Georgia State University

ScholarWorks @ Georgia State University

Chemistry Theses

Department of Chemistry

8-4-2008

Analyses and Applications of Metalloprotein Complexes

Michael Patrick Kirberger
mkirberger@student.gsu.edu

Follow this and additional works at: https://scholarworks.gsu.edu/chemistry_theses

Recommended Citation

Kirberger, Michael Patrick, "Analyses and Applications of Metalloprotein Complexes." Thesis, Georgia State University, 2008.
doi: https://doi.org/10.57709/1059235

This Thesis is brought to you for free and open access by the Department of Chemistry at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Chemistry Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

ANALYSES AND APPLICATIONS OF METALLOPROTEIN COMPLEXES

by

MICHAEL PATRICK KIRBERGER

Under the Direction of Jenny J Yang

ABSTRACT

The structural characteristics associated with the binding of beneficial metals (i.e. - Mg^{2+}, Zn^{2+} and Ca^{2+}) to natural proteins has typically received more attention than competitive binding by toxic metals (e.g. – Pb^{2+}, Hg^{2+}, Cd^{2+}, La^{3+}, etc.). In this thesis, a statistical analysis of Pb^{2+}-binding in crystallized protein structures indicates that Pb^{2+} does not bind preferentially with nitrogen, as generally assumed, but binds predominantly with oxygen, and to a lesser degree, sulfur. A comparison of Ca^{2+} and Pb^{2+} indicates that Pb^{2+} binds with a wider range of coordination numbers, with less formal change, and with less defined structure than Ca^{2+}. The Pb^{2+} ion also appears to displace Ca^{2+} with little conformational stress in calcium binding proteins (CaBP's). Experimental data from the binding of metals with engineered fluorescent proteins indicate that both Pb^{2+} and Gd^{3+} will occupy grafted calcium-binding sites with greater affinity than Ca^{2+}, and strong evidence is presented to support the hypothesis that Pb^{2+} and Gd^{3+} will bind non-specifically on the protein surface. These results suggest that toxicity is associated with two binding mechanisms: displacement of the metal cofactor which disrupts protein function, and non-specific binding which maintains higher solubility of the metal.

INDEX WORDS: metalloprotein, metal, calcium binding protein, CaBP, enhanced green fluorescent protein, EGFP, Ca^{2+}, Pb^{2+}, Gd^{3+}, database, sequence analysis, prediction, binding geometry, metal toxicity.

ANALYSES AND APPLICATIONS OF METALLOPROTEIN COMPLEXES

by

MICHAEL PATRICK KIRBERGER

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

Copyright by

Michael Patrick Kirberger and Jenny Yang

2007

ANALYSES AND APPLICATIONS OF METALLOPROTEIN COMPLEXES

by

MICHAEL PATRICK KIRBERGER

Major Professor:	Jenny J. Yang
Committee:	Giovanni Gadda
	Shahab Shamsi

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
May 2007

ACKNOWLEDGEMENTS

I would like to express my appreciation for the many people who have either contributed to the work presented in this thesis, or have supported my efforts during the course of my time at Georgia State University. First and foremost I would like to thank my research advisor, Dr. Jenny J. Yang, who helped me find my own path, and encouraged me to always reach beyond my perceived limits. I would also like to thank Dr. Keith Pascoe who has always provided me with valuable advice.

Working with a cohesive group of talented people always raises the bar of excellence, and I have been fortunate to work with such a group in Dr. Yang's lab. In particular, I wish to thank Ning Chen who gave me an intensive crash course in protein expression and purification, and Dr. Jin Zou for his help in designing methods for analyses of metal binding. Thanks also to all of the other members in Dr.Yang's lab: Dr. Wei Yang, Dr. Shunyi Li, Hsiau-wei Lee, Yubin Zhou, Yun Huang, Jingjuan Qiao, Jie Jieng, Shen Tang, Hai Deng, Julian Johnson, Yanyi Chen, David Mpofu, and Adriana Castiblanco.

Finally, I wish to thank my wife Barbara, for enduring this long process patiently, and my children, Connor, Caoilinn, and Logan, for all of the missed baseball games, soccer matches, dance classes and karate lessons.

Funding for this work was provided by NIH Grants GM070555, GM 62999-1 and MCB-0092486.

Table of Contents

List of Tables

List of Figures

1 Introduction

The biological significance of metalloproteins has become a prominent area of research, as an increasing number of new proteins are discovered that require metal cofactors, such as Ca^{2+} and Zn^{2+}, to confer function. Calcium binding proteins (CaBP's) have been studied extensively due to the tremendous variety of calcium-dependent functions in living organisms, particularly with respect to cellular functions. Conversely, the binding of toxic metals and the related toxicology has not received the same level of attention.

The increase in available information on protein structures and sequence identification has led to the evolution of online databases. Using this wealth of information, numerous efforts have been made to identify and predict metal binding sites in proteins using computational methods. Successful prediction will enable the design of proteins and macromolecules with tailored functions for a wide range of applications, including metal-selective chelation, intracellular biosensors, and contrast agents. In this thesis, work is presented detailing efforts to develop new prediction methodologies and the development and application of localized database structures to evaluate metal binding characteristics statistically. Additionally, the binding characteristics of toxic metals, particularly Pb^{2+}, are evaluated with respect to several variants of engineered proteins. Understanding general characteristics of toxic metal binding, as well as how toxic metal ions compete with beneficial metals, will lead to an improved understanding of metal toxicity, and potentially, improved methods for treatment.

1.1 Biological Roles of Metal Ions

The biological roles of proteins are inextricably linked to their three-dimensional tertiary structures, which, in many cases, are achieved as a result of binding interactions with metal ions. In fact, nearly 40% of all proteins bind metals [1-6], and a recent search for "metal binding proteins" in PubMed (http://www.ncbi.nlm.nih.gov/) returned over 87, 000 entries.

For naturally-occurring metalloproteins, the binding of Mg^{2+}, Zn^{2+} and Ca^{2+} ions has been studied extensively with respect to binding geometry, structure, coordination number, and ligand preference [4, 7-19]. Studies of peptide fragments have revealed further data regarding the binding affinities of Ca^{2+}-binding structures [7]. Despite significant variations in the binding characteristics between different metal ions, the majority of metal-binding sites are characterized by a central shell of hydrophilic ligands to chelate the ion, with a surrounding shell of hydrophobic residues [8, 9].

Ionic calcium is demonstrably one of the more relevant metals associated with biological systems. In the human body, more than 90% of the extant calcium can be found in bones and dental enamel as hydroxyapatite, and $CaCO_3$ is prevalent in the biomineralization of shells and corals [10]. Calcium ions bind to various Ca^{2+}-binding proteins, including essential trigger proteins, which results in conformational changes, thus inducing function. Calmodulin (CaM) is a very important calcium-binding protein (CaBP) widely-distributed among eukaryotes that binds up to four Ca^{2+} ions (Figure 1.1). The resulting Ca^{2+}-CaM complex mediates a variety of biological processes related to inflammation, apoptosis, muscle contraction, intracellular movement, memory, nerve growth and the immune response, and may act as a sensor and signal transducer when it

binds with proteins that are unable to bind calcium themselves [11-13]. Calcium also participates in various functions related to the cell life cycle including both cell division and apoptosis. Its prevalence in both extracellular and intracellular fluids, where fluctuations of Ca^{2+} concentrations in the latter, which normally range from sub-micromolar to millimolar levels, allows it to act as a secondary, or intracellular, messenger [10, 14-16].

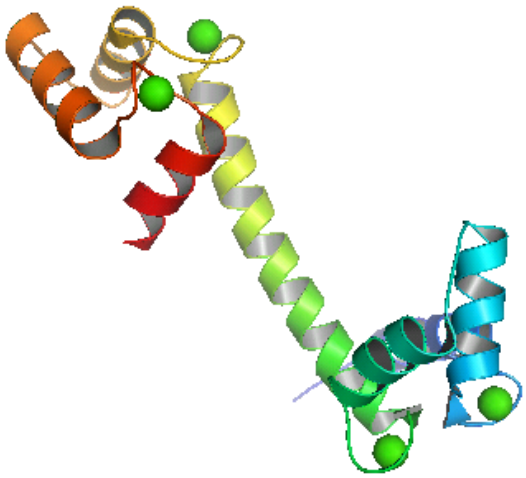


Figure 1.1: Calmodulin protein with four bound Ca^{2+} ions, from PDB File 1EXR.

Previous work by Yang *et al.* has evaluated the chemical and geometric characteristics associated with calcium and lanthanum binding in proteins [17-21], design and development of engineered CaBP's [22-24], analytical applications with engineered proteins[25, 26], prediction of calcium-binding sites from primary sequence data [27] and prediction of calcium binding sites in tertiary structures [28].

Several classes of EF-hand CaBP's capable of binding calcium, cadmium and lanthanides have already been developed by Yang *et al.*, and their relative binding

affinities reported [21, 22, 24]. These proteins were engineered using either a grafting approach, or a design approach using site-directed mutagenesis.

In the grafting approach [21], model CD2 protein variants have been designed by insertion of a 12-residue calcium-binding loop from calmodulin and additional glycine residues to link the loop to domain 1 of the CD2 variant. This grafting method has also been applied to green fluorescent protein (GFP) and enhanced green fluorescent protein (EGFP) variants. GFP and its numerous variants have been widely studied by protein chemists due to their unique chemiluminescent feature which allows them to be readily studied when used as biological tags or probes [29]. This fluorescence makes GFP an excellent and versatile tag for spectroscopic studies, particularly in metal-binding studies where the close proximity of a metal cation to a chromophore in the protein results in a detectable quenching of the fluorescent peaks during spectroscopic analysis [30]. It should be noted that GFP mutants have previously been designed for metal binding studies focusing on Cu^{2+}, Ni^{2+} and Zn^{2+} [30, 31]. The summation of these studies has provided a framework for the evaluation of toxic metal binding in proteins, as presented in this thesis. It was anticipated that the results of these studies would provide the basis for further research into the development of engineered proteins as bioremediation material with high selectivity to specific toxic metals, the development of metal-selective biosensors and elucidation of the structural role of protein-metal binding with respect to biological toxicity.

1.2 Metal Availability and Toxicity

While all metals may be considered toxic at sufficiently high concentrations, toxic metals are best defined as metals that are not chemically essential to living organisms, but may interact with biological systems, producing adverse results. These metals, which pose serious environmental and biological problems, include lead (Pb), cadmium (Cd), mercury (Hg), chromium (Cr), aluminum (Al) and the lanthanides, lanthanum (La), gadolinium (Gd), terbium (Tb) and europium (Eu).

Historically, lead (Pb^{2+}) has been, and remains, one of the most persistent toxic contaminants to enter the environment as a result of human industry and activity. Evidence of lead mining dates back to nearly 6500 BC, and lead poisoning from aqueduct pipes and cooking utensils during the Roman Empire was one of the earliest recorded instances of the toxicological implications of lead ingestion [32]. Today, EPA data indicate that lead is still one of the top 10 contaminants in waste sites identified as superfund targets [33], and the most recent survey presented by the CDC reflecting data collected from 1999-2000 indicated approximately 434, 000 children in the United States with Blood Lead Levels (BLL's) >=10μg/dL, primarily affecting children in lower socioeconomic environments where the risk to lead exposure is greater [34].

Sources of anthropogenic environmental lead include domestic wastewater, industrial effluents, leaded paint, surface runoff from mining operations, fossil fuel combustion, urban storm water runoff, and dissolution from materials in waste sites [35, 36]. Acid rain contributes to the environmental lead burden by reacting with mineralogical sources of lead and chemically releasing lead ions into soluble phase.

Lead exposure results from oral ingestion. Physiological effects of lead uptake include nerve response inhibition, CNS damage, damage to neuropsychological development, anemia, kidney damage, enzymatic interference with hemoglobin synthesis, increased blood pressure and male fertility decrease [37-43].

Lanthanides are often used as analogs to actinides in biochemical and environmental studies, due the dangers inherent in working with radioactive materials [44]. Lanthanides are introduced into the waste cycle from medical waste associated with imaging technologies. They are relatively abundant in natural sources, as well, and in the earth's crust, concentrations of lanthanum (La), cesium (Ce) and neodymium (Nd), exceed those of lead, so these rare elements are often more prevalent than more common metals [44].

Although there are natural sources of toxic metal release in the environment (e.g. – the chemical and physical weathering of rocks, decomposition of organic matter) [35], most sources are anthropogenic, including: surface runoff from mining operations; fossil fuel combustion; urban storm water runoff; leakage from landfills; leakage from agricultural activities (animal feces, fertilizers, pesticides); leakage from underground storage tanks for petroleum products and solvents; and septic system discharge [35, 45].

Eventually these metal ions enter groundwater aquifers, which provide 40% of the public water supply in the United States [45]. Once aquifers are contaminated with metal toxicants, remediation is difficult. Several studies have analyzed natural materials (e.g. - alfalfa, spruce sawdust) capable of adsorbing toxic metals [46, 47]. Diniz and Volesky reported the effective use of Sargassum biomass in the biosorption of La, Eu and Yb [48]. Similar results were found using red algae [49] and pretreated biomass of *Gracilaria*

fisheri algae [50]. In all cases, biosorption resulted primarily from the presence of carboxyl groups, and while this process is still poorly understood [51], the chelate effect exhibited by these materials is analogous to the manner in which metals are bound by proteins.

1.3 Classification of Metals

For any studies involving metals, it should be noted that metals have different properties, and react with different affinities to different ligands, so care should be taken when extrapolating generalized binding characteristics. At the most basic level, metal ions are electrophilic, Lewis acids, capable of coordinating several ligands simultaneously in acid/base reactions. These metals can be further subdivided as hard, soft or borderline acids, denoting different binding properties [4]. Table 1.1 summarizes various metal cations and the ligands they bind preferentially based on these classifications. Hard and soft acids have distinct chemical properties, as characterized in Table 1.2. Acids classified as borderline are more flexible, as they share characteristics from the other two.

Table 1.1: Hard/Soft Classifications for acids.

Classification	Lewis acid	Ligand
Hard	H^+, Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}, Mn^{2+}, Cr^{3+}, Co^{3+}	Mainly oxygen
Soft	Cu^+, Ag^+, Au^+, Pd^{2+}, Pt^{2+}, Cd^{2+}, Hg^+, Hg^{2+}	Mainly sulfur
Borderline	Zn^{2+}, Cu^{2+}, Ni^{2+}, Fe^{2+}, Fe^{3+}, Co^{2+}, Sn^{2+}, Pb^{2+}, Rh^{3+}, Ir^{3+}, Ru^{3+}	Mainly nitrogen, pyridine

Table 1.2: Characteristics of Hard and Soft acids

Hard Acids	Soft Acids
low polarizability (ion not easily deformed)	high polarizability (electron cloud more easily deformed)
High (+) charge	Small (+) charge
Small ionic radius	Larger ionic radius
Lack unshared valence e^- that can be easily excited	Several unshared valence e^- that can be easily distorted or removed
Ionic binding	Partial covalent bonds

The hard acid/base complex is formed when a metal with an empty *d*-orbital forms an electrostatic bond with a filled ligand *p*-orbital. There is generally a large difference in electronegativity between the metal-ligand ions (e.g., oxygen has an EN value of 3.5, coupled with calcium EN of 1.0). The soft acid/base complex is formed when a metal with a loosely-filled *d*-orbital forms a covalent bond with an empty ligand *p*, *d*, or π-orbital. There is generally a much smaller difference in electronegativity (e.g. - S: 2.5, Cd: 1.7) between the metal-ligand ions in this model [4]. Acids classified as borderline are more flexible, as they share characteristics from the other two classes.

1.4 Coordination Chemistry of CaBP's

The results of previous studies have provided several key characteristics associated with Ca^{2+}-binding. Calcium ions form ionic bonds predominantly with oxygen ligands from the side chains of negatively charged residues such as Asp and Glu, non-charged residues like Asn, main chain carbonyl oxygens and water. The Ca-O bond length has been previously reported to range from 2.01 Å to 3.15 Å, with a mean value of

2.4 ± 0.2 Å [4, 12, 52-54]. Observed structures frequently exhibit distorted pentagonal-bipyramidal geometries (Figure 1.2), with three monodentate ligands and one bidentate ligand providing the five planar oxygen ligands.

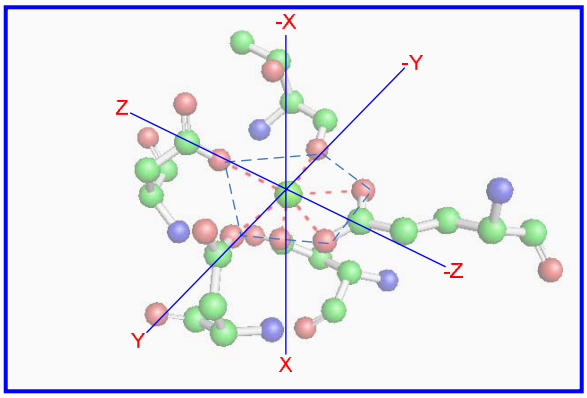


Figure 1.2: Pentagonal-bipyramid geometry surrounding Ca^{2+} ion and ligand-ion contact geometry for calmodulin EF-I binding pocket from PDB 3CLN.

The Ca^{2+} ion resides approximately 0.8 Å to 1.2 Å outside of the pentagonal plane, and less than 0.4 Å outside of the bidentate plane, with an angle relative to carbon in the bidentate plane (φ) < 30° [4, 55]. The inferior vertex (X) is occupied by a side-chain oxygen, and the superior vertex (-X) is frequently provided by a water molecule, not shown in Figure 1.2. A coordination number of 6 to 8 is most commonly observed, although coordination numbers ranging from 4 to 12 have been reported [4]. Additional structural stability is achieved through hydrogen bonding of the non-ligand Glu and Asp oxygens. The ionic radius of the calcium ion increases with increasing coordination number, but has been generally reported between 0.99Å to 1.12 Å for a coordination number of 6 to 8 [10]. As might be expected, the presence of negative charge in the binding site is relevant. Several studies have evaluated the charge environment for Ca^{2+}-binding sites, indicating higher binding affinities with a net negative charge within 5 to

15 Å of the binding microenvironment [56, 57], and the presence of 3 to 4 negative charges in the primary coordination shell may provide optimal charge configuration [7, 24].

The most common motif associated with Ca^{2+}-binding, the EF-Hand motif, was first described by Kretsinger in 1973 [58], and is observed in over 50% of all Ca^{2+}-binding proteins. This motif is characterized by a helix-loop-helix structure comprised of approximately 30 amino acids, and can be sub-divided into canonical EF-Hand (e.g. – calmodulin) and pseudo EF-Hands found in the S100 protein N-termini [27]. A thorough summary of the classification and evolution of EF-Hand proteins was presented by Kawasaki, Nakayama and Kretsinger [59].

For canonical EF-Hands, calcium ions bind in a twelve residue central loop, typically utilizing side-chain oxygen ligands from loop positions 1, 3, 5, 9, and 12, as well as a main-chain carbonyl oxygen from position 7. Ligands associated with EF-Hands are typically Asp at position 1, Asp or Asn at position 3, Asp, Ser or Asn at position 5, a water molecule at position 9, and a bidentate Glu at position 12 [7].

Pseudo EF-Hands coordinate the Ca^{2+} ion predominantly with main-chain carbonyl oxygen atoms in a 14 residue loop. Participating ligands may be contiguous or non-contiguous in the sequence.

Because the majority of ligands in contiguous EF-Hand and Pseudo EF-Hand proteins are conserved, it should be possible to develop computational methods to predict these structural motifs from the sequence of amino acids comprising the protein. It will be shown in this thesis that new patterns have been developed by Yang *et al.* that result in

better identification of both EF-Hand and Pseudo EF-Hand proteins than those presented in the Prosite online database.

1.5 Coordination Properties of Toxic Metal Ion/Protein Complexes

For Pb^{2+}, available references for proteins indicate average coordination numbers variously reported as 4 to 6 or 6 to 8 binding ligands [60, 61]. Even less clearly defined than the coordination number is the structural geometry associated with complex formation. Both Ca^{2+} and Zn^{2+} form complexes with fairly regular geometry: pentagonal-bipyramidal for Ca^{2+} and tetra- or octahedral for Zn^{2+}. If, as previously suggested, calcium- and zinc-binding proteins are the primary molecular targets for binding Pb^{2+}, it is not unreasonable to expect structural similarities upon binding with Pb^{2+}.

Divalent lead, however, appears to bind with irregular geometries, and, as will be reported, variable coordination numbers. A study by Shimoni-Livny *et al.* [62] observed two general geometries for lead binding: a holodirected geometry, where the ligands occupy a spherical globe around the bound ion; and a hemidirected geometry where the coordination ligands occupy only half of a sphere. For lead binding with low (2 to 5) coordination ligands, the hemidirected geometry is preferred, whereas for high (9 to 10) coordination ligands, the holodirected geometry is preferred. In the intermediate range of 6 to 8 ligands, which is more consistent with typical calcium binding coordination, either of the two geometries may be observed.

The Pb^{2+} ion is classified as a borderline acid with properties of both hard and soft acids. Pb^{2+} has an electron configuration identical to that of elemental Hg ($[Xe].4f^{14}.5d^{10}.6s^2$), resulting in unfilled 6*p* and 6*d* orbitals, similar to the hard acid model

(e.g. – Ca^{2+}), but with a pair of valence electrons in the 6*s* orbital. However, Pb^{2+} has a higher EN value than Ca^{2+} (1.8 vs. 1.0) which is closer to the soft acid Cd^{2+} EN value of 1.7. The combination of EN and valence places Pb^{2+} between the hard and soft acid models, and although Pb^{2+} has a putative ligand binding preference for nitrogen, binding with both oxygen (hard base) and sulfur (soft base) will be observed, as well.

Of the lanthanide ions, data from the Protein Data Bank suggest that Pr^{3+}, Tb^{3+}, and Yb^{3+} bind in small molecules with average coordination numbers of 8.3, 8.0 and 6.9 ligands, respectively, where the number of coordination binding ligands decreases relative to the size of the ion [63]. Various studies have reported the high affinity for Tb^{3+} and La^{3+} at calcium binding sites in both designed and natural proteins [24, 63, 64].

To evaluate the characteristics associated with Pb^{2+}-binding, data are presented in this work derived from a comprehensive statistical analysis of all crystal structures in the PDB that have been shown to bind Pb^{2+}.

1.6 Current Knowledge Regarding Metal Toxicity

Toxicity is dependent upon the concentration of free metal ions in the environment, which is influenced by various factors including metal solubility, salt content, redox potentials, and pH. In very acidic solutions, a high concentration of H^+ competes directly with metal cations for binding, and consequently the concentration of free metal cations tend to remain higher. The biological impact related to toxic metals directly correlates with the ions free concentration under aqueous conditions in the environment. Also, metal contaminants, unlike organic compounds, cannot be degraded over time into nontoxic constituent elements [65, 66].

Different studies have explored the mechanism of toxicity for Pb^{2+}, Cd^{2+}, and Hg^{2+} binding with metalloproteins, and the resulting conformational changes [67]. Cadmium will replace Zn^{2+}, Cu^{2+} and Ca^{2+} in proteins, enzymes and nucleic acids [68]. Cadmium is also known to interfere with calcium metabolism in the kidneys, and with calcification, decalcification and bone remodeling [40]. Lanthanides are disruptive in a similar fashion. La^{3+} and Gd^{3+} are known to block calcium channels in human and animal cells. Calcineurin is inhibited by Eu^{3+} and Tb^{3+}, while Ca^{2+}-ATPase is blocked by Dy^{3+} and La^{3+} [69].

Pb^{2+} ions also displace Zn^{2+} ions in zinc-binding sites on δ-aminolevulinic acid dehydratase (ALAD), an enzyme found in red blood cells [70-72]. More recently, several studies have indicated that a thiol-rich binding site will bind Pb^{2+} in a distorted trigonal-pyramidal geometry with 3 ligands, or alternatively with a geometry including 5-8 ligands [73-77].

The biological distribution of Pb^{2+} is closely associated with calcium metabolism, and it has been demonstrated that Pb^{2+} will compete with Ca^{2+} and Zn^{2+} in proteins [60, 74, 75, 78, 79]. More strikingly, Pb^{2+} has been shown to directly compete with Ca^{2+} ions for binding with various essential calcium-binding proteins such as protein kinase C [80] synaptotagmin I [81], and CaM [82]. CaM is an essential ubiquitous trigger protein that has been studied extensively, and a comprehensive review of the binding characteristics of CaM was recently summarized by Yang *et al.* [17]. The well-documented characteristics of CaM make it an excellent model for evaluating changes associated with the binding of a competing ion.

Although the toxicological effects of lead poisoning have been extensively studied and documented, the mechanism of toxicity relative to interaction with proteins is remains to be clearly elucidated.

1.7 Our Hypothesis

It is generally assumed that competitive binding occurs due to the similarities in atomic radii and oxidative states of the competing ions, and that competing anions tend to adopt the same binding geometry as the displaced metal. In this thesis, work is presented related to complex formation between different toxic metals and several engineered proteins to explore structural and chemical characteristics associated with toxic metal binding, and to quantitatively analyze the binding affinities of these metalloprotein complexes. From this, we attempt to further elucidate the mechanisms of metal toxicity.

We hypothesize that Pb^{2+} and other toxic metals compete with Ca^{2+} ions in CaBP's with different coordination properties that result in alterations of the conformational and functional properties of the proteins. Further, we suggest that binding may be accomplished both within a known binding site, and, with increasing concentration of the toxic metal, non-specifically on the protein surface.

Because metalloproteins have evolved to selectively bind beneficial metals, non-specific binding of these metals is not typically observed. In contrast, data will be presented in this thesis indicating that non-specific binding of toxic metals occurs readily, and the combination of these two binding mechanisms may play complementary roles in the mechanism of toxicity.

1.8 Research Objectives

The research presented in this thesis is divided into two broad categories: a) computational/statistical methods focusing on design and prediction of protein structures, and the application of localized databases for statistical analyses, and b) applications of engineered metalloproteins, particularly with respect to analyses of metal-binding characteristics. Both areas then converge in a central theme, which is an improved understanding of the nature of metalloprotein complexes, the reaction of various engineered proteins to toxic metals, and ultimately, the physiological consequences of binding with competing toxic metal ions.

In Chapter 2, two methods are explored that relate to improved prediction and design techniques, utilizing newly-developed software to facilitate accelerated research and analysis. A number of studies have reported attempts to predict Ca^{2+}-binding sites using computational methods [8, 9, 22, 28, 83-87]. The capability of predicting or designing protein structure has far-reaching implications, in that it would allow for protein design with predetermined functions.

In Chapter 3, a new statistical analysis of four datasets from previous studies is presented, utilizing different statistical methods, to establish more precise data regarding binding characteristics associated with CaBP's. These data provide a basis for comparison with the binding of Pb^{2+}.

In Chapter 4 the development of a local database is presented, which includes a developed user interface, to systematically evaluate metal-binding characteristics. This database is constructed from data available through the Protein Data Bank (PDB), and is applied in this work towards the statistical and structural evaluation of Pb^{2+}-binding

characteristics by summary and analysis of all reported crystallization data. Based on the premise that the type of metal affects the binding coordination as a condition of ligand preference, research was conducted to establish structural information on the binding of toxic metals, and to understand the determinants associated with selectivity and affinity of specific proteins to bind these metals. Analysis of this database resulted in preliminary results indicating conformational changes associated with the replacement of Ca^{2+} by Pb^{2+}.

In Chapter 5, the binding characteristics of Pb^{2+} are evaluated, both as an isolated species and in direct competition with Ca^{2+}, with different EGFP protein variants. Based on the previously reported variable coordination numbers associated with Pb^{2+} -binding, the apparent irregular complex geometry, and the borderline acidic nature of Pb^{2+}, the purpose of this line of research was to compare the empirical and structural data, and assess the validity of current assumptions regarding the binding characteristics of Pb^{2+}. In Chapter 6, conclusions of this thesis are summarized, and the implications of these studies with respect to toxicity are discussed.

The engineered CaBP's designed in our laboratory have allowed us to study conformational changes associated with metal binding, geometric and chemical features of the binding sites, and relative metal selectivity. Spectroscopic and fluorescence studies are included in this research to qualitatively assess conformational changes associated with binding of the targeted toxic metals, and from this, quantitatively assess the binding affinities of different toxic metals for a calcium binding site.

Comparative studies of different metals using EGFP variants with grafted binding loops and structural NMR studies are expected to provide future elucidation of ligand

preference resulting in predictable conformational changes, and should provide the basis for further research into the development of engineered proteins as bioremediation material with high selectivity to specific toxic metals, the development of metal-selective biosensors, and the structural role of protein/metal binding in biological toxicity.

Finally, preliminary work is discussed in Appendix B with respect to use of different EGFP variants as sensors for trypsin activity. The engineered EGFP variants produced by Yang *et al.* have been shown to react with trypsin (results not yet published), and the inherent spectroscopic capabilities of the variants suggests excellent potential for use as a sensor capable of monitoring *in vivo* trypsin activity. Consequently, the associated kinetics associated with trypsin cleavage of our variants, along with identification of the specific cleavage sites, are currently being evaluated.

2 Computational Studies

2.1 Introduction

The development of engineered proteins with pre-determined functions can be facilitated through computational methods designed to evaluate potential binding sites. Software has evolved to enhance the work of protein design, but all software requires parameter input from a researcher, so improved precision related to parametric data will result in improved design. In the case of calcium-binding proteins, the ability to more accurately predict structures from sequential data results in more narrowly-defined characteristics that serve as design input. In this chapter, two software strategies are presented to enhance the design of metal-binding sites, and the prediction of calcium-binding motifs.

2.2 PROT.E.U.S – Software for Secondary Analysis of Dezymer Data

Previously, work was completed with collaborators utilizing the Dezymer algorithm to predict potential calcium-binding sites in proteins. [86, 88]. Dezymer is a UNIX platform C++ algorithm designed to generate potential metal binding sites in a protein. The algorithm first generates a preliminary list of potential design sites based on variable parameters for: ligand geometry surrounding a metal ion; three-dimensional protein structure; and libraries of side chain rotamers of amino acids [86]. The subsequent refinement step then evaluates deviation from ideal pseudo-energy - U(P), where U(P) represents the difference between constructed sites and target sites based on bond lengths, bond angles, and sidechain clashes. All natural calcium-binding sites demonstrate a

calculated value for U(P) < 20. This value decreases depending on how closely a site conforms to ideal pentagonal-bipyramidal geometry. Combinations with higher projected U(P) (i.e.- farther from ideal geometry) can therefore be considered as unfavorable potential sites.

Both steps within the Dezymer execution incorporate variable input parameters limited predominately to molecular geometry. Corroborative research was completed in our laboratory by inserting known calcium-binding site parameters and 'creating' natural sites with expected geometries, which has validated the fundamental accuracy of the algorithm within those parameters [22, 24]. However, the text-based output data file generated during the process was frequently so large that visual analysis of the data became problematic due to the time involved and the potential for random error. For example, early executions of Dezymer for CD2 and GFP Chain A generated over 200 and 9300 potential sites, respectively.

2.2.1 Materials and Methods

In an effort to enhance the utility of Dezymer, additional refinement criteria were required to qualitatively rate the projected sites and list them in a logical order for evaluation. This additional process would eliminate those projected configurations with less favorable environmental factors from consideration, narrow the scope of potential combinations for experimental analysis to ensure a higher success yield, and accelerate the engineering design phase. To accomplish this, PROTein Engineering Utility Software (PROT.E.U.S) was developed as a secondary layer of software to perform additional refinement of the original Dezymer output data. The PROT.E.U.S software was

developed in Visual Basic 5.0 (Microsoft, Redmond, WA) to read a Dezymer output file, selectively filter potential binding sites based on criteria external to Dezymer that were stored in a local Access database (Microsoft, Redmond, WA), and provide a GUI interface whereby the filtering parameters could be changed quickly for multiple analyses.

PROT.E.U.S incorporated the following additional variable parameters conducive to metal-binding that account for environmental factors beyond the limitations of geometric configurations:

- Comparison of electrostatic potentials between native and designed sites;
- Identification of sites requiring fewer ligand mutations to ensure a native fold in the host protein;
- Identification of residues shown experimentally to be mutable;
- Good solvent accessibility (i.e. - ligands exposed for calcium ion binding).

2.2.1.1 Comparison of electrostatic potentials

Electrostatic interactions at the coordination shell of the metal binding sites and the surrounding environments have been shown to contribute to metal binding affinity and selectivity. Surveys of more than 100 calcium binding sites in EF-hand proteins have suggested that 3-4 negatively charged residues have the highest calcium binding affinity for EF-hand proteins [89]. In addition, Linse *et al.* have shown that removal of charged residues apart from the metal binding site results in a 45-fold decrease of the calcium affinity of calbindin D_{9K} [90]. Based on these observations, it was decided that the net

change of the charge numbers due to the introduction of the calcium binding sites would be incorporated into the PROT.E.U.S analysis.

2.2.1.2 Requirements for least mutations

Of primary concern in the secondary refinement analysis was the identification of those combinations requiring the least mutations, thus decreasing the probability of unfolding the protein while making structural alterations as suggested by the Ca^{2+} binding site configurations. Previous work from our laboratory and others strongly suggested that fewer mutations increase the probability of natural folding in the host protein.

The data analyzed via PROT.E.U.S for CD2 and GFP compared residues suggested by Site-Search against residues in the wild-type for a pentagonal-bipyramidal geometry. For each proposed configuration, residues that differed from the wild type were highlighted for identification, and the mutations required by site totaled and displayed in the designated grid column.

2.2.1.3 Good mutations

Previous work by Davis and Davies has identified CD2 ligands where mutations to acidic residues either allowed, or did not interfere with, binding by CD2, as well as non-conservative substitutions that interfered with, or had no effect on binding [91]. These data were incorporated into a Microsoft Access database (Microsoft Corporation, Redmond, WA). For analysis of CD2 data, PROT.E.U.S compared each instance where a mutation was suggested by the proposed configuration, and identified those sites with an asterisk where the suggested mutation residue was identical to those successful mutations identified in the Davis table (Table 2.1).

Table 2.1: Davis Table of Valid CD2 Mutations.

Residue Position	[a]Valid Mutation
12	H
17	N
26	D
34	R
36	S
37	T
45	K
46	M
47	K
51	K
52	S
56	E
70	R
79	T
83	T
86	T
90	N
94	D

[a]Indicates mutation was previously found to have little or no impact on binding.

2.2.1.4 Solvent accessibility

Negatively charged residues such as Asp or Glu are good calcium ligand residues and for a folded protein it is essential to place these charged residues on the protein surface. In addition, the majority of calcium binding sites have one or more oxygen ligand atoms from solvent H_2O. Therefore, it is crucial to ensure that the designed calcium binding sites have good solvent accessibility. Structural solvent accessibility of the scaffold protein was calculated using Lee and Richard's approach [92]. Residue locations with desired relative solvent accessibility were reported and highlighted by PROT.E.U.S.

2.2.1.5 Additional refinements

Although the PROT.E.U.S software fulfilled its primary function as a secondary-level analytical tool, limitations of the application suggested refinements that would be required to mold the existing program into a comprehensive tool capable of generating a refined list of favorable potential metal-binding sites.

First, the analysis had to be further refined by adding the capability to eliminate certain combinations based on either the presence of a specific ligand or ligands, or because one or more mutations were proposed at positions found experimentally to be unacceptable.

The original prototype was setup to perform analyses only on a pentagonal-bipyramidal geometry for binding, which is common for Ca^{2+} binding in our current research. The need to accommodate different geometric configurations had to be addressed by adding the ability to perform analysis against variable numbers of ligands. For our purposes, this range was defined from 3-8 ligands.

In addition to required changes that were specific to the analysis, functional changes were required to prepare the application for proposed future changes. A number of internal structural changes had to be made to accommodate code that would be used to generate potential binding sites. A private profile file was setup to contain data that would otherwise have to be hard-coded into the program, such as the directory path to the database tables. This file could be called by a Windows API at run-time to retrieve this data. By storing this information externally, the application would not have to be recompiled every time the data changed, since it would be dynamically linked only during execution.

The final version of the PROT.E.U.S software was modified with the following specific changes:

- Added capability to remove sites from analysis based on specific residues or positions.
- Added capability to perform analysis against different geometric configurations based on variable numbers of ligands (range = 3-8).
- Added flag to identify those proteins with available solvent accessibility data and include these data during analysis.
- Created new data structures to store original analysis parameters for subsequent trials.
- Replaced print preview function with a file export capability. This also required changes to visual layout of data not dependent on font-color since the output will go to a text file.
- Added API call to reference private profile file where default values stored.

2.2.2 Results and Discussion

Figure 2.1 summarizes the results of the analysis, with respect to required mutations, for CD2 and GFP, respectively. In each case, the majority of proposed configurations required mutations of all planar residues for binding in an acceptable geometric configuration. However, in both cases sites were identified that included residues found in the wild-type protein, reducing the number of required mutations. The 10 best Ca^{2+}-binding sites identified by the PROT.E.U.S software for CD2 and GFP Chain A are listed in Table 2.2 and Table 2.3, respectively. Column headings for Table

2.2 and Table 2.3 are defined as follows: **Row** is the row number from the original input file, **Res1**-**Res5** indicate the residues forming the binding site, **U(P) final** indicates final value for pseudo energy (i.e. - the difference between constructed sites and target sites based on bond lengths, bond angles, and sidechain clashes), **Reqd Mut** indicates the number of mutations required in the residues forming the binding site, and **ΔCharge** indicates the difference in charge between the original residues and the proposed mutations.

Table 2.2: 10 best Ca^{2+}-binding sites identified by PROT.E.U.S software for CD2.

Row	Res1	Res2	Res 3	Res4	Res5	U(P) final	Reqd Mut	ΔCharge
177	E62	N60	D54	N65	*N56	17.58	4	0
181	E62	N60	N54	N65	*N56	17.92	4	1
184	E62	N60	D54	D65	*N56	20.21	4	-1
180	E62	N60	N54	N65	*D56	23.25	4	0
175	E62	N60	D54	N65	*D56	23.59	4	-1
148	E62	*N56	D54	D65	N60	26	4	-1
90	E62	*D56	D54	D65	N60	26.37	4	-2
146	E62	*N56	D54	N65	N60	26.43	4	0
158	E62	*N56	N54	N65	N60	26.43	4	1
152	E62	*N56	N54	D65	N60	26.44	4	0

Table 2.3: 10 best Ca^{2+}-binding sites identified by PROT.E.U.S software for GFP chain A.

Row	Res1	Res2	Res 3	Res4	Res5	U(P) final	Reqd Mut	ΔCharge
4689	D86	D82	E194	K79	E5	55.63	2	-2
4693	D86	N194	D82	K79	E5	80.91	2	-1
4688	D86	D82	D194	K79	E5	82.69	2	-2
312	D16	D14	E44	F64	N121	17.91	3	-3
316	D16	N14	E44	F64	N121	18.25	3	-2
6297	D110	S123	E18	T63	T108	19.19	3	-2
5089	D96	T108	E106	T63	E98	21.17	3	-4
1736	D45	D210	E213	T217	E219	23.96	3	-3
4583	D82	D86	E5	E79	E194	24.36	3	-4
317	D16	N121	E14	E44	F64	25.82	3	-3

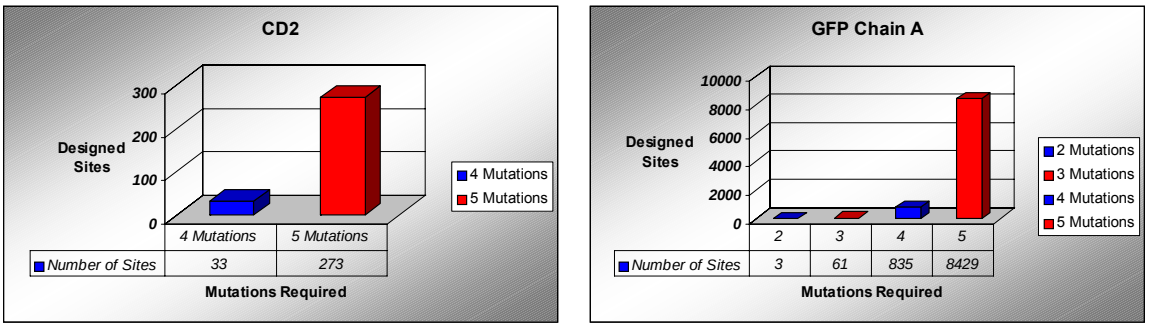


Figure 2.1: Summary of CD2 sites (left) and GFP sites (right) identified by Dezymer as potential calcium-binding sites and number of required mutations.

2.2.3 Conclusions

Preliminary studies with PROT.E.U.S demonstrated accurate function and a significant reduction in analysis time reviewing potential binding sites. To date, it has been utilized successfully for analyses related to the design of a calcium-binding CD2 cell-adhesion variant by Yang *et al*. [23].

2.3 FASTA Sequence Analysis Algorithm

A Java program was developed capable of testing FASTA sequences to identify EF-Hand and Pseudo EF-Hand binding motifs. These motifs, described in Section 1.4, are highly-conserved across genera, and represent more than half of all Ca^{2+}-binding proteins. The ability to successfully predict these motifs from the primary sequence is an initial step towards prediction of function. It also leads to better characterization of the key properties associated with binding, which will allow for improved engineering of

proteins with metal selectivity. Additionally, these data may be extrapolated to predict effects of metal toxicity resulting from displacement of beneficial ions by toxic metals.

2.3.1 Materials and Methods

Proprietary software for the sequence analysis of FASTA records was developed using Sun's Java Developer Kit and the Java Runtime Environment (http://java.sun.com/j2se/1.3/download.html). FASTA records used as input data for the sequence analysis software were obtained from the Prosite Database (http://au.expasy.org/prosite/). The FASTA sequences were manually downloaded into text files for processing.

This program was originally run with six CaBP patterns previously determined in our laboratory by conducting Multiple Sequence Alignment (MSA) analyses of several online databases [27]. These six patterns were divided into EF-Hand and Pseudo EF-Hand patterns. As seen in Tables 2.4 and 2.5, the three patterns for each type of CaBP are comprised of distinct patterns for each segment of the Helix-Loop-Helix secondary structures.

In general terms, the original program was designed to search a FASTA record first for the binding loop, and if this was found, then the program would search for each of the flanking helices. Data was written out to a set of output files, based on whether a protein was determined to be an EF Hand Protein, Pseudo EF Hand Protein, or neither.

Table 2.4: Comparison of Prosite and Yang Patterns 1 and 2 for EF-Hand.

Descriptive ID	Sequence Pattern
Prosite	
PS00018: EF-Hand	D-X-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-X(2)-[DE]-[LIVMFYW]
Yang (Pattern 1)	
EFH Helix E	X-{DNQ}-X-X-{GP}-{ENSPQ}-X-X-{DQRP}
EFH Loop	[DNS]-X-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-X(2)-[ED]
EFH Helix F	[FLMYVIW]-X-X-{NPS}-{DNEQ}-X(3)
Yang (Pattern 2)	
YY00018	X(1)-{DNQ}-X(2)-{GP}-{ENSPQ}-X(2)-{DQRP}-[DNS]-X(1)-[DNS]-{ILVFYW}-[DENSTG]-[DNQGHRK]-{GP}-[LIVMC]-[DENQSTAGC]-X(2)-[ED]-[FLMYVIW]-X(2)-{NPS}-{DNEQ}-X(3)

Table 2.5: Comparison of Prosite and Yang Patterns 1 and 2 for Pseudo EF-Hand.

Descriptive ID	Sequence Pattern
Prosite	
PS00303: S100 CaBP	[LIVMFYW](2)-x(2)-[LK]-D-x(3)-[DN]-x(3)-[DNSG]-[FY]-x-[ES]-[FYVC]-x(2)-[LIVMFS]-[LIVMF]
Yang (Pattern 1)	
PEF Helix E	X-[LIMV]-[IVLM]-X-X-[FYI]-[HQYFS]-X-[FYH]
PEF Loop	[SAVT]-X-X-X-[GDS]-[DNHS]-X-X-X-[LIVM]-X-[KRVQ]-X-[EDS]
PEF Helix F	[LFVM]-[KRL]-X-[LMF]-[LMVIF]-X-X-[EQNDS]
Yang (Pattern 2)	
YY00303	[LMVITNF]-[FY]-X(2)-[YHIVF]-[SAITV]-X(5,9)-[LIMV]-X(3)-[EDS]-[LFM]-[KRQL]-X(20,28)-[LQKF]-[DNG]-X(1)-[DNSC]-X(1)-[DKN]-X(4)-[FY]-X(1)-[EKS]

The source code for the program was comprised of four Java classes: CaSeqAnalysis.class, CaPattern.class, MetalBP.class and FASTARecord.class. The user interface is text-based from a DOS prompt. The class **CaSeqAnalysis** controls the application using input parameters from a user to specify the input data set and establish a search priority (EF-Hand versus Pseudo EF-Hand). This class also performs all IO routines and instantiates the supporting class objects. The **CaPattern** class creates

calcium-binding pattern objects that build string arrays of AA residues which correspond to a relative position within a sequence. For each relative position within a sequence, there may be one or more possible amino acids that need to be included or excluded from that position. The **FASTARecord** class converts each FASTA record into an object for sequence analysis. The **MetalBP** class creates metal-binding protein objects. The sequence analysis is actually performed within this class. This object was designed so that it can be extended for use with other metals, in addition to calcium.

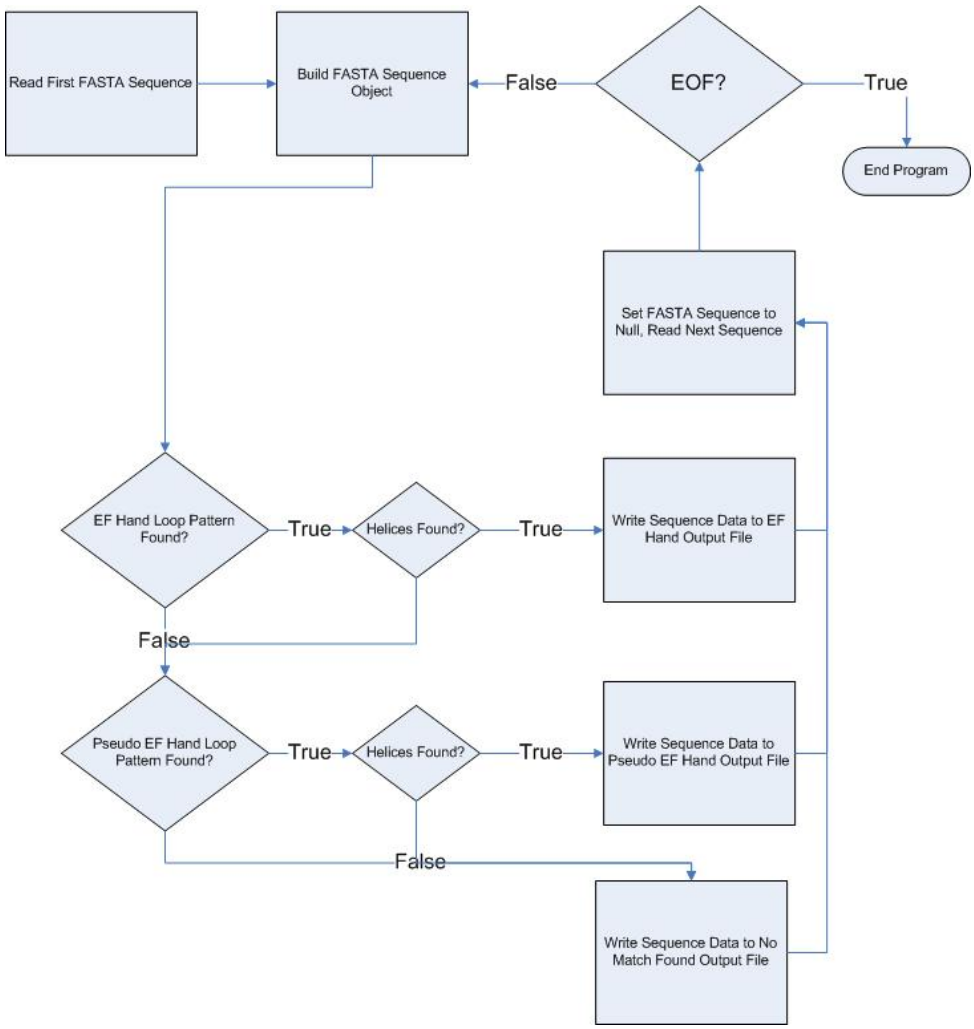


Figure 2.2: Logic flow for CaSeqAnalysis EF Hand search.

Figure 2.2 represents logic flow during program execution. In this example, EF-Hand is selected as a higher priority than Pseudo EF-Hand. Based on this parameter, the program first attempts to identify whether the protein contains all of the patterns required by an EF-Hand protein. If this search is successful, the data are written to an output file, and the program moves on to the next FASTA record.

If it is not successful, the program then searches the same record to determine whether or not it contains a Pseudo EF-Hand motif. One drawback to this method is that in cases where a protein may have more than one motif in the sequence, the program ceases to search the record as soon as a positive match is found.

Subsequent to this analysis, a new set of EF Hand patterns was developed to evaluate multiple, variable length gaps in the sequence, and a new pattern search algorithm was developed to test the new patterns. In this second version of the pattern search algorithm, CaSeqAnalysisA, the search methodology was changed from multiple patterns for each motif, to a single pattern. This also changed the logic for the search method. Instead of searching for the loop pattern followed by the helices, this algorithm performed a linear search of the sequence based on Yang Pattern 2 (Table 2.4 and Table 2.5). This search methodology was complicated by the presence of variable length gaps for the new pseudo-EF-Hand pattern. For example, the pattern value X(5,9) indicates that a gap ranging from 5-9 characters is present between identified pattern residues. A series of nested loops within the algorithm was required to reset the search parameters within the main loop body and accommodate the presence of one or more variable length gaps.

Although this increased the complexity of the pattern search, the CPU time increases were negligible.

2.3.2 Results and Discussion

Data resulting from the execution of the CaSeqAnalysis program can be observed in Table 2.6 and Table 2.7. From Table 2.6 it can be seen that the EF Hand patterns developed in our laboratory demonstrated a significant improvement over the Prosite pattern. The statistical criteria for these assessments are based on the following matrix as defined by Prosite(http://au.expasy.org/prosite):

- TP (True Positive): Protein belongs to domain and identified by pattern.
- TN (True Negative): Protein does not belong to domain, excluded by pattern.
- FP (False Positive): Protein does not belong to domain but identified by pattern.
- FN (False Negative): Protein belongs to domain but not identified by pattern.
- Potential: Protein could belong to domain but not currently established as such.

These matrix values are then used to measure Precision (Equation 1) and Positive Predictive Value, or PPV (Equation 2), where Precision indicates the ratio of True hits to all data, and PPV is the ratio of True Positive hits to all Positive hits.

$$\text{Precision} = (TP + TN)/(TP + TN + FP + FN) \qquad \text{(Equation 1)}$$

$$\text{PPV} = TP/(TP + FP) \qquad \text{(Equation 2)}$$

Table 2.6: Analysis of EF-Hand pattern matching.

	Prosite	Yang *et al.*	Yang *et al.* v.2
Total TP:	93	91	92
Total TN:	0	45	32
Total FP:	51	6	19
Total FN:	25	27	26
Total Potential:	1	1	1
	170	170	170
Precision	55.03%	80.47%	73.37%
PPV	64.58%	93.81%	82.88%

TP = True Positive. TN = True Negative. FP = False Positive. FN = False Negative. PPV = Positive Predictive Value.

Table 2.7: Analysis of Pseudo EF-Hand pattern matching.

	Prosite	Yang *et al.*	Yang *et al.*, v.2
Total TP:	70	42	80
Total TN:	0	5	5
Total FP:	5	0	0
Total FN:	10	40	2
Total Potential:	2	0	0
Total Records	87	87	87
Precision	82.35%	55.29%	97.70%
PPV	93.33%	100.00%	100.00%

TP = True Positive. TN = True Negative. FP = False Positive. FN = False Negative. PPV = Positive Predictive Value.

A comparison of the results between the two patterns shows that TP and FN between the two pattern models are close to each other. However, the Yang EF-Hand patterns successfully identify as TN 45 of the 51 sequences identified by Prosite as FP. In other words, the Yang triple pattern EF-Hand search excluded those sequences that are invalid. This latter category represents the statistical gains observed for Precision and PPV by the Yang Pattern 1 search, where Precision increased from 55.03% to 80.47%.

Figures 2.3 and 2.4 depict the statistical differences as bar charts. The original Yang Pseudo EF hand patterns, however, were statistically worse than the Prosite patterns, as seen in Table 2.7, and Figures 2.5 and 2.6, reflecting a decrease in Precision from 82.35% to 55.29%.

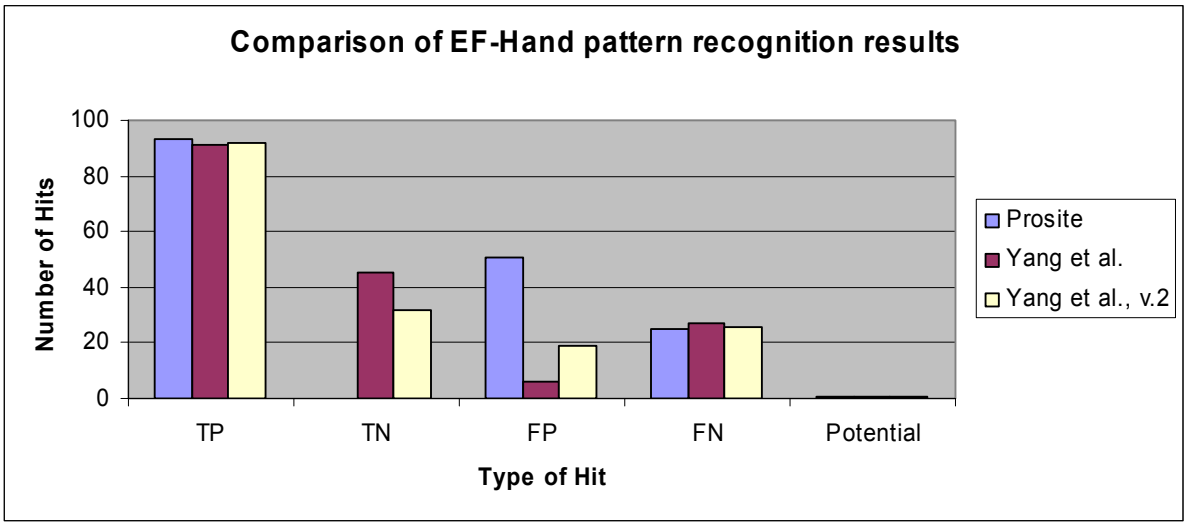


Figure 2.3: Comparison of EF-Hand pattern recognition results.

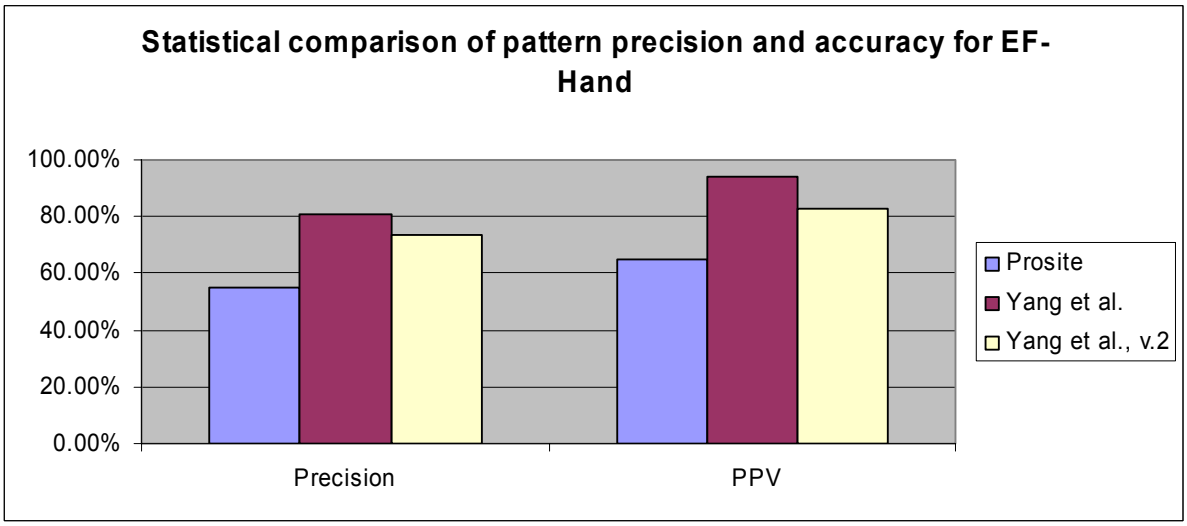


Figure 2.4: Comparison of Precision and accuracy between Prosite and Yang patterns.

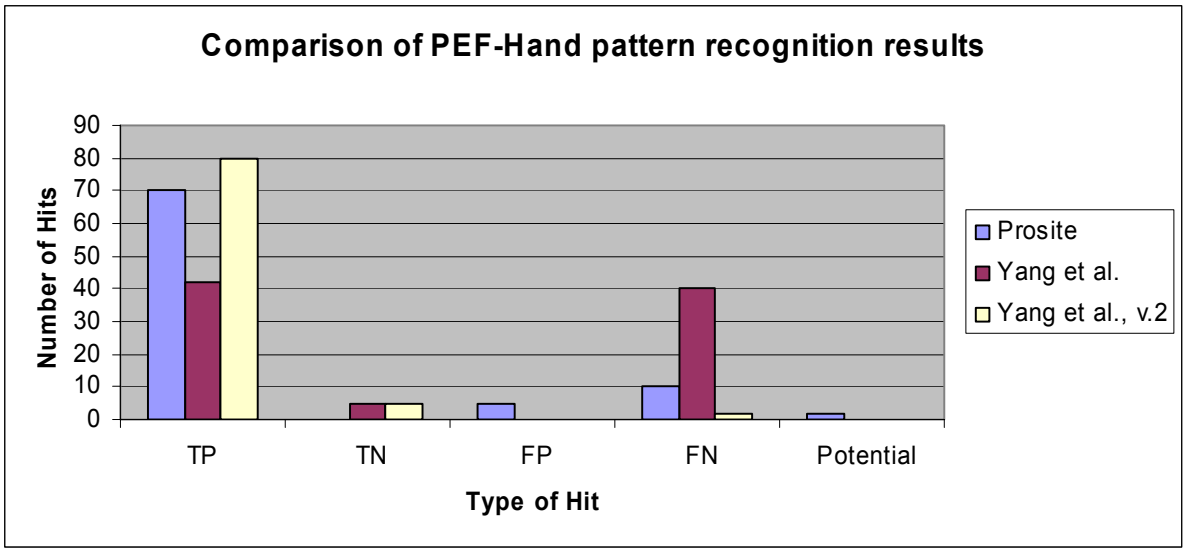


Figure 2.5: Comparison of Pseudo EF-Hand pattern recognition results.

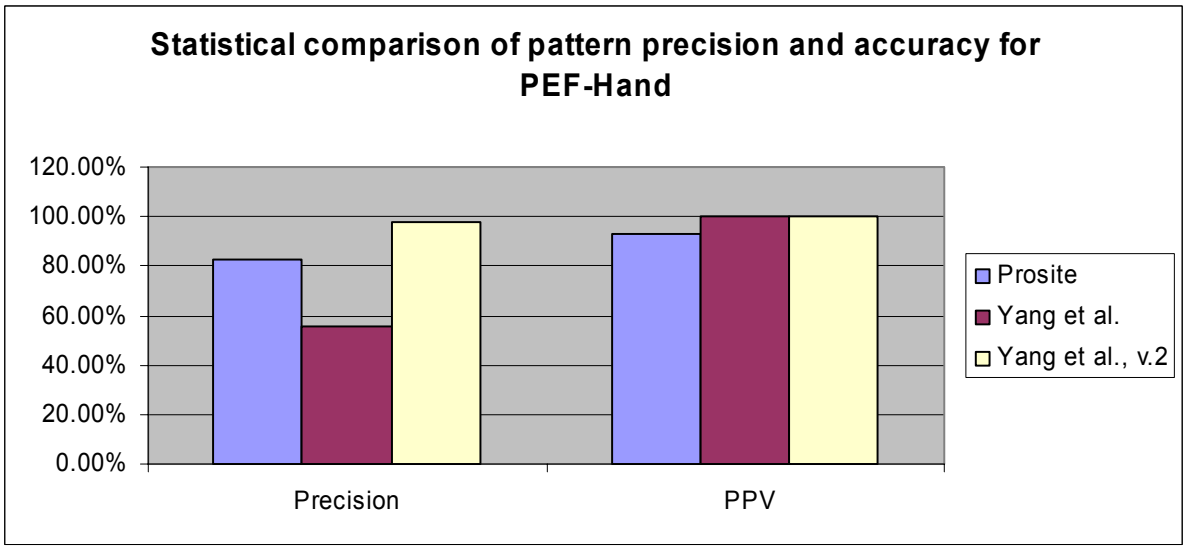


Figure 2.6: Comparison of Precision and accuracy between Prosite and Yang patterns.

Execution of CaSeqAnalysisA (version 2) produced interesting and unexpected results. The modified Pseudo-EF-Hand pattern (Yang Pattern 2), with inclusive variable-length gaps in the sequence, produced an increase in Precision to 97.70% compared to the

Prosite Precision for the same dataset of 82.35%. PPV for Yang Pattern 2 was 100.00% compared to Prosite's 93.33%.

However, the single, linear pattern search for EF-Hand using the Yang Pattern 2 resulted in a decrease in Precision in comparison with the Yang Pattern 1 results which were based on separation of the linear pattern into three smaller patterns corresponding to each component of the helix-loop-helix structure. Precision between Yang Pattern 1 and Yang Pattern 2 decreased from 80.47% to 73.37%, due primarily to a loss in TN hits (45 down to 32 hits), which were evaluated as FP (False Positives). Additional testing of the algorithm was conducted to determine if this loss could be corrected within the code, but an analysis of the reported FP hits indicates that in all cases, they conform to the pattern, so the error in this case could not be corrected programmatically. Nonetheless, even with this loss, Yang Pattern 2 still demonstrated an approximately 18% increase in Precision over Prosite's 55.03%, with a corresponding gain in PPV (82.88% vs. 64.58%, respectively).

For Pseudo-EF-Hand, Yang Pattern 2 was significantly more precise than the corresponding Prosite pattern, and came very close (97.70%) to 100% Precision. The Yang Patterns for EF-Hand were conclusively more precise than the Prosite patterns for each of the two pattern-matching algorithms tested using the linear search methodology.

2.3.3 Conclusions

Yang Pattern 2 for EF-Hand, which duplicated the Prosite pattern format for a single pattern sequence, produced better prediction results than the Prosite pattern. Some loss in precision was observed between this structure and Yang Pattern 1 (73.37% vs.

80.47%), although both outperformed the Prosite pattern (55.03%). Conversely, results were just the opposite for Pseudo EF-Hand, where the first pattern had a calculated precision of 55.29% that improved to 97.70% with the consolidated pattern. For this second Pseudo EF-Hand pattern, precision was again greater than the Prosite pattern (82.35%).

It can be concluded from these data that the highly-conserved nature of these motifs allow for reasonably accurate prediction based on the type and placement of ligands within the sequence. Although these parameters alone are insufficient for prediction in all cases, they are an essential component that can be used in conjunction with other statistical and structural parameters to develop more complex algorithms and methods for the prediction and identification of metal binding sites in proteins.

3 Statistical Analysis of Calcium-binding Proteins

3.1 Introduction

The purpose of this study was to statistically evaluate the structural characteristics associated with Ca^{2+}-binding sites, utilizing a broad sampling of known sites, to identify specific physical parameters and key characteristics associated with Ca^{2+}-binding proteins. The resulting data were expected to provide a foundation for the comparative analysis between Pb^{2+} and Ca^{2+} binding characteristics. Additionally, these results would provide the basis for new input used in the prediction and design of Ca^{2+}-binding sites in proteins.

3.2 Materials and Methods

Four datasets from previous studies were selected for analysis. These datasets were selected as they represented a diverse sampling of known Ca^{2+}-binding sites, were evaluated in previous works utilizing different methods, and contained a very limited number of EF-Hand sites. Because EF-Hand proteins represent more than half of all Ca^{2+}-binding sites, and are characterized by a more fixed geometry, it was necessary to limit the number of EF-Hand sites analyzed to obtain a general analysis of Ca^{2+}-binding sites.

As discussed in Section 1.4, the common feature for evaluating metal-binding sites in known crystal structures relies on the presence of clustered binding ligands: oxygen, nitrogen, and sulfur. In contrast to assertions noted in previous studies, the presence of identifiable oxygen clusters alone has been shown as sufficient for fast

identification of calcium-binding sites in calcium-binding proteins [28]. The four datasets presented in this work (Table A.1, Appendix A) were selected because they were produced in the past decade, and, for those involving computational predictions, they rely on clustering of ligands in the three-dimensional structure as a precursor to applying their prediction algorithms.

Dataset I was reproduced from Nayal and Di Cera [84], who evaluated 62 Ca^{2+}-binding sites from 32 Ca^{2+}-binding proteins using data from PDB files. The Nayal study suggests, in concurrence with Sodhi *et al.*, that structural data alone cannot characterize the Ca^{2+}-binding site for prediction purposes [84, 93]. The Nayal and Di Cera algorithm utilized a grid system to evaluate valences around a grid point selected based on the presence of at least 3 oxygen atoms within a probe radius of 3.4 Å. A prediction rate of 99.7% was reported using this method, based on the prediction of the Ca^{2+}-binding site within 3.5 Å of the actual site. It is important to note with respect to the Nayal study that the clustering of oxygen atoms was integral to selection of potential sites.

Dataset II was reproduced from Pidcock and Moore, who in 2001 conducted a comprehensive statistical analysis of binding sites for both Ca^{2+} and La^{3+} ions, utilizing data provided in the PDB and the Cambridge Structural Database (http://www.ccdc.cam.ac.uk/) [63]. Dataset II from Pidcock and Moore originally contained 44 proteins with 94 calcium-binding sites, obtained from a survey of 515 fully normalized crystal structures of calcium-binding proteins in PDB from 1994 to 1999 with resolution values over the range 1.0-2.5 Å. This initial dataset was reduced in our study to 44 structures with 60 sites by removing structures whose calcium-binding sites share ligand residues or ligands other than those donated by the protein and water.

Dataset III in our study was based on work done by Dudev *et al.*, who evaluated the role of second-shell atoms and their interactions with first shell atoms involved as binding ligands in metal-binding sites [83]. Because the second shell contains many of the main chain or peptide backbone atoms and structures, these constituents contribute to metal selectivity by constraining the metal-binding site with respect to ionic radii and coordination geometry. The Dudev study utilized density functional theory/continuum dielectric methods (DFT/CDM) to evaluate the dielectric medium surrounding the metal-binding site, which was determined by first identifying the oxygen, nitrogen and sulfur atoms surrounding the metal-binding site. The original Dataset consisted of 34 proteins with 81 calcium-binding sites, constrained by PDB X-ray and NMR structures with a resolution less than 3.0, and, with only one exception, no more than 30% sequence homology with the other selected protein structures.

Dataset IV contains 17 proteins with 46 calcium-binding sites, and reproduces the dataset selected by Schymkowitz *et al.* (2005), where they reported the use of the Fold-X force field for the prediction of protein metal-binding sites that include both the structural water molecules (i.e.- water molecules that interact with 2 or more polar atoms in the protein) as well as the metal ions [87]. This method, which performs fast calculations of free energy, distinguishes between high- and low-affinity binding sites, and can distinguish between binding for Ca^{2+}, Mg^{2+} and Zn^{2+}.

The characteristics of the calcium-binding sites from the selected datasets that were analyzed statistically in this study are graphically presented in Figure 3.1.

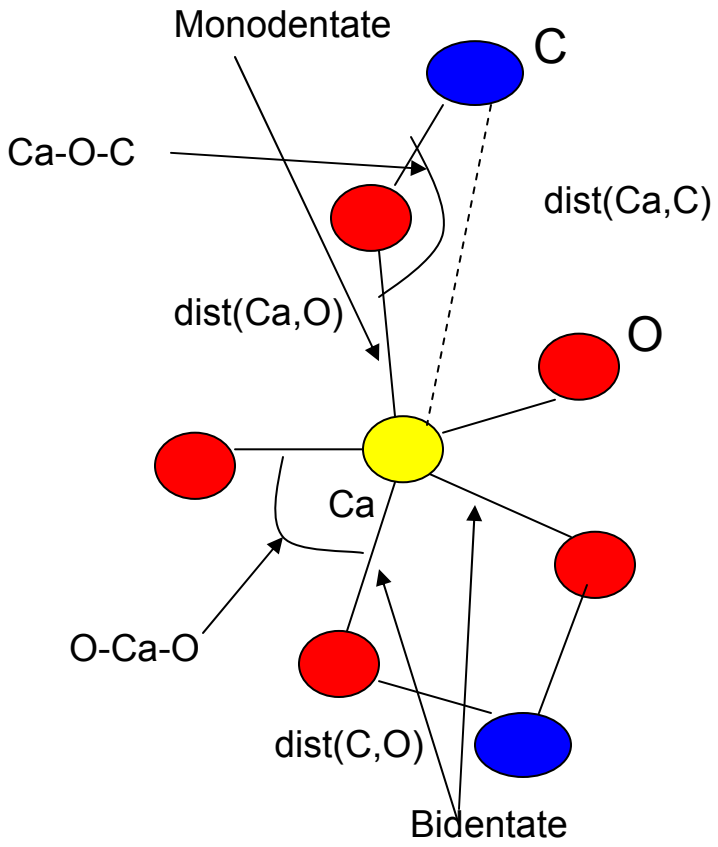


Figure 3.1: Identification of analyzed features within a Ca^{2+}-binding site. Ca-O-C is the angle between the ligand carbon, ligand oxygen, and the calcium ion. O-Ca-O is the angle between the calcium ion and two ligand oxygen atoms. The labels dist(Ca,O) dist(Ca, C), and dist(C,O) refer to distances between calcium-oxygen, calcium-carbon, and carbon-oxygen, respectively.

As seen in Figure 3.1, Ca-O-C is the angle between the central Ca^{2+} (Ca) ion, the ligand oxygen (O), and the carbon (C) bound to the ligand oxygen. O-Ca-O is the angle between two oxygen ligands (O) and the central Ca^{2+} (Ca) ion. Distances evaluated were: a) between the central Ca^{2+} (Ca) and the carbon (C) bound to the ligand oxygen (dist(Ca,C)); and b) between the ligand oxygen (O), and the carbon (C) bound to the ligand oxygen (dist(C,O)). Distinction is also made in the statistical analysis between monodentate and bidentate ligands.

The Avg Ca-O and Avg Ca-C values were calculated as follows in Equation 3 and Equation 4, respectively.

$$AvgCa-O = \frac{\sum_{1}^{k} dist(Ca,O)}{k} \quad \text{(Equation 3)}$$

$$AvgCa-C = \frac{\sum_{1}^{m} dist(Ca,C)}{m} \quad \text{(Equation 4)}$$

In Equation 3, *k* is the number of ligands in one site. In Equation 4, *m* is the number of bonded carbon atoms, and *k* must be greater than or equal to *m*. When *k* equals *m* in a single binding site, it indicates that only monodentate ligands appear in this site, otherwise *k* must be greater than *m* for polydentate ligands.

To examine the spatial relationship between the inner oxygen shell and the outer bonded carbon shell of calcium-binding sites, the ratio of Avg Ca-O and Avg Ca-C (r_RO_RC), is calculated using Equation 5. It is hypothesized that a fixed ratio should exist for this relationship, based on the average C-Ca and average O-Ca bond lengths and constraints on the C-O-Ca angle. Equation 6 calculates an adjusted ratio, where *NB* represents the number of bidentate residue(s) in a putative calcium-binding site and 0.05 is an empirically-derived coefficient of *NB* to determine the extent of the negative effect on *ar_RO_RC* produced by inclusion of the bidentate ligands.

$$r\_RO\_RC = AvgCa\text{-}O/AvgCa\text{-}C \quad \text{(Equation 5)}$$

$$ar\_RO\_RC = r\_RO\_RC - 0.05*NB \quad \text{(Equation 6)}$$

Angle and distance values were calculated using Matlab (MathWorks, Natick, MA). Statistics were compiled and graphically-rendered using MS Excel (Microsoft corporation, Redmond, WA).

3.3 Results and Discussion

3.3.1 Coordination Number and Ligand Type

Table 3.1: Coordination numbers (CN) and ligand distribution of analyzed datasets.

	Dataset I	Dataset II	Dataset III	Dataset IV	Total	% Total
Coordination Number						
Total Sites	54	78	69	44	245	100.0
CN8	0	0	2	0	2	0.7
CN7	11	12	10	8	41	14.5
CN6	20	17	12	19	68	24.0
CN5	11	28	36	13	88	31.1
CN4	12	21	9	4	46	16.3
CN3	0	0	0	0	0	0.0
Ligand Distribution						
Total Ligands	361	794	581	346	2082	
Water	110	211	214	117	652	31.3
SC Carboxylates	131	235	237	132	735	35.3
SC Amides	34	32	19	13	98	4.7
MC Carbonyl	44	55	56	42	197	9.5
SC Hydroxyl	10	14	12	14	50	2.4
Cofactors	2	0	1	0	3	0.1

CN = Coordination Number. SC = Sidechain. MC = Mainchain.

Table 3.1 summarizes the coordination numbers (CN) and distribution of ligand types for the calcium binding sites evaluated in each of the four sample datasets, and Table 3.2 summarizes mean values for all measured data. For the purpose of this study, a cutoff distance of 3.5 Å [84] was selected, so that only ligand atoms (O, N) within this distance would be considered coordinating ligands. As a consequence of this distance constraint, the CN values reported in table 3.1 differ from those reported for the original

datasets. The majority of coordination number (CN) values reported ranged from 4-7 (excluding oxygen from water), representing 86.5% of the total, with an average CN per calcium-binding site of 5.6 ± 0.2.

Table 3.2: Mean values for all calcium binding data, datasets I-IV.

	I	II	III	IV	Mean	Sdev
Average CN	5.6	5.3	5.8	5.7	5.6	0.2
Mean Ca-O Distance	2.43	2.46	2.48	2.42	2.45	0.03
Mean Ca-C Distance	3.34	3.34	3.34	3.25	3.32	0.05
Average Ca-O-C Angle	130.6	133.3	129.6	128.7	130.5	2.0
Mean r_RO_RC	0.73	0.74	0.74	0.75	0.74	0.01
Mean ar_RO_RC	0.69	0.70	0.70	0.69	0.69	0.00
Charged Residues	2.5	2.3	2.4	2.8	2.5	0.2

It should be noted that ligands can be from the protein itself, cofactors, or water molecules. With respect to ligand type, side-chain carboxylate oxygen atoms represent the major ligand contributor, followed by water molecules, which stresses the significance of solvent exposure during binding.

3.3.2 Distance Parameters

Figure 3.2 presents the average distance values for ligand oxygen atoms (first shell) and their associated carbon atoms (second shell). From Figure 3.2 it can be observed that the peak value of Avg Ca-O is 2.4, indicating an Average Ca-O per site between 2.3 and 2.5, based on 0.1 as an interval bin. Also, the peak values of Avg Ca-C are 3.2 for all datasets except IV, which peaks at 3.3. For all four datasets, the resulting curves are very similar, and exhibit a characteristic decrease as the distance for both the ligand oxygen atoms and their associated carbon atoms approaches the first/second shell interval, which is expected.

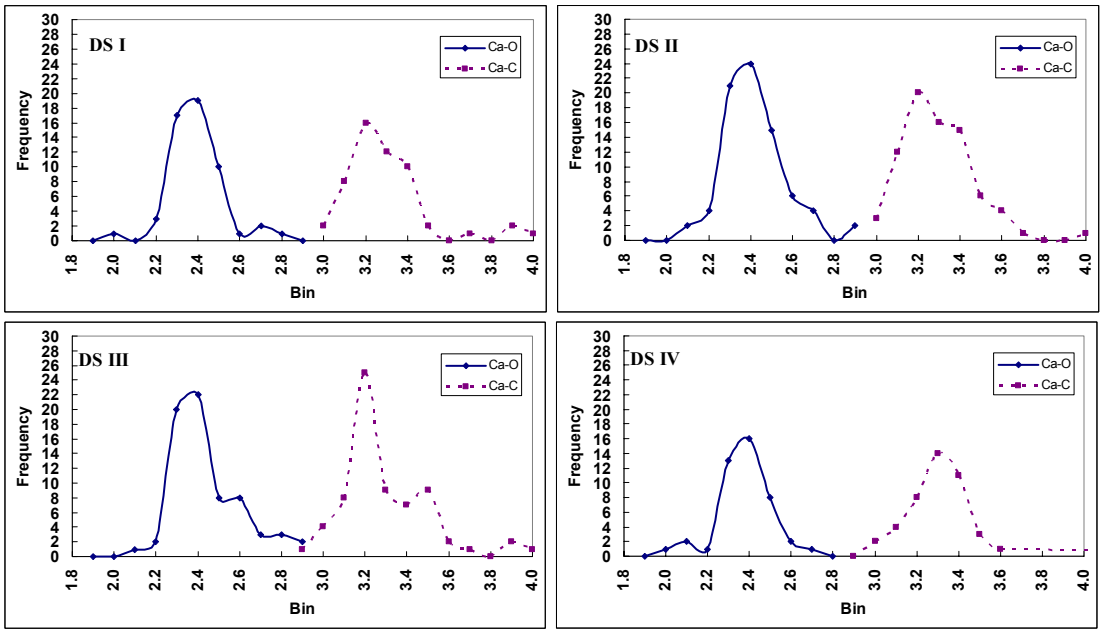


Figure 3.2: Distribution of average Ca-O (Avg Ca-O) and average Ca-C (Avg Ca-C) distances for all datasets (DS I – DS IV).

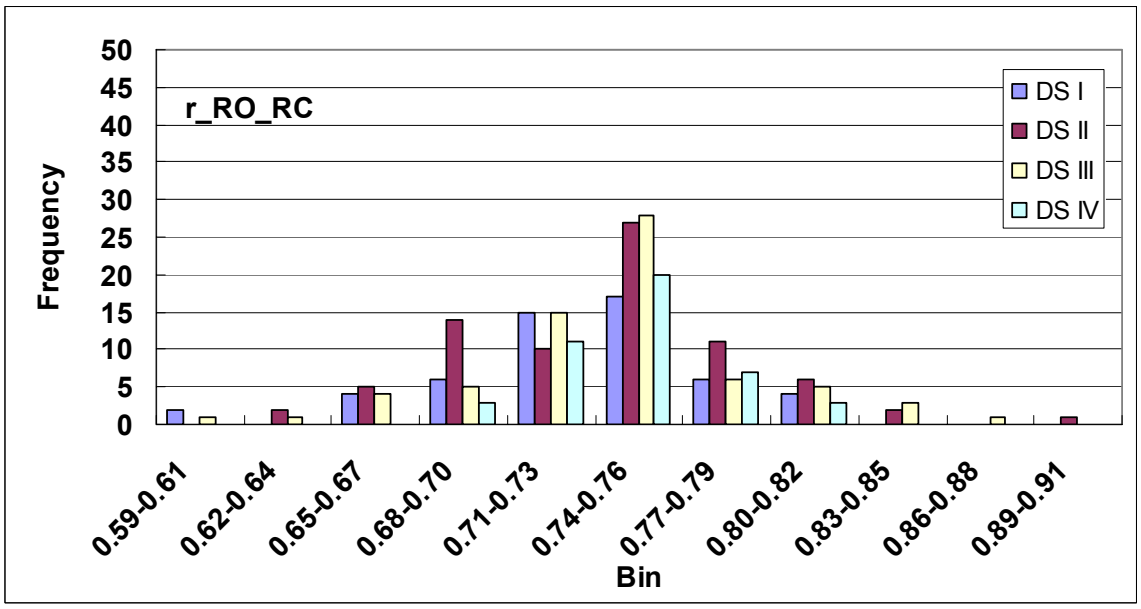


Figure 3.3: The distribution of r_RO_RC in datasets I-IV.

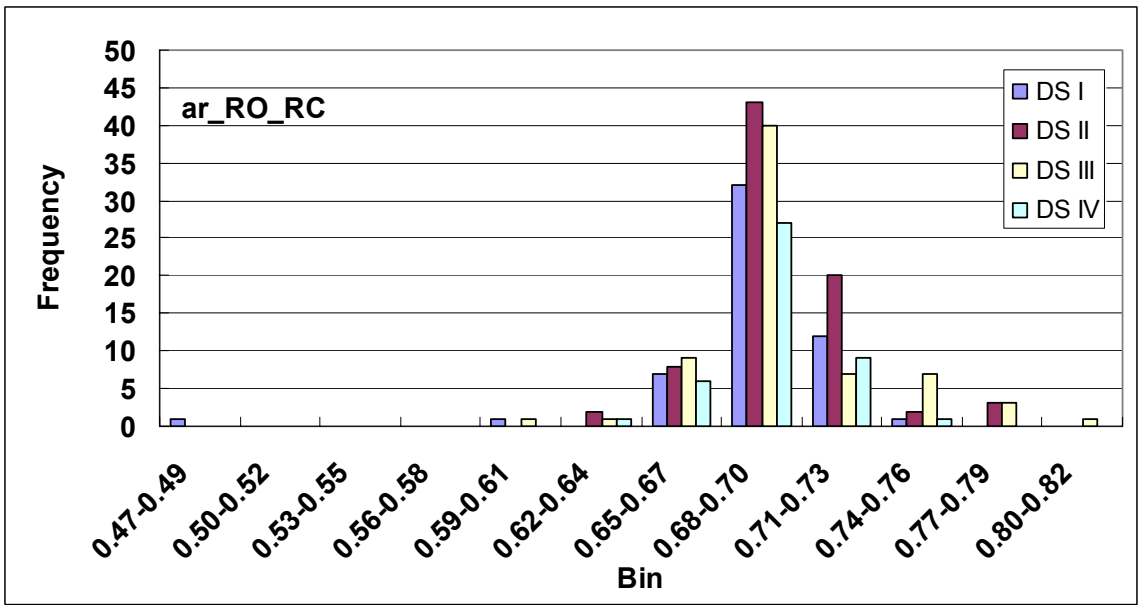


Figure 3.4: The distribution of ar_RO_RC in datasets I-IV.

The calculated ratios r_RO_RC and ar_RO_RC for the four datasets are represented in Figure 3.3 and Figure 3.4, respectively, with mean values calculated in Table 3.2. It is clear that the majority of r_RO_RC is represented by the value ranging from 0.65 to 0.82 with an interval bin of 0.03. After adjustment of r_RO_RC, the majority of ar_RO_RC values range from 0.65 to 0.76. Therefore, a relatively stable range value exists to quantitatively evaluate the first-second shell geometric properties of known calcium binding sites.

3.3.3 Bidentate Residue Effects on r_RO_RC

More interestingly, we find that the bidentate property of both Asp and Glu have an important impact on the distribution of r_RO_RC. The effect of differentiating sites based on the presence of bidentate ligands was examined for all four datasets. Similar results were obtained for all analyses. The results for only Dataset II, which included the most proteins of the four evaluated datasets, are depicted in Figure 3.5. As seen in Figure

3.5, changes in r_RO_RC based on inclusion of bidentate ligands results in a decreased ratio, which is to be expected, as the bidentate ligand distances (O-Ca) are shorter than non-bidentate distances (data not shown).

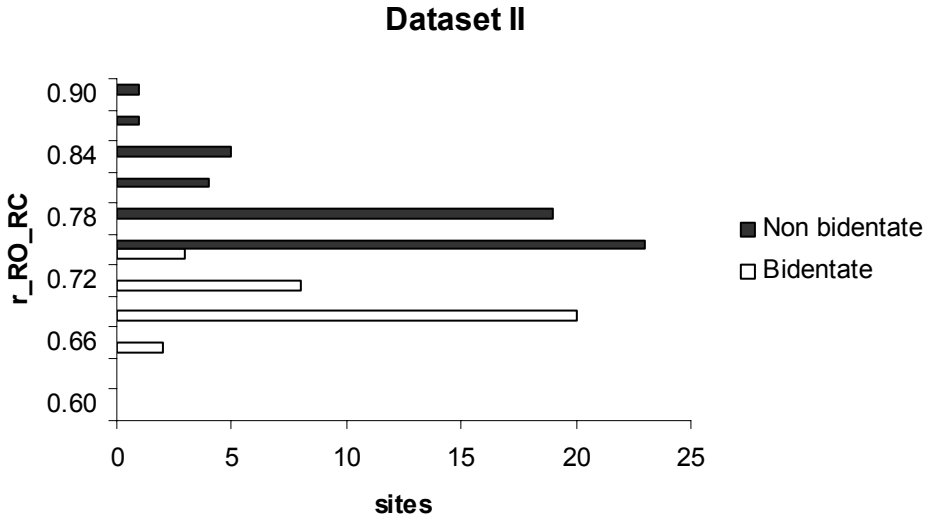


Figure 3.5: Distribution of r_RO_RC for non-bidentate (black) and bidentate (white) sites.

A clear cutoff between the two distributions is apparent at 0.75. Even though there is overlap between 0.72 and 0.75 of ar_RO_RC for both sites, the frequency (3 sites) from bidentate sites is very small. While Figure 3.5 illustrates the distinction between sites including bidentate ligands, it does not explain this phenomenon. To answer that, it is necessary to evaluate the Ca-O-C bond angle.

3.3.4 Analysis of Bond Angles

The Ca-O-C angle constrains the proximity of carbon from the calcium cation to some extent. A decrease in this angle generally results in a decrease in the C-Ca distance. It can be seen in Figure 3.6 that the bidentate Ca-O-C angles range between 100º and

150º whereas the monodentate (i.e. – non-bidentate) Ca-O-C angles range between 130º and 160º. These ranges overlap between 130º and 150º which also correspond with the highest frequency distribution across the datasets.

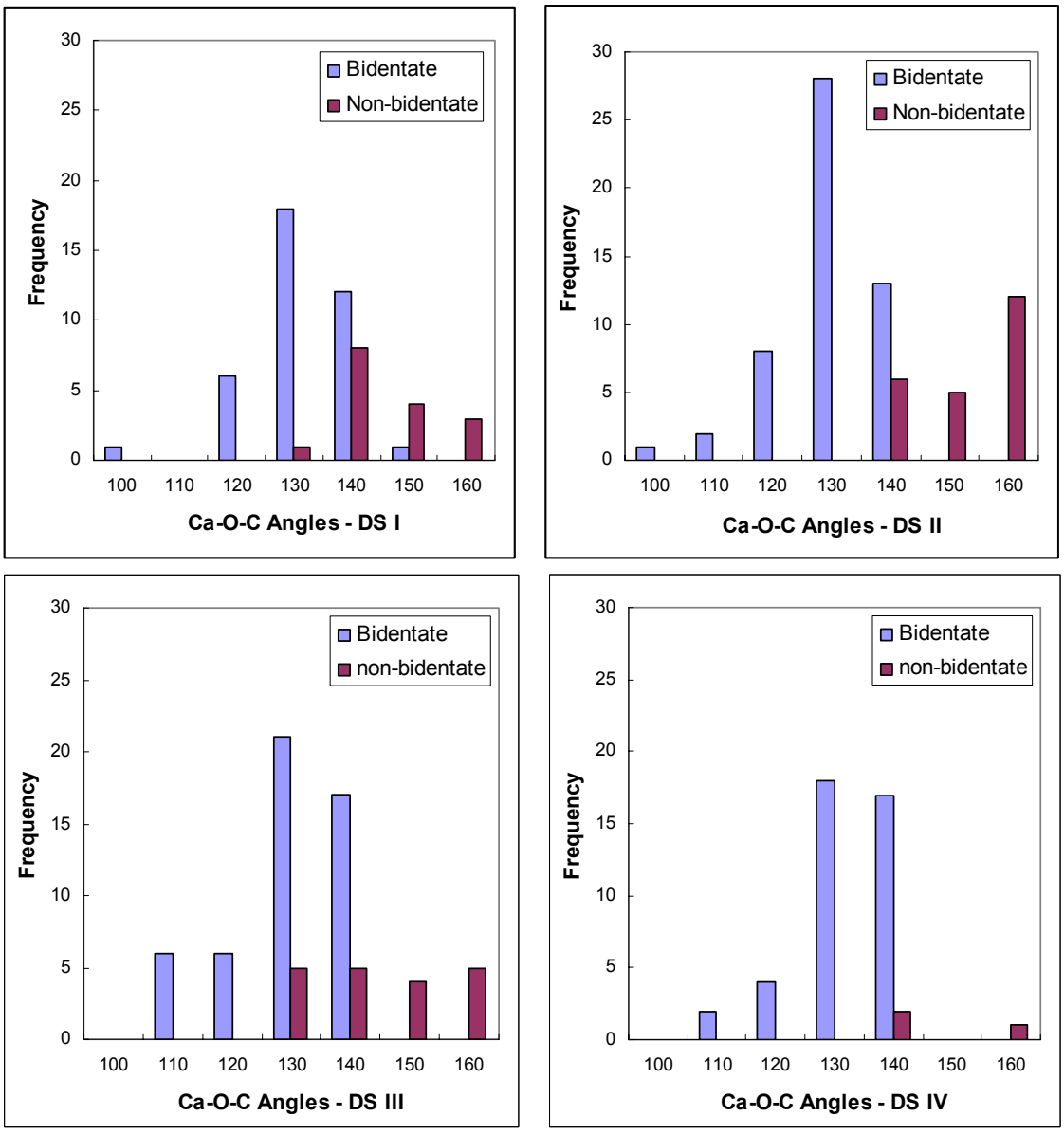


Figure 3.6: Distribution of Ca-O-C angles for bidentate and monodentate residues over datasets I-IV.

The bidentate angle properties can also be characterized by the sum of the angle pairs, as seen in Figure 3.7. Here we see a more constrained range, where the majority of the frequency distribution values fall between 180º and 200º, with a bin interval of 10º.

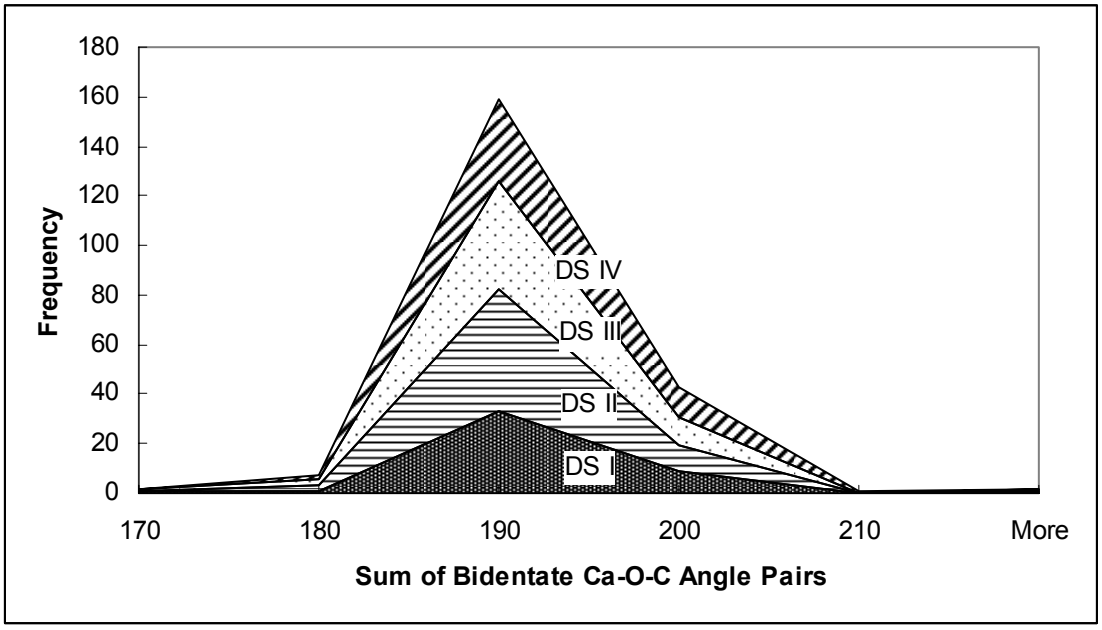


Figure 3.7: Distribution of Ca-O-C angle pair sums for bidentate residues over datasets I-IV.

3.3.5 Relationship Between r_RO_RC and ar_RO_RC

In order to relate the individual ligand Ca-O-C angles to r_RO_RC for the binding site, the average Ca-O-C (denoted as AvgCa-O-C) angle is defined by calculating the mean for all Ca-O-C angles in each individual site. In this way, both AvgCa-O-C and r_RO_RC can be compared. Figure 3.8 shows the correlation of AvgCa-O-C and r_RO_RC for Dataset II.

As seen in Figure 3.8, it can be observed that AvgCa-O-C is always greater than 90. Additionally, the coordinate convergence appeared linear, with AvgCa-O-C decreasing in inverse proportion to r_RO_RC, so a Pearson product moment correlation

coefficient (*r*) was calculated, as detailed in Equation 7, to verify the linear relationship between these variables.

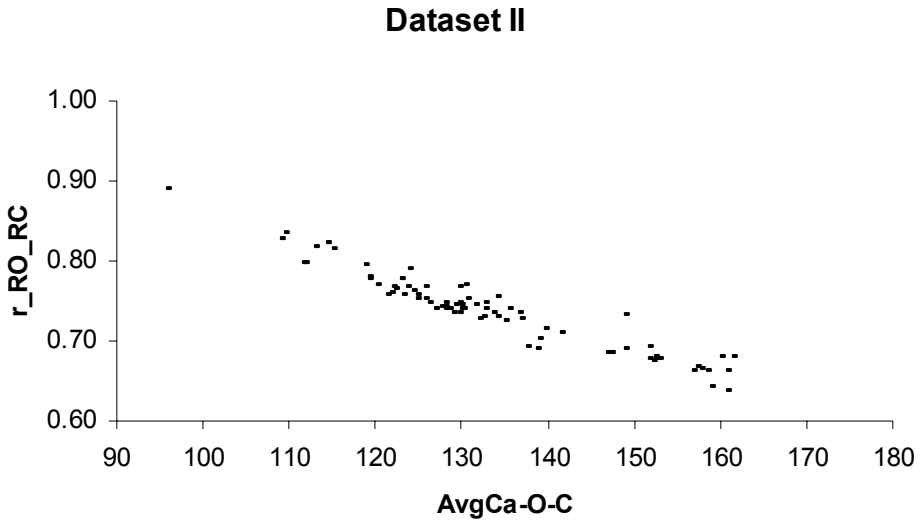


Figure 3.8: Correlation between r_RO_RC and AvgCa-O-C.

$$r = \frac{\sum (x - \overline{x})(y - \overline{y})}{\sqrt{\sum (x - \overline{x})^2 (y - \overline{y})^2}} \qquad \text{(Equation 7)}$$

In Equation 7, as defined in MS Excel, (*r*) is a measure of the correlation of two variables *X* and *Y* measured on the same object or organism; that is, a measure of the tendency of the variables to increase or decrease together. The values for x bar and y bar are the sample means AVERAGE(x) and AVERAGE(y). The value *r* indicates linearity between X and Y, where $r = -1$ indicates perfect negative correlation and $r = 1$ is perfect positive correlation. The linear model is not applicable as *r* approaches zero. For our analysis, r was found to be -0.96. This high negative correlation demonstrates that the

AvgCa-O-C angle decreases as the ligand oxygen distance from the ion increases relative to the carbon ion distance. This implies that the AvgCa-O-C angle is constrained by the C-O bond orientation relative to the calcium ion. Although the energy consequences of this orientation are not evaluated in this work, it is clear that the electrostatic force of the liganding electrons follows a vector. Second, compression of the angle can occur which may serve both to maintain the C-O bond orientation, but also may allow for continued ligand interaction without significantly altering the main chain conformation. This latter possibility has important ramifications with respect to competitive binding, and evidence reported in the structural analysis of Pb^{2+}-binding appears to support this.

3.3.6 Charge Analysis by Site

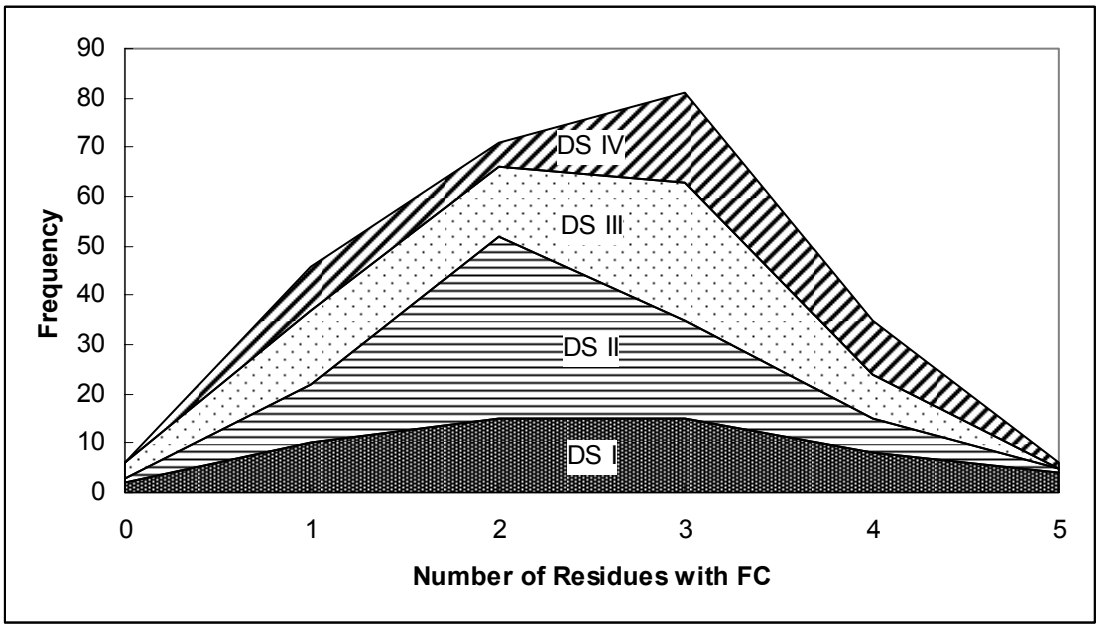


Figure 3.9: Frequency of occurrences of charged residues by site for datasets I-IV.

Because Ca^{2+} is a positively-charged ion, the microenvironment surrounding the ion must necessarily exhibit a net negative charge of -2, which may or may not

correspond to the formal charges associated with residues in the site. It was previously noted that 3-4 negative charges in the primary coordination shell may provide optimal charge configuration [7, 24]. In the studies cited for datasets used in this work, formal charge is defined where side chain carboxyl groups have a charge of -1, main chain carbonyls from Glu or Asp have a charge of -1, and all other main chain carbonyls have a formal charge of zero. Figure 3.9 for the four datasets used in this study indicates the frequency of charges follows the order -3 > -2 > -1 > -4.

Interestingly, six sites were found with a formal charge of zero, and six sites with a formal charge of 5. To further examine this, data from PDB files for 1BJR and 1HYT were obtained. Relevant data from these PDB files are summarized in Table 3.3. Models of the two binding sites were completed using Pymol (http://pymol.sourceforge.net/), and are presented in Figure 3.10.

Table 3.3: Ligand Charge Data for Two Calcium Binding Sites.

PDB_ID	HetAtm Res	ResNm	Atom CharID	RSeq	Ca Bind Dist	C-Lig-Ca Angle	Ligand source	Chg
1BJR	290	Arg	O	12	2.77	149.18	MC Carbonyl	0
1BJR	290	Ser	O	15	3.28	169.84	MC Carbonyl	0
1BJR	290	Asn	OD1	257	2.83	144.17	SC Carbonyl	0
1BJR	290	Ala	O	273	2.98	132.93	MC Carbonyl	0
1HYT	801	Asp	OD2	138	2.32	132.93	SC Carboxyl	1
1HYT	801	Glu	OE1	177	2.45	101.84	SC Carboxyl	0.5
1HYT	801	Glu	OE2	177	2.82	82.44	SC Carboxyl	0.5
1HYT	801	Asp	OD1	185	2.43	132.09	SC Carboxyl	1
1HYT	801	Glu	O	187	2.27	150.98	MC Carbonyl	1
1HYT	801	Glu	OE1	190	2.48	92.90	SC Carboxyl	0.5
1HYT	801	Glu	OE2	190	2.51	89.81	SC Carboxyl	0.5

HetAtm Res indicates the sequence number associated with the calcium ion; **ResNm** is the amino acid providing the ligand atom; **AtomCharID** indicates the atom type; **RSeq** is the primary sequence number of the residue; **Ca Bind Dist** is the distance between the ion and the ligand atom; **C-Lig-Ca Angle** is the angle between the calcium ion, the binding ligand atom and its associated carbon; **Ligand source** indicates whether the ligand atom is from the main chain or the side chain, and the functional group; and **Chg** indicates the formal charge associated with each ligand atom.

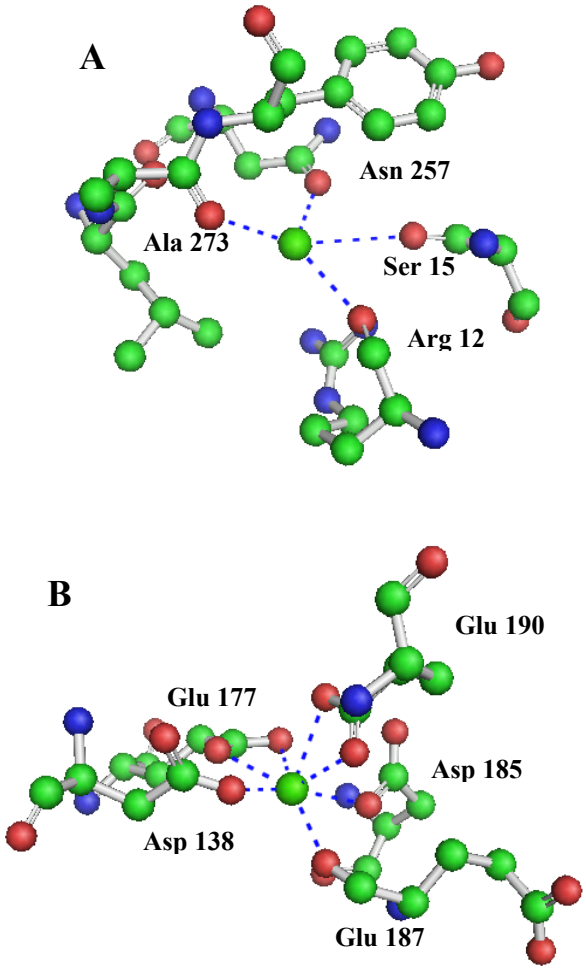


Figure 3.10: Calcium-binding sites from PDB files: (A) 1BJR, where calcium 290 has a charge of zero in the binding site, and (B) 1HYT, where calcium 801 has a charge of 5 surrounding the ion.

The structure presented in PDB file 1BJR shows the crystallized complex between proteinase K, chain E, and a buffalo lactoferrin fragment [94]. The structure presented in 1HYT is a complex formed between benzylsuccinic acid and thermolysin [95]. The calcium-binding sites in both structures can be considered atypical. In 1BJR, the binding site, which corresponds to the weaker affinity site reported by Muller *et al.*

for proteinase K [96], indicates different binding ligands in the 1BJR structure. In 1HYT we see a zinc-binding protein complex, with a bound calcium ion incidental to the structure. With respect to charge, the 4 binding ligands surrounding calcium 290 in 1BJR are all carbonyl oxygen atoms, with no formal charge (Table 3.3), which would account for weaker binding affinity. For calcium 804 in 1HYT, seven ligand atoms bind the calcium ion, comprised of two pairs of bidentate ligands that each contributes a charge of -1, two side chain carboxyl groups that each contributes a charge of -1, and a final charge of -1 from a main chain Glu carbonyl.

Figure 3.10 indicates an irregular geometry for both binding sites, as compared with the more symmetric geometry typically seen in EF-Hand calcium binding motifs (Figure 1.2). Nonetheless, the structures observed in Figure 3.10 agree with the calculated charges.

3.4 Conclusions

In general, data presented in this study show close correlation with previously reported data, but also present some important distinctions, as well as new parameters.

3.4.1 Coordination Number and Ligand Type

As previously noted, the majority of coordination number (CN) values reported in this study ranged from 4 to 8, with an average CN per calcium-binding site of 5.6 ± 0.2. These calculated mean values, which differ from those first reported by Glusker [4] (CN = 7.3), are refinements based on the increase of available data since these comprehensive works were first published, and exclude water molecules. The CN range reported is consistent with more recently published work [10, 83], although the distribution of

ligands types differs. Dudev and Lim cite a distribution of ligands, in decreasing order, as carboxylates, carbonyls, water, and hydroxyl atoms [83]. However, data presented in Table 3.1 suggest that water may be more prevalent in the binding process than previously assumed, as 31.3% of the total 2082 ligand oxygen atoms putatively are contributed by water, second only to oxygen atoms contributed by side-chain carboxylate groups, at 35.3%. This value is considerably higher than the 20% summarized by McPhalen [93], and the 22.4% reported by Dudev [83]. In the case of Dudev, this difference can be attributed to the difference in ligand distance cutoff values (2.9 Å vs. 3.5 Å in this study), however, the longer binding distance has been shown to more accurately predict Ca^{2+}-binding sites [28, 84].

3.4.2 Distance Parameters

The mean distance reported in this study for the Ca^{2+}-O bond length (2.45 ± 0.03 Å) is more precise than the previously reported mean value of 2.4 ± 0.2 Å, whereas the range (2.2 – 2.8 Å) presents a more concise distribution than previously reported values (2.01 – 3.15 Å) [4, 12, 52-54]. Coupling this mean value with the mean value derived for Ca-C bond length in a ratio defined in this study as r_RO_RC, provides a quantifiable characteristic defining the relationship between the inner and outer coordination shells which was found to be 0.65-0.82, as seen in Figure 3.3. This range, and the associated variance, is further refined to 0.65-0.76 for ar_RO_RC which accounts for variations due to the presence of bidentate ligands common in calcium-binding sites.

3.4.3 Analysis of Bond Angles

Analysis of the Ca-O-C angles indicated a range between 130°-160°, and a statistical representation of the data indicates that the AvgCa-O-C must be greater than 90°. This is important for defining possible binding geometries, as it adds an additional constraint on the ligand-ion interaction based on the orientation angle of C-O with respect to the Ca^{2+} ion. In other words, while the O-Ca distance may fall within the cutoff range, the measured angle may indicate that the proposed ligand is not involved in the binding. A good example of this would be the case where a carboxylate group has both oxygen atoms in close enough proximity for binding, but the Ca-O-C angle of one oxygen actually aligns it with the other oxygen ligand. In this case, it is more probable to assume that the electrons from the more distance oxygen shift to the nearer oxygen, resulting in a single ligand atom with greater electron density.

Finally, data summarized in this study with respect to formal charge in the binding site microenvironment indicated that the frequency of charges follows the order -3 > -2 > -1 > -4, as demonstrated in Figure 3.9. This differs slightly from previously reported values indicating that optimal charge configuration may involve 3-4 negative charges in the primary coordination shell [7, 24]. This difference can most likely be attributed to differences in the data sets evaluated, as the same criteria for determining charge were utilized.

4 Statistical and Structural Analysis of Pb^{2+}-binding Proteins

4.1 Introduction

The objective of this study was to provide a comprehensive statistical and structural analysis of Pb^{2+} binding characteristics derived from all crystalline protein structures deposited in PDB. Unlike Ca^{2+}-binding proteins, which have been studied extensively, no previous work has systematically evaluated the binding of Pb^{2+}. From the hypothesis presented in Section 1.7, it was expected that the binding of Pb^{2+} will exhibit different characteristics than those observed for Ca^{2+}. By comparing the statistical results of this analysis with the results for Ca^{2+} as presented in Chapter 3, we can begin to elucidate the key parameters associated with competitive binding, and build a foundation to further assess the physical mechanisms associated with toxicity.

4.2 Materials and Methods

Work was completed for the development of a relational database (MetalBank) and support software to create a local environment for data mining and statistical analyses. All database development was completed using Microsoft Access (Microsoft Corporation, Redmond, WA). While the long-term objective of this research project was to include all metal binding proteins, the first iteration focused on proteins known to bind Pb^{2+} only.

Data for the analyses were obtained from the RCSB Protein Data Bank (http://www.rcsb.org/pdb/). Initially, 27 PDB files for proteins known to bind lead (Pb^{2+}) were identified for analysis. Six of the original 27 were removed from the analysis as

they represented either duplicate structures or nucleic acid structures. Additionally, of the 68 Pb^{2+} ions in the 21 retained structure files, 20 ions were removed from the analysis to either eliminate redundancy associated with polymeric domains, or because the only binding ligands associated with the ion were water molecules. A summary of the number of retained Pb^{2+} ions and their associated PDB files is presented in Table 4.1.

Table 4.1: Summary list of crystallized PDB proteins found to bind Pb^{2+} ions

[a]PDBId	Description	Author	[b]Res (Å)	Total Pb^{2+} ions	Ret Pb^{2+} ions
1AFV	Hiv-1 Capsid Protein (P24) Complex	Momany	3.7	2	1
1E9N	A Second Divalent Metal Ion In The	Beernink	2.2	4	2
1FJR	Crystal Ectodomain Of Methuselah	West	2.3	4	2
1HD7	A Second Divalent Metal Ion In	Beernink	1.95	1	1
1HQJ	Crystal A De Novo Designed...	Burkhard	1.2	9	7
1IW7	Crystal The Rna Polymerase...	Vassylyev	2.6	4	2
1KA4	Pyrococcus Furiosus Carboxypeptidase	Arndt	3	1	1
1N0Y	Crystal Pb-Bound Calmodulin	Wilson	1.75	14	11
1NA0	Design Of Stable Alpha-Helical...	Main	1.6	5	3
1QNV	Yeast 5-Aminolaevulinic Acid...	Erskine	2.5	2	2
1QR7	Crystal PHE-Regulated 3-Deoxy-D-...	Shumilin	2.6	4	1
1SN8	Crystal The S1 Domain Of Rnase E...	Schubert	2	2	2
1SYY	R2 Subunit Of Ribonucleotide R...	Hogbom	1.7	1	1
1V0D	Caspase-Activated DNAse (Cad)	Woo	2.6	1	1
1XXA	E.Coli ARG Repressor ...	Van	2.2	4	2
1ZHW	Yeast Oxysterol Binding Protein Osh4	Im	1.7	1	1
1ZHY	Yeast Oxysterol Binding Protein Osh4	Im	1.6	2	2
2ANI	F127y Mutant Of Ribonucleotide...	Hogbom	2	1	1
2CH7	The Cytoplasmic Domain...	Park	2.5	2	2
2FP1	Chorismate Mutase...	Okvist	1.55	2	2
2G0A	Mouse Pyrimidine 5'-Nucleotidase	Bitto	2.35	2	1
			Total	68	48

[a]Protein DataBank Identification.
[b]Resolution

Table 4.2 lists each of the 48 retained ions, their relevant PDB identifiers, coordination number (CN) values, and Formal Charge (FC) by site. Charge by site was based on previous work related to analyses of calcium-binding sites, where sidechain carboxyl groups have a charge of (-1), main-chain carbonyls from Glu or Asp have a

charge of (-1), and all other main-chain carbonyls have a formal charge of zero [7]. Charge parameters for this study were modified to accommodate additional ligand types utilized by Pb^{2+} as follows: a charge of (-1) is assigned to main-chain carbonyls from Glu or Asp, side-chain carboxyl groups, Cys, and side-chain amide nitrogens. In the case of bidentate ligands, a charge of (-0.5) is assigned [97].

Table 4.2: PDB identifiers for retained Pb^{2+} ions.

PDB_ID	HetAtm SerNbr	ResSeq	CN	Site Charge	PDB_ID	HetAtm SerNbr	ResSeq	CN	Site Charge
1AFV	9029	1	2	-1	1N0Y	1322	508	9	-5
1E9N	4341	1319	6	-4	1N0Y	1323	509	2	-1
1E9N	4342	1320	3	-2	1N0Y	1324	510	10	-4
1FJR	3168	602	3	-1	1NA0	1973	274	2	-1
1FJR	3169	603	3	-1	1NA0	1975	276	2	-1
1HD7	2072	1319	3	-2	1NA0	1976	277	3	-2
1HQJ	1645	2003	2	-1	1QNV	2548	400	5	-3
1HQJ	1646	2004	1	-1	1QNV	2549	401	2	-1
1HQJ	1647	2005	2	-1	1QR7	10296	351	6	-3
1HQJ	1648	2006	1	-1	1SN8	1330	604	1	0
1HQJ	1649	2007	6	-2	1SN8	1331	605	2	-1
1HQJ	1650	2008	2	-1	1SYY	2617	1321	3	-1
1HQJ	1651	2009	4	-2	1V0D	1940	601	2	-1
1IW7	53577	6058	6	-3	1XXA	3715	417	4	0
1IW7	53578	6112	5	-4	1XXA	3716	418	4	-1
1KA4	4165	601	7	-2	1ZHW	3516	2001	2	-1
1N0Y	1314	500	4	-3	1ZHY	3516	2001	3	-1
1N0Y	1315	501	11	-5	1ZHY	3517	2002	2	-1
1N0Y	1316	502	2	-1	2ANI	2618	421	2	-1
1N0Y	1317	503	9	-4	2CH7	4633	1530	4	-2
1N0Y	1318	504	4	-2	2CH7	4634	1531	4	-2
1N0Y	1319	505	5	-3	2FP1	2705	501	5	-2
1N0Y	1320	506	3	-2	2FP1	2706	502	4	-2
1N0Y	1321	507	4	-3	2G0A	4655	903	7	-3

PDB_ID indicates PDB Identification. **HetAtmSerNber** is PDB serial number for ion. **ResSeq** is the sequence number for the ion. **CN** is the coordination number.

Additionally, supplemental data related to ligand contact surface area and ligand distance was located for these files from the Ligand Protein Contact Data Server (LPC) (http://bioportal.weizmann.ac.il/oca-bin/lpccsu) to verify binding ligands identified from

the database analysis. Data from LPC was manually loaded into Excel spreadsheets (Microsoft Corporation, Redmond, WA), and imported into the MetalBank database.

Data from PDB was selectively filtered to exclude residues and segments not associated with the Pb-ligand complex formation. To accelerate data filtration, a Visual Basic (Microsoft Corporation, Redmond, WA) program was developed to select the PDB-formatted text input file, specify ranges of residues to retain, and then build the retained residue segments into a comma-delimited output file for import into the MetalBank database. The graphical user interface (GUI) for this program is pictured in Figure 4.1.

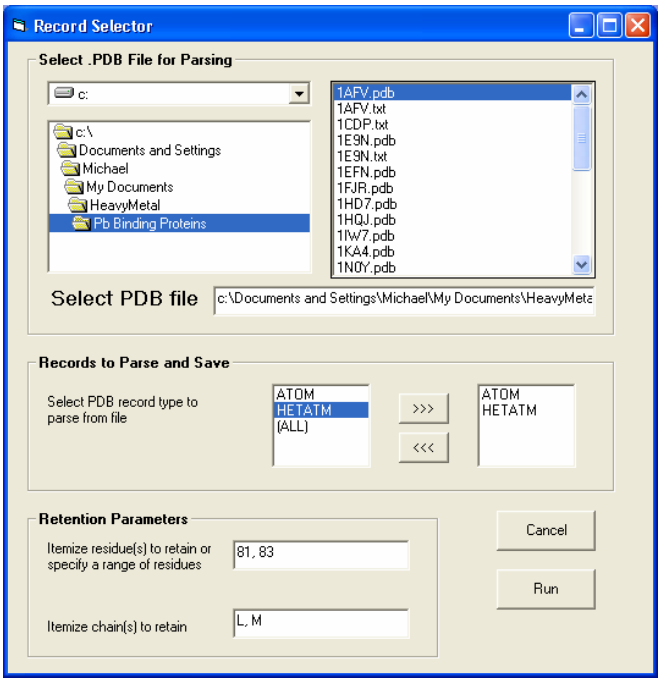


Figure 4.1: GUI interface for PDBRecordParser program.

The key MetalBank tables and their relationships are pictured in Figure 4.2. The Metalloproteins table is the parent table, with a one-to-many relationship to each of the three child tables. The LigContDataRecs table includes ligand data from LPC. The

PDBAtomRec table contains all PDB data from the selected files. The PDBHetAtomsFinal table contains the PDB data for only the 48 Pb^{2+} ions selected for the study. This latter table was separated to simplify queries. Although the present version of MetalBank contains only data for Pb^{2+} and some select calcium-binding proteins, it can be expanded in the future to accommodate other metalloproteins.

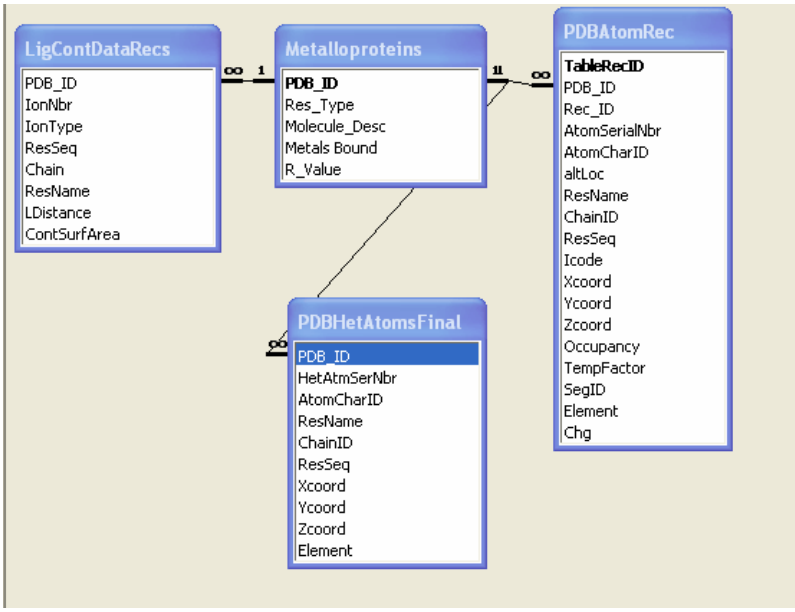


Figure 4.2: MetalBank database table relationship scheme.

Once the database tables were built, and their relationships defined, collection of relevant binding data was accomplished with the ad-hoc Structured Query Language (SQL) query, seen in Figure 4.3.

```
      SELECT PDBHetAtomsFinal.HetAtmSerNbr, PDBAtomRec.PDB_ID,
PDBAtomRec.Rec_ID, PDBAtomRec.AtomSerialNbr, PDBAtomRec.AtomCharID,
PDBAtomRec.altLoc, PDBAtomRec.ResName, PDBAtomRec.ChainID,
PDBAtomRec.ResSeq, PDBAtomRec.Icode, PDBAtomRec.Xcoord,
PDBAtomRec.Ycoord, PDBAtomRec.Zcoord, PDBAtomRec.Occupancy,
PDBAtomRec.TempFactor, PDBAtomRec.SegID, PDBAtomRec.Element,
Sqr((PDBHetAtomsFinal.XCoord-
PDBAtomRec.XCoord)^2+(PDBHetAtomsFinal.YCoord-
PDBAtomRec.YCoord)^2+(PDBHetAtomsFinal.ZCoord-PDBAtomRec.ZCoord)^2) AS
BindDist
      FROM PDBAtomRec INNER JOIN PDBHetAtomsFinal ON
PDBAtomRec.PDB_ID=PDBHetAtomsFinal.PDB_ID
      WHERE (((PDBAtomRec.AtomCharID) Like "O*" Or
(PDBAtomRec.AtomCharID) Like "S*" Or (PDBAtomRec.AtomCharID) Like "N*")
And ((Sqr((PDBHetAtomsFinal.XCoord-
PDBAtomRec.XCoord)^2+(PDBHetAtomsFinal.YCoord-
PDBAtomRec.YCoord)^2+(PDBHetAtomsFinal.ZCoord-
PDBAtomRec.ZCoord)^2))<4.2));
```

Figure 4.3: SQL query to identify Pb^{2+}-binding ligands.

This query identified all oxygen, nitrogen or sulfur ligand atoms within 4.2 Å of each of the 47 Pb^{2+} ion selected for the analysis, where Distance was calculated based on Equation 8.

$$\text{Distance} = \left((X_{Ligand} - X_{Pb^{2+}})^2 + (Y_{Ligand} - Y_{Pb^{2+}})^2 + (Z_{Ligand} - Z_{Pb^{2+}})^2\right)^{1/2} \quad \text{(Equation 8)}$$

The query also added the calculated distance into a column in the query results for eventual statistical analysis. It should be noted that the cutoff distance was selected based on previous work suggesting that Pb-O and Pb-N distances may be as long as 4.2 Å [98].

Statistical analyses, tables and graphs were completed using Microsoft Excel. Multiple Sequence Alignments for PDB proteins 3CLN, 1EXR, and 1N0Y were completed using ClustalW (http://www.ebi.ac.uk/clustalw/). Graphical representations of calcium binding ligands in 3CLN, 1EXR and 1N0Y were created using Pymol (http://pymol.sourceforge.net/).

4.3 Results and Discussion

4.3.1 Statistical Analysis of Pb^{2+}-binding

A summary of the statistical analysis results is presented in Table 4.3. For the 48 ions selected from the original 21 PDB proteins, 252 ligand atoms were identified based on a cutoff distance of 4.2 Å. However, 65 of those ligand atoms were oxygen atoms from sources other than amino acids (e.g. – water molecules or sulfate groups). These atoms were separated for calculations of bond length and excluded from determination of the coordination numbers, as their presence is an uncontrolled variable in the crystallized structure.

Table 4.3: Summary of Pb^{2+}-binding statistics based on O, N, S ligand distance < 4.2 Å.

Total PDB Proteins in study:	21
Total Pb ions evaluated:	48
Total target ligand atoms:	252
Total O Ligands, AA:	146
Total O Ligands, non-AA:	65
Total N Ligands:	28
Total S Ligands:	13
Mean atom distance, Oxygen, AA[a]:	2.87 ± 0.57
Mean atom distance, Oxygen, non-AA[b]:	3.10 ± 0.61
Mean atom distance, Nitrogen:	3.47 ± 0.71
Mean atom distance, Sulfur:	3.16 ± 0.27
Mean Coordination Number (CN)	3.9
% CN 2-5 Ligands	75.0
% CN 6-9 Ligands	17.0
Mean Charge	-1.9

[a] AA: Amino Acid
[b] non-AA: Non-Amino Acid

As seen in Table 4.3, oxygen atoms represent the major binding ligand for Pb^{2+} (146), followed by nitrogen (28) and sulfur (13). Figure 4.4 sub-divides these totals as percentages based on the source of the ligand atoms (main-chain vs. side-chain, both

oxygen and nitrogen). As seen in Figure 4.4, 70.6% of the amino acid ligands in close enough proximity to bind Pb^{2+} are side-chain oxygen atoms, with only 7% originating from main-chain oxygen atoms. Only 15% of all potential ligands are nitrogen, however, nearly a third of these originate from the main-chain, which suggests that when nitrogen is involved as a binding ligand, it may result in larger conformational changes.

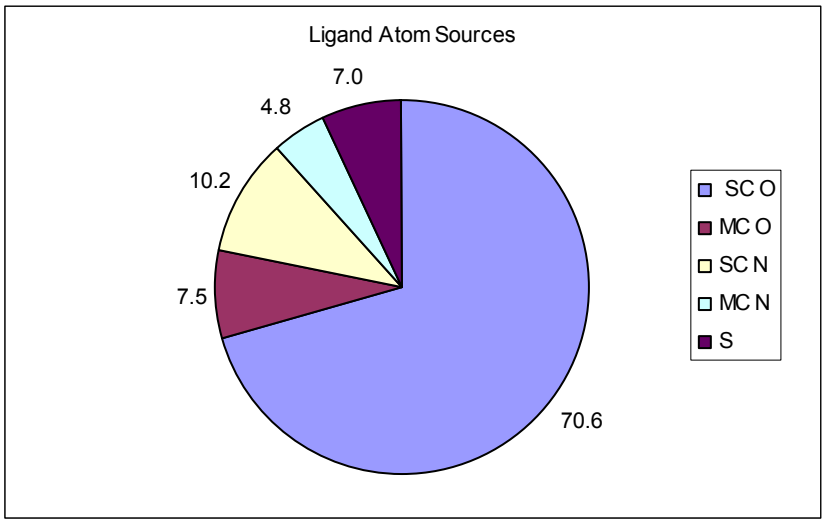


Figure 4.4: Percentage of ligand atoms based ligand atom source for 48 Pb^{2+}-binding sites. Distribution is separated based on side-chain oxygen (SC O), main-chain oxygen (MC O), side-chain nitrogen (SC N), main-chain nitrogen (MC N), and sulfur (S).

Table 4.4 and Figure 4.5 summarize the number of ligand atoms based on their amino acid residue of origin, with no distinction made between main-chain and side-chain in Figure 4.5. Glu is the predominant ligand residue, contributing 74 of the 187 total binding ligand atoms. Aside from Glu, only Asp (51), Cys (15) and Asn (11) appear to any appreciable extent.

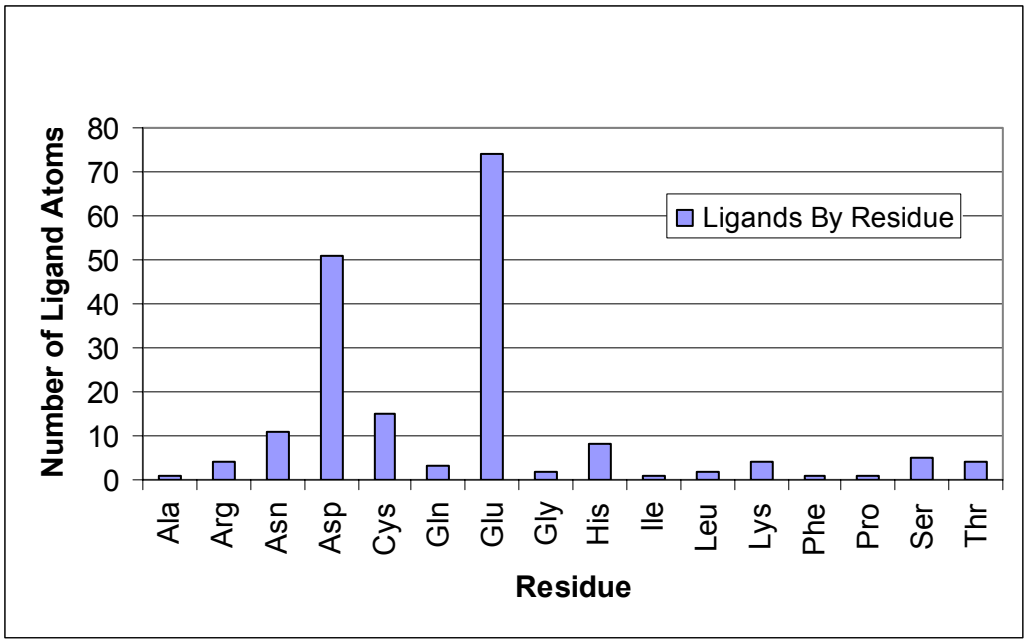


Figure 4.5: Graphical distribution of ligand atoms by residue for Pb^{2+}-binding.

Table 4.4: Summary of ligand atoms by residue.

LigResName	# Atoms	[a]SC Oxygen AA	[b]MC Oxygen AA	SC Nitrogen	MC Nitrogen	Sulfur	% By Residue
Ala	1	0	1	0	0	0	0.53
Arg	4	0	1	3	0	0	2.14
Asn	11	3	1	5	2	0	5.88
Asp	51	49	0	0	2	0	27.27
Cys	15	0	0	0	2	13	8.02
Gln	3	1	1	1	0	0	1.60
Glu	74	74	0	0	0	0	39.57
Gly	2	0	1	0	1	0	1.07
His	8	0	1	7	0	0	4.28
Ile	1	0	1	0	0	0	0.53
Leu	2	0	1	0	1	0	1.07
Lys	4	0	0	3	1	0	2.14
Phe	1	0	1	0	0	0	0.53
Pro	1	0	1	0	0	0	0.53
Ser	5	4	1	0	0	0	2.67
Thr	4	1	3	0	0	0	2.14
Total	187	132	14	19	9	13	100.00

[a]Sidechain.
[b]Mainchain.

Table 4.3 also summarizes the mean binding distance values determined from the analysis. The Pb-O distance (for amino acids) was found to be 2.87 ± 0.57 Å, which is slightly higher than the mean of 2.4 reported for Ca-O. This may be explained by the fact that Pb^{2+} has a larger ionic radius than Ca^{2+} (1.19Å vs. 0.99Å). The longer Pb-O distance for oxygen from sources other than amino acid residues (3.10 ± 0.61 Å vs. 2.87 ± 0.57 Å), could be due to the fact that the amino acids should have more restricted motion than a free molecule during crystallization.

The Pb-S and Pb-N mean distances were found to be 3.16 ± 0.27 Å and 3.47 ± 0.71 Å, respectively. A comparison of the mean binding length values summarized in Table 4.3, and data from the Sarrett study [98], indicates that the binding length distances from the database analysis for Pb-O (2.87 ± 0.57 Å), Pb-N (3.47 ± 0.71 Å) and Pb-S (3.16 ± 0.27 Å) agreed with the ranges reported by Sarret as 2.2-4.2 Å, 3.0-4.2 Å, and 2.6-3.4 Å, respectively.

The analysis produced interesting results with respect to coordination number, as summarized in Table 4.3 and Figure 4.6. A coordination number of 2 was most commonly observed, with a mean CN of 3.9, which is closer to the values suggested for Zn^{2+} (5.0) than for Ca^{2+} (6.9). Additionally, 75% of the surveyed ions had a coordination number between 2-5, with only 17.0% indicating a coordination number between 6-9, This is also inconsistent with expectations based on atomic radii, Although both Zn^{2+} and Ca^{2+} have been identified as target ions displaced by Pb^{2+}, the effective atomic radius of Pb^{2+} (1.12 Å) is much closer to Ca^{2+} (0.99 Å) than to Zn^{2+} (0.71 Å). The two sites reported in Figure 4.6 with 10 and 11 potential binding ligands are unusual, but closely parallel the high coordination geometry of 9-10 ligands previously cited [62]. Also in

Figure 4.6, three sites were reported with only a single amino acid based ligand atom, 1HQJ – 1646, 1HQJ – 1648 and 1SN8 – 1330. The sites at 1HQJ – 1646 and 1HQJ – 1648 can be explained by the fact that they include oxygen ligand atoms from water molecules. The site at 1SN8 – 1330 is less readily-explained, as the Pb^{2+} ion appears to be bound by a single His nitrogen at 2.5 Å, with no supporting water molecules to coordinate binding (Figure 4.7).

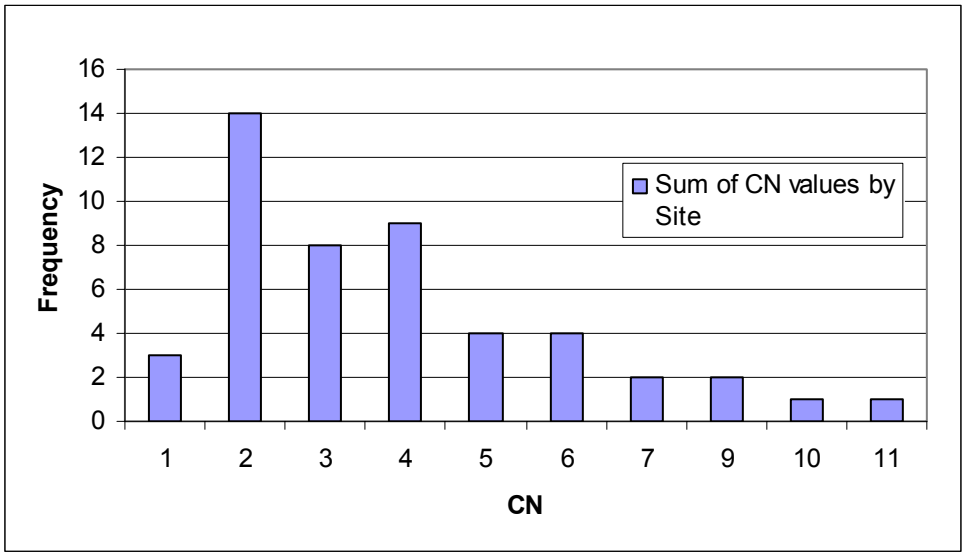


Figure 4.6: Frequency distribution of CN values for 48 Pb^{2+}-binding sites.

Figure 4.8 indicates the frequency distribution of formal charge (FC) for the selected binding sites. The majority of sites included an FC value of -1 or -2, which is lower than the -2 or -3 found during the statistical analysis of calcium. The mean FC reported in Table 4.3 of -1.9 with a majority distribution between -1 and -2 differs from data reported for Ca^{2+}. Summarily, the FC and CN data suggest that Pb^{2+} ions, unlike Ca^{2+} and Zn^{2+}, not only do not require a regular coordination structure, but can bind with very small coordination numbers, and potentially, with higher affinity. This conclusion is

further supported by the results seen in Figure 4.9 for binding with sulfur, which do not indicate a fixed geometry comprised of either 3 or 5 ligands.

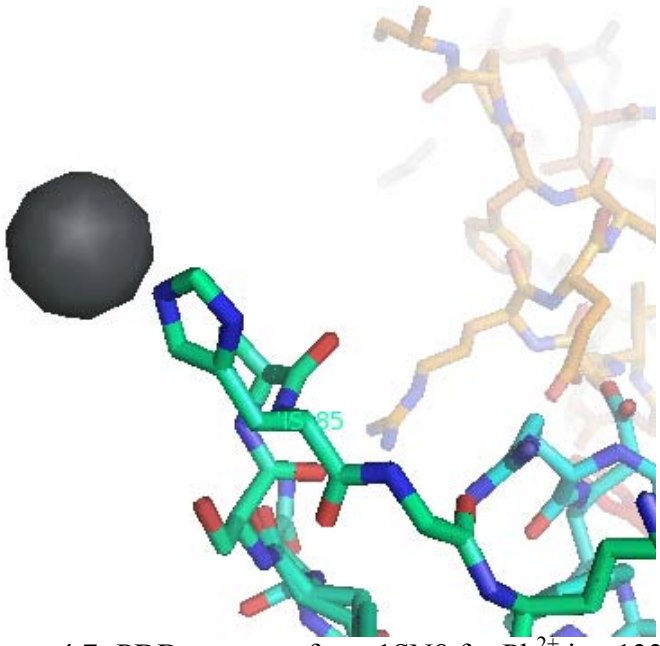


Figure 4.7: PDB structure from 1SN8 for Pb^{2+} ion 1330 where binding is observed with only a single nitrogen ligand from His 85.

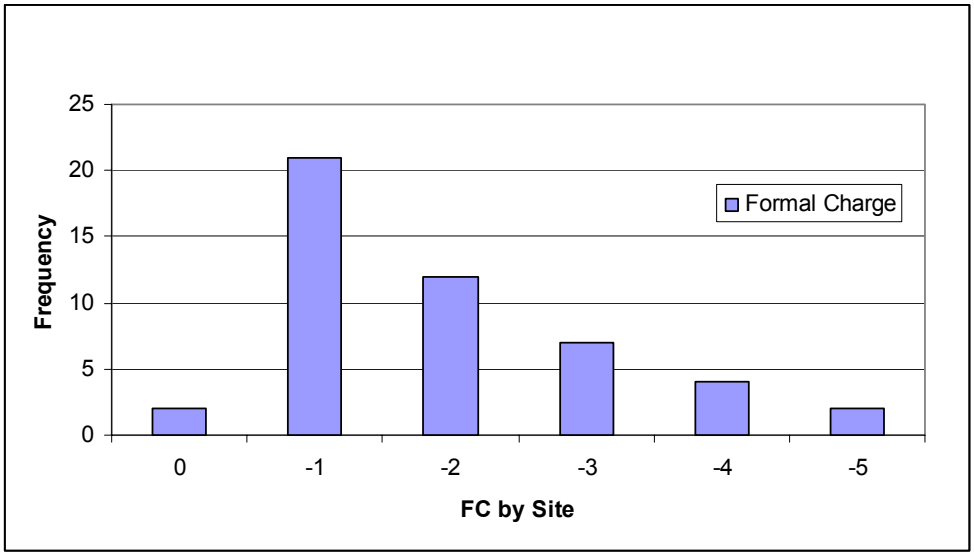


Figure 4.8: Frequency distribution of FC values for 48 Pb^{2+}-binding sites.

It was previously noted that when Pb^{2+} binds in a thiol-rich site, it reportedly tends to result in a trigonal-pyramidal geometry, with three coordinating ligands, and a fourth vertex being occupied by the $6s^2$ lone pair from Pb^{2+} [74, 75]. Only six Pb^{2+}-binding sites that included sulfur ligands were retained in this study. Of those, CN values of 2, 5, and 6 binding ligands were found, excluding 1QR7 – 10296 with two non-amino acid binding ligands, as seen in Figure 4.9.

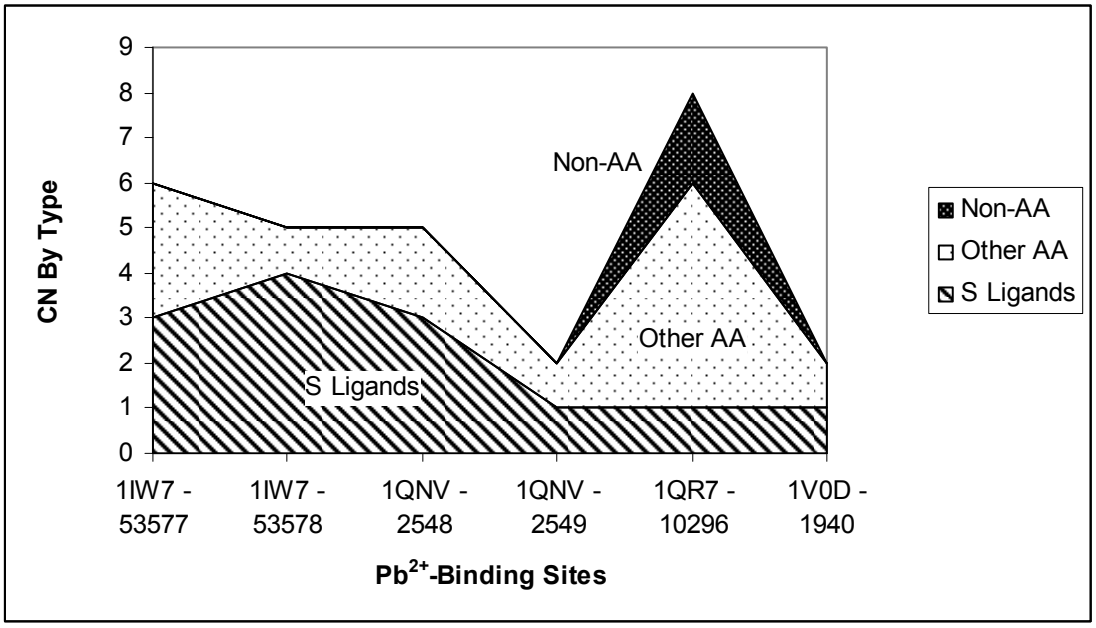


Figure 4.9: Distribution of CN values by type for six Pb^{2+}-binding sites containing sulfur ligands. Other AA represents ligand atoms other than sulfur from amino acids. Non-AA indicates oxygen ligands from sulfate or water.

4.3.2 Structural Analysis of Pb^{2+}-binding

Having summarized the data from a general statistical perspective, next we evaluated a specific instance where Pb^{2+} has replaced Ca^{2+}in a protein. Of the surveyed proteins, only one calcium-binding protein, Calmodulin (CaM), was identified where the Ca^{2+} ion had been replaced in the bound structure by Pb^{2+}. CaM is an EF-Hand protein with high sequence homology between species that includes two globular domains which

each bind two Ca^{2+} ions. The four binding pockets, which are all EF-Hand binding motifs, are identified as EF-I through EF-IV. For this study, Ca^{2+} -CaM (PDB file 1EXR) is compared with Pb^{2+}-CaM (PDB file 1N0Y): both from the species *Paramecium tetraurelia*.

Graphical representations rendered by Pymol for 1EXR and 1N0Y can be seen in Figure 4.10. Although the structures of Ca^{2+}-CaM and Pb^{2+}-CaM reported close agreement in the C-terminal domain [79], the N-terminal domain displayed significant structural changes, as seen in Figure 4.10. In addition to the obvious conformational changes, it can also be observed that Pb^{2+} ions have not only replaced Ca^{2+} ions in the binding pockets, but have also become bound to ligands outside of the binding pockets. In total, 14 Pb^{2+} ions are reportedly bound to 1N0Y. Since Pb^{2+} and Ca^{2+} have identical charges, and, as previously shown, similar ionic radii, it was hypothesized that their different acid classifications (Borderline for Pb^{2+} versus Hard for Ca^{2+}) could result in the preferential binding of different ligands when one metal replaced the other in a binding pocket, thus inducing unexpected or abnormal conformational changes.

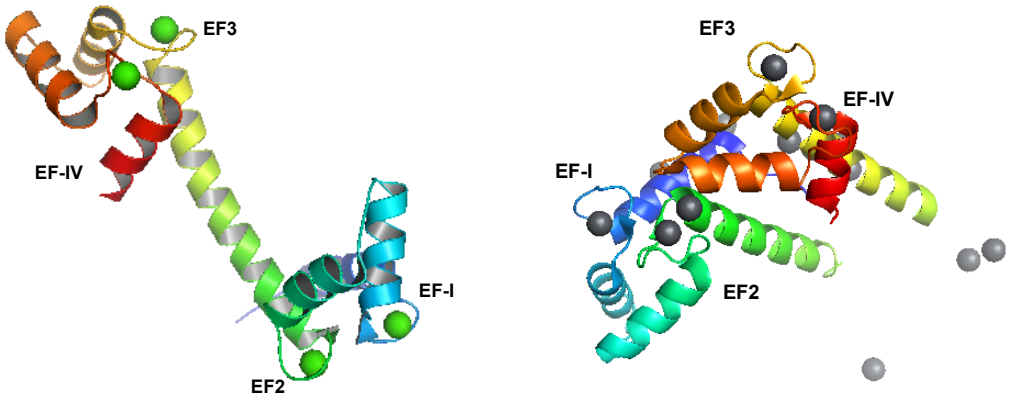


Figure 4.10: Proteins 1EXR (left), Ca-bound calmodulin, and Protein 1N0Y (right), Pb-bound calmodulin.

A multiple sequence analysis of 3CLN, 1EXR and 1N0Y using ClustalW was conducted (Figure 4.11) to establish sequence homology. 3CLN (rat CaM) was included as a control for the structural analysis. 1EXR (146 residues) and 1N0Y (148 residues), which have 100% homology, differ only in that 1N0Y has an additional Ala at the N-terminal, and a Lys at the C-terminal. Both were 88% identical to 3CLN (143 residues).

```
1exr_A     -EQLTEEQIAEFKEAFALFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGN 59
1N0Y_A     AEQLTEEQIAEFKEAFALFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGN 60
3cln_      ----TEEQIAEFKEAFSLFDKDGDGTITTKELGTVMRSLGQNPTEAELQDMINEVDADGN 56
              :************:*****************************************

1exr_A     GTIDFPEFLSLMARKMKEQDSEEELIEAFKVFDRDGNGLISAAELRHVMTNLGEKLTDDE 119
1N0Y_A     GTIDFPEFLSLMARKMKEQDSEEELIEAFKVFDRDGNGLISAAELRHVMTNLGEKLTDDE 120
3cln_      GTIDFPEFLTMMARKMKDTDSEEEIREAFRVFDKDGNGYISAAELRHVMTNLGEKLTDEE 116
           *********::******: *****: ***:***:**** ******************:*

1exr_A     VDEMIREADIDGDGHINYEEFVRMMVS- 146
1N0Y_A     VDEMIREADIDGDGHINYEEFVRMMVSK 148
3cln_      VDEMIREANIDGDGQVNYEEFVQMMTA- 143
           ********:*****::******:**.:
```

Figure 4.11: ClustalW (1.83) MSA (Multiple Sequence Alignment).

This analysis indicated 17 residue mutations between 3CLN and 1EXR, as seen in Table 4.5, the majority of which are conserved based on residue type. Five of these mutations occur within the Ca^{2+} -binding pockets: K94R, Y99L, N129D, Q135H, and V136I. Of these, three mutations are directly involved as binding ligands: Y99L, N129D, and Q135H. For mutations Y99L and Q135H, both the replaced and the new residues bind Ca^{2+} and Pb^{2+} with main-chain carbonyl oxygens, so no obvious binding changes are produced as a result of this. Mutation N129D replaces Asn in 3CLN with Asp in 1EXR and 1N0Y, however, the OD1 oxygen in each case is still shown to be the binding ligand atom.

Table 4.5: Residue Substitutions

3CLN Sequence Position (ClustalW)	3CLN Sequence Position (PDB)	3CLN	1EXR/1N0Y
[a]13	17	S	A
[a]66	70	T	S
[a]67	71	M	L
[a]74	78	D	E
75	79	T	Q
[a]81	85	I	L
82	86	R	I
[a]86	90	R	K
[a]90	94	K	R
95	99	Y	L
[a]115	119	E	D
[a]125	129	N	D
[a]131	135	Q	H
[a]132	136	V	I
[a]139	143	Q	R
[b]142	146	T	V
[a]143	147	A	S

[a]Conserved
[b]Semi-conserved

However, the Wilson paper also reported that sodium cacodylate was required in order to crystallize Pb^{2+} bound calmodulin [79]. Consequently, the presence of cacodylate in Pb^{2+} -CaM versus Ca^{2+} -CaM introduces an unknown variable that may have some effect both on the observed conformational change, as well as the coordination numbers in Pb^{2+} -CaM.

Despite these limitations it was possible to analyze comparative statistical characteristics between the binding sites found in 1EXR and 1N0Y. Data specifying the binding ligands for sites EF-I through EF-IV appear in Tables 4.6-4.9, respectively. Ligand atoms are highlighted in red. Columns in these tables are defined as follows: **ResSeqNbr** indicates residue sequence number; **Ca^{2+} LigRes** indicates ligand residue for calcium-binding; **Ca^{2+} LigAtom** is the ligand atom; **Ca^{2+} BindDist** is the distance between the Ca^{2+} ion and the ligand atom; **Ca^{2+} CLI Angle** is the carbon-ligand-ion

angle for the Ca^{2+} ion. The columns identified with Ca for 1EXR are repeated for Pb^{2+}. The difference in binding distance between Ca^{2+} and Pb^{2+} for identical ligands is reported in column **Bind Δ_{Dist}** and the difference in their respective carbon-ligand-ion angles reported in **CLI Δ_{Angle}**.

Table 4.6: Comparison of CaM EF-I binding site data between Ca^{2+} and Pb^{2+} ions.

EF1		1EXR				1N0Y					
Res Seq Nbr	Ca^{2+} Lig Res	Ca^{2+} Lig Atom	Ca^{2+} Bind Dist (Å)	Ca^{2+} CLI Angle (°)	Pb^{2+} Lig Res	Pb^{2+} Lig Atom	Pb^{2+} Bind Dist (Å)	Pb^{2+} CLI Angle (°)	Bind Δ_{Dist}	CLI Δ_{Angle}
20	ASP	OD1	2.31	144.80	ASP	OD1	2.30	146.18	-0.01	1.38
22	ASP	OD1	2.44	143.85	ASP	OD1	2.51	129.63	0.07	-14.22
24	ASP	OD1	2.35	133.89	ASP	OD1	2.62	117.80	0.27	-16.09
26	THR	O	2.35	155.06	THR	O	2.67	134.14	0.32	-20.92
26	THR				THR	OG1	3.49	130.31		
31	GLU	OE1	2.47	92.41	GLU	OE1	2.58	90.19	0.11	-2.22
31	GLU	OE2	2.48	92.02	GLU	OE2	2.54	92.06	0.06	0.04
								Mean	0.14	-8.67

ResSeqNbr indicates residue sequence number; **Ca^{2+} LigRes** indicates ligand residue for calcium-binding; **Ca^{2+} LigAtom** is the ligand atom; **Ca^{2+} BindDist** is the distance between the Ca^{2+} ion and the ligand atom, **Ca^{2+} CLI Angle** is the carbon-ligand-ion angle for the Ca^{2+} ion. Columns are repeated for Pb^{2+}. The difference in binding distance between Ca^{2+} and Pb^{2+} for identical ligands is reported in column **Bind Δ_{Dist}** and the difference in their respective carbon-ligand-ion angles reported in **CLI Δ_{Angle}**

Table 4.7: Comparison of CaM EF-II binding site data between Ca^{2+} and Pb^{2+} ions.

EF2		1EXR				1N0Y					
Res Seq Nbr	Ca^{2+} Lig Res	Ca^{2+} Lig Atom	Ca^{2+} Bind Dist (Å)	Ca^{2+} CLI Angle (°)	Pb^{2+} Lig Res	Pb^{2+} Lig Atom	Pb^{2+} Bind Dist (Å)	Pb^{2+} CLI Angle (°)	Bind Δ_{Dist}	CLI Δ_{Angle}
56	ASP	OD1	2.34	133.03	ASP	OD1	2.25	154.66	-0.09	21.63
56	ASP				ASP	OD2	4.18	45.25		
58	ASP	OD1	2.43	146.8	ASP	OD1	2.45	109.63	0.02	-37.17
58	ASP				ASP	OD2	3.13	77.27		
60	ASN	OD1	2.40	127.89	ASN	OD1	2.35	118.54	-0.05	-9.35
60	ASN				ASN	ND2	3.55	60.96		
62	THR	O	2.41	159.42	THR	O	2.52	153.61	0.11	-5.81
64	ASP	N	4.47	127.85	ASP	N	4.10	120.49		
64	ASP	OD2	4.71	87.15	ASP	OD2	3.34	112.38		
67	GLU	OE1	2.47	96.4	GLU	OE1	2.75	93.73	0.28	-2.67
67	GLU	OE2	2.60	89.39	GLU	OE2	2.82	90.49	0.22	1.1
								Mean	0.08	-5.4

ResSeqNbr indicates residue sequence number; **Ca^{2+} LigRes** indicates ligand residue for calcium-binding; **Ca^{2+} LigAtom** is the ligand atom; **Ca^{2+} BindDist** is the distance between the Ca^{2+} ion and the ligand atom, **Ca^{2+} CLI Angle** is the carbon-ligand-ion angle for the Ca^{2+} ion. Columns are repeated for Pb^{2+}. The difference in binding distance between Ca^{2+} and Pb^{2+} for identical ligands is reported in column **Bind Δ_{Dist}** and the difference in their respective carbon-ligand-ion angles reported in **CLI Δ_{Angle}**

Table 4.8: Comparison of CaM EF-III binding site data between Ca^{2+} and Pb^{2+} ions.

EF3		1EXR				1N0Y					
Res Seq Nbr	Ca^{2+} Lig Res	Ca^{2+} Lig Atom	Ca^{2+} Bind Dist (Å)	Ca^{2+} CLI Angle (°)	Pb^{2+} Lig Res	Pb^{2+} Lig Atom	Pb^{2+} Bind Dist (Å)	Pb^{2+} CLI Angle (°)	Bind Δ_{Dist}	CLI Δ_{Angle}
93	ASP	OD1	2.30	161.87	ASP	OD1	2.36	166.97	0.06	5.1
93	ASP	OD2	4.28	58.05	ASP	OD2	4.46	39.52		
95	ASP	OD1	2.33	130.85	ASP	OD1	2.35	122.79	0.02	-8.06
95	ASP	OD2	3.77	58.05	ASP	OD2	3.48	66.68		
97	ASN	OD1	2.42	132.81	ASN	OD1	2.36	136.53	-0.06	3.72
97	ASN	N	4.16	102.56	ASN	N	4.16	104.08		
97	ASN	ND2	4.08	51.62	ASN	ND2	4.01	51.7		
99	LEU	N	4.18	86.11	LEU	N	4.12	86.58		
99	LEU	O	2.28	162.46	LEU	O	2.28	157.24	0.00	-5.22
104	GLU	OE1	2.46	94.51	GLU	OE1	2.49	91.74	0.03	-2.77
104	GLU	OE2	2.54	89.99	GLU	OE2	2.54	89.17	0.00	-0.82
								Mean	0.01	-1.34

ResSeqNbr indicates residue sequence number; **Ca^{2+} LigRes** indicates ligand residue for calcium-binding; **Ca^{2+} LigAtom** is the ligand atom; **Ca^{2+} BindDist** is the distance between the Ca^{2+} ion and the ligand atom, **Ca^{2+} CLI Angle** is the carbon-ligand-ion angle for the Ca^{2+} ion. Columns are repeated for Pb^{2+}. The difference in binding distance between Ca^{2+} and Pb^{2+} for identical ligands is reported in column **Bind Δ_{Dist}** and the difference in their respective carbon-ligand-ion angles reported in **CLI Δ_{Angle}**

Table 4.9: Comparison of CaM EF-IV binding site data between Ca^{2+} and Pb^{2+} ions.

EF4		1EXR				1N0Y					
Res Seq Nbr	Ca^{2+} Lig Res	Ca^{2+} Lig Atom	Ca^{2+} Bind Dist (Å)	Ca^{2+} CLI Angle (°)	Pb^{2+} Lig Res	Pb^{2+} Lig Atom	Pb^{2+} Bind Dist (Å)	Pb^{2+} CLI Angle (°)	Bind Δ_{Dist}	CLI Δ_{Angle}
129	ASP	OD1	2.30	150.12	ASP	OD1	2.37	145.6	0.07	-4.52
131	ASP	OD1	2.34	123.46	ASP	OD1	2.47	110.61	0.13	-12.85
131	ASP	OD2	3.53	65.09	ASP	OD2	3.19	75.87		
133	ASP	OD1	2.38	127.08	ASP	OD1	2.25	134.04		
133	ASP	OD2	3.80	57.08	ASP	OD2	3.77	55.83		
135	HIS	O	2.35	149.88	HIS	O	2.52	133.14	0.17	-16.74
137	ASN	ND2	5.22	76.99	ASN	ND2	4.06	114.92		
140	GLU	OE1	2.44	96.06	GLU	OE1	2.47	100.58	0.03	4.52
140	GLU	OE2	2.55	91.05	GLU	OE2	2.78	86.05	0.23	-5
								Mean	0.13	-6.92

ResSeqNbr indicates residue sequence number; **Ca^{2+} LigRes** indicates ligand residue for calcium-binding; **Ca^{2+} LigAtom** is the ligand atom; **Ca^{2+} BindDist** is the distance between the Ca^{2+} ion and the ligand atom, **Ca^{2+} CLI Angle** is the carbon-ligand-ion angle for the Ca^{2+} ion. Columns are repeated for Pb^{2+}. The difference in binding distance between Ca^{2+} and Pb^{2+} for identical ligands is reported in column **Bind Δ_{Dist}** and the difference in their respective carbon-ligand-ion angles reported in **CLI Δ_{Angle}**

Distance and angle comparisons were evaluated only for binding ligands utilized by both ions, so in cases where Pb^{2+} appeared to have additional binding ligands, these values were not compared. Negative values appearing in columns for Δ BindDist and Δ C-Lig-Ion Angle (Tables 4.6-4.9) indicate shorter bond lengths and smaller angles for Pb^{2+} relative to Ca^{2+}, respectively. To visually compare the individual binding sites, models were built to indicate the binding ligands for Pb^{2+} and Ca^{2+} (Figure 4.12), and Table 4.10 summarizes all angle and distance data compiled between sites for the two proteins analyzed. In Figure 4.12, , the paired figures (AB, CD, EF, GH) correspond to EF sites I-IV for Ca^{2+}-binding (left) and Pb^{2+}-binding (right).

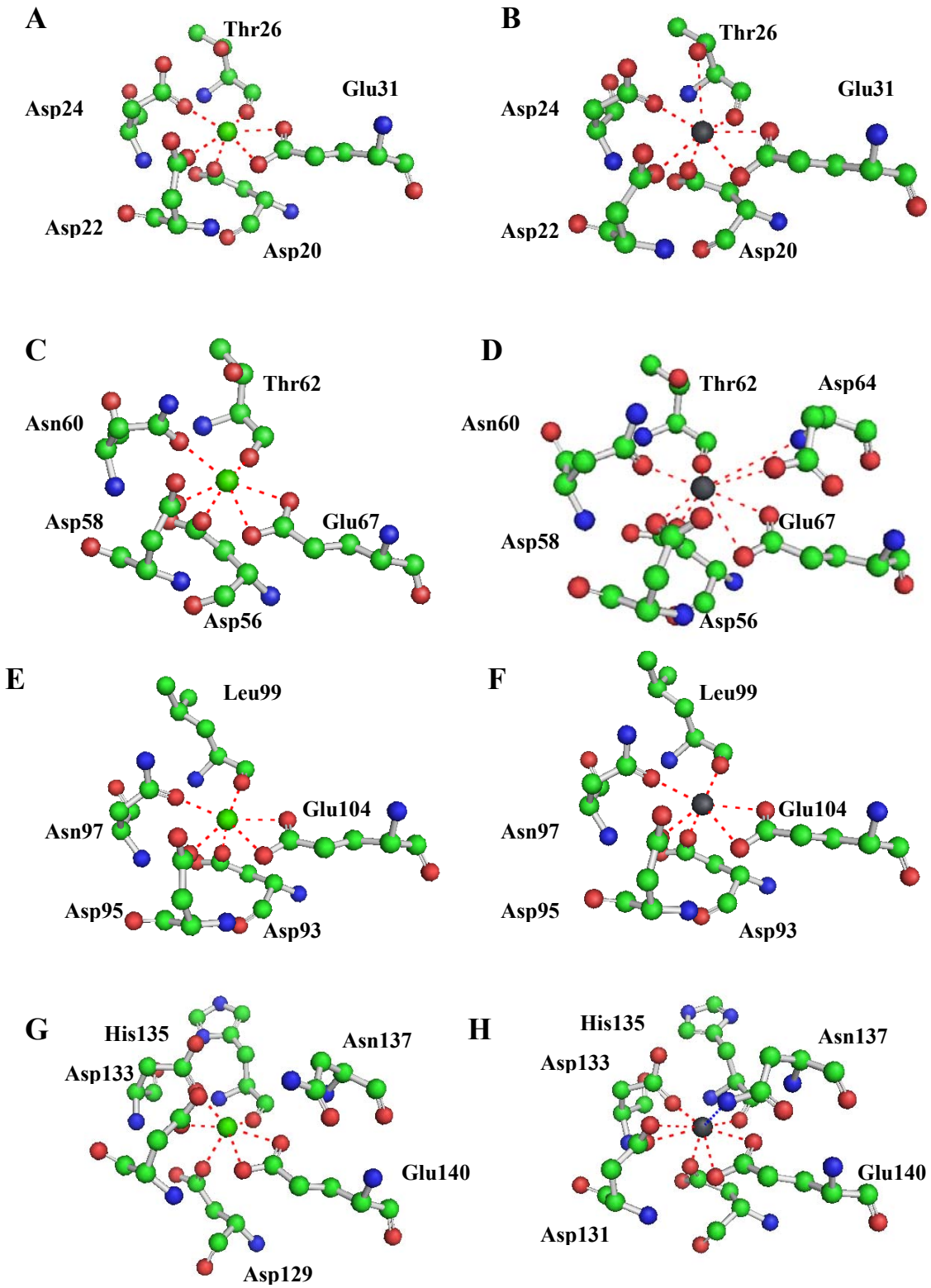


Figure 4.12: Binding site geometries for PDB proteins 1EXR (left) and 1N0Y (right).

A comparison of the models presented in Figure 4.12 indicates that the basic geometry for binding of Ca^{2+} is conserved upon binding with Pb^{2+}. Some rotation of the side-chains is evident in the net increase in binding distance and net decrease in bond angles, as summarized in Table 4.10, when binding Pb^{2+}, as well as the increase observed in the bidentate ligand dihedral (φ) angle for the Pb^{2+} ion (19.92 ± 5.93 º vs. 4.17 ± 2.98 º for Ca^{2+}). All of these changes are likely due to the larger ionic radius for Pb^{2+}, and are consistent with the mean binding distance values obtained from the database analysis.

Table 4.10: Summary of angle/distance values for 1EXR and 1N0Y.

	EF1	EF2	EF3	EF4	Mean
1EXR Nbr Ca^{2+} Ligands	6	6	6	6	6
1N0Y Nbr Pb^{2+} Ligands	7	8	6	7	7
Mean sum Δ Bind Dist Ca→ Pb (Å)	0.14	0.08	0.01	0.13	0.09
Mean sum Δ CLI Angle Ca→ Pb (°)	-8.67	-5.38	-1.34	-6.92	-5.58
Ca^{2+} planar lig-ion-lig angle sum (°)	365.58	363.67	360.57	367.39	364.30
Ca^{2+} planar lig-ion-lig % Variance from 360	1.55	1.02	0.16	2.05	1.20
Pb^{2+} planar lig-ion-lig angle sum (°)	366.65	349.69	360.23	363.79	360.09
Pb^{2+} planar lig-ion-lig % Variance from 360	1.85	2.86	0.06	1.05	1.46
Pb^{2+} planar lig-ion-lig % Variance from Ca	0.29	3.84	0.09	0.98	1.30
Ca^{2+} planar monodentate lig-ion-lig angles (°)					78.00 ± 2.37
Pb^{2+} planar monodentate lig-ion-lig angles (°)					77.64 ± 4.57
Ca^{2+} planar bidentate lig-ion-lig angles (°)					52.29 ± 0.63
Pb^{2+} planar bidentate lig-ion-lig angles (°)					49.52 ± 2.33
Ca^{2+} Ion Dihedral (°)	4.91	4.12	7.43	0.23	4.17 ± 2.98
Pb^{2+} Ion Dihedral (°)	20.76	26.03	21.10	11.79	19.92 ± 5.93
Distance of Ca^{2+} ion out of bidentate plane	0.24	0.21	0.37	0.01	0.21 ± 0.15
Distance of Pb^{2+} ion out of bidentate plane	1.01	1.36	1.02	0.61	1.00 ± 0.31

More significantly, the expected selectivity for nitrogen ions, based on the borderline acid classification of Pb^{2+}, was not observed. In EF-II the Asp64 N atom rotates closer to the Pb^{2+} ion (Figure 4.12(D)). The same is observed in EF-IV (Table 4.9 and Figure 4.12(H)), where the ND1 atom of Asn137 rotates closer to the Pb^{2+} ion. There are two possible reasons for this. First, the ND2 atom is actually involved as a binding ligand and, as seen in Figure 4.13, provide additional charge to the HOH54 water oxygen and the Asp131 OD2 oxygen superior to the pentagonal plane. The second possibility is that the ND2 OD1 ligand exchange is accidental.

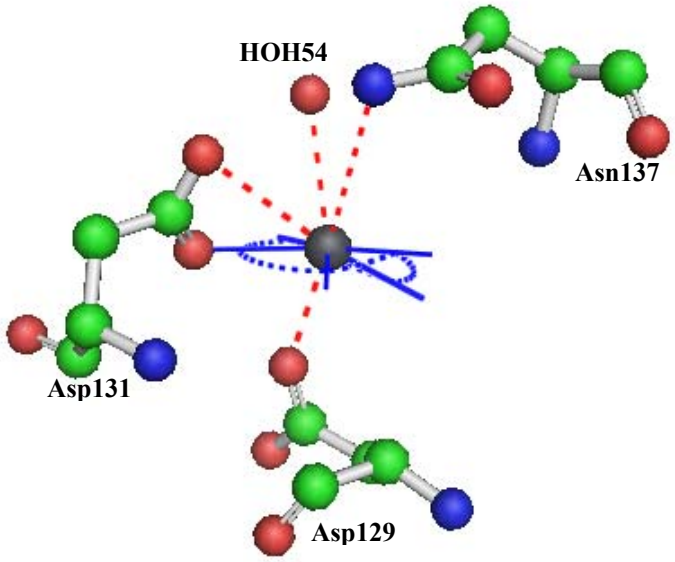


Figure 4.13: Extraplanar ligands for binding of Pb^{2+} in CaM EF-IV. The pentagonal plane around the Pb^{2+} ion is highlighted in blue.

Conversely the Asn nitrogen atoms for Asn60 and Asn97 (Figure 4.12(D) and Figure 4.12(F), respectively) are not identified as binding ligands. Despite the fact that there are no apparent barriers restricting rotation of the amide group around the Cγ-Cβ bond, it is the OD1 atom in each case that appears to be the binding ligand, not the ND2 atom which would have been expected based on the putative selectivity of Pb^{2+} for nitrogen.

Table 4.6 and Figure 4.12(B) indicate one additional ligand associated with Pb^{2+}-binding in EF-I, the Thr26 OG1 oxygen which resides 3.49 Å from the Pb^{2+} ion. Table 4.7 and Figure 4.12(D) indicate rotation of the Asp64 residue, so that the N and OD2 atoms from 4.47 Å and 4.71 Å are positioned 4.10 Å and 3.34 Å away from the Pb^{2+} ion, respectively. Data presented in Table 4.8 and Figure 4.12(F) are significant in that they indicate no essential change in the geometry or coordination resulting from the replacement of Ca^{2+} by Pb^{2+}.

For the four EF-Hand binding sites surveyed, mean values of 6 and 7 binding ligands (excluding water) were found related to the binding of Ca^{2+} and Pb^{2+} respectively (Table 4.10). A small, mean net increase in the binding distance (0.09 Å) between ligands for Ca^{2+} and Pb^{2+} was observed, as well as a mean sum decrease in the CLI angle (-5.58 º), which is consistent with the larger radius for Pb^{2+}.

The data associated with the bidentate dihedral angles for calcium-binding, as summarized in Table 4.10, correlate well with research reported by Glusker [4] who indicated a mean distance of 0.4 Å for the Ca^{2+} ion out of the carboxlyate plane, with a φ angle under 30°. For Pb^{2+}, however, the mean average distance for the lead ion increases (in Å) to 1.00 ± 0.31 from 0.31 ± 0.19 observed for Ca^{2+}, and the mean bidentate ligand

dihedral (φ) angle was much larger for the Pb^{2+} ion (19.92 ± 5.93 º) than for Ca^{2+} (4.17 ± 2.98 º) (Figure 4.14). Again, this is probably due to the larger ionic radius for Pb^{2+}.

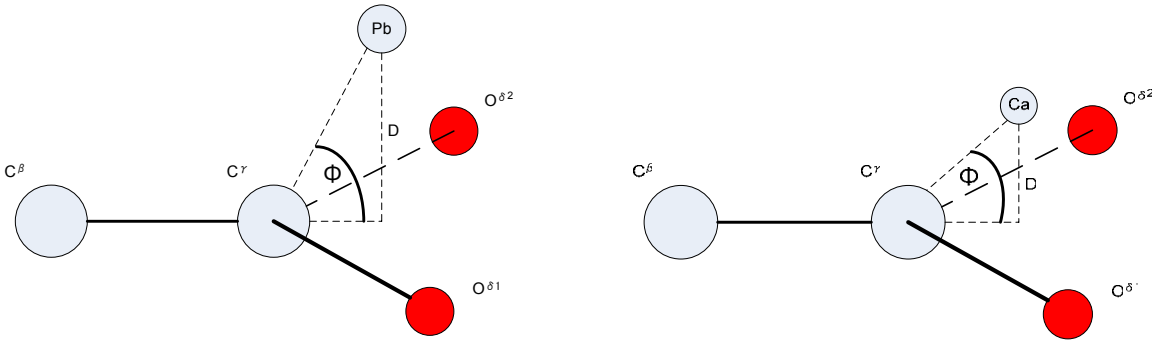


Figure 4.14: Position of the Pb^{2+} ion (left) and the Ca^{2+} (right) with respect to the bidentate plane.

The planar **lig-ion-lig angle sum (°)** for calcium and lead measured the angle sums for the ligand-ion-ligand angles in the pentagonal plane. In a perfect plane, the angle sum would equal 360 degrees. The summarized values in Table 4.10 were evaluated based on the premise that distortion resulting from the larger ionic radius of lead would be indicated by higher variance from the plane surrounding calcium. Interestingly the % planar variance surrounding Pb^{2+} was only 1.30% different from Ca^{2+}, and 1.46% relative to 360 degrees. In contrast, the mean % variance for Ca^{2+} was 1.20% out of 360 degrees. The monodentate and bidentate carbon-ligand-ion angles, as summarized in Table 4.10, were nearly identical, and in both cases, the mean differences fell within the range of the standard deviation, so they are in all probability not statistically different.

A comparison of ligands identified in Tables 4.6-4.9 indicate a number of putative ligand atoms associated with Pb^{2+}-binding that meet the 4.2 Å cutoff distance, but were not designated as binding ligands (e.g. – Asp58 OD2, Asp95 OD2, etc). In most of these

cases, the **Pb^{2+} CLI Angle** values were less than 90°. Two of these instances (97 Asn N and 99 Leu N) were excluded because they were very close to the cutoff distance for Pb^{2+} (4.16 and 4.12 Å, respectively), but more importantly, there was virtually no change in their distance from Ca^{2+} to Pb^{2+}, and it was assumed that binding should result in a more apparent spatial change relative to the ion. For comparison, Asn137 ND2, which was identified as a probable binding ligand atom (highlighted in blue in Table 4.9 and Figure 4.12(H)) actually moves 1.16 Å closer to the Pb^{2+} ion than in the calcium-bound protein. While it is possible that this isolated example is coincidental, it seems unlikely that the electrostatic force associated with binding would not result in a decrease in the distance between the two atoms. For the remaining atoms that met the cutoff distance requirement but were excluded due to the short bond angles, these cases involved carboxyl or amide groups, where one atom (N or O) was closer to the ion than the other. In these instances, the electrons would shift in the conjugated π-bond system away from the more distant atom, leaving essentially a single, more proximate atom with a higher electron density involved in the binding process, as seen in Figure 4.15.

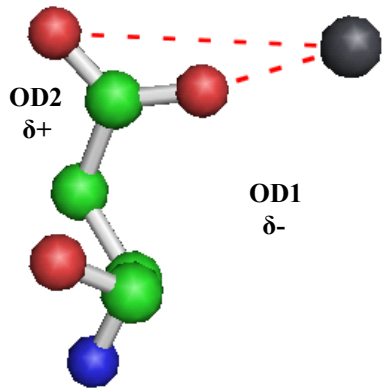


Figure 4.15: Relationship between orientation of Asp58 from 1N0Y and charge density.

The observation that the Pb^{2+} ion does not produce greater distortion – either within the pentagonal plane or with the main chain – was surprising. Figure 4.16 illustrates the differences between the two ions during binding. The Pb^{2+} ion sits higher out of the binding pocket in the pentagonal plane (Figure 4.16(C)), with less observable distortion in the pentagonal plane (Figure 4.16 (A, B)). The increase in distance between the ligand oxygen and the Pb^{2+} compared to calcium is compensated for by the C-Lig-Pb angle. Rather than resulting in main chain conformational changes, the side chain conformations absorb the changes associated with the ion displacement, which is more favorable from a thermodynamic perspective.

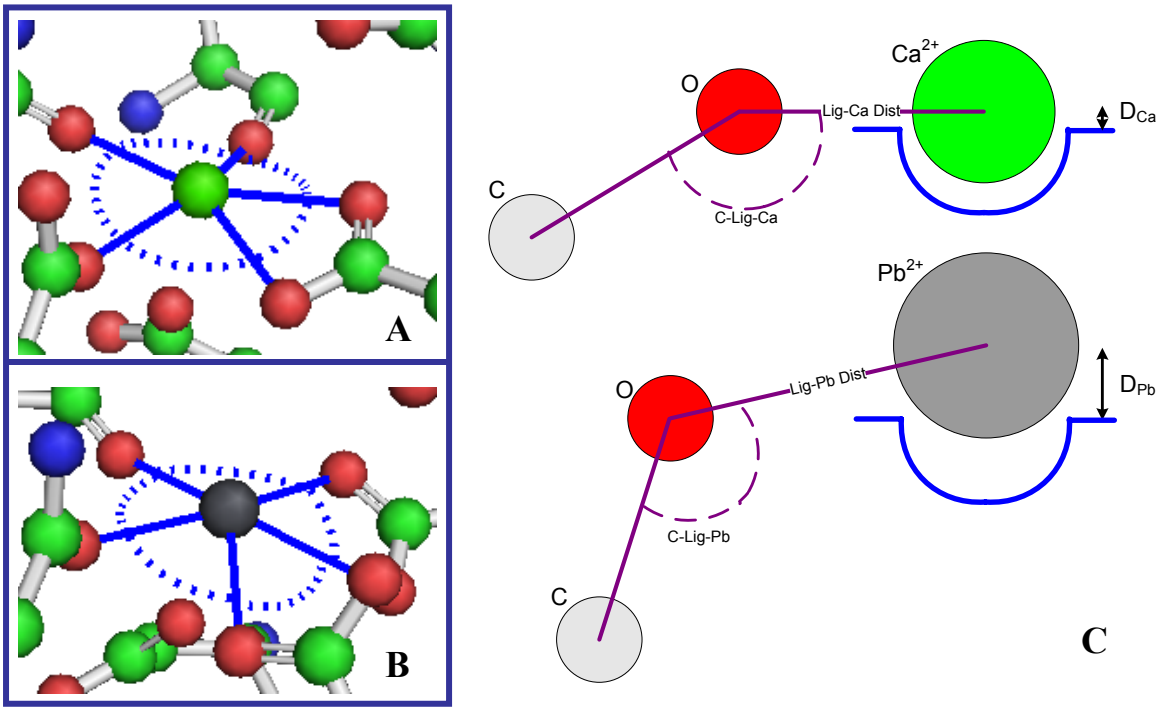


Figure 4.16: (A) Positions of Ca^{2+} ion and (B) Pb^{2+} ion with respect to pentagonal plane. (C) Comparison of Ca^{2+}- and Pb^{2+}-binding characteristics with respect to the pentagonal plane.

Because the structural analysis of the four EF-Hand binding sites between 1EXR and 1N0Y did not support the premise that Pb^{2+} would exhibit a ligand preference for nitrogen, the database was queried specifically for residues Asn and Gln containing side-chain atoms within 4.2 Å of the metal ion. A summary of this query is presented in Table 4.11.

Table 4.11: Potential binding atoms associated with asparagine, glutamine residues.

HetAtm Serial Number	PDB_ID	Atom Serial Number	Atom Character ID	Residue	Element	Binding Distance (Å)
4341	1E9N	3483	OD1	ASN	O	3.77
4341	1E9N	3484	ND2	ASN	N	3.09
4341	1E9N	2368	ND2	ASN	N	3.98
1315	1N0Y	458	OD1	ASN	O	2.35
1315	1N0Y	459	ND2	ASN	N	3.55
1322	1N0Y	1224	ND2	ASN	N	4.06
1324	1N0Y	918	OD1	ASN	O	2.36
1324	1N0Y	919	ND2	ASN	N	4.01
1976	1NA0	1404	NE2	GLN	N	3.04
2705	2FP1	1370	OE1	GLN	O	3.02
				Mean		3.32 ± 0.65
				Mean Nitrogen		3.62 ± 0.47
				Mean Oxygen		2.87 ± 0.67

A total of ten atoms (6 N, 4 O) meeting the distance criteria was found: eight from Asn and two from Gln, representing a total of only seven amino acid residues. Three of the six residues were associated with EF binding pockets in 1N0Y: Asn60 from EF-II, Asn97 from EF-III, and Asn137 from EF-IV. However, with the exception of the ND2 nitrogen from Asn137, the other nitrogen atoms in the EF binding sites for 1N0Y were shown to be unlikely binding ligands. Based on the summary analyses of all structural data associated with Pb^{2+}-binding in proteins, there is nothing to suggest that Pb^{2+} ions preferentially bind with nitrogen. In fact, Pb^{2+} ions appear more likely to bind with sulfur or oxygen, rather than nitrogen.

4.4 Conclusions

4.4.1 Statistical Analysis of Pb^{2+}-binding Characteristics

Construction of a localized database utilizing commonly available software (MS Access, MS Excel) enhances statistical analysis because it allows the researcher to define custom queries with respect to available data, an option not available with structured online databases. It also allows for subsequent expansion to include research-specific data that can be filtered to exclude non-relevant data.

In this study, several tentative conclusions can be drawn from this statistics compiled. The Pb-O, Pb-S and Pb-N mean distances found to be 2.87 ± 0.57 Å 3.16 ± 0.27 Å and 3.47 ± 0.71 Å, respectively, were consistent with previously reported ranges associated with Pb^{2+}-binding [98], however, the calculated values for coordination number were surprising. A mean CN of 3.9 was determined, ranging from 1-11 ligands, with 2 ligands dominating the binding schemes (Figure 4.6). Additionally, 75% of the surveyed ions had a coordination number between 2 to 5, with only 17.0% indicating a coordination number in the 6 to 9 range. This suggests two important characteristics. First, the binding of Pb^{2+} is more closely related to that of Zn^{2+} (5 ligands) than for Ca^{2+} (6 to 8 ligands). Second, and more importantly, the binding of Pb^{2+} does not appear to require a regular structure, and will more readily to bind to sites on a protein where other metals will not bind. A caveat to this should be noted, that because these data were compiled from the crystal structures of proteins, it is possible that some of the observed binding is an artifact of the crystallization process. Nonetheless, it suggests that further scrutiny of these phenomena in a solution environment should provide valuable data on the nature and mechanism of Pb^{2+}-binding.

With respect to the metal classification of Pb^{2+} as a borderline acid with an expected preference for nitrogen ligands, the summarized ligand data indicated that over 70% of the probable binding ligands were side-chain oxygen atoms, mostly originating from Glu residues (Figure 4.5). Of the 15% of ligands that were nitrogen, 4.8% were from main-chain nitrogen, which suggests, based on the longer Pb-N binding distance, that when nitrogen from the main-chain is involved, it may result in larger conformational changes. More significantly, however, this data suggests that there is no statistical basis for the assumption that Pb^{2+} binds preferentially to nitrogen ligands. In fact, this supposition was further challenged in the structural analysis reported in this study.

For binding sites evaluated where sulfur was present as a binding ligand, the reported CN values of 2, 5 and 6 were inconsistent with previously reported work indicating a trigonal-pyramidal geometry during displacement of Zn^{2+} by Pb^{2+}, with three coordinating ligands, but were consistent with the alternative coordination of 5-8 ligands, although it should be noted that only six Pb^{2+}-binding sites which included sulfur ligands were retained in this study, and none of those were identical to the proteins cited in Magyar and Andersen [74, 75].

4.4.2 Structural Analysis of Ca^{2+} displacement by Pb^{2+} in CaM

Two key points were elucidated in the structural analysis. First, replacement of Ca^{2+} by Pb^{2+} in CaM did not result in significant structural changes in the proteins binding site geometry. Although some rotation of the side-chains is evident in the net increase in binding distance and net decrease in bond angles, and there is an observed

increase in the bidentate ligand dihedral (φ) angle for the Pb^{2+} ion (19.92 ± 5.93 º vs. 4.17 ± 2.98 º for Ca^{2+}), these changes can be explained by adaptation to a metal ion with a larger ionic radius. This suggests that displacement can occur without disrupting the structural geometry. Also, since the ligands utilized by Ca^{2+} were also utilized by Pb^{2+}, with the addition of 1 or more supplemental ligands, these data do not indicate that additional ligands are necessary for binding, but may be utilized when available.

Second, the putative preference by Pb^{2+} for nitrogen was not observed in the structural analysis, which is consistent with the data results from the database analysis. This suggests that the borderline acid definition for Pb^{2+} should not be over generalized.

Given the summaries presented in this study, it seems apparent that Pb^{2+} does not require a fixed geometry for binding, and can bind with available ligand types (O, N, S) more readily than metals which serve as cofactors to confer function in specialized proteins, resulting in a more promiscuous ion: a property which no doubt contributes to its toxicity.

5 Use of Engineered Fluorescent Proteins with Grafted Lone metal-binding Sites to Characterize Metal-binding Properties

5.1 Introduction

Green Fluorescent Protein (GFP) is a protein found in the jellyfish *Aequorea victoria*. Wild-type GFP contains 238 residues, organized into an 11 B-sheet barrel-like structure, with a monomeric tertiary structure diameter of 24 Å, and a height of 42 Å [29] (Figure 5.1). The matured chromophore in GFP (Figure 5.2) comprising the conjugated π-bond system responsible for fluorescence is formed as a result of intramolecular cyclization of 65Ser-Tyr-Gly67 [29, 99].

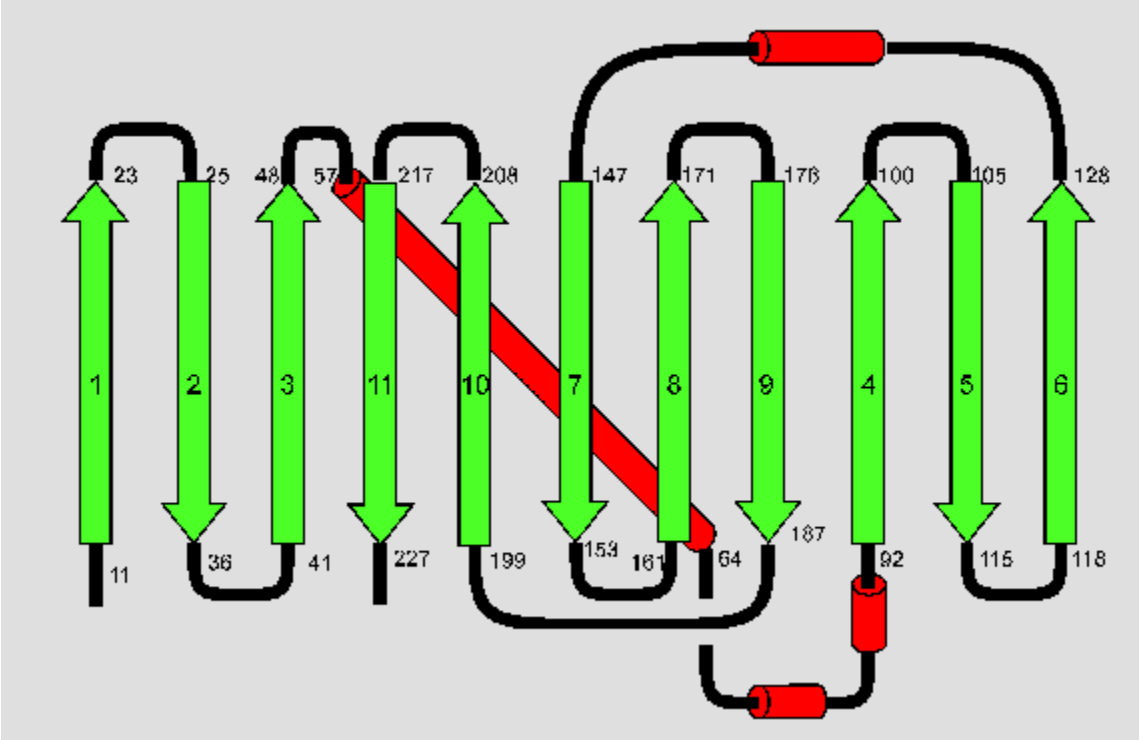


Figure 5.1: Secondary structure of GFP from http://www.conncoll.edu/ccacad/zimmer/GFP-ww/GFP3.htm.

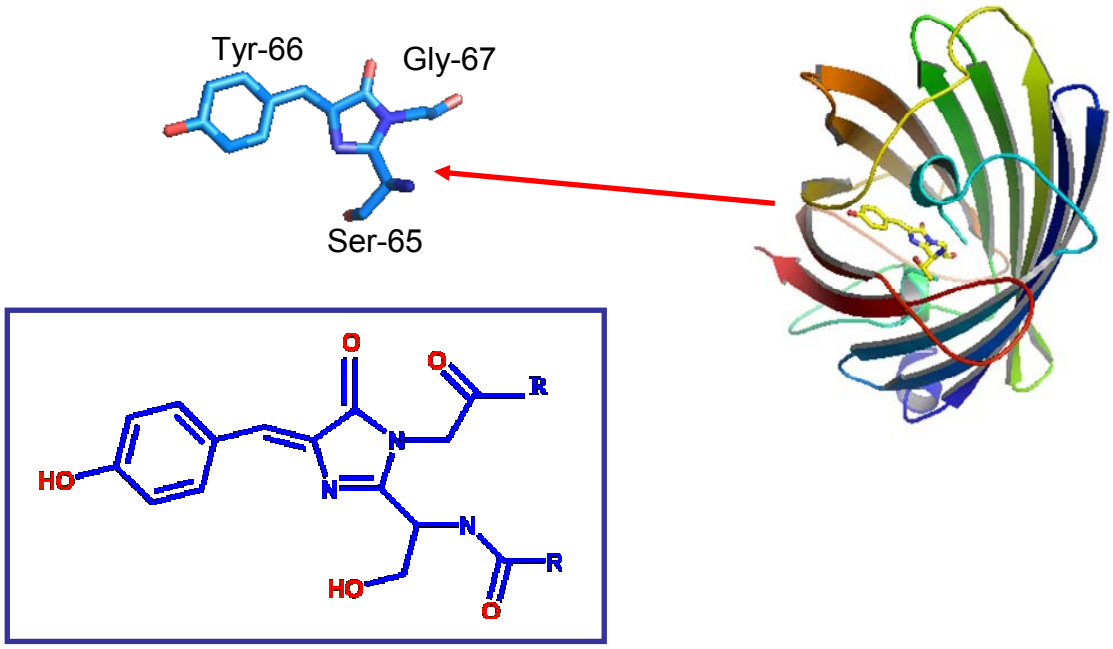


Figure 5.2: Expanded view of the mature GFP chromophore From Protein Data Bank file 1GFL.pdb. Inset panel reproduced from http://www.conncoll.edu/ccacad/zimmer/GFP-ww/GFP3.htm.

This autocyclization mechanism, as described in Figure 5.3, requires only oxygen to proceed, so that GFP requires no additional cofactors. The chromophore has a cis-geometry, and is located in the center of the barrel where it is protected from oxygen-quenching and/or attack by hydronium ions. Polar residues and H_2O form a hydrogen bonding network around the chromophore, which likely contributes to the stability of the chromophore [29].

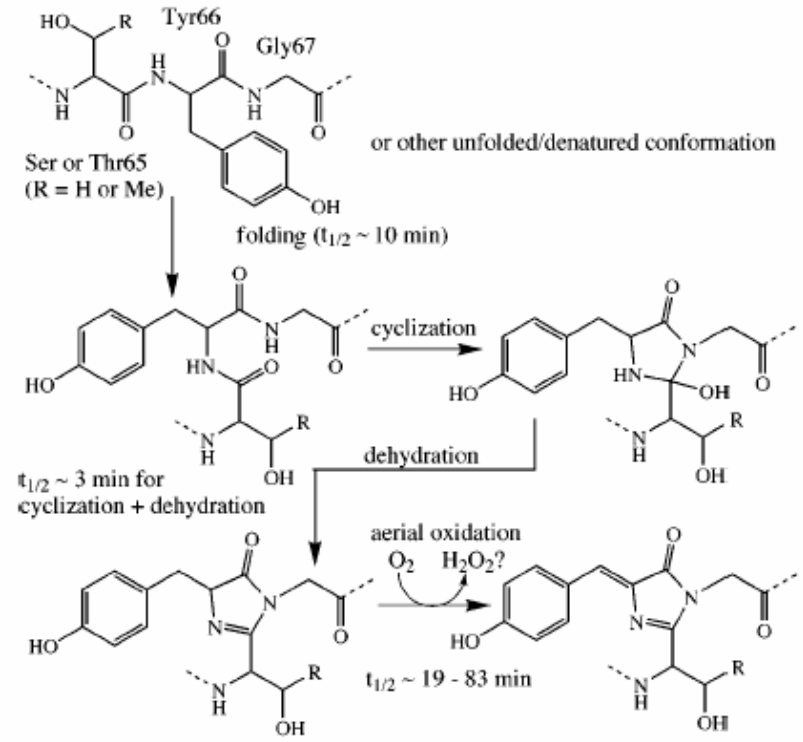


Figure 5.3: Mechanism proposed by Cubitt *et al.* [100] for the intramolecular biosynthesis of the GFP chromophore, with rate constants estimated for the Ser65Thr mutant by Reid & Flynn [101] and Heim *et al.* [102]. Figure reprinted from Tsien [29]

The relevance of GFP as an analytical tool became apparent after it was expressed in other organisms and it was found that the gene contains all of the information necessary for post-translational synthesis of the chromophore, therefore no contributions in the form of substrates or cofactors were required from the host organism [29]. Additionally, it was found that GFP: 1) could be fused to other proteins without affecting their function, 2) is generally non-toxic, and 3) is resistant to various deleterious conditions such as heat, alkaline pH, detergents, photobleaching, chaotropic salts, organic salts and many proteases [103].

Despite GFP's intrinsic properties, the scope of its versatility has been limited in biological studies due to several factors, mostly related to formation of the chromophore

[103-105]. In wild-type GFP, the length of time required for chromophore formation is ~3 hours. This is further hindered by insufficient O_2, and the higher temperature (37 °C) required for physiological conditions than that of the colder-bodied organism (*A. Victoria*) from which it is produced [103]. Additionally, the fluorescence of GFP was difficult to distinguish from background fluorescence if very low concentrations of GFP were observed [103].

To address these limitations, a number of mutant variants have been developed. Enhanced Green Fluorescent Protein (EGFP) is a GFP variant specifically engineered for accelerated fluorophore formation and increased intensity in the UV-Vis absorbance peak. Mutagenesis studies reported by Tsien indicated that the mutation F64L was found to facilitate correct folding at 37 °C in mammalian cells [29, 103, 106]. Additionally, the speed of fluorophore formation was accelerated from 2 hours to 0.45 hours following the mutation S65T [102, 106]. This latter mutation, or a substitution of Ala, Gly, Cys or Leu at residue 65, resulted in conversion of the two major absorbance peaks of wild-type GFP into a single, more intense absorbance peak, variously reported at 488 nm or 489 nm [29, 102, 106]. These mutations, along with H231L, were combined to form EGFP. The enhanced brightness of EGFP and greater stability allowed for better expression in mammalian cells and extended visualization of tagged proteins in cells with low light intensities, with very little fluorescence loss due to photobleaching.

Our current work utilizes a variant of EGFP provided by Dr. Frey's Research laboratory at Georgia State University, referenced in this thesis as EGFPwtF. In addition to the standard EGFP mutations F64L, S65T and H231L, EGFPwtF includes mutations

V21L and M218I. Additionally, a variant designated C2, which included mutations M153T and V163A, was developed for improved folding at 37 °C.

Using EGFPwtF as a scaffold, a number of variants have been produced by grafting the calcium-binding loop III (sequence DKDGNGYISAAE) and its flanking E (sequence EEEIREAFRVF) and F (sequence LRHVMTNL) helices from calmodulin onto three targeted, solvent-exposed loops on EGFPwtF, designated herein as A, B and C (Table 5.1). This grafting approach has been used to evaluate EF-hand Ca^{2+}-binding motifs [18, 20], and previous work in our laboratory has demonstrated that Ca^{2+} binding motifs maintain their binding properties when grafted onto non-Ca^{2+}-binding proteins [19, 21].

Table 5.1: Summary of EGFPwtF and C2 variant mutations.

Variant
EGFP-5G-III-5G
EGFP-B-1
EGFP-B-2
EGFP-B-3
EGFP-B-4
EGFP-B-EFM
EGFP-C-1
EGFP-C2
EGFP-C-2
EGFP-C-3
EGFP-C-4
EGFP-C-EFM
EGFPwtF

The E-III-F segments were variously combined as indicated in Figure 5.4, and will be referenced in this work as sub-variant types 1-4.

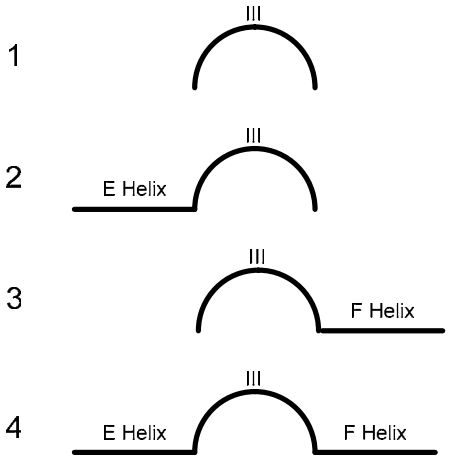


Figure 5.4: E-III-F segments grafted onto protein scaffold. Sub-variants are: (1) loop III only; (2) E-Helix and loop III; (3) loop III and F-Helix, and (4) the intact E-III-F binding site.

The addition of a metal chelating loop adjacent to the natural chromophore present in EGFPwtF provides a valuable construct for spectroscopic analysis, as binding of a metal ion in strategically-selected positions relative to the protein chromophores is expected to produce conformational changes that can alter the spectral intensity of the EGFPwtF chromophore either by quenching or enhancing the emission and absorbance spectra. The scaffold protein has the potential for use as a luminescent marker for cellular studies with respect to monitoring of contrast agents, determination of protease activity, and possibly chelation of metals for bioremediation.

5.2 Materials and Methods

5.2.1 Protein Expression and Purification of EGFPwtF and Variants

The EGFPwtF protein and designed variants utilized for metal-binding and protease studies were developed via sub-cloning through polymerase chain reaction (PCR). Proteins were prepared for subsequent purification on a Ni^{2+} chelating sepharose

column by addition of a 6x His-tag. These variants provide the scaffold for mutagenesis studies aimed towards designing proteins with high metal selectivity, and for development of a protease sensor. EMD Omnipur tris (hydroxymethyl)aminoethane (EMD Chemicals, Inc., Gibbstown, NJ), or TRIS, was utilized extensively as a buffering agent to maintain pH for the expressed proteins. Luria-Bertani media were prepared in a 2.8 L Erlenmeyer flask by combining 10 g Bacto-Tryptone (Becton, Dickinson and Co., Sparks, MD) with 5 g Bactone-yeast extract (EMD Chemicals, Inc., Gibbstown, NJ), 10 g NaCl, and then filling the flask to 1 L with ddH_2O. The pH was adjusted to 7.0 with 5 M NaOH (J.T. Baker, Phillipsburg, NJ).

5.2.1.1 Transformation

The pET28a vector used for coding EGFPwtF and all variants was transformed into *Escherichia coli* cell strain DE3, and grown on Agarose plates. During preparation, 50 μL of the appropriate cell strain were added to an autoclaved microcuvette, followed by 0.5 μL of DNA. Samples were incubated on ice for 30 min. Subsequent to incubation, the sample was subjected to heat shock for 90 s at 42 °C to allow DNA into the cell. The sample was placed on ice for 2 minutes. After cooling the sample, 50 μL of LB Medium were added, and the sample was placed in an incubator for 30 minutes at 37 °C. Culture plates were labeled and dated. Steel coils were heated in an open flame and immersed in EtOH several times for sterilization. The cell culture was then added in drops onto the agarose plate, and spread across the surface with the sterilized coil. The plate was then covered and placed in an incubator overnight at 37 °C.

5.2.1.2 Inoculation

20 mL of LB medium, pH 7.0, were transferred by pipet into a 50 mL disposable centrifuge tube, followed by 12 μL of 50 mg/mL kanamycin, for a final concentration of 0.03 mg/mL kanamycin. A cell colony was scraped from the agarose plate using an inoculation loop, and transferred to the LB media in the centrifuge tube. The sealed tube was then placed in a large beaker and packed with paper towels to prevent movement of the tube. The beaker was then placed in an incubator-shaker overnight at 37 °C.

5.2.1.3 Expression

In 1 L of autoclaved LB medium, 600 μL of 50 mg/mL kanamycin was added for a final concentration of 0.03 mg/mL. Optical density of the cell cultures was monitored using a Shimadzu UV-1601 PharmaSpec UV-Vis spectrophotometer with UV Probe software (Shimadzu North America, Columbia, MD).

Samples for the spectrophotometer were prepared in 1.0 mL plastic, disposable cuvettes. Two reference cuvettes were prepared for the baseline using 1.0 mL of the LB medium/kanamycin. Using the Bunsen burner, the neck of the 2.8 L Erlenmeyer flask was rotated in the flame to prevent bacterial growth. Next, the cell culture in the 50 mL disposable centrifuge tube was poured into the 2.8 L flask. The flask was covered with Aluminum foil, and secured in the incubator-shaker set at 200 rpm, 37 °C. The optical density of the sample was checked in the UV-Vis spectrophotometer until the absorbance reached 0.6 ± 0.1, at 600 nm. This range was previously determined for optimal induction. At the appropriate absorbance, 200 μL of Isopropyl-beta-D-thiogalactopyranoside (IPTG) were added to induce expression of the protein, for a final

IPTG concentration of 0.2 mM, and the temperature reduced to 20-25 °C, for optimal expression. Following induction, 1.0 mL samples were removed every hour for three hours, followed by a final sample on the following day, to evaluate protein expression using SDS-PAGE gels. Cell pellets were harvested the following day by centrifugation, and stored in a freezer at 4 °C until they could be purified.

5.2.1.4 Purification

To the collected cell pellet, ~20 mL of extraction buffer (20 mM TRIS, 100 mM NaCl, 0.1% Triton x-100) was added, and the sample vortexed to dissolve. The dissolved cell pellet was poured into a 50 mL plastic beaker, and the beaker placed on ice. The sample was then sonicated six times to break the cell membranes, for 30 s periods, with ~5 min intervals between sonications. Following sonication, the cell pellet solution was centrifuged for 20 min at 53, 442 x *g* to separate the protein into the supernatant. The extracted supernatant was filtered with 0.45 μm pore size filter (Whatman, Florham Park, NJ) into a 50 mL plastic centrifuge tube. Concentrated solutions were diluted with the appropriate binding buffer prior to injection into the FPLC system.

Purification of EGFPwtF and variants was completed using an Aktaprime FPLC (Amersham Biosciences, Piscataway, NJ) equipped with a UV detector and a 280 nm optical filter. Preparation of the FPLC required rinsing of both pumps A and B with 10 mL ddH_2O, twice each.

Two different columns were utilized. For most purifications, a Hitrap 5 mL HP Chelating sepharose column was used. The binding Buffer A was comprised of 1 M

K_2HPO_4, 1 M KH_2PO_4, 250 mM NaCl, pH 7.4. The elution Buffer B was comprised of Buffer A and 0.5 M imidazole.

The column was first rinsed with EDTA (Acros Organics, Geel, Belgium) solution (100 mM EDTA, 1 M NaCl, pH 8.0) to remove any metals, followed by ddH_2O. Following the EDTA rinsing step, the column was washed with 0.1 M $NiSO_4$, to bind Ni^{2+} onto the column, and rinsed again with ddH_2O to remove any unbound $NiSO_4$.

For additional purification, a Hitrap Q Ion Exchange column (GE Healthcare, Piscataway, NJ) was used. For the Q column, the binding Buffer A was comprised of 20 mM TRIS, pH 8.0. The elution Buffer B was comprised of 20 mM TRIS, 1 M NaCl, and pH 8.0.

Protein injections onto the column were limited to 5-8 mL. Once all of the protein was loaded onto the column, an elution method was run to elute the bound protein in 8 mL fractions. The collected fractions were further purified by dialysis in 2.0 L of 10 mM TRIS, 1 mM Dithiothreitol (DTT), (Inalco, Milano, Italy), pH 7.4. Protein fractions were sealed in dialysis bags (Spectrum, Rancho Dominguez, CA) with a molecular weigh cutoff value of 3,500 Da, and stirred on a stir plate for 72 hours. The dialysis solution was changed every 24 hours to remove imidazole and other impurities. Following dialysis, samples were extracted from the collected fractions and the purity evaluated using SDS-PAGE gels. Protein concentration was determined using UV-Vis Spectrometry, based on the Beer-Lambert Law

$$A = \varepsilon bc \qquad \text{(Equation 9)}$$

where b = path length (1 cm), A is the measured absorbance, and ε is the previously determined molar absorptivity found experimentally to be 21890 M^{-1} cm^{-1}. The

absorbance scan encompassed the range of 600-220 nm, and the absorbance for determining concentration was measured at 280 nm.

5.2.2 Metal-binding Studies

Previous work was completed in our laboratory to evaluate the interaction of metals (Ca^{2+}, Mg^{2+}, Gd^{3+}, Tb^{3+}, La^{3+}) with EGFPwtF and our engineered variants [25], although Pb^{2+} was not included in these analyses. To validate these previous studies, and determine if metal selectively can be analyzed ratiometrically (e.g. – correlative shifts in two related signal intensities subsequent to metal-binding), additional tests were conducted. The scaffold protein EGFPwtF and our variants contain a natural chromophore that can be evaluated by excitation using a fluorometer. The signal intensity associated with the EGFP chromophore can be altered in two ways by binding of metal ion. First, if a metal ion can be bound in close enough proximity (less than 100 Å) to the protein chromophore, the resulting second chromophore resulting from the metal-protein complex may either quench or enhance the protein chromophores signal. Second, the binding of the metal ion may produce a conformational change in the protein. Because the EGFP protein variants utilized in this research contain both a neutral and anionic species, a conformational change may either increase or decrease solvent exposure to the buried chromophore, resulting in a change in the ratio of one species to the other. For example, the anionic species which emits fluorescence when excited at 490 nm, tends to decrease relative to the neutral species when the chromophore is solvent exposed in basic pH. This ratio then can be used to analyze the protein variants ability to bind metals, and determine selectivity and affinity between competing metal ions.

Fluorometric spectral analyses of EGFPwtF and our engineered variants were conducted using two different PTI (Photon Technology International, Birmingham, NJ) Spectrofluorometers equipped with a 75 W xenon arc lamp and a model 814 Photomultiplier Tube (PMT) detector. During analyses, the excitation slit widths were set at 1 nm, to reduce potential photobleaching of the proteins, and the emission slit widths were set at 2 nm.

Previous work in our laboratory utilized excitation wavelengths of 398 nm and 488 nm, based on values reported in the literature [29], for similar analyses. To validate this, fluorometric excitation scans were run on the protein variants from emission wavelengths 507-511 nm to verify the optimal excitation wavelengths. The first excitation wavelength of 398 nm was consistently produced on both fluorometers used in the study. However, the second excitation varied between 488, 490, and 492 nm, depending on the instrument. Additionally, the emission maxima appeared at either 509 or 510 nm, so data reported in this thesis will reflect different wavelength values, in some cases. It should be noted that the ratiometric analyses compensated for this variance, as well as fluctuations in the instrument energy over time, and analyte concentrations. Data from the fluorometers were collected at 1 nm intervals, and stored on PC's running Felix32 software (Photon Technology International, Lawrenceville, NJ). Additional verification of excitation spectra was obtained using a Shimadzu spectrofluorometer (Shimadzu North America, Columbia, MD).

Absorbance data for the proteins were obtained using a Shimadzu UV-1700 PharmaSpec spectrophotometer with UVProbe software. The absorbance scan encompassed the range of 600-220 nm. Data from all spectra were saved in text files for

subsequent processing in spreadsheet software, either Microsoft Excel (Microsoft, Redmond, WA) or Kaleidagraph (Synergy Software, Reading, PA).

5.2.3 Preparation of Buffers

Metalloprotein complexes were evaluated in different buffers. EMD Omnipur tris (hydroxymethyl)aminoethane, or TRIS, was prepared at concentration of 10 mM and 100 mM in ddH_2O. HCl was used to adjust the pH to 7.4 for analysis of all metals. For Pb^{2+}, several different buffers were evaluated. Initially, 10 mM TRIS- NO_3 buffers were prepared, and the pH adjusted down to 7.0 and 7.4 using A.C.S Reagent Grade HNO_3 (EM Science, Norwood, OH). Nitric acid was used to avoid potential complications associated with the formation of $PbCl_2$, and because the source of the Pb^{2+} ion was $PbNO_3$. However, the presence of higher concentrations of NO_3 interfered with the UV-Vis spectra at shorter wavelengths, as seen in Figure 5.5.

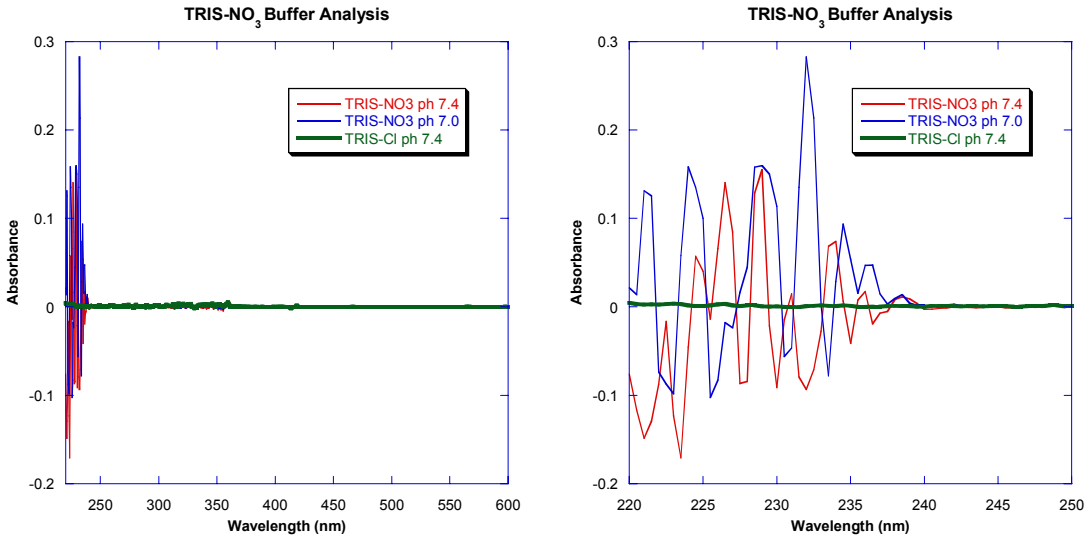
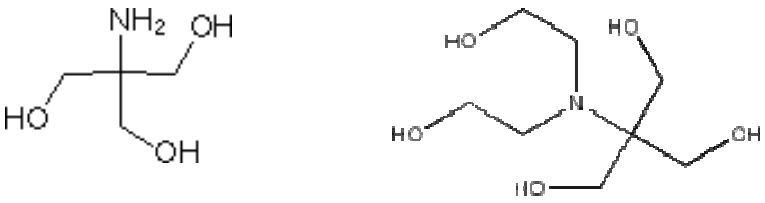


Figure 5.5: UV-Vis spectra of TRIS-NO_3 buffer.

This presented a potential problem since the concentration of the proteins was evaluated based on the absorbance at 280 nm, corresponding to the tryptophan residues in the protein. All subsequent buffers were acidified using HCl, rather than HNO_3.

Bis-TRIS (2,2-Bis(hydroxymethyl)-2,2',2"-nitrilotriethanol, 99+%, Acros Organics, Geel, Belgium) was also evaluated as a potential buffer. Bis-TRIS is similar to TRIS, but has a lower pK_a than TRIS (6.5 vs. 8.1), and previous work with Pb^{2+} was reported using Bis-TRIS [74, 78]. Bis-TRIS buffers were prepared at 10 mM and 100 mM, and for each, the pH was adjusted to 7.0 and 7.4, using HCl. Structures of the two buffers are represented in Figure 5.6.



$NH_2C(CH_2OH)_3$ TRIS $C_8H_{19}NO_5$ Bis-TRIS

Figure 5.6: Line-bond structures of TRIS (left) and Bis-TRIS (right).

To improve the accuracy of metalloprotein analyses, it was necessary to remove or control the concentration of background metals present in the buffers, particularly calcium which is often present in TRIS as an impurity. To accomplish this, the prepared buffers were treated with Analytical Grade Chelex 100 resin, 100-200 mesh Sodium Form (Bio-Rad Laboratories, Hercule, CA), hereafter referred to as Chelex. Two different methods were evaluated to determine optimal calcium removal. First, a 100 mM TRIS-Cl buffer was treated by dialysis with Chelex for three days. The Chelex concentration was 5 g/L. The buffer was placed on a stir plate at 4 °C. Next, 5 g/L of

Chelex was added directly to the 100 mM TRIS-Cl buffer, which was placed on a stir plate at 4 °C for 1 day, following which the buffer was filtered to remove the Chelex.

A comparison of the two methods can be seen in Figure 5.7. The background calcium present after treatment with Chelex was determined using 5,5',6,6'-tetrafluoro BAPTA dye (Molecular Probes, Eugene, OR).

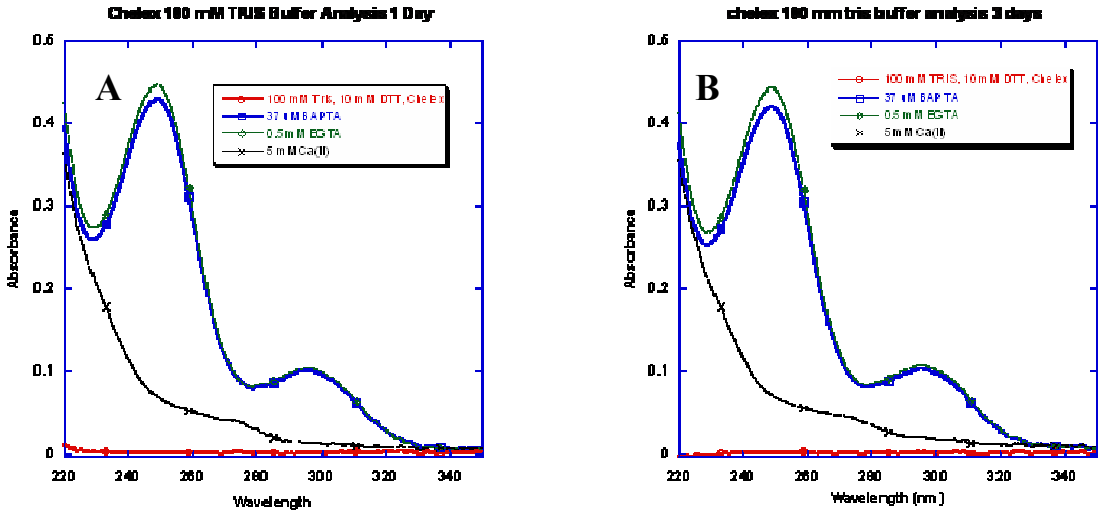


Figure 5.7: UV-Vis spectra of free Ca^{2+} in Chelex-treated 100 mM TRIS-Cl. (A) Direct addition of Chelex, 1 day. (B) Dialysis with Chelex, 3 days.

The remaining free calcium was calculated using Equation 10.

$$[Ca^{2+}]_{free} = [Dye] * ((F_{EGTA} - F_{BAPTA})/(F_{EGTA} - F_{Ca^{2+}})) \quad \text{(Equation 10)}$$

The direct addition method resulted in a $[Ca^{2+}]_{free}$ concentration of 1.58 µM, while the dialysis method resulted in a concentration of 2.31 µM. Subsequent preparation with Chelex was conducted by direct addition of Chelex, followed by filtration after 3 days, resulting in free calcium concentrations below 1 µM.

5.2.4 Analyses of EGFPwtF and Variants During Metal Complexation

The first series of analyses were conducted using 1.0 μM concentrations of EGFPwtF, EGFP-C-1, EGFP-C-3, EGFP-B-1, and EGFP-B-4. Fluorometric emission scans were conducted at excitation wavelengths 398 nm and 488 nm, corresponding to a previously reported emission maxima at 510 nm. Four metals (Eu^{3+}, Gd^{3+}, La^{3+} and Tb^{3+}) were added at 100 μM concentrations, and then subsequently removed from the solvent by addition of 1 mM ethylene glycol tetraacetic acid (EGTA). The fully-protonated structure of EGTA can be seen in Figure 5.8.

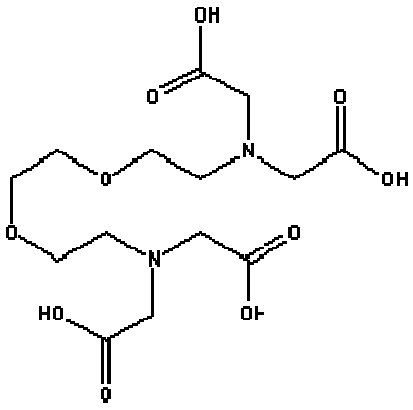


Figure 5.8: Line-bond structure of ethylene glycol tetraacetic acid (EGTA), from http://chemfinder.cambridgesoft.com/.

Emission scans were completed for each sample before addition of the metal, after addition of the metal, and after addition of the EGTA. A single analysis of each metal was completed for this initial evaluation. The Ratiometric change from the metal-free protein to the metal-protein complex was calculated by integrating the peak areas for each of the emissions scans (398 nm and 488 nm) from 500-600nm as a sum of the intensities recorded at each 1 nm interval, and then evaluating the ratio of (F398/F488), as seen in Equation 11.

$$\text{Ratiometric change} = (F398/F488) = \left(\sum_{500}^{600} Counts398 \Big/ \sum_{500}^{600} Counts488 \right) \qquad \text{(Equation 11)}$$

This change is calculated as a ratio to eliminate potential errors associated with absolute intensity values that might arise due to instrumental variations.

5.2.5 Direct Addition of Ca^{2+} and Gd^{3+} to EGFPwtF

Next, triplicate analyses of 7 μM EGFPwtF protein with Ca^{2+} and Gd^{3+} were completed. These higher concentrations were utilized both to evaluate the UV-Vis absorbance, and to evaluate the impact of the metal ions on the pH. Given that the original experiments were evaluated at physiological pH (~7.4), and since the absorbance of EGFPwtF and our variants is pH-dependent in that a pH shift alters the ratio of the two protein species (neutral/anionic), it was necessary to determine if the changes in the fluorometric spectra were related indirectly to the free metal ion concentration as an consequence of pH change. For this analysis, the ten-fold increase in protein concentration required an equivalent increase in both the added metal ion (1 mM) and EGTA (10 mM). Also, 1 mM DTT was added to the buffer, to prevent disulfide bond formation in reduced pH.

5.2.6 Competitive Titrations between Ca^{2+} and Gd^{3+}

The fluorescent ratiometric change (F398/F490) was evaluated for 1.0 μM EGFPwtF and the EGFP-C2 variant to evaluate selectivity between Ca^{2+} and Gd^{3+} in 10 mM TRIS-Cl, pH 7.4. EGFPwtF was utilized as a control because it does not include a grafted binding site. First, 1 mM Ca^{2+} was added to the protein, followed by aliquots of

Gd^{3+} covering a concentration range from 1μM to 50 μM. The total added volume represented less than 3% of the volume total, and 25-30 minutes elapsed between aliquots. Next, the procedure was repeated in triplicate with only the EGFP-C2 variant. Again, 1 mM Ca^{2+} was added to the protein, followed by aliquots of Gd^{3+} covering a concentration range from 1μM to 50 μM, but the number of data points was increased in order to calculate K_d for Gd^{3+}.

The affinity of a competing ion is assumed to be directly proportional to the change in the ratio (F398/F490), as calculated using Equation 11. To calculate K_{eq} for the competitive titration, the value for the fraction of the competing ion (F) must be first normalized across the range of concentrations evaluated. This F Normalized value (F_{Norm}) was calculated with Equation 12.

$$F_{Norm} = (F_{Ca\ initial} - F_{Gd})/(F_{Ca\ initial} - F_{Gd\ final}) \quad \text{(Equation 12)}$$

In Equation 12, $\mathbf{F_{Ca\ initial}}$ represents the initial ratio (F398/F490) following addition of Ca^{2+}, $\mathbf{F_{Gd}}$ is the ratio at each point for addition of Gd^{3+}, and $\mathbf{F_{Gd\ final}}$ is the ratio at the final concentration of Gd^{3+}. Once $\mathbf{F_{Norm}}$ was calculated, K for the competitive titration was calculated in Kaleidagraph with the following curve-fitting equation:

$$F_{Norm} = ((([P]_t + [M]_t + K) - (([P]_t + [M]_t + K)^2 - 4[P]_t[M]_t)^{1/2}) / 2[P]_t) + (C^*[M]_t) \quad \text{(Equation 13)}$$

where the final term, $(C^*[M]_t)$, accounts for non-specific binding. K_d for Gd^{3+} was then calculated using Equation 14, with K calculated from Equation 13, and K_d for Ca^{2+} which was previously determined in our laboratory for the EGFP-C2 variant to be 440 μM.

$$K_{dGd} = K/(1 + ([Ca^{2+}]/K_{dCa})) \quad \text{(Equation 14)}$$

5.2.7 Competitive Titrations between Ca^{2+} and Mg^{2+}

The fluorescent ratiometric change (F398/F490) was evaluated for 1.0 μM EGFPwtF and the EGFP-C2 variant to evaluate selectivity between Ca^{2+} and Mg^{2+} in 10 mM TRIS-Cl, pH 7.4. EGFPwtF was utilized as a control because it does not include a grafted binding site. First, 1 mM Ca^{2+} was added to the protein, followed by aliquots of Mg^{2+} covering a concentration range from 0.5 mM to 20 mM. The total added volume represented less than 3% of the volume total, and 25-30 minutes elapsed between aliquots. Next, the procedure was repeated in triplicate with only the EGFP-C2 variant. Again 1 mM Ca^{2+} was added to the protein, followed by aliquots of Mg^{2+} covering a concentration range from 0.5 mM to 20 mM. As with the Gd^{3+} selectivity titration, the number of data points was increased in order to calculate K_d for Mg^{2+}. K_d for Mg^{2+} was calculated using Equation 13 and Equation 14, following the same procedures as with Gd^{3+}.

5.2.8 Direct Addition of Pb^{2+} to EGFP-B-4 in TRIS-Cl

The direct titration of Pb^{2+} was evaluated with the EGFP-B-4, as this variant exhibited the highest ratiometric response to the lanthanides evaluated. The EGFP-B-4 variant was diluted to 1 μM in 10 mM TRIS-Cl, at pH 7.4. A single aliquot of Pb^{2+} was added, for a solution concentration of 1 μM Pb^{2+}. EGTA was added to remove the metal ion. Analyses were run in triplicate. Additionally, controls were run to evaluate the interaction between the protein and both EGTA and EDTA. The EDTA was evaluated as it has a higher affinity for Pb^{2+} than Ca^{2+}.

5.2.9 Direct Addition of Pb^{2+} to EGFPwtF and Variants in Bis-TRIS-Cl

The direct titration of Pb^{2+} was evaluated with 1-2 μM EGFPwtF and our variants in 10 mM Bis-TRIS-Cl buffer at pH 7.0. Variants evaluated included EGFP-C2, EGFP-C-2, EGFP-C-3, and a variant 5G-III-5G, which contains a grafted segment comprised of CaM loop III surrounded on either side by 5 flanking glycine linkers. The 5G variant graft was also inserted in position C. The purpose of this analysis was to assess the relative ratiometric fluorescent intensities for the different variants, as a result of binding Pb^{2+}. Single titrations were conducted with each variant, with EGFPwtF used as a negative control. Precipitation of $Pb(OH)_2$ was observed at 1.0 mM Pb^{2+}.

5.2.10 Evaluation of non-specific Pb^{2+}-binding to EGFPwtF

Previous results with Pb^{2+} and the lanthanides suggested the probability of non-specific binding of metals to both the wild type and grafting variants. To analyze this, 100 μM of Pb^{2+} was added to 1 μM EGFPwtF in 10 mM TRIS-Cl, pH 7.4, followed by addition of 10 mM KCl. This was followed by addition of 1 mM EGTA to remove the metal ions. Analyses were run in triplicate.

5.2.11 Competitive Titrations between Pb^{2+} and Ca^{2+} in TRIS-Cl Buffer

The fluorescent ratiometric change (F398/F490) was determined for 1.0 μM EGFPwtF and the EGFP-C2 variant to evaluate selectivity between Ca^{2+} and Pb^{2+} in 10 mM TRIS-Cl, pH 7.4. EGFPwtF was utilized as a negative control because it does not include a grafted binding site. Initially, 1 mM Ca^{2+} was added to the protein followed by aliquots of Pb^{2+} covering a concentration range from 1.0 μM to 500 μM to determine an approximate maximum concentration of Pb^{2+} that could be added before precipitation of

$Pb(OH)_2$. The total added volume represented less than 3% of the volume total, and 25-30 minutes elapsed between aliquots. No precipitation was observed at 100 μM, but was evident at 500 μM.

Next, the procedure was repeated in triplicate with variants EGFP-C2 and EGFP-C-4. It was assumed that these variants, which both included the complete E-III-F binding site, would exhibit the highest binding affinities. Again, 1 mM Ca^{2+} was added to the protein followed by aliquots of Pb^{2+} to obtain protein concentrations at the following points: 1.0, 2.0, 5.0, 10.0, 25.0, 50.0, 75.0 and 100.0 μM. The increased number of data points was used to calculate K_d for Pb^{2+}. K_d for Pb^{2+} was calculated using Equation 13 and Equation 14. The calcium K_d values previously established for EGFP-C2 and EGFP-C-4 were 440 μM and 800 μM, respectively.

5.2.12 Competitive Titrations between Pb^{2+} and Ca^{2+} in Bis-TRIS-Cl Buffer

The fluorescent ratiometric change (F398/F490) was determined for 1.0 μM EGFPwtF and the EGFP-C2 variant to evaluate selectivity between Ca^{2+} and Pb^{2+} in 10 mM Bis-TRIS-Cl, pH 7.0. EGFPwtF was utilized as a negative control because it does not include a grafted binding site. Results of analysis in Bis-TRIS were compared with those obtained in TRIS. Initially, 1 mM Ca^{2+} was added to the protein followed by aliquots of Pb^{2+} covering a concentration range from 1.0 μM to 100 μM. The total added volume represented less than 3% of the volume total, and 25-30 minutes elapsed between aliquots.

The procedure was then repeated in triplicate with only the EGFP-C2 variant. Again, 1 mM Ca^{2+} was added to the protein followed by aliquots of Pb^{2+} to obtain protein

concentrations at the following points: 1.0, 2.0, 5.0, 10.0, 25.0, 50.0, 75.0 and 100.0 μM. The increased number of data points was used to calculate K_d for Pb^{2+}. K_d for Pb^{2+} was calculated using Equation 13 and Equation 14.

5.3 Results and Discussion

5.3.1 Protein Expression and Purification of EGFPwtF and Variants

Optimized parameters for the expression and purification of EGFPwtF and engineered variants were previously evaluated in our laboratory [26], using established methods [29, 103]. Generally, it was found that variants where the grafted binding motif from CaM was inserted at site B displayed weaker fluorescence than site C variants.

Although previous work in our laboratory had demonstrated that purifications of the EGFP variants using the Hitrap chelating column were generally adequate, two additional efforts were made to determine if purity could be improved. First, the binding buffer was prepared as detailed in Section 5.2.3, but was modified by the addition of 50 mM Imidazole.

An SDS-PAGE gel comparing the samples from the two buffers is shown in Figure 5.9. For the protein samples observed, 16 μL sample volume was injected. Lane 1 is the (BSA/carbonic anhydrase/lysozyme) marker. Lane 2 is a sample from the cell pellet. Lanes 3 and 4 are fractions from Buffer A only, before dialysis. Lanes 5-7 are fractions from Buffer A and 50 mM Imidazole, before dialysis. Lanes 8 and 9 are fractions from Buffer A only, after dialysis. Lane 10 is a fraction from Buffer A and 50 mM Imidazole, after dialysis.

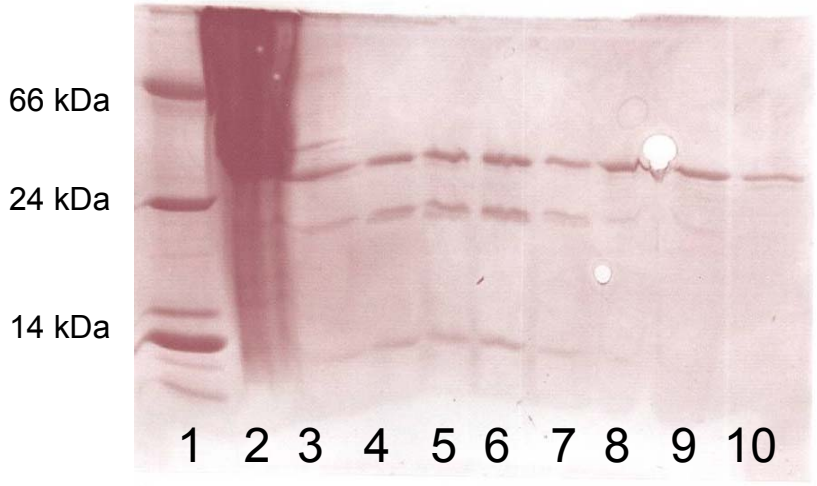
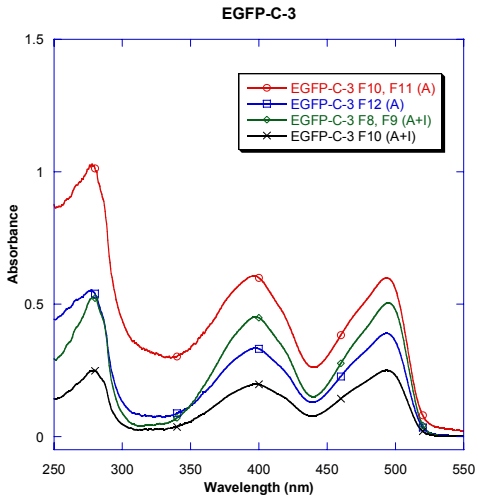


Figure 5.9: SDS-PAGE for EGFP-C-3. Fractions were collected from Binding Buffer A only (Lanes 3, 4) and Buffer A + 50 mM Imidazole (Lanes 5, 6).

The SDS-PAGE gel pictured in Figure 5.9 indicates that no apparent improvement in purity resulted from the addition of Imidazole for the purification of EGFP-C-3. However, it should be noted that the band apparent at 14 kDa in Lanes 3-6, where samples were collected prior to dialysis, are not present in Lanes 8-10, after dialysis.



Relative Concentration Values for EGFP-C-3 Purified in Different Buffers

Buffer	Fraction	Absorbance	Conc. (μM)
A	F10, F11	1.0128	46.27
A	F12	0.5397	24.66
A + 50 mM Imidazole	F8, F9	0.5232	23.90
A + 50 mM Imidazole	F10	0.2482	11.34

Figure 5.10: UV-Vis absorbance for EGFP-C-3 purified in different binding buffers. (A + I) indicates Buffer A and 50 mM Imidazole.

Additionally, a significant decrease in concentration was observed in the collected fractions, indicating that the Imidazole resulted in a loss of the protein, as seen in Figure 5.10.

The second effort involved an initial purification using the Hitrap Chelating column, followed by an additional purification using a Hitrap Q Ion Exchange column. Improved purification for EGFPwtF was observed following subsequent purification using the Q column, however, the associated absorbance spectra (Figure 5.11) indicated very low concentration for the recovered protein, with the two major fractions yielding concentrations of only 12.26 µM and 17.81 µM. In comparison with typical concentration yields exceeding 200 µM, these results indicate that the current method for the Q column requires additional changes for improved results.

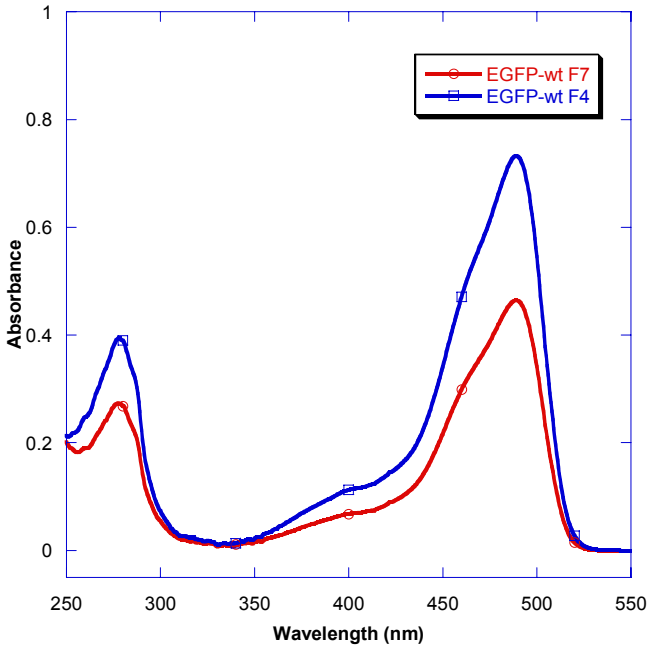


Figure 5.11: UV-Vis absorbance spectra for EGFPwtF following purification with Ion Exchange column.

5.3.2 Metal-binding Studies

5.3.2.1 Fluorometric Analyses during Metal Complexation

The ratiometric changes associated with the addition of Eu^{3+}, Gd^{3+}, La^{3+}, and Tb^{3+} are shown in Figure 5.12. As previously noted these were single trials, where 100 µM of the metal ion was added directly to 1 µM of the protein, followed by the addition of 1 mM EGTA to remove the metal ion.

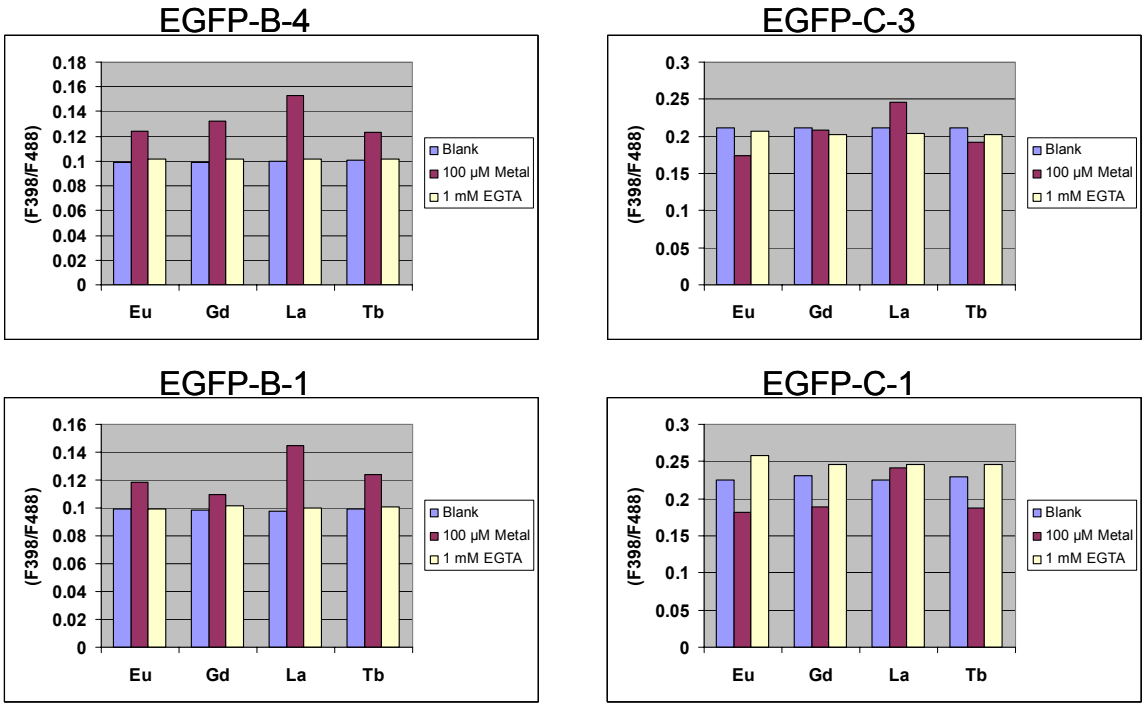


Figure 5.12: Comparison of ratiometric signal intensity changes following addition and removal of 100 µM metal ions with 1 mM EGTA. Samples consisted of 1.0 µM protein in 10 mM TRIS-Cl, pH 7.4.

As shown in Figure 5.12, addition of metals to variants B-4 and B-1 both produced an increased ratiometric change, with La^{3+} producing the most pronounced

change. However, the C-3 and C-1 variants in Figure 5.12, exhibited a different trend, where a reduction in the ratio was observed for all metals except La^{3+}.

Table 5.2: Summary of emission maxima at 398 nm and 488 nm for Fluorometric Spectra associated with EGFP variants binding different metals.

	$\lambda_{emission}$ at 488 nm			$\lambda_{emission}$ at 398 nm		
	Blank	**Metal**	**EGTA**	**Blank**	**Metal**	**EGTA**
EGFPwtF						
Eu(III)	510	512	510	509	513	509
Gd(III)	509	512	509	510	513	509
La(III)	509	512	510	509	513	509
Tb(III)	509	513	510	509	513	509
EGFP-C-1						
Eu(III)	510	512	510	510	513	510
Gd(III)	510	512	510	511	512	511
La(III)	511	513	512	509	512	510
Tb(III)	510	512	510	510	512	510
EGFP-B-1						
Eu(III)	509	512	509	509	511	509
Gd(III)	509	512	510	508	511	509
La(III)	510	512	509	510	511	510
Tb(III)	510	512	509	509	511	509
EGFP-B-4						
Eu(III)	509	512	509	510	510	509
Gd(III)	509	511	509	511	512	509
La(III)	509	512	509	509	513	509
Tb(III)	510	512	509	509	512	508
EGFP-C-3						
Eu(III)	510	512	510	510	513	511
Gd(III)	511	512	510	510	511	510
La(III)	509	512	510	510	512	509
Tb(III)	510	512	510	510	511	511

Additionally, as seen in Table 5.2, an average 2-3 nm red-shift was observed in the spectra following addition of the metal ions, which disappeared following addition of EGTA. This red-shift is clearly observable in the spectra for addition of Eu^{3+} to EGFP-C-1, excited at 488 nm (Figure 5.13).

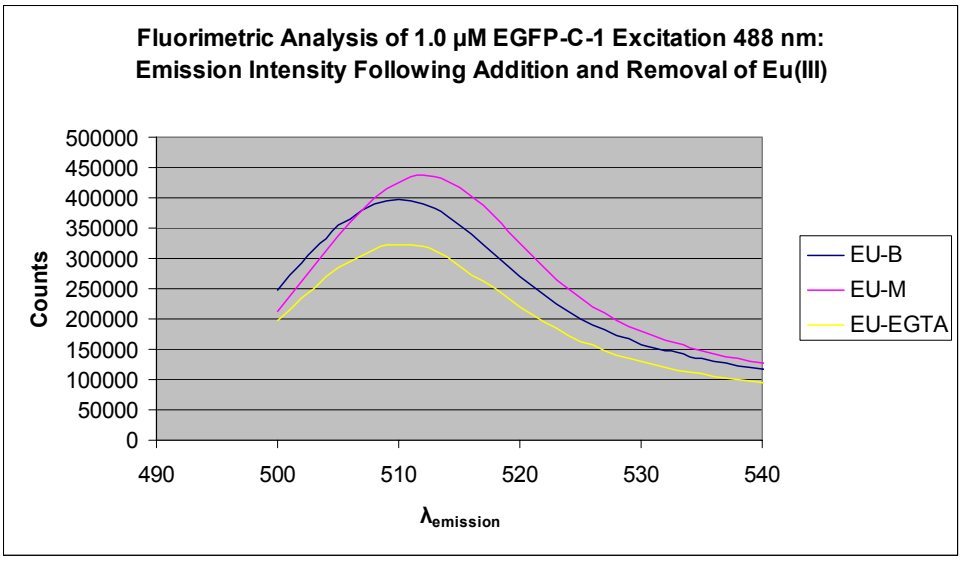


Figure 5.13: Red-shift observed following addition of Eu^{3+} to variant EGFP-C-1.

The cause of these results was not immediately apparent. The same experimental conditions were then applied to the wild-type variant, EGFPwtF, which should act as a negative control, given that it does not have the grafted calcium-binding site included with the engineered variants. The results of this analysis were perplexing, in that they suggested binding of the metal ions, as indicated by the ratiometric change following addition of the metals (Figure 5.14). Also, the red-shift observed with the EGFP variants that included a grafted binding site was also apparent in the spectra for the EGFPwtF protein (Figure 5.15), which suggests that energy is being transferred between chromophores during relaxation, and is being retransmitted at a lower energy level.

Given the absence of a binding loop in EGFPwtF, the results suggested two possible explanations. First, the method itself was invalid, although previous work in our laboratory had utilized this method to evaluate other protein graft variants [26]. The second possibility was that the observed ratiometric changes were a result of non-specific binding, which would require additional validation.

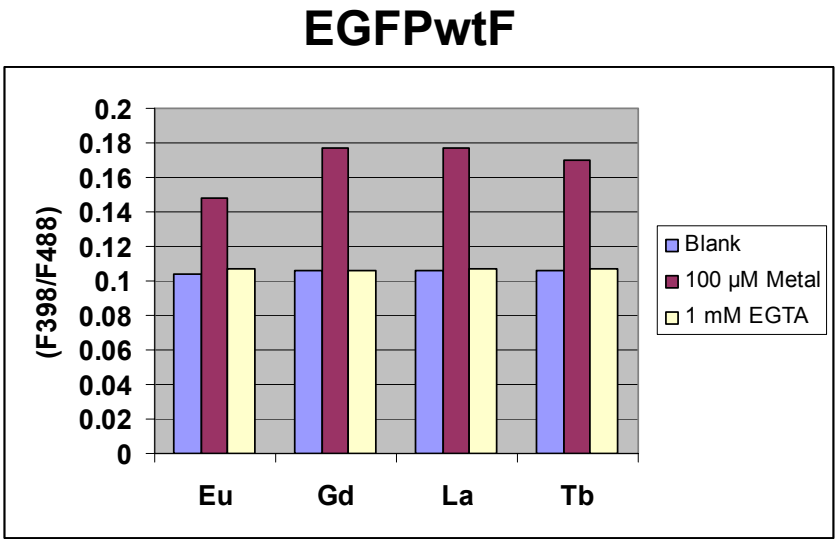


Figure 5.14: Comparison of ratiometric signal intensity changes following addition and removal of metal ions. Conditions: 1.0 µM EGFPwtF in 10 mM TRIS-Cl, pH 7.4; addition of 100 µM metal ion, followed by addition of 1 mM EGTA.

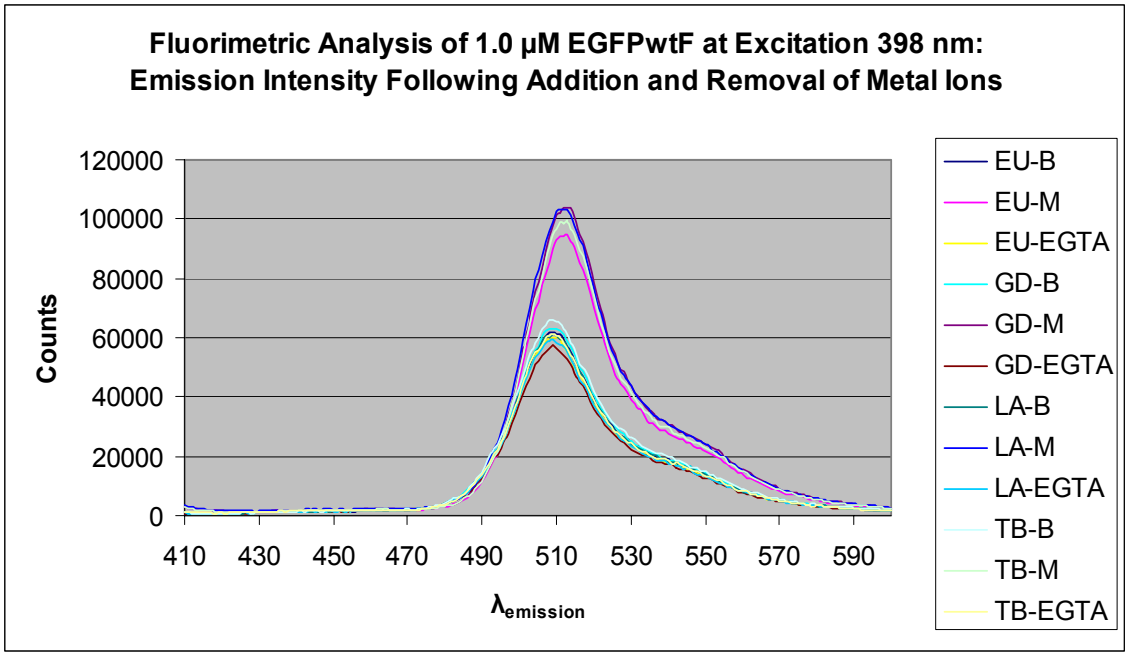


Figure 5.15: Fluorescent emission spectra of EGFPwtF associated with the metals Eu^{3+}, Gd^{3+}, La^{3+} and Tb^{3+}. A red-shift of 2-3 nm is observed upon addition of the metal ions, which reverts to the original wavelength after addition of EGTA.

5.3.2.2 Direct Addition of Ca^{2+} and Gd^{3+} to EGFPwtF

The direct addition of 1 mM Ca^{2+} and 1 mM Gd^{3+} to 7 μM EGFPwtF was evaluated based on UV-Vis absorbance, pH, and Fluorescence. The UV-Vis spectra can be observed in Figure 5.16 and Figure 5.17 for Ca^{2+} and Gd^{3+}, respectively, where B = Blank, M = Metal, and E = EGTA. Because higher concentrations of the protein and metals were utilized, while maintaining the same TRIS-Cl buffer concentration (10 mM), the effects of pH change are more apparent.

The pH was measured prior to addition of the metal ion, after addition of the metal ion, and after addition of EGTA, as seen in Table 5.3. Mean and standard deviation values were calculated for the triplicate samples evaluated for each metal. From Table 5.3 it can be seen that that the mean pH following addition of Gd^{3+} is lower than that observed for Ca^{2+}. In both cases, the pH drops to ~6 following addition EGTA. These effects are apparent in Figures 5.16 and 5.17, where the absorbance peak near 492 nm decreases significantly when EGTA is added, with a corresponding increase at 398 nm corresponding to the neutral – or protonated – form of the chromophore. Additionally, at 1 mM Gd^{3+}, precipitation was observed that is reflected in Figure 5.17. The precipitation disappeared following addition of EGTA.

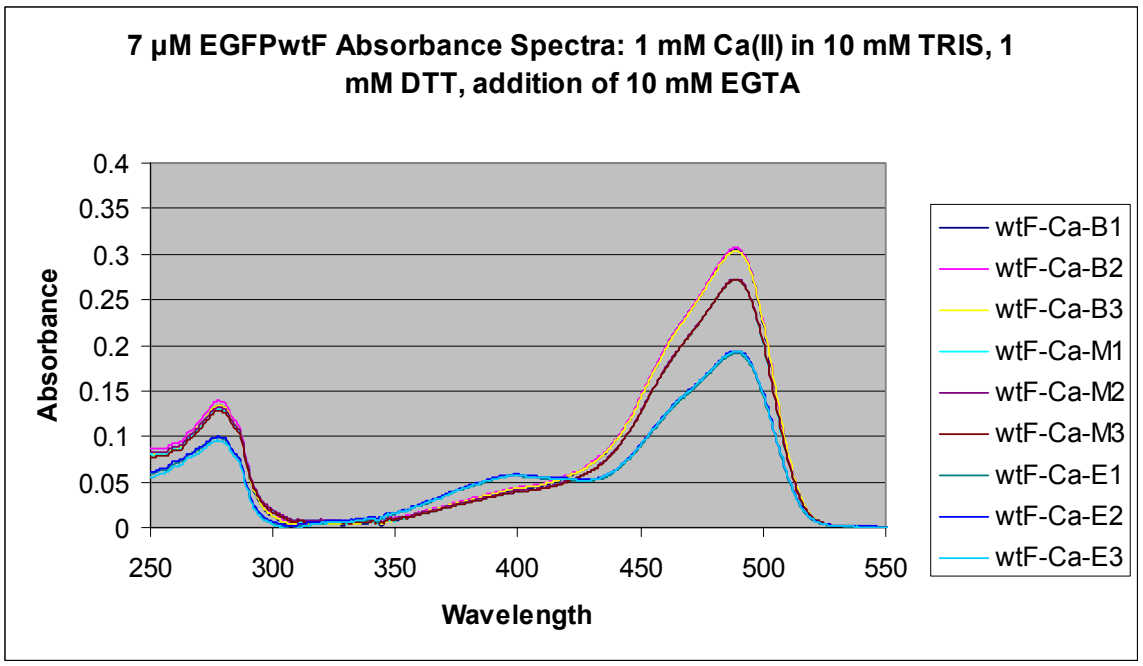


Figure 5.16: UV-Vis absorbance spectra of EGFPwtF and Ca^{2+}. In the legend, B indicates blank (protein only), M indicates addition of metal ion, and E indicates addition of EGTA.

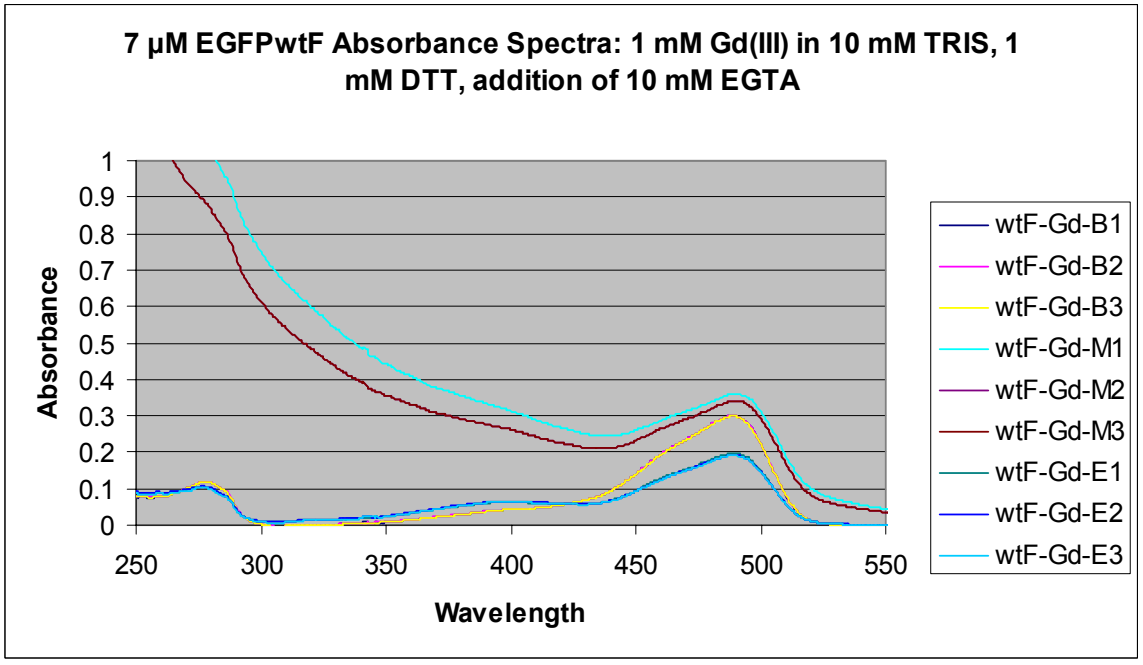


Figure 5.17: UV-Vis absorbance spectra of EGFPwtF and Gd^{3+}. In the legend, B indicates blank (protein only), M indicates addition of metal ion, and E indicates addition of EGTA.

Table 5.3: Analysis of pH changes associated with metal-binding of EGFPwtF

Sample ID	pH Blank	pH Metal	pH EGTA
Ca1	7.47	7.23	6.07
Ca2	7.47	7.26	6.16
Ca3	7.47	7.28	6.18
Mean	7.47	7.26	6.14
Std Dev	0.00	0.03	0.06
SREL	0.00	0.35	0.95
Gd1	7.56	7.01	6.04
Gd2	7.55	6.99	5.99
Gd3	7.55	6.99	6.00
Mean	7.55	7.00	6.01
Std Dev	0.01	0.01	0.03
SREL	0.08	0.17	0.44

Several important observations can be made with respect to the fluorescence signal changes, as seen in the ratiometric changes graphed in Figure 5.18, and the fluorescence spectra presented in Figures 5.19-5.22. First, the error bars in Figure 5.18, representing standard deviation, indicate low variance for addition of Ca^{2+}, but the addition of Gd^{3+} resulted in a percent relative standard deviation of 30%, indicating poor reproducibility with this experiment.

In Figure 5.18 it can be seen that no significant ratiometric change is observed with EGFPwtF following addition of Ca^{2+}. In contrast, a significant change is observed, following addition of Gd^{3+}, which is consistent with changes observed in the previous experiment when Gd^{3+} and other lanthanides were added directly to EGFPwtF. This suggests that if non-specific binding of the lanthanides tested is occurring, it is not occurring with Ca^{2+}. The importance of this becomes apparent in subsequent discussions regarding competitive binding.

Two additional points must be made, with respect to the fluorescent spectra seen in Figures 5.19-5.22. First, the red-shift previously noted with the addition of Eu^{3+}, Gd^{3+},

La^{3+} and Tb^{3+}, was confirmed for Gd^{3+} in this analysis. It was not, however, observed for Ca^{2+}. These results suggest that the observed shift to a longer wavelength may be affected by conformational changes associated with non-specific binding to the protein surface.

Second, the effects of pH are also observable in the fluorescence spectra, as seen in Figures 5.19-5.22, where the addition of EGTA results in a significant decrease in fluorescent intensity at both 398 nm and 492 nm for both Ca^{2+} and Gd^{3+}. The pH effects observed are due to the higher concentrations of metal ions and EGTA in this analysis, relative to TRIS-Cl buffer concentration of 10 mM. Subsequent analyses were conducted with higher buffering capacity.

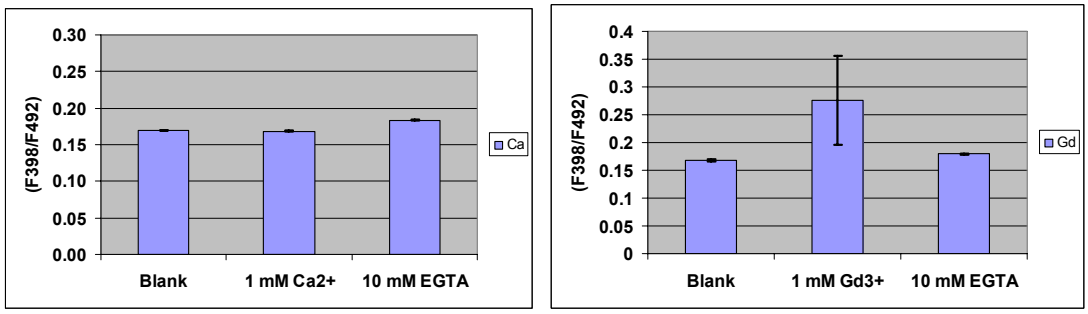


Figure 5.18: Fluorometric emission spectra at excitation wavelength 492 nm for EGFPwtF with 1 mM Ca^{2+} (left) and Gd^{3+} (right).

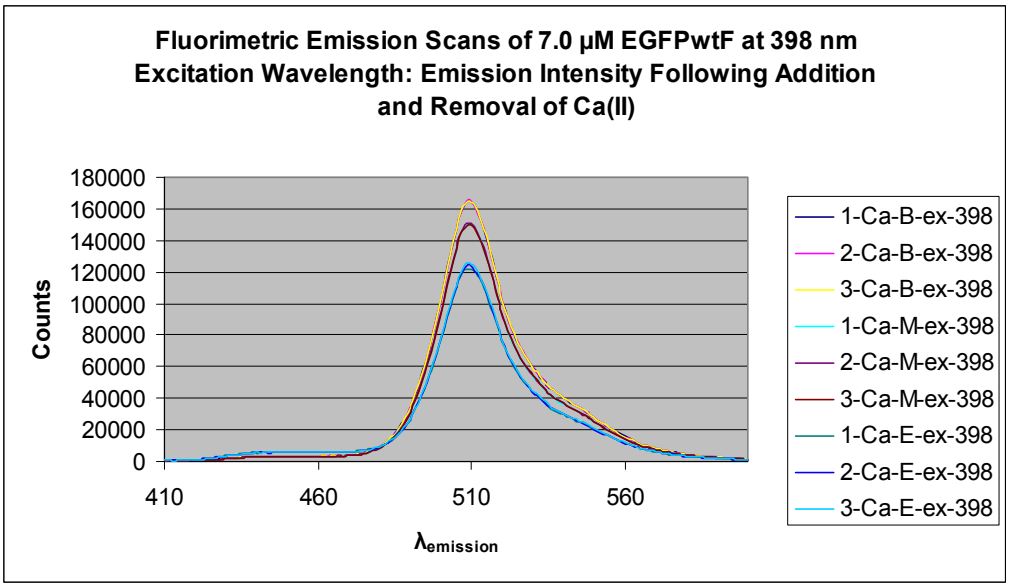


Figure 5.19: Fluorometric emission spectra at excitation wavelength 398 nm for EGFPwtF with 1 mM Ca^{2+}. In the legend, B indicates blank (protein only), M indicates addition of metal ion, and E indicates addition of EGTA.

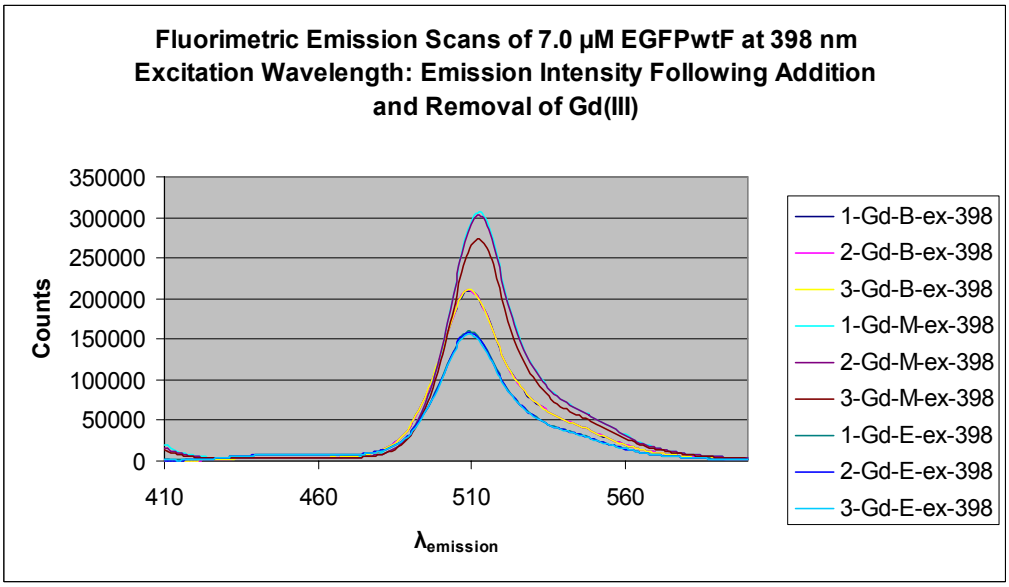


Figure 5.20: Fluorometric emission spectra at excitation wavelength 398 nm for EGFPwtF with 1 mM Gd^{3+}. In the legend, B indicates blank (protein only), M indicates addition of metal ion, and E indicates addition of EGTA.

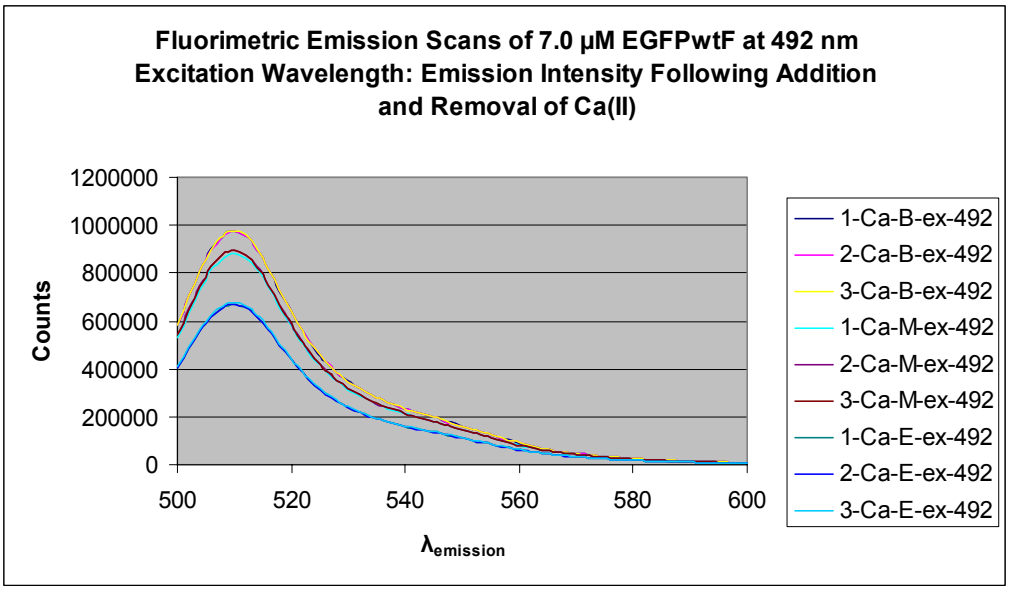


Figure 5.21: Fluorometric emission spectra at excitation wavelength 492 nm for EGFPwtF with 1 mM Ca^{2+}. In the legend, B indicates blank (protein only), M indicates addition of metal ion, and E indicates addition of EGTA.

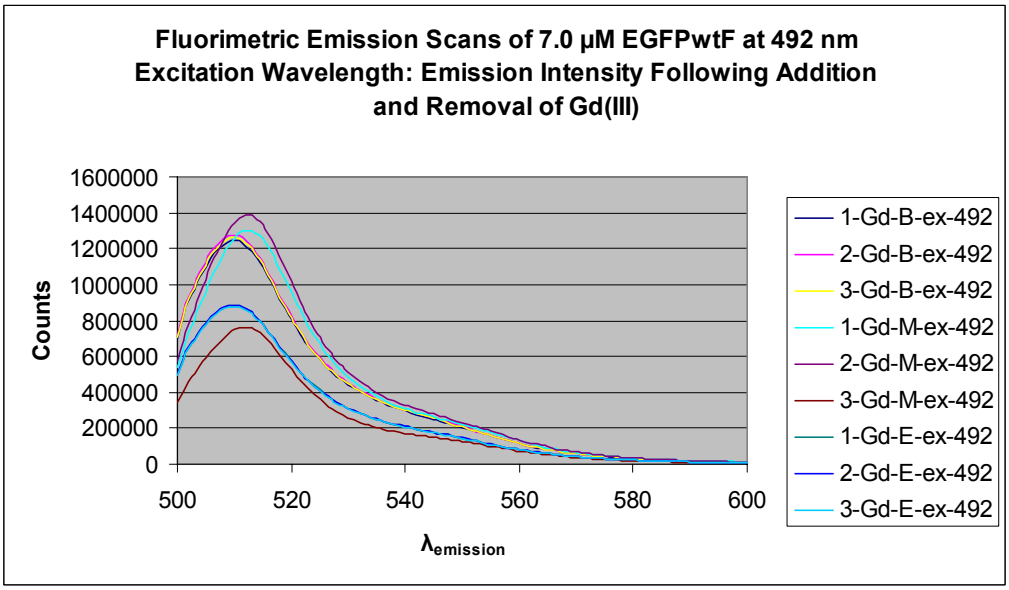


Figure 5.22: Fluorometric emission spectra at excitation wavelength 492 nm for EGFPwtF with 1 mM Gd^{3+}. In the legend, B indicates blank (protein only), M indicates addition of metal ion, and E indicates addition of EGTA.

5.3.2.3 Competitive Titrations Between Ca^{2+} and Gd^{3+}

Figure 5.23 displays the fluorescent ratiometric change (F398/F490) for 1.0 µM EGFPwtF and the EGFP-C2 variant to evaluate selectivity between Ca^{2+} and Gd^{3+} in 10 mM TRIS-Cl, pH 7.4. Three samples were run for each protein. The error bar in the figure represents standard deviation. For the C2 variant, the addition of Ca^{2+} results in a significant increase in the ratio (F398/F490), that gradually decreases following addition of Gd^{3+}. This decrease is assumed to be displacement of the Ca^{2+} ion by Gd^{3+}.

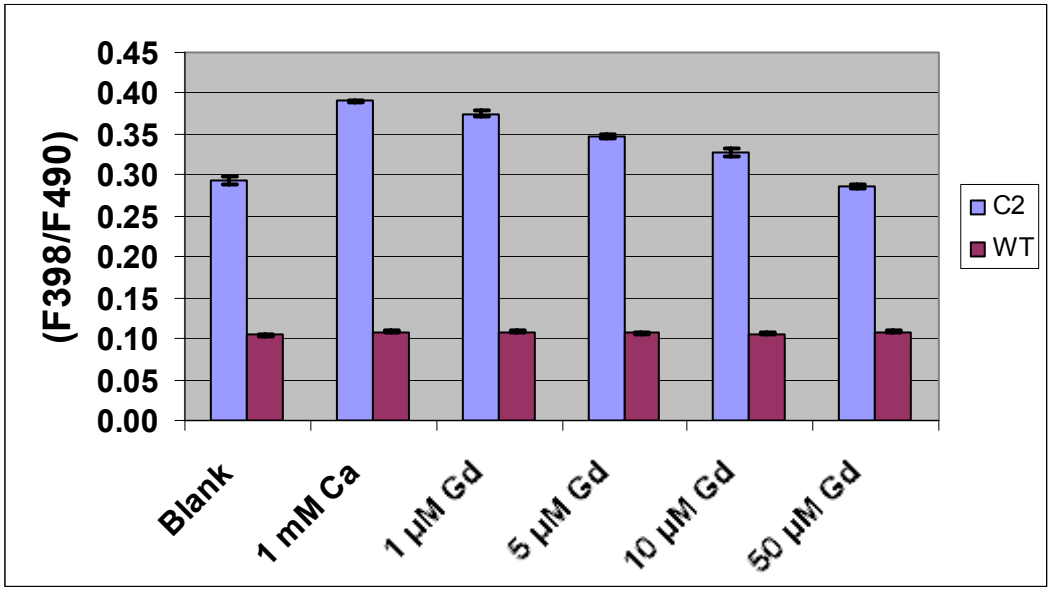


Figure 5.23: Ratiometric change (F398/F490) between 1.0 µM EGFPwtF and C2 variant to evaluate selectivity between Ca^{2+} and Gd^{3+}.

EGFPwtF was utilized as a control because it does not include a grafted binding site, and, as seen in Figure 5.23, does not change significantly with the addition of Ca^{2+}, or with the subsequent aliquots of Gd^{3+}. At first this may seem to contradict the results presented with respect to the direct addition of Gd^{3+}. However, the final concentration for Gd^{3+} presented in Figure 5.23 was 50 µM vs. 1 mM that was added by direct addition.

The red-shift that was observed following addition of 1 mM Gd^{3+} in Figures 5.20 and 5.22 did not appear in the titrations of Gd^{3+} in competition with Ca^{2+} for EGFPwtF between 0-50 µM (Figure 5.24, Figure 5.25). This shift did become apparent in the EGFP-C2 variant as the concentration of Gd^{3+} approached 50 µM, as seen in Figure 5.26 and Figure 5.27.

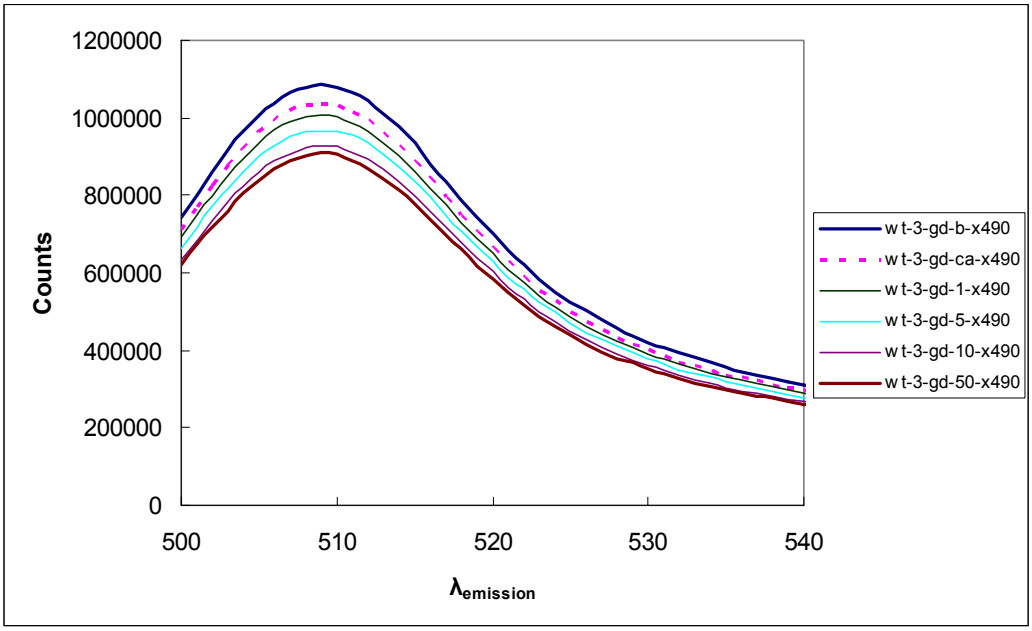


Figure 5.24: Fluorometric emission spectra of 1 µM EGFPwtF at 490 nm excitation wavelength for 1 sample titration. Competitive titration between 1 mM Ca^{2+} and 0-50 µM Gd^{3+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4.

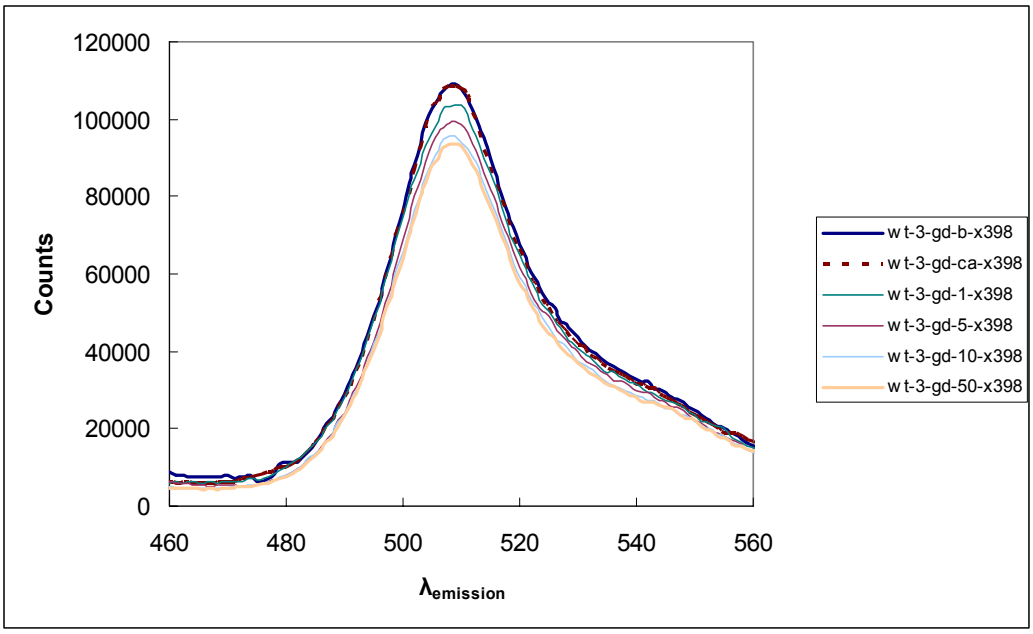


Figure 5.25: Fluorometric emission spectra of 1 µM EGFPwtF at 398 nm excitation wavelength. Competitive titration between 1 mM Ca^{2+} and 0-50 µM Gd^{3+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4.

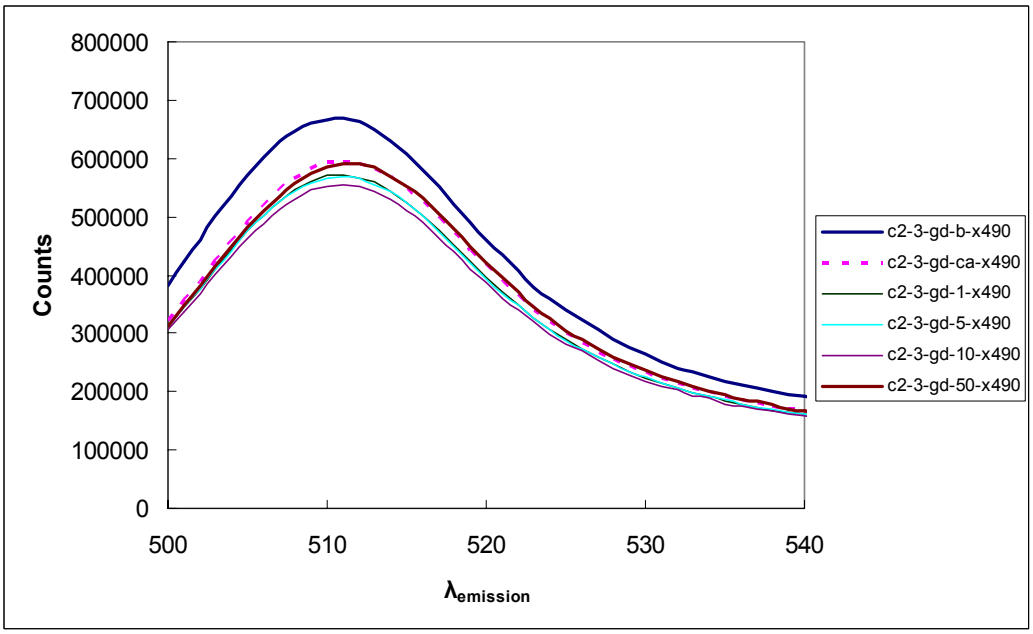


Figure 5.26: Fluorometric emission spectra of 1 µM EGFPwtF C2 variant at 490 nm excitation wavelength. Competitive titration between 1 mM Ca^{2+} and 0-50 µM Gd^{3+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4. Red-shift to 512 nm observed at 50 µM Gd^{3+}.

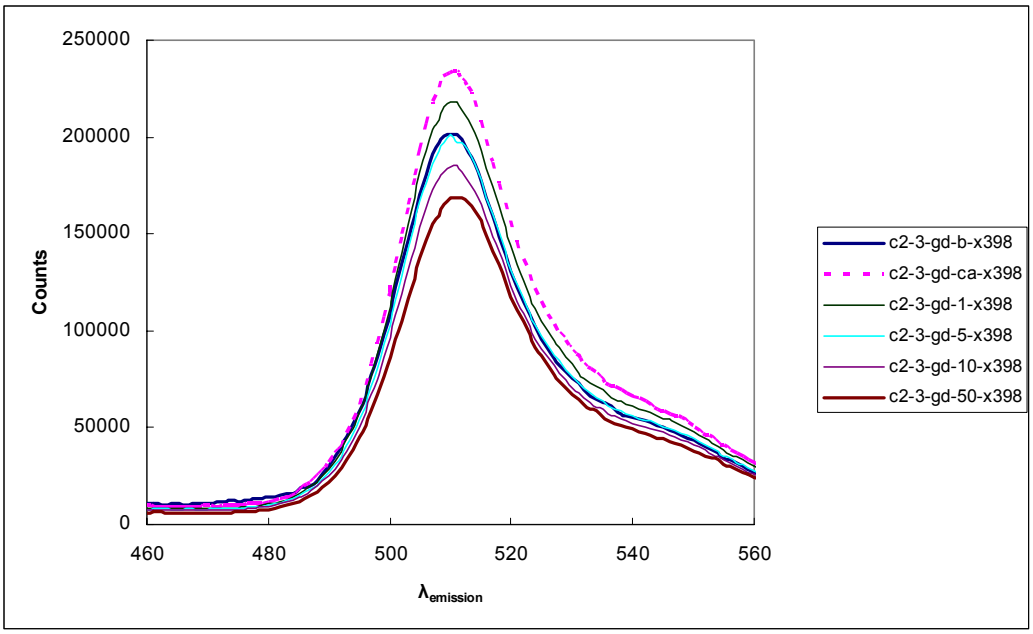
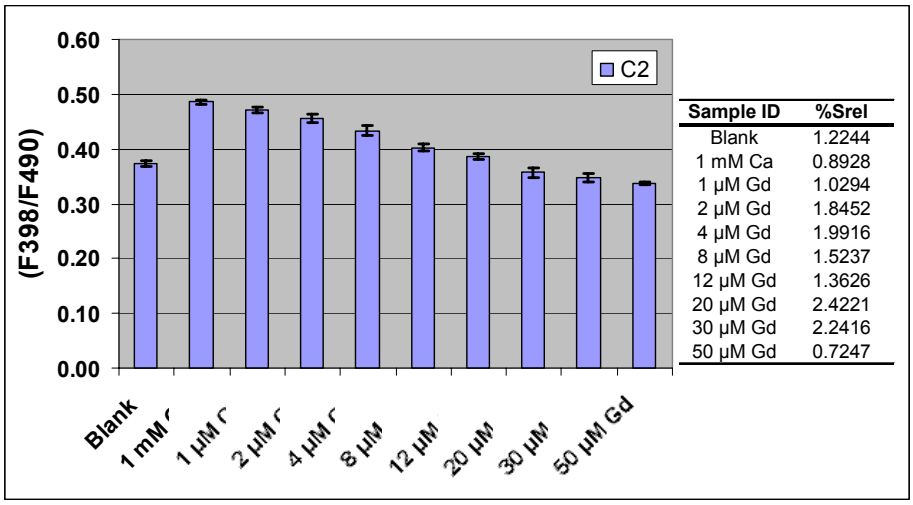


Figure 5.27: Fluorometric emission spectra of 1 µM EGFPwtF C2 variant at 398 nm excitation wavelength. Competitive titration between 1 mM Ca^{2+} and 0-50 µM Gd^{3+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4.

The fact that the red-shift was not observed with EGFPwtF in this analysis, but was observed with EGFP-C2 may indicate that the conformational change required to produce the shift occurs at lower concentration for the variant with the grafted Ca^{2+}-binding site. This is supported by data from the previous analyses indicating a red-shift with EGFPwtF at 100 µM and 1 mM with Gd^{3+}. The data also supports the idea of two binding schemes for the C2 variant: an initial change caused by displacement of the Ca^{2+} ion from the grafted binding loop, followed by non-specific binding on the protein surface.



Sample ID	%Srel
Blank	1.2244
1 mM Ca	0.8928
1 µM Gd	1.0294
2 µM Gd	1.8452
4 µM Gd	1.9916
8 µM Gd	1.5237
12 µM Gd	1.3626
20 µM Gd	2.4221
30 µM Gd	2.2416
50 µM Gd	0.7247

Figure 5.28: Ratiometric change (F398/F490) for C2 variant to evaluate Gd^{3+} affinity.

Next, the competitive titration was repeated across the same concentration range, this time with 8 plotted points for addition of Gd^{3+}, as shown in Figure 5.28. These samples were analyzed in triplicate and then averaged for the graph in Figure 5.28. As seen in the figure, the percent relative standard deviation (%Srel) was calculated to be less than 2.5%, indicating excellent reproducibility.

Data from the 8 points covering the ratiometric change from 0-50 µM (Figure 5.28) were used to plot F Normalized using Equation 12, and the data were plotted using curve-fitting based on Equation 13. Two plots were generated in Figure 5.29, where A does not include the term for non-specific binding, and B does. In Figure 5.29 A, the K_d for Gd^{3+} was calculated to be 1.59 µM, using Equation 14. In Figure 5.29 B, which factors in the final term in Equation 13 for non-specific binding, the K_d for Gd^{3+} was calculated to be 1.95 µM, using Equation 14. Both plots indicate a much higher affinity for Gd^{3+} compared to Ca^{2+} whose K_d was previously established as 440 µM. However, a better fit is observed with plot B in Figure 5.29, where non-specific binding is included in

the fit, and the correlation *R* increases from 0.98577 to 0.99691. These data suggest that Gd^{3+} will readily displace Ca^{2+} in the calcium binding site, and that non-specific binding contributes to the overall binding affinity.

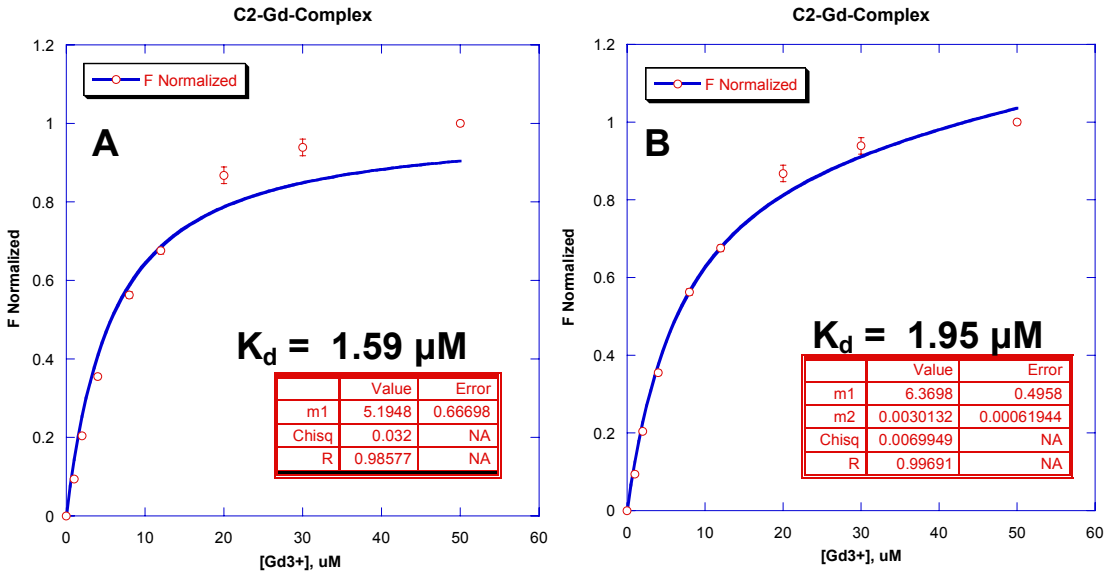


Figure 5.29: Curve fitting to determine K_d for Gd^{3+}. Plot A does not include term for non-specific binding, as shown in plot B. The value m1 represents K overall.

5.3.2.4 Competitive Titrations between Ca^{2+} and Mg^{2+}

Figure 5.30 displays the fluorescent ratiometric change (F398/F490) for 1.0 µM EGFPwtF and the EGFP-C2 variant to evaluate selectivity between Ca^{2+} and Mg^{2+} in 10 mM TRIS-Cl, pH 7.4. For the initial titration, aliquots of Mg^{2+} were added covering a concentration range from 0.5 mM to 20 mM.

For the C2 variant, the addition of Ca^{2+} results in a significant increase in the ratio (F398/F490). However, the gradual decrease that was observed immediately upon addition of Gd^{3+}, was not observed with the addition of Mg^{2+} until 10 mM Mg^{2+},

indicating that the C2 variant is much more selective for Ca^{2+} than Mg^{2+}. This result was expected, as CaBP's typically exhibit low affinity for Mg^{2+}.

As with the previous analysis of Gd^{3+}, EGFPwtF was utilized as a control because it does not include a grafted binding site, and, as seen in Figure 5.30, does not change significantly with the addition of Ca^{2+}, or with the subsequent aliquots of Mg^{2+}.

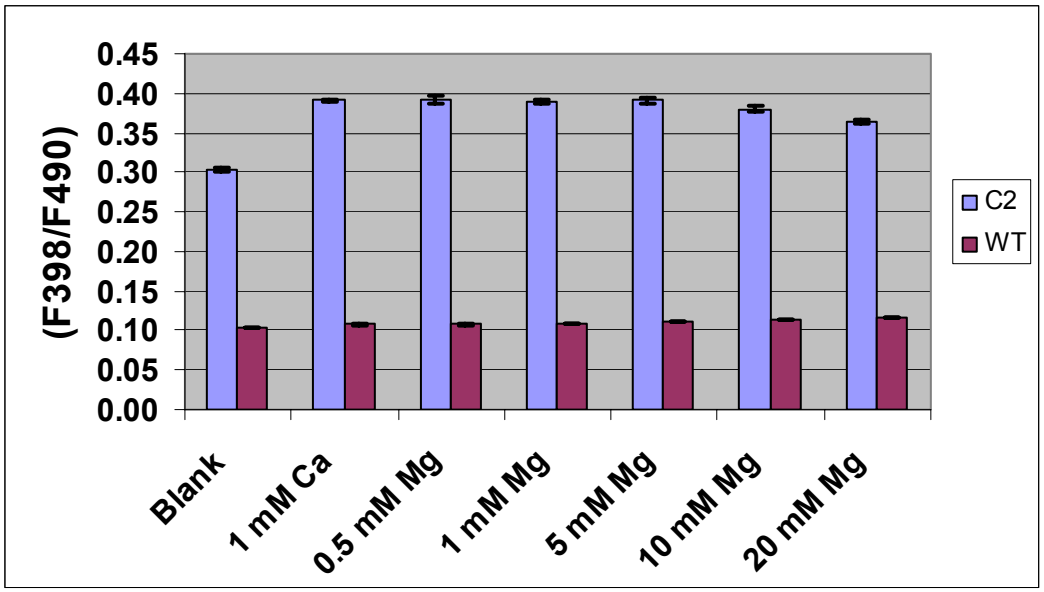


Figure 5.30: Ratiometric change (F398/F490) between 1.0 μM EGFPwtF and C2 variant to evaluate selectivity between Ca^{2+} and Mg^{2+}.

Error bars shown in Figure 5.30 indicate standard deviation. The red-shift that was previously noted following addition of Gd^{3+} did not occur in the titrations of Mg^{2+} in competition with Ca^{2+} for either EGFPwtF or EGFP-C2 as seen in Figures 5.31-5.34. In fact, very little change in fluorescence intensity is observe for Mg^{2+} at either 490 nm (Figure 5.31) or 398 nm (Figure 5.33), even at 20 mM Mg^{2+}, which is 200-fold higher than the concentration of Gd^{3+} that produced a red-shift with EGFPwtF. This strongly supports the assumption that Gd^{3+} is involved in non-specific binding at lower concentrations, whereas Mg^{2+} and Ca^{2+} are not.

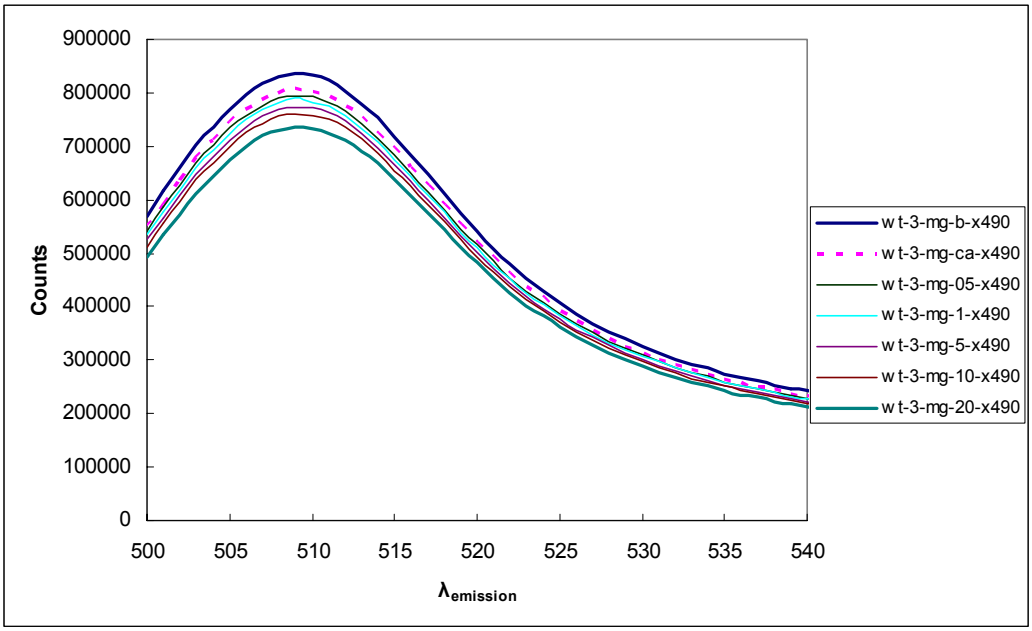


Figure 5.31: Fluorometric emission spectra of 1 µM EGFPwtF at 490 nm excitation wavelength. Competitive titration between 1 mM Ca^{2+} and 0-20 mM Mg^{2+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4.

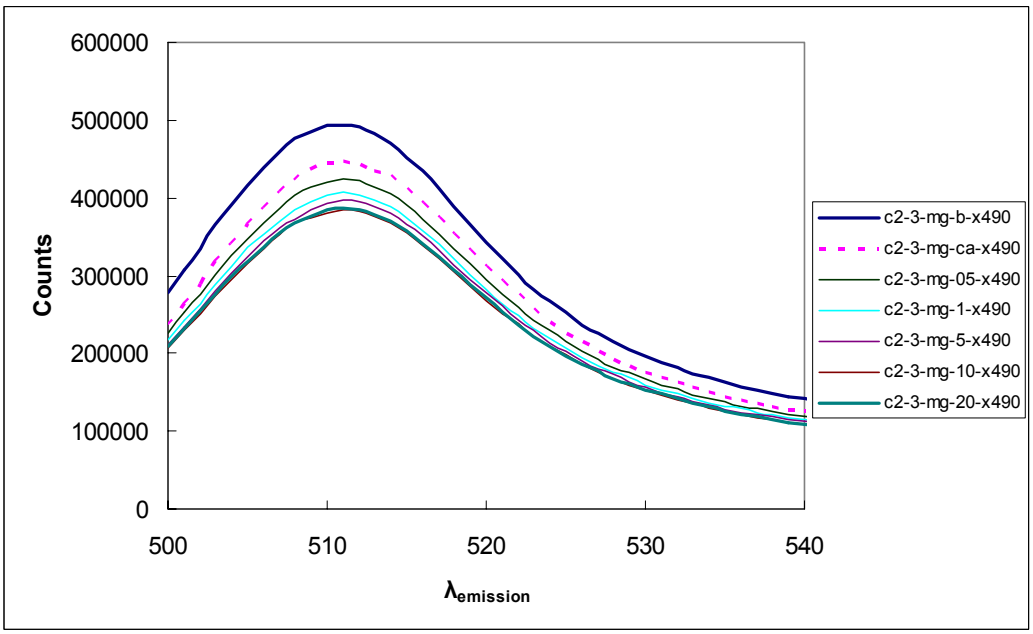


Figure 5.32: Fluorometric emission spectra of 1 µM EGFP-C2 at 490 nm excitation wavelength. Competitive titration between 1 mM Ca^{2+} and 0-20 mM Mg^{2+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4.

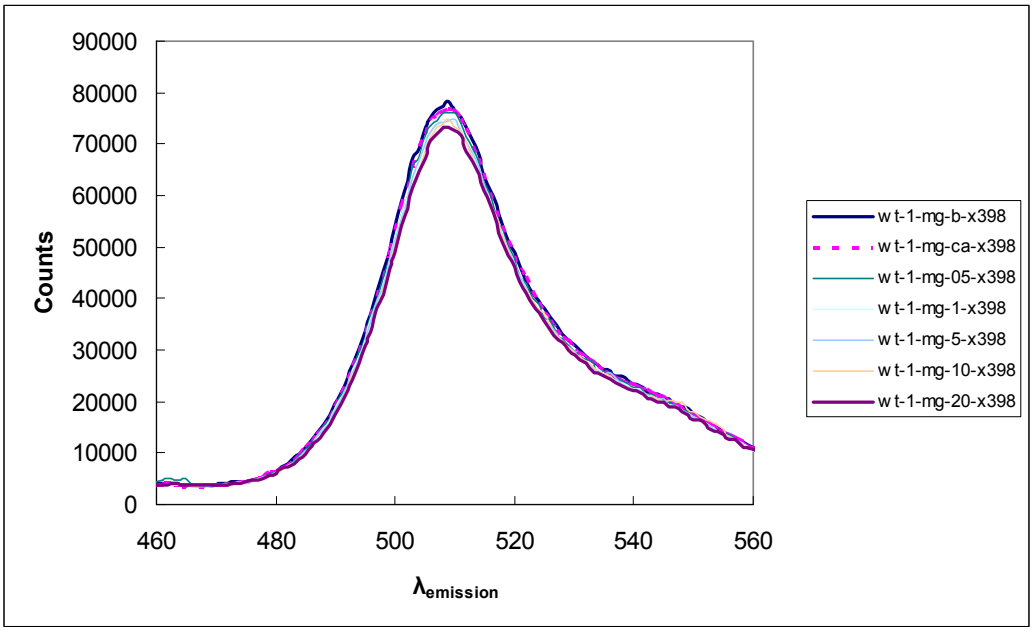


Figure 5.33: Fluorometric emission spectra of 1 µM EGFPwtF at 398 nm excitation wavelength. Competitive titration between 1 mM Ca^{2+} and 0-20 mM Mg^{2+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4.

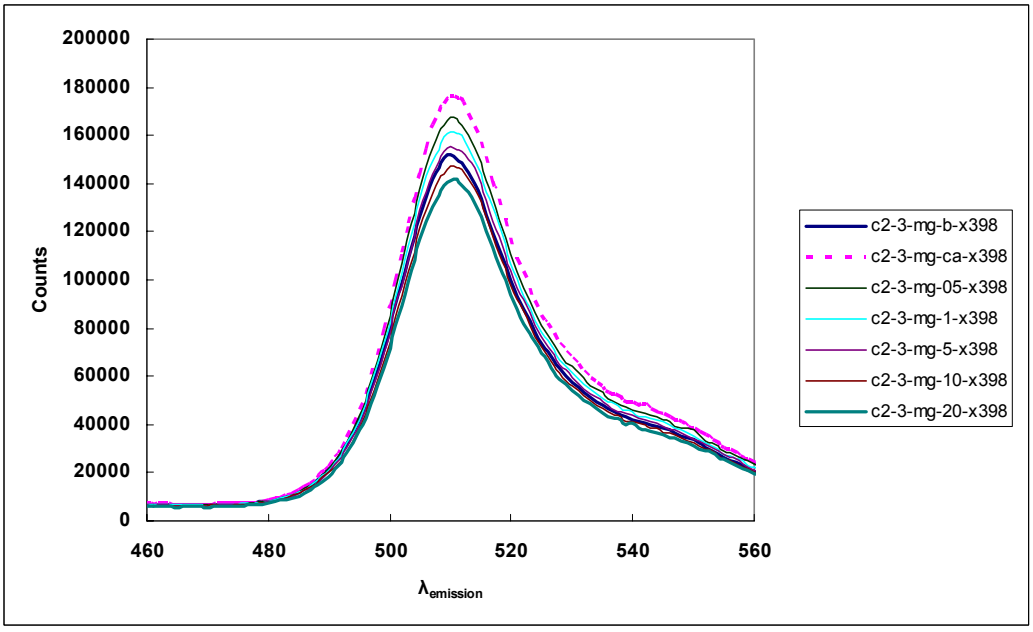
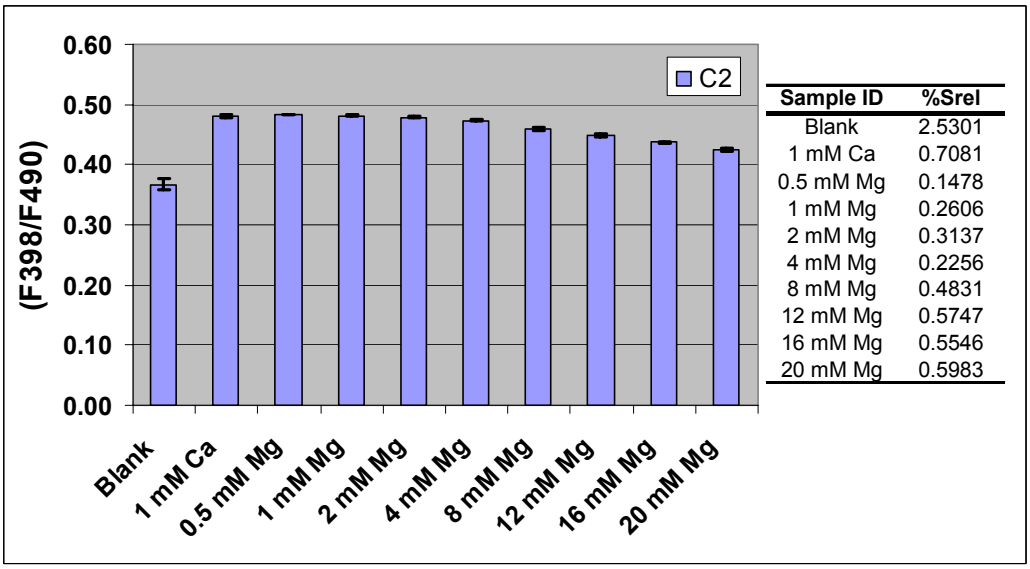


Figure 5.34: Fluorometric emission spectra of 1 µM EGFP-C2 at 398 nm excitation wavelength. Competitive titration between 1 mM Ca^{2+} and 0-20 mM Mg^{2+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4.

Next, the competitive titration was repeated across the same concentration range, this time with 8 plotted points for addition of Mg^{2+}, as shown in Figure 5.35. These trials were completed in triplicate and then averaged for the graph in Figure 5.35. As seen in the figure, the percent relative standard deviation (%Srel) was calculated to be less than 1.0% with respect to addition of Mg^{2+}, indicating excellent reproducibility.

Data from the 8 points covering the ratiometric change from 0-20 mM (Figure 5.35) were used to plot F Normalized using Equation 12. However, the data was insufficient to plot using curve-fitting based on equation 13, so K_d for Mg^{2+} was not calculated. Based on an inspection of the graph in Figure 5.35, the concentration of Mg^{2+} added was not sufficient to reduce the ratio F398/F490 down to the original value prior to addition of Ca^{2+}. This correlation indicates that very high concentrations of Mg^{2+} would need to be present to evaluate this ratiometrically, and confirms the very low affinity of the protein for Mg^{2+}.



Sample ID	%Srel
Blank	2.5301
1 mM Ca	0.7081
0.5 mM Mg	0.1478
1 mM Mg	0.2606
2 mM Mg	0.3137
4 mM Mg	0.2256
8 mM Mg	0.4831
12 mM Mg	0.5747
16 mM Mg	0.5546
20 mM Mg	0.5983

Figure 5.35: Ratiometric change (F398/F490) for 1.0 µM C2 variant to evaluate Mg^{2+} affinity.

Finally, the UV-Vis absorbance was measured for both EGFPwtF and EGFP-C2, to evaluate the change in spectra following addition of Ca^{2+}, and the subsequent addition of Mg^{2+}, as shown in Figure 5.36 and Figure 5.37, respectively. At this protein concentration (1 μM), the ratiometric change in absorbances at 398 nm and 490 nm are insufficient for use as an analytical method. Based on these data and other analyses presented in this thesis, a valid method for use of UV-Vis remains to be developed.

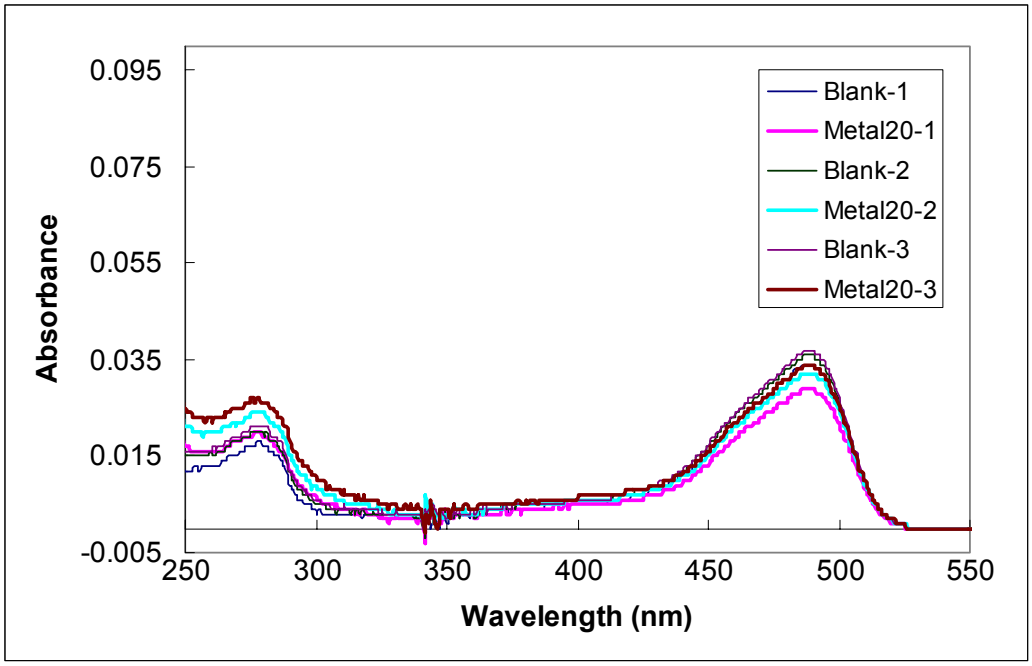


Figure 5.36: 1 μM EGFPwtF competitive titration absorbance spectra, 10 mM TRIS-Cl (pH 7.4), 1 mM Ca^{2+}, 20 mM Mg^{2+}. Blank indicates 1 mM Ca^{2+}. Metal20 indicates addition of 20 mM Mg^{2+}.

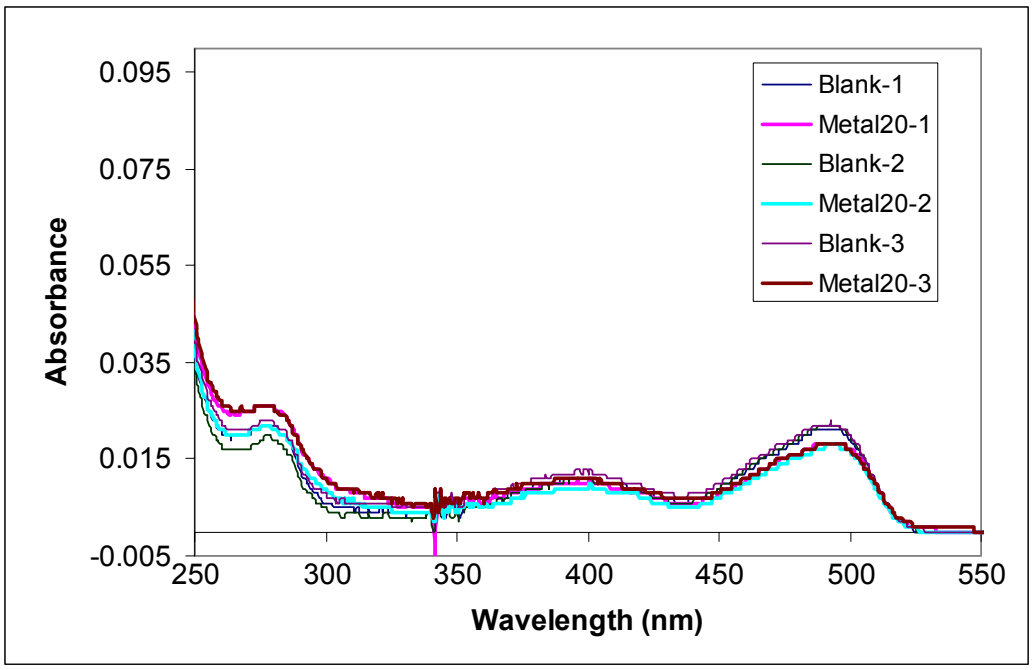


Figure 5.37: 1 µM EGFP-C2 competitive titration absorbance spectra, 10 mM TRIS-Cl (pH 7.4), 1 mM Ca^{2+}, 20 mM Mg^{2+}. Blank indicates 1 mM Ca^{2+}. Metal20 indicates addition of 20 mM Mg^{2+}.

5.3.2.5 Direct Addition of Pb^{2+} to EGFP-B-4 in TRIS-Cl

Figure 5.38 illustrates the direct titration of 1 µM Pb^{2+} to EGFP-B-4. As previously noted, this variant had exhibited the highest ratiometric response to the lanthanides evaluated. The EGFP-B-4 variant was diluted to 1 µM in 10 mM TRIS-Cl, at pH 7.4. A single aliquot of Pb^{2+} was added, for a solution concentration of 1 µM Pb^{2+}. EGTA was added to remove the metal ion. Three samples were evaluated. The addition of Pb^{2+}, even at this low concentration, resulted in a decrease in the ratio F398/F490. While it can be observed from the figure that the mean value following addition of Pb^{2+} falls within the standard deviation of the blank sample for this lower concentration of metal, this trend was later observed in the competitive titrations between Ca^{2+} and Pb^{2+},

where the addition of Pb^{2+} results in a decrease in the ratio F398/F490. It also suggests that the binding affinity of Pb^{2+} can be calculated by direct titration methods.

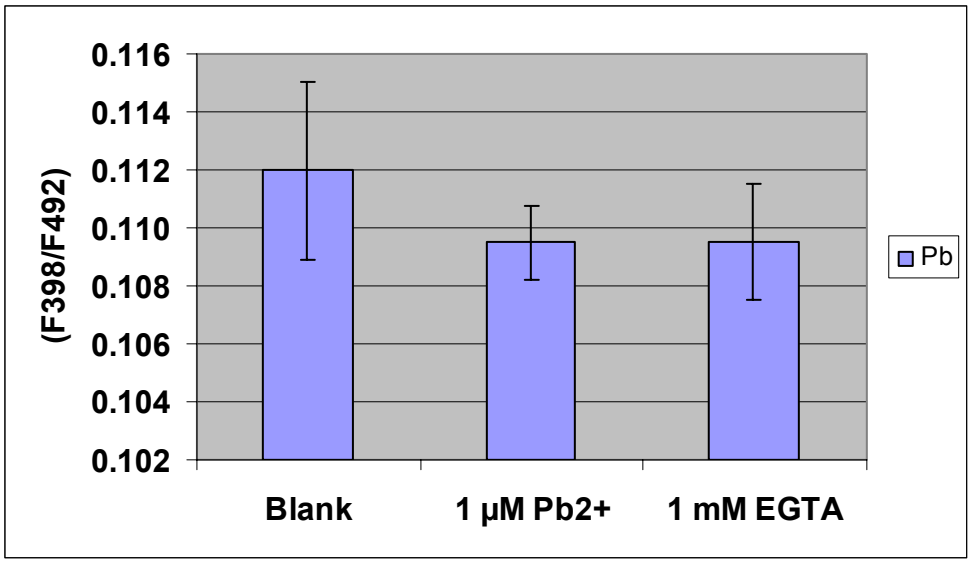


Figure 5.38: Direct addition of 1 µM Pb^{2+} to 1 µM EGFP-B-4 in 10 mM TRIS-Cl, pH 7.4.

5.3.2.6 Direct Addition of Pb^{2+} to EGFPwtF and Variants in Bis-TRIS-Cl

Results of direct titration of 1 mM Pb^{2+} with 1-2 µM EGFPwtF and variants EGFP-C2, EGFP-C-2, EGFP-C-3, and 5G-III-5G in Bis-TRIS-Cl buffer at pH 7.0 can be observed in Figure 5.39. Precipitation of $Pb(OH)_2$ was observed at 1.0 mM Pb^{2+}.As compared to the other variants evaluated, the wild type EGFPwtF exhibited little or no change in the ratio F398/F490 following addition of Pb^{2+}, which is consistent with results observed for Gd^{3+}, Eu^{3+}, La^{3+} and Tb^{3+} at lower concentrations, but not for higher concentrations. In fact, data presented in this thesis for the competitive titration of Pb^{2+} with Ca^{2+} in TRIS-Cl indicates a large increase in the ratio F398/F490 for EGFPwtF following addition of 0.5 mM Pb^{2+}. The difference between these two experiments may be due to the difference in buffers, which are later evaluated.

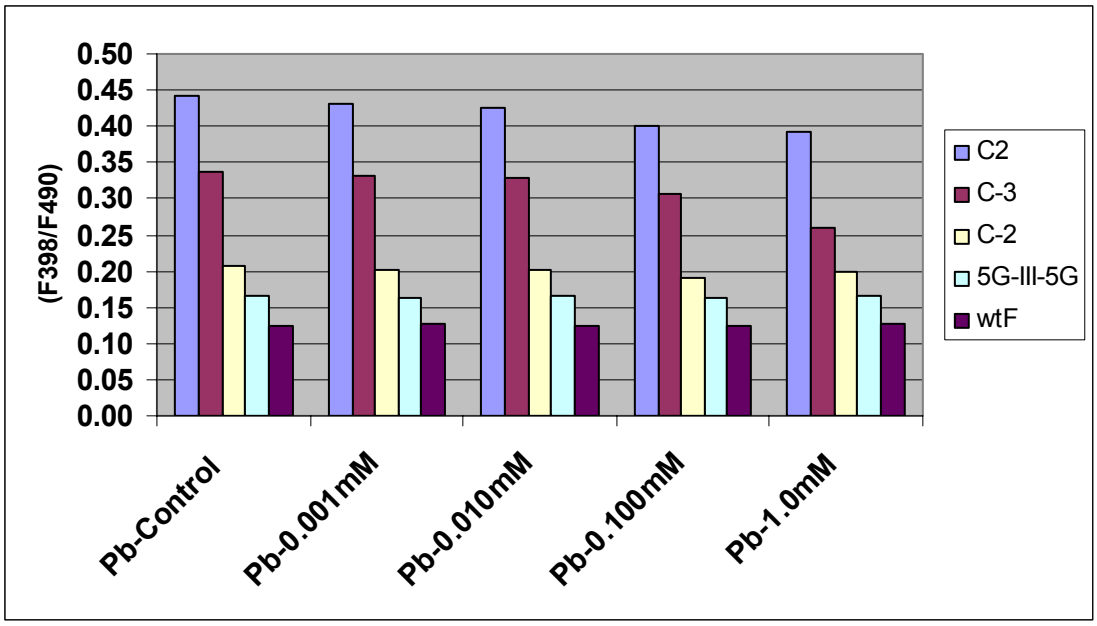


Figure 5.39: Ratiometric analysis of EGFP variants with direct titration of Pb^{2+}. $Pb(OH)_2$ precipitation observed at 1.0 mM Pb^{2+}. Protein concentrations, 1-2 μM; Bis-TRIS-Cl buffer, pH 7.0.

It can also be seen that no apparent change is observed in the 5G-III-5G variant. This is likely due to two factors. First, the calcium-binding loop is more distant from the protein surface, so the binding of a metal ion may produce less conformational change. Second, this variant only includes the grafted loop from CaM-III, not the flanking helices. Previous work has indicated that the binding affinity for the graft combinations occurs in the order E-III-F > III-F > E-III > E [7], and this trend is reflected in Figure 5.39, where the ratiometric change for EGFP-C-3 (III-F) is more significant than that for EGFP-C-2 (E-III). The C2 variant also shows a more significant ratiometric decrease.

The results of this analysis indicate that variants C2, C-2, and C-3 show potential for high binding affinities with Pb^{2+}. This is further examined in the competitive titration between Ca^{2+} and Pb^{2+}.

5.3.2.7 Evaluation of Non-specific Pb^{2+} binding to EGFPwtF

Figure 5.40 demonstrates the effect of adding 10 mM KCl to EGFPwtF following addition of 100 µM of Pb^{2+}. The ratio F398/F490 is reduced nearly to the starting level, prior to addition of Pb^{2+}. The effect presented here supports the assumption of non-specific binding on the wild-type protein surface, which is disrupted by addition of salt.

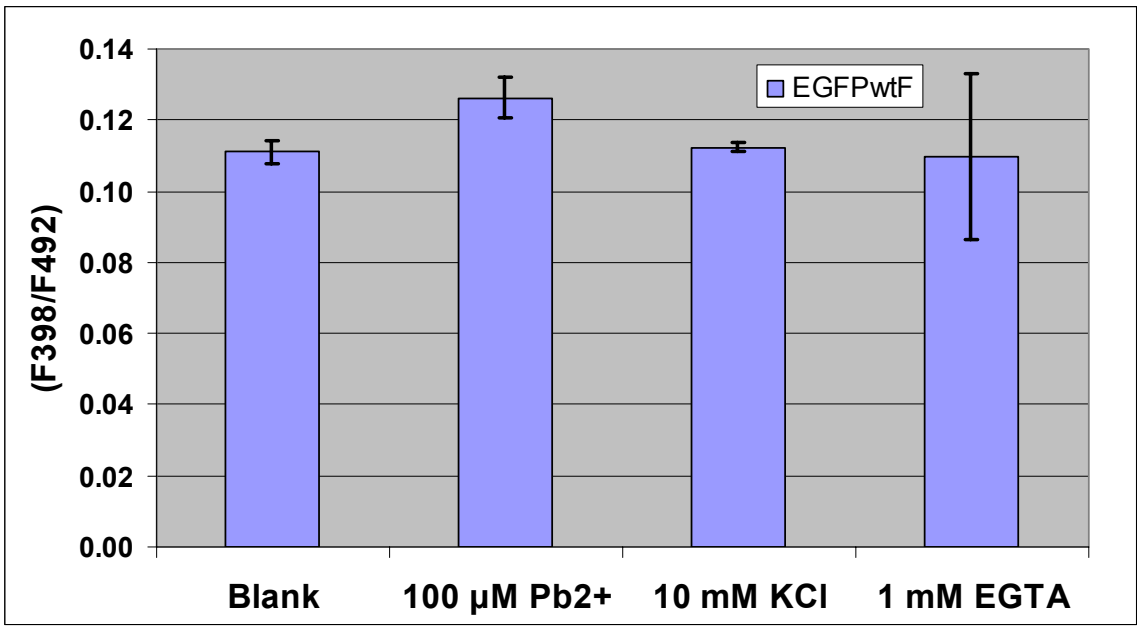


Figure 5.40: Ratiometric analysis of EGFPwtF variants with direct titration of Pb^{2+} followed by addition of 10 mM KCl. 1.0 µM EGFPwtF, 10 mM TRIS-Cl buffer, pH 7.0.

5.3.2.8 Selectivity Titrations between Pb^{2+} and Ca^{2+}

Figure 5.41 compares the fluorescent ratiometric change (F398/F490) for EGFPwtF and EGFP-C2 during the competitive titration of Ca^{2+} with Pb^{2+} in 10 mM TRIS-Cl, pH 7.4. Error bars in the figure indicate standard deviation. Mean and standard deviation values are presented in Table 5.4. As indicated in previous experiments, the wild-type was used as a control. In Figure 5.41 it can be seen that no change for EGFPwtF is apparent up to 100 µM, as was observed with the other metals evaluated.

However, at 500 µM, the ratio for EGFPwtF begins to increase, and a corresponding precipitation was observed during the analysis. A 1-2 nm red-shift was also observed at this concentration of Pb^{2+} (data not shown).

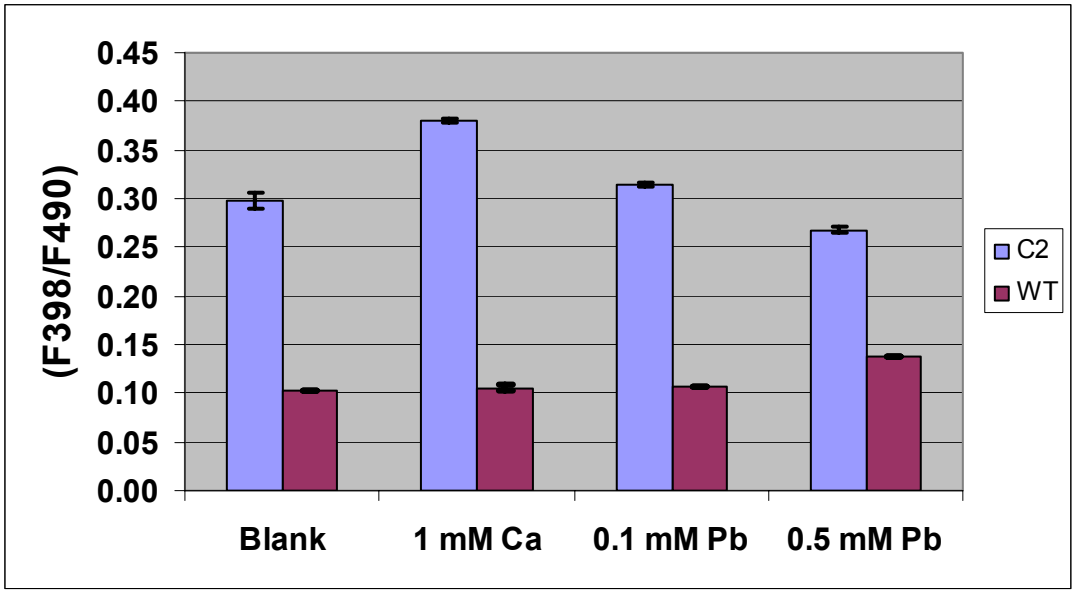


Figure 5.41: Ratiometric change (F398/F490) in 10 mM TRIS-Cl, pH 7.4, between 1.0 µM EGFPwtF and C2 variant to evaluate selectivity between Ca^{2+} and Pb^{2+}.

Table 5.4: F398/F490 Mean and Standard Deviation Values For EGFP-C2 and EGFPwtF During Competitive Titrations.

	C2	WT
Blank	0.30 ± 0.01	0.10 ± 0.00
1 mM Ca	0.38 ± 0.00	0.11 ± 0.00
0.1 mM Pb	0.31 ± 0.00	0.11 ± 0.00
0.5 mM Pb	0.27 ± 0.00	0.14 ± 0.00

A significant increase in the ratio was observed following addition of Ca^{2+}, which was reduced almost to the initial ratiometric value following addition of 100 µM Pb^{2+}. The addition of Pb^{2+} to a final concentration of 500 µM further reduced the ratio below the initial value for the blank sample. The results presented in this figure strongly indicate that Pb^{2+} displaces Ca^{2+} from the C2 variant.

This method was then duplicated with EGFPwtF and EGFP-C2 in 10 mM TRIS-Cl, pH 7.4. A total of 8 aliquots of Pb^{2+} covering a concentration range from 1.0 µM to 100 µM were added. Error bars in the figure represent standard deviation. The reduction in F398/F490 following addition of Pb^{2+} relative to the initial ratio for Ca^{2+} was considerably less in the Bis-TRIS-Cl buffer, as seen in Figure 5.42, than what was observed in Figure 5.41 for TRIS-Cl. This was consistent with the results obtained following direct addition of Pb^{2+} to other EGFP variants, as previously demonstrated in this work. Note that the standard deviation was less than 0.01 for EGFPwtF, as shown in Table 5.5.

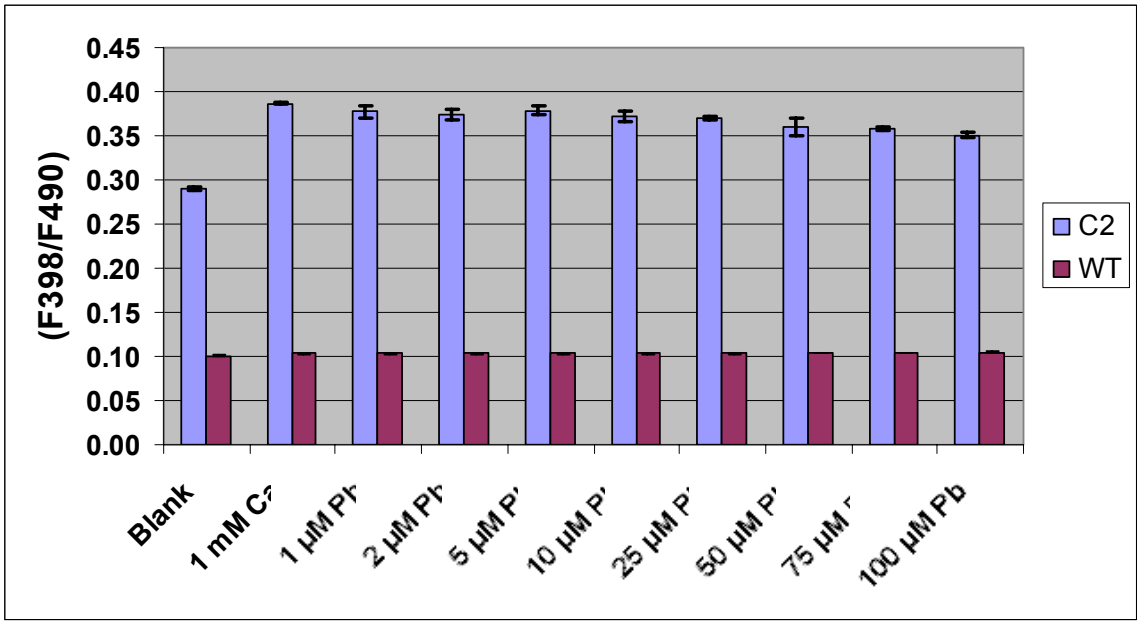


Figure 5.42: Ratiometric change (F398/F490) in 10 mM Bis-TRIS-Cl, pH 7.0, between 1.0 µM EGFPwtF and C2 variant to evaluate selectivity between Ca^{2+} and Pb^{2+}.

Table 5.5: F398/F490 Mean and Standard Deviation Values For EGFP-C2 and EGFPwtF During Competitive Titrations in 10 mM Bis-TRIS-Cl.

	C2	WT
Blank	0.29 ± 0.00	0.10 ± 0.00
1 mM Ca	0.39 ± 0.00	0.10 ± 0.00
1 µM Pb	0.38 ± 0.01	0.10 ± 0.00
2 µM Pb	0.37 ± 0.01	0.10 ± 0.00
5 µM Pb	0.38 ± 0.00	0.10 ± 0.00
10 µM Pb	0.37 ± 0.01	0.10 ± 0.00
25 µM Pb	0.37 ± 0.00	0.10 ± 0.00
50 µM Pb	0.36 ± 0.01	0.10 ± 0.00
75 µM Pb	0.36 ± 0.00	0.10 ± 0.00
100 µM Pb	0.35 ± 0.00	0.10 ± 0.00

Next, the procedure followed for analysis of EGFPwtF and EGFP-C2 in Bis-TRIS-Cl was once again repeated in TRIS-Cl, pH 7.4. In this analysis, EGFPwtF was not required, as the concentration range for Pb^{2+} had already been evaluated. Instead, two variants were evaluated in parallel: EGFP-C2 and EGFP-C-4. It was assumed that these variants, which both included the complete E-III-F binding site, would exhibit the highest binding affinities. Again, 1 mM Ca^{2+} was added to the protein followed by aliquots of Pb^{2+} to obtain protein concentrations at the following points: 1.0, 2.0, 5.0, 10.0, 25.0, 50.0, 75.0 and 100.0 µM. Results of these two analyses can be seen in Figure 5.43, with the calculated Mean, Standard Deviation, and percent Relative Standard Deviation (%Srel) values presented in Table 5.6.

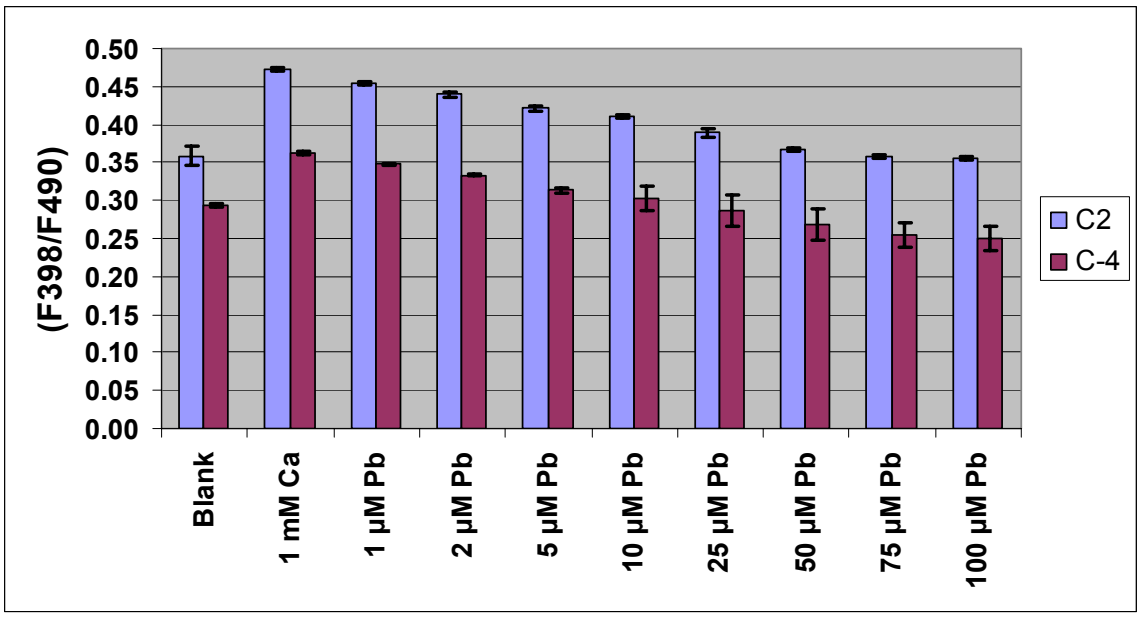


Figure 5.43: Ratiometric change (F398/F490) to evaluate selectivity between Ca^{2+} and Pb^{2+}. 1.0 µM C2 and EF172 in 10 mM TRIS-Cl buffer, pH 7.4.

Table 5.6: Mean, Standard Deviation, and Percent Relative Standard Deviation for Addition of Pb^{2+} to EGFP Variants C2 and C-4.

Sample ID	C2	C2 %Srel	C-4	C-4 %Srel
Blank	0.36 ± 0.01	3.6	0.29 ± 0.00	0.57
1 mM Ca	0.47 ± 0.00	0.52	0.36 ± 0.00	0.53
1 µM Pb	0.45 ± 0.00	0.41	0.35 ± 0.00	0.33
2 µM Pb	0.44 ± 0.00	0.80	0.33 ± 0.00	0.34
5 µM Pb	0.42 ± 0.00	0.81	0.31 ± 0.00	0.95
10 µM Pb	0.41 ± 0.00	0.63	0.30 ± 0.02	5.1
25 µM Pb	0.39 ± 0.01	1.4	0.29 ± 0.02	7.3
50 µM Pb	0.37 ± 0.00	0.45	0.27 ± 0.02	7.7
75 µM Pb	0.36 ± 0.00	0.82	0.26 ± 0.02	6.3
100 µM Pb	0.35 ± 0.00	0.73	0.25 ± 0.02	6.2

Data from Figure 5.43 covering the ratiometric change from 0-100 µM for variants C2 and C-4 was used to plot F Normalized using Equation 12, and the data was plotted using curve-fitting based on equation 13, as seen in Figure 5.44 and Figure 5.45.

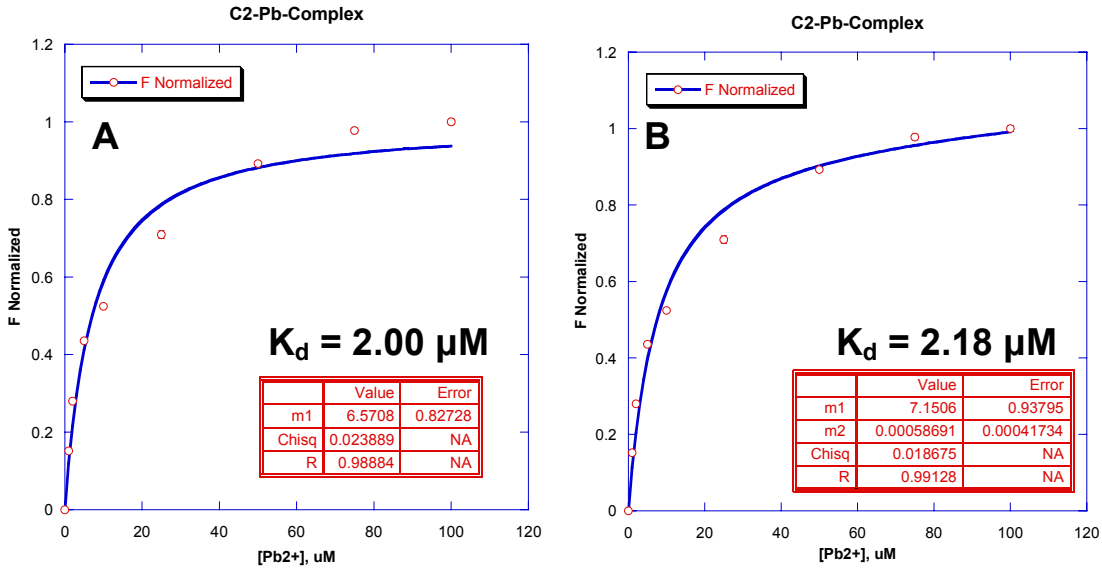


Figure 5.44: Curve fitting to determine K_d for Pb^{2+} with EGFP-C2. Plot A does not include term for non-specific binding, as shown in plot B. The value m1 represents K overall.

In Figure 5.44, K_d for Pb^{2+} with EGFP-C2 was calculated using Equation 14. Plot A in Figure 5.44 shows the curve-fitting when the term for non-specific binding is omitted from the calculation, and Plot B shows the curve when this term is included. The K_d values from A and B were calculated as 2.00 µM and 2.18 µM, respectively. Correlation (*R*) of the curve-fitting in Plot B is higher than in Plot A (0.99128 vs. 0.98884), which supports the premise of non-specific binding.

For EGFP-C-4 (Figure 5.45), the K_d for Pb^{2+} was calculated in Plot A to be 3.22 µM (excluding term for non-specific binding), and in Plot B to be 3.42 µM (including term for non-specific binding). Curve-fitting obtained in Plot B was marginally better than that observed in Plot A, with Correlation (*R*) values of 0.98976 and 0.98860, respectively. The difference in the two plots still suggests the probability of non-specific

binding, however, the results with this protein variant demonstrated less precision than with the C2 variant.

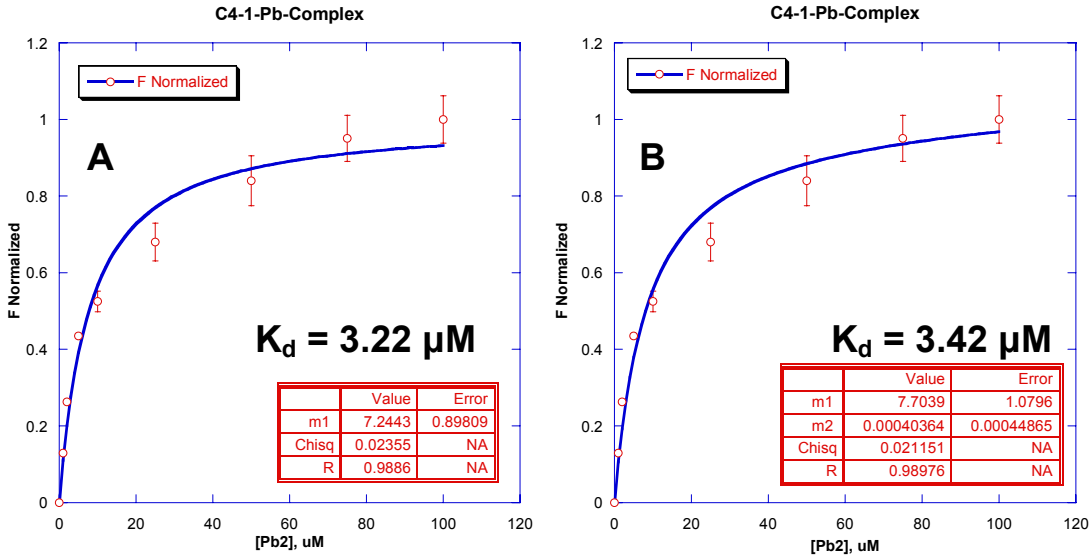


Figure 5.45: Curve fitting to determine K_d for Pb^{2+} with EGFP-C-4. Plot A does not include term for non-specific binding, as shown in plot B. The value m1 represents K overall.

In both cases (C2 and C-4), the calculated binding affinities were orders of magnitude higher than the affinity for Ca^{2+}, as the calcium K_d values previously established for EGFP-C2 and EGFP-C-4 were 440 µM and 800 µM, respectively. As was the case with Gd^{3+}, the data suggests that Pb^{2+} will readily displace Ca^{2+} in the calcium binding site, with significantly greater affinity.

Finally, the fluorescent spectra for EGFP-C2 and EGFP-C-4 at excitation wavelengths 398 nm and 490 nm are presented in Figures 5.46-5.49. The red-shift that was previously noted for Gd^{3+} began to appear at ~ 5 µM Pb^{2+}, as illustrated in Figures 5.46 and 5.48. This is also strong evidence to support a non-specific binding mechanism.

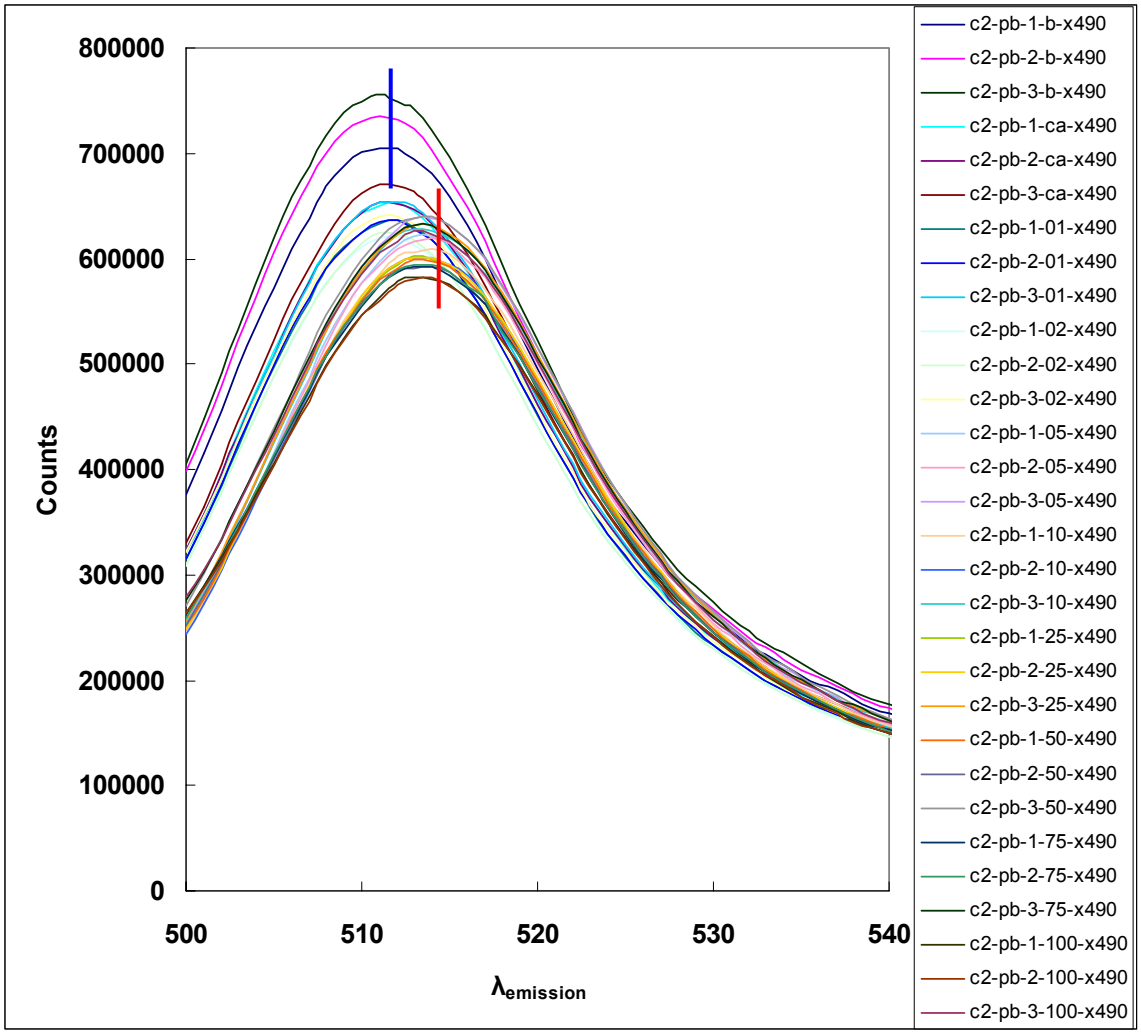


Figure 5.46: 490 nm fluorometric emission scans of 1 µM EGFP-C2 variant, competitive titration between 1 mM Ca^{2+} and 0-100 µM Pb^{2+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4. Red-shift observed at 5 µm Pb^{2+} from 511 nm to 513 nm.

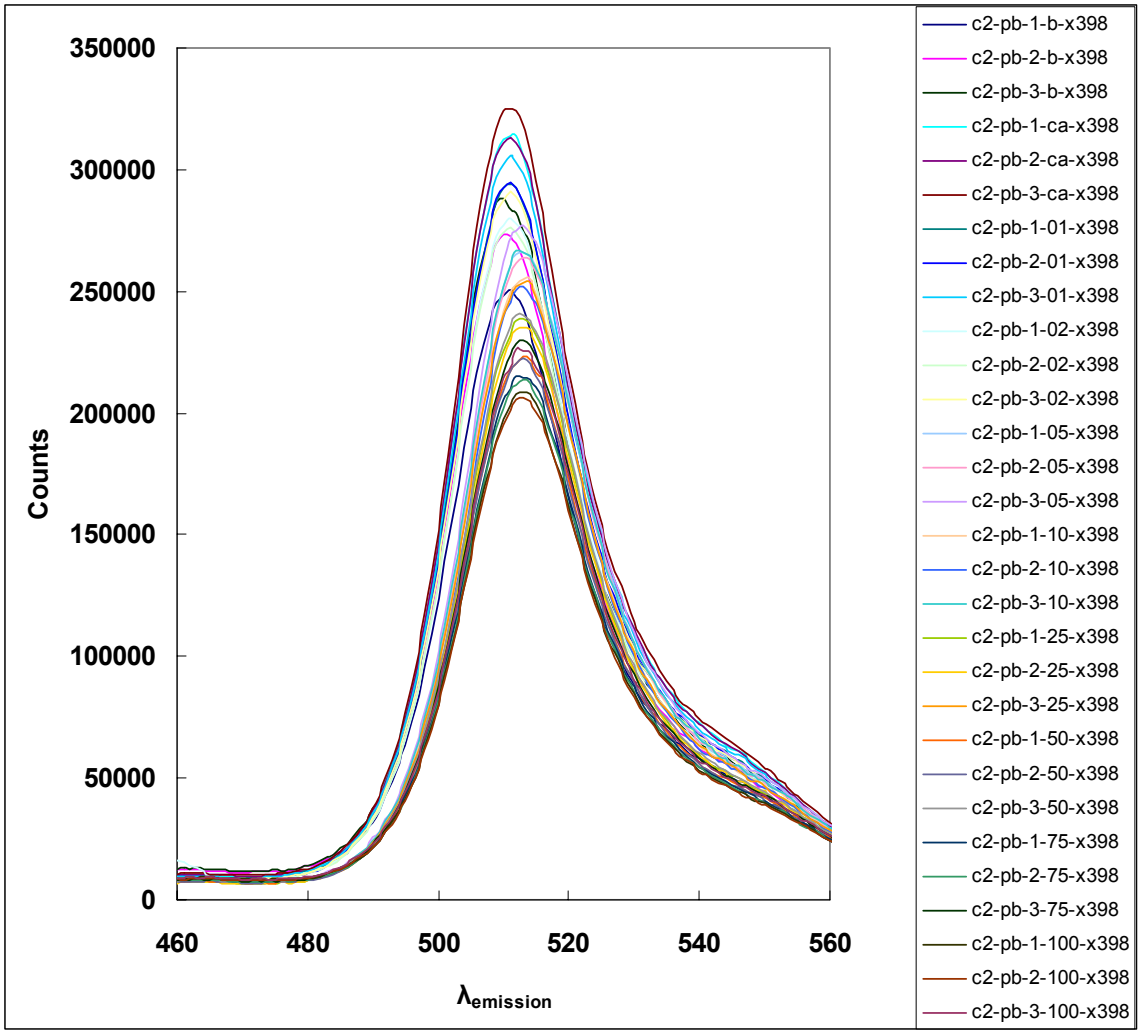


Figure 5.47: 398 nm fluorometric emission scans of 1 µM EGFP-C2 variant, competitive titration between 1 mM Ca^{2+} and 0-100 µM Pb^{2+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4. Red-shift observed at 5 µm Pb^{2+} from 511 nm to 513 nm.

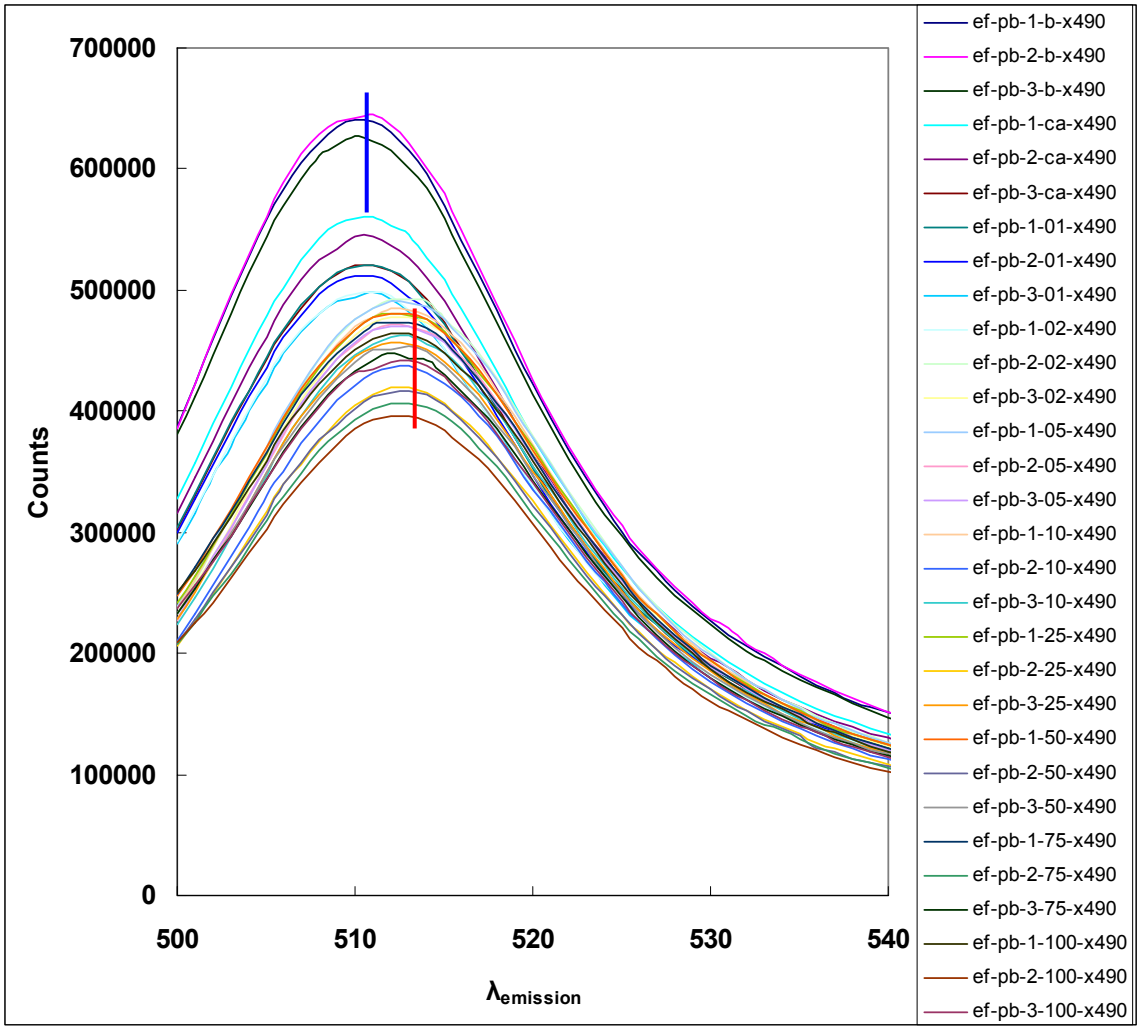


Figure 5.48: 490 nm fluorometric emission scans of 1 µM EGFPwtF EF172 variant, competitive titration between 1 mM Ca^{2+} and 0-100 µM Pb^{2+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4. Decreasing Red-shift observed at 5 µm Pb^{2+} from 511 nm to 513 nm.

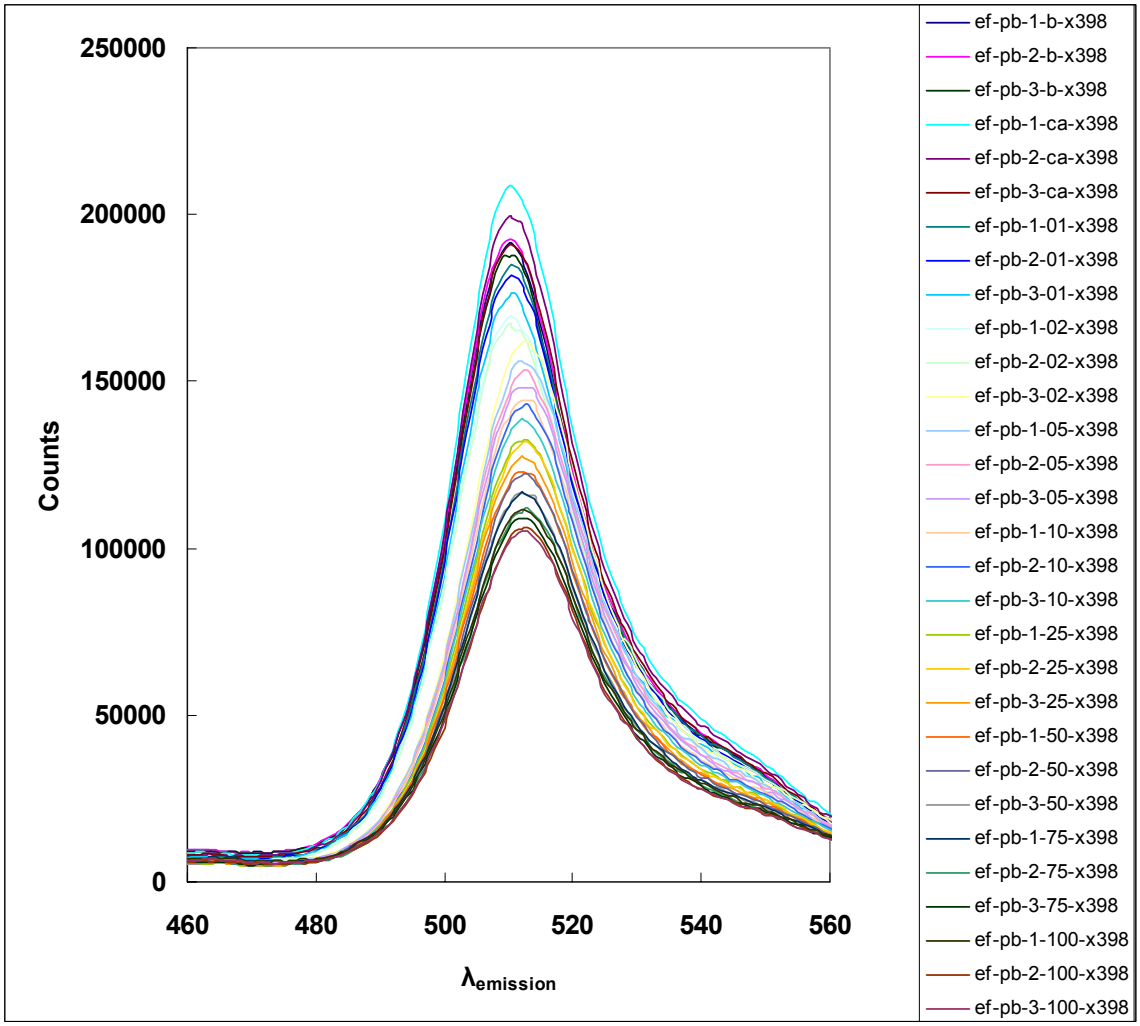


Figure 5.49: 398 nm fluorometric emission scans of 1 µM EGFP-C-4 variant, competitive titration between 1 mM Ca^{2+} and 0-100 µM Pb^{2+} in 10 mM chelexed TRIS-Cl buffer, pH 7.4. Red-shift observed at 5 µm Pb^{2+} from 511 nm to 513 nm.

5.4 Conclusions

A number of conclusions can be made from the data presented in this section regarding metalloprotein complex formation with different metals. The validity of the methods applied is supported by observations that fluorescent changes with the addition of Ca^{2+} are only observed in the engineered variants that include a calcium-binding loop, and not in the wild-type.

The ratiometric changes observed with respect to metal binding suggest that ionic radius may not be a controlling factor in binding. The ratiometric changes produced by the lanthanides, with smaller ionic radii than Pb^{2+} (Table 5.7), exhibit a similar decreasing trend following displacement of Ca^{2+}. Based on these data, it is possible that charge has more effect in this case than ionic radii, although this assumption will require further quantification.

Table 5.7: Summary of Selected Metal Properties.

Z	Ion	[a]Ionic Radii (Å)	[a]Pauling EN	Electron Configuration	Acid Type	Ligand Preference
11	Na^{+}	1.02	0.93	$[Ne]3s^{0}$	Hard	Oxygen
12	Mg^{2+}	0.72	1.31	$[Ne]3s^{0}$	Hard	Oxygen
19	K^{+}	1.38	0.82	$[Ar]4s^{0}$	Hard	Oxygen
20	Ca^{2+}	0.99	1	$[Ar]4s^{0}$	Hard	Oxygen
48	Cd^{2+}	0.97	1.69	$[Kr]4d^{10}5s^{0}$	Soft	Sulfur
57	La^{3+}	1.06	1.1	$[Xe]5p^{5}d^{1}6s^{0}$	Hard	Oxygen
63	Eu^{3+}	0.95	1.2	$[Xe]4f^{6}5s^{2}p^{6}6s^{0}$	Hard	Oxygen
64	Gd^{3+}	0.94	1.2	$[Xe]4f^{7}5p^{6}d^{0}6s^{0}$	Hard	Oxygen
65	Tb^{3+}	0.92	1.2	$[Xe]4f^{8}6s^{0}$	Hard	Oxygen
82	Pb^{2+}	1.19	2.33	$[Xe]4f^{14}5d^{10}6s^{2}6p^{0}$	Borderline	Nitrogen

[a]http://environmentalchemistry.com/yogi/periodic/

As previously noted, the calculated binding affinities for Pb^{2+} and Gd^{3+} were approximately 200 times higher than the affinities calculated for Ca^{2+}. From a toxicological perspective, this indicates that the introduction of toxic metals can significantly alter the function associated with calcium-binding proteins. Related to this were the observed red-shifts following addition of Pb^{2+} and the lanthanides. Because this red-shift was absent for both Ca^{2+} and Mg^{2+}, both of which are known to exhibit low affinity for non-specific binding, it seems apparent that these toxic metals will not only displace Ca^{2+} in the binding site, but will also bind non-specifically on the surface. This

conclusion is supported by the following: (1) the observed ratiometric change following addition of Gd^{3+}, Eu^{3+}, La^{3+}, Tb^{3+} and Pb^{2+} to EGFPwtF which has no grafted binding site; (2) the previously noted red-shift with increasing concentration of Gd^{3+} and Pb^{2+}; (3) the signal reduction observed by adding high concentration of KCl to displace 100 μM Pb^{2+} from EGFPwtF; (4) the results of the database study which demonstrated that Pb^{2+} will bind to the crystalline surface with as few as 1-2 ligands, and; (5) the absence of $Pb(OH)_2$ at high concentrations. This last example illustrates the difficulty in working with Pb^{2+}.

Lead hydroxide has a K_{sp} value at or near 1.43×10^{-20}. Because this is pH dependent, the maximum concentration for Pb^{2+} was calculated at pH 7.4 and 7.0 to be 227 nM, and 1.43 μM, respectively, using the Equation 15:

$$[Pb^{2+}] = 1.43 \times 10^{-20}/[OH^-]^2 \qquad \text{(Equation 15)}$$

Therefore, at pH 7.4 for 1 μM of protein, and assuming that Pb^{2+} forms a 1:1 complex with the protein in the same manner as calcium, precipitation is likely to be observed at or near 1.23 μM. However, in the studies presented in this thesis, precipitation was not observed in the presence of protein until after Pb^{2+} concentrations had exceeded 100 μM. This is strong evidence indicating that the higher concentration of Pb^{2+} remains soluble in the presence of proteins, and that binding is more than 1:1. It was suggested that this result may be caused by the formation of aqueous $PbCl_4^{2-}$, since acidification of the TRIS buffer was accomplished using HCl, however, there are two good arguments to refute this. First, the addition of 100 μM Pb^{2+} to the buffer, in the absence of protein, results in formation of a precipitate. Second, the formation of $PbCl_4^{2-}$ depends on increasing concentrations of Cl^- following precipitation of $PbCl_2$: in other

words, after $PbCl_2$ precipitates, an increase in the concentration of Cl^- ions creates the aqueous species $PbCl_4^{2-}$. In these analyses, however, $PbCl_2(s)$, which has a K_{sp} of 1.7 x 10^{-5} was not observed to form, and would only form after $Pb(OH)_2(s)$, since the K_{sp} for the latter is 1.43 x 10^{-20}. Furthermore, based on the K_{sp} of 1.7 x 10^{-5} for $PbCl_2(s)$, the concentration of Cl^- required for precipitation would be 412 mM. To determine the possibility of $PbCl_2(s)$ formation, the concentration of Cl^- added as HCl to adjust pH was then calculated. First, the ratio (H-TRIS/TRIS) was calculated using Equation 16, for both the initial and final pH values (10 and 7.4, respectively), based on the TRIS pK_a of 8.06, where H-TRIS represents the protonated state, and TRIS the deprotonated state.

$$pK_a = pH + \log (\text{H-TRIS/TRIS}) \qquad \text{(Equation 16)}$$

The ratio (H-TRIS/TRIS) was found to be 0.011 and 4.57 for the initial and final pH values, respectively. Because the total TRIS concentration is the sum of the two species (Equation 17), and $[TRIS]_{total}$ is known (10 mM), the ratio (H-TRIS/TRIS) is then used to calculate the mole fraction of each species.

$$[TRIS]_{total} = [\text{H-TRIS}] + [TRIS] \qquad \text{(Equation 17)}$$

At the initial pH, $[TRIS]_{total}$ was found to be 0.1 mM H-TRIS, and 9.9 mM TRIS. At the final pH, $[TRIS]_{total}$ was found to be 8.2 mM H-TRIS, and 1.8 mM TRIS. Calculating the difference between final [H-TRIS] and initial [H-TRIS], which represents total $[H^+]$, and assuming a 1:1 molar ratio between $[H^+]$ and $[Cl^-]$, gives a final concentration for $[Cl^-]$ of 8.1 mM. This is clearly insufficient for the formation of $PbCl_2(s)$.

The implication then of complex formation between proteins and toxic metals is that two potential modes of binding may be active at high metal concentrations. In the

first, competitive binding displaces the beneficial ion from the binding site, resulting in changes to the proteins function. In the second, the toxic metal may bind non-specifically to the surface, thus increasing it's mobility within the organism, and later being released from the surface under changing salt environments.

6 Conclusions

6.1 Statistical and Structural Analysis of Ca^{2+}- and Pb^{2+}-binding

It has been generally accepted that Pb^{2+}, a borderline acid, should exhibit a preference for nitrogen ligands when bound to a protein[4]. However, the data summarized in Table 4.3 identify only 28 out of 252 potential binding ligands as nitrogen, with the majority of ligands represented by oxygen and sulfur. It seems clear from the statistical analysis that Pb^{2+} does not exhibit a preference for nitrogen, as suggested, but is more likely to utilize charge Glu or Asp oxygen atoms, similar to Ca^{2+}, which is probably due to the abundance of oxygen ligands compared to nitrogen or sulfur.

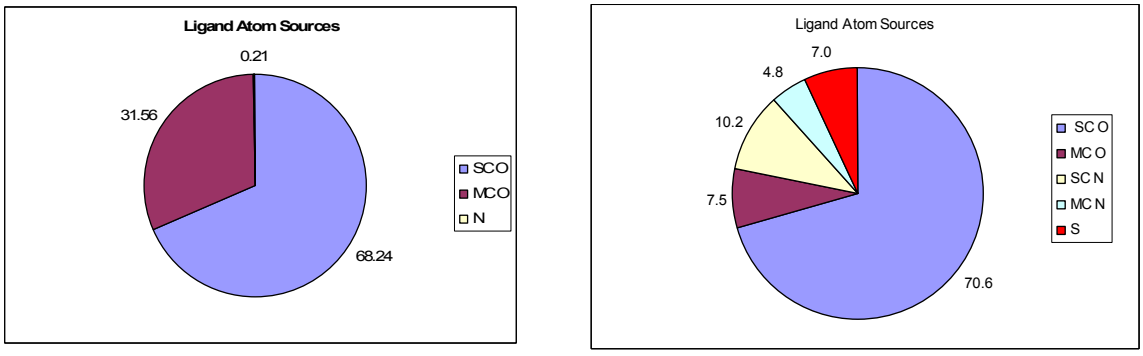


Figure 6.1: Comparison of ligand atoms sources between Ca^{2+} (left) and Pb^{2+} (right). SC indicates sidechain. MC indicates mainchain. O, N, S indicate oxygen, nitrogen, and sulfur, respectively.

The similarities is more clearly illustrated in Figure 6.1, where it is shown that, for both ions, approximately 70% of all binding ligands are sidechain oxygen atoms. Moreover, the distribution of ligand atoms from different amino acid residues is also similar, as shown in Figure 6.2. For Ca^{2+}, Asp is the most frequent ligand atom donor,

followed by Glu. For Pb^{2+}, Glu represents the majority, followed by Asp. Although the order of frequency is reversed, both ions bind to the same ligand types, which typically involve charged oxygen atoms.

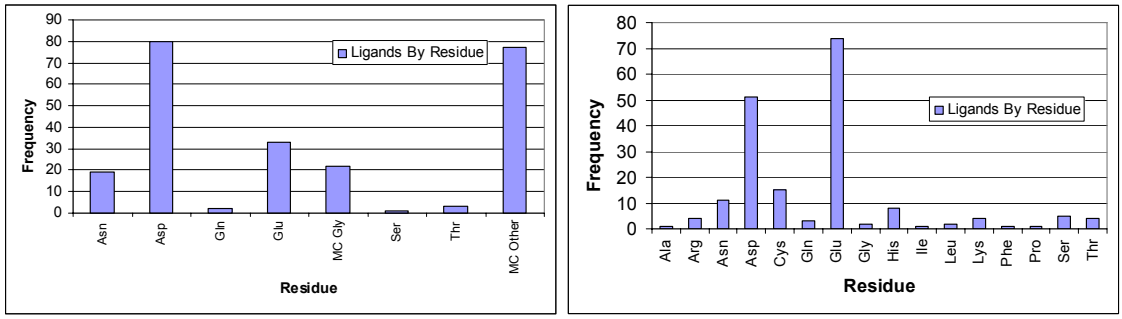


Figure 6.2: Comparison of ligand residues between Ca^{2+} (left) and Pb^{2+} (right).

From the data presented in this thesis, it is suggested that Pb^{2+} will utilized oxygen, nitrogen, or sulfur ligands when available, but the data does not support the conclusion that nitrogen is a preferred ligand.

Results of the statistical analysis of Ca^{2+} and Pb^{2+} also demonstrate that Pb^{2+} requires less charge within the binding site than Ca^{2+}, as shown in Figure 6.3. In fact, a comparison of the graphs in Figure 6.3 indicates an inverse trend between the two ions: with the frequency of charge increasing for calcium, while decreasing for lead, in the range of -1 to -3. This is probably due to the higher electronegativity of Pb^{2+} (2.33) compared to Ca^{2+} (1.0).

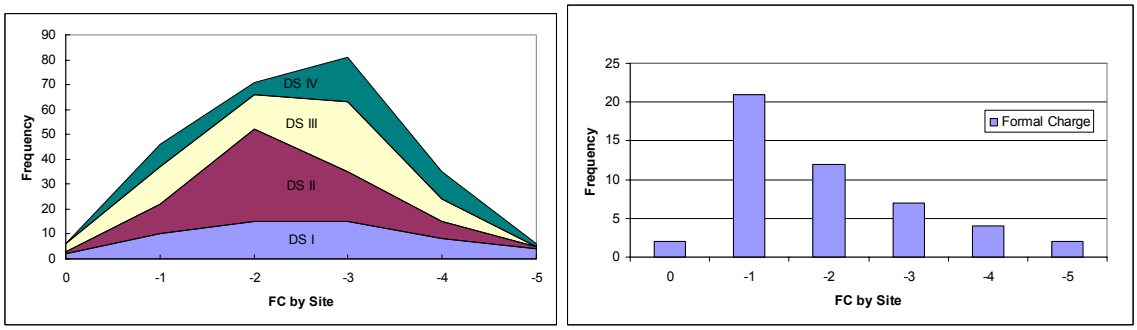


Figure 6.3: Comparison of charge between Ca^{2+} (left) and Pb^{2+} (right) in binding sites.

The observation that Pb^{2+} tends to bind with fewer ligands (2 to 5) than Ca^{2+} (5 to 7), and will bind over a wider range of ligand values, is reflected in Figure 6.4. This may also be related to the higher electronegativity of Pb^{2+}.

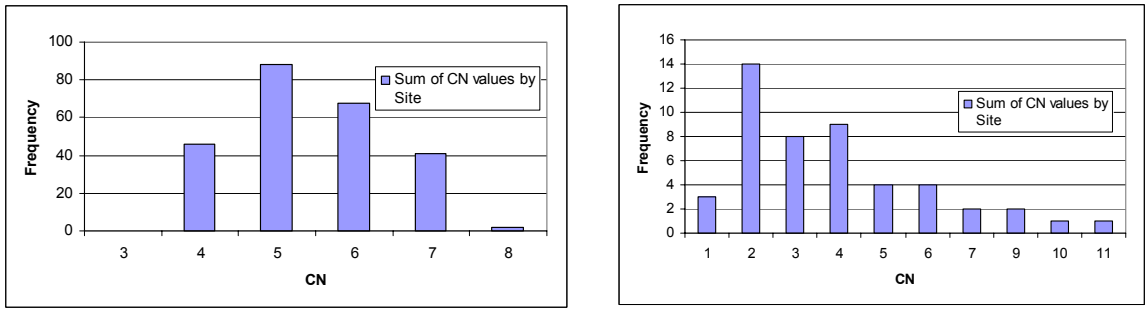


Figure 6.4: Comparison of Coordination Number values between Ca^{2+} (left) and Pb^{2+} (right) in binding sites.

Additionally, it was shown in the structural analysis (Section 4.3.2) that the Pb^{2+} ion will displace Ca^{2+} in the four EF binding sites of calmodulin, and that this displacement results in very minimal conformational stress of the binding site microenvironment. The structures shown in Figure 4.10 for 1EXR, Ca^{2+}-bound calmodulin, and 1N0Y, Pb^{2+}-bound calmodulin, indicate that the protein undergoes more pronounced conformational changes in the Pb^{2+} structure, which appears to be related to

binding of additional ions on the protein surface. Data to support this assumption were then presented in the analyses of fluorescent protein-metal complexes.

6.2 Analyses of Metal-binding Characteristics Using Engineered Fluorescent Proteins

Previous work related to the study of Pb^{2+}-binding has utilized x-ray crystallography [79], which presents a static picture of the metalloprotein complex that may differ from the same complex in solution, UV-Vis absorption spectroscopy, 1D and 2D DQ-COSY 1H NMR and circular dichroism Spectroscopy [74, 107]. The fluorescent ratiometric change method presented in this thesis provides another method for studying Pb^{2+}-binding in solution, and can be utilized to determine the relative binding affinities of different toxic metals.

New variants of EGFP were utilized to characterize the binding of different metals, which exhibited a fluorescent change subsequent to binding as a result of conformational changes that occur during binding. This approach has been utilized in our laboratory to isolate EF-Hand binding sites for study while limiting the effects of cooperativity, and in the development of sensors for *in vivo* monitoring of Ca^{2+} concentrations. Using these engineered proteins for the work reported in this thesis, the binding characteristics of different metals were compared. The wild-type variant, EGFPwtF, was utilized as a control, as this protein did not have a grafted binding site.

Verification of binding in the designated loop was presented in this thesis by the increased ratiometric change observed (398 nm/490 nm) following addition of Ca^{2+}. These changes were only observed, for Ca^{2+}, in variants that included the CaM EF-III binding loop, either with or without the flanking helices. No ratiometric change was

observed in the EGFPwtF variant with the addition of Ca^{2+}. It was also demonstrated that Pb^{2+} binds in the calcium-binding sites.

Next it was determined that Gd^{3+} and Pb^{2+} will displace Ca^{2+} in the binding sites. This was apparent in the competitive titrations discussed in sections 5.3.2.3 and 5.3.2.8. These data were then used to calculate binding affinities for both Gd^{3+} and Pb^{2+}. For the EGFP-C2 variant analyzed in this work, the binding affinities for Pb^{2+} and Gd^{3+} were both found to be 200 times higher than Ca^{2+}. This higher affinity, coupled with the conformational changes associated with the binding of these metals, suggests an important relationship to toxicity.

However, data collected from these studies also supported our hypothesis regarding non-specific binding as a second binding mechanism for both Pb^{2+} and Gd^{3+}. In the case of Pb^{2+}, an increased solubility was reported at concentrations of Pb^{2+} that far exceeded the K_{sp} for $Pb(OH)_2$. This was only observed in the presence of protein, including the wild-type variant, which exhibited ratiometric changes following the introduction of Pb^{2+}, Gd^{3+}, Eu^{3+}, La^{3+} and Tb^{3+}. Conversely, no ratiometric change was observed in the wild-type variant following the introduction of Ca^{2+} or Mg^{2+}, indicating that an additional binding mechanism was active in the case of the toxic metal evaluated. Furthermore, a 2-3 nm red-shift in the UV-Vis spectra was observed following the introduction of high concentrations of the toxic metals, which disappeared following addition of 1 mM EGTA to chelate the metal ions. Again, this phenomenon was not observed with the Ca^{2+} and Mg^{2+} ions. To further evaluate the possibility of non-specific binding, 100 μM Pb^{2+} was added to the EGFPwtF variant, which does not include a grafted binding loop. Figure 5.40 shows a reduction in the fluorescence ratiometric

change subsequent to addition of 10 mM KCl, which suggests that the Pb^{2+} ions were bound non-specifically to the protein, and the addition of salt was sufficient to remove these ions from the protein.

Finally, the curve-fitting plots for both Gd^{3+} and Pb^{2+} exhibited an improved fit when the equation included a term for non-specific binding. All of these data are consistent with the structural view of CaM depicted in 1N0Y, where the crystalline structure appears to have 7 to 10 additional Pb^{2+} ions bound to the protein opportunistically (i.e. – in non-specific binding relationships with the protein).

It is concluded from these data that the binding of the toxic metals evaluated includes two mechanisms: displacement of a beneficial metal ion in a known binding site, and non-specific binding that does not require the presence of a binding structure. These conclusions have important ramifications with respect to toxicity. First, metalloproteins may bind toxic metals with higher affinities, thus altering the function of the protein. This has been demonstrated in many studies [40, 60, 61, 67-79, 108], although affinities of metals such as Pb^{2+} and Gd^{3+} were not reported. Second, and more significantly, the potential for non-specific binding maintains higher solubility of toxic metals, as well as providing a mechanism for transport and diffusion of the metal, which may provide insight as to why increasing concentration of metals correlates with toxicity. Finally, the ability to bind non-specifically to proteins, including non-metalloproteins, increases the number of potential target proteins that may have their function altered due to the introduction of toxic metals.

7 References

1. Ibers, J.A. and R.H. Holm, *Modeling coordination sites in metallobiomolecules.* Science, 1980. **209**(4453): p. 223-35.
2. Holm, R.H., P. Kennepohl, and E.I. Solomon, *Structural and Functional Aspects of Metal Sites in Biology.* Chem Rev, 1996. **96**(7): p. 2239-2314.
3. Tainer, J.A., V.A. Roberts, and E.D. Getzoff, *Protein metal-binding sites.* Curr Opin Biotechnol, 1992. **3**(4): p. 378-87.
4. Glusker, J.P., *Structural aspects of metal liganding to functional groups in proteins.* Adv Protein Chem, 1991. **42**: p. 1-76.
5. Silva, J.R.R.F.d. and R.J.P. Williams, *The biological chemistry of the elements: the inorganic chemistry of life*. 1991, Oxford [England]
   New York: Clarendon Press;
   Oxford University Press. xxi, 561 p.
6. Lippard, S.J. and J.M. Berg, *Principles of bioinorganic chemistry*. 1994, Mill Valley, Calif.: University Science Books. xvii, 411 p.
7. Marsden, B.J., G.S. Shaw, and B.D. Sykes, *Calcium binding proteins. Elucidating the contributions to calcium affinity from an analysis of species variants and peptide fragments.* Biochem Cell Biol, 1990. **68**(3): p. 587-601.
8. Yamashita, M.M., et al., *Where metal ions bind in proteins.* Proc Natl Acad Sci U S A, 1990. **87**(15): p. 5648-52.
9. Bagley, S.C. and R.B. Altman, *Characterizing the microenvironment surrounding protein sites.* Protein Sci, 1995. **4**(4): p. 622-35.
10. Martin, R., *Bioinorganic Chemistry of Calcium.* Metal Ions in Biological Systems, ed. H. Sigel. Vol. 17. 1984, New York: Marcel Dekker. 1-49.
11. Mann, K.G., et al., *Surface-dependent reactions of the vitamin K-dependent enzyme complexes.* Blood, 1990. **76**(1): p. 1-16.
12. Herzberg, O., J. Moult, and M.N. James, *A model for the Ca2+-induced conformational transition of troponin C. A trigger for muscle contraction.* J Biol Chem, 1986. **261**(6): p. 2638-44.
13. Holmes, K.C., et al., *Atomic model of the actin filament.* Nature, 1990. **347**(6288): p. 44-9.
14. Carafoli, E., *The signaling function of calcium and its regulation.* J Hypertens Suppl, 1994. **12**(10): p. S47-56.
15. Santella, L. and E. Carafoli, *Calcium signaling in the cell nucleus.* Faseb J, 1997. **11**(13): p. 1091-109.
16. Nelson, M.R. and W.J. Chazin, *An interaction-based analysis of calcium-induced conformational changes in Ca2+ sensor proteins.* Protein Sci, 1998. **7**(2): p. 270-82.
17. Yang, J.J., A. Gawthrop, and Y. Ye, *Obtaining site-specific calcium-binding affinities of calmodulin.* Protein Pept Lett, 2003. **10**(4): p. 331-45.
18. Ye, Y., et al., *Probing site-specific calmodulin calcium and lanthanide affinity by grafting.* J Am Chem Soc, 2005. **127**(11): p. 3743-50.
19. Ye, Y., et al., *Metal binding affinity and structural properties of an isolated EF-loop in a scaffold protein.* Protein Eng, 2001. **14**(12): p. 1001-13.
20. Ye, Y., et al., *Calcium and lanthanide affinity of the EF-loops from the C-terminal domain of calmodulin.* J Inorg Biochem, 2005. **99**(6): p. 1376-83.
21. Ye, Y., et al., *A grafting approach to obtain site-specific metal-binding properties of EF-hand proteins.* Protein Eng, 2003. **16**(6): p. 429-34.
22. Yang, W., et al., *Structural analysis, identification, and design of calcium-binding sites in proteins.* Proteins, 2002. **47**(3): p. 344-56.
23. Yang, W., et al., *Design of a calcium-binding protein with desired structure in a cell adhesion molecule.* J Am Chem Soc, 2005. **127**(7): p. 2085-93.

24. Yang W, J.L., Isley L, Ye Y, Lee HW, Wilkins AL, Liu Z, Hellnga HW, Malchow R, Ghazi M, Yang JJ, *Rational Design of a Calcium-Binding Protein.* Journal of the American Chemical Society, 2003. **125**: p. 6165-6171.
25. Zou, J., et al., *Developing Sensors for Real Time Measurement of High Ca2+ Concentrations*, in *J Am Chem Soc (Pending)*. 2006.
26. Zou, J., et al., *Expression and optical properties of green fluorescent protein expressed in different cellular environments.* J Biotechnol, 2005. **119**(4): p. 368-78.
27. Zhou, Y., et al., *Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins.* Proteins, 2006. **65**(3): p. 643-55.
28. Deng, H., et al., *Predicting calcium-binding sites in proteins - a graph theory and geometry approach.* Proteins, 2006. **64**(1): p. 34-42.
29. Tsien, R.Y., *The green fluorescent protein.* Annu Rev Biochem, 1998. **67**: p. 509-44.
30. Richmond, T.A., et al., *Engineered metal binding sites on green fluorescence protein.* Biochem Biophys Res Commun, 2000. **268**(2): p. 462-5.
31. Jensen, K.K., L. Martini, and T.W. Schwartz, *Enhanced fluorescence resonance energy transfer between spectral variants of green fluorescent protein through zinc-site engineering.* Biochemistry, 2001. **40**(4): p. 938-45.
32. Needleman, H.L. *History of Lead Poisoning in the World*. [cited 08/10/2004]; Available from: http://www.leadpoison.net/general/history.htm.
33. Godwin HA, C.E., ter Horst MA, Forde CE, Stern CL, Zart MK, *207PB-1H Two-Dimensional NMR Spectroscopy: A useful New Tool for Probing Lead(II) Coordination Chemistry.* Inorganic Chemistry, 2000. **2000**(39): p. 1391-1397.
34. Todd, A.C., et al., *Unraveling the chronic toxicity of lead: An essential priority for environmental health.* Environmental Health Perspectives, 1996. **104**: p. 141-146.
35. Osmond, D.L., D.E. Line, J.A. Gale, R.W. Gannon, C.B. Knott, K.A. Bartenhagen, M.H. Turner, S.W. Coffey, J. Spooner, J. Wells, J.C. Walker, L.L. Hargrove, M.A. Foster, P.D. Robillard, and D.W. Lehning. *WATERSHEDSS: Water, Soil and Hydro-Environmental Decision Support System*. 1995 [cited; Available from: http://h2osparc.wq.ncsu.edu.
36. European Commission, D.E., *Heavy Metals in Waste, Final Report.* 2002.
37. Hernandez-Ochoa, I., et al., *Low lead environmental exposure alters semen quality and sperm chromatin condensation in northern Mexico.* Reprod Toxicol, 2005. **20**(2): p. 221-8.
38. Apostoli, P., et al., *The effect of lead on male fertility: a time to pregnancy (TTP) study.* Am J Ind Med, 2000. **38**(3): p. 310-5.
39. Despres, C., et al., *Neuromotor functions in Inuit preschool children exposed to Pb, PCBs, and Hg.* Neurotoxicol Teratol, 2005. **27**(2): p. 245-57.
40. Goyer RA, *Toxic and essential metal interactions.* Annu Rev Nutr, 1997. **17**: p. 37-50.
41. Chetty, C.S., et al., *Perinatal lead exposure alters the expression of neuronal nitric oxide synthase in rat brain.* Int J Toxicol, 2001. **20**(3): p. 113-20.
42. Reddy, G.R. and N.H. Zawia, *Lead exposure alters Egr-1 DNA-binding in the neonatal rat brain.* Int J Dev Neurosci, 2000. **18**(8): p. 791-5.
43. Moore, M.R., A. Goldberg, and A.A. Yeung-Laiwah, *Lead effects on the heme biosynthetic pathway. Relationship to toxicity.* Ann N Y Acad Sci, 1987. **514**: p. 191-203.
44. Weltje, L., *Bioavailability of Lanthanides to Freshwater Organisms*. 2002: DUP Science.
45. Drever, J.I., *The geochemistry of natural waters: surface and groundwater environments*. 3rd ed. 1997, Upper Saddle River, N.J.: Prentice Hall. xii, 436 p.
46. Marin, J. and J. Ayele, *Removal of some heavy metal cations from aqueous solutions by spruce sawdust. I. Study of the binding mechanism through batch experiments.* Environ Technol, 2002. **23**(10): p. 1157-71.
47. Sheng, P.X., et al., *Sorption of lead, copper, cadmium, zinc, and nickel by marine algal biomass: characterization of biosorptive capacity and investigation of mechanisms.* J Colloid Interface Sci, 2004. **275**(1): p. 131-41.
48. Diniz, V. and B. Volesky, *Biosorption of La, Eu and Yb using Sargassum biomass.* Water Research, 2005. **39**(1): p. 239-247.

49. Prasher, S.O., et al., *Biosorption of heavy metals by red algae (Palmaria palmata).* Environ Technol, 2004. **25**(10): p. 1097-106.
50. Chaisuksant, Y., *Biosorption of cadmium(II) and copper(II) by pretreated biomass of marine alga Gracilaria fisheri.* Environmental Technology, 2003. **24**(12): p. 1501-1508.
51. Prapaipong, P., E.L. Shock, and C.M. Koretsky, *Metal-organic complexes in geochemical processes: Temperature dependence of the standard thermodynamic properties of aqueous complexes between metal cations and dicarboxlylate ligands.* Geochimica et Cosmochimica Acta, 1999. **63**(17): p. 2547-2577.
52. Swain, A.L., R.H. Kretsinger, and E.L. Amma, *Restrained least squares refinement of native (calcium) and cadmium-substituted carp parvalbumin using X-ray crystallographic data at 1.6-A resolution.* J Biol Chem, 1989. **264**(28): p. 16620-8.
53. Vyas, M.N., B.L. Jacobson, and F.A. Quiocho, *The calcium-binding site in the galactose chemoreceptor protein. Crystallographic and metal-binding studies.* J Biol Chem, 1989. **264**(34): p. 20817-21.
54. Dudev, T. and C. Lim, *Principles governing Mg, Ca, and Zn binding and selectivity in proteins.* Chemical Reviews, 2003. **103**(3): p. 773-787.
55. Strynadka, N.C. and M.N. James, *Crystal structures of the helix-loop-helix calcium-binding proteins.* Annu Rev Biochem, 1989. **58**: p. 951-98.
56. Linse, S., et al., *The role of protein surface charges in ion binding.* Nature, 1988. **335**(6191): p. 651-2.
57. Linse, S. and S. Forsen, *Determinants that govern high-affinity calcium binding.* Adv Second Messenger Phosphoprotein Res, 1995. **30**: p. 89-151.
58. Kretsinger, R.H. and C.E. Nockolds, *Carp muscle calcium-binding protein. II. Structure determination and general description.* J Biol Chem, 1973. **248**(9): p. 3313-26.
59. Kawasaki, H., S. Nakayama, and R.H. Kretsinger, *Classification and evolution of EF-hand proteins.* Biometals, 1998. **11**(4): p. 277-95.
60. Godwin, H.A., *The biological chemistry of lead.* Current Opinion in Chemical Biology, 2001. **5**: p. 223-227.
61. Battistuzzi, G., et al., *Amide Group Coordination to the Pb(2+) Ion.* Inorg Chem, 1996. **35**(14): p. 4239-4247.
62. Shimoni-Livny, L., J.P. Glusker, and C.W. Bock, *Lone pair functionality in divalent lead compounds.* Inorganic Chemistry, 1998. **37**(8): p. 1853-1867.
63. Pidcock E, M.G., *Structural characteristics of protein binding sites for calcium and lanthanide ions.* J Biol Inorg Chem, 2001. **6**: p. 479-489.
64. Wilkins, A.L., et al., *Metal-binding studies for a de novo designed calcium-binding protein.* Protein Eng, 2002. **15**(7): p. 571-4.
65. Macalady, D.L., *Perspectives in environmental chemistry*. Topics in environmental chemistry. 1998, New York: Oxford University Press. xiv, 512 p.
66. Allen, H.E., A.W. Garrison, and G.W. Luther, *Metals in surface waters*. 1998, Chelsea, Mich.: Ann Arbor Press. xvi, 262 p.
67. Prozialeck WC, G.G., Dey PM, Reuhl KR, Parrish AR, *Cadherins and NCAM as potential targets in metal toxicity.* Toxicol Appl Pharmacol, 2002. **182**(3): p. 255-265.
68. Brzoska, M.M. and J. Moniuszko-Jakoniuk, *Interactions between cadmium and zinc in the organism.* Food Chem Toxicol, 2001. **39**(10): p. 967-80.
69. Palasz A, C.P., *Toxicological and cytophysiological aspects of lanthanides action.* Acta Biochimica Polonica, 2000. **47**(4): p. 1107-1114.
70. Millar, J.A., et al., *Lead and delta-aminolaevulinic acid dehydratase levels in mentally retarded children and in lead-poisoned suckling rats.* Lancet, 1970. **2**(7675): p. 695-8.
71. Bergdahl, I.A., *Lead-binding proteins - a way to understand lead toxicity?* Analusis, 1998. **26**(6): p. M81-M84.
72. Bergdahl, I.A., et al., *Lead binding to delta-aminolevulinic acid dehydratase (ALAD) in human erythrocytes.* Pharmacol Toxicol, 1997. **81**(4): p. 153-8.
73. Erskine, P.T., et al., *X-ray structure of 5-aminolaevulinate dehydratase, a hybrid aldolase.* Nat Struct Biol, 1997. **4**(12): p. 1025-31.

74. Magyar, J.S., et al., *Reexamination of lead(II) coordination preferences in sulfur-rich sites: implications for a critical mechanism of lead poisoning.* J Am Chem Soc, 2005. **127**(26): p. 9495-505.
75. Andersen, R.J., et al., *Characterization of the first N2S(alkylthiolate)lead compound: a model for three-coordinate lead in biological systems.* Inorg Chem, 2006. **45**(17): p. 6574-6.
76. Erskine, P.T., et al., *X-ray structure of 5-aminolevulinic acid dehydratase from Escherichia coli complexed with the inhibitor levulinic acid at 2.0 A resolution.* Biochemistry, 1999. **38**(14): p. 4266-76.
77. Jaffe, E.K., et al., *The molecular mechanism of lead inhibition of human porphobilinogen synthase.* J Biol Chem, 2001. **276**(2): p. 1531-7.
78. Ghering, A.B., et al., *Spectroscopic and functional determination of the interaction of Pb2+ with GATA proteins.* J Am Chem Soc, 2005. **127**(11): p. 3751-9.
79. Wilson, M.A. and A.T. Brunger, *Domain flexibility in the 1.75 A resolution structure of Pb2+-calmodulin.* Acta Crystallogr D Biol Crystallogr, 2003. **59**(Pt 10): p. 1782-92.
80. Markovac, J. and G.W. Goldstein, *Picomolar concentrations of lead stimulate brain protein kinase C.* Nature, 1988. **334**(6177): p. 71-3.
81. Bouton, C.M., et al., *Synaptotagmin I is a molecular target for lead.* J Neurochem, 2001. **76**(6): p. 1724-35.
82. Goldstein, G.W. and D. Ar, *Lead activates calmodulin sensitive processes.* Life Sci, 1983. **33**(10): p. 1001-6.
83. Dudev, T., et al., *First-second shell interactions in metal binding sites in proteins: a PDB survey and DFT/CDM calculations.* J Am Chem Soc, 2003. **125**(10): p. 3168-80.
84. Nayal, M. and E. Di Cera, *Predicting Ca(2+)-binding sites in proteins.* Proc Natl Acad Sci U S A, 1994. **91**(2): p. 817-21.
85. Sodhi, J.S., et al., *Predicting metal-binding site residues in low-resolution structural models.* Journal of Molecular Biology, 2004. **342**(1): p. 307-320.
86. Hellinga, H.W. and F.M. Richards, *Construction of new ligand binding sites in proteins of known structure. I. Computer-aided modeling of sites with pre-defined geometry.* J Mol Biol, 1991. **222**(3): p. 763-85.
87. Schymkowitz, J.W., et al., *Prediction of water and metal binding sites and their affinities by using the Fold-X force field.* Proc Natl Acad Sci U S A, 2005. **102**(29): p. 10147-52.
88. Clarke, N.D. and S.M. Yuan, *Metal search: a computer program that helps design tetrahedral metal-binding sites.* Proteins, 1995. **23**(2): p. 256-63.
89. Reid, R.E. and R.S. Hodges, *Co-operativity and calcium/magnesium binding to troponin C and muscle calcium binding parvalbumin: an hypothesis.* J Theor Biol, 1980. **84**(3): p. 401-44.
90. Linse, S., et al., *A calbindin D9k mutant with reduced calcium affinity and enhanced cooperativity. Metal ion binding, stability, and structural studies.* Biochemistry, 1994. **33**(41): p. 12478-86.
91. Davis, S.J., et al., *The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2.* Proc Natl Acad Sci U S A, 1998. **95**(10): p. 5490-4.
92. Lee, B. and F.M. Richards, *The interpretation of protein structures: estimation of static accessibility.* J Mol Biol, 1971. **55**(3): p. 379-400.
93. McPhalen, C.A., N.C. Strynadka, and M.N. James, *Calcium-binding sites in proteins: a structural perspective.* Adv Protein Chem, 1991. **42**: p. 77-144.
94. Singh, T.P., et al., *Crystal structure of a complex formed between proteolytically-generated lactoferrin fragment and proteinase K.* Proteins, 1998. **33**(1): p. 30-8.
95. Hausrath, A.C. and B.W. Matthews, *Redetermination and refinement of the complex of benzylsuccinic acid with thermolysin and its relation to the complex with carboxypeptidase A.* J Biol Chem, 1994. **269**(29): p. 18839-42.
96. Muller, A., et al., *Crystal structure of calcium-free proteinase K at 1.5-A resolution.* J Biol Chem, 1994. **269**(37): p. 23108-11.
97. Milner-White, E.J., *The partial charge of the nitrogen atom in peptide bonds.* Protein Sci, 1997. **6**(11): p. 2477-82.

98. Sarret G, M.A., Spadini L, Roux JC, Hazemann JL, Soldo Y, Eybert-Berard L, Menthonnex JJ, *Structural Determination of Zn and Pb Binding Sites in Penicillium chrysogenum Cell Walls by EXAFS Spectroscopy.* Environ Sci Technol, 1998. **32**: p. 1648-1655.
99. Zimmer, M. [cited; Available from: http://www.conncoll.edu/ccacad/zimmer/GFP-ww/GFP3.htm.
100. Cubitt, A.B., et al., *Understanding, improving and using green fluorescent proteins.* Trends Biochem Sci, 1995. **20**(11): p. 448-55.
101. Reid, B.G. and G.C. Flynn, *Chromophore formation in green fluorescent protein.* Biochemistry, 1997. **36**(22): p. 6786-91.
102. Heim, R., A.B. Cubitt, and R.Y. Tsien, *Improved green fluorescence.* Nature, 1995. **373**(6516): p. 663-4.
103. Zimmer, M., *Green fluorescent protein (GFP): applications, structure, and related photophysical behavior.* Chem Rev, 2002. **102**(3): p. 759-81.
104. Nagai, T., et al., *A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications.* Nat Biotechnol, 2002. **20**(1): p. 87-90.
105. Rekas, A., et al., *Crystal structure of venus, a yellow fluorescent protein with improved maturation and reduced environmental sensitivity.* J Biol Chem, 2002. **277**(52): p. 50573-8.
106. Lippincott-Schwartz, J. and G.H. Patterson, *Development and Use of Fluorescent Protein Markers in Living Cells.* Science, 2003. **300**: p. 87-91.
107. Payne, J.C., M.A. ter Horst, and H.A. Godwin, *Lead fingers: Pb2+ binding to structural zinc-binding domains determined directly by monitoring lead-thiolate charge-transfer bands.* Journal of the American Chemical Society, 1999. **121**(29): p. 6850-6855.
108. Andersen, O., *Chelation of cadmium.* Environ Health Perspect, 1984. **54**: p. 249-66.
109. Miyawaki, A., *Visualization of the spatial and temporal dynamics of intracellular signaling.* Dev Cell, 2003. **4**(3): p. 295-305.

APPENDIX A

Table A. 1: **Summary of datasets for statistical analysis of calcium binding proteins.**

Dataset I			Dataset II			Dataset III			Dataset IV		
PDB Code	Ca ID	CN	PDB Code	Ca ID	CN	PDB Code	Ca ID	CN	PDB Code	Ca ID	CN
1ALC	992	7	2TEP	7081	7	1A0T	3250	4	1ALA	2505	5
1ATN	5019	3	2TEP	7083	7	1A0T	6610	4	1ALA	2506	5
1ATN	5020	6	2TEP	7085	7	1A0T	9970	4	1ALA	2507	6
1ATN	5021	6	2TEP	7087	7	1A2X	1248	8	1ALV	1713	7
1ATN	5022	3	1BF2	5737	8	1A2X	1249	8	1ALV	1714	7
1CSE	2494	5	1BFD	3951	7	1AJJ	286	6	1ALV	1715	8
1CSE	2495	7	1BFD	3953	5	1ALV	1713	7	1ALV	1716	8
1MEE	2481	7	1BJR	2089	4	1ALV	1714	7	1ALV	3429	7
1MEE	2482	5	1BJR	2090	8	1ALV	1715	8	1ALV	3430	7
1NPC	2392	8	1EGZ	6810	6	1ALV	1716	8	1ALV	3431	7
1NPC	2393	6	1EGZ	6811	5	1ALV	3429	7	1ALV	3432	8
1NPC	2394	7	1EGZ	6812	7	1ALV	3430	7	1AUI	4393	7
1NPC	2395	7	1CE5	1631	6	1ALV	3431	7	1AUI	4394	7
1OMD	848	7	1B9Z	4130	6	1ALV	3432	8	1AUI	4395	7
1OMD	849	7	1B9O	1032	7	1AVC	5110	6	1AUI	4396	8
1OVA	12449	5	1FZC	3086	6	1AVC	5111	4	1AVS	629	7
1SNC	1183	8	1FZC	5532	5	1AVC	5112	6	1AVS	630	7
1TMN	2434	8	1FZC	8618	6	1AVC	5113	4	1AVS	1268	7
1TMN	2435	6	1FZC	11053	7	1AVC	5114	7	1AVS	1269	7
1TMN	2436	6	1KVX	956	7	1AVC	5115	6	1B9O	1032	7
1TMN	2437	7	1AG9	1418	6	1BF2	5737	8	1FJ3	2461	8
2LTN	1420	7	1AG9	2839	6	1BN8	3068	6	1FJ3	2462	6
2LTN	3237	7	1AG9	2843	7	1EDM	312	7	1FJ3	2463	7
2MSB	885	8	1AG9	2844	8	1EDM	647	6	1FJ3	2464	7
2MSB	886	6	1AI4	6086	7	1EDM	648	7	1GLG	2362	7
2MSB	887	7	1AQM	4261	8	1ESL	1268	7	1K96	715	7
2MSB	1770	8	1AX0	1981	7	1ESL	1269	4	1K96	716	7
2MSB	1771	8	1AYO	2040	5	1ESL	1270	7	1PSH	2788	7
2MSB	1772	7	1GSP	806	6	1GCA	2918	7	1PSH	2789	6
2POR	2230	7	1HEI	6434	5	1GEN	1963	7	1PSH	2790	7
2POR	2231	8	1HEI	6435	4	1HFC	1249	6	1SCD	1922	7
2POR	2232	6	1IRB	951	7	1JDC	3299	7	1SCD	1923	3
2ST1	1940	7	1JDA	3299	7	1JDC	3300	6	1SNC	1183	8
2ST1	1941	7	1JDA	3300	6	1LED	1907	7	1THM	2005	7
2TAA	3692	2	5CHY	983	5	1NBC	1219	8	1THM	2006	7
2TEC	2006	7	1TCM	5257	6	1NBC	2577	8	2PRK	2032	8
2TEC	2007	7	1TCM	5258	6	1NZY	2099	6	2PRK	2033	5
2TGT	1687	6	1TCM	10635	6	1NZY	4251	6	3CLN	1168	7
2TMV	1305	8	1TCM	10636	3	1NZY	6392	6	3CLN	1169	7
3CLN	1168	7	1OIL	2339	6	1OSA	1167	7	3CLN	1170	7

3CLN	1169	7
3CLN	1170	7
3CLN	1171	7
3CNA	1810	6
3EST	1824	7
3GBP	4706	7
3ICB	602	7
3ICB	603	7
3LHM	1033	7
3PRK	2056	8
4BP2	1010	7
4SBV	1985	5
4SBV	3985	6
4SBV	5981	6
4SGB	1710	6
5CPV	812	7
5CPV	813	7
5ENL	3291	8
5PTP	1657	6
5TNC	1283	7
5TNC	1284	7
7NN9	3401	6

1OIL	4678	6
1OBR	2584	6
1OBR	2585	8
1OBR	2586	5
1OBR	2587	6
1NBC	1219	8
1NBC	2577	8
1KIT	5861	7
1KIT	5862	6
1JS4	9517	7
1JS4	9518	7
1JS4	9519	8
1JS4	9520	7
1KBC	1296	6
1KBC	1297	6
1KBC	2623	6
1KBC	2624	6
1SBF	1711	7
1TN3	1068	8
1TN3	1069	7
1VSI	1145	5
4LIP	2335	6
4LIP	4958	6
2FIB	2037	7
1ALA	2505	5
1ALA	2506	5
1ALA	2507	6
1ATL	1608	6
1ATL	3240	7
1CLX	2702	7
1CLX	5404	7
1CLX	8106	7
1CLX	10808	7
1MMQ	1273	6
1MMQ	1274	6
1OAC	5682	6
1OAC	5683	8
1OAC	11876	6
1OAC	11877	7
1SBH	1943	7
1SBH	1944	4

1OSA	1168	7
1OSA	1169	7
1OSA	1170	7
1PAM	5314	7
1PAM	5315	8
1PAM	10629	7
1PAM	10630	8
1POA	916	8
1POA	1068	7
1REC	1512	7
1RRO	848	7
1RRO	849	7
1RRO	850	3
1RRO	851	5
1SMD	3948	8
1SMV	4516	6
1SMV	4517	7
1SMV	4518	6
1SMV	4519	0
1SRA	1264	7
1SRA	1265	7
1SRA	1266	6
1THM	2005	7
1THM	2006	7
1WDC	2903	7
1XJO	2010	7
2AYH	1703	6
2POR	2230	7
2POR	2231	8
2POR	2232	6
2SCP	1370	7
2SCP	1371	7
2SCP	1372	7
2SCP	1373	7
2SCP	1374	6
2SCP	1375	8
2STV	1519	3
2STV	1520	3
2STV	1521	3
4SGB	1710	6
5PTP	1657	6

3CLN	1171	7
3EST	1824	7
4ICB	641	7
4ICB	642	7
5PAL	843	7
5PAL	844	7

	1SRA	1264	7	9RNT	799	7	
	1SRA	1265	7				
	1SRA	1266	6				
	1CEL	7039	5				
	1ESL	1268	7				
	1ESL	1269	4				
	1ESL	1270	7				
	1GCG	2895	7				
	1HYT	2440	8				
	1HYT	2441	6				
	1HYT	2442	7				
	1HYT	2443	7				
	1IAG	1623	6				

APPENDIX B

1 Analyses of EGFPwtF and variants as potential sensors for monitoring in vivo trypsin activity

1.1 Introduction

Supplemental work was conducted to evaluate the use of our engineered EGFP variants to respond as a sensor for trypsin activity. This work is still in progress, and no data has been submitted for publication.

1.2 Materials and Methods

Analyses of EGFPwtF and variants as potential sensors for monitoring in-vivo trypsin activity were conducted on a Shimadzu UV-1700 PharmaSpec spectrophotometer (Shimadzu North America, Columbia, MD) with UVProbe software. To evaluate EGFP variants as potential sensors, the UV-Vis spectra were evaluated as the ratio between the peaks near 398 nm and 490 nm (Figure B.1), corresponding to the neutral and anionic forms of the EGFPwtF chromophore, respectively.

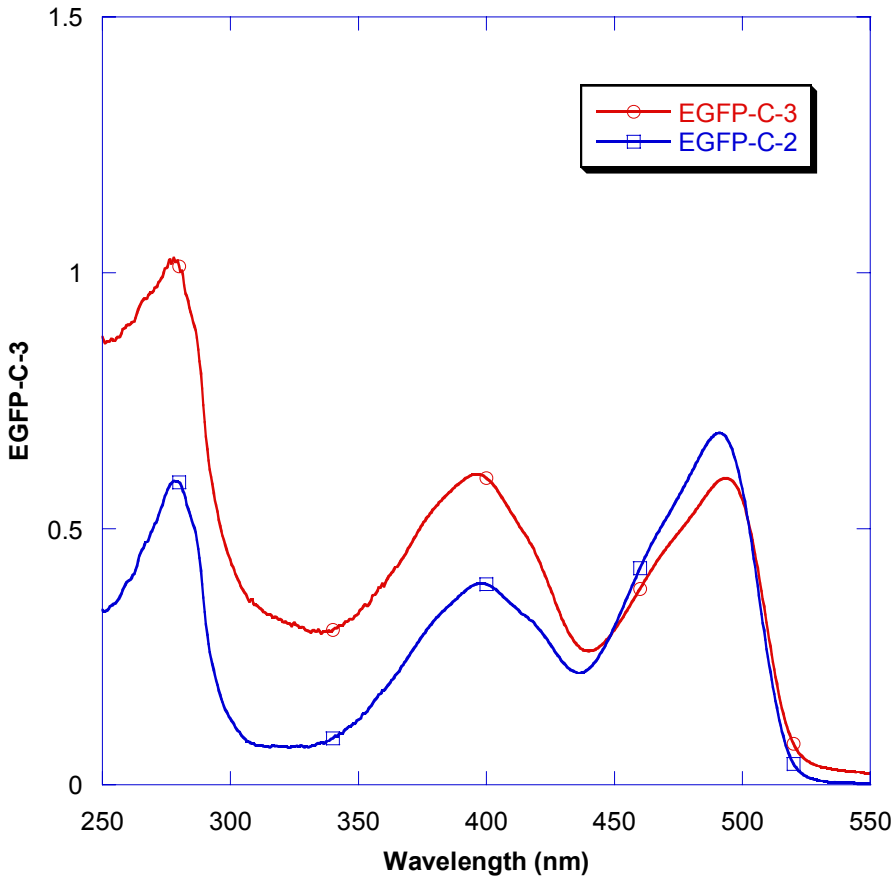


Figure B. 1: UV-Vis absorbance spectra for EGFP variants C-3 and C-2.

It has been shown (work not published) that the ratio of these two peaks changes subsequent to trypsin cleavage, indicating that conformational changes occur which alter the solvent accessibility of the chromophore. It was expected that the largest ratiometric changes would be observed when the two peaks were nearly equivalent in intensity, as is observed with the EGFP-C-3 variant in Figure B.1.

Protein samples were first prepared by dialysis in 1.5 L of 10 mM TRIS, 1 mM DTT, pH 7.4, to prevent disulfide bond formation. Because trypsin is stabilized by calcium, it was necessary to evaluate the relative effects of calcium in the buffer. For this

preliminary analysis, 20 µM EGFP-C-4 was evaluated using different buffer compositions, as seen in Figure B.2.

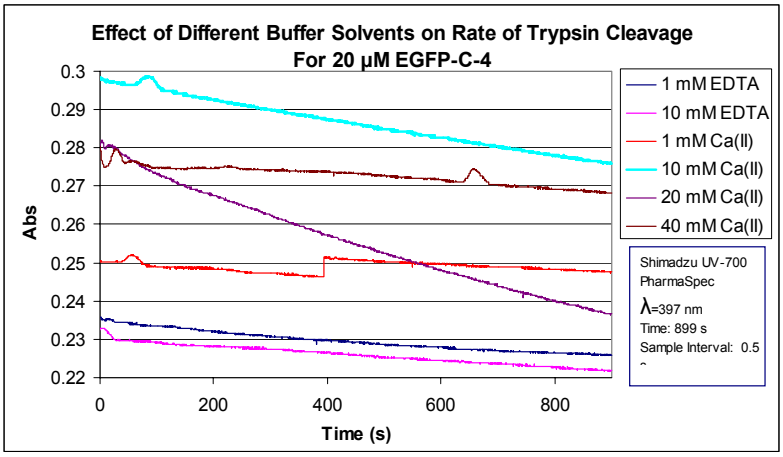


Figure B. 2: Comparison of different buffers during trypsin cleavage analysis.

Trypsin was added manually using a pipet. After addition of the trypsin, the cuvette was capped, and the solution shaken, prior to measuring absorbance. Changes in the absorption spectra as a result of trypsin activity were monitored using a time-course model: 397 nm, 899 s, with a sample interval of 0.5s.

These preliminary results indicated that the buffer containing 20 mM Ca^{2+} is close to optimal for the analysis of trypsin activity, producing the sharpest decrease in absorbance when compared to buffers containing EDTA. EDTA was evaluated as it had been previously included as a buffer component in research conducted in our laboratory. Subsequent to this analysis, 20 mM Ca^{2+} was added to the buffer for future analyses.

Protein samples were analyzed in 10 mM TRIS, 20 mM Ca^{2+}, 1 mM DTT at pH 7.4. All proteins were analyzed using a time course model on a Shimadzu UV-700 Pharmaspec spectrophotometer.

The first variant examined was EGFP-C-4. Trypsin samples were evaluated in triplicate, with a single negative control. Three 1.0 mL, 20 μM samples of EGFP-C-4 were evaluated by adding 5 μL of 2 μM trypsin for a final concentration of 10 nM. A single negative control with no trypsin addition was also analyzed. Low concentrations of trypsin were used for all analyses to reduce the initial rate (V_0) for quantification. To evaluate the kinetics associated with trypsin cleavage, the initial rate of cleavage (V_0) was calculated based on Equation 18 as the ratio of the change in absorbance over time, and the absorbance vs. time interval selected was close to linear. The initial rate (V_0) was calculated based on the change in absorbance over a 50 s interval.

$$V_0 = \Delta A/\Delta t \qquad \text{(Equation 18)}$$

Because the last set of experiments utilized Ca^{2+} in the buffer, and both trypsin and the EGFP-C-4 variant, in theory, bind Ca^{2+}, an additional experiment was designed to evaluate whether the observed absorbance changes were related to calcium-binding with the trypsin only, or whether calcium-binding with the EGFP variant contributed to this observation. To evaluate this, EGFPwtF was analyzed under the same conditions as the EGFP-C-4, because GFP and it's associated variants are not known to bind calcium [109].

The EGFP variant 5G-III-5G was also evaluated at concentrations 2, 4, 6, 10, 15, 20, 25 and 30 μM, in 10 mM TRIS, 1 mM DTT at pH 7.4. The trypsin concentration added was 5 nm. Changes in the absorption spectra as a result of trypsin activity were monitored using a time-course model: 397 nm, 899 s, with a sample interval of 0.5s. Additionally, UV-Vis absorbance spectra were collected at the initial protein addition,

following addition of 20 mM Ca^{2+}, following completion of the time-course model with added trypsin, 2 hours after initial addition of trypsin, and overnight.

1.3 Results and Discussion

Several of our engineered EGFPwtF variants have previously been shown to interact with trypsin at different points in the grafted sequence, however, the kinetics associated with this and identification of the specific cleavage sites remains to be elucidated.

The first variant examined was EGFP-C-4. The time-course spectra for triplicate samples of 20 μM EGFP-C-4 in 10 mM TRIS, 1 mM DTT and 20 mM Ca^{2+} at pH 7.4 are seen in Figure B.3. A single control sample with 0 mM Ca^{2+} was also run at this time. The three trials with 20 mM Ca^{2+} display a distinct initial drop in absorbance that is absent in the control. Data for calculation of the initial rate (V_0) can be seen in Table B.1. The initial rate (V_0) was calculated based on the change in absorbance over a 50 s interval (75-125 s) using Equation 18. As seen in Table B.1, mean V_0 was found to be $(5.3 \pm 1.0) \times 10^{-5}$, with a percent relative standard deviation of 0.2.

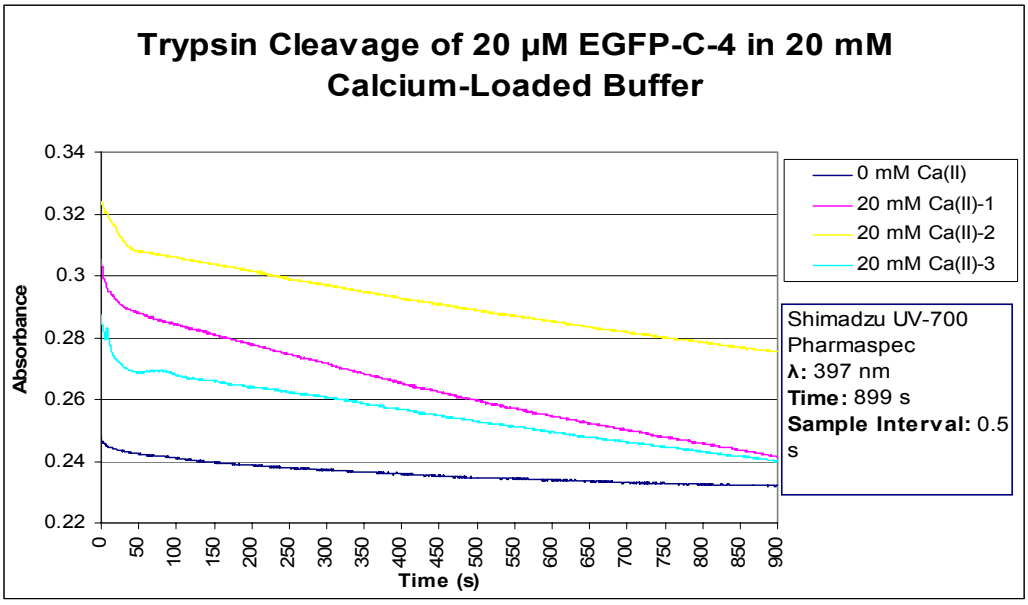


Figure B. 3: Evaluation of reproducibility of manual trypsin cleavage analysis technique.

Table B. 1: Initial rates for trypsin cleavage of 20 µM EGFP-C-4

$CaCl_2$ Concentration (mM)	ΔA	Δt	V_0
0	0.0015	50	3.0E-05
20	0.0032	50	6.4E-05
20	0.0022	50	4.4E-05
20	0.0026	50	5.2E-05
Mean			5.3E-05
Std Dev			1.0E-05
%Rel Std Dev			1.9E-01

The UV-Vis spectra showing the pre-cleavage and post-cleavage absorbances for trypsin cleavage of EGFP-C-4, trials 1 and 3 are presented in Figure B.4. These spectra clearly display a decrease in the neutral form of the chromophore, and the concomitant increase in the anionic form, subsequent to trypsin cleavage. In basic pH, these changes indicate that the conformational change associated with trypsin cleavage results in increased exposure of the chromophore to the solvent.

Additionally, Table B.2 summarizes the absorbance changes measured before addition of the trypsin (pre-cleavage) and after addition of the trypsin (post-cleavage).

The percent relative changes in absorbance for trial 1 peaks 1 and 2 were 31% and 29%, corresponding to the neutral and anionic forms, respectively. For trial 3, the percent relative changes in absorbance for peaks 1 (neutral) and 2 (anionic) were 21% and 18%. Data presented from both the UV-Vis absorbance and the time-course model indicate that trypsin cleavage can be quantified using these methods.

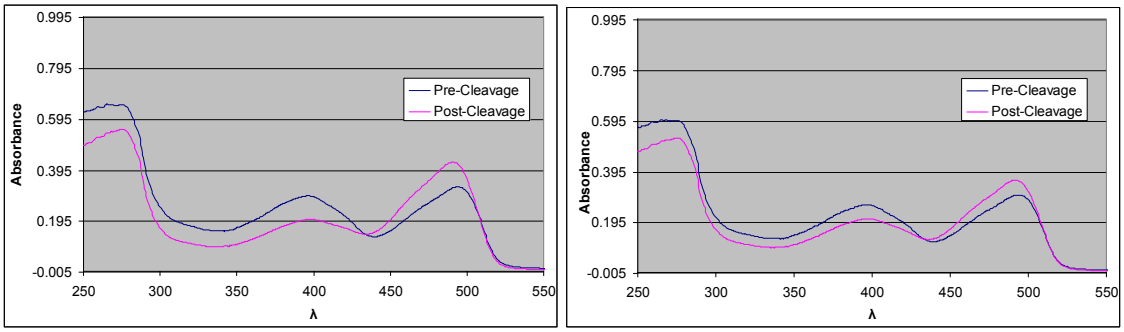


Figure B. 4: UV-Vis spectra showing the pre-cleavage and post-cleavage absorbances for trypsin cleavage of EGFP-C-4, trials 1 (left) and 3 (right).

Table B. 2: Analysis of Absorbance changes for 20 µM EGFP-C-4 in 20 mM Calcium-Loaded Buffer.

	Trial 1	Trial 3
Peak 1		
λ		
Pre-cleavage	398	397
Post-cleavage	399	397
Abs		
Pre-cleavage	0.295	0.266
Post-cleavage	0.204	0.211
Δ_{Abs}	0.091	0.055
% Rel Δ_{Abs}	31	21
Peak 2		
Λ		
Pre-cleavage	493	494
Post-cleavage	490	494
Abs		
Pre-cleavage	0.330	0.305
Post-cleavage	0.427	0.359
Δ_{Abs}	-0.097	-0.054
% Rel Δ_{Abs}	-29	-18

Next, EGFPwtF was evaluated using the same conditions and methods applied to the analysis of EGFP-C-4. This experiment produced interesting results. The time-course absorbance spectra for triplicate trials are seen in Figure B.5, with analysis of the related kinetic data presented in Table B.3. It was originally assumed that EGFPwtF would not react with trypsin, however, the spectra indicates some reaction is occurring. The initial rate (V_0) was calculated for the three trials and the control at Δt = 50 s (75 – 125 s). The mean V_0 was calculated at (3.0 ± 0.6) x 10^{-5}, which, is more than half the rate calculated for EGFP-C-4, as seen in Table B.1.

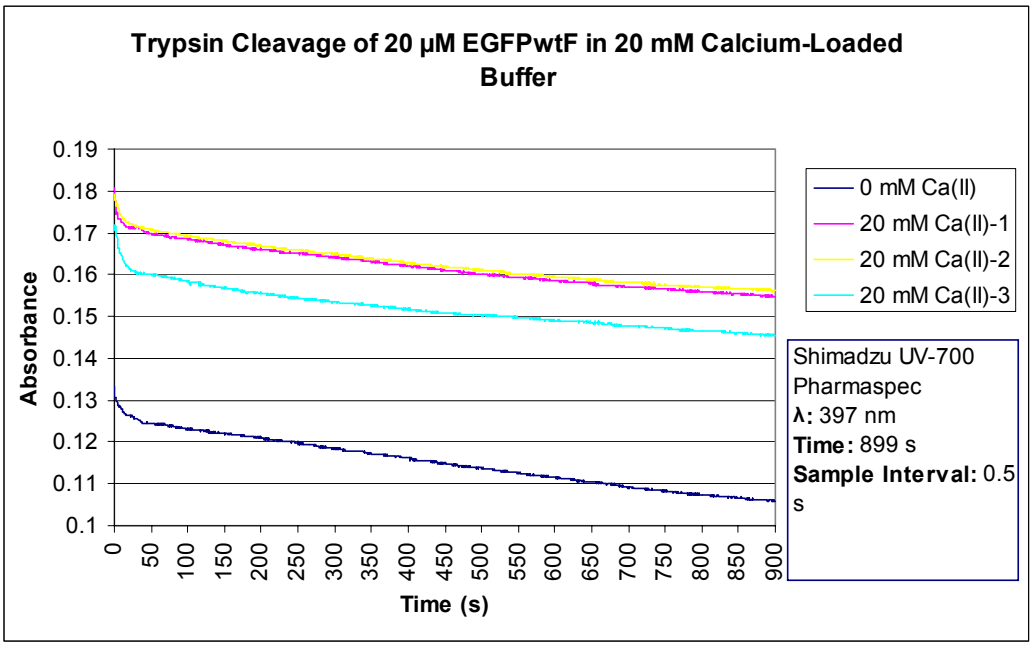


Figure B. 5: Evaluation of reproducibility of manual trypsin cleavage analysis technique.

Table B. 3: Initial rates for trypsin cleavage of 20 µM EGFPwtF

$CaCl_2$ (mM)	ΔA	Δt	V_0
0	0.0014	50	2.8E-05
20	0.0015	50	3.0E-05
20	0.0012	50	2.4E-05
20	0.0018	50	3.6E-05
Mean			3.0E-05
Std Dev			6.0E-06
%Rel Std Dev			2.0E+01

However, the changes in absorbance at 280 nm for Trp and at 488 nm for the anionic chromophore of EGFP-C-4 remain relatively constant with EGFPwtF. From Figure B.6 and B.7 which present UV-Vis absorbance spectra for EGFPwtF at 0 and 20 mM calcium, respectively, it can be seen that the absorbance at 488 nm is constant not only between the two calcium concentrations, but also between the time intervals measured with respect to addition of the trypsin. The same results are apparent at 280 nm, with the associated data summarized in Table B.4.

From Table B.4 it is clear that no change has occurred with respect to the absorbance of the protein chromophore, while the change at 280 nm indicates a 10% decrease in absorbance based on the signal associated with Tryptophan. Despite the trend observed in the time-course model in Figure B.5, the UV-Vis spectra indicate that, if trypsin cleavage is occurring in EGFPwtF, it does not result in conformational changes that change the absorbance spectra. Additionally, the small change in apparent concentration reflected at 280 nm is much smaller than the change observed for EGFP-C-4.

Table B. 4: Summary of Absorbance changes for 20 µM EGFPwtF in 20 mM Calcium-Loaded Buffer

	280nm			488nm		
Ca^{2+} (mM)	Before Trypsin	After Trypsin	Trypsin + 1 hr	Before Trypsin	After Trypsin	Trypsin + 1 hr
0	0.410	0.364	0.357	0.641	0.641	0.650
20	0.472	0.441	0.426	0.643	0.641	0.642
20	0.484	0.453	0.438	0.654	0.651	0.655
20	0.462	0.429	0.415	0.626	0.622	0.623
Mean	0.473	0.441	0.426	0.641	0.638	0.640

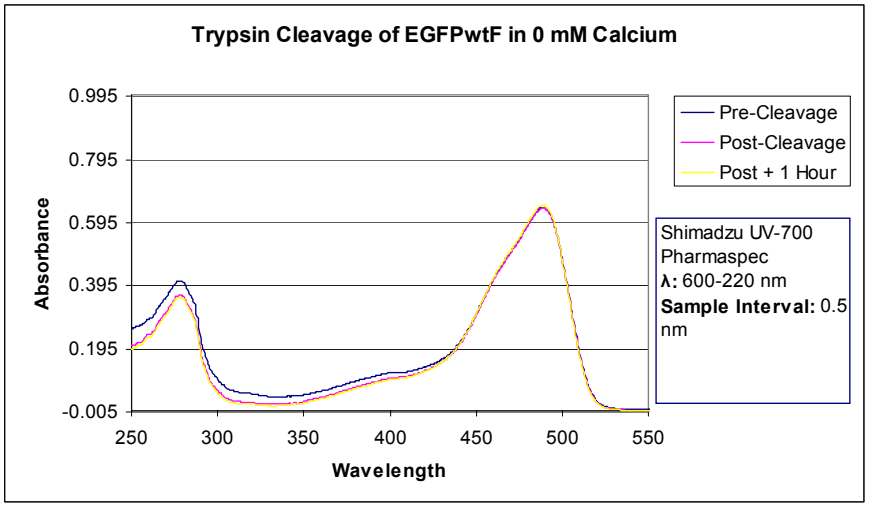


Figure B. 6: Trypsin cleavage of EGFPwtF in 0 mM Ca^{2+}.

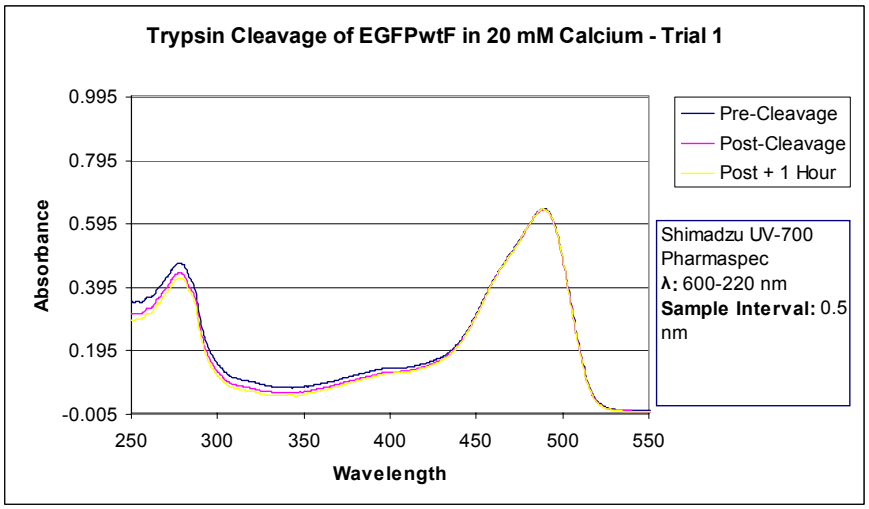


Figure B. 7: Trypsin cleavage of EGFPwtF in 20 mM Ca^{2+}.

Finally, The EGFP variant 5G-III-5G was also evaluated at concentrations 2, 4, 6, 10, 15, 20, 25 and 30 µM, in 10 mM TRIS, 1 mM DTT at pH 7.4. The trypsin concentration added was 5 nm. Changes in the absorption spectra as a result of trypsin activity were monitored using a time-course model, as seen in Figure B.8. The time-course spectra display linear absorbance values with a slope at or near zero, for all concentrations evaluated.

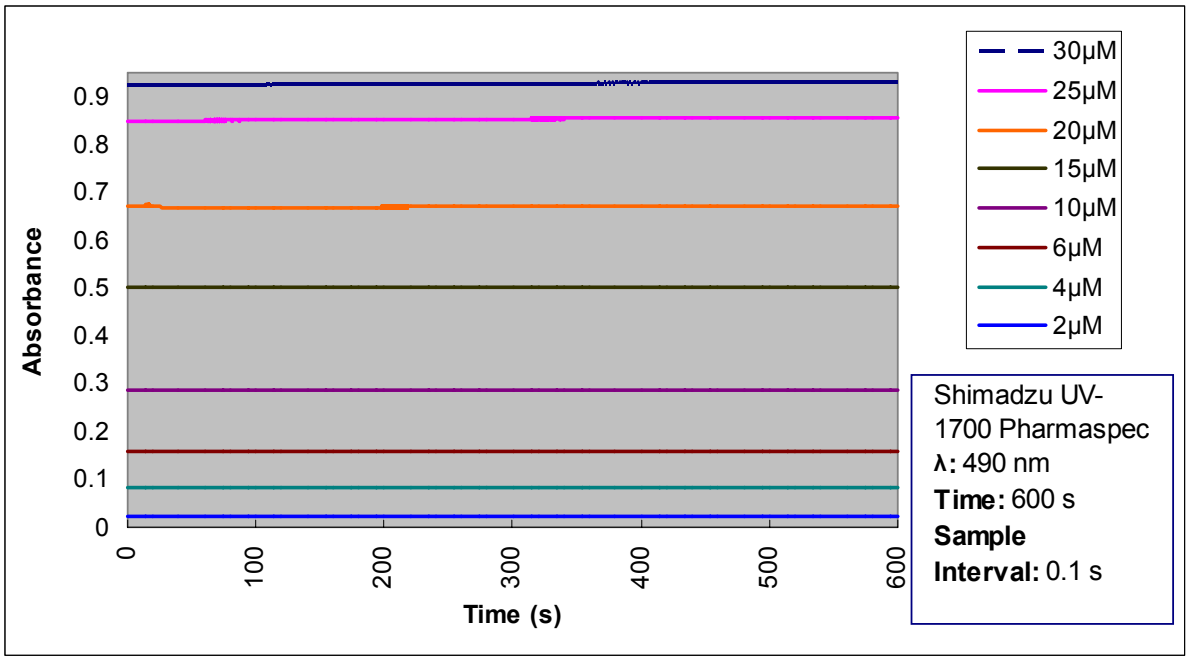


Figure B. 8: Time course model of UV-Vis spectra for trypsin cleavage of EGFP-5G-III-5G in 20 mM calcium, 10 mM TRIS, pH 7.4, over protein concentration range 2 µM – 30 µM.

Additionally, the UV-Vis absorbance spectra were collected at the initial protein addition, following addition of 20 mM Ca^{2+}, following completion of the time-course model with added trypsin, 2 hours after initial addition of trypsin, and overnight for all concentrations evaluated, as seen in Figure B.9.

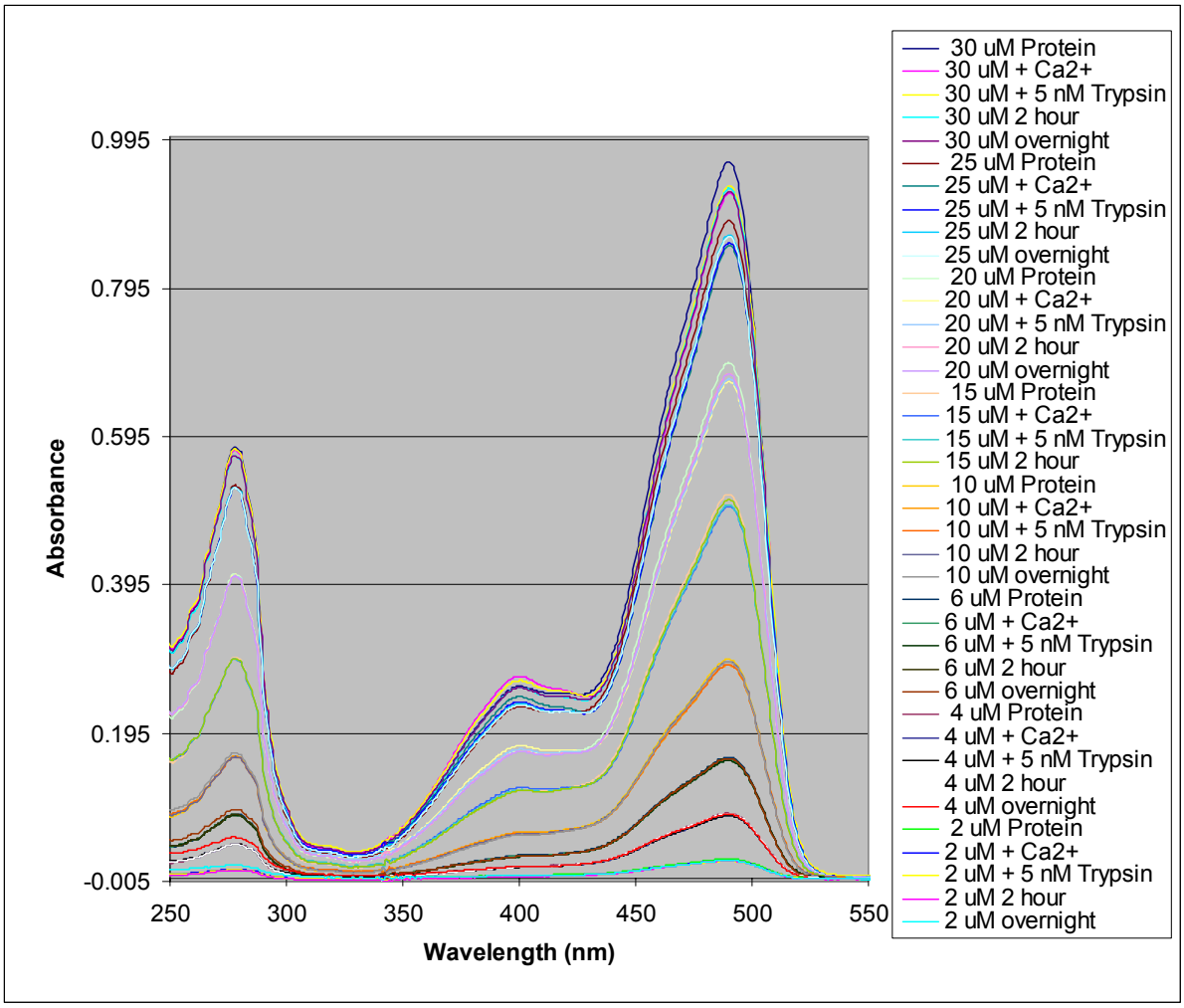


Figure B. 9: UV-Vis spectra for trypsin cleavage of EGFP-5G-III-5G in 20 mM calcium, 10 mM TRIS, pH 7.4, over protein concentration range 2 µM – 30 µM.

The UV-Vis spectra for concentrations 30, 25, 20, and 15 µM are separated, and presented individually in Figure B.10, where it can be clearly seen that no change has occurred over a 24 hour period, therefore, no cleavage has occurred. To verify this, an SDS-PAGE Gel was run (Figure B.11) which shows a strong band near 30 kDa corresponding to the protein. The gel results supported the conclusion that cleavage had not taken place, but also suggested that the protein may have dimerized based on the observed band near 66 kDa.

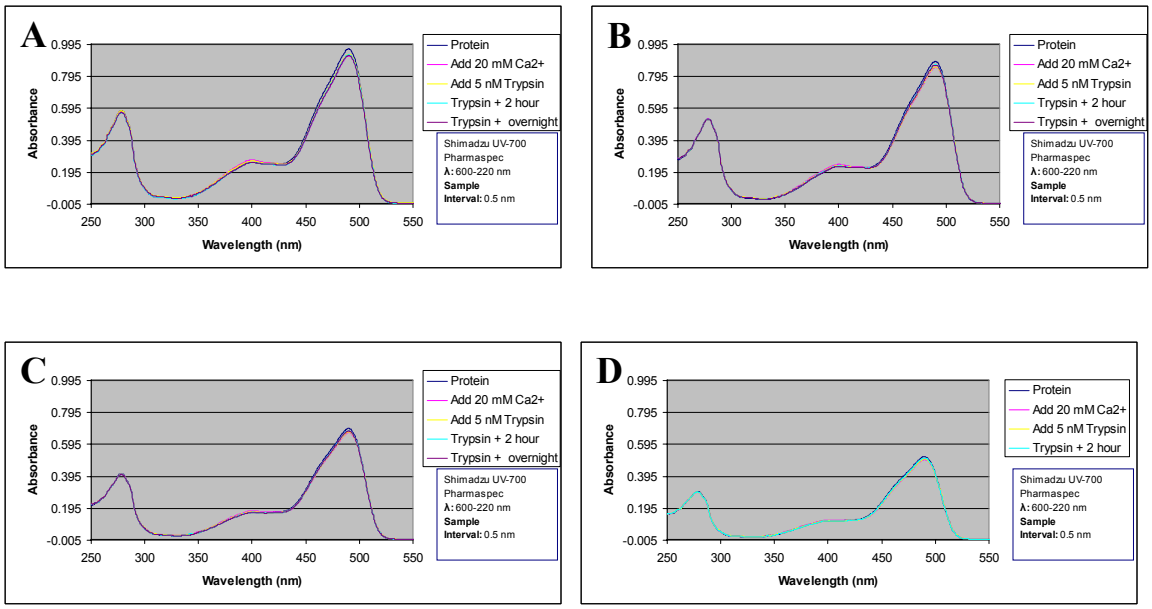
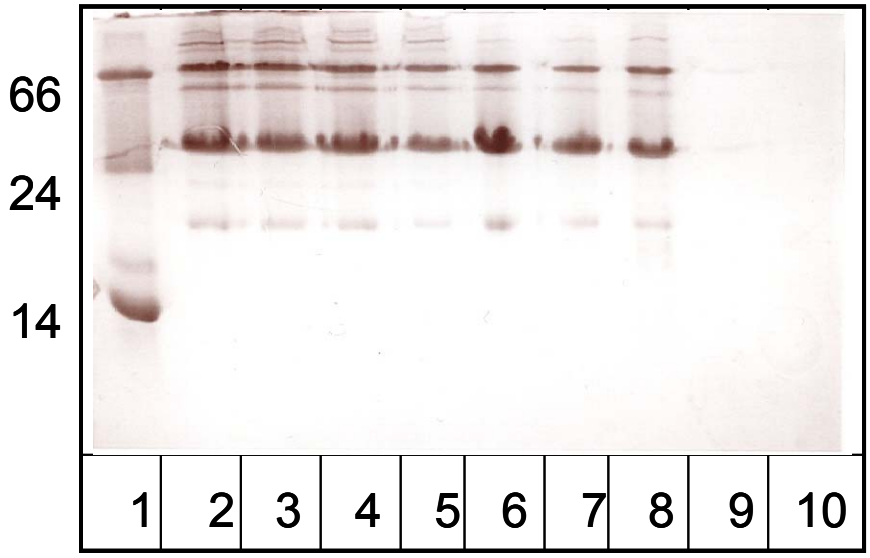


Figure B. 10: UV-Vis spectra for trypsin cleavage of EGFP-5G-III-5G in 20 mM calcium, 10 mM TRIS, pH 7.4, for concentrations (A) 30 μM, (B) 25 μM, (C) 20 μM, and (D) 15 μM.



EGFP-5G-III-5G Trypsin Cleavage	
Lane	ID
1	Marker
2	30 µm
3	30 µm
4	25 µm
5	20 µm
6	10 µm
7	6 µm
8	4 µm
9	2 µm
10	2 µm

Figure B. 11: SDS-PAGE for EGFP-5G-III-5G. Samples analyzed 24 hours after addition of trypsin.

1.4 Trypsin Cleavage Conclusions

Few conclusions are suggested at this time, with respect to application of our EGFP variants as sensors for trypsin activity. The addition of 20 mM Ca^{2+} was shown to increase trypsin cleavage activity, and the ratiometric change between the absorbance peaks at 398 nm and 488 nm can be used to quantitatively assess the cleavage based on changes from the neutral to anionic chromophores species, but further testing will be required to evaluate all of the variants currently available. The fact that the 5G-III-5G

variant appears to produce no change is consistent with other work completed in our laboratory (data not shown), indicating that cleavage occurs in the helices, but the similar results reported for the EGFPwtF variant suggests more than one cleavage site may react.