Multi-scale computational studies of heterochromatin protein-1

Jazelli Lenae Mueterthies

Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Recommended Citation
Mueterthies, Jazelli Lenae, "Multi-scale computational studies of heterochromatin protein-1" (2020). Graduate Theses and Dissertations. 18189.
https://lib.dr.iastate.edu/etd/18189

This Thesis is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
DEDICATION

I would like to dedicate this thesis to my dad, who was always supportive of me. He was unable to see me get into graduate school, but I know he would be proud.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER 1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Bio-molecular phase separation.</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Phase separating the nucleus by proteins condensates.</td>
<td>2</td>
</tr>
<tr>
<td>1.3 The HP1(Heterochromatin-1) proteins: the masters of chromatin fate.</td>
<td>3</td>
</tr>
<tr>
<td>1.4 Simulations of Intrinsically disordered proteins</td>
<td>4</td>
</tr>
<tr>
<td>CHAPTER 2. CONFORMATIONAL PROPENSITIES OF DISORDERED HP1 LINKERS</td>
<td>6</td>
</tr>
<tr>
<td>UNDER VARYING IONIC CONDITIONS</td>
<td>6</td>
</tr>
<tr>
<td>2.1 Overview</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Methods</td>
<td>7</td>
</tr>
<tr>
<td>2.3 Results</td>
<td>7</td>
</tr>
<tr>
<td>2.4 Conclusion</td>
<td>11</td>
</tr>
<tr>
<td>CHAPTER 3. DISORDER-MEDIATED OLIGOMERIZATION OF HP1 LINKERS</td>
<td>12</td>
</tr>
<tr>
<td>3.1 Methods</td>
<td>12</td>
</tr>
<tr>
<td>3.2 Results</td>
<td>13</td>
</tr>
<tr>
<td>3.3 Conclusion</td>
<td>17</td>
</tr>
<tr>
<td>CHAPTER 4. COARSE-GRAINED MODELS OF FULL-LENGTH HP1 OLIGOMERIZATION</td>
<td>18</td>
</tr>
<tr>
<td>4.1 Methods</td>
<td>18</td>
</tr>
<tr>
<td>4.2 Results</td>
<td>19</td>
</tr>
<tr>
<td>4.3 Conclusion</td>
<td>20</td>
</tr>
<tr>
<td>CHAPTER 5. GENERAL CONCLUSIONS</td>
<td>22</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>23</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Nuclear morphology image from electron microscopy.</td>
<td>3</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>A hierarchy of protein force-fields used in this thesis.</td>
<td>5</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Secondary structure, solvent-accessible surface area, end-to-end distance, and radius of gyration of HP1 linker at three ionic conditions.</td>
<td>8</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>$C_\alpha - C_\alpha$ contact frequencies for HP1 linker monomer.</td>
<td>10</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Secondary structure, radius of gyration, and solvent-accessible surface area of HP1 linker dimer at three ionic conditions.</td>
<td>15</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>$C_\alpha$ contact maps of HP1 linker dimers and inter vs intra chain contacts at three ionic conditions.</td>
<td>16</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Free energy profiles of HP1 linker dimer as a function of radius of gyration at three ionic conditions.</td>
<td>17</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>Contact maps predicted by AWSEM for three ionic conditions.</td>
<td>19</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>Overview of computational pipeline for coarse-graining HP1 linkers.</td>
<td>20</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis.
ABSTRACT

Liquid-liquid phase separation of intrinsically disordered proteins has recently become a foundational mechanism in molecular and cell biology. The emergent “liquid view” is offering a fresh perspective for understanding a longstanding problem of chromatin organization and regulation in eukaryotic nuclei. The family of heterochromatin proteins have been shown to undergo phase separation in vivo and in vitro, thereby providing clues for connecting molecular interactions of proteins, formation of heterochromatin domains, and the emergent gene regulatory processes. Understanding the molecular specificity and mechanisms by which heterochromatin proteins orchestrate nuclear processes, however, is a highly non-trivial question because of the multi-scale nature of the system, spanning from individual molecular components to large-scale nuclear domains. In this thesis, we report first steps toward underlying the multi-scale picture of heterochromatin protein actions. By carrying out simulations of heterochromatin protein components with all-atom and coarse-grained modeling, we shed light on detailed molecular features which drive the formation of aggregates and condensates in the nucleus. The study provides a stepping stone towards understanding heterochromatin formation and gene silencing from a ground-up molecular perspective.
CHAPTER 1. INTRODUCTION

This chapter provides a high-level overview of the field of chromatin biophysics, the structural and regulatory roles of heterochromatin proteins, and the nature of liquid-liquid phase separation. The latter has recently emerged as the foundation mechanism which connects chromatin biology, gene regulation and nuclear architecture. Each of the three background sections are self-contained and can be read standalone to understand the context of the results presented in the next three chapters.

1.1 Bio-molecular phase separation.

The very first glimpses of inner cellular order have come to us through the discovery of membrane-bound structures, known as organelles, in the early days of light microscopy [1]. Examples of eukaryotic organelles are the nucleus which insulates genome from the cytoplasm, ribosomes which organize protein synthesis, and vacuoles which isolate the cellular waste. The advent of high-resolution imaging and microscopy techniques in the last decade has now revealed the ubiquitous existence of organelles that altogether lack membranes [2, 3]. These membraneless organelles, also broadly referred to as biomolecular condensates, consist of proteins and nucleic acids loosely held together primarily by weak electrostatic interactions [4, 5]. Both the nucleoplasm [6, 7, 8] and cytoplasm [9, 10, 11] have now been shown to contain membraneless compartments in wide abundance. Under the microscope, membraneless condensates appear as dynamic, liquid-like droplets ranging in size from 10 to 1000 nanometers— far bigger than just a few molecules clustered together and often exceeding in size even multi-component biomolecular machines, such as the ribosome. The testament to the liquid-like quality of these droplets is their ability to fuse, grow, and dissolve by rapidly exchanging components with the outside cellular environment in response to changing external conditions [12]. Formation and behavior of these droplets, in many respects, follow
the principles of liquid-liquid phase separation well-known in physical chemistry and soft matter
science [13]. Liquid-liquid phase separation is familiar to us through the ordinary experience of
shaking a bottle of vinaigrette and watching the behavior of oil droplets as they gradually coalesce,
eventually leading to phase-separated mixture. Following the vinaigrette analogy, the emerging
new picture of intracellular order is that of a complex and out-of-equilibrium solution, which can
dynamically modulate its inner biochemistry by decomposing into protein-rich droplets embedded
in the protein-depleted environment. While established to be ubiquitous in cells, functional roles
of membraneless organelles are still far from being understood [14, 15, 16]. Current speculations
revolve around diverse regulatory purposes for phase separation such as micro-reactors, noise reduc-
ers, gene co-localizers, etc [17, 18, 19, 20]. Finally, experiments have been quick to point out that
the proteins which drive phase separation are disordered, either partially or fully [21, 22, 23, 24].
This fact suggests new roles for protein disorder, while simultaneously offering clues for the puz-
zling fact that nearly a third of proteins with highly-conserved sequences are encoding disordered
states [25, 26].

1.2 Phase separating the nucleus by proteins condensates.

As a major organelle, the nucleus can be viewed as a complex, non-equilibrium soup consisting
of long strands of DNA mixed together with various proteins and RNAs [27]. This complex mixture,
collectively referred to as chromatin, manages to regulate genome and organize a near meter-long
genomic DNA inside the micrometer-sized nucleus [1].

The organization of chromatin is neither random, nor uniformly smeared throughout the nu-
cleus. Instead, there are dense and diffuse patches of chromatin visible on micrometer scale, as
well as other hierarchically organized domains, loops, and fibers further down the nanometer
scale [28, 29, 30, 31]. The dense and slow-moving patches of chromatin correspond to geneti-
cally silenced areas known as heterochromatin, while more mobile and diffuse regions are known as
euchromatin (Fig. 1.1).
Epigenetic markings and transient protein interactions are known to dynamically regulate spatial and temporal patterns of nuclear eu- and hetero-chromatin content [32, 33]. Different chromatin patterns are strongly correlated with phenotypic identity and biochemical activities of cells [34, 35, 36].

In the recent series of exciting experiments [37, 38], liquid-liquid phase separation of HP1 (heterochromatin protein-1) proteins was found responsible for desegregation of heterochromatin and euchromatin territories, very much like how oil dissolves hydrophobic molecules into droplets, thereby separating them from water. Furthermore, it appears that, globally, protein-protein interactions and phase separation endow chromatin with viscoelastic liquid-like properties [39, 40, 41, 42].

The idea based on liquid-liquid phase separation of proteins is rapidly emerging as a powerful organization principle [43, 44, 45] capable of explaining the sub-nuclear order. For instance, in the latest computational studies [43], a simple polymer model of chromatin based on the idea of phase-separating monomeric chromatin units was able to capture experimentally-verified dynamic behaviors, such as anomalous diffusion, viscoelasticity, and coherent motion of chromosomal loci.

1.3 The HP1(Heterochromatin-1) proteins: the masters of chromatin fate.

The heterochromatin protein family HP1 have a ubiquitous presence in the nucleus with wide-ranging functions such as heterochromatin formation, gene silencing, and DNA repair [46, 47, 48, 49, 50]. In most mammalian organisms, there are three members in the HP1 family [51]: indicated by HP1α – γ in mouse cell lines, and HP1α-c in Drosophila cell lines. The three forms of HP1 show a high level of sequence similarity, yet they are distributed in distinct regions of the nucleus with distinct functions. Both HP1α and HP1β are predominantly associated with heterochromatin,
while HP1\(\gamma\) localizes in euchromatin regions of the nucleus [52]. Despite the pervasive presence and functions of HP1 proteins in the nucleus, very little is known about the molecular mechanisms that realize these versatile functions [51]. What makes a detailed molecular level study of HP1 functions challenging is the disordered nature of HP1 proteins. All HP1 proteins have a highly conserved chromodomain in the N terminus (CD) and a chromo shadow domain (CSD) at the C terminus [53], linked by a hinge region (HR). The hinge region is enriched with positively-charged amino acid residues, which make it disordered, making the overall molecules of HP1 highly flexible. The flexibility allows HP1 proteins to form various dimers, oligomers, etc., and engage in binding with multiple different protein partners [54]. Notably, the HP1 recognize epigenetic modifications on nucleosomes and specifically bind to the N-terminal tail of histone H3 methylated at Lysine-9 [55, 56], thereby promoting heterochromatin formation. Different theoretical models have been proposed to explain the rapid and significant scale spread of HP1 accumulation, modeling with and without cooperative interactions between HP1 molecules [57, 58, 59]. The realization that HP1 undergoes liquid-liquid phase separation calls for a rethinking of mechanisms and the role of HP1 in heterochromatin formation and sub-nuclear organization [44]. A central aim of this proposal is to construct detailed models which will bring clarity on how HP1 phase separation shapes intranuclear order and regulation at different scales.

1.4 Simulations of Intrinsically disordered proteins

All-atom models account for every atomic degree of freedom in a condensed matter system, whereas coarse-grained models treat solvent and ions implicitly and group a number of atoms into effectively interacting degrees of freedom, referred to here as ”beads”. In this thesis, we will be making use of several coarse-grained and all-atom models (Fig. 1.2). Some unique challenges in modeling disordered proteins come from (i) all-atom model being computationally too taxing for sampling conformations of lengthy sequences (ii) coarse-grained models lacking first principles parametrization handles, and difficulties in accounting disordered and ordered regions within a single model. After an exhaustive review of all the major coarse-grained models for intrinsically
disordered proteins reported to-date [60, 61, 62, 63, 64, 65, 66, 67, 68, 69], we have settled on a set of protein models that work best for addressing the problems of heterochromatin protein phase separation detailed in this proposal. Specifically, we have identified and tested the performance of AWSEM [70], a99SB-disp [71, 72], and HPS (Hydrophobicity-scale) models [68] for systematically and rigorously accounting for salient behaviors of HP1 proteins at different scales.

The coarse-grained force field AWSEM [70, 73] occupies a unique niche among protein models, because it can capture both ordered regions and the disordered nature of protein fragments in a data-driven way by pooling information from multiple protein sequences and all-atom simulations. Briefly, AWSEM is an implicit solvent, 3-bead-per-amino-acid resolution coarse-grained model for proteins. The Hamiltonian of AWSEM includes physically motivated terms for backbone, hydrogen bonding, burial, contact, and hydrogen bonding interactions and the knowledge-based terms, named fragment memory potentials.

The physically motivated terms capture secondary structure elements while the knowledge-based term is making use of a machine-learning pipeline based on neural networks to train its “fragment memory” potentials on either sequence data or data from all-atom simulations. Combining physically motivated and fragment potentials make AWSEM a predictive and transferable force field. A testament to the predictive ability of AWSEM has been recent work where the complete structure of large protein complex NFkB-IkB with missing N-terminal and binding domains was computationally reconstructed using only sequence information [74]. In a separate study, the bimodal conformational disorder of a well-known PEST region of IkB proteins was correctly captured by using sequence homology information [75] in agreement with the solution NMR data [76, 77].

Figure 1.2: A hierarchy of protein force-fields used in this thesis.
CHAPTER 2. CONFORMATIONAL PROPENSITIES OF DISORDERED HP1 LINKERS UNDER VARYING IONIC CONDITIONS.

This chapter describes the set-up and analysis of the all-atom simulations of disordered HP1a linkers under three different ionic conditions. These simulations serve as a stepping stone toward carrying out dimerization simulations and they pave the way for building larger-scale coarse-grained models.

2.1 Overview

As was outlined in the background section, the HP1 family of proteins are central molecular players, controlling heterochromatin formation in the nucleus [46, 47, 48, 49, 50]. The mechanistic molecular picture underlying heterochromatin formation, however, is far from being understood. In a tour de force study, it was discovered that a full dimeric HP1a protein is capable of undergoing liquid-liquid phase separation triggered by change of ionic condition alone [37, 38]. Similarly, phosphorylation serves as a trigger for HP1α to form liquid condensates. This condensation of HP1 is a key step towards forming heterochromatin territories. Therefore, understanding conditions favoring phase separation, as well as gaining a molecular level insight into driving forces, is of paramount importance for chromatin biology.

Considering the nature of liquid-liquid phase separation, which is predominantly orchestrated by protein sequences with a high degree of intrinsic disorder and abundance of charged and aromatic residues [4, 5], we set to dissect the conformational propensities, dimerization, and phase separation of HP1 by adopting a multi-scale, multi-resolution approach. In this approach, our central hypothesis is that phase separation is encoded in the disordered region, while the structured fragments of HP1 provide fine-tuning of material and physical properties of HP1 aggregates.
In this chapter, we report results obtained by carrying out three simulations using disordered linker fragment of HP1a. The objective of the simulations is to understand conformational propensities of HP1a linkers, which sheds light on understanding oligomerization tendencies in terms of detailed conformational ensemble properties. Specifically, we find a distinct transition from collapsed to more expanded forms and enhancement of beta secondary structural elements in the conformational ensemble which we argue is conducive to oligomerization, which is ultimately a cause for phase-separation under low ionic concentration environments.

2.2 Methods

The all-atom simulations are done by using AMBER99SB-ILDN force field \[78\], which is suitable for intrinsically disordered proteins, and SPC/E force-field for water and ions \[79\], which is well-suited for high ionic environments employed in the present study. Three simulations are carried out under low-salt (10 mM NaCl), medium-salt (50 mM NaCl), and high-salt (150 mM NaCl) conditions, respectively. The GROMACS2018.1 \[80\] was used for carrying out all-atom molecular dynamics simulations reported in this thesis. The linker region is constructed using PyMol \[81\] by giving the chain random coil configuration, which was then subjected to successive rounds of minimization, equilibration, and production runs.

The energy minimization was done using a steepest-descent minimization algorithm over 100000 steps. The equilibration was done in two stages, where the minimized system was subject to NVT sampling with gradual temperature ramp for 100 ps from temperature \(T = 0\) to \(T = 300K\) , followed by NPT sampling at constant temperature \(T = 300K\) . Following equilibration, each of the three systems was subject to production runs of about 3-3.5 \(\mu\)s. Convergence was assessed by contact map, principal component analysis (PCA), and radius of gyration(Rg) measures.

2.3 Results

Analysis of all-atom simulations reveals a number of interesting trends about conformational preferences of HP1 linkers in different ionic conditions.
In particular, there is a trend of going towards larger $R_g$ values with increasing salt concentration (Fig. 2.1A). This is expected, as ionic screening leads to weakening of the charged intra-residue attractions. Interestingly, there seem to be two prevalent $R_g$ values at low salt concentration. Also notable is the comparatively narrow range of $R_g$ in the medium-salt condition, compared to low- and high-salt, in which the histograms are considerably wider. The medium-salt condition gave a radius of gyration on the lower end, but not as low as the low-salt monomer at its smallest. All radii of gyration were between 10 and 15 angstroms at all times during the simulation.

Next, we analyze secondary structural preferences (Fig. 2.1A) of the linker under our three ionic conditions. While overall low, the high-salt simulation’s alpha-helical content is noticeably higher than that of low- or medium-salt. Conversely, the high-salt simulation had the lowest occurrence of loops and irregular elements, though still quite high relative to the other structural elements. The high amount of irregular elements is expected, as it is an intrinsically disordered protein. However, the difference between the different salt conditions, namely the lower occurrence at high-salt, could indicate that decreased disorder is significant to the function of HP1a at lower versus
higher salt concentration. Interestingly, the medium-salt condition had considerably less betaladder and especially hydrogen-bonded turn content, while both the high- and low-salt conditions were noticeably higher. The high- and low-salt simulations for these two measures were nearly equal in beta-ladder content, and the low-salt condition had slightly fewer turns than the high-salt. HP1a had no pi-helix content at any salt concentration we simulated. Overall, the loss of loop-content at high-salt, with low- and medium-salt being similar to each other, may be significant as a sign that either a loss of disorder-dependent functionality or the gain of more structured elements such as alpha-helical content, of which high-salt has the most, may play a role in the lack of droplet formation seen experimentally at high salt concentrations.

Analysis of solvent-accessible surface area (SASA) (Fig. 2.1B) perhaps reveals the clearest trend of how many residues are exposed to solvent. We find a clear and strong trend of increasing SASA with increasing salt concentration. Greater SASA indicates more expanded and solvent exposed conformations. The medium salt has narrowest range. The high salt has the widest range. Similarly to Rg, the adopting of a more extended configuration is indicative of a decrease in ionic interactions within HP1a in favor of interactions with the salt molecules in the system. In addition to salt, though, this more solvent-exposed configuration also allows an opportunity for more inter-molecular interactions in place of the intra-molecular interactions it lost.

Finally, we analyze contact maps to reveal the most detailed picture of the conformational ensemble. Since the system is disordered, we have computed maps showing the frequency of making contact, where contact is defined as below 6 Angstroms. The contact frequency quantifies the fraction of time the contacts persist during the simulation.

Similarly to the narrow range of SASA and Rg, the medium-salt simulation shows fewer, but stronger, contacts than either high- or low-salt. The map reveals a line of strong contacts beginning in the top-left of B in Fig. ??and proceeding diagonally down and right till the center line. While not a continuous line, it shows patches where there are multiple strong contacts nearby to each other. This is particularly true for the area showing contacts between residues 5-15, just past the N-terminus, and residues 50-55, which are about 15-20 residues from the C-terminus.
Figure 2.2: $C_{\alpha} - C_{\alpha}$ contact frequencies; the fraction of time pairs are in contact throughout time course of simulation. (A) The contact map of HP1 linker at 150mM $NaCl$ (B) The contact map of HP1 linker at 50mM $NaCl$ (C) The contact map of HP1 linker at 10mM $NaCl$
There is a similar pattern of many nearby contacts in the low-salt map, though at different residues, specifically between residues 1-10, corresponding to the N-terminus region of the linker, and residues 11-20, which are a little further downstream. This series of close-together contacts suggests a loop where the bend occurs in the vicinity of residue 10, an Alanine flanked by other Ala and Ser residues. Additionally, under low-salt conditions, there is a patch of weak interactions between residues 60-67, the C-terminus region of the linker, and residues 1-10 (N-terminus). These contacts weaken and disappear as salt-concentration increases. Contrast this with contacts between residues within the 50-60 range, just upstream from the N-terminus, where there is little or no interactions in all but the high-salt condition. From this, we conclude that the major contacts (Fig. 2.2) at low-salt occur beyond Residue 20, with very few contacts being present in the region near the N-terminus. This means that the N-terminal region is available for interactions with other molecules. The amount of contacts near the N-terminus increases at medium-salt, where some of the strongest contacts are between residues near the C-terminus and those near the N-terminus. At high-salt, it goes a step further and has most of the residues below 28 making contact with each other in a loop-like faction, as previously discussed, potentially limiting the contacts the N-terminus can make with other molecules.

2.4 Conclusion

By carrying our variety of analysis and extracting both polymeric quantities (Rg, end-to-end distance distributions), protein-specific quantities (secondary structures and contact maps), as well as generic solvent-accessible surface area, we have revealed the conformational preferences of HP1 disordered linkers. Surprisingly, we find that solvent accessible surface area is a better coordinate for describing conformations of disordered linkers in three different salt environments. The other measures, while showing weaker trends, nevertheless indicate a shift of conformational preferences with ionic strength that goes beyond the simple electrostatic picture.
CHAPTER 3. DISORDER-MEDIATED OLIGOMERIZATION OF HP1 LINKERS

This chapter describes the simulations done on dimeric disordered linkers with an objective to elucidate nature of HP1 oligomerization and binding affinity trends with respect to different ionic conditions. The objective of simulations done under different ionic conditions is to understand driving forces of HP1 phase separation where change of ionic conditions generates HP1 droplets in vitro [37]. The simulations carried out for this chapter include three constant temperature sampling runs and three constant temperature metadynamics simulations for sampling probability distributions as a function of linker dimerization measures such as radius of gyration, fraction of contacts and helicity.

3.1 Methods

The simulations for the dimerized linkers follow the same protocols as for the monomeric units. We repeat those here for the sake of completeness and to make this section self-contained.

The all-atom simulations are done by using AMBER99SB-ILDN force field [78], which is suitable for intrinsically disordered proteins, and SPC/E force field for water and ions [79], which is well-suited for high ionic environments employed in the present study. Three simulations are carried out under low-salt (10 mM NaCl), medium-salt (50 mM NaCl), and high-salt (150 mM NaCl) conditions, respectively. The linker region is constructed using PyMol [81] by giving the chain random coil configuration, which was then subjected to successive rounds of minimization, equilibration, and production runs.

The energy minimization was done using a steepest-descent minimization algorithm over 100000 steps. The equilibration was done in two stages, where the minimized system was subject to NVT sampling with gradual temperature ramp for 100 ps from $T = 0$ to $T = 300K$ temperature, followed
by NPT sampling at constant $T = 300K$ temperature. Following equilibration, each of the three systems was subject to production runs of about 3-3.5 $\mu$s. Convergence was assessed by contact map, PCA, and Rg measures.

The metadynamics simulations were used to enhance conformational sampling by biasing the system away from already-sampled configurations to explore rarely-occurring ones. All-atom metadynamics simulations of the dimerization of HP1a were performed at high, medium, and low salt concentration using GROMACS 2018.1 [80] patched with PLUMED library 3.1.2. As in the standard molecular dynamics simulations, the AMBER99SB-ILDN force field [78] and the SPC/E water model were used [79]. In our protocol, the Gaussian had an initial height of 0.5 kJ/mol and an initial width of 0.05, the unit dependent on the particular CV. The Gaussians were deposited every 1 ps.

In our metadynamics sampling, four collective variables (CVs) were used: radius of gyration (Rg), potential energy, hydrophobic contacts, and salt bridge contacts. Rg and potential energy are built-in modules for PLUMED [82], while the hydrophobic and salt bridge contacts were built from the coordination module with the following specifications. Contacts between hydrophobic residues were calculated as the number of $C_\beta$ couples closer than 6 Angstroms. For salt bridge contacts, two groups, A and B, were defined. Group A was defined as all the heavy atoms from the $R(COO)^-$ group of Asp and Glu, and Group B was defined as all the heavy atoms from the $R(NH_3)^+$ group of Lys and the $R(NHCNH_2)^{2+}$ of Arg. Contacts were calculated as couples of heavy atoms from groups A and B closer than 6 Angstroms.

### 3.2 Results

The all-atom simulations of dimeric constructs have uncovered a number of trends in conformational propensities of disordered linkers when going from low- to high-salt conditions.

Firstly, examining the secondary structure content distribution (Fig. 3.1A), it is clear to see that the dimer has significantly higher alpha-helical content at low salt concentration, compared to medium- or high-salt. The low-salt simulation, conversely, has the fewest bends, with the amount of
bend content increasing as salt concentration increases. Additionally, it is notable that the high-salt condition shows higher 3/10 helix and isolated beta-bridge content than the other two conditions, which are nearly equal to each other in both measures. As in the monomer, the medium-salt condition showed lower turn and beta-ladder content relative to low- and high-salt. Additionally, like the monomer, the low- and high-salt simulations had similar results for these two structure elements. Also following suit with the monomer simulations, the occurrence of loops and irregular elements is the lowest in high-salt condition, with medium-salt having the greatest frequency. In both monomer and dimer, the low-salt condition’s frequency of irregular elements was between that of the medium- and high-salt. One minor difference between the monomer and dimer simulations at low-salt is that the monomeric simulation had loop content closer to the medium-salt, while it was more similar to the high-salt in the simulation of the dimer. Like the monomer, the loss of loop-content could be indicative of an overall loss of disorder that may play a role in the disappearance of liquid-like droplets at high-salt concentration in experiments. However, the trends are far less clear for the dimer, where for most of the secondary structure elements, the relative amount is quite similar at all salt-concentrations.

Next, we analyze the Rg distribution of the three dimeric simulations (Fig. 3.1B). It is notable that, unlike the monomeric simulations, the histograms of all three salt-conditions have peaks centered around the same Rg range, 25-27 Angstroms. At high-salt, the Rg preference is quite narrow. This contrasts with the medium-salt, which has a smaller secondary peak at 27-28 Angstroms, and the low-salt, which has multiple overlapping peaks of similar heights in the 25-27 Angstrom range, with no clear preference between them. Overall, the radius of gyration is larger than that of the monomer, which is to be expected, but it is not twice as large, which indicates the two linkers in the dimer are still quite close together. The relative similarity of the radii at all three salt concentrations indicates that the dimer’s functionality at differing salt conditions is likely not dependent on the overall Rg. Additionally, there are many conformations that could have similar Rg, but completely different bonds and interactions.
As in the monomer, solvent-accessible surface area (SASA) shows the clearest trend, with increased salt corresponding to an increase in SASA indicative of a more elongated configuration (Fig. 3.1C). The difference between the SASA of HP1a at each salt concentration is smaller in the dimer, indicating that there is a certain portion of the dimer that is buried from the solvent regardless of salt concentration. Because the low-salt condition is where phase-separation appears in experiments, our results suggest that a more collapsed conformation may be necessary for phase-separation to occur.

In the contact maps below (Fig. 3.2), residues 1-67 correspond to the first HP1a strand, with residues 68+ comprising the second strand. This means that the top-left and bottom-right quadrants correspond to inter-molecular contacts, while the top-right indicates intra-molecular contacts within the second strand, and the bottom-left corresponds to the same in the first strand. There is a loss of intra-molecular interactions in the upper-right region (Fig. 3.2) as salt concentration increases. Within the regions of inter-molecular contacts, a place worth noting is a string of contacts between the first strand’s residues near its N-terminus and the second strand’s residues near its C-terminus. These regions contain many uncharged residues, such as Gly, Thr, and Ser, and the arrangement of contacts indicate multiple contacts in close sequence proximity. There is also a small patch of contacts just below this point of interest, in the residue range somewhat close to the second strand’s N-terminus (Res ID 100-120), that, too, is present only at high salt concentrations. Since contact maps are best served for single chains we have also quantified intra-chain and inter-chain contacts in the conformational ensemble sampled by simulations (Fig. 3.2). The results indicated
Figure 3.2: Contact maps of dimers (A-C) where each chain is indicated by a solid color bar placed on top of each contact map where (A) $C_\alpha$ contact map at 10mM NaCl (B) $C_\alpha$ contact map at 50mM NaCl (C) $C_\alpha$ contact map at 150mM NaCl. The panels (D-E) quantify inter vs intra chain contacts, showing (D) Histogram of intra-chain contact frequencies (E) Histogram of inter-chain contact frequencies

that overall low salt state tends to forms more intra-chain contacts while inter-chain contacts appear to be comparable for the three chains or within statistical noise region. This suggests that the intra-chain contacts may have an important role in phase-separation, which occurs at low-salt.

Overall, one of the biggest stand-outs is how subtle the differences are between the three salt concentrations compared to the contact maps of the monomer. Many of the contact patterns are similar, with some minor differences in strength of contact. Because of this, those few places where there is a great difference are of interest, as they may hold clues to the differences in function of HP1a at high-, medium-, and low-salt.
To understand how binding affinity is modulated by the ionic conditions we have next turned to enhanced sampling simulations. We have carried out constant metadynamics simulations of the HP1 dimer under the three ionic conditions by choosing Rg, potential energy, salt-bridge contacts, and hydrophobic contacts as our collective variables to measure.

The free energy profile of Rg, pictured in Fig. 3.3, shows to be the best reaction coordinate of those collected. A trend quickly emerges from the free energy landscapes of HP1a at the three salt concentrations we simulated. The difference between the energy minima and the flat region grows with increasing salt. This larger change in free energy suggests an increase in the strength of binding for the dimer, because a larger decrease in free energy shows a greater energetic benefit.

### 3.3 Conclusion

Overall, we find trends similar to monomeric disordered linkers which manifest in a shift of secondary structural propensities in different ionic conditions, as well as change in the overall size of dimers. In particular, solvent-accessible surface area continues to be a clear marker of the conformational changes in the HP1a at different salt conditions. The metadynamics simulations show stronger binding in dimerization as the concentration of salt increases, indicating a preference for binding in a more extended conformation.
CHAPTER 4. COARSE-GRAINED MODELS OF FULL-LENGTH HP1 OLIGOMERIZATION

This chapter describes the preliminary research done on HP1 linkers using coarse-grained force-field AWSEM (Associated memory, Water mediated Structure and Energy Model). The research reveals that while AWSEM being knowledge based and “secondary structure aware” force-field, nevertheless is not readily re-purposed for addressing more subtle conformational and secondary structural disorder displayed by Intrinsically disordered proteins. Hence the main result of this chapter is to document the research done with an aim for making more informed choices for coarse-graining shall one want to study phase-separation of disordered proteins such as HP1.

4.1 Methods

The AWSEM (Associated memory, Water mediated Structure and Energy Model) force-field based model [70, 73] represents each amino acid with three degrees of freedom or interaction ”beads” corresponding to \( C_\alpha \), \( C_\beta \), and \( O \), with the exception of Glycine, which has an H in place of \( C_\beta \). AWSEM is a knowledge based force-field which relies on information extracted from multiple sequence alignments from which secondary structural biases are extracted for short fragments. Hence the word “memory” by which each sequence is aware of secondary structural preferences based on information available in databases. The appeal for using AWSEM is its rigorous validation as a prediction of structures for proteins, dimers and larger aggregates. With one or two exceptions AWSEM however has not been used for sampling conformational ensembles of intrinsically disordered proteins. Hence the objective of research documented in this section was to carry out through investigation which attempts to find parameter regime (without modifying AWSEM’s functional form) which would be appropriate for investigating systems akin to HP1 disordered linkers.
All of our AWSEM simulations were performed using the LAMMPS molecular dynamics package, in which the AWSEM force field was implemented. All secondary structural biases were turned off for our simulations taking into account information from all-atom simulations. Twenty monomer replicates of varying Debye-Hückel screening length, beginning at 1.0 and increasing to 20.0 in increments of 1.0, were run at $T = 300K$ with a timestep of 5 fs. Coordinates were recorded every 10000 steps, and thermodynamics were output every 50000 steps. Each simulation was run for between 1.0 billion and 1.8 billion fs, or 1 to 1.8 µs.

4.2 Results

The figure Fig. 4.1 summarizes all of the results done on HP1 using all-atom and coarse-grained models and puts the findings obtained by AWSEM in appropriate context. We have chosen two resolutions of coarse-graining for studying the condensation and phase separation of HP1 components: namely, AWSEM (3 bead per residue) and also a much simpler model HPS (1 bead per residue model based on Hydrophobicity scale) [68]. The descriptions of AWSEM and HPS coarse-grained models are provided in the background section. Here we report simulation results employing these two coarse-grained force-fields. Study of full HP1 dimers with all-atom resolution is already computationally prohibitive because of the extended and disordered nature of the dimer, which requires large solvation boxes. Therefore, we have adopted coarse-grained models for studying condensation of full HP1 fragments and HP1 linkers.

First, after thorough study with the AWSEM force-field, we have found that even with no explicit secondary structural biases, the AWSEM force-field still shows a strong preference towards
Figure 4.2: Overview of computational pipeline for coarse-graining HP1 disordered linkers and final prediction of experimental trends of phase-separation forming beta secondary structural elements. The analysis of contact maps reveals patterns in large disagreement with all-atom results. Therefore, we conclude that AWSEM force-field, while known for excellent predictive capabilities for structured proteins, has rather poor performance when it comes to intrinsically disordered proteins even when explicitly removing secondary structure biasing terms. Thus, we are left with an option of either completely fine tuning AWSEM or adopting a simpler model. We have chosen the latter path for two reasons: flexibility and more straightforward interpretation. We have used 1-bead-per-residue resolution coarse-grained model, where each bead is endowed with electrostatic and excluded volume interactions. Secondary structural elements have been fixed with rigid-body constraints. In conclusion we will mentioned that a simpler 1-bead-per-residue resolution, we are able to capture two facts: contact map patterns and experimental condensation trends (Results not shown).

4.3 Conclusion

We find the performance of AWSEM force-field unsatisfactory for studying HP1 linkers, as the secondary structural bias encoded in AWSEM is predicting completely different conformational propensities, in particular beta sheets. Thus, we have decided to adopt simpler, yet more flexible 1-
bead-per-amino-acid resolution with no secondary structural biases as a good trade-off. The model is able to capture basic features of contact maps and, more importantly, predicts experimental trends of HP1 phase separation of three different ionic conditions. Thus, we conclude that there are well-defined changes in conformational preferences of HP1 linkers. Overall, the condensation is largely driven by electrostatic forces.
CHAPTER 5. GENERAL CONCLUSIONS

Our simulations show that the HP1a linker exhibits noticeable conformational shifts over a range of ionic concentrations less than an order of magnitude in size. By analyzing these differences in conformational preferences, we have determined that high ionic concentration is associated with more elongated and solvent-exposed configurations in both monomer and dimer, while lower ionic concentrations show a more collapsed configuration. Based on metadynamics simulations, we found the elongated high-salt configuration of the dimer had the strongest binding affinity, with the affinity decreasing in magnitude as the salt concentration decreased. Combining this information with the aforementioned experimental results of the Karpen and Narlikar labs, we conclude that liquid-liquid phase separation of HP1 corresponds to a less-extended conformation and a weak binding affinity when dimerizing.
BIBLIOGRAPHY


[23] Seema Qamar, Guozhen Wang, Suzanne J Randle, Francesco Simone Ruggeri, Juan A Varela, Julie Qiaojin Lin, Emma C Phillips, Akinori Miyashita, Declan Williams, Florian Ströhl, William Meadows, Rodylyn Ferry, Victoria J Dardov, Gian G Tartaglia, Lindsay A Farrer, Gabriele S Kaminski Schierle, Clemens F Kaminski, Christine E Holt, Paul E Fraser, Gerold Schmitt-Ulms, David Kleenerman, Tuomas Knowles, Michele Vendruscolo, and Peter St George-


[33] Jesse R Dixon, Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E Antosiewicz-Bourget, Ah Young Lee, Zhen Ye, Audrey Kim, Nisha Rajagopal, Wei Xie, Yaru Diao, Jing Liang,


[70] Aram Davtyan, Nicholas P Schafer, Weihua Zheng, Cecilia Clementi, Peter G Wolynes, and
Garegin A Papoian. AWSEM-MD: protein structure prediction using coarse-grained physical
8503, July 2012.

[71] Paul Robustelli, Stefano Piana, and David E Shaw. Developing a molecular dynamics force field
E4766, May 2018.

[72] Paul S Nerenberg, Brian Jo, Clare So, Ajay Tripathy, and Teresa Head-Gordon. Optimizing
solute–water van der waals interactions to reproduce solvation free energies. *The Journal of

[73] Peter G Wolynes and Garegin A Papoian. AWSEM-MD: From neural networks to protein
structure prediction and functional dynamics of complex biomolecular assemblies. In *Coarse-


[75] Davit A Potoyan, Weihua Zheng, Diego U Ferreiro, Peter G Wolynes, and Elizabeth A
Komives. PEST control of molecular stripping of NFκB from DNA transcription sites. *J.

[76] Sulakshana P Mukherjee, Pedro O Quintas, Reginald McNulty, Elizabeth A Komives, and
H Jane Dyson. Structural characterization of the ternary complex that mediates termination
of nf-kb signaling by iκbα. *Proceedings of the National Academy of Sciences*, 113(22):6212–
6217, 2016.

[77] Holly E Dembinski, Kevin Wismer, Jesse D Vargas, Gajendra W Suryawanshi, Nadja Kern,
Gerard Kroon, H Jane Dyson, Alexander Hoffmann, and Elizabeth A Komives. Functional

Dror, and David E Shaw. Improved side-chain torsion potentials for the amber ff99sb protein

[79] Pekka Mark and Lennart Nilsson. Structure and dynamics of the tip3p, spc, and spc/e water

[80] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk
Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level