8-8-2017

Biophysical Investigation of Protein Interactions to Evaluate Protein Structure and Function

Khiem Nguyen
University of Connecticut - Storrs, khiem.q.nguyen@uconn.edu

Biophysical Investigation of Protein Interactions to Evaluate Protein Structure and Function

Khiem Quoc Nguyen, PhD

University of Connecticut, 2017

Proteins are essential macromolecules for the proper function and proliferation of life in tandem with carbohydrates, lipids, and nucleic acids. Even with our current understanding, novel proteins and protein-protein interactions are continuously being discovered, providing a better and clearer picture of cellular processes. In this dissertation, the work presented will focus on tackling how proteins interact with their ligands and what this tells us about the protein's structure and function by utilizing biophysical techniques encompassing nuclear magnetic resonance spectroscopy and isothermal titration calorimetry. In chapter 2, we have studied the protein-protein interactions of cluster of differentiation 47 (CD47) and protein linking integrin associated protein to the cytoskeleton 2 (PLIC-2) and confirmed that the two do not associate *in vitro* under the conditions tested. It was anticipated that PLIC-2 could be a means for CD47 to modulate the cytoskeletal framework, but the lack of direct binding weakens that possibility. In chapter 3, the interaction between butyrophilin 3A1 (BTN3A1) and phosphoantigens (pAg) were explored, focusing specifically on the region between BTN3A1's transmembrane helix and B30.2 domain, denoted juxtamembrane (JM) region. We have found that the JM region was able to weakly associate with pAg, where the important functionality is the negatively charged diphosphate moiety. This association is potentially vital for the activation of BTN3A1, where ligand association was found to be

able to induce a conformational change within the intracellular region of BTN3A1. In chapter 4, we explored the use of small molecules in order to probe the protein surface through NMR. By using select hydrophobic probes, it is possible to quantitate the overall surface hydrophobicity of a protein, which is important for understanding protein stability as solution properties, such as aggregation, have been linked to hydrophobic interactions. I have added an appendix chapter A1, which is related to my minor contribution to the structural refinement of skelemin and understanding its interaction with integrin. All in all, the work presented here focuses on protein interactions and how it can be related to its biological role and solution properties.

Biophysical Investigation of Protein Interactions to Evaluate Protein Structure and Function

Khiem Quoc Nguyen

B.A., Clark University, 2011

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

2017

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Biophysical Investigation of Protein Interactions to Evaluate Protein Structure and Function

Presented by
Khiem Nguyen, B.A.

Major Advisor

_____

Olga Vinogradova

Associate Advisor

_____

Devendra Kalonia

Associate Advisor

_____

Andrew Wiemer

University of Connecticut

2017

Acknowledgements

The path towards accomplishing my doctorate degree at the University of Connecticut has been a long and difficult journey, but has been enjoyable nonetheless due to the thoughtful and caring individuals here in the community. First and foremost, I would like to thank my mentor, Dr. Olga Vinogradova, for her guidance. She has always pushed me to succeed and was always there for thoughtful discussion, which was a great learning experience. I will value the time spent in her lab. I would like to thank my committee members, Dr. Kalonia and Dr. Wiemer, for their help and wonderful collaboration projects. I would also like to thank my previous committee member, Dr. Anderson, who is no longer with us. You will be missed. An extra thanks for Dr. Kendall and Dr. Alder, who have agreed without hesitation to be my examiner and allowed me to do minor work within their labs.

Of course, I have to thank my friends here at the UCONN. The Vinogradova lab, who are Dr. Robbins Puthenveetil, Dr. Priya Katyal, Dr. Xiaochen Lin, and Dr. Vitaliy Gorbatyuk, have been a core part of my research career. Not only were they great colleagues that taught me the ins and outs, but they were friends that made work enjoyable and full of laughter. A special thanks to Dr. Vitaliy Gorbatyuk, the NMR facility manager, for teaching me a great deal of NMR as well as the good times working underneath him as the NMR facility assistant. I would like to thank my collaborator and friend, Dr. Zecca, for all the help and work we've done together over five years. I would also like to acknowledge everyone within the Medicinal Chemistry division for their kindness and positivity. It truly was a great environment to work in.

And last but not least, I have to show my appreciation to my family for all their support throughout these years. My dad and mom have always been supportive and believed in me. I could have only made it this far due to their hard work to provide for my siblings and me. I would like to thank my brother for the fun times we shared, which helped relieve the stressful days, as well as my sister for all the advice, help, and encouragement. I can always count on my family in rough times and this is a special thank you to you all.

Table of Contents

List of Abbreviations

A-ChytA – α-Chymotrypsinogen A

ANS – 8-Anilinonaphthalene-1-sulfonic acid

BFI – Construct of the full intracellular region of BTN3A1 (JM + B30.2 domain)

B-LgA – β-Lactoglobulin A

BSA – Bovine Serum Albumin

BTN3A1 – Butyrophilin 3A1

CD47 – Cluster of Differentiation 47

cHDMAPP – hydroxy-methyl-butyl-diphosphonate

CPMG – Carr-Purcell-Meiboom-Gill

CSA – Chemical shift anisotropy

CSP – Chemical shift perturbation

DMAPP – Dimethylallyl diphosphate

DSS – 4,4-dimethyl-4-silapentane-1-sulfonic acid

FID – Free induction decay

HIC – Hydrophobic interaction chromatography

HMBPP – (*E*)-4-hydroxy-3-methyl-but-2-enyl diphosphate

IPP – isopentyl diphosphate

ITC – Isothermal Titration Calorimetry

JM – Juxtamembrane

NMR – Nuclear magnetic resonance

PLIC – Proteins linking IAP with cytoskeleton

PRE – Paramagnetic relaxation enhancement

SAXS – Small angle x-ray scattering

T2 – Transverse relaxation time

Ub – Ubiquitin

UBA – Ubiquitin associated domain

UBL – Ubiquitin-like domain

List of Figures

List of Tables

Chapter 1 – Introduction

Proteins as a whole have diverse roles and functions within the cell, such as a cell division, enzymatic, protein expression regulation, or even apoptosis. Understanding individual biological processes is fundamental for comprehending medical problems that may arise from protein mutations and/or deletions, but is an arduous task as the mechanical workings within the cell is a complex interplay of proteins, nucleic acids, lipids, carbohydrates, and small molecules. In this dissertation, I will investigate various proteins, such as CD47 or BTN3A1, and their binding partners *in vitro* to better understand their structure and function. There are numerous ways and instrumentation for a scientist to conduct research, but this introduction will provide a brief overview of the main techniques used in this dissertation. There are multiple references if a more in-depth explanation is required [1-5].

1.1 – Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) is a technique that can be utilized to characterize the thermodynamic parameters of interactions by directly measuring the heats of binding at constant temperature. These processes may be endothermic or exothermic, depending on whether or not they use or release heat, respectively. The heat detected is of particular importance as ITC is able to directly determine the change in enthalpy (ΔH) as shown in Figure 1.1. From the titration data, the equilibrium association constant (K_a) is determined from the slope at the inflection point.

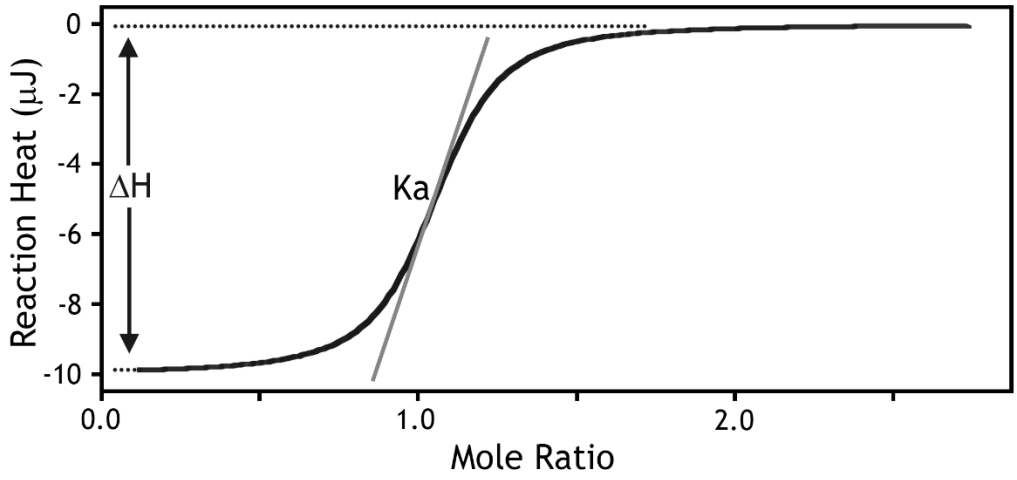

Figure 1.1. Diagram of Important Parameters Determined by ITC. The binding curve presented here is a simulated curve using an independent binding model in which the concentrations were 800 μM and 100 μM for the titrant and protein, respectively. The K_d was set to 1 μM. The ΔH is determined by measuring the maximal difference in heat (between the dotted lines) and the slope (grey line) at the inflection point is related to the binding affinity, K_a.

The determination of both the ΔH and K_a is very useful because they can be used to calculate the rest of the thermodynamic parameters through the equations:

$$\Delta G = -RTlnK_a \qquad (1)$$

$$\Delta G = \Delta H - T\Delta S \qquad (2)$$

where ΔG is the change in Gibbs free energy, R is the gas constant, T is the temperature, and ΔS is the change in entropy. Equation 1 reflects the system at equilibrium, which allows for the calculation of ΔG at the experimental temperature. Afterwards, the ΔG can be plugged in equation 2 in order to determine the ΔS. As a result, we can be able to predict the thermodynamic parameters of the system of interest, where we can observe if it is enthalpically and/or entropically favorable. This is useful is gauging the strengths of

interactions, where entropic gains are generally from hydrophobic interactions whereas favorable enthalpies can be from hydrogen bonds and charge-charge interactions.

Though ITC is a powerful tool for evaluating the thermodynamic parameters, there are several practical considerations for performing ITC runs. The ITC measures the total heat from the system, which means that it detects the formation and/or breaking of hydrogen bonds from ligand-ligand and ligand-solvent interactions on top of the ligand-protein interaction of interest. As such, this necessitates the evaluation of a titration control, where the ligand is titrated into the buffer in the absence of protein. This measures the heats of dilution, in which the ligand goes from a concentrated to a dilute state. The heat differences due to the ligand-ligand and ligand-solvent are subtracted from the actual experiment to more properly characterize the ligand-protein association. Another consideration to factor in is the c value, which is a relationship between the K_a and the concentration of the protein in the cell:

$$c = K_a n[P] \qquad (3)$$

where the n is the stoichiometry and [P] is the protein concentration in the cell. The acceptable range of the c value is generally between 10-100, which provides a reasonable binding curve as shown in Figure 1.1 (c value = 100). For the low volume NanoITC, this is appropriate for protein based interactions whose K_d values are usually in the low µM range as it translates to needing a protein concentration of approximately 100 µM. Though the ITC is able to detect bindings within the nM to mM range, obtaining workable protein concentrations would be difficult as mM affinities require very high concentrations that may not be possible due to protein stability and aggregation. On the other hand, nM affinities require lower concentrations, which consequently means lower

sensitivity due to the reduced heats from binding. However, there are workarounds, such as competitive displacement experiments.

1.2 – Nuclear Magnetic Resonance Spectroscopy

1.2.1 – Overview of Key Principles

Nuclear magnetic resonance (NMR) spectroscopy is a sensitive and robust technique that relies on a property called the nuclear spin, I, that is dependent upon the composition of the nucleus (number of protons and neutrons). When subjected to an external magnetic field, there will be a (2I + 1)-fold degeneracy. If the nuclear spin is 0 as in the case for ^{12}C, then there is only one energy level and excitation is not possible, rendering the nucleus to be undetectable by NMR. Since quadrupolar nuclei (I > 1/2) produce a much more complex spectrum due to multiple energy states, this overview will encompass the nuclei (^{1}H, ^{13}C, ^{15}N) that possess a nuclear spin of 1/2. The relative energy difference between the two energy levels is directly proportional to the gyromagnetic ratio (an intrinsic property of the nucleus), γ, and the strength of the external magnetic field, B_o. As such, higher sensitivities can be achieved by observing nuclei with a larger gyromagnetic ratio like ^{1}H as well as utilizing stronger external magnetic fields. Since there is an energy difference between two states, it is possible to excite the nuclei as:

$$\Delta E = hv \qquad (4)$$

where h is planck's constant and v is the frequency.

Excitation is performed through the application of radio-frequency pulses for a specific duration, which leads to the rotation of the bulk magnetization from the z-axis. A complete transfer of the bulk magnetization to the xy-plane would correlate with a 90^{o}

change, which is denoted as a 90^{o} pulse. The bulk magnetization would then process about the xy-plane due to its angular momentum. The analogy commonly given is the spinning top, who spins stably if the vertical axis is aligned but has a precessional motion if the vertical axis is tilted.

The importance of the magnetization in the xy-plane is because a rotating magnetic field generates an electric current that is detectable by the instrument. The output is a digitized plot of the oscillation about the x or y plane called the free induction decay (FID). The signal decays over time due to natural relaxation processes that reduce the magnetization in the xy-plane to 0 and brings the bulk magnetization back to the z-axis. The conversion from time to frequency domain is commonly done through the Fourier Transform in order to present the data as commonly seen for 1D, 2D, and 3D spectra.

1.2.2 – Chemical shifts, Chemical Exchange, and Binding

Now, there is a slew of information that can be obtained from NMR. One piece of information comes from the chemical shift values. As stated before, the frequency for excitation depends on the external magnetic field. However, there is a local magnetic field as the nuclei also behave as bar magnets, which ultimately leads to minor, but detectable, differences in the larmor frequencies. The information of the local environments is useful for determination if the protein is folded and what kind of structure it has because the chemical shifts of amino acids in α-helix, β-sheet, or random coil have been extensively tabulated throughout the years. Furthermore, following chemical shifts and chemical shift

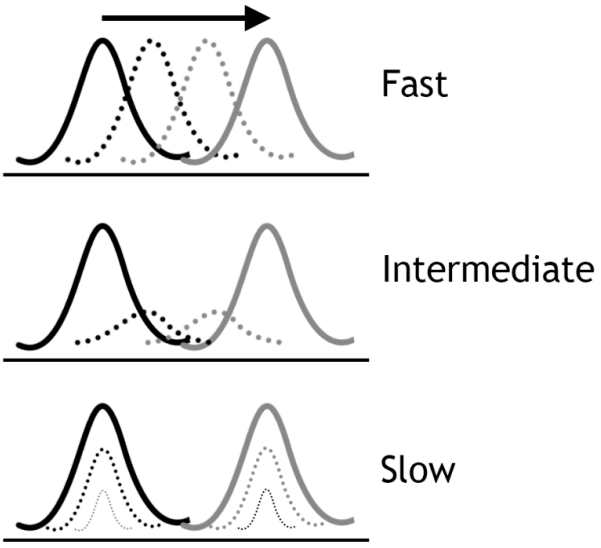

Figure 1.2. Diagram of Chemical Exchange and Motional Regimes. This depicts what we can expect to see in an NMR spectrum at the different motional regimes depending on exchange rate (k_{ex}) and chemical shift difference ($\Delta\delta$), where fast is where the $k_{ex} >> 2\pi\Delta\delta$, intermediate is $k_{ex} \sim 2\pi\Delta\delta$, and slow is $k_{ex} << 2\pi\Delta\delta$. The fast exchange regime shows an averaged peak that may show chemical shift changes when moving from 100% population A (black line) to 100% population B (grey line). The intermediate exchange also shows chemical shift changes, but is also accompanied by line broadening. Slow exchange does not show any sort of chemical shift changes, but rather shows the two distinct peaks and their relative populations.

differences are standard practices for studying protein-protein and protein-ligand interactions. This is based on the equilibrium:

$$A \rightleftarrows B \qquad (5)$$

where the protein is able to interconvert between states A and B. This can apply for multiple conformations, but the common scenario would be for binding to ligands/proteins, where A is the apo state and B is the bound state. As the chemical shift is dependent upon the localized magnetic field, any association will mean that B has different chemical shifts from A located at the direct binding site or potentially indirect conformational changes. Now, what we actually observe is based on the rate of exchange, the chemical shift difference, and populations of A and B as shown in Figure 1.2. Nonetheless, chemical shift perturbations are a useful tool for detecting binding.

1.2.3 – Paramagnetic Relaxation Enhancement

Paramagnetic Relaxation Enhancement (PRE) experiments rely on the use of paramagnetic compounds, who possess an unpaired electron. The unpaired electron, like the nucleus, is able to generate local magnetic fields. However, the unpaired electron has a much larger magnetic moment than the nucleus, so that its presence greatly enhances the relaxation rate of nearby atoms up to ~10-25Å away. Thus, one can compare the differences in spectra of a protein in the presence and absence of a paramagnetic compound, where affected residues would have a reduced signal to noise ratio due to the rapid relaxation rate. This process can be utilized so that it is possible to detect solvent exposed residues, which could be subsequently protected upon complex formation. Alternatively, covalently attaching a paramagnetic compound to the protein can provide distance restraints by deciphering which residues are nearby in space.

1.2.4 – Transverse Relaxation Time

Transverse Relaxation Time (T2) is the time it takes for the magnetization in the xy-plane to return to equilibrium. The main process that causes T2 relaxation is chemical shift anisotropy (CSA), which is the minor fluctuations in the individual localized magnetic environments. As mentioned previously, the bulk magnetization is excited and then detected through its precession within the xy-plane. However, this bulk magnetization is comprised on individual spins that are not all identical due to anisotropy, which causes the same atom (on different proteins) to precess at a different rate. This results in decoherence over time due to minor variability within the local field, where the bulk

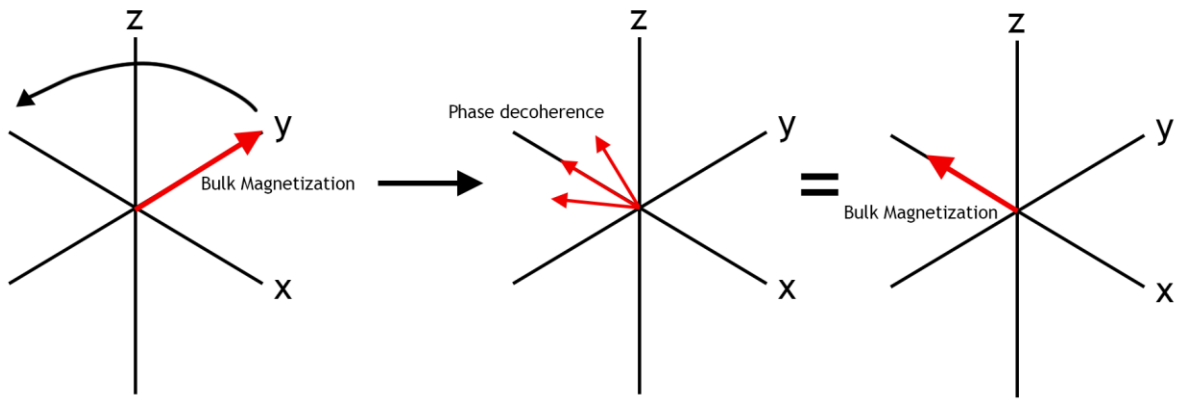

Figure 1.3. Diagram of Decoherence. This schematic shows the bulk magnetization after excitation residing on the +y axis. After a certain period of time, the spins begin to precess in the xy-plane at various frequencies, leading to phase decoherence. As a result, the bulk magnetization, which is the sum of all the individual magnetization vectors, effectively gets reduced over time.

magnetization is reduced, causing the reduction of detectable signal (Figure 1.3). This effect is related to the rotational correlation time, τ_c, which is the average time for the molecule to rotate one radian. It is important to note that τ_c is dependent upon the volume the molecule occupies, which implies that larger molecules like proteins will generally have slower correlation times and, thus, a shorter T2 in comparison to small molecules.

Another key factor that contributes to T2 relaxation is magnetic field inhomogeneities. Similarly to CSA, the precession of an atom in the xy-plane can be altered due to abnormalities in the B_o, such that the B_o felt by one atom is not the same at a different location in the NMR test tube. This is not a natural process by which relaxation occurs, but is rather due to experimental limitations in generating a uniform magnetic field through electrical currents flowing through a large superconducting coil and

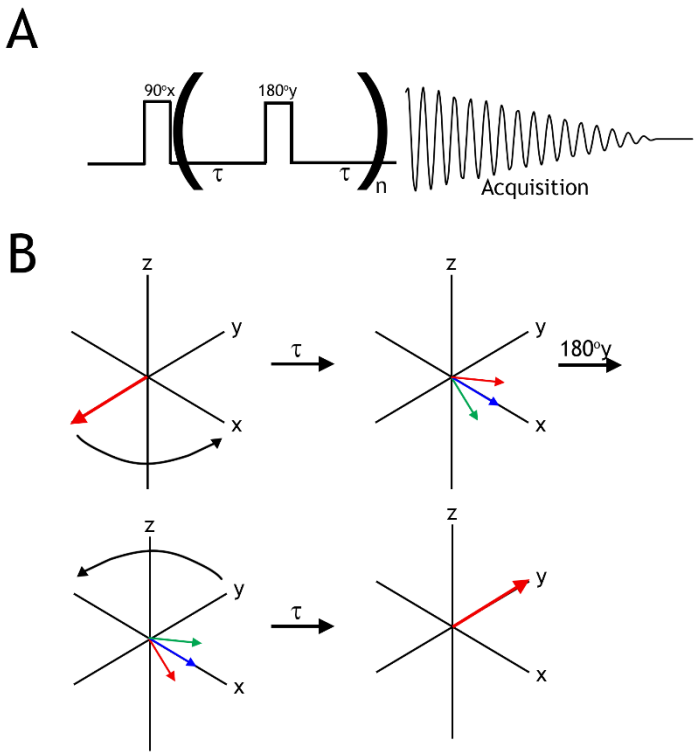

Figure 1.4. Representation of Spin Echo and Usage. (A) shows the simplest spin echo pulse sequence (CPMG), where the proton is initially excited by a 90° x-pulse. After time, τ, a 180° y-pulse is applied to invert the magnetization in the x-axis. The spins are allowed to precess for another time, τ, for refocusing before acquisition. The spin echo, τ-180°-τ, can be repeated n times as needed. **(B)** is a visual representation of the refocusing as shown by the three colored arrows, who precess at different rates.

smaller coils called shims. Fortunately, these experimental artifacts can be corrected for by applying a refocusing pulse.

A spin echo pulse sequence, where the simplest one is the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence (Figure 1.4), applies an 180° y pulse after a period of time, τ, that inverts the magnetization on the x-axis. After the spins precess for an identical period of time, τ, the magnetization will all converge to a single point (Figure 1.4). As a result, the influence of experimental artifacts are negated through the refocusing pulse and the T2 relaxation seen is the true T2 value. It is important to note that the effects of CSA are not mitigated with this pulse sequence as this is a natural process that is fundamentally tied to random molecular motions whereas field inhomogeneities are not random and dependent upon the location within the magnet.

1.3 – References

1. Holdgate, G., S. Fisher, and W. Ward, *The Application of Isothermal Titration Calorimetry to Drug Discovery*, in *Biocalorimetry 2*. 2005, John Wiley & Sons, Ltd. p. 59-79.
2. O'Brien, R. and I. Haq, *Applications of Biocalorimetry: Binding, Stability and Enzyme Kinetics*, in *Biocalorimetry 2*. 2005, John Wiley & Sons, Ltd. p. 1-34.
3. Thomson, J.A. and J.E. Ladbury, *Isothermal Titration Calorimetry: A Tutorial*, in *Biocalorimetry 2*. 2005, John Wiley & Sons, Ltd. p. 35-58.
4. Keeler, J., *Understanding NMR spectroscopy*. 2nd ed. 2010, Chichester, U.K.: John Wiley and Sons. xiii, 511 p.
5. Levitt, M.H., *Spin dynamics : basics of nuclear magnetic resonance*. 2nd ed. 2008, Chichester, England ; Hoboken, NJ: John Wiley & Sons. xxv, 714 p., 7 p. of plates.

Chapter 2 – Uncovering if PLIC-2 is a Cytoskeletal Adaptor Protein of CD47

Adapted from: **Nguyen, K.**, Puthenveetil, R., and Vinogradova, O. *Investigation of the adaptor protein PLIC-2 in multiple pathways***,** Biochem. Biophys. Rep., 2017, 9: 341-348.

2.1 – Introduction

2.1.1 – Overview of CD47

Cluster of differentiation 47 (CD47) is a ubiquitously expressed cell surface receptor [1] that consists of an extracellular IgV-like domain, five transmembrane helices, and a short unstructured cytoplasmic tail [2]. There are four different isoforms of CD47, differing only in their length of the cytoplasmic tail, where isoform 1 is the shortest and isoform 4 is the longest (Figure 2.1B). Each isoform is differentially expressed as isoform 1 is found in keratinocytes, isoform 2 is ubiquitously expressed, isoform 3 is predominantly in the liver, and isoform 4 is found within the brain and peripheral nervous system [3, 4]. Though isoform 2 is the most abundantly expressed, it is unclear if there is functional relevance between the isoforms [3].

Nonetheless, CD47 was uncovered to be a significant player in the recognition of self, a mechanism by which our immune system precludes the removal of our own cells. CD47's IgV-like domain is able to associate with signal regulatory protein-α's IgV-like domain (SIRP-α), leading to the downstream activation of protein tyrosine phosphatases

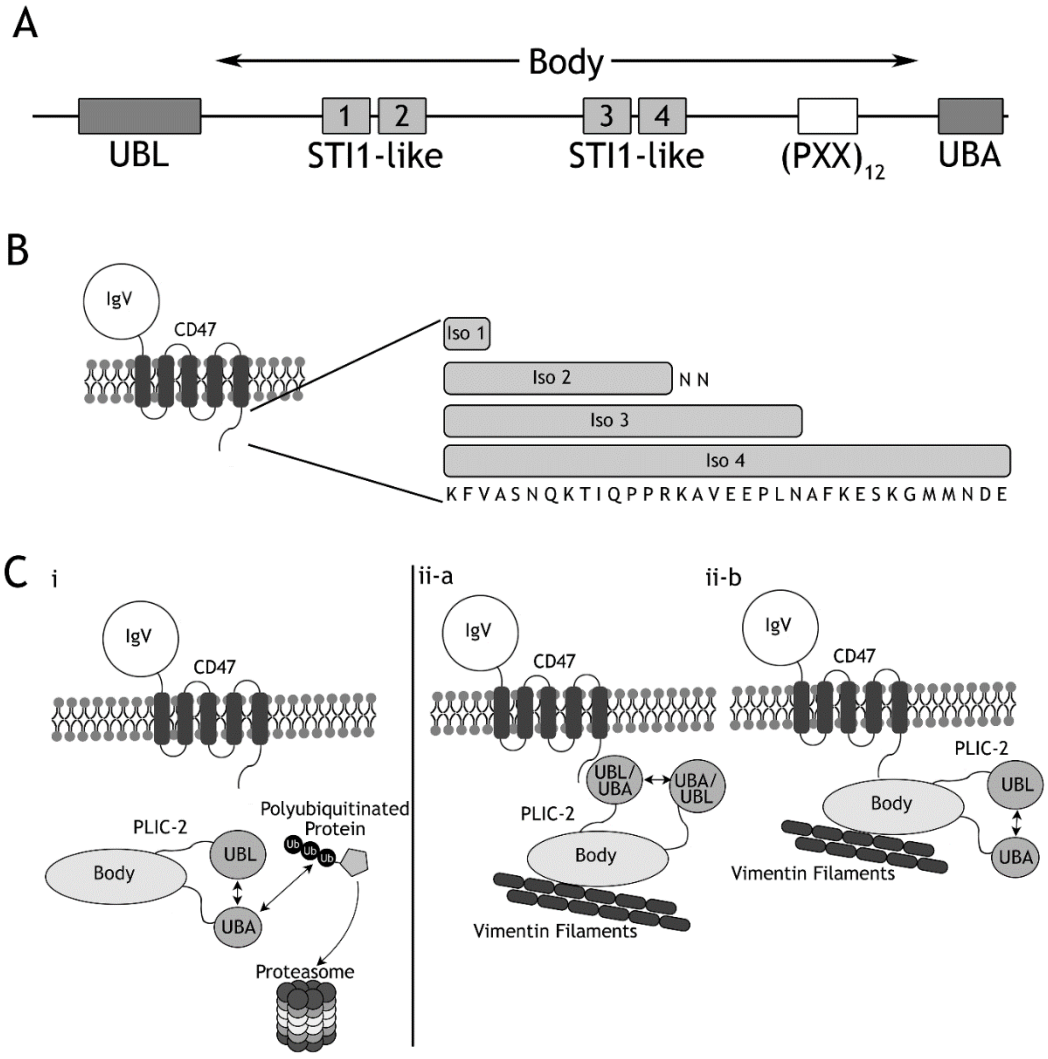

Figure 2.1. Overview of PLIC-2 and CD47 Structure as well as the Potential Pathways. (A) Schematic showing the different domains and motifs of PLIC-2. It contains a well characterized N-terminal UBL (ubiquitin like) domain and a C-terminal UBA (ubiquitin associated) domain. Interspersed between the two is the Body region which contains four Sti1 like domains and a repeat of twelve PXX motifs. **(B)** Schematic showing the overall structure of CD47. It consists of an N-terminal extracellular IgV domain, five trans-membrane helices and a C-terminal cytoplasmic tail. The tail is further differentially spliced from a single splice donor yielding the four isoforms of CD47. **(C)** A model representation showing the two pathways involving PLIC-2: **(i)** the proteasomal degradation pathway. This pathway is unrelated to the presence of CD47. Here the UBA domain of PLIC-2 binds to polyubiquitin chain bound to the substrate while the UBL domain interacts with the proteasome delivering the payload for degradation; **(ii)** the cytoskeletal regulation pathway. Central to this pathway is PLIC-2 which acts like an adaptor protein linking CD47 to Vimentin, an intermediate filament, hence by forging a link between the outer cell membrane and inner cell cytoskeleton. PLIC-2 has been proposed to interact with the cytoplasmic tail of CD47 [18]. We envision this interaction to occur in one of two ways: **(ii-a)** The UBA (or UBL) domain of PLIC-2 interacts with CD47, while the Body region interacts with vimentin filaments; **(ii-b)** the Body region of PLIC-2 interacts both with CD47 and Vimentin filaments. Meanwhile the UBL and UBA domains interact with each other.

(SHP-1 and SHP-2) and inhibition of phagocytosis, especially in the case of mature red blood cells [5-7]. Furthermore, CD47 was found to be involved in regulation of angiogenesis and blood flow through its association with thrombospondin 1 (THBS1) and vascular endothelial growth factor receptors (VEGFR) [8]. Intriguingly, CD47 is also known by the name of integrin associated protein (IAP) due to experimental evidence where it co-precipitated with integrin $\alpha_v\beta_3$ [9] and was further expanded to include integrin $\alpha_{IIb}\beta_3$, $\alpha_2\beta_1$, and $\alpha_4\beta_1$ [10]. After being found to be identical to CD47, it is no longer referred to as IAP due to the fact that it is expressed and has a functional role in erythrocytes, who lack integrins [1].

However, the molecular details of how CD47 is able to modulate cellular functions is currently unclear and has led to the interest in discovering novel proteins that may directly interact with CD47. There has not been much success, but there was a report by Wu and colleagues [11] that they had discovered two proteins that co-precipitated with the cytoplasmic tail (isoforms 2 and 4) of CD47. The proteins were found to increase CD47 related cell spreading and influenced the localization of vimentin filaments. This was a promising basis for how CD47 could modulate the cellular cytoskeleton, which is how the authors derived the name PLIC-1 and PLIC-2 (Proteins linking IAP with cytoskeleton).

2.1.2 – Who are the PLIC Proteins?

Though they were discovered and named PLIC due to aforementioned evidence, PLICs are identical to a family of proteins called ubiquilins, which are human orthologs of the yeast Dsk2 family of proteins [12]. PLIC consists of a family of four homologous

proteins, PLIC-1 through 4. All contain a ubiquitin-like (UBL) amino-terminal domain; a body region, which is predominantly predicted to be disordered by *DISOPRED* [13] and *Jpred* [14] containing STI1 (stress inducible) heat-shock protein binding motifs [15, 16]; and a carboxyl-terminal ubiquitin-associated (UBA) domain (Figure 2.1A) [17]. PLIC-1 and 2 are localized to the cytoplasm [18], PLIC-3 [19] is found in the testis and PLIC-4 is localized to the cell nucleus [20]. As mentioned previously, PLIC-1 and PLIC-2, also known as ubiquilin 1 and 2 respectively, had been shown to modulate the cytoskeleton through CD47. The direct interaction between CD47 and PLIC can be envisioned in Figure 2.1C, where it could be the individual UBL/UBA or the body region in direct association with CD47. The linkage to vimentin is most likely mediated through the body region.

However, there have been conflicting data arising, where another study showed that double labeled immune-fluorescence of PLIC-1 and vimentin filaments did not show any co-localization [21] and that PLIC-1, but not PLIC-2, binds to Gβγ subunit of G proteins and interferes with its functionality. This was based on the observation that PLIC-2 was predominantly present in the cytoplasm and never co-localized to the plasma membrane [22]. In addition, PLIC-1, but not PLIC-2, has been shown to inhibit Jurkat T-cells migration by interacting with the βγ-subunit of the heterotrimeric G-protein [22]; and stabilize intracellular $GABA_A$ receptors to promote their accumulation within the plasma membrane [23]. Interestingly, PLIC-2, but not PLIC-1, down regulates endocytosis of the two unrelated GPCRs, V2 vasopressin receptor and beta-2 adrenergic receptor [24]. Though these studies highlight PLICs involvement in diverse cellular functions as a key

regulatory molecule, they seem to contradict its primary discovered role as a structural adaptor protein linking IAP/CD47 to cell cytoskeleton.

Furthermore, PLICs have been implicated to be involved in the ubiquitin mediated proteasomal degradation pathway. The primary machinery for this is the 26S proteasomal complex that is composed of a 19S regulator complex and 20S core complex, where polyubiquitinated proteins ultimately are degraded [25]. Evidence has been shown that the S5a subunit of the 19S component binds to the UBL domain of PLIC-2 [26]. Given that PLICs contain both a UBL and UBA domain and that other UBL-UBA shuttle proteins like Rad23 and Dsk2 have been characterized [27], it may be that PLIC serves as a shuttle for proteosomal degradation (Figure 2.1C) rather than a cytoskeletal adaptor protein.

It is a bit puzzling as to how two diverse modalities are built into a single structural design of these proteins. It seems pertinent to ascertain whether the UBL/UBA system is directly involved in linking cell surface receptors to intermediate filaments or whether it acts as a shuttle protein delivering its payload to the proteasomal complex for degradation. With the aim to reconcile these differences and understand which of the two pathways PLIC actually follows, we set to investigate the individual components involved in the interactions. We focused on uncovering if there was UBL/UBA self association or UBA/Ubiquitin binding, which are commonly seen for UBL-UBA shuttle factors, and if there was direct association with the cytoplasmic tail of CD47 isoform 4.

2.2 – Materials and Methods

Cloning of PLIC-2 and CD47 Constructs

PLIC-2/UBLN-2 (Accession number: Q9UHD9) UBL domain UBLN2$^{1-107}$ present in pET23a vector, was kindly provided by Dr. Walters (NCI. USA). PLIC-2 UBA domain UBLN2$^{578-624}$ was sub-cloned into pET15b vector using BamHI and HindIII restriction sites. PLIC-2 Body region UBLN2$^{108-580}$ was sub cloned into pET15b using NdeI and BamHI restriction sites. To ensure accurate measurement of the recombinant protein concentration, a single Tryptophan residue was inserted into the corresponding vectors, for both UBL and UBA, following the six Histidine tags.

We used the longest isoform of CD47 (Isoform-4) as classified in the initial paper [18]. As per the Uniprot classification, this is annotated as Isoform-1. For this work, we will be following the classification from the original paper. Two constructs of CD47 isoform-4 (Accession no.: **Q08722-1)** were designed for this study, (i) CD47 $^{290-323}$ containing the residues comprising its cytoplasmic tail (CD47-CT) and (ii) CD47 $^{269-323}$ containing the residues comprising its last transmembrane helix along with its cytoplasmic tail (CD47-TMCT). The resulting amplicons were cloned into pET15b (for CT) (Novagen, USA) and pDB.His.MBP vector (for TMCT) (DNASU plasmid depository, USA) using the cutting sites NdeI and BamHI for the former and BamHI and HindIII for the latter. The pDB.His.MBP vector contains a Histidine-affinity tag and a MBP-fusion protein, separated from CD47's sequence by a TEV protease cleavage site. Three additional glutamates were added before the N-terminal sequence to ensure proper accessibility of the cleavage site for proteolysis.

Expression and Purification of the Recombinant Proteins

Expression

All constructs were expressed in *E. coli* BL21 (DE3) (NEB, USA). Cell cultures were grown in either LB or M9 minimal media supplemented with ^{15}N-labeled ammonium chloride as the sole nitrogen source at 37^{o}C. When the cells reached an optical density (OD_{600}) of ~0.4 to 0.6, the cultures were transferred to a shaker at room temperature (for most proteins) or a 17^{o}C shaker (for PLIC-2 Body region), and left overnight after induction with 0.5-1mM IPTG.

Purification

UBA, UBL, Body, and CD47 constructs were purified using similar protocols with procedure variations outlined below. Harvested cells were resuspended and lysed using a French Press (Thermo Electron, USA). The supernatant was bound to Ni-NTA resin (Qiagen, USA) at 4^{o}C for an hour, rinsed with ten column volumes of the wash buffer before elution. The eluate was further subjected to size exclusion chromatography (SEC) on a Superdex 75 (GE Healthcare, USA) for UBL and UBA domains, or Superdex 200 (GE Healthcare, USA) for CD47-TMCT. In addition to Ni-NTA purification the Body region was further purified using anion exchange chromatography on Resource Q (GE Healthcare, USA), with a 0-30% gradient of Buffer A (20mM Tris, pH 8.0) and Buffer B (20mM Tris, 1M NaCl, pH 8.0). The Body containing fractions were further subjected to SEC on a Superdex 75 column.

CD47-CT expressed in inclusion bodies was purified under denaturing condition. Briefly, cells were resuspended in 6M GuHCl in Tris buffer saline (TBS) buffer pH 8.0. Following lysis the supernatant was separated through centrifugation at 10,000 rpm and mixed with Ni-NTA resin at room temperature for 2 hours. The resin was rinsed first with

four column volumes of 8M urea in TBS with 10mM imidazole. Elution was carried out in the 8M Urea buffer containing 400mM imidazole. The eluate was loaded onto a Proto 300 C4 column (Higgins Analytical Inc., USA) and purified using a 10-40% gradient of (Buffer A: 90% H_2O, 10% Acetonitrile, 0.1% TFA and Buffer B is 10% H_2O, 90% Acetonitrile, 0.1% TFA). CD47-CT containing fractions were pooled and lyophilized (MillRock Tech., USA).

Nuclear Magnetic Resonance Spectroscopy

^{15}N-labeled proteins were overexpressed in M9 medium supplemented with $^{15}NH_4Cl$ as the sole nitrogen source. ^{15}N-NMR experiments were performed with a concentration of 75-150 µM on a Varian 600 MHz spectrometer equipped with an inverse triple-resonance cold probe at 25^{o}C. All spectra were processed with NMR Pipe [28] and analyzed in CcpNmr software suite [29] made available through NMRBox.

For CD47-TMCT nanodiscs experiments, ^{15}N-CD47-TMCT was titrated with unlabeled UBL at a ratio of 1:5 and Body region at 1:3. ^{15}N-UBL was titrated with unlabeled CD47-TMCT at a ratio of 1:5, whereas ^{15}N-UBA was titrated with unlabeled CD47-TMCT at a ratio of 2:5. For CD47-CT experiments, ^{15}N-UBL was titrated with unlabeled CD47-CT at a ratio of 1:10 and ^{15}N-UBA was titrated with unlabeled CD47-CT at a ratio of 1:10. For PLIC-2 self-association experiments, ^{15}N-UBA was titrated with unlabeled UBL at ratios of 1:1, 1:3, 1:5, and 1:10. For UBA and ubiquitin experiments, ^{15}N-UBA was titrated with unlabeled ubiquitin at ratios of 3:1, 2:1, 1:1, 1:2, 1:3, 1:5, 1:10. K_d values were determined form the chemical shift perturbations calculated using the following equation [30]:

$$\Delta\delta = \sqrt{0.5\left(\delta_H^2 + (\alpha\delta_N^2)\right)} \quad (1)$$

where the scaling factor, α, was set to 0.15. The plot of the chemical shift perturbation vs. molar ratio was fitted to the equation below to define the K_d [31]:

$$\Delta\delta_{obs} = 0.5\Delta\delta_{max}\left(1 + x + \frac{K_d}{[P]_0} - \sqrt{\left(1 + x + \frac{K_d}{[P]_0}\right)^2 - 4x}\right) \quad (2)$$

where $[P]_0$ is the concentration of ^{15}N-UBA PLIC-2 and x is the molar ratio of UBL or ubiquitin to UBA.

Isothermal Titration Calorimetry

All ITC experiments were performed using a low volume Nano-ITC (TA instruments, USA). The titrations were done at 25^{o}C with 200 RPM mixing, 300 s injection intervals, and 2.5 µL injections. The buffer used was 40 mM $NaPO_4$, 20mM NaCl, 0.5 mM EDTA, pH 7.0. The protein concentrations for the syringe and sample cell were as follows: 0.95 mM UBL PLIC-2 and 0.14 mM CD47-CT, 0.39 mM UBA PLIC-2 and 0.05 mM CD47-CT, 0.49 mM CD47-CT and 0.05 mM Body PLIC-2, and 0.49 mM CD47-CT and 0.04 mM Body PLIC-2. For nanodiscs, 0.35 mM UBL PLIC-2 and 0.07 mM CD47-TMCT; 0.32 mM UBA PLIC-2 and 0.07 mM CD47-TMCT. Data analysis was done with the NanoAnalyze Software (TA Instruments, USA) suite using “independent” model algorithm.

Molecular Modeling

PLIC-2 UBA was modeled through Phyre 2 [32] web server using PLIC-1 UBA domain as a template which shares 98% identity. Models of the protein complexes (PLIC-2 UBA-UBL and PLIC-2 UBA-Ubiquitin) were calculated through the HADDOCK (High

Ambiguity Driven protein-protein DOCKing) [33] webserver using the chemical shift perturbation data obtained from NMR HSQC titration experiments. The structures of the participating proteins were acquired from the RCSB database (PLIC-2 UBL: 1J8C, Ubiquitin: 2JY6, PLIC-2 UBA: 2JY5).

2.3 – Results and Discussion

2.3.1 – Determination of the UBA/UBL Self Association

We first tried to ascertain the involvement of PLIC-2 in the proteasomal degradation pathway. We initiated this by determining whether the two terminal UBA and UBL domains of PLIC-2 interact with each other (Figure 2.1C (i)). ^{15}N-HSQC based chemical shifts perturbation (CSP) experiments were employed, where ^{15}N labeled PLIC-2 UBA was titrated against unlabeled PLIC-2 UBL. Peaks representing maximum observed perturbations (marked with asterisk on Figure 2.2), were used to calculate the K_d values using the equations summarized in the methods section. The average K_d value obtained was 175±25 µM, which is comparable to the previously reported K_d value of 80±15 µM measured for Dsk2, a yeast homolog of PLIC proteins, [34]. It is known that for UBL-UBA proteins involved in proteasomal degradation, UBL interacts with the proteasome while UBA binds to polyubiquitinated substrates [17].

From reliable partial assignments transferred from PLIC-1's UBA domain [35], we identified most of the PLIC-2 UBA residues which undergo perturbations in the presence of PLIC-2 UBL. These residues either belong to the loop region connecting helices α1 and α2 (M^{592}-F^{594}) or are located within helix α3 (I^{615}-E^{616} and L^{619}) of which I^{615} showed minimal shifts. These positions match well with the corresponding Dsk2 residues M^{342}-

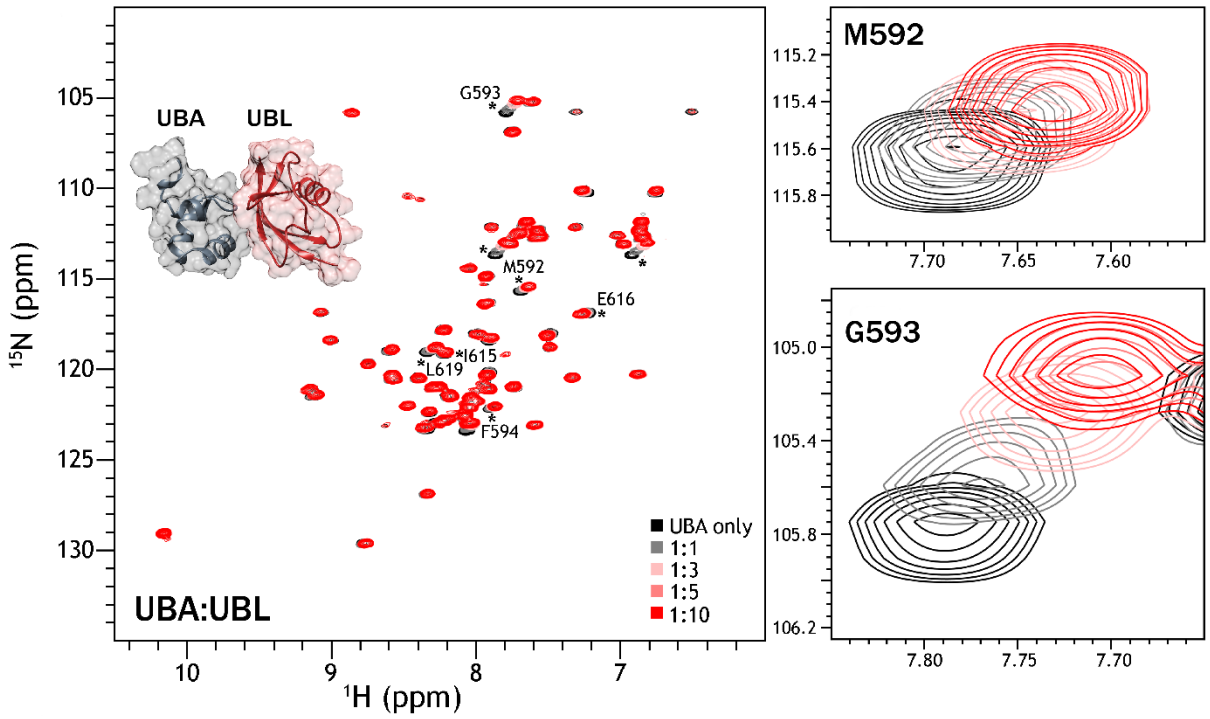

Figure 2.2. PLIC-2's UBA Domain is Able to Self-Associate with its UBL domain. ^{15}N-HSQC titration experiments of PLIC-2 ^{15}N-UBA domain in the presence of PLIC-2 UBL domain at a UBA to UBL molar ratio of 1:1, 1:3, 1:5, and 1:10. Partial amino acid assignments were achieved by aligning the ^{15}N-HSQC to that of PLIC-1 UBA domain [35]. The inset shows a molecular model displaying the interaction of UBA (grey) and UBL (red) as determined by HADDOCK. Two residues M and G showing maximum perturbation are shown as slices on the right.

F^{344} from the same loop region and L^{365}-D^{366}, L^{369} from the α3 helix. Defining the binding site are residues M$^{342/592}$, which fits snugly in the hydrophobic cavity of UBL, and L^{365}/I^{615} and L$^{369/619}$, which reinforce the hydrophobic core. It is likely that the flexibility required for the two terminals domains to interact with each other is satisfied by having a Body region that is largely disordered.

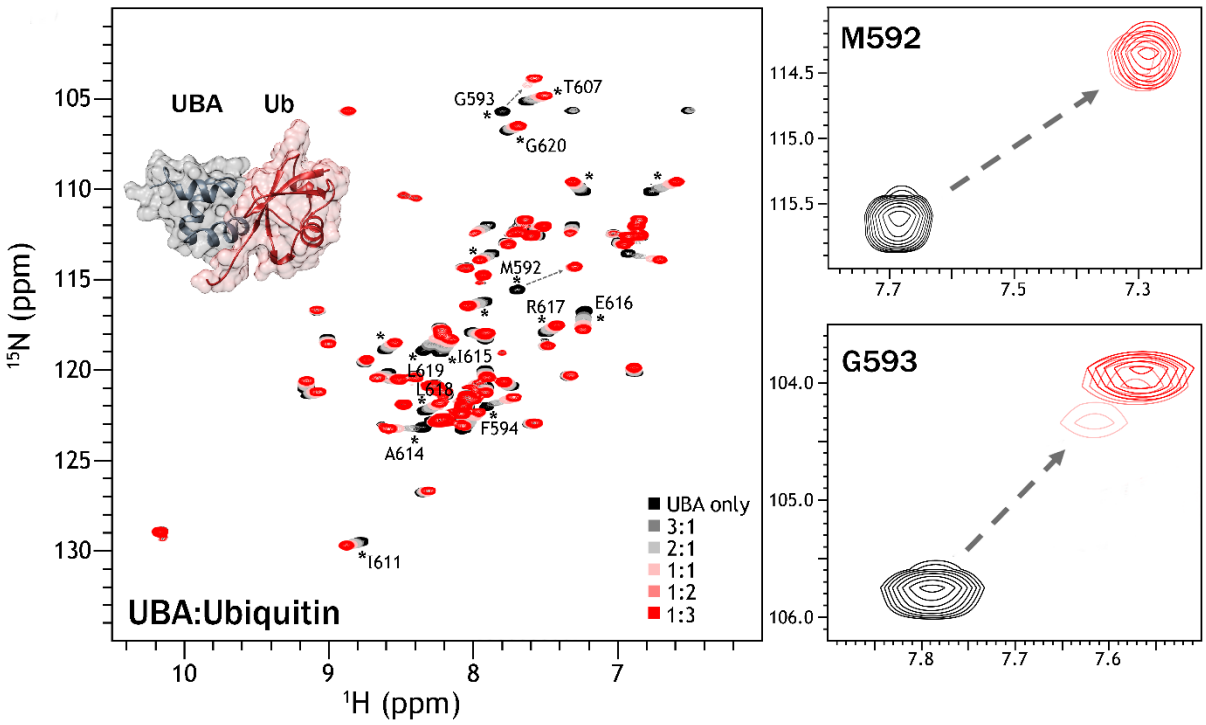

Figure 2.3. PLIC-2's UBA Domain Exhibits Large Chemical Shift Perturbations upon Titration with Ubiquitin. ^{15}N-HSQC titration experiments of PLIC-2 ^{15}N-UBA domain in the presence of ubiquitin (Ub) at a UBA to Ub molar ratios of 3:1, 2:1, 1:1, 1:2, and 1:3. Ratios 1:5 and 1:10 are not shown as the spectra reached a saturation at 1:3. A molecular model, determined by HADDOCK, of UBA PLIC-2 (grey) and ubiquitin (red) as inset in the spectra. On the right, specific peaks are shown separately to better capture the chemical shift perturbations. Two residues M and G displaying intermediate exchange are shown as slices on the right.

2.3.2 – Uncovering if UBA and Ubiquitin Interacts

Thus, we next examined the ability of PLIC-2 to bind ubiquitinated substrates through the interaction of PLIC-2 UBA domain with Ubiquitin (Ub). Our titration experiments demonstrate a stronger binding affinity of UBA for Ub, as exemplified by the pronounced CSPs in ^{15}N-HSQC spectra (Figure 2.3). The K_d value for this interaction was

determined to be 8.1±1.5 μM by averaging fits from peaks with pronounced perturbations, which is similar to PLIC-1's affinity for ubiquitin (K_d ~20 μM) [35]. CSP data was further utilized to generate a model of the binding interface between UBA-UBL and UBA-Ub using HADDOCK, which is presented as inset within the spectra in Figures 2.2 and 2.3. Intermolecular interaction between UBA and Ub is stronger than with UBL and their binding interfaces involves the same core residues in addition to other exclusive to the UBA-Ub interactions reflective of stronger binding. CSP data obtained from our titration experiments of ^{15}N UBA with Ub highlights residues M^{592}-F^{594} from the (helix α1/ α2) loop region and A^{614}-L^{619} belonging to helix α3 that are involved in this interaction. Majority of the perturbations belong to the fast exchange regime as only chemical shift perturbations were observed. However, residues M^{592}-G^{593} displayed both chemical shift perturbations and peak broadening indicative of intermediate exchange. This is part of the canonical, M-G-F containing binding site conserved throughout UBA domains of PLIC/Dsk2 and hHR23/Rad23 proteins. In addition, I^{615}-E^{616} and L^{619} from α3 showed significant perturbations, as has been observed for I^{580}-E^{581} and L^{584} form PLIC-1 where they form close contacts with the hydrophobic patches of Ub [36]. Interestingly, T^{607}, the last residue from the helix α2 showed shifts which can be attributed to its close proximity to A^{614} on α3 which exhibited strong CSP a feature also observed with PLIC-1. All together, we can conclude that the UBA-UBL and UBA-Ub interactions in PLIC-2 share the same binding interface, which is exemplified in the HADDOCK models presented as insets in Figure 2.2 and 2.3.

Now, in order for PLIC proteins to participate in the proteasomal degradation pathway, two events need to occur. First, the UBL domain needs to interact with the

proteasome and, second, the UBA domain needs to interact with ubiquitin. The former has already been established for PLIC-2, where the UBL domain has been shown to complex with hRpn13 (Adrm1) [37], which is a subunit of the 19S regulatory particle of the 26S proteasome. In this chapter, we determined that PLIC-2 UBA domain is able to associate with ubiquitin. From our K_d estimation, the PLIC-2 UBA-Ub binds stronger than UBA-UBL. Given that UBA and UBL domains associate with each other, a stronger affinity of the PLIC-2 UBA domain for ubiquitin implicates the disengagement of the UBA-UBL domains in the event of PLIC binding to the polyubiquitinated substrate en route to the proteasome. This strongly suggests the role of PLIC-2 as a shuttle protein transferring its cargo to the proteasome for degradation.

2.3.3 – Evaluating Binding of CD47 to the UBL and UBA Domains

Next, we moved on to investigate the second half of our model, testing PLIC-2 as an adaptor protein linking the cell membrane to the cytoskeleton through the cytoplasmic tail of CD47 (Figure 2.1Cii). This was achieved by determining interactions occurring between CD47 and (a) the two terminal UBA/UBL domains of PLIC-2 or (b) the Body region of PLIC-2. Two constructs of CD47 were utilized, one containing the cytoplasmic tail region (CT) and the other comprising of the last transmembrane helix in addition to the cytoplasmic tail (TMCT). This helped investigate interaction at both the aqueous and membrane/water interface. Since CD47 is a mammalian receptor protein, we assessed the possibility of posttranslational modifications (PTM) using several PTM databases from Uniprot. We found that although there are ten positions that potentially undergo PTM, the last is located at residue 206. As our longest CD47 construct (TMCT) starts from residue

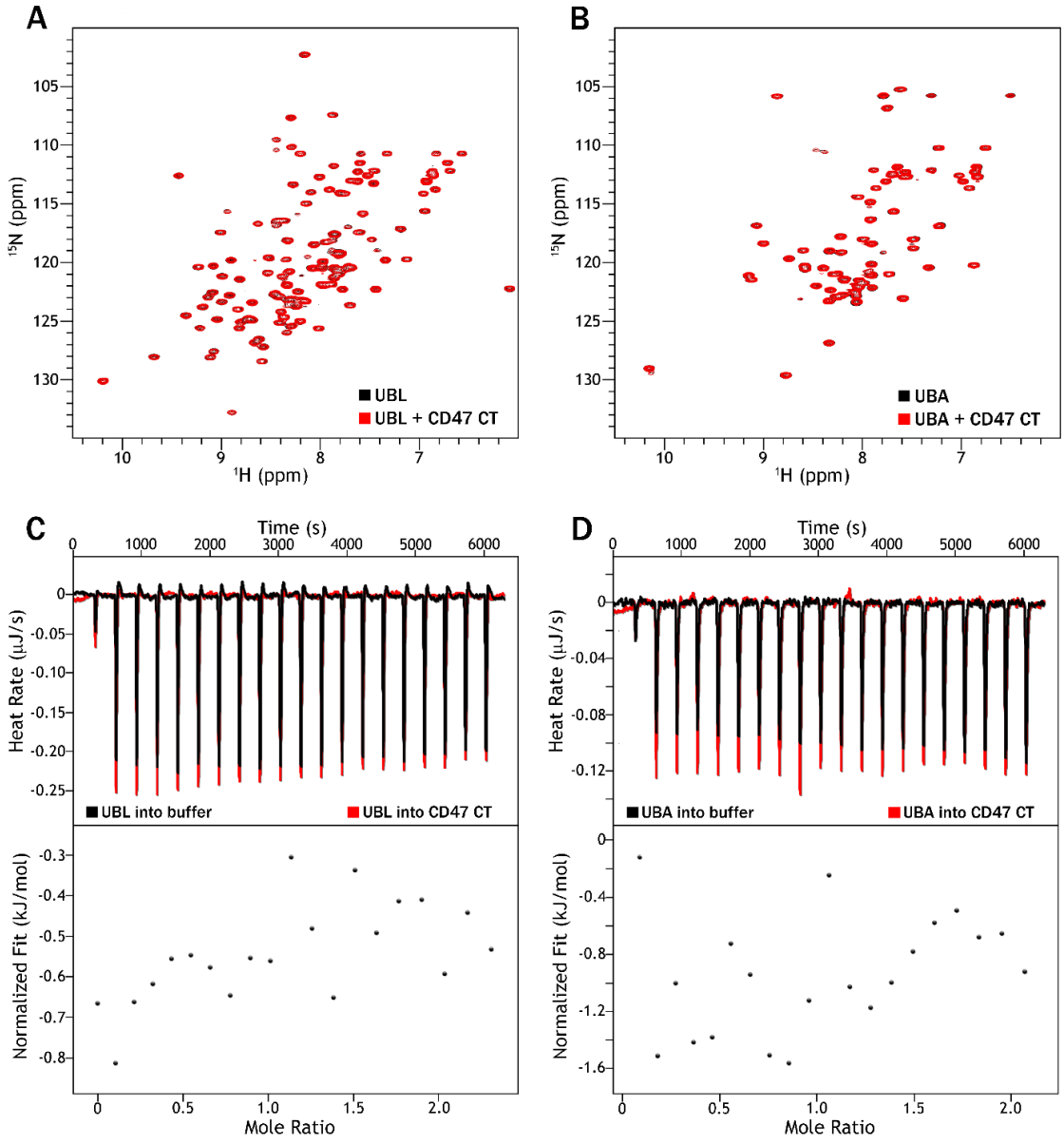

Figure 2.4. Absence of Any Observable Interaction Between CD47-CT and PLIC-2 UBA/UBL Domains. (A) ^{15}N-HSQC of PLIC-2 ^{15}N-UBL alone (black) and in the presence of CD47-CT at a UBL to CD47 ratio of 1:10 in red. **(B)** ^{15}N-HSQC of PLIC-2 ^{15}N-UBA alone (black) and in the presence of CD47-CT at a UBA to CD47 ratio of 1:10 in red. **(C)** (Top) ITC thermogram of UBL titrated into buffer (black) and into CD47-CT (red). (Bottom) Integrated heats of the titration after correction for the heat of dilution. **(D)** (Top) ITC thermogram of UBA titrated into buffer (black) and into CD47-CT (red). (Bottom) Integrated heats of the titration after correction for the heat of dilution.

number 269, no PTMs affect the outcome of our study, validating the biological significance of our *in-vitro* analysis.

In order to ascertain whether CD47 interacts with the UBL and/or UBA domains of PLIC-2, we performed ^{15}N-HSQC titration experiments with ^{15}N-UBL or ^{15}N-UBA. Figures 2.4A and 2.4B present the spectra of the UBL and UBA domains, respectively. Both remained unperturbed by the addition of the cytoplasmic tail of CD47 (CD47-CT) with no CSPs or peak broadenings. To reinforce our results, we performed Isothermal Titration Calorimetric (ITC) experiments where CD47-CT was titrated against the UBL (Figure 2.4C) or UBA domains (Figure 2.4D). The acquired thermograms reflect the heats of dilution rather than specific enthalpy driven interactions consistent with the lack of binding.

The other pertinent question to this interaction is the proximity of the tail region to the membrane surface, as in its native receptor state. To identify any interactions at this interface, we incorporated the CD47-TMCT, containing the last transmembrane helix, into small D7 nanodiscs [38]. Nanodiscs offer a soluble lipid bilayer system allowing the study of interactions at the membrane/water interface. We have previously successfully employed NMR for the study of the TMCT regions of Integrin β_3 in nanodiscs [39]. Work done by Dr. Puthenveetil showed that nanodiscs formation was successful by TEM and that there was no binding by NMR or ITC with the CD47-TMCT construct, indicating that the addition of a membrane mimetic does not produce any observable binding between CD47 and the UBL or UBA domains of PLIC-2.

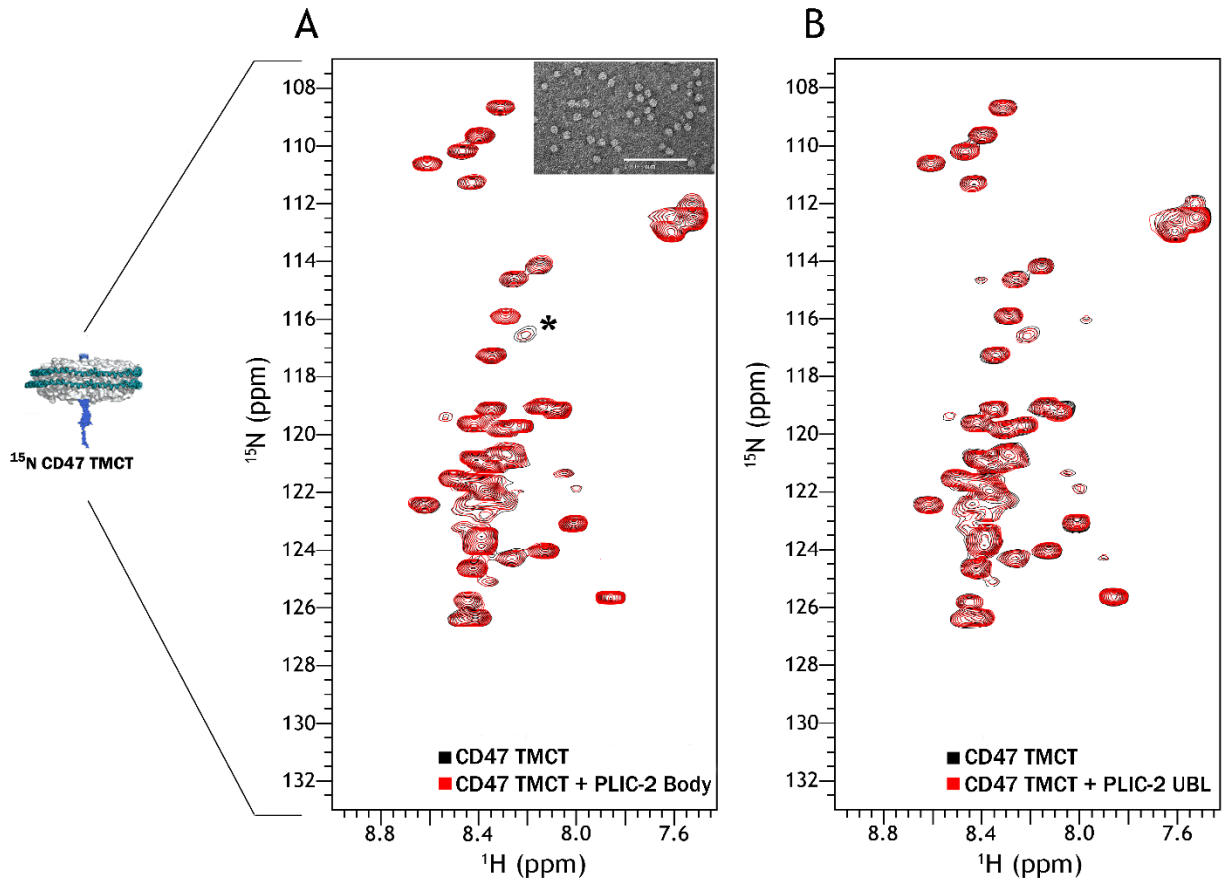

Figure 2.5. PLIC-2 Body Region Does not Bind to CD47-TMCT. ^{15}N-HSQC of labeled-CD47 TMCT. **(A)** ^{15}N-CD47 TMCT alone (black) and in the presence of PLIC-2 Body region at a CD47 to Body ratio of 1:3 in red. The amide peak with a slight intensity drop in the presence of Body is marked by the asterisk. Inset: Negatively stained TEM micrograph showing the formation of nanodiscs containing the incorporated ^{15}N labeled protein **(B)** ^{15}N-CD47 TMCT alone (black) and in the presence of PLIC-2 UBL domain at a CD47 to UBL ratio of 1:5 in red.

2.3.4 – Investigation of the CD47 and PLIC-2 Body Region

Since no interactions were observed between the PLIC-2 UBA/UBL domains and CD47, we focused our attention on the PLIC-2 Body region. This is a 51.6 kDa intrinsically disordered protein that would provide a ^{15}N-HSQC spectrum with a multitude of overlapping peaks corresponding to random coil frequencies making it difficult to observe

any CSP or peak broadenings leading to inconclusive and ambiguous data. Therefore, we tried to analyze this interaction from the side of CD47 by titrating ^{15}N-CD47-TMCT with unlabeled Body region. We also performed the same experiment with the unlabeled UBL domain serving as a control; and additionally confirming the above obtained results for the ^{15}N-UBL and unlabeled CD47-TMCT experiment. Figure 5 shows the spectra of ^{15}N-CD47-TMCT alone (black) titrated against unlabeled proteins, Body (Figure 2.5A) and UBL (Figure 2.5B) shown in red. Both titration resulted in no pronounced changes in resonance frequencies or peak intensities of CD47. There was however, a slight intensity drop observed for the amide located at 8.20/116.56 ppm in the presence of the Body region (marked with the asterisk in Figure 2.5A); but a minor intensity drop for a single peak does not substantiate a binding interaction between two proteins.

The absence of binding was further confirmed through ITC measurements using the CD47-CT construct (Figure 2.6). ITC measurements were performed to confirm the lack of binding from the NMR data between CD47 and PLIC-2 Body region. The initial titration of the Body region with CD47-CT showed some heat differences that could be interpreted as a binding curve. However, after careful analysis of the full data set, we concluded that it wasn't the case. Our rationale is based on the following arguments. First, the overall heat differences are low, so the data is close to the measurement limits and is at the noise level making each single titration unreliable. To overcome this limitation, we performed two titrations with a constant concentration of CD47-CT as a titrant into two different concentrations of the PLIC-2 Body protein present inside the measurement cell. If the observed calorimetric isotherms and the corresponding titration curves reflect a real binding event, we expect the lower concentration of PLIC-2 to

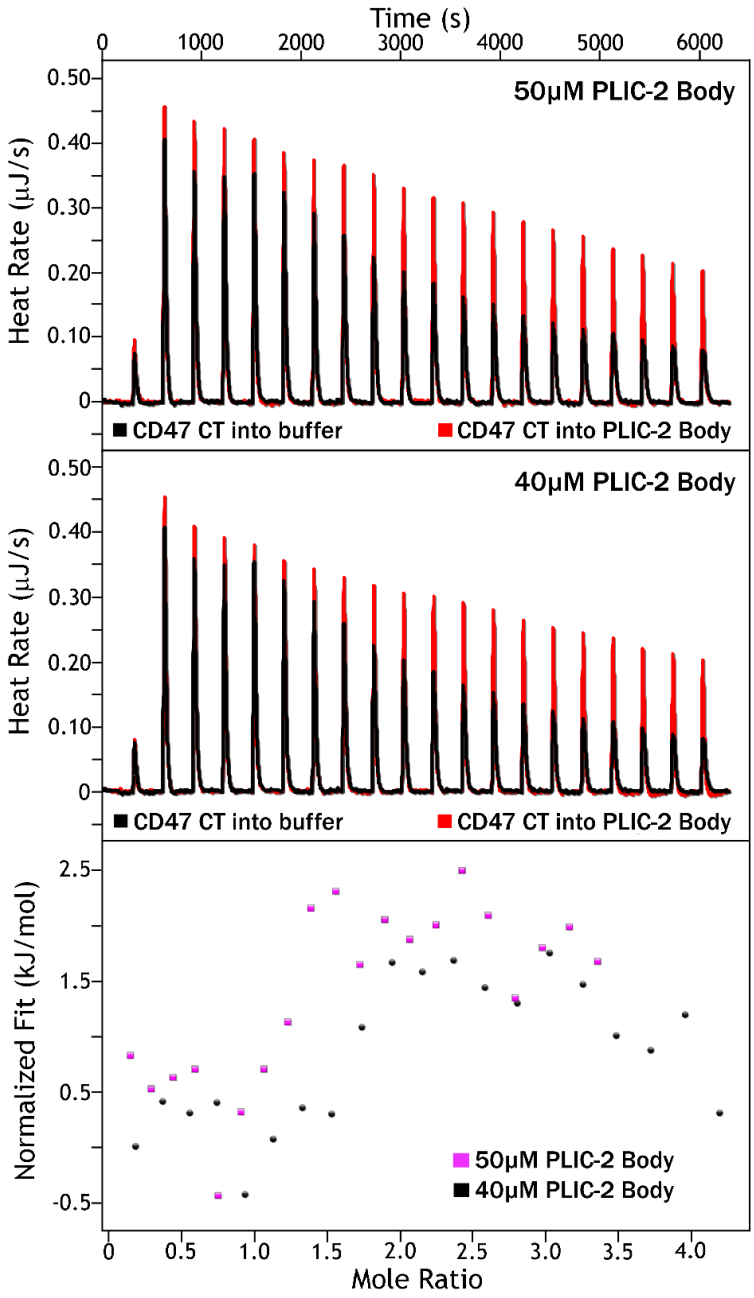

Figure 2.6. ITC Confirms the Lack of Interaction Between PLIC-2 Body Region and CD47. ITC thermograms of CD47-CT titrated into buffer (black) or into PLIC-2 Body region (red) at two different concentrations of PLIC-2 protein. The normalized fit shows the heats for both 50µM and 40µM PLIC-2 Body region.

saturate earlier owing to a lower requirement for a titrant concentration. In fact, we got opposite results, which indicates the artificial nature of observed curves as the point of saturation for 50µM of Body came sooner than the one for the 40µM titration. Lastly, if we forcefully fit the thermogram to obtain a binding constants, the data would place a K_d value into the low 100 nM range, which is a relatively strong binding affinity and should be easily detectable by NMR spectral analysis (as we demonstrated earlier for even weaker PLIC-2 inter domain interactions which were the µM range). Taken together, we believe that the observed shifts in heat values were likely produced due to proteins aggregation during titration, manifested by acquired samples cloudiness visible upon experiments completion.

2.3.5 – Does PLIC-2 Directly Associate with CD47?

The evidence presented here clearly indicates that PLIC-2 does not associate with CD47 *in vitro*. Though it seems to contradict the report from Wu and colleagues [11], the inference that the two may not be directly associated is not inconceivable when we factor in the following reasons. It is important to keep in mind that the work done here is performed on isolated domains of PLIC-2 (UBA, UBL, and Body region), where the full length protein may adopt an overall fold necessary for interacting. For instance, the UBA/UBL self association may be fairly weak as presented here (K_d = 175±25 µM), but there is evidence of PLIC-2 forming dimers/trimers [40] possibly through the Body region [41] , which may promote intermolecular UBA/UBL association.

Furthermore, it is probable that CD47 and PLIC-2 are part of a larger macromolecular assembly that may also contain, for example, integrin heterodimers [42].

The sole *in vitro* data available for PLIC's interaction is from a pull down assay using recombinant GST-fused PLIC probed against CD47 from human placenta and analyzed through immunoblots [11]. This doesn't rule out the presence of other proteins that could have been part of the pull down result and hence doesn't exclude the possibility that other protein(s) might be involved in the complex formation. Silver staining of the pull down eluate followed by subsequent LC-MS might help identify these components. Since this experiment was not performed in the original study, the direct interaction of CD47 to PLIC remains a conjecture.

On the other hand, there is a possibility that PLIC-2's role may not be a cytoskeletal adaptor protein at all, but rather serves an entirely different function altogether. Several studies have shown the involvement of PLICs in alternative cellular processes, such as proteasomal degradation [27, 43], GPCR endocytosis [24], clathrin mediated endocytosis through Eps15 and and Epsin (1 and 2) [24], and evidence in this chapter indicating that it is able to bind to ubiquitin. To complete the picture, recent evidence has arisen illustrating the functional role of the Body region, which had been elusive thus far as there was no established binding partner despite being predicted to contain six STI1 heat shock binding motifs. A recent study by Hjerpe and colleagues has implicated the Body region to associate with heat shock protein 70 (HSP70) whom is bound to an aggregated client protein. The entire complex (PLIC-2/HSP70/Target protein) is then transferred to the proteasomal complex for degradation. Specific mutations linked to amyotrophic lateral sclerosis located within the Body region were shown to abolish PLIC-2/HSP70 complex formation [40], highlighting the significance of recycling protein aggregates. The data

suggests that the Body region is responsible for payload transfer and potentially substrate recognition, further consolidating PLIC-2's role as a proteasomal shuttle factor.

2.4 – Conclusion

Even though we employed high sensitivity/ high-resolution biophysical techniques, we did not observe direct interactions between the cytoplasmic tail of CD47 and any of the individual domains of PLIC-2. Both ITC and NMR results consistently refute the notion of a binary complex. Though it is possible that the conditions for our *in vitro* analysis may not have been ideal, the evidence of UBA/UBL self association and UBA/Ubiquitin complex formation coincides with the current literature showing that PLIC-2 belongs to the proteasomal degradation pathway. This raises questions against its involvement as a protein linking the membrane to the cytoskeleton and leaves one to ponder whether the name Ubiquilins should be preferred over PLIC while referring to these adaptor proteins.

2.5 - References

1. Oldenborg, P.A., *CD47: A Cell Surface Glycoprotein Which Regulates Multiple Functions of Hematopoietic Cells in Health and Disease.* ISRN Hematol, 2013. **2013**: p. 614619.
2. Lindberg, F.P., et al., *Molecular cloning of integrin-associated protein: an immunoglobulin family member with multiple membrane-spanning domains implicated in alpha v beta 3-dependent ligand binding.* J Cell Biol, 1993. **123**(2): p. 485-96.
3. Sick, E., et al., *CD47 update: a multifaceted actor in the tumour microenvironment of potential therapeutic interest.* Br J Pharmacol, 2012. **167**(7): p. 1415-30.
4. Reinhold, M.I., et al., *In vivo expression of alternatively spliced forms of integrin-associated protein (CD47).* J Cell Sci, 1995. **108 (Pt 11)**: p. 3419-25.
5. Murata, Y., et al., *The CD47-SIRPalpha signalling system: its physiological roles and therapeutic application.* J Biochem, 2014. **155**(6): p. 335-44.
6. Oldenborg, P.A., et al., *Role of CD47 as a marker of self on red blood cells.* Science, 2000. **288**(5473): p. 2051-4.
7. Barclay, A.N. and T.K. Van den Berg, *The interaction between signal regulatory protein alpha (SIRPalpha) and CD47: structure, function, and therapeutic target.* Annu Rev Immunol, 2014. **32**: p. 25-50.
8. Soto-Pantoja, D.R., et al., *Therapeutic opportunities for targeting the ubiquitous cell surface receptor CD47.* Expert Opin Ther Targets, 2013. **17**(1): p. 89-103.
9. Brown, E., et al., *Integrin-associated protein: a 50-kD plasma membrane antigen physically and functionally associated with integrins.* The Journal of cell biology, 1990. **111**(6 Pt 1): p. 2785-94.
10. Makino, S., et al., *Cell-free protein synthesis technology in NMR high-throughput structure determination.* Methods Mol Biol, 2010. **607**: p. 127-47.
11. Wu, A.L., et al., *Ubiquitin-related proteins regulate interaction of vimentin intermediate filaments with the plasma membrane.* Mol Cell, 1999. **4**(4): p. 619-25.
12. Biggins, S., I. Ivanovska, and M.D. Rose, *Yeast ubiquitin-like genes are involved in duplication of the microtubule organizing center.* The Journal of cell biology, 1996. **133**(6): p. 1331-46.
13. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, 2004. **20**(13): p. 2138-9.
14. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic Acids Res, 2015. **43**(W1): p. W389-94.
15. Kaye, F.J., et al., *A family of ubiquitin-like proteins binds the ATPase domain of Hsp70-like Stch.* FEBS Lett, 2000. **467**(2-3): p. 348-55.
16. Johnson, B.D., et al., *Hop modulates Hsp70/Hsp90 interactions in protein folding.* J Biol Chem, 1998. **273**(6): p. 3679-86.
17. Elsasser, S. and D. Finley, *Delivery of ubiquitinated substrates to protein-unfolding machines.* Nature cell biology, 2005. **7**(8): p. 742-9.
18. Wu, A.L., et al., *Ubiquitin-related proteins regulate interaction of vimentin intermediate filaments with the plasma membrane.* Molecular cell, 1999. **4**(4): p. 619-25.

19. Conklin, D., et al., *Molecular cloning, chromosome mapping and characterization of UBQLN3 a testis-specific gene that contains an ubiquitin-like domain.* Gene, 2000. **249**(1-2): p. 91-8.
20. Davidson, J.D., et al., *Identification and characterization of an ataxin-1-interacting protein: A1Up, a ubiquitin-like nuclear protein.* Human molecular genetics, 2000. **9**(15): p. 2305-12.
21. Wu, S., et al., *Characterization of ubiquilin 1, an mTOR-interacting protein.* Biochim Biophys Acta, 2002. **1542**(1-3): p. 41-56.
22. N'Diaye, E.N. and E.J. Brown, *The ubiquitin-related protein PLIC-1 regulates heterotrimeric G protein function through association with Gbetagamma.* The Journal of cell biology, 2003. **163**(5): p. 1157-65.
23. Bedford, F.K., et al., *GABA(A) receptor cell surface number and subunit stability are regulated by the ubiquitin-like protein Plic-1.* Nat Neurosci, 2001. **4**(9): p. 908-16.
24. N'Diaye, E.N., et al., *The ubiquitin-like protein PLIC-2 is a negative regulator of G protein-coupled receptor endocytosis.* Molecular biology of the cell, 2008. **19**(3): p. 1252-60.
25. Ciechanover, A. and A.L. Schwartz, *The ubiquitin-proteasome pathway: the complexity and myriad functions of proteins death.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(6): p. 2727-30.
26. Walters, K.J., et al., *Structural studies of the interaction between ubiquitin family proteins and proteasome subunit S5a.* Biochemistry, 2002. **41**(6): p. 1767-77.
27. Grabbe, C. and I. Dikic, *Functional roles of ubiquitin-like domain (ULD) and ubiquitin-binding domain (UBD) containing proteins.* Chem Rev, 2009. **109**(4): p. 1481-94.
28. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.
29. Vranken, W., F., et al., *The CCPN data model for NMR Spectroscopy.* Proteins, 2005. **59**(59): p. 687-696.
30. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2013. **73**: p. 1-16.
31. Fielding, L., *NMR methods for the determination of protein-ligand dissociation constants.* Curr Top Med Chem, 2003. **3**(1): p. 39-53.
32. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature protocols, 2015. **10**(6): p. 845-58.
33. de Vries, S.J., M. van Dijk, and A.M. Bonvin, *The HADDOCK web server for data-driven biomolecular docking.* Nature protocols, 2010. **5**(5): p. 883-97.
34. Lowe, E.D., et al., *Structures of the Dsk2 UBL and UBA domains and their complex.* Acta Crystallogr D Biol Crystallogr, 2006. **62**(Pt 2): p. 177-88.
35. Zhang, D., S. Raasi, and D. Fushman, *Affinity makes the difference: nonselective interaction of the UBA domain of Ubiquilin-1 with monomeric ubiquitin and polyubiquitin chains.* J Mol Biol, 2008. **377**(1): p. 162-80.
36. Zhang, D., S. Raasi, and D. Fushman, *Affinity makes the difference: nonselective interaction of the UBA domain of Ubiquilin-1 with monomeric ubiquitin and polyubiquitin chains.* Journal of molecular biology, 2008. **377**(1): p. 162-80.

37. Chen, X., et al., *Structures of Rpn1 T1:Rad23 and hRpn13:hPLIC2 Reveal Distinct Binding Mechanisms between Substrate Receptors and Shuttle Factors of the Proteasome.* Structure, 2016. **24**(8): p. 1257-70.
38. Puthenveetil, R. and O. Vinogradova, *Optimization of the design and preparation of nanoscale phospholipid bilayers for its application to solution NMR.* Proteins, 2013. **81**: p. 1222-1231.
39. Puthenveetil, R., K. Nguyen, and O. Vinogradova, *Nanodiscs and solution NMR: preparation, application and challenges*, in *Nanotechnology Reviews*. 2016.
40. Hjerpe, R., et al., *UBQLN2 Mediates Autophagy-Independent Protein Aggregate Clearance by the Proteasome.* Cell, 2016. **166**(4): p. 935-49.
41. Ford, D.L. and M.J. Monteiro, *Dimerization of ubiquilin is dependent upon the central region of the protein: evidence that the monomer, but not the dimer, is involved in binding presenilins.* Biochem J, 2006. **399**(3): p. 397-404.
42. Barazi, H.O., et al., *Regulation of integrin function by CD47 ligands. Differential effects on alpha vbeta 3 and alpha 4beta1 integrin-mediated adhesion.* The Journal of biological chemistry, 2002. **277**(45): p. 42859-66.
43. Su, V. and A.F. Lau, *Ubiquitin-like and ubiquitin-associated domain proteins: significance in proteasomal degradation.* Cell Mol Life Sci, 2009. **66**(17): p. 2819-33.

Chapter 3 – Investigating the Role of the Juxtamembrane Region of BTN3A1 in Ligand Induced Activation

Adapted from: **Nguyen, K.**, Li, J., Puthenveetil, R., Lin, X., Poe, M., Hsiao, C.C., Vinogradova, O., and Wiemer, A. *The butyrophilin 3A1 intracellular domain undergoes a conformational change involving the juxtamembrane region*, FASEB J., just accepted.

3.1 – Introduction

3.1.1 – Butyrophilins, Phosphoantigens, and the Immune Response

The butyrophilins (BTN) are type I receptor glycoproteins with further classification as BTN1 (A1), BTN2 (A1, A2, A3), and BTN3 (A1, A2, A3) subfamilies [1]. They possess diverse functionalities as BTN1 is involved in the milk secretion during lactation whereas BTN2 and BTN3 were found to be able to activate/inhibit the immune system [1, 2]. In particularly, BTN3A1 has been the focus of attention due to recent studies have identifying it to be essential for the activation of Vγ9Vδ2 T cells [3-5]. These T cells, upon activation, respond by quickly lysing the target cells by generating pro-inflammatory cytokines [6, 7]. Given the clinical potential to use this mechanism as a means for the removal of cancerous entities, it is imperative to understand the activation mechanism of BTN3A1 and how it eventually translates to the activation of Vγ9Vδ2 T cells.

BTN3A1 is a ~54kDa protein, after removal of the signal peptide, that contains two extracellular Ig domains (IgC and IgV), a single transmembrane helix, an intracellular B30.2 domain, and a stretch of 68 amino acids, denoted as the juxtamembrane (JM) region, between the transmembrane helix and B30.2 domain (Figure 3.1). BTN3A1 is

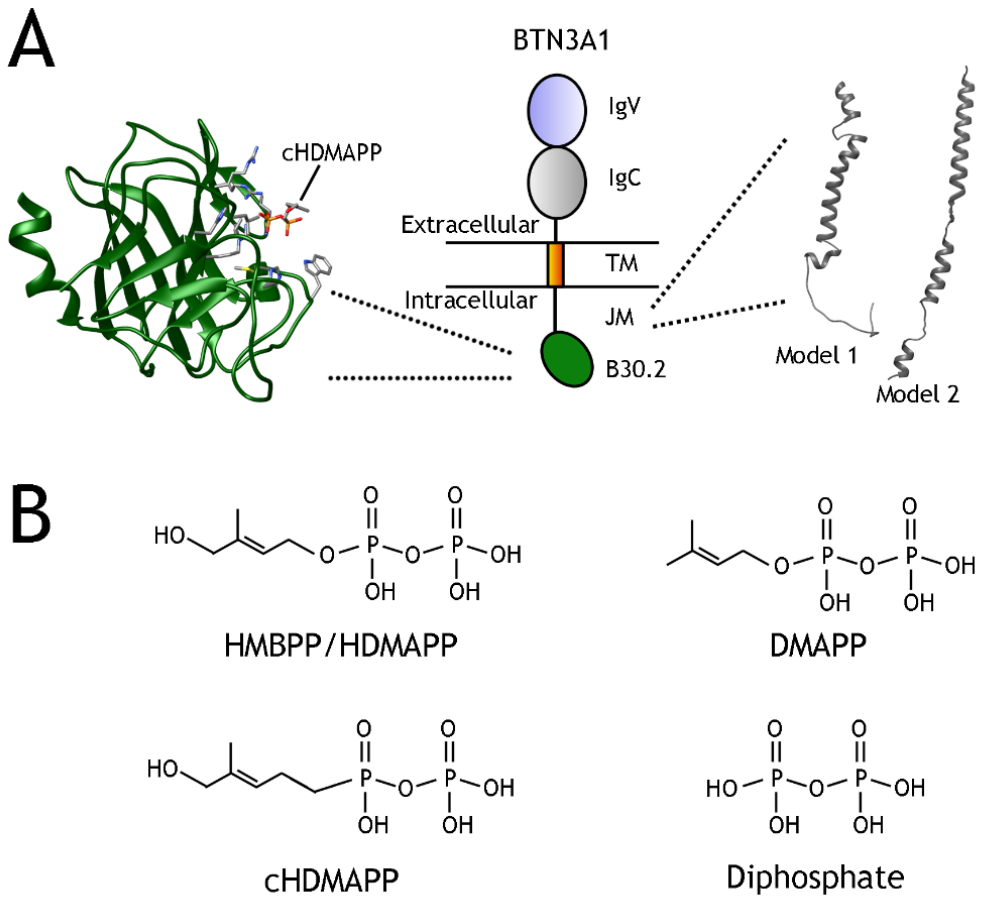

Figure 3.1. Representation of BTN3A1 and Diphosphates. In **(A)**, a simplistic diagram of BTN3A1 is shown to better visualize its domains and positioning of domains, where IgV and IgC domains reside in the extracellular region, a single transmembrane helix, and the intracellular JM region and B30.2 domain. Since the focus is on the intracellular region, the structure of B30.2 bound to cHDMAPP (PDB ID: 47NU) while showing the nearby positively charged residues. The JM region is not structurally characterized, but models (calculated through I-TASSER [24] (model 1) or MODELLER (model 2) [23]) show that a potential for a highly helical features. **(B)** shows the ligands used within this chapter as well as cHDMAPP, which was used in the crystal structure [14]. It is important to note that the nomenclature can be inconsistent at times in the literature as HMBPP can also be referred to as HDMAPP, which more clearly shows its relation to cHDMAPP and DMAPP.

activated upon association with phosphoantigens (pAg), who are small molecules that possess a diphosphate moiety. The B30.2 domain possesses a strong affinity for (*E*)-4-hydroxy-3-methyl-but-2-enyl diphosphate (HMBPP), where its K_d is ~2 μM [8]. Interestingly, HMBPP is a natural product that is known to be synthesized in bacteria, but not in humans, through the non-mevalonate pathway as a precursor for the formation of isopentyl disphosphate (IPP) [9], who has a ~71000 fold weaker EC_{50} than HMBPP for the expansion of Vγ9Vδ2 T cells [8]. This highlights the potential for targeting BTN3A1 as its activity depends on exogenous compounds and that not all disphosphate containing compounds readily activate BTN3A1, sparking the interest for synthesizing HMBPP-derviatives and prodrugs [8, 10, 11] or drugs that stimulate pAg accumulation [12, 13].

3.1.2 – The Binding Pocket and BTN3A1 Activation

Given that the drug target and ligand is known, Sandstrom and colleagues provided a crystal structure of the B30.2 domain in its apo (PDB ID:4N7I) and bound (PDB ID: 4N7U) state [14]. The bound state was crystallized through the use of a HMBPP-derivative, hydroxy-methyl-butyl-pyrophosphonate (cHDMAPP), that was chemically crosslinked through the use of glyceraldehyde. The B30.2 domain contains two β-sheets within its core with a positively charged shallow binding pocket for cHDMAPP (Figure 3.1), where mutations that reversed the charge (H351R) abrogates pAg association. Though it appears to be clear that the binding pocket resides on the B30.2 domain of BTN3A1, a comparison of the apo and bound form illustrates the lack of a conformational change upon ligand association. This is a bit troubling as it does not provide additional insight in the possible activation mechanism of BTN3A1.

BTN3A1 activity could be related to its association with protein partners. For instance, periplakin has been found to be relevant in the activation of Vγ9Vδ2 T cells [15], which is predicted to be due to the direct association to the JM region, as well as RhoB, whose association with BTN3A1 is weakened if HMBPP is present [16]. However, there is little detail of the binding interfaces and how exactly HMBPP could influence protein-protein interactions.

To examine this conundrum, a previous report by Hsiao and colleagues had investigated if there was a difference between the BFI (construct contains the JM and B30.2 domain) and the B30.2 domain upon interacting with HMBPP [8]. Though the ITC results reported indicated that the K_d and ΔH values were similar between the two constructs, the NMR spectra had shown remarkable differences in chemical shifts for the BFI/HMBPP titration that were not observed for the B30.2/HMBPP titration. Given that the difference between the two constructs is the presence of the JM region, it can be postulated that the JM region is involved in some way with ligand association. This is an important clue to understanding how BTN3A1 may be activated as the large chemical shift differences could be indicative of a conformational change that encompasses the JM region. Therefore, the work presented here will investigate the role of the JM region in HMBPP-induced activation.

3.2 - Materials and Methods

Protein Purification

Following sequencing, all the recombinant clones were transformed into *E.coli* BL-21 (DE3) cells and grown on an agar plate with antibiotics specific to the vector. A single

colony was transferred to LB broth and grown overnight at 37 °C to be transferred to LB or M9 minimal media. For M9 minimal media, $^{15}NH_4Cl$ and/or ^{13}C-glucose were added as the sole sources of nitrogen and carbon. Cells were induced with 1 mM isopropyl β-D-1-thiogalactopyranoside once they reached an optical density of 0.5 and harvested after 4 hours.

The JM region was purified under denaturing conditions from recombinant expression within inclusion bodies. A base buffer containing 8 M urea, 100 mM Tris, and 200 mM NaCl at pH 7.4 was used. Cells were resuspended in lysis buffer (base buffer + 5 mM imidazole) and lysed by three passages through French Press (Thermo Electron, USA). The lysate was centrifuged at 12,000 rpm. Supernatant containing the denatured protein was mixed with Ni-NTA resin (Qiagen, USA) at room temperature for 2 hours. The resin was washed with 10 mM imidazole and eluted with 300 mM imidazole in base buffer. The eluate was loaded onto a Proto 300 C4 column (Higgins Analytical Inc., USA) and purified in a 10-40% gradient formed by buffer A (90% H_2O, 10% acetonitrile, 0.1% trifluoroacetic acid) and buffer B (90% acetonitrile, 10% H_2O, 0.1% trifluoroacetic acid). Protein fractions were collected and lyophilized for later use. SDS-PAGE analysis was performed to analyze the purity of the purified protein.

Nuclear Magnetic Resonance Spectroscopy

All NMR experiments were performed on a Varian INOVA 600 MHz spectrometer (Agilent, USA) at 25^{o}C. Protein-protein interaction studies were carried out through ^{1}H-^{15}N HSQC experiments collected using 2048*128 increments. The interaction between JM and HMBPP, DMAPP, GDP, and sodium diphosphate was accessed in buffer

containing 100 mM MES, 50 mM NaCl, 5 mM BME, and 8% D_2O at pH 6.5. HMBPP, DMAPP, GDP, and sodium diphosphate was added in a molar excess of 10 times the concentration of the JM region protein.

Spectra for backbone assignments (HNCACB, HNCA, HN(CO)CA, CBCA(CO)NH) were collected using 1024*32*16 increments in the direct, carbon, and nitrogen dimensions, respectively. HNCO experiments were executed with lower increments in the carbon dimension (ni = 16). Experiments were performed in the presence of 5 mM dodecylphosphocholine (Avanti Polar, USA). Peak peaking was performed using CcpNmr software suite [17] and chemical shifts of assigned peaks were used to predict protein secondary structure through the TALOS+ webserver [18].

3.3 – Results and Discussion

3.3.1 – HMBPP Induced Conformational Change

Previously, published work [8] provided evidence that the JM region was affected in some way after the addition of HMBPP when you compare the bound B30.2 and BFI ^{15}N-HSQC spectra. This was a novel discovery due to the fact that it currently is unknown as to how HMBPP binding leads to the activation of BTN3A1. Given that HMBPP affects the JM region as well, it was postulated that ligand binding induces a conformational change within the intracellular region of BTN3A1.

To confirm that this does occur, SAXS (done by Dr. Puthenveetil) was used to visualize the surface envelope of BFI in the presence and absence of HMBPP as shown in Figure 3.2. Both BFI and BFI/HMBPP exhibit globular shapes, indicating that the JM region is likely to be unstructured rather than a rigid α-helix. When we compare the

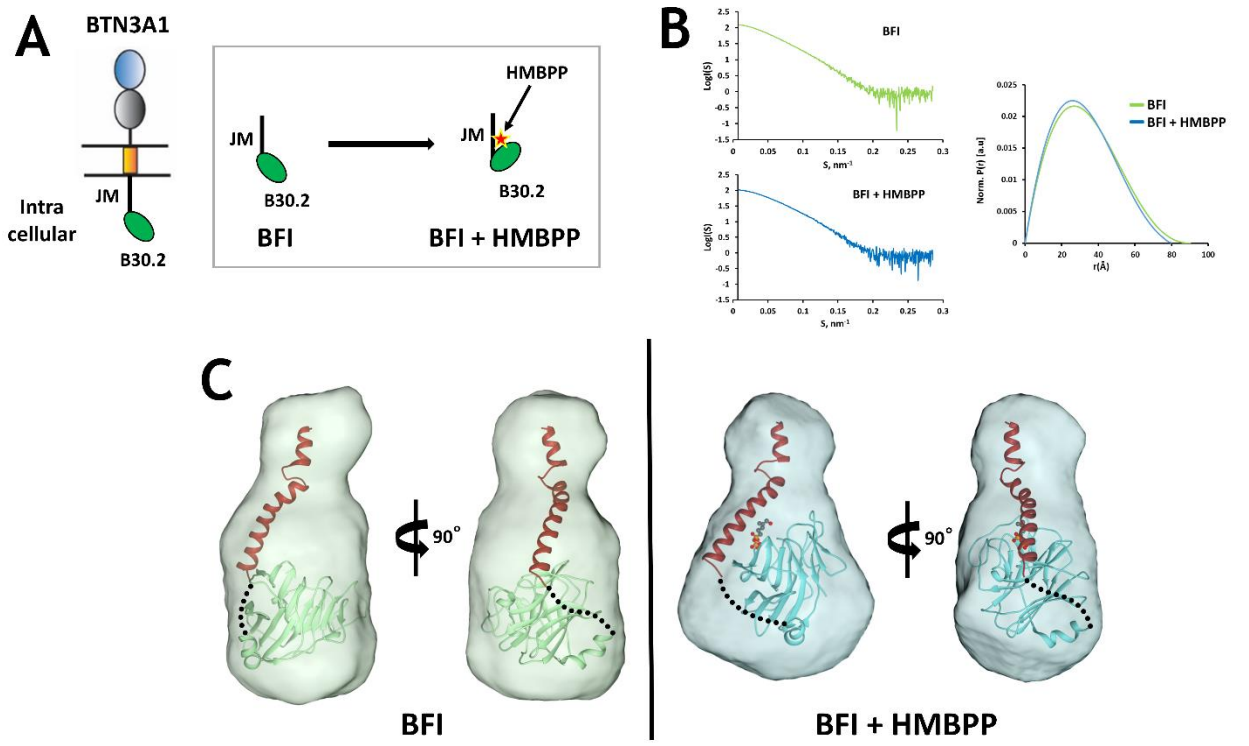

Figure 3.2. The BTN3A1 Intracellular Domain is a Globular Structure that is Compacted upon HMBPP Binding. (A) Schematic representation of the rearrangement of the BFI region (JM+B30.2 domain) in the presence of HMBPP. **(B)** Scattering curve for BFI in the presence (blue) or absence (green) of HMBPP and overlay of the normalized pair-distance distribution curves P(r) of the scattering curves. **(C)** Surface representation of the average ab initio model for BFI and BFI+HMBPP rendered through SITUS [25], where the JM is colored separately in red. The dotted lines represent the missing residues between the JM and B30.2 in our model.

differences between the two surface envelopes, we see that the addition of HMBPP causes a reduction in the radius of gyration (R_g) by ~ 2Å and D_{max} by ~8 Å. This reduction can be characterized as BFI compacting upon ligand association, which provides evidence that a conformational change does occur. Since SAXS only provides a low resolution envelope, it is difficult to define the molecular details of the conformational

change. However, we hypothesized that HMBPP binding to B30.2 leads to a reorientation of the JM/B30.2 as visualized in Figure 3.2A and 3.2C.

This hypothesis was also based on the crystal structure of the apo and bound form of B30.2 obtained by Sandstrom and colleagues [14]. The calculated backbone RMSDs between the apo and complexed B30.2 (4N7I vs. 4N7U) was 0.094Å by matching amide nitrogens using Chimera [19], which is negligible. As a result, ligand binding appears to have minimal or no affect on the overall structure of the B30.2 domain. Thus, we predict that the conformational change is related to the relative orientation of the B30.2 and JM region. However, one caveat is that the crystal structure was crosslinked with the ligand, implicating that it may not properly depict the true bound state of B30.2. It is then not possible to completely rule out that the B30.2 can undergo its own conformational changes in combination with the JM.

3.3.2 – Juxtamembrane and Diphosphates

To properly evaluate the molecular details of this conformational change, we opted to utilize an NMR approach. However, evaluation of the JM region within the BFI construct was not feasible due to serious spectral overlap. To circumvent these difficulties, we have chosen here to investigate the JM region directly and therefore we cloned the isolated BTN3A1 JM region consisting of 68 amino acid residues at position 272-339. Though we planned on studying the interaction between the isolated JM region and B30.2 domain constructs, the pH requirements for solubility of the constructs appeared to be incompatible. This led to a high propensity of the two domains to precipitate upon mixing under any buffer conditions we tested.

Therefore, we opted to test whether the JM region could directly associate with HMBPP as an avenue for facilitating the conformational change. We have found that the majority of JM amides in ^{15}N-HSQC spectrum are sharp and lay within the range of random coil chemical shifts values (8.5 -7.5 ppm), reflecting predominantly unstructured ensemble of conformers in a fast exchange. Several peaks are broader though, indicating an exchange getting to the intermediate range. As the binding is expected to be weak due to the absence of the B30.2 domain, HMBPP was titrated in at a 10 fold excess. Comparison of the JM region alone and the JM region with HMBPP shows that there are spectral differences and that the differences are not drastic, implicating that HMBPP associates weakly to the JM region (Figure 3.3A). The major effects are characterized by differences in line shape, such as peak sharpening, rather than chemical shift perturbations (Figure 3.4), suggesting a very dynamic nature of the ensemble of conformers in exchange. One affected amide, marked by the asterisk in Figure 3.3A and located within the region where serine/ threonine residues are usually found, indicates that HMBPP binding promotes one particular conformation for this amino acid from several potential conformations in intermediate exchange.

Since HMBPP was able to bind to the JM region, we sought to determine the appropriate functional group responsible for interacting. For this, we tested the effect of DMAPP, a molecule whose chemical structure is identical to HMBPP barring the absence of the allylic alcohol (Figure 3.1B). We found that DMAPP interacts with the JM region with a very similar pattern of chemical shift perturbations (Figure 3.3B) as HMBPP, indicating that the allylic alcohol is nonessential for the interaction. Interestingly, the JM region also weakly associated with GDP and sodium diphosphate (Figure 3.3C/D), two

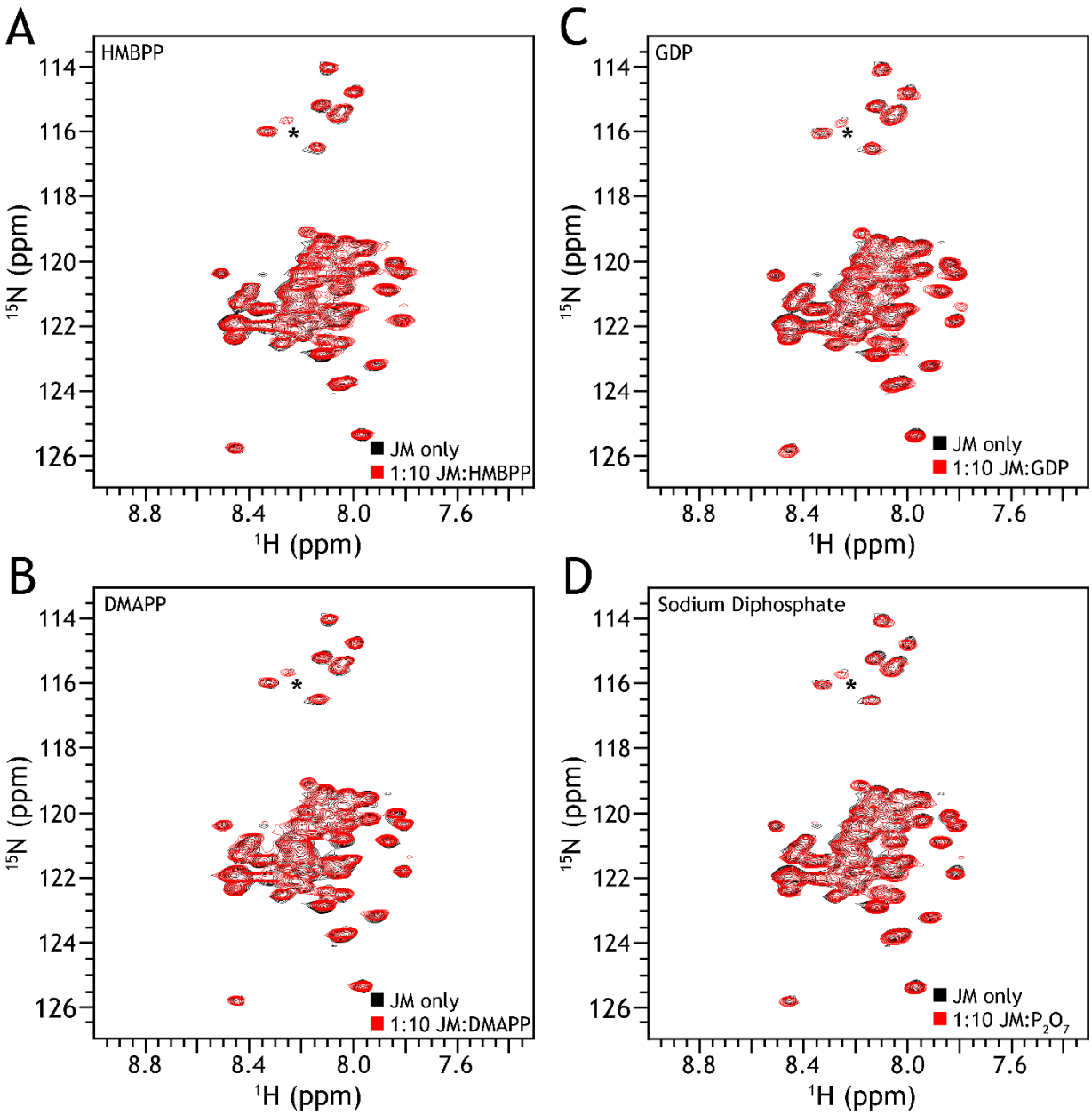

Figure 3.3. The JM Region Interacts with Various Diphosphates. (A) HSQC of JM alone and with HMBPP at a 1:10 ratio. There are minor differences as marked by the asterisks, signifying that HMBPP weakly associates with JM. The overall effect of binding is nearly identical for **(B)** JM:DMAPP, which implies that the hydroxyl of HMBPP is nonessential for binding. ^{15}N-HSQC titrations of JM with GDP **(C)** or sodium diphosphate **(D)** at a 1:10 ratio show similar differences as marked by the asterisks, which shows that GDP and sodium diphosphate are able to interact with the JM region similarly to HMBPP.

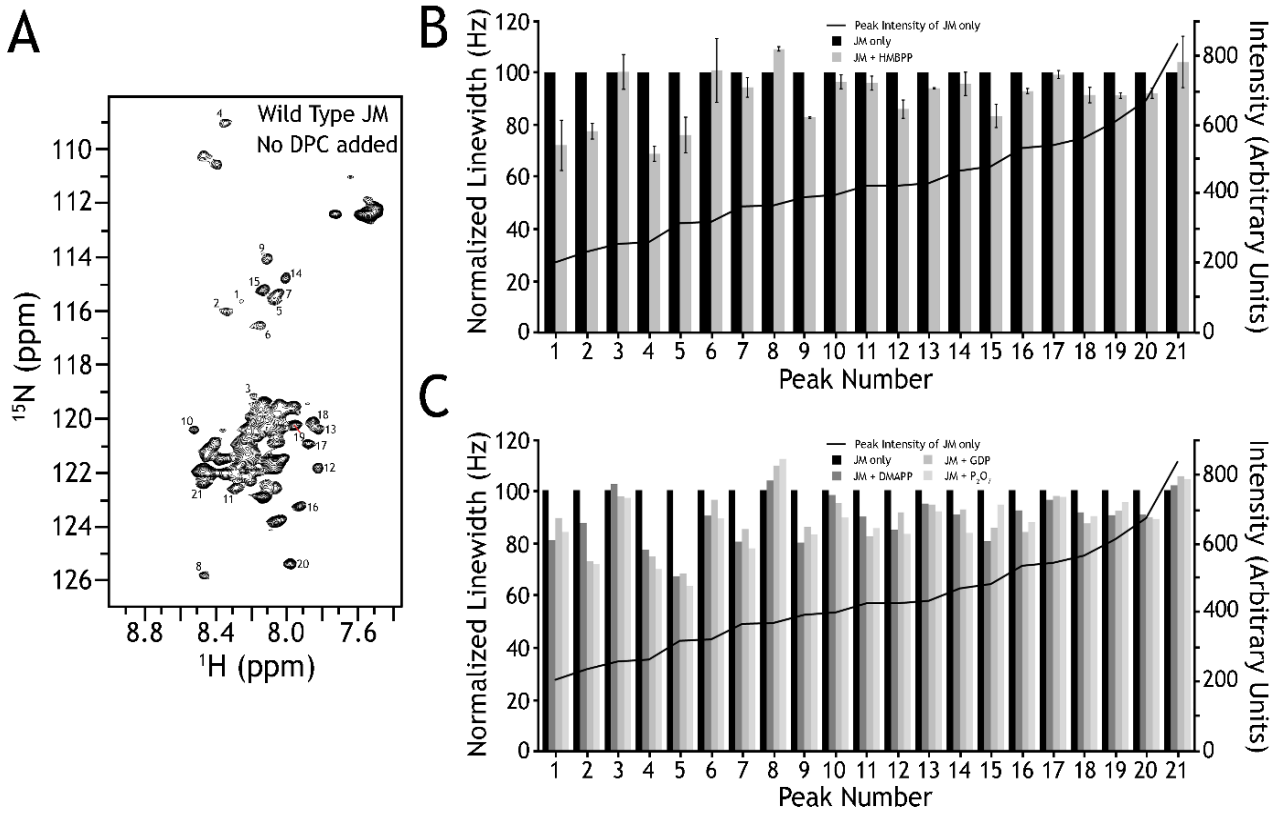

Figure 3.4. Quantification of Peak Lineshape Changes. (A) 15N-HSQC that shows all resonances for JM WT in the absence of DPC. No assignments were done, but peak numbers are shown to better visualize the peaks evaluated in **(B)** and **(C)**. In **(B)** and **(C)**, a compound graph shows normalized linewidth in relation with the peak intensity of the JM alone in the absence of DPC. Peaks with minimal or no overlap were selected and were ordered based on their peak intensity, where peaks 1, 2, and 6 correspond to S303/T304 and peaks 5 and 7 correspond with S296/T297. Quantification was performed by analyzing spectra generated without the use of apodization functions. The line width of the ^{1}H and ^{15}N were averaged ($\nu_{ave} = \sqrt{\frac{\nu_H^2+\nu_N^2}{2}}$) first and then normalized by setting the averaged linewidth of JM only to 100 for individual peaks. **(B)** shows the linewidth difference upon addition of HMBPP and the standard deviation is shown (n=2), whereas **(C)** is with DMAPP, GDP, or diphosphate.

diphosphates that do not activate BTN3A1. Similarly, the effects on peak shape were consistent between all ligands tested (Figure 3.4). Given the common functional groups of these ligands is the negatively charged diphosphate, this suggests that the diphosphate moiety participates in the interaction with the JM region while the allylic alcohol, which is necessary for high potency interactions, likely binds to the B30.2 binding pocket rather than the JM region.

However, the interactions we observed by NMR are weak with minimal spectral changes. To further confirm that the diphosphate is associating with the JM region, Dr. Poe performed competitive binding assays by ITC (Table 3.1) using the BFI and B30.2 constructs as the binding affinity of HMBPP to the isolated JM region is too weak for competitive binding. The addition of 100x molar excess of sodium diphosphate to BFI was able to reduce the ΔH from -53.7 kJ/mol to -32.9 kJ/mol while becoming more entropically favorable as TΔS increased from -21.6 kJ/mol to -1.1 kJ/mol, suggesting that diphosphate alone is able to competitively bind to BFI. In comparison, the addition of 100x molar excess of sodium diphosphate to B30.2 showed a less pronounced change in the ΔH (from -53.7 kJ/mol to -38.4 kJ/mol) and ΔS values (from -23.0 kJ/mol to -7.7 kJ/mol). Though modest, the effect of the diphosphate is lessened in the absence of the JM region, which indirectly substantiates that the JM is involved in the interaction with the diphosphate.

Table 3.1. Summary of the Thermodynamic Data Derived from ITC Measurements Under Different Conditions.

Sample	ΔH (kJ/mol)	TΔS (kJ/mol)	K_d (μM)	n
BFI	-53.7	-21.6	2.36	0.845
BFI + Diphosphate	-32.9	-1.1	2.61	0.860
B30.2	-53.7	-23.0	4.27	0.994
B30.2 + Diphosphate	-38.4	-7.7	4.10	1.070

Thus, the model that we envisioned was that HMBPP could potentially interact with the JM region as well. This notion is also consistent with the B30.2/C-HMBPP crystal structure because a closer look at the electron density of the bound ligand only shows a well-defined tetrahedral moiety suitable for fitting the β-phosphate functional group. This suggests that the other portion of HMBPP was flexible as it was not visualized by x-ray crystallography [14]. This could be due to non-optimal cross-linked conformation of the apo B30.2 used for crystallization as discussed above, which was not able to promote additional contacts, such as with H381 or other nearby residues [20], including the ones from JM region. Our data suggests that the JM region may coordinate with the second exposed phosphate given that DMAPP, which lacks the allylic alcohol, GDP and even simple diphosphate still show the same weak association with the JM region.

3.3.3 – NMR of the JM Region in the Presence of DPC Micelles

Since the JM was determined to interact with diphosphates and this could potentially key for the activation of BTN3A1, we needed to further characterize the molecular details of this interaction. Given that we have been performing NMR experiments, the logical step would be to perform 3D experiments and assign the backbone amide resonances. However, the JM region is sparingly soluble and has a tendency to aggregate, leading to inadequate quality necessary for 3D experiments. To improve protein solubility and stability, we performed backbone assignments of the JM region in the presence of dodecylphosphocholine (DPC) micelles (Figure 3.5). The quality of the spectrum was greatly improved as witnessed by the increased signal to noise ratio. Unfortunately, the degeneracy of the JM region’s primary sequence due to the large

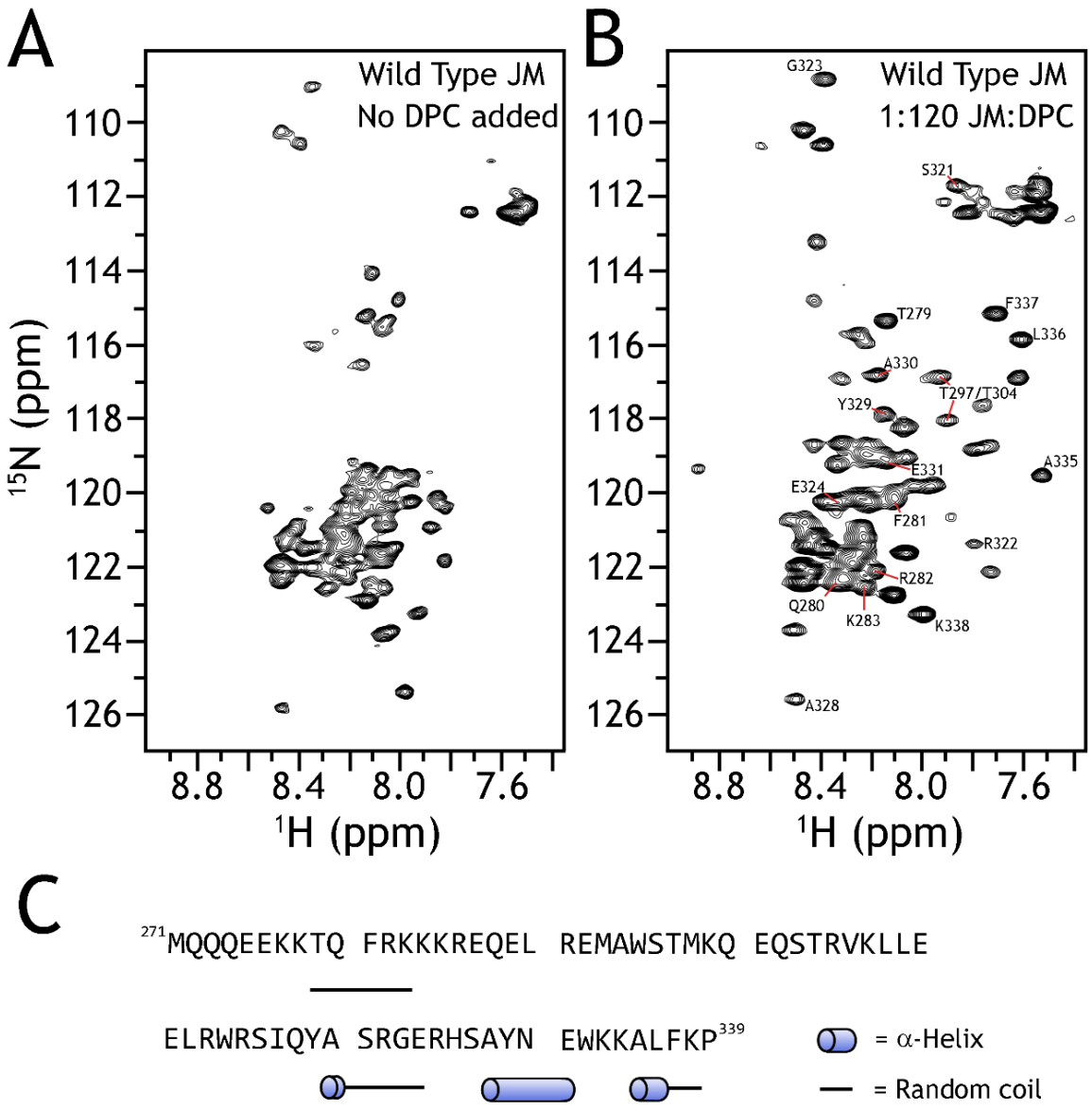

Figure 3.5. DPC Enhances Stability and Solubility to Allow for Partial Assignments. **(A)** 15N-HSQC that shows all resonances for JM WT in the absence of DPC. **(B)** 15N-HSQC of JM Wild Type in the presence of 5mM dodecylphosphocholine. Partial assignments of the protein were done and assignments are shown. For T297 and T304, conclusive assignments were not possible, but two peaks are labeled with a high probability of being a threonine. The secondary structure was predicted from the chemical shifts of the assigned residues using TALOS+ and is shown in **(C)**.

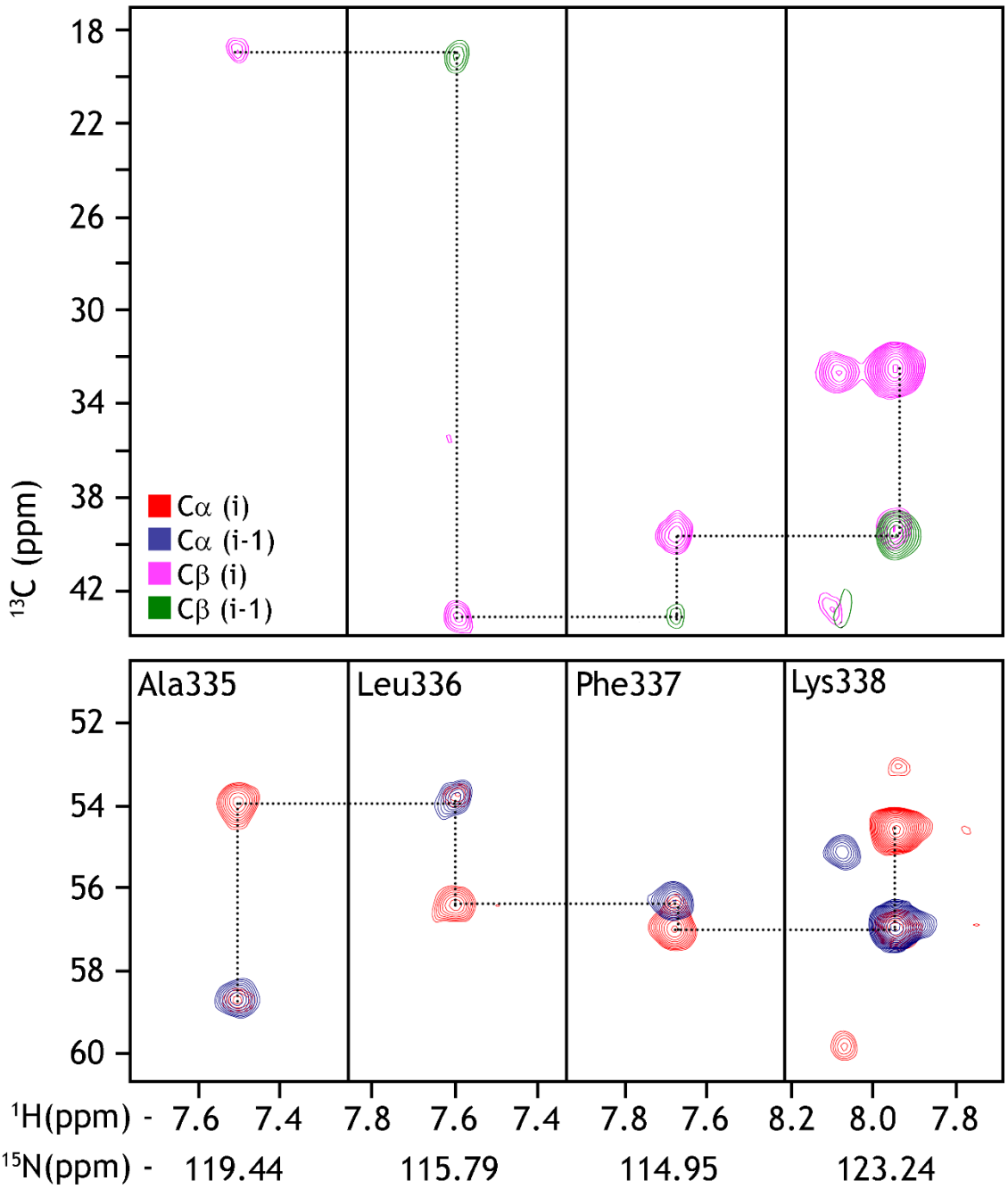

Figure 3.6. Scheme for Sequential Assignments by Matching Resonances. Illustration of the scheme behind backbone assignments, where sequential amino acids can be determined by the alignment of the Cα and Cβ shifts. The experiments provide shifts for the residue itself (i) and the previous residue (i-1). The above shows multiple strips of different ^{15}N chemical shift from HNCACB and HNCA experiments, where alignments are shown by dotted lines.

number of sequentially repeated amino acids (QQQEEKK, RKKKR, etc.) combined with the low solubility and conformational heterogeneity made it difficult to fully complete the resonance assignments. An example of the sequential assignments is shown in Figure 3.6, where the alignment of the Cα and Cβ chemical shifts determined the string of amino acids.

From the partial assignments (as marked in Figure 3.5B), we were still able to predict the secondary structure of the JM region, which shows that the C-terminus is highly helical and the N-terminus begins as a random coil (Figure 3.5C). Though the central region was not adequately assigned, the disparity of the proton shifts falls within a narrow region around ~8.2 ppm, indicating that it is likely that it is predominantly random coil. This is substantiated by our SAXS data that shows the probability of a dimeric coiled-coil structure is minimal, at least in the absence of the transmembrane linker. Also, the repeated HxxHCxC motif or a tryptophan/leucine zipper, which is typical of dimeric coiled coils, is not found in the JM sequence. This is strikingly different than most prediction servers, where it predicts that the JM is highly helical and forms a coiled coil structure [21-24]. However, it is also possible that the JM region could become helical upon ligand binding, especially due to the conformational change relative to the B30.2 domain. However, this could not be experimentally validated due to the buffer incompatibilities of the isolated B30.2 and JM region.

With the partial backbone assignments, we hoped that it would cover residues that were perturbed upon HMBPP association. We have tested whether the JM region interacts with HMBPP in the presence of detergent mimicking membrane-water interface and found that zwitterionic DPC competes with negatively charged diphosphate ligands

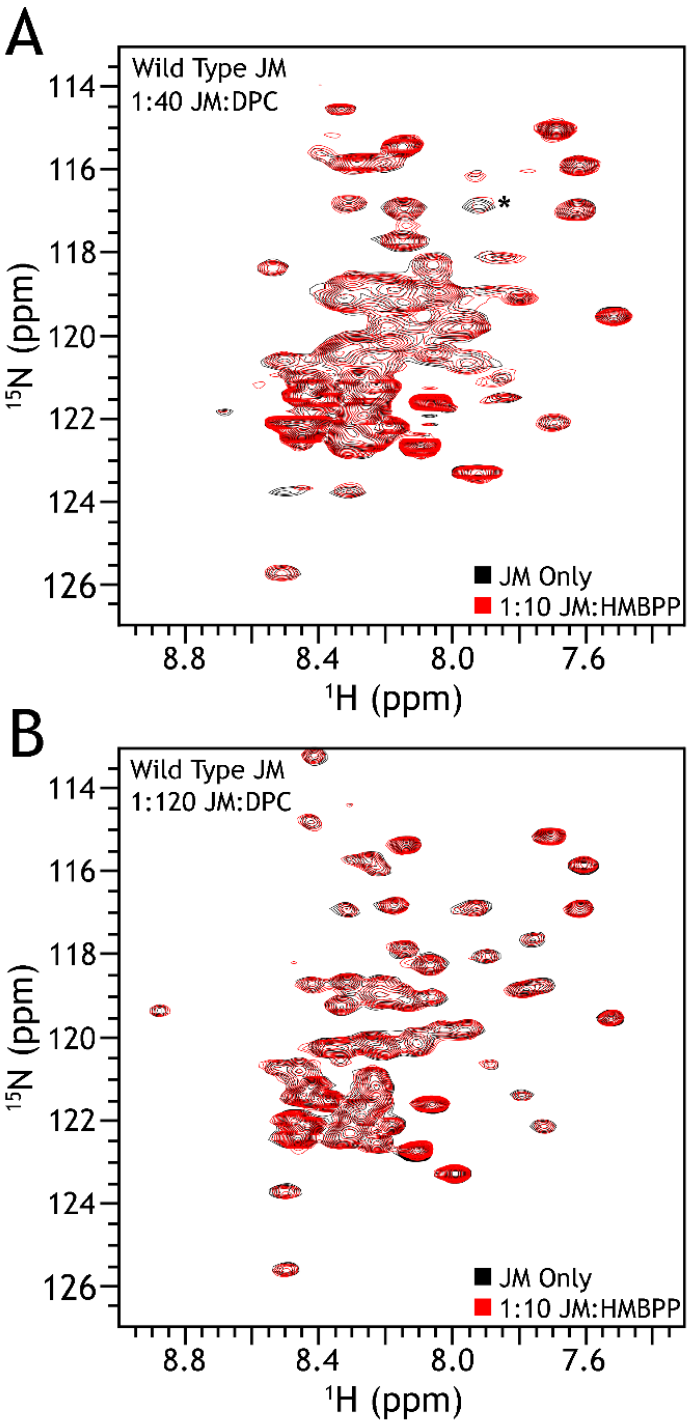

Figure 3.7. HMBPP is Able to Interact with the JM at Low Concentrations of DPC. ^{15}N-HSQC of the JM region in the presence of DPC at a ratio of 1:40 **(A)** and 1:120 **(B)**. HMBPP was titrated in at a 1:10 ratio. Minor line broadening is observed in **(A)** as marked by the asterisk while no significant differences is seen in **(B)** when the DPC ratio is much higher.

for JM binding in a concentration dependent manner. At a low JM:DPC ratio (1:40), several peaks broaden upon addition of HMBPP, indicative of binding, whereas a higher ratio (1:120) showed no significant changes due to the saturation of one peptide molecule per micelle (Figure 3.7). Based on the low JM:DPC ratio, we focused on identifying the residues involved with interacting with HMBPP. Despite the fact that these amino acids could not be conclusively assigned, unique chemical shifts indicated that at least one threonine residue was involved. The JM construct contained only three threonine residues (at positions 279, 297 and 304). Since T279 was assigned, the partial residue assignments narrowed down the HMBPP binding to the JM region to occur at one of these two positions. This also coincides well with the JM data in the absence of DPC, which shows perturbation of peaks within the spectral region that is generally serines/theronines. Thus, we focused our efforts on mutation of the three non-conserved amino acid residues found in these two ST sequences and generated a $ST^{296}AA$ mutation and a $T^{304}A$ mutation.

3.3.4 – Investigation if JM Mutants Associate with HMBPP

To further confirm our conclusions from NMR data hampered by partial assignments, we investigated JM point mutants to uncover if those two threonines had a functional role. As shown in Figure 3.8, both mutations clearly altered the NMR spectrum in the expected region as alanine mutations would greatly affect the chemical shift of the residue mutated and sequential amino acids. By looking for peaks that are greatly shifted, we identified peaks correlating with S296, T297, S303, and T304. Interestingly, the point mutations lead to pronounced chemical shift changes beyond sequentially adjacent

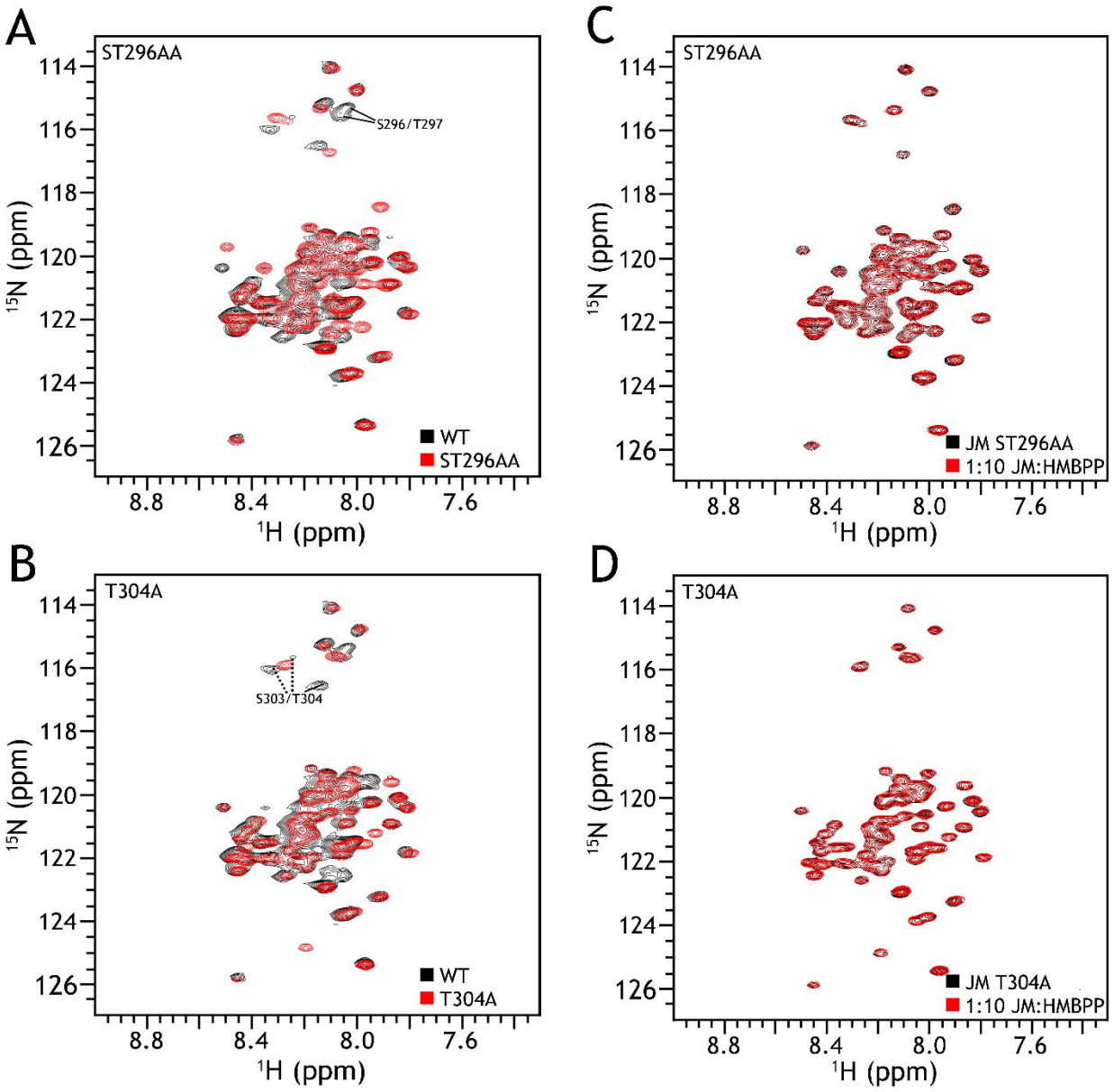

Figure 3.8. Specific Mutations Perturb the JM Threonine Spectra and Inhibit HMBPP Interaction. ^{15}N-HSQC comparison of JM and **(A)** its ST^{296}AA mutant or **(B)** its T^{304}A mutant. Based on the chemical shift changes, peaks corresponding to the mutated serine/threonine residues were identified, where solid lines represent confident assignment and dashed lines are slightly ambiguous. The mutations affect multiple peaks, suggesting that these point mutations affect the local structure. HMBPP was titrated into the mutants at a 10:1 HMBPP:JM ratio, but no interaction was seen for **(C)** ST^{296}AA or **(D)** T^{304}A.

residues, suggesting that there may be some local structure within the JM region as it affects residues through space rather than through bond.

We assessed whether HMBPP was able to interact with the JM mutants and found that there was essentially no difference between the ^{15}N-HSQC of each mutant in the presence and absence of HMBPP. Thus, we concluded that there is no association of HMBPP to the JM mutants, meaning that both threonine residues are important for HMBPP binding to the JM region.

3.3.5 – Correlation of NMR data with *In Vivo E*xperiments

Though the evidence so far shows that JM likely interacts with the diphosphate moiety, the interaction is fairly weak and it requires further investigation to properly confirm its functional significance. For this, we looked at the effect of the threonine mutants *in vivo* on the activity of BTN3A1. Cellular assays using K562 cells and interferon-γ quantification, performed by Jin Li, had shown that the mutants were able to reduce the EC_{50} of HMBPP from 0.056 µM (WT BTN3A1) to 3.1 µM and 3.8 µM for ST296AA and T304A mutants, respectively. The effect was statistically significant and was validated against control experiments (BTN3A1 surface expression levels, BTN3A1 knockout negative control, BTN3A1 knockout + WT transfected control). The activity of the mutants fell between the activity levels of the WT BTN3A1 and BTN3A1 KO, providing evidence that the threonines do serve a functional role *in vivo*.

3.3.6 – A Binding Model

Based on the SAXS, NMR, and ITC, it is clear that a conformational change occurs and that the JM is involved. The relevance of this interaction is further elevated based on the cellular assays, which showed that the mutants $ST^{296}AA$ and $T^{304}A$ showed a statistically significant reduction in BTN3A1 activity, validating that the JM region is essential for Vγ9Vδ2 T activation. Thus, the binding model that we hypothesized based on these findings (presented in Figure 3.2A) shows that the binding of HMBPP to B30.2 is likely to cause a conformational change in the orientation of the B30.2 and the JM region. Once the B30.2 and JM region come in close proximity, the exposed phosphate group could potentially interact with the JM region to form a B30.2/HMBPP/JM clamp. Our NMR data shows that one threonine (T^{304}) is more affected, indicating that this is likely close to the binding pocket. The role of $ST^{296/7}$, on the other hand, could lie in its ability to stabilize some local fold and/or its involvement in forming a binding interface with B30.2.

Nonetheless, the two threonine mutations resulted in the reduced activity of BTN3A1, emphasizing that the induced conformational change was a key step towards the activation of Vγ9Vδ2 T cells. The results point towards two possibilities that could ensue. First, the conformational change within the intracellular domain could transverse to the extracellular domain, leading to its recognition by T cells. This is consistent with the conclusions from a recent report where FRET approach was used to show that bisphosphonate treatment alters the extracellular conformation of BTN3A1 [16]. Second, the conformational change could lead to the recruitment and/or dissociation of

intracellular binding partners, for example, RhoB [16] or periplakin [15], eventually facilitating T cell activation.

3.4 – Conclusion

To conclude, our studies indicate the importance of the BTN3A1 JM region for the interaction with isoprenoid diphosphates and how its activated state is induced by ligand binding. Evidence points to the involvement of threonine residues T297 and T304, which was shown to be relevant in both *in vitro* and *in vivo* studies. Though further work will need to be done to completely understand the molecular details and interface of this interaction, it will be interesting to find out how the structural changes can impact the functions of BTN3A1, especially in regards to the full length BTN3A1. Though we present here the analysis of conformational changes within the intracellular domain, it remains to be seen how the effects propagate across membranes and to the extracellular surface.

3.5 – References

1. Rhodes, D.A., W. Reith, and J. Trowsdale, *Regulation of Immunity by Butyrophilins.* Annu Rev Immunol, 2016. **34**: p. 151-72.
2. Afrache, H., et al., *The butyrophilin (BTN) gene family: from milk fat to the regulation of the immune response.* Immunogenetics, 2012. **64**(11): p. 781-94.
3. Harly, C., et al., *Key implication of CD277/butyrophilin-3 (BTN3A) in cellular stress sensing by a major human gammadelta T-cell subset.* Blood, 2012. **120**(11): p. 2269-79.
4. Palakodeti, A., et al., *The molecular basis for modulation of human Vgamma9Vdelta2 T cell responses by CD277/butyrophilin-3 (BTN3A)-specific antibodies.* J Biol Chem, 2012. **287**(39): p. 32780-90.
5. Riano, F., et al., *Vgamma9Vdelta2 TCR-activation by phosphorylated antigens requires butyrophilin 3 A1 (BTN3A1) and additional genes on human chromosome 6.* Eur J Immunol, 2014. **44**(9): p. 2571-6.
6. Gober, H.J., et al., *Human T cell receptor gammadelta cells recognize endogenous mevalonate metabolites in tumor cells.* J Exp Med, 2003. **197**(2): p. 163-8.
7. Santolaria, T., et al., *Repeated systemic administrations of both aminobisphosphonates and human Vgamma9Vdelta2 T cells efficiently control tumor development in vivo.* J Immunol, 2013. **191**(4): p. 1993-2000.
8. Hsiao, C.H., et al., *Synthesis of a phosphoantigen prodrug that potently activates Vgamma9Vdelta2 T-lymphocytes.* Chem Biol, 2014. **21**(8): p. 945-54.
9. Wiemer, D.F. and A.J. Wiemer, *Opportunities and challenges in development of phosphoantigens as Vgamma9Vdelta2 T cell agonists.* Biochem Pharmacol, 2014. **89**(3): p. 301-12.
10. Boedec, A., et al., *Synthesis and biological activity of phosphonate analogues and geometric isomers of the highly potent phosphoantigen (E)-1-hydroxy-2-methylbut-2-enyl 4-diphosphate.* J Med Chem, 2008. **51**(6): p. 1747-54.
11. Espinosa, E., et al., *Chemical synthesis and biological activity of bromohydrin pyrophosphate, a potent stimulator of human gamma delta T cells.* J Biol Chem, 2001. **276**(21): p. 18337-44.
12. Nada, M.H., et al., *Enhancing adoptive cancer immunotherapy with Vgamma2Vdelta2 T cells through pulse zoledronate stimulation.* J Immunother Cancer, 2017. **5**: p. 9.
13. Wang, H., et al., *Indirect stimulation of human Vgamma2Vdelta2 T cells through alterations in isoprenoid metabolism.* J Immunol, 2011. **187**(10): p. 5099-113.
14. Sandstrom, A., et al., *The intracellular B30.2 domain of butyrophilin 3A1 binds phosphoantigens to mediate activation of human Vgamma9Vdelta2 T cells.* Immunity, 2014. **40**(4): p. 490-500.
15. Rhodes, D.A., et al., *Activation of human gammadelta T cells by cytosolic interactions of BTN3A1 with soluble phosphoantigens and the cytoskeletal adaptor periplakin.* J Immunol, 2015. **194**(5): p. 2390-8.
16. Sebestyen, Z., et al., *RhoB Mediates Phosphoantigen Recognition by Vgamma9Vdelta2 T Cell Receptor.* Cell Rep, 2016. **15**(9): p. 1973-85.
17. Vranken, W.F., et al., *The CCPN data model for NMR spectroscopy: development of a software pipeline.* Proteins, 2005. **59**(4): p. 687-96.
18. Shen, Y., et al., *TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts.* J Biomol NMR, 2009. **44**(4): p. 213-23.
19. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

20. Shippy, R.R., et al., *Phosphinophosphonates and their tris-pivaloyloxymethyl prodrugs reveal a negatively cooperative butyrophilin activation mechanism.* J Med Chem, 2017.
21. Wang, H. and C.T. Morita, *Sensor Function for Butyrophilin 3A1 in Prenyl Pyrophosphate Stimulation of Human Vgamma2Vdelta2 T Cells.* J Immunol, 2015. **195**(10): p. 4583-94.
22. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993. **234**(3): p. 779-815.
23. Webb, B. and A. Sali, *Comparative Protein Structure Modeling Using MODELLER.* Curr Protoc Bioinformatics, 2014. **47**: p. 5 6 1-32.
24. Zhang, Y., *I-TASSER server for protein 3D structure prediction.* BMC Bioinformatics, 2008. **9**: p. 40.
25. Wriggers, W., *Conventions and workflows for using Situs.* Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 4): p. 344-51.

Chapter 4 – NMR as a Semi-Quantitative Tool for the Evaluation of Surface Hydrophobicity

4.1 – Introduction

4.1.1 – Proteins in the Industry

Though proteins are essential components naturally produced and regulated within our body, there has been increasing interesting in the development of biologics for medicinal purposes. When we deal with these proteins in the industrial setting, there is a large focus on the overall stability of the protein, considering that they would be subjected to a number of hostile environments during processing. One of the major setbacks of protein formulations is the stability, which includes the propensity of the protein to unfold and/or aggregate [1]. This scenario has been observed with antibodies [2], which is unfortunate because of the high monetary cost and safety concerns associated with antibodies [3, 4].

The stability of a protein is a difficult parameter to assess and predict, but one key component is its hydrophobicity. Hydrophobicity has been accepted to be vital for the folding of proteins, where the hydrophobic effect leads to the internalization of nonpolar residues [5, 6], as well as protein-protein interactions between hydrophobic patches on the surface of the protein [7-9]. However, surface hydrophobicity is also responsible for nonspecific protein interactions, leading to protein aggregation, protein self-association, and surface adherence [1, 10-12]. Though hydrophobic interactions are not the sole

driving force for protein interactions, understanding this phenomenon would be beneficial in early determination if a protein would be likely to be stable.

4.1.2 – Methods for Quantification of Protein Hydrophobicity

In order to properly understand the role of hydrophobicity in the overall behavior of the protein, it becomes essential to quantify its hydrophobic character. This can be classified to be as the net hydrophobicity, which accounts for all nonpolar side chains present, or the surface hydrophobicity, which only deals with the solvent exposed nonpolar side chains [13, 14]. Though they are both important, surface hydrophobic is a major contributor to physical instabilities, such as protein self-association, aggregation, and adsorption to interfaces [1, 15, 16]. Therefore, focusing specifically on quantification of the surface hydrophobicity would be ideal for which different experimental techniques have been used.

Hydrophobic Interaction Chromatography (HIC) is a purification method that has relatively mild solution conditions compared to other chromatographic techniques such as Reverse Phase Chromatography [17-19]. In HIC, the proteins bind to a stationary phase composed of a nonpolar hydrocarbon chain. The mobile phase consists of a high concentration salt, such as ammonium sulfate, in order to promote protein/column association. As the salt concentration is gradually decreased, the proteins will begin to elute off the column in the order of their hydrophobicity, where the least hydrophobic protein would elute first. As a result, it is possible to compare the surface hydrophobicity of proteins.

Unfortunately, there are caveats associated with HIC that should be considered. The strength of the protein's association with the column is highly dependent on the experimental conditions, such as temperature, salt, and ionic strength of the mobile phase, which could considerably change the retention times of the proteins [19-23]. This makes it particularly difficult to compare hydrophobicities of proteins unless the experimental conditions are identical. The protein is also subjected to a fairly high ionic strength of salt buffer, which could screen electrostatic interactions, meaning that it may be an untrue representation of the hydrophobic characteristics of the protein as a whole. It is important to keep in mind that HIC solution conditions are not equivalent to formulation and storage buffers.

Another method is the use of extrinsic fluorescence probes, such as ANS, Bis ANS, and Prodan. This method works because these probes have differing properties in an aqueous or hydrophobic environment. In the aqueous environment, the probes have a low quantum yield. When the probes are in a hydrophobic environment, the quantum yield is much higher and the wavelength of maximum fluorescence shifts [24]. Therefore, it is possible to deduce when the probes bind to hydrophobic patches on the protein by monitoring the change in fluorescence. Although this method is simple, fast, and nondestructive, there are some concerns in regards to its reliability. The overall structure of the extrinsic dyes is quite diverse, where they can vary in size as well as aromatic, aliphatic, and charged functionalities. As a result, the mode of binding can be different between dyes, ultimately leading to differences in hydrophobicity values [25]. This is particularly important when we consider that charge-charge interactions may occur in some cases like the charged dye ANS, which could enhance or reduce its overall binding

affinity to the protein, influencing its estimation of hydrophobicity [17, 26]. As a result, the quantification of surface hydrophobicity may not be entirely accurate.

To conclude, both traditional techniques, fluorescence spectroscopy and hydrophobic interaction chromatography, have a number of experimental problems limiting their ability to provide an accurate value of hydrophobicity. Even then, the techniques only determine relative hydrophobicities, which do not provide any significant information of the protein's hydrophobic character unless compared to another protein. Therefore, this stresses the importance of establishing a multi-method quantitative measurement protocol to define surface hydrophobicity.

The alternative method that we will explore is the use of NMR to determine the surface hydrophobicity of proteins. NMR is a sensitive and robust technique that can be used to study the binding between a small molecule (probe) and a larger macromolecule (protein) by observing the transverse relaxation time (T2) [27]. Though the concept is not novel considering that drug screening by NMR has been a common approach for drug design, where fragment based drug design focuses on finding small molecule binders in specific locations on the protein [28], the investigation proposed here has a novel application by studying the interactions of small hydrophobic probes to proteins. The idea is to use an excess amount of small molecular probe, which would allow it to nonspecifically bind to many different locations of the protein. By selectively using hydrophobic probes, we can then deduce the overall surface hydrophobicity of the protein by analyzing the extent of binding. The benefit of using NMR is that we would be able to conduct experiments in buffers matching formulation and storage conditions (unlike HIC) while being unrestricted to the use of only aromatic probes (unlike fluorescence).

In this study, the sensitivity of NMR to measure surface hydrophobicity was explored by comparing the extent of binding between various probes and protein standards. The probes utilized were chosen based on the aliphatic or aromatic features. The results determined by NMR would be corroborated with HIC and fluorescence to further substantiate our findings.

4.2 – Materials and Methods

Materials

Bovine Serum Albumin (BSA), α-Chymotrypsinogen A from Bovine Pancreas (A-ChytA), and β-Lactoglobulin A (B-LgA) from Bovine Milk, N-Acetyl-L-Leucine Methyl Ester and N-Acetyl-L-Phenylalanine Ethyl Ester were purchased from Sigma (St. Louis, Mo). Acetyl-Valine-Methyl Ester was purchased from Bachem Americas Inc., N-Acetyl-L-tryptophan Ethyl Ester was purchased from TCI Chemicals and N-Acetyl-L-Tyrosine Ethyl Ester was purchased from MP Biomedicals, LLC. ANS (8-anilino-1-naphthalenesulfonic acid) was purchased from Molecular Probes. The Hiscreen Butyl HP column and Hiscreen Phenyl HP column was purchased from GE Healthcare. Solutions were prepared in a sonicated 15 mM ionic strength (8.5 buffer strength) phosphate buffer (pH 7.0). All buffers and protein stock solutions were filtered through 0.22 μm filters.

Sample Preparation

All samples were prepared in the same buffer (8.5 mM sodium phosphate, 15mM ionic strength, pH 7.0, and ~90% D_2O) using nitrogen flushed D_2O. The probe concentration was held constant at 3 mM, where molar ratios of 1:20, 1:50, 1:100, 1:150,

1:200, 1:400, and 1:1000 correspond to 150 µM, 60 µM, 30 µM, 20 µM, 15 µM, 7.5 µM, and 3 µM of protein (BSA, A-ChytA, or B-LgA), respectively. A 1:50 ratio was used (1.5 mM probe and 30 µM protein) for comparative studies between probes. The probes were also investigated alone at the concentrations of 1.5 mM, 3 mM, or 6 mM. None of the samples contained any reference compounds due to initial observations that internal references can also bind to the protein targets, which skew the measurements, while external references significantly complicated the acquisition process due to the problems associated with shimming of two different compartments simultaneously.

The probes selected to be tested belong to two major subclasses of compounds: aliphatics and aromatics. The aliphatics consisted of *tert*-butyl alcohol, 1-propanol, and 1-butanol whereas only phenol was used as an aromatic. In addition, capped amino acids were selected to mimic protein-protein interactions. Similarly, the capped amino acids were chosen based on their aliphatic and aromatic side chains, with the aliphatics being N-acetyl-L-leucine methyl ester, and N-acetyl-L-valine methyl ester, and the aromatics were N-acetyl-L-phenylalanine ethyl ester, N-acetyl-L-tryptophan ethyl ester, and N-acetyl-L-tyrosine ethyl ester.

Measuring Transverse Relaxation Time

The binding of the probe to the protein was determined by measuring the transverse relaxation time (T2) of the probe. Samples were prepared in 535-PP-7 NMR tubes purchased from Wilmad Labglass (Vineland, NJ) and experiments were performed at 25 °C. The water signal was suppressed by presaturation at power level of 6 dB for 3 seconds. The T2 was acquired from array experiments performed by using the Carr-

Purcell-Meiboom-Gill (CPMG) T2 pulse sequence without temperature compensation [29]. The acquisition delay was set to 25 seconds, no sample heating was observed under these conditions as judged by the lack of temperature-dependent perturbations in chemical shifts. No alterations in peaks shapes due to J-coupling [30] was noticed either. The bigtau parameter, which is the time between the initial pulse and data acquisition, was non-uniformly distributed with larger bias to the smaller time values to improve exponential fit. The bigtau sets were not identical between experiments due to the variability of the relaxation of the probes (see Table 4.1 for bigtau sets).

The integral for the probe peak of interest was taken in each spectrum of the array. Errors originating from overlapping peaks were minimized by base-line correction, which involves subtracting a spectrum of the protein alone from the spectrum of the probe with additional correction factors to account for concentration differences. The intensities for the peak of interest were also taken and used to compare exponential fitting parameters (see supporting information). The fit was carried out using the VnmrJ v3.2 T2 analysis module. The equation used for the fit was:

$$y = m_0 * e^{\left(\frac{-x}{T_2}\right)} + m_1 \qquad (1)$$

m_0 is a constant that corrects for the scaling factor of the integral and m_1 is a constant that corrects for baseline issues. The T2 measurements were done in triplicates on a Varian 600 MHz spectrometer equipped with a triple resonance cryogenic probe. The average T2 was calculated for further analysis.

Table 4.1. Big Tau Sets for Determining Transverse Relaxation Time

Probe	Protein	BigTau Set (s)
Tert-butyl alcohol	None B-LgA A-ChytA	0.1,0.2,0.4,0.6,0.8,1.2,2,3,4,6,8,12,20
Tert-butyl alcohol	BSA	0.1,0.2,0.4,0.8,1.2,2,3.5,6,8
Butanol	None BSA B-LgA A-ChytA	0.2,0.3,0.5,0.7,1.2,2,3.5,5.5,8,12
butanol	BSA	0.1,0.2,0.3,0.5,0.7,1.2,2,3,4.5,6
propanol	None B-LgA A-ChytA	0.2,0.4,0.6,0.8,1.2,3,6,9,12,20
propanol	BSA	0.1,0.2,0.3,0.5,0.7,1.2,2,3,4.5,6
phenol	None B-LgA A-ChytA	0.2,0.5,0.8,1.2,2,4,6,12,24
phenol	BSA	0.05,0.1,0.2,0.3,0.5,0.7,1.2,2,5
Leu	None B-LgA A-ChytA	0.1,0.2,0.4,0.8,1.2,2,4,8,16
Leu	BSA	0.05,0.1,0.2,0.3,0.5,0.7,1.2,2,5
Val	None BSA B-LgA A-ChytA	0.1,0.2,0.4,0.8,1.2,2,4,8,16
Phe	None B-LgA A-ChytA	0.1,0.2,0.4,0.8,1.2,2,4,8,16
Phe	BSA	0.05,0.1,0.2,0.3,0.5,0.7,1.2,2,5
Trp	None B-LgA A-ChytA	0.1,0.2,0.4,0.8,1.2,2,4,8,16
Trp	BSA	0.05,0.1,0.2,0.3,0.5,0.7,1.2,2,5
Tyr	None B-LgA A-ChytA	0.1,0.2,0.4,0.8,1.2,2,4,8,16
Tyr	BSA	0.05,0.1,0.2,0.3,0.5,0.7,1.2,2,5

The minimum number of points used was 9 and experiments performed at different concentrations (if applicable) all used the same bigtau set. The above table indicates the combination of protein and probe and their respective bigtau set.

Measuring the Diffusion Coefficient by DOSY

Preparation of the sample was identical to methods mentioned previously. The sample was subjected to a pulse sequence of diffusion correlation spectroscopy (DOSY) called DgcsteSL (DOSY Gradient Stimulated Echo with Spin Lock). For the experimental parameters, the diffusion delay (del) is 60 ms, total diffusion-encoding pulse width (gt1) is 2 ms, gradient stabilization delay (gstab) is 0.2 ms, relaxation delay is 10 s, and number of increments was 15 (values of the gradient level are randomly selected between 1300 and 32500 by the software). The integral area of the signal of interest was used to calculate the diffusion coefficient through the VnmrJ v3.2 DOSY module.

4.3 – Results and Discussion

4.3.1 – Experimental Considerations

Though NMR is a useful technique for analyzing both small ligands and proteins, utilizing NMR as a quantitative tool requires extra care. Proper calibration should be and needs to be done (ex - 90^{o} pulse width) on each individual sample due to sample variability. The next consideration is the presence of H_2O as the experiments are performed in an aqueous environment. Ideally, the percentage of D_2O should be close to 100%, but would be difficult to achieve as it would require the use of deuterated buffering agents as well as proteins as the amide proton is labile. The buffer itself is made by using D_2O as the solvent, which is how we estimated that the amount of D_2O is approximately 90%.

The reason as to why H_2O is important to consider is because its concentration far exceeds the concentration of the probe/proteins in solution. Given that the experiment is

a simple ^{1}H-CPMG T2, we would observe a signal for water. It would be the dominant peak by a far margin and would make it unsuitable to observe probe peaks close to 4.7 ppm due to spectral overlap. The suppression of the water signal needs to be done to obtain a proper baseline, which is achieved by a presaturation pulse prior to the excitation of the protons. The power of the presaturation pulse has to be as minimal as possible so that it would not affect the peak of interest.

The confidence in the T2 values obtained can be improved by the use of a reference compound, which would indicate if the parameters of the experiment were adequately calibrated as the reference T2 and chemical shift should be the same between runs. The reference compound can be added into the solution itself (internal reference) or in a secondary compartment separate from the solution (external reference). Though an internal reference would be straightforward as it simply requires the addition of equal volumes of reference compounds into every sample, a problem arises when the reference compound is not inert. That is, the reference compound is able to interact with the solution components like the protein. This is observed in our initial tests in which we tested the binding of *tert*-butyl alcohol and BSA at two different ratios (Table 4.2). As expected, the line width of *tert*-butyl was broadened, which is indicative of binding. However, we see that our internal reference, 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS), also showed line broadening. This makes it difficult to properly assess the extent of binding of *tert*-butyl alcohol if there are multiple ligands that may or may not bind to the same interface. Though it is possible to utilize internal references, it requires prescreens to determine if the compound is truly inert, which can become cumbersome if you are screening against a large number of proteins.

Table 4.2. Line Width Variability with an Internal Reference

	Line width of *Tert*-butyl alcohol (Hz)	Line width of DSS (Hz)
No protein	2.96	2.99
1:100 BSA:*tert*-butyl	3.64	5.54
1:50 BSA:*tert*-butyl	4.41	6.43

The line width of *tert*-butyl alcohol peak (at 1.23 ppm) and DSS (at 0.0 ppm) were calculated at half height in the presence and absence of BSA. The concentration of DSS and BSA were held constant at 2 mM and 29 µM, respectively. The concentration of *tert*-butyl alcohol was 3 mM for the no protein and 1:100 sample while being reduced to 1.5 mM for the 1:50 sample.

The alternative solution is to use an external reference. The benefits are easily visible as the external reference has no capability of interacting with the protein and can be removed and reused for a different sample, which means that the external reference is identical between runs/samples. The downside is that the sample contains two different compartments of two closely matched, but not identical solution composition. As such, it becomes troublesome to create a proper homogenized magnetic field, leading to potentially distorted line shapes. This makes the process for collecting the data much more tedious especially if the external reference is improperly aligned in the center. As a result, the T2 experiments performed here were done in the absence of a reference compound. However, the use of an external references have been used for quantitative purposes as reported in the literature [29-32] and has given reasonable results for DOSY using a Wilmad coaxial insert.

Lastly, the goal of this study is to investigate the extent of hydrophobic interactions of small probes with a protein standard. Unfortunately, the experiments have to be done in aqueous solutions so that the proteins are properly folded and behaving normally. This means that the use of entirely hydrophobic compounds, such as benzene or butane, is not feasible due to their low solubility. Thus, it is necessary that our compounds contain a hydroxyl group for increased solubility (ex – phenol and butanol). As a result, there is a caveat that the interaction may not be completely hydrophobic.

4.3.2 – Line width versus Transverse Relaxation Time

In order to determine if a small molecular weight probe interacts with a larger macromolecule target, such as protein or DNA, it is possible to track changes in the probe relaxation rate through nuclear magnetic resonance (NMR) spectroscopy [33-35]. Large molecules will have longer correlation times and relax faster in solution, which is important because binding of the probe to the protein will result in a faster relaxation rate for the probe [36]. The parameter, which correlates with relaxation rates of molecules and could be directly derived from any NMR spectrum, is the line-width of the molecules' corresponding peaks. The true line-width and T2, the transverse relaxation time, are related by the equation:

$$\vartheta_{1/2} = \frac{1}{\pi T_2} \qquad (2)$$

thus, indicating the plausibility to use either line-width or T2 to monitor the binding of the probe to the protein. However, in our initial attempt to develop a general method for measuring weak non-specific probe interactions with hydrophobic surface areas of

proteins, we have found out that line-width was not an optimal parameter to use. In practice, the measured line-width is affected by artifacts as shown by the equation:

$$\vartheta_{actual} = \vartheta_{1/2} + \vartheta_{non-hom} \qquad (3)$$

The large errors in line-width measurements are predominantly associated with magnetic field non-homogeneity and are caused by difficulties achieving exactly the same shims between samples, which would overshadow the small difference in line-width that would arise from weak binding. From this perspective, measuring T2 is a better choice as it minimizes the problem of magnetic field inhomogeneity. The T2 relaxation was determined by fitting the data points (in form of peak integrals or intensities) in time array to equation 1 as shown by the example in Figure 4.1.

4.3.3 – Intensity versus Integral

There are two possible methods that can be used to observe the exponential decay of the peak of interest. This can be either done by monitoring the intensity of the peak at a certain chemical shift or by taking the integral of a defined region. There are several benefits of using integral over intensity. First, integrals are less error prone because random noise in the defined region will be predominantly canceled out. Second, intensity is more sensitive to a number of factors not particularly related to binding, such as the line shape and possible chemical shift changes due to temperature instability.[30] Though integrals are more favorable to use over intensity, there are problems associated with overlapping peaks, especially in the case of broad protein peaks underneath the sharp probe peak. Integration over that region would combine the areas of both the signal of interest and those unwanted peaks, resulting in difficulties fitting the data and inaccurate

T2 measurements. It is necessary to do protein peaks subtraction before fitting for such cases where significant overlap is visible and it is not possible to choose probe peaks from other chemical shifts regions. An example of using integration to determine the T2 relaxation time is shown in Figure 4.1, where the peak area is taken at different time intervals for the phenol triplet peak at 7.34 ppm.

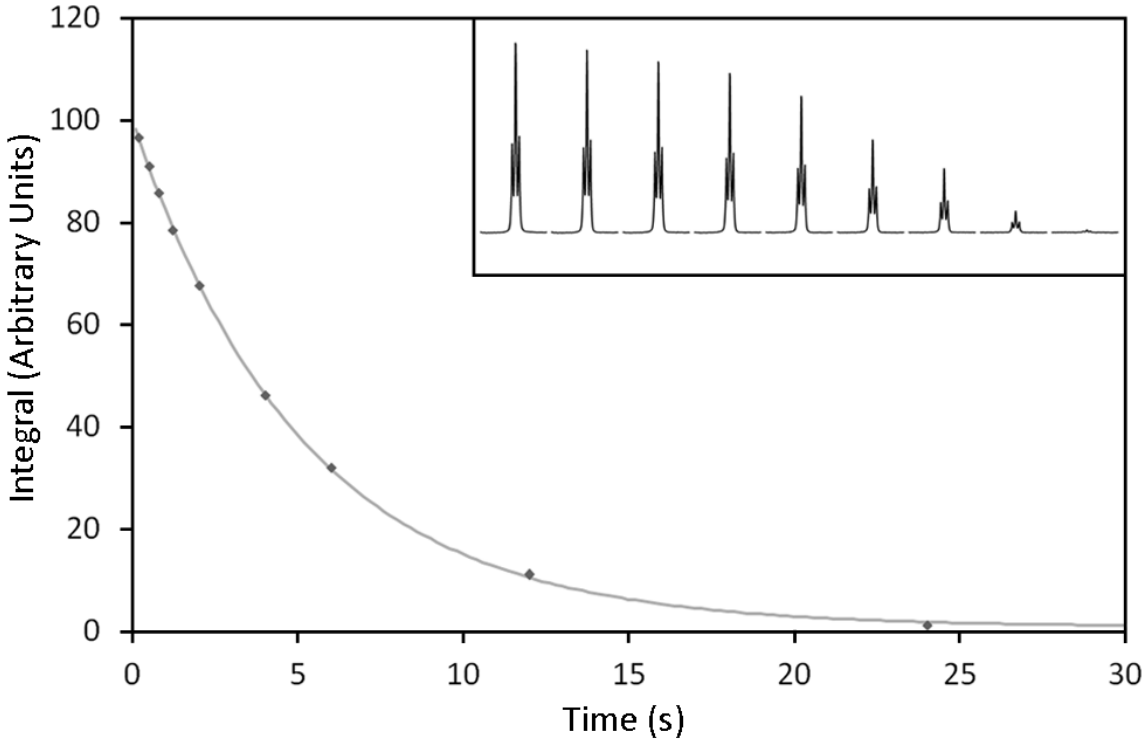

Figure 4.1. Representation of an Exponential Decay and Fit. A sample exponential decay (of 3 mM phenol) obtained from the CMPG-T2 experiment. The exponential is fit with the equation 1. The variables of the fit were m0 = 99.2, m1=0.88, and t2=5.147 s. The inset depicts the peak of interest (triplet at 7.34ppm) at consequent time points in the array.

4.3.4 – Analysis of T2 Binding

As the binding of the small probe is expected to be weak and non-specific (with a dissociation constant in the high mM range), it is reasonable to assume that the equilibrium between the free and bound state is in fast exchange (Lepre et al., 2004). Because of this, the analysis of the extent of binding becomes less complex, where the T2 observed is simply proportional to the populations of the free and bound states:

$$\frac{1}{T_{2obs}} = f_b \frac{1}{T_{2b}} + (1 - f_b) \frac{1}{T_{2free}} \quad \text{where} \quad f_b = \frac{[SP]}{[S]_{total}} \quad (4)$$

where T_{2obs} is the T2 observed (probe in the presence of protein), T_{2b} is the T2 of the bound state, which is assumed to be the T2 of the protein, T_{2free} is the T2 of the free probe, and f_b is the fraction of probe bound. The ^1H transverse relaxation times of proteins were experimentally determined to be 0.030 s, 0.020 s, and 0.003 s for B-LgA, A-ChytA, and BSA, respectively. We can then solve for the amount of fraction bound, where the equation becomes:

$$f_b = \frac{T_{2b}(T_{2f} - T_{2obs})}{T_{2obs}(T_{2f} - T_{2b})} \quad (5)$$

When we consider nonspecific interactions, the probe is likely to bind to multiple sites on the protein. To simplify matters, we assume that there are *n* equivalent binding sites, which gives use the binding model:

$$\frac{[SP]}{[P_o]} = n\left(\frac{[S]}{[S + K_d]}\right) \quad (6)$$

We can rearrange equation 6 to include the fraction bound to get:

$$f_b = \alpha - \sqrt{(\alpha^2 - \beta)} \quad (7)$$

where

$$\alpha = \frac{[S]_o + n[P]_o + K_d}{2[S]_o} \quad (8)$$

and

$$\beta = \frac{n[P]_o}{[S]_o} \quad (9)$$

From this binding model, we can see that the T2 values obtained in the experiment can be used to determine the K_d, and *n*. However, this is highly dependent on how it is fit [37] and could lead to data over-interpretation due to multiple unknown variables. Thus, quantification of the binding becomes difficult and we opted to provide a semi-quantitative approach by comparing the fraction bounds of the probes instead.

4.3.5 – Concentration Dependent Change in Fraction Bound

In order to investigate if fraction bound was a suitable measure for semi-quantitative analysis, we conducted a series of T2 experiments at different ratios of probe to protein as exemplified in Figure 4.2. The two model probes (phenol and *tert*-butyl alcohol) were tested against BSA, A-ChytA, and B-LgA, where the probe concentration was held constant and the protein concentration was increased. As shown, the increase in protein concentration causes a linear increase in the proportion of the fraction bound, at least at concentrations where the probe is still in large excess.

If we were to provide a linear fit, the steepness could be considered as a semi-quantitative measurement of protein hydrophobicity, where steeper slops are correlated to a larger degree of bound probe compared to a shallow slope. Thus, we can deduce that BSA is the most hydrophobic whereas B-LgA is the least based on phenol. *Tert*-butyl alcohol had less affinity for BSA than phenol and no significant association with A-ChytA and B-LgA.

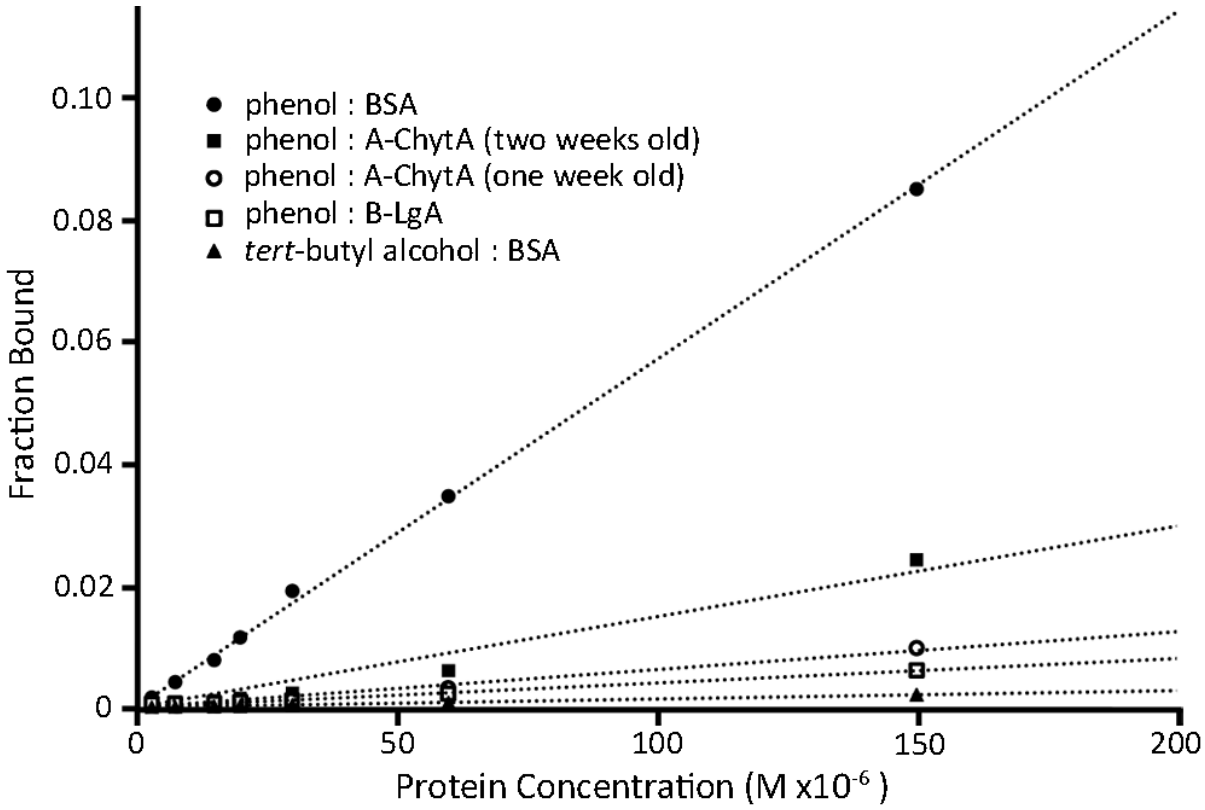

Figure 4.2. Increasing Protein Concentration Results in a Larger Population of Fraction Bound. Plot showing the fraction bound of probe vs the concentration of protein. The probe:protein mixtures are labeled and a linear fit of each data set is shown. Two sets of data are shown for phenol:A-ChytA to better represent the variability of T2s obtained, which is due to protein degradation/activation at different time points of protein preparation.

Also, Figure 4.2 illustrates that, the T2 measurements of phenol was fairly inconsistent in the presence of A-ChytA, which was quite surprising. Upon further investigation of A-ChytA's behavior in solution, we found that it is most likely due to the propensity of A-ChytA to degrade/oligomerize over time, which could be potentially related to the autocatalysis from trace impurities of chymotrypsin [38]. Dr. Zecca also performed time based analysis of A-ChytA and determined that A-ChytA does degrade

over time and begins to degrade within a day. Thus, it is imperative that proteins are prepared fresh to minimize artifacts arising from degradation products. Alternatively, this could serve as a basis for determining if there are any changes in sample integrity.

Lastly, investigation of probe-protein interactions at a multitude of ratios can be time consuming. To streamline the process, we opted to investigate the differences of fraction bound at a particular ratio. As seen in Figure 4.2, there are noticeable differences in the individual fraction bounds at a particular ratio, but this approach requires that there is a sufficient amount of probe bound to have more confidence in the measured differences in the T2 and fraction bound values. For this, we chose the 1:50 ratio and prepared fresh samples for all subsequent experiments.

4.3.6 – Comparison of Aliphatic and Aromatic Probes

As previously shown by the concentration dependent changes in the fraction bound, the extent of phenol binding to the protein standards was much higher in comparison to *tert*-butyl alcohol, suggesting that these interactions are not equivalent. This is further extended to include additional probes as shown in Table 4.3, which encompasses more aliphatic probes. Though we were not able to find another suitable aromatic probe (preliminary tests with benzyl alcohol showed an irregular chemical shift pattern with multiple unique proton peaks overlapping, which is not ideal for peak analysis through integration or intensity), the results showed similar patterns with *tert*-butyl alcohol, where 1-propanol and 1-butanol interacted with BSA. The percent decrease in the T2 was fairly higher for 1-propanol (60%) and 1-butanol (73%) in comparison to *tert*-butyl (21%). Likewise, there was no perceptible binding of the aliphatic probes to either A-ChytA or B-

Table 4.3. Percent Reduction in T2 Relaxation Times of the Probes upon Binding to the Proteins.

Protein / Probe	BSA	A-ChytA	B-LgA
Tert-butyl alcohol	21	NB	NB
1-butanol	73	NB	NB
1-propanol	60	NB	NB
Phenol	97	11	18
Capped Amino Acids			
Leucine	58	9	NB
Valine	17	NB	NB
Phenylalanine	84	NB	8
Tryptophan	94	21	19
Tyrosine	79	6	NB

The values shown above are the percent reduction in T2 when compared to the T2 of the free probe using the average T2 value obtained, where percent change $= 100 * (T_{2f} - T_{2obs})/T_{2f}$. Percent changes of less than 5% were within experimental error and are denoted with NB (no binding). The ratio for protein:probe is 1:50, where the protein concentration is 0.03 mM and the probe is 1.5 mM. T2 of the free probes and integration regions are: *tert*-butyl alcohol = 2.32s (1.28-1.18ppm), 1-butanol = 2.69 s (0.86-0.6 ppm), 1-propanol = 3.32 s (0.86-0.6 ppm), phenol = 5.16 s (7.4-7.26 ppm), leucine = 0.67 s (0.84-0.7 ppm), valine = 0.83 s (0.95-0.6 ppm), phenylalanine = 2.04 s (7.28-7.08 ppm), tryptophan = 2.20 s (7.56-7.45 ppm), tyrosine = 1.46 s (7.14-6.86 ppm).

LgA. On the contrary, phenol does interact with all three proteins and had a significant drop in T2 in the presence of BSA (97%), leading to a hydrophobicity order from most to least:

BSA > B-LgA > A ChytA

4.3.7 – Usage of Capped Amino Acids as Probes

As the goal of this study is to determine the surface hydrophobicity and attempt to correlate that to how proteins/antibodies behave in solution, we expanded the library of probes to include amino acids. The amino acids normally have charged ends due to the amino and carboxylic acid functional groups, so it was necessary to utilized capped amino acids, where the amino group is acetylated and the carboxylate is methylated. With the charges neutralized, we focused on the interaction of the side chains. Table 4.3 shows a similar pattern to our small molecular probes, where the aromatic amino acids have a large decrease in their T2 values in the presence of BSA in comparison to the aliphatic amino acids. The aliphatic amino acids exhibited no binding to A-ChytA or B-LgA. The 9% decrease for Leu can be attributed to its low T2 (0.67 s alone and 0.61 s in the presence of A-ChytA).

On the other hand, the aromatic amino acids were expected to interact with both A-ChytA and B-LgA due to similarities with phenol. Trp bound to both proteins, but both Tyr and Phe showed weak or no binding. Due to the inconsistencies in the T2, we questioned the reliability of these capped amino acids as probes. Since the side chains are most likely not the factor, we hypothesized that the capped C- and N-termini were responsible, which would be due to increased bulkiness and more steric hindrance.

Furthermore, we found that different conformations of capped amino acids exists in solution. In general, the different conformations of a compound are readily interconverted in solution, so the corresponding peaks in the ^{1}H NMR spectrum represent an average of the conformations. However, we found that capped amino acids have

different conformations in slow exchange, leading to at least two distinct conformations of non-equal populations. This is demonstrated in Figure 4.3, which depicts the aromatic region of the ^{1}H spectrum for Trp. Since Trp only has five unique aromatic protons, we can conclude that there are multiple conformations of Trp as manifested by the more complex pattern of peaks. The integrations of the peaks (as described in the figure legend) suggests that it is indeed five protons.

This is further complicated because each individual conformation behaves differently in solution as observed by the differential binding patterns (Figure 4.3). If we were to focus on the left most hydrogen, we see that the left doublet at 7.54 ppm in the presence of BSA and partially disappears with B-LgA, but remains with A-ChytA. Conversely, the right doublet at 7.48 ppm disappears in the presence of A-ChytA, but not BSA or B-LgA. This is highly undesirable as the exponential fit is complicated due to the differing binding capacities of individual conformations, which then necessitates the use of a multiple exponential fit rather than a single exponential. For example, taking the integral over the range from 7.56 ppm to 7.45 ppm to examine the T2 of one hydrogen would no longer be appropriate for analysis by equation 1 due to the presence of multiple unique protons.

Thus, the T2 values defined from these experiments becomes questionable. Though this interaction-based conformational heterogeneity is interesting on its own terms, it serves little purpose to the goal at hand. Based on these findings, we concluded that Trp is not an appropriate probe to use. Furthermore, different conformations were also observed for Leu and Tyr, but not Phe and Val. Though Phe and val do not explicitly

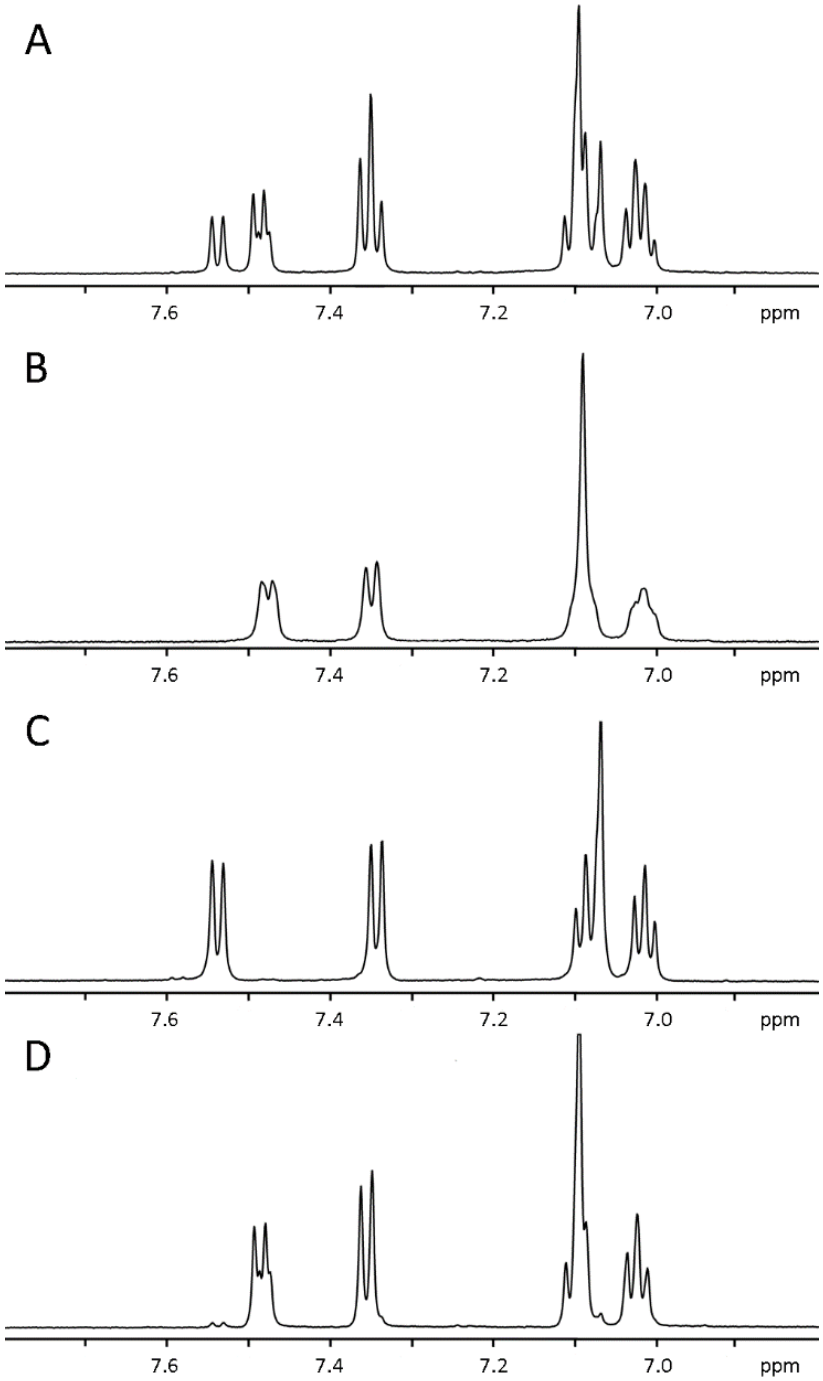

Figure 4.3. Conformations of Trp and their Behavior in the Presence of Proteins. Capped-Trp spectra are shown zoomed on the aromatic region. The concentrations of Trp (shown for Trp alone in A) are 3 mM and the ratio of Trp to proteins, (shown for BSA in **(B)**, for A-ChytA in **(C)**, and for B-LgA in **(D)**), is 1:100. Integration areas are: 7.56-7.45 ppm = 0.99, 7.4-7.3 ppm = 1.02, 7.15-7.06 ppm = 2.0, 7.05-6.98 ppm = 0.99. This correlates to the five aromatic protons of Trp.

show different conformations within the NMR spectra, it is better to err on the side of caution and avoid their use.

4.3.8 – Applicability of DOSY

Since we had found that T2 is a reasonable parameter for determination of the surface hydrophobicity of a protein, we investigated the possibility of using an alternative parameter to quantify the interaction between the probe and protein. The purpose of this was to replicate our findings in a different experiment to corroborate the surface hydrophobicity. For this, we opted to look at the diffusion coefficient, which would be determined through DOSY. The analysis of the diffusion coefficient is highly similar to T2, where the probe/protein complex are in fast exchange, leading to the dependence of the observed diffusion coefficient on the populations of the free and bound states:

$$D_{obs} = f_b D_b + (1 - f_b) D_{free} \qquad (10)$$

where D_{obs} is the observed diffusion coefficient of the probe in the presence of protein, D_b is the diffusion coefficient of the bound state, f_b is the fraction of probe bound, and D_{free} is the diffusion coefficient of the free probe.

Though we had hoped that the diffusion coefficient would provide similar results to our T2 experiments, the conclusion was that DOSY was not an adequate approach for investigating weak interactions. As presented in Table 4.4, the change on the diffusion coefficient of phenol in the presence and absence of proteins. The most significant change is anticipated to be with BSA, which showed a 97% decrease in T2. Here, we see that the diffusion coefficient dropped from 3.74×10^{-10} m^2/s to 3.59×10^{-10} m^2/s, which

Table 4.4. Determination of Diffusion Coefficients.

Sample	Diffusion coefficient ($10e^{-10}$ m^2/s)
Phenol Only	3.7400
Phenol + BSA	3.5940
Tert-butyl alcohol Only	3.1794
Tert-butyl alcohol + BSA	3.0922
DSS	2.3821±0.022

The diffusion coefficient of the samples were calculated by peak integration (phenol = 7.4-7.26 ppm, *tert*-butyl = 1.28-1.18 ppm, DSS = 0.05-$^{-}$0.05 ppm). The concentrations were 3 mM for BSA and 1.5 mM for the probe. The concentration of DSS (in an external compartment) was 5 mM. The external reference was reused for multiple runs for calibration and test runs, which is how the diffusion coefficient and standard deviation of DSS was calculated (n = 16). No error is given for the other runs as they were only run once.

amounts to a 4% decrease. This drop is very modest and falls close to experimental error. Though the probe/protein experiments were only run once as a test, we also analyzed the diffusion coefficient of DSS a multitude of times (n= 16), which was determined from be 2.382 ± 0.022 $x10^{-10}$ m^2/s, where the error itself amounts to ~1%. Though the difference is above the error, the difference is still minimal, which means that it further experiments would require a lower probe:protein ratio like 1:20 rather than 1:50. However, lower ratios are more problematic due to either the lower signal to noise of the probe (lower probe concentration) or more protein spectral overlap (higher protein concentration).

To understand what the problem could have been, we theoretically determined the diffusion coefficient of the protein (D_b) based on the Stokes-Einstein equation [39]:

$$D = \frac{kT}{6\pi\eta r_H} \quad (11)$$

where the diffusion coefficient, D, depends on the Boltzmann constant (k), temperature (T), viscosity (η), and the hydrodynamic radius (r_H). The hydrodynamic radius of a protein is variable, but we can simply assume that it adopts a spherical shape for this calculation, which means that the minimum radius [40], assuming the specific volume of a protein is 0.73 cm^3/g, is:

$$R_{min} = 0.066\sqrt[3]{M} \quad (12)$$

where R_{min} is related to the cubic root of the molecular weight (M). Once we plug in the values, we see that the estimated D_b for BSA is 0.915 x 10^{-10} m^2/s, which is markedly close to the D_{free} of phenol. This is most likely the reason why the effect of binding is minimal for DOSY as the difference between the D_{free} and D_b is much smaller than T_{2free} and T_{2b}. Thus, we were unable to use DOSY to validate our T2 experiments.

4.3.9 – Correlation with HIC and Fluorescence

Though DOSY was not an effective method for this project, investigation of the surface hydrophobicity by Dr. Zecca through HIC and Fluorescence showed promising results. The hydrophobicity order, from most to least, that was established by NMR was BSA > B-LgA > A-ChytA with phenol as the probe.

For fluorescence, an extrinsic fluorescence probe ANS was used, where its quantum yield and wavelength of maximum fluorescence depends on if it is present in an aqueous or hydrophobic environment. Through this, it was determined that the surface hydrophobicity (measured by the initial slope S_0 as shown in Table 4.5) order was BSA > B-LgA > A-ChytA, which coincides well with our NMR data. This is further validated

Table 4.5. Initial slopes (S_0) of ANS bound to BSA, B-LgA and A-ChytA.

Protein	Surface Hydrophobicity (S_0)
BSA	1355.2
B-LgA	15.4
A-ChytA	2.8

The initial slope of the relative fluorescence intensity (RFI) at 470 nm is shown above for each individual protein with ANS as the extrinsic dye, where the initial slope is representative of the surface hydrophobicity.

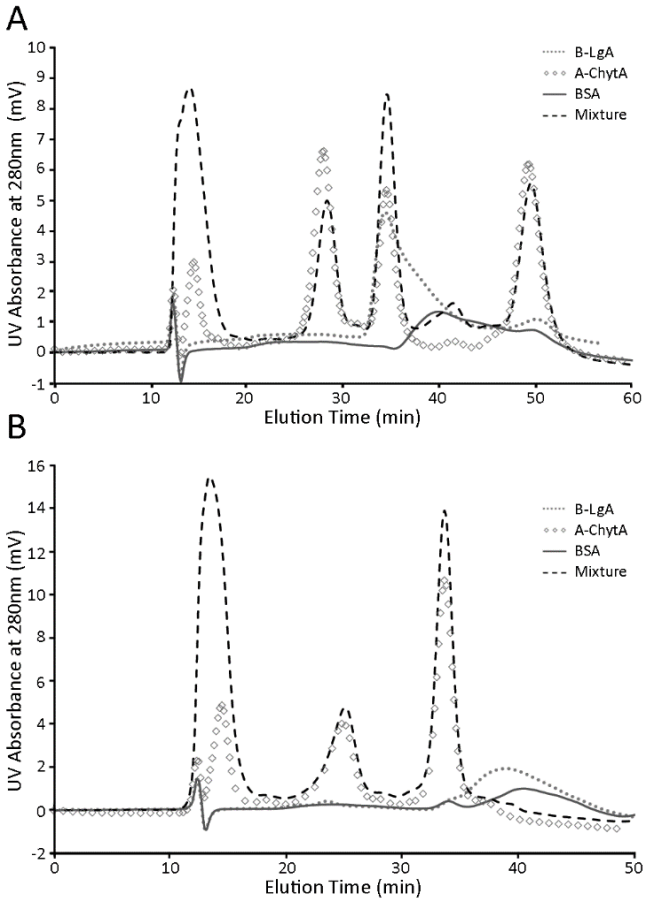

Figure 4.4. HIC Elution Profiles. The HIC elution profiles of BSA, A-ChytA, B-LgA, and Mixture on **(A)** Phenyl HP or **(B)** Butyl HP column are shown above, where the proteins were detected by monitoring the absorbance and 280 nm. The gradient was a 30 minute linear gradient of 1.0 M to 0.0 M ammonium sulfate in 20 mM sodium phosphate at a 0.5 mL/min flow rate. The column was continually washed with 20 mM sodium phosphate after 30 minutes to allow all the protein to elute and the baseline to return to normal.

through HIC, where the protein is subjected to an immobilized hydrophobic phase. The more hydrophobic proteins would remain bound to the column and would be difficult to elute, so that the elution profile would give an indication as to the relative hydrophobicities of the proteins. In Figure 4.4, the retention times showed that the hydrophobicity orders were similar, where A-ChytA was the least hydrophobic and BSA was the most. The order was the same when we utilized an immobilized aliphatic (butyl) or aromatic (phenyl) phase.

The elution profiles do show that there are multiple peaks, indicating heterogeneity in the sample arising from fragments/aggregates or different conformations. This behavior could potentially be attributed to the ammonium sulfate concentration and/or the strength of the hydrophobic stationary column [41]. In particularly, BSA has been reported to change conformations upon binding to a butyl HP column and unfolded proteins are difficult to elute off, resulting in broader peaks [42]. Therefore, it is difficult to assess the overall extent of aliphatic and aromatic interactions due to these caveats. However, we do observe that there are distinct differences in the overall elution pattern. B-LgA has a relatively sharper peak for the phenyl column, which could be due to stronger π-π interactions. On the other hand, A-ChytA has multiple peaks for both column, but possesses an additional peak for the phenyl column [43], implicating that there is a difference when comparing aliphatic and aromatic interactions, which coincides with what we observed with NMR.

4.4 – Conclusion

We experimentally determined the order of hydrophobicity to be BSA > B-LgA > A-ChytA, from most to least by NMR, HIC, and Fluorescence. This helps to prove that solution NMR is a fast and robust technique that is able to investigate the interaction between small molecular probes and proteins, where aliphatic and aromatic probes were determined to have differences in the non-specific binding capabilities. The use of capped amino acids revealed that the amino acids had multiple conformations, which is most likely due to the bulkier capped ends. Therefore, it is advised to not use capped amino acids despite its potential for mimicking protein-protein interactions.

Though we see hydrophobicity order matches between the three techniques utilized, there still remains questions that need to be answered. For instance, further studies are needed to better characterize phenol as an optimal binding probe (or possibly find another small aromatic probe), to determine a suitable method for quantification of surface hydrophobicity by NMR, to understand how and when aliphatic probes bind, how sample integrity relates to the observed differences T2, and to characterize monoclonal antibodies, whom are industrially relevant. Nonetheless, we feel that NMR is a versatile technique that is suitable for measuring surface hydrophobicity in which we hope can be used as a predictive measure for protein behavior in solution.

4.5 – References

1. Wang, W., *Instability, stabilization, and formulation of liquid protein pharmaceuticals.* Int J Pharm, 1999. **185**(2): p. 129-88.
2. Awotwe-Otoo, D., et al., *Quality by design: impact of formulation variables and their interactions on quality attributes of a lyophilized monoclonal antibody.* Int J Pharm, 2012. **438**(1-2): p. 167-75.
3. Chames, P., et al., *Therapeutic antibodies: successes, limitations and hopes for the future.* Br J Pharmacol, 2009. **157**(2): p. 220-33.
4. Farid, S.S., *Process economics of industrial monoclonal antibody manufacture.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **848**(1): p. 8-18.
5. Nicholls, A., K.A. Sharp, and B. Honig, *Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons.* Proteins, 1991. **11**(4): p. 281-96.
6. Pace, C.N., et al., *Contribution of hydrophobic interactions to protein stability.* J Mol Biol, 2011. **408**(3): p. 514-28.
7. Young, L., R.L. Jernigan, and D.G. Covell, *A role for surface hydrophobicity in protein-protein recognition.* Protein Sci, 1994. **3**(5): p. 717-29.
8. Tsai, C.J., et al., *Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect.* Protein Sci, 1997. **6**(1): p. 53-64.
9. Jones, S. and J.M. Thornton, *Principles of protein-protein interactions.* Proc Natl Acad Sci U S A, 1996. **93**(1): p. 13-20.
10. Giovambattista, N., et al., *Hydrophobicity of protein surfaces: Separating geometry from chemistry.* Proc Natl Acad Sci U S A, 2008. **105**(7): p. 2274-9.
11. Fink, A.L., *Protein aggregation: folding aggregates, inclusion bodies and amyloid.* Fold Des, 1998. **3**(1): p. R9-23.
12. Wang, W., *Protein aggregation and its inhibition in biopharmaceutics.* Int J Pharm, 2005. **289**(1-2): p. 1-30.
13. Bigelow, C.C., *On the average hydrophobicity of proteins and the relation between it and protein structure.* J Theor Biol, 1967. **16**(2): p. 187-211.
14. Kato, A. and S. Nakai, *Hydrophobicity determined by a fluorescence probe method and its correlation with surface properties of proteins.* Biochim Biophys Acta, 1980. **624**(1): p. 13-20.
15. Amin, S., et al., *Protein aggregation, particle formation, characterization & rheology.* Current Opinion in Colloid & Interface Science, 2014. **19**(5): p. 438-449.
16. Haynes, C.A., E. Sliwinsky, and W. Norde, *Structural and Electrostatic Properties of Globular Proteins at a Polystyrene-Water Interface.* Journal of Colloid and Interface Science, 1994. **164**(2): p. 394-409.
17. Cardamone, M. and N.K. Puri, *Spectrofluorimetric assessment of the surface hydrophobicity of proteins.* Biochem J, 1992. **282 (Pt 2)**: p. 589-93.
18. Roettger, B.F. and M.R. Ladisch, *Hydrophobic interaction chromatography.* Biotechnol Adv, 1989. **7**(1): p. 15-29.
19. Rassi, Z.E. and C. Horváth, *Hydrophobic Interaction Chromatography of t-RNA's and Proteins.* Journal of Liquid Chromatography, 1986. **9**(15): p. 3245-3268.

20. Melander, W. and C. Horvath, *Salt effect on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series.* Arch Biochem Biophys, 1977. **183**(1): p. 200-15.
21. Goheen, S.C. and B.M. Gibbins, *Protein losses in ion-exchange and hydrophobic interaction high-performance liquid chromatography.* J Chromatogr A, 2000. **890**(1): p. 73-80.
22. Mahn, A., M.E. Lienqueo, and J.A. Asenjo, *Optimal operation conditions for protein separation in hydrophobic interaction chromatography.* J Chromatogr B Analyt Technol Biomed Life Sci, 2007. **849**(1-2): p. 236-42.
23. Queiroz, J.A., C.T. Tomaz, and J.M. Cabral, *Hydrophobic interaction chromatography of proteins.* J Biotechnol, 2001. **87**(2): p. 143-59.
24. Hawe, A., M. Sutter, and W. Jiskoot, *Extrinsic fluorescent dyes as tools for protein characterization.* Pharm Res, 2008. **25**(7): p. 1487-99.
25. Alizadeh-Pasdar, N., E.C. Li-Chan, and S. Nakai, *FT-Raman spectroscopy, fluorescent probe, and solvent accessibility study of egg and milk proteins.* J Agric Food Chem, 2004. **52**(16): p. 5277-83.
26. Haskard, C. and E. Li-Chan, *Hydrophobicity of Bovine Serum Albumin and Ovalbumin Determined Using Uncharged (PRODAN) and Anionic (ANS-) Fluorescent Probes.* J. Agric. Food Chem., 1998. **46**: p. 2671-2677.
27. Pellecchia, M., D.S. Sem, and K. Wuthrich, *NMR in drug discovery.* Nat Rev Drug Discov, 2002. **1**(3): p. 211-9.
28. Barile, E. and M. Pellecchia, *NMR-based approaches for the identification and optimization of inhibitors of protein-protein interactions.* Chem Rev, 2014. **114**(9): p. 4749-63.
29. Meiboom, S. and D. Gill, *Modified Spin-Echo Method for Measuring Nuclear Relaxation Times.* Review of Scientific Instruments, 1958. **29**(8): p. 688-691.
30. Aguilar, J.A., et al., *Spin echo NMR spectra without J modulation.* Chem Commun (Camb), 2012. **48**(6): p. 811-3.
31. Pauli, G.F., et al., *Quantitative 1H NMR. Development and potential of an analytical method: an update.* J Nat Prod, 2012. **75**(4): p. 834-51.
32. Frank, O., et al., *Accurate determination of reference materials and natural isolates by means of quantitative (1)h NMR spectroscopy.* J Agric Food Chem, 2014. **62**(12): p. 2506-15.
33. Henderson, T.J., *Quantitative NMR spectroscopy using coaxial inserts containing a reference standard: purity determinations for military nerve agents.* Anal Chem, 2002. **74**(1): p. 191-8.
34. Bharti, S. and R. Roy, *Quantitative 1H NMR spectroscopy.* Trends in Analytical Chemistry, 2012. **35**: p. 5-26.
35. Fielding, L., *NMR methods for the determination of protein–ligand dissociation constants.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2007. **51**: p. 219-242.
36. Dubois, B.W. and A.S. Evers, *19F-NMR spin-spin relaxation (T2) method for characterizing volatile anesthetic binding to proteins. Analysis of isoflurane binding to serum albumin.* Biochemistry, 1992. **31**(31): p. 7069-76.
37. Lepre, C.A., J.M. Moore, and J.W. Peng, *Theory and applications of NMR-based screening in pharmaceutical research.* Chem Rev, 2004. **104**(8): p. 3641-76.

38. Fischer, J.J. and O. Jardetzky, *Nuclear Magnetic Relaxation Study of Intermolecular Complexes. The Mechanism of Penicillin Binding to Serum Albumin.* J Am Chem Soc, 1965. **87**: p. 3237-44.
39. Fielding, L., S. Rutherford, and D. Fletcher, *Determination of protein-ligand binding affinity by NMR: observations from serum albumin model systems.* Magn Reson Chem, 2005. **43**(6): p. 463-70.
40. Velev, O.D., E.W. Kaler, and A.M. Lenhoff, *Protein interactions in solution characterized by light and neutron scattering: comparison of lysozyme and chymotrypsinogen.* Biophys J, 1998. **75**(6): p. 2682-97.
41. Edward, J.T., *Molecular volumes and the Stokes-Einstein equation.* Journal of Chemical Education, 1970. **47**(4): p. 261.
42. Erickson, H.P., *Size and shape of protein molecules at the nanometer level determined by sedimentation, gel filtration, and electron microscopy.* Biol Proced Online, 2009. **11**: p. 32-51.
43. Wu, S.L., A. Figueroa, and B.L. Karger, *Protein conformational effects in hydrophobic interaction chromatography. Retention characterization and the role of mobile phase additives and stationary phase hydrophobicity.* J Chromatogr, 1986. **371**: p. 3-27.
44. Ueberbacher, R., et al., *Hydrophobic interaction chromatography of proteins V. Quantitative assessment of conformational changes.* J Chromatogr A, 2008. **1198-1199**: p. 154-63.
45. Fausnaugh, J.L., L.A. Kennedy, and F.E. Regnier, *Comparison of hydrophobic-interaction and reversed-phase chromatography of proteins.* J Chromatogr, 1984. **317**: p. 141-55.

Chapter 5 – Summary and Conclusions

The work presented in this dissertation has covered a wide range of goals and biological systems. However, the core that ties them all together is that investigation protein-protein and protein-ligand interactions are crucial for understanding structural features/changes as well as their functional role *in vivo*.

Chapter 2 explores the interaction between CD47 and PLIC-2 as a novel protein complex that could be related to how CD47 ultimately modulates the cytoskeleton. However, the *in vitro* results using the isolated domains of PLIC-2 and the cytoplasmic tail of CD47 isoform proved to be fruitless, where no detectable interaction was observed through ITC or NMR. The addition of a membrane mimetic (nanodiscs) did not provide positive results. On the other hand, PLIC-2, also known as ubiquilin-2, was shown to be able to self-associate (UBA/UBL complex) as well as bind to ubiquitin through its UBA domain, which is in line with the literature that have shown ubiquilin's role in ubiquitin mediated protein degradation. Though it is unlikely that CD47 and PLIC-2 associate, it may be reasonable to test other constructs (full length PLIC-2, full length CD47, CD47/integrin complexes). The additional residues could be a requirement for interacting, especially considering that our *in vitro* experiments may not be an ideal representation of *in vivo* conditions.

Chapter 3 investigates the activation mechanism of BTN3A1, who responds to intracellular pAg in order to activate Vγ9Vδ2 T cells. Association of the B30.2 and HMBPP causes minimal changes in the B30.2 domain based on the crystal structure, but there was an apparent change with the JM region by NMR. Further investigation of

the JM region specifically showed that the JM is able to interact with negatively charged diphosphate molecules in which two threonines were identified to be important. The binding itself was weak, but was corroborated with SAXS, ITC, and cellular assays. Therefore, this conformational change upon ligand association is particularly interesting to understand how pAg can activate BTN3A1. The data provided here does not provide a clear picture of the molecular details and binding interfaces as well as if this conformational change directly influences the conformation of its extracellular domain and/or binding affinity towards other protein partners. These questions will have to be addressed in future endeavors.

Chapter 4 involves the use of small hydrophobic probes as a means for evaluating the surface hydrophobicity of a protein through NMR. The work here explores the limitations and benefits of T2 as a quantitative method. The results show that the probes are able to interact with the protein, leading to a depression in their observed T2 value. The hydrophobicity order determined by phenol was consistent with current standards (HIC and fluorescence). To improve upon what has been done, it would be beneficial to expand the protein library and to include monoclonal antibodies, identifying how and when aliphatic/aromatic probes would bind, and correlation to solution properties, such as aggregation.

Appendix Chapter A1 – A Structural Model of the Skelemin and Integrin $\alpha_{IIb}\beta_3$ Complex

Adapted from: Gorbatyuk, V., **Nguyen, K.**, Podolnikova, N., Deshmukh, L., Lin, X., Ugarova, T., and Vinogradova, O. *Skelemin Association with aIIbβ3 Integrin: A Structural Model*, Biochem., 2014, 53: 6766-6775.

A1 – Introduction

It has been known that cell adhesion is primarily mediated by integrins[1] by which it connects the extracellular matrix to the cytoskeleton. It begins with the clustering of proteins to form a small network called motility induced focal complexes (FXs), which could ultimately be replaced by focal adhesions (FAs) in fully spread cells[2]. However, understanding the formation and components of FXs and FAs is crucial to understanding how the cells regulate cellular adhesion. The work presented in this chapter focuses on the role of skelemin, which is a protein found to be localized with FXs, but not FAs[3].

Skelemin, also known as myomesin 1.1, is a 185 kDa muscle M-line cytoskeletal protein that is predominantly expressed in embryonic heart[4], playing a crucial role in the early stages of cell spreading[5]. Skelemin is comprised of a unique N-terminal myosin binding domain, 5 fibronectin (FN) type III like domains, 6 immunoglobulin like C2 domains (IgC2), and a C-terminal immunoglobulin domain responsible for dimerization[6] as shown in Figure A1. The importance of skelemin in understanding FXs is that it has been shown that its IgC2 domains 4 and 5, hereafter referred to as Sk45, directly

Figure A1. Simplistic Diagram of Skelemin's Domains. It consists of a unique N-terminal myosin binding domain (MBD), seven IgC2, and five FN Type III-like domains. The IgC2 domains are numbered (shown within the ellipse) to more easily visualize the location of IgC2 domains 4 and 5.

associated with the cytoplasmic tail of integrin β_3 and integrin α_{IIb}[7,8]. Considering that integrin exists as a heterodimer of α and β subunits, this lends itself to the proposition that Sk45 is able to simultaneously bind to both integrin subunits, in which it may then modulate the cytoskeleton through contractile force and/or the recruitment of other cytoskeletal proteins to the integrin cluster[9]. Given that there is evidence that skelemin is crucial for cell spreading and is able to associate with both integrin α_{IIb} and β_3, the goal of the chapter presented here aims to deduce the overall structure of the skelemin and integrin complex.

A2 – Materials and Methods

Mutagenesis, Expression, and Purification

For these experiments, mouse skelemin was cloned into *E. coli.* Single site mutagenesis of Sk45 of K_{1424} to C was performed using the QuikChange kit (Agilent Technologies). The mutant plasmid was transformed into Rosetta (DE3) competent cells (EMD Millipore). Protein expression was carried out using LB or M9 minimal media with $^{15}NH_4Cl$ as the sole nitrogen source at 37 ^{o}C. Cultures were grown to an OD_{600} of ~0.6

and induced with 1mM of IPTG. After 4 hours, the cells were collected by centrifugation. For purification, cells were resuspended in buffer (20 mM Tris, pH 8, 300 mM NaCl, 10 mM imidazole, 1 mM DTT, 1mM PMSF, 1 tablet of complete protease inhibitor cocktail (Roche Applied Sciences) and lysed by passage through a French press (Thermo Electron). The supernatant was incubated with Ni-NTA agarose resin (Qiagen) and the protein was eluted with a buffer containing 20 mM Tris, pH 7.5, 300 mM NaCl, and 500 mM imidazole. For cleavage of the His-tag, thrombin was added directly to the Ni-NTA column, where cleavage was done at 37 ^{o}C for 4 hours with periodic mixing (50 mM Tris, 150 mM NaCl, 1 mM EDTA, pH 7.5). Thrombin was subsequently inhibited by the addition of equivalent amounts of PMSF. The eluate was further purified by size exclusion chromatography (16/60 Superdex 75 column, GE healthcare) in 50 mM NaCl, 20 mM KPO_4, pH 6.8.

Paramagnetic Relaxation Experiments

NMR experiments were performed on uniformly ^{15}N-labeled samples prepared in buffer containing 20 mM KPO_4, pH 6.8, and 7% D_2O. Spectra were recorded at 25 °C on Varian INOVA 600 MHz spectrometer equipped with cold probe. Paramagnetic relaxation enhancement (PRE) experiments were performed with Sk45 K1424C construct, where the solvent exposed K^{1424} was mutated to cysteine. The ^{1}H-^{15}N HSQC spectra were collected on the samples with and without 3-maleimido-PROXYL (m-PROXYL) spin label. The effect of the spin label on the signal intensities was estimated from the normalized intensity differences in the spectra with and without the spin label.

Isothermal Titration Calorimetry

Isothermal Titration Calorimetry was performed on a low volume Nano ITC (TA Instruments). Peptides corresponding to integrin cytoplasmic tails were solubilized in the buffer 50mM NaCl, 20mM KPO_4, and pH 6.8. All ITC experiments were performed at 25^{o}C, 300 rpm mixing, 300 seconds time intervals between injections, and 3µl injection volumes. The concentrations used are as follows: 2.5 mM C-terminal β_3 and 0.175 mM Sk4, 1.3 mM N-terminal β_3 and 0.177 mM Sk45, and 1.9 mM α_{IIb} and 0.177 mM Sk45. The analysis of the data was done in NanoAnalyze Software (TA Instruments) suite using "Independent" model.

A3 – Results and Discussion

Table A1. Thermodynamic Analysis of the Association of Skelemin and Integrin by ITC

Titrant	C-terminal β_3	N-terminal β_3	α_{IIb}
Protein	Sk4	Sk45	Sk45
Kd (µM)	16.7	37.2	14.2
ΔG (kJ/mol)	-27.3	-25.3	-27.7
ΔH (kJ/mol)	-0.6	-1.7	-1.6
-TΔS (kJ/mol)	-26.6	-23.6	-26.1
Stoichiometry, n	1.00	1.30	1.26

Skelemin IgC domains 4 and 5 have been shown to interact with β_3 cytoplasmic tail previously[10]. To address the mechanism of this interaction and to define the thermodynamic forces driving the process, we have employed ITC. We have used Sk4 and Sk45 constructs titrated with either full length α_{IIb} cytoplasmic tail or short synthetic peptides corresponding to β_3 N-or C-termini. The results, summarized in Table A1, revealed very weak interactions, which were in the tens µM range, that were

predominantly driven by entropy. For all cases, the stoichiometry of interactions was found to be one. The data also depicts that the C-terminal region of β_3 interacts with the fourth IgC domain (Sk4) whereas the N-terminal region of β_3 likely interacts with the fifth IgC domain (Sk5). Sk5 was not able to be tested independently due to stability problems in the absence of the fourth IgC domain.

It is also important to note that the affinity of the cytoplasmic tails of α_{IIb} and β_3 is around 5 μM, which is stronger than their affinity towards skelemin. This supports the notion that skelemin serves as a modulator and not an activator of integrin, where skelemin is able to associate after integrin is activated and its cytoplasmic tails are separated. The hypothesis that skelemin is able to interact with both α_{IIb} and β_3 is further corroborated as α_{IIb} and β_3 had unique binding interfaces on Sk5 as determined by chemical shift perturbation experiments (performed by Dr. Deshmuhk). The residues of Sk5 identified were 1361, 1363 and 1394 for α_{IIb} and 1368, 1370, 1372, 1374, 1382-1384, and 1411 for N-terminal β_3. With this information, we were able to dock the cytoplasmic tails of α_{IIb} and β_3 onto Sk45 to provide a model system.

However, the structure of Sk45 was not yet solved from the NMR constraints obtained due to the fact that the Sk4 and Sk5 domains were connected by a flexible linker. This resulted in inadequate resolution of the orientation between the two domains due to the lack of long distance restraints, which was then tackled through the use of residual dipolar coupling (RDC) and paramagnetic relaxation enhancement (PRE) experiments. RDC data provided additional long distance restraints for the structure calculations, which were performed by Dr. Gorbatyuk. The potential structures were thinned out by their agreement with the PRE data (Figure A2), which showed that m-proxyl attached at

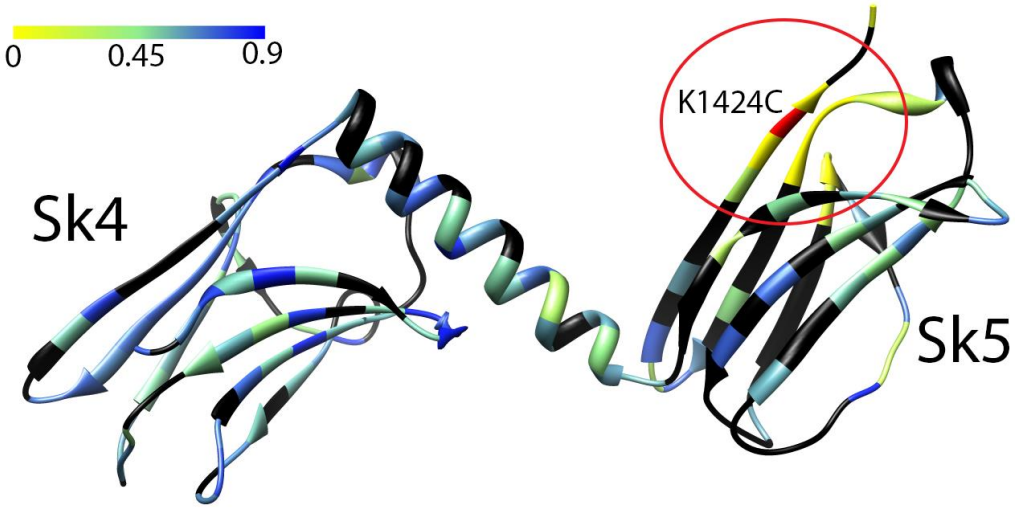

Figure A2. Visualization of PRE Affected Residues. Effects of m-proxyl on peak intensities of Sk45 were quantified by $I \div I_o$, where I and I_o is the intensity in the presence or absence of m-proxyl, respectively. Values closer to 0 represent highly affected residues. These changes are visualized on the crystal structure of human myomesin (PDB ID: 3RBS). Peaks that could not be quantified at colored black and the tag location (K1424C) is colored red. As shown, the residues that are mostly affected lies on Sk5 (circled in red) and are in close proximity to K1424C.

position 1237 of the Sk45 K1424C only had an effect on the residues within the Sk5 domain. This suggests that the two domains are likely to be far apart, where it is unlikely for the Sk5 domain to fold back to be in proximity of the Sk4 domain. Thus, the end result is an ensemble of structures (PDB ID: 4V10) of skelemin in a fairly extended state.

From there, the structure of Sk45 was used to dock the cytoplasmic tails of α_{IIb} and β_3 through the HADDOCK webserver[11], where the composite model is shown in Figure A3. The electrostatic potential was calculated by using Adaptive Poisson-Boltzmann Solver (APBS)[12] and the pdb2pqr webserver[13] using PARSE as the forcefield and PROPKA

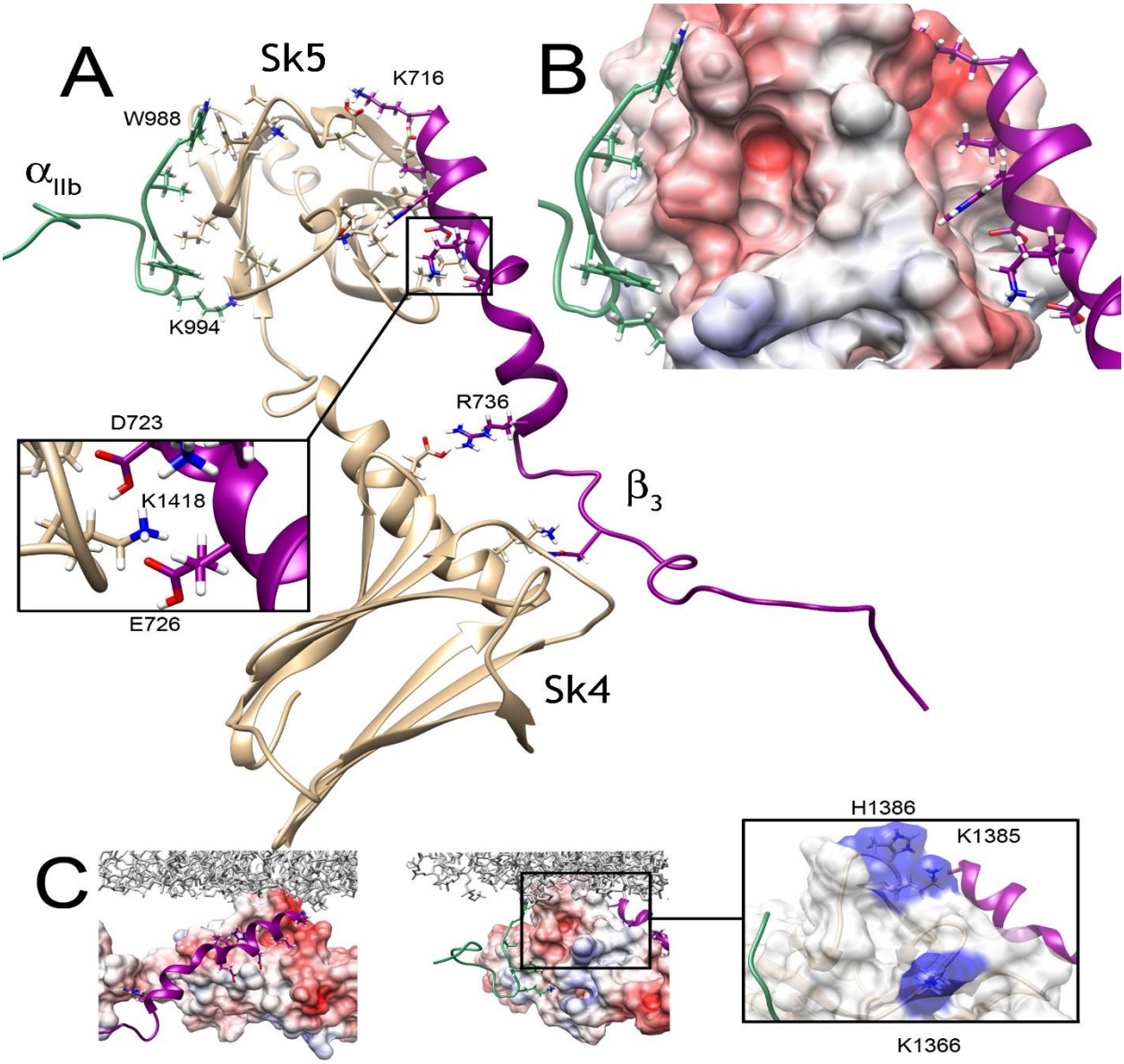

Figure A3. Model of the Integrin/Skelemin Complex. (A) Model of tertiary integrin/skelemin complex by HADDOCK: Sk45 (tan), α_{IIb} (green), and β_3 (purple). The inset shows hydrogen bonding network. (**B)** Zoomed view of the binding interface with Sk5 (colored by electrostatic potential, where the gradient is -10 (red), 0 (white), and 10 (blue)). **(C)** The tertiary complex is arbitrarily placed with respect to the lipid bilayer. Each side is shown to better visualize the binding pockets and potential orientation with the lipid bilayer. The inset depicts positively charged residues (colored in blue) that may interact with the lipid bilayer.

to assign protonation states. Based on this model, we see a number of van der-Waals interactions between α_{IIb} and Sk5's predominantly hydrophobic binding pocket. The major interactions are α_{IIb} F^{992} and F^{993} side chains with I^{1392} and the side chain of α_{IIb} W^{988} making hydrophobic contacts with F^{1388}, K^{1389} and, possibly, D^{1387} of IgC5. Additionally, a hydrogen bond is formed between K^{994} amine of α_{IIb} and the backbone oxygen of A^{1343}.

On the other hand, the β_3/skelemin interface is mostly based upon a network of hydrogen bonds between β_3 N-terminus and Sk5 as well as the linker between Sk4 and Sk5. The electrostatic surface potential of the zoomed Sk5 region presented in Figure A3B helps to visualize mostly hydrophobic cleft on the left interacting with α_{IIb} subunit and predominantly negatively charged binding site for β_3 subunit on the right. N-terminal K^{716} and K^{725} of β_3 forms hydrogen bonds with the side chain of E^{1384} and the backbone oxygen of D^{1415}, respectively. There are hydrophobic interactions between β_3 H^{722} and the side chain carbons of Sk45 K^{1413} and E^{1368}. From the skelemin side, Sk5 K^{1418} forms an extensive hydrogen bonding network, which includes β_3 residues D^{723}, through both the side chain and backbone oxygen, and β_3 E^{726}, through the side chain only. Two hydrogen bonds are found within skelemin linker helix: the side chain nitrogen of K^{1321} is connected to β_3 N^{744} while E^{1328} oxygen interacts with the side chain amine of R^{736}. Though there is no direct experimental evidence to pinpoint the residues involved for Sk4, the model demonstrates no major interactions between IgC4 and β_3, with only a couple of hydrophobic contacts found between side chains of IgC4 residues E^{1293} and N^{1294} and β_3 A^{750}.

In Figure A3C, we also show the positioning of the tertiary complex with respect to the lipid bilayer. As presented, one potential arrangement is that the skelemin Ig domains

4 and 5 may be placed parallel to the membrane surface. With this, the α_{IIb} subunit comes out of the membrane almost perpendicularly with its W^{988} side chain making contacts with lipid head-groups at the membrane-cytoplasm interface. Integrin β_3 is arranged at a sharp acute angle with its K^{716} side chain making contacts with negatively charged patch on the surface of Sk5. The top surface of Sk5 contains positively charged residues, which could associate with the negatively charged lipid membrane.

Lastly, the structure of skelemin's IgC domains were also solved by Pinotsis and colleagues[14] for the human isoform by crystallography (PDB ID: 3RBS). The structures coincided well with our NMR structure, where they adopted a well-known IgC2-fold, containing seven β-strands in two β-sheets forming two β-sandwiches. However, the major difference between the crystal and NMR structure lies within the linker region between the IgC2 domains (Figure A4). As shown, the crystal structure found that the linker region is highly helical, which depicts that the linker is rigid. In our solution structure, the N-terminal region of the linker adopts a helical fold that rests upon the Sk4 domain, which coincides with evidence that the Ig domain/helix interface area is structurally conserved throughout other Ig tandem domains of myomesin-1[14]. Strikingly, the C-terminal region of the linker was found to be flexible in solution, which allows for larger spatial orientations between the two IgC2 domains. Nonetheless, the difference between the crystal and solution structure can depict the elasticity of skelemin, where the linker region can serve as the springs/sensors of mechanical force. This is in line with current evidence as it has been shown that myomesin has similar elastic properties as titin through atomic force microscopy[14,15].

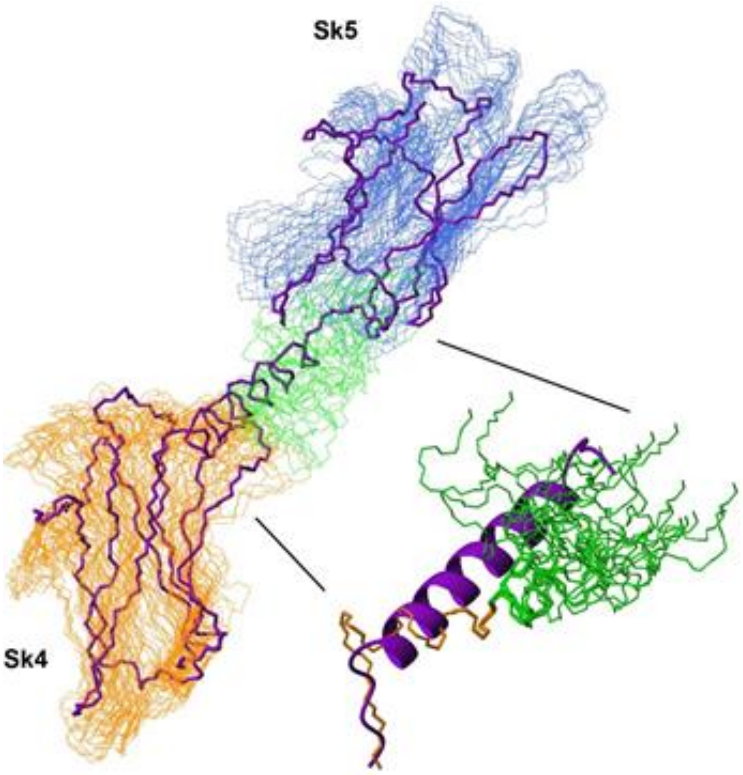

Figure A4. Comparison Between the NMR Ensemble and the Crystal Structure. The overlay of Sk45 NMR ensemble (the 20 lowest energy NMR structures with the two domains oriented by means of RDC data) with the X-ray structure of human homolog (PDB ID 3RBS). The NMR ensemble and the X-ray structure (violet) are superimposed over residues 1227-1327 (domain Sk4) and 1344-1427 (domain Sk5). The linker region is magnified and is shown in ribbon presentation for X-ray structure to highlight the difference in rigidity.

A4 – Conclusion

To conclude, we have determined by NMR a three dimensional structure of tandem IgC2-like domains 4 and 5 of skelemin, where the linker is less ordered / more dynamic than suggested by earlier X-ray studies of the human homolog myomesin-1. We show that interaction between skelemin and integrin cytoplasmic domain is weak and

predominantly entropy driven, consistent with the findings that skelemin is unable to activate the receptors. We have built a tertiary model of $\alpha_{IIb}\beta_3$ integrin cytoplasmic tails complexed with the tandem domains. While some of the interactions between K^{716} side chain of β_3 tail and skelemin may occur in the presence of α_{IIb} subunit still bound to β_3, disruption of membrane-proximal α/β clasp and release of α_{IIb} F^{992} and β_3 $H^{722}/D^{723}/E^{726}$ side chains is necessary for α_{IIb} to bind to skelemin and for β_3 to stabilize its interface with skelemin. Thus, our model favors skelemin role as a stabilizer of the activated integrins and integrin clusters by connecting α and β subunits from adjacent receptors.

A5 – References

1 Hynes, R. O. Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* **69**, 11-25 (1992).

2 Webb, D. J., Parsons, J. T. & Horwitz, A. F. Adhesion assembly, disassembly and turnover in migrating cells -- over and over and over again. *Nature cell biology* **4**, E97-100, doi:10.1038/ncb0402-e97 (2002).

3 Reddy, K. B., Bialkowska, K. & Fox, J. E. Dynamic modulation of cytoskeletal proteins linking integrins to signaling complexes in spreading cells. Role of skelemin in initial integrin-induced spreading. *The Journal of biological chemistry* **276**, 28300-28308, doi:10.1074/jbc.M102794200 (2001).

4 Steiner, F., Weber, K. & Furst, D. O. M band proteins myomesin and skelemin are encoded by the same gene: analysis of its organization and expression. *Genomics* **56**, 78-89, doi:10.1006/geno.1998.5682 (1999).

5 Price, M. G. & Gomer, R. H. Skelemin, a cytoskeletal M-disc periphery protein, contains motifs of adhesion/recognition and intermediate filament proteins. *The Journal of biological chemistry* **268**, 21800-21810 (1993).

6 Pinotsis, N., Lange, S., Perriard, J. C., Svergun, D. I. & Wilmanns, M. Molecular basis of the C-terminal tail-to-tail assembly of the sarcomeric filament protein myomesin. *The EMBO journal* **27**, 253-264, doi:10.1038/sj.emboj.7601944 (2008).

7 Deshmukh, L. *et al.* Structural insight into the interaction between platelet integrin alphaIIbbeta3 and cytoskeletal protein skelemin. *The Journal of biological chemistry* **282**, 32349-32356, doi:10.1074/jbc.M704666200 (2007).

8 Podolnikova, N. P. *et al.* Adhesion-induced unclasping of cytoplasmic tails of integrin alpha(IIb)beta3. *Biochemistry* **48**, 617-629, doi:10.1021/bi801751s (2009).

9 Li, X., Liu, Y. & Haas, T. A. Skelemin in integrin alpha(IIb)beta(3) mediated cell spreading. *Biochemistry* **52**, 681-689, doi:10.1021/bi301269a (2013).

10 Reddy, K. B., Gascard, P., Price, M. G., Negrescu, E. V. & Fox, J. E. Identification of an interaction between the m-band protein skelemin and beta-integrin subunits. Colocalization of a skelemin-like protein with beta1- and beta3-integrins in non-muscle cells. *The Journal of biological chemistry* **273**, 35039-35047 (1998).

11 de Vries, S. J., van Dijk, M. & Bonvin, A. M. The HADDOCK web server for data-driven biomolecular docking. *Nature protocols* **5**, 883-897, doi:10.1038/nprot.2010.32 (2010).

12 Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10037-10041, doi:10.1073/pnas.181342398 (2001).

13 Dolinsky, T. J. *et al.* PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research* **35**, W522-525, doi:10.1093/nar/gkm276 (2007).

14 Pinotsis, N. *et al.* Superhelical architecture of the myosin filament-linking protein myomesin with unusual elastic properties. *PLoS biology* **10**, e1001261, doi:10.1371/journal.pbio.1001261 (2012).

15 Tskhovrebova, L. & Trinick, J. Making muscle elastic: the structural basis of myomesin stretching. *PLoS biology* **10**, e1001264, doi:10.1371/journal.pbio.1001264 (2012).