Molecular and Bioinformatic Analysis of *Drosophila melanogaster* Cdt1 Phase Separation

Alvin T. Huang

Dept. of Chemistry, University of California, Berkeley

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science with Honors in Chemical Biology, under the supervision of Professor Michael Botchan

Professor Michael Botchan 

Spring 2021

Acknowledgements
I would like to first thank my research advisors Matthew Parker and Michael Botchan for mentoring me in experimental procedure and scientific analysis throughout my two-and-a-half years spent in the Botchan Lab. Without them, this thesis would not have been possible, and I would not have pursued a future in scientific research. I would also like to thank Jonchee Kao, Maren Bell, and Christi Abbate. Their mentorship and guidance in lab also taught me how to plan and run experiments properly to collect the data found in this thesis.

Abstract

For the DNA replication initiators to function correctly, they require an N-terminal intrinsically disordered region (IDR) that allows for binding to DNA and recruitment of other IDRs. This process likely occurs through DNA-induced liquid-liquid phase separation (LLPS), but the mechanism by which this occurs is not well understood. Here, I investigated the biochemical and sequence properties of the *Drosophila melanogaster* Cdt1 (DmCdt1) IDR as a model to discover which properties are needed for LLPS. I found that bulky hydrophobic residues are indispensable for DNA-induced LLPS and that aromatic residues do not significantly affect phase separation ability. Additional bioinformatic studies comparing DmCdt1 to Cdt1 orthologs and other disordered proteins from the *D. melanogaster* proteome found that there was an abundance of bulky hydrophobics across orthologs showing that LLPS is likely conserved across metazoans. Key differences in residue abundance and distribution also arose between the DmCdt1 IDR and other disordered proteins showing that these unique properties may play a role in LLPS. The results were synthesized into the design of a novel peptide that could undergo LLPS using only a small set of amino acids and focused only on bulky hydrophobic residue abundance and the distribution of amino acids. This work expands on the previous mechanistic studies on DmCdt1 LLPS and prompts further research into the role of specific amino acids in their control of protein structure and other properties to allow for DNA-induced LLPS.

Introduction

Within all eukaryotic cells, a variety of biochemical processes rely on the ability of proteins to recognize and interact with specific biomolecules. While many interactions simply involve the binding of a protein to a target, other interactions can lead to self-assembly and organization to form membraneless compartments and organelles. There are many membraneless compartments within eukaryotic cells; one example is the nucleolus which is formed from chromatin, proteins, and RNA to allow for the biogenesis of ribosomes among other macromolecules (Mitrea et al., 2016). Increasing evidence has shown that these compartments are formed through the process of liquid-liquid phase separation (LLPS) where proteins and interacting biomolecules form a denser separate phase that exists stably alongside the surrounding solution (Boeynaems et al., 2018). These phases are characterized as liquids as they share similar properties, such as forming spherical droplets due to surface tension and being able to merge with other droplets (Hyman et al., 2014). As researchers have identified more proteins that

undergo this process, it has become clear that LLPS is incredibly important for a variety of cellular functions; one such process is DNA replication initiation.

Although LLPS often creates membraneless organelles to carry out biological processes, it can also be used for the recruitment of specific proteins, as seen in DNA replication initiation. In this process, origins of replication are first licensed by the heterohexameric Origin Replication Complex (ORC) and Cdc6. Cdt1 then associates with this complex and recruits the minichromosome maintenance (MCM) complex, the replicative helicase, which ultimately allows for the recruitment of the rest of the replication machinery (Parker et al., 2017). Out of these proteins, ORC, Cdc6, and Cdt1 have been found to undergo LLPS in the presence of DNA through their intrinsically disordered regions (IDRs) (Parker et al., 2019; Hossain et al., 2021). IDRs are regions of proteins that do not form stable secondary structures but are still biologically active through their ability to recognize specific biomolecules (Oldfield et al. 2014; Uversky et al., 2017). Previous studies have shown that this IDR region is required for DNA binding by ORC and may also be important for Cdc6 and Cdt1 function (Parker et al., 2019). To understand how the IDR mediates LLPS, especially in the case of Cdt1, the major biochemical properties and sequence elements must be investigated.

Though proteins that undergo LLPS form distinctive compartments with specific resident factors and functionality, many of their IDRs seem to rely on the same properties, including the overall charge and aromatic content. These properties allow for different methods of interactions with their associated biomolecules and with each other. Since many biomolecules exist as ions in solution, such as DNA and RNA, it seems obvious that IDRs are enriched in charged residues. Proteins such as RNA chaperones contain IDRs that have long stretches of charged residues that are necessary for biomolecule binding (Das and Pappu, 2013). The Cdt1 IDR is enriched in positive residues which allows the protein to bind to negatively charged DNA and use it as a scaffold for LLPS (Parker et al., 2021). Apart from charge, some IDRs are also enriched in aromatic amino acids. These residues allow for inter-IDR π-π stacking interactions and may promote interactions with the nucleic acid bases, which are especially needed for proteins that interact with DNA and RNA (Banani et al., 2017; Boeynaems et al., 2019; Chiu et al., 2020). Other studies have also noted the ability of 1,6-hexanediol, a greasy alcohol, to break up phases created through LLPS, although it was first seen to break up the aromatic interactions in the nuclear pores (Shulga et al., 2003; Patel et al., 2007). However, the Cdt1 IDR does not have many aromatic residues and its phase is not disrupted by 1,6-hexanediol suggesting that a different biochemical mechanism underlies Cdt1 LLPS, which will be investigated further in this text (Parker et al., 2021).

Sequence elements also play an essential role since the presence of amino acid motifs in an IDR may provide the properties needed for LLPS. Oftentimes, IDRs are classified as low complexity domains which have a biased amino acid composition and often contain repetitive sequences (Boeynaems et al., 2018). These domains can be found in proteins that undergo LLPS such as Ddx4 and FUS (Han et al., 2012; Kato et al., 2012). The Ddx4 protein contains an IDR that has many *FG*-repeats that provide an aromatic bridge to phase separate with DNA while FUS has tandem *(S/G)Y(S/G)* repeats that are necessary for RNA granule formation (Nott et al., 2015; Luo et al., 2018). Additional studies in the creation of intrinsically disordered protein polymers have found that the inclusion of P-X_n-G repeats can allow for temperature-dependent LLPS (Quiroz

and Chilkoti, 2015). However, the Cdt1 IDR has high sequence complexity indicating that it either does not contain noticeable peptide motifs or relies on a different kind of sequence element for LLPS (Parker et al., 2021).

While an analysis of the biochemical properties and sequence elements is important for understanding LLPS, it is also necessary to compare these properties to orthologs as well as the organism's proteome. Many processes, such as DNA replication initiation, are conserved among eukaryotes and have orthologs that perform the same function (Parker et al, 2017). Cdt1 function is similar for many metazoans, but organisms in other kingdoms can have different Cdt1 sequences, such as *S. pombe* which has a more ordered N-terminus (Parker et al., 2017). This text will seek to analyze the sequence similarity of Cdt1 orthologs among metazoans in more depth by comparing the similarity of their IDRs. Additionally, comparing the sequence of the Cdt1 IDR to an organism's proteome and disorderome may also reveal important properties needed for LLPS.
In this text, *Drosophila melanogaster* Cdt1 (DmCdt1) was used as a model to determine the molecular basis for LLPS for metazoan replication initiator proteins. Through protein mutagenesis, I find that large, bulky hydrophobic residues are necessary for DmCdt1 IDR LLPS in the presence of DNA and crowding reagent. A Python-driven bioinformatic analysis of Cdt1 orthologs and the *D. melanogaster* proteome was also performed and showed that the amino acid composition and the distance between residues may underlie Cdt1 IDR LLPS. Additionally, I was also able to show a lack of sequence motifs that are indicative of high complexity sequences. The information gleaned from these experiments and previous studies was then used to generate a novel phase separating peptide using only a limited number of amino acids which I was able to show underwent LLPS in the presence of DNA. These studies show that the mechanism of LLPS for Cdt1 and potentially other replication initiator IDRs are based on the residue composition that allows for specific molecular interactions to occur between DNA and protein.

Data and Results

Branched hydrophobics are responsible for facilitating the phase separation of DmCdt1 in the presence of DNA and PEG.

As previously shown by Parker et al. (2021), phase separation of DmCdt1 is not inhibited by the addition of 1,6-hexanediol implying that LLPS does not occur through interactions between aromatic residues. To further test this, two DmCdt1 IDR mutants were created with one replacing all aromatics (phenylalanine) with leucine (Aromatic1) and another replacing all aromatics with alanine (Aromatic2). The two constructs allowed me to test whether the hydrophobicity of the aromatics or if the π-π interactions potentially guided phase separation. To test their LLPS ability, a depletion assay developed by Parker et al. (2019) was utilized where equal volumes of DNA or polyethylene glycol 3350 (PEG), a crowding reagent, were combined with a sample of protein, allowed to incubate at room temperature, and then centrifuged to pellet any denser droplets formed by LLPS and the supernatant assessed for protein depletion by SDS-PAGE analysis. The assay was able to show that aromatics contributed very little to the protein's phase separation ability with bands of protein being very weak for both mutants with the addition of DNA or PEG, like that of wild-type DmCdt1 (**Figure 1A**). Quantification showed that 6.57% ± 1.85% and 1.72% ± 0.74% of protein was retained

in the supernatant for wild-type DmCdt1, 13.5% ± 1.00% and 5.50% ± 0.83% for Aromatic1, and 28.8% ± 5.5% and 21.0% ± 2.8% for Aromatic2 for the DNA and PEG trials, respectively. Additionally, a Student's t-test showed that there was no significant difference in the depletion of Aromatic1 ($t = 4.7$ and 4.8) and Aromatic2 ($t = 5.5$ and 9.5) compared with wild-type DmCdt1 for DNA and PEG trials, respectively ($p = 0.05$, 1 degree of freedom). Therefore, the mutants did not abolish LLPS meaning that aromatic residues do not play a large role in the process for DmCdt1 which agrees well with the 1,6-hexanediol experiment by Parker et al (2021) (**Figure 1B**).

Other studies regarding the mechanism of LLPS for other proteins have shown that hydrophobic residues likely have some role in the process due to the hydrophobic effect (Quiroz and Chilkoti, 2015). This is similar in the context of protein folding, but this case causes the hydrophobic residues of multiple proteins to aggregate to shield these residues from the polar solvent. This in turn would cause droplets to form in solution, which is indicative of phase separation. Other aspects of DmCdt1 also potentially point towards the importance of hydrophobic residues as the IDR is highly enriched in the bulky hydrophobic residues leucine, isoleucine, and valine (around 18.5%). To test the importance of hydrophobic residues, two mutants were purified with one removing approximately half of the leucine and isoleucine residues (Leu2) and the other removing all of them (Leu1). Using the depletion assay with DNA, these mutants showed a large reduction in phase separation ability, with both Leu2 and Leu1 constructs being unable to deplete from the supernatant (**Figure 1C**). However, quantification of the results showed that the two mutants performed similarly poorly compared to the wild-type DmCdt1 at retention percentages of 79.2% ± 9.8% and 76.0% ± 11.5% compared with 14.7% ± 1.7% retention (**Figure 1D**). A depletion assay of the mutants with PEG also showed that these mutants were unable to undergo LLPS through the crowding of solution (**Figure 1E**).

Following this analysis, the ability of the Leu1 and Leu2 mutants to bind to DNA was evaluated. Different concentrations of protein were added to a set DNA concentration in an electrophoretic mobility shift assay (EMSA) to determine the binding constant (K_d) for wild-type DmCdt1 and the mutants. The K_d values of Leu1 and Leu2 were found to be approximately 106 nM and 67 nM, respectively, like the affinity of wild-type DmCdt1 for DNA ($K_d \approx 69$ nM) (**Figure 1F**). This supported the findings that the positive charge alone is necessary for DNA binding which in turn provides the scaffolding sufficient for phase separation in the absence of PEG (Parker et al., 2021).

A final analysis of the Leu1 and Leu2 mutants' ability to undergo LLPS was done using confocal microscopy. When wild-type DmCdt1 was mixed with Cy5-labeled DNA and visualized, well-defined droplets of protein and DNA formed with high intensity signals. However, these droplets were almost lost when visualizing Leu1 and Leu2. These droplets were found to be large and disperse with very little fluorescent signal (**Figure 1G**). Quantification of the droplets showed that Leu1 and Leu2 had, on average, lost more than 50% of the fluorescent signal compared to wild-type DmCdt1 (**Figure 1H**). These results, along with the depletion assay, show that a lack of hydrophobicity in the IDR can abolish a large amount of the protein's ability to undergo LLPS. This seems to indicate that the removal of hydrophobic residues destroys the need to shield these residues which in turn weakens the aggregation of droplets.

Orthologs of Cdt1 are diverse in sequence, but all share a disordered N-terminus before the Cdt1 domain and are compositionally similar.

From the results of the aromatic and hydrophobic mutants and previous findings by Parker et al. (2021), hydrophobic residues are incredibly important to the process of LLPS for DmCdt1. However, many other studies into proteins that undergo LLPS have shown that factors such as sequence motifs are also incredibly important (Quiroz and Chilkoti, 2015; Nott et al., 2015; Luo et al., 2018). To search for these motifs and other sequence elements, I decided to analyze the various orthologs of Cdt1 across other metazoans by searching for the protein in the database UniProt. A variety of ortholog protein sequences from across metazoan phyla were collected and a few were selected as "representative" organisms that came from the metazoan phyla Porifera, Cnidaria, Chordata, Arthropoda, Nematoda, Mollusca, and Annelida (**Figure 2A**). The Cdt1 sequences were then analyzed using IUPred2A, an algorithm that predicts protein disorder propensity (values below 0.5 are predicted to be ordered and scores above 0.5 indicate likely disordered regions). The program was able to show that most of these orthologs contain a disordered N-terminus IDR like that of DmCdt1 (**Figure 2B**). These orthologs were wildly variable in how the disorder measure changed over the residues and all had IDRs of different lengths. One noticeable "outlier" was that of *H. vulgaris*, a freshwater hydra, which had very little disorder in its N-terminus. However, it was not removed from consideration as the protein itself retained the two Cdt1 motifs found in all the other orthologs. Other Cdt1 ortholog sequences were also collected and were not shown graphically, but all displayed an N-terminal IDR identified using IUPred2A. All ortholog sequences can be found in the **Supplemental Table**.

After collating all the Cdt1 orthologs, Python code was written to drive a bioinformatic analysis of the sequence elements. I first analyzed the number of Cdk sites in the protein as studies on human Cdt1 have found that phosphorylation of the protein is crucial for its regulation by ubiquitination (Parker et al., 2017). By parsing each of the ortholog sequences and noting each hit of the motif *(S/T)PX(R/K)*, a plot was generated measuring the number of Cdk sites versus the length of the IDR (**Figure 3A**). The plot showed that IDRs vary quite drastically from one another regarding the number of Cdk sites with most IDRs containing between 0 and 3 Cdk sites while an outlier like DmCdt1 had 9. This general analysis in the number of Cdk sites seemed to show that the orthologs do not share similar sequence elements and are potentially unique in sequence and properties.

However, when analyzing the orthologs for general biochemical properties, similarities between the IDRs arise. As noted in previous studies by Parker et al. (2021), DmCdt1 is overall positively charged with a pI of approximately 10.2. The positive charge is what allows the protein to bind to DNA, a negatively charged polymer scaffold, during DNA replication initiation. By using the Python library BioPython, the pI's of the ortholog IDRs were all found to be between 9.75 and 11.75 which is likely a property needed to carry out DNA binding (**Figure 3B**). Additionally, the charge distributions of the ortholog IDRs were also analyzed, this time using a measure called kappa. This value between 0 and 1 describes the degree of charged residue mixing with a value close to 1 showing high mixing; this value can also be used to predict the conformation of a disordered region of a protein (Das and Pappu, 2013). The kappa values of the IDRs were found to be distributed around 0.24 showing that the positive and negative charged

residues are well separated from one another (**Figure 3C**). This implies that the IDRs likely have the same conformational shape while in solution and are all potentially able to undergo LLPS with DNA.

Additional bioinformatic studies were performed to analyze the amino acid distribution among the IDRs. To analyze all the residues, all the amino acids were split into the following categories: bulky hydrophobics (I, L, M, V), aromatics (F, W, Y), small hydrophobics (A, C, G), hydrophilic (S, T, N, Q), positive (K, R, H), negative (D, E), and proline (P). These categories were then parsed for by Python code which kept track of the number of specific residues and calculated the total percentage by dividing by the total IDR length. The results were plotted, and the distributions can be seen in **Figure 3D**. Many of the distributions are very tight about the average such as with the aromatic, positive, and negative residues while the bulky hydrophobic, proline, hydrophilic, and small hydrophobic residues are much wider in their variation. Nonetheless, it should be noted that the distributions are tight for the amino acids of interest in positive, bulky hydrophobic, and aromatic residues. The aromatic residues are of particular interest as its low composition percentage shows that it is likely not involved in any phase separation processes unlike other proteins that require aromatic amino acids. Additionally, the ortholog IDRs show similar enrichment in the amount of bulky hydrophobic residues which potentially shows the importance of those amino acids in Cdt1 function.

Distances between specific residues were also found as the spacing of specific amino acids may be important for LLPS. The distances between specific residue groups were similarly found by parsing the sequence, but this time keeping track of the number of amino acids between each instance of a specific residue. The distances collected are displayed in **Figures 3E** and **3F** with **3E** showing the average distance between amino acids and **3F** counting the number of dipeptides. The number of dipeptides was kept separate from the average distance as dipeptides could potentially be sequence elements needed for LLPS. Surprisingly, the distribution of average distance and number of dipeptides were all narrow with the values being relatively the same for all Cdt1 IDRs. The distribution of small hydrophobic and proline average distances as well as small hydrophobic and hydrophilic dipeptides were much wider compared to the others but may be a result of the varying lengths of the IDRs needing less spacer residues. Once again, these plots highlight large hydrophobic residues as the similarity in distances and number of dipeptides potentially shows a sequence element that has been retained despite evolution seemingly changing sequence similarity.

Common repetitive motifs seen in other disordered proteins are not found in DmCdt1.

Although the composition between the Cdt1 orthologs has shown the importance of specific types of amino acids, no enrichment of specific motifs, such as dipeptides, have been found in the analysis. In a study on Ddx4 by Nott et al., the researchers created a plot comparing the expected number of specific dipeptides to the actual number revealing that the protein was enriched in *FG*-repeats which were found to be needed for LLPS. A similar approach was performed on the DmCdt1 IDR in which code calculated the percent composition of each amino acid type. These percentages were used to calculate the expected probability of each dipeptide. The actual percentage of each dipeptide was then divided by its corresponding expected probability to calculate the

abundance ratio. As seen in **Figure 4**, no specific dipeptide stood out as incredibly enriched. The abundance values for dipeptides such as *HT*, *GM*, and *NR* were more enriched than expected, but this was because amino acids like histidine and methionine were very sparse in the IDR. Other dipeptides were not present in the IDR at all which could highlight their importance for a specific property although highly unlikely. Overall, the plot showed that none of the dipeptides stand out particularly well and these motifs likely do not influence the property of LLPS. This leads to the finding that composition and distribution of specific residue types is necessary for LLPS to occur.

Key properties, such as the distance between amino acids, differentiate DmCdt1 from other disordered proteins in the Drosophila proteome.

While an analysis of the sequences of Cdt1 orthologs has revealed residue composition trends, comparing the sequence of the DmCdt1 IDR to the *D. melanogaster* proteome and disorderome can reveal other factors that may be important for LLPS. This was done in a similar fashion to the ortholog analysis, although the Python code was required to parse through all the protein sequences of the proteome. To get a general idea of the scope of the proteome, the pI and kappa values of proteins in the proteome, orderome, and disorderome were found. The upper plot of **Figure 5A** shows the pI values of the proteome which shows a bimodal distribution with most proteins being either negatively or positively charged in the cell to allow for solvation. This bimodal distribution persisted for the orderome proteins, but the disorderome displayed a slightly different distribution (**Figure 5A**, bottom plot). The distribution was still bimodal, but the peaks were centered at different pI values compared to that of the proteome and orderome (pI values of 4 and 10 instead of 6 and 9). The presence of a charge difference in disordered regions compared to ordered ones likely shows that these regions are used differently within the cell. The ordered regions are likely to be the functional domains of proteins which would have a similar charge to that of the proteome while disordered regions are used as spacers or binding domains which may require high charge. However, while the pI showed a difference in values, the kappa values showed a difference in distribution (**Figure 5B**). The distribution of kappa values for the disorderome is wider compared to that of the orderome. This is likely a result of needing very specific charge mixing for ordered domains to function while disordered regions have a wider range of kappas to provide the correct structure for the domain.

An additional analysis of the amino acid composition was also performed on the proteins in the proteome and disorderome and compared to the compositions of the Cdt1 orthologs. Since the number of proteins outnumbers the number of orthologs available for analysis, an equal number of proteins from the proteome and disorderome were randomly selected and compared to the orthologs (**Figure 5C**). For many of the amino acid types, the ortholog and disorderome distributions were incredibly similar with the bulky hydrophobic, aromatic, and negative residues. For the large hydrophobic and aromatic groups, they were less abundant compared to ordered protein domains. This was likely for structural reasons as protein secondary and tertiary structures are driven by the hydrophobic effect for folding to provide a specific function. Differences in composition began to arise in the positive, small hydrophobic, hydrophilic, and proline residues. The difference in positive residues is easily explained as the orthologs are biased to allow for binding to DNA. In the proline and small hydrophobic groups, the orthologs were more

enriched compared to the disordered domains while the orthologs were less enriched in hydrophilic residues. There is potential that these differences exist due to how the regions are structured and what function they need to perform. The large presence of proline and small hydrophobic residues likely provides Cdt1 orthologs with a specific structure different from that of other disordered regions. The lower abundance in hydrophilic residues for the orthologs is harder to explain and may just provide a different protein structure.

The distances between residues were also compared, but only for the large hydrophobic and the positive amino acids. Other amino acids were not compared as the previous LLPS studies have shown that these residues are incredibly important for the process. For the large hydrophobic distances (**Figure 5D**), the average distances are much larger for disordered regions compared to ordered domains. When compared to the value of the DmCdt1 IDR (the red line), the average distance is much shorter compared to other disordered regions. Additionally, there are more large hydrophobic dipeptides in the DmCdt1 IDR than other *Drosophila* disordered regions. These differences in large hydrophobic residue distribution may be important for allowing the DmCdt1 IDR as having the correct distance between residues provides the best scaffold for interactions. For the positive residues, the average distance and number of dipeptides distributions were roughly similar for the proteome, orderome, and disorderome (**Figure 5E**). However, the DmCdt1 IDR has a shorter average distance and more dipeptides than the average disordered protein. Although charge distribution does not have a large effect on DmCdt1 IDR LLPS, the difference in charge may provide a predisposition of the protein to undergo the process in the presence of a large negative scaffold that would not be seen in other disordered proteins (Parker et al., 2021).

DmCdt1 IDR-like designer peptide only shows weak phase separation behavior in the presence of DNA.

From the analysis of DmCdt1 mutants, Cdt1 orthologs, and *D. melanogaster* disordered domains, I was able to make the following conclusions: bulky hydrophobic residues underlie DmCdt1 IDR LLPS, and composition and distance between amino acids are also important because they are conserved between orthologs. These conclusions guided the following design and testing of designer IDR proteins. Since 20 amino acids provided far too much complexity in the design of a random protein, the amino acids were simplified based on the groups of residues previously defined in the text: leucine for large hydrophobics, alanine for small hydrophobics, lysine for positive, aspartate for negative, serine for hydrophilic, and proline. These residues were selected as they have structures like that of other amino acids within those categories allowing for sufficient substitution. A sequence of 300 amino acids with similar composition to the DmCdt1 IDR was then constructed and Python code was written to shuffle the sequence and check if its residue distances and charge distribution were like that of the DmCdt1 IDR. This method eventually output a sequence (designerIDR) like that of the DmCdt1 IDR which was then purified and tested for phase separation ability.

From the depletion assay results, the designerIDR was found to undergo LLPS, but not as strongly as the DmCdt1 IDR (**Figure 6A**). The supernatant of the designerIDR trials was found to have 27.8% ± 22.6% retention of protein while the DmCdt1 supernatant only had 1.2% ± 0.9% retention in the DNA trials. The designerIDR also

showed very little LLPS ability in the presence of crowding reagent with a supernatant retention of 80.0% ± 23.7% compared to 0.26% ± 0.31% retention (**Figure 6B**). The weakening of LLPS with the designerIDR means that the protein is missing a critical interaction type that my analysis did not reveal. An EMSA of the designerIDR was also run which showed that the protein had a similar K_d to that of DmCdt1 at approximately 75 nM (**Figure 6C**). The similarity in binding to DNA once again demonstrates that overall charge is necessary and sufficient. Confocal microscopy images were also collected for DmCdt1 and the designerIDR. This found that droplets were able to be formed by both proteins, but DmCdt1 droplets were rounder than those of the designerIDR (**Figure 6D**). The DmCdt1 droplets were also found to be brighter (**Figure 6E**) but about the same size (**Figure 6F**) compared to the designerIDR droplets. This shows that the designerIDR droplets recruit much less DNA in the formed phases making them much more fluid and weaker despite being the same size. These factors in total show that the mechanism of LLPS for proteins like that of the DmCdt1 IDR has been nearly elucidated.

Discussion

Through my studies, I have shown that the DmCdt1 IDR requires the presence of bulky hydrophobic residues to undergo DNA-induced and PEG-driven LLPS. This data suggests that the process of LLPS for this type of construct is driven by the hydrophobic effect in which the polar solvent drives proteins to aggregate to shield off the hydrophobic residues to increase the overall entropy of the system. Using this knowledge and previous results stating that the positive residues are needed only for DNA binding, the process of LLPS can be summarized as such: Cdt1 is first driven to bind to DNA due to the high positive charge of the IDR (Parker et al., 2021). As more Cdt1 proteins aggregate on the DNA, the hydrophobic regions begin to bury themselves away from the polar solvent causing a liquid droplet of different density to form. It is worth noting that this specific process only occurs *in vitro* as the experiments only tested Cdt1 itself in solution. However, since other DNA replication initiator proteins have been found to undergo DNA-induced phase separation, this same process could occur with those other proteins forming a phase with DNA *in vivo*, a proposition that is supported by findings that the Orc1 N-terminal IDR is required for chromatin binding in cells (Parker et al., 2019).

However, it is likely that the mechanism of Cdt1 phase separation is still incomplete as the creation of a designer IDR using simplified residues based only on the amino acid composition of the DmCdt1 IDR was only able to phase separately weakly with DNA and crowding reagent. A potential reason is that the simplified amino acids were not able to capture a biochemical property seen in the DmCdt1 IDR. One possible factor is the lack of arginine residues because the amino acid can have π-cation interactions with DNA and other aromatic groups. The lack of arginine could also weaken the ability to rearrange and bury hydrophobic residues leading to weaker phase separation. Another issue could arise from the random distribution of alanine and proline amino acids. These residues are incredibly important for determining the overall structure of the IDR and having too much stiffness or freedom of mobility in certain regions could weaken the ability to phase separate. In fact, previous studies have shown that a scrambled DmCdt1 IDR can undergo phase separation, but only weakly recruits wild-

type DmCdt1 for LLPS (Parker et al., 2021). It is also worth mentioning that the importance of hydrophilic residues for phase separation is currently unknown. A significant portion of these amino acids belong to asparagine and glutamine which may affect how the IDR is able to rearrange itself in the hydrophobic effect, but further testing would be required for confirmation. Despite this, the experiments with the designer IDR shows that an abundance of large hydrophobic residues along with positive charge is responsible for some DNA-induced phase separation.

There is also the possibility that there could be some fundamental motif that was missing from the designer IDR that went undiscovered in my analysis of the DmCdt1 IDR. While the dipeptide study seemed to confirm that there were no dipeptide motifs, the creation of randomly shuffled designer IDRs did not have to pass a check for similar dipeptide abundances seen in DmCdt1. This leads to the potential that multiple dipeptides, possibly in pairs or larger groups, could have some role in DNA-induced phase separation. These dipeptide strings could provide a short sequence that gives the IDR a specific structure that may make phase separation stronger with DNA. Once again, a much more in-depth methodology would be required to identify these peptide strings.

My findings also revealed many of the shared characteristics between Cdt1 orthologs regarding their residue composition and their distribution of amino acids. While it is likely that including the DmCdt1 IDR residue distributions in the designer IDR helped the construct undergo weak phase separation, this cannot be fully confirmed. Since previous studies have shown that scrambling the DmCdt1 IDR does allow for phase separation (Parker et al., 2019), a designer IDR with no controls on distance would need to be tested. Perhaps a more interesting finding about the Cdt1 orthologs is how much diversity there is in their IDR sequences. The lengths of these regions have a high spread, and the amino acid composition can vary such as with small hydrophobic residues. It seems as if evolution has constantly changed these properties as phyla began to diverge from one another. If all these proteins were to be found to undergo LLPS with DNA *in vitro*, it would show that evolution has a variety of ways of creating high complexity sequences that are all capable of performing the same task.

While the Cdt1 orthologs showed similarity in important properties, a comparison of the IDRs to disordered sequences in *D. melanogaster* revealed the diversity in these regions. Although I was able to observe a variety of similarities and differences, it is likely the differences between these regions allows for DmCdt1, and potentially its orthologs, to undergo DNA-induced phase separation. The differences in the percent positive composition may allow for any region to bind to DNA, but the presence of closer bulky hydrophobic residues along with a greater number of dipeptides contributes towards the overall hydrophobicity and ability to self-associate. Additional differences in the abundance of small hydrophobic and proline residues for DmCdt1 give the region a different structure compared to other disordered regions that ultimately allows the correct conformation for phase separation. These differences in compositions and distributions may lead to differences in IDR structures, binding affinities, and hydrophobicity that ultimately determine whether a region is capable of association with other proteins for phase separation.

In conclusion, I have been able to find support for previous studies finding that the DmCdt1 IDR DNA-induced phase separation is driven largely by hydrophobics, but that future studies are needed to investigate the abundance or lack of other amino acid

types to fully elucidate the mechanism of action. Additionally, my bioinformatic study of other Cdt1 orthologs and the *D. melanogaster* proteome was able to show that evolution has biased the IDR sequences with specific residue abundances and distributions that give the sequences high complexity, diversity, and potentially the same cellular function. The evidence combined allowed for the creation of a designer IDR that was able to undergo LLPS despite some obvious shortcomings. My studies show that there is potential to design sequences for other high complexity IDRs that undergo phase separation that would allow for a much deeper understanding of the mechanism of the process and how it is important for other cellular processes.

Materials and Methods

Cloning, expression, and purification of DmCdt1, mutants, and designer peptides.

The DmCdt1 IDR, mutant constructs, and designer peptides coding sequences were each cloned into the QB3 Macrolab vector 2Cc-T which includes a tobacco etch virus (TEV) protease-cleavable hexahistidine (His6)–maltose binding protein (MBP) tag. The amino acid sequences of all constructs can be found in the **Supplemental Table**. The coding sequences were generated using IDT's Codon Optimization Tool for the K12 strain of *E. coli.* All sequences were ordered from and synthesized by IDT. The plasmids were transformed into Rosetta 2(DE3)pLysS competent cells (QB3 Macrolab) and expressed at 16℃ overnight after 1 mM IPTG induction. The liquid expression cultures were then centrifuged, and the cell pellets were harvested and frozen at -80℃ for future processing.

All constructs were purified using the following protocol. Cell pellets from 2 L of liquid culture were resuspended in 100 mL of Lysis Buffer (20 mM Tris pH 7.5, 500 mM NaCl, 30 mM imidazole, 10% glycerol, 1 mM BME, 200 µM PMSF, 0.1 mg/mL lysozyme, 1x SIGMA*FAST* protease inhibitor cocktail (Sigma-Aldrich)) and lysed by sonication. The lysate was clarified by centrifugation at 16,000 RPM for 1 hour, filtered through an aPES 0.45 µm bottle-top filter unit (Nalgene Rapid-Flow, ThermoFisher), and passed over a 5 mL His Trap HP column (GE Healthcare). The column was washed with 20 column volumes of High Salt Wash Buffer (20 mM Tris pH 7.5, 1 M NaCl, 30 mM imidazole, 10% glycerol, 1 mM BME, 200 µM PMSF) and 5 column volumes of Low Salt Wash Buffer (20 mM Tris pH 7.5, 150 mM NaCl, 30 mM imidazole, 10% glycerol, 1 mM BME, 200 µM PMSF), and eluted with 5 column volumes of Elution Buffer (20 mM Tris pH 7.5, 150 mM NaCl, 500 mM imidazole, 10% glycerol, 1 mM BME, 200 µM PMSF). The lysate was further clarified by heparin affinity chromatography (HiTrap Heparin HP, GE Healthcare), eluting with a linear gradient of increasing salt from 150 mM to 1 M NaCl. Fractions containing the desired tagged protein were combined and the tag was cleaved with TEV protease (QB3 Macrolab) at a 1:10 w/w ratio of TEV:protein in an overnight incubation at 4℃. An additional nickel affinity step was then used to remove TEV protease, uncleaved protein, and the free His6–MBP tag from the protein of interest. The clarified sample was then concentrated to 2 mL with a 10K Amicon Ultra-15 concentrator (Millipore) and loaded onto a HiPrep 16/60 Sephacryl S-300 HR column (GE Healthcare) which was equilibrated and run in Sizing Buffer (50 mM HEPES pH 7.5, 300 mM K·Glutamate, 10% glycerol, 1 mM BME). Peak fractions were combined, concentrated, aliquoted, flash-frozen in liquid nitrogen, and stored at -80℃.

Pelleting assay for liquid–liquid phase separation.

16 µM stocks of protein and dsDNA as well as a 25% PEG-3350 (Sigma 202444-250G) solution were prepared with protein in Protein Buffer (50 mM HEPES pH 7.5, 300 mM K·Glutamate, 10% glycerol, 1 mM BME) and all others in Dilution Buffer (50 mM HEPES pH 7.5, 10% glycerol, 1 mM BME). Two samples were created by mixing the protein with dsDNA and the protein with PEG in a 1:1 volumetric ratio and incubated for 30 minutes at room temperature. Following incubation, the samples were centrifuged for 10 minutes at 13,000 xg. Approximately 75% of the supernatant volume of each sample was removed and diluted 4-fold in Assay Buffer (50 mM HEPES pH 7.5, 150 mM K·Glutamate, 10% glycerol, 1 mM BME). These samples were analyzed by SDS-PAGE and Coomassie staining. For each sample, an uncentrifuged protein load and a centrifuged protein-only sample were used as controls.

Assaying DNA binding using EMSA.

A 2-fold serial dilution of a 10 µM protein stock in Protein Buffer was performed to a 1.2 nM concentration. These samples were combined with an equal volume of 20 nM IRDye800–dsDNA in Dilution Buffer and incubated at room temperature for 30 minutes. Following incubation, the samples were analyzed by running on a 7.5% PAGE gel (BioRad 4561026) pre-run at 100 V for 1 hour and visualized using an Odyssey imaging system (LICOR). The concentration at which the band shifted was identified to approximate the K_M.

Assaying phase separation by fluorescence microscopy.

10 µM stocks of protein and Cy5-dsDNA (IDT) were prepared in Protein Buffer and Dilution Buffer, respectively. Samples were made by thoroughly mixing an equal volume of protein and Cy5-dsDNA and incubated for 2 minutes. 7 µL of the sample were spotted onto a glass slide, a coverslip was placed on top, and the sample was imaged with a 60x oil objective using confocal fluorescence microscopy. Image processing and quantitation was done using FIJI by thresholding to define droplets and analyzing the intensities.

Identification and analysis of Cdt1 orthologs.

Orthologs of Cdt1 were found using the UniProt database (https://www.uniprot.org/) and completed amino acid sequences were saved. These sequences were imported into IUPred2A (https://iupred2a.elte.hu/) which calculates the disorder of the sequence. A 20-residue running average was performed on the disorder data to set up the boundaries of each Cdt1 ortholog IDR. These finalized IDR sequences were then stored as strings in a Python dictionary for further processing (Python 3.8.5). The code can be found at https://github.com/alvinhuang617/AlvinHuangSeniorThesis2021.

Analysis of the number of Cdk sites, amino acid percentages, and distances between residues was done by parsing through the IDR amino acid sequence strings. To

determine the number of Cdk sites, code was written to parse the string until it detected an ‘S’ or ‘T’. It then checked if the next three residues followed the pattern of ‘PX(R/K)’ where ‘X’ is any amino acid. If this was satisfied, the code added one to the running counter of the number of Cdk sites for the IDR analyzed. If the pattern were not followed, the code would move on to the next residue. A successful hit caused the code to skip the next three residues. This was done for every IDR in the dictionary, and the data was stored in a dictionary with the organism’s name as the key and the count as the value. This dictionary was converted into a Pandas DataFrame, and Seaborn and Matplotlib were used to graph the scatter plot in **Figure 3A**.

The grouped amino acid percentages were found by providing the code with a list of specific residues (e.g., large hydrophobics were [‘L’, ‘I’, ‘V’]). While the code parsed through the IDR sequence, if the residue was in the specified list, it added 1 to a running counter. At the end, the code divided the count by the length of the IDR to generate the percentage. This data was then compiled into a dictionary and processed using Pandas and Seaborn to plot the violin plots in **Figure 3D**.

The distance between amino acids was found by first specifying a group of residues of interest like that of the amino acid percentage code. The code started a count at 0 and parsed through the sequence until the first residue of interest was found. Every time a residue of interest was found, the current count was added to a list, and the count was reset back to 0. The first count was not kept as it represented the distance to the first instance of a residue. Parsing and counting continued until the final residue was reached. If it was not a residue of interest, this final count was not kept; otherwise, it was kept. The data for every IDR sequence was compiled as a dictionary of lists. More code took in this dictionary and noted the total number of zero counts, nonzero counts, and the length of the IDR. By finding the total number of zero and nonzero counts, these values were calculated: the number of zero counts per 100 amino acids in the IDR sequence and the average value of the nonzero distances. These values ended up being the number of adjacent amino acids per 100 residues and the average nonadjacent distance, respectively. This data was stored in a dictionary and once again processed using Pandas and Seaborn to plot the violin plots in **Figure 3F** and **3E**, respectively.

To find the pI of the sequences, the library Biopython (https://biopython.org/) was used. Kappa values were found by importing the Pappu Lab’s CIDER tool into Python 3 (http://pappulab.wustl.edu/CIDER/analysis/). Graphs were generated by processing the data using Pandas and Seaborn for **Figure 3A** and **3B**, respectively.

Identification of repetitive motifs within the DmCdt1 IDR.

The code used to generate this plot was based on the work of Nott et al. and their analysis of repetitive dipeptides in human Ddx4. Python code was written to calculate the percent composition of every amino acid in DmCdt1. The method was very similar to finding the percentage of specific amino acids seen in the IDR analysis, but this time for all 20 amino acids. For each residue found through parsing through the string, the code added 1 over the length of the IDR to the respective key in a dictionary. If any amino acids were

not present in the sequence, they were added after parsing with a value of 0 to the dictionary. The probability of finding each dipeptide was found by multiplying the percentage of one residue with another. This was done for every permutation of dipeptide, and the data was stored in a nested dictionary: the main key was 'DmCdt1' with a dictionary of every dipeptide with the associated probability as the value. The number of dipeptides was then counted by parsing through the sequence and storing each dipeptide in a dictionary with their associated counts. The code parsed the string by moving up a single residue at a time to capture all possible dipeptides. Any dipeptides that were not counted were added to the dictionary with counts of 0. The counts were then converted to observed "probabilities" by dividing each count by the total number of dipeptides. The abundance ratio comparing the observed to predicted probabilities was found dividing the observed by the expected for each pair of dipeptides. This data was then processed into a Pandas DataFrame and the resulting scatter plot in **Figure 4** was made using Seaborn.

Analysis of the Drosophila proteome and its ordered and disordered domains.

The *Drosophila* proteome, its orderome, and its disorderome were downloaded as .pkl files from the UniProt database (https://www.uniprot.org/proteomes/UP000000803). The Pickle library was then used to unpack the information as either a Python dictionary or nested dictionary. The proteome was in a dictionary with the protein names as keys and the values were the amino acid sequences. The orderome and disorderome were stored as nested dictionaries where the values were dictionaries containing the sequence of each ordered or disordered region. All code used in this section processed these dictionaries to allow for parsing of the sequence strings. Only ordered and disordered regions that were longer than 100 amino acids were processed to avoid short segments.

The code for finding the pI, the kappa values, and the amino acid percentages and distances were found using the same code as seen in the IDR analysis, albeit with slightly different dictionary processing. The strip plot seen in **Figure 5C** was created by randomly selecting 20 *Drosophila* proteome and disordered sequences that were between 150 and 350 residues. Additionally, the percent composition data of the IDRs was used. Random selection was done using Numpy and randomly selecting keys from the dictionaries. For the nested dictionary, when a key was randomly selected, any disordered sequence that obeyed the length rule was used. This data was processed using Pandas and graphed using Seaborn.

Generation of designer peptides.

To generate a designer IDR, first each amino acid group was simplified to a single residue as follows: L for large hydrophobic (including aromatics), A for small hydrophobic, S for hydrophilic, K for basic, D for acidic, and P for proline. Additionally, the percent composition and distance data for the DmCdt1 IDR was collected from the IDR data using code described in that section. The amino acid percentages determined how many residues of each type were needed for a sequence of 300 amino acids, and the count for each residue was put into a dictionary. Code was written that took the counts of

each residue and formed a 300-residue long string. This sequence was then shuffled using the Numpy Random library to create a potential designer peptide. The first check determined if the kappa value fell between 0.197 and 0.201 to be like the DmCdt1 IDR. If the kappa was acceptable, the distance between nonadjacent residues and the number of dipeptides per 100 residues were then calculated for each sequence and compared to the DmCdt1 IDR values. If it fell within the threshold of +/- 20%, the sequence would be printed out as a potential designer peptide.

Figures

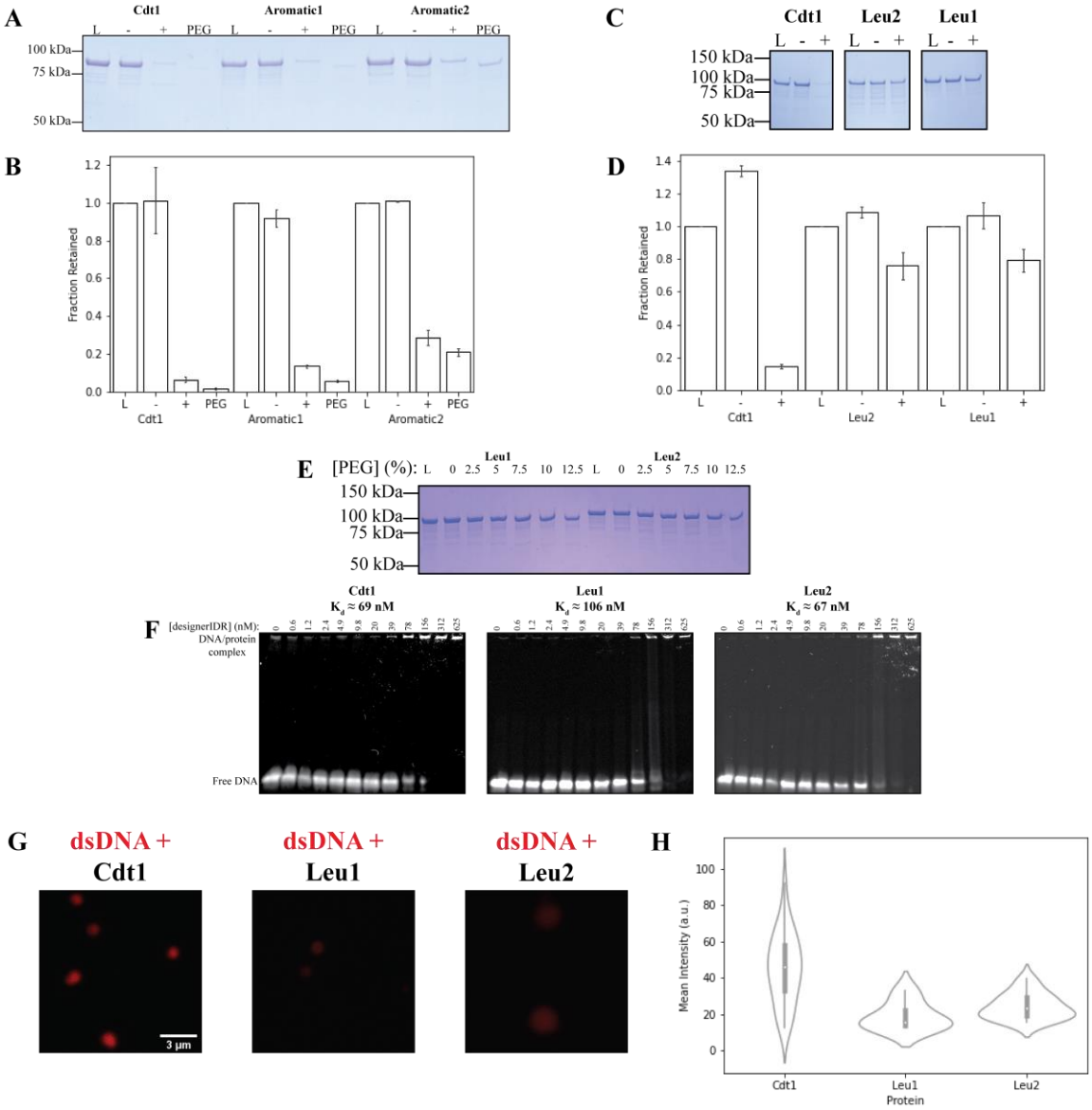

Figure 1. Aromatic residues are not needed while bulky hydrophobic residues are indispensable for LLPS. (**A**) Depletion assay results assaying wild-type DmCdt1, Aromatic1 (all Phe to Ala), and Aromatic2 (all Phe to Leu). "L" is the load control, "-" is the absence of DNA, "+" is the addition of DNA, and "PEG" is the presence of polyethylene glycol 3350 (PEG). (**B**) Quantification of depletion assay results seen in

Figure 1A. (**C**) Depletion assay results assaying wild-type DmCdt1, Leu2 (half of Leu and Ile to Ala), and Leu1 (all Leu and Ile to Ala). “L” is the load control, “-“ is the absence of DNA, and “+” is the addition of DNA (**D**) Quantification of depletion assay results seen in **Figure 1C**. (**E**) Depletion assay results assaying Leu2 and Leu1 with an increasing gradient of PEG. “L” is the load control and all other values represent the percent of PEG in solution. The first 7 lanes represent Leu1 while the next 7 are for Leu2 (**F**) Electrophoretic mobility shift assays (EMSAs) of wild-type DmCdt1 (left), Leu1 (middle), and Leu2 (right) to test DNA binding ability. Binding was assessed with 10 nM IRDye800-labeled dsDNA and a titration of Cdt1 from 0.6 to 625 nM. Each mutant showed comparable binding strength with wild-type DmCdt1. (**G**) Confocal fluorescence microscopy was used to assess DNA-dependent phase separation by combining wild-type DmCdt1 (left), Leu1 (middle), and Leu2 (right) with Cd5-dsDNA. Images were all taken at the same laser power. (**H**) Quantification of the average intensity of the droplets in **Figure 1G**.

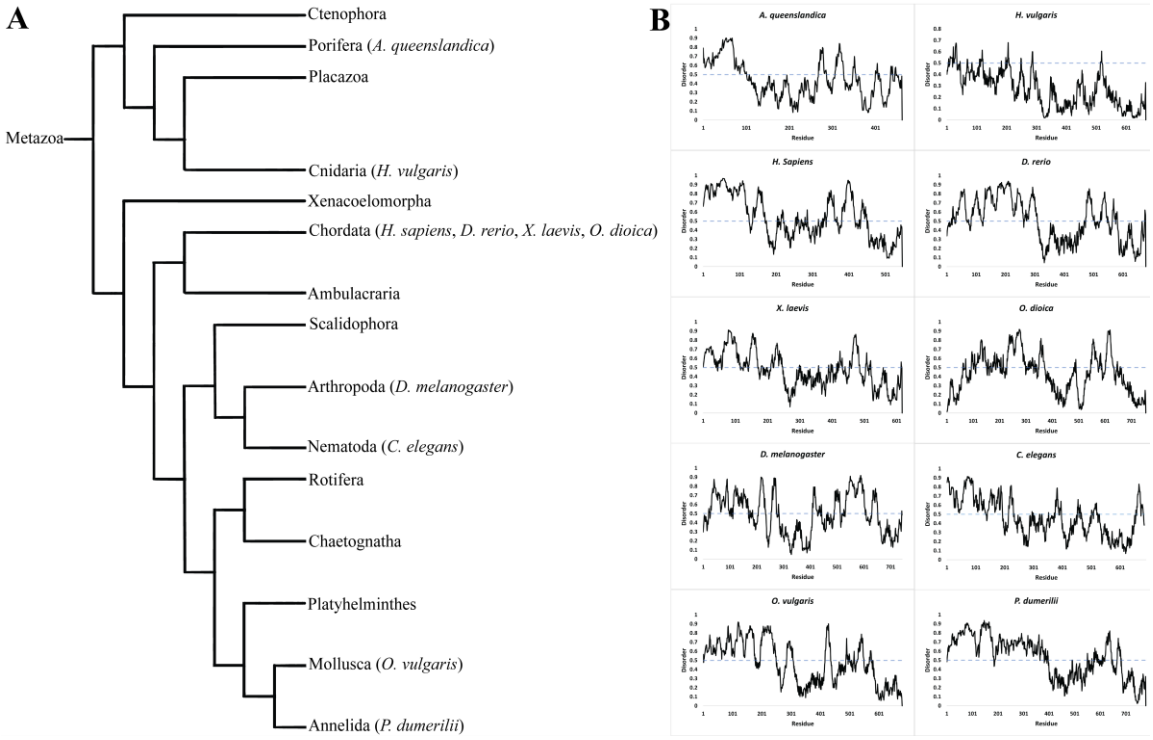

Figure 2. Evolutionary relationship between metazoan phyla and the disorder of representative metazoan Cdt1 orthologs. (**A**) Phylogenetic tree of the phyla of metazoan phyla. Figure was constructed using information from Parfrey et al. (2011). (**B**) Disorder plots of 10 representative metazoan orthologs of Cdt1 generated by the IUPred2A algorithm. Each plot shows the presence of an N-terminal intrinsically disordered region.

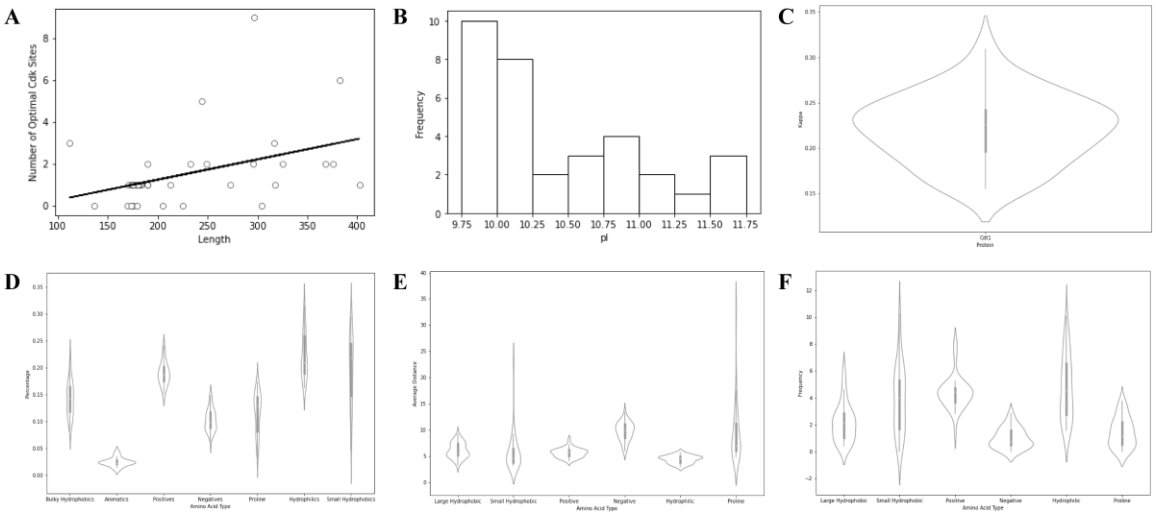

Figure 3. Cdt1 orthologs are sequentially dissimilar but share similarities in the amino acid composition and distribution of residues. (**A**) Scatterplot of the number of Cdk sites versus the length of each Cdt1 ortholog IDR. Orthologs tend to have around 0 to 3 Cdk sites despite the length of the protein and shows a weak positive linear trend (r = 0.4) (**B**) Histogram of the pI values of the Cdt1 ortholog IDRs. The pI values are distributed between 9.75 and 11.75 showing overall positive charge. (**C**) Violin plot of the kappa values of the Cdt1 ortholog IDRs. The values are distributed about a value of 0.24 showing separation of charged amino acids. (**D**) Violin plots of the composition of bulky hydrophobic, aromatic, positive, negative, proline, hydrophilic, and small hydrophobic residues. Distributions for the first 4 groups are relatively tight while others show greater variability. (**E**) Violin plots of the average distance between the large hydrophobic, small hydrophobic, positive, negative, hydrophilic, and proline residues. The distance distributions are relatively tight except for small hydrophobic and proline residues which showed greater variability. (**F**) Violin plots of the number of dipeptides for large hydrophobic, small hydrophobic, positive, negative, hydrophilic, and proline residues. The distance distributions are relatively tight except for small hydrophobic and hydrophilic residues which showed greater variability.

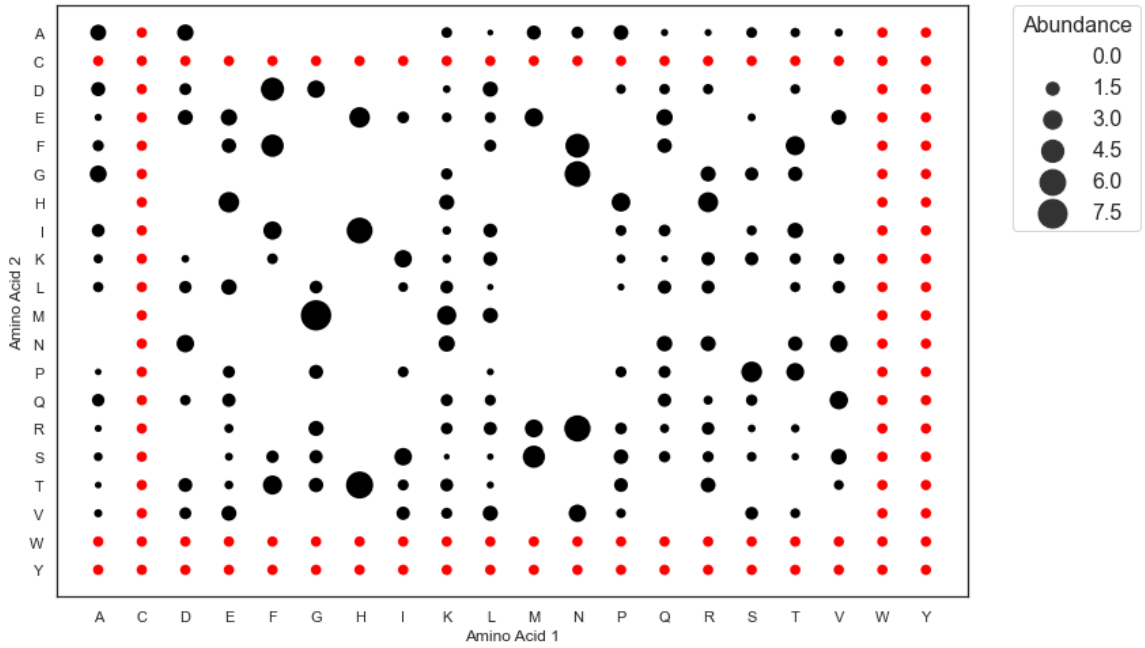

Figure 4. Scatterplot measuring the abundance of all dipeptides in the DmCdt1 IDR. Blank spots refer to missing dipeptides in the IDR while larger points show that these dipeptides are enriched in the IDR. Points that are marked red refer to dipeptides that do not exist in the IDR, but technically have abundance values of 1 as a result. Although some dipeptides were more abundant, they were not of sequence importance due to their low constituent amino acid composition.

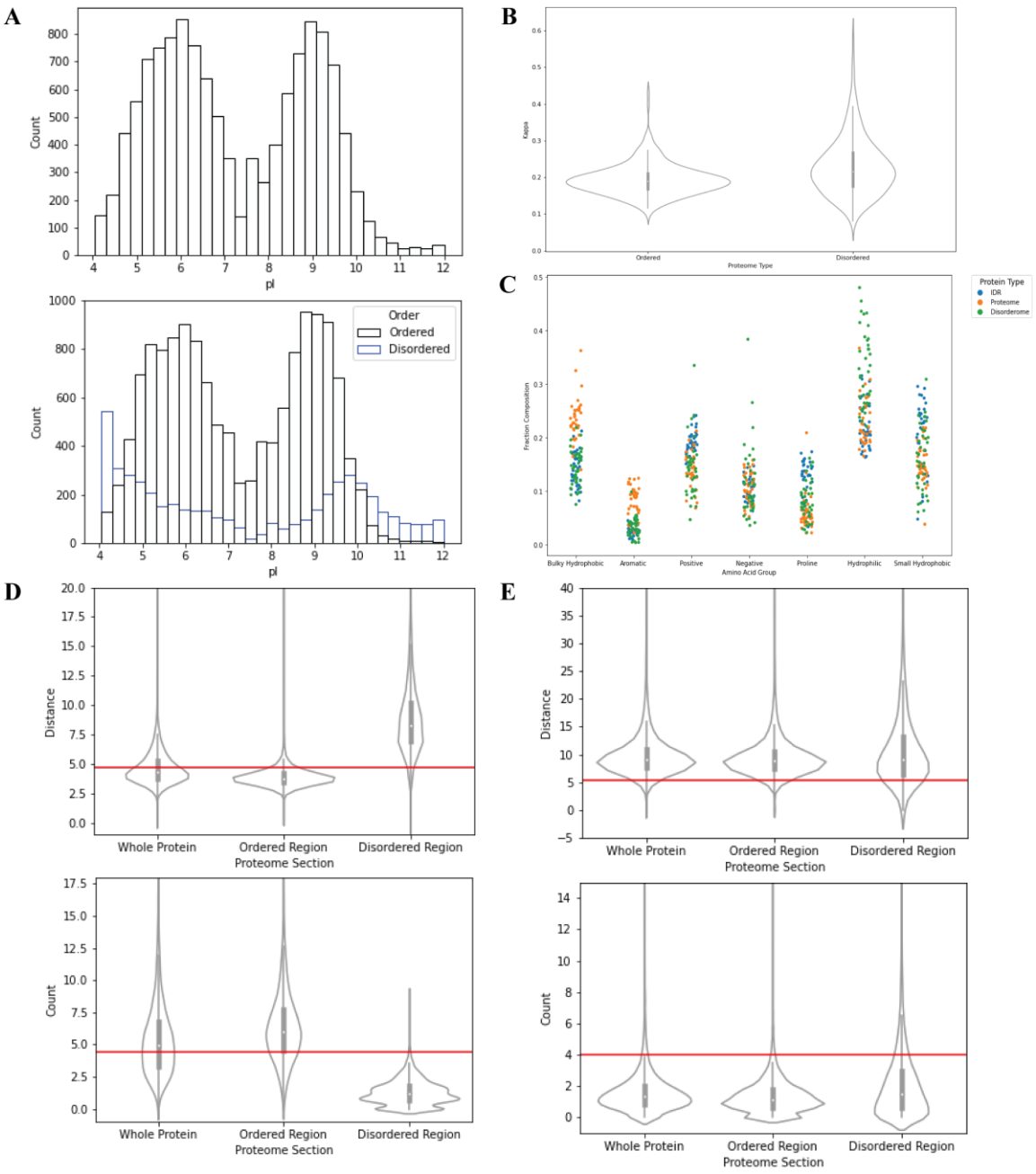

Figure 5. Analysis of the properties of proteins in the *Drosophila melanogaster* proteome, orderome, and disorderome reveal differences between the IDR and other disordered proteins. (**A**) Histogram of the pI values of proteome proteins (top) and histogram of the pI values of orderome and disorderome proteins (bottom). Bimodal distributions of the pI for the proteome and orderome are roughly the same, but the bimodal distribution of the disordeome has peaks at different pI values. (**B**) Violin plots of the kappa values of 250 randomly selected proteins from the orderome and disorderome. The distributions are centered at roughly similar values, but the distribution of the orderome kappa values is tighter. (**C**) Strip plots of the distribution of the composition of bulky hydrophobic, aromatic, positive, negative, proline, hydrophilic, and small hydrophobic residues of the Cdt1 IDR orthologs (blue), randomly selected

proteome proteins (orange), and randomly selected disorderome proteins (green). Distributions of the disordered regions and the proteome are different from one another. Differences between the disordered proteins arise in the positive, negative, proline, hydrophilic, and small hydrophobic residues. (**D**) Violin plots of the distribution of the average distance between large hydrophobic residues (top) and the number of large hydrophobic dipeptides (bottom) all compared with the value for the DmCdt1 IDR. The distance between large hydrophobic residues was found to be shorter for the IDR with many more dipeptides when compared to the disordered regions. (**E**) Violin plots of the distribution of the average distance between positive residues (top) and the number of positive dipeptides (bottom) all compared with the value for the DmCdt1 IDR. The distance between positives was found to be about the average for the IDR with many more dipeptides when compared to the disordered regions.

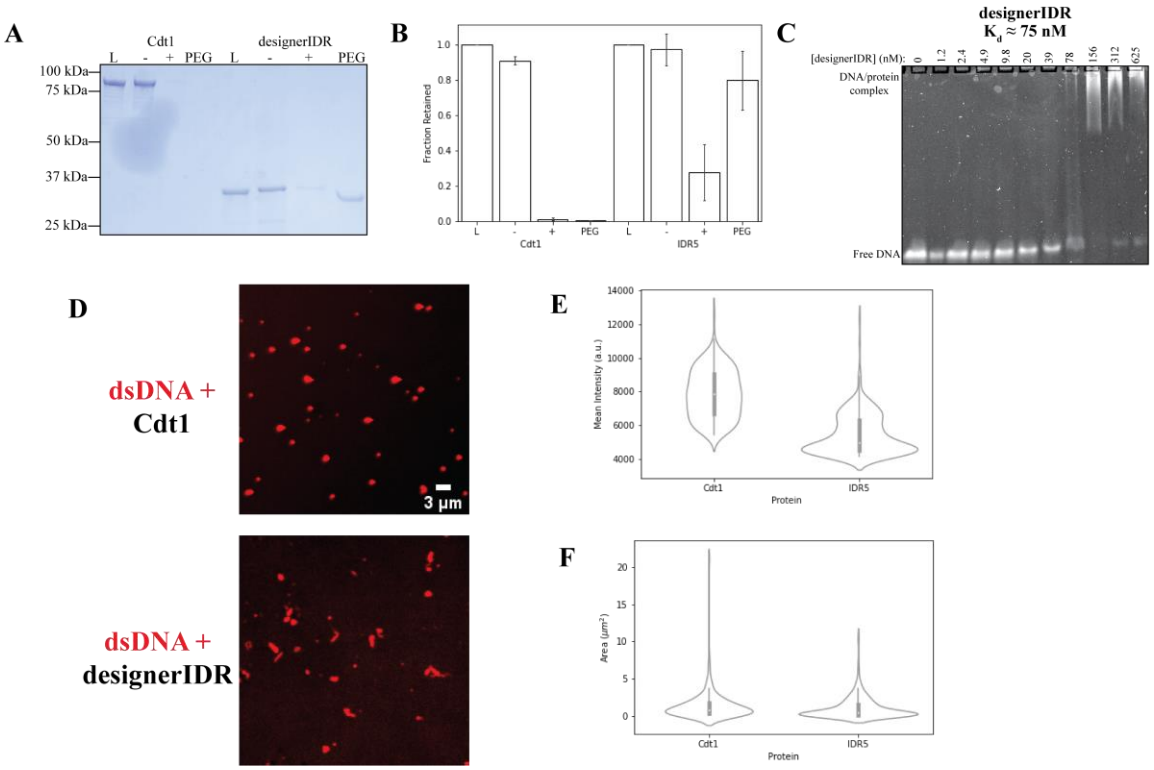

Figure 6. The designed IDR can weakly phase separate in the presence of DNA and produces weaker droplets but has similar binding to DNA. (**A**) Depletion assay results assaying wild-type DmCdt1 and designerIDR. "L" is the load control, "-" is the absence of DNA, "+" is the addition of DNA, and "PEG" is the presence of polyethylene glycol 3350 (PEG). (**B**) Quantification of depletion assay results seen in **Figure 6A**. (**C**) Electrophoretic mobility shift assay (EMSA) of designerIDR to test DNA binding ability. Binding was assessed with 10 nM IRDye800-labeled dsDNA and a titration of designerIDR from 0.6 to 625 nM. The mutant showed comparable binding strength with wild-type DmCdt1. (**D**) Confocal fluorescence microscopy was used to assess DNA-dependent phase separation by combining wild-type DmCdt1 (top) and designerIDR (bottom) with Cd5-dsDNA. Images were all taken at the same laser power. (**E**) Quantification of the average intensity of the droplets in **Figure 6D**. (**F**) Quantification of the area of the droplets in **Figure 6D**.

Supplemental Table.

Construct Name	Amino Acid Sequence
Drosophila melanogaster Cdt1	MAQPSVAAFFTNRKRAALDDAISIKNRRLVEPAETVSPASAP SQLPAGDQDADLDTLKAAATGMRTRSGRTARLIVTAAQESK KKTPAAAKMEPHIKQPKLVQFIKKGTLSPRKQAQSSKLDEEE LQQSSAISEHTPKVNFTITSQQNADNVQRGLRTPTKQILKDAS PIKADLRRQLTFDEVKTKVSRSAKLQELKAVLALKAALEQK RKEQEERNRKLRDAGPSPSKSKMSVQLKEFDTIELEVLISPLK TFKTPTKIPPPTPDKHELMSPRHTDVSKRLLFSPAKNGSPVKL VEVPAYKRYASLVESSRAGQLPLPYKYRHLLDVFKGLDSVV AMFHNRKETITFKKLKPAVQRMLRKNFTETHLAQIKHIYPDA FIFSQVKTRNFGSVSKADYFQLIIAPNVEPLPEQQQSEKPQHF TKINEDDVLASAQSTSMNPHVMTARMQRFQSLLLDRAMRA HDQFLRSQDPPIIIEKALTRWHPQFDLESCPEVELSPLPQPPNV EKYSSAKDILSTARNLFNCATPMERAMDRYEAKLESEKQQA AESNKKTEEQQAGEVTRTSTAIQTSQEVPGISGSSKNPTVPET TNQTETAKPTVKDCTVPDASSNLLKGLPKSLIEKIRAKQAAK ALEAMTRRPSQDQEATKYSRLPELARHLRNVFVTERKGVLT LEVIIKKIQNSFRANLTPQEIEAHLKLLAKELPSWASFHEVRK TMYLKVAKDMDMNKIIEKLESVANAKSN
Drosophila melanogaster Cdt1 IDR	MAQPSVAAFFTNRKRAALDDAISIKNRRLVEPAETVSPASAP SQLPAGDQDADLDTLKAAATGMRTRSGRTARLIVTAAQESK KKTPAAAKMEPHIKQPKLVQFIKKGTLSPRKQAQSSKLDEEE LQQSSAISEHTPKVNFTITSQQNADNVQRGLRTPTKQILKDAS PIKADLRRQLTFDEVKTKVSRSAKLQELKAVLALKAALEQK RKEQEERNRKLRDAGPSPSKSKMSVQLKEFDTIELEVLISPLK TFKTPTKIPPPTPDKHELMSPRHTDVSKRLLFSPAKNGSPVKL VE
Aromatic1	MAQPSVAALLTNRKRAALDDAISIKNRRLVEPAETVSPASAP SQLPAGDQDADLDTLKAAATGMRTRSGRTARLIVTAAQESK KKTPAAAKMEPHIKQPKLVQLIKKGTLSPRKQAQSSKLDEEE LQQSSAISEHTPKVNLTITSQQNADNVQRGLRTPTKQILKDAS PIKADLRRQLTLDEVKTKVSRSAKLQELKAVLALKAALEQK RKEQEERNRKLRDAGPSPSKSKMSVQLKELDTIELEVLISPLK TLKTPTKIPPPTPDKHELMSPRHTDVSKRLLLSPAKNGSPVKL VE
Aromatic2	MAQPSVAAAATNRKRAALDDAISIKNRRLVEPAETVSPASAP SQLPAGDQDADLDTLKAAATGMRTRSGRTARLIVTAAQESK KKTPAAAKMEPHIKQPKLVQAIKKGTLSPRKQAQSSKLDEEE LQQSSAISEHTPKVNATITSQQNADNVQRGLRTPTKQILKDA SPIKADLRRQLTADEVKTKVSRSAKLQELKAVLALKAALEQ KRKEQEERNRKLRDAGPSPSKSKMSVQLKEADTIELEVLISPL KTAKTPTKIPPPTPDKHELMSPRHTDVSKRLLASPAKNGSPV KLVE
Leu1	MAQPSVAAFFTNRKRAAADDAASAKNRRAVEPAETVSPAS APSQAPAGDQDADADTAKAAATGMRTRSGRTARAAVTAA

	QESKKKTPAAAKMEPHAKQPKAVQFAKKGTASPRKQAQSS KADEEEAQQSSAASEHTPKVNFTATSQQNADNVQRGARTPT KQAAKDASPAKADARRQATFDEVKTKVSRSAKAQEAKAVA AAKAAAEQKRKEQEERNRKARDAGPSPSKSKMSVQAKEFD TAEAEVAASPAKTFKTPTKAPPPTPDKHEAMSPRHTDVSKRA AFSPAKNGSPVKAVE
Leu2	MAQPSVAAFFTNRKRAAADDAASIKNRRLVEPAETVSPASA PSQAPAGDQDADLDTAKAAATGMRTRSGRTARLAVTAAQE SKKKTPAAAKMEPHIKQPKAVQFAKKGTLSPRKQAQSSKAD EEELQQSSAISEHTPKVNFTATSQQNADNVQRGARTPTKQAL KDASPIKADARRQLTFDEVKTKVSRSAKAQELKAVAALKAA AEQKRKEQEERNRKLRDAGPSPSKSKMSVQAKEFDTAELEV AISPLKTFKTPTKAPPPTPDKHEAMSPRHTDVSKRLAFSPAKN GSPVKLVE
Callithrix jacchus Cdt1 IDR	MLRLVGSARAGGRGRGARGKFGGNCARLFLSCFFLAFPTLA ARSAAMEQRRVTDFFARRRPGPRIAPPKPDCRTPSPAPRALD PATSGSRKRARPSAVPGRDQARPPARRRLRLSADAVSSPSSP GSPREAPDSPEAPGSPAGPSLCQKITTSAPAAGRPGQKDTISE LRSCLQRARQLGARVRALKASAHEAGESRTPEAEGRPEEPC GEKAPAYQRFHALAQP
Cavia porcellus Cdt1 IDR	MAQSRVTDFFLCRRPGSVAPRIKAACRTPSPTGPSPALAPGSS RKRARPAVAPGVEQPAPPARRRLRLRDADSCPGSPTPPSPPA CHLLPKKVKKISAPVPAPAARRSVCAAPPDKVPSKDTVSELQ ACLQRARELGACVQALKDRAQEVAEEPSSPEATAPAEQPCG EK
Bos taurus Cdt1 IDR	MAQRRLTDFFARRRPGGSATLPRAKPAWRTPSPAKSAPCAA APGPGSSRKRTRPPAEPTRDERVPPARKRLKLSADAVSGPSSP AARGSPEHPSPQSKETKKAARSAGGRPCLAAQENKVPSKVT LSELKSCLQRAQELGARFQELKASAQKDAGEPSAPEDEGRLE EPCGEQMP
Cricetulus griseus Cdt1 IDR	MAQSRVTDFYPRRRPGLAAPRAKSACLTPSPSGLVTPQFIRSS SRKRARPSTEPRSDQAAPPARRRLRLPGLDSCPSSPVDPSSPA DPSSPVSPSPVKKIKSTTVSEGPGLSATAQQRDKVPSKDTISEL QSCLRRARKLGAQARALKARIQENAVESSTSDAKVSTEQPC
Chlorocebus sabaeus Cdt1 IDR	MEQRRVTDFFARRRPGPHIAPPKLACRTPSPARPALRASDSA TSGSRKRARPPAAPERDQARPPARRRLRLSADAVSSPSSPEAP DSSKAPDIPACPSLGQKIKKSTSAAGQAPQLTSSQNQDTISEL ASCLQRARELGARVRVLKSSAQDAGESCTPEAKGRPEEPCG EKAPAYQRFHALAQ
Homo sapiens Cdt1 IDR	MEQRRVTDFFARRRPGPPRIAPPKLACRTPSPARPALRAPASA TSGSRKRARPPAAPGRDQARPPARRRLRLSVDEVSSPSTPEAP DIPACPSPGQKIKKSTPAAGQPPHLTSAQDQDTISELASCLQR ARELGARVRALKASAQDAGESCTPEAEGRPEEPCGEKAPAY QRFH
Ursus maritimus Cdt1 IDR	MWRAALPRRSMLASRVWPGAHTPPQTCHTPSLKHAGCHLP PCSLGGCCSPSTGVAARCPHRRRELPPPGTLPPPRGSLTGHPG

	APRVPAAEFSEAACPAAGSPAPGTRSSASSKLIFSELKSCLQR ARELGARAQELRASAQRKDAGEPSVPEDQGQPAGPCGEKAP AYQRFHALA
Gallus gallus Cdt1 IDR	MAQLRLTDFFGRTKAAVSLPAKHGARGLKAALVAPPEPRNH GAVSPPPASPRTPTRSAAPAVPRMAGRKRSRREMEAETPGGE RGGKSARKRLELPRDAGPEPASPTSLGPPQPPIASRLCTPSPEQ DSGTTPRLEQEELATLQSRLQRMRMKPPTQPPPIPAGVGAEL RSRLQSLRGLELRLRAKAAGSGTGQPGVTGTEAPIAAQNSST QL
Loxodonta africana Cdt1 IDR	MAQRRVTDFFARRRRGPPACLPRAKQALRTPSPAKPAPGVP APALSDPPGGSRPVGGEGGETEARAGGGRPRGRGTSGVPGV LTLAHCGSFPTQVPSPSSSVASGSPAHSPLSKKIKKATIPAGQL PCLAAQGDKISRKDTVSELQACLQRARELGARVQALRAHAQ EEDAEKSAMPEPERGPVEPCGEK
Myotis lucifugus Cdt1 IDR	MAQRRVTDFFLRRRPGPRAASLRAKPNWRTPSPAKSALRAQ VPTSGGSRKRARPSAEPTSDEPPRDEPTRDQPAPTARRRLRLP DDVVPGPGSPATSGSPEHSSPSKKTKKAALSGQLPKPAPQKD KVSSKVTLSELKSCLQRARELGARAQELRASAPGNDAGEPS VLEAEGHPAEPCEEKAPAYQRFH
Pelodiscus sinensis Cdt1 IDR	MAQLRVTDFFSHSKNPARDGPLGAKRRKTQPAAGAVGETR RAGRCKSRPKAAPLTPSRIPGGSDYEPRVSLAGASPGAKLIGF GDKAVGSATAGLPPLSLQQQQLLASPRTPVRPDPGSPQPPAR TLTGRKRCRQEPDPDLEAERNPGPSATLRPQENSKRAARKKL VLPSDGEPQTTAQAAPVTVSSLSSPAALRPPTPSSLDKKVKNL VNVTLSPGPESGLPSGLAALEGNISFKQVSSKELTQRLQRLHA LTRKAKLPACPSETCTDLKSRLKQVQELGSKIRNRKAEEGKE SGLQPAKETCEPAAELSEKAPAYERFHTLA
Macaca mulatta Cdt1 IDR	MEQRRVTDFFARRRPGPRIAPPKLACRTPSPARPALRAPDSA TSGSRKRARLPAAPGRDQAGPPARRRLRLSADAVSSPSSPEA PDSLEAPDIPACPSLGQKIKKSTPAAGQPRQLTSSQNQDTISEL ASCLQRARELGARVRVLKSSAQDAGESCTPEAKGRPEEPCG EKAPAYQRFHALAQ
Xenopus laevis Cdt1 IDR	MSQMRVTDFFSQSKRGTAAQNSKGRKVEAVLETRRAVTRS RAASVKAEEFLKAPCTPERASPTVSQCIGPSSKKRTRQDSDSE PLRTQQRQGKSARKKLKLPEGEHGGSVQQQLFSPCNKVALE HVTPSSLGKKIKDMVNVSLSPKFNELARNPTTPETKSPAKEN LLELKSRLQRIQELAQKVNLPAASSEGKVTITDLKARLKRAQ ELDTKIRAKAEKTETQAIDLTEQPAQESEKAPAYQRFHNLA
Macaca nemestrina Cdt1 IDR	MEQRRVTDFFARRRPGPRIAPPKLACRTPSPARPALCAPDSA TSGSRKRARPPAAPGRDQAGPPARRRLRLSADAVSSPSSPEA PDIPACPSLGQKIKKSTPAAGQPRQLTSLQNQDTISELASCLQ RARELGARVRVLKSSAQDAGESCTPEAKGRPEEPSGEKAPA YQRFHALA
Rattus norvegicus Cdt1 IDR	MAQSRVTDFYACRRPGLAIPRAKSTCLTPSPGGLVAPEFTRSS SRKRVRPHAEPGSDQPAPPARRRLRLPGLDSCPSFPADPSSPA DPSSPADPSSSVCPSPVCPSPVKRTKSTTVSRGPQQSKVPSQD

	SISELQSCLRQARKLGAQARALRARVQENVVEPSTPDTKEPT EQPCVDKAPA
Anolis carolinensis Cdt1 IDR	MAQLRVTDFFPQNKAGSSSASSRLRRKGSQPEQEPWTPAGG RKRARTESMGRGSARKRLLMAPPEGEQAAAPGPCIAPPLGP HSPSGLDKKVKDLTNVILSPGLKPETEAEGRTPAEKETSKEK LTELRSRLDKIRALTQKARSPKGSSETGVDLKSRLKRVRLLES KIRERKGSKLGKEDLGKKDLEPSGESSENVPAYQRFHA
Equus caballus Cdt1 IDR	MRPGRPCPASASARGGAGLGCGAKIGGNRGTTSARRPAGRS RPVPSGPAMAQRRVTDFFARRRPGPCAAPSRAKPAWRTPSP AKPAARAPAPAPGSRKRARPPAEPPCRESAPPARRRLQLPAE EVSGPRSPAAHDSPEHSSPSKKTKKAAASAGPPSCLASEEGK VPSKVTFSELKSCLRRARELGTRVQELRTSAQRKDAEEPSAP EAEGPPVGPCGEKAPAYQRFHALAQ
Gopherus agassizii Cdt1 IDR	MAQLRVTDFFSHRKKPARDGALGTKRRKARPAGAVAGGKR RAARSESRPNAAPPAIPGGSGHEPRASRREASPGAKLLGLEN DHRAAVSAAPALPPLSPHQQQLLASPRTPRRPDAGSPQPPAR TLTGRKRGRQELDLDRAAERNAGPPAGLQPQGNSRGAARK KLVLPADGDPQGTEQEAAVTVGTSLK
Otolemur garnettii Cdt1 IDR	MEQRRVTDFFAHRRLGPRAVPPRAKLPCRTPSPAEHAPRAPG SAPSGSRKRARPPADPGQPKAPPPARRRLRLSADAGETYSLG SPADLDSSADPRPRACSSLGTKVQKVTVSADQLSSPAASQDK VPPTDTISELQSCLQRARELGARVQALRAKKDAESLCPPETE HHAEPPCGEKL
Cebus imitator Cdt1 IDR	MEQRRVTDFFARRRPGPRIAPPKLACRTPSPAPCALDPATSGS RKRARPSAAPGRDQARPPARRRLRLSADAVSSPSSPEAPGSP AGPSLCQKIKTSAPAAGQPPHPAARRHQDTISELRSCLQRAQ QLGERVRALKASAQDAGEPRTPEAEGSPAEPCGEKAPAYQR FHALAQ
Capra aegagrus Cdt1 IDR	MGTANDPRPSEWRMRGRRSPSESGAGGPARMRPGASISRQL PWPWCQWSRPWRGRSQGAARKSAGSAAPLPPPAHRSAPAA SDPASPVMAQRRLTDFFARRRPGVSTTLPRAKPAWRTPSPAK SAACAAAPGPGSSRKRARPPAEPTRDERVPPARKRLRLPADA VSGPSSPPAPVSPEHPSPQSKETKKAARSAGGPPCLAAQENK VPSKVTLSELKSCLQRAQELGARVQELKASAQKDAGEPSAP EDEGRLEGPCGEQMPAYQRFHALAQ
Papio anubis Cdt1 IDR	MEQRRVTDFFARRRPGPRIVLPKLACRTPSPARPALRAPDSA TSGSRKRARPPAAPGRDQARPPARRRLRLSADAVSSPSSPEA PDIPACPSLGQKIKKSTAAAGQPPQLTSSQNQDTISELASCLQ RARELGARVRVLKSSAQDAGESCTPEAKGRPEEPCGEKAPA YQRFHALAQ
Pan troglodytes Cdt1 IDR	MEQRRVTDFFARRRPGPPRIAPPKLACRTPSPARPALRAPDSA TSGSRKRARPPAAPGRDQAGPPARRRLRLSVDEVSSPSTPEA PDSPEAPDIPACPSPGQKIKKSTLAAGQPPHLTSAQDQDTISEL ASCLQRARELGARVRALKASAQDAGESCTPEAEGRPEEPCG EKAPAYQRFH
Cercocebus atys	MEQRRVTDFFARRRPGPGIVLPKLACRTPSPARPALRAPDSA

Cdt1 IDR	TSGSRKRARPPAAPGRDQARPPARRRLRLSADAVSSPSSPEA PDIPACPSLGQKIKKSIPAANQPPQLTSSQNQDTISELASCLQR ARELGARVRVLKSSAQDAGESFTPEAKGRPEEPCGEKAPAY QRFHALA
Caenorhabditis elegans Cdt1 IDR	MSSRVTRSRSQTPQTAVTDFFKTTRKPRARGEKKIIDESITKT TTIKEVQIPAAPIVENKSSRTRARTPSRQAELSSSGTPPKKARE AEAEAPELPKIHEDIILENPAPQKSAIDKLMEASGPAKKKAGI RTVADLQARIAAKGAAKAIHEQNLKNKTKQVDDHVELLKSP KKPVIVEHHKSPQKSKAARNLFSPKKPVPDYIIPGFNVLKSAE KVRGEQDDAHEEEGKRMTAEFLKSSRLIDE
Chiloscyllium punctatum Cdt1 IDR	MAQARLTDFFPQAKKPKPALKGSRSPGGRRLQPRKGRSAAA PCPPPPPALPAVSEPPDSCEWPSPGQLRGCSGLTLHPEAEVSE REPALSDLVTPPRSEPLGAGSRSSTGRKRSRSASVDGAPASA VGSQRAAGVPSARRCLLLEPKEEEENPVPKQSEVEPDAQWPP KGAPEPTDLLIYEPKTLRKVISKPPGLQSSNGRNKQTGTTPAK EGLSGLKARLHRIQKLAPVPGSFASASEVDSDIKSRLKRIREL ELKVQENKEEQKKTAEQVATSGAEERVKTPAYQRFHNLAQ DV
Apostichopus japonicus Cdt1 IDR	MAQAKVTDFYAHRRSLRKIETRTMKRRKLSSSSDSLLKEVPS KRSKPEHHVQAIFQDIEDNCTSSAINIGKSISFHDFTQALNSRS KTISQTKSRVASGRKDDPKPRGRKKQETKGKVGMVGKPQAI SSSTLEQYLAKKQVNDVDVGSGKADEIKKNRETCLTPLEKTP TKRGHDQTRVPVRRKKAKEIFSEESGYSSQSSVSDNQPNERS KRMAKRCLGSQFQNGQSLEAIGTTQSSISGQDASENGRVVTP NLQKVKVKPAPTKERKERVVDSALNKSPLKDRIRQTLANQV TEKPEGGKGHKIGTRKLSPEEVKAKLAKCGKLEELQAVLGR VKRSTQVKRDEVKNPERKSEELPAYKKYRHLAV
Magallana gigas Cdt1 IDR	MAQARVTEFYATKKRNGDAQPSKRRKLQIISDVSEMIQNES VTTRATRSSARRKGISGPVSAEPENVLFTLSSSDESSKSASTK TPKQRSRRIKQVKGQQSIVDALGSRPTSPNEISNGEEVTSAW DEHDGPLTPSKKTKQDNENPTEGKTKDTKKRTRNQNVMTPL KEEKEETIKAPRSRKKLQLRANVIASSESENSENEVTEPPANS EKQKFRDALELPTTPLAISPLPETPEVVRQQMKKGPGVPTSD KVHAAMDKLKELGKEGRSLEKMSKADVKDKLQKCGKLED LKAQLLQLNEAKKQIKKIKESKTRPKPEAAQAKPDVPALDKF EALEVQVKQSDKSEPVSEPKAPAYERFHHLSTPAPPSLSLPYK
Octopus vulgaris Cdt1 IDR	MEAKQSTLTNYYNARKKNSQTKLVQDKSVNENAVAIKLTK ESSSKLSPVQSKNISRSKRYNPTRTSVVELLDTCNEQSTKLES PKSKNQIKPETPPFCHKTQDNRLICSPSKRKHTETLPDQNTRT PESNNEFTIIKPTKSARKRLPLFDEKDKETEPNELPEEKTKDIE SNIEEEKELEDSNFWELEFPLVSNIPRPISPLPETPQKLRRMRS QQMSVHSQTTESGSNDSKELQKATDPVKTTTVRQLNLTPEK VKIDLMKCSKVEELKKQLLLIKELKAATKSKEKPKSSSCKTQ PEFEKEKSDSVNKTPAYKRF
Amphimedon queensladica	MSKSEKQTVLPFQVVKKRSDDERKVKKRLTENTCIRPTTTPK STSGKKEKGKDEADVAPPPPQALQTTPTKSPSLSKFSTVEV

Cdt1 IDR	HSPLKASPHKEVKRGIKRALKYDDLTSV
Hydra vulgaris Cdt1 IDR	MEGQTTLNQYFTARKNDISMQPSKRRKLDLTTIIEPNNRNRR TCFKNKAQLEVLDLFKDSSKSVQKIEEAKSIDDRYSPKIQVEV SLKRKSTAENLEILNQELISKSISNNNDNLPSCNKNEVKCQSK DLNILRTNFTRRKLKTVSAINEQAKMTLFSRGSKLVNAGLVK SCGKPYNDKKVSNSQKIDLTTIPLLNEKEELKSQQNQNFSQL SQKNFNVEKVADISKQCNLTQFSGIIETDSAITPINKKVEPKSG LRNKYAHLLQIDMCIDEKAMADKKMDTNLCVNDAEEKPNQ PAYLKYKH
Oikopleura dioica Cdt1 IDR	MFLEYFDHLTLTKKPRHYCHSRNTDIICKFTSIRLEIGAFQGK NWGVRSLIQVSLGRPEFIPKMPRRAAGRTAALVSQPKIDVFT ERVTRRSARIKNSTLDSENENIRSIAAPTKKVLMASPKRPNQK LSVSGGSPTKMAKKEELLKNHETIQQTYAKQLELLKSPSKLG TLTPDKLKFELDLSPEAEKPLPSIVRSRRALAFDDESSGSAVL PSSFLPEVQAPVFSGSQIKRAREVDSQPNSPSTASPMYKKQLA PRMSAKEVMSSPTKPSSPRASPRKVDGRSRAERMLEAAGGD VKIGGKSISKELLKRQIEKSGKLNEIKERLKSIKTADDKLQAIR DGKIKAENLKILRKERPKPTPTPENAPCYMRFRHLVEPGRM
Platynereis dumerilii Cdt1 IDR	MAQTSVTNFFASRKRGADIHASKRRKVQRLTQDVETENVSV SEQANAELQSSRPAGKGDLKVPELQKNEVGEDSSTATKPSTR KSAPRATKTRKASISTRSSQRSKTKAAASAGQKDITQFLSSIL ESDDQTKDVLPEPEPTAAIDDHSSSPPCTPTKRESAEVPEKQG QKRSRPSIRKDLFSVKTPDAYDFKQYAEKPEEKDRLVMKGR RLSEEGTSSANLGILPEANEKTVRIKGKEAVKEKSKSKEKEIV KESKSKSKSKDSVKEGKISTKKTTKSAEVSESSSASAEKEVTK SKRQSKKVDLADKMESIRELMSTKDVPAPSNNIMDQLNEKV SAVKQKGSLSKKLPGNKLEELKQKLMQHNKQKKEVEVIKES RKGKELSMPKAPAYERFHNLAQPL
Danio rerio Cdt1 IDR	MAQARVTDYFAQSKKAGVSRSLRSKGQKVSGDVVESAVIN KPRSSSRASGSSRKATHSEITTAEPQKQTQLEFLKVIDEALSTE TAETVADSRDVNIEGLTASPRTPKRSSPEFDVCSVLFPSTAEL HSSAKKRQRLNAGHNCRSSPEERTGQKTARKKLNLLASDDK VKSTEPLASSSPQAPQQTARKESKNTVNHIANTSTVNKPNGD ENPQRATRSKKTFSREDVAALKSKLQKLKGQSENVSTPSPGP VSTLTELKARLDAAREISAKVQQRKAERVIEEAKAAETQPAT EPQEREKLPAYQRYHTLAQDVPPG
Columba livia Cdt1 IDR	MAPLLDLAAHTPASTTQSPIQGQDGGTQRGTAPPGKTRRLQ REDLAGLRSRLQKVTALAQLPPVPSGTSSDLRNHLERVRQLE LRIREKKAGCGTTSGDTGASGDTRTAAPAAEASEKPPAYQRF HTLAQELPPGLT

References

Banani SF, Lee HO, Hyman AA, Rosen MK. 2017. Biomolecular condensates: Organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**:285–298. doi:10.1038/nrm.2017.7

Boeynaems S, Alberti S, Fawzi NL, Mittag T, Polymenidou M, Rousseau F, Schymkowitz J, Shorter J, Wolozin B, Van Den Bosch L, Tompa P, Fuxreiter M. 2018. Protein Phase Separation: A New Phase in Cell Biology. *Trends Cell Biol* **28**:420–435. doi:10.1016/j.tcb.2018.02.004

Boeynaems S, Holehouse AS, Weinhardt V, Kovacs D, Van Lindt J, Larabell C, Bosch L Van Den, Das R, Tompa PS, Pappu R V., Gitler AD. 2019. Spontaneous driving forces give rise to protein−RNA condensates with coexisting phases and complex material properties. *Proc Natl Acad Sci U S A* **116**:7889–7898. doi:10.1073/pnas.1821038116

Chiu YP, Sun YC, Qiu DC, Lin YH, Chen YQ, Kuo JC, Huang J rong. 2020. Liquid-liquid phase separation and extracellular multivalent interactions in the tale of galectin-3. *Nat Commun* **11**. doi:10.1038/s41467-020-15007-3

Das RK, Pappu R V. 2013. Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues. *Proc Natl Acad Sci U S A* **110**:13392–13397. doi:10.1073/pnas.1304749110

Han TW, Kato M, Xie S, Wu LC, Mirzaei H, Pei J, Chen M, Xie Y, Allen J, Xiao G, McKnight SL. 2012. Cell-free formation of RNA granules: Bound RNAs identify features and components of cellular assemblies. *Cell* **149**:768–779. doi:10.1016/j.cell.2012.04.016

Hossain M, Bhalla K, Stillman B. 2021. Multiple, short protein binding motifs in ORC1 and CDC6 control the initiation of DNA replication. *Mol Cell* 1–19. doi:10.1016/j.molcel.2021.03.003

Hyman AA, Weber CA, Jülicher F. 2014. Liquid-liquid phase separation in biology. *Annu Rev Cell Dev Biol* **30**:39–58. doi:10.1146/annurev-cellbio-100913-013325

Kato M, Han TW, Xie S, Shi K, Du X, Wu LC, Mirzaei H, Goldsmith EJ, Longgood J, Pei J, Grishin N V., Frantz DE, Schneider JW, Chen S, Li L, Sawaya MR, Eisenberg D, Tycko R, McKnight SL. 2012. Cell-free formation of RNA granules: Low complexity sequence domains form dynamic fibers within hydrogels. *Cell* **149**:753–767. doi:10.1016/j.cell.2012.04.017

Luo F, Gui X, Zhou H, Gu J, Li Y, Liu X, Zhao M, Li D, Li X, Liu C. 2018. Atomic structures of FUS LC domain segments reveal bases for reversible amyloid fibril formation. *Nat Struct Mol Biol* **25**:341–346. doi:10.1038/s41594-018-0050-8

Mitrea DM, Kriwacki RW. 2016. Phase separation in biology; Functional organization of a higher order Short linear motifs - The unexplored frontier of the eukaryotic proteome. *Cell Commun Signal* **14**:1–20. doi:10.1186/s12964-015-0125-7

Nott TJ, Petsalaki E, Farber P, Jervis D, Fussner E, Plochowietz A, Craggs TD, Bazett-Jones DP, Pawson T, Forman-Kay JD, Baldwin AJ. 2015. Phase Transition of a Disordered Nuage Protein Generates Environmentally Responsive Membraneless Organelles. *Mol Cell* **57**:936–947. doi:10.1016/j.molcel.2015.01.013

Oldfield CJ, Dunker AK. 2014. Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu Rev Biochem* **83**:553–584. doi:10.1146/annurev-biochem-072711-164947

Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci U S A* **108**:13624–13629. doi:10.1073/pnas.1110633108

Parker MW, Bell M, Mir M, Kao JA, Darzacq X, Botchan MR, Berger JM. 2019. A new class of disordered elements controls DNA replication through initiator self-assembly. *Elife* **8**:1–35. doi:10.7554/eLife.48562

Parker MW, Botchan MR, Berger JM. 2017. Mechanisms and regulation of DNA replication initiation in eukaryotes. *Crit Rev Biochem Mol Biol* **52**:107–144. doi:10.1080/10409238.2016.1274717

Parker MW, Kao J, Huang A, Berger JM, Botchan MR. 2021. Molecular determinants of phase separation for *Drosophila* DNA replication licensing factors. *bioRxiv* 2021.04.12.439485. doi:10.1101/2021.04.12.439485

Patel SS, Belmont BJ, Sante JM, Rexach MF. 2007. Natively Unfolded Nucleoporins Gate Protein Diffusion across the Nuclear Pore Complex. *Cell* **129**:83–96. doi:10.1016/j.cell.2007.01.044

Quiroz FG, Chilkoti A. 2015. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nat Mater* **14**:1164–1171. doi:10.1038/nmat4418

Shulga N, Goldfarb DS. 2003. Binding Dynamics of Structural Nucleoporins Govern Nuclear Pore Complex Permeability and May Mediate Channel Gating. *Mol Cell Biol* **23**:534–542. doi:10.1128/mcb.23.2.534-542.2003

Uversky VN. 2017. Intrinsically disordered proteins in overcrowded milieu: Membrane-less organelles, phase separation, and intrinsic disorder. *Curr Opin Struct Biol* **44**:18–30. doi:10.1016/j.sbi.2016.10.015