Development and Application of Spectroscopic Techniques for Characterizing Turbid Solutions of Biopharmaceuticals

By
© 2020
Nicholas R. Larson

Submitted to the graduate degree program in Pharmaceutical Chemistry and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Chair: C. Russell Middaugh Ph.D.

William Picking Ph.D.

Laird Forrest Ph.D.

Jeffery Krise Ph.D.

Robert Dunn Ph.D.

Date Defended: 22 April 2020

The dissertation committee for Nicholas R. Larson certifies that this
is the approved version of the following dissertation:

Development and Application of Spectroscopic Techniques for Characterizing Turbid Solutions of Biopharmaceuticals

Chair: C. Russell Middaugh Ph.D.

Date Approved: 22 April 2020

Abstract

Characterization of biopharmaceuticals is necessary for ensuring their purity and structural integrity. Many classes of biopharmaceutical drugs are turbid solutions. Common optical characterization techniques such as absorbance and FTIR spectroscopy must be adapted for use with turbid solutions. The goal of this thesis is to develop and apply spectroscopic techniques for characterizing turbid solutions. In the first chapter we introduce the current techniques employed for characterization of biopharmaceuticals. Special attention is paid to spectroscopic techniques and possible pitfalls of their use with turbid solutions. In the second chapter we develop a new method for measuring absorbance spectra of turbid solutions. The chapters three through five we apply a variety of spectroscopic techniques to characterize three distinct turbid solutions: aggregated proteins, phase separated antibodies, and mRNA-cationic lipid nanoparticles.

Chapter two describes a label-free, direct approach to measure protein concentrations in turbid solutions using a UV-Vis integrating cavity absorbance spectrometer. Protein-particle conjugates and mixtures have been investigated extensively for their diverse applications in biotechnology. However, general methods to measure protein concentration of protein-particle solutions are lacking. Typically, proteins in turbid solutions require separation or staining with another chromophore to quantitate their concentration. Three systems are used to test the ability to measure accurate protein concentrations: proteins adsorbed to Alhydrogel, proteins in solution with gold nanoparticles and proteins encapsulated within polymeric microspheres. Protein concentrations in each of the three protein-particle systems were successfully quantified using a calibration curve created from the absorbance at 280 nm.

In chapter three we study the two main pathways of Polysorbate 80 (PS80) degradation, enzymatic ester hydrolysis and oxidation. Polysorbates are used ubiquitously in protein therapeutic drugs to help minimize adsorption to surfaces and aggregation. It has been recognized that polysorbate can itself degrade and in turn result in loss of efficacy of therapeutic proteins. Degraded polysorbates were quantified through mass spectrometry to identify the loss of individual components. Next Langmuir trough adsorption isotherms were used to characterize changes in the surface activity of the degraded polysorbates. PS80 degraded via hydrolysis results in slower surface adsorption rates, while the oxidized PS80 show increased surface activity. However, the critical micelle concentration remained unchanged. A monoclonal antibody was formulated with stock and degraded polysorbates to probe their ability to prevent aggregation. Hydrolyzed polysorbate resulted in a large increase in particle formation during shaking stress. Oxidized PS80 was still protective against aggregation for the monoclonal antibody. Monomer loss as measured by SEC was comparable in formulations without PS80 to those with esterase hydrolyzed PS80. Monomer loss for oxidized PS80 was similar to that of non-degraded PS80. Hydrolysis of PS80 resulted in free fatty acids which formed insoluble particles during mechanical agitation which stimulated protein aggregation.

In chapter four we characterize three immunoglobulins which undergo liquid-liquid phase separation with the goal of better understanding how intermolecular protein-protein interactions affect protein conformation in the protein-rich phase. We performed extensive and direct physical characterization of both the protein-rich and protein-poor phases. In comparison, the protein-rich phases displayed blue-shifted tryptophan emission spectra and their amide I infrared absorbance spectra red shifted for all three immunoglobulins. We used Raman spectroscopy to observe disulfide bonds which were isomerized to higher energy strained conformations in the protein-rich phases. Changes in the structural organization of water molecules within the protein-rich phase for all three antibodies were observed, suggesting a role for water hydrogen-bonding in phase separation. Amide I and protein fluorescence spectra differences between the phases persisted at temperatures above the critical temperature, but ceased at the temperature

where these proteins unfolded. Hydrogen exchange-mass spectrometry (HX-MS) analysis with two of the mAbs identified specific regions driving protein reversible self-association. For one of the antibodies, electrostatic interactions between two oppositely charged patches on the Fab region appeared to initiate protein self-interactions. For another antibody, a large hydrophobic patch on the IgG1 heavy chain variable domain facilitated protein self-interaction. We hypothesized that aggregation would occur at equivalent rates in both phases since the two phases have the same chemical potential. For two of the antibodies, the protein-poor phase aggregated more rapidly than the protein-rich phase, while the third antibody displayed equal rates in the two phases.

In chapter five we characterize the effect of solution pH on mRNA–cationic lipid nanoparticles (mRNA–LNPs) structure and transfection efficiency. LNPs containing mRNA are able to deliver genetic material to cells for use as vaccines or protein replacement therapies. We compared the structural and colloidal properties of LNPs not containing mRNA and mRNA free in solution with mRNA–LNPs. We used a combination of biophysical techniques to build a picture of the structure of the lipids and mRNA across pH and temperature in the form of an empirical phase diagram (EPD). A combination of FTIR spectroscopy and differential scanning calorimetry was used to investigate lipid phase behavior. The mRNA–LNPs transition from an inverse hexagonal phase at pH values below the pKa of the cationic lipid to a lamellar phase above the pKa. At higher temperatures the mRNA–LNPs also transition from an inverse hexagonal phase to a lamellar phase indicating the inverse hexagonal phase is more thermodynamically favorable. Circular dichroism was used to investigate changes to the mRNA. The mRNA within the LNP has more A form structure at pH’s below the pKa than above it. Optical density, zeta potential and dynamic light scattering measurements were used to probe the colloidal stability of the mRNA–LNPs. The particles were larger and more prone to

aggregation below the pKa. A stability study was performed to relate the biophysical characteristics to the storage of the particles in solution at 4 and 25 °C. mRNA–LNPs had the highest transfection efficiency and stability at pH values below the pKa. However, there was a trade-off between the stability and aggregation propensity since at very low pH the particles were most prone to aggregation. We performed kinetic experiments to show that the time scale of the pH-dependent phase behavior is slow (6 hr transition) and the transition from lamellar to inverse hexagonal is irreversible. This suggests that the lamellar phase is less-stable and kinetically trapped. Our findings deepen our structural understanding of mRNA–LNPs and will aid in the development of liquid formulations.

Acknowledgments

First I would like to thank my parents and brother for always supporting me throughout this process. I am grateful for their love and support, I could not have completed my thesis without them. My extended family in Kansas City – Colleen, John, Jesse, Josh, and Cody Cutter and Kathy Larson – was also supportive throughout during my time at the University of Kansas and always welcomed me with a home cooked meal.

I would like to thank my advisor Professor Russ Middaugh for his guidance, encouragement, and support during the graduate program. He went beyond expectations to assist in my improvement as a scientist and person. I am grateful for his willingness to teach and share his knowledge and experience. I would also like to thank my other lab members in the Middaugh group for their friendship and guidance – Dr. Yangjie Wei, Dr. Siva Angalakurthi, Dr. Gang Hu, and Dr. Akshay Jain. Much of what I learned at KU was taught to me by Yangjie and Siva in the lab. Yangjie and I worked together on most of my projects at KU and he was always kind and diligent with his work and he remains a good friend.

I would like to thank Professor Christian Schoneich for the collaborations between our groups and for his scientific guidance. Working with his students Indira Prajapati and Dr. Huan Kang was exciting and exposed me to a radical side of pharmaceutical chemistry I would not have otherwise seen.

I would like to thank Professor Cory Berkland for the collaborations with his group. His guidance for work with Jimmy Song and Dr. Martin Leon led to encouraging successes.

I would also like to thank Dr. Reza Esfandiary and Dr. Cavan Kalonia from MedImmune for our collaboration. Many ideas for experiments in chapters three and four came from discussions with them. I would also like to thank Dr. Anthony Tuesca for the ideas and scientific guidance for chapter five of this thesis.

Table of Contents

1 Introduction

1.1 Thesis overview

Characterization of drug products is critical to determine purity and structural integrity. Many classes of biopharmaceuticals are turbid. Some examples are: aluminum-adjuvanted vaccines, proteins at high concentration, and nano- and micro-particle-based drugs. Optical techniques for characterizing biopharmaceuticals have limitations due to the loss of scattered light from turbid solutions. In some cases, these limitations can be worked around by altering the geometry of the optical measurement. For instance, fluorescence spectroscopy of turbid solutions can be performed in a backscattering geometry so that scattered fluorescence light can reach the detector. The purpose of this thesis is to develop new techniques for turbid solution characterization and to apply a of variety of techniques to better characterize specific turbid biopharmaceutical solutions.

In chapter one we outline techniques to characterize biopharmaceutical solutions of proteins. Techniques to characterize each level of protein structure are described. We focus on techniques which can be applied with relatively high throughput for use in formulation development. Special attention is paid to spectroscopic techniques and their limitations for use with turbid solutions.

In chapter two we develop a novel method for absorbance spectroscopy measurements in turbid solutions. This method uses an integrating cavity absorbance spectrometer which is a spherical cuvette coated in a reflective material so that most or all of the scattered light reaches the detector. We show for a variety of turbid protein solutions that concentrations can be accurately quantified using integrating-cavity absorbance spectroscopy.

In chapter three we determine the effects of polysorbate 80 (PS80) degradation on the stability of a monoclonal antibody. Aggregation of proteins in solution results in particles and solution turbidity. PS80 is a nonionic surfactant which is commonly added to biopharmaceutical formulations to prevent aggregation. However, PS80 itself can degrade and in some cases has been shown to accelerate protein aggregation. We compare the effects of the two most common degradation pathways for PS80 – hydrolysis and oxidation – on protein aggregation.

In chapter four we characterize the biophysical properties of phase separated solutions of monoclonal antibodies. Previous work on phase separated solutions had been performed following dilution of the phases to circumvent having to measure the turbid protein-rich phase. Here, we use techniques capable of measuring turbid solutions directly and identify a number of differences between the protein-rich and -poor phases in terms of the secondary and tertiary structure of the constituent proteins. We apply hydrogen-deuterium exchange mass spectrometry to determine specific sequences involved in protein self-association. The differences in aggregation propensity of the antibodies is also examined.

In chapter five we focus on the pH-dependence of the physical properties of mRNA–cationic lipid nanoparticles. mRNA–lipid nanoparticles deliver coding nucleic acids for use as vaccines or protein-replacement therapies. mRNA–lipid nanoparticles are approximately one hundred nanometers in diameter. The intensity of scattered light has a power dependence on particle size (on the order of the power of six for Rayleigh scattering) and therefore solutions of mRNA–lipid nanoparticles are turbid. We use techniques capable of characterizing both the lipid phase behavior and mRNA structure within the lipid nanoparticles to determine the pH-dependence around the pKa of the cationic lipid. An accelerated short-term stability study of the mRNA–lipid nanoparticles is also performed at five pH and two temperatures.

1.2 Biophysical Characterization and the Development of Therapeutic Proteins

A critical element in the development of drugs involves their structural analysis at both the chemical and physical level. With small molecules, this can be accomplished with high precision and resolution with techniques such as crystallography, NMR, and mass spectrometry among other methods. The large size and complexity of macromolecular drugs (e.g., proteins, nucleic acids, vaccines, etc.) usually require a somewhat different approach involving multiple lower resolution methods that can be combined into a multiple faceted picture of the macromolecule. While crystallography, cryo-electron microscopy, and NMR can provide high-resolution structures of proteins, nucleic acids, and some macromolecule complexes, they are usually not applicable to pharmaceutical development due to size, physical state, and other limitations. We will here limit our discussion to methods that are usually available to the non-expert due to their simplicity of operation and availability. Furthermore, we will only consider proteins with an emphasis on monoclonal immunoglobulins with a few examples to better illustrate certain techniques, although much of what we will discuss is more widely applicable. More detailed information can be obtained in the references and in some cases in the figure legends.

When discussing the structure of proteins, it is usual to classify levels of organization into four categories. These are the primary (the amino acid sequence), secondary (reflecting local interactions such as helices, sheets, and turns), tertiary (aspects of the overall three-dimensional structure), and quaternary (subunit and aggregative nature) structures. Different methods can be used to examine and sometimes quantitate various aspects of this classification system. We will employ this convention in our discussion here.

1.2.1 Primary Structure

In the past, the sequencing of proteins was performed using the Edman degradation reaction in an automated instrument. Currently, we almost always use some form of peptide mapping. In this method, a protein is digested with one or more proteases. The digest is separated into individual peptides by chromatography (usually reversed-phase HPLC) and their molecular weights determined by high-resolution mass spectrometry. Subtle changes like those produced by deamidation and oxidation can be analyzed by an initial chromatography step to separate modified forms followed by peptide mapping. The high resolution of modern mass spectrometers (less than 10 ppm) allows single dalton changes as seen in deamidation events to be detected. Such methods have become sufficiently routine that they can be used in stability studies (Figure 1.1)[1].

It is common that lesser resolution methods are often used across all stages of development. The best known of these is SDS polyacrylamide electrophoresis (SDS-PAGE). In this method, proteins are unfolded by sodium dodecyl sulfate (SDS) into rodlike extended structures whose length is proportional to their molecular weight. The proteins are then separated by electric field-induced passage through a polyacrylamide matrix where they migrate approximately proportional to the log of their molecular weight. The proteins are stained by a dye, and their relative mobilities are obtained. This method is not usually sufficient to detect single amino acid changes but can often resolve small changes in molecular weight and larger changes such as oligomerization if these associations are not disrupted by the SDS. A reducing agent is usually included to aid in the unfolding of the proteins and to separate disulfide-linked subunits. Despite its lack of sensitivity, this method is widely used due to its speed and convenience.

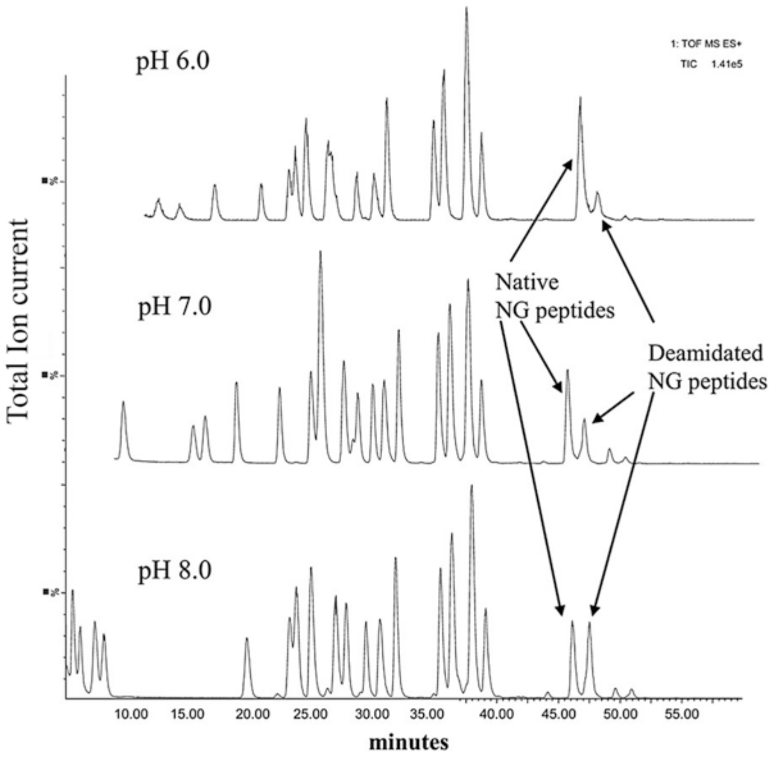


Figure 1.1. Recombinant antigen A is a 42 kDa multi-epitope antigen produced during the development of a broadly effective vaccine for group A Streptococcus. Shown are peptide maps of antigen A produced by Lys C digestion at pH 6.0, 7.0, and 8.0. Proteolysis was performed for 7 days at 4 °C and then resolved and analyzed by LC/MS as described. (See reference[1])

Isoelectric focusing is also used to characterize primary structure because it is sensitive to charge differences. Although once a cumbersome technique when performed in urea gels, it is now typically performed in ampholyte (a mixture of charged peptides) containing capillaries with laser detection. Single-site changes involving charge residues are easily seen. For example, the spectrotypes of immunoglobulins resulting from individual deamidation events and sialyation differences are well resolved[2]. Similarly, reversed-phase or ion-exchange HPLC is often sufficiently sensitive to detect small changes in sequence.

1.2.2 Secondary Structure

Circular dichroism (CD) is a method based on differences in optical absorption of left- and right-handed circularly polarized light[3]. It is particularly useful for characterization of both protein secondary and tertiary (see below) structure. During an absorption event, there is a change in the distribution of charge that can be described by a transition dipole moment. In the case of an asymmetric molecule, there is also an absorption-induced magnetic dipole moment, which is circular in nature. The interaction of these two components is through their dot product, which results in a helical circularization of charge. Thus, molecules which contain circularly averaged absorptive entities such as the peptide bonds in α-helices and to a lesser extent β-structure and various turns possess strong to weak circular dichroism.

In proteins, the interaction between the peptide bonds causes splitting of the CD signal that results in characteristic spectra (measured as the difference in the absorption of the two forms of circularly polarized light as a function of wavelength) for different secondary structure types in the 180–250 nm region. Spectra are normalized for mean residue molar concentration and can be deconvoluted to produce relative contents of the different secondary structure types based on reference spectra. This is automated in the software of modern spectropolarimeters. Changes in temperature, pH, and other variables of interest often produce alterations that involve perturbation of secondary structure as a function of various forms of stress to be quantitatively analyzed. Plots of spectral changes (typically molar ellipticity for normalization purposes to permit comparison of spectra between different proteins) can be used to detect conformational changes in terms of secondary structure alterations (Figure 1.2a, b)[4]. Contributions of other chromophores such as aromatic side chains and disulfide bonds in the 290 – 220 nm region can sometimes complicate interpretation.

There are a wide range of conditions under which CD spectra in the far UV can be obtained but also distinct limitations. A common protein concentration used is 0.1 – 0.2 mg/mL with a 0.1 cm pathlength (PL) cell. Much higher concentrations can be examined with short PLs[5]. For example, spectra of proteins at concentrations of 180 mg/mL can be obtained with 3 µm cuvettes[6]. Conversely, low concentrations can be examined with long PL cells. The only problem here is that when concentrations below 10 μg/mL are tested, a significant portion of the sample may be adsorbed to the inner surface of the cuvette which could alter their structure.

A common problem in obtaining CD spectra is absorption in the UV region by solutes. Most buffers and other solutes absorb in the far UV region. Thus, keeping the concentration of such agents low is highly desirable but may be limited by their stabilizing effects and actual interest in their effects on target proteins. Another problem is the phenomenon known as absorption flattening. When the size of investigated agents becomes larger, they begin to shadow one another, i.e., particles no longer see actual incident light. The effect is to reduce the intensity of the CD signal (it typically becomes less negative), and it shifts to the red. This is commonly seen when proteins aggregate. Thus, one should confirm a lack of aggregation by concentration independence of the observed CD signal. A less frequently encountered artifact is differential scattering of left and right circularly polarized light, but this is rarely seen. We should also mention that modern spectropolarimeters may also permit near UV CD, absorbance, scattering, and even fluorescence to be simultaneously monitored (see later). They may also possess multiple sampling capability and even be found in microtiter plate formats. Therefore, CD can in some cases be used for high-throughput screening.

A second method that is widely used for protein secondary structure analysis is Fourier transform infrared (FTIR) spectroscopy[7]. This technique is based primarily on changes in vibrations of the peptide bond. Originally, the use of dispersive instruments permitted only very high concentrations to be studied which produced low-resolution spectra. Furthermore, interference by water signals was a problem that had to be reduced by the unsatisfactory solution of the use of D_2O as the solvent. This has changed dramatically in the last 30 years with the advent of Fourier transform instruments and sampling methods such as an attenuated total reflectance (ATR) which has allowed high-resolution spectra to be measured at much lower concentration (e.g., 0.1 mg/mL). Due to the coupling of FTIR data to mathematical methods to deconvolute spectral bands, FTIR spectroscopy now rival or even exceed CD in its utility for protein secondary structure analysis.

Most studies focus on the amide I band from 1600 to 1700 cm^{-1}. This broad highly structured but poorly resolved peak arises primarily from CO stretching and to a lesser extent from CN stretching and CCN deformation. There are many additional amide bands (amide A and B and II–VII), but these are less frequently employed. The usual elements of secondary structure (e.g., α-helix, β-structure, turns, loops, and disordered regions) all produce at least one and sometimes more distinct peaks that can usually be resolved by some combination of Fourier self- deconvolution, derivative analysis, and curve fitting[7]. Less common structured forms such as the 3_{10}-helix, poly (L-Pro) helix, and the left-handed α-helix can also be detected. A very strong signal is produced in the 1610 – 1620 cm^{-1} region by aggregated protein due to the cross-beta nature of the interactions between high associated materials. This signal may be of special importance to the pharmaceutical scientist due to the need to detect aggregated protein as a degradant in protein samples (Figure 1.3)[8].

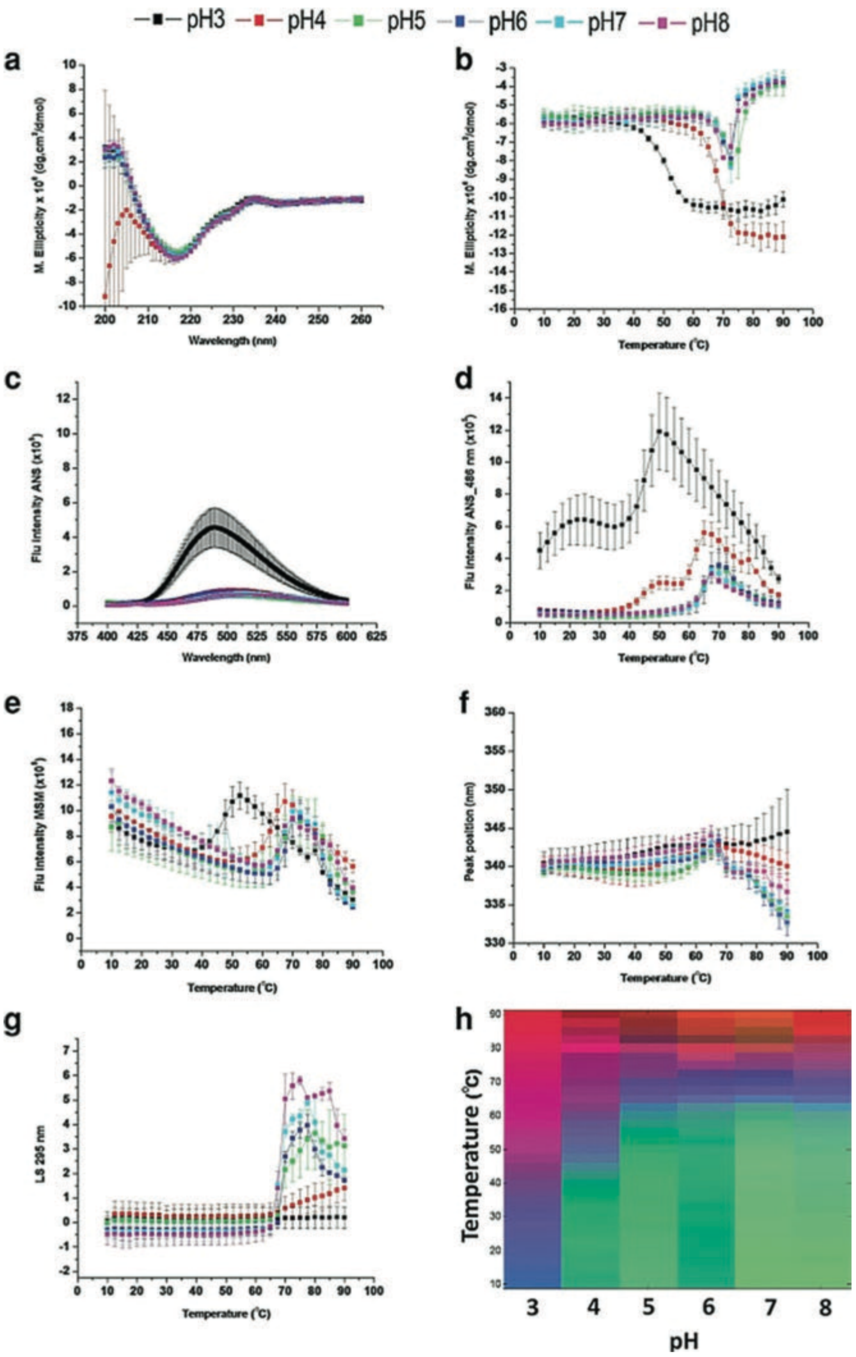


Figure 1.2. Biophysical characterization of an IgG1 mAb as a function of temperature and pH. (a) CD spectra at 10 °C, (b) CD intensity change at 217 nm with temperature, (c) ANS spectra at 10 °C, (d) ANS melting curve at 486 nm, (e) fluorescence intensity versus temperature, (f) fluorescence peak position changes with temperatures, (g) static light-scattering intensity change with temperature, and (h) empirical phase diagram (EPD) analysis of this data. Data shown for n = 3 measurements. (See reference[4])

Although transmittance was initially used as the most common sampling geometry, its application has been reduced by the use of attenuated total reflectance geometries. The very high precision of such methods allow water signals to be quantitatively subtracted and subsequently quantitative secondary structural analysis to be performed. A number of amino acid side chains also absorb in the amide I region (Asn, Glu, Arg, Lys, His, Tyr, etc. with their contribution depending on their state of ionization). These signals are usually weak and therefore ignored, but an unusually high concentration of a particular side chain may require its subtraction before secondary structure analysis is performed. Values of the extinction coefficients of side chains are available for this purpose[9, 10].

Unlike CD, FTIR spectroscopy has been routinely used to examine proteins in the solid state. This is usually done either by grinding the protein solid with IR transparent KBr and compressing a pellet or by the use of diffuse reflectance or ATR. This has been especially useful to search for protein structural changes in lyophilized formulations. As with CD, interference by solutes needs to be considered. Any substances with bonds will produce IR spectra although spectral windows in amide regions can sometimes be found. One can unfold proteins by a salt such as LiBr, but many common chaotropic agents such as urea and guanidine HCl are usually of little use due to their strong IR absorbance. The usual solution variables of interest (temperature, pH, ionic strength) can all be easily studied by FTIR spectroscopy. FTIR instruments are often significantly cheaper than spectropolarimeters further increasing the attractiveness of this technique.

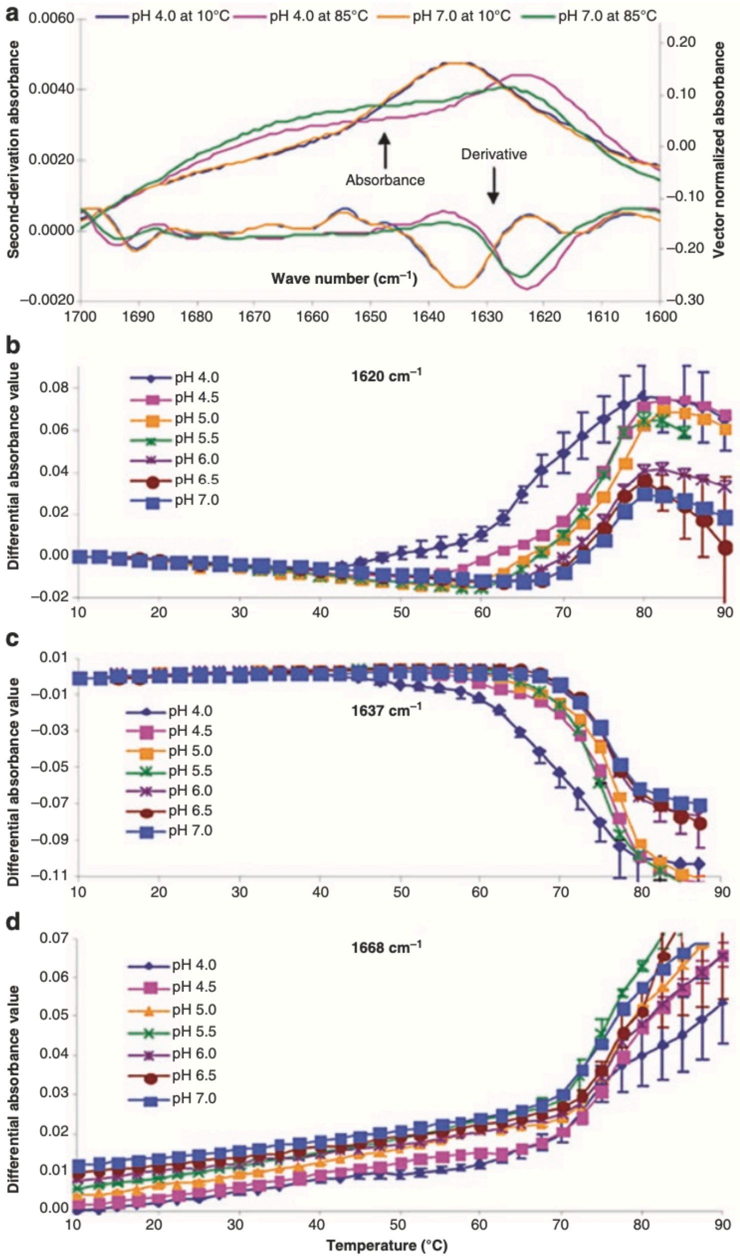


Figure 1.3. Fourier transform infrared analysis of an IgG2 mAb as a function of pH and temperature. Representative FTIR absorbance and second-derivative spectra of the IgG2 mAb at indicated solution pH before and after heat treatment (a). FTIR differential absorbance of the IgG2 mAb as (a function of pH and temperature as monitored at 1637 (b), 1620 (c), and 1668 cm^{-1} (d). FTIR spectra were baseline corrected and vector normalized between 1590 and 1710 cm^{-1}. The peak at 1620 cm^{-1} reflects aggregation of the protein. Differential absorbance data at 1668 cm^{-1} was parallelly shifted by one or multiples of 0.002 units for clearer presentation of the data. Error bars represent 1 SD with n = 2. (See reference[8])

A second vibration-based technique that can be used to examine the secondary structure of proteins is Raman spectroscopy[11]. The difference in spectra between the infrared and Raman methods lies in their unique spectral selection rules. An infrared active transition requires a change in dipole moment, while Raman involves an alteration in polarizability. Vibrational transitions can be IR active, Raman active, or both. The utility of the Raman method was dramatically enhanced by the advent of lasers, which enhanced naturally weak Raman signals. Raman spectroscopy is actually a light-scattering method in which scattered light shifts both up and down in frequency due to interactions with the vibrational transitions of target molecule. The difference in frequency of the incident (laser) light with vibrational transitions constitutes the Raman spectrum. The weak signals produced even with laser excitation require relatively high protein concentration (5–10 mg/mL). Amide bands are present, but the amide III band is often preferred due to its structure and separation of the individual secondary structure element signals. Unlike FTIR spectroscopy, amino acid side chain signals are often easily resolved making the technique of potentially broader utility. Like FTIR spectroscopy, the Raman method can be applied in both the liquid (aqueous) and solid state. Water bands are much weaker eliminating their interference as a major problem. As indicated above, however, Raman signals are intrinsically quite weak. Two modified forms of Raman spectroscopy, however, can potentially overcome this problem. The first is resonance Raman spectroscopy (RRS). In this method, an ultraviolet (UV) laser excites the sample within an absorption band(s). With proteins, this most often involves the peptide band itself in the far UV region. The electronic absorption band transitions of the peptide bond become coupled to vibrational transitions resulting in a dramatic increase in the intensity of the vibrational signals and a much more intense Raman spectrum. The two major problems with this approach are the need for expensive UV lasers and the degradation of the sample due to the intense, focused UV light. The latter can be at least partially overcome by the use of a flow cell. As appropriate instruments become more readily available, we expect to see expanded use of this method.

A second technique is known as surface-enhanced Raman spectroscopy. Spectra can be enhanced by as much as $10^{11} - 10^{12}$. The sample is usually deposited on a metallic surface such as silver or gold, often in the form of a nano- or microparticles. The mechanism is thought to involve the electric field of the surfaces and the excitation of localized plasmons. Visible or near IR light is used to excite the surface- bonded molecules of interest. The extraordinary sensitivity of the method makes it very attractive, but the need for surface localization and the particular surfaces which display the necessary properties have so far limited its use in the biopharmaceutical world.

1.2.3 Tertiary Structure-Sensitive Methods

The precise tertiary structure of a protein can only be determined by X-ray crystallography, NMR, or cryo-electron microscopy. A potential 3D structure may also be estimated by homology modeling and energy minimization. None of these methods are usually directly applicable to protein formulation and stabilization work. Thus, lower resolution methods are again typically employed with an emphasis on changes in tertiary elements rather than the actual 3D structure itself. Here we will describe the most prevalent of these methods, their routine use, as well as their advantages and disadvantages.

The simplest but not the most widely used method to examine tertiary structure changes is UV absorbance spectroscopy[12]. This method has been widely used to measure protein concentration but has been increasingly applied to gather additional information. Below 300 nm, proteins primarily display two peaks. One in the far UV arises from peptide bonds (this is the same peak used in far CD analysis) and a second from the side chains of the three aromatic amino acids with a weak contribution from disulfide bonds. In principle, the far UV peptide bond peak found between 180 and 200 nm could be used to analyze secondary structure, but it is poorly resolved, and interference in this region by oxygen and other agents is difficult to overcome with conventional spectrometers. Thus, its use has been superseded by CD analysis in the same spectral region. In contrast, the near UV aromatic region possesses great utility, which we will describe below. All three aromatic side chains feature structured spectra with the magnitude of their extinction coefficients in the order Trp > Tyr > Phe. Even a simple visual inspection of a protein's near UV spectra often permits the presence of all three side chains to be confirmed if they are present in at least moderate concentration. The high information content of a protein's spectrum can be easily seen by calculating its second derivative, which will typically display 7 – 8 distinct negative peaks. Using a diode array instrument and curve fitting (the method of splining works especially well), the position of these peaks can often be established to a precision of ± 0.02 nm. Since the Phe residues are usually buried, Tyr interfacial due to its hydroxyl group and Trp can potentially be found throughout the structure, the derivative UV spectrum of a protein can provide significant information about changes in the tertiary structure of a protein. Effects of temperature, pH, ionic strength, solutes, and long-term real-time and accelerated stability can all be probed (Figure 1.4)[13]. Furthermore, if light scattering (turbidity) is present, this is usually manifested by optical density (OD = absorbance + turbidity) above 300 nm which is proportional to a high power (3 – 4) of the wavelength (due to Raleigh scattering). Note that absorption spectra can be corrected for the presence of light scattering by a variety of different methods including extrapolation of OD value above 300 nm into absorbing regions (which is nonlinear) and the use of derivatives[14]. Using temperature-dependent second derivative absorbance spectroscopy of aromatic amino acids[15] or by analysis of the effects of cation-π interactions on absorption[16], aspects of protein dynamics can also be probed.

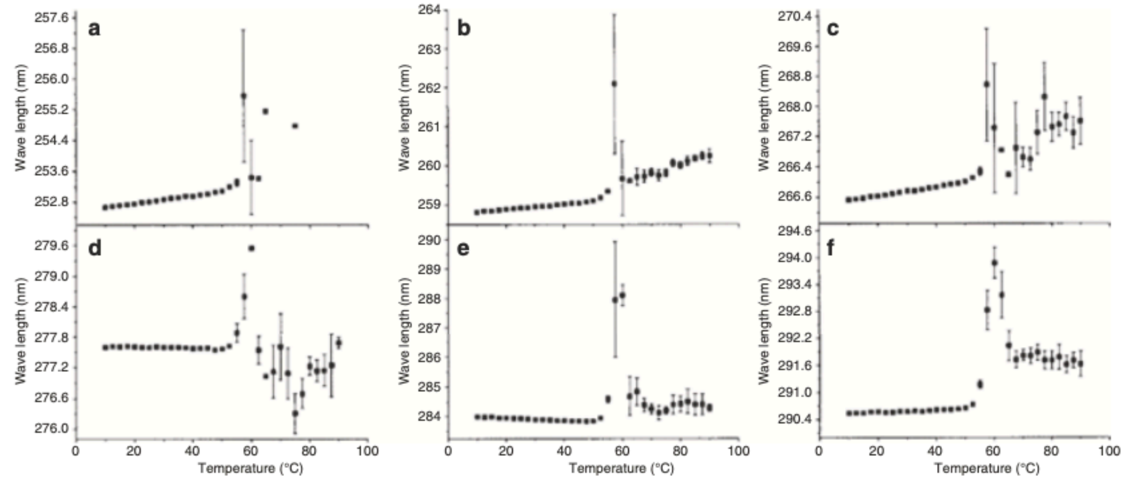


Figure 1.4. Bovine granulocyte colony-stimulating factor (bGCSF) is a member of the four helix bundle family of proteins and has veterinary therapeutic potential, whereas its human analog is already an important drug. Derivative absorbance studies of bGCSF are shown here as a function of temperature at pH 5. The wavelength positions of six negative peaks were followed as a function of temperature: (a), Phe; (b), Phe; (c), Phe; (d), Tyr; (e), Tyr/Trp; and (f), Trp. Protein concentration was 5 µM in 10 mM citrate buffer. Spectra were collected at 2.5 °C intervals, with a 5-min temperature equilibration period included before data collection. All errors are reported as standard error (n = 3). (See reference[13])

Probably the most widely used class of methods employed to examine changes in the tertiary structure of proteins are those that involve fluorescence spectroscopy[17]. The origin of the fluorescence signal used in pharmaceutical protein development activities can be either intrinsic or extrinsic. The intrinsic fluorescence from proteins arises primarily from aromatic residues. If Trp is present, this will almost always dominate the emission spectrum since Tyr and Phe are weak emitters and can lose energy by energy transfer (see later). If no Trp is present, then the fluorescence of Tyr and to a lesser extent Phe can be seen. One exception occurs when the emission of the indole side chain is quenched, perhaps by a proximate positively charged amino or guanidine group. The effect of temperature on aromatic emission results in a smooth decrease due to thermal quenching of the excited state, but this can usually be differentiated from structure changes which produce an alteration in the continuous decrease or a change in inflection of the curvilinear thermally induced decrease in intensity. Another method to resolve structural changes is to measure peak position instead of spectral intensity. Peak position can be accurately determined using derivative analysis or by determining the mean spectral center of mass. The latter is usually more precise but is redshifted 8 – 14 nm relative to actual peak position because of the asymmetry of Trp fluorescence peak (Figure 1.2e, f, and 5)[18].

The quenching of intrinsic fluorescence can be achieved by extrinsic agents[19]. The most common method employs small molecules such as acrylamide or heavy metal salts, which effectively quench some aromatic side chains. A neutral molecule- like acrylamide can actually diffuse through the dynamic protein matrix. Varying the concentration of the quencher and its effect on quenching can be used to probe the dynamic nature of a protein, although care must be taken to ensure the quenching probe does not alter the structure of the protein (especially with acrylamide). In contrast, a charged ion such as iodide or cesium can be used to selectively quench surface residues (usually Tyr).

Polarization measurements can also often be of utility in certain applications. When a protein is illuminated with polarized UV light, it will selectively excite those fluorophores whose absorption transition dipole moments are approximately parallel to that of the incident light. As a target fluorophore rotates, the polarization of the emitted light is decreased. From an analysis of such data, rotation correlation times of entire molecules can be determined. The rotational correlation time is the time required for the molecule to rotate 1/3 of its circumference. Since the rotational time is sensitive to the size and shape of a protein, the technique can be used in studies requiring such information. Furthermore, when two molecules interact, the rotation of the fluorophores is slowed, and the polarization is decreased along with the rotational correlation time. In fact, this method is infrequently used during the formulation and development of therapeutic proteins but is much more commonly used in a wide variety of binding assays. We should mention here that one is not limited to intensity and peak position measurement in fluorescence. The lifetime of the excited state can also be obtained with microtiter formats for such determination now available (Figure 1.5).

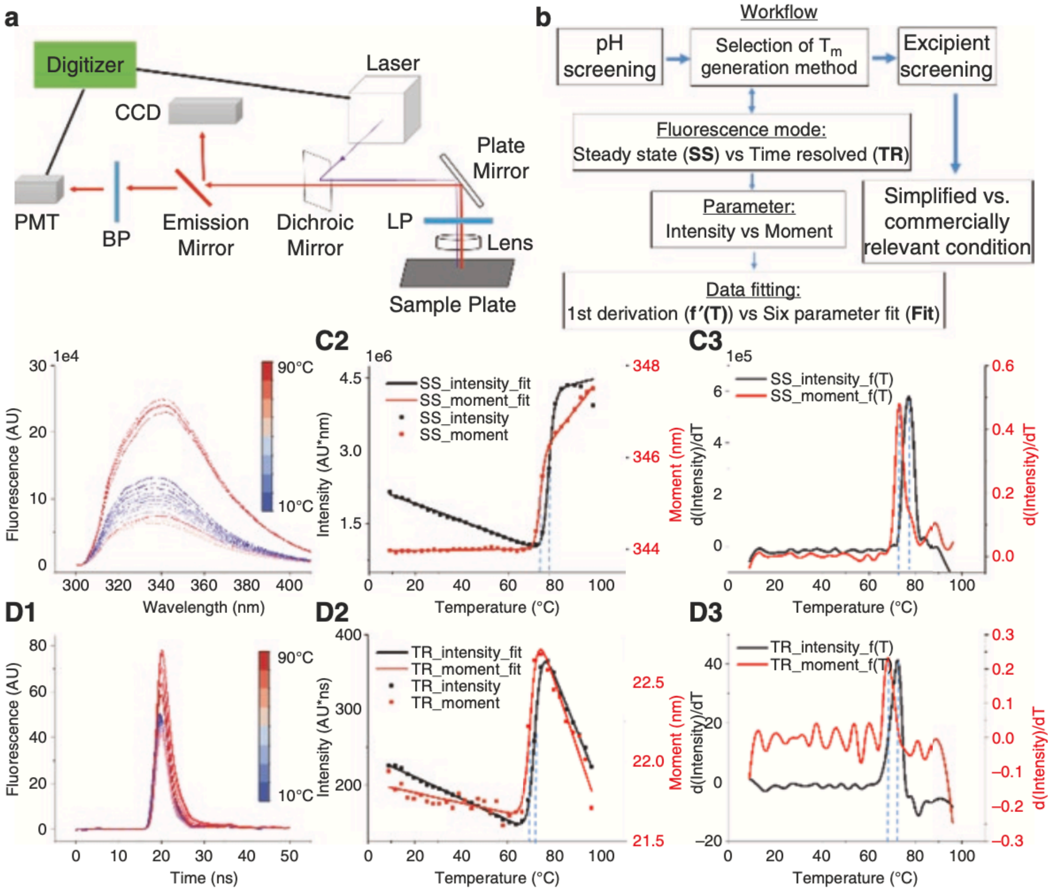


Figure 1.5. A schematic diagram of a high-throughput fluorescence plate reader (a). This instrument is capable of recording both (time-resolved fluorescence) TRF and (steady-state fluorescence) SSF. An emission mirror is placed to direct the emission signal to a CCD detector for the measurement of SSF. Otherwise, the TRF signal is recorded by a PMT. A general workflow for the formulation optimization of proteins using this fluorescence plate reader (b). Thermal melting study of protein samples using intrinsic SSF (C1–3) or TRF (D1–3). Representative temperature-dependent raw fluorescence spectra (C1) of samples; the processed melting curves are monitored by moment (red dot) or intensity (black dot). Fitted melting curves using a six-parameter fitting method are shown as a continuous line (C2). Calculation of the melting temperature (Tm) using the first derivative (C3). Representative temperature-dependent raw waveform (D1) of samples; the processed melting curves are monitored by moment (red dot) or intensity (black dot). Fitted melting curves using the six-parameter fitting method are shown as a continuous line (D2). Further calculation of the Tm using the first derivative (D3). The dashed lines indicate Tm. (See reference[18])

There are, in fact, a remarkable number of experimental approaches that employ fluorescence emission. We will very briefly mention a few more of these although most are not commonly employed during the general pharmaceutical development of proteins. Fluorescence resonance energy transfer (FRET) is a method that takes advantage of the phenomenon in which excitation of a fluorophore (the “donor”) whose emission peak overlaps the absorption peak of a second fluorophore (the “acceptor”) can produce a resonant transfer of energy such that the excitation of the donor results in a decrease of its emission and the sensitized emission of the acceptor. The efficiency of this process is dependent on the distance between the donor and acceptor and the angle between their dipoles and their spectra (in the form of a spectral overlap integral). In ideal cases, the distance between the donor and acceptor can be calculated. This usually involves the use of one or more extrinsic fluorophores covalently attached to known locations, a situation not ideal for the pharmaceutical scientist since it could significantly alter the size and solution behavior of a therapeutic protein. It is thought, however, that the energy transfer between Tyr and Trp residues is partially responsible for the weak emission of Tyr residues in most proteins. The technique is, however, widely used in fluorescence microscopy, a method of great importance, which we will not consider here. There are a wide variety of other fluorescence-based methods not routinely encountered in pharmaceutical analysis. There include fluorescence photobleaching and recovery, fluorescence correlation spectroscopy, red-edge excitation, and single-molecule fluorescence which might have some utility under special circumstances but are more in the realm of the specialist. The use of extrinsic dyes, however, opens up a wide variety of methods that are of significant importance to the pharmaceutical scientist. We will focus on just a few of these applications here based on their relative importance in the analysis of therapeutic proteins.

Extrinsic fluorescence probes have several advantages over intrinsic fluorophores including usually greater fluorescence (quantum yield), large sensitivity to their local environment, and the wide variety of such molecules with diverse properties that are available. One such use is the direct binding of probes (often called dyes for historical reasons) to particular regions of both native and conformationally altered proteins. Most commonly used probes have at least an element of apolar (hydrophobic) character. Thus, if a protein has a site or sites of an apolar nature, it is possible that a dye (depending on its structure) may bind to a protein. This is especially the case if a protein is partially unfolded or altered in such a way that binding is facilitated. For example, molten globular states of proteins which contain near-native secondary structure but a loosening of tertiary structure often bind a variety of dyes. Binding can be manifested by either an increase or decrease in the intensity of the dye's fluorescence and/or a shift in its emission wavelength optimum. Some dyes are strongly quenched by an aqueous environment, and their intensity becomes dramatically enhanced when they are located in an environment where their quenching is at least partially relieved. Therefore such dyes can be used to characterize the stability of a protein as a function of temperature, the most common perturbant. The dyes 8-anili- nonaphthalene sulfonic acid (ANS) and bis-ANS are often used for this purpose (Figure 1.2c, d, and 6)[4]. Other forms of stress (e.g., pH, a chaotropic agent, etc.) can be problematic since they can alter the interaction of a dye with a protein independent of structural perturbation. In recent years, the availability of PCR instruments in many laboratories has led to their use for fluorescence melting studies employing dyes such as SYPRO orange. If detergents are present in formulations, there are a class of dyes known as molecular rotors, which appear to not bind to detergent and can therefore be used as probes[20]. It is also possible to conjugate fluorescent dyes to proteins where they can serve as conformational or environmental probes or as FRET donors and acceptors. In pharmaceutical formulation and development, however, this produces a significantly altered protein which cannot be considered an accurate representation of a drug substance.

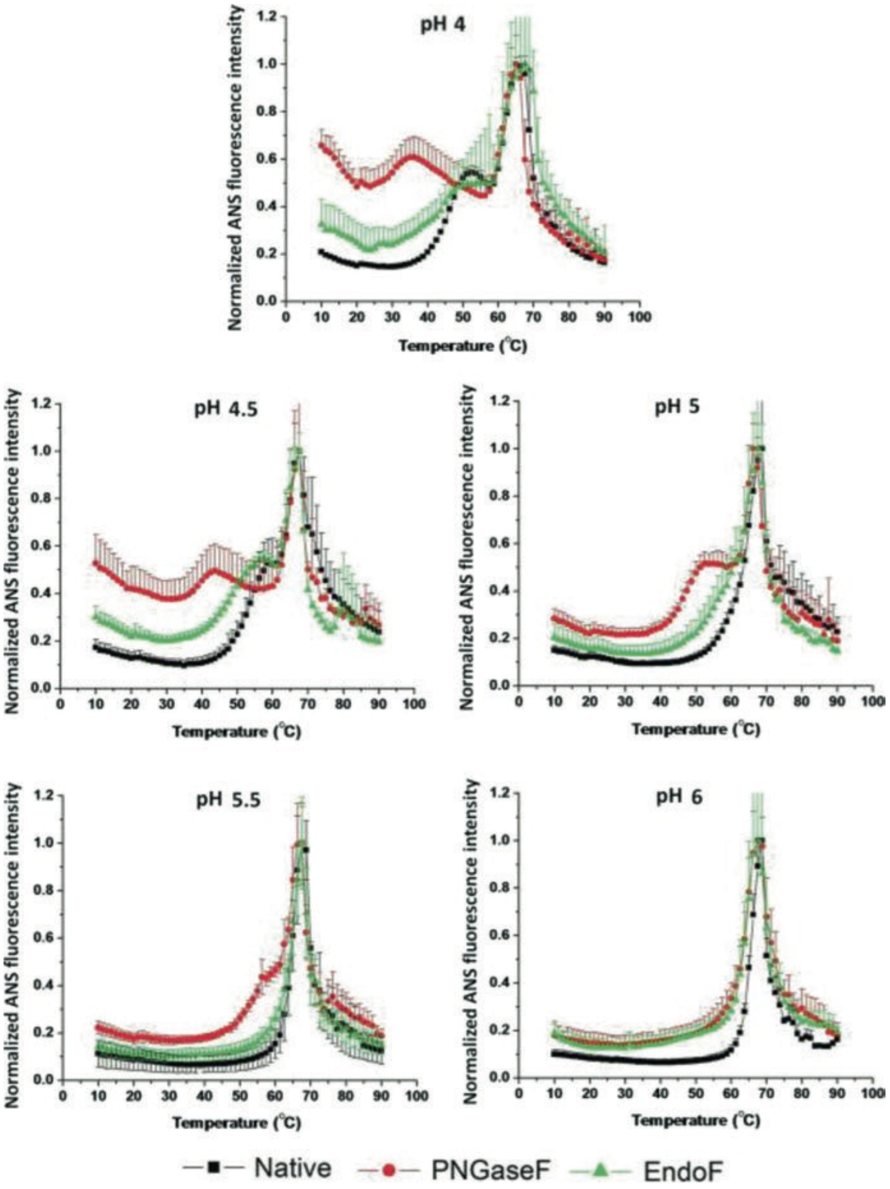


Figure 1.6. Normalized ANS fluorescence intensity change as a function of temperature in the presence of an untreated (control) IgG1 (black line), partially deglycosylated IgG1 (green line), and fully deglycosylated IgG1 (red line) from pH 4 to 6. Normalized results were generated by fitting the data to be equal to one at the maxima for incorporation into EPDs and radar charts. Curves shown here are averages of three runs. (See reference[4])

Circular dichroism can also be used to examine tertiary structure changes. While the far UV region (180 – 250 nm) reflects peptide bond chirality and secondary structure, the near UV (250 – 320 nm) contains primarily signals from the aromatic groups and disulfide bonds. These spectral features are much weaker than those in the far UV and therefore require higher concentrations and longer pathlengths. Typical conditions used might be 0.5 – 1 mg/mL protein and 0.5 – 1.0 cm pathlength. Spectra in this region are not well resolved, but the combined intrinsic and induced optical activity make aromatic rings and disulfide bonds quite sensitive to subtle changes in their immediate environment. Derivative analysis in the near UV CD region can be quite helpful in resolving spectral features, but it is generally difficult to make unambiguous assignments other than through the approximate locations of the spectral features. Both thermal studies and pH and chaotropic effects can all be explored in this region.

1.2.4 Quaternary Structure and Protein Aggregation

Proteins can self-associate into defined oligomers (where the individual proteins are known as subunits) or into larger, more random structures which we refer to as aggregates. These higher molecular weight entities can be characterized in terms of their size, size distribution, molecular weight, and shape among other properties. Such analyses have recently become especially important with the recognition of the presence of both submicron and larger particles in pharmaceutical formulations of therapeutic proteins. Such aggregates can be both immunogenic and display losses of their therapeutic activity requiring their identification and eventual removal from final formulations. As a consequence of their all-too-common presence, a wide variety of methods have been developed to facilitate their identification and characterization. We do not have sufficient space here to describe all such methods currently available, so we will briefly focus on the more commonly used techniques and their utility with regard to the formulation and development of proteins.

The most commonly used method to characterize the size of macromolecules, their oligomers, and aggregates is size exclusion (molecular sieve or gel filtration) chromatography. In this technique, the test solution is pumped through a bed of porous beads made of materials such as agarose, dextran, or polyacrylamide.

The flow of a protein or other macromolecule is impeded by its diffusion into the beads, which is dictated by the size of the channels in the beads and the dimension of the macromolecule. The result is a separation of the proteins based on their size, shape, and molecular weight with their appearance at the exit of the column detected by their optical absorbance, fluorescence (intrinsic or extrinsic), light scattering, refractive index, or some other property of the eluted protein (Figure 1.7b)[21]. It is possible to estimate the molecular weight of the chromatographed material by its elution position relative to MW standards, but a more accurate value of the hydrodynamic radius can be obtained by this method. Higher molecular weight material (e.g., aggregates) usually appears in the void volume of the column since they are unable to enter the bead pores and are not separated further. Although this is a very powerful method and is widely employed, it does suffer from a couple of potential problems. Proteins are diluted as they pass through the column and thus their oligomeric state may change. It is also possible that a protein may interact with the column matrix thereby distorting an interpretation of its behavior in terms of its size. This can sometimes be eliminated by the addition of high salt or a chaotropic agent. There is an alternative separation method that avoids the potential problems of inter- action with the column material. This is known as field flow fractionation. Here a perpendicular field is applied to sample flowing through a narrow tube. It is especially effective for larger particles but can separate proteins, their aggregates, and complexes over a very wide range. Although this method has been available for some time, it is usually not as readily available as SEC.

Probably the second most common method to size proteins is dynamic light scattering (DLS) (also known as quasielastic or photon correlation light scattering)[22]. Samples are illuminated with a laser, and the scattered light is detected at one or more angles, with 90° the most common value. The fluctuations in intensity of the scattered light due to the Brownian motion of the macromolecule or its aggregates are analyzed in the form of an autocorrelation function. This can be interpreted in terms of diffusion coefficients and converted to hydrodynamic radii by the Stokes- Einstein equation. The usual size range that can be examined by this technique is over the range of 1 − 1000 nm. Although a specialist technique, only 20 years ago, the availability of commercial instruments has made DLS a routine laboratory method. Such instruments are now often employed as "black boxes," but care is required in their use and subsequent data analysis. Because all samples display some degree of heterogeneity, the values obtained reflect a range of sizes. Such data can be viewed in a number of different formats including in terms of intensity, mass, and number averages. An averaging method known as cumulant analysis can provide a "mean" diameter and a measure of polydispersity. Intensity deconvolution methods reflect primarily scattering from larger particles and make it appear samples are dominated by the larger entities. This can be useful for detecting the presence of aggregates but presents a distorted view of the actual particle size distribution in the solution. In contrast, the number average calculated distribution reflects the actual distribution and is usually recommended if the polydispersity is low. A problem of which to beware is an apparent decrease in size at higher concentration. An observation of this type is usually caused by multiple scattering. If a photon is scattered more than once, this adds a fast component to the autocorrelation function which can be mistaken for a decrease in size. This can be partially corrected for by the use of back-scattering angles so the radiation does not penetrate deeply into the solution, but this usually is only partially effective. DLS experiments can usually be performed as a function of temperature, and the absolute intensity can also be obtained simultaneously, adding to the utility of this method (Figure 1.8)[23].

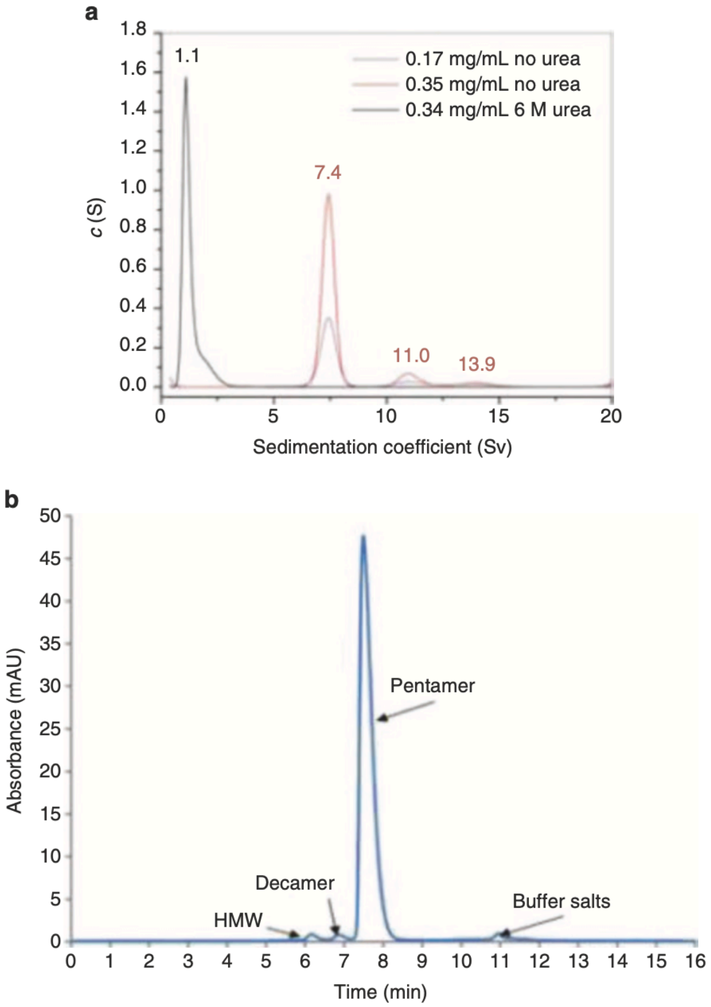


Figure 1.7 The potential human therapeutic protein pentraxin (PTX-2) is a large, glycosylated plasma protein consisting of five monomers that self-associate noncovalently into a pentameric, ringlike structure. Determination of the size distribution of rh-PTX-2 in solution. (a) Distribution of sedimentation coefficients as determined by sedimentation velocity analytical ultracentrifugation for rhPTX-2 at 0.17 mg/mL (magenta) and 0.35 mg/mL (red) in PBS buffer and at 0.34 mg/mL in PBS buffer containing 6 M urea (black). (b) A SEC chromatogram for rhPTX-2 in 10 mM sodium phosphate pH 7.5 with 5% (w/v) sorbitol was generated by dilution into the SEC mobile phase buffer before analysis. Both methods detect the subunit structure of the protein. (See reference[21])

Static light scattering (SLS) can be considered of equal importance to the dynamic form[24]. This method has been employed in a variety of forms ranging from the simple to the complex. The most sophisticated approach involves measurements of scattering intensities at multiple angles and concentrations. The scattering determinations are often in the form of chromatography detection on a SEC column and can provide detailed information about the molecular weight and radius of gyration of a protein (Figure 1.9)[25]. A setup to do this, however, can be somewhat expensive. Much simpler but less informative procedures can simply employ the scattered light at the excitation wavelength in a fluorescence experiment or the optical density (turbidity) from a simple spectrometer. These simple procedures are surprisingly powerful approaches which are usually available in a microtiter plate format making it especially useful for screening purposes. Whatever the instrument used to obtain such scattering data, it is usually used in a relative rather than absolute manner. In a common experiment, scattering intensity or turbidity is recorded as a function of time. These values are then analyzed as initial rates, delay times (reflecting nucleation events), or final values reached at longer time. When screening for inhibition of aggregation, one then looks for a decrease or complete elimination of one or all of these parameters. This approach is widely used since aggregation is often a key, undesired event in the degradation of therapeutic proteins. One should also not underestimate the utility of simple visual examination of a protein solution. This can be manifested as "cloudiness" or translucence.

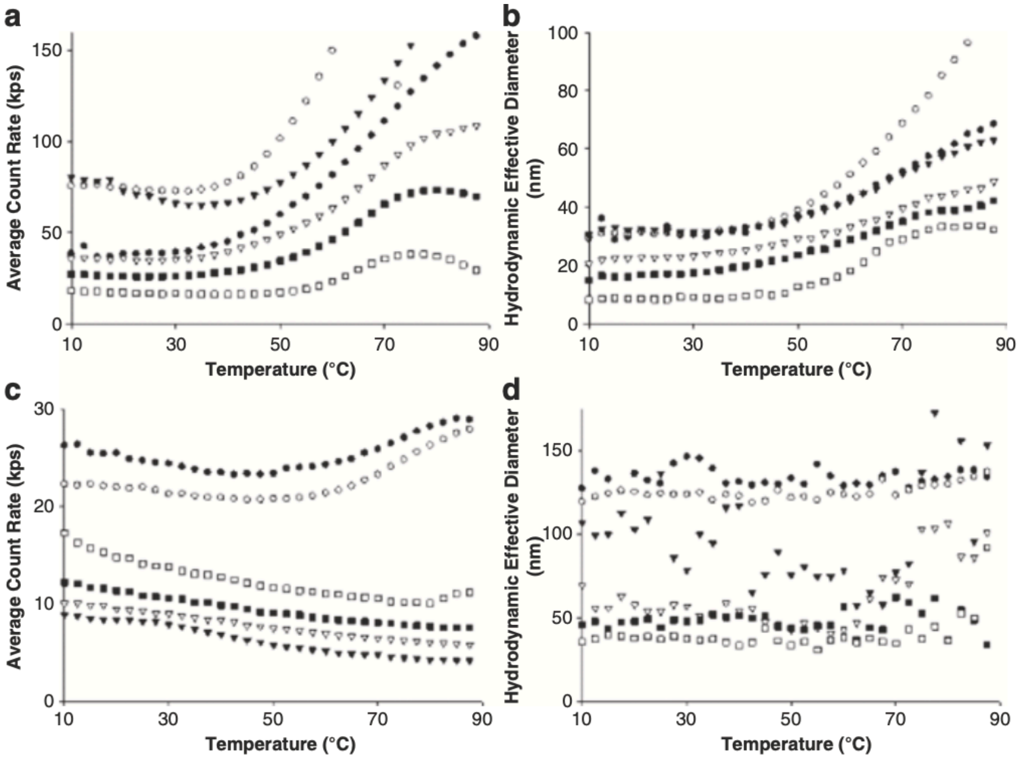


Figure 1.8. The hepatitis C virus (HCV) envelope glycoprotein E1 has been employed as a potential vaccine antigen. A truncated form (amino acids 192 – 326) of the E1 protein (E1y) was expressed in the yeast Hansenula polymorpha and purified from the cell lysate. E1y forms protein particles in the absence of detergent and remains monomeric when detergent concentration is high. Dynamic light scattering (DLS) measurements of HCV E1y (1 mg/mL) in the presence of Empigen BB from 10 to 87.5 °C. The DLS intensity at 532 nm was measured at a 90° angle to the incident beam with a 30-s integration time. The light-scattering intensity is shown in panel (a) for pH 5 and panel (c) for pH 7. The effective hydrodynamic diameter is calculated using the cumulant method and presented in panel (b) for pH 5 and panel (d) for pH 7. The mean values of five measurements are shown for the following Empigen BB concentrations (w/v): ●, no detergent; ○, 0.01%; ▾, 0.1%; ▿, 0.5%; ■, 1%; and □, 2%. (See reference[23])

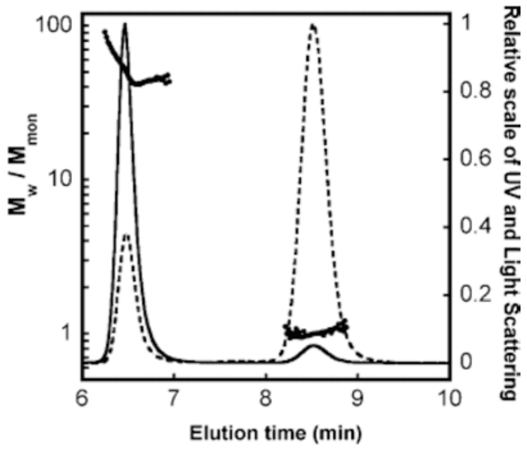


Figure 1.9. Illustrative SEC-MALS chromatogram and weight-average molecular weight (Mw) profile for an aggregated bovine α-chymotrypsinogen A (aCgn) created by incubating an initially monomer sample (initial protein concentration, C0 = 1 mg/mL) at 65 °C for 10 min to achieve approximately 30% (by mass) loss of monomer. A Protein PAK 125 SEC column is used. The peak at ca. 6.5 min is high-molecular-weight aggregate; the peak at ca. 8.5 min is monomer. Solid and dotted lines, respectively, are relative light-scattering intensity (left vertical axis, only 90° scattering angle shown) and relative UV absorbance at 280 nm (right vertical axis). Mw values (scaled by theoretical monomer Mmon = 25.7 kDa) for each 1 s "slice" of the two peaks are given by the symbols (lefthand vertical axis). (See reference[25])

Recently, it has been recognized that actual colloidal phase separation can also occur. While at first this usually appears like the above, with time the solution will actually separate into two distinct phases, one much more concentrated in protein than the other. Another observed form of precipitation appears as particles or strands of material with a wide range of physical properties.

The preceding are probably the most common methods used to characterize the size, oligomeric state, and aggregation of pharmaceutical biomolecules. There has been a recent explosion of interest in alternative techniques with individual advantages. We will briefly consider four of these. The first is known as micro-flow imaging (MFI). This method directly images particles employing a digital camera and measures both counts and size and characterizes various aspects of the morphology of the observed entities. To a limited extent, it can differentiate different types of particles (proteins, air bubbles, aggregates, etc.) and provide an accurate picture of the distribution of sub-visible particle sizes. Depending on the instrument, it can measure particle size in the range of 1 – 2 μm to 300 μm. Such instruments have become increasingly widely used to characterize protein formulations and their aggregates (Figure 1.10)[26].

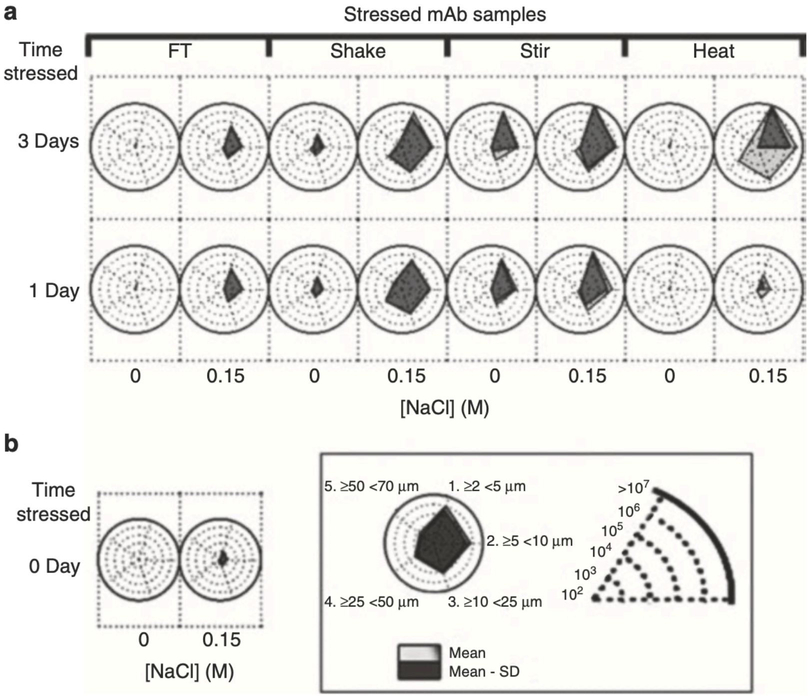


Figure 1.10 Radar plots for visualizing formation of sub-visible particles (concentration and size distributions) in IgG1 mAb solutions exposed to different stresses as measured by MFI. Radar plots show MFI particle concentration and size data distributions as generated by four indicated stresses when applied to a 1 mg/mL antibody solution in 10 mM sodium acetate, pH 5 with and without 150 mM NaCl. See reference[26] for details of radar plot analysis. The data shown are the average of three separate experiments (n = 3), and the error represents one standard deviation. The relative sizes determined by MFI are shown as the magnitude of the axes in the radar diagrams (see box). (See reference[26])

A second instrument employs a method known as nanoparticle tracking analysis (NTA). This method is effective at determining smaller sizes (10 – 2000 nm) and functions by measuring the diffusion of individual particles, in contrast to DLS. Particles are individually observed undergoing Brownian motion. Using a camera (e.g., a CCD), the motion of the particles are tracked on a frame by frame basis with the Stokes-Einstein equation used to calculate a hydrodynamic radius. An accurate particle distribution can be obtained. A third method is known as resonant mass measurement. This method employs a micro electromechanical system and is usable over the particle size range of approximately 50 nm – 5 µm. It measures both particle size and number, but in addition it can be used to determine particle surface areas, density, and dry and buoyant mass among other parameters. There are additional methods such as cell-sorting procedures available in addition to the three described above, but in an ideal situation, a combination of the methods described above can be used to build an accurate picture of a protein, its oligomers and aggregates, and the distribution of the various components and their sizes.

An older approach which is both high resolution and information rich involves the use of the analytical ultracentrifuge[27]. Two different methods are available both with their advantages: sedimentation velocity and equilibrium analysis. The instrument employed is a typical ultracentrifuge, but it is equipped with an optical device that allows one to directly monitor the behavior of a protein or other macromolecule in the presence of a centrifugal force produced by the spinning of a centrifuge's rotor. Special cells are used that permit multiple solutions to be monitored as a function of centrifugation time.

In a sedimentation velocity experiment, the rate at which a macromolecule is sedimented down a sector-shape cell is measured. This rate in the form of a sedimentation coefficient (s) is measured. This velocity normalized to the centrifugal field strength is directly proportional to the molecular weight of the particle after correction for buoyancy and the friction coefficient of the protein, both measurable parameters. Such measurements can resolve individual particles to a high resolution (Figure 1.7a)[21].

In a sedimentation equilibrium study, the sample is spun at a lower speed until an equilibrium concentration gradient of a macromolecule is created within the cell. This gradient can be analyzed to yield a molecular weight for a homogenous sample. In the case where an equilibrium between species is present, the data can be fit to a variety of association models to yield a description of any association which is occurring. Both methods are extremely powerful, but sedimentation velocity is the more commonly used method in pharmaceutical analysis. Analytical ultracentrifugation is generally not amenable to high-throughput work, and the instrument is expensive, so it is primarily used as a research tool rather than directly in formulation development.

There are a number of other methods that are commonly used in the development of protein pharmaceuticals. Of particular importance are those involving calorimetry. Differential scanning calorimetry (DSC) is the most widely applied form and is of great importance in establishing the thermal stability of proteins. In DSC experiments, the difference in energy (heat capacity) required to maintain a sample and reference at the same temperature as the overall temperature is varied and measured (Figure 1.11)[4]. When a molecule undergoes a structural transition, an exothermic peak is seen. If the transition is reversible, the area under the peak corresponds to the enthalpy of the transition. In some cases, multiple peaks may be observed corresponding to the thermal behavior of individual subunits or structural domains within individual proteins. If aggregation is occurring, a peak in the opposition direction of a structural disrupting peak may be seen, but the design of modern calorimetry cells has reduced or eliminated their magnitude. Variation in protein concentration and scan rate can also be used to probe protein-protein interactions. High-sensitivity DSC instruments are relatively expensive but are available with autosamplers making them higher throughput than previous scanning calorimeters.

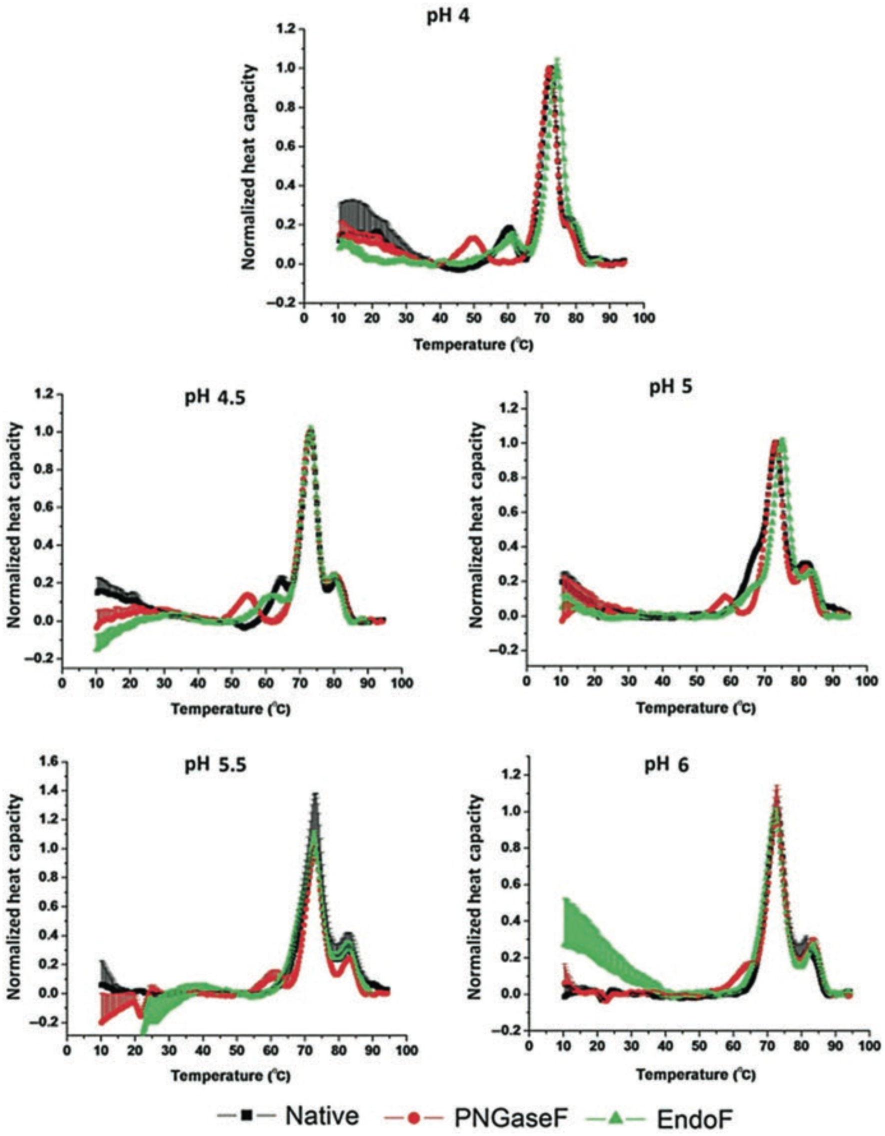


Figure 1.11. The state of glycosylation of therapeutic proteins (in this case an IgG1) is known to have a significant effect on their structure, action, and pharmacokinetics. Shown here is a differential scanning calorimetry analysis of untreated (control) IgG1 (black line), partially deglycosylated IgG1 (green line), and fully deglycosylated IgG1 (red line) from pH 4 to 6. Normalized heat capacity changes were generated by fitting the data to be equal to one at the maxima and to zero at the minima for incorporation into the EPDs and radar charts. Curves shown here are averages of three runs. (See reference[4])

Another type of calorimeter is the isothermal titration calorimeter. In these instruments, one solution is titrated into another. The most common type of experiment involves the titration of a smaller molecule (a "ligand") into a protein (a "receptor") solution or vice versa. The heat absorbed or released in the titration steps (via a syringe) can be used to obtain the enthalpy, entropy, and stoichiometry of binding based on fitting to various binding models. In a similar manner, the binding of one macromolecule to another can be characterized and both the thermodynamics and number of binding sites obtained. This is an especially nice method to examine excipient/protein interactions if they are reasonably strong. There are other calorimetric methods, but one that is occasionally useful to the pharmaceutical scientist is the dilution calorimeter in which the dissociation of oligomeric systems can be studied. It should be especially noted that lower sensitivity DSC and thermal gravimetric analysis are important in the study of protein solids.

A method that has just come into its own in the last few years is hydrogen/deuterium exchange (HDX). A protein is exposed to D_2O for various periods of time, and the partially exchanged forms are analyzed by enzymatic peptide mapping and mass spectrometry. The rate of appearance of labeled peptides serves as a measure of their exposure to solvent and provides a picture of the dynamics of a protein. Localized regions of the protein manifesting different exchange rates provide peptide level resolution which can be further enhanced to a residue level if additional information such as crystallographic temperature factors are available. Thus, the binding sites of excipients and "hot spots" of protein association can be localized by this method (Figure 1.12)[28]. Although the instrumentation to perform such studies is expensive due to the presence of a mass spectrometer, measurements are so information rich that this is becoming an important method for the pharmaceutical scientist. For example, it can be used to localize excipient binding sites as well as those of protein/protein interactions[29].

We will conclude with comments concerning three additional methods that are perhaps less commonly used but can be important in a number of specific situations. With the advent of high-concentration protein formulations (especially those of monoclonal antibodies), the need for viscosity measurement has become increasingly important. There are a variety of instruments available to measure solution viscosity. A traditional method used in protein chemistry employs a

U-shaped hollow tube. These are also known as Ostwald capillary viscometers. The temperature of the test solution is tightly controlled, and the time for the test liquid to pass through a fixed volume is determined. These instruments are not highly accurate and are rarely used today. In a falling sphere viscometer, solid balls (usually steel) are allowed to fall through the medium of interest, and the time to reach terminal velocity is measured. Other types include vibrational, oscillating and falling piston, Stabinger, and rotational viscometers. Recently a number of new methods have come into increased use. The quartz viscometer uses a hollow oscillating quartz crystal. The vibration of the sensor causes shearing of an interior fluid which is monitored by an electric signal. A recent favorite is the rectangular-slit viscometer usually employing microelectromechanical and microfluidic systems. In this method, the pressure drop in the test solution is monitored by an array of sensors. The instruments use very small volumes (microliters), can measure very high viscosities, and are relatively high throughput (Figure 1.13)[30]. There are many other types of viscometers which may be appropriate for certain applications, but the slit types seem to be the current device of choice for high-concentration protein solutions.

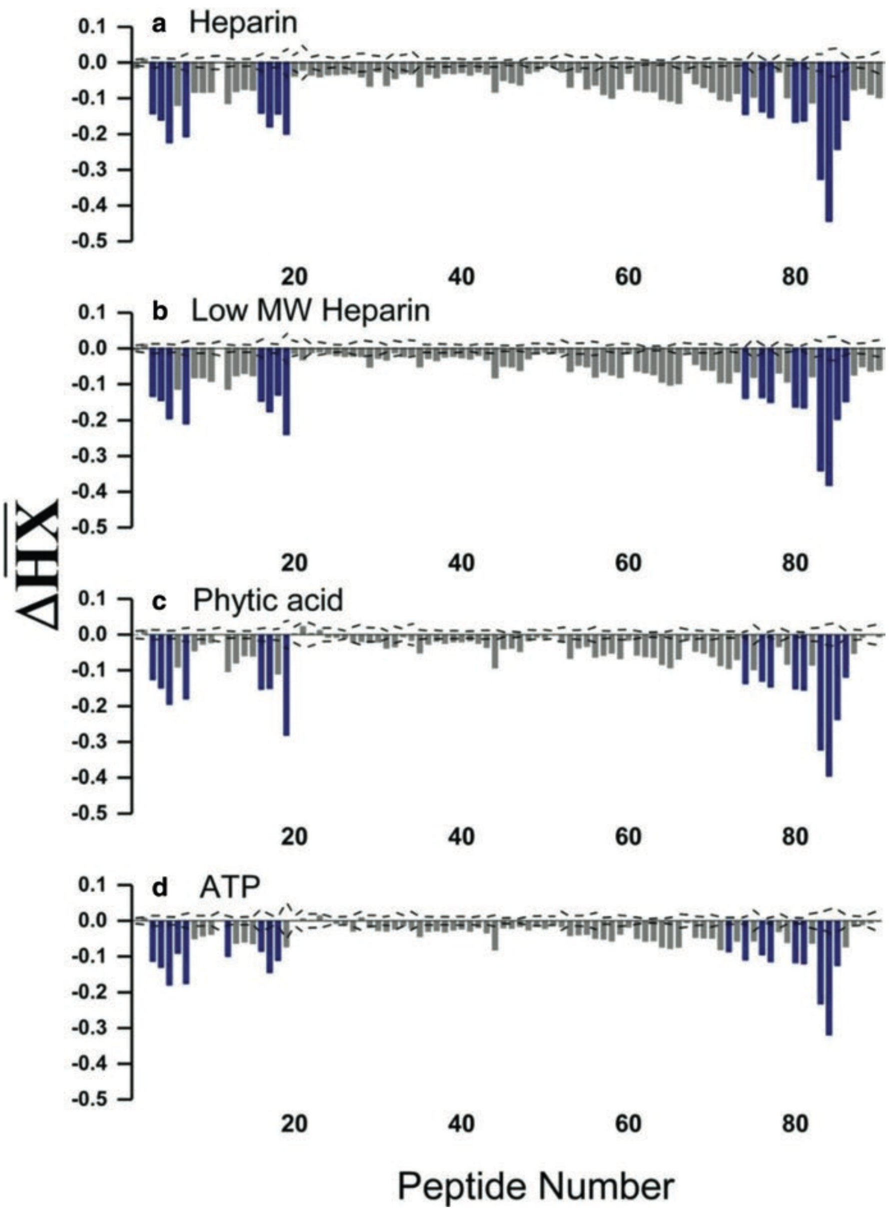


Figure 1.12. Fibroblast growth factor-1 (FGF-1) is a protein of significant clinical utility. It is dramatically stabilized by polyanions. Here isotope exchange was used to establish their binding site, and it is clearly seen that it is the same for each polyanion and its location identified. Relative protection of FGF-1 peptides by heparin, low MW heparin, phytic acid, and ATP: HX-MS was performed with FGF-1 in the presence of (a) heparin, (b) low MW heparin, (c) phytic acid, and (d) ATP. The ΔHX values for each FGF-1 peptide are shown and colored according to their k-means categorization: strong protection, deep blue; intermediate protection/insignificant, gray. (See reference[28])

As mentioned several times above, the measurement of the density of protein solutions is a critical parameter. It is possible to calculate approximate protein densities (and their reciprocal, the partial specific volume) from amino acid compositions, but actual measurements are to be preferred. Traditionally, hydrometers have been used which are based on the buoyancy of a floating glass body and the depth it sinks in a supporting liquid. Pycnometers simply measure the weight of the sample solution in a device of fixed volume. Both methods are simple and inexpensive but possess a number of undesirable properties. From the perspective of pharmaceutical protein therapeutics, hydrostatic balances are relatively accurate but involve an expensive apparatus and a large, complex air-conditioning system. In this method, a sinker is placed in the sample, and the apparent weight loss of the sinker is determined.

In pharmaceutical applications, however, none of these methods are currently used with any frequency. They have been replaced by digital density meters. Like several of the methods already described, a density meter employs the U-tube principle. A hollow tube is filled with the sample liquid. The U-tube is set in a counter mass block and can be set into oscillation. As the mass of the liquid increases, the frequency slows allowing the density of the particles in the liquid as well their partial specific volume to be determined. With care, this technique is extremely accurate and is now usually the preferred method.

Osmometry is also often used by the pharmaceutical chemist. Once upon a time, this method was used to determine molecular weight but has been superseded by many of the methods described above, especially mass spectrometry. The osmotic pressure of a protein solution is, however, an extremely important property of a pharmaceutical formulation for obvious reasons. Thus measurement of this parameter is extremely important. There are three common techniques used to measure the osmotic strength of a solution. These are vapor pressure depression, freezing point depression, and membrane osmometers. All of these work well for determining the total concentration of dissolved salt and sugar among other compounds in pharmaceutical formulations, which are typically desired to be in the physiological range (300 mOsm).

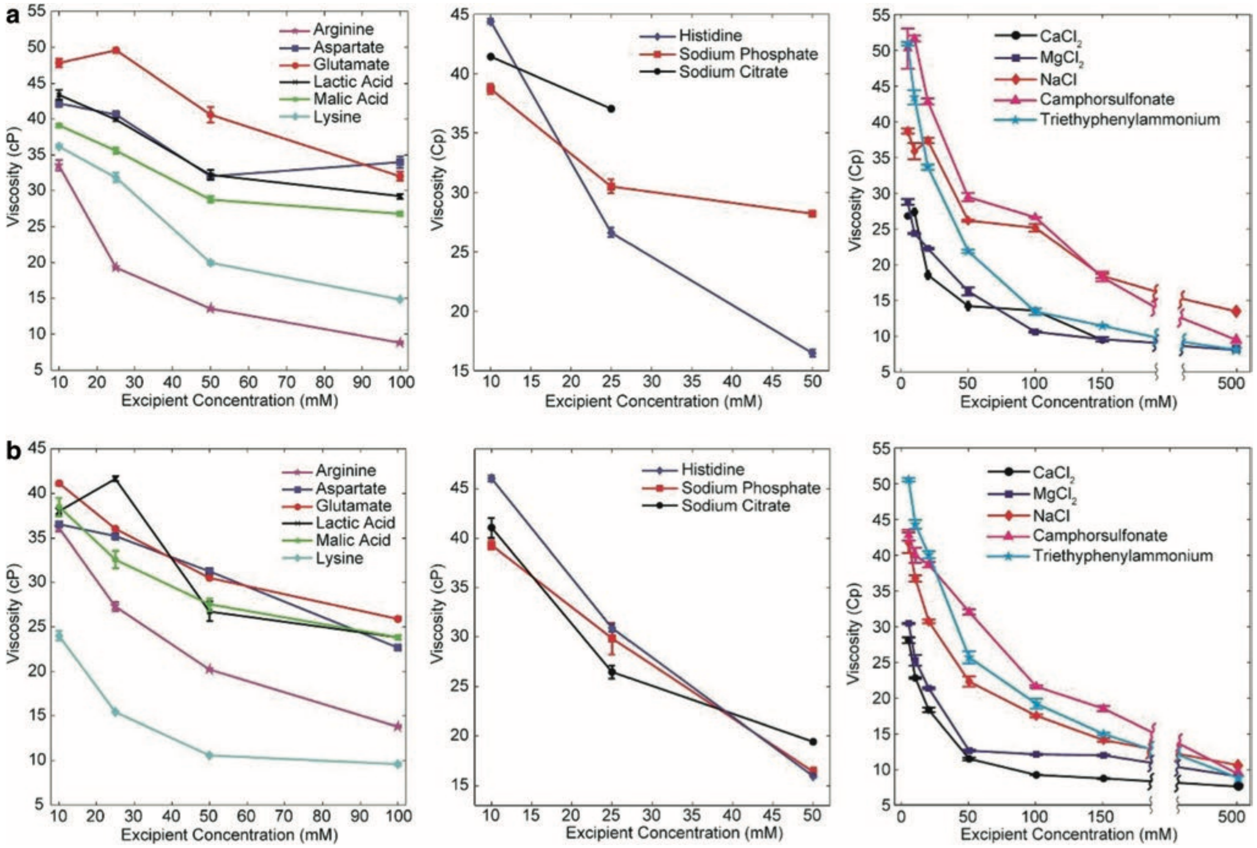


Figure 1.13. With the advent of high-concentration mAb solution, the need to lower their viscosity has become apparent. The effect of the concentration of selected excipients on the solution viscosity of 175 mg/mL solutions of (a) mAb A and (b) mAb C containing 10 mM histidine at pH 5.75 is shown here. MAb A precipitated in formulations containing sodium citrate concentrations of about 25 mM and higher. Values are average and SD from triplicate measurements. (See reference[30])

In this brief discussion, we have described many of the major methods (with an emphasis on the biophysical) that are used by pharmaceutical scientists to characterize and formulate therapeutic proteins. Similar or identical techniques can be used for the development of nucleic acid-based pharmaceuticals and many vaccines. It is possible to combine the data obtained from such methods to paint a detailed picture of target proteins using methods such as empirical phase and radar diagrams. A variety of stresses can be used including temperature, pH, buffer identity, agitation, and freeze thaw. Furthermore, as mentioned previously, multiple methods are often available in single instruments. For example, CD spectropolarimeters can combine near and far UV CD, fluorescence, absorbance, and light scattering. Certain fluorometer can monitor intrinsic and extrinsic fluorescence, lifetime, and scattering data. Thus secondary, tertiary, and quaternary (including aggregation) structure can be simultaneously characterized. This can be also done in a microtiter plate format allowing the high-throughput acquisition of data. All of this together suggests it has become possible to develop and formulate therapeutic proteins with increased speed and efficacy.

1.3 References

1. Stroop, SD. A modified peptide mapping strategy for quantifying site-specific deamidation by electrospray time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* **2007**, 21(6), 830–836.

2. Dada, O. O.; Jaya N.; Valliere-Douglass, J.; Salas-Solano, O. Characterization of acidic and basic variants of IgG1 therapeutic monoclonal antibodies based on non-denaturing IEF fractionation. *Electrophoresis.* **2015**, 36(21–22), 2695–2702.

3. Berova, N.; Nakanishi, K.; Woody, R. W.; *Circular Dichroism. Principles and Applications, Second Edition.* Wiley-VCH Publishers, New York. 2000.

4. Alsenaidy, M.A.; Kim, J. H.; Majumdar, R.; Weis, D. D.; Joshi, S. B.; Tolbert, T. J.; et al. High-throughput biophysical analysis and data visualization of conformational stability of an IgG1 monoclonal antibody after deglycosylation. *J Pharm Sci.* **2013**, 102(11), 3942–3956.

5. Guo, J.; Harn, N.; Robbins, A.; Dougherty, R.; Middaugh C.R. Stability of helix-rich proteins at high concentrations. *Biochemistry.* **2006**, 45(28), 8686–8696.

6. Harn, N.; Allan, C.; Oliver, C.; Middaugh, C. R. Highly concentrated monoclonal antibody solutions: direct analysis of physical structure and thermal stability. *J Pharm Sci.* **2007**, 96(3), 532–546.

7. Middaugh, C. R.; Mach, H.; Ryan, J. A.; Sanyal, G.; Volkin, D. B.; Infrared spectroscopy. *Methods Mol Biol.* **1995**, 40, 137–156.

8. Cheng, W.; Joshi, S. B.; Jain N. K.; He, F.; Kerwin, B. A.; Volkin, D.B.; et al. Linking the solution viscosity of an IgG2 monoclonal antibody to its structure as a function of pH and temperature. *J Pharm Sci.* **2013**, 102(12), 4291–4304.

9. Venyaminov, S.; Kalnin, N. N. Quantitative IR spectrophotometry of peptide compounds in water (H_2O) solutions. I. Spectral parameters of amino acid residue absorption bands. *Biopolymers.* **1990**, 30(13–14), 1243–1257.

10. Venyaminov, S.; Kalnin N. N. Quantitative IR spectrophotometry of peptide compounds in water (H_2O) solutions. II. Amide absorption bands of polypeptides and fibrous proteins in alpha-, beta-, and random coil conformations. *Biopolymers.* **1990**, 30(13–14), 1259–1271.

11. Nemecek, D.; Stepanek J.; Thomas G. J. Jr. *Raman spectroscopy of proteins and nucleoproteins.* Curr Protoc. Protein Sci. 2013, Chapter 17.

12. Mach, H.; Middaugh, C. R. Ultraviolet spectroscopy as a tool in therapeutic protein development. *J Pharm Sci.* **2011**, 100(4), 1214–1227.

13. Kueltzo, L. A.; Ersoy, B.; Ralston, J. P.; Middaugh, C. R. Derivative absorbance spectroscopy and protein phase diagrams as tools for comprehensive protein characterization: a bGCSF case study. *J Pharm Sci.* **2003**, 92(9), 1805–1820.

14. Mach, H.; Middaugh, C. R. Measuring protein spectra in the presence of light scattering. *BioTechniques.* **1993**, 15(2), 240–242.

15. Esfandiary, R.; Hunjan, J. S.; Lushington, G. H.; Joshi, S. B.; Middaugh, C. R. Temperature dependent 2nd derivative absorbance spectroscopy of aromatic amino acids as a probe of protein dynamics. *Protein Sci.* **2009**, 18(12), 2603–2614.

16. Lucas, L. H.; Ersoy, B. A.; Kueltzo, L. A.; Joshi, S.B.; Brandau, D. T.; Thyagarajapuram, N.; et al. Probing protein structure and dynamics by second-derivative ultraviolet absorption analysis of cation{pi} interactions. *Protein Sci.* **2006**, 15(10), 2228–2243.

17. Lakowicz, J. R. *Principles of fluorescence spectroscopy.* Springer United States; 2007.

18. Wei, Y.; Larson, N. R.; Angalakurthi, S. K.; Middaugh, C. R. Improved fluorescence methods for high-throughput protein formulation screening. *SLAS Technol.* **2018**, 23(6), 516–528.

19. Eftink, M. R.; Ghiron, C. A. Fluorescence quenching studies with proteins. *Anal Biochem.* **1981**, 114(2), 199–227.

20. Hawe, A.; Filipe, V.; Jiskoot, W. Fluorescent molecular rotors as dyes to characterize polysorbate containing IgG formulations. *Pharm Res.* **2010**, 27(2), 314–326.

21. Liu, J.; Blasie, C. A.; Shi, S.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B. Characterization and stabilization of recombinant human protein Pentraxin (rhPTX-2). *J Pharm Sci*. **2013**, 102(3), 827–41.

22. Stetefeld, J.; McKenna, S. A.; Patel, T. R. Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys Rev*. **2016**, 8(4), 409–427.

23. He, F.; Joshi, S. B.; Bosman, F.; Verhaeghe, M.; Middaugh, C. R. Structural stability of hepatitis C virus envelope glycoprotein E1: effect of pH and dissociative detergents. *J Pharm Sci.* **2009**, 98(9), 3340–3357.

24. Wyatt, P. J. Light scattering and the absolute characterization of macromolecules. *Anal Chim Acta*. **1993**, 272(1), 1–40.

25. Li, Y.; Weiss, W. F.; Roberts C. J. Characterization of high-molecular-weight nonnative aggregates and aggregation kinetics by size exclusion chromatography with inline multi-angle laser light scattering. *J Pharm Sci.* **2009**, 98(11), 3997–4016.

26. Telikepalli, S. N.; Kumru, O. S.; Kalonia, C.; Esfandiary, R.; Joshi, S. B.; Middaugh,. C. R.; et al. Structural characterization of IgG1 mAb aggregates and particles generated under various stress conditions. *J Pharm Sci.* **2014**, 103(3), 796–809.

27. Liu, J.; Yadav, S.; Andya, J.; Demeule, B.; Shire, S. J. Analytical ultracentrifugation and its role in development and research of Therapeutical proteins. *Methods Enzymol.* **2015**, 562, 441–476.

28. Angalakurthi, S. K.; Tenorio, C. A.; Blaber, M.; Middaugh, C. R. Investigating the dynamics and poly- anion binding sites of fibroblast growth factor-1 using hydrogen-deuterium exchange mass spectrometry. *Protein Sci.* **2018**, 27(6), 1068–1082.

29. Manikwar, P.; Majumdar, R.; Hickey, J. M.; Thakkar, S. V.; Samra, H. S.; Sathish, H. A.; et al. Correlating excipient effects on conformational and storage stability of an IgG1 monoclonal antibody with local dynamics as measured by hydrogen/deuterium-exchange mass spectrometry. *J Pharm Sci.* **2013**, 102(7), 2136–2151.

30. Whitaker, N.; Xiong, J.; Pace, S. E.; Kumar, V.; Middaugh, C. R.; Joshi, S. B.; et al. A formulation development approach to identify and select stable ultra-high-concentration monoclonal antibody formulations with reduced viscosities. *J Pharm Sci.* **2017**, 106(11), 3230–3241.

2 Label-free, Direct Measurement of Protein Concentrations in Turbid Solutions with a UV-Vis Integrating Cavity Absorbance Spectrometer

2.1 Introduction

The ability to measure protein concentration is fundamental to protein science. Ultraviolet absorption of light by aromatic chromophores (e.g. Phe, Trp, Tyr) and dye binding techniques are traditionally used to measure protein concetration[1]. UV concentration measurements use Beer's law ($A = \varepsilon cl$) to relate protein absorbance at 280 nm (A_{280}) to its concentration linearly through its extinction coefficient[2]. Proteins are often paired with relatively large (10 nm – 100 μm) delivery vehicles or adjuvants, in the case of vaccines, to prolong or enhance their immune response[3,4]. These additional components impose difficulty in protein concentration measurement because they can scatter, absorb or emit light which interferes with both UV absorption and dye binding measurements. The scattering of light is particularly problematic for absorption measurements in typical spectrometers since even minor amounts of scattering components may significantly increase solution optical density, which can be incorrectly reported as increased absorbance. The intensity of scattered light is proportional to particle size (d) and wavelength (λ) as $\sim d^6/\lambda^4$ for particles that do not exceed the wavelength of the light (Rayleigh scattering). For larger particles the wavelength dependence of scattered light is proportional to $1/\lambda^n$ where n ranges from 2 to 4[5]. Methods for correction of light scattering artifacts have been proposed for situations where light scattering is minimal[2]. However, for highly scattering solutions, the typical geometry employed in spectrometers is not suitable for the measurement of optically dense solutions. In many cases, to measure protein concentration, proteins must be separated from particles or fluorescently dyed. A method which can measure protein concentration without the need to separate or dye proteins is therefore of great interest.

Here we propose the use of an integrating cavity absorbance spectrometer to capture and quantitate scattered light for protein concentration measurements in turbid media. Integrating cavities are coated with a highly reflective material so light entering the cavity is reflected multiple times within the cavity before exiting towards the detector. Multiple reflections result in an increased effective path length inside the cell and the light measured by the detector is diffuse. Light scattered by materials within the cavity also contributes to diffuse measurement light. Absorbing materials not only absorb light but also decrease the effective mean path length of photons within the cavity, resulting in an attenuated apparent absorbance compared to Beer's law[6]. Absorption cavity spectrometers were initially developed to measure the absorbance coefficients of weakly absorbing samples such as water or methane[6,7]. More recently, integrating cavities have been used to measure the concentration of cellular metabolites such as iron or heme *in situ*[8,9].

We show quantification of protein concentration with an integrating cavity absorbance spectrometer is effective for three common protein + adjuvants/carrier systems: Alhydrogel (AlOH), gold nanoparticles (AuNPs) and poly lactic-co-glycolic acid (PLGA). These systems represent an optically diverse set of conditions: AlOH is purely a photon scatterer, AuNPs scatter, absorb and emit photons, and PLGA is another pure photon scatterer. The wide size ranges for AlOH (diameter 1 – 10 μm), AuNPs (of diameter 20 nm used here) and PLGA microspheres (diameter 20 – 50 μm) employed illustrate the broad applicability of the technique. AlOH and AuNPs interact with proteins similarly by adsorbing proteins to their surfaces primarily through electrostatic interactions[10,11]. Also, both AlOH and AuNPs can form covalent bonds with proteins, AlOH through ligand exchange and AuNPs through thiol-Au bonds. In contrast, PLGA encapsulates proteins within spherical matrices (pictured Figure 2.S1 in Supporting Information), which makes protein concentration measurements impossible without complete protein release by solvation of the PLGA scaffold[12].

Alhydrogel is the most commonly used vaccine adjuvant[3]. There are several methods to measure protein concentration while on the surface of Alhydrogel, but these methods require dye labeling rendering the sample unusable[13,14]. Most methods require release of the protein by pH or use of a chaotropic agent[15]. AuNPs are being used for a variety of applications including biosensing, bioimaging, therapeutic drug delivery and surface enhanced Raman spectroscopy of proteins[16]. Currently the only method for measurement of protein concentration in the presence of AuNPs is by circular dichroism, which is only applicable to relatively small (~10nm) AuNPs[17]. PLGA microspheres represent a case where there are no current methods for measuring the protein concentration without destroying the particle to release the protein. Proteins encapsulated within PLGA cannot be directly measured or even fluorescently labeled. Therefore, general methods which can measure protein concentration in turbid solutions are lacking. Herein we demonstrate that protein concentrations can be measured while adsorbed to AlOH, mixed with AuNPs or encapsulated in PLGA by using integrating cavity UV absorbance spectroscopy.

2.2 Materials and Methods

2.2.1 Materials

Lysozyme, Bovine Serum Albuminum (BSA) and Alcohol Dehydrogenase (ADH) were purchased from Sigma-Aldrich (St. Louis, MO). The IgG1 monoclonal antibody was received from Janssen. Alhydrogel was purchased from Brenntag (Mülheim, Germany). Buffer components (histidine and HEPES) were also purchased from Sigma-Aldrich (St. Louis, MO). Poly (D,L-lactide-co-glycolide) (PLGA, 85:15 lactide:glycolide, acid end group, I.V. = 0.48 dL/g, Tg, = 47.3 °C, M.W. = 62 kDa, P.I. = 1.7, Lot #LP831) was purchased from Evonik Industries (Essen, Germany). Poly (vinyl alcohol) (PVA, 88% hydrolyzed, M.W. = 20 - 30 kDa, Lot #A0320092) was obtained from Sigma Aldrich (St. Louis, MO). Methylene chloride was sourced from Fisher Chemical (Certified ACS, Lot #141990, Pittsburgh, PA).

2.2.2 PLGA Microsphere Synthesis

Protein-free microspheres, and microspheres containing BSA and lysozyme, were manufactured by Orbis Biosciences (Lenexa, KS) using a proprietary process. Briefly, 900 mg of PLGA was dissolved in 10 mL methylene chloride. BSA (100 mg) or lysozyme (100 mg) was then solubilized in DI water (pH = 6.8) at a concentration of 200 mg/mL. The albumin solution was then added dropwise to the PLGA solution, then sonicated over ice for 30 s to create an emulsion with a water:oil ratio of 1:20. The emulsion was processed through a pulseless flow pump (Harvard Apparatus PHD Ultra, Holliston, MA) and vibrating nozzle system with co-axial DI water stream, into a beaker containing a 2000 mL aqueous solution of 0.5% w/v PVA. After fabrication, the microspheres were allowed to harden in the collection beaker with stirring at 75 rpm for 3 h. The microspheres were isolated by suction filtration over a 22 μm filter (Whatman, Plc., Filter No. 541, United Kingdom), washed twice with DI water and then placed at −80 °C for 2 h. Lastly, the microspheres were lyophilized for 48 h at −38 °C and 0.038 mBar (Labconco Freezone 1, Kansas City, MO). The theoretical protein content for both albumin and lysozyme groups was formulated to be 10% w/w. PLGA-only microspheres were produced using the same manufacturing process. Electron microscopy was performed to confirm PLGA microsphere size and loading and shown in Figure 2.S1.

2.2.3 PLGA Microsphere Encapsulation Efficiency

10 mg of PLGA microparticles were dissolved into 750 μL dimethyl sulphoxide (DSMO) for 1 h then further diluted with 2 mL 0.5% SDS/0.02 N NaOH for 1 h. Concentrations were then determined with a micro BCA assay. Protein content in the BSA PLGA microspheres, as determined by μ-BCA, was 0.56 ± 0.2 % w/w. Similarly, the lysozyme PLGA microspheres were found to have a protein content of 1.26 ± 0.3 % w/w. This indicates poor encapsulation efficiencies of protein into the PLGA particles. This phenomenon can be attributed to the microsphere manufacturing process in which the oil-in-water emulsion is initially collected in a large aqueous phase. Given the solubility of albumin and lysozyme is favorable under such conditions, some diffusion of these species from a hardening PLGA matrix is likely and should be anticipated.

2.2.4 AuNP Synthesis

Sodium citrate (2.2 mM, 150 mL) placed in a 250 mL flask was heated on a hot plate with vigorous stirring. After boiling, $HAuCl_4$ (50 mM, 0.5 mL) was added and the resulting mixture was boiled for 10 min until a pinkish color was observed. The flask was then transferred to a water bath maintained at 90 °C and equilibrated for 5 min. $HAuCl_4$ (50 mM, 0.5 mL) was again added. After 30 min, the reaction was stopped. The size and absorption λ_{max} of the resulting AuNP were found to be 21.1 ± 2.9 nm by dynamic light scattering and 522 nm, respectively. The optical density (OD) of the AuNPs in this text refers to their OD at 522 nm as measured by conventional UV-Vis absorbance spectroscopy.

2.2.5 Conventional UV-Vis Absorbance Spectroscopy

Absorbance measurements were made using an Aligent 8453 UV-Vis spectrometer. Spectra for protein concentration measurements were acquired for 2 s. All concentrations were measured in triplicate.

2.2.6 Calibration Curve Preparation

AlOH and AuNPs were prepared at 2x the concentration used in experiments. Proteins were similarly prepared at 2x the concentrations used in experiments. Proteins and AlOH or AuNPs were then physically mixed (1:1 v/v) and allowed to equilibrate 1 h before measurement. The fraction of protein bound to AlOH was determined after centrifugation by measuring the A_{280} of the supernatant. For the highest concentration of AlOH (2 mg/ml), all of the protein (>95%) was determined to be bound to AlOH. For lower concentrations of AlOH (0.2 & 0.02 mg/ml), the percentage bound varied with the protein concentration and the specific protein (IgG or Lysozyme). AuNP concentrations were determined with optical densities (OD) at 522 nm as measured by conventional UV-Vis absorbance measurements. For PLGA, protein concentrations were determined by weighting the PLGA particles and using the calculated encapsulation efficiencies to calculate concentrations.

2.2.7 Integrating Cavity Absorbance Spectroscopy

A CLARiTY VF integrating cavity spectrometer (Olis, Inc., Bogart, GA) was used to collect spectra. This instrument has two deuterium lamps which illuminate the integrating cavity. The integrating cavity has four ports, two for sample illumination by the deuterium lamps, one port focused on the monochromator and the fourth port for sample loading. Spectra were acquired from 250 to 350 nm every 2 nm with an integration time of 0.2 seconds. Proteins were measured at 5 concentrations at half dilutions (1, 1/2, 1/4, 1/8, and 1/16x) ranging from an

A_{280} of 1 to 0.06. All experiments were performed in triplicate at 25 °C. A small spherical magnetic stir bar was placed inside of the cavity for continuous sample mixing. Spectra were acquired by blanking the spectrometer with a corresponding reference (e.g. AlOH, AuNPs) at the same concentration used with the protein samples. PLGA samples were blanked with water and spectra were normalized by subtraction of their absorbance at 340 nm.

2.2.8 Data Processing

All data were fitted to an empirical and theoretical relationship between the apparent absorbance (A'), as measured with the integrating cavity, and A as measured from the conventional UV-Vis spectrometer, as proposed in equation 1[20]:

$$A' = a_0 \ln(1 + a_1 A)$$

Where a_0 and a_1 are empirical constants. All concentration curves (20 curves, 5 points each) were fitted with two parameters for each equation - a_0 and a_1. All equations were fitted using a Levenberg-Marquardt algorithm to minimize χ^2 [20]. Confidence intervals of fitted parameters were calculated with an F-test as described[20]. The error of measurements and fitted parameters were propagated to the errors on the corrected absorbance (shown as 95% CI error bars in Fig. 3). All data was processed with custom scripts written in python using the scipy package libraries and matplotlib for visualization[21,22].

We compared the empirical model (eq. 1) to the theoretical model developed by Fry et al.[20] (rearranged from eq. 6 of that work) as shown by equation 2:

$$A' = -\,log\left(\frac{exp\,[-A * L_{eff}](1-K)}{1 - K * exp\,[-A * L_{eff}]}\right)$$

Where L_{eff} is the effective pathlength of the integrating cavity, and K is a constant which represents the cavity reflectivity multiplied by the fraction of the cavity not containing ports. L_{eff}

and K were fit similarly to a_0 and a_1 from the main text with the additional constraint that K was bound between 0 and 1.

2.2.10 Model Comparison

Statistical differences between the models were compared using the sum square errors (SSE) and the Mean Absolute Percentage Error (MAPE)

Mean Absolute Percentage Error:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{A_i - P_i}{A_i}$$

Where A_i is the true concentration and P_i is the predicted concentration.

2.2.11 Limit of Quantitation (LOQ) and Limit of Detection Determination (LOD)

The LOQ and LOD were determined by measuring 10 blanks, converting the apparent absorbance to conventional absorbance, multiplying absorbance by the slope of the linear calibration curve (to get units of protein concentration), and taking 10x the standard deviation for the LOD and 3.3x the standard deviation for the LOQ.

2.3 Results and Discussion

To test the ability of the integrating cavity spectrometer to measure protein concentrations in the presence of scattering materials, proteins were mixed, adsorbed or encapsulated with scattering components at known concentrations. For clear non-scattering solutions, the spectra have overall similarity to spectra measured with a conventional UV spectrometer although the spectra are compressed because of the nonlinear response of the integrating cavity. As shown in Figure 2.1, the measured absorbance becomes stronger as the concentration of protein increases in the presence of AlOH, AuNPs or PLGA. This indicates that an integrating cavity spectrometer could be used for protein concentration determination assuming the data could be fit to a model.

There have been several models proposed to convert between the absorbance measured with a conventional UV spectrometer and an integrating sphere UV spectrometer[18,19]. The empirical model described above makes no assumptions about the shape of the cavity or its contents. Initially, we compared this model with a theoretical model which makes assumptions both about the shape and contents of the cavity. The error associated with the theoretical model was worse in all cases (Table 2.S1, 2.S2, 2.S3). The mean absolute percentage error was 3.1% for the empirical model and 6.1% for the theoretical model. While the average errors are relatively low for both models, there were some cases (e.g. high AuNP concentrations) in which the theoretical model (eq. 2) performed markedly worse. For these reasons we use of the empirical model (eq. 1) for calculation of limit of quantification (LOQ) and limit of detection (LOD).

To measure the protein content of vaccines containing AlOH, fluorescence methods have been commonly used. For example, the ninhydrin or o-phthalaldehyde assays are used for the fluorescence staining of proteins adsorbed to AlOH[13,14]. Figure 2.2A, B show that the response curve of lysozyme and an IgG have different responses to the amount of Alhydrogel in solution. For lysozyme, all concentrations of AlOH produced virtually identical apparent absorbance vs concentration curves. For the IgG, as the AlOH content increased, so did the apparent absorbance at equivalent protein concentrations. We attribute this increase in apparent absorbance to be a protein specific effect because it only occurred with IgG. It is possible that upon adsorption to AlOH, the local environments of the aromatic amino acids change which causes an increase in its extinction coefficient[23]. The corrected absorbance concentration curves for Alhydrogel have near identical linearity and error as protein in transparent solutions as shown in Figure 2.3A, B.

Protein concentration measurements in the presence of AuNPs are more difficult since AuNPs simultaneously scatter, absorb, and emit photons, as well as quench the fluorescence of proteins. One recent method suggests circular dichroism can be used to measure the concentration of protein – AuNP mixtures[24]. The AuNPs used in that work are ~10 nm, which are below 1/20th of the wavelength of light used which is commonly thought of as the limit of isotropic scattering[25]. AuNPs above this size (hundreds of nanometers) are often used and there is no direct method for measuring protein concentrations in their presence. Figures 2.2C, D show the response of ADH and BSA concentration curves to different amounts of AuNPs. For both proteins the apparent absorbance response is diminished with increasing gold concentration. The change in response can be accounted for as either a change in effective path length, surface reflectivity of the cavity, or absorption coefficient. We believe this attenuation can be attributed to a decreased effective path length as the high concentration of AuNPs has a higher absorbance. It has been demonstrated that the mean path length of light travelling through media is independent of light scattering[26]. It has also been shown that for BSA/AuNP mixtures, the absorbance and scattering cross section of AuNPs increase with increasing BSA concentration[17]. These considerations make the exact mechanism difficult to assign to solely one contributor. Alternatively, a change in reflectivity of the integrating cavity surface because of nonspecific adsorption of AuNPs could also be the reason for the decreased sensitivity. The loss of sensitivity at high protein concentrations manifests as increased error bars on Figures 2.3C, D with increased AuNP concentrations.

Measurement of PLGA microspheres was made increasingly difficult by the fact that there was not a direct method for measuring the calibration curve. As opposed to AlOH and AuNPs, where the blank buffer contained the scattering species, the PLGA solutions were blanked against a transparent solution containing the buffer. Also unlike the AlOH and AuNPs, the PLGA microspheres were not a stable suspension and were therefore allowed to settle (for ~2 min) in the cavity before measurement. The concentration curves for both BSA and Lysozyme encapsulated

PLGA were attenuated similarly to the high concentration of AuNPs (Figure 2.2E, F). Nonetheless, protein concentrations were accurately quantified with the integrating cavity method (Figure 2.3E, F). To our knowledge, this is first reported method for a non-destructive measurement of protein concentration in PLGA microspheres.

To increase the utility of integrating cavity spectrometers, the ability to measure second derivative spectra in turbid media is of great interest. Attempts to interpret second derivate spectra for analysis of individual amino acids were unsuccessful (data not shown). This was attributed to inadequate signal-to-noise for quantitation of second derivative peak shifts which can be as small as ± 0.01 nm. Increased signal-to-noise could be achieved with a larger integrating cavity (increased path length) or more sensitive multiwavelength detector (multiplex advantage). Measurements of even lower concentrations (ng/ml) could theoretically be achieved with a light source which reaches into the deep UV at 205 nm[27]. More immediately applicable uses of this experimental setup include measurement of DNA or RNA encapsulated in lipids or any of the wide variety of nano/microparticle currently being investigated.

2.4 Conclusion

A label-free, direct approach to measure protein concentrations in turbid solutions was developed using an integrating cavity spectrometer. Adequate robustness of this method was shown with proteins in three model turbid systems, Alhydrogel, gold nanoparticles, and PLGA microspheres. Using calibration curves fitted to a simple nonlinear model, we were able to quantify protein concentrations in a broadly applicable manner. Our study presents a direct and non-destructive solution to quantify protein concentration in a variety of turbid solutions.

2.5 References

1. Layne, E. Spectrophotometric and Turbidimetric Methods for Measuring Proteins. *Methods Enzymol.* **1957**, *3* (C), 447–454.

2. Mach, H.; Middaugh, C. R.; Lewis, R. V. Statistical Determination of the Average Values of the Extinction Coefficients of Tryptophan and Tyrosine in Native Proteins. *Anal. Biochem.* **1992**, *200* (1), 74–80.

3. Reed, S. G.; Orr, M. T.; Fox, C. B. Key Roles of Adjuvants in Modern Vaccines. *Nat. Med.* **2013**, *19* (12), 1597–1608.

4. Makadia, H. K.; Siegel, S. J. Poly Lactic-Co-Glycolic Acid (PLGA) as Biodegradable Controlled Drug Delivery Carrier. *Polymers (Basel).* **2011**, *3* (3), 1377–1397.

5. Demchanko, A. *Ultraviolet Spectroscopy of Proteins*, 3rd ed.; Springer, Berlin, Heidelberg, 1986.

6. Hodgkinson, J.; Masiyano, D.; Tatam, R. P. Using Integrating Spheres as Absorption Cells: Path-Length Distribution and Application of Beer's Law. *Appl. Opt.* **2009**, *48* (30), 5748–5758.

7. Pope, R. M.; Fry, E. S. Absorption Spectrum (380–700 Nm) of Pure Water II Integrating Cavity Measurements. *Appl. Opt.* **1997**, *36* (33), 8710.

8. Marcero, J. R.; Piel, R. B.; Burch, J. S.; Dailey, H. A. Rapid and Sensitive Quantitation of Heme in Hemoglobinized Cells. *Biotechniques* **2016**, *61* (2), 83–91.

9. Blake, R. C.; Griff, M. N. In Situ Spectroscopy on Intact Leptospirillum Ferrooxidans Reveals That Reduced Cytochrome 579 Is an Obligatory Intermediate in the Aerobic Iron Respiratory Chain. *Front. Microbiol.* **2012**, *3* (APR), 1–10.

10. Jones, L. S.; Peek, L. J.; Power, J.; Markham, A.; Yazzie, B.; Middaugh, C. R. Effects of Adsorption to Aluminum Salt Adjuvants on the Structure and Stability of Model Protein Antigens. *J. Biol. Chem.* **2005**, *280* (14), 13406–13414.

11. Wangoo, N.; Suri, C. R.; Shekhawat, G. Interaction of Gold Nanoparticles with Protein: A Spectroscopic Study to Monitor Protein Conformational Changes. *Appl. Phys. Lett.* **2008**, *92* (13), 1–4.

12. Kirby, G. T. S.; White, L. J.; Rahman, C. V.; Cox, H. C.; Rose, F. R. A. J.; Hutmacher, D. W.; Shakesheff, K. M.; Woodruff, M. A. PLGA-Based Microparticles for the Sustained Release of BMP-2. *Eur. Cells Mater.* **2011**, *22* (SUPPL.3), 24.

13. Zhu, D.; Saul, A.; Huang, S.; Martin, L. B.; Miller, L. H.; Rausch, K. M. Use of O-Phthalaldehyde Assay to Determine Protein Contents of Alhydrogel-Based Vaccines. *Vaccine* **2009**, *27* (43), 6054–6059.

14. Brewer, J. M.; Roberts, C. W.; Stimson, W. H.; Alexander, J. Accurate Determination of Adjuvant-Associated Protein or Peptide by Ninhydrin Assay. *Vaccine* **1995**, *13* (15), 1441–1444.

15. Zhu, D.; Huang, S.; McClellan, H.; Dai, W.; Syed, N. R.; Gebregeorgis, E.; Mullen, G. E. D.; Long, C.; Martin, L. B.; Narum, D.; Duffy, P.; Miller, L. H.; Saul, A. Efficient Extraction of Vaccines Formulated in Aluminum Hydroxide Gel by Including Surfactants in the Extraction Buffer. *Vaccine* **2012**, *30* (2), 189–194.

16. Yeh, Y.-C.; Creran, B.; Rotello, V. M. Gold Nanoparticles: Preparation, Properties, and Applications in Bionanotechnology. *Nanoscale* **2012**, *4* (6), 1871–1880.

17. Xu, J. X.; Siriwardana, K.; Zhou, Y.; Zou, S.; Zhang, D. Quantification of Gold Nanoparticle UV-Vis Extinction, Absorption, and Scattering Cross-Section Spectra and Scattering Depolarization Spectra: The Effects of Nanoparticle Geometry, Solvent Composition, Ligand Functionalization, and Nanoparticle Aggregatio. *Anal. Chem.* **2017**, acs.analchem.7b03227.

18. Jávorfi, T.; Erostyák, J.; Gál, J.; Buzády, A.; Menczel, L.; Garab, G.; Razi Naqvi, K. Quantitative Spectrophotometry Using Integrating Cavities. *J. Photochem. Photobiol. B Biol.* **2006**, *82* (2), 127–131.

19. Fry, E. S.; Kattawar, G. W.; Strycker, B. D.; Zhai, P.-W. Equivalent Path Lengths in an Integrating Cavity: Comment. *Appl. Opt.* **2010**, *49* (4), 575–577.

20. Newville, M., Stensitzki, T., Allen, D. B., Ingargiola, A. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python. *Zenodo* **2014**.

21. Lima, I. Python for Scientific Computing Python Overview. *Mar. Chem.* **2006**, 10–20.

22. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9* (3), 99–104.

23. Donovan, J. W.; John, W. MACROMOLECULES : Changes in Ultraviolet Absorption Produced by Alteration of Protein Conformation Changes in Ultraviolet of Protein Conformation Absorption Produced by Alteration. **1969**, No. 8.

24. Li, S.; Peng, Z.; Leblanc, R. M. Method to Determine Protein Concentration in the Protein-Nanoparticle Conjugates Aqueous Solution Using Circular Dichroism Spectroscopy. *Anal. Chem.* **2015**, *87* (13), 6455–6459.

25. Bustamante, C.; Tinoco, I.; Maestre, M. F. Circular Differential Scattering Can Be an Important Part of the Circular Dichroism of Macromolecules. *Proc. Natl. Acad. Sci. U. S. A.* **1983**, *80* (12), 3568–3572.

26. Savo, R.; Pierrat, R.; Najar, U.; Carminati, R.; Rotter, S.; Gigan, S. Observation of Mean Path Length Invariance in Light-Scattering Media. *Science.* **2017**, *358* (6364), 765–768.

27. Scopes, R. K. Measurement of Protein by Spectrophotometry at 205 Nm. *Anal. Biochem.* **1974**, *59* (1), 277–282.

2.6 Figures and Tables

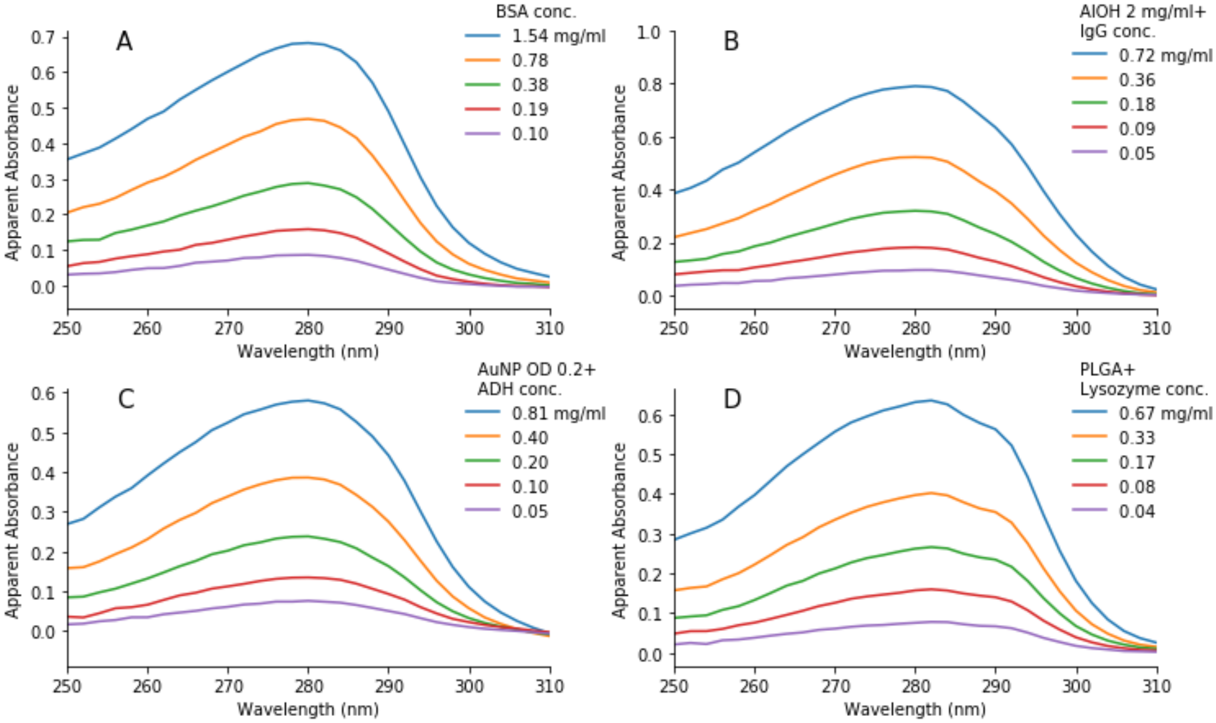


Figure 2.1. Four model proteins used in this study in various conditions. A) BSA in solution B) IgG on the surface of AlOH C) ADH with gold nanoparticles D) Lysozyme encapsulated within PLGA microspheres.

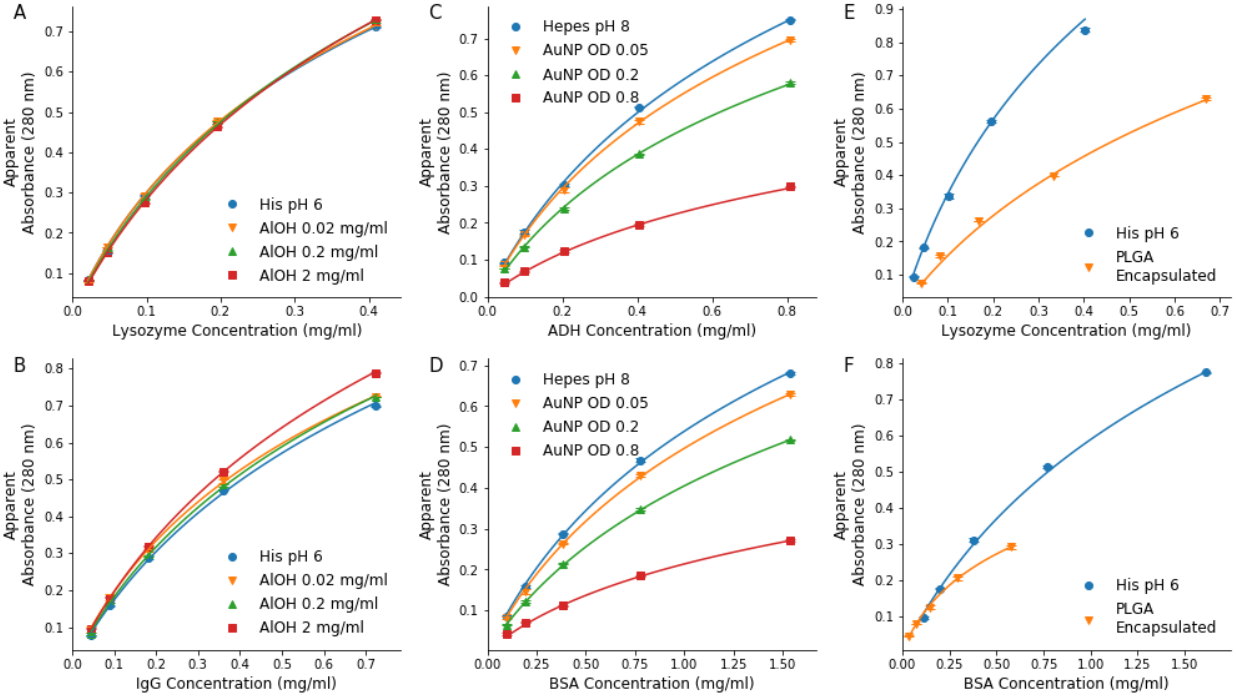


Figure 2.2. Fitted curves of protein A280 values observed using the integrating cavity. Five points were used to fit each curve. Error bars represent 95% confidence intervals on triplicate measurements. A) Lysozyme on AlOH B) IgG on AlOH C) BSA with AuNPs D) ADH with AuNPs E) Lysozyme encapsulated in PLGA F) BSA encapsulated in PGLA.

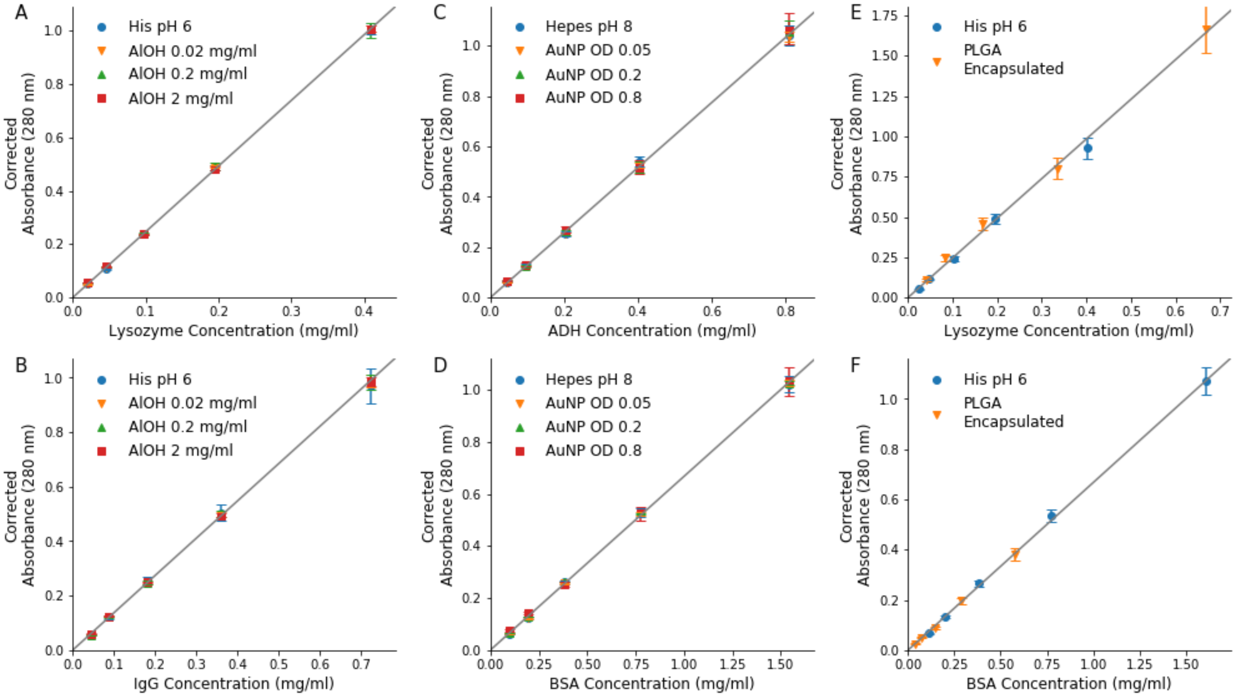


Figure 2.3. Comparison of empirically corrected absorbance spectra to conventional absorbance spectrometer. Error bars represent 95% confidence intervals on predicted absorbance. Gray lines represent A = εcl for each protein. Panels show A) Lysozyme on AlOH B) IgG on AlOH C) BSA with AuNPs D) ADH with AuNPs E) Lysozyme encapsulated in PLGA F) BSA encapsulated in PGLA.

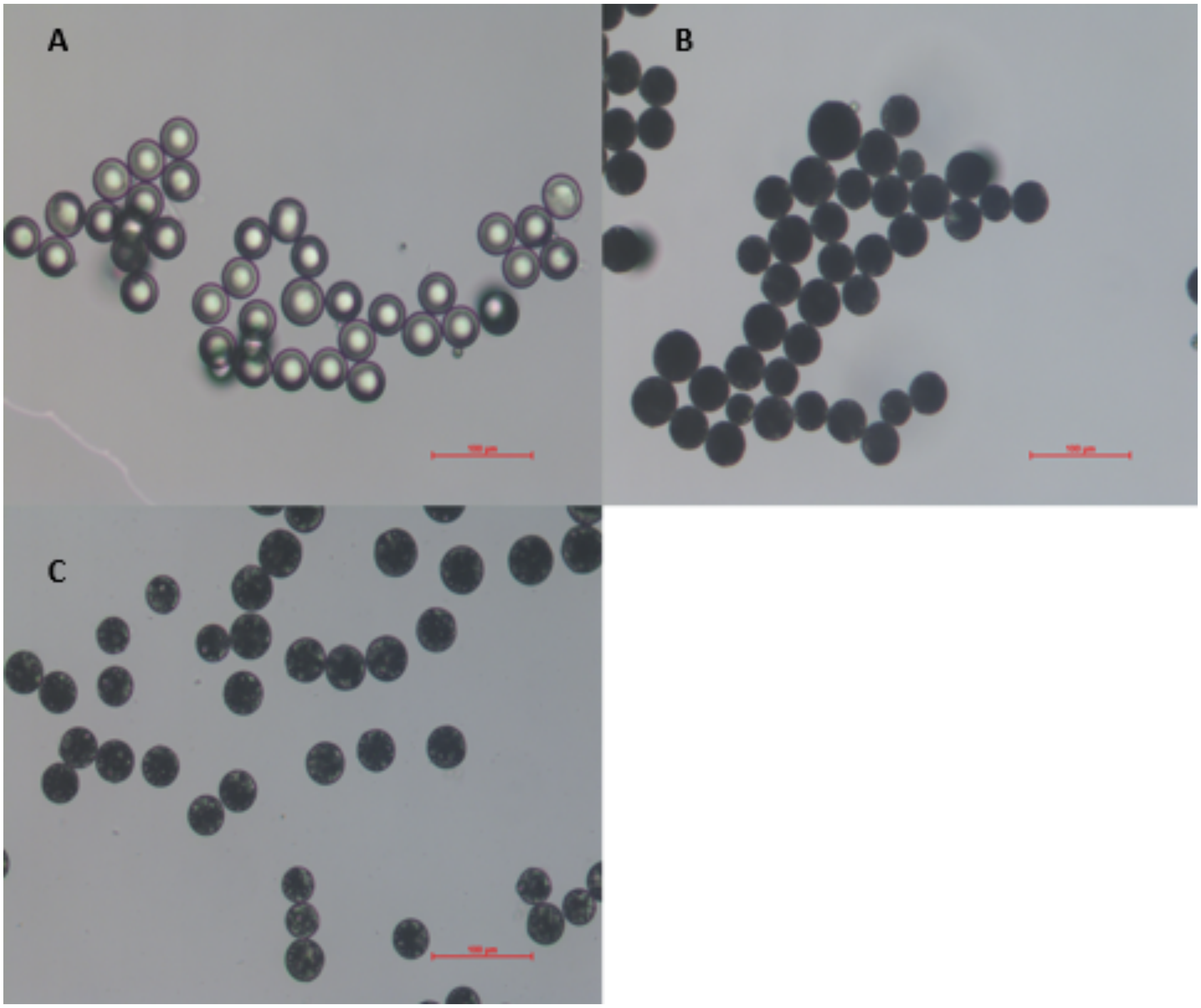


Figure 2.S1. Representative TEM images of PLGA microparticles. A) Empty microparticles B) Lysozyme loaded PLGA microparticles C) BSA loaded PLGA microparticles. The scale bars represent 100 μm.

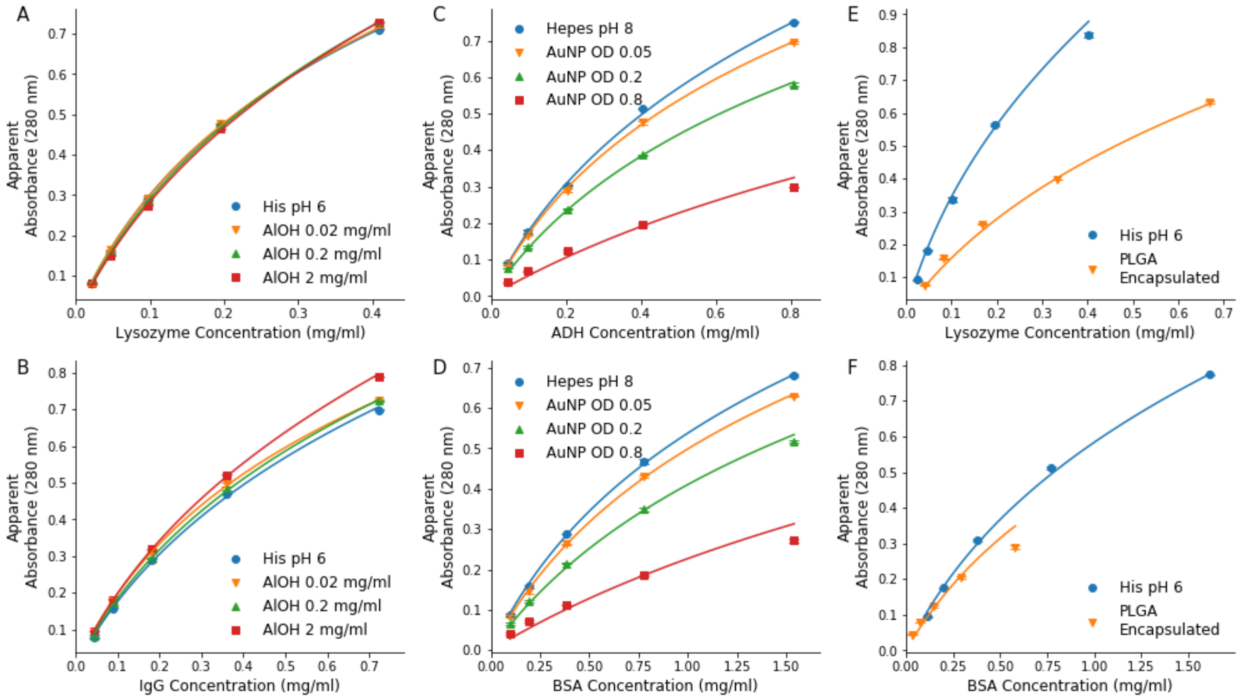


Figure 2.S2. Concentration curves fitted using the theoretical relationship derived by Fry et al.[21], five points were used to fit each curve. Error bars represent 95 % confidence intervals on triplicate measurements. A) Lysozyme on AlOH B) IgG on AlOH C) BSA with AuNPs D) ADH with AuNPs E) Lysozyme encapsulated in PLGA F) BSA encapsulated in PGLA.

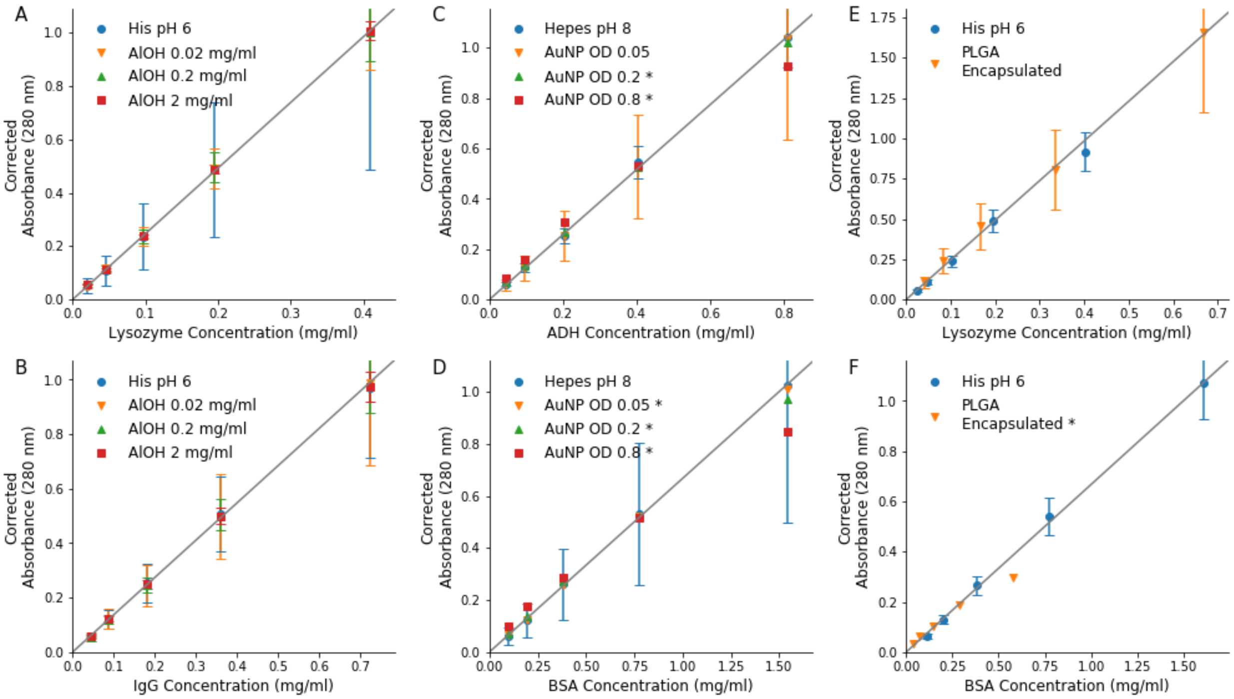


Figure 2.S3. Comparison of theoretically corrected absorbance spectra to conventional spectrometer. Error bars represent 95% confidence intervals on predicted absorbance. Gray lines represent A = εcl for each protein. An * denotes error bars were not shown for these measurements, 95% confidence intervals gave values with nonphysical meaning (reflectivity above 1 or negative effective pathlength). Panels show A) Lysozyme on AlOH B) IgG on AlOH C) BSA with AuNPs D) ADH with AuNPs E) Lysozyme encapsulated in PLGA F) BSA encapsulated in PGLA.

Table 2.S1. Comparing the theoretical and empirical fits for the highest concentration measured with each calibration curve.

Experiment	True conc. (μg/ml)	Empirical est. conc. (μg/ml)	Theoretical est. conc. (μg/ml)	Empirical % error	Theoretical % error
Lysozyme in 10 mM His pH 6	0.41	0.41 ± 0.01	0.41 ± 0.18	-0.1	-0.1
Lysozyme with 0.02 mgml AlOH	0.41	0.41 ± 0.00	0.41 ± 0.06	-0.1	-0.1
Lysozyme with 0.2 mgml AlOH	0.41	0.41 ± 0.01	0.41 ± 0.04	-0.7	-0.9
Lysozyme with 2 mgml AlOH	0.41	0.41 ± 0.01	0.41 ± 0.01	0.0	-0.1
IgG in 10 mM His pH 6	0.72	0.71 ± 0.05	0.71 ± 0.17	-2.3	-2.1
IgG with 0.02 mgml AlOH	0.72	0.72 ± 0.02	0.72 ± 0.19	-0.4	-0.4
IgG with 0.2 mgml AlOH	0.72	0.72 ± 0.02	0.72 ± 0.07	-0.7	-0.8
IgG with 2 mgml AlOH	0.72	0.72 ± 0.01	0.71 ± 0.04	-0.6	-1.7
BSA in 5 mM Hepes pH 8	1.54	1.53 ± 0.05	1.53 ± 0.65	-0.6	-0.5
BSA with GNP (OD 0.05)	1.54	1.54 ± 0.02	1.51 ± NA	-0.2	-1.7
BSA with GNP (OD 0.2)	1.54	1.54 ± 0.01	1.46 ± NA	-0.1	-5.3
BSA with GNP (OD 0.8)	1.54	1.55 ± 0.08	1.27 ± NA	0.5	-17.7
ADH in 5 mM Hepes pH 8	0.81	0.81 ± 0.03	0.81 ± 0.09	-0.1	-0.2
ADH with GNP (OD 0.05)	0.81	0.80 ± 0.01	0.80 ± 0.27	-0.6	-0.7
ADH with GNP (OD 0.2)	0.81	0.82 ± 0.03	0.79 ± NA	1.0	-1.9

ADH with GNP (OD 0.8)	0.81	0.83 ± 0.05	0.72 ± NA	2.4	-11.2
Lys in PLGA	0.67	0.68 ± 0.06	0.67 ± 0.22	1.4	1.0
Lys in 10 mM His pH 6 2	0.40	0.38 ± 0.03	0.37 ± 0.05	-6.3	-7.4
BSA in PLGA	0.58	0.57 ± 0.04	0.45 ± NA	-1.2	-23.0
BSA in 10 mM His	1.61	1.61 ± 0.08	1.61 ± 0.20	-0.3	-0.3

Table 2.S2. True concentrations (μg/ml) and predicted concentrations for both the empirical and theoretical model for the lowest concentration used in each curve.

Experiment	True conc. (μg/ml)	Empirical est. conc. (μg/ml)	Theoretical est. conc. (μg/ml)	Empirical % error	Theoretical % error
Lysozyme in 10 mM His	20.9	21.2 ± 0.3	21.0 ± 9.1	1.2	0.5
Lysozyme with 0.02 mgml AlOH	20.9	20.4 ± 0.2	20.3 ± 2.8	-2.7	-3.1
Lysozyme with 0.2 mgml AlOH	20.9	22.9 ± 0.5	22.5 ± 2.4	9.6	7.7
Lysozyme with 2 mgml AlOH	20.9	23.6 ± 0.3	23.1 ± 0.8	12.8	10.4
IgG in 10 mM His	45.2	40.0 ± 2.0	39.1 ± 9.8	-11.4	-13.5
IgG with 0.02 mgml AlOH	45.2	41.8 ± 0.8	41.4 ± 11.4	-7.5	-8.2
IgG with 0.2 mgml AlOH	45.2	40.6 ± 0.8	39.7 ± 4.4	-10.2	-12.0
IgG with 2 mgml AlOH	45.2	44.3 ± 0.5	43.3 ± 2.6	-1.9	-4.1
BSA in 5 mM Hepes pH 8	96.8	92.5 ± 2.3	91.8 ± 39.8	-4.5	-5.1
BSA with GNP (OD 0.05)	96.8	96.1 ± 0.8	98.0 ± NA	-0.7	1.2
BSA with GNP (OD 0.2)	96.8	96.7 ± 0.5	103.2 ± NA	-0.2	6.7
BSA with GNP (OD 0.8)	96.8	114.4 ± 4.5	147.0 ± NA	18.1	51.8
ADH in 5 mM Hepes	46.1	47.9 ± 1.4	47.1 ± 5.5	3.8	2.1
ADH with GNP (OD 0.05)	46.1	45.6 ± 0.6	45.4 ± 15.4	-1.2	-1.5
ADH with GNP (OD 0.2)	46.1	51.0 ± 1.7	54.4 ± NA	10.5	17.9

ADH with GNP (OD 0.8)	46.1	52.0 ± 2.2	66.4 ± NA	12.7	43.8
Lys in PLGA	41.8	44.2 ± 3.0	43.9 ± 14.9	5.7	5.1
Lys in 10 mM His	23.6	22.1 ± 1.2	21.3 ± 3.1	-6.0	-9.6
BSA in PLGA	36.3	38.3 ± 1.8	51.0 ± NA	5.7	40.6
BSA in 10 mM His	110.3	99.1 ± 4.0	96.7 ± 13.2	-10.1	-12.3

Table 2.S3. Sum square error (SSE) and Mean Absolute Percentage Error (MAPE) for empirical and theoretical equations for all fits.

Experiment	Empirical fit SSE	Theoretical fit SSE	Empirical fit MAPE	Theoretical fit MAPE
Lysozyme in 10 mM His pH 6	1.52E-05	1.73E-05	2.0	2.1
Lysozyme with 0.02 mgml AlOH	1.48E-05	1.91E-05	1.3	1.5
Lysozyme with 0.2 mgml AlOH	3.47E-05	5.14E-05	2.9	2.6
Lysozyme with 2 mgml AlOH	1.06E-05	1.62E-05	3.1	2.8
IgG in 10 mM His pH 6	3.81E-04	3.89E-04	4.0	4.1
IgG with 0.02 mgml AlOH	3.67E-05	4.34E-05	2.4	2.5
IgG with 0.2 mgml AlOH	7.74E-05	1.18E-04	2.8	3.1
IgG with 2 mgml AlOH	2.64E-05	1.68E-04	1.1	1.8
BSA in 5 mM Hepes pH 8	5.44E-04	5.91E-04	2.9	3.1
BSA with GNP (OD 0.05)	2.23E-04	8.49E-04	1.6	1.9
BSA with GNP (OD 0.2)	2.22E-04	7.37E-03	1.0	4.8
BSA with GNP (OD 0.8)	8.50E-04	8.30E-02	6.2	23.2
ADH in 5 mM Hepes pH 8	2.95E-04	4.16E-04	2.9	2.6
ADH with GNP (OD 0.05)	8.72E-05	9.81E-05	1.4	1.5
ADH with GNP (OD 0.2)	1.24E-04	3.60E-04	3.1	5.6
ADH with GNP (OD 0.8)	4.68E-04	1.07E-02	4.8	20.5

Lys in PLGA	7.44E-04	6.82E-04	7.9	7.6
Lys in 10 mM His pH 6 2	6.86E-04	9.50E-04	3.9	5.6
BSA in PLGA	1.62E-04	1.85E-02	3.4	20.1
BSA in 10 mM His	1.46E-03	2.15E-03	3.8	4.7
Average	3.23E-04	6.32E-03	3.1	6.1

Table 2.S4. Calculated LOQ and LOD for all calibration curves.

Experiment	LOQ (*µ*g/ml)	LOD (*µ*g/ml)
Lysozyme in 10 mM His pH 6	1.4	0.4
Lysozyme with 0.02 mgml AlOH	3.6	1.2
Lysozyme with 0.2 mgml AlOH	2.8	0.9
Lysozyme with 2 mgml AlOH	3.5	1.2
IgG in 10 mM His pH 6	2.8	0.9
IgG with 0.02 mgml AlOH	6.2	2.1
IgG with 0.2 mgml AlOH	4.9	1.6
IgG with 2 mgml AlOH	5.5	1.8
BSA in 5 mM Hepes pH 8	5.6	1.8
BSA with GNP (OD 0.05)	9.5	3.1
BSA with GNP (OD 0.2)	20.4	6.7
BSA with GNP (OD 0.8)	45.6	15.0
ADH in 5 mM Hepes pH 8	2.7	0.9
ADH with GNP (OD 0.05)	4.4	1.5
ADH with GNP (OD 0.2)	9.1	3.0
ADH with GNP (OD 0.8)	23.4	7.7
Lys in PLGA	3.1	1.0
Lys in 10 mM His pH 6 2	1.3	0.4
BSA in PLGA	4.3	1.4
BSA in 10 mM His	5.5	1.8

3 Comparison of Polysorbate 80 Hydrolysis and Oxidation on the Aggregation of a Monoclonal Antibody

3.1 Introduction

Polysorbate 20 and 80 are commonly used in protein therapeutics to prevent nonspecific adsorption of proteins to interfaces. Polysorbates owe their widespread prevalence to their ability to inhibit aggregation of proteins, possibly by preventing accumulation and subsequent aggregation of proteins at the interface. Polysorbate 80 (PS80) is a non-ionic surfactant with a critical micelle concentration (CMC) of 0.0017% (wt/v)[1]. The main species of PS80 contains a sorbitan head group with four chains of polyoxyethelyne (POE) extending from it. Theoretically, there are a total of twenty POE units which are attached to each head group, although in practice there may end up being more or less. Typically there is a Gaussian-like distribution in the number of POE units, resulting in a heterogeneous mixture. Of the four POE groups attached to the sorbitan head groups, one to three of them are esterified to fatty acids (FAs) at their ends which can also terminate in a primary alcohol. The fatty acids found in PS80 are 14 to 18 carbons long and can have up to three double bonds along the chain. The most abundant fatty acid is oleic acid (≥58%, 18 carbons, 1 double bond), followed by linoleic (≤18%, 18 carbons, 2 double bonds). The number of fatty acid substitutions on an individual sorbitan head group can range from zero to four. PS80 also has isosorbide head groups with zero to two fatty acid substitutions. There also exists a significant amount of POE-FAs unattached to the head groups. All of these components result in a diverse heterogeneous mixture which can vary extensively between manufacturers.

Due to its surface activity as well as small size, PS80 competes with surface active proteins such as monoclonal antibodies for the interfacial surface area[2]. By preventing protein surface adsorption, polysorbates decrease high local interfacial protein concentrations as well as adsorption-mediated conformational perturbation[3,4]. However, PS80 has been known in several cases to increase aggregation of quiescent proteins[5]. This is thought to be the result of specific interactions between polysorbate and proteins (e.g. stabilizing unfolded conformations)[6]. Although PS80 is added to prevent degradation of active pharmaceutical ingredients, PS80 itself is susceptible to degradation. There are two primary pathways through which polysorbate deteriorates, hydrolysis of the ester and oxidation.

Hydrolysis of the ester bond in PS80 can be promoted by heat, and catalyzed by acid, base, or enzymes such as esterases and lipases[7]. Enzymatic hydrolysis has been recognized as a primary pathway occurring in high concentration antibody formulations. Esterases and lipases from host-cell lines can degrade polysorbate at nanogram/mL levels, which may be under prescribed protein impurity limits in protein production processes. Phospholipase B-like 2 (PLBD2) has been identified as hydrolyzing polysorbates at concentrations as low as 200 ppm. To prevent polysorbate hydrolysis, Regeneron has developed a genetically modified host-cell line in which they engineered downregulation of PLBD2 expression[8]. Engineering out specific lipases can be effective, though identification of the exact enzyme(s) responsible may sometimes prove difficult[9]. The myriad of distinct species which make up PS80 also have different susceptibilities to degradation[10]. For instance, some esterases show preference for shorter fatty acid chains or less substituted head groups because of steric hindrance[11]. Regardless of the component degraded, free fatty acids are generated through ester hydrolysis (Figure 3.1). Free fatty acids have been implicated as initiating particle formation and identified within aggregates using FTIR microscopy[12,13].

Oxidation is the second primary route of PS80 degradation, with polysorbate heterogeneity playing a key role. Polysorbate can be oxidized through exposure to light, by trace peroxides leftover from manufacturing, or by transition metals. The unsaturated fatty acid esters are more susceptible to oxidation than their saturated counterparts[14]. The increased susceptibility to oxidation with an increasing number of double bonds stems from the stabilization of carbon-centered radicals generated by hydrogen abstraction from either allylic of bisallylic methylene groups. Figure 3.1 depicts a number of prominent degradation pathways for PS80. Oxidation of the double bond can occur in the case of oleic acid or other poly unsaturated fatty acid tails. Ultimately, oxidation results in the formation of peroxides, aldehydes, ketones and epoxides. Oxidation can also occur on the POE chain, which can result in POE-FA species in the case of substituted POE groups or in PS80 with a decreased number of POE substitutions for the unsubstituted POE groups[10]. Intermediary radicals generated from polysorbate can oxidize therapeutic proteins, for example at methionine and tryptophan residues[5,15]. This can result in alterations of protein higher order structure and generate oligomeric species.

In this work, PS80 is subjected to forced oxidation by AAPH and hydrolysis by *Candida Antarctica lipase B* (CALB). PS80 degradation is quantified with mass spectrometry to determine the fraction of native polysorbate species intact. A Langmuir trough is used to compare effects of degradation on the surface activity of the surfactant. A monoclonal antibody (mAb F) which previously exhibited microparticle formation during formulation was chosen to determine the effects of PS80 degradation on protein aggregation. Mab F was mixed with the degraded surfactants and subjected to shaking as mechanical stress. HPLC-SEC, MFI and visual inspection are used to examine the resulting aggregation. After hydrolysis, PS80 containing protein solutions resulted in increased particle counts and increased monomer loss as compared to controls. MFI morphological features differentiated particles formed in the presence of hydrolyzed PS80 as being larger and having higher aspect ratios than particles in the absence of PS80. However, even after extensive oxidation of PS80, mAb F retained similar levels of monomer to unstressed

PS80 following shaking. Thus, the current study demonstrates that PS80 hydrolysis and oxidation have distinct impacts on protein aggregation.

3.2 Materials and Methods

3.2.1 Materials

The monoclonal antibody (mAb F) was provided by Astra Zeneca (Gaithersburg, MD, USA). MAb F was dialyzed into 20 mM histidine buffer at pH 6.0 at a concentration of 10 mg/mL preceding accelerated degradation. PS80 was purchased from J.T. Baker (Radnor, PA) and diluted in water to a concentration of 10% (wt/v).

3.2.2 Polysorbate Forced Degradation

PS80 was degraded in water at a 1% (wt/v) concentration. *Candida Antarctica lipase B* (CALB) immobilized on immobead 150 was purchased from Sigma Aldrich (St. Louis, MO, USA). A concentration of 1.5 U/mL was spiked into a 1% (wt/v) PS80 stock solution. The solution was mixed using a rotating mixer for 18 h. Following hydrolysis the solution was filtered with a 0.5 µm PES filter to remove the enzymatically coated beads. This solution was stored at 4 °C in an amber bottle under nitrogen until further use.

PS80 (1%) containing solutions were spiked to a concentration of 10 mM 2,2'-azobis(2-amidinopropane) dihydrochloride (AAPH). Solutions were placed in Pyrex tubes capped with rubber stoppers. Solutions were irradiated for 6 h in a Rayonet system (Southern New England, Brandord, CT, RMA-500) with four phosphor coated low pressure mercury lamps with λ_{max} of 350 nm (RPR-3500A). This solution was stored at 4 °C in an amber bottle under nitrogen until further use.

3.2.3 UPLC-MS Analysis of PS80 and its Degradants

The separation of PS80 species and their degradants was performed using a Prevail™ C18 (Grace, Deerfield, IL) column with 150 mm × 2.1 mm dimension, 300 Å pore-size and 3.0 µm particles. The column temperature was maintained at 40 °C. One µg of samples were passed to a Waters Q-TOF Premier mass spectrometer (Milford, MA, USA) at a flow rate of 0.3 mL/min. Four solvents with different elution strength were used to elute the samples from the column; solvent A contained 20 mM ammonium formate and 0.08% formic acid in MilliQ water, solvent B was 30% methanol, 70% acetonitrile and 0.08% formic acid (v:v:v), solvent C was isopropanol, and to wash the column, solvent D with higher elution strength was prepared with 10% dioxane and 90% acetone (v:v). The PS80 species and its degradants were eluted with a starting gradient of 40% solvent A and 60% solvent B for 15 min and was then changed to 1% solvent A and 99% solvent B. At 20 min, 5% solvent C was introduced with 1% solvent A and 94% solvent B. At 21 min, the solvent gradient was set to 1%, 90% and 9% of solvent A, solvent B, and solvent C, respectively and was kept for 3 min. The column was washed with 1% solvent A and 99% solvent D for 3 min and at 28 min the column was equilibrated back to 40% solvent A and 60% solvent B for 4 min. The total run time was 32 min.

The UPLC was connected to a Water Q-TOF Premier (Milford, MA, USA) equipped with an electrospray ionization source. The samples were run in MS^E method, a data independent acquisition (DIA) mode, to record full scans in first function. PS80 species and its degradants were identified in a positive mode. The MS lenses were calibrated in positive ion mode using sodium iodide. The instrument was operated with source temperature at 120 °C, capillary voltage of 2.8 kV, and argon as a collision gas at a flow rate of 0.35 mL/min. Masslynx v.4.1 software was used to control the MS settings and analyze the data. Extracted ion chromatograms with a mass window of 0.1 Da were used to quantify the relative amounts of polysorbate 80 components in stock and degraded solutions.

3.2.4 Langmuir Trough

A Langmuir trough was used to quantify the surface pressure isotherms for PS80 and mAb containing components. Solutions of mAb F were made at 1 mg/mL with a PS80 concentration of 0.001 % (wt/v). Stock solutions of the different PS80 solutions, without the mAb were also run as controls, to study the effect of the two different degradation pathways on the surface activity of the PS80. 40 mL of solutions containing PS80 and proteins were dispensed into a Langmuir trough (Biolin Scientific, Inc, Stockholm, Sweden). Prior to use, the Langmuir trough was calibrated using water. Samples were dispensed quickly and immediately recorded to measure the adsorption to the air water interface. A paper Willhelmy plate was used to measure surface pressure over the course of two hours. Initial rate of change of surface pressure was calculated by fitting the first two minutes of the data to a linear equation and calculating the slope.

3.2.5 Critical Micelle Determination with Pyrene Fluorescence

Pyrene (Sigma Aldrich) was dissolved in DMSO and then spiked into PS80 at a concentration of 1 uM. Solutions of PS80 were prepared from 0.1% to $4.88*10^{-5}$% by two-fold serial dilutions. Fluorescence was measured at 25 °C using a Photon Technology International (PTI) spectrofluorometer (Lawrenceville, NJ) equipped with a turreted four-position Peliter-controlled cell holder and xenon lamp. Pyrene was excited by light at 335 nm and emission observed at 373 nm and 384 nm with an integration time of 5 s. The ratio of the fluorescence intensity of these two wavelengths (I1/I3) was fit to a Boltzmann-type sigmoid as described previously[16]. The fit was performed using in-house python scripts and the midpoint of the sigmoid corresponds to the CMC.

3.2.6 Mechanical Agitation Studies

Solutions of mAb F were transferred to 1 mL vials filled with 0.5 mL of 10 mg/mL mAb F in 20 mM histidine and capped with stoppers. Ten samples in total were run in duplicate during the shaking study. For each PS80 (stock, hydrolyzed, oxidized) samples were mixed to have a final concentrations of 0.01, 0.001, and 0.0001% (wt/v). The tenth sample was mAb F in the absence of PS80. Vials were placed on their side on a room temperature (22 °C) HS260 shaker (IKA Works, Inc., Wilmington, NC) for up to one week at a constant speed of 300 rpm. Vials were pulled at five time points (2 h, 8 h, 1 d, 3 d, 1 wk) for analysis by HPLC-SEC and MFI.

3.2.7 HPLC-SEC

SEC was performed using a HPLC instrument (Shimadzu, Kyoto, Japan) equipped with a photodiode array. The mobile phase was composed of 0.2 M sodium phosphate buffer (pH 6.8). Prior to injection on the HLPC, samples of mAb F were centrifuged at 13,000 xg for 10 min and the supernatant was diluted to 1 mg/mL prior to injection. Injections of 20 µL into a TSKgel SWXL guard columns (6.0 mm x 40 mm) followed by a TSKgel G3000Wxl stainless steel column (TosohBiosciences, Japan). Following this, the column was removed and twenty microliters of sample injected to quantify recovery for each sample. The SEC monomer peak area describes is the total area of the monomer peak quantified with absorbance at 280 nm. The SEC % monomer represents the fraction of the SEC trace which corresponded to the monomer peak.

3.2.8 Particle Flow Imaging

An MFI 5200 (Protein Simple, Santa Clara, California) flow imaging microscope was used to quantify particles generated during mechanical agitation. The instrument was flushed with 0.2 µm filtered water until a baseline of below 300 particles per mL was achieved. The instrument was focused with NIST traceable 25 µm Duke Standards (Thermo Fisher Scientific, Waltham,MA) prior to measurements. Samples of mAb F were diluted 10X (1 mg/mL) prior to flow imagining and run in duplicate. A volume of 1.5 mL was analyzed at a flow rate of 0.2 mL/min and capture speed of 21 frames/s. Particles only partially imaged (edge particles) were excluded from analysis. A range of 2 – 100 µm (diameter) was used to quantify the total number of particles in each sample. Particle morphology was assessed using in-house python scripts. A two dimensional histogram with 50 bins linearly spanning particle sizes from 2 – 100 µm and 50 bins from aspect ratios of 0 – 1 was used to calculate the number of particle in each of the 2500 bins.

3.3 Results

3.3.1 Polysorbate 80 Degradation and Quantitation

Following degradation by a lipase or oxidation, PS80 degradation was quantified using mass spectrometry. Figure 3.2B depicts the reverse phase chromatograms in comparison to control (non-degraded) PS80. Chromatograms of intact PS80 had eight main peaks, with the most abundant species in each peak eluting in the order of POE sorbitan linoleic acid, POE sorbitan oleic acid, POE oleic acid, POE sorbitan linolenic/oleic acid, POE sorbitan linoleic/oleic acid, POE sorbitan i-oleic acid, POE sorbitan stearic/oleic acid and POE sorbitan trioleic acid. All eight main peaks correspond to sorbitan head groups except the POE-FA peak which has no head group attached. Isosorbide species eluted in the same order about one minute before the equivalently substituted sorbitan species but were much smaller in peak area. The average length of the POE chains was ~26 in the control PS80. This did not vary significantly between species of substituted head group, however the POE chain lengths in POE-FA were much shorter at about 10 POE units.

After exposure to the esterase, PS80 shows a nearly complete loss of mono-substituted (sorbitan and isosorbide) and POE-FA species (Figure 3.2A). This is similar to a previous study[11] demonstrating that CALB preferentially degrades monosubstituted species. The di- and trisubstituted species remain almost completely intact. This enzyme did not show a preference between sorbitan and isosorbide head groups, degrading both impartially (data not shown). The primary degradation products following this reaction are free fatty acids (FFAs), POE-sorbitan, POE-isosorbides and n-POE groups (Figure 3.1, main hydrolysis products).

After exposure to AAPH, PS80 exhibited partial loss of all quantified species. The more double bonds on the PS80 FA chain resulted in increased loss, consistent with previous results[14]. The di-substituted PS80 species in Figure 3.2A substituted with oleic acid at one position and stearic, oleic, linoleic, or linolenic, (18 : 0, 18 : 1, 18 : 2, 18 : 3 – carbons : double bonds) at the other show a downward trend of the fraction remaining. As the number of substitutions on the oleic acids on sorbitan head groups increase from 1 to 3, the loss increases from 74 to 48 to 38% remaining. Oxidation of PS80 produces an extensive variety of products (Figure 3.1) originating from the oxidation of either double bonds or allylic and bisallylic methylene groups present on the fatty acids (e.g. oleic acid, lineoleic acid) or oxidation of the POE chains. The peak which forms at 17 minutes (Figure 3.2) corresponds to di-oleic sorbitan POE with a hydroperoxyl addition (+32) similar to the product upper right corner of Figure 3.1. A shift to a smaller average number of POE units was also observed, resulting from oxidative cleavage of POE chains.

3.3.2 Langmuir Trough

To determine how degradation affected the interfacial properties of PS80, a Langmuir trough was used to measure surface pressure isotherms. Figure 3.3a shows the adsorption isotherms (surface pressure vs. time) of the degraded and non-degraded PS80 solutions without the mAbs. Compared to histidine buffer alone, all PS80 solutions show significant surface activity. Interestingly, the hydrolyzed species show a slower rate of adsorption, and a lower maximum surface pressure, compared to the non-degraded PS80. On the other hand, the oxidized PS80 shows an increase in the maximum surface pressure. Figure 3.3b shows the adsorption isotherms in the presence of the mAb solutions. mAb F in the absence of PS80 had a slow increasing adsorption rate (Figure 3.3b). The mixtures of mAb F and PS80 had a higher surface affinity than mAb F alone, since the polysorbate can diffuse to the surface and absorb competitively with mAb F. For a mixture of oxidized PS80 and mAb F, the initial rate of sample adsorption was similar compared to mAb F and PS80 control. Oxidized PS80 had the equilibrium highest surface pressure compared to control and hydrolyzed PS80 whether mAb F was present or not. The initial rate of adsorption for the hydrolyzed PS80/mab F mixture was the slowest of all three samples decreasing from 2.95 nM m^{-1} min^{-1} for control PS80 to 1.76 nM m^{-1} min^{-1} for hydrolyzed PS80. This suggests that the diffusion to the surface for the hydrolyzed PS80/mab F mixture is significantly slower than that of control PS80/mAb F mixture. Both samples had similar equilibrium surface pressures suggesting the affinity for the air/water interface remained unchanged. The critical micelle concentration (CMC) for each control and degraded PS80 was measured (Table 3.1) and significant difference were not detected between samples.

3.3.3 Shaking Stress

Accelerated degradation of mAb F/PS80 mixtures was performed to probe the ability of degraded PS80 to act as a surfactant and prevent the interfacial aggregation of mAb F. HPLC-SEC was used to monitor the soluble fraction of antibodies and quantify soluble aggregates as well as the total monomer mass. During the shaking stress, mAb F lost 10% of the monomer fraction after one day and continued to lose monomer up to one week (Figure 3.4). A concentration of just 0.0001% control PS80 demonstrated protective effects, with the monomer fraction remaining intact after one full day. At concentrations of 0.01% and 0.001% PS80, a completely protective effect was provided against aggregation even for up to one week of shaking. Conversely, formulation with hydrolyzed PS80 resulted in a staggering loss of monomer at all concentrations. The monomer peak areas began to decline for all three concentrations of PS80 samples after just one day of shaking. Formulation with oxidized PS80 resulted in a monomer loss profile that looks similar to the one for control PS80, with the 0.0001% solution resulting in a significant loss of monomer. All of the samples on SEC remained at a constant monomer peak fraction of ~97.5% monomer (Figure 3.S1).

Aggregate particles were further quantified with MFI. For mAb F in the absence of polysorbate, a small number of particles (~3000) began forming after 2 h of shaking and reached a maximum number of particles after a day. For 0.0001% PS80, particles began forming after 8 h, although the total number of particles at day one was much less than solutions without PS80 (Figure 3.5). Similarly, hydrolyzed and oxidized PS80 showed particle formation after one day for 0.0001% PS80 containing solutions. For 0.001% and 0.01% PS80, both control PS80 and oxidized PS80 were protective against particle formation. However for hydrolyzed PS80, both 0.001% and 0.01% resulted in a large number of particles with 0.01% having the highest of all conditions after one week. The distribution of particles at early time points for no PS80 and

0.001% PS80 containing samples are shown as a heat map in Figure 3.6. The samples containing 0.001% hydrolyzed PS80 had the highest number of particles >25 um after 2 hr. The samples containing the PS80 control and oxidized PS80 had similar particle distributions in regard to size and aspect ratio.

Visual inspection was performed at all time points and vials after one week of shaking are pictured in Figure 3.7. At later time points (3 days, 1 week), particles in some samples began to stick to the vial. The morphology of the particles in solution also appears different between the different vials. Esterase degraded samples resulted in an opaque coating of samples on the surface of the vial, while mAb F shaken without PS80 resulted in a smaller snowflake like coating of particles. The 0.0001% oxidized polysorbate resulted in larger particles adsorbed to the surface of the vial.

3.4 Discussion

We have compared the two most common degradation mechanisms for polysorbate 80, oxidation and enzymatic hydrolysis, and their downstream effects on surfactant properties and protein aggregation. Mass spectrometry was used to quantify the degradation of the individual species of polysorbates. The CALB enzyme used here selectively cleaved monosubstituted PS80 species nearly completely. Use of immobilized enzymes allowed creation of a degraded PS80 that can be easily separated from the lipase and used for further experimentation. Similarly, the use of light and AAPH was effective for producing oxidized samples which could be stored for later use since all of the AAPH is reacted. Photolysis of AAPH results in the creation of the peroxyl radicals though the following proposed mechanism[17]: $\boxed{AAPH + h\nu \rightarrow {}^{1}AAPH \rightarrow {}^{3}AAPH}$ which in the presence of oxygen can convert to peroxyl radicals $\boxed{{}^{3}AAPH \rightarrow 2R^{\bullet} + N_2}$ and subsequently $\boxed{R^{\bullet} + O_2 \rightarrow 2ROO^{\bullet}}$. The peroxyl radicals generated showed preference for PS80 species with more double bond substitutions. Oxidized PS80 ranged from 10% to 70% intact (not degraded) depending on the species.

The rate at which PS80 diffuses to the air/water interface may be critical for prevention of antibody aggregation during shaking stress as PS80 is constantly being displaced and has to compete for interfacial surface area with proteins. The initial rate of change of surface pressure has been shown to predict the aggregation rates for a panel of antibodies[18]. For the samples described here, the degraded PS80/mAb F mixtures adsorbed had slower initial adsorption rates than control PS80, with hydrolyzed PS80 adsorbing the slowest. The decrease in initial rate of change in surface pressure could indicate hydrolyzed PS80 becomes less competitive with mAb F at the air/water interface, leading to higher mAb F interfacial concentrations which result in crowding and aggregation. Oxidized PS80 and control PS80 both had higher initial rates of change of surface pressure and were both effective at preventing mechanical aggregation. While differences were observed using the Langmuir trough, the CMC remained unchanged between intact and degraded PS80 as measured using pyrene (Table 3.1). Further, no differences were observed in the melting temperature of mAb F in the presence of control or degraded PS80 (data not shown). This suggests that differences between samples is not due to specific interaction of PS80 and mAb F.

For shaking stress, monomer loss rates were fastest for mAb F in the absence of PS80 and mAb F mixed with hydrolyzed PS80. Even at the highest concentration of 0.01%, hydrolyzed PS80 was unable to protect against monomer loss. The concentration of intact PS80 is 0.006% with 0.004% of polysorbate having been hydrolyzed. The concentration of intact PS80 should still be high enough to prevent protein aggregation as control PS80 at 0.001% was still effective at preventing monomer loss. This indicates that the hydrolyzed PS80 products are catalyzing protein aggregation. For all samples, the SEC soluble percent of dimer or higher molecular weight species percent did not increase during shaking (Figure 3.S1). This suggests that aggregation is occurring at the interface[2,19].

After just 2 h of shaking the hydrolyzed PS80 with mAb F, there is a larger number of particles over 25 μm than any of the other samples (Figure 3.6, top panel). The hydrolyzed polysorbate 80 resulted in particles with a morphology distinct from that of particles generated by samples of mAb F without PS80. The hydrophobic nature and low solubility of free fatty acids are responsible for creation of particles of this type[20]. We hypothesize these particles acted as nuclei which accelerated further protein aggregation and precipitation. The average aspect ratio for particles generated after 24 h (Figure 3.6, lower panels) of shaking in the presence of hydrolyzed PS80 was closer to 0.6 while for mAb without PS it was ~ 0.35. Further, the size range of the particles generated in the presence of hydrolyzed PS80 was wider, many particles bigger than 50 μm were observed.

The loss in the total number of particles observed by MFI (Figure 3.5, No PS80) can be explained both by a shift to larger average particle size as well as by the sticking of particles to the interior surface of the glass vials. For hydrolyzed PS80, both 0.0001% and 0.001% solutions resulted in particles sticking to the vial while 0.01% remained in solution (Figure 3.7). This corresponds well with the total particles measured per mL, where both 0.001% and 0.0001% see a drop in particle counts at three days and one week, while 0.01% continues to increase. This concentration dependence could stem from the low concentrations of PS80 inability to coat the surface of the vial. It is also possible that higher levels of intact PS80 can keep protein and/or fatty acid aggregates in solution as opposed to adsorbed on glass surfaces. While at 0.01% there is enough PS80 to prevent adsorption to the vial but aggregation still occurs. It is important to note that both MFI and SEC are unable to quantitate particles adsorbed to the surface of vails.

There has been considerable interest in designing improved surfactants as alternatives to polysorbate for protein stabilization[21,22]. While novel surfactants may be more effective than polysorbate at preventing degradation, designing a surfactant with resistance to many types of degradation will be critical. PS80 has been available in increasingly purified formulations recently and the Chinese Pharmacopeia changed its requirements form <58% oleic acid to <99% oleic acid as well as decreasing tolerances for peroxides. Linoleic and linolenic fatty acids are more susceptible to oxidative degradation than oleic acid. Therefore we anticipate that reduction in levels of linoleic and linolenic fatty acids will result in PS80 more robust to oxidative stress.

In this study, we found hydrolysis of PS80 to be more problematic for antibody formulations than PS80 oxidation. Therefore, monitoring and hindering the enzymatic hydrolysis of PS80 is critical for preventing aggregation. Methods to monitor levels of free fatty acids impurities can be used to compare different product purification steps or formulation conditions[23]. Specific enzymes which hydrolyze PS80 can be identified using proteomics approaches and knocked-out of a cell-line if necessary[9,12].

3.5 Conclusion

Polysorbate 80 is the most common excipient in protein formulation for its ability to prevent protein interfacial aggregation. To compare the effects of polysorbate oxidation and enzymatic hydrolysis, we used forced degradation. LC-MS was used to quantify the loss of individual components which make up PS80. Adsorption isotherms of PS80/mAb F mixtures were produced using Langmuir trough to investigate differences in surface activity of equivalent systems to those investigated with mechanical agitation. PS80 and PS80/mAb F mixtures degraded by hydrolysis had slower adsorption rates compared to control or oxidized PS80 and

PS80/mab F mixtures, respectively. Degraded PS80/mAbF solutions showed an increase in the equilibrium surface pressure, when the degraded through oxidation. During mechanical agitation studies of mAb F in the presence of PS80, control polysorbate was able to prevent aggregation at concentrations as low as 0.001%. After subjecting PS80 to ~40% hydrolysis a concentration of 0.01% PS80 could not protect against monomer loss, even though the concentration of intact PS80 remaining (0.006%) was higher than protective controls. Oxidized PS80 was similarly protective to control PS80, despite ~40% of PS80 being oxidized. During hydrolysis of PS80 the sole products are free fatty acids, while PS80 oxidation has several degradation pathways, with free fatty acids being a minor product. Hydrolyzed PS80 both generated more particles and had a slower initial rate of change in surface pressure compared to control and oxidized PS80. Since hydrolyzed PS80 has a slower initial rate of surface pressure change, higher concentrations mAb F can absorb to the interface and result in faster interfacial aggregation. The generation of insoluble hydrophobic particles from free fatty acids may accelerate protein monomer loss by acting as nucleation sites for protein aggregation. These effects could be synergistic, with both higher concentrations of mAb F at the interface and more interfacial nucleation sites at hydrolyzed free fatty acid particles.

3.6 References

1. Kerwin, B. A. Polysorbates 20 and 80 Used in the Formulation of Protein Biotherapeutics: Structure and Degradation Pathways. *J. Pharm. Sci.* **2008**, *97* (8), 2924–2935.

2. Khan, T. A.; Mahler, H. C.; Kishore, R. S. K. Key Interactions of Surfactants in Therapeutic Protein Formulations: A Review. *Eur. J. Pharm. Biopharm.* **2015**, *97*, 60–67.

3. Smith, C.; Li, Z.; Holman, R.; Pan, F.; Campbell, R. A.; Campana, M.; Li, P.; Webster, J. R. P.; Bishop, S.; Narwal, R.; Uddin, S.; van der Walle, C. F.; Lu, J. R. Antibody Adsorption on the Surface of Water Studied by Neutron Reflection. *MAbs* **2017**, *9* (3), 466–475.

4. Weltz, J. S.; Schwartz, D. K.; Kaar, J. L. Surface-Mediated Protein Unfolding as a Search Process for Denaturing Sites. *ACS Nano* **2016**, *10* (1), 730–738.

5. Wang, W.; Wang, Y. J.; Wang, D. Q. Dual Effects of Tween 80 on Protein Stability. *Int. J. Pharm.* **2008**, *347* (1–2), 31–38.

6. Singh, S. M.; Bandi, S.; Jones, D. N. M.; Mallela, K. M. G. Effect of Polysorbate 20 and Polysorbate 80 on the Higher-Order Structure of a Monoclonal Antibody and Its Fab and Fc Fragments Probed Using 2D Nuclear Magnetic Resonance Spectroscopy. *J. Pharm. Sci.* **2017**, *106* (12), 3486–3498.

7. Martos, A.; Koch, W.; Jiskoot, W.; Wuchner, K.; Winter, G.; Friess, W.; Hawe, A. Trends on Analytical Characterization of Polysorbates and Their Degradation Products in Biopharmaceutical Formulations. *J. Pharm. Sci.* **2017**, *106* (7), 1722–1735.

8. Burakov, D.; Goren, M.; Chen, G. Host Cell Protein Modification. **2016**, *1* (62), 37.

9. Chiu, J.; Valente, K. N.; Levy, N. E.; Min, L.; Lenhoff, A. M.; Lee, K. H. Knockout of a Difficult-to-Remove CHO Host Cell Protein, Lipoprotein Lipase, for Improved Polysorbate Stability in Monoclonal Antibody Formulations. *Biotechnol. Bioeng.* **2017**, *114* (5), 1006–1015.

10. Borisov, O. V.; Ji, J. A.; Wang, Y. J. Oxidative Degradation of Polysorbate Surfactants Studied by Liquid Chromatography-Mass Spectrometry. *J. Pharm. Sci.* **2015**, *104* (3), 1005–1018.

11. McShan, A. C.; Kei, P.; Ji, J. A.; Kim, D. C.; Wang, Y. J. Hydrolysis of Polysorbate 20 and 80 by a Range of Carboxylester Hydrolases. *PDA J. Pharm. Sci. Technol.* **2016**, *70* (4), 332–345.

12. Dixit, N.; Salamat-Miller, N.; Salinas, P. A.; Taylor, K. D.; Basu, S. K. Residual Host Cell Protein Promotes Polysorbate 20 Degradation in a Sulfatase Drug Product Leading to Free Fatty Acid Particles. *J. Pharm. Sci.* **2016**, *105* (5), 1657–1666.

13. Cao, X.; Fesinmeyer, R. M.; Pierini, C. J.; Siska, C. C.; Litowski, J. R.; Brych, S.; Wen, Z. Q.; Kleemann, G. R. Free Fatty Acid Particles in Protein Formulations, Part 1: Microspectroscopic Identification. *J. Pharm. Sci.* **2015**, *104* (2), 433–446.

14. Yao, J.; Dokuru, D. K.; Noestheden, M.; Park, S. S.; Kerwin, B. A.; Jona, J.; Ostovic, D.; Reid, D. L. A Quantitative Kinetic Study of Polysorbate Autoxidation: The Role of Unsaturated Fatty Acid Ester Substituents. *Pharm. Res.* **2009**, *26* (10), 2303–2313.

15. Singh, S. R.; Zhang, J.; Dell, C. O.; Hsieh, M.; Goldstein, J.; Liu, J.; Srivastava, A. Effect of Polysorbate 80 Quality on Photostability of a Monoclonal Antibody. **2012**, *13* (2), 422–430.

16. Aguiar, J.; Carpena, P.; Molina-Bolívar, J. A.; Carnero Ruiz, C. On the Determination of the Critical Micelle Concentration by the Pyrene 1:3 Ratio Method. *J. Colloid Interface Sci.* **2003**, *258* (1), 116–122.

17. Shao, J.; Geacintov, N. E.; Shafirovich, V. Oxidative Modification of Guanine Bases Initiated by Oxyl Radicals Derived from Photolysis of Azo Compounds. *J. Phys. Chem. B* **2010**, *114* (19), 6685–6692.

18. Geng, S. B.; Wittekind, M.; Vigil, A.; Tessier, P. M. Measurements of Monoclonal Antibody Self-Association Are Correlated with Complex Biophysical Properties. *Mol. Pharm.* **2016**, *13* (5), 1636–1645.

19. Ghazvini, S.; Kalonia, C.; Volkin, D. B.; Dhar, P. Evaluating the Role of the Air-Solution Interface on the Mechanism of Subvisible Particle Formation Caused by Mechanical Agitation for an IgG1 MAb. *J. Pharm. Sci.* **2016**, *105* (5), 1643–1656.

20. Doshi, N.; Demeule, B.; Yadav, S. Understanding Particle Formation: Solubility of Free Fatty Acids as Polysorbate 20 Degradation Byproducts in Therapeutic Monoclonal Antibody Formulations. *Mol. Pharm.* **2015**, *12* (11), 3792–3804.

21. Katz, J. S.; Nolin, A.; Yezer, B. A.; Jordan, S. Dynamic Properties of Novel Excipient Suggest Mechanism for Improved Performance in Liquid Stabilization of Protein Biologics. *Mol. Pharm.* **2018**, acs.molpharmaceut.8b00984.

22. Schiefelbein, L.; Keller, M.; Weissmann, F.; Luber, M.; Bracher, F.; Frieß, W. Synthesis, Characterization and Assessment of Suitability of Trehalose Fatty Acid Esters as Alternatives for Polysorbates in Protein Formulation. *Eur. J. Pharm. Biopharm.* **2010**, *76* (3), 342–350.

23. Cheng, Y.; Hu, M.; Zamiri, C.; Carcelen, T.; Demeule, B.; Tomlinson, A.; Gu, J.; Yigzaw, Y.; Kalo, M.; Yu, X. C. A Rapid High Sensitivity RP-UHPLC-MS Method for Assessing Polysorbate 20 Degradation in Protein Therapeutics. *J. Pharm. Sci.* **2019**, *1*, 1–7.

3.7 Figures and Tables

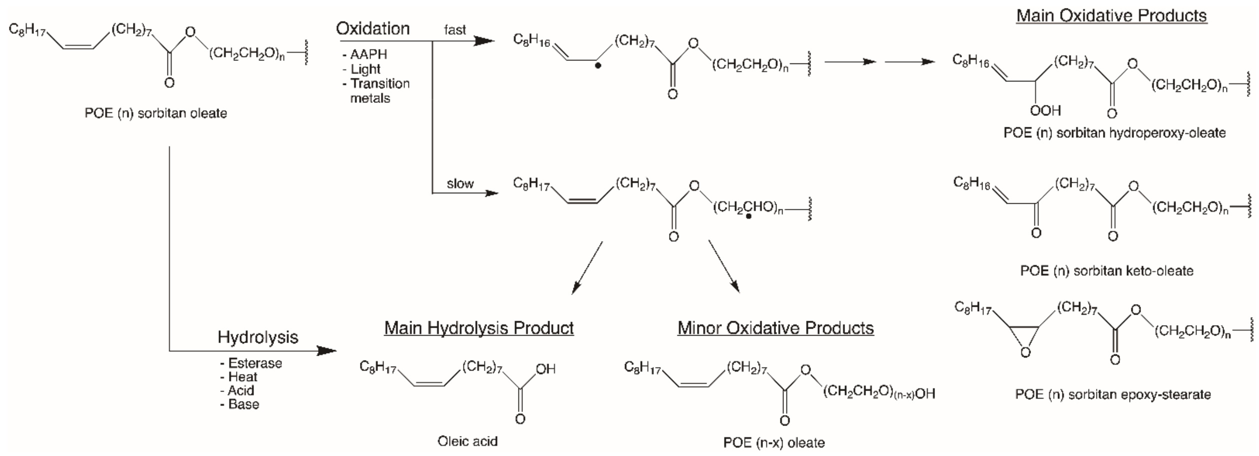


Figure 3.1. Reaction scheme for PS80 degradation, POE(n) sorbitan oleate is shown as it makes up the largest percentage of PS80. The oxidative pathway can proceed through a radical adjacent to the double bond. This pathway leads to n- POE (n)The radical can also form on the carbons in the POE chain. This results in products similar to the initial PS80 with a shorter POE chain POE (n-x) sorbitan oleate.

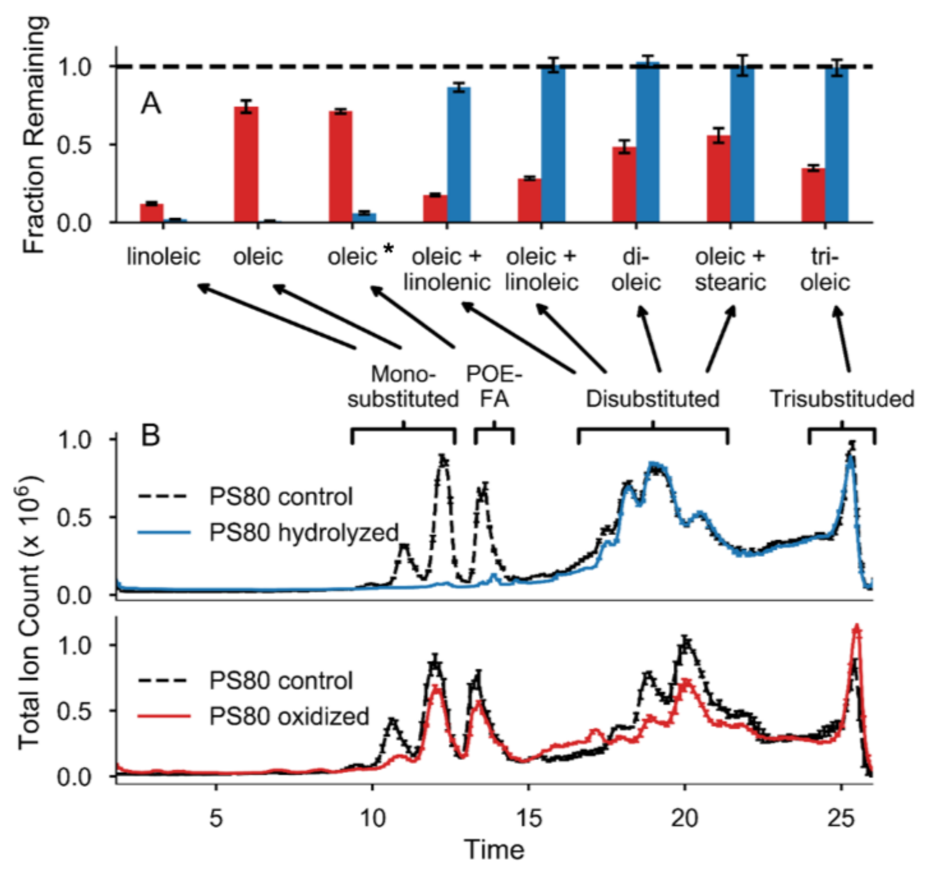


Figure 3.2. A) Fraction of PS80 species intact after forced degradation. Species depicted are POE sorbitan substituted except for POE-FA which is just POE oleic acid. The number of POE units in all cases is 26 except for POE-FA there are 10 POE units. Error bars represent the standard deviation of three replicates. B) Reverse phase chromatograms for PS80. Brackets denote the number of fatty acid substitutions on the head group (sorbitan, isosorbide). Top: Chromatograms for stock PS80 and hydrolyzed PS80. Bottom: Chromatograms for stock PS80 and oxidized PS80. Error bars represent the standard deviation of three replicates. *Oleic acid listed twice, the first oleic acid species corresponds to monosubstituted sorbitan oleic acid while second species corresponds to polyethylene-oxide oleic acid.

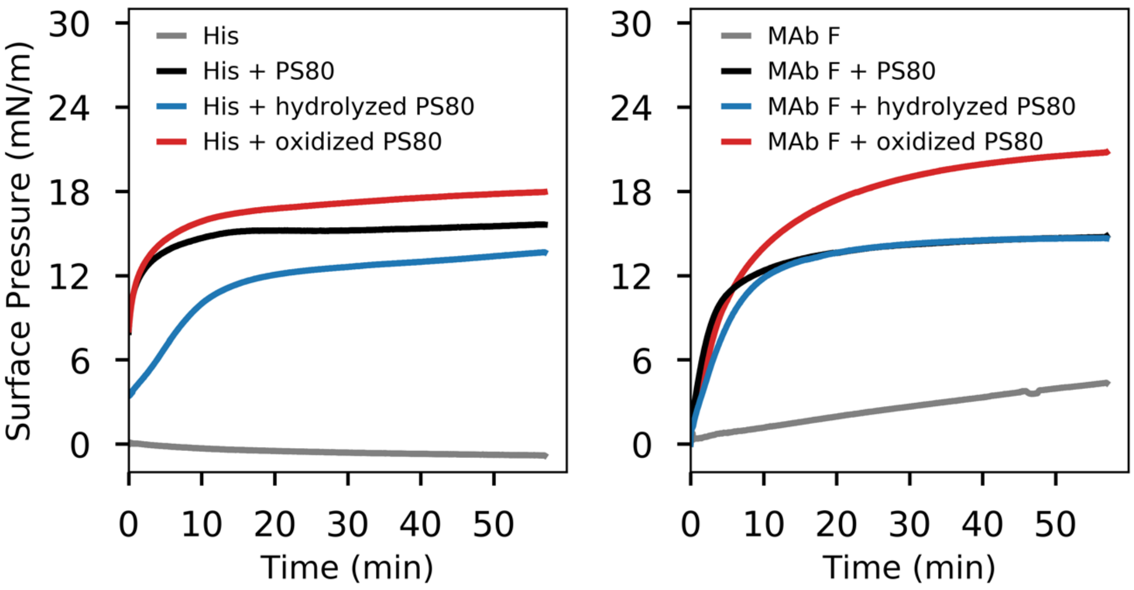


Figure 3.3. A. Surface pressure vs. time for stock and degraded PS80 solutions, as well as the histidine buffer without PS80. The graphs represent the average of two replicates. B. Surface pressure isotherms of mAb F with stock and degraded PS80 solutions. All solutions also contain 20 mM Histidine buffer.

Stock PS80	Hydrolyzed PS80	Oxidized PS80
0.0012 ± 0.0002	0.0011 ± 0.0004	0.0013 ± 0.0003

Table 3.1. Critical Micelle Concentration values (wt/v %) for polysorbate and degraded polysorbates.

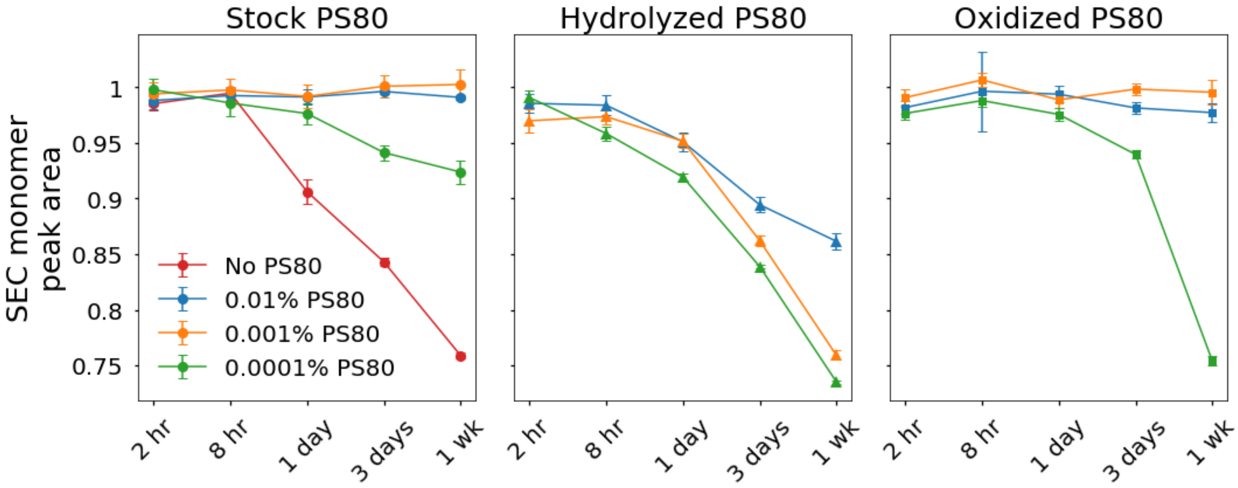


Figure 3.4. Monomer area peak fraction for the various polysorbate 80 time points. Peak areas are normalized to the sample at T0.

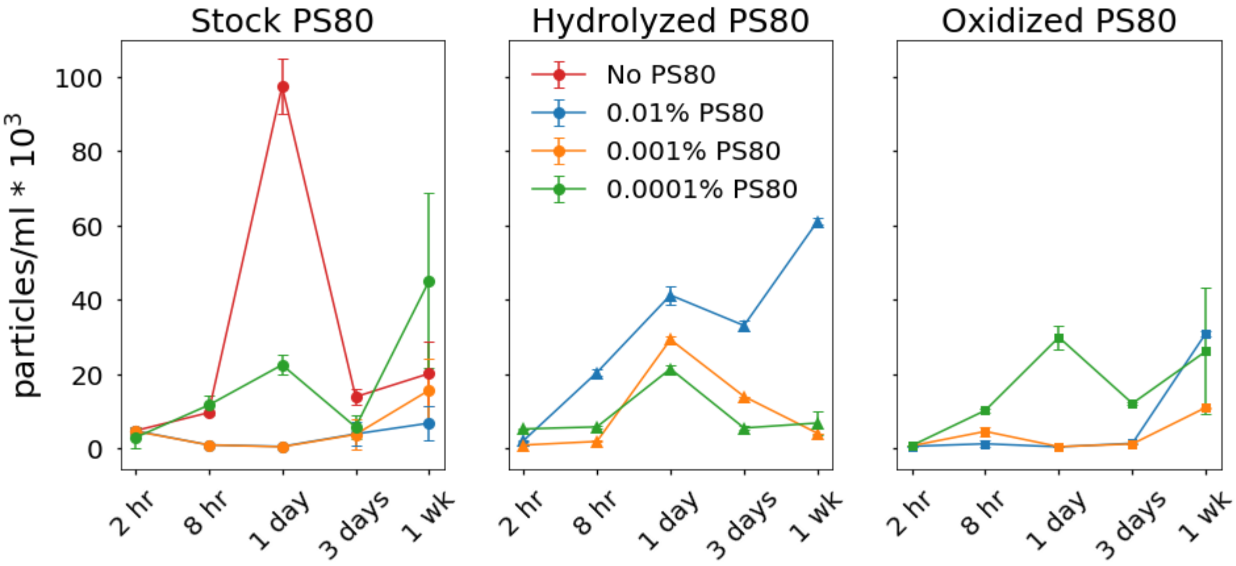


Figure 3.5. Particle counts for shaken samples at various timepoints for the different degraded polysorbates. Particle counted are in the size range of 2 – 100 µm. Error bars represent the standard deviation of two replicates.

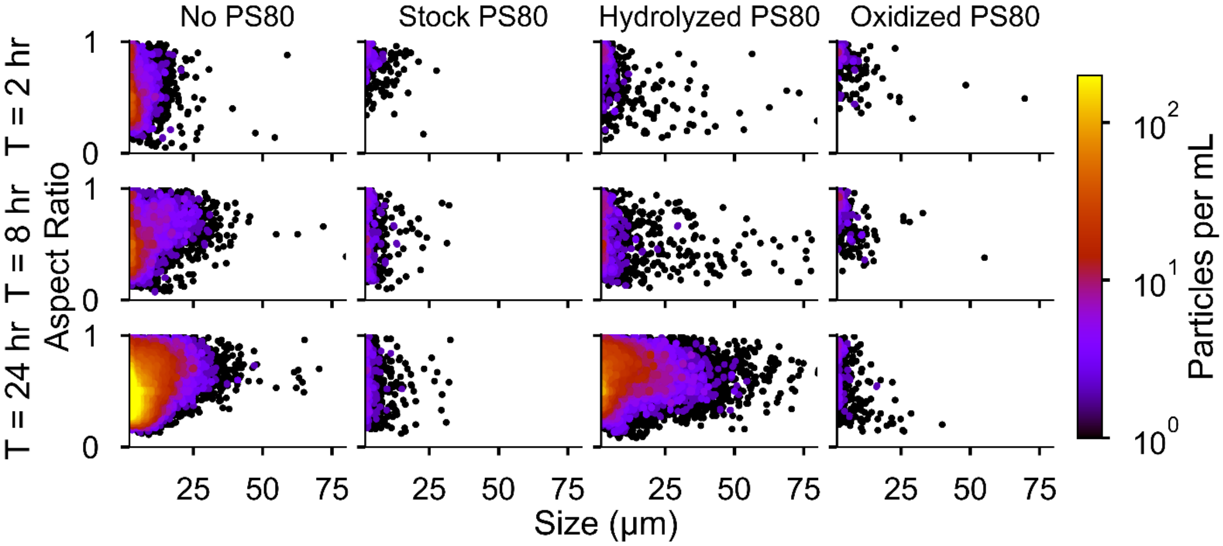


Figure 3.6. Particle morphology heat maps for the 2 h, 8 h and 24 h timepoints of the shaking study. The PS80 concentration in these samples was 0.001% (wt/v). Points on the plot represent individual measured particles. The color of each point is scaled by a 2d histogram binning of the number of particles per mL as described.

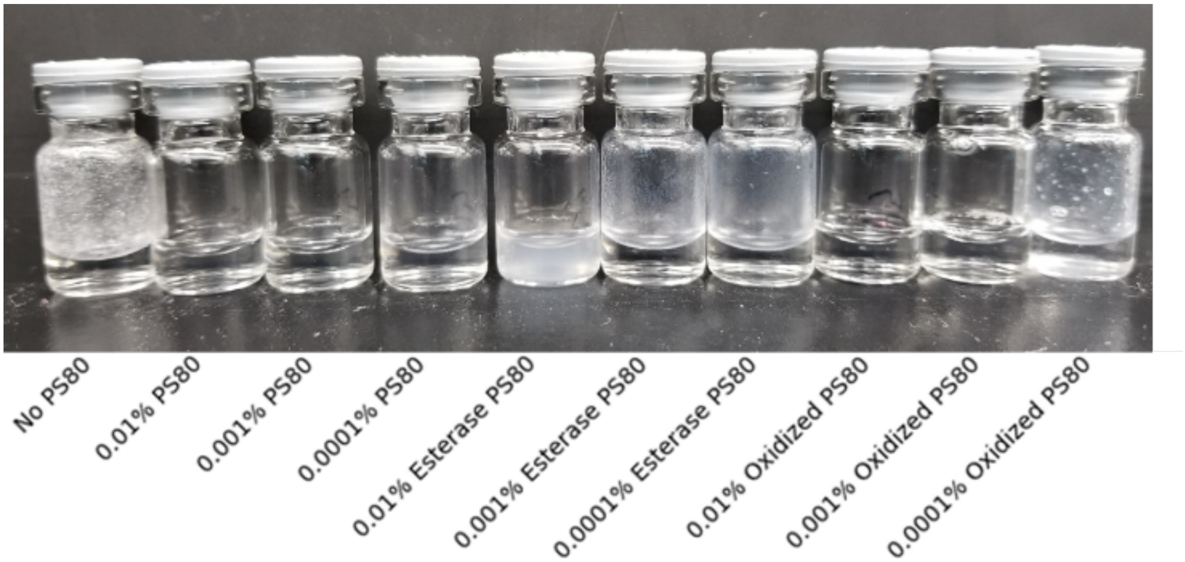


Figure 3.7. Picture of vials of mAb F after one week of shaking with various polysorbate concentrations and degraded conditions.

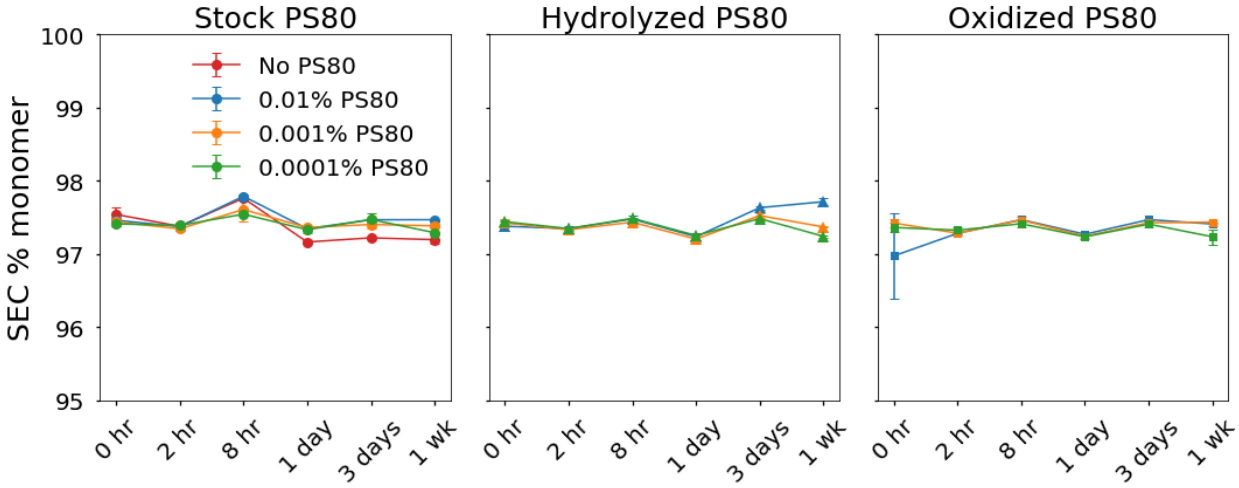


Figure 3.S1. Change of SEC determined monomer percent over time for mAb F formulated with each type and concentration of PS80.

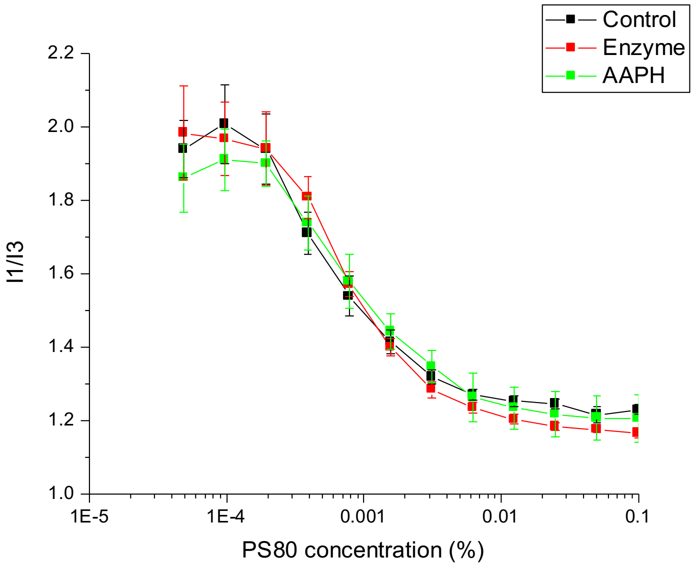


Figure 3.S2. Ratio of fluorescence pyrene peaks at 373 nm to 384 nm (I1/I3) used to determine the CMC for PS80 and its degradants.

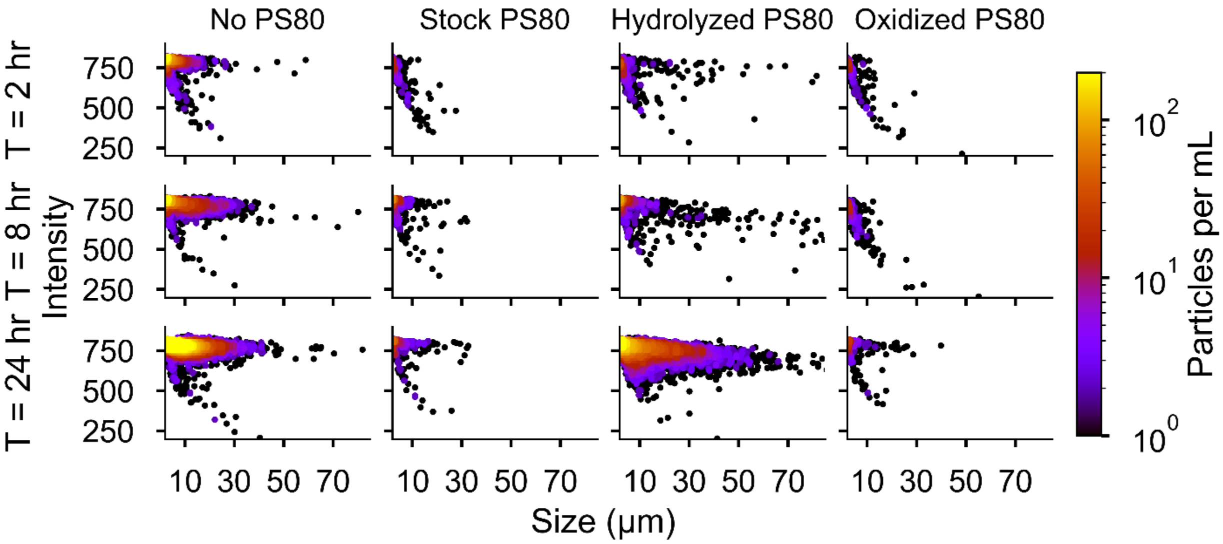


Figure 3.S3. Particle morphology heat maps for the 2 h, 8 h and 24 h timepoints of the shaking study. The PS80 concentration in these samples was 0.001% (wt/v). Points on the plot represent individual measured particles. The color of each point is scaled by a 2d histogram binning of the number of particles per mL as described.

4 Conformational Changes and Drivers of Monoclonal Antibody Liquid-Liquid Phase Separation

4.1 Introduction

Protein liquid-liquid phase separation (LLPS) occurs when a solution can achieve a lower energy state by partitioning itself into two phases[1]. An equilibrium is reached when the chemical potential is equal in the two phases. Thus, phase transitions occur about critical points of solution chemical potential. One phase is dense, viscous and protein-rich while the other phase is protein-poor[2–4]. Entropic and enthalpic effects can both promote LLPS. It is possible for non-interacting particles to exhibit phase separation solely via entropic effects (i.e. osmotic depletion)[5]. Most proteins, however, do not exhibit phase separation even at high protein volume fractions[2–4]. Therefore protein-specific interactions primarily dictate whether a protein will undergo phase separation and the shape of the liquid-liquid coexistence curve. Ionic, hydrophobic, pi-pi, van der Waals, hydrogen bonding, and other forces can all contribute to favorable protein-self interactions which drive LLPS[4].

Weak self-associative interactions at high volume fractions cause proteins in solution to form transient clusters[6]. In the case of phase separation, these clusters form larger branched networks of proteins which make up the protein-rich phase[7]. Neutron spin echo and small-angle neutron scattering experiments have identified the formation of dynamic clusters of proteins (including antibodies) in concentrated solutions[8–10]. Increases in solution viscosity are a result of higher average protein cluster size in solution[6,11]. For monoclonal antibody solutions, increases in viscosity are particularly problematic at high concentrations (>100 mg/ml), and many pharmaceutical formulation approaches have been developed to mitigate high solution viscosities[12,13].

LLPS also plays a critical role in compartmentalization within cells[4,14]. Accurate characterization of protein phase separated components in cells is performed primarily using imaging approaches. The use of many conventional biophysical tools for structurally characterizing proteins (e.g., CD, temperature melts, UV-Vis absorbance, etc.) is severely limited at such high protein concentrations. As LLPS is a thermodynamically reversible phenomenon, dilution of phases to lower protein concentrations results in solutions with equivalent physical properties (a single phase). Therefore physical characterization of LLPS cannot be performed by diluting the protein-rich and protein-poor phases to concentrations suitable for experimental characterization.

Here, we use a variety of spectroscopic techniques capable of characterizing protein phase separated components without dilution thereby preserving their actual physical state and allowing the study of protein conformational integrity in each phase. Protein structural differences between the protein-rich and protein-poor phases were measured using Raman, fluorescence and FTIR spectroscopy. Several spectroscopic observables undergo consistent shifts for all three antibody proteins in the individual phases. FTIR was also used to investigate the dynamics of water in the protein rich and protein poor phases. We determined specific sequences participating in protein-self interactions using hydrogen-deuterium exchange (HX-MS). Finally, we compared the aggregation rates for the three antibodies in the protein-rich and protein-poor phases and observed greater aggregation rates in the protein-poor phase for two of the antibodies and equivalent aggregation rates in two phases for the third antibody.

4.2 Materials and Methods

4.2.1 Materials and sample preparation

One IgG4 (mAb Z) and two IgG1 (mAbs H and I) monoclonal antibody stock solutions were provided by MedImmune LLC. These proteins were extensively dialyzed against buffers (Table 4.1) in which they could undergo LLPS at room temperature (room temperature, RT, ~ 22 °C) using dialysis cassettes (Slide-A-Lyzer™, 10 kDa MW cutoff, ThermoFisher Scientific, Waltham, MA). The individual phases could not be isolated directly from the dialysis cassettes due to high viscosity. The two resultant phases of each mAb were first mixed inside dialysis cassettes, then the mixture was transferred to a Falcon tube and kept still overnight at RT until two phases again formed with a clear boundary. Samples of each phase were obtained by decanting the protein-poor phase into a separate falcon tube. The pH and protein concentrations of the two phases (protein-poor and protein-rich) were measured using a pH meter and a Nanodrop spectrometer (Thermo Scientific), respectively (Table 4.1). These two phases of each mAb were later subjected to biophysical and analytical analyses.

To investigate the temperature effect on the LLPS of these mAbs, freshly dialyzed samples were stored in Eppendorf tubes at 25, 22, 15, and 4 °C. Concentrations of the newly formed protein-poor phases were measured every 24 h until they remained unchanged, indicative of thermodynamic equilibrium of LLPS. Upon equilibrium, concentrations of the protein-rich phases were measured using a Nanodrop spectrometer (Thermo Scientific) to construct liquid-liquid coexistence curves.

The liquid-liquid coexistence curves derived from mean field theory were fit as described[15] to:

$$|(C_c - C)/C_c| = A[(T_c - T)/T_c]^{\beta}$$

Where C_c is the critical protein concentration, T_c is the critical temperature in degrees Kelvin, A parameterizes the width of the coexistence curve, and β is the Ising exponent (0.325). The standard error of the mean on the fitted parameters was calculated as described[16].

4.2.2 Quantification of spectroscopic peaks

Spectroscopic peaks from Raman, fluorescence and FTIR were quantified using a variety of methods. The first three central moments of a peak were used to quantify peaks.

The zeroth moment is equivalent to the total peak area or total intensity:

$$\mathrm{I_{tot}} = \int \mathrm{I}(\lambda)\mathrm{d}\lambda$$

The first moment is equivalent to the peak mean over a given wavelength range:

$$\lambda_\mu = \frac{\int \lambda * I(\lambda)d\lambda}{\int I(\lambda)d\lambda}$$

The second central moment is the variance of the peak:

$$[\lambda_\sigma]^2 = \frac{\int (\lambda - \lambda_\mu)^2 I(\lambda)d\lambda}{\int I(\lambda)d\lambda}$$

The full width at half peak max (FWHM) for a Gaussian peak shape is the peak's standard deviation λ_σ multiplied by a factor:

$$FWHM = 2.355 * \lambda_\sigma$$

The peak maximum λ_{max} was also used by interpolating between points using a cubic beta spline interpolation to 0.01 cm^{-1} for determining the wavelength maximum of the amide I band by FTIR. Ratios of peak intensities of two wavelengths are donated as $I_{x/y}$, with x and y denoting the two wavelengths at which the ratio of intensities was taken.

4.2.3 Raman Spectroscopy

Raman spectra of protein samples were collected using a Zetasizer Helix (Malvern Instruments, Columbia, MD) equipped with a 785-nm laser (~280 mW). Protein samples (20 μL) were loaded into a metal micro-cuvette (Malvern) assembled with two quartz windows. For protein concentrations and buffers for each mAb, see Table 4.1. Samples were measured using 10 acquisitions of 20 sec each. Raw spectra were processed and analyzed using the Zetasizer Helix Analyze software (Malvern Instruments). Raman spectra were first buffer subtracted and normalized according to the phenylalanine peak (1003 cm^{-1}). The amide I band first moment and FWHM were calculated from 1630 – 1690 cm^{-1}.

4.2.4 FTIR Spectroscopy

FTIR analysis was performed using a Tensor-27 FTIR spectrometer (Bruker, Billerica, MA). For protein concentrations and buffers for each mAb, see Table 4.1. Protein samples were measured from 800 – 4000 cm^{-1} with a resolution of 4 cm^{-1}. A total of 256 scans were performed for both samples and buffers at 25 °C. For the thermal melting experiments, 64 scans were obtained and samples were scanned from 25 to 90 °C using an increment of 2.5 °C/step and an equilibration time of 2 min at each step. Buffer subtraction, water vapor and CO_2 compensation, baseline correction, and normalization of amide I band (1700 – 1600 cm^{-1}) were sequentially performed on the raw FTIR spectra using the OPUS V6.5 software (Bruker, Billerica, MA). Second derivative FTIR spectra were generated using a Savitzky–Golay filter with a window size of 9.

The bend + liberation FTIR band of water in these protein samples and their buffers were measured by using air as the blank. Samples were scanned from 800 to 4000 cm^{-1} using a resolution of 2 cm^{-1}. A total of 256 scans were performed. λ_{μ} was calculated between 1900 and 2300 cm^{-1} for the water band.

4.2.5 Intrinsic Fluorescence Spectroscopy

Intrinsic fluorescence spectra were obtained using a fluorescence plate reader as previously described[17,18]. Samples were loaded into a 384-well plate and silicone oil was added to avoid sample evaporation during thermal ramps. The plate was centrifuged at 2,200 × g for 1 min to remove air bubbles. For protein concentrations and buffers for each mAb, see Table 4.1. Samples were excited at 295 nm (>95% tryptophan emission) using an acquisition time of 100 ms, and emission light from 300 to 450 nm were collected. Temperature was ramped from 10 to 100 °C with an increment of 1 °C per step and an equilibration time of 2 min at each step. The mean fluorescence emission wavelength, λ_{μ}, was calculated between 300 and 400 nm. The λ_{μ} provides a better signal-to-noise ratio than the actual peak position (i.e. maximal point), although it is often 5 to 10 nm higher than the latter one due to the asymmetry of protein intrinsic fluorescence spectra. The integral of fluorescence spectra between 300 and 400 nm was taken as total fluorescence intensity (I_{tot}). Both parameters (λ_{μ} and I_{tot}) were plotted as a function of temperature to construct the melting curves. Their first derivative curves were generated using a Savizky-Golay smoothing function with a polynomial order of 2 and window size of 11 built by employing the Origin software (OriginLab Corporation, Northampton, MA). The local maximum was taken as the melting temperature.

4.2.6 Differential Scanning Calorimetry (DSC)

The overall conformational stability of proteins was investigated using a VP-Capillary micro-calorimeter (Malvern, UK). The buffers for each mAb are listed in Table 4.1. The three mAb samples (at 1 mg/mL) in their corresponding dialysis buffers were loaded into a 96-well plate kept in a 5 °C sample chamber. Samples was scanned from 10 to 100 °C with a ramp rate of 2 °C per min. Samples were pre-equilibrated for 15 min prior to each scan. Raw DSC thermograms were buffer subtracted, normalized for concentration, and baseline corrected using the DSC Origin software (OriginLab Corporation, Northampton, MA).

4.2.7 Static light scattering

Static light scattering was obtained using a fluorescence spectrometer (Photon Technology International, Birmingham, NJ) equipped with a xenon arc lamp, excitation and emission monochromators, and a temperature controllable 4-position cuvette holder. For protein concentrations and buffers for each mAb, see Table 4.1. Light at 450 nm was scattered at 90° detection angle with an acquisition time of 1 sec. Temperature was ramped from 25 to 100 °C using an increment of 1 °C per step and an equilibration time of 120 s.

4.2.8 Lyophilization of mAb H and mAb I Samples

Both mAb H and mAb I were lyophilized for use in HX-MS spectrometry studies. Samples of proteins at 5 mg/ml and 60 mg/ml were dialyzed against 40 mM phosphate and 10% trehalose at pH 7.5. A LyoStar II (SP Scientific, Warminster, PA) was employed to lyophilize the samples as described previously[19–21]. Lyophilized cakes were photographed. Reconstituted samples were passed through a 0.22 um filter and the absorbance at 280 nm was used to quantify the sample concentrations.

4.2.9 Karl-Fischer Titration

Water moisture levels of lyophilized mAb samples was determined using a C20 Compact Karl-Fischer Coulometer (Mettler-Toledo LLC., Columbus) though measurement of their reaction with AQUASTAR ® CombiCoulomat (EMD Chemicals Inc., Gibbstown, NJ). Weights of lyophilized cakes were used to calculate their water moisture content.

4.2.10 Hydrogen-Deuterium Exchange Mass Spectrometry

HX-MS was performed for both mAb H and mAb I as described perviously[19]. Lyophilized vials containing 5 and 60 mg/ml protein in 40 mM phosphate with 10% trehalose at pH 7.5 were equilibrated at 25 °C. Samples were reconstituted in an equivalent deuterium buffer for labeling times of 120, 10^3, 10^4, and 10^5 s. A quench buffer of 0.5 M tris (2-carboxyethyl) phosphine hydrochloride (TCEP), 4 M guanidine hydrochloride and 0.2 M sodium phosphate at 1 °C was used to halt deuteration by diluting 20 µL of the sample with 180 µl of quench buffer. Quenched samples were loaded into the sample loop using a LEAP H/DX PAL (LEAP Technologies, Carrboro, NC). The LC (Agilent 1200 series, Santa Clara, CA) consisted of an immobilized pepsin column, peptide desalting trap, and a C18 column. The total mass of sample injection was kept constant by modifying the injection volume. Three independent replicates were measured using a quadrupole-time-of-flight mass spectrometer (Agilent 6530, Santa Clara, CA) with an electrospray ionization source in positive mode. Complete deuterium uptake plots for all timepoints and concentrations are included in the supplemental information as PDFs, black circles represent the 5 mg/ml concentration and red triangles represent the 60 mg/ml concentration.

4.2.11 Homology Modelling

Homology models were constructed using by modeling. Homology models of the Fab domain of each antibody were generated using the Antibody Modeler within MOE[22], a computational modeling package from CCG. Templates were identified using X-ray structures of the most homologous antibodies from more than 4,000 antibody structures in the Protein Data Bank (PDB).

4.2.12 Protein aggregation in Two Phases by Size Exclusion Chromatography (SEC)

Samples were held at 40 °C in the instrument autosampler for a two-week stability study. For protein concentrations and buffers for each mAb, see Table 4.1. The protein-poor and protein-rich phases were separated using positive displacement pipettes and placed in HPLC vials with insets (150 µL total volume). To prevent sample evaporation, a small amount (15 µL) of silicone oil was overlaid on the sample. Injections were performed without diluting the sample. For the protein-rich phase, 0.2 µL of sample was injected onto the column. For the protein-poor phase, the amount injected was different for each antibody but determined so that the total protein mass of the injection was equivalent to the protein-rich phase. This was done by multiplying 0.2 µL by the concentration ratio of the protein-rich and protein-poor phases.

SEC was performed employing an HPLC instrument (Shimadzu, Kyoto, Japan) equipped with an autosampler, a column oven, and a UV detector. Samples were injected on a TSKgel SWXL guard column (6.0 mm x 40 mm) followed by a TSKgel G3000SWXL SEC column (7.8mm x 300 mm) at a flow rate of 0.7 mL/min for a total run time of 30 min. The mobile phase used was 0.2 M sodium phosphate, pH 6.8, and the column oven was set at 30 °C. SEC chromatograms were detected at 280. Total peak areas remained approximately constant through-out the stability study and protein recovery was near one hundred percent.

4.3 Results

4.3.1 LLPS of Three Different mAbs

The three mAb samples had previously been observed to undergo LLPS in low ionic strength solutions at pH values near their pI. We extensively dialyzed each of these three proteins against specific buffers in which LLPS occurred at room temperature as described in Table 4.1.

High concentrations of proteins are known to exhibit the Gibbs-Donnan effect during dialysis due to preferential retention of oppositely charged ions by charged protein molecules resulting in an offset in solution conditions (pH and ion concentrations) between dialysis buffers and dialyzed samples[23]. Therefore we measured the pH of the dialyzed mAb samples to evaluate the magnitude of the Gibbs-Donnan effect during dialysis. Both protein-rich and protein-poor phases of each mAb were found to have pH values close to the pH of their dialysis buffers (±0.04), suggesting complete dialysis and an insignificant Gibbs-Donnan effect (**Error! Reference source not found.**). This may be attributed to the net neutral charge of mAbs in their dialysis buffers with pH values close to their pI. It therefore seems reasonable to assume a similar buffer composition in the two phases for each of these three mAbs.

4.3.2 Temperature-Dependent LLPS of Three mAb Proteins

The LLPS of proteins is a temperature-dependent thermodynamic equilibrium process[24,25]. Temperature dictates solution entropy contributions and the magnitude of protein-protein interactions (PPIs), which are important driving forces for the formation of LLPS. We investigated how temperature affects the LLPS of these three mAbs in specific buffers (Table 4.1) by constructing their liquid-liquid coexistence curves with the x and y axis representing protein concentration and temperature, respectively. Most proteins show a concave-down profile[26]. Above the coexistence curve, proteins are thermodynamically stable as a homogenous solution. Beneath this curve, protein solutions form two phases with distinct concentrations (specified by the liquid-liquid coexistence curve) under equilibrium conditions. LLPS was induced for each of the three mAbs at 4, 15, 22 and 25 °C, and the protein concentrations of the rich and poor phases were measured upon equilibrium. Figure 4.1 depicts how increasing temperature resulted in an increase and decrease in concentrations of the protein-poor and rich phase, respectively.

Out of these three proteins in their respective buffers, mAb Z and I had low T_c while mAb H had the highest. Full parameters of the fit to mean field theory equations are provided in Table 4.2. The three antibodies studied here have similar critical volume concentrations near 125 mg/ml. The small critical volume fraction of antibodies is presumably a consequence of network formation at high concentrations[9,27]. Antibodies are homodimers; if any region of the Fab (antigen binding fragment) is involved in protein self-association there are two binding sites per protein. Since an antibody molecule has two binding sites, if two antibodies bind in the Fab region then the resulting dimer also has two unoccupied binding sites. This association behavior permits formation of infinitely long chains of antibody molecules.

4.3.3 Structural Comparisons of the Three mAb Proteins in the Two Phases

Next, we address whether three different mAb proteins in these two phases share identical structural properties. The overall secondary and tertiary structures of mAb proteins in the two phases were analyzed and compared using multiple biophysical techniques. Since these samples are highly concentrated (> 200 mg/mL for protein-rich phases), the use of several commonly employed spectroscopic techniques is limited. Techniques specifically suitable to examine structural properties of these highly concentrated samples were therefore used.

4.3.4 Raman spectroscopy

Raman spectroscopy was able to detect differences in protein secondary and tertiary structure between the protein-rich and protein-poor phases. Overall, however, the normalized protein spectra of the top and bottom phases were highly similar (Figure 4.S1). The largest dissimilarities were observed in the spectral region of disulfide bands (Figure 4.2). IgG1 and IgG4 antibodies both have twelve intrachain and four interchain disulfide bonds[28]. The SS stretching vibrations of disulfides in a protein provide specific information about the geometry of the dihedral angles. A disulfide bond has two CS bonds adjacent to the SS bond. The CSSC dihedral angle is typically gauche (G), with angles near 90° or −90°. The CS bonds adjacent to the SS bond usually take a gauche (G) or trans (T) conformation. Raman spectra can distinguish between GGG, GGT, and TGT conformations of the CS, SS, and SC bonds in disulfides. GGG conformers have a vibrational stretching frequency near 510 cm^{-1}, and the trans conformations in CS bonds adds roughly 15 cm^{-1} of bond strain. Therefore, GGT conformations have a stretching frequency of approximately 525 cm^{-1} while TGT have a stretching frequency near 540 cm^{-1}. These three peaks are present in the spectra shown in Figure 4.2 for all three antibodies.

We then compared the relative intensities of the TGT and GGG bands using the ratio of their intensities ($I_{540/510}$). For the protein-rich phase of mAbs I and H, the ratio of GGG to TGT conformers is half that of the protein-poor phase (Table 4.3). For mAb Z, the GGG peak is nearly absent in the protein-poor phase resulting in a $I_{540/510}$ ratio of 7.4. The changes in disulfide structures may be isotype specific since interchain disulfides in the Fab domain are distinct in IgG1 and IgG4. A peak at 490 cm^{-1} is present in all three protein-rich phases but missing in the spectra of mAb H and I protein-poor phases. A disulfide stretching vibration of 490 cm^{-1} can originate from dihedral angles of about 25°[29]. The absence of this band in the protein-poor IgG1 phases again suggests a significantly altered state of disulfide bonds in the protein-rich phases of IgG1 molecules. It is expected that the bond energies would be much higher for disulfide bonds with SS angles at 25°. The hinge regions of antibodies have long been recognized for their intrinsic flexibility[30]. Disulfide bonds covalently stabilize the Ig domains of antibodies. The additional strain in these regions suggests that domain positional adjustments may be necessary to form protein networks in solutions of antibodies.

The tyrosine ratio of the Fermi doublet ($I_{855/830}$) was used to quantify the hydrogen bond donor and acceptor preference of tyrosine in the two phases[31]. For all three proteins the $I_{855/830}$

was lower in the protein rich phase (Table 4.3) signifying a shift to tyrosine as a stronger hydrogen bond donor. This effect was most pronounced for mAb H with a shift from 1.77 to 1.57. The amide I band λ_{max} at 1670 cm^{-1} indicates the antibodies are primarily beta sheet structure as expected. The amide I band moment exhibited a blue-shift of ~1 cm^{-1} for all three antibodies in the protein-rich phase compared to the protein-poor phase (Table 4.3). Band narrowing was also observed for the protein-rich phase as well; the FWHM of the amide I band decreased by 1 – 2 cm^{-1}. Band narrowing of the amide I band could be the result of restriction of the number of accessible conformational states in the protein-rich phase.

4.3.5 FTIR Water Band

Antibodies in the protein-rich phase are thought to form extensive interaction networks in an aqueous environment. The volume fraction of mAbs at 250 mg/mL is 17.8% when calculated with a specific volume of 0.71 mL/g for proteins[2]. We hypothesized such a large change in volume fraction could affect the structure and dynamics of solvent water. The bend + liberation (BL) FTIR band of water is a sensitive intrinsic probe of water dynamics and structure[32]. The position of the BL band can be used as a qualitative measure of the rigidity of water's hydrogen bonding network. For example, high temperature weakens water's hydrogen bonding and is accompanied by a red shift in the BL band position. A kosmotrope (a solute water structure maker) often causes a blue shift in the position of the BL band, suggesting a stronger water hydrogen bonding network. Conversely, a chaotrope (water structure breaker) induces a red shift. Protein-poor phases have equivalent band positions to their corresponding buffers (Figure 4.3). This suggests that water in low-ionic strength buffers and in the protein-poor phase has similar dynamic properties. A blue shift (~2 cm^{-1}) of the BL band, however, was observed for the protein-rich phases compared to the poor phases, suggesting water in the protein-rich phase is more structured. These measurements suggest that while water is only weakly involved in protein phase separation, it may be important in the development of more sophisticated theories for LLPS. Water molecules in the two phases have different dynamic properties but reach thermodynamic equilibrium. It may be interesting to study water dynamics or simulate water diffusion across the phase boundaries to better understand the LLPS of protein solutions.

4.3.6 FTIR Amide I region

The amide I FTIR band (between 1700 and 1600 cm^{-1}) primarily derives from the stretching of the amide carbonyl group and therefore reflects the backbone pattern of hydrogen bonding of a polypeptide. Various secondary structure types show distinct amide I bands (e.g. alpha helix, intramolecular beta sheet, intermolecular beta sheet, beta turn, and random coil). The amide I band of the protein-rich phase red shifted by ~2 cm^{-1} relative to that of the more depleted phase for all three mAbs, indicative of subtle structural differences (Figure 4.4 and Figure 4.S2). The reversibility of these changes was further tested by comparing FTIR spectra of the more dilute and rich phases at a concentration equivalent to its corresponding depleted phase. They were found to share identical amide I spectra (data not shown), indicating structural changes seen in the protein-rich phase are reversible upon dilution.

Protein-rich phases showed significantly more intermolecular β sheet (1614 cm^{-1}) than the protein-poor phases (Figure 4.S2). A red shift of the stronger main peak is accompanied by an increase in intermolecular β sheet. The main peak maximum corresponds to primarily beta sheet structure. The beta sheet rich structures of antibodies participate in extensive intramolecular hydrogen bonds. This red shift in the amide I band is observed for all three proteins. The decrease in energy of the amide I band could arise from alterations in the beta sheet structure such as the sheet twist angle. The red shift could also result from small specific changes in sidechain positions. It is also possible there are some stabilizing non-bonded interactions such as $n \rightarrow \pi^*$ interactions of carbonyl lone pairs and carbonyl π^* orbitals at the protein-protein interface which lower the energy of the amide I band[33]. While we are uncertain about the exact nature of the red shift, it is consistent across the three antibodies studied as also seen with the blue shift observed in tryptophan fluorescence (see below). The amide I red-shift in FTIR and blue-shift in Raman may be opposite in direction because of the different selection rules for the two methods (dipole moment and polarizability changes for FTIR and Raman, respectively).

FTIR peak position of spectra (Figure 4.4) was followed as a function of temperature (Figure 4.5). All three mAbs for both phases exhibited large red shifts upon aggregating to a peak at 1624 cm^{-1} (Figure 4.5). This peak has previously been assigned to amorphous aggregates of IgG1 antibodies following thermal unfolding and aggregation[34]. Spectra prior to the melting temperature do not exhibit significant change, indicating that the overall secondary structure differences between the proteins in the two phases are maintained until structural alterations and aggregation initiates.

4.3.7 Intrinsic Tryptophan Fluorescence

Tryptophan residues can be used as intrinsic fluorescent probes for protein structural changes since their fluorescence is sensitive to the polarity of their immediate environment. The protein-rich phases exhibited a blue-shifted λ_μ by 1 – 2 nm compared to their corresponding depleted phases at 25 °C, suggesting Trp residues are on average more buried in the protein-rich phase. Trp residues could either be buried because of tertiary structural changes or by their involvement in protein-protein interactions at interfaces. The blue-shifts seen are not artifacts due to inner filter effects since such an artifact usually results in a red-shifted emission[35]. The twelve indole sidechains in antibodies which participate in nonbonded interactions with intrachain disulfides are designated Trp triads[36]. Conformational changes within the Trp triad could be responsible for the observed blue shift in Trp fluorescence that are accompanied by the changes in the disulfide dihedral angles seen with Raman spectroscopy.

To further probe the spectral shift observed in the protein-rich phase, fluorescence was observed during thermal ramps from 25 to 90 °C (Figure 4.6A – C). The blue-shift of λ_μ for the protein-rich phase persisted prior to the thermal transitions for all three mAbs. Typical red shifts of λ_μ were observed upon melting, although the protein-rich phases (mAb Z, mAb H, and the second transition of mAb I) exhibited stronger thermal transitions than their corresponding protein-poor phases. The post-transition baselines of the two phases nearly overlapped for each mAb. This indicates a disruption of the underlying difference between the protein-rich and protein-poor phases upon thermal unfolding.

Compared to λ_μ, total fluorescence intensity (I_{tot}) correlates linearly with the degree of unfolding of a protein and therefore was used to generate T_m values (Table 4.S1)[18]. Similar to results in Figure 4.6, mAb Z showed only one transition, while the other two mAbs manifested two (Figure 4.S3A – C). Their first derivative plots also confirm the number of transitions (Supplemental Figure 4.S3D – F). First derivative plots of the protein-poor phase samples show thermal profiles consistent with DSC thermograms (Figure 4.6. Thermal stability of the proteins in the two liquid phases of the mAb samples measured by intrinsic tryptophan fluorescence spectroscopy. The mean fluorescence emission wavelength (λ_μ) of (A) mAb Z, (B) mAb H and (C) mAb I. See Table 4.1 for protein concentrations and solution conditions. (D) DSC thermograms for mAb Z, H and I (at 1 mg/ml).D), illustrating the ability of fluorescence melts to resolve multiple thermal transitions. Hence, first derivative plots of thermal transitions were used to delineate domain specific transitions of both protein-rich and -poor phases, which could not be obtained directly using DSC due to a clogging issue at high temperature by gelled highly concentrated mAbs. The protein-rich phase of mAb Z had a lower T_m than its protein-poor phase by ~1.6 °C. The same trend was observed for mAb I, whose protein-rich phase is thermally less stable by ~ 3.9 °C (T_m1) and 1.2 °C (T_m2) than the protein-poor phase. Conversely, mAb H's enriched phase showed higher T_m values than its protein-poor phase. T_m1 of its protein-rich and protein-poor phases are 64.4 ± 0.1 and 61.1 ± 0.1 °C, respectively, although their T_m2 showed only a minor difference. The increase of T_m1 for mAb H may correspond the fact that the T_m1 of mAb H is below its critical temperature since mAb H likely has the strongest stabilizing PPIs of the three mAbs.

Differences seen in their thermal unfolding profiles can result from the relative thermal stability of each domain and/or the cooperativity of their thermal unfolding events. For instance, mAb I showed a small early transition and a large second one. The first transition is often attributed to the unfolding of the CH_2 domain because of its relatively low stability[37], and the second one probably reflects a combination of the Fab and the CH_3 domain. In contrast, the magnitude of mAb H's first transition is comparable to the second one. This may suggest that mAb H has a relatively less-stable Fab, which contributes to the first transition alone or in combination with the CH_2 domain. Since mAb H's T_m1 increases in the protein-rich phase, this indicates that PPIs in protein-rich phase likely stabilize the mAb H's Fab (supported by HDX data below). MAb Z showed one broad transition consisting of several overlapping peaks, indicating that domains of mAb Z thermally unfolds in a more cooperative manner. The first thermal transition of mAb I's two phases shared a similar magnitude (Figure 4.6C). As mentioned above, such a transition may represent the thermal unfolding of the CH_2 domain. This indicates that thermal unfolding of the CH_2 domain in mAb I did not significantly disrupt PPIs in its protein-rich phase. Therefore the CH_2 domain is probably not part of the PPI interface for mAb I (supported by HDX data below).

4.3.8 Static Light Scattering

Static light scattering (SLS) was employed to monitor the aggregation of the two phases upon heating. Dynamic light scattering could not be used because of multiple scattering in high protein concentration solutions. SLS was measured with 450 nm light to avoid any absorption of lamp light by highly concentrated chromophores (Phe, Tyr, and Trp residues). As expected, protein-rich phases had higher SLS signals than corresponding protein-poor phases at 25 °C (Figure 4.7). Increasing temperature resulted in different trends in the pre-transition baselines of these three mAbs. For instance, both phases of mAb Z showed declining linear-like baselines, with the protein-rich phase decaying slightly faster. The protein-rich phase of mAb I also showed a decreasing baseline prior to its thermal transition. Its protein-poor phase, however, had a nearly flat baseline. In contrast, both mAb H's two phases manifested flat pre-transition baselines, suggesting the interaction networks formed by mAb H are not as strongly affected by temperature. The differences seen in pre-transition baselines of the three mAbs can be attributable to the temperature dependence of their PPIs. MAb Z seems to show the highest temperature dependence and mAb H the lowest. This trend agrees with the temperature dependent phase behaviors illustrated in their coexistence curves (Figure 4.1).

Protein-poor phases showed greater increases in their SLS signals during thermal transitions than the protein-rich phase for the three mAbs. Gelation upon thermal unfolding in the protein-rich phase could be responsible for reduced transmittance and the smaller changes in SLS signals during thermal transitions. We also noticed that the protein-rich phase of mAb I had an early transition starting at ~60 °C, which was not seen in its depleted phase. We attribute this transition to the unfolding induced aggregation of the CH_2 domain, based on the DSC and fluorescence data discussed above. It suggests that the thermal unfolding of the CH_2 domain of mAb I initiates aggregation in the protein-rich phase but not in the protein-poor phase, probably because of the much higher protein concentration in the protein-rich phase. This early transition was, however, not observed for mAbs Z and H.

4.3.9 Interface Mapping with HX-MS Analysis

For mAbs H and I, we performed HX-MS measurements to identify specific locations involved in protein-protein interactions. HX-MS was recorded as a function of exchange time (120 s, 10^3 s, 10^4 s, 10^5 s). Differences in HX-MS between protein at 60 mg/ml and protein at 5 mg/ml, ΔHX, were determined (Figure 4.8). $\Delta HX < 0$ (i.e., increased protection) therefore represents regions in which the antibody backbone at 60 mg/ml is less flexible than the protein at 5 mg/ml. Sequences which exhibited substantial changes in protection are summarized in Table 4.4.

For mAb H, there were protected regions in the CDR L2 and the heavy chain framework region. Both antibodies exhibited significant protection in the variable regions. The protected regions of mAb H contain a relatively disproportionate number of charged amino acid residues (Table 4.4). The two protected sequences for mAb H contained amino acid residues of opposite charges, suggesting at least a partially electrostatically driven self-association. In a homology model, the positive and negatively charged patches reside on opposite faces of the variable domain as shown in a homology model (Figure 4.9) indicating that Fab-Fab domain interactions probably mediate reversible-self association in this case.

For mAb I, HX-MS analysis revealed protection in the V_H region (at 120 s and 10^3 s) and in the V_L and CH_1 domains at 10^4 s. The sequences involved in the variable heavy chain are relatively hydrophobic and contain a high number or Trp and Tyr residues (Table 4.4). The isolated sequences have from 20 to 70 % aromatic residues, much greater than that found overall in the antibody itself which has only 8 % aromatic residues.

Subtle differences in solution conditions may be important for the interpretation of the HX-MS results since HX-MS was performed in slightly higher ionic strength conditions than those for which proteins phase separate. A recent study performed HX-MS directly on phase separated antibody solutions and found that nearly half of the sequences in the peptide map exhibited some significant degree of protection[38]. The concentration differences from 60 mg/ml at which we performed HX-MS and the >200 mg/ml solutions may also result in increases in protection at more sites because of the increased protein volume fraction. Two previous studies from our group involving different antibodies have identified both Fab-Fab and Fab-Fc interfaces responsible for protein reversible self-association[19,39]. Here the interactions driving protein self-association with these two mAbs were primarily identified in the variable regions.

4.3.10 Size Exclusion Chromatography

Size exclusion (SEC) chromatograms of the diluted protein phases displayed a preference for more dimer in the protein-rich phase and fragments in the protein-poor phase. There was 2.2, 2.1 and 2.0 percent more dimer in the protein-rich phase than the protein-poor one for mAbs Z, H and I, respectively (Table 4.5). Fragments eluting at ~11.5 mL were observed for all three mAbs with the highest difference seen in mAb H, where 1 % more fragments were observed in the protein-poor phase. The chromatograms for mAb H and mAb Z also exhibited small shoulders on either side of the main peak (Figure 4.10). We hypothesize that the protein-rich phase contains more aggregates because dimers can participate in the dynamic networks that drive phase separation while fragments only have one binding site and therefore could only exist at the edges of a network.

4.3.11 Aggregation Rates

The two phases were separated and injected on to the column without any dilution to measure the aggregation rates for each phase at 40 °C. Since the protein-rich and protein-poor phases are in equilibrium, the two phases have identical chemical potentials. Aggregation kinetics are a function of the chemical potential of protein in solution, therefore we hypothesized that protein aggregation rates would be equivalent in both phases.

One caveat is the difference observed in HMWS when directly injecting the protein-rich phase (4.5 % HMWS for mAb I at time zero) in comparison to diluting the protein-rich phase prior to injection (2.7 % HMWS for mAb I at time zero). This difference (4.5% without dilution to 2.7% following dilution) reflects that some fraction of the aggregates observed by SEC without dilution of the protein-rich phase are reversible. This was observed for all three immunoglobulins (most drastically for mAb H). Therefore the fraction of reversible and non-reversible aggregates at each time point could not be determined by SEC. Our analysis is simplified by interpreting HMWS as the sum of reversible and non-reversible aggregates.

All three antibodies exhibited higher or equivalent aggregation rates in the protein-poor compared to the protein-rich phase (Figure 4.11). MAb Z's protein-rich phase remained approximately constant in the amount of high molecular weight species (HMWS) for the two-week period (Table 4.6). However, mAb Z's protein-poor phase had an aggregation rate of 0.84 %/wk. MAb H had two small decreases in amount of HMWS. A slower equilibration process in the protein-rich network could alter the fraction of HMWS at these time scales. The protein-poor phase of mAb H aggregated faster than the rich-phase at a rate of 0.58 %/wk. MAb I aggregated at equivalent rates in the protein-rich (0.82 %/wk) and protein-poor phases (0.78 %/wk). Typically aggregation rates increase with protein concentration since thermodynamic activity increases with concentration, however all three mAbs in this study showed slower or equivalent rates in the protein-rich phase compared to the protein-poor phase.

The clipping of antibody molecules happened more rapidly in the protein-rich phase than the protein-poor phase. Especially for mAb Z, whose protein-rich phase had a fragmentation rate of 5 %/wk and protein-poor phase a rate of 0.3 %/wk. A major degradation pathway of antibodies at pH > 6 is fragmentation caused by base catalyzed beta elimination of Xaa-Cys residues in the hinge region[40]. The hydrogen abstraction required for this mechanism may be more energetically favorable for the GGG conformation only present in the protein-rich phase of mAb Z and explain the major discrepancy in the observed rates. For mAb H, the clipping rate was only slightly faster while for mAb I they were nearly equivalent. The differences in disulfide structures in IgG4 (mAb Z) and IgG1 (mAb H and I) may be important for the differences in the observed clipping rate constants. This explanation is only speculative, however, and differences could also be the result of partitioning of some protease impurities into the protein-rich phase of mAb Z.

4.4 Discussion

Phase separation of protein solutions has become a well-recognized phenomena. Studies of proteins such as lens crystallins[2,15,41], monoclonal immunoglobulins[3,42] and a few other proteins have described the event in some detail in terms of temperature, pH and salt effects[7]. The structural state of proteins in each phase, however, has seen few studies. It might at first be assumed the state of the protein in each phase would be the same since they are present at thermodynamic equilibrium. But in a protein/water/buffer or salt system simple application of the phase rule is not possible. For example, proteins themselves exist in an equilibrium of many microstates which reflect the highly dynamic nature of protein structure. Thus, it is not a surprise that characterization of these three immunoglobulins suggest differences in the structure and behavior of the proteins in each phase. Measurements of the water bend-liberation band suggests an increases in order of the hydrogen-bonding network of a portion of the water in the more concentrated protein phase. This is also expected since the increase in protein concentration will be accompanied by more water of hydration. The magnitude of this effect, however, seems to be too small to account for the observed changes in peak shift since no changes in band position were observed for water in the protein-lean phase. Self-association of the protein in the more concentrated phase would be expected to reduce hydrating water due to reduction in immunoglobulin surface area. Entrapment of water in aggregates offers another possibility.

Is it possible that the apparent conformational changes seen are an artifact of any of the measurement processes? While we cannot entirely exclude this possibility, we think this is unlikely. Spectral artifacts such as inner filter effects in fluorescence, surface adsorption (minimal contribution at the high concentrations employed) and electrostatic effects due to unequal distribution of buffering ions all appear to be negligible. We think two other explanations are more likely. Shifts in the distribution of protein conformational states due to water activity and protein concentration could produce the effects seen and are consistent with conformational changes observed in phase separation[43]. More likely, however, is that the micro-aggregation states that appear to exist in concentrated immunoglobulin solutions[8] could lead to structural alterations. The change seen in vibrational spectra and intrinsic fluorescence are consistent with this idea. The change seen in disulfide properties (dihedral angles) are consistent with the idea that changes in the dynamic nature of immunoglobulin may be involved in the spectral changes seen. Isotope exchange identifies regions for two proteins that form contact points for self-association (we were unable to perform such studies for the third protein).

Theoretical models of protein aggregation predict increases in aggregation rates with increased protein concentration (activity) since the reaction is at least bimolecular[44,45]. Aggregation rates were slowed or equivalent for more highly concentrated solutions of all three phase separated antibodies herein. This atypical observation results from the equivalent chemical potential of the protein-rich and protein-poor phases[46]. However, if solely chemical potential dictated aggregation rates one would anticipate equivalent aggregation rates in both phases for all three antibodies. However, only for mAb I were aggregation rates equivalent in both phases while for mAb H and mAb Z the aggregation rate was slower in the protein-rich phase. If aggregation is diffusion rate-limited the higher viscosity protein-rich phase could slow aggregation. An alternate explanation for why the aggregation rates are slower in the protein-rich phase is that to form an irreversible aggregate two monomers must diffuse close together to form an encounter complex which needs to rearrange to form a stable aggregate nucleus. If the protein's native state is stabilized by self-interaction or the proteins are constrained by crowding, the kinetics of forming a stable aggregate nucleus would be expected to be slower[12].

Phase separation provides a unique opportunity to study aggregation rates at different protein concentrations but equivalent chemical potentials. Phase separation is known to promote aggregation in cells for some proteins[14]. We suspect however that in some cases the opposite may be true; the protein-rich droplets may serve to slow protein aggregation processes by slowing diffusion and/or stabilizing a proteins native state. Conformational changes of proteins in the protein-rich phase may alter other cellular reaction kinetics similar to the increased rate of clipping for mAb Z.

4.5 Conclusion

We have observed several consistent conformational changes in all three mAbs when comparing their protein-rich and protein-poor solutions of phase-separated antibodies. Tryptophan emission and vibrational spectra suggest conformational changes of proteins in dynamic clusters. Disulfide bond isomerized to higher energy conformations further confirmed changes in the dynamic nature for all three immunoglobulins in the protein-rich phase. These changes persisted in temperatures well above the coexistence curve but not above the T_m indicating that thermal unfolding disrupts the interactions governing phase separation. The hydrogen bond network of water was significantly more ordered in the protein-rich phase indicating a possible role for water in the dynamic clusters. Our measurements also highlight the utility of techniques which can characterize protein phases directly without the need for dilution. HX-MS identified specific sequences of the protein involved in the protein-protein interfaces and indicated a role for non-covalent ionic and hydrophobic interactions on the Fab for mAb H and I, respectively, leading to reversible protein-protein interactions in a concentration dependent manner. Protein aggregation was slowed in the protein-rich phase relative to the protein-poor phase for mAb's H and Z while aggregation rates were equivalent for mAb I.

4.6 References

1. Hyman, A. A.; Weber, C. A.; Jülicher, F. Liquid-Liquid Phase Separation in Biology. *Annu. Rev. Cell Dev. Biol.* **2014**, *30* (1), 39–58.

2. Schurtenberger, P.; Chamberlin, R. A.; Thurston, G. M.; Thomson, J. A.; Benedek, G. B. Observation of Critical Phenomena in a Protein-Water Solution. *Phys. Rev. Lett.* **1989**, *63* (19), 2064–2067.

3. Wang, Y.; Lomakin, A.; Latypov, R. F.; Laubach, J. P.; Hideshima, T.; Richardson, P. G.; Munshi, N. C.; Anderson, K. C.; Benedek, G. B. Phase Transitions in Human IgG Solutions. *J. Chem. Phys.* **2013**, *139* (12), 1–9.

4. Brangwynne, C. P.; Tompa, P.; Pappu, R. V. Polymer Physics of Intracellular Phase Transitions. *Nat. Phys.* **2015**, *11* (11), 899–904.

5. Dijkstra, M.; van Roij, R.; Evans, R. Phase Diagram of Highly Asymmetric Binary Hard-Sphere Mixtures. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.* **1999**, *59* (5), 5744–5771.

6. von Bülow, S.; Siggel, M.; Linke, M.; Hummer, G. Dynamic Cluster Formation Determines Viscosity and Diffusion in Dense Protein Solutions. *Proc. Nat. Acad. Sci. USA* **2019**, *In Review*.

7. Kastelic, M.; Kalyuzhnyi, Y. V.; Hribar-Lee, B.; Dill, K. A.; Vlachy, V. Protein Aggregation in Salt Solutions. *Proc. Natl. Acad. Sci.* **2015**, *112* (21), 6766–6770.

8. Godfrin, P. D.; Zarraga, I. E.; Zarzar, J.; Porcar, L.; Falus, P.; Wagner, N. J.; Liu, Y. Effect of Hierarchical Cluster Formation on the Viscosity of Concentrated Monoclonal Antibody Formulations Studied by Neutron Scattering. *J. Phys. Chem. B* **2016**, *120* (2), 278–291.

9. Yearley, E. J.; Godfrin, P. D.; Perevozchikova, T.; Zhang, H.; Falus, P.; Porcar, L.; Nagao, M.; Curtis, J. E.; Gawande, P.; Taing, R.; Zarraga, I. E.; Wagner, N. J.; Liu, Y. Observation of Small Cluster Formation in Concentrated Monoclonal Antibody Solutions and Its Implications to Solution Viscosity. *Biophys. J.* **2014**, *106* (8), 1763–1770.

10. Porcar, L.; Falus, P.; Chen, W. R.; Faraone, A.; Fratini, E.; Hong, K.; Baglioni, P.; Liu, Y. Formation of the Dynamic Clusters in Concentrated Lysozyme Protein Solutions. *J. Phys. Chem. Lett.* **2010**, *1* (1), 126–129.

11. Kastelic, M.; Dill, K. A.; Kalyuzhnyi, Y. V.; Vlachy, V. Controlling the Viscosities of Antibody Solutions through Control of Their Binding Sites. *J. Mol. Liq.* **2018**, *270*, 234–242.

12. Roberts, C. J. Therapeutic Protein Aggregation: Mechanisms, Design, and Control. *Trends Biotechnol.* **2014**, *32* (7), 372–380.

13. Whitaker, N.; Xiong, J.; Pace, S. E.; Kumar, V.; Middaugh, C. R.; Joshi, S. B.; Volkin, D. B. A Formulation Development Approach to Identify and Select Stable Ultra–High-Concentration Monoclonal Antibody Formulations With Reduced Viscosities. *J. Pharm. Sci.* **2017**, *106* (11), 3230–3241.

14. Wegmann, S.; Eftekharzadeh, B.; Tepper, K.; Zoltowska, K. M.; Bennett, R. E.; Dujardin, S.; Laskowski, P. R.; MacKenzie, D.; Kamath, T.; Commins, C.; Vanderburg, C.; Roe, A. D.; Fan, Z.; Molliex, A. M.; Hernandez-Vega, A.; Muller, D.; Hyman, A. A.; Mandelkow, E.; Taylor, J. P.; Hyman, B. T. Tau Protein Liquid–Liquid Phase Separation Can Initiate Tau Aggregation. *EMBO J.* **2018**, *37* (7), e98049.

15. Broide, M. L.; Berland, C. R.; Pande, J.; Ogun, O. O.; Benedek, G. B. Binary-Liquid Phase Separation of Lens Protein Solutions. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88* (13), 5660–5664.

16. Newville, Matthew, Stensitzki, Till, Allen, Daniel B., & Ingargiola, A. LMFIT: Non-Linear Least-Square Minimization and Curve-Fitting for Python.

17. Wei, Y.; Xiong, J.; Larson, N. R.; Iyer, V.; Sanyal, G.; Joshi, S. B.; Volkin, D. B.; Middaugh, C. R. Effect of 2 Emulsion-Based Adjuvants on the Structure and Thermal Stability of Staphylococcus Aureus Alpha-Toxin. *J. Pharm. Sci.* **2018**, *107* (9), 2325–2334.

18. Wei, Y.; Larson, N. R.; Angalakurthi, S. K.; Russell Middaugh, C. Improved Fluorescence Methods for High-Throughput Protein Formulation Screening. *SLAS Technol. Transl. Life Sci. Innov.* **2018**, *23* (6), 516–528.

19. Arora, J.; Hickey, J. M.; Majumdar, R.; Esfandiary, R.; Bishop, S. M.; Samra, H. S.; Middaugh, C. R.; Weis, D. D.; Volkin, D. B. Hydrogen Exchange Mass Spectrometry Reveals Protein Interfaces and Distant Dynamic Coupling Effects during the Reversible Self-Association of an IgG1 Monoclonal Antibody. *MAbs* **2015**, *7* (3), 525–539.

20. Hu, Y.; Arora, J.; Joshi, S. B.; Esfandiary, R.; Middaugh, C. R.; Weis, D. D.; Volkin, D. B. Characterization of Excipient Effects on Reversible Self-Association, Backbone Flexibility, and Solution Properties of an IgG1 Monoclonal Antibody at High Concentrations: Part 1. *J. Pharm. Sci.* **2020**, *109* (1), 340–352.

21. Hu, Y.; Toth, R. T.; Joshi, S. B.; Esfandiary, R.; Middaugh, C. R.; Volkin, D. B.; Weis, D. D. Characterization of Excipient Effects on Reversible Self-Association, Backbone Flexibility, and Solution Properties of an IgG1 Monoclonal Antibody at High Concentrations: Part 2. *J. Pharm. Sci.* **2020**, *109* (1), 353–363.

22. Chemical Computing Group Inc. Molecular Operating Enviornment. Montreal, QC, Canada 2017.

23. Benke, M.; Fischer, N.; Randolph, T. W.; Kendrick, B. S.; Buckel, S.; Stoner, M. R.; Austin, F.; Nixon, L. Protein−solute Interactions Affect the Outcome of Ultrafiltration/Diafiltration Operations. *J. Pharm. Sci.* **2004**, *93* (9), 2332–2342.

24. Mason, B. D.; Zhang-Van Enk, J.; Zhang, L.; Remmele, R. L.; Zhang, J. Liquid-Liquid Phase Separation of a Monoclonal Antibody and Nonmonotonic Influence of Hofmeister Anions. *Biophys. J.* **2010**, *99* (11), 3792–3800.

25. Wu, G.; Wang, S.; Tian, Z.; Zhang, N.; Sheng, H.; Dai, W.; Qian, F. Elucidating the Weak Protein-Protein Interaction Mechanisms behind the Liquid-Liquid Phase Separation of a MAb Solution by Different Types of Additives. *Eur. J. Pharm. Biopharm.* **2017**, *120*, 1–8.

26. Raut, A. S.; Kalonia, D. S. Liquid-Liquid Phase Separation in a Dual Variable Domain Immunoglobulin Protein Solution: Effect of Formulation Factors and Protein-Protein Interactions. *Mol. Pharm.* **2015**, *12* (9), 3261–3271.

27. Sun, G.; Wang, Y.; Lomakin, A.; Benedek, G. B.; Stanley, H. E.; Xu, L.; Buldyrev, S. V. The Phase Behavior Study of Human Antibody Solution Using Multi-Scale Modeling. *J. Chem. Phys.* **2016**, *145* (19).

28. Liu, H.; May, K. Disulfide Bond Structures of IgG Molecules. *MAbs* **2012**, *4* (1), 17–23.

29. Van Wart, H. E.; Lewis, A.; Scheraga, H. A.; Saeva, F. D. Disulfide Bond Dihedral Angles from Raman Spectroscopy. *Proc. Natl. Acad. Sci.* **1973**, *70* (9), 2619–2623.

30. Clark, N. J.; Zhang, H.; Krueger, S.; Lee, H. J.; Ketchem, R. R.; Kerwin, B.; Kanapuram, S. R.; Treuheit, M. J.; McAuley, A.; Curtis, J. E. Small-Angle Neutron Scattering Study of a Monoclonal Antibody Using Free-Energy Constraints. *J. Phys. Chem. B* **2013**, *117* (45), 14029–14038.

31. Wen, Z. Q. Raman Spectroscopy of Protein Pharmaceuticals. *J. Pharm. Sci.* **2007**, *96* (11), 2861–2878.

32. Verma, P. K.; Kundu, A.; Puretz, M. S.; Dhoonmoon, C.; Chegwidden, O. S.; Londergan, C. H.; Cho, M. The Bend+Libration Combination Band Is an Intrinsic, Collective, and Strongly Solute-Dependent Reporter on the Hydrogen Bonding Network of Liquid Water. *J. Phys. Chem. B* **2018**, *122* (9), 2587–2599.

33. Newberry, R. W.; Raines, R. T. The *n* →π* Interaction. *Acc. Chem. Res.* **2017**, *50* (8), 1838–1846.

34. Boulet-Audet, M.; Byrne, B.; Kazarian, S. G. High-Throughput Thermal Stability Analysis of a Monoclonal Antibody by Attenuated Total Reflection FT-IR Spectroscopic Imaging. *Anal. Chem.* **2014**, *86* (19), 9786–9793.

35. Lakowicz, J. R. *Principles of Fluorescence Spectroscopy; 3rd Edition; Springer, New York,*; 2006.

36. Ioerger, T. R.; Du, C.; Linthicum, D. S. Conservation of Cys-Cys Trp Structural Triads and Their Geometry in the Protein Domains of Immunoglobulin Superfamily Members. *Mol. Immunol.* **1999**, *36* (6), 373–386.

37. Ionescu, R. M.; Vlasak, J.; Price, C.; Kirchmeier, M. Contribution of Variable Domains to the Stability of Humanized IgG1 Monoclonal Antibodies. *J. Pharm. Sci.* **2008**, *97* (4), 1414–1426.

38. Tian, Y.; Huang, L.; Ruotolo, B. T.; Wang, N. Hydrogen/Deuterium Exchange-Mass Spectrometry Analysis of High Concentration Biotherapeutics: Application to Phase-Separated Antibody Formulations. *MAbs* **2019**, *0* (0), 1–10.

39. Arora, J.; Hu, Y.; Esfandiary, R.; Sathish, H. A.; Bishop, S. M.; Joshi, S. B.; Middaugh, C. R.; Volkin, D. B.; Weis, D. D. Charge-Mediated Fab-Fc Interactions in an IgG1 Antibody Induce Reversible Self-Association, Cluster Formation, and Elevated Viscosity. *MAbs* **2016**, *0862* (September), 0.

40. Vlasak, J.; Ionescu, R. Fragmentation of Monoclonal Antibodies. *MAbs* **2011**, *3* (3), 253–263.

41. Liu, C.; Asherie, N.; Lomakin, A.; Pande, J.; Ogun, O.; Benedek, G. B. Phase Separation in Aqueous Solutions of Lens Gamma-Crystallins: Special Role of Gamma S. *Proc. Natl. Acad. Sci.* **2002**, *93* (1), 377–382.

42. Kheddo, P.; Bramham, J. E.; Dearman, R. J.; Uddin, S.; Van Der Walle, C. F.; Golovanov, A. P. Investigating Liquid-Liquid Phase Separation of a Monoclonal Antibody Using Solution-State NMR Spectroscopy: Effect of Arg·Glu and Arg·HCl. *Mol. Pharm.* **2017**, *14* (8), 2852–2860.

43. Majumdar, A.; Dogra, P.; Maity, S.; Mukhopadhyay, S. Liquid-Liquid Phase Separation Is Driven by Large-Scale Conformational Unwinding and Fluctuations of Intrinsically Disordered Protein Molecules. *J. Phys. Chem. Lett.* **2019**, *10* (14), 3929–3936.

44. Li, Y.; Roberts, C. J. Lumry-Eyring Nucleated-Polymerization Model of Protein Aggregation Kinetics. 2. Competing Growth via Condensation and Chain Polymerization. *J. Phys. Chem. B* **2009**, *113* (19), 7020–7032.

45. Pellarin, R.; Caflisch, A. Interpreting the Aggregation Kinetics of Amyloid Peptides. *J. Mol. Biol.* **2006**, *360* (4), 882–892.

46. Ghosh, R.; Calero-Rubio, C.; Saluja, A.; Roberts, C. J. Relating Protein-Protein Interactions and Aggregation Rates from Low to High Concentrations. *J. Pharm. Sci.* **2016**, *105* (3), 1086–1096.

4.7 Figures and Tables

Table 4.1. The mAb concentrations and buffer conditions used throughout this study. Measured concentrations and pH for each phase.

Protein sample (isotype)	Dialysis buffer	Phase	Concentration (mg/mL)	pH
mAb Z (IgG 4)	20 mM histidine, pH 6.8	Protein-poor Protein-rich	35.0 ± 0.1 223.4 ± 1.3	6.80 ± 0.10 6.81 ± 0.02
mAb H (IgG 1)	20 mM sodium phosphate, pH 7.5	Protein-poor Protein-rich	8.9 ± 0.1 245.7 ± 0.4	7.51 ± 0.04 7.47 ± 0.02
mAb I (IgG 1)	5 mM sodium phosphate, pH 8.0	Protein-poor Protein-rich	19.9 ± 0.1 207.6 ± 0.6	8.00 ± 0.01 8.04 ± 0.01

Table 4.2. Critical concentration, temperature and coexistence width for antibody phase coexistence curves. Cc and Tc are the respective concentration and temperature at the critical point of the best fit coexistence curves. A parameterizes the width of the coexistence curve.

mAb (isotype)	C_c (mg/ml)	T_c (°C)	A
mAb Z (IgG 4)	133.8 ± 3.3	29.8 ± 1.9	2.14 ± 0.14
mAb H (IgG 1)	129.0 ± 1.6	85.9 ± 28.7	1.60 ± 0.16
mAb I (IgG 1)	118.9 ± 2.4	34.4 ± 2.8	2.16 ± 0.13

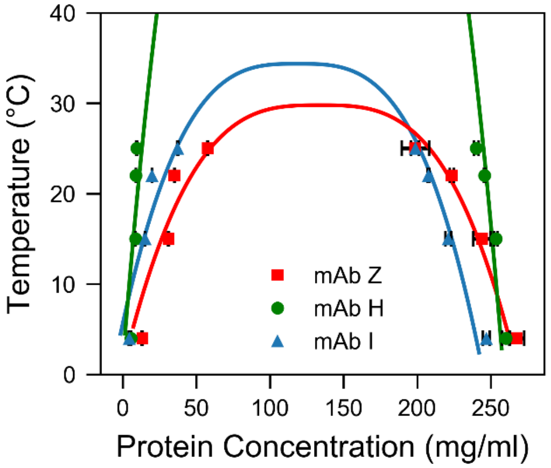


Figure 4.1. Liquid-liquid coexistence curves of mAbs Z, H, and I reflecting their temperature dependent LLPS behavior. Error bars represent the standard deviation of three experiments. Lines represent best fit to mean field theory. See Table 4.1 for protein concentrations and solution conditions.

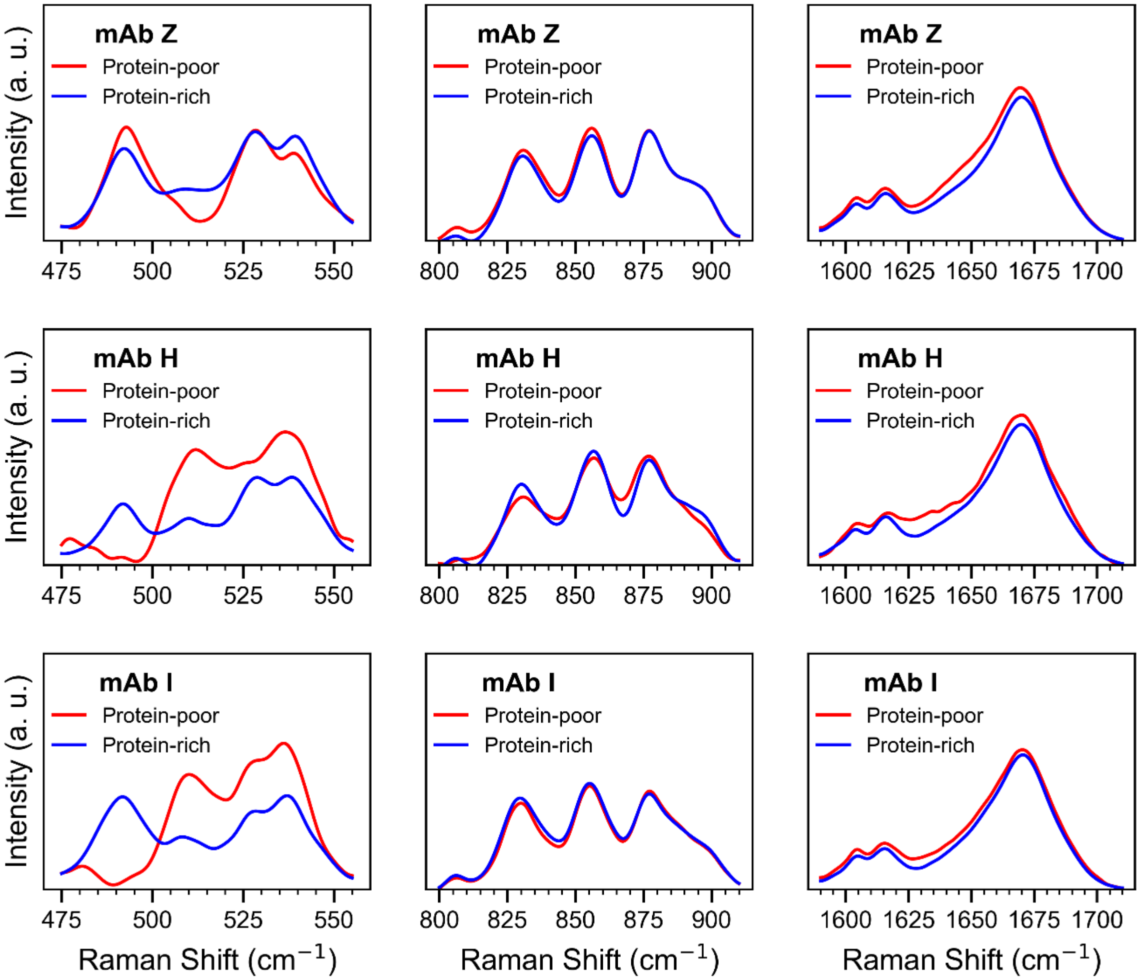


Figure 4.2. Raman spectra for mab Z (top panels), mab H (middle panels) and mab I (bottom panels). Left column displaying the spectral region associated with SS stretching vibrations, middle column – tyrosine Fermi doublet, and right column the amide I band. See Table 4.1 for protein concentrations and solution conditions.

Table 4.3. Summary of Raman data.

mAb	Phase	SS ratio $I_{540/510}$	Tyr ratio $I_{855/830}$	Amide I mean	Amide I FWHM
Z	Protein-poor	7.37 ± 0.70	1.42 ± 0.02	1663.3 ± 0.1	34.4 ± 0.1
Z	Protein-rich	2.41 ± 0.04	1.39 ± 0.02	1664.3 ± 0.1	33.7 ± 0.1
H	Protein-poor	1.15 ± 0.31	1.77 ± 0.03	1663.2 ± 0.5	35.3 ± 0.3
H	Protein-rich	2.12 ± 0.12	1.57 ± 0.01	1664.3 ± 0.1	33.7 ± 0.1
I	Protein-poor	1.11 ± 0.22	1.28 ± 0.02	1664.3 ± 0.1	33.8 ± 0.1
I	Protein-rich	2.03 ± 0.08	1.22 ± 0.02	1665.0 ± 0.1	32.6 ± 0.2

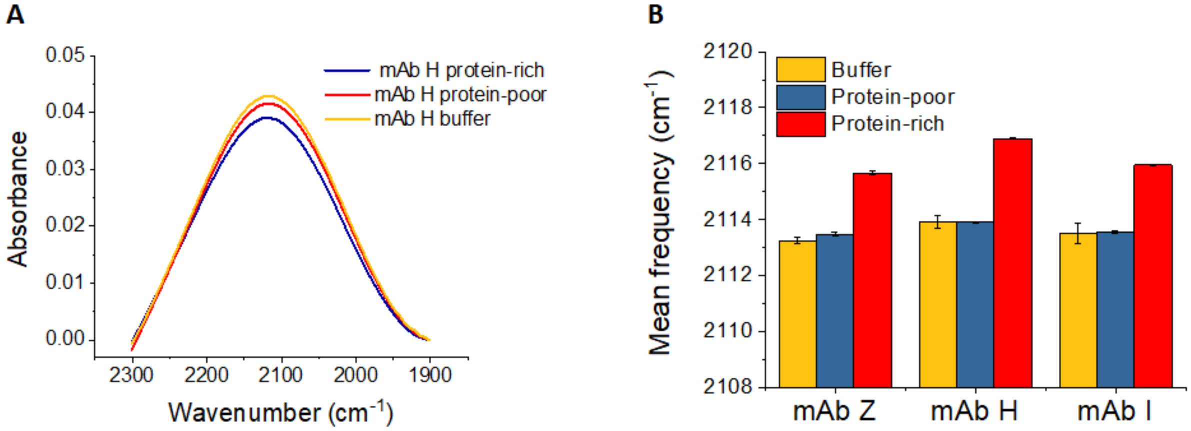


Figure 4.3. (A) FTIR bend-liberation combination bands of water in the two liquid phases of mAb samples and (B) their mean frequency between 1900 and 2300 wavenumbers cm^{-1}. See Table 4.1 for protein concentrations and solution conditions.

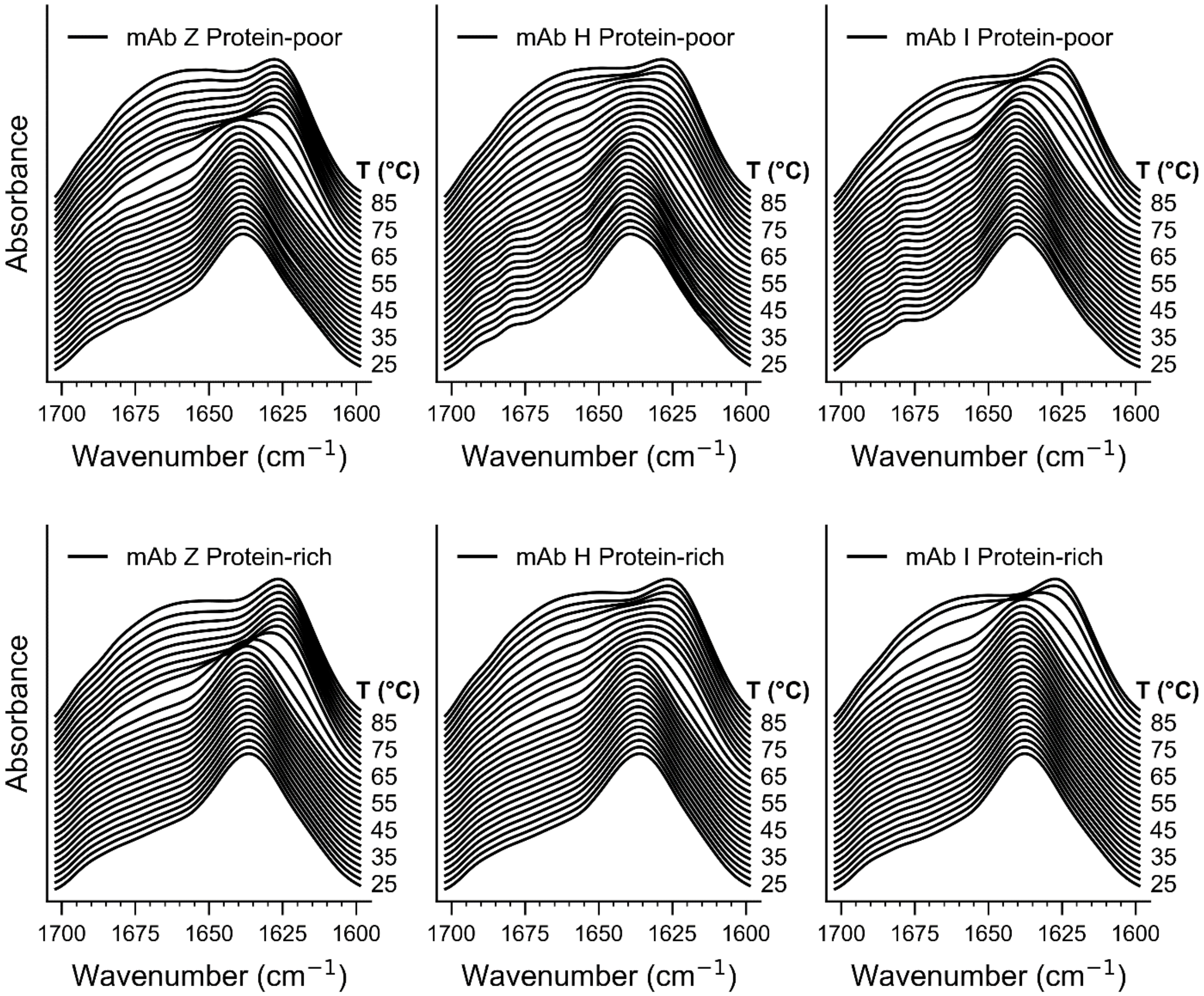


Figure 4.4. FTIR spectra vs temperature of the three antibodies for both the protein-poor and protein-rich phases. Vertically offset spectra represent increasing temperatures. See Table 4.1 for protein concentrations and solution conditions.

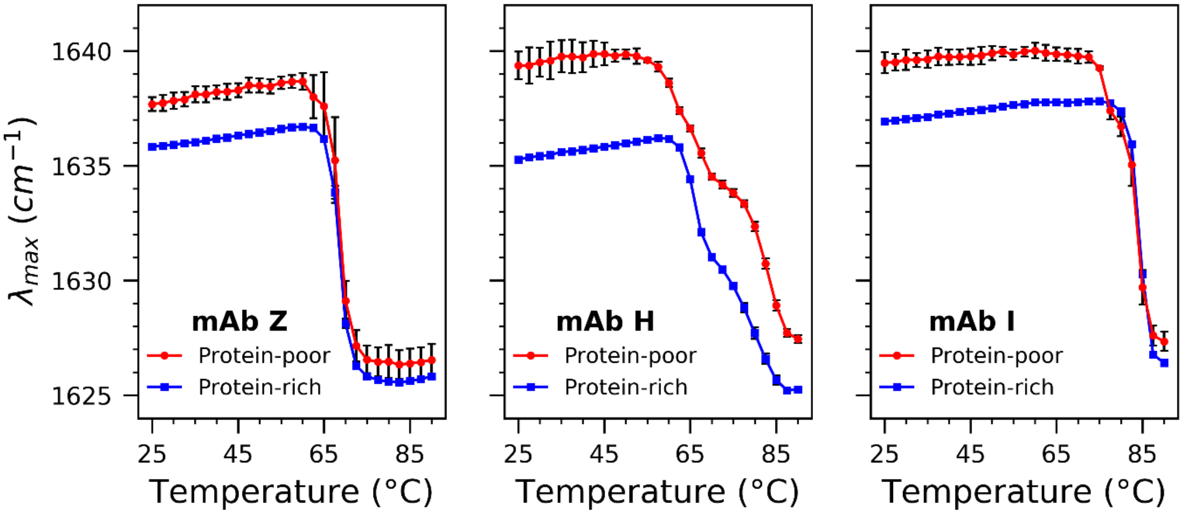


Figure 4.5. FTIR amide I peak position vs temperature for each of the antibodies in protein-poor and protein-rich phases. The error-bars represent the standard deviation of three independent measurements. See Table 4.1 for protein concentrations and solution conditions.

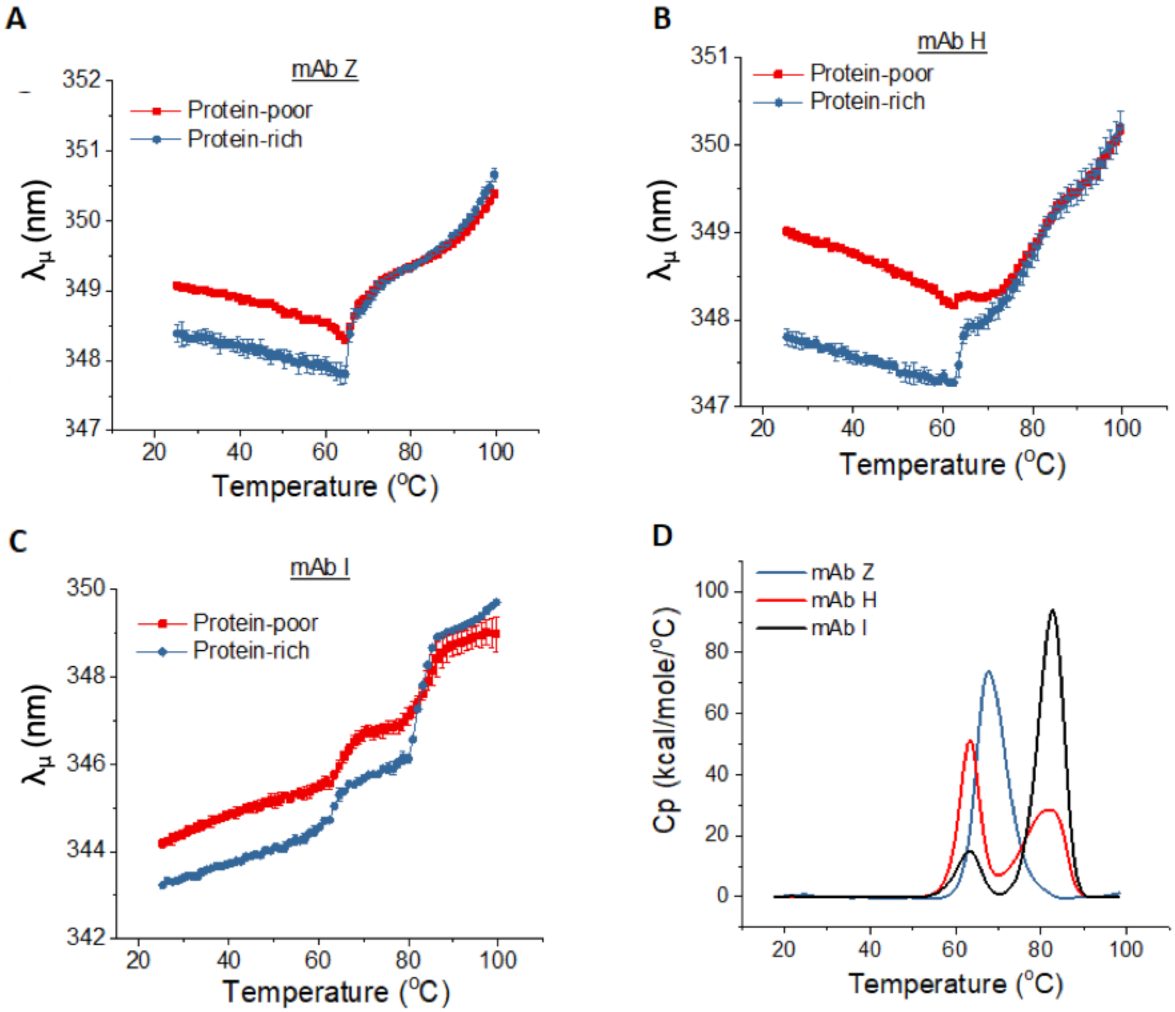


Figure 4.6. Thermal stability of the proteins in the two liquid phases of the mAb samples measured by intrinsic tryptophan fluorescence spectroscopy. The mean fluorescence emission wavelength (λ_μ) of (A) mAb Z, (B) mAb H and (C) mAb I. See Table 4.1 for protein concentrations and solution conditions. (D) DSC thermograms for mAb Z, H and I (at 1 mg/ml).

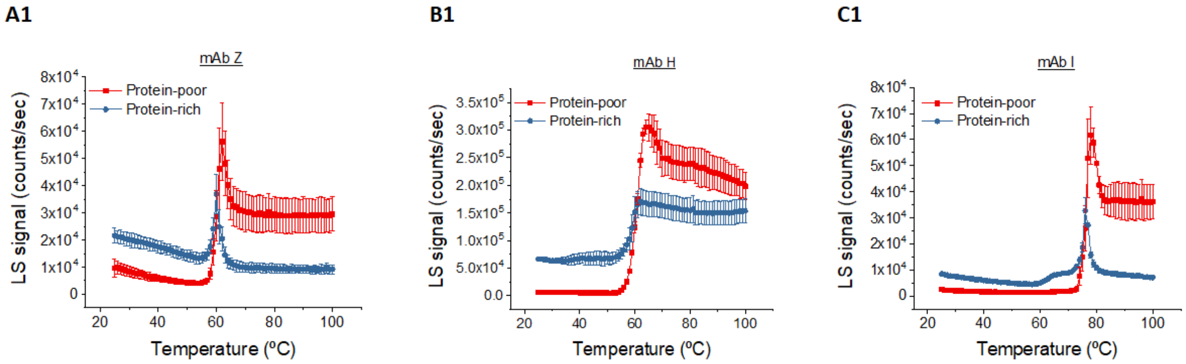


Figure 4.7. Static light scattering vs temperatureof the protein-rich and protein-poor phases of (A) mAb Z, (B) mAb H and (C) and mAb I. See Table 4.1 for protein concentrations and solution conditions.

Table 4.4. Sequences of self-associating interfaces for mAb H and mAb I as determined from HX-MS analysis (Figure 4.8).

mAb	Region	Sequence	% Positive	% Negative	% Aromatic
H	CDR L2	VLIVFDDEDRPSGIP	7	27	7
H	H chain FWR	KKPGASVKVSCKAS	29	0	0
H		Entire sequence	11	9	9
I	CDR H1	YYWSWI	0	0	67
I	CDR H2	SLYYSGSTYYNP	0	0	33
I	CDR L1	SQGISSWLAWYQQ	0	0	23
I		Entire sequence	11	9	8

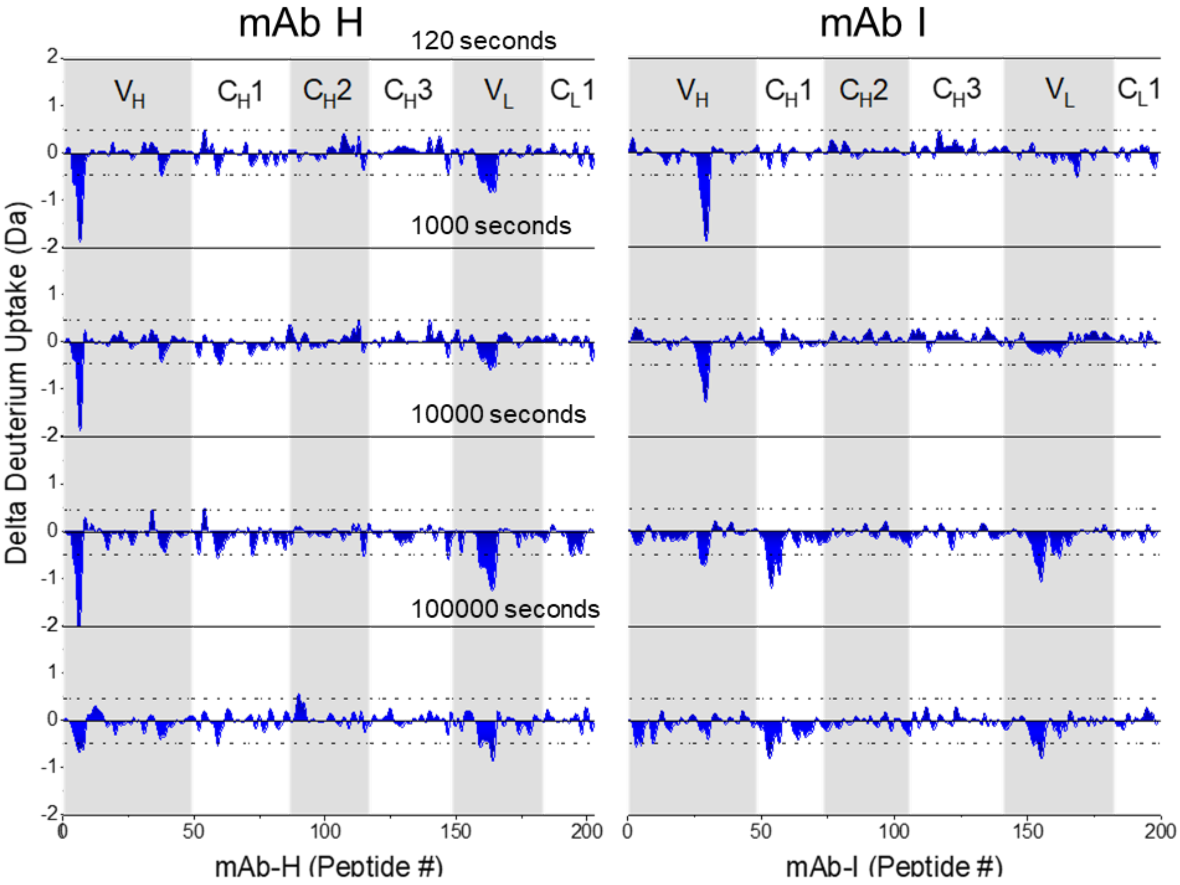


Figure 4.8. HX mass differences for mAb H and mAb I at four labeling times comparing high (60 mg/ml) and low (5 mg/ml) protein concentrations. The difference in the Y-axis displays the value of deuterium uptake at high protein concentration minus that at low protein concentration. Negative values indicate slower hydrogen exchange at high protein concentration. The dashed horizontal lines are the 99% confidence criterion. Three independent HX measurements was performed. Different domains are assigned based on peptide region and listed at the top of each plot.

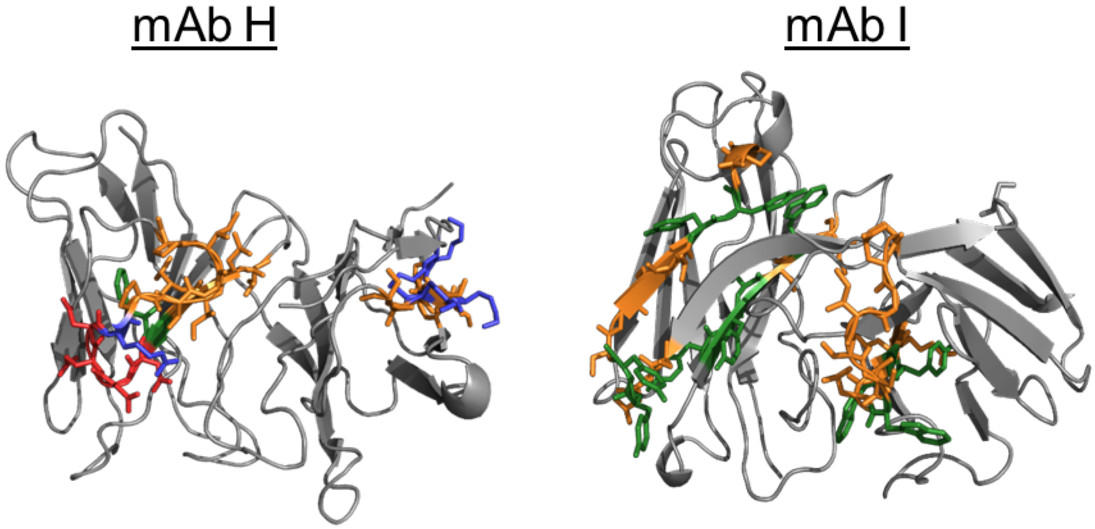


Figure 4.9. Homology models of the variable domains of mAb H (left) and mAb I (right). Residues regions that become substantially protected from HX at 60 mg/mL are shown with sidechains and in color. Regions with aromatic side-chains are shown in green, positively charged side-chains are shown in blue and negatively charged side-chains are shown in red, other protected regions are shown in orange.

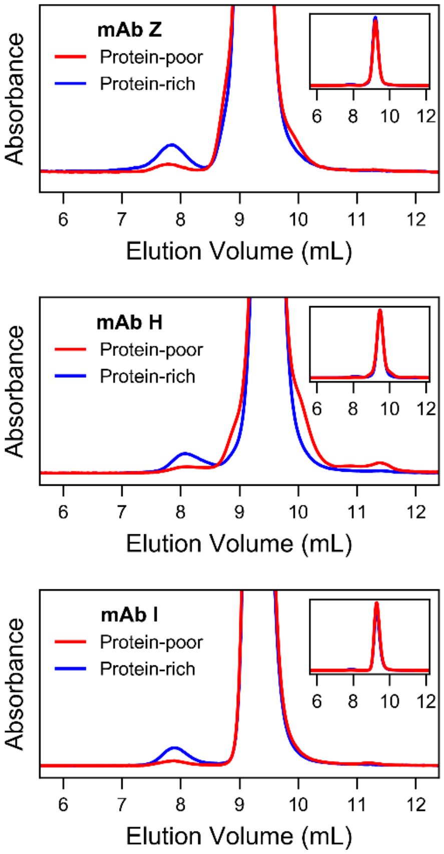


Figure 4.10. SEC chromatograms of the three mAbs in the poor and rich phases. Samples were diluted to 1 mg/ml before injection on the column. Absorbance was measured at 280 nm. The y-axis has been scaled to show the difference observed in the dimer and fragment peaks for each antibody. The inset of each panel shows the full scale of the chromatogram.

Table 4.5. Initial percent of species observed at 25 °C after phase separation by SEC. The standard deviation on these measurements is ± 0.1%.

mAb	Phase	% Monomer	% Dimer	% Fragments
Z	Protein-poor	99.4	0.5	0.1
Z	Protein-rich	97.6	2.3	0.1
H	Protein-poor	97.5	1.2	1.3
H	Protein-rich	96.3	3.3	0.4
I	Protein-poor	99.0	0.7	0.3
I	Protein-rich	97.1	2.7	0.2

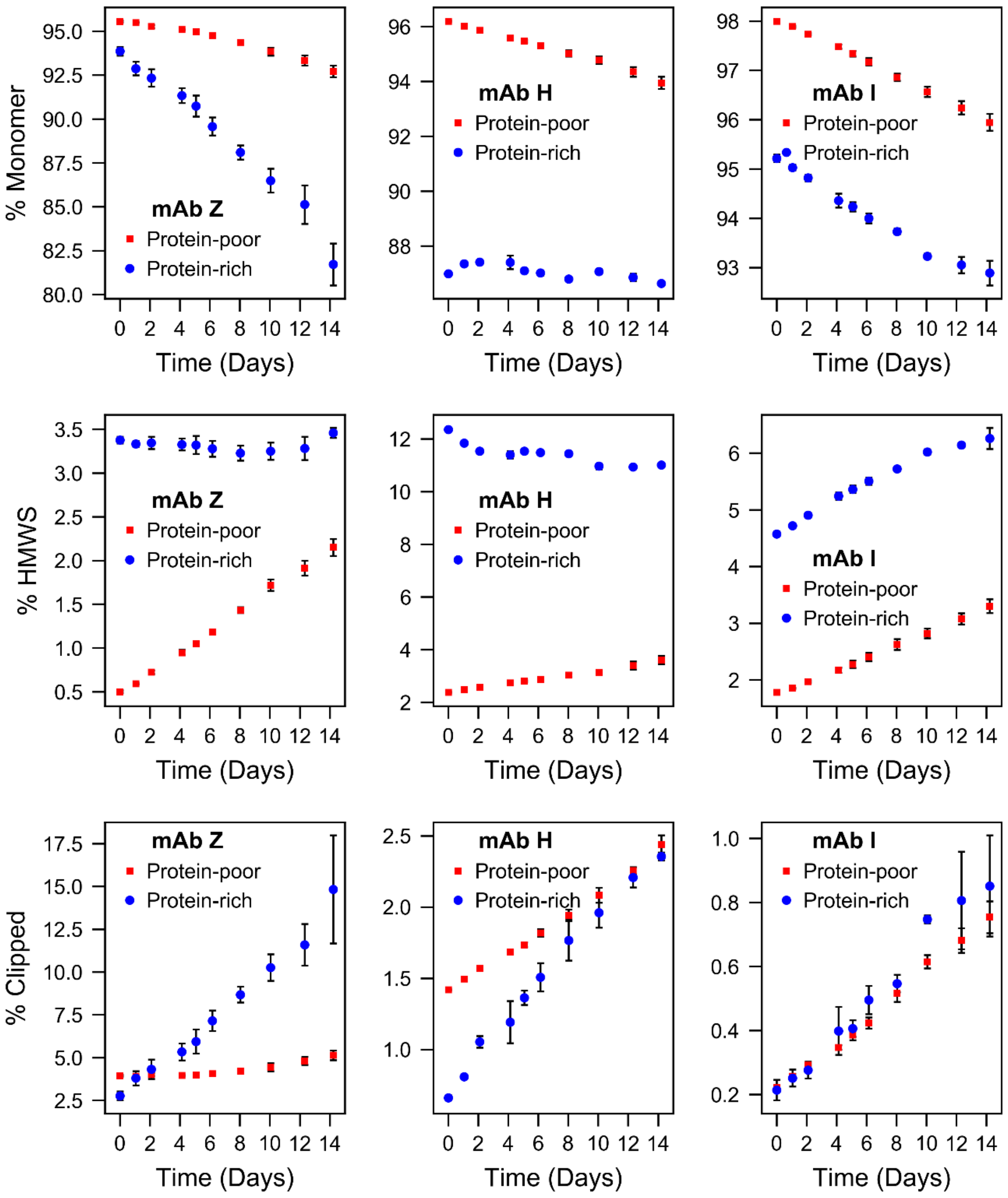


Figure 4.11. Size fractions of mAbs for the protein-poor and protein-rich phases during a two week accelerated stability study at 40 °C as measured by SEC.

Table 4.6. Rates of changes in % per week of species observed with SEC chromatography at 40 °C.

mAb	Phase	HMWS (%/ wk)	Monomer (%/wk)	Clips (%/wk)
Z	Protein-poor	0.84 ± 0.01	-1.16 ± 0.06	0.31 ± 0.06
Z	Protein-rich	-0.01 ± 0.04	-5.08 ± 0.28	5.09 ± 0.03
H	Protein-poor	0.58 ± 0.02	-1.06 ± 0.03	0.48 ± 0.01
H	Protein-rich	-0.55 ± 0.06	-0.28 ± 0.06	0.83 ± 0.03
I	Protein-poor	0.78 ± 0.02	-1.04 ± 0.03	0.27 ± 0.01
I	Protein-rich	0.82 ± 0.03	-1.25 ± 0.05	0.33 ± 0.02

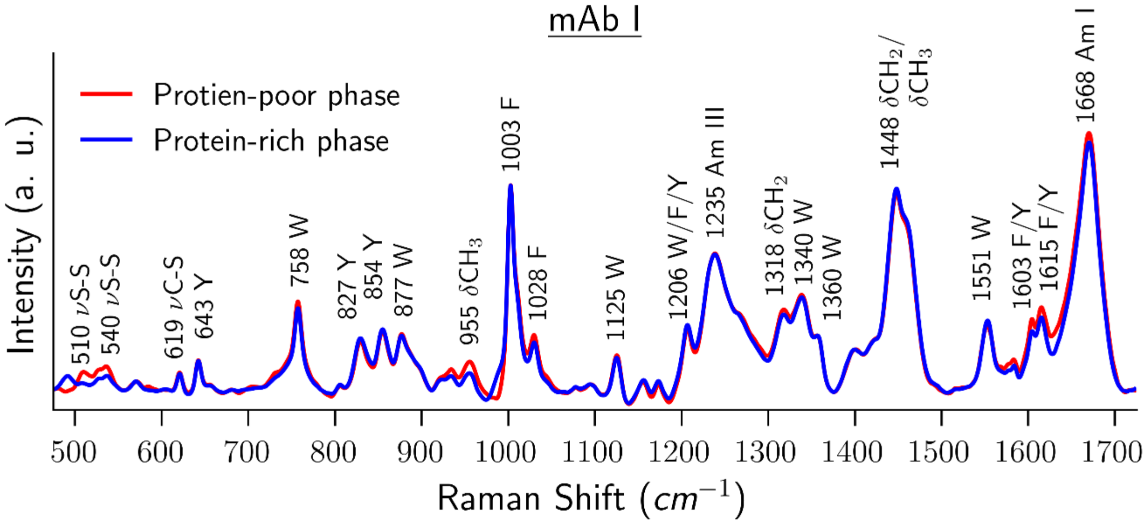


Figure 4.S1. Raman spectra for mAb I protein-poor and protein-rich phases. All major peaks are labeled with their position and assignment. See Table 4.1 for protein concentrations and solution conditions.

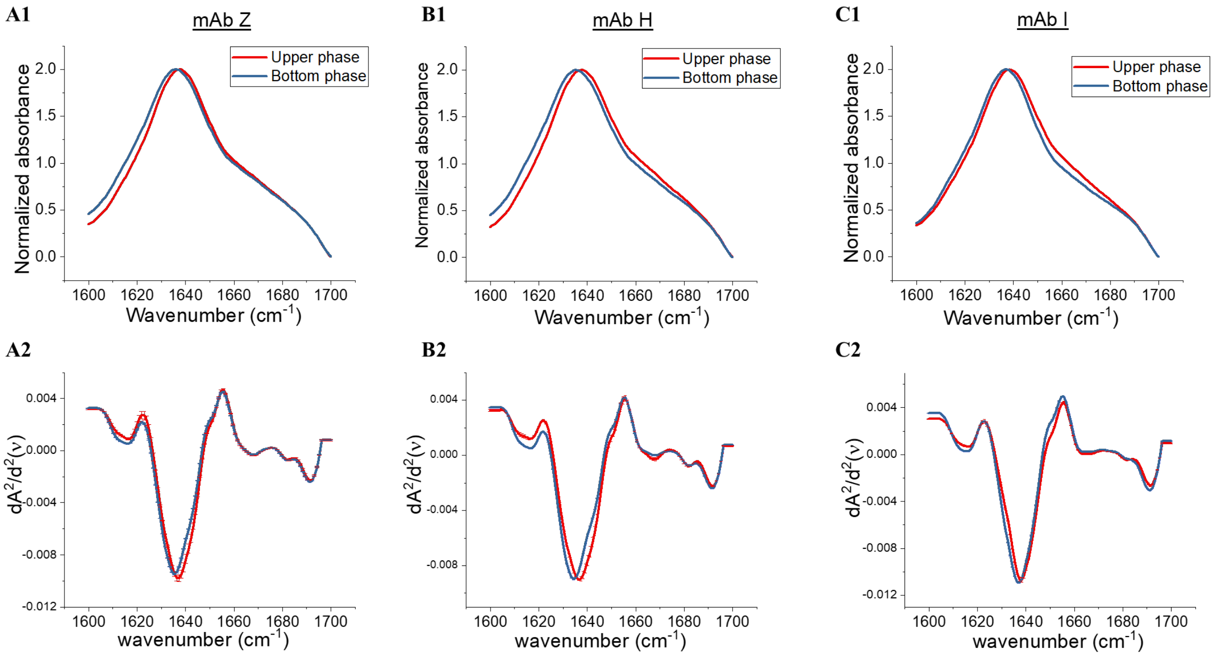


Figure 4.S2. FTIR amide I spectra of both upper and bottom phases of mAb Z (A1), mAb H (B1), and mAb I (C1). The second derivative FTIR spectra of the two phases for mAb Z (A2), mAb H (B2), and mAb I (C2). See Table 4.1 for protein concentrations and solution conditions.

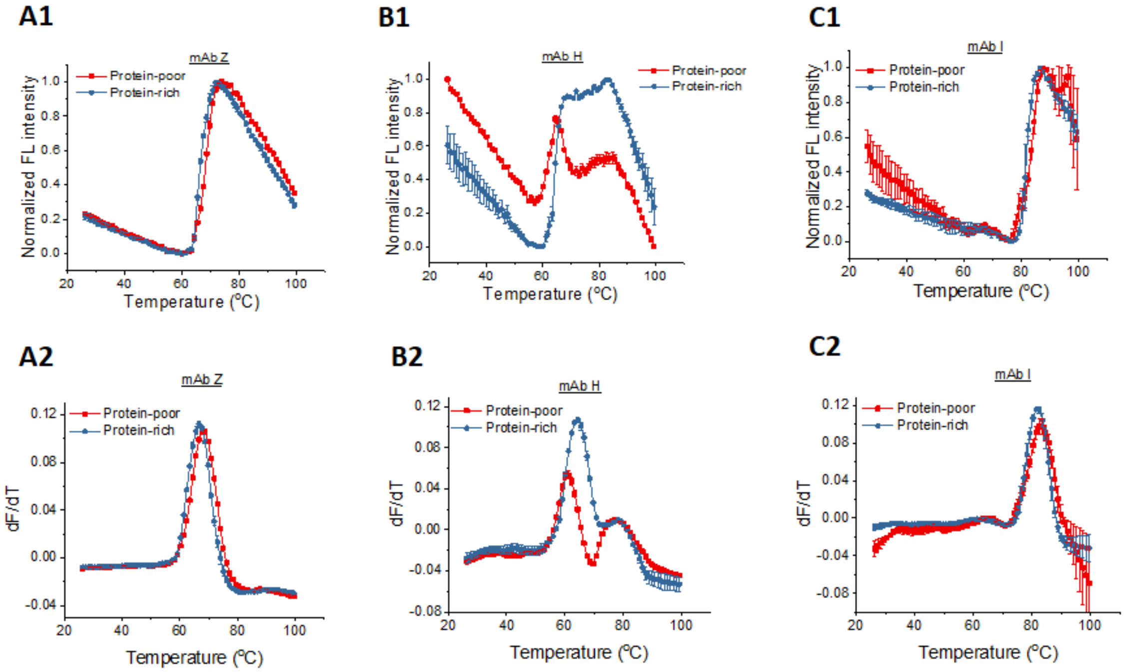


Figure 4.S3. Thermal stability of mAbs measured by intrinsic fluorescence spectroscopy. Temperature dependent normalized fluorescence intensity of mAb Z (A1), H (A2), and I (A3). The first derivative curves of mAb Z (B1), H (B2), and I (B3). Melting temperatures are listed in a tabular form (Table 4.S1). See Table 4.1 for protein concentrations and solution conditions.

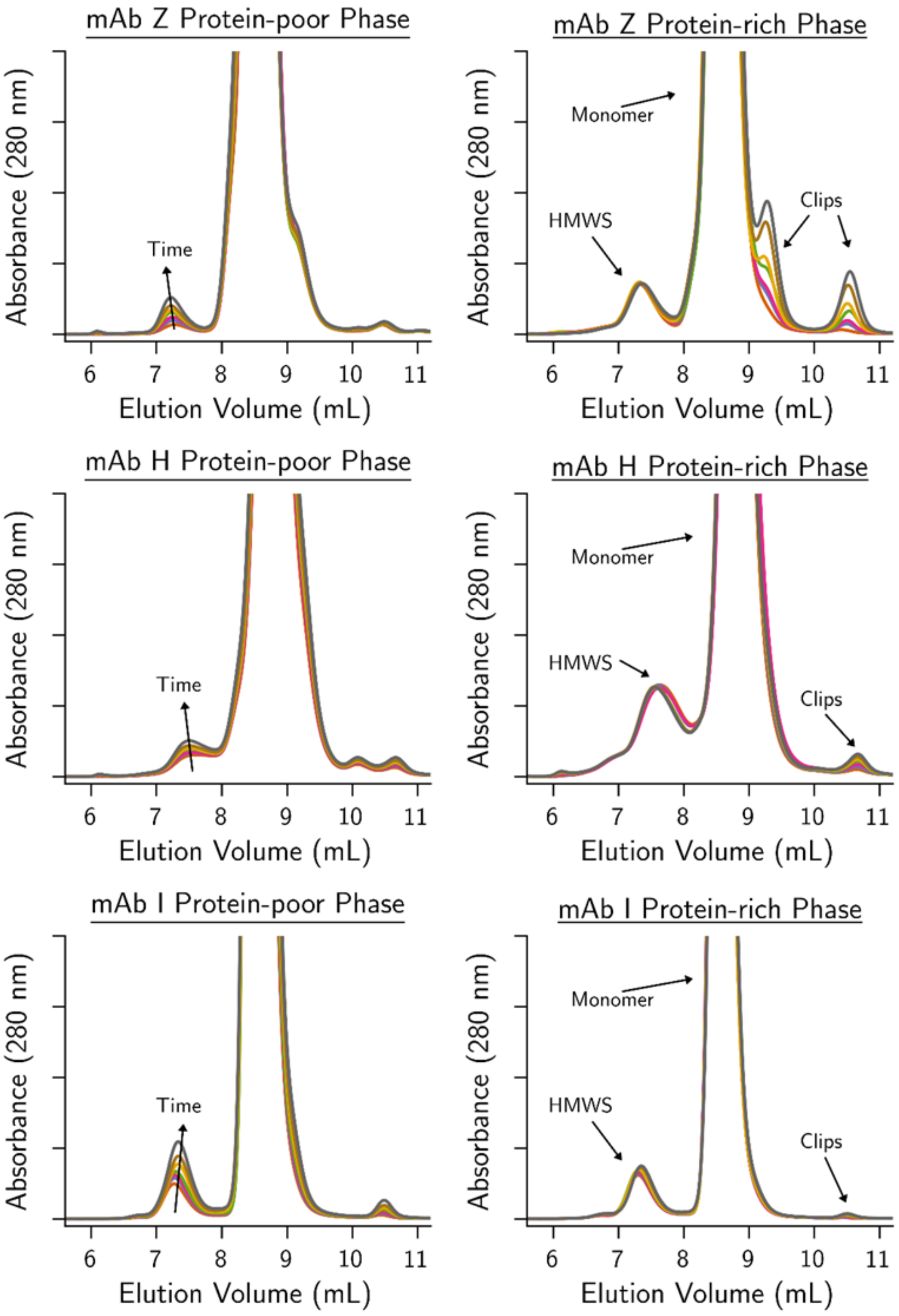


Figure 4.S4. Representative SEC chromatograms for each antibody phase over two weeks at 40 °C. The figure on the left is the protein-poor phase for each antibody. The figures on the left has an arrow indicating the direction of time (chromatograms correspond to approximately every two days). The figures on the right have arrows indicating the assignments of HMWS, monomer and clipped species. All species eluting before monomer were grouped as HMWS and all species eluting after grouped as fragmented or clipped species.

Table 4.S1. Melting temperatures for each phase of each mAb from intrinsic tryptophan fluorescence melts.

mAb	Phase	T_m1	T_m2
Z	Protein-poor	68.2 ± 0.1	NA
Z	Protein-rich	66.6 ± 0.1	NA
H	Protein-poor	61.1 ± 0.1	77.3 ± 0.2
H	Protein-rich	64.4 ± 0.1	78.3 ± 0.1
I	Protein-poor	65.1 ± 0.5	83.0 ± 0.5
I	Protein-rich	61.2 ± 1.1	81.8 ± 0.6

5 pH-Dependent Phase Behavior and Stability of Cationic Lipid–mRNA Nanoparticles

5.1 Introduction

Until recently the primary concern of gene therapy has been the ability to deliver sufficient coding nucleic acids to cells *in vivo*. The successful approval of the Onpatro lipid nanoparticle (LNP) formulation of siRNA to treat transthyretin-induced amyloidosis brings hope that nonviral vectors have begun to overcome the delivery barrier[1]. mRNA–LNPs are being investigated for a variety of disease indications from vaccines to protein replacement therapies[2–5]. Compared to viral vectors, mRNA–LNPs do not have major issues with immunogenicity of the particles and are inherently nonreplicating.

A critical step enabling therapeutic use of mRNA–LNPs was the addition of cationic lipids with a pKa between 6.2 and 7 to greatly increase their transfection efficiency[6–8]. Following cellular uptake of lipid nanoparticles into endosomes there is a decrease in their pH. The decrease in endosomal pH results in the protonation of the cation lipid and association of the cationic lipid with anionic lipids in the endosomal membrane resulting in bilayer disruption and release of RNA into the cytoplasm[3].

mRNA–LNPs are formed by the rapid mixing of an organic phase containing a cocktail of lipids with an aqueous phase containing mRNA below the pKa of the ionizable lipid (here we use dilinoleylmethyl-4-dimethylaminobutyrate (DLin-MC3-DMA)). In the organic phase there are additional lipids: 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), cholesterol, and 1,2-dimyristoyl-sn-glycerol-methoxy poly(ethylene glycol) 2000 (PEG-DMG 2000). The core of mRNA–LNPs primarily consist of mRNA complexed with the ionizable lipid along with cholesterol[9–12]. Generally the PEG lipid and DSPC decorate the outside of the particles and the

PEG lipid prevents non-specific interaction of mRNA–LNPs. The effect of pH on mRNA–LNPs structure has implications both for their cellular uptake mechanism and for the design of a stable liquid formulation.

Typically the only physical traits of mRNA–LNPs which are characterized are size, zeta potential and pKa[6,7,13]. These traits can be varied by altering the lipids or their fractional composition and screened for efficacy. There have also been some more thorough characterizations using small angle scattering methods, cryo-TEM, and NMR which provided more detailed structural information to guide the rational design and formulation of LNPs[9–12]. However, these techniques are generally low-throughput and thus cannot be easily applied to characterize mRNA–LNPs across many solution conditions. Our lab has developed an approach using lower resolution techniques with a much higher throughput to generate empirical phase diagrams (EPDs)[14,15]. An EPD is a visual summary of the behavior of a macromolecule across solution conditions of interest. This permits analysis of hundreds of distinct conditions to better understand the behavior of macromolecules as a function of solution variables such as pH, ionic strength, and temperature. Our group has previously applied this approach to characterize cationic lipid-DNA complexes, liposomes, and a variety of other macromolecules[15–17].

To investigate the effects of complexation of mRNA and cationic lipids in LNPs we here characterize: EGFP mRNA, LNPs (without mRNA), and EGFP mRNA–LNPs. First we used DLS, zeta potential, and pKa determination by TNS fluorescence to understand basic physical properties. Next, to evaluate the phase behavior of the lipids a function of pH and temperature we used a combination of FTIR and differential scanning calorimetry (DSC). Circular dichroism and DSC are employed to probe tertiary structure and stability of EGFP mRNA. Optical density measurements are used to evaluate the colloidal stability of the mRNA–LNPs. These results are then visually summarized by an EPD. A six-week stability study of the mRNA–LNPs at 4 °C and

25 °C was used to evaluate the correlation between the biophysical measurements and the stability of mRNA–LNPs. Time-dependent pH-jump experiments were performed to evaluate the rate and reversibility of pH changes on the solution structure of mRNA–LNPs.

5.2 Materials and Methods

5.2.1 Materials

Enhanced green fluorescent protein (EGFP) mRNA was purchased from Trilink Biotechnologies (San Diego, USA). The ionizable cationic lipid O-(Z, Z, Z, Z-heptatriaconta-6,9,26,29-tetraem-19-yl)-4-(N,N-dimethylamino)butanoate (DLin-MC3-DMA) was synthesized by AstraZenaca. 1,2-dis-tearoyl-sn-glycero-3-phosphocholine (DSPC) was purchased from Avanti polar lipids. 1,2-Dimyristoyl-rac-glycero-3-methylpolyoxyethylene (PEG-DMG 2000) was purchased from NOF Corporation. Water was treated with diethyl pyrocarbonate (DPEC) prior to use.

5.2.2 LNP synthesis

LNPs were prepared using a microfluidic setup previously described in greater detail[18]. Briefly, EGFP mRNA was dissolved in 20 mM acetate at pH 4. A 50:10:38.5:1.5 ratio of Dlin-MC3-DMA:DSPC:Chol:PEG-DMG$_{2000}$ was dissolved in ethanol. The lipid mixture was combined with a 50 mM citrate buffer (pH 4.0) containing mRNA at a ratio of 3:1 (aqueous : ethanol) using a microfluidic mixer (Precision Nanosystems, Vancouver, CA). Empty lipid nanoparticles were formed by using a citrate buffer without mRNA in the aqueous phase. Following mixing of the organic and aqueous phases, samples were immediately diluted 1000× into PBS (pH 7.4) with Amicon centrifugal filters (EMD Millipore). Following dialysis into PBS, samples were dialyzed against citrate-phosphate buffer at pH 4, 5, 6, 7, and 8. Citrate-phosphate (CP) buffer was prepared at 20 mM at pH 4.0, 5.0, 6.0, 7.0, and 8.0 from citric acid anhydrous and sodium phosphate dibasic anhydrous. The ionic strength of each buffer was controlled to I = 0.15 M by the addition of NaCl. Formulations were concentrated with Amicon centrifugal filters (EMD Millipore) to an RNA concentration of 0.1 mg/ml, filtered with a 0.22 µm filter and stored at 4 °C until use. LNP were concentrated to the same lipid concentration (2 mg/ml) as mRNA–LNPs. All experiments comparing mRNA–LNPs and LNPs kept the concentration of lipids constant. The encapsulation efficiency and concentration of mRNA was determined using the Ribo-Green assay (ThermoFisher, Waltham, MA).

5.2.3 In Situ *Determination of pKa with TNS*

6-(p-toluidino)-2-naphthalenesulfonic (TNS) acid sodium salt in DMSO (Sigma-Aldrich, St. Louis, MO) was prepared as a 100 µM stock solution in distilled water. An aliquot of the TNS solution was added to give a final concentration of 1 µM and following vortex mixing fluorescence intensity was measured at room temperature using a SpectraMax M5 spectrophotometer employing an excitation wavelength of 325 nm and emission wavelength of 435 nm. A sigmoidal fit to the fluorescence data was used to determine the pKa of the LNPs at the pH of the inflection point.

5.2.4 Zeta Potential

The zeta potentials of the LNPs were determined by phase analysis light scatting (PALS) at a scattering angle of 15 at 25° C with a Zeta PALS instrument (Brookhaven Instruments Corp., New York, USA). An electric field strength between 14 and 16 V/cm was used. Data were collected with 10–15 cycles of the electric field for each experiment and averaged. Zeta potential was calculated using the measured electrophoretic mobility with the Smoluchowski approximation.

5.2.5 Dynamic Light Scattering

A DynaPro II Plate Reader (Wyatt Technology, Santa Barbara, CA) was employed for dynamic light scattering measurements. Incident light was detected in a backscattering geometry and analyzed with an autocorrelator. A clear-bottomed 384 well plate holding 20 µL of sample was read at 20 °C. Samples were measured 5 times with a 15 second acquisition time. Cumulant analysis was used to determine hydrodynamic radius and polydispersity. Errors are reported as standard deviation of 3 replicates.

5.2.6 Kinetic DLS experiments

For kinetic experiments all samples were diluted 100x from a concentration of 0.1 mg/ml mRNA prior to size measurement. There were four samples in this experiment, two initially at pH 6 and two initially at pH 8. One of the pH 6 samples was diluted into pH 6 and the other diluted into pH 8. One of the pH 8 samples was diluted into pH 6 and the other diluted into pH 8. Samples were measured in triplicate. It took approximately two minutes following dilution before the first size measurement was completed. Experiments were performed in triplicate.

5.2.7 Circular Dichroism

An Applied Photophysics Chirascan (Leatherhead, UK) with a six-cylinder cuvette holder was used to record circular dichroism spectra. Two hundred microliters of sample were placed in quartz cuvettes with a 1 mm pathlength at a concentration of 0.1 mg/ml RNA. Spectra were recorded from 200 to 300 nm in 1 nm steps. Samples were baseline corrected by subtracting their corresponding buffer. For temperature melts samples were capped using Teflon stoppers and measured every 2.5 °C from 10 to 90 °C.

5.2.8 Differential Scanning Calorimetry

Differential scanning calorimetry was performed using a Malvern Microcal VP-Capillary calorimeter (Malvern, UK). Samples were placed in a 96 well plate and held in an autosampler at 4 °C prior to measurement. The temperature was increased at ramp rate of 1 °C/min from 10 to 100 °C. Samples were buffer subtracted, baselined, and normalized for concentration using Origin 7.0 (Origin Lab, Northampton, MA).

5.2.9 FTIR Spectroscopy

FTIR measurements were performed with a Tensor-27 FTIR spectrometer (Bruker, Billerica, MA) using a Bio-ATR cell. Samples were concentrated to 1 mg/ml RNA. The detector was cooled with liquid N_2 for 20 min before use. During this time the interferometer was also continuously purged with N_2 gas. Spectra were obtained from 800 – 4000 cm^{-1} with a resolution of 4 cm^{-1}. For the thermal melting experiments, 64 scans were obtained, and samples were scanned from 10 to 90 °C using an increment of 2.5 °C/step and an equilibration time of 2 min at each step. Raw FTIR spectra were buffer subtracted, water vapor and CO_2 compensated, baseline corrected, and normalized using the OPUS V6.5 software (Bruker, Billerica, MA). Peak positions were determined using in-house python scripts. For the peaks originating from CH_2 wagging, peak intensities were monitored by normalizing to the neighboring CH_3 umbrella peak.

5.2.10 Optical Density

Thermal aggregation of mRNA–LNPs was monitored using a Cary 100 UV-Visible spectrophotometer (Aligent Technologies, Santa Clara, California) equipped with a 12-cell temperature controlled Peltier device. Samples were placed in quartz cuvettes (1 cm path length) at a concentration of 0.1 mg/ml mRNA. The temperature was ramped from 10 to 90 °C in 2.5

°C increments with a 120 s equilibration time. Optical density was recorded at 450 nm (a non-absorbing region of the spectrum).

5.2.11 Phase Diagram

Phase diagrams were prepared as previously described[15,19]. Briefly, a singular value decomposition was applied to a matrix containing data from FTIR, DSC, optical density, and circular dichroism. The top three values of the SVD were each normalized and represented by a red, green, or blue color value. At each pH and temperature, a tuple of these three values was used to set an RGB color for empirical phase diagram visualization.

5.2.12 Transfection Efficiency

HeLa cells were grown in DMEM supplemented with 100 mL/L of FBS, 10 IU/mL of penicillin, and 10 mg/mL of streptomycin. The cells were incubated at 37 °C in a 5% CO_2 environment and subcultured by partial digestion with 0.25% trypsin and ethylenediaminetetraacetic acid. Cells were passaged at ~80% confluence. HeLa cells were seeded at 15,000 cells per well in a black 96-well plate. Cells were then washed off the plates and suspended in a volume of 0.2 mL media which was used for FACs and fluorescence microscopy. Agarose RNA gels (ThermoFisher Scientific) were run to determine the quality of the mRNA within the LNPs.

5.2.13 mRNA–LNP Stability

One mL of mRNA–LNPs formulated in citrate phosphate were stored at either 4 or 25 °C for up to 6 weeks. At each timepoint an aliquot was pulled from the sample to measure hydrodynamic radius, polydispersity, transfection efficiency, mRNA concentration, mRNA encapsulation efficiency, and mRNA integrity.

5.3 Results

5.3.1 pKa, Size, and Zeta Potential

We first addressed the most commonly characterized aspects of mRNA–LNP structure: pKa, size and zeta potential. The pKa of the mRNA–LNPs was found to be 6.43 in Figure 5.1A, in good agreement with previous results for DLin-MC3-DMA[7]. The hydrodynamic radius of the mRNA–LNPs was similar for pH 4, 5 and 6 at ~46 nm, while the mRNA–LNPs at pH 7 and 8 were ~35 nm (Figure 5.1B). LNPs were smaller, with radii near 28 nm which was fairly consistent across pH. EGFP mRNA alone had a radius ~ 12 nm which did not change significantly with pH (Figure 5.S1). The mRNA–LNPs zeta potential decreased from 20 to -20 mV as pH increased from 4 to 8. Zeta potential valus shown in Figure 5.1C was similar for mRNA–LNPs and LNPs. The LNPs decreased in zeta potential faster as pH increased but at pH 4 and 8 the LNPs and mRNA–LNPs had equivalent zeta potential.

5.3.2 Circular Dichroism

Circular dichroism (CD) of mRNA arises from the double stranded regions which are distinctly coiled depending on the interactions of base pairs. The CD spectra of EGFP mRNA has maxima at 262 and 223 nm and minima at 209 and 247 nm (Figure 5.2A). This is consistent with the A form of mRNA which has a maximum around 260 nm, a shoulder at 220 nm and a minimum near 210 nm[20,21]. mRNA does not typically adopt the B form because of the unfavorable steric clash of the 2-OH with the 3`-OH. mRNA can adopt B form conformations but because it is unfavorable typically less than 5% of structures are in conformations similar to B form[22]. The minima at 247 nm becomes more pronounced and red shifts ~3 nm for EGFP mRNA at higher pH. This indicates a conformational equilibrium between the A and B forms with the percentage of B-form like structures increasing at higher pH for EGFP mRNA free in solution.

mRNA–LNPs also show a pH dependent shift in mRNA structure (Figure 5.2B). At pH below the pKa of the mRNA–LNPs the CD signal becomes more negative. This shift indicates an increasingly favored A form mRNA at lower pH. At higher pH the RNA CD spectra are similar to those of EGFP mRNA in solution. This suggests that ionic interactions of Dlin-MC3-DMA and RNA induce more A form in EGFP mRNA. Measurements of EGFP LNPs below 220 nm were hindered by the CD signal from the lipids (absorbing double bonds) while above 220 nm the spectra of LNPs did not exhibit signal as shown in Figure 5.S2.

CD melts performed to compare the stability of mRNA tertiary structure in solution with the structure inside the mRNA–LNPs are shown in Figure 5.2C and 5.2D. mRNA in solution exhibited a gradual loss of structure as temperature was increased at pH 6, 7 and 8. For pH 4 and 5, along with the gradual loss of structure there were sharper transitions at ~ 40 °C and 85 °C, respectively. For mRNA–LNPs at all pH conditions there were several gradual transitions as a function of temperature. For pH 4 and 5 there was a large increase in the CD signal which corresponded to the aggregation of the nanoparticles (discussed in later the optical density section)

5.3.3 Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) was employed to understand the thermally inducd structural transitions of mRNA and lipids. mRNA typically has thermodynamic transitions associated with the unraveling of hydrogen-bonded regions. DSC can detect the coming apart of these regions as enthalpic peaks in a DSC thermogram. EGFP mRNA exhibited many distinct and overlapping thermal transitions at each pH as shown in Figure 5.3A. The mRNA unfolding curves for each pH differed significantly in the transition temperatures and enthalpies indicative of highly pH dependent stability. DSC was able to resolve multiple discrete transitions that CD was not sensitive enough to identify. pH 4 and 5 exhibited sharp peaks at 40 and 85 °C which were distinct from the thermograms of pH 6, 7 and 8. These peaks agree well with the sharp transitions observed for pH 4 and 5 using circular dichroism shown in Figure 5.2C.

DSC of LNPs (without mRNA) as shown in Figure 5.3C exhibited no transitions except at pH 5 and 6. For pH 5 there was a large broad transition with a maximum at ~65 °C. At pH 6 there is a similar transition at 25 °C. For mRNA–LNPs (Figure 5.3B) only one transition was observed at pH 7 and 8, while pH 6 had two transitions, where pH 5 had one transition, and pH 4 lacked transitions. The transition for pH 5 and the first transition for pH 6 are similar to those observed for the empty LNPs. We assign these peaks to inverse hexagonal to lamellar phase transitions (further supported by FTIR, see below). There is a transition of similar width and magnitude for pH 6, 7, and 8 at 72, 68, and 65 °C, respectively. This transition may arise from the dissociation of mRNA and cationic lipids. This hypothesis is also supported by the small negative feature seen by CD in Figure 5.2D for pH 6, 7 and 8 at similar temperatures since one might expect a change in mRNA structure following disassociation. The cationic lipid would become increasingly charged at low pH which could explain the increase in transition temperature and eventual disappearance of this transition since at pH 4 and 5 the ionic interactions between DLin-MC3-DMA and mRNA may be strong enough prevent dissociation at high temperature. For all pH conditions the individual transitions of the RNA were no longer observed, suggesting the RNA and lipid phase transitions become coupled upon complexation.

5.3.4 Optical Density

The colloidal stability of the mRNA–LNPs was assessed by measuring solution optical density as a function of pH and temperature. Increases in optical density in Figure 5.4 correspond to aggregation of the mRNA–LNPs. Optical density was monitored at 450 nm so that all observed optical density changes could be attributed to scattering. Aggregation began at a lowest temperature at pH 4 of near 65 °C. At pH 5 this increased to around 73 °C and at the higher pH values of 6, 7 and 8 the mRNA–LNPs did not aggregate until approximately 80 °C.

5.3.5 FTIR Spectroscopy

FTIR is able to probe the phase behavior and conformation of lipids in bilayers[23]. FTIR spectra for mRNA–LNPs from pH 4 to 8 are shown in Figure 5.4A. We assume the FTIR signal arises from lipids although a small fraction of the signal does originate from the mRNA (20 : 1 mass lipids to mRNA). The C-N stretching band near 1200 cm^{-1} blue shifts with increasing pH as shown in Figure 5.4B. This peak is indicative of the pKa of the amine group and did not change significantly with temperature. To evaluate the lipid phase organization and geometry, the C=O stretching peak was followed as a function of pH and temperature in Figure 5.4C. An increase in C=O stretching frequency of approximately 4 cm^{-1} indicates an inverse hexagonal to lamellar phase transition[23]. At pH 4 and 5 the C=O stretching frequency has a value of ~1729.5 cm^{-1} while at pH 7 and 8 the initial value of the C=O stretching frequency is 1732.5 cm^{-1}. This suggests that at pH 4 and 5 the organizational structure of the lipids is inverse hexagonal while at pH 7 and 8 the structure of the lipids is lamellar. At pH 6 the initial value of the C=O stretching frequency is 1731 cm^{-1} which suggests a mixture of the two phases. As temperature increases, a transition between inverse hexagonal phase and a lamellar phase was observed at pH 5. At pH 6 a similar transition was observed but the magnitude of the shift was smaller since the initial C=O stretching band position was already somewhere in-between the two. This is in good agreement with the transition observed by DSC at pH 5 and 6 since at pH 5 the transition has a large magnitude and at pH 6 a similar but smaller magnitude transition was observed.

The CH_2 scissoring band position and contour is determined by the lateral interchain interactions of methylene chains and is sensitive to the organizational structure of lipids in assemblies. The CH_2 scissoring band observed for the mRNA–LNPs has one main peak (not split) which is consistent with inverse hexagonal or lamellar structure. The peak position was 1466 cm^{-1} at pH 4 and gradually decreased to 1464.5 cm^{-1} at pH 8. For all pH conditions there is a gradual decrease in the peak position with an increase in temperature as shown in Figure 5.4D. Though this band can sometimes be used to directly assign the organizational structure lipid subcells, it can be interfered with by end-methyl groups in hydrocarbon chains. Thus, we relied more heavily on C=O and DSC to assign the lipid phase organizational structure. The fact that this peak is not split to supports our phase assignments.

The CH_2 wagging peaks are indicative of the nonplanarity of the CH_2 bonds of the carbon chains. The three bands which arise from non-planarity are kink (gtg), double gauche (gg), and end gauche (tg). In Figure 5.4E the intensity of the gtg peak at 1367 cm^{-1} has been normalized to the methylene umbrella at 1378 cm^{-1} for all pH and temperatures. At low pH the gtg peak height is ~0.75 while at high pH it is ~0.92 indicating a higher fraction of kinked gauche CH_2 groups at higher pH. As temperature increased both pH 5 and 6 saw a gradual increase in the fraction of kinked CH_2 groups which coincided with the transition from inverse hexagonal to lamellar. The transition from inverse hexagonal to lamellar also coincided with the observation that mRNA–LNPs are approximately 14 nm smaller in diameter at pH values below the pKa suggesting that the phase of the lipids and their size are linked. The increased kinks in the CH_2 groups may be responsible for the increase in size of mRNA–LNPs at low pH since the apolar chains cannot pack together as tightly.

The methylene symmetric (${}_{s}\upsilon_{CH2}$) and asymmetric stretching modes (${}_{as}\upsilon_{CH2}$) give rise to IR-active bands at 2850 and 2920 cm^{-1}. The methylene stretching modes are shown for all pH conditions in Figure 5.S2A. The $_{s}\nu_{CH2}$ band is relatively free from contributions of CH_3 groups and is sensitive to the conformal disorder of trans methylene chains. A conformational disordering of the lipid structure is associated with the increase is this peak position. Figure 5.S2B depicts the $_{s}\nu_{CH2}$ peak position as a function of temperature and pH. At 10 °C the $_{s}\nu_{CH2}$ band has a maximum of 2852.5 cm^{-1} for pH 4 and increases with pH to 2853.5 cm^{-1} at pH 8, indicating an increase in conformational disorder at higher pH. Since DLin-MC3-DMA will be less ionized at higher pH, the nonionic fraction will act as a solvent to some extent, which leads to the increase in disorder of the system. All pH conditions saw a gradual increase in peak position as with increasing temperature as well.

5.3.6 Empirical Phase Diagram

An EPD was used to visually summarize the solution behavior of mRNA–LNPs from 15 to 90 °C and pH 4 to 8. The EPD for mRNA–LNPs in Figure 5.5 has at least six distinct regions. Tentative assignments for the regions are as follows: (I) inverse hexagonal phase and more A form RNA, (II) lamellar phase, (I / II) transition region from region I to region II, (III) aggregated particles with inverse hexagonal structure, (IV) a lamellar lipid phase with disassociated RNA and cationic lipids, and (V) aggregated particles with lamellar phase lipids. It is important to recognize that the EPD is not meant to reflect equilibrium since some of the process involved here are not reversible.

5.3.7 Accelerated Stability Study

To understand the effect of pH on lipid nanoparticle stability an accelerated stability study was performed. Samples were held at 4 and 25 °C for up to six weeks with timepoints every week. Transfection efficiency of HeLa cells was assessed using fluorescence activated cell sorting (FACs) and fluorescence microscopy. Transfection efficiencies varied significantly with pH even at time zero (Figure 5.6A, E). Transfection efficiency at time zero decreased as pH increased from 85 % at pH 4 and 5 to a pH 8 mRNA–LNPs transfection efficiency of just under 40%. Differences in transfection efficiency as a function of pH at time zero suggest the inverse hexagonal structure mRNA–LNPs at low pH are able to traffic more mRNA to the cytosol. In general FACs results were well supported by the fluorescence microscopy data, examples of microscopy results are shown in Figure 5.S7.

For storage at 4 °C, pH 4 and 5 slowly lost transfection efficiency over the course of six weeks but still retained transfection efficiency near 50% at the final six-week timepoint (Figure 5.6A). For pH 6, 7, and 8 mRNA–LNPs failed to transfect a significant number of cells at the four-week timepoint. To understand the loss in transfection efficiency a ribogreen assay was performed to analyze RNA concentration and encapsulation efficiency, an RNA gel was run to assay the integrity of the RNA, and dynamic light scattering was performed to measure the colloidal stability at each timepoint. For the mRNA–LNPs at 4 °C the mRNA concentration and encapsulation efficiency were constant throughout the study (Figure 5.6B, C). The RNA gel for each timepoint are shown in Figure 5.4. Marker bands from ribosomal RNA in the far-left lane are the 18S (1.8 kilobases) band and approximately twice as dark 28S (4.7 kilobases) band. At time zero the RNA gel is similar for all pH values, showing one main band just past the 18S marker, in good agreement with the size of EGFP size of 996 bases. After one week a second band appeared at all pH near the 28S marker. This band could be from an mRNA multimer or lipid-bound mRNA. The total area of the RNA band was integrated in Figure 5.S6 for each timepoint which remained approximately constant for pH 4 and 5 but for pH 6, 7 and 8 decreased and the loss of the RNA increased with pH. The size and polydispersity of the mRNA–LNPs remained approximately constant over the six-week period as shown in Figure 5.6D and Figure 5.S8, respectively.

Storage of the mRNA–LNPs at 25 °C for one week resulted in a complete loss of transfection efficiency for pH 6, 7, and 8. At pH 4 after one week the complexes could only transfect ~10% of cells while at pH 5 they could only transfect ~5% of cells. For both pH 4 and 5 all transfection efficiency was lost after two weeks. At 25 °C there was a significant loss of mRNA as measured by the ribogreen assay (Figure 5.6F). This loss correlated well with mRNA gels shown in Figure 5.S5 (quantified in Figure 5.S6B). However, at pH 5, 6, and 7 the mRNA that remained was still encapsulated within the LNP (Figure 5.6G). At pH 4 significant aggregation of the particles was observed after four weeks (Figure 5.6H). For pH 7 and 8 an increase of approximately 10 nm in radius was observed. Overall the mRNA–LNPs exhibited poor stability at 25 °C.

5.3.8 Kinetics of mRNA–LNPs Phase Transitions

To understand why a higher transfection efficiency occurred below the pKa of the mRNA–LNPs, we performed a kinetic analysis of the phase transition upon pH change using DLS to track the size change associated with the inverse hexagonal to lamellar phase transition. Figure 5.8 shows the time dependent size of mRNA–LNPs at pH 6, pH 8, following dilution from pH 6 to pH 8, and following dilution from pH 8 to pH 6. At pH 6 and pH 8 the mRNA–LNPs are approximately 32 nm and 46 nm in radius, respectively, and remain at that size over the course of the experiment. Upon dilution from pH 8 into pH 6 the mRNA–LNPs instantly (within 2 min of dilution) increased in size from 32 to 36 nm and steadily increased in size to 45 nm after 6 hrs. Following dilution from pH 6 to pH 8 there was no significant change in size. The irreversible nature of this process may signify that the lamellar mRNA–LNPs at pH 8 is a kinetically trapped phase. This is consistent with the FTIR and CD data which show that the lipids are more organized and the mRNA more structured at lower pH, probably indicating a more thermodynamically stable complex. It is also implied that inverse hexagonal phase is more stable since at pH 5 and 6 there is a transition from an inverse hexagonal to lamellar phase as temperature increases.

5.4 Discussion

mRNA–LNPs have been studied by several groups previously using a combination of SAXS and cryo-TEM to better understand their structure[9–12]. LNPs of a similar composition have been previously examined by SAXS and determined to possess lipid inverse hexagonal structure at low pH in ethanol citrate buffer for both empty and siRNA encapsulating LNPs[9]. The inverse hexagonal structure was determined by three peaks in a q-position ratio of $1:\surd 3:\surd 4$. A q peak at 1 nm^{-1} corresponded to a cylinder center to center distance of ~6 angstroms for the inverse hexagonal packing arrangement[9]. However, following dialysis into a PBS buffer only the q peak at 1 nm^{-1} remained suggesting a lamellar or multilamellar LNP structure. Here we showed mRNA–LNPs of the same composition also undergo a phase change upon change of solution pH and temperature. We did not perform SAXS experiments so our phase assignments are only tentative but we have a good deal of confidence since the particles used in that previous work and the particles used here have the same composition and the only difference is the particles here encapsulated EGFP mRNA (996 bases) while those particles encapsulated poly A (600 – 4000 bases). The carbonyl stretching observed using FTIR was also suggestive of an inverse hexagonal to lamellar phase transition as a function of pH, further solidifying our confidence in the phase assignments.

The phase transition from an inverse hexagonal structure (below pKa) to a lamellar structure (above pKa) perturbed many physical characteristics of the mRNA–LNPs. The RNA within the particles assumed a more A form tertiary structure below the pKa. The lipids within the bilayers showed an increase in kinked methylene groups above their pKa. Since kinked lipids pack less tightly in bilayers we hypothesize that more loosely packed lipids caused the increase in size of the mRNA–LNPs above the pKa. Perhaps the most striking was the fact that particles at lower pH were more efficient at transfection of HeLa cells. This is somewhat surprising since upon dilution of the particles in the HeLa cell media they are all at the same pH and thus would assume the same internal structural organization. This indicated that conversion between inverse hexagonal and lamellar mRNA–LNPs is a kinetic process which happens on a similar or longer time scale as cellular transfection. We confirmed this using pH-jump experiment in Figure 5.8 to show that following a change in pH from 8 to 6 the mRNA–LNPs take approximately 6 hrs to undergo a phase change from the lamellar to inverse hexagonal structure. However, solution conditions are significantly more crowded within liposomes than the solutions used which may alter the kinetics of the phase transition. Wittrup et al.[24] have shown that the transfection process begins after ~1 hr with a $t_{1/2}$ of 2 hrs. They were also able to show the time from endocytosis to release in the cytosol takes approximately 15 minutes. Our findings suggest that using mRNA–LNPs already in the inverse hexagonal form may increase endosomal escape. This is consistent with previous work by Koltover et al.[25] on cationic liposome-DNA complexes which demonstrated that inverse hexagonal complexes more rapidly fuse with anionic vesicles than lamellar complexes and thus release DNA more readily. We hypothesize that since the endocytic process may be more rapid than the phase transition for some fraction of the mRNA–LNPs. This is part of the reason cells were more efficiently transfected at pH 4. This is consistent with previous findings of Gilleron et al.[26] that only 1 – 2% of siRNAs delivered by LNPs escape from the endosomes into the cytosol.

Transfection efficiencies were equivalent using pH 4 and pH 5 buffered mRNA–LNPs and steadily decreased as the pH of the buffer increased. The phase transition from inverse hexagonal to lamellar had many disparate characteristics such as the particle size, the amount of methylene kinks, and the tertiary structure of the mRNA within the particles. The methylene stretching frequency varied more gradually with pH suggesting more solvent-like properties of the lipids at higher pH. The ability of mRNA to change conformation while inside LNPs may depend on the solvent-like properties of the lipids. The kinetic process of a lamellar to inverse hexagonal phase transition seems to be able to occur even at 25 °C as supported by the size increase in the particles stored at pH 7 and 8 after four weeks. This further suggests the lamellar phase is kinetically trapped. Transfection efficiency appeared most correlated with carbonyl stretching. Observed carbonyl stretching frequencies increased in a sigmoidal fashion with pH. Approximately 40% of DLin-MC3-DMA is expected to be protonated at pH 6 and thus a mixture of the two phases should be expected at time zero. The carbonyl stretching frequency at pH 6 is 1731 cm^{-1} indicating a partially lamellar and partially inverse hexagonal structure. Future experiments could use the carbonyl stretching frequency (in concert with other FTIR peaks) as a rapid method for phase order determination of mRNA–LNPs.

mRNA–LNPs are typically lyophilized for storage and reconstituted in PBS for use in clinical applications[3,27]. Developing a stable liquid formulation could significantly reduce process development costs. However, the poor stability of mRNA–LNPs observed here indicates this is a difficult task. When mRNA–LNPs had an inverse hexagonal structure they were significantly more stable with time. Therefore, development of a liquid formulation designed to maintain an inverse hexagonal structure may be critical. Since the inverse hexagonal structure appears to be more thermodynamically stable, it may be possible to perform two dialysis and reconcentration steps: first to pH 5 to form the inverse hexagonal structure and then to a second pH for the formulation. Even at a pH above the pKa the mRNA–LNPs should maintain the inverse hexagonal structure. Formulation at a pH greater than 5 could be beneficial since aggregation propensity from optical density experiments was highest at pH 4 and 5. Understanding the determinants of aggregation propensity of mRNA–LNPs is an exciting avenue for future research.

5.5 Conclusion

To elucidate structural and colloidal properties of mRNA–LNPs across a range of solution conditions, we applied an array of biophysical techniques. The mRNA–LNPs have inverse hexagonal structure below their pKa and above the pKa have a lamellar structure. At higher temperatures the low pH inverse hexagonal phase transitioned to a lamellar structure, indicating the inverse hexagonal phase is more stable. Overall there was significantly more structure at low pH as indicated by the increased A form mRNA and the lipid methylene stretching vibrations. We summarized the phase behavior of the mRNA–LNPs with an empirical phase diagram which captured at least six distinct phases. We then performed short-term stability at 4 and 25 °C at five different pH conditions. At time zero particles formulated at the lowest pH had the highest transfection efficiency. mRNA–LNPs formulated at pH 4 and 5 were most stable, retaining up to approximately 70% of their transfection ability after 6 weeks at 4 °C. At pH 6, 7, and 8 all transfection efficiency was lost after 4 weeks at 4 °C. We hypothesized that the difference in transfection efficiency at time zero between different pH values was due to the kinetics of the mRNA–LNP phase transition. pH-jump kinetic experiments suggested that the lamellar phase is kinetically trapped since a jump from a pH above the pKa to below resulted in a lamellar to inverse hexagonal phase transition but the opposite pH-jump did not result in an opposite transition.

5.6 References

1. Adams, D.; Gonzalez-Duarte, A.; O'Riordan, W. D.; Yang, C.-C.; Ueda, M.; Kristen, A. V.; et al. Patisiran, an RNAi Therapeutic, for Hereditary Transthyretin Amyloidosis. *N. Engl. J. Med.* **2018**, *379* (1), 11–21.

2. Pardi, N.; Hogan, M. J.; Pelc, R. S.; Muramatsu, H.; Andersen, H.; DeMaso, C. R.; et al. Zika Virus Protection by a Single Low-Dose Nucleoside-Modified MRNA Vaccination. *Nature* **2017**, *543* (7644), 248–251.

3. Cullis, P. R.; Hope, M. J. Lipid Nanoparticle Systems for Enabling Gene Therapies. *Mol. Ther.* **2017**, *25* (7), 1467–1475.

4. Bahl, K.; Senn, J. J.; Yuzhakov, O.; Bulychev, A.; Brito, L. A.; Hassett, K. J.; et al. Preclinical and Clinical Demonstration of Immunogenicity by MRNA Vaccines against H10N8 and H7N9 Influenza Viruses. *Mol. Ther.* **2017**, *25* (6), 1316–1327.

5. Hajj, K. A.; Whitehead, K. A. Tools for Translation: Non-Viral Materials for Therapeutic MRNA Delivery. *Nat. Rev. Mater.* **2017**.

6. Hassett, K. J.; Benenato, K. E.; Jacquinet, E.; Lee, A.; Woods, A.; Yuzhakov, O.; et al. Optimization of Lipid Nanoparticles for Intramuscular Administration of mRNA Vaccines. *Mol. Ther. - Nucleic Acids* **2019**, *15* (April), 1–11.

7. Jayaraman, M.; Ansell, S. M.; Mui, B. L.; Tam, Y. K.; Chen, J.; Du, X.; et al. Maximizing the Potency of SiRNA Lipid Nanoparticles for Hepatic Gene Silencing in Vivo. *Angew. Chemie - Int. Ed.* **2012**, *51* (34), 8529–8533.

8. Alabi, C. A.; Love, K. T.; Sahay, G.; Yin, H.; Luly, K. M.; Langer, R.; Anderson, D. G. Multiparametric Approach for the Evaluation of Lipid Nanoparticles for SiRNA Delivery. *Proc. Natl. Acad. Sci.* **2013**, *110* (32), 12881–12886.

9. Yanez Arteta, M.; Kjellman, T.; Bartesaghi, S.; Wallin, S.; Wu, X.; Kvist, A. J.; et al. Successful Reprogramming of Cellular Protein Production through MRNA Delivered by Functionalized Lipid Nanoparticles. *Proc. Natl. Acad. Sci.* **2018**, *115* (15), E3351–E3360.

10. Viger-Gravel, J.; Schantz, A.; Pinon, A. C.; Rossini, A. J.; Schantz, S.; Emsley, L. Structure of Lipid Nanoparticles Containing SiRNA or MRNA by Dynamic Nuclear Polarization Enhanced NMR Spectroscopy. *J. Phys. Chem. B* **2018**, acs.jpcb.7b10795.

11. Kulkarni, J. A.; Darjuan, M. M.; Mercer, J. E.; Chen, S.; Van Der Meel, R.; Thewalt, J. L.; et al. On the Formation and Morphology of Lipid Nanoparticles Containing Ionizable Cationic Lipids and SiRNA. *ACS Nano* **2018**, *12* (5), 4787–4795.

12. Patel, S.; Ashwanikumar, N.; Robinson, E.; Xia, Y.; Mihai, C.; Griffith, J. P.; et al. Naturally-Occurring Cholesterol Analogues in Lipid Nanoparticles Induce Polymorphic Shape and Enhance Intracellular Delivery of MRNA. *Nat. Commun.* **2020**, *11* (1), 1–13.

13. Kauffman, K. J.; Dorkin, J. R.; Yang, J. H.; Heartlein, M. W.; Derosa, F.; Mir, F. F.; Fenton, O. S.; Anderson, D. G. Optimization of Lipid Nanoparticle Formulations for MRNA Delivery in Vivo with Fractional Factorial and Definitive Screening Designs. *Nano Lett.* **2015**, *15* (11), 7300–7306.

14. Maddux, N. R.; Joshi, S. B.; Volkin, D. B.; Ralston, J. P.; Middaugh, C. R. Multidimensional Methods for the Formulation of Biopharmaceuticals and Vaccines. *J. Pharm. Sci.* **2011**, *100* (10), 4171–4197.

15. Wei, Y.; Joshi, S. B.; Bhambhani, A.; Zeng, Y.; Larson, N. R.; Hu, G.; Deeds, E. J.; Middaugh, C. R. Chapter 7: An Empirical Phase Diagram: High-Throughput Screening Approach to the Characterization and Formulation of Biopharmaceuticals; 2020; pp 139–185.

16. Middaugh, C. R.; Ramsey, J. B. Analysis of Cationic-Lipid-Plasmid-DNA Complexes. *Anal. Chem.* **2007**, *79* (19), 7240–7248.

17. Ruponen, M.; Braun, C. S.; Middaugh, C. R. Biophysical Characterization of Polymeric and Liposomal Gene Delivery Systems Using Empirical Phase Diagrams. *J. Pharm. Sci.* **2006**, *95* (10), 2101–2114.

18. Zhigaltsev, I. V.; Belliveau, N.; Hafez, I.; Leung, A. K. K.; Huft, J.; Hansen, C.; Cullis, P. R. Bottom-up Design and Synthesis of Limit Size Lipid Nanoparticle Systems with Aqueous and Triglyceride Cores Using Millisecond Microfluidic Mixing. *Langmuir* **2012**, *28* (7), 3633–3640.

19. Kim, J. H.; Iyer, V.; Joshi, S. B.; Volkin, D. B.; Middaugh, C. R. Improved Data Visualization Techniques for Analyzing Macromolecule Structural Changes. *Protein Sci.* **2012**, *21* (10), 1540–1553.

20. Demidov, V. V. *Nucleic Acids: Structures, Properties and Functions*; 2001; Vol. 6.

21. Kypr, J.; Kejnovská, I.; Renčiuk, D.; Vorlíčková, M. Circular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37* (6), 1713–1725.

22. Sedova, A.; Banavali, N. K. RNA Approaches the B-Form in Stacked Single Strand Dinucleotide Contexts. *Biopolymers* **2016**, *105* (2), 65–82.

23. Lewis, R. N. A. H.; McElhaney, R. N. Membrane Lipid Phase Transitions and Phase Organization Studied by Fourier Transform Infrared Spectroscopy. *Biochim. Biophys. Acta - Biomembr.* **2013**, *1828* (10), 2347–2358.

24. Wittrup, A.; Ai, A.; Liu, X.; Hamar, P.; Trifonova, R.; Charisse, K.; Manoharan, M.; Kirchhausen, T.; Lieberman, J. Visualizing Lipid-Formulated SiRNA Release from Endosomes and Target Gene Knockdown. *Nat. Biotechnol.* **2015**, *33* (8), 870–876.

25. Koltover, I.; Salditt, T.; Rädler, J. O.; Safinya, C. R. An Inverted Hexagonal Phase of Cationic Liposome-DNA Complexes Related to DNA Release and Delivery. *Science (80-.).* **1998**, *281* (5373), 78–81.

26. Gilleron, J.; Querbes, W.; Zeigerer, A.; Borodovsky, A.; Marsico, G.; Schubert, U.; et al. Image-Based Analysis of Lipid Nanoparticle-Mediated SiRNA Delivery, Intracellular Trafficking and Endosomal Escape. *Nat. Biotechnol.* **2013**, *31* (7), 638–646.

27. Kulkarni, J. A.; Cullis, P. R.; Van Der Meel, R. Lipid Nanoparticles Enabling Gene Therapies: From Concepts to Clinical Utility. *Nucleic Acid Ther.* **2018**, *28* (3), 146–157.

5.7 Figures and Tables

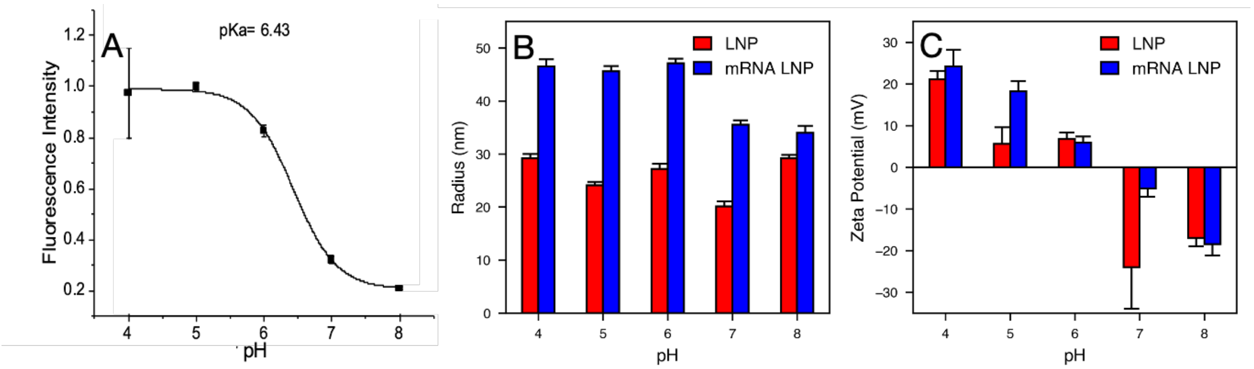


Figure 5.1. (A) Experimental pKa measurement by a TNS fluorescence assay. (B) size measurement of LNPs and mRNA–LNPs using DLS as a function of pH. (C) zeta potential measurement of LNPs and mRNA–LNPs as a function of pH.

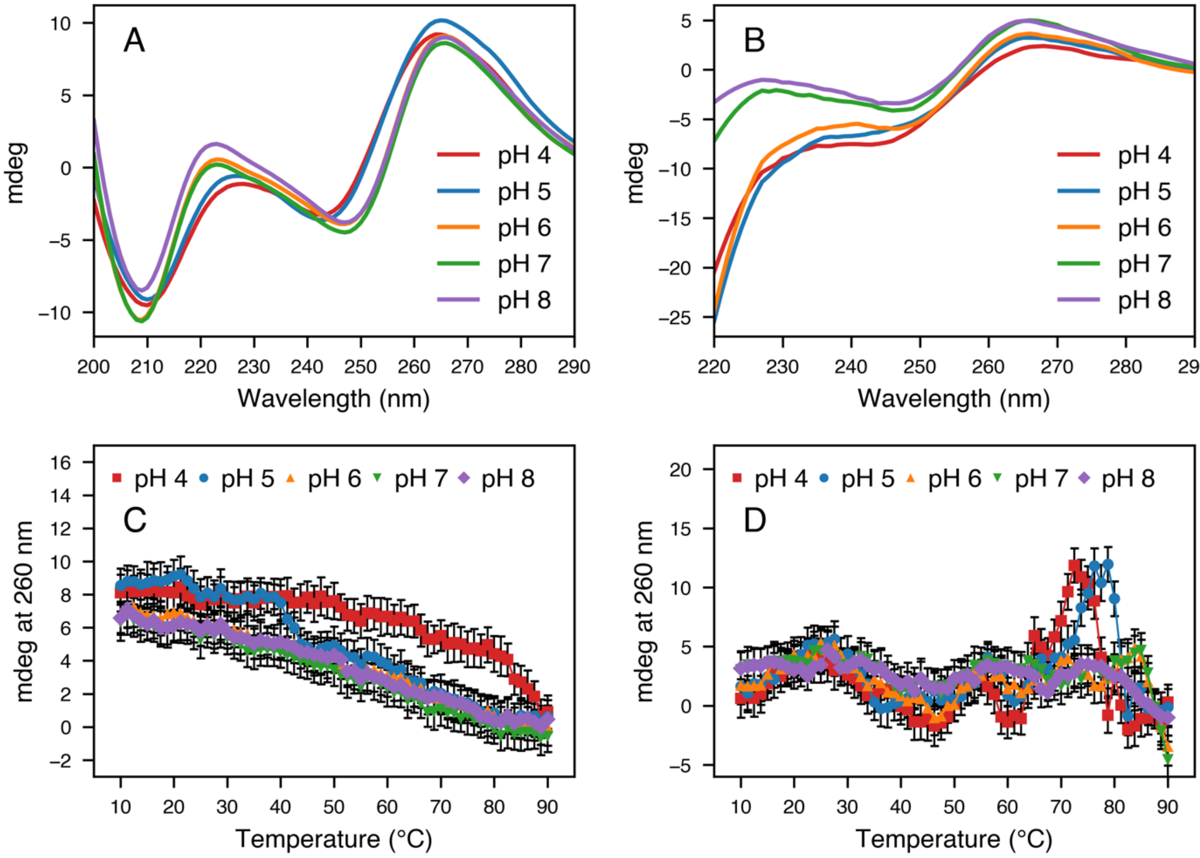


Figure 5.2. Circular dichroism of (A) EGFP mRNA and (B) mRNA–LNPs as a function of pH. Circular dichroism at 260 nm for (C) EGFP mRNA and (D) mRNA–LNPs as a function of pH and temperature.

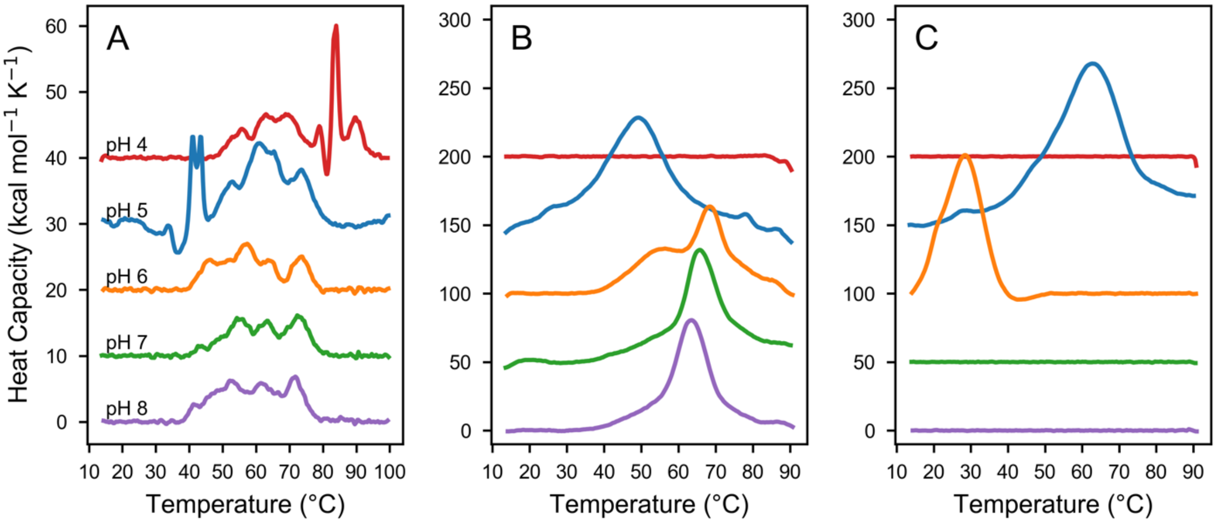


Figure 5.3. DSC chromatograms for mRNA, mRNA–LNPs, and empty LNPs. (A) DSC thermograms of EGFP mRNA from pH 4 to 8. (B) DSC thermograms of mRNA–LNPs from pH 4 to 8. (C) DSC thermograms LNPs from pH 4 to 8. For all figures thermograms are offset.

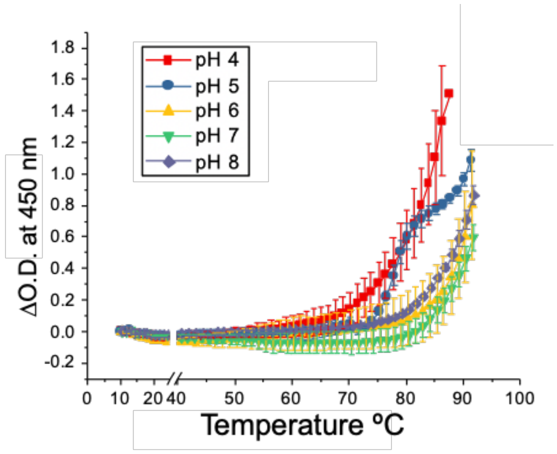


Figure 5.4. Optical density measurements as a function of pH and temperature for mRNA–LNPs.

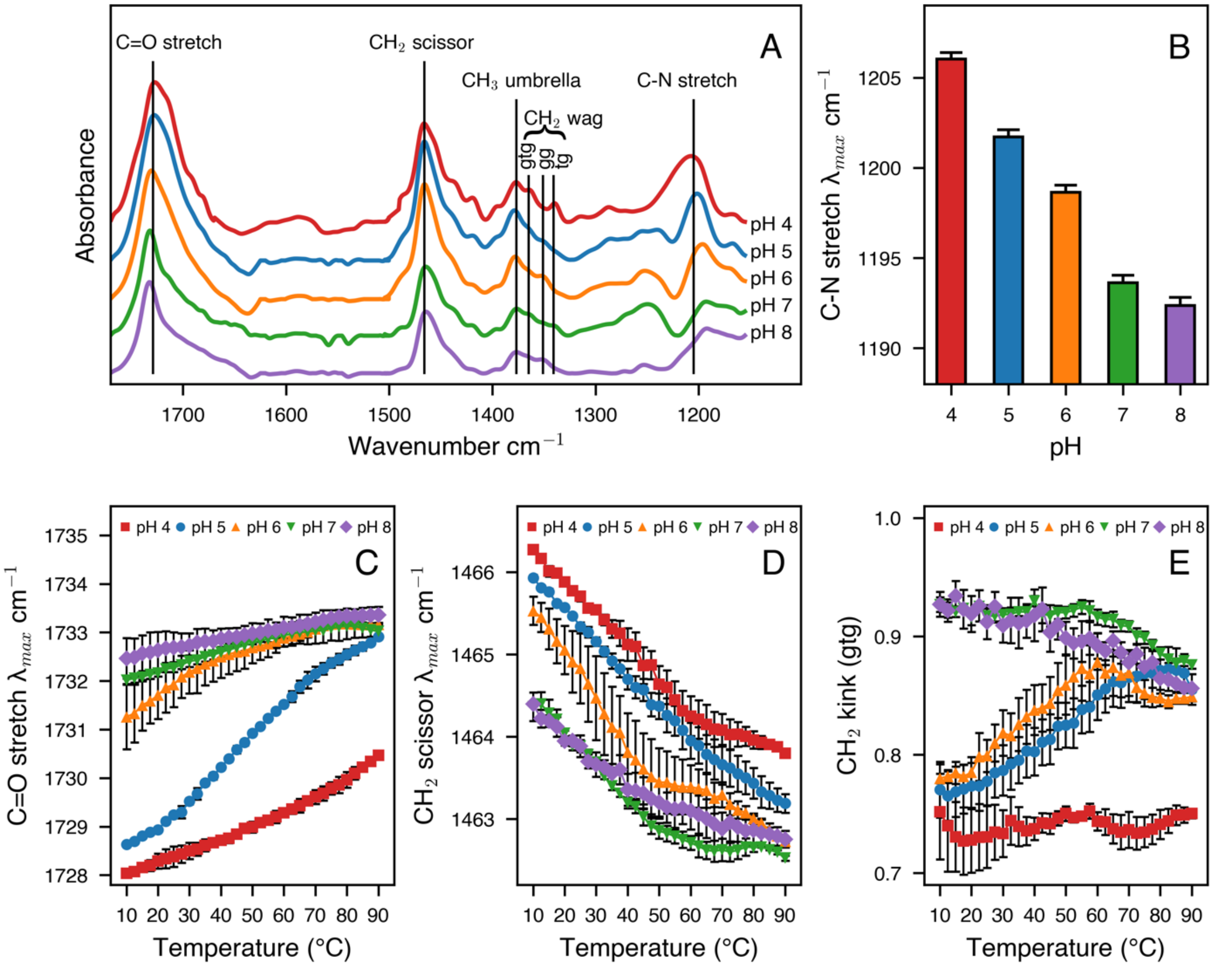


Figure 5.5. (A) FTIR spectra for mRNA–LNPs at each pH. (B) Peak position of C-N stretching peak. (C) C=O stretching vibration peak maximum of mRNA–LNPs at pH 4 to 8 from 10 to 90 °C. (D) CH2 scissor band peak maximum of mRNA–LNPs at pH 4 to 8 from 10 to 90 °C. (E) Peak intensity of the CH2 kink band normalized to the CH3 umbrella for mRNA–LNPs from pH 4 to 8 at temperatures of 10 to 90 °C.

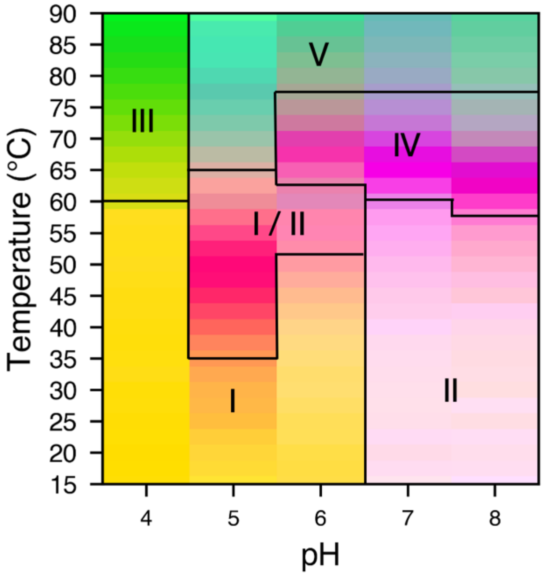


Figure 5.6. Empirical phase diagram for mRNA–LNPs with at least six distinct structural regions enumerated[15,19].

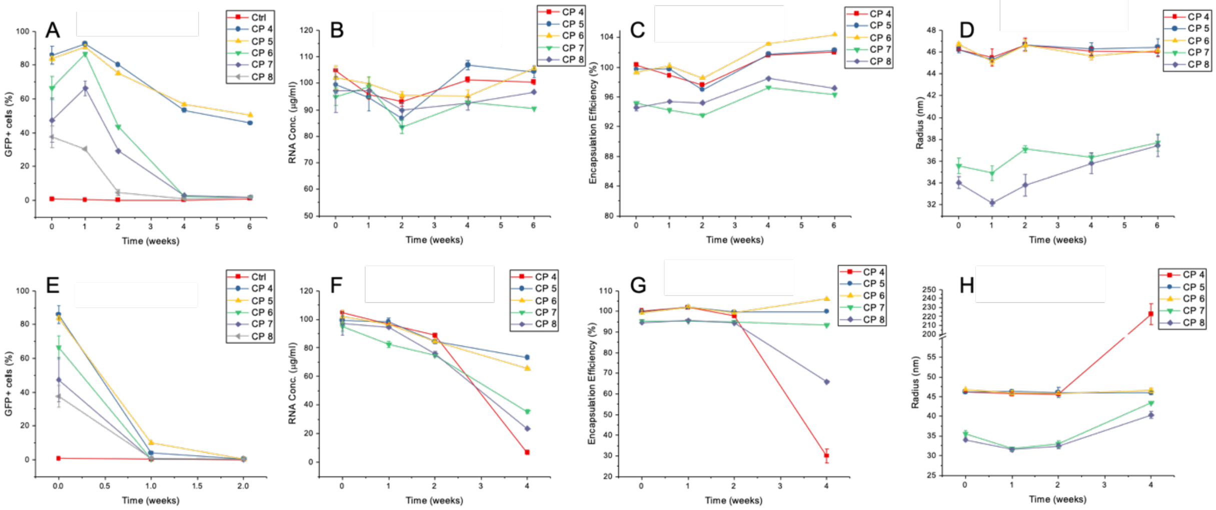


Figure 5.7. Measurement of % GFP positive cells of mRNA–LNPs during storage at (A) 4 °C and (E) 25 °C. RNA concentrations measured using ribogreen assay at of mRNA–LNPs during storage at (B) 4 °C and (F) 25 °C. Encapsulation efficiencies measured using ribogreen of mRNA–LNPs during storage at (C) 4 °C and (G) 25 °C. Hydrodynamic radius of mRNA–LNPs during storage at (D) 4 °C and (H) 25 °C.

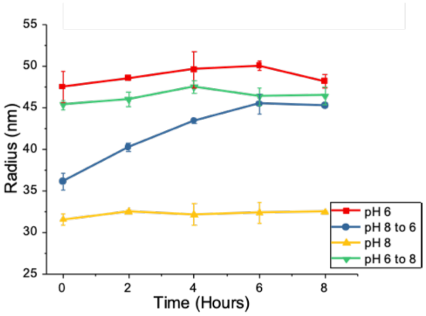


Figure 5.8. Kinetic analysis of pH-dependent size of mRNA–LNPs following dilution.

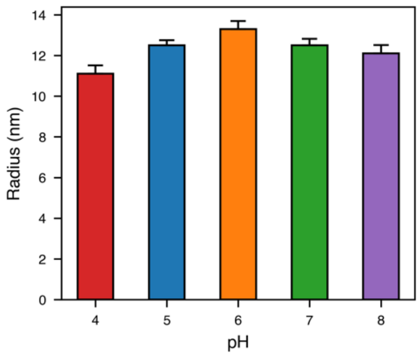


Figure 5.S1. Hydrodynamic radius of the EGFP mRNA as a function of pH.

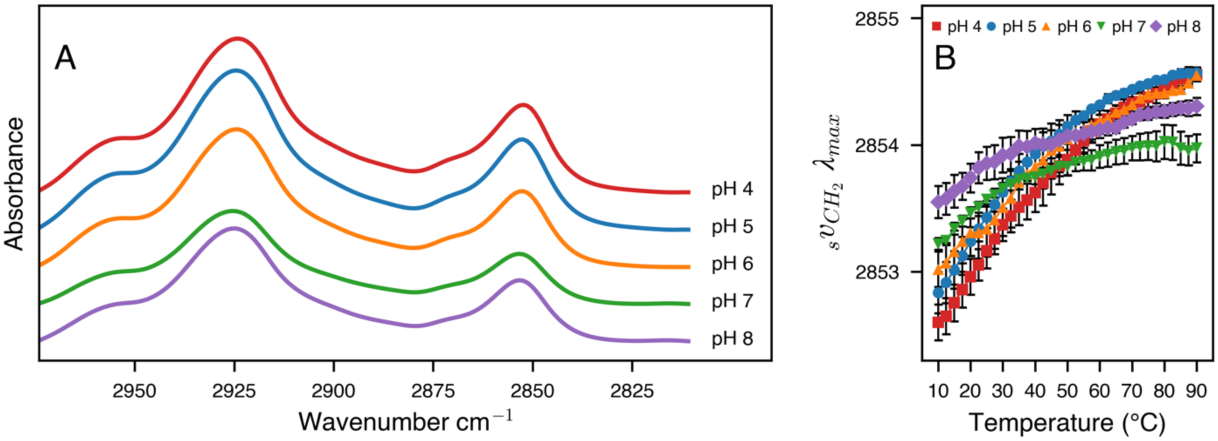


Figure 5.S2. (A) FTIR spectra of the methylene symmetric and asymmetric stretch for mRNA–LNPs from pH 4 to 8. (B) Peak position of the methylene symmetric stretching band as a function of pH and temperature for the mRNA–LNPs.

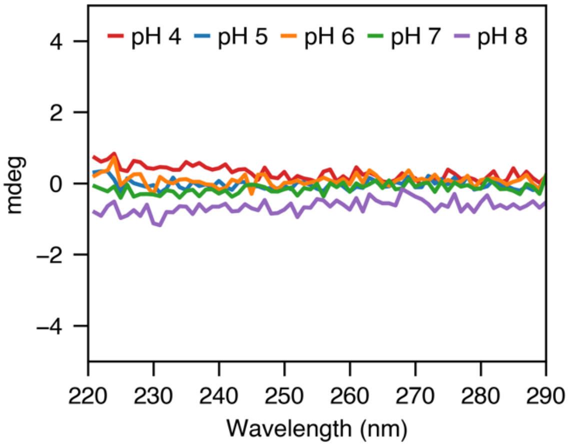
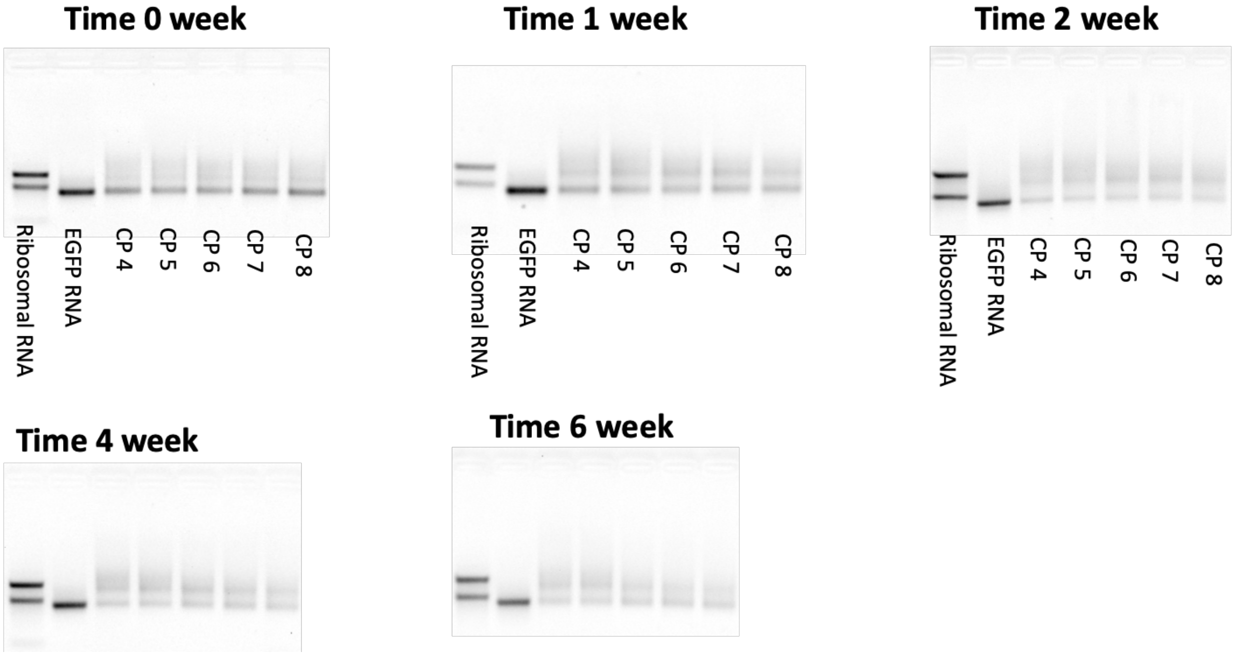


Figure 5.S3. Circular dichroism of empty LNPs as a function of pH.


Time 0 week

Time 1 week

Time 2 week

Time 4 week

Time 6 week

Figure 5.S4. Agarose gels of mRNA–LNPs as a function of time and pH during storage at 4 °C.

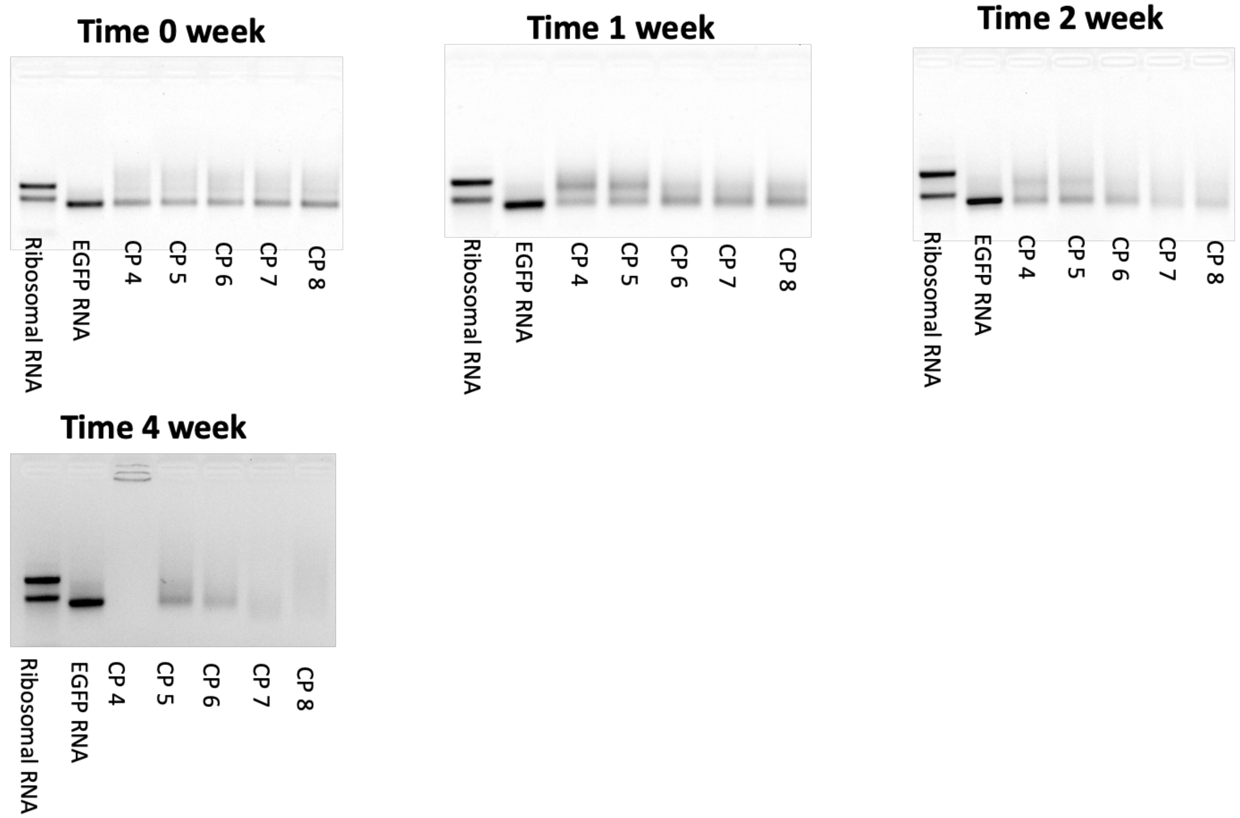


Figure 5.S5. Agarose gels of mRNA–LNPs as a function of time and pH during storage at 25 °C.

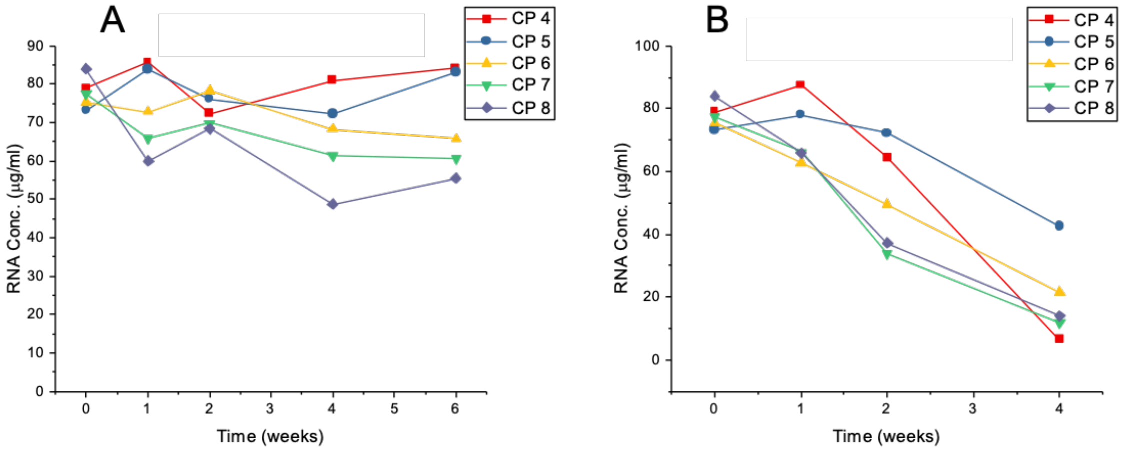


Figure 5.S6. EGFP mRNA concentration quantified by agarose gel for storage at (A) 4 °C and (B) 25 °C.

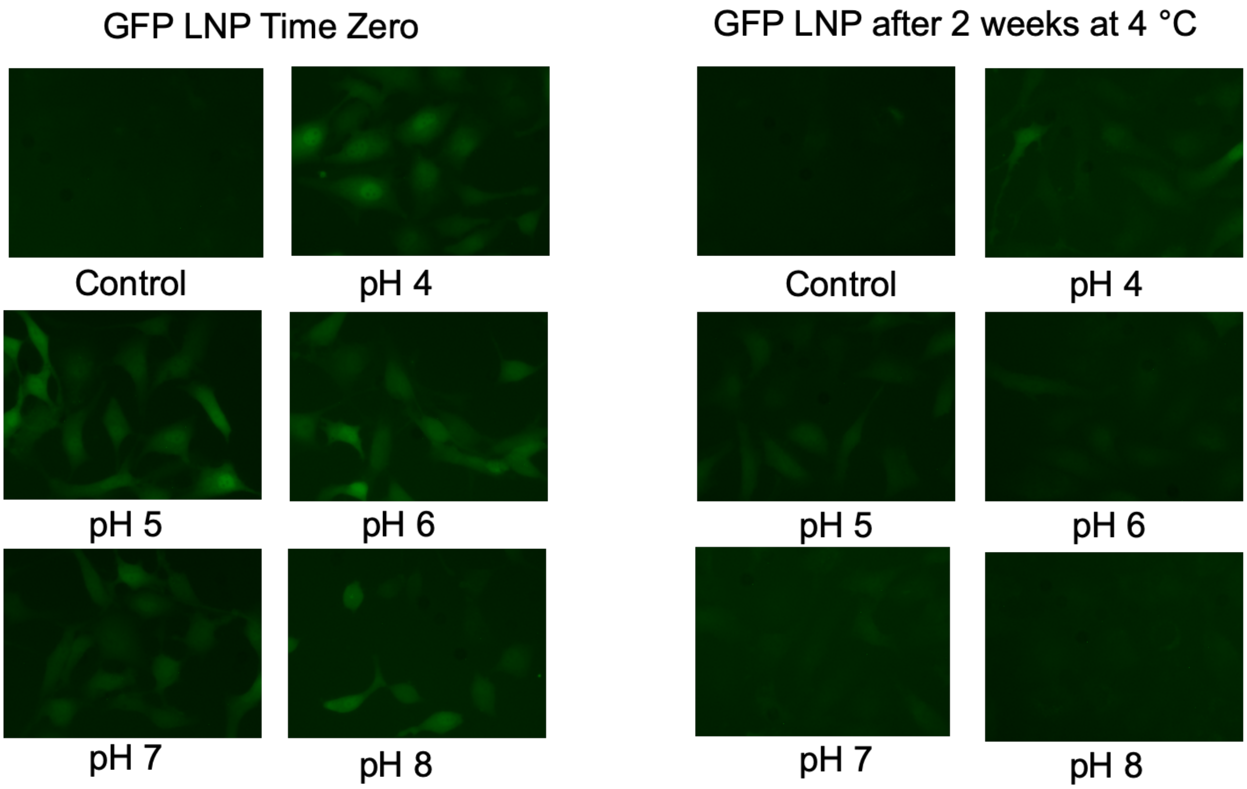


Figure 5.S7. Example fluorescence microscopy data of transfection efficiency of HeLa cells

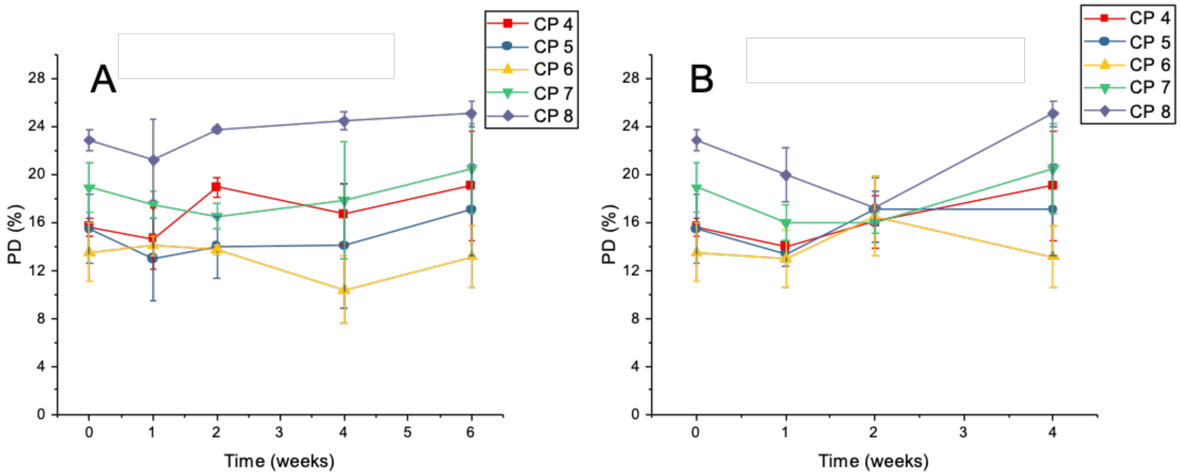


Figure 5.S8. Polydispersity of mRNA–LNPs as a function of time for storage at (A) 4 °C and (B) 25 °C.

6 Conclusion

Formulation development of turbid solutions is becoming necessary in the biopharmaceutical industry. Almost all drugs in the gene delivery field are turbid since encasing genetic material (RNA, DNA) for delivery requires particles of a fairly large size (at least 20 nm in diameter). Formulations of monoclonal antibodies are at higher and higher concentrations to achieve infrequent dosing which can hopefully be applied in the home. Vaccine adjuvants are often turbid as well, for example the most common adjuvant Alhydrogel is made up of micron sized particles. In this thesis we both adopted existing techniques and developed novel ones for characterization of turbid solutions.

In chapter two we developed a novel technique for measurement of absorbance of turbid solutions. We used an integrating cavity absorbance spectrometer to collect scattered light for quantification. We applied this to quantify the concentration of proteins in solutions with gold nanoparticles, Alhydrogel, and within polymeric microparticles. The pathlength of the integrating cavity depends on the absorbance of the solution in a nonlinear manner. The relationship between absorbance on concentration was quantified using a nonlinear fit. Overall this technique makes it possible to measure the concentration of turbid solutions directly without requiring the addition of other labels or dyes.

While we focused on measuring the concentration of protein solutions this technique could be expanded to measure the concentration of DNA and RNA in turbid solutions. We have done preliminary experiments which demonstrated this is possible for mRNA-lipid nanoparticles. However, the current incarnation of this instrument requires large sample volumes (~1 mL) of which we did not often have sufficient quantity. Miniaturizing this instrument, similar to the method used by nanodrop to miniaturize a classic cuvette-based absorbance spectrometer, could greatly benefit the widespread adoption of this technique. Furthermore, the current light source does not emit sufficient light in the deep UV. This prevents measurement of the peptide bond for concentrations of proteins low in aromatic amino acids.

In chapter three we examined how PS80 degradation effects protein aggregation. Polysorbate was degraded either by oxidation or hydrolysis and mixed with a monoclonal antibody. Mechanical agitation stability studies showed protein aggregation was significantly accelerated when mixed with hydrolyzed PS80. Aggregate particles showed a different morphology in the presence of hydrolyzed PS80. Lipid precipitation was identified as accelerating aggregation since the free fatty acids created in hydrolyzed PS80 have a relatively low solubility. Hydrolyzed PS80 also exhibited distinct surface properties with a lower rate of surface absorption and lower surface pressure at equilibrium than control or oxidized PS80.

In the future we would like to understand how oxidation of PS80 may result in protein oxidation. While PS80 oxidation may not have a large impact on protein aggregation, it is the possible creation of radicals in solution which could oxidize proteins. Oxidation of proteins can result in the loss of three-dimensional structure and thus result in loss of therapeutic efficacy. Methods for detection of PS80 hydrolysis at low levels are also worth developing since they could be early indicators of formulation instability.

In chapter four we applied a variety of characterization techniques to understand protein phase separation induced structural changes in monoclonal antibodies. We used three different monoclonal antibodies which exhibited phase separation and created phase diagrams for all of them. We showed that tryptophan is more buried in the protein-rich phase using fluorescence spectroscopy. Differences in secondary structure and the disulfide bond conformations were quantified using Raman spectroscopy and FTIR. For two of the proteins we used HX-MS to identify specific sequences responsible for protein phase separation. For one antibody oppositely charged positive and negative patches were responsible protein-self association. For another antibody a large hydrophobic region on the variable domain exhibited protection at high protein concentrations. Since protein-rich and -poor phases have the same chemical potential we hypothesized they would aggregate at the same rate. For all three antibodies, the protein-rich phase aggregated at the same rate or slower than the protein-poor phase.

In the future it would be worthwhile to compare antibodies which do not exhibit phase separation to those which do to see how the structure of antibodies depends on concentration independent of solution phase behavior. Repulsive interactions between proteins could also result in structural changes at high solution volume fractions of proteins. To develop a better understanding of the concentration effects of antibody solution structure a large panel of different antibodies would be required. Even the behavior we observed had some amount of specificity among the individual proteins examined such as the disulfide ratios seen by Raman spectroscopy, aggregation rates, and the sites of contact based on HDX studies. Another future experiment would be to use another commonly studied system such as lens crystallines to see if structural changes occur following phase separation in other types of proteins.

In chapter five we characterized the pH-dependent properties and stability of mRNA–cationic lipid nanoparticles. mRNA-LNPs showed a shift from inverse hexagonal structure below the pKa of the cationic lipid to lamellar above. To characterize the lipid phase structure, we used FTIR and DSC. To characterize the mRNA within the particle we used circular dichroism. In the inverse hexagonal phase the mRNA showed an increase in A form geometry. The particles were approximately 10 nm greater in radius below the pKa of the cationic lipid. We measured the stability of the mRNA–LNPs at 4 and 25 °C and at low pH the particles exhibited much greater stability. We attributed this to the increased organizational structure and thermodynamic stability of the inverse hexagonal phase. We also showed that following a pH-jump from 8 to 6, the particles underwent a lamellar to inverse hexagonal phase change over 6 hours but the opposite pH-jump did not result in a phase change. This suggested that the lamellar phase is kinetically trapped and less thermodynamically stable than the inverse hexagonal phase.

mRNA–LNPs are a relatively new therapeutic modality and we still have many unanswered questions following our work. Small angle scattering experiments would be useful to confirm our phase behavior assignments. Determination of how mRNA size and/or sequence effect the phase behavior of mRNA–LNPs is also an unexplored topic. Most pressing for formulation development is a more sophisticated understanding of mRNA–LNP degradation. Mass spectrometry methods must be developed both for the lipids and mRNA and specific degradation products identified. Once this has been completed, the specific impacts of various types of degradation on mRNA–LNP structure and efficacy could be assessed.

Appendix

Chapter 1, 2, and 3 have been published and are available:

1. Wei Y., Larson N.R., Hu G., Kumar P., Middaugh C.R. (2020) Chapter 8: Biophysical Characterization and the Development of Therapeutic Proteins. In: Jameel F., Skoug J., Nesbitt R. (eds) Development of Biopharmaceutical Drug-Device Products. AAPS Advances in the Pharmaceutical Sciences Series, vol 35. Springer, Cham

2. Larson N.R., Wei Y., Middaugh C.R, Label-Free, Direct Measurement of Protein Concentrations in Turbid Solutions with a UV−Visible Integrating Cavity Absorbance Spectrometer *Anal. Chem.* **2018**, 90, 4982-4986

3. Larson N.R., Wei Y. Prajapati I., et al. Comparison of polysorbate 80 hydrolysis and oxidation on the aggregation of a monoclonal antibody, *J Pharm. Sci.*, **2020**, 109, 633-639