AN ABSTRACT OF THE DISSERTATION OF

Nathan E. Jespersen for the degree of Doctor of Philosophy in Biochemistry and Biophysics presented on September 10, 2019.

Title: Viral Targeting of Host Hubs: Interactions between the Rabies Virus and LC8

Abstract approved:

_____

Elisar J. Barbar

As obligate intracellular parasites, viruses rely on host machinery for their own reproduction. Viruses are therefore required to interact with a wide variety of host proteins, despite limitations in viral genome size. The most parsimonious method of success is to hijack central and essential proteins, known as hub proteins. LC8 is a notable example of a hub protein, and has been shown to interact with more than 100 eukaryotic partners in various unrelated pathways. An increasing number of studies have noted interactions between LC8 and viruses, including the Ebola, rabies, and rotaviruses; however, LC8's role within these systems is unclear.

This thesis examines the structural and functional properties of hub proteins, with a particular focus on LC8 and its interactions with the rabies virus. Three chapters of original work include two primary research reports and one review/opinion piece. The first research report, Chapter 2, structurally characterizes the interaction between LC8 and the rabies virus phosphoprotein (RavP). We use insights gleaned from our structural studies to predict and test potential roles for LC8 in the rabies virus infection cycle, and demonstrate that LC8 is important for efficient viral polymerase activity. Chapter 3 is an in-depth study of the LC8 recognition motif, where we examine both the structural and functional plasticity of LC8, and identify and validate many new LC8 interactions. We also develop a tool that can be used to predict LC8 binding motifs in proteins of interest, which will greatly improve the ability of those outside of the LC8 field to recognize, test, and validate partner proteins. In Chapter 4 we present a variety of new ideas about what qualities describe a linear motif-binding hub

protein. This work provides ideological and linguistic suggestions for important structural and functional features of hubs that underlie their plasticity. It then further describes how viruses take advantage of these features to more efficiently hijack host pathways. Finally, Chapter 5 discusses the impacts of my thesis work, and outlines some potential future projects. Each individual chapter builds on the previous one to create an expanding view of the importance of linear motif-binding hubs for infectious viruses.

Viral Targeting of Host Hubs: Interactions between the Rabies Virus and LC8

by
Nathan E. Jespersen

A DISSERTATION

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented September 10, 2019
Commencement June 2020

Doctor of Philosophy dissertation of Nathan E. Jespersen presented on September 10, 2019

APPROVED:

Major Professor, representing Biochemistry and Biophysics

Head of the Department of Biochemistry and Biophysics

Dean of the Graduate School

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

Nathan E. Jespersen, Author

ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Elisar Barbar, for all of your insight, assistance, mentoring, and prodding over the years. I have no idea how many hours I spent chatting with you about whatever new theory was on my brain, and your enthusiasm in those situations was always so wonderful. Sarah and I used to joke about how if we were ever feeling nervous during a presentation, we could glance over at you and see you smiling and supportive, and it would perk us right up. Your willingness to let me look into a new idea allowed me to get a very broad skillset, and you always encouraged me to consider some grant or fellowship. During the few occasions where I was feeling really down about science, you were marvelously supportive. I am super fortunate to have had you as an advisor!

I have had some terrific collaborators over the years, many of whom are my French compatriots, including Dr. Marc Jamin, Dr. Cedric Leyrat, and particularly Dr. Danielle Blondel and everyone else in Gif-sur-Yvette. You all opened up a whole new set of experiments to me, and were *extremely* patient (especially considering my very poor French). I'm not sure how my RavP work would have come together without you.

The Biochemistry and Biophysics Department at OSU is great. Regardless of the professor, I was never reticent to go by their office to pester them about science. I would say the best thing about graduate school is that there is *probably* somebody right down the hall who knows a lot about any given subject, and our department is tight-knit enough to swing by for a chat.

I would like to thank all of my committee members. Dr. Theo Dreher, your virology course was one of my favorites I have ever taken, and I think that your input during my prelims strongly impacted the way my future experiments proceeded and my success with the Chateaubriand Fellowship. My initial GCR, Dr. Sandra Loesgen, Thanks for taking the time to sit down with me and iron out some of the finer points of my fellowship application. Dr. Brian Dolan, you saved the day by agreeing to be my GCR at the last moment! Dr. Afua Nyarko – not only have you mentored me as a member of my committee, but you also taught me many of the wet-lab protocols that I still follow. Even though I've had to resort to trickery to learn anything about your life, I still really appreciate all of our interactions. Dr. Ryan Mehl, I've had so many great conversations with you about science, life, and work. I often sought you out when I was feeling particularly uncertain, and you always lent me an ear and gave me some good advice. Thanks!

The Barbar Besties (otherwise known as the current Barbar lab members Aidan Estelle, Kayla Jara, and Heather Forsythe) are quite a group. I can't say I fully understand most of the conversations that happen in the lab (What's Aidan hating on now? Pomegranates?), but it's always entertaining! Sanjana Saravanan, you were the best undergrad I ever had. Patrick Reardon, you don't *quite* go in the Barbar lab section, but with the amount of time you spend in our lab helping us with NMR questions, I think it's pretty appropriate. My ol' teammate Dr. Sarah Clark, with whom I have so many great memories in the lab and outside of it. Whether it was commiserating or co-celebrating, it was always better with you!

I also want to thank my family and friends (some of which are included in other categories). Mom and Dad, you guys should win awards for selfless support. I know I moved across the country and don't get home that frequently, but you guys (almost) never complain about that, and always make sure to remind me how proud of me you are. It makes me feel nice. Amongst my friends, I'd like to give a particular shout-out to Dr. Andrew Brereton, who is one of the most interesting people to chat with I have ever met. Our graveyard of forgotten theories would be something to behold! Last but not least, my wonderful and intelligent wife Michelle. Your calm mien relaxed me when I was manic, and your dedication to sleep kept me healthy. Without your regular and on-point reminders, who knows how many things would have been left undone.

CONTRIBUTION OF AUTHORS

Cedric Leyrat contributed to the molecular dynamics simulations in Chapter 2. Marc Jamin, Jeam-Marie Bourhis, and Francine Gérard contributed to SAXS data collection and analysis in Chapter 2. Danielle Blondel guided in cell experiments in Chapter 2. Ylva Ivarsson and Cecilia Blikstad performed phage display experiments from Chapter 3. Anna Akhmanova and York-Christoph Ammon collected cell images shown in Chapter 3. Aidan Estelle contributed to the writing and production of the LC8Pred algorithm in Chapter 3. Sarah Clark contributed to figure design throughout my thesis. Elisar Barbar was involved in all steps of this thesis work.

TABLE OF CONTENTS

Page

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

Viral Targeting of Host Hubs: Interactions between the Rabies Virus and LC8

Chapter 1

Introduction

Dynein Light Chain 1: More Than a Motor Protein Subunit

History and Overview

Dynein light chain 1 (LC8) has a long and often revised history. It was first described as a component of axonemal dynein in 1982, where it was purified from *Chlamydomonas reinhardii*[1]. The name LC8 is derived from this study, as the light chain migrated consistent with an 8 kDa protein on an SDS-PAGE gel (it is actually 10 kDa). It has gone by various other names, such as Dyn2 in yeast, Cut-up in *Drosophila*, dlc1 (dynein light chain 1), and PIN (protein inhibitor of nNOS[2], though it is no longer considered an nNOS inhibitor[3]). LC8 was purported to be simply a component of axonemal dynein within cilia and flagella until 1995, when it was published that LC8 is conserved in highly divergent organisms, including plants, which lack dynein[4,5]. It is now known that LC8 is present in all examined eukaryotes[6], including plants, and is highly conserved in sequence[5]. For example, the unicellular eukaryote *Giardia lamblia* encodes an LC8 that is 93% sequence identical to human LC8, differing in only 6 of 89 positions. The presence of LC8 in organisms lacking dynein suggests an additional function for the light chain.

LC8 also interacts with myosin Va[7], another motor protein involved in intracellular transport along actin filaments, as well as a variety of proteins demonstrably transported along actin and/or microtubules[8–10]. This led to a hypothesis that LC8 was acting as a cargo adaptor, connecting the motor proteins to a wide variety of different cargoes[11–14]. In support of this idea, structural analyses of LC8 indicate that it forms a stable dimer under cellular conditions, with an association constant of around 80 nM[15,16]. The dimeric structure creates two symmetrical binding grooves, which in theory could allow LC8 to bind to both dynein and cargo concurrently. However, the homodimeric nature of dynein and myosin Va, as well as the fact that there are 2 identical binding sites on the LC8 homodimer, suggested that LC8 was instead facilitating dynein dimerization[17]. Further studies demonstrated that there is a significant pool of cytoplasmic LC8 not associated with motor proteins[18], while structural and thermodynamic analyses indicated that LC8 bound to both chains of the dynein intermediate chain contemporaneously (Figure 1.1A)[19,20]. In another experiment, a synthetic peptide capable of binding LC8 was attached to a fluorescent label and examined for localization with LC8 and dynein. Although the peptide bound to LC8, and a pool of LC8 was localized with dynein, these were separate pools, suggesting that LC8 does indeed have functions irrespective of dynein[21].

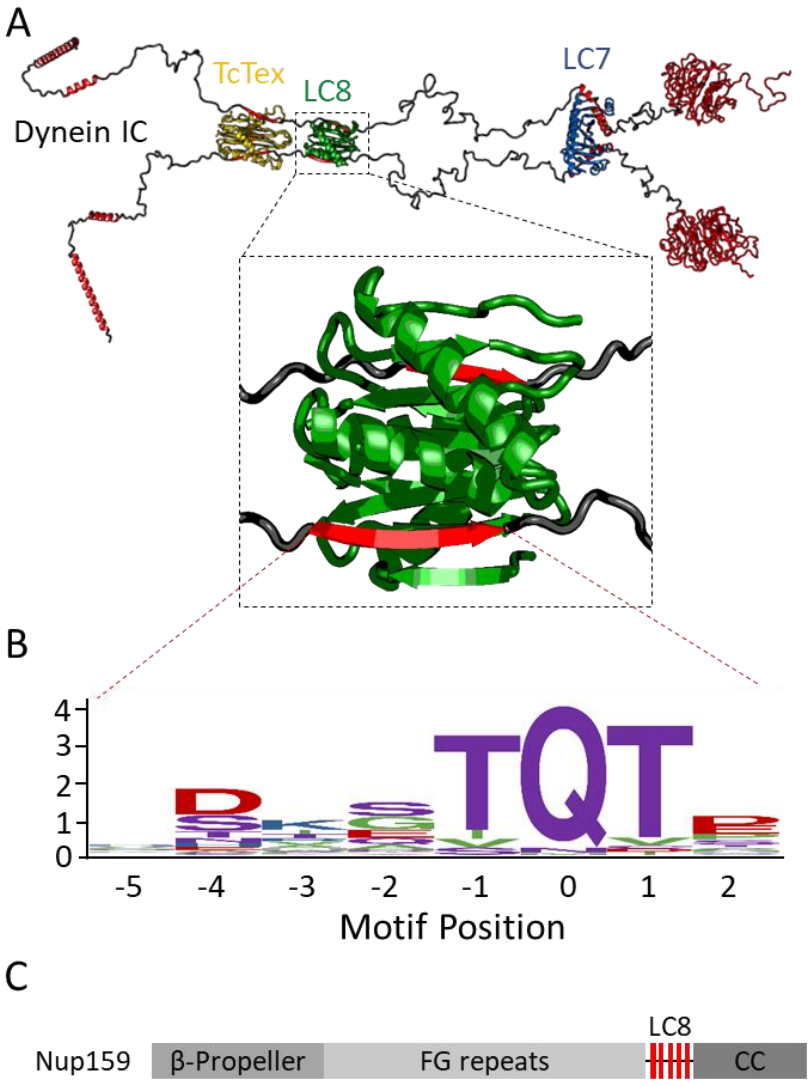


Figure 1.1: The dynein light chain 8 structure and interactions.
(A) The dynein intermediate chain complex is composed of two copies of each: IC, LC7, TcTex, and LC8. Interactions between LC8 and IC take place at the dimer interface, where IC forms a β-strand upon binding (expansion). (B) Amino acid enrichment is shown for each position within the LC8 binding motif, calculated from 79 known binder motifs on the LC8 database (http://lc8hub.cgrb.oregonstate.edu)[22]. Amino acid heights represent relative enrichment of amino acids. (C) Domain architecture for Nup159, denoting the presence of 5 LC8 motifs within a short region between the C-terminal coiled-coil domain, and the Phe-Gly repeat region[23].

Structural studies on interactions between LC8 and the bicoid RNA-binding protein Swallow are some of the first to suggest a role for LC8 outside of the dynein complex[24]. Although it was still unclear at this point, the Barbar lab noted that LC8 bound to IC and Swallow at the same interface[25]. Additionally, LC8 seemingly stabilized the dimeric structure of both proteins, by facilitating interactions at downstream self-association domains. On the cellular side, yeast nucleoporin Nup159-LC8 interactions are an early example of LC8 function separate from motor proteins. Nup159 is an essential structural component of the cytoplasmic face of the nuclear pore complex (NPC)[26], and is involved in nuclear export of preribosomes, as well as mRNA[27,28]. Interestingly, it contains five LC8 binding sites concentrated within a 100 amino acid stretch of the 160 kDa protein[29]. Oligomerization and phase separation of the NPC is facilitated by FG repeats and a coiled-coil domain that flank the LC8 binding region (Figure 1.1C)[30]. Deletion of the LC8 binding region inhibits dimerization of the Nup159 complex, and deletion of LC8 *and* Nup159 complex components are synthetically lethal[29]. On the other hand, deletion of the dynein heavy chain did not show any synergistic growth retardation with Nup159 complex proteins. The discovery of LC8 within the nuclear pore complex reinforced the hypothesis of LC8 as a "molecular glue,"[29] "molecular Velcro,"[31] or "dimerization engine"[32]. In essence, LC8 functions by sticking two monomers together and stabilizing the dimeric form of its partners.

As the number of LC8 binding partners increased (now with more than 100 described interactions[22]), this hypothesis of molecule glue further evolved to define LC8 as a central hub protein, involved in a wide variety of different cellular pathways[32]. However, despite a preponderance of evidence, it is still quite common for papers to suggest that LC8 is a link to dynein, and therefore a mechanism of transport[33,34].

LC8 Structure

There are 22 solution and crystal structures available for LC8 in the apo state, or bound to various peptides[20,35–44]. LC8 forms a 21 kDa dimer in solution, and is only known to form a monomer at low pH[15]. It is part of a unique fold family, and the dimeric structure contains two five-stranded anti-parallel β-sheets (Figure 1.1A). Each β-sheet contains four strands from one protomer, and one strand from the other protomer. On the opposite face of the dimer interface is a pair of α-helices. Because the binding groove is localized to the dimeric interface, monomeric versions of LC8 are inactive and incapable of binding to partners[45].

LC8 interacts with an eight amino-acid recognition motif, also termed a linear motif (LM; analysis and characterization of this motif is the basis of Chapter 3)[22]. Sequences bound to LC8

form a single β-strand structure integrated into an LC8 antiparallel β-sheet[35]. While there is some variation in the binding motif, it is most frequently anchored by a TQT sequence[44]. The TQT anchor is highly enriched among known LC8 partners and is therefore termed the "motif anchor" (Figure 1.1B)[44]. Interestingly, the vast majority of LC8 binding partners are significantly enriched in disorder, and almost all known interactions occur in so called intrinsically disordered regions[46].

Intrinsic Protein Disorder

The unfolded state of a protein has often been considered as an intermediate on the path to a folded and therefore functional conformation. However, within the past two decades it has become apparent that protein intrinsic disorder, i.e. a lack of stable secondary or tertiary structure, is of paramount importance for a wide variety of functions within a plethora of systems. It is now estimated that approximately 30% of eukaryotic proteins contain intrinsically disordered regions (IDRs) that are over 30 amino acids long[47]. Interestingly, the fraction of disorder within the proteome varies between domains of life, with 7% of residues within IDRs for archaea, 8% in bacteria, 13% in viruses, and 20% in eukaryotes[48–50]. IDRs are characterized by low sequence diversity, an abundance of charged residues, and few hydrophobic residues, resulting in open and highly flexible structures best described by an ensemble of conformations. For proteins containing IDRs, known as intrinsically disordered proteins (IDPs), these regions confer a variety of benefits. Firstly, IDRs are highly accessible, meaning that they are ideal locations for binding interactions or impactful post-translational modifications[47]. Secondly, the flexibility in IDRs allows a protein to form a variety of conformations, avoiding undesirable steric clashes and conserving entropy[51]. Thirdly, due in part to their highly charged physiochemical properties, as well as the abundance of short linear motifs, disordered regions are much more promiscuous in their binding interactions, and greatly facilitate the formation of large, complex assemblies[52,53]. Finally, while ordered proteins are greatly constrained by the requirement for residues that conserve their 3-dimensional structure, disordered proteins are under no such pressures. They therefore display much higher mutation rates, providing an efficient means for evolutionary adaptability[54].

These functional benefits have led to the presence of IDPs within essentially all cellular pathways and compartments; however, they are most commonly found in signaling pathways[55] and regulatory processes[51], and are significantly enriched in the nucleolus, cell-cell junctions, and cytoskeletal macromolecular complexes[56]. IDPs are also the dominant protein class for non-

communicable diseases[57,58]. IDP misregulation or aggregation can result in Alzheimer's disease, associated with the beta-amyloid[59] and tau proteins[60,61], or breast cancer, associated with BRCA1/2[62]. Surprisingly, despite their prevalence and importance to biology, IDPs are perhaps the most poorly understood system, mechanistically, in all of molecular biology. This is due to inherent difficulties with structural characterization and restrictions in the available methods to provide quantitative descriptions of dynamics/heterogeneity. Therefore, any studies on IDPs require a combination of complementary methods in order to efficiently and accurately model IDP conformational ensembles.

LC8 selectively binds to the disordered regions in other proteins in part because of the increased accessibility of motifs within this region; however, that does not explain the benefit of LC8 interactions for these multitudinous partners.

LC8 Function

There are more than 100 experimentally verified binary interactions for LC8[22]. Partners include a large number of IDPs performing multifarious functions, including intracellular transport (dynein intermediate chain)[25,63,64], nuclear pore formation(Nup159)[65], host-virus interactions (rabies phosphoprotein)[66], and transcription (ASCIZ)[67–69]. Additionally, LC8 is highly expressed across a wide variety of cell types[70], and LC8 and its partners are broadly distributed within individual cells (Figure 1.2) [22,71]. Studies on the precise importance of LC8 in various cellular functions have been hampered by the fact that LC8 knockouts are lethal, leading to mortality in mice early in embryogenesis from incomplete embryonic turning, and severe neural tube and cardiac defects[72]. Comparable experiments in *Drosophila melanogaster* and *Caenorhabditis elegans* were also embryonically lethal, or led to severe pleiotropic effects[73,74]. Fortunately, recent work by King *et al*.[72] has succeeded in generating a limb bud mesoderm-restricted conditional knock-out of LC8, resulting in the first available animal (and cell-line) model. Although these cells have not been thoroughly studied, preliminary analyses indicate that knockouts of LC8 lead to impairments in the Hippo signaling pathway[72], amongst other things. This is consistent with previous data suggesting a strong enrichment in LC8 binding proteins within this pathway[75]. Future work will take advantage of these cells and model organisms to acutely understand the role LC8 plays within different pathways.

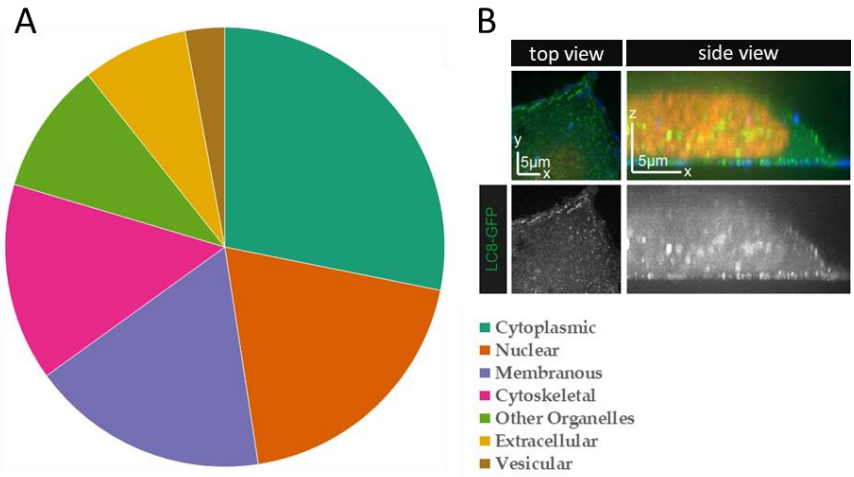


Figure 1.2: LC8 and its binding partners display broad cellular localization.
(A) Localization information derived from the COMPARTMENTS program demonstrates that LC8 binding partners are localized to all cellular compartments. (B) Live HeLa cells stably expressing LC8-GFP (green) were transiently transfected with the focal adhesion marker TagBFP-paxillin (blue) and the nucleus marker dsRed-tagged histone H2B (red). The top view images shown on the left represent an optical section located next to the coverslip. LC8-GFP is present throughout the cell, but forms puncta at the cell cortex.

The fact that LC8 is active in a variety of unrelated pathways led to the hypothesis that LC8 acts as a central "hub" protein[32]. Unlike most proteins, which have a small number of interaction partners, hub proteins sit at the center of highly complex interaction networks[76]. Hubs are typically classified as either static (or "party" hubs) and dynamic (or "date" hubs). Party hubs bind a large number of partners simultaneously at different sites, for example BRCA2[77], while date hubs typically bind to linear motifs within IDRs, and have multiple partners that compete for the same site[78,79]. Well-known examples of date hubs include the calmodulin and 14-3-3 proteins [80–82]. LC8 also falls into this category. A description of some defining features for linear motif-binding date hubs, particularly LC8, is the subject of Chapter 4.

LC8 is non-enzymatic, therefore its role in complex assembly is purely structural. There are three different structural roles that LC8 plays within the area of macromolecular assembly. These roles are outlined in Figure 1.3. Known examples include situations where LC8 pulls two partner protomers together and facilitates dimerization via a downstream self-association domain (as in dynein)[17,25,83], or coiled-coil (as in the protein swallow)[20,24,84]. It can also act as staples, by binding to multiple sites on a partner (as in Nup159 or Chica)[44,65]. There are no known instances

of it binding to two different sequences or binding out of register in multivalent binding interactions. Regardless of the structural role, LC8 always stabilizes its partners and facilitates their interactions with downstream partners[65,85,86].

It was first noted in 2000 that LC8 was appropriated by the rabies virus for an unknown function[66,87]. Since then, there have been a steadily increasing number of viruses reported to hijack LC8, such as the human immunodeficiency virus[88], the African swine fever virus[89], the Ebola virus[90], the human papillomavirus[91], and the porcine circovirus[34]. It was common for these studies to predict a role for LC8 in intracellular trafficking of viral components through dynein[34]; however, work on the rabies virus demonstrated that there were no direct links between the virus, LC8, and dynein[92]. Further description of this interaction, as well as an examination of the role for LC8 in the rabies virus infection cycle is the subject of Chapter 2.

LC8 Feedback Regulation

One of LC8's most interesting traits is its ability to self-regulate its intracellular concentrations via direct interactions with the transcription factor ASCIZ (ATMIN-Substrate Chk-Interacting Zn^{2+} finger)[68,69,93]. The human version of ASCIZ is an 88 kDa protein that, astoundingly, contains 11 different LC8 recognition motifs located within its disordered C-terminal tail[69]. *In vivo* knockouts of ASCIZ display a phenotype remarkably similar to that of LC8 knockouts (*i.e.* death during embryogenesis with severe developmental defects)[67,72,94], with the main difference being a slight attenuation in the deleterious effects (late embryogenesis death as opposed to early embryogenesis). This suggests that the main cellular role for ASCIZ is to regulate LC8 concentrations[72].

Although ASCIZ encodes 11 LC8 binding motifs, biophysical analyses of the ASCIZ-LC8 interaction identified multiple stable complexes, including a fully occupied and a partially occupied state[69]. *In vivo* experiments demonstrated that overexpression of LC8 leads to a decrease in its own transcription via an ASCIZ-mediated pathway[68]. Taken together, these data suggest that ASCIZ acts as a sensor of LC8 concentrations, and regularly adjusts transcription of LC8 to fit the cell's needs. We believe that this self-regulatory pathway is particularly important for linear motif-binding date hubs like LC8, and expand upon this idea in Chapter 4.

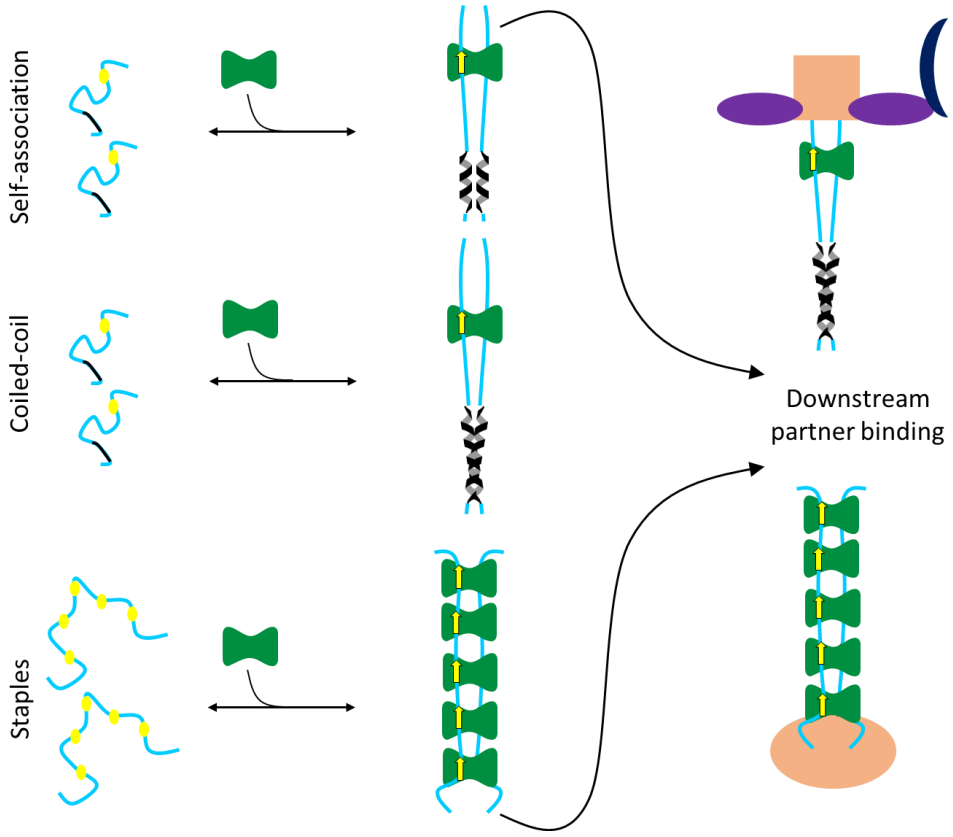


Figure 1.3: Structural roles of LC8 interactions.
Interactions between LC8 dimers (green) and partners lead to stabilizing changes in partner proteins. Many partners have a self-association domain or coiled-coil region proximal to the LC8 motif, and binding by LC8 leads to formation of this structure. On the other hand, proteins with multiple motifs in close proximity are sutured together by LC8 interactions, resulting in a more rigid, ladder-like structure (as determined by Cryo-EM)[69]. Regardless of the type of structural changes, the result is stabilization of a larger complex, often facilitating interactions with various downstream binding partners via the bivalency effect (orange, purple, dark blue shapes).

The Rabies Virus

History and Epidemiology

The Laws of Eshnunna are some of the oldest known writings in the world, dating back to approximately 1950 B.C. They dictate that if a man owns a "mad" dog, and it bites another man and kills him, then the owner must pay 2/3rds of a mina in fines (estimates vary dramatically, but this equates to 6 months to 4 years-worth of wages for a typical agricultural worker)[95]. This suggests that even in ancient times, people were aware of the *effects* of the rabies virus, and that they knew the madness could somehow be passed from dogs to humans. In fact, the name rabies, as well as the name of the rabies genus (*Lyssavirus*) originated from the word for anger and madness. Lyssa is classically the Greek primordial spirit of rage and frenzy.

Zoonotic pathogens, meaning those that can be passed across the species barriers, are responsible for some of the most deadly and infectious diseases. They are also often significantly more difficult to control than host-specific pathogens due to the unpredictability in their transmission and their ability to escape into other species. The frequency with which they cross the species barrier is highly variable, and some viruses, like the human immunodeficiency virus (HIV), permanently adapt to a new host after zoonosis. Others, like the rabies virus, play jump rope with the barrier and constantly move between host organisms. Although rabies is often associated with either bats or dogs, any warm blooded animal can theoretically contract the disease. There have even been cases of rabid seals and polar bears[96]. It is therefore logical that rabies would interact with highly conserved proteins like LC8, which would be accessible in a variety of different hosts.

Symptoms of a rabies virus infection are numerous. The more dramatic ones include aggression, resistance to pain, paralysis, hypersensitivity, photosensitivity, hypersexuality, insomnia, and hydrophobia (or a fear of water), which is thought to be unique to rabies[97,98]. These vivid symptoms have cemented its place in the cultural mind and may have spawned the zombie, werewolf, *and* vampire mythos (although this is pure conjecture)[99]. It was once said that "if a man could look at his reflection, then he is not rabid," which may be the origin for vampires classically not having a reflection[97]. Our long history with the virus has resulted in a correspondingly long time researching it, and work by Louis Pasteur led to one of the earliest successful vaccines nearly 150 years ago[99]. Despite the fact that there is a very effective vaccine, as well as treatment options via a post-exposure prophylaxis, there are still approximately 55,000 annual deaths from rabies[100]. This is due, in part, to the virus's unusual transmission and infection cycle.

The path the rabies virus (RAV) takes through the host can be broken up in 5 steps (Figure 1.4). 1) An infected mammal bites another mammal. The rabies virus is highly concentrated in saliva and salivary glands during late phases of the infection, therefore the most common form of transmission is via bite[100]. 2) RAV enters the new host and localizes to exposed muscle cells. Evidence supporting muscle cells as the primary site of infection comes from studies demonstrating RAV interactions with a nicotinic acetylcholine receptor (nAchR) restricted to muscle cells[100–102]. 3) RAV buds from muscle cells into the neuromuscular junction, and enters the peripheral nervous system (PNS) at motor neurons[103]. The virus is considered a 'neurotropic' virus, meaning it preferentially moves through neuronal cells. 4) RAV is transported in a unidirectional, retrograde fashion across the axons of neuronal cells, through the PNS and into the central nervous system (CNS). During steps 2-4, infected individuals are completely asymptomatic, and RAV is non-neurotoxic (does not kill neuronal cells) [100,103]. 5) Once the virion reaches the brain, there is a change in viral replication and tropism caused by an unknown stimulus, leading the virus to radiate out from the brain, and into peripheral motor and sensory neurons. Eventually, RAV is localized to hair follicles, the heart, salivary glands, the tongue, adrenal glands, and the skin[103–106]. At this stage, the individual begins to show symptoms, which typically facilitate spread of the virus. For example, the hydrophobia is caused by painful spasms occurring when the infected individual considers drinking[107]. This is believed to increase transmission, because virus in the salivary glands is allowed to accumulate, rather than being washed away. The infection cycle then repeats.

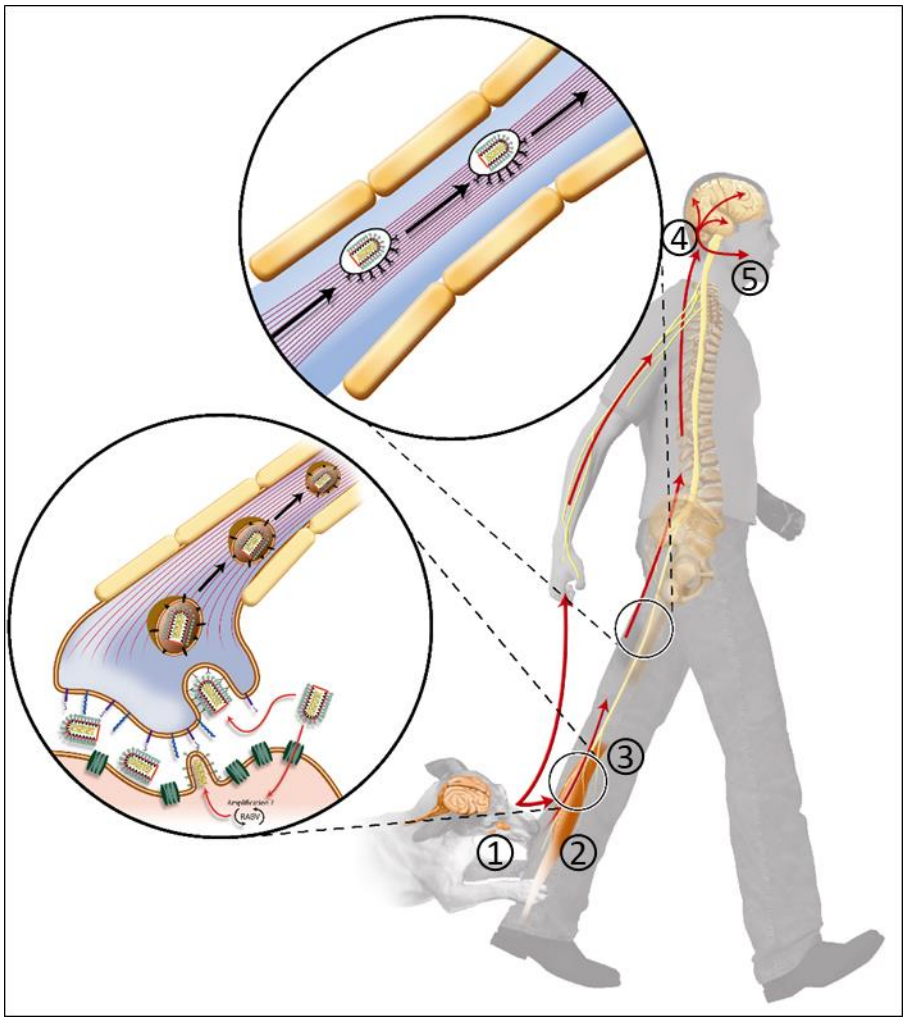


Figure 1.4: The rabies virus path of infection.
RAV infections begin with exposure of muscle tissue to RAV by an animal bite or scratch (1). The virus proliferates in muscle cells (2), and infects the peripheral nervous system through neuromuscular junctions (bottom inset, 3). Virions move through neurons via dynein-mediated retrograde transport (top inset, 4), and upon reaching the brain, radiate out, infecting the entire CNS and localizing to peripheral sensory and motor neurons (5). This figure has been adapted with permission from[100].

There is a huge amount of variation in the timespan of the transmission cycle, with reports typically ranging from two weeks to two years[98]. There has even been one report of a 25 year incubation period[108]. The long incubation time, neurotropism, and non-cytotoxic nature of RAV make it one of the rare viruses that can be treated via a post-exposure vaccination (prophylaxis); however, the virus has a horrifying ~100% lethality rate in humans once symptoms arise. Interestingly, it has been noted during autopsies of RAV victims, that their neuronal cells are still functional, and the virus is beginning to be cleared. This suggests that their immune system eventually responded, but too slowly[109,110]. After RAV arrival at the CNS, there are no effective methods of treatment. The previously proposed 'Milwaukee Protocol' method, *i.e.* inducing a coma in an attempt to allow the host immune system to catch up, ended in failure[111,112]; however, this has led to the realization that host immune suppression and carefully controlled viral replication are imperative qualities for RAV lethality. Therefore, treatment will require an understanding of not just the transmission cycle, but also information about the structure and interactions of viral proteins.

Structure

The *Mononegavirales* order is made up of 5 families, with many relatively well known and infectious viruses. These include the Ebola, measles, mumps, Nipah, vesicular stomatitis, and human parainfluenza viruses. Viruses in this order are nonsegmented, meaning their genetic material is on a single strand of RNA. They produce only 5-10 mRNA, including one that encodes their own RNA dependent RNA polymerase (RdRp, labeled the "L" protein), and they protect their genetic material by coating the RNA with many virally encoded nucleoproteins (N, Figure 1.5A). A matrix protein (M) surrounds the ribonucleoparticle, and acts as a bridge to viral envelope, containing trimeric glycoprotein (G). Additionally, many *Mononegavirales* viruses encode another essential noncatalytic subunit for the RdRp which is frequently referred to as the phosphoprotein (P). In numerous *Mononegavirales*, such as the Nipah virus, measles, and RAV, P is also responsible for immune suppression via a direct interaction with host Signal Transducer and Activator of Transcription 1 (STAT1)[113]. In RAV, these essential roles in both immune suppression and viral transcription/replication have made RavP a common target for studies, revealing interactions with many host and viral proteins, including host LC8[66,87].

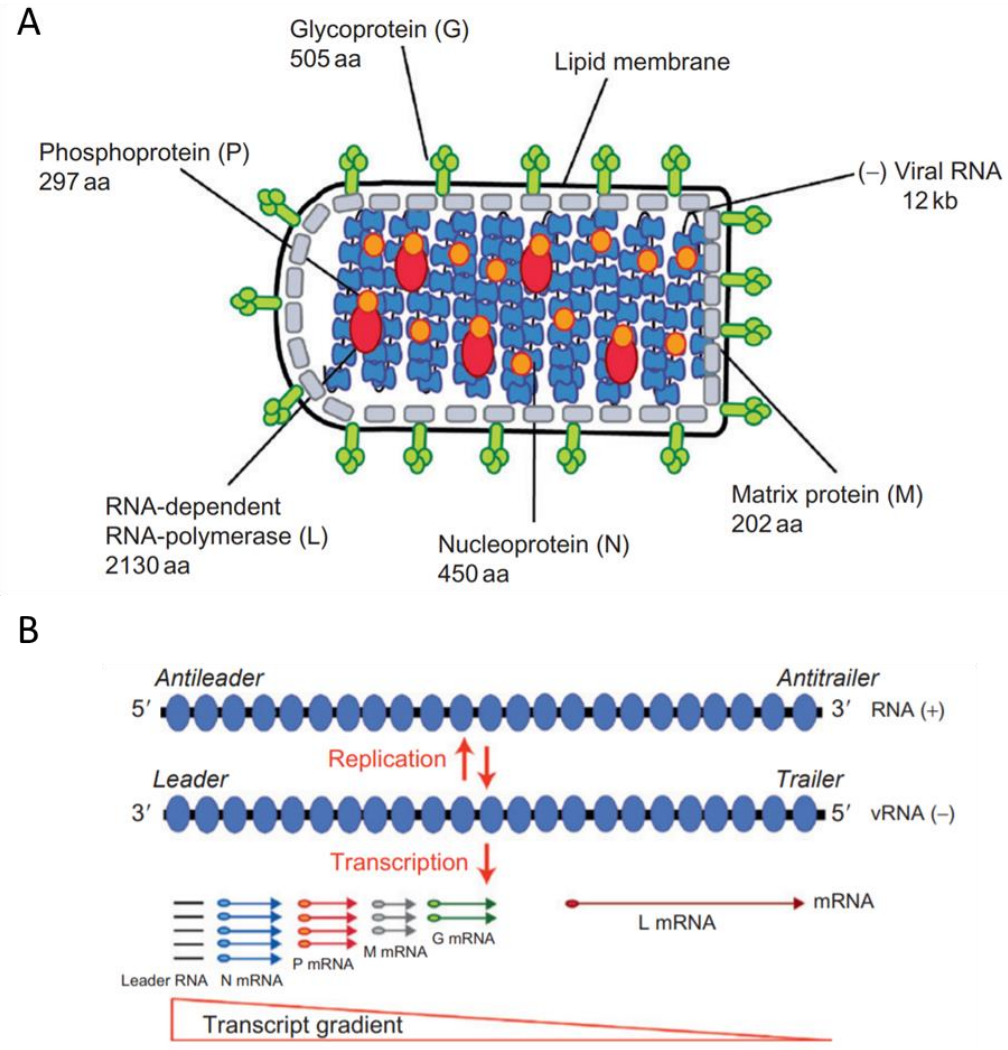


Figure 1.5: Rabies virion structure and transcriptional regulation system.
The rabies virus encode 5 proteins, present in the mature, bullet-shaped virion (A). The relative concentrations of each of these proteins is regulated by the order in which they appear on the viral –ssRNA genome (B). In between each gene is a short sequence where the polymerase complex disengages, and then perhaps (or perhaps not) rebinds. This results in cellular concentrations of rabies proteins decreasing as: N>P>M>G>L. This figure has been revised with permission from [114].

For RAV, replication and transcription of the viral genetic material is carried out by the RdRp complex, composed of L+N+P+RNA. Interestingly, the concentrations of all five RAV proteins (N, P, L, M, and G) are regulated at the transcriptional level, based on the order in which they appear in the viral genome (Figure 1.5B). Protein encoded towards the 3'-end are therefore more highly expressed than those towards the 5' end (N>P>M>G>L). This decline in protein concentrations is caused by an odd (and cool!) feature of the polymerase complex, where it detaches from the RNA at each intergenic region, and may or may not reengage, resulting in fewer transcripts of each gene as the complex moves more 5'[114]. Once a certain threshold of viral M protein has been produced, the polymerase switches from transcription to replication[115].

Rabies and LC8

The rabies phosphoprotein (RavP) contains the 'TQT' LC8 binding motif, and interactions between RavP and LC8 have already been verified[66,116]. Incredibly, an *in vivo* experiment that deleted the LC8 binding motif in RavP resulted in a completely non-lethal virus (as compared to 100% lethality in the wild type virus)[92]. Although the initial hypothesis was that LC8 was important for intracellular transport of viral components via interactions between dynein and LC8, this option was disproved by data suggesting: 1) transport still occurred in a dynein mediated fashion, and 2) virions still travelled up from the inoculation point to the brain in infected mice[66,92,116]. While it is clear that this interaction is important for virus virulence, there is disagreement in the field concerning the role LC8 plays in RAV lethality. In one study, a group used a minigenome system (luciferase gene under control of the viral polymerase complex) to say that LC8 was not involved in viral transcription or replication[66]. However, this was an early version of the minigenome system which lacked internal controls to account for differences in cell concentrations and expression levels. In a second experiment, Tan *et al.*[92] claimed that LC8 did play a role in viral transcription using qRT-PCR, which was unable to distinguish between replication and transcription with their experimental design. A structural analysis of this interaction, as well as an improved methodology to test LC8's role in the polymerase complex, is the subject of Chapter 2.

A recent paper by Bauer *et al.*[117] revealed the presence of a potential LC8 binding site in RavL. In this study, mutations to the LC8 binding site on RavL resulted in attenuated viral reproduction; however, no experiments were run to verify a direct interaction, as opposed to an abrogation in RavL structure. Moreover, the predicted LC8 binding site does not fit accepted

criteria present in all other known LC8 sites[22]. Chapter 2 also covers the role of the RavL motif in LC8 interactions.

Biophysical Techniques

Proteins containing a mixture of ordered and disordered domains are typically not amenable to analysis by any single biophysical method. It is therefore important to use a wide variety of techniques in combination to study protein structure. I was very 'fortunate' in that many of the proteins I worked with were very poorly behaved, requiring me to have a good understanding of diverse methods. Although I have used others during my tenure as a graduate student, the main techniques presented in my thesis are those described below. I will only give a sentence or two description of the method itself since much more detailed descriptions are available online, and then I will describe their utility for my experimentation.

Isothermal Titration Calorimetry (ITC)

ITC is a technique used to measure interactions between molecules. Its basic premise is to measure the heat given off or required for a binding event. This allows one to measure and/or calculate a variety of useful thermodynamic parameters associated with the interaction, such as change in enthalpy, or the dissociation constant (K_d). While this was somewhat useful for the rabies project, it was essential for our ability to test binding between LC8 and a whole slew of peptides (close to 70). In hindsight, perhaps a high-throughput method would have been preferable; however, it was important for us to be able to reliably compare changes in enthalpy upon binding of LC8 to various peptides in order to understand how variable the LC8 binding interaction can be.

Circular Dichroism (CD)

CD takes advantage of the fact that certain molecules differentially absorb right vs. left-handed light. For protein structural work, the great benefit is that various secondary structures (α-helices, β-strands, and random coils) preferentially and differentially absorb right or left-handed light at wavelengths between ~180 nm and 230 nm. This means that one can recognize the relative contributions of each secondary structure to the final tertiary protein structure. I used CD to determine the stability of my constructs, and their similarity compared to the full-length protein, as well as to verify that there were very minor changes in secondary structure upon LC8 binding.

Also, I'm fond of CD for its relative simplicity. If your sample is pure, what you see is what you get, and what you get is what you have.

MultiAngle (Laser) Light Scattering (MALLS or MALS)

MALLS is a method used to determine the molecular weight of a protein/complex using relative scattering of the molecule at different angles. In my work, this was used to determine the multimerization states of RavP and the complex with LC8.

Small-Angle X-ray Scattering (SAXS)

Although I personally did not perform any of SAXS experiments (they were done via collaboration with Marc Jamin), they were quite integral to our findings on RavP-LC8 interactions. SAXS uses elastic scattering of X-rays (at small angles) to provide information about the shape, size, and density of a molecule of interest. In our case, RavP contains a large amount of disorder and was therefore not amenable to crystallization of the full-length protein. On the other hand, the size of the dimer (66 kDa) made NMR description at the high-resolution level infeasible. SAXS is ideal in cases like this, where the protein is a mixture of ordered and disordered regions, and when domain structures are available. In our case, we were able to combine it with molecular dynamics simulations and NMR data to generate real ensembles for RavP and the RavP-LC8 complex.

Nuclear Magnetic Resonance Spectroscopy (NMR)

NMR is our lab's bread and butter. It uses radio waves in the presence of a strong external magnetic field to excite various NMR-active nuclei. The magnitude of the response, as well as the rate at which the excitation decreases, can provide a large amount of insight into the physicochemical environment of a given nucleus. It is easily one of the most effective methods for looking at disordered proteins, because it actually often gives *better* signals for disordered protein regions, and it can provide information about ensemble averages in addition to infrequently occupied states. For me, it was also useful for the study of RavP-LC8 interactions, as RavP was surprisingly NMR-visible for a protein of its size, due to its disordered regions. I also believe that reading and learning about NMR is intellectually the most difficult thing I have studied over the course of my PhD.

Dissertation Contents

This dissertation contains three chapters of original completed work, as well as a series of partially completed projects. Not included are two second author papers on unrelated subjects. The first describes the importance of multivalency in IDP complexes, using LC8 as a particular example system, and is published in *FEBS Letters*[46]. The second describes the structure and interactions of the N-terminal portion of the intermediate chain of dynein from *Chaetomium thermophilum,* and is currently under review at *JBC*[118]. While these projects are not included in my thesis because they lack a direct connection, they certainly impacted the way my research developed and represent a significant investment of time, effort, and a corresponding intellectual growth.

Chapter 2 describes interactions between the rabies virus phosphoprotein and LC8. This is an in depth structural analysis, and is linked to a variety of in cell experiments that I was able to complete in France as a part of the Chateaubriand fellowship. These data show that LC8 restricts the RavP conformational ensemble, facilitating a structure conducive for efficient viral polymerase activity. This work is currently under review at *J. Mol Bio*. It also serves as a specific example of an LC8 interaction with a stable pre-formed dimer, and prepares the reader for the following 2 chapters, which are broader in scope.

Chapter 3 is a manuscript published in *Life Science Alliance*, and looks at all of the traits important to the 'hub' nature of LC8. In addition to describing LC8's structural characteristics, we also identify ~20 novel binding partners, and develop an algorithm that can effectively predict additional LC8 interactions based on sequence alone. This tool is available online, alongside a curated database of known LC8 partners. These data greatly improve our understanding of LC8 and what makes it such an important hub protein.

Chapter 4 is a "Perspectives" or an "idea" paper, which pulls together a wide variety of details about hub proteins in order to describe some of their defining features, such as their dynamic binding interface and self-regulation. We then proceed to describe how viruses take advantage of the highly connected hub proteins in order to efficiently

hijack hosts. While this is not currently published (planned submission to *TIBS* in September), we believe that it will be a significant contribution to our understanding of host-viral interactions.

Chapter 5 summarizes the key findings from the other sections. I also present a wide variety of either partially completed projects, projects in early stages, or ideas that I would consider pursuing if I were at the front end of my PhD rather than the finish.

Chapter 2

The LC8-RavP Ensemble Structure Evinces a Role for LC8 in Regulating *Lyssavirus* Polymerase Functionality

Nathan E. Jespersen, Cedric Leyrat, Francine C. Gérard, Jean-Marie Bourhis, Danielle Blondel, Marc Jamin, and Elisar Barbar

Under review at the *Journal of Molecular Biology*, August 2019

Abstract

The rabies and Ebola viruses recruit the highly conserved host protein LC8 for their own reproductive success. *In vivo* knockouts of the LC8 recognition motif within the rabies virus phosphoprotein (RavP) result in completely non-lethal viral infections. In this work, we examine the molecular role LC8 plays in viral lethality. We show that RavP and LC8 co-localize in rabies infected cells, and that LC8 interactions are essential for efficient viral polymerase functionality. NMR, SAXS, and molecular modeling demonstrate that LC8 binding to a disordered linker adjacent to an endogenous dimerization domain results in restrictions in RavP domain orientations. The resulting ensemble structure of RavP-LC8 tetrameric complex is similar to that of a related virus phosphoprotein that does not bind LC8, suggesting that with RavP, LC8 binding acts as a switch to induce a more active conformation. The high conservation of the LC8 motif in *Lyssavirus* phosphoproteins and its presence in other analogous proteins such as the Ebola virus VP35 evinces a broader purpose for LC8 in regulating downstream phosphoprotein functions vital for viral replication.

Introduction

The replication and transmission of a virus is dependent on a multitude of interactions between host and viral proteins. In general, proteins that are frequently co-opted by viruses are key to viral growth or host defense. One such protein is the highly conserved hub LC8, which binds to multiple viral partners. The rabies virus phosphoprotein[66,87], the Ebola virus VP35[119], and the human immunodeficiency virus integrase[88], represent a small percentage of the more than 100 proteins known or predicted to bind LC8 in a wide array of cellular systems[31,32]. LC8 binds its partners at a well-characterized and conserved recognition motif[32,38,64]. Intriguingly, a null mutation in the LC8 binding motif of the rabies virus phosphoprotein (RavP) results in completely non-lethal viral infections in mice, compared to 100% lethality for rabies infections involving wild type RavP[92].

The rabies virus (RAV) is a nonsegmented, negative-sense single-stranded RNA virus of the *Rhabdoviridae* family and *Mononegavirales* order, which has afflicted humans since before its first mention in the Codex of Eshnunna in 1930 B.C.[95]. Although the disease is preventable with vaccination and treatable with a post-exposure prophylaxis, the virus has a nearly 100% lethality rate once symptoms arise. Impressively, this rate is achieved with only 5 virally encoded genes, including an RNA-dependent RNA polymerase (RavL), a nucleoprotein that encapsulates and protects the viral RNA (RavN), and a non-catalytic phosphoprotein (RavP). RavP contains an LC8 binding motif, and is essential for viral transcription/replication via interactions with RavL and RavN as well as for host immune suppression by inhibiting host Signal Transducer and Activator of Transcription 1 (STAT1)[120–123]. These essential roles make RavP an ideal target for elucidating the importance of LC8 interactions with viral proteins.

LC8 is an 89 amino acid dimeric protein (Figure 2.1A) that is conserved in eukaryotes as evolutionarily distant as *Homo sapiens* and *Giardia lamblia*. Its numerous binding partners include the dynein intermediate chain[63], the transcription factor ASCIZ[68], and the tumor suppressor p53 binding protein[9]. LC8 recognizes a conserved 8-10 amino acid motif in intrinsically disordered proteins (IDPs) and facilitates partner dimerization by forming a tetrameric complex of two IDP chains and a folded LC8 dimer[32]. While LC8 can accommodate a wide variety of residues, binding sequences often contain a KxTQT motif, where x is a small, hydrophobic amino acid[20]. Amino acids within the LC8 binding motif adopt a β-strand conformation, and are integrated into LC8's central β-sheet (Figure 2.1A)[35,44], while segments neighboring the LC8 site remain largely disordered. Interestingly, whereas previously described

LC8 partners are primarily monomeric until bound by LC8, RavP is a dimer in the absence of LC8 with residues 90-133 corresponding to the dimerization domain[124]. The atypical oligomerization state of RavP suggests that LC8 is required for some other function than to facilitate dimerization.

The LC8 recognition motif is located within a disordered linker connecting the RavP dimerization domain (DD) with the C-terminal domain (CTD), and contains the canonical "TQT" binding motif (Figure 2.1B). Viral transcription and replication require that RavP imposes the correct spatiotemporal arrangement on RavN and RavL at the N-terminus, and $RavN^{RNA}$ at the C-terminus[125]. Therefore, elucidating the structure of RavP and how it is modulated by LC8 is key to understanding the viral replication complex.

In vivo, inhibiting LC8 binding to RavP does not stop viral infection and replication, but rather limits it to levels that do not kill the host organism[92]. Here we show, via a minigenome luciferase assay, that a loss of LC8 interaction leads to a 70% decrease in viral polymerase activity. While it is clear that LC8 interactions with RavP are essential for rabies virus virulence, the molecular basis for why this interaction leads to increased lethality is unknown. To help address this question, we report here the high-resolution structural ensemble of the 90 kDa RavP-LC8 complex using nuclear magnetic resonance (NMR), small angle X-ray scattering (SAXS), and molecular dynamics simulations. We show that RavP is an IDP duplex, with two chains held together by the DD, and separated from the CTD by a long disordered linker that allows for the sampling of multiple conformations. LC8 binds tightly to a disordered linker adjacent to the DD, effectively lengthening the DD and shifting the population of the ensemble towards structures with more restricted CTDs, and affecting the mobility of both ordered and disordered domains distant from the binding site. This effect of DD elongation on domain mobility and CTD orientation could be a shared mechanism among viral phosphoproteins that bind LC8, such as the Ebola virus VP35. We provide testable structural models that explain the consequences of viral incorporation of LC8.

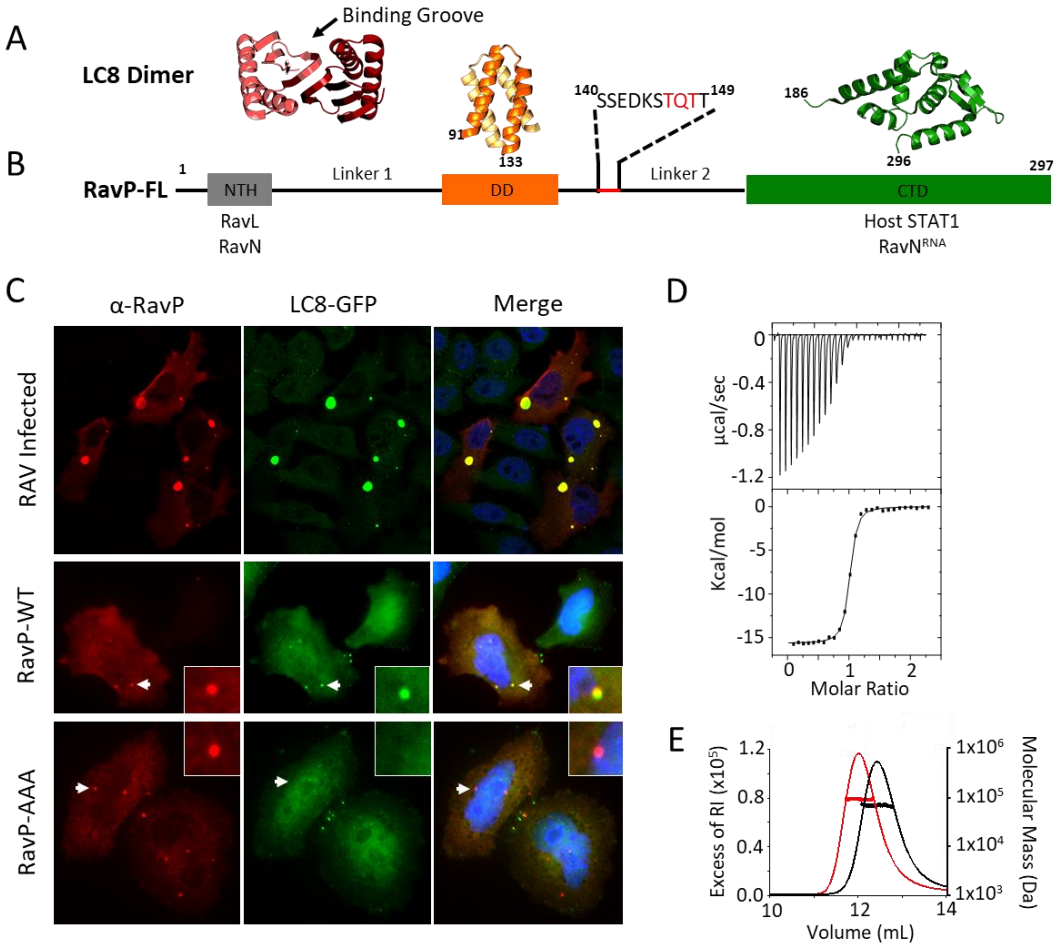


Figure 2.1: LC8 and RavP colocalize in cells and form a tight complex in vitro.
(A) Representative crystal structure of an LC8 dimer (protomers shown in shades of red). Partners bind along the noted groove and form a β-strand in each groove. (B) Domain organization for RavP showing the dimerization domain (DD, PDB 3L32) and the C-terminal domain (CTD, PDB 1VYI) in orange and green, respectively, as well as the predicted N-terminal helix (NTH, grey). Interacting partners are listed below the domain architecture diagram. The recognition motif for LC8 is underscored in red, and expanded above. Black lines denote predicted disorder (Linkers 1 and 2). (C) Confocal analysis of representative HeLa-LC8-GFP cells with RavP staining. (Top) Cells were infected with RAV-CVS11 at an MOI of 10 for 48 hours. Co-transfection of pTIT-RavN and pTIT-RavL with pTIT-RavP-WT (Middle) or pTIT-RavP-AAA (Bottom) proceeded for 48 hours. P was stained with a rabbit polyclonal antibody followed by incubation with Alexa-568 goat anti-rabbit IgG. White arrows denote the smaller Negri body-like structures seen during transfections, which are magnified in the inset. DAPI was used to stain the nuclei (Blue). (D) ITC thermogram for LC8 interactions with RavP, collected at 25 °C in 50 mM NaCl, 50 mM sodium phosphate, 1 mM NaN_3, pH 7.5, and fit to a single-site binding model. (E) Molecular mass of RavP and LC8-RavP complex measured by SEC-MALLS. The lines show the elution profile monitored by refractive index (left axis) for RavP (black) and the RavP-LC8 complex (red). The crosses show the molecular mass (right axis) derived from MALLS and refractometry measurements. The molecular mass calculated for dimeric RavP is 70.7 kDa (black line and crosses) and the heterotetrameric LC8-RavP is 94.8 kDa (red line and crosses).

Results

LC8 and RavP colocalize in cells and form a tight complex in vitro

The LC8 binding site on RavP was identified by deletion experiments [92] and homology modeling [126] to correspond to amino acid positions 140-149, a region that contains a TQT motif; however, no experimental characterizations of this interaction in cells or *in vitro* was previously reported. Here, we demonstrate the co-localization of LC8 and RavP in both RAV infected and RavP transfected cells. In infected cells, RAV forms large ribonucleoparticle inclusions known as Negri bodies (NBs), which contain RavL, RavN, and RavP, as well as a variety of host proteins [127]. NBs are sites for viral transcription and replication, and similar inclusions can be formed from minimal transfections with RavN and RavP [127].

Using confocal microscopy, we show that LC8 is localized inside NBs in infected cells (Figure 2.1C, Top). Additionally, transfection using only RavP, RavN, and RavL led to *puncta* typical of smaller NB-like structures that exhibited colocalization between RavP and LC8 (Figure 2.1C, Middle). To verify that the colocalization represented a direct interaction, we made a triple mutant of RavP where the TQT motif was replaced with three alanines (RavP-AAA). Although transfections with RavP-AAA still led to NB-like inclusion formation, LC8 was no longer localized to these regions, confirming that the TQT motif is the only site of interaction with LC8 (Figure 2.1C, Bottom). Additionally, this suggests that although RavP localizes LC8 to NBs, LC8 is not essential for NB formation.

In addition to the cellular assays, ITC and multi-angle laser light scattering (MALLS) show that LC8 binds RavP with an apparent 1:1 ratio (Figure 2.1D), and forms a complex with a MW of approximately 95 kDa (Figure 2.1E), indicating that the dimeric LC8 (~24 kDa) binds tightly to a RavP dimer (~66 kDa). Interestingly, the interaction between RavP and LC8 is among the tightest yet described for a naturally occurring LC8 partner, with a K_d of 82 ± 8 nM (Table 2.1). Unlike most LC8 binding partners, RavP forms a stable dimer in solution, and therefore the entropic cost of bridging two chains is already paid. The tight binding of LC8 to RavP is likely due to the lower entropic cost of binding compared to other, primarily monomeric partners.

Construct	K_d (nM)	ΔG (kcal/mol)	ΔH (kcal/mol)	$-T\Delta S$ (kcal/mol)
RavP-FL	81.6 ± 8.1[a]	-9.7 ± 0.1	-15.2 ± 0.4	5.5 ± 0.4
RavP-N	29.5 ± 5.3	-10.2 ± 0.1	-15.6 ± 1.3	5.3 ± 1.3
RavP-C	4190 ± 289	-7.3 ± 0.1	-18.8 ± 0.5	11.5 ± 0.5

Table 2.1: Thermodynamic parameters for LC8 interactions with RavP constructs.
[a] Errors are calculated as the s.d. from 2 separate sample preparations or the errors in the fit, whichever was larger. Fits to a single-site binding model gave a stoichiometry of 1 (one LC8 protomer to one RavP chain).

RavP construct design and domain structure

The high percentage of disorder in RavP makes it not amenable to study via crystallography, leaving NMR as the only viable method for residue specific structural information. NMR spectra of the 90 kDa LC8/RavP complex, however, show considerably broadened peaks that are difficult to assign using standard NMR techniques. Additionally, the large number of residues (297 amino acids per monomer) leads to significant peak overlap. Therefore, to simplify the resonance assignment process, we created constructs containing either the N-terminal half of the protein (RavP-N, residues 1-152), or the C-terminal half (RavP-C, residues 140-297; Figure 1B, 2A). Both constructs include the LC8 binding site (residues 140-149). RavP-N contains the DD and the RavL and RavN interaction sites[121], and RavP-C contains the STAT1 and RavN^{RNA} interaction sites[128,129] (Figure 2.1B,2.2A).

Sequence analysis indicates that amino acids 135-155 are part of the disordered Linker 2, and previous work has demonstrated that the DD and CTD are both independently folding domains [130], justifying the design of smaller constructs that begin or end in the Linker 2 region. The CD spectra of both constructs corroborate the predicted structure via the characteristic negative ellipticity around 205 and 222 nm, indicative of both disordered and helical regions. The summation of the CD spectra for the individual RavP-N and RavP-C constructs yields a spectrum that matches one recorded for full-length RavP, indicating that these constructs collectively contain the same amount of secondary structure as full-length RavP (Figure 2.2C). RavP-N and RavP-C have similar number of residues (155 versus 158), but size exclusion chromatograms show that RavP-N migrates closer to RavP-FL than to RavP-C. This demonstrates that RavP-N, which contains the DD and the long intrinsically disordered N-terminal region, forms a dimer with highly flexible linkers in solution, whereas RavP-C is a monomer with a shorter flexible tail (Figure 2.2E).

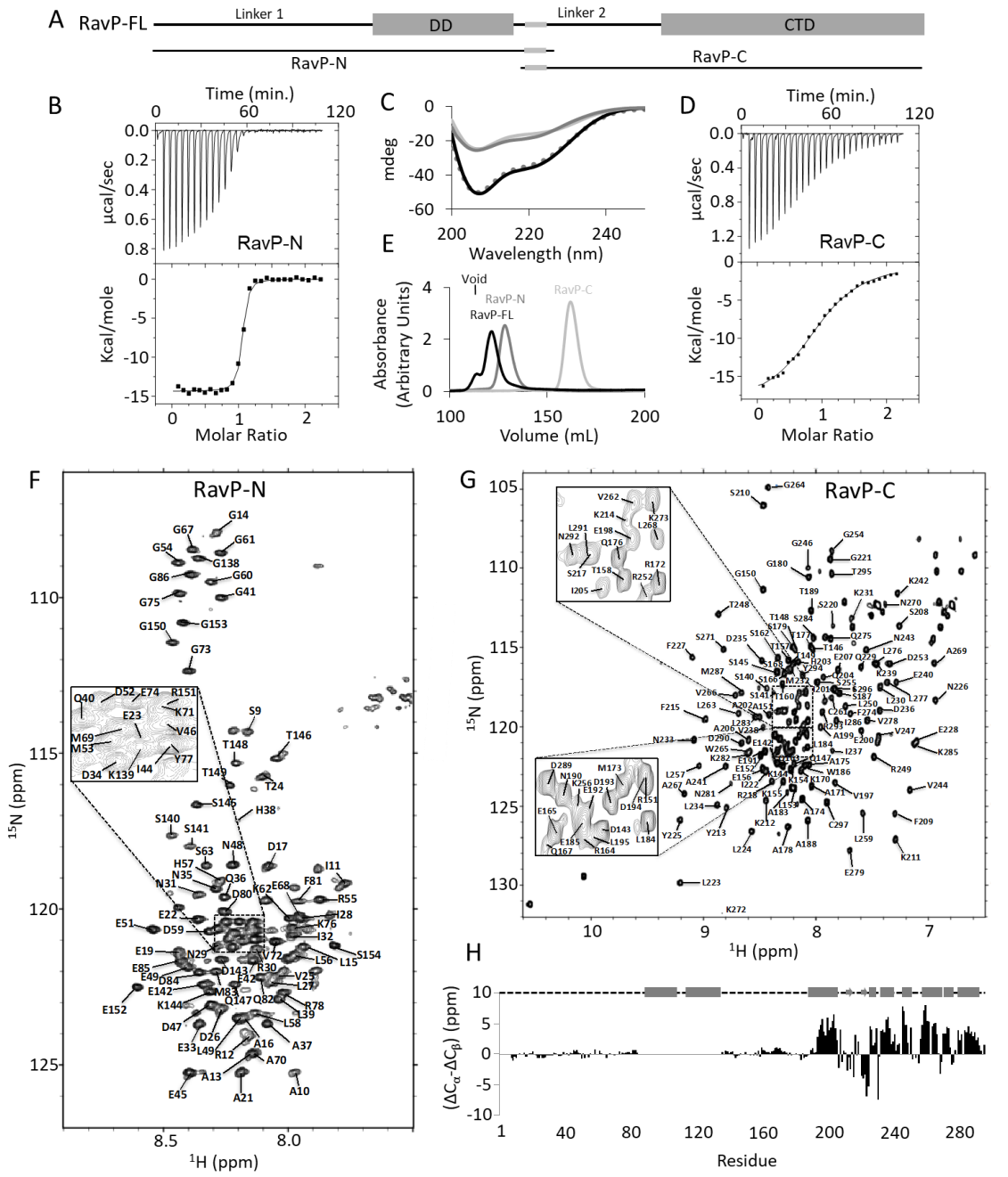


Figure 2.2: Structural analysis and assignments of RavP constructs.
(A) Domain organization for RavP-N (1-152) and RavP-C(140-297) constructs. (B) ITC binding thermogram for RavP-N binding to LC8, collected at 25 °C in 50 mM NaCl, 50 mM sodium phosphate (pH7.5), 1 mM NaN_3, and fit to a single-site binding model. (C) Far UV CD spectrum of RavP-N (grey), RavP-C (light grey), RavP-FL (black), and an additive spectrum for RavP-N + RavP-C (dashed grey), showing the similarity between the constructs and the full length protein. (D) ITC binding thermogram for RavP-C binding to LC8. (E) Size exclusion chromatogram for RavP-N (grey), RavP-C (light grey), and RavP-FL (black) demonstrating that the full-length and N-terminal constructs both migrate as dimers, while RavP-C is a monomer. ^{15}N,^{1}H-HSQC spectra

for RavP-N (F) and RavP-C (G) showing assignments. The spectra were acquired at 25 ^{o}C on a 700 MHz Bruker NMR spectrometer. (H) Secondary chemical shift values determined from Cα and Cβ shifts, showing two long disordered linkers and a well-ordered CTD, consisting of a mixture of α-helices and β-sheets. The structural organization shown above was obtained from crystal structures of the CTD (PDB 1VYI) and dimerization domain (PDB 3L32), where cylinders represent α-helices; arrows represent β-strands, and dashed lines represent disordered regions.

As with RavP-FL, we used ITC to determine binding affinities for LC8 with RavP-N (Figure 2.2B), and RavP-C (Figure 2.2D). RavP-FL and RavP-N, which are dimers, have a much higher affinity for LC8 than the monomeric RavP-C (81.6 nM, 29.5 nM, and 4190 nM respectively; Table 2.1). The higher affinity for the dimeric constructs demonstrates the contribution of bivalency to affinity, as described previously for LC8 binding partners[84], and shown here by the significant contribution of entropy to the difference in affinities ($T\Delta S$ of 5.5 and 5.3 kcal/mol for RavP-FL and RavP-N respectively, compared to 11.5 kcal/mol for RavP-C).

Resonance assignments of the 66 kDa RavP dimer

Backbone assignments of all observable peaks of RavP-N reveal that peaks corresponding to the helical dimerization domain and the eight amino acids at the N-terminus are not visible (Figure 2.2F), and that the observed peaks correspond largely to disordered regions. Extreme peak broadening leading to complete loss of intensity has also been observed for the dimerization domain of a closely related vesicular stomatitis virus phosphoprotein[125]. RavP-C spectra, on the other hand, show wide proton chemical shift dispersion, with a large number of peaks in the 8-8.5 ppm region (Figure 2.2G), consistent with a protein containing both structured and disordered regions. By overlaying the spectra for RavP-FL with those of the constructs, we assigned 191 of 282 non-Proline residues (Figure 2.3A,B). The missing assignments are for residues that have considerably broadened peaks: those within the dimerization domain (residues 87-137), near the C-terminus (residues 287-293) and at the N-terminus (residues 1-8). Assignments were confirmed by overlaying HNCO spectra of the constructs with that of the full-length protein. Secondary chemical shifts for RavP (Figure 2.2H; SI Figure 2.1) identify the folded CTD and two long disordered linkers (Linker 1, residues 9-86; Linker 2, residues 138-190). Interestingly, there is no evidence for the N-terminal helix, predicted for positions 8-31 by PSIPRED [131]. This is in contrast to what is seen in the closely related vesicular stomatitis virus phosphoprotein (VSVP), in which residues 2-16/25-31 show a helical propensity[132]. The difference between RavP and VSVP is accurately recognized by the program Agadir[133], which predicts significantly lower helical propensities for the N-terminus of RavP than VSVP. A

comparison of the NMR generated secondary chemical shift values to the known structure of the CTD shows very strong agreement; however, it is of note that peaks from 14 residues that either shift or disappear in RavP-FL compared to RavP-C, are primarily localized to a helix between residues 247 and 253, suggesting some structural rearrangement of these residues in RavP-FL compared to the monomeric CTD (Figure 2.3A-C). RavP-FL peaks that do not overlay with the RavP-N peaks were not observed in 3D spectra, and were therefore not assigned. These peaks presumably represent residues in the DD or the first eight amino acids.

Dynamics measurements and structural heterogeneity in free and LC8 bound RavP

Dynamics measurements for RavP-FL were determined using T_1, T_2, steady-state het-NOE, peak intensity, and amide proton H/H exchange CLEANEX experiments (Figure 2.3E, black). Due to peak overlap, heteronuclear NOE and CLEANEX experiments were collected using an HNCO-based pulse sequence (SI Figure 2.2) in addition to the more sensitive HSQC-based experiment necessary for observing lower intensity signals arising from the CTD. T_1 values do not show any variation within the linkers or CTD (average of 0.7 s for the linkers, and approximately 1.0 s for the CTD). The longer relaxation time for the structured region is expected for a large protein at 800 MHz field strength[134]. T_2 on the other hand, shows heterogeneity in the linkers, suggesting multiple conformations (Figure 2.3E; SI Figure 2.3). As expected, T_2 values decrease for residues closer to the structured regions, and become too low for reliable fits for the ordered CTD. Het-NOE experiments corroborate the disorder in the linker regions, with values near 0.2-0.3. Despite relatively large error bars, HSQC-based het-NOEs for the ordered CTD have values distinctly higher than those of the linkers, closer to 1, as expected for folded proteins (Figure 2.3E). RavP-FL normalized intensities (Figure 2.3B) for CTD peaks show greatly diminished intensities compared to the linkers. T_1, T_2, and Het-NOE data are further analyzed using spectral density mapping[135]. In this case, variabilities in the spectral density function values, in particular J_0, support our conclusion that the linkers in RavP are likely in exchange between multiple conformations on an intermediate timescale (Figure 2.3E; SI Figure 2.3). For the CLEANEX experiment, which measures fast amide proton exchange with water, peaks are observed only for amide protons that are readily exchangeable[136]; therefore, higher peak intensities identify residues that are most solvent exposed. Linker 1 residues are all visible in the spectrum, but residues 55-70 show lower intensity peaks. In the CTD, peaks for residue 221 and the C-terminal residue 297 are both visible. Residue 221 is part of the longest turn in the structure, and residue 297 is disordered. Linker 2 peaks vary in intensity; residues 160-180 are

either absent or much less intense, indicating some order. In summary, for free RavP, variations in CLEANEX intensities, peak intensities, T_2 relaxation times, and J_0 values (Figure 2.3) indicate that the linkers are largely disordered, but have significant structural heterogeneity on the microsecond to millisecond timescale. As a final measure of structural propensities within the linkers, we collected ^{1}H-^{15}N RDCs, which can provide long-ranged information on relative orientations of backbone bond vectors within the protein. While the CTD peaks are not visible due to the substantially decreased peak intensities, $^{1}D_{NH}$ values for the linkers cluster around 0, confirming that the disordered linkers are sampling many conformations with no significant secondary structure (SI Figure 2.3).

A comparison of HSQC peak intensities between free and LC8-bound RavP-FL reveals peak attenuation for all residues within the LC8 binding motif (residues 140-149, Figure 2.3), similar to other LC8 binding partners[64,69]. Indeed, the only peaks that disappear in the HSQC are those within the LC8 binding region, and there are no peaks that shift (Figure 2.3D,E). Carbonyl chemical shifts reiterate these findings, and additionally show some peak disappearance between residues 160-170. Further, RDC experiments indicate that $^{1}D_{NH}$ values for the linkers are unchanged by LC8 interactions (SI Figure 2.3). To test the effect of LC8 on RavP-FL dynamics, we collected the same experiments as above on the LC8-RavP complex (Figure 2.3). T_1, T_2, and het-NOE values are largely unchanged; however, spectral density analyses suggest that the N-terminus of Linker 1 is more restricted and undergoing slower exchange in the complex, evidenced by the increased J_0 in this region (Figure 2.3E). As expected, residues immediately adjacent to the LC8 binding site have higher J_0 values, evincing more restriction in the complex. Additionally, intensity ratios of CLEANEX values for residues 55-70, 140-180, and 221 in the CTD are significantly less than 1. While we would expect this change at the LC8 binding site, the long-range effects on the N-terminus of Linker 1 and the CTD suggest that they also decrease in flexibility, but do not change in secondary structure, as shown by similar carbonyl chemical shifts (SI Figure 2.2). Taken together, these data demonstrate that LC8 binding does not change the secondary structure of RavP (aside from the binding site), but instead may impact domain orientations and flexibility within the conformational ensemble.

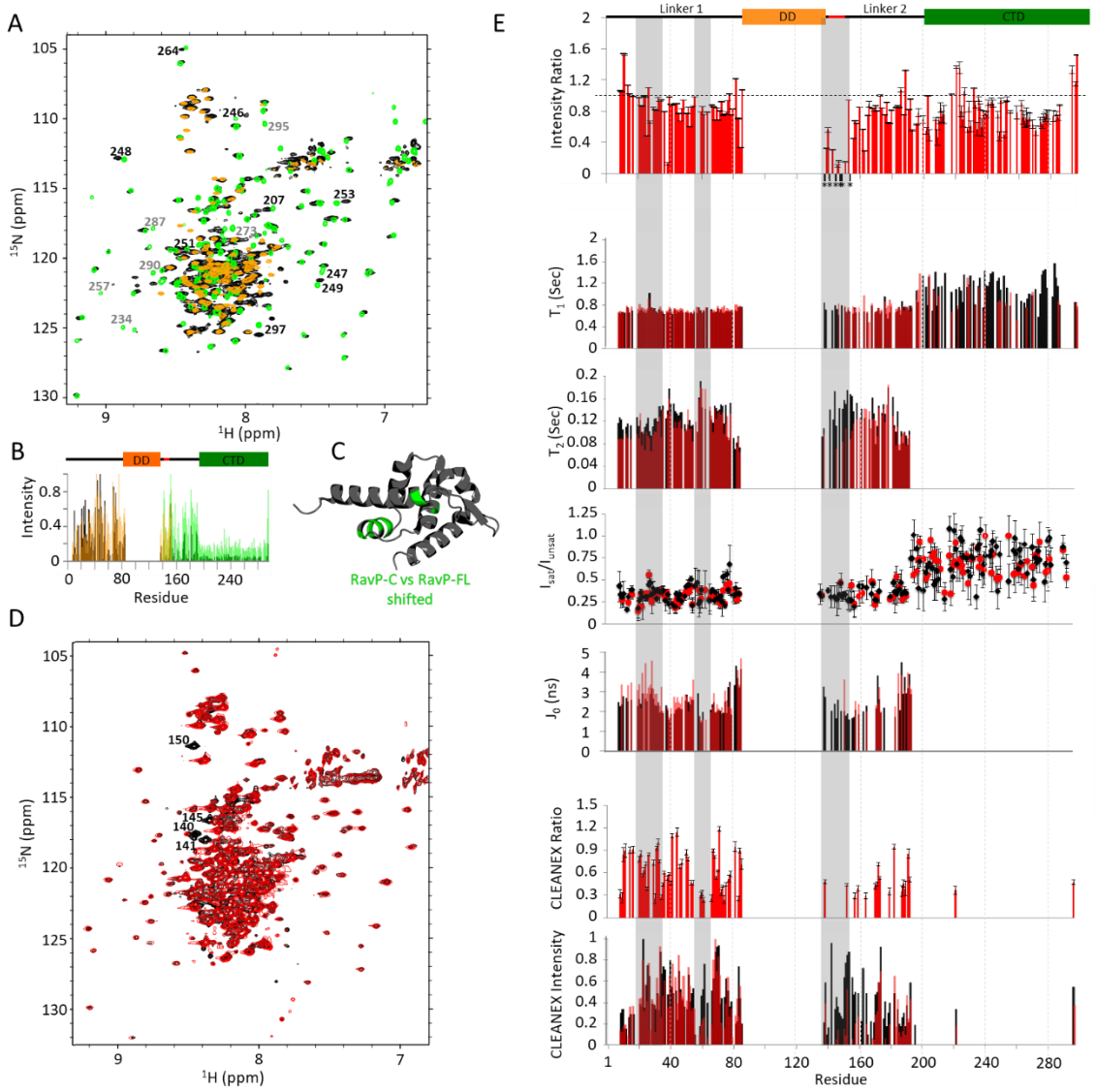


Figure 2.3: RavP NMR relaxation experiments.
(A) ^{1}H-^{15}N TROSY-HSQC spectral overlay of RavP-FL (black) with RavP-C (green) and RavP-N (orange) demonstrating the strong agreement in chemical shifts between the constructs and the full-length protein. Peaks that shift between RavP-FL and RavP-C spectra are labelled in black, and peaks that disappear or decrease in intensity are labeled in grey. Normalized intensities for all three proteins are shown with the domain architecture in (B). (C) Crystal structure for the RavP CTD (PDB 1VYI) highlighting residues with peaks that shift in HSQCs for RavP-C and RavP-FL (green). (D) BEST ^{1}H-^{15}N TROSY-HSQC for free (black) and LC8 bound (red) RavP-FL collected at 800 MHz. Example peaks showing large attenuations in peak intensity are labelled. (E, Top) Ratios of Bound:Free peak intensities, generated using the spectra shown in (D). Error bars were calculated using the intensities of the baseline noise. Peaks that disappear in the bound sample in (D) are marked by an *, and are localized to the LC8 binding site. Relaxation times (T_1, and T_2) for free (black) and bound (red) are shown overlaid on the same axis, as are J_0 values from the spectral density analysis. Residues with fit errors greater than 25% of their value are not shown. HSQC-based ^{1}H-^{15}N NOEs (I_{sat}/I_{unsat}) demonstrate the difference

between linker and CTD residues. The CLEANEX intensities (E, Bottom) for free and bound RavP show regional differences, and were compared as a CLEANEX ratio (bound:free) in the graph above. Note that LC8 interactions lead to decreased CLEANEX intensities from residues 55-70, and across the majority of Linker 2. The segments most changed by LC8-binding are highlighted in grey.

Ensemble structure from SAXS and molecular dynamics simulations

To assess the effect of LC8 on RavP domain orientations, we performed SAXS experiments on RavP and the RavP-LC8 complex (Figure 2.4A). The SAXS-determined molecular weights of 69 kDa and 91 kDa for free and bound RavP are consistent with the theoretical MW for a dimer (68 kDa with 6xHis tag) and a heterotetramer (92 kDa) respectively (Table 2.2). The radius of gyration (R_g) calculated using the Guinier approximation for RavP-free was 4.9 ± 0.1 nm, and 5.0 ± 0.1 nm for the RavP-LC8 complex (Figure 2.4B). Interestingly, although the R_g values are approximately the same for bound and free, the D_{max} is significantly smaller for the complex (210 Å for RavP and 173.5 Å for the complex). This indicates that although the structures occupy a similar volume on average, RavP-free can inhabit more extended conformations (Figure 2.4C). The Kratky plots for both free and bound RavP are typical of a multi-domain protein with flexible linkers (Figure 2.4D), presenting a maximum at low q values, followed by a gradual decrease as q increases.

A mixture of structure-based models (SBM) and molecular dynamics simulations (MDS) was used to create an ensemble to fit the SAXS profile. In classical MDS, the LC8-RavP interaction was found to be stable over the course of the simulation. With the GAJOE program, the number of conformers was decreased until an ensemble containing the minimum number of structures to accurately fit the SAXS data was obtained. This led to optimized weighted ensembles of five conformers each for RavP-free and the RavP-LC8 complex (Figure 2.4E-H). The ensemble predicted SAXS scattering profile fit well to the data, with χ_{EOM} values of 0.64 for the RavP-free ensemble, and 0.49 for the complex. The R_g range for RavP-free and bound is very similar in the selected ensemble, and both can access an extended (R_g range 5-6.8 nm) and a compact (R_g range 3.5-4.2 nm) conformation. There are very few structures in the selected ensemble with an intermediate R_g, indicating that there is a thermodynamic bias for the extended and compact conformations. Interestingly, the gap between these two broadly defined conformations is much more dramatic for the complex, signifying that some intermediate conformations are prohibited in the presence of LC8. It is of note that the selected ensemble for both samples has a strong enrichment for the compact structure when compared to the initial pool, demonstrating the importance of the compact structure within the ensemble.

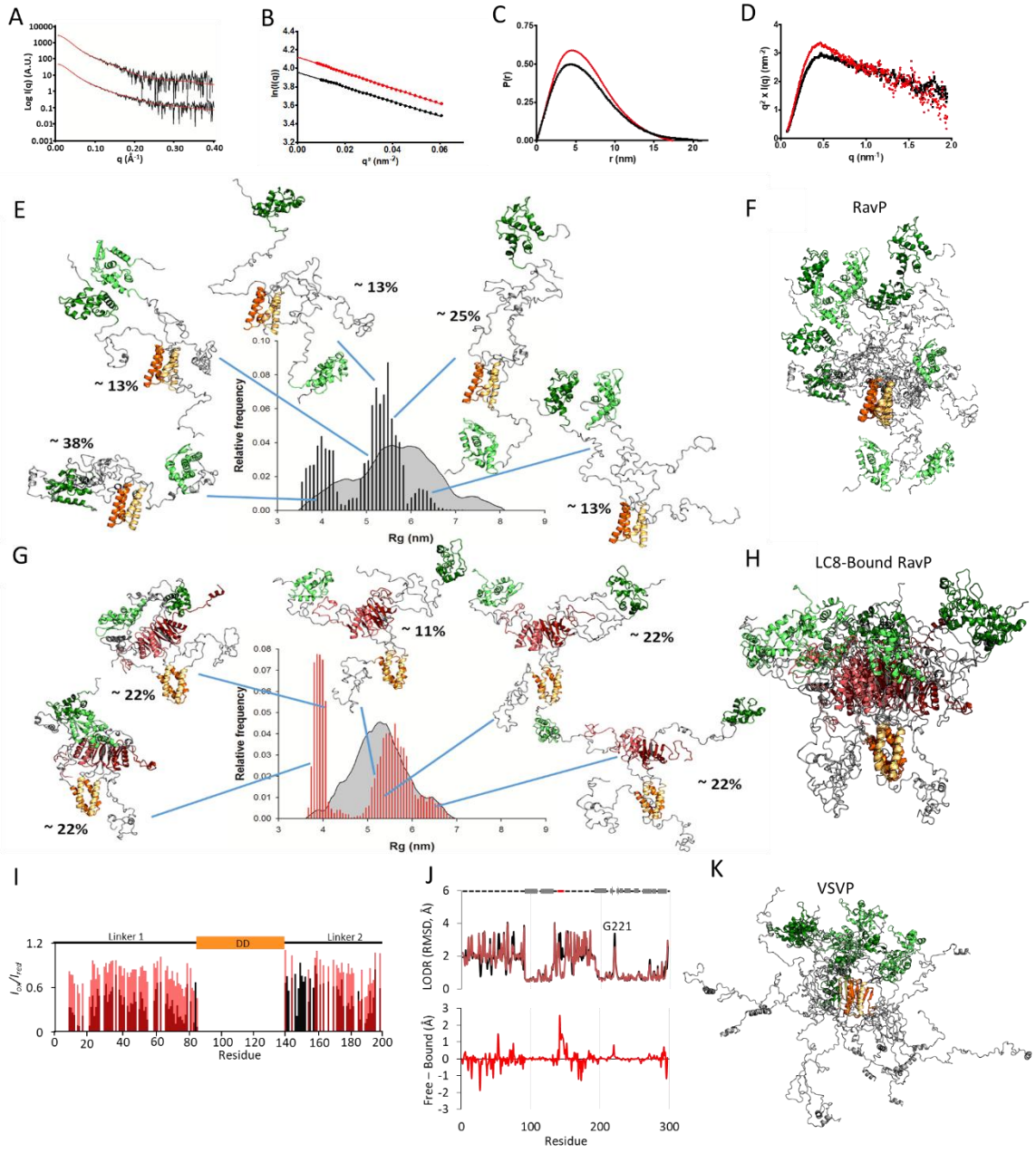


Figure 2.4: SAXS data and selected ensembles analysis.
(A) Fitted SAXS profiles for RavP (bottom) and LC8-RavP (top). The experimental curve is shown as a black line and the EOM selected ensemble as a red line. (B) Guinier plot for RavP (in black) and LC8-RavP complex (in red). (C) Pair distribution function for RavP (in black) and LC8-RavP complex (in red). (D) Kratky plot for RavP (in black) and LC8-RavP complex (in red). (E) Radius of gyration distributions for the initial pool (grey) and EOM-selected RavP ensembles (black histogram). Representative RavP models from the EOM are shown as cartoon with their central dimerization domains colored in orange and their C-terminal domains in green. The intrinsically disordered regions are shown in grey. The relative contribution of each model to the EOM calculated SAXS profile is indicated as a percentage. The RavP selected ensemble is

overlaid and aligned on the DD in (F). (G) Radius of gyration distributions for the initial pool (grey) and EOM-selected LC8-RavP ensembles (red histogram). Representative models are shown with LC8 in red, and overlaid in (H). (I) PRE intensity ratios of the paramagnetic (oxidized) to diamagnetic (reduced) of MTSL labelled Rav-C261S free (black) or RavP-C261-LC8 complex (red). Values are only shown for residues 1-196, as the CTD peaks were not visible in this experiment. (J, Top) Ensemblator analysis of models generated via SBM and MDS for RavP (black) and LC8-RavP (red). Models for each sample were compared in one group, with automatic clustering turned off. Residues with a higher LODR in the complex represent regions that are less restricted in motion when LC8 is bound to RavP. Lower LODR regions, on the other hand, are those that are more rigidified within the complex ensemble. A comparison of the differences is shown below. Note the large decrease in variability for residues in the LC8 binding site (residues 140-149). Differences in the CTD, particularly in the loop containing residue 221, are distal from the LC8 site and therefore unexpected. (K) VSVP ensemble overlay of 5 structures, adapted from Leyrat *et al.*, 2012[125], with conformers aligned using the DD.

	$RavP_{FL}$[a]	$RavP_{FL}$ – LC8
Data collection parameters		
Instrument	ESRF – BM 29	
Beam size at sample (µM)	700 x 700	
Wavelength (Å)	0.9919	
q range ($Å^{-1}$)	0.20 - 33	
Detector	Pilatus 1 M	
Detector distance (m)	2.867	
Time per frame (s)	2	
Column	S200 increase 10/300 GL	
Flow rate ($mL.min^{-1}$)	0.5	
Injected sample concentration ($mg.mL^{-1}$)	8	9
Injected volume (µL)	500	
Temperature (K)	293	
Structural parameters		
Rg (Å) [from Guinier]	48.7	50.4
Rg (Å) [from P(r)]	51.6	49.7
Dmax (Å)	210	173.5
Molecular mass M_R (kDa) from Volume of correlation	69	91
Calculated M_R (kDa) from sequence	68.4	93.5
EOM χ_{exp}	1.45 (SBM), 1.01 (MDS), 0.64 (SBM+MDS)	0.67 (SBM), 0.50 (MDS), 0.49 (SBM+MDS)

Table 2.2: SAXS data collection and scattering-derived parameters.
[a] Data processed with Primus/ScÅtter.

Weights for selected structures are shown by a percentage next to a representative structure. A comparison of these weighted structures reveals that the CTDs of unbound RavP are free to rotate around the DD, while the CTDs within the complex are kept in closer proximity to one another (Figure 2.4E-H). Additionally, CTDs in the complex have a relatively fixed orientation with respect to the DD, and the distance between the CTDs is the main distinction between the compact (R_g 3.8-4.1 nm) and extended (5.1-6.8 nm) conformations in the complex. The R_g gap between 4.1 nm and 5.1 nm may therefore represent the relative benefits of a slight attraction between the CTDs *versus* an increased freedom of motion. Another interesting point is that Linker 1 is often excluded from the CTD region in the bound complex, most likely because the presence of LC8 creates a situation where there is no space for this linker.

In order to test this model, where LC8 interactions lead to increased proximity of the CTDs to each other, and decreased proximity of the CTDs to Linker 1, we performed Paramagnetic Relaxation Enhancement (PRE) NMR experiments by attaching an MTSL spin label to C297. RavP contains two endogenous Cys, C261 on the surface of the CTD, and the C-terminal residue C297 adjacent to the CTD. C261 was mutated to a Serine (C261S) to simplify the PRE analysis. Effects of the mutagenesis were monitored by NMR, which indicated no change in structure based on chemical shifts. In paramagnetic samples, CTD peaks were not visible due to the innate broadness of the peaks, as well as the MTSL attached to this domain; however, we were able to monitor the impact of labelling on the linker regions (Figure 2.4I, SI Figure 2.3). Supporting the MS generated models, linker intensities are significantly lower in the free RavP sample, particularly at the N-terminus of Linker 1, indicating that LC8-binding results in the exclusion of Linker 1 from the binding region. It is not possible to verify from our PRE data whether the CTDs are held in closer proximity in the complex, as those residues were not visible in either spectrum.

To overlay and compare structures within the ensembles, we used the Ensemblator program [137] which has a built-in Locally Overlaid Dipole Residual (LODR) tool to evaluate residue-level structural heterogeneity by comparing individually overlaid dipeptides (Figure 2.4J). In general, peptides with perfect alignment will have an RMSD of 0, and peptides that show an approximately 180° rotation in the ensemble will have an RMSD of ~5 Å. For RavP, the Ensemblator recognizes the linkers as highly variable, and the domains as mostly static structures (Figure 2.4J).

Although these models are independent from the NMR data, with the exception that the linkers were modeled as disordered segments due to the near-zero $\Delta C\alpha$-$\Delta C\beta$ and $^1D_{NH}$ values

(Figure 2.2H, SI Figure 2.3), they accurately recapitulate multiple findings from NMR. First, the difference curve between RavP-free and the complex (Figure 2.4J) reveals a large stabilizing effect on the LC8 binding site (Residues 140-149), consistent with the formation of a β-strand. These peaks are missing in NMR spectra (Figure 2.3D) due to increase in ordered structure. Additionally, the sole residue visible in the CTD in the CLEANEX spectrum (Figure 2.3E, G221) is pinpointed in the models as the most variable portion of the CTD (Figure 2.4J). These data lend credence to the models as a valuable source of information about the ensembles at the residue level. With that in mind, examination of CTD residues reveals that changes are localized to the ends of helices and the turn containing residue G221 (Figure 2.4J).

Conservation of modular structure and LC8 binding

Phosphoprotein (P) expression and function is conserved across viruses in the *Rhabdoviridae* family[130]. Previous work comparing Ps in the *Mononegavirales* order, which includes the *Rhabdoviridae* family, has demonstrated that, although there is little sequence conservation across families, Ps are consistently organized in a modular manner, with disordered linkers connecting independently folding domains[130]. In order to test if LC8 binding C-terminal of the DD is also conserved, we scanned phosphoproteins from different genera in the *Rhabdoviridae* family, as well as different *Lyssaviruses* (the rabies genus), for potential binding motifs. The LC8 motif in *Lyssavirus* phosphoproteins is stringently conserved (Figure 2.5A,B); however, no phosphoproteins from other *Rhabdoviridae* genera contain the motif (data not shown).

We used a representative protomer from our RavP SAXS ensemble to map the various conserved regions of the RavP sequence within the *Lyssavirus* genus (Figure 2.5A). As expected the structured DD and CTD are both relatively conserved (blue), while the linkers are generally more variable (red and blue). Interestingly, within Linker 2 only the LC8 binding sequence is conserved, indicating that there is evolutionary pressure selecting for this binding sequence. Additionally, the N-terminal region of Linker 1 is highly conserved. This region contains binding sites for L and N, which are both essential for viral proliferation[121].

It is surprising that *Lyssaviruses* are the only genus within the *Rhabdoviridae* family to appropriate LC8. One possible explanation for this is that *Lyssavirus* phosphoproteins have a domain architecture distinct from other genera, which requires stabilization via LC8. Using IUPred, we determined propensities for order and disorder in phosphoproteins across the different genera in the *Rhabdoviridae* family (Figure 2.5C). The majority of genera displayed similar

domain architecture, consisting of two disordered linkers (humps in prediction values) between ordered domains. This suggests that LC8 interactions are not required for stabilization of the RavP dimeric structure, but have instead evolved as an additional switch to regulate downstream functions for RavP. Future work on the pathway(s) affected by LC8 binding can illuminate not only the importance of LC8 for the whole *Lyssavirus* genus, but also a key difference between viruses within the *Rhabdoviridae* family.

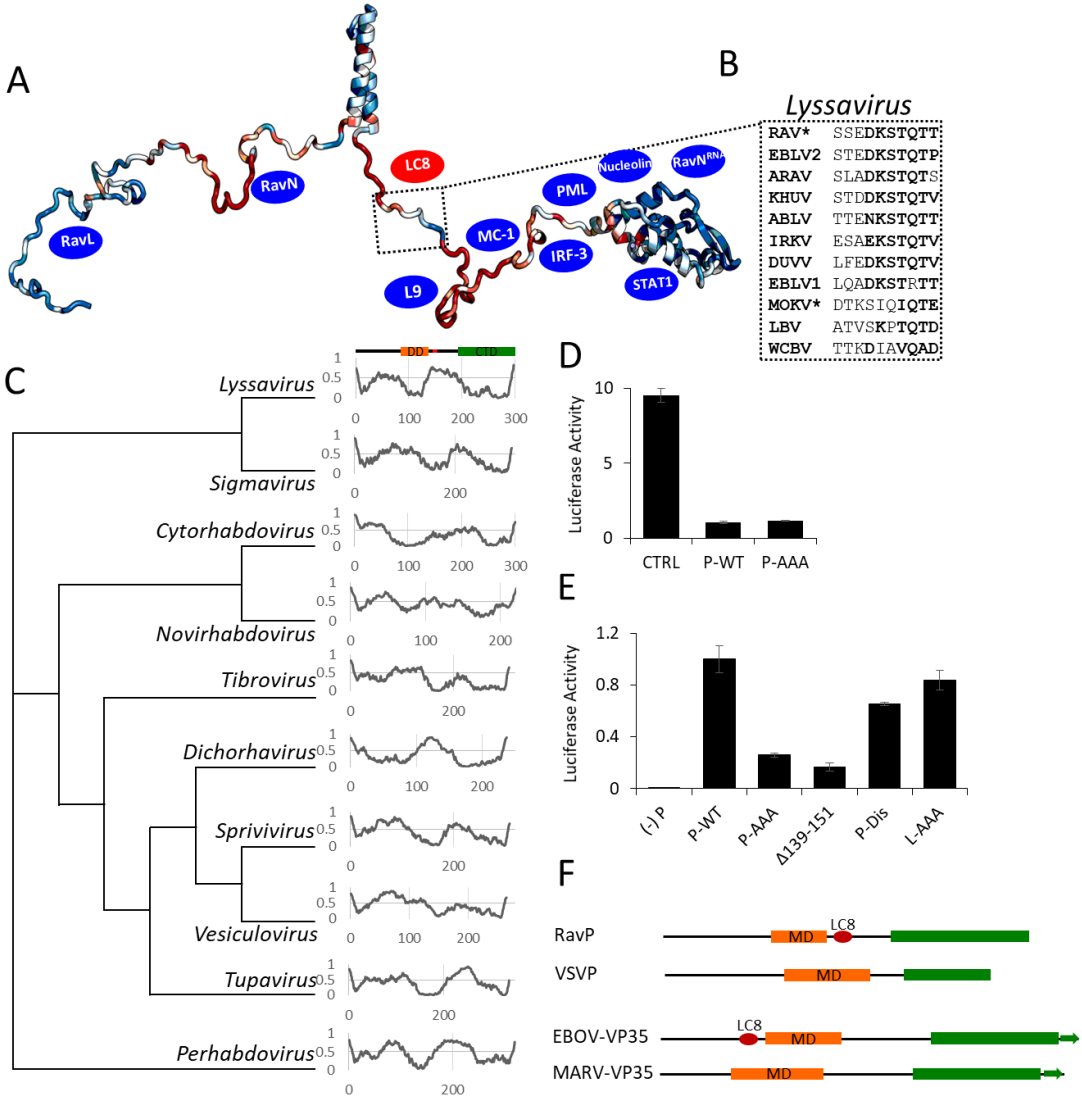


Figure 2.5: Phylogenetic and functional analysis of the Rhabdoviridae family and Lyssavirus genus Phosphoprotein.
(A) ConSurf analysis of P proteins within the *Lyssavirus* genus, with conserved residues shown in blue, and variable residues shown in red. The LC8 binding sequence is highlighted, with example sequences from viruses within the *Lyssavirus* genus shown in (B). Residues in bold are amino acids frequently present in LC8 binding partners. *Denotes sequences verified to bind LC8 [87].

Blue ovals denote approximate binding sites of known partners (see Discussion). (C) Phylogenetic tree for phosphoproteins in different genera of the *Rhabdoviridae* family. Genera which show a different domain architecture were excluded from the analysis. Phosphoproteins for representative viruses within each family were used for the comparison (see Methods). Predicted order/disorder was calculated using IUpred2A's long disorder analysis, where higher values indicate a probable disordered region, and lower values indicate a structured segment (Y-axis: 0-1, with a value of >0.5 denoting a predicted disordered region). X-axes are labeled by residue numbers, and represent the entire phosphoproteins from representative viruses. (D) Luciferase reporter assays for IFNα-treated Hek-293T cells (2,000 U/mL, 6 hours). Cells were transfected with luciferase assay components, as well as a plasmid encoding RavP-WT, RavP-AAA, or an empty vector control. (E) Luciferase expression from the RAV minigenome system. BSR-T7 cells were transfected with either RavP constructs shown, as well as RavN, RavL or RavL-AAA, and luciferase reporter assay components. All luciferase values are normalized to the RavP-WT transfection, with error bars shown as the standard deviation from the mean for two independent experiments, run in triplicate. A negative control is shown ((-) P), in which the RavP plasmid was not transfected. (F) Predicted domain architectures for RavP, VSVP, Ebola virus VP35, and Marburg virus VP35. LC8 sites are marked by red ovals.

LC8 is not involved in RAV suppression of the STAT1-mediated immune response

STAT1 is one potential downstream RavP binding partner that may be impacted by LC8 interactions. STAT1 plays an essential role in the innate immune response by regulating the transcriptional response to Interferons (IFNs). During a typical innate immune response, activated STAT1 translocates to the nucleus and binds to IFN-activated sequences within target gene promoters, stimulating transcription of these genes. During a RAV infection, on the other hand, RavP binds to cytosolic STAT1, inhibiting STAT1 translocation and activity[138]. Here we use an IFN-meditated luciferase assay to probe the effect of LC8-binding to RavP on the suppression of the host immune system. The exposure of HEK-293T cells to IFNα in control, RavP-WT, or RavP-AAA transfected cells reveals that LC8 interactions do not affect the ability of RavP to suppress STAT1 activity (Figure 2.5D). While control cells demonstrated a potent response to IFNα, transfections with either RavP-WT or RavP-AAA containing plasmids significantly suppressed this response.

LC8 is required for efficient polymerase function via interactions with RavP, but not via RavL

The presence of LC8 within the ribonucleoparticle inclusions (Figure 2.1C) potentiates a role for LC8 in viral polymerase activity. In order to test this hypothesis, we used a minireplicon system, wherein the viral polymerase components (RavP, RavN, and RavL) are transfected in eukaryotic cells, alongside a luciferase reporter gene framed by the RAV 3' leader and 5' trailer sequences[139]. Therefore the production of luciferase, measured via a luminescence assay, is directly related to viral polymerase activity. A comparison of transfections using WT RavP or

RavP-AAA shows that mutations of the LC8 motif lead to a 70% decrease in luciferase intensity (Figure 2.5E). While these data suggest that LC8 is impacting polymerase activity, it is unclear whether LC8 is directly facilitating viral transcription. One potential alternative is that a different partner binds RavP in competition with LC8, *e.g.* binds to the same region of the linker.

To test for the presence of a competitive binder, in addition to the WT and AAA mutant, we also deleted the LC8 binding region (RavP Δ139-151), or replaced the full motif with a disordered sequence of similar length (RavP-Dis). Figure 2.5E demonstrates that the complete removal of the LC8 motif results in a decrease in luciferase intensity similar to that of the RavP-AAA construct, while replacement with a random coil sequence leads to a partial, but incomplete attenuation in signal (Figure 2.5E). It is interesting that mutation of a smaller region (RavP-AAA compared to RavP-Dis) results in a larger change in luciferase signal. One possible explanation is that there are proteins that compete with LC8 for a similar binding region of RavP. These proteins may be able to bind in the RavP-AAA construct, but not in the RavP-Dis construct, resulting in abrogated RavP-AAA localization or function.

Previous work has suggested that RavL also interacts with LC8, and that this interaction is required for efficient polymerase activity[117]; however, this assertion was based largely on focus forming unit assays, which do not directly measure polymerase activity. To directly test the importance of the LC8 motif in RavP, we mutated the TQT motif in RavL to an AAA (RavL-AAA), and performed a minigenome luciferase assay with RavP-WT, RavN, and RavL-AAA (Figure 2.5E). These data indicate that the LC8 binding sequence in RavL is not important for polymerase activity, as there is no statistically significant difference between these samples (Figure 2.5E). Consistently, mutation of the LC8 site in RavP eliminates LC8 localization to Negri Bodies (Figure 2.1C), suggesting that RavL is insufficient for LC8 recruitment to the sites of viral transcription.

Discussion

RavP is an essential component of the rabies virus, as it interacts with RavL[121], RavN[120], $RavN^{RNA}$ [120,123], and a multitude of host proteins[140]. These interactions implicate RavP in many diverse roles, including viral transcription, controlled replication, immune evasion[122,138], and nucleocytoplasmic transport[141], as well as host apoptosis/autophagy regulation[142] and mitochondrial dysfunction[143]. Prior research demonstrates that interactions between RavP and LC8 are imperative for viral lethality[92]. Cell-based experiments indicate that LC8 may have roles

in primary transcription[92] and intracellular localization[141]; however molecular level descriptions of the role of LC8 in viral lethality have been hampered by the structural complexity of the multi-domain, partially disordered RavP.

Here we show by CD, SAXS, MALLS, and NMR, that RavP is an IDP duplex composed of two structured domains and two long disordered linkers. Linker 1 (residues 1-86) and Linker 2 (residues 138-191, connecting the central DD to the CTD) are highly dynamic on the nanosecond to picosecond timescale (homogeneous T_1 and het-NOEs), but have considerable structural heterogeneity on the millisecond to microsecond time scale, demonstrated by the variable amide proton solvent exchange, transverse relaxation rates, and J_0 values (Figure 2.4E). Taken together, these results suggest that the linkers are interconverting between disordered and partially compact structures. While the structure of the CTD in full length RavP matches crystallographic studies on the individual domain, the SAXS selected ensemble of the full length RavP emphasizes the motional freedom available to the CTDs: either proximal (CTDs are essentially adjacent), or distal (CTDs are on opposite sides of the dimerization domain). These multiple orientations contrast with the homologous VSVP, for which the CTD conformations are primarily proximal[125], while the N-terminal Linker 1 samples a wide range of conformations. Regulation of these CTD orientations by LC8 is important for Rabies viral polymerase function, but not for immune evasion via the STAT1 pathway (Figure 2.5D).

LC8 restricts dynamics of disordered linkers and ordered CTD

Using a wide variety of dynamics experiments that probe for small differences between bound and free RavP, we demonstrate that LC8 binding to RavP does not change RavP secondary structure, but rather alters the dynamics of Linkers 1 and 2, and rigidifies portions of the CTD. The decrease in peak intensity within the LC8 binding site (residues 140-149) is typical for LC8 binding proteins, presumably due to formation of the interfacing β-strand; however, long-range effects on Linker 1, Linker 2 residues 150-180, and CTD residue 221, indicate that LC8 alters dynamics outside of the binding site. Dynamic changes in the millisecond to microsecond timescales are seen as decreased intensities in CLEANEX experiments, and dynamic changes in the microsecond to nanosecond as increased values of J_0, most prominently for residues 10-40 and 150-160.

The linkers in RavP are the sites for interaction with a wide variety of proteins (Figure 2.5A) including the ribosomal protein L9[140], mitochondrial complex I[143], IRF-3[138], STAT1[122], RavN^{RNA123}, PML[144], and nuclear import and export factors[141]. Additionally, a number of proteins

bind RavP at unknown sites (e.g. BECN1 [142] and HSP90/Cdc37[145]), and may also depend on the accessibility of these linkers. Therefore, the changes in linker dynamics upon LC8 binding may lead to changes in binding interactions with downstream binding partners. There is indeed evidence that LC8 binding causes changes in the nuclear/cytoplasmic localization of RavP[141], suggesting changes in accessibility of nuclear import and export sequences within these linkers. We have tested the impact of LC8 binding on two downstream interactions: the innate immune response protein STAT1, and the viral polymerase components. While LC8 does not affect RavP-STAT1 interactions, it does play a role in the polymerase complex by facilitating viral polymerase activity. Further studies will determine how RavP interactions with LC8 affect other downstream partners, particularly those that bind within the linkers.

LC8 elongates the RavP DD, facilitating CTD orientations conducive to transcription

Molecular dynamics simulations based on SAXS data, and supported by PRE experiments, reveal that the CTDs are more constricted upon LC8 binding, resulting in a more compact structural ensemble. A comparison of RavP to the vesicular stomatitis virus phosphoprotein (VSVP), another member of the *Rhabdoviridae* family that shares a common domain architecture[130], suggests an intriguing new role for LC8 in the restriction of the conformational ensemble, and potentially in the regulation of downstream interactions common to both. In the VSVP ensemble the CTD is more restricted with respect to the DD[125] (Figure 2.4K) favoring orientations where both CTDs of the dimer are constrained on one side of the DD, similar to the LC8-bound RavP ensemble (Figure 4G,H), and different from the free RavP ensemble (Figure 2.4E,F). Contributing to this difference is the longer dimerization domain in VSVP, and the shorter linker 2. Interestingly, LC8-binding in RavP would result in a dimerization domain that is lengthened to a similar extent as the VSV-DD, suggesting that the lengthening of the DD restricts CTD domain orientation. Another difference is in the topology of the peptide chains in the DD. In RavP, Linkers 1 and 2 extend from the same side of the DD, resulting in steric hindrance between the N-terminal and C-terminal region. In VSVP the linkers are on opposite faces of the DD. Our PRE data support a model where LC8 binding leads to increased distance between Linker 1 and the CTD of RavP, resulting in linker localizations to opposite faces of the central DD, similar to the native state of VSVP.

The striking similarity in tertiary orientations and CTD restriction between LC8/RavP and VSVP suggests that LC8 binding facilitates a phosphoprotein function common between VSVP and RavP. One common, essential P function is to connect the viral polymerase, L, to the

nucleoprotein-RNA complex, N^{RNA}, during viral transcription and replication. It is proposed that CTD binding N^{RNA} could trigger some conformational change in N^{RNA}, resulting in loosening the grip of N on the RNA to allow the polymerase access to the RNA[123]. The polymerase requires the cooperation of P to synthesize long RNA molecules and/or to scan intergenic regions during viral transcription and replication. Two models have been proposed to explain how P assists the polymerase in its motion along the template RNA: (i) In the first model, P remains attached to L and walks along the N-RNA template with its CTD continuously and alternatively associating with and dissociating from N[146]. (ii) In the second model, multiple P molecules are positioned through their CTD at regular intervals along the N-RNA template, and the polymerase jumps from one to another[147].

Our data support a model for the replication complex in which both CTDs of a phosphoprotein dimer must be oriented on the same side of the molecule for efficient activity of the replication complex (Figure 2.6). In VSV, the architecture of P creates this situation in the free protein, so that VSVP is fully functional in its free form. In RAV, the higher flexibility of Linker 2 and the steric hindrance induced by the presence of the two N-terminal linkers in close proximity of the CTD results in a situation where the CTDs are further apart in the free form of the protein. LC8 binding would then be necessary to bring the CTDs in close proximity and to stimulate the activity of the replication complex. Consistent with this model, our minigenome assay has demonstrated that abrogation of the LC8-RavP interaction leads to a severe attenuation in polymerase activity. We propose that RavP movement along the nucleocapsid is a rate-limiting step in viral transcription and replication, and LC8 facilitates this process by restricting the CTDs to a more ideal “walking” conformation (Figure 2.6). However, another possibility is that a common orientation of the CTD could help RavP to release the RNA from adjacent RavN molecules to provide RavL access to the RNA. Additionally, our minigenome data suggest that, while LC8 directly affects polymerase functionality, there are supplementary factors at play such as competitive binding between LC8 and a potential inhibitor of polymerase activity, evinced by the three levels of polymerase activity: RavP-WT (fast), RavP-Dis (intermediate), RavP-AAA (slow).

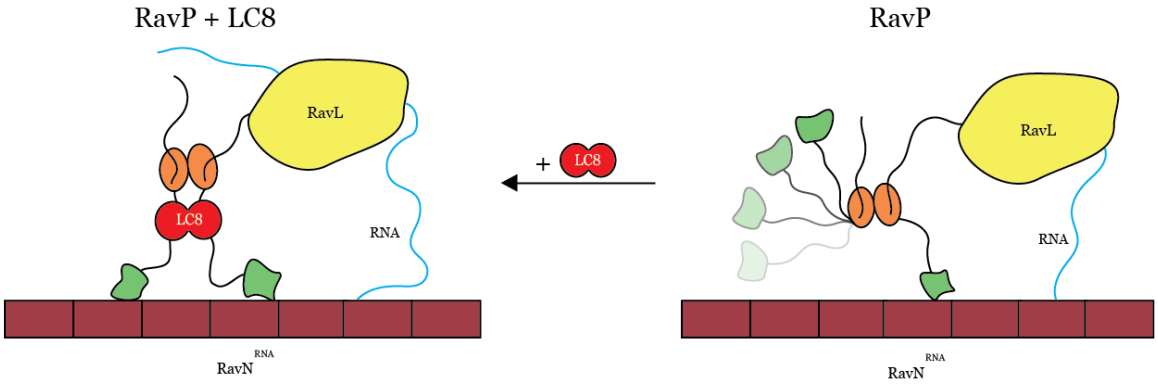


Figure 2.6: Models depicting how LC8 binding facilitates polymerase activity.
In the absence of LC8 (right), the CTDs (green ovals) occupy both distal and proximal conformations, leading to non-ideal movement of the polymerase along the $RavN^{RNA}$ complex. When RavP is bound to LC8 (left), CTDs are forced into a proximal conformation, facilitating polymerase movement along the $RavN^{RNA}$ complex and leading to increased polymerase efficiency, evinced by the minigenome assay comparison of RavP-WT to RavP-Dis (Figure 2.5E). The two levels of polymerase activity shown above are: fast (left), when RavP is bound to LC8 in the RavP-WT construct; and medium (right), when the entire LC8 motif in RavP is mutated to a disordered sequence, as in the RavP-Dis construct.

In support of our dimer elongation hypothesis, it is interesting to note that an analogous pair of related viral proteins, the VP35s of the Ebola and Marburg viruses, show a strikingly similar pattern in their LC8 binding sites. The Ebola and Marburg viruses are members of the *Filoviridae* family, and are in the same *Mononegavirales* order as RAV and VSV. The VP35 protein expressed in the *Filoviridae* family serves analogous functions to P, including suppression of the host immune response and bridging the viral polymerase and nucleoprotein to facilitate viral transcription and replication[148]. While the Ebola virus VP35 is known to bind LC8[119], the Marburg virus version does not contain the TQT motif necessary for LC8 binding (Figure 2.5F). Additionally, the LC8-binding site occurs in precisely the region occupied by the Marburg virus's longer multimerization domain. This pair of examples may suggest that, while RavP and Ebola VP35 can perform the same functions as VSVP and Marburg VP35 when bound to LC8, they have some additional viral functions in their free form. We propose that the lengthening of the dimerization domain, and subsequent restrictions in the CTD, allows RavP to switch between selected conformations within the ensemble, affecting various downstream functions. The high conservation of the LC8 motif within the *Lyssavirus* genus, implies that LC8 performs an important function not operating in other genera.

Previous work on 53BP1, an essential protein in the double strand break repair pathway, has demonstrated the importance of an LC8 binding site adjacent to a self-association domain[149].

While only one multimerization region is required to form oligomers, both sites are required for efficient localization and double strand break repair. These results parallel our own, and indicate that LC8 extension of dimerization to restrict conformations of folded domains is a novel function of LC8.

Perspective

We propose two potential functions for RAV incorporation of LC8. The first possibility, directly stemming from our structural findings and supported by our minigenome luciferase assay, is that the relative orientations of the two CTDs regulate viral transcription and replication. The structural ensemble of RavP-LC8 is highly biased towards conformations where the CTDs are oriented on the same side of the protein, as in VSVP. Efficient transcription and replication may require the correct orientation of these CTDs. Second, as RavP has a complex interactome, the LC8-induced dynamics restrictions could cause significant changes to the binding affinities of many downstream partners; LC8 may provide a mechanism for spatial or temporal regulation of these multifarious interactions. Structural, dynamics, thermodynamics, and minigenome data presented here suggest that LC8 binding results in a combination of both of these possibilities, nevertheless these models lay the groundwork for future innovative investigations in the LC8 viral infection-cycle.

Materials and Methods

Bacterial expression constructs and protein production

RavP constructs containing the N-terminal region (residues 1-152, RavP-N), the C-terminal region (residues 140-297, RavP-C) and full-length RavP (residues 1-297, RavP-FL) were generated using PCR, and then cloned into a modified pET24d expression vector with an N-terminal His_6 tag followed by a tobacco etch virus (TEV) protease cut site. LC8 was also sub-cloned into a pET24d expression vector with N-terminal His_6 tag and TEV cutting site. RavP-FL constructs with the TQT residues mutated to AAA (RavP-AAA), or the Cys at residue 261 mutated to a Ser (RavP-C261S), were generated using a QuikChange Lightning mutagenesis kit (Agilent, Santa Clara, California). RavP constructs where the LC8 motif is either deleted (RavP-Δ139-151) or mutated to a random coil sequence (RavP-Dis; $_{138}$RRSSEDKSTQTTGR$_{151}$ → $_{138}$SDNSNQEEPDASLQ$_{151}$) were generated using a Gibson assembly kit (New England Biolabs, Ipswich, Massachusetts), with a bridging primer. DNA sequences were verified by automated

sequencing. The recombinant vectors were transformed into BL21 Rosetta (DE3) cell lines for protein expression.

Bacterial cells were grown in LB medium at 37 ^{o}C to an optical density (A_{600}) of 0.6, followed by induction of protein synthesis with 0.5-0.8 mM isopropyl β-D-thiogalactopyranoside for 8-14 hours at 25 ^{o}C. Labeled protein was harvested from cells grown in M9 minimal media supplemented with ^{12}C or ^{13}C glucose and $^{15}NH_4Cl$ as the carbon and nitrogen sources. Cells were lysed via sonication in 50 mM sodium phosphate at pH 8.0 containing 300 mM sodium chloride, 1 mM sodium azide and 10 mM imidazole. The proteins were then purified by affinity purification on TALON metal affinity resin (Takara Bio USA, Mountain View, California), and eluted in 50 mM sodium phosphate at pH 8.0 containing 300 mM sodium chloride, 1 mM sodium azide, and 300 mM imidazole. The His_6 tag was cleaved using TEV protease before final purification on a Superdex^{TM} 75 HiLoad (26/600, GE Healthcare, Chicago, Illinois) size exclusion chromatography column in 20 mM sodium phosphate at pH 7.5 containing 150 mM sodium chloride and 1 mM sodium azide. Protein concentrations were determined from absorbance at 280 nm using molar extinction coefficients computed with the ProtParam tool on the ExPASy website[150].

Eukaryotic cell culture, transfection, and viral infection

HeLa Kyoto cells stably expressing LC8-GFP[151] were a kind gift from I. Poser and A. Hyman (Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany). The cells were cultured in DMEM medium with 10% (v/v) fetal calf serum and with 1% (v/v) penicillin/streptomycin.

pTIT plasmids [152] encoding RavL from RAV strain SAD-L16 (0.5 µg), RavN from CVS-11 (1.5 µg), and RavP-WT/RavP-AAA from CVS-11 (1.0 µg) were transfected into HeLa-LC8-GFP cells ($2*10^6$ cells) using Lipofectamine 2000 (Invitrogen, Carlsbad, California). These plasmids were cotransfected with 0.5 µg of a plasmid encoding the T7 RNA polymerase, as previously described[127].

The challenge virus standard (CVS, French CVS 1151) strain of rabies virus was grown in N2A cells. HeLa-LC8-GFP cells ($2*10^6$ cells) were infected with CVS at an MOI of 10, and collected for fluorescent imaging at 48 hours post-infection.

Minigenome luciferase reporter assay

We used the minigenome system as previously described[139,153], with slight modifications. BSR cells stably expressing the T7 RNA polymerase (BSR-T7) were transfected with pRL-TK (0.1 µg, encoding the control R*enilla* luciferase gene regulated by a thymidine kinase promoter), pDI-Luc (0.8 µg, encoding the firefly luciferase gene under control of the RAV polymerase), and pTIT plasmids encoding RavL or RavL-AAA (0.33 µg), RavN (1 µg), and RavP-WT, RavP-AAA, RavP-Δ139-151, or RavP-Dis (0.5 µg). Forty-eight hours after transfection, firefly and *Renilla* luciferase activities were measured using a dual-luciferase assay kit as described by the manufacturer (Promega, Madison, Wisconsin). The *Renilla* luciferase gene is constitutively expressed at low levels, and provides an internal control to account for differences in cell expression rates and concentrations. Therefore, our luminescence readings are a ratio of Firefly/*Renilla* luciferase activity. Experiments were performed in triplicate on two biological replicates.

IFNα-mediated luciferase assay

In IFNα-dependent signaling assays, HEK-293T cells were transfected with pRL-TK (0.75 µg), pISRE-Luc (2.5 µg), and pCDNA3.1 encoding RavP-WT, RavP-AAA, or an empty vector control (2.5 µg)[154]. Twenty-four hours after transfection, 'activated' cells were treated with 2,000 U/mL of human recombinant IFNα for six hours at 37 ºC, while 'non-activated' controls were not. Cells were then harvested, and firefly and *Renilla* luciferase activities were measured using a dual-luciferase kit (Promega). Values displayed in Figure 2.5D are a normalized ratio of activated to non-activated cells, and are the result of two biological replicates performed in triplicate.

Immunofluorescence and confocal microscopy

Cells were fixed for 10 min with 4% paraformaldehyde (PFA) and permeabilized for 10 min with 0.1% TritonX-100 in PBS. Cells were incubated for 1 hour at RT with the rabbit polyclonal anti-RavP antibody (Dilution of 1/1000) previously described [155], then washed and incubated for 1 hour with Alexa fluor conjugated secondary antibodies (Thermo Fisher Scientific, Waltham, Massachusetts). Following washing, cells were mounted with ImmunoMount (Thermo Fisher Scientific) containing DAPI. Images were captured using a Leica SP8 confocal microscope.

Isothermal titration calorimetry

ITC experiments for the interactions of LC8 with RavP-N, RavP-C, and RavP-FL were performed using a Microcal (North Hampton, MA) VP-ITC microcalorimeter at 25 °C in buffer composed of 50 mM sodium phosphate, 50 mM sodium chloride, 1 mM sodium azide, 5 mM β-mercaptoethanol, pH 7.5. In all experiments, an initial 2 µl injection was followed by 26 injections of 10 µl LC8 (300 µM) into 25 µM RavP constructs in the sample cell. Protein samples and buffer were degassed before data collection. The data were processed using Origin 7.0 (Malvern Panalytical, Malvern, UK) and fit to a single-site binding model. The recorded data are the averages of three independent experiments, with uncertainty reported as the standard deviation from the mean or the error in the fit, whichever was greater.

Paramagnetic labelling

Purified RavP-C261S was labelled with MTSL (*S*-(1-oxyl-2,2,5,5-tetramethyl-2,5-dihydro-1H-pyrrol-3-yl)methyl methanesulfonothioate; Toronto Research Chemicals, Toronto, Canada) at the sole remaining Cysteine, C297, as described in[156]. Briefly, SEC-purified, 100 µM RavP-C261S was incubated with 10 mM DTT for 2 hours at 25 °C to ensure that C297 was reduced. The sample was then buffer exchanged using a HiTrap Desalting column (GE Healthcare) into an MTSL labelling buffer composed of 50 mM Tris, 100 mM sodium chloride, 1 mM sodium azide, pH 7.5. A stock of 50 mM MTSL in ethanol was added to RavP-C261S at a 15X molar excess (1.5 mM MTSL in 100 µM RavP-C261S), and labelling proceeded in the dark at 4 °C for 16 hours. Labeled samples were then buffer exchanged into NMR buffer (50 mM sodium phosphate, 100 mM sodium chloride, 1 mM sodium azide, 5 mM ethylenediaminetetraacetic acid, pH 6.5) to remove any excess MTSL label, and concentrated to 150 µM RavP-C261S. LC8-bound samples were made by adding 1.8 mM LC8 in NMR buffer directly to pre-labelled RavP-C261S at a molar ratio of 1.2:1 LC8:RavP-C261S. After paramagnetic NMR spectra were collected, samples were reduced by the addition of 2 mM Ascorbate (100 mM stock in NMR buffer), followed by a 2 hour incubation period to ensure complete reduction before diamagnetic spectra were collected.

Circular dichroism

Spectra were recorded on a JASCO 720 spectropolarimeter using a 1 mm cell at a concentration of 10 µM in 20 mM sodium phosphate, 1 mM sodium azide, pH 7.5, at 25 °C.

Spectra for samples of LC8-bound proteins were collected using a 1:1 molar concentration of binding partners. Difference spectra were obtained by subtracting the 10 μM LC8 spectrum from the spectra of the bound form.

NMR measurements

NMR measurements were collected at 25 ºC using 150 μM $^{13}C/^{15}N$ labeled RavP-FL, RavP-N, or RavP-C in a buffer containing 50 mM sodium phosphate (pH 6.5), 100 mM sodium chloride, 1 mM sodium azide, 5 mM β-mercaptoethanol, 5 mM ethylenediaminetetraacetic acid, a protease inhibitor mixture (Roche Applied Science), 2,2-dimethylsilapentane-5-sulfonic acid for ^{1}H chemical shifts referencing, and 10% D_2O (vol/vol). Experiments for the LC8-bound samples were collected using a 1:1 molar ratio of 150 μM LC8:RavP construct. Band selective excitation short transient (BEST) ^{1}H-^{15}N TROSY-HSQC[157] spectra were compared before and after 3D experiments to verify sample stability. Additionally, samples were checked for degradation using sodium dodecyl sulfate polyacrylamide gel electrophoresis.

Backbone resonance assignments for constructs were determined using triple resonance experiments, including HNcoCACB, HBHANH, HBHAcoNH, HNCO, HNCA, HNcoCA, HNcaNNH, HNcocaNNH[158], and hCCcoNHs[159]. These experiments were carried out on a Bruker 700-MHz spectrometer. Backbone resonance assignments for RavP-FL were determined from comparison of HSQC spectra of the constructs with BEST ^{1}H-^{15}N TROSY-HSQC spectra of RavP-FL, and verified using a BEST TROSY-HNCO experiment collected on a Bruker 800-MHz spectrometer.

Longitudinal (T_1) and transverse (T_2) relaxation times were determined at 800-MHz using standard HSQC-based pulse sequences[160]. T_1 measurements involved eight time points that ranged from 20 ms to 1200 ms, and were collected as an interleaved experiment. The T_2 experiments had nine time points ranging from 16.96 ms to 271.36 ms. All T_1 and T_2 experiments used a recovery delay of 2 s. Two-dimensional TROSY-based steady-state ^{1}H-^{15}N NOE experiments were collected using a simultaneous recovery delay and saturation period of 6 s. Additionally, a three-dimensional HNCO-based ^{1}H-^{15}N NOE experiment was used to measure heteronuclear NOE values for peaks that overlapped in the two-dimensional experiment.

Phase modulated CLEAN chemical exchange (CLEANEX-PM) experiments were run using Fast-HSQC detection on an 800-MHz Spectrometer at 25 ºC, pH 6.5, and with a mixing time of 50 ms[136]. HNCO-based CLEANEX spectra were also collected for RavP-FL samples

under the same conditions. A recovery delay of 1.7 s was used for the HSQC-based experiment, and 2 s for the HNCO-based CLEANEX [159]. All dynamics data were collected at 800-MHz.

Residual dipolar couplings were collected at 800-MHz for linker residues using the *amide RDCs by TROSY spectroscopy* (ARTSY) method, with a ^{1}H dephasing duration of 10.75 ms[161]. This method utilizes a ratio of interleaved reference and attenuated spectra to determine the $^{1}J_{NH}$ and $^{1}D_{NH}$ values, rather than traditional splitting experiments, and therefore bypasses some of the issues with extreme peak broadening in the upfield component of the split peaks. Free and LC8-bound RavP samples were run both in standard NMR buffer, as well as in pf1 phage aligned solution (Asla Biotech, Riga, Latvia). A 50 mg/mL stock of Pf1 phage was added into samples to a final concentration of 10 mg/ml, where significant ^{2}H splitting was seen.

PRE data were collected at 800-MHz using a ^{1}H-^{15}N HSQC-TROSY pulse sequence, and data are presented as a ratio of the oxidized to reduced peak intensities. Spectra were compared to unlabeled spectra to verify that the MTSL label was not impacting RavP structure.

NMR data analysis

All spectra were processed with either TopSpin (Bruker BioSpin, Billerica, Massachusetts) or NMRPipe[162], and analyzed using Sparky[163]. C_α, C_β, and C' chemical shifts were compared to random coil values at the specified temperature and pH to determine the secondary structure[164,165].

T_1 and T_2 relaxation experiments were analyzed using Sparky[163]. Errors for ^{1}H-^{15}N NOE and peak intensity ratios were calculated from the intensities of the baseline noise. Errors for RDCs were calculated as described in[161]. Reduced spectral density mapping was performed on the NMR relaxation data as previously described[135,166] to yield values for spectral densities $J(0)$, $J(\omega_N)$, and $J(0.87\omega_H)$.

SEC-MALLS

Size-exclusion chromatography (SEC) combined with multi-angle laser light scattering (MALLS) and refractometry was performed at 20 °C with a flow rate of 0.5 mL min^{-1} on a Superdex 200 HR 10/300 GL column (GE Healthcare) equilibrated with 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 50 mM arginine, 50 mM glutamate and 0.2 mM TCEP buffer supplemented with a protease inhibitor cocktail (Complete EDTA-free, Roche). MALLS detection was performed with a DAWN-HELEOS II detector (Wyatt Technology, Santa Barbara, California) using a 690 nm laser light source. The protein concentration was measured with an Optilab T-rEX

detector (Wyatt Technology) using a refractive-index increment, *dn*/*dc*, of 0.185 mL g^{-1}. Weight-averaged molar masses (MW) were calculated with ASTRA (Wyatt Technology)[167].

Small-angle x-ray scattering

SAXS data were collected at the BioSAXS beamline BM29 at the European Synchrotron utilizing an in-line analytical SEC S200 column (10/300 GL, GE Healthcare). SAXS measurements were performed with a Pilatus 1M detector at a distance of 2.867 m allowing a q range of 0.03 to 4.94 nm with a wavelength of 0.09919 nm. The scattering from the buffer before and after each sample measurement was used for background subtraction. Data were analyzed using Primus (ATSAS)[168] and ScÅtter (Bioisis) program suites[169]. The volume of correlation method was used to determine the molecular mass[170]. The radius of gyration (R_g) was calculated using both the Guinier approximation and the pair distribution function. SAXS datasets were scaled and averaged to produce one unique $I(q)$ curve.

Model generation and molecular dynamics simulation

A combined atomistic coarse-grained and classical explicit solvent molecular dynamics simulations approach was used to generate ensembles of conformers suitable for fitting the SAXS data for both the free and bound complex. Models of the LC8-RavP complex were built with the rigid body SAXS modeling software CORAL[168] using the structures of the DD and CTD domains[124,129], as well as a representative LC8-peptide structure. As CORAL builds the regions for which no high-resolution structure are available as beads, the resulting model was then processed to generate an all-atom model. The all atom model of the LC8-RavP complex was then used as input for molecular dynamics simulations. To generate the starting model for RavP, LC8 was deleted from the model of the complex. Models of RavP and LC8-RavP were then simulated in GROMACS 5[171] using either an atomistic coarse-grained structure-based model[172] or explicit solvent classical molecular dynamics simulations (MDS). In the case of the structure-based model MDS, a time step of 0.0005 time units was used and the simulation was coupled to a temperature bath via Langevin dynamics. A single 100 ns trajectory was obtained for both RavP and LC8-RavP, and snapshots were extracted every 50 ps to create two ensembles with 2000 models each. In the case of classical MDS, we generated multiple trajectories for an aggregated simulation time of ~ 250 ns for RavP and ~ 400 ns for LC8-RavP. MDS was performed using charmm22*[173] to simulate collapsed states, and the amber03ws and amber99SBws forcefields[174] to reproduce the properties of intrinsically disordered proteins (SI Figure 2.4). At the beginning of each simulation, the protein was immersed in a box of SPC/E water, with a minimum distance of 1.0

nm between protein atoms and the edges of the box. The genion tool was used to add 150 mM NaCl[171]. Long range electrostatics were treated with the particle-mesh Ewald summation[175]. Bond lengths were constrained using the P-LINCS algorithm. The integration time step was 5 fs. The v-rescale thermostat and the Parrinello–Rahman barostat were used to maintain a temperature of 300 K and a pressure of 1 atm. Each system was energy minimized using 1,000 steps of steepest descent and equilibrated for 500 ps with restrained protein heavy atoms prior to production simulations. Radius of gyration versus time, RMSF and interdomain distances were calculated using GROMACS routines. Snapshots were extracted every 200 ps from each trajectory, leading to the generation of ~ 1250 models of RavP and ~ 2000 models of LC8-RavP.

For each model from the RavP and LC8-RavP ensembles, theoretical SAXS patterns were calculated with CRYSOL[176] and ensemble optimization fitting was performed with GAJOE[177]. The optimum selected ensemble size and relative weights of the models were determined automatically by GAJOE. Goodness of fit was demonstrated by the low χ values using the ensemble optimization method (χ_{EOM}).

By comparison with other methods for constructing ensembles (e.g. Ensemble[178] or Asteroids[179]), simulations of trajectories can provide an easier way to generate all-atom models of physically accessible conformers for multimeric molecules or multi-molecular complexes. The possible drawback with classical MDS that intrinsically disordered proteins are too compact can be avoided by combining different methods and different force fields, in particular by using force fields specifically devised for disordered proteins[174]. Clearly, our different trajectories started from an initial conformation that has an R_g value larger or equivalent to the experimental R_g value, and although some led to more compact conformations (those using Charmm force fields), others led to larger conformers (those using SBM or amber force field; SI Figure 2.4). Extracting conformers from these different trajectories generates ensembles with great diversity of compactness. Modeling SAXS data with ensembles generated by SBM or MDS simulations has been reported for different systems (e.g.[180–182]).

Structure prediction and ensemble visualization

Secondary structure predictions were obtained via PSIPRED v3.3[131] and Agadir[133]. Regions of order and disorder were predicted using DISOPRED3[183] and IUPred2A[184]. Domain boundaries were predicted using DomPred[131]. Ensemble analysis of the models generated from the SAXS profile was carried out using the Ensemblator program[137,185], which aligns structures based on common regions, and describes regions of variability using a variety of complementary

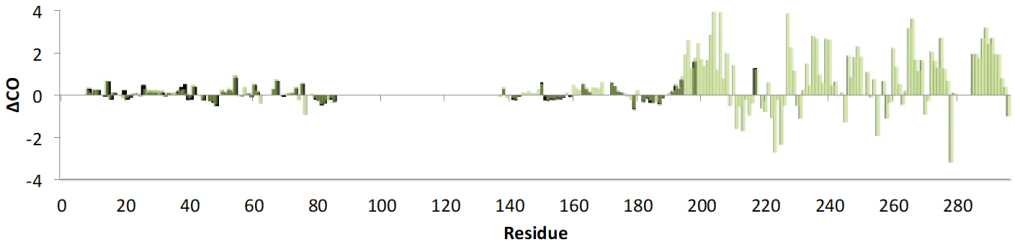
methods. Here we use the locally overlaid dipole residual (LODR) function to get information on residue-level backbone similarity from a pairwise comparison of all models in the ensemble. The LODR analyzes differences in local structure by aligning each dipeptide within the protein and calculating the RMSD for the backbone of the next amino acid within the protein sequence.

Phylogenetics

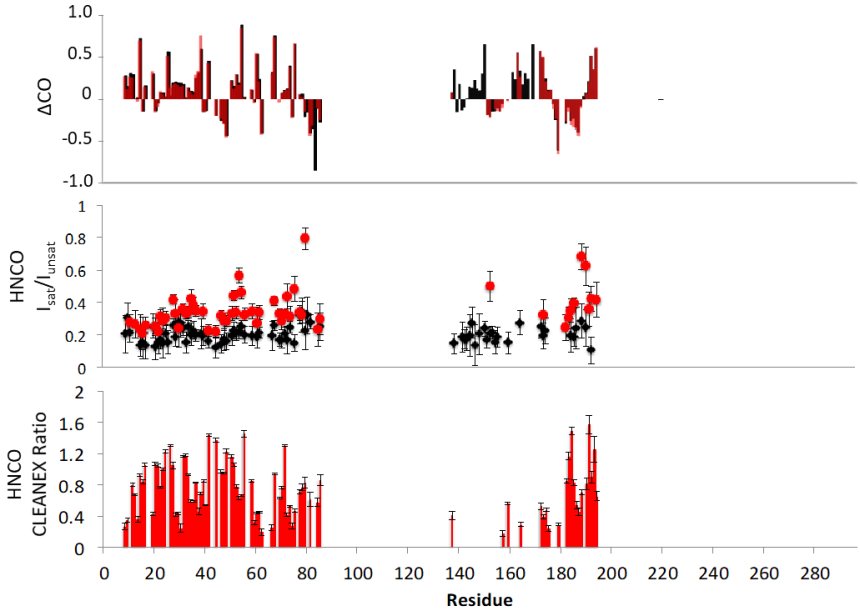
Phosphoprotein sequences for viruses in the *Rhabdoviridae* family were obtained from the UniProt server (accessed 03/2018). Alignments and trees were generated with Clustal Omega[186], by comparing representative phosphoproteins for viruses within different genera in the *Rhabdoviridae* family. Representative viruses were lettuce necrotic yellow virus (*Cytorhabdovirus*), orchid fleck virus (*Dichorhavirus*), rabies virus (*Lyssavirus*), viral hemorrhagic septicemia virus (*Novirhabdovirus*), perch rhabdovirus (*Perhabdovirus*), Drosophila melanogaster sigmavirus (*sigmavirus*), spring viremia of carp virus (*Sprivivirus*), Tibrogargan virus (*Tibrovirus*), Durham virus (*Tupavirus*), vesicular stomatitis Indian virus (*Vesiculovirus*). Phosphoproteins from three genera that did not show a similar domain architecture were excluded from the analysis (*Ephemerovirus*, *Varicosavirus*, and *Nucleorhabdovirus*). Sequence conservation analysis for phosphoproteins related to RavP was carried out using the ConSurf server[187], with the final conservation coloring based on 48 unique sequences with a maximum percent identity of 85%. Conservation scores were then mapped onto a representative protomer from within the SAXS ensemble.

Acknowledgements

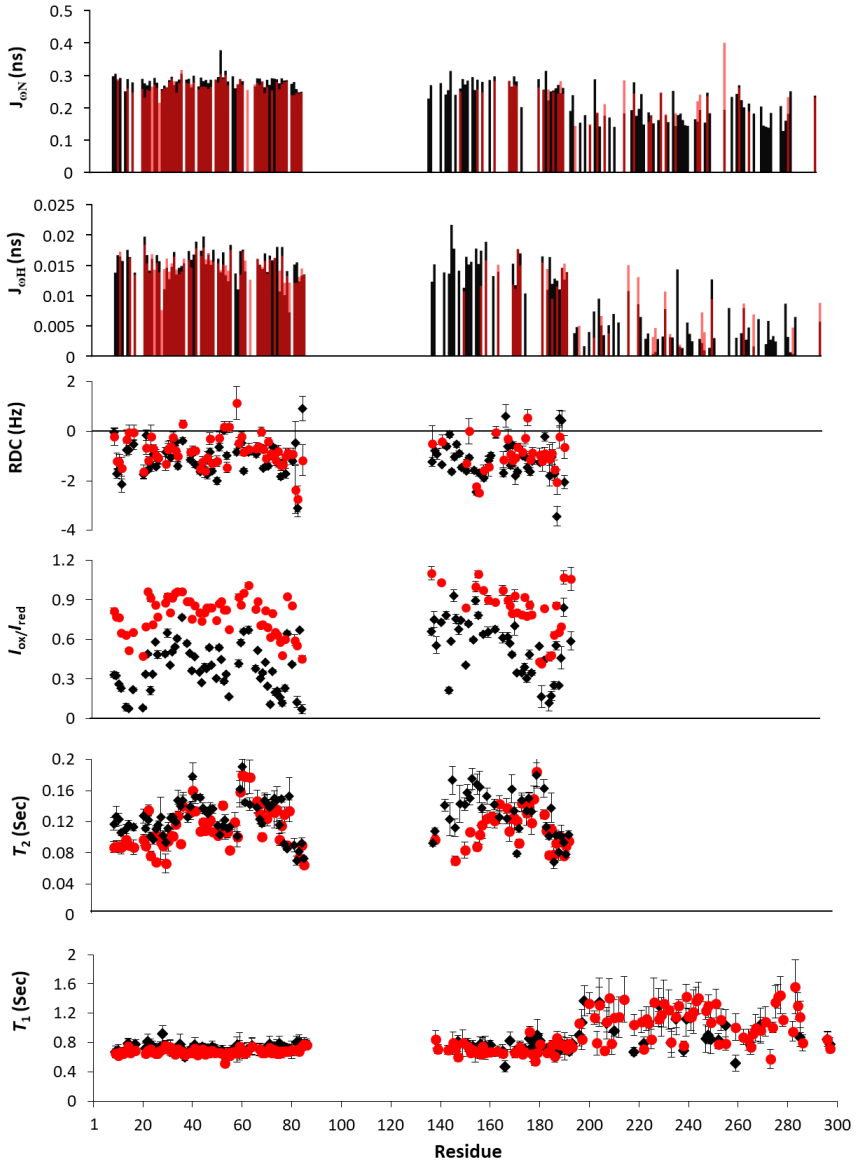
N.J. was supported in part by the Chateaubriand Fellowship of the Office for Science & Technology of the Embassy of France in the United States. This work was supported by National Institutes of Health Grant GM 084276 to E.B. and by an Impact Award to E.B. from the College of Science at Oregon State University. The authors also acknowledge Helena Kovacs and Rainer Kuemmerle at Bruker BioSpin for help with NMR data collection, Nathalie Scrima for technical support, Grant Larson for figure development assistance, and Andrew Brereton, Nikolaus Loening, and Patrick Reardon for insightful conversations and data interpretation. Support to facilities includes the Oregon State University NMR Facility funded in part by the National Institutes of Health, HEI Grant 1S10OD018518, and by the M. J. Murdock Charitable Trust grant # 2014162. This work was also supported by the Fondation de la Recherche Médicale (Equipe FRM DEQ20170336754) and used the platforms of the Grenoble Instruct Center (ISBG: UMS 3518 CNRS-CEA-UJF-EMBL) with support from FRISBI (ANR-10-INSB-05-02) and GRAL (ANR-10-LABX-49-01) within the Grenoble Partnership for Structural Biology (PSB).



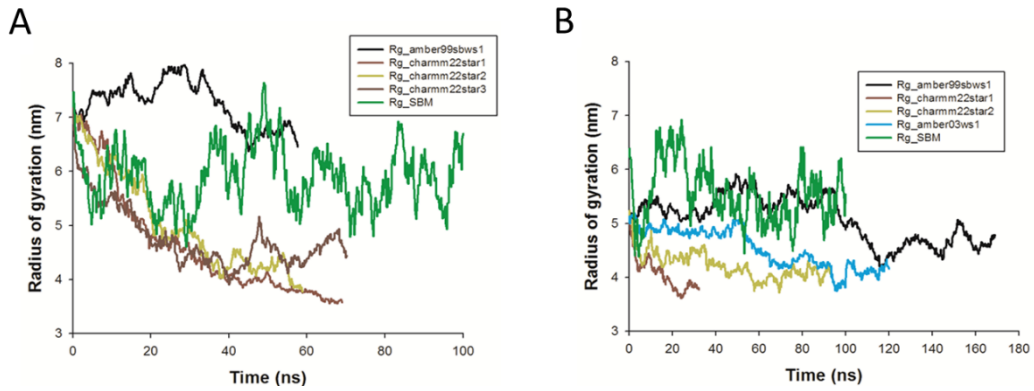
SI Figure 2.1: Change in carbonyl secondary chemical shifts for RavP-FL and RavP constructs. ΔCO was calculated as the difference between CO values for predicted random coils and RavP-FL (black) and RavP-C/RavP-N constructs (green). CO random coil chemical shifts were obtained from the Poulsen IDP/IUP generator[164,165]. C-terminal domain peaks were not visible in RavP-FL HNCO spectra. See also Figure 2.2.



SI Figure 2.2: HNCO-based experiments for RavP-FL apo (black) and bound to LC8 (red). Bound samples contained a 1:1.2 ratio of RavP-FL:LC8. ΔCO was calculated as the difference between CO values for experimental and predicted random coils for RavP (black) or RavP-LC8 (red). CLEANEX ratios are calculated as the intensity ratios between free and bound proteins, and were collected under identical conditions. In CLEANEX experiments, residues for which the error rates were greater than 25% of their value were removed from the graph. See also Figure 2.3.



SI Figure 2.3: Fits for relaxation experiments and spectral density function analyses. Spectral density function $J_{\omega N}$ and $J_{\omega H}$ values indicate there is very little difference between apo RavP-FL (black), and RavP bound to LC8 (red) at these timescales. $^{1}D_{NH}$ values are unchanged between free and bound ravp, and lie near 0, suggesting disorder. PRE values are shown as intensity ratios for oxidized (paramagnetic) to reduced (diamagnetic) samples, with error bars calculated from the background noise. T_1 and T_2 relaxation data are also shown with error bars. Spectra were collected in 150 μM RavP, in 150 mM NaCl, 20 mM MES, 1 mM NaN_3, 5 mM EDTA, pH 6.5. Relaxation times (T_1 and T_2) for RavP-free (black) and RavP-bound (red). Error bars represent the error in the fit, and residues with fit errors greater than 25% of their T_1/T_2 value are not shown. See also Figure 2.3E.



SI Figure 2.4: Molecular dynamics trajectories.
Radius of gyration versus simulation time extracted from the molecular dynamics trajectories that were used in the ensemble generation procedure for RavP (A) and LC8-RavP (B). In each case, the variations of the radius of gyration over the course of the trajectory is shown as a coloured line as indicated (trajectories obtained using coarse-grained structure-based models are shown in green, amber99sbws classical M.D. trajectories are shown in black, amber03ws classical M.D. trajectories in cyan, and charmm22* classical M.D. trajectories in brown and yellow).

Chapter 3

Systematic Identification of Recognition Motifs for the Hub Protein LC8

Nathan Jespersen, Aidan Estelle, Nathan Waugh, Norman E Davey, Cecilia Blikstad, York-Christoph Ammon, Anna Akhmanova, Ylva Ivarsson, David A. Hendrix, and Elisar Barbar

Published in *Life Science Alliance* (2019) 2, e201900366

Abstract

Hub proteins participate in cellular regulation by dynamic binding of multiple proteins within interaction networks. The hub protein LC8 reversibly interacts with more than 100 partners through a flexible pocket at its dimer interface. To explore the diversity of the LC8 partner pool, we screened for LC8 binding partners using a proteomic phage display library composed of peptides from the human proteome, which had no bias towards a known LC8 motif. Of the identified hits, we validated binding of 29 peptides using isothermal titration calorimetry. Of the 29, 19 were entirely novel, and all had the canonical TQT motif anchor. A striking observation is that numerous peptides containing the TQT anchor do not bind LC8, indicating that residues outside of the anchor facilitate LC8 interactions. Using both LC8-binding and non-binding peptides containing the motif anchor, we developed the 'LC8Pred' algorithm that identifies critical residues flanking the anchor, and parses random sequences to predict LC8 binding motifs with ~78% accuracy. Our findings significantly expand the scope of the LC8 hub interactome.

Introduction

Most proteins interact with few partners, but a class of proteins referred to as hubs interact with a large number of partners in complex protein-protein interaction networks[76,188]. Hubs can be static or dynamic. Static hubs bind a large number of partners simultaneously at different sites, for example BRCA2[77]. Dynamic hubs bind multiple partners that compete for the same site[78,79]. Well-known examples of dynamic hubs include calmodulin and 14-3-3 proteins[80–82]. A more recently discovered member of dynamic hub proteins is the dynein light chain LC8[32].

There are more than 280 binary interactions for human LC8 in the Mentha database[189], some of which have been extensively studied, including the dynein intermediate chain[25,63,64] and the transcription factor ASCIZ[67–69]. Additionally, expression patterns show that LC8 is highly expressed across a wide variety of cell types[70], and is broadly distributed within individual cells[71,190].

LC8 is an 89 amino acid homodimeric protein first identified as a subunit of the dynein motor complex. Co-localization and binding studies with dynein led to a common perception that LC8 functions as a dynein 'cargo adaptor' to facilitate transport of dynein cargo[12,34]. However, further studies have shown that LC8 interacts with many proteins not associated with dynein at the same symmetrical grooves in the LC8 dimer interface (Fig 1A). Because of the symmetry of the binding sites of the LC8 dimer, and its association with dimeric proteins, it is now generally accepted that LC8 serves not as a cargo adaptor in the dynein machinery, but rather as a dimerization hub in a variety of systems[32].

LC8 interacts with an eight amino-acid recognition motif within intrinsically disordered regions of its partners. Sequences bound to LC8 form a single β-strand structure integrated into an LC8 antiparallel β-sheet[35]; Figure 3.1). While there is some variation in the binding motif, it is most frequently anchored by a TQT sequence [44]. The glutamine in the TQT anchor is typically numbered as position 0, because it is the most highly conserved amino acid[20]. The flanking threonines are therefore defined as positions 1 and -1. The TQT anchor is highly enriched among known LC8 partners and will be referred to in this paper as the "motif anchor" (Figure 3.1B)[44].

A dynamic binding interface, determined from nuclear magnetic resonance (NMR) relaxation and hydrogen/deuterium exchange experiments[20,191,192], allows for large sequence variation in LC8 binding partners; however, several steric and enthalpic restrictions are placed on binding sequences. One restriction is inferred from analysis of solvent accessible surface areas of peptides bound to LC8 (Figure 3.1C)[44]. The side chains of the amino acids at positions –1 and 1

of the peptide (both threonines in Figure 3.1C) are completely buried, leading to a strong preference for amino acids with branched side chains that are either hydrophobic or, as is the case for threonine, participate in hydrogen bonding. In fact, these two positions are the only side chains that are completely buried (Figure 3.1C, orange vs pink side chains), suggesting that these residues are under more stringent selective pressures. Interestingly, even though the amino acids on both sides of the anchor are highly variable, their side chains are easily fit within discrete pockets (Figure 3.1B). In contrast, outside of the eight amino-acid LC8 binding motif there is higher variability in amino acid sequence and in side chain rotamer conformations (Figure 3.1B). Analysis of these structures explains the preference for the 'TQT' anchor within the LC8 recognition motif, but falls short of capturing the spectrum of amino acids that can flank the anchor in potential binding sequences.

In an effort to determine a consensus binding motif, Rapali *et al.*[38] used phage display and randomized all eight amino acid positions of the motif except for the conserved glutamine at position 0, and determined VSRGTQTE to be the most thermodynamically favorable binding sequence[38]. While this experiment led to the discovery of multiple LC8 binding partners, the idea of a specific "consensus sequence" belies the dynamics of the LC8 binding site. Additionally, by selecting for the tightest binder, many weaker binders were likely outcompeted and therefore not visible in their study. Our goal in this work is to determine the extent of the variability in LC8 binding sequences flanking the motif anchor.

LC8 motif prediction analyses have increased the number of known binding sequences, and enhanced our understanding of the motif specificity[38,75]; however, algorithms generated in these studies were designed for initial screening and are therefore not sufficiently stringent for general use, nor made publicly available. Here, we use a combination of proteomic peptide phage display (ProP-PD) technology and position-specific scoring matrices to determine likely LC8 binding sequences. Interestingly, although our methods were unbiased with respect to the presence of a TQT anchor, sequences experimentally validated to bind LC8 all contained a TQT or variation of the TQT triplet. A database that includes partners identified in this work along with published interactions is now available and contains all 82 validated LC8 interactions. Finally, we used this database to develop an algorithm that utilizes both binding and non-binding sequences to effectively predict LC8 partners, and define rules for LC8 partner recognition that underscore the plasticity of the LC8 binding pocket.

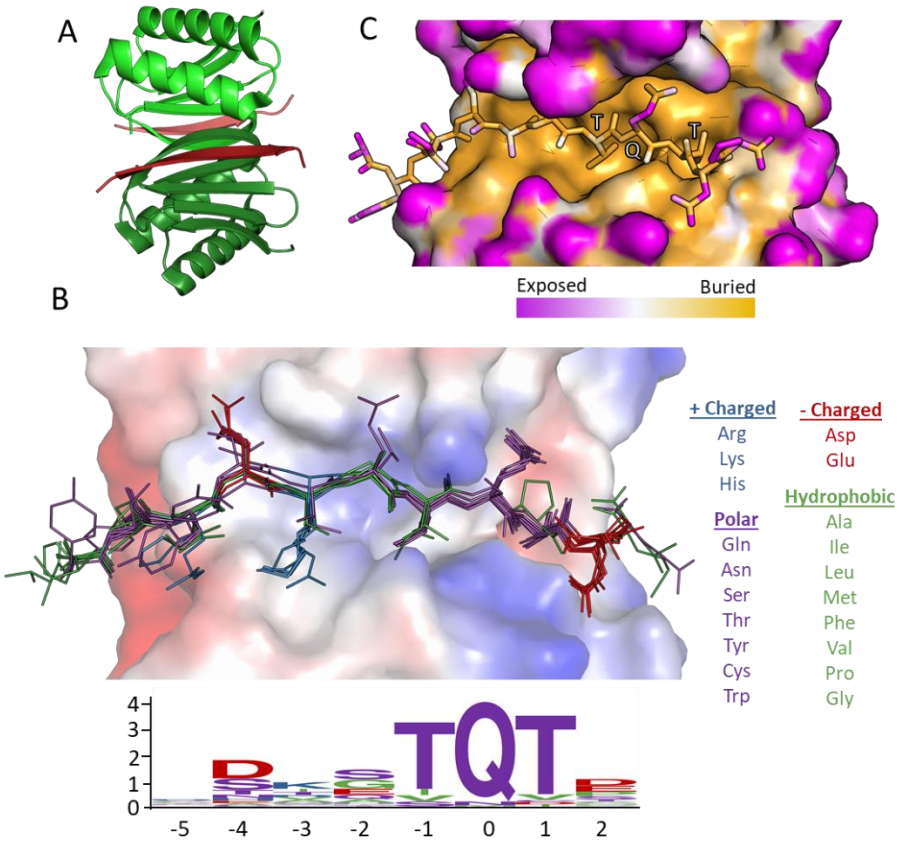


Figure 3.1: Motif sequence logo and surface analysis of LC8.
(A) Crystal structure of a representative LC8 dimer (protomers shown in shades of green) bound to a peptide (shades of red). (B) Electrostatic charge potential for a representative LC8 structure using PyMOL's smoothed-charge potential calculator, with positive potentials shown in blue, negative in red, and neutral in white. Peptides from available crystal structures of bound LC8 are shown, and colored based upon amino acid chemical characteristics (right). Amino acid enrichment is shown below each position within the LC8 binding motif, calculated from 79 known binder motifs listed on the LC8 database (http://lc8hub.cgrb.oregonstate.edu). Amino acid heights represent relative enrichment of that amino acid. (C) Solvent accessible surface area depiction of the same LC8/peptide pair shown in A. Color scheme was defined at the atom level using the GetArea program[32], with magenta representing more solvent exposed, and orange regions more buried atoms.

Results

LC8 is broadly distributed in cells

To examine the subcellular distribution of LC8, we used HeLa cells stably expressing endogenous levels of a C-terminally tagged LC8-GFP fusion generated by BAC TransgeneOmics[151]. In interphase cells, LC8-GFP is present throughout the cytoplasm, within the cell nucleus, and enriched in patch-like structures at the cell cortex located in the vicinity of focal adhesions (Figure 3.2A). In mitotic cells, LC8-GFP is present at the spindle and enriched at the spindle poles and prometaphase kinetochores (Figure 3.2B). While the mitotic LC8-GFP localization is consistent with LC8 being part of the cytoplasmic dynein complex[193], the nuclear localization and cortical accumulations in interphase cells are not observed for the other core subunits of cytoplasmic dynein, such as the heavy, intermediate or light intermediate chains, which were previously tagged with GFP and detected in a similar manner[194].

In agreement with the imaging data, analysis of localization patterns for known LC8 partners performed using the COMPARTMENTS database[195] show that they can be found in multiple cellular compartments, such as the cytoplasm, nucleus and vesicles (Figure 3.2C). Surprisingly, LC8 partners can even be found in extracellular space. Localization within the various categories of subcellular structures supports the conclusion that LC8 is broadly distributed throughout the cell irrespectively of dynein, and concentrates at certain subcellular sites where LC8 partners are enriched.

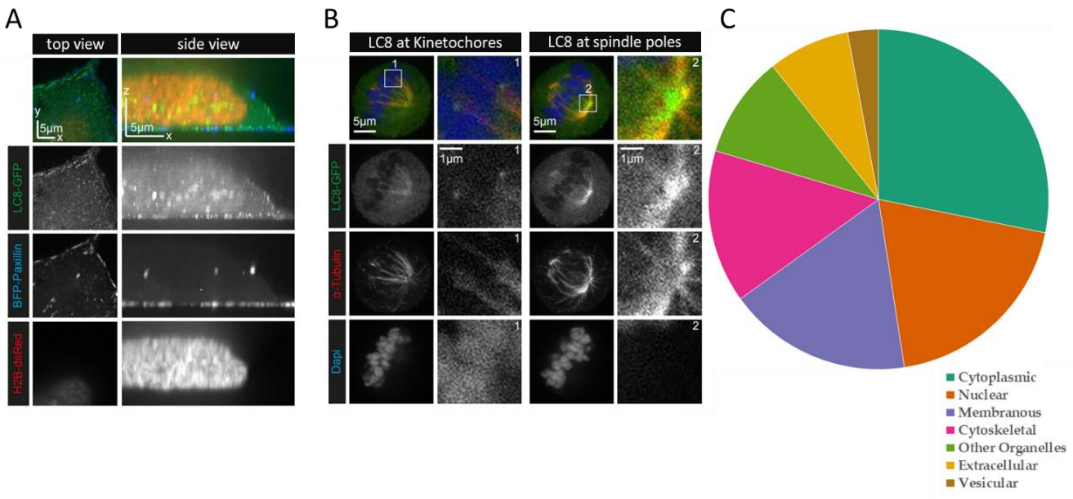


Figure 3.2: LC8 and its binding partners are broadly distributed.
(A) Live HeLa cells stably expressing LC8-GFP (green) were transiently transfected with the focal adhesion marker TagBFP-paxillin (blue) and the nucleus marker dsRed-tagged histone H2B (red). The top view images shown on the left represent an optical section located next to the coverslip. LC8-GFP is present throughout the cell, but forms puncta at the cell cortex. (B) HeLa cells stably expressing LC8-GFP were fixed with paraformaldehyde (PFA) and stained for the endogenous α-tubulin (red) and with DAPI (blue) to visualize the DNA. LC8 accumulates at the kinetochores (box 1) and at the spindle poles (box 2). For both A and B, images were acquired using confocal spinning disc microscopy. (C) Localization information derived from the COMPARTMENTS program demonstrates that LC8 binding partners are localized to all cellular compartments.

ProP-PD selections identify 16 new LC8 interactions

For broad mapping of the LC8 interaction network, we used a Proteomic Peptide Phage Display assay (ProP-PD), in which a library is created that encodes sequences for peptides drawn from a genome of interest. These peptides are then synthesized by bacteriophages and displayed externally on bacteriophage coats. Our ProP-PD library is composed of 479,846, 16 amino-acid long peptides with overlapping segments designed from the disordered regions of 18,692 human proteins[196]. Through phage display selections using immobilized LC8 as bait protein we pulled down and sequenced 53 potential binding partners with highly variable sequences, including 5 previously identified partners (SI Table 3.1).

To validate the binding of these partners, we synthesized 14 amino-acid long peptides containing the motif anchor at positions 10-12, and tested their binding by isothermal titration calorimetry (ITC). Of the 53 synthesized peptides, 16 interacted with LC8 to varying degrees,

while 31 peptides showed no binding under our experimental conditions (Table 3.1). Binding of the remaining peptides was not tested due to their poor solubility.

UniProt	Gene	Sequence	Start	End	Binder?	Citation
O43521	BCL2L11	APMSCDKSTQTPSP	108	117	Y	[10]
Q9UPA5	BSN[II]	PRATAEFSTQTPSP	1498	1511	Y	[197]
Q86VQ1	GLCCI1	SSSTRSIDTQTPSV	340	353	Y	[198]
Q96R06	SPAG5	HPETQDSSTQTDTS	463	476	Y	[199]
Q6IMN6	CAPRIN2	NQSFTTASTQTPPQ	791	804	Y	
O75665	OFD1	AKESCNMETQTSST	153	166	Y	[200]
Q02505	MUC3A	PVLTSATGTQTSPA	1800	1813	Y	
Q9UBY0	SLC9A2	*DD*HSREKGTQTSG*D*	749	804	Y	
Q9Y2F5	ICE1	EKELRHIGTQISSD	181	194	Y	
Q9ULV3	CIZ1	ARAGRSVSTQTGSM	13	26	Y	[201]
Q99102	MUC4	SQNHWTRSTQTTRE	200	213	Y	
Q99102	MUC4	*DD*NHWTRSTQTTRE	200	213	Y	
P07359	GP1BA	GQGAALTTATQTTHLE	566	581	Weak[b]	
Q9Y4F4	TOGARAM1	SKTQQTFGSQTECT	788	801	Weak[b]	
Q8WWN8	ARAP3	SPSPTGLPTQTPGF	1514	1527	Weak[b]	
Q01973	ROR1	*DD*SGGNATTQTTS*D*	760	774	Weak[b]	
Q8NEZ4	KMT2C	IVSCVSVSTQTASD	205	218	N	
Q9UPA5	BSN-shift[a]	STQTPSPAPASDMP	1505	1518	N	
Q7Z2Z2	EFL1	*D*ERLMCTGSQTFD*D*	375	386	N	
Q02817	MUC2	TPTPTPTGTQTPTT	2000	2013	N	
Q9HC84	MUC5B-3	SMATPSSSTQTSGT	2673	2686	N	
Q8TEC5	SH3RF2	TLVSTASGTQTVFP	714	727	N	
Q9P2G1	ANKIB1	RGDGSDVSSQTPQT	1065	1078	N	
O43526	KCNQ2	*DD*PMYSSQTQTYG*D*	370	380	N	
P14859	POU2F1	ESGDGNTGTQTNGL	13	26	N	
P35568	IRS1	LPRKVDTAAQTNSR	841	854	N	
Q2KHR3	QSER1	KTLTFSGSSQTVTP	374	387	N	
Q99814	EPAS1	TEAKDQCSTQTDFN	509	522	N	
Q9Y4K1	CRYBG1	RSFVLPVESTQDVSSQ	550	565	N	
P49862	KLK7	SFRHPGYSTQTHVN	98	111	N	
Q92904	DAZL	TQDDYFKDKRVHHFRRS	272	288	N	
Q96FV2	SCRN2	VRTLPRFQTQVDRR	342	355	N	
Q96FV2	SCRN2	*DD*TLPRFQTQVDRR	344	355	N	
Q7Z589	EMSY	KITFTKPSTQTTNT	261	274	N	
Q13952	NFYC	CLKETLQITQTEVQ	289	302	N	
Q9HC84	MUC5B-1	TTLPVLTSTATKST	3049	3062	N	
P53350	PLK1	AASLIQKMLQTDPTAR	278	293	N	
Q92499	DDX1	*DD*HSGNAQVTQTKF*D*	271	282	N	
Q8NBH2	KY	ITSYNSQGTQLTVE	81	94	N	
Q06190	PPP2R3A	LQETLTTSSQANLS	625	638	N	
Q13618	CUL3	KHSGRQLTLQHHMG	542	555	N	
Q9H4B6	SAV1	NQSFLRTPIQRTPH	70	83	N	
Q2TV78	MSTL1	EGYRGTANTTTAAYLA	259	274	N	

Q6ZU65	UBN2	PLQATISKSQTNPV	942	955	N
Q96SC8	DMRTA2	SSRSAFSPLQPNAS	433	446	N
Q6ZRI0	OTOG	TLQQPLELTASQLPAG	1541	1556	N
Q96JG9	ZNF469	RAAALPEETRSSRR	1014	1027	N

Table 3.1. Peptides synthesized based on their binding in phage display experiments. Anchor motifs are underlined. Aspartates shown in italics were added to increase solubility. [a] BSN-ProP-PD is the Bassoon sequence pulled down by phage display, without shifting the TQT sequence into the correct position. [b] Peptides that displayed an interaction with LC8 via ITC, but the data were not of sufficient quality to obtain reliable K_d measurements.

Binding dissociation constants ranged from 0.16 μM for BCL2L11, to affinities too weak to be reliably determined by ITC (K_d greater than 25 μM cut-off). Representative thermograms for strong binders (data fit with high accuracy), weak binders (not fit), and non-binders are shown in Fig 3.3A. Under our experimental conditions, LC8 is a dimer[15]; therefore the 1:1 binding ratio in all ITC thermograms corresponds to 2 peptides binding to an LC8 dimer, as expected. Although the ProP-PD method has no innate bias toward anchor-containing sequences, all of the LC8-binding peptides contain a TQT-like sequence (Table 3.1). On the other hand, many anchor containing peptides failed to bind in our ITC assays, suggesting that the achor motif is strongly predictive of high-affinity binding, but insufficient to guarantee interactions with LC8.

There are several plausible explanations for the lack of binding of ProP-PD-derived peptides to LC8. In the hybrid M13 system used here, each phage particle displays hundreds of peptide copies on its surface. This avidity allows the capture of low/medium-affinity interactions, in the range of 40-150 μM[196,202]. Additionally, because LC8 binds much more tightly to dimeric than to monomeric partners due to the bivalency effect[192], the presence of multiple peptides in close proximity on the phage surface could facilitate binding to otherwise weak motifs. Thus, it is possible that the anchor-containing peptides that failed to bind in our ITC experiments, in fact do bind LC8, but with affinities weaker than 40 μM.

Available crystal structures of LC8 bound to partner peptides indicate that partners require amino acids N-terminal of the TQT anchor motif for the necessary backbone H-bonds to form a β-strand. Therefore, for the phage display identified peptides that had the TQT anchor at the N-terminus, the corresponding synthesized peptide sequence was shifted to have the TQT motif at the 10-12 position, and the flanking regions were replaced with the actual protein sequence at these positions to allow for at least five amino acids on the N-terminal side of the TQT sequence.

This design may partially explain why some peptides that were pulled down in the ProP-PD experiments did not bind via ITC, as the sequences used in these experiments were not

identical. Many of the nonbinding ProP-PD hits have their TQT anchor located near the N-terminus of the peptide. As each ProP-PD peptide is N-terminally flanked by an SSSG linker, their binding behavior might be expected to differ from native sequences, wherein flanking regions are different and might contain amino acids that impair LC8 binding. This effect is most clearly demonstrated by our results for two partially complementary peptides derived from the protein bassoon (BSN). Two versions of this peptide were synthesized for this experiment: one version exactly matching the ProP-PD hit with the TQT motif at the N-terminus (minus the SSSG tag, STQTPSPAPASDMP), and a modified version with the TQT motif near the C-terminus (PRATAEFSTQTPSP). The modified BSN peptide with a C-terminal TQT demonstrated strong binding to LC8, while the original phage-determined peptide with N-terminal TQT failed to demonstrate even weak binding to LC8. This example provides strong evidence that LC8-partner binding interactions are sensitive to the positioning of the TQT anchor within the full binding motif, and validate our concern that the SSSG linker erroneously facilitated binding for some sequences in the ProP-PD experiment.

Position-specific scoring and disorder prediction identifies 7 new LC8 interactions

We scanned *Homo sapiens* and viral proteomes for potential LC8-binding sequences based upon: 1) their propensity for disorder, 2) their sequence conservation across related species, and 3) the sequence's similarity to known motifs. This final parameter used a position-specific scoring matrix (PSSM) based on the relative enrichment or depletion of each residue in a potential sequence, weighted by amino acid frequency. Enrichment and depletion PSSMs were populated using sequences that interact with LC8, verified either by mutagenesis or *in vitro* assays. ITC experiments were performed on 19 synthetic peptides with high-scoring sequences, but only seven bound (Table 3.2). These include the human papilloma virus E4 protein, the rotaviral VP4, and human CCL2, CCL7, SON, MAST1, and ZFPM1 proteins. Notably, a previous study also predicted a binding site within the rotavirus A VP4 protein[203] at residue position 644-651 (IDMSTQIG); however, the synthesized peptide sequence tested in that study did not bind LC8. Here, we predict an alternative site at positions 605-612 (NDISTQTS) based on disorder propensity and motif similarity. ITC experiments confirmed that the site predicted here binds with a K_d of 4.2 μM (Figure 3.3B).

UniProt	Gene	Sequence	Start	End	Binder?	Citation
Q6LCS3	E4 (HPV)	YLQGRQEDKQTQTPPP	16	30	Y	
Q8IX07	ZFPM1	PAPPSYSDKGVQTPSK	947	962	Y	
P11193	VP4 (Rotavirus A)	YVTNSLNDISTQTSTI	600	614	Y	
P13500	CCL2	YDSMDHLDKQTQTPKT	85	99	Y	
P18583	SON	YSRKSRCVSVQTDPT	87	100	Y	
Q9Y2H9	MAST1²⁰⁴²⁰⁴²⁰⁴²⁰⁴²⁰⁴²⁰⁴²⁰⁴²⁰⁴	YGCTRHQSVQTEDG	1387	1399	Y	[204]
P80098	CCL7	QDFMKHLDKKTQTPKL	84	99	Y	
Q5K4E3	PRSS36	YGPDGEETETQTCPP	468	581	N	
P20702	ITGAX	YGQIAPENGTQTPSP	1146	1159	N	
P03586	MT/HEL (TMV)	AQPKQKLDTSIQTEYP	1305	1320	N	
Q8IYH5	ZZZ3	KSVAENGDTDTQTSMF	237	252	N	
Q5DMI6	DNLJ2 (phage T5)	YKIEIPTQCPSCGSK	2	15	N	
Q92904	DAZL	YPQKKSVDRSIQTVVS	243	257	N	
Q9NZ56	FMN2	YHHRILEAKSIQTSPT	735	749	N	
Q13418	ILK	MDDIFTQCREGN	1	12	N	
O43432	EIF4G3	DFTPAFADFGRQTPGG	676	691	N	
Q99613	EIF3C	YELMASLDQPTQTVVM	830	844	N	
O15444	CCL25	NKVFAKLHHNTQTFQA	94	109	N	
P20042	EIF2S2	KPFMLDEEGDTQTEET	21	36	N	

Table 3.2: Peptides synthesized based on favorable sequence comparisons.
Anchor motifs are underlined. Tyrosines shown in italics were added for accurate concentration determination.

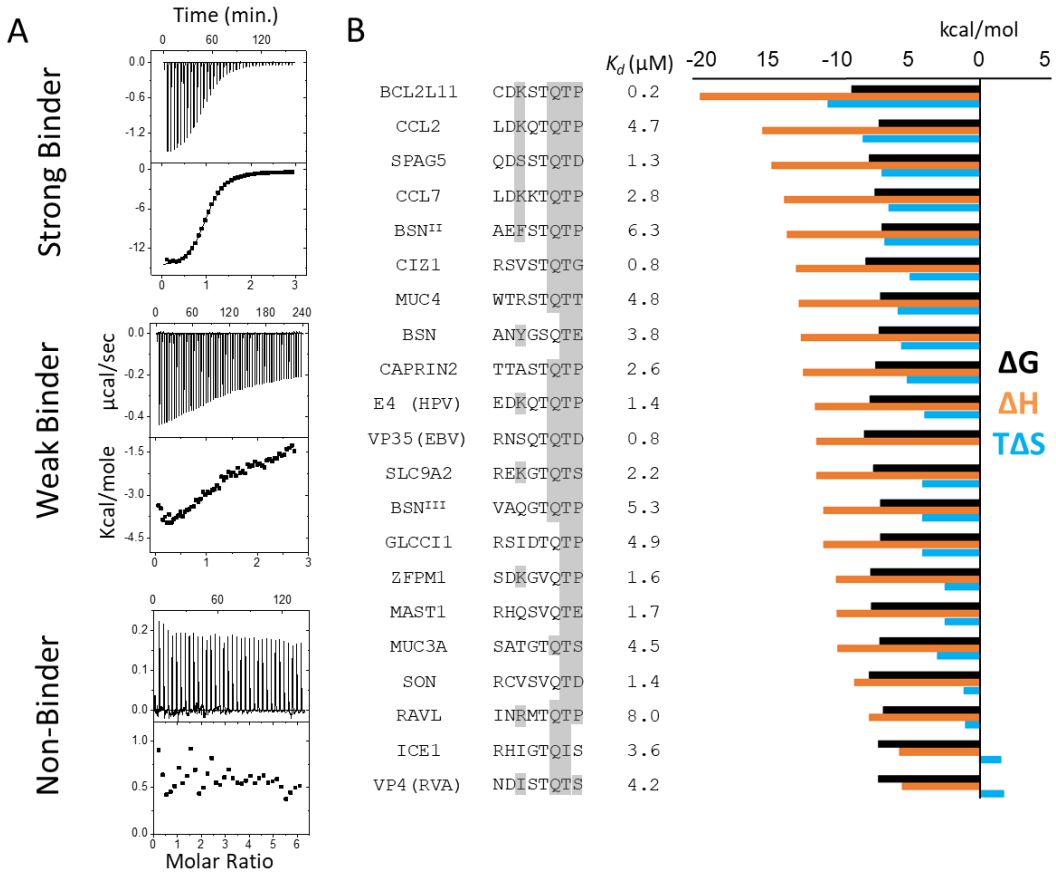


Figure 3.3: LC8-peptide binding thermodynamics.
(A) Representative ITC thermograms of LC8 bound to 14 amino acid-long synthetic peptides. The data were collected at 25 ºC in 50 mM NaCl, 50 mM NaPO4, 1 mM NaN_3, pH 7.5. Experiments were performed in triplicate. We categorized peptides as strong binders (K_d reliably determined; SPAG5), weak binders (heat generated, but unable to fit the data; TOGARAM1), and non-binders (QSER1). Weak binders are those with affinities >25 µM. (B) Binding affinities and thermodynamic parameters for strong LC8 binders identified in this study. Thermodynamic parameters for all binding peptides at 25 ºC are shown. ΔG (black), ΔH (orange), and TΔS (blue) kcal/mol values are the average of 2-3 independent ITC experiments. K_ds are shown in µM. Eight amino acid motifs are shown, with residues capable of making conserved hydrogen bonding interactions highlighted in grey. Sequences are ordered by descending ΔH values.

Common motif features that promote LC8 binding

Since only 7 of the 19 PSSM-predicted binding sequences actually interact with LC8, it is clear that an additional filtering method should be introduced to minimize false positives. To assess common features for binding from this growing dataset of interactions, we overlaid all known tight binding partners (50 sequences with K_ds <10 µM, Figure 3.4B), as well as all non-binding sequences (determined here, Figure 3.4C). This comparison revealed some conspicuous

differences between binders and non-binders, allowing for the determination of the position-based rules that follow (Figure 3.4A,D).

The anchor is extremely well conserved in both amino acid type and volume. There is a strong preference for a mid-sized H-bonding/hydrophobic residue at positions -1 and +1, and a clear preference for a glutamine at position 0. Any deviation from this anchor, such as the RQT seen in EIF4G3, leads to a non-binding sequence. Both threonines are completely buried in crystal structures (Figure 3.1C), and therefore deviations to a charged group are highly unfavorable (Figure 3.4D, Poor Anchor).

Position +2, which has no β-strand backbone interactions in any crystal structures, shows a large preference for proline, aspartate, and glutamate residues. Interestingly, these three residues are classically depleted in β-strands[205,206], providing a potential explanation for these residues acting as "strand-breaking" amino acids at the periphery of the LC8 binding pocket. An alternative explanation for their enrichment is that the negative charge for E and D can interact with the positive electrostatic charge on LC8 (Figure 3.1B). Proline, on the other hand, might energetically assist in binding by reducing the change in entropy, as both proline and pre-proline residues are conformationally restricted[207]. Hydrophobic amino acids are not well accommodated at this position (Figure 3.4D, Hydrophobic +2).

Position -2 shows little charge preference and allows positive, negative, polar, and hydrophobic residues; however, there are no examples of bulky aromatic side chains at this position among the tightly binding peptides, indicating that there are some steric constraints (Figure 3.4D, Bulky hydrophobic -2).

Position -3 favors large side chains as nearly all tight binders contain an amino acid at least as large as valine at this position, with only two occurrences of an alanine. Fig 1B reveals a binding pocket where large side chains can fit, which is often occupied by lysines or arginines. A small side chain at the -3 position does not immediately exclude a sequence from binding, as in CAPRIN2 (Table 1), but seems to be less favorable based on the depletion of these residues (Figure 3.4D, Small -3).

Position -4 favors amino acids capable of making a polar contact, such as aspartate, and no sequences identified to date have hydrophobic residues larger than alanine at this position (Figure 3.4D, Hydrophobic -4). Finally, the -5 position shows a slight bias towards positively charged residues (Figure 3.4B,C), but it is unclear whether this effect is significant.

In general, partners must bind within a deep hydrophobic pocket and form a β-strand structure; therefore, multiple similar charges within a peptide, or sterically challenging prolines at

any internal position, makes binding unfavourable. Even with this systematic comparison, a number of the non-binding sequences could not be categorized (Figure 3.4D).

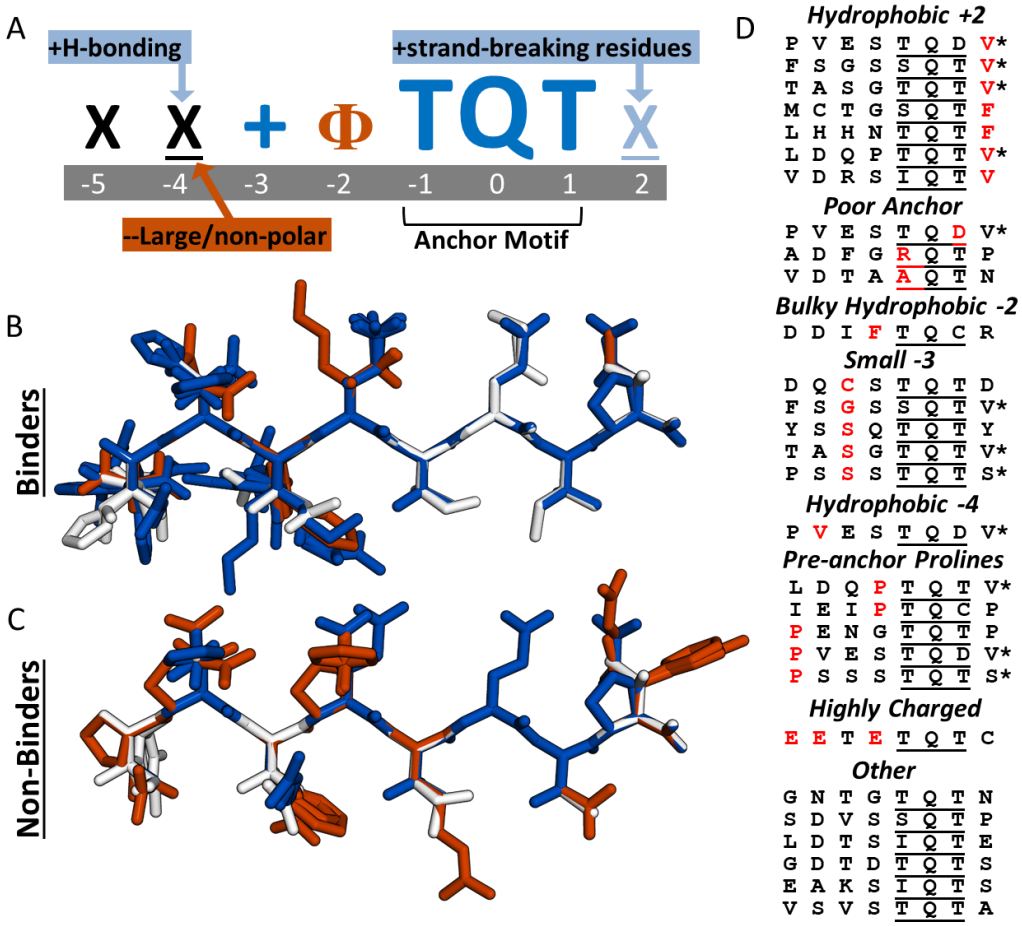


Figure 3.4: Analysis of LC8 binding and nonbinding motifs reveals positional preferences. (A) Motif preferences for LC8 binding partners. "Φ" denotes hydrophobic residues; "X" signifies any residue (unless certain residues are disfavored); underlined "X" signifies any residue but with strong preferences for particular residues; "+" denotes positively charged amino acids. Physiochemical properties beneficial for binding are colored dark blue or light blue, based on magnitude, and deleterious properties are colored in red. (B) All known tightly binding sequences (K_d < 10 μM) are cropped to eight amino acid motifs, and built using the Chimera molecular modeling software. This includes LC8 sequences found on the LC8Hub database, as well as those determined in this paper. (C) Overlay of all non-binding peptides used in this study. Residues are colored based upon whether they are beneficial (blue), deleterious (red), or neutral for binding (white), using the amino acid enrichment and depletion in known motifs (Figure 3.6A). (D) Categories of non-binding sequences. Residues highlighted in red depict what caused the sequence to be placed within a given category. * Denotes sequences that have been placed in multiple categories.

The partner-binding pocket is conserved in LC8 sequences but is structurally variable

A comparison of LC8 amino acid sequences from 58 different eukaryotic species using the ConSurf program [187] reveals that the partner binding site is strictly conserved across these diverse organisms (Figure 3.5A). Interestingly, the conservation of residues creates a noticeable gradient pattern that radiates out from the dimeric interface/partner binding site, with the most conserved residues near the core (maroon), and the least conserved residues at the peripheries (blue).

We used the Ensemblator program[137], which aligns independently determined 3D structures and identifies regions of structural conservation or plasticity, to visualize how a sequence that is strictly conserved is capable of binding such a wide variety of sequences. By overlaying the protomers from 5 published crystal and NMR structures of free LC8, we observed that the β-strand that directly binds to partners is highly variable (Figure 3.5B)[20,191,192], and has the highest RMSD values between structures. It is of note that the most sequence-conserved region is also the most structurally variable part of the protein. This structural plasticity allows accommodation of a diverse set of partners with a wide range of properties, and sheds light on why definitive identification of LC8-binding motifs is such a difficult task.

Enthalpic and entropic modes of binding

Crystal structures of LC8 bound to different peptides reveal surprisingly few conserved backbone and side chain H-bonds. Backbone H-bonding between the antiparallel β-strands occurs only for residues between positions -5 and -1 (Figure 3.5C,D). Tellingly, there are only five frequently observed side chain H-bonds, and four of these occur within the anchor (Figure 3.5E). The remaining interaction occurs at position -4, which is often an aspartate residue. The conspicuous lack of conserved side chain polar contacts, along with LC8's dynamic binding interface, suggests that most disordered (or extended) anchor-containing sequences should be capable of binding LC8.

Thermodynamic data obtained by ITC demonstrate that all of our peptides bind to LC8 in an enthalpically driven reaction (Figure 3.3B). Analysis of all peptides with strong anchors, high binding affinities, and similar K_ds, shows large differences in their ΔH and TΔS values, spanning 15 kcal/mol. In general, peptide sequences that contain TQT anchor sequence and the capacity for a polar contact at the -4 position have the largest ΔH (*e.g.* BCL2L11 and CCL2), while sequences like MAST1 or ICE1 that lack one or more of these interactions have a lower ΔH (Figure 3.3B). Some outliers (such as VP4, BSN, and RavL) have lower ΔH values. One possible explanation is that these peptides have predicted helical structures, which require additional energy to unfold

before binding, and thus result in smaller overall ΔH values. Indeed, the LC8-recognition motif within VP4 has the smallest ΔH (-5.6), and is predicted by IUpred to be ordered (average disorder propensity of 0.37).

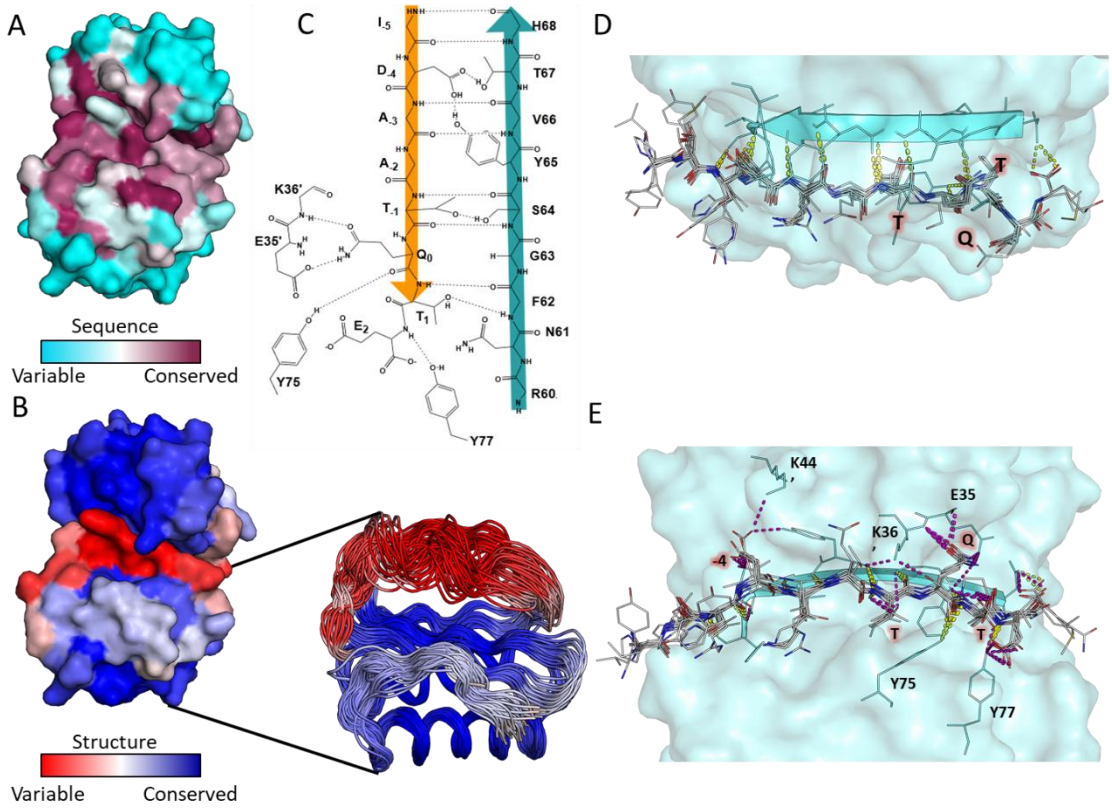


Figure 3.5: LC8 is conserved in sequence but structurally variable.
(A) Surface representation of LC8 colored by sequence conservation using ConSurf. More sequence conserved regions are shown in magenta, less sequence conserved regions are shown in cyan. Highly conserved residues map to those within the LC8 binding site. (B) Surface representation of LC8 colored by *structural* conservation in the free protein using the Ensemblator. Regions that are more structurally variable are shown in red, while more structurally conserved regions are shown in blue. An overlay of NMR and crystal structure protomers used for the structural analysis is shown as a cut-out in B. (C) 2D depiction of the binding interface between an example peptide (orange) and the binding β-strand within LC8 (Teal). Polar Bonds between LC8 and peptides from crystal structures are shown in D (top down view, only backbone interactions) and E (pocket view). Colors of polar contacts are based on whether the polar contacts stem from backbone (yellow) or side chain (purple) residues *on the peptide*. Peptide residues with frequent side chain interactions are labeled in red. Residues outside of the binding β-strand that are important interaction sites shown in C, are labeled in E.

Incorporation of physicochemical features and non-binder data improves binding predictions

Based on position preferences described above, we developed an LC8Pred algorithm that captures common features observed in binding peptides, including size and charge preferences, as well as features present in the 32 anchor-containing non-binding peptide sequences (Figure 3.6A). For each matrix, positive values within the matrix indicate that the given amino acid is enriched in binding sequences and depleted in non-binding sequences, while high negative values signify depletion of that amino acid in binding sequences, and enrichment in non-binding sequences. The addition of non-binding sequence information significantly improved the algorithm's capacity to differentiate between binding and non-binding sequences; however, with only 32 non-binding sequences, our data was notably sparse, and separation between the two groups was incomplete. To improve our differentiation capacity, we binned the 20 amino acids into four categories, and developed additional PSSMs utilizing these bins, thereby reducing the overall number of matrix terms. The first PSSM separated amino acids based on polarity, while the second separated according to volume.

These matrices largely confirm groupings as described in the 'common feature' section above, but with some exceptions. Notably, while there is a preference for large amino acids at the -3 position, the polarity matrix also shows an enrichment in positively charged residues. In addition, although the -5 position is the most varied in the matrix, it has a high score for positively charged residues (Figure 3.6A). This discrepancy is due to the lack of positively charged residues at -5 in the non-binding sequences, rather than from any strong enrichment of positive charge in the binding sequences. The -5 position also shows a slight enrichment for very large amino acids, and is the only position to do so. Crystal structures show that the -5 position is not buried within LC8's binding groove, and therefore experiences much less steric restriction (Figure 3.1B).

Using the described matrices, we scored all known binders and non-binders to determine the discriminatory capabilities of the PSSMs (Figure 3.6B). Although the amino acid, volume, and polarity matrices were each moderately successful at separating binding from non-binding sequences in isolation, the best separation was achieved when every matrix was combined. We combined the volume and polarity matrices to determine a volume and polarity score, and the amino acid matrix was used to determine an amino acid score (Figure 3.6B).

Since our goal is to predict partners with high reliability, strict thresholds were used to determine what constitutes a binder and a non-binder. A minimal score of 12.9 on the amino acid matrix, and 0.1 on the volume and polarity matrix is used to determine whether a sequence is

likely to be considered a binder. These thresholds result in only four false positives, and 20 false negatives with our available data-set, corresponding to a 75% true positive rate, and an 88% true negative rate (Figure 3.6B). Interestingly, although the volume and polarity matrices only provide a small increase in accuracy overall at these thresholds, they are extremely proficient at separating binders from non-binders when applied stringently. A threshold of 2.7 on the volume and polarity matrix alone results in a 0% false positive rate, while retaining 57% of the true positives (Figure 3.6B).

Although we achieve an accuracy of 78%, there are a number of outliers: both high-scoring non-binders, and low-scoring binders. Within the binders, the first sequence, DDKNTMTD, is from Myosin Va (Figure 3.6B). It is unsurprising that this sequence scores poorly, as it is the only 'TMT' anchor with verified binding data, and therefore has a low score due to the M instead of Q. On the other hand, binding is likely salvaged by the presence of the highly favourable amino acids at the other positions, and by the presence of adjacent coiled-coil domains in the full length protein. The remaining three lowest scores belong to proteins with multiple LC8-binding sequences proximal to one another (namely ASCIZ, ATMIN, and BSN), which would facilitate binding of weaker motifs due to bivalency. Within the nonbinders, three of the four well-scoring non-binders are listed in Figure 3.4D as "other," indicating that there is consistency between algorithm predictions, and our ability to recognize binders/non-binders based on sequence. This also suggests that there are some deleterious interactions that we have yet to understand, and will require more data to decipher. The fourth sequence contains a hydrophobic valine at the +2 position (Figure 3.6B, sequence 8), which is very rare, as this position is often fully solvent exposed and prefers β-strand breaking residues (Figure 3.1B). While LC8Pred weights valine at +2 negatively (Figure 3.6A), the remaining residues score well enough to result in the erroneous categorization of this sequence as a binder. Further accumulation of LC8 binding and non-binding sequences will no doubt help to clarify the importance of one poorly scoring residue, and improve LC8Pred accuracy. Our LC8 motif algorithm is available on the database web-page for public use (http://lc8hub.cgrb.oregonstate.edu/finder.html) for any sequence of interest.

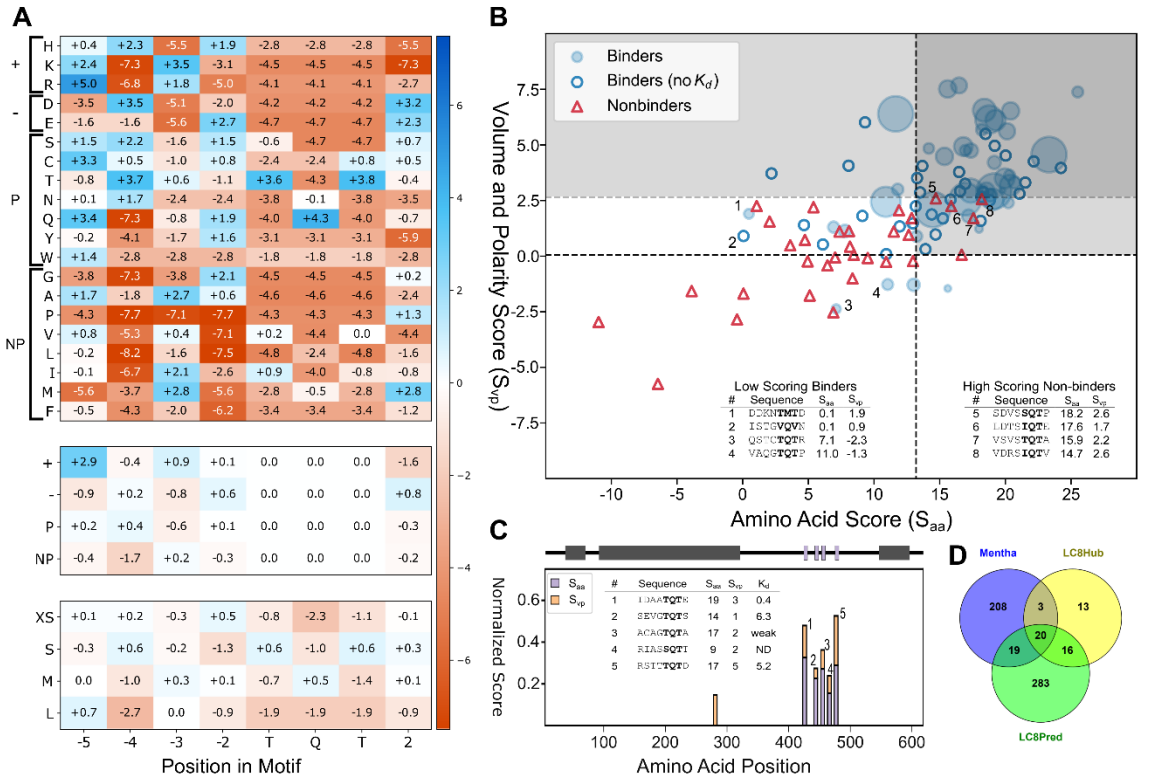


Figure 3.6: Generation and testing of the LC8Pred algorithm.
Position specific scoring matrices (PSSMs) for amino acids (A, top), bins by chemical property – positively charged, negatively charged, polar, or nonpolar (middle), and bins by volume – less than 106 A^3, 122 to 142 A^3, 155 to 171 A^3 and greater than 200 A^3 (bottom). Values correspond to the combined weight at a given position for the binder only matrix and the non-binder-normalized matrix. (B) Scatterplot of available sequences scored using a leave-one-out method of cross validation. For binders with a known K_d, the size of the bubble was varied inversely with the K_d, with binders with a K_d below 0.5 µM represented as the maximum possible dot size. Binder sequences with an unknown binding affinity were plotted as hollow circles, and non-binders as red triangles. The light grey box denotes predicted binding sequences using this scoring system. A second threshold for the volume and polarity axis indicates the very high confidence region, above which the specificity is unity. Outliers are noted in the tables, and numbered in the figure. (C) Normalized scores from matrices used to evaluate known LC8-binding protein Chica, where a score of one equates to the ideal amino acids of physicochemical properties at all positions. A sliding window to evaluate Chica for predicted binding sites across the protein was used, with the '0' position within the motif plotted (i.e. at 400, the 0 position is the 400th amino acid within Chica). A diagram of Chica showing secondary structure prediction (grey) and LC8 binding sites (purple) is above, and sequences predicted to be likely to bind at the right, along with their corresponding scores. (D) Venn diagram of human proteins in the LC8Hub database, proteins which contain at least one LC8-binding sequence as determined by LC8Pred, and proteins reported to bind LC8 in the protein-protein interaction database Mentha[189].

Predictive scores for the human protein Chica: a known LC8 binder

To test the ability of LC8Pred to identify binding sequences, we scored a test protein on each matrix using a sliding window. For this test we selected Chica, a protein which contains a series of LC8-binding sequences between residues 400 and 475[44]. To prevent algorithmic bias, peptides from Chica were not used in the development of our scoring matrix. Upon applying the LC8Pred algorithm, six positive scores were returned within Chica (Figure 3.6C). One of these scores fell far below threshold and was ignored. The remaining five were within the LC8 binding region; four of which have previously been determined experimentally to bind LC8[44]. The other is an SQT-containing sequence that scored below the designated threshold in the amino acid matrix, indicating that while this particular sequence may bind LC8, the prediction is of low confidence (Figure 3.6C). These test results provide strong evidence of the discriminatory power of our algorithm, as it can successfully recognize sequences that bind LC8 while excluding those that do not.

Human proteome scan identifies 374 potential binding sequences

After determining LC8Pred's reliability and ability to distinguish potential motifs, we used it to scan the human proteome to identify high-confidence binding partners. In total, 785 sequences scored above our PSSM thresholds. These sequences were then further filtered using IUpred to eliminate motifs within ordered regions. This process yielded 374 high-confidence hits from 338 proteins (SI Table 3.2 – See manusript[22]). Of these, 36 have been previously described in direct interaction studies and are listed on our LC8Hub database (Figure 3.6D). A further 19 partners have been identified in high throughput proteomics studies, such as pull-down mass spectrometry, including the highest scoring hit (FAM117B)[200,208]. Our data validate these interactions, and define likely binding regions within these partners. It is of note that several of the identified partners contain multiple putative LC8 sites in close succession. The ability of LC8 to "zip up" partners with multiple recognition motifs has been described for both Nup159[65] and ASCIZ[69], and it is possible that many partners within this list contain weaker LC8 sites proximal to these tight-binding motifs.

Prior studies on LC8 interactions have noted an enrichment in LC8 partners within the Hippo signalling pathway[75]. Our proteome scan has identified these same partners (*e.g.* AMOT, WWC1, WWC2), as well as additional novel binders from the hippo pathway, such as STK4 and DLG5. Interestingly, this pathway is the only 'biological process' significantly enriched in LC8 binding partners, based on gene ontology analysis using the WebGestalt program[209].

In order to verify that LC8Pred is correctly predicting partners, we synthesized three peptides from SI Table 3.2 (See manuscript[22]) and tested their capacity to interact with LC8 via ITC. The three peptides were derived from the human proteins: HTATSF1, a cofactor required for the Tat protein activation of human immunodeficiency virus transcription; OTOF, a calcium ion sensor involved in vesicle-plasma membrane fusion and neurotransmitter release, associated with hearing loss; and NIN, a component of the core centrosome and a dynein activator protein. These peptides were selected based on their mid-level scores and lack of prior data detailing LC8 interactions (SI Table 3.2 – see manuscript[22]). All three peptides bound to LC8 (SI Table 3.4 – see manuscript[22]), although only HTATSF1 was a 'strong' binder with a fittable thermogram (K_d of 10 μM). These data support the effectiveness of our LC8Pred algorithm, and demonstrate that it is capable of predicting binding partners of varying affinities despite non-canonical motifs (SI Table 3.3).

Discussion

Hub proteins are essential for cell viability as they are central in protein-protein interaction networks. Dynamic hubs such as LC8 often have a recognizable binding motif, which should allow for the prediction of binding partners without the need for exhaustive testing of each individual interaction[210]; however, no such program is available for LC8. Instead, binding partners are often identified via high-throughput pull-down experiments. For example, the interaction between LC8 and OFD1 was initially identified via pull-down mass spectrometry study in cilia[200]. In most cases, follow-up experiments for validation of direct binding are not performed, as it is prohibitively expensive to verify these interactions in a systematic fashion. Here, we validate purported and previously unreported LC8 binding partners (including OFD1), measure their binding affinities and thermodynamic properties, and establish a database of known LC8-partner interactions in order to define and describe generalizable requirements for LC8 motif recognition. We use these rules, along with amino acid preferences in non-binding sequences, to develop an algorithm that effectively distinguishes between binding and non-binding sequences, with the aim of facilitating *a priori* prediction and discovery of LC8-partner interactions with much greater confidence and accuracy than has been possible before now. Furthermore, we validate interactions that reinforce the importance of LC8 within a wide variety of systems, and demonstrate that LC8 is both localized ubiquitously throughout the cell, and enriched in distinct regions unrelated to the dynein complex.

Out of 72 synthesized tetradecameric peptides, we verified binding for 29 peptides derived from 27 distinct proteins (SI Table 3.3). Of these 27 proteins, 19 are newly identified LC8 binding partners. It is of note that all of our validated sequences contain the canonical TQT anchor (or variation thereof) at the C-terminus of the peptide, supporting the idea that a C-terminal anchor is vital for LC8 binding. Although the LC8 binding site is structurally dynamic, there are distinct preferences and exclusions for each position within the binding motif (Figure 3.4). In addition to the presence of an anchor, binders often have -4 positions capable of H-bonding, larger positive side chains at -3 positions, and strand breaking +2 positions. On the other hand, the presence of pre-anchor prolines, a high concentration of charges, or bulky hydrophobic groups at the -2 position will each limit the likelihood that a sequence will bind LC8 (Figure 3.4).

Algorithms for motif identifications have been developed for both 14-3-3 and calmodulin to efficiently predict potential binding partners. In the case of calmodulin, its diverse set of binding motifs has led to multiple programs[211–213], which predict potential binding partners via a mixture of sequence similarity to known binders, α-helical propensity, or the number of canonical calmodulin binding motifs within a given sequence. In the case of 14-3-3, which binds phosphorylated sequences within disordered segments of proteins, the algorithm makes use of support vector machines and artificial neural networks[210] and scores potential binding sequences using a PSSM. Here we succeeded in generating LC8Pred, an algorithm with a 78% accuracy rate, by incorporating non-binder data and by reducing the PSSM dimensionality from 20 amino acids to 4 physiochemical categories, based on either polarity or volume. We have tested LC8Pred on the known LC8 binder Chica, as well as by scanning the human proteome. In the case of Chica, LC8Pred efficiently recognized known binding sites and excluded all other regions (Figure 3.6C). Our proteome scan identified 338 potential LC8 binding partners, including 19 which have been identified previously via high-throughput proteomics studies (Figure 3.6D, SI Table 3.2 – see manuscript[22]), providing a new set of high-confidence LC8-interacting proteins. Three peptides were selected from these potential partners and shown to indeed bind LC8.

The ability to bind a wide variety of sequences despite an extremely conserved binding interface is a hallmark of dynamic hubs, as exemplified by calmodulin[214] and 14-3-3 proteins[215]. Crystal and NMR structures for LC8 show that the β3 strand at the partner binding interface has the highest sequence conservation (Figure 3.5A), and surprisingly, is also the most dynamic region (Figure 3.5B). Consistent with the dynamic nature of the binding grooves, thermodynamic analyses of tight binding sequences demonstrate a wide range of entropy/enthalpy compensation, including some sequences that bind with a favorable change in entropy, such as ICE1 and VP4.

Previous studies on LC8 dynamics of binding to dynein intermediate chain (IC) and the protein swallow (Swa) show that increases in ordered structure upon binding are peptide-dependent[192]. With Swa, the complex is more compact, rigid and homogeneous than with IC, indicating that the IC peptide retains more freedom of motion in the bound state than does the Swa peptide. Consistent with these observations, IC binds with a favorable entropy, while Swa does not. Our work here demonstrates that these different modes of binding are not limited to IC and Swa, but rather that entropic factors commonly modulate LC8 binding to accommodate extraordinary variation in binding sequences.

Hub proteins like LC8 are essential for cell homeostasis as they sit at the center of complex interaction networks; therefore, it is imperative to understand the rules that govern hub protein interactions. The dynamic nature of the LC8 pocket, and entropic contributions to binding, make it difficult to predict partners with high confidence, and yet it is this very dynamic characteristic that makes LC8 such a powerfully effective hub protein. Here we have amalgamated our experimentally verified LC8-binding sequences with all previously described binding sequences, and developed an algorithm that significantly advances our ability to predict LC8 partners based solely on sequence. Confidence in a potential LC8-binding sequence can be further improved by considering the structure and conservation of the binding site, and we have therefore linked LC8Pred to ProViz, a tool that analyzes protein structure and conservation. Additionally, it is important to note that LC8Pred is optimized for stringency and predicting tight binding interactions, and does not account for adjacent oligomerization sites, which would increase binding affinities. Future versions of the algorithm will incorporate parameters to account for other factors impacting binding, such as oligomerization state or subcellular localizations. We also anticipate that the predictive power of our algorithm will improve dramatically as more LC8 binding and non-binding sequences are identified and deposited in the LC8hub database, resulting in a comprehensive view of the LC8 hub interaction network.

Materials and Methods

Cell culture and transfection

HeLa Kyoto cells stably expressing LC8-GFP were a kind gift from I. Poser and A. Hyman (Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany)[151]. The cells were cultured in DMEM medium with 10% (v/v) fetal calf serum and with 1% (v/v) Penicillin/Streptomycin. The cell line was routinely checked for mycoplasma contamination

using LT07-518 Mycoalert assay (Lonza, Switzerland). The identity of the cell line was monitored by immunofluorescence-staining-based analysis with multiple markers.

Microscopy and image analysis

Live and fixed samples were imaged with spinning disk microscopy, which was performed on an inverted research microscope Eclipse Ti-E with the Perfect Focus System (Nikon), equipped with Nikon Plan Apo VC 100x N.A. 1.40 oil objective, Yokogawa CSU-X1-A1 spinning disc, Vortran Stradus 405 nm (100 mW), Cobolt Calypso 491 nm (100 mW) and Cobolt Jive 561 nm (100 mW) lasers, Chroma emission filters ET460/50m (part of 49021 filter set), ET525/50m (part of 49002 filter set) and ET630/75m (part of 49008 filter set), ASI motorized stage MS-2000-XYZ with Piezo Top Plate (ASI), Photometrics Evolve 512 EMCCD camera (Photometrics), and controlled by MetaMorph 7.7 software (Molecular Devices). Images were projected onto the camera chip with intermediate lens 2.0x (Edmund Optics) at a magnification of 0.067 mm/pixel. To keep cells at 37°C we used the stage top incubator INUBG2E-ZILCS (Tokai Hit). Z-series of live and fixed samples were acquired using a 0.1-µm-step confocal-based scan. Side views were reconstructed by projecting maximum fluorescence intensities of 24x12-µm side view slices.

Alternatively, fixed samples were imaged using wide field fluorescence illumination on a Nikon Ni upright microscope equipped with DS-Qi2 camera (Nikon), Intensilight C-HGFI illuminator (Nikon), ET-DAPI, ET-EGFP and ET-mCherry filters (Chroma), Nikon NIS Br software, and a Plan Apo Lambda 100X oil NA 1.45 (Nikon) objective. For presentation, images were adjusted for brightness and contrast using ImageJ 1.47v (NIH).

Localization prediction

Localization information is derived from the COMPARTMENTS program[195], using the curated "Knowledge-based" evidence category. The list of LC8 binding proteins used matches the curated list on the LC8 database, described in this paper. Only data with confidence scores of three or higher (out of five) are included. Cellular compartments are simplified for depiction purposes (e.g., "Other Organelles" includes Golgi bodies, mitochondria, and so on).

Proteomic peptide phage display (ProP-PD) selections

Phage display selections were performed using a proteomic library designed from the disordered regions of the human proteome described in Davey *et al.*[196]. Selections were

performed with minor adjustments. GST-LC8 (0.1 mg/mL in 100 µL TBS, 50 mM Tris-HCl, 150 mM NaCl, pH 7.4) was coated on a Maxisorp 96-well plate (Nunc) via overnight shake-incubation at 4 ºC. Plates were blocked with 0.5% BSA in TBS for 1 h at 4 ºC, and washed with TBS. The phage library was added to the well (100 µl), and incubated for 2 h at 4 ºC. Unbound phages were removed by washing plates 5 times with 300 µL TBS + 0.05% Tween. Bound phages were eluted by infection into 100 µL log-phase *E. coli* Omnimax cells (Invitrogen; OD: 0.3-0.8) in 2xYT media (10 g bacto-yeast extract, 16 g bacto-tryptone, 5 g NaCl per liter) supplemented with 10 µg/ml tetracyclin. After a 30 min shake-incubation at 37 ºC, the bacteria were hyperinfected with M13K07 helper phages for 45 min to allow phage production. Cultures were transferred into 5 mL 2xYT, 0.3 mM IPTG, and grown overnight with antibiotics (25 µg/ml kanamycin and 100 µg/ml carbenicillin). The bacteria were pelleted by centrifugation. One mL of the phage supernatant was extracted and heat inactivated at 65 ºC for 20 min. Finally, the solution was pH neutralized using 10x TBS, and the phage pool was used in the next round of selection. Five rounds of phage selections were performed in total. The phage pool from the fourth day of selection was used for clonal phage ELISAs and sequencing. For next-generation sequencing, 5 µl of the phage pool from the fourth day of selection was used as template in a barcoding PCR reaction. The sample was prepared and analyzed as described in detail elsewhere[202].

Peptide synthesis

A total of 72 putative binding partners identified from ProP-PD selections and algorithm predictions were commercially synthesized from either Genscript (Piscataway, NJ), or Synpeptide (Shanghai, China), as 14-16 amino acid sequences. Non-native residues were added to the termini of some peptides to facilitate solubility and peptide concentration determination (Tables 3.1 and 3.2, italics). All peptides were derived from either human or viral proteins.

Isothermal titration calorimetry

ITC experiments for the interactions of LC8 with peptides were performed using a Microcal (North Hampton, MA) VP-ITC microcalorimeter at 25 °C in buffer composed of 50 mM sodium phosphate, 50 mM NaCl, 1 mM sodium azide, 5 mM β-mercaptoethanol, pH 7.5. Some peptides contained cysteine residues, so 5 mM β-mercaptoethanol was included in all solutions for consistency. In all experiments, an initial 2 µl injection was followed by 26-50 injections of 3-10 µl peptide (500 µM) into 25 µM LC8 in the sample cell. Number and volume of injections were adjusted for each experiment to minimize ambiguity in the shape behaviour of

isotherms and thermograms. Peptide concentrations were determined from absorbances at 280 nm using molar extinction coefficient values computed with the Protparam tool on the ExPASy website[150]. Peptides lacking aromatic residues were weighed and resuspended in the proper volumes to ensure 500 μM final concentrations. Protein samples and buffer were degassed before data collection. Data were processed using Origin 7.0 (Microcal) and fit to a single-site binding model. Final values for binding parameters are averages of 2-3 independent experiments.

LC8Pred algorithm generation

The LC8Pred algorithm was developed using 79 LC8 binding sequences and 32 anchor-containing non-binding sequences (SI Table 3.4 – see manuscript[22]). We selected sequences that bind LC8 with high-confidence, on which direct interaction data are available. In addition, all sequences with a K_d above 25 μM were not included. The TQT (or variation thereof) anchor-containing non-binders were those peptides shown by ITC to have no binding to LC8.

In addition, a new series of matrices were developed which binned amino acids into categories based on physicochemical properties. Specifically, a matrix that separates amino acids into positively charged, negatively charged, hydrophobic, or polar and uncharged, and a matrix that separates amino acids into four groups based on volume, with volume bins being selected to minimize the range of volumes within each bin. We built these matrices to overcome the limitation of our small dataset, as reducing the number of groups from 20 amino acids to 4 possible properties improves the likelihood that some information is available for a given position and a given property within the motif.

In total, 6 matrices were developed, two for each set of bins (amino acid, polarity, volume). For a given bin, one matrix was normalized to the background frequency of a given amino acid or a given property within the disordered eukaryotic proteome taken from the DisProt database of intrinsically disordered regions[216]. For the other matrix, normalization was done for the frequency of a given amino acid or property in the non-binder dataset. As non-binding sequences were selected based on the presence of an anchor, there is no enrichment or depletion at the anchor positions of -1 to +1. These positions were therefore ignored in these matrices.

To simplify our scoring system, we combined the matrices into two simple scoring metrics, S_{aa} and S_{vp}, where S_{aa} is a combination of the two matrices that use amino acid type bins, and S_{vp} is a combination of the four matrices that use volume or polarity bins. To determine how effective each individual matrix was at separating binding and non-binding sequences, we scored our available sequences using leave-one-out cross validation, where a given sequence was

excluded from the matrix, and then scored. The leave-one-out approach was used to combat the difficulty of our limited dataset.

We used receiver operating characteristic (ROC) curves (SI Figure 3.1) as a metric of the effectiveness of each score. The area under these curves corresponds to the ability of each matrix to separate binding sequences from non-binding sequences. We then combined scores into the S_{aa} and S_{vp} scores described above, where each individual matrix score was weighted through a grid-search of possible weights, where the largest area under the ROC was taken to be the optimal weight for each score. Surprisingly, the area under the ROC curve was highest when the binder-only polarity matrix was removed from the S_{vp} Score. Positions -1, 0 and 1 are therefore not weighted in the polarity matrix (Figure 3.6A), because the non-binder normalized matrix was also excluded at those positions due to a lack of anchor enrichment, as discussed above.

Improved Algorithm Generation

We aligned 79 experimentally confirmed LC8 binding sequences (Table S4). Only 'strong' binding sequences (K_d < 25 μM) were selected so that the final algorithm provided only high-confidence hits, rather than being exceptionally permissive; however, this means that many weak binding sequences will be overlooked. Aligned motifs were defined as an eight amino-acid peptide sequence p, stretching from the -5 to the +2 positions relative to the anchor Q residue, thereby defining the "0" position of each binding sequence as the Q in the canonical TQT motif. We used these binding sequences to generate a PSSM with weights, W_{ai}^{B}, for a given amino acid, a, at position, i, such that $p[i] = a$ within each instance of the motif. These weights (eq. 1) define a log-likelihood score comparing the frequency of the amino acid at that position in the binders, B_{ai}, to the background frequency of amino acid, P_a, across disordered eukaryotic proteins found in DisProt, a database of intrinsically disordered proteins and regions [216]. A positive weight at a given position and amino acid indicates that that amino acid is enriched at that position, relative to a random occurrence rate. Likewise, a negative weight indicates that the amino acid is depleted at that residue.

$$W_{ai}^{B} = \log_2\left(\frac{B_{ai}}{P_a}\right) \qquad \text{eq. 1}$$

B_{ai} is taken as the number of counts C_{ai} plus a pseudo-count D, over the total number of counts at position i. Pseudo-counting was primarily used to fill gaps in the matrix, and a value of 5 was selected for D based on prior work in the field (eq. 2)[75].

$$B_{ai} = \frac{C_{ai} + \frac{D}{20}}{D + \sum_{a=1}^{20} C_{ai}} \quad \text{eq. 2}$$

The frequency matrix B_{ai} is then used to compute W_{ai}^{B} as in eq. 1. In addition to a matrix informed by known binding sequences, we assembled a matrix normalized to nonbinding sequences, each determined via ITC to not bind to LC8. We aligned 32 anchor-containing sequences in the same manner as above and used them to generate a PSSM with weights W_{ai}^{N}, incorporating both binding and nonbinding sequences. For this matrix, instead of the background frequency of a given amino acid, the frequency within the binders was normalized to the frequency N_{ai} of the same amino acid in the non-binders (eq. 3).

$$W_{ai}^{N} = \log_2\left(\frac{B_{ai}}{N_{ai}}\right) \quad \text{eq. 3}$$

N_{ai} was assembled in the same manner as B_{ai}, using non-binding sequences, and the resultant PSSM provides log-scale scores that reflect relative enrichments for each amino acid at a given position in the known binding sequences, versus in the known non-binding sequences. A positive W_{ai}^{N} value corresponds to an amino acid at a given position being relatively enriched in the binding sequences, while a negative value corresponds to it being enriched in non-binding sequences.

To further improve our ability to differentiate binders and non-binders, we built additional PSSMs that binned amino acids into four categories. First, PSSMs were assembled based on properties of charge and polarity, with a bin for positively charged (H,K,R), negatively charged (D,E), polar (C,N,Q,S,T,W,Y), and hydrophobic amino acids (A,F,G,I,L,M,P,V). In addition, another set of matrices was assembled based on amino acid volume, also with four bins: very small: less than 106 Å^3 (A,C,G,S); small: 122 to 142 Å^3 (D,N,P,T,V); medium: 155 to 171 Å^3 (E,H,I,K,L,M,Q); and large: greater than 200 Å^3 (F,R,W,Y;[217]. Groupings for amino acid volume were chosen to minimize the range of volumes within a given group.

In this manner, a total of 6 PSSMs were constructed, two each corresponding to amino acid bins, polarity/charge based bins, and volume based bins. For a given input sequence, each matrix returns a score, equivalent to the sum of the weights at the relevant amino acids and position (eq. 4). For clarity, each matrix was given a subscript, S_B for the matrices built using only binding sequences, and S_N for matrices using both binder and non-binder data. The polarity

and charge matrices were indicated with a P ($S_{PB},\ S_{PN}$), and the volume matrices were indicated with a V (S_{VB}, S_{VN})

$$S(p) = \sum_{i=-5}^{2} W_{P[i],i} \qquad \text{eq. 4}$$

To simplify the output of our scoring system, we simplified our output to two scores. The first metric used the first two matrices described, which scored input peptides on their amino acid sequence (eq. 6). The resultant score, $S_{aa}(p)$, is the sum of the scores for both the binder-only matrix W_{ai}^{B} and the non-binder normalized matrix W_{ai}^{N} across all positions i in the motif. Both scores are modified by weight factors (α_B, α_N, respectively) that were optimized using a ROC curve computed using leave-one-out cross validation, whereby a separate set of weights $W_{p[i],i}^{B}, W_{p[i],i}^{N}$ was calculated at the exclusion of each peptide.

$$S_{aa}(p) = \sum_{i=-5}^{2} (\alpha_B W_{p[i],i}^{B} + \alpha_N W_{p[i],i}^{N}) \qquad \text{eq. 5}$$

The second metric scored sequences based on their volume and polarity properties, using values from our four additional PSSMS score (eq. 7). The score $S_{vp}(p)$ combines scores from a background frequency normalized matrix with polarity bins W_{ai}^{PB}, a non-binder normalized matrix with polarity bins W_{ai}^{PN}, as well as equivalent matrices using volume bins W_{ai}^{VB}, W_{ai}^{VN}. These were also modified by weight factors, α_{PB}, α_{PN}, α_{VB}, α_{VN}.

$$S_{VP}(p) = \sum_{i=-5}^{2} (\alpha_{PB} W_{p[i],i}^{PB} + \alpha_{PN} W_{p[i],i}^{PN} + \alpha_{VB} W_{p[i],i}^{VB} + \alpha_{VN} W_{p[i],i}^{VN}) \qquad \text{eq.6}$$

To test our approach, we scored our training sequences using leave-one-out cross validation, where one sequence was omitted from the training set used to generate each scoring matrix. The omitted sequence was then scored on the PSSMs generated, with this process repeated for every peptide used. Scoring was evaluated via an ROC curve. The area under the ROC curve (AUROC) corresponds with the score's ability to separate true positives – sequences which do bind LC8 and which also score as binders – from false positives – sequences that do not bind, but still return a positive score. The larger the AUROC, the greater the ability of the score to differentiate binding from non-binding sequences (SI Figure 3.1). Using the AUROC as a metric for PSSM effectiveness, we did a grid search of possible weight factor values for equations 5 and 6. The arrived at α values are as follows: α_B = 1, $\alpha_N = 0.78$, $\alpha_{PB} = 0, \alpha_{PN} = 0.6$, α_{VB} = 0.37, α_{VN} =

0.22. Notably, the settled on α for the polarity matrix that used only binding sequences was optimized to 0, indicating that the binder-only polarity matrix did not improve the ability of S_{VP} to separate binding and non-binding sequences.

Additionally, we used the AUROC to determine that the non-binder normalized scores W_{ai}^{N}, W_{ai}^{PN}, W_{ai}^{VN} had a negative impact on the effectiveness of our scoring when included at the -1, 0 and +1 position in the matrices. As the non-binding sequences were selected for the presence of a TQT anchor, there is effectively no relative enrichment for any amino acid or physiochemical property at those positions in the associated PSSMs, and we therefore excluded those scores at the positions in question.

The LC8 Motif Repository

We have manually curated a database that compiles information for all known LC8 binding partners. Including the 19 binding partners identified in this work, there are currently 80 experimentally confirmed LC8 interacting partners containing 116 individual anchor motifs. Of these binding motifs, 98 have been confirmed by *in vivo* or *in vitro* experiments, with a further 18 identified through biochemical screening methods. The database serves to (1) provide a source of up-to-date information on LC8 and its cellular role, (2) organize and classify LC8 binding proteins in an easily searchable manner, and (3) list the sequences of all TQT motifs to aid in identification of new binding partners. Access to the motif repository is available at http://LC8hub.cgrb.oregonstate.edu. For each protein, the following information is provided: the species, TQT peptide sequence, number of motifs in the protein, Protein Data Bank (PDB) ID (if a structure exists), reference link, and interaction type. The interaction type has three levels of classification, depending on the method by which the LC8-partner interaction was identified: (1) High throughput biochemical method, such as yeast-2-hybrid, where the interaction has not been confirmed by *in vivo* or *in vitro* experiments; (2) *in vivo* experiments, such as mutation or knockout experiments, where a function for the LC8-partner complex has been identified; and (3) *in vitro* experiments that determine the binding affinity, structure, or other information about the LC8-partner interaction. Additionally, sequences of interest can be tested at LC8Hub by inputting a .fasta file or a string of letters corresponding to the protein sequence of interest. Output provides both the S^{aa} and the S^{vp} scores, as well as indicating sequences that are likely to bind LC8 according to available data. Lastly, sequences determined to either bind or not bind LC8 despite the presence of an anchor sequence can be submitted for incorporation into the database. It is our

hope that the information in this database will facilitate research on LC8 and, by enhancing our understanding of the TQT motif, enable more robust prediction of new binding partners.

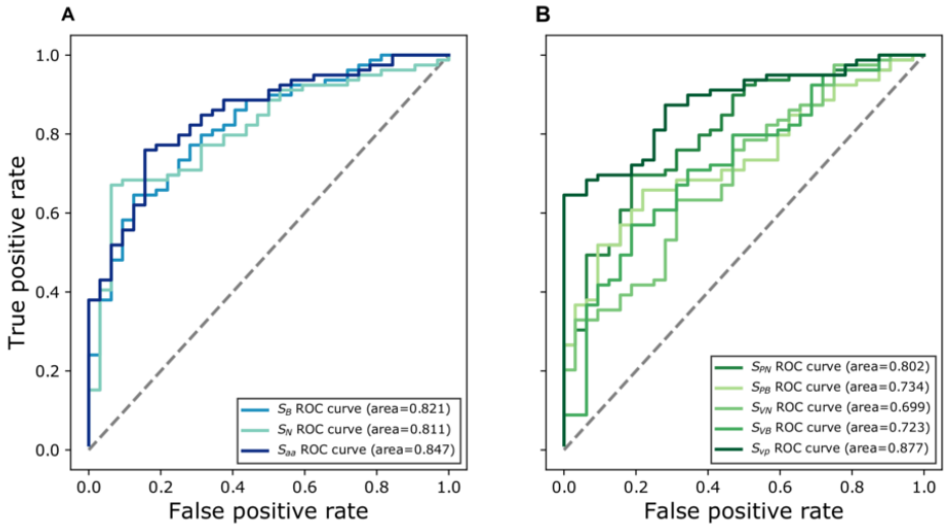
Structure and motif analysis

Structures of LC8 were obtained from the Protein Data Bank (free LC8 PDB codes: 1PWJ, 1PWK, 1RE6, 3BRI, 5WOF; bound to peptides: 2XQQ, 4QH7, 3E2B, 2P2T, 3BRL, 3DVP, 3P8M, 3ZKE, 4D07, 4HT6, 5E0M). All images were generated using PyMol[218]. Peptides without structures available were built *in silico* using Chimera[219]. Peptides in Figure 3.4 are colored according to enrichment and depletion tables for amino acids, shown in Figure 3.6A (blue for scores > 1, white for scores between 1 and -1, and red for scores < 1). Solvent accessible surface area analysis was performed using a representative LC8 crystal structure (2XQQ) with the GETAREA program [220]. Protein charge potential was calculated for LC8 using PyMol's built-in charge-smoothed potential calculator. Two-dimensional lig-plots were generated using ChemDraw.

Alignment of LC8 structures was done using the Ensemblator[137] program, and the root mean squared deviation (RMSD) for residues in free LC8 structures (listed above) was calculated using the built-in local alignment tool. This tool works by aligning each dipeptide within the protein and calculating the RMSD for the next amino acid within the protein sequence. A representative structure was then colored based on these values to demonstrate structural conservation. Sequence-based conservation was performed using ConSurf[187], with LC8 sequences from 58 different eukaryotic species.

Acknowledgements

This work was supported by National Institutes of Health Grant GM 084276 to EB, by the Swedish Research Council (C0509201) to YI, by Lennanders foundation and Ingegerd Bergh's foundation to CB. The NGS was performed in collaboration with Eduard Resch, Fraunhofer Institute for Molecular Biology and Applied Ecology IME TMP, Frankfurt am Main, Germany. Y.-C. A. is supported by the MARIE SKŁODOWSKA-CURIE ACTIONS Innovative Training Network (ITN) 675407 PolarNet.



SI Figure 3.1: Optimization of matrix weights.
(A) ROC curves of both amino acid matrices and S_{aa}. The larger the area under the curve (AUROC), the more effective the curve is at separating binder sequences from non-binder sequences. (B) ROC curves of each volume and polarity matrix, and S_{vp}. Notably, the S_{vp} curve performed substantially better than each volume and polarity matrix individually, which suggests that volume and polarity are both essential to understanding the preferences within the LC8 motif.

Uniprot	Gene	Peptide
Q8WWN8	ARAP3	TQTPGFPTQPPATSSP
O43521	BCL2L11	DKSTQTPSPPAQAFNH
Q96G01	BICD1	KLNGDYRTPTLRKGES
Q92904	DAZL	KRVHHFRRSRAMLKSV
Q92499	DDX1	TQTKFLPNAPKALIVE
Q96SC8	DMRTA2	LSSRSAFSPLQPNASH
Q7Z589	EMSY	TQTTNTTTQKVIIVTT
Q9Y5Z7	HCFC2	VIPETSVSSTVSSTQT
Q9Y2F5	ICE1	ELRHIGTQISSDSYGS
P35568	IRS1	LPRKVDTAAQTNSRLA
O43526	KCNQ2	VPMYSSQTQTYGASRL
Q9P266	KIAA1462	TQTSFSEEPQSSQLLP
P49862	KLK7	STQTHVNDLMLVKLNS
Q8NEZ4	KMT2C	SVSTQTASDDQAGKLW
Q2TV78	MST1L	TANTTTAAYLASVGTR
Q02505	MUC3A	LTSATGTQTSPAPTTV
Q9HC84	MUC5B	LTSTATKSTATSVTPI
Q9HC84	MUC5B	SSSTQTSGTPPSLTTT
Q6ZRI0	OTOG	LTASQLPAGPTESPAS
Q8NDX5	PHC3	SSVSTTTSSTTTTTIT
P14859	POU2F1	TGTQTNGLDFQKQPVP
Q9ULJ8	PPP1R9A	TDGSVVKLESSVSERI

Q01973	ROR1	SGGNATTQTTSLSASP
Q9H4B6	SAV1	FLRTPIQRTPHEIMRR
Q96FV2	SCRN2	LPRFQTQVDRRHTLYR
Q8TEC5	SH3RF2	TLVSTASGTQTVFPSK
Q96R06	SPAG5	STQTDTSHSGITNKLQ
Q96JG9	ZNF469	AAALPEETRSSRRRRL
Q99814	EPAS1	STQTDFNELDLETLAP
Q86VQ1	GLCCI1	TRSIDTQTPSVQERSS
P98088	MUC5AC	VTSSITSTTQTSTTSA
Q2KHR3	QSER1	SQTVTPENQTLNYSSN
Q5VU36	SPATA31A5	GTVPQSLSPHEDLVAS
Q9H2K8	TAOK3	PSMSVSTGSQSSSVNS
Q6ZU65	UBN2	SQTNPVVKLSNNPQLS
Q9Y4F4	FAM179B	QQTFGSQTEATSSNGQ
Q13618	CUL3	SGRQLTLQHHMGSADL
Q99102	MUC4	STQTTRESQTSTLTHR
Q9P2G1	ANKIB1	SSQTPQTSSDWLEQVH
P07359	GP1BA	TATQTTHLELQRGRQV
Q9HC84	MUC5B	LTSTATKSTATSFTPI
Q02817	MUC2	GTQTPTTTPITTTTTV
Q9BWV3	CIZ1	VSTQTGSMTGQIPRLS
Q8NBH2	KY	TSYNSQGTQLTVEVHP
O75665	OFD1	SANMETQTSSTFNRDS
Q7Z2Z2	EFTUD1	TGSQTFDSFPPETQAL
A0FGR9	ESYT3	VSRSTTTTTSATTVAT
Q13952	NFYC	ITQTEVQQGQQQFSQF
Q9Y4K1	AIM1	SFVLPVESTQDVSSQV
Q6IMN6	CAPRIN2	STQTPPQCQLPSIHVE
P53350	PLK1	LQTDPTARPTINELLN
Q9UBY0	SL9A	GTQTSGLLQQPLLSKD
P48436	SOX9	SQYDYTDHQNSSSYYS

SI Table 3.1: Binding enriched peptides from proteomic peptide phage display.
[a] Peptides identified from next generation sequencing of binding enriched phage pools as well as peptides identified by ELISA experiments are shown. Phage display hits validated via ITC are highlighted in grey. Motif anchors are underlined.

UniProt	Gene	8AA Sequence
Q6LCS3	E4 (HPV)	EDKQTQTP
O43521	BCL2L11*	CDKSTQTP
O43719	HTATSF1	KDGDTQTD
O75665	OFD1*	CNMETQTS
P07359	GP1BA	LTTATQTT
P11193	VP4 (Rotavirus A)	NDISTQTS
P13500	CCL2	LDKQTQTP
P16289	L (RAV)*[a]	ISRMTQTP
P18583	SON	RCVSVQTD
P80098	CCL7	LDKKTQTP
Q01973	ROR1	GNATTQTT
Q02505	MUC3A	SATGTQTS
Q05127	VP35 (EBV)*[a]	RNSQTQTD
Q6IMN6	CAPRIN2	TTASTQTP
Q86VQ1	GLCCI1	RSIDTQTP
Q8IX07	ZFPM1	SDKGVQTP
Q8N4C6	NIN	RTSETNTP
Q8WWN8	ARAP3	TGLPTQTP
Q96R06	SPAG5*	QDSSTQTD
Q99102	MUC4	WTRSTQTT
Q9HC10	OTOF	SSTEVQVE
Q9UBY0	SLC9A2	REKGTQTS
Q9ULV3	CIZ1*	RSVSTQTG
Q9UPA5	BSN^{II}*	AEFSTQTP
Q9UPA5	BSN*[a]	ANYGSQTE
Q9UPA5	BSN^{III}*[a]	VAQGTQTP
Q9Y2F5	ICE1	RHIGTQIS
Q9Y2H9	MAST1*	RHQSVQTE
Q9Y4F4	TOGARAM1	QTFGSQTE

SI Table 3.3: LC8 binding sequences identified and verified in this study.
[a] Peptides synthesized to validate previous data, not based on predictions. Motif anchors are underlined. * denotes sequences previously identified.

Chapter 4

Linear Motif-Binding Hub Interaction Networks: The Multifunctional Role of LC8 as a Structural LMB-hub.

Nathan Jespersen and Elisar Barbar

In preparation for submission to *Trends in Biological Sciences*, September 2019

Introduction

An organism's interactome describes the entire set of protein-protein interactions that occur within a cell. For example, in yeast 6,000 proteins participate in approximately 16,000-40,000 binary interactions[221,222]. These interactions are not evenly distributed, and are best described by a power law, where-in most proteins interact with a few partners, whilst some sit at the center of complex interaction networks[223]. The centrality of a protein arises from both its essential function and its pleiotropy. Knock-outs of central proteins, termed 'hubs', are either lethal, or lead to multiple unrelated deleterious effects[224].

Hubs can be static ("party" hubs) or dynamic ("date" hubs). Static hubs bind many partners simultaneously at different sites, for example BRCA2[77]. Dynamic hubs typically bind to linear motifs within intrinsically disordered regions (IDRs), and have multiple partners that compete for the same site[78,79]. While dynamic/date hubs are the most commonly used terms to define this class of proteins, we suggest using the more descriptive and specific term 'linear motif-binding hubs' (LMB-hubs)[225], due to their propensity to interact with short linear motifs at a single, highly conserved binding interface[226]. Well-known examples of LMB-hubs include calmodulin, 14-3-3, and SH3 domains[80–82,227]. A more recently discovered and characterized member of this class is the dynein light chain LC8[22,32].

The more than 100 experimentally verified binary LC8-partner protein interactions[22] include a large number of intrinsically disordered proteins (IDPs) regulating multifarious functions, including intracellular transport (dynein intermediate chain)[25,63,64], nuclear pore formation (Nup159)[65], and transcription (ASCIZ)[67–69]. Additionally, LC8 is highly expressed across a wide variety of cell types[70], and LC8 partners are broadly distributed within individual cells[22,71]. In addition to broad expression patterns and diverse functions, steadily increasing evidence indicates the importance of LC8 for a variety of viral infection cycles[66,88,89,119].

As obligate intracellular parasites, viruses are incapable of producing all of the components necessary for their own replication; their success is based on an ability to infect and utilize host machinery. In the most limited circumstance, this entails initial implantation and endocytosis of the virus, appropriation of replicative molecules such as ribosomes or polymerases, silencing or avoidance of host immune responses, the formation and packaging of nascent viruses, and dissemination of the new viruses to additional cells or hosts. Each step requires a multitude of interactions with host proteins and pathways, and incredibly, often proceeds efficiently despite the severe discrepancy in genome and proteome sizes between the

virus and its host. While the intricacies of every host-virus pair are different, the importance of linear motif binding-hubs (LMB-hubs) in this process is rapidly becoming apparent. Here we outline some of the defining features of LMB-hubs using LC8 as our model, and describe how viruses take advantage of these features to efficiently hijack their hosts.

Key Features of LMB-Hubs

LMB-hubs have dynamic binding interfaces in otherwise rigid structures

A defining feature of LMB-hubs is that one specific binding site accommodates many diverse partners. This requires either a very open, nonspecific interface, or a highly plastic binding groove. Prime examples are major histocompatibility complexes I and II (MHCs), a set of proteins integral to the adaptive immune system. They function by extracellularly presenting antigens from potential pathogens for recognition by the appropriate T-cells. This involves cutting pathogen proteins into short peptides, and loading them onto a variety of different MHC isoforms. Although MHCs are polygenic and polymorphic[228], it is infeasible for such a small number of proteins to bind and present the wide variety of peptides derived from foreign proteins without considerable promiscuity in their recognized linear motifs[229]. Studies on the cornucopia of MHC crystal structures have noted that B-factors, which can act as a proxy for dynamics, are surprisingly high within the MHC binding groove[230]. This suggests that MHC-binding of diverse peptides is facilitated by a plastic binding interface. Indeed, analysis of the stability of the interface before and after binding indicates that free MHC molecules are significantly more dynamic, and have structures stabilized by bound peptides[231]. Although LMB-hub plasticity is perhaps best characterized in MHCs, this general feature is observed in many LMB-hubs[232–234].

LC8 interacts with an eight amino-acid recognition motif within intrinsically disordered regions of its partners. While there is substantial variation in this recognition motif, it is most frequently anchored by a TQT sequence (termed the "motif anchor")[44]. A thorough analysis of the motif preferences of LC8 can be found in Jespersen *et al.*[22]. As in the case of MHCs, an abundance of structural data facilitates analysis of the structural heterogeneity in free and bound LC8 (Figure 4.1). We used the Ensemblator[137] to align independently determined 3D structures and identify regions of structural conservation or plasticity. By overlaying the protomers from five published crystal and NMR structures of free LC8[20,191,192], we observed that the β-strand that directly binds to partners is highly variable (Figure 4.1), and has the highest RMSD value between structures. On the other hand, an overlay of 17 structures of LC8 bound to peptides

demonstrates that partners greatly stabilize and regularize the binding interface (Figure 4.1). Interestingly, despite the plasticity of the binding groove, a comparison of regions of conservation within LC8 reveals that the general features of the groove are strictly conserved across organisms as evolutionarily distant as *Giardia lamblia* and *Homo sapiens* (Figure 4.1)[22]. Residues near the dimeric interface/partner binding site are noticeably more conserved than peripheral regions. In fact, due to hub protein centrality and essentiality, these binding sites are often very highly conserved[235]. This leads to an intriguingly contrasting idea that LMB-hubs are structurally plastic, but conserved in sequence. It therefore appears that the conserved attribute is the ability of the hub-partner complex to dock onto the hub and/or partner protein in/near the interaction site.

In a parallel example, the catalytic subunit of the serine/threonine protein phosphatase calcineurin (PP2B or PPP3) contains an LMB site favoring a PxIxIT motif (Figure 4.1)[236,237]. Partners encoding this motif target calcineurin to a wide variety of substrates and regulators. Upon binding, disordered partners form a β-strand and fit within a small binding groove[238]. Like LC8, the available crystal structures indicate that the binding groove is the most plastic region of the free protein (Figure 4.1). Unlike LC8, calcineurin is an enzyme, and displays broad conservation throughout the active site, as well as within the binding groove (Figure 4.1). It is remarkable that some of the most sequence-conserved regions of calcineurin and LC8 are also the most structurally variable parts, and supports the hypothesis that this structural plasticity allows accommodation of a diverse set of partners with a wide range of properties.

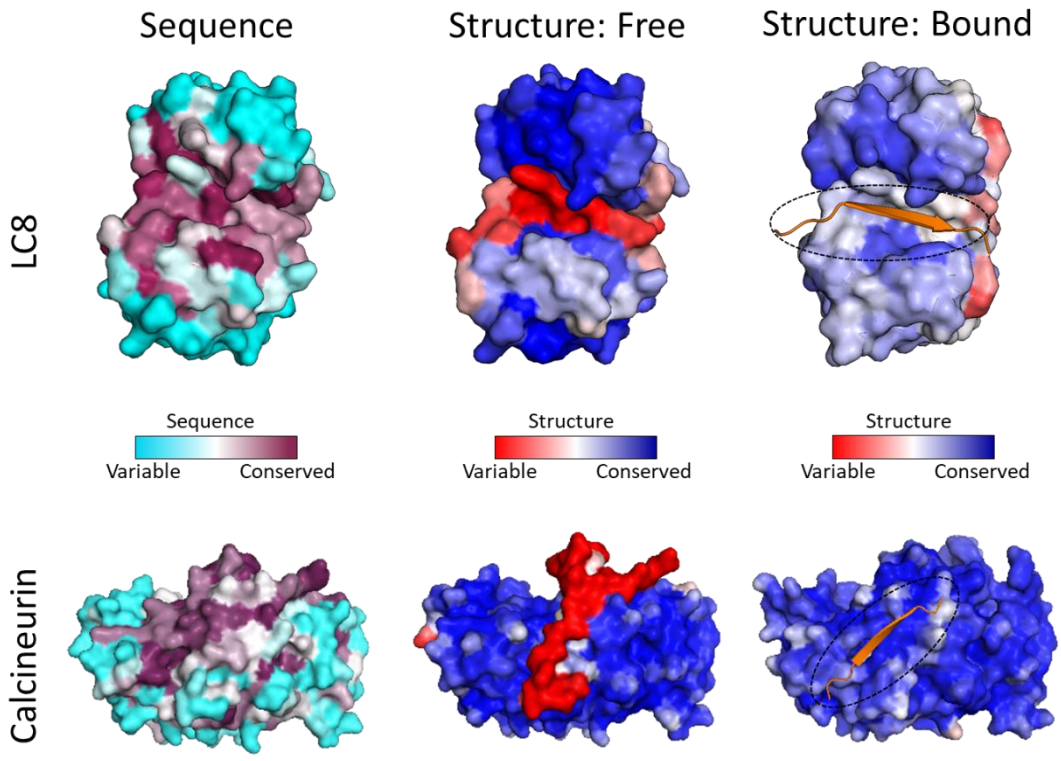


Figure 4.1: Sequence conservation and structural variability in LMB-hubs.
A comparison of the sequence (left), and structural conservation for free (middle) and bound (right) LMB-hubs. Some of the most conserved regions in terms of sequence, are some of the most variable structurally for both free LC8 (top) and calcineurin (bottom). The binding region is significantly stabilized when bound to peptides, suggesting that the plasticity is functionally useful for facilitating binding of diverse sequences. Conservation analyses were performed using ConSurf[187] on 500 sequences with sequence identities ranging from 40% to 95%. Structural conservation was measured using the Ensemblator's global overlay tool[137] on available structures for free calcineurin (4 structures), bound calcineurin (8 structures), free LC8 (5 structures), and bound LC8 (17 structures).

Dividing LMB-hubs into enzymatic and structural categories

While LMB-hubs are defined by their interactions with short motifs typically situated in disordered regions of proteins, hub-partner interactions are amenable to the same functional categorization as other proteins. In particular, segregation of enzymatic proteins (such as phosphatases and kinases) from structural, chaperoning proteins (such as heat shock proteins and PZD domains). We refer to these two groups as enzymatic LMB-hubs and structural LMB-hubs, respectively. The key point of differentiation is the transient nature of the enzymatic interaction, in which recognition sequences are bound until modified in the enzyme reaction. The transient nature of their interaction means that many diverse partners can be bound to the same hub protein

in a short time. One well-known example is the serine/threonine-specific protein kinase B (PKB, also known as AKT1). PKB is a key basophilic kinase that plays a role in cell survival, cell cycle progression, metabolism, and apoptosis[239] by targeting and phosphorylating proteins with a minimal consensus sequence of RxRxxS/T[240,241]. While it is far from the fastest enzyme[242], the entire binding, modification, and dissociation process occurs on the order of milliseconds[243].

Structural LMB-hubs are concerned with formation and stabilization of higher order protein associations (Figure 4.2). This function is comparable to roles described for intracellular chaperones like heat shock proteins (HSPs), which stabilize protein tertiary structure. While the most noted function of HSPs is the facilitation of protein folding, they are also involved in protein degradation[244] and prevention of deleterious protein aggregation[245]. "Consensus" motifs are often highly variable, and indeed, the heat shock cognate 71 kDa protein (HSC70) has two different described consensus motifs. Proteins containing an accessible KFREQ-like sequence interact with the peptide-binding domain of HSC70, leading to their translocation into lysosomes and eventual degradation[246,247]. On the other hand, sequences rich in accessible bulky hydrophobic groups and positively charged residues, but lacking negatively charged residues, can bind to HSC70 to activate its folding activity[248,249].

While HSPs stabilize tertiary protein structure, PDZ domains are involved in stabilization of quaternary protein structure. PDZ domains are 80-90 amino acid domains that occur within a wide variety of proteins. While there is variability in the binding motif based on PDZ class, they all bind to the four C-terminal residues of the protein, typically consisting of an xxΦ* motif (where 'Φ' is a hydrophobic residue, and '*' is the C-terminus)[250]. Interestingly, PDZ domains often arise as repeating clusters within a protein, providing scaffolding to construct large, macromolecular assemblies[251]. For example, photoreception in the eye of *Drosophila* species is dependent on the INAD protein[252], which encodes 5 separate PDZ domains. These PDZ domains work together to recruit various key components of the photoreceptor transducisome, as well as to promote oligomerization of INAD[253,254], making them integral to transducisome structure. Unlike enzymatic LMB-hubs, stabilization of large molecular assemblies often requires integration of structural LMB-hubs like PDZ domains within the complex, and partners of structural LMB-hubs are consequently competing directly for limited hub pools.

At the systems level, network analyses indicate that LMB-hubs act as stabilizing chaperones by organizing and connecting biological processes and pathways. Removal of hub protein "nodes" in interaction networks tend to make systems much more sensitive to additional perturbations[255]. This suggests that the structural plasticity of LMB-hubs binding sites endows a

functional adaptability that allows cells to efficiently respond to various requirements. We therefore propose that LMB-hubs are particularly important to a cell's ability to respond effectively to external stimuli such as increased salinity or temperature. Unfortunately, knock-outs (or even knock-downs) of hubs are often lethal[76,256,257], which complicates experimental tests of the importance of hubs in the response to external stimuli. Intriguingly, recent work by King *et al*.[72] reports the generation of a limb bud mesoderm-restricted conditional knock-out of LC8, resulting in a potentially very informative animal (and cell-line) model. Although these cells are not yet fully characterized, preliminary analyses indicate that knockouts of LC8 lead to impairments in the Hippo signaling pathway[72], amongst other things. This is consistent with previous data suggesting an strong enrichment in LC8 binding proteins within the Hippo pathway[75], which is partially responsible for the regulation of cell growth and apoptosis[258]. Further studies are warranted to identify the broader impacts of removing highly central hub proteins.

LC8 can be firmly placed within the category of structural LMB-hubs (Figure 4.2). As an example, LC8 is perhaps best known for its role as a light chain in the dynein complex[1], wherein it brings two protomers of the dynein intermediate chain together, stabilizing the essential quaternary motor complex[19,259]. Experiments with the yeast dynein complex found that deletion of LC8 led to a significant decrease in complex formation, decreased velocity and processivity for the motor, as well as increases in dynein heavy chain aggregation[41]. Notably, the phenotype for a knockout of lc8 was equivalent to a knockout of both LC8 and the dynein intermediate chain, indicating that LC8 acts as a small lynch-pin for the whole complex. Parallel examples can be seen with the centriolar protein Ana2, and the syntaphilin protein involved in mitochondrial mobility and inhibition of syntaxin-1. In the case of Ana2, LC8 interactions lead to tetramerization of Ana2, facilitating interactions with downstream partners like Sas6 and mud, and eventually contributes to correct spindle orientation[37,260]. For syntaphilin, interactions with LC8 stabilize a coiled coil structure proximal to the LC8 binding site, and allows it to recruit and anchor axonal mitochondria for synaptic regulation[71].

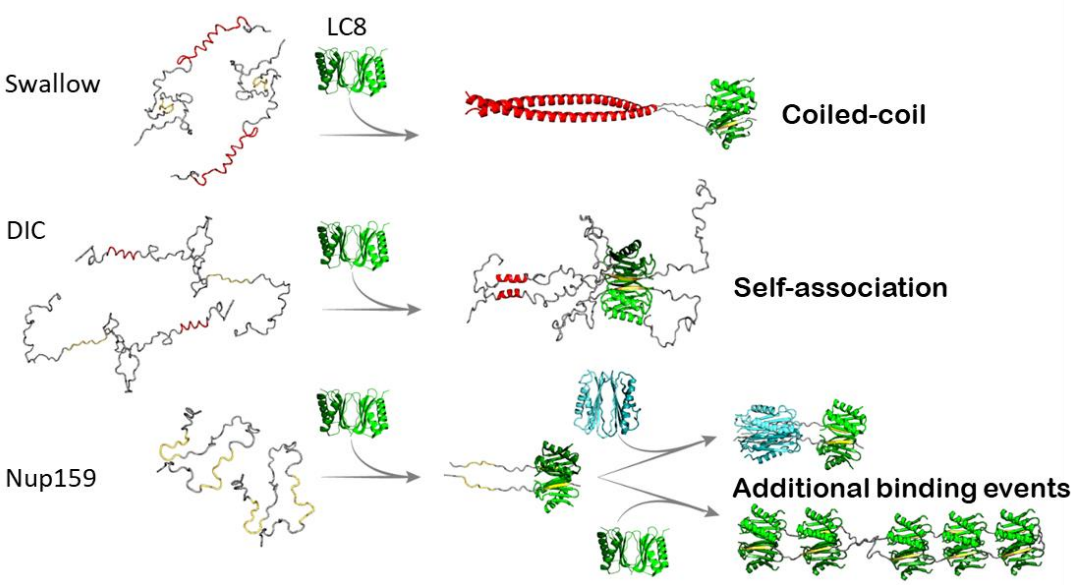


Figure 4.2: Functions for structural LMB-hubs.
Models depicting the role for structural LMB-hubs, using LC8 interactions as an example system. LC8 binds to disordered chains at the recognition motif, causing the formation of an adjacent coiled-coil (in swallow, top)[84], a self-association domain (in the dynein intermediate chain, middle)[25], or an additional binding event (in IC and Nup159, bottom)[29,259].

LMB-hubs concentrations are tightly-controlled and feedback-regulation may be common

Since LMB-hubs form 1:1 complexes with many partners, several questions arise: How can many different partners bind the same LMB-hub when they all compete for a common binding site? With a limited pool of each hub, how does a cell allocate sufficient protein for each function? For enzymatic LMB-hubs, the case can be made that transient binding facilitates the maintenance of sufficient pools of hub protein. However, structural LMB-hubs tend to be integrated into high order complexes where they are inaccessible to other partners. Interestingly, for many hub proteins there is evidence for "just-in-time" synthesis[257], *i.e.* protein expression is commensurate with the cell's needs, suggesting that hub protein concentrations are delicately balanced and exquisitely controlled. For the vast majority of hub proteins, such regulatory systems are not well described; however, we hypothesize that feedback loops allowing for LMB-hub self-regulation are common. A prime example of hub protein self-regulation can be seen for LC8 regulation by the transcription factor ASCIZ (ATMIN-Substrate Chk-Interacting Zn^{2+} finger)[69].

ASCIZ is an 88 kDa protein identified as a transcription factor for the hub protein LC8[68]. Recent mouse experiments demonstrate that LC8-1 (the more highly expressed of two LC8 isoforms) germline knock-outs lead to death early in embryogenesis, while a conditional knockout can proceed with severe developmental defects[72]. Additionally, ASCIZ knockouts display identical, but attenuated, phenotypes, suggesting that the most essential role for ASCIZ is regulation of LC8 concentrations within the cell[72]. Structural work on the LC8-ASCIZ complex reveals a remarkable number of LC8 binding sites (11 within human ASCIZ) and shows that this multivalency is important for regulation of LC8 concentrations within the cell[69]. Although the specifics of this regulation still require further research, it appears that when LC8 concentrations are high, the full complement of ASCIZ binding sites are occupied, leading to inhibition of transcription of the LC8-1 gene (Figure 4.3). Conversely, when LC8 concentrations are low, the linear motifs on ASCIZ are largely unoccupied, leading to LC8 transcription and translation[69]. Seemingly, LC8 is part of a feedback loop tightly regulating its own concentrations within the cell.

Although the LC8-ASCIZ interaction is one of the most explicit examples of LMB-hub feedback regulation, indirect routes of self-regulation can be identified for many other hubs. For example, the plant protein RCD1 (Radical-induced Cell Death 1) interacts with a (DE)x(1,2)(YF)x(1,4)(DE)L motif in disordered regions in many transcription factors[261]. Knock-outs of RCD1 result in pleiotropic effects in a variety of stress and developmental pathways. One particularly well studied system involves the generation of ROS (reactive oxygen species) from mitochondrial dysfunction. High concentrations of ROS cause inactivation, aggregation, and the eventual degradation of RCD1. RCD1, in turn, binds to and inhibits various transcription factors tasked with production of MDS (mitochondrial dysfunction stimulon) genes, whose role is to ameliorate the production, and control the impact, of ROS[262]. Thus, when RDC1 concentrations are high, production of MDSs is decreased, leading to increased ROS, which in turn facilitate RDC1 degradation. We hypothesize that as the regulatory pathways for various LMB-hubs are more adequately described, similar feedback loops will be found, and illustrate the importance of self-regulation for these highly sought after hubs. Due to their importance, LMB-hub motif preferences are often thoroughly studied, leading to the generation of partner prediction tools[22,210,212]. It is therefore possible to combine information about gene regulation with partner predictions to identify potential feedback loops.

14-3-3s are a family of structural LMB-hub proteins involved in cell-cycle control, signal transduction, protein trafficking, and apoptosis, among other things[263,264]. They are often dimeric,

and usually bind to phosphorylated partners at a basic and proline-rich motif[215,265]. Binding facilitates partner protein folding, localization, and stimulation or inhibition of downstream interactions[80]. Previous work describing motif preferences for 14-3-3 has led the development of a motif predictor, 14-3-3-Pred, which uses a combination of position-specific scoring matrices, support vector machines, and artificial neural networks to predict likely 14-3-3 binding sites[210]. Importantly, it has been effectively used to identify actual 14-3-3 motifs in various partners/systems[266,267]. We have therefore examined ATF-1, a cyclic-AMP response element binding transcription factor known to regulate intracellular 14-3-3ζ concentrations[268], and identified a high-probability 14-3-3 binding motif within this transcription factor. Additionally, the likely binding motif occurs at a known phosphorylation site (S63, UniProt code: P18846) previously associated with enhanced ATF-1 transcriptional activity[269,270]. These data suggest that 14-3-3s may also have a self-regulatory mechanism (Figure 4.3), supporting our hypothesis that LMB-hub self-regulation is a common requirement for effective regulation of intracellular hub concentrations.

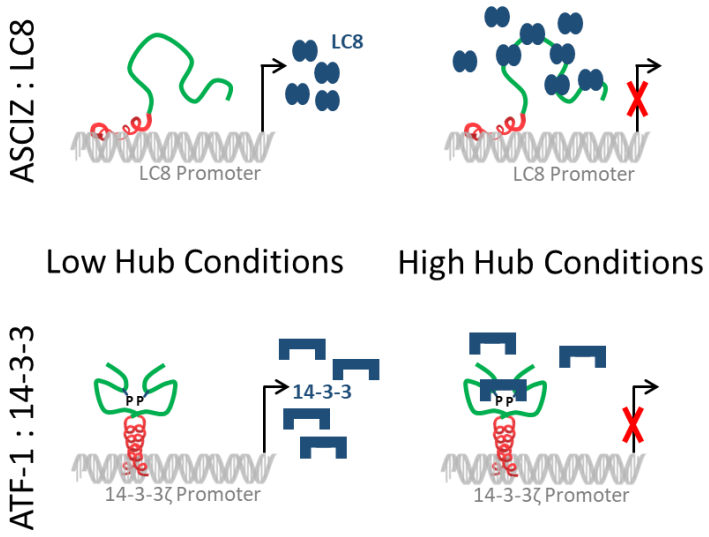


Figure 4.3: Structural LMB-hub transcriptional self-regulation.
In the proposed models of LC8 regulation via the transcription factor ASCIZ, and 14-3-3 regulation via the transcription factor ATF-1, low concentrations of the structural LMB-hub would result in a free version of the transcription factor, facilitating increased hub protein production. When hub concentrations are high, binding to the transcription factor would turn off hub production. We hypothesize that this a common regulatory tactic for structural LMB-hubs. Red helices represent DNA-binding domains, green lines are IDRs, grey double-helices are DNA, Small 'Ps' are phosphorylated residues on ATF-1, and LMB-hubs are shown in dark blue.

Viral Hijacking of Host LMB-Hubs

In the first section, we described how structural LMB-hubs have dynamic binding interfaces, conserved motif binding sites, integral roles in stabilizing proteins and macromolecular complexes, and are tightly regulated within the cell. We propose that these traits make LMB-hubs efficient targets for viruses looking to hijack a cell. The feasibility and advantages of viral hijacking of host LMB-hubs are outlined below.

Fast viral mutation rates facilitate the evolution of linear motif mimicry in viruses

One of the qualifying features of a linear motif is its short length, from three amino acids in the case of KEN-box binders like CDH1[271], to a rare but variable 18 amino acids in WH2 motif bound by actin[272]. Therefore, generation of a novel LM in a viral protein often requires only a couple of missense mutations. In contrast to eukaryotes, whose emphasis is largely conservation of sequence with slow, stochastic mutagenesis (10^{-8} mutations per position per cellular generation for humans)[273], viruses often encode low fidelity polymerases and emphasize adaptability (10^{-5}-10^{-3} mutations per position per generation)[274]. A high mutation rate and short sequence requirement results in a high concentration of LMs. In fact, it has been demonstrated that viruses participate in more domain-motif interactions, bind more unique LMB-domains/residue, and have a higher concentration of unique LMs/residue than eukaryotic proteins[225].

Due to the ease of LM genesis, convergent evolution of viral LM-LMB-domain interactions is common[275,276], even within a single viral family. The papillomavirus protein E7 is a short, oncogenic sequence that affects everything from cell entry to cell death[277]. Most saliently, it contains a high concentration of LMs[278,279], with seven known motifs within the ~100 amino acid sequence, including a C-terminal PDZ binding motif[280]. Interestingly, PDZ-binding motifs are only present in a few phylogenetically distant clades, suggesting that it has evolved separately on multiple occasions[281].

Another example is illustrated by comparison of rabies virus phosphoprotein to Ebola virus protein VP35. Both viruses are (-)ssRNA viruses in the *Mononegavirales* order, and both have demonstrated LC8 interactions[66,90]. While the LC8 motif is highly conserved within the VP 35 protein of *Ebolaviruses*, the closely related *Marburgvirus* VP35 lacks the motif. The same pattern is seen for the phosphoproteins of the *Lyssavirus* genus (rabies), where all other closely related genera do not have LC8 binding motifs (Chapter 2). This leads to an interesting scenario, where LC8 interactions have cropped up in two analogous proteins[282], but are missing from all of

the intervening viruses. Additionally, available data suggest that the function of LC8 is conserved between rabies and Ebola viruses, as LC8 interactions appear to facilitate viral polymerase activity[90,92]. It is of note that this pattern of LM convergent evolution in viruses is a common occurrence, suggesting that many LMB-hub interactions are actively under strong selective pressures. Indeed, the E7 proteins of papillomaviruses have lost the requisite LxSxE motif necessary for binding the retinoblastoma protein within some clades[279,283], and even display the selective loss and reoccurrence of E2F motifs within the same protein[279,284]. These example provide evidence that LMs are under constant selective pressures, both positive *and* negative. LMs therefore provide viruses with a simple and efficient method for evolutionary plasticity, allowing them to gain and lose interactions and functions much more quickly.

Viral linear motifs often outcompete host linear motifs

The complex interaction network surrounding hub proteins necessitates an intricate control system, where binding to various partners is regulated by spatial, temporal, and affinity-based differences in partners. For example, there is no competition between the pool of LC8 that binds ASCIZ in the nucleus, and the pool of LC8 interacting with dynein in the cytoplasm. Similarly, many hubs show periodicity in their expression and/or interactions, as seen in cyclin proteins, where different partners are bound depending on the cell cycle stage[79,285]. On the other hand, in cases where many partners are temporally and spatially in sync, there is a delicate balance between interactions with binding partners, which must be regulated by slight variations in affinity[285].

While evolution has favored a fine-tuning of affinities in host interaction networks, viral proteins often adopt a high-affinity, idealized LM in order to outcompete host proteins for the LMB domain[275]. A particularly compelling example of this can be seen in rabies viral hijacking of LC8, where the affinity of the rabies phosphoprotein for LC8 (80 nM, Chapter 2), is 2 times tighter than the highest affinity eukaryotic partner yet tested (160 nM for Nek9)[286]. Additionally, the Nef protein of HIV binds to the SH3 domain of Hck with a higher affinity than any other known partner (0.44 μM)[275,287].

Linear motifs are the most parsimonious method to appropriate host systems

Viral genomes are greatly constrained in size by a variety of factors, including: 1) an advantage in replication rates provided by smaller genomes, 2) restrictions in space caused by viral capsids, and 3) restrictions in fidelity caused by a lack of proof-reading in RNA viruses,

leading to an increased rate of deleterious mutations in large genomes[288]. Due to these size restrictions, and the requirements for a large number of interactions in the viral infection cycle, viruses often have a higher number and concentration of interaction sites per protein[225]. LMs provide a unique and parsimonious opportunity to bind many partners, and to exert control by hijacking a highly central protein. Indeed, it is commonly found that viruses favor LMs for host interactions[225,275,289], and even in cases of mimicry of an entire interface, the virus interface is typically smaller than the endogenous interaction[290]. This suggests that viruses are actively taking advantage of the low LM requirements to efficiently appropriate a wide variety of host systems.

Conclusions

In this work, we describe a newly defined category of hub proteins, the "linear motif binding-hubs." We are by no means the first to recognize that dynamic/date hubs typically bind to linear motifs at a single interface, but we believe that this method of classification is much more clear and concise when compared to the alternatives in common use. We also suggest further vernacular changes in order to differentiate between structural and enzymatic LMB-hubs, as these types of proteins will be treated and used very differently within cells. Current data indicate that structural LMB-hubs like calcineurin and LC8 often have dynamic binding interfaces in order to accommodate many motif variants, but are surprisingly conserved in sequence, suggesting the conserved feature is binding site plasticity. Although thorough analyses of LMB-hub regulatory systems are scarce, available data on LC8 and RCD1, along with predictions for 14-3-3, lead us to speculate that self-regulation is an important and common feature for LMB-hubs specifically. Further work on LMB-hub transcriptional regulation is required to validate this hypothesis, but we propose using the existing motif-predictors on known transcription factors to facilitate this process.

Another emergent feature appears to be susceptibility to appropriation by intracellular parasites like viruses. The relatively short evolutionary space between non-binding and binding sequences means that viruses can, in theory, develop new LMB-hub interactions in a few generations, providing a huge amount of evolutionary plasticity. Separate studies have noted the increased tendency of viruses to hijack hubs[289,290], as well as to be rich in LMs[225,275]. Here we suggest that viruses target structural LMB-hubs in particular, because these proteins are both more easily obtainable due to the short motif, and less transient than other types of hubs (such as enzymatic hubs, where it is difficult to irreversibly bind to the active site using the LM).

Replacement of virally hijacked structural LMB-hubs would require *de novo* synthesis of new protein. This also suggests that viral motifs may represent an important source of information about “optimized” motifs, because while eukaryotic proteins have to balance LMB-hub-partner interactions with many proteins, viruses simply have to outcompete other partners. Future studies characterizing of LM preferences in viruses could be very useful to determine whether viruses hijack structural LMB-hubs to facilitate their own function, or to simply inhibit other host-hub interactions via the most parsimonious method.

Chapter 5

Conclusion

Impact

The studies presented here take an expanding approach to understand the importance of linear motif-binding hubs in host-virus interactions, with a particular focus on the hub protein LC8. We start by describing the role of a hub in one specific system, follow with a description of how to characterize hub interactions, and finish by describing the qualities of both hubs and viruses that make these interactions common. My work utilizes a variety of biophysical techniques (including NMR, ITC, CD, and SAXS), computational analyses, and biochemical methods, and represents years of experimentation and literature review. These data greatly increase our understanding of the rabies polymerase complex functionality, interactions preferences for LC8, and traits important for viral hijacking of host systems. The following section highlights important conclusions from each chapter. The remainder of this chapter describes projects that are nearing completion, under development, or of particular interest for future work.

Highlights of Reported Work

In Chapter 2, we present the first structural ensemble of RavP, and demonstrate that it is a highly flexible protein dimer composed of ordered domains and long disordered linkers. We also present the structural ensemble of the 90 kDa LC8-RavP complex. Interestingly, LC8 binding does not significantly alter the secondary or tertiary structure of RavP, but instead changes the dynamics and orientations of the domains that are distant from the LC8 site. The ensemble for the RavP-LC8 complex is populated with more compact structures that have folded C-terminal domains with dramatically restricted mobility, and as such, mirrors the ensemble of a related virus phosphoprotein that lacks the LC8 binding site. This suggests that LC8 binding is the switch that makes RavP function similarly to other viral phosphoproteins. This novel idea is based on the comparison of the structures of the free RavP to the RavP-LC8 complex, which we report using an integrative approach including NMR, SAXS and molecular dynamics simulation. A combination of methods is necessary for a complex system that is both highly dynamic and highly structured.

Previous work on LC8 interactions with any of its 100+ partners has been limited to either short peptide motifs or domains of the partner proteins, meaning we have produced the first structural ensemble for LC8 with *any* full-length partner. The restricted domain motions upon LC8 binding could also apply to LC8 interactions the Ebola viral VP35s, which have a similar

domain architecture to RavP, suggesting that LC8 may serve a common function within LC8-binding members of the *Mononegavirales* order. The implications are far reaching for drug discovery as they position LC8 to be targeted for similar treatments for rabies and Ebola. On the cellular side, we used targeted cell culture assays to show that LC8 interactions with RavP directly affects viral polymerase activity, but do not suppress the host innate immune response, a possibility that has been suggested in the literature. We further expand our analysis to include a purported LC8 binding site within RavL, and demonstrate that the LC8 site on RavL is not important for viral polymerase activity.

In Chapter 3 we examine the motif preferences for LC8 and identify 29 new binding sequences, 19 of which occurred within entirely undescribed LC8 partners. We initially use proteomic peptide phage display technology to identify potential partners from a large library of peptides. Importantly, this library had no innate bias for anchor motif sequences, an issue with previous studies which may have impacted their attempts at motif description. We then use our algorithms to identify sequences based on their comparison to known LC8 binders. We synthesize ~80 14-amino acid peptides, test their ability to bind LC8 by ITC, and identify new binding partners with various affinities. Particularly innovative in this work is the development of an algorithm to compute recognition motifs that predict LC8 binding with a reliability that is significantly higher than previously reported approaches. This improvement was only possible because we were 'fortunate' enough to test many TQT anchor-containing sequences that do not bind to LC8. While prior studies were limited to assuming that residues were untenable at certain positions based on depletion data, we were able to use non-binders to normalize a PSSM, providing weights where a positive weight is more frequently found in the binder dataset, and a negative weight is more frequently found in the non-binder dataset. The combination of "positive" and "negative" binders thereby enhances the discriminatory power of our model. In addition, we establish a curated, publicly available database of known interactions (http://lc8hub.cgrb.oregonstate.edu), with general requirements for LC8 motif recognition. All of the interactions in the database are either experimentally validated by us or by others in the literature. As more LC8 binding sequences are tested, we can further improve our algorithm in order to predict LC8 partners with higher confidence.

Chapter 4 is a combination of literature review and an opinion/perspectives piece. Previous work has sought to categorize hubs into dynamic/date and party/static. I am of the opinion that these terms are not explicit or obvious (especially since static and dynamic do not describe the protein's structure). We therefore propose the new term "linear motif binding-hubs,"

or LMB-hubs in place of date/dynamic hubs, as it much more directly describes the most common feature of this type of hub protein. We further segregate LMB-hubs into either enzymatic or structural, a distinction which not only describes the function for the hub within the cell, but also can be used to predict some regulatory and binding features. For example, examination of self-regulation (or feedback regulation) in multiple structural LMB-systems suggests that this a common tactic to tacitly control intracellular hub concentrations. On the structural side, we show that LMB-hubs often have dynamic binding interfaces in otherwise static structures, a feature which is highly conserved. These findings and vernacular distinctions allow us to more succinctly describe and understand the roles that structural LMB-hubs play within various essential cellular pathways.

Viral proteins are significantly enriched in linear motifs. We explore this in Chapter 4 by detailing some of the more potent benefits of viral hijacking of host LMB-hubs, including increased evolutionary plasticity due to quick mutation rates, and parsimonious appropriation of essential host pathways. Although previous work has noted the enrichment of linear motifs in viral proteins, we have taken it a step further to describe how LMB-hub features are utilized by viruses to efficiently hijack a host, despite their limited genome size.

When I presented to my committee for my program of study meeting, I tentatively titled my work “the role of intrinsic disorder in macromolecular assemblies,” and I intended to use ribosomal biogenesis, dynein complexes, and viral polymerase complexes as my example systems. Ribosomal biogenesis did not go very well. Despite not being in my thesis, I did work on dynein complexes quite a lot, and will eventually have a paper published on that work. But it turns out that I’m quite into viruses, and think they are really interesting. Two out of three isn’t so bad!

Ongoing Work & Future Directions

I expect that I am not unique in having a huge number of projects that are either partially completed, or in their early stages. Frankly, I could write an entire thesis entitled “Experiments I think would be cool to do.” This section describes a few of those experiments and where things stand with them currently.

Identification of downstream RavP interactions

In Chapter 2 we show that LC8 interactions are important for polymerase activity based on minigenome data, demonstrating a 70% reduction in transcription rates when the LC8 anchor motif is mutated to AAA. In an attempt to better understand this interaction, we further mutated the entire 14 amino-acid binding region to both a 2-heptad repeat (in order to simulate an extended dimerization domain, although NMR data indicate it does not for a coiled-coil), and a random disordered sequence of equivalent length. Surprisingly, *both* of these mutations resulted in a partial amelioration of polymerase activity. The accumulated data therefore indicate that there may be another binding partner that shares LC8's binding site. In the RavP-WT case, LC8 binds very tightly and outcompetes this partner. When the LC8 site is mutated, as in RavP-AAA, this competitive partner may bind RavP more frequently, impacting RavP's cellular localization (or functionality) in a way that inhibits polymerase activity. In the 3rd case, when the entire LC8 site is mutated so that (presumably) both LC8 and partner "X" cannot bind, polymerase functionality is only partially inhibited, suggesting that LC8 is both directly and indirectly involved in supporting the polymerase complex.

In order to test whether LC8 is indeed competitive with another RavP binding partner, we performed HA/FLAG-tag pull downs of RavP-WT and RavP-AAA from HEK293T cells. In theory, the RavP-AAA mutant will pull down more of the hypothetical partner than the RavP-WT because of a lack of competition from LC8. Pull-downs were performed in triplicate, and analyzed via a Mass Spectrometry proteomics analysis. With this method, we identified 17 proteins that are significantly enriched in the mutant compared to wild-type, indicating that these proteins bind better when LC8 is not present. On the other hand, there are two proteins (including LC8), which are reduced in their binding when LC8 is absent. Many previously described RavP partners (such as STAT1) were pulled down in equal proportions. Although none of the enriched hits are known binders, one is of particular note.

Work by Kammouni *et al.*[143] demonstrated that RavP interacts with mitochondrial Complex I, leading to increased Complex I activity and production of reactive oxygen species. They also localized the binding region between RavP and the unknown Complex I to residues 139-172, exactly the area bound by LC8 (the LC8 motif spans residues ~140-150). It is therefore feasible that LC8 interactions with RavP would be in direct competition with those from a subunit of Complex I. Indeed, one of our enriched partners in the pull-down analysis is NADH Dehydrogenase [Ubiquinone] 1 Alpha subcomplex subunit 5 (NUDFA5), an accessory subunit to

complex I. I believe that this is a strong hit, and supports our hypothesis that there are proteins directly competing with LC8 for pools of RavP.

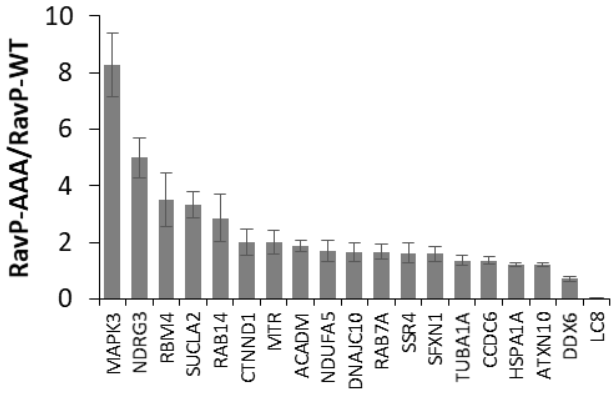


Figure 5.1: Mass spectrometry pull-downs of RavP partners.
Pull-downs were performed in triplicate using α-FLAG beads, and analyzed via Mass Spectrometry. The above graph depicts proteins that were significantly enriched or depleted based on a student's T-test, compared to an empty vector control transfections. Additionally, only those that were increased or decreased by more than 20% are shown.

As of now, we have ordered 8 of the more promising hits (based on their size and known solubility/expression information; RBM4, ACADM, NDUFA5, RAB7A, SFXN1, DDX6, SUCLA2, and SSR4), and are cloning them into eukaryotic expression vectors. The next step will be to perform specific pull-downs of these partners during cotransfections with RavP to verify binding. After verifying that they do indeed bind, we will be able to take advantage of the plethora of NMR data collected on RavP to determine their binding sites, and to characterize their impact on RavP structure. The final stage of this project would be to re-visit the minigenome luciferase assays with cotransfections of verified partners. In this case, we would expect that overexpression of LC8 competitors would decrease polymerase activity, similar to what happens when the LC8 site is mutated.

LC8 interactions with Ebola VP35 and Rotavirus VP4

Unlike typical LC8-binding proteins, viral partners studied thus far have a strong self-association domain proximal to the LC8 binding site, indicating that LC8 is not required for their dimerization. It is therefore unlikely that LC8 is performing the same function for viral proteins as it does for eukaryotic proteins (*i.e.* facilitating dimerization).

Ebola viruses (EBOV) are particularly well known viruses, due to the recent outbreaks in West Africa (2013-2016)[291] and the Democratic Republic of Congo (2018)[292]. It has an up to 90% fatality

rate, because of the virus's tendency to globally suppress the host's immune system[293]. EBOV have –ssRNA genomes (*Filoviridae* family), and consequently belong to the same *Mononegavirales* order as rabies and VSV.

VP35 is structurally related to RavP (Chapter 2) and serves analogous functions including suppression of the host immune system and bridging the viral polymerase and nucleoprotein to facilitate viral transcription and replication[148]. Unlike RavP which is a tight dimer, VP35 forms a higher order oligomer, and recent data indicate that EBOV VP35 has equally stable trimeric and tetrameric structures[294]. On the other hand, the similar Marburg virus (*Filoviridae*) VP35 lacks the LC8 site, and is unequivocally a trimer[295]. It is unclear whether the trimeric and tetrameric forms of EBOV VP35 facilitate separate functions, but the easy interplay between these two states provides an interesting opportunity for LC8 to mediate selection of association states. LC8 binds all known partners in a bivalent fashion[296], therefore, we hypothesize that interactions with LC8 will favor tetrameric VP35 (Figure 5.2A).

Experimental procedures for the production of hexahistidine tagged EBOV VP35 (residues 83-145) were published earlier this year[294], and led to the generation of crystal structures of the trimeric and tetrameric forms. We have followed a similar protocol for protein production using a construct containing residues 65-145, which includes the LC8 binding site (residues 69-76) and the multimerization domain (MD). To identify the effect of LC8 on the association state, we have used analytical ultracentrifugation (AUC), SEC, and MALS. Interestingly, our SEC and AUC data indicate that there are two equally stable peaks for VP35, *both* of which interact with LC8 and migrate at the same place on an SDS gel; however, these two peaks appear to be non-interconverting, as isolation of one peak does not result in the re-appearance of two peaks after three days. Future work on this project would seek to describe the difference between peaks one and two, as well as to directly understand the importance of the LC8 interaction in the Ebola virus life cycle. We would hypothesize that LC8 is also important for viral transcription in Ebola, which could be tested using a similar minigenome experiment as was described for rabies.

Rotaviruses cause acute diarrhea with an estimated 215,000 deaths annually in children under 5, largely in developing countries[297]. Rotaviruses are not restricted to humans, and are often problematic for large cattle herds, disproportionately affecting young calves[298]. The virus uses dsRNA as its genetic material, and is a member of the *Reoviridae* family. As our ITC data in Chapter 3 demonstrate, the rotavirus protein VP4 is an LC8 interacting partner. Although LC8 may serve a common function for Ebola and rabies, there is little functional similarity with VP4. While VP35 and RavP are important for viral polymerase activity, rotavirus VP4 is essential for viral

penetration, and forms pseudo-trimeric spikes on the outer face of the mature virion. To become fully infectious, the VP4 pseudo-trimers are digested by trypsin-like proteases from the intestinal lumen, which stabilizes the spike structure and facilitates viral penetration (Figure 5.2B)[299].

In addition to roles in viral penetration, VP4 also contains major antigenic sites targeted in vaccination[300]. Because of this, many studies have focused on the efficient recombinant production of VP4 in bacterial and baculovirus systems[301,302], and multiple structural analyses of VP4 exist. Cryo-EM studies of the mature virion revealed an atypical structure for VP4, where the foot domain displays homotrimeric symmetry, whilst the two other domains are homodimeric (Figure 5.2B,C)[303,304]. Crystal structures of the central antigenic domain indicate that it can easily form either dimeric or trimeric structures, depending on the conditions[305]. On the other hand, AUC studies on purified VP4 show that it is monomeric in solution[302]. While these data appear somewhat contradictory, we hypothesize that they are each capturing a snapshot of the oligomerization states that VP4 occupies during an infection cycle. The pseudo-trimeric/dimeric structure has been thoroughly described, and is essential for VP4's role in cell penetration.

During the viral replication process, VP4 is localized throughout the cytoplasm, and is not yet interacting with nascent virions. Very little is known about VP4 structure or function during this time, but confocal microscopy imaging demonstrates that it is highly expressed, and particularly concentrated to cytoskeletal regions[306]. Solution experiments indicate that monomeric VP4 is predisposed towards digestion by a variety of proteases, including trypsin, at locations that are protected in oligomeric VP4[302]. Residue 582 within the 'foot' region of VP4 is even more sensitive to digestion than the intended trypsin site (Figure 5.2C). Interestingly, our data show that there is at least one LC8 site just C-terminal of the trypsin cleavage site (residues 604-611), suggesting a potentially intriguing role for LC8 in the stabilization of the homodimer and protection from trypsinization. A future study could examine the oligomerization state of VP4, and the structural impact of LC8 interactions, using AUC, size exclusion chromatography, and negative stain electron microscopy. In addition to the biophysical analysis, we could also expose the free and LC8-bound forms of VP4 to a series of proteolytic digestions. We hypothesize that LC8 interactions are protective for VP4, and that the heterocomplex will be less prone to proteolysis. This study is in its early stages, because we have had limited success expressing the full-length constructs in *E. coli*. However, cellular data that I collected in France strongly support that LC8 plays a role in the rotavirus infection cycle, evidenced by abrogated LC8 localization in rotavirus infect cells (Figure 5.2D).

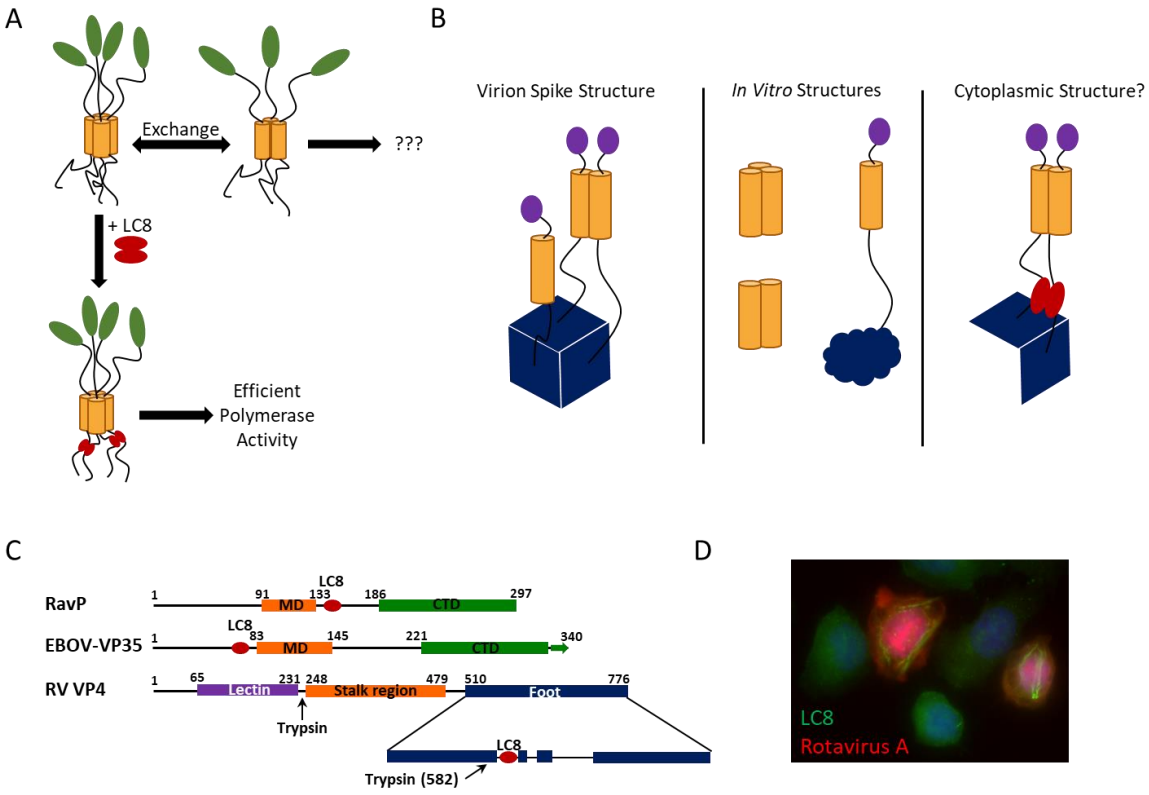


Figure 5.2: LC8 interactions with Rotavirus VP4, and Ebolavirus VP35.
(A) VP35 can form equally stable trimeric or tetrameric oligomers. We hypothesize that LC8 stabilizes the tetrameric structure, facilitating viral polymerase activity. Domains with solved structures are shown in green (CTD) and orange (oligomerization domain). LC8 is shown in red. (B) VP4 structure is context dependent. Within the virion, VP4 forms a pseudo-trimeric spike structure with a trimeric foot domain (blue), a dimeric stalk domain (orange) and separate lectin head domains (purple). Monomers, dimers, and trimers have been validated *in vitro,* and the foot domain (blue) is significantly more accessible in the monomeric structure. Cytoplasmic accumulation of VP4 has been noted, but no structural information is available; however, abrogated LC8 localizations within the cytoplasm during RV infections suggests interactions occur between LC8 and VP4. (C) Domain architectures for RavP (top), VP35 (middle), and VP4 (bottom). Colored blocks represent structured domains, and black lines denote disordered segments. The VP4 foot domain has a predicted disorder at the LC8 site, and the trypsin site. (D) Modified HEK cells expressing GFP-LC8 (green) under the endogenous regulatory system, infected with rotavirus A virions containing an mCherry gene (red). Note that uninfected cells show broad LC8 cytoplasmic localization, while infected cells have a tubular LC8 localization that is often perinuclear.

Phylogenetic analysis of Mononegavirales and LC8 conservation

In Chapter 4, we discuss how linear motif development patterns in viruses evince both positive and negative selective pressures. Chemes *et al.*[279,284] present data on LMs in the Papillomavirus protein E7, demonstrating three types of LM evolution (Figure 5.3). 1) All motifs present are homologous. In this case, LMs are lost in some intervening clades of E7, but ancestral reconstruction suggests that a single appearance event occurred. 2) Some motifs are homologous, and some motifs are convergently evolved. 3) All motifs are convergently evolved. I believe that this analysis is extremely powerful, because it demonstrates positive and negative selection, gives a time-scale for LM evolution, and can even describe some differences between different papillomavirus clades.

I would propose a similar analysis be done within the Mononegavirales Order for LC8 motifs. This would provide a variety of benefits, the most obvious being identification of other viruses utilizing LC8; however, it is also likely that we will glean additional information about LC8's role in Mononegavirales viruses based on which genera contain conserved LC8 motifs. Indeed, I have already obtained a list of phosphoproteins and VP35s from UniProt, and tested the presence of LC8 motifs using LC8Pred (described in Chapter 3). In addition to *Lyssavirus* and *Ebolavirus* genera, LC8 motifs are present in many *Sprivivirus*es, and some *Respiroviruses* and *Avulaviruses.* Future work would phylogenetically categorize these viruses, potentially using the nucleoprotein or polymerase protein as a proxy (due to poor sequence conservation in the phosphoprotein). Describing known similarities in binding location and functions of these viruses could help us understand whether LC8 is being used for the same purpose in all motif-containing Mononegavirales viruses. A final point of interest would be to search *other* proteins in viruses not encoding a phosphoprotein LC8 motif. If viruses are using LC8 for a more general 'control' function, then it is possible that the motif only needs to be present in *some* viral protein, rather than specifically the phosphoprotein.

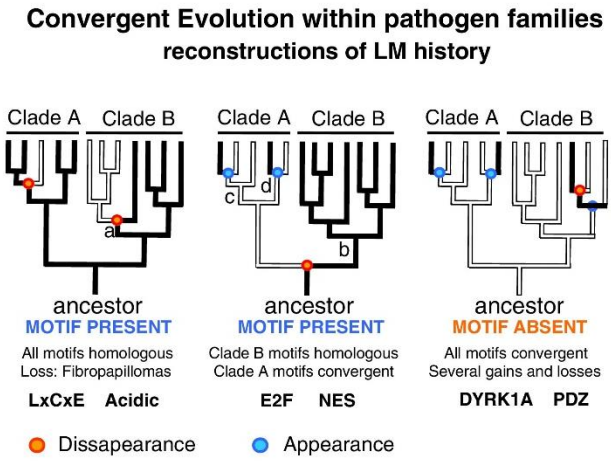


Figure 5.3: Convergent evolution of linear motifs in E7.
Phylogenetic analysis of E7 linear motifs reveals different evolutionary patterns. On the left, all occurrences are homologous. Examples include the LxCxE motif bound by Rb proteins. In the middle, some points are homologous (like b), and some are convergent (like c and d). E2F motifs display a conservation pattern like this. On the right, all instances are convergent evolution. PDZ domains in E7 exemplify this pattern of evolution. This figure was modified with permission from [281].

Differences between LC8 isoforms

In *Drosophila melanogaster*, there is a single LC8 gene. On the other hand, most mammals have two isoforms of LC8 that are greater than 90% sequence identical (83 of 89 residues). The majority of my thesis has focused on the more highly expressed isoform: LC8-1. Work characterizing LC8 regulation within the cell determined that, while LC8-1 is under control of an ASCIZ-mediate promoter, LC8-2 is constitutively expressed at ~5% of LC8-1 levels[72]. During early embryogenesis, qRT-PCR data suggest that ratio is even more dramatic, with LC8-2 being expressed at approximately 0.2% of the LC8-1 levels. On the other hand, the Human Protein Atlas indicates that LC8-2 is most highly expressed in the brain (data not shown). Despite the differences in regulation and expression, studies testing affinities and avidities of LC8-1 vs. LC8-2 have reported no, or minor, differences[307]. Out of curiosity, I have tested binding of RavP to both isoforms, and noted an approximately 2-fold change in affinity (Figure 5.4). Although this is not a huge difference, if we consider the delicate balance that must be struck between numerous hub partners, this could cause large, cascading effects. The difference in binding could be of particular interest when we contemplate the rabies life cycle. Rabies is known to replicate particularly slowly in the muscle and peripheral nervous system, and then ramp up expression when it arrives in the brain[308]. It is of note that LC8-2, which binds more tightly to RavP, is more

highly expressed in the brain where rabies replication is fastest. It is particularly interesting when viewed through the lens of our Chapter 2 data, demonstrating that LC8 plays a role in the polymerase complex. Future work could examine whether LC8-2 and LC8-1 stimulate RAV polymerase activity differently using a minigenome luciferase reporter assay.

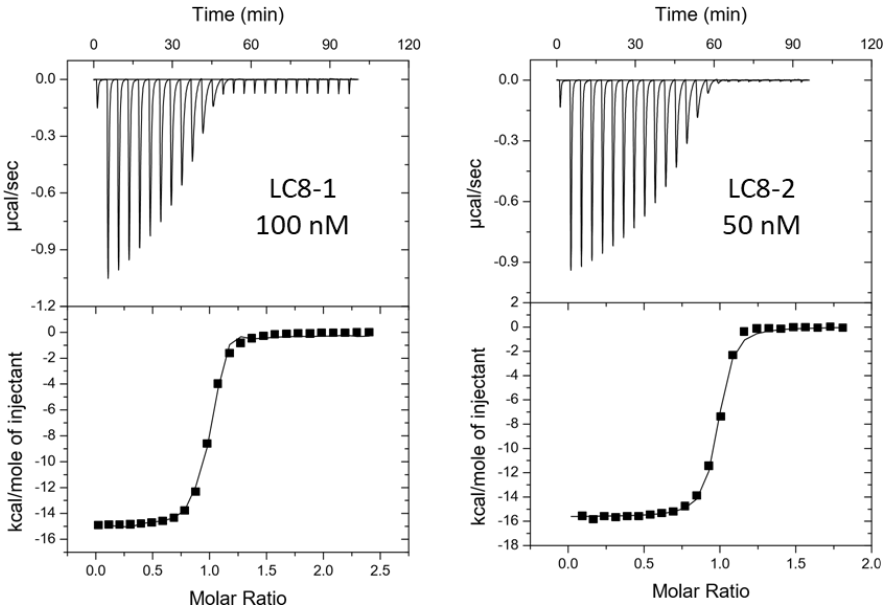


Figure 5.4: RavP interactions with LC8 isoforms.
Binding thermograms for RavP-FL interactions with LC8-1 (left) and LC8-2 (right) isoforms at 25° C. LC8-2 binds to RavP twice as tightly as LC8-1.

Examination of LC8-1 KO cells

Careful control of cellular concentrations of hub proteins is essential for cell survival and homeostasis. Unfortunately, it is extremely difficult to study regulation for LMB-hubs, because they have very complex regulation. For example, Calmodulin is expressed from 3 different gene loci, which are all under control of separate promoters and poorly defined regulatory systems[309]. Likewise, 14-3-3 has seven different isoforms which are variably regulated, both transcriptionally and post-translationally[310].

Recently, the Heirhorst lab has successfully developed both LC8-1 and ASCIZ knockout cell lines[72], which provide us with a unique opportunity to study cellular interactomes in an organism where a hub is both unregulated and significantly depleted. Unlike Calmodulin and 14-3-3, the pool of active LC8 seems to be regulated by a single transcription factor (ASCIZ)[72], and

there is little evidence for post-translational regulatory mechanisms. The second isoform of LC8 (LC8-2) is constituently expressed at low levels.

Future studies could take advantage of these cells to understand both the role of LC8, and the phenotypic response of central hub protein knockouts, knockdowns, and misregulation. If I proceeded with this research, I would perform transcriptomics studies on these knockout cells to determine the wide-spread importance of LC8. Additionally, it has been noted that hub proteins are mediators of cellular homeostasis, suggesting that *viable* knockouts may only be viable under ideal conditions. I would therefore expose cell cultures to a variety of stressful external stimuli, such as high salinity, altered pH, altered temperature, partial anoxia, and nutrient deficiencies. Previous tests have been performed to describe the impact of stressors like osmotic pressure on protein folding[311], and our work could piggyback off of this to determine whether removal of a structural LMB-hub impacts intracellular folding under different conditions. I believe that these cells present a unique opportunity to understand the role of LC8 (and by proxy structural LMB-hubs) in the mediation of cell homeostasis.

References

1. Pfister, K. K., Fay, R. B. & Witman, G. B. Purification and polypeptide composition of dynein ATPases from Chlamydomonas flagella. *Cell Motil.* **2,** 525–47 (1982).
2. Jaffrey, S. R. & Snyder, S. H. PIN: An Associated Protein Inhibitor of Neuronal Nitric Oxide Synthase. *Science (80-.).* **274,** 774–777 (1996).
3. Parhad, S. S., Jaiswal, D., Ray, K. & Mazumdar, S. The protein inhibitor of nNOS (PIN/DLC1/LC8) binding does not inhibit the NADPH-dependent heme reduction in nNOS, a key step in NO synthesis. *Biochem. Biophys. Res. Commun.* **472,** 189–193 (2016).
4. King, S. M. & Patel-King, R. S. The M(r) = 8,000 and 11,000 outer arm dynein light chains from Chlamydomonas flagella have cytoplasmic homologues. *J. Biol. Chem.* **270,** 11445–52 (1995).
5. Wickstead, B. & Gull, K. Dyneins across eukaryotes: A comparative genomic analysis. *Traffic* **8,** 1708–1721 (2007).
6. Pfister, K. K., Shah, P. R., Hummerich, H., Russ, A., Cotton, J., Annuar, A. A., King, S. M. & Fisher, E. M. C. Genetic Analysis of the Cytoplasmic Dynein Subunit Families. *PLoS Genet.* **2,** e1 (2006).
7. Espindola, F. S., Suter, D. M., Partata, L. B. E., Cao, T., Wolenski, J. S., Cheney, R. E., King, S. M. & Mooseker, M. S. The light chain composition of chicken brain myosin-Va: Calmodulin, myosin-II essential light chains, and 8-kDa dynein light chain/PIN. *Cell Motil. Cytoskeleton* **47,** 269–281 (2000).
8. Lee, K. H., Lee, S., Kim, B., Chang, S., Kim, S. W., Paick, J.-S. & Rhee, K. Dazl can bind to dynein motor complex and may play a role in transport of specific mRNAs. *EMBO J.* **25,** 4263–4270 (2006).
9. Lo, K. W. H., Kan, H. M., Chan, L. N., Xu, W. G., Wang, K. P., Wu, Z., Sheng, M. & Zhang, M. The 8-kDa dynein light chain binds to p53-binding protein 1 and mediates DNA damage-induced p53 nuclear accumulation. *J. Biol. Chem.* **280,** 8172–8179 (2005).
10. Puthalakath, H., Huang, D. C. ., O'Reilly, L. A., King, S. M. & Strasser, A. The Proapoptotic Activity of the Bcl-2 Family Member Bim Is Regulated by Interaction with the Dynein Motor Complex. *Mol. Cell* **3,** 287–296 (1999).
11. Navarro, C., Puthalakath, H., Adams, J. M., Strasser, A. & Lehmann, R. Egalitarian binds dynein light chain to establish oocyte polarity and maintain oocyte fate. *Nat. Cell Biol.* **6,**

427–435 (2004).

12. Rodríguez-Crespo, I., Yélamos, B., Roncal, F., Albar, J. P., Ortiz de Montellano, P. R. & Gavilanes, F. Identification of novel cellular proteins that bind to the LC8 dynein light chain using a pepscan technique. *FEBS Lett.* **503,** 135–141 (2001).
13. Schnorrer, F., Bohmann, K. & Nüsslein-Volhard, C. The molecular motor dynein is involved in targeting Swallow and bicoid RNA to the anterior pole of Drosophila oocytes. *Nat. Cell Biol.* **2,** 185–190 (2000).
14. Hays, T. & Karess, R. Swallowing dynein: a missing link in RNA localization? *Nat. Cell Biol.* **2,** E60–E62 (2000).
15. Barbar, E., Kleinman, B., Imhoff, D., Li, M., Hays, T. S. & Hare, M. Dimerization and folding of LC8, a highly conserved light chain of cytoplasmic dynein. *Biochemistry* **40,** 1596–605 (2001).
16. Benison, G. & Barbar, E. in *Methods Enzymol.* **455,** 237–258 (2009).
17. Nyarko, A., Hare, M., Hays, T. S. & Barbar, E. The intermediate chain of cytoplasmic dynein is partially disordered and gains structure upon binding to light-chain LC8. *Biochemistry* **43,** 15595–15603 (2004).
18. King, S. M., Barbarese, E., Dillman, J. F., Patel-King, R. S., Carson, J. H. & Pfister, K. K. Brain Cytoplasmic and Flagellar Outer Arm Dyneins Share a Highly Conserved M_r 8,000 Light Chain. *J. Biol. Chem.* **271,** 19358–19366 (1996).
19. Williams, J. C., Roulhac, P. L., Roy, A. G., Vallee, R. B., Fitzgerald, M. C. & Hendrickson, W. A. Structural and thermodynamic characterization of a cytoplasmic dynein light chain-intermediate chain complex. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 10028–10033 (2007).
20. Benison, G., Karplus, P. A. & Barbar, E. Structure and Dynamics of LC8 Complexes with KXTQT-Motif Peptides: Swallow and Dynein Intermediate Chain Compete for a Common Site. *J. Mol. Biol.* **371,** 457–468 (2007).
21. Bergen, J. M. & Pun, S. H. Evaluation of an LC8-Binding Peptide for the Attachment of Artificial Cargo to Dynein. *Mol. Pharm.* **4,** 119–128 (2007).
22. Jespersen, N., Estelle, A., Waugh, N., Davey, N. E., Blikstad, C., Ammon, Y.-C., Akhmanova, A., Ivarsson, Y., Hendrix, D. A. & Barbar, E. Systematic identification of recognition motifs for the hub protein LC8. *Life Sci. alliance* **2,** e201900366 (2019).
23. Gaik, M., Flemming, D., von Appen, A., Kastritis, P., Mücke, N., Fischer, J., Stelter, P., Ori, A., Bui, K. H., Baßler, J., Barbar, E., Beck, M. & Hurt, E. Structural basis for

assembly and function of the Nup82 complex in the nuclear pore scaffold. *J. Cell Biol.* **208,** 283–97 (2015).

24. Wang, L., Hare, M., Hays, T. S. & Barbar, E. Dynein light chain LC8 promotes assembly of the coiled-coil domain of swallow protein. *Biochemistry* **43,** 4611–4620 (2004).
25. Benison, G., Nyarko, A. & Barbar, E. Heteronuclear NMR Identifies a Nascent Helix in Intrinsically Disordered Dynein Intermediate Chain: Implications for Folding and Dimerization. *J. Mol. Biol.* **362,** 1082–1093 (2006).
26. Kraemer, D. M., Strambio-de-Castillia, C., Blobel, G. & Rout, M. P. The Essential Yeast Nucleoporin NUP159 Is Located on the Cytoplasmic Side of the Nuclear Pore Complex and Serves in Karyopherin-mediated Binding of Transport Substrate. *J. Biol. Chem.* **270,** 19017–19021 (1995).
27. Gleizes, P.-E., Noaillac-Depeyre, J., Léger-Silvestre, I., Teulières, F., Dauxois, J.-Y., Pommet, D., Azum-Gelade, M.-C. & Gas, N. Ultrastructural localization of rRNA shows defective nuclear export of preribosomes in mutants of the Nup82p complex. *J. Cell Biol.* **155,** 923–936 (2001).
28. Fernandez-Martinez, J., Kim, S. J., Shi, Y., Upla, P., Pellarin, R., Gagnon, M., Chemmama, I. E., Wang, J., Nudelman, I., Zhang, W., Williams, R., Rice, W. J., Stokes, D. L., Zenklusen, D., Chait, B. T., Sali, A. & Rout, M. P. Structure and Function of the Nuclear Pore Complex Cytoplasmic mRNA Export Platform. *Cell* **167,** 1215-1228.e25 (2016).
29. Stelter, P., Kunze, R., Flemming, D., Höpfner, D., Diepholz, M., Philippsen, P., Böttcher, B. & Hurt, E. Molecular basis for the functional interaction of dynein light chain with the nuclear-pore complex. *Nat. Cell Biol.* **9,** 788–796 (2007).
30. Lemke, E. A. The Multiple Faces of Disordered Nucleoporins. *J. Mol. Biol.* **428,** 2011–2024 (2016).
31. Rapali, P., Szenes, Á., Radnai, L., Bakos, A., Pál, G. & Nyitray, L. DYNLL/LC8: a light chain subunit of the dynein motor complex and beyond. *FEBS J.* **278,** 2980–2996 (2011).
32. Barbar, E. Dynein Light Chain LC8 Is a Dimerization Hub Essential in Diverse Protein Networks†. *Biochemistry* **47,** 503–508 (2008).
33. Kreko-Pierce, T. & Eaton, B. A. The Drosophila LC8 homolog cut up specifies the axonal transport of proteasomes. *J. Cell Sci.* **130,** 3388–3398 (2017).
34. Theerawatanasirikul, S., Phecharat, N., Prawettongsopon, C., Chaicumpa, W. & Lekcharoensuk, P. Dynein light chain DYNLL1 subunit facilitates porcine circovirus type

2 intracellular transports along microtubules. *Arch. Virol.* **162,** 677–686 (2017).

35. Clardy, J., Liang, J., Jaffrey, S. R., Guo, W. & Snyder, S. H. Structure of the PIN/LC8 dimer with a bound peptide. *Nat. Struct. Biol.* **6,** 735–740 (1999).
36. Bodor, A., Radnai, L., Hetényi, C., Rapali, P., Láng, A., Kövér, K. E., Perczel, A., Wahlgren, W. Y., Katona, G. & Nyitray, L. DYNLL2 Dynein Light Chain Binds to an Extended Linear Motif of Myosin 5a Tail That Has Structural Plasticity. *Biochemistry* **53,** 7107–7122 (2014).
37. Slevin, L. K., Romes, E. M., Dandulakis, M. G. & Slep, K. C. The mechanism of dynein light chain LC8-mediated oligomerization of the Ana2 centriole duplication factor. *J. Biol. Chem.* **289,** 20727–20739 (2014).
38. Rapali, P., Radnai, L., Süveges, D., Harmat, V., Tölgyesi, F., Wahlgren, W. Y., Katona, G., Nyitray, L. & Pál, G. Directed evolution reveals the binding motif preference of the LC8/DYNLL hub protein and predicts large numbers of novel binders in the human proteome. *PLoS One* **6,** (2011).
39. Regué, L., Sdelci, S., Bertran, M. T., Caelles, C., Reverter, D. & Roig, J. DYNLL/LC8 protein controls signal transduction through the Nek9/Nek6 signaling module by regulating Nek6 binding to Nek9. *J. Biol. Chem.* **286,** 18118–18129 (2011).
40. Lightcap, C. M., Kari, G., Arias-Romero, L. E., Chernoff, J., Rodeck, U. & Williams, J. C. Interaction with LC8 Is Required for Pak1 Nuclear Import and Is Indispensable for Zebrafish Development. *PLoS One* **4,** e6025 (2009).
41. Rao, L., Romes, E. M., Nicholas, M. P., Brenner, S., Tripathy, A., Gennerich, A. & Slep, K. C. The yeast dynein Dyn2-Pac11 complex is a dynein dimerization/processivity factor: structural and single-molecule characterization. *Mol. Biol. Cell* **24,** 2362–2377 (2013).
42. Benison, G., Karplus, P. A. & Barbar, E. The Interplay of Ligand Binding and Quaternary Structure in the Diverse Interactions of Dynein Light Chain LC8. *J. Mol. Biol.* **384,** 954–966 (2008).
43. Myllykoski, M., Eichel, M. A., Jung, R. B., Kelm, S., Werner, H. B. & Kursula, P. High-affinity heterotetramer formation between the large myelin-associated glycoprotein and the dynein light chain DYNLL1. *J. Neurochem.* **147,** 764–783 (2018).
44. Clark, S., Nyarko, A., Löhr, F., Karplus, P. A. & Barbar, E. The Anchored Flexibility Model in LC8 Motif Recognition: Insights from the Chica Complex. *Biochemistry* **55,** 199–209 (2016).
45. Song, Y., Benison, G., Nyarko, A., Hays, T. S. & Barbar, E. Potential Role for

Phosphorylation in Differential Regulation of the Assembly of Dynein Light Chains. *J. Biol. Chem.* **282,** 17272–17279 (2007).

46. Clark, S. A., Jespersen, N., Woodward, C. & Barbar, E. Multivalent IDP assemblies: Unique properties of LC8-associated, IDP duplex scaffolds. *FEBS Lett.* **589,** 2543–2551 (2015).
47. Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C. & Obradovic, Z. Intrinsically disordered protein. *J. Mol. Graph. Model.* **19,** 26–59 (2001).
48. Fukuchi, S., Hosoda, K., Homma, K., Gojobori, T. & Nishikawa, K. Binary classification of protein molecules into intrinsically disordered and ordered segments. *BMC Struct. Biol.* **11,** 29 (2011).
49. Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. Prediction and Functional Analysis of Native Disorder in Proteins from the Three Kingdoms of Life. *J. Mol. Biol.* **337,** 635–645 (2004).
50. Peng, Z., Yan, J., Fan, X., Mizianty, M. J., Xue, B., Wang, K., Hu, G., Uversky, V. N. & Kurgan, L. Exceptionally abundant exceptions: comprehensive characterization of intrinsic disorder in all domains of life. *Cell. Mol. Life Sci.* **72,** 137–151 (2015).
51. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradović, Z. Intrinsic disorder and protein function. *Biochemistry* **41,** 6573–6582 (2002).
52. van der Lee, R., Buljan, M., Lang, B., Weatheritt, R. J., Daughdrill, G. W., Dunker, A. K., Fuxreiter, M., Gough, J., Gsponer, J., Jones, D. T., Kim, P. M., Kriwacki, R. W., Oldfield, C. J., Pappu, R. V., Tompa, P., Uversky, V. N., Wright, P. E. & Babu, M. M. Classification of intrinsically disordered regions and proteins. *Chem. Rev.* **114,** 6589–6631 (2014).
53. Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I. & Wright, P. E. Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci.* **93,** 11504–11509 (1996).
54. Brown, C. J., Johnson, A. K., Dunker, A. K. & Daughdrill, G. W. Evolution and disorder. *Curr. Opin. Struct. Biol.* **21,** 441–446 (2011).
55. Tantos, A., Han, K.-H. & Tompa, P. Intrinsic disorder in cell signaling and gene transcription. *Mol. Cell. Endocrinol.* **348,** 457–465 (2012).

56. Anurag, M., Singh, G. P. & Dash, D. Location of disorder in coiled coilproteins is influenced by its biological role and subcellular localization: a GO-based study on human proteome. *Mol. BioSyst.* **8,** 346–352 (2012).
57. Babu, M. M., van der Lee, R., de Groot, N. S. & Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* **21,** 432–440 (2011).
58. Uversky, V. N., Oldfield, C. J. & Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D ^{2} Concept. *Annu. Rev. Biophys.* **37,** 215–246 (2008).
59. Howlett, D. R., Jennings, K. H., Lee, D. C., Clark, M. S., Brown, F., Wetzel, R., Wood, S. J., Camilleri, P. & Roberts, G. W. Aggregation state and neurotoxic properties of Alzheimer beta-amyloid peptide. *Neurodegeneration* **4,** 23–32 (1995).
60. Iqbal, K., del C. Alonso, A., Chen, S., Chohan, M. O., El-Akkad, E., Gong, C.-X., Khatoon, S., Li, B., Liu, F., Rahman, A., Tanimukai, H. & Grundke-Iqbal, I. Tau pathology in Alzheimer disease and other tauopathies. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1739,** 198–210 (2005).
61. Alonso, A. C., Grundke-Iqbal, I. & Iqbal, K. Alzheimer's disease hyperphosphorylated tau sequesters normal tau into tangles of filaments and disassembles microtubules. *Nat. Med.* **2,** 783–787 (1996).
62. Mark, W.-Y., Liao, J. C. C., Lu, Y., Ayed, A., Laister, R., Szymczyna, B., Chakrabartty, A. & Arrowsmith, C. H. Characterization of Segments from the Central Region of BRCA1: An Intrinsically Disordered Scaffold for Multiple Protein–Protein and Protein–DNA Interactions? *J. Mol. Biol.* **345,** 275–287 (2005).
63. Makokha, M., Hare, M., Li, M., Hays, T. & Barbar, E. Interactions of cytoplasmic dynein light chains Tctex-1 and LC8 with the intermediate chain IC74. *Biochemistry* **41,** 4302–4311 (2002).
64. Nyarko, A. & Barbar, E. Light chain-dependent self-association of dynein intermediate chain. *J. Biol. Chem.* **286,** 1556–1566 (2011).
65. Nyarko, A., Song, Y., Nováček, J., Žídek, L. & Barbar, E. Multiple recognition motifs in nucleoporin Nup159 provide a stable and rigid Nup159-Dyn2 assembly. *J. Biol. Chem.* **288,** 2614–2622 (2013).
66. Raux, H., Flamand, A. & Blondel, D. Interaction of the rabies virus P protein with the LC8 dynein light chain. *J. Virol.* **74,** 10212–10216 (2000).
67. Zaytseva, O., Tenis, N., Mitchell, N., Kanno, S. I., Yasui, A., Heierhorst, J. & Quinn, L. M. The novel zinc finger protein dASCIZ regulates mitosis in Drosophila via an essential

role in dynein light-chain expression. *Genetics* **196,** 443–453 (2014).

68. Jurado, S., Conlan, L. A., Baker, E. K., Ng, J. L., Tenis, N., Hoch, N. C., Gleeson, K., Smeets, M., Izon, D. & Heierhorst, J. ATM substrate Chk2-interacting Zn 2+ finger (ASCIZ) is a bi-functional transcriptional activator and feedback sensor in the regulation of dynein light chain (DYNLL1) expression. *J. Biol. Chem.* **287,** 3156–3164 (2012).
69. Clark, S., Myers, J. B., King, A., Fiala, R., Novacek, J., Pearce, G., Heierhorst, J., Reichow, S. L. & Barbar, E. J. Multivalency regulates activity in an intrinsically disordered transcription factor. *Elife* **7,** (2018).
70. Petryszak, R., Keays, M., Tang, Y. A., Fonseca, N. A., Barrera, E., Burdett, T., Füllgrabe, A., Fuentes, A. M.-P., Jupp, S., Koskinen, S., Mannion, O., Huerta, L., Megy, K., Snow, C., Williams, E., Barzine, M., Hastings, E., Weisser, H., Wright, J., Jaiswal, P., Huber, W., Choudhary, J., Parkinson, H. E. & Brazma, A. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res.* **44,** D746–D752 (2016).
71. Chen, Y.-M., Gerwin, C. & Sheng, Z.-H. Dynein Light Chain LC8 Regulates Syntaphilin-Mediated Mitochondrial Docking in Axons. *J. Neurosci.* **29,** 9429–9438 (2009).
72. King, A., Hoch, N. C., McGregor, N. E., Sims, N. A., Smyth, I. M. & Heierhorst, J. Dynll1 is essential for development and promotes endochondral bone formation by regulating intraflagellar dynein function in primary cilia. *Hum. Mol. Genet.* **28,** 2573–2588 (2019).
73. Dick, T., Ray, K., Salz, H. K. & Chia, W. Cytoplasmic dynein (ddlc1) mutations cause morphogenetic defects and apoptotic cell death in Drosophila melanogaster. *Mol. Cell. Biol.* **16,** 1966–1977 (1996).
74. Kamath, R. S., Fraser, A. G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D. P., Zipperlen, P. & Ahringer, J. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature* **421,** 231–237 (2003).
75. Erdős, G., Szaniszló, T., Pajkos, M., Hajdu-Soltész, B., Kiss, B., Pál, G., Nyitray, L. & Dosztányi, Z. Novel linear motif filtering protocol reveals the role of the LC8 dynein light chain in the Hippo pathway. *PLoS Comput. Biol.* **13,** e1005885 (2017).
76. Jeong, H., Mason, S. P., Barabási, A.-L. & Oltvai, Z. N. Lethality and centrality in protein networks. *Nature* **411,** 41–42 (2001).
77. Komurov, K. & White, M. Revealing static and dynamic modular architecture of the

eukaryotic protein interaction network. *Mol. Syst. Biol.* **3,** 110 (2007).

78. Patil, A., Kinoshita, K. & Nakamura, H. Hub promiscuity in protein-protein interaction networks. *Int. J. Mol. Sci.* **11,** 1930–1943 (2010).
79. Wu, X., Guo, J., Zhang, D.-Y. & Lin, K. The properties of hub proteins in a yeast-aggregated cell cycle network and its phase sub-networks. *Proteomics* **9,** 4812–4824 (2009).
80. Aitken, A. 14-3-3 proteins: A historic overview. *Semin. Cancer Biol.* **16,** 162–172 (2006).
81. Uhart, M., Flores, G. & Bustos, D. M. Controllability of protein-protein interaction phosphorylation-based networks: Participation of the hub 14-3-3 protein family. *Sci. Rep.* **6,** 26234 (2016).
82. Uchikoga, N., Matsuzaki, Y., Ohue, M. & Akiyama, Y. Specificity of broad protein interaction surfaces for proteins with multiple binding partners. *Biophys. physicobiology* **13,** 105–115 (2016).
83. Hall, J., Song, Y., Karplus, P. A. & Barbar, E. The Crystal Structure of Dynein Intermediate Chain-Light Chain Roadblock Complex Gives New Insights into Dynein Assembly. *J. Biol. Chem.* **285,** 22566–22575 (2010).
84. Kidane, A. I., Song, Y., Nyarko, A., Hall, J., Hare, M., Löhr, F. & Barbar, E. Structural features of LC8-induced self-association of swallow. *Biochemistry* **52,** 6011–6020 (2013).
85. Asthana, J., Kuchibhatla, A., Jana, S. C., Ray, K. & Panda, D. Dynein light chain 1 (LC8) association enhances microtubule stability and promotes microtubule bundling. *J. Biol. Chem.* **287,** 40793–40805 (2012).
86. Kirkham, J. K., Park, S. H., Nguyen, T. N., Lee, J. H. & Günzl, A. The dynein light chain LC8 is required for RNA polymerase I-mediated transcription in Trypanosoma brucei, facilitating assembly and promoter binding of class I transcription factor A. *Mol. Cell. Biol.* **36,** MCB.00705-00715 (2015).
87. Jacob, Y., Badrane, H., Ceccaldi, P. E. & Tordo, N. Cytoplasmic dynein LC8 interacts with lyssavirus phosphoprotein. *J. Virol.* **74,** 10217–22 (2000).
88. Jayappa, K. D., Ao, Z., Wang, X., Mouland, A. J., Shekhar, S., Yang, X. & Yao, X. Human immunodeficiency virus type 1 employs the cellular dynein light chain 1 protein for reverse transcription through interaction with its integrase protein. *J. Virol.* **89,** 3497–3511 (2015).
89. Alonso, C., Miskin, J., Hernáez, B., Fernandez-Zapatero, P., Soto, L., Cantó, C., Rodríguez-Crespo, I., Dixon, L. & Escribano, J. M. African swine fever virus protein p54

interacts with the microtubular motor complex through direct binding to light-chain dynein. *J. Virol.* **75,** 9819–9827 (2001).

90. Luthra, P., Jordan, D. S., Leung, D. W., Amarasinghe, G. K. & Basler, C. F. Ebola virus VP35 interaction with dynein LC8 regulates viral RNA synthesis. *J. Virol.* **89,** 5148–53 (2015).
91. Schneider, M. A., Spoden, G. A., Florin, L. & Lambert, C. Identification of the dynein light chains required for human papillomavirus infection. *Cell. Microbiol.* **13,** 32–46 (2011).
92. Tan, G. S., Preuss, M. A. R., Williams, J. C. & Schnell, M. J. The dynein light chain 8 binding motif of rabies virus phosphoprotein promotes efficient viral transcription. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 7229–7234 (2007).
93. Rapali, P., García-Mayoral, M. F., Martínez-Moreno, M., Tárnok, K., Schlett, K., Albar, J. P., Bruix, M., Nyitray, L. & Rodriguez-Crespo, I. LC8 dynein light chain (DYNLL1) binds to the C-terminal domain of ATM-interacting protein (ATMIN/ASCIZ) and regulates its subcellular localization. *Biochem. Biophys. Res. Commun.* **414,** 493–498 (2011).
94. Jurado, S., Smyth, I., van Denderen, B., Tenis, N., Hammet, A., Hewitt, K., Ng, J.-L., McNees, C. J., Kozlov, S. V., Oka, H., Kobayashi, M., Conlan, L. A., Cole, T. J., Yamamoto, K., Taniguchi, Y., Takeda, S., Lavin, M. F. & Heierhorst, J. Dual Functions of ASCIZ in the DNA Base Damage Response and Pulmonary Organogenesis. *PLoS Genet.* **6,** e1001170 (2010).
95. Kienzle, T. E. *Rabies*. (Chelsea House, 2007).
96. Mørk, T. & Prestrud, P. Arctic rabies--a review. *Acta Vet. Scand.* **45,** 1–9 (2004).
97. Gómez-Alonso, J. Rabies: a possible explanation for the vampire legend. *Neurology* **51,** 856–859 (1998).
98. ANDERSON, L. J., Nicholson, K. G., Tauxe, R. V & Winkler, W. G. Human Rabies in the United States, 1960 to 1979: Epidemiology, Diagnosis, and Prevention. *Ann. Intern. Med.* **100,** 728 (1984).
99. Wasik, B. & Murphy, M. *Rabid: A Cultural History of the World's Most Diabolical Virus*. (Viking Press, 2012).
100. Davis, B. M., Rall, G. F. & Schnell, M. J. Everything You Always Wanted to Know About Rabies Virus (But Were Afraid to Ask). *Annu. Rev. Virol.* **2,** 451–471 (2015).
101. Lafon, M. Rabies virus receptors. *J. Neurovirol.* **11,** 82–87 (2005).

102. Albuquerque, E. X., Pereira, E. F. R., Alkondon, M. & Rogers, S. W. Mammalian nicotinic acetylcholine receptors: from structure to function. *Physiol. Rev.* **89,** 73–120 (2009).

103. Ugolini, G. Use of rabies virus as a transneuronal tracer of neuronal connections: implications for the understanding of rabies pathogenesis. *Dev. Biol. (Basel).* **131,** 493–506 (2008).

104. Macedo, C. I., Carnieli Jr, P., Brandão, P. E., Rosa, E. S. T. da, Oliveira, R. de N., Castilho, J. G., Medeiros, R., Machado, R. R., Oliveira, R. C. de, Carrieri, M. L. & Kotait, I. Diagnosis of human rabies cases by polymerase chain reaction of neck-skin samples. *Brazilian J. Infect. Dis.* **10,** 341–345 (2006).

105. Jogai, S., Radotra, B. D. & Banerjee, A. K. Rabies viral antigen in extracranial organs: a post-mortem study. *Neuropathol. Appl. Neurobiol.* **28,** 334–338 (2002).

106. Jackson, A. C., Ye, H., Phelan, C. C., Ridaura-Sanz, C., Zheng, Q., Li, Z., Wan, X. & Lopez-Corella, E. Extraneural organ involvement in human rabies. *Lab. Invest.* **79,** 945–51 (1999).

107. Despond, O., Tucci, M., Decaluwe, H., Grégoire, M.-C., S Teitelbaum, J. & Turgeon, N. Rabies in a nine-year-old child: The myth of the bite. *Can. J. Infect. Dis.* **13,** 121–5 (2002).

108. Shankar, S. K., Mahadevan, A., Sapico, S. D., Ghodkirekar, M. S. G., Pinto, R. G. W. & Madhusudana, S. N. Rabies viral encephalitis with proable 25 year incubation period! *Ann. Indian Acad. Neurol.* **15,** 221–223 (2012).

109. Hooper, D. C., Roy, A., Kean, R. B., Phares, T. W. & Barkhouse, D. A. Therapeutic immune clearance of rabies virus from the CNS. *Future Virol.* **6,** 387–397 (2011).

110. Mahadevan, A., Suja, M. S., Mani, R. S. & Shankar, S. K. Perspectives in Diagnosis and Treatment of Rabies Viral Encephalitis: Insights from Pathogenesis. *Neurotherapeutics* **13,** 477–492 (2016).

111. Manesh, A., Mani, R. S., Pichamuthu, K., Jagannati, M., Mathew, V., Karthik, R., Abraham, O. C., Chacko, G. & Varghese, G. M. Case Report: Failure of Therapeutic Coma in Rabies Encephalitis. *Am. J. Trop. Med. Hyg.* **98,** 207–210 (2018).

112. Pathak, S., Horton, D. L., Lucas, S., Brown, D., Quaderi, S., Polhill, S., Walker, D., Nastouli, E., Núñez, A., Wise, E. L., Fooks, A. R. & Brown, M. Diagnosis, management and post-mortem findings of a human case of rabies imported into the United Kingdom from India: a case report. *Virol. J.* **11,** 63 (2014).

113. Brzózka, K., Finke, S. & Conzelmann, K.-K. Inhibition of interferon signaling by rabies virus phosphoprotein P: activation-dependent binding of STAT1 and STAT2. *J. Virol.* **80,** 2675–2683 (2006).

114. Albertini, A. A. V, Ruigrok, R. W. H. & Blondel, D. Rabies virus transcription and replication. *Adv. Virus Res.* **79,** 1–22 (2011).

115. Finke, S., Mueller-Waldeck, R. & Conzelmann, K.-K. Rabies virus matrix protein regulates the balance of virus transcription and replication. *J. Gen. Virol.* **84,** 1613–1621 (2003).

116. Jacob, Y., Real, E. & Tordo, N. Functional interaction map of lyssavirus phosphoprotein: identification of the minimal transcription domains. *J. Virol.* **75,** 9613–9622 (2001).

117. Bauer, A., Nolden, T., Nemitz, S., Perlson, E. & Finke, S. A Dynein Light Chain 1 Binding Motif in Rabies Virus Polymerase L Protein Plays a Role in Microtubule Reorganization and Viral Primary Transcription. *J. Virol.* **89,** 9591–9600 (2015).

118. Loening, N., Saravanan, S., Jespersen, N. & Barbar, El. A nascent helix in dynein intermediate chain directly interacts with dynactin p150Glued in the initiation of stable dynein/dynactin complexes. *J. Biol. Chem.* (2019).

119. Kubota, T., Matsuoka, M., Chang, T.-H., Bray, M., Jones, S., Tashiro, M., Kato, A. & Ozato, K. Ebolavirus VP35 Interacts with the Cytoplasmic Dynein Light Chain 8. *J. Virol.* **83,** 6952–6956 (2009).

120. Chenik, M., Chebli, K., Gaudin, Y. & Blondel, D. In vivo interaction of rabies virus phosphoprotein (P) and nucleoprotein (N): existence of two N-binding sites on P protein. *J. Gen. Virol.* **75,** 2889–2896 (1994).

121. Chenik, M., Schnell, M., Conzelmann, K. K. & Blondel, D. Mapping the interacting domains between the rabies virus polymerase and phosphoprotein. *J. Virol.* **72,** 1925–1930 (1998).

122. Vidy, A., Chelbi-Alix, M. & Blondel, D. Rabies virus P protein interacts with STAT1 and inhibits interferon signal transduction pathways. *J. Virol.* **79,** 14411–20 (2005).

123. Albertini, A. A. V, Wernimont, A. K., Muziol, T., Ravelli, R. B. G., Clapier, C. R., Schoehn, G., Weissenhorn, W. & Ruigrok, R. W. H. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* **313,** 360–363 (2006).

124. Ivanov, I., Crépin, T., Jamin, M. & Ruigrok, R. W. H. Structure of the dimerization domain of the rabies virus phosphoprotein. *J. Virol.* **84,** 3707–3710 (2010).

125. Leyrat, C., Schneider, R., Ribeiro, E. A., Yabukarski, F., Yao, M., G??rard, F. C. A.,

Jensen, M. R., Ruigrok, R. W. H., Blackledge, M. & Jamin, M. Ensemble structure of the modular and flexible full-length vesicular stomatitis virus phosphoprotein. *J. Mol. Biol.* **423,** 182–197 (2012).

126. Poisson, N., Real, E., Gaudin, Y., Vaney, M.-C. C., King, S., Jacob, Y., Tordo, N. & Blondel, D. Molecular basis for the interaction between rabies virus phosphoprotein P and the dynein light chain LC8: Dissociation of dynein-binding properties and transcriptional functionality of P. *J. Gen. Virol.* **82,** 2691–2696 (2001).
127. Nikolic, J., Le Bars, R., Lama, Z., Scrima, N., Lagaudrière-Gesbert, C., Gaudin, Y. & Blondel, D. Negri bodies are viral factories with properties of liquid organelles. *Nat. Commun.* **8,** 58 (2017).
128. Wiltzer, L., Okada, K., Yamaoka, S., Larrous, F., Kuusisto, H. V., Sugiyama, M., Blondel, D., Bourhy, H., Jans, D. A., Ito, N. & Moseley, G. W. Interaction of rabies virus P-protein with STAT proteins is critical to lethal rabies disease. *J. Infect. Dis.* **209,** 1744–1753 (2014).
129. Mavrakis, M., McCarthy, A. A., Roche, S., Blondel, D. & Ruigrok, R. W. H. Structure and function of the C-terminal domain of the polymerase cofactor of rabies virus. *J. Mol. Biol.* **343,** 819–831 (2004).
130. Gerard, F. C. A., Ribeiro, E. de A., Leyrat, C., Ivanov, I., Blondel, D., Longhi, S., Ruigrok, R. W. H. & Jamin, M. Modular Organization of Rabies Virus Phosphoprotein. *J. Mol. Biol.* **388,** 978–996 (2009).
131. Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292,** 195–202 (1999).
132. Yabukarski, F., Leyrat, C., Martinez, N., Communie, G., Ivanov, I., Ribeiro, E. A., Buisson, M., Gerard, F. C., Bourhis, J. M., Jensen, M. R., Bernad??, P., Blackledge, M. & Jamin, M. Ensemble Structure of the Highly Flexible Complex Formed between Vesicular Stomatitis Virus Unassembled Nucleoprotein and its Phosphoprotein Chaperone. *J. Mol. Biol.* **428,** 2671–2694 (2016).
133. Muñoz, V. & Serrano, L. Elucidating the folding problem of helical peptides using empirical parameters. *Nat. Struct. Mol. Biol.* **1,** 399–409 (1994).
134. Frueh, D. P. Practical aspects of NMR signal assignment in larger and challenging proteins. *Prog. Nucl. Magn. Reson. Spectrosc.* **78,** 47–75 (2014).
135. Farrow, N., Zhang, O., Szabo, A., Torchia, D. & Kay, L. Spectral density function mapping using 15N relaxation data exclusively. *J. Biomol. NMR* **6,** 153–162 (1995).

136. Hwang, T. L., van Zijl, P. C. M. & Mori, S. Accurate quantitation of water-amide proton exchange rates using the Phase-Modulated CLEAN chemical EXchange (CLEANEX-PM) approach with a Fast-HSQC (FHSQC) detection scheme RID B-8680-2008. *J. Biomol. NMR* **11,** 221–226 (1998).

137. Brereton, A. E. & Karplus, P. A. Ensemblator v3: Robust atom-level comparative analyses and classification of protein structure ensembles. *Protein Sci.* **27,** 41–50 (2018).

138. Chelbi-Alix, M. K., Vidy, A., El Bougrini, J. & Blondel, D. Rabies viral mechanisms to escape the IFN system: the viral protein P interferes with IRF-3, Stat1, and PML nuclear bodies. *J. Interferon Cytokine Res.* **26,** 271–280 (2006).

139. Fouquet, B., Nikolic, J., Larrous, F., Bourhy, H., Wirblich, C., Lagaudrière-Gesbert, C. & Blondel, D. Focal Adhesion Kinase Is Involved in Rabies Virus Infection through Its Interaction with Viral Phosphoprotein P. *J. Virol.* **89,** 1640–1651 (2015).

140. Li, Y., Dong, W., Shi, Y., Deng, F., Chen, X., Wan, C., Zhou, M., Zhao, L., Fu, Z. F. & Peng, G. Rabies virus phosphoprotein interacts with ribosomal protein L9 and affects rabies virus replication. *Virology* **488,** 216–224 (2016).

141. Oksayan, S., Wiltzer, L., Rowe, C. L., Blondel, D., Jans, D. A. & Moseley, G. W. A novel nuclear trafficking module regulates the nucleocytoplasmic localization of the rabies virus interferon antagonist, P protein. *J. Biol. Chem.* **287,** 28112–28121 (2012).

142. Liu, J., Wang, H., Gu, J., Deng, T., Yuan, Z., Hu, B., Xu, Y., Yan, Y., Zan, J., Liao, M., DiCaprio, E., Li, J., Su, S. & Zhou, J. BECN1-dependent CASP2 incomplete autophagy induction by binding to rabies virus phosphoprotein. *Autophagy* 00–00 (2017). doi:10.1080/15548627.2017.1280220

143. Kammouni, W., Wood, H., Saleh, A., Appolinario, C. M., Fernyhough, P. & Jackson, A. C. Rabies virus phosphoprotein interacts with mitochondrial Complex I and induces mitochondrial dysfunction and oxidative stress. *J. Neurovirol.* **21,** 370–382 (2015).

144. Blondel, D., Regad, T., Poisson, N., Pavie, B., Harper, F., Pandolfi, P. P., De Thé, H. & Chelbi-Alix, M. K. Rabies virus P and small P products interact directly with PML and reorganize PML nuclear bodies. *Oncogene* **21,** 7957–7970 (2002).

145. Xu, Y., Liu, F., Liu, J., Wang, D., Yan, Y., Ji, S., Zan, J. & Zhou, J. The co-chaperone Cdc37 regulates the rabies virus phosphoprotein stability by targeting to Hsp90AA1 machinery. *Sci. Rep.* **6,** 27123 (2016).

146. Curran, J. A role for the Sendai virus P protein trimer in RNA synthesis. *J. Virol.* **72,** 4274–4280 (1998).

147. Ribeiro, E. de A., Leyrat, C., Gérard, F. C. A. A., Albertini, A. A. V. V, Falk, C., Ruigrok, R. W. H. H. & Jamin, M. Binding of rabies virus polymerase cofactor to recombinant circular nucleoprotein-RNA complexes. *J. Mol. Biol.* **394,** 558–575 (2009).

148. Takamatsu, Y., Kolesnikova, L. & Becker, S. Ebola virus proteins NP, VP35, and VP24 are essential and sufficient to mediate nucleocapsid transport. *Proc. Natl. Acad. Sci. U. S. A.* **115,** 1075–1080 (2018).

149. Becker, J. R., Cuella-Martin, R., Barazas, M., Liu, R., Oliveira, C., Oliver, A. W., Bilham, K., Holt, A. B., Blackford, A. N., Heierhorst, J., Jonkers, J., Rottenberg, S. & Chapman, J. R. The ASCIZ-DYNLL1 axis promotes 53BP1-dependent non-homologous end joining and PARP inhibitor sensitivity. *Nat. Commun.* **9,** 5406 (2018).

150. Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics Protoc. Handb.* 571–607 (2005). doi:10.1385/1-59259-890-0:571

151. Poser, I., Sarov, M., Hutchins, J. R. A., Hériché, J.-K., Toyoda, Y., Pozniakovsky, A., Weigl, D., Nitzsche, A., Hegemann, B., Bird, A. W., Pelletier, L., Kittler, R., Hua, S., Naumann, R., Augsburg, M., Sykora, M. M., Hofemeister, H., Zhang, Y., Nasmyth, K., White, K. P., Dietzel, S., Mechtler, K., Durbin, R., Stewart, A. F., Peters, J.-M., Buchholz, F. & Hyman, A. A. BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat. Methods* **5,** 409–415 (2008).

152. Finke, S. & Conzelmann, K. K. Virus promoters determine interference by defective RNAs: selective amplification of mini-RNA vectors and rescue from cDNA by a 3' copy-back ambisense rabies virus. *J. Virol.* **73,** 3818–3825 (1999).

153. Le Mercier, P., Jacob, Y., Tanner, K. & Tordo, N. A novel expression cassette of lyssavirus shows that the distantly related Mokola virus can rescue a defective rabies virus genome. *J. Virol.* **76,** 2024–2027 (2002).

154. Lieu, K. G., Brice, A., Wiltzer, L., Hirst, B., Jans, D. A., Blondel, D. & Moseley, G. W. The rabies virus interferon antagonist P protein interacts with activated STAT3 and inhibits Gp130 receptor signaling. *J. Virol.* **87,** 8261–8265 (2013).

155. Lahaye, X., Vidy, A., Pomier, C., Obiang, L., Harper, F., Gaudin, Y. & Blondel, D. Functional characterization of Negri bodies (NBs) in rabies virus-infected cells: Evidence that NBs are sites of viral transcription and replication. *J. Virol.* **83,** 7948–58 (2009).

156. Venditti, V. & Fawzi, N. L. in 243–255 (2018). doi:10.1007/978-1-4939-7386-6_12

157. Favier, A. & Brutscher, B. Recovering lost magnetization: Polarization enhancement in

biomolecular NMR. *J. Biomol. NMR* **49,** 9–15 (2011).

158. Yoshimura, Y., Kulminskaya, N. V. & Mulder, F. A. A. Easy and unambiguous sequential assignments of intrinsically disordered proteins by correlating the backbone 15N or 13C' chemical shifts of multiple contiguous residues in highly resolved 3D spectra. *J. Biomol. NMR* **61,** 109–121 (2015).
159. Grzesiek, S. & Bax, A. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.* **114,** 6291–6293 (1992).
160. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. & Kay, L. E. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. *Biochemistry* **33,** 5984–6003 (1994).
161. Fitzkee, N. C. & Bax, A. Facile measurement of 1H–15N residual dipolar couplings in larger perdeuterated proteins. *J. Biomol. NMR* **48,** 65–70 (2010).
162. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. & Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6,** 277–293 (1995).
163. Lee, W., Tonelli, M. & Markley, J. L. NMRFAM-SPARKY: Enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* **31,** 1325–1327 (2015).
164. Kjaergaard, M., Brander, S. & Poulsen, F. M. Random coil chemical shift for intrinsically disordered proteins: effects of temperature and pH. *J.Biomol.NMR* **49,** 139–149 (2011).
165. Kjaergaard, M. & Poulsen, F. M. Sequence correction of random coil chemical shifts: Correlation between neighbor correction factors and changes in the Ramachandran distribution. *J. Biomol. NMR* **50,** 157–165 (2011).
166. Barbar, E., Hare, M., Makokha, M., Barany, G. & Woodward, C. NMR-detected order in core residues of denatured bovine pancreatic trypsin inhibitor. *Biochemistry* **40,** 9734–42 (2001).
167. Wyatt. Submicrometer Particle Sizing by Multiangle Light Scattering following Fractionation. *J. Colloid Interface Sci.* **197,** 9–20 (1998).
168. Franke, D., Petoukhov, M. V., Konarev, P. V., Panjkovich, A., Tuukkanen, A., Mertens, H. D. T., Kikhney, A. G., Hajizadeh, N. R., Franklin, J. M., Jeffries, C. M. & Svergun, D. I. *ATSAS 2.8* : a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J. Appl. Crystallogr.* **50,** 1212–1225 (2017).

169. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D. T., Konarev, P. V. & Svergun, D. I. New developments in the *ATSAS* program package for small-angle scattering data analysis. *J. Appl. Crystallogr.* **45,** 342–350 (2012).

170. Rambo, R. P. & Tainer, J. A. Accurate assessment of mass, models and resolution by small-angle scattering. *Nature* **496,** 477–481 (2013).

171. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4,** 435–447 (2008).

172. Noel, J. K., Whitford, P. C., Sanbonmatsu, K. Y. & Onuchic, J. N. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* **38,** W657-61 (2010).

173. Piana, S., Lindorff-Larsen, K. & Shaw, D. E. How Robust Are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **100,** L47–L49 (2011).

174. Best, R. B., Zheng, W. & Mittal, J. Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **10,** 5113–5124 (2014).

175. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **103,** 8577–8593 (1995).

176. Svergun, D., Barberato, C., Koch, M. H. J. & IUCr. *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Crystallogr.* **28,** 768–773 (1995).

177. Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2,** 207–217 (2015).

178. Choy, W.-Y. & Forman-Kay, J. D. Calculation of ensembles of structures representing the unfolded state of an SH3 domain. *J. Mol. Biol.* **308,** 1011–1032 (2001).

179. Nodet, G., Salmon, L., Ozenne, V., Meier, S., Jensen, M. R. & Blackledge, M. Quantitative Description of Backbone Conformational Sampling of Unfolded Proteins at Amino Acid Resolution from NMR Residual Dipolar Couplings. *J. Am. Chem. Soc.* **131,** 17908–17918 (2009).

180. Jamros, M. A., Oliveira, L. C., Whitford, P. C., Onuchic, J. N., Adams, J. A. & Jennings, P. A. Substrate-Specific Reorganization of the Conformational Ensemble of CSK

Implicates Novel Modes of Kinase Function. *PLoS Comput. Biol.* **8,** e1002695 (2012).

181. Chumnarnsilpa, S., Robinson, R. C., Grimes, J. M. & Leyrat, C. Calcium-controlled conformational choreography in the N-terminal half of adseverin. *Nat. Commun.* **6,** 8254 (2015).
182. Leyrat, C., Renner, M., Harlos, K., Huiskonen, J. T. & Grimes, J. M. Drastic changes in conformational dynamics of the antiterminator M2-1 regulate transcription efficiency in Pneumovirinae. *Elife* **3,** (2014).
183. Jones, D. T. & Cozzetto, D. DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31,** 857–863 (2015).
184. Mészáros, B., Erdős, G. & Dosztányi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46,** W329–W337 (2018).
185. Clark, S. A., Tronrud, D. E. & Andrew Karplus, P. Residue-level global and local ensemble-ensemble comparisons of protein domains. *Protein Sci.* **24,** 1528–1542 (2015).
186. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D. & Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7,** 539–539 (2014).
187. Ashkenazy, H., Abadi, S., Martz, E., Chay, O., Mayrose, I., Pupko, T. & Ben-Tal, N. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44,** W344-50 (2016).
188. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407,** 651–654 (2000).
189. Calderone, A., Castagnoli, L. & Cesareni, G. mentha: a resource for browsing integrated protein-interaction networks. *Nat. Methods* **10,** 690–691 (2013).
190. Wang, X., Olson, J. R., Rasoloson, D., Ellenbecker, M., Bailey, J. & Voronina, E. Dynein light chain DLC-1 promotes localization and function of the PUF protein FBF-2 in germline progenitor cells. *Development* **143,** 4643–4653 (2016).
191. Fan, J.-S., Zhang, Q., Tochio, H. & Zhang, M. Backbone dynamics of the 8 kDa dynein light chain dimer reveals molecular basis of the protein's functional diversity. *J. Biomol. NMR* **23,** 103–114 (2002).
192. Hall, J., Hall, A., Pursifull, N. & Barbar, E. Differences in Dynamic Structure of LC8 Monomer, Dimer, and Dimer−Peptide Complexes †. *Biochemistry* **47,** 11940–11952

(2008).

193. Cianfrocco, M. A., DeSantis, M. E., Leschziner, A. E. & Reck-Peterson, S. L. Mechanism and regulation of cytoplasmic dynein. *Annu. Rev. Cell Dev. Biol.* **31,** 83–108 (2015).

194. Splinter, D., Razafsky, D. S., Schlager, M. A., Serra-Marques, A., Grigoriev, I., Demmers, J., Keijzer, N., Jiang, K., Poser, I., Hyman, A. A., Hoogenraad, C. C., King, S. J. & Akhmanova, A. BICD2, dynactin, and LIS1 cooperate in regulating dynein recruitment to cellular structures. *Mol. Biol. Cell* **23,** 4226–4241 (2012).

195. Binder, J. X., Pletscher-Frankild, S., Tsafou, K., Stolte, C., O'Donoghue, S. I., Schneider, R. & Jensen, L. J. COMPARTMENTS: unification and visualization of protein subcellular localization evidence. *Database* **2014,** bau012–bau012 (2014).

196. Davey, N. E., Seo, M.-H., Yadav, V. K., Jeon, J., Nim, S., Krystkowiak, I., Blikstad, C., Dong, D., Markova, N., Kim, P. M. & Ivarsson, Y. Discovery of short linear motif-mediated interactions through phage display of intrinsically disordered regions of the human proteome. *FEBS J.* **284,** 485–498 (2017).

197. Fejtova, A., Davydova, D., Bischof, F., Lazarevic, V., Altrock, W. D., Romorini, S., Schöne, C., Zuschratter, W., Kreutz, M. R., Garner, C. C., Ziv, N. E. & Gundelfinger, E. D. Dynein light chain regulates axonal trafficking and synaptic levels of Bassoon. *J. Cell Biol.* **185,** 341–55 (2009).

198. Hutchins, J. R. A., Toyoda, Y., Hegemann, B., Poser, I., Hériché, J.-K., Sykora, M. M., Augsburg, M., Hudecz, O., Buschhorn, B. A., Bulkescher, J., Conrad, C., Comartin, D., Schleiffer, A., Sarov, M., Pozniakovsky, A., Slabicki, M. M., Schloissnig, S., Steinmacher, I., Leuschner, M., Ssykor, A., Lawo, S., Pelletier, L., Stark, H., Nasmyth, K., Ellenberg, J., Durbin, R., Buchholz, F., Mechtler, K., Hyman, A. A. & Peters, J.-M. Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* **328,** 593–599 (2010).

199. Schmidt, J. C., Kiyomitsu, T., Hori, T., Backer, C. B., Fukagawa, T. & Cheeseman, I. M. Aurora B kinase controls the targeting of the Astrin–SKAP complex to bioriented kinetochores. *J. Cell Biol.* **191,** 269–280 (2010).

200. Boldt, K., van Reeuwijk, J., Lu, Q., Koutroumpas, K., Nguyen, T.-M. T., Texier, Y., van Beersum, S. E. C., Horn, N., Willer, J. R., Mans, D. A., Dougherty, G., Lamers, I. J. C., Coene, K. L. M., Arts, H. H., Betts, M. J., Beyer, T., Bolat, E., Gloeckner, C. J., Haidari, K., Hetterschijt, L., Iaconis, D., Jenkins, D., Klose, F., Knapp, B., Latour, B., Letteboer, S. J. F., Marcelis, C. L., Mitic, D., Morleo, M., Oud, M. M., Riemersma, M., Rix, S., Terhal,

P. A., Toedt, G., van Dam, T. J. P., de Vrieze, E., Wissinger, Y., Wu, K. M., Apic, G., Beales, P. L., Blacque, O. E., Gibson, T. J., Huynen, M. A., Katsanis, N., Kremer, H., Omran, H., van Wijk, E., Wolfrum, U., Kepes, F., Davis, E. E., Franco, B., Giles, R. H., Ueffing, M., Russell, R. B., Roepman, R., Al-Turki, S., Anderson, C., Antony, D., Barroso, I., Bentham, J., Bhattacharya, S., Carss, K., Chatterjee, K., Cirak, S., Cosgrove, C., Danecek, P., Durbin, R., Fitzpatrick, D., Floyd, J., Reghan Foley, A., Franklin, C., Futema, M., Humphries, S. E., Hurles, M., Joyce, C., McCarthy, S., Mitchison, H. M., Muddyman, D., Muntoni, F., O'Rahilly, S., Onoufriadis, A., Payne, F., Plagnol, V., Raymond, L., Savage, D. B., Scambler, P., Schmidts, M., Schoenmakers, N., Semple, R., Serra, E., Stalker, J., van Kogelenberg, M., Vijayarangakannan, P., Walter, K., Whittall, R., Williamson, K. & Williamson, K. An organelle-specific protein landscape identifies novel diseases and molecular mechanisms. *Nat. Commun.* **7,** 11491 (2016).

201. den Hollander, P. & Kumar, R. Dynein Light Chain 1 Contributes to Cell Cycle Progression by Increasing Cyclin-Dependent Kinase 2 Activity in Estrogen-Stimulated Cells. *Cancer Res.* **66,** 5941–5949 (2006).
202. Wu, C.-G., Chen, H., Guo, F., Yadav, V. K., Mcilwain, S. J., Rowse, M., Choudhary, A., Lin, Z., Li, Y., Gu, T., Zheng, A., Xu, Q., Lee, W., Resch, E., Johnson, B., Day, J., Ge, Y., Ong, I. M., Burkard, M. E., Ivarsson, Y. & Xing, Y. PP2A-B' holoenzyme substrate recognition, regulation and role in cytokinesis. *Cell Discov.* **3,** 17027 (2017).
203. Martínez-Moreno, M., Navarro-Lérida, I., Roncal, F., Albar, J. P., Alonso, C., Gavilanes, F. & Rodríguez-Crespo, I. Recognition of novel viral sequences that associate with the dynein light chain LC8 identified through a pepscan technique. *FEBS Lett.* **544,** 262–267 (2003).
204. Navarro-Lérida, I., Martínez Moreno, M., Roncal, F., Gavilanes, F., Albar, J. P. & Rodríguez-Crespo, I. Proteomic identification of brain proteins that interact with dynein light chain LC8. *Proteomics* **4,** 339–346 (2004).
205. Ashok Kumar, T. CFSSP: Chou and Fasman Secondary Structure Prediction server. (2013). doi:10.5281/ZENODO.50733
206. Chou, P. Y. & Fasman, G. D. Prediction of protein conformation. *Biochemistry* **13,** 222–245 (1974).
207. Ho, B. K. & Brasseur, R. The Ramachandran plots of glycine and pre-proline. *BMC Struct. Biol.* **5,** 14 (2005).
208. Hein, M. Y., Hubner, N. C., Poser, I., Cox, J., Nagaraj, N., Toyoda, Y., Gak, I. A.,

Weisswange, I., Mansfeld, J., Buchholz, F., Hyman, A. A. & Mann, M. A Human Interactome in Three Quantitative Dimensions Organized by Stoichiometries and Abundances. *Cell* **163,** 712–723 (2015).

209. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45,** W130–W137 (2017).
210. Madeira, F., Tinti, M., Murugesan, G., Berrett, E., Stafford, M., Toth, R., Cole, C., MacKintosh, C. & Barton, G. J. 14-3-3-Pred: improved methods to predict 14-3-3-binding phosphopeptides. *Bioinformatics* **31,** 2276–2283 (2015).
211. Yap, K. L., Kim, J., Truong, K., Sherman, M., Yuan, T. & Ikura, M. Calmodulin target database. *J. Struct. Funct. Genomics* **1,** 8–14 (2000).
212. Abbasi, W. A., Asif, A., Andleeb, S. & Minhas, F. ul A. A. CaMELS: *In silico* prediction of calmodulin binding proteins and their binding sites. *Proteins Struct. Funct. Bioinforma.* **85,** 1724–1740 (2017).
213. Mruk, K., Farley, B. M., Ritacco, A. W. & Kobertz, W. R. Calmodulation meta-analysis: predicting calmodulin binding via canonical motif clustering. *J. Gen. Physiol.* **144,** 105–14 (2014).
214. Frederick, K. K., Marlow, M. S., Valentine, K. G. & Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature* **448,** 325–329 (2007).
215. Johnson, C., Crowther, S., Stafford, M. J., Campbell, D. G., Toth, R. & MacKintosh, C. Bioinformatic and experimental survey of 14-3-3-binding sites. *Biochem. J.* **427,** 69–78 (2010).
216. Piovesan, D., Tabaro, F., Mičetić, I., Necci, M., Quaglia, F., Oldfield, C. J., Aspromonte, M. C., Davey, N. E., Davidović, R., Dosztányi, Z., Elofsson, A., Gasparini, A., Hatos, A., Kajava, A. V., Kalmar, L., Leonardi, E., Lazar, T., Macedo-Ribeiro, S., Macossay-Castillo, M., Meszaros, A., Minervini, G., Murvai, N., Pujols, J., Roche, D. B., Salladini, E., Schad, E., Schramm, A., Szabo, B., Tantos, A., Tonello, F., Tsirigos, K. D., Veljković, N., Ventura, S., Vranken, W., Warholm, P., Uversky, V. N., Dunker, A. K., Longhi, S., Tompa, P. & Tosatto, S. C. E. DisProt 7.0: a major update of the database of disordered proteins. *Nucleic Acids Res.* **45,** D219–D227 (2017).
217. Richards, F. M. Areas, Volumes, Packing, and Protein Structure. *Annu. Rev. Biophys. Bioeng.* **6,** 151–176 (1977).
218. Schrodinger, L. The Pymol Molecular Graphics System. (2010).

219. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

220. Fraczkiewicz, R. & Braun, W. Exact and efficient analytical calculation of the accessible surface areas and their gradients for macromolecules. *J. Comput. Chem.* **19,** 319–333 (1998).

221. Ekman, D., Light, S., Björklund, A. K. & Elofsson, A. What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae? *Genome Biol.* **7,** R45 (2006).

222. Grigoriev, A. On the number of protein-protein interactions in the yeast proteome. *Nucleic Acids Res.* **31,** 4157–4161 (2003).

223. He, X. & Zhang, J. Why do hubs tend to be essential in protein networks? *PLoS Genet.* **2,** e88 (2006).

224. Yu, H., Braun, P., Yildirim, M. A., Lemmens, I., Venkatesan, K., Sahalie, J., Hirozane-Kishikawa, T., Gebreab, F., Li, N., Simonis, N., Hao, T., Rual, J.-F., Dricot, A., Vazquez, A., Murray, R. R., Simon, C., Tardivo, L., Tam, S., Svrzikapa, N., Fan, C., de Smet, A.-S., Motyl, A., Hudson, M. E., Park, J., Xin, X., Cusick, M. E., Moore, T., Boone, C., Snyder, M., Roth, F. P., Barabási, A.-L., Tavernier, J., Hill, D. E. & Vidal, M. High-quality binary protein interaction map of the yeast interactome network. *Science* **322,** 104–10 (2008).

225. Garamszegi, S., Franzosa, E. A. & Xia, Y. Signatures of Pleiotropy, Economy and Convergent Evolution in a Domain-Resolved Map of Human–Virus Protein–Protein Interaction Networks. *PLoS Pathog.* **9,** e1003778 (2013).

226. Kim, P. M., Lu, L. J., Xia, Y. & Gerstein, M. B. Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* **314,** 1938–1941 (2006).

227. Hu, G., Wu, Z., Uversky, V. N. & Kurgan, L. Functional Analysis of Human Hub Proteins and Their Interactors Involved in the Intrinsic Disorder-Enriched Interactions. *Int. J. Mol. Sci.* **18,** (2017).

228. Charles A Janeway, J., Travers, P., Walport, M. & Shlomchik, M. J. in *Immunobiol. Immune Syst. Heal. Dis. 5th Ed.* (Garland Science, 2001). at <https://www.ncbi.nlm.nih.gov/books/NBK27156/>

229. Rapin, N., Hoof, I., Lund, O. & Nielsen, M. MHC motif viewer. *Immunogenetics* **60,** 759–65 (2008).

230. Martin Werner, J., Mellins, E., Sethi, D., Wieczorek, M., Abualrous, E. T., Sticht, J.,

Álvaro-Benito, M., Stolzenberg, S., Noé, F. & Freund, C. Major Histocompatibility Complex (MHC) Class i and MHC Class ii Proteins: Conformational Plasticity in Antigen Presentation. *Article* **8,** 1 (2017).

231. Reich, Z., Altman, J. D., Boniface, J. J., Lyons, D. S., Kozono, H., Ogg, G., Morgan, C. & Davis, M. M. Stability of empty and peptide-loaded class II major histocompatibility complex molecules at neutral and endosomal pH: Comparison to class I proteins. *Proc. Natl. Acad. Sci.* **94,** 2495–2500 (1997).
232. Agniswamy, J., Fang, B. & Weber, I. T. Plasticity of S2-S4 specificity pockets of executioner caspase-7 revealed by structural and kinetic analysis. *FEBS J.* **274,** 4752–4765 (2007).
233. Fujiwara, Y., Goda, N., Tamashiro, T., Narita, H., Satomura, K., Tenno, T., Nakagawa, A., Oda, M., Suzuki, M., Sakisaka, T., Takai, Y. & Hiroaki, H. Crystal structure of afadin PDZ domain-nectin-3 complex shows the structural plasticity of the ligand-binding site. *Protein Sci.* **24,** 376–385 (2015).
234. Chan, P. M., Ng, Y.-W. & Manser, E. A robust protocol to map binding sites of the 14-3-3 interactome: Cdc25C requires phosphorylation of both S216 and S263 to bind 14-3-3. *Mol. Cell. Proteomics* **10,** M110.005157 (2011).
235. Holland, D. O., Shapiro, B. H., Xue, P. & Johnson, M. E. Protein-protein binding selectivity and network topology constrain global and local properties of interface binding networks. *Sci. Rep.* **7,** 5631 (2017).
236. Roy, J., Li, H., Hogan, P. G. & Cyert, M. S. A Conserved Docking Site Modulates Substrate Affinity for Calcineurin, Signaling Output, and In Vivo Function. *Mol. Cell* **25,** 889–901 (2007).
237. Wigington, C. P., Roy, J., Damle, N. P., Yadav, V. K., Blikstad, C., Resch, E., Wong, C. J., Mackay, D. R., Wang, J. T., Krystkowiak, I., Stearns, T., Gingras, A.-C., Ullman, K. S., Ivarsson, Y., Davey, N. E. & Cyert, M. S. Systematic discovery of Short Linear Motifs decodes calcineurin phosphatase signaling. *bioRxiv* 632547 (2019). doi:10.1101/632547
238. Grigoriu, S., Bond, R., Cossio, P., Chen, J. A., Ly, N., Hummer, G., Page, R., Cyert, M. S. & Peti, W. The Molecular Mechanism of Substrate Engagement and Immunosuppressant Inhibition of Calcineurin. *PLoS Biol.* **11,** e1001492 (2013).
239. Hanada, M., Feng, J. & Hemmings, B. A. Structure, regulation and function of PKB/AKT—a major therapeutic target. *Biochim. Biophys. Acta - Proteins Proteomics* **1697,** 3–16 (2004).

240. Alessi, D. R., Caudwell, F. B., Andjelkovic, M., Hemmings, B. A. & Cohen, P. Molecular basis for the substrate specificity of protein kinase B; comparison with MAPKAP kinase-1 and p70 S6 kinase. *FEBS Lett.* **399,** 333–338 (1996).

241. Obata, T., Yaffe, M. B., Leparc, G. G., Piro, E. T., Maegawa, H., Kashiwagi, A., Kikkawa, R. & Cantley, L. C. Peptide and protein library screening defines optimal substrate motifs for AKT/PKB. *J. Biol. Chem.* **275,** 36108–36115 (2000).

242. Berg, J. M., Tymoczko, J. L. & Stryer, L. Making a Fast Reaction Faster: Carbonic Anhydrases. (2002). at <https://www.ncbi.nlm.nih.gov/books/NBK22599/>

243. Zhang, X., Zhang, S., Yamane, H., Wahl, R., Ali, A., Lofgren, J. A. & Kendall, R. L. Kinetic mechanism of AKT/PKB enzyme family. *J. Biol. Chem.* **281,** 13949–13956 (2006).

244. Höhfeld, J., Cyr, D. M. & Patterson, C. From the cradle to the grave: molecular chaperones that may choose between folding and degradation. *EMBO Rep.* **2,** 885–890 (2001).

245. Artigues, A., Crawford, D. L., Iriarte, A. & Martinez-Carrion, M. Divergent Hsc70 binding properties of mitochondrial and cytosolic aspartate aminotransferase. Implications for their segregation to different cellular compartments. *J. Biol. Chem.* **273,** 33130–33134 (1998).

246. Cuervo, A. M. Chaperone-mediated autophagy: selectivity pays off. *Trends Endocrinol. Metab.* **21,** 142–50 (2010).

247. Fred Dice, J. Peptide sequences that target cytosolic proteins for lysosomal proteolysis. *Trends Biochem. Sci.* **15,** 305–309 (1990).

248. Takenaka, I. M., Leung, S. M., McAndrew, S. J., Brown, J. P. & Hightower, L. E. Hsc70-binding peptides selected from a phage display peptide library that resemble organellar targeting sequences. *J. Biol. Chem.* **270,** 19839–44 (1995).

249. Fourie, A. M., Sambrook, J. F. & Gething, M. J. Common and divergent peptide binding specificities of hsp70 molecular chaperones. *J. Biol. Chem.* **269,** 30470–8 (1994).

250. Vaccaro, P. & Dente, L. PDZ domains: troubles in classification. *FEBS Lett.* **512,** 345–9 (2002).

251. Harris, B. Z. & Lim, W. A. Mechanism and role of PDZ domains in signaling complex assembly. *J. Cell Sci.* **114,** 3219–3231 (2001).

252. Shieh, B. H. & Zhu, M. Y. Regulation of the TRP Ca2+ channel by INAD in Drosophila photoreceptors. *Neuron* **16,** 991–998 (1996).

253. Paulsen, R., Bähner, M. & Huber, A. The PDZ assembled "transducisome" of microvillar photoreceptors: the TRP/TRPL problem. *Pflügers Arch. - Eur. J. Physiol.* **439,** r181–r183 (2000).

254. Ranganathan, R. & Ross, E. M. PDZ domain proteins: Scaffolds for signaling complexes. *Curr. Biol.* **7,** R770–R773 (1997).

255. Han, J.-D. J., Bertin, N., Hao, T., Goldberg, D. S., Berriz, G. F., Zhang, L. V., Dupuy, D., Walhout, A. J. M., Cusick, M. E., Roth, F. P. & Vidal, M. Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature* **430,** 88–93 (2004).

256. Zotenko, E., Mestre, J., O'Leary, D. P. & Przytycka, T. M. Why Do Hubs in the Yeast Protein Interaction Network Tend To Be Essential: Reexamining the Connection between the Network Topology and Essentiality. *PLoS Comput. Biol.* **4,** e1000140 (2008).

257. Batada, N. N., Hurst, L. D. & Tyers, M. Evolutionary and physiological importance of hub proteins. *PLoS Comput. Biol.* **2,** e88 (2006).

258. Pan, D. The Hippo Signaling Pathway in Development and Cancer. *Dev. Cell* **19,** 491–505 (2010).

259. Hall, J., Karplus, P. A. & Barbar, E. Multivalency in the assembly of intrinsically disordered dynein intermediate chain. *J. Biol. Chem.* **284,** 33115–33121 (2009).

260. Wang, C., Li, S., Januschke, J., Rossi, F., Izumi, Y., Garcia-Alvarez, G., Gwee, S. S. L., Soon, S. B., Sidhu, H. K., Yu, F., Matsuzaki, F., Gonzalez, C. & Wang, H. An Ana2/Ctp/Mud Complex Regulates Spindle Orientation in Drosophila Neuroblasts. *Dev. Cell* **21,** 520–533 (2011).

261. O'Shea, C., Staby, L., Bendsen, S. K., Tidemand, F. G., Redsted, A., Willemoës, M., Kragelund, B. B. & Skriver, K. Structures and Short Linear Motif of Disordered Transcription Factor Regions Provide Clues to the Interactome of the Cellular Hub Protein Radical-induced Cell Death1. *J. Biol. Chem.* **292,** 512–527 (2017).

262. Shapiguzov, A., Vainonen, J. P., Hunter, K., Tossavainen, H., Tiwari, A., Järvi, S., Hellman, M., Aarabi, F., Alseekh, S., Wybouw, B., Van Der Kelen, K., Nikkanen, L., Krasensky-Wrzaczek, J., Sipari, N., Keinänen, M., Tyystjärvi, E., Rintamäki, E., De Rybel, B., Salojärvi, J., Van Breusegem, F., Fernie, A. R., Brosché, M., Permi, P., Aro, E.-M., Wrzaczek, M. & Kangasjärvi, J. Arabidopsis RCD1 coordinates chloroplast and mitochondrial functions through interaction with ANAC transcription factors. *Elife* **8,** (2019).

263. Hermeking, H. & Benzinger, A. 14-3-3 proteins in cell cycle regulation. *Semin. Cancer Biol.* **16,** 183–192 (2006).

264. Stevers, L. M., Sijbesma, E., Botta, M., MacKintosh, C., Obsil, T., Landrieu, I., Cau, Y., Wilson, A. J., Karawajczyk, A., Eickhoff, J., Davis, J., Hann, M., O'Mahony, G., Doveston, R. G., Brunsveld, L. & Ottmann, C. Modulators of 14-3-3 Protein-Protein Interactions. *J. Med. Chem.* **61,** 3755–3778 (2018).

265. Yaffe, M. B., Rittinger, K., Volinia, S., Caron, P. R., Aitken, A., Leffers, H., Gamblin, S. J., Smerdon, S. J. & Cantley, L. C. The structural basis for 14-3-3:phosphopeptide binding specificity. *Cell* **91,** 961–971 (1997).

266. Bose, A. & Dalal, S. N. 14-3-3 proteins mediate the localization of Centrin2 to centrosome. *J. Biosci.* **44,** 42 (2019).

267. Yuan, L., Barbash, S., Kongsamut, S., Eishingdrelo, A., Sakmar, T. P. & Eishingdrelo, H. 14-3-3 signal adaptor and scaffold proteins mediate GPCR trafficking. *Sci. Rep.* **9,** 11156 (2019).

268. Kasinski, A., Dong, X., Khuri, F. R., Boss, J. & Fu, H. Transcriptional regulation of YWHAZ, the gene encoding 14-3-3ζ. *PLoS One* **9,** e93480 (2014).

269. Gupta, P. & Prywes, R. ATF1 Phosphorylation by the ERK MAPK Pathway Is Required for Epidermal Growth Factor-induced c- *jun* Expression. *J. Biol. Chem.* **277,** 50550–50556 (2002).

270. Kobayashi, M., Shimomura, A., Hagiwara, M. & Kawakami, K. Phosphorylation of ATF-1 enhances its DNA binding and transcription of the Na,K-ATPase alpha 1 subunit gene promoter. *Nucleic Acids Res.* **25,** 877–882 (1997).

271. Pfleger, C. M. & Kirschner, M. W. The KEN box: an APC recognition signal distinct from the D box targeted by Cdh1. *Genes Dev.* **14,** 655–665 (2000).

272. Dominguez, R. The WH2 Domain and Actin Nucleation: Necessary but Insufficient. *Trends Biochem. Sci.* **41,** 478–490 (2016).

273. Roach, J. C., Glusman, G., Smit, A. F. A., Huff, C. D., Hubley, R., Shannon, P. T., Rowen, L., Pant, K. P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L. B., Hood, L. & Galas, D. J. Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. *Science (80-.).* **328,** 636–639 (2010).

274. Jenkins, G. M., Rambaut, A., Pybus, O. G. & Holmes, E. C. Rates of Molecular Evolution in RNA Viruses: A Quantitative Phylogenetic Analysis. *J. Mol. Evol.* **54,** 156–165 (2002).

275. Davey, N. E., Travé, G. & Gibson, T. J. How viruses hijack cell regulation. *Trends*

Biochem. Sci. **36,** 159–169 (2011).

276. Kadaveru, K., Vyas, J. & Schiller, M. R. Viral infection and human disease--insights from minimotifs. *Front. Biosci.* **13,** 6455–6471 (2008).
277. Roman, A. & Munger, K. The papillomavirus E7 proteins. *Virology* **445,** 138–68 (2013).
278. Leonardo G. Alonso, ‡, Maria M. García-Alai, ‡, Alejandro D. Nadra, ‡, Alicia N. Lapeña, ‡, Fabio L. Almeida, §, Peter Gualfetti, ∥ and & Gonzalo de Prat-Gay*, †. High-Risk (HPV16) Human Papillomavirus E7 Oncoprotein Is Highly Stable and Extended, with Conformational Transitions that Could Explain Its Multiple Cellular Binding Partners†. (2002). doi:10.1021/BI025579N
279. Chemes, L. B., Glavina, J., Faivovich, J., de Prat-Gay, G. & Sánchez, I. E. Evolution of Linear Motifs within the Papillomavirus E7 Oncoprotein. *J. Mol. Biol.* **422,** 336–346 (2012).
280. Tomaić, V., Gardiol, D., Massimi, P., Ozbun, M., Myers, M. & Banks, L. Human and primate tumour viruses use PDZ binding as an evolutionarily conserved mechanism of targeting cell polarity regulators. *Oncogene* **28,** 1–8 (2009).
281. Chemes, L. B., de Prat-Gay, G. & Sánchez, I. E. Convergent evolution and mimicry of protein linear motifs in host-pathogen interactions. *Curr. Opin. Struct. Biol.* **32,** 91–101 (2015).
282. Jamin, M. & Yabukarski, F. Nonsegmented Negative-Sense RNA Viruses—Structural Data Bring New Insights Into Nucleocapsid Assembly. *Adv. Virus Res.* **97,** 143–185 (2017).
283. Wang, J., Zhou, D., Prabhu, A., Schlegel, R. & Yuan, H. The Canine Papillomavirus and Gamma HPV E7 Proteins Use an Alternative Domain to Bind and Destabilize the Retinoblastoma Protein. *PLoS Pathog.* **6,** e1001089 (2010).
284. Chemes, L. B., de Prat-Gay, G. & Sánchez, I. E. Convergent evolution and mimicry of protein linear motifs in host–pathogen interactions. *Curr. Opin. Struct. Biol.* **32,** 91–101 (2015).
285. Goel, A. & Wilkins, M. R. Dynamic Hubs Show Competitive and Static Hubs Non-Competitive Regulation of Their Interaction Partners. *PLoS One* **7,** e48209 (2012).
286. Gallego, P., Velazquez-Campoy, A., Regué, L., Roig, J. & Reverter, D. Structural analysis of the regulation of the DYNLL/LC8 binding to Nek9 by phosphorylation. *J. Biol. Chem.* **288,** 12283–12294 (2013).
287. Stangler, T., Tran, T., Hoffmann, S., Schmidt, H., Jonas, E. & Willbold, D. Competitive

displacement of full-length HIV-1 Nef from the Hck SH3 domain by a high-affinity artificial peptide. *Biol. Chem.* **388,** 611–5 (2007).

288. Holmes, E. C. Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol.* **11,** 543–546 (2003).
289. Dyer, M. D., Murali, T. M. & Sobral, B. W. The Landscape of Human Proteins Interacting with Viruses and Other Pathogens. *PLoS Pathog.* **4,** e32 (2008).
290. Franzosa, E. A. & Xia, Y. Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci.* **108,** 10538–10543 (2011).
291. Lo, T. Q., Marston, B. J., Dahl, B. A. & De Cock, K. M. Ebola: Anatomy of an Epidemic. *Annu. Rev. Med.* **68,** 359–370 (2017).
292. Ebola | Ebola situation reports: Democratic Republic of the Congo. *WHO* (2019). at <https://www.who.int/ebola/situation-reports/drc-2018/en/>
293. Cross, R. W., Mire, C. E., Feldmann, H. & Geisbert, T. W. Post-exposure treatments for Ebola and Marburg virus infections. *Nat. Rev. Drug Discov.* **17,** 413–434 (2018).
294. Zinzula, L., Nagy, I., Orsini, M., Weyher-Stingl, E., Bracher, A. & Baumeister, W. Structures of Ebola and Reston Virus VP35 Oligomerization Domains and Comparative Biophysical Characterization in All Ebolavirus Species. *Structure* **27,** 39-54.e6 (2019).
295. Bruhn, J. F., Kirchdoerfer, R. N., Urata, S. M., Li, S., Tickle, I. J., Bricogne, G. & Saphire, E. O. Crystal Structure of the Marburg Virus VP35 Oligomerization Domain. *J. Virol.* **91,** (2017).
296. Barbar, E. & Nyarko, A. Polybivalency and disordered proteins in ordering macromolecular assemblies. *Semin. Cell Dev. Biol.* **37,** 20–25 (2015).
297. WHO | Rotavirus. *WHO* (2018). at <https://www.who.int/immunization/diseases/rotavirus/en/>
298. Gomez, D. E. & Weese, J. S. Viral enteritis in calves. *Can. Vet. J. = La Rev. Vet. Can.* **58,** 1267–1274 (2017).
299. Rodríguez, J. M., Chichón, F. J., Martín-Forero, E., González-Camacho, F., Carrascosa, J. L., Castón, J. R. & Luque, D. New Insights into Rotavirus Entry Machinery: Stabilization of Rotavirus Spike Conformation Is Independent of Trypsin Cleavage. *PLoS Pathog.* **10,** e1004157 (2014).
300. Li, Y., Xue, M., Yu, L., Luo, G., Yang, H., Jia, L., Zeng, Y., Li, T., Ge, S. & Xia, N. Expression and characterization of a novel truncated rotavirus VP4 for the development of a recombinant rotavirus vaccine. *Vaccine* **36,** 2086–2092 (2018).

301. Trask, S. D. & Dormitzer, P. R. Assembly of Highly Infectious Rotavirus Particles Recoated with Recombinant Outer Capsid Proteins. *J. Virol.* **80,** 11293–11304 (2006).

302. Dormitzer, P. R., Greenberg, H. B. & Harrison, S. C. Proteolysis of Monomeric Recombinant Rotavirus VP4 Yields an Oligomeric VP5* Core. *J. Virol.* **75,** 7339–7350 (2001).

303. Chen, J. Z., Settembre, E. C., Aoki, S. T., Zhang, X., Bellamy, A. R., Dormitzer, P. R., Harrison, S. C. & Grigorieff, N. Molecular interactions in rotavirus assembly and uncoating seen by high-resolution cryo-EM. *Proc. Natl. Acad. Sci.* **106,** 10644–10648 (2009).

304. Li, Z., Baker, M. L., Jiang, W., Estes, M. K. & Prasad, B. V. V. Rotavirus Architecture at Subnanometer Resolution. *J. Virol.* **83,** 1754–1766 (2009).

305. Yoder, J. D. & Dormitzer, P. R. Alternative intermolecular contacts underlie the rotavirus VP5* two- to three-fold rearrangement. *EMBO J.* **25,** 1559–1568 (2006).

306. Nejmeddine, M., Trugnan, G., Sapin, C., Kohli, E., Svensson, L., Lopez, S. & Cohen, J. Rotavirus spike protein VP4 is present at the plasma membrane and is associated with microtubules in infected cells. *J. Virol.* **74,** 3313–3320 (2000).

307. Radnai, L., Rapali, P., Hódi, Z., Süveges, D., Molnár, T., Kiss, B., Bécsi, B., Erdödi, F., Buday, L., Kardos, J., Kovács, M. & Nyitray, L. Affinity, Avidity, and Kinetics of Target Sequence Binding to LC8 Dynein Light Chain Isoforms. *J. Biol. Chem.* **285,** 38649–38657 (2010).

308. Prosniak, M., Hooper, D. C., Dietzschold, B. & Koprowski, H. Effect of rabies virus infection on gene expression in mouse brain. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 2758–2763 (2001).

309. Rhoads, F. F. A. R. & Rhoads, A. R. Evolutionary Aspects of Calmodulin. *IUBMB Life (International Union Biochem. Mol. Biol. Life)* **51,** 215–221 (2001).

310. Fu, H., Subramanian, R. R. & Masters, S. C. 14-3-3 Proteins: Structure, Function, and Regulation. *Annu. Rev. Pharmacol. Toxicol.* **40,** 617–647 (2000).

311. Wang, Y., Sukenik, S., Davis, C. M. & Gruebele, M. Cell Volume Controls Protein Stability and Compactness of the Unfolded State. *J. Phys. Chem. B* **122,** 11762–11770 (2018).