1-10-2017

Structure-Function Investigation of Membrane Proteins Using Nanodiscs

Robbins J. Puthenveetil
University of Connecticut, robbins.puthenveetil@uconn.edu

Structure-Function Investigation of Membrane Proteins using Nanodiscs

Robbins Puthenveetil, Ph.D.

University of Connecticut, [2017]

Several membrane proteins, comprising of either α-helical or β-barrel structures have been successfully studied in vitro using several membrane mimetic systems. Nanodiscs, a new class of discoidal nanoscale lipoprotein complex have been used extensively in the present study to investigate different membrane proteins. We begin by delineating the development of small (D7) nanodiscs through its rigorous characterization and demonstrate advantages in solution NMR applications. In addition, we show the development of an on-column method to generate nanodiscs through the selective use of variable protein tags. D7 discs have been further used to visualize the downstream phosphorylation events of activated integrin β_3 through Src kinase. The C-terminal domains, Tpr^C and $MOSP^C$, two β-barrel membrane porins, were shown to trimerize when visualized through negatively stained transmission electron microscopy images in nanodiscs. The periplasmic and membrane conformers of native MOSP in *Treponema denticola* was established along with the partial identification of their multimeric complexes through a series of immuno-chemical experiments. Also discussed are two specific studies which revisit previously published papers where, (a) the direction of the cytoplasmic tail of integrin β_3 bound to the SH3 domain of Src kinase was shown to be in the reverse orientation w.r.t the published crystal structure; and (b) PLIC proteins, [Protein

Linking the cell membrane (IAP/CD47) to the Cytoskeleton] was shown to be involved in proteasomal degradation with no binding ability to the cytoplasmic tail of CD47. Small angle X-ray scattering studies of BTN3A1, an immune modulator, revealed the rearrangement of the intracellular cytoplasmic B30.2 domain with respect to the juxta membrane region in the presence of a phosphoantigen HMBPP. Lastly, a novel finding is briefly described involving orthlogs of TamB, a component of the transloaction and assembly module system, comprising of the TamA-TamB complex in proteobacteria, interacting with BamA (Beta-barrel assembly machinery protein A) in bacterial diderms where TamA is absent.

Structure-Function Investigation of Membrane Proteins using Nanodiscs

Robbins Puthenveetil, Ph.D.

M.Sc., Nirma University, 2006

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Structure-Function Investigation of Membrane Proteins using Nanodiscs

Presented by

Robbins Puthenveetil, M.Sc.

Major Advisor

_____

Dr. Olga Vinogradova, Ph.D.

Associate Advisor

_____

Dr. Nathan Alder, Ph.D.

Associate Advisor

_____

Dr. Andrew Wiemer, PhD.

University of Connecticut
2017

Acknowledgements

It is humbling to use the accomplishments of a thriving scientific community as a stepping stone to further my own progress in the field. Of course a diverse study such as this one comes with its own triumphs and tribulations. I sincerely thank the collaborative efforts of several people, who have helped me through my journey.

First and foremost, I would like to thank my advisor Dr. Olga Vinogradova for allowing me to be a part of her team and offering me the freedom to work on my own ideas. Her hands-free approach has helped me become an independent thinker. I will be taking with me, her child like curiosity, and her fearless foray into new fields. I truly appreciate her willingness to explore ideas that may not always appear feasible. I will always fondly remember the innumerable get togethers with her family members.

I would like to thank my committee members, Dr. Nathan Alder, for introducing the concept of nanodiscs which became the focus of my doctoral work; Dr. Arlene Albert, for her expertise and guidance in the field; Dr. Andrew Wiemer, for his extensive discussions, patience and collaboration on his project; Dr. Amy Anderson and Dr. Debra Kendall. I will remain indebted to Dr. Kendall for stepping in for Dr. Anderson, who unfortunately passed away. Dr. Anderson was an excellent researcher, mentor and pillar of the scientific community here at UConn. Her humility, intellectual prowess and willingness to help others made her a true inspiration.

Academic research is conducted in a confined environment, and the people in that environment contribute immensely to the success of the undertaken work. Thank you lab members, Xiaochen, Priya, Khiem and Lalit for all the troubleshooting suggestions,

discussions and pranks. I would specially like to thank Dr. Vitaliy Gorbatyuk, whom I have always looked up to as a father figure. He has taught me everything there is to do with NMR and guided me through every experiment. I truly admire his constant patience and support. I would also like to thank Oksana Gorbatyuk for help with troubleshooting certain protocols and for delighting us with her decadent “banana cake” on more than one occasion.

Special thanks is due to Dr. Justin Radolf, for allowing me to work in his lab for a year, and helping me expand my skillset. His mentorship and guidance has helped me gain confidence while working in the field of microbiology. A special thanks is also due to his lab members Sanjiv, Arvind, Abhishek, Ashley, Morgan, Carson and Dr. Melissa Caimano for welcoming me to their lab and making me feel at home.

I would like to thank the Dept. of Molecular and Cell Biology for providing teaching assistantships throughout my entire Ph.D. career. Thanks is due to Dr. Linda Guilani, Maryke and Jane for their hard work and assistance making my experience as a teaching assistant very pleasant. Special thanks to Anne St. Onge, for her help with the completion of all my official documentation and requirements.

I would like to thank my family members for having unstinting belief in my abilities, and for their unending support and prayers. Finally, I would like to thank all my friends, the bajrangi’s: Sharad, Vineet, Suarab, Abhinav, Gaurav and Ujjwal. I would especially like to thank Bushra Husain for being there for me through the thick and thin of times. Apurba, Ananya, Rajiv and Sambhavi for all the outdoor excursions and trips. Above all, I would like to thank the grace of GOD for all the opportunities I have had in life.

Table of Contents

LIST OF FIGURES

CHAPTERS

Chapter 1

Chapter 2

Chapter 3

Chapter 4

Chapter 5

APPENDIX CHAPTERS

Chapter A1

Chapter A2

Chapter A3

Chapter 1

Introduction

Adapted from:

Nanodiscs and Solution NMR: preparation, application and challenges.
Puthenveetil, R., Nguyen, K., and Olga Vinogradova, (2017).
Nanotechnology Reviews 6 (1), 111-125

General overview

Membrane proteins act like a conduit communicating the cell cytoplasm with its external environment. They make up for roughly 30% of the eukaryotic genome [1] and 50% of all drug targets [2]. Membrane proteins have long been the focal point of both academic and pharmaceutical research. Not surprisingly, their dysregulation, misfolding and/or mutations have been associated with numerous diseases [3, 4]. Despite their importance and relative abundance, very few have been structurally and functionally characterized, which reflects through the poor representation (less than 3%) of their available structures in the protein database. Membrane proteins are notorious for the difficulties associated with their overexpression, purification, low yield and stability. The inability to obtain protein samples which are stable, pure and in large quantities for X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo electron microscopy (cryo-EM) greatly hampers the structural elucidation of these elusive proteins. Initial success in isolating membrane proteins was offered through the use of detergents, which helped in the biochemical investigation of several proteins. Detergents, however, do not provide an ideal solvent condition mimicking a “native like” environment for membrane proteins [5] and may not be optimal for studying signaling across the membrane as they may lead to the unfolding of the soluble interacting partners. Thus, there stands a need for a proper membrane mimetic that stably houses membrane proteins and makes them amenable to in-vitro investigations.

Solution NMR has been successfully used to study membrane proteins solubilized in different membrane mimicking systems, including organic solvent mixtures, amphipols, micelles, and bicelles [6, 7]. These media, though useful, present themselves with a

number of caveats, including surface curvature artifacts, limited diversity of detergent or lipid molecules, heterogeneity of the sample preparation and a debilitating inability to study any interaction with their soluble binding partners. These problems were remedied by the introduction of nanodiscs, a class of soluble membrane mimetic which provided a stable lipid bilayer system quite close to a native membrane environment. The success of nanodisc has been vividly exemplified through the biophysical characterization of several receptors, enzymes, channels, and transporters [8, 9]. In this chapter, we discuss the application of nanodiscs to solution NMR. We also provide interesting strategies for incorporating membrane proteins into nanodisc along with discussing other potential alternative nanoscale phospholipid bilayer systems.

What are Nanodiscs?

Nanodiscs are discoidal nanoscale lipoprotein complexes that are composed of a phospholipid bilayer held together by two anti-parallel strips of amphipathic helical Membrane Scaffold Protein (MSP) (Figure 1A). Nanodiscs were originally developed by Sligar and colleagues in the late nineties [10-13]. MSP is derived from Apolipoprotein A-1 (apoA-1), which normally functions to transport cholesterol from the tissue cells to the liver for excretion. The concept of a disc shaped structure came from the very nature of Apolipoprotein A-1 which forms nascent discoidal particles by utilizing phospholipids from the serum. These phospholipid particles later on ingests cholesterol from the tissue cells with the help of transporter ABCA1, growing from a discoidal to a spherical form and eventually ending up in the liver [14]. Full length apoA-1 is a helical protein with an N-terminal four helix bundle and two C-terminal helices (PDB: 2A01). An N-terminal deletion

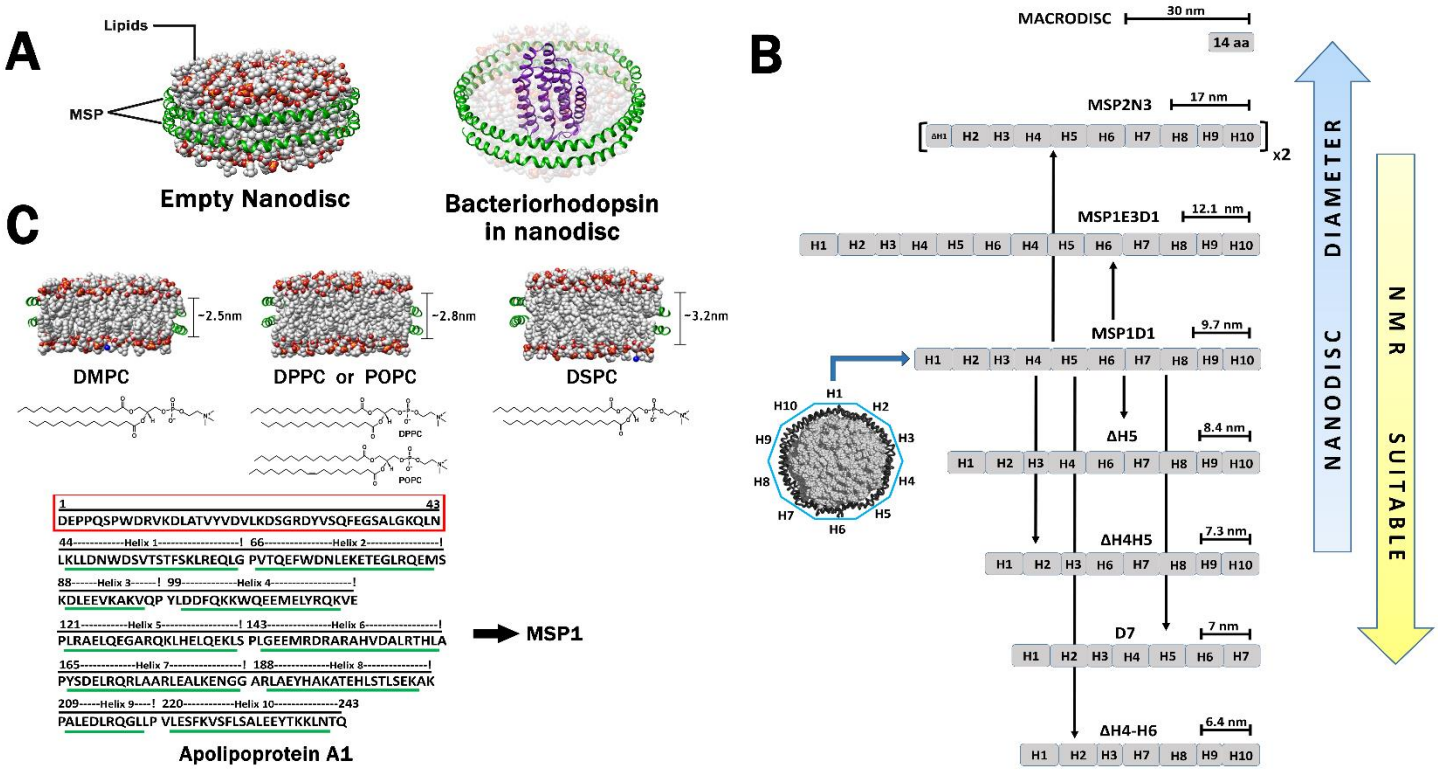


Figure 1. Nanodiscs models were constructed through the nanodisc builder from CHARMM GUI [15]. (A) Visual representation of a nanodisc that is either empty or containing bacteriorhodopsin (PDB: 1R2N). The outer belt protein MSP is shown in green while the lipid molecules are colored by atom type, where carbon is grey, oxygen is red, phosphorus is orange, and nitrogen is blue. The lipids are rendered partially transparent in order to distinguish bacteriorhodopsin from nanodisc. (B) Schematic representation of the overall length and nanodisc diameters for several MSP1D1 variants. The arrow indicates the scheme for addition or deletion of helices from the core MSP1D1 leading to MSP with varying lengths. Nanodiscs with smaller diameters are preferred for NMR studies. The largest "MACRODISC" is obtained from a singular 14 amino acid peptide producing a disc of 30nm in diameter while the smallest disc, with a diameter of 6.4 nm, is obtained by deleting of helices 4 through 6 from the MSP1D1 construct. (C) Depiction of the approximate hydrocarbon thickness of nanodisc bilayers composed of commonly used phosphatidylcholines. As shown, DMPC is the smallest, DPPC and POPC have similar thickness, and DSPC is the largest [16]. The chemical structure of each lipid is presented underneath (Top Panel). The complete sequence of the native protein, Apolipoprotein A1, is shown as well as each individual helices. MSP1 construct, represented by the green lines, is created by deleting the N-terminal region illustrated in the red box (Bottom Panel).

construct apoA-1 (Δ1-43) yielded a structure that was arranged in a circular belt of amphipathic helices (PDB: 1AV1) [17], which lead to the advent of the present nanodisc technology. This deletion construct consisting of ten alpha helical repeats was referred to as MSP1, which (Figure 1C: Bottom panel) formed a belt encompassing lipid molecules. Another larger construct MSP2 was designed by joining two MSP1 molecules through a stable linker. MSP1, MSP2 [11] and other variants are broadly referred to as MSP. The variants are denoted by a letter and a number that follows the parent MSP1 or MSP2 from which they are derived through either deletions or insertions of helices. The presence of a protein belt constrains the dimensions of the bilayer, trapping lipid molecules within its central core. This way, the system ensures a monodisperse size distribution. MSP variants differ in the size of their total length resulting in discs of different diameters. MSP with varying length in permutation with different types of lipid molecules, including DMPC (dimyristoyl-phosphocholine), DMPG (dimyristoyl-phosphorylglycerol), POPC (palmitoyl-oleoyl-phosphocholine), DPPC (dipalmitoyl-phosphocholine) and *E. coli* lipid extracts, etc. [18], offer a wide variety of customizable discs catering to the stability requirements of different membrane proteins. Lipid molecules differing in their acyl chain lengths provide a bilayer with altered thickness (Figure 1C: Top panel) which may be required for membrane proteins that either differ in their expanse of the bilayer region or undergo relative movements in the transmembrane region as a result of their activation. Nanodiscs provide a competitive edge over other systems due to their soluble nature, ease of concentration, monodispersity, temperature stability, and compatibility with cell

free expression system offering themselves as a more viable alternative for investigation of membrane proteins.

Size variations in nanodisc

NMR is a powerful high resolution technique that provides molecular details delineating both the structural and dynamic properties of a protein. The most prominant advantage lies in its ability to investigate the protein in solution conditions closely resembling a native environment, which is in sharp contrast to X-ray crystallography. However, the single most important limitation resides with the overall size of the protein. Previously, the size limitation for NMR susceptibility was capped at 50 kDa [19], but with multiple technological advances, such as TROSY [20]-based triple-resonance experiments at high magnetic fields strengths combined with extensive deuteration and innovative isotopic labeling schemes in sample preparations [21, 22], this size has been pushed to 100 kDa. Even with these advancements, it still remains challenging to study larger complexes at higher resolution.

1. Small discs

The original MSP construct, MSP1, that contained ten amphipathic helical repeats was modified at its first helix through partial deletion and further addition of a Histidine-tag to obtain the MSP1D1 construct [23], which formed nanodisc with a diameter of ~10 nm and MW of ~150 kDa. MSP1D1 became the base construct from which most of the variant discs were generated (Figure 1B). For NMR applications, a larger disc size is unfavorable as its transverse relaxation rate is faster, leading to peak broadening and lower resolution. The problem was ameliorated by creating truncated constructs of MSP,

which yielded smaller discs. Sligar's group had previously developed MSP1E3D1 [23] which are larger discs of ~13 nm diameter, obtained through additional insertion of helices 4-6 (Figure 1B). This proved that repetition of helices in the middle region of the protein does not affect its property to form discs, which also indirectly meant that the middle region is amenable to alterations without impacting the formation of disc. Wagner's group adopted this idea and successively removed the same three helices 4 through 6 obtaining smaller nanodiscs. The smallest obtained disc ΔH4-H6 was of ~6.4 nm, but was unstable and formed larger assemblies of ~11 nm over time. The next smaller disc, which was stable enough to retain its small size, was from the MSP construct ΔH4H5 with a diameter of ~7.3 nm and MW of ~70 kDa (Figure 2A) [24].

At the same time, our group independently arrived at a similar conclusion through a different scheme of deletion constructs. We successively removed individual helix from the C-terminal end and obtained the smallest disc from the deletion of the last three helices ΔH8-H10, which we refer to as D7 discs, since this deletion construct contained the first seven helices. D7 disc have a diameter of ~ 7nm with a molecular weight of ~ 62 kDa (Figure 2A). In the MSP1D1 construct, all helices are roughly of similar lengths except for helix 3 and 9 which are half the standard length [Figure 2B: top panel (MSP1D1 or D10)]. Both D7 and ΔH4-H6 constructs were devoid of three helices, with our deletion ΔH8-H10 containing one short helix along with two standard helices while Wagner's deletion removed three standard length helices. This translates into our D7 construct being slightly larger than the ΔH4-H6 construct though both have deletions by three helices. Interestingly, this was quite accurately corroborated in the distribution of disc sizes where our D7 (ΔH8-H10) disc size of ~7 nm falls right in between Wagner's discs

ΔH4H5 of ~7.3 nm and ΔH4-H6 of ~6.4 nm. Furthermore, we found that any successive reduction beyond seven helices lead to the formation of discs with larger diameters. For example, our D5 construct, which contains the first five helices, produced a size similar to that of the full length MSP1D1 (D10) construct with ten helices. This can be conceptualized as four D5 molecules coming together as two dimers of D5 form either half of the circular belt. These rings present an edge to edge arrangement mimicking a single ring of 10 helices. The most intriguing finding from our D7 construct was that the terminal helices do not regulate the formation of nanodiscs, which we believe might have been the initial concern that led to the addition or deletion of helices exclusive to the middle region. A point to be noted here is that though Wagner's ΔH4H5 gave the smallest disc, the ΔH5 construct with a slightly larger diameter of ~8.4 nm (Figure 2A: Left Panel) provided the best NMR spectrum and was adopted for structure determination.

Once we reached the smallest possible size through the deletion constructs, we still argued if we could possibly break the size barrier and force the system to even smaller diameters. A closer look at the protein sequence of MSP1D1 showed the strategic presence of proline residues present in between helices (Figure 2B: Top Panel). We hypothesize whether the inter-helical turns are an outcome of proline residues. Additionally, we observed that most helices had a standard length of ~ 22 amino acids (aa) with the smallest helix being 11 aa long. The natural presence of smaller helices led us to wonder whether reducing the helix length would facilitate further size reduction. Using all of this information, we attempted to redesign our D5 MSP construct, which previously gave us larger discs. D5 contains five helices with one 11 aa and four 22 aa blocks. Our idea was to break down D5 into blocks of 11 aa each, such that it gives us a

total of 9 helices roughly resembling the natural 10 helix arrangement in MSP1D1. We achieved this by adding proline residues breaking the 22 aa bocks down to 11 aa and left the naturally occurring 11 aa block untouched (Figure 2B: Bottom Left Panel). Unfortunately, this approach was unsuccessful as can be seen from the SEC data (Figure 2B: Bottom Right Panel), where the elution volumes of D5 and D5P (proline mutant) are nearly identical, indicating a similar stokes radius and hence similar diameters. In effect, we were unable to decrease the size of nanodisc any further through selective placement of additional prolines. While we tried several permutations of prolines insertion, we only present the case with maximum number of inserted prolines for the sake of simplicity.

2. Large discs

In regard to size, the other end of the spectrum has also been worked out through the development of discs with larger diameters. Sligar's lab had successfully developed large discs by fusing two MSP1 molecules together to obtain discs of 17 nm diameter, which were named MSP2N3 [25]. This is most likely useful for incorporating large macromolecular complexes. A separate use of larger nanodisc in NMR has been demonstrated by Opella's group. They developed a 30 nm MACRODISC using multiple copies of an amphipathic 14-residue peptide, which formed a discoidal bilayer particle when mixed with phospholipids at a certain ratio (Figure 2A: Right Panel). The authors found that increasing the discs diameter slows their reorientation rate and generates sufficient magnetic susceptibility anisotropy, thus allowing them to align in an external magnetic field [26]. These discs can be used for residual dipolar coupling measurements [27] by providing partial alignment of macromolecules. A summary of all discs relevant to

the above discussion are presented in Figure 1B as a quick comparison of their diameters and suitability to solution NMR applications.

3. Disc thickness

Another key characteristic for consideration is the thickness of the lipid bilayer, which is defined by the length of the aliphatic chain from the participating lipid molecules. It is important to have a bilayer thickness appropriate enough to accommodate the entire hydrophobic length of a membrane protein. Mismatches may occur when the hydrophobic thickness is lower or higher than the lipid bilayer. This anomaly may cause curvature and/or disorder of the bilayer in close proximity to the protein's core surface, leading to distortions in the lipid molecules around the hydrophobic area [28]. In few certain cases, this could be favorable for studying the folding of β-barrel protein [29]. In most general cases, the functional states of channels or receptors may be affected through the distortion of their structure [30].

The appropriate thickness required for the protein of interest would have to be experimentally explored by testing various lipid compositions. Nanodiscs are resilient towards accommodating a wide range of lipids, which could differ in their alkyl chain length, degree of saturation, and polarity of head groups. A complex lipid composition might be essential to mimic natural bilayers and maintain the functional activity of particular membrane protein [31].

4. Alternative bilayer systems

Besides MSP derived nanodiscs, other existing nanoscale phospholipid bilayer systems can serve as potential alternatives. Here we discuss two such systems: SMALPs

(Styrene Maleic Acid Lipid Particles) or Lipodisq ® and Saposin-A lipoprotein disc (Salipro). Of the two, Saposin-A lipoprotein disc might offer a greater potential for NMR applications.

SMALPs or Lipodisq® are discoidal lipid-polymer complexes, where the outer annulus is formed by Styrene-Maleic acid (SMA) copolymer composed of Styrene and Maleic acid in ratios of either 2:1 or 3:1 [32]. These self-assembling water-soluble particles contain about 140 lipids and have a median diameter of ~10 nm for an empty disc (Figure 2C: Bottom Panel). SMA is required in very low concentrations for the formation of lipid particles and is reflected in the requirement of extremely low molar ratios of lipid to polymer. SMALPs have been successfully used with bilayer forming lipids like POPC [33] and DMPC [34]. An intriguing feature of SMA copolymer lies in its ability to directly solubilize membrane proteins from its native membrane, forming nanoscale discs with endogenous lipids surrounding the protein. This ability has been exemplified through the incorporation of mitochondrial respiratory Complex IV, which was shown to retain its enzymatic activity [35]. SMA precipitates at pH values lower than its pKa of 6.5. Thus, NMR experiments involving SMALPs need to be executed within a narrow pH range of 6.8-7.5. Additionally, the use of divalent cations should be omitted as they destabilize the SMA copolymer causing them to dissociate from the lipid molecules. The major drawback of using SMALPs in NMR studies resides in its ability to form heterogeneous disc population as can be visualized from the TEM data referenced in [34]. The presence of a relatively lighter copolymer belt, in comparison to protein nanodiscs, might offer some advantages with respect to its relaxation properties. It still remains to be seen whether SMALPs incorporating a deuterated protein can lead to an encouraging NMR spectrum.

Saposin-A lipoprotein discs (Salipro), a nanoscale disc, where two Saposin-A molecules assemble together to produce a disc like system with lipid molecules in its core. The difference between Saposin-A disc and nanodisc lies in the arrangement of the outer annulus protein surrounding the lipid molecules. For nanodisc, the outer protein forms a continuous belt surrounding the lipid molecules while Saposin-A proteins are present in a head to tail arrangement forming a discontinuous envelope around lipids. The structure of Saposin-A bound to Lauryldimethylamine-N-Oxide (LDAO) solved at 1.9 Å shows that it forms a closed structure in its apo form (detergent free) and becomes extended on binding to LDAO molecules [36]. Two such detergent bound Saposin-A proteins are brought together through the core hydrophobic interaction with the acyl chains, forming a discontinuous structure with a distorted disc shape (Figure 2C: Top Panel). Saposin-A disc has a diameter of ~3.2 nm with a molecular weight of 43 kDa. In total, there are 40 detergent molecules arranged asymmetrically with 24 in the upper and 16 in the lower leaflet of the bilayer as seen in case of LDAO. This system is also compatible with a wide range of lipid molecules. Recently, Saposin-A discs were utilized for cryo-EM study of large membrane protein complexes like archaeal mechanosensitive channel T2 (32.9 kDa), with four predicted transmembrane helices existing as a putative homo pentamer, and bacterial peptide transporter $PepT_{So2}$ (56 kDa), with 14 transmembrane helices existing as a homo tetramer [37].

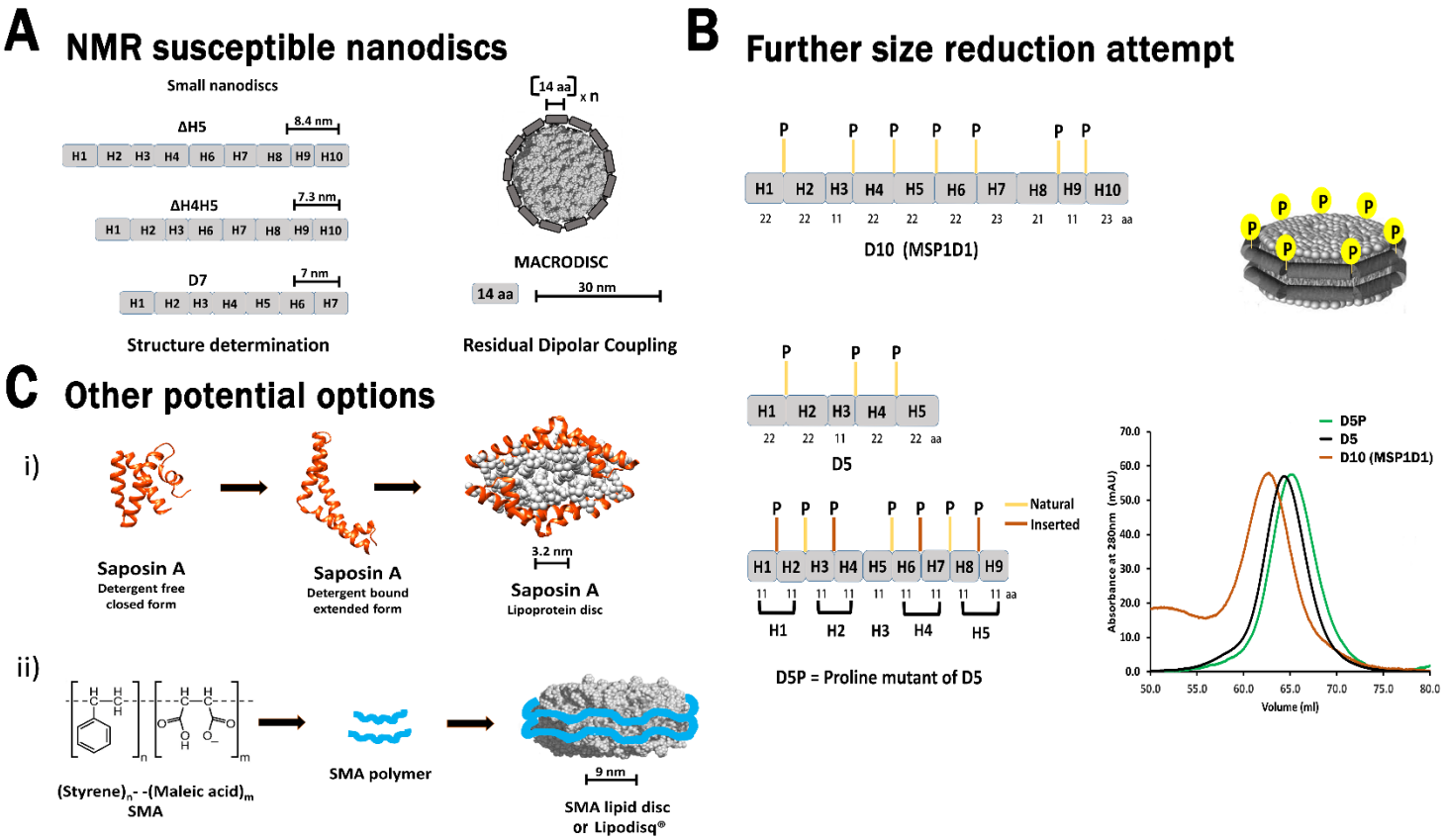


Figure 2. (A) Small nanodiscs that have been employed in solution NMR studies. ΔH5, ΔH4H5 and D7 provide discs with smaller diameters that can be used for structural investigation of encapsulated proteins (Left Panel). Pictorial representation of larger "MACRODISC" is shown where the outer annulus is formed by a 14 amino acid amphipathic peptide. The large 30 nm discs has been used to orient molecules for residual dipolar coupling (RDC) experiments (Right Panel). (B) Schematic representation of MSP1D1/D10 with the positioning of prolines and the number of amino acids present in each helix. Pictorial representation of a nanodisc showing the presence of prolines at helix turns (Top Panel). A scheme for incorporating proline residues further forcing nanodisc to a smaller size is shown. An altered construct D5P, is created from D5 through the strategic placement of additional prolines, increasing the number of individual helices and potentially offering additional turns at proline residues. Each individual helix in the D5P construct is of identical length and is maintained at the smallest size (11 amino acids) that is naturally present in the MSP1D1 construct (Bottom Left Panel). The elution profile from the size exclusion chromatography indicating that the D5P nanodisc (green) is approximately the same size as D5 disc (black). Furthermore, D5 and D5P nanodiscs are not significantly smaller than the original MSP1D1/D10 nanodiscs (brown) (Bottom Right Panel). (C) Alternative nanoscale phospholipid bilayer systems: Saposin-A lipoprotein disc formed by the addition of lipid molecules is shown. Saposin-A in its detergent free form has a closed conformation and becomes extended when bound to lipid molecules (PDB ID: 4DDJ). Saposin-A lipoprotein disc, with a diameter of 3.2 nm, are formed when two Saposin-A proteins are brought together by its lipid core (Top Panel). Styrene-Maleic Acid lipid particle (SMALP) is shown where the synthetic Styrene-Maleic Acid copolymer forms a disc by encapsulating lipids within its central cavity. SMA lipid disc or Lipodisq® have a diameter of 9nm (Bottom Panel).

This paper brings to light certain key features of Saposin-A disc that could be useful for NMR applications. First, Saposin-A system provides a wide range of pH applicability: reactions can be carried out at physiological pH or, as shown in the original paper [36], at acidic pH. Second, it has exceptional temperature stability: Saposin-A discs can withstand freeze-thaw cycles, extending the storage capability of a sample. They are highly thermostable (0-95°C), allowing for NMR experiments to be conducted at higher temperatures as is required for larger systems to improve spectral quality. Third, sample homogeneity: The negatively stained TEM image shows the presence of homogeneous population of Saposin-A disc containing the incorporated protein. However, this system also presents two major concerns: (a) the propensity to allow spontaneous oligomerization of inserted proteins. The cryo-EM paper demonstrates more than two Saposin-A molecules coming together to accommodate a large membrane complex. NMR studies are preferentially carried out on monomeric proteins as oligomeric forms of proteins increase the ambiguity of the results; (b) potential interaction between the incorporated protein and Saposin-A. The presence of low number of lipid molecules might lead to protein-protein interaction between the incorporated protein and Saposin-A. Moreover, there seems to be a gap between the two discontinuous Saposin-A proteins, which may allow for the loss of lipid molecules at higher temperatures, reducing their effective concentration. If these concerns contribute to the final dataset collected from NMR experiments, the resulting structure may not represent its true form in a native environment. Nevertheless, these are conjectures and need to be confirmed (or rejected) through rigorous experimentation. In summary, Saposin-A discs have a significantly smaller size with higher pH variability, thermostability and homogeneous distribution.

Taken together, they might offer an advantage for Saposin-A discs over nanodisc for solution NMR applications. During the preparation of this review, we did not come across any NMR data using Saposin-A disc, but it won't be surprising to see one in the very near future.

Preparation of nanodiscs

For NMR experiments, optimal sample preparation is paramount to getting good data. The following simple but key concepts should be taken into consideration before preparing nanodisc samples.

1. Expression and Purification of MSP

Expression and purification of MSP is similar to any typical His-tag protein. We find "The QIAexpressionist™ (Qiagen, USA) as a good general handbook for His-tag protein purification. MSP in pET28a vector is transformed into *E.coli* BL-21(DE3) (Agilent, USA), grown to an optical density of 0.6, and induced with 1mm IPTG. The induction is continued for 4-6 hours. All expression steps are carried out at 37°C. We have found that overnight expression at room temperature leads to an increase in aggregation and impurities, reducing the overall yield of monomeric MSP. Post harvesting, the cells are purified on Ni-NTA resin (Qiagen, USA) following the protocol referenced in the methods paper [38]. Briefly, the purification process is performed in Tris-NaCl buffer at pH 8 and has four steps (three wash and one elution). The first buffer contains 1% Triton X-100 and the second contains 50mM Sodium Cholate as detergents. The second and third wash buffers each contain 20 and 50 mM Imidazole, respectively. Elution is achieved with 400mM Imidazole.

MSP purification can be carried out both at room temperature and 4°C. A key consideration is to check the pH of the buffers at 4°C, since Tris is temperature sensitive and a change in pH might affect the final yield. We find that the wash steps along with removing impurities also washes out reasonable amount of MSP. This can be mitigated by reducing the salt concentration by half and having a single 20 mM Imidazole wash instead of 50 mM in buffer three. Imidazole is removed from the final elution and buffer exchanged through overnight dialysis in Tris-NaCl buffer at pH 7.5. For NMR applications, we perform additional Size Exclusion Chromatography (SEC) on a Superdex S-75 (GE, USA) column and discard the tailing end of the peak that contains degraded MSP. This helps us get rid of MSP molecules smaller than the optimal D7 construct which may form larger discs, increasing the heterogeneity of the sample. A further elaboration on this rationale can be found in the section (Size variation in nanodiscs). After completion of the purification process, MSP is concentrated to roughly 4 mg/ml and stored at 4°C. For long term storage, lyophilization has been recommended [38] .

2. Incorporation into nanodisc

2.1. Preparation of reagents

The primary reagents consists of two Tris-NaCl buffers with 20 mM Tris-HCl and 100 mM NaCl at pH 7.5. The second buffer, in addition, contains 60 mM Sodium Cholate detergent. NaCl concentration in the above buffers can be significantly reduced for membrane proteins that are sensitive to salt. The lipid stocks used for nanodisc preparations comes either in a soluble form, dissolved in chloroform or as solid powders (Avanti polar lipids, USA). Chloroform is removed from the solubilized lipid stock using a stream of nitrogen gas and later left in a vacuum desiccator to eliminate any leftover

residual amounts. Alternatively, one could use the powdered stock that can be directly weighted in appropriate quantities for disc preparation. Though we haven't directly compared the soluble and the powder form, we routinely use the powdered form which has worked well without any drawbacks. Protein incorporation into disc can be carried out at different temperatures which may improve their stability or reduce protein degradation. Lipid molecules that differ in their phase transition temperatures viz., POPC at ~ 4°C, DMPC at ~ 25°C, and DPPC at ~ 37°C [39] are chosen based on the desirable operating temperature required for disc preparation. The most popular utilization has been with using DMPC at room temperature. Alternatively, one could also prepare or use lipid extracts from multiple sources to ensure proper stability of specific membrane proteins.

2.2. Preparation of nanodiscs

Lipid molecules are reconstituted in the cholate containing buffer. The amount of lipids required depends on the type of MSP molecule that is utilized. Generally, the reconstituted lipid molecules are added in molar excess as follows: 100-120 molar excess for MSP1E3D1, 80 molar excess for MSP1D1, 50 molar excess for ΔH5 and 20 molar excess for the smaller constructs, D7 or ΔH4H5. A typical reaction mixture consists of, MSP and lipid molecules for preparing empty discs; MSP, lipids, and target membrane protein for protein incorporated discs. After the addition of MSP and lipids in the desired molar ratios, the volume is made up by the aforementioned two buffers such that the final concentration of cholate in the reaction volume is between 12-40 mM. This reaction mixture is kept on a rocker for an hour. Meanwhile, pretreatment of Bio-beads SM-2 (Bio-Rad, USA), which functions to pull out the detergent and initiate nanodisc formation, are carried out. The beads are washed with Methanol (1x), Water (3x) and buffer one (1x)

from the above two buffers. Bio-beads are added to the reaction mixture in a ratio of 0.5 to 1.5 grams per ml of volume and left rocking at room temperature, such that the beads remain gently suspended in the solution, for 3-4 hours or overnight for samples with high amount of total detergents. Later, the sample is separated from beads, concentrated, and run on an SEC column to obtain purified nanodisc. For incorporating protein into nanodiscs, the reaction mixture typically contains MSP at a concentration that is twice in molar excess of the target protein. A protein embeds itself into disc by occupying the volume of the lipid molecules that it displaces. Hence, the addition of lipids can be done at a reduced molar ratio than what is required for making empty disc. Empirical optimization to ensure maximum incorporation of the target protein into nanodisc is achieved by monitoring peaks in the void volume arising from the aggregated or unincorporated membrane protein, and at the position of free MSP, implicating inefficient disc formation. Both can be overcome by varying the concentration of lipids, MSP, or target protein.

2.3 On-column nanodisc formation

As with any project, a typical NMR sample is prepared in several iterations to obtain the best NMR spectrum. Preparing nanodisc samples for NMR applications are time intensive and requires multiple purifications to separate empty discs from the incorporated ones. We developed a quick, alternative, on-column method where the protein incorporated disc was formed on the resin either through the bound MSP (approach-1) or the bound target protein (approach-2) [40]. The advantages to this method were several-fold, including reduced preparation time, concomitant separation of empty and incorporated discs and better yields, which collectively improved the quality of our NMR

spectrum. For on-column preparation, the required reagents remain the same as in the original protocol while we explore the differential permutation of affinity tags to achieve our goals. Briefly, for approach-1, the harvested cells containing over expressed His-tagged MSP is bound to Ni-NTA resin in a gravity column and washed as per the regular protocol. Prior to elution, the resin is mixed with lipids and target protein in the same column for an hour before the addition of Bio-beads. After passage of the stipulated time, discs are formed on the resin attached through MSP (shown as inset in Figure 3A for approach- 1).Once the reaction is over, elution is carried out in the same elution buffer previously used for the purification of MSP. The eluate contains a mixture of empty and protein incorporated discs, further separated through chromatography. For approach-2, we took the advantage of the MBP (Maltose binding protein) tag by engineering the transmembrane and cytoplasmic tail regions of Integrin-α_{IIb} fused to MBP (in pMAL-C2 vector) with a TEV cut site present between the tag and protein. MBP helps solubilize the otherwise insoluble transmembrane region of Integrin and the TEV cut site helps us in the removal of the MBP tag after thier incorporation into discs using TEV protease. Alternatively, other tags could be used in combination with a TEV cut site. The MBP-fused construct was expressed in *E.coli* BL-21 (DE3) similarly to MSP. Harvested cells were lysed and mixed with Amylose resin (NEB, USA), which is later washed with Tris-NaCl buffer containing 20 mM Tris-HCl and 100 mM NaCl at pH 7.5. Amylose resin retains the MBP fused protein through its affinity for MBP. MSP and lipids are added to the resin, followed by the addition of Bio-beads. Discs are formed around the MBP fused protein attached to the Amylose resin (shown as inset in Figure 3A for approach-2). After the formation of discs, the resin is rinsed with Tris-NaCl buffer, which washes out the empty

discs leaving behind the protein incorporated ones. Elution is carried out in Tris-NaCl buffer with 20mM maltose and the eluate is further subjected to SEC for buffer exchange and removal of unincorporated Integrin. For NMR sample, elution was carried out in the presence of TEV to help remove the MBP tag. Approach-2 is faster and more efficient of the two approaches. A pictorial representation summarizing the different approaches along with their relative time scale is shown in Figure 3A. There are two caveats to the on-column method. (i) Estimation of the protein concentrations: we usually express MSP/target protein in large volumes of LB and aliquot 1L harvests in falcon tubes. From one such tube, we estimate the protein concentration and use that to decide the MSP to target protein or vice versa ratios. We admit that this may not be accurate, but the convenience of sampling through several proteins grossly outweighs the effects of inaccuracy arising from minor discrepancies in protein estimations. This problem is usually alleviated by the addition of MSP slightly in excess to its required concentration. (ii) Retrieval of chromatography resin after mixing with Bio-beads: a simple solution to this problem is to use magnetic beads in lieu of agarose beads, which can be easily recovered from the solution mixture using any of the commercial magnetic apparatus. Alternatively, a net filter can be utilized with a pore size large enough for the resin but small enough for the beads, since both differ in size by several orders of magnitude.

An extension of approach-2, though not useful for NMR applications lies in the use of antibodies to pull larger protein receptors/complexes from any cell lysate. This strategy may find wide applicability among general and pharmaceutical researchers. Antibody based pull down is a routinely used well-established technique. We propose to utilize antibodies in conjunction with our on-column method by immobilizing antibodies through

covalent crosslinking to the resin bead using Crosslink IP/Crosslink Magnetic IP Kit (Pierce, USA). The crosslinked antibodies (specific to the protein or a small tag) can then be used to fish out native receptors expressed in a mammalian cell line. This can be further washed and mixed with MSP and lipid molecules. Nanodiscs with the incorporated pulled receptor can be obtained either with the appropriate elution buffer or through TEV protease cutting at a TEV site engineered between the tag and the protein. A schematic representation, comparing the antibody versus the MBP tag using approach-2, is depicted in Figure 3B.

3. Considerations for NMR sample preparation

Deuteration is the sole key towards successful implementation of nanodiscs in solution NMR. The target receptor protein needs to be expressed in M9 Minimal media supplemented with [U-99%^{15}N] NH_4Cl and [U-99% ^{2}H, ^{13}C] Glucose in [99%] D_2O (Cambridge Isotope Labs, USA). Nanodiscs should be prepared with deuterated lipids [eg. 1,2-dimyristoyl-d54-sn-glycero-3-phosphocholine; d54-DMPC (Avanti Polar, USA)] using any of the methods explained above with one exception. Typically, MSP is added twice in molar excess of the target protein. In case of NMR sample preparation, MSP should be added three–four times in molar excess of the target protein since the aim is to get monomeric incorporation of the target protein into nanodisc. The empty discs are successively removed thorough the employment of different chromatographic techniques.

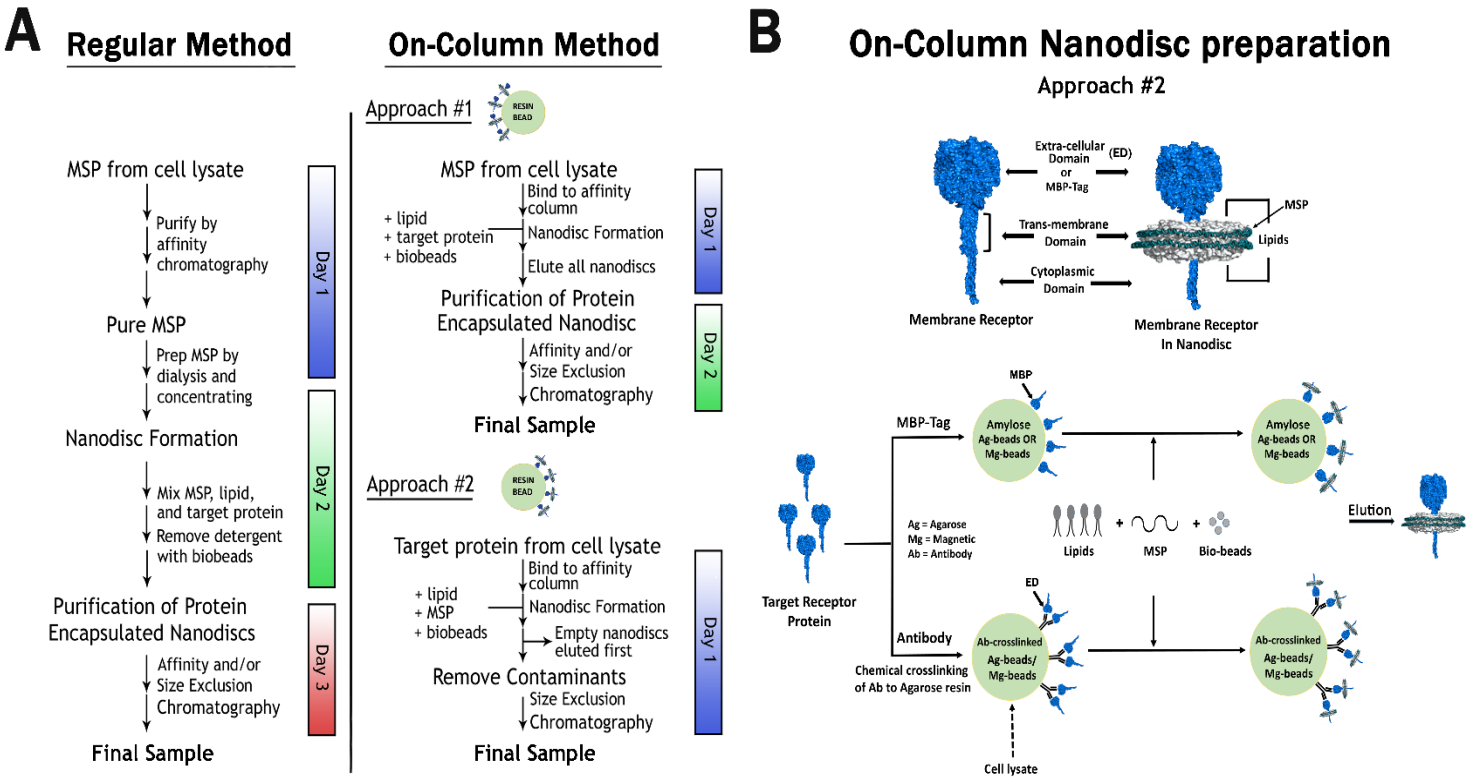


Figure 3. General workflow for nanodisc preparation. (A) Comparison of the regular method to the on-column method along with the approximately required time scales. The on-column method is split into two approaches, depending on the immobilization of either MSP or the target protein onto the resin. (B) Schematic representation of a target protein with its corresponding soluble Extra cellular/MBP tag domain, a transmembrane hydrophobic domain, and soluble cytoplasmic domain. Also shown is a nanodisc containing the encapsulated target protein with MSP and lipid molecules (Top Panel). A more detailed workflow for approach #2 from the on-column method is illustrated. The target protein is bound to the resin either through its affinity for the MBP tag or through the crosslinked antibody specific to the extracellular domain (ED). The addition of MSP, lipids, and bio-beads lead to the encapsulation of the target protein into nanodisc while it is still bound to the resin. This nanodisc encapsulated protein is finally eluted from the resin through their respective elution buffers (Bottom Panel).

Technical Challenges and possible solutions associated with the NMR application of nanodiscs

Despite the advantage of using nanodisc, there are certain challenges that needs to be overcome. We categorize them into: a) preparation of protein incorporated nanodisc, and b) uncertainties with experimental setups. We have previously discussed potential solution to the former in the section “Preparation of nanodiscs”. We now address some strategies to overcome the latter.

Smaller discs are definitely more advantageous than larger ones. One can easily employ the small nanodiscs developed specifically for this purpose either through our (D7) or Wagner’s (ΔH4H5 or ΔH5) constructs. If the plasmid for MSP1D1 is readily available, D7 construct can be easily generated by placing a stop codon at the end of the seventh helix using site directed mutagenesis. The ~7 nm diameter discs reflects sufficient lateral space to incorporate smaller porins, such as 8-stranded OmpX [24], or helical membrane proteins with at least two or three transmembrane helices, as has been demonstrated above through α_{IIb} homo-dimers and β_3 homo-trimers. To make assignments feasible, complete deuteration of the system is required along with the use of deuterated lipids to minimize dipolar-dipolar relaxations. NMR experiments need to be carried out on high field magnets like 800 MHz or above to take full advantage of the TROSY-based experiments for complete backbone assignments [41]. Furthermore, novel pulse sequences specifically designed for Ile/Leu/Val side-chain assignments in perdeuterated samples with selectively protonated methyl groups can be used [22, 42]. These assigned methyls, in addition to the backbone amides utilized for the structural

characterization of β-barrel porins [24], are absolutely necessary for the high resolution structure determination of α–helical proteins as the number of long distance restraints obtained exclusively from HN-HN would be insufficient. Interestingly, chemicals shifts of membrane protein acquired in other membrane mimetic systems, such as micelles, bicelles or amphipols, often overlap (or are similar) with corresponding chemical shifts in nanodisc, especially for the residues located within the hydrophobic core, and could potentially be utilized to expedite the assignment process [24, 43].

Nanodiscs are especially advantageous in their ability to achieve high concentrations in the high μM to low mM ranges. A sample with higher concentration drastically increases the signal to noise ratio and provides better spectrum quality. Nevertheless, proper care should be taken to maintain the homogeneity of these concentrated samples. We have shown that smaller constructs of MSP form larger discs [40]. This suggests that the purity of your MSP preparation is very crucial. Impure preparations with degraded MSP might lead to a heterogeneous population consisting of a mixture of small and large discs. This will effectively reduce the concentration of small discs, compromising sensitivity. An easy way to check this is through negative stained TEM (Transmission Electron Microscopy) images, which clearly displays sample heterogeneity. Moreover, if your sample contains a beta barrel porin, you can check the number of pores present in the population of your disc which will manifest itself as black dots within your disc's image. This will help ascertain the oligomeric state of a porin within a disc. As an example, demonstrating the visualization of pores in nanodiscs, we previously carried out image analysis of a trimeric beta barrel protein TprC (C-Terminal

domain) in large MSP1E3D1 discs. The image averaged result shows the presence of three pores in the backdrop of a nanodisc [44].

Large macromolecular systems, including nanodiscs, are notorious for peak broadenings in most of the triple-resonance NMR experiments, resulting in a low signal–to-noise ratio. Even though nanodisc with incorporated membrane protein can be easily concentrated, the problem of sensitivity still remains. This can be partially alleviated by data acquisition at higher temperatures with longer time durations. The downside of higher temperatures lies in the detrimental effect on protein stability. Another approach to increase the sensitivity (and resolution if necessary) without significantly raising acquisition times is through the utilization of non-uniform sampling (NUS) [45, 46]. NUS has recently emerged as a powerful techniques to speed up the acquisition of multidimensional NMR experiments. It is based on the rationale where a subset of the linearly sampled data in the Nyquist grid is measured. For this reason, NUS requires processing algorithms different from the normal Fourier transform. A number of sophisticated methods have been developed for reconstructing sparsely sampled NMR data with minimal artifacts. NMRPipe (http://www.nmrscience.com/nmrpipe.html), a UNIX-based collection of programs and scripts developed by Frank Delaglio [47], provides tools for reconstruction of NUS data, including Iterative Shrinkage Thresholding (IST) decomposition and options for Maximum Entropy Reconstruction (MEM). The NMRPipe compatible scripts for istHMS, developed by Wagner's group, could be found at their website: http://gwagner.med.harvard.edu/intranet/istHMS/index.html. The IST software itself may be requested from the same group. MEM processing script generator, developed by Jeff Hoch and coworkers is available as part of RNMRTK

(http://rnmrtk.uchc.edu/rnmrtk/RNMRTK.html) and can be found at the website: http://sbtools.uchc.edu/nmr/nmr_toolkit/.

References

1. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* Journal of molecular biology, 2001. **305**(3): p. 567-80.
2. Drews, J., *Drug discovery: a historical perspective.* Science, 2000. **287**(5460): p. 1960-4.
3. Myers, J.K., L.A. Beihoffer, and C.R. Sanders, *Phenotology of disease-linked proteins.* Hum Mutat, 2005. **25**(1): p. 90-7.
4. Stenson, P.D., et al., *The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine.* Human genetics, 2014. **133**(1): p. 1-9.
5. Yang, Z., et al., *Membrane protein stability can be compromised by detergent interactions with the extramembranous soluble domains.* Protein science : a publication of the Protein Society, 2014. **23**(6): p. 769-89.
6. Sanders, C.R. and K. Oxenoid, *Customizing model membranes and samples for NMR spectroscopic studies of complex membrane proteins.* Biochimica et biophysica acta, 2000. **1508**(1-2): p. 129-45.
7. Liang, B. and L.K. Tamm, *NMR as a tool to investigate the structure, dynamics and function of membrane proteins.* Nature structural & molecular biology, 2016. **23**(6): p. 468-74.
8. Denisov, I.G. and S.G. Sligar, *Nanodiscs for structural and functional studies of membrane proteins.* Nature structural & molecular biology, 2016. **23**(6): p. 481-6.
9. Viegas, A., T. Viennet, and M. Etzkorn, *The power, pitfalls and potential of the nanodisc system for NMR-based studies.* Biological chemistry, 2016.
10. Carlson, J.W., A. Jonas, and S.G. Sligar, *Imaging and manipulation of high-density lipoproteins.* Biophysical journal, 1997. **73**(3): p. 1184-9.
11. Bayburt, T.H., Y.V. Grinkova, and S.G. Sligar, *Self-Assembly of Discoidal Phospholipid Bilayer Nanoparticles with MembraneScaffold Proteins.* Nano letters, 2002. **2**(8): p. 853-856.
12. Sligar, S.G., *Finding a single-molecule solution for membrane proteins.* Biochemical and biophysical research communications, 2003. **312**(1): p. 115-9.
13. Bayburt, T.H. and S.G. Sligar, *Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers.* Protein science : a publication of the Protein Society, 2003. **12**(11): p. 2476-81.
14. Ohashi, R., et al., *Reverse cholesterol transport and cholesterol efflux in atherosclerosis.* QJM, 2005. **98**(12): p. 845-56.
15. Jo, S., et al., *CHARMM-GUI: a web-based graphical user interface for CHARMM.* Journal of computational chemistry, 2008. **29**(11): p. 1859-65.
16. Kucerka, N., M.P. Nieh, and J. Katsaras, *Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature.* Biochim Biophys Acta, 2011. **1808**(11): p. 2761-71.

17. Nath, A., W.M. Atkins, and S.G. Sligar, *Applications of phospholipid bilayer nanodiscs in the study of membranes and membrane proteins.* Biochemistry, 2007. **46**(8): p. 2059-69.
18. Bayburt, T.H. and S.G. Sligar, *Membrane protein assembly into Nanodiscs.* FEBS Lett, 2010. **584**(9): p. 1721-7.
19. Clore, G.M. and A.M. Gronenborn, *Determining the structures of large proteins and protein complexes by NMR.* Trends Biotechnol, 1998. **16**(1): p. 22-34.
20. Pervushin, K., et al., *Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution.* Proceedings of the National Academy of Sciences of the United States of America, 1997. **94**(23): p. 12366-71.
21. Kay, L.E., *NMR studies of protein structure and dynamics - a look backwards and forwards.* Journal of magnetic resonance, 2011. **213**(2): p. 492-4.
22. Tugarinov, V. and L.E. Kay, *Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods.* Journal of the American Chemical Society, 2003. **125**(45): p. 13868-78.
23. Denisov, I.G., et al., *Directed self-assembly of monodisperse phospholipid bilayer Nanodiscs with controlled size.* Journal of the American Chemical Society, 2004. **126**(11): p. 3477-87.
24. Hagn, F., et al., *Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins.* J Am Chem Soc, 2013.
25. Grinkova, Y.V., I.G. Denisov, and S.G. Sligar, *Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers.* Protein Eng Des Sel, 2010. **23**(11): p. 843-8.
26. Park, S.H., et al., *Nanodiscs versus macrodiscs for NMR of membrane proteins.* Biochemistry, 2011. **50**(42): p. 8983-5.
27. Bax, A., G. Kontaxis, and N. Tjandra, *Dipolar couplings in macromolecular structure determination.* Methods in enzymology, 2001. **339**: p. 127-74.
28. Fernandez, C., et al., *Lipid-protein interactions in DHPC micelles containing the integral membrane protein OmpX investigated by NMR spectroscopy.* Proceedings of the National Academy of Sciences of the United States of America, 2002. **99**(21): p. 13533-7.
29. Tamm, L.K., H. Hong, and B. Liang, *Folding and assembly of beta-barrel membrane proteins.* Biochimica et biophysica acta, 2004. **1666**(1-2): p. 250-63.
30. Lee, A.G., *How lipids affect the activities of integral membrane proteins.* Biochimica et biophysica acta, 2004. **1666**(1-2): p. 62-87.
31. Roos, C., et al., *Characterization of co-translationally formed nanodisc complexes with small multidrug transporters, proteorhodopsin and with the E. coli MraY translocase.* Biochimica et biophysica acta, 2012. **1818**(12): p. 3098-106.
32. Knowles, T.J., et al., *Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer.* Journal of the American Chemical Society, 2009. **131**(22): p. 7484-5.
33. Vargas, C., et al., *Nanoparticle self-assembly in mixtures of phospholipids with styrene/maleic acid copolymers or fluorinated surfactants.* Nanoscale, 2015. **7**(48): p. 20685-96.

34. Orwick, M.C., et al., *Detergent-free formation and physicochemical characterization of nanosized lipid-polymer complexes: Lipodisq.* Angewandte Chemie, 2012. **51**(19): p. 4653-7.
35. Long, A.R., et al., *A detergent-free strategy for the reconstitution of active enzyme complexes from native biological membranes into nanoscale discs.* BMC Biotechnol, 2013. **13**: p. 41.
36. Popovic, K., et al., *Structure of saposin A lipoprotein discs.* Proc Natl Acad Sci U S A, 2012. **109**(8): p. 2908-12.
37. Frauenfeld, J., et al., *A saposin-lipoprotein nanoparticle system for membrane proteins.* Nat Methods, 2016. **13**(4): p. 345-51.
38. Ritchie, T.K., et al., *Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs.* Methods Enzymol, 2009. **464**: p. 211-31.
39. Shaw, A.W., M.A. McLean, and S.G. Sligar, *Phospholipid phase transitions in homogeneous nanometer scale bilayer discs.* FEBS letters, 2004. **556**(1-3): p. 260-4.
40. Puthenveetil, R. and O. Vinogradova, *Optimization of the design and preparation of nanoscale phospholipid bilayers for its application to solution NMR.* Proteins, 2013. **81**(7): p. 1222-31.
41. Yang, D. and L. Kay, *Improved HN-detected triple resonance TROSY-based experiments.* Journal of biomolecular NMR, 1999(13): p. 3-10.
42. Yang, D., et al., *Sequence-specific assignments of methyl groups in high-molecular weight proteins.* Journal of the American Chemical Society, 2004. **126**(12): p. 3710-1.
43. Yu, T.Y., et al., *Solution NMR spectroscopic characterization of human VDAC-2 in detergent micelles and lipid bilayer nanodiscs.* Biochim Biophys Acta, 2012. **1818**(6): p. 1562-9.
44. Anand, A., et al., *Bipartite Topology of Treponema pallidum Repeat Proteins C/D and I: OUTER MEMBRANE INSERTION, TRIMERIZATION, AND PORIN FUNCTION REQUIRE A C-TERMINAL beta-BARREL DOMAIN.* The Journal of biological chemistry, 2015. **290**(19): p. 12313-31.
45. Hyberts, S.G., H. Arthanari, and G. Wagner, *Applications of non-uniform sampling and processing.* Topics in current chemistry, 2012. **316**: p. 125-48.
46. Hoch, J.C., M.W. Maciejewski, and B. Filipovic, *Randomization improves sparse sampling in multidimensional NMR.* Journal of magnetic resonance, 2008. **193**(2): p. 317-20.
47. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* Journal of biomolecular NMR, 1995. **6**(3): p. 277-93.

Chapter 2

On-column method for nanodisc preparation and their optimized design for solution NMR applications

Adapted from:

Optimization of the design and preparation of nanoscale phospholipid bilayers for its application to solution NMR.
Puthenveetil, R. and Olga Vinogradova (2013). Proteins, 81(7):1222-31.

Abstract

Despite arduous efforts and recent technological developments structural investigation of integral membrane proteins remains a challenge. The primary deterrents include difficulties with their expression, low inherent solubility, and problems associated with existing membrane mimicking systems. A relatively new class of membrane mimetics, nanodiscs, is emerging as a promising alternative. Although nanodiscs have been proven successful for several biophysical applications, they yet remain to become the system of preferred choice for structure determination. We have hereby made nanodiscs more suitable for solution NMR applications by reducing the diameter of the self-assembly complex to its potential limit. We achieved a noticeable improvement in the quality of NMR spectra obtained for the transmembrane and cytoplasmic domains of integrin αIIb incorporated into these smaller discs rendering them susceptible for a thorough structural investigation. In addition, we also present an on-column method for a rapid, efficient, single-step preparation of protein incorporated nanodiscs at high concentrations. These discs have been fully characterized by transmission electron microscopy, dynamic light scattering, and differential scanning calorimetry.

Materials and Methods

The buffers listed hereunder are the same as used in the original protocol for nanodiscs preparation [1]: binding buffer: 20 mM $NaH_2PO_4*H_2O$, 1 mM PMSF, pH 7.4; wash buffer-1: 40 mM Tris/HCl, 0.3 M NaCl, 1% Triton X-100, pH 8.0; wash buffer-2: 40 mM Tris/HCl, 0.3 M NaCl, 50 mM Na-cholate, 20 mM imidazole, pH 8.0; wash buffer-3: 40 mM Tris/HCl, 0.3 M NaCl, 50 mM imidazole, pH 8.0; elution buffer: dialysis buffer + 0.4 M imidazole ; dialysis buffer: 20 mM Tris/HCl, 0.1 M NaCl,0.5 mM EDTA, pH 7.4. Additional buffers are detailed within corresponding sections.

Cloning and Expression: From the several available MSP constructs, we chose MSP1E3D1 and MSP1D1, pET28a vectors with N-terminal His-tag and kanamycin resistance, generously provided by Dr. Nathan Alder. MSP1D1 vector was used as template to generate the several deletion mutants, D7, D6 and D5. Mutants were created by successively deleting single helices from the template till the MSP was reduced to five helices using the site directed mutagenesis kit (Agilent, USA). The construct for integrin α_{IIb} (A2b) fused to maltose binding protein (MBP) with ampicillin resistance was gifted by Dr. Jun Qin. Integrin α_{IIb} consisting of a single trans-membrane and the C-terminal cytoplasmic domains (TMCD), 54 amino acids in total, are inserted into a modified pMAL-C2 (NEB, USA) vector containing a 6X-His tag and TEV cutting site in the order of MBP-HIS-TEV- α_{IIb} [2]. All constructs were transformed into *E. coli* BL21 (DE3) cells and induced with 1mM IPTG in LB for four hours. ^{15}N α_{IIb} was prepared by induction with 1mM IPTG in minimal media for six hours.

Purification: We found that 0.5 L of frozen cells gave us an average of 12 to 13 mg of MSP1E3D1. Two liters of culture was lysed and split into different parts giving us 1.65,

3.3, 6.6, 12 and 24 mg protein. 1ml of Ni-NTA (QIAGEN, USA) resin was centrifuged and mixed with each of the above protein concentrations. This gave us five samples labeled as NDMT1, NDMT2, NDMT3, NDMT4, NDMT5 having respectively 1.65, 3.3, 6.6, 12 and 24 mg of protein bound per ml of Ni-NTA resin. In conclusion, we have 5 samples doubling in the concentration of protein bound per ml of resin. This resin was washed according to Sligar's protocol with wash buffers 1, 2 and 3, and additionally with five column volumes of three buffers: (1) 40mM Tris, 200mM NaCl, 50mM Imidazole, pH 8, (2) 40mM Tris, 200mM NaCl, pH 8, and (3) dialysis buffer. MSP mutants D5, D6 and D7 were purified from 0.5 L culture mixed with 1 ml of Ni-NTA resin. Each mutant preparation was washed with the above five buffers and later used for the preparation of empty nanodiscs. Bacteriorhodopsin in purple membrane was generously provided by Dr. Robert Birge and was purified using the protocol described elsewhere [3]. An average of 10 – 11 mg of Integrin α_{IIb} was obtained from 1 liter of culture, which was lysed and mixed overnight with amylose resin (NEB, USA) in a buffer containing 20mM Tris, 0.2M NaCl, 1mM EDTA, pH7.4 and complete protease inhibitor tablet (Roche, USA). This integrin bound resin was washed with 10 column volumes of the same buffer plus additional 5 column volumes of dialysis buffer before being subjected to nanodisc preparation. MSP proteins (MSP1E3D1, and D7) used for the formation of integrin containing nanodiscs were purified using the standard solution protocol [1].

On-column formation of Nanodiscs: Powdered version of DMPC (dimyristoyl-sn-glycero-3-phosphocholine) (Avanti Polar, USA) lipid was used for preparing the several nanodiscs. Nanodiscs were prepared using the "*General Method for On-column Nanodisc Formation*" given hereunder. Empty discs containing MSP1E3D1, was prepared at a

protein to lipid ratio of 1:160; MSP1D1, at a protein to lipid ratio of 1:80 [1]; and D5, D6, D7, at an optimized (data not shown) protein to lipid ratio of 1:30. Bacteriorhodopsin (bR) in purple membrane was mixed overnight with 4% Triton X-100. Discs were prepared by mixing bacteriorhodopsin with MSP (bound at 6.6 mg/ml of Ni-NTA resin) and solubilized DMPC at a MSP1E3D1:bR:DMPC molar ratio of 2:3:160 [4]. Integrin α_{IIb} (A2b) containing nanodisc was prepared by mixing MSP with the resin bound α_{IIb} at a MSP1E3D1:A2b:DMPC molar ratio of 1:0.5:160 and D7:A2b:DMPC molar ratio of 1:0.5:20. Once nanodiscs were formed free lipids and unbound proteins were washed out with 2 column volumes of dialysis buffer. Empty and bacteriorhodopsin containing discs were then eluted with 4 column volumes of elution buffer while α_{IIb} containing discs were eluted with 4 column volumes of 20 mM maltose, 20mM Tris, 0.2M NaCl, 1mM EDTA, pH 7.4. All the eluates were loaded onto 16/60 Superdex 200 column pre-equilibrated with dialysis buffer from which the relevant nanodisc fractions were pooled and used for further analysis. Nanodisc containing ^{15}N labeled α_{IIb} was eluted using 10mM KPi, 10mM KCl, 0.05mM EDTA, pH 6.5 buffer containing TEV protease and concentrated for NMR experiments.

General Method for On-column Nanodisc Formation: All experiments were done at room temperature. Powdered version of DMPC was solubilized in dialysis buffer containing twice the concentration of sodium cholate in a glass vial by sonication. Solubilized lipids were added at an appropriate ratio to the resin containing bound MSP1E3D1 or target protein or both in a short flex column using a Hamilton syringe (Hamilton, USA). The volume of the reaction mixture was adjusted with dialysis buffer to get a crucial final cholate concentration between 12-15 mM. The column was capped and kept on a nutator

mixer for one hour. To this mixture wet BioBeads SM-2 (Sigma, USA) were proportionally added at a ratio of 0.6 gm/ml of the reaction volume. Prior to adding the beads they were preconditioned as recommended by the manufacturer and finally washed with dialysis buffer. This mixture was left horizontally shaking, suspending the BioBeads in solution for 3 hours after which the column was washed with appropriate buffer and eluted. Ni-NTA resin was salvaged from Bio-Beads by using a fine mesh sieve thereby making it available for regeneration.

Nuclear Magnetic Resonance (NMR)

HSQC experiments were run on Varian 800 MHz spectrometer (Agilent, USA) equipped with the cryoprobe using TROSY pulse sequence [5] at 40°C for 16 hours (256 scans in t2 and 128 increments in t1) in the above mentioned buffer of pH 6.5. Line-width analysis of the peaks obtained from the spectra was performed using the CCPN Analysis software [6]. T1, T2 relaxation experiments were carried out using the INVREC and CPMGT2 pulse sequences at 40°C.

Transmission Electron Microscopy (TEM) and Imaging

NDMT samples were adjusted to an absorbance A_{280} value of around 0.46. Each sample was diluted 80 times with Milli Q water (Millipore, USA) and prepared on a glow discharged copper grid (mesh 400) covered with carbon film. Grids were negatively stained with aqueous 1% uranyl acetate and air dried after wicking away excess stain. Electron micrographs were taken with AMT XR-40 camera slide mounted on a Tecnai Biotwin G2 Spirit Transmission Electron Microscope (FEI, USA). Images were taken at a direct magnification of 180000x. NDbR and NDA2b were adjusted to an absorbance A_{280} value of 1 and 0.5 respectively and diluted 40 times with Milli Q water (Millipore, USA).

ImageJ (NIH, USA) software was used for measuring the diameters of the two extreme NDMT samples NDMT1 and NDMT5. Electron micrograph images were reverse contrasted and converted to a binary format. The diameter was automatically counted using the Analyze Particles macro, after setting the scale to the calibrated value on the micrograph. The table obtained was exported and plotted using Origin. A total of 646 and 443 nanodiscs were counted for NDMT1 and NDMT5 respectively.

Dynamic Light Scattering (DLS)

All experiments were done at 25°C with Zetasizer NanoS (Malvern Instruments, USA) detecting backscatter at 173°. Six scans were each taken for sixty seconds and averaged to get the mean diameter. The particle size distribution by volume data was exported, fitted and overlaid in Origin.

Differential Scanning Calorimetry (DSC)

The most monodisperse sample, NDMT5, at a concentration of 0.95 mg/ml was used for the experiment. This sample was referenced against the buffer obtained as a filtrate from the concentration step. Prior to the run, both were filtered with 0.22 µm and degassed using vacuum. Thermograms were obtained by scanning from 10 °C to 50 °C at 35 °C/hour at a constant pressure of 3 atm [7]. In total, 3 sets of heating, cooling cycles were done using NanoDSC (TA Instruments, USA). The data was corrected for baseline with a thermogram obtained from buffer scans and fitted with residuals of 1.6 using the two state scaled model through the NanoAnalyze (TA Instruments, USA) software.

Results

On-Column Preparation of Empty Nanodiscs

In this method nanodiscs are formed bound to the resin through their outer belt protein MSP (Figure 1A). We chose MSP1E3D1 and DMPC lipids as they have been extensively used in reconstituting several membrane proteins within nanodiscs. The histidine tagged MSP1E3D1 was used for its ability to bind Ni-NTA resin. Discs were formed at the recommended MSP1E3D1:DMPC ratio of 1:160. Five samples, NDMT1 through NDMT5, were prepared successively doubling the protein concentration giving 1.65, 3.3, 6.6, 12 and 24 mg MSP bound per ml of resin. We stopped at 24 mg/ml to avoid crossing the binding capacity of the resin. This prevents any loss of protein in the flow-through, which could perturb the final MSP to DMPC ratio. All the five concentrations of MSP was prepared from a single cell lysate and mixed with solubilized DMPC obtained from the same stock preparation. An overlay of the chromatograms obtained from each of these samples through size-exclusion chromatography (SEC) on Superdex 200 column is shown in Figure 1B. The seemingly homogenous major peaks, from each of these samples, spread over a total difference of 5 ml in their elution volumes, with little to no free MSP. Interestingly, with increasing protein concentration bound per ml of resin, we saw a decrease in the amount of aggregates and a shift in the elution volumes. Nanodiscs were also prepared using the original solution method [1] as a control to compare with discs obtained from the on-column method. The inset in Figure 1B shows a perfect overlay of the chromatograms from the control and one of our preparations, NDMT4, and the free MSP protein eluting at 83 ml. The elution volume of the rightmost peak in our preparation is around 61 ml corresponding to a molecular weight of about 275 kDa. This

is in good accordance with 268.6 kDa calculated from the molecular weight of two MSP1E3D1 (32.6 kDa each) and a bilayer of 300 DMPC molecules (0.67 kDa each) [8].

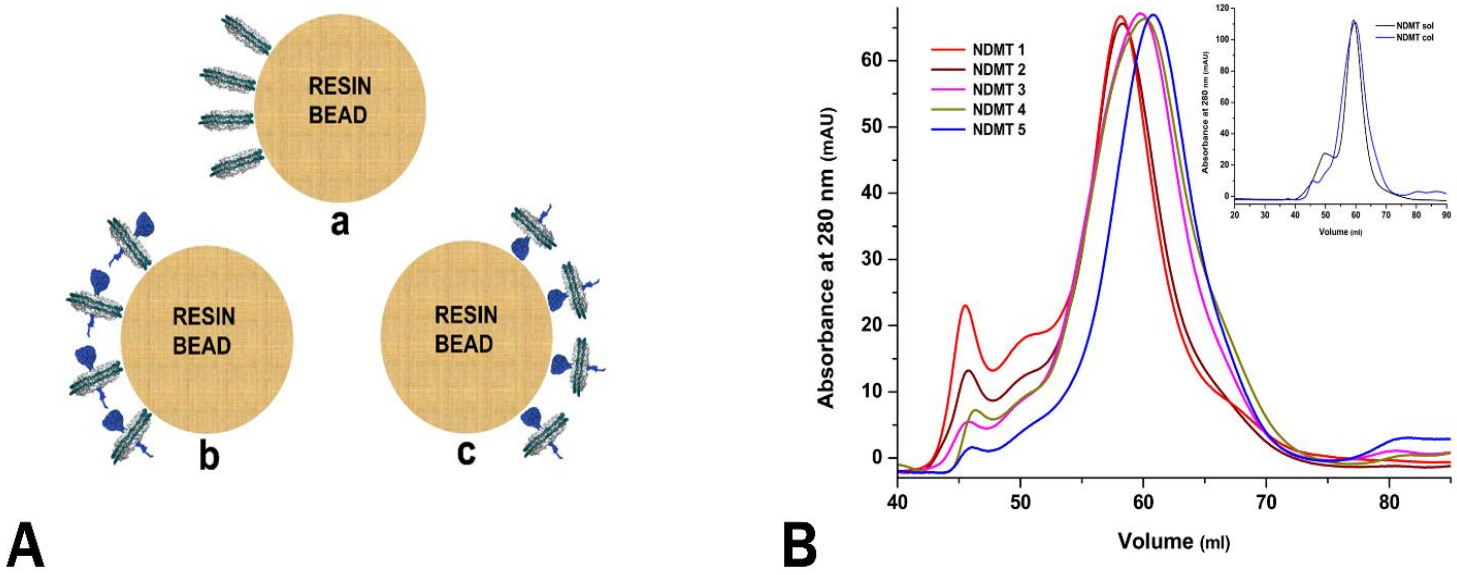


Figure 1. A. Model diagram showing the different components of a protein incorporated nanodisc and different schemes of the on-column method: (**a**) Nanodiscs bound to the resin through MSP, for example, NDMT; (**b**) Nanodisc with target protein bound to the resin through MSP, for example, NDbR; (**c**) Nanodisc with target protein bound to the resin through the target protein, for example, NDA2b.
B. Chromatogram overlay from SEC of the different NDMT samples: NDMT1 (red), NDMT2 (brown), NDMT3 (magenta), NDMT4 (khaki), and NDMT5 (blue). Inset: SEC comparing nanodiscs made in solution (NDMT sol: black) and on-column (NDMT col: blue), also shown is the free protein MSP1E3D1 (green).

TEM and DSC were used to determine the identity of nanodiscs. Electron micrographs obtained from TEM of the five NDMT samples [Figure 2B(i)] show discoidal nano-lipoproteins against a negative background. NDMT5 was further used to check for the characteristic transition temperatures obtained for nanodiscs through DSC. The thermogram in Figure 2B(ii) shows a broad phase transition, typical for nanodiscs, with a pre-transition temperature at 24°C and transition temperature at 27.6°C. Both

temperatures corresponded well with the phase transition temperatures previously reported for DMPC nanodiscs [7]. DLS was further used to confirm the purity and to quantitatively differentiate our five preparations. An overlay of the normalized particle size distributions by volume is shown in Figure 2A(i) with impressive samples homogeneity ranging from 100% to 99.8%. The table shown as inset in Figure 2A(i) gives the mean Stokes diameters ranging from 13.99 $\pm$ 2.02 nm for NDMT1 (1.65 mg/ml) to 11.99 $\pm$ 1.65 nm for NDMT5 (24 mg/ml). This correlates well with the originally reported diameter of 12 nm for MSP1E3D1 discs [1].

The above results raise an interesting question as to why discs, prepared by the same method varying only in the amount of MSP bound per ml of resin, gave different mean diameters. Hence, we sought to analyze the dispersity of our samples by measuring the diameters of the individual discs over an entire electron micrograph using ImageJ software [9]. 646 and 443 nanodiscs were respectively used for the two extreme samples, NDMT1 and NDMT5. Figure 2A(ii) shows the frequency distribution data, and Figure 2A(iii) shows the normalized curves fitted to this data. This clearly indicates that the NDMT5 preparation is more monodisperse than NDMT1. In electron micrographs the orientation of nanodiscs, deposition of stain around them along with the granular negative background leads to artifacts in defining the edges of the specimen at higher magnifications [10]. Hence, the populations were distributed around slightly bigger diameters as seen from the x-axis. None the less, these systematic artifacts do not affect the assessment of population dispersity.

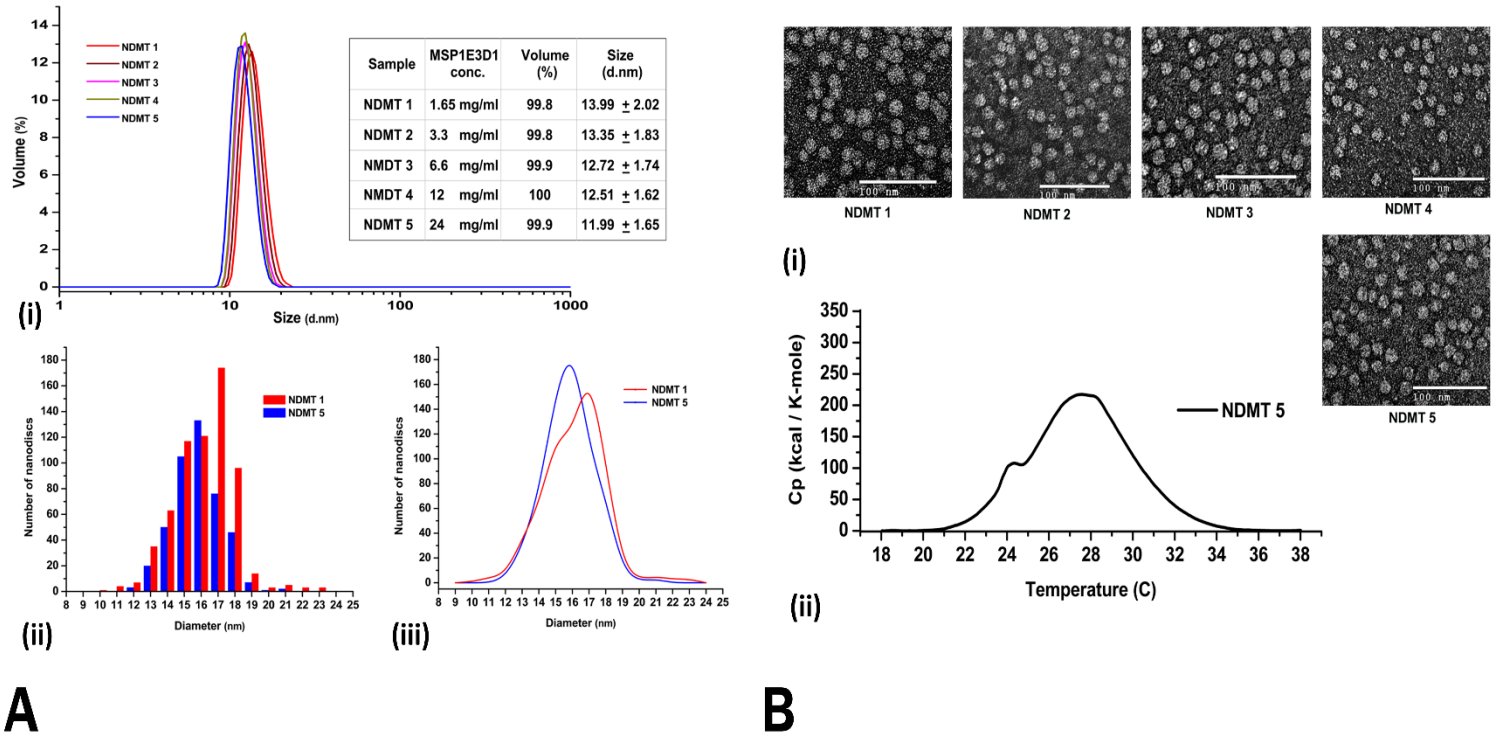


Figure 2: A. (a) TEM images of the different samples, NDMT1 through NDMT5, negatively stained with 1% uranyl acetate. (b) DSC thermogram showing the pretransition temperature (small shoulder) at 24 ºC and transition temperature at 27.6 ºC for NDMT5.**B.** (**a**) Overlay of the particle size distribution by volume for different nanodisc samples obtained through DLS: NDMT1 (red), NDMT2 (brown), NDMT3 (magenta), NDMT4 (khaki), and NDMT5 (blue). Inset: Table showing the mean Stokes diameters for the different NDMT samples. (**b**) Frequency distribution of nanodisc diameters for the two extreme samples NDMT1 (red) and NDMT5 (blue). (**c**) Normalized curve fitted to the same frequency distribution data.

On-Column Incorporation of Target IMPs into Nanodiscs, Two Case Studies

The primary aim behind making nanodiscs was to incorporate target membrane proteins. This was efficiently achieved by using the on-column method. Two types of proteins are involved in the process, MSP, which forms the outer layer covering the disc; and a target protein, which is encapsulated within the disc. Thus two potential approaches could be envisioned: (i) resin immobilized MSP is mixed with the target protein forming nanodiscs bound to the resin through MSP as shown in Figure 1A(b), or (ii) resin

immobilized target protein is mixed with MSP forming nanodiscs bound to the resin through the target protein as shown in Figure 1A(c). We have used bacteriorhodopsin to demonstrate the first approach and MBP-fused integrin α_{IIb} for the second. It is worthwhile to note that the second approach has been uniquely applied before for reconstituting anthrax toxin pore into nanodiscs [11].

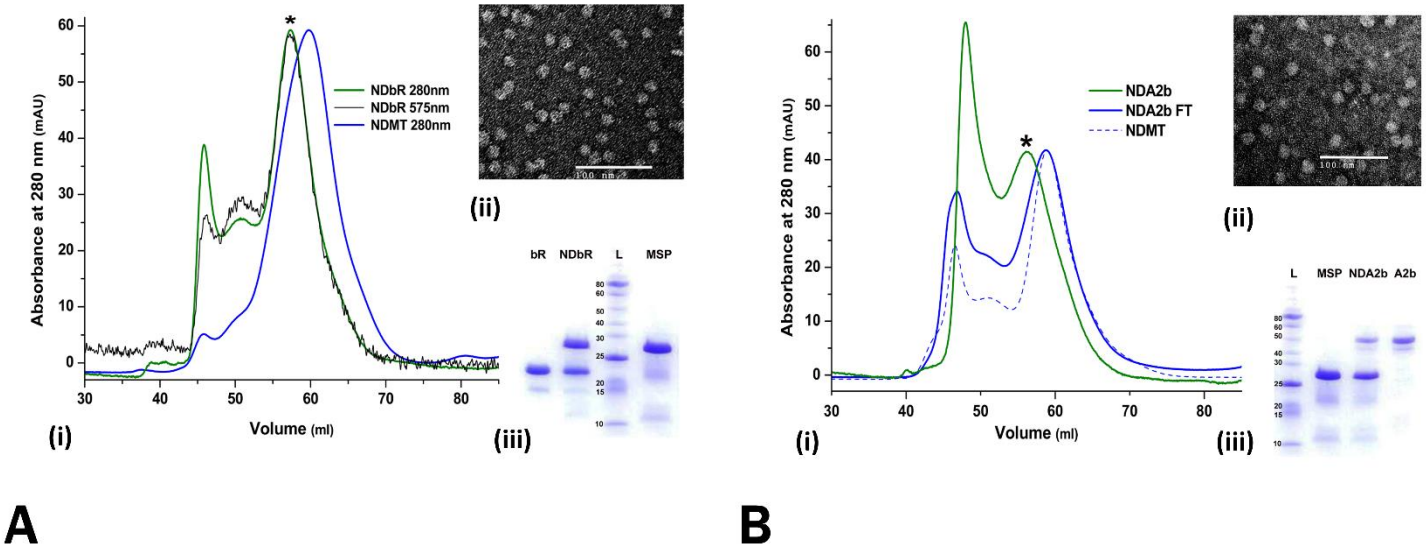
1. Bacteriorhodopsin in nanodiscs (NDbR)

Bacteriorhodopsin (bR) in purple membrane was purified as described elsewhere [3] and solubilized overnight in 4% Triton X-100 [12]. It was mixed with MSP bound to Ni-NTA resin in a previously optimized MSP1E3D1:bR:DMPC ratio of 2:3:160 [4]. Figure 3A(i) presents the overlay of the chromatograms obtained from SEC on Superdex 200 column for NDbR and NDMT monitored at 280 nm. Additionally, NDbR was monitored at 575 nm, a characteristic absorption wavelength for bacteriorhodopsin. The peak containing bacteriorhodopsin in nanodisc is shown by the asterisk sign to separate it from the other smaller peaks corresponding to higher order aggregates. The shift in elution volumes for the two chromatograms is around 2 ml. This corresponds to a difference in molecular weight of approximately 75 kDa, roughly indicating a homotrimeric arrangement of bacteriorhodopsin molecules within a single nanodisc as has been previously reported [4]. Additionally, when the fractions were pooled and concentrated, the sample exhibited a light purple color with two absorbance maxima at 280 and 556 nm as compared to the empty discs with a single absorption maximum at 280 nm (data not shown). Furthermore, TEM confirmed the formation of discs as indicated by the presence of discoidal particles in the electron micrograph [Figure 3A(i)]. SDS-PAGE analysis confirmed the presence of bacteriorhodopsin within nanodiscs as seen from lane two in

Figure 3A(iii) showing two bands: one, with MW of about 23 kDa, corresponding to bacteriorhodopsin shown also as a control in lane one; and the second, with a MW of about 28 kDa, corresponding to MSP1E3D1 also shown as a control in lane four. Each lane was compared against lane three representing molecular weight markers. In conclusion, we were successful in reconstituting bacteriorhodopsin within nanodiscs on Ni-NTA resin.

2. Integrin α_{IIb} Trans-Membrane and Cytoplasmic Domains in Nanodisc (NDA2b)

We used MBP fused transmembrane and cytoplasmic domains of integrin α_{IIb} (A2b). After lysing the cells the protein was left bound to the amylose resin to which we directly added MSP and lipids in a MSP1E3D1:A2b:DMPC ratio of 1:0.5:160. This ratio was calculated on the basis of having one α_{IIb} molecule per disc. Two types of discs were formed during this process: empty discs and protein incorporated discs. The discs containing α_{IIb} stayed bound to amylose resin through the MBP tag while empty discs which were unable to bind got washed out in the flow-through. Next, discs containing α_{IIb} were eluted using maltose. The flow-through and elution fractions were run on Superdex 200 column. As a control, empty discs prepared using Sligar's protocol was also run on the column. Figure 3B(i) presents an overlay of the chromatograms obtained from these samples. The peak containing α_{IIb} discs are shown by the asterisk sign to separate it from the other higher order aggregates. As expected, both the flow-through peak and the empty disc prepared as a control had the same elution volume. The peak corresponding to the α_{IIb} disc was shifted from the flow-through by approximately 3 ml, correlating with an increase in MW of about 100 kDa. This suggests that there are roughly two α_{IIb} (52



A **B**

Figure 3: A. (**a**) Size exclusion chromatography showing NDbR (green) and NDMT (blue) at 280 nm and NDbR (black) at 575 nm. (**b**) TEM image of NDbR negatively stained with 1% uranyl acetate. (**c**) SDS-PAGE analysis showing: lane 1, bacteriorhodopsin (bR); lane 2, nanodiscs containing bacteriorhodopsin (NDbR); lane 3, protein ladder (L); and lane 4, MSP1E3D1 (MSP). **B.** (**a**) Size-exclusion chromatography showing NDA2b (green), NDA2b flow-through (solid blue) and the control NDMT (dashed blue). (**b**) TEM image of NDA2b negatively stained with 1% uranyl acetate. (**c**) SDS-PAGE analysis showing: lane 1, protein ladder (L); lane 2, MSP1E3D1 (MSP); lane 3, nanodisc containing MBP-integrin α_{IIb} (NDA2b); and lane 4, MBP-integrin α_{IIb} (A2b)

kDa each) molecules per nanodisc, an interesting observation consistent with the previous reports for α_{IIb}-TMCD homo dimerization in dodecyl-phosphocoline micelles [13]. Again, TEM confirmed the formation of discs as indicated by the presence of discoidal particles in the electron micrograph [Figure 3B(ii)]. SDS-PAGE analysis confirmed the presence of α_{IIb} within nanodiscs as seen from lane three in Figure 3B(iii) showing two bands: one, with MW of about 50 kDa, corresponding to MBP-integrin α_{IIb} shown also as a control in lane four; and the second, with a MW of about 28 kDa, corresponding to MSP1E3D1, shown also as a control in lane two. Each lane was compared against lane one representing molecular weight markers. In conclusion, we successfully incorporated

MBP-fused integrin α_{IIb}-TMCD into nanodiscs on amylose resin, with the concurrent separation of empty and protein incorporated discs, in a single purification step.

Minimal Nanodiscs

1. Preparation

The overall diameter of nanodiscs is governed by the length of the outer coat protein, MSP. We created smaller discs by shortening the length of MSP. The popularly employed constructs of MSP, MSP1D1 and MSP1 E3D1, differ in the addition of three helical repeats in the center of the protein insinuating the possible interaction of terminal helices in the formation of nanodisc. Hence we created C-terminal deletion mutants to additionally ascertain the importance of these interactions. For the sake of simplicity, we will refer to the different sized MSPs as “D” followed by the number of helical repeats. So, the two types of nanodiscs made from the existing MSP1E3D1 and MSP1D1 are called D13 and D10 as they have thirteen and ten helical repeats respectively. We will focus on the MSP truncation mutants with five, six and seven helical repeats (Figure 4A(i)), where major differences in the disc size were observed. Discs were prepared from each of these mutants in an optimized (data not shown) MSP to DMPC ratio of 1:30 and analyzed on Superdex 200 column. The chromatograms obtained were overlaid for comparison. Interestingly, instead of continuous disc size reduction following the truncation of MSP length, the limit was reached at D7. After this point, the discs sizes increased with any successive deletion of helices. D7 discs, with the smallest size, have the lowest elution volume followed by D6 discs with two peaks, one close to the elution volume of D7 and the other to D5 [Figure 4A(ii)]. In order to confirm that the peaks fractions indeed contained nanodiscs we performed TEM. Figure 4B shows the presence of discoidal

nano-lipoproteins against a negative background for each of the D5, D6 and D7 discs. The D7 discs were further characterized by DLS giving a mean diameter of 7 nm (Figure 6B(i)); and SEC on a Superdex 75 column giving a molecular weight of 62 kDa [Figure 6B(ii)], which is in good accordance with the molecular weight of two D7 MSP (17.7 kDa each) and a bilayer of 40 DMPC molecules (0.67 kDa each).

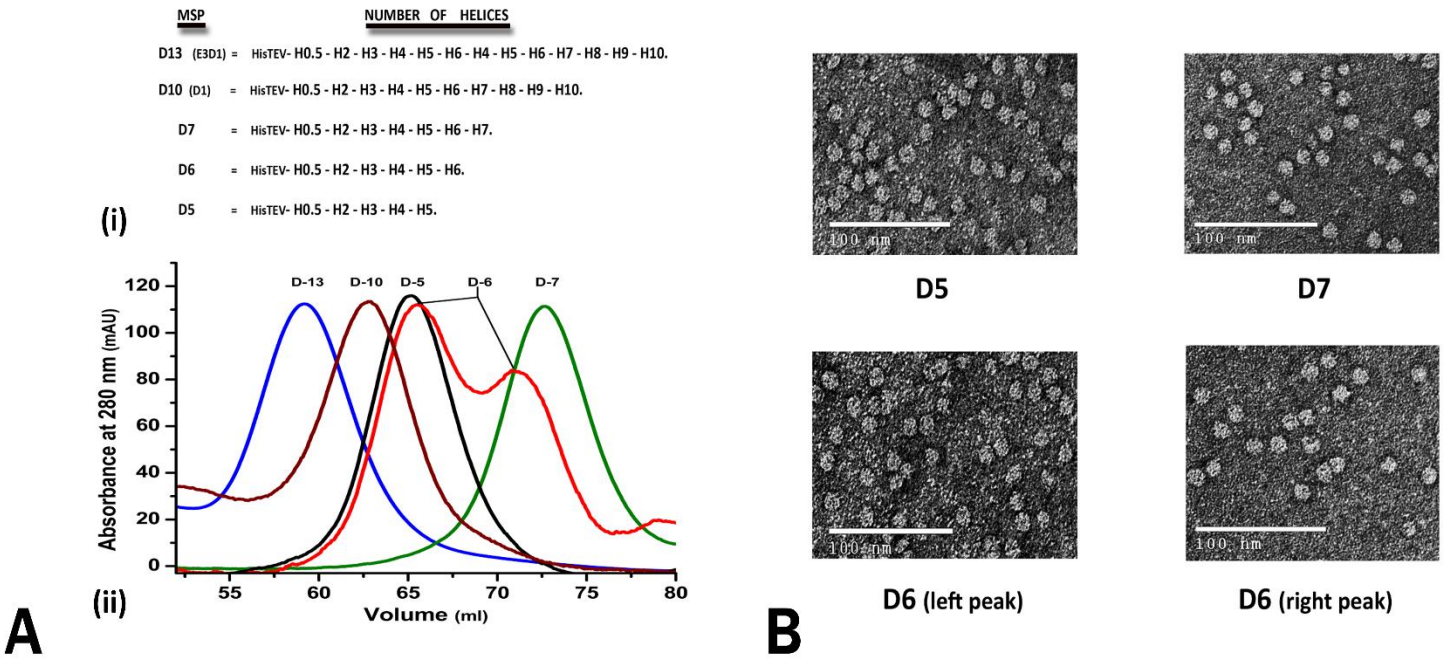


Figure 4: A. (**a**) Comparison of the different MSP mutants created by deletion of its terminal helices. (**b**) Overlay of the chromatograms obtained from size exclusion chromatography of discs prepared with different MSP mutants. **B.** TEM images of the major peak fractions from D5, D7 discs and the two peaks from the D6 disc, negatively stained with 1% uranyl acetate.

2. NMR Spectra

Our major goal behind developing smaller discs was to use them for studying integral membrane proteins by solution NMR. We incorporated integrin α_{IIb} into our smallest D7 discs using the on-column method at an optimum MSP: α_{IIb}: DMPC ratio of 1:0.5:20. In order to show that nanodiscs can be successfully used for solution NMR we obtained ^{15}N-

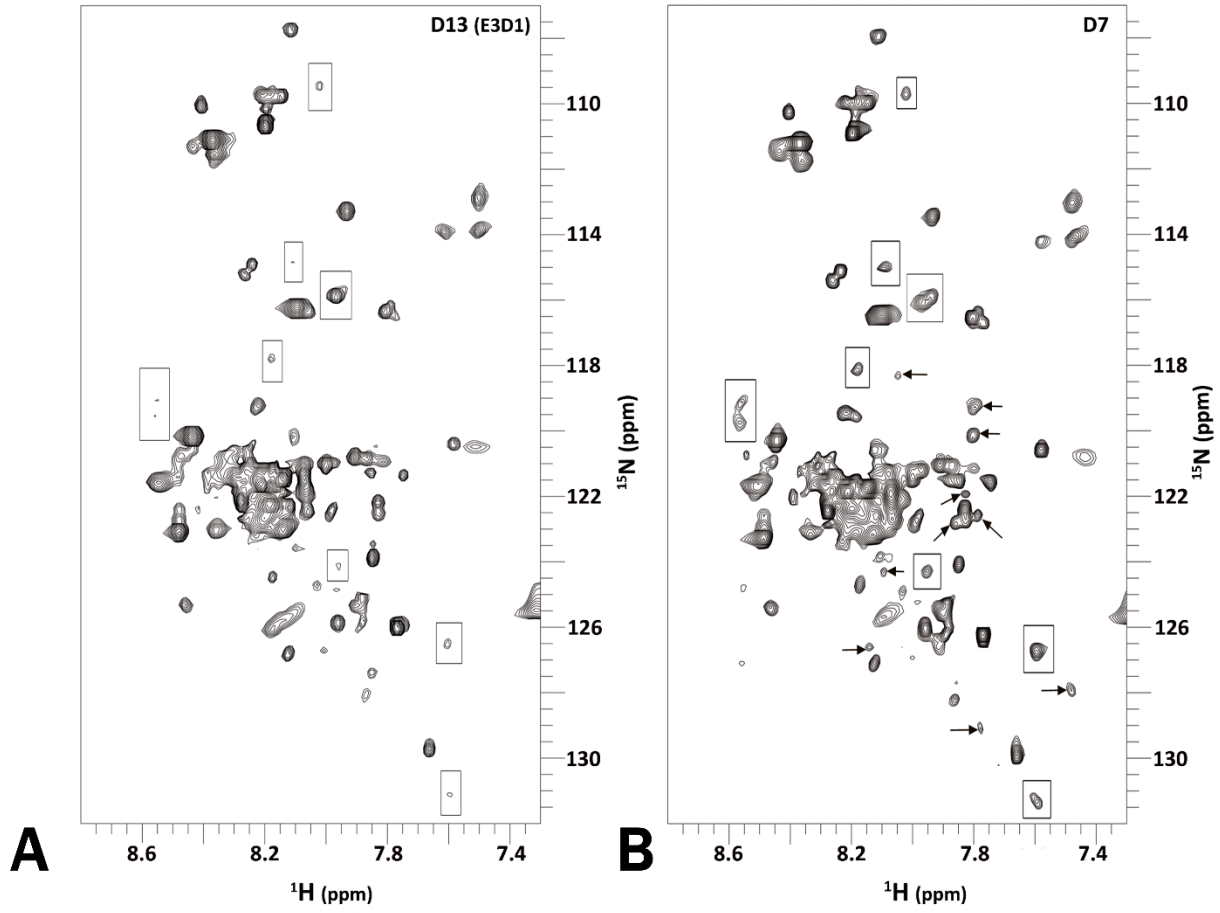


Figure 5: ^{15}N TROSY-HSQC obtained from integrin α_{IIb}-TMCD incorporated into larger D13 (MSP1E3D1) discs **(A)** and smaller D7 discs **(B)**. Boxes represent peaks with increased intensity while arrows represent new peaks.

TROSY-HSQC spectra for integrin α_{IIb}-TMCD incorporated into D13, D10, and D7 discs on an 800 MHz spectrometer. Although we did not see much difference between the D13 and D10 spectra (data not shown), significant improvement in spectrum quality was observed for D7 when compared to D13 discs. All amide peaks in the spectra could be divided in two major groups. Group I peaks represent the amides belonging to the cytoplasmic (~13 amino acids) and extracellular (9 amino acids) parts of the protein which, as a result of fast local motion, are equally sharp in both spectra (Figure 5). Group II peaks, corresponding to ~20 trans-membrane and ~8 membrane-proximal amino acids, are much broader, reflective of the longer overall rotational correlation time of the larger disc complex. The spectrum improvement obtained in smaller discs is associated with this group: some peaks, which are almost scarcely visible in D13 discs (shown by boxes), become much more prominent in D7 discs; while the others (shown by arrows) only appear in D7 discs bringing the overall number of peaks to an expected value of ~ 50. Interestingly, even among these broader peaks in D7 discs we observe noticeable heterogeneity in the peak widths, which is possibly an indication of some residual local motion associated with the depth of membrane insertion. We have further analyzed the line-width differences in ^{1}H dimension for the two spectra using CCPN Analysis software [6], and confirmed the improvement as exemplified by the table presented in Figure 6A for several well resolved peaks. Though significant, the accomplishments with the smaller discs come with the caveat of having an incomplete understanding of its morphology and bilayer properties at such low lipid concentrations. We additionally measured the longitudinal and transverse relaxation rates of the acyl chains in the DMPC lipids within nanodiscs (-$(CH_2)_n$- peak at ~ 1.3 ppm). We found approximately a 1.6 times reduction in

the overall rotational correlation times (τ_C) of the two disc complexes (T1/T2 ratio in D13 discs of ~55 is reduced to ~35 in D7 discs).

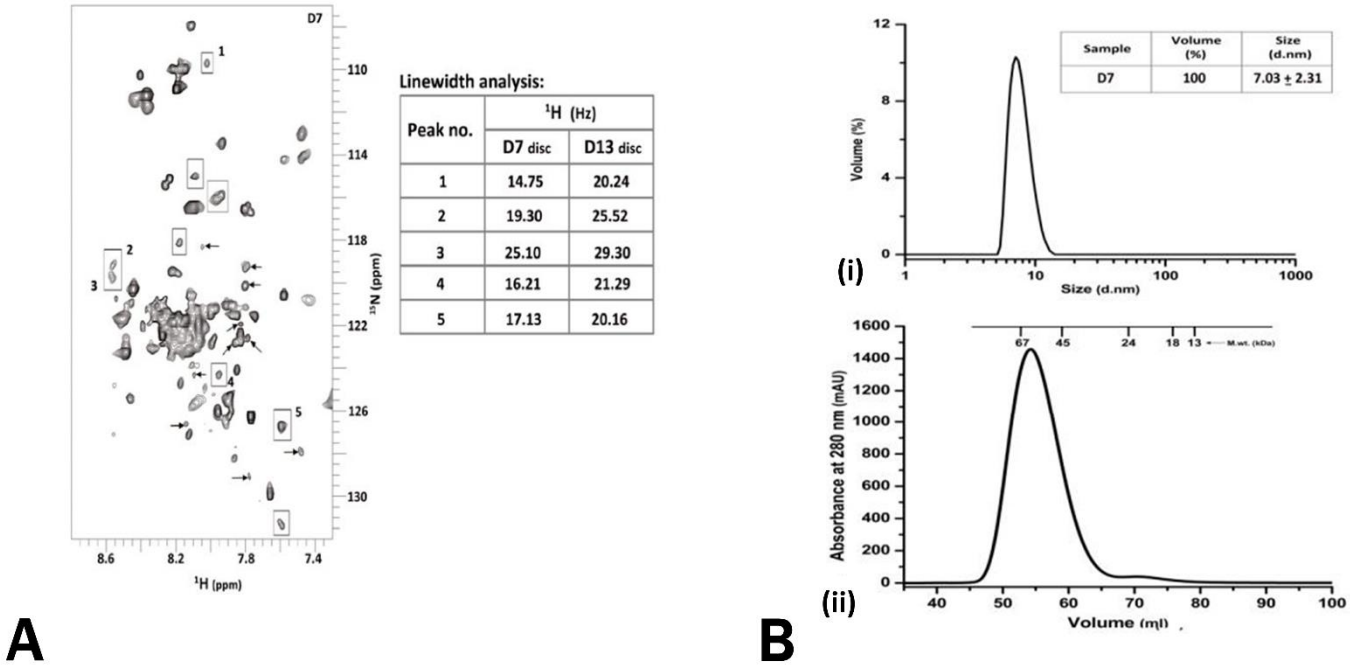


Figure 6: A. (a) DLS data showing the particle size distribution by volume for the D7 disc. Inset: Table showing the mean Stokes diameter. (b) Chromatogram of the D7 discs obtained from size exclusion chromatography on Superdex 75. **B.** A comparison of the line-width analysis in 1 H dimension of the five selected nicely separated peaks from 15N TROSY-HSQC of integrin αIIb-TMCD in D7 and D13 discs.

Discussion

Our interest in nanodiscs as a suitable membrane mimetic for studying membrane proteins using solution NMR led us to develop a simplified method for their preparation. In this method the entire preparation occurs in a single purification step. Nanodiscs are formed bound to the resin and are eluted directly from the column. MSP's tendency to aggregate at higher concentrations in solution [1] is circumvented in our method where

high concentration of MSP is achieved due to the large binding capacity of the resin. As the result, more concentrated discs samples are prepared in lower reaction volumes. Though the binding of MSP to the resin is random, its distribution across the volume of resin is directly proportional to the concentration. Thus it would be logical to imagine that some MSP molecules are held further apart at lower concentrations than at higher ones. We suspect that this forced physical separation might lead to the formation of some discs with slightly larger diameters than regular optimal ones. NDMT1 and NDMT5 samples have respectively the lowest and highest concentration of protein bound per ml of resin. The frequency distribution data obtained for NDMT1 clearly shows the presence of a heterogeneous population distribution when compared to NDMT5. This explains why the average mean diameter obtained from the DLS experiment is higher for NDMT1 over NDMT5. Though more rigorous experimentation could explain what exactly led to this population dispersity, our experiments clearly indicate that NDMT5 (having the highest protein concentration per volume of resin) produces more homogenous and monodisperse nanodisc preparation than NDMT1. Thus, the use of higher protein concentration bound per ml of resin is preferable for obtaining a homogenous preparation of nanodiscs. DSC thermogram obtained for NDMT5 shows the characteristic broad phase transition [Figure 2B9ii)], with transition temperatures identical to the ones previously reported for nanodiscs [7]. Since there is no observed difference in bilayer characteristics of nanodiscs produced by the two methods, replacement of the chloroform solubilized DMPC with its powdered form worked really well and seems like a feasible alternative. Given the stability of resin beads at different temperatures, this method can

be extended to other lipids like DPPC or POPC; which in combination with different MSP constructs, will make our method applicable over the entire catalog of nanodiscs.

The major advantage of our on-column method is highlighted while incorporating proteins within nanodiscs. Not only do we achieve a rapid, single step, homogenous preparation but we do so with the concurrent separation of protein incorporated discs from the empty discs. This is beneficial as it improves the yield by reducing the number of purification steps, providing a more pure homogeneous sample, typically required for structural studies, in a shorter period of time. A quick graphical comparison of the two methods as shown in Chapter-1 Figure 3A clearly delineates the advantages of our method. In order to test the robustness of the presented method we used different target proteins to exemplify the two strategies for incorporating membrane proteins within nanodiscs.

The first strategy entails the formation of protein incorporated discs bound to the resin through MSP. We used bacteriorhodopsin, a well characterized protein within nanodiscs, as a target protein to demonstrate the feasibility of this approach. The washing step removed free lipids and any unincorporated bacteriorhodopsin. Following elution, SEC confirmed the absence of the empty discs through the comparison of the UV absorption profiles at 280 and 575 nm.

However, it was difficult to separate the reconstituted discs from the empty ones due to their similar sizes. This drawback was overcome by our second strategy where the protein incorporated discs are formed bound to the resin through the target protein. We used MBP tagged transmembrane and cytoplasmic domains of integrin α_{IIb} as our target protein to demonstrate this case. The amylose resin selectively binds to the MBP tagged

protein. Since MSP lacks the MBP tag it cannot bind to the resin. This feature helped in the removal of unincorporated MSP and empty discs along with free lipids during the washing step. Integrin α_{IIb} containing discs could then be eluted either by using maltose or by cutting the MBP tag with TEV protease. The latter approach was used when preparing samples for solution NMR studies.

A limiting factor affecting the study of membrane proteins within any system using solution NMR lies in the effective size of the entire complex. Thus we developed smaller discs to achieve shorter rotational correlation time. Since all the deletion mutants formed nanodisc, it implies that interaction of terminal helices are not crucial for the formation of discs. MSP mutant containing seven helical repeats produced the smallest nanodisc. Any further reduction in the helical repeats led to a successive increase in the size of the nanodisc. Interestingly, D5 discs, formed with MSP containing five helical repeats, were similar in size to D10 discs, containing MSP with ten helical repeats. The larger size of D5 discs insinuates the possibility of having more than two MSP proteins per nanodisc. The above pattern of disc formation with several separate helical peptides derived from MSP possibly explains the self-assembly of 30 nm “macrodiscs” developed as an alignment media for residual dipolar coupling experiments [14]. The advantage of our smallest D7 discs, with a size of 7 nm and molecular weight of 62 kDa, as compared to D13(MSP1E3D1) discs, (size of 12 nm and molecular weight of 275 kDa), is clearly evident from the presented ^{15}N TROSY-HSQC data. The improvements in line-width are associated with the broadest peaks in the spectrum, corresponding to the amides within the lipid bilayer as confirmed by relaxation measurements showing the reduction in overall rotational correlation time. Clearly, with such relatively large system the deuteration of

lipids and incorporated proteins would be required to avoid unwanted dipolar-dipolar interactions. In our view, D7 discs provide a suitable system for structural investigation of small integral membrane proteins as exemplified by their application to integrin α_{IIb} transmembrane and cytoplasmic domains.

References

1. Ritchie, T.K., et al., *Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs.* Methods Enzymol, 2009. **464**: p. 211-31.
2. Yang, J., et al., *Structure of an integrin {alpha}IIb{beta}3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation.* Proc Natl Acad Sci U S A, 2009. **106**(42): p. 17729-34.
3. Oesterhelt, D. and W. Stoeckenius, *Isolation of the cell membrane of Halobacterium halobium and its fractionation into red and purple membrane.* Methods Enzymol, 1974. **31**(Pt A): p. 667-78.
4. Bayburt, T.H., Y.V. Grinkova, and S.G. Sligar, *Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs.* Arch Biochem Biophys, 2006. **450**(2): p. 215-22.
5. Salzmann, M., et al., *TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins.* Proc Natl Acad Sci U S A, 1998. **95**: p. 13585-90.
6. Vranken, W., F., et al., *The CCPN data model for NMR Spectroscopy.* Proteins, 2005. **59**(59): p. 687-696.
7. Shaw, A.W., M.A. McLean, and S.G. Sligar, *Phospholipid phase transitions in homogeneous nanometer scale bilayer discs.* FEBS Lett, 2004. **556**(1-3): p. 260-4.
8. Bayburt, T.H. and S.G. Sligar, *Membrane protein assembly into Nanodiscs.* FEBS Lett, 2010. **584**(9): p. 1721-7.
9. Abramoff, M.D., P.J.Magalhaes, and S.J Ram., *"Image Processing with ImageJ".* Biophotonics International, 2004. **11**(7): p. 36-42.
10. Pyrz, W.D. and D.J. Buttrey, *Particle size determination using TEM: a discussion of image acquisition and analysis for the novice microscopist.* Langmuir, 2008. **24**(20): p. 11350-60.
11. Katayama, H., et al., *Three-dimensional structure of the anthrax toxin pore inserted into lipid nanodiscs and lipid vesicles.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3453-7.
12. Dencher, N.A. and M.P. Heyn, *Formation and properties of bacteriorhodopsin monomers in the non-ionic detergents octyl-beta-D-glucoside and Triton X-100.* FEBS Lett, 1978. **96**(2): p. 322-6.
13. Li, R., et al., *Oligomerization of the integrin alphaIIbbeta3: roles of the transmembrane and cytoplasmic domains.* Proc Natl Acad Sci U S A, 2001. **98**(22): p. 12462-7.
14. Park, S.H., et al., *Nanodiscs versus macrodiscs for NMR of membrane proteins.* Biochemistry, 2011. **50**(42): p. 8983-5.

Chapter 3

Application of nanodiscs to α helical and β barrel membrane proteins

Adapted from:

Nanodiscs and Solution NMR: preparation, application and challenges.
Puthenveetil, R., Nguyen, K., and Olga Vinogradova (2017).
Nanotechnology Reviews 6 (1), 111-125.

&

Bipartite Topology of Treponema pallidum Repeat Proteins C/D and I: OM Insertion and Porin Function Require a C-terminal β-barrel Domain. Anand A, LeDoyt M, Karanian C, Luthra A, Koszelak-Rosenblum M, Malkowski MG, Puthenveetil R, Vinogradova O, Radolf JD (2015). J Biol Chem., 290(19): 12313-31

Introduction

Nanodiscs have been successfully used for a myriad of proteins. (i) Blood coagulation and human tissue factor: Tissue factor (TF), an integral membrane protein complexes with Factor VIIa (FVIIa) to initiate the blood coagulation cascade. FVIIa generally requires acidic phospholipids like phosphatidylserine (PS) that tends to form clusters in large bilayers. This problem was circumvented by the smaller size of nanodiscs bilayers and was used to study the binding of Factor VIIa using SPR [1, 2]. (ii) Bacteriorhodopsin: Native trimeric bacteriorhodopsin was studied through the excitation of chromophores in nanodisc [3]. Monomeric forms of bR were also utilized to address structure and function aspects in disc [4]. (iii) Cytochrome P450: Using purified affinity tags multiple integral membrane proteins CYP3A4 [5] and a cinnamate hydroxylase CYP73A5 [6] were assembled along with their redox partners Cyt P450 into nanodiscs. CYP3A4 was additionally studied with ^{13}C,^{15}N-labeled proteins using solid state NMR [7].(iv) SecYEG translocon complex: Heterotrimeric 15 transmembrane SecYEG incorporated into nanodisc was shown to competitively to Syd which displaced the previously bound SecA [8]. (v) β2 adrenergic receptor: The binding 7 pass transmembrane GPCR β2AR to its agonist and antagonist has been successfully studied in nanodisc [9]; though the affinity of Gs (G-protein subunit) was markedly reduced in the receptor nanodisc system [10]. (vi) Epidermal growth factor receptor (EGFR): The kinase activity of EGFR was demonstrated in nanodisc was found to be better than in detergent solutions [11]. (vii) TPRV1: recently, a Cryo-EM structure of rat TPRV1, an ion channel, was obtained in large MSP2N2 nanodisc revealed its mechanism of ligand and lipid action [12].

Though there are very few structure papers at this point, there still are several examples of successful nanodiscs application with solution NMR. The initial use of nanodisc was demonstrated for two membrane proteins: (i) CD4 mutant, containing a single transmembrane and cytoplasmic tail, where the aliphatic resonances in ^{1}H-^{13}C HSQC spectra were compared to that in DPC micelles [13]; and (ii) VDAC-1, human anion channel protein, where ^{1}H-^{15}N TROSY HSQC spectra were compared between LDAO isotropic bicelles and MSP1D1 nanodiscs [14]. Additionally, the occurrence of Chemical shift perturbations (CSP) due to NADH binding was demonstrated. A similar study was later performed with VDAC-2 [15]. Following this, several other proteins were investigated in nanodiscs. The Voltage sensing domain (VSD) of KvAP channel, containing four consecutive transmembrane helices, was incorporated into (MSP1D1) nanodisc with different types of lipid molecules and was found to maintain a proper conformation only in zwitterionic environments [16]. A 30 nm MACRODISC, discussed earlier, was developed as a medium for residual dipolar coupling measurements [17]. Bacteriorhodopsin, a hepta-helical transmembrane protein, was studied comparatively across different mimetics comprising of micelles, amphipols, and (MSP1D1) nanodiscs [18]. ^{1}H-^{15}N TROSY HSQC spectra of the trans-membrane and cytoplasmic domains of Intergrin α_{IIb} spectra was obtained in smaller D7 nanodisc [19]. YgaP, an E.coli integral membrane protein with its two transmembrane and cytoplasmic Rhodanese domain, was studied in mixed micelles, FC12 micelles, and (MSP1D1) nanodiscs. The study demonstrated the improvement in quality of ^{1}H-^{15}N TROSY HSQC spectrum in the presence of deuterated d54-DMPC lipids in nanodisc. Interestingly, the spectral quality in nanodisc was relatively better in comparison to FC12, which reflects the perturbed

structural integrity of Rhodanese domain in micelles [20]. Sequential assignments of OmpA, an *E. coli* beta barrel protein, was achieved in (MSP1D1) nanodisc albeit with the help of assignments from Fos-10 micelles. Additionally, application of micro-coil NMR was demonstrated to find optimized solution conditions for OmpA in nanodisc [21]. ^{1}H-^{15}N TROSY HSQC spectra of two beta-barrel outer membrane proteins OprG and OprH from *P. aeruginosa* were compared between DHPC micelles and different truncation mutants of MSP ΔH4, ΔH5, ΔH4H5 and ΔH4-H6 (developed by Wagner's group [22]). In agreement with Wagner's findings, it was found that ΔH5 offered the best spectrum [23]. Protein – protein interaction have also been studied through the incorporation of P450-cyt$_b$ in a 22 amino acid long peptide based nanodisc [24]. Another approach, using HADDOCK simulations [25] with distance restraints derived from paramagnetic relaxation enhancement (PRE), allowed to distinguish the reorientation of the effector binding site on K-RAS4B GTPase anchored to (MSP1D1) nanodisc, that was previously occluded through the anionic membrane [26]. The most rigorous and compelling study came from Wagner's group, where they solved the structure of a beta barrel protein, OmpX, in nanodisc using small nanodiscs formed by the ΔH5 deletion construct of MSP [22].

Results and Discussion

Alpha helical proteins in nanodisc

With the aim to demonstrate the application of nanodiscs applications beyond structural studies, we hereby demonstrate how a nanodisc can be used to study the functional activity of a membrane protein. To demonstrate this we used the major platelet

Integrin $\alpha_{IIb}\beta_3$ as an example. Integrins represents a class of human cell surface receptor heterodimers formed by a combination of alpha and beta subunits. Upon stimulation of the extracellular domain, Integrin undergoes activation, which results in the phosphorylation at the β3 cytoplasmic tail of two tyrosine residues by Src kinase as shown in Figure 1B (Top Panel). Integrin constructs, used in our study, consisted of an MBP tag (maltose binding protein) fused to the transmembrane and cytoplasmic domain (TMCD) of α_{IIb} or β_3 [27]. MBP tag enhances the solubility of the TMCD region, which aids in the purification and incorporation of β_3 into nanodisc. Once Integrin subunits are incorporated into the disc, the MBP tag is cut off with TEV protease through the TEV cut site engineered between MBP and the rest of the protein, providing nanodiscs containing just the TMCD region (Figure 1A). We used our small D7 MSP construct (roughly similar to Wagner's ΔH4H5 construct) to form nanodisc. By themselves, α_{IIb} forms homo-dimers and β_3 forms homo-trimers [28, 29]. The ^{1}H-^{15}N TROSY HSQC spectra of the individual α_{IIb} or β_3 incorporated into discs were collected on an 800 MHz magnet (Agilent, USA) and are shown in Figure 1B. Thus, at least two or three transmembrane helixes fit comfortably within the internal lipid area of the disc having an estimated inner diameter of ~ 60 Å. The overall size of the assembly, even within the smallest nanodiscs, is still quite large, resulting in broader peaks corresponding to the transmembrane portion of the protein. Additional local motion of cytoplasmic tails, which protrudes out from the disc, were manifested by much sharper peaks. We exploited this advantage to study the receptor activation of Integrin β_3 on a low field 600 MHz magnet (Agilent, USA). Lack of deuteration makes it difficult to observe peaks from the transmembrane region, a phenomenon that could be exploited to our advantage to focus on the cytoplasmic tail. We simply mixed the

^{15}N-labled β_3 in discs with Src kinase along with ATP molecules in an NMR tube and collected a ^{1}H-^{15}N TROSY HSQC spectrum. By overlaying the un-phosphorylated and bi-phosphorylated spectra, one can clearly see shifts arising due to the phosphorylation of β_3 tails in nanodiscs (Figure 1C: Right Panel). The shifts correspond to the region around the tyrosine residues at the C-terminus of the cytoplasmic tail where phosphorylation occurs as shown from our previous studies using just the short cytoplasmic tail regions [30]. Assignments will further be required to exactly pinpoint the specific residues perturbed in the process. Thus, here we show a simple application, free from the need for deuteration, which can be routinely applied with a relatively low field NMR magnet to study receptors signaling. In addition to integrins we have also incorporated the last transmembrane and cytoplasmic tail of CD47 into nanodiscs (Figure 1B) the application of which is detailed in Chapter 4.

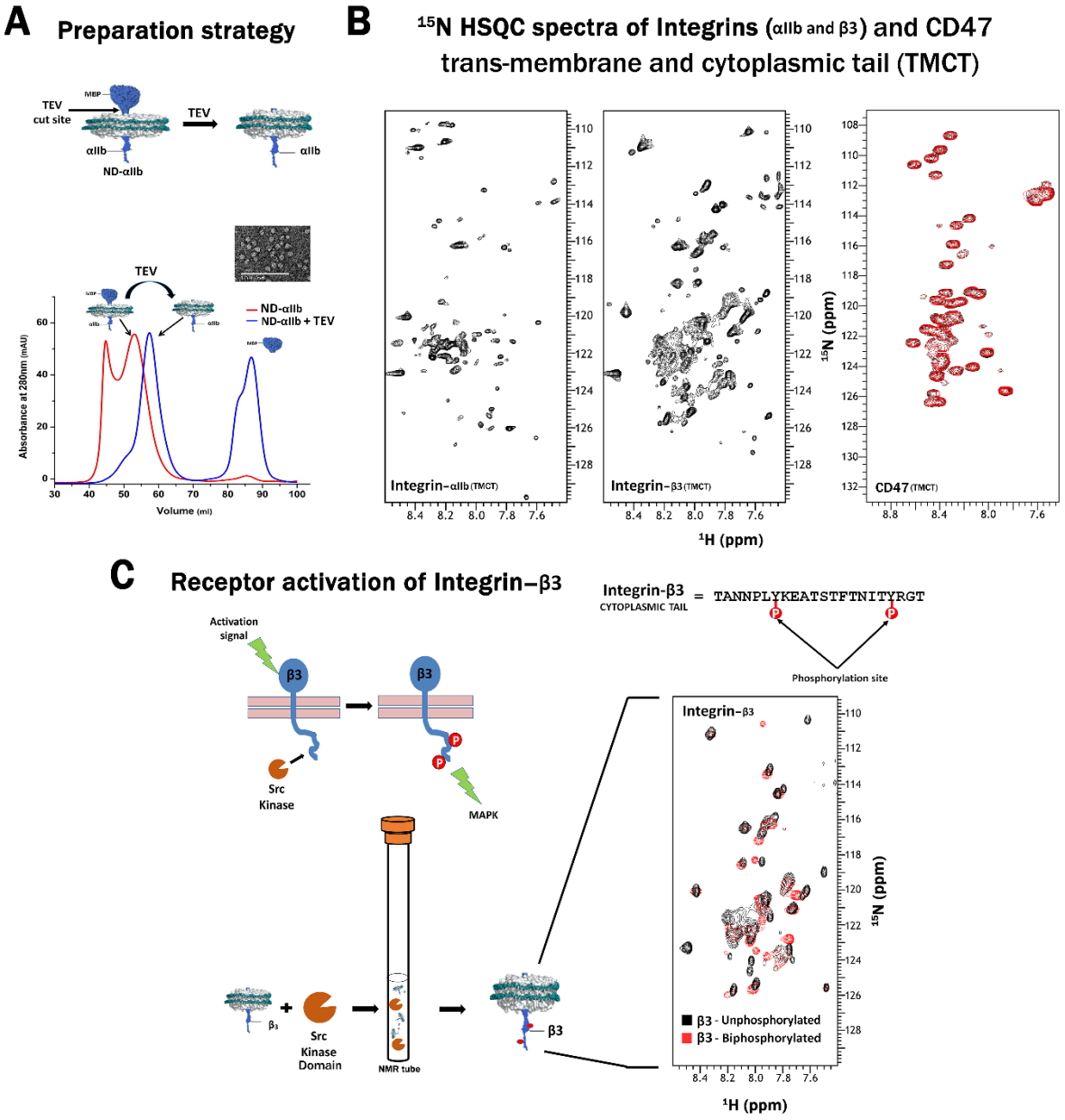


Figure 1. (A) Overview of the procedure to obtain Integrin α_{IIb} incorporated nanodiscs using an MBP-tag. After successful incorporation of MBP-α_{IIb} into nanodiscs, TEV protease treatment is performed which cuts at the TEV cut site previously engineered between MBP and our protein of interest, this allows for the removal of the MBP tag. The α_{IIb} incorporated nanodiscs can then be purified from the rest of the cut products using SEC. Inset: Negatively stained TEM image of α_{IIb} containing nanodiscs displaying a monodisperse population of discs. The exact same procedure can also be extended to obtain β_3 and CD47 incorporated nanodiscs. (B) ^{1}H-^{15}N- TROSY HSQC spectra of Integrin α_{IIb} and β_3 (TMCT) acquired on 800 MHz magnet and CD47 (TMCT) acquired on 600MHz magnet (Agilent, USA). The sharp peaks corresponds to the cytoplasmic tail region while for Integrins the broader peaks correspond to the transmembrane region. (C) Applicability of nanodisc systems for studying signaling pathways. A schematic showing the activation events of Integrin β_3 which is relayed intracellularly through the phosphorylation of its cytoplasmic tail by Src kinase (Top Panel). This event is mimicked (*in vitro*) in an NMR tube by mixing ^{15}N labeled β_3

incorporated nanodiscs with the purified kinase domain of Src kinase in the presence of ATP molecules. Phosphorylation is manifested through chemical shifts perturbations (CSP) observed through the overlay of the ^{1}H-^{15}N- TROSY HSQC spectra from the un-phosphorylated (black) and bi-phosphorylated (red) β_3 collected on a 600 MHz magnet (Agilent, USA). Also shown is the sequence of the β_3 tail where the phosphorylation occurs at the two tyrosine residues.

Beta barrel proteins in nanodisc

Treponema pallidum, a gram negative spirochete is the causative agent of the sexually transmitted disease, syphilis. It is a highly motile, extracellular bacterium renowned for its invasiveness. The proteins that assemble into the *T. pallidum* outer membrane (OM) determines the bacterium's ability to obtain nutrients, break tissue and endothelial barriers, and accomplish the many facets of its complex infectiveness. Previously, using bioinformatics-based approaches, several rare OMPs were identified based on their ability to form β-barrel structures [31], many of which are members of the paralogous *T. pallidum* repeat family (Tpr) [32]. The most prominent identified OMP was TprC (TP0117). Subsequently [33], TprC was demonstrated to possess most properties of a typical OMP (*i.e.* β-barrel structure, amphiphilicity, low abundance, and surface exposure) with the ability to form channels in large unilamellar liposomes (LUVs). Biophysical analysis revealed $TprC^{C}$, but not $TprC^{N}$, can form an amphiphilic β-barrel with porin activity equivalent to that of the full-length polypeptide. Based on these findings it was proposed that $TprC^{C}$ was sufficient to inserts into the OM bilayer while $TprC^{NT}$ resides in the periplasm giving the protein a distinct bipartite topology [34]. Both recombinant $TprC^{C}$ and $MOSP^{C}$ were incorporated into nanodiscs and analyzed for their ability to form porins. Both recombinant $TprC^{C}$ and $MOSP^{C}$ denatured in 8M Urea and refolded in 1% DDM (n-Dodecyl-beta-Maltoside) (Avanti Polar, USA). $rTprC^{C}$/$rMOSP^{C}$ was incorporated into large MSP1E3D1 discs and negatively stained using 0.75% uranyl

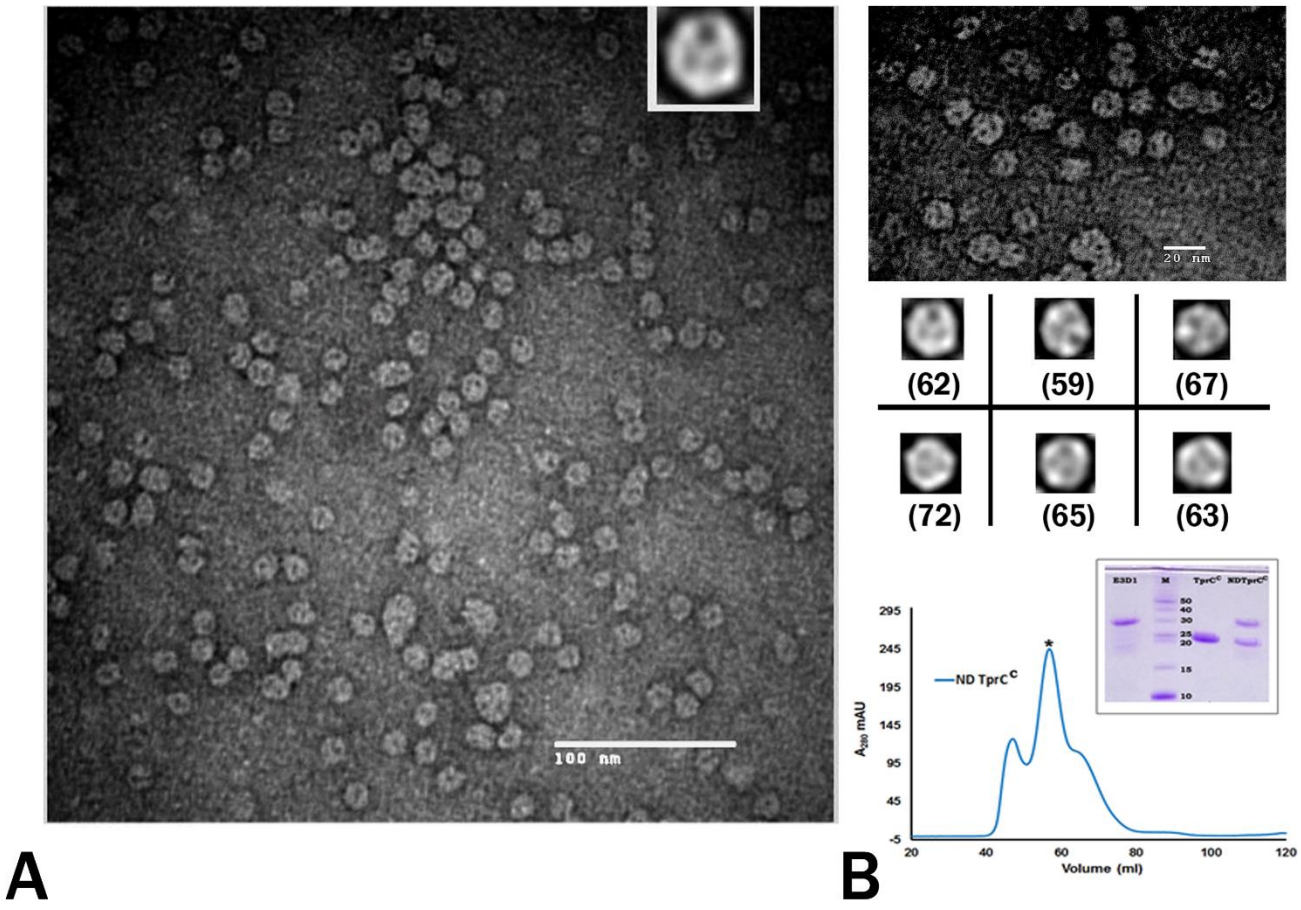


Figure 2: TprC^C in MSP1E3D1 nanodiscs: **(A)** Negatively stained TEM image of TprC^C discs taken at 250K magnification with a defocus of ~1.6. **(B) Top:** Negatively stained TEM image at 360K magnification; **Middle:** Image averaged discs done on six TEM images with the number of discs used per image for averaging shown in parenthesis; **Bottom:** Size Exclusion chromatogram showing the peak containing TprC^C incorporated discs in asterisk. SDS-PAGE of the chromatogram peak showing incorporation of TprC^C in nanodiscs as seen by the presence of two bands in lane 4. Lanes 1 and 3 serve as controls with molecular weight markers shown in lane 2.

formate (SPI-Chem, USA). Images were observed under electron microscope (FEI Tecnai G2 Spirit BioTWIN) using a low dose method. Multiple discs images were further averaged, using XMIPP, to obtain a final representative image displaying the arrangement of porins in the discs. Our analysis showed that the both TprC^C (Figure 2)

and MOSP^{C} (Figure 3) formed porins in nanodiscs as can be adjudicated from the presence of small pores envisioned by the visualization of three black dots in the center of the discs. These dots are an outcome of the deposition of the stain in the cavity formed by the beta barrel porins. Interestingly, we can see that they form trimers in the discs, an oligomeric state native to these proteins as demonstrated through western blotting of the lysate from *T. pallidum and T. denticola* [35]. MOSP is more extensively discussed in Chapter 5

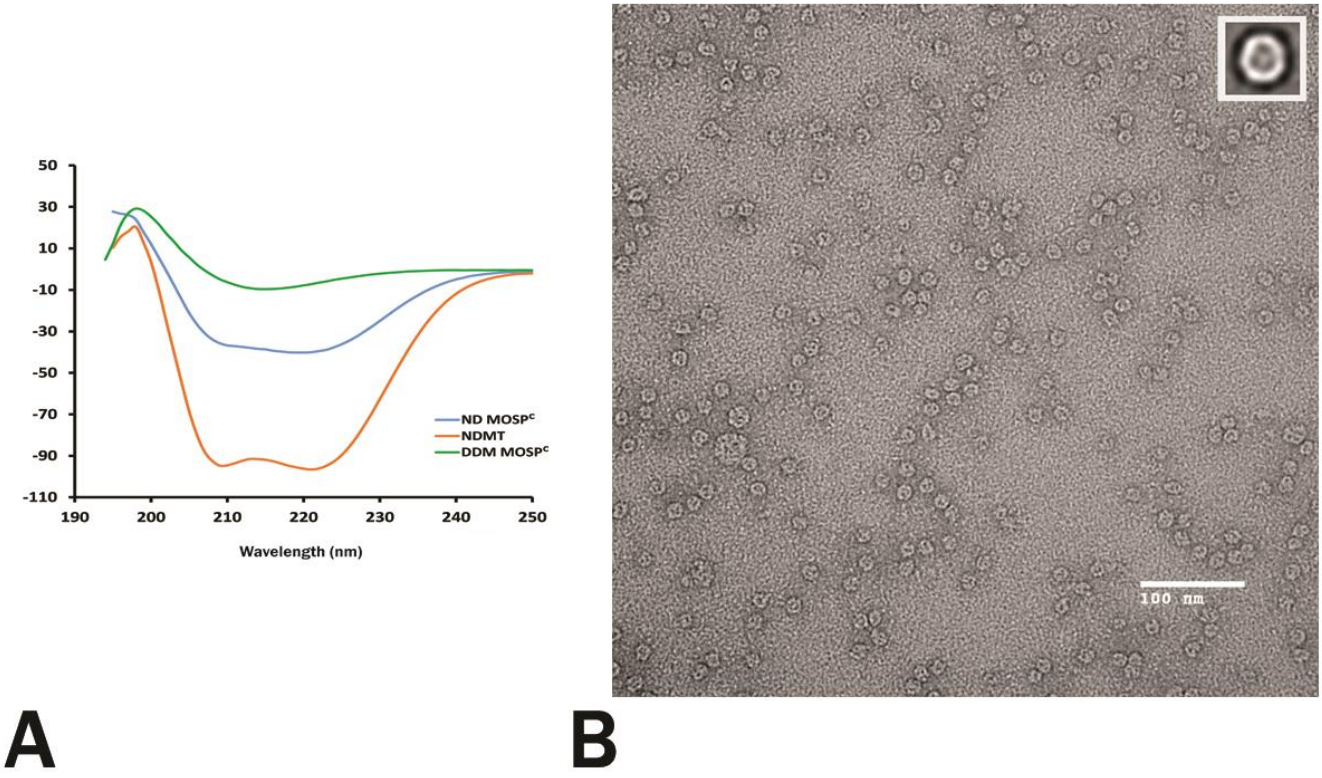


Figure 3: MOSP^{C} in MSP1E3D1 nanodiscs **(A)** Circular Dichroism spectra of refolded MOSP^{C} (green) representing beta barrel structure; empty nanodisc (NDMT) (brown) representing alpha helical structure arising from MSP which forms the outer belt; and MOSP^{C} containing discs which gives an average spectrum demonstrating the presence of both beta barrel and alpha helical structures. **(B)** Negatively stained TEM image of MOSP^{C} discs taken at 150K magnification with a defocus of ~1.6.

References

1. Yang, L. and M. Glaser, *Formation of membrane domains during the activation of protein kinase C.* Biochemistry, 1996. **35**(44): p. 13966-74.
2. Shaw, A.W., et al., *The local phospholipid environment modulates the activation of blood clotting.* The Journal of biological chemistry, 2007. **282**(9): p. 6556-63.
3. Heyn, M.P., R.J. Cherry, and N.A. Dencher, *Lipid--protein interactions in bacteriorhodopsin--dimyristoylphosphatidylcholine vesicles.* Biochemistry, 1981. **20**(4): p. 840-9.
4. Bayburt, T.H. and S.G. Sligar, *Self-assembly of single integral membrane proteins into soluble nanoscale phospholipid bilayers.* Protein science : a publication of the Protein Society, 2003. **12**(11): p. 2476-81.
5. Denisov, I.G., et al., *Cooperativity in cytochrome P450 3A4: linkages in substrate binding, spin state, uncoupling, and product formation.* The Journal of biological chemistry, 2007. **282**(10): p. 7066-76.
6. Duan, H., et al., *Co-incorporation of heterologously expressed Arabidopsis cytochrome P450 and P450 reductase into soluble nanoscale lipid bilayers.* Archives of biochemistry and biophysics, 2004. **424**(2): p. 141-53.
7. Kijac, A.Z., et al., *Magic-angle spinning solid-state NMR spectroscopy of nanodisc-embedded human CYP3A4.* Biochemistry, 2007. **46**(48): p. 13696-703.
8. Alami, M., et al., *Nanodiscs unravel the interaction between the SecYEG channel and its cytosolic partner SecA.* The EMBO journal, 2007. **26**(8): p. 1995-2004.
9. Leitz, A.J., et al., *Functional reconstitution of Beta2-adrenergic receptors utilizing self-assembling Nanodisc technology.* BioTechniques, 2006. **40**(5): p. 601-2, 604, 606, passim.
10. Whorton, M.R., et al., *A monomeric G protein-coupled receptor isolated in a high-density lipoprotein particle efficiently activates its G protein.* Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(18): p. 7682-7.
11. Mi, L.Z., et al., *Functional and structural stability of the epidermal growth factor receptor in detergent micelles and phospholipid nanodiscs.* Biochemistry, 2008. **47**(39): p. 10314-23.
12. Gao, Y., et al., *TRPV1 structures in nanodiscs reveal mechanisms of ligand and lipid action.* Nature, 2016. **534**(7607): p. 347-51.
13. Gluck, J.M., et al., *Integral membrane proteins in nanodiscs can be studied by solution NMR spectroscopy.* Journal of the American Chemical Society, 2009. **131**(34): p. 12060-1.
14. Raschle, T., et al., *Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs.* J Am Chem Soc, 2009. **131**(49): p. 17777-9.

15. Yu, T.Y., et al., *Solution NMR spectroscopic characterization of human VDAC-2 in detergent micelles and lipid bilayer nanodiscs.* Biochimica et biophysica acta, 2012. **1818**(6): p. 1562-9.
16. Shenkarev, Z.O., et al., *NMR structural and dynamical investigation of the isolated voltage-sensing domain of the potassium channel KvAP: implications for voltage gating.* Journal of the American Chemical Society, 2010. **132**(16): p. 5630-7.
17. Park, S.H., et al., *Nanodiscs versus macrodiscs for NMR of membrane proteins.* Biochemistry, 2011. **50**(42): p. 8983-5.
18. Etzkorn, M., et al., *Cell-free expressed bacteriorhodopsin in different soluble membrane mimetics: biophysical properties and NMR accessibility.* Structure, 2013. **21**(3): p. 394-401.
19. Puthenveetil, R. and O. Vinogradova, *Optimization of the design and preparation of nanoscale phospholipid bilayers for its application to solution NMR.* Proteins, 2013. **81**(7): p. 1222-31.
20. Tzitzilonis, C., et al., *Detergent/nanodisc screening for high-resolution NMR studies of an integral membrane protein containing a cytoplasmic domain.* PloS one, 2013. **8**(1): p. e54378.
21. Susac, L., R. Horst, and K. Wuthrich, *Solution-NMR characterization of outer-membrane protein A from E. coli in lipid bilayer nanodiscs and detergent micelles.* Chembiochem : a European journal of chemical biology, 2014. **15**(7): p. 995-1000.
22. Hagn, F., et al., *Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins.* Journal of the American Chemical Society, 2013. **135**(5): p. 1919-25.
23. Kucharska, I., et al., *Optimizing nanodiscs and bicelles for solution NMR studies of two beta-barrel membrane proteins.* Journal of biomolecular NMR, 2015. **61**(3-4): p. 261-74.
24. Zhang, M., et al., *Reconstitution of the Cytb5-CytP450 Complex in Nanodiscs for Structural Studies using NMR Spectroscopy.* Angewandte Chemie, 2016. **55**(14): p. 4497-9.
25. de Vries, S.J., M. van Dijk, and A.M. Bonvin, *The HADDOCK web server for data-driven biomolecular docking.* Nat Protoc, 2010. **5**(5): p. 883-97.
26. Mazhab-Jafari, M.T., et al., *Oncogenic and RASopathy-associated K-RAS mutations relieve membrane-dependent occlusion of the effector-binding site.* Proc Natl Acad Sci U S A, 2015. **112**(21): p. 6625-30.
27. Yang, J., et al., *Structure of an integrin {alpha}IIb{beta}3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation.* Proc Natl Acad Sci U S A, 2009. **106**(42): p. 17729-34.
28. Li, R., et al., *Oligomerization of the integrin alphaIIbbeta3: roles of the transmembrane and cytoplasmic domains.* Proc Natl Acad Sci U S A, 2001. **98**(22): p. 12462-7.
29. 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.
30. Deshmukh, L., et al., *Tyrosine phosphorylation as a conformational switch: a case study of integrin beta3 cytoplasmic tail.* J Biol Chem, 2011. **286**(47): p. 40943-53.

31. Tommassen, J., *Assembly of outer-membrane proteins in bacteria and mitochondria.* Microbiology, 2010. **156**(Pt 9): p. 2587-96.
32. Lafond, R.E. and S.A. Lukehart, *Biological basis for syphilis.* Clinical microbiology reviews, 2006. **19**(1): p. 29-49.
33. Anand, A., et al., *TprC/D (Tp0117/131), a trimeric, pore-forming rare outer membrane protein of Treponema pallidum, has a bipartite domain structure.* Journal of bacteriology, 2012. **194**(9): p. 2321-33.
34. Anand, A., et al., *Bipartite Topology of Treponema pallidum Repeat Proteins C/D and I: OUTER MEMBRANE INSERTION, TRIMERIZATION, AND PORIN FUNCTION REQUIRE A C-TERMINAL beta-BARREL DOMAIN.* The Journal of biological chemistry, 2015. **290**(19): p. 12313-31.
35. Anand, A., et al., *The major outer sheath protein (Msp) of Treponema denticola has a bipartite domain architecture and exists as periplasmic and outer membrane-spanning conformers.* Journal of bacteriology, 2013. **195**(9): p. 2060-71.

Chapter 4

Identifying the role of PLIC-2 in proteasomal degradation and cytoskeletal regulation

Adapted from:

Investigation of the adaptor protein PLIC-2 in multiple pathways.
Nguyen, K.#, Puthenveetil, R.#, and Olga Vinogradova (2017).
Biochemistry and Biophysics Reports 9, 341-348

 # Equal contribution

Introduction

PLIC (or Protein Linking IAP and Cytoskeleton) proteins, also referred to as Ubiquilins, are human orthologs of the yeast Dsk2 family of proteins [1]. 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* [2] and *Jpred* [3]containing STI1 (stress inducible) heat-shock protein binding motifs [4, 5]; and a carboxyl-terminal ubiquitin-associated (UBA) domain (Figure 1A) [6]. PLIC-1 and 2 are localized to the cytoplasm [7], PLIC-3 [8] is found in the testis and PLIC-4 is localized to the cell nucleus [9]. PLICs were identified and subsequently named based on their ability to mediate an association between CD47 and Vimentin, an intermediate filament [10], forging a link between the outer membrane and cell cytoskeleton. CD47 (cluster of differentiation 47) is a ubiquitously present cell surface receptor. It is also known as Integrin associated protein (IAP) since it was initially identified through its association with integrin αvβ3 [11]. CD47 is primarily a cell surface receptor for thrombospondin-1 (THBS1) and signal regulatory protein-α (SIRPα). The engagement of CD47 with SIRPα leads to a “don’t eat me” signal that prevents cells from undergoing phagocytosis [12]. CD47 consists of an extracellular, glycosylated N-terminal IgV-like domain, five predicted transmembrane (TM) helices, and a short C-terminal cytoplasmic tail (CT) [13]. This tail is further alternatively spliced to produce four isoforms (Figure 1B) with differential tissue expression [14]. Wu et al. had previously reported that PLIC-1 and 2 directly interact with the cytoplasmic tails of CD47. This interaction eventually enables Integrins to indirectly engage with the cell cytoskeleton and modulate cell spreading.

Since then, these complexes remain uncharacterized and the exact details of their binary interfaces remains largely unknown.

The ubiquitin mediated proteasomal degradation pathway is a controlled mechanism for the proteolysis of cellular proteins. Key to this system is the comradery of three enzymes E1, E2 and E3; together they prime a protein for degradation through the addition of a polyubiquitin chain. The primary machinery of this mechanism resides in a 26S proteasomal complex, where target proteins meet their terminal fate. This complex is compartmentalized into two components: the 19S regulatory complex and the 20S core complex [15]. S5a/Rpn10, a subunit of the 19S component has been shown to bind polyubiquitin chains. Additionally, and important to this study, are a class of proteins containing an N-terminus UBL (Ubiquitin Like) domain and a C-terminus UBA (Ubiquitin associated) domain. The UBL domain reversibly binds to the proteasome while the UBA domain binds to poly-ubiquitin chains. These UBL-UBA proteins shuttle their payloads to the 26S proteasome for degradation (Figure 1C (i)). It has been shown that the S5a subunit of the 19S component binds to the UBL domain of hHR23a (mammalian homolog of Rad23) [16] and hPLIC-2 (mammalian homolog of Dsk-2) [17].

PLIC proteins derive its name from their purported role of being an adaptor protein linking the cell surface to the cell cytoskeleton. However, it has also been shown to be a proteasome shuttle factor for polyubiquitinated proteins, thereby regulating protein turnover [18]. It is puzzling 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. 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 [19]; and stabilize intracellular $GABA_A$ receptors to promote their accumulation within the plasma membrane [20]. Interestingly, PLIC-2, but not PLIC-1, down regulates endocytosis of the two unrelated GPCRs, V2 vasopressin receptor and beta-2 adrenergic receptor [21]. Though these studies highlight PLICs involvement in diverse cellular functions as a key regulatory molecule, they seem to further contradict its primary discovered role as a structural adaptor protein linking IAP/CD47 to cell cytoskeleton.

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. First, we focused on the UBA domain of PLIC-2; since the structure of the UBL domain was already solved (PDB: 1J8C) and its interaction with the proteasome complex was clearly established [17]. Our data shows that there are inter-domain interactions within PLIC-2 where UBA binds weakly to the UBL domain which is further disrupted by its stronger binding affinity for Ubiquitin. Second, we sought to ascertain the primary binding interface between the cytoplasmic tail of CD47 and PLIC-2 as was initially observed [7]. We envisioned this interaction to occur via the three separate domains of PLIC-2 as demonstrated in Figure 1C (ii-a and ii-b). We approached this problem in two ways, (i) we investigated the interaction of just the soluble cytoplasmic tail (CT) with the separate domains of PLIC-2 and (ii) we introduced the last transmembrane helix along with its cytoplasmic tail into nanodiscs to decipher any binding that might occur at the membrane/water interface in a near native membrane environment. Despite our rigorous multi-pronged approach, we were, quite disappointingly, unable to observe any specific

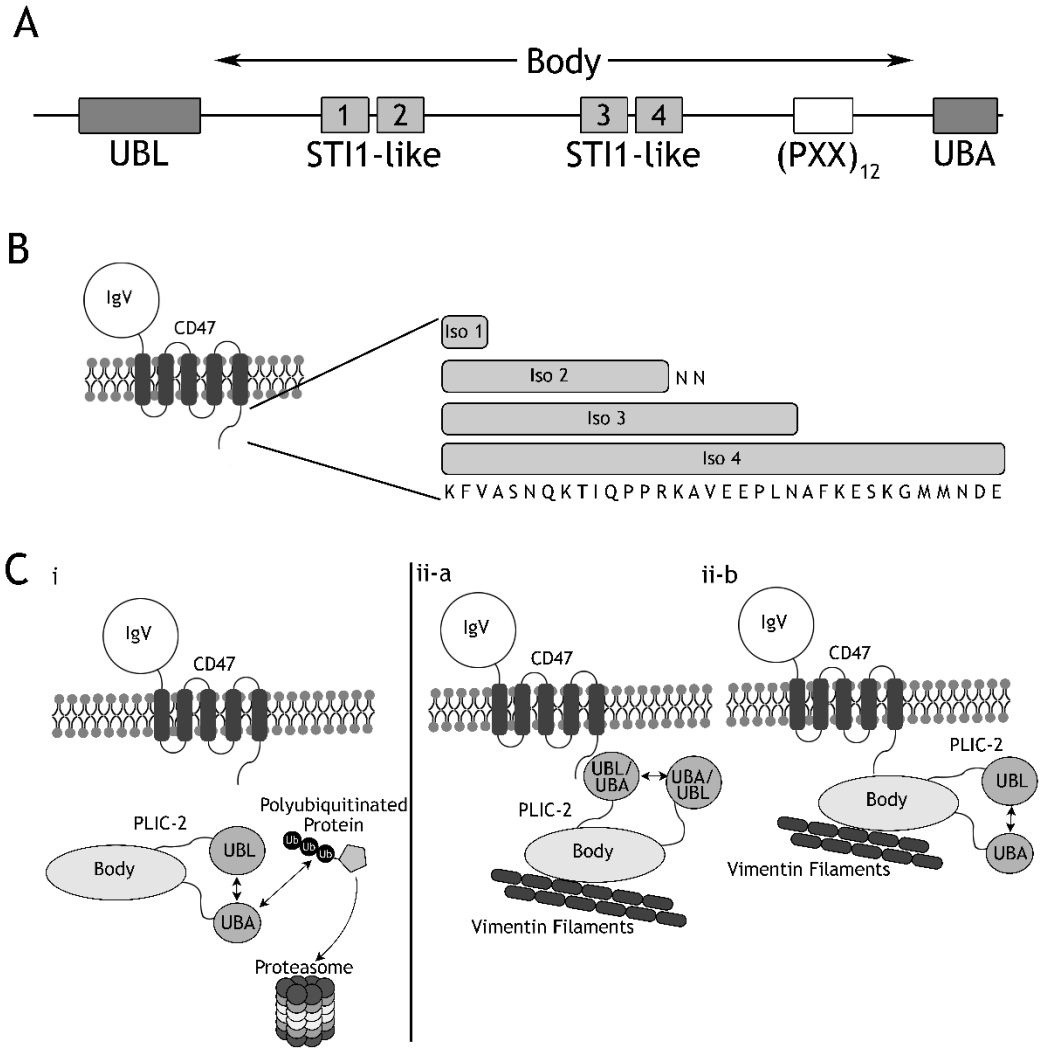


Figure 1: (A) Schematic showing the different domains and motifs of PLIC-2. It contains a well characterized N-terminal UBL (ubiquitin like) domain and a C-terminal UBA (ubiquitin associated) domain. Interspersed between the two is the Body region which contains four Sti1 like domains and a repeat of 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 [7]. 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.

interactions between the two proteins using both ITC and NMR. A full report of our findings is discussed in this chapter.

Materials and Methods

Cloning of PLIC-2 and CD47constructs

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 [7]. 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.

Nanodiscs Preparation and Transmission Electron Microscopy (TEM)

Nanodiscs were prepared from a smaller truncated construct (D7) of MSP1D1 previously developed in our lab [22] as smaller discs are more conducive for NMR investigations. D7 was purified in the same manner as other MSP proteins forming nanodiscs [23]. D7 was mixed with purified ^{15}N-CD47-TMCT in a molar ratio of two to one along with a twenty-fold molar excess of DMPC lipids (Avanti polar, USA) solubilized in a buffer containing sodium cholate. The reaction mixture was mixed for an hour at room temperature and later on mixed with wet Bio-beads SM2 (BIO-RAD, USA) for four hours. The resultant solution was loaded onto Superdex 200, to separate empty discs from protein incorporated ones. The MBP tag on the incorporated protein was cleaved using TEV protease and repurified on S-200 column to obtain discs containing the incorporated trans-membrane and cytoplasmic tail of CD47 Isoform-4. For TEM, ^{15}N-CD47 TMCT nanodiscs were diluted several fold in water to obtain a well-dispersed population, applied to a glow discharged carbon-coated 400-mesh copper grid (Ted Pella Inc., USA), and negatively stained with freshly prepared 0.75% uranyl formate. Images were taken on a Tecnai G2 Spirit BioTWIN microscope (FEI, USA) at an accelerating voltage of 80 kV with a defocus of ~−1.6.

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 [24] and analyzed in CcpNmr software suite [25] 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 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.

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, 300s injection intervals, and 2.5µL 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: 0.35 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 [28] 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) [29] 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).

Results

Solution NMR has been successfully applied to study protein interactions for numerous intrinsically disordered and membrane proteins [30-33]. In our present study, we use NMR to investigate the interaction of PLIC-2 in two different pathways. PLIC-2 was chosen as a candidate over PLIC-1 since PLIC-1 exists in four isomeric forms (Accession No. Q9UMX0), potentially making the analysis cumbersome and inconclusive. PLIC-2 was compartmentalized into its individual domains: UBL, Body, and UBA. Each domain was used to assess their binding ability to its different partners.

PLIC-2 (UBA) – PLIC-2 (UBL) interaction

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 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, were used to calculate the Kd values using the equations summarized in the methods section. The average Kd value obtained was 175±25 µM, which is comparable to the previously reported Kd 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 [6].

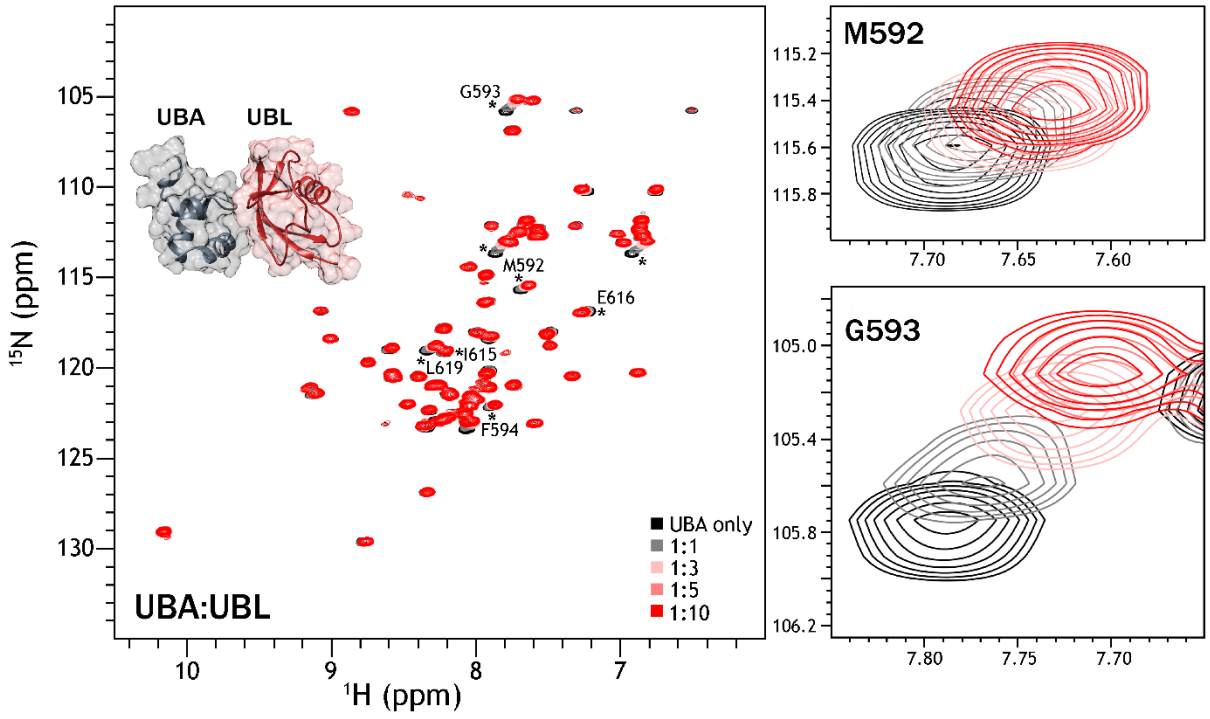


Figure 2: ^{15}N-HSQC titration experiments of PLIC-2 ^{15}N-UBA domain in the presence of PLIC-2 UBL domain at a UBA to UBL molar ratio of 1:1, 1:3, 1:5, and 1:10. Partial amino acid assignments were achieved by aligning the ^{15}N-HSQC to that of PLIC-1 UBA domain [35]. The inset shows a molecular 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.

PLIC-2 (UBA) – Ubiquitin interaction

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 3). The Kd 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 (Kd ~20 µM) [35]. Amide resonances were

identified based upon assignments available for the almost identical UBA domain from PLIC-1 (PDB: 2JY5). 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 and 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. 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.

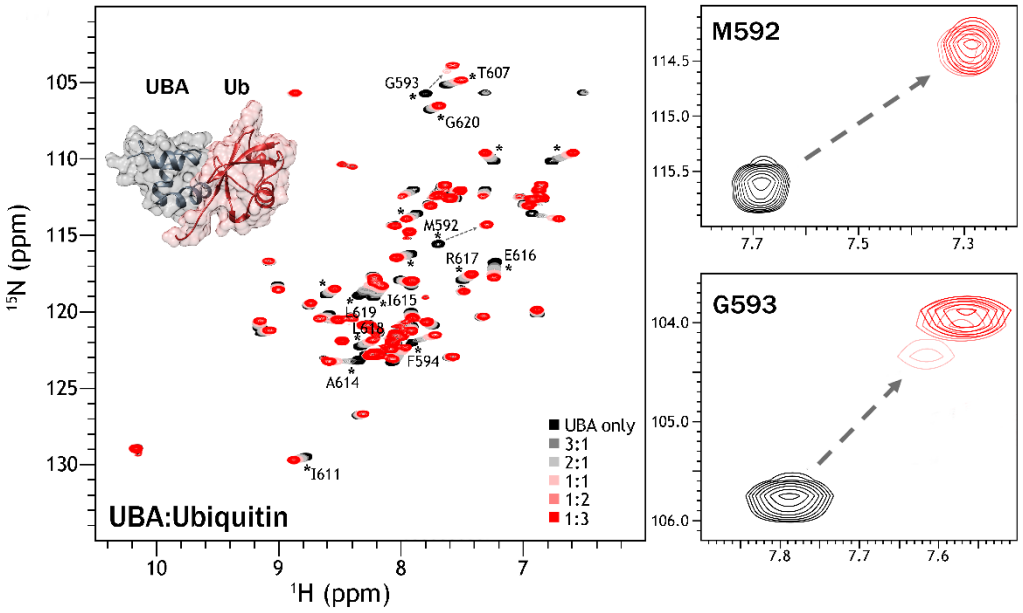


Figure 3: ^{15}N-HSQC titration experiments of PLIC-2 ^{15}N-UBA domain in the presence of ubiquitin (Ub) at a UBA to Ub molar ratios of 3:1, 2:1, 1:1, 1:2, and 1:3. Ratios 1:5 and 1:10 are not shown as the spectra reached a saturation at 1:3. A molecular model, determined by HADDOCK, of UBA PLIC-2 (grey) 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.

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

CD47 (TMCT) – PLIC-2 (UBA/UBL) interaction

A 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 [22]. Nanodiscs offer a soluble lipid bilayer system allowing the study of interactions at the membrane/water interface. We have previously successfully demonstrated the application of the TMCT regions of Integrin β_3 in nanodiscs [36]. The formation of discs were confirmed through TEM (Figure 5A, inset). . ^{15}N-HSQC spectra were collected for labeled ^{15}N-UBL (Figure. 4A) and ^{15}N-UBA (Figure. 4B) in the presence of unlabeled CD47-TMCT followed by a further evaluation through ITC. No perceptible

perturbations/peak broadenings were observed which was further corroborated by the ITC data where the heat differences were at the noise level for both UBL (Figure. 4C) and

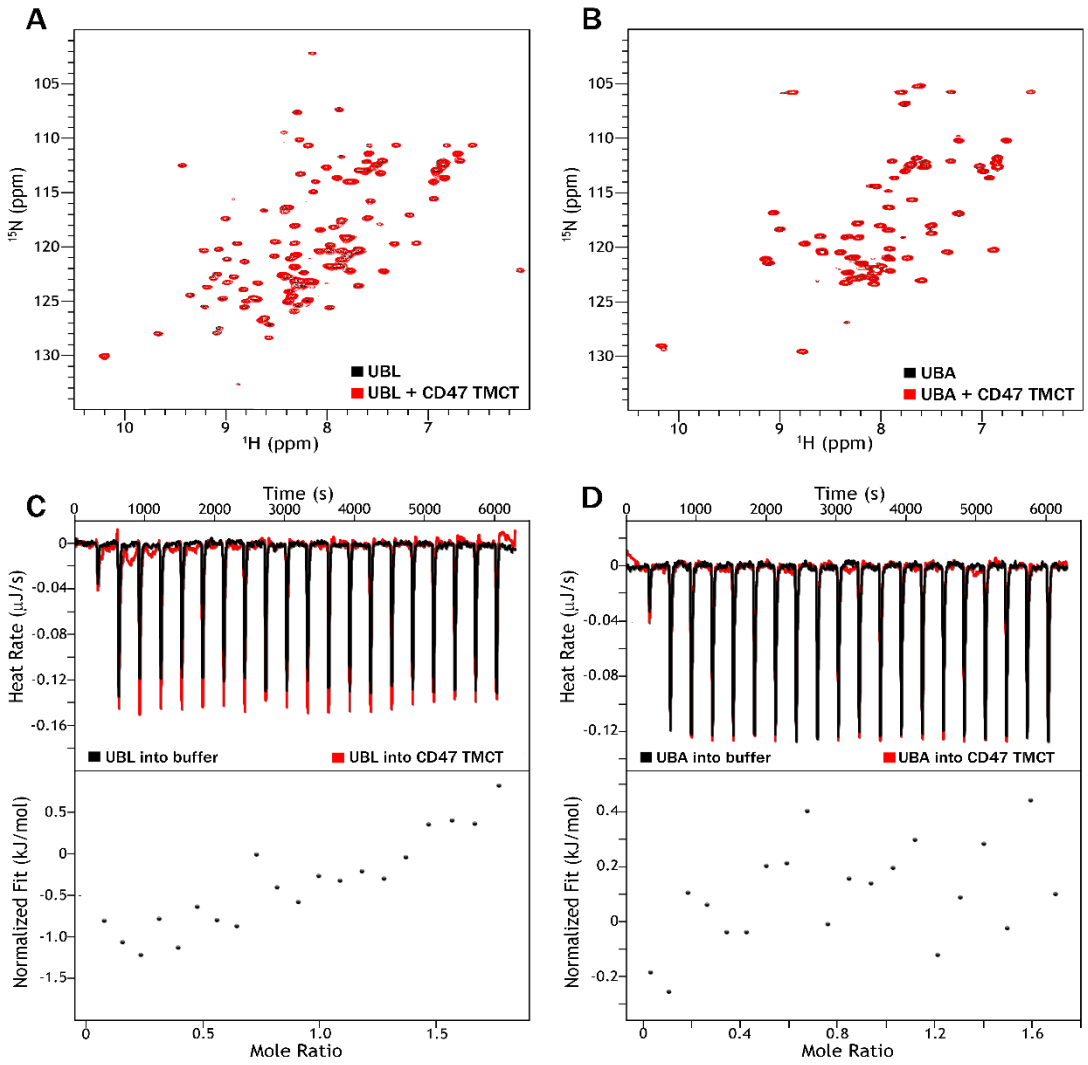


Figure 4: (A) ^{15}N-HSQC of PLIC-2 ^{15}N-UBL alone (black) and in the presence of CD47-TMCT at a UBL to CD47 ratio of 1:5 in red. **(B)** ^{15}N-HSQC of PLIC-2 ^{15}N-UBA alone (black) and in the presence of CD47-TMCT at a UBA to CD47 ratio of 2:5 in red. **(C)** ITC thermogram of UBL titrated into buffer (black) and into CD47-TMCT (red). **(D)** ITC thermogram of UBA titrated into buffer (black) and into CD47-TMCT (red).

UBA (Figure. 4D). The results show no discernable interactions occurring between CD47 and the UBL or UBA domains of PLIC-2.Additionally we also performed ^{15}N-HSQC titration experiments with ^{15}N-UBL or ^{15}N-UBA and unlabeled CD47-CT which also yielded negative results for both NMR and ITC measurements.

CD47 (TMCT) – PLIC-2 (Body region) interaction

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, which 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 5A) and UBL (Figure 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 5A); but a minor intensity drop for a single peak does not substantiate a binding interaction between two proteins.

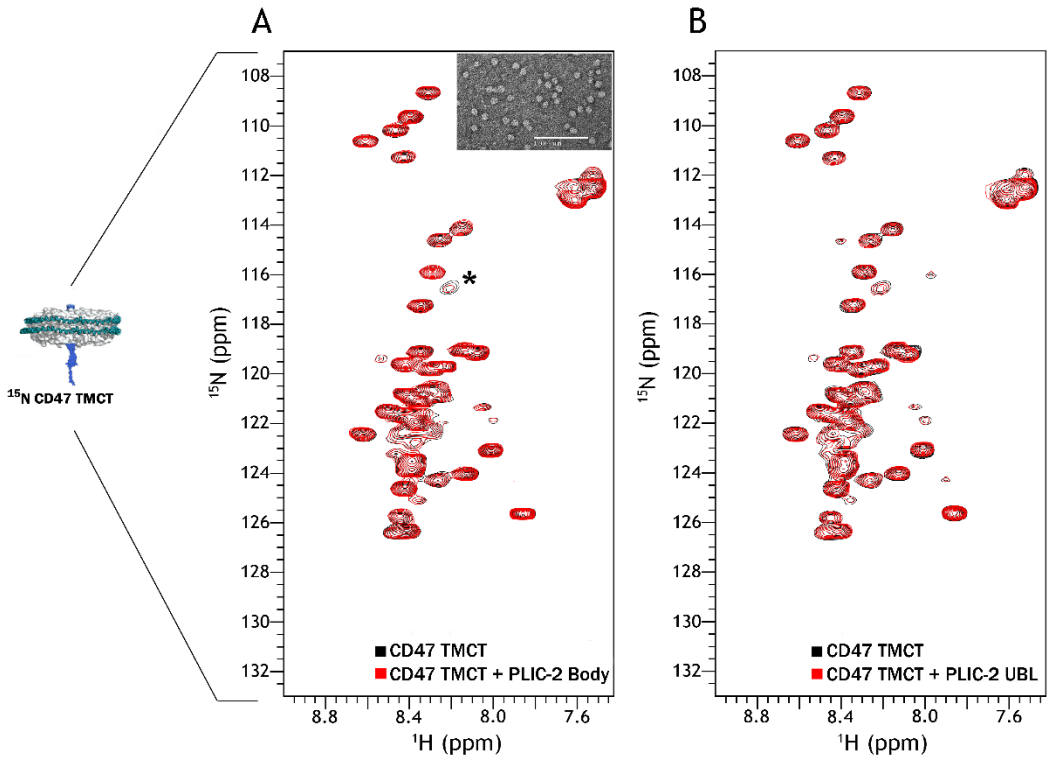


Figure 5: ^{15}N-HSQC titration experiments of ^{15}N-CD47 TMCT. **(A)** ^{15}N-CD47 TMCT alone (black) and in the presence of PLIC-2 Body region at a CD47 to Body ratio of 1:3 in red. The amide peak with a slight intensity drop in the presence of Body is marked by the asterisk. Inset: Negatively stained 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.

Discussion

Since the structure of the UBL domain from PLIC-2 was already available, we focused our attention on the UBA domain of PLIC-2. The human PLIC-2 UBA domain can be contrasted to its ortholog *S. cerevisiae* Dsk2 where similar UBA domain interactions have been observed. The UBA domains of these two proteins do not share much of a sequence homology, but they do have the same fold. Dsk2 UBA forms a dimer or higher order polymers [37], which is attributed to the presence of positive and negatively charged

regions on the surface of the protein. PLIC-2 UBA, in contrast, lacks strong charges and is monomeric, as can be concluded from the SEC data (Figure 6). From reliable partial

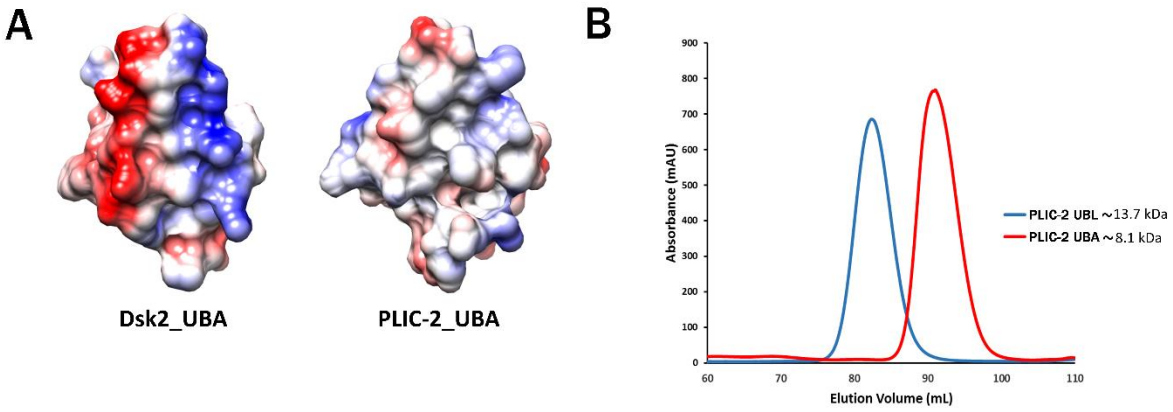


Figure 6: (A) Electrostatic surface potential showing charge distribution on the surface of the UBA domains of Dsk2 and PLIC-2. (B) SEC chromatogram showing the monomeric form of PLIC-2 UBA (~8.1kDa) which elutes at a later volume than the monomeric PLIC-2 UBL (~13.7kDa).

assignments, 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}-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.

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) [38], which is a subunit of the 19S regulatory particle of the 26S proteasome. Here we analyze the binding of the PLIC-2 UBA domain to ubiquitin. From our Kd estimation, the PLIC-2 UBA-Ub binds stronger than UBA-UBL. 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 [39]. 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 and 3. This proves the role of PLIC-2 as a shuttle protein transferring its cargo to the proteasome for degradation.

The latter part of our study, which we hoped would lead to insightful results, was defining the role of PLIC-2 as an adaptor protein linking the outer cell membrane and cell cytoskeleton. The name PLIC is based on its function of being on a Protein that Links IAP (integrin associated protein/CD47) to Cell cytoskeleton, Vimentin, through its cytoplasmic

tail. The only reported record identifying this association comes from a singular paper, where the name PLIC was conferred upon these adaptor proteins [10]. Almost two decades later, we tried to investigate this interaction with the aim of identifying the critical domains and residues involved in the process. Since the binding was shown to occur through the cytoplasmic tail of CD47, we employed two constructs of the tail region: (i) the tail by itself (CT), and (ii) along with the last transmembrane helix (TMCT) to capture interactions that might occur at the membrane water interface. The latter construct was inserted into small nanodiscs [40]. Even though we employed high sensitivity/ high-resolution biophysical technique like NMR, 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. The inference that the two may not be directly associated is not inconceivable when we factor in the following reasons.

First, several studies have shown the involvement of PLICs in alternative cellular processes. In particular, PLIC-2 has been linked to proteasomal degradation pathways [18, 41] and GPCR endocytosis [21], which are seemingly unrelated to cytoskeletal modulation. This potentially means that the purported role of PLIC linking CD47 to vimentin filaments is not its major cellular function. This, in part, is supported by another study where double labeled immune-fluorescence of PLIC-1 and vimentin filaments did not show any co-localization [42]. Additionally, the same group which published the original PLIC-2 paper later found 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 [19]. They also reported that PLIC-2 binds to Eps15 and Epsin (1 and 2), all of which are UIM-containing adaptor proteins involved in clathrin mediated endocytosis and thereby in protein turnover [21]. Second, it's likely that CD47 and PLIC-2 are part of a larger macromolecular assembly that may also contain, for example, integrin heterodimers [43]. 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 [10]. This doesn't rule out the presence of other proteins which 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. Moreover, we were unable to find any interaction between the cytoplasmic tails of Integrin β3 and CD47 (data not shown) which are known to form a complex, most likely associating through their transmembrane region. This further rules out the possibility of CD47's cytoplasmic tail acquiring a secondary structure through its interaction with the tail of Integrin β3 which would then bind to PLIC-2. Considering all the data, we are certain that PLIC-2 does not bind to the cytoplasmic tail of CD47.

In conclusion, we show that PLIC-2 belongs to the proteasomal degradation pathway and it does not interact with the cytoplasmic tail of CD47. 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.

References

1. 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.
2. Ward, J.J., et al., *The DISOPRED server for the prediction of protein disorder.* Bioinformatics, 2004. **20**(13): p. 2138-9.
3. Drozdetskiy, A., et al., *JPred4: a protein secondary structure prediction server.* Nucleic Acids Res, 2015. **43**(W1): p. W389-94.
4. 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.
5. Johnson, B.D., et al., *Hop modulates Hsp70/Hsp90 interactions in protein folding.* J Biol Chem, 1998. **273**(6): p. 3679-86.
6. Elsasser, S. and D. Finley, *Delivery of ubiquitinated substrates to protein-unfolding machines.* Nature cell biology, 2005. **7**(8): p. 742-9.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. Oldenborg, P.A., et al., *Role of CD47 as a marker of self on red blood cells.* Science, 2000. **288**(5473): p. 2051-4.
13. 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.
14. 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.
15. 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.
16. Mueller, T.D. and J. Feigon, *Structural determinants for the binding of ubiquitin-like domains to the proteasome.* The EMBO journal, 2003. **22**(18): p. 4634-45.
17. 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.

18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. Ritchie, T.K., et al., *Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs.* Methods Enzymol, 2009. **464**: p. 211-31.
24. Delaglio, F., et al., *NMRPipe: a multidimensional spectral processing system based on UNIX pipes.* J Biomol NMR, 1995. **6**(3): p. 277-93.
25. Vranken, W., F., et al., *The CCPN data model for NMR Spectroscopy.* Proteins, 2005. **59**(59): p. 687-696.
26. Williamson, M.P., *Using chemical shift perturbation to characterise ligand binding.* Progress in Nuclear Magnetic Resonance Spectroscopy, 2013. **73**: p. 1-16.
27. Fielding, L., *NMR methods for the determination of protein-ligand dissociation constants.* Curr Top Med Chem, 2003. **3**(1): p. 39-53.
28. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis.* Nature protocols, 2015. **10**(6): p. 845-58.
29. 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.
30. Tyukhtenko, S., et al., *Characterization of the neuron-specific L1-CAM cytoplasmic tail: naturally disordered in solution it exercises different binding modes for different adaptor proteins.* Biochemistry, 2008. **47**(13): p. 4160-8.
31. Vinogradova, O., et al., *A structural mechanism of integrin alpha(IIb)beta(3) "inside-out" activation as regulated by its cytoplasmic face.* Cell, 2002. **110**(5): p. 587-97.
32. Yang, J., et al., *Structure of an integrin {alpha}IIb{beta}3 transmembrane-cytoplasmic heterocomplex provides insight into integrin activation.* Proc Natl Acad Sci U S A, 2009. **106**(42): p. 17729-34.
33. Sahu, D., M. Bastidas, and S.A. Showalter, *Generating NMR chemical shift assignments of intrinsically disordered proteins using carbon-detected NMR methods.* Anal Biochem, 2014. **449**: p. 17-25.
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. Puthenveetil, R., Nguyen, K. and Olga Vinogradova, *Nanodiscs and Solution NMR: preparation, application and challenges.* Nanotechnology Reviews, 2016. **under review**.
37. Lowe, E.D., et al., *Structures of the Dsk2 UBL and UBA domains and their complex.* Acta crystallographica. Section D, Biological crystallography, 2006. **62**(Pt 2): p. 177-88.
38. 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.
39. 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.
40. Puthenveetil, R. and O. Vinogradova, *Optimization of the design and preparation of nanoscale phospholipid bilayers for its application to solution NMR.* Proteins, 2013. **81**(7): p. 1222-31.
41. 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.
42. Wu, S., et al., *Characterization of ubiquilin 1, an mTOR-interacting protein.* Biochim Biophys Acta, 2002. **1542**(1-3): p. 41-56.
43. 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.

Chapter 5

The outer membrane and periplasmic conformers of the bipartite MOSP forms multimeric complexes

Adapted from:

The major outer sheath protein in *T. denticola* forms multimeric conformers

Puthenveetil R, Kumar S, Anand A, Dey A, Vinogradova O and Radolf JD (*in prep*)

Introduction

Periodontitis, a chronic inflammation of gingiva, is caused by the overgrowth of complex polymicrobial consortium part of the oral microbiota [1, 2]. Among the several treponemes resident to the oral cavity, *Treponema denticola* has been implicated as an etiological agent of periodontal disease. *T. denticola* along with *Porphyromonas gingivalis* and *Tannerella forsythia* form the core "Red Complex" responsible for the progression of chronic periodontitis; and are commonly isolated from chronic periodontal lesions in subgingival plaques [3]. T. denticola is a gram negative, anerobic spirochete with an unusual outer membrane (OM) devoid of lipopolysaccharide (LPS) and a lipid composition similar to lipoteichoic acid commonly found in gram positive bacteria [4].

Membrane Outer Sheath Protein (MOSP/TDE0405) is the most abundant protein present in the OM and a prominent determinant of virulence in *T. denticola* [2, 5]. MOSP is ubiquitous among oral treponemes with considerable sequence variation among different species [6, 7]. MOSP is involved in a myriad of function, viz., induces cytopathic effect on erythrocytes and epithelial cells [8]; releases metalloprotease by activating neutrophils[9]; modulates the regulation of actin filaments and intracellular calcium in fibroblast [10, 11]. Like any typical outer membrane protein, MOSP has been shown to form pore forming channels [12]. MOSP ($MOSP^{FL}$) exists in a well-defined bipartite topology where the C-terminal domain is membrane embedded ($MOSP^{C}$) with the N-terminus ($MOSP^{N}$) tethered to it [13]. Alternatively, it was previously proposed that $MOSP^{FL}$ might exist exclusively as a single OM porin [14], though there is now evidence of a bipartite classification through the Conserved Domain Database (CDD) [15] and Pfam

[16]; along with biochemical evidence through the demonstration of porin activity by $MOSP^C$ in liposomes [13]. Treponema pallidum repeat (Tpr) proteins found in *T. pallidum* subspecies *pallidum* are a collection of proteins (Trp A-L) that are orthologs of MOSP from *T.denticola* [17-19]. Tprs are cognate proteins that mostly harbor the canonical bipartite domain architecture with the exception of TprF which only contains the $MOSP^N$ domain and is localized to the periplasm [20]. Sequence variability and cellular localization of TprK, containing both domains, remains a point of contention among different research groups [21-24]. TprC/D/I are confirmed bonafied OM paralogs with a bipartite topology and porin activity [20, 25]. Bipartite Tprs (B, E, G, H, J, and L) have been proposed to be porins though presently there is no experimental evidence to support the claim [18, 19]. *T. denticola* serves as a predecessor to *T. pallidum,* which is believed to have evolved by reductive evolution [26, 27]. Taken together, MOSP could serve as a common parental ortholog to the different Tpr proteins.

For the progression of periodontitis, it is important for the bacteria to penetrate the periodontal tissue. *T. denticola* contains several proteases [28] which may aid in this process. Two such well-studied proteases are the trypsin like protease and dentilisin. Two trypsin like proteases (OpdB) (TDE2140 and TDE1195) have been identified. TDE2140 has been shown to cleave C-terminal to the Arg residue [29] while TDE1195, a putative OpdB, is postulated to be a lysine specific protease [28]. Dentilisin is a protease present in the outer membrane of *T. denticola* and is speculated to help penetrate epithelial cells by degrading intracellular adhesion proteins [30]. Dentilisin is a complex of several proteins PrtP, PrcA1, PrcA2 and PrcB. PrtP is a protease [31]; PrcA is hydrolyzed to give PrcA1 and PrcA2 [32] and PrcB associates with PrtP [33]. The biological roles of PrcA

and PrcB remain unknown. MOSP has been shown to associate only with the PrcA2 component of dentilisin and all components have been implicated to be surface exposed [34]. PrtP has also been shown to be essential for the cleavage of factor H [35]. Proteolytic cleavage of periodontal tissue will provide a plethora of peptides that may serve as a source of nutrient for *T. denticola*. Central to peptide intake in bacteria is a family of ABC transporters called the Opp system [36]. OppA is the key peptide binding protein which transfers the peptide to the remainder of the Opp system to be transported intracellularly through an active intake. *T. denticola* has several OppA proteins specific for peptides varying in their length. A 70 kDa OppA was previously identified as a surface exposed lipoprotein anchored to the outer leaflet of the OM in *T. denticola.* The OppA gene was found contiguous to the permease and nucleotide binding domain jointly forming the classical oligopeptide permease system found in bacteria. Contrary to its expected role of oligopeptide transport, OppA in *T.denticola* was proposed to be an adhesin and help in the delay of immunological recognition [37].

In this study, we show that MOSP exists both as a membrane and periplasmic conformer each forming individual multimeric complexes. MOSP belonging to both conformers exists as stable trimers with each conformer forming further multimeric complexes. Mass spec analysis has shed some light on these potential binding partners. We also provide an indirect evidence towards the binding of the periplasmic conformer to the peptidoglycan sacculus. Contrary to MOSP binding to just the PrcA2 component of dentilisin [34] we shown that dentilisin binds in its entirety to MOSP. We further demonstrate that the dentilisin-MOSP association is weak. Surface localization experiments highlight the membrane accessibility of PrcA2 and PrcB while PrtP and

PrcA1 remain inaccessible, differing from the previous finding where they are all surface accessible [34]. Additionally, we provide direct evidence for the formation of trimeric pores by recombinant $MOSP^C$ in artificial nanoscale phospholipid bilayer system called nanodiscs while recombinant $MOSP^N$ exists as a monomer with an extended conformation. Finally we also expressed MOSP in *E. coli* using a PelB signal sequence. Unlike *T. denticola*, in *E. coli* MOSP was recovered only with OMs where it forms trimers maintaining the bipartite topology as was observed in *T. denticola*. Based on our results, MOSP forms stable trimers through its $MOSP^C$ domain while the extended $MOSP^N$ may facilitate its association with multiple binding partners. The multimeric conformers of MOSP is a feature unique to *T. denticola* whose periplasmic milieu might facilitate the uptake of complex nutrients from the host tissue. In addition it also sheds light on a unique OMP biogenesis pathway through the bifurcation of MOSP into two cellular locales.

Materials and methods

Ethics Statement: Animal protocols described in this work strictly follow the recommendations of the Guide for Care and Use of Laboratory Animals of the National Institutes of Health and were approved by the University of Connecticut Health Center Animal Care Committee under the auspices of Animal Welfare Assurance A347-01.

Propagation of *Treponema denticola*: *T. denticola* (ATCC 35405) was cultivated to their mid to late log phase of the growth curve (2×10^8-10^9) in new oral spirochete broth (NOS) [38] supplemented with 10% heat inactivated normal rabbit serum. The culture was kept at 37°C in a closed chamber with a GasPak (Becton, Dickinson, Cockeysville, USA) providing an anaerobic environment.

Cloning of MOSP, TprF and SKP genes: Recombinant $MOSP^{FL}$ (TDE0405), $MOSP^{N}$, $MOSP^{C}$ from *T. denticola* (ATCC 35405) and TprF from *T. pallidum* (Nichols) were cloned into pET23b, without the signal sequence, using the protocol previously described in Anand et al. [13, 20]. Additionally $MOSP^{FL}$ was subcloned into pET20b (EMD Millipore, USA) vector using the restriction sites BamHI and HindIII. The resultant protein contained an N-terminal pelB signal sequence which facilitated the incorporation of $MOSP^{FL}$ into the outer membrane of *E. coli* (K12). DNA encoding *T. denticola* SKP (TDE2602) was PCR-amplified and cloned into the BamHI and HindIII restriction sites of pET23b (Novagen, USA).

Expression, purification and refolding of recombinant proteins: All proteins were expressed in *E.coli* (DE3) (Agilent, USA) at 37°C. Proteins were induced with 1mM IPTG once they reached a cell density of 0.5. Post induction cells were harvested at 8,000 rpm for 20 min at 4°C. The pellet was lysed in Tris-NaCl buffer (20mM Tris, 150mM NaCl, pH 8) supplemented with protease inhibitor cocktail (PIC; SIGMA, USA). The lysate was centrifuged at 12,000 rpm for 40 min at 4°C to separate insoluble proteins and cell debris. The supernatant was mixed with Ni-NTA resin (QIAGEN, USA) for an hour and washed with Tris-NaCl buffer followed by another wash with the same buffer containing 20mM Imidazole. Elution was carried out in Tris-NaCl buffer with 400mM Imidazole at pH 8. For denatured protein purification, the insoluble pellet from cell lysate was dissolved in Tris-NaCl buffer containing 6M GuHCL and left on a shaker for two hours at room temperature. Any insoluble material was separated by centrifugation at 12,000 rpm for 40 min at 4°C and the supernatant was mixed with Ni-NTA resin for two hours. The resin was washed in Tris-NaCl buffer containing 8M Urea, pH 8.0 and further washed with the same buffer

containing 20mM Imidazole. Elution was carried out in Tris-Urea buffer (20mM Tris, 8M Urea, 400mM Imidazole, pH 8).

Refolding of denatured $MOSP^{C:}$ Eluate from the $MOSP^{C}$ purification was concentrated to 0.1mM and refolded in a dropwise fashion into 20 fold excess of 1.5% DDM (n-Dodecyl-β-D-Maltoside; Anatrace, USA) dissolved in 15mM Tris buffer at pH 9.0 with 2mM EDTA. Refolding was carried out at room temperature and its efficiency was accessed using CD spectroscopy. Refolding of denatured $MOSP^{N}$: Eluate from $MOSP^{N}$ purification was concentrated to 0.05mM and refolded through dialysis in a stepwise manner using a gradient of buffer containing (8,6,4,2,1,0.5 and 0M Urea) in 20mM Tris, 250mM NaCl, 5% glycerol, 2mM EDTA with PIC, pH 8.5 at 4°C. Following refolding the protein was further purified through Size Exclusion Chromatography (SEC) on a S200 column (GE, USA) equilibrated with Tris-NaCl buffer, pH 7.5 to separate high molecular weight oligomers form the refolded monomers.

Far-UV circular dichroism spectroscopy: Far-UV circular dichroism (CD) spectroscopy was performed using a JASCO J-715 spectral polarimeter as described previously [13].

Nanodiscs preparation and Transmission Electron Microscopy: Small and large nanodiscs were prepared using a truncated construct (D7) obtained from MSP1D1 previously developed in our lab [39] and MSP1E3D1. Both D7 and E3D1 in pET28a vector were purified using the protocol developed by Sligar's group [40]. Nanodiscs were formed in the presence of DMPC (1, 2-dimyristoyl-sn-glycero-3-phosphocholine; Avanti Polar, USA) lipids at room temperature. D7 and E3D1 were mixed with detergent refolded $MOSP^{C}$ in a D7/E3D1 to $MOSP^{C}$ ratio of 2:1. DMPC lipids were added to this mixture in a molar excess of 20 for D7 and 100 for E3D1. The reaction mixture was kept on a shaker

for an hour, after which SM2 Bio-Beads (Bio-Rad, USA) were added and left overnight gently suspending the beads in the solution. Beads were filter separated from the reaction mixture and further subjected to SEC on S200 column (GE, USA) equilibrated with Tris-NaCl buffer at pH 7.5. Peak fractions containing $MOSP^C$ incorporated discs were used for Transmission Electron Microscopy (TEM).

$MOSP^C$ containing discs were diluted several fold in water to obtain a well-dispersed population, applied to a glow discharged carbon-coated 400-mesh copper grid (Ted Pella Inc., USA), and negatively stained with freshly prepared 0.75% uranyl formate (SPI-Chem, USA). Images were taken on a Tecnai G2 Spirit BioTWIN microscope (FEI, USA) at an accelerating voltage of 80 kV with a defocus of ~−1.6. Class averaging was performed by averaging 70-80 discs/image using XMIPP [41].

Analytical Ultracentrifugation: Sedimentation velocity experiments were performed using a Beckman-Coulter XL-I analytical ultracentrifuge at 20°C and 40,000 rpm. Samples containing approximately 0.05mM recombinant $MOSP^N$ and TprF in Tris-NaCl buffer pH 7.5 were loaded into double sector cells equipped with 1.2 cm Spin60 charcoal-epon centerpieces (Spin Analytical, Berwick, ME) and quartz windows. The raw scans were recorded using absorbance optics at 280 nm. The buffer densities and viscosities were calculated using SEDNTERP [42]. Initially, the sedimentation velocity data were analyzed using the time derivative method [43] in DCDT+ [44] to ensure that the samples contained homogenous single species. The data was then fit to a non-interacting discrete single species model using a c(s) distribution model with Sedfit [45] to obtain the sedimentation coefficients and the buoyant molar masses of the proteins.

Triton X-114 phase partitioning: 10µg of recombinant proteins or 2x 10^9 cells of *T. denticola* were added to 2% Triton X-114 (TX-114) in PBS supplemented with 0.5% PIC and incubated overnight at 4°C. The insoluble fraction was separated by centrifugation at 12,000 rpm for 20 min. at 4°C while the supernatant was further subjected to phase separation. The detergent-enriched and aqueous phases were extracted four times with PBS and 2% TX-114, respectively [46]. All samples were precipitated overnight with 10 volumes of acetone at -20°C and subsequently analyzed through SDS-PAGE and immunoblotting.

Preparation of Membrane and Aqueous conformers of MOSP from *T. denticola*: In order to obtain Detergent/ Membrane conformer of MOSP, freshly harvested *T. denticola* (2x10^9 cells/ml) were dissolved in IP Lysis buffer (PIERCE, USA) and left on a shaker for two hours at 4°C. This buffer contains 1% NP-40 detergent which aids in selective solubilization of the outer membrane of *T. denticola*. The detergent solubilized mixture was subjected to ultracentrifugation at 100,000 xg for 45 min at 4°C. The supernatant was clarified on a Superdex S200 (10/300) column (GE, USA) where the MOSP containing fraction was obtained in the void volume separating it from empty micelles and degradative products. The Soluble/Aqueous conformer of MOSP was obtained through TX-114 phase partitioning, where several aqueous phases were pooled, concentrated and clarified through S200 column obtaining the higher molecular weight MOSP in the void volume. For DDM solubilized membrane fraction, *T.denticola* cells were resuspended in Tris buffer (20 mM Tris-HCl, 100 mM NaCl, pH 8.0) and disrupted by sonication for three 20 s pulses interspersed with 30 s of rest on ice. The lysate was centrifuged at 6000 xg for 20 min at 4°C to remove cell debris and further fractionated

using ultracentrifugation at 100,000 xg for 45 min at 4°C. The pellet containing the membrane fraction was solubilized in 2% DDM and used for TX-114 phase partitioning.

Blue Native-polyacrylamide gel electrophoresis (BN-PAGE): BN-PAGE of recombinant and native proteins was performed as previously described [25].

Immunologic reagents and Immunoblot analysis: Rat polyclonal antisera directed against MOSP and isolated *T. denticola* endoflagellar filaments were described previously [47]. Rat antisera directed against Msp^N, Msp^C, and SKP were generated in female Sprague-Dawley rats by intraperitoneal injection with 30 µg of purified protein in a 1:1 mixture of phosphate-buffered saline (PBS) (pH 7.4) and complete Freund's adjuvant; 4 and 6 weeks later, animals received 15-µg booster doses in a 1:1 mixture of PBS and incomplete Freund's adjuvant. Rabbit PrtP, PrcA1, PrcA2 and PrcB antibodies were a generous gift from Dr. Christopher Fenno (University of Michigan, Ann Arbor).

Immunoblot analysis were performed on samples that were separated through SDS or BN-PAGE. Proteins were transferred from the gel to nitrocellulose membrane (0.45 µM pore size, GE, USA) using a semidry apparatus (Bio-Rad, USA)) or a wet apparatus (XCell surelock blot module; Invitrogen, USA). Membranes were blocked for 1 h in PBS with 5% nonfat dry milk, 3% BSA, and 0.1% Tween 20 and probed overnight at 4°C with primary antibodies directed against different proteins. Rat primary antibodies directed against SKP, $MOSP^{FL/C/N}$ were used at a dilution of 1:5,000; Rabbit antibodies directed against PrtP, PrcA1, PrcA2 and PrcB were used at a dilution of 1:10,000. After washing with PBS and 0.1% Tween 20 (PBST), the membranes were incubated for 1 hour at RT with a horseradish peroxidase (HRP)-conjugated goat anti-rat (Southern Biotech, USA) or anti-rabbit (Bio-Rad, USA) antibody at a dilution of 1:30,000. Following

washes with PBST, the blots were developed using the SuperSignal West Pico chemiluminescent substrate (Thermo Fisher Scientific).

Heat-sensitivity assay and Proteinase K treatment: Duplicate samples of the detergent and soluble MOSP fraction were mixed with 1x Laemmli sample buffer (Bio-Rad, USA). One set was boiled for 10 min while the other was left at RT. All samples were subsequently separated through SDS-PAGE and blotted with rat primary antibodies directed against $MOSP^{FL}$ as described above.

$2x10^9$ *T.denticola* cells were harvested and reconstituted in 50mM Tris (pH 8), 10mM $CaCl_2$. Proteinase K (Promega, USA) was added at a working concentration of 100µg/ml and left at 40°C for 2 hours. The reaction was quenched with 5mM PMSF and separated on SDS-PAGE after quickly heating the sample for 10min. Subsequent immunoblot analysis was carried out using rat primary antibodies directed against $MOSP^{FL/C/N}$.

Co-immunoprecipitation (Co-IP) Assay: Detergent/Soluble MOSP fractions were mixed with rat $MOSP^{FL}$ antibodies with 10µl of PIC and left overnight at 4°C. The reaction was further mixed with Protein G magnetic beads (EMD-Millipore, USA) followed by elution with 0.2M glycine-HCl, pH 2.2. Similar steps were also performed for $MOSP^{FL}$ antibodies crosslinked to agarose resin obtained using the Crosslink IP kit (PIERCE, USA). Eluted samples were neutralized by 1M Tris pH 8.5 and analyzed with SDS-PAGE and immunoblotting. Immunoprecipitated proteins from the detergent fraction were probed with various rabbit polyclonal antibodies followed by detection with HRP conjugated Clean-Blot IP detection reagent (Thermo Scientific, USA). Additionally, eluate from the agarose crosslinked immunoprecipitation were subjected to SDS-PAGE analysis

Immuno-fluorescence Assay: Mid-late logarithmic-phase *T. denticola* cells were encapsulated in gel-microdroplets of low-melting-point agarose as previously described [48]. Encapsulated organisms were probed with rat antiserum directed against $MOSP^{FL/C/N}$, FlaA, domain or rabbit PrtP, PrcA1, PrcA2 and PrcB (each diluted 1:200). Encapsulated organisms was extensively washed with DMEM prior to the addition of primary antibodies. Beads were resuspended in DMEM with 2%BSA (0.2 to 0.3 ml) and primary antibody in the presence or absence of 0.05% (v/v) Triton X-100. Samples were incubated overnight with gentle mixing at 4°C. Beads were washed three times by low-speed centrifugation (500 xg) and mixed in DMEM (Thermo-Fisher, USA) with 2%BSA and 1 µg/ml of goat anti-rat Alexa Fluor 488 or Alexa Fluor 594 conjugates (Invitrogen, USA) for 2 hours at RT. Following further washing with DMEM, microdroplet gels were visualized under the microscope.

10µl of *T.denticola/ E. coli* cells in Vectashield® antifade reagent (Invitrogen) containing DAPI were applied to glass slides. Fluorescent images were acquired on an epifluorescence Olympus BX- 41 microscope using 100X (*T. denticola*) (1.4 NA) oil immersion objective equipped with Retiga Exicharge-coupled-device (CCD) camera (Q Imaging, USA) and fluorescein isothiocyanate (FITC) and rhodamine omega filter sets. The data were analyzed using Cell M (Olympus) and ImageJ.

Results

Recombinant C-terminal MOSP ($MOSP^{C}$) forms trimeric pores while the N-terminal domain ($MOSP^{N}$) exhibits an extended conformation together conferring a bipartite topology to full length MOSP ($MOSP^{FL}$).

The InterPro server [49] predicts MOSP (TDE0405) to contain two distinct domain, an N-terminal MOSP (113-348) domain and a C-terminal MOSP (364-542) domain as depicted in Figure 1A. The protein structure predicted through Iterative Threading ASSEmbly Refinement (I-TASSER) [50] displayed a beta barrel architecture for $MOSP^{C}$ while full length MOSP ($MOSP^{FL}$) was bipartite with $MOSP^{C}$ forming the potential pore attached to the extended $MOSP^{N}$ (Figure 1B). TX-114 phase partitioning, a routine technique used to segregate proteins based on their ability to partition differentially either in a detergent enriched or aqueous phase was used to access the amphiphilic character typical of an outer membrane spanning protein [51]. Recombinant full length MOSP ($MOSP^{FL}$) and C-terminal domain of MOSP ($MOSP^{C}$) partitioned into the detergent enriched phase while the hydrophilic N-terminal region ($MOSP^{N}$) remained in the aqueous phase (Figure 1C); SKP, a periplasmic protein served as a control. These results matched perfectly with previous study [13]. The results from phase partitioning can be interpreted to envision $MOSP^{C}$ as being embedded in the outer membrane (OM) while $MOSP^{N}$ remains tethered in the periplasm. Additionally, Arvind et al. previously demonstrated $MOSP^{C}$ to contain a secondary structure rich in β-sheets [13], a feature typical to β-barrel OM proteins. Herein we show how $MOSP^{C}$ can, by itself, form trimeric pores in artificial lipid bilayer nanodiscs, a system most suitable for the structure-function investigation of membrane proteins [52, 53]. We proved this by using a combination of

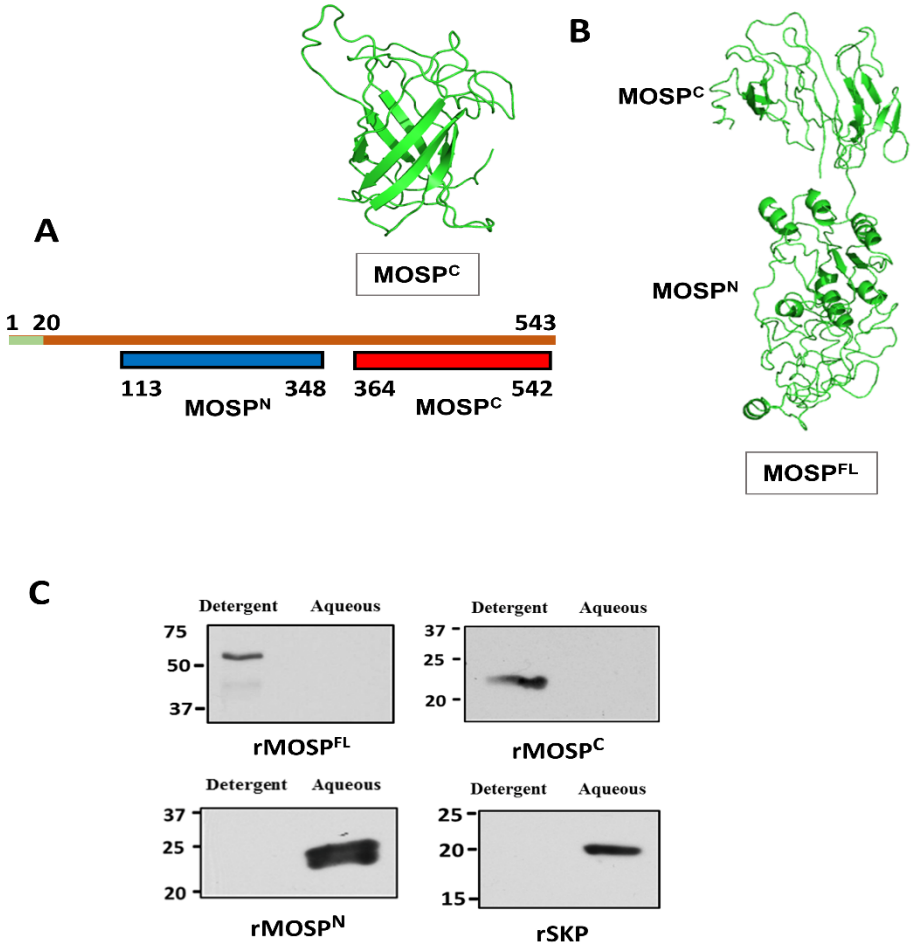


Figure 1: (A) Domain architecture of MOSP. First 20 amino acids form the signal sequence (green). The portion of MOSP colored in red denotes the C-terminal domain, MOSPC and blue denotes the N-terminal domain, MOSPN. **(B)** Predicted protein structure for MOSPC and MOSPFL generated using I-TASSER. **(C)** 10 µg of recombinant proteins MOSPFL, MOSPC, MOSPN and SKP were phase partitioned in TX-114 and separated on SDS-PAGE. Lanes show the detergent enriched and aqueous phase probed with antisera directed against the recombinant proteins.

small and large nanodiscs. D7 disc (small nanodisc), developed by Puthenveetil et al. [54] and MSP1E3D1 disc (large nanodisc) were used to incorporate detergent (DDM; n-Dodecyl-beta-Maltoside) refolded $MOSP^C$. Refolding efficiency was accessed through far-UV CD spectroscopy which gave us a broad minimum centering on 218 nm, indicating

a preponderance of β-structure where as empty nanodisc gave their signature α-helical minima at 208 and 222 nm arising from the outer belt membrane scaffold protein. As expected $MOSP^{C}$ incorporated discs gave an averaged spectrum arising from the presence of both α and β secondary structures (Figure 2A). MSP1E3D1 discs have a large diameter that can easily accommodate oligomeric membrane proteins [55]. These discs were utilized to ascertain the trimeric nature of $MOSP^{C}$. Electron micrographs were averaged using the Xmipp software suite [41]. The averaged representative TEM image majorly showed the presence of perfect trimers for $MOSP^{C}$ manifested through the black dots in the backdrop of a nanodisc (Figure 2A). In addition to trimers a few monomers were also observed which led us to consider the possibility of the entrapment of monomers owing to the large available space in the MSP1E3D1 discs. This question was resolved through the employment of small nanodiscs. Interestingly, the size exclusion chromatogram obtained for small discs produced three peaks (Figure 2B). SDS-PAGE analysis further revealed that peaks-1 and 2 contained a band for $MOSP^{C}$ in addition to the D7 protein while peak 3 only contained a band for D7. This meant that both peaks-1 and 2 contained $MOSP^{C}$ incorporated discs while peak-3 contained empty discs which also overlapped with the peak for empty discs prepared separately as a control. The representative image averaged TEM micrograph showed the presence of monomeric $MOSP^{C}$ in both peaks 1and 2 with the additional presence of trimers in peak-1 (Figure 2B). The result conclusively proved that $MOSP^{C}$ formed core trimers forcing the formation of large discs from the D7 construct that otherwise exclusively forms smaller discs. The pliability of smaller nanodisc constructs to form discs with larger diameters has been

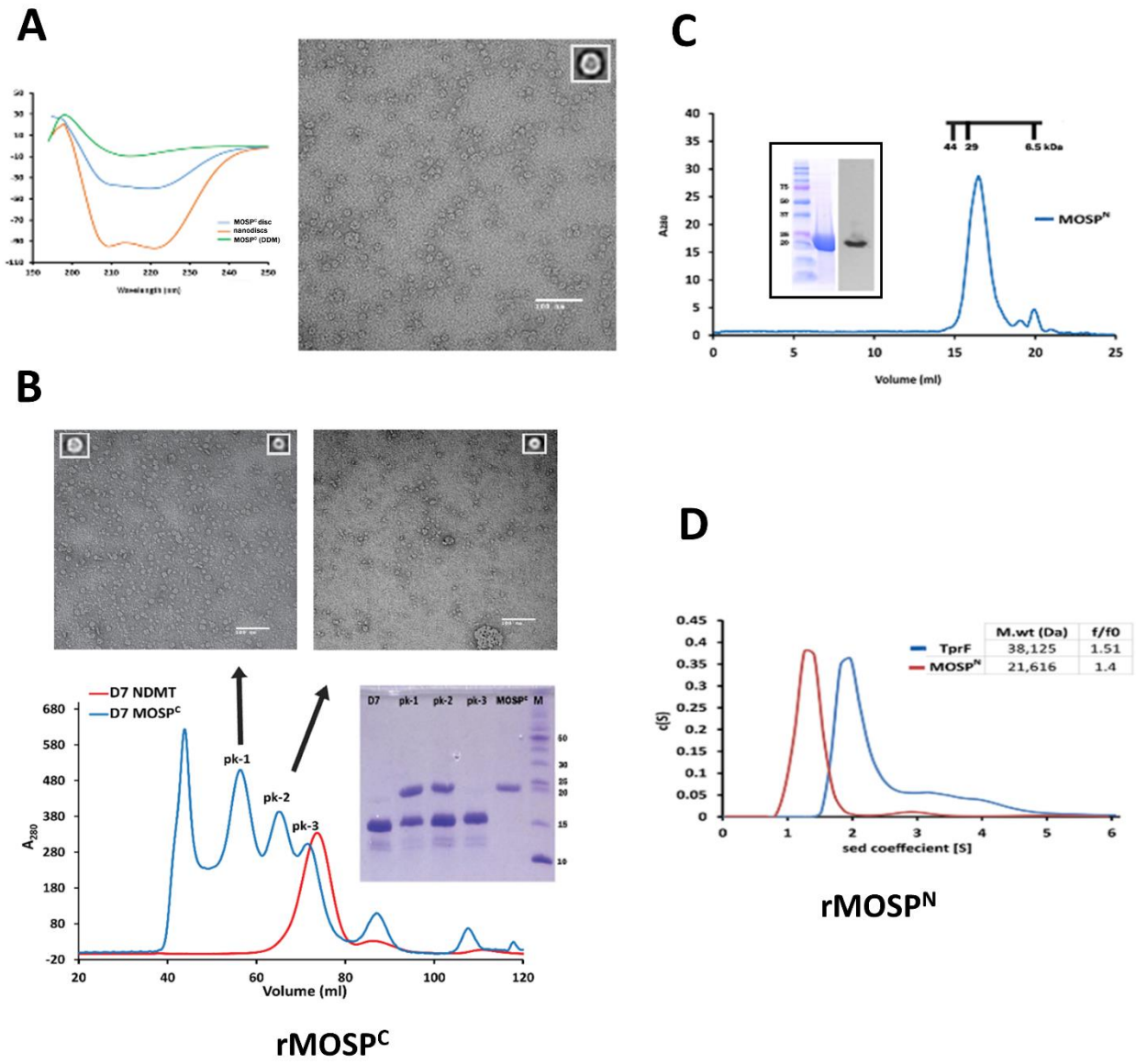


Figure 2: (A) Large E3D1 discs, CD spectra of MOSP^{C} refolded in DDM (green), E3D1 empty nanodiscs (orange) and MOSP^{C} encapsulated discs (blue). Negatively stained TEM image of MOSP^{C} in E3D1 discs with a representative average image. **(B) Small D7 discs,** SEC chromatogram of MOSP^{C} incorporated D7 discs separated on S200 column (blue). The chromatogram in red represents empty discs run as a control. SDS-PAGE gels analyzing peaks 1 through 3 in lanes 2-4. Lanes 1 and 5 contain D7 and MOSP^{C} for reference with lane 6 contains molecular weight markers. Also shown are TEM images of samples from peak 1 and 2 with a representative averaged image. **(C)** SEC chromatogram of refolded recombinant MOSP^{N} with SDS-PAGE and immunoblots shown as inset. **(D)** Sedimentation velocity experiments performed on MOSP^{N} and TprF. The molecular weight and frictional coefficient as determined through c(S) analysis using SEDFIT.

previously demonstrated in several instances [54, 56, 57]. Next, we moved onto the recombinant MOSP N-terminal ($MOSP^N$) domain. Size exclusion chromatography of refolded $MOSP^N$ revealed a monomeric form, as can been seen by the peak corresponding to a molecular weight of ~23 kDa in Figure 2C. The peak fraction were further subject to Analytical ultracentrifugation (AUC) estimating the molecular weight to that of a monomer corroborating the observation from SEC. Additionally, the f/f0 ratio closely matched the value obtained for TprF (Figure 2D), a *T. pallidum* repeat (Tpr) protein, indicating the presence of an extended shape as was also observed in the $MOSP^{FL}$ model in Figure 1B. TprF is a truncated ortholog of MOSP comprising only of its N-terminal domain and has been previously shown to have an extended structure through small angle X-ray scattering [20] . All the above data, taken together strongly establishes the recombinant MOSP to comprise of a trimeric C-terminal domain and a monomeric N-terminal domain, the latter of which could associate with other proteins in the periplasm forming a large complex.

Native *T.denticola* MOSP also exists as a bipartite protein with a putative periplasmic $MOSP^N$ and a surface accessible, membrane embedded $MOSP^C$ domain.

After conclusively establishing a bipartite topology for the recombinant protein, we moved onto investigate the native structure of MOSP in *T. denticola*. Indirect immunofluorescence analysis of native organisms, probed with antisera directed against $MOSP^C$ and $MOSP^N$ was utilized to establish the cellular localization of the two domains. Agarose beads (gel microdroplets) encapsulating *T. denticola*, served as an ideal system,

as it not only maintained the integrity of the OM but also helped in its selective removal through detergent solubilization. $MOSP^{C}$ was labeled in intact organism, while both $MOSP^{C}$ and $MOSP^{N}$ were labeled in permeabilized organisms (data not shown). Success of detergent solubilization was accessed with antiserum directed against *T. denticola* endoflagella, exclusively present in the periplasm. Labeling of $MOSP^{C}$ with permeabilized organism could be either because of a partially solubilized OM or through the accessibility to an alternate form of full-length MOSP present in the periplasm as has be reported before [8, 13, 58]. As a second approach, we performed proteinase K (PK) treatment of intact *T. denticola* to corroborate our findings from the IFA experiment. Previous exposure of *T. denticola* to PK retained full motility, and selectively exposed the surface accessible proteins leaving the periplasmic proteins unaffected [13] . This provided a proper setting for us to determine the topology of MOSP through its exposure to PK. SDS-PAGE gel images in Figure 3 compares PK treated to untreated samples probed with antisera directed against $MOSP^{FL}$, $MOSP^{C}$ and $MOSP^{N}$. No proteolytic product was detected by $MOSP^{C}$ antibodies while a cleaved band at ~ 25 kDa was detected by both $MOSP^{FL}$ and $MOSP^{N}$ antibodies. $MOSP^{C}$ being surface accessible underwent proteolytic degradation leaving behind an intact $MOSP^{N}$ domain in the periplasmic space unavailable for PK degradation. Interestingly, all blots showed the presence of a band at ~ 52 kDa, the position for full- length MOSP, indicating the presence of a periplasmic form that was also inaccessible to PK treatment.

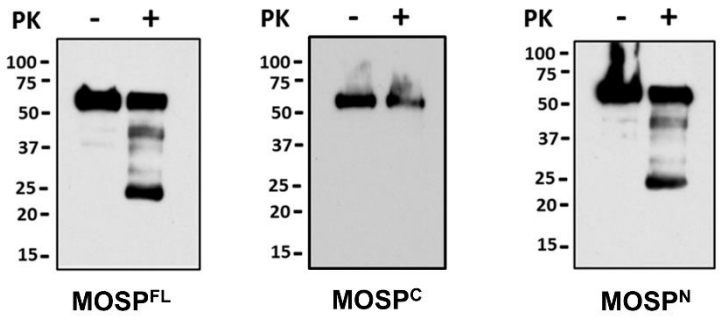


Figure 3: Surface proteolysis of T. denticola exposed to proteinase K (PK) for 1 hour. Immunoblot analysis of MOSP in the presence (+) and absence (-) of PK, detected with $MOSP^{FL}$, $MOSP^{C}$ and $MOSP^{N}$ antisera. Molecular mass standards (kDa) are indicated on the left of each panel.

The membrane and periplasmic conformers of MOSP are native only to *T. denticola*, with the periplasmic form likely associating with the peptidoglycan sacculus.

Both the above two approaches implicated the presence of two conformers of full-length MOSP residing in the OM and periplasmic space. TX-114 phase partitioning, a simple and effective technique to segregate proteins based on their amphiphilic or hydrophilic nature was performed [46]. Phase partitioning was carried out on two samples, intact and lysed *T. denticola*. The gel picture in Figure 4A shows that the membrane form was recovered in the detergent enriched phase while the periplasmic conformer was recovered in the aqueous phase for intact organism. The presence of a band in the insoluble lane indicated a possible weak association with the peptidoglycan sacculus that was mostly disrupted during the phase partitioning process. The success of phase partitioning was demonstrated by the periplasmic protein SKP that exclusively partitioned into the aqueous phase. Detergent solubilized *T. denticola* membrane fraction,

was obtained from the lysate following the protocol in Figure 4B. *T. denticola* lysate in tris buffer was fractionated to separate the soluble from the membrane fraction. The pellet, containing the membrane, was solubilized in 2% DDM and subjected to TX_114 phase partitioning. As shown in Figure 4C MOSP was completely present in the pellet and interestingly partitioned both in the detergent and aqueous phase. This suggested that the membrane pellet brought down the periplasmic conformer which was later solubilized by the detergent. DDM has been previously utilized to disengage TprC, an OM protein, from the peptidoglycan in *T. pallidum* [25]. DDM solubilized pellet, before being phase partitioned was assessed for its ability to form heat sensitive trimers. The ability to form trimers (Figure 4D) discounted the possibility of an artifact introduced through the use of detergent leading to an erroneous phase partitioning result. It is important to note that the sample used for the trimerization assay contained both forms of MOSP.

We further questioned whether the ability to form different conformers was a feature intrinsic to MOSP, or was it an outcome of the complex periplasmic milieu developed to assist the survival of *T. denticola* residing within a unique nutritional environment. To answer this question we expressed MOSP with a signal sequence localizing it to the pertinent cellular compartment in *E.coli*, a genetically malleable surrogate. The gene encoding MOSP was placed downstream of the PelB signal sequence in pET vector and expressed under the control of IPTG, later analyzed through cell fractionation, BN-PAGE and IFA. MOSP was exclusively recovered with the OM following fractionation; and was found to retain its native trimeric form as was shown through BN-PAGE (done by Dr. Arvind Anand). OmpA, SKP and α-ATPB were used to

access the integrity of fractionation. Representative IFA images (data not shown) of intact and permeabilized *E. coli*, probed with antisera directed against $MOSP^{C}$ and $MOSP^{N}$

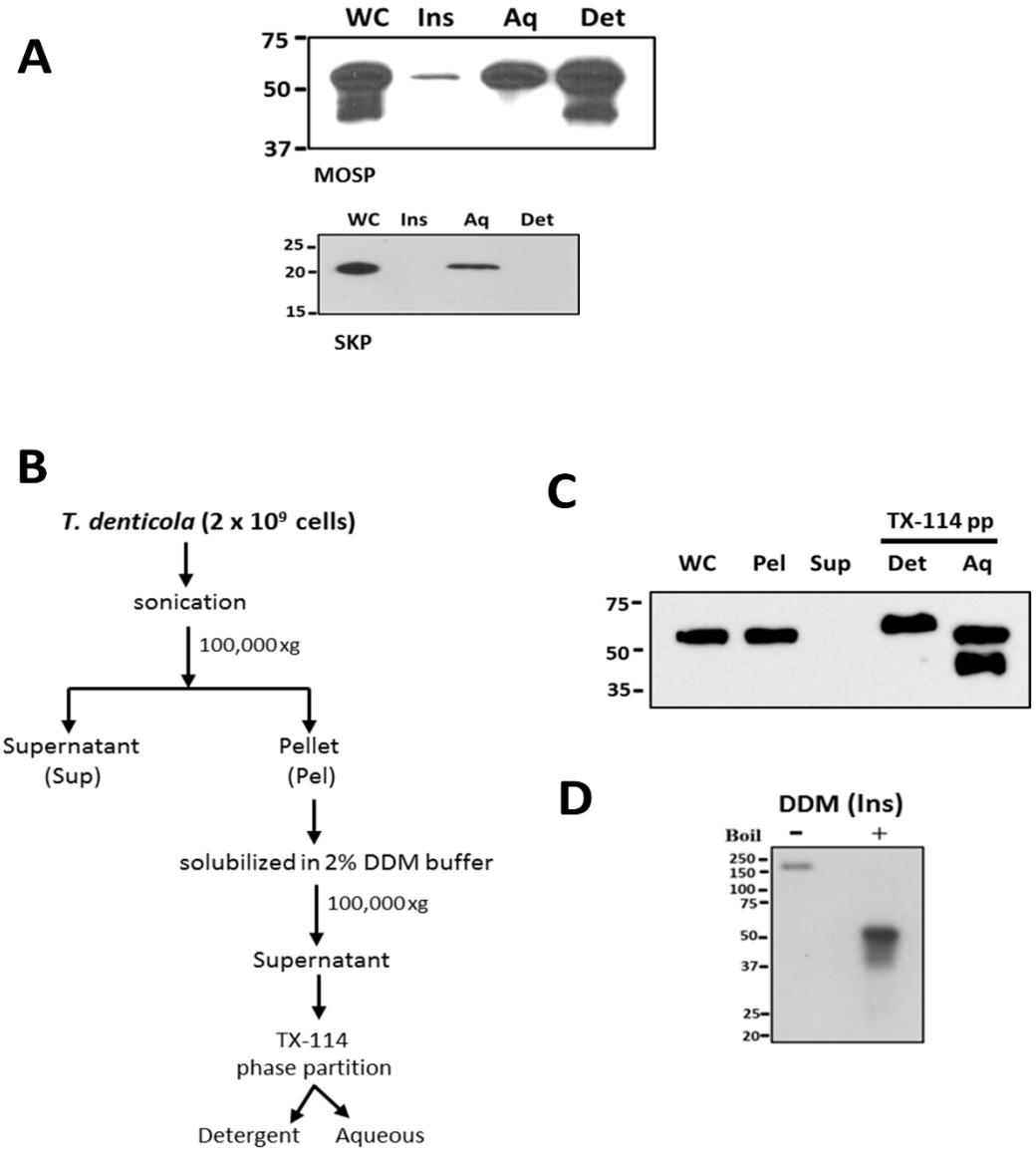


Figure 4: (A) Detergent enriched and aqueous phase from TX-114 phase partitioning of 2x 10^9 *T. denticola* cells immunoblotted with $MSOP^{FL}$ antiserum following SDS-PAGE. Also show are whole cell (WC) and insoluble material (Ins) from phase partitioning. **(B)** Protocol delineating the fractionation of *T.denticola* to obtain membrane pellet further solubilized with 2% DDM and phase partitioned using TX-114. **(C)** Immunoblot analysis of the pellet (pel) and supernatant (sup) from initial cell fractionation shown with respect to the whole cell (wc), followed by the detergent and aqueous phase from TX-114 phase partitioning of the DDM solubilized pellet. **(D)** DDM solubilized pellet was separated by SDS-PAGE with (+) and without (-) boiling followed by immunoblot analysis with anti-$MOSP^{FL}$ antiserum. Molecular mass standards (kDa) are indicated on the left of each panel.

positively labeled $MOSP^C$ while $MOSP^N$ was labeled only in permeabilized organism. Antiserum directed against the *E. coli* periplasmic protein SKP served to confirm the successful permeabilization of E.coli. The results indicated that MOSP by itself formed a bonafied, bipartite OM trimer allowing us to suspect either the periplasmic composition of *T. denticola* or the association of MOSP with different binding partners, to regulate the formation of multiple conformers.

Both MOSP conformers are present as individual multimeric complexes

An obvious succession to the identification of the conformers was to ascertain if they existed as individual trimers or as multimeric complexes. BN-PAGE was employed for the determination of their multimeric states. Typically TX-114 phase partitioning ends with the precipitation of proteins using acetone. This step was omitted since we wanted to retain the native conformation of MOSP. The periplasmic form was isolated from the aqueous phase of TX-114 phase partition. The detergent phase containing the membrane conformer was resistant to separation on BN-PAGE, since the detergent TX-114 formed

micellar aggregates due to a low cloud point when loaded into the gel. Moreover, the detergent phase was present as a viscous plug making it additionally difficult to traverse the gel. The problem was circumvented by isolating the membrane form using NP-40, a mild detergent belonging to the triton regime. Fractionation with NP-40 was thoroughly analyzed using the protocol outlined in Figure 5A. Briefly, intact *T.denticola* was solubilized in 1%NP-40 buffer with mixing followed by separating the detergent solubilized fraction from the insoluble pellet through ultracentrifugation. Both fractions were phase partitioned and analyzed through SDS-PAGE. Bands corresponding to MOSP were identified in the supernatant and pellet. DDM containing supernatant exclusively partitioned into the detergent phase while the pellet partitioned into the aqueous phase (Figure 5B). This proved that the NP-40 isolated conformer was a true representation of the membrane form. A brief outline of the procedure to obtain the membrane and periplasmic conformer is outlined in Figure 5C. Both samples were accessed, through SDS-PAGE, for their ability to form heat sensitive trimers, confirming the presence of MOSP in its native form; while BN-PAGE analysis demonstrated that each conformer, individually existed as a higher molecular weight complex (Figure 5C).

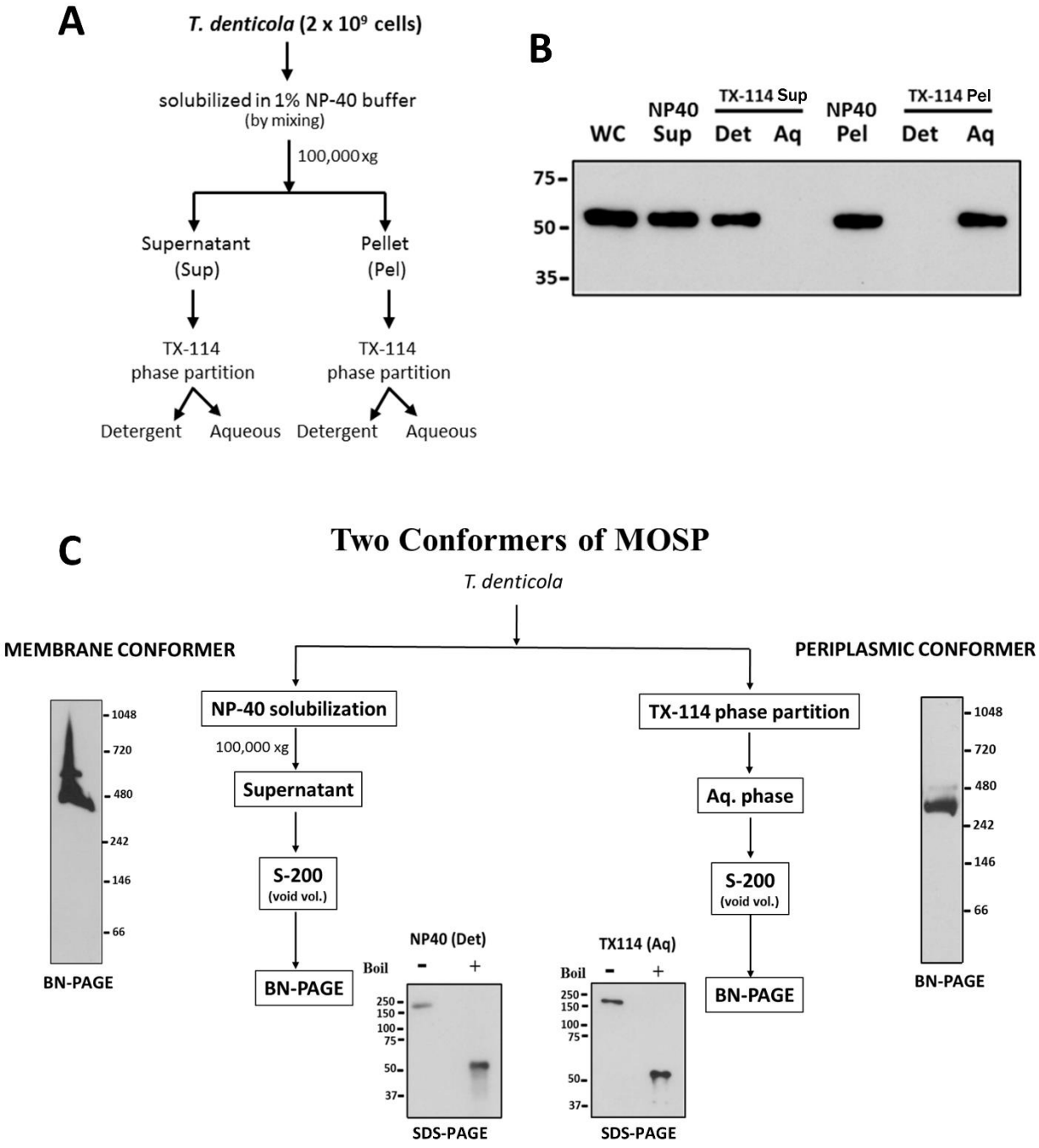


Figure 5: (A) Protocol outlining TX-114 phase partitioning of the fractions obtained from OM solubilization with the 1% NP-40 detergent. **(B)** Immuoblot of the individual steps outlined in the protocol detected with anti-MOSP^{FL} antiserum. Lanes: whole cell (wc), NP40 supernatant (Sup), detergent (Det) and aqueous (Aq) phase from TX-114 phase partitioning of supernatant; NP40 pellet (Pel) followed by detergent (Det) and aqueous (Aq) phase from TX-114 phase partitioning of pellet. **(C)** BN-PAGE of the membrane and periplasmic conformer obtained using the outlined protocol. Also shown are SDS-PAGE gels with (+) and without (-) boiling followed by immunoblot analysis with anti-MOSP^{FL} antiserum for the two conformers. Molecular mass standards (kDa) are indicated on the left of each panel.

Dentilisin forms a detergent sensitive complex only with the membrane conformer of MOSP in *T.denticola*.

Godovikova et al. had previously shown dentilisin to form a complex with MOSP [34]. We tried to determine whether dentilisin was associated with either or both conformers. Our study was initiated by first identifying the solvent characteristics of dentilisin. Dentilisin is a detergent stable complex of four components, PrtP, PrcA1, PrcA2 and PrcB of which PrtP acts as the primary protease [33, 59, 60]. The outcome of TX-114 phase partitioning, using *T.denticola*, clearly demonstrated that the individual components of dentilisin partition into the detergent enriched phase (Figure 6A, done by Dr. Dey). Of the four components only PrtP and PrcA1 have a discernable spII cutting site and hence are anchored to the membrane. The reason for PrcA2 and PrcB also partitioning with the dentilisin complex resided in the finding that dentilisin formed an intact SDS stable complex (Figure 6B) aiding in the co-partitioning of the other components to the detergent phase. As a final analysis we performed co-immunoprecipitation with $MOSP^{FL}$ antisera

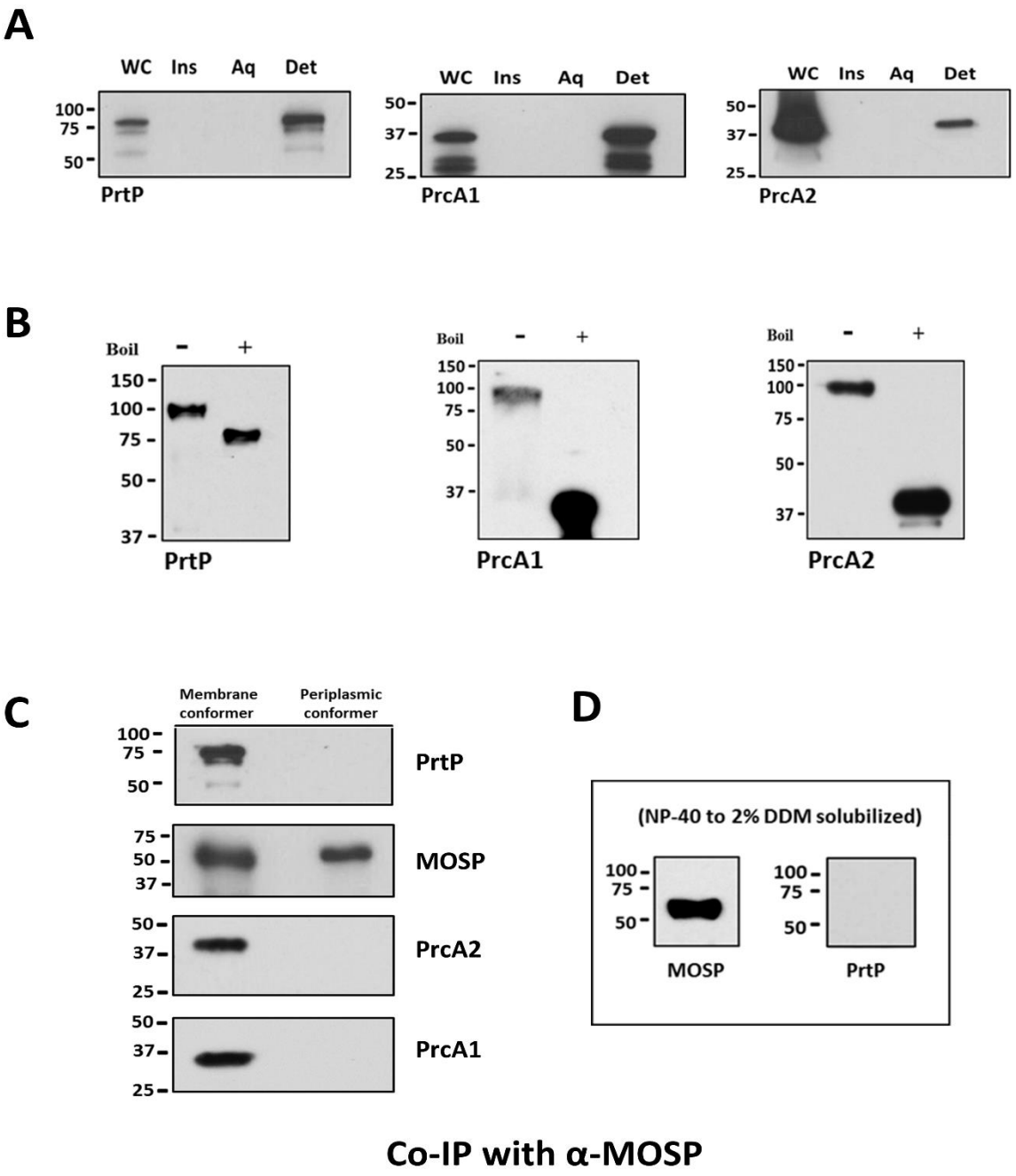


Figure 6: (A) Immunoblot, with antisera directed against dentilisin components (PrtP, PrcA1, and PrcA2), of the detergent and aqueous phase from TX-114 phase partitioning of 2x 10^9 *T. denticola* organisms. Lanes: whole cells (WC), Tx-114 insoluble material (Ins), aqueous (Aq) and detergent (Det) phase. Molecular mass standards (kDa) are indicated on the left of each panel. **(B)** SDS-PAGE gels of the membrane conformer with (+) and without (-) boiling followed by immunoblot analysis with anti- PrtP, PrcA1,

PrcA2 antisera. **(C)** Eluate from co-immunoprecipitation (Co-IP) using $MOSP^{FL}$ antiserum of the membrane and periplasmic conformers, immunoblotted with antisera against the dentilisin components. **(D)** Co-IP of the membrane conformer further solubilized with 2% DDM and blotted with ant- $MOSP^{FL}$ and PrtP antisera.

and were able to identify all dentilisin components only with the membrane conformer (Figure 6C). Though both dentilisin and MOSP individually occur as detergent stable oligomers, their association with each other could be easily abrogated with the use of strong detergents. Shown in Figure 6D is the outcome of Co-IP performed with $MOSP^{FL}$ antisera on the NP-40 solubilized membrane conformer which was further mixed with 2% DDM. No band for PrtP was observed, the band for MOSP served to demonstrate the success of Co-IP. Similar results were obtained for Co-IP performed with DDM solubilized pellet from Figure 4B (data not shown).

A relevant follow up, to the identification of dentilisin being associated with the membrane form of MOSP, would be to identify its native topology within *T. denticola*. Immunolabeling experiments performed on *T. denticola* trapped within agarose gel microdroplets were probed with antisera directed against different components of dentilisin. OM permeabilization was achieved with TX-100 detergent. Normal rabbit serum served to eliminate the possibility of any non-specific interactions. Only PrcB and PrcA2 were surface accessible whereas all components were labeled in the permeabilized organism (Figure 7). The results were in contrast to previous finding of all components being surface accessible in intact organism [34]. The inability to immunolabel PrtP and PrcA1 was substantiated with both proteins, containing a spII cutting site, being a lipoproteins anchored to the inner leaflet of the OM and hence not surface accessible in intact organisms. The results were further reproduced by Dr. Sanjiv Kumar in the presence of antibodies detecting the *Td* Bam POTRA domain as an additional control

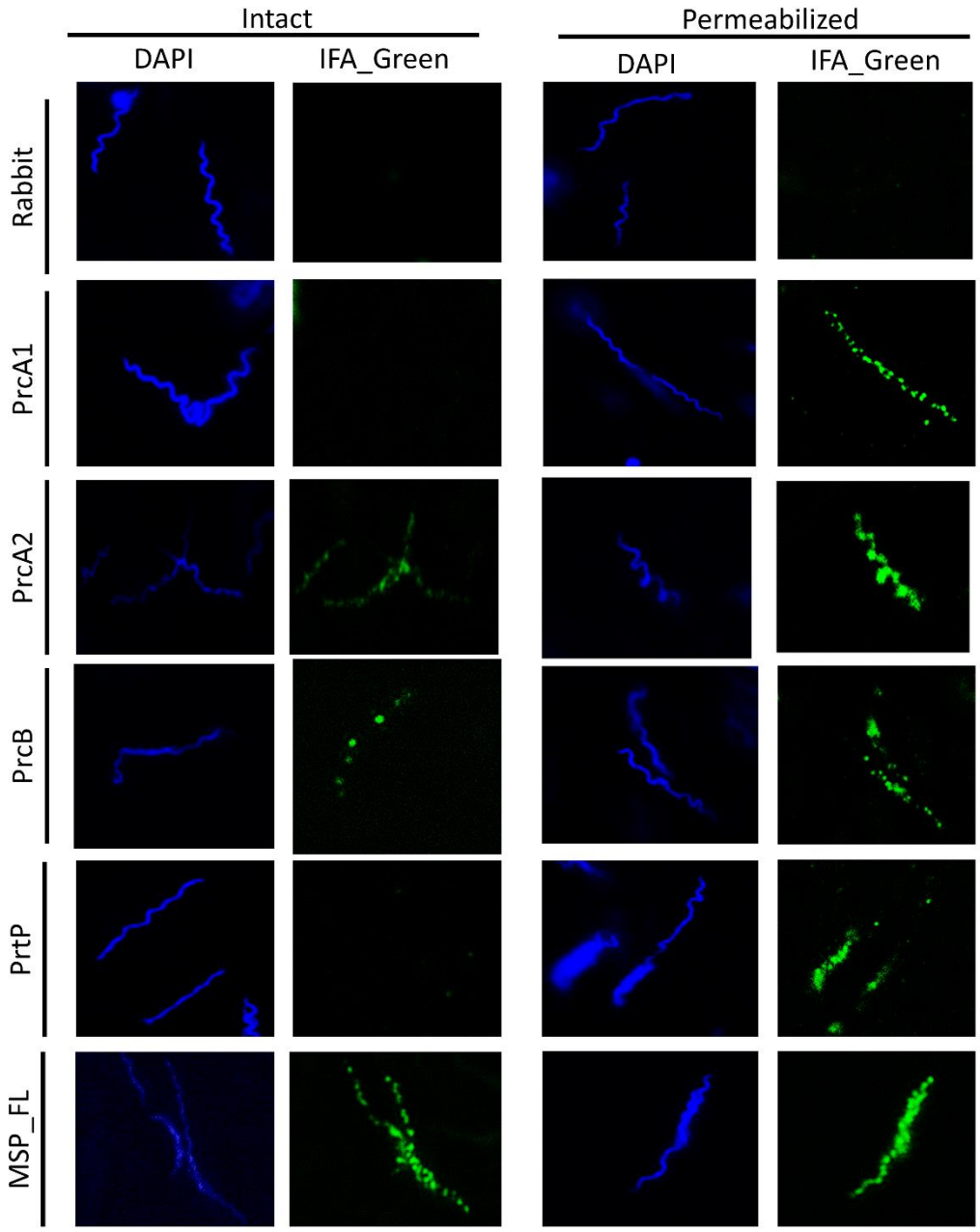


Figure 7: Intact and 0.05% TX-100 permeabilized *T. denticola* encapsulated in agarose gel microdroplets were probed with rabbit antisera against $MOSP^{FL}$, PrtP, PrcA1, PrcA2 and PrcB. Normal rabbit serum (Rabbit) served as a control to monitor background, non-specific labeling. Antibody binding was detected with goat anti-rabbit Alexa Fluor 488 (green).

Identification of potential candidates associated with the different conformers of MOSP.

The most intriguing observation was found with native gel analysis of the MOSP membrane conformer. Having established dentilisin as part of the MOSP complex, we further immublotted the BN-PAGE gel with antisera against the different dentilisin components. Dentilisin was present as part of a minor complex at ~ 600 kDa, while the majority of MOSP, at ~ 450 kDa, remained associated with other protein(s) (Figure 8). The periplasmic conformer was found to have a single band at ~ 360 kDa (Figure 5C). The major bands from the native gels were further analyzed for the identification of their protein IDs though LC-MS/MS.

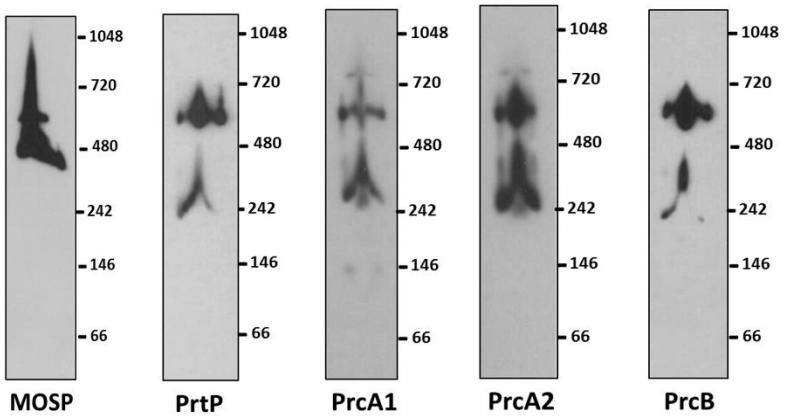


Figure 8: BN-PAGE of the membrane conformer immunoblotted with antisera directed against MOSP^{FL} and the individual components of dentilisin (PrtP, PrcA1, PrcA2 and PrcB). Molecular mass standards (kDa) are indicated on the right of each panel.

Discussion

Phylogenetic analysis show treponemes to form a single cluster separate from gram negative and positive bacteria [61]. The oral gram-negative treponeme, *T. denticola*, possess an outer membrane characteristic of gram positive bacteria [4] facilitating the notion of having proteins anchored to the OM facing the external environment. MOSP, with considerable sequence variation, is ubiquitously present across oral treponemes [6]. MOSP is one of the dominant immune responsive antigen from T. denticola [2]. A foray into the understanding of the structural architecture and the functional relevance of a prominent OM protein should offer insights into the survival of the organism is an ever varying environment.

Independent studies have confirmed that MOSP is an archetypal OM trimer with well-established porin activity, either through the entire protein forming one giant pore akin to OmpF [14], or having a bipartite topology with the C-terminal domain ($MOSP^{C}$) forming the pore and being tethered to its N-terminal domain ($MOSP^{N}$) like OmpA [13]. Evidence for both has been provided majorly through surface accessibility from IFA analysis. The use of antisera directed against the full-length protein fails to differentiate between the domains, a problem circumvented through the use of antisera directed specifically against the individual domains. Moreover, the display of porin activity with just the $MOSP^{C}$ domain, makes a stronger argument in favor of a two domain architecture [13]. The ability of recombinant $MOSP^{C}$ to form trimeric pores in artificial membranes with an extended monomeric $MOSP^{N}$ domain, demonstrated in the present study provides further impetus to a bipartite topology with the functional implication of the N-terminus potentially binding to other proteins. The antigenic determinant of MOSP has been

arguably narrowed down to its $MOSP^N$ domain [62] which remains inaccessible due to its periplasmic localization.

The surface accessibility of $MOSP^C$ in *T. denticola* through immune labeling and PK experiments indirectly highlighted the presence of the different conformers of MOSP. A possibility demonstrated in previous studies [8, 13, 58]. In addition, the periplasmic MOSP was also shown to associate with the peptidoglycan layer comfortably placing it in its hypothesized position of being a parental ortholog to the *Treponema pallidum* repeat proteins (Tpr) [63]. Tprs consist of a family of twelve paralogs (Tpr A-L) containing sequence homology with MOSP [17, 63]. TprC and TprI are bonaifed OM proteins sharing the bipartite topology with the membrane conformer of MOSP. Interestingly, they can also associate with the PG layer [20, 25]. TprF, a truncated Tpr lacking the C-terminal domain has an extended shape [20] analogous to the N-terminal $MOSP^N$ domain. Of special note is TprK, which even though harboring both domains remains localized to the periplasmic space [23] resembling the periplasmic conformer of MOSP. The divergent nature of Tprs can be easily visualized through sequence based clustering in CLANS [64]. Tprs with classical OM characteristics cluster together, while Tprs (A,F,K) which don't follow the canonical OM features fall outside the core cluster (Figure 9B). Thus the ability of MOSP to exist as a membrane and periplasmic conformer with its association to the PG sacculus encompasses the individual characteristics of the disparate Tpr paralogs.

Central to the biogenesis of β-barrel proteins in the OM is BamA, a component of the beta-barrel assembly machinery complex, and two periplasmic chaperones SKP and SurA [65]. We identified orthologs of BamA (TDE2601) and SKP (TDE2602) in *T. denticola* present in a genetic arrangement where the genes are contiguous to each other,

as is classically observed in proteobacteria [66]. We have also identified a putative ortholog of SurA (TDE1658). Nascent OM polypeptide associated with the periplasmic chaperones is transported to the Bam complex and threaded into the OM through a process called β-augmentation [67]. The initiation of this process occurs at the POTRA-1 domain in the flexible POTRA arm of BamA, which recognizes the C-terminal insertional signal present in the unfolded OM protein [68, 69]. The presence of an insertional signal in the C-terminal end of MOSP confirms that it utilizes the Bam machinery to enter the OM. Despite the distant phylogenetic relationship between spirochetes and proteobacteria [70], the heterologous expression of MOSP in the OM of *E.coli* clearly demonstrates the inability of the Bam machinery to differentiate OM protein from different species. MOSP maintains its trimeric, bipartite topology in *E.coli* substantiating the use of a genetically modifiable surrogate for the structural and functional investigation of OM proteins from organisms that are either non cultivable or remain resilient to genetic manipulation like the commensal treponemes [71]. The absence of a periplasmic conformer in *E. coli* allowed us to hypothesize on the involvement of the periplasmic milieu in *T. denticola* to differentially segregate the proteins into different cellular locales (Figure 9A). The most likely explanation resides with the formation of a multimeric complex which sequesters away MOSP from the Bam pathway potentially by occulting its C-terminal insertional sequence. Periplasmic proteins with diminished activity [72] or with partial chaperoning capabilities through the presence of peptidyl-prolyl cis-trans isomerase domains could serve as potential binding candidates stabilizing the periplasmic

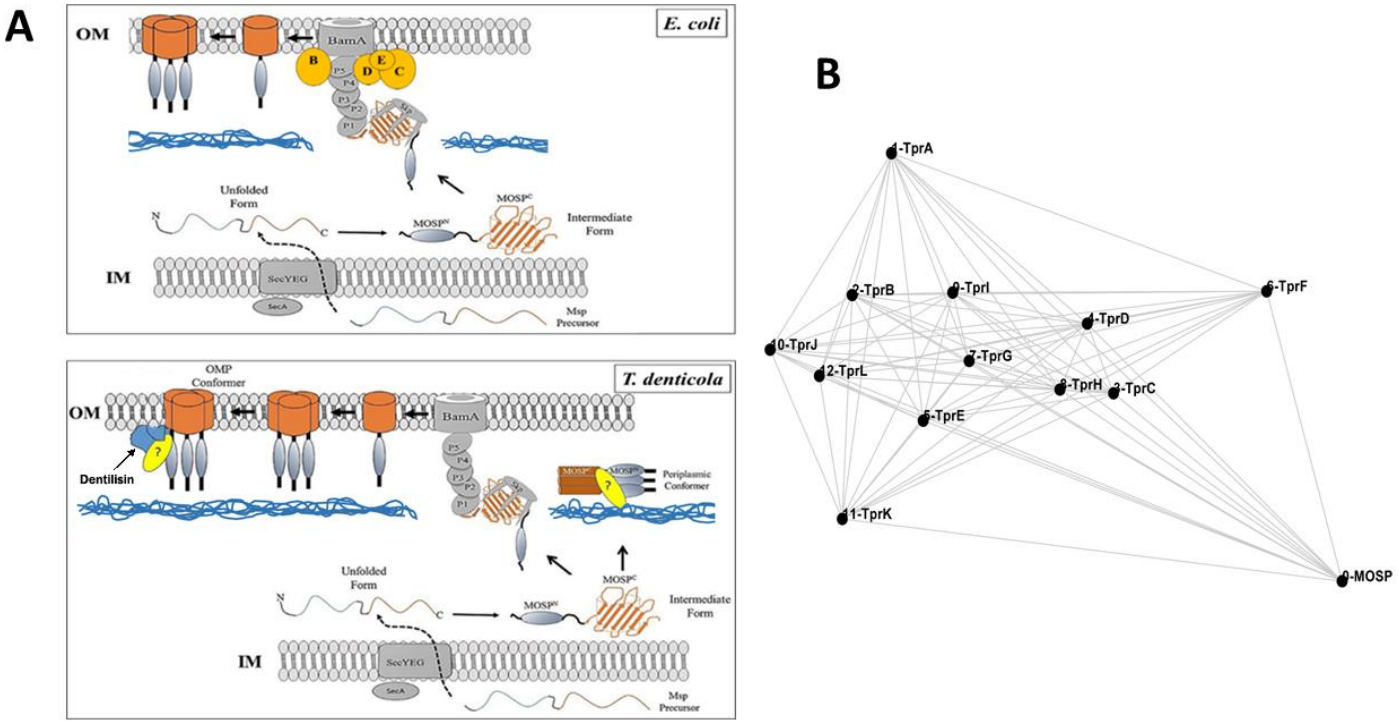


Figure 9: (A) Proposed model depicting the different pathways followed by MOSP in *E. coli* and *T. denticola*. MOSP precursor is exported into the periplasm through the SEC translocon where they are received by the periplasmic chaperones. The chaperone protected nascent polypeptide is transported to the POTRA arm of the Bam complex and threaded into the OM. The periplasmic conformer only found in *T. denticola* is predicted to bind to the peptidoglycan sacculus. The unknown proteins bound to MOSP are shown in yellow. Dentilisin is depicted in blue associated to the membrane conformer. **(B)** Cluster analysis based on alignment and phylogenetic relationship between MOSP and the 12 Treponema repeat protein (Tpr A-L) paralogs from *T. pallidum*.

conformer. We believe that this conformer may either serve a specific role or may remain poised in the periplasm, ready to enter the OM in response to a cellular shock.

The formation of a dentilisin complex with MOSP has been previously demonstrated [34] and challenged [30]. We lay this contention to rest by conclusively demonstrating the association of dentilisin to the membrane conformer of MOSP. Immuno-labeling experiments showed two (PrcA2 and PrcB) out of the four components

to be surface accessible. The other two lipoproteins, PrtP and PrcA1, harboring the cleavable SP-II cutting site remained inaccessible since they anchor to the inner leaflet of the OM. These results were in contrast to previous observations where all four components were surface exposed [34]. A likely explanation to this discrepancy might come from the presence of two populations of dentilisin, of which one associates with MOSP. We base our hypothesis on the following two reasons. First, MOSP deletion mutants did not affect dentilisin activity [73, 74] whereas defects in PrtP abolished the protease activity but did not affect MOSP expression [30]. Second, MOSP-Dentilisin is a weakly associating complex, easily abrogated through the use of DDM (Figure 6D). Moreover the BN-PAGE in Figure 8 indicates a possible presence of dentilisin at ~ 300 kDa. *T.denticola* has over 25 peptidases, hydrolases and putative degradative enzymes. Other than PrtP, a three membered family of internalin related proteins (TDE0593, TDE2003 and TDE2231) has been strongly implicated in disrupting epithelial barriers [27, 75]. For tissue penetration, in addition to proteases, adhesins are required to successfully bind a mammalian cell. In-silico analysis helped us identify several *T.denticola* orthologs of putative adhesins from *T.pallidum*, TDE0015 (TP0136) [76], TDE2318 (TP0155) [77] and pallilysin, a complex formed by TDE0841 (TP0750) and TDE0840 (TP0751) of which TP0751 functions as a metalloproteinase [78]. Future work on these proteins should reveal their precise role in the dissemination and attachment to the extracellular matrix.

Majority of the membrane conformer of MOSP from native gel analysis (Figure 9A) was present at a position different from dentilisin. MS/MS analysis identified the proteins in this fraction to contain a solute binding domain (SBD) canonical to the protein OppA from the ATP binding cassette ATP transporters (re). Of the three OppA identified, two

contained a complete set of the OppA permease system with the Solute binding protein, two nucleotide binding domain (NBD) and two permeases [79, 80] . TDE1071 is a 70kDa classical OppA reported before and has been implicated to act as an adhesin exposed on the surface of *T.denticola* [37]. The other TDE1072, is an unusually large protein with a relatively small portion of the protein comprising of SBD unlike the classical OppA where it forms the majority of the protein. In a previous study, deletion of the MOSP gene also abolished the expression of TDE1072 [74]. It is alluring to speculate that the conglomerate formed by MOSP, protease and peptide transport system might work in concert to uptake the products of intense proteolytic activity in periodontal lesion during the progression of disease. To further confirm the different binding partners of MOSP associating with the different MOSP conformers, MS analysis of the Co-IP eluate from pull down with MOSP antibodies is underway.

References

1. Paster, B.J., et al., *Bacterial diversity in human subgingival plaque.* Journal of bacteriology, 2001. **183**(12): p. 3770-83.
2. RP., E., *Virulence determinants of oral treponemes*, in *Pathogenic Treponema*, L.S. Radolf JD, Editor 2006, Caister Academic Press: Norwich, United Kingdom. p. 357-386.
3. Socransky, S.S., et al., *Microbial complexes in subgingival plaque.* Journal of clinical periodontology, 1998. **25**(2): p. 134-44.
4. Schultz, C.P., et al., *Evidence for a new type of outer membrane lipid in oral spirochete Treponema denticola. Functioning permeation barrier without lipopolysaccharides.* The Journal of biological chemistry, 1998. **273**(25): p. 15661-6.
5. Fenno, J.C. and B.C. McBride, *Virulence factors of oral treponemes.* Anaerobe, 1998. **4**(1): p. 1-17.
6. Fenno, J.C., et al., *Conservation of msp, the gene encoding the major outer membrane protein of oral Treponema spp.* Journal of bacteriology, 1997. **179**(4): p. 1082-9.

7. Lee, S.H., K.K. Kim, and B.K. Choi, *Upregulation of intercellular adhesion molecule 1 and proinflammatory cytokines by the major surface proteins of Treponema maltophilum and Treponema lecithinolyticum, the phylogenetic group IV oral spirochetes associated with periodontitis and endodontic infections.* Infection and immunity, 2005. **73**(1): p. 268-76.
8. Fenno, J.C., et al., *Cytopathic effects of the major surface protein and the chymotrypsinlike protease of Treponema denticola.* Infection and immunity, 1998. **66**(5): p. 1869-77.
9. Ding, Y., et al., *Release and activation of human neutrophil matrix metallo- and serine proteinases during phagocytosis of Fusobacterium nucleatum, Porphyromonas gingivalis and Treponema denticola.* Journal of clinical periodontology, 1997. **24**(4): p. 237-48.
10. Visser, M.B., et al., *Treponema denticola major outer sheath protein induces actin assembly at free barbed ends by a PIP2-dependent uncapping mechanism in fibroblasts.* PloS one, 2011. **6**(8): p. e23736.
11. Wang, Q., et al., *A spirochete surface protein uncouples store-operated calcium channels in fibroblasts: a novel cytotoxic mechanism.* The Journal of biological chemistry, 2001. **276**(25): p. 23056-64.
12. Mathers, D.A., et al., *The major surface protein complex of Treponema denticola depolarizes and induces ion channels in HeLa cell membranes.* Infection and immunity, 1996. **64**(8): p. 2904-10.
13. Anand, A., et al., *The major outer sheath protein (Msp) of Treponema denticola has a bipartite domain architecture and exists as periplasmic and outer membrane-spanning conformers.* Journal of bacteriology, 2013. **195**(9): p. 2060-71.
14. Egli, C., et al., *Pore-forming properties of the major 53-kilodalton surface antigen from the outer sheath of Treponema denticola.* Infection and immunity, 1993. **61**(5): p. 1694-9.
15. Marchler-Bauer, A., et al., *CDD: NCBI's conserved domain database.* Nucleic acids research, 2015. **43**(Database issue): p. D222-6.
16. Finn, R.D., et al., *The Pfam protein families database: towards a more sustainable future.* Nucleic acids research, 2016. **44**(D1): p. D279-85.
17. Gray, R.R., et al., *Molecular evolution of the tprC, D, I, K, G, and J genes in the pathogenic genus Treponema.* Molecular biology and evolution, 2006. **23**(11): p. 2220-33.
18. Centurion-Lara, A., et al., *Fine analysis of genetic diversity of the tpr gene family among treponemal species, subspecies and strains.* PLoS neglected tropical diseases, 2013. **7**(5): p. e2222.
19. Radolf, J.D., et al., *Treponema pallidum, the syphilis spirochete: making a living as a stealth pathogen.* Nature reviews. Microbiology, 2016. **14**(12): p. 744-759.
20. Anand, A., et al., *Bipartite Topology of Treponema pallidum Repeat Proteins C/D and I: OUTER MEMBRANE INSERTION, TRIMERIZATION, AND PORIN FUNCTION REQUIRE A C-TERMINAL beta-BARREL DOMAIN.* The Journal of biological chemistry, 2015. **290**(19): p. 12313-31.

21. Giacani, L., et al., *Antigenic variation in Treponema pallidum: TprK sequence diversity accumulates in response to immune pressure during experimental syphilis.* Journal of immunology, 2010. **184**(7): p. 3822-9.
22. Centurion-Lara, A., et al., *Treponema pallidum major sheath protein homologue Tpr K is a target of opsonic antibody and the protective immune response.* The Journal of experimental medicine, 1999. **189**(4): p. 647-56.
23. Hazlett, K.R., et al., *The TprK protein of Treponema pallidum is periplasmic and is not a target of opsonic antibody or protective immunity.* The Journal of experimental medicine, 2001. **193**(9): p. 1015-26.
24. Cox, D.L., et al., *Surface immunolabeling and consensus computational framework to identify candidate rare outer membrane proteins of Treponema pallidum.* Infection and immunity, 2010. **78**(12): p. 5178-94.
25. Anand, A., et al., *TprC/D (Tp0117/131), a trimeric, pore-forming rare outer membrane protein of Treponema pallidum, has a bipartite domain structure.* Journal of bacteriology, 2012. **194**(9): p. 2321-33.
26. Norris SJ, W.G., *Comparative genomics of spirochetes*, in *Pathogenic Treponema*, L.S. Radolf JD, Editor 2006, Caister Academic Press: Norwich, United Kingdom. p. 19-38.
27. Seshadri, R., et al., *Comparison of the genome of the oral pathogen Treponema denticola with other spirochete genomes.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(15): p. 5646-51.
28. Veith, P.D., et al., *Major proteins and antigens of Treponema denticola.* Biochimica et biophysica acta, 2009. **1794**(10): p. 1421-32.
29. Fenno, J.C., et al., *The opdB locus encodes the trypsin-like peptidase activity of Treponema denticola.* Infection and immunity, 2001. **69**(10): p. 6193-200.
30. Ishihara, K., et al., *Dentilisin activity affects the organization of the outer sheath of Treponema denticola.* Journal of bacteriology, 1998. **180**(15): p. 3837-44.
31. Miao, D., et al., *The Treponema denticola chymotrypsin-like protease dentilisin induces matrix metalloproteinase-2-dependent fibronectin fragmentation in periodontal ligament cells.* Infection and immunity, 2011. **79**(2): p. 806-11.
32. Lee, S.Y., et al., *Cleavage of Treponema denticola PrcA polypeptide to yield protease complex-associated proteins Prca1 and Prca2 is dependent on PrtP.* Journal of bacteriology, 2002. **184**(14): p. 3864-70.
33. Godovikova, V., et al., *Treponema denticola PrcB is required for expression and activity of the PrcA-PrtP (dentilisin) complex.* Journal of bacteriology, 2010. **192**(13): p. 3337-44.
34. Godovikova, V., M.P. Goetting-Minesky, and J.C. Fenno, *Composition and localization of Treponema denticola outer membrane complexes.* Infection and immunity, 2011. **79**(12): p. 4868-75.
35. McDowell, J.V., et al., *Analysis of a unique interaction between the complement regulatory protein factor H and the periodontal pathogen Treponema denticola.* Infection and immunity, 2009. **77**(4): p. 1417-25.
36. Berntsson, R.P., et al., *A structural classification of substrate-binding proteins.* FEBS Lett, 2010. **584**(12): p. 2606-17.

37. Fenno, J.C., et al., *Identification of a Treponema denticola OppA homologue that binds host proteins present in the subgingival environment.* Infection and immunity, 2000. **68**(4): p. 1884-92.
38. Haapasalo, M., et al., *Characterization, cloning, and binding properties of the major 53-kilodalton Treponema denticola surface antigen.* Infection and immunity, 1992. **60**(5): p. 2058-65.
39. 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.
40. Ritchie, T.K., et al., *Chapter 11 - Reconstitution of membrane proteins in phospholipid bilayer nanodiscs.* Methods Enzymol, 2009. **464**: p. 211-31.
41. Scheres, S.H., et al., *Image processing for electron microscopy single-particle analysis using XMIPP.* Nat Protoc, 2008. **3**(6): p. 977-90.
42. *Computer-aided interpretation of sedimentation data for proteins, in Analytical Ultracentrifugation in Biochemistry and Polymer Science*. 1992, Royal Society of Chemistry. 90-125.
43. Stafford, W.F., 3rd, *Boundary analysis in sedimentation transport experiments: a procedure for obtaining sedimentation coefficient distributions using the time derivative of the concentration profile.* Analytical biochemistry, 1992. **203**(2): p. 295-301.
44. Philo, J.S., *Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques.* Analytical biochemistry, 2006. **354**(2): p. 238-46.
45. Schuck, P., *Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling.* Biophysical journal, 2000. **78**(3): p. 1606-19.
46. Brusca, J.S. and J.D. Radolf, *Isolation of integral membrane proteins by phase partitioning with Triton X-114.* Methods Enzymol, 1994. **228**: p. 182-93.
47. Caimano, M.J., et al., *The Treponema denticola major sheath protein is predominantly periplasmic and has only limited surface exposure.* Infect Immun, 1999. **67**(8): p. 4072-83.
48. Cox, D.L., et al., *Treponema pallidum in gel microdroplets: a novel strategy for investigation of treponemal molecular architecture.* Molecular microbiology, 1995. **15**(6): p. 1151-64.
49. Mitchell, A., et al., *The InterPro protein families database: the classification resource after 15 years.* Nucleic acids research, 2015. **43**(Database issue): p. D213-21.
50. Yang, J., et al., *The I-TASSER Suite: protein structure and function prediction.* Nature methods, 2015. **12**(1): p. 7-8.
51. Mathias, R.A., et al., *Triton X-114 phase separation in the isolation and purification of mouse liver microsomal membrane proteins.* Methods, 2011. **54**(4): p. 396-406.
52. Bayburt, T.H. and S.G. Sligar, *Membrane protein assembly into Nanodiscs.* FEBS Lett, 2010. **584**(9): p. 1721-7.
53. Puthenveetil, R., K. Nguyen, and O. Vinogradova, *Nanodiscs and Solution NMR: preparation, application and challenges*, in *Nanotechnology Reviews*2016.

54. Puthenveetil, R. and O. Vinogradova, *Optimization of the design and preparation of nanoscale phospholipid bilayers for its application to solution NMR.* Proteins, 2013. **81**(7): p. 1222-31.
55. Bayburt, T.H., Y.V. Grinkova, and S.G. Sligar, *Assembly of single bacteriorhodopsin trimers in bilayer nanodiscs.* Archives of biochemistry and biophysics, 2006. **450**(2): p. 215-22.
56. Hagn, F., et al., *Optimized Phospholipid Bilayer Nanodiscs Facilitate High-Resolution Structure Determination of Membrane Proteins.* J Am Chem Soc, 2013.
57. Park, S.H., et al., *Nanodiscs versus macrodiscs for NMR of membrane proteins.* Biochemistry, 2011. **50**(42): p. 8983-5.
58. Caimano, M.J., et al., *The Treponema denticola major sheath protein is predominantly periplasmic and has only limited surface exposure.* Infection and immunity, 1999. **67**(8): p. 4072-83.
59. Bian, X.L., et al., *Mutagenesis of a novel gene in the prcA-prtP protease locus affects expression of Treponema denticola membrane complexes.* Infection and immunity, 2005. **73**(2): p. 1252-5.
60. Bamford, C.V., et al., *The chymotrypsin-like protease complex of Treponema denticola ATCC 35405 mediates fibrinogen adherence and degradation.* Infection and immunity, 2007. **75**(9): p. 4364-72.
61. Henz, S.R., et al., *Whole-genome prokaryotic phylogeny.* Bioinformatics, 2005. **21**(10): p. 2329-35.
62. Edwards, A.M., et al., *Binding properties and adhesion-mediating regions of the major sheath protein of Treponema denticola ATCC 35405.* Infection and immunity, 2005. **73**(5): p. 2891-8.
63. Fraser, C.M., et al., *Complete genome sequence of Treponema pallidum, the syphilis spirochete.* Science, 1998. **281**(5375): p. 375-88.
64. Frickey, T. and A. Lupas, *CLANS: a Java application for visualizing protein families based on pairwise similarity.* Bioinformatics, 2004. **20**(18): p. 3702-4.
65. Hagan, C.L., T.J. Silhavy, and D. Kahne, *beta-Barrel membrane protein assembly by the Bam complex.* Annual review of biochemistry, 2011. **80**: p. 189-210.
66. Rowley, G., et al., *The periplasmic chaperone Skp is required for successful Salmonella Typhimurium infection in a murine typhoid model.* Microbiology, 2011. **157**(Pt 3): p. 848-58.
67. Heuck, A., A. Schleiffer, and T. Clausen, *Augmenting beta-augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins.* Journal of molecular biology, 2011. **406**(5): p. 659-66.
68. Paramasivam, N., M. Habeck, and D. Linke, *Is the C-terminal insertional signal in Gram-negative bacterial outer membrane proteins species-specific or not?* BMC genomics, 2012. **13**: p. 510.
69. Bennion, D., et al., *Dissection of beta-barrel outer membrane protein assembly pathways through characterizing BamA POTRA 1 mutants of Escherichia coli.* Molecular microbiology, 2010. **77**(5): p. 1153-71.

70. Paster, B.J., and Dewhirst, F.E., *The phylogenetic diversity of the genus Treponema* in *Pathogenic Treponema: Molecular and Cellular Biology*, J.D. Radolf, and Lukehart, S.A., Editor 2006, Caister Academic Press: Norwich, UK. p. 9-18.
71. Walfield, A.M., P.A. Hanff, and M.A. Lovett, *Expression of Treponema pallidum antigens in Escherichia coli.* Science, 1982. **216**(4545): p. 522-3.
72. Giuseppe, P.O., et al., *The crystal structure of the leptospiral hypothetical protein LIC12922 reveals homology with the periplasmic chaperone SurA.* Journal of structural biology, 2011. **173**(2): p. 312-22.
73. Cogoni, V., et al., *Treponema denticola chymotrypsin-like proteinase (CTLP) integrates spirochaetes within oral microbial communities.* Microbiology, 2012. **158**(Pt 3): p. 759-70.
74. Abiko, Y., et al., *Characterization of Treponema denticola mutants defective in the major antigenic proteins, Msp and TmpC.* PloS one, 2014. **9**(11): p. e113565.
75. Schubert, W.D., et al., *Structure of internalin, a major invasion protein of Listeria monocytogenes, in complex with its human receptor E-cadherin.* Cell, 2002. **111**(6): p. 825-36.
76. Brinkman, M.B., et al., *A novel Treponema pallidum antigen, TP0136, is an outer membrane protein that binds human fibronectin.* Infection and immunity, 2008. **76**(5): p. 1848-57.
77. Cameron, C.E., et al., *Treponema pallidum fibronectin-binding proteins.* Journal of bacteriology, 2004. **186**(20): p. 7019-22.
78. Houston, S., et al., *Activation and proteolytic activity of the Treponema pallidum metalloprotease, pallilysin.* PLoS pathogens, 2012. **8**(7): p. e1002822.
79. Higgins, C.F. and M.M. Hardie, *Periplasmic protein associated with the oligopeptide permeases of Salmonella typhimurium and Escherichia coli.* Journal of bacteriology, 1983. **155**(3): p. 1434-8.
80. Klepsch, M.M., et al., *Escherichia coli peptide binding protein OppA has a preference for positively charged peptides.* Journal of molecular biology, 2011. **414**(1): p. 75-85.

Appendix

Chapter A1

Investigating the interaction of the SH3 domain of Src kinase with the cytoplasmic tail of Integrin β3

Adapted from:

Structural insights into the recognition of β3 integrin cytoplasmic tail by the SH3 domain of Src kinase

Katyal P, Puthenveetil R and Vinogradova O, Protein Sci. 2013 Oct; 22(10):1358-65

Introduction

Integrins comprise a large family of bidirectional signaling receptors that tether extracellular matrix to the cytoskeleton and regulate cellular behavior via classical signaling pathways [1]. These major cell surface receptors mediate cell adhesion, migration, proliferation, differentiation, and programmed cell death through 'inside-out' and 'outside-in' signaling mechanisms. Over the last decade, substantial progress has been made in structural characterization of Integrins within these signaling cascades. Inside-out signaling is initiated by the separation of heterodimeric cytoplasmic tails causing an overall conformational change that propagates across the membrane to the extra-cellular domain of the receptor [2]. In contrast, ligand binding to the extracellular domain leads to conformational changes across the receptor transducing the signal across the membrane, to the cytoplasm further leading to a series of downstream signaling eventsreferred to as outside-in signaling [3]. Major platelet integrin $\alpha_{IIb}\beta_3$, an archetypical representative of the class, regulates platelet aggregation and serves as a molecular scaffold interacting with various extracellular and intracellular binding partners. Shattil and co-workers [4, 5] have shown that outside-in signaling in platelets was mediated by a direct interaction of carboxyl terminus β_3 integrin with Src kinase.

In humans, c-Src, a founding member of Src family non-receptor tyrosine kinases (SFKs), is a 60 kDa membrane anchoring cytosolic protein. The role of c-Src (hereafter referred to as Src) has been implicated in various cellular functions including cell division, motility, and cell survival. It is abundantly expressed in platelets and is vital to many vascular processes [6]. Src is a multi-domain protein composed of: (i) an N-terminal myristoylation site , (ii) SH4 domain, (iii) SH3 domain, (iv) SH2 scaffold domain, and (v)

catalytic kinase domain, followed by a tyrosine (Y^{527}) containing C-terminal regulatory sequence. In resting cells, intra- and inter-molecular interactions mediated by Src SH2 domain and Csk (C-terminal Src kinase) respectively, maintain Src kinase in its auto-inhibited state, which is characterized by the phosphorylation of its Tyr^{527}. This inactive form of Src is proposed to bind the resting β_3 integrin cytoplasmic tail through its SH3 domain [7]. Once Integrins engage with their extracellular ligands, sequential $\alpha_{IIb}\beta_3$ microclustering induces the Src activation by a series of events including auto-phosphorylation of Tyr^{416} in the activation loop, dissociation of Csk and recruitment of phosphatase (PTP-1B) responsible for dephosphorylating Tyr^{527} [7]. Thus structural characterization of β_3 integrin interaction with Src should help in the comprehension of outside-in signaling events. This is particularly true for platelets, osteoclasts and endothelial cells, where both proteins are found in abundance [4, 7].

Src SH3 domain, recognized as the key integrin binding site [5], has been extensively studied and structurally characterized in both free and ligand bound states (PDB: 1RLQ, 1QWF, 1NLP). It consists of two antiparallel β sheets positioned at right angles to one another. The β strands are linked via RT loop, n-Src loop, distal loop and a short 3_{10}-helix [8, 9]. In general, SH3 domains are known to favor peptides bearing a PxxP motif [10]. The PXXP motif adopts a polyproline type II helix and binds between the RT loop and n-Src loop. The selectivity of these peptides is further enhanced by basic residues, arginines or lysines, which flank the core motif. In addition, other mechanisms may contribute towards the specificity of interactions involving SH3 domain [11]. Although two classes of peptides, class I - (R/K)xxPxxP and class II - PxxPx(R/K) [10-12], are considered as canonical SH3 binding targets, accumulating evidence suggests that SH3

can also recognize non-PxxP motifs [10, 13-15]. The exact molecular detail of such recognition still remains unclear, however some non-PXXP peptides have been shown to occupy the same interface as PXXP motifs [16].

In this chapter, we present NMR data and a docked model of SH3 with cytoplasmic integrin β_3. We also show that the C-terminal RGT motif of β_3 adopts a partial PPII type helix facilitating its interaction with the RT and n-Src loops of SH3 in an orientation opposite to that obtained from X-ray studies [17].

Result Summary

Previous studies have identified the C-terminal RGT^{762} motif of β_3 integrin [4, 5, 18] as the binding site for Src kinase through its SH3 domain. Here we map the binding site of this interaction onto the surface of SH3 using solution NMR. First, the full length unlabeled β_3 cytoplasmic tail (hereafter referred to as β_3CT) was titrated against ^{15}N-labeled SH3. The ^{15}N-HSQC spectra of SH3 clearly demonstrated a significant chemical shift differences when compared to the spectrum from the free protein. The major amino acids undergoing perturbations were from the RT loop (residues R^{95} to L^{100}) and the n-Src loop (E^{115} to W^{118}) (Figure 1A, right panel). The details of this interaction was further corroborated by the use of a β3 heptapeptide, comprising the last seven β_3 residues including the RGT motif. Chemical shift perturbation pattern obtained with β_3 heptapeptide closely matched with the full-length β_3CT. The similar binding interfaces between the two confirm that the C-terminal RGT motif of β_3 is sufficient to bind the Src SH3 domain.

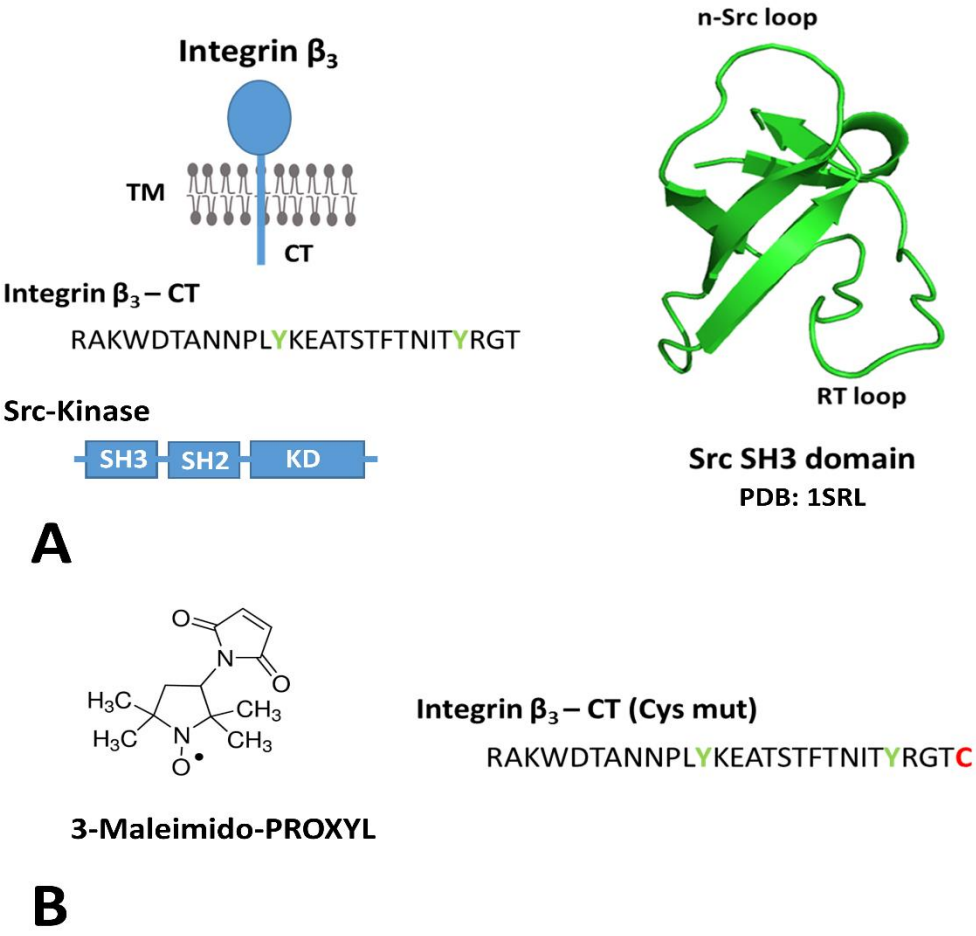


Figure 1: (A) Left panel: Schematic representation showing the modular structure of Integrin β_3 and Src Kinase along with the sequence of the Integrin β_3 cytoplasmic tail where the two tyrosine residues undergoing phosphorylation are highlighted in green. **Right panel:** Structure of the SH3 domain of Src Kinase displaying the two loop regions where chemical shift perturbations were observed in the presence of β_3 tail. **(B)** Structure of the m-PROXYL tag used in the PRE experiments and the sequence of β_3 mutant with the additional cysteine shown in red which facilitated the attachment of the tag.

In order to ascertain the orientation of the RGT motif at the binding site of the SH3 domain we carried out Paramagnetic Relaxation Enhancement (PRE) experiments. In order to attach a spin label (PRE tag), a β_3 mutant with an additional cysteine residue after the terminal YRGT sequence (Figure 1B) was introduced using QuikChange site-

directed mutagenesis kit (Agilent, USA). Prior to the reaction, the mutant was reduced by 1 mM TCEP and subsequently removed by PD-10 column (GE healthcare, USA). The reduced β_3 mutant was allowed to react overnight with either an excess of cysteine specific spin label, 3-maleimido-PROXYL (Figure 1B), hereafter referred to as mProxyl (Sigma-Aldrich, USA), or 1-Oxyl-2,2,5,5-tetramethylpyrroline-3-methyl) methanethiosulfonate, hereafter referred to as MTSL (Alexis Biochemicals, USA). The reactions were performed at room temperature at a pH of 7.0. The unreacted spin label was separated from the tagged mutant by reverse phase HPLC on PROTO C4 column (The Nest Group, USA). Attachment of the tag was confirmed through mass spectrometry. ^{15}N-HSQC spectrum of SH3 in the presence of tagged β_3 mutant, along with untagged mutant as a control, were collected on a 600 MHz magnet (Agilent, USA) equipped with cold probe. The nitroxide radical from the spin label attenuates peak intensities of nearby residues allowing direct mapping of its location onto the binding surface. Since the label is attached next to the RGT motif, maximum attenuation is expected for SH3 residues adjacent to the ones involved in binding of this motif. To test whether the addition of a paramagnetic tag affected the interaction, we compared the spectra of SH3 in the presence of wild type β_3CT and its tagged mutant (Figure 2). Since no significant differences were observed we proceeded with the PRE experiments. ^{15}N-HSQC spectra were collected for labeled SH3 in the presence of untagged PRE-tagged β_3CT. Maximum attenuation was observed for the residue N^{135} with a moderate reduction for Y^{136} suggesting their close proximity to the tag. A moderate decrease in the intensities of the n-Src loop also confirmed its positioning near-to the β_3 RGT motif, while the RT loop was less affected, as can be observed from a comparison of the intensity map

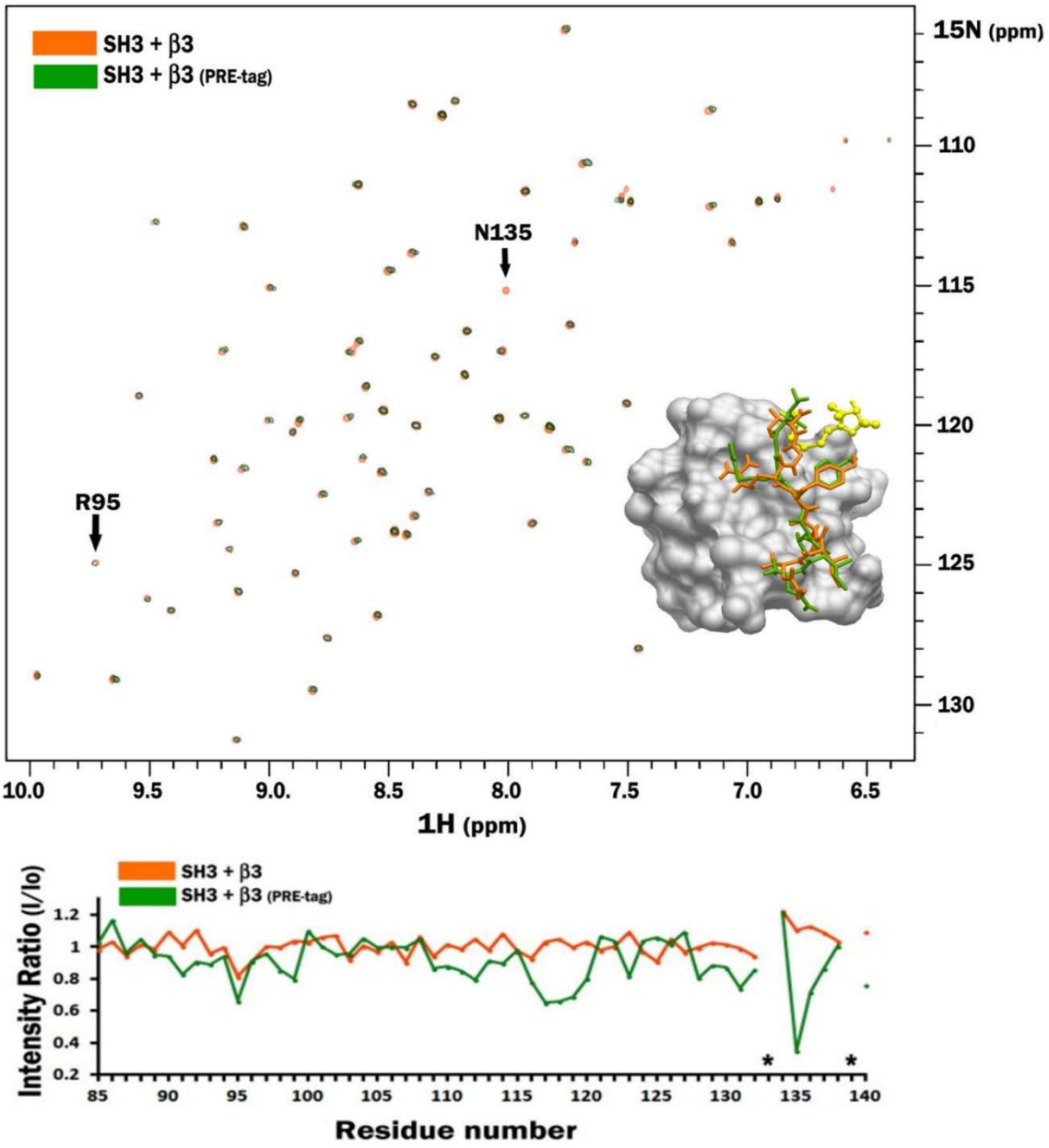


Figure 2: (A) Top: Superimposition of SH3 ^{15}N-HSQC spectra in the presence of wild type β_3CT (orange) and with paramagnetic tag attached to its C-terminus end (green). The inset shows surface representation of the docked SH3 model with the heptapeptide (orange) and heptapeptide with PRE tag (green, attached tag is shown in yellow). **Bottom:** Difference in intensity ratios from PRE experiments of the backbone amide groups of SH3 in the presence of tagged β_3CT (green) and untagged β_3CT (orange). The asterisk mark represents proline residues.

obtained for SH3 in the presence of tagged and untagged (control) β_3CT (Figure 2, bottom panel). The N^{135} residue from the SH3 domain in close proximity to the nitric oxide from the tag, allowed us to reliably predict the orientation of the RGT motif within the SH3 binding interface.

Conclusion

We deduced the molecular details of the interaction between the Src SH3 domain and Integrin β_3 using NMR. Chemical shifts mapping experiments depict the C-terminus of β_3 located between the RT and n-Src loops of SH3 domain. According to our PRE results, the RGT motif binds to the pocket formed by the n-Src loop. The preceding NITY motif which makes intermediate contacts with the RT loop. The experimental data allowed us to obtain a reliable docked model of the complex using the HADDOCK web server. Our docking model provides a more accurate representation of the binding interface (Figure 3B, left panel) than the recent X-ray model containing the RGT tri-peptide [17] (Figure 3B, right panel. The attachment of the paramagnetic tag to the C-terminal end of RGT sequence allowed us to orient the RGT motif unambiguously with respect to its SH3 binding pocket (Figure 3A). If the orientation from the X-ray structure were true we would have seen the PRE effect opposite to the current position on SH3 surface. Moreover, the weak electron density map, used to fit the small RGT tri-peptide, could easily lead to an ambiguous orientation. (iii) Our model also supports the biochemical studies that have clearly shown the importance of R^{760} from integrin β_3 YRGT motif and D^{117}, W^{118} and Y^{131} residues from the SH3 domain [17]. We also found that the SH3 domain doesn't bind to

the phosphorylated Integrin β_3, suggesting that Src can be associated with integrin through its SH3 domain only in the resting non-phosphorylated state of the receptor. We presume this weak but constitutive interaction would keep Src tethered to the integrin and consequently bring two or more Src kinases in close proximity upon integrin activation via clustering [5]; which in turn, would aid in trans-activation of the kinase through auto-phosphorylation of Y^{418} in the Src activation loop.

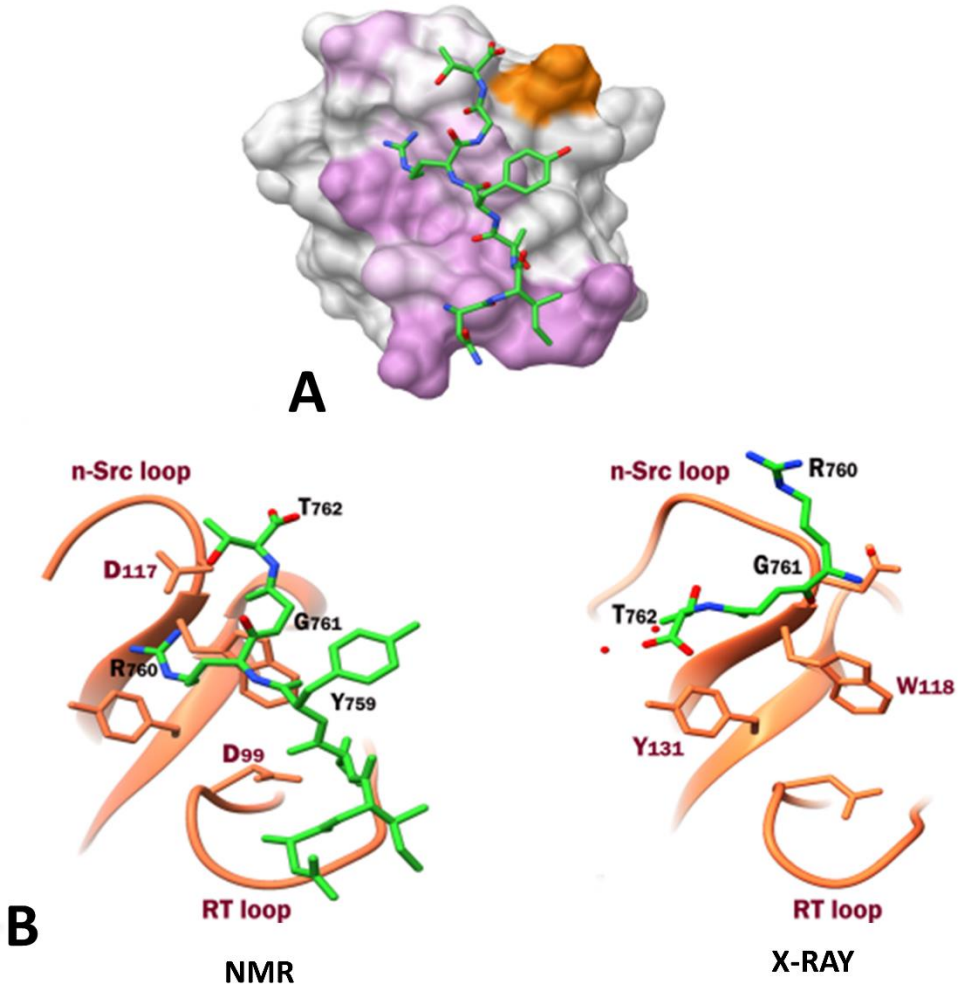


Figure 3: (A) β_3 heptapeptide with spin label is docked onto Src SH3 domain using the HADDOCK program. The decreases in intensity, associated with paramagnetic spin label, is mapped with orange on the surface of SH3. **(B)** Structural comparison showing the difference in the orientation of the RGT motif obtained from the NMR study with respect to reported X-ray structure (PDB:4HXJ) [17].

References

1. Hynes, R.O., *Integrins: bidirectional, allosteric signaling machines.* Cell, 2002. **110**(6): p. 673-87.
2. Vinogradova, O., et al., *A structural mechanism of integrin alpha(IIb)beta(3) "inside-out" activation as regulated by its cytoplasmic face.* Cell, 2002. **110**(5): p. 587-97.
3. Liu, S., D.A. Calderwood, and M.H. Ginsberg, *Integrin cytoplasmic domain-binding proteins.* J Cell Sci, 2000. **113 (Pt 20)**: p. 3563-71.
4. Arias-Salgado, E.G., et al., *Specification of the direction of adhesive signaling by the integrin beta cytoplasmic domain.* J Biol Chem, 2005. **280**(33): p. 29699-707.
5. Arias-Salgado, E.G., et al., *Src kinase activation by direct interaction with the integrin beta cytoplasmic domain.* Proc Natl Acad Sci U S A, 2003. **100**(23): p. 13298-302.
6. Roskoski, R., Jr., *Src kinase regulation by phosphorylation and dephosphorylation.* Biochem Biophys Res Commun, 2005. **331**(1): p. 1-14.
7. Shattil, S.J., *Integrins and Src: dynamic duo of adhesion signaling.* Trends Cell Biol, 2005. **15**(8): p. 399-403.
8. Cordier, F., et al., *Ligand-induced strain in hydrogen bonds of the c-Src SH3 domain detected by NMR.* J Mol Biol, 2000. **304**(4): p. 497-505.
9. Yu, H., M.K. Rosen, and S.L. Schreiber, *1H and 15N assignments and secondary structure of the Src SH3 domain.* FEBS Lett, 1993. **324**(1): p. 87-92.
10. Mayer, B.J., *SH3 domains: complexity in moderation.* J Cell Sci, 2001. **114**(Pt 7): p. 1253-63.
11. Zarrinpar, A., R.P. Bhattacharyya, and W.A. Lim, *The structure and function of proline recognition domains.* Sci STKE, 2003. **2003**(179): p. RE8.
12. Feng, S., et al., *Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions.* Science, 1994. **266**(5188): p. 1241-7.
13. Kang, H., et al., *SH3 domain recognition of a proline-independent tyrosine-based RKxxYxxY motif in immune cell adaptor SKAP55.* EMBO J, 2000. **19**(12): p. 2889-99.
14. Kishan, K.V., et al., *The SH3 domain of Eps8 exists as a novel intertwined dimer.* Nat Struct Biol, 1997. **4**(9): p. 739-43.
15. Mongiovi, A.M., et al., *A novel peptide-SH3 interaction.* EMBO J, 1999. **18**(19): p. 5300-9.
16. Kami, K., et al., *Diverse recognition of non-PxxP peptide ligands by the SH3 domains from p67(phox), Grb2 and Pex13p.* EMBO J., 2002. **21**(16): p. 4268-76.
17. Xiao, R., et al., *Structural framework of c-Src activation by integrin beta3.* Blood, 2013. **121**(4): p. 700-6.
18. Ablooglu, A.J., et al., *Antithrombotic effects of targeting alphaIIbbeta3 signaling in platelets.* Blood, 2009. **113**(15): p. 3585-92.

Appendix

Chapter A2

Binding of HMBPP promotes a conformational change to the internal domain of BTN3A1 mediated through an alteration of the B30.2 domain with respect to the JM region

Adapted from:

HMBPP binds to the juxta membrane region of BTN3A1 and together with B30.2 promotes a conformational change to its intracellular domain
Nguyen K, Puthenveetil R, Li J, Lin X, Hsiao CH, Vinogradova O, and Wiemer AJ. (*in prep*)

Introduction

Butyrophilins are immune modulators belonging to the B7 family of co-stimulatory molecules that enhance/inhibit T cell activation [1, 2]. They are implicated in several diseases viz. type-1 diabetes [3], ovarian cancer [4], dyslipidaemia [5], myocardial infarction [6], rheumatoid arthritis [7], sarcoidosis [8] and chronic kidney disease [9] to name a few. B7 family of proteins typically contain an extracellular IgV and IgC domain followed by a transmembrane region and an intracellular cytoplasmic tail [10]. Butyrophilin (BTN) and butyrophilin like (BTNL) proteins are members of the broad B7 family with the difference of having an intracellular signaling domain B30.2 in lieu of the cytoplasmic tail (Figure 1A). Contrary to the known binding ligands of B7 proteins (B7.1 and B7.2 on Antigen Presenting Cells, APC) like CD28/CTLA4 (on T cells) the ligands for butyrophilins remain elusive. Butyrophilins influence a subset of T cell lymphocytes made up of γ and δ heterodimer called the γδ TCR [11]. This class of T cells are less understood than their αβ counterpart. BTN homologs are divided into three types (BTN1, 2 and 3) where each type further has either one or three subtypes [BTN1A1, BTN2 (A1, A2, A3) and BTN3 (A1, A2, A3)] with 90% sequence identity within each subtype and 50% identity across types [12]. BTN3 family of butyrophilins are exclusively found in humans and are of the best studied thus far. Antibodies raised against BTN3A shed some light on the possible roles of the different subtypes of this family. Both, 232.5 and 103.2 antibodies can mediate inhibitory effect on T cell activation and proliferation. The modulation of the cellular activation machinery was achieved through an agonistic effect on BTN3A3 and antagonistic effect on BTN3A1 brought about by 232.5 and 103.2 antibodies respectively [13, 14]. Additionally, it has been shown that antibody 20.1 can differentially regulate the

production of IFNγ depending on the engagement of either BTN3A1 or BTN3A2 [15]. The agonistic binding of 20.1 antibody to a small subset of γδ T cells (Vγ9Vδ2 T cells) mimicked how the cells responds to phosphoantigen (pAg), which was abrogated by the antagonist 103.2 antibody [16]. BTN3A1 was specifically identified as the butyrophilin which led to the activation of Vγ9Vδ2 T cells in the presence of a pAg [17].

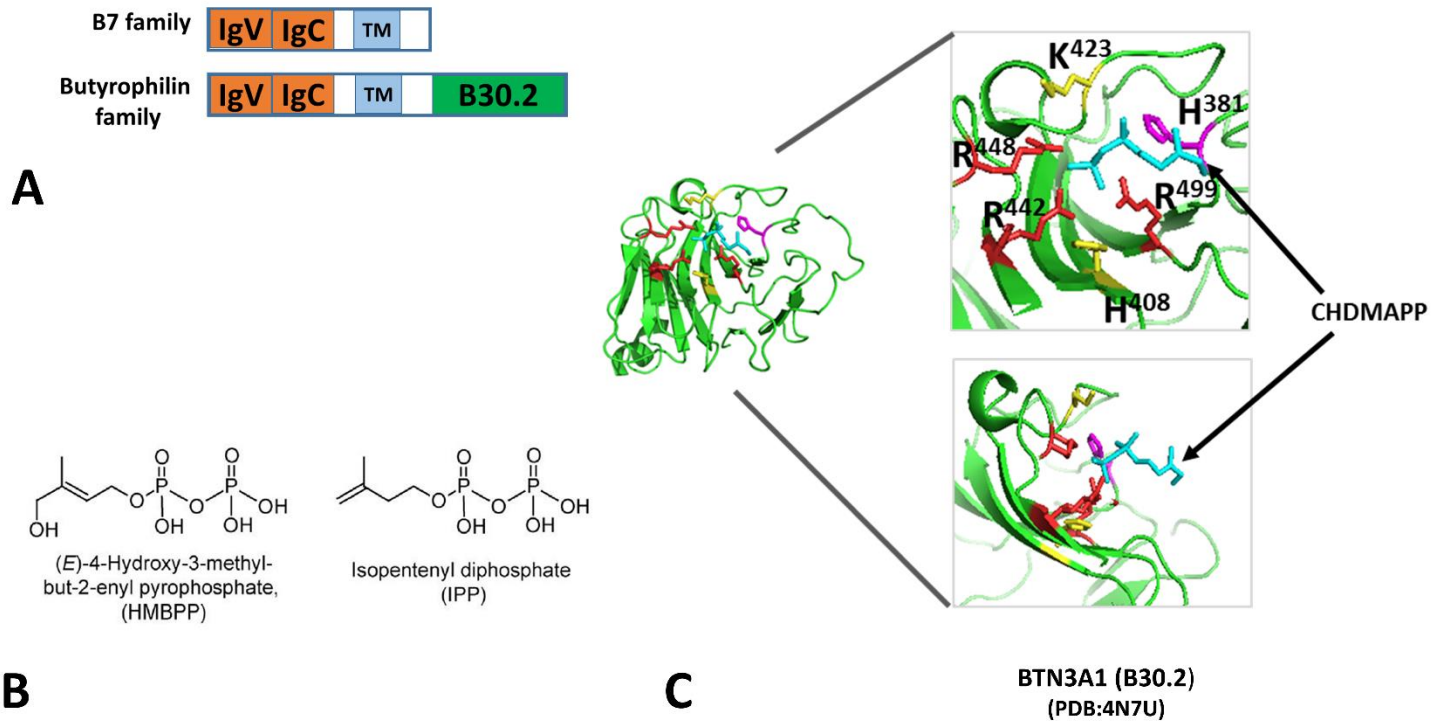


Figure 1: (A) Schematic displaying the difference in the organization of different domains in the B7 and classical Butyrophilin family. **(B)** Molecular structure of two phosphoantigens HMBPP, used in this study, and endogenous IPP. **(C)** Atomic structure of B30.2 domain from BTN3A1 highlighting the residues critical for interaction with phosphoantigen (CHDMAPP) in blue. The core interactions are formed by three Arginine residues colored in red and further supported by Histidine (yellow/magenta) and Lysine (yellow) residues. Previous singular mutation of H^{381} (magenta) to arginine has demonstrated a loss or gain of function in activating γδT cells.

Phosphoantigens are intermediates of isoprenoid bio-synthesis. IPP (Isopentyl dipyrophosphate) is a pAg that is formed as a metabolic intermediate of the mevalonate

or cholesterol synthesis pathway found in most prokaryotic and eukaryotic cells. HMBPP [(E)-4-hydroxyl-3-methyl-but-2-enyl-pyrophosphate], also a pAg produced as a metabolic intermediate of the non-mevalonate pathway is exclusively found in Gram-negative and Gram-positive bacteria (Figure 1B). HMBPP, with an affinity of <1μM is a more potent activator Vγ9Vδ2 T cells than IPP having an affinity in the mM range [18, 19]. pAg (CHDMAPP) was later shown to bind to the B30.2 intracellular domain (4N7U) of BTN3A1 [18], using X-ray crystallography along with the juxta-membrane (JM) region through NMR using the pAg, HMBPP [14]. The intracellular B30.2 domain consists of a SPRY domain with two sets of β sheets (A and B sheets) and PRY β strands at the N-terminus. A positively charged binding pocket for the pAg has been identified on this domain involving residues R^{442}, R^{448}, R^{449}, H^{381}, H^{408} and K^{423} (Figure 1C). It has been speculated that the binding of the pAg brings about a change in the extracellular domain of BTN3A1 leading to the activation of the Vγ9Vδ2 T cells. In this chapter, we highlight the role of the JM region in propagating the effect of the intracellular pAg binding to the extracellular domain of BTN3A1. We achieved this through the elegant use of SAXS (Small angle x-ray scattering) and NMR methodologies in association with mutagenesis of the specific residues in the juxta membrane region further corroborated through functional studies.

Result Summary

Two Human BTN3A1 (Accession No.: O00481) constructs representing the JM+B30.2 domain (BFI): Q^{272} - A^{513}; and JM region: Q^{272} – P^{339} were cloned into 6X His-tag pET21b vector [19]. The plasmids were transformed into *E.coli* BL21 (DE3) and grown in either LB for SAXS or minimal media supplemented with $^{15}NH_4Cl$ (Cambridge Isotope Lab., USA) for NMR experiments. For SAXS experiment, BFI protein was grown in LB at 37°C and induced with 1mM IPTG. Harvested cells were lysed and centrifuged to remove insoluble material by centrifugation at 15,000 rpm. The supernatant was mixed with Ni-NTA resin (QIAGEN, USA), washed with TBS pH 7.5 containing 20mM Imidazole and eluted in the same buffer supplemented with 400mM Imidazole. The eluate was subjected to SEC in buffer containing 20 mM TRIS, 100mM NaCl, 5 mM β-mercaptoethanol, 5% glycerol, pH 7.5 on S75 column (GE Healthcare, USA). Peak fractions containing pure protein were concentrated to three concentrations in the range of (0.6-3 mg/ml). Duplicates of BFI samples were prepared with and without HMBPP. SAXS was performed at the CHESS G1 station using a 9.8 keV 250 μm × 250 μm Xray beam. Data were collected on a Pilatus 100 K detector with a sample-to-detector distance of 1.5 m. 15s exposure with 1 s intervals were taken for the protein and its corresponding buffer samples loaded into an in vacuo oscillating flow cell. Radiation damage was accessed by overlaying the first and the last exposure profile obtained from the sample data. Buffer measurements were performed twice, before and after the sample, to monitor changes in the background scattering. Two dimensional scattering from the detector was converted

to one dimensional scattering image as a function of momentum transfer *q* ($q = 4\pi\sin(\theta)/\lambda$, where 2π is the scattering angle and λ is the

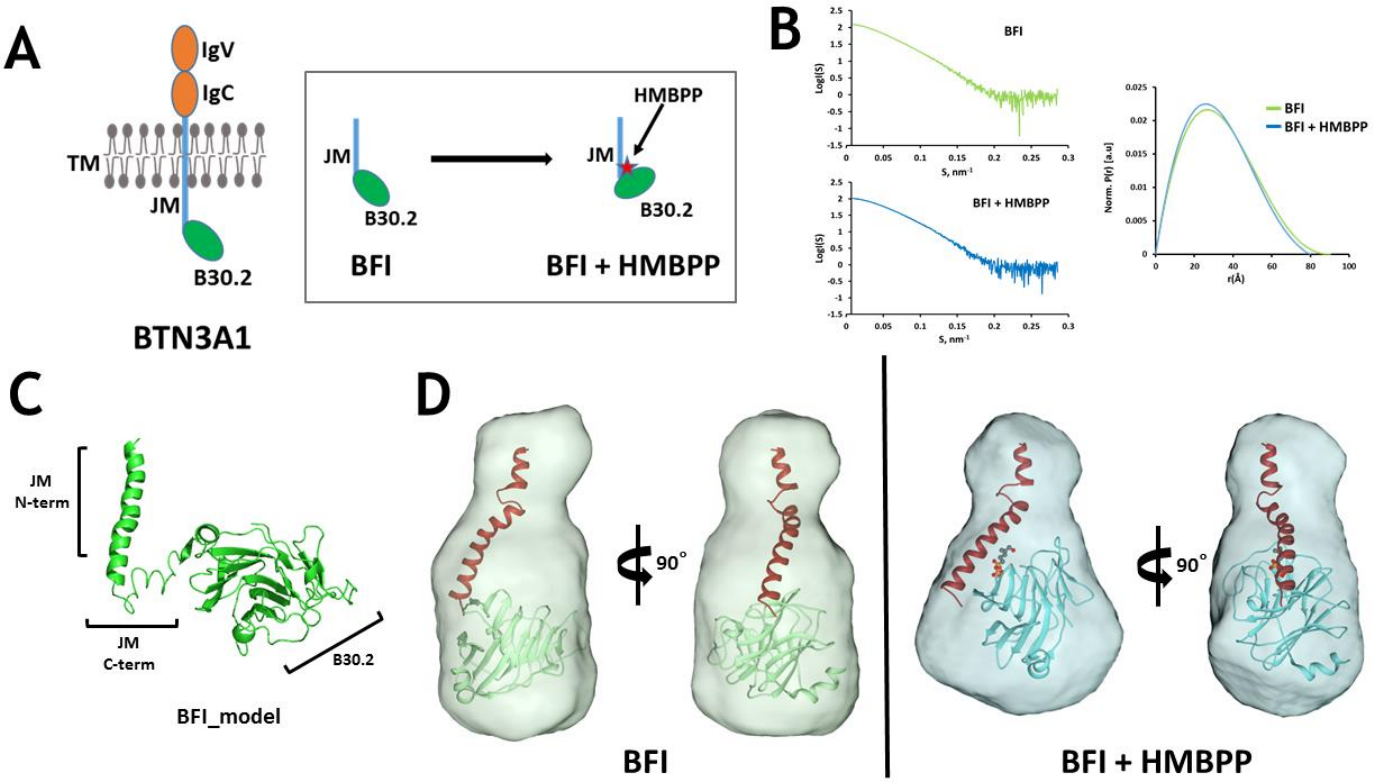


Figure 2: (A) Schematic representation of the BTN3A1 receptor. Box: Visual aid displaying the rearrangement of BFI region (JM+B30.2 domain) in the presence of HMBPP. **(B)** Scattering curve for BFI in the presence (blue) and absence (green) of HMBPP. Overlay of the normalized pair-distance distribution curves P(r) of the scattering curves. **(C)** Model of BFI generated through homology modelling exemplifying the flexible C-terminal end of the JM region. **(D)** Surface representation of the average ab initio model for BFI and BFI+HMBPP rendered through SITUS.

wavelength) using the software BioXTAS RAW and PRIMUS [20]. Individual scattering profiles were averaged for both, buffer and sample. Background scatter, from the buffer, was subtracted from the sample profile to obtain the final protein scattering image. The radii of gyration (*Rg*) were determined from Guinier plots using AUTORG [21] and compared with the values calculated from pair distance distribution functions [P(r)] using

the program AUTOGNOM [22]. Fifteen independent runs in DAMMIN [23] were performed and the most probable model was computed with DAMAVER [24]. The final envelope representation was obtained using SITUS [25].

SAXS allows for the determination of low resolution structural envelope of proteins in solution. In contrast to models that predicts an extended helical conformation of the JM region [26], our SAXS data clearly suggests a relative globular shape than what can be achieved by rigid, extended coiled-coil structures. This observation can easily be attributed do the flexible, unstructured C-terminal end of JM, as shown in Figure 2C, which may act like a hinge helping the B30.2 domain to reorient itself with respect to the JM region. Our hypothesis is further bolstered by the envelope model of BFI obtained in the presence of a phosphoantigen HMBPP. Overlay of the P(r) distance distribution curve obtained for BFI clearly shows a reduction in the radius of gyration (R_g) by ~ 2Å in the presence of HMBPP(Figure 2B), which was further corroborated by an identical difference in the R_g values obtained from Guinier analysis. A more striking difference was also noted in the D_{max} values which reduced by ~8 Å. Taken together these differences manifest into a smaller envelope for BFI in the presence of pAg (Figure 2C). Thus, HMBPP binding to BTN3A1 triggers a measurable conformational change that results in the movement of B30.2 domain along the JM region, potentially in the direction toward the membrane, shown schematically in Figure 2A. A visual understanding of the event is illustrated by the homology modeled N-terminal JM region and B30.2 domain (4N7I/4N7U) fitted using SITUS [25] into the averaged ab initio envelope and aligned together in CHIMERA [27] for both the apo and ligand-bound states. Overall, the SAXS results are consistent with our hypothesis of BTN3A1 intracellular conformational changes.

In addition to SAXS experiments, ^{1}H-^{15}N HSQC NMR experiments with ^{15}N-labeled JM region along with two mutants JM1(T304A) and JM2(S296A,T297A) were carried out in the presence and absence of HMBPP. All experiments were carried out on 600 MHz (Agilent, USA) equipped with a cold probe. The experiments reveal that both $T^{297/304}$ peaks experienced broadening along with the adjoining serine displaying perturbations. Identification of serine/threonine were achieved by overlaying the JM HSQC spectra with the JM mutants. Moreover HMBPP addition to the ^{15}N- labeled JM mutants display no perturbation/broadenings confirming the binding of HMBPP to the JM region. Moreover the binding was further confirmed through *in vivo* experiment where wild type K562 cells were transfected with BTN3A1 and two mutants (BTN3A1-T304A and BTN3A1-S296A, T297A).These cells were mixed with HMBPP and washed to remove unbound HMBPP. Post washing the cells were mixed with Vγ9Vδ2 T cells and monitored for IFNγ through ELISA to access the activation of T cells. Both mutants show reduced activation of Vγ9Vδ2 T cells after treatment with HMBPP. The major difference between IPP and HMBPP lies in the presence of a hydroxyl group at the opposite to the pyrophosphates. Efforts are underway to ascertain the effect of the hydroxyl group on the binding of pAg to the JM region of BTN3A1 through the use of DMAPP (dimethylallyl pyrophosphate) an isomer of IPP lacking the hydroxyl group, and a metabolic intermediate of the mevalonate pathway.

Conclusion

A combination of SAXS, NMR and in vivo experimentation successfully demonstrates the binding of the phosphoantigen HMBPP to the intracellular juxta membrane and B30.2 domain of the receptor protein BTN3A1. The interaction was briefly noted in a previous publication by Hsiao et al. [19] and the present work provides more concrete evidence supporting this association. Bioinformatic analysis shows that the B30.2 domain is not conserved in the pAg binding region as seen by the magenta and blue colors (Figure 3A) which sheds some light on the evolution of the Butyrophilin receptors were the distal C-terminal region of the receptor is modified to interact with different binding partners. Additionally, sequence comparison of the putative BTN3A3 secretory protein (Accession Id: A0A089GIA6) with BTN3A1/A3 reveals the potential importance of the JM region which is absent in the secretory form (Figure 3B). Though experimental evidence is required to prove the existence of this form. It would be interesting to observe weather the R^{381} to H^{381} mutation would still lead to a gain of function, through the activation of γδ T cells for these secretory BTN3A3 proteins.

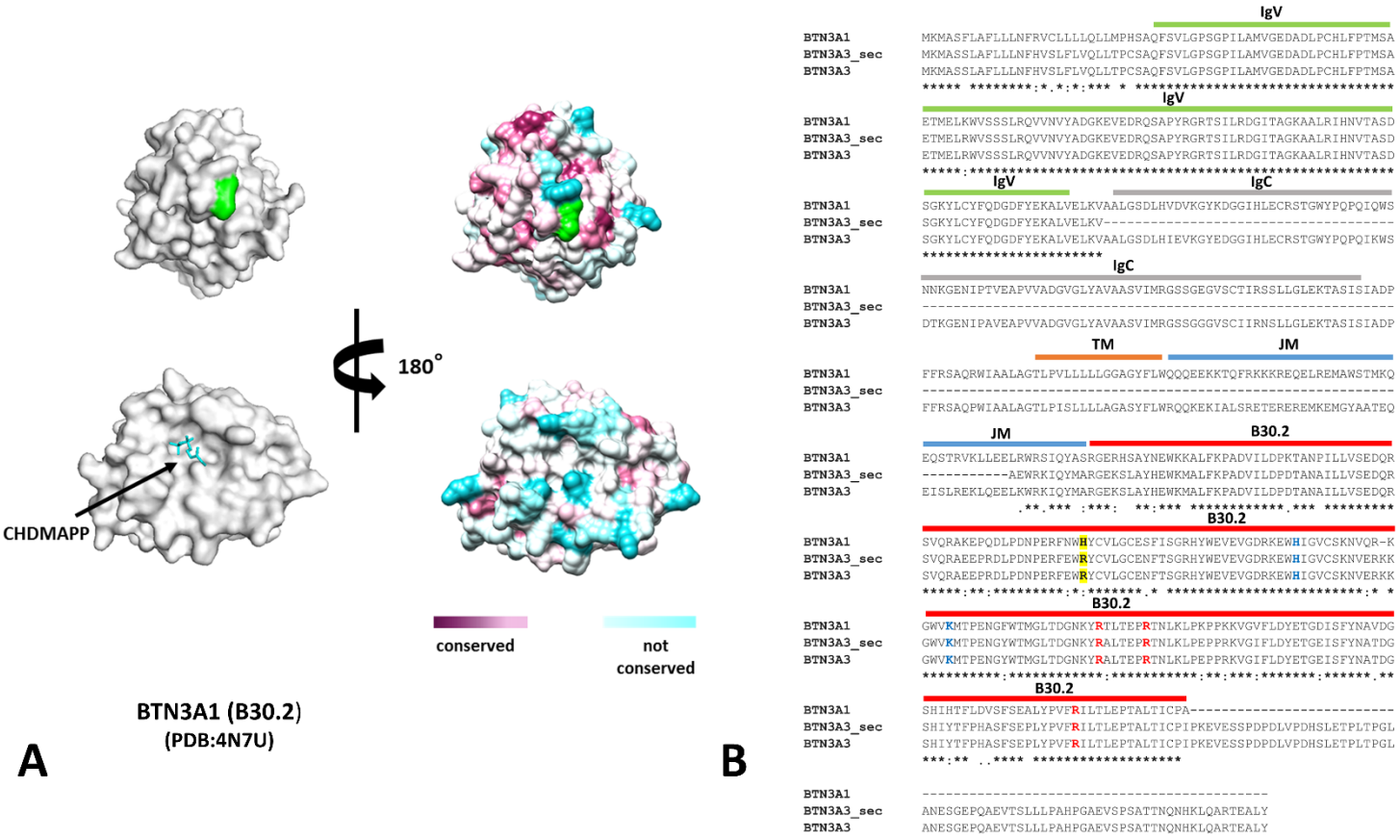


Figure 3: (A) Conservation of amino acid residues in the B30.2 domain of BTN3A1. The color coding from magenta to cyan represents a reduction in conservation with stronger colors constituting higher conservation or the lack of it. The first three residues are highlighted in green to provide a reference to the orientation of the molecule. Bound phosphoantigen (CHDMAPP) is shown to indicate the binding pocket in the domain. **(B)** Sequence alignment of BTN3A3 secretory molecule (BTN3A_sec) with BTN3A1 and BTN3A3. Different domains are accentuated by lines with distinct colors. The histidine residue conferring a "loss or gain of function" is highlighted in yellow.

References

1. Arnett, H.A., S.S. Escobar, and J.L. Viney, *Regulation of costimulation in the era of butyrophilins.* Cytokine, 2009. **46**(3): p. 370-5.
2. Abeler-Dorner, L., et al., *Butyrophilins: an emerging family of immune regulators.* Trends Immunol, 2012. **33**(1): p. 34-41.
3. Viken, M.K., et al., *Reproducible association with type 1 diabetes in the extended class I region of the major histocompatibility complex.* Genes Immun, 2009. **10**(4): p. 323-33.
4. Peedicayil, A., et al., *Risk of ovarian cancer and inherited variants in relapse-associated genes.* PLoS One, 2010. **5**(1): p. e8884.
5. Yoshida, T., et al., *Association of a polymorphism of BTN2A1 with chronic kidney disease in individuals with or without hypertension or diabetes mellitus.* Exp Ther Med, 2011. **2**(2): p. 325-331.
6. Yamada, Y., et al., *Association of a polymorphism of BTN2A1 with myocardial infarction in East Asian populations.* Atherosclerosis, 2011. **215**(1): p. 145-52.
7. Mitsunaga, S., et al., *Exome sequencing identifies novel rheumatoid arthritis-susceptible variants in the BTNL2.* J Hum Genet, 2013. **58**(4): p. 210-5.
8. Valentonyte, R., et al., *Sarcoidosis is associated with a truncating splice site mutation in BTNL2.* Nat Genet, 2005. **37**(4): p. 357-64.
9. Yoshida, T., et al., *Association of a genetic variant of BTN2A1 with chronic kidney disease in Japanese individuals.* Nephrology (Carlton), 2011. **16**(7): p. 642-8.
10. Ikemizu, S., et al., *Structure and dimerization of a soluble form of B7-1.* Immunity, 2000. **12**(1): p. 51-60.
11. Vantourout, P. and A. Hayday, *Six-of-the-best: unique contributions of gammadelta T cells to immunology.* Nat Rev Immunol, 2013. **13**(2): p. 88-100.
12. Henry, J., et al., *Cloning, structural analysis, and mapping of the B30 and B7 multigenic families to the major histocompatibility complex (MHC) and other chromosomal regions.* Immunogenetics, 1997. **46**(5): p. 383-95.
13. Yamashiro, H., et al., *Stimulation of human butyrophilin 3 molecules results in negative regulation of cellular immunity.* Journal of leukocyte biology, 2010. **88**(4): p. 757-67.
14. Palakodeti, A., et al., *The molecular basis for modulation of human Vgamma9Vdelta2 T cell responses by CD277/butyrophilin-3 (BTN3A)-specific antibodies.* The Journal of biological chemistry, 2012. **287**(39): p. 32780-90.
15. Messal, N., et al., *Differential role for CD277 as a co-regulator of the immune signal in T and NK cells.* European journal of immunology, 2011. **41**(12): p. 3443-54.
16. 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.
17. Wang, H., et al., *Butyrophilin 3A1 plays an essential role in prenyl pyrophosphate stimulation of human Vgamma2Vdelta2 T cells.* Journal of immunology, 2013. **191**(3): p. 1029-42.

18. 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.
19. Hsiao, C.H., et al., *Synthesis of a phosphoantigen prodrug that potently activates Vgamma9Vdelta2 T-lymphocytes.* Chemistry & biology, 2014. **21**(8): p. 945-54.
20. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* Journal of applied crystallography, 2003. **36**(5): p. 1277-1282.
21. Petoukhov, M.V. and D.I. Svergun, *Analysis of X-ray and neutron scattering from biomacromolecular solutions.* Current opinion in structural biology, 2007. **17**(5): p. 562-71.
22. Svergun, D.I., *Determination of the regularization parameter in indirect-transform methods using perceptual criteria. .* Journal of Applied Crystallography, 1992. **25**: p. 495-503.
23. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.* Biophysical journal, 1999. **76**(6): p. 2879-86.
24. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering.* Journal of applied crystallography, 2003. **36**(3 Part 1): p. 860-864.
25. Wriggers, W., *Conventions and workflows for using Situs.* Acta crystallographica. Section D, Biological crystallography, 2012. **68**(Pt 4): p. 344-51.
26. 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.
27. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.

Appendix

Chapter A3

Investigation of the interaction between orthologs of TamB and BamA along with the identification of an accessory protein of BamA from *T. pallidum* and *denticola* species

Adapted from:

Comparative characterization of the accessory protein of BamA and its interaction with TamB from two Treponemal species.
Kumar, S, Puthenveetil R, Dey A, Vinogradova O, Radolf JD *(to be written)*

Introduction

Gram-negative bacterial diderms, esp. Proteobacteria, typically contain an outer membrane (OM) connected to lipo-polysaccharides (LPS) on its outer leaflet and an inner membrane (IM) separated by a periplasmic space. The OM with an asymmetrical lipid composition houses several porin forming β barrel proteins, with anti-parallel β strands [1]. An outer membrane protein (OMP) is initially synthesized in the cytoplasm, reaching the OM traversing the IM through the SecYEG translocon [2]. OMP precursors in the cytoplasm are brought to the translocase through one of the following two mechanisms: (i) through the SecAB chaperones which binds to the unfolded OMPs [3], or (ii) co-translationally from the ribosome [4]. Nascent OMPs enter the intervening periplasmic space where they bind to chaperones, SurA [5] or SKP [6], preventing misfolding and aggregation in the periplasm. Chaperone bound OMP is delivered to the β-barrel assembly machinery (BAM) complex, primarily consisting of BamA in the OM and two or more accessory proteins in the periplasm anchored to the inner leaflet of the outer membrane [7]. The transfer might occur through a direct contact demonstrated for SurA *in vivo* [8]; and as has been speculated for SKP, which can bind through its positively charged head region to the negatively charged patches on the periplasmic domains of BamA [9]. BamA is finally responsible for inserting OMP into the outer membrane through lateral opening of its β-barrel domain [10]. A pictorial representation of OMP biogenesis is depicted in Figure 1A (top panel).

BAM machinery is a complex composite of several proteins, named BamA through BamE, present in the periplasmic space of *E.coli*. The primary component of the BAM machinery is BamA, an OMP with 16 stranded β-barrel with five polypeptide transport-

associated (POTRA) domains. POTRA domains help thread nascent OMP into the β-barrel subunit of BamA through a process called β augmentation [11]. The accessory proteins (BamB-E) are anchored to the inner leaflet of the outer membrane, through the post-translational lipid modification of their N-terminus region [12]. BamD and BamB interact directly with BamA, while BamC and E interact indirectly to BamA through BamD [13]. The accessory protein composition varies with bacterial species, for example *N. meningitidis* lacks BamB [14]. BamB has an eight bladed beta propeller fold and interacts with the POTRA 2/3 domains of BamA. It functions to enhance the incorporation of nascent OMPs into the outer membrane [11]. Crystal structure of the intact *E.coli* BAM complex with all accessory components has been solved (Figure 1A, bottom panel) shedding light on their typical domain architecture and function [15]. BamA belongs to an archetypal protein transport system classified through the presence of an outer membrane domain typically followed by POTRA domains, collectively classified under the Omp85 superfamily of proteins [16]. Proteins belonging to this family are involved in protein incorporation into the membrane (eg. BamA/TamA) or translocation across the membrane (eg. FhaC). Eukaryotic members of the Omp85 superfamily, found in mitochondria and plastids [17], are descendants from the bacterial BamA protein. The other well characterized member of the Omp85 family is TamA (Translocation and assembly module), consisting of a C-terminal outer membrane domain and three N-terminal POTRA domains (Figure1B, bottom panel) [18, 19]. Interestingly, TamA binds to TamB, a protein containing a conserved DUF490 domain at its C-terminus and is anchored to the inner membrane through its N-terminus, together spanning the entire periplasmic region (Figure 1B, top panel) [20]. It has been

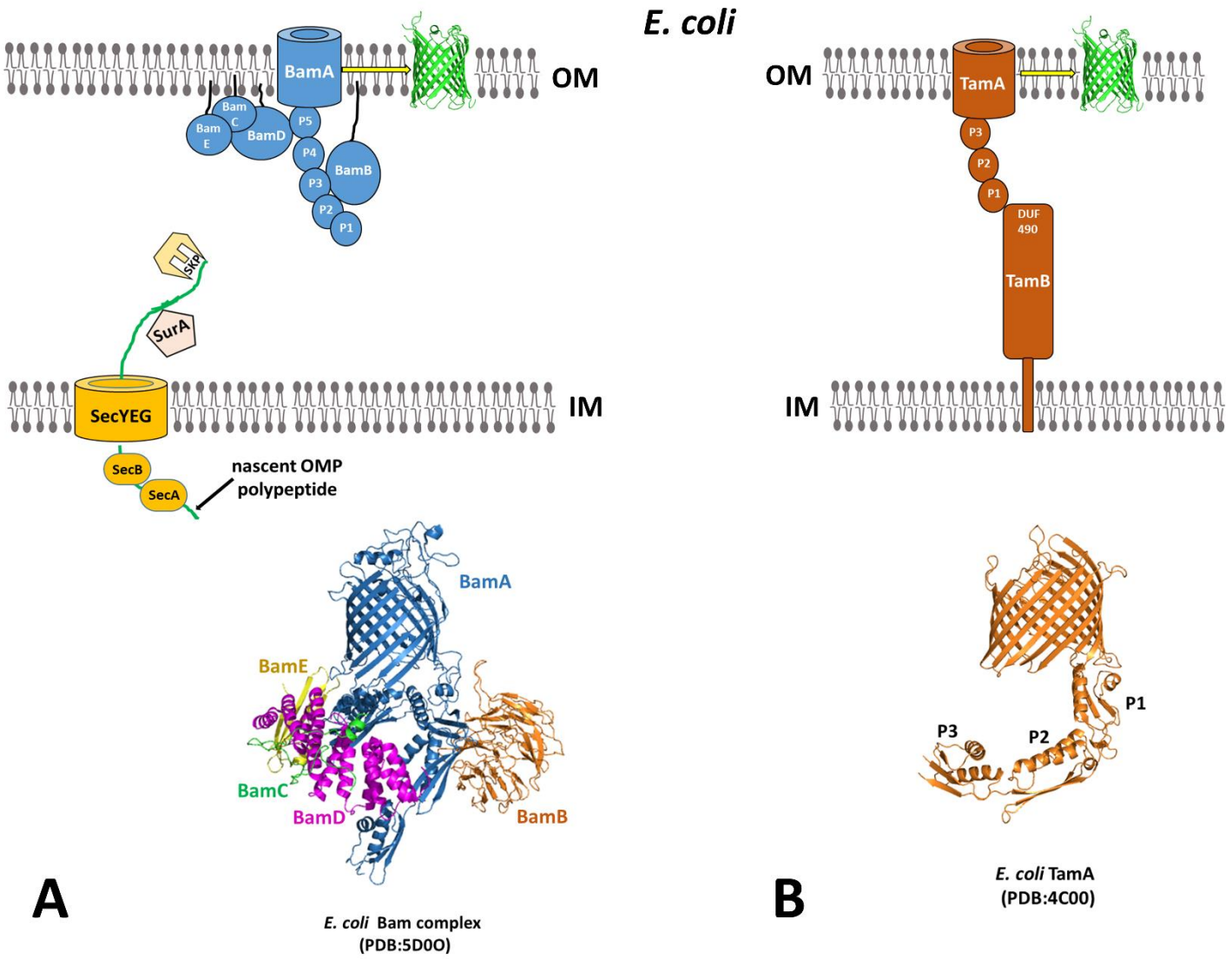


Figure 1: (A) Top: Model exemplifying the Outer-membrane Protein (OMP) biogenesis in *E.coli*: Briefly, the nascent unfolded OMP polypeptide is translocated through the SecYEG pathway into the periplasm where they are transferred to chaperones SurA or SKP and threaded into the outer membrane through the five POTRA domains (P1-P5) of BamA. Proteins BamBCDE are accessory proteins that expedite the process of membrane insertion. **Bottom:** Crystal structure of the Bam complex (BamABCDE). **(B) Top:** Schematic diagram showing, the more recently discovered, TamA-TamB complex in *E.coli*. TamA similar to BamA is involved in the insertion of proteins into the outer membrane. TamB, containing its signature C-terminal DUF490 domain, is anchored to the inner membrane and interacts with TamA at its POTRA P1 domain. **Bottom:** Crystal structure of the TamA complex showing a similar organization of its three POTRA domains as BamA.

shown that the TamA-TamB system is required for the specific incorporation of auto-transporters in the OM [19]. The genetic locus of TamB is contiguous to proteins from the Omp85 family, in proteobacteria it is found in an operon with TamA while in organisms from early branching phyla it's found close to BamA [21]. The following chapter discusses the identification of BamB orthologs and the interaction of TamB with BamA from two treponemal species *T. pallidum* and *T. denticola*.

Result summary

The genomes for *T. pallidum* (strain Nichols) and *T. denticola* (strain ATCC 35405) were scouted for protein candidates with the signature PQQ domain found in BamB (3Q7N) [22] and DUF490 found in TamB [19] using Pfam/CDD/InterProScan. The search resulted in singular candidates for the TamB ortholog (TP_0325 and TDE_2600) and close to five candidates for the BamB ortholog. The TamB candidates were further accessed for the position of their genetic loci, and were found contiguous to BamA (Figure 2A), as is expected for early branching phyla lacking the gene for TamA [21]. BamB candidates were further accessed for their signal sequences using LipoP. This helped us narrow down to a single candidate (TDE_1930 and TP_0783) with an N-terminal signal sequence that would enable the protein's transport across the inner membrane. In addition, the presence of a transmembrane region was predicted at its N-terminus [23] which further supports the anchoring of BamB to the inner leaflet of the OM, as has been observed for all BamB orthologs (Figure 2B). Both *T. pallidum* and *denticola* BamB candidates were modelled using different homology modelling software (SWISS-MODEL,

PHYRE2 and I-TASSER). The template with the highest confidence, identified by the modelling software belonged to BamB orthologs. The final model showed the signature β-propeller architecture found in BamB (Figure 2C). In order to validate the identified candidates we further carried out co-immunoprecipitaiton (Co-IP) experiments from *T. pallidum* and *T. denticola microorganisms. T. pallidum* (Nichols) was propagated through intra-testicular inoculation of white rabbits following the protocol in [24]. *T. denticola* (ATCC 35405) was grown to the mid to late log phase in new oral spirochete (NOS) broth [25] supplemented with 10% heat-inactivated normal

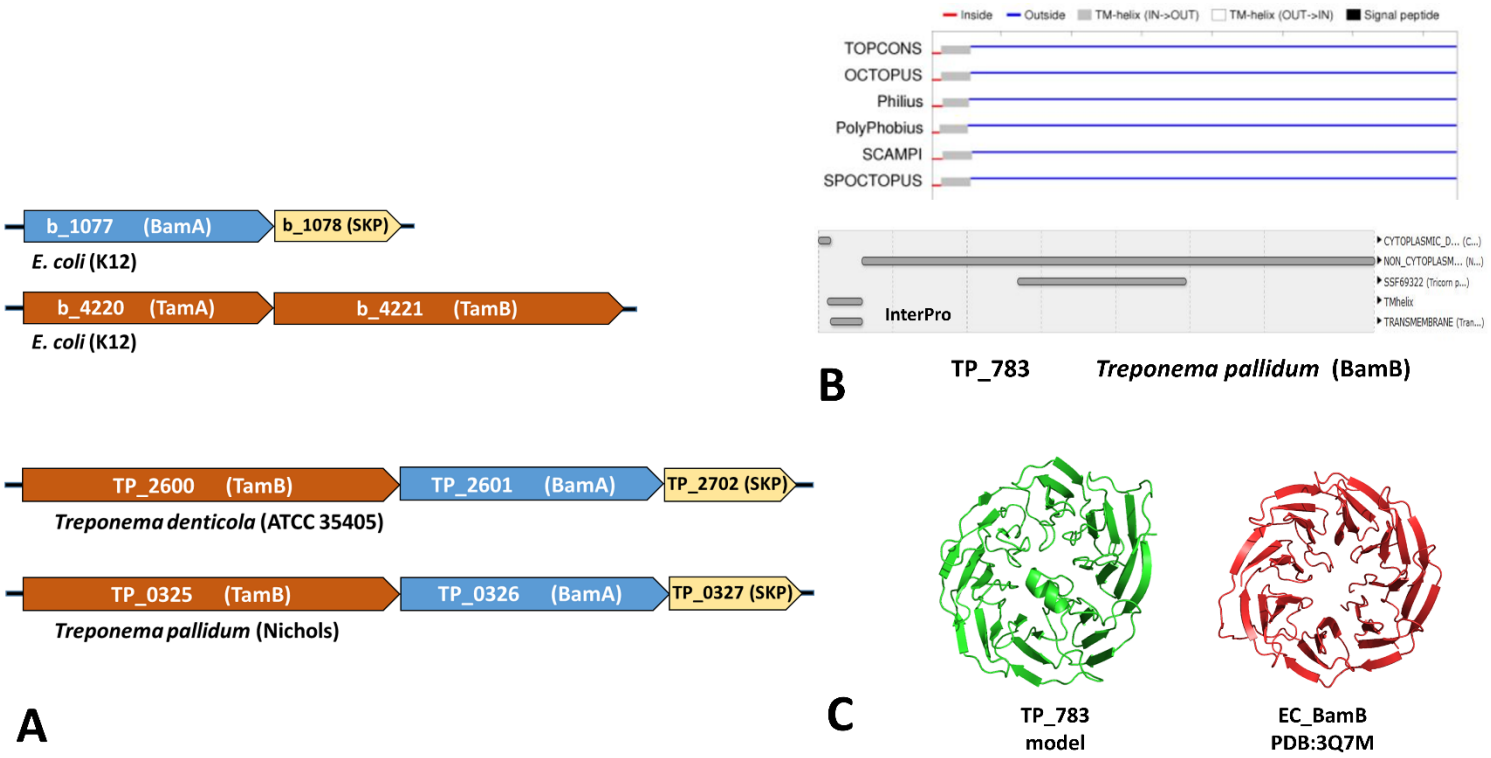


Figure 2: (A) Genetic loci showing the difference in the arrangement of BamA, TamA and TamB in *E.coli* and Treponeamal species. **(B)** Bioinformatic analysis identifying the presence of a transmembrane region at the N-terminal end of the putative BamB ortholog (TP_783) in *Treponema pallidum*. **(C)** Homology model revealing the β-propeller structure typically found in BamB proteins.

rabbit serum, using the GasPak Plus anaerobic system (Becton, Dickinson, Cockeysville, MD). The propagation of treponemal species typically achieved a cell density of $2x10^9$ cells per ml. For Co-IP experiments harvested cells were resuspended in PBS buffer (Bio-Rad, USA) supplemented with protease cocktail (Sigma, USA) and lysed using mild pulse sonication for 10 seconds in two cycles. Insoluble cell debris was removed by centrifugation at 6000 xg and the supernatant was mixed at 4°C with antibodies raised against BamB and TamB orthologs in rats. Affinity purification was performed using Protein-G magnetic beads (EMD Millipore, USA) as described in the manufacturer's protocol. The eluate was separated through SDS-PAGE and blotted with antibodies specific to the POTRA domain of BamA (raised in rabbits/rats) either using the regular protocol for Rat antibodies or the clean-blot IP detection kit (PIERCE, USA) for Rabbit antibodies. The pull down eluates were positive for the presence of BamA, as seen by the band close to the 100kDa molecular weight marker (Figure 3A). Analogous experiment for the BamB ortholog from *T.denticola* is underway along with the production of rabbit antibodies against the POTRA domain of BamA from T. *denticola*. In addition, native page analysis has also confirmed the presence of the proteins with BamA since they co-migrate in the gel. Further experiments demonstrating inter-domain interaction *in vitro* through the use of recombinant proteins are planned. This will assist in corroborating the identification of the two orthologs through their interaction with the POTRA domain of BamA.

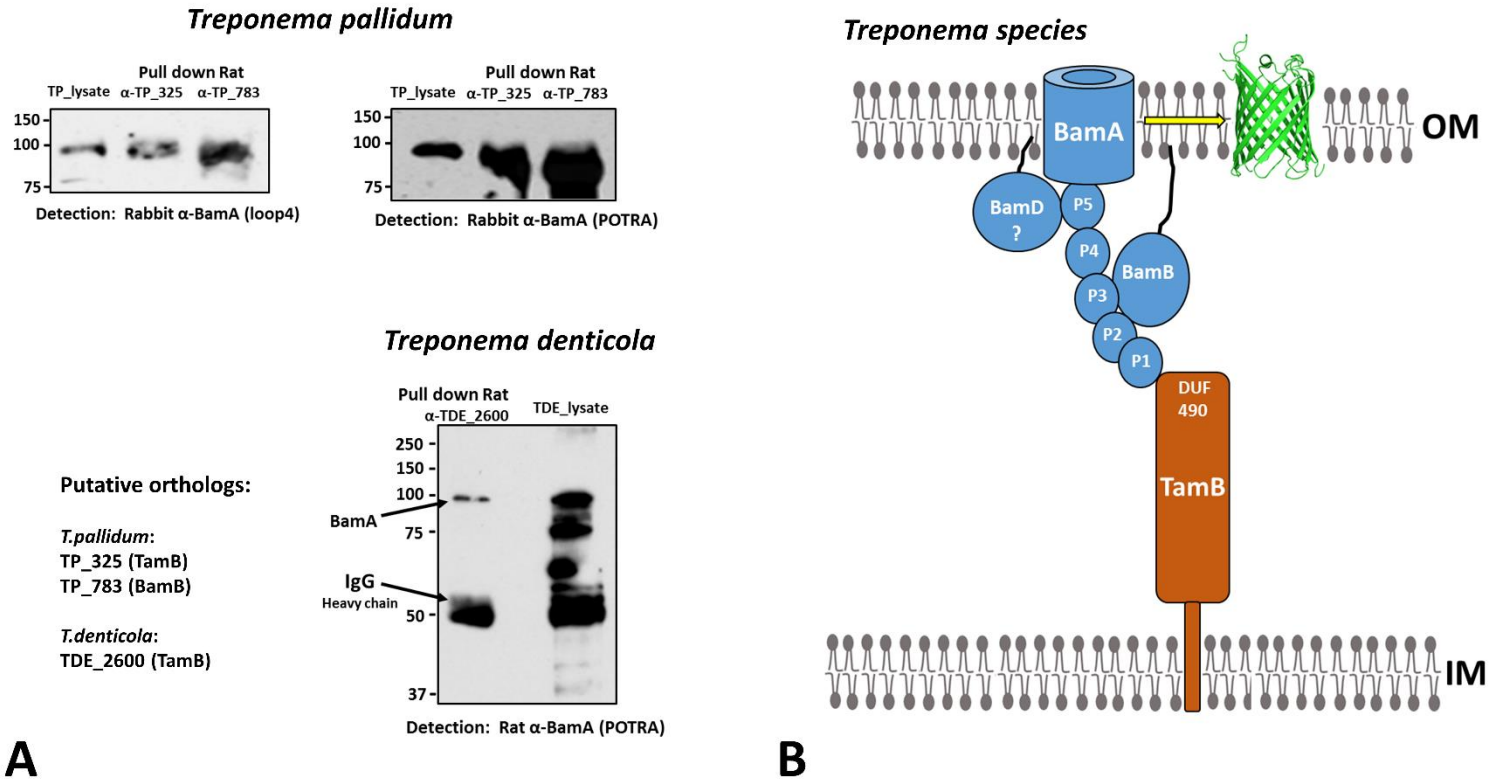


Figure 3: (A) Co-Immuno preceipitation (Co-IP) using rat antibodies for the putative TamB orthologs (TP_325 and TDE_2600) and BamB ortholog (TP_783) from native T. *pallidum* and *denticola* cultures. Blots are developed with POTRA [rat (TDE) and rabbit (TP)] antibodies additionally confirmed with rabbit BamA Loop4 antibodies for T *pallidum*. **(B)** Schematic diagram demonstrating the unusual interaction of BamA with TamB which is in contrast to the TamA-TamB association found in E. coli (Figure 1).

Conclusion

We have successfully identified BamB and TamB orthologs in the two treponemal species *T. pallidum* (strain: Nichols) and *T. denticola* (strain: ATCC 35405). This was achieved through a series of computational, *in vivo* and *in vitro* (planned) experimentation. Bio-informatics analysis helped us with the initial identification of BamB and TamB orthologs through (i) the conserved domain analysis; (ii) prediction of the sorting signal

sequences and (iii) homology modelling displaying the typical domain architectures observed for the proteins. *In vivo* experimentation of these potential candidates through Co-IP and BN-PAGE further confirmed their association with BamA. Novel to this finding is the unusual interaction of the BamA and TamB proteins (Figure 3B). Recently, the BamA-TamB interaction has been reported for Borrelia burgdoferi in a paper [26] by Iqbal, H et al. Future *in vitro* experimentation using recombinant BamB,

TamB (DUF490) and the BamA (POTRA) domains through ITC and co-elution using SEC will help corroborate the interaction of these proteins. Additionally, experiments are planned to identify the specific POTRA domain binding to BamB and TamB through the utilization of different assortments of the individual POTRA domains (eg. POTRA 1-2, 1, 2-3 for TamB and POTRA 1-2, 2-3, 4-5 for BamB). This is based on the rationale from previous findings where TamB has been reported to bind the POTRA-1 domain of TamA [20] while BamB has been shown to bind POTRA-2/3 domains of BamA [15].

References

1. Gan, L., S. Chen, and G.J. Jensen, *Molecular organization of Gram-negative peptidoglycan.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(48): p. 18953-7.
2. Van den Berg, B., et al., *X-ray structure of a protein-conducting channel.* Nature, 2004. **427**(6969): p. 36-44.
3. Economou, A. and W. Wickner, *SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion.* Cell, 1994. **78**(5): p. 835-43.
4. Matlack, K.E., W. Mothes, and T.A. Rapoport, *Protein translocation: tunnel vision.* Cell, 1998. **92**(3): p. 381-90.
5. Behrens, S., et al., *The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity.* The EMBO journal, 2001. **20**(1-2): p. 285-94.

6. Qu, J., et al., *The trimeric periplasmic chaperone Skp of Escherichia coli forms 1:1 complexes with outer membrane proteins via hydrophobic and electrostatic interactions.* Journal of molecular biology, 2007. **374**(1): p. 91-105.
7. Hagan, C.L., T.J. Silhavy, and D. Kahne, *beta-Barrel membrane protein assembly by the Bam complex.* Annual review of biochemistry, 2011. **80**: p. 189-210.
8. Sklar, J.G., et al., *Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli.* Genes & development, 2007. **21**(19): p. 2473-84.
9. Gatzeva-Topalova, P.Z., T.A. Walton, and M.C. Sousa, *Crystal structure of YaeT: conformational flexibility and substrate recognition.* Structure, 2008. **16**(12): p. 1873-81.
10. Noinaj, N., et al., *Structural insight into the biogenesis of beta-barrel membrane proteins.* Nature, 2013. **501**(7467): p. 385-90.
11. Heuck, A., A. Schleiffer, and T. Clausen, *Augmenting beta-augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins.* Journal of molecular biology, 2011. **406**(5): p. 659-66.
12. Noinaj, N., S.E. Rollauer, and S.K. Buchanan, *The beta-barrel membrane protein insertase machinery from Gram-negative bacteria.* Current opinion in structural biology, 2015. **31**: p. 35-42.
13. Kim, S., et al., *Structure and function of an essential component of the outer membrane protein assembly machine.* Science, 2007. **317**(5840): p. 961-4.
14. Volokhina, E.B., et al., *The beta-barrel outer membrane protein assembly complex of Neisseria meningitidis.* Journal of bacteriology, 2009. **191**(22): p. 7074-85.
15. Gu, Y., et al., *Structural basis of outer membrane protein insertion by the BAM complex.* Nature, 2016. **531**(7592): p. 64-9.
16. Gentle, I.E., L. Burri, and T. Lithgow, *Molecular architecture and function of the Omp85 family of proteins.* Molecular microbiology, 2005. **58**(5): p. 1216-25.
17. Gentle, I., et al., *The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria.* The Journal of cell biology, 2004. **164**(1): p. 19-24.
18. Gruss, F., et al., *The structural basis of autotransporter translocation by TamA.* Nature structural & molecular biology, 2013. **20**(11): p. 1318-20.
19. Selkrig, J., et al., *Discovery of an archetypal protein transport system in bacterial outer membranes.* Nature structural & molecular biology, 2012. **19**(5): p. 506-10, S1.
20. Selkrig, J., et al., *Conserved features in TamA enable interaction with TamB to drive the activity of the translocation and assembly module.* Scientific reports, 2015. **5**: p. 12905.
21. Heinz, E., et al., *Evolution of the Translocation and Assembly Module (TAM).* Genome biology and evolution, 2015. **7**(6): p. 1628-43.
22. Noinaj, N., J.W. Fairman, and S.K. Buchanan, *The crystal structure of BamB suggests interactions with BamA and its role within the BAM complex.* Journal of molecular biology, 2011. **407**(2): p. 248-60.
23. Tsirigos, K.D., et al., *The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides.* Nucleic acids research, 2015. **43**(W1): p. W401-7.
24. Hazlett, K.R., et al., *TP0453, a concealed outer membrane protein of Treponema pallidum, enhances membrane permeability.* Journal of bacteriology, 2005. **187**(18): p. 6499-508.

25. Haapasalo, M., et al., *Characterization, cloning, and binding properties of the major 53-kilodalton Treponema denticola surface antigen.* Infection and immunity, 1992. **60**(5): p. 2058-65.
26. Iqbal, H., et al., *The TamB ortholog of Borrelia burgdorferi interacts with the beta-barrel assembly machine (BAM) complex protein BamA.* Molecular microbiology, 2016.