

Cloning, Expression, and Biophysical Investigations of Truncated Inclusion Protein A (IncA)  
from *Chlamydia trachomatis*

Catrina Alicia Campbell  
Longmont, Colorado

Bachelor of Science, University of Texas at Arlington, 2012  
Master of Arts, University of Alabama, 1992  
Bachelor of Arts, University of Colorado, 1989

A thesis presented to the Graduate Faculty of the University of Virginia in Candidacy for the  
Degree of Master of Science

Department of Chemistry

University of Virginia  
August 2014

Approved by:

---

Dr. Linda Columbus

---

Dr. Cameron Mura

---

Dr. Jill Venton

---

Dr. James Demas

*To my mom and dad, who gave me a thirst for knowledge and the desire to learn.*

## Acknowledgements

I would like to thank Dr. Linda Columbus for allowing me to work in her lab, study inclusion proteins, and pushing my limits.

Thank you Columbus and Mura Lab members for all their technical assistance and feedback.

Dr. John D. Shannon for his incredible patience and assistance with MALDI-TOF MS and circular dichroism.

Many thanks to Jessica Sarver in the Cafiso lab for allowing me to think out loud about EPR.

Thank you to Sandy, Vicki, and Kevin in Facilities Management who kept me laughing, updated me on the news and “goings on” in the department in the “wee” hours of the morning.

Special thanks to Tsega Solomon who gave me invaluable technical advice, interesting philosophical debates over lunch, and a renewed interest in shopping. How I will miss you!

Special thanks to Dr. Kim Bassett for her listening ear and unconditional friendship.

Special thanks to Heather and Addison Ely for providing me shelter in the storm.

Thanks to my family for their support, ability to just listen, and giving the best pep talks!!

## Abstract

*Chlamydia (C.) trachomatis* is an obligate bacterial pathogen of eukaryotic cells that, despite over a century of scientific inquiry, has avoided eradication.<sup>1, 2</sup> *C. trachomatis* replicates within an intracellular vacuole called an inclusion body. Inclusion membrane protein A (IncA), an effector protein expressed by *C. trachomatis*, is secreted by a type III secretory system into the host-derived membrane of the inclusion. IncA is proposed to mediate homotypic fusion of inclusion bodies. Inclusion proteins are characterized by a bilobed trans-membrane segment with a C-terminal domain that extends into the host cytoplasm. Sequence analysis of IncA reveals a “soluble N-ethylmaleimide-sensitive attachment protein receptors” (SNAREs) motif in the C-terminal region which is found in fusion-facilitating proteins.<sup>3</sup> A signature of the SNARE motif is the leucine zipper which may initiate the formation of homo-dimers or hetero-dimers. In this study, a truncated construct with an N-terminal histidine tag was generated which yielded the SNARE soluble domain without the transmembrane region. Size exclusion chromatography (SEC) demonstrated the hexa-histidine tag disrupted dimer formation; however, after cleavage of the tag with thrombin, dimer was reformed. Circular dichroism of the construct confirmed the dimer was all  $\alpha$ -helical as expected for a SNARE motif. Finally, two single cysteine mutants were prepared and spin-labeled. Continuous wave electron paramagnetic resonance confirmed the sites were spin-labeled and the lineshapes were consistent with moderately immobilized spin labels expected for a tertiary contact. Future double electron-electron resonance experiments will measure the distance between the spin labels, which will elucidate the dimer orientation.

## Table of Contents

Acknowledgements .....	iii
Abstract .....	iv
Table of Contents .....	v
Chapter 1 (Introduction) .....	vi
List of Figures .....	vii
<b>Chapter 1 (Introduction).....</b>	<b>1</b>
1.1 Overview .....	1
1.2 The <i>Chlamydial</i> Genus.....	2
1.2.1 The Biphasic Lifecycle.....	3
1.3 Inclusion Proteins .....	5
1.4 Inclusion Protein A .....	6
1.5 IncA from <i>C. trachomatis</i> shows Similarity to SNARE Proteins .....	7
1.6 Research Purpose .....	9
<b>Chapter 2 (Theoretical Concepts) .....</b>	<b>11</b>
2.1 Cloning Overview .....	11
2.2 PIPE Cloning and Mutagenesis .....	11
2.3 Expression and Purification .....	14
2.3.1 <i>Escherichia coli</i> expression systems .....	14
2.3.2 The pET expression system .....	15
2.3.3 Immobilized Metal Affinity Chromatography (IMAC) .....	17
2.3.4 Dialysis and Thrombin Cleavage .....	18
2.3.5 Size Exclusion Chromatography (SEC) .....	19
2.4 Matrix Assisted Laser Desorption Ionization-Time of Flight Mass Spectrometry (MALDI-TOF MS) .....	21
2.4.1 Overview .....	21
2.4.2 Hard versus Soft Ionization.....	22
2.4.3 Mechanism of Action .....	22
2.5 Circular Dichroism (CD).....	23
2.6 Electron Paramagnetic Resonance (EPR) .....	26
2.6.1 EPR Overview .....	26
2.6.2 Continuous Wave EPR.....	27
2.6.3 Site-Directed Spin Labels.....	30

<b>Chapter 3 (Materials and Methods)</b>	<b>32</b>
3.1 Cloning and Expression	32
3.2 Purification	34
3.3 Dialysis and Thrombin Cleavage	35
3.4 SEC	35
3.5 MALDI-TOF MS	35
3.6 CD	36
3.7 Site-Directed Spin Labeling of IncA Mutants for EPR Structural Studies	36
3.8 CW-EPR	37
<b>Chapter 4 (Results and Discussion)</b>	<b>38</b>
4.1 Cloning	38
4.2 Expression and Purification	38
4.3 Size Exclusion Chromatography	39
4.4 CD	43
4.5 MALDI-TOF MS	44
4.6 Site-Directed Spin Labeling of $\Delta 85$ IncA141C and $\Delta 85$ IncA252C and EPR	45
4.6.1 Conclusions	48
4.7 Future Directions	49
<b>References</b>	<b>53</b>

## List of Figures

Figure 1.1	Biphasic lifecycle of <i>C. trachomatis</i>	<b>3</b>
Figure 1.2	Type III secretory system illustration	<b>4</b>
Figure 1.3	Proposed structure of the IncA monomer	<b>6</b>
Figure 1.4	Illustration of the heptameric leucine zipper motif	<b>7</b>
Figure 1.5	<i>C. trachomatis</i> IncA primary sequence	<b>9</b>
Figure 2.1	Illustration of PIPE cloning methodology	<b>13</b>
Figure 2.2	Rendering of IPTG function in pET expression systems	<b>16</b>
Figure 2.3	Depiction of size exclusion chromatography	<b>20</b>
Figure 2.4	MALDI-TOF MS sample ionization and source region	<b>24</b>
Figure 2.5	Characterization of secondary structure in circular dichroism	<b>25</b>
Figure 2.6	Two states of electron alignment in a magnetic field	<b>28</b>
Figure 2.7	Hyperfine splitting	<b>29</b>
Figure 2.8	Free spin label in an applied magnetic field	<b>30</b>
Figure 2.9	Methanethiosulfonate reaction	<b>30</b>
Figure 2.10	Illustration of structural information generated from EPR spectra	<b>31</b>
Figure 3.1	Depiction of PIPE cloning & PIPE mutagenesis constructs	<b>32</b>
Figure 4.1	Purification results of $\Delta 85$ IncA	<b>39</b>
Figure 4.2	Initial SEC elution profile of $\Delta 85$ IncA	<b>40</b>
Figure 4.3	SEC chromatogram of $\Delta 85$ IncA with standards	<b>41</b>
Figure 4.4	SEC chromatogram of thrombin cleaved $\Delta 85$ IncA with standards	<b>42</b>
Figure 4.5	CD of $\Delta 85$ IncA	<b>43</b>
Figure 4.6	CD of $\Delta 85$ IncA with $\Delta 85$ IncA mutants	<b>44</b>
Figure 4.7	MALDI-TOF MS of $\Delta 85$ IncA141R1 and $\Delta 85$ IncA252R1	<b>46</b>
Figure 4.8	X-Band (9GHz) EPR spectra of $\Delta 85$ IncA141R1	<b>47</b>
Figure 4.9	X-Band (9GHz) EPR spectra of $\Delta 85$ IncA252R1	<b>47</b>
Figure 4.10	Literature proposed orientations of the IncA dimer.	<b>49</b>
Figure 4.11	Possible DEER experiment with $\Delta 85$ IncA141R1.	<b>50</b>
Figure 4.12	Possible DEER experiment with $\Delta 85$ IncA252R1.	<b>51</b>





## Chapter 1

### Introduction

#### 1.1 Overview

In 1907, Czech parasitologist Stanislaus von Prowazek discovered the bacterium *Chlamydia trachomatis*.<sup>4</sup> He misnamed it *chlamydozoa*, mistakenly believing it to be a virus.<sup>5</sup> Later, the bacteria would be classified in the genus *clamys*, which, in ancient Greece, was a cloak or costume worn by young men.<sup>6</sup> *Chlamydia* is an appropriate name due to the pathogen's clandestine methods entering host cells, preventing attack by the host's immune system and subsequent degradation, and hiding such that they can flourish without scrutiny from lysosomes. *C. trachomatis* causes disease stealthily. Over a century after Prowazek gave the bacteria its name, *C. trachomatis* endures as a significant agent of disease and has yet to be eradicated.<sup>7, 1, 8</sup> *C. trachomatis* is the most common sexually transmitted infection in the world as well as in the United States, and is the number one preventable cause of blindness.<sup>1, 9, 10, 2, 11, 12</sup> In women, this bacterium can cause pelvic inflammatory disease, ectopic pregnancy, tubal infertility and chronic pelvic pain.<sup>7, 10</sup> In men, *C. trachomatis* can cause epididymitis as well as severe tissue scarring that can lead to infertility.<sup>7, 13</sup> Despite disease control strategies in developed nations starting in the late 1980's, the incidence of *C. trachomatis* has increased since the mid 1990's and remains undeterred.<sup>2, 10</sup> Interest in *C. trachomatis* extends beyond *Homo sapiens* to the veterinary world, infecting ruminants, birds, toads, and the marsupial koalas.<sup>14</sup>

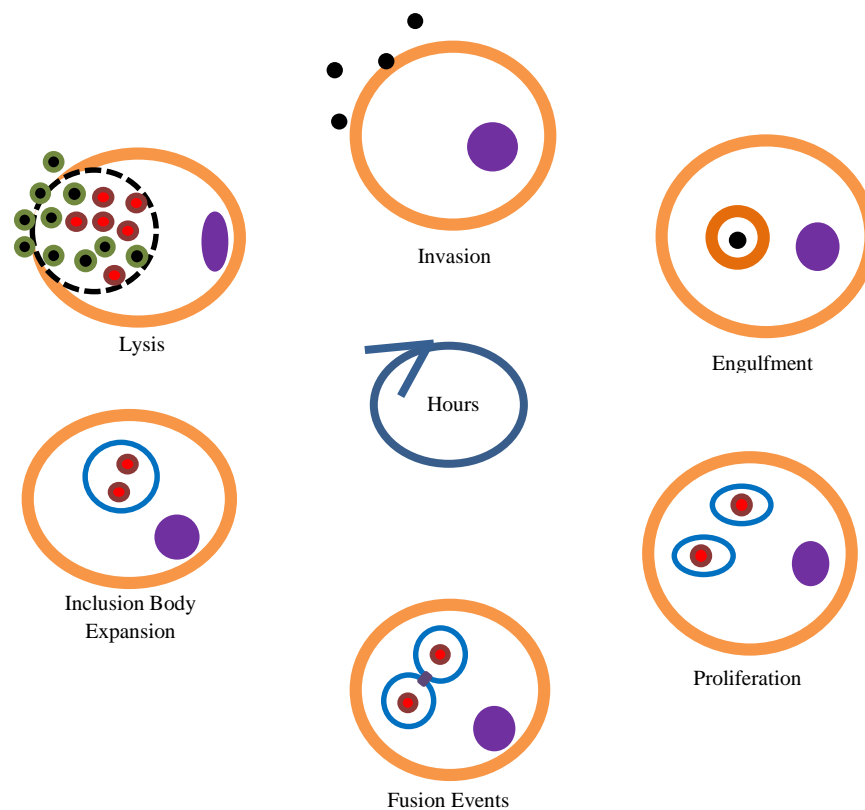
## 1.2 The Genus *Chlamydiae*

Controversy surrounds *chlamydial* taxonomy. Many bacteriologists disagree with the current classification of nine chlamydial pathogens and suggest these species share more commonalities than differences. Currently, three species are included in the genus *Chlamydiae* (*trachomatis*, *suis*, and *muridarum*), and six species are included in *Chlamydiophila* (*pneumonia*, *pecorum*, *psittaci*, *abortus*, *felis*, and *caviae*). According to some bacteriologists, species in *Chlamydiae* diverged evolutionarily from *Chlamydiophila* based on 16S ribosomal RNA sequence similarities.<sup>15</sup> *Chlamydiophila* species share >90% 16S rRNA sequence homology while species in *Chlamydiae* share between 80 and 90% sequence similarity. For purposes of this study, all nine species of *Chlamydiae* will be classified below. Each species has different properties, causes a wide range of diseases, and presents differently in host tropism. The commonalities they share, however, are significant and unique to the genus.<sup>16</sup>

*Chlamydiae* have 560 conserved genes and share approximately 90% homology in their 16s ribonucleic acid or small unit (RNA).<sup>14</sup> Although the genome sequence is available for several species, these pathogens lack a genetic transfer system, so functional studies require the use of indirect methods.<sup>9, 17, 16, 18, 19</sup> *Chlamydial* species are gram negative and sensitive to antibiotics such as penicillin and Brefeldin A.<sup>4, 14, 16</sup> Specific antigenic major outer membrane proteins (MOMPs), cross-linked by disulfide bridges, give the bacteria their shape<sup>4</sup>; however, the key feature of all *Chlamydiae* is their biphasic life cycle development.<sup>4, 5, 7, 11, 12, 14, 16, 18, 20, 21</sup>

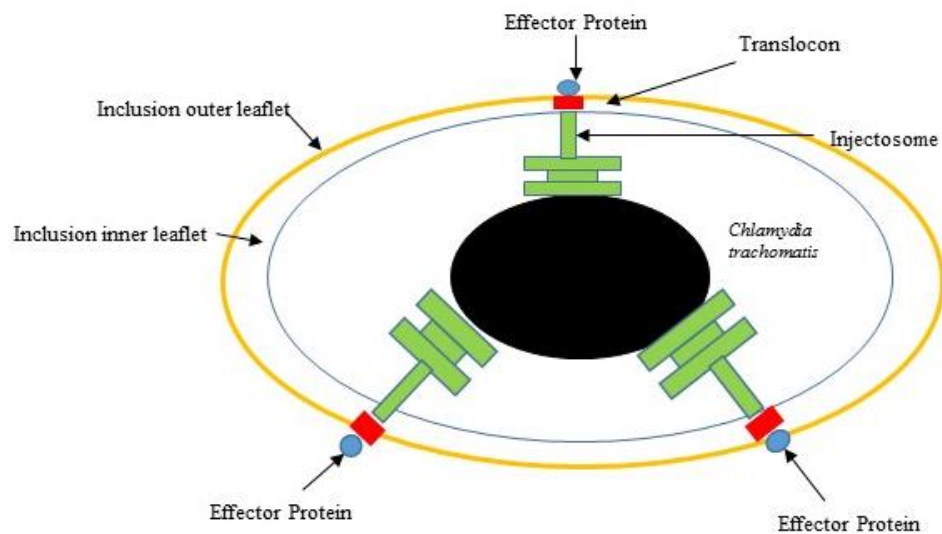
### 1.2.1 Biphasic Development Cycle

When *Chlamydial* species exist outside the host, they are known as elementary bodies (EBs) (**Figure 1.1**). EBs are metabolically inert, have a dense nucleus, can withstand harsh conditions, and are relatively small (about 0.3 to 0.5  $\mu\text{m}$ ).<sup>4, 16, 22</sup> Once EBs gain entry into host epithelial cells via protein transporters, they cloak themselves in the host's plasma membrane (**Figure 1.2**).



**Figure 1.1.** Biphasic lifecycle of *Chlamydia trachomatis*. **Elementary bodies** invade **epithelial cells**. The pathogen is engulfed, the plasma membrane becomes modified, and the *C. trachomatis* differentiates into the **reticulate body**. **Inclusion bodies** fuse when the host is multiply infected. The **inclusion body** expands to accommodate the progeny. Reticulate bodies return to the elementary state and the host cell is lysed when approximately 1000 progeny have been produced.

Protein synthesis begins between 30 minutes and one hour after invasion. A type III secretory system (T3SS) secretes effector proteins, thereby modifying the host plasma membrane.<sup>20, 23</sup> Besides organisms in the non-proteobacteria phylum, *Chlamydiae* are the only bacteria to possess a T3SS.<sup>12, 14</sup> The plasma membrane of the host becomes a non-acidified vacuole or an inclusion body that envelopes the bacteria.<sup>2, 4, 5, 7, 9-14, 16, 19-21, 23</sup>



**Figure 1.2** Type III secretory system illustration. Molecular syringes called injectosomes secrete effector proteins to the outer the inclusion membrane.

The inclusion body has two functions: (1) to protect *Chlamydiae* during replication and (2) to serve as a gate while the pathogen interacts with its host.<sup>16, 24</sup> Using the inclusion, the pathogen successfully eludes endosomes that target foreign matter for degradation by separating from the endocytic pathway early in its development. *Chlamydial* species are true obligate pathogens because they are unable to synthesize adenosine triphosphate (ATP). Described as “energy pathogens,” *Chlamydiae* import ATP into the inclusion body

as it simultaneously engages with endoplasmic reticulum and the Golgi apparatus to procure its nutritional needs and lipids.<sup>4, 12, 22, 23, 25, 26</sup> The EB, now safely tucked away in the inclusion body, reduces the cross-links between MOMP, doubles its size, and begins replicating by binary fission.<sup>4, 16</sup> The mature, metabolically active EB is now called a reticulate body (RB). RBs multiply exponentially, and the inclusion body expands to accommodate up to 1,000 RB progeny.<sup>27</sup> RBs eventually differentiate back to the EB state between 48 and 72 hours post infection and after the host cell is lysed.<sup>9, 11, 12, 25</sup>

### 1.3 Inclusion Proteins

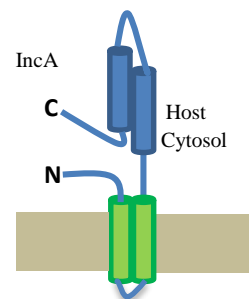
Inclusion (Inc) proteins are unique to *Chlamydiae*, yet inclusion protein sequences are not highly conserved throughout the genus.<sup>3, 23, 27</sup> Despite 36 to 59 putative inclusion proteins identified in the *C. trachomatis* genome, only 23 have been identified between five species (*muridarum*, *caviae*, *felis*, *pneumoniae*, and *trachomatis*).<sup>23, 24</sup> Effector proteins that modify the inclusion body, and are secreted by T3SS, include IncA proteins. IncA from *C. trachomatis* is the focus of this investigation and described later. Recognized by the host cell as “normal”, inclusion proteins undergo post-translational modifications.<sup>28</sup>

The common feature distinguishing inclusion proteins, however, is their 50- to 80-amino-acid, bi-lobed, hydrophobic domain that is localized to the inclusion body membrane.<sup>19, 23, 24, 27, 29</sup> Currently, at least 9 proteins are identified with this feature, and each inclusion protein has a unique function in the development process of the pathogen.<sup>16, 28</sup> Several inclusion proteins have been investigated in terms of function and localization. One major difficulty is the inability to manipulate *Chlamydia* genetically leaving only two

approaches 1) over expression in the host and 2) in vitro with recombinant protein. Despite these challenges, three Inc proteins have postulated functions. IncG and CT229 have been shown to interact directly with the host. IncG interacts with the host 14-3-3 $\beta$ , a regulatory molecule that binds to signaling proteins. 14-3-3 $\beta$  is postulated to interact with other proteins that facilitate vesicular transport processes.<sup>28</sup> CT229 has been shown to recruit host Rab4A GTPase to the inclusion membrane.<sup>30</sup> IncA from *C. trachomatis*, has been identified as a fusion promoting protein and is the focus of this thesis.

#### 1.4 Inclusion Protein A

IncA was the first inclusion protein to be identified and accumulates on the membrane of the inclusion body in all species. IncA has two trans-membrane helical segments with a C-terminal domain extending to the host cytoplasm (**Figure 1.3**).<sup>16, 31</sup> Current hypotheses indicate that IncA from *C. trachomatis* is the only *chlamydial* inclusion protein that facilitates homotypic fusion of inclusion membranes in multiply infected cells;<sup>16, 29</sup> however, approximately 2 percent



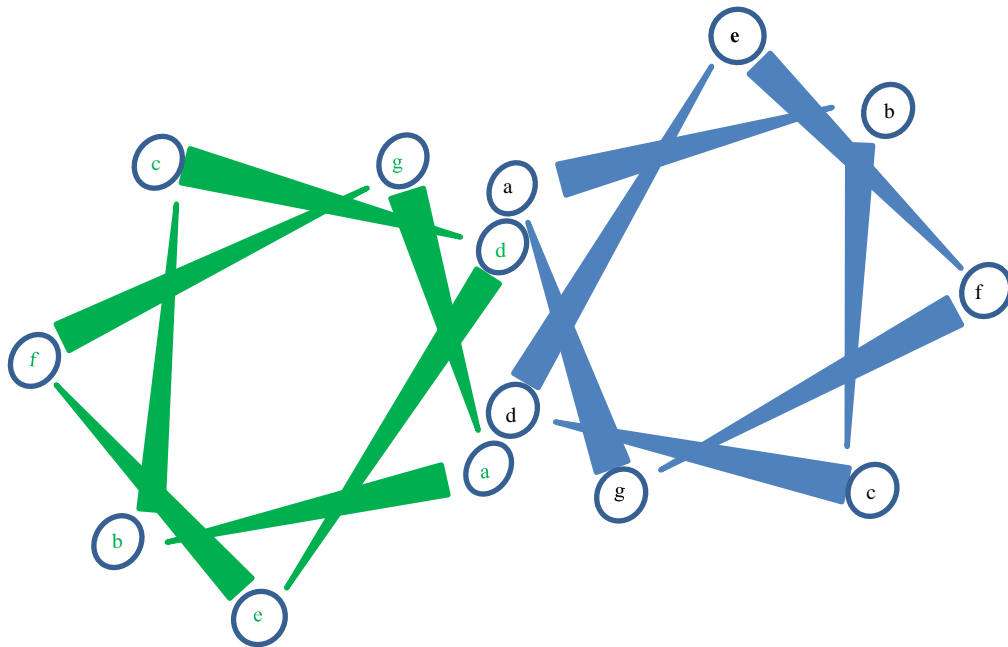
**Figure 1.3.** Proposed structure of the IncA monomer and location in the inclusion membrane. The C-terminal  $\alpha$ -helical domain faces the cytosol and the bi-lobed transmembrane region embedded in the inclusion membrane.

of *C. trachomatis* strains demonstrate non-fusogenicity as a result of aberrant *incA* DNA transcription.<sup>32</sup> Single base frame shifts and larger deletions lead to truncated IncA proteins that either partially fuse or lack fusion capabilities altogether.<sup>32, 33</sup> Phenotypes of IncA-positive and IncA-negative strains present differently; specifically, IncA-positive *C. trachomatis* serovars have increased virulence.<sup>18, 34</sup> Virulence differences may be attributed to (1) inclusions formed by IncA-negative *C. trachomatis* do not divide nor do they fuse,

hence, the pathogen has limited genetic exchange reducing antigenic variation and (2) increasing inclusion body membrane surface area may allow the pathogen to procure more host cell nutrients and thus, produce more RB's.<sup>32</sup> Structural features shared with SNARE proteins are what make IncA from *C. trachomatis* unique.

### 1.5 IncA from *C. trachomatis* shows similarity to SNARE Proteins

Soluble N-ethylmaleimide Attachment Receptor (SNARE) proteins are a family of proteins that promote fusion events in yeast and mammalian cells. Synaptobrevin,



**Figure 1.4.** Illustration of the heptameric leucine zipper motif. Top down view of a leucine zipper motif heptameric sequence. Positions A and D have a hydrophobic residue and a leucine, respectively.

synaptotagmin, and synaptosomal-associated protein 25 (SNAP-25) are the primary proteins involved in fusion events.<sup>35</sup> Sequence similarity is low in IncA proteins between species, however, sequence analysis of IncA proteins found in 6 different *chlamydial* species contain SNARE-motifs suggesting IncA and SNARE proteins share a common ancestor.<sup>28, 36</sup>

SNAREs and IncA share several features. First, the leucine-zipper heptad repeat which is defined by leucine as the first amino acid of each repeat and the fourth amino acid is a hydrophobic residue (**Figure 1.5**). The leucine-zipper motif is conserved proximal to the C-terminal domain and forms coiled-coil structures for both SNAREs and IncA. Genotypically, the *C. trachomatis* genome codes for 50 proteins to putatively engage in coiled-coil interactions.<sup>24</sup> Second, bioinformatics and modeling suggest that the polar residue located in the center of the motif may organize the coiled coils in the hydrophilic domain into a single amphipathic structure.<sup>27, 37</sup> SNAREs have a helix formed with either a glutamine or arginine residue located in the center.<sup>28, 36</sup> The structure of IncA is postulated to form a helix with a threonine residue located in the center. Finally, IncA proteins in *C. trachomatis* dimerize, and tetramers of IncA may form stable structures similar to the SNARE complex, thereby facilitating fusion events.<sup>27, 28, 38</sup> Previous studies show only the  $\alpha$ -helical region facing the host cytoplasm is necessary for fusion<sup>27</sup>, however, biochemical properties and the structure of the soluble C-terminal domain of the IncA dimer have yet to be characterized.



M T T P T L I V T P P S P P A P S Y S A N R V P Q P S L M D K I K  
 K I A A I A S L I L I G T I G F L A L L G H L V G F L I A P Q I T  
 I V L L A L F I I S L A G N A L Y L Q K T A N L H L Y Q D L Q R E  
 V G S L K E I N F M **L S V L Q K E F L H L S K E F A T T S K D**  
**L S A V S Q D F Y S C L Q G F R D N Y K G F E S L L D E Y K N**  
 S T E E M R K L F S Q E I **I A D L K G S V A S L R E E I R F L**  
**T P L A E E V R R L A H N Q Q S L T V V I E E L K T I R D S L**  
**R D E I G Q L S Q L S K T L T S Q I A L Q R K E S S D L C S Q**  
 I R E T L S S P R K S A S P S T K S S

**Figure 1.5.** *C. trachomatis* IncA primary sequence. The primary structure of IncA is illustrated by two transmembrane segments at the N-terminal domain, the central polar residue, the coiled-coiled domain, and leucine zipper. Larger font indicates the A and D residues with the leucine and hydrophobic residues. The underlined cysteine residues are at position 141 and 252.

## 1.6 Research Purpose

The overall goal of this investigation was to determine the orientation of the IncA dimer using in direct methods, specifically electron paramagnetic resonance (EPR) and site-directed spin labeling (SDSL). Polymerase Incomplete Primer Extension (PIPE) cloning methods retained the cleavable hexa-histidine affinity tag (his-tag) while the N-terminal transmembrane region (residues 1-85) was successfully deleted ( $\Delta 85$  IncA). PIPE mutagenesis changed one of two native cysteine residues to alanine at position 141 to create  $\Delta 85$  IncA252C. Position 252 was mutated to alanine creating mutant  $\Delta 85$  IncA141C. Both cysteine residues were removed to generate the  $\Delta 85$  IncA “cysless” mutant and protocols were established to express and purify all IncA mutants. Previous studies have shown full length, wild type IncA resistant to denaturation by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE).<sup>39</sup> In contrast, SDS-PAGE analyses of truncated IncA and the respective IncA mutants consistently observed monomer suggesting the bi-lobed transmembrane domain may influence folding and stability of  $\alpha$ -helices on the

C-terminal domain. In order to simplify purification, the his-tag was removed from  $\Delta 85$  IncA and  $\Delta 85$  IncA mutants. Size exclusion chromatography (SEC) shows cleavage of the his-tag on the N-terminal domain facilitates dimer formation. The soluble IncA dimer for all constructs was assessed with circular dichroism and all presented  $\alpha$ -helical character. With properly folded IncA mutants, biophysical studies were initiated. SDSL protocols were optimized for  $\Delta 85$  IncA141C and  $\Delta 85$  IncA252C. EPR and SDSL on truncated IncA mutants show moderately immobilized spin labels suggesting global tumbling. The results of this study provide a road map to prepare samples for distance analysis studies using Double Electron-Electron Resonance (DEER) experiments.

## **Chapter 2**

### **Theoretical Concepts**

#### **2.1 Cloning Overview**

In the early 1970's, scientists developed technology to manipulate nucleic acids and, subsequently, to incorporate non-native genes into a foreign host.<sup>40</sup> Polymerase chain reaction (PCR) is a technique that revolutionized molecular biology, biochemistry, and cell biology research by designing heating and cooling cycle regimens to amplify a single sequence of DNA into millions of copies.<sup>40</sup> Amplification, combined with the ability to isolate sequences of DNA and/or make alterations to these sequences led to a new field called "recombinant DNA technology." Using recombinant-DNA technology, proteins could be replicated identical to the original molecule. Systematically perturbing sequences allowed scientists to create sequences of interest and provide a new approach to study individual segments of DNA, the impact of mutations, and the structure and function of proteins manufactured downstream.

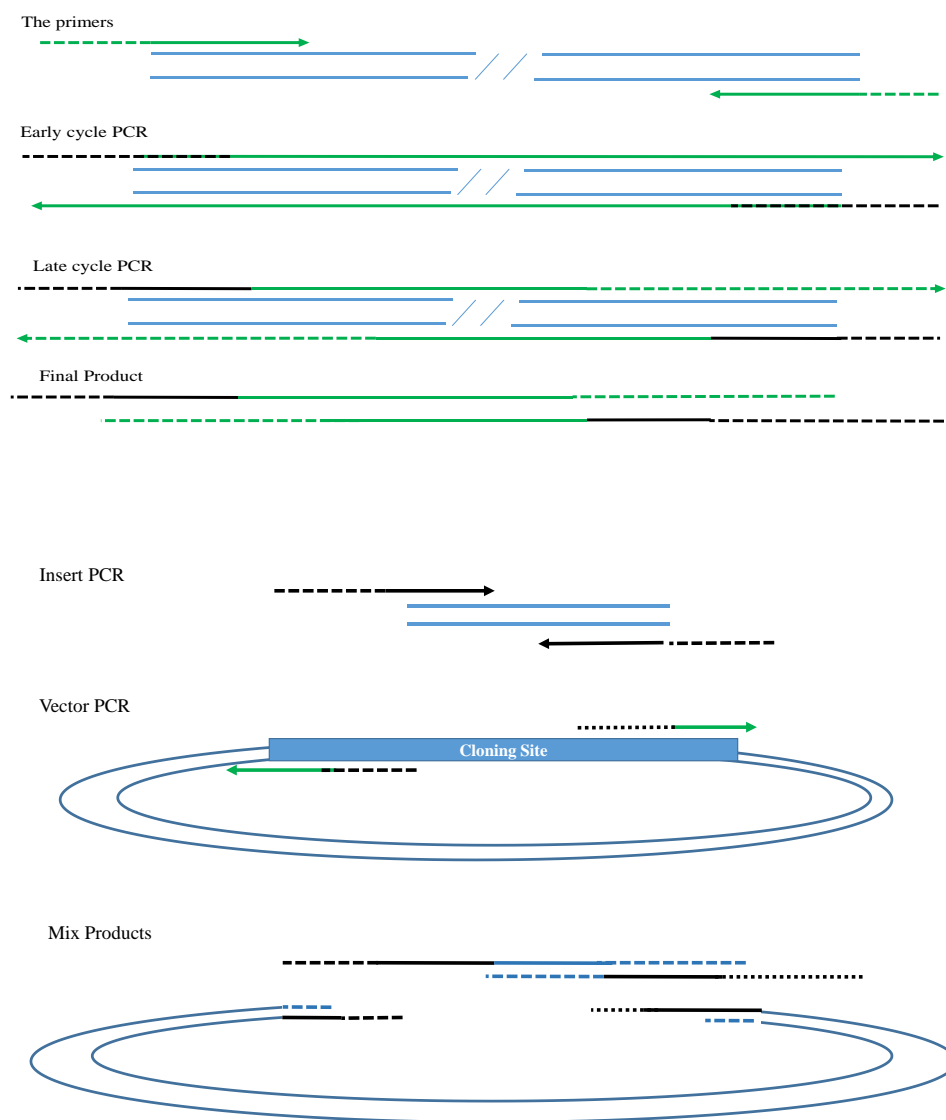
#### **2.2 PIPE Cloning and Mutagenesis**

The success of recombinant DNA technology and the subsequent development of high-throughput methods to characterize proteins, influenced scientists to expand their research to characterizing the proteome of entire organisms. In 2000, the Joint Center for Structural Genomics (JCSG) was formed with a mission to strategically expand protein structure characterization in known organisms.<sup>41</sup> Their efforts delocalized into Protein Structure Initiative Centers that have amassed a considerable amount of expertise in all

steps of structure characterization, ranging from target selection to structure determination.<sup>42</sup> Simultaneously, the JCSG worked to increase pipeline efficiency (stages in which a protein is identified as a suitable target to final structure determination) and reduce costs. Thermo-cycling, however, produces a mixture of DNA products. Cloning protocols after amplification, however, are tedious, laborious, and time consuming. Cloning required labor-intensive purification of the PCR product and lacked flexibility to alter and/or insert DNA sequences into any vector.<sup>42-44</sup> Restriction enzyme and recombinatorial cloning methods were flawed because they required expensive enzymes, were prone to error, and created unwanted mutations. Polymerase Incomplete Primer Extension (PIPE) cloning was conceived to meet these needs and cost constraints.

Fully formed DNA is not always created and PIPE cloning takes advantage of this imperfection. In the final stages of PCR, the 5' end is left incomplete due to sequence-specific stalling or insufficient availability of deoxynucleotide triphosphate (dNTP) or DNA building blocks (**Figure 2.1**).<sup>43</sup> Primers designed for the incomplete 5' ends of DNA can control cloning and mutagenesis with simpler methodology than classic cloning techniques.<sup>43</sup> In order for *Escherichia coli* (*E. coli*) to produce the protein of interest, both an insert gene to the DNA of the protein and the DNA of the vector are required. Inserts are amplified using designed primers that anneal to the first 25 base pairs of the protein's DNA sequence (i.e., forward primer) and the reverse complement of last 25 base pairs (i.e., reverse primer) of the 5' end of the opposite strand. The primers create annealing templates for the 5' end for combining PCR fragments.<sup>45</sup> The insert and vector are amplified with PCR separately, and the two unpurified PCR products from the vector and insert are

combined in different ratios (e.g., vector to insert 1:1, 2:1, 1:2). The insert and vector complementary termini regions create the expression plasmid when annealed.<sup>44</sup> The resulting solution is transformed into chemically competent cells.<sup>44</sup>



**Figure 2.1.** Illustration of PIPE cloning methodology. A) Primers with 5' complementary extensions represented by dashed green lines anneal to template DNA. Extensions are favored and additional template DNA is amplified early in the PCR cycle. B) During later PCR cycles strand synthesis varies and incomplete extension of the primer occurs. Left over PCR single stranded templates serve as annealing sites.

Two to four colonies are selected to evaluate cloning success via DNA sequencing. Literature suggests that the efficacy of PIPE cloning and PIPE mutagenesis is 85%. PCR mutations and primer mutations rarely occur.<sup>43, 45</sup> PIPE cloning and mutagenesis provides a less labor-intensive, cost-effective, and accurate method to quickly create large numbers of protein constructs without extra manipulations required by traditional cloning methods.

## **2.3 Expression and Purification**

### **2.3.1 *Escherichia coli* expression systems**

*E. coli* expression systems are cost efficient, have a short doubling time, and produce yields up to 30% of their biomass. *E. coli* expression systems are a convenient option for soluble proteins between 10 and 50 kilodaltons (kDa) and with few disulfide bonds. *E.-coli*-derived-expression systems, however, are not without drawbacks. High-level expression leads to accumulation of insoluble proteins that form inclusion bodies. Also, *E.-coli*-expression systems lack adequate mechanisms for disulfide-bond formation and have limited post-translational modifications. Expression with affinity or fusion tags on the N-terminal or C-terminal region enhances solubility and purification.<sup>46</sup>

Codon bias is another concern with high-level expression. Codons (three letter nucleotide sequences that code for amino acids) occur at different frequencies and rare codons tend to be expressed at low levels.<sup>46, 47</sup> Many copies of heterologous mRNA are produced with the addition of an inducing agent that depletes the pool of corresponding transfer RNAs (tRNAs).<sup>48</sup> Certain organism genomes are biased to the rare codons that occur in the *E. coli* hosts.<sup>48</sup> Frame shift errors, early translation termination, and mis-

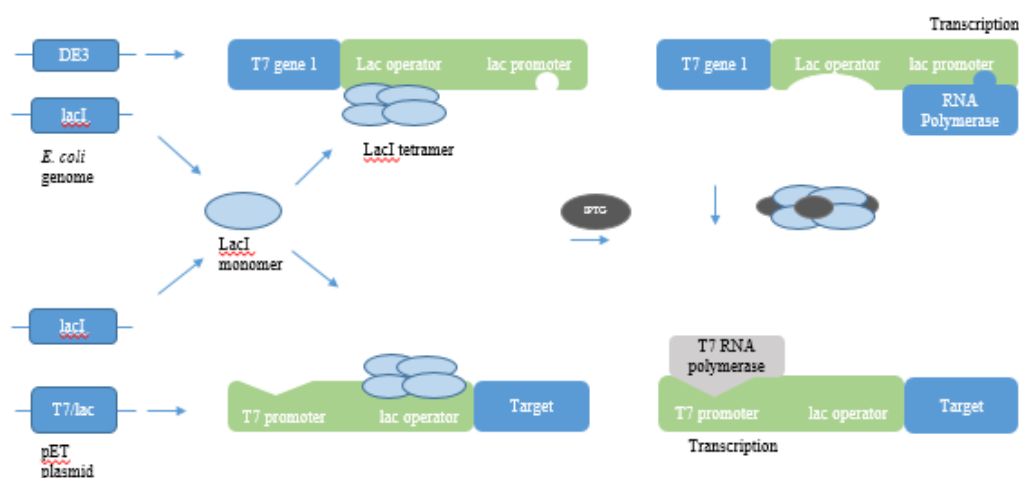
incorporation of amino acids occur when rare codons are clustered near the N-terminal region.<sup>46, 47</sup> Adenosine-Thymine (AT) rich genomes require chemically competent cells containing extra copies of *argU*, *ileY* and *leuW* tRNA genes that recognize AGA/AGG, AUA, and CUA codons, respectively, to alleviate codon bias.<sup>48</sup> Interestingly, proteins produced in low amounts (i.e., membrane proteins) tend to have more rare codons in their protein sequences and could significantly affect protein expression.

### 2.3.2 The pET expression system

A widely used *E. coli* expression system is the pET expression system derived from the pBR322 plasmid and described by Studier and Moffatt in 1986.<sup>49, 50</sup> Features that make this system appealing include the specificity of the RNA polymerase from bacteriophage T7 to a T7 promoter region. T7 RNA polymerase generates complete transcripts and the polymerase to elongate chains much faster than *E. coli*-RNA polymerase.<sup>49, 50</sup> Finally, gene expression is controlled by the experimenter (inducible) usually introduced at the optimal development of *E. coli* or until *E. coli* growth reaches an optical density at 600 nm of 0.6 or 0.8.

The pET expression system is controlled by the LacUV5 promoter and is isopropyl  $\beta$ -D-1thiogalactopyranoside (IPTG) inducible (**Figure 2.2**). The *lacI* gene represses both the LacUV5 promoter in *E. coli* and the T7/Lac-hybrid promoter located on the plasmid.<sup>51</sup> The T7 promoter and the lac operator are located 5' to the gene for the target protein. When the Lac operator is not repressed in the presence of the T7 RNA polymerase, transcription of the target gene is initiated.<sup>51</sup> The pET expression system has a modified Lac promoter

sequence where the native *E. coli* RNA polymerase can bind. The LacI tetramer is released from the Lac operator of *E. coli* when IPTG is introduced to the system. Native RNA polymerase binds to the lac promoter on the *E. coli* genome and T7 polymerase is transcribed, translated in high numbers, and expressed. IPTG also binds to the LacI tetramer bound to the lac operator on the plasmid. The IPTG bound LacI tetramer is released from the lac operator on the vector, the newly expressed T7 RNA polymerase binds to the T7/lac hybrid promoter sequence, and transcription of the target gene into mRNA begins.<sup>51</sup>



**Figure 2.2.** Rendering of IPTG function in pET expression systems. During expression, the LacI tetramer represses the Lac promoter in both the *E. coli* genome and the pET plasmid. IPTG releases the LacI tetramer inducing expression of the target protein.

The quantity of protein expression is reflected by the increase in transcription of mRNA. The beauty of the pET expression system is that T7 RNA polymerase is indigenous only to genetically engineered cells integrated with the genes for T7 RNA polymerase, the lac promoter, and the lac operator in its genome.<sup>51</sup> The T7 promoter, therefore, is not



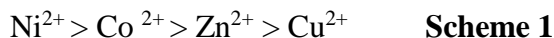
recognized by the host cell RNA polymerase. Target genes are mostly silent when not induced and, upon induction, only the target gene is transcribed.

### **2.3.3 Immobilized Metal Affinity Chromatography (IMAC)**

Immobilized Metal Affinity Chromatography (IMAC) capitalizes on the interactions between divalent metal ions in the stationary phase and histidine residues on the target protein in solution (i.e., the mobile phase). Genes for affinity tags are inserted onto the plasmid and proteins are expressed with the tag on either the N-terminal or C-terminal domains. An affinity tag with 6 to 14 histidine residues is commonly referred to as a “his”-tag. Increasing the number of histidine residues increases the binding strength of the affinity tag to the IMAC resin and may enhance separation of the target protein from contaminants. Histidine residues have two features that make them the best choice for an affinity tag: (1) due to their hydrophilic and flexible features, the his-tag seldom interferes with the function of the protein,<sup>52</sup> and (2) the imidazole side chain allows reversible coordinate bonds to divalent metal ions.

The his-tag is expressed on the target protein and binds to the metal complex on the stationary phase of the column while protein without the his-tag and other contaminants pass through the column. Target proteins are eluted with a high concentration of imidazole. Imidazole competes for coordination sites on the metal ions and displaces the tagged protein from the stationary column. Lowering the pH also elutes protein by protonating the imidazole moiety of histidine, decreasing the affinity to metal ions, and eluting the “his-tagged” proteins. This investigation used cobalt (II); however, a number of divalent ions can be used in IMA chromatography such as nickel (II), copper (II), cobalt (II), and zinc

(II). The affinity of divalent metals is conferred by their stability constants. Nickel has the highest stability constant and has a greater affinity to histidine than the other divalent metals. Decreasing affinity of typical divalent metals used in IMAC is shown in **Scheme 1**.



Although cobalt has a lower stability constant and, therefore, less affinity for histidine than nickel, cobalt has better selectivity.<sup>53</sup> The lower affinity of cobalt has been shown to be better at removing background proteins.<sup>53</sup> Sequentially adjacent or specially positioned juxtaposed histidine residues enhance cobalt's ability to bind to the target protein while nickel is promiscuous and binds to the imidazole moiety indiscriminately.<sup>54</sup>

### 2.3.4 Dialysis and Thrombin Cleavage

Although dialysis is a technique often confused as a purification step, the technique is specifically used to exchange buffers and removes small contaminants. The sample and the buffer solution (dialysate) are placed on opposite sides of a membrane that contains pores of a specific size. The smallest pore size that retains 90% of the sample is commonly referred to as “molecular weight cut off” (MWCO). Simple diffusion keeps larger molecules inside while allowing reagents and small contaminants to exit. Net movement of smaller molecules across the membrane ceases when equilibrium is reached.

In this case, the imidazole concentration following affinity elution must be significantly decreased for two reasons. First, imidazole inhibits thrombin cleavage. Thrombin is an active site-specific protease, it recognizes the consensus sequence

LVPRGS, and it cleaves between the peptide bond between arginine and glycine. Thrombin is so specific that mutations in its sequence decreases the enzyme's efficacy between 200 and 400 percent.<sup>55</sup> Second, imidazole is strongly absorbed in the ultraviolet region in the electromagnetic radiation spectrum and interferes with measuring the protein concentration using  $A_{280}$  as well as interfering with circular dichroism measurement (see below). Many vector systems have thrombin cleavage sites encoded in their DNA such that the his-tag can be cleaved, if necessary. In this investigation, the his-tag is cleaved in order to assess dimer formation. The protein sample, however, requires an additional purification step (such as size exclusion or ion-exchange chromatography) once cleavage is complete to remove thrombin and N-terminal cleaved product.

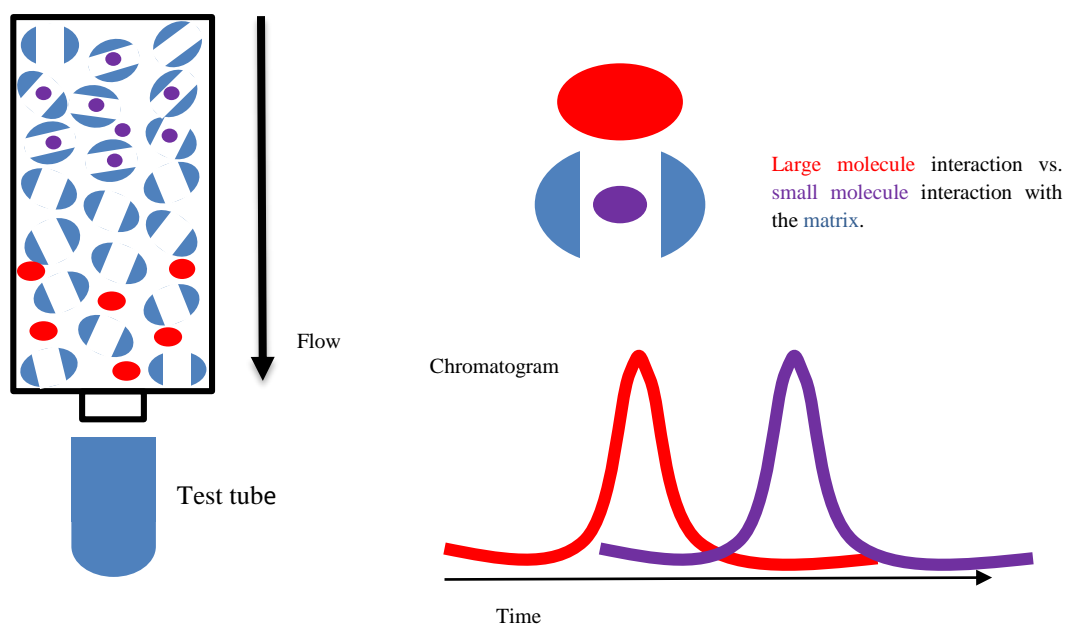
### **2.3.5 Size Exclusion Chromatography (SEC)**

Size exclusion chromatography (SEC) separates proteins based on their mass and shape. Swedish scientists Jerker Porath and Per Flodin optimized the use of dextran in the technique in 1959 and established reliable prediction of molecule size by the fraction in which the analyte elutes from the column.<sup>56</sup> Several molecular properties exist that hamper separation: dissociation into subunits, high concentrations of carbohydrates, and formation of weak complexes with dextran.<sup>57</sup> How exactly SEC separates polymers has been debated for the past 50 years. Literature indicates that separations are based on more factors than just molecular weight—the size and configuration must also be considered. According to Sun, *et. al.*, “The ability of a molecule to enter a pore depends on the way it is oriented with respect to the pore opening; the cross section of the chain perpendicular to the opening

may be the correct measure and not the average dimension.”<sup>58</sup> Conversely, some scientists suggest that particles separate, despite architecture and chemical composition, based on

$$V_H \propto M[\eta] \quad \text{Equation 2.1}$$

hydrodynamic volume ( $V_H$ ), where  $M$  is the molecular weight and  $\eta$  is the intrinsic viscosity (**Equation 2.1**).<sup>59</sup> Currently, no theory conclusively explains how SEC separates molecules, so the debate continues. Nevertheless, some principles are conserved throughout the scientific literature.



**Figure 2.3.** Depiction of size exclusion chromatography. Larger molecules do not interact with the matrix eluting faster than smaller molecules in the SEC column. Far left: Illustration of large molecules, small molecules, and the Sephadex matrix in a SEC column. Top right: Enhanced depiction of large and small molecules interacting with the Sephadex. Bottom right: Chromatogram depicting the elution behavior of large molecules and small molecules.

In SEC, water is the mobile phase, and the stationary phase is a “molecular sieve” composed of polymeric carbohydrates and acrylamides that make a cross-linking network of polymeric chains.<sup>60</sup> Absorption of water causes the resin to swell and create openings,

or pores, in the cross-links. Molecules that are greater in size than the largest pores do not interact with each other and molecules pass through the column relatively quickly. Smaller molecules will pass through the pores, and take a longer time to elute. In summary, particles elute from the column according to decreasing molecular size.<sup>60</sup> SEC can be used to separate and obtain the molecular weight of proteins, enzymes, nucleic acids, and even larger macromolecules such as hormones and polysaccharides. The maximum molecular weight of a molecule that can pass through a column is resin-specific and known as the exclusion limit. Exclusion limit is a property specific to the columns stationary phase. Similarly, the permeation limit defines the minimum molecular weight that is too small to pass through the cross-linking matrix and is a property of the mobile phase (**Figure 2.3**).<sup>60</sup>

The most popular matrix to separate proteins is Sephadex. Sephadex contains hydroxyl groups contained on the polymer's chain, providing polar properties and allow the carbohydrate material adsorb water. Cross-linking can be controlled to provide different pore sizes as well as exclusion and permeation limits. The matrix's name indicates the resin's ability to hydrate. For example, Sephadex G-200 has a water-regaining value of about 20 milligrams per milliliter (mg/mL). Resins are stable to mild redox agents, stable to bases and weak acids, and insoluble in water.<sup>60</sup>

## **2.4 Matrix Assisted Laser Desorption Ionization Time of Flight Mass Spectrometry (MALDI-TOF-MS)**

### **2.4.1 Overview**

Before recombinant DNA technology, PCR, or bioinformatics, traditional chemical and enzymatic methods were used to determine structures of proteins. Peptide sequences

were determined by Edman degradation or stepwise chemical degradation of amino acids from the N terminus to the C terminus, detecting changes through ultraviolet absorption or fluorescence spectroscopy.<sup>61</sup> The drive to accurately identify primary protein structure has brought significant improvements to mass spectrometry limits of detection, resolving power, and mass accuracy.<sup>61</sup> The accuracy of traditional techniques, such as chromatography and SDS-PAGE techniques, are limited. Current chromatography methods are still unable to separate molecules with similar properties (e.g., isozymes), and the mass determination derived by gel electrophoresis yields errors up to several percent. Both techniques are laborious and time consuming compared to matrix-assisted laser-desorption-ionization time-of-flight mass spectrometry (MALDI-TOF-MS).<sup>62</sup> MALDI-TOF-MS offers “soft” ionization to make polypeptides accessible to mass spectrometry analysis and mass accuracy in the range of 0.01 percent.<sup>62</sup>

#### **2.4.2 Hard vs. Soft Ionization**

Hard ionization produces ions by breaking bonds.<sup>60</sup> Before the late 1980's, mass spectrometry used this technique and was limited to small, thermo-stable compounds.<sup>61</sup> Soft ionization techniques, such as electrospray and chemical ionization (e.g., MALDI), form ions without breaking covalent bonds.<sup>61</sup> Effective development of methods to “softly” ionize molecules and transfer them from the liquid phase to the gas phase (without significant fragmentation) made MS suitable for polypeptide analysis.<sup>62</sup>

#### **2.4.3 Mechanism of Action**

The target analyte is co-spotted (the protein and the matrix are spotted in the same area such that the two compounds mix) with a weak organic compound or matrix that

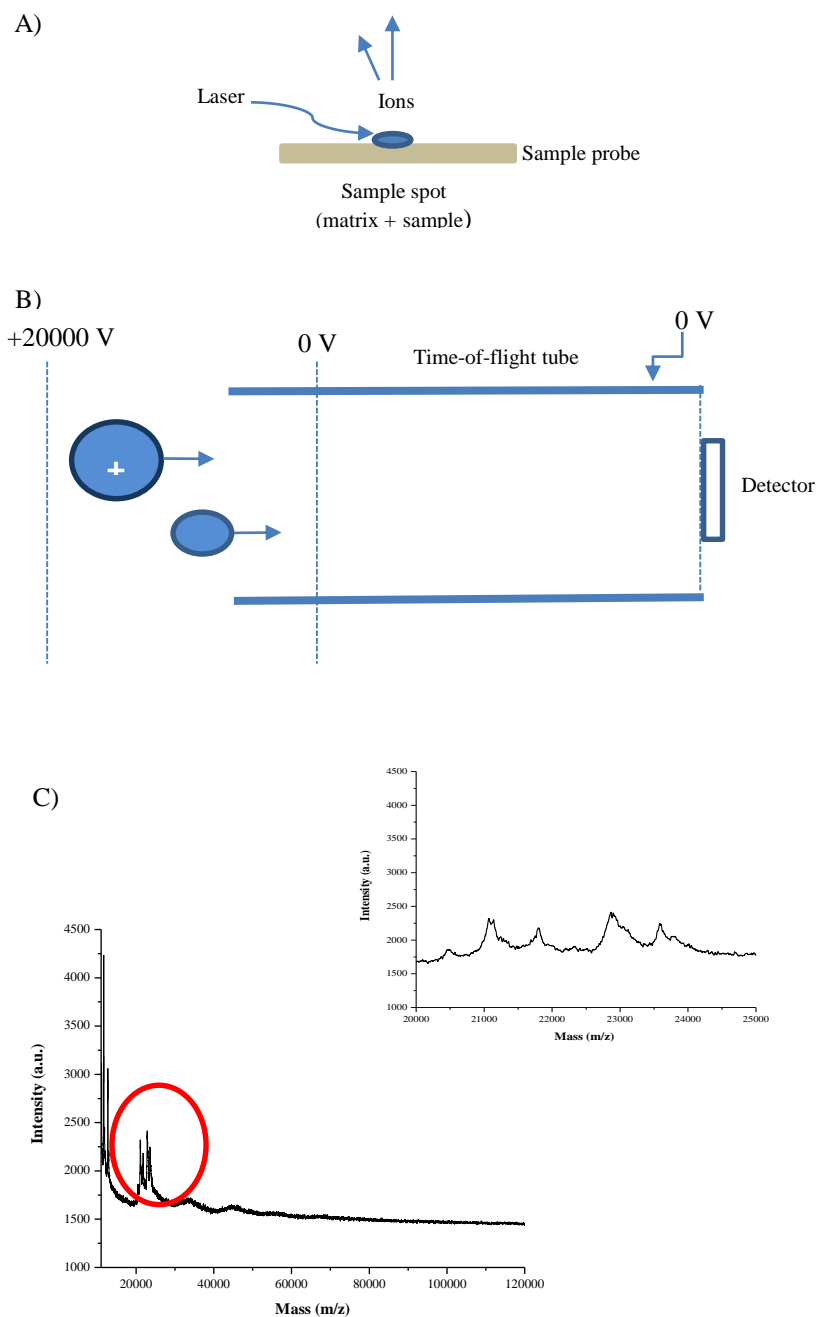
crystallizes the molecule, facilitates ablation and ionization, and prevents the destruction of the analyte (**Figure 2.4A**).<sup>61, 63, 64</sup> The matrix and the analyte absorb energy at the same wavelength.<sup>61</sup> The crystallized compound is exposed to a short pulse of light from a laser. The matrix volatilizes the sample, transfers protons from the matrix to the molecule, and forms ions.<sup>64</sup> The charged molecules are accelerated into the source region of specified length under vacuum in the TOF-MS, where they are separated according to their kinetic energies as shown in **Figure 2.4B**.<sup>62, 64</sup>

$$KE = zeV = \frac{1}{2}mv^2 \quad \text{Equation 2.2}$$

KE is expressed with respect to the potential applied or with respect to the velocity of ions.<sup>64</sup> The relationships are given by the **Equation 2.2**: where  $z$  is the charge,  $e$  is the charge of the electron ( $1.6022 \times 10^{-19}$  C),  $V$  is the potential drop the ions experience before entering the flight tube,  $m$  is the mass of the ion, and  $v$  is the velocity of the ion. **Figure 2.4C** illustrates a typical chromatogram obtained by MALDI-TOF MS.

## 2.5 Circular Dichroism (CD)

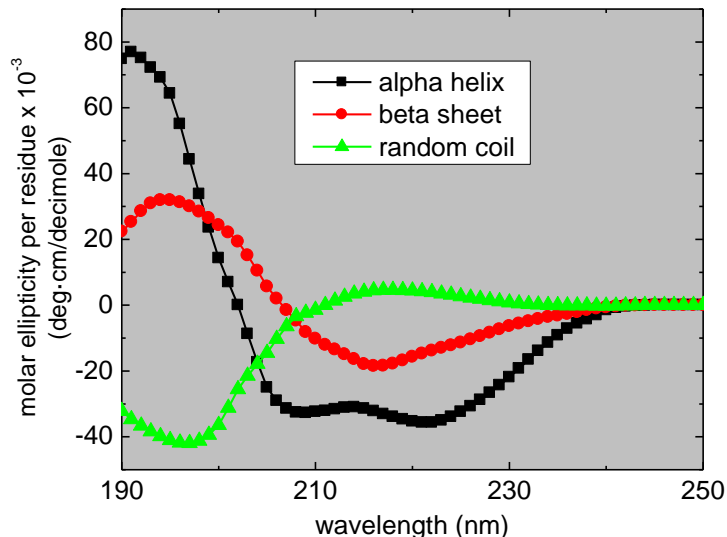
Driven by improvements in recombinant DNA technology, bioinformatics, and data analyses, protein structure and function has been a significant research topic since the late 1980's.<sup>65</sup> The increasing number of definitively characterized protein structures has led to the need to investigate proteins in native conditions as well as their rates of structural changes.<sup>65</sup> Circular dichroism applications include secondary structure determination (i.e.,  $\alpha$ ,  $\beta$ , and random coils), whether a purified protein is folded or whether multi-domains are expressed correctly, and whether a mutation affects its conformation or stability.<sup>65, 66</sup>



**Figure 2.4.** MALDI TOF sample ionization and source region. A) A sample co-spotted with matrix allowing ionization without destructing the analyte. B) As molecules enter the source region, smaller cations travel faster than larger ones towards the detector where they are separated according to their mass to charge ratio. C) Representative MALDI-TOF MS spectrum of  $\Delta 85$  IncA with  $m/z$  fragments ranging from 21.23.6. The inset is the enhanced spectrum region from the red circle delineating the four peaks.



When linearly polarized light passes through a chiral molecule, the left and right polarizations travel at different speeds. If the left and right components are absorbed at different magnitudes, the two circularly polarized components are absorbed differently. Circular dichroism is the difference in the left and right absorption. The electric component of the plane of polarized light creates an electric dipole moment, and the magnetic component creates a magnetic dipole moment. These motions combine to create both translational and rotational excitation of an electron. The sample is scanned in the far



**Figure 2.5.** *Characterization of secondary structure in circular dichroism spectroscopy.*<sup>1</sup> Rendering of secondary structure signatures found in protein. The  $\alpha$ -helices have minimums at 210 and 208 nm, beta sheets have minimums located approximately 218 nm. Disordered proteins or random coils have the most negative points between 195 and around 200 nm. Graphic reprinted with express written consent from Allied Protein Laboratories, Inc.

ultraviolet region (240 nm and below) where broad  $n \rightarrow \pi^*$  transitions are located at 190 and 220 nm. The chromophore for the secondary structure is the peptide bond, but aromatic side chains can also contribute to the spectral signal.<sup>65</sup> Based on data derived from crystal structures, signals are compared to a library of secondary structures determined by X-ray

crystallography.<sup>65</sup> For example,  $\alpha$ -helices are characterized by a negative signal with minima at 222 and 208 nm, the  $\beta$ -pleated-sheet signature is a negative signal near 218 nm, and disordered regions exhibit negative signals at 195 and 200 nm (**Figure 2.5**).

## 2.6 Electron Paramagnetic Resonance (EPR)

Electron Paramagnetic Resonance (EPR) is a spectroscopic technique that detects species with unpaired electrons. Many books and journal articles have attempted to make the technique accessible to life sciences.<sup>67-70</sup> A concise description of continuous-wave (CW) EPR and spin labels is offered below.

### 2.6.1 EPR Overview

In 1944, Yevgeny Zavoisky, a physicist at Kazan State University in the former Soviet Union, was the first to see the possibilities for EPR.<sup>71</sup> During World War II, the development of RADAR required reliable and tunable microwave sources, thus making components for EPR relatively inexpensive and reliable. Instruments were exclusively constructed in-house until the 1980's when the Bruker Biosciences Corporation began making them commercially available. Initially, EPR was nearly exclusively used to characterize transition metals but fell out of favor due to its complexity.<sup>67</sup> As a result of the ability to add spin labels and, subsequently, a lone electron onto polypeptides, the technique has experienced a renaissance within structural biology research.<sup>70</sup>

X-ray crystallography has been lauded as the gold standard for protein architectural studies, however, this technique provides only structural data *in the conditions of the crystal*.<sup>68</sup> Proteins, especially membrane proteins, are not static entities. Structure

determination provides perspective on dynamics, which in turn, facilitates function characterization. The uniqueness of EPR stems from its ability to detect only unpaired electrons, in any phase, over a wide range of temperatures *a priori*. One of the best features about EPR is its ability to characterize proteins in an environment that mimics its native, dynamic environment.

### 2.6.2 Continuous-Wave (CW) EPR

Electrons possess a magnetic dipole attributed mostly to spin angular momentum. Experiencing an applied magnetic field splits electrons into two populations due to the Boltzmann distribution. The amount of splitting depends on the strength of the magnetic field as well as electron-nuclei interactions.<sup>70</sup> Spins are aligned either parallel or anti-parallel to the applied magnetic field and the difference in energy is called “Zeeman splitting,” as characterized by the **Equation 2.3**:

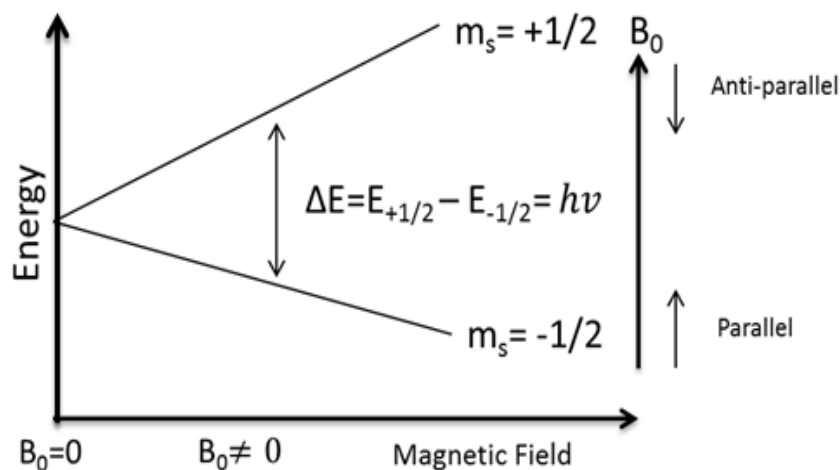
$$\Delta E = h\nu = g\beta_e B_0 \quad \textbf{Equation 2.3}$$

Where  $h$  is Planck’s constant,  $\nu$  is frequency of microwave radiation,  $\beta_e$  is the Bohr magneton, and  $B_0$  is the applied magnetic field. The g-factor ( $g$ ) defines resonance position and is define by the **Equation 2.4**:

$$g = \frac{h\nu}{\beta B_0} \quad \textbf{Equation 2.4}$$

The g-value for unpaired electrons is 2.0023<sup>67, 70</sup> and deviations in this value provides molecular structural information near the unpaired electron. In a CW experiment, the magnetic field is swept while a constant microwave frequency is applied, thereby inducing

a spin flip (**Figure 2.6**).<sup>72</sup> An EPR spectrum can be observed when the magnetic field and the microwave frequency are equal and is said to exhibit resonance. A myriad of microwave frequencies are used in scientific research (i.e., S band (2 – 4 GHz), X-band (9-12 GHz), W Band (35 GHz)), however, X band remains the optimal frequency for continuous wave protein structural studies. Hyperfine splitting can be observed due to the interaction of electrons with nuclei that possess nuclear-spin angular momentum (**Figure 2.7**).



**Figure 2.6.** *Two States of Electron Alignment in a Magnetic Field.* An applied magnetic field produces electrons in two populations: against the applied magnetic field or anti-parallel and aligned with the applied magnetic field or parallel. The Zeeman splitting is the difference in energy states.

The Hamiltonian, which describes both electron and nuclei interactions, is defined by

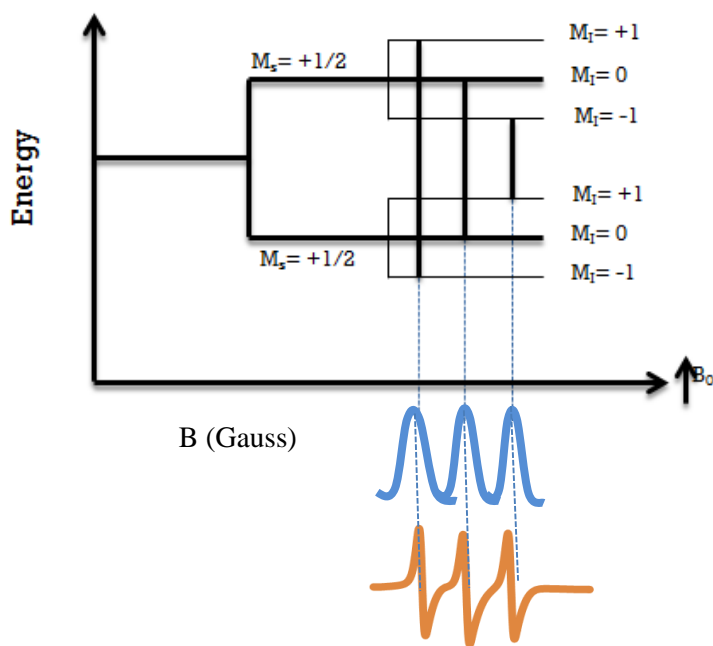
**Equation 2.5:**

$$\mathcal{H} = g_e |\beta_e| \hat{S} B_0 + \hat{S} A \hat{I} - g_N \beta_N B_0 \hat{I} \quad \text{Equation 2.5}$$

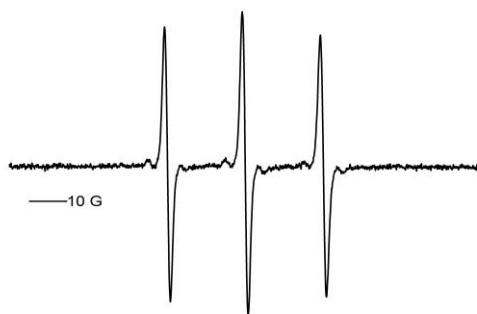
Where  $\hat{S}$  is the electron spin angular momentum,  $\hat{I}$  is the nuclear spin angular momentum,  $A$  is the hyperfine tensor, and nuclear Zeeman term is  $g_N \beta_N B_0 \hat{I}$ . These mathematical

expressions describe the molecular interactions of different components in biological systems, however, it is the EPR signature absorption spectrum that provides detailed structural information.

Absorption energy is measured and the first derivative is detected as a spectra (Figure 2.9). Phase sensitive detection, requires the first derivative to be recorded to enhance sensitivity, minimize noise from the detection diode, and eliminate baseline instabilities attributed to direct current electronics.<sup>72</sup>



**Figure 2.7.** *Hyperfine splitting.* The absorption peak is measured and the EPR signature spectra is the first derivative of that signal. Three peaks are observed due to the interaction of electronic spin with nuclear spin resulting in the splitting pattern shown here.



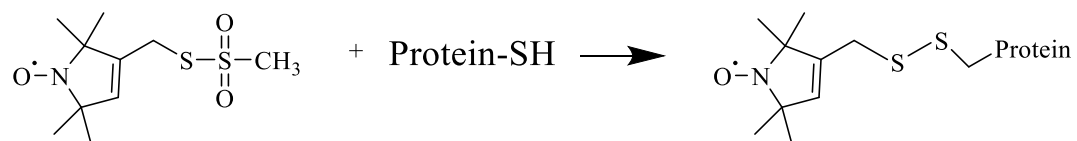
**Figure 2.8.** *Illustration of free electron in an applied magnetic field.* Lacking attachment to the protein via disulfide bonds, MTSSL demonstrates the three peaks that characterize nitroxide spin label free electron spin signatures. Deviations from this pattern yield information regarding the electron environment.

EPR when coupled with site-directed spin labeling provides information regarding protein folding, conformational changes during protein function, geometry of assembled subunits in multi-component systems, and the arrangement of proteins in membranes (Figure 2.10).<sup>68, 70</sup>

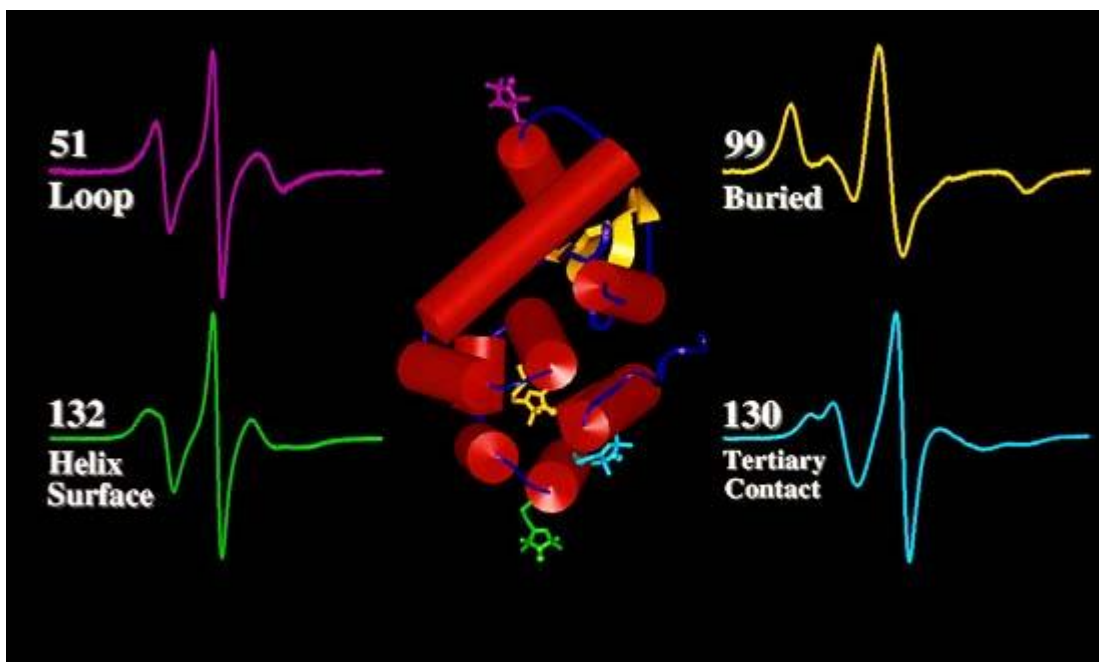
### 2.6.3 Site-directed Spin Labels

Most molecules lack unpaired electrons. Spin labels, however, provide a lone, stable electron added to a protein via a  $\text{SN}_2$ -like reaction to cysteine side-chains creating a disulfide bond. The most commonly used spin label in structural biology is 1-(oxy-2, 2, 5, 5 tetramethylpyrroline-3-methyl) methanethiosulfonate (MTSSL), and the resulting spin labeled side chain is referred to as R1 (Figure 2.9). For example, 141R1 implies that the native cysteine located on residue 141 is modified with MTSSL. MTSSL added to the protein using existing cysteine residues or by strategically creating cysteine mutations on a protein, allow R1 to provide subtle structural details found in EPR spectra on a

nanosecond timescale at X-band (9.5 gigahertz) due to sensitivity to secondary structure and protein backbone motion coupled with an applied magnetic field.<sup>73</sup>



**Figure 2.9** *Methanethiosulfonate reaction.* Methanethiosulfonate (MTSSL) reacts with cysteine residues and attach to the protein as the side chain known as R1.



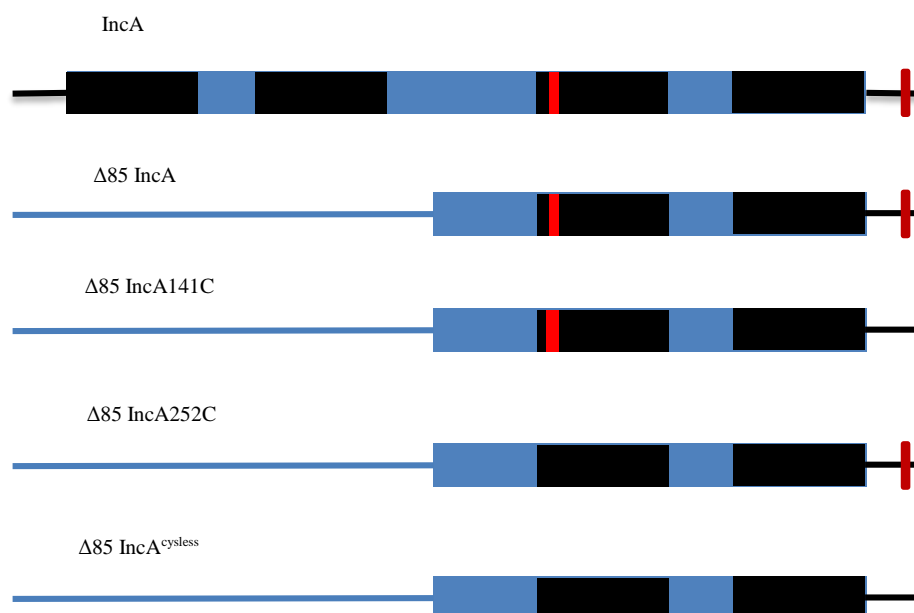
**Figure 2.10** *Illustration of structural information generated from EPR spectra.* Color coded R1 labels are matched with their respective spectra indicating the structural information gained from EPR spectra. Graphic reprinted with expressed written consent from Wayne Hubbell.

## Chapter 3

### Materials and Methods

#### 3.1 Cloning and Expression

PCR and PIPE cloning were used to generate the soluble C-terminal domain of IncA using the wild-type *incA* gene (Functional Genomic Resource Center, Rockville, MD) as a template. The soluble C-terminal domain of IncA is referred to as  $\Delta 85$  IncA and corresponds to residues 86-273. Primers were purchased from Novagen (San Diego, CA). Thermo cycler settings were: initial denaturation at 93°C (3 minutes), subsequent denaturing at 93°C (1 minute), annealing at 66°C (1 minute), elongation at 70°C (48 seconds), final elongation at 70°C (24 seconds), and final hold at 4 °C (until removal). The insert was introduced into a pET28b vector. The DNA sequence was confirmed by Gene Wiz, Inc. (South Plainfield, New Jersey).



**Figure 3.1** Depiction of PIPE cloning and PIPE mutagenesis constructs. Constructs prepared for IncA oligomeric studies. Shown here: helical regions, cysteine residues, and random coil/disordered chains.



Three additional  $\Delta 85$  IncA mutants were generated using PIPE mutagenesis:  $\Delta 85$  IncA141C,  $\Delta 85$  IncA252C, and  $\Delta 85$  IncA “cysless”. All mutants are summarized in **Figure 3.1**. The number corresponds to the location of the cysteine residue. For example,  $\Delta 85$  IncA141C corresponds to truncated IncA with a cysteine residue at position 141. Cysteine residues mutated to alanine residues are  $\Delta 85$  IncA<sup>cysless</sup> was prepared using  *$\Delta 85$  incA252C* gene and the primers for  $\Delta 85$  IncA141C. All mutants were expressed and purified in the same manner.

Truncated IncA ( $\Delta 85$  IncA) was amplified by transforming the resulting DNA into Top 10 competent cells and plated onto a Luria Broth (LB) kanamycin agar plate, and allowed to grow overnight at 37°C. A single colony was placed into a 5 mL LB kanamycin solution and shaken overnight (~12 hours) at 37°C at 225 revolutions per minute (RPM). Next, the solution was centrifuged, the supernatant removed, and the DNA was extracted and purified using a Qiagen Kit (Germantown, MD). The IncA plasmids were introduced to *E. coli* bacteria by transformed into chemically competent BL21-DE3-RIL cells. The cells were incubated at 37°C for one hour. The solution was streaked onto a double antibiotic (kanamycin and chloramphenicol, 50 mg/mL each) LB plate overnight at 37 °C. Overnight (O/N) cultures were prepared by introducing single colonies into 5-mL LB, kanamycin, and chloramphenicol culture tubes; incubated at 37 °C; and shaken at 225 RPM overnight. The O/N cultures were inoculated into 500 mL of Terrific Broth, kanamycin (50 mg/mL), and chloramphenicol (50 mg/mL). Cells were allowed to grow for 4 hours in Terrific Broth until the optical density at 600 nanometers was between 0.6 and 0.8. The culture was induced with 1 mM isopropyl  $\alpha$ -D-1-thiogalactopyranoside (IPTG; Sigma

Aldrich) and allowed to grow for an additional 5 hours. The cells were harvested by centrifugation at 4,000 RPM (1,790 x g) for 20 minutes at 22°C.

### **3.2 Purification**

The cell pellet was re-suspended in lysis buffer (150 mM NaCl, 150 mM Tris, pH 7.8). Protease inhibitor (Complete Protease Inhibitor, Roche Diagnostics, Indianapolis, Indiana) was added to the solution, and the cells were lysed using the microfluidizer (Microfluidics, Newton, Massachusetts). Cell debris and protein were separated by centrifugation at 11,000 RPM (9,480 x g) for 45 minutes. Immobilized metal affinity chromatography (IMAC) was used to purify the  $\Delta 85$  IncA from the supernatant. The column was prepared by washing 2 mL of chelating sepharose resin (GE Healthcare, Uppsala, Sweden) with 3 column volumes of Milli-Q water, 10 mL of 100 mM cobalt chloride ( $\text{CoCl}_2$ , Sigma Aldrich), washed again with 3 column volumes of water, and then equilibrated with 20 mL of lysis buffer. The subsequent supernatant from the centrifugation was introduced to the column, washed with 30 mL of wash buffer (150 mM sodium chloride (NaCl), 150 mM Tris, 1 mM tris (2-carboxyethyl) phosphine (TCEP), and 20 mM imidazole at pH 7.8) and eluted in two equal fractions of 15 mL of buffer (150 mM NaCl, 150 mM Tris, 1 mM TCEP, and 500 mM imidazole at pH 7.8). Sodium dodecyl polyacrylamide electrophoresis (SDS-PAGE) was used to analyze the purification process and NanoDrop <sup>TM</sup> ultraviolet visible spectrophotometer (NanoDrop Technologies, Inc, Wilmington, DE) was used for protein quantification.

### 3.3 Dialysis and Thrombin Cleavage

The elution fractions were combined and dialyzed against 4 L of dialysis buffer (1 mM TCEP, 75 mM Tris, and 150 mM NaCl at pH 7.8) two times for three hours each. Thrombin was introduced to the resulting sample, shaken at 100 RPM at 22°C, and incubated overnight. Cleavage was analyzed by SDS-PAGE. Thrombin was removed by introducing the resulting solution onto a benzamidine column (D-amino-benzamidine agarose cross-linked 4% beaded agarose, cyanogen-bromide activated, Sigma Aldrich, St. Louis, Missouri). The flow through containing  $\Delta 85$  IncA was concentrated using an Amicon 20-mL concentrator (Millipore, Darmstadt, Germany) with a molecular weight cut-off of 10,000 Da.

### 3.4 Size Exclusion Chromatography (SEC)

A HR Sephacryl 26/60 200 S column (G.E. Healthcare, Uppsala, Sweden) was washed with Milli-Q water then equilibrated with buffer (1 mM TCEP, 75 mM Tris, and 150 mM NaCl at pH 7.8). Cytochrome c (12 kDa), carbonic anhydrase (29 kDa), and bovine serum albumin (66 kDa) were prepared in 4 mg/mL concentration in buffer (1 mM TCEP, 75 mM Tris, and 150 mM NaCl at pH 7.8). Five runs were performed in the preparative column with the three standards and two samples of  $\Delta 85$  IncA.

### 3.5 MALDI-TOF-Mass Spectrometry

Soluble C-terminal IncA was analyzed on a Bruker Microflex MALDI. Using a MSP 96 target ground steel plate (Bruker Daltonics, Billerica, MA). IncA mutants were co-spotted with a zip tip and sinapinic acid in 50% acetonitrile and 0.05% trifluoroacetic acid (TFA). The instrument was calibrated by the close external method using a series of high-

molecular-weight-protein calibrants (Trypsinogen (24 kDa), and Chymotrypsin (25.6 kDa), and Bovine Albumin (43 kDa)). Two hundred laser shots were averaged to procure spectra with the following settings: positive ion, linear mode, and m/z range 4,000—20,000 or 20,000—100,000, and grid voltage 40 to 75%. Data were analyzed using OriginPro.

### **3.6 Circular Dichroism**

CD data were collected using an AVIV-CD Model 410 spectrophotometer. Spectra were acquired using a 1-mm path length quartz cuvette and the temperature set point was 4.0°C. Buffer spectra taken from four consecutive scans (260–200 nm, 2-second averaging time, and 0.5-nanometer [nm] steps) were averaged and then subtracted from the average of three consecutive scans of sample spectra. Data were analyzed using OriginPro, and  $\alpha$ -helix content was estimated using mean residue molar ellipticity at 208 and 222 nanometers ( $[\theta]_{208}$  and  $[\theta]_{222}$ ).

### **3.7 Site-Directed Spin-Labeling of IncA Mutants ( $\Delta 85$ , $\Delta 85$ IncA141C, $\Delta 85$ IncA252C) for EPR Studies.**

Two mutants used in this investigation required different means of preparation.  $\Delta 85$  IncA141C failed to bind adequately to methanethiosulfonate spin label (MTSSL) using standard protocols. TCEP was removed from the  $\Delta 85$  IncA141C (truncated IncA with a cysteine residue mutated to alanine at position 252) by introducing 2.5 mL of concentrated eluent from the SEC step to a PD-10 desalting column.  $\Delta 85$  IncA141C was eluted with 3.5 mL of spin-labeling buffer (75 mM Tris and 150 mM NaCl at pH 7.8). MTSSL was added at 10-times molar excess to  $\Delta 85$  IncA141C mutant, and the sample was allowed to incubate

overnight (12-15 hours) at ambient temperature.  $\Delta 85$  IncA141C was concentrated to 2.5 mL, loaded onto the PD-10 column, and eluted with 3.5 mL of labeling buffer. Finally,  $\Delta 85$  IncA141C was concentrated to approximately 250  $\mu$ L ( $\sim 400$   $\mu$ M). The protocol for  $\Delta 85$  IncA252C (truncated IncA with a cysteine residue mutated at position 252) mutant had three significant changes. First, after the TCEP was removed,  $\Delta 85$  IncA141C was allowed to incubate 24 hours at ambient temperature. Second,  $\Delta 85$  IncA141C was incubated with 5-times molar excess of MTSSL. The  $\Delta 85$  IncA141C mutant was concentrated to 2.5 mL and then introduced onto a PD-10 column a second time. Third, the sample was concentrated to 500  $\mu$ L, 2 mL of labeling buffer was added, and the sample was subsequently concentrated to 500  $\mu$ L. This process was repeated three more times. Finally,  $\Delta 85$  IncA141C was concentrated to a final volume of 500  $\mu$ L ( $\sim 400$   $\mu$ M).

### **3.8 Continuous Wave Electron Paramagnetic Resonance Spectroscopy**

Room-temperature X-band EPR spectroscopy was performed on a Bruker EMX spectrometer with an ER 4123D dielectric resonator (Bruker Bioscience Corporation, Billerica, MA) or a modified Varian E-line 102 Century series spectrometer with a loop gap resonator (Medical Advances, Milwaukee, Wisconsin). All X-band spectra were taken using 2-milliwatt (mW) incident microwave power, 1 gauss (G) field modulation, and a scan range of 100 G. Sample volumes were 8  $\mu$ L and were loaded into round boro capillaries (0.60 mm inner diameter x 0.84 mm outer diameter; Vitrocom, Mountain Lakes, New Jersey). A protein concentration of approximately 250 mM was used in all EPR experiments. Data were analyzed using OriginPro.

## Chapter 4

### Results and Discussion

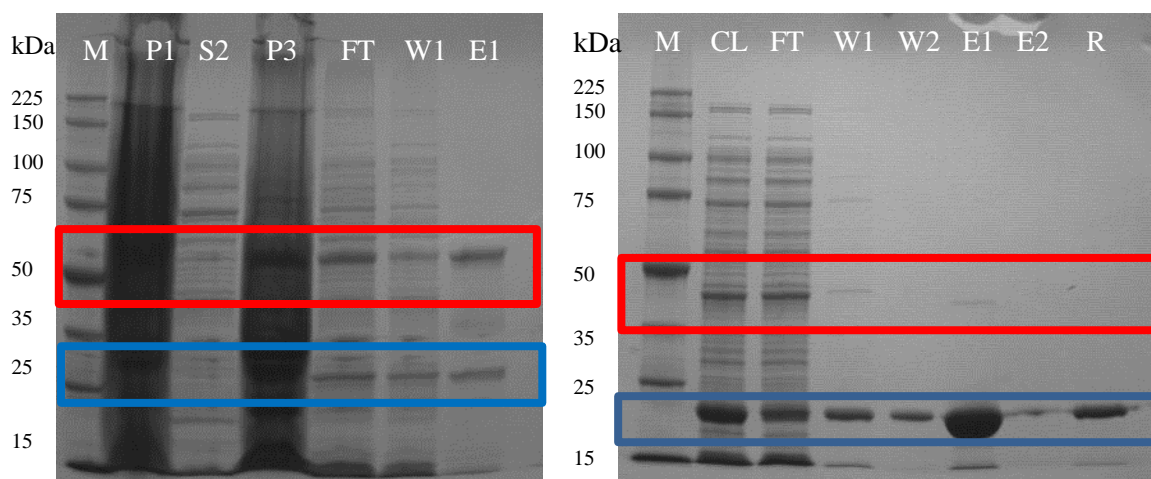
#### 4.1 Cloning and mutagenesis

The PIPE cloning technique successfully inserted the truncated *incA*-DNA gene into the pET28b vector and PIPE mutagenesis successfully removed selected native cysteine residues (see **Figure 3.1**). All sequence results were verified using GeneWiz, Inc.

#### 4.2 Expression and purification

According to ExPasy, the theoretical molar mass of wild-type IncA from *C. trachomatis* is approximately 30.3 kDa.<sup>74</sup> Wild type IncA and truncated IncA ( $\Delta 85$  IncA) were analyzed by Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). In each gel, wild type IncA,  $\Delta 85$  IncA and truncated mutants exhibited substantial protein quantities in each fraction. **Figure 4.1** (left) shows SDS-PAGE results from wild type IncA extraction fractions and IMAC purified samples. Full-length, wild type IncA presents two bands: a thick band at approximately 60 kDa as well as a faint band at approximately 30 kDa in the elution fraction 1 lane. These results suggest wild-type IncA is resistant to denaturation and most likely biologically stable as a dimer. This result is congruent with previous findings.<sup>39</sup> Truncated IncA ( $\Delta 85$  IncA) has a theoretical molecular weight of 21.3 kDa (23.3 with the his-tag).<sup>74</sup>  $\Delta 85$  IncA presents as a thick band around 23 kDa consistent with monomer (**Figure 4.1, right**). Significantly less dimer is represented by the faint band near 45 kDa. SDS-PAGE results demonstrate the  $\Delta 85$  IncA construct lacks the denaturation resistance property. This data was an unexpected development and shows truncated IncA is stable as a monomer. Previous studies have shown the bi-lobed hydrophobic region is not necessary to maintain the deleterious function of IncA and fusion

can take place with the soluble helices alone.<sup>27</sup> SDS-PAGE data suggests that the bi-lobed, hydrophobic N-terminal region may confer as yet uncharacterized properties to the protein as a whole and contributes to the stability of the protein fold despite exposure to a denaturant.

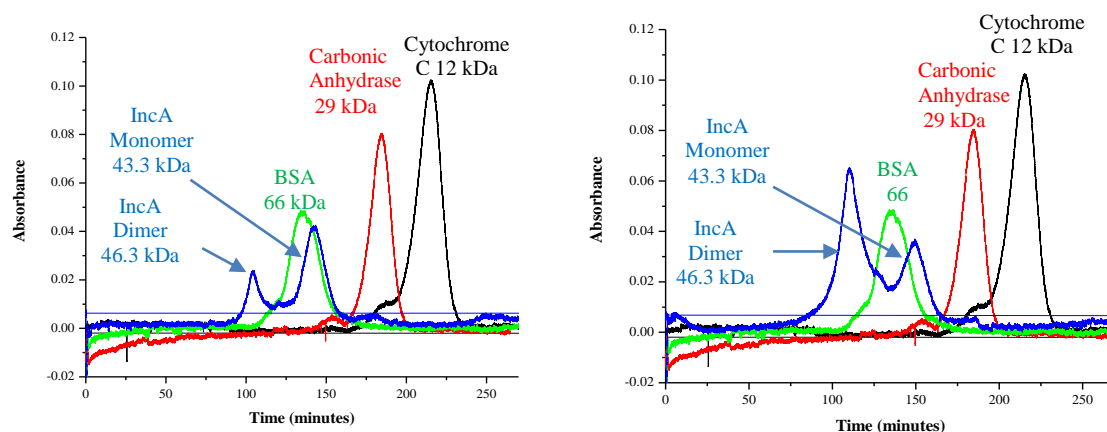


**Figure 4.1.** Purification results of wild type and  $\Delta 85\text{IncA}$  using IMAC. SDS-PAGE using a 10% Protean (Bio-Rad) gel was used to analyze  $\text{Co}^{2+}$  affinity chromatography fractions: **Left:** Marker (M), Pellet 1 (P1), Supernatant 2 (S2), Pellet 3 (P3), Flow through (FT), Wash 1 (W1), Elution 1 (E1). SDS PAGE analysis shows wild type IncA exists in two states as both a **dimer** (60 kDa) and a **monomer** (30 kDa). However, the thicker band at 60 kDa indicating resistance to denaturation and is most stable as a dimer. The fainter band in the blue box is the monomer. **Right:** Marker (M), Cell Lysate (CL), Flow through (FT), Wash 1 (W1), Wash 2 (W2), Elution 1 (E1), Elution 2 (E2), and Resin (R). The IncA soluble C-terminus has a theoretical molecular weight of 21.3 kDa; the bands in the red box indicate that the monomer was successfully expressed and not resistant to denaturation.

### 4.3 Size Exclusion Chromatography

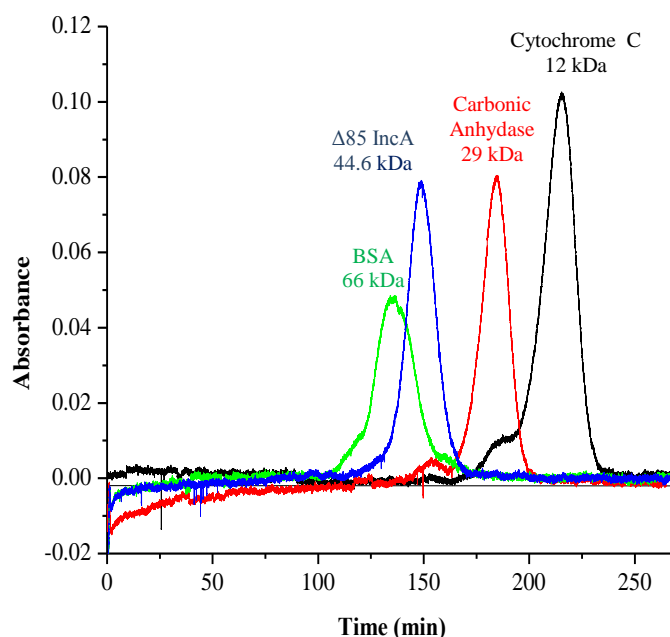
In order to ascertain the oligomeric state of  $\Delta 85\text{ IncA}$ ,  $\Delta 85\text{ IncA141C}$ , and  $\Delta 85\text{ IncA252C}$  mutants, samples were applied to a preparative gel filtration size exclusion chromatography. The Sephadex S200 column was calibrated with molecular standards cytochrome c (12 kDa), carbonic anhydrase (29 kDa), and bovine serum albumin (66 kDa).

$\Delta 85$  IncA presented an aberrant elution pattern (**Figure 4.2**). Both runs generated two peaks. The first peak, representative of the dimer fraction, eluted before BSA. The first peak should have eluted between BSA and carbonic anhydrase. The second peak, representative of monomer, should have eluted between carbonic anhydrase and cytochrome C. The SEC data is consistent with earlier reports with full length wild-type IncA.<sup>39</sup> In that study, the elution shift was attributed to detergents used to mimic the native environment of IncA. Due to the significant elution shift, accurate molecular weights of the  $\Delta 85$  IncA could not be calculated. Surprisingly, the fractions in each run show two appreciable concentration differences in monomer and dimer populations, suggesting that the two oligomeric states exist in dynamic equilibrium. Perron-Savard et.al., found similar behavior with transport protein PhoP found in salmonella enterica.<sup>75</sup>



**Figure 4.2.** Initial SEC elution profile  $\Delta 85$  IncA. Three standards were applied to a Sephadex S200 column: bovine serum albumin (66 kDa), carbonic anhydrase (29 kDa), and cytochrome c (12 kDa).  $\Delta 85$  IncA shows an aberrant elution profile consistent with previous studies.<sup>40</sup>

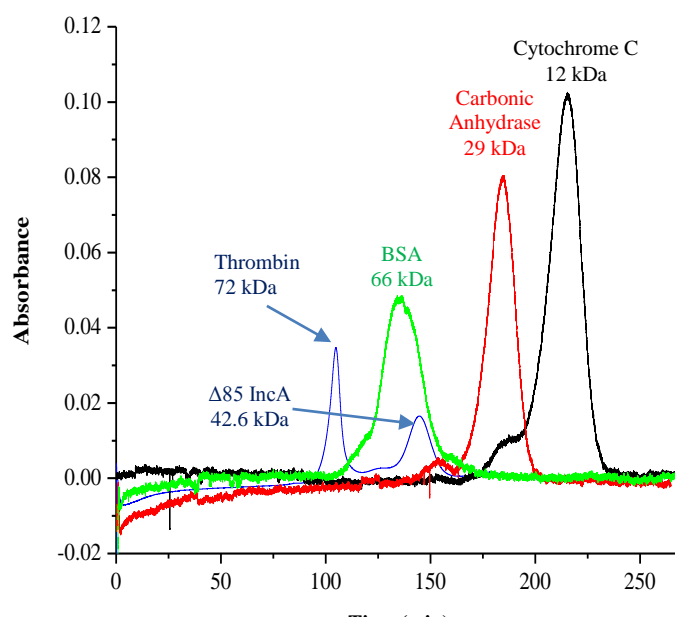




**Figure 4.3.** SEC chromatogram of  $\Delta 85$  IncA with standards. Dimer and monomer populations of  $\Delta 85$  IncA were separated. Concentrated dimer population re-introduced to the SEC column with standards Cytochrome c (12 kDa), carbonic anhydrase (29 kDa), and bovine serum albumin (66 kDa) showed a single peak consistent with dimer formation. Standard concentrations were 4 mg/mL while the  $\Delta 85$  IncA was 9.2 mg/mL (46.2 mg) and were prepared in dialysis buffer (1 mM TCEP, 75 mM Tris, 150 mM NaCl at pH 7.8).

The Perron-Savard group attributed the two populations to the C-terminal his-tag used to purify the sample. In this case, when the monomer fraction was removed and the fraction consistent with dimer was concentrated and introduced to the SEC column a second time, soluble  $\Delta 85$  IncA dimer fraction eluted consistent with the purity and elution pattern of the molecular standards (**Figure 4.3**). A large peak eluted very late in the run and was attributed to his-tag degradation when SDS-PAGE electrophoresis revealed no protein in the corresponding fractions. In order to streamline purification and generate a pure, stable dimer population with one SEC run, the his-tag was cleaved with thrombin. His-tag removal produced two well resolved elution peaks corresponding to the molecular weights

of thrombin and soluble  $\Delta 85$  IncA dimer, respectively (**Figure 4.4**). In the case of PhoP, removal of the affinity tag promoted dimer formation.<sup>75</sup> Additionally, affinity tag removal from the N-terminal domain of Platelet Activating Factor proteins have facilitated heterodimer formation.<sup>76</sup> Without the his-tag, the  $\Delta 85$  IncA dimer is 42.6 kDa, and the elution pattern was consistent with expectations (**Figure 4.4**).<sup>74</sup> SEC data presented here suggest specificity in structural alignment exists to promote dimer formation and the monomer population is produced by interference by the flexible his-tag.

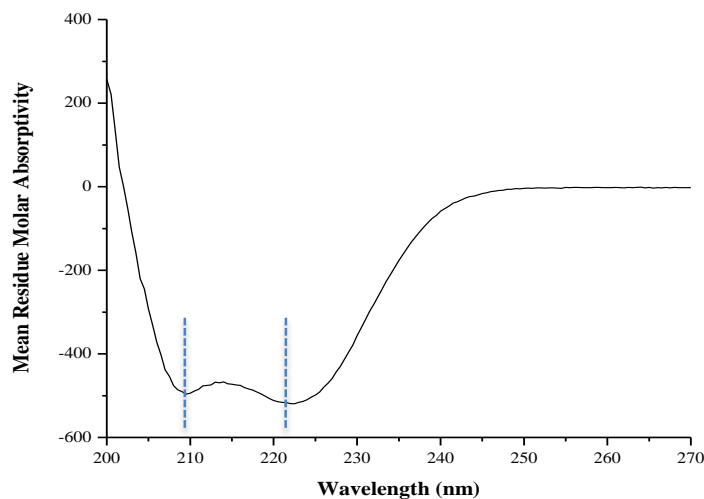


**Figure 4.4.** SEC chromatogram of thrombin cleaved  $\Delta 85$  IncA with standards. Cleaved  $\Delta 85$  IncA (21.3 kDa) and thrombin (72 kDa) is consistent with expectations eluting earlier than  $\Delta 85$  IncA. Cleaved  $\Delta 85$  IncA (21.3 kDa) and thrombin (72 kDa) are compared with standard concentrations cytochrome c 12 kDa, carbonic anhydrase 29 kDa, and bovine serum albumin (66 kDa). Standard concentration were 4 mg/mL and  $\Delta 85$  IncA sample was 5.9 mg/mL (14.7 mg) and were prepared in dialysis buffer (1 mM TCEP, 75 mM Tris, 150 mM NaCl at pH 7.8).

His-tag removal facilitates dimer formation for  $\Delta 85$  IncA but was not without consequence. The smaller  $\Delta 85$  IncA peak is due to a smaller protein concentration (**Figure 4.4**) introduced to the column (14.7 mg). Furthermore,  $\Delta 85$  IncA and  $\Delta 85$  IncA252C remained stable in dilute fractions, however, the  $\Delta 85$  IncA141C mutant began precipitating between 24 and 48 hours.

#### 4.4 CircularDichroism

In order to assess the secondary fold of  $\Delta 85$  IncA relative to the wild type sequence, CD spectroscopy was used to characterize the construct's  $\alpha$ -helical character. CD was performed on  $\Delta 85$  IncA with the affinity tag to provide a baseline for other constructs. A 1 mm path length generated a spectra with minima near 208 and 222 nm<sup>77</sup> revealing classic  $\alpha$ -helical signature consistent with CD spectrum of GCN4 (**Figure 4.5**).

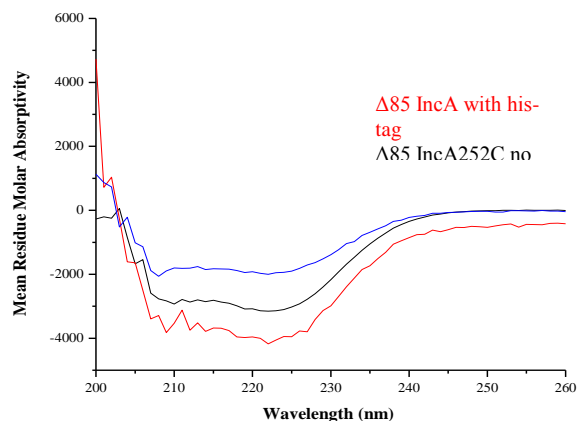


**Figure 4.5.** *CD of  $\Delta 85$  IncA.* Representative data of  $\Delta 85$  IncA conducted with 4 replicates, a 1 mm path length, and a wavelength step of 0.05 nm. The plot is consistent with a characteristic  $\alpha$ -helical signature with absorption minimums at 208 and 222 nm.

CD was also used in order to assess the secondary fold on the mutants. Although characteristic minima are observed at 208 and 210 nm, cleaved mutants  $\Delta 85$  IncA252C and  $\Delta 85$  IncA141C present more shallow spectra than  $\Delta 85$  IncA (**Figure 4.6**). The concentration for  $\Delta 85$  IncA (with tag),  $\Delta 85$  IncA252C, and  $\Delta 85$  IncA141C were 0.45 mg/mL, 0.10 mg/mL, and 0.06 mg/mL, respectively. Formation of a stable coiled-coil structure may prove to be essential for IncA proteins. In order to obtain a definitive conclusion about the effects of the his-tag on secondary structure of  $\Delta 85$  IncA, the spectra for each construct with and without the his-tag should be compared.

#### 4.5 MALDI-TOF Mass Spectrometry

In order to assess spin label efficiency, MALDI-TOF MS was performed on  $\Delta 85$  IncA141C and counterpart  $\Delta 85$  IncA141R1 as well as  $\Delta 85$  IncA252C and its counterpart



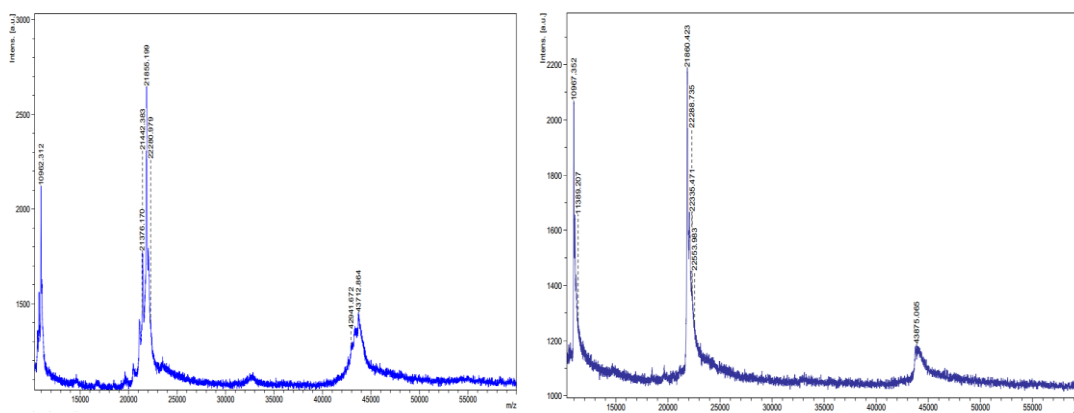
**Figure 4.6.** CD of  $\Delta 85$  IncA with  $\Delta 85$  IncA mutants. Representative data of  $\Delta 85$  IncA with his-tag cleaved  $\Delta 85$  IncA252C and  $\Delta 85$  IncA141C mutants conducted with 3 replicates, a 2 mm path length, and a wavelength of 0.05 mm. The plot is consistent with  $\alpha$ -helical signatures. Concentrations of  $\Delta 85$  IncA,  $\Delta 85$  IncA252C, and  $\Delta 85$  IncA141C were 0.45 mg/mL 0.10 mg/mL and 0.06 mg/mL, respectively.

$\Delta 85$  IncA252R1.  $\Delta 85$  IncA141R1 represents a cysteine label at position 141 and a cysteine mutated to an alanine at position 252.  $\Delta 85$  IncA252R1 represents a cysteine label at position 252 with a cysteine mutated to an alanine at position 141. Mass error for MALDI-TOF MS is 0.01%.<sup>62</sup> The theoretical molecular weight for each cleaved construct is 21.6 kDa ( $\pm 0.2$ ) while the molecular weight for each spin-labeled, cleaved construct is 21.8 kDa ( $\pm 0.2$ ). MALDI-TOF MS shows the molar mass for  $\Delta 85$  IncA141C and  $\Delta 85$ IncA141R1 to be 21.4 and 21.8 kDa, respectively (**Figure 4.7**). The molar mass of the matrix is 224 g/mol and the mass discrepancy may due to sinapinic acid adducts. The area underneath the curve for  $\Delta 85$  IncA141C was 143,731 and the area underneath the curve for 141-R1 was 201,785. Using the area corresponding to each mass the spin label efficiency was calculated to be 87.2% for  $\Delta 85$  IncA141R1 and 34.9% for  $\Delta 85$  IncA252R1 using **Equation 4.1**.

$$SLE = \frac{\text{Area of } \Delta 85 \text{ IncA141-R1}}{\text{Area of } \Delta 85 \text{ IncA141C} + \text{Area of } \Delta 85 \text{ IncA141-R1}} \times 100 \quad \text{Equation 4.1}$$

#### 4.6 CW EPR & Site-Directed Spin-Labeling of $\Delta 85$ IncA252C and $\Delta 85$ IncA141C

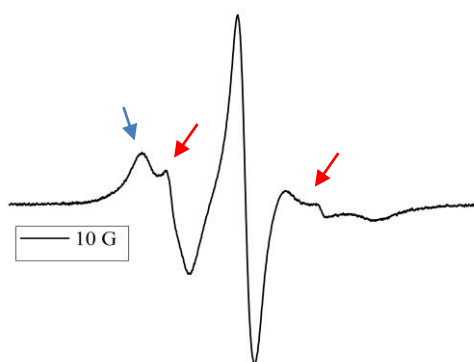
PIPE mutagenesis was used to selectively remove one the native cysteine residues located at positions 141 and 252 on the primary sequence of  $\Delta 85$  IncA.  $\Delta 85$  IncA141R1 represents a cysteine label at position 141 and a cysteine mutated to an alanine at position 252.  $\Delta 85$  IncA252R1 represents a cysteine label at position 252 with a cysteine mutated to an alanine at position 141. Both mutants have the his-tag removed. Methanethiosulfonate (MTSSL) successfully bound to the protein as shown by the generation of “Lorentzian-like” line shape (**Figure 4.8 and Figure 4.9**).



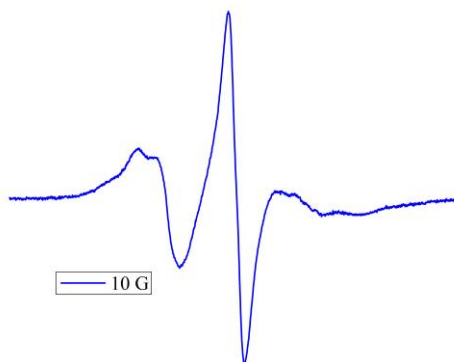
**Figure 4.7.** MALDI-TOF MS of  $\Delta 85$  IncA141R1 &  $\Delta 85$  IncA252R1. Left: The  $\Delta 85$  IncA141R1 sample was co-spotted with sinapinic acid, allowed to dry, and ablated with 200 laser shots. To assess spin label efficiency, the area underneath the +1 charge state most closely corresponding to the construct's molecular weight (21,442.4) was compared to the +1 charge state most closely corresponding with the MTSSL labeled construct (21,855.2). Spin label efficiency was calculated to be 87%. Right:  $\Delta 85$  IncA 252R1 was co-spotted with sinapinic acid, allowed to dry, and ablated with 200 laser shots. To assess spin label efficiency, the area underneath the +1 charge most closely corresponding to the constructs molecular weight (21, 860.2) was compared to the +1 state most closely corresponding with the MTSSL labeled construct (22, 062.4). Spin label efficiency was calculated to be 34.9%.

EPR lineshapes are effected by three types of motion (1) motion of the spin label relative to the protein backbone, (2) movement of the bonds that attach MTSSL to the protein and, (3) tumbling of protein in solution.<sup>78</sup> Comparing the spectra to previous characterized proteins, the outer extrema on both  $\Delta 85$  IncA141R1 and  $\Delta 85$  IncA252R1 spectra indicate a moderately immobilized spin label.<sup>69, 73</sup> The absence of line broadening and the shape of the outer extrema are consistent with system with at least two components 25Å apart.<sup>70</sup> Exposure to other secondary or tertiary structure (i.e., coiled-coils) could explain also the immobility of the spin label. EPR spectra analysis also suggest  $\Delta 85$  IncA is tumbling in solution and this data is consistent with the SEC data. The shape of the outer extrema and the central peak shape data supports dimer formation. Little discernable differences

between the  $\Delta 85$  IncA141R1 and the  $\Delta 85$  IncA252R1 lineshapes, however, their spectra alone cannot elucidate the structure of the dimer. Introducing the sample to the PD-10 desalting column or washing the sample could remove the excess spin-label indicators. EPR analysis of the protein with and without the his-tag would help elucidate if the tag has significant structural implications.



**Figure 4.8.** X-Band (9 GHz) EPR spectra of 141-R1. Methanethiosulfonate (MTSL) was added 10 times excess to the  $\Delta 85$  IncA141C mutant and incubated overnight at ambient temperature. Excess spin label was removed by eluting twice through a PD-10 desalting column. The Lorentzian shape indicates MTSL successfully bound to  $\Delta 85$  IncA141. The spectra suggests an moderately immobilized R1 spin label and **excess spin label**. Absences of line broadening and **outer extrema** are consistent with at least a two component system 25 Å apart.



**Figure 4.9.** X-Band (9 GHz) EPR Spectra of  $\Delta 85$ IncA252-R1. Methanethiosulfonate (MTSSL) was added 5 times excess to the  $\Delta 85$  IncA<sup>141</sup> mutant and incubated 24 hours at ambient temperature. The sample was introduced to a PD-10 desalting column and washed 4 times with labeling buffer (75 mM Tris, 150 mM NaCl, pH 7.8). The Lorentzian-like line shape indicates MTSSL successfully bound to  $\Delta 85$  IncA252C. The spectra suggests a single component spin label system with that is moderately immobilized. Lack of line broadening and outer extrema is consistent with at least two components 25 Å apart.

#### 4.6.1 Conclusions

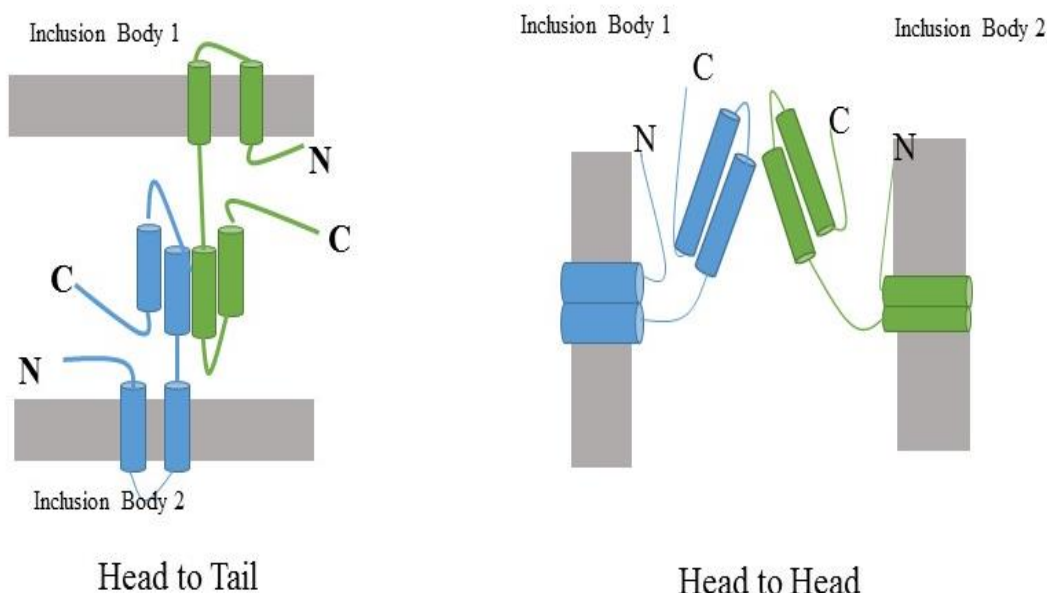
Four truncated mutants of the IncA protein have been successfully cloned, expressed, and purified. SDS-PAGE data following IMAC shows  $\Delta 85$  IncA and mutants lack the denaturation resistance property of full-length wild-type IncA. These data suggests that the hydrophobic region confers protein fold stability for the C-terminal domain that faces the host cytoplasm. SEC data indicate oligomers exist in dynamic equilibrium and his-tag cleavage on the N-terminal domain facilitates dimer formation. SEC results also suggest that removal of the his-tag creates solubility issues with the  $\Delta 85$  IncA141C mutant. Circular dichroism of affinity “tag-less” mutants ( $\Delta 85$  IncA141C and  $\Delta 85$  IncA252C) presented minima more shallow relative to  $\Delta 85$  IncA with the affinity tag. Concentration differences or the removal of the his-tag may contribute to these differences. Therefore, CD analysis with and without the his-tag should be repeated on all IncA mutants to confirm these results and ascertain the extent of the loss of  $\alpha$ -helical character, if any. MALDI-TOF MS indicates spin labeling efficiency in this investigation was 87.2% for  $\Delta 85$  IncA141C mutant and 34.9% for the  $\Delta 85$  IncA252C mutant. Changes in the spin label protocol may improve these percentages as well as enhance spectra resolution. EPR spectra show strongly immobilized unpaired electrons consistent with a two component system, 25Å apart. The spectra could also indicate tumbling proteins in solution as well as spin-label interaction with another structure. Preliminary biophysical studies have been initiated. Additional evidence for the oligomeric state of IncA can be obtained by conducting distance measurements using double electron-electron resonance (DEER) as well as generating a model using bioinformatic tools to assess the viability of distance constraints.



When Stanislaus von Prowazek first discovered *C. trachomatis*, he probably never imagined that the bacteria would continue to cause problems into the 21<sup>st</sup> century. Scientists find inclusion protein A from *C. trachomatis* an important subject to study due to its similarity with SNARE proteins, however, understanding IncA proteins may serve as a guideline to develop targeted drug delivery systems because of the fusion events they facilitate as well as their specificity. Hopefully, scientists will understand the pathogenicity of *C. trachomatis* long before the next century begins.

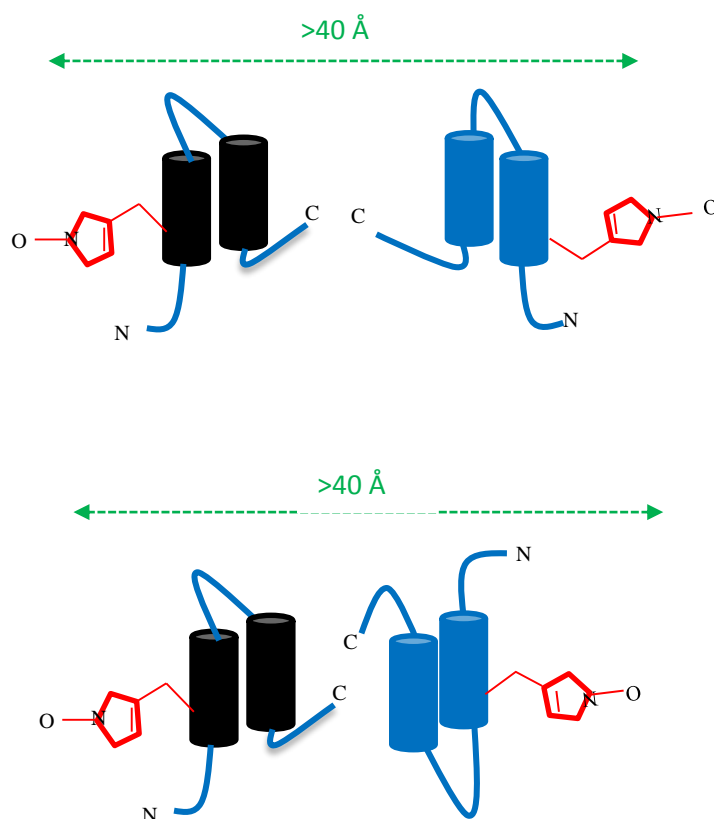
#### 4.7 Future Directions

Additional EPR studies are necessary to confirm the orientation of the IncA dimer (**Figure 4.10**). First,  $\Delta 85$  IncA should be labeled using the two native cysteine residues. Literature suggests dimer orientation is head to head or head to tail with the C-terminal



**Figure 4.10** Literature proposed orientations of the IncA dimer. **Left:** The conformer has the N-terminal regions diagonal to each other while the C-terminal domain faces the host cytoplasm. This orientation is called head to tail dimerization. **Right:** After dimerization, the N-terminal domains are localized on the same plane while the C-terminal region faces the host cytoplasm. This orientation is called head to head dimerization.

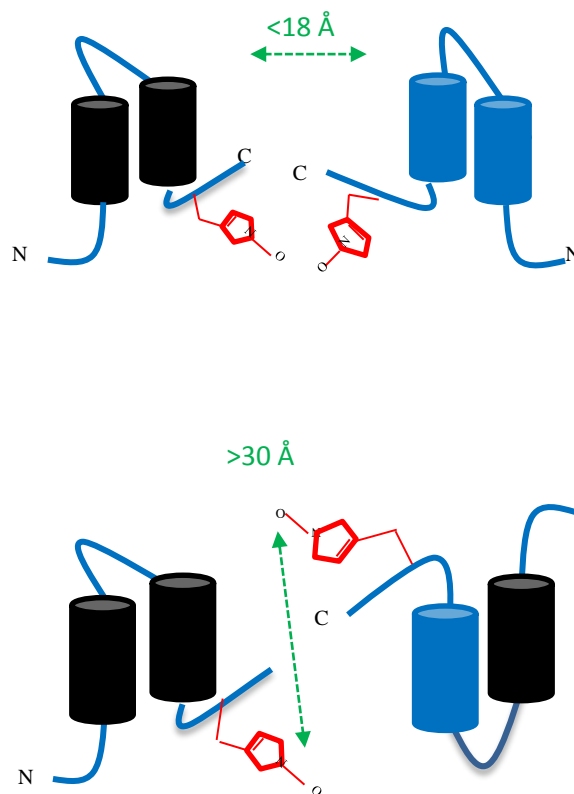
regions opposite each other. CW EPR data here suggest that  $\Delta 85$  IncA141-R1 has its label on the  $\alpha$ -helix surface facing the solvent and the  $\Delta 85$ IncA252-R1 spin label (located on the unstructured region of the C-terminal domain) has tertiary contact.



**Figure 4.11.** Possible DEER experiment with  $\Delta 85$  IncA141-R1. Experiments with DEER should be consistent with distances greater than 40 Å apart if  $\Delta 85$  IncA dimerizes in the head to head (top) or head to tail (bottom) configuration.

EPR spectroscopy can be used deduce the dimer orientation by measuring the distance between the two spin labels across the dimer interface. DEER spectroscopy can measure distances greater than 18 Å and less than 60 Å.<sup>79</sup> If the distance between the two spin labels is less than 18 Å the distances could not be measured with DEER spectroscopy but could be measured with CW approaches.

The diameter of a right-handed  $\alpha$ -helix is 12 Å. If IncA dimerizes in the head to head or head to tail orientation with juxtaposed C-terminal regions, four alpha helices should separate the spin labels by ~48 Å (**Figure 4.11**). DEER experiments on  $\Delta 85$  IncA141-R1 dimer should reveal a distance greater than 40 Å between the two spin labels. In the same scenario, if  $\Delta 85$  IncA252R1 dimerizes in the head to head configuration, a DEER signal should not be observed because the two labels should be less than 18 Å apart (**Figure 4.12**).



**Figure 4.12.** Possible DEER experiment with  $\Delta 85$  IncA252R1. Top: Experiments with DEER measurements with  $\Delta 85$  IncA252R1 should be consistent with distances less than 18 Å apart if IncA dimerizes in the head to head configuration and no signal should be observed. Bottom: DEER measurements with  $\Delta 85$  IncA252R1 should be greater than 30 Å and less than 40 Å apart if  $\Delta 85$  IncA dimerizes in the head to tail configuration.

If  $\Delta 85$  IncA252R1 dimerizes in the head to tail configuration, a DEER signal should indicate the two labels are 30 – 40 Å apart. Additional cysteine mutants should be designed based on these distance measurements to further investigate the dimeric structure of IncA.

## References

1. Control, C. o. D., Incidence, Prevalence, and Cost of Sexually Transmitted Infections in the United States. National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention: 2013; Vol. United States Government pp 1 - 4.
2. Wilson, D. J., Insights from Genomics into Bacterial Pathogen Populations *Plos Pathogens* **2012**, 8, 1 - 9.
3. Scidmore-Carlson, M. A.; Shaw, E. I.; Dooley, C. A.; Fischer, E. R.; Hackstadt, T., Identification and characterization of a *Chlamydia trachomatis* early operon encoding four novel inclusion membrane proteins *Molecular Microbiology* **1999**, 33, 753-765.
4. Pospischil, A., From disease to etiology: historical aspects of *Chlamydia*-related diseases in animals and humans. *Drugs of today*, **2009**, 45, 141.
5. Jaenicke, L., Stanislaus von Prowazek - Prodigy between Working Bench and Coffee House *Protist* **2001**, 152, 157-166.
6. Whitaker, J. R., Merriam-Webster Unabridged Dictionary. Merriam-Webster, Incorporated: Springfield, 2014. Accessed 5 May 2013.
7. Ronzone, E.; Paumet, F., Two Coiled-Coil Domains of *Chlamydia trachomatis* IncA Affect Membrane Fusion Events during Infection. *Plos One* **2013**, 8, 1 - 11.
8. Control, C. f. D., CDS STD facts - chlamydia 2012.
9. Dautry-Varsat, A.; Subtil, A.; Hackstadt, T., Recent insights into the mechanisms of Chlamydia entry. *Cellular Microbiology* **2005**, 7, 1714 - 1722.
10. Gottlieb, S. L.; Brunham, R. C.; Byrne, G. I.; Martin, D. H.; Xu, F.; Berman, S. M., Introduction: The Natural History and Immunobiology of *Chlamydia trachomatis* Genital Infection and Implications for Chlamydia Control. *The Journal of Infectious Diseases* **2010**, 201, 2.
11. Bannantine, J. P.; Griffiths, R. S.; Viratyosin, W.; Brown, W. J.; Rockey, D. D., A secondary structure motif predictive of protein localization to the chlamydial inclusion membrane *Cellular Microbiology* **2000**, 2, 35-47.
12. Hackstadt, T.; Fischer, E. R.; Scidmore, M. A.; Rockey, D. D.; Heinzen, R. A., Origins and functions of the *chlamydial* inclusion. *Trends in Microbiology* **1997**, 5, 288-293.
13. Stephens, R. S.; Myers, G.; Eppinger, M.; Bavoli, P. M., Divergence without difference: phylogenetics and taxonomy of *Chlamydia* resolved *FEMS Immunol. Med. Microbiology* **2009** 55, 115-119.
14. Roulis, E.; Polkinghorne, A.; Timms, P., *Chlamydia pneumoniae*: modern insights into an ancient pathogen *Trends in Microbiology* **2013**, 21, 120 - 128.
15. Everett, K. D. E.; Bush, R. M.; Anderson, A. A., Emended description of the order Chlamydiales, proposal of Parachlamydiaceae fam. nov. and Simkaniaceae fam. nov., each containing one monotypic genus, revised taxonomy of the family Chlamydiaceae, including a new genus and five new species, and standard for the identification of organisms *International Journal of Systematic Bacteriology* **1999** 49, 415-440.
16. Rockey, D. D.; Scidmore, M. A.; Bannantine, J. P.; Brown, W. J., Proteins in the chlamydial inclusion membrane. *Microbes and Infection* **2002**, 4, 333-340.

17. Stephens, R. S.; Kalman, S.; Lammel, C.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.; Olinger, L.; Tatusov, R. L.; Zhao, Q.; Koonin, E. V.; Davis, R. W., Genome Sequence of an Obligate Intracellular Pathogen of Humans: *Chlamydia trachomatis*. *Science* **1998**, 282, 754-760.
18. Suchland, R. J.; Jeffrey, B. M.; Bhatia, M. X. A.; Chu, H. G.; Rockey, D. D.; Stamm, W. E., Identification of Concomitant Infection with *Chlamydia trachomatis* IncA-Negative Mutant and Wild-Type Strains by Genomic Transcriptions, and Biological Characterizations *Infection and Immunity* **2008**, 76, 5438-5446.
19. Subtil, A.; Delevoye, C.; Balana, M.-E.; Tastevin, L.; Perrine, S.; Dautry-Varsat, A., A directed screen for chlamydial proteins secreted by a type III mechanism identifies a translocated protein and numerous other new candidates *Molecular Microbiology* **2005** 56, 1636-1647.
20. Peters, J.; Wilson, D. P.; Myers, G.; Timms, P.; Bavoil, P. M., Type III secretion in *Chlamydia*. *TRENDS in Microbiology* **2007**, 15, 241-251.
21. Lorenzini, E.; Singer, A.; Singh, B.; Lam, R.; Skarina, T.; Chirgadze, N. Y.; Savchenko, A.; Gupta, R. S., Structure and Protein-Protein Interaction Studies on *Chlamydia trachomatis* Protein CT670 (YscO Homolog) *Journal of Bacteriology* **2010** 192, 2746-2756.
22. Elkadi, O. A., MDR-selective microbial-based therapy: A novel approach to cancer treatment *Medical Hypotheses* **2013**, 81, 207-211.
23. Lutter, E. I.; Martens, C.; Hackstadt, T., Evolution and Conservation of Predicted Inclusion Membrane Proteins in Chlamydiae *Comparative and Functional Genomics* **2011**, 2012, 1 - 13.
24. Li, Z.; Chen, C.; Chen, D.; Wu, Y.; Zhong, Y.; Zhong, G., Characterization of Fifty Putative Inclusion Membrane Proteins Encoded in the *Chlamydia trachomatis* Genome *Infection and Immunity* **2008** 76, 2746-2757.
25. Zhong, G., Killing me softly: chlamydial use of proteolysis for evading host defenses *Trends in Microbiology* **2009**, 17, 467-474.
26. Saka, H. A.; Valdivia, R. H., Acquisition of nutrients by *Chlamydiae*: unique challenges of living in an intracellular compartment *Current Opinion in Microbiology* **2010** 13, 4 - 10.
27. Delevoye, C.; Nilges, M.; Dautry-Versat, A.; Subtil, A., Conservation of the Biochemical Properties of IncA from *Chlamydia trachomatis* and *Chlamydia caviae* *The Journal of Biological Chemistry* **2004**, 279, 46896-46906.
28. Kostyukova, E. S.; Lazarev, V. N.; Govorum, V. M., Inclusion Membrane Proteins of Chlamydiae. *Biomeditsinskaya Khimiya* **2008** 2, 148-159.
29. Hackstadt, T.; Scidmore-Carleson, M. A.; Shaw, E. I.; Fisher, E. R., The *Chlamydia trachomatis* IncA protein is required for homotypic vesicle fusion *Cellular Microbiology* **1999** 1, 119-130.
30. Rzomp, K. A.; Moohead, A. R.; Scidmore, M. A., The GTPase Rab4 Interacts with *Chlamydia trachomatis* Inclusion Membrane Protein CT229. *Infection and Immunity* **2006** 74, 5362-5373.

31. Rockey, D. D.; Rosquist, J. L., Protein Antigens of *Chlamydia psittaci* Present in Infect Cells but Not Detected in the Infectious Elementary Body *Infection and Immunity* **1994** 62, 106-112.
32. Suchland, R. J.; Rockey, D. D.; Bannantine, J. P.; Stamm, W. E., Isolates of *Chlamydia trachomatis* That Occupy Nonfusogenic Inclusions Lack IncA, A Protein Localized to the Inclusion Membrane *Infection and Immunity* **2000**, 68, 360-367.
33. Rockey, D. D.; Viratyosin, W.; Bannantine, J. P.; Suchland, R. J.; Stamm, W. E., Diversity within inc genes of clinical *Chlamydia trachomatis* variant isolates that occupy non-fusogenic inclusions. *Microbiology* **2002**, 148, 2497-2502.
34. Geisler, W. M.; Suchland, R. J.; Rockey, D. D.; Stamm, W. E., Epidemiology and Clinical Manifestations of Unique *Chlamydia trachomatis* Isolates That Occupy Nonfusogenic Inclusions. *Journal of Infectious Disease* **2001**, 184 879-884.
35. Boron, W. F.; Boulpaep, E. L., *Medical Physiology: A Cellular and Molecular Approach* 2nd ed.; Elsevier Philadelphia 2012.
36. Delevoye, C.; Nilges, M.; Dehaoux, P.; Paumet, F.; Perrinet, S.; Dautry-Versat, A.; Subtil, A., SNARE Protein Mimicry by an Intracellular Bacterium *Plos Pathogens* **2008** 4, 1 - 14.
37. Paumet, F.; Wesolowski, J.; Garcia-Diaz, A.; Delevoye, C.; Aulner, N.; Shuman, H. A.; Subtil, A.; Rothman, J. E., Intracellular Bacteria Encode Inhibitory SNARE-like Proteins *Plos One* **2009** 4, 1- 9.
38. Moore, E. R.; Mead, D. J.; Dooley, C. A.; Sager, J.; Hackstadt, T., The trans-Golgi SNARE syntaxin 6 is recruited to the chlamydial inclusion membrane *Microbiology* **2011**, 157, 830-838.
39. Keller, A. Structural Investigations of Inclusion Membrane Protein A (IncA) of *Chlamydia Trachomatis*, University of Virginia Charlottesville, **2011**.
40. Garrett, R. H.; Grisham, C. M., *Recombinant DNA: Cloning and Creation of Chimeric Gene*. Brooks/Cole Cengage Learning Boston, MA 2010.
41. Verenini, J. Joint Center for Structural Genomics Developing HT Methods for Gene to Structure and Function <http://www.jcsg.org/prod/newscripts/about.cgi>.
42. Elsliger, M.-A.; Deacon, A. M.; Godzik, A.; Lesley, S. A.; Wooley, J.; Wuthrich, K.; Wilson, I. A., "The JCSG high-throughput structural biology pipeline." *Acta Crystallographica Section F* **2010** F66, 1137-1142.
43. Klock, H. E.; Lesley, S. A., *The Polymerase Incomplete Primer Extension (PIPE) Method Applied to High-Throughput Cloning and Site-Directed Mutagenesis*. Springer Science + Business Media: **2009**; Vol. 498.
44. Raymond, A.; Haffner, T.; Ng, N.; Lorimer, D.; Staker, B.; Stewart, L., Gene design, cloning and protein-expression methods for high-value targets at the Seattle Structural Genomics Center for Infectious Disease *Acta Crystallographica Section F* **2011**, F67 992-995.
45. Klock, H. E.; Loesema, E. J.; Knuth, M. W.; Lesley, S. A., Combining the polymerase incomplete primer extension method for cloning and mutagenesis with microscreening to accelerate structural genomics efforts *Proteins Structure Function Bioinformatics* **2007**, 982-994.

46. Brondyk, W. H., *Selecting An Appropriate Method for Expressing Recombinant Protein*. Elsevier: Framingham, **2007**; Vol. 463, p 17.
47. Kane, J., Effects of rare codon clusters on high-level expression of heterologous proteins in *Escherichia coli*. *Current Opinion in Biotechnology* **1995**, 6, 494-500.
48. Carstens, C.-P.; Bonnaardel, J.; Allen, R.; Waesche, A., BL21-Codon Pluss Cells Correct Expression Problems Caused by Codon Bias Strategene, Ed.
49. Tabor, S.; Richardson, C. C., A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes *Proceedings of the National Academy of Sciences* **1985**, 82, 5.
50. Studier, W.; Moffatt, B., Use of Bacteriophage T7 RNA polymerase to Direct Selective Highlevel Expression of Cloned Genes *Journal of Molecular Biology* **1986** 189, 18.
51. Sorensen, H. P.; Sorensen, K. K., Advanced genetic strategies for recombinant protein expression in *Escherichia coli* *Journal of Biotechnology* **2005**, 115, 18.
52. Porath, J.; Olin, B., Immobilized Metal Ion Affinity Adsorption and Immobilized Metal Ion Affinit Chromatography of Biomaterials. Serum Protein Affinities for Gel-Immobilized Iron and Nickel Ions *Biochemistry* **1983**, 22, 10.
53. Steinert, K.; Wulbeck, M.; Ribbe, J., Ni-NTA resins--you key to efficient purification of 6XHis-tagged proteins GmbH, Q., Ed. Hilden, Germany.
54. Recombinant Protein Purification Principles and Methods In *General Electric Company*, AB, G. E. H. B.-s., Ed. General Electric Company Uppsala **2000**; Vol. 1142-75, pp 115-119.
55. Gallwitz, M.; Enoksson, M.; Thorpe, M.; Hellman, L., The Extended Cleavage Specificity of Human Thrombin *Plos One* **2012**, 7, 16.
56. Eisenstein, M., A look back: adventures in the matrix *Nature Methods* **2006**, 3 (5), 410.
57. Whitaker, J. R., Determination of Molecular Weights of Proteins by Gel Filtration on Sephadex *Analytical Chemistry* **1963**, 35, 1950-1953.
58. Sun, T., A study of the Separation Principle in Size Exclusion Chromatography *Macromolecules* **2004** 37, 9.
59. Wang, Y.; Teraoka, I.; Hansen, F. Y.; Peters, G. H.; Hassager, O., A Theoretical Study of the Separation Principle in Size Exclusion Chromatography *Macromolecules* **2010** 43, 1651-1659.
60. Christian, G. D., *Analytical Chemistry* 5th ed.; John Wiley & Sons New York, **1994**, p12.
61. Domon, B.; Aebersold, R., Mass Spectrometry and Protein Analysis, *Science* **2006**, 312, 212-217.
62. Bonk, T.; Humeny, A., MALDI-TOF-MS Analysis of Protein and DNA *Neuroscientist* **2001**, 7, 6-12.
63. Caprioli, R. M.; Farmer, T. B.; Gile, J., Molecular Imaging of Biological Samples: Localization of Peptides and Proteins Using MALDI-TOF MS *Analytical Chemistry* **1997**, 69, 4751-4760.



64. Counterman, A. E.; Thompson, M. S.; Clemmer, D. E., Identifying a Protein by MALDI-TOF Mass Spectrometry: An Experiment for the Undergraduate Laboratory *Journal of Chemical Education* **2003**, 80, 177-180.
65. Kelly, S. M.; Jess, T. J.; Price, N. C., How to study proteins by circular dichroism *Biochimica et Biophysica Acta* **2005** 1751, 119-150.
66. Greenfield, N. J., Using circular dichroism spectra to estimate protein secondary structure *National Protoco.* **2006** 1, 2876-2890.
67. Eaton, G. R.; Eaton, S. S.; Salikhov, K. M., *Foundations of Modern EPR* World Scientific Publishing Co. Pte. Ltd. : River Edge **1998**.
68. Eaton, S. S.; Eaton, G. R., *Biological Magnetic Resonance* Kluwer Academic/Plenum Publishers New York 2000; Vol. 19.
69. Humphries, G. M. K.; McConnell, H. M., *Nitroxide Spin Labels* Academic Press, Inc New York, **1982**, Vol. 20, p 564.
70. Swartz, H. M.; Swartz, S. M., Biochemical and Biophysical Applications of Electron Spin Resonance *Methods of Biochemical Analysis* **1983**, 29, 117.
71. Salikhov, K., Modern Development of Magnetic Resonance. *Applied Magnetic Resonance* **2009** 35, 361-362.
72. Weber, R. T., EMX User's Manual 2.0 ed.; Bruker Instruments, I., Ed. EPR Division Billerica 1998.
73. Kroncke, B. M.; Horanyi, P. S.; Columbus, L., Structural Origins of Nitroxide Side Chain Dynamics on Membrane Protein  $\alpha$ -Helical Sites. *Biochemistry* **2010**, 49, 10045-10060.
74. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Bairoch, A. *Protein Identification and Analysis Tools on the ExPASy Server* Humana Press **2005**.
75. Perron-Savard, P.; Crescenzo, G. D.; Mauual, H. L., Dimerization and DNA binding of the *Salmonella enterica* PhoP response regulator are phosphorylation independent *Microbiology* **2005**, 151, 3979-3987.
76. Arachiche, A.; Mumaw, M. M.; de la Fuente, M.; Nieman, M. T., Protease-activated receptor 1 (PAR1) and PAR4 heterodimers are required for PAR1-enhanced cleavage of PAR4 by  $\alpha$ -thrombin. *The Journal of Biological Chemistry* **2013**, 288, 32553-32562.
77. Lumb, K. J.; Carr, C. M.; Kim, P. S., Subdomain Folding of the Coiled Coil Leucine Zipper from the bZIP Transcriptional Activator GCN4. *Biochemistry* **1994**, 33, 7361-7367.
78. Miltshetyn, E. Characterizing the Intrinsically Disordered Protein IA3 Using Site Directed Spin Labeling and Electron Paramagnetic Resonance Honor's Thesis University of Florida, Gainesville 2013.
79. Jeschke, G., DEER Distance Measurements on Proteins *Annual Reviews in Physical Chemistry* **2012**, 63, 419-446.