactions) within the protein itself. For spin-labeled IC$_{1LL}$-S60C, paramagnetic broadening is also seen within residues 104–116 (ratios as low as ~0.4) and 124–133 (ratios as low as ~0.65), indicating
transient long-range interactions between this part of ‘Region 2’ and the light chains binding segments in IC.

Attachment of a spin-label at residue 84 (IC1LL-S84C, Figure 4.5d) produced a small reduction in intensity ratios for the first 30 residues (values of ~0.8 or less) within ‘Region 1’ as well as decreased peak intensity ratios in the light chains binding segments. The spin label attached at residue 108 (IC1LL-L108C, Figure 4.5e) is immediately preceded by Proline107 in the primary sequence, and is just N-terminal to the first LC8 binding site (IC1LL residues 111–120). Spin-label attachment at this site produced small decreases in peak intensity ratios for certain segments within ‘Region 1’ as well as noticeable decrease (intensity ratios ranging from ~0.3–0.7) in essentially the whole of ‘Region 2.’ In summation, several transient medium- and long-range interactions are observed within IC1LL, with reciprocal corroboration and complementarity in PRE effects seen among the five different spin-labeled variants. These experiments provide evidence for long-range intra-molecular interactions within IC; paramagnetic peak broadening due to inter-molecular interactions between different protein molecules can be discounted due to a lack of significant broadening observed in control experiments in which HSQC spectra of 15N-labeled IC1LL were recorded in the presence of excess non-isotope-labeled paramagnetic MTSL-IC1LL-K11C (Figure A1.7).

**Characterization of Long-Range Interactions in Assembled IC1LL.**

To investigate whether the transient tertiary structure observed for IC1LL in its free state (vide supra) is retained or altered for IC in complex with binding partners, NMR-PRE experiments were collected for 15N IC1LL-R25C with excess unlabeled LC8 or nNudE, or with both (Figure 4.6). As observed previously (Morgan et. al. 2011), peaks corresponding to IC residues in the binding interface with the light chains LC8 (IC1LL residues 110–120 and 124–138) are absent from the NMR spectrum and thus the NMR-PRE peak intensity profile when LC8 is bound (Figure 4.6a, green bars). For the remaining peaks in the LC8-bound state of 15N IC1LL-R25C, the PRE profile is highly similar to that of apo 15N IC1LL-R25C, including the significant decrease (compared to what is expected for a statistical coil model of IC1LL-R25C) in peak intensity ratios within residues 50–76, indicating that the transient long-range intra-molecular contact between ‘Region 1’ and ‘Region 2’ of IC is retained in the LC8-bound state.

When in complex with nNudE (Figure 4.6c, blue bars), the peaks corresponding to the first 41 residues of 15N IC1LL-R25C disappear from the NMR spectrum (in agreement with previous results that mapped the binding site of NudE to the first 41 residues of IC;
Figure 4.6. Characterization of transient tertiary structure sampled by IC1LL in its bound states. Intensity ratio plots for MTSL spin-labeled $^{15}$N IC1LL-R25C protein at pH 6.5: (A) Apo IC1LL-R25C (orange bars) compared with IC1LL-R25C with excess non-isotopically-labeled LC8 bound (green bars), both at 20˚C; (B) Apo IC1LL-R25C (orange bars, as above in panel ‘A’) compared with 0.15 mM IC1LL-R25C with 0.9 mM non-isotopically-labeled nNudE and excess non-isotopically-labeled LC8 bound (maroon bars), both at 20˚C; (C) Apo IC1LL-R25C (magenta bars) compared with IC1LL-R25C with excess non-isotopically-labeled nNudE bound (blue bars), both at 5˚C. In each case, the ratio of peak intensities in the $^1$H-$^{15}$N HSQC spectra of the oxidized (paramagnetic) and reduced (diamagnetic) states of the MTSL spin-label are plotted as a function of residue number; vertical arrows in each plot indicate the site of MTSL spin label attachment for IC1LL-R25C. Blank spaces in the plots correspond either to the presence of proline residues (positions 65, 89, 107, 121, and 122) or to non-assigned residues. The dotted lines (black) represent the PRE intensity ratios simulated for IC1LL-R25C by the flexible-meccano program for a purely random coil ensemble of conformations. A diagram is shown (top) depicting the sub-domains ‘Region 1 (R1)’ and ‘Region 2 (R2)’ (light gray shading) and the crystallographically-determined (Hall et al. 2009) dynein light chain LC8 binding sites (dark gray shading) in the IC1LL construct. Corresponding secondary structure elements are indicated below the IC1LL diagram as described in the legend to Figure 4.5.
Figure 4.6. (Continued)
Nyarko et al. 2012). In contrast to the significant paramagnetic broadening seen within residues 50–76 for the free state of IC1LL-R25C (Figure 4.6c), in the nNudE-bound state only a small decrease (relative to the simulated random statistical coil profile) is seen and is limited to residues 59–66. Overall, there is marked increase in peak intensity ratios across ‘Region 2’ for nNudE-bound IC compared to the free state, indicating that this whole region of the protein is more distant from ‘Region 1’ (spin-label attachment is at residue 25) and that IC may be sampling more extended conformations within its N-terminal 100 residues when NudE is bound. Furthermore, this same pattern is observed when the IC-NudE sub-complex is combined with LC8 in a ternary complex (Figure 4.6b), indicating that in its assembled state (which is biologically relevant to the assembled dynein motor complex as it exists and functions within the cell), binding of NudE results in IC sampling more extended conformations within its N-terminal ~100 residues.

Unfortunately, the cognate experiment with dynactin p150Glued (to assess the arrangement of Regions 1 and 2 relative to one another in the bound state with this particular regulator) could not be successfully performed due to the significant (essentially total) disappearance of peaks that occurs in the p150Glued-bound state of IC (Morgan et al. 2011); the remaining peak intensities for residues in Regions 1 and 2 of IC were not sufficient in the p150Glued-bound state to permit reliable assessment via the NMR-PRE experiments that could be performed successfully for the IC-NudE complex.

Discussion

In our ongoing efforts to characterize the structural and dynamical features of the cargo-attachment domain of the dynein intermediate chain, we present here the most detailed NMR analysis and structural characterization performed to-date of the IC chain peptide comprising residues 84–143 (IC_{TL}), containing the binding sites for the dynein light chains Tctex1 and LC8. Whereas previous examination evidenced a lack of secondary structure or rigidity in this segment of IC (Benison et al. 2006), the more detailed examination presented here indicates notable deviation from random coil behavior in the form of enhanced polyproline II conformational sampling for the light chains binding segments of IC—particularly for the region that binds LC8. We also provide the first evidence of significant transient tertiary structure within the unbound form of IC residues 1–143—the N-terminal portion of which, contains the overlapping binding recognition sites identified for the dynein regulator component proteins dynactin p150Glued and NudE (Morgan et al. 2011; McKenney et al. 2011; Nyarko et al. 2012). This unexpected
transient tertiary structure presents functional implications for *Drosophila* dynein IC interactions with competing regulatory binding partners dynactin p150Glued and NudE when both are present in the cell.

**Residual Secondary Structure in IC$_{TL}$—Enhanced Polyproline II Content.**

To probe the structure of IC$_{TL}$, a number of NMR observables were measured including chemical shifts, scalar couplings, and RDCs; all experimentally determined values demonstrate that this protein is highly disordered, lacking fixed secondary structure. Though not consistent with fixed secondary structure, the observed experimental values do evidence deviations from expected random coil values, suggesting the presence of residual or transient elements of various secondary structures in certain regions of the protein. However, it is important to note that measured chemical shifts, scalar couplings, and RDCs each represent ensemble averages for conformers inter-converting on a timescale faster than the millisecond. When deviations from expected random coil values are relatively small (compared to values expected for relatively static, fully-formed secondary structural elements) for a given observed NMR parameter, these values represent an arithmetic average over contributions made by all molecules in the ensemble, potentially corresponding to several different possible combinations of deviations in secondary structure populations. In other words, deviation from expected random coil values gives indication that Ramachandran space is not being sampled according to a random statistical coil description, but it does not necessarily give direct indication of the particular types and quantities of the deviations in conformational sampling. For instance, a $\Delta \delta C^\alpha$ value of +3.0 ppm would give clear indication of significantly populated $\alpha$-helical structure (largely to the exclusion of other conformations), while a $\Delta \delta C^\alpha$ value of +0.5 ppm also indicates deviation from random coil conformational sampling, but this relatively small value could be correlated with a small deviation in the form of small increased $\alpha$-helical conformational sampling of a residue (relative to a random coil), or possibly a larger increase in $\alpha$-helical conformational sampling accompanied by increased (relative to random coil) $\beta$-strand conformational sampling (whose negative contribution to the observed $\Delta \delta C^\alpha$ value would help to offset the increased positive contribution from the $\alpha$-helical conformations).

A major advantage of molecular ensemble calculations like those of the FM/ASTEROIDS approach (Ozenne *et. al.* 2012b; Nodet *et. al.* 2009), is the combination and
simultaneous consideration (and endeavor to achieve agreement with all of the experimental data within the level of experimental uncertainty) of multiple different experimental parameters whose values have different backbone $\phi/\psi$ dihedral angle conformational dependencies that are complementary in terms of mapping the different regions of Ramachandran space. It was previously demonstrated that the combination of $^{13}$C\[^{\alpha}\], $^{13}$C\[^{\beta}\], $^{13}$C\[^{\prime}\], $^{15}$N, and $^1$H\[^N\] chemical shifts (CSs) with $^{1}$D\[^{HN}\] RDCs for a polypeptide could unambiguously distinguish populations of the $\beta_P$ (polyproline II), $\beta_S$ (beta-sheet), $\alpha_R$ (right-handed $\alpha$-helix), and $\alpha_L$ (left-handed $\alpha$-helix) regions of Ramachandran space (Ozenne et al. 2012a). Application of the FM/ASTEROIDS CS-RDC approach to structural characterization of IC\[^{TL}\] included these same five chemical shift types and also 4 different types of RDC values for each residue in the protein. Deviations of the determined IC\[^{TL}\] conformational ensemble from the predicted random statistical coil behavior occurred as generally decreased $\beta_S$ sampling coupled with increased sampling of $\beta_P$ conformations and two short regions of consecutive residues with increased $\alpha_R$ conformational sampling (Figure 4.3). By themselves, while not providing the same conformational resolving power as the molecular ensemble calculations, the scalar coupling data collected for IC\[^{TL}\] do corroborate and are consistent with the results of the ensemble calculations (which did not include scalar coupling constant data, thus making these values an independent source of information) vis-à-vis deviations of IC\[^{TL}\] from random statistical coil behavior. For example, there is a dearth of positive $\Delta^{3}$H\[^{NNH}\] values, and also of negative $\Delta^{1}$J\[^{C\alpha\eta\eta}\] values for this protein (Figure 4.2a,c). In both cases, the signs of the deviations from random coil scalar coupling values are consistent with overall decreased $\beta$-strand content and with increased sampling of polyproline II and/or $\alpha$-helix ($\alpha_R$) conformations.

Dynamics data also suggest non-uniformity in the conformational and motional rigidity of residues in IC\[^{TL}\]. For instance, values of the $J(\omega_{N})/J(0.87\omega_{H})$ ratios reveal three separate regions in IC\[^{TL}\] with localized maxima corresponding to increased order relative to the rest of the chain (Figure 4.4d); like the scalar couplings (above), these dynamics data are also separate from the NMR parameters used in the ensemble calculations. Within each region (proceeding from the N- to C-terminus of IC\[^{TL}\]) more locally elevated values are seen for Ser92; Leu110, Val112, and Tyr113; Lys123, Leu126, and Gln131. Incidentally, the residues listed for the first two regions correlate with segments of polypeptide containing multiple sequential residues that were determined from the ensemble calculations to have increased $\alpha_R$ (right-handed $\alpha$-helical)
conformation relative to the statistical coil description, with the span of residues 91–94 and 110–113 each containing central pairs of residues whose \( p(\alpha_R) \) values exceed 0.5 (or, >50% \( \alpha \)-helical conformational content) (Figure 4.3). The occurrence of greater motional restriction in certain regions of the polypeptide backbone could be expected to correlate with increased sampling of secondary structures whose intra-chain backbone hydrogen-bonds (as seen in \( \alpha \)-helices) would produce increased ordering of backbone \( H-N \) vectors on the pico- to nanosecond time scale. Even residues not directly engaged in the intrachain hydrogen bond(s) of a single loop/turn of \( \alpha \)-helix (such as residues \( i+1 \), \( i+2 \), and \( i+3 \)) would be expected to exhibit some amount of increased ordering due to conformational constraints imposed by the hydrogen bond interaction between residues \( i, i+4 \) (and propagation of this rigidity to other residues within the single loop/turn).

The third set of residues mentioned above (Lys123, Leu126, and Gln131) fall within a region of IC\(_{TL}\) (residues 123–136) for which the ensemble calculations showed significantly and continuously elevated (relative to the statistical coil description) sampling of polyproline II structure with an averaged \( p(\beta_P) \) value of about 0.5 (Figure 4.3). Whereas an \( \alpha \)-helix contains stabilizing and rigidifying intrachain backbone hydrogen bonds, polyproline II helices possess an extended conformation that precludes the formation of main-chain intramolecular hydrogen bonds (Rath et. al. 2005). However, the side chains of certain types of amino acid residues (Glutamine, Asparagine, Arginine, Lysine, Threonine, and Histidine) can participate in main-chain/side-chain interactions in PPII structure, donating or accepting hydrogen bonds from their side-chain atoms to backbone carbonyl oxygens and amides; Glutamine is the most prevalent in this particular group of residue types, most commonly donating a hydrogen bond from its side chain amide to the backbone carbonyl of the proceeding residue (e.g. an \( i, i+1 \) interaction) (Stapley et. al. 1999). Incidentally, the residue with the global maximum value of the \( J(\omega_N)/J(0.87\omega_H) \) ratio is Glutamine131 (Figure 4.4d). Several of the above-listed residue types occur along the primary sequence of IC\(_{TL}\), including the region of increased polyproline II structure near the C-terminus, potentially providing stability as well as an explanation for the increased backbone rigidity inferred from the dynamics data.

Alternatively, as noted by Cho et. al., 2007, for an unfolded protein or a polypeptide behaving as a random coil, certain NMR observables (including dynamics measurements and RDCs) can exhibit noticeable variation in values along the polypeptide chain that can be attributed to and clearly predicted by amino acid ‘bulkiness’ (the ratio of the side chain volume to its length), without the need of invoking transient secondary structure as an explanation (Cho et. al. 2007).
al. 2007); local steric interactions between side chains and the backbone can restrict motions on the pico- to nanosecond time scale. Also, the side chain ‘bulkiness’ of an amino acid residue correlates with its hydrophobicity (Zimmerman et al. 1968), and in unfolded proteins—such as the case of unfolded apomyoglobin—some of the regions observed to have more restricted backbone motions on the pico- to nanosecond time scale (as evidenced by locally-elevated $J(\omega_N)/J(0.87\omega_H)$ ratios) were attributed to local hydrophobic collapse rather than formation of transient secondary structure (Yao et al. 2001; Schwarzinger et al. 2002). However, for the case of IC$_{TL}$, local maxima in the ‘bulkiness’ function (Figure 4.4d, right ordinate) do not strongly align with the local maxima in the $J(\omega_N)/J(0.87\omega_H)$ ratios (Figure 4.4d, left ordinate), suggesting that IC$_{TL}$ is not behaving purely as a random coil, and that localized restrictions in its backbone motions may indeed correlate with local elements of transient secondary structure, and are not ascribed to simple amino acid ‘bulkiness.’

**Functional Implications of Enhanced Polyproline II Content in IC$_{TL}$**

Many intrinsically disordered proteins or protein regions exhibit a coupled folding-upon-binding mechanism of interaction with their binding partners (Wright et al. 2009), and this phenomenon is observed in the case of the dynein intermediate chain in binding to its light chain partners LC7, LC8, and Tctex1, wherein the predominately disordered IC assumes α-helical or β-strand structure in its bound state with LC7 or with LC8 and Tctex1, respectively (Benison et al. 2006; Hall et al. 2009, 2010). Another common occurrence among intrinsically disordered protein regions is the preexistence of some degree of preformed residual or nascent localized structure—in interacting regions termed ‘MoRFs’ (Molecular Recognition Features)—that bears resemblance to the structure adopted by the IDP in its partner-bound state (Mohan et al. 2006). The LC7-binding region of IC exhibits nascent α-helicity in its apo state (Benison et al. 2006; Nyarko et al. 2011), which then translates into well-defined α-helical structure in the LC7-bound state (Hall et al. 2010). However, the Tctex1 and LC8 binding regions in IC do not exhibit nascent β-strand structure in the apo state which might presage the β-strand structure that these regions of IC assume in their light chains-bound state (Benison et al. 2007b; Hall et al. 2009). Rather, results of the ensemble calculations (based upon NMR data) reported here reveal an overall decreased sampling (relative to the random statistical coil description) of $\beta_S$ conformations, including regions of continuously decreased $\beta_S$ populations that coincide with the Tctex1 and LC8 binding sites in IC (Figure 4.3, top panel). Rather than preferentially sampling
β$_S$-type conformations in the apo state (as a prelude to the β-strand structure in the bound state), residues in the LC8 light chain binding segment of apo IC exhibit continuously elevated β$_P$ conformational sampling, while the Tctex1 light chain binding site exhibits shorter stretches of increased β$_P$ sampling (relative to the random statistical coil description) and a region of increased α$_R$ sampling at its N-terminal end (residues 110–113). However, it is interesting to note that, when IC is bound to both Tctex1 and LC8 (or for an IC$_{1L}$-type construct bound with 2 molar equivalents of LC8), in addition to the residues Proline121 and Proline122, the short linker region between the light chains binding sites (IC residues 123–125) also adopts an extended polyproline II structure (Hall et al. 2009). Thus, for these particular IC residues, the elevated polyproline II conformational content in the free state translates into the same type of secondary structure in the light chains-bound state.

Given the enhanced and substantial polyproline II (PPII) conformations exhibited by the LC8 recognition site of apo IC, we propose that this conformational character may potentially facilitate the “docking” of this segment of IC on the surface of its LC8 binding partner prior to assumption of its final β-strand bound-state structure (and perhaps likewise for the case of Tctex1, though to a lesser extent). The importance of extended polyproline II structure in protein–protein recognition elements stems from its particular molecular architecture wherein the backbone carbonyl oxygen atoms and the amide N–H protons (of non-Proline residues), as well as both the hydrophobic and polar portions of residue sidechains, are exposed on the surface and free to participate in hydrogen bonds and other inter-molecular interactions across the binding interface with a partner (Siligardi et al. 1995; Rath et al. 2005). While a single, isolated β-strand may also possess some of these same molecular topological characteristics (e.g. unsatisfied backbone hydrogen bonds and exposure of side chain atoms), a polyproline II helix enjoys the advantage of potential intra-chain (e.g. side-chain/main-chain, discussed above) hydrogen bonds that can contribute stabilization to the free form of the polypeptide. The preexistence of some ordering in the apo polypeptide backbone of a disordered segment can also help to decrease the entropic penalty that usually accompanies the folding of a disordered protein region upon binding to its partner (Wright et al. 1999).

With its three-fold symmetry (e.g. a left-handed 3$_1$ helix) along the helical axis and its high surface exposure of backbone and sidechain atoms, polyproline II structure can be thought of as presenting different ‘faces’ to a potential binding partner (Rath et al. 2005). The enhanced sampling of transient of polyproline II structure by the LC8-binding region of IC$_{1L}$, coupled with
the potential for stabilizing intra-chain hydrogen bonds within the segment, could result in the presentation of a particular arrangement of specific molecular fragments that could help confer some degree of preference or specificity in its molecular recognition and initial docking on the surface of its binding partner, LC8. Indeed, examples of the rapid formation of initial transient encounter complexes followed by evolution of the flexible polypeptide chain from this intermediate stage to its final bound-state structure have begun to emerge for the case of IDPs engaging in folding transitions coupled to partner binding (Sugase et. al. 2007; Narayanan et. al. 2008). On average, IDPs bind faster to their targets than ordered proteins, and in a critical assessment study of this kinetic advantage, it was determined that this advantage stems largely from the fewer total encounter times required before the formation of the final binding complex and from reduced binding free-energy barriers in evolving from the encounter complex to the final bound state—thus accelerating the folding-coupled binding process for an IDP with its partner (Huang et. al. 2009). In the proposed mechanism for $IC_{\text{T}}$ binding to LC8, while the enhanced and significant polyproline II (PPII) structure may facilitate initial recognition and docking of this region of IC on the target binding partner (LC8) surface, the retained flexibility in this segment (as inferred from NMR dynamics experiments, Figure 4.4) may allow it to “slip” into its binding groove in LC8, transitioning into its final β-strand structure in the process. The relatively open PPII conformation progresses easily to other conformational states (Williamson et. al. 1994), and in the case of the LC8 recognition sequence in IC, it’s as if this segment is ‘primed’ for binding its target, with PPII structure needing to undergo only a relatively small φ-angle transition within Ramachandran space to reach β-strand structure.

The conformational transition of intrinsically disordered protein regions from significant PPII conformation in the free state, to β-strand structure in an assembled (or, ‘bound’) state has been an emergent mechanistic theme in the amyloidogenic transition of certain proteins (Blanch et. al. 2000), including those involved in the major human neurodegenerative diseases Alzheimer’s and Parkinson’s. The human Tau (associated with Alzheimer’s disease) and α-synuclein (associated with Parkinson’s disease) proteins transition from soluble forms to large arrays of ordered β-structure (β-structured oligomeric aggregates) that compose neurofibrillary tangles and amyloid plaques in the brain (von Bergen et. al. 2005; Conway et. al. 2000). Using approaches similar to those reported here for the characterization of $IC_{\text{T}}$ (representative molecular ensemble selection on the basis of NMR data), recent detailed analyses of conformational preferences of human Tau and α-synuclein have revealed enhanced polyproline II conformational sampling, most notably in aggregation-nucleation sites for each protein.
We propose that the interaction of IC with its partner LC8 may represent another example (albeit one that is non-amyloidogenic and non-pathogenic) of an IDP segment that transitions from significant PPII structure in its free state, to β-strand structure in the bound (assembled) state—also adding another strand to the β-sheet structure at the core of its binding partner LC8, somewhat akin to the molecular assembly process in β-amyloid growth.

**Significant Transient Tertiary Structure in apo IC1LL.**

While the ensemble calculations performed for IC1LL evidenced local conformational preferences and deviations from the random statistical coil description, we also sought to probe for more long-range interactions within the larger construct corresponding to the first 143 residues of IC—a fragment that constitutes a ‘hotbed’ for binding activity, binding several different dynein as well as non-dynein binding partners and cargoes.

The PRE profiles observed in the immediate vicinity of MTSL spin-label attachment sites at residues 60, 84, and 108 of IC1LL are largely consistent with the (FM) simulated PRE ratios for a random coil ensemble, and are thus also consistent with previous NMR characterization of IC in which these regions were shown to be devoid of any rigid, fixed secondary structure (Benison et al. 2006; Morgan et al. 2011). In the case of spin-label attachment at residue 108, the immediately preceding residue in the sequence is Proline107, and the combination of a bulky MTSL label alongside a preceding Proline residue may account for the deviations from expected random coil values in the residues immediately N-terminal to this spin-label attachment site (Figure 4.5e). Among the PRE profiles for the five IC1LL variants, numerous medium- and long-range intra-molecular interactions are indicated, with reciprocal corroboration of these intra-chain contacts among the data collected for placement of a single spin-label at five different sites along the protein’s length. Perhaps the most salient feature among the PRE profiles for apo IC1LL (Figure 4.5), is the evidence of significant transient contact between ‘Region 1’ and ‘Region 2’ in IC. Whereas the intervening residues 42–45 (previously referred to as ‘Linker 1’; Morgan et al. 2011) have been described and conceptualized as a flexible linker between two helical segments, in light of results in the present study, this segment might perhaps be conceptualized as having some ‘hinge-like’ character as well. In addition, one particularly unexpected result is the interaction of the LC8 light chains binding segments (near the C-terminus of IC1LL) with other portions of IC, including Regions 1 and 2 located in the N-terminal half of this protein construct.

Incidentally, sites of intra-molecular contact within IC are seen to coincide with regions that contain certain secondary structural elements or propensities (Figure 4.5, top), such as the
well-formed $\alpha$-helix that is ‘Region 1’, the nascent helix at the N-terminus of ‘Region 2’, and also the enhanced and significant polyproline II conformations (determined from our ensemble calculations on IC$_{TL}$ in this study); there is also some evidence to suggest (for example, in the NMR-PRE profile for the IC$_{1L}$-K11C variant, Figure 4.5a) that the short regions of enhanced $\alpha$-helical ($\alpha_{R}$) conformations (such as for residues 110–113, Figure 4.3) identified from the ensemble calculations on IC$_{TL}$ participate to a small extent in these transient intra-molecular contacts. In a recent published report, there was definitive demonstration of the coupling between transient secondary and tertiary structure in the intrinsically disordered protein domain of the activator for thyroid hormone and retinoid receptors (ACTR) protein (Iesmantavicius et. al. 2013). In light of such an observation, as well as the results from our NMR-PRE measurements on apo IC$_{1L}$, we propose that secondary structural elements (either well-formed or nascent) may mediate the transient medium- and long-range intra-molecular contacts seen for apo IC. This insight might possibly explain differences observed in the pico-to-nanosecond timescale dynamics of (wild-type) IC:1–143 compared to those observed for the same residues in IC:84–143 (e.g. ‘IC$_{TL}$’), when both datasets were collected under identical solution and temperature conditions and on the same 600 MHz NMR instrument (see Figure A1.8). Transient tertiary intra-molecular contacts that are mediated by elements of nascent secondary structure might be expected to exhibit some stabilization or enhancement of the nascent secondary structure in these interacting segments, thus leading to more positive steady-state $^1$H–$^{15}$N heteronuclear NOE values for certain residues (Figure A1.8) in a longer construct (IC:1–143) compared to values for the same residues in a smaller construct (IC$_{TL}$) that cannot engage in the same intra-molecular contacts (because the part of the protein with which these segments might interact is simply not present in the smaller construct).

**Altered Sampling of Transient Tertiary Structure in Assembled IC$_{1L}$**

In light of the significant transient tertiary structure exhibited by apo IC, we sought to determine whether these types of medium- and long-range intra-molecular contacts persisted or were somehow altered when IC was bound to certain partners in its assembled and more biologically-relevant state. When LC8 is bound to IC$_{1L}$-R25C, the significant transient tertiary contacts that are seen between Regions 1 and 2 of apo IC are retained (Figure 4.6a). In the LC8-bound state of IC$_{1L}$-R25C, the PRE intensity ratio plots also evidence subtle but noticeable decreases in the region of residues 76–87 and 92–106, relative to the free state of IC$_{1L}$-R25C.
NMR-PRE experiments on apo IC1\textsubscript{LL} variants (Figure 4.5d) indicate weak interactions between this general region of IC and the light chains binding segments; in the presence of LC8, the light chains binding segments of IC are essentially sequestered by that binding interaction, such that contacts made with other regions of IC are abrogated. Having been thusly released from weak transient interactions with the light chains binding segments of IC, the spans of residues 76–87 and 92–106 might then become more accessible to the spin-label at residue 25, thus explaining the small relative decrease in peak intensity ratios seen for these residues in the LC8-bound state of IC1\textsubscript{LL}-R25C. Further support of this interpretation comes from the data collected for the sample in which nNudE was added to the IC/LC8 sub-complex; the small decrease (relative to apo IC1\textsubscript{LL}-R25C) in peak intensity ratios seen for residues 76–87 and 92–106 in LC8-bound IC1\textsubscript{LL}-R25C is reversed when nNudE is added (compare Figure 4.6a, green bars with Figure 4.6b, maroon bars), likely due to the sequestration of the protein region containing the spin-label that results when ‘Region 1’ of IC is bound to NudE, thus limiting the access of the spin-label to other regions within the protein. But the most prominent result of the NMR-PRE experiments conducted on assembled IC1\textsubscript{LL}, is the observation of decreased contact (relative to apo IC) between Regions 1 and 2 of IC when NudE is bound (Figure 4.6c), and the fact that this pattern persists in the ternary complex between IC, NudE, and LC8 (Figure 4.6b)—a complex that represents the assembled state of these proteins as they would exist within a functional dynein motor complex in the cell.

**Functional Aspects and Implications of Transient Tertiary Structure in IC1\textsubscript{LL}.**

The presence of medium- and long-range tertiary intra-molecular contacts, even when only transient, might seem puzzling in the context of an intrinsically disordered protein. However, increasing evidence is emerging for the existence of this type of structure within several different IDPs that have been studied, suggesting that these non-local contacts may play particular functional roles in IDPs (Bertoncini et. al. 2005; Mukrasch et. al. 2009; Platzer et. al. 2011). For instance, in the case of α-synuclein, long-range intra-molecular contacts between its C-terminus and the more centrally-located ‘fibril core region’ of the protein are proposed to (competitively) prevent inter-molecular binding interactions of the ‘fibril core region’ between several different α-synuclein molecules, which leads to formation of aggregates in the brain that are associated with Parkinson’s disease (Bertoncini et. al. 2005; also see Figure 2.4 in Chapter 2).

More generally, the relatively extended conformation of the polypeptide backbone in IDPs (and unfolded proteins) results in a larger hydrodynamic radius, relative to a folded protein
of the same mass (Uversky 2002). This biophysical feature of IDPs would tend to significantly slow and impede their diffusion through the crowded cytoplasm, impairing their ability to find and arrive at their molecular recognition targets (i.e. binding partners). Intra-molecular contacts in an IDP (like those observed in this study for the first 143 residues of IC), would help to make the disordered protein less extended and more compact, thus speeding its diffusion through the cell, on the way to arrival at a target binding partner. In addition, many IDPs exhibit remarkable binding promiscuity, being commonly involved in a wide range of inter-molecular interactions. Many IDPs exhibit transient elements of pre-formed secondary structure that are highly interaction-prone (e.g. “sticky”) and most often used by the IDP for binding to specific partners, as seen in the case of dynein IC:1–143. The inherent flexibility in these “sticky” transient pre-formed elements of secondary structure can potentiate unwanted interactions with non-native binding partners of the IDP—of which, there are many to choose from within the cell. Thus, it has been proposed that IDPs can undergo ‘functional mis-folding’ wherein the polypeptide chain uses intra-molecular contacts to help sequester these “sticky” elements within itself, thereby decreasing the occurrence of unnecessary and unwanted interactions with non-native binding partners (Uversky 2011b). To be functional, these intra-molecular contacts would need to be transient and only partially-populated (as opposed to very strong, static, constitutive interactions), so as to still permit the IDP to use these sequestered “sticky” regions for native binding contacts once it does encounter the correct binding partner. Indeed, transient and more weakly-populated intra-molecular contacts are the emerging theme in tertiary structures observed in IDPs.

In the case of assembled IC with dynein light chains bound, the significant transient intra-molecular contact between ‘Region 1’ and ‘Region 2’ in the apo IC protein is also present in the light chains-bound state. These two regions in the N-terminal half of IC:1–143 are significant in light of their participation in binding interactions with dynein regulatory proteins dynactin p150Glu, NudE/EL, and ZW10/RZZ (reviewed in Chapter 1). Upon binding interaction with NudE, the transient intra-molecular contact between ‘Region 1’ and ‘Region 2’ is significantly decreased, with IC sampling more extended conformations within its N-terminal ~100 residues, relative to the free state of IC. This finding has direct bearing upon previous experimental observations made on the interactions of Drosophila melanogaster NudE and dynactin p150Glu with IC. First, steady-state $^1$H–$^{15}$N heteronuclear NOE dynamics data showed that residues corresponding to the nascent helix at the N-terminus of ‘Region 2’ in IC (NudE only binds to ‘Region 1’ and not to ‘Region 2’) become slightly more disordered when NudE is bound than when IC is free (Nyarko et al. 2012). This hitherto puzzling observation can be readily
explained by the fact that there is decreased transient packing of ‘Region 1’ with ‘Region 2’ when NudE is bound which, given our proposed model wherein nascent secondary structures in IC mediate and are stabilized by long-range intra-molecular contacts, would explain the small increase in disorder of the nascent helix in ‘Region 2’ when NudE is bound.

Second, despite having comparable binding affinities, when p150\(^{\text{Glued}}\) is in excess, it can displace NudE from binding to IC, but the converse is not true—\(i.e.,\) excess NudE cannot displace p150\(^{\text{Glued}}\) bound to IC; as an explanation for these observations, it was proposed that some population within the IC–NudE ensemble might exist with ‘Region 2’ more exposed, and that this sub-population might bind more readily to p150\(^{\text{Glued}}\) and then, by mass action in a rapidly equilibrating mixture of IC, p150\(^{\text{Glued}}\), and NudE, p150\(^{\text{Glued}}\) would proceed to become the dominant species bound to IC (Nyarko et al. 2012). In essence, the results of our current study support this previous prediction vis-à-vis the existence of some population of the IC–NudE ensemble having a more exposed ‘Region 2.’ More specifically, the results of the present study actually demonstrate a shift (relative to free IC) in the populations of IC conformations assumed when NudE is bound. Binding of NudE to ‘Region 1’ of IC effectively helps to release ‘Region 2’ from this interaction, thus making it more accessible to other potential binding partners, such as dynactin p150\(^{\text{Glued}}\). In a sense, the significant release of ‘Region 2’ from its ‘sequestering’ interaction with ‘Region 1’ due to NudE binding, may provide dynactin p150\(^{\text{Glued}}\) with a sort of ‘foothold’ to initially engage in binding interaction with IC, ultimately displacing NudE, particularly if the initial p150\(^{\text{Glued}}\) binding to ‘Region 2’ event were to produce alterations in the conformational sampling in an adjacent region of IC that would make NudE binding less favorable.

**Conclusions**

In summation, ensemble calculations based upon a collection of RDC and chemical shift NMR data (which are corroborated and bolstered by independent NMR scalar coupling and dynamics data) evidence deviations from statistical coil behavior in IC\(_{\text{TL}}\) in the form of regions with enhanced secondary structure, with the most pronounced feature being the increased polyproline II conformational sampling near the C-terminus of IC\(_{\text{TL}}\). The enhanced polyproline II conformations exhibited by the LC8-binding region of IC are proposed to mediate its initial docking at the surface of LC8, followed by transition to its final β-strand structure in the bound state. Regions of either well-defined or nascent secondary structure are observed to coincide with
the multiple segments of IC that engage in medium- and long-range intra-molecular interactions within the first 143 residues of dynein IC (detected using NMR-PRE experiments); these regions of secondary structure are proposed to mediate the transient tertiary intra-molecular contacts observed, with nascent secondary structure segments becoming stabilized and more structured due to long-range transient tertiary contacts. In both the apo and light chains-bound state, IC Regions 1 and 2 exhibit significant transient intra-molecular contact with one another, that is disrupted by the binding of NudE to ‘Region 1.’ In summary, intrinsic protein disorder underlies the versatility of IC in binding to several different regulators (some even having overlapping binding sites in IC), however, deviations from random coil behavior (in the form of secondary structure content as well as transient tertiary structure) underlie and mediate interactions of IC with binding partners, and also furnish IC with additional functionality in the context of the cellular milieu.

**Acknowledgements**

We thank Yujuan Song for molecular biology work that created the IC1_{LL} single-cysteine mutant constructs, Jonathan Yih and Afua Nyarko for contribution of some of the IC_{TL} protein samples used in NMR experiments, and Dr. Gregory Benison for collecting the IC_{TL} NMR dynamics data. We acknowledge the support of the nucleic acid and protein core and the mass spectrometry facilities and services core in the OSU Environmental Health Sciences Center (National Institutes of Health NIEHS Grant00210). This work was supported, in whole or in part, by National Institutes of Health Grant GM084276.
Chapter 5

Concluding Discussion and Outlook
Summary

The intermediate chain (IC) is the central component of the cargo attachment sub-domain of the cytoplasmic dynein motor complex, and is an IDP exemplar of the ‘flexible molecular scaffold’ class, forming a flexible polybivalent scaffold in association with its dimeric binding partners. The studies described herein were aimed at elucidation of key structural and dynamical properties of a critical segment of ‘N–IC’ (residues 1–143 in *D. melanogaster* sequence numbering) that represents a ‘hotbed’ of binding activity, containing the binding sites for two of the dynein light chains (Tctex1 and LC8), as well as the binding recognition sites for several non-dynein proteins that are part of other macromolecular complexes (such as dynactin, NudE/Lis1, and ZW10/RZZ) that regulate the targeting and function of the cytoplasmic dynein motor within the cell. These studies were aimed not only at characterization of the intermediate chain IDP itself, but also at determination of the molecular details of IC interactions with protein components of certain complexes that regulate the dynein motor, and perhaps more importantly, at understanding the molecular basis for regulation of these binding interactions. In essence, by regulating the association state of IC with these particular non-dynein proteins, the association state of dynein with these regulatory protein complexes (dynactin, NudE/Lis1, ZW10/RZZ, etc.) is determined which, in turn, determines where the dynein motor complex will go in the cell, which function(s) it will perform, and whether or not it will be carrying certain cargoes on the way to its destination. In a sense, the dynein intermediate chain represents a key switch point (to use railroad parlance) for the regulation of dynein activity within the cell.

In this final chapter, I summarize the main conclusions of my dissertation work presented in Chapters 3 and 4. In addition, an assessment is made of outstanding questions that remain to be addressed on our way to understanding regulation of the dynein complex via the intermediate chain, as well as proposed approaches that may yield answers to these questions (‘Future Work’ section).

Main Conclusions

*Dynactin p150Glued exhibits a multi-site binding ‘footprint’ upon the Drosophila Dynein Intermediate Chain, providing insights into regulation of this binding interaction in other (mammalian) species.*

Dynactin is one of the most important and prolific regulators of the cytoplasmic dynein motor complex, being implicated in, and required for, nearly all functional activities of dynein in
the cell. In light of the fundamental functional importance of the dynein–dynactin connection (and the involvement of defects in these two complexes and their interaction in a variety of human diseases), there has been great interest in understanding the biophysical basis for the dynein–dynactin interaction and how this interaction might be regulated within the cell.

Using solution-state protein NMR spectroscopy, a bi-segmental binding ‘footprint’ for dynactin p150Glued in the predominately disordered Drosophila dynein intermediate chain was identified (Chapter 3). The two non-contiguous recognition regions in the N-terminus of IC were identified as: ‘Region 1’ (corresponding to IC residues 1–41) and ‘Region 2’ (corresponding to IC residues 46–75), separated from one another by a short, flexible 4-residue linker segment, ‘Linker 1’ (corresponding to IC residues 42–45). ‘Region 1’ consists of a well-defined and stable α-helix, while ‘Region 2’ contains nascent helicity at its N-terminal end (IC residues 48–60). Isothermal Titration Calorimetry experiments demonstrated that both Regions 1 and 2 are required for full binding affinity with dynactin p150Glued. Through inter-species sequence alignment and extrapolation of the results obtained in D. melanogaster to mammalian (rat) species, ‘Region 2’ of the p150Glued binding ‘footprint’ on IC overlaps with an alternative splice site R. norvegicus IC (Figure 1.3); given the observation in the D. melanogaster system that both Regions 1 and 2 are required for full binding affinity, this suggests that alternative splicing of IC within ‘Region 2’ could alter the amino acid sequence in this region and modulate the binding affinity between different isoforms of IC and p150Glued in mammalian species.

Further, NMR dynamics data collected for D. melanogaster IC in complex with p150Glued or in a ternary complex with p150Glued and the light chains bound, demonstrated retention of disorder in regions of IC outside of its interaction sites with binding partners. Perhaps most notably, there is retained disorder in a segment dubbed ‘Linker 2’ that is a flexible segment of IC that spans the separation between ‘Region 2’ and the light chains binding sites. This linker segment contains at its N-terminus, several phosphorylatable residues in a ‘serine-rich’ region that is known to undergo regulatory phosphorylation by cellular kinases; the flexibility of this segment in the assembled state of IC suggests that the phosphorylatable residues might be accessible to kinases (or phosphatases) even when IC is simultaneously bound to p150Glued and the light chains. In addition, some of the residues in the ‘serine-rich’ region partially overlap the C-terminal end of ‘Region 2’, including Serine-84 (R. norvegicus IC-2C isoform residue numbering), the phosphorylation of which, is known to abolish the IC–p150Glued binding interaction. Thus, phosphorylation of such a residue within ‘Region 2’ might potentially interfere with the binding of this region to p150Glued, which is required (along with ‘Region 1’) for
achievement of full binding between IC and p150\textsuperscript{Glued}—as demonstrated for the \textit{D. melanogaster} proteins. Lastly, alternative splicing at a second site in mammalian IC (Figure 1.3) can produce variable ‘Linker 2’ lengths. Via the ‘multivalency effect,’ this could also potentially affect whether or not binding of the light chains to IC is able to enhance the binding interaction ‘upstream’ between IC and p150\textsuperscript{Glued}.

In summation, the results of the published work presented in Chapter 3 provides key insights into the physical basis of the binding interaction between dynein IC and dynactin p150\textsuperscript{Glued} (and thus between dynein and dynactin) and suggests physical bases and mechanisms by which the affinity of this binding interaction might be modulated, so as to control whether or not the dynein motor complex is associated with dynactin and thus, the behavior of the dynein motor complex within the cell.

\textit{Deviations of Drosophila Dynein Intermediate Chain from random coil behavior underlie its interactions with binding partners and also suggest a biophysical basis for displacement of NudE from IC by dynactin p150\textsuperscript{Glued} when both are present in the cell.}

Deviations of the predominately disordered N–IC protein from random coil behavior in the form of either well-formed or residual (transient) secondary structures coincide with and seem to underlie its interactions with binding partners. Regions 1 and 2 of IC contain helical structure: Region 1 consists of a well-defined and stable \(\alpha\)-helix (residues 1–40), and Region 2 contains nascent helicity at its N-terminal end (IC residues 48–60) (Chapter 3). Both of these regions participate in the binding interaction with dynactin p150\textsuperscript{Glued}, while binding to NudE is mediated by ‘Region 1’ alone (Morgan \textit{et al.} 2011; Nyarko \textit{et al.} 2012). It was also determined from the work presented in Chapter 4, that the light chains binding segments of IC exhibit deviations from random coil behavior in the form of enhanced polyproline II conformational sampling relative to the statistical coil description. Thus, there are deviations (from random coil behavior) in the local conformational sampling of certain segments within IC:1–143, and these are seen to correlate with binding sites for IC partner proteins.

The work presented in Chapter 4 revealed the surprising result that IC:1–143 exhibits several significant transient intra-molecular contacts, with the regions of interaction largely coinciding with those that exhibit either well-defined or residual secondary structure. The most significant transient intra-molecular contact made within IC:1–143 is that between ‘Region 1’ and ‘Region 2’ in the N-terminal half of IC:1–143. Intriguingly, this transient intra-molecular contact is significantly decreased when NudE is bound to IC; the binding of NudE causes IC to sample
more extended conformations within its first ~100 residues, thus making ‘Region 2’ potentially
more exposed and accessible to other binding partners such as dynactin p150Glued. This
observation provides insight into the previously reported result that, despite have similar binding
affinities for IC, when p150Glued is in excess, it can displace NudE that is already bound to IC, but
the reverse is not true—i.e. excess NudE cannot displace p150Glued from IC (Nyarko et. al. 2012).
Results presented in Chapter 4 suggest that p150Glued might be able to use ‘Region 2’ (to which it
also binds, whereas NudE only binds ‘Region 1’) as a ‘foothold’ for initial binding interaction
with IC—given that this region actually becomes more exposed when NudE is bound—on the
way to displacing NudE from IC.

Future Work

The studies described in this dissertation work have provided structural and dynamical
insights into the dynein intermediate chain that can be built upon to further develop our
understanding of the interactions between dynein IC and proteins of dynein regulatory
complexes—and ultimately, the biophysical basis for regulation of these interactions in the cell,
as selection of binding to a particular regulator determines dynein targeting and function in that
cell. In this section I present a collection of specific extensions of my results as open questions
that address the biophysical bases of these protein-protein interactions and regulation thereof.

What is the biophysical basis of Drosophila dynein IC binding to other regulatory proteins,
such as ZW10?

Some very preliminary NMR results (Appendix 2, Figure A2.4) suggest that Drosophila
ZW10 may exhibit a bi-segmental binding footprint on IC, much like that observed for dynactin
p150Glued (Figure 1.4). Technical difficulties prevented completion of the NMR-based protein
titration of Drosophila IC:1-143 with full-length ZW10, however, these experiments could be
attempted again and possibly with modified protein constructs of ZW10 that may exhibit greater
solubility and stability. Assessment of the necessity (or, energetic contribution) to binding made
by particular segments of IC (e.g. perhaps Regions 1 and 2) could be made through application of
ITC experiments using smaller constructs of IC, as described in Chapter 3.

Previously published work in a mammalian species (rat) identified phosphorylation at a
particular residue in the ‘serine-rich’ region of dynein IC (Threonine-89, rat IC–2C isoform
residue numbering) as a requirement for efficient binding interaction with ZW10 (Whyte et. al.
2008). Interestingly, non-phosphorylated Drosophila IC:1-143 exhibited binding interaction with
ZW10 (see Appendix 2). The Drosophila cognate of Threonine-89 (numbering in rat IC–2C) is not immediately obvious from interspecies sequence comparisons. Furthermore, there is still no evidence that regulatory IC phosphorylation occurs in Drosophila, though, the conservation of a general ‘serine-rich’ region in the same basic part of the protein across species and throughout evolution, suggests that it serves some function. In any case, ITC experiments could be performed to assess the apo-IC binding affinity of ZW10, while IC-binding competition assays (among ZW10, p150Glued, and NudE) could be conducted using Native-PAGE and NMR-based titration experiments, in order to assess the hierarchy of binding interactions that occur when these three regulatory proteins are all simultaneously present in the Drosophila cell. Mammalian species are believed to use alternative splicing and phosphorylation of IC residues in the ‘serine-rich’ region to achieve differential binding to regulatory proteins that compete for overlapping binding sites on IC. In the absence of regulatory IC phosphorylation and alternative splicing in this region, fruit flies may be relying on other biophysical mechanisms to regulate binding interactions of IC with different dynein regulators.

Are there direct parallels in mammalian species?

Determining the specific roles played by the various mammalian splice isoforms of cytoplasmic dynein IC remains a critical unanswered question and challenge in understanding the complex regulation of cytoplasmic dynein function in mammalian cells. Simpler model organisms serve a variety of critical and indispensable purposes in scientific research, but there can be limitations when working with ‘lower’ and less complex species. For instance, yeast (S. cerevisiae) don’t possess ZW10/RZZ, and thus would not be useful if one were trying to study the regulation of cytoplasmic dynein by ZW10/RZZ. Likewise, Drosophila simply do not possess the same level of complexity seen in mammalian species, in terms of isoform diversity in the N-terminal-most region of IC. For this and other reasons, I believe that the future direction of work presented in this dissertation lies largely in its extension to the proteins in mammalian species, where application of protein biophysical techniques can to help elucidate details underlying the more complex regulation of dynein observed in these species.

Structure and dynamics in apo mammalian ICs?

The structure and dynamics of Drosophila apo IC:1-143 have been determined with residue-level resolution, however, it is unknown whether the same features are exhibited in mammalian ICs. Sequence-based secondary structure prediction suggests regions of α-helical
content in the N-terminal portion of mammalian (rat) IC that are analogous to those predicted and experimentally-determined in *Drosophila* IC. Alpha-helical structure is predicted for residues 7–41 (in rat IC-1A isoform) and for residues 7–43 (in rat IC-2C isoform); shorter α-helices are also predicted (with prediction strength equal to that seen for the little α-helix predicted for residues 49–59 in *Drosophila* IC) in residues 53–64 (rat IC-1A isoform) and residues 54–65 (rat IC-2C isoform). Thus, the N-terminal-most predicted α-helices in mammalian (rat) IC are coincident (through inter-species sequence alignment) with ‘Region 1’ defined in *Drosophila* IC, and there are also shorter predicted α-helices in a region of rat IC that would coincide with the N-terminal part of ‘Region 2’ as defined in *Drosophila* IC. The question of structure and dynamics in mammalian IC isoforms could be addressed directly and determined experimentally through application of protein NMR spectroscopy (as described in Chapter 3).

The presence of any (transient) tertiary structure could also be determined directly through NMR-PRE experiments (as seen in Chapter 4) for the mammalian IC isoforms. *Drosophila* IC:1-143 exhibited significant transient tertiary structure in the apo form, with packing between ‘Region 1’ and ‘Region 2’ (putatively facilitated through the transient packing of helical structures in the two regions) perhaps being the most prominent interaction. This ‘bending’ motion of *Drosophila* IC is also facilitated by its regions of disorder, particularly that seen in the ‘Linker 1’ segment (residues 42–45). This span of 4 amino acids contains 3 Glycine residues, which tend to have greater backbone conformational flexibility compared to other amino acid types, making them amenable to participation in protein turn-type conformations that would allow Regions 1 and 2 to pack together. No such collection of Glycine residues is seen in the analogous linker region in mammalian ICs, though there are 2 Proline residues in close proximity to one another in the intervening (putative) linker segment of rat IC-1A (Figure 1.3a) that may facilitate a turn conformation, allowing the putative analogous Regions 1 and 2 to pack together in rat IC-1 isoforms. However, the presence of consecutive Proline residues in this region would be expected to produce a turn that is less flexible and more rigid than that containing several consecutive Glycine residues.

Determination of structure (secondary and tertiary) and dynamics in mammalian ICs is requisite for understanding the biophysical basis of interactions with binding partners. In the case of *Drosophila* dynein IC, it was seen that elements of secondary structure (either well-formed or nascent) underlie the interaction of certain regions of the protein with binding partners; and deviation from random coil behavior (vis-à-vis transient tertiary structure, for example) may lend itself to regulation of the binding interactions with non-dynein regulatory proteins in the
Determination of mammalian IC binding interfaces with non-dynein regulatory proteins.

Though the binding sites for such proteins as dynactin p150Glued and NudE have been mapped with residue-level resolution via NMR spectroscopy for the case of Drosophila dynein IC, any parallels drawn to mammalian species are, at this point, speculative and based in part on extrapolations from inter-species alignment of sequences and sequence features (as seen in Figure 1.3a). It is not known whether certain regulatory proteins (such as dynactin p150Glued) exhibit a multi-segmental binding footprint on mammalian IC. The binding interfaces of regulatory proteins on mammalian IC isoforms could be mapped at residue-level resolution by NMR spectroscopy; confirmation that certain regions of an IC isoform contribute favorably to the energetics of the binding interaction (and are thus likely to be necessary for the full binding interaction) could be garnered through ITC experiments conducted with appropriate IC segmental constructs (as described in Chapter 3). Elucidation of binding interfaces in IC is important in that, it could help to shed light on mechanistic details underlying the selection of one IC binding partner vs. another when both are simultaneously present in the same vicinity or cellular compartment.

Experiments aimed at determining binding affinities of regulatory proteins for the different mammalian IC isoforms should also be conducted with light chains Tctex1 and LC8 pre-bound to IC, as this could potentially affect observed binding affinities via the ‘multivalency effect.’ In vivo, dynein IC is likely to exist in an assembled state with the light chains bound. So, in order to understand the binding selection of one regulatory protein vs. another as it occurs in the cell, the presence of the light chains on IC should be taken into account in assessing binding affinities of different regulatory proteins for the different mammalian IC isoforms.

What are the biophysical bases for the impact of IC phosphorylations seen to regulate binding interactions with non-dynein regulatory proteins in mammalian systems?

Previously published work in a mammalian species (rat) identified phosphorylation at Serine-84 and Threonine-89 (rat IC-2C isoform residue numbering) in the ‘serine-rich’ region of dynein IC as modulators of its binding interactions with proteins dynactin p150Glued and ZW10. Phosphorylation of Serine-84 and Threonine-89 each individually prevent binding of dynactin p150Glued to IC, while phosphorylation of Threonine-89 is required for the binding interaction of IC with ZW10. The effect of phosphorylation on the structure and dynamics of mammalian IC
itself is unknown, but could be readily probed and determined via NMR spectroscopy as was described for the characterization of apo Drosophila IC in Chapter 3. It might be the case that phosphorylation modulates the ordering or (secondary) structure of the region of IC in which it occurs, or it could possibly even affect the global (tertiary) structure and conformation of mammalian IC; these structural changes could in-turn be responsible for the changes in binding affinity with partners dynactin p150\textsuperscript{Glued} and ZW10. Several published examples exist demonstrating the impact that phosphorylation has on the structure and dynamics of disordered proteins and regions: in some cases, phosphorylation induces ordering or increased structure in the modified region (or nearby regions) of the protein; in other instances, phosphorylation is seen to disrupt or otherwise decrease secondary structure in the modified region of the protein; there are also many examples in which phosphorylation has little or no effect on the structure or dynamics of the protein, as is the case for the Sic1 yeast protein mentioned in Chapter 1 (in the section ‘Intrinsically Disordered Proteins—A Brief Introduction’). Any of these three scenarios are possible for the case of IC phosphorylations. It may be the case that the physical basis underlying the effect of phosphorylation on IC binding affinities has no conformational underpinning (i.e., is not affected by changes in the conformation of IC), but is simply just a matter of electrostatics.
Bibliography


Conway, K., Harper, J. D., and Lansbury, P. T. (2000) “Fibrils Formed In Vitro from α-Synuclein and Two Mutant Forms Linked to Parkinson’s Disease are Typical


Free and a Phosphopeptide-Complexed Src Homology 2 Domain Studied by $^{15}$N NMR Relaxation.” Biochemistry 33, 5984-6003.


Biol. 15, 386-392.


and may tether microtubules at adherens junctions.” Nat. Cell Biol. 3, 913-917.


Panchal, S. C., Bhavesh, N. S., and Hosur, R. V. (2001) “Improved 3D triple resonance experiments, HNN and HN(C)N, for H$^{1}$ and $^{15}$N sequential correlations in ($^{13}$C, $^{15}$N) labeled proteins: Application to unfolded proteins.” *J. Biomol. NMR* **20**, 135-147.


J. Biomol. NMR 51, 369-378.


Stuchell-Bereton, M. D., Siglin, A., Li, J., Moore, J. K., Ahmed, S., Williams, J. C., and Cooper, J. A. (2011) “Functional interaction between dynein light chain and intermediate chain...


Vuister, G. W., Delaglio, F., and Bax, A. (1993b) “The use of $^{1}J_{\text{C} \alpha \text{H} \alpha}$ coupling constants as a probe for protein backbone conformation.” J. Biomol. NMR 3, 67-80.

Wang, A. C., and Bax, A. (1995) “Reparameterization of the Karplus Relation for $^{3}J(H-C-N)$ and $^{3}J(H^N-C^\prime)$ in Peptides from Uniformly $^{13}$C/$^{15}$N-Enriched Human Ubiquitin.” J. Am. Chem. Soc. 117, 1810-1813.


Appendices
Appendix 1

The Role of Residual Structure in Dynein Intermediate Chain in Complex Assembly and Regulation—supplemental material

Jessica L. Morgan, Valéry Ozenne, Malene Ringkjøbing Jensen, Martin Blackledge, and Elisar Barbar
Figure A1.1. Comparison of ensemble-calculated RDCs with experimental values for IC$_{TL}$. $^1$D$_{\text{N-HN}}$ (a.k.a. $^1$D$_{\text{HN}}$), $^1$D$_{\text{C$_\alpha$-H$_\alpha$}}$, $^1$D$_{\text{C$_\alpha$-C'}}$, and $^4$D$_{\text{HN-H$_\alpha$(-1)}}$ RDCs measured for IC$_{TL}$ aligned in liquid crystalline polyethylene glycol/alcohol phase at 800 MHz. Comparison of experimental values (red) and values back-calculated from the final ASTEROIDS-selected ensemble (blue) with statistical coil FM predictions (black dotted line; these are RDCs calculated from a statistical coil ensemble with no particular secondary structure).
Figure A1.2. Comparison of ensemble-calculated secondary chemical shifts with experimental values for IC\textsubscript{TL}. Experimental $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, $^{13}\text{C}'$, $^1\text{H}_N$, and $^{15}\text{N}$ secondary chemical shifts (relative to random coil values from RefDB; Zhang \textit{et. al.} 2003) target data (\textit{red bars}); secondary chemical shift values back-calculated from the from the final ensemble of conformers (\textit{blue bars}) selected using the \textit{flexible-meccano}/ASTEROID\textsc{s} CS-RDC approach (Ozenne \textit{et. al.} 2012a).
Figure A1.3. Plots of reduced spectral density mapping data for apo IC$_{TL}$. Plots of the spectral density at: (A) the zero frequency, J(0); (B) the $^{15}$N frequency, J($\omega_N$); and (C) 0.87 of the $^1$H frequency, J(0.87$\omega_H$), for IC$_{TL}$ at 20°C derived from the relaxation data $T_1$, $T_2$, and steady-state $^1$H-$^{15}$N heteronuclear NOEs acquired at 600 MHz. Horizontal dashed lines in each plot indicate the mean values of J(0) [1.32 ns-rad$^{-1}$], J($\omega_N$) [0.312 ns-rad$^{-1}$], and J(0.87$\omega_H$) [0.049 ns-rad$^{-1}$].
Figure A1.4. Retained α-helical structure in the MTSL spin-labeled K11C and R25C mutants of IC1_{LL}. Comparison plot of secondary $^1\text{H}^N$ chemical shifts for residues 1 – 41 of the diamagnetic MTSL-labeled $^{15}\text{N}$ IC1_{LL}-K11C (red bars) and $^{15}\text{N}$ IC1_{LL}-R25C (green bars) constructs used in this study with those from the $^{15}\text{N}$ IC1_{LL} (blue bars) construct; data were collected for the proteins in 10 mM sodium phosphate (pH 6.5) with 50 mM NaCl, 1 mM NaN$_3$, at 5°C. Blank spaces in the plot correspond to non-assigned residues in each protein. A black dashed line is shown with ordinate value –0.19 ppm, which corresponds to the average secondary chemical shift value for residues in an α-helix reported by Wishart and Sykes (Wishart et. al. 1994a). Nearly all residues exhibit $^1\text{H}^N$ secondary chemical shift values that are negative, with most being below –0.19 ppm, indicating significant α-helical structure. The presence of the MTSL spin-labels attached via cysteine mutation of the protein sequence are seen to introduce only modest perturbations in the $^1\text{H}^N$ secondary chemical shifts compared to those for the non-mutant IC1_{LL}, indicating that the α-helical structure of ‘Region 1’ is largely retained in these spin-labeled single-cysteine mutants.
Figure A1.5. Retention of binding function of MTSL spin-labeled IC1$_{LL}$-K11C and IC1$_{LL}$-R25C constructs with dynein regulatory proteins dynactin p150$^{Glue}$ and NudE. Plots of relative peak intensity ($I_{\text{bound}} / I_{\text{free}}$) versus residue number; relative peak intensity is defined as the ratio of peak heights in the HSQC spectrum of the spin-labeled (diamagnetic state) $^{15}$N IC construct in complex with non-isotopically-labeled binding partner (NudE or p150$^{Glue}$) to the those in the spectrum of the free diamagnetic spin-labeled IC protein. To account for differences in IC protein concentration between the free and bound samples, a normalization factor was determined from the relative peak intensities for IC residue 143, which is not altered by the binding of p150$^{Glue}_{221-509}$ or nNudE. (A and B) 0.6 mM diamagnetic MTSL-labeled $^{15}$N IC1$_{LL}$-K11C (A) and $^{15}$N IC1$_{LL}$-R25C (B), each with non-isotopically-labeled p150$^{Glue}_{221-509}$ in a 1:0.5 molar ratio at 5˚C. The pattern and magnitude of peak disappearances (for residues 1–41) and significant attenuations (for residues 46–75) are entirely consistent with previous observations for the p150$^{Glue}_{221-509}$ binding interaction with wild-type IC:1–143 (Morgan et. al. 2011), suggesting that the spin-labeled IC1$_{LL}$-K11C and IC1$_{LL}$-R25C constructs constitute a reasonably faithful representation of the wild-type IC protein. This is further supported by the pattern and magnitude of peak attenuations seen for 0.5 mM diamagnetic MTSL-labeled $^{15}$N IC1$_{LL}$-R25C combined with a three-fold molar excess of nNudE at 5˚C (C), including the previously-observed disappearance of peaks corresponding to IC ‘Region 1’ residues 1–40 (C and D), and noticeable attenuation of peaks corresponding to the approximate range of residues 53–64 in IC ‘Region 2’ (C), that is not present for the previously cited lower 0.15 mM IC and 0.9 mM nNudE concentrations (panel D) (Nyarko et. al. 2012). Thus, this particular MTSL spin-labeled IC cysteine mutant exhibits similar behavior to wild-type IC in binding nNudE, including disappearance of peaks corresponding to residues 1–40, as well as the reproduction of decreased peak intensities in ‘Region 2’ that were previously attributed to non-specific artifacts due to overall high protein concentrations (Nyarko et. al. 2012).
Figure A1.5. (Continued)
Figure A1.6. Overlay of the $^1$H–$^{15}$N HSQC spectra of $^{15}$N IC1$_{LL}$-S60C with MTSL spin-label attached in its paramagnetic (orange contours) and diamagnetic (black contours) states. The spectral data for the IC1$_{LL}$-S60C construct is shown as an example of the data collected for the five single-cysteine mutants used in this study. Several peaks corresponding to residues in the vicinity of the site of spin-label attachment (residue S60C) are indicated (black text). Several additional peaks are labeled (orange text) for residues distant in the primary sequence from the site of spin-label attachment, whose relative peak intensities ($I_{\text{paramagnetic}}/I_{\text{diamagnetic}}$) are reduced to 50% or less; these residues are all within the ‘Region 1’ domain of dynein IC. NMR spectra were recorded on a Bruker 700 MHz spectrometer at 5°C in 10 mM sodium phosphate, pH 6.5.
Figure A1.7. Control PRE experiment to probe for inter-molecular interactions between IC molecules. No significant deviations from unity are observed for $I_{\text{paramagnetic}} / I_{\text{diamagnetic}}$ intensity ratios collected at 20°C for $^{15}$N-labeled IC1$_{LL}$ in the presence of excess non-isotope-labeled IC1$_{LL}$-K11C with MTSL spin-label attached and in its paramagnetic state. This indicates that the PRE effects seen for the single-cysteine IC1$_{LL}$ mutants (Figure 4.5 in Chapter 4) are not due to inter-molecular interactions between different IC molecules or interactions between the spin-label and the protein, and can be attributed instead to long-range intra-molecular contacts within IC1$_{LL}$. 

![Graph showing intensity ratios](image-url)
Figure A1.8. The impact of long-range intra-molecular interactions on IC protein dynamics. Steady-state $^1$H–$^{15}$N heteronuclear NOE data collected for IC:1–143 (grey bars) and IC:84–143 (e.g. ‘IC$_{TL}$’, red circles) in 10 mM sodium phosphate buffer (pH 6.5) at 20°C, recorded on the same 600 MHz Bruker NMR instrument; values are shown as $I_{\text{sat}}/I_{\text{unsat}}$ and NOE values lower than negative 1.5 corresponding to residues at the C-termini were truncated in these plots. In general, the steady-state $^1$H–$^{15}$N heteronuclear NOE values collected for IC$_{TL}$ are lower (more negative) than the values observed for the corresponding residues in IC:1–143. Regions of enhanced secondary structural content (identified from our ensemble calculations) in IC$_{TL}$ cannot participate in certain long-range intra-molecular contacts that are seen for the same span of residues in the larger IC:1–143 construct, since the N-terminal 83 residues of IC are essentially ‘missing’ from the IC$_{TL}$ construct. For transient tertiary intra-molecular contacts that are mediated by elements of nascent secondary structure, long-range intra-molecular contacts might be expected to provide some stabilization or enhancement of the nascent secondary structure, leading to greater ordering and more positive steady-state $^1$H–$^{15}$N heteronuclear NOE values for these residues in the context of the larger protein (IC:1–143) compared to those observed for the same residues in the smaller construct (IC$_{TL}$), which cannot engage in these same stabilizing long-range intra-molecular contacts. Though an overall general increase is seen for the values of most residues in IC:1–143 compared to their counterparts in IC$_{TL}$, it is notable that regions corresponding to residues with enhanced $\alpha$-helical ($\alpha_R$) content (residues 91–94 and 110–113 in IC$_{TL}$, indicated with red vertical arrows in the plot) transition from negative (in IC$_{TL}$) to noticeably positive values in the context of IC:1–143; the longest contiguous region of enhanced polyproline II content (residues 123–136 in IC$_{TL}$) are demarcated with a green bracket in the plot. Data presented for IC:1–143 were previously reported in Morgan et. al., 2011 (Morgan et. al. 2011). Data for IC$_{TL}$ are from this study (see Figure 4.4c in Chapter 4).
Figure A1.8. (Continued)
Appendix 2

Cloning and preparation of full-length *Drosophila melanogaster* Zeste-white 10 (ZW10) protein and its interaction with the Dynein Intermediate Chain

Jessica L. Morgan
Overview
The RZZ (Rod-ZW10-Zwilch) complex is one of several regulatory complexes of the cytoplasmic dynein motor, and implicated in targeting dynein to kinetochores during mitosis (reviewed in Kardon et. al. 2009b). Previously published work identified a direct interaction between dynein IC and ZW10 of RZZ in a mammalian species, thus mediating direct linkage between the two macromolecular complexes; more specifically, the published work established the necessity of phosphorylation of a particular residue (Threonine-89, rat IC–2C isoform residue numbering) within the ‘serine-rich’ region of the intermediate chain for binding to ZW10, suggesting that this region of IC might be involved in the binding interaction (Whyte et. al. 2008). This project was initiated to determine molecular details of the interaction between Drosophila dynein IC and ZW10—in a similar vein to my previous characterization of the interaction between Drosophila dynein IC and dynactin p150Glued (reported in Chapter 3).

Experimental Procedures

Construct Cloning and Protein Preparation.
A clone of the Drosophila melanogaster gene for ZW10 [gene Dmel\mit(1)15 clone #SD07771 obtained from the Drosophila Genomics Resource Center, Bloomington, Indiana] was used as the template in multiple polymerase chain reactions to remove a retained intron within the supplied clone, using the primers 5’-TACTTCCAATCCAAATGCAATGGAGGAAGAGGC-3’ (‘primer #1’) and 5’-CCATTAGTTTGACAAAGTCCTGAGCACTTTTCG-3’ (‘primer #2’) to amplify the first 1,350 bp of the ZW10 gene, and primers 5’-CGAAAAGTGCTCAGGACTTTGCTAAACTAATGG-3’ (‘primer #3’) and 5’- TTATCCACTTCCAATGCTATACAAATCTGCG-3’ (‘primer #4’) to amplify the last 816 bp (including the stop codon) of the ZW10 gene; the product fragments of these first two PCR reactions were then excised from an agarose gel, purified (Gel Extraction Kit, Qiagen), and joined together in a third PCR reaction using primers #1 and 4 to create a ZW10 cDNA which was subsequently cloned into a pMCSG9 expression vector (Stols et. al. 2002) using published protocol for ligation-independent cloning (Eschenfeldt et. al. 2009). This yielded a vector from which an N-terminal His\_6-tagged maltose-binding protein (MBP) fusion with full-length ZW10 protein (721 amino acid residues) could be expressed, with a tobacco etch virus (TEV) protease cleavage site intervening the C-terminus of MBP and the N-terminus of the ZW10 protein. The sequence of this ZW10 construct was verified by automated DNA sequencing prior to transformation into an Escherichia coli BL21 Rosetta host cell line for protein expression.
Transformed cells containing the His<sub>6</sub>-MBP-ZW10 construct were grown at 37 °C in Terrific broth (TB) media to an optical density (A<sub>600 nm</sub>) of ~0.4, followed by protein induction for 18 hours with 0.05 mM isopropyl 1-thio-β-D-galactopyranoside (IPTG) at 18 °C (Figure A2.1a). Cells were harvested, lysed, centrifuged to remove cellular debris, and the soluble fraction in the supernatant was purified using Qiagen nickel-nitrilotriacetic acid (Ni-NTA) affinity chromatography protocol (Qiagen, Valencia, CA). The His<sub>6</sub>-MBP moiety was removed with TEV protease and the post-cleavage mixture was subjected to additional Ni-NTA affinity chromatography (Figure A2.1b). The ZW10 protein was further purified via size-exclusion chromatography (SEC) on a Superdex<sup>TM</sup> 200 (16:100) gel filtration column (GE Healthcare) with a running buffer of 50 mM sodium phosphate (pH 7.3) with 0.2 M sodium sulfate and 1mM NaN<sub>3</sub> (Figure A2.2). <sup>15</sup>N-labeled IC:1-143 protein was expressed and purified as described previously (Morgan et al. 2011). For NMR and other experiments, protein concentrations were determined from sequence-based calculated molar extinction coefficients at 280 nm (IC:1-143, 2,980 M<sup>-1</sup>cm<sup>-1</sup>; ZW10, 57,300 M<sup>-1</sup>cm<sup>-1</sup>).

**NMR Spectroscopy.**

The <sup>15</sup>N IC:1-143 sample was prepared with 0.6 mM concentration in 10 mM sodium phosphate (pH 6.5) with 50 mM NaCl, 10 mM DTT (dithiothreitol), 1 mM NaN<sub>3</sub>, 10% <sup>2</sup>H<sub>2</sub>O, a mixture of protease inhibitors (Roche Applied Science), and 1 mM 2,2-dimethyl-2-silapentane-5-sulfonic acid (DSS). Chemical shifts were referenced with internal 2,2-dimethyl-2-silapentane-5-sulfonic acid (Markley et al. 1998). <sup>1</sup>H–<sup>15</sup>N HSQC spectra of apo<sup>15</sup>N IC:1-143 and <sup>15</sup>N IC:1-143 with non-isotopically-labeled ZW10 at a ~1:0.2 molar ratio were recorded using echo-anti-echo phase discrimination of 256 increments with 1024 points at 5 °C on a Bruker 700 MHz spectrometer equipped with room-temperature BBI probe (Figure A2.3). Additional steps in the IC:ZW10 NMR titration process could not be successfully obtained due to technical difficulties (see below, ‘Summary and Outlook’). NMR spectra were processed with NMRPipe (Delaglio et al. 1995) and analyzed using NMRView (Johnson 2004). Backbone chemical shift assignments for IC:1-143 at 5 °C were those reported previously (Morgan et al. 2011). The relative intensity (I<sub>bound</sub>/I<sub>free</sub>) of each peak was calculated as the ratio of the intensity of the peak (determined as peak height) in the spectrum of bound IC to the intensity of the same peak in the spectrum of unbound IC (Figure A2.4); to account for small differences in IC concentration in the collected spectra, a normalization factor was determined from the peak intensity of residue 143, which is presumably not altered by ZW10 binding.
**Figure A2.1.** Expression and purification of *Drosophila melanogaster* ZW10 protein. (A) SDS-PAGE analysis of His$_6$-MBP-ZW10 expression over time (lanes 1-4, highlighted in green). His$_6$-MBP-ZW10 (theoretical mass ~128 kDa) runs right alongside the molecular weight marker of 130 kDa (lane 5, highlighted in red). The vast majority of His$_6$-MBP-ZW10 is in the soluble fraction (lane 6). (B) Affinity chromatography purification of His$_6$-MBP-ZW10 (lanes 1-6), TEV protease cleavage reaction (pre- and post-cleavage reaction in lanes 8 and 9, respectively), and second affinity chromatography purification (lanes 10-12) to yield purified ZW10 (lane 10). Molecular weight markers (MW, lanes 3, 7, and 13) are labeled with the mass of each band in units of kilodaltons. Protein bands are highlighted, corresponding to His$_6$-MBP-ZW10 (highlighted in green, theoretical mass ~128 kDa), His$_6$-MBP (highlighted in blue, theoretical mass ~45.5 kDa), and ZW10 (highlighted in magenta, theoretical mass ~82.6 kDa).
Figure A2.2. Size exclusion chromatography (SEC) of *Drosophila melanogaster* ZW10 protein. (Top) Chromatogram of ZW10 purification via SEC on a Superdex™ 200 (16:100) gel filtration column with flow-rate of 60 mLs/hour and a running buffer of 50 mM sodium phosphate (pH 7.3) with 0.2 mM sodium sulfate and 1 mM Na$_2$S$_3$ (with no reductant present in the running buffer). Fractions were collected from elution times: 68–78 minutes (red bracket), 78–86 minutes (blue bracket), and 100–116 minutes (purple bracket) and were analyzed via SDS-PAGE. (Bottom) SDS-PAGE analysis of ZW10 fractions from SEC purification. Samples from each of the indicated elution fractions (lanes 1–6) were prepared with sample loading dye that contained either 0 (‘-’) or 10 mM (‘+’) β-mercaptoethanol (‘βME’). Molecular weight markers (MW, lane 7) are labeled with the mass of each band in units of kilodaltons.
Figure A2.3. Overlay of $^1$H–$^{15}$N HSQC spectra of apo IC:1-143 (black) and IC:1-143 in the presence of unlabeled ZW10 at a relative molar ratio of ~1:0.2 (purple). Several peaks are labeled [corresponding to residues within ‘Region 1’ (Morgan et. al. 2011)] whose relative intensities are attenuated by more than 80% in the presence of ZW10 (black text); additional peaks are labeled [corresponding to residues within ‘Region 2’ (Morgan et. al. 2011)] whose relative intensities are decreased by 60% or greater when ZW10 is present (purple text). The spectra were recorded at 5 °C with ~0.6 mM $^{15}$N IC:1-143 in 10 mM sodium phosphate buffer (pH 6.5) on a 700 MHz Bruker NMR instrument with room-temperature BBI probe.
Figure A2.4. Numerical plot of relative peak intensities ($I_{\text{bound}}/I_{\text{free}}$) versus residue number for the spectra shown in Figure A2.3. Relative peak intensity is defined as the ratio of the peak intensity in the spectrum of IC:1-143 with ZW10 (molar ratio ~1:0.2) to the peak intensity in the spectrum of free IC:1-143 protein. Two distinct series of peaks (encompassed within ‘Region 1’ and ‘Region 2’; Morgan et. al. 2011) in IC:1-143 exhibited significant attenuation upon interaction with ZW10.
Summary and Outlook

Summary.

The full-length *Drosophila melanogaster* ZW10 protein (721 amino acids residues, theoretical mass of ~82.6 kDa) is able to be expressed recombinantly and purified under native (non-denaturing) conditions (Figure A2.1). In addition, it can be isolated in monomeric form (Figure A2.2) and is soluble in the absence of its cohorts (Rod and Zwilch) normally present in the ‘RZZ’ complex (Karess 2005).

During the very first protein purification performed on the first large-scale culture of ZW10--particularly during size exclusion chromatography--it was discovered that full-length ZW10 has a propensity to form intermolecular disulfide-linked oligomers. The formation of these disulfide-linked oligomers was likely hastened during the process of sample concentration prior to SEC. In Figure A2.2, it can be seen that different populations of ZW10 exist and elute from the gel filtration column at different times, with larger molecular species eluting earlier in time. No significant peaks were observed in the chromatogram prior to 60 minutes of elution time, with the first 60 minutes of elution encompassing and corresponding to the excluded volume of this particular SEC column at this particular flow-rate. Given that the molecular weight range of this particular gel filtration resin is 10–600 kDa, any disulfide-linked oligomers are likely of size ~600 kDa or less. Given the SDS-PAGE analysis of the different eluted fractions from SEC (Figure A2.2, bottom), it seems likely that the intermolecular disulfide-linked oligomers of ZW10 are greater than dimeric—e.g. if disulfide-linked dimers were forming in significant quantity, one would expect to see visible bands in lanes 1 and 3, that are at the level of, or just slightly above, the 170 kDa (a *Drosophila* ZW10 disulfide-linked dimer would have mass ~165.1 kDa) molecular weight marker band in lane 7 of the SDS-PAGE gel. But rather, significantly larger bands than that are see in lanes 1 and 3, with these bands running near to the very top edge of the SDS-PAGE gel. The primary sequence of full-length *Drosophila* ZW10 contains 15 cysteine residues. Based upon the SEC and SDS-PAGE analyses (Figure A2.2), I surmise that at least 2 of these cysteine residues must be present on the surface of the ZW10 protein, or are otherwise sufficiently accessible to permit formation of intermolecular disulfide bonds between individual ZW10 molecules.

The peak with elution center at ~106 minutes is likely consistent with monomeric (non-disulfide-linked) ZW10 (based upon the elution profiles of other globular proteins of known weight on this particular column at this particular flow-rate). In all subsequent preparations and
purifications of the ZW10 protein, reductant (usually ≥ 10 mM βME) was present in all buffers used, including buffers used for cell lysis as well as affinity and size exclusion chromatography purifications. Inclusion of reductant in the purification process significantly decreased the fraction of protein eluting earlier on the SEC column and increased the fraction eluting with peak center at ~106 minutes (data not shown). Additional observations include: once disulfide-linked oligomers have formed, the addition of reductant (even high concentrations thereof) to the preparation is unable to reduce the disulfide bonds formed between individual molecules of ZW10 (chromatography data not shown); the monomeric (non-disulfide-linked) form of ZW10 would seem to be the preferable form with which to work, as it is more amenable to staying in solution and can be concentrated more than oligomeric disulfide-linked ZW10, when tested in a variety of buffers; a preparation of initially monomeric (non-disulfide-linked) ZW10 protein, even in the presence of reductant, even when kept at protein concentrations ≤ 100 µM, will evidence precipitation (the precipitate is primarily disulfide-linked ZW10 oligomer, judged by SDS-PAGE analysis) over the course of a few days.

To identify the ZW10 binding interface of IC:1-143, an NMR titration experiment was attempted wherein ¹H–¹⁵N HSQC spectra of ¹⁵N IC:1-143 would be recorded with stepwise addition of unlabeled ZW10. To this end, an initial sample was prepared with an IC:ZW10 molar ratio of ~1:0.2; only reduced, monomeric ZW10 protein was used to prepare this sample. A superimposition of the ¹H–¹⁵N HSQC spectra of ¹⁵N-labeled IC:1-143, collected in the absence and presence of this initial small amount of ZW10 (Figure A2.3), reveals no novel peaks for IC:1-143 upon its interaction with ZW10. Instead, numerous peaks for IC:1-143 are significantly attenuated, presumably due to intermediate chemical exchange effected by the binding association with ZW10, leading to significant peak broadening and apparent disappearance of the peaks from the spectrum. In fact, the profile of peak disappearances (Figure A2.4) is highly reminiscent of that seen for the interaction of IC:1-143 with p150Glued (Morgan et. al. 2011).

Unfortunately, I did not have a sufficient quantity of monomeric ZW10 to continue further with the NMR titration on the same day that this initial ~1:0.2 (IC:ZW10) molar ratio spectrum was collected, but given the appearance of the spectrum from this first step in the titration, I endeavored to prepare more monomeric ZW10 and to resume the NMR titration before the ‘window of opportunity’ closed [i.e. we would only have the 700 MHz NMR instrument available to us for a few days]. Unfortunately, during the days between collection of the initial spectrum and the time when I had more monomeric ZW10 available, the majority of the protein in my original NMR sample had come out of solution, and the majority of the ZW10 protein in
that initial NMR sample had become disulfide-linked and oligomeric. Indeed, there seem to be some cysteine residues on the surface of full-length ZW10 that render the full-length version of the protein difficult to work with.

**Outlook.**

Preliminary NMR data provides evidence of binding between IC:1-143 and full-length ZW10 protein, with an initial pattern of peak disappearances reminiscent of that seen for the binding of p150Glued to IC:1-143 (Morgan *et. al.* 2011). In addition to the evidence of binding from NMR, Native PAGE analysis and GST-pulldown assays have both demonstrated binding interaction between IC:1-143 and full-length ZW10 protein (data not shown).

Under ideal circumstances, it may be possible to perform a complete NMR titration of ^15^N IC:1-143 with unlabeled full-length monomeric ZW10 protein, but all steps of the titration would probably need to be performed within the same day, given the difficulties encountered with the full-length ZW10 protein. However, I would point out that, in mapping the binding interface of IC with p150Glued via NMR, full-length *Drosophila melanogaster* p150Glued protein (consists of 1,265 amino acid residues) was not used; rather, a 289-residue fragment of p150Glued was used—a fragment which had previously been determined (by more coarse-grained methods) to contain the binding site for IC. For the NMR mapping of the IC binding interface for NudE, a 174-residue fragment of NudE was used, not the full-length protein (Nyarko *et. al.* 2011). Thus, it would not be unreasonable to try to find and use smaller functional fragments of the 721-residue ZW10 protein, rather than working with the full-length version. One benefit in identifying smaller soluble and functional pieces of ZW10, is that the total number of cysteine residues will be reduced in the individual pieces, and the particular cysteines that caused disulfide-linked oligomerization (and subsequent protein precipitation) in the full-length protein, might not be present in the fragment of ZW10 that might ultimately be used for NMR titration and other experiments where higher protein concentration and greater sample stability are required.

Potential domains of ZW10 might be identified through use of bioinformatics tools. Alternatively, limited proteolysis might be another route to identify domains or smaller soluble pieces of ZW10. Binding activity of smaller ZW10 fragments with IC could initially be determined using relatively simple coarse-grained techniques, such as simple PAGE analysis or affinity chromatography (such as GST-pulldowns); these types of coarse-grained approaches were used in the initial identification of minimal interacting domains of dynein IC and dynactin.
p150\textsuperscript{Glu} (\textit{cf.} Vaughan and Vallee 1995, Table 1).

\textbf{Acknowledgements}

The author would like to thank Dr. Nazish Abdullah for technical advice on the removal of unspliced introns from genomic template DNA. The ZW10 gene clone was obtained from the Drosophila Genomics Resource Center (Bloomington, IN), supported by NIH grant 2P40OD010949-10A1.
Appendix 3

Preparation, characterization, and crystallization of the dynein light intermediate chain (LIC) protein from *Chaetomium thermophilum*.

Jessica L. Morgan
Figure A3.1. Expression and solubility of the *Chaetomium thermophilum* dynein Light Intermediate Chain (‘CT-LIC’) protein. (A) SDS-PAGE gel (12%) of CT-LIC (in a pET15b expression plasmid transformed into Rosetta BL21(DE3) *E. coli* cells) expression in LB media; induction was at OD$_{600nm}$ ~ 0.4 with +0.4 mM IPTG, induction at 17°C for 17.0 hours. (B) Solubility test samples of CT-LIC from the large-scale 17°C culture. (C) SDS-PAGE gel (12%) of Ni-NTA purification fractions collected from the large-scale CT-LIC culture.
Figure A3.2. Size exclusion chromatography (SEC) of the CT-LIC protein. (A) Preparative scale SEC purification of CT-LIC on a Superdex200 column using flow rate = 60 mLs/hour and a running buffer of 20 mM Tris.Cl, pH 7.5 with 200 mM Sodium Sulfate and 1 mM NaN$_3$. The purest fractions of CT-LIC were collected and pooled and dialyzed into crystallization buffer [12.5 mM Tris.Cl, 5 mM NaCl, 1 mM NaN$_3$, pH 7.5]. (B) Analytical-scale SEC (Superdex75 resin) chromatogram of the CT-LIC protein sample to be used for crystallization screening. Vertical blue line indicates the mean elution time of a BSA standard (~67 kDa molecular weight). CT-LIC has predicted molecular weight 61.6 kDa. These chromatography results suggest that the LIC protein from this species (Chaetomium thermophilum) is monomeric—in contrast to the dimeric LIC protein found in mammalian species.
Figure A3.3. Far UV-CD spectroscopic characterization of CT-LIC thermal denaturation. Spectra were acquired at the indicated temperatures for a 4 μM protein sample in 12.5 mM Tris.Cl, 5 mM NaCl, pH 7.5; 0.1 cm cell path length. Values plotted are those with HT[V] values of 600 or less.
Figure A3.4. Far UV-CD thermal unfolding profile of CT-LIC (red, left ordinate). The ratio $\Theta_{222\text{nm}}/\Theta_{208\text{nm}}$ (blue) for data presented in Figure A3.3 is plotted along the secondary ordinate (right-hand side).
Figure A3.5. Time-series of crystal formation of the CT-LIC protein. Crystallization buffer conditions: 0.1 M MES pH 6.0 + 20% w/v PEG-6,000 @ 20°C. The time series is from left-to-right, top-to-bottom, and spanned approximately 1 month of time. These crystals diffracted to 2.7 Å and had space group C222_1.
Appendix 4

NMR backbone chemical shift assignments and dynamics of the *Drosophila melanogaster* dynein LC7 protein—free and IC7-bound state NMR data

Jessica L. Morgan
Figure A4.1. $^{13}$C ($C_\alpha$, $C_\beta$) strip plot for residues Gln42–Gln49 taken from the HNCACB experiment collected for *D. melanogaster* LC7. Data were recorded at 30°C in 12.5 mM Tris•Cl (pH 7.5), 5 mM NaCl, 1 mM NaN$_3$. 
Figure A4.2. $^1$H–$^{15}$N HSQC spectrum of the *Drosophila melanogaster* dynein LC7 protein with backbone chemical shift assignments. The spectrum was recorded at 30°C in 12.5 mM Tris·Cl (pH 7.5), 5 mM NaCl, 1 mM NaN$_3$. 
Figure A4.3. $^1$H–$^{15}$N HSQC spectrum of the *D. melanogaster* dynein IC-bound LC7 with backbone chemical shift assignments. The spectrum was recorded for $^{15}$N-labeled LC7 with excess unlabeled ‘IC’ (e.g. IC:212–260) at 30°C in 12.5 mM Tris-Cl (pH 7.5), 5 mM NaCl, 1 mM NaN$_3$. 
Figure A4.4. NMR dynamics data collected for apo (A–C) and IC7-bound (D) LC7. Plots of $^1$H–$^{15}$N steady-state heteronuclear NOEs (A and D), $^{15}$N $T_2$ (B), and $T_1$ (C) values for apo $^{15}$N LC7 (A–C) or $^{15}$N LC7 with excess unlabeled IC7 present (D); data were recorded at 30°C in 12.5 mM TrisCl (pH 7.5), 5 mM NaCl, 1 mM NaN$_3$. 