Characterization of Influenza H5N1 Nucleocapsid Protein for Potential Vaccine Design

By: Adam Buffone

A Thesis Submitted to: The Faculty of Graduate and Post Doctoral Studies

University of Ottawa

In Partial Fulfillment of the Requirements for the Degree of:

Master of Science in Biochemistry

Department of Biochemistry, Microbiology and Immunology

Faculty of Medicine

Abstract

Avian influenza H5N1 causes occasional but serious infections in humans and efforts to produce vaccines against this strain continue. Current influenza vaccines are prophylactic and utilize the two major antigens, hemagglutinin and neuraminidase. NP is an attractive alternative antigen because it is highly conserved across all influenza strains, has been shown to increase the rate of viral clearance, and potential therapeutic vaccines would elicit cytotoxic T lymophcyte responses in an infected person. The NP antigen from H5N1 was characterized using a variety of physiochemical methods to gain insights into both the biological and physical properties of the antigen which are important from a regulatory viewpoint when considering therapeutic vaccines. Results obtained to date show that NP is relatively unstable and indicate that the conformation of the H5N1 NP antigen is highly dependent upon purification procedure, buffer conditions, pH and the presence or absence of RNA. These factors will need to be clearly defined and taken into consideration when manufacturing and regulating NP vaccine preparations.

-I dedicate this thesis to my grandmothers Josephine Buffone and Mary Paciocco who left this world at the same time this manuscript was being completed.

Acknowledgments

Completion of my master's thesis in biochemistry has been one of the most challenging yet most rewarding tasks I have ever engaged in. At times, it seemed impossible, but looking back on my journey; I would not have changed a thing. There are many people who have helped me along the way and I am forever grateful.

First and foremost, I would like to thank Dr. Mary Alice Hefford for supervising my research project. Her patience, guidance, support and scientific expertise were instrumental both in and out of the lab. Dr. Hefford hired me to work as a student, under the FSWEP campaign, and after a year of training and instruction she extended to me the opportunity to complete a master's degree. Dr. Hefford set a precedence to aspire to, and I will always remember her eager attitude to teach and conduct research with integrity.

Thank you to the Hefford lab team, especially Dr. John Mark Kim, Sophie Smith and Sylvie Fournier. Their support, knowledge and friendship throughout the last few years were greatly appreciated.

I want to acknowledge the entire CVE division at Health Canada. Managers, research scientists and graduate students throughout CVE were extremely helpful and always full of encouragement. Since beginning this journey in 2009, they accepted me into the CVE family and truly provided me a positive work environment for which I thrived in.

I also express my sincerest gratitude to the members of my TAC. Thank you Dr. Earl Brown and Dr. Natalie Goto, for all of your guidance, support and suggestions throughout this process. I appreciate the time you both took to come out to Tunney's Pasture to sit in on my meetings.

Finally, I would like to thank my family. Mom and Dad, without your love, encouragement and financial support, I would not be where I am today. You provided me with a solid foundation of morals and values by which I am able to build a life upon and for this I am forever grateful. To my siblings, Jessica and Lucas, thank you for the endless support, extensive proof reading and constant reassurance. I would also like to thank my uncle John Paul for his advice and council throughout my academic career.

LIST OF ABBREVIATIONS

ε = molar absorbtivity / extinction coefficient (L mol^{-1} cm^{-1})

A_{260} = Absorbance measured at wavelength 260 nm

A_{280} = Absorbance measured at wavelength 280 nm

APS = Ammonium Persulfate

BCA = Bicinchoninic Acid Protein Assay

c = concentration

CD = Circular Dichroism

CTL = Cytotoxic T Lymphocyte

DEPC = Diethylpyrocarbonate

DLS = Dynamic Light Scattering

FPLC = Fast Protein Liquid Chromatography

HA = Hemagglutinin

HPLC = High Performance Liquid Chromatography

IPTG = Isopropyl β-D-1-thiogalactopyranoside

l = pathlength (cm)

MAC = Metal Affinity Chromatography

MALS = Multi Angle Light Scattering

MBP = Maltose-Binding Protein

MHC I = Major Histocompatibility Complex Class One

NA= Neuraminidase

NEP = Nuclear Export Protein

NP = Nucleocapsid Protein

Nts = Nucleotides

OD = Optical Density

PMSF = Phenylmethanesulfonylfluoride

SDS = Sodium Dodecyl Sulfate

SDS-PAGE = Sodium Dodecyl Sulfate- Polyacrylamide Gel Electrophoresis

SEC = Size Exclusion Chromatography

T_M = Melting Temperature

vRNA = Viral RNA

vRNP = Viral Ribonucleoprotein

VSV = Vesicular Stomatitis Virus

List of Figures

List of Tables

Table of Contents

1.0 Introduction

Influenza, a highly contagious virus of the respiratory tract, threatens the health of people worldwide. It has been the responsible agent in several annual epidemics and sporadic pandemics. According to the World Health Organization (WHO), three to five million cases of severe illness and close to 250-500 thousand deaths occur annually due to influenza outbreaks [1]. The sudden onset of an influenza infection includes signature symptoms such as chills, aches, pain and malaise. In more severe cases, bacterial pneumonia can occur requiring immediate hospitalization and sometimes result in death [2]. Efforts to produce vaccines against influenza continue. Current vaccines against influenza are strain specific and prophylactic, utilizing the two major antigens of the influenza virus, Hemagglutinin (HA) and Neuraminidase (NA) in order to prevent infection. Variability in HA and NA proteins requires yearly identification of the three most common internationally circulating strains and production, distribution and testing of an annual vaccine. The viral Nucleocapsid protein (NP), however, is also an attractive antigen because it is a well conserved internal viral protein with the capacity to elicit a Cytotoxic T Lymphocyte (CTLs) response. Given the universal potential of NP as a vaccine candidate, this thesis focuses on characterizing the NP antigen from an avian strain (H5N1) in order to gain insights into the properties of that antigen that might influence the quality, safety and efficacy of a potential therapeutic vaccine.

1.1. The Influenza Virus

1.1.1. General Viral Structure

Influenza is a member of the *orthomyxoviridae* family and consists of three genera: A, B and C, with A being the best characterized. Of the three genera, A

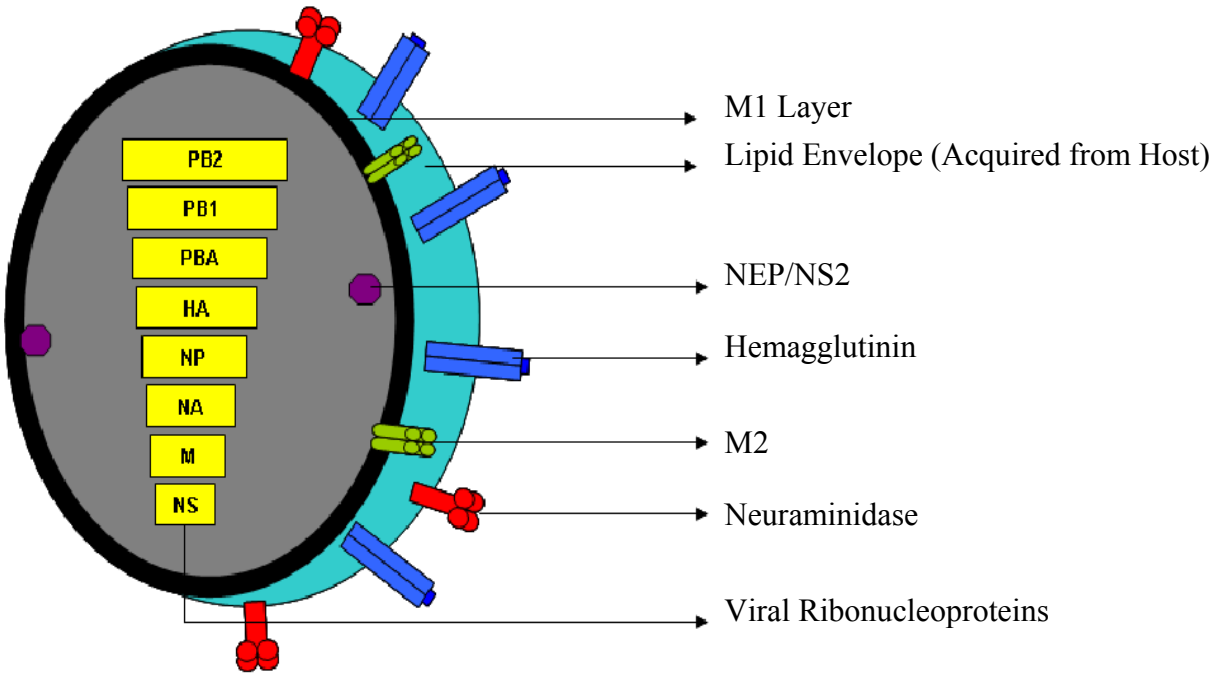
also poses the most serious threat to public health [1] and has been the root cause of all recorded epidemics and pandemics [3]. The global structure of the entire virus particle has been described as pleomorphic [2], in that it has various appearances, existing as either spherically with a diameter of 100nm [4] or as long, filamentous in gross morphology. It has been estimated that the viral particle (or virion) consists of 1% RNA, 5 to 8% carbodydrate, 20% lipid and 70% protein [4]. The influenza virion contains a lipid envelope acquired from the host parental cell upon budding. Protruding from this lipid envelope are three major surface antigens: hemmaglutin (HA) and neuraminidase (NA), which are critical determinants of influenza pathogenicity [5], and the M2 protein ion channel. Underneath the envelope lies the matrix protein (M1) which forms a core that houses the viral ribonucleoproteins (vRNP) [4]. The vRNP, in turn, consists of condensed, segmented genomic viral RNA wrapped around nucleocapsid protein subunits which self-associate into helical nucleocapsids held together by the polymerase complex -- PA, PB1 and PB2 [6], (Figure 1A).

1.1.2. Genome Organization

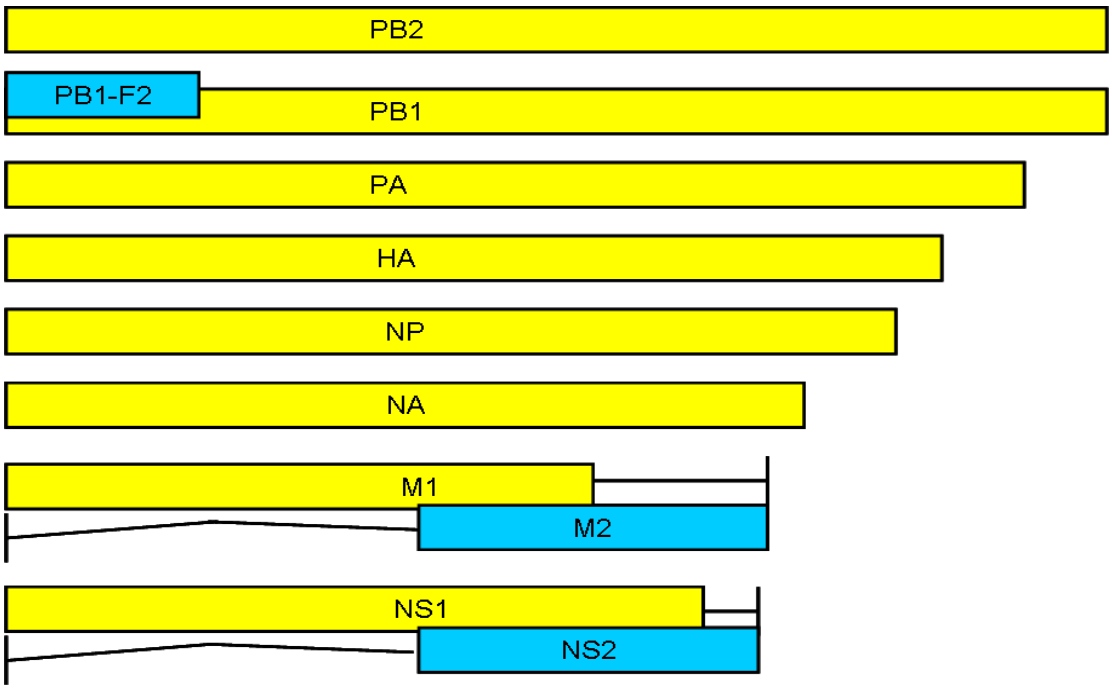
The influenza genome is comprised of 8 separate, segmented negative-sense RNA strands (totalling 14, 000 nucleotides) which encode 11 viral proteins (Figure 1B) [7]. Segments one through three of the RNA encode the viral polymerase subunits: PB1, PB1-F2, PB2, and PA. (PB2 encodes a protein involved in 'cap snatching', while PB1 possesses endonuclease activity and is involved in chain elongation during transcription). PB1-F2 is a recently discovered protein resulting from an open reading frame on the PB1 gene which yields an 87 amino acid long accessory protein involved in promoting apoptosis, and PA is a protein for which no

Figure 1: The influenza virus and the influenza viral genome. (A) Cross sectional cartoon depiction of a typical influenza virus particle. (B) Organization of the influenza genome (adapted from Fields Virology p 1651). As depicted in blue, PB1-F2 is created by a +1 frame shift, while M2 and NS2 are encoded by mRNAs which have been spliced [4]. The nucleocapsid protein is encoded by viral gene segment 5.

A


PB2
PB1
PBA
HA
NP
NA
M
NS
M1 Layer
Lipid Envelope (Acquired from Host)
NEP/NS2
Hemagglutinin
M2
Neuraminidase
Viral Ribonucleoproteins

B


PB2
PB1-F2
PB1
PA
HA
NP
NA
M1
M2
NS1
NS2

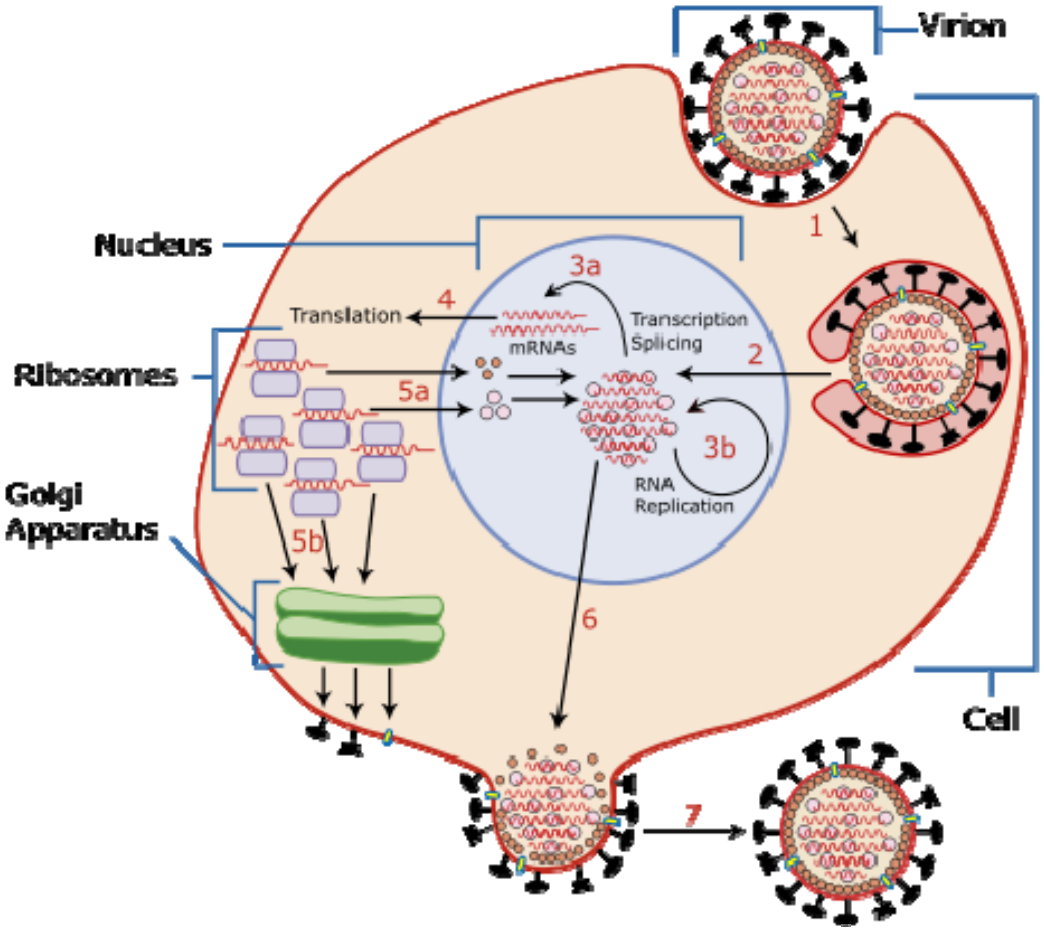
specific function yet determined, but mutation studies suggest a definite role in transcription and replication). Segment four encodes the major surface glycoprotein HA, which is responsible for binding to siaylated host cell receptors. The nucleoprotein (NP) is encoded by segment 5, and is a structural protein that encapsulates the viral genome providing stability and protection to the genome. NP also associates with the viral polymerase subunits to modulate replication and transcription. Segment seven encodes the two matrix proteins M1 (a protein involved in viral assembly and budding, and interactions with vRNPs) and M2 (a proton channel that acidifies during endosomal digestion). Finally, the last and shortest strand, segment eight encodes two non-structural proteins, NS1 (a protein that hinders immune interferon response) and NS2 (also known as NEP which plays a role in the export of vRNPs) [48].

1.1.3. Influenza Virus Life Cycle

Transmission of influenza occurs through three routes: direct contact with infected individuals, inhalation of virus-charged aerosols and contact with fomites [9]. Regardless of transmission route, the viral infection begins only after the influenza viral particle adheres to the cellular membrane on the epithelial surface of the body which is exposed to the environment. Typically, influenza infects cells of the upper respiratory tract and the onset of the flu is manifested by the presence of symptoms such as body chills, muscle aches, malaise, fever and an overall decrease in well-being.

Figure 2 presents a schematic representation of the influenza life cycle. HA and NA surface proteins perform two very distinct functions that must be carefully

Figure 2: The influenza virus life cycle. The infectious life cycle begins with binding of influenza viral particle to sialic acid via its hemagglutinin glycoprotein receptor. The viral particle is endocytosed, and fuses with an endosome (1). In the late endosome, a pH drop occurs causing the fusion of the viral membrane to the endosomal membrane, subsequently releasing the viral RNA into the cytoplasm (2). The viral genome is then transported into the host nucleus, where it is processed in two ways: the genome is transcribed (3a) into vmRNA which will be exported to the cytoplasm and translated into viral proteins (4) or the genome is replicated for progeny virus (3b). Newly synthesized viral proteins either traverse back into the nucleus to associate with the genome to form viral ribonucleoprotein complexes (5a) or they are sent through the secretory pathway and end up on the viral membrane waiting to be incorporated into newly formed viruses (5b). The new genome associated with the viral proteins (Nucleocapsid proteins + subunits of the viral polymerase) are exported to the site on the cell membrane that has acquired the remainder of viral proteins (6). Mature virus buds off from the parent cell, acquiring the appropriate surface glycoprotein (HA, NA and M2) (7). Figure taken from: <http://commons.wikimedia.org/wikiFile:Virus_Replication_zh-hant.svg>



<http://commons.wikimedia.org/wikiFile:Virus_Replication_zh-hant.svg>

balanced to achieve infection and optimal replication. Infection is facilitated initially by the binding events coordinated by attachment of the viral surface glycoprotein hemagglutinin (HA) to host sialylated cell receptors. Upon binding to the receptor sialic acid on the cell membrane, the viral particle is internalized by receptor mediated endocytosis. Acidification in the endosome induces a conformational change in HA leading to the fusion of the viral membrane to the endosomal membrane [10]. Specifically, the lowered pH results in the full exposure of the fusion peptide contained in HA2, which is extended into the endosomal membrane and helps pull the viral and endosomal membranes together [11]. The drop in pH internally results from the opening of the M2 ion channels. The influx of protons facilitates the uncoating of viral ribonucleoprotein or vRNP complexes, liberating them into the cytoplasm of the host cell. (The viral ribonucleoprotein complex – M1 protein interaction is disturbed upon decreasing the pH.) Viral replication requires transcription of the genome in the host nucleus but vRNPs are about 10-20 nm wide and, therefore, too big to passively diffuse into the nucleus. Nuclear localisation signals are primarily located on NP, and function to recruit impotin α and β, and together, in an energy dependent process, these proteins deliver the vRNPs to the nucleus [4, 11, 12].

Genome transcription is carried out by the virally encoded RNA-dependent RNA polymerase once the vRNP complexes have been translocated to the nucleus [13]. In order to transcribe the viral RNA, the influenza virus utilizes a 'cap-snatching' mechanism by which 10-13 nucleotides from cellular pre-mRNA transcripts are captured by viral PB2, cleaved by viral endonuclease (PB1) and used as primer at the 5'end of viral RNA [13, 14]. The nascent mRNA, therefore, contains a 5'cap and is

also polyadenylated at the 3'end as a result of a 'stutter' that occurs when the polymerase complex can not read past the polyuridine stretch on the vRNA [4]. The newly synthesized mRNA is spliced and sent to the cytoplasm where it is translated on the ribosomes to produce viral proteins.

Eventually, the viral process switches from transcription to viral genome replication. Replication involves the production of a full length (+) sense complementary strand to the (-) sense RNA strand, known as cRNA. Unlike transcription, this process does not require a cap, nor requires the cap snatching function of PB2 and the endonuclease activity of PB1 [4]. Negative sense vRNA is then produced from the cRNA template and associated with soluble NP and newly-formed viral proteins PB1, PB2, PA and NP are transported back into the nucleus where they assemble into vRNPs for progeny virus [15]. Newly-formed vRNPs associate with M1 and NEP (also known as NS2) to relocate from the nucleus to the site of virus assembly and budding at the apical plasma membrane (the membrane facing the lumen of a polarized cell) [4, 13].

While the vRNPs are transported to the site of assembly, other major surface proteins such as HA, NA and M2 traverse through the secretory pathway (endoplasmic reticulum and Golgi) and are post-translational modified as they make their way to the apical plasma membrane. Accumulation of M1 protein marks the beginning of the budding process, as an outer curvature of the membrane appears. Further accumulation of viral genome along with viral proteins results in the pushing out of the membrane until the inner core is fully enveloped, with the process terminating once the membrane pinches off near the base of the budding virion.

The final step of virion replication involves the removal of any sialic acid still bound to HA by NA, thus allowing full release of the new virion from the parental cell [4].

1.1.4. The Integral Membrane Proteins Hemagglutinin (HA) and Neuraminidase (NA)

The continuous antigenic change amongst HAs and NAs give rise to the different influenza subtypes, with 15 known HAs and 9 different NAs identified to date [5]. Of the 15 circulating HAs and 9 NAs, H1, H2, H3, NA1 and NA2 have been able to sufficiently adapt to infect humans [3]. Structurally, HA is an elongated homotrimer consisting of three identical subunits [16]. It is initially synthesized as a precursor protein (HA0) [17], requiring cleavage which liberates the two domains of each subunit, HA1 and HA2. These subunits consist of a long helical stalk (HA2) embedded into the viral membrane, and a large globular domain (HA1) containing the sialic acid binding pocket [16].

HA is the leading determinant of tissue tropism, based on its preference to bind either α2-3 or α2-6 sialic acid linkages [18]. Many strains of influenza do not infect humans but persist in other species, such as aquatic birds. The main differentiating determinant is that, in humans, the majority of sialylated cells in the upper respiratory tract contain α2-6 linkages, while those in birds contain α2-3 sialic acid linkages. This difference in glycosidic linkage acts as a species barrier preventing the widespread infectability of most influenza strains. However, there are species, such as pigs, that contain cells with both sets of linkages. These animals serve as a mixing pot for various influenza strains, as they can be infected with both α2-6 and α2-3 specific influenza viruses simultaneously and thus provide an opportunity for genetic reassortment [3]. Reassortment may result in new strains that

possess characteristic gene segments from avian strains that were not once able to infect humans, along with the glycan specificity that allows for binding to α2-6 linkages like those found in human-adapted viruses. The fact that these new viruses can 'jump' the species barrier and acquire the ability to infect humans is cause for concern as it provides an entry into the general human population for new and potentially more deadly influenza strains [3]. A prime example of jumping of the species barrier occurred with the 1957 pandemic strain, H2N2, in which three avian gene segments were reassorted into existing influenza strains capable of infecting humans [19].

Neuraminidase (NA), the other major integral membrane glycoprotein, is a large protein with a molecular weight of 200-240 kDa [20]. The gross structural form of NA resembles that of a mushroom containing a cubed-shaped head domain attached to a long slender helical stalk which is anchored by a small hydrophobic sequence of 29 amino acid residues near the N terminal of the protein [20, 21]. The crystal structure of NA revealed six, 4-stranded, anti-parallel β-sheets arranged in a symmetrical manner, much like 'propeller blades' [21]. The folded conformation is stabilized by the presence of 9 disulfide bonds. The highly conserved active site of NA is located centrally within a deep pocket where sialic acid (N-acetylneuraminic acid), the primary substrate, binds. NA specifically cleaves the ketosidic linkage between the terminal sialic acid and adjacent monosaccharide, usually galactose [2, 21]. This cleavage is necessary for the release of progeny virus. Deficiency in NA activity compromises vitality of the virus and results in its aggregation near the cell surface [22]. In comparison to HA, the roles of NA, aside from its sialidase activity, remain elusive.

1.1.5. Influenza Nucleocapsid Protein

Gene segment five of the influenza virus encodes for the nucleocapsid protein (NP). NP is approximately 498 amino acids long and each NP protomer has a molecular size of approximately 56 kDa. The predicted pI (the pH at which the net charge of the protein is zero) is 9.8, making it a positively charged protein at neutral pH, as is characteristic of most proteins that bind nucleic acids. NP's primary biological role is to condense the viral RNA into ordered viral nuclear complexes by binding to vRNA and then associating with 3 components of the viral polymerase, PB1, PB2 and PA to form the viral ribonuecleoprotein complex. NP has also been shown to interact with various cellular factors during influenza viral infection and to be a key modulator in viral transcription and genome replication [23] [24].

Two crystal structures have been solved for NP from different influenza strains. First, the H1N1 NP was solved to a resolution of 3.2 Å by Yizhi Jane Tao et al in 2006 [25], and then in 2008, Pang-Chui Shaw and collaborators solved the crystal structure of H5N1 NP to a resolution of 3.3 Å [6] . Taken together, these crystals revealed valuable structural information at the secondary, tertiary and quaternary levels and have further enhanced the current functional knowledge surrounding NP.

Both crystal structures show that NP is fairly helical and contains three domains: the head domain, the body domain and the tail loop/linker region [6, 25]. NP appears to look like a half moon, with the bulkiest domains being the head and body; the extending tail loop region interacts with neighbouring NP to facilitate oligomerization. The folded protein sequence begins in the body domain, with the N terminus containing helices 1 to 4. It then folds to the head domain, which consists of helices 5-11 and helix 19. From the head domain, helices 12-17 are looped back

to the body domain and then extend outwards to the tail loop/linker region which contains helix 18 and interacts with helix 19 which, as already mentioned, is located in the head domain [26] (Figure 3).

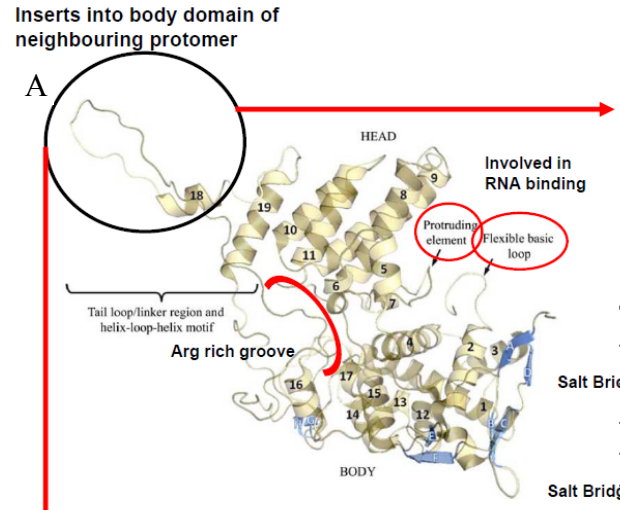
As noted above, NP is a largely helical protein consisting of 19 α-helices (48.6% of the protein), and 8 β-strands (4.8%) with the remaining amino acids (46.6%) in 'loop' or other non-regular secondary structural elements [26]. Alignment studies of over 2500 NP sequences from various influenza viral strains revealed that, in total, the helices are the most conserved secondary structural elements with 72.7% conservation. Loop sequence is 68.1% conserved while strand is only 58.3% conserved across the aligned sequences. Even more intriguing is the fact that helix 6 (responsible for RNA binding) and 7 (involved in homo-oligomerization) of the head domain show 100% conservation across all strains.

The crystal structure data for H5N1 NP, in conjunction with the sequence alignments, has allowed for a more in depth characterization of each domain of NP. The head domain (helices 5-11 and 19) is 80.6% helical, contains 19.4% loop, no β-sheet and displays 80.7% sequence conservation. There is only one amino acid that is hypervariable in this domain [26]. The body domain (helices 1-4 and 12-17) is less helical (42.9%), contains 7.9% β-sheet, 49.2% loop and has an overall sequence conservation of 66.1%. There are eight hypervariable residues in the body domain. The tail loop/linker region contains 20% helix (the contribution of helix 18), no β-strands and is 80% loop. This region of the protein shows 67.5% sequence conservation across strains and contains 3 hypervariable residues [26].

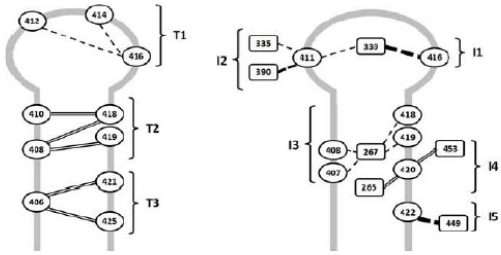
One of the major roles of NP is to encapsidate viral RNA. In order to accomplish this task, it must be able to homo-oligomerize and form higher-ordered,

Figure 3: The structural characteristics of the nucleocpasid protein from influenza H5N1. (A) X-Ray crystal structure adapted from Shaw [26]. NP is composed of a head domain, a body domain and a linker tail/loop region. Oligomerization is facilitated by the insertion of the circled region into the body domain of a neighbouring protomer. Crystal structure also displays the protruding element, flexible loop and arginine rich groove which are critical for RNA binding. **(B) Schematic depiction of the tail loop/linker region examined using mutational analysis (lower chart) adapted from Shaw [26]**. Oligomerization was drastically reduced when amino acids involved in salt bridges where disrupted. Other notable results include the abolition in activity in the double mutant L418S P419S, indicating these two amino acids have structural roles that are important for NP function. **(C) Schematic representation of the amino acids comprising the tail loop/linker region from amino acid 397-451**. The tail loop (amino acids 402-428) is flanked with two linkers (397-401 and 429-436) that contain variable amino acids conferring flexibility. Within the tail loop lies R416, which is involved in forming a salt bridge with E339 of a neighbouring protomer and Helix 18 also contained in the this domain. Helix 18 contains an R at position 422 which forms an intramolecular salt bridge with E449, stabilizing the helix into a fixed position.

Structural Characteristics of H5N1 NP

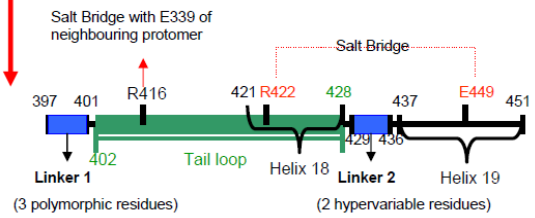


B



Mutation	Effect
E339A / R416A (Salt Bridge)	No activity with either residue mutated- incapable of oligomerizing – Exist as monomers rather than trimers/oligomers as seen with WT.
R267A	50% Decrease in Activity
R422A / E449A (Salt Bridge)	50% Decrease in Activity –mixture of unstable oligomers when either residue mutated
T390S/ S407A/ T411S/ F420S	No change in activity
L418S P419S	No activity
L418S	No change in activity from WT
P419S	No change in activity from WT
V408S P410S	No activity
I406S	50% decrease in activity
V408S/ P410S/ I425S	No change in activity

C



more compact structures known as the ribonucleoprotein complexes (vRNPs) that further associate with subunits of the viral polymerase. Homo-oligomerization is facilitated by the insertion of the tail linker/loop region of one NP (amino acid 402-428) protomer into the “insertion groove” within the body domain of a neighbouring NP protomer (Figure 3) [6, 25-27]. The tail loop region is conserved across strains, but the linker regions flanking the tail loop vary considerably leading to the differences in orientation of this domain relative to the head and body domains [26]. In a study in which various amino acids within the tail loop of NP were mutated and functional analysis of these mutants was conducted, amino acids directly involved in homo-oligomerization were identified. Most notably, dynamic light scattering indicated that when the intermolecular salt bridge E339-R416 between protomers was disrupted, homo-oligomerization could not occur and these mutants remained in the monomeric state. In contrast, the wild type NP existed exclusively as trimers and higher-order oligomers [27]. This data suggest that, while the tail loops may still have the capacity to access the groove, the interaction between one protomer and the next is no longer maintained, resulting in the monomeric rather than oligomeric state. The residues E339 and R416 involved in this intermolecular salt bridge are 100% conserved across influenza strains surveyed in the sequence alignment studies [27]. The results of this mutational analysis (summarized in Figure 3) delineated a number of other residues important for NP structure-function. For example, hydrophobic interactions in the stem region of NP also play a role in NP activity, but to a lesser degree than the salt bridges. NP appeared to be functional when single point mutations were introduced to the hydrophobic amino acids: V408S, P410S, L418S and P419S. However, double mutants such as V408S

P410S or L418S P419S abolished RNP activity [27]. Homo-oligomerization was also compromised when an I406S mutation was introduced at the base of the tail loop structure, and alterations to the interacting salt bridge at R422-E449 (another apparent inter-chain interaction between protomers) resulted in a 50% decrease in activity and the resulting NP variants existed as unstable higher-ordered oligomers [27]. The authors of this study (who are also responsible for solving the crystal structure of H5N1 NP) propose three factors that regulate NP oligomerization events: 1) Tail loop/insertion groove interaction (salt bridge at E339-R416), 2) Tail loop conformation stabilization (as shown by the hydrophobic amino acid mutations), 3) Maintenance of the homo-oligomer to ensure RNP activity [27].

In vRNP, each protomer/monomer of NP binds to the phosphate backbone of roughly 24 vRNA nucleotides with high affinity in what appears to be a non-sequence specific fashion (K_d=20nM) [26,28]. Florence Baudin and coworkers have used differential enzymatic digestion of RNA fragments in a 'panhandle' secondary structural motif both in the presence and absence of NP to demonstrate that the secondary structure of RNA 'melts' as it associates with NP making it more accessible for viral transcription; in the process, the nitrogenous bases of the bound RNA are exposed to the environment [28].

It has been shown that NPs from other negative-sense segmented RNA viruses like rabies and vesicular stomatitis virus (VSV) do not change conformation upon binding RNA [29, 30]. However, this has not yet been proven in influenza NP and should not be assumed as there are differences between the NPs from these three viruses. NP from influenza is longer in primary sequence, with 498 amino acids versus 398 residues in rabies and 422 residues in VSV. The three proteins also

show little homology [6]. Although all three show the same global conformation (a crescent-like shape) and contain a head, body and linker tail/loop domain, there remain three distinct features that differentiates the influenza NP from that of rabies or VSV. Firstly, the way in which RNA is encapsidated is different. Rabies and VSV isolate the RNA at the inside of their capsid formations, while in influenza, the RNA bound to NP is solvent exposed [26, 29-30]. Secondly, influenza NP has a 'protruding element' comprised from amino acids 167-168 [26] situated in a groove between the head and the body domains of the protein which is absent in rabies and VSV NPs. This protruding element is rich in positively charged arginine and lysine residues which are all important for RNA binding and completely exposed when NP is in oligomeric state further demonstrating how RNA is wrapped on the outside of influenza NP [6]. Finally, NP contains a highly basic flexible loop (residues 72-91) region which is also in the groove between the head and body domain. Others have suggested that RNA binding is initiated by this basic loop (residues 72-91) which surveys the immediate environment for any viral RNA, and eventually binds that RNA. Since this basic RNA-binding loop is in close proximity to the arginine rich groove, RNA is drawn in to the groove and then locked or clamped into place by the protruding arginine side chains which point towards each other [6]. These three structural features, exclusive to the NP from influenza, are highly conserved across influenza strains and it is possible, perhaps even likely, that NP may change conformation upon RNA binding. The idea that the flexible loop will be "locked" (via the bound RNA) in close proximity to the arginine/lysine region suggests that, like intrinsically disordered proteins [31] at least this part of the molecule is unstructured in the unbound state but possibly becomes relatively more fixed when RNA is bound[31].

1.2. The Immune System

Both existing prophylactic and potential therapeutic influenza vaccines are designed to stimulate the immune system to mount a defence against infecting influenza virus. The human immune system is a complex network of cells, tissues and organs responsible for conferring protection from invading pathogens such as bacteria, viruses and worms. It can be divided into two main arms, the innate immune system and the adaptive immune system [32]. The innate immune system is more 'primitive' in that it relies on physical barriers such as the skin, acts almost immediately with maximum exertion in response to infection and does not confer memory or utilize antibodies. It is the first line of defence of our bodies[32]. Occasionally, a virus is capable of evading the innate immune system and eventually infects a host cell. When the innate immune system cannot cope with the infection, a stronger adaptive immune response is mounted. The adaptive immune response takes more time to mount as it is antigen specific. Adaptive immune response results in immunological memory and thus is considered to be a more advanced response [32].

The adaptive response can be further divided into cellular and humoral immune responses. Humoral immunity involves the action of B cells, a class of lymphocytes that are derived from and matured in the bone marrow. B cells are responsible for the production of antibodies and, ultimately, for immuno-memory [32]. Cell-mediated adaptive immune response, on the other hand, elicits the T cells, a class of lymphocytes that are derived in the bone marrow and mature in the thymus. T cells play a role in assisting B cells in expressing antibodies, interact with phagocytes to enhance the destruction of pathogens and finally destroy host cells

that have been infected with a virus or other pathogen. T cells exert their effects by releasing cytokines in order to attract other immune cells and by cell-cell interactions [32].

Cytotoxic T Lymphocytes (CTLs) fall into the category of T cells involved in the cell-mediated adaptive immune response. Cytotoxic T Lymphocytes are concerned primarily with mounting immune responses to viruses and other intracellular pathogens. CTLs can be characterized by their CD8+ surface co-receptors and their T cell receptor (TCR) both of which coordinate the recognition of antigens presented on antigen-presenting cells. When the antibody-mediated response in a host is evaded, the T cells are activated. However, the T cells require a signalling mechanism to indicate that a cell is infected. This is accomplished through antigen presentation. Antigen presenting cells such as dendritic cells, macrophages and B cells take the endogenous pathogen and process it. The viral protein is sent to the proteasome, where it is digested into peptides and then associates with Major Histocompatibility Complex (MHC) class I before translocating to the cell membrane for presentation. The TCR on a CTL binds to the MHC class 1 containing the antigen and is subsequently 'primed' and can now recognize other cells that have the same antigen and essentially kill them. This killing occurs when Cytotoxic T cells release granules known as perforins that perforate the cell membrane of the targeted cells [32].

CD8+ T cells (CTLs) interact exclusively with MHC class I, and, originally, this mode of presentation was believed to arise from processing of endogenous proteins/antigens. Any exogenous proteins were assumed to interact with MHC class II, and thus would elicit an antibody response. However, the mechanism of

cross presentation precludes this way of thinking. Cross presentation is a pathway by which exogenous proteins/antigens can be endocytosed by a dendritic cell, processed by the proteasome and associated with MHC class I for presentation. This cross presentation mechanism would then allow the CTLs to respond to an infection without the infection of an antigen presenting cell. This is a point of contact for therapeutic vaccines, in that exogenous antigens such as NP can be injected into a patient and expands the CD8+ T cell population [32].

1.3. NP as a Universal Vaccine Candidate

Most current vaccines against influenza are prophylactic and utilize HA and NA antigens to induce humoral immune responses. Because these antigens are different in the different influenza strains, seasonal flu vaccines based on HA and NA must be prepared annually to target the most prevalent circulating strain. NP, on the other hand, presents an attractive universal vaccine candidate that, while not prophylactic, would increase in the viral clearance after infection, thus functioning as a therapeutic vaccine. NP is attractive because it is relatively conserved across all known influenza strains (as described above) and has low drift rate in its sequence [25, 33]. In addition, more recent research indicates that recombinant NP has the potential to promote cross-reactivity and therefore may be a valuable constituent within existing vaccines[34-37]. NP has been shown to stimulate an immune response by eliciting the cytotoxic T Lymphocytes in both human and mice [37, 38] although it appears that recombinant NP requires the cytokine IL-2 (either endogenous or added) in order to enhance proliferation of both CD4+ and CD8+ T cells [37]. This cellular response to NP inoculation has been shown to result in a quicker viral clearance and faster recovery from influenza infection [37]

Up until recently, most research has been focussed on targeting NP antigens which interact with MHC class I to elicit a T cell response because this was believed to be the main mechanism of protection. However, recent studies have suggested that, in addition to activating T cell response, recombinant NP has the capacity to induce substantial titers of anti-NP antibodies. These anti-NP antibodies, while not neutralizing, are capable of increasing viral clearance and confer hetero-subtypic immunity[39]. For example, in mice vaccinated with recombinant NP, the injection not only induced a cellular response (marked by an expanded CD8+ T cell population), but could also induce a humoral response that significantly contributed to viral clearance and conveyed cross-protection from different subtypes of influenza [39-41]. Anti-NP antibodies appear to engage in antiviral activity early on in infection and are speculated to play roles such as inhibiting replication through interaction with macrophages, stimulating the production of cytokines and chemokins and promoting the production of cytolic molecules from natural killer (NK) cells [40] . Clearly NP's ability to elicit CTLs coupled with its ability to induce the production of anti-NP antibodies displaying antiviral properties and conferring heterosubtypic immunity make this conserved protein an extremely attractive vaccine candidate.

1.4. Canada's Regulation Governing Biologics

In the event that a manufacturer decided to create a recombinant NP vaccine, based on it's therapeutic potential, regulatory agencies in the developed world would require an in-depth characterization of the properties of this biologic and impose stringent regulatory framework on manufacturers. In Canada, Health Canada is the federal regulatory agency that is responsible for regulating all drugs and vaccines for human use, both innovator products and subsequent entry biologics or biosimilars.

Biologics, including vaccines, are the purview of the Biologics and Genetic Therapies Directorate at Health Canada. (A biologic is considered any vaccine, antibody, blood or tissue therapeutic product which is derived from a biological source, whether naturally-occurring or expressed through the use of biotechnology). Unlike their small-molecule counterparts, biologics are subjected to regulation of both the manufacturing process and the final product. Regulation of the manufacturing process is necessary in the realm of biologics because they have complex structures which are extremely easy to alter and sensitive to the manufacturing processes. Even the smallest changes in production process could result in alteration of the biologics' efficacy and/or its overall safety profile [42, 43].

1.5. Vaccine Development and Regulation: Current Approaches

In the development of a vaccine, two major considerations must be taken into account. The first is the discovery of the biologic which elicits the desired therapeutic response. Discovering highly immunogenic antigens is a major task, as is the subsequent vigorous lab testing required to verify the therapeutic response being elicited is in fact due to the antigen in question [44]. The second and equally important aspect in vaccine development is the translation of the biologic (the isolated antigen or modified/attenuated organism presenting that antigen) alone into a formulation which, while appropriate for both safe injection and bio-distribution throughout the body, also confers stability to the antigen, giving it a shelf life of an extended period of time. In examining these two aspects of vaccine design, both manufacturers and regulators must consider the relationship between structural integrity and biological activity [44].

Vaccines date back to the 1770's when Edward Jenner developed the small pox vaccine [45]. The concept of vaccination, therefore, is not new and, consequently, much of the thinking surrounding the creation and testing of vaccines remains rooted in the traditional (and rudimentary) microbiology and immunology of the earliest effective vaccines. For historical reasons, many of the methods and end-points used to characterize and define vaccines are similar to those employed back in the late 1700s. For instance, a vaccine's utility and efficacy are still defined by its ability to induce an immune response using biological agents with similar immunological identity as the pathogen of interest [45]. While these end-points are still biologically relevant, it is now apparent that many other factors may be critical to vaccine function. Merely measuring cytokine response and antibody titres is no longer considered accurate enough or precise enough to evaluate the safety and efficacy of a vaccine completely and physiochemical methods are gaining acceptance as the ideal way to characterize the structural behaviour of vaccine antigens under a variety of stress conditions. Indeed, it has been suggested that both manufacturers and regulators should start viewing a vaccine as a physicochemical entity (much like a therapeutic protein or other non-vaccine biologic) and must consider integrating 'low resolution' analytical procedures when testing potential vaccine formulations to get a 'true' picture of the limitations to an antigen in a formulation in terms of both safety and efficacy [44].

1.6. Physicochemical Approaches to Vaccine Assessment

A variety of physicochemical approaches and analytical techniques are amenable to vaccine characterization, whether the vaccine is a complex mixture (like a split virion or a live, attenuated bacterium) or a simple, isolated (and perhaps

recombinant) antigen [44]. Physicochemical methods, however, are more commonly used with component vaccines (vaccines based on a single antigen or only a few components/antigens of the infective agent) where the information gleamed can be more descriptive and contribute to the scientific understanding of structure-function relationships in the antigen [46-47] .

One useful physicochemical technique (and one used in the thesis work described here) is circular dichroism spectropolarimetry (CD). Far-UV CD is a rapid technique capable of evaluating secondary structural information of a protein/ biologic and giving insight into the stability of the protein or protein mixture in a variety of stress conditions (exposure to denaturatants, osmolytes, heat and pH change). The technique is useful in observing the structural consequences a mutation has on overall protein folding and stability [48, 49]. It is important to note that CD is largely an empirical method (though some theoretical basis for observed spectra has been developed and, while it gives estimates of proportions of each secondary structural element, is not strictly quantitative [50]). Circular dichroism involves the differential absorption of the left and right circularized components of plane polarized light through a protein sample. In proteins, the amide bond along with side chain chromophores are the most optically active. The difference between the degree of left and right polarized light absorbed is known as circular dichroism [48, 49] and is reported as differences in molar absorbance (molar CD):

$$\Delta\varepsilon(\lambda)=\varepsilon_R(\lambda) - \varepsilon_L(\lambda);\ \ cm^{-1}\ M^{-1} \qquad (1)$$

Where ε_R and ε_L are the extinction co-efficient of the right and left polarized lights at wavelength (λ) [51], or in terms of mean residue ellipticity (deg cm^2 $dmol^{-1}$) given by:

$$[\theta]_{mrw}\lambda=MRW\ \theta/10\ (l)(c) \qquad (2)$$

where $[\theta]_{mrw}\lambda$ is mean residue ellipticity at wavelength λ, [θ] is observed ellipticity (degrees), l is pathlength (cm) and c= concentration in g/mL [52, 53]. The relationship between the (1) and (2) is given by

$$[\theta]_{mrw} = 3298\ \Delta\varepsilon \qquad (3)$$

When regions of the polypeptide adopt highly ordered arrays, their optical transitions can be altered (shift or split) as a result of excitation. The wavelength of these transitions can therefore be increased or decreased while the intensity of the transition can be concurrently enhanced or diminished. Taken together, different structural elements within a protein results in characteristic CD spectra [48, 54]. For example, the repeating φ and ψ angles of a helical peptide yield a characteristic CD spectrum containing a negative band at both 222nm and 208nm, and a positive band at 192nm [51]. A purely β-sheet conformation, on the other hand, shows a negative band at 215nm and a positive band at 198nm [51] while disordered (or "random coil") proteins show low ellipticity at 210nm and a negative band near 195nm[48,55].

As stated above, CD can be used to examine the structural implications certain stresses impose upon a protein. The structural changes in the protein can be followed while changing the pH, ionic strength or buffer conditions. Thermal denaturation can also be performed and followed on CD to reveal the temperature at which the protein "melts", as well as reveal the thermodynamics surrounding folding and unfolding [56]. For example, suppose a protein of interest is thermally denatured, CD can follow the transition between folded, F, and unfolded protein, U, at any temperature, T, and can give the equilibrium constant of folding, K_f [56]. Using the Pace method or Greenfield method to analyze the CD data obtained at various

temperatures [57], the fraction of unfolded protein can be determined by the following equation, assuming a two-state unfolding mechanism occurs:

$$F(\text{folded}) \longleftrightarrow U\ (\text{unfolded})$$

$$f_F + f_U = 1 \qquad (4)$$

The constants f_F and f_U represent the fraction of protein present in either the folded or unfolded conformation, respectively. Since the system is assumed at any time (or temperature, in thermal denaturations) to contain only folded and unfolded protein in various quantities, and no intermediates, the system is set to equal 1. An observed value of y (CD signal, mean residue ellipticity or molar ellipticity) at any point will be equal to

$$y = y_F f_F + y_U f_U \qquad (5)$$

where y_F and y_U represent the folded and unfolded components of y, respectively. Equations (4) and (5) can be combined to give:

$$f_U = (y_F - y)/(y_F - y_U) \qquad (6)$$

From equation (6) it is possible to determine the melting temperature of a protein, T_M, which is equal to the temperature at which 50% of the protein is in the unfolded state (f_U = 50%).

While far-UV CD is commonly used to examine protein secondary structure, fluorescence spectroscopy is a tool that reveals some detailed information on the tertiary structural characteristics of a protein using the aromatic residues of the protein itself as intrinsic probes for folding. Proteins contain natural chromophores in the form of their aromatic residues (ie. tryptophan, tyrosine and phenylalanine). Tryptophan residues absorb and emit the most light, and are therefore used most frequently as probes for the microenvironment within a protein. Briefly, fluorescence

spectroscopy involves the excitation of a protein at a given wavelength that causes the valence electrons to move temporarily into a higher electronic state. Upon returning to ground state, energy is released in the form of light. The amount of energy released is equal to the amount of fluorescence observed at higher wavelengths [58]. As a protein is unfolded by heat or by chaotropic agents, the fluorescence wavelength increases to higher wavelengths — a phenomenon called a “red shift”. This red shift is indicative of the tryptophan residues becoming more exposed to the hydrophilic environment. When the protein folds, and becomes more compact, the fluorescence wavelength often decreases — a phenomenon known as a “blue shift”. A blue shift indicates that the protein is becoming more compact, and that the exposed aromatic residues are becoming buried within the hydrophobic core of the protein [58]. Fluorescence spectroscopy, in combination with varying experimental conditions, therefore, can give information about the response of the changes in the microenvironment of the aromatic residues in response to stress conditions. When used in conjunction with other physicochemical techniques (like far-UV CD), it can give valuable information pertinent to the safety and efficacy assessment of vaccine antigens.

2.0. Rationale and Statement of Objectives

As noted in the above introduction, current vaccines against influenza are prophylactic and target the two main surface glycoproteins hemagglutinin and neuraminidase. Unfortunately, these vaccines must be re-formulated yearly, to reflect the predicted threat of what is to come the following year. This approach, based on eliciting antibodies towards strain-specific antigens, is effective but very cumbersome, costly and leaves the population completely vulnerable to any potential pandemic strain which may have acquired the capacity to jump the species barrier and for which there is no previous immunity to [59]. Moreover, the production time for each annual or new pandemic is roughly nine months, and therefore precludes any modification to more appropriately fit a current, but unexpected, influenza landscape of the population. It is for these reasons that other approaches to conferring protection against influenza are being developed.

One alternate route currently considered is to explore cellular-immunity based strategies for targeting multiple influenza strains. This master's thesis will focus on such an approach involving one antigen in particular, the nucleocapsid protein (NP) — a protein which has been shown to be highly conserved across influenza strains [26]. NP is highly immunogenic and capable of eliciting a cytotoxic T lymphocyte response and, more recently has been shown to aid in the production of non-neutralizing antibodies which have demonstrated an increased viral clearance capacity [36, 40, 41, 60]. It is for these reasons that the NP antigen is a potential candidate for a universal therapeutic influenza vaccine. Any new vaccine based on the influenza NP antigen will be subjected to regulation by the Biologics and Genetic Therapies Directorate at Health Canada, in whose laboratories this master's thesis

work was conducted. Health Canada is interested in characterizing NP antigens from avian influenza H5N1 to gain insights into both the biology and the physicochemical properties of this antigen from a regulatory perspective and to build on the scientific knowledge and expertise needed to develop policy for regulation of similar therapeutic vaccines.

It has previously been shown that the nucleocapsid protein from severe acute respiratory syndrome (SARS coronavirus) is somewhat "unstable" [61, 62]. N-proteins from other viruses [63, 64, 65] have also been shown to be both conformationally and chemically unstable — a trend that is not surprising given that the nucleocapsid protein needs, at different times in the viral life cycle, to bind or encapsulate RNA or DNA and, at other times, to expose it so it can be transcribed or replicated. The flexibility of NPs to perform the diametrically opposed functions of binding and releasing the viral genome suggests conformation flexibility and perhaps even instability in the protein itself. **It is therefore hypothesized that, similar to the NP of SARS, the NP from influenza H5N1 will show low conformational stability that will be critically important to consider when making NP-based vaccines and in assessing them for safety and efficacy.**

The objective of this master's thesis is to extensively characterize the NP antigen from H5N1 in order to gain insights into both the biology of the virus and the properties of the antigen that might influence the quality, safety and efficacy of a therapeutic vaccine. In order to achieve these objectives, a variety of physicochemical methods will be used to assess NP conformation and conformational stability under a variety of conditions.

3.0. Materials and Methods

3.1. Chemicals and Reagents

Reagents including growth media LB broth and agar, imidazole, phenylmethanesulfonylfluoride (PMSF), sodium phosphate monobasic monohydrate, sodium phosphate dibasic heptahydrate, BCA reagents such as bichionic acid and copper (II) sulphate, 10M stock NaOH, RNaseA, lysozyme, NaCl, KCl, adenosine, RNA from baker's yeast, *Saccharomyces cerevisiae* and diethylpyrocarbonate (DEPC) were purchased from Sigma-Aldrich (St. Louis, USA). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was purchased from Roche (Indianapolis, IN, USA). The antibiotic carbenicillin used during protein preparation and the Sypro® ruby red protein gel stain were obtained from Invitrogen (Carlsbad, CA, USA). Sodium dodecyl sulfate (SDS) and ammonium persulfate (APS) were purchased from BioRad (Mississauga, ON, Canada). Stock titrants such as 10M HCl were from EMD Biosciences (Gibbstown, NJ, USA), while 18M H_2SO_4 was from BDH (West Chester, PA, USA). Vanadyl sulfate was purchased from Fisher (Fair Lawn, NJ, USA). Anti-His antibody for Western blot analysis was purchased from Amersham Biosciences as was the secondary horse radish peroxidase-linked goat anti-rabbit secondary antibody. Anti-NP, a polyclonal antibody raised in rabbits using isolated NP, was a gift from Dr. Sean Li (Ottawa, ON, Canada). . All solutions were prepared using distilled and deionised water (NANOpure Diamond system, Barnstead International (Asheville, NC, USA). Buffers were filtered through 0.2μm membrane filters prior to use in FPLC/HPLC.

3.2.1. Expression System

A glycerol stock of *Escherichia coli* BL21 cells containing the complete sequence for the nucleocapsid gene from A/duck/yokohama/aq10/2003(H5N1) was sub-cloned into expression vector pQE2 (Qiagen),a gift from the laboratory of Dr. Sean Li. The pQE2 vector provided the 6XHis-tag located at the N-terminal of the protein necessary for affinity chromatography. Competent *E. coli* BL21 cells were transformed with the pQE2 vector. The transformed BL21 cells were added to a 50% glycerol solution and stored at -80°C.

3.2.2. Protein Expression and Purification

Growth media (Lauria-Bertani (LB) EZ Broth Mix from Sigma / LB Agar Powder from Sigma)for overnight cultures, large scale growth culture and Petri plates were prepared, autoclaved and cooled. 100µL of 50 mg/mL carbenicillin (Invitrogen) was added to the cooled LB agar. Freshly made LB + carbenicillin plates were streaked with recombinant H5N1 NP glycerol stock BL21 cells and incubated overnight at 37°C. The plates were then stored at 4°C for up to 3 weeks and served as a working stock for preparation of cultures. Individual, well-isolated colonies from these plates were used to inoculate 12mL of LB supplemented with 12µL of 50 mg/mL carbenicillin, and the culture was incubated overnight (18-20 hours) at 37°C on a bench top shaker (Innova 4000 Incubator Shaker) at 250 RPM.

To prepare the large scale growth media, 1mL of 50 mg/mL carbenicillion was added to 1L of sterile LB in a 2.8L Fernbach flask. 2mL of this broth was removed and set aside to be used as a blank for optical density measurements. The remainder of the broth was inoculated with 10mL of the overnight culture previously prepared and incubated at 37°C on the bench top shaker at 150 RPM for 2-3 hours.

The OD_{600} was monitored using the BioRad Smart Spec 3000 spectrophotometer throughout the incubation until the culture reached an OD_{600} of 1.0 at which point recombinant gene expression was induced with a final concentration of 2.82μM IPTG (isopropyl β-D-1-thiogalactopyranoside) and allowed to incubate overnight at 37°C /150 RPM.

The following day, the cultures were removed from the incubator and transferred to two large 500mL centrifuge bottles. The balanced bottles were centrifuged for 30 minutes at 8,000 x g and 4°C on the Beckman Avanti J-251 Centrifuge using the JA-10 rotor. After centrifugation the supernatants were discarded and the pellets were either processed immediately or stored at -80°C for later use.

Processing of the pellet required an ice bucket to keep the fresh pellet cool, or to thaw the previously made pellets that had been stored. The pellet was resuspended with 25mL Buffer A (50mM Na_3PO_4 / 200mM KCl, pH 7.4). Once the pellet was fully resuspended, it was transferred into a small 30mL centrifuge tube and 12.5μL of 10mg/mL lysozyme was added to the suspension immediately (to weaken the cell walls) and incubation allowed to proceed at room temperature for 30 minutes at which time the test tube was placed back on ice and sonicated (Fisher Scientific Sonic Dismembrator model 500) for 5 minutes. Sonication parameters were as follows: 5 minute run time – 1.0 second on, 1.0 second off (10 minutes total), (Amplitude 28%). Immediately after sonication, 1mM PMSF a serine protease inhibitor was added to the sonicated suspension and the sample centrifuged for 30 minutes at 25,000 x g at 4°C.

After centrifugation, the supernatant (the soluble fraction containing the nucleocapsid protein) was collected then loaded onto a 50mL superloop and applied

to a His-Trap™ column (GE Healthcare) pre-equilibrated with Buffer A (50mM Na_3PO_4 / 200mM KCl, pH 7.4) attached to an Amersham Biosciences AKTA Fast Performance Liquid Chromatography (FPLC). Nucleocapsid protein (NP) was eluted from the His-Trap column using a stepwise imidazole gradient : 10% Buffer B (50mM Na_3PO_4 / 200mM KCl/ 200mM Imidazole, pH 7.4 and 90% Buffer A for 25 minutes, 35% Buffer B for 30 minutes (to elute any non-specific binding) and then 100% Buffer B for 25 minutes (to fully elute bound NP)). The flow rate was set to 2.5mL/min. Aliquots (10mL) of individual FPLC fractions were mixed 1:1 with 2X loading buffer (62.5mM Tris-HCl, 25% Glycerol, 2%SDS, 0.01% Bromophenol Blue, 710mM β-mercaptoethanol) and electrophoresed on SDS-PAGE to acertain which fractions contained NP. The remainder of each 10mL FPLC fraction was diluted in half using Buffer B and those containing NP were combined and dialyzed in 25,000 kDa molecular weight cut-off dialysis tubing against 3 changes (2h per equilibration) of 4.0L Buffer A at 4°C in order to remove imidazole. One final dialysis against fresh 4.0L Buffer A was then performed overnight at 4°C. Protein concentration was then determined using a standard bicinchoninic acid (BCA) assay. Absorbance at 280nm was also recorded.

3.3. SDS-PAGE / Densitometry

The purity of NP was estimated using densitometry. NP containing solutions were mixed 1:1 with 2X SDS-loading buffer (BioRad) and analyzed using 12% SDS–PAGE using a Mini-Protean 3 cell (Bio-Rad) running at 400 mV for 45 minutes. Gels were stained overnight using Sypro® stain. The NP band density was assessed as a percentage of the total lane density using a GelDoc XR with Quantity One4.5.2 software (BioRad).

3.4. Western Blot Analysis

NP identity was further verified by Western blot analysis. Briefly, two SDS gels were epectrophoresed in the same running tank. One gel was loaded with sample and later stained with Sypro® stain. The other gel was loaded twice with the same sample as in the first gel. The doubly loaded gel was electrophoresed, and then cut in two down the center. One gel was washed with anti-His antibody at a diulution of 1:5000, while the other half was simultaneously washed with anti-NP antibody at a dilution of 1:10,000. Following the antibody washes, the membranes where wrapped in saran wrap and exposed to Kodak X-ray film for detection. Films were then developed in the darkroom after 30 second exposures.

3.5. Protein Concentration Determination

Protein concentration was determined by BCA assay for each NP preparation prior to use in further experimentation. Also, optical density (OD_{280}) was also recorded to provide a concentration calculation using the theoretical molar absorptive coefficient (ε) of 54235 M^{-1} cm^{-1} (Expasy ProtParam). Concentration was also determined for one NP preparation using quantitative amino acid analysis performed by Dr. Rey Interior at the Sick Children's Hospital (Toronto, ON).

3.6. Circular Dichroism Spectropolarimetry (CD)

Protein samples were dialyzed against 10mM Na_3PO_4 at pH 7.4 (physiological conditions). 3.0mL aliquot samples were placed in a 1.0cm quartz cuvette (Hellma™ Analytics) and analyzed by CD with constant stirring. NP concentration varied from 0.03mg/mL to 0.07mg/mL depending on the preparation. CD spectral readings at room temperature were taken using a JASCO J-815 Spectrometer under nitrogen purge. Parameters during spectral acquisition were as

follows: band width: 1nm, response: 1 sec, sensitivity: standard, measurement range: 290-195nm, pitch: 0.1nm and scanning speed: 50nm/min. Temperature was controlled (at 25^{o}C for room temperature scans) using a Peltier unit (Jasco PFD-425S/15). Spectra represent the cumulative average of 5 spectral accumulations, and this raw spectral data (in mdeg) was then converted to molar ellipticity and/or mean residue ellipticity. In order to compute the amount of secondary structural elements within NP, specialized online algorithms on Dichroweb http://dichroweb.cryst.bbk.ac.uk/html/home.shtml or K2D2 (*BMC Structural Biology* 2008, 8:25) (http://www.ogic.ca/projects/k2d2/) were used.

3.6.1. Thermal Denaturations Followed by Circular Dichroism

Protein samples were thermally denatured and secondary structural changes followed on CD. Cuvettes of differing path lengths were used depending on the concentration of the NP prep (1.0 and 0.1cm pathlengths). Samples in the 1.0cm cuvette were constantly stirred. Regardless of the cuvette size, all sample preparations were capped to prevent evaporation of NP. Again the CD was flushed with nitrogen gas, and the parameters were as follows: band width: 1nm, response: 1sec, sensitivity: standard, measurement range: 20-85°C, data pitch: 1°C, monitor wavelength: 222nm and temperature slope: 1°C/min. Full length CD spectra were recorded at 5°C intervals as temperature increased from 20-85°C. Data at 222nm was recorded for every 1°C increase and deposited into graphing software Sigma Plot for melting temperature determination.

3.7. Fluorescence Spectroscopy

Fluorescence readings of NP samples were conducted using a Varian Cary Eclipse Fluorescence Spectrophotometer. 2.5mL samples of NP were placed into

1.0cm quartz cuvettes from Hellma™ Analytics. The cuvette was then dusted with a Kimwipe® and placed in the multi-cell holder. Initial readings were always taken with the photomultiplier (or gain) set to medium. Depending on the signal, the gain was adjusted to give optimal signal with minimal noise. Normally, a gain setting between 500-700au sufficed. The excitation wavelength was determined from an initial wavelength scan from 300-400nm with an arbitrary excitation wavelength (280nm) to obtain the emission maxima. This excitation wavelength was used for all fluorescence measurements pertaining to that particular sample. This method was used to determine exact excitation wavelengths for every NP sample.

3.7.1. Thermal Denaturations Followed by Fluorescence

Protein samples were thermally denatured and followed by fluorescence in 1.0 cm Hellma™ cuvettes with constantly stirring. All NP sample preparations were capped to prevent evaporation of NP. Prior to beginning the thermal denaturation, the emission maximum was determined from initial room temperature wavelength scans as described above. Thermal denaturation data were collected under a variety of pH and salt concentration conditions. Data accumulated were deposited into Excel™ and analyzed using the method of Pace et al.[57] to determine fraction unfolded as a function of temperature.

3.8. pH Titrations

Stock protein solution was gently mixed on a Rotator prior to performing the titration. The concentration of NP (10mM Na_3PO_4) varied depending on the preparation analyzed but was always between 0.03-0.07mg/mL. The pH meter was calibrated using standards at pH 7.0 and 10.0 and the probe was placed in the CD cuvette containing sample. (2.5mL of protein was added into a 1.0cm Hellma™

quartz cuvette before taking measurements). A small stir bar was dropped into the suspension for constant mixing, throughout the experiment. Freshly prepared stock solutions of acid (6N, 1N HCl, 1N H_2SO_4) and base (1N, 0.5N NaOH) were used in each titration. CD readings starting at pH 7.4 were recorded. CD measurement parameters were as follow: band width: 1nm, response: 1sec, sensitivity: standard, measurement range: 290-195nm, pitch: 0.1nm, scanning speed: 50nm/min, stir bar speed: medium. Immediately after recording the CD spectrum at the desired pH, the optical density at both 280nm and 214nm was recorded using a BioRad Smart Spec 3000 spectrophotometer and then fluorescence emission spectra were recorded using a Varian Cary Eclipse Fluorescence Spectrophotometer. After recording all measurements at the initial pH value, the pH was manually adjusted to the next desired value (in increments of 0.5 pH units) by adding small aliquots of acid. Once the desired pH had been reached, the sample was allowed to sit for 5 minutes with constant stirring. The process of recording CD, $OD_{280/214}$ and fluorescence was repeated. The process was repeated until the pH had been adjusted to pH 3.0.

Once pH of 3 had been reached, the entire protocol was repeated using base to bring the pH back to 7.5 in increments of 0.5 pH units. The pH probe was rinsed after every addition of acid or base with distilled water. The resultant spectra obtained were corrected using blanks prepared to account for volume and effects of acid/base. Analysis was done using Jasco Spectra Manager's "Spectra Analysis" program.

3.9. Effects of Ionic Strength (NaCl Titration)

The effects of salt on NP secondary structure was assessed using CD in a similar manner to that previously described. Stock protein solution was gently mixed

on a rotator prior to performing NaCl titration. During this time, a fresh stock of 5.0M NaCl was prepared. 2.5mL of NP sample (0.03-0.07mg/mL) was placed into a 1.0cm Hellma™ cuvette and sequentially titrated from 0.0M – 1.0M NaCl in 0.1M intervals with constant stirring.

3.10. High Performance Liquid Chromatography

In order to gauge the size of the NP, size exclusion chromatography was performed using a Thermo Finnegan High Performance Liquid Chromatography (HPLC) consisting of a P4000 pump, AS3000 sampler, vacuum and UV detector. Eluted material was monitored at 280nm and 214nm, and data was analyzed using "Chromquest" computer software. Samples were applied to a Tosch G3000 (7.8mm (ID) X 30.0cm (L)) column using a running buffer comprised of 50mM Na_3PO_4, 150mM NaCl and adjusted to pH 7.2. It is important to note that the HPLC was run with 0.1% SDS to prevent the protein from sticking to the column. The column was calibrated using Sigma Aldrich's protein standard kit prior to running samples.

3.11. Dynamic Light Scattering

Particle sizing using dynamic light scattering (performed in the laboratory of Dr. Michael Johnston, Health Canada) provided an alternate approach to determine the size/aggregation state of NP. Freshly prepared samples (400uL) of NP immediately after purification were analyzed with a zeta potential/particle sizer (PSS/Nicomp 380 ZLS Dynamic light scatterer). Samples analyzed were either undiluted or diluted with buffer to the following ratios: 1:2, 1:3, 1:4.

3.12. Examination of the Effects of RNA on NP Structure – RAaseA Digest after 'Anomalous' Spectra Recorded on CD

When an ‘anomalous’ CD spectra was obtained after dialyzing into 10mM phosphate buffer, the solution containing NP was divided into two equal volumes and each sub-fraction placed in a 15mL Falcon tube. One of these sub-fractions was incubated overnight at room temperature with gentle agitation, while the other sub-fraction was similarly incubated in the presence of RNAseA (at approximately 1/50 of the NP concentration as determined by BCA). After approximately 18 hours after RNaseA addition, comparison CD spectra were obtained for the two sub-fractions. The pH of each sub-fraction was then dropped to 3.0 using concentrated HCl (6.0M) and the spectrum recorded. The pH was slowly adjusted up to 7.0 using NaOH (1.0M) and the CD spectrum at this pH was obtained.

3.13. Deliberate Addition of RNaseA During NP Preparation

NP purification was carried out exactly as described above. However, 10µL of 10mg/mL RNase A was added after sonication and the suspension allowed to incubate for 2 hours at 37°C prior to centrifugation.

3.14. RNase Free Work

3.14.1. Establishment of RNase Free Working Conditions

In order to keep RNase contamination to a minimum, preventative measures were taken. Clean gloves were worn at all times. Diethylpyrocarbonate (DEPC)-treated water was produced by adding 10mL of DEPC to 20L of double distilled water and allowed to sit overnight in a 20L carboy. The next morning the carboy was autoclaved for 90 minutes to kill any residual RNase enzymes and eliminate trace DEPC. During buffer preparation, sterile techniques were followed when handling all reagents such as using DEPC-treated water, autoclaving glassware prior to use and ensuring that all metal spatulas and weigh boats, etc. were

thoroughly wiped with RNase Erase® (MP Biomedicals) and then rinsed with DEPC-treated water. Before each buffer pH measurement, the probe was rinsed with RNase Erase® as well and rinsed to prevent any contamination. In addition, the entire bench, along with all pipettes were thoroughly sprayed with RNase Erase® and then rinsed with DEPC-treated water before experiments were conducted. RNase free, sterile pipette tips from Eppendorf were used, as well as pre-sterilized 50mL and 15mL Falcon tubes. Plastic ware was treated with 0.1N NaOH/1mM EDTA overnight, and then rinsed thoroughly with DEPC-treated water before use.

3.14.2. Preperation of a Small Molecule RNase Inhibitor: Vanadyl Complex

The small molecule RNase inhibitor was created as outlined [66]. Briefly, 2M vanadium sulphate ($VOSO_4$) was prepared by mixing 2.6g $VOSO_4$ in 8.0mL DEPC-treated water. Adenosine (0.25M) was separately prepared by mixing 3.34g adenosine into 48mL of DEPC-treated water in a 100mL beaker. (The adenosine solution required heat to be fully dissolved.) The solution pH was monitored using an RNase Erase® treated pH probe. 6mL of 2M $VOSO_4$ was added drop wise to the adenosine solution. A color change (dark green) and a pH drop to pH 2.5 had signalled the formation of vanadyl-adenosine complex. Immediately upon formation of this complex, the pH was adjusted to 6.5-7.0 with 10M NaOH. The volume was adjusted to 60mL with DEPC-treated water and the complex was stored in 1mL aliquots at -80°C. When needed, individual aliquots were thawed and diluted 1:10 in the lysis buffer for NP purification.

3.14.3. Verification of RNase Inhibition of Vanadium-Adenosine Complex

In order to verify that the vanadyl complex was formed and active, an RNase Alert®Lab test kit (Ambion) was used. (This kit utilizes a fluorescent substrate

containing a fluorophore linked to a quencher via a small RNA linker. If the RNA is cleaved by RNase enzymes, the quencher is released and will no longer quench the fluorescence from the fluorophore, resulting in fluorescent spectra indicative of the presence of RNase.) Briefly, 45µL of sample was incubated with the substrate for half an hour at 37°C. After incubation the sample was added to 2.0mL DEPC-treated water and readings were taken on the fluorimeter using an excitation wavelength of 490nm and following the emission at 520nm. A positive control in which RNase A was deliberately added to a test aliquot was included, as was a negative control containing solely DEPC-treated water.

3.14.4. NP Preparation in the Presence of RNase Inhibitor: Vanadyl Complex in Lysis Buffer

After establishing an RNase free working area, the freshly made NP pellet was resuspend in Buffer A (50mM Na_3PO_4 / 200mM KCl, pH 7.4) (~25mL/pellet) on ice. The re-suspension was then transferred into a 25mL centrifuge tube which had been pre-treated with DEPC or RNase Eraser. To the re-suspension lysis buffer (25 mL), vanadyl complex (2.5mL of 0.2M) was added using an autoclaved RNase-free graduated cylinder. A stock solution of lysozyme was prepared in DEPC-treated water and added to the vanadyl complex lysis buffer (to a final concentration of 5ug/mL) and the resultant solution incubated for 30 minutes at room temperature. The reaction tubes were then placed on ice and sonicated for 5 minutes (10 minutes: 1second on, 1second off) using Fisher Scientific Sonic Dismembrator model 500 sonicator at 28% amplitude. Prior to sonication, the sonicator probe was thoroughly rinsed with RNase Erase® and DEPC-treated water. Periodically, sonication was interrupted so that an additional 100uL of vanadyl complex could be

added to the lysis suspension. Immediately after sonication, 250uL of 100mM PMSF was added to each 25mL sample before centrifugation at 25,000x g for 30 minutes at 4°C. The soluble fraction was collected and used for NP isolation using FPLC as previously described.

3.14.5. NP preparation in the Presence of RNase Inhibitor: Vanadyl Complex in all Buffers

The protocol described above was repeated however, instead of adding vanadyl complex to solely the lysis buffer, all buffers used in the NP preparation contained vanadyl complex. All buffers used in FPLC and dialysis were made with DEPC-treated water in an RNase-free manner, and all contained vanadyl complex.

3.15. NP Titration with Eukaryotic RNA

NP was purified and digested with RNaseA as described above. The RNaseA was then removed from the sample by dialysis using a 30kDa molecular weight cut-off membrane (4 changes of 2.0L per change of 10mM Na_3PO_4 in DEPC treated RNase free water). A stock solution of RNA (RNA from baker's yeast, *S. cerevisiae* from Sigma) was prepared at a working concentration of 1mg/mL. NP was placed in a rectangular CD cuvette (1cm path length) and RNA was added in successive, small aliquots until the concentration of RNA was 10 X more than the measured concentration of NP (based on the assumption that 24 nucleotides of RNA bind to each NP monomer) in an attempt to drive binding. After each addition of 2.5uL of RNA, the sample was allowed to stir for 5 minutes then a CD spectrum was acquired using the protocol described above for the pH titrations. NP was titrated with RNA at both pH 7.0 and pH 5.5.

4.0 Results

4.1. Isolation and Verification of Nucleocapsid Protein

NP recombinantly produced in IPTG-induced BL21 cells was isolated using immobilized metal affinity chromatography (IMAC) and FPLC fractions from this isolation were analyzed using SDS-PAGE. A band with an electrophoretic mobility consistent with protein of a molecular size of 60kDa was routinely observed in specific column fractions, which accurately compared with the molecular weight of 56kDa for the expected $(His)_6$-tagged NP as theoretically determined. Using densitometry, it was determined that NP was isolated to a purity of greater than 95% (Figure 4 and Figure 5A). Western blot analysis of these FPLC fractions using either anti-NP or anti-His antibody as a probe further verified the identity of NP (Figure 5B and 5C, respectively). Western blotting with anti-NP antibody revealed a band corresponding to the expected 60kDa molecular size, with some degradation product in the more concentrated samples. Anti-His antibody probing showed histidine containing protein fragments throughout all of the fractions. Occasionally, such analysis of an NP preparation revealed a second band on Sypro™-stained SDS-PAGE at 30kDa which, according to Western blot analysis, was rich in histidines as shown by the presence of bands in the blot probed with the anti-His antibody (Figure 5C).

4.2. Verification of Proper Folding of NP (Circular Dichroism)

Every NP preparation was analyzed by CD spectropolarimetry (after dialysis of the sample into 10mM Na_3PO_4) to ensure proper protein folding. In total, fifteen preparations produced in the absence of RNase inhibitors were analyzed. Twelve of the fifteen gave CD spectra at pH 7.4 that resembled the blue trace shown in Figure

Figure 4: A representative SDS-PAGE analysis of purified NP (elution fractions from Ni-NTA column, Fractions A11-B10) of protein preparation #1. A protein band migrating with the expected molecular size (56 KDa) and of purity > 95% is evident in all five fractions. L represents the molecular size ladder, Invitrogen's Benchmark™ product.

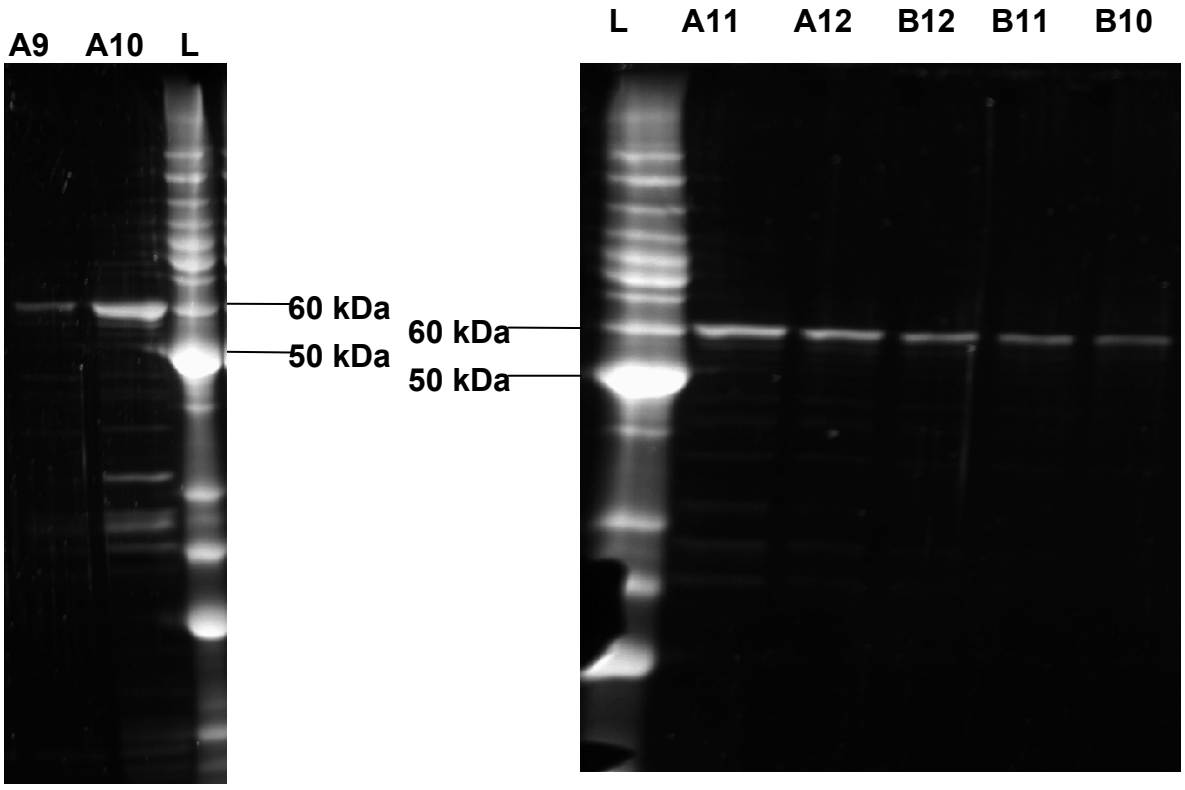

A9
A10
L
60 kDa
50 kDa
L
A11
A12
B12
B11
B10
60 kDa
50 kDa

Figure 5: (A) Sypro™-stained SDS-PAGE analysis of purified NP (elution fractions from Ni-NTA column, Fractions A9-A12) of protein preparation #1 (one month after purification stored at 4°C). A band at expected molecular size (56kDa) and of purity > 95% is evident. **(B and C) Western blot analysis of a replicate SDS-PAGE electrophoresed simultaneously with the gel shown in (A)**. (B) A band at the expected molecular size (56kDa) is evident in all fractions when the anti-NP antibody is used. Fractions A10 and A11 show some faster migrating material (potentially a degradation product). (C) Several bands representing proteins of various molecular size in addition to that expected for NP protein can be seen in the Western blot analysis using anti-His antibody.

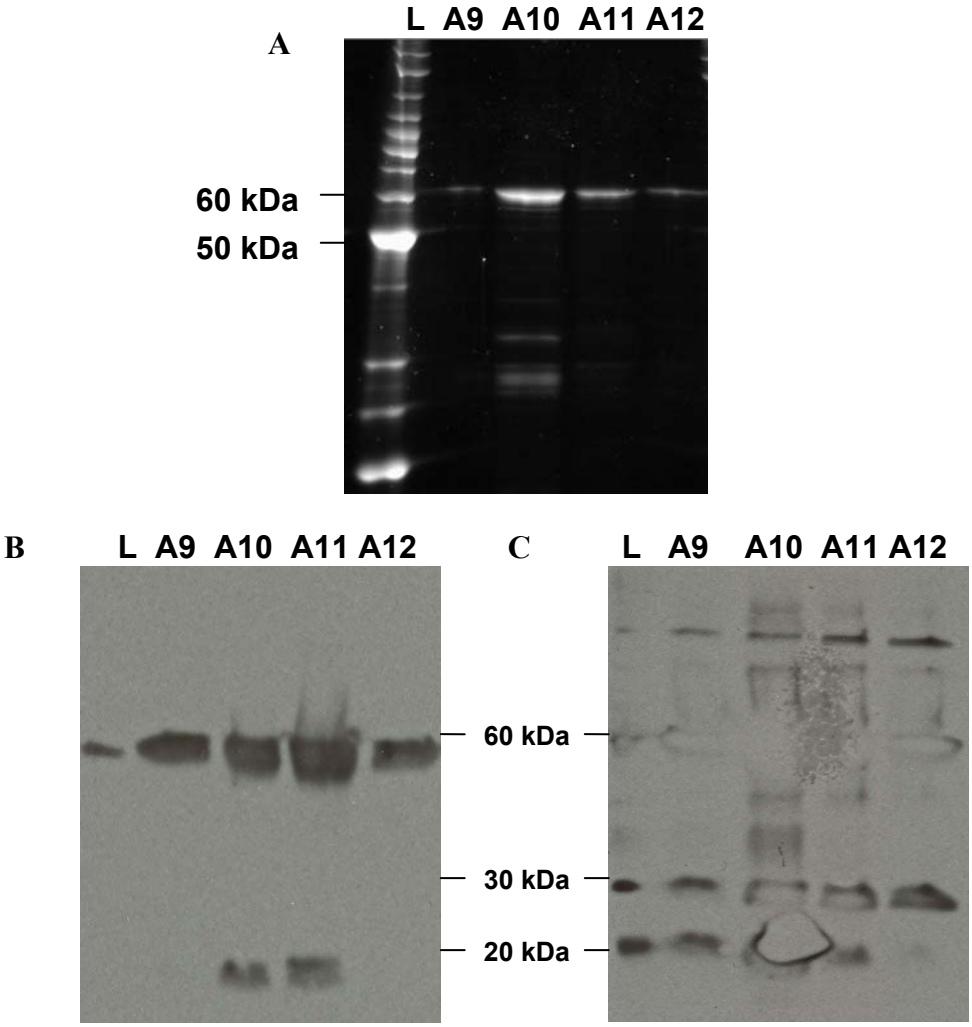
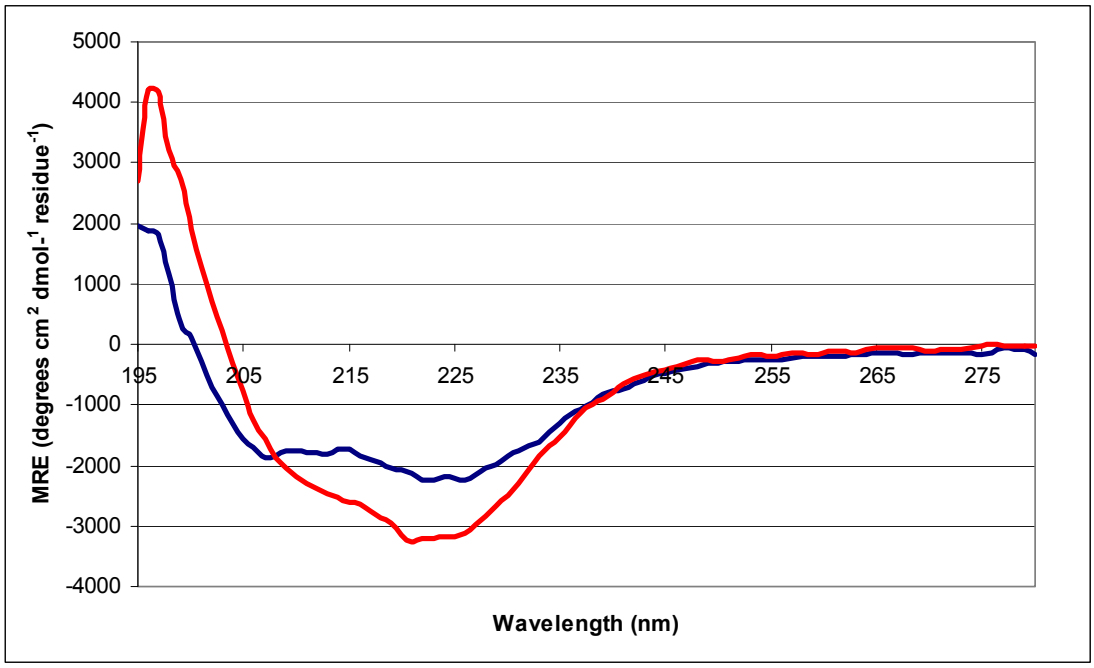

A
L A9 A10 A11 A12
60 kDa
50 kDa
B
L A9 A10 A11 A12
C
L A9 A10 A11 A12
60 kDa
30 kDa
20 kDa

Figure 6: Representative CD spectra obtained for predominant (blue) and anomalous (red) NP preparations at room temperature in 10mM Na_3PO_4, pH 7.4. Twelve of the fifteen preparations consistently gave CD spectra at pH 7.4 that resembled the blue trace. This trace was dubbed the, 'predominant' spectrum and contains two distinct minima at 222nm and 208nm. However, with 3 of the 15 preps (or 20% of the time), a CD spectrum that looked quite different than the 'predominant' spectrum was obtained dubbed the 'anomalous' spectrum, and is depicted by the red trace. The anomalous spectrum appears to have only one minimum at 222nm.

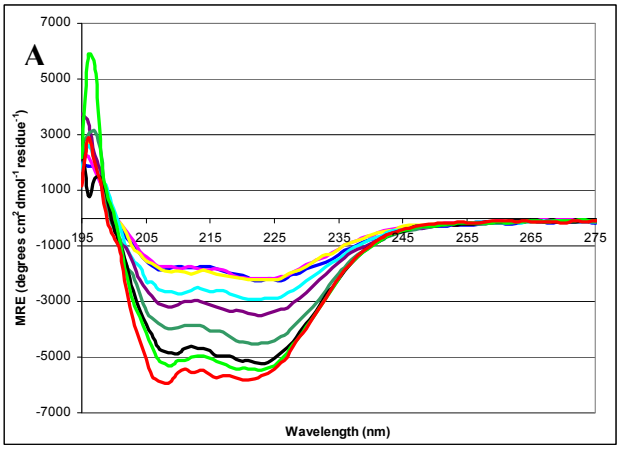

5000
4000
3000
2000
1000
0
-1000
-2000
-3000
-4000
195
205
215
225
235
245
255
265
275
MRE (degrees cm2 dmol-1 residue-1)
Wavelength (nm)

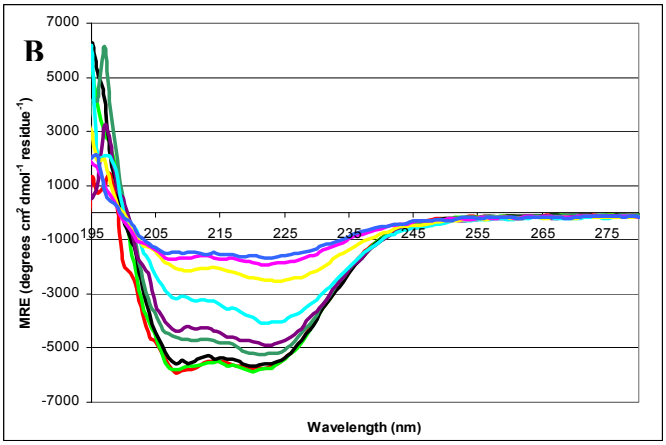
6. This trace was dubbed the, 'predominant' spectrum. However, with three of the 15 fifteen preparations (20%), a CD spectrum that looked quite different than the 'predominant' spectrum was obtained. This spectrum was dubbed the 'anomalous' spectrum and is depicted by the red trace in Figure 6. Interestingly, there were no obvious deviations or errors in the production process (in terms of growth conditions, isolation methods etc.) that could account for the 'anomalous' result. The spectrum for the 'predominant' preparations appears to have two distinct minima at roughly 222nm and 208nm, while the 'anomalous' spectrum appear to have only one minimum at 222nm. Raw data was converted to mean residue ellipticity (MRE) using concentrations determined by BCA and/or A_{280} measurements and submitted to secondary structure determination algorithms (CDSSTR, Selcon 3, K2D [67] and K2D2 [68]). Unfortunately, none of the four deconvolution programs used could give an estimate of secondary structure content that produced a satisfactory fit with the experimental data from either the 'predominant' or 'anomalous' preparations (Appendix 1).

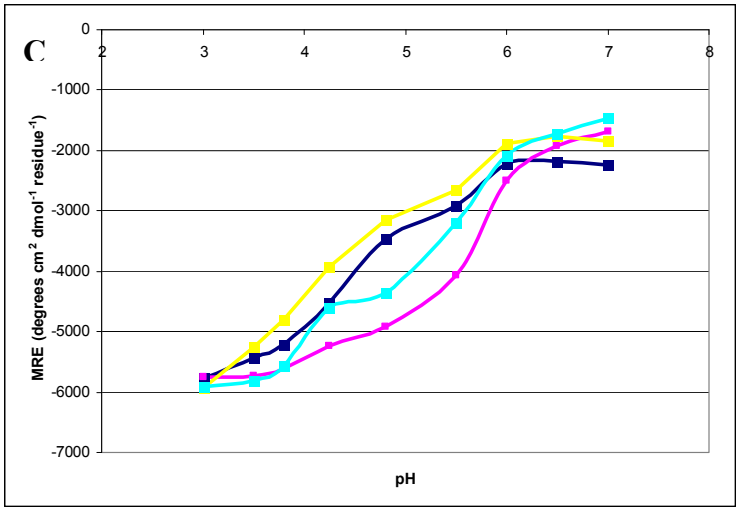
4.3. Secondary and Tertiary Structure Estimation of 'Predominant' Preparations

Circular dichroism spectropolarimetry indicates that, upon decreasing the pH from near-physiological (pH 7.4) to acidic conditions (pH 3.0) using HCl, NP adopts an altered conformation(s). The increase in negative ellipticity around 222nm observed in the CD spectrum at more acidic pH values (compare the spectrum of pH 7.4 to 3.0) suggests that the altered conformation may contain more helical or helix-like secondary structural elements than the conformation assumed at neutral pH. Upon returning to original pH, the degree of negative ellipticity at approximately 222nm (and thus presumed helical

Figure 7: Representative CD spectra obtained for predominant NP preparations when titrated with 1M HCl (A) and 1M NaOH (B). Upon adjusting the pH from 7.0 to 3.0 with 1M HCl, an increase in mean residue ellipticity (MRE) occurs (A). Upon returning to pH 7.0 using 1M NaOH, a decrease in MRE occurs (B). (A and B) pH: 7.0 (—), 6.5 (—), 6.0 (—), 5.5 (—), 5.0 (—), 4.5 (—), 4.0 (—), 3.5 (—), 3.0 (—). **(C) Cross sectional view of maxima at 222nm and 208nm throughout the titration.** (C) (▪) HCl titration following 222nm, (▪) NaOH back titration following 222nm, (▪) HCl titration following 208nm, (▪) NaOH back titration following 208nm.


A
7000
5000
3000
1000
-1000
-3000
-5000
-7000
195
205
215
225
235
245
255
265
275
MRE (degrees cm2 dmol-1 residue-1)
Wavelength (nm)


B
7000
5000
3000
1000
-1000
-3000
-5000
-7000
195
205
215
225
235
245
255
265
275
MRE (degrees cm2 dmol-1 residue-1)
Wavelength (nm)


C
0
-1000
-2000
-3000
-4000
-5000
-6000
-7000
2
3
4
5
6
7
8
MRE (degrees cm2 dmol-1 residue-1)
pH

content) decreases consistently as the pH is incrementally increased (Figure 7A and B). Cross sectional views at both 222nm and 208nm demonstrate that the conformational change is completely reversible, in that both points have the same ellipticity before and after the pH titration (Figure 7C). Similar, yet less dramatic results were obtained when freshly prepared NP is titrated with 1M H_2SO_4. When titrating from pH 7.4 down to pH 3.0 an increase in negative ellipticity is observed at both 208nm and 222nm, however, the depth of the through is not as deep as when titrated with HCl. When returning to pH 7.0 from pH 3.0 with NaOH, negative ellipticity is lost at 222nm and 208nm (Figure 8A and 8B). Cross sectional analysis also reveals that the change induced by the addition of sulphuric acid is reversible (Figure 8C).

Fluorescent emission spectra were recorded throughout the titration to acid pH values in order to assess tertiary structure of the predominant NP preparations. When NP was titrated from pH 7.0 to pH 3.0, a slight red shift was observed, as the emission maximum shifted from 336nm to 338nm (Figure 9A). Upon back titration, a blue shift was observed, as the emission maximum shifted from 340nm at pH 3.0 to 337nm at pH 7.4 (Figure 9B). Characteristic fluorescence profiles of NP during titration with sulphuric acid and the reverse titration with NaOH are shown (Figure 10A and 10B). In a similar fashion, decreasing the pH with sulphuric acid again resulted in a blue shift being observed, with the wavelength of the emission maximum shifting from 337nm to 340nm. Then, upon addition of base, the maximum shifted back from 338nm to 335nm (a red shift). The changes in emission intensity at 340nm (the emission maximum at acid pH) are shown cross-sectionally in Figure 10C.

Absorbance at 280nm and 214nm (A_{280} and A_{214}) were recorded to gauge the concentration of samples used to obtain the CD and fluorescence spectra. During the

Figure 8: Representative CD spectra obtained for predominant NP preparations when titrated with 1M H_2SO_4 (A) and 1M NaOH (B). Upon adjusting the pH from 7.0 to 3.0 with 1M H_2SO_4, an increase in mean residue ellipticity (MRE) occurs (A). Upon returning to 7.0 using 1M NaOH, a decrease in MRE occurs (B). (A) pH: 7.4 (—), 7.0 (—), 6.5 (—), 6.0 (—), 5.2 (—), 4.6(—), 4.0 (—), 3.5 (—), 3.0 (—). (B) pH: 7.5(—), 7.0(—), 6.5 (—), 6.0 (—), 5.5 (—), 4.5(—), 3.5 (—), 3.0 (—). **Cross sectional view of maxima at 222nm and 208 nm through out the titration (C)**. (▪)H_2SO_4 titration following 222nm, (▪) NaOH back titration following 222nm, (▪)H_2SO_4 titration following 208nm, (▪)NaOH back titration following 208nm.

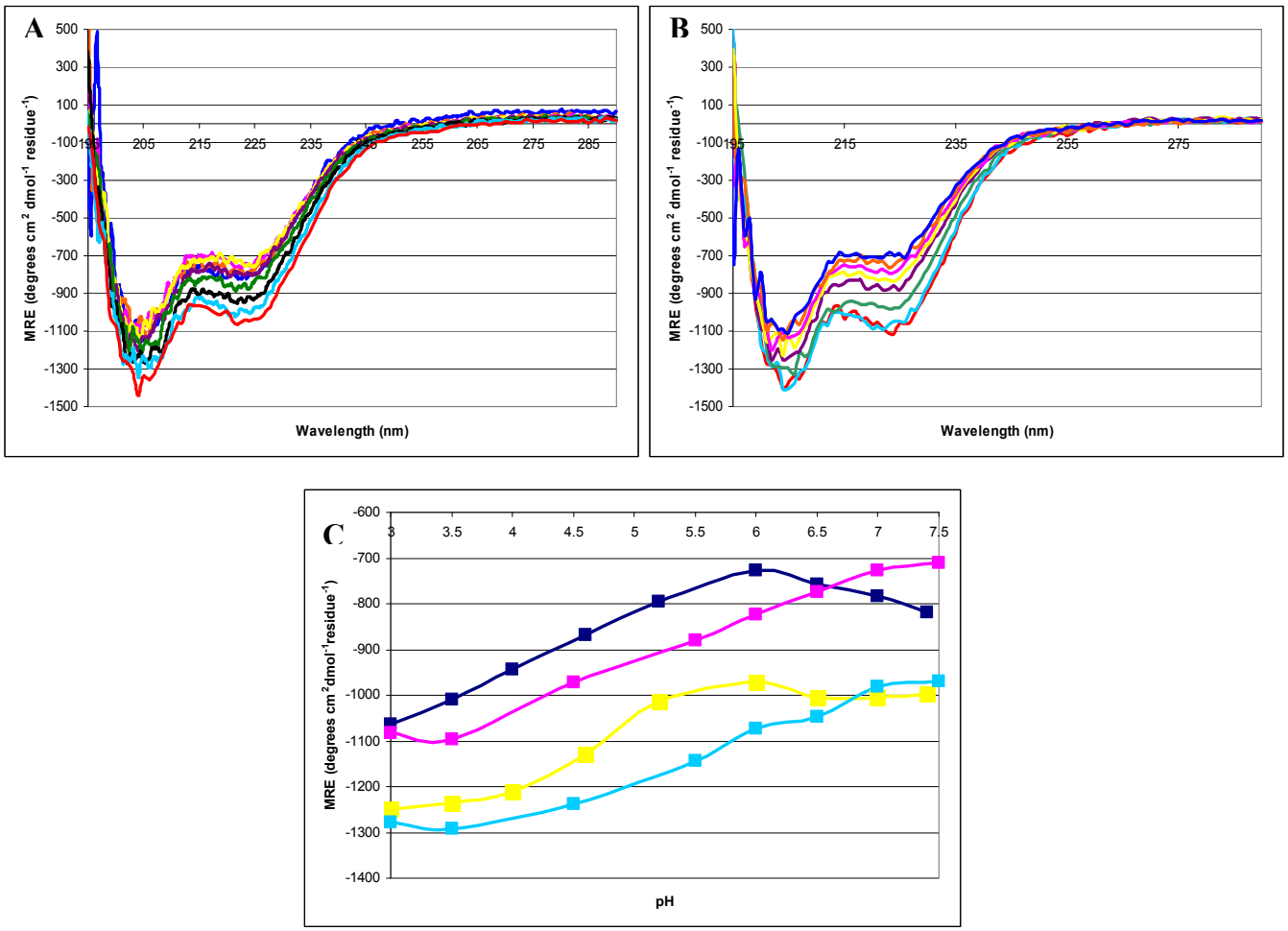
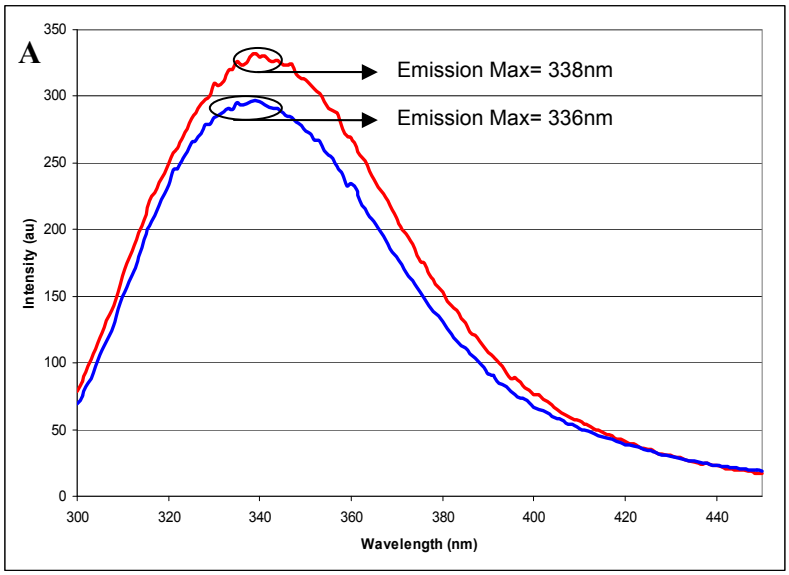

A
500
300
100
-100
-300
-500
-700
-900
-1100
-1300
-1500
205
215
225
235
255
265
275
285
MRE (degrees cm 2 dmol-1 residue-1)
Wavelength (nm)
B
500
300
100
-100
-300
-500
-700
-900
-1100
-1300
-1500
215
235
255
275
MRE (degrees cm 2 dmol-1 residue-1)
Wavelength (nm)
C
-600
-700
-800
-900
-1000
-1100
-1200
-1300
-1400
3
3.5
4
4.5
5
5.5
6
6.5
7
7.5
MRE (degrees cm2dmol-1residue-1)
pH

Figure 9: Characteristic fluorescence profiles of NP during titration with HCl (A) and back titration with NaOH (B). (A) Titration performed on a predominant NP preparation from pH 7.4 to pH 3.0 followed by fluorescence spectroscopy. The fluorescence emission maximum observed at pH 7.4 occurs at 336nm. Upon adjusting the pH, a red shift is observed, with the wavelength shifting to 338nm. (B) Titration of a normal NP preparation from pH 3 to pH 7 followed by fluorescence spectroscopy. The fluorescence emission maximum occurs at 340nm at pH 3.0. Upon returning to neutral pH, a blue shift occurs with the emission maximum shifting to 337nm. pH: 7.4 (—), 3.0 (—).


A
Emission Max= 338nm
Emission Max= 336nm
Intensity (au)
Wavelength (nm)
0
50
100
150
200
250
300
350
300
320
340
360
380
400
420
440

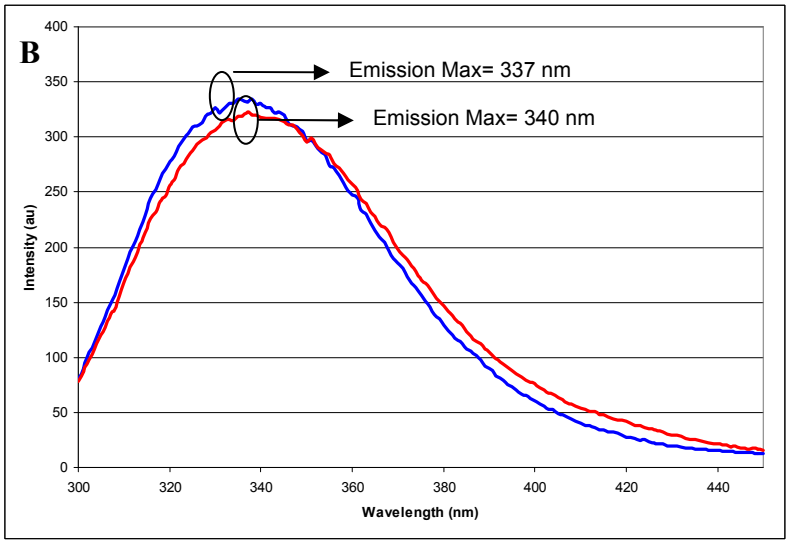
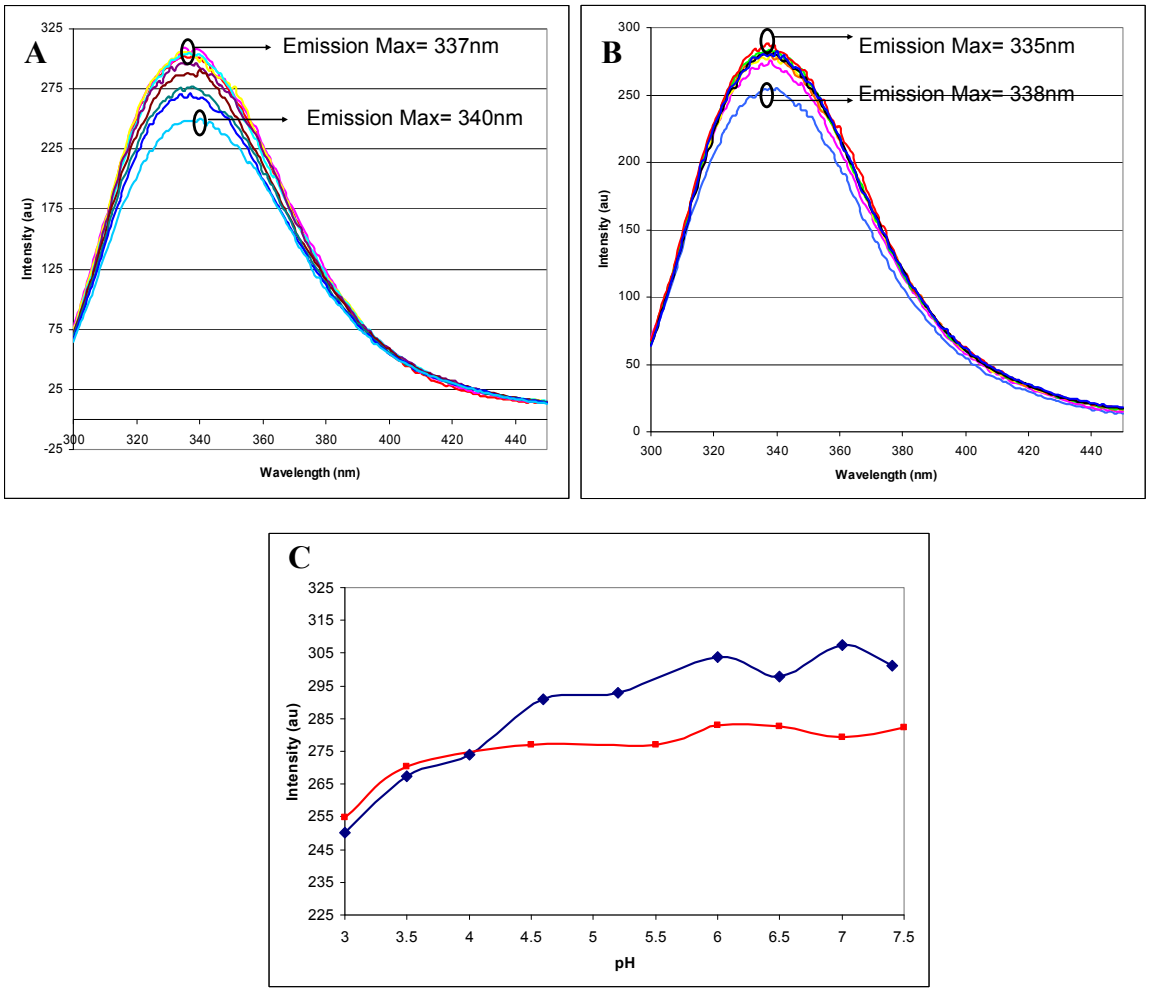

B
Emission Max= 337 nm
Emission Max= 340 nm
Intensity (au)
Wavelength (nm)
0
50
100
150
200
250
300
350
400
300
320
340
360
380
400
420
440

Figure 10: Characteristic fluorescence profiles of NP during titration with H_2SO_4 (A) and back titration with NaOH (B). (A) Titration of a predominant NP preparation from pH 7.4 to pH 3.0 followed by fluorescence spectroscopy. The fluorescence emission observed at pH 7.4 is at 340nm. (B) Titration of a normal NP preparation from pH 3.0 to pH 7.0 followed by fluorescence spectroscopy. pH: 7.4 (—), 7.0 (—), 6.5 (—), 6.0 (—), 5.5 (—), 4.5 (—), 4.0 (—), 3.5 (—), 3.0 (—). **(C) Cross section following at 340 nm during acid (—) and base titration (—).**


A
Emission Max= 337nm
Emission Max= 340nm
Intensity (au)
Wavelength (nm)
B
Emission Max= 335nm
Emission Max= 338nm
Intensity (au)
Wavelength (nm)
C
Intensity (au)
pH

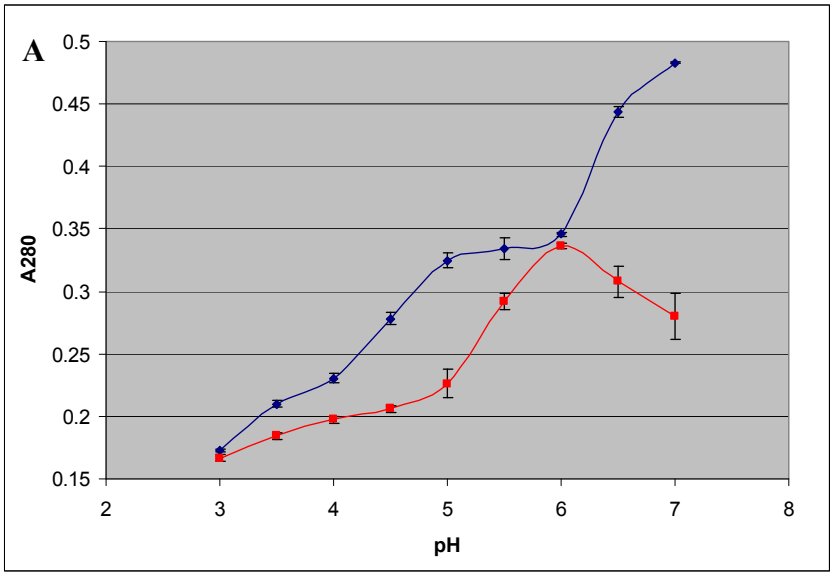
titrations, the absorbance values fluctuated as the pH was varied **(Figures 11A and 11B)** and never returned to their original starting value upon reverse titration.

In order to ensure that the dramatic increase in CD signal during acid titration was indeed due to the change in pH and not to the increased concentration of chloride ions that resulted upon titration with HCl, two additional titrations were preformed: a salt titration and a titration to acid pH using H_2SO_4 as the titrant (data not shown). Titrating with NaCl (Figure 12) revealed a slight increase in negative ellipticity (and presumably helical secondary structural content) as the salt concentration increased, but nothing as dramatic as what was observed upon acid titration with HCl (Figure 7*)*.

4.4. Secondary and Tertiary Structure Estimation of 'Anomalous' Preparations

Occasionally, NP preparations (dubbed 'anomalous' preparations) gave CD spectra that looked very different than what was normally observed for NP at pH 7.4 before acid titration. The shape of this "anomalous spectrum" persisted during sample acidification until pH 5.5 although the intensity of the signal (degree of negative ellipticity) decreased with increased solution acidity. The data thus indicate that NP protein was behaving as one would expect for most proteins upon acidification and gradually unfolding as the positive charge on the protein increased. However, once pH 5.5 was reached, the anomalous NP spectra reverted to a 'normal' profile resembling that seen with the "predominant" preparations at lower pH values and steady increases in negative ellipticity at 222nm with decreasing pH were again observed. The original anomalous spectrum is not seen after reverse titration with NaOH, rather the spectra appeared similar to those obtained for a 'predominant' preparation (Figure 13A-C).

Figure 11: Absorbance measured at 280nm (A) and 214nm (B) during predominant NP titration with HCl (—) and NaOH (—). Triplicate measurements were taken at each pH value in a 1.0cm cuvette.


A
A280
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
2
3
4
5
6
7
8
pH

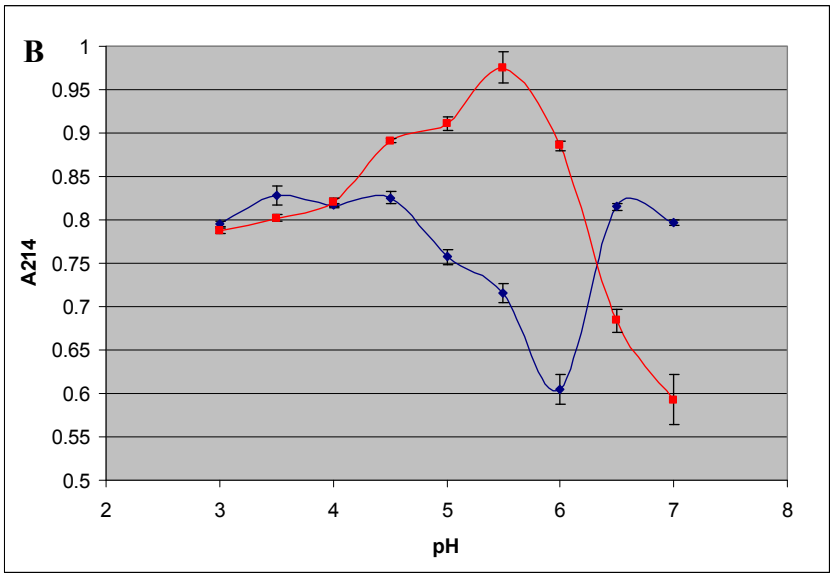

B
A214
1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
2
3
4
5
6
7
8
pH

Figure 12: NP from a 'predominant' preparation titrated with 5M NaCl at pH 7.4. Final NaCl concentrations were as follows: 0.0M (—), 0.1M (—), 0.2M (—), 0.3 M (—), 0.4M (—), 0.5M (—), 0.6M (—), 0.7M (—), 0.9M (—), 1.0M (—). Samples were allowed to equilibrate five minutes prior to taking CD measurements.

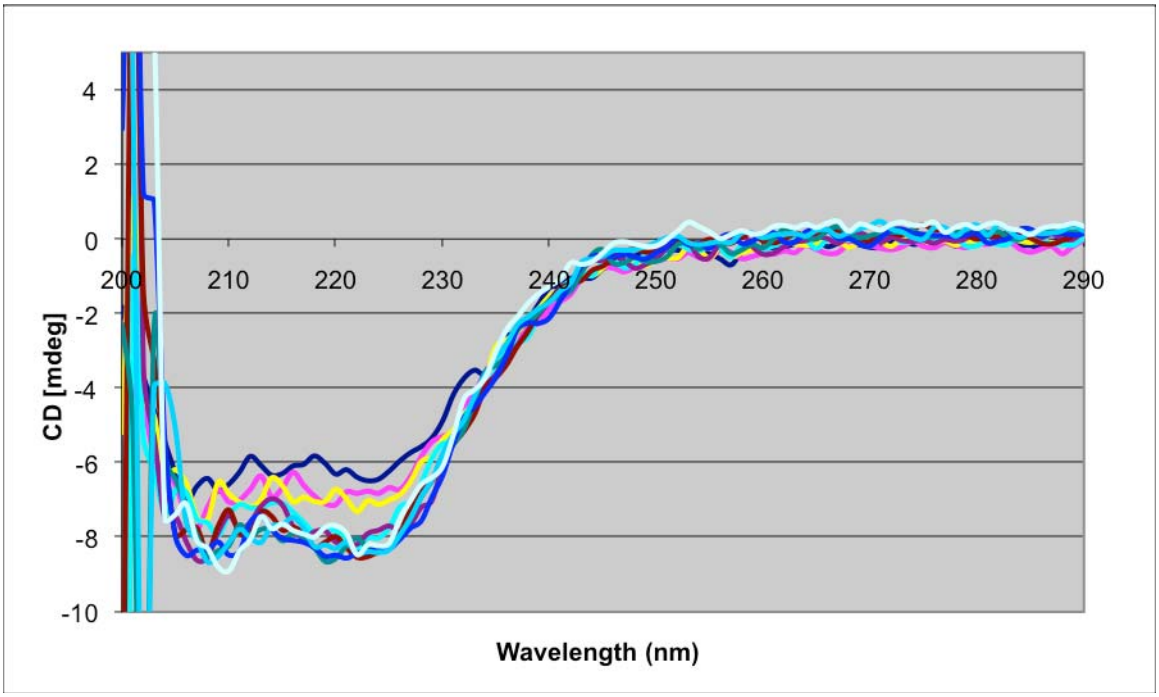
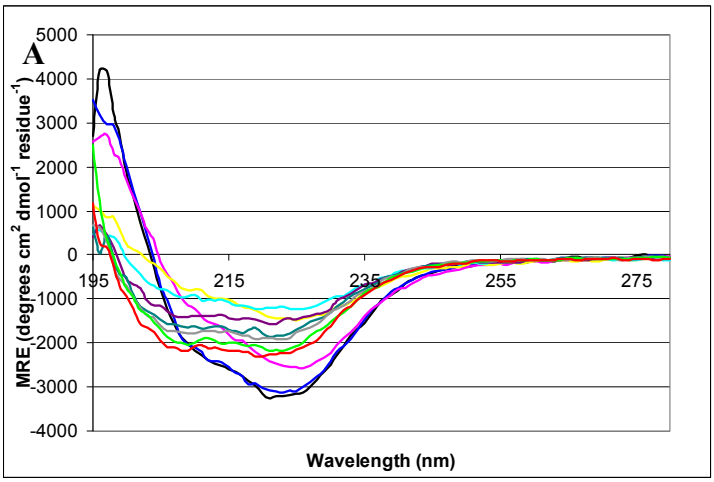
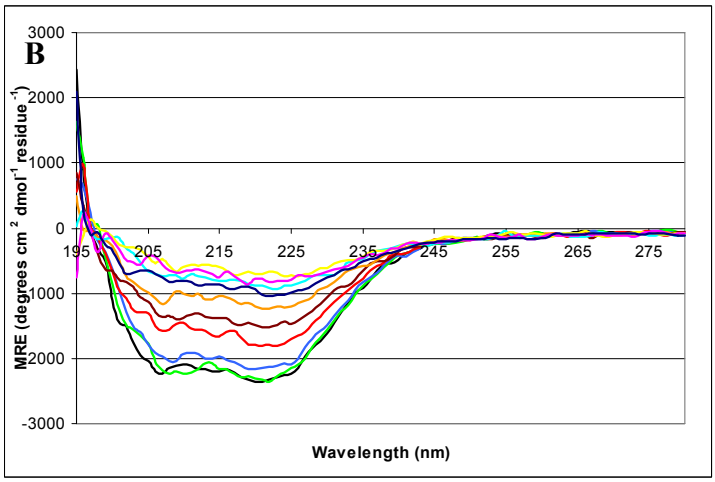
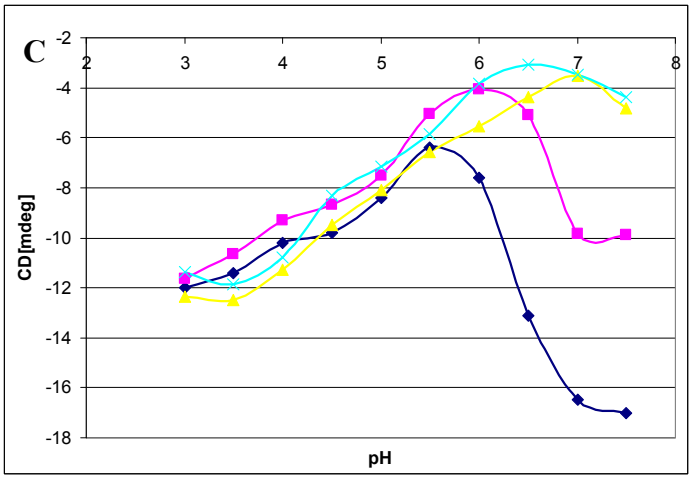

4
2
0
-2
-4
-6
-8
-10
200
210
220
230
240
250
260
270
280
290
CD [mdeg]
Wavelength (nm)

Figure 13: Representative CD spectra obtained for anomalous NP preparations when titrated with 1M HCl (A) and 1M NaOH (B). (A) Upon adjusting the pH from 7.5 to 5.5 a decrease in mean residue ellipticity (MRE) was observed. Once the pH was adjusted below 5.5, an increase in MRE was observed as previously described. (B) Upon back titration with NaOH, a decrease in MRE is observed as the pH is raised from 3.0 to 7.4. The original anomalous spectrum initially observed is never recovered. (A) pH: 7.5 (—), 7.0 (—), 6.5 (—), 6.0 (—), 5.5 (—), 5.0 (—), 4.5(—), 4.0 (—), 3.5 (—), 3.0 (—). (B) pH: 3.0 (—),3.5 (—) 4.0 (—), 4.5(—), 5.0 (—), 5.5 (—), 6.0 (—), 6.5 (—), 7.0 (—), 7.5 **(—). (C) Cross sectional view of maxima at 222nm and 208 nm throughout 'anomalous' NP titration.** (▪) HCl titration following 222nm, (▪) NaOH back titration following 222nm, (▪) HCl titration following 208nm, (▪) NaOH back titration following 208nm.


A
MRE (degrees cm 2 dmol-1 residue-1)
5000
4000
3000
2000
1000
0
-1000
-2000
-3000
-4000
195
215
235
255
275
Wavelength (nm)


B
MRE (degrees cm 2 dmol-1 residue-1)
3000
2000
1000
0
-1000
-2000
-3000
195
205
215
225
235
245
255
265
275
Wavelength (nm)


C
CD[mdeg]
-2
-4
-6
-8
-10
-12
-14
-16
-18
2
3
4
5
6
7
8
pH

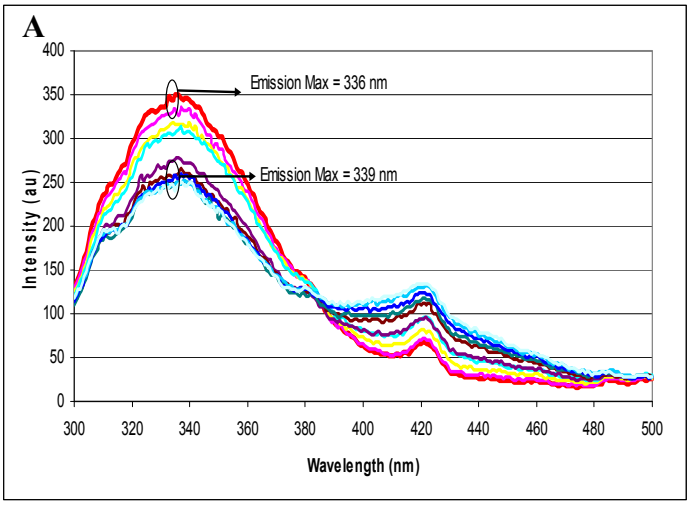
Fluorescent emission spectra were recorded throughout the titration in order to assess tertiary structure of the anomalous NP preparations. Characteristic fluorescence profiles of NP during titration with sulphuric acid from pH 7.4 to pH 3.0 and then back to neutral pH with NaOH is shown (Figure 14A and 14B). As was the case for the predominate preparation, a red shift is observed upon decreasing the pH, with the wavelength shifting from 336nm to 339nm. Upon addition of base, however, the maximum did not shift at all, and stayed at 338nm throughout the titration. The changes in emission intensity at 340nm are shown cross-sectionally in Figure 14C.

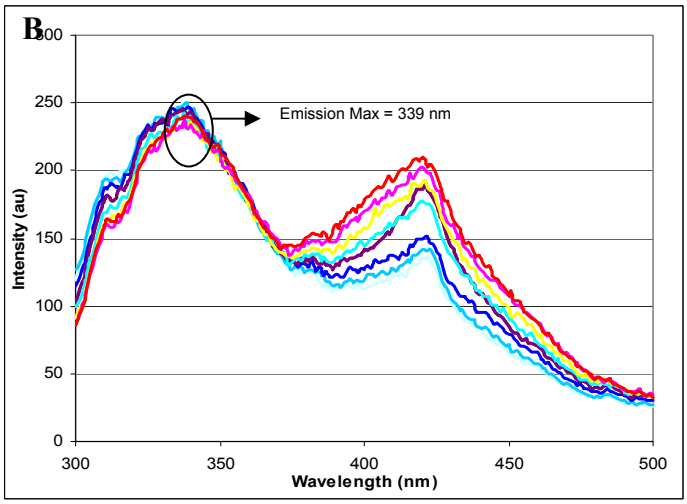
Absorbance at 280nm and 214nm for each of the samples was recorded to gauge the concentration after completing CD and fluorescence measurements. During the titrations, the absorbance value obtained at either wavelength fluctuated as the pH varied (Figure 15A and 15B). The absorbance values never returned to their original starting value even when the original pH was once again obtained.

4.5. RNA Interaction with NP: RNase A Digest

Comparison of the predominant to the anomalous spectrum indicates approximately the same degree of ellipticity at 208nm (Figure 6); however, the anomalous form has much deeper signal at approximately 220nm. An increase in helical content in a protein would be expected to increase negative ellipticity at 220nm with the concomitant increase at 208nm and increase in sheet or random coil content would be expected to increase negative ellipticity at lower wavelengths (202-210nm) and decrease the depth of signal at 220nm. Thus, the ‘anomalous’ shape can not entirely be attributed simply to change in the relative proportions of typical protein secondary structural elements and alternate explanations for the anomalous spectrum needed to be considered. NP functions to bind RNA and it has been

Figure 14: Characteristic fluorescence profiles of NP from an 'anomalous' preparation during titration with H_2SO_4 (A) and back titration with NaOH (B). (A) Titration of an anomalous NP preparation from pH 7.4 to pH 3.0 followed fluorescence spectroscopy. (B) Titration of an 'anomalous' NP preparation from pH 3.0 to pH 7.0 followed by fluorescence spectroscopy. pH: 7.4(—), 7.0(—), 6.5(—), 6.0(—), 5.5(—), 5.0(—), 4.5(—), 4.0(—), 3.5(—), 3.0(—). **(C) Cross section following at 340nm during acid (—) and base titration (—).**


A
Emission Max = 336 nm
Emission Max = 339 nm
Intensity (au)
Wavelength (nm)
400
350
300
250
200
150
100
50
0
300
320
340
360
380
400
420
440
460
480
500


B
Emission Max = 339 nm
Intensity (au)
Wavelength (nm)
300
250
200
150
100
50
0
300
350
400
450
500

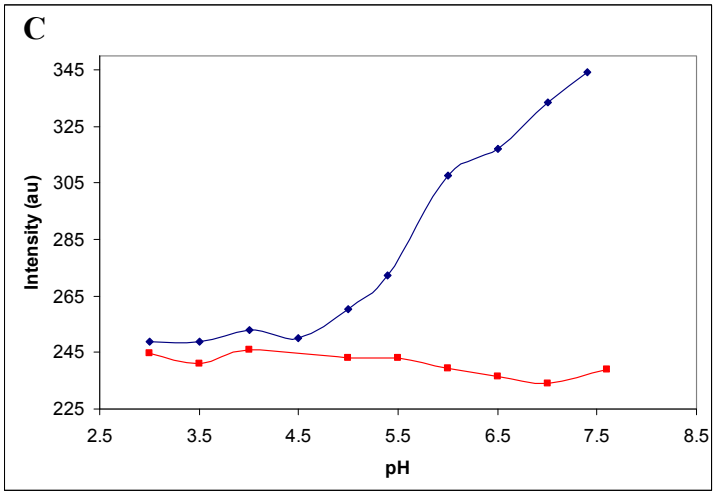
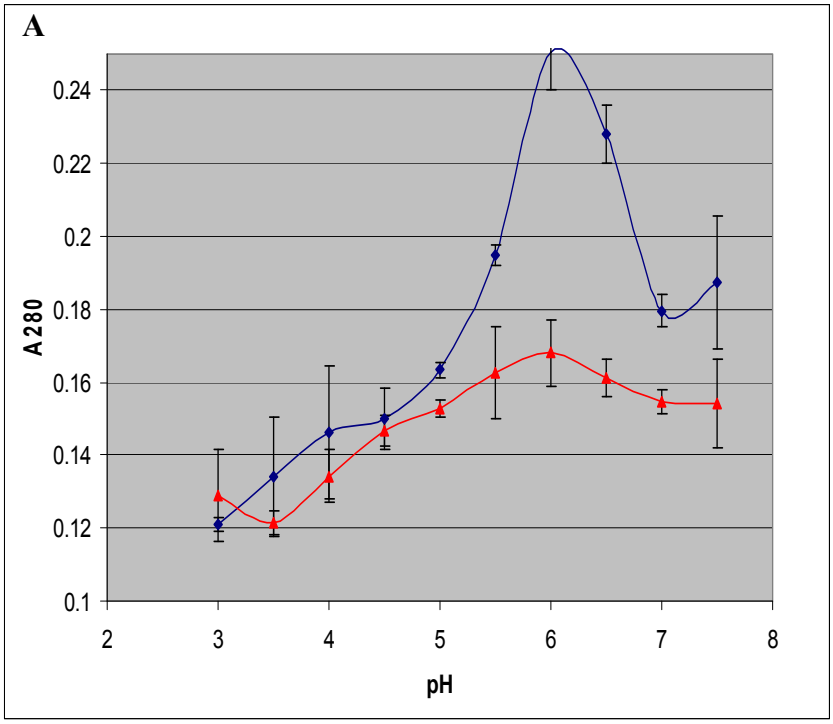
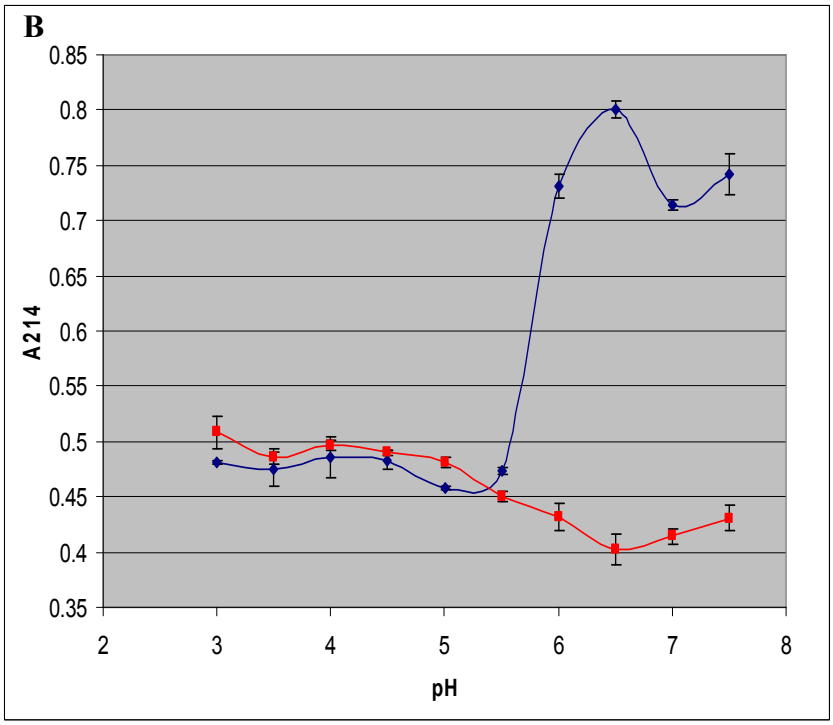

C
Intensity (au)
pH
345
325
305
285
265
245
225
2.5
3.5
4.5
5.5
6.5
7.5
8.5

Figure 15: Absorbance measured at 280nm (A) and 214nm (B) during 'anomalous' NP titrations with HCl (—) and NaOH (—). Measurements were recorded three times per sample at each pH value using a 1.0 cm path length cuvette.


A
A280
0.24
0.22
0.2
0.18
0.16
0.14
0.12
0.1
2
3
4
5
6
7
8
pH


B
A214
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
0.45
0.4
0.35
2
3
4
5
6
7
8
pH

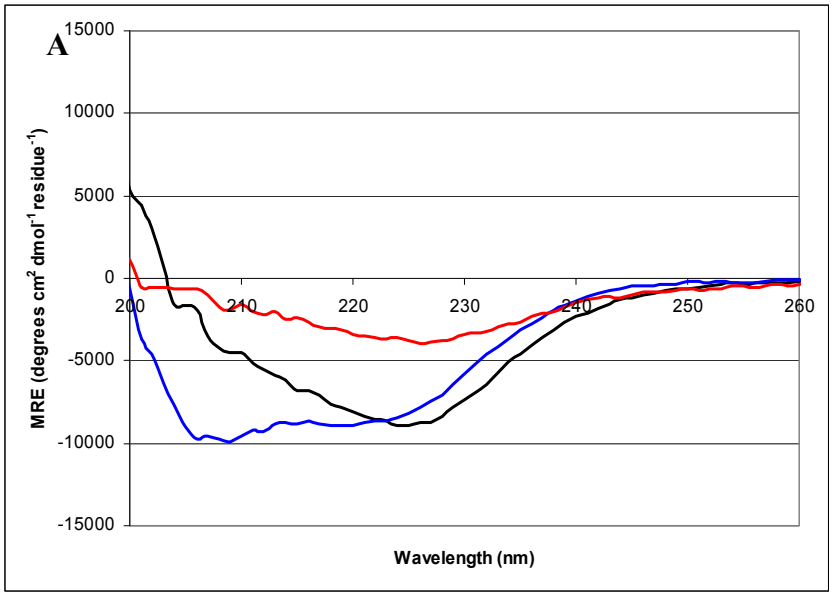
shown that some conformations of nucleic acids can give CD spectra with negative ellipticity in the region observed in spectrum of anomalous NP preparations [69]. Therefore, it was hypothesized that the 'anomalous' preparations result from cellular RNA binding NP during recombinant protein production or isolation process.

This hypothesis was tested when a second preparation gave an anomalous CD spectrum at neutral pH: RNase A was added to part of this preparation to determine if the RNA was indeed present and bound to NP. Comparison of the CD spectra obtained for an 'abnormal' preparation with and without RNase A treatment (Figure 16A and 16B) suggest that the atypical CD spectrum results from binding of NP protein to RNA.

4.6. RNase Free Preparations

To further test the hypothesis that RNA binds to NP and induces a conformational change (or at least results in the anomalous spectra observed at pH 7.4 for some preparations of NP), a small molecule RNase inhibitor was used during the purification of recombinant NP to inhibit the action of endogenous RNases during that process. The rationale for this experiment was that the RNA would stay intact, resulting in a protein conformation/complex that yields a CD spectrum reminiscent of the 'anomalous' preparation. RNA-vanadyl complex was added into preparations in two different ways. The first set of experiments added the vanadyl complex into the lysis buffer, while the second set of experiments had the vanadyl complex in all buffers used throughout the preparation. In both cases, an expected band at the corresponding molecular weight of 56kDa was obtained (Figure 17A and 17B). In the one comparison made, it was observed that vanadyl complex added only to the lysis buffer, resulted in less NP (Figure 17A) than when the inhibitor was added into all the buffers during isolation (Figure 17B). Addition of vanadyl complex at any stage during preparation appeared to result in

Figure 16: Effects of RNA on NP secondary structure. An anomalous preparation at neutral pH as indicated by initial CD room temperature scan was divided into two samples: one untreated sample (A) and one treated with RNase A to remove any residual RNA (B). After 18 hours, the spectra were recorded (—). The pH was then adjusted and a scan was taken at pH 3.0 using concentrated HCl (—). The pH was then re adjusted to neutral pH before reading the final CD spectrum (—).


A
15000
10000
5000
0
-5000
-10000
-15000
200
210
220
230
240
250
260
MRE (degrees cm2 dmol-1 residue-1)
Wavelength (nm)

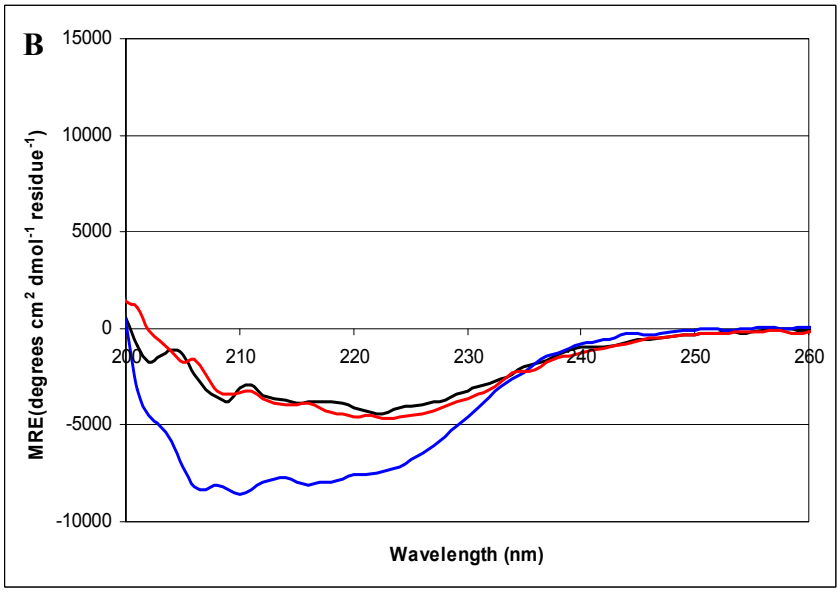
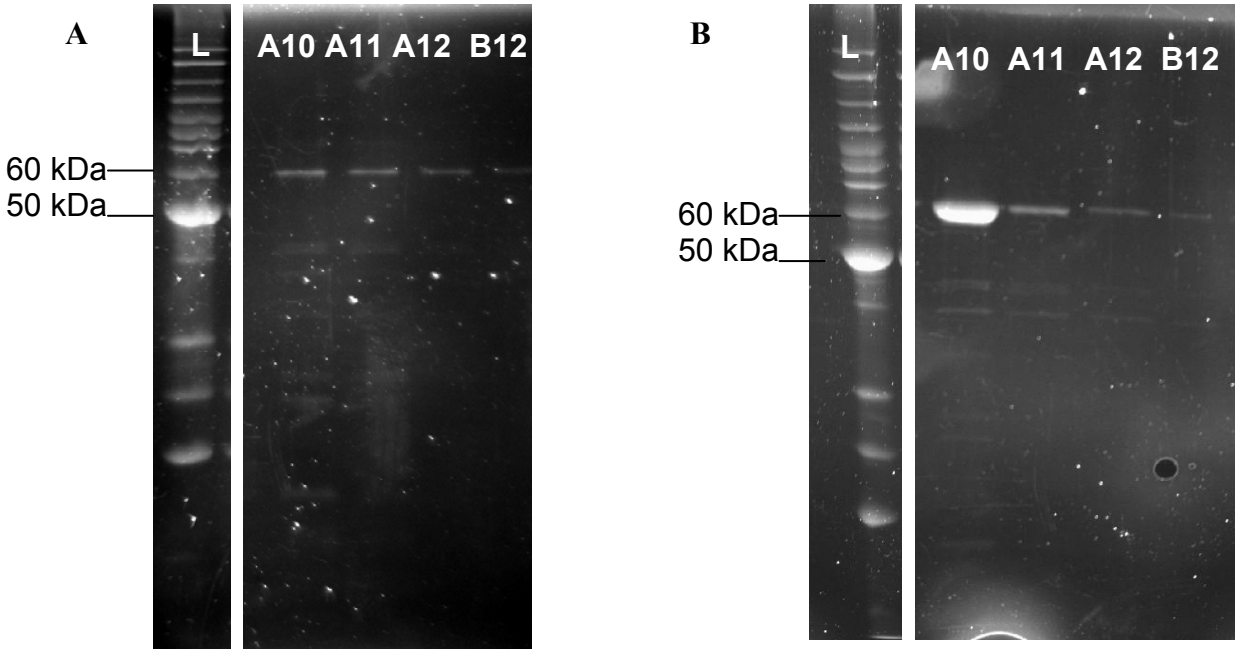
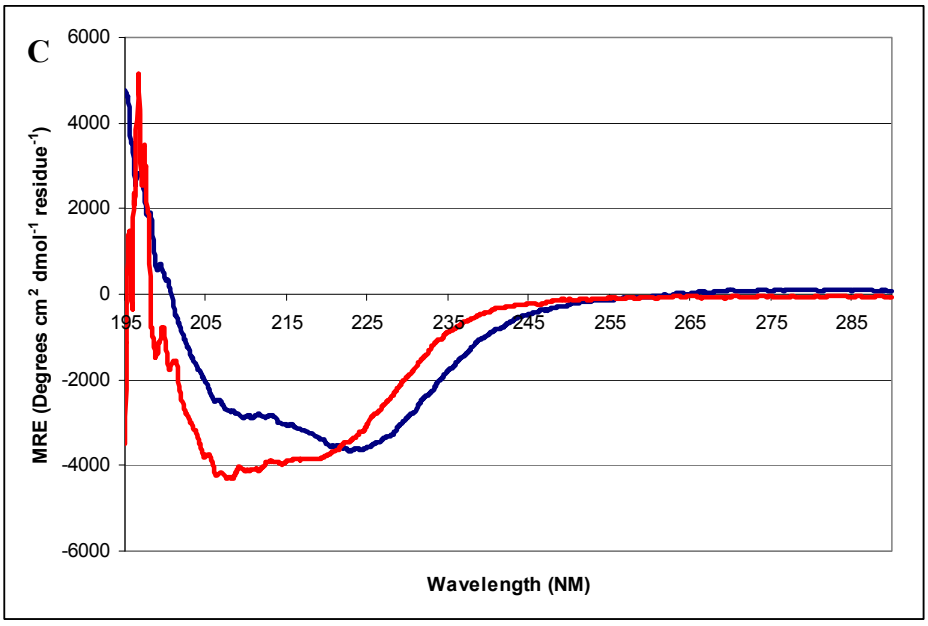

B
15000
10000
5000
0
-5000
-10000
200
210
220
230
240
250
260
MRE(degrees cm2 dmol-1 residue-1)
Wavelength (nm)

Figure 17: Sypro™ stained SDS-PAGE analysis of purified NP (elution fractions from Ni-NTA column, fractions A10-B12) prepared with vanadyl complex in only the lysis buffer (A) and in all buffers (B) during isolation. Both methods result in a band at appropriate molecular weight. **(C)** Representative CD spectrum of undigested NP after isolation in the presence of vanadyl complex (—). (—) Post RNase A digest results in a spectrum resembling the 'predominant preparation'.


A
L
A10 A11 A12 B12
60 kDa
50 kDa
B
L
A10 A11 A12 B12
60 kDa
50 kDa


C
6000
4000
2000
0
-2000
-4000
-6000
195
205
215
225
235
245
255
265
275
285
MRE (Degrees cm2 dmol-1 residue-1)
Wavelength (NM)

higher yields of NP than in preparations without the vanadyl complex. Moreover, when the vanadyl complex was added at any point in the preparation, the anomalous spectrum with a deep valley at 222 nm was obtained on CD analysis after isolation and purification (Figure 17C), a result consistent with the hypothesis that the anomalous spectrum was the result of RNA binding NP. A sample of the anomalous preparation obtained was then digested overnight with RNase A as described earlier, and the CD spectrum after digestion recorded. This resultant spectrum was observed to resemble those obtained from the predominant preparations (Figure 17C), again consistent with the hypothesis that anomalous preparations contained bound RNA**.**

4.7. Deliberate Addition of RNaseA to NP Preperations

To further investigate the interaction between NP and RNA, the reverse experiment described above was performed. The addition of RNaseA was expected to yield NP that gave a 'predominant' CD spectrum. When RNase A was added to the lysis solution to a final concentration of 4µg/mL and allowed to incubate for two hours after sonication, there was no recovery of NP as determined by SDS-PAGE (Data not shown). The chromatogram from the FPLC did not show a UV peak indicating that no protein had bound to the nickel column. This isolation procedure with RNase A was attempted four times, and every attempt resulted in a failure to isolate NP.

4.8. RNA Titration

RNA-free preparations giving a 'predominant' CD spectrum were titrated with eukaryotic RNA to see if the NP would bind the nucleic acid and an 'anomalous' CD spectrum could thereby be obtained. NP was titrated with increasing concentrations of RNA at both pH 7.0 and pH 5.5 (neutral pH and the point at which the anomalous spectrum began to behave as a predominant preparation, respectively). In both

titrations, no change in the shape of the CD spectrum was observed even at concentrations of RNA that were calculated to result in 10-fold excess of nucleic acid binding units. Thus, NP folded in the absence of RNA did not appear to be able to bind any RNA. (Data not shown).

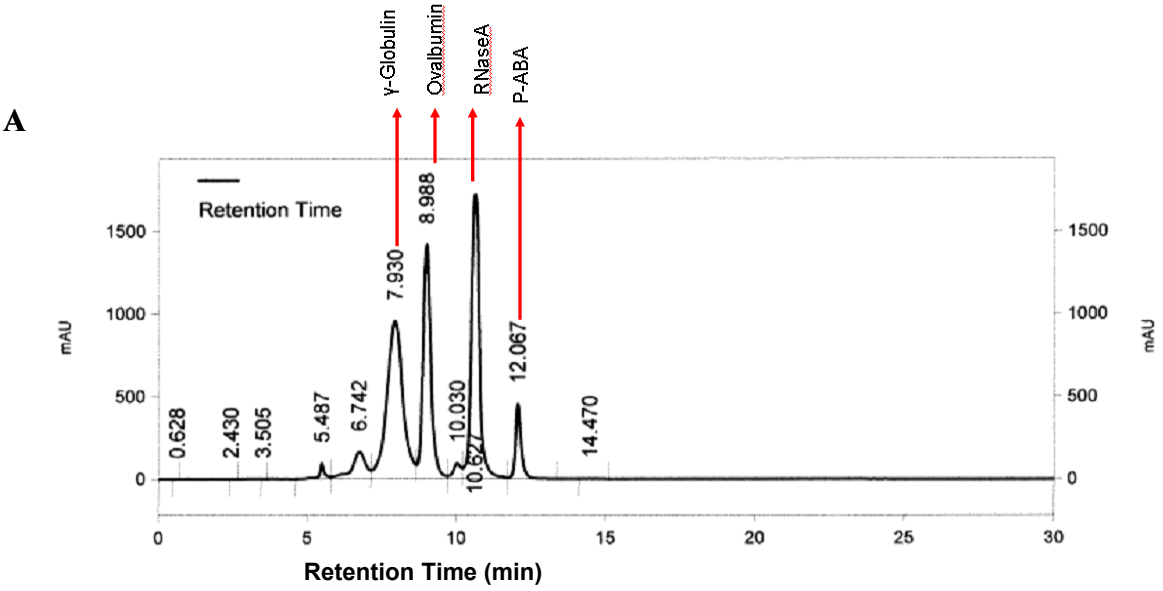
4.9. Quaternary Structural Determination of NP

In order to assess the quaternary structure under the experimental conditions used throughout the titrations and RNase A digests, SEC-HPLC and dynamic light scattering were used. Size exclusion chromatography was performed on a High Performance Liquid Chromatography (HPLC). Results from size exclusion chromatography of fresh NP preparation gave only one peak with a retention time between that observed for RNaseA and P-ABA. These two molecules are substantially smaller than the monomeric form of the NP protein at 56kDa. When the same chromatography was executed with 0.1% SDS present in the running buffer (to prevent the protein from sticking to the column), the same results were obtained (Figure 18). The unexpected retention time on SEC-HPLC could not be attributed to sample degradation as SDS-PAGE analysis of the same SEC-HPLC sample resulted in the observation of a band corresponding to a protein of size of 56kDa and no observable breakdown products. These data suggest that perhaps, even in the presence of 0.1% SDS, NP is sticking to the column or that it is assuming an extremely compact conformation with a much smaller than expected Stoke's radius.

Particle sizing using dynamic light scattering was utilized as an alternate approach to determine the size of NP. Freshly prepared NP was immediately sent for particle sizing at various concentrations. As shown in (Figure 19 A-D), the sizes

Figure 18: **Quaternary structure determination of NP at pH 7.4 through SEC-HPLC. (A)** Various protein standards applied to an equilibrated SEC column. Larger proteins had the shortest retention time (ie γ-Globulin) while smaller molecules took more time to traverse the column (ie P-ABA). (B) NP gives a peek between RNaseA (molecular weight of 14 kDa) and P-ABA, two molecules substantially smaller than the expected molecular weight of 56 kDa. There appears to be no sign of multimers in the chromatogram.

A


γ-Globulin
Ovalbumin
RNaseA
P-ABA
Retention Time
mAU
1500
1000
500
0
0.628
2.430
3.505
5.487
6.742
7.930
8.988
10.030
10.527
12.067
14.470
0
5
10
15
20
25
30
Retention Time (min)

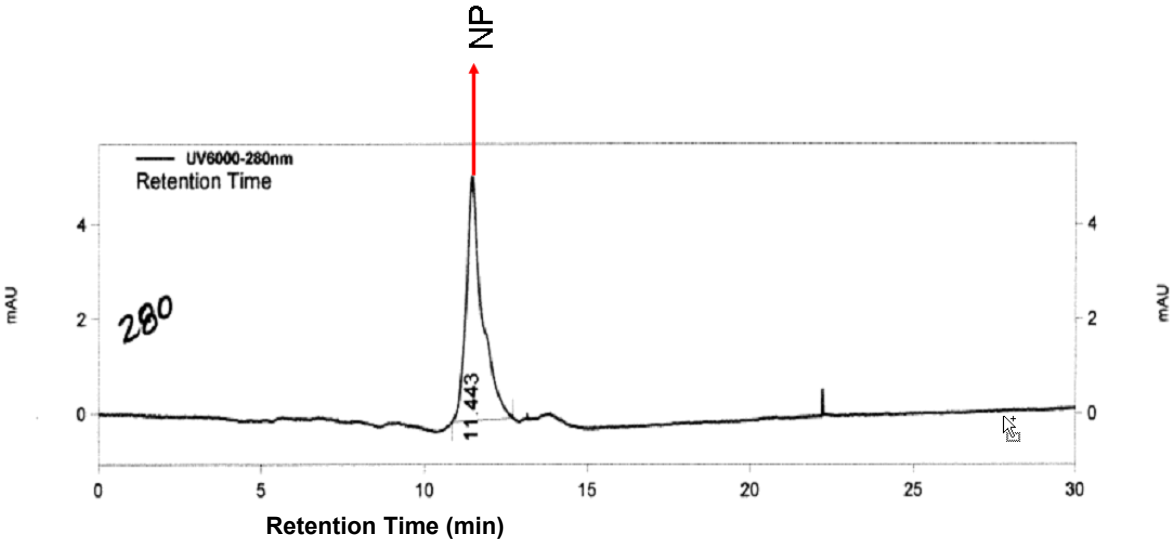
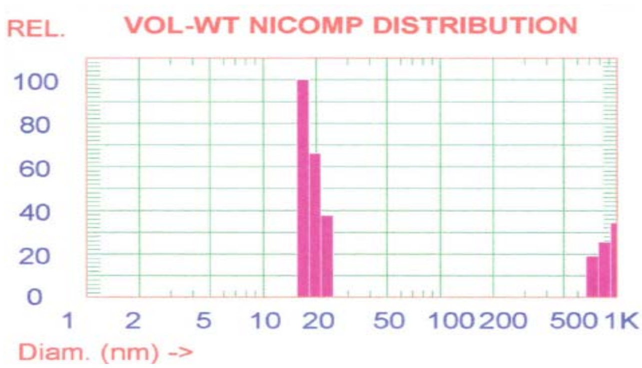
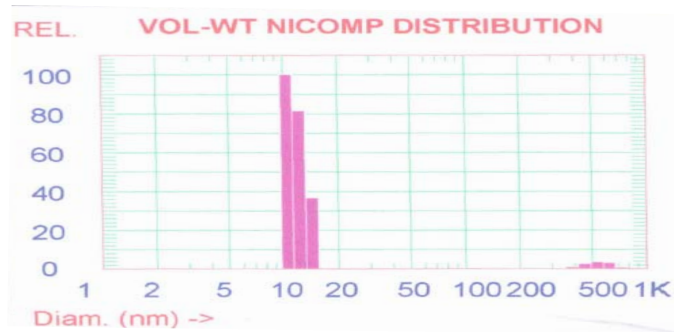

NP
UV6000-280nm
Retention Time
280
mAU
4
2
0
11.443
0
5
10
15
20
25
30
Retention Time (min)

Figure 19: Dynamic light scattering conducted at varying concentrations of NP. (A) Undiluted (B) 1:2 dilution (C) 1:3 Dilution (D) 1:4 Dilution. In all cases particles substantially larger than what was expected for NP was observed. As the sample became more diluted, the population of larger molecules decreased, while an increase in smaller molecules was observed. Samples were diluted with 10mM Na_3PO_4 at pH 7.4

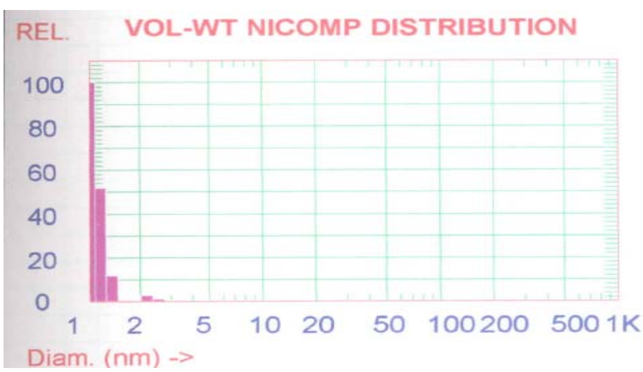
A Undiluted



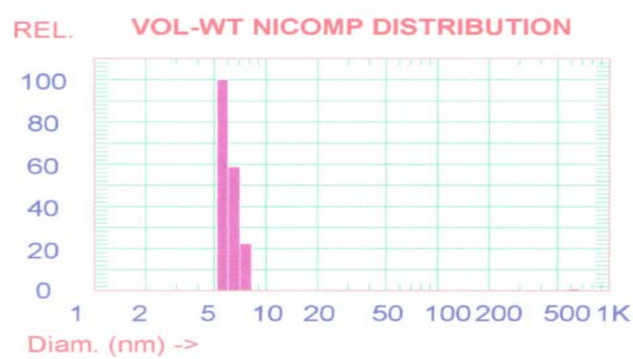
B 1:2 Dilution



C 1:3 Dilution



D 1:4 Dilution

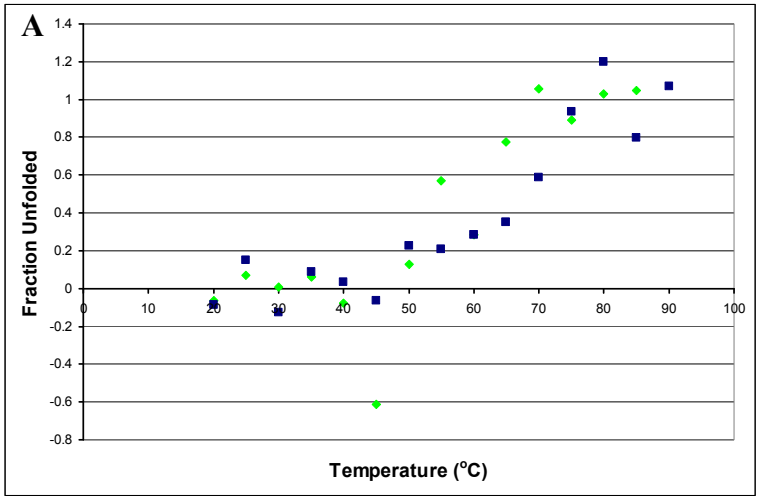


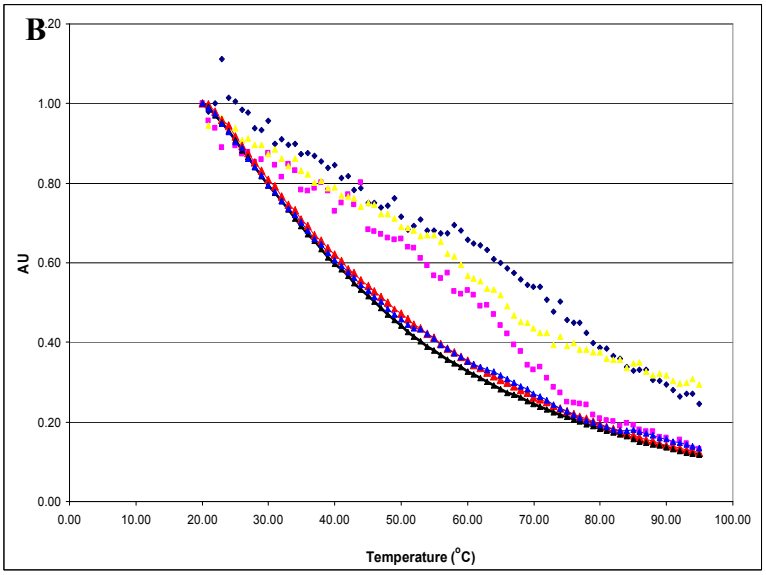
obtained are substantially greater than that anticipated for a protein. Interestingly, a large population at 500nm observed at the 1:1 ratio begins to decrease upon dilution, while the population at the smaller diameters begin to increase. This data has not yet been reproduced.

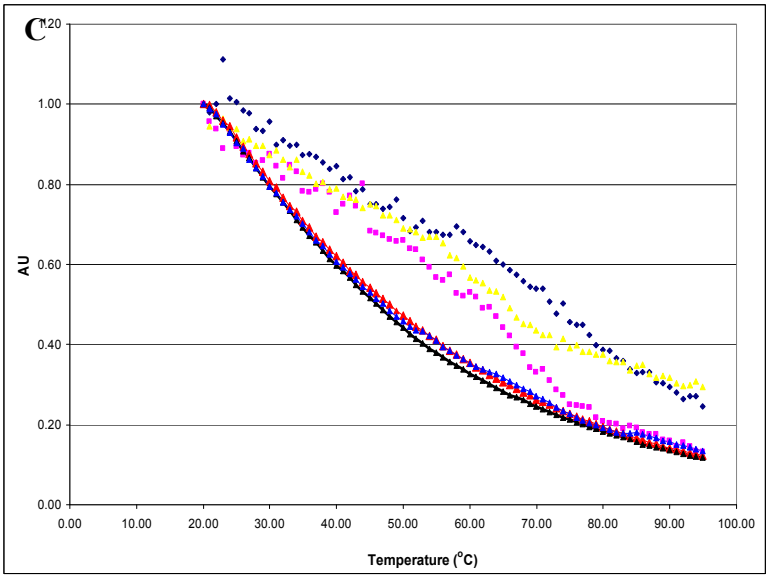
4.10. Stability of Secondary and Tertiary Structure Analyzed Through CD and Fluorescence

The effect of salt on the secondary and tertiary structure on both the 'predominant' and 'anomalous' forms of NP was examined using thermal denaturation assays followed by CD and fluorescence, respectively. Far UV CD (secondary structure) studies showed that the T_M of the 'predominant' preparation at pH 7.0 decreased when salt (NaCl) was added to a concentration of 150mM (Figure 20 A). Normalized protein tryptophan fluorescence emission intensity at this same pH decreased with increasing temperature. However, it is well known that the quantum yield of Trp fluorescence is temperature dependent, even for Trp free in solution [70]. Therefore the effect of salt on quantum yield with increasing temperature must be viewed relative to the loss seen with free Trp in solution. With tryptophan free in solution, a consistent, but non-linear loss of quantum yield with increasing temperature is observed (Figure 20 B, black line). On the other hand, the plot of quantum yield as a function of increasing temperature from tryptophan in NP of the 'predominant' form at pH 7.0 produces a biphasic curve. At lower temperatures, where the protein is expected to be folded, the loss in quantum yield is actually retarded relative to that seen with tryptophan free in solution. At higher temperatures, however, the loss of quantum yield for NP in the 'predominant' form is

Figure 20: Secondary and tertiary structural analysis of 'predominant' NP preparations analyzed through CD (A) and fluorescence spectroscopy (B and C). (A) Thermal denaturation were performed on 'predominant' NP preparations under varying salt conditions ((■) 0mM NaCl and (■) 150mM NaCl)). Samples were heated from 20°C to 95°C. (B and C) Normalized tryptophan fluorescence emission intensity as a function of heat (25°C to 95°C) for 'predominant' NP preparations at pH 7.0 (B) and pH 3.0 (C) at varying salt concentrations (♦ 0M NaCl, ■0.5 M NaCl, ▲1.0M NaCl). NP samples at both pH levels were compared to free tryptophan in solution (0.014mg/mL) at varying salt concentrations (►0M NaCl, ►0.5M NaCl, ►1.0M NaCl).


A
Fraction Unfolded
Temperature (°C)


B
AU
Temperature (°C)


C
AU
Temperature (°C)

actually accelerated relative to that seen with free tryptophan in solution (Figure 20B, navy line).

For tryptophan free in solution, increasing the salt concentration from 0M to 0.5M NaCl produces a slight delay in loss of quantum yield with temperature (Figure 20B, black line to red line). A further increase in salt concentration from 0.5M to 1.0M NaCl had little further effect (Figure 20B, red line to light blue line). For Trp in NP in the "predominant" form, the addition of salt has very little effect on the quantum yield at low temperatures. However, at intermediate temperatures and at higher temperatures, increasing salt appears to increasingly shield the tryptophan fluorescence from a loss in quantum yield (Figure 20B, pink squares, yellow triangles). At 0.5M and 1.0M NaCl, the break in the quantum yield curve for NP in the 'predominant' form occurs at lower temperatures than in the absence of salt.

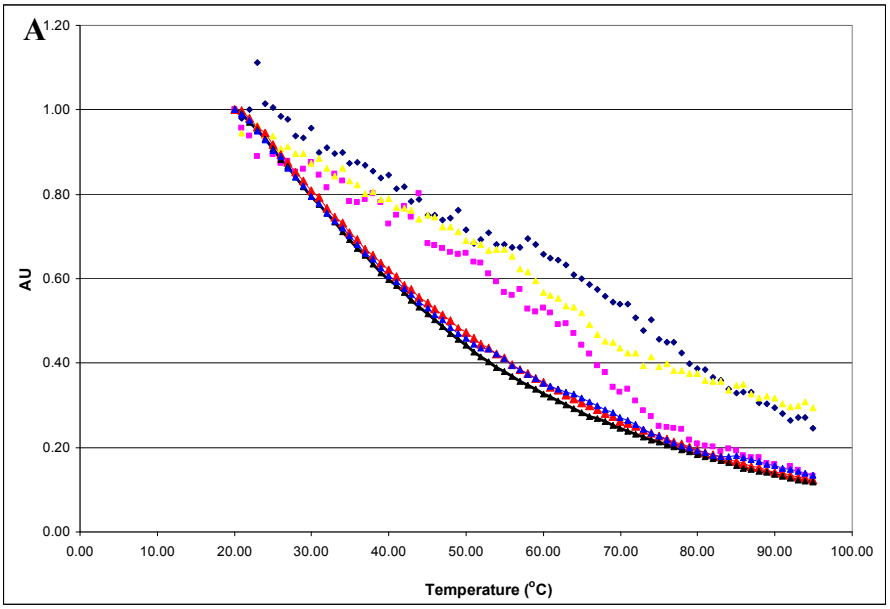
At low temperatures, the fluorescence results at pH 3.0 are very similar to those seen for the 'predominant' form at pH 7.0: the loss of quantum yield with increasing temperature is less dramatic than that seen for tryptophan free in solution. However, at higher temperatures there is much less loss of quantum yield at pH 3.0 than at pH 7.0 and the observed fluorescence remains higher than for tryptophan free in solution. As was the case for the 'predominant' form at pH7.0, salt appears to have little effect at lower temperatures at pH 3.0. However, at higher temperatures, addition of 0.5M NaCl delayed the temperature-induced onset of loss of tryptophan emission intensity relative to that seen in the absence of salt (Figure 20 C, pink squares vs navy diamonds). However, further addition of salt (1.0M NaCl final, Figure 20 C, yellow triangles) seemed to reverse this effect to some extent.

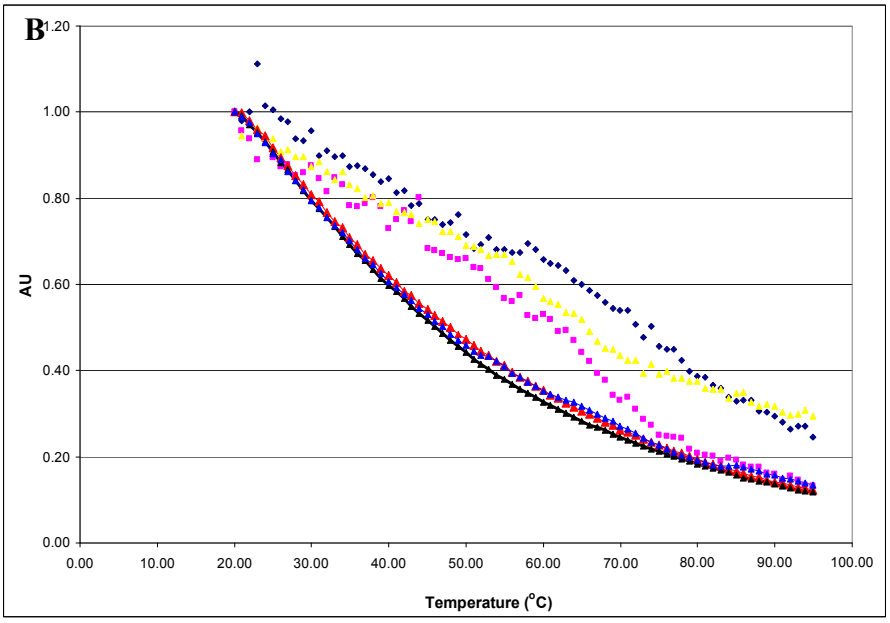
The effect of salt on the temperature dependence of tryptophan fluorescence emission intensity in 'anomalous' preparations was quite different than that observed for 'predominant' preparations. At neutral pH, addition of 0.5M NaCl hastened the onset of loss of emission intensity relative to preparations in which there was no salt (0M NaCl) (Figure 21 A). An intermediate effect was observed at 1.0M NaCl. At pH 3.0, the intensity versus temperature curve (Figure 21 B) of anomalous preparation with 0.5M NaCl was very similar to that containing no salt, whereas, the addition of 1.0M NaCl shifted the curve to the left (onset of loss of emission intensity occurring at lower temperatures). At neural pH (0.5M NaCl, pH 7.0, for example), the plots of quantum yield versus temperature for the 'anomalous' preparation actually appeared to be triphasic rather than biphasic. The same effect is not seen when the pH is adjusted to 3.0.

4.11. Concentration Estimation of NP Preparations

As noted above, the estimation of concentration using absorbance measurements at either 280nm (aromatic residue absorbance) or 214nm (peptide bond) for any one sample varied when the pH of that sample was changed. In addition, absorbance values for the same sample at the same pH value (and hence the concentration estimates based on those values) were different before and after of titration of the sample with acid and restoration of neutral pH by the addition of base (Figures 11A and B and 15 A and B for 'predominant' and 'anomalous' preparations, respectively). Estimations of concentration based on absorbance values and theoretical extinction coefficients also differed from those obtained from BCA assay and sometimes those estimates differed by almost a factor of ten (Table 1). In order to determine which, if any, estimation was more accurate, one sample

Figure 21: Tertiary structural analysis of 'anomalous' NP preparations analyzed through fluorescence spectroscopy at pH 7.0 (A) and pH 3.0 (B). Normalized tryptophan fluorescence emission intensity as a function of heat (25°C to 95°C) for 'anomalous' NP preparations at pH 7.0 (A) and pH 3.0 (B) at varying salt concentrations (♦ 0M NaCl, ■0.5 M NaCl, ▲1.0 M NaCl). NP samples at both pH levels were compared to free tryptophan in solution (0.014mg/mL) at varying salt concentrations (►0M NaCl, ►0.5M NaCl, ►1.0M NaCl).


A
1.20
1.00
0.80
0.60
0.40
0.20
0.00
AU
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
Temperature (°C)


B
1.20
1.00
0.80
0.60
0.40
0.20
0.00
AU
0.00
10.00
20.00
30.00
40.00
50.00
60.00
70.00
80.00
90.00
100.00
Temperature (°C)

was sent for quantitative amino acid analysis. The result (Appendix 3) indicated that BCA assay gave a more reliable estimation of NP concentration. Thus BCA assay results were used in concentration-dependent calculations.

Table 1: Comparison table displaying the different estimations of NP concentration based on BCA assay versus absorbance values using a theoretical extinction coefficient.

NP concentration determination by BCA vs A_{280} *		
	Concentration determined by BCA	**Concentration determined by A_{280}**
NP Prep 18 Fraction A 10	**0.05 mg/mL**	**0.31mg/mL**
NP Prep 18 Fraction A 11	**0.01mg/mL**	**0.17mg/mL**
NP Prep 18 Fraction A 12	**0.00mg/mL**	**0.08mg/mL**

***A_{280}=εcl where ε = 54235, c = concentration, l = pathlength of cuvette (1.0 cm)**

5.0 Discussion

The main objective of this thesis, in keeping with the regulatory mandate of the Center for Vaccine Evaluation at Health Canada, was to characterize the Nucleocapsid Protein (NP) antigen from the avian H5N1 influenza strain using physico-chemical methods in order to gain insights into the properties of the antigen that might influence the quality, safety and efficacy of a therapeutic vaccine. NP is an attractive therapeutic antigen because it is an internal protein that, because it is fairly conserved across all influenza strains, presents a potential for a universal influenza vaccine. A new, therapeutic vaccines based on the viral NP candidate antigen, while not prophylactic, would aid in the treatment of infected individuals by eliciting a Cytotoxic T Lymphocyte (CTLs) response [38, 60, 71].

In order to fully characterize NP, the recombinant protein was first produced in *E. coli.* The protein purification process yielded small amount of protein ranging from 0.02mg/mL to 0.07mg/mL. Regardless of the concentration, freshly prepared NP easily aggregated and formed precipitate if stored at 4^{o}C for periods exceeding 10 days. Removal of salt (KCl) from the storage buffer accelerated aggregation. Aggregation and low yield also precluded the routine cleavage of the histidine tag necessary for affinity purification prior to NP characterization. Therefore characterization was continued with the His-tag intact. The propensity of NP to form aggregates is not surprising as the protein's main biological function in the influenza virus is to homo-oligomerize as it binds RNA [24], [23]. Others [6] have succeeded in purifying NP from H5N1 at much higher concentrations (2.5mg/mL) without reported precipitation issues using a maltose-binding protein (MBP) tag and a similar buffer system to that used here but one that included 150 mM NaCl throughout. It may be

that their use of a larger MBP rather than a His-Tag prevents aggregation during the concentration step. NP from influenza A/Hong Kong/1074/99 (H9N2) with a His-tag, on the other hand, was isolated at concentrations sufficient for U.V. detection on FPLC [25]. The ability to retain His-tagged H9N2 NP in concentrated solution when His-Tagged H5N1 NP precipitated may reflect slight differences in the trimeric structures of the two proteins as observed from crystal data [25]. Alternatively, it may be simply that the removal of salt (to facilitate CD analysis for this thesis work) resulted in precipitation of His-tagged H5N1 NP at higher concentrations.

In all, 20 preparations of NP were produced over the course of this thesis work. In each case a band corresponding to a protein of roughly 56kDa was observed on denaturing SDS-PAGE analysis of freshly purified protein (Figure 4 and 5A). Western blot analysis using both anti-His and anti-NP antibodies confirmed the identity of this band as NP protein (Figure 5 A and B). Occasionally, more concentrated fractions of NP contained some anti-NP reactive material of lower molecular size which had previously been demonstrated to be smaller fragments of NP presumably resulting from the breakdown of protein (Bozena Jaentschke, personal communication). In some preparations, other bands were observed at approximately 30 and 20kDa; these bands could also be detected on Western blot analysis with anti-His antibody but not with anti-NP antibody. Excision and analysis by mass spectroscopy indicated these to be proteins originating from *E. coli* and naturally rich in histidine residues (Figure 5C).

After removal of imidazole and KCl used in the preparatory phase, each preparation of NP was immediately analysed by CD spectropolarimetry to ensure that the refolding process used was successful. Twelve of the fifteen (80%) initial

NP preparations (produced without RNase inhibition) resulted in CD spectra with two distinct minima at approximately 222nm and 208nm, reminiscent of spectra from proteins with high α-helical content [56](Figure 6, blue trace). This trace was dubbed the 'predominant' spectrum. Despite the fact that secondary structure deconvolution programs failed to give an estimate of secondary structure content that produced a satisfactory fit to the experimental data, these CD data do indicate that the purified protein was successfully refolded. In contrast to the two peaks of negative ellipticity observed in the spectra of NP samples, CD spectra of random coil/extended protein chains exhibit little or slightly positive ellipticity between 210-240nm and either a single, sharp minimum at approximately 198nm or a single, more shallow minimum centered between 195-205nm [56]. Spectra resembling those expected for random coil conformation could be obtained from NP protein through thermal denaturation (see below) indicating that the spectrum in Figure 6 was the result of a folded structure that was lost upon heating. The available X-ray crystal structure of NP [6] also indicates that the protein is rich in α-helix (48.6 % α-helix, 4.8% β-strand and 46.6% loop). The remaining three preparations (20%) of NP produced without deliberate RNase inhibition resulted in CD spectra with one distinct minimum at 222nm (Figure 6 red trace). The negative ellipticity at 222nm suggests that, similar to the 'predominant' preparations, the protein is largely helical, however the absence of a minimum at 208nm suggests that the conformation of these 'anomalous' preparations differ from that of the 'predominant' preparation. As was the case with the 'predominant' preparations, the available deconvolution programs failed to give accurate estimates of the secondary structural content from the experimental CD data for the 'anomalous' preparations.

As previously noted (Figure 6 and Figure 11A), the large peak in negative ellipticity at approximately 222nm without a concomitant increase in negative ellipticity at 208nm can not be explained simply by increasing the helical content of a protein. Nor can the observed spectral differences between the 'anomalous' and 'predominant' preparations be explained by changing the relative proportion of other common protein secondary structural elements (β-sheet, turn or random coil), all of which give significant negative ellipticity at lower wavelengths [54]. The fact that NP's primary biological function is to bind viral RNA and research indicating that nucleotides (RNA/DNA) in certain conformations can give a characteristic negative ellipticity between 230-250nm [69] led to the hypothesis that the spectrum obtained for 'anomalous' preparations was due to RNA bound to NP, acquired during purification. In order to test this hypothesis, the next NP preparation that gave an 'anomalous' CD spectrum immediately after purification was subjected to RNaseA digestion. Overnight incubation of 'anomalous' NP samples with RNaseA (at a concentration approximately one-fiftieth the concentration of NP determined by BCA assay) resulted in a CD spectrum that was reminiscent of the predominant preparation with a minima at both 222nm and 208nm (FIGURE 11B). The results from RNaseA digestion thus support the hypothesis that NP has the propensity to bind RNA during the purification process, and that RNA binding impacts the CD spectrum and possibly the secondary structure of NP.

To further corroborate the RNaseA digestion data and the notion that RNA was the culprit for the 'anomalous' preparations, two other experimental approaches were pursued involving either the addition of RNase or the inhibition of RNase enzymes during the preparatory phase of NP. In the first approach, RNaseA was

deliberately added into the lysis buffer prior to sonication to completely digest any free RNA. This strategy was expected to result in a preparation that gave the 'predominant' CD spectrum. Interestingly, addition of RNaseA to the lysis buffer precluded the isolation of NP, as indicated by the absence of a band corresponding to the 56kDa MW band on SDS-PAGE immediately following affinity chromatography (data not shown). Four preparations were attempted with this initial RNaseA treatment, and all four yielded no protein, suggesting that at least small amounts of RNA may be required for successful isolation of NP. Conversely, four preparations were executed in an RNase-free environment utilizing DEPC treated water, sterile pipette tips, an established RNase-free working area and the addition of a small molecule RNase inhibitor (vanadyl ribonucleoside complex) [66] to the lysis solution. It was expected that, with the inhibition of RNase enzymes, the CD spectrum recorded immediately after isolation would be the 'anomalous' spectrum. This was indeed the case: 100% of the preparations including the inhibitor gave 'anomalous' spectra, further validating the propensity of NP to bind RNA and giving an alternate NP conformation detectable by CD (Figure 17). Attempts to reproduce the "anomalous" form of NP by adding total yeast RNA to RNA-free NP were unsuccessful. While it has been suggested that influenza NP binds RNA in a non-sequence-specific fashion [26], demonstrations of RNA binding to influenza NP by this group [27] and others [25] have been accomplished only when using a virus-specific oligonucleotide sequence. Further experiments using influenza-specific RNA as a titrant may resolve this issue. The conclusion from the RNaseA and RNase-free experiments that the anomalous spectrum results from the presence of RNA, however, is bolstered by literature indicating that recombinant NP from other

negative stranded RNA viruses (Ebola, Vesicular Stomatitis Virus, Rabies) also have the propensity to bind RNA of their expression system [72-74]. As shown from the data in this thesis, the potential for association between NP and RNA seems to result in observable changes in influenza NP structure and, from a regulatory perspective, must be tightly controlled when making vaccines for human use.

Regulatory guidance suggests the full characterization of the quality (physico-chemical) attributes of therapeutic proteins including the protein's response to stress conditions like exposure to extremes of temperature and pH [75] where proteins are likely to unfold. Acid-induced unfolding of a protein is driven by the protonation of ionizable side chains of amino acids comprising the polypeptide. This protonation leads to intramolecular charge repulsion and results in protein unfolding and subsequent loss of secondary and tertiary structure [76]. The CD spectrum of the 'predominant' form of NP in its RNA free state at pH 7 (Figure 3 blue trace) contains two distinct negative peaks at 222nm and 208nm suggesting high α-helical content. Rather than seeing the intensity of these peaks decrease upon adjusting the pH to 3 with HCl, a dramatic increase in negative ellipticity was unexpectedly observed. Thus 'predominant' NP preparations when titrated with acid can be driven into an acid-induced state which exhibits increased negative ellipticity, correlating with increased helical structure compared to the native state (Figure 7 A). Although these results are surprising, there are some studies reported in the literature in which proteins do not unfold as expected when exposed to acidic conditions but rather assume higher ordered acid-induced conformations [76, 77]. The acid-induced conformational change in the 'predominant' form of NP is globally reversible: upon return to pH 7, a loss in ellipticity is observed at 222nm and 208nm, and the final

spectrum at pH 7 is virtually identical to the initial spectrum at pH 7 indicating that NP has reverted back to its original native conformation (Figure 7B). As indicated by the cross-sectional analysis following the ellipticity at 222nm and 208nm (Figure 7C), the process or pathway of conformational change is not reversible: the spectra at intermediate pH values are not the same for the same pH on the forward and reverse titrations. This micro-irreversibility is known as hysteresis.

These same titrations were also carried out for the 'anomalous' preparations. As demonstrated before, the 'anomalous' preparations yielded CD spectra consisting of a distinct minimum at 222nm and the absence of a second minimum at 208nm (Figure 6 red trace). Addition of either HCl (Figure 13) or H_2SO_4 (data not shown) resulted in the retention of the overall anomalous spectral profile (single minimum at 222nm) until pH of 5.5 although the intensity of the signal decreased at all wavelengths as the solution became more acidic. Thus, until this point, NP exhibited typical unfolding one might expect during acid titration, and lost ellipticity as the pH decreased implying a loss in regular secondary structure elements. However, once the pH was adjusted below 5.5, the trend was interrupted and NP displayed instead the profile seen with the 'predominant' form at lower pH values: one marked by the increased negative ellipticity and defined minima at both 222nm and 208nm. This change in 'anomalous' to 'predominant' profile suggests that, during acidification, NP releases its bound RNA and takes on a conformation (with increased helical content) similar to that of the predominant preparation at pH 3.0. In addition, the 'anomalous' spectrum initially observed at pH 7 is not seen upon reverse titration with NaOH; rather the spectrum achieved on back titration appears to be similar to that observed for a 'predominant' preparation. Indeed, the spectrum

at pH 7 after acidification of an 'anomalous' preparation overlays closely with the spectrum of a 'predominant' NP sample prior to acidification, leading one to believe that, upon acidification, the 'anomalous' preparation releases its RNA, and, upon refolding as the pH is raised, fails to re-incorporate RNA yielding a conformation reminiscent of that seen for the 'predominant' preparation.

During the course of this thesis work, the lab group of Pang-Chui Shaw and collaborators published an extensive analysis of the link between NP oligomerization and its ability to form ribonucleoprotein (RNP) particles. Building upon their original work showing the first X-Ray crystal structure of H5N1 NP, these researchers constructed and tested a series of NP variants in which certain key residues within the tail loop domain were mutated. The original paper containing the structure [6] showed that recombinant NP (at 2.5mg/mL), in the absence of RNA, existed as a mixture of trimer and tetramer. They propose that homo-oligomerization is facilitated through the insertion of the tail loop domain of one NP protomer into the body domain of a neighboring NP protomer at the 'insertion groove'. The more recent work indicates that mutations which disrupt the tail loop structure or abrogate the ability to form a salt bridge between R416 in the tail loop of one protomer and E339 in the insertion groove of another protomer also abrogates the ability to oligomerize, and these variants remain in the monomeric form. They are also deficient in RNP activity (transcription-replication activity). Size exclusion chromatography experiments [27] also indicate that, in the presence of RNA, the elution peak representing the trimeric/tetrameric forms of wild type NP is greatly diminished, and largely replaced by an earlier eluting peak presumably representing higher oligomeric forms of NP complexed with RNA. An RNP-deficient variant in which the

tail structure has been destabilized but the ability to form the salt bridge remains intact (monomeric in the absence of RNA) can also be driven to form the same higher oligomeric forms of NP by addition of a 1850-nucleotide RNA.

The ability to form a salt bridge depends upon the positive charge on R416 and a negative charge on E339, a condition which presumably existed in both 'predominant' and 'anomalous' preparations at pH 7.4. Acid titration of either form should result in the protonation and concomitant loss of charge on E339. Analogous to Shaw's E339A mutant, deprotonated E339 can no longer engage in a salt bridge and NP at pH 3.0 should be monomeric thus explaining the observed change in CD spectrum for the 'predominant' form of NP on acid titration. The results presented in this thesis in conjunction with Shaw's published work would indicate that the monomeric form of NP is more helical than the trimeric/tetrameric form of NP in the absence of RNA. Using a similar argument, the 'anomalous' spectrum at pH 7.4 is postulated to represent higher oligomeric forms of NP complexed with RNA. Adjusting the pH to 3.0 results in protonation of E339, disruption of the salt bridge maintaining the interaction of NP protomers in this higher oligiomeric state as well as loss of RNA from the complex and formation of the same monomeric NP seen on acid titration of the predominant form. Because experimental conditions used in this thesis work did not preclude the activity of advantageous RNases, the RNA once released from the NP-RNA complex is presumably digested and no longer available for reincorporation when the pH is returned to neutral. Thus back titration of the 'anomalous' preparation results in a CD spectrum reminiscent of the 'predominant' form at pH 7.4 rather than the original 'anomalous' spectrum.

Definitive conclusions that the 'predominant' form of NP at pH 7.0 represents the same mixture of trimers and tetramers as reported by Shaw and Ye and their collaborators [6, 25, 27] under similar conditions and that this same form is monomeric at pH 3.0 depend upon the ability to accurately size the NP protein produced in the current thesis under these various pH conditions. Attempts were made to gauge the quaternary structure/ molecular size of the "predominant" form of NP using two very distinct methods: size-exclusion HPLC and particle sizing through dynamic light scattering. Unfortunately, neither method gave easily interpretable results. The groups of both Shaw and Ye used SEC-FPLC on a Superdex™ 200 column using a concentration of NP at 2.5mg/mL or higher [6, 27]. Because protein preparations undertaken for this thesis work yielded NP at concentrations almost 100 fold less and attempts to concentrate often resulted in precipitation, size-exclusion chromatography was attempted on an analytical HPLC column using U.V. detection rather than an FPLC column using MALS (multi angle light scattering). As seen in Figure 18, the only protein eluting from the NP preparation appears to have a molecular size similar to that of RNase A, which is much smaller than expected even for the monomeric form of NP. It is well known that many proteins have the propensity to 'stick' to SEC-HPLC columns and thus elute much later than expected [78]. However, the addition of small amounts of SDS which has been shown to effectively eliminate this problem did not have a significant effect on the retention time. Conversely, DLS on concentrated (using Centricon™) samples gave estimates of molecular sizes that were significantly higher than what would be expected for even multimeric forms of NP. Others [79] have shown that protein aggregates that are abnormally retained or not visible in SEC-HPLC can be

observed and sized successfully by SEC-FPLC, presumably because of the different nature of the column matrix. A MALS detector has recently been ordered and future experiments using SEC-FPLC in conjunction with MALS detection are planned. FPLC experiments will require higher concentrations of protein than what were used in the CD experiments described here and the effect of that concentration change will need to be examined.

During the course of the acid titrations described above, changes in tertiary structure of each type of NP preparation ('predominant' and 'anomalous') was monitored using the molecule's tryptophan residues as probes. (An excitation wavelength of 280nm was used to specifically excite tryptophans rather than all aromatic residues within NP.) Using these intrinsic probes, information regarding the micro-environment surrounding the tryptophan aromatic amino acids can be elucidated. Titration with both HCl and H_2SO_4 of the 'predominant' preparations resulted in a red shift (shifting of the emission maximum to higher wavelengths). As shown in Figure 9A and 10A, the emission maxima of 'predominant' preparations shifted from 336nm to 338nm during HCl titration and from 337nm to 340nm when titrated with H_2SO_4. This red shift indicates that the aromatics which were buried inside the hydrophobic pocket of the protein are becoming more exposed to the hydrophilic (aqueous solvent) environment as the pH is dropped. Bringing the pH back to 7.0 with NaOH results in a blue shift as demonstrated in Figure 9B and 10B, with the wavelengths shifting from 340nm to 337nm and 338nm to 335nm after HCl titration and H_2SO_4 titration, respectively. This blue shift indicates that as the pH was adjusted, NP adopts a global fold that results in the burial of the aromatics into the hydrophobic core of NP.

Fluorescence profiles of the ‘anomalous’ preparations displayed different fluorescent profiles throughout the titrations. When the ‘anomalous’ preparations were titrated down with either HCl or H_2SO_4 (Figure 14A) a red shift was observed (336nm to 338nm with HCl and 336nm-339nm with H_2SO_4). The ’anomalous’ preparation also gave a peak around 420nm that appeared to increase in intensity as the pH was decreased. (This peak can not be attributed to the presence of RNA in the sample, as nucleic acids are weak intrinsic probes and most nucleic acids require modification by extrinsic probes in order to be visualized by fluorescence. Further experimentation will be needed to determine the source of the emission at 420nm and to rule out the possibility that it results from contamination of the solutions and/or plastic and glassware used.) Upon returning the solution pH to 7.0 using NaOH after either HCl titration (Appendix 2) or H_2SO_4 titration (Figure 14B) , there was no shift in emission maxima, contrary to what was observed during the back titration with the ‘predominant’ preparation. The additional peak at 420nm did not decrease in intensity, as one may have expected but rather it continued to increase in intensity, until it reached almost an equivalent intensity as the observed emission maximum at 339nm (H_2SO_4) and surpassed the intensity emission maximum at 338nm (HCl). While the source and behavior of the fluorescence emission at 420nm is not clear, the results fluorescence data from the acid titration of the ‘anomalous’ preparations confirms the conclusions of the CD data for this moiety: the conformation of the ‘anomalous’ preparation before and after acid titration is not the same and neutralizing the pH does not restore the original ‘anomalous’ conformation and/or RNA binding. These data, viewed in conjunction with acidification experiments performed using ‘anomalous’ preparations that had

been RNase A-digested, lend further credence to the postulation that acidification of 'anomalous' preparations leads to the irreversible loss of bound RNA and a concomitant conformational change.

Both CD data (an indicator of secondary structure) and normalized tryptophan fluorescence emission intensity (one indicator of tertiary structure) suggested that the addition of salt (0.15M in the case of CD and 0.5M NaCl in fluorescence) may decrease the stability of the local environment at neutral pH (Figure 20 A and B). Despite the 'chatter' (likely the result of a very low protein concentration and increased signal to noise), the CD data (Figure 20 A) clearly indicates that addition of salt shifts the melting curve to the left, indicating protein destabilization. Fluorescence data at neutral pH indicates that tertiary structure is lost concomitantly with secondary structure both in the presence and absence of salt. At lower temperatures (where the protein is expected to be folded) in the absence of salt (0M NaCl, Figure 20 B, navy diamonds) the loss in quantum yield is actually retarded relative to that seen in Tryptophan free in solution indicating that the protein structure is shielding the tryptophan from solvent quenching. At higher temperatures, the loss of quantum yield actually occurs more rapidly than for tryptophan free in solution, indicating that in the unfolded protein, side chains from other residues are also contributing to quenching. The fact that the discontinuity in the rate of loss of quantum yield (break in the curve) occurs at approximately the same temperature as the T_M observed for NP in the 'predominant' form upon thermal denaturation followed by CD (Figure 20A, navy squares) reinforces the conclusion that the discontinuity in the quantum yield curve is due to concerted protein unfolding upon increasing temperature. For tryptophan in NP in the 'predominant'

form, the addition of salt has very little effect on the quantum yield at low temperatures, where the protein is expected to be folded. However, at intermediate temperatures, where the protein is beginning to unfold, and at higher temperatures, where the protein is likely to be completely unfolded, increasing salt appears to increasingly shield the tryptophan fluorescence from quenching (Figure 20B, pink squares, yellow triangles). At 0.5M and 1.0M NaCl, the break in the quantum yield curve for NP in the 'predominant' form occurs at lower temperatures than in the absence of salt, indicating that salt is destabilizing the protein. While technical limitations precluded the use of higher concentrations of salt, the CD data at 0.15M NaCl also indicates that salt destabilizes NP in the 'predominant' form. The fact that the effects of salt on the 'predominant' form of NP were different at pH3.0 compared to those observed at pH7.0 (Figure 20C and Figure 20B, respectively) reinforces the conclusion that the folded structure of NP is very different in the presence of acid.

Fluorescence data indicate that the effects of salt in combination with temperature for the 'anomalous' preparation of NP at neutral pH are quite complex. Not only is the effect of salt different than for the 'predominant' preparation at this same pH, the curves resulting from plotting normalized fluorescence intensity against temperature appear to be triphasic for the 'anomalous' preparation rather than biphasic. These data are consistent with the conclusion (from the RNaseA digestion experiments) that the 'anomalous' form of NP results from binding of RNA to the protein. As the temperature increases, in the absence of salt, one might expect that dissociation of the RNA from the protein (with a concomitant conformational change) occurs before (at lower temperatures) than the complete unfolding of the protein. The presence of salt would be expected to influence the

binding of the negatively charged RNA to the positively charged arginine residues in the binding groove between the head and body domain of NP. Theory suggests [80] that lower salt concentrations might be expected to be stabilizing whereas higher salt concentrations would be expected to be interrupting the charge-charge interactions and thus destabilizing the binding. While the data indicate differences between the fluorescence intensity vs temperature plots for 'anomalous' preparation at pH 7.0 in the presence of various amounts of salt, it is difficult to draw conclusions as one does not know *a priori* what the effect of RNA binding to NP will be on the quantum yield of tryptohpan fluorescence. Further experiments with smaller graduations in concentration covering a wider range of salt concentrations would need to be performed to elucidate this effect. The data here, however, do help support the presence of RNA in the 'anomalous' form at pH 7.0 but not pH 3.0.

6.0 Conclusions

The research presented in this master's thesis is, to the best of my knowledge, the first attempt to physicochemically characterize the influenza H5N1 NP antigen using CD, fluorescence, SEC-HPLC and dynamic light scattering. Overall, the data presented here demonstrate that NP is a dynamic protein that can adopt several different conformations. As observed, even when the expression and purification protocol remain ostensibly unchanged, the NP product can give at least two different CD spectra ('predominant' and 'anomalous'), suggesting it can adopt at least two different conformations. Acidification of NP solutions ('predominant' or 'anomalous') results in yet a third, structured conformation. The data also show that the 'anomalous' spectrum/conformation is most likely the result of RNA binding to NP. Since the 'anomalous' spectrum cannot be reproduced by mathematically summing the CD spectrum of RNA and that of NP in the 'predominant' form, these findings refute the notion [26] that the nucleocapsid protein does not change conformation when binding RNA. Taken in conjunction with the published literature, the data also suggest that the monomeric form of NP (achieved here by acidification rather than mutation to interrupt an essential salt bridge) is conformationally distinct and, indeed, more highly helical than the trimer/tetramer in the absence of RNA (here, the 'predominant' form) of NP. Previous researchers have suggested that, despite structural differences among individual NP molecules in the unit cell, the structure of one NP protomer in the trimeric crystal is representative of NP structure in both its monomeric and trimeric/tetrameric/higher oligomeric forms [6]. While the oligomeric state of the 'predominant' and acidified forms of NP remains to be unequivocally demonstrated, the data in this thesis suggest that this is not the case

but instead self-association of NP may be accompanied by a fairly major structural change. These data, therefore, form the basis of a different and perhaps more complete understanding of how the influenza virus achieves nucleocapsid particle formation and binds RNA for packaging or exposes it to replication during the viral life cycle.

In addition to adding to our understanding of the biology of the influenza virus, the data in this thesis also have ramifications for the regulation and review of potential therapeutic vaccines that use NP protein as the antigen. Vaccines, and indeed all biologics, are subjected to strict regulatory scrutiny. Manufactures must demonstrate that the material in lot released for sale today is for all practical purposes identical to the lot that was demonstrated to be safe and effective in clinical trials performed to achieve licensure. For vaccines, even those that are therapeutic rather than prophylactic, the conformation of the antigen or protein component is extremely important. NP, when used in trial vaccines, has been shown to produce antibodies [40, 41] . Conformational changes in proteins can expose different epitopes and result in changes in the immune response to the protein. NP is an internal protein and not exposed on the surface of the virus and the antibodies elicited by NP appear to be non-neutralizing (i.e., they do not result in capture and death of the whole virus). One might argue, therefore, that a conformational change, even one that results in the production of a different antibody, may not alter the therapeutic effect of eliciting CTL response. Even if this argument proves to be true, efficacy is only one part of the regulatory consideration. Safety is an equally, if not more important, consideration. When an NP-based vaccine is eventually licensed, the manufacturers will have had to demonstrate, through clinical trials, that

the non-neutralizing antibody response raised by the NP antigen does not cause untoward effects (adverse reactions) and that the vaccine is safe. Subsequent lots of vaccine that present a different conformation of NP (because of different isolation procedures, differences in oligomeric state or the amount of RNA bound or differences in the pH or salt concentration of the buffer) run the risk of eliciting a different antibody response – one that has not been demonstrated to be safe through clinical trials.

The characterization experiments presented in this thesis also demonstrate the utility of physicochemical methods in the complete characterization of potential biologics in the form of vaccines from a regulatory perspective. While physicochemical characterization is the norm for other biologics using recombinant proteins, these more “modern” techniques are only beginning to see application in vaccine evaluation. With the findings presented in this thesis, Health Canada can now consider the utility of physicochemical methods in the evaluation and quality assurance of vaccines, the importance of preparation and solution conditions to NP conformation and the stringencies that must be imposed on manufacturers when making any vaccine with a substantial NP component.

7.0 References

1. Influenza - Seasonal Fact Sheet 211. Available at: http://www.who.int/mediacentre/factsheets/fs211/en/index.html. , 2011.

2. Gong J, Xu W, Zhang J. Structure and functions of influenza virus neuraminidase. *Curr Med Che*m 2007;14(1):113-22.

3. Stevens J, Blixt O, Paulson JC, Wilson IA. Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat Rev Microbio*l 2006; Nov;4(11):857-64.

4. D.M. Knipe and P.M. Howley. Fileds Virology 2006;2:3177.

5. Zambon MC. Epidemiology and pathogenesis of influenza. *J Antimicrob Chemothe*r 1999; Nov;44 Suppl B:3-9.

6. Ng AK, Zhang H, Tan K, Li Z, Liu JH, Chan PK, et al. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB* J 2008; Oct;22(10):3638-47.

7. Du L, Zhou Y, Jiang S. Research and development of universal influenza vaccines. *Microbes Infec*t 2010; Apr;12(4):280-6.

8. Steinhauer DA, Skehel JJ. Genetics of influenza viruses. *Annu Rev Gene*t 2002;36:305-32.

9. Hall CB. The spread of influenza and other respiratory viruses: complexities and conjectures. *Clin Infect Di*s 2007; Aug 1;45(3):353-9.

10. Huang Q, Sivaramakrishna RP, Ludwig K, Korte T, Bottcher C, Herrmann A. Early steps of the conformational change of influenza virus hemagglutinin to a fusion active state: stability and energetics of the hemagglutinin. *Biochim Biophys Acta* 2003; Jul 11;1614(1):3-13.

11. Samji T. Influenza A: understanding the viral life cycle. *Yale J Biol Me*d 2009; Dec;82(4):153-9.

12. Pinto LH, Lamb RA. The M2 proton channels of influenza A and B viruses. *J Biol Che*m 2006; Apr 7;281(14):8997-9000.

13. Kash JC, Goodman AG, Korth MJ, Katze MG. Hijacking of the host-cell response and translational control during influenza virus infection. *Virus Re*s 2006; Jul;119(1):111-20.

14. Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, et al. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. *Nature* 2009; Apr 16;458(7240):914-8.

15. Julkunen I, Sareneva T, Pirhonen J, Ronni T, Melen K, Matikainen S. Molecular pathogenesis of influenza A virus infection and virus-induced regulation of cytokine gene expression. *Cytokine Growth Factor Rev* 2001; Jun-Sep;12(2-3):171-80.

16. Gamblin SJ, Skehel JJ. Influenza hemagglutinin and neuraminidase membrane glycoproteins. *J Biol Chem* 2010; Sep 10;285(37):28403-9.

17. Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem* 2000;69:531-69.

18. Chandrasekaran A, Srinivasan A, Raman R, Viswanathan K, Raguram S, Tumpey TM, et al. Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat Biotechnol* 2008; Jan;26(1):107-13.

19. Hsieh YC, Wu TZ, Liu DP, Shao PL, Chang LY, Lu CY, et al. Influenza pandemics: past, present and future. *J Formos Med Assoc* 2006; Jan;105(1):1-6.

20. Varghese JN, Laver WG, Colman PM. Structure of the influenza virus glycoprotein antigen neuraminidase at 2.9 A resolution. *Nature* 1983; May 5-11;303(5912):35-40.

21. Webster RG, Air GM, Metzger DW, Colman PM, Varghese JN, Baker AT, et al. Antigenic structure and variation in an influenza virus N9 neuraminidase. *J Virol* 1987; Sep;61(9):2910-6.

22. Palese P, Tobita K, Ueda M, Compans RW. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. *Virology* 1974; Oct;61(2):397-410.

23. Portela A, Digard P. The influenza virus nucleoprotein: a multifunctional RNA-binding protein pivotal to virus replication. *Journal of General Virology* 2002; April 01;83(4):723-34.

24. Skorko R, Summers DF, Galarza JM. Influenza A virus in vitro transcription: roles of NS1 and NP proteins in regulating RNA synthesis. *Virology* 1991; Feb;180(2):668-77.

25. Ye Q, Krug RM, Tao YJ. The mechanism by which influenza A virus nucleoprotein forms oligomers and binds RNA. *Nature* 2006; Dec 21;444(7122):1078-82.

26. Ng AK, Wang JH, Shaw PC. Structure and sequence analysis of influenza A virus nucleoprotein. *Sci China C Life Sci* 2009; May;52(5):439-49.

27. Chan WH, Ng AK, Robb NC, Lam MK, Chan PK, Au SW, et al. Functional analysis of the influenza virus H5N1 nucleoprotein tail loop reveals amino acids that are crucial for oligomerization and ribonucleoprotein activities. *J Viro*l 2010; Jul;84(14):7337-45.

28. Baudin F, Bach C, Cusack S, Ruigrok RW. Structure of influenza virus RNP. I. Influenza virus nucleoprotein melts secondary structure in panhandle RNA and exposes the bases to the solvent. *EMBO* J 1994; Jul 1;13(13):3158-65.

29. Albertini AA, Wernimont AK, Muziol T, Ravelli RB, Clapier CR, Schoehn G, et al. Crystal structure of the rabies virus nucleoprotein-RNA complex. *Science* 2006; Jul 21;313(5785):360-3.

30. Green TJ, Zhang X, Wertz GW, Luo M. Structure of the vesicular stomatitis virus nucleoprotein-RNA complex. *Science* 2006; Jul 21;313(5785):357-60.

31. Uversky VN. Intrinsically disordered proteins from A to *Z. Int J Biochem Cell Bio*l 2011; Aug;43(8):1090-103.

32. Male D, Brastoff J, Roth D, Roitt I. *Immunology.* Seventh ed.Mosby Elsevier; 2006.

33. Roy S, Kobinger GP, Lin J, Figueredo J, Calcedo R, Kobasa D, et al. Partial protection against H5N1 influenza in mice with a single dose of a chimpanzee adenovirus vector expressing nucleoprotein. *Vaccine* 2007; Sep 28;25(39-40):6845-51.

34. Wiley JA, Hogan RJ, Woodland DL, Harmsen AG. Antigen-specific CD8(+) T cells persist in the upper respiratory tract following influenza virus infection. *J Immuno*l 2001; Sep 15;167(6):3293-9.

35. Rimmelzwaan GF, Osterhaus AD. Cytotoxic T lymphocyte memory: role in cross-protective immunity against influenza?. *Vaccine* 1995;13(8):703-5.

36. McMichael AJ, Askonas BA. Influenza virus-specific cytotoxic T cells in man; induction and properties of the cytotoxic cell. *Eur J Immuno*l 1978; Oct;8(10):705-11.

37. Gschoesser C, Almanzar G, Hainz U, Ortin J, Schonitzer D, Schild H, et al. CD4+ and CD8+ mediated cellular immune response to recombinant influenza nucleoprotein. *Vaccine* 2002; Nov 1;20(31-32):3731-8.

38. Taylor PM, Askonas BA. Influenza nucleoprotein-specific cytotoxic T-cell clones are protective in vivo. *Immunology* 1986; Jul;58(3):417-20.

39. Carragher DM, Kaminski DA, Moquin A, Hartson L, Randall TD. A novel role for non-neutralizing antibodies against nucleoprotein in facilitating resistance to influenza virus. *J Immuno*l 2008; Sep 15;181(6):4168-76.

40. LaMere MW, Lam HT, Moquin A, Haynes L, Lund FE, Randall TD, et al. Contributions of antinucleoprotein IgG to heterosubtypic immunity against influenza virus. *J Immuno*l 2011; Apr 1;186(7):4331-9.

41. Lamere MW, Moquin A, Lee FE, Misra RS, Blair PJ, Haynes L, et al. Regulation of antinucleoprotein IgG by systemic vaccination and its effect on influenza virus clearance. *J Viro*l 2011; May;85(10):5027-35.

42. Release of a revised version of the Draft Guidance for Sponsors: Information and Submission Requirements for Subsequent Entry Biologics (SEBs). Available at: http://www.hc-sc.gc.ca/dhp-mps/consultation/biolog/2009-03-seb-pbu-notice-avis-eng.php. , 2011.

43. Canada's Regulation of Subsequent Entry Biologics. Available at: http://www.torys.com/Publications/Documents/Publication%20PDFs/FDR2009-2.pdf.

44. Volkin DB, Middaugh CR. Vaccines as physically and chemically well-defined pharmaceutical dosage forms. *Expert Rev Vaccine*s 2010; Jul;9(7):689-91.

45. Morgan AJ, Parker S. Translational mini-review series on vaccines: The Edward Jenner Museum and the history of vaccination. *Clin Exp Immuno*l 2007; Mar;147(3):389-94.

46. Middaugh CR, Joshi SB. Spectroscopy of Vaccines. Development of VaccinesJohn Wiley & Sons, Inc.; 2011. p. 261-92.

47. Malyala P, O'Hagan D, Singh M. Biophysical Characterization of Protein Antigens within Vaccine Formulations. Development of VaccinesJohn Wiley & Sons, Inc.; 2011. p. 293-303.

48. Greenfield NJ. Using circular dichroism spectra to estimate protein secondary structure. *Nat Protoc* 2006;1(6):2876-90.

49. Greenfield NJ. Determination of the folding of proteins as a function of denaturants, osmolytes or ligands using circular dichroism. *Nat Protoc* 2006;1(6):2733-41.

50. Bloemendal M, Johnson WC,Jr. Structural information on proteins from circular dichroism spectroscopy possibilities and limitations. *Pharm Biotechno*l 1995;7:65-100.

51. Sreerama N, Woody RW. Computation and analysis of protein circular dichroism spectra. *Methods Enzymo*l 2004;383:318-51.

52. Kelly SM, Jess TJ, Price NC. How to study proteins by circular dichroism. *Biochim Biophys Acta* 2005; Aug 10;1751(2):119-39.

53. Kelly SM, Price NC. The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sc*i 2000; Dec;1(4):349-84.

54. Greenfield NJ. Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal Biochem* 1996; Mar 1;235(1):1-10.

55. Greenfield N, Fasman GD. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistr*y 1969; Oct;8(10):4108-16.

56. Greenfield NJ. Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat Protoc* 2006;1(6):2527-35.

57. Measuring the conformational stability of a protein. Available at: http://dasher.wustl.edu/bio5325/reading/pace-review-87.pdf.

58. Lakowicz JR. *Principles of fluorescence spectroscopy.* 3rd ed.Springer; 2006.

59. Livingston BD, Higgins D, Van Nest G. Evolving strategies for the prevention of influenza infection: potential for multistrain targeting. *BioDrugs* 2006;20(6):335-40.

60. Yewdell JW, Bennink JR, Smith GL, Moss B. Influenza A virus nucleoprotein is a major target antigen for cross-reactive anti-influenza A virus cytotoxic T lymphocytes. *Proc Natl Acad Sci U S* A 1985; Mar;82(6):1785-9.

61. Wang Y, Wu X, Wang Y, Li B, Zhou H, Yuan G, et al. Low stability of nucleocapsid protein in SARS virus. *Biochemistr*y 2004; Aug 31;43(34):11103-8.

62. Mark J, Li X, Cyr T, Fournier S, Jaentschke B, Hefford MA. SARS coronavirus: unusual lability of the nucleocapsid protein. *Biochem Biophys Res Commun* 2008; Dec 12;377(2):429-33.

63. Ihnat PM, Vellekamp G, Obenauer-Kutner LJ, Duan J, Han MA, Witchey-Lakshmanan LC, et al. Comparative thermal stabilities of recombinant adenoviruses and hexon protein. *Biochim Biophys Acta* 2005; Nov 15;1726(2):138-51.

64. Lidon-Moya MC, Barrera FN, Bueno M, Perez-Jimenez R, Sancho J, Mateu MG, et al. An extensive thermodynamic characterization of the dimerization domain of the HIV-1 capsid protein. *Protein Sci* 2005; Sep;14(9):2387-404.

65. Yelamos B, Nunez E, Gomez-Gutierrez J, Datta M, Pacheco B, Peterson DL, et al. Circular dichroism and fluorescence spectroscopic properties of the major core protein of feline immunodeficiency virus and its tryptophan mutants. Assignment of the individual contribution of the aromatic sidechains. *Eur J Biochem* 1999; Dec;266(3):1081-9.

66. Brittan L. Pasloske. Chapter 8: Ribonuclease Inhibitors. In: C.H. Schein, editor. Methods in Molecular Biology: Nuclease Methods and Protocols Totowa, NJ: Humana Press Inc. p. 105.

67. Whitmore L, Wallace BA. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* 2008; May;89(5):392-400.

68. Perez-Iratxeta C, Andrade-Navarro MA. K2D2: estimation of protein secondary structure from circular dichroism spectra. *BMC Struct Biol* 2008; May 13;8:25.

69. Barría MI, González A, Vera-Otarola J, León U, Vollrath V, Marsac D, et al. Analysis of natural variants of the hepatitis C virus internal ribosome entry site reveals that primary sequence plays a key role in cap-independent translation. *Nucleic Acids Research* 2009; February 01;37(3):957-71.

70. GALLY JA, EDELMAN GM. The effect of temperature on the fluorescence of some aromatic amino acids and proteins. *Biochim Biophys Acta* 1962; Jul 16;60:499-509.

71. Tite JP, Russell SM, Dougan G, O'Callaghan D, Jones I, Brownlee G, et al. Antiviral immunity induced by recombinant nucleoprotein of influenza A virus. I. Characteristics and cross-reactivity of T cell responses. *J Immunol* 1988; Dec 1;141(11):3980-7.

72. Iseni F, Barge A, Baudin F, Blondel D, Ruigrok RW. Characterization of rabies virus nucleocapsids and recombinant nucleocapsid-like structures. *J Gen Virol* 1998; Dec;79 (Pt 12)(Pt 12):2909-19.

73. Noda T, Hagiwara K, Sagara H, Kawaoka Y. Characterization of the Ebola virus nucleoprotein-RNA complex. *J Gen Virol* 2010; Jun;91(Pt 6):1478-83.

74. Green TJ, Macpherson S, Qiu S, Lebowitz J, Wertz GW, Luo M. Study of the assembly of vesicular stomatitis virus N protein: role of the P protein. *J Virol* 2000; Oct;74(20):9515-24.

75. Food and Drug Administration, HHS. International Conference on Harmonisation; guidance on Q5E Comparability of Biotechnological/Biological Products Subject to Changes in Their Manufacturing Process; availability. Notice. *Fed Regist* 2005; Jun 30;70(125):37861-2.

76. Fink AL, Calciano LJ, Goto Y, Kurotsu T, Palleros DR. Classification of acid denaturation of proteins: intermediates and unfolded states. *Biochemistry* 1994; Oct 18;33(41):12504-11.

77. Ahmad E, Fatima S, Khan MM, Khan RH. More stable structure of wheat germ lipase at low pH than its native state. *Biochimie* 2010; Jul;92(7):885-93.

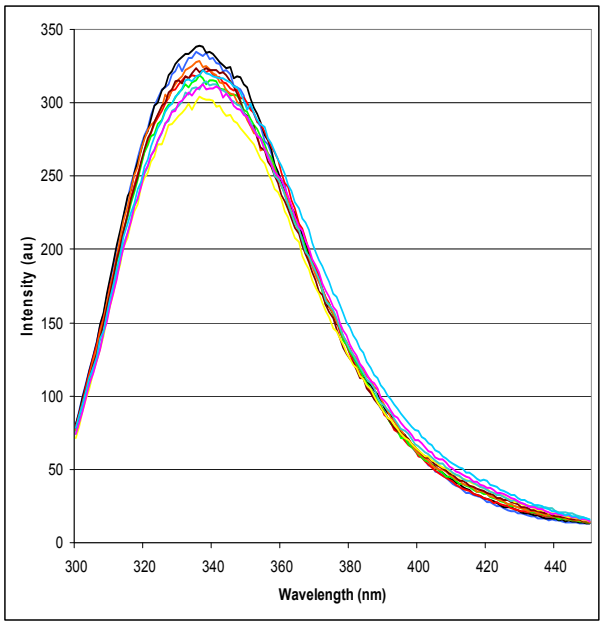
78. Garcia-Canas V, Lorbetskie B, Cyr TD, Hefford MA, Smith S, Girard M. Approach to the profiling and characterization of influenza vaccine constituents by the combined use of size-exclusion chromatography, gel electrophoresis and mass spectrometry. *Biologicals* 2010; Mar;38(2):294-302.

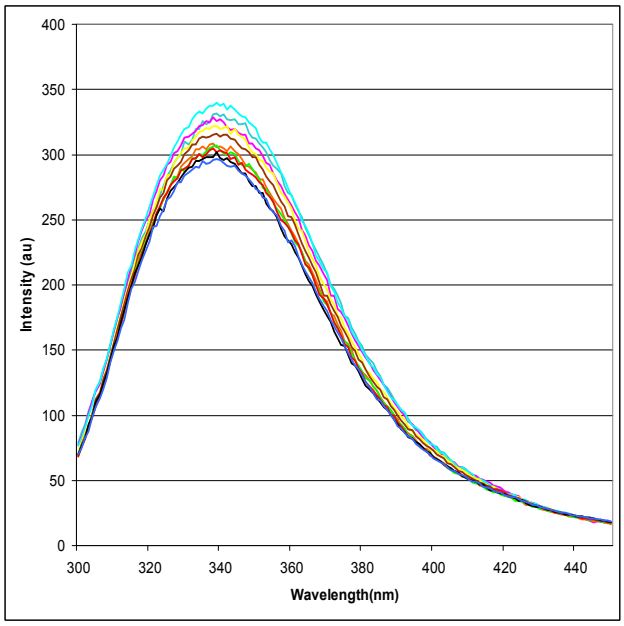
79. Carpenter JF, Randolph TW, Jiskoot W, Crommelin DJ, Middaugh CR, Winter G. Potential inaccurate quantitation and sizing of protein aggregates by size exclusion chromatography: essential need to use orthogonal methods to assure the quality of therapeutic protein products. *J Pharm Sci* 2010; May;99(5):2200-8.

80. Creighton T. *Proteins: Structures and Molecular Properties.* 2nd ed. New York: W.H. Freeman and Company; 1997.

Appendix 1: **Characteristic fluorescence profiles of NP during titration with HCl (A) and back titration with NaOH (B).** (**A**) Titration of a predominant NP preparation from pH 7.0 to pH 3.0 followed by fluorescence spectroscopy. The fluorescence emission observed at neutral pH is at 336nm. Upon adjusting the pH, a red shift is observed, with the wavelength shifting to 338nm. (B) Titration of a 'predominant' NP preparation from pH 3.0 to pH 7.0 followed by fluorescence spectroscopy. The fluorescence emission observed at pH 3.0 is 340 nm. Upon returning to neutral pH, a blue shift occurs with the maximum shifting to 337nm. pH: 7.4 (—), 7.0 (—), 6.5 (—), 6.0 (—), 5.5 (—), 5.0 (—), 4.5 (—), 4.0 (—), 3.5 (—), 3.0 (—).

Appendix 1


Intensity (au)
Wavelength (nm)
0
50
100
150
200
250
300
350
300
320
340
360
380
400
420
440


Intensity (au)
Wavelength(nm)
0
50
100
150
200
250
300
350
400
300
320
340
360
380
400
420
440

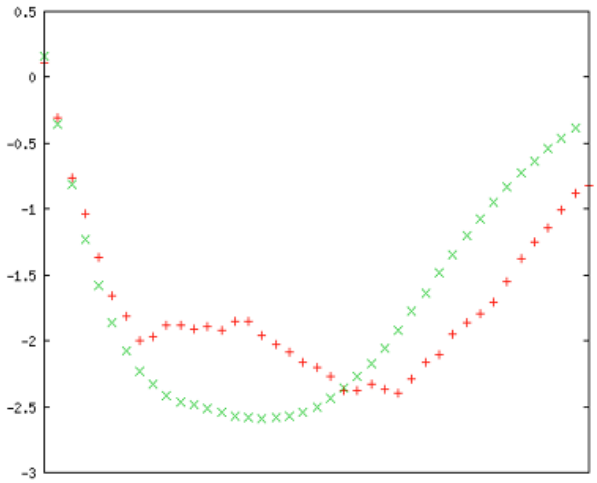
Appendix 2: Representative depiction of the secondary structural determination printouts from (A) K2D2 deconvolution program for 'predominant preparation at pH 7.0 and (B) K2D2 deconvolution with an 'anomalous' preparation at pH7.0. CD data were converted to mean residue ellipticity and then uploaded to the designated program. Inputted data is shown by the red dots, while the computed fits are depicted by the green dots. As shown in A and B, a satisfactory fit was not achieved.

Appendix 2

predicted &alpha helix %: 20.46
predicted &beta strand %: 29.8

max error: 0.4

your input spectrum (txt)
predicted spectrum (txt)

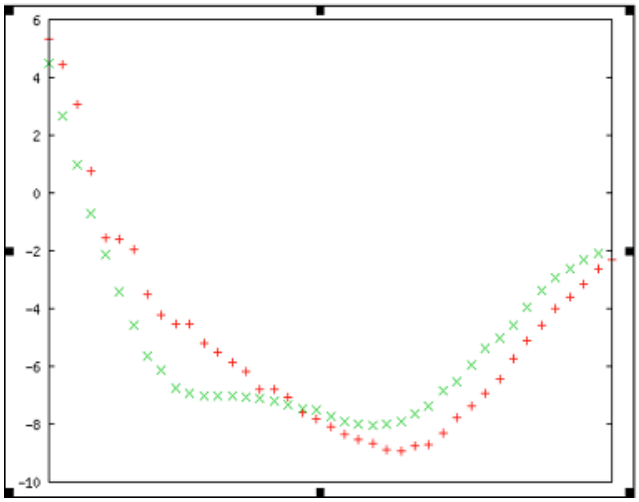


K2d2 predictions Anomalous sample
\\pH7 without RNaseA

predicted &alpha helix %: 60.86
predicted &beta strand %: 4.33

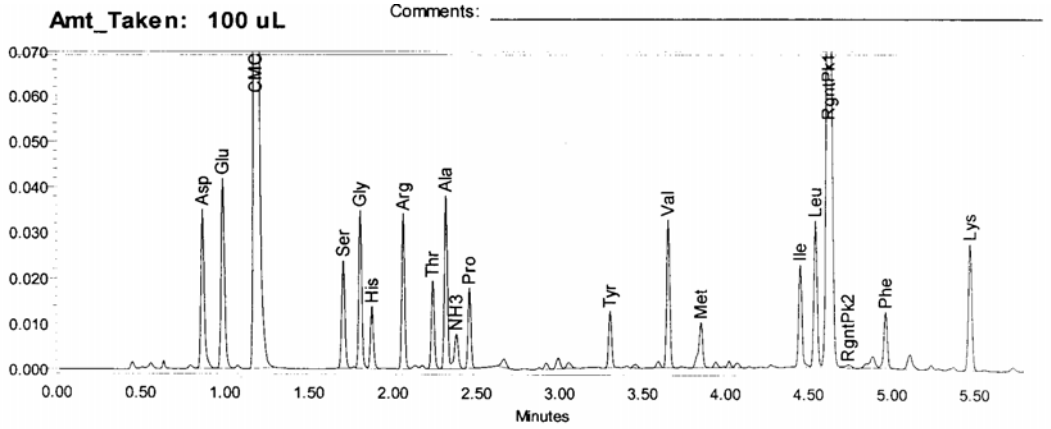
max error: 0.4

your input spectrum (txt)
predicted spectrum (txt)



Appendix 3: Results from amino acid analysis for NP preparation 16 fraction A 10. The data indicates that the concentration obtained by BCA is accurate.

Appendix 3



	Pk Name	RT	Area	Amount	Units
1	Asp	0.870	48817	901.380	pmoles
2	Glu	0.990	59761	1140.132	pmoles
3	CMC	1.187	551631	11752.898	pmoles
4	Ser	1.705	31918	591.440	pmoles
5	Gly	1.806	44654	832.526	pmoles
6	His	1.877	17340	323.350	pmoles
7	Arg	2.064	40449	814.619	pmoles
8	Thr	2.241	24268	439.540	pmoles
9	Ala	2.319	47765	870.732	pmoles
10	NH3	2.382	12871		pmoles
11	Pro	2.460	21690	371.973	pmoles
12	Tyr	3.307	15781	279.099	pmoles
13	Val	3.657	41890	709.830	pmoles
14	Met	3.855	17272	298.921	pmoles
15	Ile	4.451	31094	499.698	pmoles
16	Leu	4.544	44332	754.169	pmoles
17	RgntPk1	4.626	274418		pmoles
18	RgntPk2	4.741	2102		pmoles
19	Phe	4.965	17990	305.769	pmoles
20	Lys	5.476	39654	375.944	pmoles