STRUCTURAL AND INTRINSIC DISORDER IN THE REGULATION OF PROTEIN-PROTEIN INTERACTIONS

A THESIS SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA BY

Michael E. Fealey

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Dr. David D. Thomas

June, 2019

Acknowledgements

First, thanks are owed Dr. David D. Thomas. Dave has been an excellent mentor these past several years. He's allowed me to grow and develop as a scientist, to explore areas of research outside of muscle biophysics (including my favorite area – intrinsically disordered proteins), and to have the freedom to develop my skills as a teacher (knowing full well the time-commitment required). Ultimately, Dave's flexibility is what helped me acquire the skills I was hoping to obtain in graduate training and for that, I am deeply appreciative.

Second, I'd like to thank the other members of my PhD committee. Dr. Wendy Gordon, Dr. Paolo Provenzano, Dr. Doug Ohlendorf and Dr. Jeongsik Yong, thank you all for taking the time to review my preliminary exam proposal and provide feedback. It's part of what helped me turn that research proposal into one of the chapters of this thesis. And thank you for continuing to serve as advisors throughout the course of my training and into my thesis defense.

Third, thanks are owed to all of lab PIs who I was lucky to have as collaborators on the research projects that make up this thesis including: Dr. Thomas S. Hays, Dr. Adam W. Avery, Dr. Edward H. Egelman, Dr. Anne Hinderliter, and Dr. Alessandro Cembran. Adam, I am especially grateful for your invitation to work on studies of β-III-spectrin. Your enthusiasm, endurance, and commitment was most inspiring.

Fourth, I'd like to thank all of my past students at Concordia University. I am so grateful for having had the opportunity to be a part of your undergraduate education. Your positivity, curiosity, willingness to rise to challenges, and acceptance of me as an instructor were a huge help getting through graduate school. I always felt uplifted after

class and being a part of your learning really helped get me through the tougher periods of my graduate training experience. So thank you. I wish you all the best. I'd also like to thank the faculty members in the Science Department for letting me teach there in the first place! Dr. Leanne Bakke, Dr. Mandy Brosnahan, Dr. Taylor Mach, Dr. Benjamin Harrison, Dr. Mary Ann Yang, and Dr. Carolyn Wanamaker, you all create a wonderful work and learning environment in the Science Department. I will very much miss getting to work in it. Dr. Bakke, thank you for giving me the opportunity to teach A&P. I never thought that would ever be an option for me when I first applied for the adjunct position, but I was so happy you let me go for it. Dr. B and Dr. Mach, thank you both for supporting my in-class salivary cortisol research with students. It was a lot of fun working on that and it gave me a new appreciation for research as an instructional tool.

Last but certainly not least, I'd like to thank all the members of Dave's lab for their help along the way. I doubt there's a single person in the lab who hasn't helped me in some way or another, but I'm particularly indebted to following people: Dr. Andrew Thompson, for training in pulsed EPR; Mr. Jonathan Crain, for showing me the ropes on everything when I first joined DDT lab; Dr. Benjamin Binder, for helping me with many teaching endeavors and also the simulation work that makes up a segment of Chapter 4; Dr. Robyn Rebbeck, for countless discussions about science, teaching and training philosophy, and just general support. I would not have gotten far without any of you.

Thesis Abstract

This thesis applied spectroscopy and molecular dynamics simulation to study the structural biology of actin-binding domains (ABDs) from the spectrin superfamily of proteins as well as an intrinsically disordered region (IDR) of an integral membrane protein called synaptotagmin 1. In the former case, the structural hypothesis being tested was that actin-binding domains exist in distinct conformational states that are either permissive to or inhibitory towards binding of actin filaments. This question was probed using pulsed-EPR, which measured distances between the calponin homology (CH) domains that make up the ABD as proxy for conformation in the presence or absence of actin or with and without disease-causing mutation. The initial hypothesis of a closed compact state being unable to bind actin and an open extended state being binding-competent was largely supported by the data. However, the hypothesis was ultimately refined to conclude that an "open" state is likely to still be a fairly collapsed structure that is dynamically disordered. With this model, future efforts will be able use the model to look for small molecules that perturb the conformational equilibrium of ABDs harboring disease-causing mutations in potentially therapeutically efficacious ways. Moreover, the model can be tested in other ABDs of the protein superfamily to assess similarities and differences in mechanism.

In the case of the intrinsically disordered region of synaptotagmin 1, it was hypothesized that a post-translational modification, specifically phosphorylation of a threonine residue, caused a structural change in the IDR that then results in a change in neurotransmitter release. This hypothesis was also tested with spectroscopic methods, mainly FRET and circular dichroism, but also with molecular dynamics. It was found that

mimicking the low dielectric environment of the membrane with co-solvents in solution and artificially in silico caused the synaptotagmin 1 IDR to fold into helical structure. The post-translational modification, however, was found to interfere with the formation of helical structure, providing a still incomplete but novel molecular explanation for the effect it has on potentiation of neurotransmitter release observed in vivo. At the very least, the structural model provides a working hypothesis that can be further explored in further work.

Table of Contents

List of Tables

List of Figures

terminus, in addition to transiently interacting with the acidic C-terminus of the IDR sequence, also interacts with acidic lipids in a synaptic vesicle membrane and likely has two competing interactions (lipid – intermolecular; acidic C-terminus – intramolecular) that could influence propensity to exist in compact structural state.

List of Abbreviations

ATP – adenosine triphosphate
ABD – actin-binding domain
BMD – Becker muscular dystrophy
CH – calponin homology
DEER – double electron-electron resonance
DMD – Duchenne muscular dystrophy
EM – electromagnetic or electron microscopy
EPR – electron paramagnetic resonance
FRET – Förster resonance energy transfer
IDPs – intrinsically disordered proteins
IDRs – intrinsically disordered regions
NMR – nuclear magnetic resonance
UV – ultraviolet

CHAPTER 1: PROTEIN STRUCTURAL DYNAMICS AND METHODS TO STUDY THEM

1.1 - Protein folding, structure and conformational change in relation to function

1.1.1 – Traditional structure-function relationship of proteins

Proteins are a class of biomolecules that generally serve as major workhorses for molecular processes that underlie cellular function. This includes processes such as the catalysis of metabolism, ion transport of second messenger signaling and alteration of membrane potential, maintenance of cell shape and integrity with cytoskeletal protein filaments, and regulation of gene expression. Typically, proteins are able to carry out these distinct functions by virtue of having unique three-dimensional folded structures that are dictated by their primary amino acid sequence. A newly synthesized chain of amino acids folds into secondary structural elements that generally include α-helices, β-sheets, and turns, which then form additional intramolecular contacts to stabilize a more globular, tertiary structure. In some cases, a protein in its tertiary structure combines with others to form a larger protein molecule that is then referred to as being in its quaternary structure. This then represents the final form of the molecule that consists of multiple subunits. Ultimately, these final unique three-dimensional structures are then able interact with specific ligands, ions, or macromolecule binding partners such as other proteins, nucleic acids, or lipids as a way of carrying out their functions.

The three-dimensional structures of proteins are, however, not static. Proteins can exist in multiple conformational states. In many instances, distinct conformational subsets of a protein's overall ensemble are what actually enable execution of functional activity and the transitioning between these conformations is what imparts regulation of function.

As a classic example, the ubiquitous class of motor proteins known as myosins generate the mechanical force of many molecular movements carried out in cells, but most famously movement in muscle cell contraction. Myosins undergo large-scale changes in their conformation that are tightly coupled to the binding, hydrolysis and subsequent release of an ATP molecule, ultimately allowing them to convert the chemical potential energy of a phosphodiester bond into mechanical energy that enables ratchetting along the length of actin filaments. [1-3] Another classic example where conformation of a protein regulates function is hemoglobin. Hemoglobin has a quaternary structure consisting of four subunits, each of which contains an iron-containing heme prosthetic group for binding molecular oxygen. Hemoglobin protein structure is designed to promote efficient loading and unloading of O_2 as the red blood cell shuttling it passes from alveolar to tissue capillaries. This loading and unloading is mediated by binding cooperativity and allostery that are linked to each subunit's conformation.[4] When a molecule of O_2 binds to one subunit, it causes small structural distortions to that subunit's secondary and tertiary structure. Those distortions propagate to the adjacent subunits, altering their conformations in such a way that their subsequent binding of molecular O_2 becomes easier. These O_2 binding-induced changes that increase regional affinity for the ligand in adjacent subunits continue until hemoglobin is saturated at all four sites. In both examples, myosin and hemoglobin, changes in protein conformation ultimately change the way in which each molecule can bind their chemical substrate, which in turn enables execution of their molecular-level function.

1.1.2 – Intrinsically disordered proteins and protein regions

While the above examples illustrate roles of conformational change in regulating protein function, it is important to note that not all protein molecules exist in a stable and well-folded three-dimensional structure under native conditions. About 30-40% of eukaryotic genomes encode for proteins that are categorized as being intrinsically disordered or partially intrinsically disordered[5], meaning their entire amino acid sequence or significant parts of their sequence lack stable tertiary and/or secondary structure most of their cellular lifetime and instead they more frequently occupy random coil-like conformations. When intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) were first recognized as a separate entity within this class of bio-molecule[6, 7], it brought up several important questions. How do IDPs interact with their target macromolecule if they don't have a well-folded surface that can form residue contacts at a molecule-molecule interface? Can something that mostly lacks three-dimensional structure participate in important protein behaviors like allostery if there isn't a clear occupation of discrete conformations in their ensemble? In the roughly two decades development of the IDP field, generalized answers to the above questions have begun to emerge. Consequently, our view of the protein structure-function relationship has expanded. What constitutes a functional protein "fold" is now viewed as a broader continuum ranging from highly stable ordered structure to complete disorder.[8, 9]

A large determinant of why one protein molecule adopts a stable, folded state and another does not is the amino acid composition of the primary sequence. Typically, proteins that are intrinsically disordered have a notably low frequency of bulky, hydrophobic or aromatic amino acids and an enrichment in polar, charged or structure-

breaking amino acids.[10] A large driving force for protein folding is the burial of bulky, hydrophobic residues, where burial of residue side chains shields hydrophobic side chains from the surrounding polar solvent. Amino acid sequences of IDPs do not typically have this driving force because of their bias towards charged and polar residues. This does not mean, however, that disordered proteins never fold. Quite the contrary, IDPs will often undergo folding as part of the biological function they carry out in the cell.[8] For example, IDPs often serve as hub proteins where they use short linear motifs in their primary sequence to bind multiple different folded proteins and, in doing so, help to organize the folded proteins into one central location.[11, 12] The interaction of these short linear motifs with their target protein partners can often associated with regional folding into short helices or sheets. In this regard, the unfolded-to-folded transition frequently seen in IDPs is similar to the examples described above for well-folded proteins in that it is just another type of conformational transition that regulates function.

In this thesis, research questions on the two ends of the protein structure-function continuum are examined. On one end, the very well-folded but conformationally flexible actin-binding domains from two representative members of the spectrin superfamily are probed to understand how conformation regulates their binding to filamentous actin (Chapters 2 and 3). On the other end, an intrinsically disordered linker region in the integral membrane protein synaptotagmin is examined, asking a more fundamental question about molecular determinants of secondary structure formation (Chapters 4). The relevance of these distinct avenues of research are further introduced in two subsequent sections below, but first methods for measuring folding and conformational change will be discussed.

1.2 – Spectroscopy as a measure of protein structure and conformation

1.2.1 – Fluorescence and resonance energy transfer

The conformational change of a protein is often studied through use of spectroscopy, a discipline primarily concerned with how different types of electromagnetic (EM) radiation interacts with matter. In the case of this thesis, we're specifically concerned about the interaction of EM radiation (particularly ultraviolet and microwave regions of the spectrum) with protein molecules or probes that we attach to the protein molecules. Spectroscopy can be broadly categorized as being absorption-based or emission-based. One of the main absorption-based spectroscopies used in Chapter 4 is circular dichroism. While perhaps simpler compared to other spectroscopies, circular dichroism is quite useful for studying folding reactions of IDPs and is always a good place to start study of an IDP before moving on to more laborious methods.

EM radiation used in spectroscopy consists of waves that have electric (E) and magnetic (B) field components that oscillate in magnitude perpendicularly to each other and the direction of the wave's propagation.

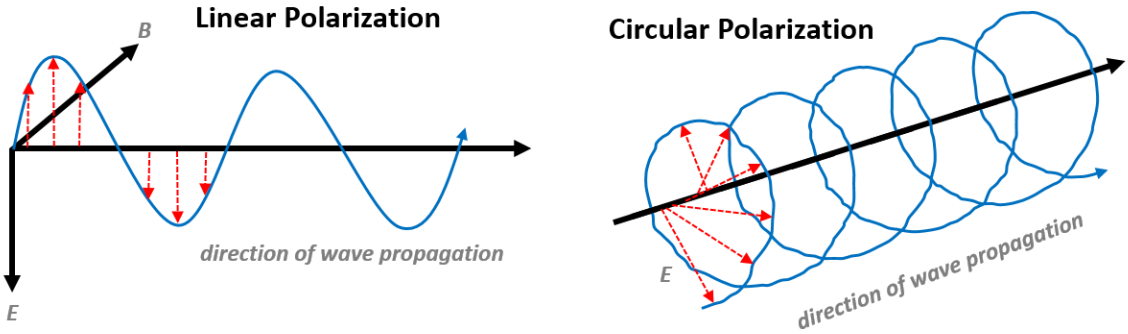

Figure 1.1. Different polarizations of light used in this thesis.

Most EM radiation sources will emit waves whose oscillations occur in any direction about the axis of propagation and these waves are said to be unpolarized. When waves have their electric field component constrained to a single plane, they are said to be linearly polarized (Figure 1.1, left). In the circular dichroism spectrometer used in Chapter 4, waves have a constant electric field vector that rotates around the axis of wave propagation. These light waves are circularly polarized (Figure 1.1, right). The directionality of circular polarization can be either clockwise or counterclockwise.

Ultimately, the relevance of light polarization in examining protein structure is that protein molecules differentially absorb right and left circularly polarized light. This stems from several factors including the chirality of their amino acids, the folded secondary structure of those amino acids and the wavelength of the incident light. Both right and left forms are directed at solutions containing the protein of interest. In the absence of any absorption of either form of circularly polarized light, the waves sum to a linearly polarized wave because each electric field component is unmodified. However, when greater absorption of one polarization occurs relative to the other, summation of right and left results in elliptically polarized light because of the difference in electric field component magnitudes resulting from differential absorption. This leads to the degree of ellipticity, θ, that is measured in the experiment as a function of wavelength:

$$\tan(\theta) = (E_l + E_r)/(E_l - E_r) \qquad \text{Eq. 1.1}$$

where E_l and E_r represent the magnitudes of the electric field vectors in left and right circularly polarized light, respectively. Measurement of θ as a function of wavelength spanning from approximately 190-260 nm results in characteristic absorption profiles depending on the type of secondary structure present in the protein molecule (Figure 1.2).

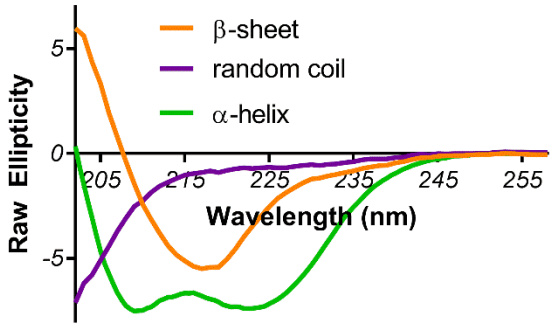

Figure 1.2. Basis absorption spectra for random coil (purple), β-sheet (orange) and α-helical protein secondary structures, where raw ellipticity is in degrees.

Different types of secondary structure result in unique absorption profiles; this forms the basis for quantifying the relative amounts of secondary structure in regular protein samples. If a protein has a mixture of secondary structures, the absorption profile is just a linear combination of the individual components. Using basis data sets of purely α-helical, purely β-sheet and purely random coil as references, a non-linear least squares fit of experimental data can be performed using coefficients for each form of secondary structure as the floating parameters:

$$\theta_\lambda = \sum f_\alpha * S_{\lambda\alpha} + f_\beta * S_{\lambda\beta} + f_{rc} * S_{\lambda rc} \qquad \text{Eq. 1.2}$$

where θ_λ is ellipticity at a given wavelength in the spectrum, $S_{\lambda\alpha}$, $S_{\lambda\beta}$, $S_{\lambda rc}$ are the α-helix, β-sheet, and random coil basis spectra at that given wavelength, and f_α, f_β, f_{rc} are the weighting factors for the basis spectra. Fitting data sets in this manner allows for the relative percentages of structure to be estimated in a sample.[13]

Ultimately, this form of spectroscopy was used to study folding of an intrinsically disordered protein region in Chapter 4. Because of the absorption profile shifts with changes in protein secondary structure, addition of co-solvents or other chemical reagents that perturb the folding equilibrium of an IDP allows for this form of spectroscopy to

extract equilibrium thermodynamics parameters. These are useful for learning about global protein behavior and energetic barriers to folding and further details on how free energies are obtained from circular dichroism data is detailed in the supporting materials of Chapter 4.

Aside from circular dichroism, another major type of spectroscopy used in this work is fluorescence, a form of emission spectroscopy. In fluorescence, EM radiation is directed at a protein molecule to excite an electronic transition that will soon after result in emission of a detectable photon of longer wavelength (Figure 1.3, blue and green transitions).

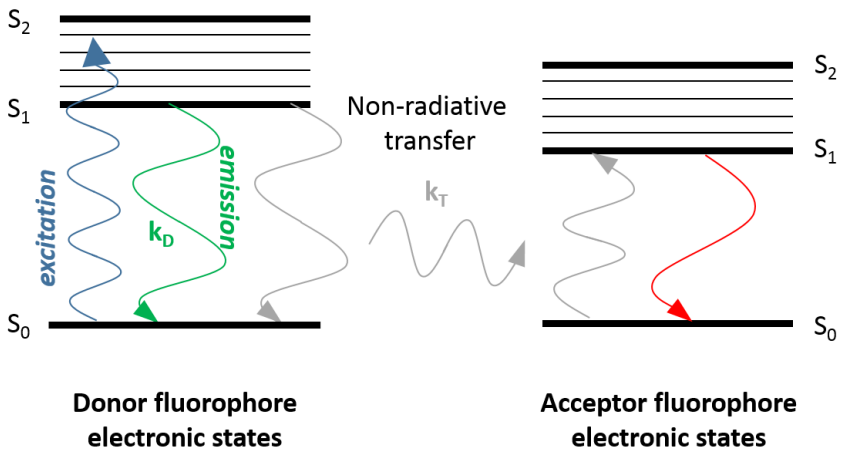

Figure 1.3. Jablonski diagram of electronic states important for fluorescence and resonance energy transfer in distance measurements.

The type of radiation used for fluorescence could be UV in origin if, for example, the intent is to excite natural fluorophores within the protein like tryptophan. Alternatively, it could come from the visible part of the electromagnetic spectrum if excitation of an extrinsic fluorophore is intended. In either case, the applied radiation excites an electron from its ground state within the fluorophore to a higher energy electronic state.

Occupancy of this higher energy level is short lived, due to loss of energy through vibrational relaxation which eventually brings the electron down to its lowest energy excited state. From here, the electron relaxes back to ground state and emits a photon in the process (Figure 1.3). The length of time the electron spends in an excited state is referred to as its lifetime, τ, which is typically only a few nanoseconds.

Fluorescence spectroscopy can ultimately be used for studying protein conformation because it is capable of making distance measurements through the process of resonance energy transfer. Fluorescence resonance energy transfer or Förster resonance energy transfer (FRET) is a form of spectroscopy that uses two fluorophores to make molecular measurements on the order of a few nanometers, a scale suitable for protein molecules. In FRET, these two fluorophores can be attached to the same molecule if the goal is to make intramolecular distance measurements, or they can be attached on two separate molecules to measure intermolecular distance. In either case, one fluorophore acts as a donor, the fluorescent probe being excited with some wavelength of light and then either emitting a photon or transferring energy non-radiatively to a second chromophore that acts as an acceptor. An acceptor is a molecule whose spectroscopic properties allow it to undergo electronic excitation to a higher energy level by non-radiatively receiving transferred energy from the first fluorophore's excited state (Figure 1.3, gray transitions). This non-radiative transfer of energy has a strong dependence on distance between the two fluorophores, as defined by the Förster equation:

$$k_T/k_D = (R/R_0)^{-6} \qquad \text{Eq. 1.3}$$

Here k_T is the rate of non-radiative energy transfer from donor fluorophore to acceptor chromophore, k_D is the natural rate of relaxation of the donor excited state to ground

state, R is the interprobe distance, and R_0 is the Förster distance or the distance at which the two fluorophores in question have an energy transfer efficiency of 50%. The rate of donor relaxation when in the presence of an acceptor (k_{DA}) can be represented as the sum of the intrinsic donor relaxation rate and the rate of non-radiative energy transfer as both processes contribute to depletion of the excited state:

$$k_{DA} = k_D + k_T \qquad \text{Eq. 1.4}$$

Both the donor relaxation rate without (k_D) and with (k_{DA}) acceptor are inversely proportional to their respective excited-state lifetimes, which are readily measurable experimental parameters:

$$k_D = 1/\tau_D \qquad \text{Eq. 1.5}$$

$$k_{DA} = 1/\tau_{DA} \qquad \text{Eq. 1.6}$$

If we substitute a rearranged Eq. 1.4 into Eq. 1.3 and move the exponent,

$$(k_{DA} - k_D/k_D)^{-1/6} = (R/R_0) \qquad \text{Eq. 1.7}$$

Based on Eq. 1.5 and Eq. 1.6, Eq. 1.7 becomes an expression for interprobe distance as a function of experimentally measurable donor lifetimes:

$$[(1/\tau_{DA} - 1/\tau_D)/(1/\tau_D)]^{-1/6} = (R/R_0) \qquad \text{Eq. 1.8}$$

Factoring out $1/\tau D$ and simplifying results in a more accessible expression:

$$[(\tau_D/\tau_{DA}) - 1]^{-1/6} = (R/R_0) \qquad \text{Eq. 1.9}$$

Eq. 1.9 shows that a simple measurement of donor lifetime in the presence and absence of acceptor should enable determination of R, the interprobe distance. The only other factor in Eq. 1.9 that must be determined is the Förster distance R_0, which relies on several factors, as shown below:

$$R_0 = (9.78 \times 10^3)*(\kappa^2 n^{-4} \phi_D J(\lambda))^{1/6} \qquad \text{Eq. 1.10}$$

Here κ is the relative orientation of fluorophore transition dipoles, n is the solution's refractive index, ϕ_D is the quantum yield of the donor fluorophore, and J(λ) is the overlap integral, which is mathematically shown below and conceptually illustrated in Figure 1.4:

$$J(\lambda) = \int F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda \,/\, \int F_D(\lambda)d\lambda \qquad \text{Eq. 1.11}$$

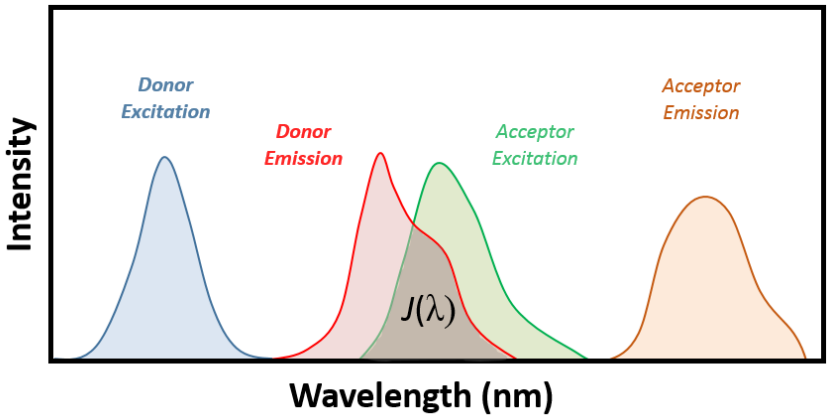

Figure 1.4. Use of excitation and emission spectra of donor and acceptor fluorophores in the design of FRET experiments. Note regional overlap of donor emission and acceptor excitation spectrums.

When all factors contributing to R_0 are known, measurements of fluorescence lifetimes with and without an acceptor probe can be made for the purpose of inferring conformational change in a protein. If we consider an example of intramolecular FRET (Figure 1.5), where the protein of interest has two folded domains and a donor probe in one of those domains and an acceptor in the other, we can predict the expected outcome based on a structural hypothesis. If the example protein binds actin and, upon doing so, its domains become further separated from one another compared to the unbound state, then the interprobe distance will increase, resulting in both a decrease in energy transfer efficiency and an increase in the apparent donor lifetime (Figure 1.5). From these lifetime

measurements, a distance change can be determined using Eq. 1.9, and conformation change can subsequently be inferred.

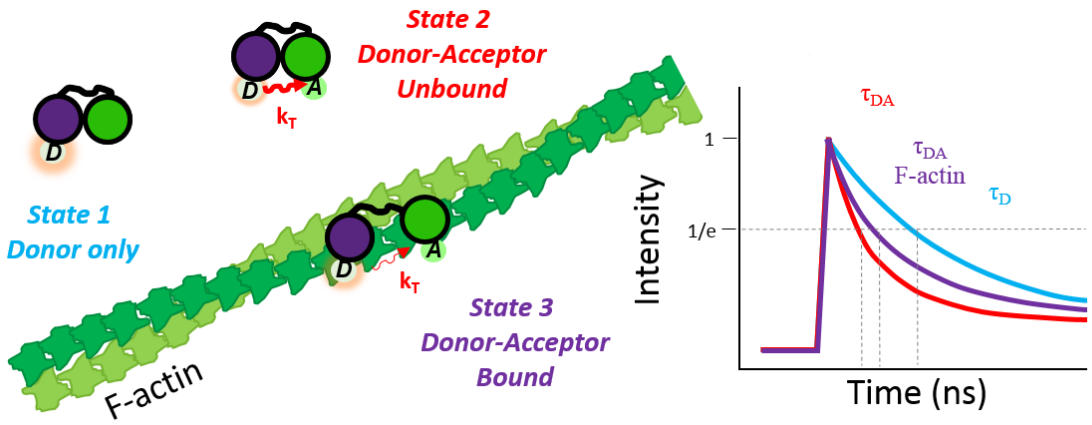

Figure 1.5. Use of lifetime-based FRET for measuring conformational change in a protein upon binding its target.

1.2.2 – Electron paramagnetic resonance

While both circular dichroism and FRET are using UV light (at least, in the chapter where these methods are relevant, Chapter 4), radiation from other regions of the electromagnetic spectrum can be used to assess protein conformation. The other major form of spectroscopy used in this thesis (Chapters 2 and 3) is a form of electron paramagnetic resonance (EPR). In this form of spectroscopy, rather than using high-energy UV waves for exciting an electronic transition, low-energy microwaves are used to excite a spin transition of a free radical.

Since most proteins do not have an unpaired valence electron that would make them EPR active, we instead use an attachable probe that contains a stable free radical. These probes are referred to as "spin labels" and are covalently linked to a specific region of interest in the protein under study. Spin labels generally consist of a heterocyclic ring that contains a nitroxide bond (Figure 1.6). On the carbon atoms adjacent to the nitroxide

nitrogen are methyl groups that flank and help stabilize the free radical when in solution. Opposite the nitroxide is some type of reactive functional group such as maleimide (as used in MSL), iodoacetamide (as used in IASL) or methanesulfonothioate (as used in MTSSL) that serves the role of chemically attaching the probe to the protein's sulfhydryl side chains found in cysteine residues. Once a spin label is attached to the protein, EPR spectroscopy can be performed.

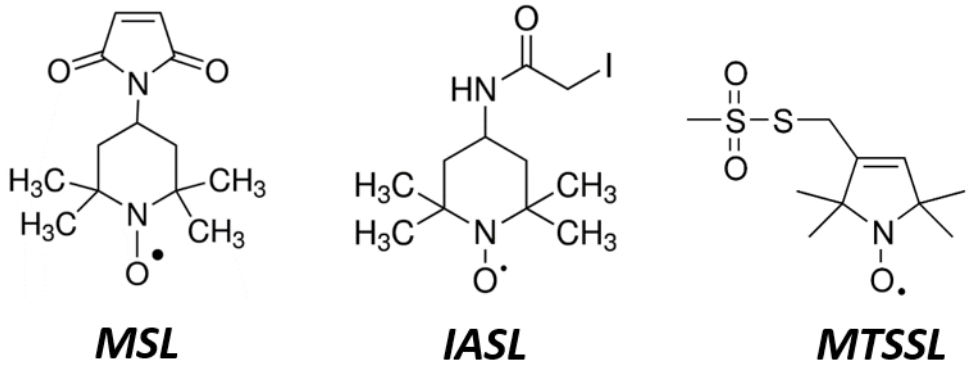

Figure 1.6. Commonly used nitroxide spin labels. The two left-most labels, MSL and IASL, are spin labels that become irreversibly covalently attached to the target protein. The right-most label, MTSSL, forms a disulfide linkage that can be reversed with reducing agent. MSL is the label used in this thesis.

In EPR, the end goal is to excite a spin transition in the nitroxide radical. These types of transitions have lower energy differences between states and thus require lower frequency radiation for excitation. This same reasoning dictates use of even lower frequency radiation in nuclear magnetic resonance (NMR), where the energy gap between nuclear spin states is smaller still compared to electron spin states. In either form of magnetic resonance, however, before you can excite the spin transition the sample must be subjected to an external magnetic field (B_0). In the absence of this applied external magnetic field, the spin quantum states available for the nitroxide's electron (m_s

= +½, m_s = -½) are degenerate, meaning there is no energy gap for excitation. However, in the presence of the external magnetic field, the Zeeman Effect causes splitting of spin state energy levels so that one becomes higher in energy. This occurs because one of the spin aligns parallel with B_0 (resulting in lower energy) whereas the other aligns antiparallel (resulting in higher energy) (Figure 1.7). With this newly generated energy gap, application of microwaves can then excite the spin transition if the conditions for resonance are met:

$$\Delta E = h\nu = g_e \mu_B B_0 \qquad \text{Eq. 1.12}$$

ΔE is the energy difference between spin states, h is Planck's constant (6.626×10^{-34} m^2kg/s), ν is the applied microwave frequency, g_e is the g-factor (2.0023 for a free electron), μ_B is the Bohr magneton (9.274×10^{-24} J/T) and B_0 is the external magnetic field strength. Absorption of incident microwaves that are being continuously applied in the EPR experiment (frequency of microwaves is usually ~9.5 GHz) only occurs at specific magnetic field strengths that generate energy gaps that exactly match that of the applied radiation frequency.

One of the consequences of using a nitroxide as the source of a stable electron, however, is that the nuclear spin states of the ^{14}N nucleus in the nitroxide bond influence the local magnetic field of the electron. This influence of the nitrogen's nuclear spin on the electron magnetic field is referred as the hyperfine interaction and ultimately results in further splitting of the electron spin energy levels generating the characteristic three resonances typically seen in nitroxide spin label's field swept spectrum (Figure 1.7).

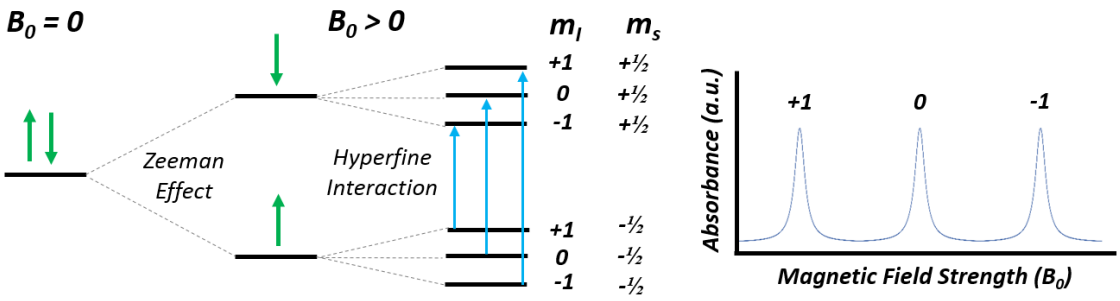

Figure 1.7. Spin transitions of the nitroxide radical. Degeneracy of spin states is eliminated with application of an external magnetic field. The two spin states become further split by the local field of the nitrogen nucleus whose spin is I=1, resulting in possible m_I spin states of -1, 0, and +1 which interact with the field experienced by the electron that causes hyperfine splitting.

While the resulting spectra described above can provide information on protein structure and dynamics[14], the work in this thesis was focused more on use of a pulsed EPR experiment called double electron-electron resonance (DEER), rather than the continuous wave EPR experiment. In DEER experiments, the goal is to measure interprobe distance just like FRET in fluorescence spectroscopy. While making the same kind of distance measurement and inferring conformational change in a protein from it, DEER has a few advantages over FRET that make it an attractive spectroscopy to utilize. (1) DEER only requires use of a single label type, which makes labeling a protein, particularly for intramolecular distance measurements, considerably more straightforward. (2) A spin label is generally much smaller than a fluorescent probe, which usually makes it less perturbing to the protein's structure and function. (3) DEER is more sensitive to conformational heterogeneity in the sample and can quantify degree of static disorder in the sample. And lastly, DEER can be done on more complex samples. DEER does have two principal disadvantages relative to FRET: (1) It is less sensitive, requiring much higher concentrations (micromolar vs nanomolar) and

acquisition time (hours vs seconds). (2) It's distance dependence is less steep (R^{-3} vs R^{-6}), making background correction a more significant problem.

The way DEER measures an interprobe distance ultimately stems from a dipolar interaction between two spins that are in close proximity (R<10nm, as in FRET). The interaction is generated with a series of microwave pulses applied at different frequencies. The main pulse sequence used in the DEER experiment is a spin echo sequence.[15] The spin echo pulse sequence first applies a 90° pulse to the sample to shift the net magnetization vector into the transverse plane. Since the sample is once again in an external magnetic field so that the Zeeman Effect can provide the necessary splitting for excitation of the spin transition, the net magnetization vector is initially going to be aligned with the external field. After the 90° pulse places the net magnetization vector in the transverse plane, different spins with different chemical environments begin to precess at different frequencies, and coherence is lost. This represents transverse, T_2, relaxation. After some period of time, τ, has passed, a second pulse is applied at the same frequency though this time it's a 180° pulse that inverts magnetization vectors. After the 180° flip, each spin's precession results in a building up and subsequent loss of magnetization along the axis of detection in the transverse plane, which constitutes the echo (Figure 1.8, top). This spin echo pulse sequence is only applied to a subpopulation of spins in the sample and the frequency at which it is applied is referred to as the observe frequency, ν_{obs}. However, as this observe frequency pulse sequence is being applied, a second frequency referred to as the pump frequency, ν_{pump}, is being used for application of additional pulses that invert spins of a different subpopulation of spins in the sample. This spin inversion is accomplished through application of a 180° pulse at ν_{pump}. If the

inverted spins are in close proximity to the spins being subjected to the spin echo pulse sequence, the precession of the spin echo spins is altered through the dipolar interaction, ultimately resulting in a decrease in the amplitude of the measured spin echo. The strength of the dipolar interaction is inversely proportional to the cube of the distance between spins, which is where interprobe distance will be derived. To ultimately extract the distance between spins from the change in echo amplitude, however, the ν_{pump} 180° pulse has to be applied at varied time points during the spin echo pulse sequence to assemble a DEER waveform. The DEER waveform is a plot of the echo amplitude as a function of time at which the 180° pump pulse is applied. As the time at which the pump pulse is increased, disruption of spin precession eventually reaches a maximum modulation depth at which point subsequent increases in the time at which the pump pulse is applied only promotes echo amplitude oscillation about a mean value (Figure 1.8). The initial rate of echo amplitude decay and the frequency of its modulation are directly related to distance between two spins. The closer two spins are, the faster the decay and the higher the frequency of oscillations. The oscillations in the waveform are also a readout on order within the protein. If a protein is conformationally heterogeneous, the oscillations are dampened out. If the protein is highly ordered, oscillations are more prominent.

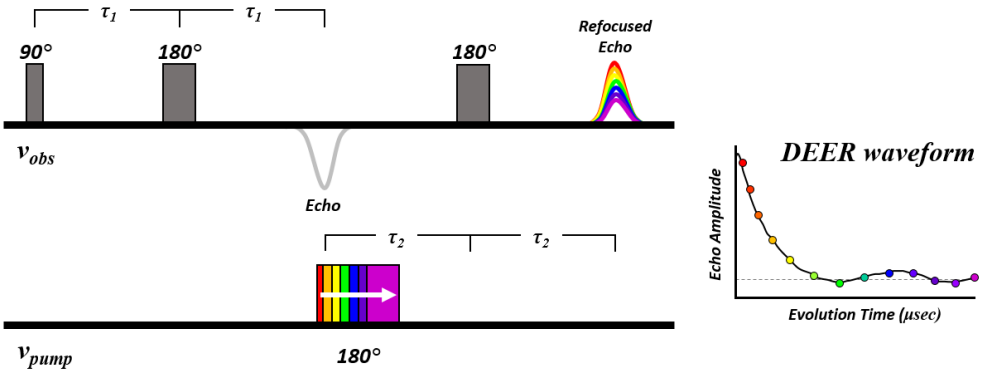

Figure 1.8. Four pulse DEER experiment.[15] The observe frequency (ν_{obs}) moves the net magnetization vector of a subpopulation of spins to the transverse plane and during time τ_1, coherence is lost. A refocusing pulse is then applied such that coherence transiently returns to generate a spin echo. It is during this period that the pump pulse at ν_{pump} is applied to invert spins in a second subpopulation of spins in the sample. Varying the time at which the pump pulse is applied during time period τ_2 results in differing degrees of dipolar disruption to observe spin precession, resulting in loss of echo amplitude that generates the DEER waveform.

When a protein sample is labeled with two spin labels, performing DEER to obtain a waveform allows for a distance measurement to be made and for conformational changes to be subsequently inferred. In we consider the same example discussed above for FRET, if our two-domain protein exists in a compact conformation when not bound to actin filaments, the two spins will be in close proximity resulting in stronger dipolar interactions. This will lead to a DEER waveform with a faster decay and higher frequency oscillations (Figure 1.9). If the protein's two domain then undergo separation upon binding actin, the distance between spins will increase and the dipolar interaction strength will decrease. Consequently, the rate of echo amplitude decay will slow and oscillations will have a lower frequency (Figure 1.9).

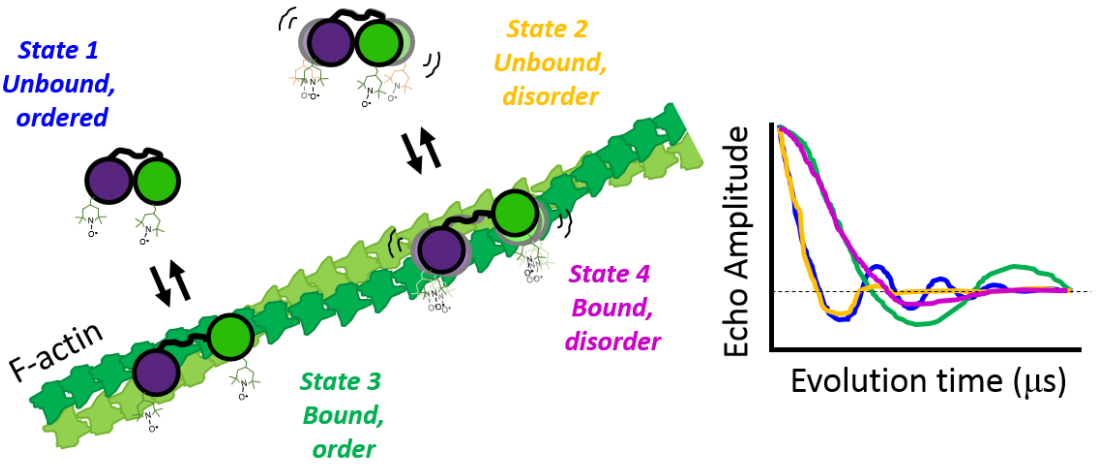

Figure 1.9. Example of DEER-based distance measurement to determine conformation of a protein and the amount of conformational heterogeneity.

1.3 – Structural biology questions relevant to this thesis

1.3.1 – The spectrin superfamily of proteins and their actin-binding domains

Actin-binding domains (ABDs) are found in several different cytoskeletal proteins found throughout many different cell types. Some notable examples of such proteins include dystrophin, utrophin, α-actinin, β-spectrin, fimbrin, plectin and nesprin.[16] These proteins all generally function as cytoskeletal integrators, meaning they bind some cytoskeletal structure such as actin or microtubules and link them to either other filaments or to membrane-embedded proteins which could be localized to either the plasma or nuclear membrane.[16] The relevance of such linkages depends on the protein in question. In the case of dystrophin, its linkage of the muscle actin cytoskeleton to the dystroglycan complex in the sarcolemma[17, 18] aids in the transmission of force from the sarcomere to the extracellular matrix through its indirect linkage to proteins like laminin.[19, 20] The importance of this dystrophin linkage is readily apparent in the types of muscular dystrophy known as Duchenne and Becker muscular dystrophy (DMD and BMD, respectively), where missense or nonsense mutations lead to a loss of functional dystrophin resulting in loss of sarcolemma integrity and ultimately muscle fiber degeneration.[21] In the case of β-spectrin (particularly βIII-spectrin), the protein links cortical actin to the plasma membrane of Purkinje cells in the cerebellum. Spectrin plays a similar membrane stabilizing role that, when disrupted, leads to fragility of membrane structures.[22] Within Purkinje cells specifically, βIII-spectrin seems to be particularly important for maintenance of dendrite membrane structures, as an in-cell deletion of the protein results in dendrite degeneration.[23] This functional role, similar the observation made for loss of dystrophin, is also readily appreciated in a disease context. Mutations

that alter βIII-spectrin's interaction with actin change its ability to maintain dendrite processes[24], resulting in spinocerebellar ataxia type 5 (SCA-5), a condition where motor coordination mediated by the cerebellum is impaired. Without the Purkinje cell's normal expansive dendrite tree, the capacity for integration of motor sensory inputs is significantly diminished. In the case of the protein α-actinin, two α-actinin protein molecules form antiparallel dimers so that both ends of the molecule have ABDs that can bind actin filaments. This allows for α-actinin to act as an actin filament cross linker in several cell types.[25]

Each of the proteins mentioned above ultimately binds to actin by utilizing calponin-homology (CH) domains termed CH1 (most N-terminal in sequence) and CH2 (more C-terminal in sequence) which are located in tandem. The tandem CH domains thus comprise all ABDs found in the spectrin superfamily and are found at the N-terminus of their filamentous structures. While other regions of certain spectrin superfamily proteins (particularly dystrophin and utrophin) do have other segments capable of binding actin despite not being made up of CH domains[26, 27], these regions are found within the spectrin-like repeats that make up the rod domain of each protein and are not the focus of the thesis research. The focus of Chapters 2 and 3 will only be on the ABDs composed of tandem CH domains.

Given their ubiquity and diverse roles in cytoskeletal functions, ABDs became of significant interest to biochemists and structural biologists and the mechanism by which these protein domains bound actin became the focus of intense research. In some of the earliest studies elucidating the mechanism of interaction between ABDs and actin, solution state NMR was the method of choice and it used small peptide fragments from

the putative ABD under study to define regions that associate with F-actin. These studies focused on dystrophin's ABD and came up with the initial identification of residues that seemed important for association.[28, 29] However, binding studies that systematically deleted these potential binding interface residues showed that not all residue regions contributed to binding and it was instead suggested that segments within the first 90-residues were most important for the interaction with actin.[30] When one of the first crystal structures of an ABD was solved from the protein fimbrin[31] (Figure 1.10, left), the regions identified in these earlier peptide and regional deletion studies did not form a particularly obvious binding surface on the ABD molecule. The implicated regions were somewhat discontinuous and few of the residue segments were either contributing to the core fold of a CH domain or were partially buried at the interface between CH1 and CH2.

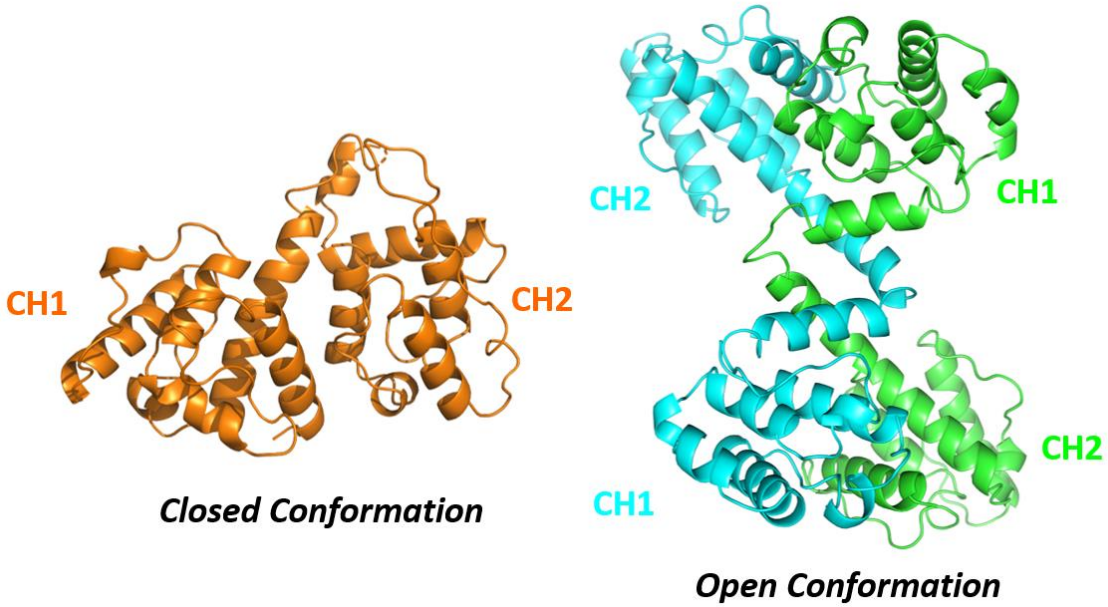

Figure 1.10. Earlier crystal structures of ABDs showing possible variability in conformation of tandem CH domains. Fimbrin on the left (1AOA) and utrophin on the right (1QAG) are shown as examples.

Not long after this structure, however, came the crystal structure of utrophin's first ABD (Figure 1.10, right).[32] Unlike fimbrin's ABD, which crystalized as a compact monomer, utrophin's ABD crystalized as a domain-swapped dimer where the two CH domains were in an extended conformation. Interestingly, the CH1 and CH2 domain interface identified in fimbrin was also found in the utrophin structure, it just involved CH1 and CH2 from different ABD molecules within the dimer. The structural comparison between fimbrin and utrophin lead to the hypothesis that utrophin's ABD, and possibly dystrophin's which crystalized in a similar manner[33], might be more flexible compared to fimbrin and undergo some sort of conformational change in solution. What was not clear, however, was whether or not a more "open" extended conformation was what bound actin filaments or if the original, more compact "closed" conformation (as represented by fimbrin's crystal structure) was binding competent. To answer this question, electron microscopy reconstructions of ABD-actin complexes were performed for a few ABDs. The results varied depending on the ABD studied with fimbrin appearing to bind in its "closed" conformation[34] whereas utrophin exhibited a greater degree of diversity in its binding mode, though binding in an "open" conformation appeared to possible.[35] As information accumulated on other ABDs, particularly disease-causing mutations in α-actinin, electron microscopy was revisited to get a better mechanistic picture of binding and led to the proposal that CH2 sterically hinders binding and thus requires opening of the ABD for CH1-association.[36] This was later supported by DEER experiments that, like the example depicted in Figure 1.9, used spin labels to measure inter-CH domain distance in utrophin's ABD as a function of actin binding.[37] It was found that, in the absence of actin, the interprobe distance was short, consistent with

CH1 and CH2 being in a "closed" state. However, when increasing concentrations of actin were added, the interprobe distance increased indicating separation of CH domains upon binding filaments.

While the DEER experiments on utrophin's ABD started to contribute more to the binding mechanism and were consistent with the CH domain-opening hypothesis, it was still unclear whether or not the binding mechanism was universal for all ABDs or even just similar to some ABDs. A continuation of this research question is what constitutes Chapter 3 of this thesis. DEER spectroscopy is once again used examine the ABD-actin binding mechanism but this time for utrophin's homologue, dystrophin.

The other element of the CH domain-opening hypothesis proposed from cryo-EM structures[36] was in relation to the mechanism of disease-causing missense mutations. In α-actinin and many other ABDs of other spectrin superfamily of proteins[38-40], several missense mutations are localized to the CH2 subdomain, the domain that was proposed to be a negative regulator of CH1 domain forming its binding interface with actin through steric hindrance. Importantly, most of these mutations result in a gain-of-function, where their affinity for actin is increased. Given this high frequency of mutations in CH2 and their common pattern of increasing ABD affinity for actin in vitro, it was proposed that these mutations promote opening of the CH domains, such that CH2 is less of a steric hindrance to CH1's binding of actin.

The content of Chapter 2 in this thesis contributes to the mechanistic understanding of these CH2-localized mutations. This missense-mutation opening hypothesis did not seem to match well with what crystallographers were seeing in their structural models. In α-actinin, for example, one of its gain-of-function mutations

(K255E) was shown to still crystalize in the canonical "closed" conformation[41], ultimately being at odds with the proposed mechanism of disease.[36] In Chapter 2, we try to help resolve this question by examining a different gain-of-function mutation but one that is from βIII-spectrin. This disease-causing mutation is localized CH2 and results in dramatically increased affinity for actin[42], making it a good test subject for structural hypothesis testing. A combination of DEER spectroscopy, cryo-EM, and circular dichroism are all used synergistically in an attempt to elucidate the mechanism of CH2 mutation-induced disease.

1.3.2 – Synaptotagmins in exocytosis

Synaptotagmins are a class of integral membrane proteins that regulate vesicle trafficking events. The best-studied isoform of the family, synaptotagmin 1 (Syt 1), is specifically responsible for regulating synchronicity of neuronal exocytosis. Neuronal exocytosis of neurotransmitters is a complex process involving numerous synaptic proteins. Chief among the dozen or so proteins that promote docking, priming and ultimately fusion of a synaptic vesicle with the plasma membrane are synaptobrevin, SNAP-25, syntaxin, Munc18-1, Munc13, synaptotagmin 1 and complexin.[43] The first three proteins mentioned are all SNARE proteins (soluble N-ethylmaleimide-sensitive factor attachment protein receptor) which zipper together into a four-helix bundle that's thought to bring membranes into close proximity for fusion (Figure 1.11).[44] Munc18-1 is crucial for helping the proper assembly of the SNARE complex and it works together with Munc13 to help position the docked vesicle. Complexin is a small IDP that contains regions that fold upon binding various targets (including other proteins and also

membranes). In part of its central sequence is a SNARE complex-binding region that tightly associates with assembled SNARE complexes and is thought to stabilize it before it progresses to its completely zippered state that promotes membrane merger.[45] Complexin works in some fashion with Syt 1 such that complexin becomes displaced when Syt 1 binds calcium ions (Ca^{2+}).[46]

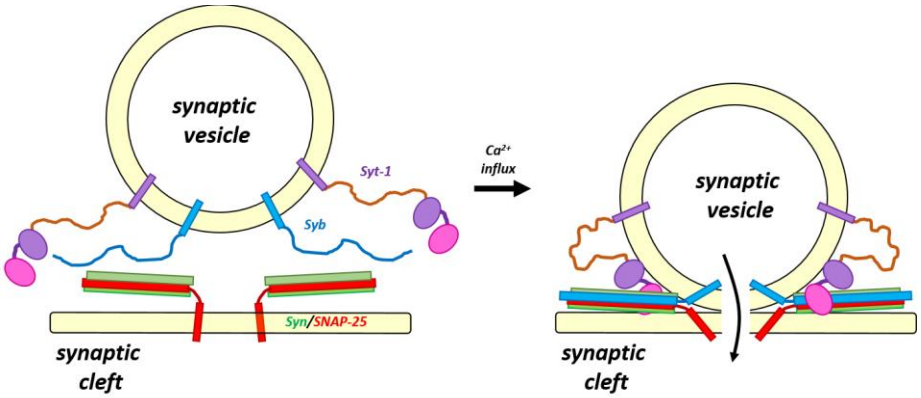

Figure 1.11. Simplified model of SNARE-synaptotagmin-mediated fusion of a synaptic vesicle with the plasma membrane. As the vesicle docks (left) the v-SNARE synaptobrevin (Syb) folds upon binding syntaxin(Syn)-SNAP-25 complexes that were assembled through use of Munc proteins. Complete zippering of the SNARE complex to its state of promoting membrane fusion is promoted upon action potential-induced influx of Ca^{2+} (right).

Syt 1 itself seems to play a role in sensing the Ca^{2+} influx that ultimately triggers complete fusion of vesicle and plasma membranes, but the mechanism is not entirely clear. Syt 1 contains a single-pass transmembrane helix that anchors tandem C2 domains, Ca^{2+} and lipid-binding β-sandwiches, to the vesicle membrane (structure detailed further in Chapter 4). Binding lipids seems to be promoted by Ca^{2+} association to each C2 domain, so one hypothesis for Syt 1's promotion of fusion is that each of its two C2 domains binds to opposing membrane surfaces and helps pull them closer together for SNARE-mediated fusion. In conjunction with lipid interactions, several structural studies suggest that the C2 domain, specifically the C2B domain, can bind to SNARE complexes

directly and that this interaction works together with membrane bridging to promote fusion.[47, 48]

While much work has been done on the structural biology of Syt 1 C2 domains, considerably less effort has been put into studying the structure and function of the ~60 residues that span between the transmembrane helix and the tandem C2 domains. The main assumption in the field is that the sequence is inert and just a flexible tether. This region is intrinsically disordered in solution, but initial studies that looked at its interaction with synthetic membrane suggest structural potential.[49] Chapter 4 of this thesis aims to address the question of what this 60-residue Syt 1 IDR might be doing for function by examining the structural biology of a post-translational modification that occurs in its central region of sequence.

CHAPTER 2: STRUCTURAL BASIS FOR HIGH-AFFINITY ACTIN BINDING REVEALED BY A B-III-SPECTRIN SCA5 MISSENSE MUTATION*

NOTE: My contribution to this project was the pulsed-EPR experiments that measured the closed-open equilibrium in the ABD with and without mutation, circular dichroism measuring secondary structure and stability, and helping write the first draft and all subsequent drafts of the manuscript during the process of publishing. Project conceptualization and initiation was carried out by Adam W. Avery and Thomas S. Hays. Co-sedimentation experiments were carried out by Adam W. Avery. All cryo-EM analysis was carried out by Edward H. Egelman, Fengbin Wang, and Albina Orlova.

2.1 – Section Summary

Spinocerebellar ataxia type 5 (SCA5) is a neurodegenerative disease caused by mutations in the cytoskeletal protein β-III-spectrin. Previously, a SCA5 mutation resulting in a leucine-to-proline substitution (L253P) in the actin-binding domain (ABD) was shown to cause a 1000-fold increase in actin-binding affinity. However, the structural basis for this increase is unknown. Here we report a 6.9 Å cryo-EM structure of F-actin complexed with the L253P ABD. This structure, along with co-sedimentation and pulsed-EPR measurements, demonstrates that high-affinity binding caused by the CH2-localized mutation is due to opening of the two CH domains. This enables CH1 to bind actin aided by an unstructured N-terminal region that becomes α-helical upon binding. This helix is required for association with actin as truncation eliminates binding. Collectively, these results shed light on the mechanism by which β-III-spectrin, and likely similar actin-binding proteins, interact with actin, and how this mechanism can be perturbed to cause disease.

*This chapter was reproduced in its entirety with permissions from the following article: Avery AW, Fealey ME, Wang F, Orlova A, Thompson AR, Thomas DD, Hays TS, Egelman EH. Structural basis for high affinity actin binding revealed by a β-III spectrin SCA-5 missense mutation. *Nat Comm.* 2017 8:1350.

2.2 – Introduction

Spinocerebellar ataxia type 5 (SCA5) is a neurodegenerative disease that stems from autosomal dominant mutations in the cytoskeletal protein β-III-spectrin.[50, 51] SCA5 pathogenesis results from a functional deficit in Purkinje cells, in which the expression of β-III-spectrin is required for normal cerebellar control of motor coordination.[52] β-III-spectrin is thought to form a heterotetrameric complex with α-II-spectrin, and to cross-link actin filaments to form a cytoskeleton localizing to the shafts and spines of Purkinje cell dendrites. β-III-spectrin is required for normal dendrite structure[23] and synpatic transmission[53, 54]. Recently, our group reported that a SCA5 missense mutation, L253P, localized to the β-III-spectrin N-terminal actin-binding domain (ABD), causes a ~1000-fold increase in actin-binding affinity.[42] Here, we probe the structural mechanism of this mutation by studying the ABD with complementary biophysical techniques.

The β-III-spectrin ABD is comprised of tandem calponin homology (CH) domains and the L253P mutation is localized to the second subdomain (CH2). Very little is known about the structural biology of β-III-spectrin's ABD, with the closest related atomic model being the isolated CH2 domain of β-II-spectrin.[55] Crystal structures of N-terminal ABDs from the spectrin superfamily, including α-actinin, dystrophin and utrophin, invariably show that extensive contacts are made between CH1 and CH2, suggesting a tendency to exist in a "closed" conformation without actin.[32, 33, 56, 57] A cryo-EM structure of the fimbrin ABD shows that it associates with actin in a closed structural state.[58] In contrast, cryo-EM showed that α-actinin associates with actin in an "open" structural state in which only a single CH domain is bound to the filament and the second domain is structurally disordered on account of it being dissociated from the interacting

CH domain.[36] A similar conclusion was reached for filamin[59], another member of the spectrin superfamily. Binding studies suggested that the CH1 domain of α-actinin has greater intrinsic affinity for actin in isolation[60] and this suggested that it was CH1 bound in the cryo-EM structure. This led to the hypothesis that the CH2 domain functions to regulate the actin-binding function of CH1 through steric hindrance when the two domains are associated. Consistent with this, many mutations in the CH2 domains of both α-actinin and filamin impart modest gains in ABD affinity for actin.[39, 61] Collectively, these studies suggest that the L253P mutation of β-III-spectrin, which is similarly localized to CH2, causes high-affinity actin binding by disrupting a regulatory mechanism that shifts the ABD structural equilibrium from a closed to more open binding-competent state. Here, we report cyro-EM, co-sedimentation and pulsed EPR data consistent with such a mechanism.

2.3 – Results

2.3.1 – Structure of L253P β-III-spectrin ABD bound to actin

To begin testing our hypothesis, we first performed cryo-EM on the β-III-spectrin ABD bound to actin filaments. The actin-binding affinity of the wild-type (WT) β-III-spectrin ABD is low (K_d =75μM), resulting in poorly decorated actin filaments that were of insufficient quality for analysis (Fig. 2.4a). The L253P ABD yielded high quality complexes of decorated filaments (Fig. 2.4b,c), enabling a three-dimensional reconstruction (Fig. 2.1a) of the mutant ABD-actin complex at 6.9 Å resolution (Fig. 2.5). This represents a substantial improvement over previous ABD-actin reconstructions[36, 58, 62], the best of which was 12 Å. The reconstruction and resulting

atomic model provide several mechanistic insights. First, the density map reveals that only a single CH domain is bound to actin, as observed previously for the α-actinin ABD. Second, the bound CH domain has an additional N-terminal helix that is tightly associated with actin (Fig. 2.1b, red). This helix was not identified in other ABD-actin cryo-EM complexes. However, reexamination of the α-actinin-actin reconstruction[36] suggests that extra density is present, consistent with such an N-terminal helical extension. By comparison, the higher resolution fimbrin-actin reconstruction[58], containing closed CH domains, shows no extra density. The presence of this contiguous N-terminal helix unambiguously identifies the bound domain as CH1. Thus, high-affinity actin binding, caused by the L253P mutation in the CH2 domain, is mediated through the CH1 domain. The L253P mutation does not expose or generate a *de novo* high-affinity actin binding site in the CH2 domain, as has been suggested previously.[63]

All N-terminal ABDs contain amino acid sequences of variable length and composition preceding the conserved CH1 domain. However, a structured N-terminal region preceding the globular fold of a CH domain has not been previously observed in most reported ABD crystal structures. This reflects either disorder in this region or the intentional truncation of the region based on predicted intrinsic disorder.[64] However, when calmodulin was crystallized with the plectin ABD, calmodulin was bound to the N-terminal region which had become α-helical.[64] Solution studies confirmed that in the absence of calmodulin, the plectin N-terminal region is unstructured. The β-III-spectrin CH1 domain with the extended N-terminal helix built into the cryo-EM map superimposes very well with the corresponding plectin CH1 domain with calmodulin (Fig. 2.6), and shows that calmodulin would be involved in massive clashes with actin.

As proposed this explains how calmodulin, in the presence of Ca^{2+}, dissociates the plectin ABD from actin since the binding of actin and calmodulin is competitive.[65]

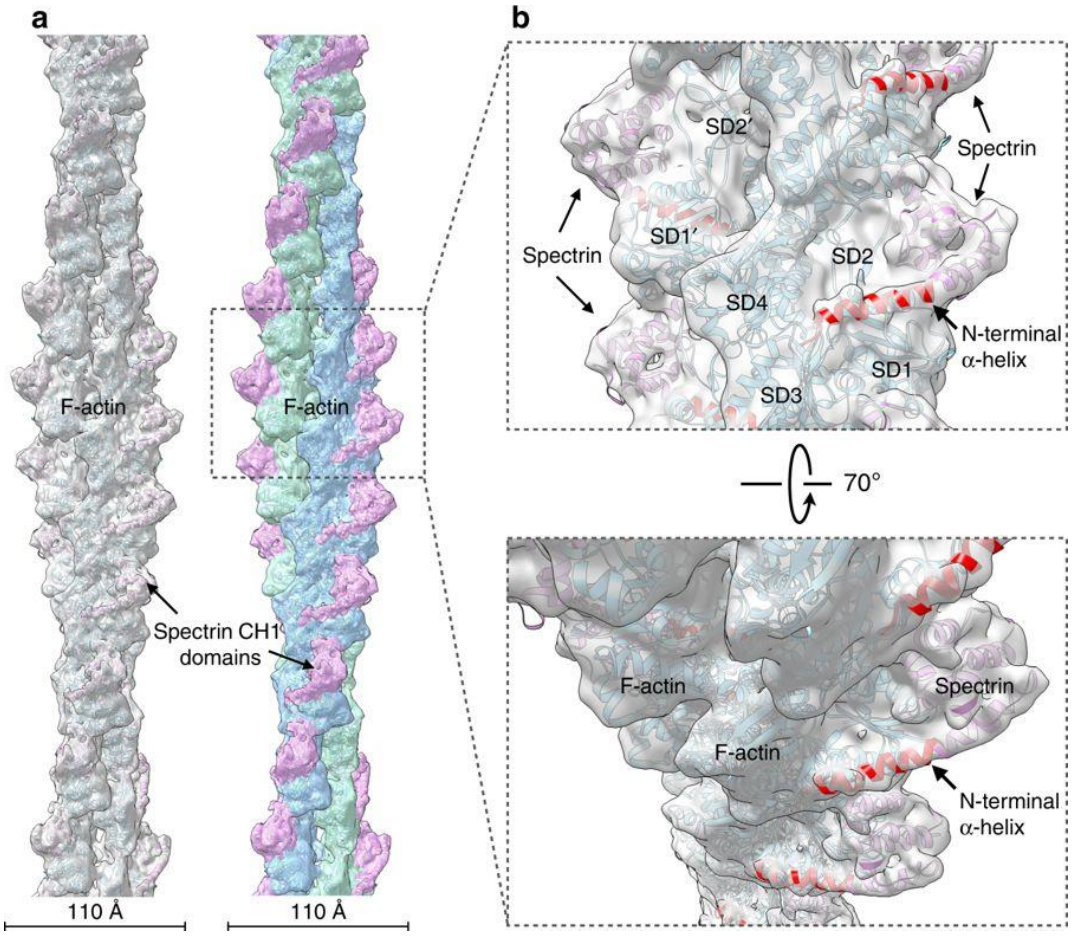

Figure 2.1. Cryo-EM map and model of L253P β-III-spectrin ABD bound to actin. (**a**) The map (left, gray transparent surface) has been fit with a model for actin (cyan) and the β-III-spectrin ABD (magenta). On the right, the surface of the reconstruction has been color coded for the two actin strands (blue and green) and the β-III-spectrin ABD (magenta). (**b**) Close-up view of (a) showing that the CH1 domain has an additional N-terminal helix (red) interacting with F-actin. The actin subdomains (SD1, SD2, SD3 and SD4) have been labeled on one actin subunit, while SD1' and SD2' are labeled on a different subunit.

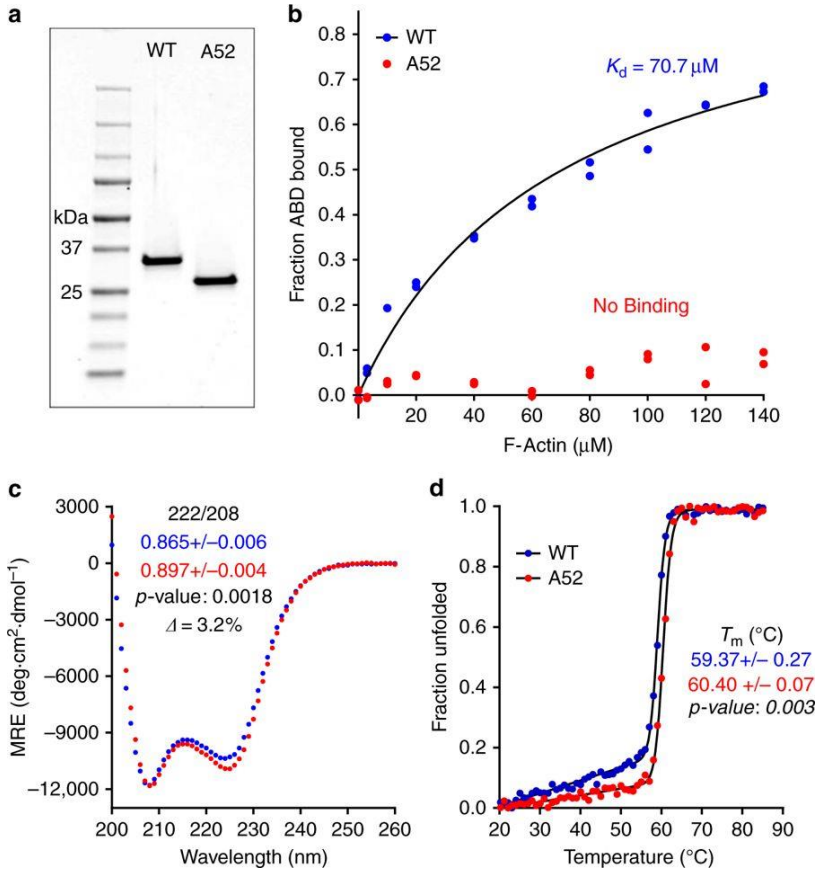

Figure 2.2. The β-III-spectrin N-terminus is required for actin binding. (**a**) Coomassie blue stained gel of purified wild-type ABD or wild-type ABD without the N-terminal 51 amino acids (A52). (**b**) F-actin co-sedimentation assays showing that the N-terminal truncation abolishes actin affinity. (**c**) CD spectra demonstrating α-helical absorption profiles. The A52 ABD has a statistically significant increase in helicity (n=3). (**d**) CD denaturation at 222nm. The A52 ABD has a statistically significant increase in T_m (n=3).

2.3.2 – N-terminal residues of β-III-spectrin ABD are essential for actin binding

The β-III-spectrin cryo-EM structure showing the N-terminal helix bound to actin suggests that the helix must contribute to binding affinity. To test this, we measured affinity of WT ABD with and without the first 51 amino acids (A52). Strikingly, truncation of the N-terminal sequence abolished binding of the ABD to actin (Fig. 2.2a,b). Circular dichroism (CD) indicates that this loss in binding is not due to

misfolding (Fig. 2.2c,d). On the contrary, the A52 ABD showed small but reproducible increases in helicity and stability, suggesting that the N-terminal residues contain intrinsic disorder, which we verified by CD (Fig. 2.7). Collectively, these data sets, combined with the cryo-EM structural model, indicate that the N-terminal sequence is critical to binding, both directly by interacting with actin and potentially indirectly through allosteric destabilization of the ABD, the latter of which may alter the ABD conformational ensemble to impact binding. Evidence supporting an allosteric contribution comes from recent folding and binding studies on the utrophin ABD1 which showed that the utrophin N-terminal residues destabilize ABD1 and are required for full ABD1 binding activity. The utrophin N-terminal residues alone do not bind actin.[66]

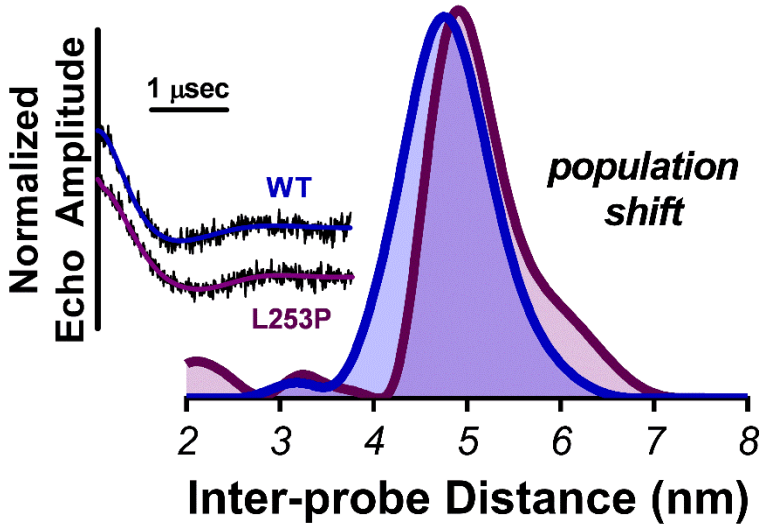

Figure 2.3. DEER measurement showing the L253P mutation opens the β-III-spectrin ABD structure. Echo amplitude decays of WT ABD (blue) and L253P ABD (purple) along with their corresponding Tikhonov fits are shown on the left. The inter-probe distances derived from Tikhonov regularization (Fig. 2.9) for both WT and L253P ABDs are shown on the right. The WT ABD distance distribution is centered at 4.8 nm, consistent with the distance predicted in the homology model of the closed state shown in Fig. 2.8. Upon introduction of the L253P mutation, the distance distribution undergoes a shift to populate a longer inter-probe distance, visible as a shoulder to the right of the 4.8 nm peak, consistent with structural opening of the ABD.

2.3.3 – L253P mutation in β-III-spectrin ABD promotes structural opening of CH domains

Previously we demonstrated that the L253P mutation substantially destabilizes the ABD (ΔT_m = -14.8 °C). If the open structural state of the ABD is responsible for high-affinity actin binding, then decreased stability may facilitate opening of the CH domains from a closed state. In our previous structural homology model of the β-III-spectrin ABD in the closed structural state[42], L253 is positioned at the CH domain interface (Fig. 2.8), suggesting that the L253P mutation could also perturb CH1-CH2 interactions that stabilize the closed structural state. To test this hypothesis, we used double electron-electron resonance (DEER) to measure inter-CH domain distance with and without the L253P mutation. We exploited the native cysteine residues at positions 76 and 231 for irreversible attachment of spin labels.

For WT ABD in the absence of actin, clear oscillations were present in the echo amplitude decay (Fig. 2.3). Analysis (Fig. 2.9) revealed an inter-probe distance centered at 4.8 nm (Fig. 2.3) which agrees well with the distance predicted for the closed structural state (Fig. 2.8). Upon introduction of the L253P mutation, the distance distribution undergoes a shift to populate a longer inter-probe distance, visible as a shoulder to the right of the 4.8 nm peak. This indicates that the ABD undergoes an opening as a result of the mutation (Fig. 2.3), consistent with the cryo-EM structure showing that CH1 separates from CH2 upon associating with actin (Fig. 2.1).

In the context of other ABDs harboring disease-causing mutations, a similar structural mechanism has been proposed for the K255E mutation of α-actinin-4, a

missense mutation also located at the CH domain interface. This mutation caused increased affinity for actin, but the crystal structure showed the ABD of mutant α-actinin-4 to be in a closed state.[41] While seemingly counter to our proposed mechanism here, the DEER-derived distance distributions indicate that the shift to an open conformer is not complete (Fig. 2.3). Only a small portion of the population exists in the more open state in the absence of actin, with the remaining ensemble occupying the closed conformer. Given that crystallographic conditions favor more stable conformers, the K255E mutant may open, similar to L253P β-III-spectrin, but crystallize in its more stable closed state because it is more significantly populated.

2.4 – Discussion

SCA5 pathology is characterized by atrophy of the cerebellum[67], likely reflecting degeneration of dendritic arbors extended by Purkinje cells. Within dendrites, β-III-spectrin binds to actin filaments to form a spectrin-actin skeleton underlying the plasma membrane.[68] The low affinity of WT β-III-spectrin for actin suggests that normal membrane function requires a dynamic spectrin-actin cytoskeleton in which spectrin-actin linkages form and dissociate. We suggest that the high affinity of L253P β-III-spectrin for actin decreases dynamics of spectrin-actin linkages, resulting in reduced plasticity of the spectrin-actin cytoskeleton. We speculate that spectrin-actin cytoskeleton plasticity is important for the cytoskeleton to expand or retract within structurally dynamic regions of the dendritic arbor, such as growing or remodeling dendrites and spines. Recent work has highlighted the requirement of β-III-spectrin to support dynamic spine structure and post-synaptic signaling.[54] In addition, disrupted microtubule-based

transport has been reported for the L253P mutation[69], and these transport defects may disrupt arborization and contribute to SCA5 pathogenesis. Disrupted transport may be secondary to defects in microtubule tracts that are organized by the spectrin-actin cytoskeleton[70-73], and/or result from the direct impact of high-affinity binding of L253P β-III-spectrin to the actin-related protein, ARP1[74], a component of the dynactin complex that facilitates cargo transport by microtubule motor proteins.[75, 76]

Collectively, the 6.9 Å cryo-EM structure, binding studies, and DEER distance distributions converge on a structural mechanism for disease. The CH2 domain-localized L253P mutation perturbs a closed-open structural equilibrium in β-III-spectrin's ABD by lowering the energetic barrier between structural states. The ABD is then relieved of its regulatory mechanism allowing for the CH1 domain to interact with actin filaments, aided by an additional N-terminal unstructured region that becomes helical upon binding actin.

2.5 – Materials and Methods

2.5.1 – Protein Purification

For cryo-EM analyses, the wild-type or L253P human β-III-spectrin ABD coding sequences contained in pET-30a-ABD WT or pET-30a-ABD L253P vectors were expressed in *E. coli* BL21(DE3) (Novagen). ABD proteins were both purified using a HiTrap Q 5 mL ion-exchange column followed by a Superdex 200 size exclusion column (GE Healthcare Life Sciences).[42] Elution fractions of the Superdex 200 column containing pure ABD proteins as assessed by SDS-PAGE were pooled and concentrated (Amicon Ultra-4 Centrifugal Filter, 10 K MWCO). A Bradford assay (Biorad) was then

used to determine protein concentrations which equaled 44.0 and 40.6 µM for wild-type and L253P, respectively. ABD proteins were stored on ice until preparation of ABD-F-actin complexes.

To test the contribution of the N-terminus to actin-binding affinity, the coding sequences for wild-type ABD (amino acids 1-284) or truncated wild-type ABD (amino acids 52-284) were PCR amplified using the forward primer AAACACCTGCAAAAAGGTATGAGCAGCACGCTGTCACCC or AAACACCTGCAAAAAGGTGCAGATGAACGAGAAGCTGTGC and reverse primer AAATCTAGACTACTTCATCTTGGAGAAGTAATGGTAGTAAG. PCR products were digested with AarI and Xba1 restriction enzymes and ligated into the BsaI site of pE-SUMOpro (LifeSensors) containing His and SUMO tags. The final constructs pE-SUMO-ABD WT and pE-SUMO-A52-ABD WT were sequence verified and transformed into *E. coli* BL21 (DE3)pLysS (Agilent). Transformed bacteria were incubated with rotation at 27°C in flasks containing 1L LB media with 100 µg per mL ampicillin and 50 µg per mL chloramphenicol until an absorbance of 0.5 at 550 nm was reached. Then flasks were placed in ice for 10 min before addition of IPTG to 0.5 mM final. The flasks were then incubated with rotation for 4 h in a 22°C water bath. Bacteria were harvested at 5,000 x g and pellets stored at -20°C. Bacteria were lysed with lysozyme (Sigma) for 1 h at 4°C in buffer containing 50 mM Tris, pH 7.5, 300 mM NaCl, and 25% sucrose with protease inhibitors (Complete Protease Inhibitor tablet, EDTA-free, Roche), followed by a freeze-thaw cycle using an isopropanol-dry ice bath. Then $MgCl_2$ to 10 mM final and DNase1 (Roche) to 7.5 U per mL final were added and lysate incubated for 1 h at 4°C. Lysate was clarified at 40,000 x g at 4°C for 30 min. Supernatants were collected and

passed through a 0.45 µm syringe filter before loaded onto a Poly-Prep (Biorad) chromatography column containing 1 mL Ni-NTA agarose (Qiagen) equilibrated in buffer containing 50 mM Tris, pH 7.5, 300 mM NaCl and 20 mM imidazole. The column was washed with buffer containing 50 mM Tris, pH 7.5, 300 mM NaCl and 20 mM imidazole, and proteins eluted in buffer containing 50 mM Tris, pH 7.5, 300 mM NaCl and 150 mM imidazole. Fractions containing ABD proteins were pooled and loaded into a Slide-a-Lyzer, 10K MWCO, dialysis cassette (ThermoScientific), and dialysis performed at 4°C in buffer containing 25 mM Tris, pH 7.5, 150 mM NaCl and 5 mM β-mercaptoethanol. To cleave off the SUMO tag, Ulp1 SUMO protease was added to dialyzed ABD proteins at a 1:14 (protease:ABD) mass ratio, and digests incubated for 1.5 h at 4°C. To separate ABD proteins from the cleaved His-SUMO tag and His-tagged SUMO protease, ABD proteins were loaded onto a Poly-Prep chromatography column containing 0.5 mL Ni-NTA agarose equilibrated in 25 mM Tris, pH 7.5, 150 mM NaCl, and 5 mM β-mercaptoethanol. Elution fractions containing ABD proteins were collected and then loaded onto a gel filtration column (Sephadex S100, GE) equilibrated in buffer containing 10 mM Tris, pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$ and 1 mM DTT at 4°C. Fractions were analyzed by SDS-PAGE and Coomassie blue staining, and fractions enriched with ABD proteins were pooled and concentrated (Amicon Ultra-15 Centrifugal Filter, 10 K MWCO).

For DEER analyses, the wild-type and L253P ABD proteins were modified to substitute a serine residue in place of cysteine 115. Spin labeling was performed using native cysteines of which the β-III-spectrin ABD contains four (C76, C115, C186, C231), and C76, C115 and C231 are all solvent exposed. Residues C76 and C231 were best

suited for DEER distance measurements, so C115 was mutated to serine to prevent non-specific labeling at that site. The C186 site did not require mutagenesis because it is naturally buried in the core of CH2 and would thus not be accessible to free spin label in solution. PCR site-directed mutagenesis was performed on pET-30a-ABD WT and pET-30a-ABD L253P vector templates using the oligonucleotides CATGCGGATCCACTC-CCTGGAGAACGTG and CACGTTCTCCAGGGAGTGGATCCGCATG. The resulting constructs pET-30a-ABD WT C115S and pET-30a-ABD L253P C115S were sequence verified, and then transformed into *E. coli* BL21(DE3). The ABD C115S proteins were purified as described above.[42] For structural studies of the β-III-spectrin N-terminal residues, a small peptide corresponding to residues SSTLSPTDFDSLEIQGQYSDINNRWDLPDSDWDNDSSSARLFERSRIKALA was produced via solid-state synthesis through Selleck Chemicals LLC.

2.5.2 – Cryo-EM ABD-actin

5 µM of rabbit skeletal muscle G-actin was polymerized in 15 mM Hepes-HCl buffer, pH 7.5, 75 mM KCl, 1 mM $MgCl_2$, and 0.5 mM ATP for 2 h at room temperature. For negatively stained samples 2 µM F-actin was incubated with 5-10 µM wild-type β-III-spectrin or with 2-5µM β-III-spectrin mutant L253P for 2-20 min. For cryo-samples 1.5 - 2µL of the mixture was applied to lacey carbon grids that were plasma cleaned (Gatan Solarus) and vitrified in a Vitrobot Mark IV (FEI, Inc.). Grids were imaged in a Titan Krios at 300 keV, and recorded with a Falcon II direct electron detector at 1.05 Å per pixel, with seven "chunks" per image. Each chunk, containing multiple frames, represented a dose of ~ 20 electrons per $Å^2$. A total of 586 images (each 4k x 4k) were

selected that were free from drift or astigmatism, and had a defocus less than 3.0 μm. The program CTFFIND3[77] was used for determining the Contrast Transfer Function (CTF) and the range used was from 0.6 to 3.0 μm. The SPIDER software package[78] was used for most subsequent steps. The CTF was corrected by multiplying each image by the theoretical CTF, both reversing phases where they need to be reversed and improving the Signal-to-Noise ratio. The program e2helixboxer within EMAN2[79] was used for boxing long filaments from the micrographs. Overlapping boxes, 384 px long with a 40 px shift between adjacent boxes (~ 1.5 times the axial rise per subunit) were extracted from these long filaments, yielding ~ 60,000 segments that were padded to 384 x 384 px. The CTF determination and particle picking came from the integrated images (all seven chunks), while the segments used for the initial alignments and reconstruction came from the first two chunks.

An initial reconstruction using the IHRSR method[80] showed clear decoration of actin, but the mass density due to β-III-spectrin was lower than that from the actin. This appeared to arise from incomplete occupation. We therefore used atomic models of pure actin and actin decorated with α-actinin[36] to sort the segments. Only ~ one-third of the segments showed a higher cross-correlation with the decorated filament, and 20,340 segments were used for further processing. The IHRSR cycles converged to a rotation of -166.9° and an axial rise of 27.3 Å per subunit. After excluding segments with a large out-of-plane tilt or poor orientation against the reference, 12,443 segments were used in the final reconstruction.

2.5.3 – Model building

An actin-spectrin asymmetric unit was segmented from the filament map in Chimera.[81] Model building began by docking cryo-EM structure of actin (5JLH)[82] and a predicted model of spectrin CH1 domain into the experimental density data. This predicted spectrin model was generated by I-TASSER[83] based on a crystal structure of plectin (1MB8)[84]. Then the actin-spectrin complex was rebuilt with RosettaCM protocols.[85] A total of 1,500 models were generated, and the 10 best models (selected based on Rosetta's energy function) were combined into one model by manual editing in Coot[86] to yield the best overall fit to the density map. A filament model was subsequently generated from this and refined by Phenix real-space refinement.[87] MolProbity[88] was used to evaluate the quality of the model (Table 2.1). The MolProbity scores for the actin-spectrin filament models compare favorably (99th percentile) with structures of similar resolution.

Although segments were sorted to exclude naked actin, Phenix refinement of the actin-spectrin reconstruction clearly shows that the occupancy by spectrin is not 100% percent, and the actual occupancy is ~ 75%. Therefore, the threshold chosen for the filament needed to show the full volume for spectrin shows a somewhat larger and lower resolution actin.

2.5.4 – Co-sedimentation assays

Actin was purified from acetone powder derived from the psoas muscle of New Zealand white rabbit.[89] Acetone powder was hydrated in 4°C water for 30 minutes to extract actin. The resultant slurry was passed through a Whatman filter paper and 30 mM

KCl was added to the filtrate to polymerize actin for a period of 1 hr at room temperature. Filamentous actin was then pelleted by 30 minute centrifugation at 80,000 rpm in a TLA 100.3 rotor. The actin pellet was resuspended in buffer containing 5 mM Tris pH 7.6, 0.5 mM ATP, 0.2 mM $MgCl_2$ and homogenized on ice. A 10 minute clarifying spin at 70,000 rpm was then performed to pellet aggregate proteins. The G-actin containing supernatant was then isolated and polymerization was initiated by addition of 2 mM $MgCl_2$ with a 30 minute incubation at room temperature. The purified ABD proteins were clarified at 100,000 x g at 4°C for 30 min prior to setting up binding assays. A Bradford assay was performed to determine F-actin and clarified ABD protein concentrations. Binding assays were performed in F-buffer containing 10 mM Tris, pH 7.5, 150 mM NaCl, 0.5 mM ATP, 2 mM $MgCl_2$ and 1 mM DTT. Binding reactions contained 3 µM ABD protein and F-actin ranging from 0 to 140 µM in a total volume of 60 µL. Binding reactions were incubated at room temperature (23-24°C) for 30 min to reach equilibrium, and then F-actin pelleted by centrifugation at 100,000 x g at 25°C for 30 min. Unbound ABD was measured by combining 45 µL of binding reaction supernatant with 15 µL 4x Laemmli sample buffer and performing SDS-PAGE followed by Coomassie blue staining. After destaining, gels were scanned using the 700 nm channel in an Odyssey Imager (LI-COR Biosciences). The fluorescence intensities of the ABD protein bands were quantified in Image Studio Lite Ver 5.2 software (LI-COR Biosciences).[90] Individual ABD band fluorescence intensities were converted to amount ABD protein. This conversion was performed using a standard curve generated by linear regression (Prism 5 software) on ABD Coomassie blue fluorescence intensity values attained from a SDS-PAGE gel

loaded with varying amounts of ABD in F-buffer. To determine the dissociation constant (Kd) value, data were fit by non-linear regression in Prism 5 software to the equation:

$$Y = X/(K_d + X) \quad \text{Eq. 2.1}$$

Where Y equals fraction ABD bound and X equals free F-actin concentration.

2.5.5 – Circular dichroism

ABD proteins were clarified at 100,000 x g for 20 min at 4°C. A Bradford assay was performed to determine ABD protein concentrations, and ABD proteins were diluted to 250 ng/µL in buffer containing 10 mM Tris, pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT. CD spectra were acquired in a Jasco J-815 Spectropolarimeter equipped with a Peltier temperature controller. Immediately before analysis, the instrument was baseline-corrected using ABD protein buffer. For secondary structure analyses, CD spectra were measured from 200 nm and 260 nm at 25°C. Thermal unfolding of the ABD protein sample was analyzed by recording CD spectra at 222 nm over the temperature range of 20°C-85°C. CD analyses were performed three times for each ABD protein. Non-linear regression analysis was performed in Prism 5 (GraphPad Software, Inc.) to determine the melting temperature using the following equation for a two-state transition, as reported previously[91]:

$$Y = (\alpha_N + \beta_N T)/(1 + e^{4T}) \quad \text{Eq. 2.2}$$

For secondary structure analysis of the N-terminal peptide, a lyophilized powder was reconstituted in the same buffer system as that described for the ABD proteins. The reconstituted peptide was then diluted to a concentration of 99 ng per µL and subsequently scanned over the same wavelength range described above. In the case of both ABDs and peptide, raw ellipticity was normalized to each sample's respective concentrations according to the following equation:

$$MRE = [\theta(MW/N - 1)]/(lc) \quad \text{Eq. 2.3}$$

Where θ represents the raw ellipticity, MW represents the protein molecular weight, N is the number of amino acids, l is the path length, and c is the concentration in milligrams per milliliter.

2.5.6 – Statistical analyses

Unpaired t-tests were performed in Prism 5 software to determine whether significant differences existed in ABD protein melting temperatures or 222 nm/208 nm absorbance ratios determined by CD. The n value was equal to three in all cases.

2.5.7 – Spin labeling

In the β-III-spectrin ABD constructs, 500 µM of the spin label 4-maleimido-TEMPO (MSL, 4-maleimido-2,2,6,6-tetramethyl-1-piperidinyloxy; Sigma-Aldrich) was added to 25 µM protein and equilibrated on a rocker for 3 hours at 4 °C. Prior to addition

of MSL, the protein solution had been run over a Zeba desalting column pre-equilibrated with 10 mM Tris pH 7.5 and 150 mM NaCl to remove most of the 1 mM DTT left over from size exclusion. After the spin-label incubation period, the protein was once again subjected to a Zeba desalting column to remove any unreacted spin label. To ensure complete removal, however, the spin labeled protein was then subjected to three 4 h rounds of dialysis in 4 liter solutions contain 10 mM Tris pH 7.5, 150 mM NaCl and 1 mM DTT. MSL was ultimately chosen over the more commonly used (1-oxyl-2,2,5,5,-tetramethylpyrroline-3-methyl) methanethiosulfonate (MTSSL) because spin labeling of the β-III-spectrin ABD constructs was incomplete, requiring inclusion of DTT reducing agent post-labeling to prevent undesired ABD cross-linking. Incubation of the ABD constructs with spin label for periods longer than 3 hours resulted in significant protein loss due to precipitation. The spin labeled WT and L253P β-III-spectrin ABD constructs were concentrated down to 230 μM and 175 μM, respectively, prior to spin counting and DEER sample preparation.

2.5.8 – EPR spectroscopy

To verify labeling, a continuous wave (CW) electron paramagnetic resonance (EPR) spectrum was acquired with sample temperature of 296 K on the E500 Bruker EPR spectrometer operating at X-band (9.5 GHz) and equipped with a super high Q (SHQ) cavity. The derivative spectrum was then doubly integrated to determine spin concentration by comparing with the double integral of a 100 μM TEMPOL standard (Fig. 2.8b and 2.8c). For WT and L253P β-III-spectrin ABD constructs, spin concentrations were determined to be 98 μM and 75 μM, respectively, indicating a

labeling efficiency of ~43% for both. After spin counting, we performed double electron-electron resonance (DEER) on β-III-spectrin ABD constructs doubly labeled with MSL. ABD samples were prepared by adding 7% v/v glycerol (as a cryoprotectant), loading samples into quartz capillaries (1.1 mm i.d., 1.6 mm o.d., 15 μL sample volume) and subsequently flash freezing samples in liquid nitrogen after which samples were stored at -80 °C until use. A Bruker E580 spectrometer operating at Q-band (34 GHz) with an EN5107 resonator was then used to implement a four-pulse DEER protocol with a π/2 pulse width of 12 ns and an electron double resonance (ELDOR) pulse width of 24 ns.[15] The ELDOR frequency was set to the absolute maximum of the field swept nitroxide absorption spectrum and the observe position was set 24 G up-field. Experiments were run at a temperature 65 K. After data acquisition, background-corrected DEER decays were analyzed using the Tikhonov regularization method provided in DeerAnalysis2016 to extract distance distributions encoded in the waveform (Fig. 2.9).[92] To determine the stable components in the resulting Tikhonov distributions, we examined the impact of a range background models whose starting fit positions varied from 0.5 μs to 2.4 μs on the uncorrected waveform. The components that were invariable were used for structural interpretation of the ABD constructs. The component that was not stable (the peak at >7.0 nm) was excluded from structural interpretation as it represents an artifact of background subtraction.

2.6 – Supporting Information

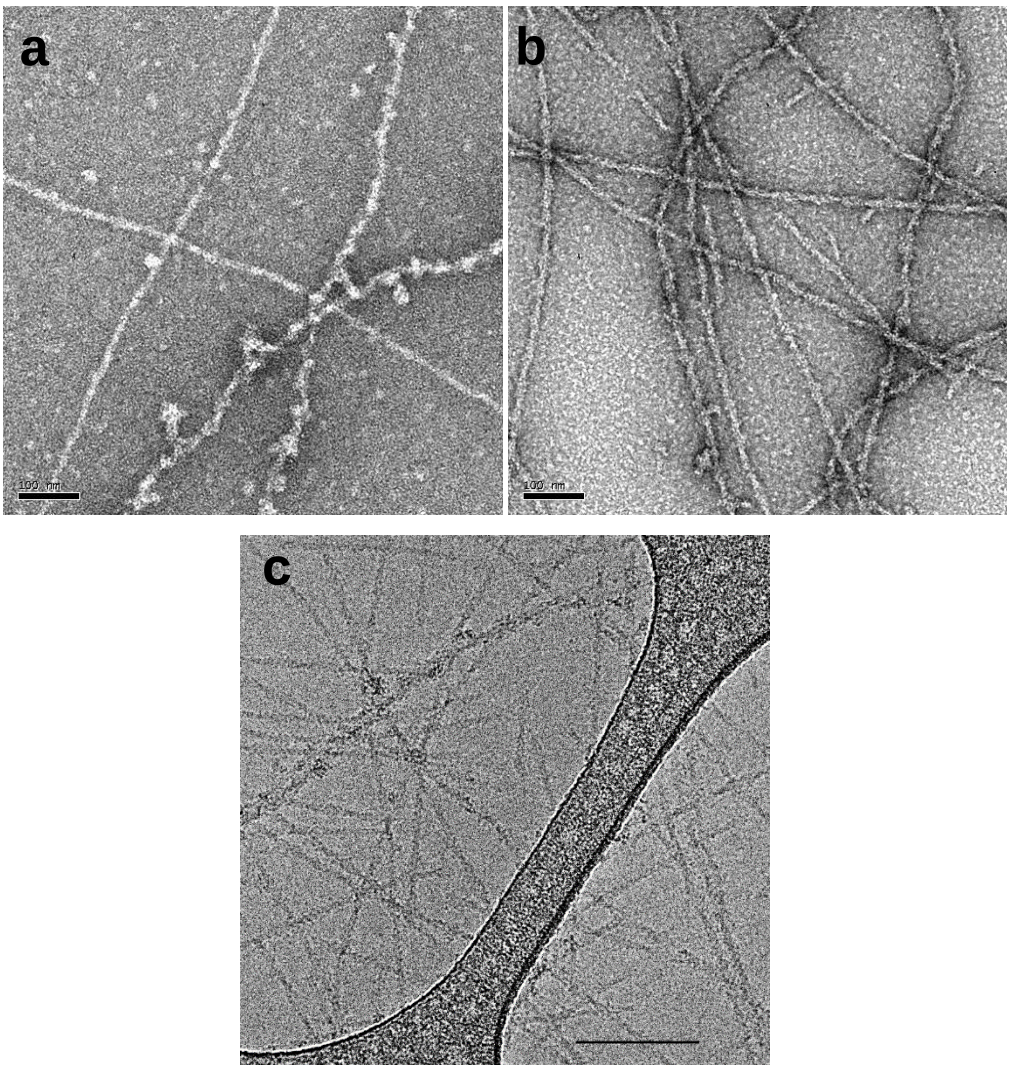

Figure 2.4. Electron microscopy of β-III-spectrin ABD-actin complexes. Negative staining of (**a**) WT ABD-actin showing poor decoration of filaments and (**b**) L253P ABD-actin. Cryo-EM was used (**c**) for the 3D reconstruction of the L253P ABD decorated actin filaments. Lacey carbon grids were used, with the filaments suspended in the holes within this carbon film. The carbon film is seen near the center of the image, with holes on both sides of it. The space bars in (**a**), (**b**) and (**c**) are 1,000 Å.

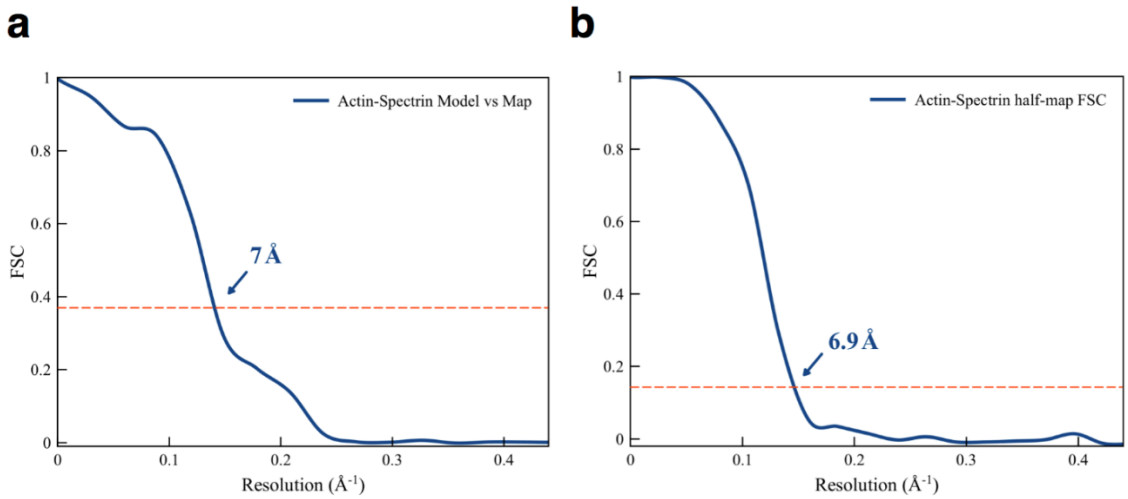

Figure 2.5. Resolution of the Actin-spectrin filament reconstruction, estimated two different ways. (**a**) The resolution was derived from Fourier Shell Correlation (FSC) calculation between the refined atomic model and the map. (**b**) A "gold-standard" FSC is also calculated between two half maps. Given the relatively small number of segments used in the final reconstruction (~12k), splitting this into two halves to generate a conventional "gold standard" map:map FSC can be problematic, as the result from the half-maps is biased to a lower resolution than the full map.[93] Nevertheless, this method yielded an estimate of 6.9 Å resolution at FSC=0.143. For the map:model comparison, we used FSC=0.38 where 0.38=√0.143.

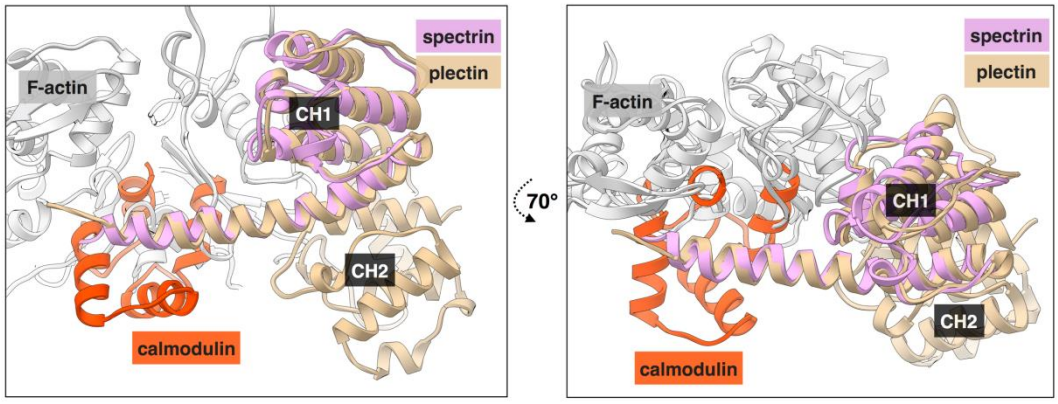

Figure 2.6. Comparison with calmodulin-bound plectin structure (4Q57). F-actin (gray), β-III-spectrin (pink), plectin (brown) and calmodulin (orange) are shown. Plectin model is aligned with spectrin using only the N-terminal helices (rmsd=0.8 Å), and a major clash would result between calmodulin and F-actin.

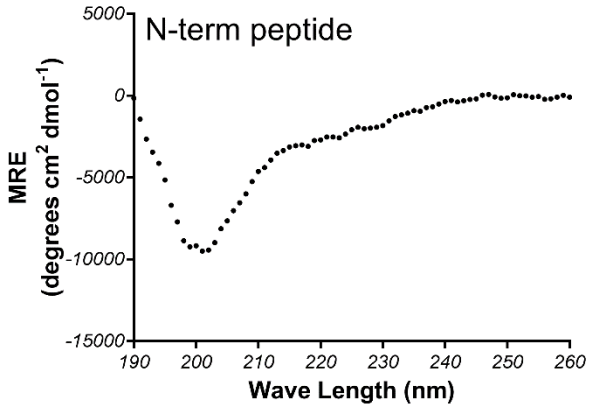

Figure 2.7. Circular dichroism spectra on a peptide corresponding to the β-III-spectrin N-terminus. The absorption profile is consistent with a disordered structure.

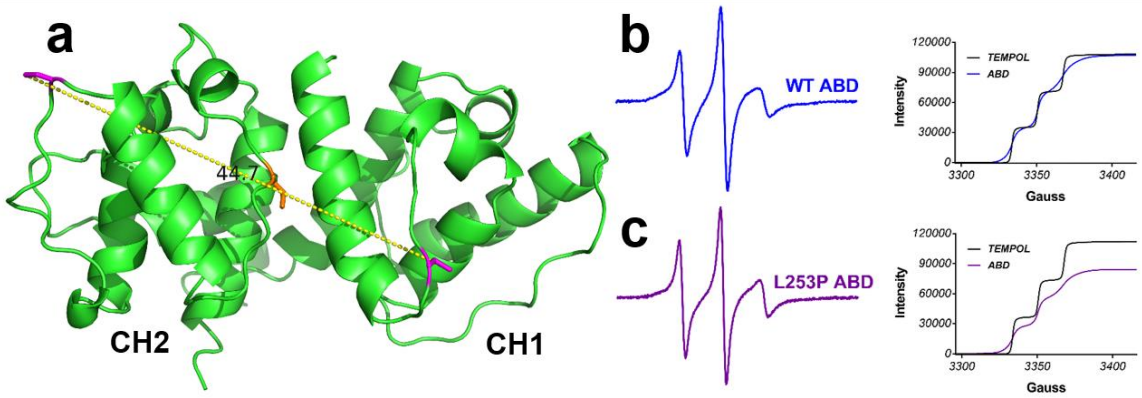

Figure 2.8. Site-directed spin labeling of β-III-spectrin ABD. (**a)** Homology model generated for β-III-spectrin ABD from crystal structure of the closed state of α-actinin, indicating the approximate inter-probe distance (side chain-side chain distance of 4.5 nm, yellow dotted line) between two native cysteine residues (C76 and C231, magenta residues) and highlighting L253 in orange at the CH domain interface. (**b** and **c**) CW EPR spectrum (left) of MSL-labeled β-III-spectrin ABD and example spin count (right) for WT (blue) and L253P mutant (purple). Note probe mobility is as expected for cysteine residues located in loops.

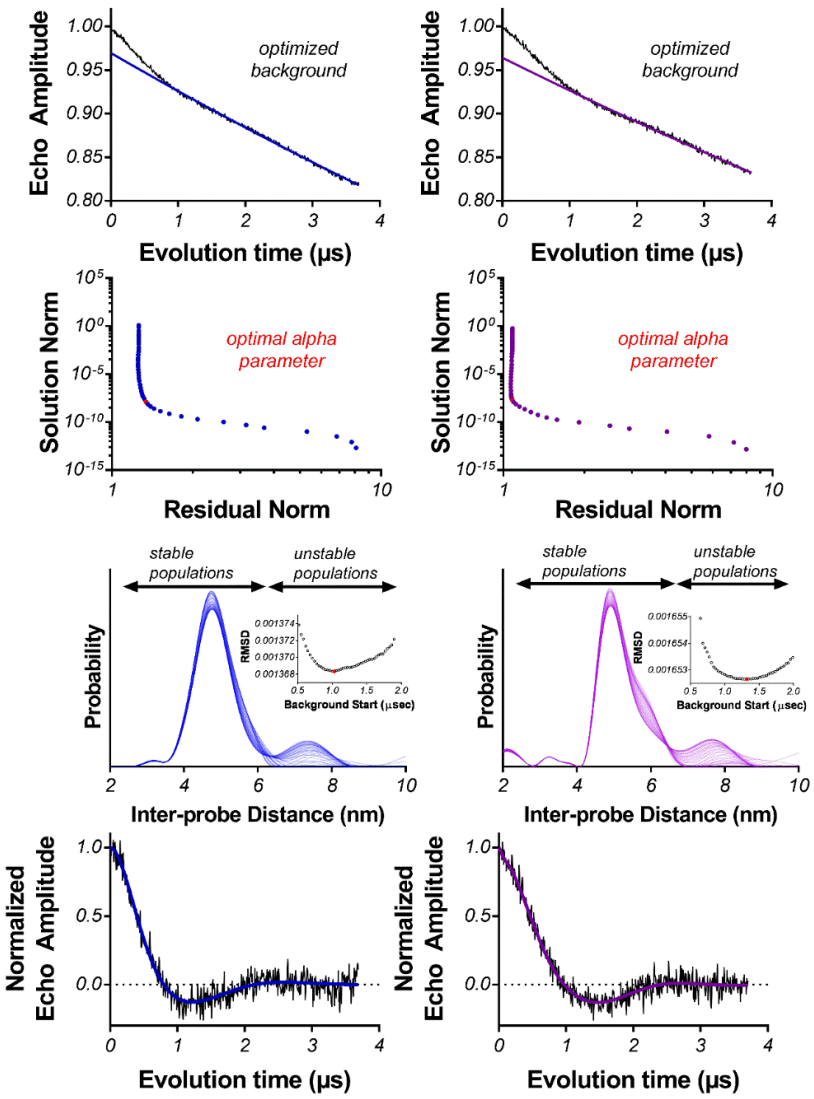

Figure 2.9. DEER analysis of WT (blue) and L253P (purple) β-III-spectrin ABDs. (top row) Raw echo amplitude decays with optimal background. (middle rows) L-curve analysis indicating smoothing parameter chosen from Tikhonov fit of the echo amplitude decay. As the distance distributions from fitting the DEER waveform are highly sensitive to the choice of the background component, we examined the impact of a range background fits (varying the start of the fit from 0.5 μsec to 2.4 μsec on the uncorrected waveform). The components that were invariable with respect to choice of background were used for structural interpretation of the ABD constructs. The component that was not stable, and in some cases disappeared entirely (the peak(s) at >7.0 nm), was excluded from structural interpretation as it represents an artifact of background subtraction. (bottom row) Normalized echo amplitude decays fit with the Tikhonov model corresponding to the lowest RMSD.

	Actin-spectrin filament
Clash score, all atoms	6.6
Protein geometry	
Ramachandran favored (%)	92.7
Ramachandran outliers (%)	0
Rotamer outliers (%)	0
Cβ deviations > 0.25 Å (%)	0
RMS deviations	
Bonds (Å)	0.01
Angles (°)	1.00
MolProbity score	1.82 (99^{th}, 3.25 Å – 7.25 Å)
PDB ID	6ANU

Table 2.1 – Refinement statistics for the actin-spectrin filament model by MolProbity.[88]

CHAPTER 3: DYNAMICS OF DYSTROPHIN'S ACTIN-BINDING DOMAIN*

NOTE: MF and BH equally contributed. Benjamin Horn contributed molecular dynamics along with Alessandro Cembran whereas Michael Fealey contributed the spectroscopy and wrote the manuscript.

3.1 – Section Summary

We have used pulsed EPR, calorimetry, and molecular dynamics simulations to examine the structural mechanism of binding for dystrophin's N-terminal actin-binding domain (ABD1) and compare it to utrophin's ABD1. Like other members of the spectrin superfamily, dystrophin's ABD1 consists of two calponin-homology domains, CH1 and CH2. Several mutations within dystrophin's ABD1 are associated with development of severe degenerative muscle disorders Duchenne and Becker muscular dystrophies, highlighting the importance of understanding its structural biology. To investigate structural changes within dystrophin ABD1 upon binding to actin, we labeled the protein with spin probes and measured changes in inter-CH domain distance using double electron-electron resonance (DEER). Previous studies on the homologous protein utrophin showed that actin binding induces a complete structural opening of the CH domains, resulting in a highly ordered ABD1-actin complex. In the present study, DEER shows that dystrophin ABD1 also undergoes a conformational opening upon binding F-actin, but this change is less complete and significantly more structurally disordered than observed for utrophin. Using molecular dynamics simulations, we identified a hinge in the linker region between the two CH domains that grants conformational flexibility to ABD1. The conformational dynamics of both dystrophin's and utrophin's ABD1 showed

*This work was reproduced in its entirety with permission from the following article: Fealey ME, Horn B, Coffman C, Miller R, Lin AY, Thompson AR, Groth E, Hinderliter A, Cembran A, Thomas DD. Dynamics of dystrophin's actin binding domain. *Biophys J*. 2018 115:445-454.

that compact conformations driven by hydrophobic interactions are preferred, and that extended conformations are energetically accessible through a flat free energy surface. Considering that the binding free energy of ABD1 to actin is on the order of 6-7 kcal/mole, our data are compatible with a mechanism in which binding to actin is largely dictated by specific interactions with CH1, but fine tuning of the binding affinity is achieved by the overlap between conformational ensembles of ABD1 free and bound to actin.

3.2 – Introduction

Dystrophin (Dys) is a large muscle cytoskeletal protein of 427 kDa. Structurally, Dys consists of an N-terminal actin-binding domain (ABD1), 24 spectrin-like repeats that form the rod domain and house the second actin binding domain (ABD2), and a C-terminal region containing a cysteine-rich domain that binds the dystroglycan complex.[94] The N- and C-terminal ends of the protein provide a biochemical picture of how Dys can link the costameric F-actin network to both the sarcolemmal membrane and extracellular matrix via the dystroglycan complex. This location in the cytoskeletal network of muscle is crucial for maintaining membrane and costamere integrity, a fact that is best appreciated in the context of Duchenne and Becker muscular dystrophies (DMD and BMD, respectively).[95] In DMD and BMD, a multitude of mutations (missense and nonsense alike) in Dys lead to the development of muscle tissue degeneration. Patients with these forms of muscular dystrophy have elevated cytosolic calcium levels, indicative of an extracellular leakage of the ion through the sarcolemma due to focal membrane tears. Similarly, Dys-deficient model organisms, particularly *mdx* mice, show the same elevated cytosolic calcium phenotype.[96]

The biophysical mechanism by which Dys protects the sarcolemma from mechanical stress is still an active research question. Dys contains several spectrin-like repeat domains (triple helix bundles) between the N-terminal ABD1 and C-terminal ZZ domain that anchors it to the dystroglycan complex. Atomic force microscopy has been used to unfold these types of protein domains from spectrin, showing that they have a low energy barrier to unfolding, which probably serves as a mechanical means to dissipate transduced force propagating through the costamere.[97] However, independent of such a

mechanism is the impact Dys has on the actin cytoskeleton alone. Previously, it was shown that the large-scale bending and twisting motions of F-actin are greatly restricted when both Dys and its homologous partner utrophin (Utr) bind.[98, 99] Intriguingly, the amplitude of these large scale motions is reduced, but the rate of motion is increased indicating that both Dys and Utr binding impart resilience to the actin cytoskeleton. Such an effect could be another means by which the two proteins dampen the laterally transduced force of contraction within the costamere.

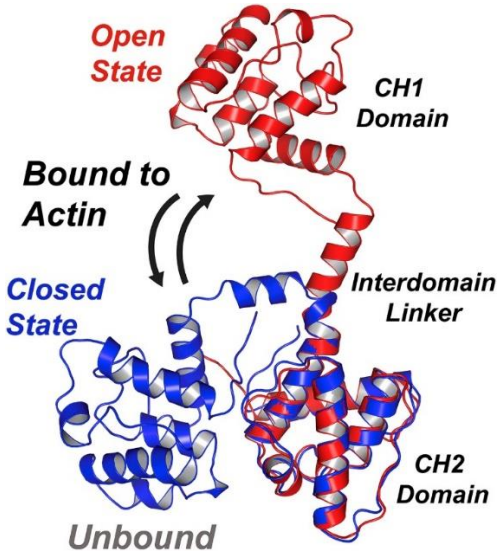

Figure 3.1. Proposed structural model for Dys ABD1 upon binding F-actin. In the absence of actin, the two adjacent CH domains are closely packed (closed state, blue). In the presence of actin, CH1 and CH2 can become more separated (open state, red) and sample multiple structural states. To see figure in color go on-line.

While both Dys and Utr were found to restrict the amplitude of motion, the proteins were found to do so differentially, with Utr being the more restrictive of the two proteins.[98, 99] Although it is unclear how such functional differences arise, a possibility is that they originate from differences in the structural binding modes of each protein's

actin-binding domains. Previously, the Utr ABD1 was shown to undergo a conformational opening upon association with F-actin; Utr ABD1's adjacent calponin-homology (CH) domains transitioned from a more closed compact state to an open extended state forming a well-ordered and stable complex with F-actin.[37] We propose here that the Dys ABD1 can also undergo a similar structural transition (Fig. 3.1), but with distinct nuances that could contribute to a less stable ABD1-F-actin complex and, in turn, a smaller restrictive contribution to overall filament motions. Indeed, spectroscopic evidence in support of Dys ABD1 having a distinct binding mode has been reported.[100]

Several ABDs from proteins in the spectrin superfamily have been characterized in both open[32, 33] and closed[38, 41, 56] conformations using x-ray crystallography. Solution studies have converged on a similar finding; analytical ultracentrifugation[41] and pyrene excimer fluorescence studies[100] have identified ABDs in a closed conformation whereas pulsed-EPR experiments show an equilibrium between closed and more open conformations, specifically for utrophin's ABD. This open/closed dichotomy persists when ABDs are bound to actin. Both mutation studies[101] and cosedimentation assays[102] suggest a predominant role played by CH1 to binding affinity, with CH2 adopting a regulatory role. Excimer fluorescence suggests that Dys ABD1 binds to actin in a closed conformation[100], but DEER shows clearly that Utr ABD1 binds to actin in an open conformation[37], and cryo-EM models suggest an open conformation for actin-bound α-actinin and β-III-spectrin ABDs.[36, 103] These observations suggest that ABDs are malleable domains whose conformations are susceptible to actin binding and, potentially, even to the small perturbations introduced by different experimental techniques. It is thus

appealing to hypothesize that the observed differences in Dys and Utr ABD1 interactions with actin stem from their distinct structural dynamics.

To test this hypothesis in the present study, we have used a combination of double electron-electron resonance (DEER) experiments and molecular dynamics (MD) simulations to examine the conformational dynamics of Dys ABD1. By applying these methods in parallel, we were able to define a mechanism of Dys ABD1 binding to actin that is distinct from that of Utr ABD1, providing insight into the different physical properties of these two proteins that may contribute to physiological function.

3.3 – Materials and Methods

3.3.1 – Protein mutagenesis, expression, purification, and labeling

A plasmid encoding the mouse Dys ABD1 protein fragment (residues 8-246) was subjected to mutagenesis using a Q5 ® Site-Directed Mutagenesis Kit (New England BioLabs) to remove the protein's native cysteines and incorporate new cysteines at positions 120 and 239. These positions were chosen to mimic our previous study on Utr ABD1 and also because they were suitable for measuring the proposed distance changes. The mutated plasmid was then transformed into BL21 DE3 *E. coli* and subsequently grown to an OD of 0.7, at which point protein expression was induced with 2 mM IPTG per liter of growth. Use of mouse Dys ABD1 was chosen so that we could compare spectroscopic measurements to that of Utr ABD1 published previously, which was also from mouse.[37] Expression was allowed to occur overnight at room temperature. After expression, cells were pelleted at 10,000 rpm. Cells were subsequently lysed in 50 mM Tris pH 8, 20% w/v sucrose, 1 mM EDTA and 1 mM DTT by incubation with lysozyme

for 1 hour at 4 °C, followed by freeze-thaw cycling in a dry ice isopropanol bath. The resulting lysate was then incubated with DNase I for 1 hour at 4 °C. All cell debris was removed from solubilized protein by centrifugation at 18,500 rpm in a SS-34 Sorvall rotor and was subsequently sterile-filtered through a 0.2 μm filter to ensure complete removal of particulates. The lysate was then loaded onto a GE 5 mL SP ion-exchange column and eluted over a linear gradient (0-500 mM) of NaCl buffered by 10 mM Tris pH 7.5, 1 mM EGTA, and 1 mM DTT. Elution fractions containing the ABD1 fragment were then concentrated and further purified using an S100 sepharose gel-filtration column. Protein fractions pooled after size exclusion were then dialyzed in buffer containing 10 mM Tris, 100 mM NaCl, 2 mM $MgCl_2$ and 1 mM DTT at pH 8, buffer conditions under which actin-bundling does not occur.[104] Final protein purity was verified by SDS-PAGE.

For DEER experiments, a 5-fold excess (to cysteine residue concentration) of 4-maleimido-TEMPO (MSL, 4-maleimido-2,2,6,6-tetramethyl-1-piperinyloxy; Sigma-Aldrich) was added to a DTT-free Dys ABD1 sample and allowed to incubate for 3 h at 25 °C. Free label was subsequently removed by 4 rounds of dialysis in 4 liters of 10 mM Tris, 100 mM NaCl, 2 mM $MgCl_2$ and 1 mM DTT at pH 7.5. The resultant labeled protein was spin counted to determine the label concentration which, when compared to the protein concentration, resulted in ~84% labeling efficiency (Fig. 3.6). Spin labeled samples were then loaded into quartz capillary tubes (1.1 mm ID, 1.6 mm OD, 15 μL sample volume) containing 7% v/v glycerol as a cryo-protectant and flash frozen in liquid nitrogen and subsequently stored at -80 °C until use.

3.3.2 – Double electron-electron paramagnetic resonance (DEER)

We performed DEER experiments to measure interprobe distances in the range of 2-6 nm. Measurements were made on a Bruker E580 spectrometer (Billerica, MA) operating at Q-band (34 GHz) with an EN5107 resonator. A four-pulse DEER protocol with a $\pi/2$ pulse width of 12 ns and an electron double resonance (ELDOR) pulse width of 24 ns was implemented. The ELDOR frequency was placed on the pump position which corresponded to the absolute maximum of the nitroxide absorption spectrum. The observe position was placed 24 Gauss higher than the pump position on the field-swept spectrum. Experiments were performed at 65 K. The resulting DEER waveform was analyzed using the model-independent Tikhonov fit provided in DeerAnaylsis2013.2. The Tikhonov distribution was then fit to multiple Gaussian distance distributions (equations 3.1 and 3.2) assuming the existence of discrete structural states as described previously[105]:

$$\rho(r) = 1/(\sigma\sqrt{2\pi}) \exp((r-R)^2/(2\sigma^2)) \quad \text{Eq. 3.1}$$

$$\sigma = (FWHM)/(2\sqrt{\ln(2)}) \quad \text{Eq. 3.2}$$

3.3.3 – Differential scanning calorimetry (DSC)

DSC experiments were performed on a NanoDSC (TA Instruments, New Castle, DE) at a scan rate of 1 °C/min using micromolar protein concentrations. Buffers consisted of 20 mM MOPS, 100 mM KCl at a pH of 7.5. Solutions were extensively degassed under vacuum with gentle stirring prior to loading into the calorimeter. This helps prevent release of air bubbles from solution during the experiment. From the

thermodynamics parameters enthalpy of unfolding (ΔH), melting temperature (T_m), and change in baseline heat capacity (ΔC_p), we calculated a free energy of unfolding (ΔG) using the Gibbs-Helmholtz equation:

$$\Delta G = \Delta H(1-T/T_m) + \Delta C_p(T-T_m-T\ln(T/T_m) \quad \text{Eq. 3.3}$$

Further details of our analysis using multiple approaches to constrain each parameter are discussed in the Supporting Materials and Methods.

3.3.4 – Molecular Dynamics Simulations

Molecular dynamics simulations were performed using GROMACS 5.0.6[106] starting from chain A of the crystal structure of Dys ABD1 (PDB 1DXX[33]), corresponding to residues 9-246 of human Dys (Uniprot P11532). Similarly, for Utr the starting structure was taken from chain A of the PDB 1QAG[32] of human utrophin. For the latter, using the Uniprot P46939 sequence, both the N- and C-termini were elongated with PyMOL by 6 and 5 residues, respectively, so that the resulting sequence 25-261 aligns well with the one simulated for dystrophin (Fig. 3.7). For Dys ABD1, amino acids S10 and S188 were mutated back to cysteine residues using CHARMMGUI[107], and for both systems water seen in crystals was preserved. All histidine residues were set to neutral charge with hydrogen on the Nε, except for Utr ABD1 His 88 and 190, which were neutral with proton on Nδ. Periodic boundary conditions were implemented through a rhombic dodecahedron box with a box vector length of 11.6 nm. The system was solvated in TIP3P water[108] and KCl was added to achieve electroneutrality and an ionic

strength of 150 mM. Covalent hydrogen bonds were constrained with the LINCS[109] algorithm, and the equations of motion were propagated with a time step of 2.0 fs. Electrostatic interactions were treated with the particle-mesh Ewald[110] algorithm with a real cut-off and a grid spacing of approximately of 1.2 nm and 0.12 nm, respectively. Van der Waals interactions were switched off between 1.2 nm and 1.0 nm. A constant temperature of 300.0 K was maintained with the V-rescale algorithm[111] applied to the protein and solvent groups independently, and a constant pressure of 1 atm was maintained with the Parrinello-Rahman barostat.[112] Simulations were run using two flavors of the CHARMM force field family: CHARMM36[113] and CHARMM22*[114, 115]. For each system (Dys and Utr), 5 simulations were run with che CHARMM36 force field, and 5 with the CHARMM22* force field, for a total of 20 independent simulations. Comparisons of structural parameters derived from each force field are summarized in Fig. 3.8-3.11. For both force fields, the system was initially minimized to remove bad contacts, and then it was equilibrated for 1 ns at constant volume and temperature, followed by 3 ns at constant pressure and temperature during which the harmonic restraints on the protein's heavy atoms were gradually reduced from 239 kcal mol^{-1} nm^{-2} to zero. At this point, for each force field 5 independent simulations were started by randomizing the atoms' velocities, and continued for 500 ns, resulting in a total simulation time of 5 μs for the ten total trajectories for each system. Coordinates of the trajectories were saved every 5 ps and additional analyses (see Supporting Information) indicated that to improve statistics the trajectories obtained with the two force fields could be pooled together. The trajectories were processed with GROMACS, data analysis

and plotting was performed with the software R[116], and structures were visualized with PyMOL[117].

Principal component analysis (PCA)[118] was performed on the protein's backbone atoms to identify the main conformational changes that characterize ABD1's dynamics. To directly compare the conformational states of Dys and Utr, the principal components were calculated by pooling the trajectories of both systems together, and then each system's trajectories were projected on the common principal components. Residue Q189 for Dys was omitted from the analysis because sequence alignment of Dys to Utr (Fig. 3.7) shows a gap in Utr in that position. The trajectories were then binned along the first two principal components, and the relative density of the histograms was used to calculate the free energy difference through the Boltzmann distribution. To identify flexible regions in the protein backbone, the conformations of the Cα of 4 consecutive residues were binned using the MK32K25 structural alphabet[119] and the corresponding Shannon entropy was calculated using the method of Pandini et al[120].

As a proxy for the distance between the two spin labels on residues 120 and 239, we monitored the distance between the two Cα carbons of these residues. Additionally, solvent accessible surface area (SASA)[121] was measured for the total protein, and separated in the hydrophobic and hydrophilic contributions. The differences in SASA between the open conformation of ABD1 and the closed conformations were plotted on the closed conformation's representative structure using PyMOL.

3.4 – Results

3.4.1 – Dys ABD1's structure transitions from a closed to open state, but with significant structural disorder

To test the hypothesis that the CH domains of Dys ABD1 undergo a conformational opening upon binding of F-actin, we performed DEER experiments on a spin labeled ABD1 fragment. Application of DEER allows us to make a direct comparison to existing structural data on the homologous Utr ABD1 fragment.[37] Using site-directed mutagenesis, we removed all native cysteine residues in the Dys ABD1 construct and engineered in new cysteines at residue positions 120 and 239, positions similar to those used in Utr ABD1 in ref. 59. Based on the domain-swapped dimer crystal structure previously determined for Dys ABD1[33], these residue locations would be suitable for detecting a change in interdomain distance upon F-actin association.

After spin-labeling the protein, we measured interprobe distances with and without varying concentrations of F-actin. In the absence of F-actin, DEER measurements identified a short interprobe distance, indicating that the CH1 and CH2 domains of Dys ABD1 are in a more compact closed state (Fig. 3.2, top). This is apparent in the DEER echo amplitude, which shows a rapid decay and mild oscillations in the time domain, both of which indicate a short interprobe distance. In samples containing F-actin, the DEER echo amplitude decays more slowly, indicating longer distances. The oscillations initially present in the absence of F-actin are dampened out in its presence, indicating actin-induced structural disorder (Fig. 3.2, middle and lower panels).

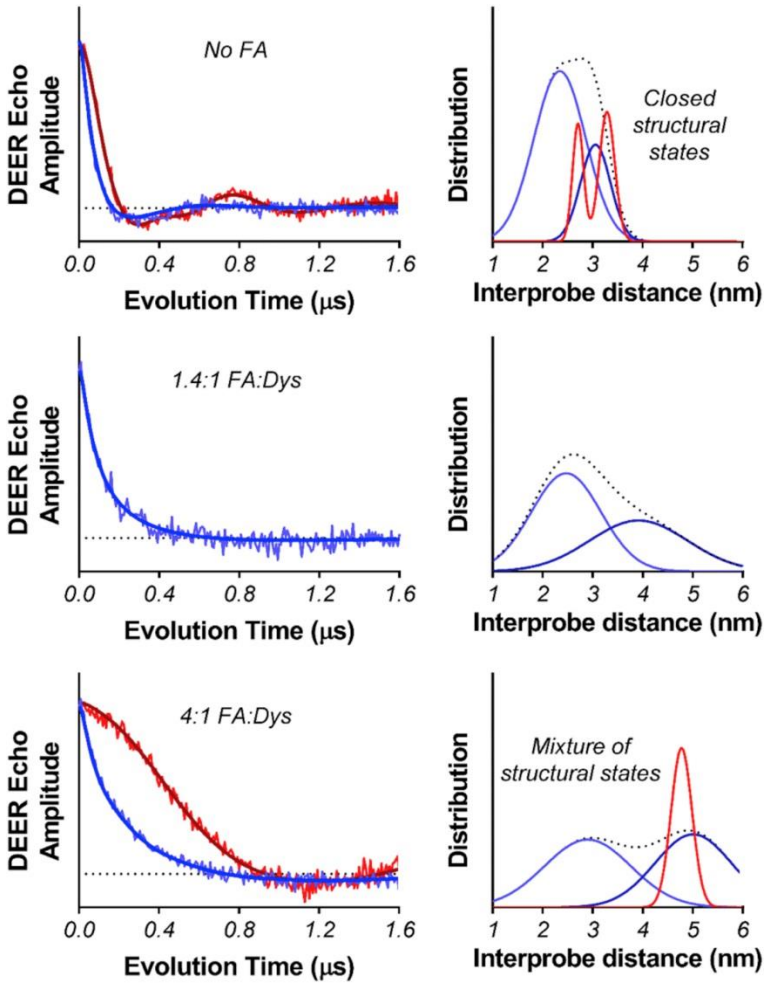

Figure 3.2. DEER data on 80μM MSL-labeled Dys ABD1 (blue data sets) in the presence of increasing F-actin from top to bottom. Molar ratio of F-actin (FA) to Dys ABD1 is shown above time domain data. Utr ABD1 DEER reported previously in ref[37]. (red data sets) is overlaid for comparison. Left: time-domain decays. Right: derived distance distributions. Tikhonov distributions for Dys ABD1 (black dotted lines) were fit to two discrete Gaussian distributions (blue solid lines) and indicate that F-actin shifts interprobe distance toward a more open structural state but there is considerable structural disorder. Note that this contrasts with Utr ABD1's Gaussian distributions (shown as red solid lines) which are ordered. Moreover, Utr ABD1's structural opening is complete. For complete time domains see Fig. 3.12. To see figure in color go on-line.

When comparing these structural results to the previously examined Utr ABD1, several features stand out. First, both ABDs share a common mechanism of opening upon binding to F-actin. Second, the Dys ABD1 contains much more structural disorder than

seen previously in Utr ABD1. This is evident in the mild or complete lack of oscillations present in the echo amplitude decay of Dys which, in Utr ABD1, were very well resolved. Lastly, unlike the Utr ABD1, Dys ABD1 does not undergo a complete shift to the open structural state. This is evident by the fact that a shorter interprobe distance is still sampled even in the presence of excess of actin. This latter observation is consistent with previous work examining the DysABD1 open-closed structural transition upon binding actin[100]; Pyrene excimer fluorescence using residues C10 and C188 indicated that Dys ABD1 occupied a closed conformation when bound to actin. This method, however, would not have been able to detect the open conformers identified in this study by DEER. Collectively, the pulsed EPR data identifies a common binding mode for Dys and Utr ABD1, but simultaneously highlights distinct structural dynamics in the ABD-actin interaction.

3.4.2 – MD simulations characterize a compact & conformationally heterogeneous ABD1

To gain atomistic insight into the factors affecting the compact vs. extended conformational equilibrium, we performed MD simulations starting from the Dys and Utr ABD1 crystal structures, both of which are in an open conformation.[32, 33] Ten 500 ns simulations were independently run for each ABD, and in 18 out of 20 simulations the ABD structures similarly collapsed to compact conformations. The results show that not all simulations sampled the same conformation and a small number of stable compact conformations were identified for both Dys and Utr using principal component analysis (PCA) (Fig. 3.3, minima B-F). The free energy profiles in Fig 3.3 show that for both Dys and Utr at least 5-6 distinct compact and one extended conformations exist. Moreover,

these conformations have similar energies and the different structures can easily interconvert into each other by overcoming small free energy barriers (3-4 kcal/mol).

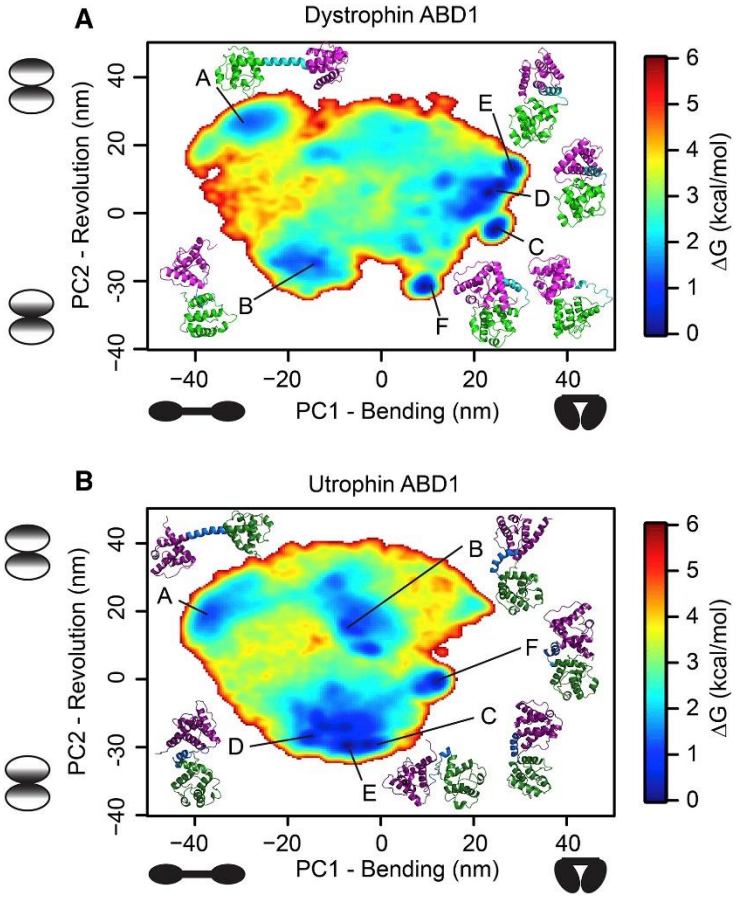

Figure 3.3. Free energy landscape of Dys ABD1 (A), and Utr ABD1 (B) projected on the first two principal components. The two principal components describe a "bending" motion of the two CH domains around a central swivel (PC1) that allows the extended-to-compact transition, and the "revolution" motion of one CH domain around the other (PC2). A representative structure for each of the major conformational minima are also plotted. Structural heterogeneity correlates with the measured structural disorder present in DEER distributions and low unfolding free energy. To see figure in color go on-line.

Using the method of Pandini et al.[120], we calculated the backbone Shannon entropy as a proxy of its conformational flexibility and projected the calculated entropy on the structure (Fig 3.4 A). The entropy of the linker region spanning from Gly-130 to Asn-135 is comparable to that of the disordered N- and C-termini, indicating that this

region is characterized by a great degree of conformational flexibility, as recently suggested by Chakravarty et al.[122]. This suggests that the linker region closer to the CH2 domain acts as a swivel, allowing the conformational transitions between compact and extended conformations, as well as the revolution of one CH domain around the other. This role is supported by monitoring the change in φ/ψ dihedral angles between the extended (minimum A in Fig 3.3 A) and a compact conformation (minimum F in Fig 3.3 A), which highlights the very same region as the one in which dihedrals change the most (Fig. 3.13).

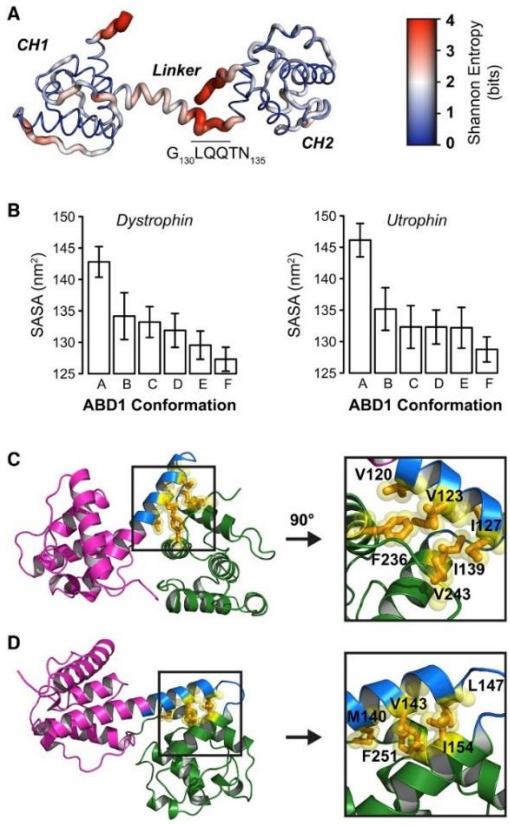

Figure 3.4. Structural collapse of dystrophin and utrophin ABD1, derived from MD simulations. (A) Dys ABD1 Shannon entropy calculated for macrodiehedrals formed by 4 consecutive Cα carbons. (B) SASA averaged over the ensemble of structures defining each minimum identified in Figure 3.3, with error bars representing one standard deviation. Representative closed structural states of (C) Dys and (D) Utr ABD1. Hydrophobic residues that promote closure are highlighted. To see figure in color go on-line.

Next, we aimed at identifying the driving force inducing the transition to compact conformations. Upon further analysis, all compact structures display a smaller solvent accessible surface area when compared to the extended conformation (Fig. 3.4 B). Although some of the residues involved may vary in each different compact structure, we observed that a core of hydrophobic residues on the same face of the amphipathic linker helix and a cluster of residues on the CH2 domain are predominantly involved in hydrophobic interactions in all compact conformations (Fig 3.4 C and D, Fig. 3.14). Such a pattern could indicate that the closed structural states are stabilized in part by interactions involving hydrophobic side chains. A similar mechanism was recently proposed for the ABD of β-III-spectrin[42], and it is also supported by the fact that the same hydrophobic interactions are present in the extended crystal structure of Dys ABD1 (1DXX), although because of the dimeric nature of the crystal they are domain-swapped (Fig. 3.15).

3.4.3 – Experimental validation of MD simulations

To directly test the MD results that the distribution of closed conformers is promoted by hydrophobic contacts, differential scanning calorimetry (DSC) experiments were performed. The increase in a protein's heat capacity (ΔC_p) during an unfolding transition correlates directly with increased exposure of apolar amino acids, based empirically on the free energy of transfer between organic and aqueous phases (Supporting Materials and Methods,). Comparison of this calculated ΔC_p with that experimentally measured by DSC indicates the extent of hydrophobic residue exposure to water in the folded state. We found that the measured value of ΔC_p for thermal unfolding of Dys ABD1 (4.09 ± 0.07 kcal/mol) was substantially less than the calculated value (4.7

kcal/mol) (Table 3.1). This is consistent with the MD conclusion that the native state of Dys ABD1 has an unusually large fraction of hydrophobic residues exposed to solvent.

Next, to compare the ensemble of conformations generated by MD simulations to the experimental data obtained by DEER, we plotted for Dys ABD1 the probability density of the distance between the Cα of Val-120 and Leu-239, the residue positions where spin labels were inserted (Fig. 3.16). The plot shows a broad peak centered at 1.2 nm with a shoulder at 2.0 nm, corresponding to the closed conformations. The features of the peak match the features of the same peak in the DEER experiments in absence of actin (Fig 3.2). We characterized the two underlying distributions as both due to compact conformations, differing only by the relative orientation of the two CH domains (Fig. 3.16). In the DEER data, the peak is centered at about 1 nm longer distance than in the data from the simulations. This could be explained by the fact that in the simulations the distance was mapped between the two Cα rather than from the ends of the long TEMPO spin label. In addition, the two mutations introduced to add the spin labels may somewhat interfere with the hydrophobic patch between the linker and the CH2 domain (in particular perturbing Val-120 and Phe-236 which is only 3 residues away from Leu-239). Yet, it was reassuring to notice that the *change* in distances metric was preserved. In fact, the DEER data indicates an increase of about 2.5 nm upon transition from the compact to the extended conformations (Fig. 3.2 top and bottom panels) and a similar trend is observed in the simulated data when considering the small shoulder at 4.0 nm, which represents the extended conformation.

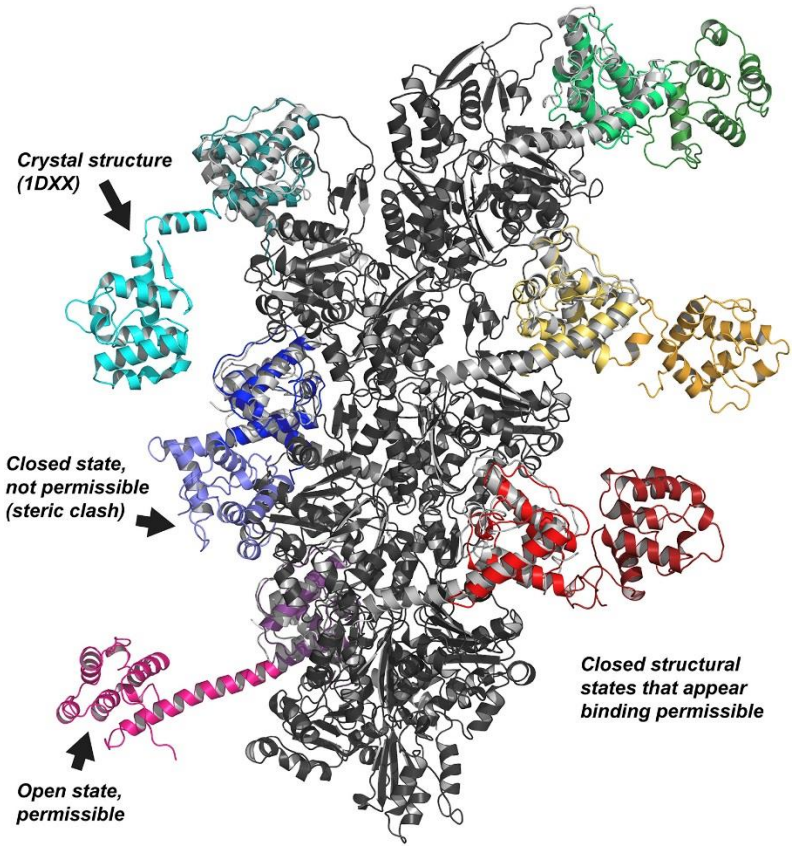

Figure 3.5. Modeling of MD simulation-derived Dys ABD1 conformers on actin filament. When the CH1 domain of Dys ABD1 conformers in free energy minima of Fig. 3.3 are aligned with the CH1 domain of β-III-spectrin ABD from a recent 6.9 Å cyro-EM structure (6ANU), some of the structural models are devoid of steric clashes (green, yellow, red). The open conformation (magenta) is similarly free of steric clashes. While some closed conformations of Dys ABD1 exhibit significant steric clashes with actin (blue), the fact that others do not suggests there are binding-compatible closed states for Dys ABD1 in agreement with previous structural measurements using pyrene excimer fluorescence.[100] To see figure in color go on-line.

3.5 – Discussion

The diverse biophysical methods employed in this study were geared towards identifying structural differences in Dys and Utr ABD1-actin complexes and understanding how they, in turn, may contribute to differences in function. We found that, similar to Utr, the Dys ABD1 can exist in two main structural states: a compact state

where the CH domains are in close apposition and a more extended open state that is stabilized by actin. Unlike our previous study of Utr ABD1[37], Dys ABD1 did not shift completely to its extended open state; a significant mole fraction of ABD1 continued to sample compact states even with an excess actin concentration (Fig. 3.2). Moreover, the Dys ABD1-actin complex is much more structurally heterogeneous than Utr ABD1. This indicates that the two protein domains, though homologous, may functionally diverge in part due to distinct structural dynamics.

Support for the distinct structural dynamics come not only from DEER (Fig. 3.2) where the full-width half max of Dys ABD1 distance distributions far exceed that of Utr ABD1, but also MD simulations (Fig. 3.3 and Fig. 3.4). In the latter case, for instance, an overlay of structural states derived from Fig. 3.3 indicates that Dys and Utr ABD1 are sampling distinct conformational spaces. If the CH1 domains of each ABD1 are aligned in pairwise fashion and the positions of the linker and CH2 are examined, not a single pair of structural states from the free energy minima orient exactly the same way (Fig. 3.17). Although we observed several general similarities, even in the most favorable of cases the orientation of the two CH domains were still distinct such that no structure for the Dys ABD1 could be mapped onto a structure for Utr ABD1.

When considering how these distinct ABD1 structural dynamics could impact function, two possibilities come to mind: alteration of ABD1 affinity for actin and alteration of actin structural dynamics when bound to filaments. With regard to modulation of affinity, recent work on Dys and Utr ABD1 indicate that actin binding is largely mediated by the CH1 domain[102, 123] as has been suggested in other binding studies of ABDs or through inference from CH2 domain mutations that increase affinity for

actin.[60, 103] We used this evidence to justify aligning our simulation-derived Dys ABD1 structures with that of the CH1 domain of β-III-spectrin in a recently reported cryo-EM structure and assessed if any of the stable conformers were capable of docking without steric hindrance (Fig. 3.5).[103] From this analysis, we found that docking the open extended conformers from the MD simulations is compatible with binding because the linker and CH2 domain protrude and point away from actin, avoiding any potential steric clash. Steric clash was one of the main reasons originally proposed for ABDs needing to open upon binding.[36] We also found, however, that while many of the closed compact conformers result in steric clashes between the CH2 domain and actin, at least some compact conformations for either Dys or Utr are compatible with binding to actin (structure B in Fig. 3.3 A for Dys ABD1, and structures B, D, and E in Fig. 3.3 B for Utr ABD1). This is in agreement with the recent work of Shams et al.[124], which suggests that various binding modes between ABD1 and actin could be accessible. Therefore, the conformational ensembles computationally characterized could explain how both compact and extended conformations of ABD1 are sampled when bound to actin.

The notion that Dys and Utr ABD1 closed structural states have high degeneracy and that some are potentially "binding competent" whereas others are not suggest that ABD1's structural dynamics could directly impact affinity for actin. If, for example, the most stable closed conformers of ABD1 are those that give raise to steric clashes when binding to actin, shifting the equilibrium to a binding competent conformer of higher free energy will result in a reduced binding affinity. Alternatively, if multiple closed conformers are free of steric clashes and thus binding competent, affinity for actin would

be increased. In this way, the shifting of conformational equilibria in the structurally dynamic unbound state could play a role in fine-tuning affinity for actin.

It is important to point out that the energetics at play in such a binding model are compatible with existing measurements of ABD1 affinity. In Dys ABD1, for instance, the ΔG of actin binding derived from reported K_d that range from 10-50 μM[123, 125] is an energetically favorable 6-7 kcal/mole. Our calculations show that the energy difference between the various closed and open ABD1 conformers should be smaller than this range (Fig. 3.3) and thus be capable of an actin-induced redistribution of the structurally dynamic ensemble. Unfortunately, the computational precision of the free energy profile constructed from our unbiased simulations does not allow any further extrapolation. While there appears to be more than one conformer in the ABD1 ensembles that is potentially binding competent (Fig. 3.5), in the absence of simulation data that includes actin filaments, no conclusions can be drawn about which conformation is the most stable when bound. As such, it is not possible to directly explain the differences in the experimentally detected spin label distances in Dys and Utr (Fig. 3.2) as due to the binding of one specific more stable conformer identified by MD simulations. However, DEER measurements indicate that it is primarily an open conformation for Utr and a heterogeneous mixture for Dys. Future work could potentially probe this structural heterogeneity further using pressure perturbation EPR methods.[126]

With regard to how Dys and Utr differentially alter actin dynamics, the structural behavior measured by DEER provides insights into a mechanistic contribution (Fig. 3.2). When full-length Dys and Utr bind actin, they each restrict the amplitude of bending and twisting motions of the filament, although Utr does so to a much greater degree. Part of

the difference is probably explained by binding of each protein's ABD2[99]. Each ABD2 varies in its proximity to ABD1[26, 27]. However, another contributing factor may be distinct structural dynamics within each protein complex. Because actin itself can adopt many conformations upon twisting and bending[127], binding of ABDs in general may impose on the filament's conformational exploration. If bound ABDs differ, however, in their ability to conformationally adapt to actin fluctuations, they may consequently have distinct restrictive capacities. If these dynamic differences persist in their full-length structures, they may locally contribute to the overall global differences measured previously.[98, 99] Additionally, if bound ABDs differ in the degree to which they can adapt to structural changes of actin, distinct ABD dynamics may also contribute to ABD's ability to remain bound under mechanical stress experienced by the filament. The fact that the computationally characterized ensembles of Dys and Utr ABD1 have similar general features, but differ in the specific orientation of the CH domains, supports this as a viable adaptation mechanism.

Overall, the results presented in this study agree well with previous work on Dys ABD1 and other ABDs from the spectrin superfamily and simultaneously provide new insights. In a prior study examining the question of whether or not the Dys ABD1 opens upon binding, for instance, pyrene excimer fluorescence data indicated that the ABD1 could bind in a closed conformation.[100] While these results could not determine the distribution of conformational states when bound to actin, they are in agreement with our current findings. We showed in Fig. 3.2 that even in a 4:1 excess of actin, about 50% of the ABD1 is bound in a compact conformation. However, what our results show in addition is that the remaining ~50% of the bound population is present in an extended

conformation, making our findings complementary to that of Singh et al.[100]. We have already pointed out that different experimental techniques can capture ABD domains in both compact and extended conformations. Such differences, when considered in conjunction with a flat free energy landscape of easily interconvertible conformers, suggest that ABDs are malleable proteins responsive to small external perturbations. A perturbation on the order of just 1 kcal/mole could shift the equilibrium from one conformation to another (Fig. 3.3).

This proposed mechanism may not only contribute to our understanding of ABDs being found in a variety of conformations, but also the impact of disease-causing mutations. If the conformational ensembles of ABDs are easily perturbed, defects such as missense mutations could significantly alter the distribution of structural states. If such mutations redistributed the conformational ensemble in a way that more readily populates binding competent conformers, for instance through the mechanism discussed above, it could help explain the modest to significant gains in affinity observed in ABDs associated with disease.[39, 42] In addition, mutation-induced redistribution may alter the conformational ensemble in such a way that there is a gain conformational adaptability that allow binding to persist despite structural fluctuations of the filament.

Application of this generalized ABD model to DMD and BMD mutations in Dys requires some caution. Many DMD and BMD missense mutations that have been studied previously are mutations that have substituted CH domain core hydrophobic residues for charged or polar residues.[33] When such mutations were studied in the isolated Dys ABD1 fragment, the main affect seemed to be misfolding of the domain rather than change in binding affinity.[128] Indeed, when many of the same missense mutations were examined in

full-length Dys, changes in binding affinity were modest or statistically insignificant.[129] This suggests that for DMD and BMD, missense mutations that perturb the core fold of ABD1 CH domains are promoting disease via a misfolding-induced degradation mechanism.[130] For some missense mutations, perhaps those that are more superficially located within the ABD1 or those that do not misfold the ABD1 but perturb stability, a functional impact like that discussed above for ABDs in general may become relevant. However, without clear examples of DMD and BMD missense mutations that alter binding affinity while simultaneously preserving the ABD1 fold, this potential mechanism will remain speculative and bare most relevance to other ABDs in which missense mutation stability-structure-affinity relationships are better established.[103]

One of the observations made here points to an interesting feature of Dys ABD1 that will probably require further investigation in future studies. While DSC measurement of heat capacity helped validate the MD-suggested use of surface hydrophobicity in modulating closed conformer sampling, it also revealed a rather complex endothermic transition (Fig. 3.18). The absolute heat capacity of a protein is a measure of underlying structural diversity in the folded and unfolded states, but can also be used to evaluate the presence of intermediate states. It is not known whether the distribution of conformers identified for Dys ABD1 reflect a distribution of equivalent distinct states or arise as a distribution or sampling of partially folded intermediate states. We find that the diversity of open and closed conformers corresponds energetically to a two-state model comprised of a folded and unfolded state. The unfolded state retains a surprising amount of order, consistent with retaining hydrophobic contacts even under extreme conditions. This suggests that the energetic functional landscape of Dys ABD1 may be capable of

absorbing much more energy than anticipated from the distribution of open and closed conformers identified. However, further quantitative assessment of such a thermodynamic prediction will require application of additional biophysical techniques.

3.6 – Conclusions

We have probed Dys ABD1 dynamics with complementary experimental and computational techniques. We have characterized free ABD1 in solution as adopting a compact conformation. With computer simulations we have shown that both Dys and Utr ABD1 are conformationally malleable proteins for which low-energy compact and extended conformations are easily accessible through a flat free energy landscape. We have identified the region of the linker that grants conformational flexibility to the protein and the hydrophobic patches responsible for stabilizing the compact conformations. We have characterized the conformational ensemble of actin-bound Dys ABD1 as originating from an equal contribution of compact and extended structures. Finally, we have linked our findings and proposed a model for how unique structural dynamics of Dys and Utr ABD1 contribute to their distinct functions.

3.7 – Supporting Materials and Methods

3.7.1 – Multifaceted approach to DSC analysis

MD simulations revealed that surface exposed hydrophobic residues drive the open to closed conformational fluctuations of ABD1. The extent of hydrophobic and hydrophilic exposure is directly proportional to the change in heat capacity.[131] Differential scanning calorimetry (DSC) is a direct means to measure the change in heat capacity and relate it to the solvent exposed surface area.[131-137] DSC analysis also provided a free energy of unfolding of 4.45 ± 0.07 kcal/mol for DysABD1. This energy sets an upper limit to the free energy difference between different folded conformations, as higher energies would start depleting the folded population in favor of the unfolded. Comparing the experimental to the maximal change in heat capacity, we determined the percent unfolding of Dys ABD1 to be 87% or, conversely, that 13% residual structure remains in the unfolded state. Because simulations have shown that surface exposed hydrophobic residues are present in Dys ABD1 and play a key role in defining its tertiary structure, it is consistent to think that the structure of this protein is more resilient to unfolding. That is, because the folded structure is relatively stable with exposed hydrophobic residues, the thermal unfolding, which coincides with the exposure of core hydrophobic residues, may bring a lesser perturbation to the structure than in other proteins that have lower content of exposed hydrophobic residues. As a result, a large portion of residual structure is then likely resultant of hydrophobic contacts recapitulating the role of hydrophobic contacts as a driving force for Dys ABD1 reconfiguration.

MD revealed the Dys ABD1 ensemble of closed conformations is stabilized by hydrophobic interactions involving mainly the CH2 domain and the flexible linker. We

tested whether Dys ABD1 is enriched in hydrophobic residues that affect the change in heat capacity between the folded and unfolded states when compared to other proteins. Specifically, calorimetric values were compared to predictive models relating changes in solvent exposed surface areas to changes in heat capacity using DSC. The last three columns of Table 3.1 contain thermodynamic values obtained by fitting our DSC data using three different methods, and in the first two columns are theoretical predictions obtained using models from Spolar[131] and Freire[132]. Although each fitting method is discussed in detail in the following sections of the supplement, it is relevant to point out here that the Spolar[131] method predicts the change in heat capacity from the molecular weight of the protein, whereas the Freire[132] model utilizes the protein amino acid sequence.

Table 3.1 shows that the change in heat capacity of Dys ABD1 upon unfolding fitted from DSC data is greater than theoretical prediction based on molecular weight using the method outlined by Spolar[131], but it is smaller than the heat capacity predicted by Freire's model that includes the protein's sequence. These deviations are consistent with the high concentration of aromatic residues (~3 times expected for a protein of this size) while simultaneously having some of the aromatic residues already solvent exposed in the 1DXX crystal structure reported by PDB.

3.7.2 – Text S1, Predicting ΔC_p and ΔH_{Tm} from molecular weight and sequence

Sequence based change in heat capacity prediction

In a two-state unfolding process, the initial state is the folded state (or ensemble of states) and the final state is the unfolded state. The unfolded state is relatively, but not

completely, devoid of higher order structure. Thus, it is necessary to estimate the residual structure in the unfolded state to characterize the unfolding process. Using models for determining the absolute heat capacity of previously studied proteins, the maximum change in heat capacity for Dys ABD1 was compared to determine residual structure from the experimentally determined change in heat capacity.[132] The maximum change in heat capacity was predicted to be 5.9 kcal/mol/°C.[132-135] This overestimation reflects that hydration effect does not take the role of configurational entropy of the unfolded protein into consideration, as has been noted in reference[138]. Another oversimplification is that the geometry of Dys ABD1 is not globular.[132] Adjustments were made to the absolute folded heat capacity based on the change in surface areas of a sphere (approximating globular geometry) and two smaller spheres of equal volume (approximating Dys ABD1 geometry). This adjustment resulted in a predicted change in heat capacity of 4.7 kcal/mol/°C. Then, by normalizing the experimental change in heat capacity to the theoretical maximum change in heat capacity we can determine the role of configurational entropy in the unfolded state. Analysis yielded 87% unfolding for Dys ABD1 and is calculated by the following equation.

$$unfolded = \frac{\Delta C_{p,experimental}(T)}{C_{p,u}(T) - C_p(T)} * 100 \qquad \text{Eq. 3.4}$$

Where $C_{p,u}(T)$ and $C_{p,f}(T)$ were determined using previously established models. The heat capacity of any amino acid residue or backbone species (x) has been estimated as a third order polynomial.[132]

$$C_{p,x} = A_x + B_x T + C_x T^2 + D_x T^3 \quad \text{Eq. 3.5}$$

The total heat capacity of the unfolded state will then be the additive heat capacities of all components

$$C_{p,u}(T) = \sum_{x=1}^{20} n_x C_{p,x} + (N-1)C_{p,bb} + C_{p,NH_2} + C_{p,COOH} \quad \text{Eq. 3.6}$$

$$C_{p,f}(T) = (a + bT)M_W \quad \text{Eq. 3.7}$$

Where the summation from 1 to 20 represents a summation over the 20 amino acid residue types, $C_{p,bb}$ represents the heat capacity of the backbone species, and the last two terms in equation 3.29 account for the N- and C- termini, respectively. All coefficients A, B, C, D, a, and b have been determined and tabulated.[132]

Predictions based on Molecular Weight

The overall change in heat capacity (ΔC_p) was estimated from molecular weight, utilizing the fact that the change in hydrophobic and hydrophilic surface areas dictate changes in heat capacity and are linearly dependent on molecular weight for proteins spanning 4 kDa to 40 kDa given by the set of equations listed below[131]:

$$\Delta A_{NP} = 0.636 M_W - 2120 \text{Å}^2 \quad \text{Eq. 3.8}$$

$$\Delta A_P = 0.309 M_W - 40 \text{Å}^2 \quad \text{Eq. 3.9}$$

$$\Delta C_{P,Calc} = -0.32 \Delta A_{NP} + 0.14 \Delta A_P \, cal mol^{-1} K^{-1} \quad \text{Eq. 3.10}$$

Where equations 3.8 and 3.9 calculate the expected change in exposed nonpolar and polar surfaces, respectively and are used in equation 3.10 to calculate the resulting change in heat capacity between folded and unfolded states. For Dys ABD1 which has a molecular weight of 27.355 kDa (determined via ExPASy ProtParam), the change in heat capacity was predicted to be 3.7 ± 0.7 (kcal/mol°C). Because this model is based on globular proteins, the values reported are interpreted as an upper estimate for the total change in heat capacity of a protein, because globular proteins are mostly spherical and therefore have the highest volume to surface area ratio. However, since the prediction based off of molecular weights assumes a normal distribution of amino acids (whereas, Dys ABD1 has ~3x higher content of aromatic residues) Dys ABD1 is likely to deviate from this value.

The enthalpy of unfolding was estimated with respect to the same work noting that enthalpy was similarly proportional to molecular weight and temperature given the following equation[131],

$$\Delta H^{\circ}_{fold}(T) = (-0.161TM_W) + 47.24M_W + 672.8T - 232010calg^{-1} \qquad \text{Eq. 3.11}$$

where, T is the transition temperature of the protein, and M_W is the molecular weight of the protein. For Dys ABD1 the model yielded a value of 118.5 ± 0.4 kcal/mol at 55°C. Once again, the enthalpy predicted serves as an upper limit, as the predictive model was developed largely off globular proteins, whereas Dys ABD1 is non-globular.

3.7.3 – Text S2, Calorimetric features of the experimental endotherm

To further narrow down the parameter space for the Gibbs-Helmholtz equation to model heat capacity it can be seen that the three necessary variables (ΔH_{Tm}, ΔCp, and T_m) all manifest in distinct parts of the endotherm (Fig. 3.18A). The enthalpy of transition is contained within the large spike in heat capacity represented by the shaded red region in the plot. The region has an upper boundary described by the raw instrumental readout, whereas the lower boundary is described by the shift in populations from the folded state to the unfolded state, which is represented by the sigmoidal baseline (i.e. the shift in equilibrium). Comparison of transition enthalpy between the first and second scans on the same sample allow for the determination of reversibility, which was 15%. The sigmoidal baseline is representative of the population shift from folded to unfolded ($\frac{d\theta_u}{dT}$) and when scaled by the change in heat capacity between the two states it gives the lower bound for the excess heat capacity. Then, the temperature at which the enthalpy equals half the complete excess enthalpy of transition (ΔH_{Tm}) is the melting point (T_m). This also corresponds to the point at which the baseline change in heat capacity takes its half-maximal value.

The change in heat capacity (ΔC_p)

After baseline subtraction, the baselines should be parallel to the x-axis. The folded baseline was extrapolated beyond the transition to the same region used for the unfolded baseline definition. The difference in heat capacity is then the change in heat capacity (ΔC_p) given by the following equation:

$$\Delta C_p(T) = C_{p,u}(T) - C_{p,f}(T) \quad \text{Eq. 3.12}$$

where the heat capacities on the right hand side of the equation are the heat capacities after baseline subtraction, thus $C_{p,f}(T) = 0$ by definition.

The excess transition enthalpy (ΔH_{Tm})

The excess enthalpy of transition was analyzed utilizing the approach developed by Freire.[132] The Enthalpy of transition (Fig. 3.18A) represents the area defined by endotherm and the baseline heat capacities of the two states. To determine the sigmoidal baseline, the total change in enthalpy ($\Delta H_{tot}(T)$) was determined using Riemann sums.

$$\Delta H_{tot}(T) = \sum_{T=T_L}^{T} C_{p,data}(T) - C_{p,f}(T) \quad \text{Eq. 3.13}$$

Where $\Delta H_{tot}(T)$ describes the total change in enthalpy due to both the excess heat capacity (ΔH_{Tm}) and the basal heat capacity ($C_{p,basal}$) defined by the lower transition temperature (T_L) and T. The maximal change in enthalpy is the summation from the lower limit to the upper limit of the transition (T_U):

$$\Delta H_{max} = \sum_{T=T_L}^{T_U} C_{p,data}(T) - C_{p,f}(T) \quad \text{Eq. 3.14}$$

Normalizing $\Delta H_{tot}(T)$ to the total change in ΔC_p (3.5 ± 0.3 kcal/mol°C) yields the following equation for approximating the basal heat capacity ($C_{p,basal}$) of the protein as a function of temperature:

$$C_{p,basal}(T) = \frac{\Delta H_{tot}(T)}{\Delta H_{max}} * \Delta C_p$$ Eq. 3.15

Then, the excess enthalpy is the difference between the total change in enthalpy ($\Delta H_{\max}$) and basal change in enthalpy:

$$\Delta H_{Tm} = \Delta H_{max} - \Delta H_{basal}$$ Eq. 3.16

Which can be solved with the above described parameters as an area integral bounded by the experimental heat capacity and the basal heat capacity:

$$\Delta H_{Tm} = \int_{T_L}^{T_U} C_{p,tot}(T) - C_{p,basal}(T) dT$$ Eq. 3.17

The sigmoidal baseline and ΔH_{Tm} will vary minimally on the definition of the temperature bounds (T_U and T_L).

Melting temperature (T_m)

The melting temperature is the point along the transition at which the distribution between folded and unfolded populations is 1:1. Therefore, this lies at the temperature for which:

$$\Delta H(T_m) = \frac{\Delta H_{Tm}}{2}$$ Eq. 3.18

Where

$$\frac{\Delta H_{Tm}}{2} = \int_{T_U}^{T_m} (C_{p,tot}(T) - C_{p,basal}(T))dT \quad \text{Eq. 3.19}$$

And can be solved directly by finding the solution to the following equation:

$$\Delta H_{Tm} = 2(\Delta H_{tot}(T_m) - \Delta H_{basal}(T_m)) \quad \text{Eq. 3.20}$$

Where T_m will also correspond to the half max of $\Delta H_{basal}(T)$.

3.7.4 – Text S3, Nonlinear least squares regression

The values from analysis of the calorimetric features (Text S2) and constraints from predictive models (Text S1), along with a comparison to the Van't Hoff fit strictly describe where in the three-dimensional parameter space to fit to the Gibbs-Helmholtz model. From this, the nonlinear regression analysis was performed with equation 3.31 and run in R using a Gauss-Newton algorithm to minimize the squared residuals. This yielded the following model fit (Fig. 3.18B). The next sections describe how the fitting is performed.

3.7.5 – Two-state model analysis of heat capacity

As a protein denatures in solution, thereby exposing hydrophobic residues buried within the interior of the protein, the heat capacity ($C_p(T)$) of the sample cell increases according to the following expression:

$$Cp(T) = \theta_u H(T) \frac{d}{dT} = (\theta_u) \frac{d}{dT} H(T) + \theta_u (H(T)) \frac{d}{dT} \quad \text{Eq. 3.21}$$

For Dys ABD1, the simplest model that can be applied is a two-state model from the native/folded (θ_f) state to denatured/unfolded (θ_u) state as Dys ABD1 underwent thermal denaturation. This means that the occupancy of intermediate states is negligible, therefore simplifying an equilibrium expression to two discrete ensembles:

$$K = \frac{[u]}{[f]} \quad \text{Eq. 3.22}$$

$$K = e^{\left(\frac{-\Delta G}{RT}\right)} \quad \text{Eq. 3.23}$$

Where [u] and [f] are concentrations of folded and unfolded ensembles and K is the equilibrium constant between the two states. Because the entire population, [P], of the protein is distributed between the unfolded and folded state the total protein concentration and fraction (θ) of each state can be written as follows:

$$[P] = [u] + [f] \quad \text{Eq. 3.24}$$

$$\theta_u = \frac{K}{K+1} \quad \text{Eq. 3.25}$$

$$\theta_N = \frac{1}{K+1} \quad \text{Eq. 3.26}$$

The free energy may be represented as both a function of the temperature dependent enthalpic and entropic contributions (ΔH(T) and ΔS(T), respectively) and as a consequence of the two-state model the change in heat capacity (ΔC_p) simplifies to a discrete value as shown in equation 3.30[132, 135]:

$$\Delta H(T) = \Delta H_{Tm} + \int_{Tm}^{T} C_p(T)dT \quad \text{Eq. 3.27}$$

$$\Delta G(T) = -RTln(K) \quad \text{Eq. 3.28}$$

$$\Delta G(T) = \Delta H(T) - T\Delta S(T) \quad \text{Eq. 3.29}$$

$$\Delta G(T) = \Delta H_{Tm}(1 - \frac{T}{T_m}) + \Delta C_p(T - T_m - T\left(\ln\left(\frac{T}{T_m}\right)\right)) \quad \text{Eq. 3.30}$$

Where the limit as ΔC_p approaches 0 is the Van't Hoff equation for free energy. By substituting equations 3.30 and 3.23 into equation 3.21 and differentiating with respect to temperature the final fit expression of heat capacity as a function of temperature is as follows:

$$C_p(T) = \Delta H(T)\frac{(\Delta Cp - \frac{\Delta CpTm}{T} + \frac{\Delta H}{T})\frac{K}{RT}}{(K+1)^2} + \frac{\Delta C_p K}{K+1} + FBL \quad \text{Eq. 3.31}$$

Where FBL (folded baseline) describes the initial baseline heat capacity of the folded state, which we model using either a first or second order polynomial. The difference between the endotherm and the baseline heat capacity is equivalent to the instantaneous rate of change of the fraction folded (or unfolded) with respect to temperature $\frac{d\theta}{dT}$. Therefore, the two-state equilibrium can be modeled by three thermodynamic parameters: the excess enthalpy needed to transition a population from a folded to an unfolded state

(ΔH_{Tm}), the overall change in heat capacity between the folded and unfolded states (ΔC_p), and the melting temperature (T_m). For a two-state transition model, the largest rate of change occurred at the melting temperature ($\frac{d\theta}{dT_m}$), where both θ_f and θ_u equal 0.5. The maximal rate of change of folded proteins relates to the overall enthalpy of the unfolding process in that,

$$\Delta H_{VH} = 4RT_m^2(\frac{d\theta_u}{dT}) \quad \text{Eq. 3.32}$$

where ΔH_{VH} is the Van't Hoff enthalpy. Thus, through comparison of the Van't Hoff, it is possible to determine the validity of a two-state model.[135] For Dys ABD1, the Van't Hoff enthalpy (110 kcal/mol) was not extremely different from the enthalpy determined via a nonlinear regression, NLS, model (118.5 ± 0.4 kcal/mol) described later in the supplement, thus supporting the use of a two-state model. However, the Van't Hoff comparison is not sufficient for defining a two-state model. Further support for the two-state model comes from the comparison of the changes in heat capacity calculated between the different folded conformers of ABD1, which varied by approximately 0.05 kcal/mol°C between open conformers and approximately 0.1 kcal/mol°C between open and closed ensembles, which MD simulations characterized as having free energy differences smaller than a couple kcal/mol. Finally, error surface and confidence region mapping were used to estimate nonlinear confidence regions to address the uncertainty in the folded baseline as has been noted in protein folding energetics studied by DSC.[139, 140]

3.7.6 – Text S4, Disorder Prediction Modeling

In order to extend our results to WT Dys ABD1, we first needed to assess the role of cloning material to the folding energetics. This was done using predictive models to determine how ordering within the construct would be affected by cloning material left behind after purification. Because the construct was separated from the 6x His tag using a TEV protease, it is expected that two extra residues, glycine and serine, were left over after TEV cleavage. Disorder calculations were compared between the WT sequence and the construct containing the extra two amino acids.[141, 142] There is minimal predicted change in disorder with the addition of the two extra amino acids (Fig. 3.19).

3.7.7 – Analysis of the Molecular Dynamics Trajectories

In our simulations, two versions of the CHARMM force field were used because evidence shows that the standard CHARMM36 force field may not perform as well as the CHARMM22* force field when simulating intrinsically disordered protein.[143] Although ABD1 is not an IDP, our calorimetric measurements indicated that it may be considered marginally stable. We chose to use the CHARMM22* force field to be able to capture any region of the protein that showed marginal stability. At the same time, we recognize that the CHARMM22* is a more specialized force field, and thus has not been field-tested as extensively as CHARMM36. We performed additional analyses with the goal of testing whether significant differences in the secondary structure, backbone dynamics, and conformational space sampled arise between the two force fields.

To evaluate differences in secondary structure sampled by the two force fields, Figure 3.8 shows a comparison of the secondary structure elements calculated with the DSSP algorithm.[144] The figure shows that, for both Dys and Utr, in the flexible regions

CHARMM36 displays a higher α-helical content than CHARMM22*, which in turn displays a higher component of turn, bend, and coil.

To evaluate difference in the backbone dynamics, Figure 3.9 shows a comparison of the backbone φ/ψ dihedral order parameter calculated accordingly to the method of van der Spoel and Berendsen.[145] Dihedral angles autocorrelation functions were calculated over each individual 500 ns trajectory, and the mean of the φ/ψ order parameters was averaged over all 5 trajectories for a given system and force field. For each amino acid, t-test statistics was performed between the 2 sets of 5 trajectories to evaluate whether the observed backbone order parameters were significantly different between the two force fields. In Figure 3.9, all amino acids that resulted in a p-value lower than 0.05 are marked with a black dot. We notice that the majority of the amino acids showing a statistically significant difference are those belonging to ordered regions, for which a negligible absolute difference in the order parameter between the two force fields is present. The less ordered regions show larger differences between the two force fields, but the larger variability is present also among trajectories of the same force field. As a result, from these analyses it is not possible to establish whether the disordered regions are characterized by a statistically significant difference in dynamics between the two force fields.

As an alternative approach to using the backbone order parameter to map the conformational flexibility of different regions of the protein we use the Shannon entropy as described in the main text. Figure 3.10 shows a comparison of the Shannon Entropy for the two force fields, which suggests that the flexible linker region could experience a higher conformational dynamics in the CHARMM22* simulations.

The previous analyses showed some differences between the CHARMM36 and CHARMM22* trajectories both in terms of secondary structure and in terms of backbone conformational dynamics. To test whether these differences significantly affect the conformational space sampled by the trajectories, we used the root mean square deviation (RMSD) as a metric to test the difference between the various trajectories. In Figure 3.11 (panels A and C) we have plotted the RMSD matrices between all the trajectories. In short, each frame of each trajectory is here compared (by calculating the RMSD) with each frame of each trajectory. The bottom-left quadrant compares the 5 CHARMM36 trajectories with each other, and the top-right quadrant compares the 5 CHARMM22* trajectories. The other two quadrant identically compare the CHARMM36 trajectories with the CHARMM22* ones. The figure shows that for both Dys and Utr there are trajectories within the same force field that are similar to each other (for example Dys c36(2), c36(3), and c36(5)), as well as very dissimilar (for example Dys c36(1) with c36(4)). When comparing trajectories obtained with different force fields, we see a similar distribution of similar and dissimilar trajectories as those within the same force field, which suggests that the two force fields do not greatly differ in the conformational space sampled. To better exemplify this, we plotted (panels B and D) the probability density of the RMSD for each subset (within c36, within c22, and the cross-comparison c36-c22). Again, what we observe is that the distribution of the cross-comparison c36 vs c22 is similar to those within the c36 or within the c22 force fields. Together, these results support the view that the differences between the force fields highlighted by the secondary structure and backbone dynamics analyses do not significantly affect the conformational space sampled by the trajectories. In view of these observations, to

improve the statistics of our principal component analyses, we have pooled together the trajectories obtained with the two force fields.

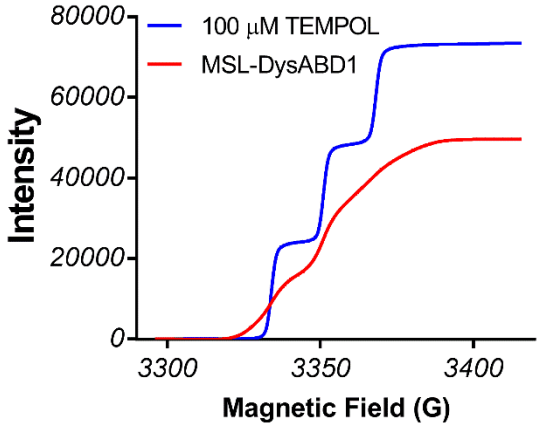

Figure 3.6. Spin labeling Dys ABD1 (120:239) with 4-maleimide-TEMPO. Shown are the double integrals of 100 μM TEMPOL standard (blue) and DysABD1 labeled with MSL (red). This sample was used to determine the spin concentration of Dys ABD1 (120:239), where the ratio of max intensities from the double integral indicate a spin concentration of approximately 67.6 μM. When compared to the Bradford determined protein concentration of 80 μM, this corresponds to a labeling efficiency near 84%.

Dys #									9	10	11	12	13	14	15	16	17	18	19	20
Dys									D	C	Y	E	R	E	D	V	Q	K	K	T
Utr									R	S	D	E	H	N	D	V	Q	K	K	T
Utr #									25	26	27	28	29	30	31	32	33	34	35	36

Dys #	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Dys	F	T	K	W	V	N	A	Q	F	S	K	F	G	K	Q	H	I	E	N	L
Utr	F	T	K	W	I	N	A	R	F	S	K	S	G	K	P	P	I	N	D	M
Utr #	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56

Dys #	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Dys	F	S	D	L	Q	D	G	R	R	L	L	D	L	L	E	G	L	T	G	Q
Utr	F	T	D	L	K	D	G	R	K	L	L	D	L	L	E	G	L	T	G	T
Utr #	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76

Dys #	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Dys	K	L	P	K	E	K	G	S	T	R	V	H	A	L	N	N	V	N	K	A
Utr	S	L	P	K	E	R	G	S	T	R	V	H	A	L	N	N	V	N	R	V
Utr #	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96

Dys #	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
Dys	L	R	V	L	Q	N	N	N	V	D	L	V	N	I	G	S	T	D	I	V
Utr	L	Q	V	L	H	Q	N	N	V	E	L	V	N	I	G	G	T	D	I	V
Utr #	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116

Dys #	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120
Dys	D	G	N	H	K	L	T	L	G	L	I	W	N	I	I	L	H	W	Q	V
Utr	D	G	N	H	K	L	T	L	G	L	L	W	S	I	I	L	H	W	Q	V
Utr #	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136

Dys #	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140
Dys	K	N	V	M	K	N	I	M	A	G	L	Q	Q	T	N	S	E	K	I	L
Utr	K	D	V	M	K	D	V	M	S	D	L	Q	Q	T	N	S	E	K	I	L
Utr #	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156

Dys #	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160
Dys	L	S	W	V	R	Q	S	T	R	N	Y	P	Q	V	N	V	I	N	F	T
Utr	L	S	W	V	R	Q	T	T	R	P	Y	S	Q	V	N	V	L	N	F	T
Utr #	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176

Dys #	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180
Dys	T	S	W	S	D	G	L	A	L	N	A	L	I	H	S	H	R	P	D	L
Utr	T	S	W	T	D	G	L	A	F	N	A	V	L	H	R	H	K	P	D	L
Utr #	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196

Dys #	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
Dys	F	D	W	N	S	V	V	C	Q	Q	S	A	T	Q	R	L	E	H	A	F
Utr	F	S	W	D	K	V	V	K	-	M	S	P	I	E	R	L	E	H	A	F
Utr #	197	198	199	200	201	202	203	204		205	206	207	208	209	210	211	212	213	214	215

Dys #	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220
Dys	N	I	A	R	Y	Q	L	G	I	E	K	L	L	D	P	E	D	V	D	T
Utr	S	K	A	Q	T	Y	L	G	I	E	K	L	L	D	P	E	D	V	A	V
Utr #	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235

Dys #	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240
Dys	T	Y	P	D	K	K	S	I	L	M	Y	I	T	S	L	F	Q	V	L	P
Utr	R	L	P	D	K	K	S	I	I	M	Y	L	T	S	L	F	E	V	L	P
Utr #	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255

Dys #	241	242	243	244	245	246														
Dys	Q	Q	V	S	I	E														
Utr	Q	Q	V	T	I	D														
Utr #	256	257	258	259	260	261														

Figure 3.7. Alignment of Dys and Utr ABD1 sequences. Green, yellow, and red shadings indicate conservative, semi-conservative, and non-conservative mutations. Cell border colors indicate the domains: CH1 (purple), linker (blue), CH2 (green).

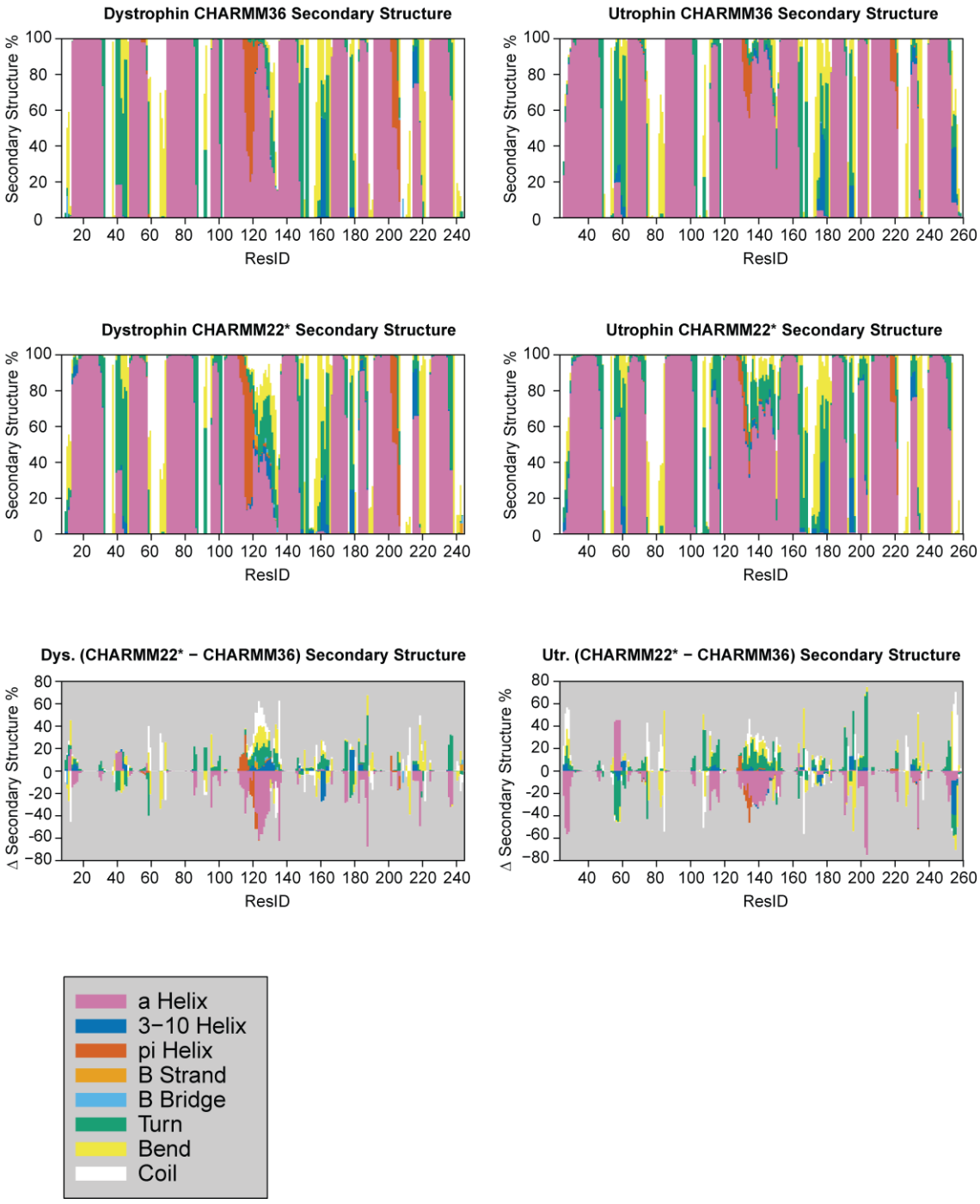

Figure 3.8. Structure comparison between CHARM22* and CHARM36 force fields.

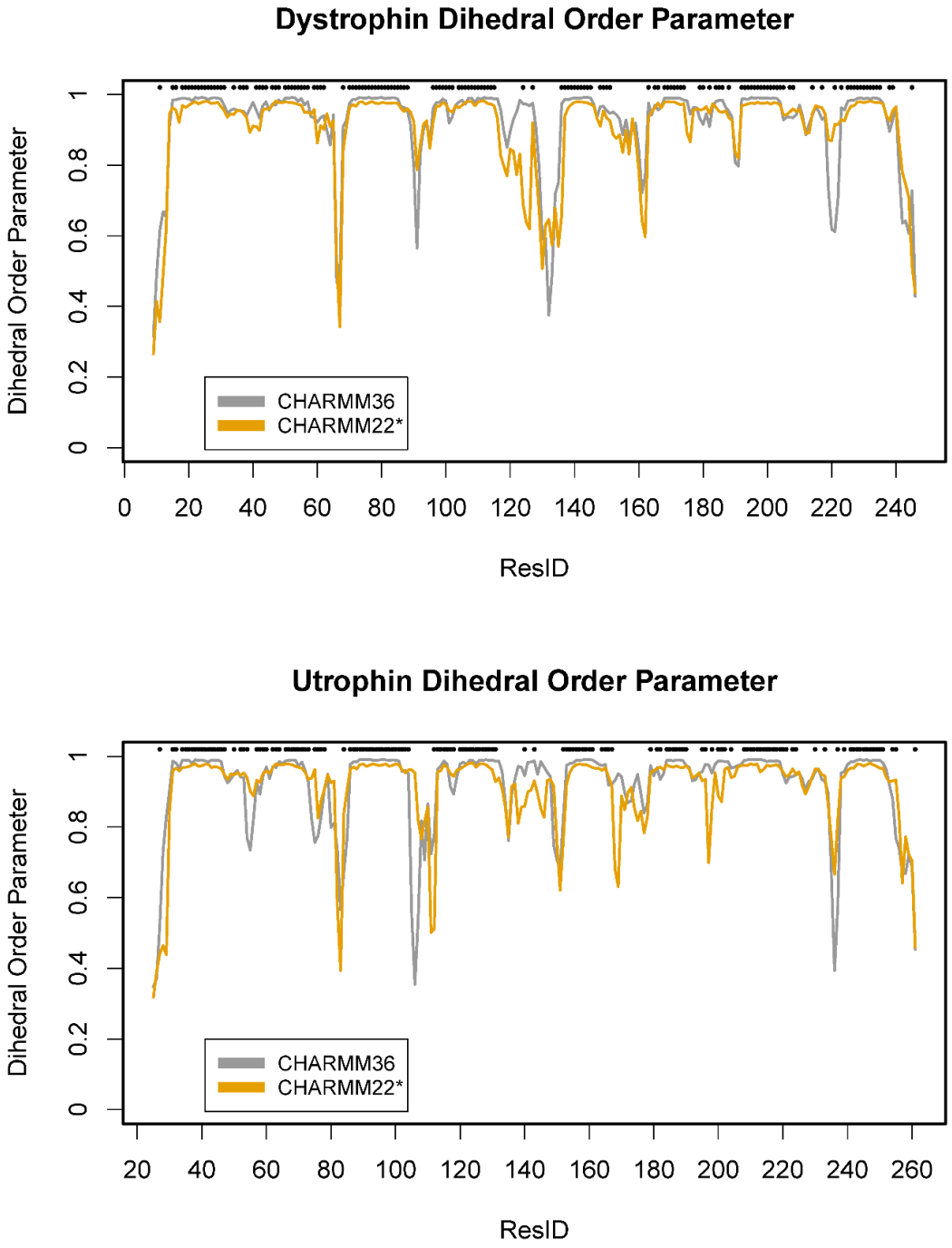

Figure 3.9. Dihedral order parameter. Black dots mark the amino acids for which a t-test between the order parameter of the two force fields resulted in a p-value of less than 0.05.

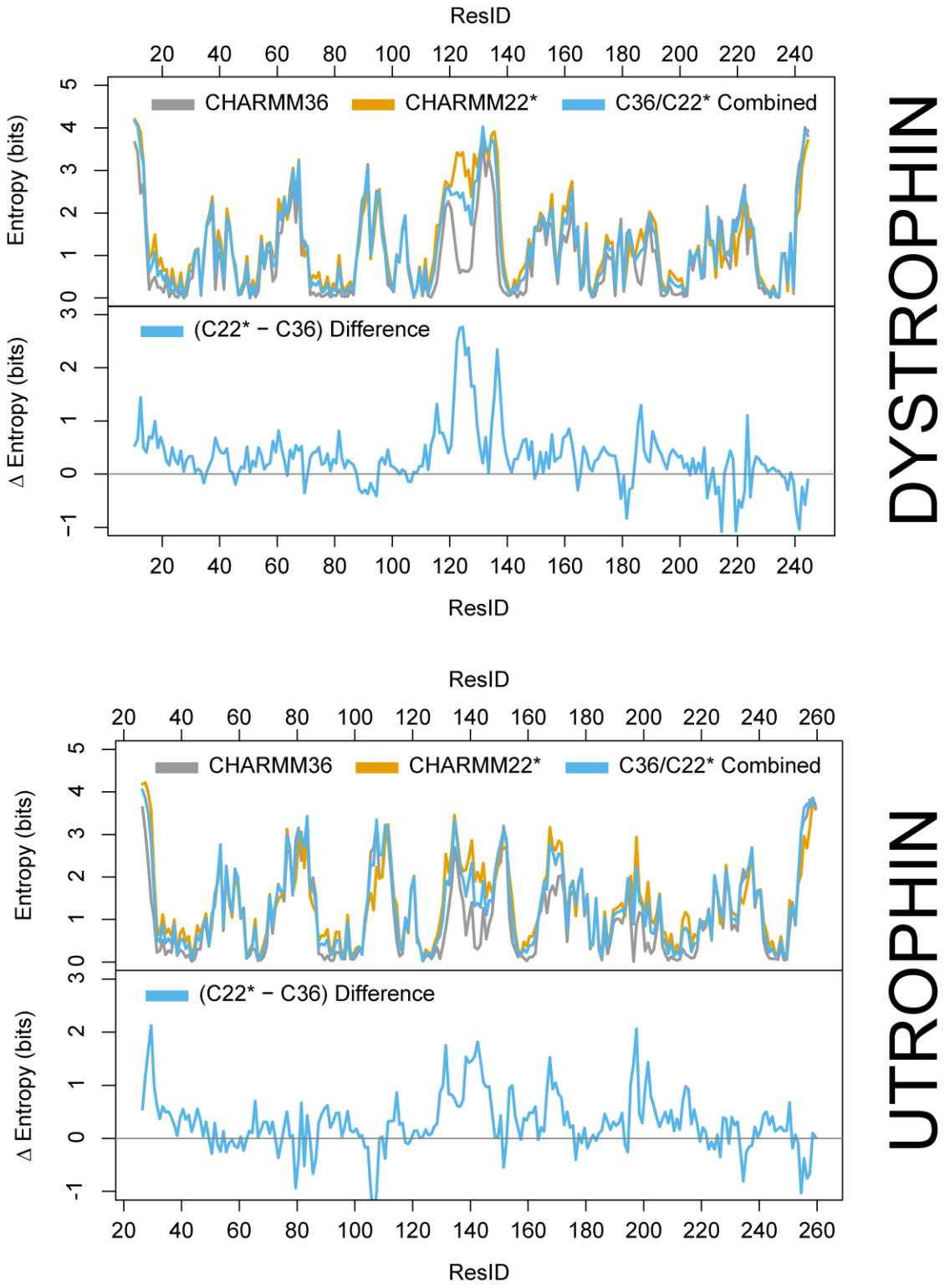

Figure 3.10. Shannon Entropy comparison between force fields.

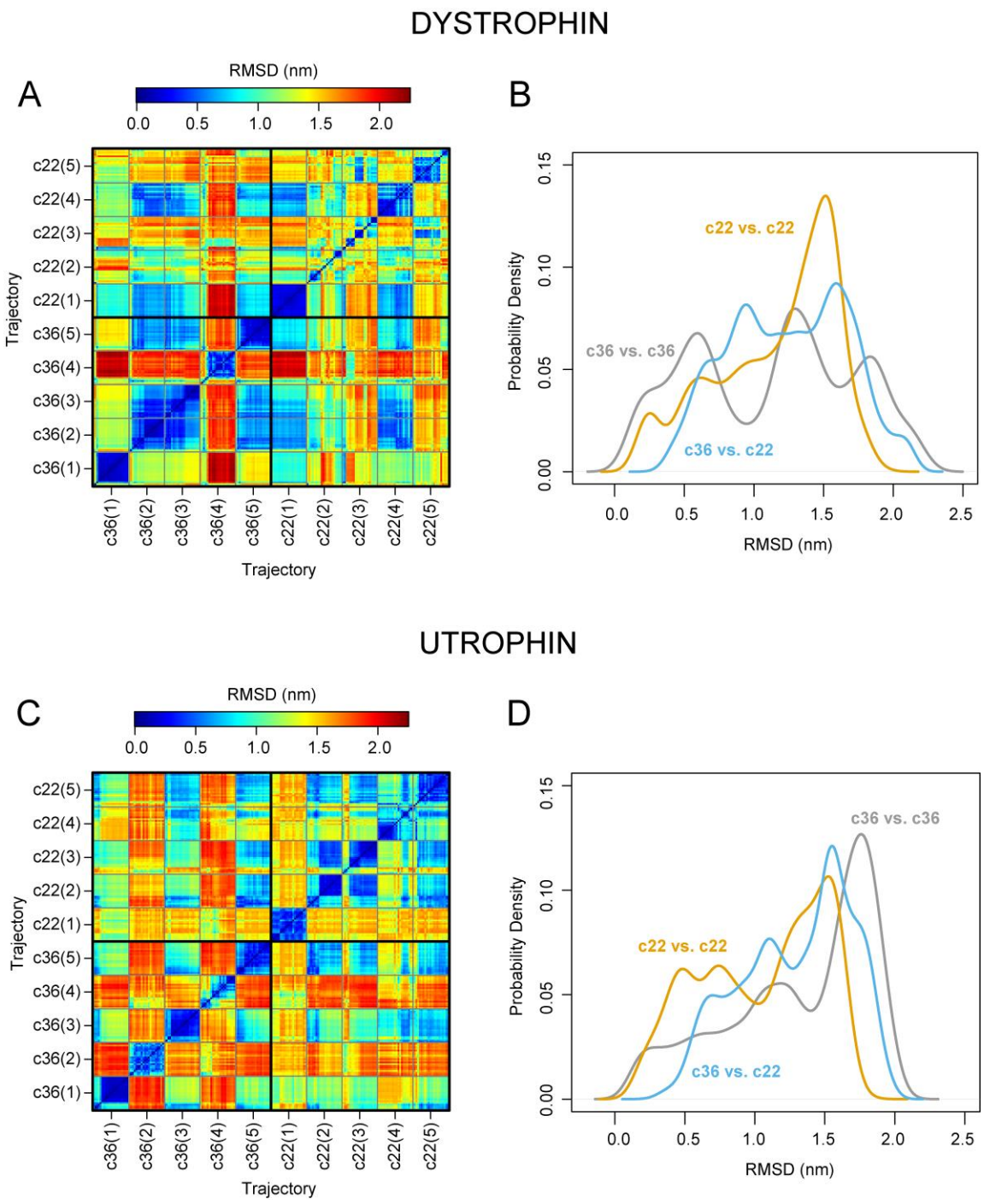

Figure 3.11. (A, C): Pairwise RMSD matrices for the 5 CHARMM36 and the 5 CHARMM22* trajectories. (B, D): Probability density plots of the RMSD from panels (A, C), calculated within a single force field (orange and gray lines), and across force fields (blue lines).

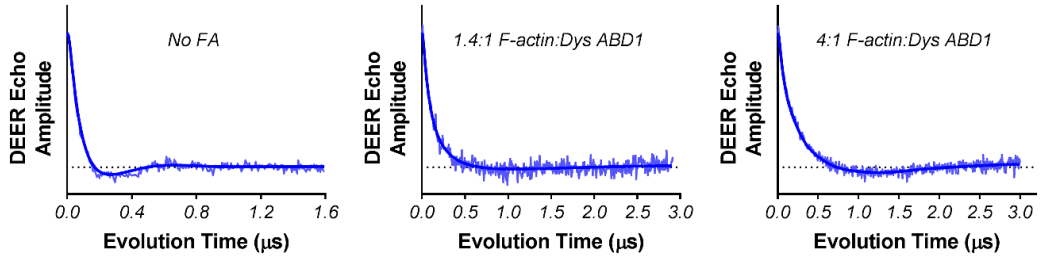

Figure 3.12. Double electron-electron resonance (DEER) on Dys ABD1 (120:239) at various actin concentrations. Shown here are the complete time domains of each background-corrected DEER waveform. The Fig. 3.2 shown in the main text shows the waveforms up to 1.6 μs to help emphasize the changes in waveform decay with increasing actin concentrations.

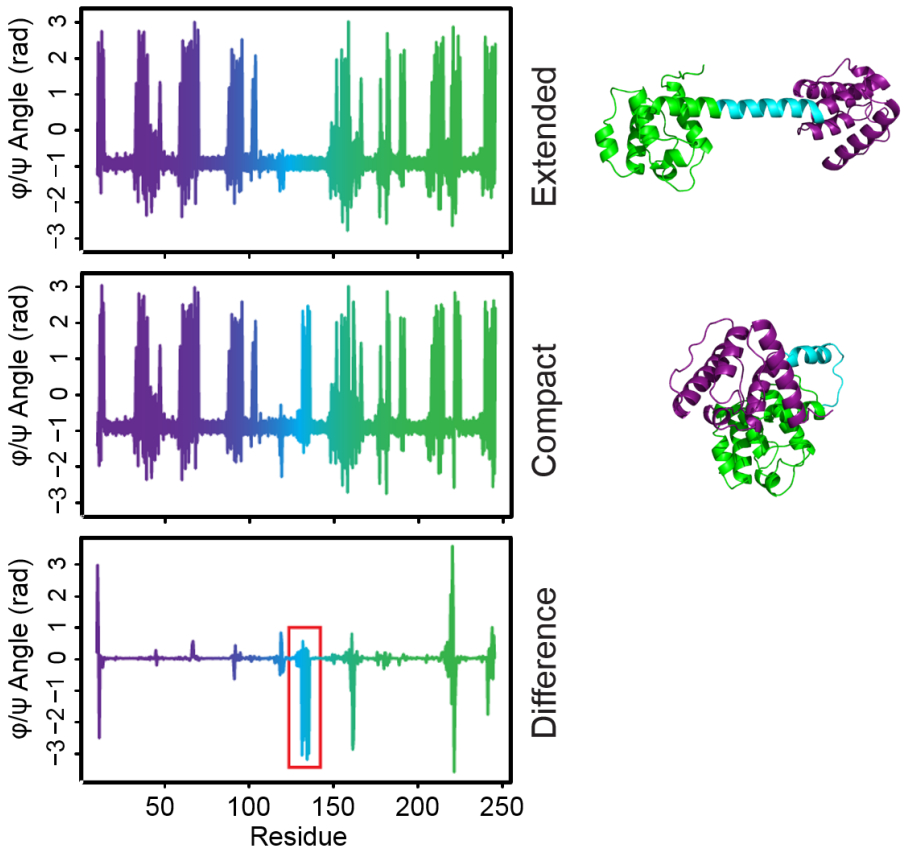

Figure 3.13. Dihedral angle analysis for Dys ABD1 extended state (top panel, conformation A in Figure 3.3) and for the most compact states based on SASA analysis (middle panel, conformation F in Figure 3.3). The bottom panel represents the difference between the dihedral angles of the two conformations. The red box highlights the linker region where the backbone is changing conformation, which overlaps with the region identified in Figure 3.4A.

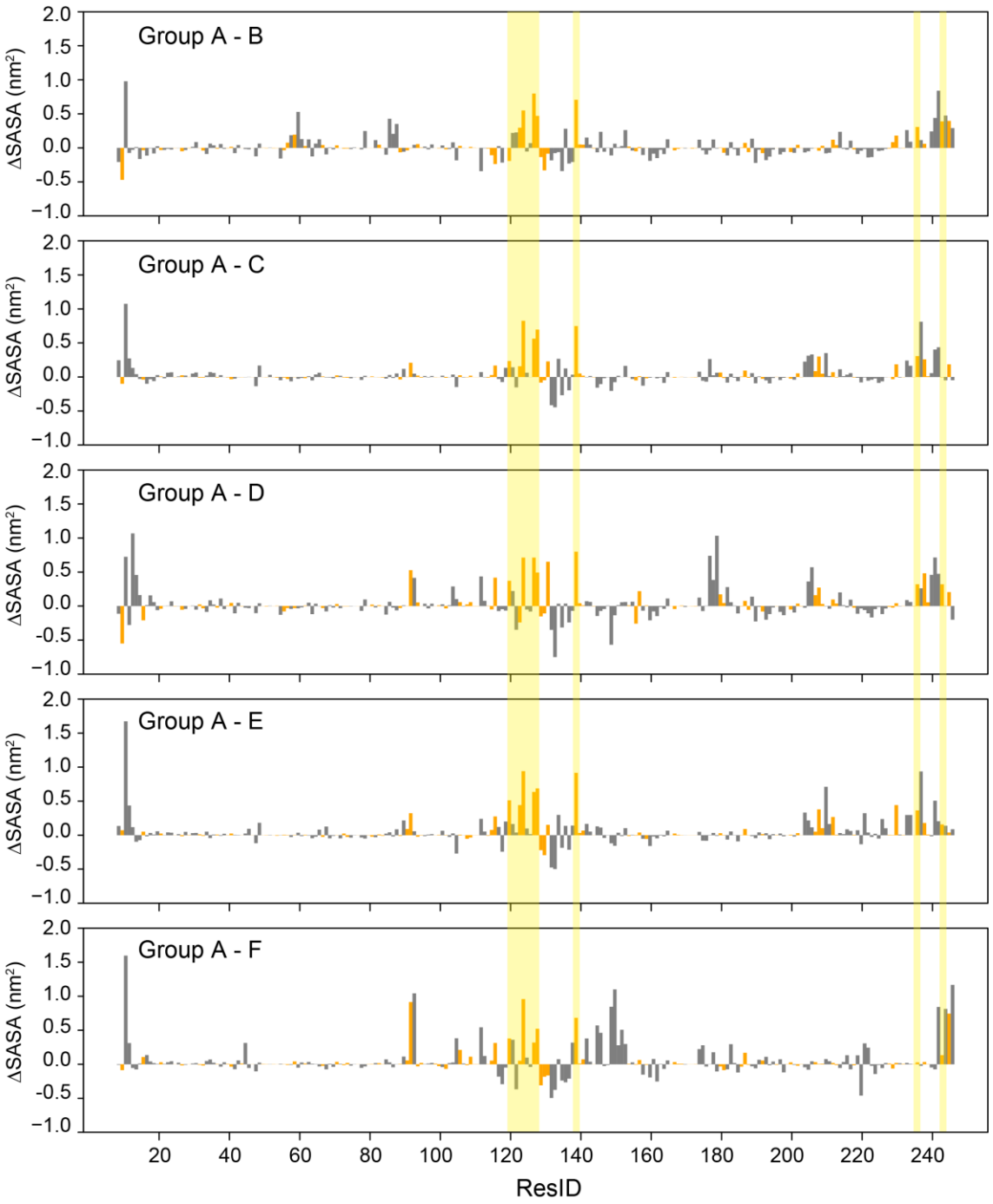

Figure 3.14. Difference in solvent accessible surface area (SASA) between the main open conformer of Dys ABD1 (Group A) and each major closed conformer (Groups B-F), where Groups A-F are those identified in Fig. 3.3 of the main text. Hydrophobic residues are shown as yellow bars whereas all other residues are gray. The specific residues pointed out in Fig. 3.4C of the main text are highlighted in the panels above with a light yellow strip.

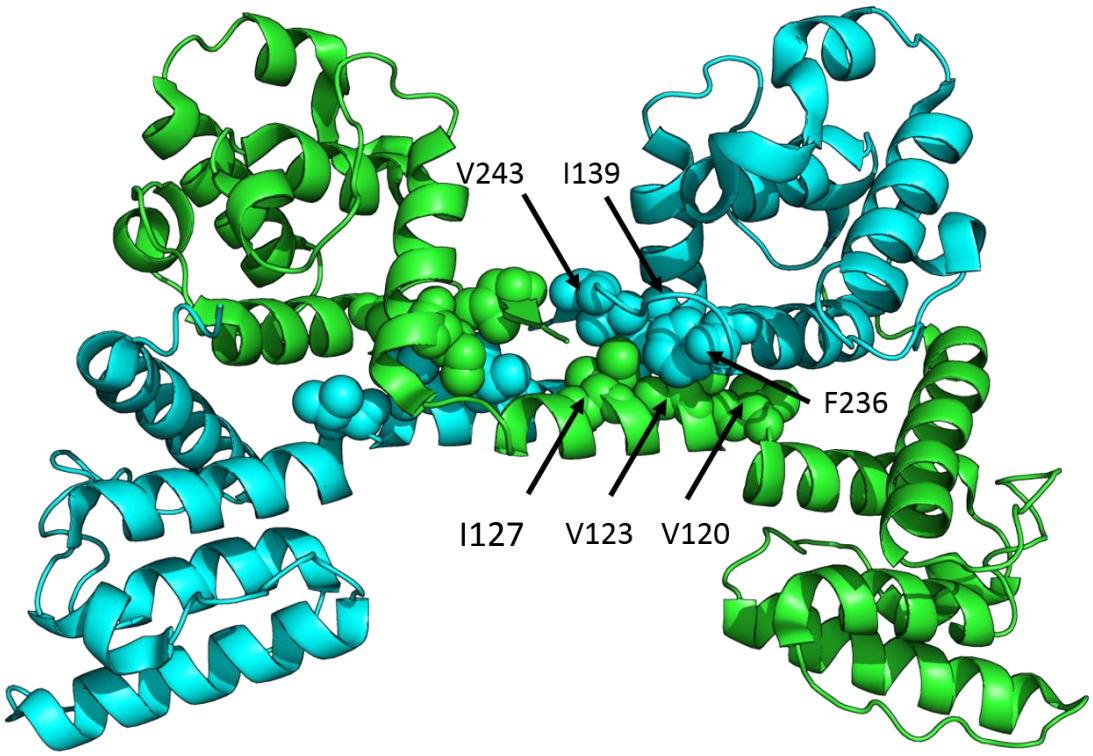

Figure 3.15. Dys ABD1 crystal structure 1DXX showing the swapped domains between chains A and B. In spheres are the hydrophobic residues identified as forming interactions in the compact structures identified by MD simulations in Figure 3.4.

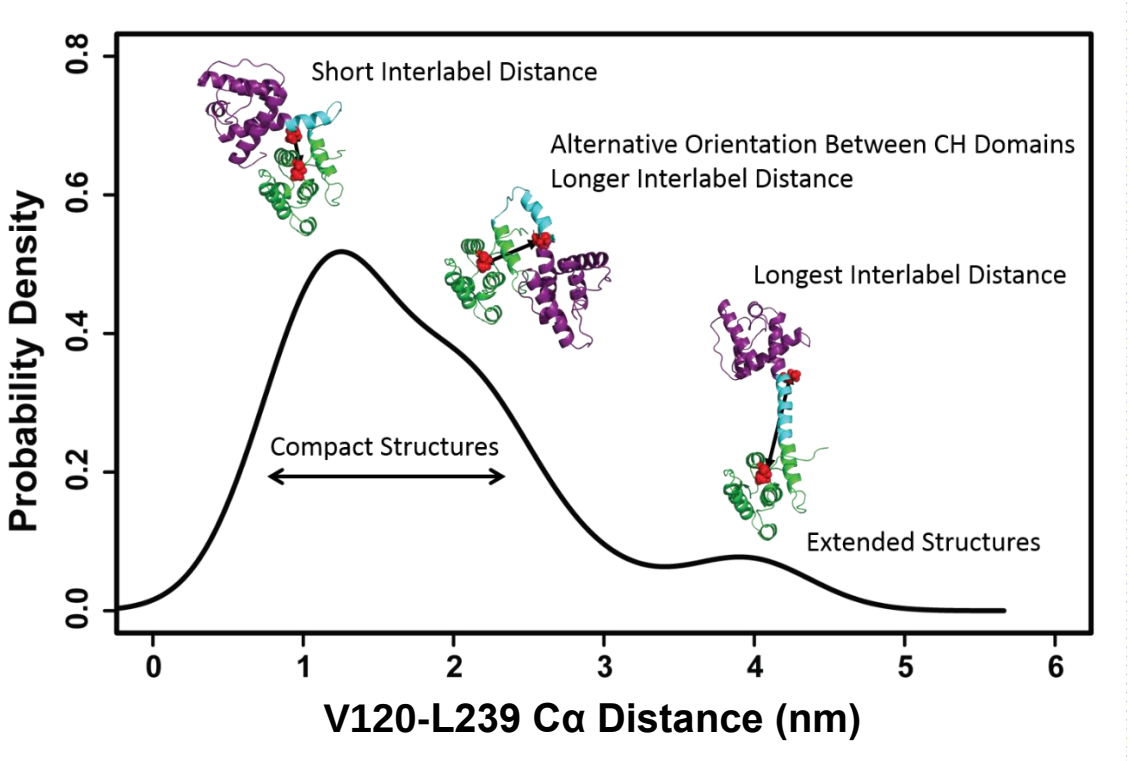

Figure 3.16. Probability density plot of the distance between the Cα atoms of the two spin labels V120 and L239 in Dys ABD1, calculated over the entire conformational ensemble. Representative structures for selected distances are also shown.

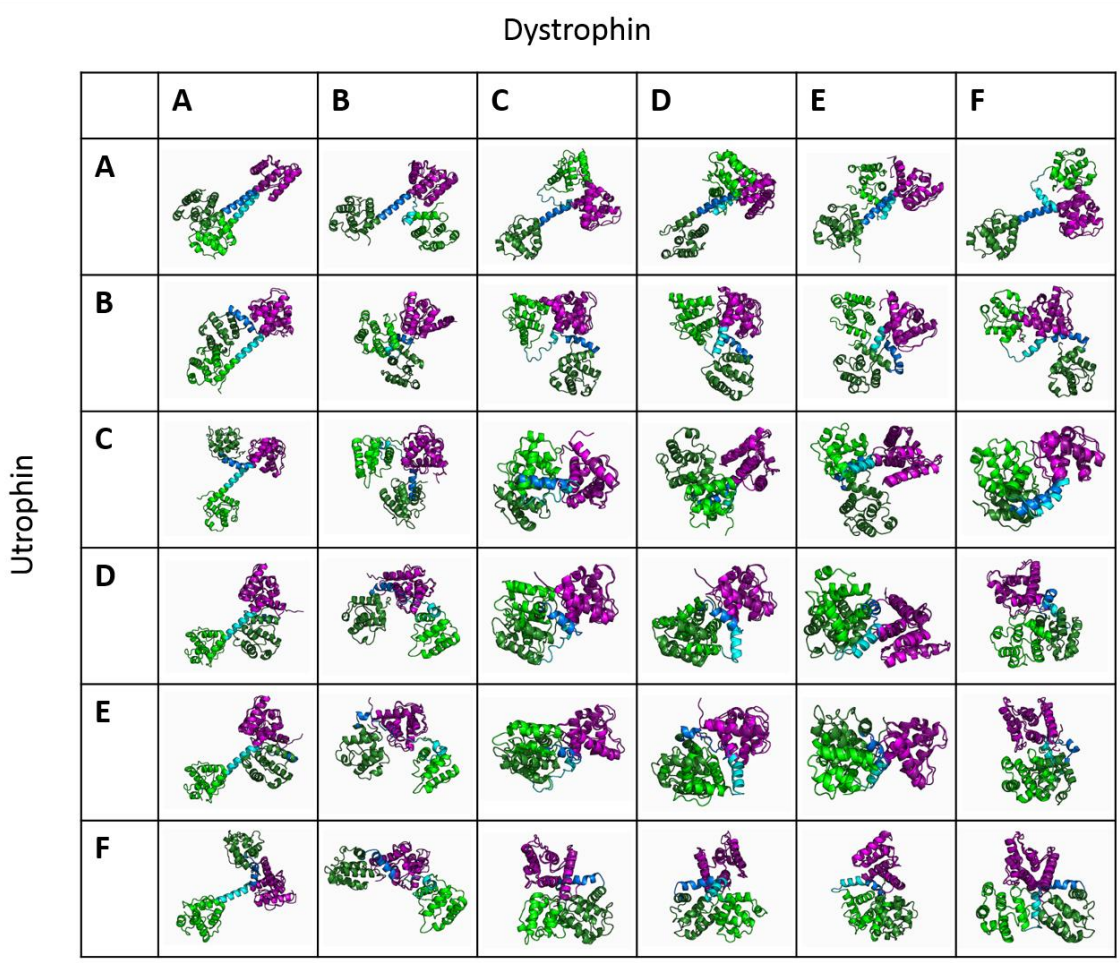

Figure 3.17. Overlap between Dys ABD1 and Utr ABD1 conformational ensembles. Structures were aligned using the CH1 domain as a reference (purple). Lighter cyan and green colors are used for Dys ABD1, darker colors for Utr ABD1. The two axes label the specific conformer as described in Figure 3.3 of Dys ABD1 (horizontally) and of Utr ABD1 (vertically).

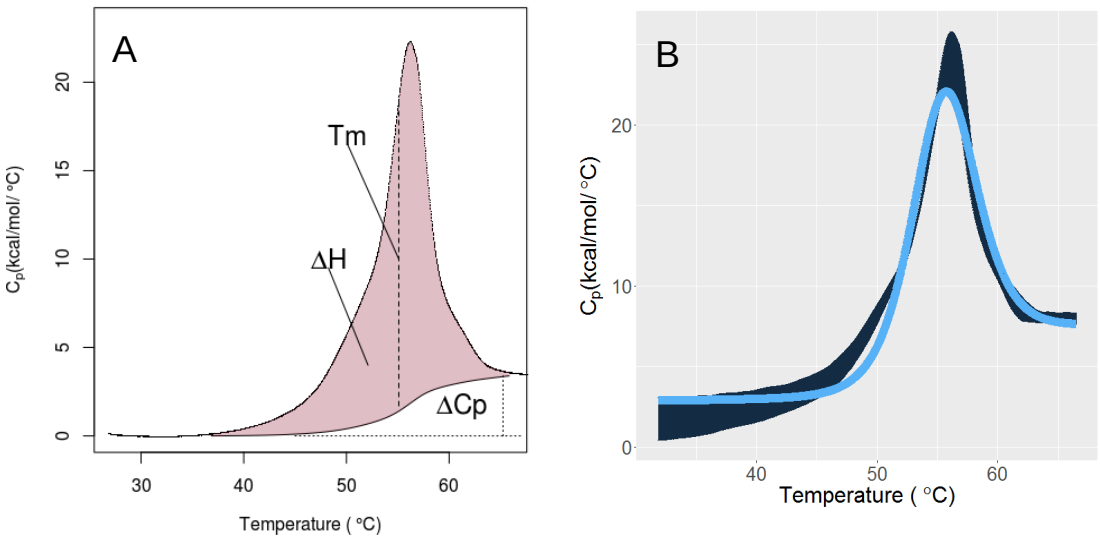

Figure 3.18. Graph A illustrates the calorimetric features of the thermal denaturation of Dys ABD1 using averaged DSC data from Dys ABD1. The key features ΔH, T_m, and ΔC_p are representative of the total area under the curve (shaded red), the temperature at the half integral, and the overall heat capacity baseline shift of the process, respectively. Graph B represents the averaged triplicate scans (where the width represents the standard deviation) in dark blue and the model fit in light blue.

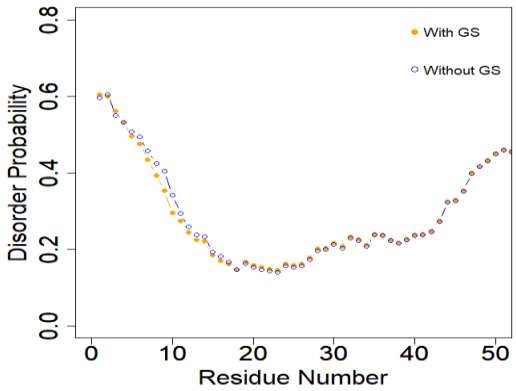

Figure 3.19. Theoretical calculations from the Uversky lab determining the likelihood of disorder throughout the Dys ABD1 construct.[141, 142] This was utilized to provide a metric with which to determine the effect of cloning material left behind (which was a glycine and a serine after TEV cleavage) on the purified construct. The similarities between ABD1 with cloning remnants (orange) and without cloning remnants (blue) is predicted to be minimal based on these calculations.

Dys ABD1	Prediction (Freire)	Prediction (Spolar)	Calorimetric	Van't Hoff	NLS Fit
ΔH_{Tm} (kcal/mole)	NA	160 ± 30	142 ± 3	110	118.5 ± 0.4
ΔC_p (kcal/mole/°C)	4.7	3.7 ± 0.7	3.5 ± 0.5	NA	4.09 ± 0.07
T_m (°C)	NA	NA	55.13± 0.02	NA	55.07±0.02
ΔG(37°C) (kcal/mole)	NA	NA	6.1 ±0.4	NA	4.45 ± 0.07

Table 3.1. List of fitted thermodynamic parameters for Dys ABD1 with associated errors. From left to right, the columns contain the list of thermodynamic values being represented, values from the nonlinear least squares (NLS) model fit, thermodynamic values taken from the calorimetric features of the raw data, excess enthalpy from the Van't Hoff analysis, predictions for excess enthalpy and change in heat capacity based on molecular weight, and a prediction for the change in heat capacity based on differences in the predicted absolute heat capacities of the folded and unfolded states. Values that were not described by the given model are listed as 'NA'. Errors for the Van't Hoff enthalpy and Freire prediction model are not included.

CHAPTER 4: STRUCTURAL IMPACT OF PHOSPHORYLATION AND DIELECTRIC CONSTANT VARIATION ON SYNAPTOTAGMIN'S INTRINSICALLY DISORDERED REGION*

4.1 – Section Summary

We used time-resolved Förster resonance energy transfer, circular dichroism, and molecular dynamics simulation to investigate the structural dependence of synaptotagmin 1's intrinsically disordered region (IDR) on phosphorylation and dielectric constant. We found that a peptide corresponding to the full-length IDR sequence, a ~60-residue strong polyampholyte, can sample structurally collapsed states in aqueous solution, consistent with its κ-predicted behavior, where κ is a sequence-dependent parameter that is used to predict IDR compaction. In implicit solvent simulations of this same sequence, lowering the dielectric constant to more closely mimic the environment near the surface of a lipid bilayer promoted further sampling of collapsed structures. We next examined the structural tendencies of central region residues of the IDR in isolation. We found that the exocytosis-modulating phosphorylation of Thr112 disrupts a local disorder-to-order transition induced by trifluoroethanol/water mixtures that decrease the solution dielectric constant and stabilize helical structure. Implicit solvent simulations on these same central region residues testing the impact of dielectric constant alone converge on a similar result, showing that helical structure is formed with higher probability at a reduced dielectric. In these helical conformers, lysine-aspartic acid salt bridges contribute to stabilization of transient secondary structure. In contrast, phosphorylation results in formation of salt bridges unsuitable for helix formation. Collectively, these results suggest a model in which phosphorylation and compaction of the IDR sequence regulate structural transitions that in turn modulate neuronal exocytosis.

*This chapter was reproduced in its entirety with permission from the following article: Fealey ME, Binder BP, Uversky VN, Hinderliter A, Thomas DD. Structural impact of phosphorylation and dielectric constant variation on synaptotagmin's IDR. *Biophys J.* 2018 114:550-561.

4.2 – Introduction

Synaptotagmin 1 (Syt 1) is the calcium ion (Ca^{2+}) sensor for neurotransmitter release, chelating Ca^{2+} upon action potential-induced entry and consequently triggering fast synchronous release.[146] The underlying mechanisms by which Syt 1 mediates this biological event remain incompletely understood. Our goal has been to investigate the allosteric network of Syt 1 to elucidate regulatory mechanisms that underlie controlled neurotransmitter release.[49, 147, 148] Our group found that an intrinsically disordered region (IDR) within Syt 1 (residues ~80-141) exerts allosteric influence over the adjacent C2 domain referred to as C2A (Fig. 4.1*A*). Until recently, this IDR had received little attention and its structural biology still represents a significant gap in our understanding of Syt 1 function. Because the Syt 1 IDR exerts allosteric influence over C2A, structural or disordered ensemble transitions that occur there may be important for modulating C2 domain functions that in turn influence exocytosis.[149] Additionally, the IDR was recently shown to bind to the pleckstrin homology (PH) domain of dynamin 1, implicating its sequence in an important protein-protein interaction of endocytosis.[150]

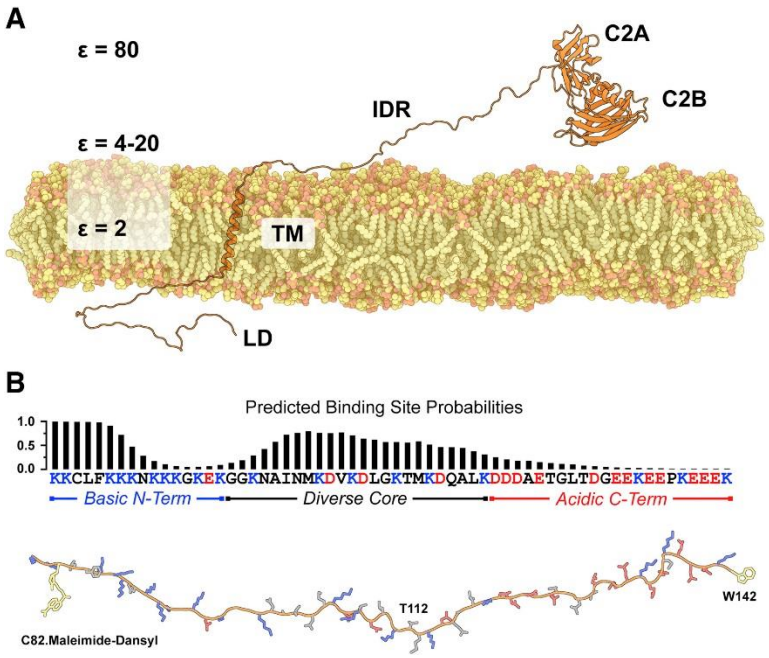

Figure. 4.1. Model of Syt 1 and amino acid sequence of its IDR. (*A*) Syt 1 consists of a short luminal domain (LD), a single transmembrane helix (TM), a ~60 residue IDR, and two calcium ion- and phospholipid-binding C2 domains in tandem, C2A and C2B. Approximate dielectric constants for the bilayer core (ε=2), interfacial region (ε=4-20), and bulk solution (ε=80) are shown to indicate environments the polyampholytic IDR may experience. (*B*) The IDR sequence (residues 80-141) is shown with the basic residues in blue and the acidic residues in red. The distribution of charged residues partially delineates the sequence into three segments as indicated. Above the sequence is the ANCHOR predicted probability for each residue potentially contributing to a binding site as described in ref. [151] and [152].The modeled peptide below the sequence shows the location for covalent attachment of acceptor dyes as well as location of the added tryptophan (yellow) used in FRET experiments. Thr112, the phosphorylation site examined in this study, is also indicated.

When examining the amino acid sequence of this IDR to make initial structural predictions, it appears to be a strong polyampholyte with ~50% of its residues consisting of an almost equal number of oppositely charged amino acids (Fig. 4.1 *B*).[153, 154] However, the distribution of charged residues divides the IDR into three main segments: a highly positive N-terminal region (~80-96), a more sequence-diverse central region with dispersed lysine-aspartic acid repeat motifs (~97-120), and a highly negative C-

terminal region (~121-141) (Fig. 4.1 *B*). Intrinsically disordered proteins (IDPs) that have a high degree of charge separation sample more compact structural states, and the degree of compactness can be predicted with the calculated κ. This κ varies from 0 (less compact) to 1 (most compact) and, in model 50-mer peptides consisting of an equal number of lysine and glutamic acid residues, correlates to radii of gyration (R_g) that range from 3.0 nm to 1.7 nm.[155] When we performed this calculation for Syt 1's IDR sequence, we determined a κ value of 0.22, suggesting that the IDR can sample more compact structural states whose R_g approaches 2.3 nm (Tables 4.3 and 4.4, Supporting Materials and Methods Chapter 4). Further application of the sequence-dependent analysis software, CIDER, places the Syt 1 IDR in a chimeric region of the IDP phase diagram, predicting exploration of both coil and hairpin structures (Fig. 4.9).[156] Recently it was proposed that this IDR functions in Syt 1 as an "electrostatic zipper" that opens and closes to mediate docking of synaptic vesicles and fusion pore opening, respectively, though no distance measurement was made.[157] Chimeric behavior of the IDR would be appropriate for such opening and closing and can be experimentally measured using Förster resonance energy transfer (FRET), an approach we have explored here (Fig. 4.1 *B*).

While the calculated κ term for the Syt 1 IDR probably stems from the N- and C-terminal regions with high charge density, the central region has its charged residues more evenly distributed, suggestive of more local coil disorder. Within this central region there is a series of KD-repeats with the exception of a KT sequence at residue positions 111 and 112. Many of these charged residues are in *i*,*i*+4 positions, suggesting they could stabilize helical secondary structure through salt bridge formation. Intriguingly, Thr112 is

a site of phosphorylation that was recently shown to promote PKC-mediated potentiation of synaptic transmission (Fig. 4.1 *B*).[158] Addition of a phosphate at this site would complete a positive-negative repeat pattern in this central region of the IDR, impact the local electrostatics of the region, and probably alter its structural propensity. Furthermore, this central region contains a predicted binding site that is adjacent to Thr112 (Fig. 4.1 *B*), suggesting that phosphorylation of Thr112 could initiate a structural transition that modulates binding.[151, 152] Such a mechanism could help explain how phosphorylation *in vivo* promotes synaptic plasticity. Collectively, this differential distribution of charged residues suggests that electrostatics serve more than one structural role in IDR-modulation of neuronal exocytosis.

The polyampholyte nature of this IDR is of further significance because it is part of an integral membrane protein. The lipid bilayer core and interfacial regions have low dielectric constants (ε=2-20) compared to bulk aqueous solution (ε=80), so the potential structural impact of charged residues in the IDR sequence is likely to be more pronounced in Syt 1's native local environment (Fig. 4.1 *A*). In the current study, we aimed to answer the above questions in an attempt to establish basic principles of this IDR's behavior.

4.3 – Materials and Methods

4.3.1 – Reagents

Potassium chloride (KCl), sodium phosphate monobasic (NaH_2PO_4) and sodium phosphate dibasic (Na_2HPO_4) were obtained from ThermoFisher Scientific. 2,2,2-trifluoroethanol (TFE) ≥99.0% was obtained from Sigma-Aldrich. *N*-[2-

(dansylamino)ethyl]maleimide and N-(iodoacetaminoethyl)-1-naphthylamine-5-sulfonic acid (IAEDANS) acceptor labels were obtained from Sigma-Aldrich. Three peptides corresponding to the full-length Syt 1 IDR (residues 80-141) with a C-terminal tryptophan as well as the core region of the IDR with some of the acidic C-terminus (residues 97-130) in both unphosphorylated and phosphorylated form (phosphorylated at threonine 112) were obtained from Selleck Chemicals at a purity of 95% based both on mass spectrometry and HPLC. Numbering of residues assumes start codon methionine is removed. The amino acid sequences for each of the three peptides correspond to the following: (full-length IDR) KKCLFKKKNKKKGKEKGGKNAINMKDVKDLGKT-MKDQALKDDDAETGLTDGEEKEEPKEEEKW; (unphosphoryl-ated core of IDR) GKNAINMKDVKDLGKTMKDQALKDDDAETGLTDG; (phosphorylated core of IDR) GKNAINMKDVKDLGK(pT)MKDQALKDDDAETGLTDG. The core region peptides included a few residues from the acidic terminus of the IDR to help ensure adequate solubility. Despite this inclusion of C-terminal residues, these peptides are referred to as the "core region" throughout the manuscript. The peptides did not include N- or C-terminal caps in an attempt to avoid potential perturbations to helicity.[159]

In the full-length IDR construct, the tryptophan was added to both enable UV measurement for concentration determination and also to act as a donor probe in time-resolved FRET experiments. Use of tryptophan and dansyl was also chosen as a way of minimizing structural perturbation to system with fluorescent probes, only requiring addition of one residue and exploiting the single native cysteine.

4.3.2 – Tryptophan Lifetime FRET

For the time-resolved FRET, we first reconstituted full-length IDR peptide in 10 mM NaH_2PO_4/Na_2HPO_4 buffer with the pH adjusted to 7.4. Measuring the A280 and using a calculated extinction coefficient of 5600 $cm^{-1}M^{-1}$, the concentration of peptide was determined. A 4-fold excess of maleimide-dansyl (in the case of the full-length synthetic peptide) or AEDANS (in the case of the recombinant full-length IDR) acceptor probe that had been reconstituted in dimethylformamide (DMF) was then added to the peptide solution and allowed to incubate with the full-length IDR for 3 hours at room temperature (~23 °C) in the presence of 1 mM tris(2-carboxyethyl)phosphine (TCEP) reducing agent to prevent cysteine cross-linking. During the reconstitution of full-length IDR, the peptide forms a dimer that is in equilibrium with its monomeric form. To isolate the acceptor-labeled monomer for FRET and simultaneously separate labeled peptide from free dye, we subjected the peptide solution to size exclusion chromatography using a pre-packed S100 sephadex column from G&E. Because of the large molecular weight difference between the peptide dimer (~14 kDa), peptide monomer (~7 kDa) and free maleimide-dansyl probe (~370 Da), all three species can be resolved into separate fractions. The elution fractions containing just acceptor-labeled full-length IDR peptide were identified by UV absorption, pooled, and concentrated using an Amicon Ultra-4 centrifugal filter unit from Millipore whose molecular weight cut-off was 3 kDa. The final concentration of full-length peptide as well as the extent of labeling with maleimide-dansyl were determined with UV absorption using the approach described by Gustiananda et al.[160]. In this approach, the peptide concentration is determined by

measuring the donor-acceptor peptide and subtracting the contribution from dansyl with a correction factor previously described[161]:

$$[\text{IDR peptide}] = (A280 - (A331 \times 0.386))/5500 \quad \text{Eq. 4.1}$$

$$[\text{maleimide-dansyl}] = A331/4000 \quad \text{Eq. 4.2}$$

The ratio of [IDR peptide] and [maleimide-dansyl] enables an assessment of labeling efficiency and for the full-length IDR peptide, this ratio was near unity indicating complete labeling. This seemed reasonable considering the IDR is unfolded, allowing for the cysteine to be well exposed to free acceptor dye in bulk solution.

Tryptophan lifetimes were measured on a custom made spectrometer from Fluorescence Innovations[147, 162], which uses an Nd:Yag laser for excitation of pyrromethene 597 with a 532 nm wavelength. The emission from pyrromethene passes through a doubling crystal to create a 295 nm wavelength specific for tryptophan excitation. Syt 1 IDR with a C-terminal tryptophan was synthesized (or expressed recombinantly) and subjected to lifetime measurements at 23°C in the absence and presence of dansyl (or AEDANS in the case of the recombinant IDR) acceptor probe irreversibly linked to the N-terminal cysteine residue present in the natural sequence (Fig. 4.1 *B*). Tryptophan emission from samples at concentrations of 1-3 μM was monitored at 340 nm.

Tryptophan waveforms with and without covalently attached acceptor dye were fit to multiexponential decays using a custom-made fitting program FargoFit.[163] The general description of the waveform is a sum of exponentials:

$$F(t) = \sum_{i=1}^{n} a_i e^{-t/\tau i} \quad , \qquad \text{Eq. 4.3}$$

where a_i represents the amplitude of each i^{th} component, n represents the number of components, and τ_i is the fluorescence lifetime. The tryptophan waveforms were sufficiently described by two exponentials (Fig. 4.2 *A*). To determine the distance *R* between tryptophan and dansyl probes, amplitude-weighted average lifetimes $\langle\tau\rangle$ were calculated and subsequently used to determine the efficiency of energy transfer (E) which was then used in a rearranged Förster equation with 2.1 nm for the tryptophan-dansyl R_0:

$$E = 1 - \langle\tau_{DA}\rangle/\langle\tau_D\rangle, \quad R = R_0(1/E - 1)^{1/6} \qquad \text{Eq. 4.4}$$

4.3.3 – Circular Dichroism

Peptides were reconstituted in buffer containing 10 mM sodium phosphate (pH 7.4), and stock solution concentrations were determined using BCA microplates and UV absorption (for full-length peptides). TFE was added at incremental percentages v/v over a range of 0-60%. Each raw data set of absorption had a corresponding buffer scan

subtracted from it to remove any absorption contributions from buffer or TFE co-solvent. Resulting data sets were plotted as mean residue ellipticity (MRE) according to the following equation:

$$MRE = \theta x(MW/(N - 1))/lc \,, \qquad \text{Eq. 4.5}$$

where [θ] is raw ellipticity, MW is the molecular weight of the peptide, N is the number of residues, l is the cuvette path length, and c is the peptide concentration in mg/mL. Concentrations of core region peptide samples ranged from 18-20 μM for all experiments whereas full-length IDR peptides were at a concentration of 10 μM. To extract the free energy of folding for Syt 1 IDR peptides, a simultaneous non-linear least squares fit of 198 nm and 222 nm wavelengths was performed (mathematical derivation in Supporting Materials and Methods Chapter 4). Four replicate measurements of TFE gradients for each Syt 1 IDR peptide were individually fit and then averaged to determine standard deviations. Estimates of helicity were obtained by fitting data sets to linear combinations of α-helix, β-sheet, and random coil secondary structures as described previously.[13, 164]

4.3.4 – Atomistic MD simulation of Syt 1 IDR

Peptides corresponding to both the full-length Syt 1 IDR and the core region were generated in Discovery Studio 2.0. Molecular Dynamics (MD) simulations were performed with NAMD 2.9[165] using the CHARMM36 force field[166], the Generalized Born Implicit Solvent (GBIS) model[167], and 2-fs time steps. Peptide termini were left

charged to mimic the state of the peptides used in CD experiments. The peptide systems were maintained at constant temperature with a Langevin thermostat (298 K). Peptides corresponding to residues 80-141 in the unphosphorylated state or residues 97-129 of the Syt 1 IDR with and without phosphorylation at Thr112, were subjected to energy minimization and gradually warmed up over a period of 200 ps prior to equilibration. All systems subsequently achieved RMSD convergence within 50 ns (Fig. 4.10). We performed a total of eight different simulations. For the full-length Syt 1 IDR sequence, the peptide was subjected to two different salt concentrations (10 mM and 100 mM) at two GBIS solvent dielectric constants ($\varepsilon = 80$ and $\varepsilon = 20$). For the core region residues alone, we simulated the peptide with and without phosphorylation of Thr112 at the two dielectric constant values described above at a salt concentration of 100 mM. A ε value of 80 was used to mimic an aqueous solution environment, while a value of 20 was used to mimic the upper limit of the membrane interfacial region.[168, 169] The full-length IDR peptide was simulated at $\varepsilon = 80$ for 970 ns with either a 10 mM salt concentration or a 100 mM salt concentration. At $\varepsilon = 20$, the full-length peptide was simulated for 480 ns and 760 ns at salt concentrations of 10 mM and 100 mM, respectively. The unphosphorylated core region peptide was simulated with $\varepsilon = 20$ for 1 μs and with $\varepsilon = 80$ for 1.08 μs. The phosphorylated core region peptide was simulated with $\varepsilon = 20$ for 1.0 μs and with $\varepsilon = 80$ for 1.1 μs. Trajectories were analyzed using VMD 1.9.1.[170] The program STRIDE, included within VMD, was used to monitor secondary structure throughout each trajectory.[171] When calculating the probability of existing in different types of secondary structure (coil, beta sheet, turn, and helix), we combine 3_{10}-helicies with α-

helices under the assumption that 3_{10}-helical structure is an on-path transition point between coil and α–helix.[172]

4.3.5 – Protein Purification

For FRET studies on the recombinant IDR construct, a nucleotide sequence encoding residues 80-141 with a C-terminal tryptophan followed by a His-tag, was cloned into a pET28a vector and subsequently transformed into BL21 *E. coli* using the New England BioLabs heat-shock protocol. Transformed cells were plated onto a kanamycin-containing agar plate with an antibiotic concentration of 50 μg/mL and subsequent colonies were used to generate inoculate for large-scale bacterial growth and protein expression, induced with 1 mM IPTG. After 12 hours at 25 °C, cells were harvested by centrifugation. Cell pellets were resuspended in 20 mM MOPS, 200 mM KCl, 4 M urea, 1 mM PMSF at pH 7.4 (4 °C), then lysed by sonication. The resulting lysate was treated with DNAse and 1 mM $MgCl_2$ and allowed to stir at 4 °C for 1 hour, followed by centrifugation to remove cellular debris. The supernatant was passed through a ThermoScientific Nalgene prefilter plus GFP sterile filter with a pore size of 0.2 μm to remove further particulate matter. The filtered solution was then slurried with TALON metal affinity resin buffered in 20 mM MOPS, 200 mM KCl, 4 M urea, pH 7.4. The slurry was gently stirred at 4 °C for 2 hours, then transferred into a column to remove unbound protein and nucleic acid. Settled resin was subsequently rinsed with four column volumes of buffer containing 20 mM MOPS, 200 mM KCl, 4 M urea, and 10 mM imidazole (to remove non-specifically bound proteins) at pH 7.4. Resin was then rinsed with a high salt buffer (20 mM MOPS, 1 M KCl, 4 M urea, 10 mM imidazole at pH 7.4) to separate nucleic acid bound to linker region peptide (specifically, the region containing

a high number of lysine residues). The resin was then rinsed with four column volumes of buffer containing 20 mM MOPS, 200 mM KCl, 4 M urea, and 10 mM imidazole at pH 7.4. The linker region protein was then eluted with buffer containing 20 mM MOPS, 100 mM KCl, 4 M urea and 300 mM imidazole at pH 7.4. Fractions of eluted protein were run on SDS-PAGE and pure fractions whose 260 nm/280 nm ratio was <0.7 (which indicates >95% purity with respect to nucleic acid contamination) were pooled and concentrated. Syt 1 IDR was then dialyzed against 4 liters of buffer containing 10 mM sodium phosphate at pH 7.4 four times, each with an equilibration time of 6 hours to ensure complete exchange. The Syt IDR was then subjected to size-exclusion chromatography as described above for the Syt IDR synthesized peptide to isolate monomers.

4.4 – Results

4.4.1 – Full-length IDR peptide can sample compact structural states in aqueous solution

To first test the hypothesis that the Syt 1 IDR can sample compact structural states consistent with an "electrostatic zipper", we used tryptophan (Trp)-dansyl FRET, placing the aromatic amino acid at the C-terminus of an IDR peptide and dansyl at the N-terminus using the sequence's single native cysteine for irreversible attachment (Fig. 4.1 *B*). The Trp-dansyl probe pair has an R_0 of 2.1 nm, a value suitable for detecting IDR compaction based on the κ prediction if it occurs. We measured the Trp lifetime decay with and without acceptor dansyl and found the lifetime to be shortened by proximity of the acceptor probe. When fit to bi-exponential functions (Fig. 4.2 *A*), it was found that the

Trp and dansyl probes were a mean distance of 2.40±0.05 nm apart (Fig. 4.2 *B*), which represents the average inter-probe distance but only for compact conformers detected by this dye pair.

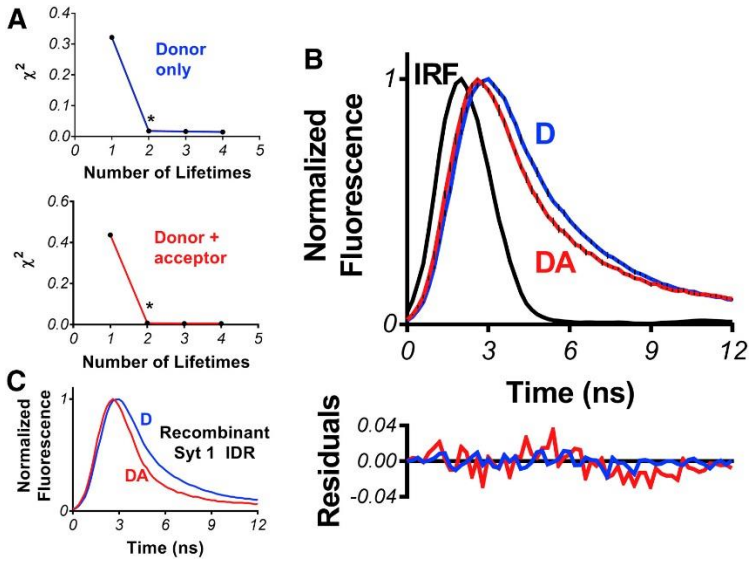

Figure. 4.2. Sampling of a structurally collapsed state by the Syt 1 IDR in aqueous solution. (*A*) Determining the number of lifetimes needed to fit tryptophan decays without (blue) and with (red) acceptor dyes. In both cases, two lifetimes were considered necessary and sufficient (asterisk). (*B*) Time-resolved FRET measurement of Syt 1 IDR synthesized peptide showing Trp-dansyl undergoing energy transfer in 10 mM sodium phosphate buffered to a pH of 7.4. Blue: donor-only. Red: donor-acceptor. Error bars: SD of 4 replicates. Black: instrument response function (IRF). The distance derived from the change in average lifetime ($\langle\tau_D\rangle$ = 1.53±0.11, $\langle\tau_{DA}\rangle$ = 1.06±0.10) is 2.40±0.05 nm (Eq. 4.4). Shown below the time-resolved data are residuals of fitting to bi-exponential functions. (*C*) A similar FRET result was obtained from a recombinantly expressed Syt 1 IDR construct that was labeled with AEDANS acceptor. The slight increase in FRET in the recombinant IDR is thought to be due to inclusion of a his-tag, which tend to promote compaction in IDPs.[173]

Given the ~60 residue separation between the two probes, such a measurement indicates that the IDR peptide can sample compact structural states in aqueous solution, though this measurement cannot differentiate between hairpin and globular structures. Additionally, by using our minimally perturbing probe pair, we are not able to measure

occupation of conformers adopted beyond a ~3.4 nm inter-probe distance, where FRET efficiency between Trp-Dansyl is below 5%. Thus, to complement our FRET measurements characterizing the structural behavior of the IDR in an aqueous environment with low salt, we next performed MD simulations on a peptide corresponding to the full-length Syt 1 IDR sequence (residues 80-141). We used an implicit solvent model with a dielectric constant of $\varepsilon = 80$ and a 10 mM salt concentration. The simulated IDR structure sampled a wide range of conformers under these conditions as indicated by its fluctuating R_g during the trajectory (Fig. 4.3A). Intriguingly, while FRET indicates that sampling of compact conformers does occur (Fig. 4.2), MD simulations suggest that compact conformers are not the dominant structural species in the overall aqueous environment ensemble (Fig. 4.3B). On average, the IDR sequence samples extended coils more frequently (Fig. 4.3C). Collectively, FRET and MD results correlate with the IDR's predicted behavior based on amino acid sequence alone and is consistent with the previously proposed "electrostatic zipper" model in which the IDR samples both open (extended) and closed (compact) conformers.[156, 157]

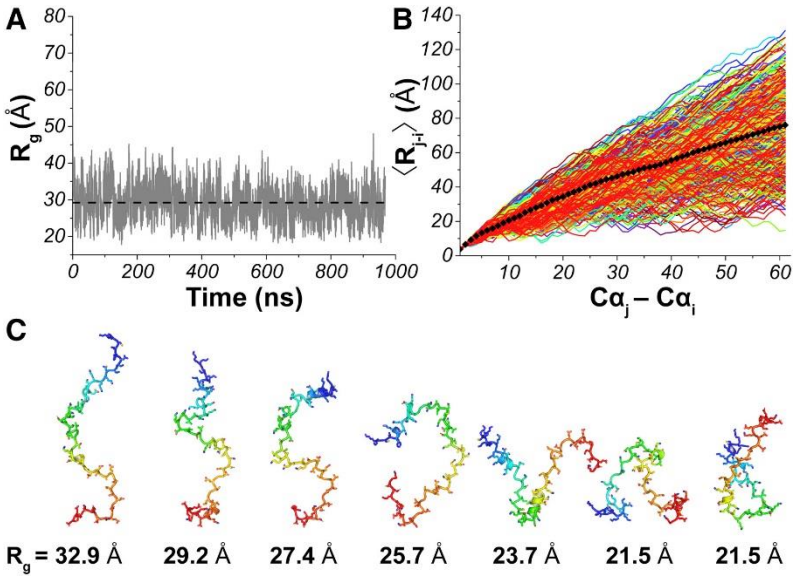

Figure. 4.3. MD simulation of the full-length Syt 1 IDR (residues 80-141) with an implicit solvent (ε=80) and salt concentration of 10 mM. (*A*) Radius of gyration (R_g) as a function of simulation time indicates the IDR sequence samples both extended and compact states. Dotted line indicates average R_g. (*B*) Inter-residue distance ($<R_{j-i}>$) between first (i) and each subsequent (j) residue ($C\alpha_j$-$C\alpha_i$) in the Syt 1 IDR plotted as a function of simulation time. Plots suggest weak long-range interactions promote sampling of more compact conformations. However, compact conformers are not the dominant structural species in an aqueous environment as indicated by the location of the average inter-residue distances (black diamonds). (*C*) Representative conformers of different R_g. Structures, N- to C-terminus, are color-coded blue to red.

4.4.2 – TFE-water mixtures promote partial helicity in core IDR residues

Next, to assess the structural potential of the more sequence-diverse central region, we used circular dichroism and the helix-promoting co-solvent trifluoroethanol (TFE) (Fig. 4.4 and Fig. 4.11). Several of the residues in the central region, because they have hydrophobic side chains, would likely have low free energy barriers to desolvation, which often occurs upon protein-protein binding. This, coupled with the core region's predicted binding site and phosphorylation (Fig. 4.1 *B*), make it a potential site of ordering. Moreover, TFE/water mixtures allow for creation of low dielectric environments that more closely mimic that found near a lipid bilayer surface. For these

reasons, TFE is a useful perturbing agent to study structural transitions of a membrane-localized polyampholyte IDR.[174] In the unphosphorylated state, increasing TFE concentration results in partial folding into helical structure (Fig. 4.4 *A*). This transition exhibits 2-state behavior exemplified by the isodichroic point (vertical dotted line). When we fit the transition to a 2-state model (Fig. 4.4 *D*, Supporting Material and Methods) we measured a folding free energy (ΔG_0) of 1.0±0.2 kcal/mole (Table 4.1).

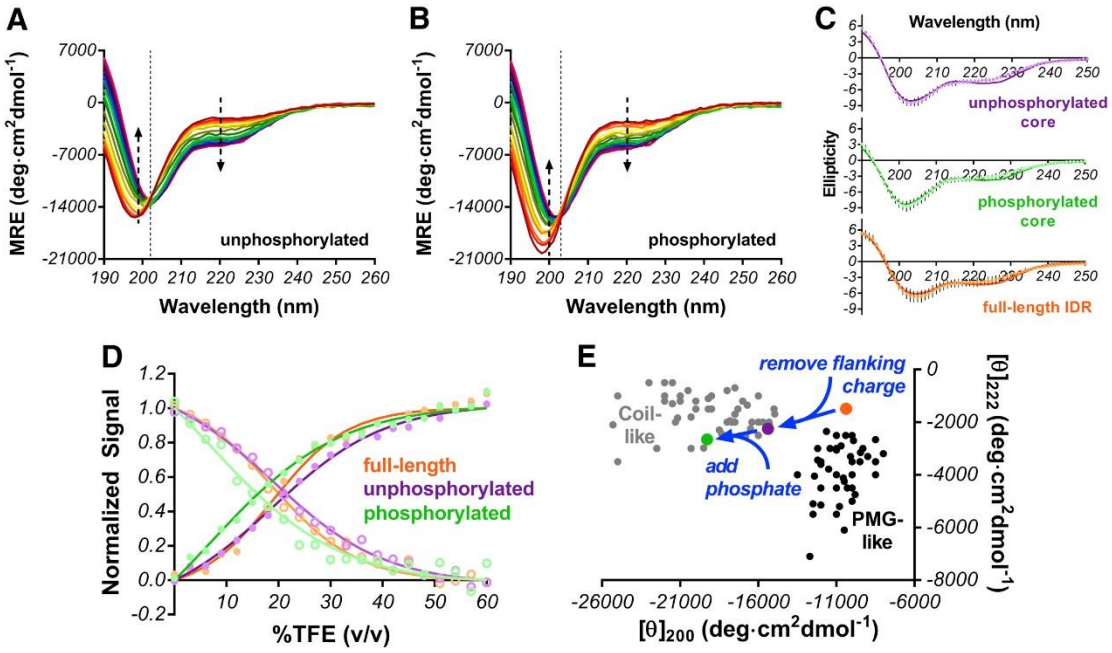

Figure. 4.4. Impact of Thr112 phosphorylation on Syt 1 IDR core region (GKNAINMKDVKDLGKTMKDQALKDDDAETGLTDG). (*A*) MRE of unphosphorylated Syt 1 IDR core and (*B*) MRE of phosphorylated Syt 1 IDR core as a function of increasing TFE (0-60% v/v going from red to violet) in 10 mM sodium phosphate at a pH of 7.4. (*C*) Fitting of absorption profiles to linear combinations of α-helix, β-sheet and random coil in each IDR construct at a maximal TFE concentration of 60% v/v. (*D*) Simultaneous fitting of the 198 nm coil minimum (solid circles, dark line) and 222 nm (open circles, lighter line) signals during folding transition for unphosphorylated (purple), phosphorylated (green) and full-length (orange) peptides. (*E*) Comparison of 222 nm/200 nm ratio from full-length (orange), unphosphorylated core (purple) and phosphorylated core (green). Removal of flanking charge shifts IDR from a pre-molten globule-like (PMG-like) state to a more coil-like state and phosphorylation accentuates this effect. Figure modified from reference[175].

To assess the impact of phosphorylation, we repeated the TFE titration with a phosphorylated peptide (phosphothreonine, pThr112) of identical sequence (Fig. 4.4 *B*). This approach of using TFE-induced folding has been successfully applied in other disordered systems to discern differences in folding propensity upon residue phosphorylation.[176, 177] Upon inclusion of the post-translational modification, TFE-induced folding was somewhat restricted as evident in the absorption profile (Fig. 4.4 *B*, Fig. 4.11). When the structural transition was fit to a 2-state model as before, a ΔG_0 of 0.36±0.09 kcal/mole was found (Fig. 4.4 *D*) indicating that phosphorylation increases the ease with which helical structure forms with increasing TFE concentration (the $\Delta\Delta G_0$ is equal to -0.6±0.2 kcal/mole). However, when we quantified the helical content of unphosphorylated and phosphorylated peptides in the presence of maximal TFE (Fig. 4.4 *C*), the phosphorylated peptide did not reach the same level of helicity as that of the unphosphorylated peptide (Table 4.2), giving rise to an apparent lowering of the folding free energy. Collectively, these findings indicate that phosphorylation interferes with TFE-induced helix formation in the core region residues.

	***m*-value (cal/mol·TFE)**	**ΔG_0 (cal/mol)**	**$[\theta]_U$ (198nm)**	**$[\theta]_F$ (198nm)**	**$[\theta]_U$ (222nm)**	**$[\theta]_F$ (222nm)**
Unphosphorylated	-55±2	1000±200	-12±1	-4.9±0.2	-0.8±0.2	-4.3±0.2
Phosphorylated	-48±4	360±90	-14.9±0.9	-7.3±0.9	-1.1±0.5	-4.2±0.7
Full-length IDR	-82±20	1500±300	-9.3±0.5	-3.6±0.3	-2.1±0.4	-4.8±0.2

Table 4.1. Fit parameters from TFE-induced folding of Syt 1 IDR constructs.

When examining the CD absorption profile of IDPs, another useful parameter in assessing structure is the 222 nm/200 nm ratio, which can help discern whether or not a

particular IDP is in a more compact pre-molten globule state or in a disordered coil state.[175, 178] When we compare the full-length Syt 1 IDR with the unphosphorylated and phosphorylated core region peptides, a trend becomes apparent (Fig. 4.4 *E*).

In the full-length IDR, the 222 nm/200 nm ratio is consistent with its predicted chimeric behavior existing in a region of space between purely coil and purely pre-molten globule, which is consistent with the FRET measurement of compactness (Fig. 4.2) and MD simulation sampling of extended coils (Fig. 4.3). However, when the flanking positive and negative regions are largely removed to generate a peptide that includes core region residues, the 222 nm/200 nm ratio is shifted more definitively in the direction of coil disorder and phosphorylation shifts the ratio further still (Fig. 4.4 *E*). This indicates that the central region, where positive and negative charges are more evenly distributed, imparts more coil behavior and phosphorylation enhances that behavior. However, with the flanking densely positive and densely negative regions, structural states intermediate between pre-molten globule and coil are preferred.

This suggests that, if the central region transitions to structural ensembles containing more ordered elements, the flanking charged regions that promote sampling of compact structural states may exert regulatory control over accessibility of those core region conformers. TFE-induced folding of full-length Syt 1 IDR has a larger ΔG_0 of 1.5±0.3 kcal/mole compared to the unphosphorylated core peptide, which supports the hypothesis that flanking basic and acidic termini can limit core region folding (Fig. 4.4 *D*, orange; Table 4.1). Also consistent with this hypothesis is the observation that estimated helical content of the full-length peptide is statistically indistinguishable from that of the unphosphorylated core peptide (fraction helical in 60% v/v TFE, full-length 0.45±0.02 vs

core 0.42±0.01). However, because we did not measure TFE-induced folding of the flanking basic and acidic regions in isolation, we cannot conclusively say the core region residues are the only residues that become helical in the full-length IDR peptide. Regardless, mechanisms where structural compaction in IDPs limit accessibility of internal regions of sequence have been proposed in other systems[179, 180] and may still bear relevance to the Syt 1 IDR structure within the context of the whole protein embedded in a synaptic vesicle.

	Fraction helix (experiment)	**Fraction total helix (MD sim)**	**Fraction α-helix (MD sim)**	**Fraction 3_{10}-helix (MD sim)**
Unphosphorylated	0.42±0.01	0.16	0.12	0.04
Phosphorylated	0.37±0.03	0.09	0.04	0.05

Table 4.2. Estimated helical content of unphosphorylated and phosphorylated Syt 1 IDR core regions under maximum (60% v/v) TFE concentration (experiment) or lowest (ε=20) dielectric constant (MD sim).

4.4.3 – Implicit solvent simulations show increased helicity in unphosphorylated core residues upon reducing dielectric constant

While CD provides global information on Syt 1 IDR core region structure, we also sought atomistic details from MD simulation similar to our FRET experiments.[165, 166] In the *in vitro* studies, both unphosphorylated and phosphorylated peptides formed helical structures in the lower dielectric environments of TFE-water mixtures (Fig. 4.4). While TFE is known to stabilize helical structure[181, 182], helical conformers would be a reasonable form of secondary structure for this IDR to adopt. Several of its residues have high to intermediate helical propensity.[183, 184] Moreover, several oppositely charged residues in the core region are positioned in *i*, *i*+4 locations, suggesting a role for salt

bridges in the stabilization of helical structure. In the low dielectric environment of the interfacial region of the bilayer, there should be a higher propensity to form such salt bridges. We tested this hypothesis using an implicit solvent, first simulating the IDR core region at ε=80 (similar to that of water) and then with ε=20 (closer to that of the interfacial region). In both simulations, salt concentration was 100 mM to more closely mimic *in vivo* conditions. At ε=80, the unphosphorylated peptide largely occupied coil or turn secondary structure (Fig. 4.5 *A* and Table 4.5), consistent with CD measurements of the synthetic peptide (Fig. 4.4). When the unphosphorylated peptide was simulated again with ε=20, there was a large increase in the probability of helical conformers (Fig. 4.5 *B* and Table 4.5). When examining the 1.0 μsec trajectory, the helical conformers formed were often stabilized by lysine-aspartic acid salt bridges (Fig. 4.5 *C*). In particular, K111 and D115 as well as K104 and D108 were among the longer-lived interactions within continuous helices (Fig. 4.5 *D*) though several other electrostatic interactions were also observed (Fig. 4.5 *E*), not all of which were helix stabilizing.

4.4.4 – Implicit solvent simulations indicate Thr112 phosphorylation disrupts formation of helical conformers in IDR core residues

To assess the impact of phosphorylation on the structural propensity of the IDR core residues, we repeated the implicit solvent simulations on a phosphorylated peptide of identical sequence. At ε=80 and 100 mM salt, the peptide was largely disordered and residual helical content was minimal (Fig. 4.6 *A* and Table 4.5). When the phosphorylated peptide was simulated at ε=20, helical structure formed but to a lesser degree compared to the unphosphorylated peptide (Fig. 4.6 *B* and Table 4.5) which is consistent with CD

data (Fig. 4.4). This difference is more pronounced when the total helical content is separated back into α- and 3_{10}-helices. The unphosphorylated peptide's total helical content is largely dominated by α-helix (Table 4.2). In the phosphorylated peptide, the ratio of α- to 3_{10}-helices is closer to 50:50 with a slight preference toward the 3_{10}-helix (Table 4.2).

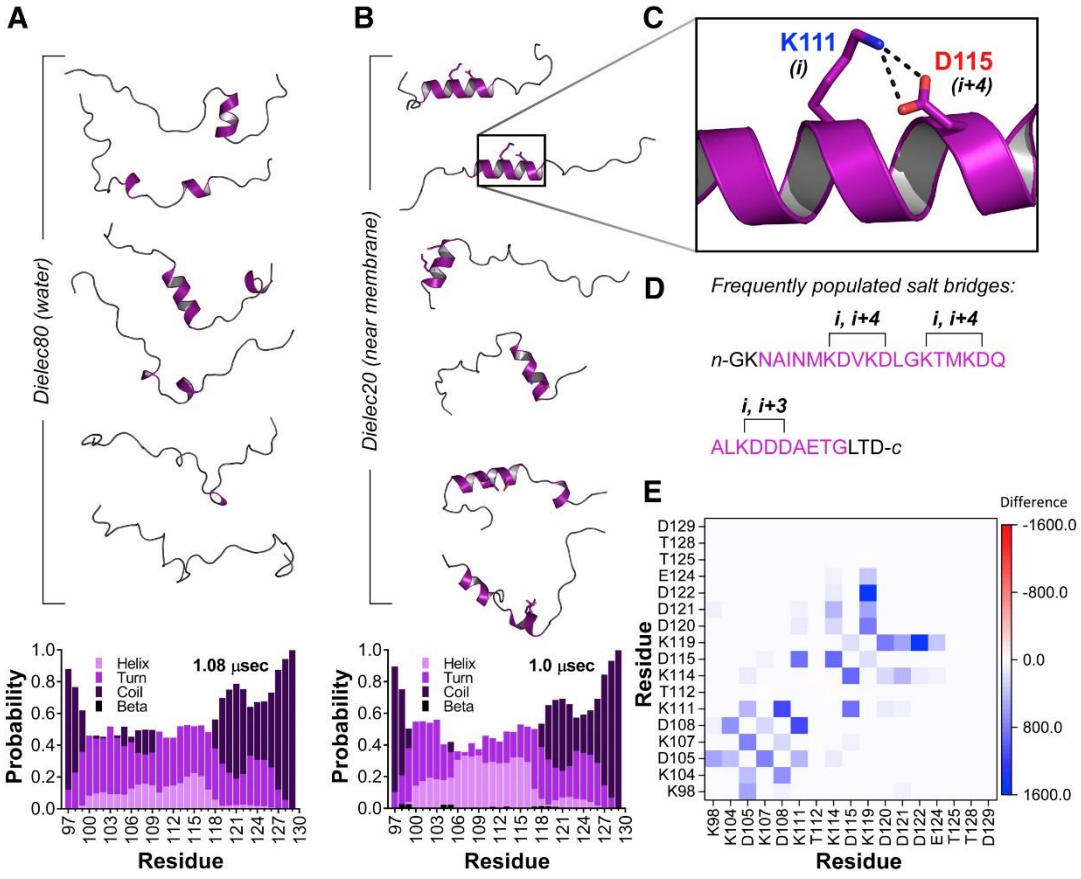

Figure. 4.5. Simulated structural ensemble of unphosphorylated Syt 1 IDR core region. (*A*) Representative conformers of the peptide at a dielectric constant of 80 with secondary structure probability per residue is shown below (histogram). (*B*) Representative conformers of the peptide at a dielectric constant of 20 and secondary structure probability per residue shown below (histogram). Note increase in helical probability. (*C*) Close-up of example salt bridge that forms to stabilize helix in low dielectric environment. (*D*) IDR core region sequence showing sample salt bridge interactions found in helices at a dielectric constant of ε=20. Residues highlighted in purple occupy helical secondary structure. (*E*) Heat map showing change in all salt bridge interactions in going from dielectric 80 to dielectric 20. An increased frequency of interaction between two charged residues is indicated by gradations of blue whereas decreased interactions are indicated by gradations of red. Note increased frequency of *i*, *i*+3 and *i*, *i*+4 contacts.

The region of highest helicity in the unphosphorylated peptide was significantly disrupted by phosphorylation, as indicated by a sharp rise in turn/coil secondary structures in the residues immediately adjacent to and including Thr112 (Fig. 4.6 *A* and *B*). Similar to the unphosphorylated peptide under ε=20 simulation conditions, several salt bridges were prevalent during the course of the 1.0 μsec trajectory. In this case, however, the salt bridges involved the phosphate and several different lysine residues, none of which were in optimal positions for stabilizing helical structure (Fig. 4.6 *C* and *D*). For example, K111 (*i*, *i*+1), K107 (*i*, *i*+5), K104 (*i*, *i*+8) and K98 (*i*, *i*+14) all form salt bridges with the phosphate (Fig. 4.6 *E*). In addition to these unfavorable intramolecular interactions, phosphothreonine itself is thought to have a lower helical propensity which may further contribute to reduced helix formation.[185] Moreover, in other model helical peptides, phosphorylation events that are centrally localized in the helix are more destabilizing to structure.[186]

After examining unphosphorylated and phosphorylated peptides individually, we next considered them simultaneously as part of a thermodynamic cycle for the core region structural states (Fig. 4.7). Using the simulation-derived secondary structure probabilities, we calculated energetics for structural transitions that could then be compared to those derived from CD. Overall, the energetic trends were similar to those extracted from TFE folding experiments. Transitioning from high to low dielectric constant was energetically favorable for both unphosphorylated and phosphorylated states. However, like the CD experiments, the phosphorylated peptide had a more energetically favorable transition ($\Delta\Delta G_{U\rightarrow F}$ = -0.88 kcal/mole) despite having less overall

helicity in each dielectric environment. In both dielectric environments, phosphorylation was energetically unfavorable for helix formation.

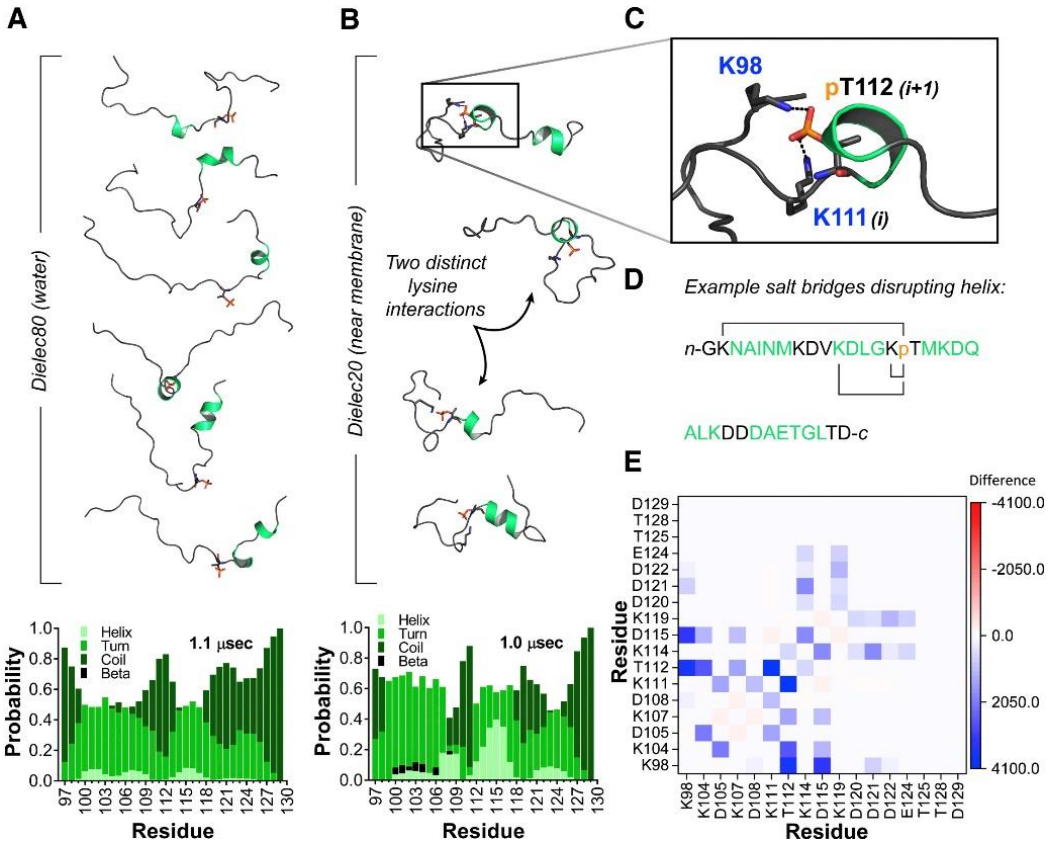

Figure. 4.6. Structural ensemble of phosphorylated Syt 1 IDR core. (*A*) Representative conformers of the peptide at a dielectric constant of 80. Secondary structure probability per residue is shown below (histogram). Note reduced helicity compared to unphosphorylated peptide. (*B*) Representative conformers of the peptide at a dielectric constant of 20. Secondary structure probability per residue is shown below (histogram). Note that residues 107-118, which formed a continuous helix in the unphosphorylated peptide, is disrupted by coil secondary structure immediately adjacent to phosphorylation site. (*C*) Close-up of example salt bridge that distorts peptide backbone and limits helix formation in low dielectric environment. (*D*) IDR core region sequence showing salt bridge interactions inhibitory toward helices at a dielectric constant of ε=20. Green residues are those that occupy helical conformers during the trajectory. (*E*) Heat map showing change in salt bridge interactions in going from dielectric 80 to dielectric 20. An increased frequency of interaction between two charged residues is indicated by gradations of blue whereas decreased interactions are indicated by gradations of red.

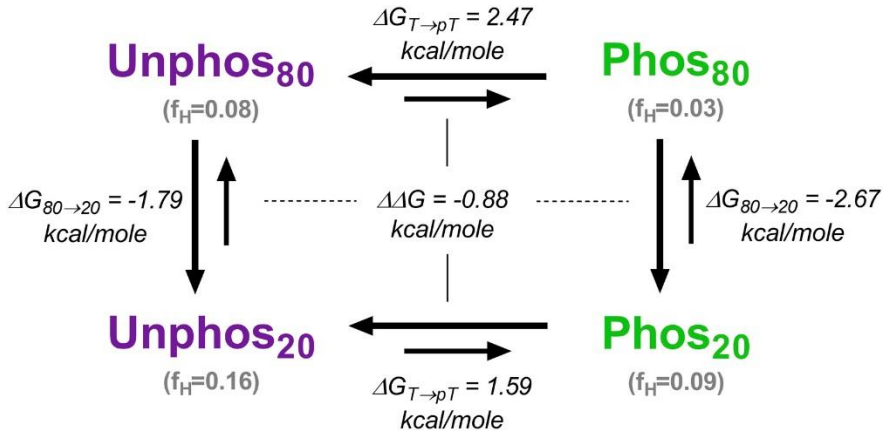

Figure. 4.7. Simulation-derived thermodynamic cycle of structural states adopted by Syt 1 IDR core region (residues 97-129). Folding free energies of each structural state (Unphos80, Phos80, Unphos20, Phos20) were determined by applying the equation $\Delta G = -RT\ln(f_H/f_{NH})$, where f_H is the fraction of helical conformers and f_{NH} is the fraction of non-helical conformers. The free energy changes associated with transitioning between each state were then calculated and are shown in the figure above. Note that for both unphosphorylated and phosphorylated peptides, lowering the dielectric constant is favorable toward helix formation. The fraction of helical content is shown parenthetically below each structural state. At a given dielectric constant, phosphorylation is generally unfavorable toward helix formation. Note that the apparent $\Delta\Delta G$ computed from MD structural data is of similar magnitude to that derived from CD.

4.4.5 – Sampling compact structural states in full-length IDR simulations can limit core region folding

After obtaining data on full-length IDR compaction and helix formation in the core region residues separately, we next considered how sampling of compact structural states may influence core region helix propensity. We simulated the full-length IDR sequence at ε=80 and ε=20 at high (100 mM) and low (10 mM) salt concentrations and compared their structural behaviors. As suggested by the CD TFE-folding data, inclusion of the flanking N- and C-terminal regions seem to reduce probability of core region helix formation (Fig. 4.8A). This conclusion stems in part from comparison of the helix probability seen for the core region alone (Fig. 4.5A) with the same core region residues

within the full-length sequence. Comparison of core region and full-length peptides at ε=20 was less supportive in this regard (Fig. 4.8B). However, when full-length IDR peptides at ε=80 and ε=20 are examined at high and low salt concentrations, the correlation becomes more apparent. At ε=80 and 100 mM salt, there is a decrease in the sampling of compact structural states (Fig. 4.12) which coincides with a modest increase in helix probability in the core region residues (Fig. 4.13). The decrease in sampling of compact structural states is still subtle at ε=80 (Fig. 4.8C-E), which may in part explain why we did not measure a significantly different change in FRET efficiency under increasing salt concentration (Fig. 4.14A). This same salt-induce change was, however, more pronounced in simulations performed at ε=20 (Fig. 4.12 and 4.13). In general, the IDR more readily samples compact/hairpin structural states at ε=20 as evidenced by R_g and polymer chain behaviors (Fig. 4.8C-E).

While salt influences compaction at ε=20, another likely competitor for intramolecular interactions within the IDR is intermolecular interactions with the lipid bilayer. Previous EPR spin label accessibility measurements indicate that the basic N-terminal region can partially penetrate into a POPC:POPS lipid bilayer when reconstituted.[187] This suggests that the basic N-terminus is the IDR region most likely responsible for direct lipid binding. To assess this potential contribution, we measured membrane binding of IDR peptides representing either the full-length IDR sequence or core region (residues 97-130) to large unilamellar vesicles whose lipid composition mimics that of a synaptic vesicle. Consistent with our hypothesis, only the full-length IDR peptide that includes the basic N-terminus was found to bind membrane (Fig. 4.14B and Fig. 4.14C, Supporting Materials and Methods Chapter 4). Though the core region

undergoes structural changes when the full-length IDR is associated with a synaptic vesicle mimic[49], these co-sedimentation results suggest that the basic N-terminus is largely responsible for lipid binding. Because of this binding and the high local concentration of acidic phospholipids normally experienced by Syt 1 in the bilayer, lipid binding may be able to compete with intramolecular association of N- and C-terminal regions. Collectively, the simulations of the full-length IDR and binding results point to the possibility of a structural interplay between IDR compaction and occupation of transient helical conformers in the core region residues.

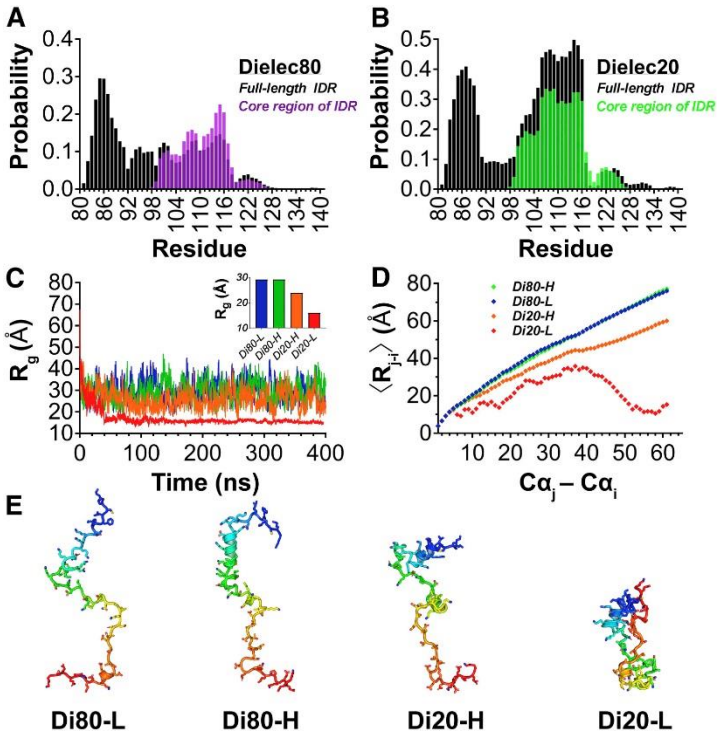

Figure. 4.8. Structural impact of dielectric constant and salt on full-length Syt 1 IDR. Comparison of core region helix propensity in full-length (black histograms) and core region (purple or green histograms) peptides at (A) ε = 80 and (B) ε = 20 when salt concentration is 100 mM. The full-length IDR shows (C) variable R_g depending on simulation conditions where L and H represent 10 mM and 100 mM salt. Inset shows average R_g for entire trajectory. The average inter-residue distance for all conditions is shown in (D) along with representative conformers in (E). Structures in (E), N- to C-terminus, are color-coded blue to red.

4.5 – Discussion

In the current study, we probed the structural propensity of Syt 1's IDR, examining the dependence on dielectric constant and phosphorylation. These are important goals because, in the former case, the type of intrinsic disorder exploited by Syt 1 (strong polyampholyte) is naturally subjected to a low dielectric environment by virtue of being embedded in a lipid bilayer. This makes the impact of electrostatics more pronounced, enhancing any effect charged residues exert on local protein structure. In the latter case, phosphorylation at Thr112 has been shown to potentiate synaptic transmission in hippocampal neurons[158], but a molecular rationale for this cellular change is lacking.

With regard to dielectric constant, we found both through experiment (Fig. 4.4) and simulation (Fig. 4.5 and Fig. 4.6) that reduced dielectric constant promotes helix formation in core region residues. These results are of biological interest because we previously reported that the IDR remains largely disordered when associated with a lipid bilayer.[49] While we could not rule out the possibility of the IDR peptide undergoing oligomerization on the membrane surface, which could alter structural propensity, another possibility for lack of structure is lipid-induced unfolding. While the bilayer provides a constant low dielectric environment that should promote folding in the core region of the IDR, other chemical factors undoubtedly still influence secondary structure. The vesicles we used previously were composed of nine lipid species in specific mole fractions designed to mimic the outer leaflet of the native organelle. IDR interaction with this cocktail of lipid headgroups may favor a more disordered state as part of a regulatory mechanism. Many IDPs and IDRs that interact with membranes are sensitive to lipid composition and have their secondary structure altered by different lipid mixtures.[188] In

the single-pass integral membrane protein phospholamban, for example, this type of lipid modulation of structure has been directly linked to altered functionality. In response to cationic phospholipids, phospholamban's cytoplasmic domain becomes more disordered and this structural shift consequently decreases the protein's inhibitory potency towards the sarcoplasmic reticulum Ca^{2+}-ATPase (SERCA) that it normally regulates.[189]

Another possible explanation for the full-length IDR being more disordered when associated with a membrane could relate to an increased tendency to sample compact structural states (Fig. 4.8). Increased population of compact states due to low dielectric constant was inversely correlated with core region helix formation. This suggests that, in addition to possible lipid-regulation of folding, there may also be suppression of structure from intramolecular interactions between basic and acidic termini.

Another finding of biological interest in this study is the site of apparent helix formation (Fig. 4.5 and Fig. 4.6). The position where helical structure forms is intriguing because it correlates with a disorder-based predicted binding site (Fig. 4.1). In a study by de Jong et al[158], residues 109-116 were identified as being inhibitory toward synaptic release when unphosphorylated and permissive when phosphorylated. Our results suggest a potential structural model for this altered function. In the absence of phosphorylation, these residues (which form part of a predicted binding site) have a tendency to form an α-helix which may be necessary for association with an as yet unidentified partner. When phosphorylated, however, there is a local disruption of helix formation that may then limit binding. This mode of regulation for binding interactions is common. There are examples in other IDP systems where single phosphorylation events similarly break helices[190] or, in extreme cases, sequester a partially helical binding site into a different

form of secondary structure.[191] In some instances, phosphorylation does not have a significant impact on secondary structure, but rather shifts local and global disordered ensembles to regulate interactions.[179] And in other cytoplasmic regions of integral membrane proteins, phosphorylation can rearrange salt bridge interaction networks to switch local structure.[177] Alternatively, a core region helix may form part of an interaction surface that contributes to Syt oligomerization, in which case helix disruption could alter the oligomeric state.[187] These possibilities are also worth considering in the context of other synaptotagmin isoforms. In synaptotagmin 7, for instance, phosphorylation of Ser103 (a residue in its putative IDR) also potentiates exocytosis, in this case, of insulin.[192] Such a similarity suggests that findings on Syt 1's IDR may help inform future study of other synaptotagmin IDR isoforms.

The structural differences between unphosphorylated and phosphorylated core region residues in Syt 1 also provide potential insight into a recent report of the IDR interacting with the PH domain of dynamin 1, a GTPase known for facilitating fission of vesicles that are being endocytosed.[150] That study showed that a T112E missense mutation in the IDR abolishes the binding interaction, suggesting that phosphorylation is a regulatory mechanism for the interaction. However, that study also showed that deletion of residues 117-119 (ALK) or 111-119 (KTMKDQALK) did not disrupt binding of the IDR to the PH domain. Given that the KTMKDQALK residues are among those that adopt helical conformers in the absence of phosphorylation, the binding interaction with dynamin's PH domain may not involve a specific structured helix in the Syt 1 IDR, but instead rely on a mechanism involving local electrostatic interactions.

On a more global level, the full-length IDR seems to sample collapsed structures (Fig. 4.2 and Fig. 4.8), consistent with both computational predictions[155, 156] and effects of site-directed mutagenesis on the basic and acidic regions of the IDR in vesicle-vesicle docking/fusion assays.[157] The other two sites of phosphorylation within the IDR (Thr125 and Thr128), are located in the acidic C-terminal region, suggesting their placement could enhance compaction.[193, 194] This may be particularly relevant in a low dielectric environment like that near the membrane surface, which already enhances occupation of compact conformations (Fig. 4.8).

When comparing the CD and MD estimates of helical structure in lowest dielectric environments, there are differences in the absolute fractions determined by each method. While CD and MD do not quantitatively converge, they both indicate that phosphorylation is inhibitory toward helix formation. The lack of quantitative agreement is not surprising, since the MD simulations consider only the change in dielectric constant, ignoring other helix-stabilizing properties of TFE that involve alterations of peptide-water interactions which would likely require explicit treatment of both TFE and water.[195-197]

4.6 – Conclusions

We examined the structural impact of dielectric constant and phosphorylation on Syt 1's IDR. The general sequestration of positive and negative residues at the N- and C-termini promotes sampling of collapsed structural states in aqueous solution (Fig. 4.2) although compact states are more minor compared to extended coil states (Fig. 4.3). The more sequence-diverse core region of the IDR, where charge is more uniformly

distributed, can form short helices when subjected to TFE (Fig. 4.4) or a simulated low dielectric environment (Fig. 4.5). Phosphorylation at Thr112, however, disrupts this helical tendency both folding experiments in TFE and in implicit solvent simulations (Fig. 4.4 and Fig. 4.6). In addition to phosphorylation alteration of folding (Fig. 4.7), the helical propensity of these same core region residues may also be limited in the full-length sequence as compact structural states are sampled (Fig. 4.8). Collectively, these results argue that the Syt 1 IDR is influenced by the low dielectric environment near the membrane surface and that altered charge distribution through post-translational modification can exert local alterations to its structural propensity.

4.7 – Supporting Material

4.7.1 – Calculation of Compaction Predictor κ

Polyampholytes are a class of intrinsically disordered proteins that have a high frequency of positively and negatively charged amino acids. The Syt 1 IDR falls into this category with ~56% of the amino acids being charged. In the predictive framework developed by Das and Pappu[155], a polyampholyte IDP will adopt linear/extended or compact/globular random coils, depending on the distribution of charged side chains in sequence. The physical observable for measurement of compaction in this model is radius of gyration R_g, which is correlated with a computable predictive parameter called κ. The κ term depends on two main terms: σ, the overall charge asymmetry, and σ_i, the local charge asymmetry within a "blob" of linear sequence (usually 5 or 6 residues). How κ was calculated and used to predict the possible compactness of the Syt 1 IDR is shown below (Table 4.3 and Table 4.4). The fraction of positive ($f+$) and negative ($f-$) within each 5 or 6 residue *i*-th blob was used to determine σ_i:

$$\sigma_i = (f_+ - f_-)_i^2/(f_+ + f_-)_i \quad \text{Eq. 4.6}$$

Each blob's charge asymmetry was compared to the IDR's overall charge asymmetry (Eq. 4.7) to determine the squared deviation (δ) using Eq. 4.8:

$$\sigma = (f_+ - f_-)^2/(f_+ + f_-) \quad \text{Eq. 4.7}$$

$$\delta = \sum_{i=1}^{Nblob} (\sigma_i - \sigma)^2/Nblob \quad \text{Eq. 4.8}$$

Where N_{blob} is the number of blobs in the IDR sequence. Next, δ was determined for a rearranged primary sequence of the IDR amino acids that represents the maximum possible charge segregation. This value (δ_{max} = 0.503) was then used to calculate κ, the parameter used to predict degree of compaction based on Figure 2 from[155]:

$$\kappa = (\delta/\delta_{max}) \quad \text{Eq. 4.9}$$

4.7.2 – IDR peptide binding to large unilamellar vesicles (LUVs)

To test if the IDR core region bound LUVs whose lipid composition mimicked that of a synaptic vesicle, we prepared liposomes using our previously developed lipid cocktail. This lipid composition was designed to mimic the outer leaflet of a synaptic vesicle organelle: 1-stearoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (18:0-18:1 PE); 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (16:0-18:1 PE); 1-stearoyl-2-docosahexaenoyl-*sn*-glycero-3-phosphoethanolamine (18:0-22:6 PE); 1-stearoyl-2-docosahexaenoyl-*sn*-glycero-3-phosphoserine (18:0-22:6 PS); 1-stearoyl-2-oleoyl-*sn*-glycero-3-phosphoserine (18:0-18:1 PS); 1-stearoyl-2-arachidonoyl-*sn*-glycero-3-phospho-(1'-myo-inositol-4',5'-bisphosphate) (18:0-20:4 PI(4,5)P2); 1,2-dioleoyl-*sn*-glycero-3-phospho-(1'-myo-inositol-4',5'-bisphosphate) (18:1-18:1 PI(4,5)P2); 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoinositol (16:0-18:1 PI); cholesterol. For further details preparation, see references[49, 198].

Co-sedimentation with synaptic vesicle mimic LUVs were performed by incubating 15 μM of either the full-length IDR (a synthetic peptide corresponding to residues 80-156) or the IDR core with part of the acidic C-terminus (residues 97-130)

with LUVs for 30 minutes at 22 °C. The concentration of LUVs for each sample ranged from 0-6 mM and total sample volume was 60 μL. After incubation, samples were subsequently spun down in a TLA 100 rotor for 1 hour at 22 °C to pellet LUVs and bound peptide. Depletion of peptide from the resultant supernatants was then used to assess membrane binding, where 20 μL aliquots were taken from the supernatants and run on SDS-PAGE (Fig. 4.14).

4.7.3 – Nonlinear fitting to two-state folding model

To determine the free energy of folding, we fit CD signal change to a two-state folding model. In a two-state model, there is equilibrium between unfolded (U) and folded (F) states of the protein:

$$K_{eq} = [U]/[F] \qquad \text{Eq. 4.10}$$

The equilibrium constant is directly related to free energy through the Gibbs equation (Eq. 4.11):

$$\Delta G = -RT\ln(K_{eq}) \qquad \text{Eq. 4.11}$$

In cases where protein folding equilibria are monitored as a function of osmolyte, co-solvent (such as trifluoroethanol, TFE), or denaturant it can be assumed that the equilibrium between unfolded and folded states is linearly dependent on the added solute. This is an assumption of the linear extrapolation method[199] which we employ here:

$$\Delta G_{\%TFEv/v} = m \times (\%TFE\ v/v) + \Delta G_0 \quad \text{Eq. 4.12}$$

Where $\Delta G_{\%TFEv/v}$ is the free energy of the folded-unfolded equilibrium at some specified volume percentage of TFE, m is the proportionality constant by which the folding equilibrium free energy changes with addition of TFE, and ΔG_0 is the equilibrium folding free energy in the absence of any TFE. In the case of circular dichroism measurements, there is a change in absorption of circularly polarized light with a change in structure. This signal change, when plotted as a function of % TFE v/v, creates a sigmoidal transition from which the fraction of unfolded species (f_U) can be represented:

$$f_U = ([\theta]_{\%TFEv/v} - [\theta]_{min})/([\theta]_{max} - [\theta]_{min}) \quad \text{Eq. 4.13}$$

Where $[\theta]_{\%TFEv/v}$ is the absorption of circularly polarized at different volume percentages of TFE, $[\theta]_{min}$ is the absorption in the absence of TFE where there is no folded structure, and $[\theta]_{max}$ is the absorption where there is constant helical structure with addition of TFE. By calculating f_U at all volume percentages of TFE, equilibrium constants ($K_{eq,\ \%TFE}$) and folding free energies ($\Delta G_{\%TFEv/v}$) can be calculated for each data point:

$$K_{eq,\ \%TFE} = f_U/(1-f_U) \quad \text{Eq. 4.14}$$

$$\Delta G_{\%TFEv/v} = -RT\ln(K_{eq,\ \%TFE}) \quad \text{Eq. 4.15}$$

A plot of $\Delta G_{\%TFEv/v}$ from Eq. 4.15 versus % TFE v/v yields a linear correlation upon which linear regression analysis can be performed to obtain the slope (m) and intercept (ΔG_0) of Eq. 4.12 above.

While the above approach provides the desired value for the folded-unfolded equilibrium in the absence of TFE, we further analyzed the data with a non-linear least squares regression fitting of the raw data so that we could simultaneously fit two wavelengths reporting on the transition (198 nm and 222 nm) to a single set of m-values and ΔG_0. The signal measured by circular dichroism is dependent on the fraction of unfolded protein. To model the folding transition, we use the following expression to describe the experimental data:

$$[\theta]_{fit} = f_U \text{ x } ([\theta]_F - [\theta]_U) + [\theta]_U \qquad \text{Eq. 4.16}$$

Where $[\theta]_{fit}$ is the modeled signal, $[\theta]_F$ is the absorption for maximally folded protein, and $[\theta]_U$ is the absorption for unfolded protein. In the non-linear least squares regression fitting, both $[\theta]_F$ and $[\theta]_U$ are fit parameters. However, absorption values taken directly from the raw data both in the absence of TFE (for $[\theta]_U$) and when there is no additional gain of folded structure upon addition of TFE (for $[\theta]_F$) can be used as close starting values. These parameters did not change substantially during the fitting process (Table 4.1). Since f_U is directly related to the folded-unfolded equilibrium constant (Eq. 4.14) and the equilibrium constant is related to free energy in the presence of TFE (Eq. 4.15) which in turn is linearly related to % TFE v/v (Eq. 4.12), then simple rearrangement and

substitution of equations will give a modified expression for $[\theta]_{fit}$ in terms of the four fit parameters used for non-linear regression:

$$[\theta]_{fit} = ((e^{-(m*(\%TFEv/v)+\Delta G0)/(RT)})/(e^{-(m*(\%TFEv/v)+\Delta G0)/(RT)} + 1)) \times ([\theta]_F - [\theta]_U) + [\theta]_U \quad \text{Eq. 4.17}$$

With Eq. 4.17 used to calculate the theoretical absorption of circularly polarized light as a function of % TFE v/v, we used the solver function of Excel to minimize the sum of the square differences between experimental and calculated fit values by manipulating the four parameters of m, ΔG_0, $[\theta]_F$ and $[\theta]_U$ in a manner similar to that described previously.[147, 162] In addition to using absorption values from the raw data for $[\theta]_F$ and $[\theta]_U$ as described above, we also used m and ΔG_0 parameters obtained from the initial linear regression analysis of the linear-extrapolation method. Using these four starting parameters in Eq. 4.17 already does a reasonable job of describing the data. However, application of solver for further error minimization improves the total sum of the square differences between raw and fit values. After fitting the raw data sets, both the data and fit were normalized for side-by-side comparison in the main text.

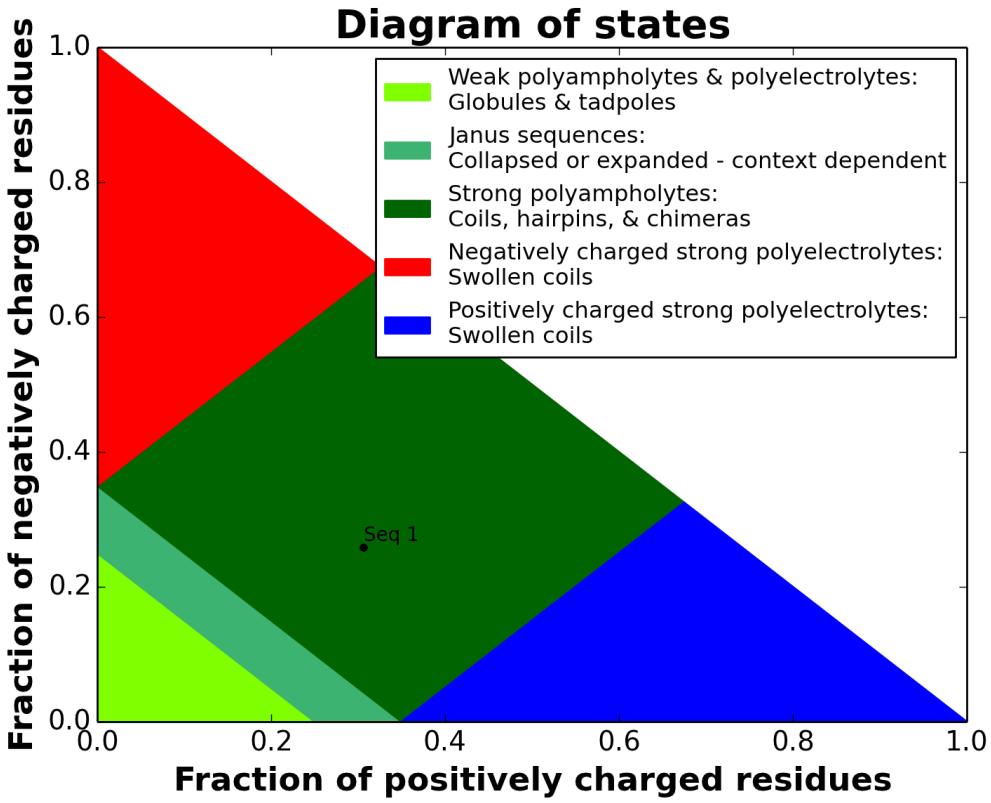

Figure 4.9. Position of Syt 1 IDR in sequence charge-dependent phase diagram of structural states. Note that location of the Syt 1 IDR, indicated by Seq 1 data point, suggests it likely samples coil and/or hairpin-like structural states in an environment whose dielectric constant is similar to water.

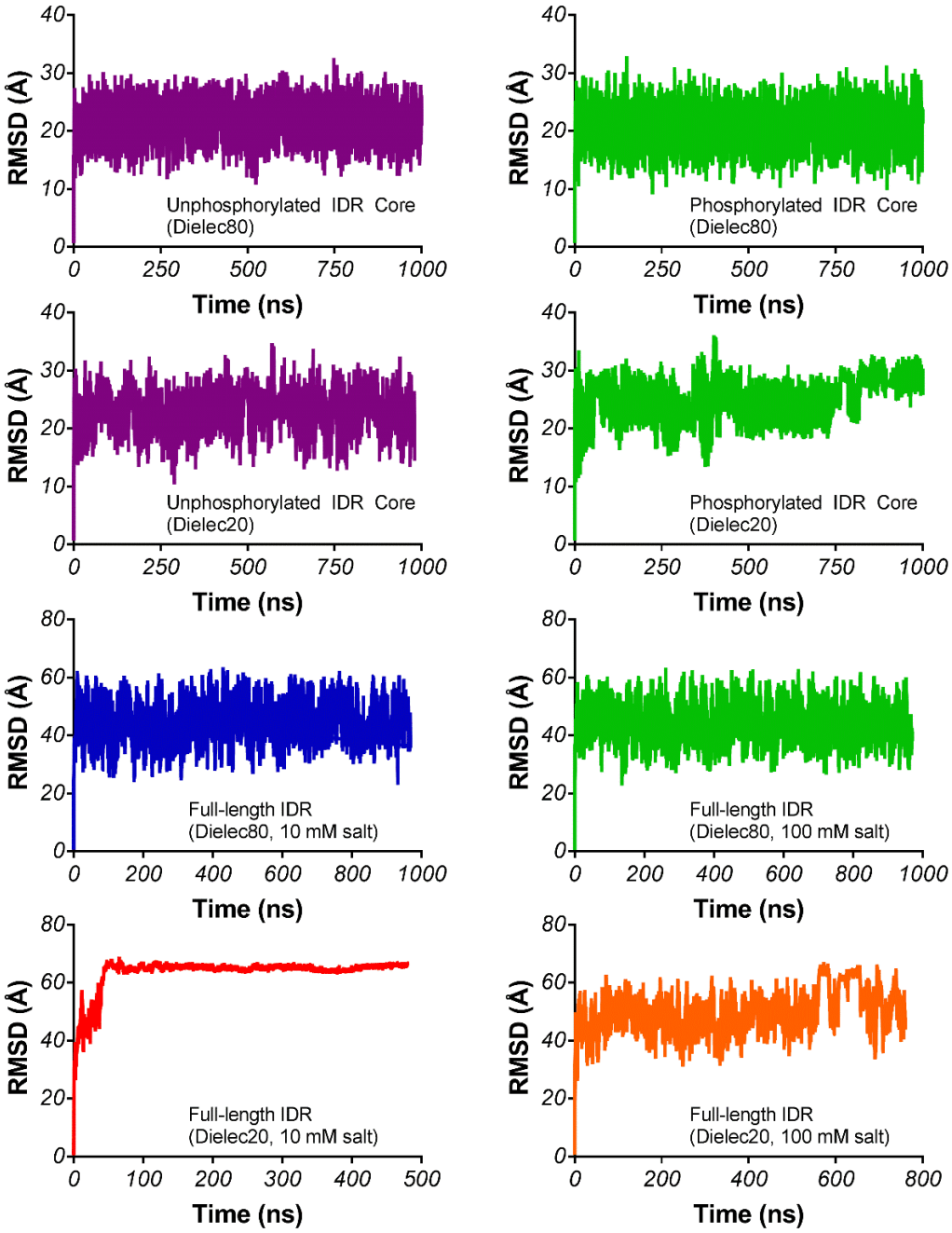

Figure 4.10. RMSD of Syt 1 IDR constructs used in MD simulations. Core region peptides (residues 97-129) in unphosphorylated (purple) and phosphorylated (green) states and at dielectric constants of both ε=80 and ε=20 are shown in top four panels. The full-length peptides (residues 80-141) at dielectric constant and salt concentration of ε=80 and 10 mM (blue), ε=80 and 100 mM (green), ε=20 and 10 mM (red), ε=20 and 100 mM (orange) are shown in the bottom four panels. Note that each peptide reaches equilibrium rapidly.

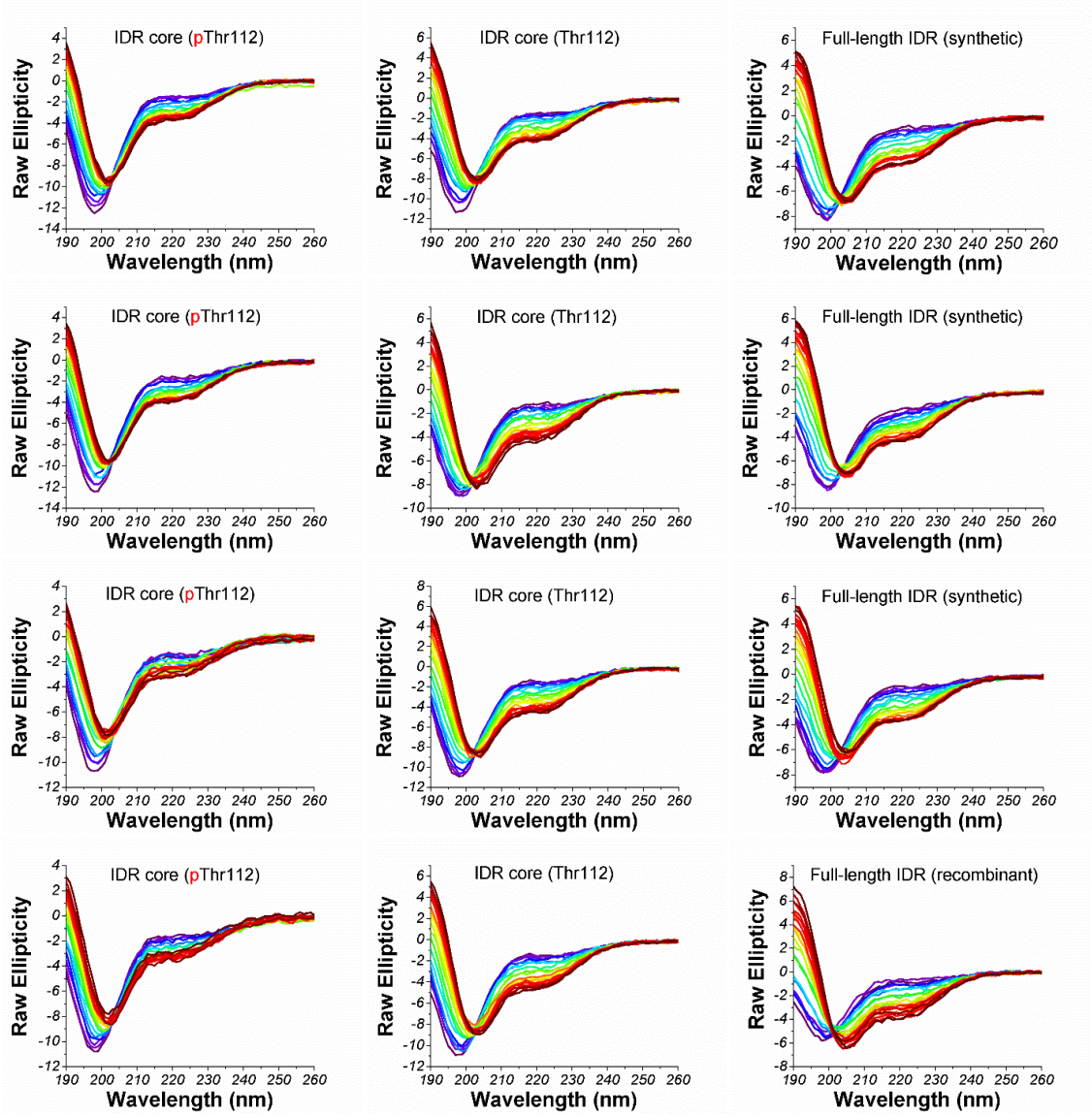

Figure 4.11. TFE-induced folding of Syt1 IDR. The left column shows partial folding of the IDR core residues (97-130) with Thr112 phosphorylated. Note its restricted transition. The middle column shows partial folding of IDR core residues (97-130) with Thr112 in its unphosphorylated state. The right column shows partial folding of the full length IDR (residues 80-140). Note that the top three panels in this column are replicates of the synthesized peptide and the bottom panel is the recombinantly expressed IDR that includes a C-terminal his-tag. Violet to red corresponds to a 0-60% v/v TFE range in increments of 3%.

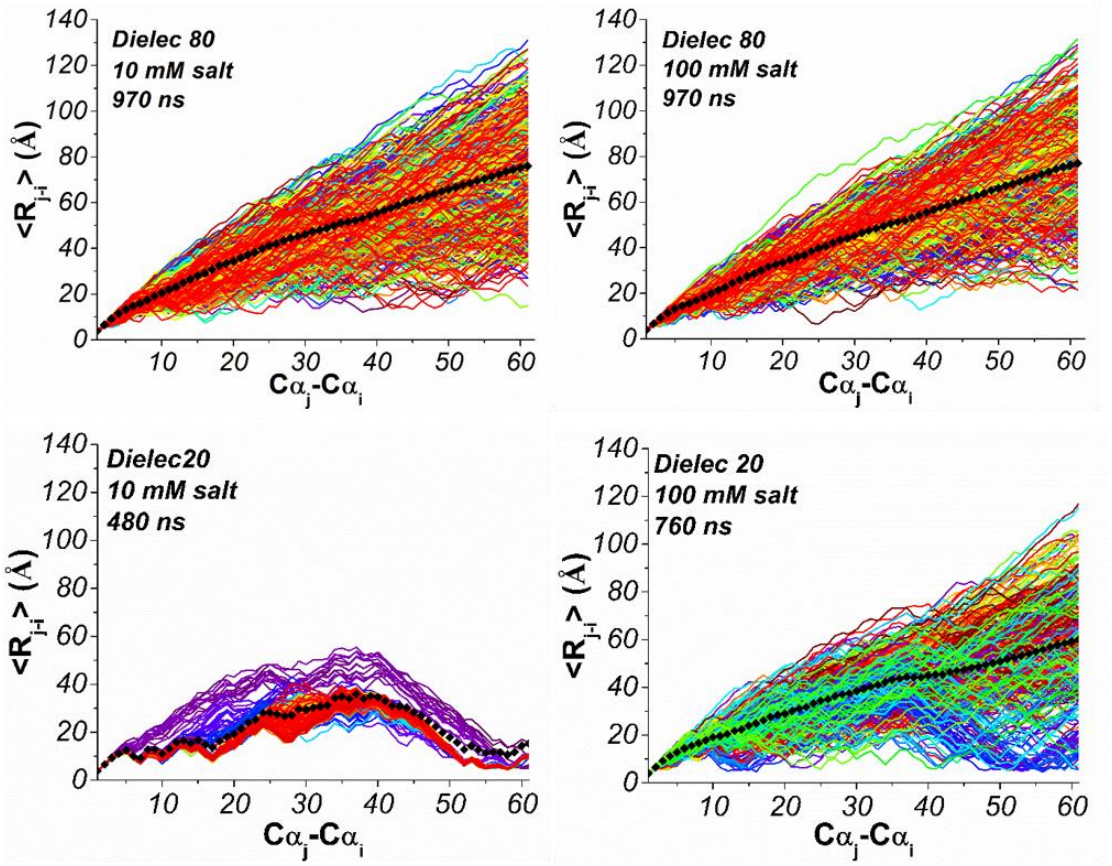

Figure 4.12. Simulated inter-residue distances of full-length Syt 1 IDR under varied dielectric constant and salt conditions during the course of each trajectory. Black diamonds indicate the average structure.

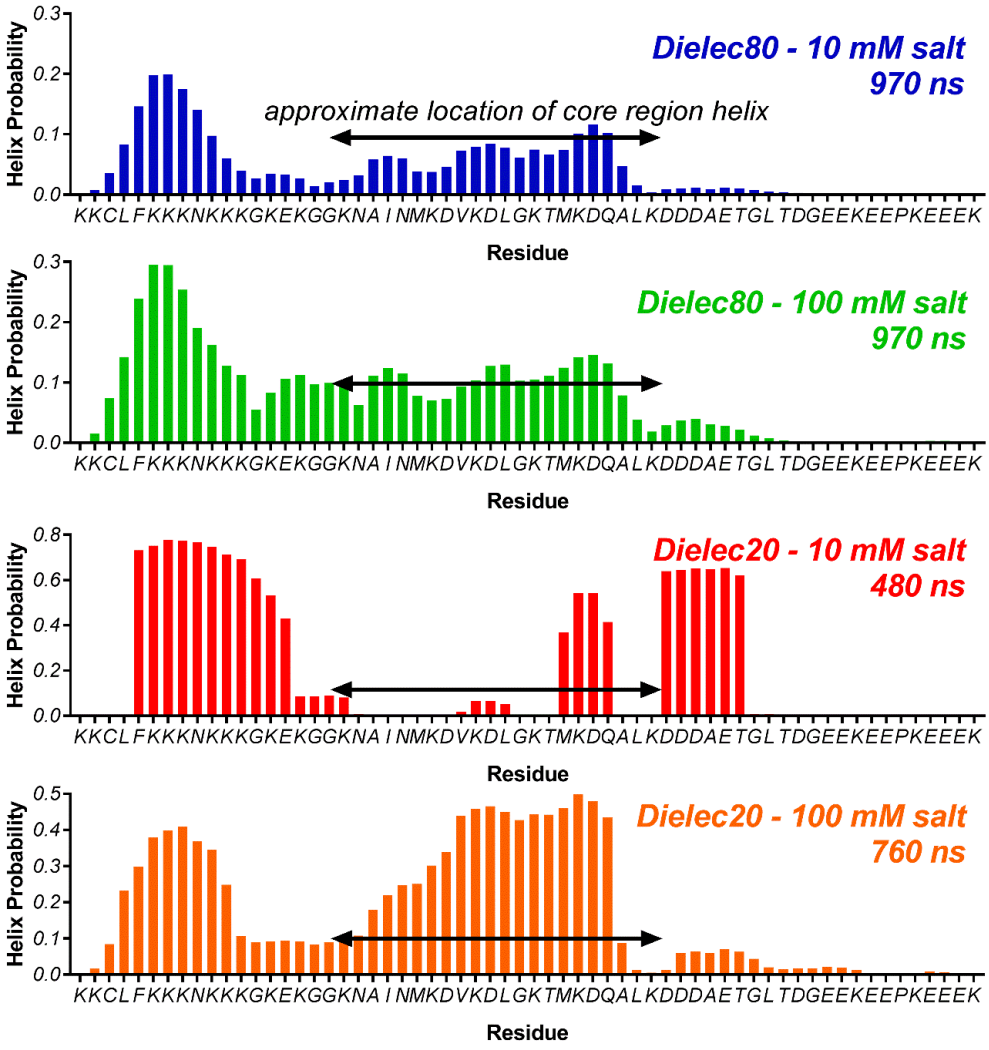

Figure 4.13. Helix probability histograms (determined as described in main text) for full-length Syt 1 IDR sequence. Black arrows for all four histograms indicate approximate core residue region. In all cases, arrows align with 0.1 on the probability axis to aid comparison. Also noteworthy is that, under all four conditions, helical content in the acidic C-terminus was markedly low.

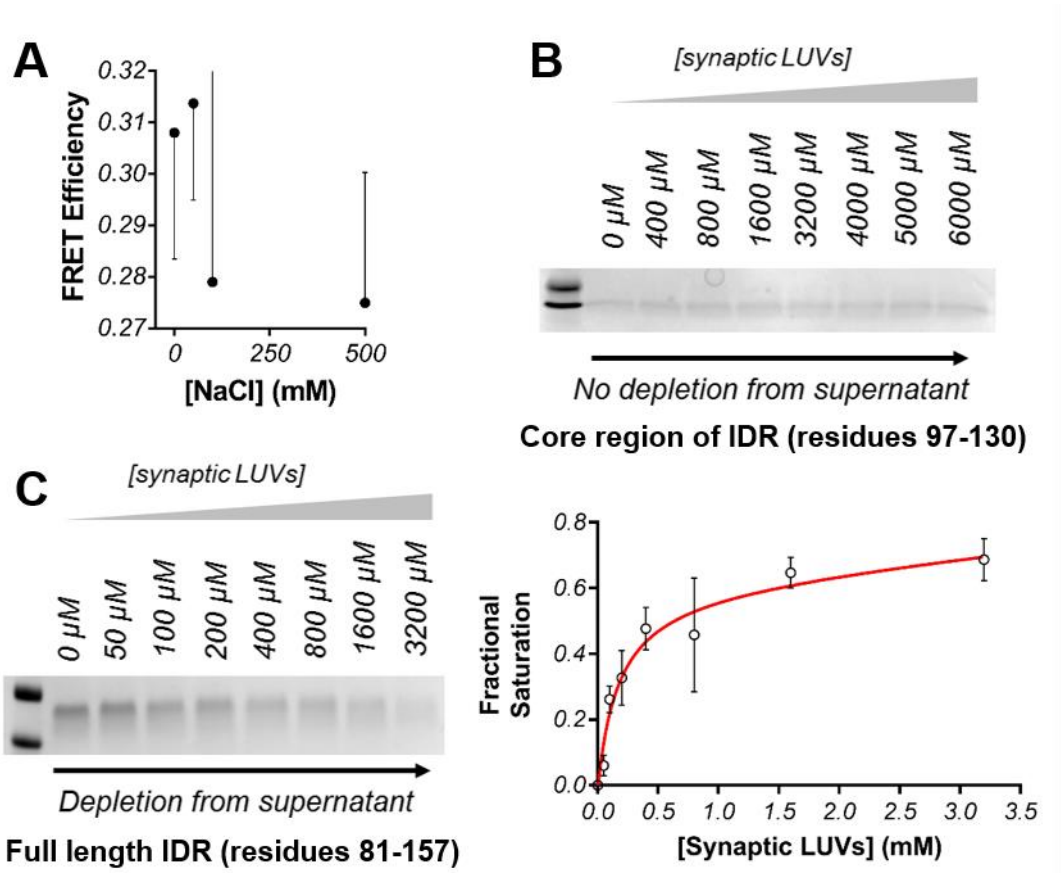

Figure 4.14. Potential factors influencing full-length IDR compaction. (A) FRET efficiency was not significantly affected by salt. (B) Core region Syt 1 IDR (representative of n=2) as a synthetic peptide in a co-sedimentation assay with synaptic vesicle mimic LUVs. (C) Full-length Syt 1 IDR (left, representative of n=3) as a synthetic peptide in a co-sedimentation assay with synaptic vesicle mimic LUVs and (right) the resultant binding curve (K_D = 169±82 µM). Collectively, the presence of either peptide in the supernatant indicates that the peptide containing the polybasic N-terminus is binding competent. This suggests that the basic N-terminus, in addition to transiently interacting with the negative C-terminus of the IDR sequence, also interacts with acidic lipids in a synaptic vesicle membrane and likely has two competing interactions (lipid – intermolecular; negative C-terminus – intramolecular) that could influence propensity to exist in compact structural state.

Table 4.3. Calculated kappa for blob g=5.

For g = 5										
i	**+**	**-**	$\mathbf{f_+}$	$\mathbf{f_-}$	$\boldsymbol{\sigma_i}$	$\boldsymbol{(\sigma_i - \sigma)^2}$	$\boldsymbol{\sum N_{blob}}$	**# of blobs**	**δ**	$\boldsymbol{\kappa_{g=5}}$
1	2	0	0.4	0	0.400	0.157	6.927	58	0.119	0.238
2	2	0	0.4	0	0.400	0.157				
3	2	0	0.4	0	0.400	0.157				
4	3	0	0.6	0	0.600	0.355				
5	3	0	0.6	0	0.600	0.355				
6	4	0	0.8	0	0.800	0.633				
7	4	0	0.8	0	0.800	0.633				
8	4	0	0.8	0	0.800	0.633				
9	3	0	0.6	0	0.600	0.355				
10	4	0	0.8	0	0.800	0.633				
11	3	1	0.6	0.2	0.200	0.038				
12	3	1	0.6	0.2	0.200	0.038				
13	2	1	0.4	0.2	0.067	0.004				
14	2	1	0.4	0.2	0.067	0.004				
15	2	1	0.4	0.2	0.067	0.004				
16	2	0	0.4	0	0.400	0.157				
17	1	0	0.2	0	0.200	0.038				
18	1	0	0.2	0	0.200	0.038				
19	1	0	0.2	0	0.200	0.038				
20	0	0	0	0						
21	1	0	0.2	0	0.200	0.038				
22	1	1	0.2	0.2	0.000	0.000				
23	1	1	0.2	0.2	0.000	0.000				
24	2	1	0.4	0.2	0.067	0.004				
25	2	2	0.4	0.4	0.000	0.000				
26	1	2	0.2	0.4	0.067	0.004				
27	1	1	0.2	0.2	0.000	0.000				
28	2	1	0.4	0.2	0.067	0.004				
29	1	1	0.2	0.2	0.000	0.000				
30	1	0	0.2	0	0.200	0.038				
31	2	0	0.4	0	0.400	0.157				
32	2	1	0.4	0.2	0.067	0.004				
33	1	1	0.2	0.2	0.000	0.000				
34	1	1	0.2	0.2	0.000	0.000				
35	1	1	0.2	0.2	0.000	0.000				
36	1	1	0.2	0.2	0.000	0.000				
37	1	1	0.2	0.2	0.000	0.000				
38	1	2	0.2	0.4	0.067	0.004				
39	1	3	0.2	0.6	0.200	0.038				
40	1	3	0.2	0.6	0.200	0.038				
41	0	4	0	0.8	0.800	0.633				

42	0	3	0	0.6	0.600	0.355	
43	0	2	0	0.4	0.400	0.157	
44	0	1	0	0.2	0.200	0.038	
45	0	1	0	0.2	0.200	0.038	
46	0	1	0	0.2	0.200	0.038	
47	0	1	0	0.2	0.200	0.038	
48	0	2	0	0.4	0.400	0.157	
49	0	3	0	0.6	0.600	0.355	
50	1	3	0.2	0.6	0.200	0.038	
51	1	3	0.2	0.6	0.200	0.038	
52	1	4	0.2	0.8	0.360	0.127	
53	1	3	0.2	0.6	0.200	0.038	
54	2	2	0.4	0.4	0.000	0.000	
55	1	3	0.2	0.6	0.200	0.038	
56	1	3	0.2	0.6	0.200	0.038	
57	1	3	0.2	0.6	0.200	0.038	
58	2	3	0.4	0.6	0.040	0.001	

Table 4.4. Calculated kappa for g=6.

for g=6										
i	+	-	f_+	f_-	σ_i	$(\sigma_i - \sigma)^2$	$\sum N_{blob}$	# of blobs	δ	$\kappa_{g=6}$
1	3	0	0.500	0.000	0.500	0.246	6.170	57	0.108	0.215
2	3	0	0.500	0.000	0.500	0.246				
3	3	0	0.500	0.000	0.500	0.246				
4	3	0	0.500	0.000	0.500	0.246				
5	4	0	0.667	0.000	0.667	0.439				
6	5	0	0.833	0.000	0.833	0.688				
7	5	0	0.833	0.000	0.833	0.688				
8	4	0	0.667	0.000	0.667	0.439				
9	4	0	0.667	0.000	0.667	0.439				
10	4	1	0.667	0.167	0.300	0.088				
11	4	1	0.667	0.167	0.300	0.088				
12	3	1	0.500	0.167	0.167	0.026				
13	2	1	0.333	0.167	0.056	0.003				
14	3	1	0.500	0.167	0.167	0.026				
15	2	1	0.333	0.167	0.056	0.003				
16	2	0	0.333	0.000	0.333	0.108				
17	1	0	0.167	0.000	0.167	0.026				
18	1	0	0.167	0.000	0.167	0.026				
19	1	0	0.167	0.000	0.167	0.026				
20	1	0	0.167	0.000	0.167	0.026				
21	1	1	0.167	0.167	0.000	2.E-05				
22	1	1	0.167	0.167	0.000	2.E-05				
23	2	1	0.333	0.167	0.056	0.003				
24	2	2	0.333	0.333	0.000	2.E-05				
25	2	2	0.333	0.333	0.000	2.E-05				
26	1	2	0.167	0.333	0.056	0.003				
27	2	1	0.333	0.167	0.056	0.003				
28	2	1	0.333	0.167	0.056	0.003				
29	1	1	0.167	0.167	0.000	2.E-05				
30	2	0	0.333	0.000	0.333	0.108				
31	2	1	0.333	0.167	0.056	0.003				
32	2	1	0.333	0.167	0.056	0.003				
33	1	1	0.167	0.167	0.000	2.E-05				
34	1	1	0.167	0.167	0.000	2.E-05				
35	2	1	0.333	0.167	0.056	0.003				
36	1	2	0.167	0.333	0.056	0.003				
37	1	2	0.167	0.333	0.056	0.003				
38	1	3	0.167	0.500	0.167	0.026				
39	1	3	0.167	0.500	0.167	0.026				
40	1	4	0.167	0.667	0.300	0.088				
41	0	4	0.000	0.667	0.667	0.439				

42	0	3	0.000	0.500	0.500	0.246	
43	0	2	0.000	0.333	0.333	0.108	
44	0	1	0.000	0.167	0.167	0.026	
45	0	2	0.000	0.333	0.333	0.108	
46	0	1	0.000	0.167	0.167	0.026	
47	0	2	0.000	0.333	0.333	0.108	
48	0	3	0.000	0.500	0.500	0.246	
49	1	3	0.167	0.500	0.167	0.026	
50	1	4	0.167	0.667	0.300	0.088	
51	1	4	0.167	0.667	0.300	0.088	
52	1	4	0.167	0.667	0.300	0.088	
53	2	3	0.333	0.500	0.033	0.001	
54	2	3	0.333	0.500	0.033	0.001	
55	1	4	0.167	0.667	0.300	0.088	
56	1	4	0.167	0.667	0.300	0.088	
57	2	3	0.333	0.500	0.033	0.001	

Table 4.5. MD-derived secondary structure for Syt 1 IDR core region. Turn (T), β-sheet (B), helix (H) and coil (C) are shown as fractions for each residue.

	Unphosphorylated Dielec80				Unphosphorylated Dielec20				Phosphorylated Dielec80				Phosphorylated Dielec20			
Res	T	B	H	C	T	B	H	C	T	B	H	C	T	B	H	C
97	0.120	0.000	0.000	0.880	0.104	0.000	0.000	0.896	0.126	0.000	0.000	0.874	0.270	0.000	0.000	0.730
98	0.229	0.000	0.005	0.766	0.208	0.027	0.008	0.751	0.244	0.002	0.003	0.750	0.320	0.001	0.003	0.676
99	0.359	0.000	0.020	0.621	0.380	0.027	0.040	0.505	0.379	0.002	0.014	0.605	0.652	0.002	0.006	0.340
100	0.457	0.000	0.083	0.460	0.547	0.000	0.144	0.305	0.500	0.000	0.064	0.435	0.676	0.084	0.039	0.202
101	0.442	0.000	0.097	0.460	0.551	0.000	0.173	0.277	0.484	0.000	0.077	0.438	0.686	0.099	0.045	0.169
102	0.443	0.000	0.104	0.454	0.540	0.000	0.194	0.265	0.486	0.000	0.078	0.436	0.711	0.089	0.064	0.135
103	0.519	0.000	0.092	0.389	0.559	0.000	0.185	0.253	0.550	0.000	0.045	0.405	0.614	0.118	0.061	0.208
104	0.445	0.000	0.093	0.462	0.404	0.024	0.179	0.348	0.474	0.002	0.030	0.494	0.682	0.107	0.052	0.159
105	0.394	0.000	0.089	0.517	0.355	0.024	0.194	0.419	0.448	0.002	0.035	0.515	0.604	0.000	0.048	0.348
106	0.418	0.000	0.128	0.454	0.361	0.000	0.307	0.331	0.482	0.000	0.059	0.459	0.664	0.084	0.034	0.218
107	0.352	0.000	0.153	0.495	0.355	0.000	0.335	0.310	0.440	0.000	0.073	0.487	0.612	0.094	0.181	0.113
108	0.342	0.000	0.158	0.499	0.412	0.000	0.327	0.261	0.414	0.000	0.065	0.522	0.228	0.193	0.170	0.409
109	0.354	0.000	0.149	0.497	0.373	0.000	0.336	0.287	0.359	0.000	0.044	0.597	0.234	0.115	0.174	0.478
110	0.486	0.000	0.102	0.412	0.458	0.000	0.287	0.252	0.331	0.000	0.009	0.660	0.220	0.000	0.000	0.780
111	0.448	0.001	0.141	0.411	0.445	0.006	0.295	0.253	0.193	0.000	0.007	0.799	0.087	0.000	0.033	0.879
112	0.448	0.001	0.147	0.404	0.422	0.010	0.285	0.282	0.160	0.000	0.010	0.830	0.502	0.000	0.123	0.374
113	0.520	0.000	0.162	0.317	0.485	0.011	0.288	0.215	0.324	0.000	0.026	0.650	0.612	0.000	0.221	0.167
114	0.527	0.000	0.203	0.270	0.484	0.010	0.323	0.182	0.462	0.000	0.057	0.482	0.621	0.000	0.343	0.035
115	0.520	0.000	0.226	0.255	0.531	0.000	0.324	0.146	0.492	0.000	0.086	0.421	0.578	0.000	0.398	0.024
116	0.525	0.001	0.205	0.269	0.516	0.000	0.295	0.190	0.522	0.000	0.083	0.395	0.591	0.000	0.350	0.059
117	0.482	0.001	0.122	0.395	0.501	0.011	0.124	0.362	0.484	0.000	0.053	0.464	0.623	0.000	0.120	0.258
118	0.380	0.000	0.058	0.562	0.388	0.015	0.060	0.534	0.384	0.000	0.022	0.593	0.393	0.000	0.018	0.589
119	0.284	0.000	0.019	0.697	0.309	0.013	0.024	0.652	0.296	0.000	0.008	0.696	0.243	0.000	0.007	0.750
120	0.228	0.000	0.022	0.750	0.257	0.005	0.052	0.685	0.242	0.000	0.011	0.747	0.330	0.000	0.013	0.658
121	0.190	0.000	0.024	0.787	0.237	0.000	0.070	0.693	0.211	0.000	0.017	0.773	0.305	0.000	0.067	0.628
122	0.220	0.000	0.026	0.754	0.268	0.000	0.073	0.659	0.237	0.000	0.019	0.744	0.349	0.000	0.084	0.567
123	0.335	0.000	0.023	0.642	0.363	0.000	0.061	0.574	0.333	0.000	0.016	0.650	0.446	0.005	0.091	0.459
124	0.309	0.000	0.019	0.672	0.350	0.000	0.056	0.593	0.312	0.000	0.016	0.672	0.467	0.006	0.079	0.449

125	0.308	0.000	0.016	0.676	0.338	0.001	0.040	0.620	0.312	0.000	0.013	0.676	0.420	0.003	0.060	0.518
126	0.260	0.000	0.009	0.731	0.288	0.000	0.018	0.691	0.255	0.000	0.006	0.739	0.349	0.000	0.021	0.630
127	0.122	0.000	0.004	0.875	0.146	0.000	0.008	0.845	0.126	0.000	0.002	0.872	0.150	0.000	0.005	0.846
128	0.057	0.000	0.000	0.943	0.067	0.000	0.000	0.932	0.053	0.000	0.000	0.947	0.065	0.000	0.000	0.935
129	0.001	0.000	0.000	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.000	0.999	0.001	0.000	0.000	0.999

CHAPTER 5: FUTURE DIRECTIONS

5.1 – Actin-binding domains

The results presented in Chapters 2 and 3 of this thesis provide several potential avenues for future studies. One of the first is exploration of the N-terminal residues upstream the CH1 domain. In Chapter 2, the truncation of the 50 N-terminal disordered region abolished binding. This finding is intriguing, because the ABD is typically defined as consisting of the tandem CH domains. Yet, in this example, the CH domains are present but no binding occurs. This opens the question about how significantly the N-terminal residues of other ABDs in the spectrin superfamily contribute to their respective binding affinities. Similar truncation experiments could address this. Another facet to this question is post-translational modifications. The N-terminal residues sometimes contain phosphorylation sites within the sequence. If the N-terminal sequence contributes to the actin-binding interface and PTMs occur there, is affinity being tuned up or down? Some research groups have already begun probing this question with simulation approaches[200], but it is well known that phosphorylation is typically an on/off switch for tuning functions.

Another possible question to explore is whether or not the βIII-spectrin mutation could be used to help in the generation of more high-resolution cryo-EM models, but for other ABDs. The wild-type affinity of the βIII-spectrin ABD was too low to decorate filaments and generate a model. Other ABDs may face similar problems, but if their affinities could be increased through introduction of mutations that increase affinity by perturbing the closed-open equilibrium (but the not fold) of CH domains, perhaps high-

resolution models would become attainable. Indeed, a more recent cryo-EM structural study[201] that generated ABD-actin models of even higher resolution than that of βIII-spectrin utilized ABD mutants that had elevated affinity. Those structural models also stabilized F-actin with phalloidin, which may contribute to improved resolution in the model. Either way, the cryo-EM models could reveal more information on ABD-actin interactions, particularly the N-terminal residues, which might otherwise be difficult to study structurally when bound to actin filaments.

5.2 – The IDR of synaptotagmin 1

For Chapter 4, the results of the study are of interest because they suggest that structural modifications outside the C2 domains are biologically relevant to the function of the protein. How, exactly, is still not known and will require design of exploratory experiments to generate new hypotheses. It is possible that the structural change is having an allosteric effect that impacts Ca^{2+}-sensing, but it also seems plausible that this region of synaptotagmin might form a binding site for some other exocytosis effector protein. Ultimately, this topic will be something I will explore during my postdoc.

References

[1] Lymn, R. W., and Taylor, E. W. (1971) Mechanism of adenosine triphosphate hydrolysis by actomyosin, *Biochemistry 10*, 4617-4624.

[2] Geeves, M. A. (2016) Review: The ATPase mechanism of myosin and actomyosin, *Biopolymers 105*, 483-491.

[3] Houdusse, A., and Sweeney, H. L. (2016) How myosin generates force on actin filaments, *Trends Biochem Sci 41*, 989-997.

[4] Perutz, M. F., Fermi, G., Luisi, B., Shaanan, B., and Liddington, R. C. (1987) Stereochemistry of cooperative mechanisms in hemoglobin, *Cold Spring Harb Symp Quant Biol 52*, 555-565.

[5] Dunker, A. K., Obradovic, Z., Romero, P., Garner, E. C., and Brown, C. J. (2000) Intrinsic protein disorder in complete genomes, *Genome Inform Ser Workshop Genome Inform 11*, 161-171.

[6] Kriwacki, R. W., Hengst, L., Tennant, L., Reed, S. I., and Wright, P. E. (1996) Structural studies of p21Waf1/Cip1/Sdi1 in the free and Cdk2-bound state: conformational disorder mediates binding diversity, *Proc Natl Acad Sci U S A 93*, 11504-11509.

[7] Dunker, A. K., Lawson, J. D., Brown, C. J., Williams, R. M., Romero, P., Oh, J. S., Oldfield, C. J., Campen, A. M., Ratliff, C. M., Hipps, K. W., Ausio, J., Nissen, M. S., Reeves, R., Kang, C., Kissinger, C. R., Bailey, R. W., Griswold, M. D., Chiu, W., Garner, E. C., and Obradovic, Z. (2001) Intrinsically disordered protein, *J Mol Graph Model 19*, 26-59.

[8] Dyson, H. J., and Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat Rev Mol Cell Biol 6*, 197-208.

[9] Wright, P. E., and Dyson, H. J. (2015) Intrinsically disordered proteins in cellular signalling and regulation, *Nat Rev Mol Cell Biol 16*, 18-29.

[10] Romero, P., Obradovic, Z., Li, X., Garner, E. C., Brown, C. J., and Dunker, A. K. (2001) Sequence complexity of disordered protein, *Proteins 42*, 38-48.

[11] Fung, H. Y. J., Birol, M., and Rhoades, E. (2018) IDPs in macromolecular complexes: the roles of multivalent interactions in diverse assemblies, *Curr Opin Struct Biol 49*, 36-43.

[12] Olsen, J. G., Teilum, K., and Kragelund, B. B. (2017) Behaviour of intrinsically disordered proteins in protein-protein complexes with an emphasis on fuzziness, *Cell Mol Life Sci 74*, 3175-3183.

[13] Greenfield, N., and Fasman, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry 8*, 4108-4116.

[14] Hubbell, W. L., Cafiso, D. S., and Altenbach, C. (2000) Identifying conformational changes with site-directed spin labeling, *Nat Struct Biol 7*, 735-739.

[15] Jeschke, G. (2012) DEER distance measurements on proteins, *Annu Rev Phys Chem 63*, 419-446.

[16] Liem, R. K. (2016) Cytoskeletal integrators: the spectrin superfamily, *Cold Spring Harb Perspect Biol 8*.

[17] Campbell, K. P., and Kahl, S. D. (1989) Association of dystrophin and an integral membrane glycoprotein, *Nature 338*, 259-262.

[18] Suzuki, A., Yoshida, M., Yamamoto, H., and Ozawa, E. (1992) Glycoprotein-binding site of dystrophin is confined to the cysteine-rich domain and the first half of the carboxy-terminal domain, *FEBS Lett 308*, 154-160.
[19] Dickson, G., Azad, A., Morris, G. E., Simon, H., Noursadeghi, M., and Walsh, F. S. (1992) Co-localization and molecular association of dystrophin with laminin at the surface of mouse and human myotubes, *J Cell Sci 103 (Pt 4)*, 1223-1233.
[20] Ervasti, J. M., and Campbell, K. P. (1993) A role for the dystrophin-glycoprotein complex as a transmembrane linker between laminin and actin, *J Cell Biol 122*, 809-823.
[21] Davies, K. E., and Nowak, K. J. (2006) Molecular mechanisms of muscular dystrophies: old and new players, *Nat Rev Mol Cell Biol 7*, 762-773.
[22] Zhang, R., Zhang, C., Zhao, Q., and Li, D. (2013) Spectrin: structure, function and disease, *Sci China Life Sci 56*, 1076-1085.
[23] Gao, Y., Perkins, E. M., Clarkson, Y. L., Tobia, S., Lyndon, A. R., Jackson, M., and Rothstein, J. D. (2011) beta-III spectrin is critical for development of purkinje cell dendritic tree and spine morphogenesis, *J Neurosci 31*, 16581-16590.
[24] Avery, A. W., Thomas, D. D., and Hays, T. S. (2017) beta-III-spectrin spinocerebellar ataxia type 5 mutation reveals a dominant cytoskeletal mechanism that underlies dendritic arborization, *Proc Natl Acad Sci U S A 114*, E9376-E9385.
[25] Ribeiro Ede, A., Jr., Pinotsis, N., Ghisleni, A., Salmazo, A., Konarev, P. V., Kostan, J., Sjoblom, B., Schreiner, C., Polyansky, A. A., Gkougkoulia, E. A., Holt, M. R., Aachmann, F. L., Zagrovic, B., Bordignon, E., Pirker, K. F., Svergun, D. I., Gautel, M., and Djinovic-Carugo, K. (2014) The structure and regulation of human muscle alpha-actinin, *Cell 159*, 1447-1460.
[26] Rybakova, I. N., and Ervasti, J. M. (2005) Identification of spectrin-like repeats required for high affinity utrophin-actin interaction, *J Biol Chem 280*, 23018-23023.
[27] Rybakova, I. N., Humston, J. L., Sonnemann, K. J., and Ervasti, J. M. (2006) Dystrophin and utrophin bind actin through distinct modes of contact, *J Biol Chem 281*, 9996-10001.
[28] Levine, B. A., Moir, A. J., Patchell, V. B., and Perry, S. V. (1990) The interaction of actin with dystrophin, *FEBS Lett 263*, 159-162.
[29] Levine, B. A., Moir, A. J., Patchell, V. B., and Perry, S. V. (1992) Binding sites involved in the interaction of actin with the N-terminal region of dystrophin, *FEBS Lett 298*, 44-48.
[30] Corrado, K., Mills, P. L., and Chamberlain, J. S. (1994) Deletion analysis of the dystrophin-actin binding domain, *FEBS Lett 344*, 255-260.
[31] Goldsmith, S. C., Pokala, N., Shen, W., Fedorov, A. A., Matsudaira, P., and Almo, S. C. (1997) The structure of an actin-crosslinking domain from human fimbrin, *Nat Struct Biol 4*, 708-712.
[32] Keep, N. H., Winder, S. J., Moores, C. A., Walke, S., Norwood, F. L., and Kendrick-Jones, J. (1999) Crystal structure of the actin-binding region of utrophin reveals a head-to-tail dimer, *Structure 7*, 1539-1546.
[33] Norwood, F. L., Sutherland-Smith, A. J., Keep, N. H., and Kendrick-Jones, J. (2000) The structure of the N-terminal actin-binding domain of human dystrophin and how mutations in this domain may cause Duchenne or Becker muscular dystrophy, *Structure 8*, 481-491.

[34] Hanein, D., Volkmann, N., Goldsmith, S., Michon, A. M., Lehman, W., Craig, R., DeRosier, D., Almo, S., and Matsudaira, P. (1998) An atomic model of fimbrin binding to F-actin and its implications for filament crosslinking and regulation, *Nat Struct Biol 5*, 787-792.
[35] Galkin, V. E., Orlova, A., VanLoock, M. S., Rybakova, I. N., Ervasti, J. M., and Egelman, E. H. (2002) The utrophin actin-binding domain binds F-actin in two different modes: implications for the spectrin superfamily of proteins, *J Cell Biol 157*, 243-251.
[36] Galkin, V. E., Orlova, A., Salmazo, A., Djinovic-Carugo, K., and Egelman, E. H. (2010) Opening of tandem calponin homology domains regulates their affinity for F-actin, *Nat Struct Mol Biol 17*, 614-616.
[37] Lin, A. Y., Prochniewicz, E., James, Z. M., Svensson, B., and Thomas, D. D. (2011) Large-scale opening of utrophin's tandem calponin homology (CH) domains upon actin binding by an induced-fit mechanism, *Proc Natl Acad Sci U S A 108*, 12729-12733.
[38] Sawyer, G. M., Clark, A. R., Robertson, S. P., and Sutherland-Smith, A. J. (2009) Disease-associated substitutions in the filamin B actin binding domain confer enhanced actin binding affinity in the absence of major structural disturbance: Insights from the crystal structures of filamin B actin binding domains, *J Mol Biol 390*, 1030-1047.
[39] Weins, A., Schlondorff, J. S., Nakamura, F., Denker, B. M., Hartwig, J. H., Stossel, T. P., and Pollak, M. R. (2007) Disease-associated mutant alpha-actinin-4 reveals a mechanism for regulating its F-actin-binding affinity, *Proc Natl Acad Sci U S A 104*, 16080-16085.
[40] Murphy, A. C., Lindsay, A. J., McCaffrey, M. W., Djinovic-Carugo, K., and Young, P. W. (2016) Congenital macrothrombocytopenia-linked mutations in the actin-binding domain of alpha-actinin-1 enhance F-actin association, *FEBS Lett 590*, 685-695.
[41] Lee, S. H., Weins, A., Hayes, D. B., Pollak, M. R., and Dominguez, R. (2008) Crystal structure of the actin-binding domain of α-actinin-4 Lys255Glu mutant implicated in focal segmental glomerulosclerosis, *J Mol Biol 376*, 317-324.
[42] Avery, A. W., Crain, J., Thomas, D. D., and Hays, T. S. (2016) A human beta-III-spectrin spinocerebellar ataxia type 5 mutation causes high-affinity F-actin binding, *Sci Rep 6*, 21375.
[43] Rizo, J., and Xu, J. (2015) The synaptic vesicle release machinery, *Annu Rev Biophys 44*, 339-367.
[44] Sutton, R. B., Fasshauer, D., Jahn, R., and Brunger, A. T. (1998) Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 A resolution, *Nature 395*, 347-353.
[45] Chen, X., Tomchick, D. R., Kovrigin, E., Arac, D., Machius, M., Sudhof, T. C., and Rizo, J. (2002) Three-dimensional structure of the complexin/SNARE complex, *Neuron 33*, 397-409.
[46] Tang, J., Maximov, A., Shin, O. H., Dai, H., Rizo, J., and Sudhof, T. C. (2006) A complexin/synaptotagmin 1 switch controls fast synaptic vesicle exocytosis, *Cell 126*, 1175-1187.

[47] Brewer, K. D., Bacaj, T., Cavalli, A., Camilloni, C., Swarbrick, J. D., Liu, J., Zhou, A., Zhou, P., Barlow, N., Xu, J., Seven, A. B., Prinslow, E. A., Voleti, R., Haussinger, D., Bonvin, A. M., Tomchick, D. R., Vendruscolo, M., Graham, B., Sudhof, T. C., and Rizo, J. (2015) Dynamic binding mode of a Synaptotagmin-1-SNARE complex in solution, *Nat Struct Mol Biol 22*, 555-564.

[48] Zhou, Q., Lai, Y., Bacaj, T., Zhao, M., Lyubimov, A. Y., Uervirojnangkoorn, M., Zeldin, O. B., Brewster, A. S., Sauter, N. K., Cohen, A. E., Soltis, S. M., Alonso-Mori, R., Chollet, M., Lemke, H. T., Pfuetzner, R. A., Choi, U. B., Weis, W. I., Diao, J., Sudhof, T. C., and Brunger, A. T. (2015) Architecture of the synaptotagmin-SNARE machinery for neuronal exocytosis, *Nature 525*, 62-67.

[49] Fealey, M. E., Mahling, R., Rice, A. M., Dunleavy, K., Kobany, S. E., Lohese, K. J., Horn, B., and Hinderliter, A. (2016) Synaptotagmin I's intrinsically disordered region interacts with synaptic vesicle lipids and exerts allosteric control over C2A, *Biochemistry 55*, 2914-2926.

[50] Dick, K. A., Ikeda, Y., Day, J. W., and Ranum, L. P. (2012) Spinocerebellar ataxia type 5, *Handb Clin Neurol 103*, 451-459.

[51] Ikeda, Y., Dick, K. A., Weatherspoon, M. R., Gincel, D., Armbrust, K. R., Dalton, J. C., Stevanin, G., Durr, A., Zuhlke, C., Burk, K., Clark, H. B., Brice, A., Rothstein, J. D., Schut, L. J., Day, J. W., and Ranum, L. P. (2006) Spectrin mutations cause spinocerebellar ataxia type 5, *Nat Genet 38*, 184-190.

[52] Stankewich, M. C., Gwynn, B., Ardito, T., Ji, L., Kim, J., Robledo, R. F., Lux, S. E., Peters, L. L., and Morrow, J. S. (2010) Targeted deletion of betaIII spectrin impairs synaptogenesis and generates ataxic and seizure phenotypes, *Proc Natl Acad Sci U S A 107*, 6022-6027.

[53] Perkins, E. M., Clarkson, Y. L., Sabatier, N., Longhurst, D. M., Millward, C. P., Jack, J., Toraiwa, J., Watanabe, M., Rothstein, J. D., Lyndon, A. R., Wyllie, D. J., Dutia, M. B., and Jackson, M. (2010) Loss of beta-III spectrin leads to Purkinje cell dysfunction recapitulating the behavior and neuropathology of spinocerebellar ataxia type 5 in humans, *J Neurosci 30*, 4857-4867.

[54] Efimova, N., Korobova, F., Stankewich, M. C., Moberly, A. H., Stolz, D. B., Wang, J., Kashina, A., Ma, M., and Svitkina, T. (2017) Beta III spectrin is necessary for formation of the constricted neck of dendritic spines and regulation of synaptic activity in neurons, *J Neurosci 37*, 6442-6459.

[55] Banuelos, S., Saraste, M., and Djinovic Carugo, K. (1998) Structural comparisons of calponin homology domains: implications for actin binding, *Structure 6*, 1419-1431.

[56] Borrego-Diaz, E., Kerff, F., Lee, S. H., Ferron, F., Li, Y., and Dominguez, R. (2006) Crystal structure of the actin-binding domain of alpha-actinin 1: evaluating two competing actin-binding models, *J Struct Biol 155*, 230-238.

[57] Franzot, G., Sjoblom, B., Gautel, M., and Djinovic Carugo, K. (2005) The crystal structure of the actin binding domain from alpha-actinin in its closed conformation: structural insight into phospholipid regulation of alpha-actinin, *J Mol Biol 348*, 151-165.

[58] Galkin, V. E., Orlova, A., Cherepanova, O., Lebart, M. C., and Egelman, E. H. (2008) High-resolution cryo-EM structure of the F-actin-fimbrin/plastin ABD2 complex, *Proc Natl Acad Sci U S A 105*, 1494-1498.

[59] Orlova, A. A., Galkin, V. E., Nakamura, F., and Egelman, E. H. (2011) Filamin CH domains bind to F-actin in an open conformation, *Mol. Biol. Cell 22*, 794a.
[60] Way, M., Pope, B., and Weeds, A. G. (1992) Evidence for functional homology in the F-actin binding domains of gelsolin and alpha-actinin: implications for the requirements of severing and capping, *J Cell Biol 119*, 835-842.
[61] Duff, R. M., Tay, V., Hackman, P., Ravenscroft, G., McLean, C., Kennedy, P., Steinbach, A., Schoffler, W., van der Ven, P. F. M., Furst, D. O., Song, J., Djinovic-Carugo, K., Penttila, S., Raheem, O., Reardon, K., Malandrini, A., Gambelli, S., Villanova, M., Nowak, K. J., Williams, D. R., Landers, J. E., Brown, R. H., Jr., Udd, B., and Laing, N. G. (2011) Mutations in the N-terminal actin-binding domain of filamin C cause a distal myopathy, *Am J Hum Genet 88*, 729-740.
[62] Sutherland-Smith, A. J., Moores, C. A., Norwood, F. L., Hatch, V., Craig, R., Kendrick-Jones, J., and Lehman, W. (2003) An atomic model for actin binding by the CH domains and spectrin-repeat modules of utrophin and dystrophin, *J Mol Biol 329*, 15-33.
[63] Nakamura, F., Hartwig, J. H., Stossel, T. P., and Szymanski, P. T. (2005) Ca^{2+} and calmodulin regulate the binding of filamin A to actin filaments, *J Biol Chem 280*, 32426-32433.
[64] Song, J. G., Kostan, J., Drepper, F., Knapp, B., de Almeida Ribeiro, E., Jr., Konarev, P. V., Grishkovskaya, I., Wiche, G., Gregor, M., Svergun, D. I., Warscheid, B., and Djinovic-Carugo, K. (2015) Structural insights into Ca^{2+}-calmodulin regulation of plectin 1a-integrin beta4 interaction in hemidesmosomes, *Structure 23*, 558-570.
[65] Kostan, J., Gregor, M., Walko, G., and Wiche, G. (2009) Plectin isoform-dependent regulation of keratin-integrin alpha6beta4 anchorage via Ca2+/calmodulin, *J Biol Chem 284*, 18525-18536.
[66] Singh, S. M., Bandi, S., and Mallela, K. M. G. (2017) The N-terminal flanking region modulates the actin binding affinity of the utrophin tandem calponin-homology domain, *Biochemistry 56*, 2627-2636.
[67] Burk, K., Zuhlke, C., Konig, I. R., Ziegler, A., Schwinger, E., Globas, C., Dichgans, J., and Hellenbroich, Y. (2004) Spinocerebellar ataxia type 5: clinical and molecular genetic features of a German kindred, *Neurology 62*, 327-329.
[68] Han, B., Zhou, R., Xia, C., and Zhuang, X. (2017) Structural organization of the actin-spectrin-based membrane skeleton in dendrites and soma of neurons, *Proc Natl Acad Sci U S A 114*, E6678-E6685.
[69] Lorenzo, D. N., Li, M. G., Mische, S. E., Armbrust, K. R., Ranum, L. P., and Hays, T. S. (2010) Spectrin mutations that cause spinocerebellar ataxia type 5 impair axonal transport and induce neurodegeneration in Drosophila, *J Cell Biol 189*, 143-158.
[70] Pielage, J., Cheng, L., Fetter, R. D., Carlton, P. M., Sedat, J. W., and Davis, G. W. (2008) A presynaptic giant ankyrin stabilizes the NMJ through regulation of presynaptic microtubules and transsynaptic cell adhesion, *Neuron 58*, 195-209.
[71] Koch, I., Schwarz, H., Beuchle, D., Goellner, B., Langegger, M., and Aberle, H. (2008) Drosophila ankyrin 2 is required for synaptic stability, *Neuron 58*, 210-222.
[72] Stephan, R., Goellner, B., Moreno, E., Frank, C. A., Hugenschmidt, T., Genoud, C., Aberle, H., and Pielage, J. (2015) Hierarchical microtubule organization controls

axon caliber and transport and determines synaptic structure and stability, *Dev Cell 33*, 5-21.
[73] Eaton, B. A., Fetter, R. D., and Davis, G. W. (2002) Dynactin is necessary for synapse stabilization, *Neuron 34*, 729-741.
[74] Holleran, E. A., Ligon, L. A., Tokito, M., Stankewich, M. C., Morrow, J. S., and Holzbaur, E. L. (2001) beta III spectrin binds to the Arp1 subunit of dynactin, *J Biol Chem 276*, 36598-36605.
[75] Lees-Miller, J. P., Helfman, D. M., and Schroer, T. A. (1992) A vertebrate actin-related protein is a component of a multisubunit complex involved in microtubule-based vesicle motility, *Nature 359*, 244-246.
[76] Schlager, M. A., Hoang, H. T., Urnavicius, L., Bullock, S. L., and Carter, A. P. (2014) In vitro reconstitution of a highly processive recombinant human dynein complex, *EMBO J 33*, 1855-1868.
[77] Mindell, J. A., and Grigorieff, N. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy, *J Struct Biol 142*, 334-347.
[78] Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields, *J Struct Biol 116*, 190-199.
[79] Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007) EMAN2: an extensible image processing suite for electron microscopy, *J Struct Biol 157*, 38-46.
[80] Egelman, E. H. (2000) A robust algorithm for the reconstruction of helical filaments using single-particle methods, *Ultramicroscopy 85*, 225-234.
[81] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis, *J Comput Chem 25*, 1605-1612.
[82] von der Ecken, J., Heissler, S. M., Pathan-Chhatbar, S., Manstein, D. J., and Raunser, S. (2016) Cryo-EM structure of a human cytoplasmic actomyosin complex at near-atomic resolution, *Nature 534*, 724-728.
[83] Zhang, Y. (2008) I-TASSER server for protein 3D structure prediction, *BMC Bioinformatics 9*, 40.
[84] Garcia-Alvarez, B., Bobkov, A., Sonnenberg, A., and de Pereda, J. M. (2003) Structural and functional analysis of the actin binding domain of plectin suggests alternative mechanisms for binding to F-actin and integrin beta4, *Structure 11*, 615-625.
[85] Wang, R. Y., Kudryashev, M., Li, X., Egelman, E. H., Basler, M., Cheng, Y., Baker, D., and DiMaio, F. (2015) De novo protein structure determination from near-atomic-resolution cryo-EM maps, *Nat Methods 12*, 335-338.
[86] Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.
[87] Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution, *Acta Crystallogr D Biol Crystallogr 66*, 213-221.

[88] Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S., and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Crystallogr D Biol Crystallogr 66*, 12-21.
[89] Prochniewicz, E., Zhang, Q., Howard, E. C., and Thomas, D. D. (1996) Microsecond rotational dynamics of actin: spectroscopic detection and theoretical simulation, *J Mol Biol 255*, 446-457.
[90] Butt, R. H., and Coorssen, J. R. (2013) Coomassie blue as a near-infrared fluorescent stain: a systematic comparison with Sypro Ruby for in-gel protein detection, *Mol Cell Proteomics 12*, 3834-3850.
[91] Legardinier, S., Legrand, B., Raguenes-Nicol, C., Bondon, A., Hardy, S., Tascon, C., Le Rumeur, E., and Hubert, J. F. (2009) A Two-amino Acid Mutation Encountered in Duchenne Muscular Dystrophy Decreases Stability of the Rod Domain 23 (R23) Spectrin-like Repeat of Dystrophin, *J Biol Chem 284*, 8822-8832.
[92] Jeschke, G., Koch, A., Jonas, U., and Godt, A. (2002) Direct conversion of EPR dipolar time evolution data to distance distributions, *J Magn Reson 155*, 72-82.
[93] Subramaniam, S., Earl, L. A., Falconieri, V., Milne, J. L., and Egelman, E. H. (2016) Resolution advances in cryo-EM enable application to drug discovery, *Curr Opin Struct Biol 41*, 194-202.
[94] Koenig, M., Monaco, A. P., and Kunkel, L. M. (1988) The complete sequence of dystrophin predicts a rod-shaped cytoskeletal protein, *Cell 53*, 219-228.
[95] Ervasti, J. M. (2007) Dystrophin, its interactions with other proteins, and implications for muscular dystrophy, *Biochim Biophys Acta 1772*, 108-117.
[96] Turner, P. R., Westwood, T., Regen, C. M., and Steinhardt, R. A. (1988) Increased protein degradation results from elevated free calcium levels found in muscle from mdx mice, *Nature 335*, 735-738.
[97] Rief, M., Pascual, J., Saraste, M., and Gaub, H. E. (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles, *J Mol Biol 286*, 553-561.
[98] Prochniewicz, E., Henderson, D., Ervasti, J. M., and Thomas, D. D. (2009) Dystrophin and utrophin have distinct effects on the structural dynamics of actin, *Proc Natl Acad Sci U S A 106*, 7822-7827.
[99] Lin, A. Y., Prochniewicz, E., Henderson, D. M., Li, B., Ervasti, J. M., and Thomas, D. D. (2012) Impacts of dystrophin and utrophin domains on actin structural dynamics: implications for therapeutic design, *J Mol Biol 420*, 87-98.
[100] Singh, S. M., and Mallela, K. M. (2012) The N-terminal actin-binding tandem calponin-homology (CH) domain of dystrophin is in a closed conformation in solution and when bound to F-actin, *Biophys J 103*, 1970-1978.
[101] Kuhlman, P. A., Hemmings, L., and Critchley, D. R. (1992) The identification and characterisation of an actin-binding site in alpha-actinin by mutagenesis, *FEBS Lett 304*, 201-206.
[102] Singh, S. M., Bandi, S., Winder, S. J., and Mallela, K. M. (2014) The actin binding affinity of the utrophin tandem calponin-homology domain is primarily determined by its N-terminal domain, *Biochemistry 53*, 1801-1809.
[103] Avery, A. W., Fealey, M. E., Wang, F., Orlova, A., Thompson, A. R., Thomas, D. D., Hays, T. S., and Egelman, E. H. (2017) Structural basis for high-affinity actin

binding revealed by a beta-III-spectrin SCA5 missense mutation, *Nat Commun 8*, 1350.
[104] Orlova, A., Rybakova, I. N., Prochniewicz, E., Thomas, D. D., Ervasti, J. M., and Egelman, E. H. (2001) Binding of dystrophin's tandem calponin homology domain to F-actin is modulated by actin's structure, *Biophys J 80*, 1926-1931.
[105] Colson, B. A., Thompson, A. R., Espinoza-Fonseca, L. M., and Thomas, D. D. (2016) Site-directed spectroscopy of cardiac myosin-binding protein C reveals effects of phosphorylation on protein structural dynamics, *Proc Natl Acad Sci U S A 113*, 3233-3238.
[106] Abraham, M. J., Murtola, T., Schulz, R., Pall, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX 1-2*, 19-25.
[107] Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: a web-based graphical user interface for CHARMM, *J Comput Chem 29*, 1859-1865.
[108] Jorgensen, W. L., Chandrasekhar, J., and Madura, J. (1983) Comparison of simple potential functions for simulating liquid water, *J Chem Phys 79*, 926-935.
[109] Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997) LINCS: A linear constraint solver for molecular simulations, *J Comput Chem 18*, 1463-1472.
[110] Essmann, U., Perera, L., and Berkowitz, M. L. (1995) A smooth particle mesh Ewald method, *J Chem Phys 103*, 8577-8593.
[111] Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling, *J Chem Phys 126*, 014101.
[112] Parrinello, M., and Rahman, A. J. (1981) Polymorphic transitions in single crystals: a new molecular dynamics method, *J Appl Phys 52*, 7182-7190.
[113] Klauda, J. B., Venable, R. M., Freites, J. A., O'Connor, J. W., Tobias, D. J., Mondragon-Ramirez, C., Vorobyov, I., MacKerell, A. D., Jr., and Pastor, R. W. (2010) Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types, *J Phys Chem B 114*, 7830-7843.
[114] MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins, *J Phys Chem B 102*, 3586-3616.
[115] Piana, S., Lindorff-Larsen, K., and Shaw, D. E. (2011) How robust are protein folding simulations with respect to force field parameterization, *Biophys J 100*, L47-L49.
[116] (2013)., R. C. T. R: A language and environment for statistical computing., R Foundation for Statistical Computing.
[117] The PyMOL Molecular Graphics System, Version 2.0 (Schrödinger,LLC).
[118] Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins, *Proteins 17*, 412-425.
[119] Pandini, A., Fornili, A., and Kleinjung, J. (2010) Structural alphabets derived from attractors in conformational space, *BMC Bioinformatics 11*, 97.

[120] Pandini, A., Fornili, A., Fraternali, F., and Kleinjung, J. (2012) Detection of allosteric signal transmission by information-theoretic analysis of protein dynamics, *FASEB J 26*, 868-881.
[121] Eisenhaber, F., Lijnzaad, P., Argos, P., Sander, C., and Scharf, M. (1995) The double cubic lattice method: efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies, *J Comput Chem 16*, 273-284.
[122] Chakravarty, D., Chakraborti, S., and Chakrabarti, P. (2015) Flexibility in the N-terminal actin-binding domain: clues from in silico mutations and molecular dynamics, *Proteins 83*, 696-710.
[123] Singh, S. M., Bandi, S., and Mallela, K. M. (2015) The N- and C-terminal domains differentially contribute to the structure and function of dystrophin and utrophin tandem calponin-homology domains, *Biochemistry 54*, 6942-6950.
[124] Shams, H., Golji, J., Garakani, K., and Mofrad, M. R. (2016) Dynamic regulation of α-actinin's calponin homology domains on F-actin, *Biophys J 110*, 1444-1455.
[125] Renley, B. A., Rybakova, I. N., Amann, K. J., and Ervasti, J. M. (1998) Dystrophin binding to nonmuscle actin, *Cell Motil Cytoskeleton 41*, 264-270.
[126] McCoy, J., and Hubbell, W. L. (2011) High-pressure EPR reveals conformational equilibria and volumetric properties of spin-labeled proteins, *Proc Natl Acad Sci U S A 108*, 1331-1336.
[127] De La Cruz, E. M., Roland, J., McCullough, B. R., Blanchoin, L., and Martiel, J. L. (2010) Origin of twist-bend coupling in actin filaments, *Biophys J 99*, 1852-1860.
[128] Singh, S. M., Kongari, N., Cabello-Villegas, J., and Mallela, K. M. (2010) Missense mutations in dystrophin that trigger muscular dystrophy decrease protein stability and lead to cross-beta aggregates, *Proc Natl Acad Sci U S A 107*, 15069-15074.
[129] Henderson, D. M., Lee, A., and Ervasti, J. M. (2010) Disease-causing missense mutations in actin binding domain 1 of dystrophin induce thermodynamic instability and protein aggregation, *Proc Natl Acad Sci U S A 107*, 9632-9637.
[130] Talsness, D. M., Belanto, J. J., and Ervasti, J. M. (2015) Disease-proportional proteasomal degradation of missense dystrophins, *Proc Natl Acad Sci U S A 112*, 12414-12419.
[131] Spolar, R. S., Livingstone, J. R., and Record, M. T., Jr. (1992) Use of liquid hydrocarbon and amide transfer data to estimate contributions to thermodynamic functions of protein folding from the removal of nonpolar and polar surface from water, *Biochemistry 31*, 3947-3955.
[132] Freire, E. (1995) Differential scanning calorimetry, *Methods Mol Biol 40*, 191-218.
[133] Loladze, V. V., Ermolenko, D. N., and Makhatadze, G. I. (2001) Heat capacity changes upon burial of polar and nonpolar groups in proteins, *Protein Sci 10*, 1343-1352.
[134] Makhatadze, G. I., and Privalov, P. L. (1990) Heat capacity of proteins. I. Partial molar heat capacity of individual amino acid residues in aqueous solution: hydration effect, *J Mol Biol 213*, 375-384.
[135] Privalov, P. L. (1979) Stability of proteins: small globular proteins, *Adv Protein Chem 33*, 167-241.

[136] Privalov, P. L., and Makhatadze, G. I. (1990) Heat capacity of proteins. II. Partial molar heat capacity of the unfolded polypeptide chain of proteins: protein unfolding effects, *J Mol Biol 213*, 385-391.
[137] Robinson, C. R., Liu, Y., O'Brien, R., Sligar, S. G., and Sturtevant, J. M. (1998) A differential scanning calorimetric study of the thermal unfolding of apo- and holo-cytochrome b562, *Protein Sci 7*, 961-965.
[138] Makhatadze, G. I., and Privalov, P. L. (1996) On the entropy of protein folding, *Protein Sci 5*, 507-510.
[139] Naganathan, A. N., Sanchez-Ruiz, J. M., and Munoz, V. (2005) Direct measurement of barrier heights in protein folding, *J Am Chem Soc 127*, 17970-17971.
[140] Zhou, Y., Hall, C. K., and Karplus, M. (1999) The calorimetric criterion for a two-state process revisited, *Protein Sci 8*, 1064-1074.
[141] Atkins, J. D., Boateng, S. Y., Sorensen, T., and McGuffin, L. J. (2015) Disorder prediction methods, their applicability to different protein targets and their usefulness for guiding experimental studies, *Int J Mol Sci 16*, 19040-19054.
[142] Uversky, V. N., Oldfield, C. J., and Dunker, A. K. (2005) Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling, *J Mol Recognit 18*, 343-384.
[143] Rauscher, S., Gapsys, V., Gajda, M. J., Zweckstetter, M., de Groot, B. L., and Grubmuller, H. (2015) Structural ensembles of intrinsically disordered proteins depend strongly on force field: a comparison to experiment, *J Chem Theory Comput 11*, 5513-5524.
[144] Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers 22*, 2577-2637.
[145] van der Spoel, D., and Berendsen, H. J. (1997) Molecular dynamics simulations of Leu-enkephalin in water and DMSO, *Biophys J 72*, 2032-2041.
[146] Sudhof, T. C. (2013) A molecular machine for neurotransmitter release: synaptotagmin and beyond, *Nat Med 19*, 1227-1231.
[147] Fealey, M. E., Gauer, J. W., Kempka, S. C., Miller, K., Nayak, K., Sutton, R. B., and Hinderliter, A. (2012) Negative coupling as a mechanism for signal propagation between C2 domains of synaptotagmin I, *PLoS One 7*, e46748.
[148] Fealey, M. E., and Hinderliter, A. (2013) Allostery and instability in the functional plasticity of synaptotagmin I, *Commun Integr Biol 6*, e22830.
[149] Hilser, V. J., and Thompson, E. B. (2007) Intrinsic disorder as a mechanism to optimize allosteric coupling in proteins, *Proc Natl Acad Sci U S A 104*, 8311-8315.
[150] McAdam, R. L., Varga, K. T., Jiang, Z., Young, F. B., Blandford, V., McPherson, P. S., Gong, L. W., and Sossin, W. S. (2015) The juxtamembrane region of synaptotagmin 1 interacts with dynamin 1 and regulates vesicle fission during compensatory endocytosis in endocrine cells, *J Cell Sci 128*, 2229-2235.
[151] Meszaros, B., Simon, I., and Dosztanyi, Z. (2009) Prediction of protein binding regions in disordered proteins, *PLoS Comput Biol 5*, e1000376.
[152] Dosztanyi, Z., Meszaros, B., and Simon, I. (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins, *Bioinformatics 25*, 2745-2746.

[153] Dobrynin, A. V., Colby, R. H., and Rubinstein, M. (2004) Polyampholytes, *J Polym Sci Pol Phys 42*, 3513-3538.
[154] Higgs, P. G., and Joanny, J. F. (1991) Theory of polyampholyte solutions, *J Chem Phys 94*, 1543-1554.
[155] Das, R. K., and Pappu, R. V. (2013) Conformations of intrinsically disordered proteins are influenced by linear sequence distributions of oppositely charged residues, *Proc Natl Acad Sci U S A 110*, 13392-13397.
[156] Holehouse, A. S., Das, R. K., Ahad, J. N., Richardson, M. O., and Pappu, R. V. (2017) CIDER: resources to analyze sequence-ensemble relationships of intrinsically disordered proteins, *Biophys J 112*, 16-21.
[157] Lai, Y., Lou, X., Jho, Y., Yoon, T. Y., and Shin, Y. K. (2013) The synaptotagmin 1 linker may function as an electrostatic zipper that opens for docking but closes for fusion pore opening, *Biochem J 456*, 25-33.
[158] de Jong, A. P., Meijer, M., Saarloos, I., Cornelisse, L. N., Toonen, R. F., Sorensen, J. B., and Verhage, M. (2016) Phosphorylation of synaptotagmin-1 controls a post-priming step in PKC-dependent presynaptic plasticity, *Proc Natl Acad Sci U S A 113*, 5095-5100.
[159] Chakrabartty, A., Doig, A. J., and Baldwin, R. L. (1993) Helix capping propensities in peptides parallel those in proteins, *Proc Natl Acad Sci U S A 90*, 11332-11336.
[160] Gustiananda, M., Liggins, J. R., Cummins, P. L., and Gready, J. E. (2004) Conformation of prion protein repeat peptides probed by FRET measurements and molecular dynamics simulations, *Biophys J 86*, 2467-2483.
[161] Haughland, R. P. (2001) *Handbook of fluorescent probes and research products.*, Molecular Probes.
[162] Gauer, J. W., Sisk, R., Murphy, J. R., Jacobson, H., Sutton, R. B., Gillispie, G. D., and Hinderliter, A. (2012) Mechanism for calcium ion sensing by the C2A domain of synaptotagmin I, *Biophys J 103*, 238-246.
[163] Kast, D., Espinoza-Fonseca, L. M., Yi, C., and Thomas, D. D. (2010) Phosphorylation-induced structural changes in smooth muscle myosin regulatory light chain, *Proc Natl Acad Sci U S A 107*, 8207-8212.
[164] Adler, A. J., Greenfield, N. J., and Fasman, G. D. (1973) Circular dichroism and optical rotatory dispersion of proteins and polypeptides, *Methods Enzymol 27*, 675-735.
[165] Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD, *J Comput Chem 26*, 1781-1802.
[166] Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., and Mackerell, A. D., Jr. (2012) Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles, *J Chem Theory Comput 8*, 3257-3273.
[167] Tanner, D. E., Chan, K. Y., Phillips, J. C., and Schulten, K. (2011) Parallel generalized born implicit solvent calculations with NAMD, *J Chem Theory Comput 7*, 3635-3642.
[168] Cherepanov, D. A., Feniouk, B. A., Junge, W., and Mulkidjanian, A. Y. (2003) Low dielectric permittivity of water at the membrane interface: effect on the energy coupling mechanism in biological membranes, *Biophys J 85*, 1307-1316.

[169] Raudino, A., and Mauzerall, D. (1986) Dielectric properties of the polar head group region of zwitterionic lipid bilayers, *Biophys J 50*, 441-449.
[170] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: visual molecular dynamics, *J Mol Graph 14*, 33-38, 27-38.
[171] Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment, *Proteins 23*, 566-579.
[172] Millhauser, G. L. (1995) Views of helical peptides: a proposal for the position of 3(10)-helix along the thermodynamic folding pathway, *Biochemistry 34*, 3873-3877.
[173] Marsh, J. A., and Forman-Kay, J. D. (2010) Sequence determinants of compaction in intrinsically disordered proteins, *Biophys J 98*, 2383-2390.
[174] Uversky, V. N. (2015) Protein misfolding in lipid-mimetic environments, *Adv Exp Med Biol 855*, 33-66.
[175] Uversky, V. N. (2002) Natively unfolded proteins: a point where biology waits for physics, *Protein Sci 11*, 739-756.
[176] Zor, T., Mayr, B. M., Dyson, H. J., Montminy, M. R., and Wright, P. E. (2002) Roles of phosphorylation and helix propensity in the binding of the KIX domain of CREB-binding protein by constitutive (c-Myb) and inducible (CREB) activators, *J Biol Chem 277*, 42241-42248.
[177] Pavsic, M., Ilc, G., Vidmar, T., Plavec, J., and Lenarcic, B. (2015) The cytosolic tail of the tumor marker protein Trop2--a structural switch triggered by phosphorylation, *Sci Rep 5*, 10324.
[178] Li, J., Motlagh, H. N., Chakuroff, C., Thompson, E. B., and Hilser, V. J. (2012) Thermodynamic dissection of the intrinsically disordered N-terminal domain of human glucocorticoid receptor, *J Biol Chem 287*, 26777-26787.
[179] He, Y., Chen, Y., Mooney, S. M., Rajagopalan, K., Bhargava, A., Sacho, E., Weninger, K., Bryan, P. N., Kulkarni, P., and Orban, J. (2015) Phosphorylation-induced conformational ensemble switching in an intrinsically disordered cancer/testis antigen, *J Biol Chem 290*, 25090-25102.
[180] Stott, K., Watson, M., Bostock, M. J., Mortensen, S. A., Travers, A., Grasser, K. D., and Thomas, J. O. (2014) Structural insights into the mechanism of negative regulation of single-box high mobility group proteins by the acidic tail domain, *J Biol Chem 289*, 29817-29826.
[181] Myers, J. K., Pace, C. N., and Scholtz, J. M. (1998) Trifluoroethanol effects on helix propensity and electrostatic interactions in the helical peptide from ribonuclease T1, *Protein Sci 7*, 383-388.
[182] Roccatano, D., Colombo, G., Fioroni, M., and Mark, A. E. (2002) Mechanism by which 2,2,2-trifluoroethanol/water mixtures stabilize secondary-structure formation in peptides: a molecular dynamics study, *Proc Natl Acad Sci U S A 99*, 12179-12184.
[183] Luque, I., Mayorga, O. L., and Freire, E. (1996) Structure-based thermodynamic scale of alpha-helix propensities in amino acids, *Biochemistry 35*, 13681-13688.
[184] Pace, C. N., and Scholtz, J. M. (1998) A helix propensity scale based on experimental studies of peptides and proteins, *Biophys J 75*, 422-427.
[185] Szilak, L., Moitra, J., Krylov, D., and Vinson, C. (1997) Phosphorylation destabilizes alpha-helices, *Nat Struct Biol 4*, 112-114.

[186] Andrew, C. D., Warwicker, J., Jones, G. R., and Doig, A. J. (2002) Effect of phosphorylation on alpha-helix stability as a function of position, *Biochemistry 41*, 1897-1905.
[187] Lu, B., Kiessling, V., Tamm, L. K., and Cafiso, D. S. (2014) The juxtamembrane linker of full-length synaptotagmin 1 controls oligomerization and calcium-dependent membrane binding, *J Biol Chem 289*, 22161-22171.
[188] Bartels, T., Ahlstrom, L. S., Leftin, A., Kamp, F., Haass, C., Brown, M. F., and Beyer, K. (2010) The N-terminus of the intrinsically disordered protein alpha-synuclein triggers membrane binding and helix folding, *Biophys J 99*, 2116-2124.
[189] Li, J., James, Z. M., Dong, X., Karim, C. B., and Thomas, D. D. (2012) Structural and functional dynamics of an integral membrane protein complex modulated by lipid headgroup charge, *J Mol Biol 418*, 379-389.
[190] Steinmetz, M. O., Jahnke, W., Towbin, H., Garcia-Echeverria, C., Voshol, H., Muller, D., and van Oostrum, J. (2001) Phosphorylation disrupts the central helix in Op18/stathmin and suppresses binding to tubulin, *EMBO Rep 2*, 505-510.
[191] Bah, A., Vernon, R. M., Siddiqui, Z., Krzeminski, M., Muhandiram, R., Zhao, C., Sonenberg, N., Kay, L. E., and Forman-Kay, J. D. (2015) Folding of an intrinsically disordered protein by phosphorylation as a regulatory switch, *Nature 519*, 106-109.
[192] Wu, B., Wei, S., Petersen, N., Ali, Y., Wang, X., Bacaj, T., Rorsman, P., Hong, W., Sudhof, T. C., and Han, W. (2015) Synaptotagmin-7 phosphorylation mediates GLP-1-dependent potentiation of insulin secretion from beta-cells, *Proc Natl Acad Sci U S A 112*, 9996-10001.
[193] Hilfiker, S., Pieribone, V. A., Nordstedt, C., Greengard, P., and Czernik, A. J. (1999) Regulation of synaptotagmin I phosphorylation by multiple protein kinases, *J Neurochem 73*, 921-932.
[194] Vrljic, M., Strop, P., Hill, R. C., Hansen, K. C., Chu, S., and Brunger, A. T. (2011) Post-translational modifications and lipid binding profile of insect cell-expressed full-length mammalian synaptotagmin 1, *Biochemistry 50*, 9998-10012.
[195] Fioroni, M., Diaz, M. D., Burger, K., and Berger, S. (2002) Solvation phenomena of a tetrapeptide in water/trifluoroethanol and water/ethanol mixtures: a diffusion NMR, intermolecular NOE, and molecular dynamics study, *J Am Chem Soc 124*, 7737-7744.
[196] Starzyk, A., Barber-Armstrong, W., Sridharan, M., and Decatur, S. M. (2005) Spectroscopic evidence for backbone desolvation of helical peptides by 2,2,2-trifluoroethanol: an isotope-edited FTIR study, *Biochemistry 44*, 369-376.
[197] Kentsis, A., and Sosnick, T. R. (1998) Trifluoroethanol promotes helix formation by destabilizing backbone exposure: desolvation rather than native hydrogen bonding defines the kinetic pathway of dimeric coiled coil folding, *Biochemistry 37*, 14613-14622.
[198] Rice, A. M., Mahling, R., Fealey, M. E., Rannikko, A., Dunleavy, K., Hendrickson, T., Lohese, K. J., Kruggel, S., Heiling, H., Harren, D., Sutton, R. B., Pastor, J., and Hinderliter, A. (2014) Randomly organized lipids and marginally stable proteins: a coupling of weak interactions to optimize membrane signaling, *Biochim Biophys Acta 1838*, 2331-2340.
[199] Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of

phenylmethanesulfonyl alpha-chymotrypsin using different denaturants, *Biochemistry 27*, 8063-8068.

[200] Travers, T., Shao, H., Joughin, B. A., Lauffenburger, D. A., Wells, A., and Camacho, C. J. (2015) Tandem phosphorylation within an intrinsically disordered region regulates ACTN4 function, *Sci Signal 8*, ra51.

[201] Iwamoto, D. V., Huehn, A., Simon, B., Huet-Calderwood, C., Baldassarre, M., Sindelar, C. V., and Calderwood, D. A. (2018) Structural basis of the filamin A actin-binding domain interaction with F-actin, *Nat Struct Mol Biol 25*, 918-927.