# EXPLORING THE POTENTIAL OF PROTEIN CAGES AS MRI CONTRAST AGENTS WITH AN EMPHASIS ON PROTEIN CAGE CHARACTERIZATION BY MASS SPECTROMETRY TECHNIQUES

by

Lars Otto Liepold

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

Doctor of Philosophy

In

Chemistry

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2009

APPROVAL

of a dissertation submitted by

Lars Otto Liepold

This dissertation has been read by each member of the dissertation committee and has been found to be satisfactory regarding content, English usage, format, citation, bibliographic style, and consistency, and is ready for submission to the Division of Graduate Education.

Dr. Trevor Douglas

Approved for the Department of Chemistry

Dr. David Singel

Approved for the Division of Graduate Education

Dr. Carl A. Fox

# STATEMENT OF PERMISSION TO USE

In presenting this dissertation in partial fulfillment of the requirements for a doctoral degree at Montana State University, I agree that the Library shall make it available to borrowers under rules of the Library. I further agree that copying of this dissertation is allowable only for scholarly purposes, consistent with "fair use" as prescribed in the U.S. Copyright Law. Requests for extensive copying or reproduction of this dissertation should be referred to ProQuest Information and Learning, 300 North Zeeb Road, Ann Arbor, Michigan 48106, to whom I have granted "the exclusive right to reproduce and distribute my dissertation in and from microform along with the non-exclusive right to reproduce and distribute my abstract in any format in whole or in part."

Lars Otto Liepold

November 2009

# ACKNOWLEDGEMENTS

During my work in the Douglas/Young Lab I have been fortunate in that the time I have spent excited about my work far exceeds down times that graduate students often experience. This fortune was made possible by a number of factors namely a healthy community of colleges, family and friends. I will always be in debt to Trevor and Mark for providing an exciting lab environment, affective advice, and for permitting a flexible work schedule which made the work always seem fresh. I would like to thank the governmental agencies and the US tax payers for funding my efforts. Debbie Willits helped in many ways but I would like to thank her most for listening to my questions and for her friendship. I would like to thank Peter Suci for listening to many ideas, specifically quantitative aspects of my research, and for correcting me in how I thought about these ideas. Luke Oltrogge was great to work with both in the lab and in the mountain surrounding Bozeman. I will miss the collaboration I have with Joynal Abedin. His passion for synthesis complemented my passion of analysis resulting in a very effective melding of two skills. Mark Allen and Masaki Uchida always took the time to answer my questions and kept the lab operational and for that I would like to thank them. There are too many good friends of Bozeman and beyond that have made life outside of work enjoyable and I hope you all know I how much I appreciate your friendship. Thank you Cory and Darren for all your distracting visits. Thank you Mom and Dad for all your support. Finally, I would like state how fortunate I feel for being able to work in this truly amazing and inspiring land which continually instills a sense of exploration in me.

TABLE OF CONTENTS

TABLE OF CONTENTS- CONTINUED

TABLE OF CONTENTS- CONTINUED

# LIST OF TABLES

# LIST OF FIGURES

LIST OF FIGURES - CONTINUED

LIST OF FIGURES - CONTINUED

# ABSTRACT

Described here is the development of a protein cages as efficient and potentially relevant MRI contrast agents. Three approaches are outlined to fuse high affinity $Gd^{3+}$ chelating moieties to the surfaces of protein cages. In the first approach, a metal binding peptide has been genetically engineered into the subunit of Cowpea chlorotic mottle virus (CCMV) and to the small heat shock protein cage from *Methanococcus jannaschii* (HSP). The genetic fusion resulted in a 200x binding enhancement of $Gd^{3+}$ to CCMV in comparison with wild type CCMV and metal binding functionality was added to the HSP protein cage. In a second approach DOTA-Gd was attached to CCMV by reactions with endogenous lysine residues on the surface of the viral capsids and resulted in $r_1$ = 2,806 at 61 MHz for the 28nm diameter particle.

Directed by the results of earlier generations of protein cage based contrast agents a next generation MRI contrast agent was designed. In this work a DTPA-Gd containing polymer was grown in the interior of HSP resulting in $T_1$ particle relaxivities of 4,200$mM^{-1}$ $sec^{-1}$ for the 12nm particle. Relaxivity parameters were determined and this analysis suggests that the rotational correlation time of the $Gd^{3+}$ chelate has been optimized while the exchange life time of $Gd^{3+}$-bound water is slower than optimal. This synthetic approach holds much promise for the development of future generations of contrast agents.

Throughout the evolution of the protein cage based contrast agents there has also been and evolution of our ability to characterize these cages with mass spectrometric techniques. Specifically refined methodologies are presented for QTof characterization of protein cage at the level of amino acids, protein subunits, protein complexes and their cellular expression. Furthermore, correct charge state assignment is crucial to assigning an accurate mass to supramolecular complexes such as protein cages analyzed by electrospray mass spectrometry. Conventional charge state assignment techniques fall short of reliably and unambiguously predicting the correct charge state for many supramolecular complexes. We provide an explanation of the shortcomings of the conventional techniques and have developed a robust charge state assignment method that is applicable to all spectra.

# CHAPTER 1

# INTRODUCTION AND BACKGROUND

## Protein Cages

The work presented in this thesis covers an investigation into the structure and utility of a wide range of protein cage architectures for use as magnetic resonance imaging (MRI) contrast agents. This work includes the synthesis of a variety of protein cage based contrast agents that were designed by both chemical and genetic methods, the physical characterization of these modified protein cages by mass spectrometry, and the characterization of the relaxivity properties of these next generation imaging agents.

Protein cages are a diverse group of quaternary protein assemblies that share a common architecture and provide a diverse range of biological functions from viruses to DNA binding proteins. We use these biotemplates as the starting point in the design of novel materials with functions other than what nature had originally intended. In order for this approach to be successful, a detailed understanding of the structure – function relationships of these cages is required. We have decided to work with protein cage architectures because they provide a platform which can be altered by design using the tools of chemistry and molecular biology to impart novel functionality including catalysis for hydrogen generation, magnetic behavior for potential memory storage, drug and imaging agent encapsulation for applications in biomedical diagnostics and therapeutics.

There are common features among all protein cages including their ability to self-assemble from a distinct number of subunits to form precisely defined nano-containers in

the 5-100 nm size regime. These architectures are unique structures in that they have multiple and spatially distinct surfaces including the interior, exterior and subunit interface to which functional groups can be attached (Figure 1). Furthermore, the interior cavity is often the focal point of functional modification due to the high volume to surface area ratio that these sphere-like particles provide. Finally the atomic positions of protein cages provided by x-ray crystallography allow us to take full advantage of genetic and chemical modification such that these modifications are manifested in predictable three dimensional orientations about the protein cage.

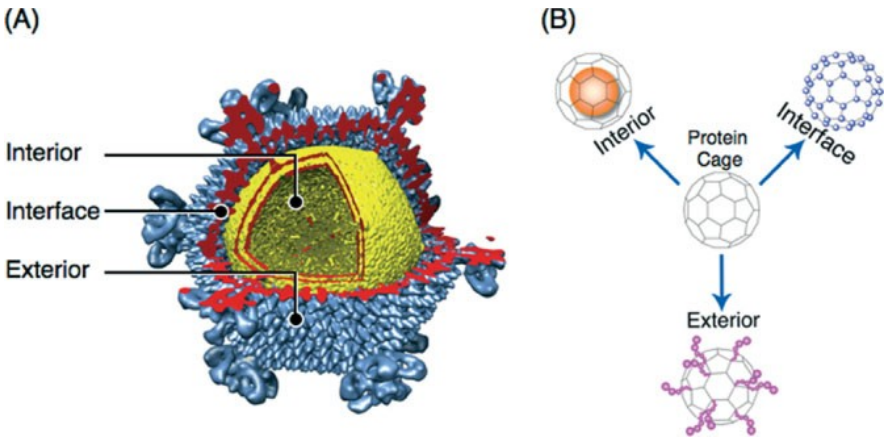



Figure 1 Surfaces of a protein cage
A) Cryo-electron microscopy imaging reconstruction of Sulfolobus turreted icosahedral virus (75 nm in diameter) isolated from a boiling, acidic environment in Yellowstone National Park. B) Schematic illustration of the three interfaces in a protein cage architecture available for chemical or genetic modification. The rendering of STIV was provided by Jonathan Hilmer.[1]

In the Douglas and Young labs we employ a library of protein cages, which support a range of synthetic chemistry and genetic manipulations useful for creating high performance imaging agents (Figure 2). The biological roles these protein cages carry

out in nature varies greatly including chaperonin functionality in the case of heat shock proteins, iron storage and DNA binding are among the functionality of Dps and ferritin protein cages, finally viral capsids can provide for us similar function which nature originally intended such as delivering material to targets inside a host organism.

Four protein cages were explored in depth in the work presented in this thesis (Figure 2). The capsid of the Cowpea chlorotic mottle virus (CCMV) was genetically and chemically modified to impart higher affinity Gadolinium (Gd) binding functionality which is important for MRI contrast.[2] A smaller and more porous protein cage, small heat shock protein from *Methanococcus jannaschii* (HSP), was also genetically modified resulting in Gd binding functionality. Furthermore, this protein cage was used to contain the growth of an internally directed branched polymer.[3] Then, a high affinity Gd chelate was attached to the HSP protein cage – branched polymer hybrid resulting in a high performance MRI contrast agent. Finally the smallest protein cage, Dps (DNA binding protein from starved cells) from both *Listeria innocua* and *Sulfolobus solfataricus*, were used as a standards for the development of mass spectrometry techniques that detect intact protein cage complexes.[4]

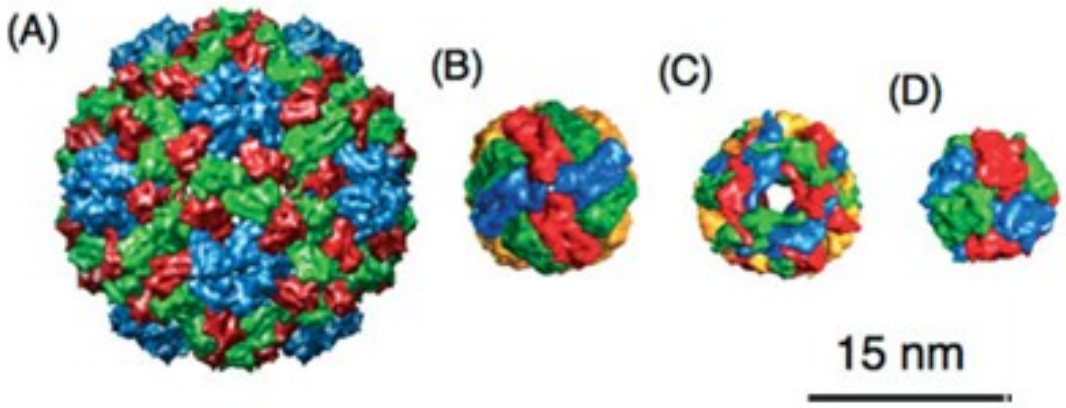



Figure 2 A Library of Cages
Space-filling images of protein-cage architectures from near atomic resolution X-ray structures. A) Cowpea chlorotic mottle virus 28 nm in diameter (1cwp), B) H chain of human ferritin 12 nm in diameter (1fha), C) small heat shock protein 12 nm in diameter (1shs), D) DNA binding protein from starved cells 9 nm in diameter (1qgh).

Cowpea Chlorotic Mottle Virus (CCMV) Background

The cavity of Cowpea chlorotic mottle virus (CCMV) is an ideal size-constrained environment where the interior surface can direct the attachment or nucleation of molecular or nanomaterials. Viruses package their viral nucleic acid within the capsid architecture and our understanding of this process has been used to direct packaging of non-viral payloads. CCMV is an RNA-containing plant virus, composed of 180 identical coat protein subunits that self-assemble to form an icosahedral cage architecture.[5, 6] CCMV is approximately 28 nm in diameter with a 24 nm diameter interior cavity (Figure 2). In the wild-type protein, the interior surface carries a high positive charge density that interacts with the negatively charged viral RNA. The positively charged interior surface has been used for directing the nucleation of mineralization reactions to form spatially constrained nanoparticles of polyoxometalate salts.[7] Subsequently, through a

combination of protein design and genetic engineering, the charge on the interior surface of the CCMV capsid has been altered from positive to negative (by glutamic acid substitution) without disrupting the ability of the CCMV cage to assemble. The interior interface of this engineered cage architecture provides a rich, highly symmetrical, and repetitive surface for directing the mineral nucleation of transition metal oxides.[8]

Modification of CCMV subunits results in multivalent presentation of up to 180 copies of any genetically engineered, or naturally occurring, functional group. These engineered functional groups can be used as attachment sites for ligand presentation on the exterior surface. It has been shown that amines, carboxylic acids, and thiol groups can react with activated small molecules. For example, using activated fluorescein (maleimide or *N*-hydroxysuccinimide (NHS)), quantitative labeling of these sites can be achieved without disrupting the overall cage architecture.[9] In addition, a small peptide of 11 amino acids has been effectively chemically linked to the exterior surface of CCMV, clearly illustrating that chemical modification is a generic approach to surface modification of the virus. [9]

Human H Ferritin (HFn) Background

Ferritins are a family of proteins found in all domains of life, which function primarily to sequester and store iron.[10, 11] These proteins are composed of 24 subunits that self-assemble to form a cage of 12 nm diameter with an interior cavity of 8 nm diameter (Fig. 1F), in which hydrated ferric oxide (or phosphate) is formed.[12] This biomineralization reaction is a multistep process involving uptake of Fe(II), oxidation to Fe(III), hydrolysis, nucleation, and crystal growth.[10-13] Historically, ferritins were the

first protein cages to be used as templates for synthesis of inorganic nanoparticles.[14-17] In a biomimetic synthesis Mann and co-workers showed that $Fe_3O_4$ nanoparticles could be artificially synthesized within empty ferritin (apo-ferritin) cages under conditions of elevated temperature and pH.[14-17] Transmission electron microscopy (TEM) revealed homogeneously sized nanoparticles commensurate in size with the interior diameter of ferritin (7.3 ± 1.4 nm).[14] Inorganic nanoparticles (other than iron oxides) can be prepared within apo-ferritin by using similar biomimetic strategies. Recently, an approach to using ferritin for molecular entrapment has been demonstrated where the cage was disassembled under low pH conditions and subsequently reassembled at near neutral pH in the presence of an magnetic resonance imaging (MRI) contrast agent, Gd-$HPDO_3A$ (gadolinium-[10-(2-hydroxypropyl)-1,4,7,10-tetraazacyclododecane-1,4,7-triacetic acid]).[18] On average about ten Gd complexes were entrapped within the cage, resulting in a complex exhibiting high relaxivity of water protons with potential applications in MRI.

The outer surface of ferritin has also been used as a template for chemical and genetic modification. Antibody conjugation and surface-charge modification to ferritin have been used as stains for transmission electron microscopy.[19] Both processes use amide bond formation in the coupling, an approach that has also been exploited for attachment of biotin and long chain alkyls to the cage. Long chain primary amines (such as nonyl amine) were chemically linked to the surface of ferritin through attachment via surface exposed carboxylic residues.[20-22] This resulted in the formation of a hydrophobic ferritin that could be transferred from aqueous solution into organic solvents without

apparent aggregation or disruption of the quaternary structure of the protein. Genetic introduction of a titanium binding peptide (RKLPDA), identified by phage display, to the N-terminus of the ferritin subunit imparts specific affinity to the engineered ferritin for titanium metal.[23] It was suggested that the specificity of binding was due to electrostatic interactions between peptide and surface groups of the titanium substrate.[24] This N-terminal extension is uniquely presented on the exterior surface of the ferritin cage and is readily altered by genetic manipulation.

Small Heat Shock Protein (HSP) Background

The small heat-shock protein (HSP) originally isolated from the hyperthermophilic archaeon *Methanococcus jannaschii* is another cage like protein that has been utilized for nanoparticle and molecular encapsulation. The 24 subunit HSP cage has an exterior diameter of 12 nm with a 6.5 nm interior cavity, and assembles with the same overall symmetry as ferritin.[25-27] HSP differs from ferritin (and Dps) in that it has large pores (3 nm diameter), which allow easy access between the interior and exterior environments of the cage (Figure. 2 C). HSP has been used as a container for controlled encapsulation and release of a chemotherapeutic agent through site directed chemical attachment of an organic drug molecule.[28, 29] A mutant of the protein (G41C), with cysteine residues introduced on the interior surface, provided a site for spatially selective attachment of the anticancer drug doxorubicin[28] It has been demonstrated that doxorubicin can be chemically attached to engineered thiols on the interior surface of HSP, giving rise to loading of 24 doxorubicin molecules per cage. The doxorubicin was attached via a hydrazone linker to the cage and shown to be released under acidic

conditions, such as those found in an endosome, illustrating a real potential for these cages as (targeted) drug-delivery vehicles[28, 30, 31] The interior of these cage architectures provides a well defined, constrained molecular container that isolates any encapsulated cargo from the external environment. However, these protein-cage architectures are not rigid, inert reaction vessels, but rather dynamic systems that allow molecular exchange between inside and outside.[32] The cage therefore acts not only as a size-constrained environment but also as a dynamic scaffold for chemical reactivity.

The exterior surface of HSP has also proven to be a versatile nanoscale platform for genetic and chemical modification.[28-30] Endogenous lysine residues (found predominantly on the exterior surface) and genetically engineered cysteines have been chemically modified for the attachment of fluorescent probes and cell-specific targeting antibodies.[28] An anti-CD4 monoclonal antibody conjugated to a fluorescently labeled HSP mutant enabled targeting of the HSP to CD4+ lymphocytes within a population of splenocytes. In addition, cancer-cell specific targeting peptides have been genetically incorporated onto the exterior of HSP and shown to effectively target tumor cells in vitro.[28] These cell-specific targeted protein cages would be useful vessels to deliver imaging and therapeutic agents to desired tissue and highlight the potential biomedical importance of these architectures.

DNA Binding Protein from Starved Cells (Dps) Background

Another member of the ferritin family of proteins are the smaller cage-like architectures found in Dps proteins (DNA binding proteins from starved cells), which are capable of similar mineral entrapment to that of ferritins.[33] Several different Dps proteins

have been discovered in both the bacterial and archeal domains of life but appear to be absent in eukaryotes[34-38] Dps protein cages are structurally similar to ferritins but comprise only 12 subunits with an exterior diameter of 9 nm and interior diameter of 5 nm (Fig. 1H).[33] While the true biological function of Dps proteins is not yet completely understood, the Dps protein cages from the bacterium *Listeria innocua* and the archaea *Sulfolobus solfataricus* have proven useful as a size constrained reaction vessels for nanomaterials synthesis.

Protein Cages as Next Generation Therapeutic Materials

Using approaches that mimic the structure-function relationships in virus particles, i.e., selective encapsulation, homogeneous size, and targeted cell delivery, creates the potential for protein cage nano-particles for imaging and therapeutic applications. Key to the success of this approach is an ability to produce and to modify protein cages to impart novel functionality. The chemical plasticity of these protein cages is an advantage not available in other nano-particle systems. The highly symmetrical, spatially controlled, presentation of functional groups has allowed us to attach organic species such as targeting peptides, fluorescent labels and inorganic species such as magnetic nano-materials and gadolinium (Gd) ions specifically to either the interior or exterior of the protein cages. A major advance has been the ability to purify milligram quantities of the protein cages expressed in bacterial, insect, plant, or yeast expression systems. These recombinant proteins have identical properties to the native proteins. There is considerable interest in the controlled formation of size constrained, nano-phase inorganic materials for medical diagnostics. In this regard the protein cage

acts as a capping layer to a magnetic nano-particle and which can be functionalized (either chemically or genetically) for attachment to cell surfaces or for tailoring particle-particle interactions. Magnetic materials associated with protein cage architectures have an inherent advantage over other polymer systems. The relative rigidity of the cage architecture lends itself to enhanced activities as a macromolecular contrast agent platform. Also, since the properties of nano-phase materials are intimately related to their dimensions, any heterogeneity in size is reflected as heterogeneity in their physical properties. A significant advantage to the use of protein cages is the ability to produce mono-dispersed nano-particles with a precisely defined size and shape. Looking forward to the development of new multifunctional biomedical materials, the successful demonstration that therapeutic agents can be incorporated into the protein cages in combination with cell targeting and MR imaging capacity will have a significant impact in medical imaging.

A lack of sensitivity is a limitation of MR however this can be directly addressed by depositing high quantities of chelated Gd groups of the surfaces and cavity of protein cages. Furthermore, the ease of chemical and genetic manipulation on the platforms provides for the engineering of multi-functional particles. Targeting, drug delivery and imaging could all be added to one particle, which is desired by clinicians as next generation therapeutics.

## The Rationale for Developing Gadolinium-Containing Contrast Agents on a Protein Cage Platform

A major focus of research presented in this thesis work has been creating MRI contrast agents with the specific interest in designing Gd-based contrast agents on protein cage platforms. The application of protein cages as a MRI contrast agents addresses fundamental issues in contrast agent development. This is true since parameters of Gd based magnetic resonance contrast that must be optimized for maximal contrast include the requirement that Gd be anchored to a large (> 4nm), rigid but porous structure and protein cages are among the best candidates to simultaneously provide these characteristics.

The library of protein cages described previously (Figure 2) provides a range of materials in which these critical parameters can be experimentally explored. For example CCMV is known to have a flexible N-terminus but the endogenous metal binding site is known to be rigid. Ferritin is extremely rigid but it has the disadvantage, with respect to Gd based relaxivity, of reduced flux of molecular species through the protein cage between the interior and exterior environments. Conversely, HSP has many large pores that allow for molecular traffic to easily access the interior of the cage. This range of size, rigidity and molecular accessibility provide us with a suite of platforms, which can be used to determine the important parameters of Gd-based relaxivity.

## The Rationale for Studying Protein Cages by Mass Spectrometry

There are two underlying research objectives of the Douglas and Young lab; the first objective is the discovery and understanding of novel protein cages in their natural setting. We are interested in understanding the structure – function relationship of these biological nano-containers and the depth of our understanding of this relationship is related to our ability to probe various details of these protein cage structures. Therefore, a suite of characterization techniques are required to yield an accurate understanding of how the structures of these cages influence their native function in biology.

The second major focus of the Douglas and Young lab, which is the design of functional materials based on protein cage templates, is heavily dependent on the ability to detect details of these engineered materials. In this area of research, the complexity of the proposed modification to the protein cage is limited by the ability to accurately detect this added complexity. Therefore, our desire to make more sophisticated protein cage based materials is satisfied, in part, by acquiring more powerful characterization tools. Among the most useful characterization tools with regard to protein cage analysis are, TEM, dynamic light scattering (DLS), size exclusion chromatography (SEC) and mass spectrometry (MS).

Mass spectrometry is a very informative analytical technique that has aided scientist in characterization of materials for over one century.[39, 40] MS analysis of protein samples has recently been made possible by advances in the "soft ionization" techniques that successfully transfer intact protein molecules, as ions into the gas phase for subsequent mass analysis inside the mass spectrometer.[41-43]

Mass spectrometry affords information that is unique when compared to other spectroscopic techniques. For the purpose of identifying the atomic or molecular species present in a sample this technique requires only the comparison of experimentally derived masses with theoretical mass calculations directly derived from chemical formula. This is in contrast to other techniques that probe physical characteristics of matter which are not necessarily uniquely associated with an atomic or molecular species. These techniques must assume that the physical property that is detected results from the presence of a particular species. Therefore the presence of a species is assumed by association between a species and detectable signal and this in contrast to MS where the signal itself can identify the chemical species present.

Since the readout of mass spectrometry is fundamentally related to chemical formula, MS yields the possibility to simultaneously detect the presence of multiple species in one sample. Therefore, given a high enough mass resolution, only in the case of isomers is the detection of two different species convoluted into one mass spectrometric peak.

For these fundamental reasons MS analysis is an effective method for accurate characterization of protein cages. There are many questions that can be addressed with MS analysis including exact mass identification of protein cage preparations. Modifications to protein cages are characterized by MS including chemical modifications to reactive functional groups (amine, carboxylic acid and sulfhydryl containing residues). The confirmation of genetic mutants can be determined by MS analysis. MS analysis of the intact cage allows insight into the distribution of modifications to the entire protein

cage assembly. Also MS is used for the discovery and understanding of protein cages in their native environment.

MS analysis of protein cages modified for MRI contrast functionality is crucial to accurately determining the end products before the contrast functionality is accessed. In this research area MS provided important information involving the characterization of branched polymers grown in the interior of protein cage and the determination of the presence of metal ions bound on chelators that have been chemically attached to the protein cage.

## Research Objectives

My work as outlined in this thesis is the development of next generation magnetic resonance contrast agents. Two important recurring questions arise in the construction of these novel imaging materials; "*what do we intend to build and why?"* and "*how do we know what we have constructed*? "

In chapters four and five, which tracks the development of these imaging materials, the question of *what do we intend to build and why*? is addressed. In these chapters a research objective has been the construction of Gd-based MR contrast agents on a protein cage platforms. This objective has expanded to determining the physical parameters that lead to the highest relaxivities possible in these systems.

The second objective of the research outlined in this thesis was to develop mass spectrometry methods to help address the second reoccurring question of; *how do we know what we have constructed*? Specifically there are two general objectives related to

this question. The first objective is: what are the most appropriate configurations of a QTof mass spectrometer that allow for; 1) the most flexibility in the type of questions that can be addressed while, 2) also providing the highest possible performance (sensitivity, accuracy, resolution) and finally, 3) a general ease of use? This objective is outlined in chapter 2. Chapter 3 addresses a specific question related to the relatively new field of mass spectrometric analysis of intact complexes such as protein cages, called noncovalent MS (NCMS). In this chapter three novel methods were tested for accurate determination of the charge state of an ion with the objective of developing a deconvolution strategy to identify the correct charge state of ions from noncovalent spectra and therefore accurate mass assignment of these complexes.

# CHAPTER 2

# QUADRUPOLE TIME OF FLIGHT ANALYSIS OF PROTEIN CAGES

## Introduction

A focus of the research presented in this thesis is adding Gd based MR contrast functionality to protein cages by chemical and genetic means. The ability to characterize these functional modifications is paramount in the development of these materials. Mass spectrometry is well suited for this task and can provide details of protein cages at the level of amino acids, protein subunits and the entire complex. Obtaining this broad range of detail requires a suite of instrument configurations with each configuration optimized for a particular mass range. Size exclusion coupled to MS (SEC-MS) for example, is a liquid chromatography / MS (LC/MS) technique that provides detail of the subunit mass of the protein. After digesting the protein cage into smaller peptide molecules, reverse phase LC/MS can be used to interrogate the state of each amino acid and can determine what type and where modifications occur. This instrumental configuration can be extended to separations of extremely complex peptide mixtures ($>10^4$ peptides) resulting in protein identifications from cellular digests. Finally, the entire native complex of protein cages can be detected by MS yielding details of protein cages which cannot be extracted from other methods and this technique is referred to as native spray MS or noncovalent MS (NCMS).

Access to a high performance mass spectrometer capable of addressing the analytical questions mention above has facilitated the research presented in this chapter.

However, the acquisition of this instrument was an experiment in itself since the desire was not to become a mass spectrometry lab but rather use this high performance instrumentation as a tool to interrogate the central themes of the lab including protein cage discovery and application. In an attempt to answer that question, a summary of methodologies for various configurations of a Quadrapole Time-of-flight liquid chromatography mass spectrometry system (QTof LC/MS) is provided below with the primary focus on the use of this instrumentation in the analysis of protein cages, their modifications and cellular expression. Furthermore, these configurations and methodologies continue to be developed and their current state reflects years of testing new ideas concerning the instrumentation setup and configurations that provide the best results. These results are judged on the basis of flexibility in the type of questions that can be addressed, performance (sensitivity, accuracy, resolution) and ease of use.

QTof Instrumentation

QTof mass spectrometers consist of a quadrupole mass filter, a collision cell and a time of flight mass analyzer. Detected ions traverse these components in this order and this instrument setup allows for the mass interrogation of a wide range of species ranging from small molecules to intact protein complexes with masses in the mega-Dalton range. While the mass resolution of the Tof analyzer is less than other mass analyzer such as ion cyclotron or Orbitrap instruments, Tof analyzers routinely produce ions with high enough resolution for accurate mass assignment of small molecules and peptides. Also, Tof systems are unique since they produce high resolution ion peaks and also have a very large mass range. Finally the QTof instrument is a tandem MS since it can use the

quadrupole to select parent ions for subsequent fragmentation in the collision cell. Then the Tof produces accurate masses for both parent and fragment ions. This feature makes Tof analyzers particularly useful in protein identification by peptide mass fingerprinting methods since accurate masses, in the low ppm range, are produced for peptide parent and fragment ions. Finally ,the Qtof mass spectrometer can be directly coupled to an high performance liquid chromatography (HPLC) system and in doing so mixtures can be separated into individual species prior to sample infusion which allows the investigator to address more experimental questions.

Electrospray Ionization (ESI)

Two commonly used MS ionization techniques result in intact protein molecular ions and these include matrix assisted laser desorption ionization (MALDI) and electrospray ionization (ESI). ESI is readily coupled to LC systems since this ionization process effectively transfers analytes from the solution phase to the gas phase. This process occurs by creating a spray of charged droplets that, through the interplay of evaporation and reduction in droplet surface charge, become increasingly smaller. At a critical droplet size analyte ions are either ejected from these small droplets or the solvent is completely evaporated from the analyte creating an ion.

As a protein molecule is transferred from the solution phase to the gas phase the individual protein molecules have a differing amount of charges and this results in an ESI protein mass spectrum with a series of ion peaks. Each peak in the spectrum differs only in the number of charges and the m/z position is determined by the equations shown below:

$$m/_Z = \frac{mass}{charge} \qquad \text{eq. 1}$$

$$m/_Z = \frac{M+n \cdot Mass_n}{n} \qquad \text{eq. 2}$$

$$m/_Z = \frac{M+H}{H} \qquad \text{eq. 3}$$

In these equations *M* is the mass of the protein, *n* is the number of charge carriers present on the protein and $Mass_n$ is the mass of the charge carrier. When hydrogen (*H*) is the charge carrier equation 3 is used to determine the peak position of ESI produced ions from proteins since the mass of hydrogen can be estimated to be equal to 1 Dalton. Figure 4B shows three protein mass spectra produced from the ESI process. In these spectra multiple ion peaks result and as is characteristic of this multiple-charging, the ions are spaced closer at low m/z (high charge state) and further apart at high m/z (low charge state).

ESI is a quantitative technique and can yield reproducible results for all detectable analytes. The ESI efficiency of a particular protein analyte is dependent on solvent and the analyte. Therefore, the instrument setup, the analyte and the solvent present during ESI are parameters that must be constant when comparing two spectra for quantitative purposes.

Another consideration for quantitative analysis of ESI spectra is that the distribution of ion peaks for a particular protein analyte can change according to the structural state of the protein. In Figure 4B the ion peak distribution for all three spectra are multi-modal. This multi-modal nature arises from the presence of more than one protein isoform. Tracking only one ion in a protein spectrum could lead to inaccurate quantitative conclusions. This is true since a change in the abundance of a single protein

ion peak might not reflect the overall concentration of the protein in solution but rather the change in the concentration of a particular isoform of the protein. Therefore comparison of solution concentrations of a particular protein from two sources is only valid if the ion peak distribution is similar for both spectra. However, ion peak distributions that are different for two spectra of the same proteins sprayed in the same solution arise from differences in the abundances of protein isoforms. Therefore, the ability to detect these isoforms makes it possible to probe details of protein tertiary structure.

SEC-MS will be discussed first, followed by tandem LC/MS experiments to determine modifications to amino acid residues contained on protein cages and for protein identifications in complex mixtures. Finally NC/MS analysis on a QTof instrument will be described.

## SEC-MS

### Instrument Setup

Size exclusion chromatography coupled to a mass spectrometer (SEC-MS) is primarily used to determine the subunit or monomer mass of protein cages. SEC-MS routinely reports exact mass identification of protein cage preparations including genetic mutant preparations. This technique can also monitor chemical modifications to reactive functional groups (amine, carboxylic acid and sulfhydryl containing residues) that reside on protein cages. SEC-MS returns masses for protein cages containing branched

polymers grown in their interiors. Finally, the presence of metal ions bound on chelators that have been chemically attached to the protein cage can be determined by SEC-MS.

An advantage of SEC-MS is that is a relatively simple and quick analysis compared to other protein cage MS analyses which will be discussed below. Protein analysis by ESI-MS often requires a transfer of the protein of interest to an ESI friendly solution and SEC-MS does this by separating the larger protein species from the smaller interfering molecules prior to MS infusion. Effectively this LC separation is an automated protein desalting procedure. Another attractive feature of this method is the long LC column lifetime, on the order of several thousand injections which makes the analysis relatively inexpensive to perform. For the reasons mention above, SEC-MS is a very effective analytical method for mass interrogation of protein cage subunits.

An important feature of this method is the type of SEC column employed. The separation media in these columns are composed of silica beads containing divots rather than the polymer matrix which is used in Fast protein liquid chromatography (FPLC) applications. The resolution provided by the silica bead type of media is lower than what can be achieved by the polymer matrix FPLC media. However, the SEC-MS column can operate at high back-pressures and this is important since the ESI source imparts a high pressure environment on the column. Also a feature that is unique to this type of column is that it can repeatedly endure harsh cleaning protocols such as injecting strong acids or bases, which is used to remove bound protein and therefore increases the column lifetime.

As MS methods are developed, the sensitivity towards analytes of interest must be considered and the two LC/MS components that are extremely influential on the sensitivity of a method are the ESI source and the LC separation technique. Two categories of ESI sources exist; a standard source which can handle higher sample infusion rates and the more sensitive nano-flow ionization source. The advantage of the standard source is that the LC flow rates that can be infused are in a regime (10μL/minute to > 1mL/minute) that allows for ease of LC method development, employment and troubleshooting. Furthermore, the spray emitted from the ESI source is very stable and rarely fails. As the flow rates are reduced below 10μL/minute the LC employment and troubleshooting typically requires more user attention when compared to higher flow rate configurations. Also, the spray needle of nano-flow sources are more likely to foul during routine operation. Another important point is that the sensitively of a LC/MS experiment has an inverse square relationship with the column diameter. As the column diameter is reduced 50% the sensitivity is increased by a factor of four and this makes the nano-flow MS methods extremely sensitive. For the reasons mentioned above there exists a trade-off in LC/MS method development between ease of use and sensitivity when choosing between a standard flow and nano-flow method.

Despite the higher sensitivity of nano-flow or nanospray ESI, standard ESI is used in SEC-MS due to the ease of operation. However, conditions were chosen to optimize the sensitivity of this standard ESI method; such as a 1mm diameter for the SEC column. This is the smallest practical column diameter that can be employed with a standard

ionization source since the maximum flow rates for this column are in the 10 to 50µL/minute regime which is at the lower limit for the standard ESI source.

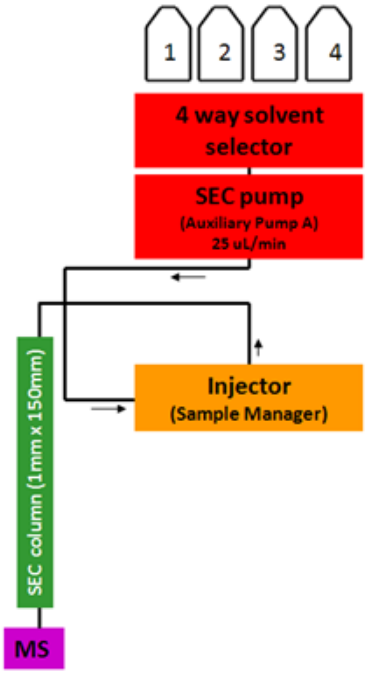



Figure 3 SEC-MS setup
The auxiliary pump is used for the isocratic gradient with flow rates of 25µL/minute. A four way solvent selector which has been manually installed before the auxiliary pump can be used to select a buffer of choice through the software.

Figure 3 shows the instrument setup for SEC-MS. The LC runs an isocratic gradient from the auxiliary pump A. Typically, a solution of 40% isopropanol and 0.1% formic acid in water is used since isopropanol is known to minimize the unfolding of proteins, which helps to maintain the integrity of the protein cage complex as it elutes through the column.

The Chromatography of SEC-MS

Despite the low resolution of the SEC column there is sufficient separation to "clean-up" and separate the protein cages from salt and buffer molecules, which are detrimental to ESI. Figure 4A shows Total Ion Chromatograms (TIC) of three separate SEC-MS injections of the protein cages: Heat Shock Protein from *Methanococcus jannaschii* (HSP), Human Ferritin (HFn) and the Cowpea chlorotic mottle virus (CCMV). The TIC plots the summation of all the ion intensities for a given scan as a function of time. Therefore these traces yield information on the total amount of ions infused into the mass spectrometer at a given time. From an examination of the traces of these three injections it becomes apparent that two major peaks are present in all injections. A protein spectrum can be produced for each of the three protein cage subunits by summing the region corresponding to the first peak and these spectra are shown in Figure 4B. Next, a specific protein ion(s) corresponding to each protein cage can be used to create Extracted Ion Chromatograms (EIC) for each subunit protein. In these chromatograms the intensity of the ion(s) or a m/z range is plotted as a function of time and therefore the region where the subunits were infused into the mass spectrometer can be precisely determined and these chromatograms are shown in Figure 4C. The second peak in the three injections contains small molecules and EIC are shown of the m/z range containing these small molecules (m/z = 50 to 600, Figure 4C). The ability of the column to separate interfering small molecules from the protein cage subunits prior to ESI ionization can be determined by looking at the EIC of the subunits and the small molecules.

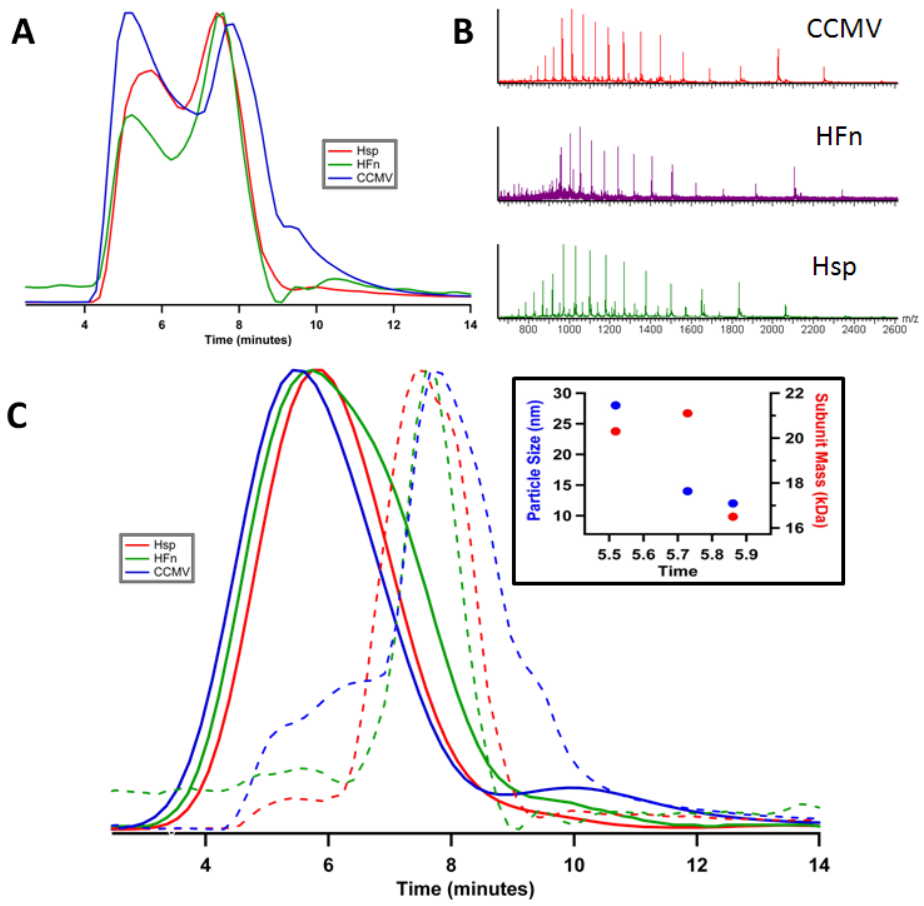



Figure 4 SEC-MS of protein cages
(A) TIC of HSP (red), HFn (green) and CCMV (blue), (B) spectra of the three protein cages and (C) EIC of HSP (red), HFn (green) and CCMV (blue) ions and EIC for salt and buffer molecules for each injection are shown as dashed lines. Inset the retention time (peak top) dependence on particle size and subunit mass.

It is worth mentioning that the retention times, or time at the top of each chromatographic peak, are slightly different for each protein cage (Figure 4C) and this could result from separation by the size of the cage if the cages elute as intact complex or by the size of the subunit protein if the complex falls apart during the SEC separation. Inset in Figure 4C is the dependence of retention time for the three cages on particle size

and subunit mass. Since the plot of particle size vs. retention time produces a more continuous trend compared to subunit mass vs. retention time it is more likely that these cages elute as intact complexes and are subsequently broken into subunits at the ion source. Due to non-specific interactions between the protein cage and the column, larger complexes do not always elute earlier than smaller complexes, however this is generally the case. Dynamic light scattering (DLS) or transmission electron microscopy (TEM) analysis of the sample collected as it elutes of the SEC column can be performed to verify the quaternary nature of these cages when eluted in a particular buffer. It is known that the column itself does not disrupt the structure of protein cages since SEC-noncovalent MS (SEC-NCMS) analysis in volatile buffers such as ammonium acetates, ammonium bicarbonate, triethylammonium acetate and triethylammonium bicarbonate all yield intact cages as determined by MS.

The low resolution of the SEC limits the use of these columns as an analytical separation tool. However, an advantage of low chromatographic resolution is that all proteins elute at a similar time and therefore undergo the ESI process in a solution that has a consistent make-up. When making quantitative estimates of unlabeled and labeled subunit proteins, similar elution times become an advantage since the efficiency of ESI is highly dependent on the solvent. This is contrary to reverse phase separations where slight differences in the protein molecule can result in complete separation of the two molecular species. In this latter case, the species are ionized in different solutions (different ratio of aqueous to organic solvents) which can lead to variations in ionization efficiencies which further convolutes the inherent flight efficiency differences of the two

species. Another practical advantage of the low chromatographic resolution of SEC-MS is that the protein always elutes in a predicted region and therefore the investigator can easily find the protein of interest.

An extreme example of how robust this method is at desalting a "dirty" sample is shown in the detection of the CCMV cage in rabbit plasma. The CCMV samples, in this experiment, were prepared by spiking the protein cage into a undiluted solution of rabbit plasma to mimic an *in vivo* CCMV detection experiment. Figure 5 is a plot of four summed ions vs. CCMV concentration, from an ESI spectrum of CCMV, that were successfully detected after separation from a rabbit plasma matrix. Here the extremely high salt concentration and other interfering molecules were adequately separated from the protein cage. The equilibration time after a plasma injection is longer than what is observed after a typical sample injection however the column does recover in a relatively short period and is stable after multiple plasma injections. Furthermore, the CCMV subunit response is almost linear and could be used to make quantitative inferences regarding the concentration of CCMV subunit in a plasma matrix. The ability to accept sample matrixes such as plasma is another advantage of SEC separations over reverse phase techniques.

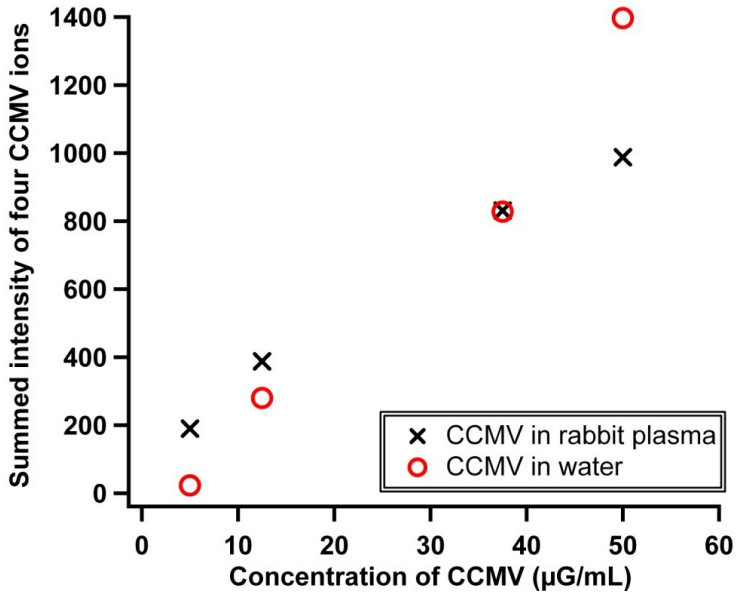



Figure 5 SEC-MS of CCMV in rabbit plasma
The summed peak areas of 4 protein ions from a CCMV subunit mass spectrum in water and rabbit plasma vs. the CCMV concentration in the injected sample. The onset of the CCMV chromatographic peak was averaged to produce the mass spectra. Included in the plot is a calibration curve of CCMV in water. This shows that there is not a drastic change in the ionization efficiency of the CCMV ions due to the plasma matrix.

Deconvolution of SEC-MS Mass Spectra and Monitoring Protein Labeling

Raw mass spectra of protein cage subunits must be deconvoluted to obtain mass information. This process in initiated by combing the appropriate, individual spectra. The appropriate spectra can be determined by making an EIC, as outlined above (Figure 4), to show precisely where the ions of interest elute and this process is also shown in Figure 6. Once the appropriate spectra are combined there are two methods that can be used to produce a mass from the multiple charged ion series. The first and more traditional technique, which will be elaborated on in Chapter 3, uses a relatively simple algebraic relationship between m/z values of two adjacent ion peaks to calculate the

charge state of each ion peak. Once the charge state is known for the ion peaks, a protein mass can be calculated. A mass value can be calculated for each ion peak. Modern software platforms provide this deconvolution method and these calculations are performed automatically, however sometimes a manually peak picking step is required. Another strategy uses the MaxEnt deconvolution method and this technique is almost completely automatic with only a few manual parameters that are required prior to deconvolution.[44] This method is preferred if the user would like to estimate the relative ratios of modified to unmodified subunits since the MaxEnt algorithm takes into account the quantitative data contained in the raw protein mass spectra.

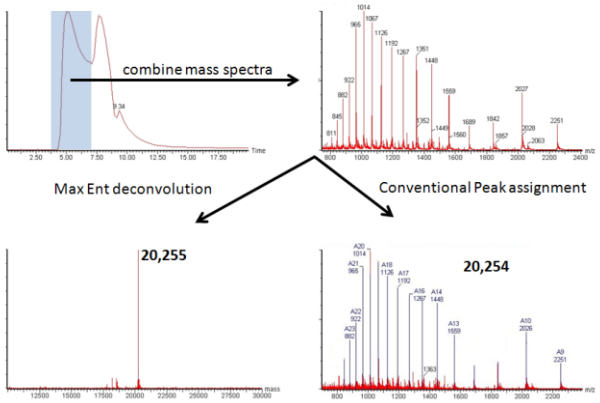



Figure 6 Deconvolution strategies
First, individual mass spectra are combined. Then a conventional or MaxEnt deconvolution step is used to determine the mass. In this example the conventional method determine the correct mass for the wild type CCMV subunit (calculated mass = 20,254Da) while the MaxEnt was 1Da above the correct mass. The charge state of the ion peaks are listed above each peak in the spectrum corresponding to the conventional method.

An example of using SEC-MS with MaxEnt deconvolution to estimate the relative abundance of modifications to protein cages is shown in Figure 7. Here a branched polymer was grown in the interior of a protein cage and a series of four

successive reactions were successfully characterized.[3] It is important to note that none of these reactions result in 100% product yield or the product results in multiple molecular species. Both of these reaction outcomes have the effect of decreasing the molar concentration of the molecular species present. So, these reactions not only make the sample more complex but they also decrease the concentration of the molecular species. Although the quantitative information from these analyses is an estimation, since various molecular species have different flight efficiencies, the information gained is critical to characterization of these complex synthetic products.

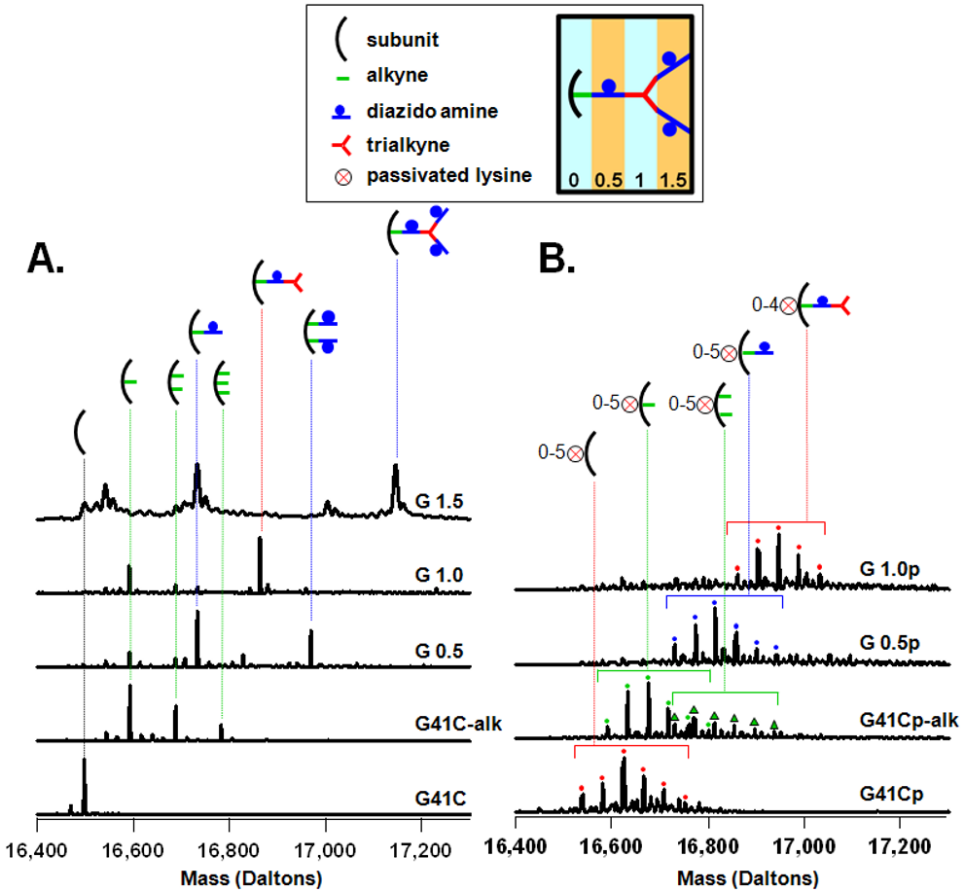



Figure 7 SEC-MS of growth of branched polymer in HSP
Characterization of the first steps in synthesis using LC/MS to determine added mass to protein cage subunits (A,B). Deconvoluted mass spectra of the non-passivated (A) and passivated (B) preparations. The horizontal brackets in B indicate a group of subunits bound to the same click reaction product, but containing 5 to 6 amines passivated with the acetyl group.[3]

For less complex labeling characterization, the flight efficiency differences between various molecular species can be accurately determined and therefore accurate quantification of labeling is possible by SEC-MS. This analysis is initiated by creating a LC/MS calibration curve of unlabeled subunit by standard SEC-MS methods. Next an independent protein quantitative method such as a Bradford or BCA analysis is performed on a labeled protein mixture. The reaction conditions used to create the

mixture should ideally be adjusted to result in a mixture with both the presence of mono-labeled and unlabeled species. Then the flight efficiency can be calculated by determining the molar concentration of the labeled species. This is achieved by subtracting the unlabeled protein concentration (determined by the LC/MS calibration curve) from the total protein concentration (determined by the Bradford analysis of the mixture). The flight efficiency of the two species can be compared since the unlabeled species relationship between LC/MS response and concentration is equal to the slope of the calibration curve. This same relationship for the labeled species can be calculated from the LC/MS analysis of the mixture, containing the labeled species response, which can be divided by the concentration of this species and calculated as outlined above. Mixtures with more than one label per subunit could be calculated in a similar manner but this would be increasingly convoluted and would require more assumptions. However the determination of the flight efficiency difference of a mono-labeled subunit is straightforward.

### Insight into the Degree of Protein Unfolding and Quaternary Structure Reveled by SEC-MS

Finally, SEC-MS can provide insight into the degree of protein unfolding. Consider two isoforms of the same protein where one isoform is more unfolded compared to the other. The more unfolded isoform has a higher amount of exposed charged residues resulting in a higher charge state therefore a lower m/z value. Therefore spectra containing a higher abundance of highly charged ions (lower m/z) are thought to come from protein isoforms that are more unfolded than ions with lower charge states.

Often the ion distribution for a protein is bi- or multi-modal suggesting various stable isoforms states. This effect is shown in an example of a monoclonal antibody incubated in an increasingly reducing environment. Figure 8A shows the MaxEnt deconvolution of the 1mM, 20minute sample indicating the presence of a tetramer and trimer majority and a lower abundance of dimers and monomers. As the environment becomes more reducing there is a shift from tetramers to monomers with both the presence of trimers and dimers in the intermediate reducing environment regime. At the highest DTT concentration (100mM) there are no hetrotetramer ions remaining and the heavy subunit is highly charged indicating its more unfolded nature. However the light subunit has a bimodal ion distribution suggesting that it is present as two isoforms (Figure 8A).

As the reducing environment is increased there is an overall trend towards lower m/z ions, which is a result of both a break-up of tertiary structure and the unfolding of protein resulting in higher charge states. This trend is plotted in Figure 8C by tracing three regions of the spectrum over the changing reducing environment. Also, the ratio of highly charged ions to lower charged ions for the light chain increases with increasing reducing environment indicating a change in the relative abundance of the two isoforms (Figure 8D). It should be noted that a multimodal ions series could result from a distribution isoforms developing inside the mass spectrometer, however in this case (Figure 8) the isoforms are dependent on the reducing environment and therefore the spectra are likely to reflect the isoform state in solution. In this example both the composition of the complex (tetramer, trimer, dimer or monomer, Figure 8C) and the

degree of protein subunit unfolding (Figure 8D) can be obtained from SEC-MS analysis. This example shows the ability of SEC-MS to provide structural information of proteins.

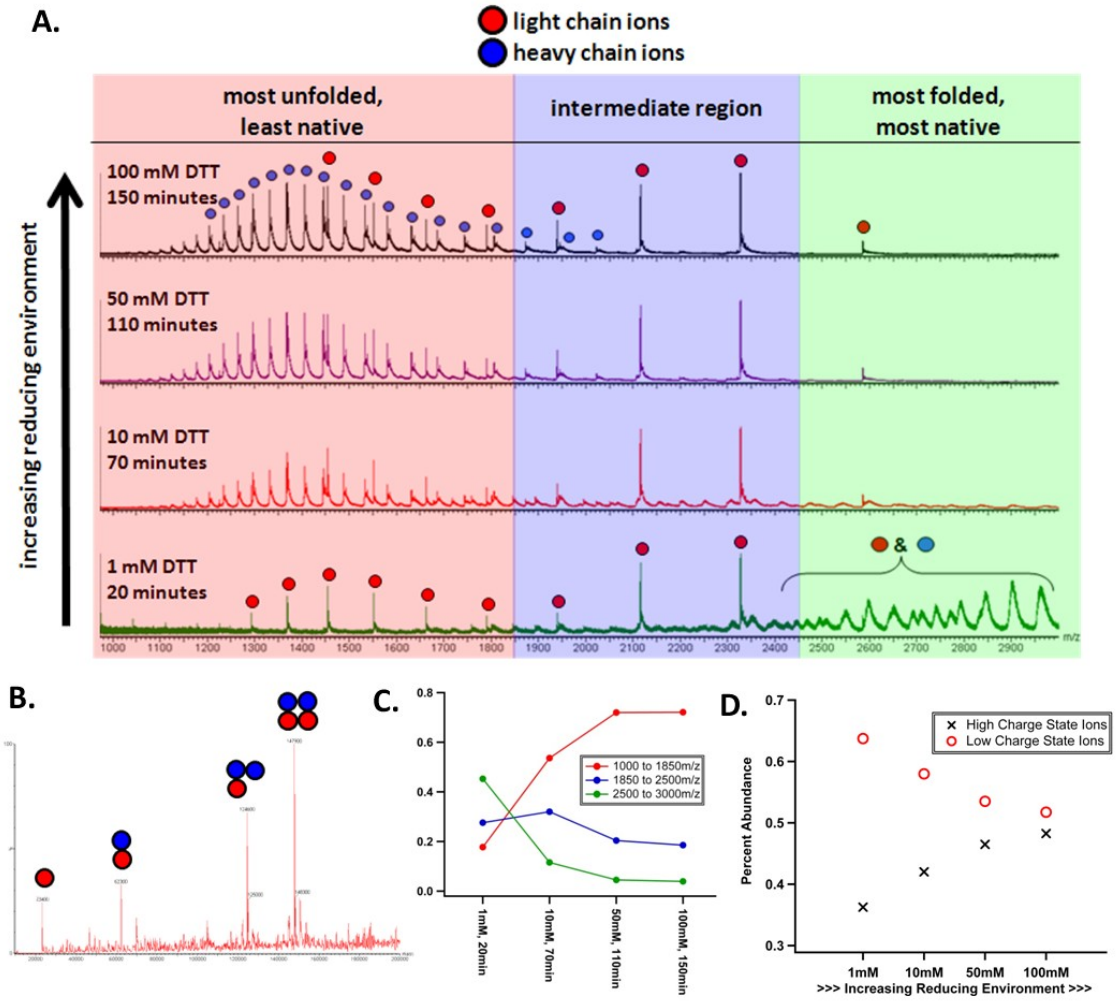



Figure 8 SEC-MS to gain insight into the degree of protein folding
(A) Four combined spectra of IgG that were produce by SEC-MS analysis and were incubated in an increasing reducing environment. Heavy and Light chain ions are labeled. (B) MaxEnt deconvolution of the data corresponding to the least reducing environment. (C) A plot of the three region of the combined spectra vs. reducing environment. (D) A plot of the percent abundance of the highly charged and lower charge state ions vs. reducing environment.

## Nano Reverse Phase LC/MS to Gain Details of Modifications to the Amino acid of Protein Cages

Often SEC-MS or other analytical techniques cannot address questions related to the state of a particular amino acid. However nano reverse phase LC/MS is capable of answering these questions and therefore is critical to the in depth characterization of modifications to protein cages. Reverse phase LC/MS analysis of parent and fragment ions of peptide molecules, which result from proteolytic digestion of proteins, is generally used for addressing two questions; 1) what proteins are present in a complex protein mixture as identified through a characteristic "fingerprint" for each protein or 2) what and where are the modified amino acids of a particular protein? Analysis of proteins at the amino acid level will be discussed first followed by protein identification and quantitation of complex mixtures.

### Instrument Setup

The LC configurations used in most experiments that focus on protein identification or amino acid modification consist of a low flow rates typically between 100nL/min to 1000nL/min. These low flow rates are a necessary parameter that allows for the coupling of the more sensitive nanospray ionization source to the LC. A sensitive technique is critical in this analysis since a high level of peptide coverage is desired. Therefore detection of all peptides including those peptides with low ESI flight efficiency is important. Another important feature of the setup is the ability to "wash-off" buffer and salt molecules while retaining the peptides on the column. By removing the salt before ionization the ESI efficiency is drastically improved. This washing step is

performed by loading the sample which contains peptides, salt and buffer molecules onto the trapping column. This step is performed at relatively high flow rates of 20µL/minute and the solvent is directed to the waste as shown in Figure 9A. After a few minutes of washing the flow rate is decreased to 900nL/min and then directed to the analytical column for separation and MS detection (Figure 9B).

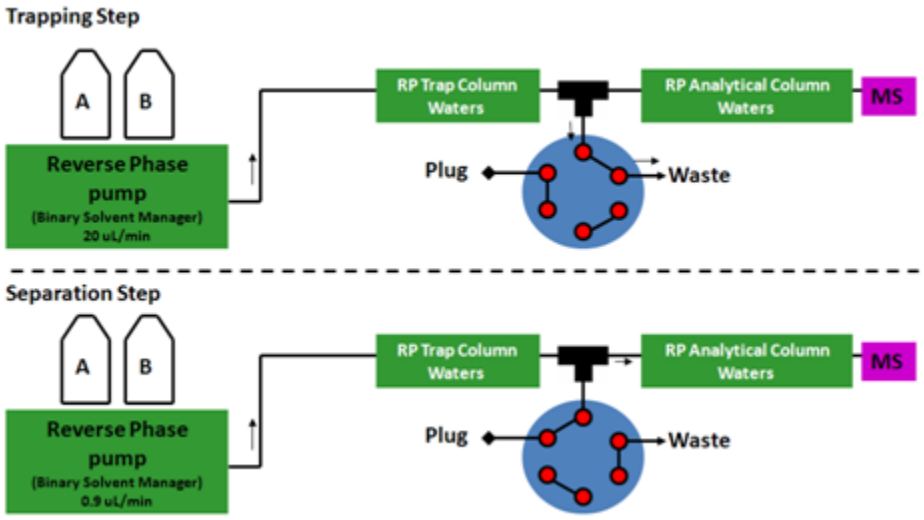



Figure 9 nanoLC/MS setup
Setup for nanoLC/MS showing the two configurations, the trapping step that washes off buffer and salt molecules. Then the separation and subsequent MS analysis step is shown.

Overview of Tandem LC/MS Experiment

"Shotgun proteomics" is a common approach in LC/MS configurations and consists of simultaneously digesting all proteins in a mixture and then the peptide mixture is separated and analyzed by LC/MS techniques.[45] It may seem counter intuitive that prior to MS analysis complex protein samples are made over an order of magnitude more complex by proteolytic digestion to yield a mixture of peptides. However, peptides are

desired over proteins for MS analysis since the fragments produced from peptides are more informative than those from proteins. Fragment ions of peptides produced from collision induced dissociation (CID) are useful since they are produced in a predictable manner as shown in Figure 10 and due to their predictable nature theoretical fragments can be generated for a given peptide sequence. Protein identifications are produced by comparing experimentally determined parent and fragment masses to a list of theoretically generated parent and fragment masses which are produced from a user defined protein database.

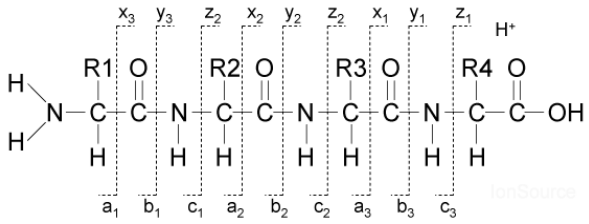



Figure 10 CID of peptides
The peptide fragmentation pattern as a result of CID. Typically y and b type ions are most abundant in CID spectra from QTof mass spectrometers.

Scan Types for Tandem MS

Tandem mass spectrometers are capable of performing MS/MS experiments which consists of isolating a single parent ion, then collision induced dissociation (CID) is performed on the parent ion and the fragment ions are then analyzed by a mass analyzer. Three types of tandem MS or MS/MS scans will be discussed here for generation of parent/fragment masses. In the QTof instrument the quadrupole is used to isolate the parent ions prior to CID in the collision cell and subsequent mass analysis is

carried out in the time of flight (TOF) mass analyzer. In a directed MS/MS scan the user defines what parent ion to isolate and the quadrupole will only pass this mass. This scan type results in high signal to noise ratio for the fragment ions since 100% of the duty cycle is devoted to fragmenting the ion of interest. As mentioned above multiple scans can be combined in a single acquisition which provides the user with the ability to monitor multiple ions. The analysis of multiple directed MS/MS experiments is referred to as multiple reaction monitoring (MRM) and is used in screening newborns for metabolite abnormalities.[46] In these experiments each tandem MS scan tracks an individual metabolite. As more scans are added into one scanning protocol the sensitivity is reduced and there becomes a point if too many ions are set to be monitored it becomes likely that some ions will be completely missed since the chromatographic peak width is smaller than the time it takes to scan all the ions of interests. Data dependent MS/MS and $MS^E$ scans have been developed to maximize the number of peptide fragmentations in a given period of time.

Data dependent scans are the most commonly used scan for shotgun proteomics. Here the mass spectrometer computer automatically chooses which ions to fragment based on user-defined criteria such as the abundance and charge state of an ion. In this scan it is important for the mass spectrometer to spend a minimum amount of time acquiring fragmentation data but the quadrupole must be locked on a particular parent ion long enough to obtain adequate signal to noise for the fragment ions and this parameter can be controlled by user defined settings. Although this is the most common choice for protein identification, the entire parent ion chromatographic peak is not collected.

Therefore the ability to perform quantitative analysis is lost. However, there are pre-analysis, peptide labeling methods that allow for robust relative quatitation of two sample types with data dependent acquisitions.[47]

LC/MS$^E$ is a scan that was developed to increase the number of peptides that are detected and fragmented.[48] Also this technique allows for label free quantitative analysis. This scanning protocol cycles through two scan types; a low energy scan that collects all the parent ions and then a high energy scan than simultaneously collects all the fragment ions. A major disadvantage of this scan is that the fragment ions are decoupled from the parent ions. So a trick must be employed to match the parents with its fragment ions and this is performed by matching the two ion types based on similar retention times. This disadvantage must be weighted with the advantages offered by LC/MS$^E$ such as an improved duty cycle such that many more peptide parent ions are collected. Another advantage is that the entire peak width is collected which allows for relative quantitation of two sample types along with the ability to estimate the absolute concentration of proteins in a mixture and this will be discussed later.

The work flow for analysis of tandem MS data usually includes automated spectral processing. When scanning in data dependent MS mode the processing software matches theoretical parent/fragment masses with experimental parent/fragment for protein identification. An addition step is required in the LC/MS$^E$ scanning mode since the parent masses and fragment masses are not directly coupled. This done by making EIC of both low energy parent ions and high energy fragment ions. Then parents and fragments are matched based on the similarity of their chromatographic peaks. When

parent masses are matched with fragments they are referred to as "accurate mass retention times" (AMRT). This matching is not as robust as standard MS/MS experiments and therefore multiple injections are required to improve statistics of the matching process.

As samples become complex as is the case for proteolytic digest of the entire population of cellular proteins, the majority of the processing is automated. In this scenario the task of the task is to determine the appropriate processing software parameters such that the protein identifications are accurate. A more manual approach produces the best results when looking for specific details of amino acid modifications in less complex samples.

Searching for Gd-Containing Peptides in CCMV-DOTA-Gd Construct

An example of using $MS^E$ and directed MS/MS is outlined below and involves identifying lysine residues in CCMV that were chemically modified with a Gd chelate (Chapter 4). After the modification of CCMV was complete, SEC-MS was performed to determine the extent of labeling (Figure 33). To gain information on the labeling sites a nano reverse phase LC/MS approach was undertaken. After a trypsin digest of the Gd labeled CCMV, an $LC/MS^E$ experiment was performed since this technique would provide the best chance of detection all peptides. After $LC/MS^E$ analysis, manual inspection of a combined spectrum of the entire twenty minute run revealed the presence of five candidate peptides that may contain the Gd chelate (Figure 11A). The candidate peptides were manually found by looking for the unique isotope distribution of Gd. Next EIC are made of each candidate ion to determine when they elute off the column (Figure

11B). Directed MS/MS scans were then carried out for each ions as shown in Figure 11C and from these analyses, two peptides that contain the Gd chelate were identified. Furthermore the position of the modified amino acid was determined by the fragment ion data (Figure 11 D). Currently no software platforms exist that can perform these steps automatically therefore it is important that the user has the basic knowledge of the steps outlined above.

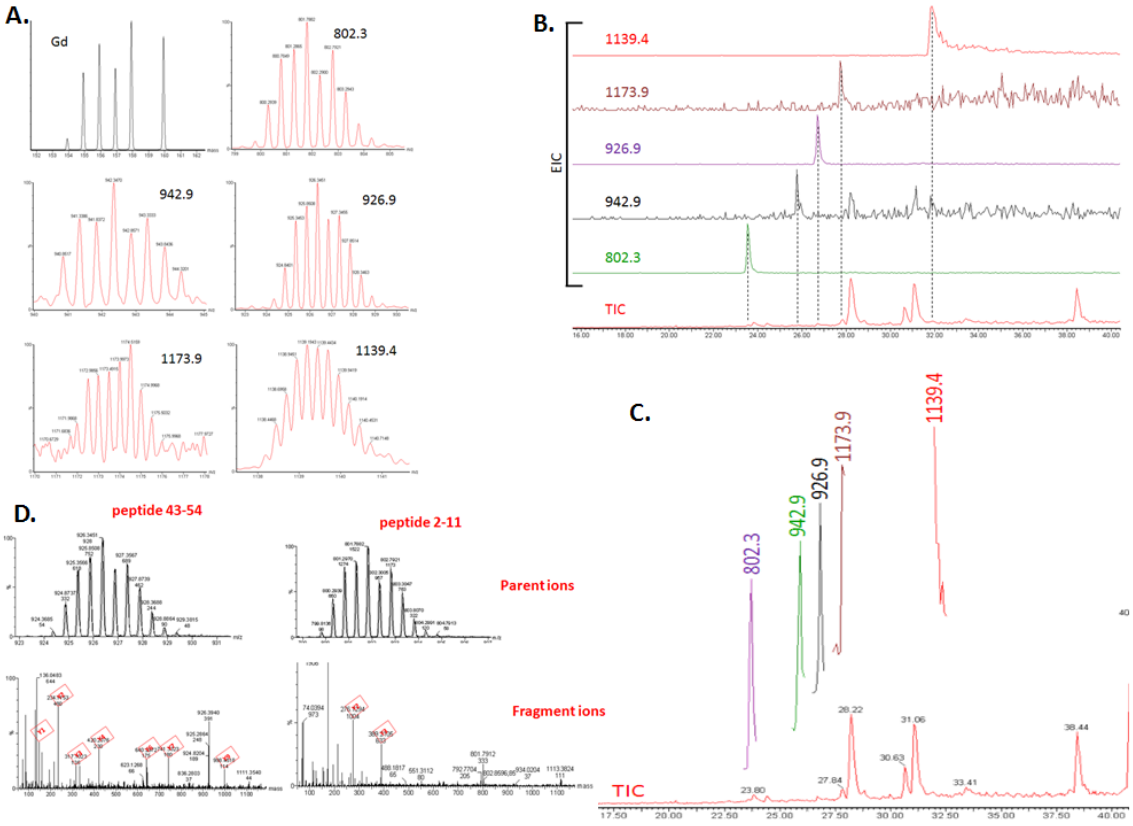



Figure 11 Determination of Gd-chelate containing peptides by LC/MS
(A) Isotope distribution of Gd and five candidate ions that may contain Gd from a trypsin digest of CCMV sample labeled with a Gd chelate. (B) LC/MS$^{E}$ analysis of a this sample showing the EIC of ions shown in (A). (C) EIC of a directed LC/MS/MS analysis. (D) Resulting MS/MS data from the analysis shown in (C) for two ions.

## NanoLC/MS$^{E}$ for Protein ID and Quantitation in Complex Mixtures

When engineering additional functionality to protein cages we are oftentimes inspired by their native function. Therefore, it is helpful to have tools to address questions related to their native function and among the simplest questions asked is; when are these protein cages expressed and what other proteins are they expressed with and these questions pertain to the field referred to as proteomics? With these questions in mind we were interested in developing a tool to quantitatively monitor protein expression on a global scale. To perform these experiments additional separation capacity is needed, with regard to the nano reverse phase setup mentioned above, since there is a limit of how many ions can be successfully analyzed by the MS at any given moment.

### Instrument Setup

One way to reduce the number of peptides that are infused into the instrument is to load the total peptide population onto an ion exchange column and then elute partitions of the total peptide population onto the reverse phase trapping column for subsequent washing, reverse phase separation and MS analysis. This step is then repeated until all the peptides are eluted off the ion exchange column. This can be performed by placing an ion exchange column in series before the trapping column. The method is referred to as two dimensional (2D) LC/MS and is shown in the Figure 12 A-C.[49] The majority of the peptides will be trapped on the ion exchange peptide and can be eluted off is a stepwise fashion by running increasing concentration of volatile salts over the ion

exchange column and which will partition the peptide population onto the trapping column for subsequent LC/MS analysis.

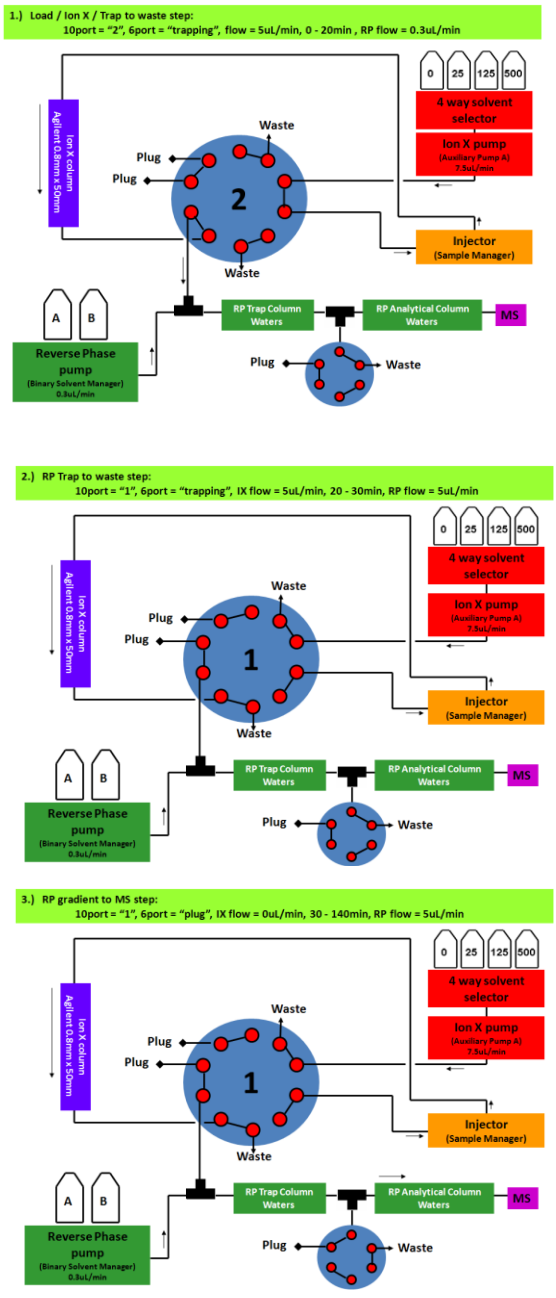



Figure 12 2D LC/MS setup
Setup for the analysis of a trypsin digest of a cellular sample by 2D nanoLC/MS$^{E}$ methods. Three steps are shown; the ion exchange step, the reverse phased trapping and the reverse phase and MS analysis step.

Example of 2D-LC/MS

The conventional method for data collection in LC/MS based proteomic experiments is to save each reverse phase run as a single file. The disadvantage of this is that peptides from a single protein will be spread out over several salt steps and therefore several data files which will each be individually processed. The LC and MS method files can be configured to run several salt steps and collect this in a single data file. Figure 13 shows one file that contains 3 salt steps and plots the chromatogram for the reverse phase gradient as percent acetonitrile, the pressure of the reverse phase pumps, the TIC and the AMRTs that we determined by the processing software. In Figure 13 the peptide signal (AMRT) corresponds to the onset of the reverse phase gradient as expected. This method is very useful as it automates the 2D LC separation and produces more protein identifications since the peptides from a single protein are contained in a single file.

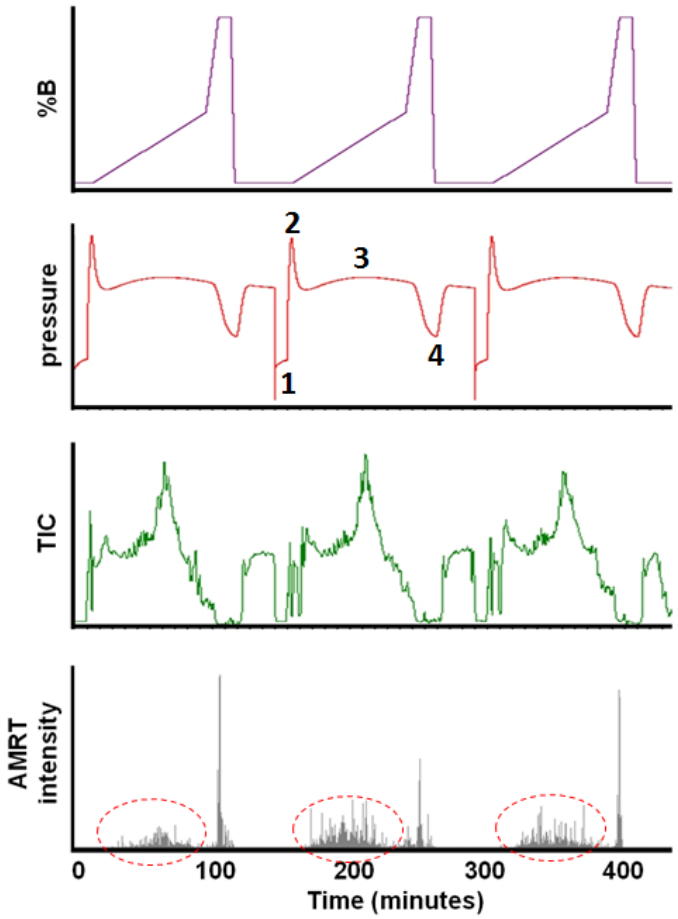



Figure 13 2D LC/MS in a single data file
The gradient as percent of the organic solvent is shown along with the pressure, TIC trace and AMRT are plotted vs. time. The pressure traces highlights four regions 1-the ion exchange step, 2-the trapping step, 3-the gradient region and the 4-the cleaning region. The possible peptides are enclosed by red, dashed circles in the AMRT plot.

Quantitation with LC/MS$^E$

An advantage of LC/MS$^E$ is that relative quantitation, quantitation between various samples for the same protein, can be performed without a protein or peptide labeling step. In addition to relative quantitation with LC/MS$^E$ data sets, there are reports of calculating absolute quantitative data from this technique. The assumption of this method is that the average intensity of the top three peptides from any protein is

proportional to its molar concentration. Then, by spiking the sample with a known amount of a known protein this method can be used to return the molar concentration of a protein. This method is easy to perform however at best it returns approximate molar concentrations of identified proteins.

This analysis was performed on a digested solution of a *Sulfolobus* turreted icosahedral virus (STIV) preparation and Figure 14 shows the resulting normalized protein concentration. It is interesting to note that genes that are next to each other in the viral genome have similar protein expression levels and there are two occurrences of this kind in this STIV data set. The first group contains three structural proteins and the second group consists of two proteins that have putative functions involved with the turret structure. This suggests that the protein contained in these two groups may be co expressed. This analysis alone may not lead to hard conclusions however the ease of performing this experiment and the data that is generated both suggest that this type of analysis is worthwhile.

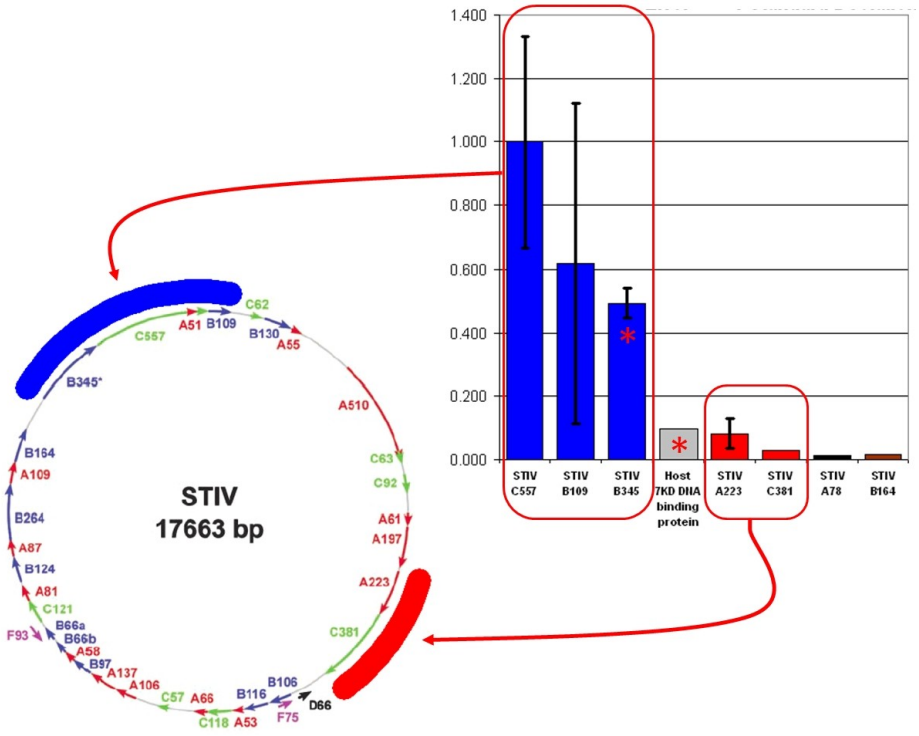
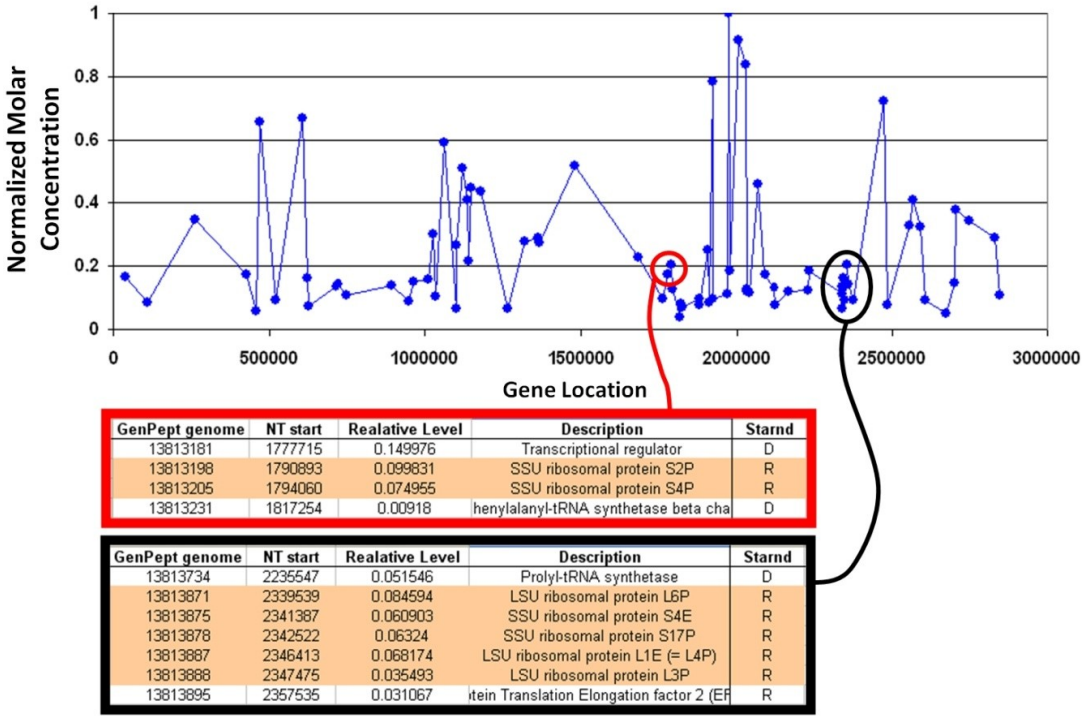



Figure 14 LC/MS$^{E}$ Quantitation of STIV proteins
The comparison of gene location and protein expression is made. The expression level of these proteins are plotted as an estimate of relative molar protein concentration determined from LC/MS$^{E}$.

A study of proteins from a *Sulfolobus solfataricus* P2 (SS P2) cellular digest produced similar results to those of the STIV study that was mention above. Figure 15 A is a plot of normalized gene expression vs. gene location. There are two examples of a group of ribosomal protein subunits, which are grouped since they are near each other in the genome, that have similar levels of protein expression within in each of these "operons". The operon examples from the virus and cellular digest both suggest that the LC/MS$^{E}$ method can yield meaningful absolute quantitative estimates on a global scale.




| GenPept genome | NT start | Realative Level | Description | Starnd |
|---|---|---|---|---|
| 13813181 | 1777715 | 0.149976 | Transcriptional regulator | D |
| 13813198 | 1790893 | 0.099831 | SSU ribosomal protein S2P | R |
| 13813205 | 1794060 | 0.074955 | SSU ribosomal protein S4P | R |
| 13813231 | 1817254 | 0.00918 | henylalanyl-tRNA synthetase beta cha | D |

| GenPept genome | NT start | Realative Level | Description | Starnd |
|---|---|---|---|---|
| 13813734 | 2235547 | 0.051546 | Prolyl-tRNA synthetase | D |
| 13813871 | 2339539 | 0.084594 | LSU ribosomal protein L6P | R |
| 13813875 | 2341387 | 0.060903 | SSU ribosomal protein S4E | R |
| 13813878 | 2342522 | 0.06324 | SSU ribosomal protein S17P | R |
| 13813887 | 2346413 | 0.068174 | LSU ribosomal protein L1E (= L4P) | R |
| 13813888 | 2347475 | 0.035493 | LSU ribosomal protein L3P | R |
| 13813895 | 2357535 | 0.031067 | tein Translation Elongation factor 2 (EF | R |

Figure 15 LC/MSE Quantitation of SS proteins
The protein expression is plotted vs. gene location. The expression level of these proteins are plotted as an estimate of relative molar protein concentration determined from LC/MS$^{E}$. Two groups of proteins are highlighted since they are ribosomal subunit proteins that are likely to be expressed in a similar molar ratio.

The SS P2 LC/MS$^{E}$ data set can be compared to a study of the yeast proteome in which the authors used an affinity library where each open reading frame was tagged with a high-affinity epitope and expressed from its natural chromosomal location to determine quantitative protein information on a global scale.[50] It should be noted that obtaining the yeast data was a major undertaking compared to the one day analysis of the SS P2 proteome. In Figure 16 we can see that both organisms have a similar trend in protein expression level when the proteins are plotted in decreasing abundance. This plot is organized by decreasing protein abundance so proteins at similar abundance between the two species don't necessarily have the same function.

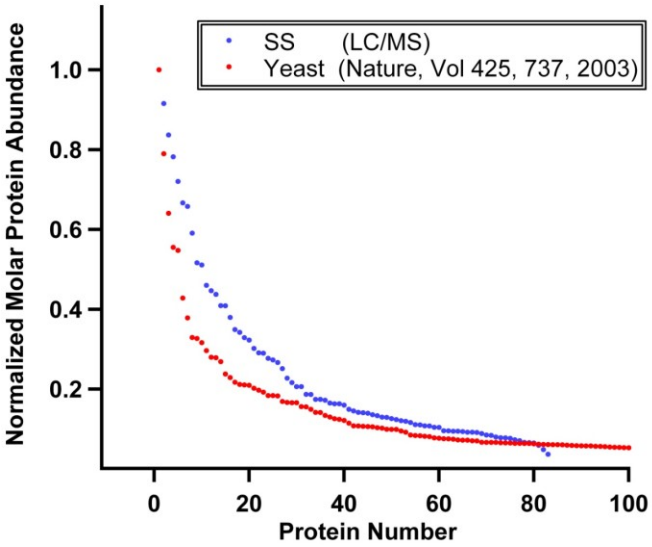



Figure 16 Comparison of yeast and SS protein expression
Relative molar concentration estimates for Yeast and SS proteins plotted together and organized by decreasing abundance.

Another way to look at this data is to calculate the number protein molecules present in each cell. The SS P2 LC/MS$^{E}$ data can be converted to the number of protein molecules expressed per cell by scaling the data with quantitative information from the thermosome protein of SS that has been previously described. First, the mass of the SS P2 cell can be assumed to be equal to $1.7 \cdot 10^{-23}$ grams on the basis that it is a sphere with at density of 1g/mL. Another assumption is that 20% of the mass is from protein and that the thermosomes subunits A and B contribute to 4% of the total protein mass.[51] We can put the above assumptions into equation 2.1 and calculate the number of proteins per cell for the thermosome proteins.

$$\frac{Thermosome\ Subunits}{Cell} = \frac{Mass_{Cell} \cdot Percent_{Total\ Protein} \cdot Percent_{Thermosome(A,B)} \cdot 6.022 \cdot 10^{23}}{MW_{Protein}} \qquad \text{eq. 4}$$

When the thermosome data is put into equation 2.1 a value of 7,956 is returned. In equation 2.2 *k* is introduced and is the universal scaling constant for all proteins in this data set. We then can set equation 2.2 equal to 7,956 proteins per cell and solve for *k* buy inputting the LC/MS$^E$ data in for the thermosomes. Specifically, *k* equals the protein count divided by arbitrary value obtained from the MS$^E$ analysis and this was fixed to the constant value of 8,572 for the analysis.

$$\frac{Molecules_{Protein}}{Cell} = (k \cdot MS^E_{Protein}) \qquad \text{eq. 5}$$

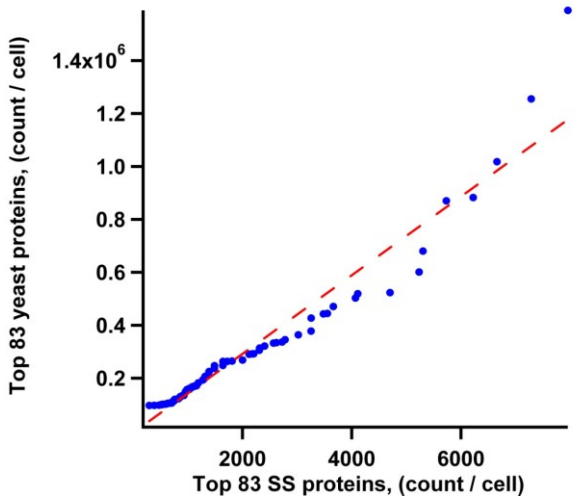



Figure 17 Comparison of yeast and SS protein per cell
Estimates for protein count per cell for yeast and SS proteins graphed in an x-y plot with yeast data on the y-axis and SS data on the x-axis. A linear fit to this data is shown (y = 148.67(x) – 5417).

Since there have been no similar reports of this kind, for the global expression of SS P2 proteins, I was interested in finding a way to verify the accuracy of the data. With this in mind Figure 17 was generated and this is an x-y plot of the number of proteins per

cell for two species. The y-axis contains the yeast data from a previous report vs. the x-axis that contains the SS P2 data. Although the two organisms are extremely different, it was assumed that the protein per unit volume should be of a similar magnitude. Therefore the slope of a linear fit to Figure 17 should be similar to the relative differences in the two cellular volumes. The experimentally determined slope is similar (slope = 149) to the relative difference of the cellular volume (relative volume difference = 125). These analyses highlight the ease in which global quantitative data can be estimated.

## Noncovalent Mass Spectrometry

MS analysis of protein cages in their native form has become feasible in the past decade due to advances in the field.[52-54] This new characterization technique is extremely beneficial in the design of complex protein cage constructs. In the case of the branched polymer grown inside HSP, which will be discussed in Chapter 5, this technique was critical in the complete characterization of these constructs. In this case, SEC-MS could only provide informative data up to the midpoint in the polymer growth since subunit-subunit cross linking occurred. Furthermore the polymer didn't contain any readily detectable molecular signatures that could be easily exploited by other analytical techniques for the quantitation of polymer growth. Native gels were unable to provide meaningful characterization of polymer growth since the size of these particles was unchanging. Although NC/MS analysis of the HSP branched polymer hybrid was unable to produce resolved ion peaks, this analysis allowed for complete characterization of polymer growth.

Instrument Setup, Online and Offline

The advantages of the TOF analyzer are most obvious for analysis of protein complexes in their native states. In this application it is necessary to transfer the complex from atmospheric pressure to a vacuum high enough for efficient TOF operation. The conventional operation pressures of commercial mass spectrometers are not conducive to maintaining noncovalent interactions. By increasing the pressure inside the instrument complexes undergo more collisions with nitrogen and argon molecules. This has the effect of reducing the kinetic energy of the complex by transferring it into vibrational modes within the complex. A balance must be made between the pressure of the instrument and electronics to reduce the kinetic energy sufficiently for proper control of the complex while not transferring to much energy into vibrational modes which could break the noncovalent interactions. On the QTof Premier there is a valve which connects the first turbo pump 1 (Figure 18A) with the backing pump. Closing this valve has the effect of increasing the pressure to the proper range for detection of noncovalent complexes.

Another important parameter for NC/MS analysis is the proper solution conditions. A critical experimental parameter is that protein complexes must be thoroughly desalted prior to the ESI process. Residual, non volatile small molecules will concentrate within the droplets during the ESI process and then be transferred to the complex. This has the effect of making the spectrum very broad due to a large distribution of salt or buffer molecules - complex combinations and therefore these small molecules can decrease the signal until the complex becomes undetectable. A nanospray

source is the most commonly used source of NC/MS since the droplets begin the ESI process much smaller than compared to the standard ESI source and therefore there is less of a concentration effect for the small molecules. Also, the smaller droplets transfer more complexes into the instrument in the case of nano ESI. Figure 18A shows the typical setup for nanospray NC/MS experiments.

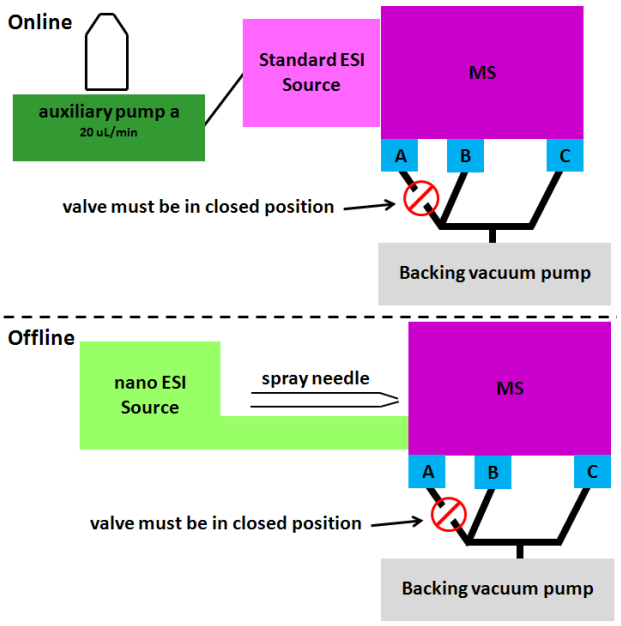



Figure 18 NCMS setup
Online and Offline instrumentation setup for NCMS analysis. The valve that must be closed for proper pressure conditions for NCMS analysis in shown.

Offline Example

Nanospray NC/MS was used to track the synthesis of HFn conjugated with IgG, Figure 19. NC/MS analysis has the advantage of low sample consumption typically 10uL

of a 1mg/mL sample was used to create a spectrum. Also when the conjugated complex is detected the number of IgG molecules per cage is given.

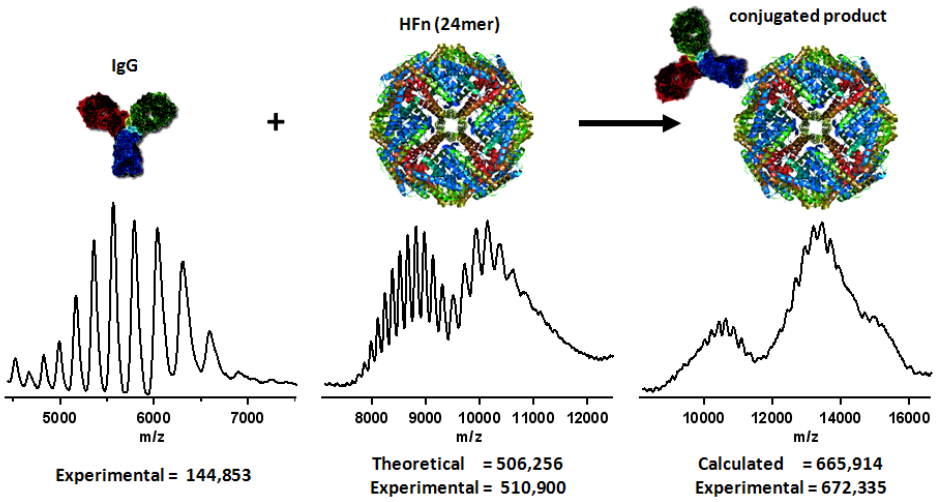



Figure 19 NC/MS characterization of IgG conjugated to HFn
Monitoring the synthesis of a conjugation of HFn + IgG via an SPDP/SMCC coupling reaction. Three separate nanospray needles were loaded with 100mM ammonium acetate with each tip containing either IgG, HFn or the synthesis product. The concentration of protein was approximately 1mg/mL for all three needles. The charge state envelope on the left for the HFn spectrum is for a 24mer while the charge state envelope on the right is for a 22mer. In the spectrum of the conjugated product, the left charge state envelope is for unlabeled HFn while the conjugates envelope is on the right.

Online Example

While the best resolution and sensitivity results from nanospray ESI sources, the standard sources offers ease of use, a predictable signal and the ability to couple LC for direct injection or online desalting. So for smaller complexes the conventional source is desired over the nanospray source. Shown in Figure 20 are two NCMS spectra of the Dps protein cage from *Listeria innocua* (LiDps), one taken before mineralization and one after. After mineralization there is a shift to higher m/z values which corresponds to an

increase in mass. Since the ion peaks are not resolved an estimate for mass can be calculated by assuming the mineralized cage has the same charge state as the empty cage. This figure highlights the ability of NC/MS to obtain quantitative data on protein cages that have undergone modification which results in heterogeneous products.

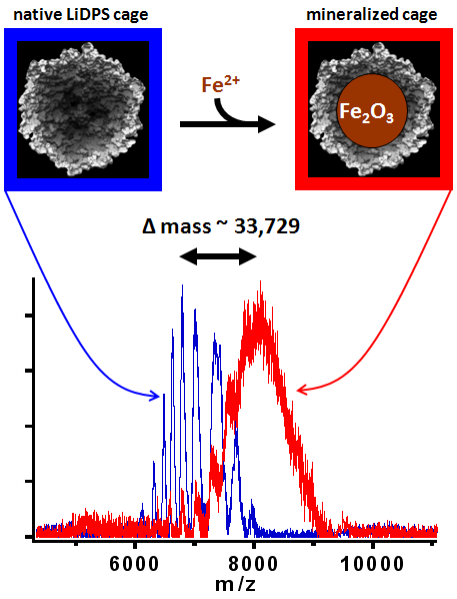



Figure 20 NCMS characterization of mineralized LiDps
Two spectra are shown of LiDps, the blue trace is of empty protein cage and the red trace is a spectrum of iron-loaded LiDps.

## Conclusions

Methods have been developed for the high resolution characterization of protein cages by numerous mass spectrometry techniques. While, none of these methods were designed with the sole intention to perform at the highest level of sensitivity, accuracy (resolution) and speed that was capable with the available instrumentation. The current instrumentation setup achieves an extremely high level of flexibility in the types of

analytical questions that can be addressed. Currently it is possible to automatically toggle between analyses that yield information on small molecules such as chelator bound metals, proteomic peptide analysis, subunit mass interrogation and native spray analysis to detect intact protein cages. Furthermore, with minor user input it is possible to perform high performance experiments such as nanospray analysis of intact protein cages and nanospray 2D/LC proteomic analysis of complex protein mixtures.

# CHAPTER 3

# CORRECT CHARGE STATE ASSIGNMENT OF NATIVE ELECTROSPRAY SPECTRA OF PROTEIN COMPLEXES

## Introduction

The information available from mass analysis techniques, including native gel electrophoresis, size exclusion chromatography, ultra centrifugation, and mass spectrometry (MS) increases dramatically with resolution. Mass spectrometry currently provides the highest mass resolution. The application of MS to the analysis of large intact biomolecular complexes has become a reality through the introduction of electrospray ionization (ESI) and the improvement of large-ion transmission through mass analyzers.[55-66] This type of MS, termed native electrospray mass spectrometry, maintains weak intermolecular protein–protein, protein–ligand, and complex–adduct interactions allowing for mass interrogation of biomolecular complexes. To date, this analytical approach has provided details on the stoichiometry and 3D organization of complexes, subunit exchange kinetics, and thermal dissociation equilibrium values.[67-70] In the future, increasing instrument performance, as well as a more thorough understanding of the interaction between complexes and adducts, will greatly improve the utility of this technique. Here we describe a feature of some native electrospray spectra that invalidates the application of the conventional charge state assignment method to these spectra. Furthermore, we exploit this feature in a method we have developed for charge state

assignment. This method is robust, applicable to all native electrospray spectra, and results in more accurate mass assignments.

A characteristic of ESI, for large analytes, is that the molecule or complex of interest takes up multiple protons. This results in the formation of multiple ions, each differing in charge state (*Z*) and having a unique A characteristic of ESI, for large analytes, is that the molecule or complex of interest takes up multiple protons. This results in the formation of multiple ions, each differing in charge state (*Z*) and having a unique *m/z* value as shown in equation 6 where *M* is the mass of the analyte. Two tasks have to be completed in the deconvolution of raw ESI data. First, the charge states must be calculated for each ion, which can be accomplished by inputting the *m/z* values of two adjacent peaks into equation 7. The second task is to calculate a mass (*M*) for all ions by using equation 6.

$$(m/z)_z = \frac{M + Z}{Z} \quad \Rightarrow \quad M = (m/z)_Z \cdot Z - Z \qquad \text{eq. 6}$$

$$Z = \frac{(m/z)_{Z-} - }{(m/z)_{Z-} - m/z)_Z} \qquad \text{eq. 7}$$

Reported masses from ESI spectra are calculated by averaging the values of *M* for all ions. Complications are common with charge state assignment of "native electrospray" spectra. This is in contrast to standard ESI, where the assignment of charge states is unambiguous. In these spectra, noncovalent interactions are not maintained, and individual molecular species are detected. These complications stem from a few qualities unique to native electrospray. First, the ion peaks are broad relative to conventional ESI, making the peak centers more ambiguous.[71] A second complication arises because of the

relatively uniform ion peak spacing over the envelope of charge states for native electrospray spectra, and this effect will be discussed in the Results section. Finally, native electrospray spectra typically have fewer ion peaks than spectra from conventional ESI. Therefore, equation 7 is applied fewer times, resulting in greater uncertainty in the charge state assignment.

Robinson and coworkers have shown that an iterative approach can aid in the assignment of charge states for native electrospray spectra.[61, 65, 72] In their approach, they choose a range of candidate charge states and calculated masses for all associated ions at every charge state iteration. A standard deviation of the resulting calculated masses is produced at each iteration. The charge state series that produces the smallest standard deviation is deemed correct. This method has been successfully applied in other native electrospray studies. We refer to this charge state assignment strategy as the *Conventional ESI* method.

The conventional ESI method is extremely useful; however its application to some spectra results in incorrect charge state assignments. We found that the source of this ambiguity was that the mass of a complex ($M$), for a subset of spectra (we term non-ideal), changes for ions of different charge states. More specifically, $M$ increases as the charge state decreases. Since $M$ is not constant the application of equation 7 is no longer valid and therefore there is no analytical solution to the charge state assignment. Therefore, we were interested in developing a general data analysis technique applicable to all native electrospray spectra for charge state assignment.

## Methods

### Sample Preparation

*SsDps-L and LiDps* were heterologously expressed and purified as described elsewhere.[38, 73] Before mass spectrometry analysis, samples where desalted and transferred to an appropriate buffer by using P-30 Micro Bio-Spin columns (BioRad, Hercules, CA). The samples were transferred into 10 mM solutions of either ammonium acetate (AA) from Aldrich (St. Louis, MO), ReagentsPlus, 99.99% (catalog no. 431311); ammonium bicarbonate (AB) from Sigma (St. Louis, MO), 99.0% (catalog no. A6141); triethylammonium acetate (TEAA) buffer from Fluka, St. Louis, MO (catalog no. 90,357). All stock buffer solutions were made fresh and prepared with water from Fisher Scientific (Pittsburgh, PA), environmental grade (catalog no. W11). The pH of the buffers was adjusted by adding small volumes of dilute solutions of either ammonium hydroxide or acetic acid.

### Instrumental Configuration

The mass spectrometer used was a QTOF Premier (Waters, Milford, MA). Higher than standard operating pressures were maintained within the instrument by closing the speedy valve, which is positioned between the backing pump and the turbomolecular pump located in the source region. Typical pressures were 8mbars, backing; $1.2 \cdot 10^{-2}$ mbars, collision cell; $3.1 \cdot 10^{-6}$ mbars, TOF. Typical voltages used where 2000 V, capillary; 70 V, sample cone; 50 V, collision cell. Samples were introduced into the

standard source of the mass spectrometer by direct injection using the Aquity liquid chromatography platform (Waters) with a flow rate of 15 µL per minute.

## Results

### The Effect of the Charge State on the Spacing of Adjacent ESI Ion Peaks

Native electrospray spectra typically have a few properties that can lead to ambiguous charge state assignments. These properties are broad ion peaks, a low number of ions, and finally the uniform ion peak spacing across a charge state envelop. The relationship of the absolute value of the charge on this last property, the spacing of the ion peaks, is outlined below. To determine the relationship of the ion peaks spacing and the absolute value of the charge state for associated ions of ESI spectra, equations 8 to 11 can be utilized to describe the *m/z* difference between two adjacent ion peaks ($\Delta(m/z)_{Z\leftrightarrow Z+1}$). In these equations $M$ is the mass of the analyte and $Z$ is the charge state of the ion.

$$\Delta m/z_{Z\leftrightarrow Z+1} = (m/z)_{Z} - (m/z)_{Z+1} \qquad \text{eq. 8}$$

$$\Delta m/z_{Z\leftrightarrow Z+1} = \frac{M+Z}{Z} - \frac{M+Z+1}{Z+1} \qquad \text{eq. 9}$$

$$\Delta m/z_{Z\leftrightarrow Z+1} = \frac{(M+Z)(Z+1)}{Z(Z+1)} - \frac{M*Z+Z^2+Z}{Z(Z+1)} \qquad \text{eq. 10}$$

$$\Delta (m/z)_{Z\leftrightarrow Z+1} = \frac{M}{Z^2+Z} \qquad \text{eq. 11}$$

When this relationship is plotted the inverse relationship between $\Delta(m/z)_{Z\leftrightarrow Z+1}$ and $Z$ becomes obvious(Figure 21a).

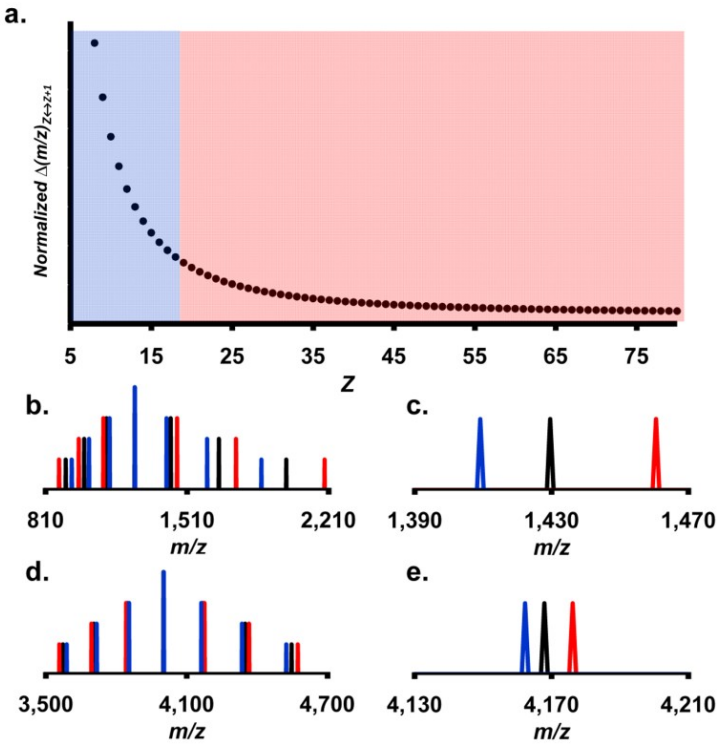



Figure 21 Ions spacing dependence on charge state
(a) A plot of normalized $\Delta(m/z)_{Z\leftrightarrow Z\_1}$ versus *Z*. Typical charge state ranges for standard ESI spectra are shown as the blue region while the region for typical charge states for native electrospray spectra are shown in red. (b)–(e) Generated spectra of a 10kDa analyte (standard ESI spectra) (b), (c) and of a 100 kDa analyte (native electrospray spectra) (d), (e). Plots (c) and (e) are zoomed in views of the peaks directly to the right of the base peaks from plots (b) and (d), respectively. The x-axis in plots (c) and (e) are presented at the same scale (range = 80 *m/z*). The correct charge states were used to calculate ion peak positions shown in black (b)–(e). The blue ion peaks were generated by using the correct charge state values plus one while the red ion peaks were calculated using the correct charge state values minus one (b)–(e).

The relationship between predicted maximum charge state and the mass of the complex produced from the lab of de la Mora is given in equation 12

$$Z = 0.078 \cdot M^{1/2} \qquad \text{eq. 12}$$

and was used to determine typical *Z* values for conventional ESI spectra (mass = 10,000 to 50,000 Da) and native electrospray spectra (mass = 50,000 to 1 M Da).[53, 74] By

indicating typical charge state ranges for both conventional and native electrospray spectra in the plot of normalized $\Delta(m/z)_{Z\leftrightarrow Z+1}$ versus $Z$ it becomes clear that the spacing of the ion peaks is less variant for ion peaks from native electrospray spectra. The y-axis in this plot was normalized by setting $M$, in equation 11, equal to unity.

To show the outcome of the charge state's effect on the ion peak's spacing, two sets of generated mass spectra were calculated and are presented in Figure 21b–e. In Figure 21b and c, the spectrum generated was calculated for an analyte that has a theoretical mass of 10 kDa (representative of standard ESI spectra) in contrast to Figure 21d and e, where the spectrum represents an analyte with a mass of 100 kDa (representative of native electrospray spectra). This figure illustrates that the spacing between ions across the charge state envelope in the case of the 10 kDa spectrum varies much more than the spacing for the 100 kDa example. Two additional spectra were produced for both the 10 and 100 kDa examples by using charge state values of +1 and -1 of the correct charge state value. In the 100 kDa example, the incorrect spectra align much better with the correct spectrum when compared with the 10 kDa example. The inverse relationship between the absolute value of the charge state and the ion peak spacing makes charge state assignment more difficult at high charge state values.

The Mass Dependence on Charge State

Native electrospray analysis of the *Ss*Dps-L protein cage (DNA binding protein from starved cells-like protein) provided a model system for investigating accurate charge state determination. *Ss*Dps-L is derived from the hyperthermophilic archaeon

*Sulfolobus solfataricus* and is a spherical complex that is assembled from 12 identical protein subunits with an o.d. ~10 nm (Figure 22).[38, 75]

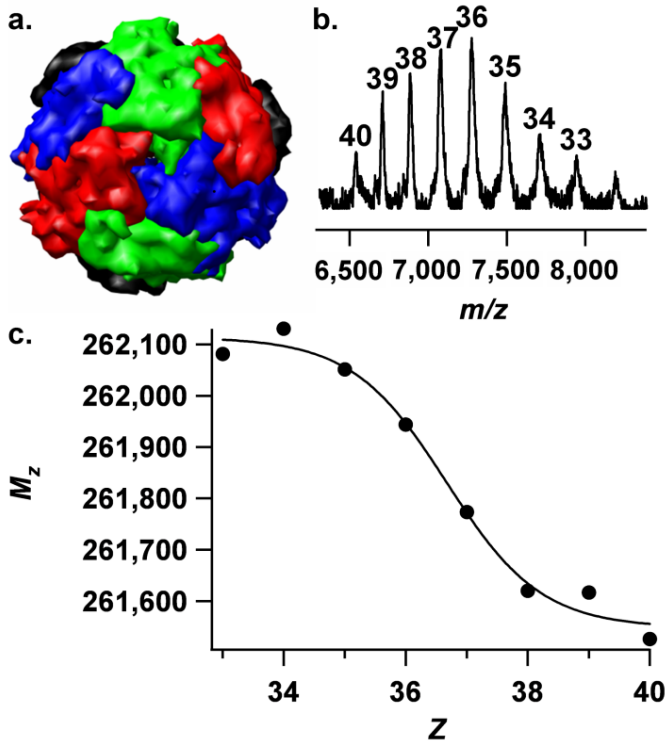



Figure 22 NCMS of SsDps
(a) The structure of the intact *Ss*Dps-L complex, from the hyperthermophilic archaeon *Sulfolobus solfataricus,* oriented about the 3-fold axis.[75] (b) The prototypical native electrospray spectrum of the *Ss*Dps-L complex with the correct charge states (*Z*) labeled above the ion peaks. (c) A plot of $M_Z$ versus *Z* for ions from the prototypical *Ss*Dps-L spectrum.

The native mass spectrum of this intact complex (Figure 22b) is typical in that it has a narrow charge state envelope (from 33 to 40) and relatively broad ion peaks compared with conventional ESI spectra. Using the known mass of *Ss*Dps-L, a charge state was assigned to each ion peak. A mass ($M_Z$) was then calculated for each ion peak by using equation 6. A plot of $M_Z$ versus *Z* (Figure 22c) shows that the mass varies with

the charge in a sigmoidal fashion and approaches both an upper and a lower limit at the charge state extremes.

This sigmoidal behavior is also apparent for other complexes that we have analyzed in our lab, in addition to spectra that have been reported.[53, 76, 77] Figure 23 shows examples of sigmoid trends of data obtained from spectra from three complexes; another Dps protein from *Listeria innocua* (*Li*Dps), the α6-glutamate synthase complex from *Azosprillum brasilense* and the urease complex from *Helicobacter pylori*. [35, 53, 73, 76, 77]

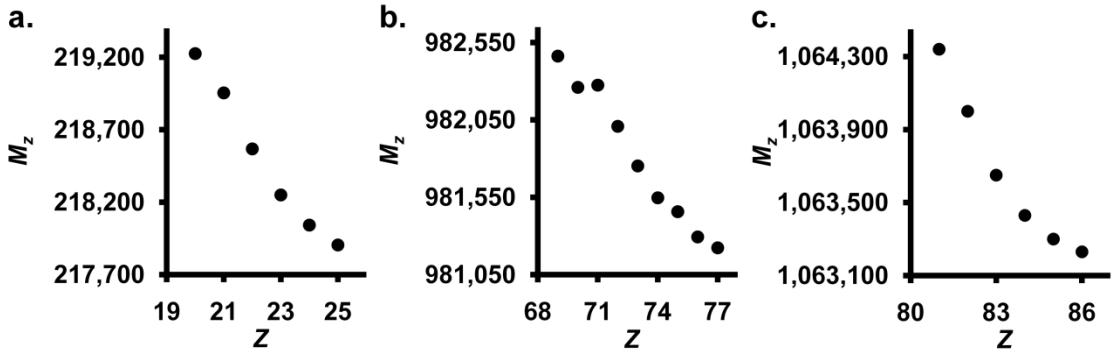



Figure 23 NC/MS of LiDps, glutamate synthase complex and urease
Plots of $M_z$ versus $Z$ for three complexes. Data from an *Li*Dps spectrum (a), data from a previously reported spectrum of the α6-glutamate synthase complex (b), and an urease complex, which has been previously reported (c).

To aid in the comparison of the sigmoid trends shown in Figure 23 and others, normalized mass and centered charge values were calculated with the following procedure to allow for one plot containing sigmoid trends from a multitude of spectra. Data shown in Figure 23 along with data from other spectra, in the form of $M_Z$ versus $Z$ were fit with the sigmoid equation shown in equation 13.

$$M_z = M_C + \left\lceil \frac{M_{CMA}}{1 + e^{(Z - C_S) \cdot R}} \right\rceil$$ eq. 13

The $M_C$ term is a variational parameter that describes the mass of a complex stripped of all but the most tightly bound adducts and is an asymptotic limit of the sigmoid function. The mass of the complex with a maximal amount of adduct loading (complex mass plus the mass of the adducts) is described by the variational parameter $M_{CMA}$. $C_S$ is the point of maximal slope for the change in mass as the charge state changes. Finally, $R$ is a floating term that dictates the overall rate of mass change as the charge varies. The normalized mass term (normalized $M_Z$) was then calculated with equation 14.

$$NormilizedM_Z = \frac{M_Z - M_C}{M_{CMA} - M_C}$$ eq. 14

The centered $Z$ term was calculated by subtracting the charge state of the ion peak from the variational parameter $C_S$. Figure 24 shows the similarity in the sigmoid nature across complexes and analysis conditions.

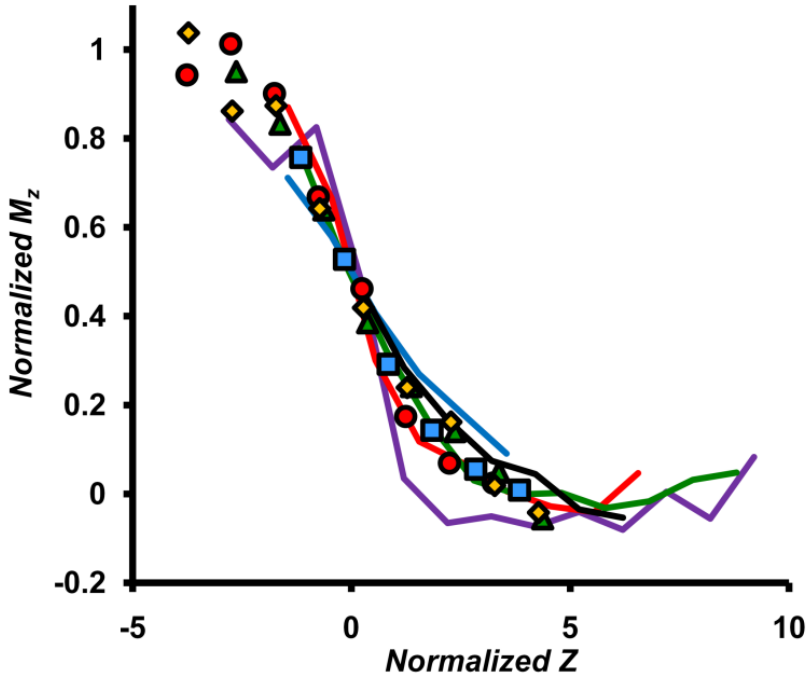



Figure 24 Normalized plot of Mass vs. Centered Charge State
Plots of *Normalized* $M_Z$ versus *Centered Z* for a variety of complexes analyzed under various conditions. Data from the prototypical *Ss*Dps-L spectrum (Figure 2.2c) is plotted as red circles, data from a spectrum of the urease complex, which has been previously reported is plotted as blue squares, and data from another previously reported spectrum of the α6-glutamate synthase complex is plotted as yellow diamonds. Data from a native electrospray spectrum of IgG is shown as green triangles. Data from *Li*Dps and *Ss*Dps-L spectra are plotted as solid lines. *Li*Dps in ammonium acetate pH 6.8 (blue), ammonium acetate pH 8.7 (purple), ammonium bicarbonate pH 8.2 (green), triethylammonium acetate pH 6.8 (red), *Ss*Dps-L in ammonium acetate pH 8.2 (black).

Development and Evaluation of Four Charge State Assignment Methods

Charge state assignment of native electrospray spectra can be ambiguous for the reasons discussed above, here we evaluated four charge state determination strategies on three complexes, *Ss*Dps-*L, Li*Dps, and the α6-glutamate synthase complex. Both Dps complexes were analyzed in a variety of buffer conditions to assess the method's robustness as the solution conditions are varied. One of the methods, the conventional

ESI method, was produced elsewhere as described above.[61, 65, 72] Three of the methods were developed in this work and are described below. We found the most reliable assignment results from a method that uses the ion's peak width to aid in the determination of the correct charge state. This technique corrects for the variation of mass for ion peaks of different charge states.

*R2 method.* We observed that the charge-state-corrected ion peak width increases as the charge decreases in non-ideal native electrospray spectra. This peak width can be calculated by dividing the ion peak width (fwhm) by the *m/z* at the maximum peak intensity and is referred to as the corrected peak width (*CPWz*).[72, 78] As suggested in previous works, a plot of *MZ* versus *CPWZ* should produce a linear relationship.[72, 78] If the charge state is selected incorrectly, either too high or too low, the linear correlation between *Mz* and *CPWz* breaks down (Figure 25). The point at which the $R^2$ value is at a maximum corresponds to the correct charge state.

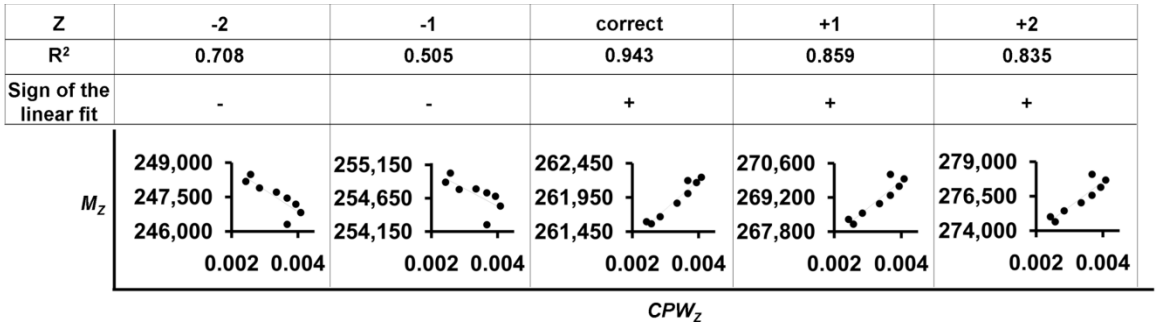



Figure 25 Plots of Mass vs. Corrected Peak Width for various charge states
Plots of $M_Z$ versus C$PW_Z$ for data from the prototypical *Ss*Dps-L spectrum (Figure 22b), where $M_Z$ has been calculated with various charge states. The charge state assignment, $R^2$ value, and sign of the slope of a linear fit to the plots are shown above each plot.

*Absolute value of the slope method.* It has been previously established that a linear relationship of percent mass increase

$$\%MI_Z = \frac{M_Z - M_{Theoretical}}{M_{Theoretical}} \times 00 \qquad \text{eq. 15}$$

*and CPWz* holds true for ions from different complexes.[72, 78] Plots of %*MIz* versus *CPWz*, or similar plots, shown in these publications, are comprised of data points produced from ion peaks from various spectra.[72, 78] These spectra were of different complexes or the same complex measured under different conditions. This suggests that the peak width of an ion is linearly related to the percent mass increase for ions from different spectra. We wanted to determine if this linear relationship was obtained for ion peaks produced from the same spectrum with different charge states. If this linear relationship is general, we could exploit the slopes of these plots as another method to determine the correct charge state. The plots of $\%MI_Z$ versus *CPWz* for *Ss*Dps-L, *Li*Dps and the α6-glutamate synthase complex measured under different buffer conditions do exhibit a linear behavior suggesting that this linear relationship is general (Figure 26a–c).[77] Plots of $\%MI_Z$ versus *CPWz* for *Ss*Dps-L and *Li*Dps measured under various buffer conditions also produced linear trends (data not shown).

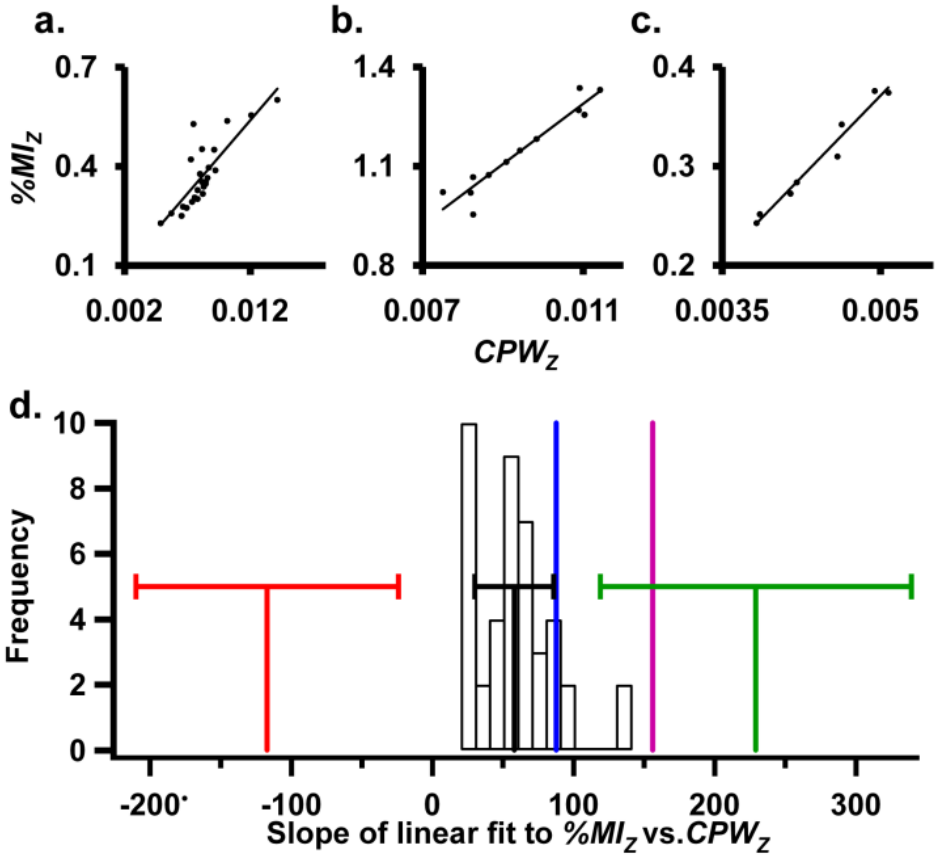



Figure 26 Plots of Percent Mass Increase vs. Corrected Peak Width
(a)–(c) Plots of $\%MI_Z$ versus $CPW_Z$ for three complexes. Data from *Ss*Dps spectra in ammonium acetate pH 8.2 (triplicate injections) (a), data for *Li*Dps in triethylammonium acetate pH 6.8 (triplicate injections) (b), and data from the α6-glutamate synthase complex spectrum (single analysis) (c). All the linear fits from data listed in Table 2.1 are shown in histogram form (d) with the average slope ± one standard deviation as a black line. The average slope produced from picking the charge both -1 and +1 from the correct charge state are shown as a red line (-1) and a green line (+1) with error bars for both of ± one standard deviation. The slope previously reported, which was produced from a plot of $\%MI_Z$ versus $CPW_Z$, for different complexes, is shown as a purple line (d). The slope produced from the α6-glutamate synthase complex is shown as a blue line (d).

In addition to the observation that the plots in Figure 26a–c have a linear relationship, the absolute value of their slopes is also similar. This prompted us to look at the similarity of slopes for a larger dataset, including data from *Ss*Dps-L and *Li*Dps

analyzed under various buffer conditions. A histogram containing the slope values from all the Dps datasets shows this similarity in Figure 26d. The slope values range from 20 to 140 with an average value of 58 ± 28 (±one standard deviation). The slope produced from the α6-glutamate synthase spectrum (87.8) is similar to the average slope produced for the two Dps complexes (Figure 26d). However, the slope value of a linear fit of a previously reported dataset containing data points obtained from different spectra of various complexes or the same complexes ran under different conditions was higher than (156) the average Dps value.[72] The difference in the slope value could be due to different experimental conditions used.

Also included in the histogram, shown in Figure 26d, are the slope values produced from masses calculated from the *Ss*Dps-L and *Li*Dps datasets, with *Z* values that are both -1 and +1 of the correct value. When the incorrect charge states were used slope values of-117 ± 93 and 229 ± 110, respectively were produced (Figure 26d). The slope value from the correct dataset (58) is significantly different from the values from the two datasets produced with the incorrect charge states (-117 and 229). This suggests that the comparison of experimental absolute values of the slope to a predetermined average value for similar spectra can be a predictive tool for charge state assignment.

A potential complication with this method is that the mass of the complex must be known. However, if the theoretical mass of the complex is unknown $\%MI_Z$ values can be approximated by setting the mass of the complex equal to the y-intercept of the linear fit to a plot of $M_Z$ versus $CPW_Z$.

*First positive slope method.* The value of the slope of a linear fit to the plot of $\%MI_Z$ versus $CPW_Z$ (or $M_Z$ versus $CPW_Z$) can also be utilized as another method to determine the correct charge state. This method tracks when the value of the slope of the linear fit switches from a negative value to a positive value as the initial guess of the charge state is iterated with increasing $Z$ values. The first slope with a positive value corresponds to the correct charge state (Figure 25 and Figure 26).

We tested two complexes (*Ss*Dps-L and *Li*Dps) under a variety of conditions and found that the first positive slope method was the most reliable in predicting the correct charge state (100%), followed by the conventional ESI method (87%), then the absolute value of the slope method (80%), and the $R^2$ method was the least reliable (35%). Table 2.1 summarizes the four methods in their ability to pick the correct charge state for these complex/conditions. Furthermore, the first positive slope method correctly predicted the charge state for the spectrum of the α6-glutamate synthase complex (produced outside our lab) while the conventional method picked the incorrect charge state.[77] Based on the samples analyzed in this work, we suggest that the first positive slope method should be used for charge state determination.

Table 3.1. Results of Charge State Assignment Methods

| Complex (buffer) [Collision energy] | $R^2$ | Absolute value of the slope | Conventional ESI | First positive slope | Replicates |
|---|---|---|---|---|---|
| *Ss*Dps-L (TEAA pH 6.8) [CE 50] | 1 | 2 | 0 | 3 | 3 |
| *Ss*Dps-L (AA pH 6.8) [CE 50] | 1 | 3 | 3 | 3 | 3 |
| *Ss*Dps-L (AA pH 8.2) [CE 50] | 0 | 3 | 3 | 3 | 3 |
| *Ss*Dps-L (AB pH 6.8i 7.4f) [CE 50]* | 1 | 0 | 3 | 3 | 3 |
| *Ss*Dps-L (AB pH 8.3) [CE 50]** | 3 | 0 | 3 | 3 | 3 |
| *Ss*Dps-L (TEAA pH 8.3i 7.7f) [CE 50]* | 0 | 1 | 0 | 2 | 2 |
| *Li*Dps (AA pH6.8) [CE 50] | 1 | 3 | 3 | 3 | 3 |
| *Li*Dps (AB pH8.2) [CE 50] | 0 | 3 | 3 | 3 | 3 |
| *Li*Dps (AA pH8.7) [CE 50] | 0 | 3 | 3 | 3 | 3 |
| *Li*Dps (TEAA pH8.3) [CE 50] | 1 | 3 | 2 | 3 | 3 |
| *Li*Dps (TEAA pH6.8) [CE 50] | 2 | 2 | 2 | 2 | 2 |
| *Li*Dps (TEAA pH 6.8) [CE 30] | 0 | 3 | 3 | 3 | 3 |
| *Li*Dps (TEAA pH 6.8) [CE 50] | 1 | 3 | 3 | 3 | 3 |
| *Li*Dps (TEAA pH 6.8) [CE 70] | 2 | 3 | 3 | 3 | 3 |
| *Li*Dps (TEAA pH 6.8) [CE 90] | 2 | 2 | 2 | 2 | 2 |
| *Li*Dps (TEAA pH 6.8) [CE 110] | 1 | 2 | 2 | 2 | 2 |
| *Li*Dps (TEAA pH 6.8) [CE 130] | 0 | 1 | 2 | 2 | 2 |
| All data | Correct for 35% | Correct for 80% | Correct for 87% | Correct for 100% | 46 |

* "i" Indicates the pH prior to injection of the first replicate and "f" indicates the pH after the final injection.
** Data not used to produce the histogram in Figure 6d.

The consequence of a misassigned charge state is a dramatic decrease in the accuracy of the mass measurement. For example, considering all the spectra summarized in Table 2.1, the average ppm mass error of the spectra in which the conventional ESI method misassigned the charge states is 33,920 compared with the ppm error of 6,052 for these same samples, where the charge states were selected correctly with the first positive slope method. This corresponds to an absolute mass error of 8,611 Da for the conventional ESI method and 1,479 Daltons for the first positive slope method. These mass errors were calculated by comparing the experimental mass values to calculated mass values, which were calculated from the protein sequences and subunit compositions of the complexes.

Suggested Charge State Assignment Workflow

A macro that automatically performs the first positive slope method's steps is available for download at http://chemistry.montana.edu/douglasgroup/software.

1. Determine all peak positions (*m/z* units) and peak widths (fwhm) for all associated ions.

2. Calculate the $CPW_Z$ by dividing the peak width of each ion by its peak position.

3. Plot the $CPW_Z$ versus ion number (the ion number is a number starting at 1 to *n* assigned to each ion. The ion with the lowest charge state (right-most ion in the spectrum) should be set to 1. Does the graph show sigmoidal character? Are there larger $CPW_Z$ values on the left side of the plot?

- Yes, the spectrum is non-ideal. Use the *First positive slope* method, step 5.
- No, the spectrum is ideal. Use the conventional ESI method, step 6.

4. Pick a reasonable range of charge states,

- A guess charge state can be calculated with equation 12 by imputing a guess mass for the complex.
- Determine the correct charge state by iterating across the entire range of guess charge states.
- Apply the following charge state assignment methods for all guess charge states. For non ideal spectra weight your charge state assignment decision on step 5, and for ideal weight your decision based on 6. Both steps 7 and 8 can aid in the understanding of charge state assignment.

5. First positive slope of a linear fit to a graph of $\%MI_Z$ versus $CPW_Z$ (or $M_Z$ versus $CPW_Z$). In some cases a positive slope results from noise in the data and not a true positive linear correlation between $\%MI_Z$ versus $CPW_Z$. When this is the case, the correct charge state is usually the determined charge state + 1. The $R^2$ of this linear fit can be used to indicate a true correlation and therefore the true first positive slope.

6. Conventional ESI

7. Absolute value of the slope of a linear fit to $\%MI_Z$ versus $CPW_Z$

8. Highest R2 for a linear fit of the plot $\%MI_Z$ versus $CPW_Z$

## Discussion

Two properties of native electrospray spectra must be mentioned in the discussion of assigning charge states to these spectra. First, it has been previously established that complexes comprise a continuum in which some complexes can be detected almost completely devoid of buffer adducts at one extreme, and there are other complexes that are only detected when heavily loaded with adducts. At the onset of analysis of an uncharacterized complex, the investigator does not have the luxury of knowing the abundance of adducts present. The second property, which has been presented in this work, is that some native electrospray spectra contain ions that result in masses that decrease as the charge state increases (termed here as non-ideal). The conventional charge state assignment methodology will correctly predict the charge state in two types of native electrospray spectra. The first type is the ideal spectra in which the mass is invariant for ions of all charge states. The second type of spectra in which the

conventional methodologies will correctly predict the charge state is the non-ideal spectra that contain a low abundance of adducts present. As the abundance of adducts increases, for non-ideal spectra, there comes a point where the conventional charge state assignment methodologies incorrectly predict the charge state. The charge state assignment methodologies presented in this work are most helpful to these spectra and must be applied. Finally, for complexes that have not been previously characterized both properties of the complex, the amount of adducts present and the type of spectrum, ideal or non-ideal, are unknowns and, therefore, these uncharacterized complexes should have their charge states assigned with the methodologies presented in this work.

The relationship between the absolute value of the ion peak's charge state and the spacing of adjacent ion peaks should be included in the discussion of reasons why charge state assignment can be ambiguous for native electrosprayed complexes. In equation 11 it is evident that the spacing becomes relatively uniform as the charge state increases, as this may be the most problematic aspect of charge state assignment for large supramolecular complexes.

The normalized sigmoidal plots shown in Figure 24 provide insight into how the abundance of adducts change for ion peaks with different charge states. These plots obtained from different spectra follow a relatively similar sigmoid trend. This suggests that the change in the proportion of adducts present as the charge state is changed is similar for data from different complexes/conditions. Also, the overall range of mass difference between associated ions can vary widely between datasets, but the change in the mass occurs over a similar range of charge states. This is an interesting result and

may provide clues for a more thorough understanding of the relationship of the adduct abundance and the ion peak's charge state.

## Conclusions

We have described how the mass and, therefore, the adduct composition of large macromolecular complexes can change across associated ions of different charge states. Due to variation of mass for associated ions, we have shown that the application of the conventional charge state assignment method to all native electrospray spectra will result in incorrect charge state assignments. Therefore, we have developed a deconvolution strategy that is applicable to all native electrospray spectra and that accurately assigns the charge state of ions and, therefore, produces a more accurate mass assignment. A macro that will automatically perform the suggested native electrospray deconvolution steps is available for download at http://chemistry.montana.edu/douglasgroup/software.

# CHAPTER 4

# METAL BINDING AND MRI PROPERTIES OF ENGINEERED PROTEIN CAGE CONSTRUCTS

## Introduction

MRI is one of the most utilized imaging techniques in medicine since it is noninvasive and provides comparatively high-resolution information. The usefulness of the technique is augmented by the use of contrast agents that increase the rate of water proton relaxation and can therefore increase contrast between tissues. Gadolinium ($Gd^{3+}$) chelates are commonly used as contrast agents in clinical settings. [79, 80] In general, there are two ways to improve the imaging sensitivity using contrast agents, by either increasing the relaxivity of water protons through direct interaction with the contrast agent or by targeted delivery of the agent to specific locations within the body.

### Protein Cage Background

Viral capsids are multimeric protein assemblies that form cage architectures and can be generally categorized as protein cages. Other, non viral protein cages include heat shock proteins, ferritins, and vault ribonucleoprotein particles among others. These protein cages can serve as robust synthetic platforms that are chemically and genetically malleable and can be readily modified. Previous studies have explored the use of protein cages as therapeutic or imaging delivery agents.[81-83] Cell targeting has been achieved by utilizing capsids with natural affinities for cellular receptors or by chemically linking peptides or antibodies to protein cage architectures. [30, 84] In addition, targeted protein

cages incorporating a therapeutic payload (doxorubicin) have been constructed, demonstrating the multifunctional capacity for biomedical applications.[30]

Protein cages have the potential to serve as extremely efficient contrast agents for the following reasons: 1) they are large, commonly between 18 - 100nm in diameter, and relatively rigid molecular structures with large rotational correlation times, resulting in increased relaxivity rates; 2) protein cages can serve as robust platforms where multiple functional motifs can be added through genetic or chemical modifications.[9, 29, 73, 85-89] These modifications could potentially result in the attachment of both $Gd^{3+}$ binding and site specific targeting functionalities and 3) protein cages can potentially carry hundreds (if not thousands) of $Gd^{3+}$ ions and the contrast from an individual capsid will increase significantly with the number of $Gd^{3+}$ ions it carries. Due to these three reasons, protein cages have been investigated as MRI contrast agents including the Cowpea chlorotic mottle virus (CCMV) capsid with bound $Gd^{3+}$ at endogenous metal bind sites and the MS2 virus capsid with GdDTPA chemically attached.[81, 90] Dendrimers, liposomes as well as other supramolecular structures maintain properties 2 and 3 mentioned above and therefore have also been developed as potential contrast agents.[91, 92]

In this study, potential MRI contrast agents based on the multivalent protein cage architecture of the Cowpea chlorotic mottle virus (CCMV) capsids and the heat shock protein from *Methanococcus jannaschii* (HSP) have been developed. The CCMV viral capsid is assembled from 180 identical protein subunits while HSP is composed of 24 subunits. Both cages form symmetrical cage architectures through noncovalent interactions between the subunits. A consequence of the cage-like architecture

assemblies is that amino acid residues are organized and presented in a repeating and symmetrical fashion over the protein cage's surface. Therefore, genetic and chemical modifications of amino acids within the subunit results in the presentation of the modification at all 180 symmetrically related sites on the protein cage. The protein capsid of CCMV has an exterior diameter of 28nm and an interior diameter of 22nm. The CCMV capsid undergoes pH dependent swelling, which opens sixty, 2nm pores. HSP has an exterior diameter of 12nm and an interior diameter of 8nm. There are eight, 3nm pores and six, 2nm pores present on the HSP protein cage. The structures of these cages are shown in (Fig. 27).[6]

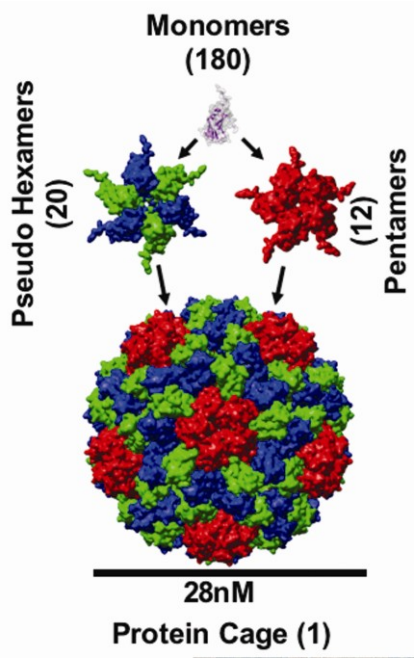



Figure 27 The structure of CCMV
The 28nm viral capsid of CCMV is made of 180 individual subunit proteins (20 hexamers and 12 pentamers).

Relaxivity Background

MRI images are produced by first placing a sample in a magnetic field. When the sample is placed in a magnetic field the nuclear spin of protons residing in the sample align with this external field. The MRI scanner produces a pulse of energy at the correct orientation and frequency which tilts the spin of the protons out of alignment with the external field. The scanner then measures how long it takes for the "protons" (nuclear spin of the protons) to "relax" to their equilibrium state. These are the fundamental steps that are required to make an MRI image. Bright objects in the image have short relaxation times while dark areas have long relaxation rates and the bulk of the image is produced from water protons.

MR images result from probing one of two relaxation processes either $T_1$ relaxation (spin-lattice or longitudinal) or $T_2$ relaxation (spin-spin or transverse). Gd agents typically are used to enhance the $T_1$ relaxation process and $T_1$ specifically refers to a time in units of seconds equal to a realignment decay constant for the protons in a given sample. The $T_1$ of water, for example, is approximately 3 seconds.

There are a few parameters that must be optimized with regard to Gd based contrast to achieve maximum relaxivity per Gd ion.[79, 93-97] First of all, Gd is toxic and therefore it must be in the form of a metal ion chelate. The number of water ions that can occupy a coordination site of Gd, this number is referred to as "q". In general, as q increases the stability of the chelator-metal interaction decreases. The majority of the relaxivity affect of Gd occurs when a water molecule becomes bound to the Gd ion (inner-sphere relaxation) so a balance must be made between increasing the number of

water coordination sites and stability of the chelator-metal interaction. Another parameter that must be optimized is the residence time of the Gd bound water referred to as $\tau_M$. This parameter can be controlled by selecting the proper Gd chelator. The final parameter that is important for efficient PRE is the rotational correlation time of the Gd ion ($\tau_R$). Larger Gd complexes result in greater $\tau_R$ values which yield more efficient PRE.

The PRE of a contrast agent is given in units of seconds$^{-1}$ · mM$^{-1}$ and is referred to as *relaxivity*. The Gd concentration (mM) is used for the calculation of *ionic relaxivity* while the concentration of a particle (mM) containing one or more Gd ions is used to calculate the *particle relaxivity*. For example a 1 mM aqueous solution of Magnevist, which is a clinically employed contrast agent and has a relaxivity of $r_1 \approx 5$ seconds$^{-1}$ mM$^{-1}$, has a $T_1$ of 0.19 seconds based on equation 16.

$$T_{1\,observed} = \left[\frac{1}{T_{1\,sample}} + \wedge R_{1\,contrast\ agent}\left(Concentration_{contrast\ agent}\right)\right]^{-1} \qquad \text{eq. 16}$$

The utility of this technique is greatly limited the inability of current scanners to detect localization of low concentrations of contrast agents *in vivo*. Therefore *in vivo* contrast agent targets are only successfully detected when these targets are larger than the resolution of the MR image. The size and copy number of targets that can be detected can be reduced by increasing the efficiency of detecting these contrast agents. Designing a contrast agent on the scaffold of a protein cage holds much promise in increasing the relaxivity of the contrast agent and therefore the overall sensitivity of this technique.

CCMV has been shown to contain an endogenous metal binding site. In a previous study characterizing the metal binding by FRET, tryptophan residues proximal

to the endogenous metal binding site were used to determine $Tb^{3+}$, $Gd^{3+}$ and $Ca^{2+}$ dissociation constants ($K_d$ = 19μM, 31μM and 2mM respectively).[98] Further studies indicated that $Gd^{3+}$- bound CCMV (CCMV-Gd) capsids exhibited high ionic relaxivity rates ( $r_1$ = 202, $r_2$ = 376$mM^{-1}$ $sec^{-1}$, at 61 MHz ) and extremely high relaxivity rates per capsid ($r_1$ = 28,482 and $r_2$ = 53,016 $mM^{-1}$ $sec^{-1}$, 61 MHz).[90] However $Gd^{3+}$ binding to the endogenous sites was too weak for the CCMV-Gd to be useful as a clinical MRI contrast agent, since free $Gd^{3+}$ is toxic *in vivo*.[99]

The purpose of this study was to add high affinity gadolinium binding moieties to the surface of the three protein cage architectures. The first approach was to genetically incorporate a nine residue peptide sequence, from the $Ca^{2+}$ binding protein calmodulin, as a genetic fusion to the N-terminus of the CCMV subunit (CCMV-CAL), to the C-terminus of the HSP subunit (HSP-CAL). Characterization of the metal binding to the genetically engineered capsid was undertaken using FRET analysis. The second approach was to covalently attach the clinically relevant contrast agent DOTA-Gd to reactive lysine residues on CCMV via an NHS ester coupling reaction (CCMV-DOTA). Both $r_1$ and $r_2$ relaxivity data were measured as a function of field strength for the genetic and chemically modified protein cage contrast agents.

## Methods

### Engineering CCMV and HSP to Express the Metal Binding Sequence of Calmodulin

The SubE/R26C/K42R gene, a mutant of the coat protein of the CCMV capsid, cloned into the *Pichia pastoris* vector; pPICZA (Invitrogen) was used as the template.[100] QuikChange site-directed mutagenesis (Stratagene) using the primer; (5'cgaggaattcatgtctacagacaaagatggtgatggatggttagaattcgaagagggtgggggcgaagagaacgaggag aacac3'), and its reverse complement was used to insert the calmodulin DNA sequence into the N terminus of the coding region of the capsid protein. This modified CCMV protein was called CCMV-CAL.

The HSP gene cloned into the *E. coli* vector; pET-30a (Novagen) was used as the template. QuikChange site-directed mutagenesis (Stratagene) using the HSP template along with the primer;

(5'ccaaaggcagaatcctcaattaagaaaggagacaaagatggagatggttggttagaattcgaagaaggataaggatccga attcgagctccgtcgacaagcttg3'), was use to insert the DNA sequence into the coding region of the C terminus of the protein subunit. This modified HSP protein was called HSP-CAL. The mutangenized vectors from CCMV and HSP were both confirmed by DNA sequencing (Applied Biosystems).

### Expression and Purification of Protein Cages

The mutagenized CCMV capsid protein gene was expressed and purified in a *Pichia pastoris* heterologous protein expression system as previously described.[100] High

levels of coat protein expression were induced and yielded assembled viral protein capsids devoid of nucleic acid.

These viral capsids were purified to near homogeneity by lysis of cells, followed by ion exchange chromatography. Size exclusion chromatography (SEC) was used to further purify the capsid and to eliminate any aggregates or subunit disassembly products potentially present in the samples (Superose 6, Amersham Biosciences; 50 mM HEPES, pH 6.5). Both ion exchange and SEC were performed on an Amersham Akta purifier FPLC.

The twelve residue binding sequence and three glycine residues, shown in (Fig. 28), replaced the original residues 4-18 of the SubE/R26C/K42R CCMV mutant. This replacement was confirmed at the protein level by liquid chromatography / mass spectrometry (LC/MS) of the purified protein (Agilent Technologies 1100 LC system coupled to an Esquire 3000 ion trap mass spectrometer, Bruker Daltonics). The theoretical mass was calculated by considering loss of the N-terminal methionine and acetylation of the second residue (serine). Protein concentration was determined by the absorbance at 280 nm ($\varepsilon$ = 29280 $M^{-1}$ $cm^{-1}$ for CCMV-CAL, $\varepsilon$ values were calculated by inserting amino acid sequences into the ProtParam tool at: (http://ca.expasy.org/tools/protparam.html)).

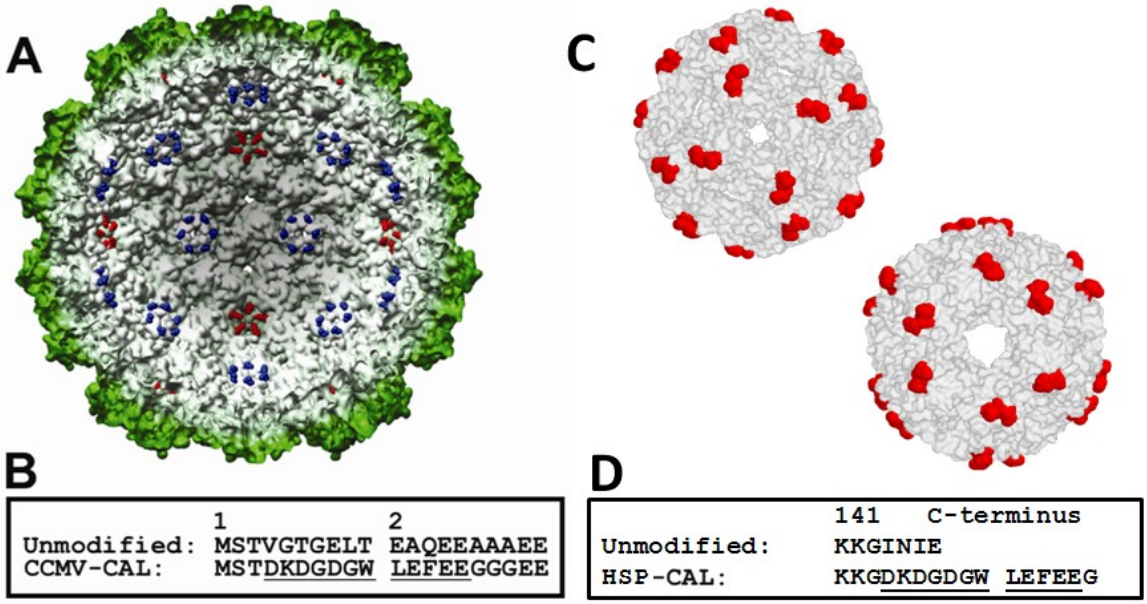



Figure 28 Inside view of CCMV
Inside view of (A) the viral capsids of CCMV and view of (C) HSP about the 4-fold and 3-fold axis. Blue highlights are residue 27 in the 6-fold environment of CCMV. Red highlights are residue 42 in the 5-fold environments of CCMV. (B) The first twenty amino acids are shown for both the unmodified viral subunit and genetically modified subunit. The underlined residues are responsible for metal binding. (D) the C-terminus is shown for the unmodified HSP and HSP-CAL.

The mutagenized HSP coat protein gene was expressed in an *E. coli* heterologous protein expression system. High levels of coat protein was expressed and yielded assembled protein cages. These protein cages were purified and characterized in a similar manner as the CCMV-CAL protein cage (above). The twelve residue binding sequence shown in bold and one glycine shown here; "DKDGDGWLEFEEG" replaced residues 144-147 and added an additional nine residues to the HSP subunit. This modified HSP protein cage was called "HSP-CAL".

Protein concentration was determined by the absorbance at 280 nm ( $\varepsilon$ = 13940 $M^{-1}$ $cm^{-1}$ for HSP-CAL and $\varepsilon$ = 29280 $M^{-1}$ $cm^{-1}$ for CCMV-CAL, $\varepsilon$ values were

calculated by inserting amino acid sequences into the ProtParam tool at: http://ca.expacy.org/cgi-bin/protparam).

Synthesis of CCMV- DOTA Particles

The following buffers were used in the synthesis of CCMV-DOTA; labeling buffer (100mM HEPES, 100mM NaCl, pH 7.2), and storage buffer (100mM HEPES, 100mM NaCl, pH 6.5). A lysine reactive form (NHS-ester) of the metal chelator DOTA was used in the synthesis (Macrocyclics, B-280). (Fig. 29) outlines the general reaction scheme. The K42R mutant of the CCMV (0.5 – 3mg/mL, 25 - 150μM subunit) virus particle was purified from infected plants (as previously described) and dialyzed into 500mL of labeling buffer for 3 hours.[101] The concentration of plant virus was calculated by multiplying the $A_{260}$ by 6.4 to yield a concentration of CCMV subunit in μM units. A concentration of 1 to 2 mg/mL (~50 to 100μM subunit) was typically used in the reaction. A 20X (mole:mole) of NHS-ester DOTA was added next. The pH was maintained at 7.0 by additions (1-5μL) of 0.5M NaOH. The reaction mixture was monitor by LC/MS (both Standard and Nano Aquity LC systems and both Q-Tof Micro and Q-Tof Premier mass spectrometers were used). SEC and reverse phase separation techniques were used. The deconvolution program MaxEnt1 (Waters) was used to determine the percent of subunits with DOTA covalently linked to them. Equation 17 was used to approximate the average labeling of subunits with DOTA, where $D$ is the number of DOTA(s) attached to the subunit and $I_D$ is the intensity of the ion corresponding to a subunit with $D$ DOTA(s) attached.

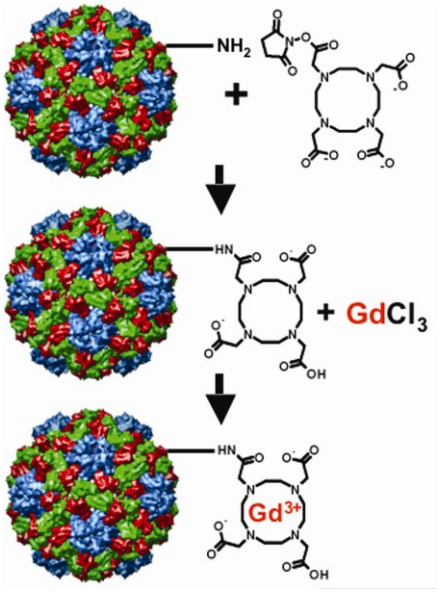



Figure 29 CCMV DOTA-Gd reaction
The reaction scheme to attach DOTA-$Gd^{3+}$ to the CCMV viral capsid. Endogenous lysines on the viral capsid are reacted to a DOTA / NHS conjugation. Next $GdCl_3$ is added to produce a viral capsid conjugated with $Gd^{3+}$ ions.

$$\frac{AverageDOTA}{Subunit} = \frac{\sum_{0}^{D} DI_D}{\sum_{0}^{D} I_D}. \qquad \text{eq. 17}$$

The reaction was allowed to proceed for 2 hours at 25°C. At that point, LC/MS analysis revealed that the majority of NHS-DOTA reactant was hydrolyzed. The reaction was repeated (with additions of NHS-ester DOTA) until there was, on average, one DOTA covalently attached per subunit. Unreacted DOTA was removed by dialyzing the reaction mixture into labeling buffer. Next the CCMV-DOTA was dialyzed into labeling buffer with 10X $GdCl_3$ (moles $Gd^{3+}$ : moles subunit) and a pH of 7.0. The progression of

the metal loading onto the CCMV-DOTA was monitored by LC/MS. Free $Gd^{3+}$ was separated from the CCMV-DOTA-Gd by dialyzing the labeled capsid into storage buffer that contained 5mM EDTA. Then multiple dialysis steps were performed into storage buffer. Alternatively the CCMV-DOTA-Gd was separated from free $Gd^{3+}$ by running the reaction product over SEC using storage buffer as the eluent.

Characterization of the Modified Protein Cages

Characterization of both purified capsids was performed by SEC, dynamic light scattering (DLS), transmission electron microscopy (TEM) and LC/MS. Typical data from these characterization techniques is shown in (Fig. 30). DLS analysis was performed on a ZetaPlus (Brookhaven Instruments). TEM (Leo 912AB) of negatively stained samples (1% uranyl acetate) was performed. The synthesis of HSP-CAL was similar to that of CCMV-CAL.

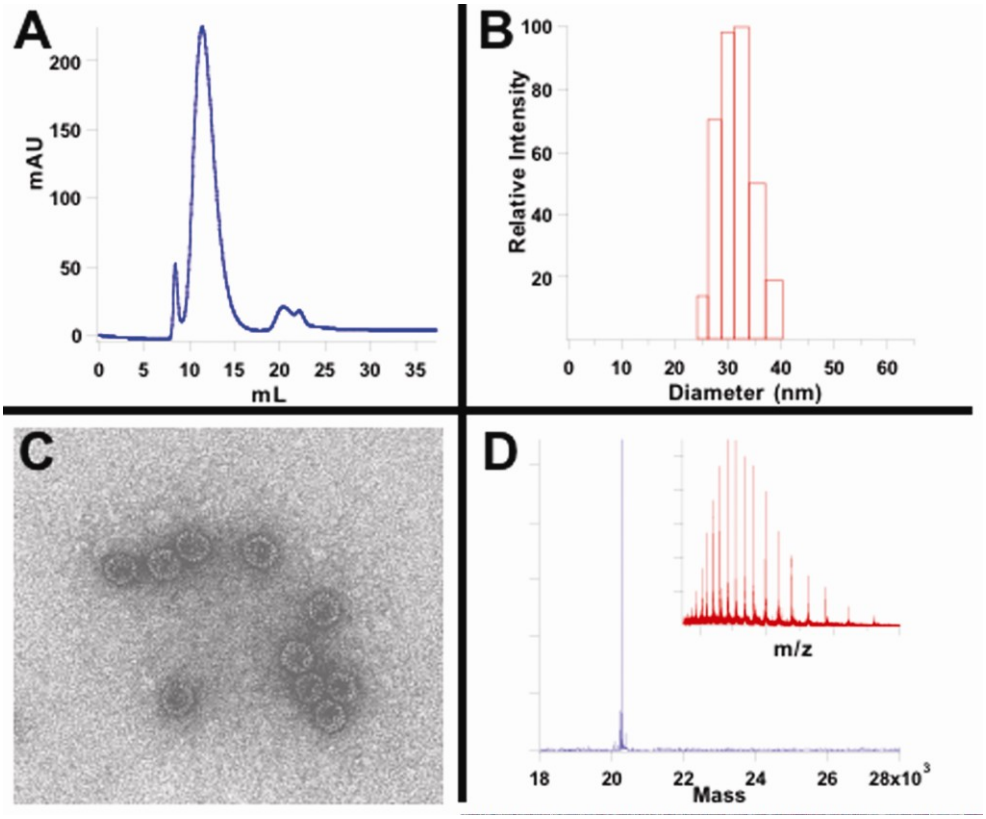



Figure 30 Characterization of CCMV-CAL
Data from routine virus capsid characterization of CCMV-CAL or CCMV-DOTA. (A) Size exclusion chromatogram showing the 280nm absorbance of the three main components. The small left peak is aggregated virus capsids. The large middle peak is intact capsids eluting at the correct retention volume and the peaks to the right are buffer molecules. (B) Dynamic light scattering indicates a viral capsid mean diameter of 30nm. (C) Transmission electron micrograph of negatively stained viral particles. (D) The deconvoluted spectrum shows the correct subunit mass. Inset is the raw electrospray mass spectrum of viral capsid subunits.

Characterization DOTA-Gd Reaction Site in CCMV-DOTA-Gd

After removal of unbound $Gd^{3+}$ from a solution of CCMV-DOTA-Gd, inductively coupled plasma mass spectrometry (ICP-MS) analysis was used to determine the total concentration of $Gd^{3+}$ bound to the viral capsid (7500, Agilent Technologies). Protein concentrations of the CCMV-DOTA-Gd samples were determined by the BCA assay

(Pierce). Protease digestion (Trypsin Gold, Promega) of CCMV-DOTA-Gd was carried out (1mg/mL protein subunit, 1:200 subunit:trypsin, 37$^{O}$C, 12 hours) and LC/MS/MS was performed on the digested sample (nanoAquity coupled to a Q-Tof Premier). Data was analyzed with PLGS2 and MassLynx (Waters).

Generation of Binding Isotherms for CCMV-CAL and HSP-CAL

The lanthanides ions $Gd^{3+}$ and terbium ($Tb^{3+}$) are known to bind with similar affinities, however, $Gd^{3+}$ does not undergo fluorescence resonance energy transfer (FRET) and therefore cannot be used to probe metal binding. Therefore $Tb^{3+}$ was used as a $Gd^{3+}$ mimic to study metal binding. All fluorescence experiments were performed on a (SPEX Fluorolog) spectrophotometer at 25$^{o}$C. The fluorescence spectrum of CCMV-CAL-Tb and HSP-CAL-Tb were measured ($\lambda$ max = 340 nm) following excitation at 295nm, with a $Tb^{3+}$ emission max was near 550nm. The 340nm peak was recorded by scanning from 305nm to 575nm in 1nm steps. The 550 peak was monitored by scanning from 525nm to 575nm in 0.2nm steps. Excitation and emission slit widths were set to 4 and 8 nm, respectively. Fluorescence was measured on 500µL solutions of viral capsid to which metal ion was added in 5 – 20 µL increments from 10 - 500µM solution standards. Subunit protein concentrations of 0.05µM and 0.1µM for the CCMV-CAL and HSP-CAL mutants were used.

Calculation of $K_d$

In analyzing the data for $Tb^{3+}$ binding to the CCMV-CAL capsid, we have assumed that when metal ions are titrated to the CCMV-CAL capsid, they first bind completely to the inserted sites. Once the inserted sites are completely occupied then the endogenous sites start to bind the additional metal ions that are added. Equation 18 was used to analyze the data where $\theta$ is the fraction of capsid subunits with bound $Tb^{3+}$, $[Tb^{3+}{}_{Free}]$ is the free $Tb^{3+}$ concentration, and $K_d$ is the dissociation constant for subunit binding $Tb^{3+}$ ions. HSP-CAL had only one binding site so the $K_d$ calculation was less convoluted but similar CCMV-CAL.

$$\theta = \frac{\left[ Tb^{3+}_{Free} \right]}{[Tb^{3+}_{Free}] + K_d} \qquad \text{eq. 18}$$

Baseline correction was performed on the 550nm $Tb^{3+}$ spectra. To find the maximum intensity of the 550nm peak a Gaussian function was individually fit to each spectra. The maximum intensity of the 550nm peak ($I_{550}$) vs. the total $Tb^{3+}$ in solution ( $[Tb^{3+}{}_{total}]$ ) was plotted and fit to Equation. 19 where $K_{d\_Initial}$ and $I_{550Limit}$ were the fitting parameters.

$$I_{550} = I_{550Limit}\left( \frac{[Tb^{3+}_{Total}]}{[Tb^{3+}_{Total}] + K_{d\_Initial}} \right) \qquad \text{eq. 19}$$

Fractions bound terms ($\theta$) were calculated by dividing $I_{550}$ values by $I_{550Limit}$ determined from Equation 19. Values for $[Tb^{3+}{}_{Free}]$ were then calculated by Equation. 20.

$$[Tb^{3+}_{Free}] = [Tb^{3+}_{Total}] - (\theta[subunit_{Total}]) \quad \text{eq. 20}$$

A plot of $\theta$ vs. $[Tb^{3+}{}_{free}]$ was then fit with the Equation. 18 and $K_d$ was determined along with an error associated with the fit.

Stiochiometric Titration of CCMV-CAL

Terbium ions ($Tb^{3+}$) were used as $Gd^{3+}$ ion mimics. $Tb^{3+}$ titrations were performed with the condition of [subunit] >> $K_d$ to determine the number of $Tb^{3+}$ bound per CCMV-CAL subunit. Two titrations were performed in which the protein concentration used was 2.6 μM and 10μM. The capsid was titrated to ~20μM total $Tb^{3+}$ in both experiments. The data, from both the beginning and end portions of the titration, were fit to linear functions.

Relaxometry and $Gd^{3+}$ Quantitation

For relaxometry experiments, fully assembled CCMV-CAL (60μM subunit) containing 200 μM $GdCl_3$ was prepared in pH 6.5 buffer (50mM HEPES, 150 mM NaCl). As a control, 200μM $GdCl_3$ was prepared in the same buffer. CCMV-DOTA-Gd was prepared in pH 6.5 buffer (100mM HEPES, 100 mM NaCl) with a subunit concentration of 101 μM (determined by the BCA assay) and 34μM $Gd^{3+}$ (determined by ICP-MS). Using a custom-designed variable field relaxometer, $T_1$ relaxivity was measured using a saturation recovery pulse sequence with 32 incremental τ values. The range of Larmor frequencies was 2–62 MHz (0.05–1.5 T) and the measurements were carried out at a temperature of 23°C. $T_1$ values were determined by fitting data into Equation. 21 with *A* and *B* as fitting parameters.

$$f(Seconds) = A\left(1 - e\left(\frac{-Seconds}{\frac{1}{T1}}\right)\right) + B \qquad \text{eq. 21}$$

$T_2$ was measured using a CPMG pulse sequence with 500 echoes and an interecho time of 2 ms. $T_2$ values were determined by fitting data into Equation. 22 with *A*, *B* and *N* as fitting parameters.

$$f(Seconds) = \sqrt{A\left(e\left(\frac{-2Seconds}{\frac{1}{T2}}\right) + N\right) + B} \qquad \text{eq. 22}$$

Since the $r_1$ and $r_2$ relaxivities are expressed in units of ($mM^{-1}$ of bound $Gd^{3+}$ * $seconds^{-1}$), it was necessary to determine the mM concentration of bound $Gd^{3+}$. The calculation of the fraction of CCMV-CAL with bound $Gd^{3+}$ turned into an approximation since this capsid contains two types of binding sites. First the parameters for the higher affinity binding site or "CAL" were input into Equation. 23. The concentration of binding sites [*BS*], [$Gd^{3+}{}_{Total}$] and the $K_d$ are all values input into this equation.

$$\theta = \frac{([BS] + [Tb^{3+}_{Total}] + K_d - \sqrt{([BS] + [Tb^{3+}_{Total}] + K_d)^2 - (4[BS][Tb^{3+}_{Total}])}]}{2[BS]} \qquad \text{eq. 23}$$

From this, the fraction bound term for the CAL binding site ($\theta_{CAL}$) could be determined and was further used (Equation. 24) to calculate the [$Gd^{3+}{}_{Free}$] left after the CAL site was maximally bound.

$$Gd^{3+}_{Free} = Gd^{3+}_{Total} - ([CAL] \cdot \theta_{CAL}) \qquad \text{eq. 24}$$

The fraction of $Gd^{3+}$ bound to the endogenous site ($\theta_{Endogenous}$) was calculated by setting [$Gd^{3+}{}_{Total}$], which was input into Equation. 23, equal to [$Gd^{3+}{}_{Free}$], which was determined by Equation. 24. $K_d$ and [$BS$] values for the endogenous site were input into Equation. 23 to determine the fraction bound for the endogenous site ($\theta_{Endogenous}$). Finally, the faction bound terms for both the endogenous site ($\theta_{Endogenous}$) and the engineered site ($\theta_{CAL}$) were multiplied by the mM concentration of their respective binding sites and then added together resulting in an mM concentration of total bound $Gd^{3+}$. The calculation $Gd^{3+}$ bound to HSP-CAL was less convoluted than for CCMV-CAL since HSP-CAL had only one binding site.

## Results

### CCMV and HSP Modified Architectures

The CCMV and HSP protein cages (Fig. 27) have been modified for enhanced $Gd^{3+}$ binding, using two complementary approaches, while maintaining the advantages of the large molecular platform. In the first approach, a $Gd^{3+}$ binding peptide from calmodulin, was genetically introduced onto the N-terminus of the CCMV and HSP subunits (CCMV-CAL-Gd and HSP-CAL). CCMV-CAL has an increased affinity for $Gd^{3+}$ in comparison with wild type CCMV and metal binding functionality was added to the HSP protein cage. In a second approach, DOTA-Gd was conjugated to CCMV resulting in high affinity $Gd^{3+}$ binding and imparting highly efficient relaxivity properties to the CCMV capsid (CCMV-DOTA-Gd).

Genetic Modification of CCMV and HSP – Attachment of Calmodulin Peptide

A peptide sequence from the $Ca^{2+}$ binding portion of the protein calmodulin (DKDGDGWLEFEEGGG) was genetically fused to the N-terminus of CCMV and HSP (Fig. 28). Interestingly, this construct with a nine residue peptide incorporated as an N-terminal fusion did not disrupt the ability of CCMV to self assemble as shown by SEC, DLS or TEM analyses (Fig. 30). This was also the case for the C-terminal insertion in HSP-CAL. The mutation of the coat protein subunit gene was confirmed by DNA sequencing. LC/MS of the purified viral capsid subunit produced an experimental average mass of 20234 Da compared to a calculated average mass of 20232 Da for the CCMV-CAL and 17460 (experimental) compared to 17461 (calculated) for HSP-CAL confirming the identity of these recombinant proteins.

$K_d$ Determination of CCMV-CAL and HSP-CAL

The metal binding affinity of the CCMV-CAL and HSP-CAL mutants were probed by FRET using excitation of endogenous tryptophan residues. The lanthanide ions $Gd^{3+}$ and terbium ($Tb^{3+}$) are known to bind with similar affinities and both show preference for $Ca^{2+}$ binding sites in proteins although with significantly higher affinities than $Ca^{2+}$ binding.[90, 102] Titrations of the CCMV-CAL and HSP-CAL mutants with increasing $Tb^{3+}$ revealed a decrease in the tryptophan fluorescence (340 nm) and concomitant increase in the $Tb^{3+}$ fluorescence (at 550 nm) indicating energy transfer between these sites (Fig. 31). The complete data set was fit to Equation. 18 and an average $K_d$ of 82 ± 14 nM for $Tb^{3+}$ binding to CCMV-CAL and a $K_d$ of 1 μM for $Tb^{3+}$ binding to HSP-CAL were determined. This indicates an enhancement in the metal

binding affinity of 232 fold over binding to endogenous sites in the wild type CCMV ($K_d$ of 19,000 nM for $Tb^{3+}$).[98]

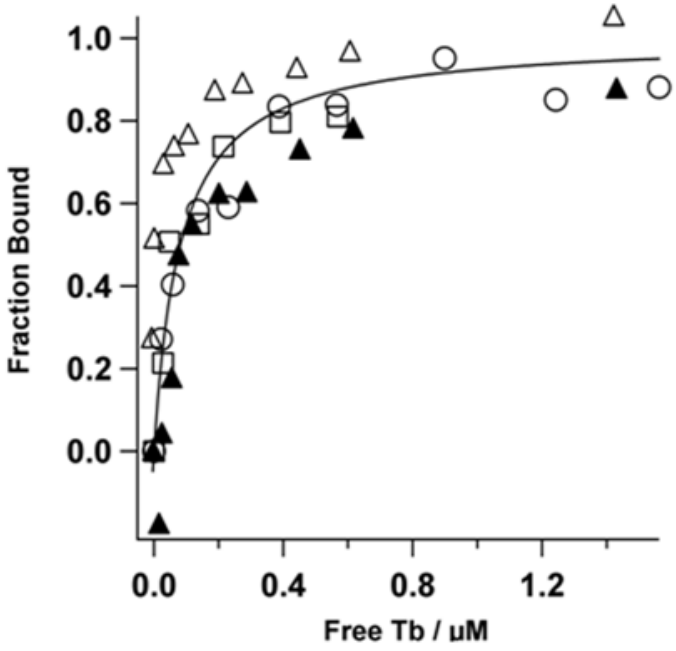



Figure 31 Bind isotherms for CCMV-CAL
Binding isotherms for CCMV-CAL binding $Tb^{3+}$ ions. The open triangle, square and circle represent three replicate experiments for the increase of the 550nm peak. The solid triangle represents the decrease in the 340nm peak from the first experiment. All four data sets were used for the fit.

Stoichiometric Titrations of CCMV-CAL

A stoichiometric titration of $Tb^{3+}$ was performed to determine the number of ions bound per CCMV-CAL subunit. When the normalized fluorescence intensity was plotted against the ratio of subunit:$Tb^{3+}$ ([subunit] / [$Tb^{3+}{}_{Total}$]) as shown in (Fig. 32), the data shows two distinct regions. The fluorescence response during the first part of the titration ( [binding site] >> [metal] ) increases linearly with $Tb^{3+}$ addition. The second linear

portion of the titration ( [metal >> [binding site] ) shows a relatively constant fluorescent response with a smaller slope. The x-intersection of the fits to each of these two regions indicates the point at which the binding sites are maximally occupied. Our data indicate a value of 7.3 ± 1.1 ([subunit] / [$Tb^{3+}_{Total}$]) for this intersection. At the end of the $Tb^{3+}$ titration, there is one $Tb^{3+}$ ion for every 7.3 ± 1.1 subunits or each CCMV capsid has approximately 25 metal ions bound at the introduced peptide sites.

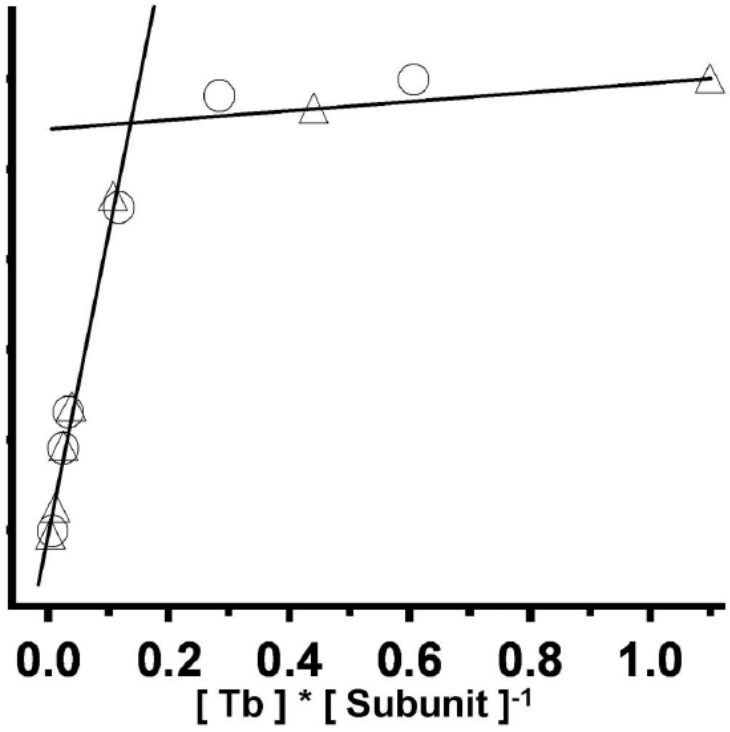



Figure 32 A stoichiometric titration of CCMV CAL with $Tb^{3+}$ ions
Two replicate data sets are shown here, one with circles and the other with triangles.

Chemical Modification of CCMV – Attachment of DOTA-Gd

DOTA-Gd was covalently attached to the surface of CCMV via reaction with surface exposed lysine residues (Fig 29). A deconvoluted mass spectrum after a routine CCMV-DOTA-Gd synthesis and removal of unbound $Gd^{3+}$showed a distribution of 0 to 3 DOTA-Gd per subunit (Fig. 33). The addition of a single DOTA-Gd added an experimental mass of 544 Daltons to the subunit molecular weight, which corresponds well with the theoretical value of 543.7 Daltons per DOTA-Gd. It was qualitatively observed by LC/MS that up to two DOTA on average per subunit produced a stable, fully assembled capsid yielding up to 360 DOTA-Gd per capsid.

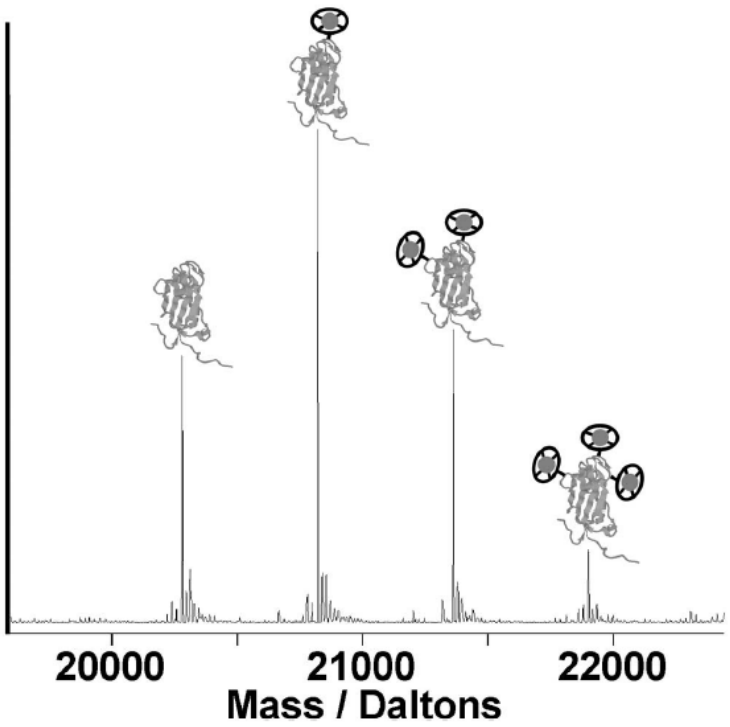



Figure 33 SEC-MS of CCMV-DOTA-Gd
Typical deconvoluted electrospray mass spectra of CCMV capsid subunits. CCMV capsids were reacted with NHS-DOTA and then $GdCl_3$. Unlabelled subunits and subunits with one to three DOTA-Gd were detected.

Using mass spectrometry we were able to map the lysines which are the sites of labeling. LC/MS/MS analysis indicates that DOTA-Gd is primarily attached through Lys 8 and Lys 45 (shown in Fig11). The distribution of DOTA-Gd labeling on CCMV-DOTA-Gd suggests that there is not complete occupancy of either Lys 45 or Lys 8 within the capsid and that additional unidentified residues are also labeled. However, residues 8 and 45 are likely the most prevalently labeled lysines in CCMV. Figure 34 shows the inside view of the CCMV capsid with DOTA-Gd modeled onto residue 45. The position of lysine 45 in the structure suggests that the attached DOTA-Gd resides on the interior of the capsid. Residue 8 is not shown since the N-terminus is disordered in the X-ray crystal structure of CCMV and its position is therefore uncertain.

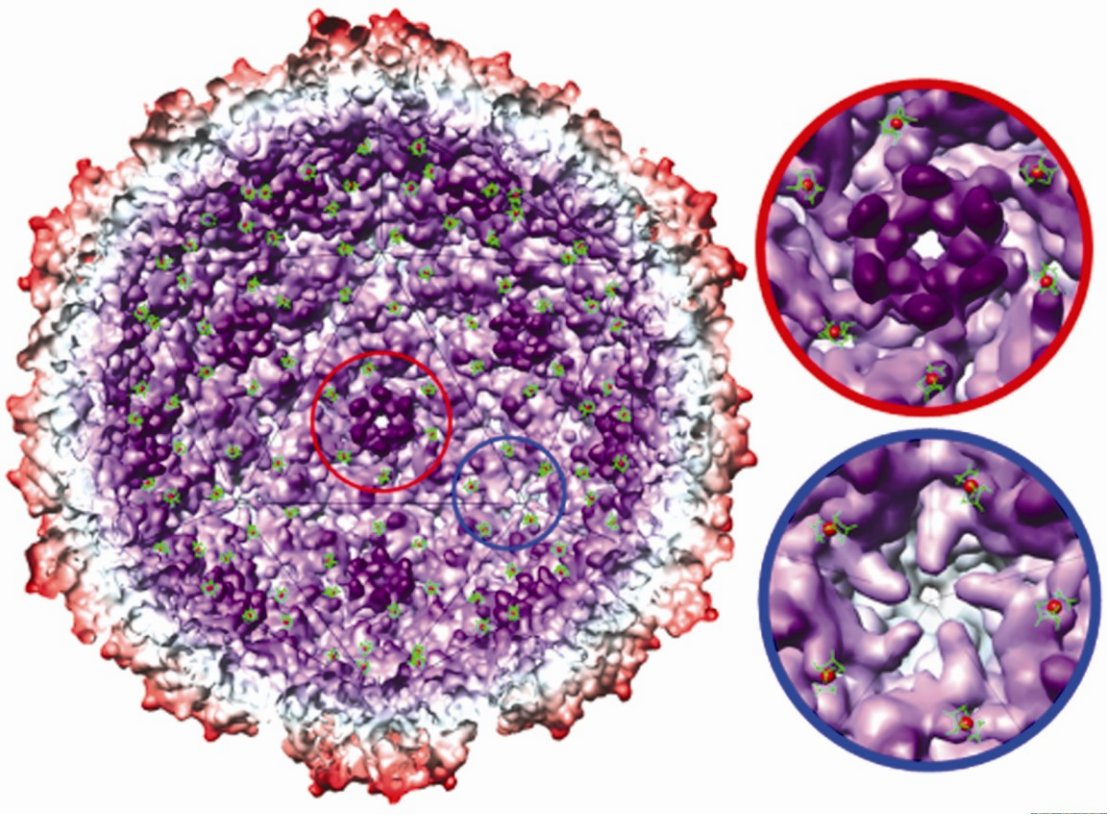


Figure 34 Model Structure of CCMV containing DOTA-Gd
Cutaway view of the interior of CCMV. DOTA-Gd was modeled to be attached through Lys 45. The red highlighted view is a close up of Lys 45 in the 6-fold environment while blue highlighted view is of Lys 45 in a 5-fold environment.

Relaxometry of CCMV-CAL-Gd and CCMV-DOTA-Gd

Highly efficient $T_1$ and $T_2$ relaxivity properties were observed in both capsids (CCMV-CAL-Gd and CCMV-DOTA-Gd). Table 1 summarizes $r_1$ and $r_2$ values for $Gd^{3+}$ / virus systems including; wild type CCMV-Gd, CCMV-CAL-Gd, CCMV-DOTA-Gd and MS2-DTPA-Gd.[81, 90] The ionic $r_1$ and $r_2$ trends are CCMV-CAL-Gd > CCMV-Gd > CCMV-DOTA-Gd > MS2-DTPA-Gd. The $r_1$ and $r_2$ trends relative to the particle are CCMV-CAL-Gd > wild type CCMV-Gd > MS2-DTPA-Gd > CCMV-DOTA-Gd . In the genetic approach, the CCMV-CAL mutant exhibited approximately the same $T_1$ and $T_2$

relaxivity as wild type CCMV-Gd previously reported.[90] Figure 35 A and B shows the similarities between the $T_1$ and $T_2$ relaxivity values of wild type CCMV-Gd and CCMV-CAL-Gd for field strengths ranging from 0.2 to 1.5 T. This construct not only has an increased affinity for $Gd^{3+}$, but also maintains the very high relaxivity required for clinically relevant contrast agents.

Obtaining relaxivity data for the HSP-CAL construct was problematic due to protein instability at higher concentrations of the protein cage and $Gd^{3+}$ which were required for the relaxivity experiments. Therefore relaxivity results are not present here for this construct.

The increased $T_1$ and $T_2$ relaxivity afforded by the chemical conjugation of DOTA-Gd to the capsid of CCMV is shown (Fig. 35 A and B). This conjugation resulted in increased $T_1$ and $T_2$ relaxivities by a factor of ten relative to free DOTA-Gd. By chemically attaching DOTA-Gd to the CCMV capsid we have engineered a viral capsid with clinically relevant binding and high relaxivity.

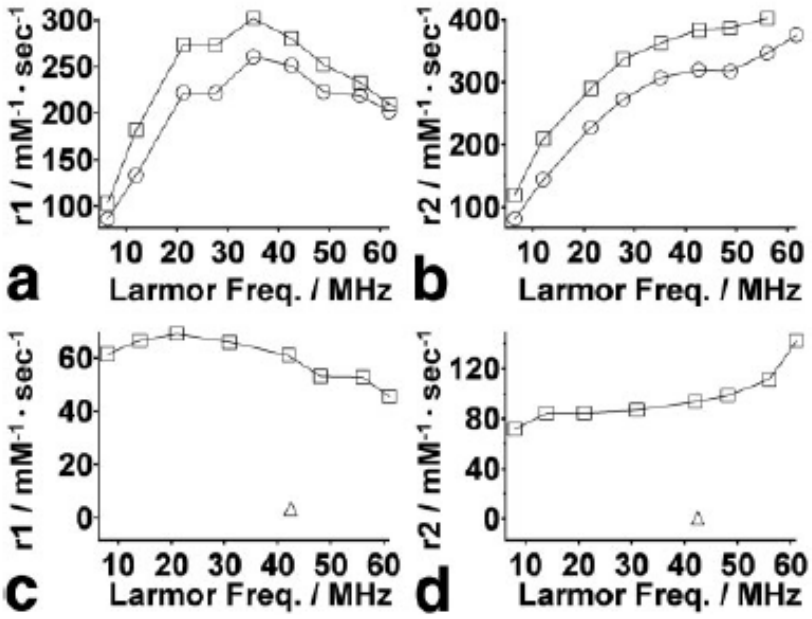



Figure 35 Relaxivity measurements of CCMV-CAL-Gd and CCMV-DOTA-Gd
(A) $r_1$ values for two CCMV capsids conjugated with $Gd^{3+}$ ions, wild type CCMV-Gd (circles) and CCMV-CAL-Gd (squares). (B) $r_2$ values for two CCMV capsids conjugated with $Gd^{3+}$ ions, wild type CCMV-Gd (circles) and CCMV-CAL-Gd (squares). (C) $r_1$ values for DOTA-Gd (triangle) and CCMV-DOTA-Gd (squares). (D) $r_2$ values for DOTA-Gd (triangle) and CCMV-DOTA-Gd (squares).

The $r_1$ dependence on field strength, for both wild type CCMV-Gd and CCMV-CAL-Gd, varies in manner typical of nanoscale systems with a relaxivity maximum near 1 T.[103] CCMV-DOTA-Gd has $r_1$ field strength dependence more similar to small molecule systems with a maximum near 0.5 T. All CCMV / Gd systems exhibit a positive correlation between field strength and $T_2$ relaxivity values as expected.

## Discussion

The major achievement of this study is demonstration that the CCMV viral capsid is a robust platform for development of MR contrast agents. Specifically we have shown that a genetically engineered CCMV capsid binds $Gd^{3+}$ (CCMV-CAL-Gd) with a significantly higher affinity than wild type CCMV. In a second approach, the CCMV capsid has been chemically modified resulting in the attachment of multiple DOTA-Gd (CCMV-DOTA-Gd). These amine conjugated DOTA-Gd groups have been previously shown to bind with a dissociation constant of $10^{-20}$ which corresponds to $10^{15}$ improvement in metal affinity compared to wild type CCMV-Gd. Both these nanoparticle contrast agents are water soluble and have high ionic and particle relaxivities. These viral capsids have multiple $Gd^{3+}$ ions attached per capsid resulting in the potential for high contrast associated with tissue specific localization. Localization of the viral capsid could be accomplished by active targeting or by passive means.

It is surprising that we observed sub-stoichiometric binding of $Gd^{3+}$ to the introduced CAL peptide in the CCMV-CAL-Gd construct. The data suggest that only 25 out of 180 sites bind $Gd^{3+}$ at saturating conditions. A likely possibility is that not all the introduced sites are accessible to bind metal ions due to different chemical environments of the CAL peptides. It is also worth mentioning that the N-terminus is grouped in two environments, twenty pseudo 6-fold and twelve 5-fold environments (Fig. 28). The combination of these groupings yields an average of 5.6 subunits per N terminus grouping. This value is reasonably close to the experimentally determined value of 7.3 ± 1.1 CAL peptides per bound metal ion. It is possible that single metal ions are bound by

multiple CAL peptides grouped at these 5 and pseudo 6-fold environments. The dissociation constant of 82 ± 14 nM for CCMV-CAL-$Gd^{3+}$ in comparison to dissociation constants in the range of $10^{-20}$ molar for approved contrast agents indicates that the CAL peptide binds $Gd^{3+}$ too weakly for clinical application. Therefore, this approach of genetically attaching metal binding peptides to the CCMV capsid is not likely to be as fruitful as chemical modification approaches.

It was expected that all three CCMV / Gd constructs (CCMV-Gd, CCMV-CAL-Gd, CCMV-DOTA-Gd) would have similar relaxivity values since size was thought to be the dominate factor in determining relaxivity rates. However, CCMV-DOTA-Gd has ionic relaxivity values that are approximately 25% of the values for CCMV-Gd and CCMV-CAL-Gd, indicating that factors other than size can influence the relaxivity of these capsids. The endogenous $Gd^{3+}$ binding pocket in wild type CCMV is at the interface of three subunits and contains side chains from each subunit. The result of this inter-subunit binding pocket is that the overall motion of the $Gd^{3+}$ ion is identical to the overall motion of the entire capsid and there is no additional motion of the $Gd^{3+}$ ion. This is in contrast to the DOTA bound $Gd^{3+}$ of CCMV-DOTA-Gd in which the flexibility of the linker connecting Gd to the capsid adds additional motion to the $Gd^{3+}$ ion. In addition, LC/MS/MS data of a CCMV-DOTA-Gd trypsin digest indicates that the DOTA-Gd is primarily attached on the N-terminus end of the capsid subunit through Lys 8 and Lys 45. It is known that the N terminus of CCMV is mobile and can occupy both the interior and exterior of the CCMV capsid architecture.[32, 104] This increased local mobility of the region of the viral capsid that is labeled with DOTA-Gd may cause a

reduction in relaxivity. EPR studies of CCMV labeled with another small molecule concluded that there was local mobility within the spin label itself which decreased the rotational correlation time when compared to the predicted value for the CCMV capsids.[105-109] This local mobility of the label may also exist in CCMV-DOTA-Gd that would lead to an additional decrease in relaxivity rates. Local mobility from flexible regions of the protein as well as the DOTA-Gd itself could explain the lower relaxivity values of CCMV-DOTA-Gd compared to CCMV-Gd and CCMV-CAL-Gd.

The consideration of the gadolinium ligands offers another explanation for the relaxivity differences between CCMV-DOTA-Gd and the two other CCMV / $Gd^{3+}$ systems (CCMV-Gd and CCMV-CAL-Gd). Multiple water molecules are thought to be ligands of $Gd^{3+}$ in both CCMV-Gd and CCMV-CAL-Gd in contrast to the DOTA-Gd system where a single water molecule is coordinated by the chelated $Gd^{3+}$ atom. The increased number of bound waters makes CCMV-Gd and CCMV-CAL-Gd more efficient at relaxation of water molecules compared to CCMV-DOTA-Gd. Also, it was found that a peptide similar to the genetically fused sequence used in this work, resulted in unexpectedly high relaxivity values. These authors suggested that the high relaxivity values were a result of having the $Gd^{3+}$ coordinated primarily with oxygen atoms since oxygen ligand systems resemble the aqua $Gd^{3+}$ complex which has near optimal lifetimes for bound water molecules.[110-112] This ideal oxygen environment is found in CCMV-Gd and CCMV-CAL –Gd, whereas in the DOTA system one of the four carboxylate groups is converted to an amide bond in the conjugation reaction. This conversion increases the lifetime of the bound water molecule from approximately 250 nanoseconds to greater

than 1 microsecond which is further from the theoretical ideal lifetime of 20-30 nanoseconds.[113] The larger value for the lifetime of bound water in the DOTA system probably accounts for some reduction of the relaxivity rate of CCMV-DOTA-Gd compared to CCMV-CAL-Gd.

Another possible example of how local mobility can affect relaxivity rates is revealed by comparing two different, $Gd^{3+}$ chelated, viral capsid systems (CCMV-DOTA-Gd and MS2-DTPA-Gd). Anderson and co-workers have attached up to 500$Gd^{3+}$ ions to the MS2 capsid yielding $T_1$ ionic relaxivity rates that are approximately three times lower than the $T_1$ ionic relaxivity rates for CCMV-DOTA-Gd (Table 1). This deviation is larger than expected since the relaxivity difference between DOTA and DTPA is small, the two capsids are the same size and in both systems the chelators are attached to endogenous lysines. Furthermore the isothiocyanate linked DTPA results in a lifetime for bound waters closer to the ideal value when compared to the NHS ester linked DOTA system, so the MS2/DTPA system should result in higher relaxivity rates.[103] A comparison of the length of the resulting linkers could explain the difference in observed relaxivity rates between the two cages. The NHS-ester used in the DOTA system results in a short linker with three rotatable bonds (Figure 36).

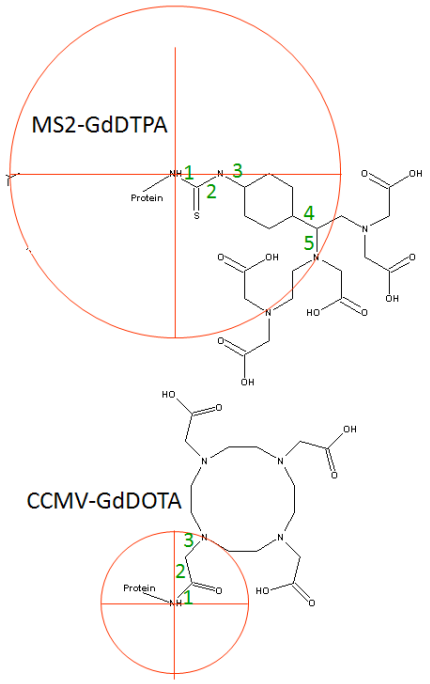



Figure 36 Difference in Gd-chelate linker length between MS2-GdDTPA and CCMV-DOTA-Gd
The radius of the circle in these two models corresponds to the distance from the side chain amine nitrogen to the closest coordinating atom, which is a nitrogen atom in both cases. There are five rotatable bonds in this distance in the case of MS2-GdDOTA and only 3 from CCMV-GdDOTA.

This is in contrast to the DTPA system where the linker is longer and has four rotatable bonds (The linker was measured from the lysine's amine nitrogen to the nitrogen that coordinates the $Gd^{3+}$ ion on either chelator.). The longer and more rotatable linker in the DTPA system may result in more mobility and this could account for the higher ionic relaxivity rates for CCMV-DOTA-Gd compared to MS2-DTPA-Gd. Finally, $Gd^{3+}$ loading of up to 360 $Gd^{3+}$ ions per capsid has been achieved in the CCMV-DOTA-Gd construct, which yields a calculated particle relaxivity values of $r_1$ = 16,560 and $r_2$ = 51,120 $sec^{-1}$ $mM^{-1}$ at 1.5T.

Table 4.1 Relaxivities of Capsid-$Gd^{3+}$ Constructs at 1.5 T

| Construct | Gd/particle[a] | r1 $(s^{-1})^*$ $(mM\ Gd^{3+})^{-1}$ | r2 $(s^{-1})^*$ $(mM\ Gd^{3+})^{-1}$ | r1 $(s^{-1})^*$ $(mM\ particle)^{-1}$ | r2 $(s^{-1})^*$ $(mM\ particle)^{-1}$ |
|---|---|---|---|---|---|
| Wild-type CCMV-Gd | 141[a] | 202 | 376 | 28,482 | 53,016 |
| CCMV-CAL-Gd | 172[a,b] | 210 | 402[c] | 36,120 | 69,144 |
| CCMV-DOTA-Gd | 61[a] | 46 | 142 | 2806 | 8662 |
| MS2-DTPA-Gd | 514[a] | 14 | NA | 7196 | NA |
| DOTA-Gd | 1 | 3[d] | 5[d] | 3[d] | 5[d] |

[a]Gd/particle (mole/mole) of the samples used in the relaxometry experiment.
[b]Calculated by setting CAL-sites/Gd = 7.3.
[c]Relaxivity measured at 1.3 T.
[d]Relaxivity measured at 1.0 T.
NA = Not available.

Protein cages offer advantages over other macromolecular contrast agents. These cages are generally more rigid than liposome or dendrimer systems. Protein cages are homogeneous whereas both dendrimers and liposomes are heterogeneous mixtures. Also the ability to accurately determine which residues are modified along with the availability of the near atomic resolution crystal structure provides information about the arrangement and environment of each modification within the protein cage structure. This information aids in the design of the agent since the spatial arrangement, structural rigidity and chemical environment of these modifications can be taken into account. In conclusion, the work presented here shows the potential of protein cages as MRI contrast agents and will direct the design of the next generation of these imaging agents.

# CHAPTER 5

## SUPRAMOLECULAR PROTEIN CAGE COMPOSITE MR CONTRAST AGENTS WITH EXTREMELY EFFICEINT RELAXIVITY PROPERTIES

### Introduction

Designing a magnetic resonance (MR) contrast agent (CA) on the scaffold of a protein cage has the potential for increasing image contrast of target tissues. While the high resolution of MR affords excellent anatomical detail, the ability to detect low copy number tissues is limited by the low sensitivity of the technique. CA are used to increase the sensitivity of MR to tissues of interest by accelerating the nuclear magnetic resonance (NMR) relaxation rate of water protons in the vicinity of the target tissue[79, 93, 111, 113] The utility of CA is currently limited by the inability to detect nanomolar concentrations of CA *in vivo*. Designing contrast agents that are more sensitive to MR scanners will allow for the detection of smaller and lower copy number targets.

Through a process known as paramagnetic relaxation enhancement (PRE), Gd ions carry out relaxation acceleration of protons and therefore are commonly used as $T_1$ weighted MR agents. Three parameters are important for the optimization of PRE, 1) the number of metal-bound water molecules (q), 2) the water exchange lifetime ($\tau_M$) and 3) the rotational correlation time of the Gd ion ($\tau_R$) (Figure 37).[79, 93, 99, 111, 113, 114] Since PRE is a spin-lattice effect, direct water interaction with the Gd ion is necessary and therefore ionic relaxivity increases with greater q values. It is also important for the water to remain bound to Gd long enough for relaxation to occur. However, when $\tau_M$ is too high,

the water molecule remains bound to Gd long after the spins have relaxed, this results in an effective decrease in the "duty cycle" of the Gd ion. The final parameter that is important for efficient PRE is $\tau_R$ and larger Gd complexes result in greater $\tau_R$ values, which yield more efficient PRE.

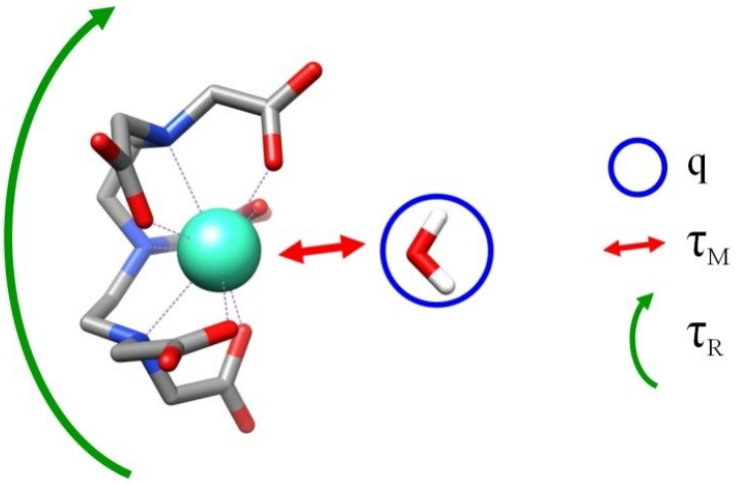



Figure 37. An schematic of DTPA-Gd and the important physical parameters that effect the relaxivity
Parameters that effect relaxivity including the number of water molecules that directly interact with the Gd ion (q), the exchange lifetime of the metal bound water molecule ($\tau_M$), and the rotational correlation time of the Gd chelate ($\tau_R$).

Protein cages, such as viral capsids, heat shock proteins (HSP), and ferritin-like proteins are good candidates as scaffolds to which chelated Gd groups can be anchored resulting in designer contrast agents with near ideal relaxivity properties.[85, 90, 98, 115-117] The large surface area and interior volume of protein cages can accommodate numerous Gd ions[2, 81, 90, 94, 118, 119] and this directly increases the *particle relaxivity* (relaxivity per 1mM of the particle). Also, by anchoring the Gd chelate to the protein cage the $\tau_R$ or

mobility of the Gd ion is decreased and this increases the *ionic relaxivity* (relaxivity per 1mM of Gd ion) (Figure 37[2, 94] Finally, targeting groups can be presented on the exterior of the protein cage to direct localization of the protein cages to receptors of interest resulting in novel MR probes to increase conspicuity of tissues of interest.[30, 87, 120-131]

To date, several groups have reported the utilization of viral capsids or other protein cages structures as MR contrast agents which can be divided into three categories, 1) endogenous metal binding sites[90], 2) genetic insertion of a metal binding peptide[2], 3) chemical attachment of small molecule chelates.[2, 81, 94, 118, 119] Of these three categories the first two are less significant from a clinical standpoint due to their lower Gd binding affinities since unbound Gd is toxic. The third category is the most common approach with three groups reporting similar results. These reports of protein cage – Gd based CA have inspired us to develop a novel protein cage branched polymer hybrid system that resulted in $T_1$ ionic relaxivities of 25$mM^{-1}$ $sec^{-1}$ and that contain up to 159 Gd per particle. These CA have high particle relaxivities (4,200$mM^{-1}$ $sec^{-1}$) and have extremely high relaxivity densities ($r_1$ / particle volume).

## Synthesis of the Branched Polymer – Protein Cage Hybrid

The synthesis of this novel Gd containing protein cage polymer hybrid begins with polymer growth on the interior of the protein cage. This step consists of synthesizing a branched polymer structure, via azide - alkyne "click" cycloaddition reactions in the interior cavity of the HSP cage from *Methanococcus jannaschii*, Figure 38.[3, 27] Recombinant HSP self-assembles in an *E. coli* expression system from 24

identical protein subunits forming a near spherical protein cage of 12 nm exterior diameter and 6.5 nm interior diameter.[25] The HSP cage is relatively porous and robust making it a good model system for internal modification.[3, 28, 132] The presence of eight, 3 nm pores at the 3-fold and six, 1.7 nm pores at the 4-fold axes allows free access of small molecules to the interior cavity. The cage is stable up to ~60 °C and in the pH range of 5-8. The G41C mutant of HSP introduces a unique cysteine residue at position 41 resulting in an internally directed initiation point for polymer growth.[3, 28, 132]

A branched polymer was grown in a stepwise fashion within the internal cavity of the HSP cage (Figure 38A). The polymer was initiated by reaction of N-propargyl bromoacetamide (bromo-alkyne) with the genetically inserted cysteine. This alkyne derivative is referred to as generation zero (G0.0). The exposed alkyne was subsequently reacted with 2-azido-1-azidomethyl ethylamine (diazido-amine) via a Cu(I) catalyzed 'click' reaction to yield G0.5. Exposed azide functional groups on G0.5 were subsequently 'clicked' with tripropargyl amine (tri-alkyne) to generate a branched structure (G1.0). Iterative stepwise reactions with diazido-amine and tri-alkyne were then undertaken to produce generations G1.5, G2.0, G2.5, G3.0 and G3.5.

These protein cage - branched polymer hybrid materials (HSP-BP) are thermally stable, to >120ºC, due to the internal cross-linking of the branched polymer. Every half generation (Gx.5) of this branched polymer adds additional derivatizable amine functional groups, up to 200 on the fully crosslinked cage. After the polymer synthesis the HSP-BP cages maintain their native shape and size distribution.[3] This increased stability dramatically expands the synthetic range and the potential *in vivo* utility of these

biological templates. Surface exposed amines on lysine residues were reacted, in one preparation, with N-hydroxysuccinimide acetate. This reaction "passivated" the endogenous amines and in doing so we hoped to decouple the relaxivity properties of Gd chelates that were attached to either the endogenous amines or to the amines introduced during the polymer growth.

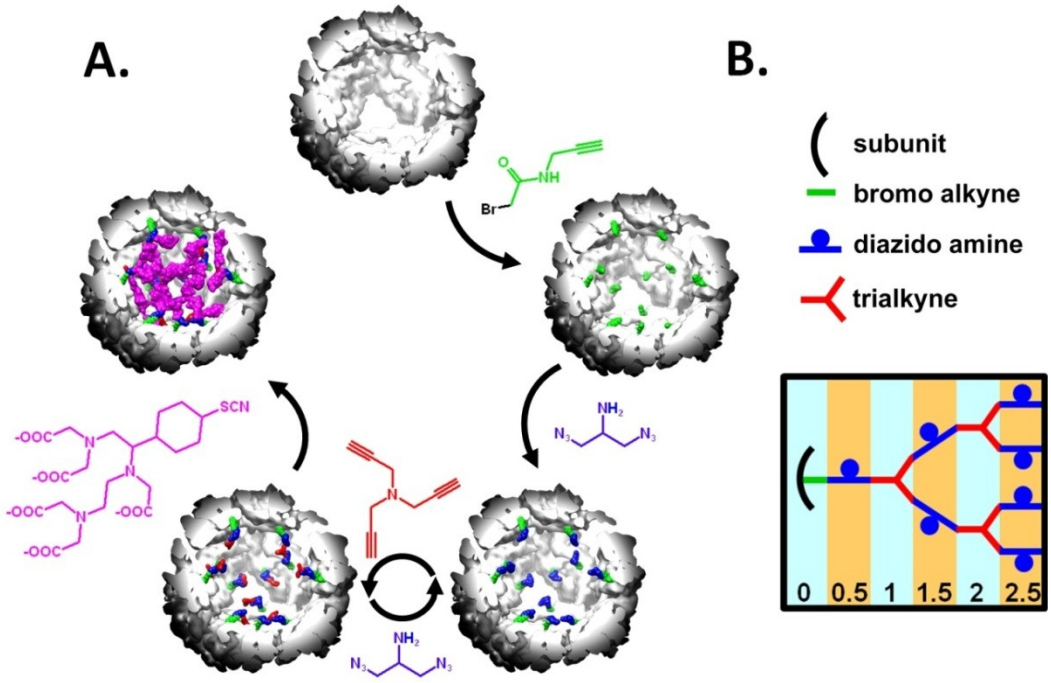



Figure 38. The synthetic scheme to generate a branched polymer inside the HSP cage (A) This synthesis is shown to scale with a cut-away view of HSP. HSP, G0.0, G0.5, G1.0 and G1.5 with DTPA are shown starting at the top and moving clockwise. First, an alkyne containing (green) molecule is anchored to the interior of the protein cage. Next, a diazido amine compound (blue) is clicked to the alkyne molecule on the interior of the cage. This is followed by a series of click reactions with the tri-alkyne (red) and diazido amine molecules. These reactions create a series of increasing generations of the HSP – branched polymer hybrid. Next, the introduced amines on the diazido amine compound were reacted with an amine reactive form of the DTPA-Gd chelate. (B) This represents an idealized growth of the polymer at each generation and assumes 100% reaction yields and no crosslinking between various polymer branches

Relaxivity of the Branched Polymer – Protein Cage Hybrid

The diazido-amine, which is incorporated into the branched polymer, contains a primary amine which has been shown to be chemically addressable after polymer formation on the interior of a protein cage.[3] After the polymer was grown in the cage an amine reactive form of the diethylene triamine pentaacetic acid - Gd (DTPA-Gd) contrast agent (2-(4-isothiocyanatobenzyl)-DTPA) was reacted with G0, G0.5, G1.5, G2.5 and G3.5. Higher concentrations of the Gd were detected with increasing generations up to a maximum of 159 Gd per cage at G2.5 (Figure 40A). This data is in general agreement with a similar branched polymer preparation that was labeled with a fluorescein isothiocyanate and produced 208 labels per cage at G2.5.[3] In both cases the maximum labeling was observed at G2.5 with no increase in labeling at G3.5. Steric crowding, shown in the modeled structures (Figure 39), is the mostly likely explanation for the loss of amine reactivity in G3.5.

To aid in the understanding of the polymer packing in the interior cavity of the protein cage a physical model of the branched polymer structure was constructed within structurally defined HSP cage (pdb file: 1shs) for generations G0.0, G0.5 (± DTPA-Gd), G1.0 and G1.5 + DTPA-Gd, Figure 39. The degree of polymer growth was determined from mass spectrometry and this data guided the creation of the branched polymer network model (cf. Supporting Information). While only 15% of the interior volume of the cage is occupied at G1.5 + DTPA-Gd, visual inspection shows that the interior is significantly crowded.

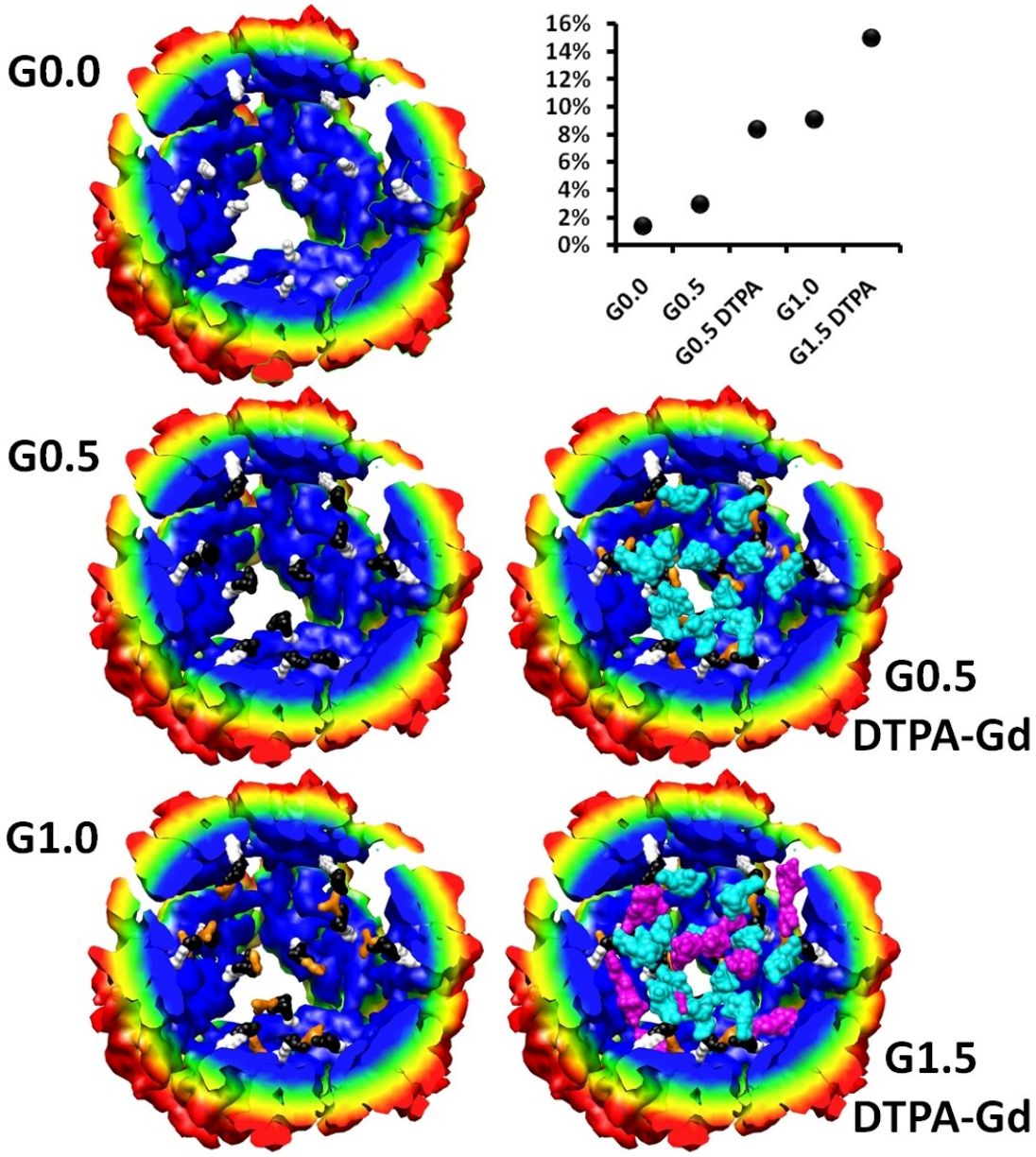



Figure 39 Structural models of branched polymer inside HSP
Models for generations G0.0, G0.5, G1.0 shown on the left. Shown on the right are models of G0.5 and G1.5 with DTPA-Gd coupled through the primary amine of the branched polymer. Here the bromo-alkyne molecule is shown in white, the diazido amine (black), tri-alkyne (orange), DTPA-Gd attached G0.5 (cyan) and DTPA-Gd attached G1.5 (purple). The degree of polymerization and DTPA-Gd labeling was based on mass spectrometry data. The percent of the interior volume of HSP that is occupied by the polymer and Gd-DTPA is plotted as a function of polymer growth points on the upper right corner of the figure. (cf. Supporting Information for details on the construction of this model).

Relaxivity measurements were collected on DTPA-Gd coupled to the protein cage – polymer hybrid (HSP-BP-DTPA-Gd) for generations G0.0, G0.5, G1.5, G2.5 and G3.5. The particle relaxivity increased as additional Gd was loaded into each successive generation to a maximum value of 4,200 $sec^{-1}$ $mM^{-1}$ (31MHz), Figure 40B. This corresponds to the second highest value for a protein cage construct with clinically relevant binding. Dramatic changes in the ionic relaxivity were not observed across the generations and the generational average value was 25$mM^{-1}$ $sec^{-1}$ (31MHz). This ionic relaxivity value is near the average value for all protein cage constructs (cf. Supporting Information) and is much higher than DTPA-Gd which has a $r_1$ = 4$mM^{-1}$ $sec^{-1}$, Figure 40C.[133]

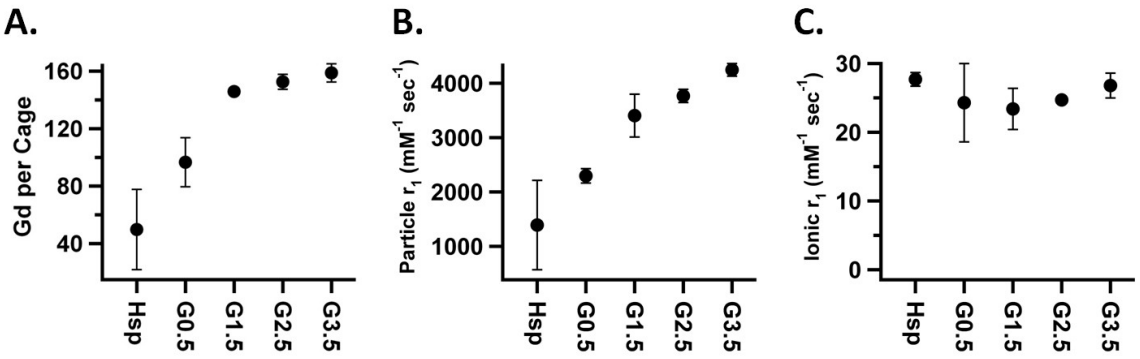



Figure 40. Generational Data for HSP-BP-DTPA-Gd
Gd ions per HSP cage for the cage hybrid construct (A), particle $r_1$ relaxivity (B) and ionic $r_1$ relaxivity (C) for the hybrid system at 31MHz (0.73 T) for generations G0.0 to G3.5. The three plots were produced by averaging the data from Batch 1 and 2 of the non passivated constructs. Gd and protein concentration determinations are described in supporting information.

Fitting Relaxivity Data to the SBM model for Relaxivity

To optimize the relaxivity properties of Gd based contrast agents it is helpful to have quantitative information of the parameters important to PRE. Since $\tau_R$, $\tau_M$ and q have different dependencies on the external magnetic field, it is possible to extract information of these values from a plot of relaxivities obtained for a contrast agent taken over a range of field strengths. These plot are referred to as nuclear magnetic resonance dispersion (NMRD) plots and they can be fit to the Solomon-Bloembergen-Morgan (SBM) analytical model for PRE[96, 111, 118] This type of analysis has been performed on a protein cage-Gd system and values for $\tau_R$, $\tau_M$ and q were obtained.[96]

We were also interested in determining the parameters that influence relaxivity ($\tau_M$, $\tau_R$ and q) and therefore $r_1$ NMRD profiles of the branched polymer constructs were collected at field strengths ranging from 4 to 62 MHz. Twenty individual NMRD profiles were collected containing two preparations of generations G0.0, G0.5, G1.5, G2.5 and G3.5 of the passivated and non-passivated hybrid constructs.

The individual NMRD profiles were similar for all preparations (cf. Supporting Information) indicating that no drastic changes in the relaxivity occurred as the chelator was attached to increasing generations of the polymer. This also suggests that there is not a drastic difference between the relaxivity properties of the Gd chelates that are anchored to endogenous lysine residues or to the polymer backbone. Therefore, the twenty NMRD profiles were averaged (Figure 41). This averaged NMRD profile is indicative of Gd attached to a relatively large species with a peak at 31 MHz, consistent with the intact HSP-branched polymer conjugate. By fitting this profile to the SBM model for Gd based

PRE values for $\tau_M$ , $\tau_R$ and q were obtained ($\tau_M = 5.6 \cdot 10^{-7}$ seconds, $\tau_R = 7.6 \cdot 10^{-9}$ seconds and q = 1.1) (cf. Supporting Information).

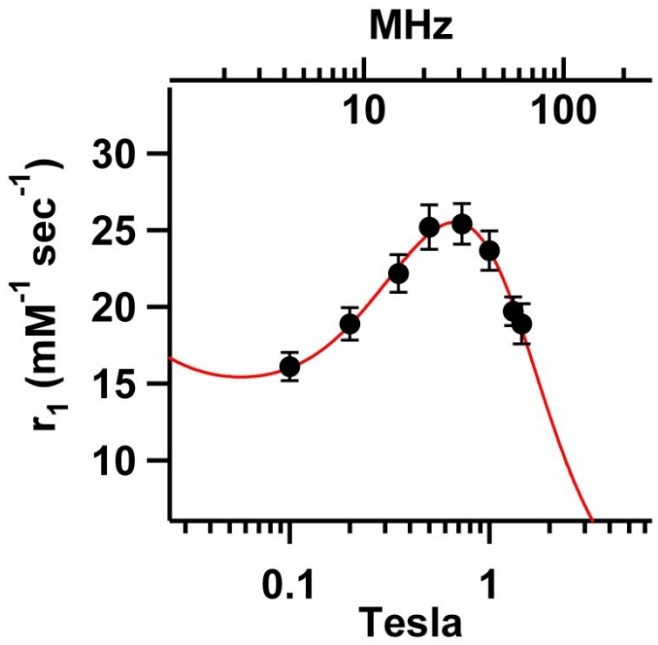



Figure 41 NMRD profile for the cage polymer hybrid construct
Twenty measurements were averaged for each data point consisting of two preparations of both passivated and unpassivated samples for the following generations G0.0, G0.5, G1.5, G2.5, G3.5. The error bars represent 1 standard deviation. A SBM fit was performed on these data resulting in $\tau_M = 5.6 \cdot 10^{-7}$ seconds, $\tau_R = 7.6 \cdot 10^{-9}$ seconds and q = 1.1.

The experimentally determined $\tau_R$ value ($7.6 \cdot 10^{-9}$ seconds) was lower than what was expected for the entire HSP cage, which can be calculated from the Einstein Stokes equation ($\tau_R = 2.2 \cdot 10^{-7}$ seconds) (cf. Supporting Information). This result is consistent with the idea that the linker connecting the metal to the cage is flexible. However, by tethering the chelated metal to the cage the mobility is significantly decreased when compared to the $\tau_R$ obtained for DTPA-Gd alone ($\tau_R \approx 7 \cdot 10^{-11}$ seconds).[95] The SBM model predicts that increasing the $\tau_R$ values above $10^{-8}$ seconds does not have a dramatic

effect on PRE (cf. Supporting Information). Therefore the $\tau_R$ for HSP-BP-DTPA-Gd is within an optimal range for these materials. This effect is shown Figure 42A-B where no significant difference in PRE is observed when $\tau_R$ is longer than the value determined by the SBM fit to the experimental data. However when $\tau_R$ is set to a smaller value a reduction in PRE is predicted by the SBM model.

The fit of the NMRD data for the HSP-BP-DTPA-Gd construct indicates a water exchange lifetime of $5.6 \cdot 10^{-7}$ seconds. By comparison to DTPA-Gd ($\tau_M = 2.4 \cdot 10^{-7}$) the water exchange for the chelate is significantly slower when docked on the polymer-cage hybrid system.[97] The SBM model predicts that the water exchange lifetime of the HSP-BP-DTPA-Gd is not optimal (too long) and significant PRE increases would result from a faster water exchange rate. When the $\tau_M$ value is set to a larger value the PRE is reduced and as the water exchange rate is increased a dramatic increase in PRE is predicted by the SBM model.

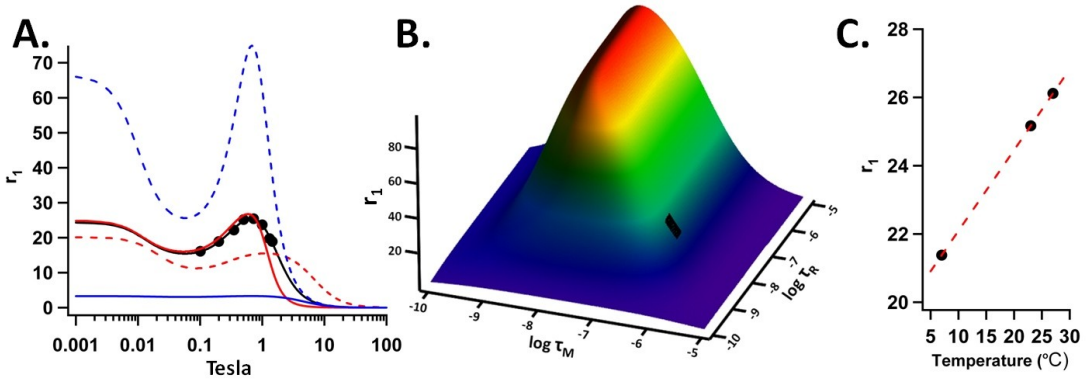



Figure 42 Relaxivity of HSP-BP-DTPA-Gd
(A) The experimental relaxivity data (black points) and a SBM fit (solid black line) to these data. NMRD profiles with the q and $\tau_M$ equal to the fit values and a lower $\tau_R$ value ($\tau_R = 7.6 \cdot 10^{-10}$, red dotted line) and higher value ($\tau_R = 7.6 \cdot 10^{-8}$, red solid line). Also profiles with the q and $\tau_R$ equal to the fit values and a lower $\tau_M$ value ($\tau_M = 7.6 \cdot 10^{-10}$, blue dotted line) and higher value ($\tau_R = 7.6 \cdot 10^{-8}$, blue solid line). (B) A plot of the relaxivity as calculated by the SBM model as a function of $\tau_R$ and $\tau_M$ at 0.73Tesla (31MHz) and q = 1.1. The black rectangle indicates the region corresponding to $\tau_R$ and $\tau_M$ values returned by the SBM fit to the experimental data. (C) A plot of $r_1$ vs. temperature for the HSP hybrid construct at 0.5Tesla (21MHz). Each experimental data point in (A) and (C) was produced by averaging both batches of the passivated and non-passivated preparations for G0.0, G0.5, G1.5, G2.5 and G3.5 and $r_1$ is the ionic relaxivity with units of $mM^{-1}$ $sec^{-1}$.

Water exchange lifetimes for many Gd chelators are longer than the optimal value for maximum PRE as predicted by analytical models for PRE and this is the case for DTPA-Gd. The water exchange rate can be increased by raising the temperature of the sample and this should lead to a positive effect on PRE. However heating the sample also results in a shorter rotational correlation time, which should negatively impact PRE. Since heating the sample has these two opposite effects on PRE, it was possible to ascertain the limiting factor for efficient PRE, $\tau_R$ or $\tau_M$, in the HSP-BP-Gd-DTPA construct. As shown in Figure 42C, a direct correlation between relaxivity and temperature was observed for the HSP-BP-Gd-DTPA construct, suggesting that the

positive effect of a shorter exchange lifetime out-competed the negative effect of a shorter rotational correlation time. The analysis of the temperature dependence is in agreement with the SBM fitting results since both analyses suggest the $\tau_R$ is at an optimal value while the $\tau_M$ is higher than the ideal value for maximum PRE.

Although drastic changes in the NMRD profiles were not observed across generations there were slight differences. The most obvious difference was that the peak centered at 31MHz was shifted slightly to the left as the generations increased (cf. Supporting Information). Therefore, we analyzed the SBM fits to individual generations to glean additional information from this system. Both $\tau_M$ and q remained relatively constant across all generations, Figure 43A,B. This suggests that there is not a drastic decrease in the permeability of water through the protein cage polymer hybrid construct with increasing polymer generation.

An increase in $\tau_R$ was observed as the polymeric network was grown (Figure 43C) through increasing generations. Previous studies suggest that local mobility in the linker has a large role in determining the relaxivity of protein cage – Gd chelate systems.[2, 94] The generational increase in $\tau_R$ shown in Figure 43C could result from decreased local mobility due to increasing rigidity of the polymer imparted by cross-linking at higher generations. It is known that the hybrid protein cage branched polymer constructs become more stable to heat treatment at higher generations. It is possible that local rigidity of the polymer coincides with the global increase in stability of the higher generations.[3]

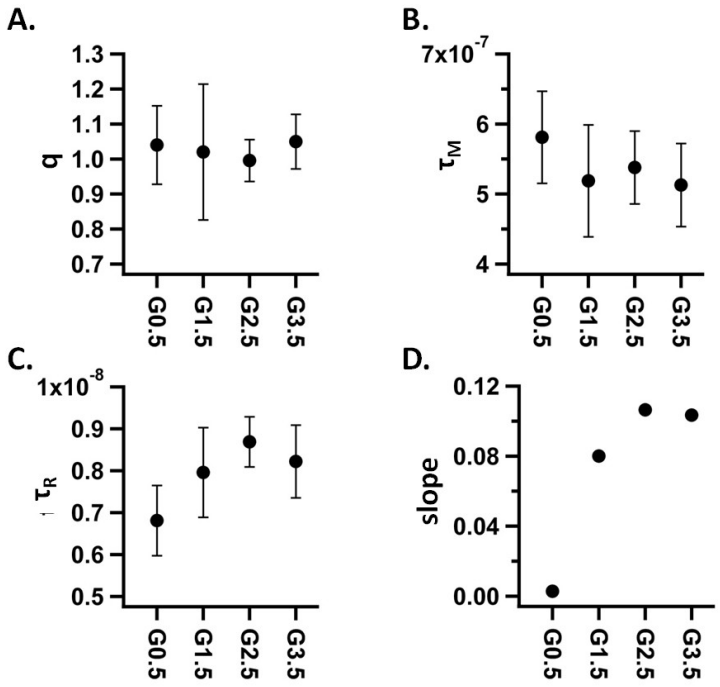



Figure 43. Results of SBM fits of individual generations
showing the generational change for q (A), $\tau_M$ (B), $\tau_R$ (C) and the slope of a linear fit to a plot of $r_1$ vs. temperature (D) for generations G0.5, G1.5, G2.5 G3.5. Four experimental points were averaged for each data point shown in plots A-C, comprised of two experimental points from the non-passivated preparations and two from passivated preparations. The error bars are ± one standard deviation for plots A-C.

Furthermore, the relaxivity showed an altered dependence on temperature across the generations. A positive dependence on temperature holds true for all generations however, the slope of this positive dependence ($r_1$ vs. temperature) increased from G0.0 to a plateau at G2.5 (Figure 43D) (cf. Supporting Information). Two physical models could explain this observation. First, since $\tau_R$ is higher for higher generations, heating the lower generation samples may cause $\tau_R$ to decrease past a critical point where the relaxivity is significantly decreased. In this scenario the higher generation samples may have a $\tau_R$ value that is further up the plateau in the $r_1$ vs. $\tau_R$ plot such that a reduction in $\tau_R$

due to heating would not significantly decrease the relaxivity. A second possible explanation for this change of slope ($r_1$ vs. temperature) for the generations was that at higher generations there was a decrease in the water flux through the protein cage and this results in decreased relaxivity. Increasing the temperature allows more water to pass through the protein cage that, in this scenario, has more of an affect for the higher generation samples. The higher generations would therefore show a greater dependence on an increase in temperature when compared to the lower generations. This second explanation is not corroborated by the SBM fitting results since q and $\tau_M$ do not change which would be expected if the flux of water were changing across generations. However, the accuracy and resolution of the values obtained from the SBM fit may be too low to conclusively rule out a generational change in q or $\tau_M$ and this second explanation should not be disregarded. It should be noted that although the SBM fitting picked up generational changes in $\tau_R$ for HSP-BP-DTPA-Gd, these changes do not correspond to significant changes in $r_1$ values.

## Conclusions

The most promising aspect of this work is the development of a method to take advantage of the interior volume of the protein cage for development of a $T_1$ CA. To date, only the surface of protein cages including viral capsids have been modified to carry Gd chelators by single step chemical labeling. Here we expand the chemical modification to fill the interior volume of the protein cage with a DTPA-Gd labeled polymer network, which results in a particle that has a per particle relaxivity of 4,200$mM^{-}$

$^{1}$ $sec^{-1}$. While this is lower than the report of the particle relaxivity achieved with the MS2 viral capsid - DTPA-Gd construct[81] (7,200$mM^{-1}$ $sec^{-1}$), the MS2 construct is significantly larger than the HSP, occupying twelve times the volume per particle. When the relaxivity behavior of these materials are compared on a volume basis, the HSP construct exhibits a relaxivity that is seven times that of the MS2 construct (cf. Supporting Information). Until now the extremely efficient relaxivity properties of Gd bound to the low-affinity endogenous metal binding sites contained in the capsids of the Cowpea chlorotic mottle virus (CCMV) seemed unattainable in systems with high affinity chelators.[90] However, in the present work we have achieved this by filling the interior cavity of HSP with DTPA-Gd and the per volume relaxivity, or the "relaxivity density", values are are among the highest reported for protein cage – Gd constructs (cf. Supporting Information). Furthermore, the HSP-BP-DTPA-Gd construct has more efficient relaxivity properties per unit mass when compared to DTPA-Gd. This suggests that the high Gd loading and the increased ionic relaxivity of HSP-BP-DTPA-Gd adequately make up for the extra mass of the protein cage when compared to DTPA-Gd (cf. Supporting Information).[134] This is the first example of the protein cage – Gd chelate construct having a higher relaxivity per unit mass than the Gd – chelate itself and this example highlights the advantages of designing MR CA on nanoscale platforms.

This work represents a paradigm shift in the employment of protein cages as functional medical materials. We show that there are significant benefits, namely increased stability and efficient particle relaxivity properties that result from a synthetic approach that takes full advantage of the interior volume of protein cages. This strategy

can be extended to larger protein cages and it will be interesting to see if the ionic relaxivity remains constant in larger systems. Currently, we are developing similar protein cage-polymer hybrid constructs with Gd chelators that have more desirable q and $\tau_M$ values to see how these particles perform. Also, using this strategy of labeling protein cages with branched polymers containing moieties with other desired properties such as drug delivery, radiation dose and hyperthermia, should be pursued in an attempt to develop high performance clinical agents. Finally, successful application of these functionalized protein cages as imaging or drug delivery agents is likely to require coupling of *in vivo* targeting functionality and therefore we are hopeful that targeting functionality of protein cages is developed further.

In conclusion we report the successful generation of a Gd chelate-containing branched polymer on the interior of a protein cage. Parameters important to relaxivity were determined by fitting NMRD profiles to the SBM model. These results suggest that while the exchange lifetime is lower than optimal, the rotational correlation time of the Gd ions has been optimized for HSP-BP-DTPA-Gd.

# REFERENCES CITED

1. Fulton, J.; Bothner, B.; Lawrence, M.; Johnson, J. E.; Douglas, T.; Young, M., Genetics, biochemistry and structure of the archaeal virus STIV. *Biochem Soc Trans* 2009, *37* (Pt 1), 114-7.

2. Liepold, L.; Anderson, S.; Willits, D.; Oltrogge, L.; Frank, J. A.; Douglas, T.; Young, M., Viral capsids as MRI contrast agents. *Magn Reson Med* 2007, *58* (5), 871-9.

3. Abedin, M. J.; Liepold, L.; Suci, P.; Young, M.; Douglas, T., Synthesis of a cross-linked branched polymer network in the interior of a protein cage. *Journal of the American Chemical Society* 2009, *131* (12), 4346-54.

4. Liepold, L.; Oltrogge, L. M.; Suci, P. A.; Young, M. J.; Douglas, T., Correct charge state assignment of native electrospray spectra of protein complexes. *J Am Soc Mass Spectrom* 2009, *20* (3), 435-42.

5. Johnson, J. E.; Speir, J. A., Quasi-equivalent viruses: A paradigm for protein assemblies. *Journal of Molecular Biology* 1997, *269* (5), 665-675.

6. Speir, J. A.; Munshi, S.; Wang, G.; Baker, T. S.; Johnson, J. E., Structures of the native and swollen forms of cowpea chlorotic mottle virus determined by X-ray crystallography and cryo-electron microscopy. *Structure* 1995, *3* (1), 63-78.

7. Douglas, T.; Ripoll, D., Electrostatic gradients in the iron storage protein ferritin. *Protein Science* 1998, *7*, 1083-1091.

8. Douglas, T.; Strable, E.; Willits, D.; Aitouchen, A.; Libera, M.; Young, M., Protein Engineering of a Viral Cage for Constrained Nano-Materials Synthesis. *Adv Mater* 2002, *14* (6), 415-418.

9. Gillitzer, E.; Willits, D.; Young, M.; Douglas, T., Chemical modification of a viral cage for multivalent presentation. *Chem. Commun.* 2002, (20), 2390-1.

10. Chasteen, N. D.; Harrison, P. M., Mineralization in ferritin: An efficient means of iron storage. *Journal of Structural Biology* 1999, *126* (3), 182-194.

11. Harrison, P. M.; Arosio, P., Ferritins: Molecular properties, iron storage function and cellular regulation. *Biochim. Biophys. Acta-Biomembr.* 1996, *1275* (3), 161-203.

12. Lawson, D. M.; Artymiuk, P. J.; Yewdall, S. J.; Smith, J. M. A.; Livingstone, J. C.; Treffry, A.; Luzzago, A.; Levi, S.; Arosio, P.; Cesareni, G.; Thomas, C. D.;

Shaw, W. V.; Harrison, P. M., Solving the Structure of Human H-Ferritin by Genetically Engineering Intermolecular Crystal Contacts. *Nature* 1991, *349* (6309), 541-544.

13. Hwang, J.; Krebs, C.; Huynh, B. H.; Edmondson, D. E.; Theil, E. C.; Penner-Hahn, J. E., A short Fe-Fe distance in peroxodiferric ferritin: Control of Fe substrate versus cofactor decay? *Science* 2000, *287* (5450), 122-125.

14. Douglas, T.; Mann, S., Biomolecules in the synthesis of inorganic solids. In *Molecular Biology and Biotechnology*, Meyers, R. A., Ed. VCH publishers: New York, 1995; pp 466-469.

15. Mann, S.; Archibald, D. D.; Didymus, J. M.; Douglas, T.; Heywood, B. R.; Meldrum, F. C.; Reeves, N. J., Crystallization at inorganic-organic interfaces: biominerals and biomimetic synthesis. *Science* 1993, *261* (5126), 1286-92.

16. Mann, S.; Meldrum, F. C., Controlled Synthesis of Inorganic Materials Using Supramolecular Assemblies. *Adv Mater* 1991, *3* (6), 316-318.

17. Mann, S.; Webb, J.; Williams, R. J. P., *Biomineralization: Chemical and Biochemical Perspectives*. VCH: New York, 1989.

18. Aime, S.; Frullano, L.; Crich, S. G., Compartmentalization of a gadolinium complex in the apoferritin cavity: A route to obtain high relaxivity contrast agents for magnetic resonance imaging. *Angewandte Chemie-International Edition* 2002, *41* (6), 1017-+.

19. Danon, D.; Skutelsk.E; Marikovs.Y; Goldstei.L, Use of Cationized Ferritin as a Label of Negative Charges on Cell Surfaces. *Journal of Ultrastructure Research* 1972, *38* (5-6), 500-&.

20. Wong, K. K. W.; Colfen, H.; Whilton, N. T.; Douglas, T.; Mann, S., Synthesis and characterization of hydrophobic ferritin proteins. *J Inorg Biochem* 1999, *76* (3-4), 187-195.

21. Wong, K. K. W.; Douglas, T.; Gider, S.; Awschalom, D. D.; Mann, S., Biomimetic snthesis and characterization of magnetic proteins (magnetoferritin). *Chem. Mater.* 1998, *10*, 279-85.

22. Wong, K. K. W.; Whilton, N. T.; Colfen, H.; Douglas, T.; Mann, S., Hydrophobic proteins: synthesis and characterisation of organic-soluble alkylated ferritins. *Chemical Communications* 1998, (16), 1621-1622.

23. Sano, K.; Ajima, K.; Iwahori, K.; Yudasaka, M.; Iijima, S.; Yamashita, I.; Shiba, K., Endowing a ferritin-like cage protein with high affinity and selectivity for certain inorganic materials. *Small* 2005, *1* (8-9), 826-832.

24. Hayashi, T.; Sano, K.; Shiba, K.; Kumashiro, Y.; Iwahori, K.; Yamashita, I.; Hara, M., Mechanism underlying specificity of proteins targeting inorganic materials. *Nano letters* 2006, *6* (3), 515-519.

25. Kim, K. K.; Kim, R.; Kim, S. H., Crystal structure of a small heat-shock protein. *Nature* 1998, *394* (6693), 595-9.

26. Kim, K. K.; Yokota, H.; Santoso, S.; Lerner, D.; Kim, R.; Kim, S. H., Purification, crystallization, and preliminary X-ray crystallographic data analysis of small heat shock protein homolog from Methanococcus jannaschii, a hyperthermophile. *J Struct Biol* 1998, *121* (1), 76-80.

27. Kim, R.; Kim, K. K.; Yokota, H.; Kim, S. H., Small heat shock protein of Methanococcus jannaschii, a hyperthermophile. *Proceedings of the National Academy of Sciences of the United States of America* 1998, *95* (16), 9129-9133.

28. Flenniken, M. L.; Liepold, L. O.; Crowley, B. E.; Willits, D. A.; Young, M. J.; Douglas, T., Selective attachment and release of a chemotherapeutic agent from the interior of a protein cage architecture. *Chem Commun (Camb)* 2005, (4), 447-9.

29. Flenniken, M. L.; Willits, D. A.; Brumfield, S.; Young, M. J.; Douglas, T., The small heat shock protein cage from Methanococcus jannaschii is a versatile nanoscale platform for genetic and chemical modification. *Nano letters* 2003, *3* (11), 1573-1576.

30. Flenniken, M. L.; Willits, D. A.; Harmsen, A. L.; Liepold, L. O.; Harmsen, A. G.; Young, M. J.; Douglas, T., Melanoma and lymphocyte cell-specific targeting incorporated into a heat shock protein cage architecture. *Chem Biol* 2006, *13* (2), 161-70.

31. Willner, D.; Trail, P. A.; Hofstead, S. J.; King, H. D.; Lasch, S. J.; Braslawsky, G. R.; Greenfield, R. S.; Kaneko, T.; Firestone, R. A., (6-Maleimidocaproyl)hydrazone of doxorubicin--a new derivative for the preparation of immunoconjugates of doxorubicin. *Bioconjug. Chem.* 1993, *4* (6), 521-7.

32. Liepold, L. O.; Revis, J.; Allen, M.; Oltrogge, L.; Young, M.; Douglas, T., Structural transitions in Cowpea chlorotic mottle virus (CCMV). *Phys Biol* 2005, *2* (4), S166-72.

33. Grant, R. A.; Filman, D. J.; Finkel, S. E.; Kolter, R.; Hogle, J. M., The crystal structure of Dps, a ferritin homolog that binds and protects DNA. *Nat. Struct. Biol.* 1998, *5* (4), 294-303.

34. Bozzi, M.; Mignogna, G.; Stefanini, S.; Barra, D.; Longhi, C.; Valenti, P.; Chiancone, E., A novel non-heme iron-binding ferritin related to the DNA-binding proteins of the Dps family in Listeria innocua. *J. Biol. Chem.* 1997, *272* (6), 3259-3265.

35. Ilari, A.; Stefanini, S.; Chiancone, E.; Tsernoglou, D., The dodecameric ferritin from Listeria innocua contains a novel intersubunit iron-binding site. *Nat. Struct. Biol.* 2000, *7* (1), 38-43.

36. Stefanini, S.; Cavallo, S.; Montagnini, B.; Chiancone, E., Incorporation of iron by the unusual dodecameric ferritin from Listeria innocua. *Biochemical Journal* 1999, *338*, 71-75.

37. Su, M. H.; Cavallo, S.; Stefanini, S.; Chiancone, E.; Chasteen, N. D., The so-called Listeria innocua ferritin is a Dps protein. Iron incorporation, detoxification, and DNA protection properties. *Biochemistry* 2005, *44* (15), 5572-5578.

38. Wiedenheft, B.; Mosolf, J.; Willits, D.; Yeager, M.; Dryden, K. A.; Young, M.; Douglas, T., An archaeal antioxidant: Characterization of a Dps-like protein from Sulfolobus solfataricus. *Proceedings of the National Academy of Sciences of the United States of America* 2005, *102* (30), 10551-10556.

39. Glish, G. L.; Vachet, R. W., The basics of mass spectrometry in the twenty-first century. *Nat Rev Drug Discov* 2003, *2* (2), 140-150.

40. Griffiths, J., A Brief History of Mass Spectrometry. *Analytical Chemistry* 2008, *80* (15), 5678-5683.

41. El-Aneed, A.; Cohen, A.; Banoub, J., Mass Spectrometry, Review of the Basics: Electrospray, MALDI, and Commonly Used Mass Analyzers. *Applied Spectroscopy Reviews* 2009, *44* (3), 210 - 230.

42. Gaskell, S. J., <B>Electrospray: Principles and Practice</B>. *Journal of Mass Spectrometry* 1997, *32* (7), 677-688.

43. Rockwood, A. L., The Expanding Role of Mass Spectrometry in Biotechnology. Gary Siuzdak. San Diego, CA: MCC Press, 2003, 286 pp., $49.00, softcover. ISBN 0-9742451-0-0. *Clin Chem* 2004, *50* (6), 1108-1109.

44. Ferrige, A. G.; Seddon, M. J.; Green, B. N.; Jarvis, S. A.; Skilling, J.; Staunton, J., Disentangling electrospray spectra with maximum entropy. *Rapid Communications in Mass Spectrometry* 1992, *6* (11), 707-711.

45. Fournier, M. L.; Gilmore, J. M.; Martin-Brown, S. A.; Washburn, M. P., Multidimensional Separations-Based Shotgun Proteomics. *Chem Rev* 2007, *107* (8), 3654-3686.

46. Chace, D. H.; Kalas, T. A.; Naylor, E. W., Use of Tandem Mass Spectrometry for Multianalyte Screening of Dried Blood Specimens from Newborns. *Clin Chem* 2003, *49* (11), 1797-1817.

47. Wu, W. W.; Wang, G.; Baek, S. J.; Shen, R. F., Comparative study of three proteomic quantitative methods, DIGE, cICAT, and iTRAQ, using 2D gel- or LC-MALDI TOF/TOF. *J Proteome Res* 2006, *5* (3), 651-8.

48. Geromanos, S. J.; Vissers, J. P. C.; Silva, J. C.; Dorschel, C. A.; Li, G.-Z.; Gorenstein, M. V.; Bateman, R. H.; Langridge, J. I., The detection, correlation, and comparison of peptide precursor and product ions from data independent LC-MS with data dependant LC-MS/MS. *Proteomics* 2009, *9* (6), 1683-1695.

49. Delahunty, C.; Yates, J. R., 3rd, Protein identification using 2D-LC-MS/MS. *Methods* 2005, *35* (3), 248-55.

50. Ghaemmaghami, S.; Huh, W. K.; Bower, K.; Howson, R. W.; Belle, A.; Dephoure, N.; O'Shea, E. K.; Weissman, J. S., Global analysis of protein expression in yeast. *Nature* 2003, *425* (6959), 737-41.

51. Trent, J. D.; Kagawa, H. K.; Yaoi, T.; Olle, E.; Zaluzec, N. J., Chaperonin filaments: The archaealâ€‰cytoskeleton? The National Academy of Sciences of the USA.

52. Benesch, J. L. P.; Ruotolo, B. T.; Simmons, D. A.; Robinson, C. V., Protein Complexes in the Gas Phase: Technology for Structural Genomics and Proteomics. *ChemInform* 2007, *38* (47).

53. Heck, A. J. R.; Heuvel, R. H. H. v. d., Investigation of intact protein complexes by mass spectrometry. *Mass Spectrometry Reviews* 2004, *23* (5), 368-389.

54. Ruotolo, B. T.; Benesch, J. L. P.; Sandercock, A. M.; Hyung, S.-J.; Robinson, C. V., Ion mobility-mass spectrometry analysis of large protein complexes. *Nat. Protocols* 2008, *3* (7), 1139-1152.

55. Chernushevich, I. V.; Thomson, B. A., Collisional cooling of large ions in electrospray mass spectrometry. *Analytical Chemistry* 2004, *76* (6), 1754-1760.

56. Fitzgerald, M. C.; Chernushevich, I.; Standing, K. G.; Whitman, C. P.; Kent, S. B. H., Probing the oligomeric structure of an enzyme by electrospray ionization time-of-flight mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 1996, *93* (14), 6851-6856.

57. Ganem, B.; Li, Y. T.; Henion, J. D., Observation of Noncovalent Enzyme Substrate and Enzyme Product Complexes by Ion-Spray Mass-Spectrometry. *Journal of the American Chemical Society* 1991, *113* (20), 7818-7819.

58. Green, B. N.; Bordoli, R. S.; Hanin, L. G.; Lallier, F. H.; Toulmond, A.; Vinogradov, S. N., Electrospray ionization mass spectrometric determination of the molecular mass of the similar to 200-kDa globin dodecamer subassemblies in hexagonal bilayer hemoglobins. *J. Biol. Chem.* 1999, *274* (40), 28206-28212.

59. Loo, J. A., Studying noncovalent protein complexes by electrospray ionization mass spectrometry. *Mass Spectrometry Reviews* 1997, *16* (1), 1-23.

60. Rostom, A. A.; Fucini, P.; Benjamin, D. R.; Juenemann, R.; Nierhaus, K. H.; Hartl, F. U.; Dobson, C. M.; Robinson, C. V., Detection and selective dissociation of intact ribosomes in a mass spectrometer. *Proceedings of the National Academy of Sciences of the United States of America* 2000, *97* (10), 5185-5190.

61. Sanglier, S.; Leize, E.; Van Dorsselaer, A.; Zal, F., Comparative ESI-MS study of similar to 2.2 MDa native hemocyanins from deep-sea and shore crabs: From protein oligomeric state to biotope. *J Am Soc Mass Spectr* 2003, *14* (5), 419-429.

62. Siuzdak, G.; Bothner, B.; Yeager, M.; Brugidou, C.; Fauquet, C. M.; Hoey, K.; Chang, C. M., Mass spectrometry and viral analysis. *Chemistry & biology* 1996, *3* (1), 45-48.

63. Smith, R. D.; Lightwahl, K. J.; Winger, B. E.; Loo, J. A., Preservation of Noncovalent Associations in Electrospray Ionization Mass-Spectrometry - Multiply Charged Polypeptide and Protein Dimers. *Org Mass Spectrom* 1992, *27* (7), 811-821.

64. Tang, X. J.; Brewer, C. F.; Saha, S.; Chernushevich, I.; Ens, W.; Standing, K. G., Investigation of Protein-Protein Noncovalent Interactions in Soybean Agglutinin by Electrospray-Ionization Time-of-Flight Mass-Spectrometry. *Rapid Communications in Mass Spectrometry* 1994, *8* (9), 750-754.

65. Tito, M. A.; Tars, K.; Valegard, K.; Hajdu, J.; Robinson, C. V., Electrospray time-of-flight mass spectrometry of the intact MS2 virus capsid. *Journal of the American Chemical Society* 2000, *122* (14), 3550-3551.

66. Van Berkel, W. J. H.; Van Den Heuvel, R. H. H.; Versluis, C.; Heck, A. J. R., Detection of intact megaDalton protein assemblies of vanillyl-alcohol oxidase by mass spectrometry. *Protein Science* 2000, *9* (3), 435-439.

67. Benesch, J. L. P.; Sobott, F.; Robinson, C. V., Thermal dissociation of multimeric protein complexes by using nanoelectrospray mass spectrometry. *Analytical Chemistry* 2003, *75* (10), 2208-2214.

68. Ilag, L. L.; Videler, H.; McKay, A. R.; Sobott, F.; Fucini, P.; Nierhaus, K. H.; Robinson, C. V., Heptameric (L12)(6)/L10 rather than canonical pentameric complexes are found by tandem MS of intact ribosomes from thermophilic bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 2005, *102* (23), 8192-8197.

69. Sobott, F.; Benesch, J. L. P.; Vierling, E.; Robinson, C. V., Subunit exchange of multimeric protein complexes - Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J. Biol. Chem.* 2002, *277* (41), 38921-38929.

70. Taverner, T.; Hernandez, H.; Sharon, M.; Ruotolo, B. T.; Matak-Vinkovic, D.; Devos, D.; Russell, R. B.; Robinson, C. V., Subunit architecture of intact protein complexes from mass spectrometry and homology modeling. *Accounts Chem. Res.* 2008, *41* (5), 617-627.

71. Sobott, F.; Robinson, C. V., Characterising electrosprayed biomolecules using tandem-MS - the noncovalent GroEL chaperonin assembly. *Int J Mass Spectrom* 2004, *236* (1-3), 25-32.

72. Mckay, A. R.; Ruotolo, B. T.; Ilag, L. L.; Robinson, C. V., Mass measurements of increased accuracy resolve heterogeneous populations of intact ribosomes. *Journal of the American Chemical Society* 2006, *128* (35), 11433-11442.

73. Allen, M.; Willits, D.; Young, M.; Douglas, T., Constrained synthesis of cobalt oxide nanomaterials in the 12-subunit protein cage from Listeria innocua. *Inorg. Chem.* 2003, *42* (20), 6300-6305.

74. Fernandez de la Mora, J., Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. *Analytica Chimica Acta* 2000, *406* (1), 93-104.

75. Gauss, G. H.; Benas, P.; Wiedenheft, B.; Young, M.; Douglas, T.; Lawrence, C. M., Structure of the DPS-Like Protein from Sulfolobus solfataricus Reveals a Bacterioferritin-Like Dimetal Binding Site within a DPS-Like Dodecameric Assembly. 2006; Vol. 45, pp 10815-10827.

76. Pinkse, M. W. H.; Maier, C. S.; Kim, J. I.; Oh, B. H.; Heck, A. J. R., Macromolecular asserably of Helicobacter pylori urease investigated by mass spectrometry. *Journal of Mass Spectrometry* 2003, *38* (3), 315-320.

77. van Breukelen, B.; Barendregt, A.; Heck, A. J. R.; van den Heuvel, R. H. H., Resolving stoichiometries and oligomeric states of glutamate synthase protein complexes with curve fitting and simulation of electrospray mass spectra. *Rapid Communications in Mass Spectrometry* 2006, *20* (16), 2490-2496.

78. Benmelouka, M.; Borel, A.; Moriggi, L.; Helm, L.; Merbach, A. E., Design of Gd(III)-based magnetic resonance imaging contrast agents: static and transient zero-field splitting contributions to the electronic relaxation and their impact on relaxivity. *J Phys Chem B* 2007, *111* (4), 832-40.

79. Aime, S.; Cabella, C.; Colombatto, S.; Crich, S. G.; Gianolio, E.; Maggioni, F., Insights into the use of paramagnetic Gd(III) complexes in MR-molecular imaging investigations. *Journal of Magnetic Resonance Imaging* 2002, *16* (4), 394-406.

80. Meade, T. J.; Taylor, A. K.; Bull, S. R., New magnetic resonance contrast agents as biochemical reporters. *Curr Opin Neurobiol* 2003, *13* (5), 597-602.

81. Anderson, E. A.; Isaacman, S.; Peabody, D. S.; Wang, E. Y.; Canary, J. W.; Kirshenbaum, K., Viral nanoparticles donning a paramagnetic coat: conjugation of MRI contrast agents to the MS2 capsid. *Nano letters* 2006, *6* (6), 1160-4.

82. Lewis, J. D.; Destito, G.; Zijlstra, A.; Gonzalez, M. J.; Quigley, J. P.; Manchester, M.; Stuhlmann, H., Viral nanoparticles as tools for intravital vascular imaging. *Nature Medicine* 2006, *12* (3), 354-360.

83. Yamada, T.; Iwasaki, Y.; Tada, H.; Iwabuki, H.; Chuah, M. K.; VandenDriessche, T.; Fukuda, H.; Kondo, A.; Ueda, M.; Seno, M.; Tanizawa, K.; Kuroda, S., Nanoparticles for the delivery of genes and drugs to human hepatocytes. *Nat Biotechnol* 2003, *21* (8), 885-90.

84. Singh, P.; Destito, G.; Schneemann, A.; Manchester, M., Canine parvovirus-like particles, a novel nanomaterial for tumor targeting. *J Nanobiotechnology* 2006, *4*, 2.

85. Douglas, T.; Young, M., Viruses: Making friends with old foes. *Science* 2006, *312* (5775), 873-875.

86. Klem, M. T.; Willits, D.; Young, M.; Douglas, T., 2-D array formation of genetically engineered viral cages on Au surfaces and imaging by atomic force microscopy. *Journal of the American Chemical Society* 2003, *125* (36), 10806-10807.

87. Raja, K. S.; Wang, Q.; Gonzalez, M. J.; Manchester, M.; Johnson, J. E.; Finn, M. G., Hybrid virus-polymer materials. 1. Synthesis and properties of PEG-decorated cowpea mosaic virus. *Biomacromolecules* 2003, *4* (3), 472-6.

88. Sen Gupta, S.; Kuzelka, J.; Singh, P.; Lewis, W. G.; Manchester, M.; Finn, M. G., Accelerated bioorthogonal conjugation: A practical method for the Ligation of diverse functional molecules to a polyvalent virus scaffold. *Bioconjugate chemistry* 2005, *16* (6), 1572-1579.

89. Varpness, Z.; Peters, J. W.; Young, M.; Douglas, T., Biomimetic synthesis of a H-2 catalyst using a protein cage architecture. *Nano letters* 2005, *5* (11), 2306-2309.

90. Allen, M.; Bulte, J. W. M.; Liepold, L.; Basu, G.; Zywicke, H. A.; Frank, J. A.; Young, M.; Douglas, T., Paramagnetic viral nanoparticles as potential high-relaxivity magnetic resonance contrast agents. *Magnet Reson Med* 2005, *54* (4), 807-812.

91. Kobayashi, H.; Brechbiel, M. W., Dendrimer-based macromolecular MRI contrast agents: characteristics and application. *Mol Imaging* 2003, *2* (1), 1-10.

92. Mulder, W. J.; Strijkers, G. J.; van Tilborg, G. A.; Griffioen, A. W.; Nicolay, K., Lipid-based nanoparticles for contrast-enhanced MRI and molecular imaging. *NMR Biomed* 2006, *19* (1), 142-64.

93. Caravan, P., Strategies for increasing the sensitivity of gadolinium based MRI contrast agents. *Chem Soc Rev* 2006, *35* (6), 512-23.

94. Datta, A.; Hooker, J. M.; Botta, M.; Francis, M. B.; Aime, S.; Raymond, K. N., High relaxivity gadolinium hydroxypyridonate-viral capsid conjugates: nanosized MRI contrast agents. *Journal of the American Chemical Society* 2008, *130* (8), 2546-52.

95. Eldik, R. v., *Advances in Inorganic Chemistry: Relaxometry of water-metal ion interactions*. Academic Press: 2005; Vol. 57, p 536.

96. Helm, L., Relaxivity in paramagnetic systems: Theory and mechanisms. *Progress in Nuclear Magnetic Resonance Spectroscopy* 2006, *49* (1), 45-64.

97. Helm, L.; Nicolle, G. M.; Merbach, A. E., Water and Proton Exchange Processes on Metal Ions. *ChemInform* 2006, *37* (14).

98. Basu, G.; Allen, M. A.; Willits, D.; Young, M.; Douglas, T., Metal binding to Cowpea Chlorotic mottle Virus using Tb(III) Fluorescence. *Journal of Bioinorganic Chemistry* 2003, *8*, 721-725.

99. Rocklage, S. M.; Worah, D.; Kim, S. H., Metal-Ion Release from Paramagnetic Chelates - What Is Tolerable. *Magnet Reson Med* 1991, *22* (2), 216-221.

100. Brumfield, S.; Willits, D.; Tang, L.; Johnson, J. E.; Douglas, T.; Young, M., Heterologous expression of the modified coat protein of Cowpea chlorotic mottle bromovirus results in the assembly of protein cages with altered architectures and function. *Journal of General Virology* 2004, *85*, 1049-1053.

101. Bancroft, J. B.; Wagner, G. W.; Bracker, C. E., Self-Assembly of a Nucleic-Acid Free Pseudo-Top Component for a Small Spherical Virus. *Virology* 1968, *36* (1), 146-&.

102. Vazquez-Ibar, J. L.; Weinglass, A. B.; Kaback, H. R., Engineering a terbium-binding site into an integral membrane protein for luminescence energy transfer. *Proc Natl Acad Sci U S A* 2002, *99* (6), 3487-92.

103. Laus, S.; Sour, A.; Ruloff, R.; Toth, E.; Merbach, A. E., Rotational dynamics account for pH-dependent relaxivities of PAMAM dendrimeric, Gd-based potential MRI contrast agents. *Chem-Eur J* 2005, *11* (10), 3064-3076.

104. Speir, J. A.; Bothner, B.; Qu, C.; Willits, D. A.; Young, M. J.; Johnson, J. E., Enhanced local symmetry interactions globally stabilize a mutant virus capsid that maintains infectivity and capsid dynamics. *J Virol* 2006, *80* (7), 3582-91.

105. Hemminga, M. A.; Faber, A. J., Analysis of Anisotropic Spin-Label Motion in Saturation-Transfer Electron-Spin-Resonance Spectra of Spin-Labeled Cowpea Chlorotic Mottle Virus. *J Magn Reson* 1986, *66* (1), 1-8.

106. Vriend, G.; Hemminga, M. A.; Haasnoot, C. A. G.; Hilbers, C. W., A Two-Dimensional Nuclear Overhauser Enhancement Nmr-Spectroscopy Study at 500 Mhz on Cowpea Chlorotic Mottle Virus Protein Assembled in Spherical Capsids. *J Magn Reson* 1985, *64* (3), 501-505.

107. Vriend, G.; Hemminga, M. A.; Verduin, B. J. M.; Schaafsma, T. J., Swelling of Cowpea Chlorotic Mottle Virus Studied by Proton Nuclear Magnetic-Resonance. *Febs Lett* 1982, *146* (2), 319-321.

108. Vriend, G.; Schilthuis, J. G.; Verduin, B. J. M.; Hemminga, M. A., Saturation-Transfer Electron-Spin-Resonance Spectroscopy on Maleimide Spin-Labeled Cowpea Chlorotic Mottle Virus. *J Magn Reson* 1984, *58* (3), 421-427.

109. Vriend, G.; Verduin, B. J. M.; Hemminga, M. A., Role of the N-Terminal Part of the Coat Protein in the Assembly of Cowpea Chlorotic Mottle Virus - a 500 Mhz Proton Nuclear-Magnetic-Resonance Study and Structural Calculations. *Journal of Molecular Biology* 1986, *191* (3), 453-460.

110. Caravan, P.; Greenwood, J. M.; Welch, J. T.; Franklin, S. J., Gadolinium-binding helix-turn-helix peptides: DNA-dependent MRI contrast agents. *Chemical Communications* 2003, (20), 2574-2575.

111. Lauffer, R. B., Paramagnetic Metal-Complexes as Water Proton Relaxation Agents for Nmr Imaging - Theory and Design. *Chem Rev* 1987, *87* (5), 901-927.

112. Raymond, K. N.; Pierre, V. C., Next Generation, High Relaxivity Gadolinium MRI Agents. *Bioconjugate Chem.* 2005, *16* (1), 3-8.

113. Caravan, P.; Ellison, J. J.; McMurry, T. J.; Lauffer, R. B., Gadolinium(III) chelates as MRI contrast agents: Structure, dynamics, and applications. *Chem Rev* 1999, *99* (9), 2293-2352.

114. Penfield, J. G.; Reilly, R. F., Jr., What nephrologists need to know about gadolinium. *Nat Clin Pract Nephrol* 2007, *3* (12), 654-68.

115. Douglas, T.; Young, M., Host-Guest Encapsulation of Materials by Assembled virous protein cages. *Nature* 1998, *393*, 152-155.

116. Uchida, M.; Klem, M. T.; Allen, M.; Suci, P.; Flenniken, M.; Gillitzer, E.; Varpness, Z.; Liepold, L. O.; Young, M.; Douglas, T., Biological containers: Protein cages as multifunctional nanoplatforms. *Adv Mater* 2007, *19* (8), 1025-1042.

117. Uchida, M.; Willits, D. A.; Muller, K.; Willis, A. F.; Jackiw, L.; Jutila, M.; Young, M. J.; Porter, A. E.; Douglas, T., Intracellular Distribution of Macrophage Targeting Ferritin-Iron Oxide Nanocomposite. *Adv Mater* 2009, *21* (4), 458-+.

118. Hooker, J. M.; Datta, A.; Botta, M.; Raymond, K. N.; Francis, M. B., Magnetic resonance contrast agents from viral capsid shells: a comparison of exterior and interior cargo strategies. *Nano letters* 2007, *7* (8), 2207-10.

119. Prasuhn, D. E., Jr.; Yeh, R. M.; Obenaus, A.; Manchester, M.; Finn, M. G., Viral MRI contrast agents: coordination of Gd by native virions and attachment of Gd complexes by azide-alkyne cycloaddition. *Chem Commun (Camb)* 2007, (12), 1269-71.

120. Destito, G.; Yeh, R.; Rae, C. S.; Finn, M. G.; Manchester, M., Folic acid-mediated targeting of cowpea mosaic virus particles to tumor cells. *Chem Biol* 2007, *14* (10), 1152-62.

121. Kaiser, C. R.; Flenniken, M. L.; Gillitzer, E.; Harmsen, A. L.; Harmsen, A. G.; Jutila, M. A.; Douglas, T.; Young, M. J., Biodistribution studies of protein cage nanoparticles demonstrate broad tissue distribution and rapid clearance in vivo. *International journal of nanomedicine* 2007, *2* (4), 715-33.

122. Kaltgrad, E.; O'Reilly, M. K.; Liao, L.; Han, S.; Paulson, J. C.; Finn, M. G., On-virus construction of polyvalent glycan ligands for cell-surface receptors. *Journal of the American Chemical Society* 2008, *130* (14), 4578-9.

123. Koudelka, K. J.; Destito, G.; Plummer, E. M.; Trauger, S. A.; Siuzdak, G.; Manchester, M., Endothelial targeting of cowpea mosaic virus (CPMV) via surface vimentin. *PLoS pathogens* 2009, *5* (5), e1000417.

124. Koudelka, K. J.; Rae, C. S.; Gonzalez, M. J.; Manchester, M., Interaction between a 54-kilodalton mammalian cell surface protein and cowpea mosaic virus. *J Virol* 2007, *81* (4), 1632-40.

125. Kovacs, E. W.; Hooker, J. M.; Romanini, D. W.; Holder, P. G.; Berry, K. E.; Francis, M. B., Dual-surface-modified bacteriophage MS2 as an ideal scaffold for a viral capsid-based drug delivery system. *Bioconjugate chemistry* 2007, *18* (4), 1140-7.

126. Rae, C. S.; Khor, I. W.; Wang, Q.; Destito, G.; Gonzalez, M. J.; Singh, P.; Thomas, D. M.; Estrada, M. N.; Powell, E.; Finn, M. G.; Manchester, M., Systemic trafficking of plant virus nanoparticles in mice via the oral route. *Virology* 2005, *343* (2), 224-35.

127. Raja, K. S.; Wang, Q.; Finn, M. G., Icosahedral virus particles as polyvalent carbohydrate display platforms. *Chembiochem* 2003, *4* (12), 1348-51.

128. Shriver, L. P.; Koudelka, K. J.; Manchester, M., Viral nanoparticles associate with regions of inflammation and blood brain barrier disruption during CNS infection. *Journal of neuroimmunology* 2009, *211* (1-2), 66-72.

129. Singh, P.; Prasuhn, D.; Yeh, R. M.; Destito, G.; Rae, C. S.; Osborn, K.; Finn, M. G.; Manchester, M., Bio-distribution, toxicity and pathology of cowpea mosaic virus nanoparticles in vivo. *J Control Release* 2007, *120* (1-2), 41-50.

130. Steinmetz, N. F.; Manchester, M., PEGylated viral nanoparticles for biomedicine: the impact of PEG chain length on VNP cell interactions in vitro and ex vivo. *Biomacromolecules* 2009, *10* (4), 784-92.

131. Uchida, M.; Flenniken, M. L.; Allen, M.; Willits, D. A.; Crowley, B. E.; Brumfield, S.; Willis, A. F.; Jackiw, L.; Jutila, M.; Young, M. J.; Douglas, T., Targeting of cancer cells with ferrimagnetic ferritin cage nanoparticles. *Journal of the American Chemical Society* 2006, *128* (51), 16626-33.

132. Flenniken, M. L.; Uchida, M.; Liepold, L. O.; Kang, S.; Young, M. J.; Douglas, T., A library of protein cage architectures as nanomaterials. *Curr Top Microbiol Immunol* 2009, *327*, 71-93.

133. Powell, D. H.; Dhubhghaill, O. M. N.; Pubanz, D.; Helm, L.; Lebedev, Y. S.; Schlaepfer, W.; Merbach, A. E., Structural and Dynamic Parameters Obtained from 17O NMR, EPR, and NMRD Studies of Monomeric and Dimeric Gd3+ Complexes of Interest in Magnetic Resonance Imaging: An Integrated and Theoretically Self-Consistent Approach1. *Journal of the American Chemical Society* 1996, *118* (39), 9333-9346.

134. Livramento, J. B.; Helm, L.; Sour, A.; O'Neil, C.; Merbach, A. E.; Toth, E., A benzene-core trinuclear GdIII complex: towards the optimization of relaxivity for MRI contrast agent applications at high magnetic field. *Dalton Trans* 2008, (9), 1195-202.

# APPENDICES

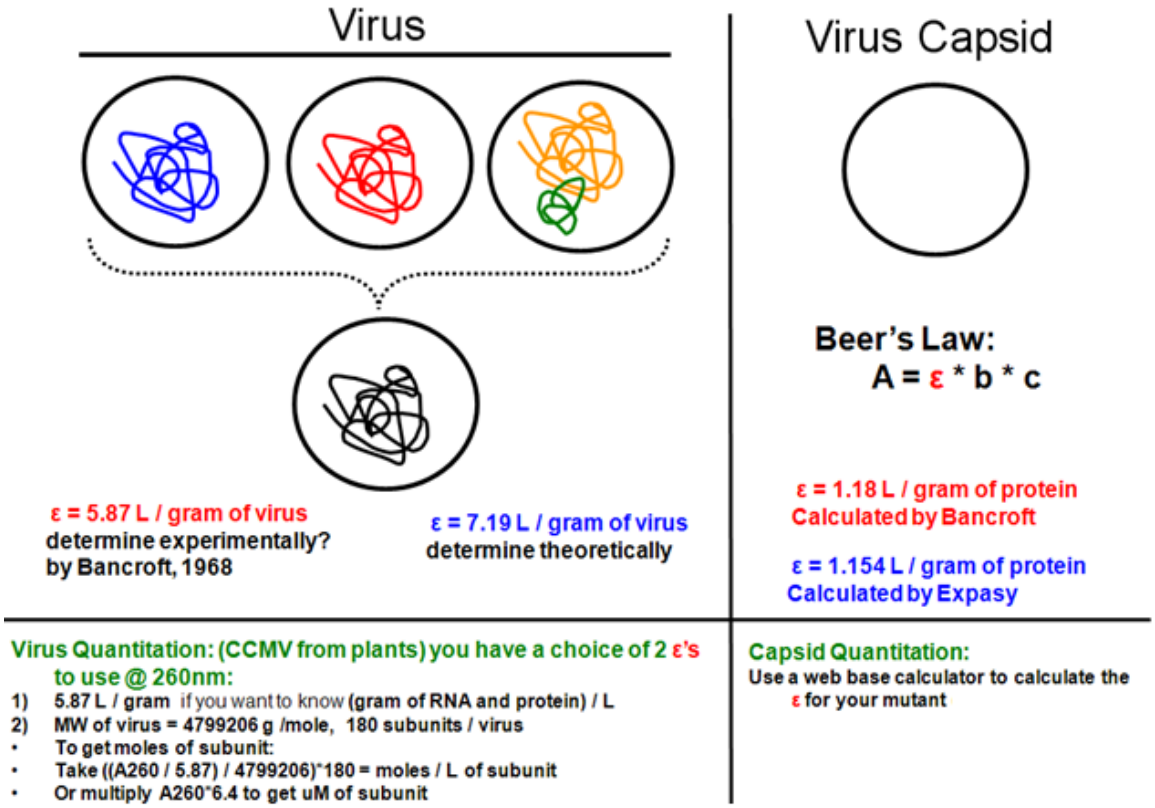
# APPENDIX A

## SUPPORTING INFORMATION FOR CHAPTER 4

# APPENDIX A

## Quantitation of CCMV Virus by absorbance at 260 or 280nm

There are three types of CCMV virus particles each filled with a unique set of RNA molecules.




## Auto-cleavage sites of various mutants of HSP

Below are the cleavage sites of HSP as determined by LC/MS/MS.

```
                MGDKDGDGWLEF/EEGGGGTGIQISGKGFMPIS..... wtHSPdelta2-15,hand16-32
MFGRDPFDSLFERMFKEFFATPMTGTT/MIQSSTGIQISGKGFMPIS..... wtHSP
MFGRDPFDSLFERMFKEFFATPMTGTT/MIQSSTGIQISGKCFMPIS..... G41C
MFGRDPFDSLFERMFKEFFATPMTGTT/MIQSSTGIQISGKCFMPIS..... RGD WXE
```

# APPENDIX B

## SUPPORTING INFORMATION FOR CHAPTER 5

# APPENDIX B

## Relaxivity Equations

Analytical equation containing the decay constant, $T_1$, for the recovery of the net nuclear spin magnetization for a sample placed in a magnetic field which has be tilted out of equilibrium:

$$M_z(t) = M_{z\ Eqilibrium}\left(1 - e^{\frac{-t}{T_1}}\right)$$

| | |
|---|---|
| $M_z(t)$ | nuclear spin magnetization in the z axis at time t (seconds) |
| $M_{z\ Equilibrium}$ | equilibrium state of the nuclear spin magnetization in the z axis |
| $T_1$ | decay constant for the recovery of spin (seconds) |

Observed $T_1$ of a specific sample type with a contrast agent present:

$$T_{1\ observed} = \left[\frac{1}{T_{1\ sample}} + (r_1)[Contrast\ Agent]\right]^{-1}$$

| | |
|---|---|
| $r_1$ | relaxivity of a contrast agent ($mM^{-1}$ $seconds^{-1}$) |
| [*Contrast Agent*] | concentration of the contrast agent (mM) |

*Solomon-Bloembergen-Morgan (SBM) model for PRE:*

Relaxivity of contrast agent including the dipolar, scalar and Currie relaxation mechanisms:

$$r_1 = \frac{q \cdot [Constrast\ Agent]}{[Water]}\left[\frac{1}{T_{1M} + \tau_M}\right]$$

$$r_1 = \frac{q \cdot [Constrast\ Agent]}{[Water]}\left[\frac{1}{\left(\frac{1}{T_{1M}^{dipolar}} + \frac{1}{T_{1M}^{scalar}} + \frac{1}{T_{1M}^{currie}}\right)^{-1} + \tau_M}\right]$$

| | |
|---|---|
| $[Water] = 55.6$ | Molar concentration of water in units of (moles / liter) |
| q | number of inner sphere waters that bind to the Gd ion, (fitting parameter) |
| $T_{1M}^{dipolar}$ | dipolar contribution to the relaxation time (seconds) |
| $T_{1M}^{scalar}$ | scalar contribution to the relaxation time (seconds) |
| $T_{1M}^{currie}$ | Currie contribution to the relaxation time (seconds) |
| $\tau_M$ | residence time for the Gd bound water molecule, (seconds) (fitting parameter) |

Relaxivity of contrast agent considering only the dipolar relaxation mechanism (the dipolar mechanism was only considered in the fitting of the NMRD profiles in this work):

$$r_1 = \frac{q \cdot [Constrast\ Agent]}{[Water]}\left[\frac{1}{T_{1M}^{dipolar} + \tau_M}\right]$$

SBM analytical description of the dipolar relaxation time:

$$T_{1M}^{dipolar} = \frac{2C_{dd}}{15r_{IS}^6}[3J(\omega_I, \tau_{d1}) + 7J(\omega_s, \tau_{d2})]$$

Prefactor for relaxation:

$$\boldsymbol{C_{dd} = \gamma_I^2 \gamma_S^2 \hbar^2 S(S+1)\left(\frac{\mu_0}{4\pi}\right)^2}$$

$\gamma_I = 2.675 \cdot 10^8$ nuclear gyromagnetic ratio (second$^{-1}$ Tesla$^{-1}$)

$\gamma_S = -1.760859778 \cdot 10^{11}$ electronic gyromagnetic ratio (second$^{-1}$ Tesla$^{-1}$)

$\hbar = 1.054571628 \cdot 10^{-34}$ Planck constant (Joules · seconds)

$S = 7/2$ spin quantum number for the $Gd^{3+}$ ion

$\mu_0 = 4\pi \cdot 10^{-7}$ magnetic permeability of free space (Newton · Amps$^{-2}$)

$r_{IS}^6 = 3 \cdot 10^{-10}$ distance between the nuclear and the electronic spin (meters)

Spectral density function:

$$J(\omega, \tau) = \frac{\tau}{1 + \omega^2 \tau^2}$$

| | |
|---|---|
| $\omega$ | Larmor frequency of nuclear or electric spin |
| $\tau$ | correlation time where $\tau$ is either $\tau_{d1}$ or $\tau_{d2}$ |

Correlation times in units of seconds:

$$\tau_{d1} = \left(\frac{1}{\tau_R} + \frac{1}{\tau_M} + \frac{1}{T_{1e}}\right)^{-1}$$

and

$$\tau_{d2} = \left(\frac{1}{\tau_R} + \frac{1}{\tau_M} + \frac{1}{T_{2e}}\right)^{-1}$$

| | |
|---|---|
| $\tau_R$ | rotational correlation time for the Gd ion (seconds) (fitting parameter) |
| $T_{1e}$ | longitudinal electronic relaxation time (seconds) |
| $T_{2e}$ | transverse electronic relaxation time (seconds) |

Electronic relaxation time (longitudinal and transverse) (seconds):

$$T_{1e} = \left[\frac{2\Delta^2}{50}(4S(S+1) - 3)\left(\frac{\tau_v}{1 + \omega_S^2\tau_v^2} + \frac{4\tau_v}{1 + 4\omega_S^2\tau_v^2}\right)\right]^{-1}$$

and

$$T_{2e} = \left[\frac{\Delta^2}{50}(4S(S+1)-3)\left(3\tau_v + \frac{5\tau_v}{1+\omega_S^2\tau_v^2} + \frac{2\tau_v}{1+4\omega_S^2\tau_v^2}\right)\right]^{-1}$$

$\tau_v = 1.4 \cdot 10^{-11}$ correlation time for instantaneous distortions of the metal complex polyhedron (seconds)

$\Delta^2 = 9 \cdot 10^{18}$ mean square fluctuation of the zero-field splitting

**Fitting scheme**

Four fitting parameters were used ($\tau_R$, $\tau_M$, $\tau_V$ and $\Delta$) to individually fit all twenty NMRD profiles while q was held at a value of 1 in these SBM fits. (20 profiles result from 2 preparations for both non-passivated and passivated batches for G0.0, G0.5, G1.5, G2.5 G3.5) Next average values for $\tau_V$ and $\Delta$ were calculated from the 20 individually fit profiles. This resulted in: $\Delta = 3.0 \cdot 10^9 \pm 0.3 \cdot 10^9$ and $\tau_V = 1.4 \cdot 10^{-11} \pm 0.2 \cdot 10^{-11}$.

The average values for $\tau_V$ and $\Delta$ were used to fit the averaged NMRD profile (all twenty data sets averaged) with q, $\tau_R$ and $\tau_M$ all set as fitting parameters. The individual NMRD profiles were also fit in this manner.

**Mathcad File Pages for Fitting $T_1$ Relaxivity Data to the SBM Model**

Shown below are screen images of the file used to fit relaxivity data to the SBM model for relaxivity. Mathcad, version 14, was used for this fitting. Colored text is not used by the program for any calculations but is shown as a guide.

$\gamma I := 2.675 \cdot 10^{8}$ nuclear gyromagnetic ratio ( second^-1 * Tesla^-1)

$\gamma S := -1.760859778 \cdot 10^{11}$ electronic gyromagnetic ratio ( second^-1 * Tesla^-1)

$\mu 0 := 4\pi \cdot 10^{-7}$ magnetic permeability of free space (Newton * Amps^-2)

$SGd := 3.5$ electron spin quantum number fpr Gd (unitless)

$h := 1.054571628 \cdot 10^{-34}$ Planck constant (Joules * seconds)

$kb := 1.3806504 \cdot 10^{-23}$ Boltzmann constant (Joules * Kelvin^-1)

$Temp := 310.15$ Temperature the relaxivity experiments were carried out (Kelvin)

$Ris := 3 \cdot 10^{-10}$ Distance between Gd ion and bound water (meters)

$WaterConcentration := 55.6$ Concentration of water (Moles / Liter)

$M := 1 \cdot 10^{-3}$ Millimolar concentration of water (1/1000 * Moles / Liter)

$\Delta := 3 \cdot 10^{9}$ Magnitude of ZFS interaction (unitless)

$tv := 1.4 \cdot 10^{-11}$ Correlation time for transient ZFS fluctuation (seconds)

$\tau R := 4.259 \cdot 10^{-6}$ Rotational correlation of order k (seconds)

$\tau M := 10 \cdot 10^{-9}$ mean water exchange lifetime for Gd bound waters (seconds)

$q := 1$ Number of waters bound (unitless)

Input relaxivity data below:

$$AveR1 := \begin{pmatrix} 18.896 \\ 19.711 \\ 23.668 \\ 25.403 \\ 25.169 \\ 22.175 \\ 18.889 \\ 16.105 \end{pmatrix} \quad Field := \begin{pmatrix} 61.7 \\ 56.0 \\ 42.6 \\ 31.0 \\ 21.3 \\ 14.9 \\ 8.5 \\ 4.3 \end{pmatrix} \quad FieldT := \frac{Field}{42.57} \quad FieldT = \begin{pmatrix} 1.449 \times 10^{0} \\ 1.315 \times 10^{0} \\ 1.001 \times 10^{0} \\ 7.282 \times 10^{-1} \\ 5.004 \times 10^{-1} \\ 3.5 \times 10^{-1} \\ 1.997 \times 10^{-1} \\ 1.01 \times 10^{-1} \end{pmatrix} \quad AveR2 := 32.8 \quad FieldT2 := 1$$

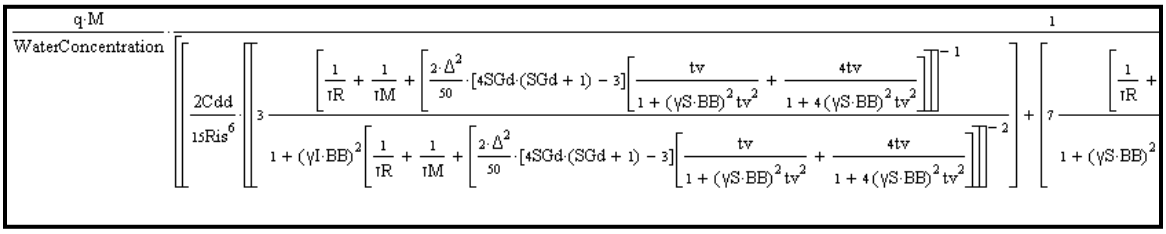
Relaxivity consider only dipolar interaction = interaction between nuclear magnetic moment and the electrons outside the nucleus.

$Cdd := \gamma I^2 \gamma S^2 h^2 SGd \cdot (SGd + 1) \left(\frac{\mu 0}{4\pi}\right)^2$ Prefactor for dipolar relaxivity mechanism

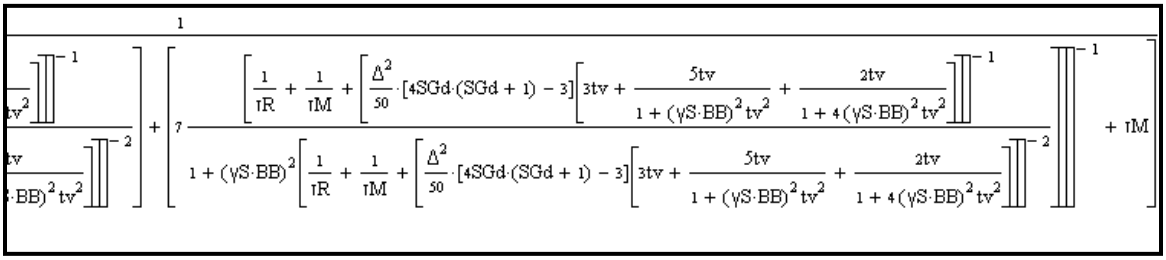
$Cdd = 3.89 \times 10^{-42}$

$$R1\_dipolar(BB, \tau R, \tau M, q) := \frac{q \cdot M}{WaterConcentration} \left[ \frac{2Cdd}{15Ris^6} \left[ 3 \frac{\left[ \frac{1}{\tau R} + \frac{1}{\tau M} + \left[ \frac{6\Delta^2}{50} [4SGd \cdot (SGd + 1) - 3] \left[ \frac{tv}{1 + (\gamma S \cdot BB)^2 tv^2} + \frac{4tv}{1 + 4(\gamma S \cdot BB)^2 tv^2} \right] \right]^{-1}}{1 + (\gamma I \cdot BB)^2 \left[ \frac{1}{\tau R} + \frac{1}{\tau M} + \left[ \frac{2\Delta^2}{50} [4SGd \cdot (SGd + 1) - 3] \left[ \frac{tv}{1 + (\gamma S \cdot BB)^2 tv^2} + \frac{4tv}{1 + 4(\gamma S \cdot BB)^2 tv^2} \right] \right]^{-2}} + 7 \frac{\left[ \frac{1}{\tau R} + \frac{1}{\tau M} + \left[ \frac{\Delta^2}{50} [4SGd \cdot (SGd + 1) - 3] \left[ 3tv + \frac{5tv}{1 + (\gamma S \cdot BB)^2 tv^2} + \frac{2tv}{1 + 4(\gamma S \cdot BB)^2 tv^2} \right] \right]^{-1}}{1 + (\gamma S \cdot BB)^2 \left[ \frac{1}{\tau R} + \frac{1}{\tau M} + \left[ \frac{\Delta^2}{50} [4SGd \cdot (SGd + 1) - 3] \left[ 3tv + \frac{5tv}{1 + (\gamma S \cdot BB)^2 tv^2} + \frac{2tv}{1 + 4(\gamma S \cdot BB)^2 tv^2} \right] \right]^{-2}} \right]^{-1} + \tau M \right]$$

Zoomed in view of the left side of the above equation:



Zoomed in view of the right side of the above equation:



"Mathcad Solve Block" to fit relaxivity data to the SBM model:

**Step 1 - Input guess values for q, tR, tM:**

$q\_G := 2 \times 10^{0}$

$\tau R\_G := 3.631 \times 10^{-10}$

$\tau M\_G := 1.122 \times 10^{-9}$

resid_dipolar(FieldT, τR_G, τM_G, q_G) := $\overrightarrow{\text{AveR1 - R1\_dipolar(FieldT, τR\_G, τM\_G, q\_G)}}$

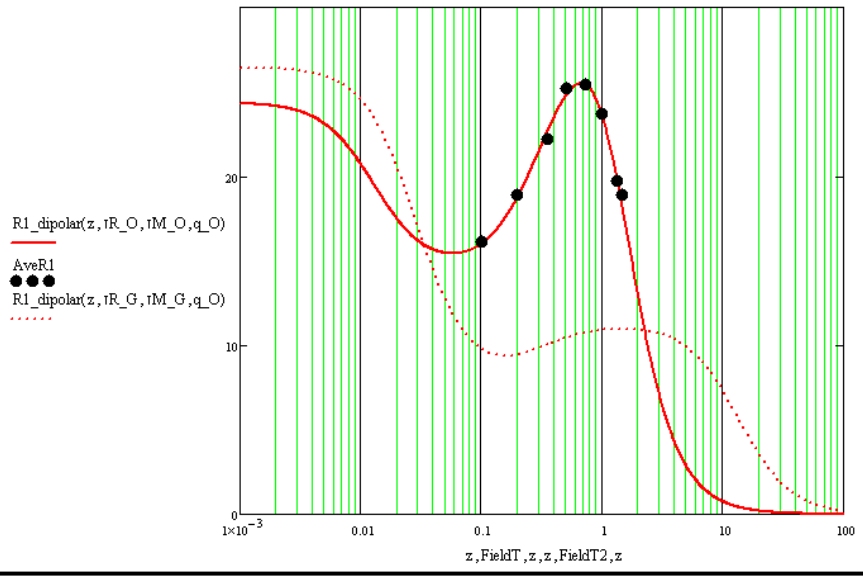
Given

0 = resid_dipolar(FieldT, τR_G, τM_G, q_G)

$\begin{pmatrix} \tau R\_O \\ \tau M\_O \\ q\_O \end{pmatrix}$ := Minerr(τR_G, τM_G, q_G)

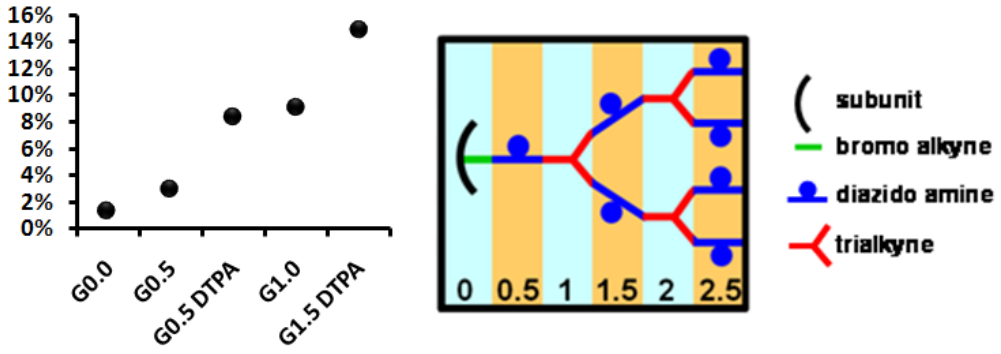
**Fitting Results:**

$\begin{pmatrix} \tau R\_O \\ \tau M\_O \\ q\_O \end{pmatrix} = \begin{pmatrix} 7.62 \times 10^{-9} \\ 5.577 \times 10^{-7} \\ 1.069 \times 10^{0} \end{pmatrix}$



## Details on the Construction of the Branched Polymer Modeled into the HSP Structure

| Generation | Volume addition at each gen (A^3) | Total volume added | % interior occupied | Added groups |
|---|---|---|---|---|
| G0.0 | 1332 | 2664 | 1.3% | 24 bromo alkynes |
| G0.5 | 1568 | 5800 | 2.8% | 24 diazido amines |
| G0.5 DTPA | 5542 | 16884 | 8.3% | 24 DTPA(Gd)s |
| G1.0 | 723 | 18330 | 9.0% | 20 total (4 are in xlinks, 16 are not xlinked) trialkynes |
| G1.5 DTPA | 5982 | 30294 | 14.9% | 20 DTPA(Gd)s |




## Protein and Gd Quantitation

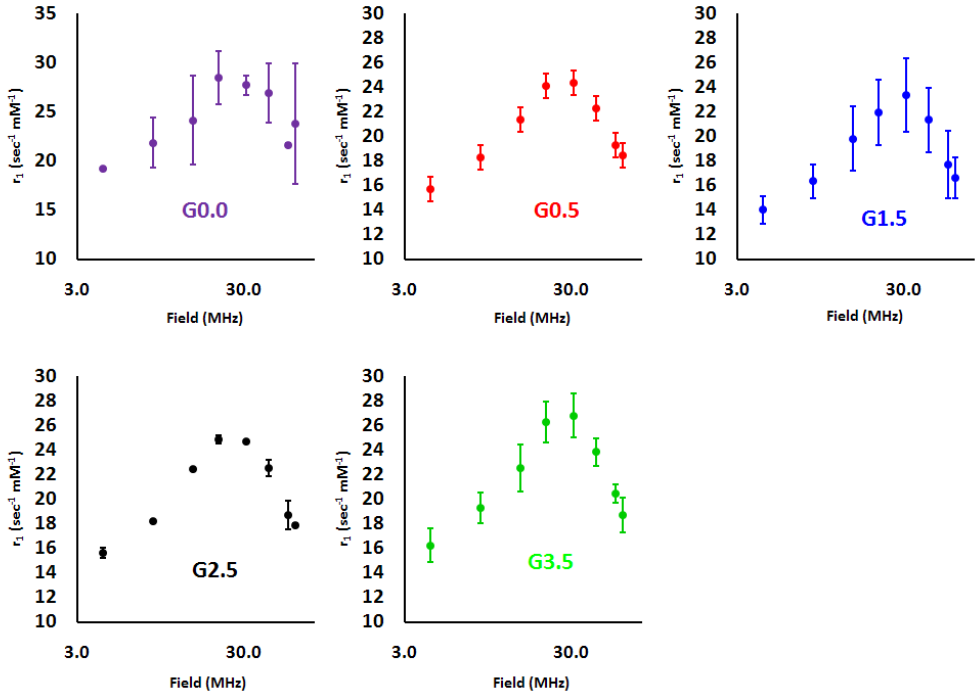
A BCA Protein Assay (bicinchoninic acid) from Pierce was used to quantitate the protein (HSP) concentration (www.piercenet.com) and the protocol provided by Pierce was used. Protein samples were analyzed in triplicate resulting in an average relative standard deviation of 2.1% for the twenty samples for the BCA assay. Energy Laboratories, Inc. performed the ICP-MS quantitation of the the Gd ions (www.energylab.com).

## Comparison of Protein Cage – Gd Based Contrast Agents

| Cage / Chelator | Ionic $r_1$ (MHz) | Particle $r_1$ | Cage Diameter (nm) | Particle $r_1$ per volume ($sec^{-1} mM^{-1} nm^{-3}$) | Particle $r_1$ per mass[2] ($L g^{-1} sec^{-1}$) | Clinically relevant binding | Group Reference |
|---|---|---|---|---|---|---|---|
| CCMV / endogenous binding site | 202 (62) | 28,482[a] | 28 | 2.5 | 7.8 | No | Douglas / Young [3] |
| CCMV / metal binding peptide – genetic fusion | 210 (62) | 36,120[a] | 28 | 3.1 | 9.9 | No | Douglas / Young [4] |
| MS2 / DTPA-ITC | 16.9 (64) | 7,200 | 27 | 0.7 | 2.9 | Yes | Kirshenbaum [5] |
| CPMV / DOTA-click | 15.5 (64) | 4,150 | 30 | 0.3 | 1.1 | Yes | Finn [6] |
| CCMV / DOTA-NHS ester | 46 (62) | 2,806 | 28 | 0.2 | 0.8 | Yes | Douglas / Young [4] |
| MS2 / bis(HOPO)-TAM | 31 (60) | 2,900 | 27 | 0.3 | 1.2 | Yes | Francis [7, 8] |
| HSP-BP-DTPA-Gd | 19 (62) | 3,450 | 14 | 2.4 | 8.7 | Yes | Douglas / Young this work |
| DTPA-Gd | 4 (20) | 4 | | 12.2 | 7.3 | Yes | [9] |

**NMRD Profile and SBM fit of Generations G0.0, G0.5, G1.5, G2.5 and G3.5**

Four experimental data points are averaged for all points in the plots below. These four points are comprised of two points from the passivated preparations and two from the non-passivated preparations. The error bars are plus and minus one standard deviation.



**Einstein Stokes Relation to Estimate the Rotational Correlation Time of HSP in Water**

$$\tau_R = \frac{4\pi\eta r^3}{3kT}$$

$\tau_R$ = rotational correlation time (seconds)

$\eta$ = Viscosity of water ( water @ 20 $^0$C = 0.001002 Pa sec = 0.001002 kg m$^{-1}$ sec$^{-1}$)

r = radius of the particle (HSP = 6nm = 6 · 10$^{-9}$ m)

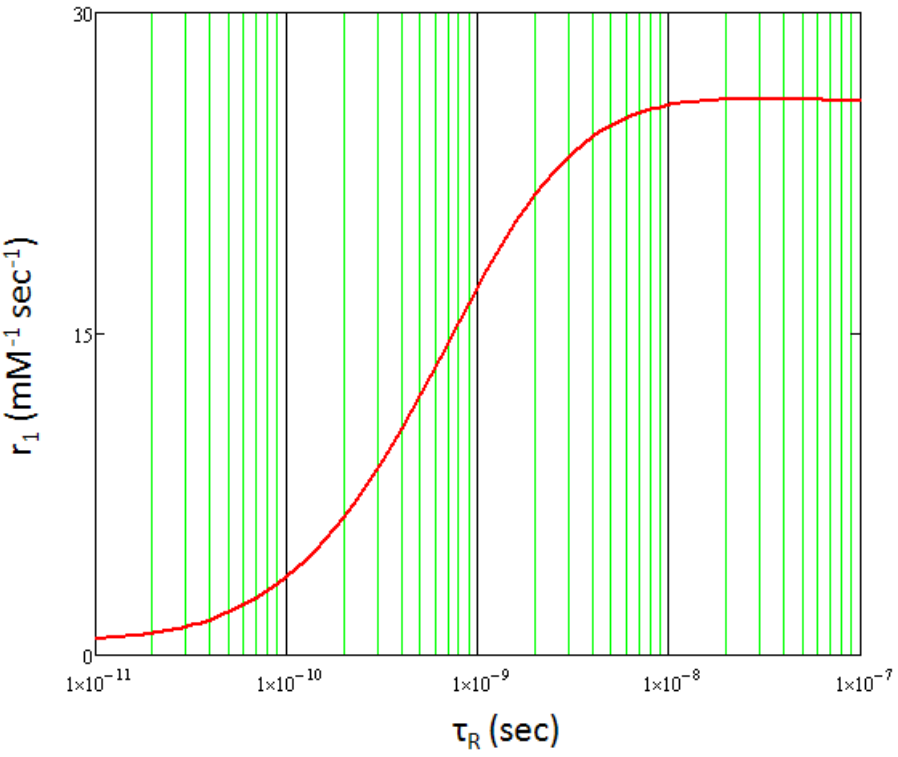
k = boltzman constant (1.3806504 · 10$^{-23}$ kg m$^2$ sec$^{-2}$ K$^{-1}$)

K = temperature in Kelvin

Excel equation:
=(4*PI()*0.001002*(6*10^-9)^3)/(3*(1.38*10^-23)*296) = 2.2 *10$^{-7}$

Units:
=(kg m$^{-1}$ sec$^{-1}$*(m)^3) / ( ( kg m$^2$ sec$^{-2}$ K$^{-1}$)*K) = seconds

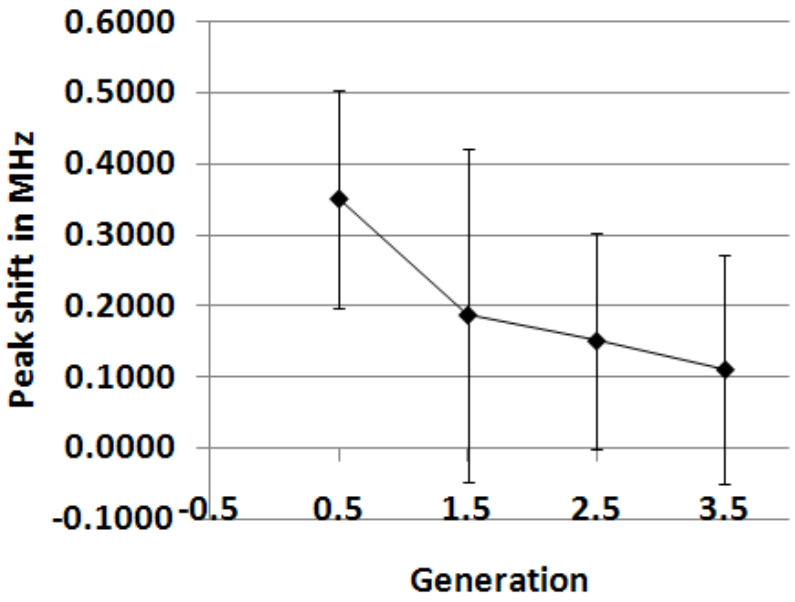
**Relationship of $r_1$ and $\tau_R$ Determined by the SBM Model**




q = 1.1, $\tau_M = 5.6*10^{-7}$ seconds and the magnetic field = 31MHz (0.73 Tesla):

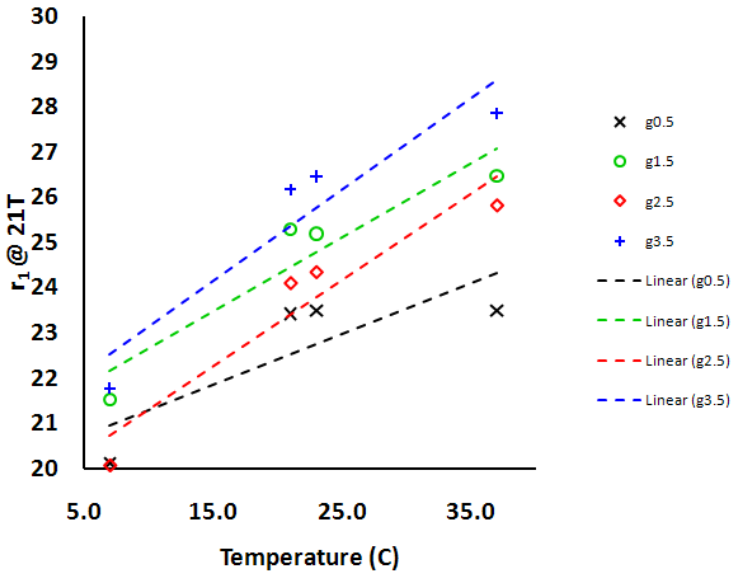
**Peak Shift in MHz for the Peak Located at Approximately 31MHz**

Four experimental data points are averaged for the points below. These four points are comprised of two points from the passivated preparations and two from the non-passivated preparations. The error bars are plus and minus one standard deviation.



**Plots of $r_1$ vs. Temperature for G0.5, G1.5, G2.5 and G3.5**

Four experimental data points are averaged for the points below. These four points are comprised of two points from the passivated preparations and two from the non-passivated preparations.



# APPENDIX C

## IGOR MACRO FOR CHARGE STATE ASSIGNMENT OF NONCOVALENT ESI MASS SPECTRA

# APPENDIX C

```
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Macro LoadSpectrum()

killwaves/Z wave0
killwaves/Z wave1

DoWindow/K IonSummary
DoWindow/K RawData
DoWindow/K Summary0
DoWindow/K WidthVsIonNumber
DoWindow/K ChargestateIteration
DoWindow/K MassVsChargestate
DoWindow/K MassVsCorrectedIonPeakWidth

LoadWave/G/D/A=wave

make/O/N=(numpnts(wave0)) RawMZ
make/O/N=(numpnts(wave0)) RawInt



RawMZ = wave0
RawInt = wave1

Display/N=RawData RawInt vs RawMZ

SetAxis/W=RawData/A

ModifyGraph highTrip(bottom)=1e+07;DelayUpdate

Label bottom "m/z"

ModifyGraph width=300,height=200

RemoveFromGraph/W=RawData/Z CS0
RemoveFromGraph/W=RawData/Z CS1
RemoveFromGraph/W=RawData/Z CS2
```

```
RemoveFromGraph/W=RawData/Z CS3
RemoveFromGraph/W=RawData/Z CS4
RemoveFromGraph/W=RawData/Z CS5
RemoveFromGraph/W=RawData/Z CS6
RemoveFromGraph/W=RawData/Z CS7
RemoveFromGraph/W=RawData/Z CS8
RemoveFromGraph/W=RawData/Z CS9
RemoveFromGraph/W=RawData/Z CS10
RemoveFromGraph/W=RawData/Z CS11
RemoveFromGraph/W=RawData/Z CS12
RemoveFromGraph/W=RawData/Z CS13
RemoveFromGraph/W=RawData/Z CS14
RemoveFromGraph/W=RawData/Z CS15
RemoveFromGraph/W=RawData/Z CS16
RemoveFromGraph/W=RawData/Z CS17
RemoveFromGraph/W=RawData/Z CS18
RemoveFromGraph/W=RawData/Z CS19
RemoveFromGraph/W=RawData/Z CS20
RemoveFromGraph/W=RawData/Z fit_RawInt

make/O/N=(200) res
make/O/N=(2) Mr

Doalert 1,"Do you want to see suggestion boxes during this analysis session?"
res(199) = V_flag

if (res(199) ==1)
Doalert 0,"To zoom into the area of interest:\rLeft click and make a box around the area
of interest, then right click inside the box and select [Expand].\rTo zoom out press
Ctrl+A"
endif

EndMacro

//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Macro InitialMassGuess(yval,zval,aval,bval,cval)
Variable yval=res(21), zval=res(9), aval=res(10), bval = res(111), cval = res(112) //
declare numeric params
Prompt yval, "guess mass"
Prompt zval, "number of ions left of biggest ion" // set prompt for xval param
```

```
Prompt aval, "number of ions right of biggest ion"
Prompt bval, "low m/z limit in the range of interest"
Prompt cval, "high m/z limit in the range of interest"
res(21) = yval
res(9) = zval
res(10) = aval
res(111) = bval
res(112) = cval

DoWindow/K IonSummary
DoWindow/K Summary0
DoWindow/K WidthVsIonNumber
DoWindow/K ChargestateIteration
DoWindow/K MassVsChargestate
DoWindow/K MassVsCorrectedIonPeakWidth

Tag/W=RawData/K/N=text0

RemoveFromGraph/W=RawData/Z PeakInt
RemoveFromGraph/W=RawData/Z SmoothRawInt

RemoveFromGraph/W=RawData/Z CS0
RemoveFromGraph/W=RawData/Z CS1
RemoveFromGraph/W=RawData/Z CS2
RemoveFromGraph/W=RawData/Z CS3
RemoveFromGraph/W=RawData/Z CS4
RemoveFromGraph/W=RawData/Z CS5
RemoveFromGraph/W=RawData/Z CS6
RemoveFromGraph/W=RawData/Z CS7
RemoveFromGraph/W=RawData/Z CS8
RemoveFromGraph/W=RawData/Z CS9
RemoveFromGraph/W=RawData/Z CS10
RemoveFromGraph/W=RawData/Z CS11
RemoveFromGraph/W=RawData/Z CS12
RemoveFromGraph/W=RawData/Z CS13
RemoveFromGraph/W=RawData/Z CS14
RemoveFromGraph/W=RawData/Z CS15
RemoveFromGraph/W=RawData/Z CS16
RemoveFromGraph/W=RawData/Z CS17
RemoveFromGraph/W=RawData/Z CS18
RemoveFromGraph/W=RawData/Z CS19
RemoveFromGraph/W=RawData/Z CS20
RemoveFromGraph/W=RawData/Z fit_RawInt
```

```
variable LowRange, HighRange

findlevel RawMZ res(111)
LowRange = V_LevelX

findlevel RawMZ res(112)
HighRange = V_LevelX

res(113) = LowRange
res(114) = HighRange

SetAxis/W=RawData left 0,1
SetAxis/W=RawData bottom res(111),res(112)

wavestats/R=(LowRange,HighRange) RawInt

Variable BiggestIon, IntensityOfBiggestIon

BiggestIon = RawMZ(V_maxloc)
IntensityOfBiggestIon = V_max

res(89) = BiggestIon

Variable GuessChargestate

GuessChargestate = round( (yval/BiggestIon) )
res(89) = GuessChargestate

res(0)=1 //resMS
res(2)= (1+res(9)+res(10)) //number_of_Charges
res(1)= (GuessChargestate-res(10))//first_charge
res(3)=6 //ChargestateIntensityDistribution
res(4)= GuessChargestate//round((Mr(0)/1000)^(2/3))//MaxCharge

RawInt = RawInt / IntensityOfBiggestIon

Mr(0) = ((GuessChargestate*BiggestIon)-GuessChargestate)

make/O/N=(res(2)) Charges
make/O/N=(res(2)) PeakWidth

Redimension/N=(res(2)) Charges
Redimension/N=(res(2)) PeakWidth
```

```
Variable i

do
Charges(i)= i + res(1)
i += 1 // increment our loop variable
while(i < (numpnts(Charges)) )

make/O/N=(res(2)) CSI
Redimension/N=(numpnts(Charges)) CSI

CSI = 1

wavestats RawMZ
make/O/N = (round(V_max)) m_z_range

m_z_range = x

make/O/N=(numpnts(m_z_range)) m
make/O/N=(numpnts(m_z_range)) n

Redimension/N=(numpnts(m_z_range)) m
Redimension/N=(numpnts(m_z_range)) n

m = 0
n = 0

RemoveFromGraph/W=RawData/Z n
AppendToGraph/W=RawData/L n vs m_z_range
ModifyGraph/W=RawData rgb(n)=(0,0,52224)

Variable j

do
m= exp (-( (( m_z_range - ( (((res(5)*res(10))+Mr(0)+Charges(j)-((j)*res(5)))/Charges(j))
) ) / res(0)) )^2)*CSI(j)
n=m+n
j += 1
while(j <(numpnts(Charges)))

if (res(199) ==1)
Doalert 0,"Does the calculated m/z values (blue) match the raw spectrum well?\r-Yes,
select [Macro] > [FindWidthAndCenter]\r-No: If the calculated m/z values are too wide,
try a larger guess mass. If the calculated m/z values are too narrow try a smaller guess."
endif
```

```
EndMacro

//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Macro FindCenterAndWidth(zval,yval)
Variable zval = res(106), yval = res(55)
Prompt zval, "Enter smoothing value, bigger numbers result in more smoothing."
Prompt yval, "Enter peak offset"
res(55) = (yval)
res(106) = zval
variable width, y, j, x,left,right,s,PeakCenter,LeftSide, RightSide

DoWindow/K IonSummary
DoWindow/K Summary0
DoWindow/K WidthVsIonNumber
DoWindow/K ChargestateIteration
DoWindow/K MassVsChargestate
DoWindow/K MassVsCorrectedIonPeakWidth

Tag/W=RawData/K/N=text0

wavestats/R=(res(113),res(114)) RawInt
RawInt = RawInt + V_min

wavestats/R=(res(113),res(114)) RawInt
RawInt = RawInt / V_max

Duplicate/O RawInt,SmoothRawInt;DelayUpdate
Smooth/B res(106), SmoothRawInt

wavestats/R=(res(113),res(114)) SmoothRawInt
SmoothRawInt = SmoothRawInt - V_min - res(55)
RawInt = RawInt - V_min - res(55)

wavestats/R=(res(113),res(114)) SmoothRawInt
SmoothRawInt = SmoothRawInt / V_max *1.01
RawInt = RawInt / V_max *1.01

RemoveFromGraph/W=RawData/Z SmoothRawInt
AppendToGraph/W=RawData/L SmoothRawInt vs RawMZ
```

```
ModifyGraph/W=RawData rgb(SmoothRawInt)=(0,39168,0)
ModifyGraph/W=RawData lsize(SmoothRawInt)=1.5

make/O/N=(numpnts(Charges)) PeakMaxGuess
redimension/N =(numpnts(Charges)) PeakMaxGuess
PeakMaxGuess = 1

make/O/N=(numpnts(Charges)) PeakMax
redimension/N =(numpnts(Charges)) PeakMax
PeakMax = 1

make/O/N=(numpnts(Charges)) PeakWidth
redimension/N =(numpnts(Charges)) PeakWidth
PeakWidth = 1

make/O/N=(numpnts(Charges)) mass
redimension/N =(numpnts(Charges)) mass
mass = 1

make/O/N=(numpnts(Charges)) PeakInt
redimension/N =(numpnts(Charges)) PeakInt
PeakInt = 1

make/O/N=(numpnts(Charges)) IonNumber
redimension/N =(numpnts(Charges)) IonNumber

variable iii
iii=0

do
IonNumber(iii) = iii
iii = iii+1
while(iii < numpnts(Charges))

IonNumber = IonNumber +1

RemoveFromGraph/W=RawData/Z PeakInt
AppendToGraph/W=RawData/L PeakInt vs PeakMax
ModifyGraph/W=RawData
marker(PeakInt)=42,msize(PeakInt)=10,mrkThick(PeakInt)=1
ModifyGraph/W=RawData mode(PeakInt)=3,rgb(PeakInt)=(0,0,0)

do
PeakMaxGuess(j) = (((res(5)*res(10))+Mr(0)+Charges(j)-((j)*res(5)))/Charges(j))
```

```
j = j+1
while(j <= numpnts(Charges))

RemoveFromGraph/W=RawData/Z CS0
RemoveFromGraph/W=RawData/Z CS1
RemoveFromGraph/W=RawData/Z CS2
RemoveFromGraph/W=RawData/Z CS3
RemoveFromGraph/W=RawData/Z CS4
RemoveFromGraph/W=RawData/Z CS5
RemoveFromGraph/W=RawData/Z CS6
RemoveFromGraph/W=RawData/Z CS7
RemoveFromGraph/W=RawData/Z CS8
RemoveFromGraph/W=RawData/Z CS9
RemoveFromGraph/W=RawData/Z CS10
RemoveFromGraph/W=RawData/Z CS11
RemoveFromGraph/W=RawData/Z CS12
RemoveFromGraph/W=RawData/Z CS13
RemoveFromGraph/W=RawData/Z CS14
RemoveFromGraph/W=RawData/Z CS15
RemoveFromGraph/W=RawData/Z CS16
RemoveFromGraph/W=RawData/Z CS17
RemoveFromGraph/W=RawData/Z CS18
RemoveFromGraph/W=RawData/Z CS19
RemoveFromGraph/W=RawData/Z CS20

variable RangeOfPeakCenterGuess, RRR, LLL,ION

wavestats PeakMaxGuess
ION = V_npnts

findlevel RawMZ PeakMaxGuess(ION-1)
RRR = V_LevelX

findlevel RawMZ PeakMaxGuess(ION-2)
LLL = V_LevelX
RangeOfPeakCenterGuess = (LLL-RRR)/3

y=0

do
findlevel RawMZ PeakMaxGuess(y)
wavestats/R = ( ( V_LevelX - RangeOfPeakCenterGuess ) , ( V_LevelX +
RangeOfPeakCenterGuess ) ) SmoothRawInt //This finds the peak top and center
PeakMax(y) = RawMZ(V_maxloc)
```

```
res(107) = V_max
y = y+1
while(y <= numpnts(Charges))

s=0

do
findlevel RawMZ PeakMax(s)
wavestats/R = ( ( V_LevelX - RangeOfPeakCenterGuess ) , ( V_LevelX +
RangeOfPeakCenterGuess ) ) SmoothRawInt //This finds the peak top and center
PeakCenter = V_LevelX
res(108) = PeakCenter
PeakInt(s) = SmoothRawInt(PeakCenter)
res(109) = PeakInt(s)
findlevel/Q/R = ( ( PeakCenter - 1 ) , ( PeakCenter - RangeOfPeakCenterGuess*3 ) )
SmoothRawInt (PeakInt(s)/2)
LeftSide = V_LevelX
findlevel/Q/R = ( ( PeakCenter + 1 ) , ( PeakCenter + RangeOfPeakCenterGuess*3 ) )
SmoothRawInt (PeakInt(s)/2)
RightSide = V_LevelX
Peakwidth(s) = RawMZ(RightSide) - RawMZ(LeftSide)
if (s == 0)
killwaves/Z CS0
Make/N=2/D/O CS0
killwaves/Z CS0r
Make/N=2/D/O CS0r
CS0 = PeakInt(s)/2
CS0r(0) = RawMZ(RightSide)
CS0r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS0 vs CS0r
ModifyGraph/W=RawData lsize(CS0)=2,rgb(CS0)=(0,0,0)
endif

if (s == 1)
killwaves/Z CS1
Make/N=2/D/O CS1
killwaves/Z CS1r
Make/N=2/D/O CS1r
CS1 = PeakInt(s)/2
CS1r(0) = RawMZ(RightSide)
CS1r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS1 vs CS1r
ModifyGraph/W=RawData lsize(CS1)=2,rgb(CS1)=(0,0,0)
endif
```

```
if (s == 2)
killwaves/Z CS2
Make/N=2/D/O CS2
killwaves/Z CS2r
Make/N=2/D/O CS2r
CS2 = PeakInt(s)/2
CS2r(0) = RawMZ(RightSide)
CS2r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS2 vs CS2r
ModifyGraph/W=RawData lsize(CS2)=2,rgb(CS2)=(0,0,0)
endif

if (s == 3)
killwaves/Z CS3
Make/N=2/D/O CS3
killwaves/Z CS3r
Make/N=2/D/O CS3r
CS3 = PeakInt(s)/2
CS3r(0) = RawMZ(RightSide)
CS3r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS3 vs CS3r
ModifyGraph/W=RawData lsize(CS3)=2,rgb(CS3)=(0,0,0)

endif
if (s == 4)
killwaves/Z CS4
Make/N=2/D/O CS4
killwaves/Z CS4r
Make/N=2/D/O CS4r
CS4 = PeakInt(s)/2
CS4r(0) = RawMZ(RightSide)
CS4r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS4 vs CS4r
ModifyGraph/W=RawData lsize(CS4)=2,rgb(CS4)=(0,0,0)
endif

if (s == 5)
killwaves/Z CS5
Make/N=2/D/O CS5
killwaves/Z CS5r
Make/N=2/D/O CS5r
CS5 = PeakInt(s)/2
CS5r(0) = RawMZ(RightSide)
```

```
CS5r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS5 vs CS5r
ModifyGraph/W=RawData lsize(CS5)=2,rgb(CS5)=(0,0,0)
endif

if (s == 6)
killwaves/Z CS6
Make/N=2/D/O CS6
killwaves/Z CS6r
Make/N=2/D/O CS6r
CS6 = PeakInt(s)/2
CS6r(0) = RawMZ(RightSide)
CS6r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS6 vs CS6r
ModifyGraph/W=RawData lsize(CS6)=2,rgb(CS6)=(0,0,0)
endif

if (s == 7)
killwaves/Z CS7
Make/N=2/D/O CS7
killwaves/Z CS7r
Make/N=2/D/O CS7r
CS7 = PeakInt(s)/2
CS7r(0) = RawMZ(RightSide)
CS7r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS7 vs CS7r
ModifyGraph/W=RawData lsize(CS7)=2,rgb(CS7)=(0,0,0)

endif
if (s == 8)
killwaves/Z CS8
Make/N=2/D/O CS8
killwaves/Z CS8r
Make/N=2/D/O CS8r
CS8 = PeakInt(s)/2
CS8r(0) = RawMZ(RightSide)
CS8r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS8 vs CS8r
ModifyGraph/W=RawData lsize(CS8)=2,rgb(CS8)=(0,0,0)
endif

if (s == 9)
killwaves/Z CS9
Make/N=2/D/O CS9
```

```
killwaves/Z CS9r
Make/N=2/D/O CS9r
CS9 = PeakInt(s)/2
CS9r(0) = RawMZ(RightSide)
CS9r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS9 vs CS9r
ModifyGraph/W=RawData lsize(CS9)=2,rgb(CS9)=(0,0,0)
endif

if (s == 10)
killwaves/Z CS10
Make/N=2/D/O CS10
killwaves/Z CS10r
Make/N=2/D/O CS10r
CS10 = PeakInt(s)/2
CS10r(0) = RawMZ(RightSide)
CS10r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS10 vs CS10r
ModifyGraph/W=RawData lsize(CS10)=2,rgb(CS10)=(0,0,0)
endif

if (s == 11)
killwaves/Z CS11
Make/N=2/D/O CS11
killwaves/Z CS11r
Make/N=2/D/O CS11r
CS11 = PeakInt(s)/2
CS11r(0) = RawMZ(RightSide)
CS11r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS11 vs CS11r
ModifyGraph/W=RawData lsize(CS11)=2,rgb(CS11)=(0,0,0)
endif

if (s == 12)
killwaves/Z CS12
Make/N=2/D/O CS12
killwaves/Z CS12r
Make/N=2/D/O CS12r
CS12 = PeakInt(s)/2
CS12r(0) = RawMZ(RightSide)
CS12r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS12 vs CS12r
ModifyGraph/W=RawData lsize(CS12)=2,rgb(CS12)=(0,0,0)
endif
```

```
if (s == 13)
killwaves/Z CS13
Make/N=2/D/O CS13
killwaves/Z CS13r
Make/N=2/D/O CS13r
CS13 = PeakInt(s)/2
CS13r(0) = RawMZ(RightSide)
CS13r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS13 vs CS13r
ModifyGraph/W=RawData lsize(CS13)=2,rgb(CS13)=(0,0,0)
endif

if (s == 14)
killwaves/Z CS14
Make/N=2/D/O CS14
killwaves/Z CS14r
Make/N=2/D/O CS14r
CS14 = PeakInt(s)/2
CS14r(0) = RawMZ(RightSide)
CS14r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS14 vs CS14r
ModifyGraph/W=RawData lsize(CS14)=2,rgb(CS14)=(0,0,0)
endif

if (s == 15)
killwaves/Z CS15
Make/N=2/D/O CS15
killwaves/Z CS15r
Make/N=2/D/O CS15r
CS15 = PeakInt(s)/2
CS15r(0) = RawMZ(RightSide)
CS15r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS15 vs CS15r
ModifyGraph/W=RawData lsize(CS15)=2,rgb(CS15)=(0,0,0)
endif

if (s == 16)
killwaves/Z CS16
Make/N=2/D/O CS16
killwaves/Z CS16r
Make/N=2/D/O CS16r
CS16 = PeakInt(s)/2
CS16r(0) = RawMZ(RightSide)
```

```
CS16r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS16 vs CS16r
ModifyGraph/W=RawData lsize(CS16)=2,rgb(CS16)=(0,0,0)
endif

if (s == 17)
killwaves/Z CS17
Make/N=2/D/O CS17
killwaves/Z CS17r
Make/N=2/D/O CS17r
CS17 = PeakInt(s)/2
CS17r(0) = RawMZ(RightSide)
CS17r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS17 vs CS17r
ModifyGraph/W=RawData lsize(CS17)=2,rgb(CS17)=(0,0,0)
endif

if (s == 18)
killwaves/Z CS18
Make/N=2/D/O CS18
killwaves/Z CS18r
Make/N=2/D/O CS18r
CS18 = PeakInt(s)/2
CS18r(0) = RawMZ(RightSide)
CS18r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS18 vs CS18r
ModifyGraph/W=RawData lsize(CS18)=2,rgb(CS18)=(0,0,0)
endif

if (s == 19)
killwaves/Z CS19
Make/N=2/D/O CS19
killwaves/Z CS19r
Make/N=2/D/O CS19r
CS19 = PeakInt(s)/2
CS19r(0) = RawMZ(RightSide)
CS19r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS19 vs CS19r
ModifyGraph/W=RawData lsize(CS19)=2,rgb(CS19)=(0,0,0)
endif

if (s == 20)
killwaves/Z CS20
Make/N=2/D/O CS20
```

```
killwaves/Z CS20r
Make/N=2/D/O CS20r
CS20 = PeakInt(s)/2
CS20r(0) = RawMZ(RightSide)
CS20r(1) = RawMZ(LeftSide)
AppendToGraph/W=RawData CS20 vs CS20r
ModifyGraph/W=RawData lsize(CS20)=2,rgb(CS20)=(0,0,0)
endif

s = s+1

while(s <= numpnts(Charges))

wavestats/R=(res(113),res(114)) RawInt
RawInt = RawInt + V_min

wavestats/R=(res(113),res(114)) RawInt
RawInt = RawInt / V_max

SetAxis/W=RawData left 0,1.1

Duplicate/O Charges CorrectPeakWidth

CorrectPeakWidth = (PeakWidth ) / PeakMax

Display/N=WidthVsIonNumber CorrectPeakWidth vs IonNumber
ModifyGraph/W=WidthVsIonNumber width=300,height=206.5
ModifyGraph/W=WidthVsIonNumber mode=3,marker=19,msize=3,rgb=(0,0,0)
ModifyGraph/W=WidthVsIonNumber highTrip(left)=1e+07
Label/W=WidthVsIonNumber left "\\f03CPW\\BZ "
Label/W=WidthVsIonNumber bottom "\f01Low\f03 Z\f01 <<< \f03Ion Number \f01>>>
High\f03 Z"
TileWindows/O=1/A=(3,3)

if (res(199) ==1)
Doalert 0,"Make sure that the peaks are smoothed enough to produce true peak tops, if
not redo this step."
Doalert 0,"Does (CPWz) decrease as the (Ion Number) increases?.\rYes, the spectrum is
non-ideal.\rNo, the spectrum is idea."
Doalert 0,"Once adequate smoothing has been performed select [Macro] >
[DetermineChargestate]."
endif

End macro
```

```
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Macro DetermineChargestate()

variable
f,g,CorrectMaxChargestate,q,MaxPercentMassIncrease,MinPercentMassIncrease,MaxPe
akWidth,MinPeakWidth
variable TrueComplexMassGuess

Duplicate/O Charges TestChargestate
Duplicate/O Charges TestMass
Duplicate/O Charges PercentMassIncrease
Duplicate/O Charges CorrectPeakWidth
Duplicate/O TestMass CorrectMass

CorrectPeakWidth = 0
TestChargestate = 0
CorrectMass = 0

Duplicate/O TestChargestate GuessChargestate1
redimension/N=21 GuessChargestate1
GuessChargestate1 = 0

Duplicate/O GuessChargestate1 Error1
Error1 = 0

Duplicate/O GuessChargestate1 Slope
Slope = 0

Duplicate/O GuessChargestate1 RSquared
RSquared = 0

Duplicate/O TestChargestate Zero
redimension/N=21 Zero
Zero = 0

Duplicate/O TestChargestate Correct
redimension/N=2 Correct
Correct = 0
```

```
Duplicate/O TestChargestate CorrectY
redimension/N=2 CorrectY
CorrectY = 0

Duplicate/O TestChargestate CorrectRaw
redimension/N=2 CorrectRaw
CorrectRaw = 0

Duplicate/O TestChargestate CorrectYRaw
redimension/N=2 CorrectYRaw
CorrectYRaw = 0

DoWindow/K IonSummary
DoWindow/K Summary0
DoWindow/K ChargestateIteration
DoWindow/K MassVsChargestate
DoWindow/K MassVsCorrectedIonPeakWidth

Tag/W=RawData/K/N=text0

Display/N=ChargestateIteration Error1 vs GuessChargestate1
ModifyGraph/W=ChargestateIteration mode=3,marker=19,msize=3
AppendToGraph/R/W=ChargestateIteration Slope vs GuessChargestate1
ModifyGraph/W=ChargestateIteration
mode=3,marker=19,msize=3,rgb(Slope)=(0,0,52224)
Label/W=ChargestateIteration bottom "\f03Z"
Label/W=ChargestateIteration left "\\f01\\K(65280,0,0)Stadard Deviation of
\\f03M\\BZ\\M \\f01& \\f03M\\BESI\\M \\f01(Conventional ESI Method)"
Label/W=ChargestateIteration right "\\f01\\K(0,0,52224)Slope of \\f03%MI\\BZ
\\M\\f00\\f01vs. \\f03CPW\\BZ \\M\\f01(First Positive Slope Method)"
ModifyGraph/W=ChargestateIteration width=270,height=206.5
ModifyGraph/W=ChargestateIteration
nticks(bottom)=21,sep(bottom)=1,highTrip(bottom)=1,lowTrip(bottom)=1
AppendToGraph/R/W=ChargestateIteration Zero vs GuessChargestate1
ModifyGraph/W=ChargestateIteration rgb(Zero)=(0,0,0)
AppendToGraph/R/W=ChargestateIteration CorrectY vs Correct

f = 0

TestChargestate = Charges - 11
GuessChargestate1(f) = res(89) - 11
CorrectPeakWidth = (PeakWidth ) / PeakMax

wavestats CorrectPeakWidth
```

```
MaxPeakWidth = V_max
MinPeakWidth = V_min

Do
TestMass = 0
TestChargestate = TestChargestate + 1
GuessChargestate1(f) = GuessChargestate1(f-1) + 1
TestMass = ( TestChargestate*PeakMax ) - TestChargestate
wavestats TestMass
Error1(f) = V_sdev
CurveFit line TestMass /X=CorrectPeakWidth /D
TrueComplexMassGuess = W_coef(0)
res(91) = TrueComplexMassGuess
PercentMassIncrease = ( ( TestMass - TrueComplexMassGuess ) / TestMass ) * 100
wavestats PercentMassIncrease
MaxPercentMassIncrease = V_max
MinPercentMassIncrease = V_min
CurveFit line PercentMassIncrease /X=CorrectPeakWidth /D
Slope(f) = W_coef(1)
RSquared(f) = (V_Pr)^2
f=f+1
while (f<21)

res(90) = 0
g=0
q=0

Do
if (Slope(g)>0)
CorrectMaxChargestate = GuessChargestate1(g)
res(90) = CorrectMaxChargestate
q=1
endif
g=g+1
while (q==0)

wavestats Slope
Correct(0) = CorrectMaxChargestate
CorrectY(0) = Slope(V_minloc)
Correct(1) = CorrectMaxChargestate
CorrectY(1) = Slope(V_maxloc)

Tag/W=ChargestateIteration/C/N=text0 CorrectY, 0,"charge state predicted from 1st
Positive slope method"
```

```
Tag/W=ChargestateIteration/C/N=text0/X=14/Y=26

wavestats/R =(res(113),res(114)) RawInt
CorrectRaw(0) = RawMZ(V_maxloc)
CorrectYRaw(0) = RawInt(V_maxloc)

RemoveFromGraph/W=RawData/Z CorrectYRaw
AppendToGraph/W=RawData CorrectYRaw vs CorrectRaw
String fitText1
sprintf fitText1, "\\f01\Z081st Posivitive Slope charge state = %g",
CorrectMaxChargestate
Tag/W=RawData/C/N=text0 CorrectYRaw, 0,fitText1
Tag/W=RawData/C/N=text0/X=20/Y=8
ModifyGraph/W=RawData
mode(CorrectYRaw)=2,rgb(CorrectYRaw)=(65535,65535,65535)

Charges = x - res(10)
Charges = Charges + CorrectMaxChargestate

wavestats Error1
res(97) = GuessChargestate1(V_minloc)
CorrectMass = ( PeakMax * Charges ) - Charges
TestMass = CorrectMass

CurveFit line TestMass /X=CorrectPeakWidth /D

res(98) = W_coef(0)//y intercept of M Vs PW
res(197) = (V_Pr)^2

Display/N=MassVsCorrectedIonPeakWidth TestMass vs CorrectPeakWidth
ModifyGraph/W=MassVsCorrectedIonPeakWidth width=250,height=206.5
ModifyGraph/W=MassVsCorrectedIonPeakWidth
mode=3,marker=19,msize=3,rgb=(0,0,0)
ModifyGraph/W=MassVsCorrectedIonPeakWidth highTrip(left)=1e+07

Label/W=MassVsCorrectedIonPeakWidth left "\\f03M\\BZ"
Label/W=MassVsCorrectedIonPeakWidth bottom "\\f03CPW\\BZ "

if (res(197)<0.5)
TileWindows/O=1/A=(2,3)
Doalert 1,"The 1st positive slope may result from niose.\r Is there a linear correlation
between Mz and CPWz?"

if (V_flag == 2)
```

```
Charges = Charges + 1
CorrectMaxChargestate = CorrectMaxChargestate + 1
CorrectMass = ( PeakMax * Charges ) - Charges
TestMass = CorrectMass

CurveFit line TestMass /X=CorrectPeakWidth /D
RemoveFromGraph/W=MassVsCorrectedIonPeakWidth/Z fit_TestMass

res(98) = W_coef(0)//y intercept of M Vs PW
res(197) = (V_Pr)^2

Tag/W=ChargestateIteration/K/N=text0

wavestats Slope
Correct(0) = CorrectMaxChargestate
CorrectY(0) = Slope(V_minloc)
Correct(1) = CorrectMaxChargestate
CorrectY(1) = Slope(V_maxloc)

Tag/W=ChargestateIteration/C/N=text0 CorrectY, 0,"charge state predicted from 1st
Positive slope method"
Tag/W=ChargestateIteration/C/N=text0/X=14/Y=26
Tag/W=RawData/K/N=text0

String fitText2
sprintf fitText2, "\\f01\Z081st Posivitive Slope charge state = %g",
CorrectMaxChargestate
Tag/W=RawData/C/N=text0 CorrectYRaw, 0,fitText2
Tag/W=RawData/C/N=text0/X=20/Y=8

if (res(197)<0.5)
Doalert 0,"The 1st positive slope may not be reliable"
endif

if (res(197)>0.5)
Doalert 0,"The 1st positive slope resulted from noise and not a linear correlation between
(Mz vs. CPWz).\rTherefore the second positive slope is the true charge state."
endif
endif
endif

wavestats TestMass
res(99) = round(V_max - V_min)//Delta of Complex masses
PercentMassIncrease = ( ( TestMass - W_coef(0) ) / TestMass ) * 100
```

```
CurveFit line PercentMassIncrease /X=CorrectPeakWidth /D
res(92) = W_coef(0)
res(93) = W_coef(1)
res(94) = (V_Pr)^2
res(95) = ((MaxPercentMassIncrease-MinPercentMassIncrease)/numpnts(Charges))
res(96) = ((MaxPeakWidth-MinPeakWidth)/numpnts(Charges))

Display/N=MassVsChargestate TestMass vs Charges
ModifyGraph/W=MassVsChargestate width=250,height=206.5
ModifyGraph/W=MassVsChargestate mode=3,marker=19,msize=3,rgb=(0,0,0)
ModifyGraph/W=MassVsChargestate highTrip(left)=1e+07
Label/W=MassVsChargestate left "\\f03M\\BZ"
Label/W=MassVsChargestate bottom "\\f03Z "

Make/N=15/T/O SummaryText
Make/N=15/D/O Summary

Edit/K=0/N=Summary 'SummaryText';DelayUpdate
AppendToTable/W=Summary0 'Summary';DelayUpdate
ModifyTable/W=Summary0 width(SummaryText)=320

SummaryText(0) = "Predicted Max Charge State for 1st Positive Slope Method"
Summary(0) = CorrectMaxChargestate
SummaryText(1) = "Predicted Max Charge State for Conventional ESI Method"

wavestats Error1
variable ChargestateESI
ChargestateESI = GuessChargestate1(V_minloc)
Summary(1) = GuessChargestate1(V_minloc)
SummaryText(2) = "calculated average mass (1st Positive Slope)"

wavestats TestMass
Summary(2) = V_avg
SummaryText(3) = "Standard Deviation of calculated average mass (1st Positive Slope)"
Summary(3) = V_sdev

SummaryText(7) = ""
SummaryText(8) = ""
SummaryText(9) = ""
SummaryText(10) = ""
SummaryText(11) = ""
SummaryText(12) = ""
SummaryText(13) = ""
```

```
SummaryText(14) = ""
Summary(7) = 0
Summary(8) = 0
Summary(9) = 0
Summary(10) = 0
Summary(11) = 0
Summary(12) = 0
Summary(13) = 0
Summary(14) = 0

Duplicate/O TestMass TestMassESI
Duplicate/O Charges ChargesESI

ChargesESI = Charges + (ChargestateESI-CorrectMaxChargestate)
TestMassESI = ( PeakMax * ChargesESI ) - ChargesESI
SummaryText(4) = "calculated average mass (Conventional ESI)"

wavestats TestMassESI
Summary(4) = V_avg
SummaryText(5) = "Standard Deviation of calculated average mass (Conventional ESI)"
Summary(5) = V_sdev
SummaryText(6) = "Rsquared of a linear fit to (Mz vs. CPWz)"
Summary(6) = res(197)

TileWindows/O=1/A=(2,3)

end macro

//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Macro SigmiodFit()

SummaryText(7) = ""
SummaryText(8) = ""
SummaryText(9) = ""
SummaryText(10) = ""
SummaryText(11) = ""
SummaryText(12) = ""
SummaryText(13) = ""
SummaryText(14) = ""
Summary(7) = 0
```

```
Summary(8) = 0
Summary(9) = 0
Summary(10) = 0
Summary(11) = 0
Summary(12) = 0
Summary(13) = 0
Summary(14) = 0

wavestats Charges
SetAxis/W=MassVsChargestate bottom (Charges(0) -3),(Charges(V_npnts-1)+3)
CurveFit/X=1 Sigmoid TestMass /X=Charges /D
AppendToGraph/W=MassVsChargestate fit_TestMass

RemoveFromGraph/W=MassVsCorrectedIonPeakWidth/Z fit_TestMass
SetAxis/A/W=MassVsCorrectedIonPeakWidth

SummaryText(7) = "Mass Complex (sigmoid fit)"
Summary(7) = W_coef(0)+W_coef(1)
SummaryText(8) = "standard deviation of Mass Complex (sigmoid fit)"
Summary(8) = W_sigma(0)+W_sigma(1)
SummaryText(9) = "Mass Complex Max Adducts (sigmoid fit)"
Summary(9) = W_coef(0)
SummaryText(10) = "standard deviation of Mass Complex Max Adducts (sigmoid fit)"
Summary(10) = W_sigma(0)
SummaryText(11) = "Cs (sigmoid fit)"
Summary(11) =W_coef(2)
SummaryText(12) = "standard deviation of Cs (sigmoid fit)"
Summary(12) = W_sigma(2)
SummaryText(13) = "Rate (sigmoid fit)"
Summary(13) =W_coef(3)^-1
SummaryText(14) = "standard deviation of Rate (sigmoid fit)"
Summary(14) = W_sigma(3)^-1

end macro

//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Macro ReportIonData()

DoWindow/K IonSummary

Edit/K=0/N=IonSummary 'IonNumber';DelayUpdate
```

```
AppendToTable PeakMax
AppendToTable PeakInt
AppendToTable Charges
AppendToTable CorrectMass
AppendToTable PeakWidth
AppendToTable CorrectPeakWidth
AppendToTable PercentMassIncrease

end macro

//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

Macro ViewSuggestions()

Doalert 1,"Do you want to see suggestion boxes during this analysis session?"
res(199) = V_flag
end macro

//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
//XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```