August 2022

Identification of Secondary Structural Elements Contained Within the Intrinsically Disordered N-Terminal Tail of the Bloom's Syndrome Helicase.

Vivek Somasundaram
University of South Florida

Identification of Secondary Structural Elements Contained Within the Intrinsically Disordered N-Terminal Tail of the Bloom's Syndrome Helicase.

by

Vivek Somasundaram

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
with a concentration in Cell and Molecular Biology
Department of Cell Biology, Microbiology, and Molecular Biology
College of Arts and Sciences
University of South Florida

Major Professor: Kristina Schmidt, Ph.D.
Gary Daughdrill, Ph.D.
Ernst Schönbrunn, Ph.D.
Meera Nanjundan, Ph.D.

Date of Approval:
September 22, 2022

Keywords: BLM, Intrinsically Disordered Proteins (IDPs), RecQ Helicases, protein NMR, alpha helices

ACKNOWLEDGEMENTS

First and foremost, I offer thanks to my PI for granting me the opportunity to work on this project. I must also thank my committee members for their contributions as well. I would also like to thank Dr. Wade Borcherds, Serena Higbee, Malissa Fenton, and Emily Gregory for training me to use the FPLC and NMR instruments both of which were essential for this project. Also, I offer thanks to Dr. Elizabeth Jones-Mason for all of her assistance during my time at USF, especially during the challenges to teaching posed by the pandemic. I would also like to thank the CMMB Department for their financial support during my time at USF. In addition, I wish to thank Dr. Mark Kearley and Dr. Brian Miller, both of whom were essential professors and committee members during my undergraduate days. I must also extend my thanks to my previous mentors Dr. Laura Osteen and Mr. Brad Fantle, whose guidance helped me greatly during my high school and college years. Finally, I offer the greatest thanks to my mother Gnanabhanu Somasundaram and my father Dr. Thayumanasamy Somasundaram, not only for inspiring me to become a scientist, but to keep my head up during even the most difficult times.

TABLE OF CONTENTS

LIST OF FIGURES ii

ABSTRACT iii

CHAPTER 1: INTRODUCTION
- Defining protein disorder 1
- Function of disorder in RecQ helicases 3
- Disordered domains in RecQ helicases 4
- Approaches to characterize disordered proteins and their findings 9
- Computational methods 10
- Nuclear magnetic resonance spectroscopy 12
- Förster resonance energy transfer 14
- Circular dichroism 15

CHAPTER 2: IDENTIFYING ALPHA HELICES IN BLM 1-100 USING NMR 18

CHAPTER 3: CLONING AND EXPRESSION OF THE FIRST 380 RESIDUES OF BLM 1-380 29

CHAPTER 4: DISCUSSION 33

CHAPTER 5: MATERIALS AND METHODS 36

REFERENCES 42

LIST OF FIGURES

Figure 1: All known domains present within the five RecQ helicases 5

Figure 2: IUPred disorder tendency predictions for all five human RecQ helicases 7

Figure 3: BLM-protein-protein interaction map 8

Figure 4: Agadir predicted output of BLM N-terminus 19

Figure 5: Optimal induction parameters for induction of expression of BLM fragments grown in small scale cultures 20

Figure 6: SDS-PAGE of IPTG induced lysate of BLM fragment bound to ProBond Resin 21

Figure 7: Optimal induction parameters for induction of expression of BLM fragments grown in large scale cultures 22

Figure 8: Flowthrough for IMAC purification of His-tagged BLM 1-100 from E. coli lysate 24

Figure 9: BLM 1-100 resonance assignments 25

Figure 10: NMR obtained change in chemical shifts 26

Figure 11: Overlay of Agadir predictions with NMR obtained change in chemical shifts 27

Figure 12: Gel confirmation of cloning steps 30

Figure 13: Confirmation of induction of expression and integrity of GST tag for two BLM 1-380 clones 31

Figure 14: Anti-GST Western blot 32

ABSTRACT

The Bloom's Syndrome Helicase (BLM) is one of five human RecQ helicases and is necessary for maintenance of genome stability. Whilst a crystal structure exists for the C-terminal domain of BLM, very limited structural knowledge is known about the intrinsically disordered N-terminal tail. This lack of insight exists, despite the fact that the N-terminus of BLM is essential for the overall biological activity of BLM. Here we provide an Nuclear Magnetic Resonance spectroscopy based approach that we used to identify two distinct α-helices contained within the first 100 residues of BLM. In addition, we propose a mutagenesis-based approach involving rationally designed proline mutants to determine the biological function of these α-helices We also provide an experimental framework to characterize the remainder of the BLM N-terminus. Taken together the experiments described here will enable us to identify α-helices in a mostly disordered region and develop a structure and biological function of the BLM N-terminus.

CHAPTER 1: INTRODUCTION

Defining protein disorder

Protein structure is generally thought of as unique and highly specific conformation that a protein adopts for its biological function. The notion that “protein structure dictates function” has been a central pillar of biochemistry, but while this philosophy applies to structurally ordered proteins-structurally disordered proteins do not entirely fit into this framework. Unlike structurally ordered proteins, which possess a fixed structure, structurally disordered proteins display conformational flexibility and can adopt multiple structures *in vivo.*[1] This in turn means that structurally disordered proteins have a dynamic structure that differentiates them from their more rigid ordered counterparts.

Disorder in proteins is generally denoted through low hydrophobicity, the lack of fixed angles on a Ramachandran plot, and transient secondary structures.[2] The low level of hydrophobicity makes intrinsically disordered proteins (IDPs) less bulky, thereby making them more flexible and more favorable for disordered proteins. [2] Low hydrophobicity also reduces collapse of the polypeptide chain due to the hydrophobic effect. In addition, the increased net charge in disordered regions creates a force to oppose collapse of the polypeptide chain. As Ramachandran plots show possible amino acid conformations in a peptide, and disordered proteins have a dynamic structure, this dynamism is reflected by an abundance of allowed conformations. This conformational flexibility, results in secondary structural elements that have a lifespan and are regarded as “transient”. Hence this has given rise to the term transient secondary structures, to reflect the ephemeral nature of these elements in IDPs.

IDPs contain Intrinsically Disordered Regions (IDRs) which are amino acid regions approximately 40 residues long that lack a rigid structure.[3] It is important to note here that the term IDP is an all-encompassing term both referring to proteins that are entirely disordered or proteins that contain numerous IDRs within them.

The extent of disorder is also not a purely qualitative designation, as the extent of disorder protein can be predicted with high accuracy. One such program, IUPred, enables the user to determine disorder tendency of a protein simply by supplying its primary sequence.[4] The resultant output not only includes locations of predicted IDRs, but also supplies a disorder tendency value (scoring from 0-1) indicating predicted regions of low or high disorder. Disordered proteins have also been shown to elicit characteristics quite differently from their ordered counterparts, as they have a wide range of isoelectric points that deviate more than one pH unit from physiological pH.[5] This in turn can make identifying them on SDS-PAGE somewhat challenging as they tend to run at apparent molecular weights different from their actual molecular weight. The primary sequence of disordered proteins is also quite unique as soluble amino acids (acidic, basic, and polar residues) are abundant in IDPs.[6] This is turn means that disordered proteins also have a lower aggregation propensity than their ordered counterparts. Taken together although the criterion for being classified as an IDP is relatively generic, the IDP proteome has a high degree of diversity ranging from tumor suppressor genes like BRCA to the HIV-1 transcriptional regulator protein tat.[7] The common features that they all share are low hydrophobicity and lack of a fixed structure.

From an evolutionary standpoint disorder regions confer conformation flexibility due to their lack of a fixed structure.[8] Specifically as protein-protein interactions within IDPs usually occur within small regions (motifs), this enables the region to be specifically tuned for an interaction through post-translational modifications or adoption of specific conformations.[8] The lack of rigidity

allows for disordered proteins to adopt conformations that are more conducive to protein-protein interactions. It has even been shown that upon interaction, disordered proteins undergo conformational changes.[9] This flexibility coupled with their other unique features not only dismissed the previously held notion that proteins require a fixed structure, but also demonstrates that these proteins exist as a unique subclass of proteins.

Conservatively, recent estimates suggest that the IDP proteome contains more than 100,000 short linear motifs that act as binding sites.[10] As the lack of a fixed secondary structure makes IDPs more amenable for interaction, this same feature also confers IDPs to have crucial roles in biological processes. Cellular signaling processes are significantly assisted by IDPs as their flexibility allows for interaction with multiple signaling molecules.[11] Moreover IDPs have fast association rates which enables for rapid activation of signaling pathways and allow for multiple interactions.[12] [13] Moreover as IDPs are the target of extensive PTMs, these modifications enable them to act as switches or regulators of cell signaling.[14]

Function of disorder in RecQ Helicases

RecQ helicases use ATP hydrolysis to unwind DNA and are integral to maintenance of genome stability through a variety of mechanisms. For instance, BLM achieves this through double strand break repair and homologous recombination, whereas WRN does it through homologous recombination and non-homologous end joining.[15] RECQ1 uses checkpoint mediated activation to regulate genome stability while RECQ4L and RECQ5 use homologous recombination and disruption of Rad51 respectively. [16] [17]

Loss of function mutations in the three helicases BLM, WRN, and RECQ4L result in Bloom's syndrome (BS), Werner's syndrome (WS), and Rothmund-Thompson syndrome (RTS) respectively.[16] [17] [18] All three of these diseases are rare autosomal recessive disorders with

very distinct phenotypes and prognoses.[19] Both BS and WS patients produce a premature aging phenotype, whereas RTS patients have poikiloderma, short stature and both skeletal and dental abnormalities.[9] Specifically, all three diseases cause patients to have an elevated risk to developing cancers, although certain cancers are associated with each syndrome.[20] For example BS patients have an elevated risk to most cancers, WS and RTS patients are mostly susceptible to skin cancers and osteosarcomas, respectively.[21] This may suggest that although all three helicases are implicated in homologous recombination (HR), the function of BLM is more critical for the process of HR. Additionally, it has been reported that cells deficient in WRN do not exhibit any more sensitivity to gamma irradiation than WT cells.[22] Several murine RECQ4L models have been reported, but with quite different outcomes. Whereas one was lethal, in another model the mice survived to adulthood showing symptoms of RTS, whilst in a third model, the mice also survived to adulthood with 5% of the population not showing any malignancies. [23] [24] [25]

As disordered regions are more flexible, the advantages of disordered regions in interaction prone RecQ helicases are obvious. The tendency for IDPs to bind to proteins with fast association rates often means that interactions in the disordered regions of RecQ helicases are not subject to orientation restraints that can occur when two ordered proteins interact. [26] The kinetics of IDPs are also crucial as intrinsic disorder minimizes the likelihood of very low rate constants, which can hinder protein-protein interactions from taking place. [27] Taken together disorder in proteins is favorable for protein-protein interactions from both structural and thermodynamic perspectives.

Disordered domains in RecQ helicases.

All five RECQ helicases share the feature of an ordered helicase domain and an ordered RQC domain (Figure 1). However, as all five RecQ helicases possess unique biological functions and

mechanisms for maintaining genome stability, they possess distinct domains some of which are found in disordered regions (Figure 1 and 2).

All five RecQ helicases contain disordered domains, but the extent of disorder differs amongst the five helicases. Whilst BLM, RECQ4, and RECQ5 contain N or C-termini that are almost entirely disordered RECQ1 and WRN only contain small, disordered regions (Fig 3). Of note is that the region predicted to be disordered in WRN overlaps with the WRN-RPA binding site at residues 422-484.[28] In addition the exonuclease domain in WRN, which is just upstream of the disordered tail, has been shown to be involved in protecting nascent DNA from MRE11/EXO1 degradation.

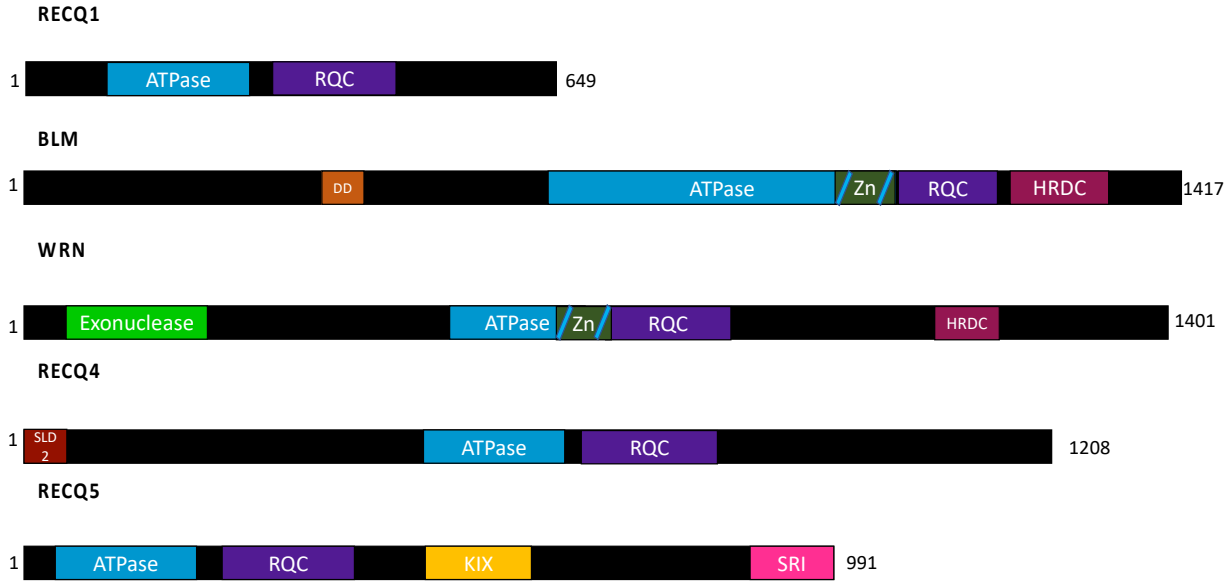

Figure 1: All known domains present within the five RECQ helicases. Helicase core for each is denoted by ATPase and RQC domain unique to all RECQ helicases. Zn-finger is shown independently but is part of ATPase domain. HRDC domain found in BLM and WRN functions to recognize quaternary DNA structures. Domains unique to each helicase are shown in distinct colors. Dimerization domain oligomerizes BLM, Exonuclease prevents MRE11/EXO1 degradation, SLD2 is interaction site for TopBP1, KIX domain is needed for binding to Mus81, while SRI domain is necessary for interaction with RNAPII. Figure adapted from Hamedeh and Lansdorp, 2020.

The disordered C-terminal region of RECQ1 (residues 592-649) has been shown to efficiently bind to PARP-1.[29] RECQ5, whose C-terminus is predicted to be disordered, has two domains in its C-terminus which have been shown to be responsible for protein-protein interactions for RECQ5.[30] The disordered C-terminus also contains the site where Rad51 interacts with RECQ5 as well as the SRI domain which interacts with RNA polymerase II (Figure 1).[17] As for RECQ4, the disordered N-terminus has been shown to be both a DNA interaction motif as well as a region for multiple protein-protein interactions.[31] [32]

BLM, has been shown to interact with TopIα[33], TopIIα[34], TopIIIα[35], p53[36], TopBP1[37], RPA[38], RECQ4[39], Mcm6[40], Rad51[41], Rad54[42], MLH1[43], ATM[44], WRN[45], TRF1[46], TRF2[46], and Mus81[47]. The bulk of BLM protein-protein interactions take place within the BLM N-terminus (Figure 3). This explains why although the N-terminus is dispensable for enzymatic activity, it is needed for the biological activity of BLM.

In considering the role of disorder in RecQ helicases, specifically for protein-protein interaction, the presence of disordered regions is a crucial feature for the biological activity of all RecQ helicases. Specifically, these IDRs confer the ability for protein-protein interactions to take place within RecQ helicases and is not merely a stochastic property borne out of the sequence of these helices.

Role of ordered domains in RECQ helicases

While the disordered sequences of RecQ helicases themselves lack a significant amount of conservation, the domains present in all five RecQ helices are quite conserved, with the average shared identity approximately 40% (Figure 3). Curiously though while RECQ1, WRN, RECQ4, and RECQ5 all have helicase cores contained within the N-terminal tail, the helicase core of BLM is in the C-terminus

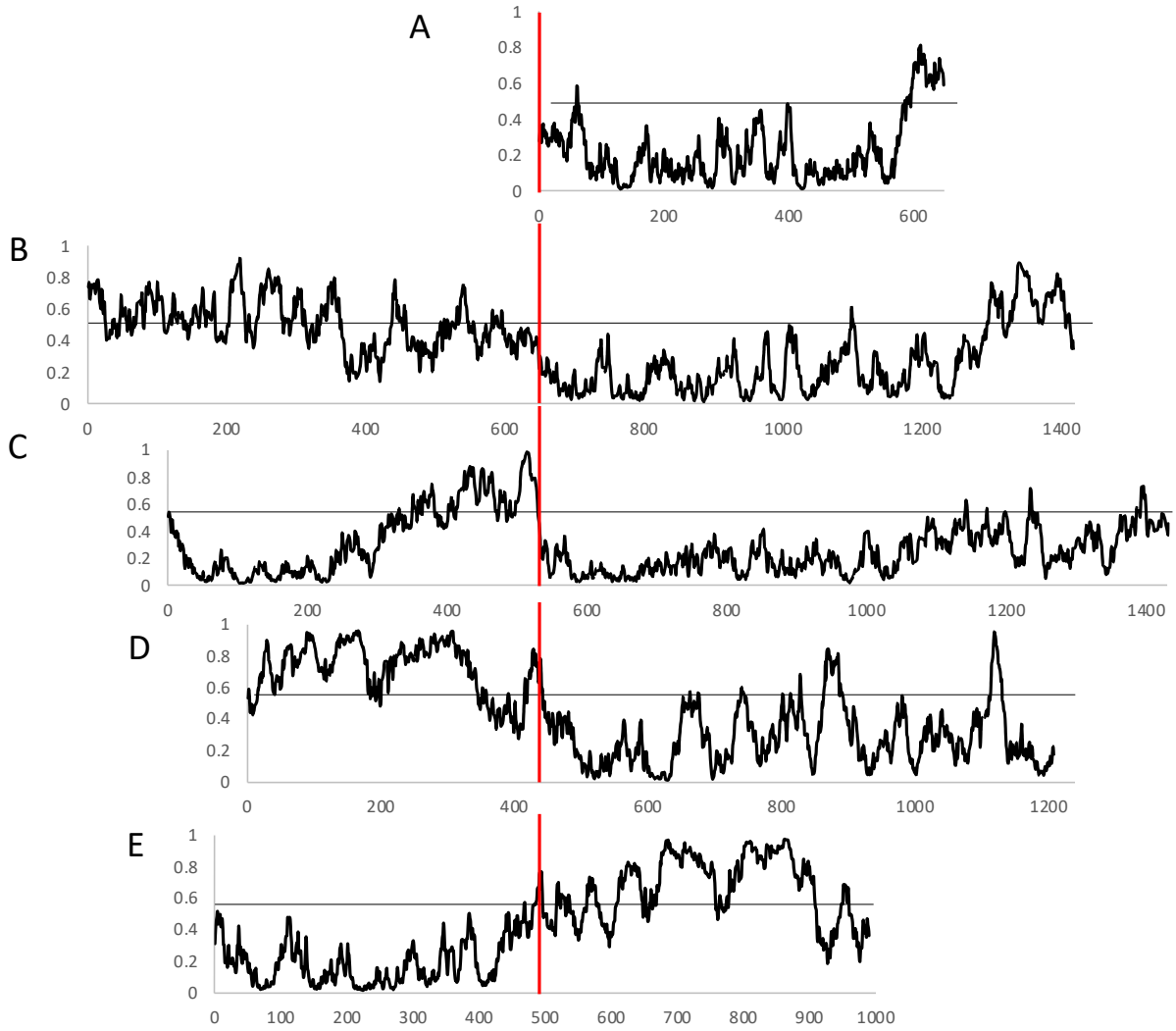

Figure 2: IUPred disorder tendency predictions for all five human RecQ helicases. Threshold between predicted order and disorder indicated by black horizontal line at 0.5. Demarcation between N and C-terminus is illustrated by black perpendicular lines on each chart. A-E: RECQ1, BLM, WRN, RECQ4, and RECQ5 respectively. BLM, WRN, and RECQ4L all contain long disordered N-terminal tails, a feature not shared by either RECQ1 and RECQ5.

An additional crucial difference is that BLM and WRN both possess signature tandem domains all of which are contained in the C-terminus which are absent from REC1, REC4L, and RECQ5.[48] One such domain is the RecQC (RQC) which has unique functions in each helicase.

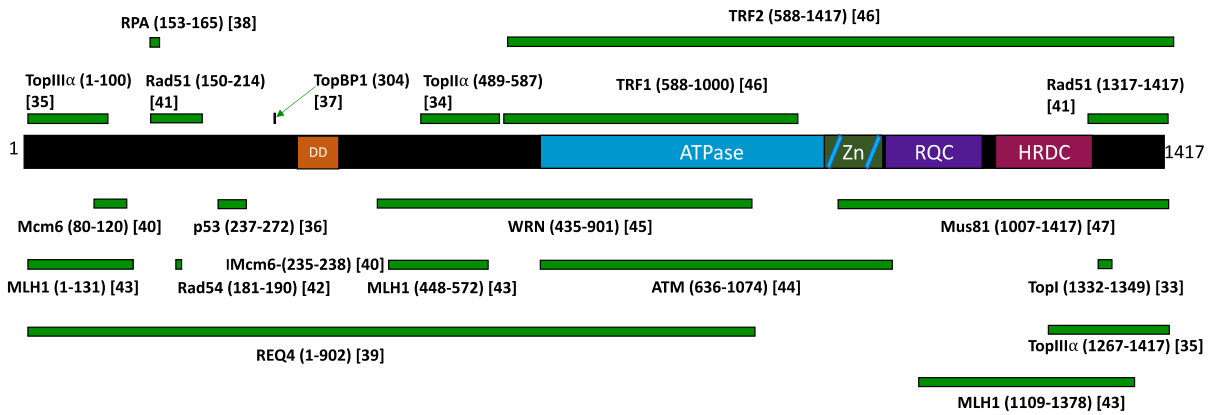

Figure 3: BLM-protein-protein interaction map. Shows known interacting proteins and the sites/regions they have been experimentally shown to bind to. Dimerization domain (362-414) ATPase (642-1068), Zn-finger (994-1068) RecQC domain (1074-1194), and Helicase and RNaseD C-terminal domain (1208-1290). Citation numbers for each site shown in square brackets. Domain sizes/locations taken from Kitano et al., 2014.

The surface of the RQC domain of BLM and WRN have both shown to interact with dsDNA.[49] This is distinguishable from the RQC domain in RECQ1 which has been shown to contain an α-helix that prevents strand annealing from taking place in this region and promotes DNA unwinding.[50]Deletion of this helix or alanine mutagenesis was shown to improve strand annealing, but reduced DNA unwinding. [50]

The second ordered domain is called the helicase and RNAseD C-terminal domain. The role of the HRDC in both BLM and WRN is somewhat less clear. In BLM the HRDC has been shown to be needed for BLM to dissolve Holliday Junctions, an event that is crucial for BLM to prevent aberrant recombinant events.[51] In WRN the HRDC is thought to potentially facilitate protein binding, as the linker region upstream of it is thought to be unstructured and therefore a potential site for protein-protein interactions.[52] It should be noted that the linker regions between the RQC and HRDC BLM and WRN are distinguishable by length, as the linker for WRN is considerably longer.[52] This longer length offers support to the idea that the HRDC in WRN may be favorable for protein-protein interactions to take place. This conclusion is

additionally supported by a small region of predicted disorder in WRN that is just upstream (residues 1069-1144) of the linker region of WRN (Figure 2C).

When considering RECQ helicases, what is most arguably most essential is the phenotype of cells where these individual helicases are deficient. All five helicases are necessary for maintenance of genome stability, and downregulation of any of the five helicases predisposes cells to cancers. However, deficiencies in RECQ1 and RECQ5 are not associated with any specific disease state, whereas deficiencies in BLM, WRN, and RECQ4 all have specific malignancies identified. Taken together, within the context of necessity, this suggests that RECQ helicases possess a hierarchy with BLM, WRN, and RECQ4 more necessary than RECQ1 and RECQ5 for maintenance of genome stability.

Approaches to characterize disordered proteins and their findings.

Lack of a fixed secondary structure, the key feature of disordered proteins, limits structural characterization. As such, this cannot be characterized by X-ray diffraction (XRD) due to the lack of a stable diffraction pattern.[53] X-ray crystallography is one of the most common methods used to developed structural models of proteins. Unfortunately, the crystallization of most IDPs is incredibly challenging, which also hinders characterization since a crystallized protein is needed for XRD.[54] Even deletion of the disordered fragments from an IDP and subsequent characterization of the ordered fragments yields limited success.[55] These limitations, coupled with the plethora of IDPs present in the proteome mean that alternate methods that relay structural information are crucial to understanding this class of proteins. Thus, in attempting to characterize these proteins it is crucial to understand the merits and limitations of techniques that can characterize IDPs. These merits and limitations are summarized in Table 1.

Computational methods

Arguably the simplest method for characterizing IDPs is to simply bypass the wet lab and use software as a means of analysis. We have already discussed IUPred, a software that can predict disorder tendency to a residue specific resolution from a proteins primary structure. IUPred predictions also provide users with the option to obtain ANCHOR plots, a feature that enables the user to identify sites that may function as sites for protein-protein interactions. Using the same parameters as IUPred, ANCHOR uses the estimated energies of residues in a supplied primary structure to identify regions likely to gain energetically through protein interactions.[56] Currently this methodology can predict disordered binding regions with an accuracy of 70%.[57]

Two other types of software, Disprot and Spot-Disorder2, also generate multiple sequence alignment. [58] [59] Disprot is a regularly updated database that effectively serves as a repository for published findings about IDPs. It contains more than 800 entries about IDPs/IDRs grouped into 7 major classes and 35 subclasses. This enables the user to quickly obtain all relevant published information about a specific IDP/IDR.[60] The predecessor to Spot-Disorder2, namely Spot-Disorder, used Long-short term memory networks in a bidirectional recurrent neural network (BRNN).[61] Spot-Disorder2 uses this same methodology but couples it with both evolutionary information and predicted 1-D structures to predict disorder in proteins and can even identify semi-disordered regions.[62]

Agadir can be used to predict the presence and location of α-helices in a protein. This software, based on helix-coil theory, was developed from data collected from 423 peptides analyzed by circular dichroism.[63] [64] The output of Agadir provides helical propensity by yielding numerical values for each residue within the sequence. Residues with a high value for helical propensity are identified as residues more likely to adopt a helical conformation. In general, helical

propensity values >5 are regions highly likely to adopt a helical conformation.[65] Although Agadir calculations are valid for IDPs, it is best suited for making predictions for smaller peptides (80-100 residues) rather than large proteins.

Additional software exists to determine pathogenicity which is defined as the likelihood a point mutation within a disordered region will impact the functionality of a protein. This is important because traditional software that can do this, like Polyphen and SIFT, are not reliable at making such predictions for disordered regions as these software algorithms are derived from structured proteins.[66] PON-DISO, uses available AAindex data from 685 known biochemical properties of amino acids in conjunction with evolutionary features from existing MSA data. Six total features, two AAindex and four evolutionary features. Taken together this software enables the user to ascertain the pathogenicity of an amino acid substitution within the disordered region of a protein with a success rate of at least 50%. [67] A similar software PON-P2, also employs MSA and AAindex data but couples with random forest probability that bolsters accuracy to 61.7%. [68]

Molecular Dynamics (MD), a simulation that predicts the motion of atoms in a protein based on established knowledge of interatomic interactions, has also been used to characterize IDPs.[69] [70] In previous years MD characterization of IDPs has been limited by the fact that parameters used for ordered proteins were completely unsuitable for IDPs and the high amount of charged residues in IDPs created artifacts of electrostatic potentials. Recent advances in computational processing and changes in approach by unifying force fields for IDPs have increased the viability of MD as a means for characterizing IDPs[71] the limitation of MD and most other software that offers insight into protein structure and function, is that these algorithms are based on existing knowledge of structured proteins which are not applicable for IDPs. [72] [73] [74]

Nuclear magnetic resonance spectroscopy

Nuclear magnetic resonance (NMR) is arguably the best experimental means of characterizing structure in disordered proteins, as it allows the determination of secondary structure at a residue specific level.[75] Specifically this technique circumvents the limitations imposed by the structural flexibility of disordered proteins, while providing data at an atomic level.[76] X-ray diffraction, involves bombarding a sample with an X-ray beam that triggers an elastic scattering of particles. The resultant diffraction pattern can then be used to calculate a protein structure. Unlike XRD, NMR involves the usage of chemical shifts to provide information about the molecular structure and environment of amino acid residues.[77] Acquisition of NMR spectra of a protein requires that the protein be composed of amino acids that contain nuceli that have a magnetic spin such as ^{1}H, ^{13}C, and ^{15}N.[78] This in turn requires that the protein be synthesized by a host organism in the presence of media supplemented by these isotopes.

Protein NMR involves bombarding a purified peptide (concentration ranging from 100-250 mM) with radio waves thereby triggering the nuclei to resonate at specific frequencies. These specific frequencies, referred to as chemical shifts, are influenced by the identity of neighboring atoms (or residues in the case of protein NMR).[79] This specificity is what enables the identification of specific residues, although it must be noted that identical residues with the same neighboring environment will be more difficult to distinguish from each other.[80] The resultant spectra provides peaks, which can then be identified using a reference of known chemical shifts for each residue.

In conjunction with assigning residues, a random coil library for each residue must be made by entering the peptide sequence into a database of known chemical shifts collected from protein or peptide samples that are devoid of structure, including transient secondary structure.[81] This provides expected chemical shifts for each residue based on its neighboring residues, referred

to as random coil shifts.[82] Upon assigning residues the corresponding chemical shifts are then compared to a library of known chemical shifts and the change in chemical shift is calculated by subtracting the random coil library shifts from the chemical shifts of the NMR assigned residues.[83] This process is typically performed on the backbone nuclei for every single assigned residue. Regions with consecutive positive α-carbon shifts and consecutive negative β-carbon shifts correspond to locations of α-helices.[84] This is because the backbone chemical shifts are sensitive to torsional constraints and can be used to calculate dihedral angles.[85] These constraints are then filtered with a torsional angle tolerance range of 20° - 35°, and the conformation with the lowest root-mean square deviation is selected so that β -sheets are most easily distinguishable under these conditions.[86] Aside from structure NMR can also be used to identify specific residues where binding takes place either as protein-protein interactions or protein-DNA unwinding. [87]

In general protein NMR experiments first proceed with Heteronuclear Single Quantum Correlation spectra (HSQC) to identify optimal experimental parameters (such as temperature) and to assess if residues can be accurately assigned using NMR. From there HNCACB and HNCO experiments are used to assign alpha-carbon, beta-carbon, and carbonyl carbons atoms.[88]

Although NMR provides high resolution data and the precise location of secondary structural elements, there are important limitations that prevent it from the being a universal method for characterizing IDPs. First, each residue produces a peak on the spectra, and repeat residue-residue pairs are more difficult to distinguish and assign. As such, a large protein will produce a spectrum with numerous overlapping peaks thereby complicating accurate assignment. Moreover, the spectra of disordered proteins tend to more clustered and narrower than the spectra of ordered proteins, which further exacerbates the problem of overlapping peaks. Thus,

NMR is best suited for characterizing smaller peptides ranging from about 80-120 residues, rather than larger proteins.[89] For an entire protein to be characterized via NMR, it is advisable to truncate the peptide into pieces and characterize each fragment individually.

Protein degradation is also detrimental to NMR experiments, as degradation products will produce distinct signals that will be difficult to identify and distinguish from the undegraded peptide.[90] The timeframe of NMR experiments is important to consider as well, as NMR experiments will need to proceed for several days for complete data acquisition.[91] When combined with the time constraints associated with purifying and preparing sufficient peptide for the experiment, the total timescale can range from weeks to a full month. Given the timeframe of NMR experiments peptides must be sufficiently stable enough to last the duration and physiological conditions of the experiment.

Förster resonance energy transfer

Förster Resonance Energy Transfer (FRET) is another technique that has had some recent success with characterizing the structure of IDPs. Briefly, this technique requires a protein of interest to be tagged with a fluorescent molecule whereupon one fluorophore acts as an acceptor, and the other a donor.[92] Light is then used to transfer energy to the donor fluorophore which then transfers the energy to the acceptor fluorophore converting it to an excited energy state. Subsequent emission of the photon can then be measured as energy transfer efficiency (ET), which is proportional to the distance between the two fluorophores. This in turn can be used to identify the distances between two amino acids which can provide limited insight into the structure of the protein.[93] FRET has also been used to profile the energetic landscape of IDPs, with findings demonstrating that they can stochastically jump between different conformation states.[94] FRET has also been used to characterize changes in IDPs upon undergoing post-translational modifications. [95] [96] [97] [98]

The most common methodology to label proteins for FRET analysis involves cysteine-maleimide chemistry, whereupon native cysteines are replaced with cysteines conjugated with maleimide.[99] This in turn can limit applications of FRET as a protein lacking cysteines may behave differently when non-native cysteines are introduced into the protein being characterized.[100] To overcome some of these limitations it has recently been proposed that FRET be paired with NMR and Small-angle X-ray scattering (SAXS) to develop a more complete structural model of IDPs.[101]

Circular dichroism

Circular Dichroism (CD) is another common technique used to evaluate protein secondary structure through the usage of polarized light. Briefly polarized light has different orientations, and the guiding principle behind CD is that a mixture of asymmetric molecules will absorb left- and right-handed polarized light at different amounts. This difference, termed DE, can then be used to obtain information about the conformation of an optically active analyte such as a protein.[102]

In proteins, the amino portion of the polypeptide backbone contains chromophores which in turn results in structural elements possessing unique CD spectra.[103] α-helices have a positive band at 193 nm and negative bands at 208 and 222 nm, and b-sheets have positive bands at 195 nm and negative bands at 218 nm. [104] [105] Experimentally CD is advantageous for characterizing proteins as only approximately 50μg of analyte is needed for sample analysis, with results being obtained within a few hours.[106] This is quite advantageous compared to NMR which can require 1-2 mg of protein, and experiments potentially taking several days to complete.

Disordered proteins specifically have unique CD spectra with negative bands at 195 nm and low ellipticity beyond 210 nm.[107] In addition to possessing unique spectra, the dynamic nature of IDPs is also advantageous for characterization through CD. By collecting multiple CD measurements of an IDP at different physiological conditions, multiple conformations can be induced.[108] Combining the resultant datasets can provide a more complete picture of the structures of the IDP and the conformations they can undergo.

Compared with NMR and FRET, the versatility of protein CD is quite limited by its lack of residue specificity. Although structural elements are identified and distinguished by unique spectra, CD cannot provide the specific location of these structures. In essence CD cannot provide information about specific residues within a protein, it can only relay general structural information about a protein. As such, recent studies have proposed pairing CD with computations methods like Molecular Dynamic simulations (MD) to generate IDP ensembles that provide insight into IDP protein structure.[109]

Characterizing RecQ Helicases using assays specific for disorder

As the computational approaches discussed require no wet lab work, the starting point for characterizing disordered domains in RecQ helicases is best begun here. Not only is most software free to use, but it also does not require purified peptide as a starting point. In characterizing the disordered N-terminal tail of the BLM yeast homologue Sgs1, Agadir calculations were first performed to obtain preliminary data to identify if any helical content could be predicted prior to performing NMR analysis.[89] Furthermore Agadir has also been used to identify domains necessary for maintenance of Rmi1 stability, a protein that is required during the decatenation of Holliday junctions.[110]

Table 1: Summary of advantages and disadvantages of methods to characterize intrinsically disordered proteins.

Technique	Advantages	Drawbacks
Computational methods	• No experimental work necessary • Small or no cost for usage. • Minimal time needed to obtain results.	• Results provided are based upon existing algorithms. • Less reliable than experimentally obtained results.
Nuclear Magnetic Resonance (NMR)	• Residue specific resolution. • Specific location for structural elements.	• Long duration for experiments • Requires large amounts of purified peptide. • Significant cost for usage.
Förster Resonance Energy Transfer (FRET)	• High sensitivity/resolution. • Provides location of protein structures.	• Requires labelled and purified protein. • Significant cost for usage.
Circular Dichroism (CD)	• Small amount of protein needed. • Short time frame for experiments. • Indicates presence of secondary structure	• Cannot provide specific location for structural elements.

Additionally numerous experimental approaches have been used to inform our understanding of disordered domains in RecQ helicases. In RECQ4L, NMR characterization was used to identify disordered DNA binding domains shown to participate in strand exchange. [111] NMR has also been used to obtain the structure of the C-terminal HRDC of BLM although, most of this domain has been shown to be ordered rather than disordered. [112] Here, NMR was used to identify protein chemical shift perturbations upon titration with DNA, demonstrating the HRDC affinity for ssDNA binding.[113]

CHAPTER 2: IDENTIFYING ALPHA-HELICES IN BLM 1-100 USING NMR

Upon identifying the BLM N-terminus as the region of interest, an experimental methodology was developed. As the N-terminus of BLM (residues 1-647) contains numerous sites of protein-protein interactions (Figure 3), we sought to understand how the disordered structure of BLM may enable protein-protein interactions. As α-helices have been shown to facilitate protein-protein interactions, and these α-helices form a component of protein secondary structure, identification of α-helices in the BLM N-terminus was prioritized. Moreover, the yeast homologue of BLM, Small growth suppressor 1 (Sgs1), contains an α-helix in its N-terminus that is required for binding toTop3.[89]

From previous work with Sgs1, a transient α-helix in Sgs1 spanning from residues D25-A38 was identified.[89] As Sgs1 is the yeast homologue of BLM, we hypothesized that α-helices may also form in the N-terminus of BLM and serve an analogous role. We first began by performing calculations using Agadir. This software, developed from helix-coil theory, enables prediction of helix content and location from the primary sequence of a peptide. The resultant Agadir output predicted the presence of multiple helices distributed along the entirety of the N-terminus (Figure 4). The strongest predicted helices were contained within the first 380 residues. Six 100-residue fragments were designed to partition the N-terminus into separate fragments that could be individually characterized (Figure 4). Hundred residue fragments were selected as our previous work successfully characterized an 80-residue fragment, but a 125 residue fragment was found to be too difficult to characterize due to resonance overlap. [89] Thus, 100 residue fragments were chosen as an optimal size.

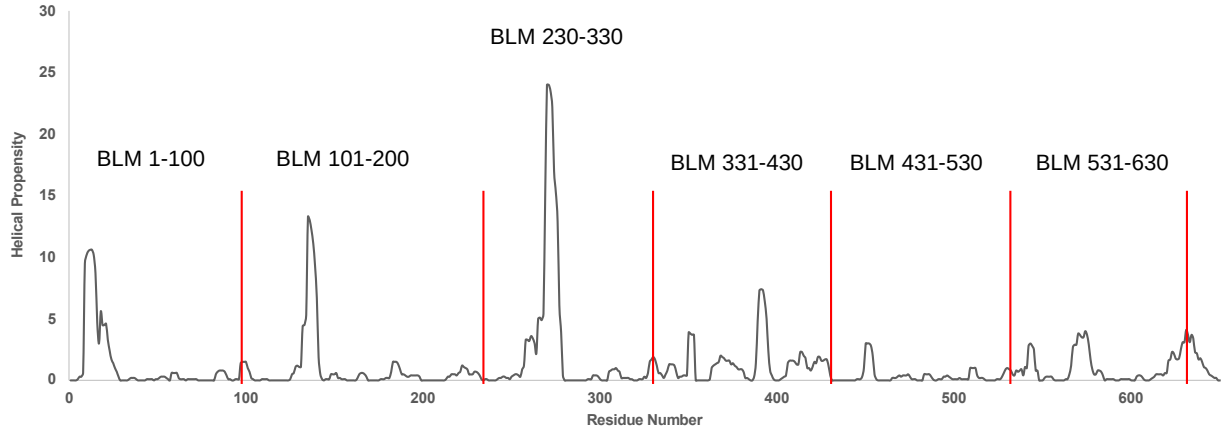

Figure 4: Agadir prediction of helical propensity of BLM N-terminus. The entire sequence of BLM 1-647 was used as the software input, and helical propensity values for each residue were calculated for the entire region. Red lines correspond to 100-residue fragment boundaries that were designed for NMR analysis. Predicted alpha helices are indicated by a region of at least 7 consecutive residues with helical propensity scores > 5.

These 100 residue fragments were cloned into the pET-28a(+) vector (Genscript) which confers an N-terminal polyhistidine tag, cleavable by thrombin. These plasmids were then transformed into chemically competent *E. coli* and expression parameters for the fragments were trialed and identified for five of the six fragments (Figure 5). Appropriate expression of BLM 431-530 could not be identified, so sequencing of the expression plasmid was completed to ensure an absence of truncations or mutations were present in the plasmid DNA. However, no such deletions or mutations could be identified

To ensure that induced BLM fragments contained a functional His-tag, lysate was incubated with Nickel Pro Bond and eluted with high concentration imidazole buffer (Figure 6). For this characterization the four best expressing fragments (BLM 1-100, 101-200, 331-430, and 531-630) were selected.

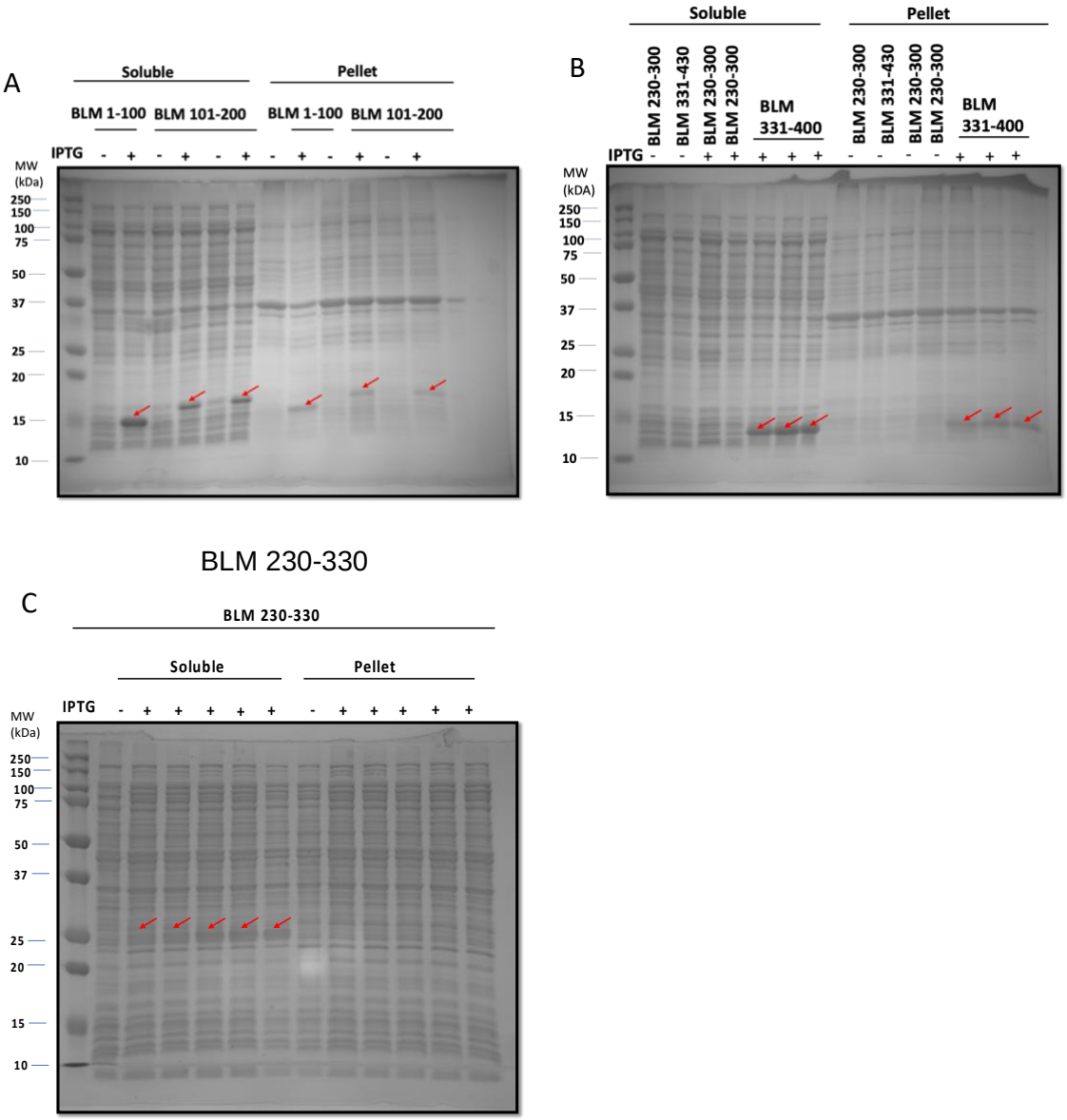

Figure 5: Optimal induction parameters for induction of expression of BLM fragments grown in small scale cultures (25-50 mL of LB). Transformants generated in BL21(DE3) and grow to mid-log phase and spiked with IPTG. Negative control (IPTG deficient) shown by – and IPTG spiked cultures denoted by +. Cell lysates prepared from 5 OD of each culture. Identified parameters are: BLM 1-100 (OD_{600}= 0.6, 0.5 mM IPTG, 3 hr., 37°C; BLM 101-200 (OD_{600}= 0.6, 1.0 mM IPTG, 3 hr., 37°C); BLM 230-330 (OD600= 0.4, 0.5 mM IPTG, 16 hr., 16°C); BLM 331-430 (OD_{600}= 0.6, 0.5 mM IPTG, 3 hr., 37°C).

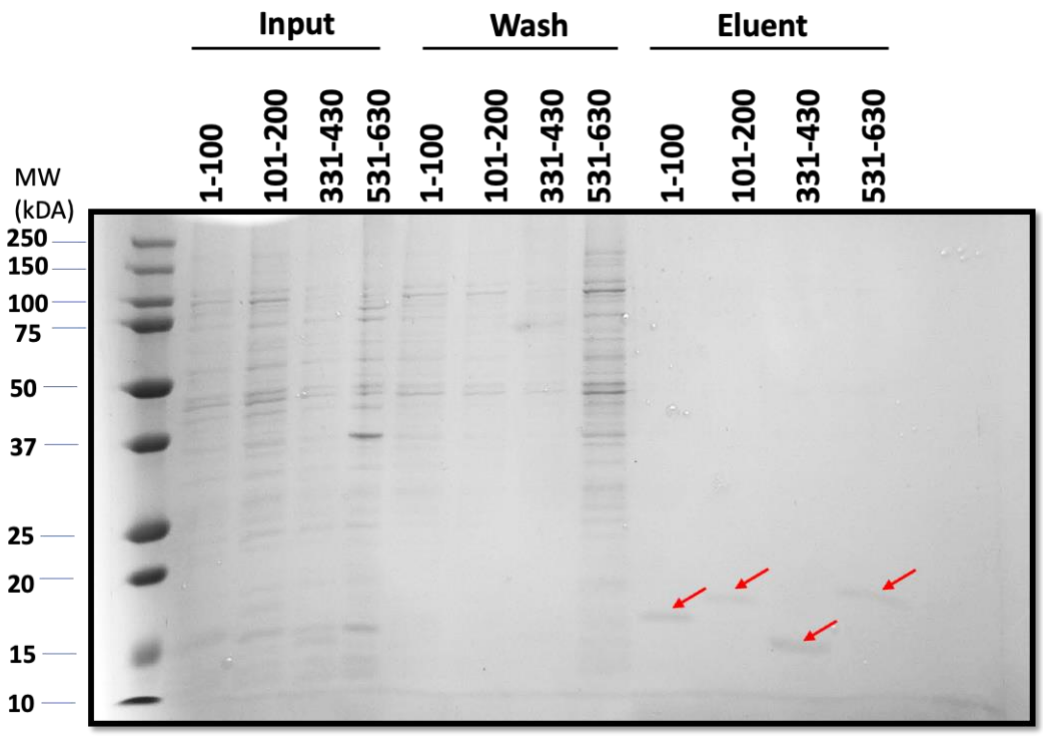

Figure 6: SDS-PAGE of IPTG induced lysate of BLM fragment bound to ProBond Resin. Four highest expressing BLM fragments were transformed in BL21(DE3) and induced with IPTG. Lysate was prepared and incubated with ProBond Nickel resin. Low concentration imidazole buffer was used to wash resin prior to elution with high concentration imidazole. BLM peptide enrichment through binding to Nickel resin is confirmed by presence of singular peptide band in eluent.

In preparation for the acquisition of NMR data, it was necessary for us to determine if induction parameters identified for small scale cultures were suitable for large scale (two-liter cultures). SDS-PAGE analysis revealed that small scale parameters identified for each respective BLM fragments were unsuitable due to an excessive amount of insoluble peptide being produced. As such, new induction parameters for large scale cultures were identified.

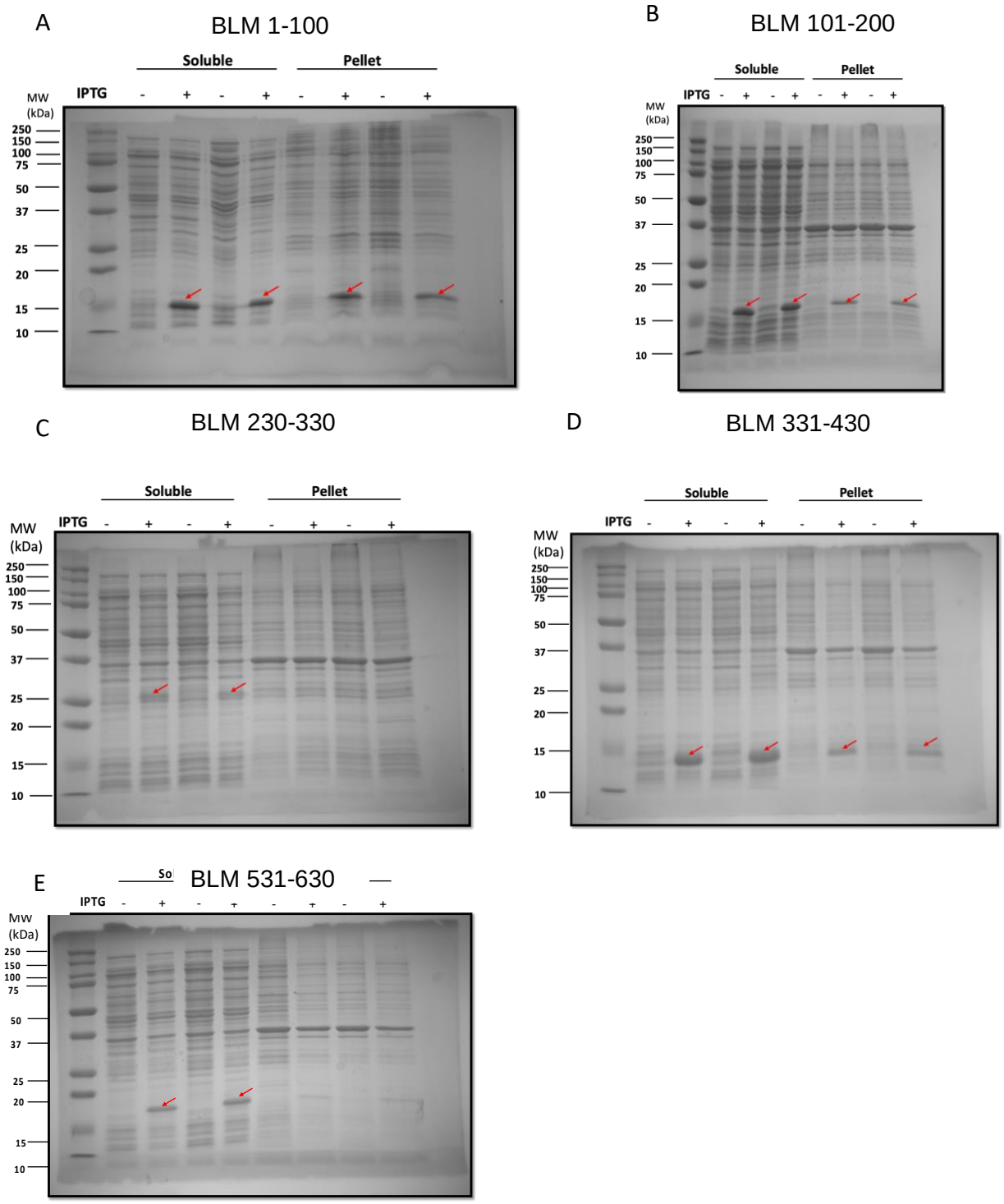

Figure 7: Optimal induction parameters for induction of expression of BLM fragments grown in large scale cultures. (2 liters M9 minimal media). Parameters are: A) BLM 1-100 (OD600= 0.4, 0.5 mM IPTG, 3 hr., 37°C); B) BLM 101-200 (OD600= 0.6, 0.3 mM IPTG, 1 hr.,37°C); C) BLM 230-300 (OD600= 0.4, 0.5 mM IPTG, 16 hr., 16°C); D) BLM 331-430 (OD600= 0.4, 0.5 mM IPTG, 3 hr., 37°C);E) BLM 531-630 (OD600= 0.4, 0.5 mM IPTG, 3 hr., 37°C).

From the Agadir prediction (Figure 4), a helix spanning from residues 7-28 was predicted within the first 100 residues of the BLM N-terminus (Figure 4). This region not only contained a large, predicted helix, but also sites of protein-protein interactions. As such this region was identified as a candidate for purification and NMR analysis. BLM 1-100 was transformed into BL21(DE3) and grown in M9 minimal media as a two-liter culture. Expression of BLM 1-100 was induced using IPTG and confirmed via SDS-PAGE (Figure 8A). Cultures were pelleted and lysed at 1500 psi and cleared via centrifugation at 16000 rpm. Lysate was loaded onto a HisPrep FF16/10 Nickel column, and peptide was isolated to apparent purity using a gradient elution. Successful purification of peptide was confirmed via SDS-PAGE (Figure 8B), and fractions were pooled and dialyzed overnight into Size Exclusion Column (SEC) buffer. Following dialysis purified peptide was incubated with Thrombin Clean Cleave resin to remove the N-terminal 6x-His tag from the peptide (Figure 8C). Following confirmation of successful cleavage, peptide and tag were separated using Size Exclusion Chromatography (Figure 8D). Peptide fractions were pooled and concentrated to a volume of 540 µL (Figure 8E), mixed with deuterated water and NMR analysis was performed using a Varian VNMRS 600 MHz and HNCO, HSQC, and $HNC_{\alpha}C_{\beta}$ data was collected.

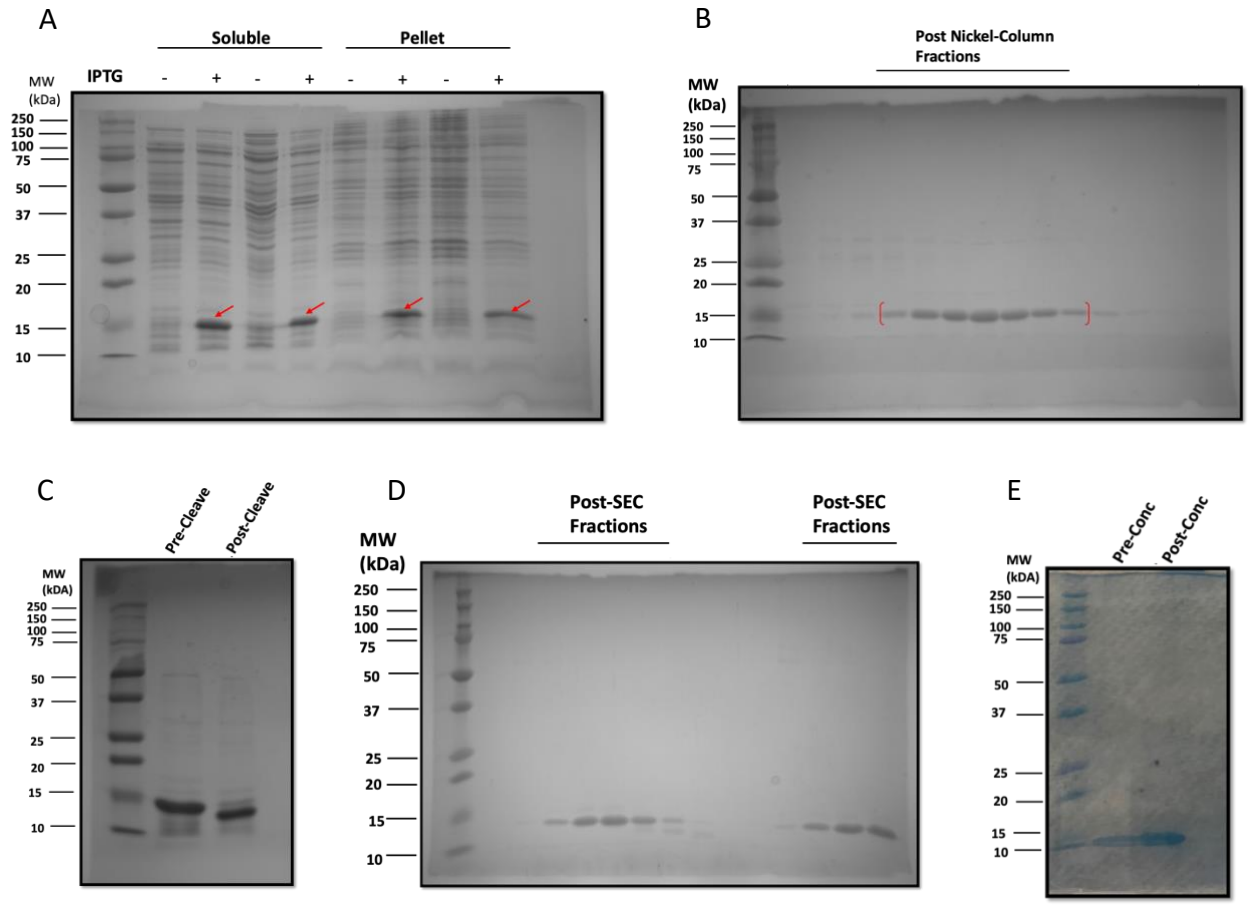

Figure 8: Flowthrough for IMAC purification of His-tagged BLM 1-100 from E. coli lysate. A) SDS-PAGE gel to confirm successful IPTG induction of expression of BLM 1-100 peptide. B) SDS-PAGE after Nickel-FPLC to ensure successful isolation of BLM 1-100 from E. coli lysate. C)SDS-PAGE gel to confirm successful thrombin cleavage of 6x-His tag from BLM 1-100 peptide. D) SDS-PAGE to confirm SEC separation of cleaved 6x-His tag from BLM 1-100 and complete purity of peptide pool, E) Concentration of BLM 1-100 peptide pool to small volume suitable for NMR analysis.

Single label (^{15}N) data was collected, and Heteronuclear Single Quantum Correlation (HSQC) spectra were run at 30, 25, 20, and 16°C. NMR peak intensity was found to increase at lower temperatures, which led us to conclude that running future spectra at lower temperatures would yield the best quality spectra. Double label (^{15}N, ^{13}C) data was collected as HSQC, $HNC_{\alpha}C_{\beta}$, and HNCO spectra. Resonances were assigned using established chemical shifts for each amino acid (Figure 9).[91] In total 86% of the residues in BLM 1-100 could be assigned from the collected spectra (Figure 9).

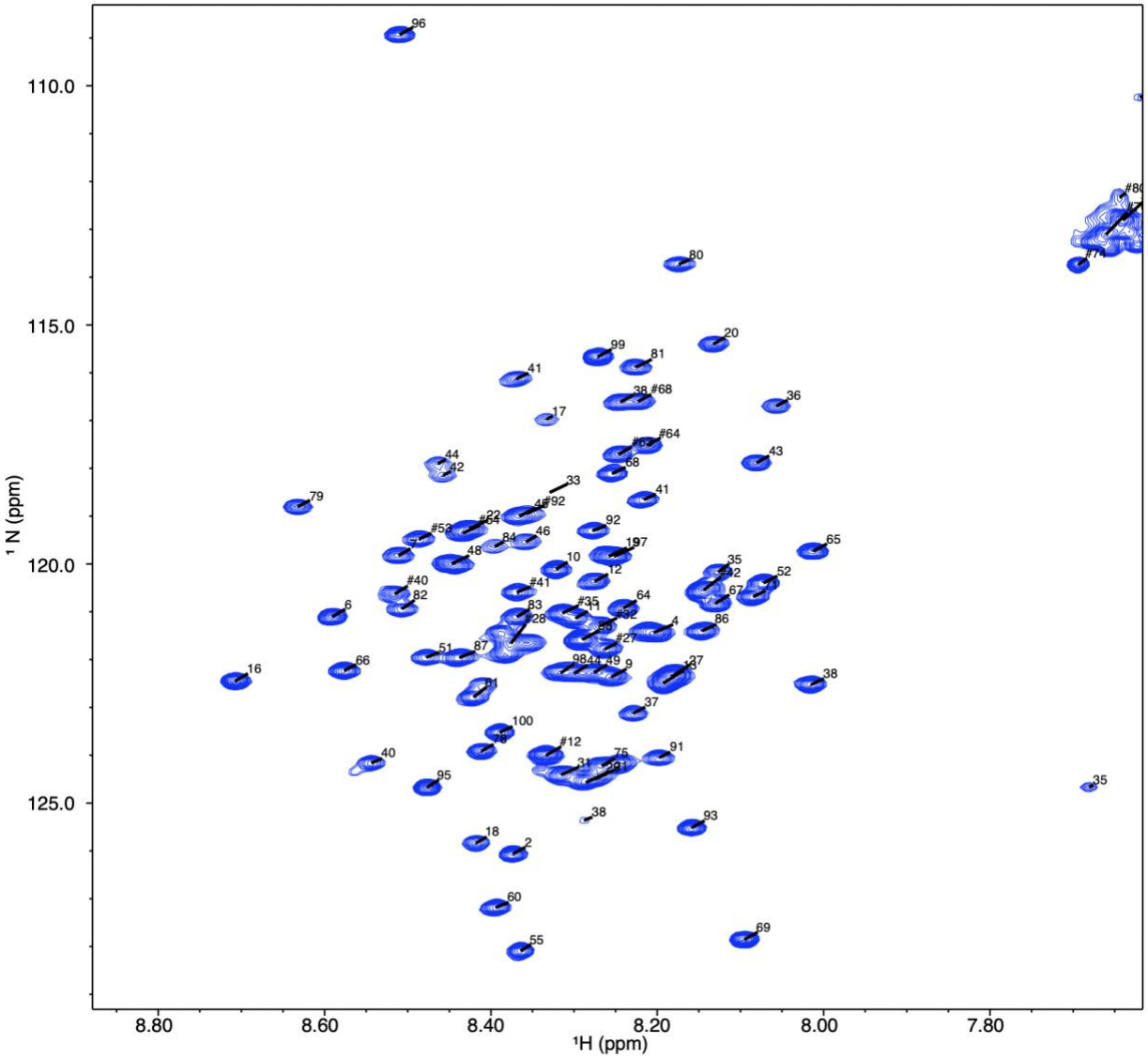

Figure 9: BLM 1-100 resonance assignments. Assigned HSQC spectra of BLM 1-100 taken at 16°C. 86% of residues could be assigned for this specific peptide.

From assignments, both β-carbon and α-carbon chemical shifts were obtained and compared to values calculated for each residue from a random coil library (NcIDP).[82] The difference in chemical shift was calculated by subtracting the NMR obtained shifts from the values obtained from the random coil library. The differences were plotted, with regions of consecutive positive α-shifts (Figure 10A) and consecutive negative β-shifts (Figure 10B) being identified as regions where α-helices were present.

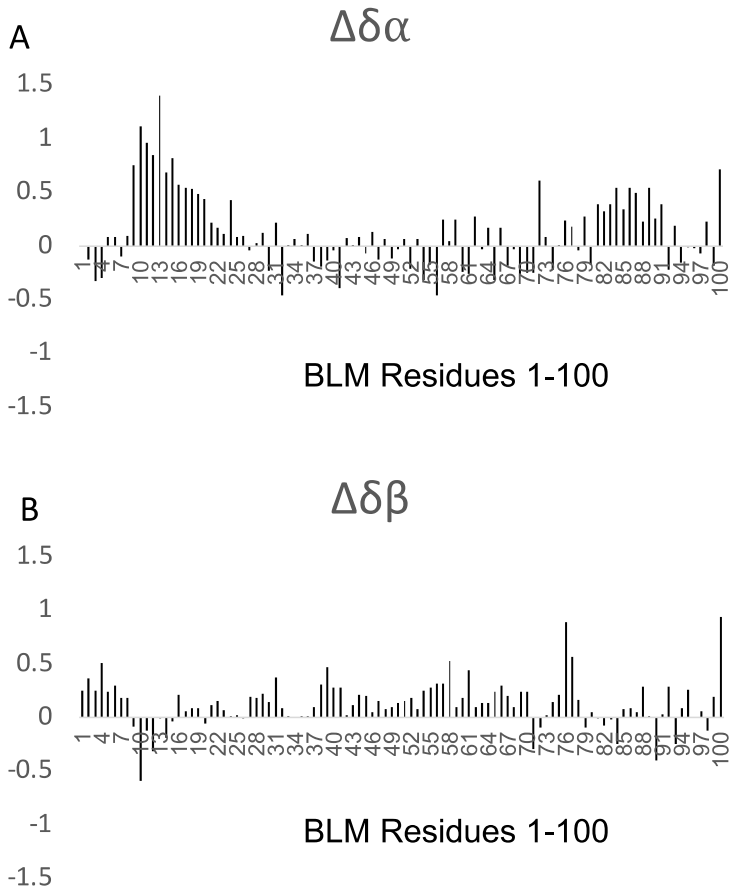

Figure 10: NMR obtained change in chemical shifts. Calculated by subtracting random coil shifts from NMR assignment chemical shifts. A) change in α-carbon shifts, consecutive positive shifts from residues 8-13; 15-22; and 82-86 indicated presence of helices. B) Changes in β-carbon shifts, consecutive negative shifts indicate presence of helices.

From change in chemical shift data an α-helix region spanning from residues 7-28 was identified, and concurrent with Agadir helix predictions (Figure 4). Notably another helix spanning from residues 83-86 was identified from NMR data, that was not predicted from Agadir output. Of note, this identified site corresponds to QVFF region, which has previously been shown to be a binding site on BLM for Mcm6 during G1-phase.[40] Taken together this data coupled with the finding of a helix in this region, further suggests the importance of the role of α-helices in protein-protein interactions. Proline mutagenesis of this region will be necessary to verify the necessity of the α-helix for Mcm6 binding in this region. If this helix is shown to be

necessary for the G1 BLM-Mcm6 interaction, the transience of this helix may account for why Mcm6 has multiple binding sites on BLM.

To confirm the reliability of our Agadir output we overlaid the calculated helical propensity for the first hundred residues with the chemical shift differences obtained from our NMR spectra of the BLM 1-100 region (Figure 11A and 11B).

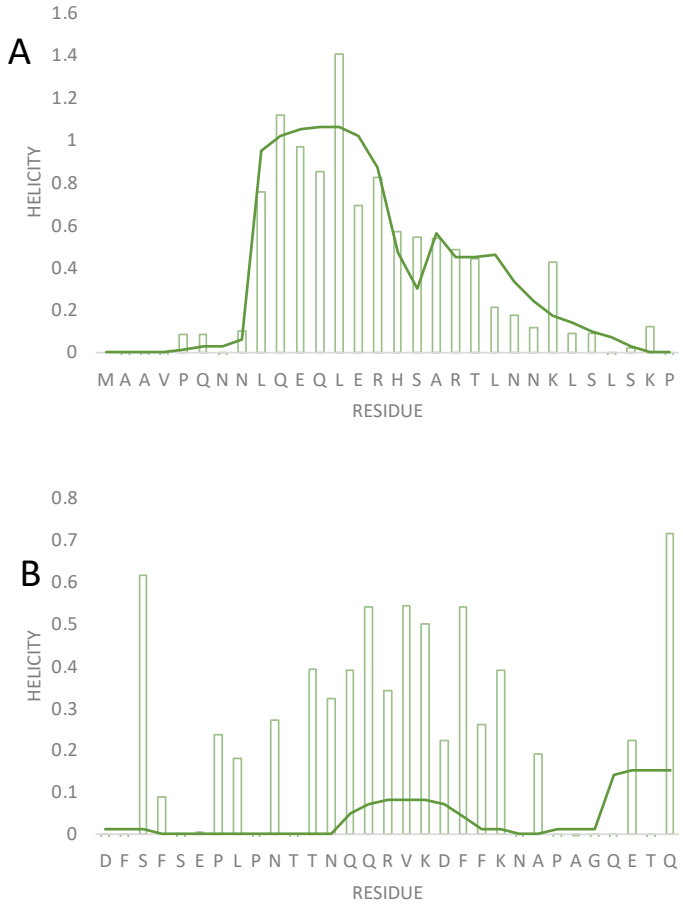

Figure 11: Overlay of Agadir predictions with NMR obtained change in chemical shifts. First identified (Q10-L25) helix overlaps well with prediction, but Agadir did not predict second identified helix (T81-K91).

We found that that the predicted output from Agadir possesses a significant amount of overlap with the BLM 1-100 NMR spectra. This outcome is also concurrent with our previously reported findings considering the overlap between NMR spectra and Agadir predictions with Sgs1.[89]

One crucial difference is that Agadir failed to strongly predict a helix for the 82-86 residue region which was identified by NMR. Taken together with our previous findings this suggests that while Agadir is reliable for predicting helical content, predictions regarding strength and position of the helix are not as reliable.

CHAPTER 3: CLONING AND EXPRESSION OF THE FIRST 380 RESIDUES OF BLM

Upon confirming the presence of helical content in the first hundred residues of BLM and the relative reliability of Agadir as a tool to predict the presence of helical content, molecular cloning was performed to clone a relevant segment of BLM into p-GEX-6p-2 vector for additional experiments. As most of the helical content is contained within the first 380 residues of BLM, the region of BLM 1-380 was selected for cloning. p-GEX vector was specifically selected as it enables expression of the protein of interest along with an N-terminal Glutathione-S transferase (GST) tag, which improves protein solubility and is easily detectable via a GST antibody.

Moreover, due to concerns about peptide solubility we opted to clone BLM 1-380 rather than the entire BLM N-terminus (residues 1-647). To clone our desired 1-380 fragment, we designed primers that annealed to the start codon of BLM and the end of the nucleotide sequence of the first 380 residues of BLM. Forward and reverse primers contained BamHI and EcoRI sites respectively. Following confirmation of presence of desired PCR product (Figure 12A), this resultant product was cloned into TOPO 2.1 vector using TOPO TA Cloning Kit (Invitrogen) (Figure 12B). A sequential diagnostic digest was performed first using EcoRI (Figure 12C) followed by BamHI (Figure 12D) to confirm integration of PCR product into vector, and positive clones were sequenced to ensure insert was mutation free.

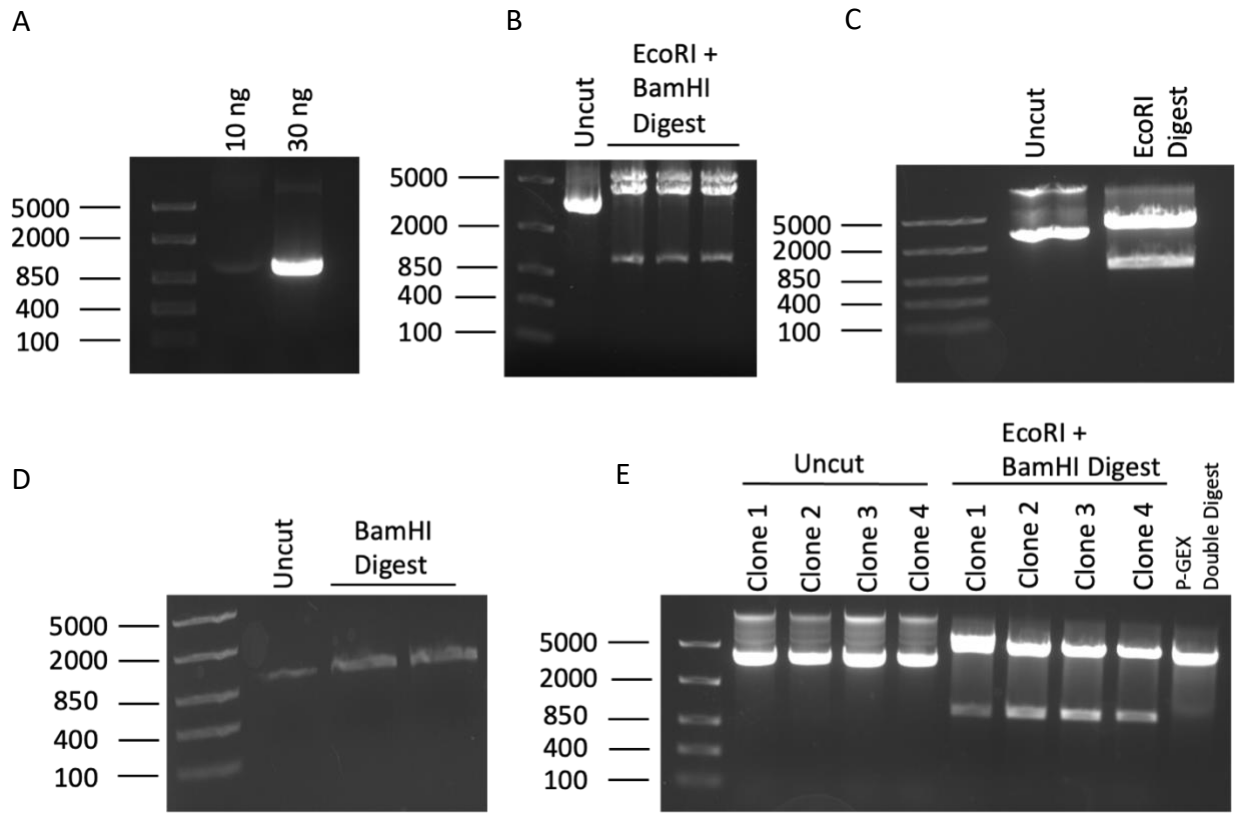

Figure 12: Gel confirmation of cloning steps A) PCR amplification with two amounts of BLM 1-380 product with BamHI and EcoRI ends. B) BamHI + EcoRI digest of Topo 2.1 clones with BLM 1-380 insert. C) Single digest of Topo clone with EcoRI. D) BamHI digest of previous EcoRI digested clones. E) BamHI + EcoRI diagnostic digest of p-GEX + BLM 1-380 clones.

p-GEX-6p-2 was doubly digested with BamHI and EcoRI and incubated with BLM 1-380 insert in a 10:1 insert:vector ratio. Presence of insert was confirmed with a diagnostic digest and clones were sequenced to confirm absence of mutations (Figure 12E). Following successful cloning we confirmed presence of GST-BLM 1-380 via SDS-PAGE (Figure 13). To confirm functionality of the of the GST tag a binding assay was performed using glutathione magnetic agarose beads. Additionally the integrity of the GST tag was additionally confirmed through Western Blotting with a GST-antibody (BioLegend) (Figure 14).

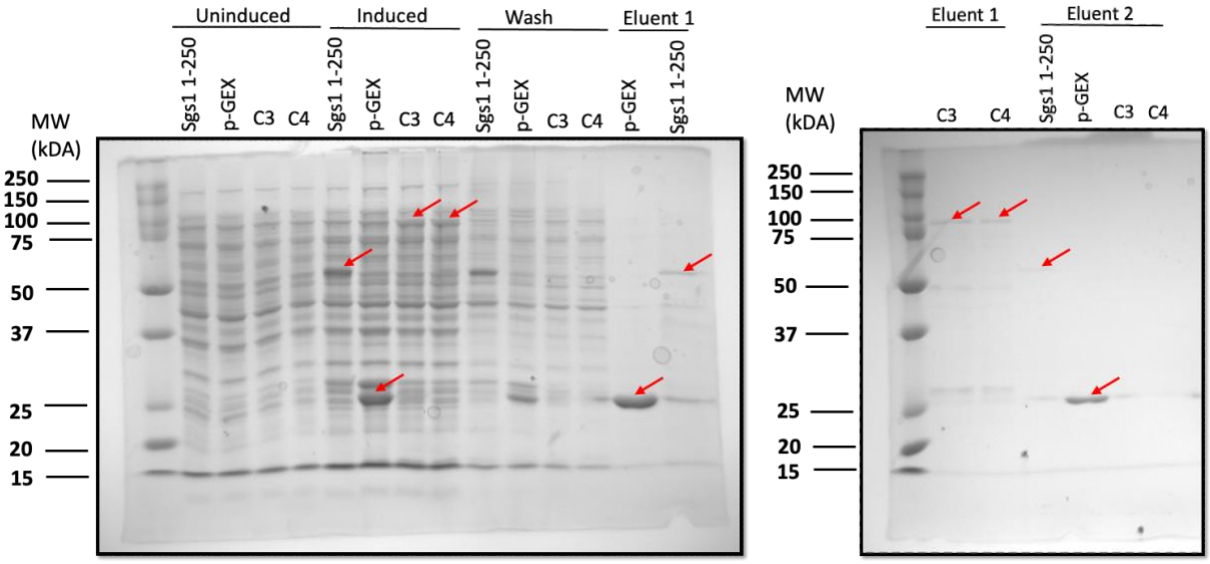

Figure 13: Confirmation of induction of expression and integrity of GST tag for two BLM 1-380 clones. Lysates generated by transforming vectors into BL21(DE3) and inducing expression with IPTG. Lysates were incubated with GST magnetic beads, washed with low salt buffer, and eluted with buffer containing reduced glutathione. GST tagged Sgs1 1-250 and empty p-GEX-6p-2 vector are used as positive and negative controls respectively. Location of each peptide is shown in red brackets. BLM 1-380 clones run at approximately 80 kDa, Sgs 1-250 runs at 50 kDa, and p-GEX runs at 25 kDa.

From the Agadir output, locations of predicted α-helices were identified. The output allowed us to determine the residues of highest helix propensity-thereby allowing us to identify single residues most needed for helix formation. Thus, these residues were selected as candidates for mutagenesis, with the rationale that targeting these residues would have the highest likelihood of collapsing/destroying the helix.

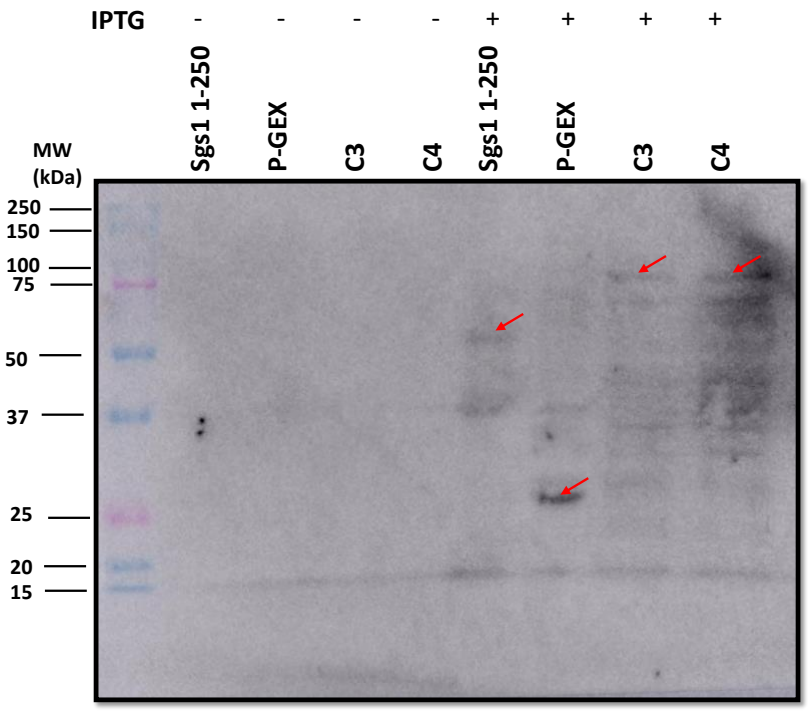

Figure 14: Anti-GST Western blot. GST Sgs1 1-250 is a previously made fusion protein, p-GEX empty vector encodes GST tag, and C3 and C4 are GST BLM 1-380 positive clones. Expected weights are 50 (Sgs1 1-250), 26 (p-GEX), and 80 (C3 and C4) kDa respectively.

From the data N7P, N8P, L13P, A135P, and S269P, K271P, K272P were all identified as proline mutants with the intent of breaking α-helices that spanned in these regions. L13P and A135P were both residues with the highest helical propensity. By contrast N7P; N8P; and S269P were all proposed helix N-caps, referring to them being the first residue where the helix forms. As it has been shown that proline mutations made within the first turn of the helix (3-4 residues) can be tolerated we sought to design combined mutants by targeting residues with the highest propensity as well as mutagenizing the N-cap. Using the GST BLM 1-380 clone cDNA as a template, proline mutagenesis will be performed to generate BLM mutants with disrupted helices. Upon making these mutants, pulldowns will be performed to determine if abolishment of α-helices in these mutants results in reduced protein-protein interactions compared to endogenous BLM.

CHAPTER 4: DISCUSSION

Here we have used a methodology previously used to characterize the yeast helicase Sgs1 and have used this to characterize its human homologue BLM. We used two computational methods (Agadir and IUPred/ANCHOR) for the reliable prediction of binding regions and α-helix position simply by knowing the sequence of the protein of interest. Of note, in our previous work with Sgs1 we used two fragments sized 80 and 125 residues.[89] Upon determining that a 125 residue fragment was too large to be accurately assigned, this fragment was truncated to 80 residues instead.[89] Here we have used a 100-residue fragment and have shown that it can be accurately assigned with minimal overlapping residues complicating analysis. Thus, additional NMR experiments can be performed using 100-residue peptides.

From our calculated chemical shifts we have identified two distinct regions of α-helical content, one of which (8-26) overlaps well with our Agadir prediction (Figure 4). This trend follows with our conclusions from our studies with Sgs1 where we show that Agadir predictions overlap almost identically with NMR obtained structural models.[89] Taken together our findings indicate that our approach of using Agadir to rationally design small peptides which are then characterized via NMR is an efficient and reliable method for determining structural elements for IDPs. As Agadir predictions were calculated using the primary sequence of the BLM N-terminus, we wanted to ensure that partitioning the N-terminus into 100-residue fragments did not disrupt structural elements in this fragment. Our findings that our Agadir predictions mirror the NMR obtained data of BLM 1-100 demonstrate that truncating BLM into its first 100 residues does not interfere with structural elements contained in this region. Thus, our approach of designing fragments using residue boundaries of low helical propensity seems to be the most optimal

means of partitioning a protein for NMR analysis. From our previous work, we have shown that the interaction between BLM and Mcm6 is cell cycle specific, as BLM has been shown to bind to Mcm6 in multiple sites.[40] Given one of our NMR identified helices-which was not predicted by Agadir- overlaps with the Mcm6 binding region, this helix is likely to be necessary for the BLM-Mcm6 interaction. Moreover, the small magnitude of this helix, suggests its presence is transient.[65] This transience may in turn account for why Mcm6 interacts with multiple binding sites in BLM rather than a fixed site, as this identified helix may not be adopted during all phases of the cell cycle. Additional experiments involving FACS sorting cells in G1 and S-phase and monitoring protein conformations in these live cells through FRET may enable us to identify changes in BLM protein conformation during the cell cycle. [114]

Given the relative success of findings with BLM 1-100, additional characterization of remaining BLM fragments is merited, as this will provide structural information into the remainder of the N-terminus. The two best candidates for subsequent analysis are BLM 101-200 and BLM 230-330 due to their predicted helices. Specifically, BLM 101-200 contains numerous sites of protein-protein interactions while BLM 230-330 contains the largest predicted helix. Although the goal is to characterize the entire N-terminus, a clear hierarchy exists in terms of fragment priority as these two fragments most merit characterization. In addition, induction parameters for BLM 431-530 have yet to be identified, and the consequent lack of peptide expression prevents obtaining sufficient peptide for NMR characterization.

Although induction parameters have been identified for both BLM 101-200 and BLM 230-330, purification of sufficient peptide for NMR analysis has posed two distinct challenges. Expression of BLM 230-330 has only been obtained at 16°C and expression levels are considerably lower than other BLM fragments. As approximately 2 mg of protein are needed for NMR runs, a simple optimization of increasing starting culture sizes will be essential to ensure sufficient

peptide can be obtained. The major limitation is that the BLM 230-330 peptide has a weight of 13 kDa, but SDS-PAGE reveals a peptide that runs at approximately twice this molecular weight (Figure 5C). As the dimerization domain of BLM is thought to be contained in this region, this may account for the reason the peptide dimerizes.[48] Such dimerization was not expected as this is merely a 100-residue fragment of BLM, but the observation that this domain retains functionality further demonstrates that our rationally designed BLM fragments retain endogenous structure and function.

BLM 101-200 presented additional challenges; specifically endogenous degradation during the induction step. Although we identified parameters that minimized the degradation of BLM 101-200 peptide, we could not successfully eliminate it entirely. As NMR requires purified peptide, any degradation products needed to be removed to avoid peaks overlapping. We engineered a C-terminal streptavidin tag onto our BLM 101-200 peptide and performed a dual affinity purification by passing peptide first through a His and then through a Streptavidin column and confirmed successful purification through SDS-PAGE. Successful characterization of this fragment using this methodology will rely on increasing starting culture size to maximize yield.

From the Agadir predictions (Figure 4) we have designed proline mutants to disrupt predicted helices in the BLM N-terminus. Upon creating these mutants, we would anticipate proteins whose binding sites overlap with these helices will no longer be able to interact with these BLM proline mutants. To test this hypothesis, pulldowns will need to be performed using these mutants, to see if helix destruction results in downregulation of binding between BLM and its interacting proteins. Finally, to ascertain the biological significance of BLM protein-protein interactions we want to establish stable mammalian cell lines to observe how these BLM mutants behave *in vitro*. Upon establishing cell lines, we will perform sensitivity assays to see how these mutations impact DNA damage response and overall BLM function.

CHAPTER 5: MATERIALS AND METHODS

Plasmid design

To identify potential helical content, a helix prediction software called Agadir was used.[63] The primary sequence of BLM was entered, and helical propensity values were obtained and plotted against the residue number (Figure 1). From the output, regions containing helices were separated into 100 residue fragments, with fragments being designed such that the predicted helices were close to the center of the fragment (Figure 4). This minimizes the likelihood of disrupting a predicted helix due to partitioning of the N-terminus. Following successful design of 100 residue fragments, they were then cloned into pET-28a(+) vector conferring encoded fragments with an N-terminal 6X His-tag (Genscript).

Confirming induction parameters for BLM fragments

To identify optimal parameters for BLM peptide expression fragments 20 µL of BL21(DE3) was transformed with 2 ng of plasmid.[89] Upon addition of plasmid DNA cells were allowed to incubate at 4°C for minutes and then heat shocked for 30 seconds at 42°C and rescued through addition of 100mL of prewarmed SOC (0.5 % Yeast extract, 2% Tryptone, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM Glucose). This mixture was then shaken at 37°C and plated in its entirety on a prewarmed LB-kan plate. Colonies were allowed to grow overnight at 37°C and were then streaked out to obtain singles. Single colonies were then inoculated in LB-kan overnight. These overnight cultures were then used to inoculate a fresh culture at OD_{600}=0.04 in a volume of 50 mL and grown at 37°C to OD_{600}=0.6.[89] A 5 OD aliquot of this culture was then collected as a control, with the remainder of the culture being spiked with Isopropyl ß-D-1-thiogalactopyranoside (IPTG). To identify optimal parameters, induction

conditions were parameterized by IPTG concentration, induction time, and temperature. To confirm induction success an SDS-PAGE was performed using a 16% gel and the most optimal induction conditions were identified for five out of the six fragments (Figure 7).

To identify induction conditions for large scale cultures, transformants were first grown overnight in 50 mL of M9-Minimal media (2 mM Na2HPO4, 22 mM KH2PO4, 8 mM NaCl, 2 mM MgSO4, 11 mM D-glucose, 0.1 mM CaCl2, 10 µM FeCl3, 1 mg of Vitamin B1/L, pH=7.35, 10mg kanamycin). Cultures were inoculated in two 1-liter cultures at OD_{600}=0.04 and grown to an OD_{600}=0.4. A 5 OD aliquot was collected for SDS-PAGE analysis, and the remaining cultures were spiked with IPTG and allowed to grow at 37°C.[89] To confirm the success of induction, SDS-PAGE analysis was performed using a 16% gel.[89]

Pro-bond resin

To confirm integrity of the 6X-His tag, the four best expressing fragments were selected for binding to Nickel Pro Bond Resin. 2ng of plasmid was transformed into 20 µL of chemically competent BL21(DE3) cells, via heat shock at 42°C for 30 seconds. Transformants were rescued with 100 µL of prewarmed (37°C) SOC broth and shaken at 37°C for 1 hour. These transformants were plated after shaking on a prewarmed LB-kanamycin plate and allowed to grow overnight at 37°C. Transformants were then streaked out onto a fresh prewarmed LB-kan plate and allowed to grow overnight at 37°C. Single colony transformants were inoculated into 50 mL of LB-kan (10mg/mL) at an OD_{600}=0.04 and grown to OD_{600}=0.6.[89] Cultures were then spiked with 1.0 mM IPTG and induced at 37°C for 3 hours.[89] A 5 OD aliquot of these cultures was then collected and centrifuged at 3400 rpm for 10 minutes with the resultant supernatant removed. The pellet was then resuspended in A1 lysis buffer (50 mM Na_2HPO_4,300 mM NaCl, 10 mM Imidazole) and sonicated three times over thirty second intervals pulsing on and off every second. Sonicated lysate was then centrifuged at 14,000 rpm for 10 minutes with

resultant supernatant collected. This supernatant was then incubated with Pro Bond Nickel Resin (Invitrogen) for 60 minutes.[89] The supernatant was then separated from the resin, and the resin was washed with wash buffer (50 mM NaH_2PO_4, pH 8.0 0.5 M NaCl) and then incubated for 10 minutes with elution buffer (50 mM NaH_2PO_4, pH 8.0 0.5 M NaCl, 250 mM Imidazole). flowthrough, wash and eluents were then run on an SDS-PAGE to confirm successful binding to the nickel resin (Figure 6).

Purification of BLM 1-100 peptide

BLM 1-100 was cloned in the pET-28a(+) vector (GenScript). 2ng of plasmid was transformed into 20 µL of chemically competent BL21(DE3) cells, via heat shock at 42°C for 30 seconds. Transformants were rescued with 100 µL of prewarmed (37°C) SOC broth and shaken at 37°C for 1 hour. These transformants were plated after shaking on a prewarmed LB-kan plate and allowed to grow overnight at 37°C. Transformants were then streaked out onto a fresh prewarmed LB-kan plate and allowed to grow overnight at 37°C. Single colony transformants were inoculated in 50mL of M9 media and allowed to grow overnight at 37°C. For NMR single label (^{15}N) runs isotopic ammonium chloride was used, and for double label (^{15}N and ^{13}C) isotopic ammonium chloride and D-glucose was used.

From the 50 mL overnight cultures, two one-liter cultures were inoculated at OD_{600}=0.04 and grown at 37°C to OD_{600}=0.4. A 5 OD aliquot (Uninduced) was collected, and the remaining culture was spiked with 0.5 mM IPTG and allowed to grow at 37°C for 3 hours. The 5 OD aliquot was centrifuged at 4°C for 3400 rpm at 10 minutes, and the resultant supernatant was decanted. The pellet was then resuspended in 200 µL of A1 Lysis buffer (50 mM Na_2HPO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) on ice.[89] This lysate was then sonicated (QSonica) with 3 rounds of 30 seconds set to pulse on and off at one second intervals at an amplitude of 30%. The resultant sonicated lysate was then centrifuged at 4°C for 10 minutes at 14000 rpm. The

supernatant (soluble) was separated from the pellet (insoluble) and both were resuspended in 2X Laemlli buffer. For the soluble fraction the supernatant was mixed in a 1:1 ratio with 2X Laemlli, whereas the pellet was resuspended with 200 µL of 2X Laemmli. Samples were boiled at 95°C for five minutes and stored at -20°C for subsequent SDS-PAGE analysis.

After induction another 5OD aliquot (Induced) was collected and lysate was prepared as previously described. The remainder of the two-liter cultures were pelleted by centrifugation at 4000xg at 4°C, stored at -80°C. Pellet was subsequently resuspended in 25mL Lysis buffer (300 mM NaH_2PO_4, 500 mM NaCl, Pierce Protease Inhibitor Tablet) and lysed at 16,000 psi using a French Press. Lysate was cleared by centrifuging at 18000 rpm at 4°C for 60 minutes. Supernatant (soluble) was separated from pellet and was injected into a 50 mL superloop. The superloop enabled the lysate to be processed by an AKTA FPLC (GE) loaded onto a HisPrep FF16/10 Nickel column. Column was washed with Wash Buffer (300 mM NaH_2PO_4, 500 mM NaCl, 10 mM Imidazole, pH=8.00) and protein was eluted with elution buffer (300 mM NaH_2PO_4, 500 mM NaCl, 500mM Imidazole, pH=8.00) using a gradient elution (gradient from 50-500 mM). 4 mL fractions were collected, and peptide containing fractions were pooled and dialyzed overnight in SEC buffer (300 mM NaH_2PO_4, 500 mM NaCl, 1mM EDTA, pH=7.00) at 4°C.[89] N-terminal His-tag was cleaved using Sigma RECOMT Clean-cleave kit and cleaved product was loaded onto HiLoad 16/600 Superdex 75 pg SEC column. This purified cleaved peptide pool was passed over the column in 2 mL injections and 2.5 mL fraction were collected from the column. 180 mL of SEC Buffer was used as the liquid phase for each injection and the range from 40-60 mL was collected.[89] The peptide containing pool was concentrated from approximately 60 mL to 540 µL 60 µL of deuterated water was added to this sample for NMR analysis.

NMR analysis

Data was collected using a Varian VNMRS 600 MHz spectrometer. Data collection was performed in conjunction with our collaborator Dr. Gary Daughdrill. For the single label (^{15}N) run HSQC, $HNC_{\alpha}C_{\beta}$, and HNCO data was collected at 15°C. Data was processed using NMRFx and analysis and assignments were done with NMRViewJ. Secondary shift assignments were completed by calculating chemical shifts from a random coil library (ncIDP) and subsequently subtracted from the NMR obtained chemical shifts for the corresponding residue.[82] Consecutive positive -shifts and consecutive negative b-shifts were assigned as α-helices.[84]

Cloning BLM 1-380 into p-GEX-6p-2 and GST binding assay

BLM cDNA was previously cloned into pcDNA3 mammalian expression vector. Using this cDNA as a template, BLM 1-380 was cloned with BamHI and EcoRI sites flanking the 5' and 3' ends respectively. Forward: 5' CTTA GGATCC atggctgctgttcctcaaaataatc 3', Reverse: 5' CTTA GAATTC TTA acagatgtgctccatcacatg 3'. PCR was performed using 30ng of template, 10 mM forward and reverse primers, 2.5 mM dNTP's, and 0.625 units Pfu Ultra (Agilent). This PCR product was subsequently cloned into TOPO 2.1 Vector using TOPO Cloning Kit (Invitrogen) and digested with EcoRI. This product was subsequently gel extracted with Qiaquick Gel Extraction Kit (Qiagen) and re-digested with BamHI. This newly digested product is then gel extracted once more. p-GEX 6p-2 (GE Healthcare) was double digested with BamHI and EcoRI and dephosphorylated with Antarctic Phosphatase Kit (NEB). Subsequently BLM 1-380 insert was ligated with this digested and dephosphorylated p-GEX-6p-2 vector using DNA ligase kit (Takara). Using this clone as a template, site directed mutagenesis was performed using forward and reverse primers with nucleotide mutations made in the middle of the primer. Mutagenesis products were digested with 20 units of DpnI (New England BioLabs) for 1 hour at 37°C.

100mL of DH5a was mixed with 100mL of 40% glycerol, and 5 mL of Dpn1 digested product was added to this mix. 100mL of this was then transferred to an electroporation cuvette (FisherBrand) and electroporated. Following electroporation this product was mixed with 1 mL of prewarmed LB and transferred back to original mix. This mix was shaken for 1 hour at 37°C with 250 µL of this culture plated onto four LB-amp plates (100 mg/mL) and allowed to grow overnight. Colonies were then streaked out and single colonies were inoculated overnight in LB-amp, and miniprepped using a plasmid mini-prep kit (Qiagen). Positive clones were then used for subsequent assays.

GST-binding assay was completed by transforming GST BLM 1-380 into BL21(DE3) and expression was induced as previously described. Cells were lysed in GST binding buffer using sonication as previously described. Lysate was incubated with Glutathione Magnetic Agarose Beads (ThermoFisher) and washed with wash buffer (125 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH=7.4). Beads bound with GST-BLM 1-380 were eluted with 250 µL of Wash buffer (125 mM Tris-HCl, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, 50 mM reduced glutathione, pH=8.0)

REFERENCES

[1] Radivojac, P., Z. Obradovic, D. K. Smith, G. Zhu, S. Vucetic, C. J. Brown, J. D. Lawson, and A. K. Dunker. "Protein Flexibility and Intrinsic Disorder." *Protein Sci* 13, no. 1 (Jan 2004): 71-80.

[2] Uversky, V. N. "Intrinsically Disordered Proteins from a to Z." *Int J Biochem Cell Biol* 43, no. 8 (Aug 2011): 1090-103.

[3] Romero, P., Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guilliot, and A. K. Dunker. "Thousands of Proteins Likely to Have Long Disordered Regions." *Pac Symp Biocomput* (1998): 437-48.

[4] Romero, P., Z. Obradovic, C. R. Kissinger, J. E. Villafranca, E. Garner, S. Guilliot, and A. K. Dunker. "Thousands of Proteins Likely to Have Long Disordered Regions." *Pac Symp Biocomput* (1998): 437-48.

[5] Fuxreiter, M. "Fuzziness: Linking Regulation to Protein Dynamics." *Mol Biosyst* 8, no. 1 (Jan 2012): 168-77.

[6] Tedeschi, G., M. Mangiagalli, S. Chmielewska, M. Lotti, A. Natalello, and S. Brocca. "Aggregation Properties of a Disordered Protein Are Tunable by Ph and Depend on Its Net Charge Per Residue." *Biochim Biophys Acta Gen Subj* 1861, no. 11 Pt A (Nov 2017): 2543-50.

[7] Shojania S, O'Neil JD. Intrinsic disorder and function of the HIV-1 Tat protein. Protein Pept Lett. 2010 Aug;17(8):999-1011.

[8] Bah, A., & Forman-Kay, J. D. (2016). Modulation of Intrinsically Disordered Protein Function by Post-translational Modifications. The Journal of biological chemistry, 291(13), 6696–6705.

[9] Arai, M., K. Sugase, H. J. Dyson, and P. E. Wright. "Conformational Propensities of Intrinsically Disordered Proteins Influence the Mechanism of Binding and Folding." *Proc Natl Acad Sci U S A* 112, no. 31 (Aug 4 2015): 9614-9.

[10] Tompa, P., N. E. Davey, T. J. Gibson, and M. M. Babu. "A Million Peptide Motifs for the Molecular Biologist." *Mol Cell* 55, no. 2 (Jul 17 2014): 161-9.

[11] Wright, P. E., and H. J. Dyson. "Intrinsically Disordered Proteins in Cellular Signalling and Regulation." *Nat Rev Mol Cell Biol* 16, no. 1 (Jan 2015): 18-29.

[12] Pontius, B. W. "Close Encounters: Why Unstructured, Polymeric Domains Can Increase Rates of Specific Macromolecular Association." *Trends Biochem Sci* 18, no. 5 (May 1993): 181-6.

[13] Stein, A., R. A. Pache, P. Bernado, M. Pons, and P. Aloy. "Dynamic Interactions of Proteins in Complex Networks: A More Structured View." *FEBS J* 276, no. 19 (Oct 2009): 5390-405.

[14] Van Roey, K., T. J. Gibson, and N. E. Davey. "Motif Switches: Decision-Making in Cell Regulation." *Curr Opin Struct Biol* 22, no. 3 (Jun 2012): 378-85.

[15] Cobb, J. A., L. Bjergbaek, and S. M. Gasser. "Recq Helicases: At the Heart of Genetic Stability." *FEBS Lett* 529, no. 1 (Oct 2 2002): 43-8

[16] Ellis, N. A., J. Groden, T. Z. Ye, J. Straughen, D. J. Lennon, S. Ciocci, M. Proytcheva, and J. German. "The Bloom's Syndrome Gene Product Is Homologous to Recq Helicases." *Cell* 83, no. 4 (Nov 17 1995): 655-66

[17] Yu, C. E., J. Oshima, Y. H. Fu, E. M. Wijsman, F. Hisama, R. Alisch, S. Matthews*, et al.* "Positional Cloning of the Werner's Syndrome Gene." *Science* 272, no. 5259 (Apr 12 1996): 258-62.

[18] Kitao, S., N. M. Lindor, M. Shiratori, Y. Furuichi, and A. Shimamoto. "Rothmund-Thomson Syndrome Responsible Gene, Recql4: Genomic Structure and Products." *Genomics* 61, no. 3 (Nov 1 1999): 268-76.

[19] Oshima, J., H. Kato, Y. Maezawa, and K. Yokote. "Recq Helicase Disease and Related Progeroid Syndromes: Recq2018 Meeting." *Mech Ageing Dev* 173 (Jul 2018): 80-83.

[20] Lu, L., W. Jin, H. Liu, and L. L. Wang. "Recq DNA Helicases and Osteosarcoma." *Adv Exp Med Biol* 804 (2014): 129-45

[21] Balajee AS. Human RecQL4 as a Novel Molecular Target for Cancer Therapy. Cytogenet Genome Res. 2021;161(6-7):305-327.

[22] Lebel, M., and P. Leder. "A Deletion within the Murine Werner Syndrome Helicase Induces Sensitivity to Inhibitors of Topoisomerase and Loss of Cellular Proliferative Capacity." *Proc Natl Acad Sci U S A* 95, no. 22 (Oct 27 1998): 13097-102.

[23] Ichikawa, K., T. Noda, and Y. Furuichi. "[Preparation of the Gene Targeted Knockout Mice for Human Premature Aging Diseases, Werner Syndrome, and Rothmund-Thomson Syndrome Caused by the Mutation of DNA Helicases]." *Nihon Yakurigaku Zasshi* 119, no. 4 (Apr 2002): 219-26.

[24] Mann, M. B., C. A. Hodges, E. Barnes, H. Vogel, T. J. Hassold, and G. Luo. "Defective Sister-Chromatid Cohesion, Aneuploidy and Cancer Predisposition in a Mouse Model of Type Ii Rothmund-Thomson Syndrome." *Hum Mol Genet* 14, no. 6 (Mar 15 2005): 813-25.

[25] Hoki, Y., R. Araki, A. Fujimori, T. Ohhata, H. Koseki, R. Fukumura, M. Nakamura*, et al.* "Growth Retardation and Skin Abnormalities of the Recql4-Deficient Mouse." *Hum Mol Genet* 12, no. 18 (Sep 15 2003): 2293-9.

[26] Zhou, H. X., X. Pang, and C. Lu. "Rate Constants and Mechanisms of Intrinsically Disordered Proteins Binding to Structured Targets." *Phys Chem Chem Phys* 14, no. 30 (Aug 14 2012): 10466-76

[27] Zhou, H. X. "Intrinsic Disorder: Signaling Via Highly Specific but Short-Lived Association." *Trends Biochem Sci* 37, no. 2 (Feb 2012): 43-8

[28] Yeom, G., J. Kim, and C. J. Park. "Investigation of the Core Binding Regions of Human Werner Syndrome and Fanconi Anemia Group J Helicases on Replication Protein A." *Sci Rep* 9, no. 1 (Sep 30 2019): 14016.

[29] Sharma, S., P. Phatak, A. Stortchevoi, M. Jasin, and J. R. Larocque. "Recq1 Plays a Distinct Role in Cellular Response to Oxidative DNA Damage." *DNA Repair (Amst)* 11, no. 6 (Jun 1 2012): 537-49.

[30] Newman, J. A., H. Aitkenhead, P. Savitsky, and O. Gileadi. "Insights into the Recq Helicase Mechanism Revealed by the Structure of the Helicase Domain of Human Recql5." *Nucleic Acids Res* 45, no. 7 (Apr 20 2017): 4231-43.

[31] Ohlenschlager, O., A. Kuhnert, A. Schneider, S. Haumann, P. Bellstedt, H. Keller, H. P. Saluz*, et al.* "The N-Terminus of the Human Recql4 Helicase Is a Homeodomain-Like DNA Interaction Motif." *Nucleic Acids Res* 40, no. 17 (Sep 1 2012): 8309-24.

[32] Xu, X., C. W. Chang, M. Li, C. Liu, and Y. Liu. "Molecular Mechanisms of the Recq4 Pathogenic Mutations." *Front Mol Biosci* 8 (2021): 791194.

[33] Tangeman, L., M. A. McIlhatton, P. Grierson, J. Groden, and S. Acharya. "Regulation of Blm Nucleolar Localization." *Genes (Basel)* 7, no. 9 (Sep 21 2016).

[34] Russell, B., S. Bhattacharyya, J. Keirsey, A. Sandy, P. Grierson, E. Perchiniak, J. Kavecansky, S. Acharya, and J. Groden. "Chromosome Breakage Is Regulated by the Interaction of the Blm Helicase and Topoisomerase Iialpha." *Cancer Res* 71, no. 2 (Jan 15 2011): 561-71.

[35] Hu, P., S. F. Beresten, A. J. van Brabant, T. Z. Ye, P. P. Pandolfi, F. B. Johnson, L. Guarente, and N. A. Ellis. "Evidence for Blm and Topoisomerase Iiialpha Interaction in Genomic Stability." *Hum Mol Genet* 10, no. 12 (Jun 1 2001): 1287-98.

[36] Garkavtsev, I. V., N. Kley, I. A. Grigorian, and A. V. Gudkov. "The Bloom Syndrome Protein Interacts and Cooperates with P53 in Regulation of Transcription and Cell Growth Control." *Oncogene* 20, no. 57 (Dec 13 2001): 8276-80.

[37] Sun, L., Y. Huang, R. A. Edwards, S. Yang, A. N. Blackford, W. Niedzwiedz, and J. N. M. Glover. "Structural Insight into Blm Recognition by Topbp1." *Structure* 25, no. 10 (Oct 3 2017): 1582-88 e3.

[38] Kang, D., S. Lee, K. S. Ryu, H. K. Cheong, E. H. Kim, and C. J. Park. "Interaction of Replication Protein a with Two Acidic Peptides from Human Bloom Syndrome Protein." *FEBS Lett* 592, no. 4 (Feb 2018): 547-58.

[39] Singh, D. K., V. Popuri, T. Kulikowicz, I. Shevelev, A. K. Ghosh, M. Ramamoorthy, M. L. Rossi*, et al.* "The Human Recq Helicases Blm and Recql4 Cooperate to Preserve Genome Stability." *Nucleic Acids Res* 40, no. 14 (Aug 2012): 6632-48.

[40] Shastri, V. M., V. Subramanian, and K. H. Schmidt. "A Novel Cell-Cycle-Regulated Interaction of the Bloom Syndrome Helicase Blm with Mcm6 Controls Replication-Linked Processes." *Nucleic Acids Res* 49, no. 15 (Sep 7 2021): 8699-713.

[41] Bergeron, K. L., E. L. Murphy, L. W. Brown, and K. H. Almeida. "Critical Interaction Domains between Bloom Syndrome Protein and Rad51." *Protein J* 30, no. 1 (Jan 2011): 1-8.

[42] Srivastava, V., P. Modi, V. Tripathi, R. Mudgal, S. De, and S. Sengupta. "Blm Helicase Stimulates the Atpase and Chromatin-Remodeling Activities of Rad54." *J Cell Sci* 122, no. Pt 17 (Sep 1 2009): 3093-103.

[43] Langland, G., J. Kordich, J. Creaney, K. H. Goss, K. Lillard-Wetherell, K. Bebenek, T. A. Kunkel, and J. Groden. "The Bloom's Syndrome Protein (Blm) Interacts with Mlh1 but Is Not Required for DNA Mismatch Repair." *J Biol Chem* 276, no. 32 (Aug 10 2001): 30031-5.

[44] Beamish, H., P. Kedar, H. Kaneko, P. Chen, T. Fukao, C. Peng, S. Beresten*, et al.* "Functional Link between Blm Defective in Bloom's Syndrome and the Ataxia-Telangiectasia-Mutated Protein, Atm." *J Biol Chem* 277, no. 34 (Aug 23 2002): 30515-23.

[45] Myung, K., A. Datta, C. Chen, and R. D. Kolodner. "Sgs1, the Saccharomyces Cerevisiae Homologue of Blm and Wrn, Suppresses Genome Instability and Homeologous Recombination." *Nat Genet* 27, no. 1 (Jan 2001): 113-6.

[46] Lillard-Wetherell, K., A. Machwe, G. T. Langland, K. A. Combs, G. K. Behbehani, S. A. Schonberg, J. German*, et al.* "Association and Regulation of the Blm Helicase by the Telomere Proteins Trf1 and Trf2." *Hum Mol Genet* 13, no. 17 (Sep 1 2004): 1919-32.

[47] Zhang, R., S. Sengupta, Q. Yang, S. P. Linke, N. Yanaihara, J. Bradsher, V. Blais, C. H. McGowan, and C. C. Harris. "Blm Helicase Facilitates Mus81 Endonuclease Activity in Human Cells." *Cancer Res* 65, no. 7 (Apr 1 2005): 2526-31.

[48] Kitano, K. "Structural Mechanisms of Human Recq Helicases Wrn and Blm." *Front Genet* 5 (2014): 366.

[49] Kitano, K., S. Y. Kim, and T. Hakoshima. "Structural Basis for DNA Strand Separation by the Unconventional Winged-Helix Domain of Recq Helicase Wrn." *Structure* 18, no. 2 (Feb 10 2010): 177-87.

[50] Mukhopadhyay, S., T. Das, M. Bose, C. K. Jain, M. Chakraborty, S. Mukherjee, K. Shikha, A. K. Das, and A. Ganguly. "Residues at the Interface between Zinc Binding and Winged Helix Domains of Human Recq1 Play a Significant Role in DNA Strand Annealing Activity." *Nucleic Acids Res* 49, no. 20 (Nov 18 2021): 11834-54.

[51] Wu, L., K. L. Chan, C. Ralf, D. A. Bernstein, P. L. Garcia, V. A. Bohr, A. Vindigni*, et al.* "The Hrdc Domain of Blm Is Required for the Dissolution of Double Holliday Junctions." *EMBO J* 24, no. 14 (Jul 20 2005): 2679-87.

[52] Kitano, K., N. Yoshihara, and T. Hakoshima. "Crystal Structure of the Hrdc Domain of Human Werner Syndrome Protein, Wrn." *J Biol Chem* 282, no. 4 (Jan 26 2007): 2717-28.

[53] Chouard, T. "Structural Biology: Breaking the Protein Rules." *Nature* 471, no. 7337 (Mar 10 2011): 151-3.

[54] Romero, P., Z. Obradovic, X. Li, E. C. Garner, C. J. Brown, and A. K. Dunker. "Sequence Complexity of Disordered Protein." *Proteins* 42, no. 1 (Jan 1 2001): 38-48.

[55] Bandaru, V., W. Cooper, S. S. Wallace, and S. Doublie. "Overproduction, Crystallization and Preliminary Crystallographic Analysis of a Novel Human DNA-Repair Enzyme That Recognizes Oxidative DNA Damage." *Acta Crystallogr D Biol Crystallogr* 60, no. Pt 6 (Jun 2004): 1142-4

[56] Dosztanyi, Z., V. Csizmok, P. Tompa, and I. Simon. "The Pairwise Energy Content Estimated from Amino Acid Composition Discriminates between Folded and Intrinsically Unstructured Proteins." *J Mol Biol* 347, no. 4 (Apr 8 2005): 827-39.

[57] Dosztanyi, Z., B. Meszaros, and I. Simon. "Anchor: Web Server for Predicting Protein Binding Regions in Disordered Proteins." *Bioinformatics* 25, no. 20 (Oct 15 2009): 2745-6.

[58] Hanson, J., K. K. Paliwal, T. Litfin, and Y. Zhou. "Spot-Disorder2: Improved Protein Intrinsic Disorder Prediction by Ensembled Deep Learning." *Genomics Proteomics Bioinformatics* 17, no. 6 (Dec 2019): 645-56.

[59] Hatos, A., B. Hajdu-Soltesz, A. M. Monzon, N. Palopoli, L. Alvarez, B. Aykac-Fas, C. Bassot*, et al.* "Disprot: Intrinsic Protein Disorder Annotation in 2020." *Nucleic Acids Res* 48, no. D1 (Jan 8 2020): D269-D76.

[60] Piovesan, D., F. Tabaro, I. Micetic, M. Necci, F. Quaglia, C. J. Oldfield, M. C. Aspromonte*, et al.* "Disprot 7.0: A Major Update of the Database of Disordered Proteins." *Nucleic Acids Res* 45, no. D1 (Jan 4 2017): D219-D27.

[61] Hanson, J., K. Paliwal, T. Litfin, Y. Yang, and Y. Zhou. "Accurate Prediction of Protein Contact Maps by Coupling Residual Two-Dimensional Bidirectional Long Short-Term Memory with Convolutional Neural Networks." *Bioinformatics* 34, no. 23 (Dec 1 2018): 4039-45.

[62] J. Hanson, K. Paliwal, T. Litfin, and Y. Zhou, “Enhancing protein intrinsic disorder prediction by utilizing deep squeeze and excitation residual inception and long short-term memory networks.”, Genomics, Proteomics & Bioinformatics, 17: 645-656 (2019).

[63] V. Munoz and L. Serrano, "Elucidating the folding problem of helical peptides using empirical parameters. II. Helix macrodipole effects and rational modification of the helical content of natural peptides," *J Mol Biol* 245, no. 3 (Jan 20 1995): 275-96.

[64] Munoz, V., and L. Serrano. "Development of the Multiple Sequence Approximation within the Agadir Model of Alpha-Helix Formation: Comparison with Zimm-Bragg and Lifson-Roig Formalisms." *Biopolymers* 41, no. 5 (Apr 15 1997): 495-509.

[65] Pace, C. N., and J. M. Scholtz. "A Helix Propensity Scale Based on Experimental Studies of Peptides and Proteins." *Biophys J* 75, no. 1 (Jul 1998): 422-7.

[66] Vacic, V., P. R. Markwick, C. J. Oldfield, X. Zhao, C. Haynes, V. N. Uversky, and L. M. Iakoucheva. "Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder." *PLoS Comput Biol* 8, no. 10 (2012): e1002709.

[67] Ali, H., S. Urolagin, O. Gurarslan, and M. Vihinen. "Performance of Protein Disorder Prediction Programs on Amino Acid Substitutions." *Hum Mutat* 35, no. 7 (Jul 2014): 794-804.

[68] Niroula, A., S. Urolagin, and M. Vihinen. "Pon-P2: Prediction Method for Fast and Reliable Identification of Harmful Variants." *PLoS One* 10, no. 2 (2015): e0117380.

[69] Karplus, M., and J. A. McCammon. "Molecular Dynamics Simulations of Biomolecules." *Nat Struct Biol* 9, no. 9 (Sep 2002): 646-52.

[70] Kasahara, K., H. Terazawa, T. Takahashi, and J. Higo. "Studies on Molecular Dynamics of Intrinsically Disordered Proteins and Their Fuzzy Complexes: A Mini-Review." *Comput Struct Biotechnol J* 17 (2019): 712-20.

[71] Robustelli, P., S. Piana, and D. E. Shaw. "Developing a Molecular Dynamics Force Field for Both Folded and Disordered Protein States." *Proc Natl Acad Sci U S A* 115, no. 21 (May 22 2018): E4758-E66.

[72] Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. "Improved Side-Chain Torsion Potentials for the Amber Ff99sb Protein Force Field." *Proteins* 78, no. 8 (Jun 2010): 1950-8.

[73] Best, R. B., N. V. Buchete, and G. Hummer. "Are Current Molecular Dynamics Force Fields Too Helical?". *Biophys J* 95, no. 1 (Jul 2008): L07-9.

[74] Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling. "Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters." *Proteins* 65, no. 3 (Nov 15 2006): 712-25.

[75] Sibille, N., and P. Bernado. "Structural Characterization of Intrinsically Disordered Proteins by the Combined Use of Nmr and Saxs." *Biochem Soc Trans* 40, no. 5 (Oct 2012): 955-62.

[76] Kosol, S., S. Contreras-Martos, C. Cedeno, and P. Tompa. "Structural Characterization of Intrinsically Disordered Proteins by Nmr Spectroscopy." *Molecules* 18, no. 9 (Sep 4 2013): 10802-28.

[77] Wüthrich, Kurt. *Nmr of Proteins and Nucleic Acids.* The George Fisher Baker Non-Resident Lectureship in Chemistry at Cornell University. New York: Wiley, 1986.

[78] Howard, M. J. "Protein Nmr Spectroscopy." *Curr Biol* 8, no. 10 (May 7 1998): R331-3.

[79] Case, D. A. "Interpretation of Chemical Shifts and Coupling Constants in Macromolecules." *Curr Opin Struct Biol* 10, no. 2 (Apr 2000): 197-203.

[80] Clore, G. M., and A. M. Gronenborn. "Determining the Structures of Large Proteins and Protein Complexes by Nmr." *Trends Biotechnol* 16, no. 1 (Jan 1998): 22-34.

[81] Schwarzinger, S., G. J. Kroon, T. R. Foss, J. Chung, P. E. Wright, and H. J. Dyson. "Sequence-Dependent Correction of Random Coil Nmr Chemical Shifts." *J Am Chem Soc* 123, no. 13 (Apr 4 2001): 2970-8.

[82] Chen, T. C., C. L. Hsiao, S. J. Huang, and J. R. Huang. "The Nearest-Neighbor Effect on Random-Coil Nmr Chemical Shifts Demonstrated Using a Low-Complexity Amino-Acid Sequence." *Protein Pept Lett* 23, no. 11 (2016): 967-75.

[83] De Simone, A., A. Cavalli, S. T. Hsu, W. Vranken, and M. Vendruscolo. "Accurate Random Coil Chemical Shifts from an Analysis of Loop Regions in Native States of Proteins." *J Am Chem Soc* 131, no. 45 (Nov 18 2009): 16332-3.

[84] Mielke, S. P., and V. V. Krishnan. "Characterization of Protein Secondary Structure from Nmr Chemical Shifts." *Prog Nucl Magn Reson Spectrosc* 54, no. 3-4 (Apr 5 2009): 141-65.

[85] Vila, J. A., Y. A. Arnautova, and H. A. Scheraga. "Use of 13c(Alpha) Chemical Shifts for Accurate Determination of Beta-Sheet Structures in Solution." *Proc Natl Acad Sci U S A* 105, no. 6 (Feb 12 2008): 1891-6.

[86] Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci U S A. 2007 Jun 5;104(23):9615-20.

[87] Ketkar, A., M. Voehler, T. Mukiza, and R. L. Eoff. "Residues in the Recq C-Terminal Domain of the Human Werner Syndrome Helicase Are Involved in Unwinding G-Quadruplex DNA." *J Biol Chem* 292, no. 8 (Feb 24 2017): 3154-63.

[88] Bax, A., and M. Ikura. "An Efficient 3d Nmr Technique for Correlating the Proton and 15n Backbone Amide Resonances with the Alpha-Carbon of the Preceding Residue in Uniformly 15n/13c Enriched Proteins." *J Biomol NMR* 1, no. 1 (May 1991): 99-104.

[89] Kennedy, J. A., G. W. Daughdrill, and K. H. Schmidt. "A Transient Alpha-Helical Molecular Recognition Element in the Disordered N-Terminus of the Sgs1 Helicase Is Critical for Chromosome Stability and Binding of Top3/Rmi1." *Nucleic Acids Res* 41, no. 22 (Dec 2013): 10215-27.

[90] Bramham, J. E., A. Podmore, S. A. Davies, and A. P. Golovanov. "Comprehensive Assessment of Protein and Excipient Stability in Biopharmaceutical Formulations Using (1)H Nmr Spectroscopy." *ACS Pharmacol Transl Sci* 4, no. 1 (Feb 12 2021): 288-95.

[91] Heikkinen, H. A., S. M. Backlund, and H. Iwai. "Nmr Structure Determinations of Small Proteins Using Only One Fractionally 20% (13)C- and Uniformly 100% (15)N-Labeled Sample." *Molecules* 26, no. 3 (Feb 1 2021).

[92] Metskas, L. A., and E. Rhoades. "Single-Molecule Fret of Intrinsically Disordered Proteins." *Annu Rev Phys Chem* 71 (Apr 20 2020): 391-414.

[93] Ray, K., R. Badugu, and J. R. Lakowicz. "Distance-Dependent Metal-Enhanced Fluorescence from Langmuir-Blodgett Monolayers of Alkyl-Nbd Derivatives on Silver Island Films." [In English]. *Langmuir* 22, no. 20 (Sep 26 2006): 8374-78.

[94] Choi, U. B., J. J. McCann, K. R. Weninger, and M. E. Bowen. "Beyond the Random Coil: Stochastic Conformational Switching in Intrinsically Disordered Proteins." *Structure* 19, no. 4 (Apr 13 2011): 566-76.

[95] Monahan, Z., V. H. Ryan, A. M. Janke, K. A. Burke, S. N. Rhoads, G. H. Zerze, R. O'Meally*, et al.* "Phosphorylation of the Fus Low-Complexity Domain Disrupts Phase Separation, Aggregation, and Toxicity." *EMBO J* 36, no. 20 (Oct 16 2017): 2951-67.

[96] Traynelis, S. F., L. P. Wollmuth, C. J. McBain, F. S. Menniti, K. M. Vance, K. K. Ogden, K. B. Hansen*, et al.* "Glutamate Receptor Ion Channels: Structure, Regulation, and Function." *Pharmacol Rev* 62, no. 3 (Sep 2010): 405-96.

[97] Choi, U. B., S. Xiao, L. P. Wollmuth, and M. E. Bowen. "Effect of Src Kinase Phosphorylation on Disordered C-Terminal Domain of N-Methyl-D-Aspartic Acid (Nmda) Receptor Subunit Glun2b Protein." *J Biol Chem* 286, no. 34 (Aug 26 2011): 29904-12.

[98] Choi, U. B., R. Kazi, N. Stenzoski, L. P. Wollmuth, V. N. Uversky, and M. E. Bowen. "Modulating the Intrinsic Disorder in the Cytoplasmic Domain Alters the Biological Activity of the N-Methyl-D-Aspartate-Sensitive Glutamate Receptor." *J Biol Chem* 288, no. 31 (Aug 2 2013): 22506-15.

[99] Kim, Y. G., S. O. Ho, N. R. Gassman, Y. Korlann, E. V. Landorf, F. R. Collart, and S. Weiss. "Efficient Site-Specific Labeling of Proteins Via Cysteines." [In English]. *Bioconjugate Chemistry* 19, no. 3 (Mar 2008): 786-91

[100] LeBlanc, S. J., P. Kulkarni, and K. R. Weninger. "Single Molecule Fret: A Powerful Tool to Study Intrinsically Disordered Proteins." *Biomolecules* 8, no. 4 (Nov 8 2018).

[101] Naudi-Fabra, S., M. Tengo, M. R. Jensen, M. Blackledge, and S. Milles. "Quantitative Description of Intrinsically Disordered Proteins Using Single-Molecule Fret, Nmr, and Saxs." *J Am Chem Soc* 143, no. 48 (Dec 8 2021): 20109-21.

[102] Greenfield, N. J. "Using Circular Dichroism Spectra to Estimate Protein Secondary Structure." *Nat Protoc* 1, no. 6 (2006): 2876-90.

[103] Sreerama, N., and R. W. Woody. "Computation and Analysis of Protein Circular Dichroism Spectra." *Methods Enzymol* 383 (2004): 318-51.

[104] Holzwarth, G., and P. Doty. "The Ultraviolet Circular Dichroism of Polypeptides." *J Am Chem Soc* 87 (Jan 20 1965): 218-28.

[105] Greenfield, N., and G. D. Fasman. "Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation." *Biochemistry* 8, no. 10 (Oct 1969): 4108-16.

[106] Kelly, S. M., T. J. Jess, and N. C. Price. "How to Study Proteins by Circular Dichroism." *Biochim Biophys Acta* 1751, no. 2 (Aug 10 2005): 119-39.

[107] Venyaminov, SYu, I. A. Baikalov, Z. M. Shen, C. S. Wu, and J. T. Yang. "Circular Dichroic Analysis of Denatured Proteins: Inclusion of Denatured Proteins in the Reference Set." *Anal Biochem* 214, no. 1 (Oct 1993): 17-24.

[108] Chemes, L. B., L. G. Alonso, M. G. Noval, and G. de Prat-Gay. "Circular Dichroism Techniques for the Analysis of Intrinsically Disordered Proteins and Domains." *Methods Mol Biol* 895 (2012): 387-404.

[109] Ezerski, J. C., P. Zhang, N. C. Jennings, M. N. Waxham, and M. S. Cheung. "Molecular Dynamics Ensemble Refinement of Intrinsically Disordered Peptides According to Deconvoluted Spectra from Circular Dichroism." *Biophys J* 118, no. 7 (Apr 7 2020): 1665-78.

[110] Kennedy, J. A., S. Syed, and K. H. Schmidt. "Structural Motifs Critical for in Vivo Function and Stability of the Recq-Mediated Genome Instability Protein Rmi1." *PLoS One* 10, no. 12 (2015): e0145466.

[111] Keller, H., K. Kiosze, J. Sachsenweger, S. Haumann, O. Ohlenschlager, T. Nuutinen, J. E. Syvaoja*, et al.* "The Intrinsically Disordered Amino-Terminal Region of Human Recql4: Multiple DNA-Binding Domains Confer Annealing, Strand Exchange and G4 DNA Binding." *Nucleic Acids Res* 42, no. 20 (Nov 10 2014): 12614-27.

[112] Kim, S. Y., T. Hakoshima, and K. Kitano. "Structure of the Recq C-Terminal Domain of Human Bloom Syndrome Protein." *Sci Rep* 3 (Nov 21 2013): 3294.

[113] Kim YM, Choi BS. Structure and function of the regulatory HRDC domain from human Bloom syndrome protein. Nucleic Acids Res. 2010 Nov;38(21):7764-77.

[114] Doerr A. Tracking protein conformation in live cells. Nat Methods. 2021 Dec;18(12):1451.