

Georgia State University

ScholarWorks @ Georgia State University

Chemistry Dissertations

Department of Chemistry

8-4-2008

Using Protein Design to Understand the Role of Electrostatic Interactions on Calcium Binding Affinity and Molecular Recognition

Lisa Michelle Jones

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

 Part of the [Chemistry Commons](#)

Recommended Citation

Jones, Lisa Michelle, "Using Protein Design to Understand the Role of Electrostatic Interactions on Calcium Binding Affinity and Molecular Recognition." Dissertation, Georgia State University, 2008.
doi: <https://doi.org/10.57709/1059259>

This Dissertation 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 Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

USING PROTEIN DESIGN TO UNDERSTAND THE ROLE OF ELECTROSTATIC
INTERACTIONS ON CALCIUM BINDING AFFINITY AND MOLECULAR
RECOGNITION

by

LISA MICHELLE JONES

Under the Direction of Jenny J. Yang

ABSTRACT

Calcium regulates many biological processes through interaction with proteins with different conformational, dynamic, and metal binding properties. Previous studies have shown that the electrostatic environment plays a key role in calcium binding affinity. In this research, we aim to dissect the contribution of the electrostatic environment to calcium binding affinity using protein design. Many natural Ca^{2+} binding proteins undergo large conformational changes upon Ca^{2+} binding which hampers the study of these proteins. In addition, cooperativity between multiple Ca^{2+} binding sites makes it difficult to study site-specific binding affinity. The design of a single Ca^{2+} binding site into a host system eliminates the difficulties that occur in the study of Ca^{2+} binding affinity. Using a computer algorithm we have rationally designed several calcium binding sites with a pentagonal bipyramidal geometry in the non-calcium dependent cell adhesion protein CD2 (CD2-D1) to better investigate the key factors that affect calcium binding affinity. The first generation proteins are all in varying

electrostatic environments. The conformational and metal binding properties of each of these designed proteins were analyzed.

The second generation designed protein, CD2.6D79, was designed based on criteria learned from the first generation proteins. This protein contains a novel calcium binding site with ligands all from the β -strands of the non-calcium dependent cell adhesion protein CD2. The resulting protein maintains native secondary and tertiary packing and folding properties. In addition to its selectivity for Ca^{2+} over other mono and divalent metal ions, it displays strong metal binding affinities for Ca^{2+} and its analogues Tb^{3+} and La^{3+} . Furthermore, our designed protein binds CD48, the ligand binding partner of CD2, with an affinity three-fold stronger than CD2. The electrostatic potential of the Ca^{2+} binding site was modified through mutation to facilitate the study of the effect of electrostatic interactions on Ca^{2+} binding affinity. Several charge distribution mutants display varying metal binding affinities based on their charge, distance to the Ca^{2+} binding site, and protein stability. This study will provide insight into the key site factors that control Ca^{2+} binding affinity and Ca^{2+} dependent biological function.

INDEX WORDS: Ca^{2+} binding proteins, Ca^{2+} binding affinity, protein design, electrostatic potentials, protein stability, protein folding

USING PROTEIN DESIGN TO UNDERSTAND THE ROLE OF ELECTROSTATIC
INTERACTIONS ON CALCIUM BINDING AFFINITY AND MOLECULAR
RECOGNITION

by

LISA MICHELLE JONES

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

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2006

Copyright by
Lisa Michelle Jones
2006

USING PROTEIN DESIGN TO UNDERSTAND THE ROLE OF ELECTROSTATIC
INTERACTIONS ON CALCIUM BINDING AFFINITY AND MOLECULAR
RECOGNITION

by

LISA MICHELLE JONES

Major Professor: Jenny J. Yang
Committee: Jenny J. Yang
Giovanni Gadda
Alfons Baumstark

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2006

Acknowledgments

First and foremost I would like to thank my Lord and Savior Jesus Christ for his grace and mercy in my life. I would not have made it this far if not for his strength. His grace has been sufficient for any situation I have found myself in.

To my parents, thank you for all of the love and support throughout my life. I appreciate the both of you. Thanks for the encouragement on the days I called crying my eyes out. I love you! I would like to thank my sisters, Lauren, Kim, and Kelly, for always being there for me. I appreciate you guys and I love you. To Bishop Michael and Pastor Tanda Canon, thank you for all of the prayers, love, and support. You have been my Atlanta parents and I appreciate you both for every word you have spoken into my life. I would have been a much lesser person if God had not brought the both of you into my life.

To Dr. Jenny Yang, thank you for all of the support. I have learned a lot from you. Even when you think I was not listening, I was. I appreciate everything that you have done for me and for my career. You took me in when I was out in the cold and I will forever be thankful to you for that. I am the scientist I am today because of you and your constant pushing to be the best. Thank you, thank you, thank you!

To the three most beautiful, awesome, women on the planet, Donna, Lucretia, and Teree. You guys have been great friends. We are the four musketeers. Donna, thanks for the many characters. Lucretia, thanks for being my music compatriate. Teree, you have completely changed my style and I love you for it. To Amy, the coolest, funniest person I know. We have been through a lot together but we are both still standing.

Never let anyone or anything break that beautiful spirit of yours. I am looking forward to watching you graduate and show the world what a great scientist you are. Charmita, Sandra, Janet, and Jennifer, I love you guys. I cannot wait to read all of your dissertations.

To the members of the Yang lab, I thank all of you for all of your help and support. Special thanks to Dr. Wei Yang for always knowing the answer. To Anna Wilkins-Maniccia, thanks for being a good friend and for all of the help you have given me. Thanks to April Ellis for all the help. To Jack, thank you for the silly times (remember when we first started in 560), all the help with NMR, and computer stuff. I appreciate you! To Kendra and Dan, thank you for all of your help. To Julian, the little brother I never wanted, thank you for expressing my proteins for me. To Shawn, I appreciate all of your help with purification

I would also like to thank Dr. Stuart Allison for his help with calculations and for allowing me to use his program surfprot and Dr. Donald Hamelburg for his helpful advice with calculations. Thank you to Dr. Russell Malchow for the ICP-MS data. I would also like to thank Dr. Alfons Baumstark for all of his encouragement especially when I need it most. I would like to send a special thanks to Dr. Dabney Dixon for all of her helpful advice and support. Thanks to Dr. Gadda for always asking questions at my presentations.

To my Assembly of Truth church family I appreciate all of you. You have been my family here in Atlanta. Sisters and brothers I could count on in any situation. I love all of you. To Minister Johnson, you have spoiled me rotten and I have enjoyed every

minute of it. I love you! Cyrelle, what can I say, I love you girl! Timeka, you were always there with a smile and a word of encouragement. Thank you!

This work is supported by the grants received to Dr. Jenny J. Yang from GSU, the GAANN fellowship, the National Science Foundation (NSF), and the National Institutes of Health (NIH).

Table of Contents

Acknowledgements	iv
List of Tables	xv
List of Figures	xvii
List of Abbreviations	xxiv
Chapter 1 Introduction	1
1.1 The role of calcium in biological systems	1
1.2 General structural properties of calcium binding proteins	10
1.3 Calcium binding affinity and metal selectivity of proteins	19
1.4 Local factors that play a role in calcium binding affinity	22
1.5 Hydrogen bonding that contributes to Ca^{2+} binding affinity	27
1.6 Effect of electrostatic interactions on Ca^{2+} binding affinity	29
1.7 Previous approaches used to study Ca^{2+} binding affinity	34
1.8 Role of electrostatic interactions in molecular recognition of proteins	35
1.9.1 Cell adhesion function of cluster of differentiation 2 (CD2)	37
1.10.1 Our design approach and previous studies	40
1.10.2 CD2 and excellent host system	41
1.11.1 Objective and significance of this study	43
Chapter 2 Materials and Methods	47
2.1 Rational design of calcium binding proteins	47
2.2 Protein engineering	48

2.3	Sub-cloning of first generation proteins into pET-32a	50
2.4	Protein expression	52
2.4.1	Expression in the pGEX-2T vector	52
2.4.2	Expression in the pET32a vector	54
2.5	Protein purification	54
2.5.1	Purification of pGEX-2T expressed proteins	54
2.5.2	Purification of the pET32a expressed proteins	55
2.6	Far UV CD	55
2.7	Fluorescence	57
2.7.1	Trp fluorescence	57
2.7.2	FRET experiments	57
2.7.3	Rhodamine-5N (Rhod-5N) experiments	58
2.8	Nuclear magnetic resonance spectroscopy (NMR)	59
2.8.1	1D NMR	59
2.8.2	2D NMR	60
2.9	Surface plasmon resonance	61
2.10	ICP-MS experiments	62
2.11	Electrostatic potential calculations and visualizations	63
Chapter 3	Analysis of Electrostatic Interactions on CD2 and its Variants	65
3.1	Current methods for studying electrostatic interactions	65
3.1.1	The dielectric constant	67
3.1.2	The Poisson-Boltzmann equation	68

3.1.2.2	Increasing the precision of the PB equation	69
3.1.3	Debye-Huckel theory	71
3.1.4	Monte Carlo simulations	72
3.2.1	Calculating electrostatic potentials using DelPhi	72
3.2.2	Visualization of electrostatic surfaces	76
3.3	Electrostatic “hot spots” in CD2	77
3.4	Environment of the designed Ca^{2+} binding sites	84
3.5	Designing charge distribution variants of 6D79	86
3.6	Predicting Ca^{2+} binding affinity from free energy calculations	95
3.6.1	Free energy calculations	95
3.6.2	Predicting Ca^{2+} binding affinity through Debye-Huckel methods	100
3.7	Calculating pKa's with MCCE	106
3.8	Predicting stability using the Fold-X force field	114
Chapter 4	Conformational Analyses and Metal Binding of Designed Ca^{2+} binding proteins without a Stable Structure	120
4.1	The rationale for studying the first generation proteins	120
4.1.1	Cloning, expression, and purification of the first generation designed proteins	125
4.2	Conformational analyses of the first generation proteins	129
4.2.1	Circular dichroism (CD)	129
4.2.2	Fluorescence	131

4.2.3	The conformation of the apo form of the first generation proteins	132
4.2.4	The effects of metals on the conformation of the first generation proteins	134
4.3	Metal binding studies of the first generation proteins	134
4.3.1	Far UV CD	137
4.3.2	Fluorescence resonance energy transfer (FRET)	137
4.3.3	Inductively coupled plasma mass spectrometry (ICP-MS)	144
4.3.4	Metal selectivity	145
4.4	Comparison of the metal binding affinities of the three designed proteins	147
4.5	Reducing protein degradation by protease inhibitors	149
4.6	Rationale for the N80/89 mutant	156
4.7	Conformational analyses of the N80/89 mutant	157
4.8	Folding properties of CD2	157
4.9	Rationale for using the pET28AM1 vector to express and purify proteins	159
4.10	The pET32a vector	166
4.11	Expression and purification of N80/89 in the pET32a vector	166
4.12	Tb ³⁺ binding of N80/89 fusion protein	170
4.13	Second generation variants	172
4.13.1	Second generation designed protein 7E54	174

Chapter 5	Rational design of a novel calcium binding site adjacent to the ligand binding site on CD2 increases its adhesion function	177
5.1	Cell adhesion proteins	177
5.2	Electrostatic interactions and Ca^{2+} binding affinity	178
5.3	Rational design and selection of the novel Ca^{2+} binding site	180
5.4.1.1	Conformational analyses of CD2.6D79	182
5.4.1.2	High resolution NMR	184
5.4.2	CD2.6D79 has a reduced thermal stability	187
5.5	CD2.6D79 binds to Ca^{2+} and Ln^{3+}	191
5.6	Further structural validation of our designed Ca^{2+} binding site	203
5.7	CD2 binds CD48 with an enhanced affinity to the target protein	205
5.8	Common metal binding properties for naturally occurring calcium binding proteins	205
5.9	Implications of the designed novel Ca^{2+} binding protein	207
5.10	Implication of the electrostatic contribution on the cell adhesion	210
Chapter 6	Conformational analyses and stability of the charge distribution variants of CD2.6D79	216
6.1	The rationale for studying the charge distribution variants	216
6.2	Selection of the charge distribution variants	220
6.3	The 6D79.K91 variants	224

6.3.1	Conformational analyses of the 6D79.K91 variants	224
6.4	The 6D79.D94A variants	235
6.4.1	Conformational analyses of the 6D79.D94 variants	235
6.5	The 6D79.T75 variants	240
6.5.1	Conformational analyses of the 6D79.T75 variants	240
6.6	Stability of the charge distribution variants	243
6.6.1	Calculations of stability using DelPhi	243
6.6.2	Measurement of thermal stability in the apo form of the CD2.6D79 variants	245
6.6.3	Comparison of protein yields to stability	254
6.6.4	Comparison of calculated and experimental stabilities of the CD2.6D79 variants	254
6.6.5	Measurement of thermal stability in the metal bound form of the CD2.6D79 variants	256
6.7	Conclusions	257
Chapter 7	Metal binding properties of the 6D79 charge distribution variants	262
7.1	Rationale for studying the metal binding affinity of the charge distribution variants	262
7.2	The 6D79.K91 variants	263
7.2.1	Tb ³⁺ binding affinity of the K91 variants	264
7.2.2	Ca ²⁺ binding affinity of the K91 variants	264

7.3	The 6D79.D94 variants	273
7.3.1	Tb ³⁺ binding affinity of the 6D79.D94 variants	273
7.3.2	Ca ²⁺ binding affinity of the 6D79.D94 variants	273
7.4	Objective and significance of this study	278
7.4.1	Tb ³⁺ binding affinity of 6D79.T75K	278
7.4.2	Ca ²⁺ binding of T75K	278
7.5	The role of electrostatic interactions around the coordination shell on Ca ²⁺ binding affinity	282
7.6	The comparison of binding affinities to protein stability	282
Chapter 8	Conformational properties of the 6D79.R34 mutants	285
8.1	Introduction	285
8.2	Conformational analysis of the R34 mutants	287
8.3	pH titration using far UV CD and fluorescence	287
8.4	Removing the charge by R34I	293
8.5	Reducing charge repulsion by reversing charge interactions (the R34E-D72K mutant)	296
8.6	Comparison of the purification yields of the R34 mutants	300
8.6.1	Disruption of the native non-covalent interactions of CD2	302
Chapter 9	Significance of this study	305
9.1	Review	305
9.2	The structural requirements of Ca ²⁺ binding sites	306

9.3	The effect of the electrostatic environment on molecular recognition	308
9.4	The role of electrostatic interactions around the coordination shell on thermal stability and Ca^{2+} binding affinity	309
9.5	The role of electrostatic interactions on protein folding	311
9.6	Protein design as an important tool for studying the properties of proteins	311
	Publications	312
	Appendix	313
	References	354

List of Tables

Table 1.1	Radii of divalent ions	23
Table 1.2	The variant proteins of CD2 engineered for the research study	46
Table 2.1	Primer table	49
Table 3.1	Electrostatic potential of CD2 under different salt conditions	78
Table 3.2	Electrostatic potentials of charged residues	80
Table 3.3	Electrostatic environment of CD2	85
Table 3.4	Charged residues on the surface of 6D79	89
Table 3.5	The charge distribution variants	93
Table 3.6	Calculated binding affinities of the variants	97
Table 3.7	Comparison of pKa values of CD2	103
Table 3.8	Comparison of pKa values of calbindin D _{9k}	110
Table 3.9	Comparison of pKa values of CD2 calculated by MCCE	113
Table 3.10	Comparison of pKa values with different dielectric constants	116
Table 3.11	Predicted conformational stabilities	118
Table 4.1	Ca ²⁺ binding ligands of the first generation proteins	124
Table 4.2	Bacterial expression media recipes	127
Table 4.3	Metal binding affinities of the first generation proteins	148

Table 5.1	Summary of metal binding affinity and thermal stability of CD2.6D79 to different metal ions	193
Table 6.1	Selection of the charge distribution variants	221
Table 6.2	The primary sequence of the charge distribution variants	223
Table 6.3	Comparison of the chemical shifts of the K91 variants	229
Table 6.4	Melting temperatures and electrostatic free energy of The CD2.6D79 variants	246
Table 6.5	Purification yields of the charge distribution variants	255
Table 7.1	Binding affinities of the charge distribution variants	283
Table 8.1	Proteins studied for protein folding	288

List of Figures

Figure 1.1	Ca ²⁺ regulation in different cellular compartments	2
Figure 1.2	Conformational change of recoverin	4
Figure 1.3	The trigger protein calmodulin	5
Figure 1.4	The Ca ²⁺ dependent cell adhesion protein E-cadherin	7
Figure 1.5	The thermal stable Ca ²⁺ binding protein thermolysin	8
Figure 1.6	Geometry of a Ca ²⁺ binding site	11
Figure 1.7	Secondary structure elements of Ca ²⁺ binding liganding residues	12
Figure 1.8	The EF-hand motif	14
Figure 1.9	Ca ²⁺ binding site of calbindin D _{9K}	16
Figure 1.10	C-terminal domain of calmodulin	17
Figure 1.11	The semicontinuous binding site of the galactose binding protein	18
Figure 1.12	The discontinuous binding sites of E-cadherin	20
Figure 1.13	Variations in calcium binding affinity	21
Figure 1.14	The structure of thermitase	26
Figure 1.15	An Asx turn in site IV of turkey troponin C	28
Figure 1.16	Ca ²⁺ acts as an electrostatic switch in the C ₂ A domain of synaptotagmin	30
Figure 1.17	The electrostatic environment in Ca ²⁺ binding	32
Figure 1.18	Surface charges of calbindin D _{9K}	33

Figure 1.19	Topology of CD2 compared to cadherin	38
Figure 1.20	Structure of CD2	42
Figure 2.1	The pET32a vector	51
Figure 2.2	The pGEX-2T vector	53
Figure 3.1	Boundary conditions	70
Figure 3.2	DelPhi charge file	74
Figure 3.3	Descriptive charge file	75
Figure 3.4	Electrostatic potentials of all of the residues in CD2	79
Figure 3.5	Charged surface of CD2	82
Figure 3.6	Charged clusters in different regions of CD2	83
Figure 3.7	Model structure of 6D79	87
Figure 3.8	Electrostatic potential changes	90
Figure 3.9	Position of the charge distribution variants	94
Figure 3.10	Schematic diagram of the Debye-Huckel model	102
Figure 3.11	pKa values of CD2	104
Figure 3.12	pKa values of calbindin D _{9k}	109
Figure 3.13	pKa values of CD2 calculated by MCCE	112
Figure 3.14	pKa calculations with an increased dielectric constant	115
Figure 4.1	The first generation designed proteins	123
Figure 4.2	Optimization of expression of the first generation proteins	128
Figure 4.3	Expression of the first generation proteins	130
Figure 4.5	Conformational analyses of the first generation proteins	133

Figure 4.6	Conformational analysis of 206 in the presence of different metal ions	135
Figure 4.7	Conformational analysis of 6775 in the presence of different metal ions	136
Figure 4.8	Far UV CD of 5606 with increasing additions of Ca^{2+}	138
Figure 4.9	Enhancement of Tb^{3+} fluorescence emission upon the addition of protein	141
Figure 4.10	Tb^{3+} binding affinity of the first generation proteins	142
Figure 4.11	La^{3+} binding affinity of the first generation variants	143
Figure 4.12	Selectivity of the first generation designed proteins	146
Figure 4.13	Cleavage of the first generation proteins	150
Figure 4.14	Optimization of protease inhibitors	153
Figure 4.15	SDS-PAGE analysis of 6775 with different inhibitors	154
Figure 4.16	SDS-PAGE analysis of 6775 with thrombin and different inhibitors	155
Figure 4.17	Conformational analyses of N80/89	158
Figure 4.18	Hydrophobic interior of CD2	160
Figure 4.19	Optimization of competent cells for expression	162
Figure 4.20	Optimization of expression media	163
Figure 4.21	Optimization of IPTG concentrations for expression	165
Figure 4.22	Expression in the pET32a vector	167
Figure 4.23	Expression in the pET32a vector using 0.2 mM IPTG	168

Figure 4.24	Purification of N80/89 expressed in the pET32a vector	169
Figure 4.25	Further purification of N80/89 thioredoxin	171
Figure 4.26	Tb ³⁺ binding affinity of N80/89 thioredoxin	173
Figure 4.28	Expression and purification of 7E54	175
Figure 4.29	Conformational analyses of 7E54 in the presence and absence of Ca ²⁺	176
Figure 5.1	Structure of CD2.6D79	181
Figure 5.2	Conformational analyses of CD2.6D79	183
Figure 5.3	1D NMR of CD2.6D79	186
Figure 5.4	TOCSY of CD2.6D79	188
Figure 5.5	Chemical shift changes upon metal binding	189
Figure 5.6	Thermal denaturation of CD2.6D79	190
Figure 5.8	HSQC assignment of CD2.6D79	194
Figure 5.9	HSQC Ca ²⁺ titration	195
Figure 5.10	Ca ²⁺ binding of CD2.6D79	199
Figure 5.11	Lanthanide binding to CD2.6D79	200
Figure 5.12	La ³⁺ binding to CD2.6D79	201
Figure 5.13	Selectivity of CD2.6D79	202
Figure 5.14	Mn ²⁺ relaxation of the HSQC spectra	204
Figure 5.15	CD48 binding of CD2.6D79	206
Figure 5.16	Chemical shift changes of Ca.CD2	209
Figure 5.17	Surface potential of the CD48 binding surface	212

Figure 5.18	Electrostatic potential of R87	213
Figure 6.1	Positions of the charge distribution variants	222
Figure 6.2	Proximity of Asp2 to Lys91	225
Figure 6.3	Conformation of the 6D79.K91 variants	226
Figure 6.4	NMR of the 6D79.K91 variants amide region	228
Figure 6.5	The side chain of Lys91 is near the hydrophobic core of the protein	230
Figure 6.6	NMR of the K91 variants side chain region	231
Figure 6.7	Amide region of the K91 variants	233
Figure 6.8	HSQC spectra of 6D79 and 6D79.K91E	234
Figure 6.9	Hydrogen bonding of D94 to R96	236
Figure 6.10	Conformation of the 6D79.D94 variants	237
Figure 6.11	NMR of the 6D79.D94 variants	238
Figure 6.12	Proximity of T75 to D94	241
Figure 6.13	Conformation of the 6D79.T75 variants	242
Figure 6.14	NMR of 6D79.T75K	244
Figure 6.15	Thermal denaturation of 6D79.K91E	247
Figure 6.16	Thermal denaturation of 6D79.K91I	248
Figure 6.17	Thermal denaturation of 6D79.D94A	250
Figure 6.18	Thermal denaturation of 6D79.D94K	251
Figure 6.19	Formation of a salt bridge between R96 and E92	252
Figure 6.20	Thermal denaturation of 6D79.T75K	253

Figure 7.1	Tb ³⁺ binding of 6D79.K91E	265
Figure 7.2	Tb ³⁺ binding of 6D79.K91I	266
Figure 7.3	HSQC titration of 6D79.K91E	268
Figure 7.4	1D ¹ H NMR titration of 6D79.K91E	269
Figure 7.5	Chemical shift changes of 6D79.K91E	270
Figure 7.6	1D ¹ H NMR titration of 6D79.K91I	271
Figure 7.7	Chemical shift changes of 6D79.K91I	272
Figure 7.8	Tb ³⁺ binding of 6D79.D94K	274
Figure 7.9	Tb ³⁺ binding of 6D79.D94A	275
Figure 7.10	1D ¹ H NMR titration of 6D79.D94K	276
Figure 7.11	Chemical shift changes of D94K	277
Figure 7.12	Tb ³⁺ binding of 6D79.T75K	279
Figure 7.13	1D ¹ H NMR Ca ²⁺ titration of 6D79.T75K	280
Figure 7.14	Chemical shift changes of 6D79.T75K	281
Figure 8.1	Position of residue 34	286
Figure 8.2	Conformational analysis of the R34 mutants	289
Figure 8.3	pH titration of 6D79.R34E using far UV CD	291
Figure 8.4	pH titration of 6D79.R34E using Trp fluorescence	292
Figure 8.5	Conformational analysis of the 6D79.R34I variants	294
Figure 8.6	Thermal denaturation of 6D79.R34I	295
Figure 8.7	Residues R34 and D72 face the interior of the protein	297
Figure 8.8	Conformational analysis of 6D79.R34E-D72K	298

Figure 8.9	Space-filling model of CD2.6D79	299
Figure 8.10	Conformational analyses of 6D79.R34D-D72K	301
Figure 8.11	The hydrogen bonding network of CD2	303

List of Abbreviations

CaM	Calmodulin
CD	Circular Dichroism
CD2	Cluster of differentiation 2
DH	Debye-Huckel methods
EGTA	Ethylene Glycol-bis(β -aminoethyl Ether)
FRET	Fluorescence resonance energy transfer
FPLC	Fast performance liquid chromatography
GS4B	Glutathione sepharose 4B
GST	Glutathione-S-transferase
HSQC	Heteronuclear single quantum coherence
ICP-MS	Inductively coupled plasma mass spectrometry
Ig fold	Immunoglobulin fold
IPTG	Isopropyl β -D-thiogalactoside
NMR	Nuclear magnetic resonance
NOESY	Nuclear Overhauser spectroscopy
PB	Poisson-Boltzmann methods
PCR	polymerase chain reaction
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
T _m	Melting temperature
TOCSY	Total correlation spectroscopy
U(p)	Pseudo-energy
UV	Ultraviolet

Chapter 1: Introduction

1.1 The role of calcium in biological systems

Calcium is an alkaline earth metal whose ions are ubiquitous in biological organisms, processes, and structures. Calcium ions are essential structural component in the biomineralization of teeth, bones, and shells (McPhalen et al., 1991). As shown in Figure 1.1, in solution Ca^{2+} interacts with numerous Ca^{2+} binding proteins in different cellular compartments to mediate a wide range of physiological processes. Regulation is achieved through Ca^{2+} interacting with proteins of varying conformation, dynamic and metal binding properties. The majority of calcium-mediated biochemical processes result from the binding of Ca^{2+} to specific and selective sites on the surface of intra- and extracellular protein molecules (Strynadka and James, 1989).

Intracellularly, Ca^{2+} has an extraordinary number of diverse roles in metabolic regulation, nerve transmission, muscle contraction, cell motility, cell division and growth, secretion, and membrane permeability (McPhalen et al., 1991). Ca^{2+} acts as an intracellular second messenger with its concentration levels changing rapidly in response to extracellular stimuli. The increase in Ca^{2+} concentrations in cells causes the trigger proteins, calmodulin (CaM) and troponin C (TnC) among others, to bind Ca^{2+} . Upon Ca^{2+} binding the trigger proteins undergo a large conformational change which allows for the regulation of a substantial number of target proteins (Nelson and Chazin, 1998; She et al., 1998). For example Tanaka et al. have shown that Ca^{2+}

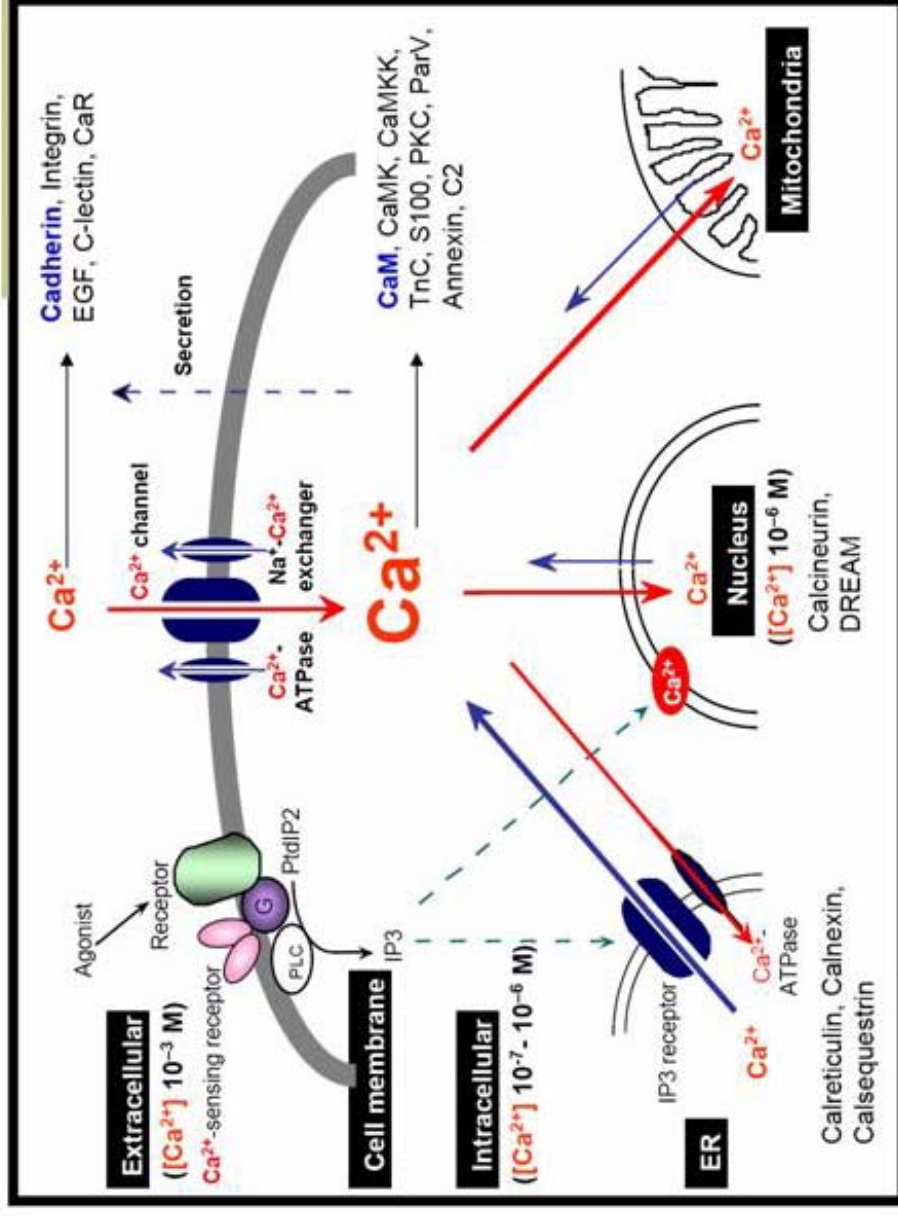


Figure 1.1. Ca^{2+} regulation in different cellular compartments. In solution Ca^{2+} interacts with numerous Ca^{2+} binding proteins in different cellular compartments to mediate a wide range of physiological processes. Modified from Schafer and Heizmann (Schafer and Heizmann, 1996).

binding to the intracellular retinal protein recoverin causes the exposure of an N-terminal group, which is buried in the Ca^{2+} free state of the protein, enabling membrane binding (Figure 1.2) (Tanaka et al., 1995).

The trigger protein calmodulin (CaM) is a ubiquitous, acidic protein found in almost all eukaryotic cells, in organisms ranging from yeast to plants to humans (Vogel, 2002). CaM can bind four Ca^{2+} ions, one in each of its EF-hand calcium binding motifs (Figure 1.3). CaM mediates more than 100 different target proteins in both the calcium-free and calcium-loaded forms (Yang, 2005). CaM binds to different target molecules at different cellular environments with varying Ca^{2+} concentrations. It affects numerous cellular processes including smooth muscle contraction, biosynthesis of other messenger molecules, protein phosphorylation and dephosphorylation, gene expression, and cell-cycle control (Vogel, 2002). Some proteins including neuromodulin and neurogranin, bind to the Ca^{2+} free form of CaM instead of the Ca^{2+} bound form (Chin and Means, 2000; Soderling and Stull, 2001).

Ca^{2+} levels in the extracellular matrix are several orders of magnitude higher than in the cytoplasm (2 mM versus 0.2 μM) (Vogel, 2002). The Ca^{2+} flux modulates Ca^{2+} regulatory proteins at key control points in essential physiological pathways. During development, differentiation of cell structures, tissue formation, and coordinated dynamic interactions of cells in various processes, require direct contact between cell surfaces and between cells and the extracellular matrix (Chothia and Jones, 1997). Ca^{2+} interacts with many of these cell adhesion molecules through which these contacts are made. One such

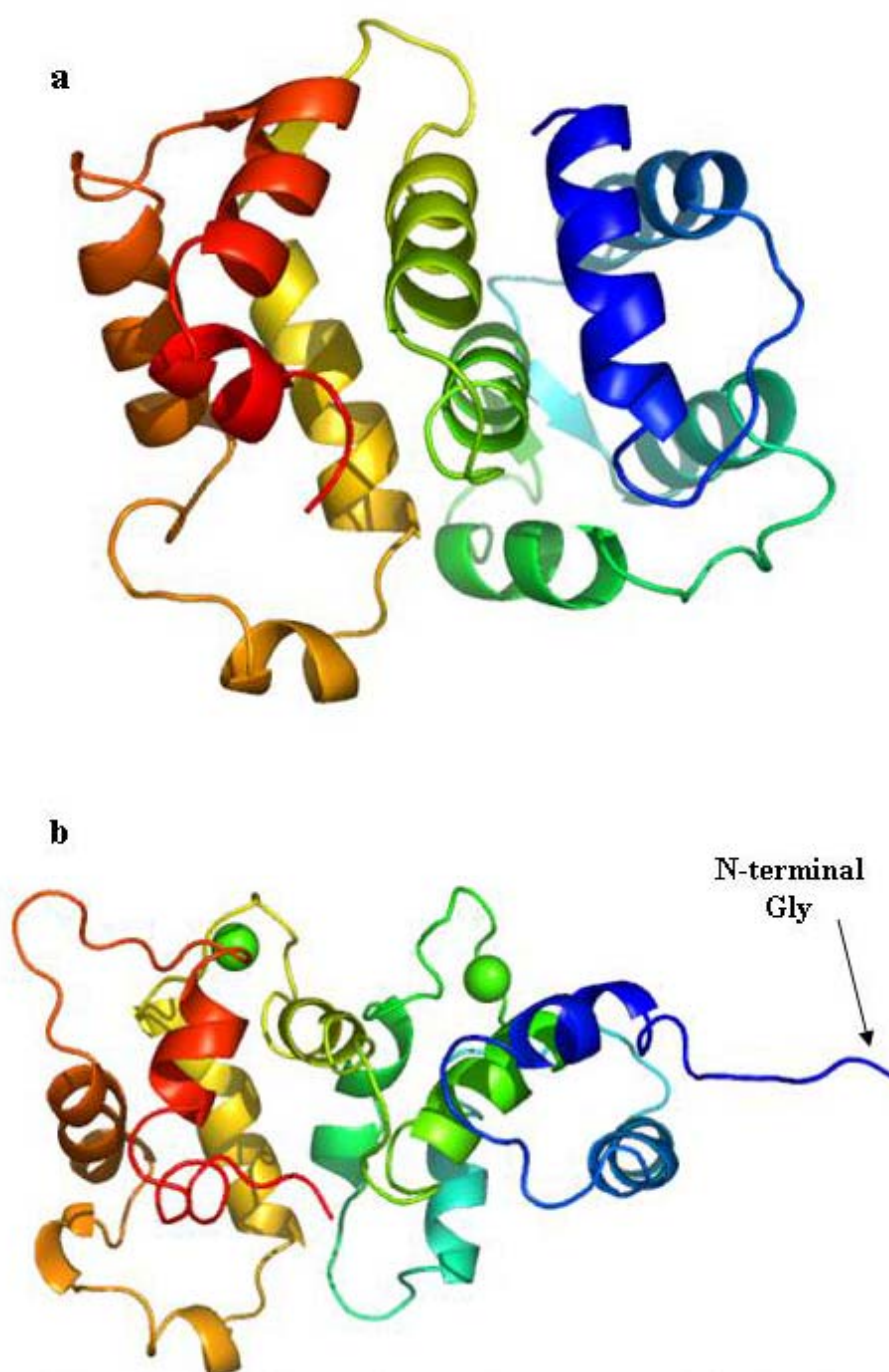


Figure 1.2. Conformational change of recoverin. The structure of the intracellular retinal protein recoverin in the Ca^{2+} free (1IKU (Tanaka et al., 1995)) (a) and Ca^{2+} bound (1JSA (Ames et al., 1997)) (b) forms shows that Ca^{2+} binding exposes an N-terminal group which enables membrane binding of the protein. Calcium shown in green balls.

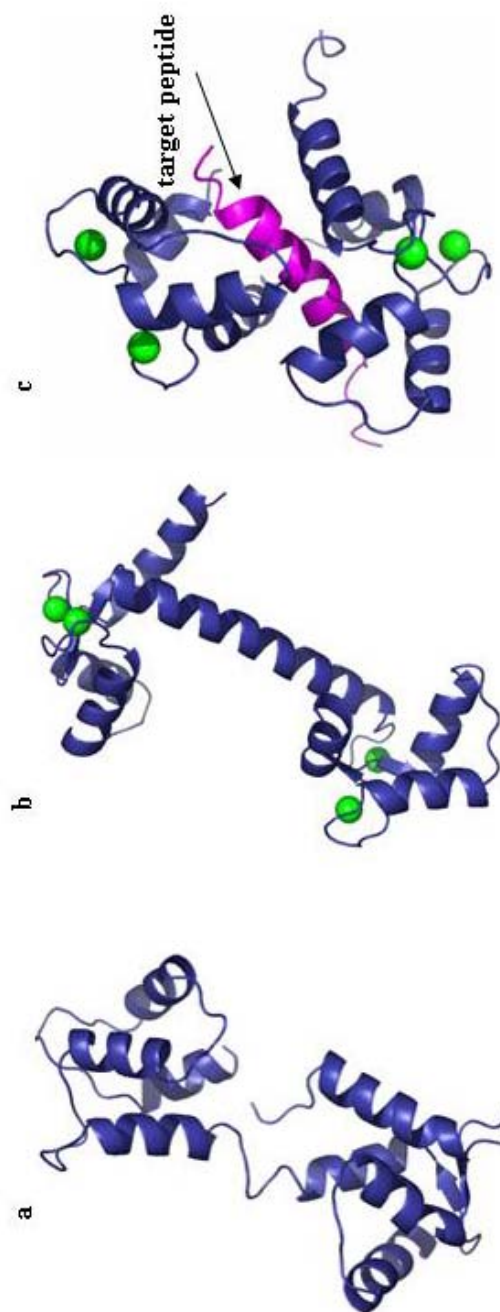


Figure 1.3. The trigger protein calmodulin. The trigger protein calmodulin changes its conformation upon binding four Ca^{2+} ions. (a) Ca^{2+} free form of calmodulin (1CFD (Kuboniwa et al., 1995)) (b) Ca^{2+} bound form of calmodulin (1CLL (Chattopadhyaya et al., 1992)) (c) Ca^{2+} bound to a target peptide, the calmodulin binding domain of myosin light chain kinase (shown in magenta) (2BBM (Ikura et al., 1992)). Calcium shown in green.

group of Ca^{2+} dependent cell adhesion molecules are cadherins which are homophilic cell-cell adhesive proteins (Figure 1.4). Additionally, Ca^{2+} plays a crucial role in the blood-clotting process through specific binding to enzymes that have post-translationally modified γ -carboxyglutamate (Gla) residues (Selander-Sunnerhagen et al., 1992). It is unclear as to whether the binding of Ca^{2+} to the Gla residues forms a bridge that allows these proteins to bind directly to the phospholipid membranes or if Ca^{2+} binding allows for the exposure of hydrophobic residues on the protein which brings about membrane binding (Sunnerhagen et al., 1995). However, it is known that the absence of Ca^{2+} severely limits blood clotting (Sunnerhagen et al., 1995).

Several proteolytic enzymes, including some members of the subtilisin and trypsin family of serine proteinases, the heat stable metalloproteinase thermolysin, DNase I, amylase, and dihydrofolate reductase, have been shown to require specific Ca^{2+} concentrations for maintenance of structural integrity and/or protection against proteolytic digestion (McPhalen et al., 1991). Since Ca^{2+} does not participate directly in catalysis, in fact in the majority of cases the Ca^{2+} binding sites are at least 10 Å away from the catalytic residues and lie within flexible loops (Leszczynski and Rose, 1986), it is presumed the Ca^{2+} ions serve to order the residues in the floppy loops (Strynadka, 1993). This aids in the maintenance of secondary structure and reduces the susceptibility to proteolytic digestion at these exposed areas of the enzyme. This is observed in the metalloproteinase thermolysin where the four Ca^{2+} ions are 17.8, 19.1, 19.4, and 31.4 Å from the catalytic site which consists of residues Glu143, His231, and Asp226 (Figure 1.5).



Figure 1.4. The Ca^{2+} dependent cell adhesion protein E-cadherin. Structure of the homophilic cell adhesion protein E-cadherin bound to three Ca^{2+} ions (1Q1P (Haussinger et al., 2004)). Calcium shown in green.

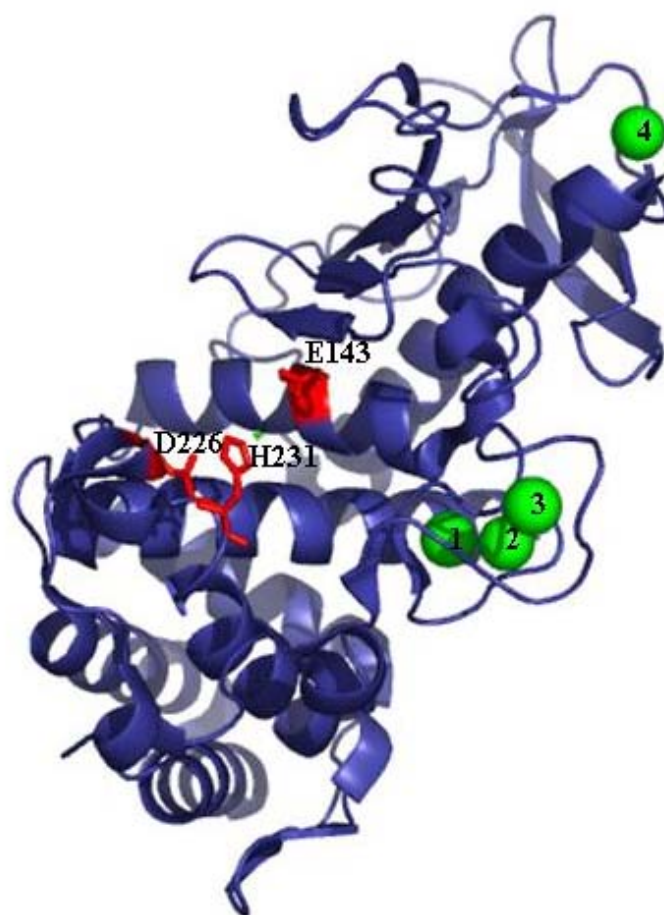


Figure 1.5. The thermal stable Ca^{2+} binding protein thermolysin. Structure of the thermal stable Ca^{2+} binding protein thermolysin. The catalytic residues Glu143, His231, and Asp 226 are highlighted in red. Ca^{2+} ions 1, 2, 3, and 4 (shown in green) are 17.8, 19.1, 19.4, and 31.4 Å from the catalytic site, respectively, and lie within the flexible loops of the protein (1FJV (English et al., 2001)).

Improper Ca^{2+} binding affinities of proteins such as calmyrin, calreticulin, and the parathyroid calcium sensing receptor have been implicated in the role of many diseases including Alzheimer's disease, cardiovascular disease, and osteoporosis (Nash et al., 1994; Nemeth et al., 1998; Stabler et al., 1999). Marfan syndrome is an autosomal dominant disease of connective tissue that occurs at a frequency of at least 1 in 10,000 in the population (Lee et al., 1991). Marfan syndrome is characterized by mutations in the gene encoding the human fibrillin-1 protein, a 350 kDa glycoprotein that is a major structural component of connective tissue microfibrils that have an average diameter of 10 nm (Sakai et al., 1986). Fibrillin-1 has a modular structure that includes 47 epidermal growth factor-like (EGF-like) domains, 43 of which contain a consensus sequence associated with calcium binding. A mutation causing an Asn 2144 to Ser amino acid change in one of the potential calcium binding residues has been observed in a patient with Marfan syndrome. Studies by Handford et al. on a mutant EGF-like domain containing the A2144S mutation show that the mutant protein exhibits a > 5-fold reduction in calcium binding affinity compared to the wild type protein as measured by ^1H -NMR (Handford et al., 1995). The mutation does not prevent assembly of fibrillin into microfibrils in Marfan patients but the appearance of the bead region is altered. A model developed by Handford et al. suggests that the fibrillin-1 EGF-like domains adopt a helical arrangement stabilized by calcium and the decreased calcium binding caused by the mutation results in distortion of the helix (Handford et al., 1995).

1.2 General structural properties of calcium binding proteins

The significant diversity of calcium action is possible because of the many ways in which the proteins of these systems can effectively bind the ion. Analysis of 22 Ca^{2+} binding proteins for which high resolution structures are available reveals that there are at least 18 markedly different ways to fold a polypeptide chain around a Ca^{2+} ion, based on the number of coordinating ligands, coordinating waters, and bidentate ligands (Strynadka, 1993).

Ca^{2+} ions are able to accommodate 4 – 12 ligands in their primary coordination sphere but coordination numbers of six or seven are the most common (Brown, 1988). The most frequent assembly is seven oxygen atom ligands arranged in a pentagonal bipyramidal fashion around the Ca^{2+} ion at an average Ca^{2+} -O distance of 2.4 Å (Figure 1.6). The majority of Ca^{2+} ligands in X-ray structures are oxygen atoms, although nitrogen ligands have been observed. A recent survey by Pidcock and Moore of a total of 515 X-ray structures of Ca^{2+} containing proteins indicate that the amino acid Asp has the highest calcium binding preference (Pidcock and Moore, 2001). The relative Ca^{2+} ligand population is acidic residues (Glu or Asp) > water > main chain carbonyl > Asn > Ser > Thr > Gln > Tyr. Analysis of the structural characteristics of protein Ca^{2+} binding sites have shown that these sites often lie within loops on or near the protein surface. The relative population of protein derived Ca^{2+} ligand residues are 72.6, 13.9, and 13.5%, respectively, for turn/loop, helix, and sheet secondary structure (Pidcock and Moore, 2001) (Figure 1.7). The abundance of ligand residues from turn/loop structure is most likely due to its flexible nature and its ability to supply a large number of amino acids

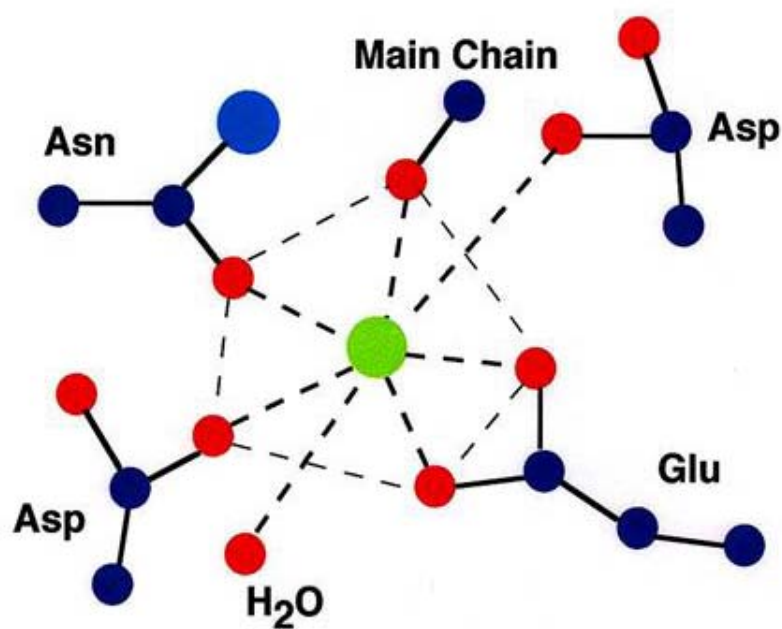


Figure 1.6. Geometry of a Ca^{2+} binding site. A common geometry found in Ca^{2+} binding proteins is the pentagonal bipyramidal geometry. This particular site has a coordination number of 7 with Glu as the bidentate ligand (Yang et al., 2002).

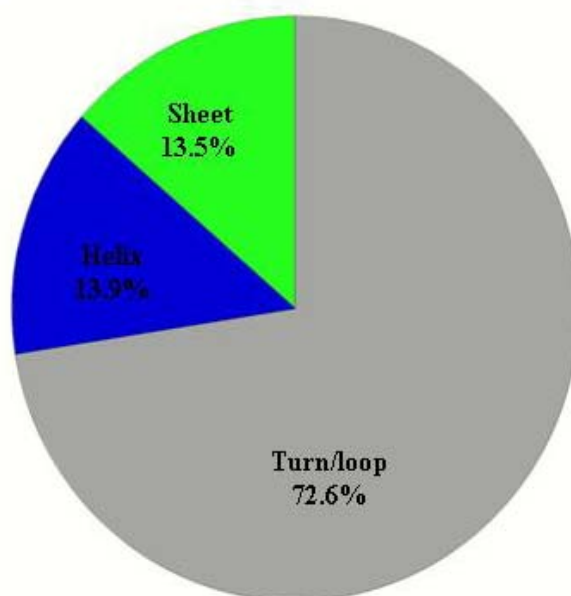


Figure 1.7. Secondary structure elements of Ca²⁺ binding liganding residues. Pie chart displays the prevalence of Ca²⁺ binding ligands from different secondary structural elements (Pidcock and Moore, 2001). The largest percentage of Ca²⁺ binding ligands are derived from turn/loop secondary structure while ligands from helices and sheets are not as prevalent.

from a short sequence. To the best of our knowledge there is not a natural Ca^{2+} binding site that has been found to have all of its ligands from β -sheet secondary structure presumably due to the rigid nature of β -sheets.

Ca^{2+} binding sites are generally classified into three types (Einspahr, 1984; McPhalen et al., 1991; Pidcock and Moore, 2001). Class I sites are characterized by all of the Ca^{2+} ligands being supplied from a continuous short length primary sequence of amino acids as seen in EF-hand proteins. The EF-hand, first described by Kretsinger (Kretsinger and Nockolds, 1973), has a conserved helix-loop-helix motif and is one of nature's most prevalent substrate binding motifs (Falke et al., 1994). Ca^{2+} binding ligands of EF-hand proteins are from a calcium-binding loop consisting of twelve consecutive residues. The conserved region of the EF-hand site extends for 29 consecutive residues including 9 residues from the first helix, followed by the 12 residue Ca^{2+} binding sequence, then by 8 residues from the second helix (Falke et al., 1994). In a classical EF-hand motif the residues located at positions 1, 3, 5, 7, and 12 of the loop provide the oxygen atoms that bind to the Ca^{2+} with position 12 as the bidentate ligand (Kawasaki et al., 1998) (Figure 1.8). Positions 1, 3, and 5 are highly conserved with Asp occurring at these sites most often (~59%, 76%, and 52%, respectively) (Falke et al., 1994). Position 7 is one of the most variable in the loop. The anchor residue at position 12 is highly conserved with the amino acid Glu (92%) the most prevalent residue (Falke et al., 1994). Two non coordinating positions, Gly at position 6 and a hydrophobic side chain at position 8 are also conserved. Position 6 has shown to be critical for the fold of loop and important to the hydrogen bonding network while position 8 is buried for loop

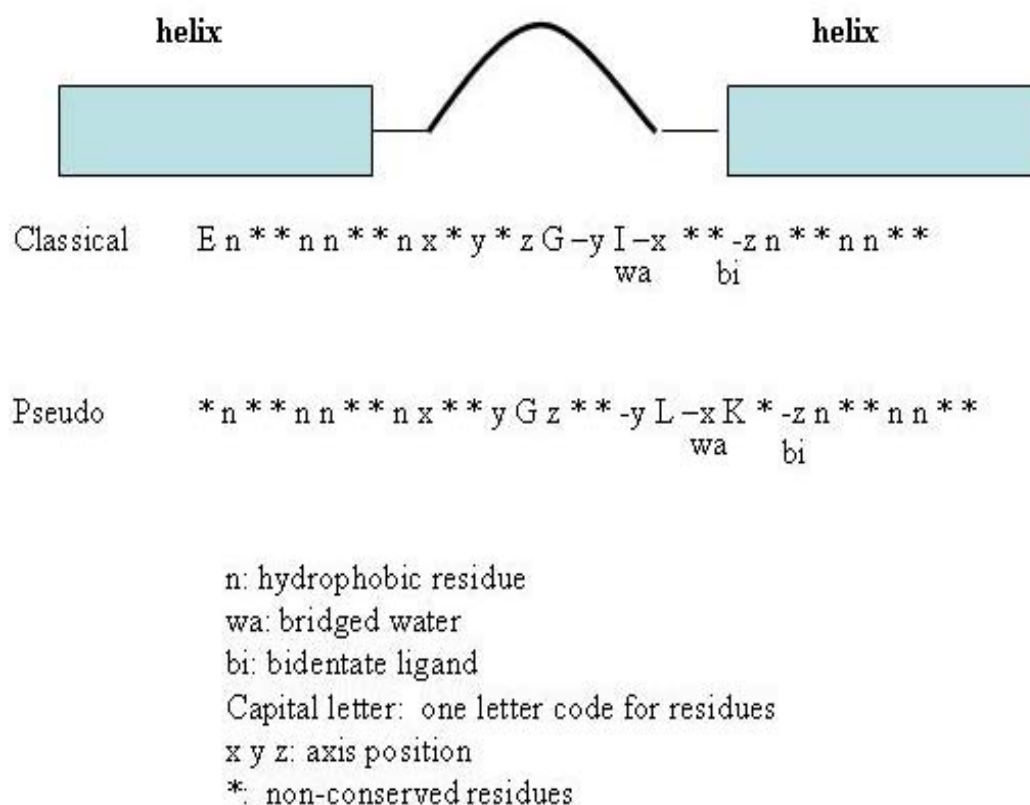


Figure 1.8. The EF-hand motif. The sequence patterns of the classical and pseudo EF-hand motifs. The pseudo EF-hand has two additional residues with the bidentate ligand at position 14 compared to position 12 for the classical motif. Modified from Yang and Yang (2005).

stabilization. The pseudo EF-hand motif has two additional residues as compared to the classical EF-hand with liganding residues at positions 1,4,6, and 9 and a bidentate carboxyl side chain Glu at position 14 (Yang, 2005) (Figure 1.8). In addition, residues at 1,4,6, and 9 use main-chain oxygen atoms for Ca^{2+} binding instead of side chain oxygens. All proteins from the S100 family have pseudo EF-hand motifs in site I. The buffering protein calbindin D_{9k} is a member of the S100 family and is the smallest EF-hand protein. It contains two Ca^{2+} binding sites, one has a classical EF-hand motif at the C-terminus and the other has a pseudo EF-hand motif at the N-terminus (Figure 1.9).

In native proteins the EF-hand motif are always paired, typically with conserved hydrogen bonding and hydrophobic interactions between the two EF-hand sites providing stabilization and a potential bridge for cooperativity (Strynadka and James, 1989). In addition, the EF-hand may be repeated in the same molecule yielding single or multi-domain structures containing two, four, or six Ca^{2+} binding sites (Moncrief et al., 1990; Nakayama et al., 1992). This phenomenon is observed in calmodulin which has two pairs of classical EF-hand motifs. Figure 1.10 shows the paired EF-hands in the C-terminus of calmodulin.

Class II sites, termed semi-continuous, have one ligand supplied by a part of the amino acid sequence far removed from the main binding sequence. Examples of class II sites include the galactose binding protein (Figure 1.11), site I of subtilisin and site I of thermitase.

Class III sites, where all of the ligands are supplied by amino acids remote from one another in the primary sequence, are also termed discontinuous sites. Examples of

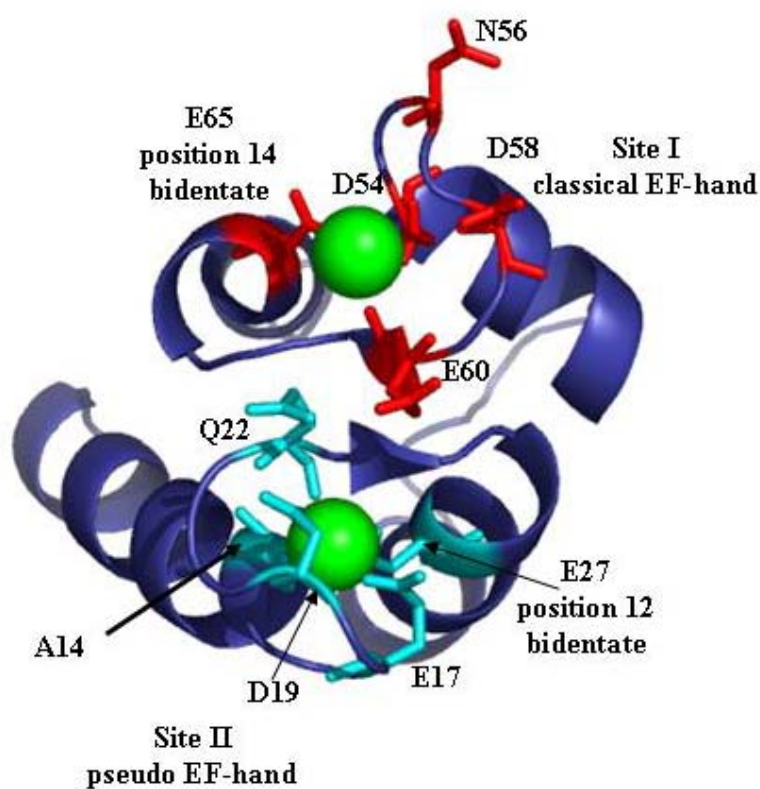


Figure 1.9. Ca^{2+} binding sites of calbindin D_{9K} . The two Ca^{2+} binding sites of calbindin D_{9K} . Site I (ligands shown in cyan) has a classical EF-hand motif with its bidentate ligand, E27, at position 12. Site II (ligands shown in red) has a pseudo EF-hand motif with its bidentate ligand, E65, at position 14 1B1G (Kordel et al., 1993). Calcium shown in green.

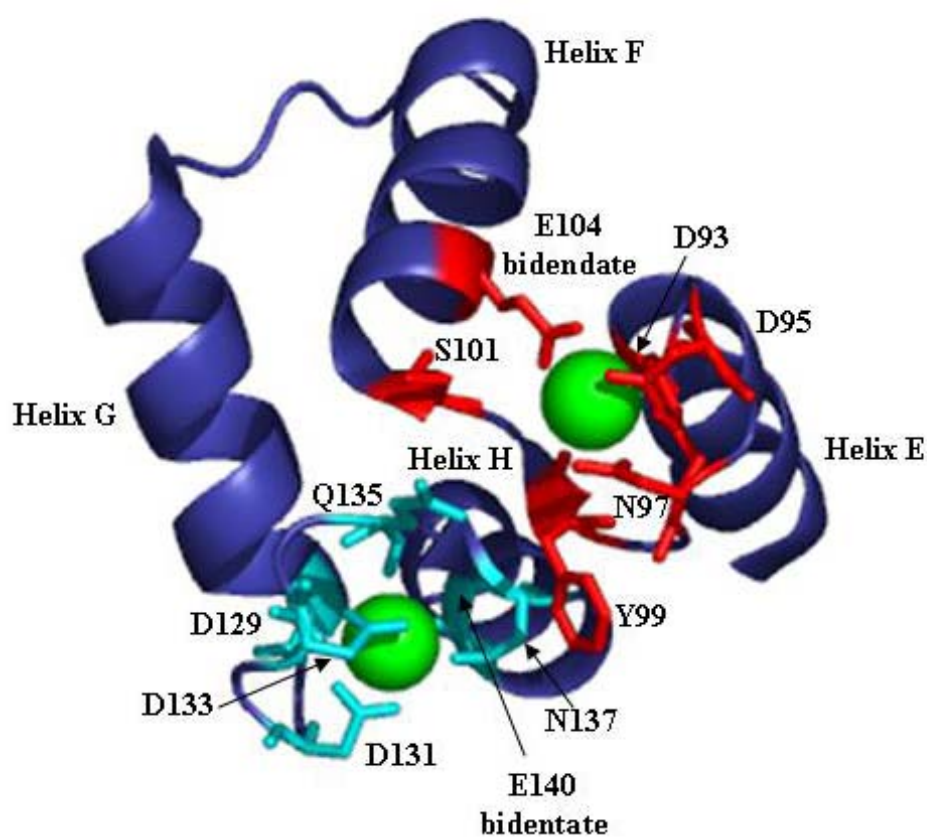


Figure 1.10. C-terminal domain of calmodulin. Paired EF-hand motifs (site III and site IV) in the C-terminal domain of calmodulin. The Ca^{2+} binding ligands are highlighted (red for site III and blue for site IV) and the bidentate ligands have been identified 1J7P (Chou et al., 2001). Calcium ions are shown in green.

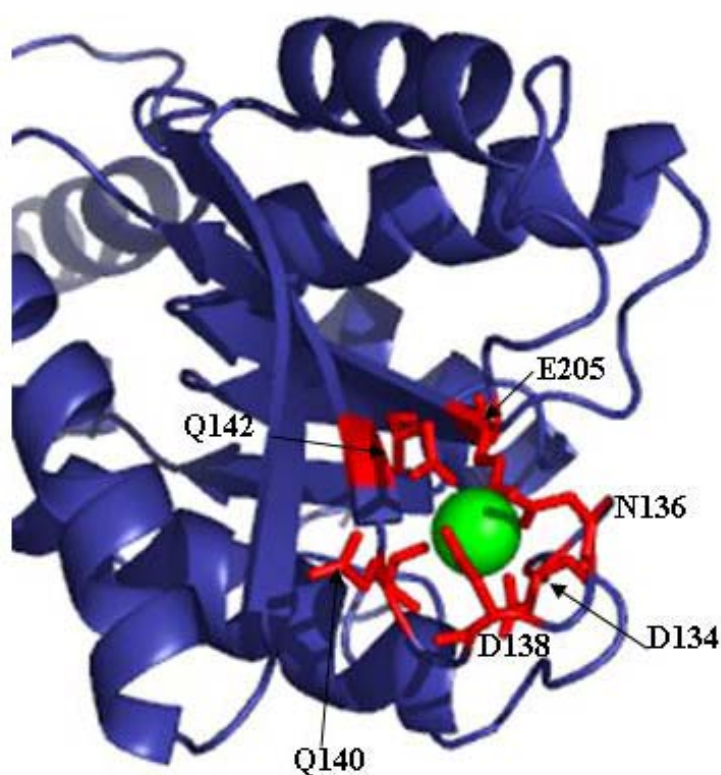


Figure 1.11. The semicontinuous binding site of the galactose binding protein. The Ca^{2+} binding site of the galactose binding protein is an example of a semicontinuous site. One of the ligands (E205) is supplied by a part of the amino acid sequence that is far removed from the other ligand residues in the primary sequence (2GBP (Vyas et al., 1987)). Calcium ions are shown in green.

discontinuous Ca^{2+} binding sites include site I of trypsin, sites I and II of proteinase K, and the cadherin family. Cadherins are a group of calcium dependent homophilic cell adhesion molecules. They are the major constituent of the cell-cell adhesive junctions, functioning together with numerous other proteins such as catenins. The structure of E-cadherin, so named because it is expressed in epithelial cells, is shown in Figure 1.12. The three Ca^{2+} ions share some of the same ligands but the binding sites are discontinuous. Interestingly, amongst all three classes of sites the pentagonal bipyramidal is the most common geometry utilized for Ca^{2+} binding.

1.3 Calcium binding affinity and metal selectivity of proteins

Ca^{2+} binding proteins have varying tertiary structures, Ca^{2+} affinities, and roles in the biological system. They have very little in common besides their ability to bind Ca^{2+} ions. The natural environment puts certain limitations to the range of Ca^{2+} affinities which are compatible with the particular biological role of each protein (Linse and Forsen, 1995) (Figure 1.13). The serine binding protein thermitase has an extremely strong binding affinity ($K_d = 10^{-10}$ M (Gros et al., 1991)) while the binding affinity of the buffering protein calbindin D_{9k} is four orders of magnitude weaker ($K_d = 10^{-6}$ M (Kesvatera et al., 2001a)) and the binding affinity of concanavalin A is 7 orders of magnitude weaker ($K_d = 10^{-3}$ M (Bouckaert et al., 2000)).

In addition, the selectivity of Ca^{2+} binding proteins over other physiologically relevant metals is important for their function. The high level of intracellular Mg^{2+} , for example, as compared to Ca^{2+} , imposes the necessity of discrimination against Mg^{2+} for Ca^{2+} binding proteins operating inside the cell. In fact Mg^{2+} is very similar to Ca^{2+} in that

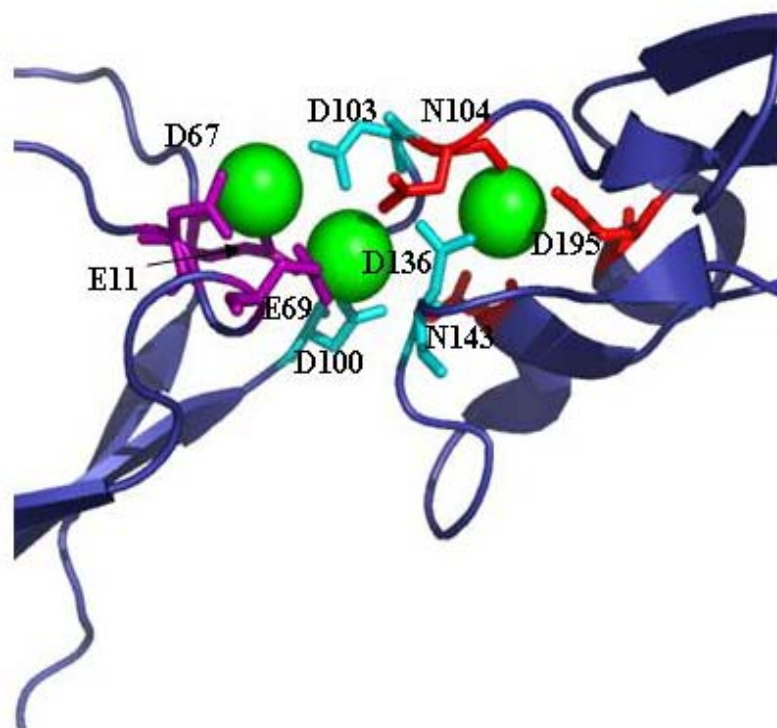


Figure 1.12. The discontinuous binding sites of E-cadherin. The three Ca^{2+} binding sites of E-cadherin located at the linker region between two domains are discontinuous. The ligand residues of sites I (red), II (cyan), and III (purple) are highlighted. All of the ligand residues are from far apart in the primary sequence. Site II and III share one ligand, D203 (1Q1P (Haussinger et al., 2004)). Calcium ions are shown in green.

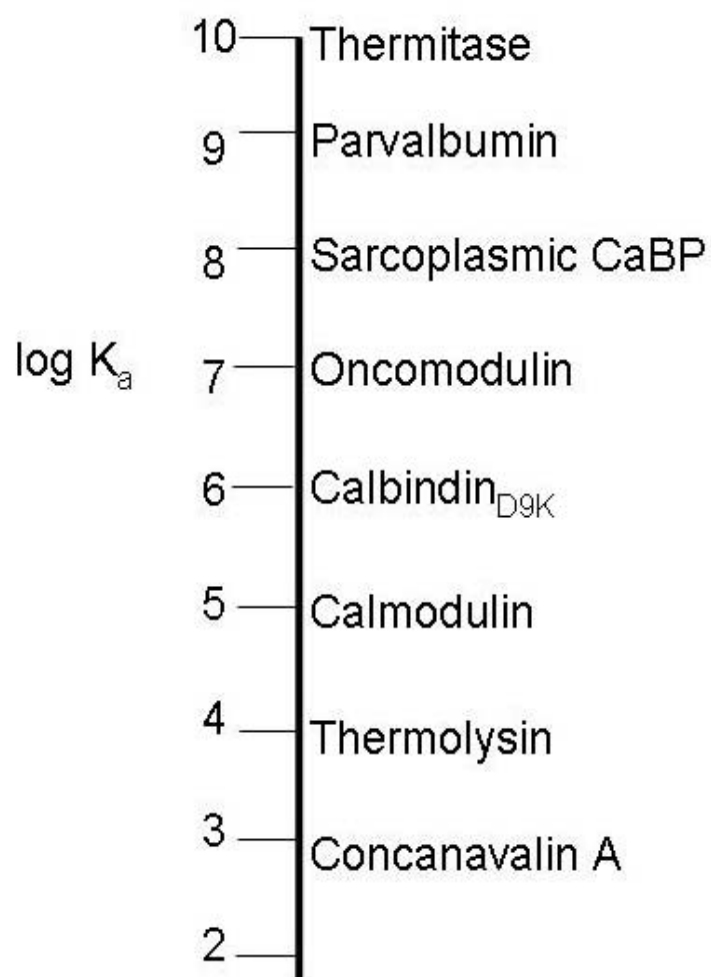


Figure 1.13. Variations in calcium binding affinity. Calcium binding proteins have a wide range of binding affinities depending on their biological function (Linse and Forsen, 1995).

it is a divalent alkaline earth metal that favors oxygen ligands. However, an important difference is that Mg^{2+} has a much slower rate for loss of water molecules from its hydration shell (Frausto da Silva, 1991), and its complexes are generally only six coordinate, with water molecules occupying at least two sites, which makes Mg^{2+} unfavorable for binding to the highly coordinate irregular binding sites of Ca^{2+} binding proteins (McPhalen et al., 1991). It is these types of sites that produce the greatest structural change upon metal binding enabling Ca^{2+} binding to act as a conformational trigger. The high net negative charge of the ligands at most Ca^{2+} binding sites favors divalent and trivalent cation binding over monovalent ions like Na^+ . In addition, the cavity size of a Ca^{2+} binding site affects the size range of ions than can be accommodated within it. The radius of Ca^{2+} is 1.0 Å, whereas other divalent ions Mn^{2+} , Fe^{2+} , Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+} , and Cd^{2+} have radii of 0.75, 0.70, 0.68, 0.65, 0.60, 0.65, and 0.90 Å, respectively (Table 1.1). Cd^{2+} which has the most similar radius to Ca^{2+} has a preferential binding to strong sulfur containing proteins (Williams, 2002). Another factor in metal selectivity may be the free energy of metal-ion dehydration which varies with hydration number of the free ion and ionic radius.

1.4 Local factors that play a role in calcium binding affinity

Local factors such as charge in the coordination shell and ligand type have been shown to be important for Ca^{2+} binding affinity. The net charge in the coordination sphere refers to the number of acidic residues directly involved in binding to the Ca^{2+} binding site. A detailed survey of 276 sequences of EF-hand proteins has shown that 64% of the Ca^{2+} binding loops have four acidic residues and 27% of loops have three

Divalent ion	Radius Å ^a
Ca ²⁺	1.0
Mn ²⁺	0.75
Fe ²⁺	0.70
Co ²⁺	0.68
Ni ²⁺	0.65
Cu ²⁺	0.60
Zn ²⁺	0.65
Cd ²⁺	0.90

^a(Williams, 2002)

Table 1.1. Radii of divalent ions

residues (Marsden et al., 1990). It has been suggested that four acidic residues provide the most stable arrangement of anionic charges (Reid and Hodges, 1980) and that reducing or increasing the number of acid residues at chelating positions will decrease the stability of the Ca^{2+} binding complex (Falke et al., 1994; Reid and Hodges, 1980).

Ca^{2+} binding sites also display a preference for ligand types at different liganding positions for the different classes of Ca^{2+} binding sites. For discontinuous Ca^{2+} binding sites, a detailed survey has shown that Asp is most prevalent residue utilized as the bidentate ligand (53%) (Pidcock and Moore, 2001). However, for the EF-hand motif, for the bidentate ligand at position 12 the most commonly utilized amino acid is Glu (92%) (Falke et al., 1994). Marsden et al. have examined the frequency of each amino acid in each of the 12 positions of the EF-hand motif (Marsden et al., 1990). Among 276 EF-hands they found 162 of them have Asp in position 1. High degrees of conservation are found at positions 3 (Asp in 120 cases, Asn in 40), 6 (Gly in 147) and 8 (Ile in 116 and Leu or Val in 36) (Marsden et al., 1990). This conservation of certain ligand residues indicates that ligand type is important for Ca^{2+} binding affinity.

Very few structural characteristics of Ca^{2+} binding sites correlate with Ca^{2+} binding affinities. Nevertheless, one trend in structure and affinity appears to involve the number of water ligands in the calcium binding site. The highest affinity sites use zero or one water ligand. Parvalbumins, the EF-hand protein with the greatest affinity for calcium (10^{-9} M (Linse and Forsen, 1995)), are also the only EF-hand with a “waterless” calcium binding loop (Strynadka, 1993). Moreover, in both the mannose binding lectin (Weis et al., 1992) and phospholipase A_2 (Scott et al., 1990) structures, a water ligand in

the calcium binding site of the native enzyme is replaced by an oxygen of the substrate when it bound to the enzyme. Both of these proteins exhibit higher Ca^{2+} affinities in the substrate bound form. The Ca^{2+} binding protein thermitase has three Ca^{2+} binding sites (Figure 1.14) with varying affinities of $< 10^{-10}$, 10^{-10} , and 10^{-4} M, for site I, II, and III, respectively (Gros et al., 1991). Both site I and III utilize the same pentagonal bipyramidal geometry but they have a six-fold difference in affinity. Site I does not use any water molecules as ligands whereas for site III, four out of seven ligands are oxygens from water molecules. Site II has a distorted square antiprism geometry with one of its eight ligands supplied by a water molecule. Gros et al. hypothesize the strong binding affinity of site I is due to a favorable entropy effect because Ca^{2+} is totally depleted of water ligands when bound to the protein (Gros et al., 1991).

Other local factors found to be important in Ca^{2+} binding affinity include minor changes in the average conformation upon Ca^{2+} binding (binding to a “preformed” site is most favorable) (Linse and Forsen, 1995). Ion binding to a “preformed site”, sites requiring a minimal conformational change, is tighter and more favorable entropically than binding to a more flexible site. Three dimensional NMR structures of apo and Ca^{2+} bound calbindin D9K, a high affinity Ca^{2+} binding protein (10^{-6} M (Linse et al., 1988)), indicate the Ca^{2+} binding site is largely in place in the Ca^{2+} free form (Skelton et al., 1990). The ability of the protein to provide all Ca^{2+} coordinating oxygen atoms without invoking strain in the polypeptide chain is also an important factor for Ca^{2+} binding affinity.

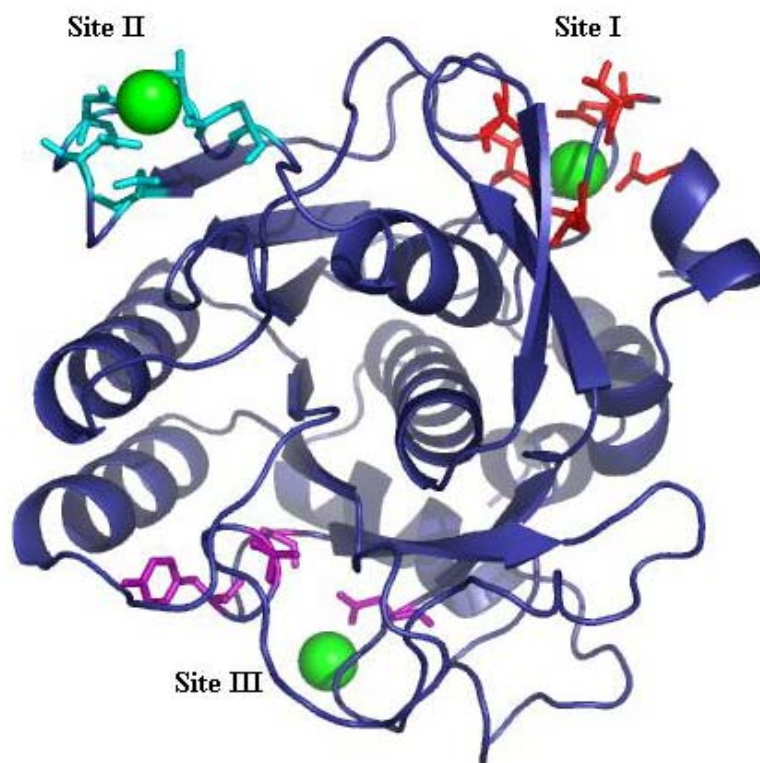


Figure 1.14. The structure of thermitase. The three Ca^{2+} binding sites of thermitase have drastically different binding affinities. The Ca^{2+} binding ligands are highlighted in red (site I), cyan (site II), and purple (site III) (3TEC (Gros et al., 1989)). Calcium ions are shown in green.

1.5 Hydrogen bonding that contributes to Ca^{2+} binding affinity

An extensive network of hydrogen bonding interactions among ligand and binding loop residues are also found in the majority of Ca^{2+} binding sites. Ca^{2+} liganding Asp and Glu residues form favorable hydrogen bonds with main-chain NH groups, with amide, hydroxyl, or basic groups on side-chains, and with ordered solvent molecules. These hydrogen bonding networks probably serve to position side-chain groups effectively to coordinate the Ca^{2+} ion, stabilize adjacent Ca^{2+} ligands (especially water molecules), stabilize changes in direction for the main-chain of the Ca^{2+} binding loop to bring other ligands into Ca^{2+} binding position, and aid in counteracting the repulsive electrostatic forces associated with juxtaposing two or more negatively charged amino acids within the relatively limited volume of the Ca^{2+} coordination sphere (Strynadka and James, 1989; Strynadka, 1993; Vyas, 1994). The importance of hydrogen bonding in Ca^{2+} binding affinity is evident with the strong affinity Ca^{2+} binding site (site I) of thermitase. Studies of the crystal structure of thermitase indicate residues from the N-terminal form a number of contacts with the Ca^{2+} binding loop of site I (Teplyakov et al., 1990). There are six direct hydrogen bonds between the N-terminus and the Ca^{2+} binding loop and it has been suggested these bonds give extra stability to Ca^{2+} site I which confers tighter Ca^{2+} binding (Teplyakov et al., 1990).

A common secondary structural element of many Ca^{2+} binding proteins is the Asx turn. The Asx turn, first described for Ca^{2+} binding proteins in the crystallographic analysis of turkey troponin C (Herzberg and James, 1988) (Figure 1.15), is a secondary structural interaction involving hydrogen bonding between the side-chain oxygen atom of

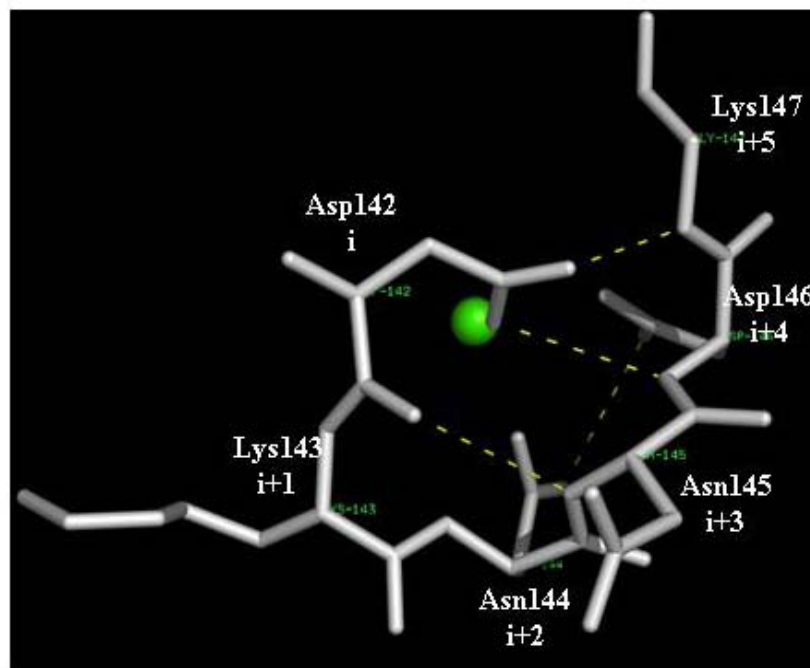


Figure 1.15. An Asx turn in site IV of turkey troponin C. The hydrogen bond network is shown in yellow. The Asx turn involves hydrogen bonding between the side chain oxygen atom of an aspartate, asparagine, serine or threonine at position i with the main chain NH of a residue at position $i+2$. This is evident in this site with the side chain of Asp142 hydrogen bonded to the main chain of Asn 144. Additionally, the side chain of Asn144 is hydrogen bonded to the main chain of Asp146 (STNC (Herzberg and James, 1988)).

an aspartate, asparagine, serine, or threonine residue at position i with the main-chain NH of a residue at position $i+2$ (Baker and Hubbard, 1984; Richardson, 1981). Asx turns occur within each of the 12 residue high affinity Ca^{2+} binding loops in the troponin C EF-hand structure (Strynadka, 1993). The Asx turn is a highly conserved feature of Ca^{2+} binding sites in the EF-hand family and is also prevalent in class II and III binding sites that are found in loop secondary structure. The role of the Asx turn in Ca^{2+} binding proteins appears to be both stabilization of the main-chain conformation and correct positioning of a ligand oxygen atom.

1.6 Effect of electrostatic interactions on Ca^{2+} binding affinity

Electrostatic interactions have been shown to be important in Ca^{2+} dependent molecular recognition and for Ca^{2+} binding affinity. Studies by Ubach et al. have shown that Ca^{2+} can act as an electrostatic switch in protein-protein interactions. The synaptic vesicle protein synaptotagmin I is composed of a C_2A domain which binds Ca^{2+} ions as well as phospholipids in a Ca^{2+} dependent manner. The binding of the C_2A -domain of synaptotagmin I to its target protein syntaxin is triggered by a change in the electrostatic potential around the calcium binding site (Figure 1.16). According to the electrostatic switch model, the C_2A domain is in close proximity to syntaxin before calcium influx, but electrostatic repulsion between the two proteins prevents fusion from proceeding. Upon calcium influx and the sequential binding of three Ca^{2+} ions to the C_2A domain, syntaxin is attracted through electrostatic interactions and fusion is triggered (Ubach et al., 1998).

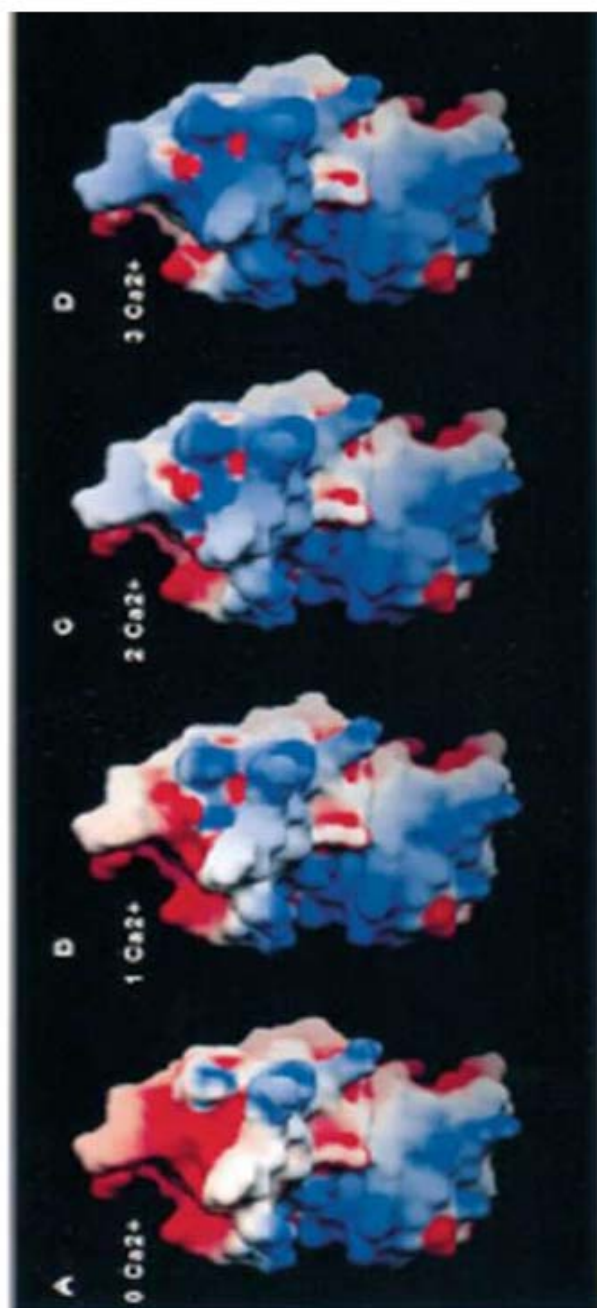


Figure 1.16. Ca^{2+} acts as an electrostatic switch in the C_2A domain of synaptotagmin. Changes in the electrostatic potential of the C_2A domain upon progressive binding of Ca^{2+} to sites $\text{Ca}1$, $\text{Ca}2$, and $\text{Ca}3$ (A-D). Reproduced from Ubach et al. (Ubach et al., 1998)

The importance of a negative electrostatic potential around a Ca^{2+} binding site is evident in the Ca^{2+} binding ability of α -lactalbumin. Alpha lactalbumin, found in milk, regulates lactose biosynthesis (McKenzie and White, 1987). The protein has significant sequence homology with lysozyme (Nitta and Sugai, 1989), a non- Ca^{2+} binding protein. The three-dimensional structures of the two proteins are also very similar (Stuart et al., 1986), Figure 1.17a and b. Lysozyme even contains a loop analogous to the Ca^{2+} binding loop in α -lactalbumin. It is hypothesized that the two protein-encoding genes share a common ancestor (Hopper and McKenzie, 1974). A further examination of the electrostatic surface of the two proteins reveals the explanation for their differences in Ca^{2+} binding properties. The Ca^{2+} binding surface of α -lactalbumin has a highly negative potential, as indicated by the red color in Figure 1.17c. Conversely, the analogous surface on lysozyme is highly positive (Figure 1.17d). This positively charged environment is not conducive for Ca^{2+} binding.

The electrostatic distribution outside of the coordination shell of the metal binding site is considered to be a key determinant that governs high affinity Ca^{2+} binding (Linse and Forsen, 1995). Previous analysis has shown that screening of electrostatic interactions of proteins by added inorganic salts reduces calcium binding affinity (Linse et al., 1988). The X-ray structure of bovine calbindin $\text{D}_{9\text{K}}$ reveals a cluster of negatively charged amino acid side chains on the protein surface around the calcium sites. The contribution of individual charges to the free energy of calcium binding to calbindin $\text{D}_{9\text{K}}$ was investigated by Linse and coworkers. They have shown that removal of three negative surface charges Glu, Asp, and Glu at positions 17, 19, and 26 (Figure 1.18) in

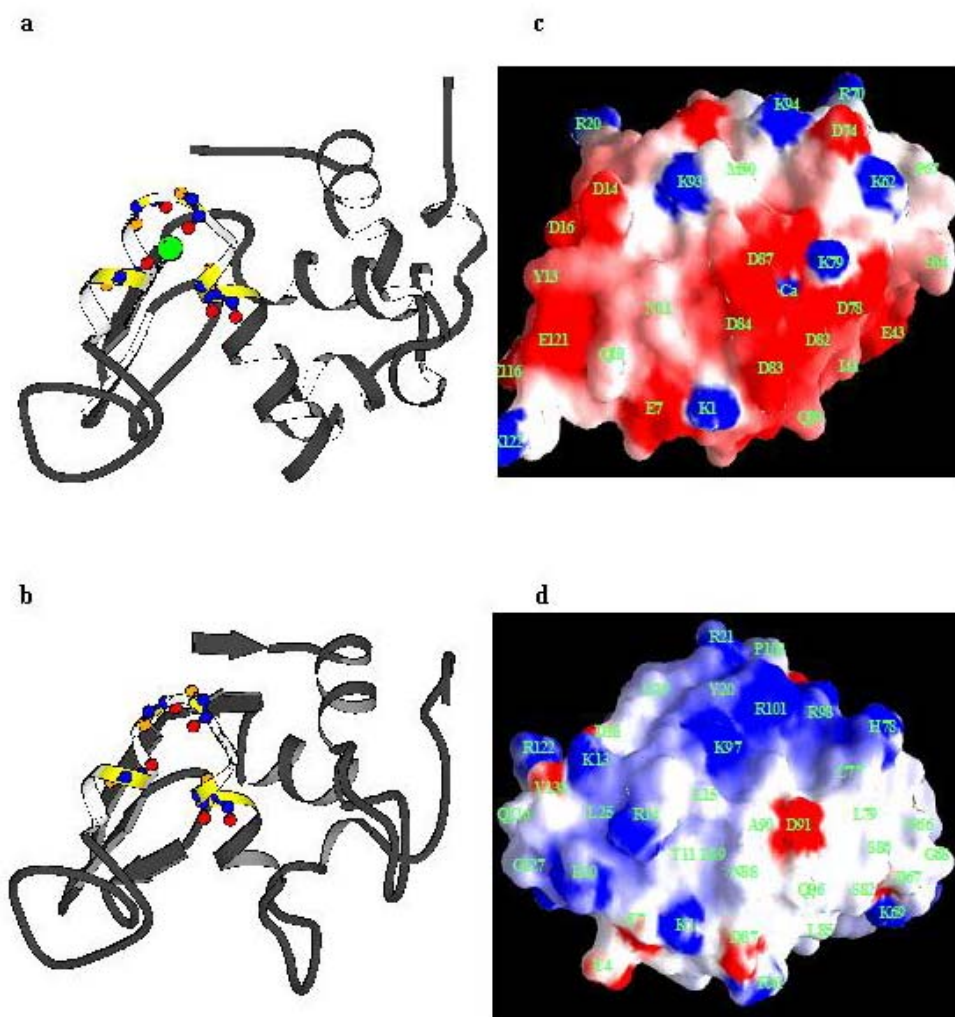


Figure 1.17. The electrostatic environment in Ca^{2+} binding. Comparison of the structure (a) and electrostatic surface (c) of bovine α -lactalbumin (1F6R (Chrysina et al., 2000)) to the structure (b) and electrostatic surface (d) of hen egg-white lysozyme (1E8L (Schwalbe et al., 2001)).

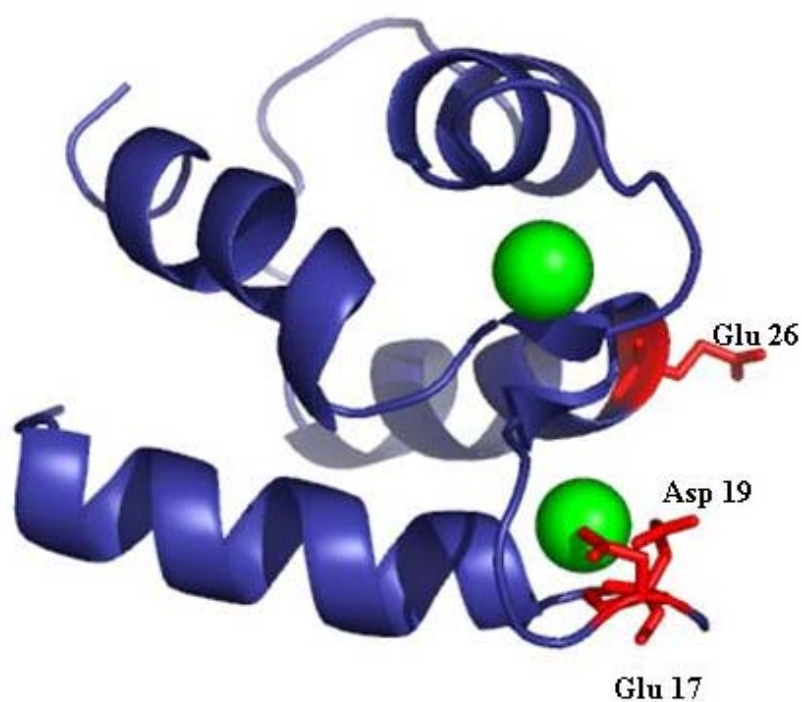


Figure 1.18. Surface charges of calbindin D_{9k}. Structure of calbindin D_{9k} with 3 surface charged amino acids highlighted that are important for Ca²⁺ binding affinity though not directly involved as Ca²⁺ ligands (1B1G (Kordel et al., 1993)).

the vicinity of the calcium binding sites of this protein leads to a 45 fold decrease in average affinity per site at low ionic strength and a five-fold reduction at 0.13 M KCl indicating the importance of electrostatic distribution (Linse et al., 1988; Linse et al., 1991). These residues are not directly involved as Ca^{2+} ligands but do have a significant effect on Ca^{2+} binding affinity. However, calbindin $\text{D}_{9\text{k}}$ has two Ca^{2+} binding sites and current methods used to study Ca^{2+} binding affinity do not allow for the study of affinity of a single site so the effect of electrostatic interactions on site-specific Ca^{2+} binding affinity is still not well understood. It is the focus of this dissertation to understand the effect of electrostatic interactions on Ca^{2+} binding.

1.7 Previous approaches used to study calcium binding affinity

Intensive studies have been carried out to understand the key factors that govern Ca^{2+} binding affinity. Several approaches have been utilized including site-directed mutagenesis, peptide models, and structural analysis. Mutagenesis studies on natural, intact Ca^{2+} binding proteins have been useful in determining the key residues important for Ca^{2+} binding affinity particularly for the identification of conserved residues in the EF-hand motif (Falke et al., 1994) (Henzl et al., 1996). Peptide models and fragments have been valuable in studying calcium binding affinity for one Ca^{2+} binding site at a time (Niki et al., 1996) (Marsden et al., 1988). Synthetic peptides have provided a plethora of insight into the Ca^{2+} binding properties of proteins in the EF-hand family of Ca^{2+} binding proteins. Studies of site III of calmodulin reveal that the Ca^{2+} binding affinity is increased 40-fold by replacing Asn97 with Asp underscoring the importance of ligand types and charge for binding affinity (Reid, 1990). Nevertheless, the peptide

model approach is only useful for the study of a continuous Ca^{2+} binding site. In addition, peptides lack stable structures in solution and therefore the metal binding affinity is dramatically reduced compared to an intact protein (Malik et al., 1987; Reid et al., 1980).

Although a great deal of research has been carried out to understand Ca^{2+} binding affinity there are many limitations to the current approaches. First, the cooperativity between multiple Ca^{2+} binding sites makes it difficult to study site-specific binding affinity. Second, many Ca^{2+} binding proteins undergo a large, global conformational change upon Ca^{2+} binding. In these cases it is difficult to separate the binding energy from the conformational change energy.

1.8 Role of electrostatic interactions in molecular recognition of proteins

Numerous experimental and theoretical studies have established connections between protein electrostatic properties and function (Nakamura, 1996; Perutz, 1994; Sharp and Honig, 1990). In the interior of a protein back-bone, polar atoms compensate their charge desolvation by forming favorable electrostatic interactions as main chain-main chain hydrogen bonds, resulting in the formation of secondary structures. The specificity of protein tertiary structure, binding, and their unique function stems from the spatial arrangement of polar and charge atoms, and by formation of electrostatic interactions, mainly salt-bridges, and main chain-main chain, main chain-side chain, and side chain-side chain hydrogen bonds (Sinha and Smith-Gill, 2002). For example, in thermophilic proteins, the driving force towards thermostability is through the

optimization of electrostatic interactions by increasing the number of salt-bridges (Kelly et al., 1993; Knapp et al., 1997; Spassov et al., 1995).

Analyses have shown that salt-bridges can be stabilizing or destabilizing, playing important structural and functional roles in both cases. Stabilizing salt-bridges rigidify local regions for a better fit due to functional requirements while destabilizing salt-bridges provide conformational specificity for the fold or function. High affinity and very specific binding of proteins also involves polar and charge interactions. Many enzyme inhibitor complexes have been shown to have minimal mean hydrophobicities, and the interactions are mainly through polar and charged residues (Jones and Thornton, 1996). This charge complementarity leads to specific binding of high affinity. Favorable electrostatic interactions have also been shown to be responsible for fast association rate constants due to simple diffusion between two proteins. Studies of the extracellular ribonuclease barnase and its natural inhibitor barstar show that electrostatic forces enhance their association rate by 500-fold (Gilboa et al., 2001). Both long range electrostatic interactions followed by short range electrostatic interactions during docking were proposed to be the mechanism behind the high affinity binding (Buckle et al., 1994). A good example for the role of electrostatic interactions in molecular recognition is the adhesion protein cluster of differentiation 2 (CD2) and its binding partner CD48.

1.9 Cell adhesion function of cluster of differentiation 2 (CD2)

Cluster of differentiation 2 (CD2) plays an important role in mediating both cell adhesion in T-lymphocytes and in signal transduction by interacting with its binding partner molecules, CD48 in rats or CD58 in humans (Arulanandam et al., 1993b) (Davis et al., 1998b). The CD2/CD58 system was the first heterophilic protein/protein interaction identified at the cell surface (Selvaraj et al., 1987). In addition, CD2 was the first protein discovered to be important in promoting intercellular adhesion and signal transduction. To date, CD2 is one of the most extensively studied cell adhesion proteins and has served as a model system to identify the key determinants involved in molecular recognition.

CD2 consists of two extracellular domains and an intracellular C-terminal domain linked by a single transmembrane region. The CD2 extracellular domains belong to the immunoglobulin superfamily (IgSF) (Chothia and Jones, 1997) (Figure 1.19). Nine β -sheets are sandwiched in two layers of GFCC'C'' and BED strands. Several structures of CD2 have been solved by NMR and X-ray crystallography (Jones et al., 1992); (Bodian et al., 1994). The folding, dynamics, and stability of this protein have been extensively studied ; (Yang et al., 2000b).

The monomeric binding affinities of human CD2 to CD58 and rat CD2 to CD48 are low and characterized by a fast off rate ($k_{\text{off}} > 4 \text{ s}^{-1}$) and weak affinity ($K_a = 10^4 - 10^5$) (Davis et al., 1998a). The fast off rate allows for transient reversible cell-cell interactions. The ligand binding site, on the GFCC'C'' face of CD2, is much flatter than binding sites in other proteins suggesting that the CD2-ligand interface has less

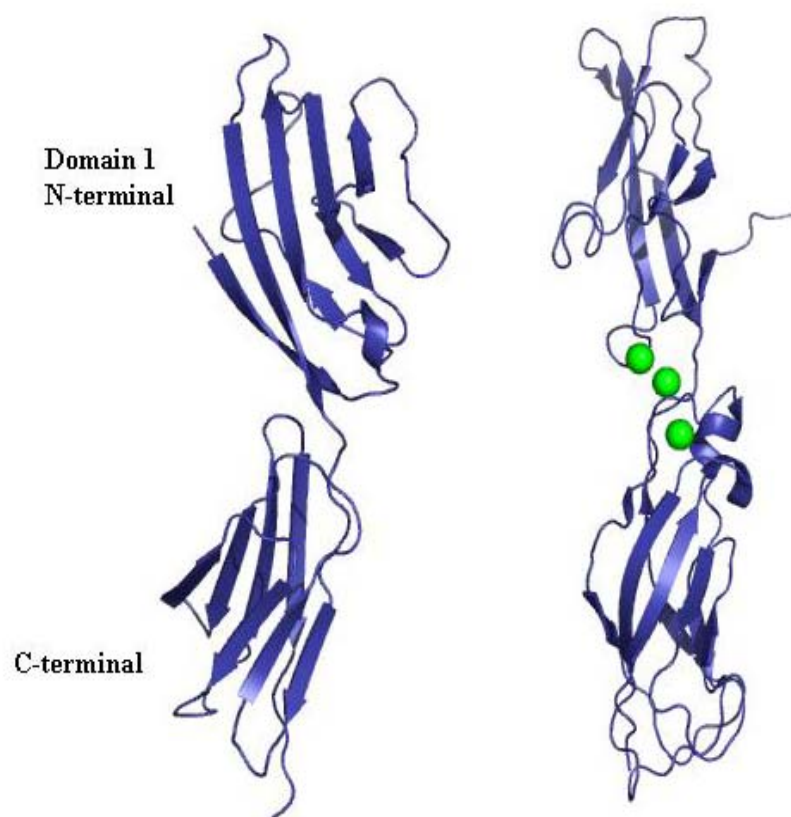


Figure 1.19. Topology of CD2 compared to cadherin. CD2 (1hng (Driscoll et al., 1991))(left) and the Ca^{2+} dependent cell adhesion proteins cadherin (1Q1P (Haussinger et al., 2004)) (right) have a similar β -sheet architecture. Both proteins belong to the immunoglobulin superfamily. Calcium shown in green.

surface complementarity (McAlister et al., 1996). Additionally, studies have shown that the GFCC'C'' surface of CD58 and CD48 are used for ligand binding. This is consistent with the possibility that they, along with CD2, evolved from a homophilic precursor.

The molecular recognition properties of CD2 have been extensively studied. The CD48 binding surface of CD2 contains several charged residues (70%). Many of these charged residues are highly conserved between species. The role of these charged residues in mediating this low affinity binding was investigated using site-directed mutagenesis studies (Davis et al., 1998a). Twenty-eight single mutations were engineered to examine which residues are important for binding. The results demonstrated that electrostatic interactions are not important for binding affinity. Rather the aromatic and aliphatic residues Y81, F49 and L38, are responsible for the ligand binding energy (Davis et al., 1998a). It was shown that the charged residues confer specificity for binding which accounts for their conservation between species.

Kim *et al.* have determined through site-directed mutagenesis studies that human CD2 interacts with CD58 using a "hot spot" that correspond to Y86 on CD2 and K34 on CD58 (Kim et al., 2001). A single mutation at Y86 to alanine decreased the binding interaction by ~1000 fold but a Y86F mutant had little or no effect upon binding, indicating the importance of an aromatic residue in binding (Kim et al., 2001). A similar "hot spot" of binding observed in human CD2 was revealed for rat CD2 where the structurally analogous aromatic residue Y81 has been shown to be important for CD48 binding (Davis et al., 1998a). In addition, the aromatic residue F49 and the aliphatic residue L38 are important for binding affinity.

1.10.1 Our design approach and previous studies

The study of site specific Ca^{2+} binding affinities have been hampered in naturally occurring Ca^{2+} binding proteins due to the cooperativity between multiple bound Ca^{2+} ions as well as Ca^{2+} induced conformational change. To better understand the role of long range electrostatic interactions on Ca^{2+} binding affinity and molecular recognition we have designed single Ca^{2+} binding sites into a host protein using the computer algorithm Dezymer (Yang et al., 2001b; Yang et al., 2002b). The use of rational design will allow us to examine the structural factors that control calcium binding affinity in an isolated system without the complications of cooperative metal-metal interactions and global conformational changes. The geometric parameters of a Ca^{2+} binding site including geometry, bond lengths, and bond angles were input into design algorithms such as Dezymer (Yang et al., 2001b; Yang et al., 2002b).

We have previously shown that this design method has been successful in designing Ca^{2+} binding proteins (Yang et al., 2003). The designed protein CD2.Ca1 is the first successful metalloprotein design that has a high coordination number (seven) metal binding site constructed into a host protein (Yang et al., 2003) (chapter 4). CD2.Ca1 binds Ca^{2+} and its analogs Tb^{3+} and La^{3+} with strong affinities and displays selectivity over other physiologically relevant metals Mg^{2+} and K^{+} . The results demonstrate the viability of designing a single calcium-binding site into a host protein, taking into account only local properties of a calcium-binding site (Yang et al., 2003).

The viability of this method was further validated with the design of the protein Ca·CD2. Similar to CD2.Ca1, this protein binds and is selective for Ca^{2+} and its analogs

over other mono and divalent cations (Yang et al., 2005). In addition, the NMR structure of this protein has been solved and the protein structure is similar to wild type CD2. Further, the validity of the design process was confirmed using a paramagnetic NMR technique based on the relationship between relaxation effects and distance. The study established that Ca^{2+} was indeed binding to the designed ligand residues. Surface plasmon resonance studies of this protein were carried out to examine the functional properties of this protein. This protein retains the ability to bind CD48 with a comparable binding affinity to CD2. The Ca^{2+} binding site of Ca·CD2 is on the opposite surface of CD2 than the CD48 binding site and therefore the mutated ligands do not play a role in CD48 binding.

Finally the design approach has been used to examine the role of charge and ligand type in Ca^{2+} binding affinity. The designed protein 7E15 binds Ca^{2+} with a strong affinity ($100 \pm 50 \mu\text{M}$) and was used as a model for the study of charge and ligand in Ca^{2+} binding (Maniccia, 2006). Several variants of this protein were engineered to alter the charge in the Ca^{2+} coordinating shell (for -2 to -5) as well as the ligand type (comparison of Asn to Gln).

1.10.2 CD2 and excellent host system

The host system for our design approach is domain 1 of the rat cell surface adhesion receptor CD2 (Figure 1.20). The β -sheet architecture of CD2 is similar to the extracellular Ca^{2+} dependent cell adhesion molecule cadherin (Nagar et al., 1996) (Figure 1.19). Therefore, information obtained by designing Ca^{2+} binding sites in CD2 can provide insight into Ca^{2+} dependent cell adhesion properties. Previous studies on CD2

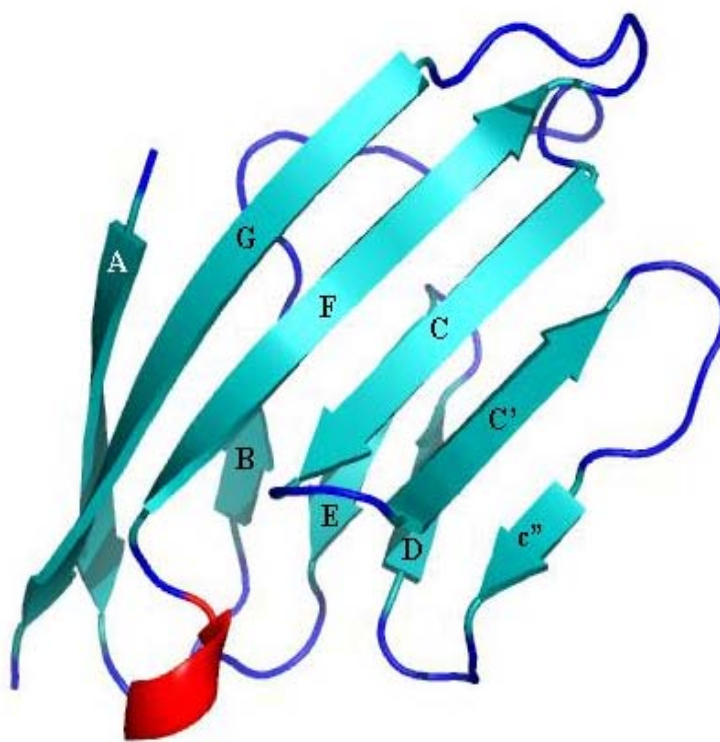


Figure 1.20. Structure of CD2. Domain 1 of the rat cell adhesion protein CD2. CD2 has 9 β -sheets (shown in cyan), one α -helix (shown in red), and several loop regions (shown in blue) (1hng (Driscoll et al., 1991)). Each β -sheet is labeled.

shows that the protein is amenable to protein engineering with over 40 single mutations being generated, which do not exhibit significant effects on the structure of the protein (Arulanandam et al., 1993a). CD2 is stable against a wide range of pH, temperature, and salt concentrations (Yang et al., 2000a). These characteristics of CD2 meet the criteria for an excellent host protein, therefore CD2 is an ideal model system to understand the effect of protein environment on calcium binding affinity and molecular recognition.

1.11 Objective and significance of this study

Long range electrostatic interactions have been shown to be important for Ca^{2+} binding affinity (Linse et al., 1988). However their role in site specific binding has not been elucidated due to the prevalence of proteins with multiple Ca^{2+} binding sites. The cooperativity between the multiple Ca^{2+} ions hampers the study of site specific binding (Linse and Forsen, 1995). In addition, many natural Ca^{2+} binding proteins undergo large conformational changes upon Ca^{2+} binding which makes it difficult to distinguish the binding energy from the energy of conformational change. A structured host protein will provide the stability for both Ca^{2+} free and Ca^{2+} bound forms which will allow for a focus on local factors including the effect of charge distribution around the Ca^{2+} coordination shell on Ca^{2+} binding affinity. The goals of this study are to examine the effect of electrostatic interactions within 5-15 Å of the Ca^{2+} coordination shell on Ca^{2+} binding affinity. This study will help gain insight into the factors that control Ca^{2+} binding affinity. An establishment of rules for Ca^{2+} binding affinity will aid in understanding the mechanisms of Ca^{2+} affinity related diseases. In addition, the structural similarities between CD2 and the Ca^{2+} dependent cell adhesion molecule cadherin facilitates the

understanding of other Ca^{2+} dependent processes including cell adhesion. The knowledge obtained is valuable for the development of therapeutic treatments of diseases caused by the abnormal functions of cell adhesion proteins.

Understanding the factors which control Ca^{2+} binding affinity will allow for the modulation of Ca^{2+} binding affinities through design. The ability to design Ca^{2+} binding proteins with varying affinities will be useful for the development of Ca^{2+} sensors. Different compartments of the cell and the extracellular matrix can be targeted based upon their levels of Ca^{2+} concentrations. It will aid in the development of designed Ca^{2+} binding proteins as contrast agents and useful reagents for diagnostic tests.

In this study, the role of electrostatic interactions in Ca^{2+} binding affinity will be approached from two different directions. First, several proteins were designed with Ca^{2+} binding sites in different regions of CD2. The first generation proteins 206, 5606, and 6775, have Ca^{2+} binding sites that are in varying electrostatic environments. However, although these proteins bind Ca^{2+} and its analogs Tb^{3+} and La^{3+} , they have a conformation that is different from wild type CD2 due to mutations of residues that were involved in the formation of the hydrophobic core of the protein. Therefore, it became necessary to design second generation proteins with refined design criteria. The second generation protein CD2.6D79 required two solvent exposed residues to be mutated to form the metal binding site and also utilizes two wild type residues as ligands. CD2.6D79 has a conformation similar to wild type CD2 and displays a strong metal binding affinity as well as metal selectivity. The electrostatic environment of this Ca^{2+} binding site was altered through site-directed mutagenesis of residues within 15 Å. Table

1.2 shows each designed protein that is studied in this research and the rationale for their design.

Chapter 2 of this dissertation lists all of the materials and methods used in this study including the molecular cloning, expression, and purification of designed proteins, and the techniques used for monitoring the conformation and metal binding of these proteins.

Chapter 3 focuses on the electrostatic calculations used to predict Ca^{2+} binding affinity and calculate pK_a values of CD2 and calbindin D9k as well as to predict the charge distribution variants that will have the greatest effect on Ca^{2+} binding affinity.

Chapter 4 focuses on the conformation and metal binding affinity of the first generations proteins 206, 5606, 6775, and N80/89.

Chapter 5 focuses on the conformational, metal binding, and molecular recognition properties of the second generation designed protein 6D79.

Chapter 6 discusses the effect of close and long range electrostatic interactions on protein stability through analysis of the charge distribution variants of 6D79.

Chapter 7 discusses the effect of charge around the coordination shell on Ca^{2+} binding affinity through the analysis of the charge distribution variants of 6D79.

Chapter 8 focuses on the pH dependent folding of the R34 charge distribution variants.

Chapter 9 discusses the significance of this work.

Designed Site	Rationale for Design	Chapter
206 5606 6775 (CD2.Ca1)	To study the effect of varying protein environments on Ca^{2+} binding affinity	4
Partial Site N80/89	To study the effect the introduced mutations have on the fold of the protein	
7E54 6D79	To design Ca^{2+} binding proteins that can be used a model systems for studying the role of electrostatic interactions	5
6D79 variants K91E K91I D94K D94A T75K T75D	To determine the role of electrostatic interactions around the coordination shell on protein stability through site-directed mutagenesis of non-liganding residues.	6
	To determine the role of electrostatic interactions around the coordination shell on Ca^{2+} binding affinity through site-directed mutagenesis of non-liganding residues.	7
6D79 variants R34E R34I	To investigate the folding properties of CD2	8

Table 1.2. The variant proteins of CD2 engineered for the research study.

Chapter 2: Material and Methods

2.1 Rational design of calcium binding proteins

The design of calcium binding proteins was accomplished using the computer algorithm Dezymer on an SGI O2 computer. A geometric description of the ligands around a Ca^{2+} ion, and a library of side chain rotamers of amino acids were input to identify potential Ca^{2+} binding sites (Yang et al., 2003; Yang et al., 2002b). The parameters used for bond lengths (Ca-O bond lengths, ideal value is 2.4 Å) and angles (O-Ca-O bond angles ideal values are 72°, 144°, and 90°, respectively) were based on a pentagonal bipyramidal geometry, which is the most common geometry of natural Ca^{2+} binding sites (Pidcock and Moore, 2001). In the design process, a calcium atom is attached to the side chain of an anchor glutamate residue with distance of 2.4 Å to both oxygen atoms. This pseudo-residue is floated first along the backbone of the host protein and four other residues around it are mutated to Asp, Asn and/or Glu, which are common Ca^{2+} binding ligands, to look for the combination that satisfies the geometric requirement (Yang et al., 2003; Yang et al., 2002b).

The proteins were designed based on several criteria. First, the side chain of the introduced residues should not clash with the pre-existing atoms. Second, the disruption of the native hydrogen bond network and van der Waal interactions should be minimized to ensure proper folding of the protein. Third, the designed proteins should have a high number of negative charges to ensure strong binding. Most natural Ca^{2+} binding sites have four negative charges.

2.2 Protein engineering

The second generation sites were engineered using site-directed mutagenesis. The first step in creating these mutants was designing primers for the polymerase chain reaction (PCR). All PCR primers were generated based on several criteria including number of base pairs and T_m values. The optimal characteristics of primers for PCR are that they consist of greater than 21 base pairs and preferably had a T_m of 66°C or greater. Table 2.1 lists all of the primers designed for the engineering of each protein used in this study as well as their corresponding T_m 's. All primers were synthesized by Sigma Genosys (Pittsburgh, PA). First the primers were phosphorylated using T4 polynucleotide kinase (PNK) (New England BioLabs). The phosphorylation reaction consisted of 6 μ L of primer, 2 μ L of T4 PNK buffer, 1 μ L of 100 mM ATP, 1.5 μ L of T4 PNK, and 10.5 μ L of H₂O. The reaction was incubated at room temperature for 30 min and then incubated at 65° C for 15 min.

In this work *Pfu* DNA polymerase (Promega) was used to generate the required mutations. *Pfu* is an excellent polymerase because of its proofreading properties. The amount of primer required for the PCR reaction was calculated using equation 2.1.

$$primervolume = \frac{(\#ofbasepairs / 21)500}{primerconcentration(ug)} \quad (\text{Eq. 2.1})$$

The PCR reaction consisted of 10 μ L *Pfu* buffer, 10 μ L 3 mM dNTP mix, 1 μ L template DNA, 1 μ L *Pfu* polymerase, and H₂O to bring the total volume to 100 μ L.

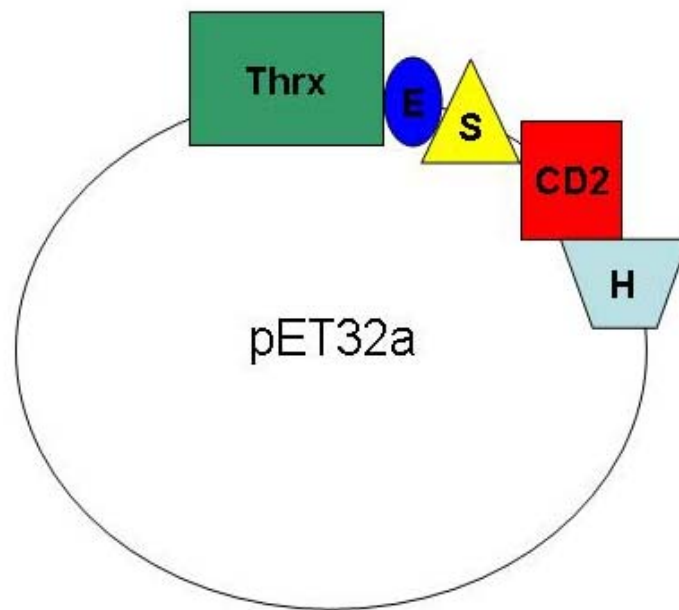
Protein	Primers	T _m (°C)
R34E	F – CGA TGG GAG GAG GGG AGC ACC CT	68
	R – CAC CTC ATC AAT ATC ATC AGT CAT	66
R34I	F - CGA TGG GAG ATT GGG AGC ACC CT	68
	R - CAC CTC ATC AAT ATC ATC AGT CAT	66
R34E-D72K	F - CTG ACA AGA GAT AAG AGT GGC ACC T	66
	R - ATT CTT TAT CTT CAA GTC TCC ATT TG	68
R34I-D72K	F - CTG ACA AGA GAT AAG AGT GGC ACC T	66
	R - ATT CTT TAT CTT CAA GTC TCC ATT TG	68
T75K	F – GAC AGT GGC AAA TAT AAT GTA GAC G	64
	R – ATC TCT TGT CAG ATT CTT TAT CTT C	66
T75D	F- GAC AGT GGC GAC TAT AAT GTA GAC G	64
	R – ATC TCT TGT CAG ATT CTT TAT CTT C	66
K91E	F – ATC CTG AAC GAG GAA CTG GAC TTG AG	68
	R – ACG TGT CCC ATT TGT GCT GTA TAC	70
K91I	F – ATC CTG AAC ATC GAA CTG GAC TTG AG	68
	R – ACG TGT CCC ATT TGT GCT GTA TAC	70
D94K	F – C AAG GAA CTG AAG TTG AGG ATT CTA G	66
	R – TT CAG GAT ACG TGT CCC ATT TGT	66
D94A	F- C AAG GAA CTG GCC TTG AGG ATT CTA G	66
	R - TT CAG GAT ACG TGT CCC ATT TGT	66

Table 2.1. Primer table. Forward and reverse primers used for the engineering of the charge distribution variants of 6D79 along with their corresponding T_m values.

All PCR experiments were done in the pGEX-2T vector. After PCR, the DNA was run on a 0.8% agarose gel. The relevant band (based on molecular weight) as identified by UV light was excised from the gel and ligated overnight with T4 DNA ligase (New England BioLabs) at 4 °C. The ligation reactions consisted of 25 µL of DNA, 1 µL of ligase, 3 µL of ligase buffer, and 1 µL H₂O. The reaction was then transformed into DH5α competent cells (Stratagene) and then grown in 10 mL of Luria-Bertani (LB) broth. The DNA was purified by a DNA mini prep kit (Qiagen) and sequenced (Biology Core Facility) to determine if the PCR was successful.

2.3 Sub-cloning of first generation proteins into the pET-32a vector

Sub-cloning of the first generation proteins into pET32a (Novagen) (Figure 2.1) was carried out using BamHI and EcoRI (New England BioLabs) restriction enzymes. Both the pGEX-2T and the pET32a vectors have these restriction sites. Both vectors were digested with the following reactions: 1 µL BamHI buffer, 1 µL BamHI, 1 µL EcoRI buffer, 1 µL EcoRI buffer, and 6 µL DNA. The reactions were incubated at 37°C for 2 hr. A 0.8% agarose gel was run and the relevant bands (DNA that encoded the designed protein for the pGEX-2T reaction and digest pET32a vector) were excised. The DNA was purified via a gel extraction kit (Qiagen). The DNA of the designed protein was then ligated into the pET32a vector using T4 DNA ligase (New England BioLabs). The reactions contained the following: 10 µL pET32a vector DNA, 10 µL designed protein encoding DNA, 1.5 µL T4 DNA ligase, 1.5 µL ligase buffer, and 1 µL of 100 mM ATP. The reactions were incubated at 4°C for 18 hr. The DNA was transformed



Thrx – thioredoxin sequence
E – enterokinase cleavage sequence
S – S-tag sequence
CD2 – CD2 sequence
H – His tag sequence

Figure 2.1. The pET32a vector. The vector plasmid pET32a utilizes thioredoxin as a fusion partner for expression with an enterokinase cleavage site. In addition the vector has a N-terminal S-tag and a C-terminal His-tag.

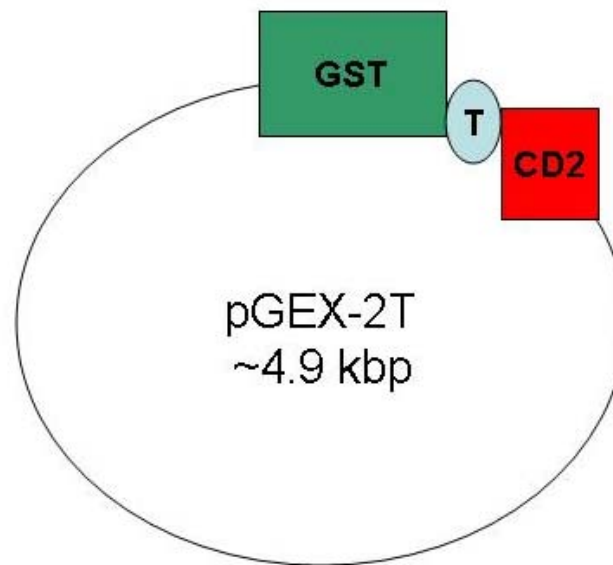
into DH5 α competent cells and then grown in 10 mL of Luria-Bertini (LB) broth. The DNA was purified by a mini prep kit (Qiagen) and sequenced (Biology Core Facility).

2.4 Protein expression

2.4.1 Expression in the pGEX-2T vector

The pGEX-2T vector (Figure 2.2) was a gift from Dr. Zhi-ren Liu at Georgia State University. Expression in the pGEX-2T vector was carried out using Luria-Bertini (LB) broth. The DNA was transformed into BL21(DE3) competent cells and plated onto LB plates containing 100 mg/ml ampicillin. One colony was used to inoculate 500 mL of LB media with ampicillin. The 500 mL culture was grown overnight at 37 °C with agitation. Four 1 L flasks of LB media with ampicillin were inoculated with 100 mL of the overnight culture each. The 1 L cultures were grown at 37 °C with agitation until the optical density at 600 nm (OD₆₀₀) reached 0.8. The culture was then induced with 0.1 mM isopropyl-beta-D-thiogalactopyranoside (IPTG). The culture was induced for 3 hr and then centrifuged at 7K rpm for 20 min. The cell pellet was collected and frozen.

For heterogeneous labeling of protein SV minimal media was used for expression. The DNA was transformed into BL21(DE3) competent cells and plated onto LB plates with 100 mg/ml ampicillin and 0.3% glucose. The plates were incubated at 37 °C for 16 hr. Several 2 mL samples of SV media containing 100 mg/ml ampicillin, 0.5% glucose, and 0.05% ¹⁵NH₄Cl were inoculated with single colonies. The samples were grown with agitation at 37 °C for 8 hr. The four 2 mL samples with the best growth (based on turbidity) were then used to inoculate four 100 mL of SV media containing the same concentrations of ampicillin, glucose and ¹⁵NH₄Cl. These samples were grown overnight



GST – Glutathione S-transferase sequence

T – thrombin cleavage site

CD2 – CD2 sequence

Thrombin cleavage site

Leu Val Pro Arg ↓ Gly Ser
CTG GTT CCG CGT GGA TCC

Figure 2.2. The pGEX-2T vector. The vector plasmid pGEX-2T utilizes GST as a fusion partner for expression with a thrombin cleavage site.

at 37 °C with agitation. The overnight culture was used to inoculate a four 1 L culture of SV containing ampicillin, glucose, and $^{15}\text{NH}_4\text{Cl}$. The 1 L cultures were grown at 37 °C with agitation until the $\text{OD}_{600} = 0.8$. The culture was then induced with 0.1 mM IPTG and expressed for 3 hr at 37 °C with agitation. The cultures were centrifuged at 7K rpm for 20 min and the cell pellets were collected and frozen.

2.4.2 Expression in the pET32a vector

The pET32a vector was a gift from Dr. Zhi-ren Liu at Georgia State University. For expression in the pET32a vector the DNA was transformed into BL21(DE3) competent cells and plated onto LB ampicillin plates. Two 10 mL culture of LB with 100 mg/mL of ampicillin was inoculated with a single colony from the plate and grown overnight at 37 °C with agitation. Two 1 L flasks of LB with ampicillin were inoculated with a 10 mL culture. The culture was grown at 37 °C with agitation until the OD_{600} reached 0.6. The culture was then induced with 0.2 mM IPTG and grown for 3 hr. The culture was centrifuged at 7K rpm for 20 min.

2.5 Protein purification

2.5.1 Purification of pGEX-2T expressed proteins

The cell pellet was re-suspended in lysate buffer (PBS buffer containing N-lauryl sarcosine, DTT, and EGTA). The protein was blended 3x for 20 s each. The protein was then sonicated 6x for 20 s. The protein was centrifuged at 17K rpm for 30 min. The supernatant was syringe filtered and loaded onto a column containing glutathione sepharose 4 beads through sedimentation. The column was washed with PBS buffer to remove any excess proteins from the cells. Thrombin (1 unit/ μL) (Amersham Pharmacia)

was added to each column and the fusion protein was cleaved at 4 °C overnight. The designed protein was eluted from the columns and run on a Sephadex 75 gel filtration column. Further purification was done with an ion exchange column. An SP column (Amersham Pharmacia), a cation exchange column, was used for purification. The pH of the protein was lowered through dilution with pH 2 acetic acid. The column was equilibrated in 20 mM sodium acetate buffer pH 4.5. The protein bound at low pH and then a gradient was run with 20 mM Tris buffer with a pH of 8. The protein eluted around pH 7.4. After purification the purity of our protein was verified by SDS-PAGE and mass spectrometry analyses.

2.5.2 Purification of the pET32a expressed proteins

Proteins expressed in the pET32a vector were purified 1 L at a time. The protein was re-suspended in extraction buffer (20 mM Tris, 100 mM NaCl, 0.1% Triton X-100). The protein was sonicated 6x for 20 s each and centrifuged at 17K rpm for 20 min. The protein was bound to a nickel chelating column in a phosphate buffer with 250 mM NaCl. The protein was eluted with an imidazole gradient. The protein was dialyzed to remove the imidazole and further purified using an ion exchange column.

2.6 Far UV CD

The far-UV CD spectra of the designed proteins were obtained from 190 to 260 nm with a bandwidth of 0.5 nm at room temperature using a Jasco 810 spectropolarimeter. A rectangle cell with light path of 1 cm was used. The protein was diluted to 3 μ M in 10 mM Tris pH 7.4. The protein stock was separated and diluted with

metals at the same ratio to final concentrations of 1 mM EGTA, 10 mM Ca^{2+} , 100 μM Tb^{3+} , and 100 μM La^{3+} .

For metal titrations a 4.2 ml stock of 3 μM protein was made up in 10 mM Tris pH 7.4. From this stock 2.25 mL of protein was diluted with 250 μL of H_2O . Also from this stock 1.35 ml of protein was diluted with 150 μL of 10 mM Tb^{3+} , 10 mM La^{3+} or 100 mM Ca^{2+} . The metal titrations were carried out by gradually adding metal to the protein samples. To increase the sensitivity, 10 min collection of 1000 points at 205 nm was averaged at each metal concentration. This time averaging method gives the resolution up to 0.01 mdeg. To obtain the metal binding affinity, the fractional change of the CD signal at 205 was fit using equation 2.2:

$$F = \frac{([P]_T + [M]_T + K_d) - \sqrt{([P]_T + M_T + K_d)^2 - 4([P]_T * M_T)}}{2[P]_T} \quad (\text{Eq. 2.2})$$

Where F is the fractional change, P is the total protein concentration, and M is the total metal concentration.

For the thermal denaturation studies a rectangle cell with light path of 1 mm was used. For each protein a 20 μM protein sample was prepared in 10 mM Tris, pH 7.4. Excess Ca^{2+} or excess EGTA was added to the protein to final concentrations of 10 and 1 mM, respectively, with the same dilution factor. The samples were equilibrated for 20 min at room temperature. The sample was recorded at 222 nm from 10 to 90°C with a 5 min equilibration time at each temperature.

2.7 Fluorescence

2.7.1 Trp fluorescence

All Trp fluorescence experiments were performed on a PTI fluorometer with excitation and emission slit widths of 2 nm. The emission spectra were collected from 300 – 400 nm with the excitation wavelength at 282 nm. A 3 μ M stock of protein was made up in 10 mM Tris, pH 7.4. The protein stock was separated and diluted with metals at the same ratio to final concentrations of 1 mM EGTA, 10 mM Ca^{2+} , 100 μ M Tb^{3+} , and 100 μ M La^{3+} .

2.7.2 FRET experiments

All Tyr/Trp-sensitized fluorescent resonance transfer experiments were performed on a PTI fluorometer with an excitation slit width of 6 nm and an emission slit width of 8 nm. The emission spectra were collected from 500 to 600 nm with the excitation wavelength at 282 nm. A 3 μ M stock of protein was made up in 20 mM PIPES, 10 mM KCl pH 6.8 (10 mM Tris, pH 7.4 for first generation proteins). From this stock 2.25 ml was diluted 250 μ L of H_2O for the protein sample. Also from this stock 1.35 ml was diluted with 150 μ L of 10 mM Tb^{3+} and 100 mM Tb^{3+} to make two metal stocks of 1 mM and 10 mM Tb^{3+} , respectively. Three 800 μ L protein samples were used for the titration. As Tb^{3+} was added in increments to each of the samples 10 min equilibrium times were allowed for each point. The terbium binding affinities were obtained directly by a terbium titration. First the data was normalized to decrease the baseline slope using an exponential equation. Then the blank Tb^{3+} intensity was subtracted. The dissociation

constant for Tb^{3+} was determined using the fractional change of the fluorescence intensity at 545 nm using equation 2.2 and assuming 1:1 binding.

For the competition assay of other metal ions a 4 ml protein- Tb^{3+} solution was made up with concentrations of 3 and 30 μM , for protein and Tb^{3+} , respectively. This sample was equilibrated for 15 min to allow for Tb^{3+} binding. The metal of interest was added to each protein-Tb sample to final concentrations of 100 μM La^{3+} , Mn^{2+} , Ni^{2+} , and Co^{2+} and 10 mM Ca^{2+} , Mg^{2+} , and K^{+} . Metal concentrations of 100 μM were used to minimize protein precipitation. The solutions were equilibrated for 20 min. All binding affinities determined using this method were fit using Spec-fit (Binstead, 1998) with a 2 ligand model taking into account the binding affinity of Tb^{3+} .

2.7.3 Rhodamine-5N (Rhod-5N) experiments

The Molecular Probes dye Rhod-5N binds Ca^{2+} with a rather weak affinity. In order to obtain the Ca^{2+} binding affinity of 6D79 the Rhod-5N was used in a competition assay measured by fluorescence. The buffer used was 20 mM PIPES, 20 mM KCl pH 6.8. The protein concentration was 20 μM and the dye concentration was 50 μM . The Ca^{2+} titration was performed in concurrence with an EGTA titration in order to obtain the concentration of Ca^{2+} in our sample. This was done in response to Ca^{2+} contamination of our sample which alters the binding affinity that we obtain. Using this method we found that the initial Ca^{2+} contamination of my samples were approximately 8 μM . The data was fit using Spec-fit.

2.8 Nuclear magnetic resonance spectroscopy (NMR)

2.8.1 1D NMR

One-dimensional ^1H NMR spectra were recorded on a Varian Unity plus 500 MHz NMR spectrometer with a spectral width of 6600 Hz and complex data points of 17K using a modified WATERGATE pulse sequence at 25 °C. The volume of all of the freshly prepared samples was 500 μL with concentrations greater than 100 μM in 10% D_2O . The buffer used for all NMR experiments was 20 mM PIPES, 10 mM KCl pH 6.8. Samples were dialyzed against 1 L of this buffer 3x for 1 hr each. The data was processed with the program FELIX98 (MSI). After Fourier transformation, a 90° squared-sinebell window function was applied. For samples with higher noise a window function of 1 or 2 Hz exponential was used. Samples were reference according to the dioxane peak with a chemical shift of 3.743 ppm. For Ca^{2+} titrations an initial point contained 25 μM EGTA to ensure a Ca^{2+} free sample. Increasing additions of Ca^{2+} , based on the ration of metal to protein concentration, were added from stocks of 0.01, 0.05, 0.1, 0.5, and 1 M Ca^{2+} . Large concentrations of Ca^{2+} stocks were used in order to minimize volume increase so that changes in protein concentration would be minimal. The changes in chemical shifts of particular residues were measured by picking the middle of the peaks. The fractional change of these chemical shift changes were fit to equation 2.2 in order to obtain a K_d .

2.8.2 2D NMR

All 2D NMR experiments were recorded on a Varian Unity plus 600 MHz NMR spectrometer with a spectral width of 8000 Hz. The samples were in 20 mM PIPES 10 mM KCl, pH 6.8, with either EGTA (5 mM) or Ca^{2+} (10 mM). The protein concentrations were 1 mM for the homonuclear 2D total correlation spectroscopy (TOCSY) and nuclear Overhauser spectroscopy (NOESY) experiments and 250 μM for ^{15}N HSQC experiments in 10% D_2O . TOCSY spectra were recorded with mixing times of 36, 60 and 90 ms. NOESY spectra were recorded with mixing times of 100 and 150 ms. The assignment of NMR spectra was achieved by standard methods and procedures (Yang et al., 2005). The data was first processed using Felix(MSI) with a 90° sinebell window function. The spectra was referenced using spectral widths of 8000 in both the x and y directions. The spectra was then assigned using the program Sparky. The resonances of wild type CD2 were reassigned at a pH of 7.4. For Ca^{2+} titrations, increasing additions of Ca^{2+} , based on the ratio of metal to protein concentration (for example 0.5:1, Ca^{2+} :protein concentration), was added from stocks of 0.01, 0.05, 0.1, 0.5, and 1 M Ca^{2+} . Assignments of TOCSY, NOESY, and HSQC experiments are detailed in the Appendix.

Paramagnetic relaxation NMR experiments have been used in accurately determining the location of Ca^{2+} binding sites (Ubach et al., 1998). The strong relaxation effects of the paramagnetic Ca^{2+} analogue Mn^{2+} can be used to determine where the Ca^{2+} binding site is located. The relaxation rate of a protein amide proton induced by a bound

Mn^{2+} ion is proportional to r^{-6} (where r is the distance between the amide proton and the Mn^{2+} ion). The relationship between the two variable can be fit with equation 2.3.

$$\text{inverse of } \Delta \equiv \frac{\text{inverse manganese bound resonance intensities}}{\text{inverse manganese free resonance intensities}} \quad (\text{Eq. 2.3})$$

These experiments were done using a 150 μM ^{15}N labeled sample. An ^{15}N -HSQC pulse sequence was used for the titration. For Mn^{2+} titrations, the protein was first incubated with 10 mM Ca^{2+} in order to reduce non-specific binding of the Mn^{2+} ions which would lead to over relaxation in the spectra. Then increasing amounts of Mn^{2+} was added to final concentrations of 10, 20, and 30 μM . The change in intensities of all of the peaks was measured using Sparky and fit to equation 2.3.

2.9 Surface plasmon resonance

All surface plasmon resonance (SPR) experiments were done by Alice Kearney at Oxford University in the lab of Dr. P. Anton van der Merwe. SPR studies were performed at 25 °C and a flow rate of 10 $\mu\text{l}/\text{min}$ using a Biacore 2000 (Biacore AB, Stevenage, Herts., UK) in HBS buffer (10 mM HEPES, 150 mM NaCl, pH 7.4) containing either 1 mM EGTA or 1 mM CaCl_2 . The anti-GST antibody was coupled to a CM5 sensor chip (Biacore) using the Amine Coupling Kit (Biacore) as directed, with the following modifications: after an activation step of 300 s, anti-GST antibody (30 $\mu\text{g}/\text{ml}$ in sodium acetate pH 5.0) was injected for 300 s. Immobilisation levels were approximately 17000 Response Units (RUs). The anti-GST antibody was regenerated with a 180 s injection of 10 mM glycine-HCl pH 2.0 (Biacore). The CD2-GST fusion proteins were purified as detailed in section 2.5.1. However, the fusion proteins were not cleaved using

thrombin. The proteins were eluted from the GS4B beads with 20 mM Glutathione pH 11. The samples were then dialyzed against 10 mM Tris pH 7.4 3x. Each protein was 1/10 in HBS buffer and injected for 180 s over immobilised anti-GST antibody. Soluble rat CD48 was diluted to 70 μ M and seven 2-fold dilutions thereof were injected for 30 s. Affinity constants were determined by non-linear fitting of the Langmuir binding isotherm to the data.

2.10 ICP-MS experiments

All ICP-MS experiments were run on a Finnigan MAT Element2 double focusing instrument with reverse Nier Johnson geometry by Dr. Russell Malchow. Protein was dialyzed against 1 L of 1 mM Tris pH 7.0 with 5 μ M of metal at 4 °C for 36 hr with 3 changes of buffer. The Ca^{2+} contents of dialyzing buffer and protein solution were examined after acidification with 1% HNO_3 by monitoring the content of ^{42}Ca , ^{43}Ca , and ^{44}Ca . A concentration of 1 ppb of ^{115}In was added as an internal standard in all samples and standard solutions to compensate for sample-to-sample variations in instrument performance (Yang et al., 2003). Calibration curves were established for each isotope and the absolute concentrations of each isotope was determined. The dissociation constant was determined from equation 2.3.

$$K_d = \frac{[Ca]_{free}[P]_{free}}{[Ca]_{bound}} \text{ (Eq. 2.3)}$$

2.11 Electrostatic potential calculations and visualizations

All electrostatic potentials were calculated using the program DelPhi (Gilson and Honig, 1988a; Honig and Nicholls, 1995; Nicholls A and Honig, 1991; Rocchia et al., 2001; Yang et al., 1993). Protein coordinates from the solution structure of CD2, 1hng (Jones et al., 1992), were used for the calculations. A new pdb file for site 6D79 was generated by replacing the wild type CD2 with the designed Ca^{2+} binding ligands using the computer algorithm Dezymer. The file was modified in Sybyl to generate coordinate files for all of the charge distribution variants. The hydrogen atoms were built in and all of the structures were minimized with an Amber force field using Sybyl. For the minimization, a dielectric constant of 4 was used. A maximum energy change of 0.001 kcal/mol was imposed and 1000 iterations were used. For the DelPhi calculations, interior and exterior dielectric constants of 4 and 80, respectively, were utilized. A grid size of 165 was used. The salt concentration was set to 0 and the linear solution of the Poisson Boltzmann equation was imposed until convergence was reached. All run files for DelPhi are in the Appendix. The pdb files of CD2.6D79 variants were generated using Sybyl. The individual amino acids were mutated and the protein was minimized using the Amber force field.

To visualize the electrostatic surface, the Graphical Representation and Analysis of Structural Properties (GRASP) program was used (Nicholls et al., 1991). The molecular surfaces of CD2 and the designed proteins were built from the coordinate files. The potential map calculated by DelPhi was uploaded and the electrostatic surface was

visualized. Potentials less than $-7kT$ are red, those greater than $7kT$ are blue, and neutral potentials are white.

Chapter 3. Analysis of Electrostatic Interactions on CD2 and its Variants

3.1 Current methods for studying electrostatic interactions

As discussed in **section 1.6 and 1.8**, electrostatic interactions are essential for the folding, structure, stability, and biological functions of proteins. One of the key factors involved in Ca^{2+} binding affinity is the role of electrostatic interactions around the coordination shell. A net negative charge within 5 – 15 Å of the Ca^{2+} coordination shell has been shown to be important for strong Ca^{2+} binding affinity (Linse et al., 1988; Linse and Forsen, 1995) (Figure 1.18). Therefore, Ca^{2+} binding affinity could be fine-tuned by modifying aspects of the protein's electrostatic environment. By introducing or removing charged residues around Ca^{2+} binding sites to induce different electrostatic arrangements, we can further our understanding of Ca^{2+} binding affinity. Moreover, the ability to estimate electrostatic contribution to calcium affinity through calculation will provide guidance for the design of Ca^{2+} sites in proteins with a desired binding affinity, which could be useful for the design of Ca^{2+} sensors. In this chapter, first different computational methods for estimating electrostatic contributions will be discussed. Second, the “hot spots” of CD2 will be identified. Lastly, the effect of electrostatic interactions on Ca^{2+} binding sites will be estimated.

Extensive studies have been carried out using computational modeling of electrostatic interactions for studying protein properties and in protein design. At the most basic level, Coulomb's law for a pair of charges describes the energy of all electrostatic interactions (equation 3.1).

$$V_e = Z_1 Z_2 e^2 / \epsilon r \quad (\text{Equation 3.1})$$

The electrostatic potential is inversely proportional to the dielectric constant, ϵ , of the medium and the distance, r , separating the two charged species with charged numbers Z_1 and Z_2 (van Holde, 1998).

For proteins, the knowledge of the solute, the solvent, and the interactions between them is required because values of charges and dielectric constants of any atom in the protein vary with the environment. Coulomb's law is not descriptive enough for the study of complex macromolecular systems. Since it is not possible to include these interactions in detail, models needed to be developed that represent these effects as an average or continuum. Several continuum models have been developed to understand electrostatic interactions including finite difference methods that solve the Poisson-Boltzmann equation, Coulombic methods, and Debye-Hückel methods. The validity of these models for calculating electrostatic energies in proteins is usually established by comparison of measured and calculated pK_a values.

3.1.1 The dielectric constant

The energies associated with long range interactions are dependent on the intervening medium (van Holde, 1998). For example, the interaction of two charged atoms becomes shielded in a polar medium. The polarizability of a medium is defined as its dielectric constant (ϵ) relative to that of a vacuum, which is the least polarizable medium. The expressions for the energy of long range interactions are all inversely related to the dielectric of the medium and are therefore weakened in a highly polarizable medium such as water (van Holde, 1998). The dielectric constant of the environment plays a central role in stabilizing the conformation of a macromolecule.

Coulomb's law only takes into account one dielectric constant, that of water. However, since water is more easily polarized by an electric field than are most solutes, at least two dielectric constants are required to capture the underlying physics of polar molecules in solution (Honig and Nicholls, 1995). The two dielectric constants correspond to the molecular surface, defined as the contact surface formed between the van der Waals envelope of the molecule and a probe solvent molecule (Richards, 1977). Previous studies on the calculations of folded proteins yield dielectric constants in the range of 2-5 (Gilson and Honig, 1986). Therefore, in most studies, all regions inside the surface are assigned a low dielectric constant value of 2 to 4 (Harvey, 1989) and the exterior regions are assigned the dielectric constant of water (80).

3.1.2 The Poisson-Boltzmann equation

The classical treatment of electrostatic interactions in solution is based on the Poisson-Boltzmann (PB) equation, which takes into account the effect of mobile ions and a non-uniform dielectric constant (equation 3.2).

$$0 = \nabla \cdot [\epsilon(r) \nabla \phi(r)] - \epsilon(r) \kappa(r)^2 \sin[\phi(r)] + 4\pi \rho^f(r) / kT \quad (\text{Eq. 3.2})$$

where $\phi(r)$ is the dimensionless electrostatic potential in units of kT/q (k is the Boltzmann constant, T is the absolute temperature, and q is the charge on a proton), ϵ is the dielectric constant, ρ^f is the fixed charge density (in proton charge units), and the term $\kappa(r)^2 = 1/\lambda^2 = 8\pi q^2 I / ekT$, where λ is the Debye length and I is the ionic strength of the bulk solution. The variables ϕ , ϵ , κ , and ρ are all functions of the position vector r . The PB equation can be reduced to the simpler but less descriptive Coulomb's law when ϵ is uniform in space.

There are a number of numerical methods that can be used to solve the PB equation, nearly all of which make some type of simplifying assumptions concerning the definition of r (van Holde, 1998). For example, the Born model for transferring an ion from a medium of dielectric D_1 to another of dielectric D_2 uses a cavity radius of r to define the volume of solvent excluded by the ion (van Holde, 1998). One method that requires no assumption for r is the finite difference solution to the PB equation (van Holde, 1998). In this method, the macromolecule and any solvent is mapped onto a three-dimensional grid. Each grid point is a well-defined positional vector in space and

can be assigned a partial charge, ionic strength, and dielectric constant. The PB equation is numerically evaluated for each grid point. An initial guess of the solution is then refined successively through an iterative procedure.

The computer algorithm DelPhi employs finite difference methods to solve the non-linear and linear forms of the PB equation (equation 3.2) (Nicholls A and Honig, 1991; Rocchia et al., 2001). The charges and dielectric properties of the system are mapped onto a lattice with a defined grid size and the dielectric properties are separated by the interior and exterior dielectric constants of a molecule taking into account the polar molecules in aqueous solution.

3.1.2.2 Increasing the precision of the PB equation

Recently, there have been attempts to increase the precision of the PB equation. The introduction of an electrostatic model that more accurately represents the complexity of electrostatic interactions in a heterogeneous protein environment by using multiple (up to 24) parameters that take into account variations in protein geometry, local structure, and the type of interacting residues has been established (Wisiz et al., 1998). These parameters can be divided into three classes, contributions from dipolar solvent effects, relaxation effects in the protein interior, and non-covalent interactions. Within each class, the parameters are further subdivided based on the region in which the groups are located (surface, boundary, or core) and interaction type (charge-charge or charge-polar). Figure 3.1 shows the classification as either core, boundary, or surface residues based on the depth below the surface and solvent-accessible surface area. The classification was used to capture difference in solvent and interior relaxation effects. All residues were

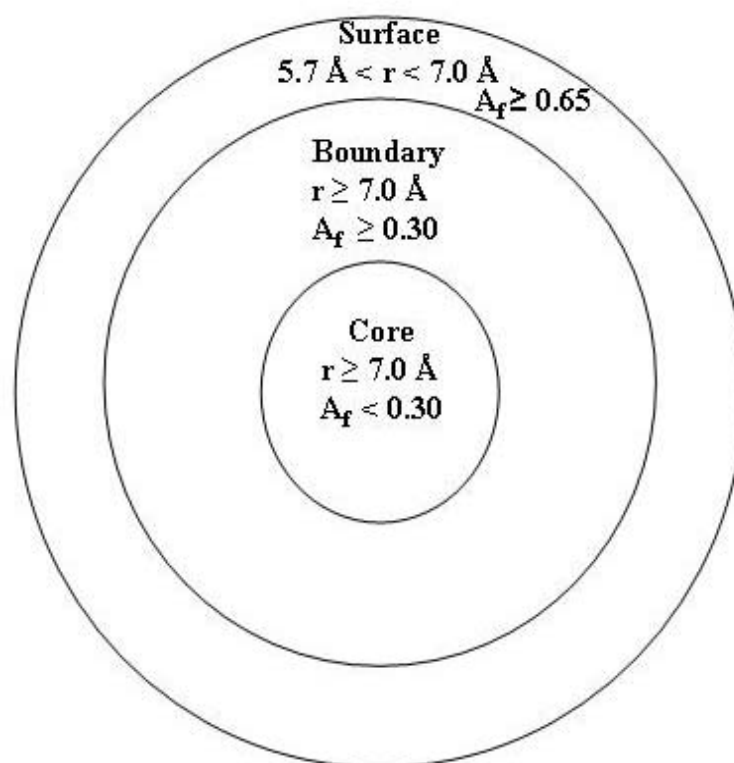


Figure 3.1. Boundary conditions. Definition of the surface, boundary, and core dielectric regions used to enhance the precision of the Poisson-Boltzmann equation where r is the distance of their C_β atom to a pseudo-accessible surface area and A_f is the fractional solvent accessible surface area of atoms involved in the ionization process (Wisz et al., 1998).

classified according to the distance of their C_β atom to a pseudo solvent-accessible surface (r), calculated only in the presence of backbone and C_β atoms, and the fractional solvent-accessible surface area (Af) of atoms involved in the ionization process (Wisiz and Hellinga, 2003). These three regions allows for more descriptive and precise calculations than the use of two dielectric constants for the PB equation. The 24 unknown variables are solved iteratively using continuum methods. This method is time consuming but the empirically determined values of the parameters correlate well with experimentally observed contributions of ion pairs compared with other methods.

3.1.3 Debye-Hückel theory

A second method used to understand electrostatic interactions is the Debye-Hückel method, which is based on a very simplified approach but has been shown to be a highly accurate method for predicting pK_a values using equation 3.3 (Lee et al., 2002).

$$1.36\Delta pK_a = -\Delta G_{ij} = -332z_i\Delta\psi_j = \frac{332z_i\Delta z_j}{\epsilon_{eff}} \left(\frac{e - \kappa r_{ij}}{r_{ij}} \right) \quad \text{Eq. 3.3}$$

ΔG_{ij} is the interaction energy, z_i is the charge at site I, ψ_j is the electrostatic potential of site I caused by charge j , z_j is the charge of site j , r is the distance between site i and j , and κ is the Debye-Hückel screening length. This method demonstrates that complex interactions, including the deprotonation of a charge residue on a protein surface upon association, can be approximated using simple Debye-Hückel potential arguments. Although this method does not take into account interior and exterior dielectric constants but rather assumes that all amino acid residues are equally exposed to solvent, it has been successful in providing data of pK_a values that are comparable to experimental data.

3.1.4 Monte Carlo simulations

A third method used to calculate electrostatic interactions is Monte Carlo simulations (MC). Monte Carlo methods are based on the use of random numbers and probability statistics to investigate problems (van Holde, 1998). With MC methods, a large system can be sampled in a number of random configurations, and that data can be used to describe the system as a whole using an equation of choice to describe the electrostatic energy of a system. The advantage of using Monte Carlo simulations instead of solving the Debye–Huckel or Poisson–Boltzmann equation is that the approximations inherent in the two latter approaches are avoided. The computational efforts are roughly the same and for weakly charged systems the different approaches give identical results (Fushiki et al., 1991). The drawback of the MC approach is that the inclusion of a protein with a low dielectric interior, different from that of the solution, significantly increases the computation time (Da Silva et al., 2001).

3.2.1 Calculating electrostatic potentials using DelPhi

Through the use of rational design we have designed several single Ca^{2+} binding sites into a host protein, the rat cell adhesion molecule CD2. To further examine the electrostatic environment of these sites the electrostatic “hot spots” of CD2 were examined using the computer algorithms DelPhi (version 4.0) and GRASP. As discussed in **section 2.11**, a pdb file downloaded from the protein data bank is required to have hydrogens attached and minimized using Sybyl. The first step in this process is calculating a potential map of the entire protein using DelPhi, which utilizes finite difference methods to solve the Poisson-Boltzmann equation (equation 3.2). The charge

distribution of the molecule is represented by point charges located at atomic nuclei in the form of a charge file that can be as descriptive as possible based on the calculation. For example, a charge file can contain only the charges on the four charged amino acids (Figure 3.2) or it can be very descriptive and contain charges and partial charges on all atoms in the system (Figure 3.3).

DelPhi can calculate several different energies including the grid energy, solvation energy, and coulombic energy. The grid energy is obtained from the product of the potential at each point on the grid and the charge at that point, summed over all points on the grid. The grid energy is only useful when comparing two DelPhi runs with exactly the same grid conditions. The difference can be used to extract solvation energies, binding energies, and salt effects.

The coulombic energy, generated by other fixed charges at the j -th charge position, is calculated using Equation 3.4

$$\varphi_{coul}(r_j) = \sum_{i \neq j} \frac{q_i}{4\pi\epsilon_0\epsilon_l |r_j - r_i|} \quad \text{Eq. 3.4}$$

It is defined as the energy required to bring charges from infinite distance to their resting position within the dielectric specified for the molecule. The solvation energy is obtained from the product of the potential due to induced surface charges with all fixed charges of the solute molecule using Equation 3.5

$$\varphi_{solv}(r_j) = \sum_p \frac{\delta_p}{4\pi\epsilon_0 |r_j - r_p|} \quad \text{Eq. 3.5}$$

where p goes over the locations of polarization surface charges and δ_p is the polarization charge at position r_p on the surface. The solvation energy is the energy of transferring the

atom	resnumbc	charge
nz	lys	1.0
nh1	arg	0.5
nh2	arg	0.5
oe1	glu	-0.5
oe2	glu	-0.5
od1	asp	-0.5
od2	asp	-0.5
oxt		1.00
ot		-1.00
cal		2.00

Figure 3.2. DelPhi charge file. Example of a charge file used for a DelPhi run which only contains the charges for the charged atoms of the amino acids Lys, Arg, Glu, and Asp as well as the terminal charges and the charge of calcium (cal).

atom_	resnumbc_	charge_	
N	ALA	-0.400	
CA	ALA	0.000	
HN	ALA	0.400	
HA	ALA	0.000	
C	ALA	0.550	
O	ALA	-0.550	
CB	ALA	0.000	
HB1	ALA	0.000	
HB2	ALA	0.000	
HB3	ALA	0.000	
N	ARG	-0.400	! charged arginine
CA	ARG	-0.000	
HN	ARG	0.400	
HA	ARG	0.000	
C	ARG	0.550	
O	ARG	-0.550	
CB	ARG	0.000	
HB1	ARG	0.000	
HB2	ARG	0.000	
CG	ARG	0.000	
HG1	ARG	0.000	
HG2	ARG	0.000	
CD	ARG	0.350	
HD1	ARG	0.000	
HD2	ARG	0.000	
NE	ARG	-0.350	
CZ	ARG	0.350	
NH1	ARG	-0.700	
NH2	ARG	-0.700	
HE	ARG	0.450	
HH1	ARG	0.400	
HH2	ARG	0.400	
HH11	ARG	0.400	
HH12	ARG	0.400	
HH21	ARG	0.400	
HH22	ARG	0.400	

Figure 3.3. Descriptive charge file. Example of a descriptive charge file used for a DelPhi run. This abbreviated file shows the charges used for all of the atoms for Ala and a charged Arg residue. This full length file contains the charges for a neutral Arg also as well as charges for every atom of all the amino acids.

molecule from a medium equal to the interior dielectric of the molecule into a medium of external dielectric of the solution. In addition, DelPhi can output a potential map with the electrostatic potentials of the entire protein calculated. This file is compatible with visual programs such as GRASP that will allow the potential map to be viewed as an electrostatic surface. Files with the calculated potential of particular atoms or residues can also be generated based on specific coordinates of the pdb file that are input into the program.

3.2.2 Visualization of electrostatic surfaces

Perhaps the most extensive application of continuum electrostatics in structural biology has been the visual representation of $\phi(r)$ for proteins (Honig and Nicholls, 1995). These representations yield information about protein function that is not evident from the three-dimensional structure alone. The GRASP program is utilized extensively to display electrostatic surfaces (Nicholls et al., 1991). GRASP contains extremely rapid algorithms for the construction of rendered molecular surfaces and for solving the Poisson-Boltzmann equation. The molecular surface constructed by GRASP can be molecular or accessible and can be color coded by electrostatic potential derived from its internal Poisson-Boltzmann solver or external programs such as Delphi. Potential maps calculated using DelPhi can be transferred into GRASP. In addition to electrostatic potentials, surfaces can also be colored by other properties, such as any of those of the underlying atoms (e.g. hydrophobicity) or by its own intrinsic properties, such as local curvature. For electrostatic potentials, positive potentials are colored blue, negative potentials are colored red, and neutral potentials are colored white.

3.3 Electrostatic “hot spots” in CD2

DelPhi was used to calculate the electrostatic potential of domain 1 of CD2 and identify the electrostatic “hot spots” of the protein using the pdb file 1hng, an X-ray structured by Jones et al. which was edited to contain only residues 1-99 (Jones et al., 1992). The energy terms of the whole protein were calculated. CD2 has an overall net charge of -1, subsequently, it is expected the net potential would be negative. The electrostatic potential of the whole molecule was -483.8 kt which was then converted to -286.9 kcal/mol. The effect of salt in the system was also measured. DelPhi allows the input of salt concentrations to better simulate experimental conditions, which often requires salt. The addition of salt will shield the charged residues interacting with the solvent, therefore, a decreased potential is expected. The potential of CD2 was calculated at salt concentrations of 0, 130 and 500 mM. Although both salt concentrations, 130 and 500 mM, had an effect on the protein (compared to 0 mM salt), differences between the concentrations of 130 and 500 mM of salt were minimal, -225.41 and -225.59 kcal/mol, respectively (Table 3.1). The addition of salt is screening the surface charges and therefore the electrostatic potential is decreased. However, this effect is not significant for the total charge potential for salt concentrations of 130-500 mM salt.

The potential of the individual amino acids was examined in order to identify charged clusters within the protein. Figure 3.4 shows the potentials of each amino acid in CD2 at 0 mM salt. As expected, the charged residues have significant potentials (> than 20 kcal/mol) as compared to neutral residues (Table 3.2). Among all of the positive charged residues R44 has the largest potential. For the negative charged residues D62

Conditions	Electrostatic Potential (kcal/mol)
CD2 0 salt	-286.91
CD2 130 mM salt	-225.41
CD2 500 mM salt	-225.59

Table 3.1. Electrostatic potential of CD2 under different salt conditions.

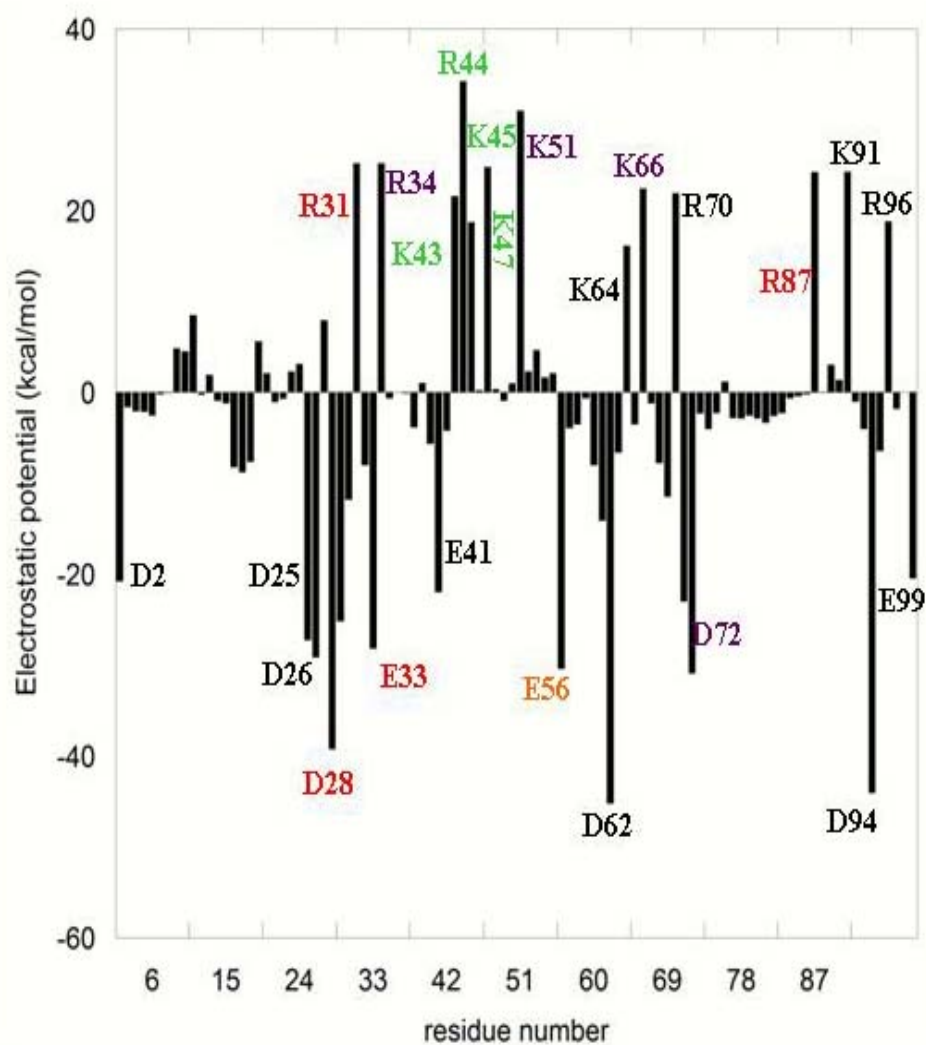


Figure 3.4. Electrostatic potentials of all of the residues in CD2. The charged residues have larger potentials compared to the non-charged residues. The residues are color coded by their proximity to one another in the tertiary folding of the protein (within 12 Å).

Residue	Electrostatic Potential (kcal/mol)
D2	-20.764
D25	-27.260
D26	-29.151
D28	-39.247
D29	-25.198
R31	25.186
E33	-28.169
R34	25.198
E41	-21.950
K43	21.590
R44	34.206
K45	18.698
K47	24.778
K51	30.971
E56	-30.408
D62	-45.250
K64	16.122
K66	22.419
R70	21.914
D72	-30.892
R87	24.236
K91	24.277
D94	-44.048
R96	18.811

Table 3.2. Electrostatic potentials of charged residues.

and D94 have the largest potential. This plot vs. sequence does not reveal any significant information about the varying electrostatic environment of the protein. However, when the electrostatic distribution was visualized in three-dimensions, several charged clusters in CD2 were identified which give insight into the different electrostatic environments of the protein. First of all CD2 has two significant surfaces based on the charge. Surface BEG is not charged whereas surface GFCC'C'' is largely charged (Figure 3.5). Four charged clusters, KKRK, KREEED, DREKKK, and KRRDD, have been identified in different regions of CD2 based on distance cutoffs of 12 Å (Figure 3.6). Cluster KKRK, so named because it consists of K43, K45, R44, and K47, is a positively charged cluster with an electrostatic potential of 68.15 kcal/mol. The residues of this particular cluster are close together in the primary sequence of the protein. Residues K43 and R44 are found in the C' β -sheet of the protein while residues K45 and K47 are in the loop regions between the C' and C'' sheets. Cluster KREEED (K43, R31, E33, E28, D29) has a mix of negatively and positively charged amino acids and is made up of residues from different regions of the primary structure. The electrostatic potential of this cluster is -41.86 kcal/mol indicating the significant influence the negatively charged amino acids residues have on the potential. In fact, the potentials of the individual positive residues in the cluster are considerably decreased compared to the potentials of the individual residues of cluster KKRK (> than 4-fold in some cases) as shown in Table 3.2. This cluster is located on the CD48 binding surface and is involved in the selectivity of binding of CD2 to CD48. The individual potentials of these residues are not as important because their role is to interact with oppositely charged residues on the CD48 surface.

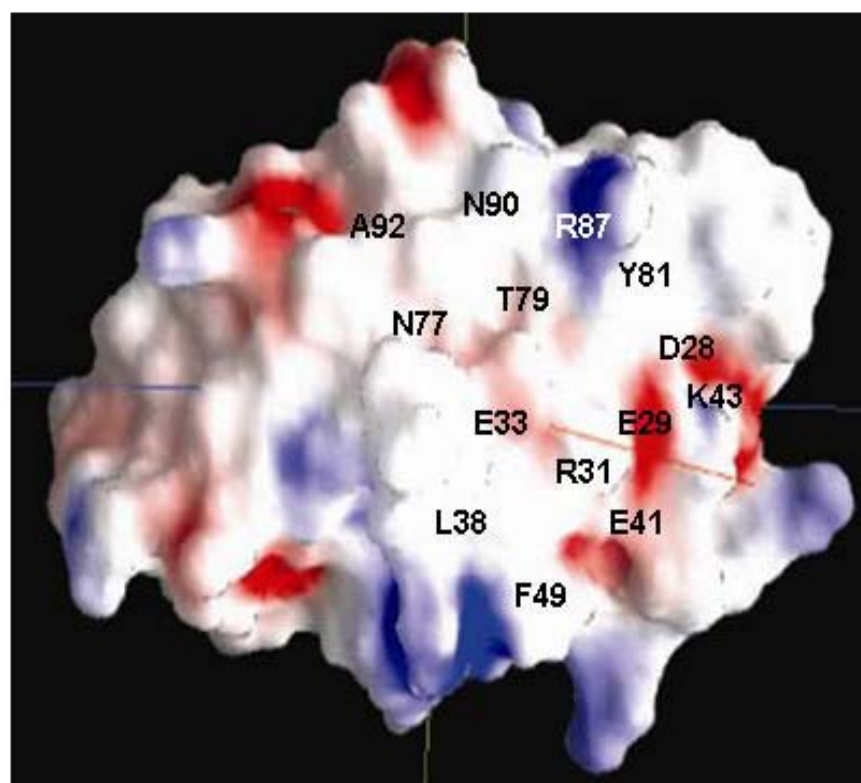


Figure 3.5. Charged surface of CD2. The GFCC'C" surface of CD2 is highly charged. Seventy percent of the residues are charged.

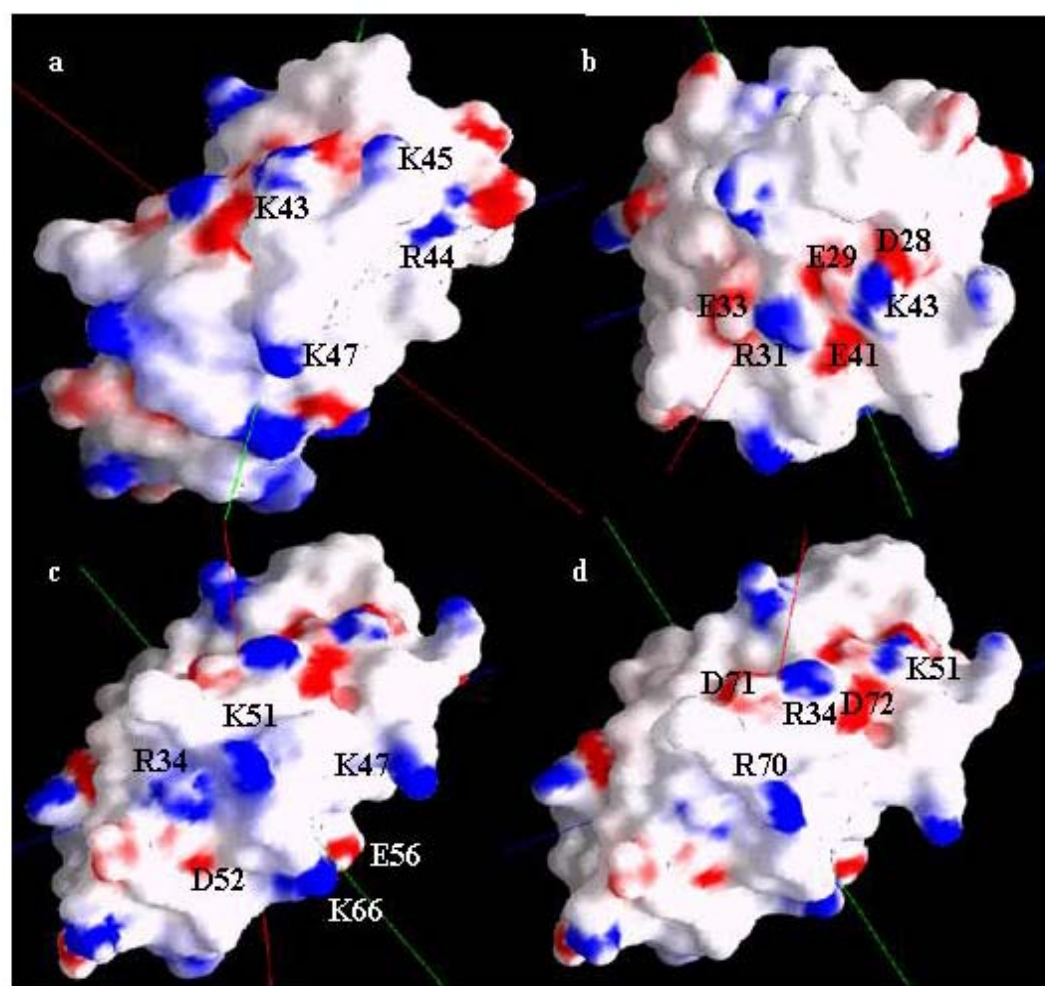


Figure 3.6. Charged clusters in different regions of CD2. (a) Cluster KKRK (b) Cluster KREEED (c) Cluster DREKKK (d) Cluster KRRDD

The potential of the cluster DREKKK (D52, R34, E56, K47, K51, and K66) is 49.44 kcal/mol and for KRRDD (K51, R34, R70, D71, and D72) is -10.21 kcal/mol.

3.4 Environment of the designed Ca^{2+} binding sites

Several Ca^{2+} binding sites were designed into CD2 based on a geometric method (section 2.1). Table 3.3 shows the distances of the designed proteins to each of the charged clusters. The distances were measured using the pdb files generated from the design program Dezymer. The Ca^{2+} distance was measured to the central residues of the cluster.

Each of the first generation designed proteins (206, 5606, and 6775) with Glu as a bidentate are in varying electrostatic environments. Designed site 5605 is 11 Å from the DREKKK cluster which has a strong positive potential whereas sites 206 and 6775 are in close proximity to the KRRDD and KREEED cluster, respectively, which both have negative potentials. However there is a 3-fold difference between the electrostatic potentials of those two clusters. Therefore, the studies on these first generation sites presented in the following chapter will provide insight into whether these clusters have a significant effect on calcium-binding affinity.

The second generation site 7E54 is situated on a surface that does not have a significant number of charges. The closest charge cluster to the binding site is KKRK which is 17 Å away. Previous studies on the distance dependence of electrostatics on protein interactions indicate that electrostatic interactions > than 15 Å from a particular site do not have a significant effect on the properties of that site (Gilson and Honig, 1988a). Therefore, cluster KKRK may not have a significant effect on site 7E54. On the

Designed Protein	Ligands	Distance from KKRK Å	Distance from KREEED Å	Distance from DREKKK Å	Distance from KRRDD Å
206 (1 st generation)	V39E, I65E, D72E, Y67E, L63Q	13	24	9	4
5606 (1 st generation)	V80E, F21E, K91D, L89D, V78N	23	29	11	12
6775 (1 st generation)	L68E, D72, I65E, I14Q, L95D	21	11	20	27
6D79 (2 nd generation)	T79D, A92E, N77, N90	20	9	16	17
7E54 (2 nd generation)	A54E, I65D, T69D, N67, D72	17	19	23	19

Table 3.3. Electrostatic environment of CD2. Each designed calcium binding site is in a different electrostatic environment in CD2.

other hand, site 6D79 is located on a highly charged surface (> than 70% of the residues are charged). It is situated near the CD48 binding surface and is in close proximity to the KREEED cluster (9 Å). The role of the electrostatic environment on the Ca^{2+} binding affinity of these two proteins can be investigated by mutating residues around the Ca^{2+} binding sites to introduce or remove charged residues.

3.5 Designing charge distribution variants of 6D79

Study of the first generation proteins will provide insight into the role of the electrostatic environment on Ca^{2+} binding affinity via three different proteins in three different electrostatic environments. The second generation protein 6D79 allows for the study to be carried out from a different standpoint. The alteration of the electrostatic environment of a single Ca^{2+} binding site, 6D79, through site-directed mutagenesis of charged residues around the Ca^{2+} binding site provides a system for studying the role of non-local electrostatic interactions.

The second generation designed site 6D79 is located on a surface in close proximity to the CD48 binding site of CD2 (Figure 3.7). The CD48 binding site is a highly charged surface where greater than 70% of the residues are polar or charged. In comparison, less than 50% of residues forming the binding sites of antibodies and proteases are polar or charged (McAlister et al., 1996). This protein is ideal to study the effects of electrostatic environment on Ca^{2+} binding affinity through single mutations around the coordination shell that change the electrostatic environment of the Ca^{2+} binding site.

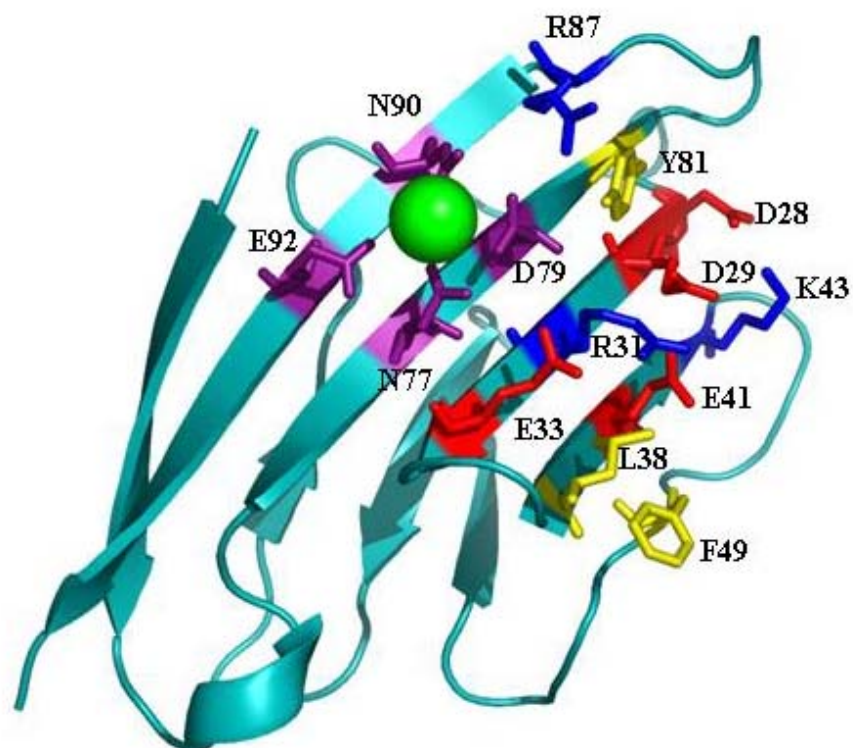


Figure 3.7. Model structure of 6D79. Second generation designed site 6D79 (Ca²⁺ binding ligands shown in purple) is located on the CD48 binding surface of CD2. The residues involved in CD48 binding are highlighted. The positive charged residues are shown in blue, negative charged residues are in red, and the aromatic and aliphatic residues are in yellow.

To facilitate the design of the charge distribution variants, several criteria were applied including distance of the residue to the Ca^{2+} binding site, electrostatic potential change of the Ca^{2+} binding site, solvent accessibility of the residue side chain, and the role of the residue in CD48 binding. For distance, a cutoff value of 15 Å was used due to previous work that demonstrated the range of electrostatic interactions using several systems including the bovine pancreatic trypsin inhibitor (BPTI) and rhodanese (Gilson and Honig, 1988b; Sharp and Honig, 1990). The effect of the interactions of charge residues at a distance of 15 Å drops by 79%. After examining the distance of the charged residues around the 6D79 binding site eight residues, E29, R31, R34, E41, K51, R87, K91, and D94 are within the distance restraint (Table 3.4).

The next step was determining what potential change would occur if these residues were mutated to an opposite charge. The pdb file for 6D79, which was generated by Dezymer, was modified to reflect a point mutation at each residue to its opposite charge. The files were minimized using a Kollman all-atom force field with 1000 iterations. DelPhi was utilized to calculate the potential of the Ca^{2+} binding site with each introduced mutation with a grid size of 65 with no salt. Table 3.4 shows the electrostatic potential changes that were observed at the Ca^{2+} binding site with each charge reversal mutation. All of the mutations triggered varying changes in the electrostatic potential. As expected, the change from positive charged amino acids to negative charged (R34D, K91D, R87D, R31D, and K51E) ones produced negative potential shift changes while the opposite was true for negative charged amino acids changed to positive charged ones (Figure 3.8). These changes were distance dependent

Residue	Distance from Ca ²⁺ binding site Å	Δelectrostatic potential kcal/mol	% solvent accessibility
E29	12.5	R 38.0	24.5
R31	8.0	D -56.7	64.8
R34	9.5	D -39.7	62.3
E41	13.6	R 36.6	19.4
K51	15.0	E -19.4	89.5
R87	6.9	D -73.8	95.8
K91	11.4	D -41.6	36.1
D94	14.1	K 30.3	43.4

Table 3.4. Charged residues on the surface of site 6D79 which have been identified as possible charge distribution mutants.

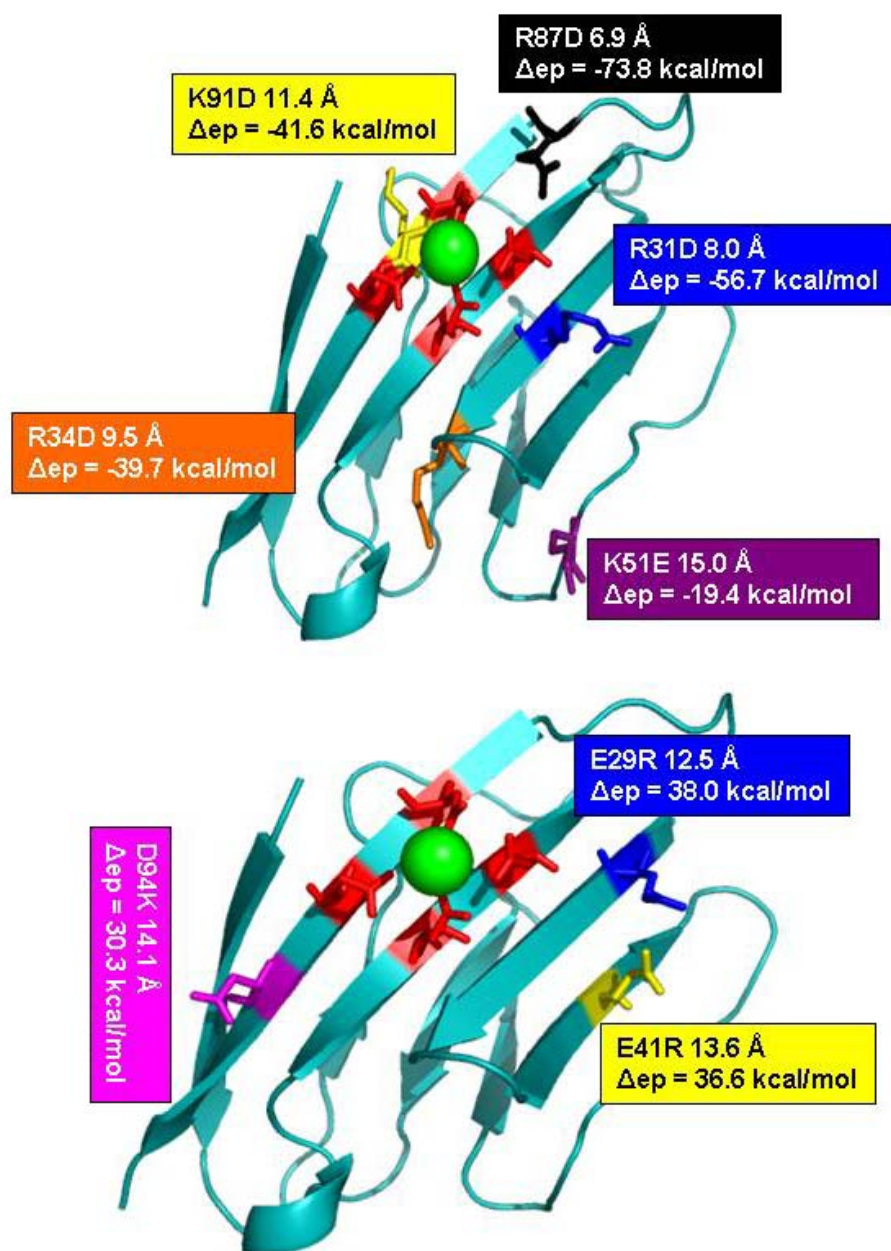


Figure 3.8. Electrostatic potential changes. The electrostatic potential changes of the Ca^{2+} binding site of CD2.6D79 upon single mutations around the coordination shell.

with R87D (6.9 Å) having the most substantial effect and K51E (15.0 Å) having the least effect. The discriminating factor for these calculations was an electrostatic potential change greater than 18 kcal/mol. This value was based on a study by Gilson and Honig where they examined the range of electrostatic interactions (Gilson and Honig, 1988b). In this study, the range of interactions influencing the potential at the active site of the sulfur-transferase, rhodanese were examined. This enzyme has a number of features that make it interesting for studying electrostatic interactions. First, it contains a network of salt-bridges, some of which are buried at the subunit interface. Second, it has two positively charged side chains that are important in binding the enzyme's negatively charged substrates. Third, the key side chain in catalysis is a cysteine whose pK is shifted downward to 6.5, and which must be in the ionized thiolate form for catalysis to occur (Ploegman et al., 1979). The contribution of various elements of the protein structure to the stabilization of the thiolate ion were calculated by performing a detailed Poisson-Boltzmann calculation with Cys247's sulfur charged (-1), and determining the potential at each other charged atom in the protein (Gilson and Honig, 1988b). The study established that for a surface that falls into an electrostatic potential range between 10 – 100 kcal/mol, a distance dependent electrostatic potential range of 18 - 20 kcal/mol was observed. The calculation was also performed with the protein BPTI, which yielded a distance dependent electrostatic potential range of 19 - 20 kcal/mol. At this point, residue K51 was eliminated as a possible choice due to its small potential change.

The next evaluation was based on solvent accessibility. Several studies have demonstrated that residues that have a greater solvent accessibility are least likely to

undergo a significant conformational change upon mutation (Mercier et al., 2003); (Golinelli et al., 1998; Kapila et al., 2001). The Naccess program calculates the atomic accessible surface defined by a rolling probe of a given size around a van der Waals surface. The program uses the Lee & Richards method (Lee and Richards, 1971), where a probe of a given radius is rolled around the surface of the molecule, and the path traced out by its center is the accessible surface. Typically, the probe has the same radius as water (1.4 Å) and therefore the surface described is referred to as the solvent accessible surface. The calculation also makes successive thin slices through the three-dimensional molecular surface to calculate the accessible surface of individual atoms. Naccess can be used to calculate both atomic and residue accessible surfaces for amino and nucleic acids. Previous mutagenesis studies with CD2 have shown that mutations of residues with side chain accessibilities of 20% do not trigger conformational changes to occur, as a result, 20% was used as the cutoff factor for solvent accessibility (Jones, 2006). The % solvent accessibility of the possible residues is shown in Table 3.4. In this case all of the residues possessed a % solvent accessibility greater than 20%.

The final criteria, and most stringent, was the filtering out of residues important for CD48 binding. Site 6D79 is adjacent to the CD48 binding surface and many of the charged residues on this surface are essential for binding. Previous studies have shown that CD48 is either completely eradicated or significantly diminished as a consequence of point mutations of residues D29, E41, R31, and R87 (Davis et al., 1998a), hence, these residues were eliminated. Table 3.5 reveals the residues chosen for mutation and the difference in their primary sequence as compared to CD2 and 6D79 and Figure 3.9

Protein	34	75	79	91	92	94
CD2	R	T	T	K	A	D
6D79	R	T	D	K	E	D
K91E	R	T	D	E	E	D
K91I	R	T	D	I	E	D
D94K	R	T	D	K	E	K
D94A	R	T	D	K	E	A
R34E	E	T	D	K	E	D
R34I	I	T	D	K	E	D
T75D	R	D	D	K	E	D
T75K	R	K	D	K	E	D

Table 3.5. The residues chosen for the charge distribution variants.

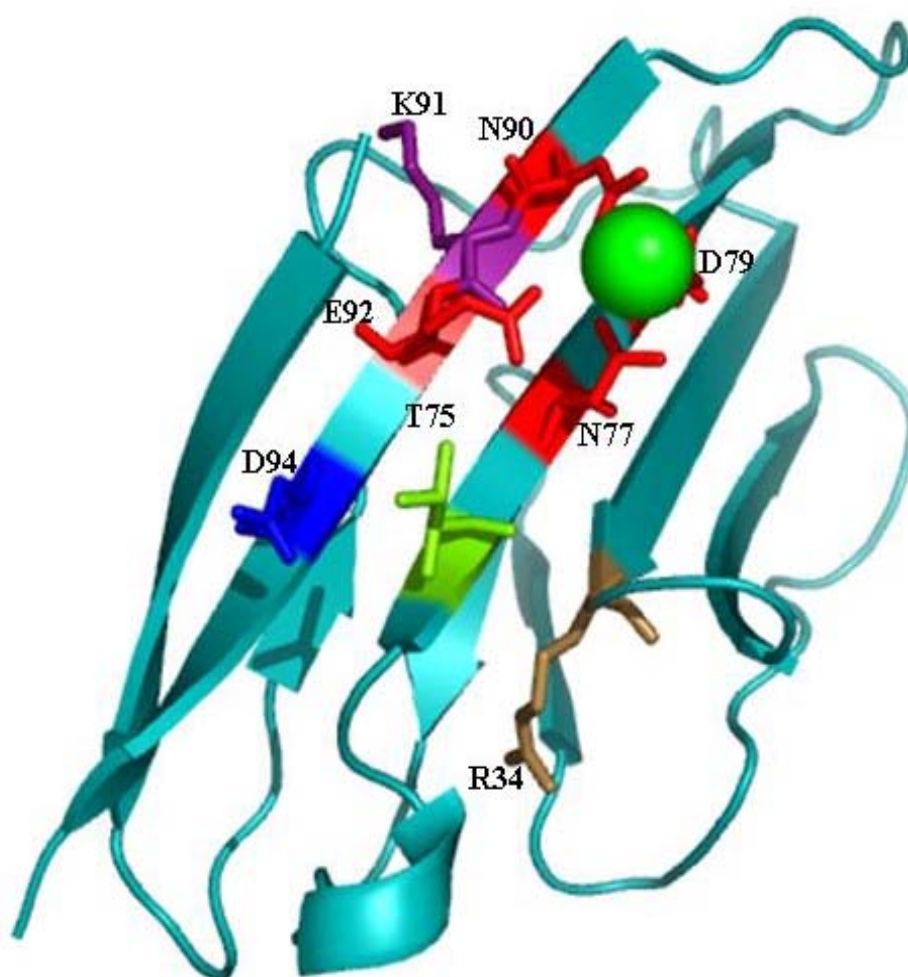


Figure 3.9. Position of charge distribution variants. Position of the residues chosen for mutation as part of the charge distribution study in respect to the Ca²⁺ binding site 6D79 (Ca²⁺ binding ligands are shown in red).

reveals their position in respect to the Ca^{2+} binding site. For each residue, mutants will be engineered to the opposite charged residue and to a neutral residue. This allows for a two-fold study, the examination of the effect of merely removing the charge and also the effect of replacing it with an opposite charge. In addition to the charged residues, the neutral residue T75 was also chosen to be mutated. This residue allows for the opportunity to examine what effect the introduction of either a positive or a negative charge has on the Ca^{2+} binding site. This residue is 8.5 Å from the Ca^{2+} binding site, has a solvent accessibility of 62% and is not involved in CD48 binding, hence it meets our criteria for a charge distribution mutant.

3.6 Predicting Ca^{2+} binding affinity from free energy calculations

The prediction of binding affinities of proteins could be an important aspect of protein design. Two different methods have been employed to predict the binding affinity of the charge distribution variants. These methods include free energy calculations and Debye-Hückel methods.

3.6.1 Free energy calculations

As mentioned earlier, the comparison of grid energies from several DelPhi runs with the same grid conditions can yield solvation energies, binding energies, and salt effects. The grid energy is obtained from the product of the potential at each point on the grid and the charge at that point, summed over all points on the grid. The use of grid energy calculations has been successful in measuring the binding energy of the barnase-barstar complex (Rocchia et al., 2001). The binding energy can be calculated using equation 3.6.

$$\Delta G_{\text{binding}} = \Delta G_{\text{gridP} \cdot \text{Ca}} - \Delta G_{\text{gridP}} - \Delta G_{\text{gridCa}} \quad \text{Eq. 3.6}$$

The grid conditions were kept constant for every run. These conditions include the grid size and percentage fill of the lattice. The grid size is an odd integer number of point per side of the cubic lattice. A larger grid size gives a better resolution of the molecules on the lattice and results in more accurate potentials. The grid size for the free energy runs was set at 65, which was more than sufficient for the rather small CD2 protein. The percentage fill of the lattice is the percentage of the objects longest linear dimension to the lattice linear dimension. This will affect the scale of the lattice which is measured in grids/angstrom. A large percentage fill will provide a more detailed mapping of the molecular shape onto the lattice and was set at 20% for these calculations. In addition, the x, y, and z coordinates of each protein were centered at the same position on the lattice for all of the calculations. The interior and exterior dielectric constants were set at 4 and 80, respectively. The ΔG_{grid} is calculated by DelPhi in units of kT , where k is Boltzmann's constant and T is the absolute temperature, and can be converted into kcal/mol using the relationship $1 \text{ kT} = 0.582 \text{ kcal/mol}$. Furthermore, the free energy was converted to joules using the relationship $1 \text{ cal} = 4.184 \text{ J}$ and the K_d of each variant was calculated using equation 3.7.

$$\Delta G = RT \ln K_d \quad \text{Eq. 3.7}$$

where the gas constant R was $8.31451 \text{ J/K}\cdot\text{mol}$ and the temperature was 298 K . The calculated grid energy of Ca^{2+} alone was 26.35 kT (15.34 kcal/mol). This number was used for the calculations of the $\Delta G_{\text{binding}}$ for all of the variants. Table 3.6 reveals the

Protein	$\Delta G_{\text{grid PCa}}$ kT	$\Delta G_{\text{grid P}}$ kT	$\Delta G_{\text{binding}}$ (kcal/mol)	K_d (mM)
6D79	115.24	101.55	-7.37	51.02
T75K	171.00	115.69	16.86	9.01×10^2
T75D	116.61	106.23	-9.30	23.47
D94K	115.50	105.61	-9.58	20.94
D94A	106.87	98.12	-10.24	16.02
R34E	151.67	155.26	-17.43	8.81×10^{-4}
R34I	111.83	105.19	-11.47	9.75×10^{-5}
K91E	118.71	117.87	-14.85	2.49×10^{-3}
K91I	116.11	99.63	-5.75	98.39

Table 3.6. Calculated binding affinities of the charge distribution variants.

Grid energies of the charge distribution variants in the Ca^{2+} free and Ca^{2+} bound forms compared to 6D79. The free energy of binding was calculated using equation 3.2. The binding affinities based on these free energies was also calculated using equation 3.3.

calculated ΔG_{grid} for the Ca^{2+} bound and Ca^{2+} free forms of the variants, the $\Delta G_{\text{binding}}$, and the predicted K_d of each variant.

The free energy of binding and binding affinity follows an interesting trend. Additionally, the predicted binding affinities for the variants are high compared to natural extra-cellular Ca^{2+} binding proteins. For the T75 variants, T75D has a slightly more negative binding energy as compared to 6D79 indicating the protein may bind Ca^{2+} stronger than 6D79. Further, for the K_d calculation, a 2-fold difference is observed between the binding of 6D79 and T75D. These results are expected with the introduction of a negative charge around the Ca^{2+} coordination shell. On the other hand, the T75K variant displays an opposite effect. Both the $\Delta G_{\text{binding}}$ and K_d change indicate decreased Ca^{2+} binding compared with 6D79. In fact, the $\Delta G_{\text{binding}}$ is positive (16.86 kcal/mol), which may indicate binding is very weak. The predicted K_d value for T75K is 17-fold weaker than 6D79 indicating the introduction of a positive charge around the Ca^{2+} binding site has a considerable effect on the Ca^{2+} binding affinity. It is interesting to note that T75K has a more dramatic effect on the Ca^{2+} binding affinity than T75D. This result is peculiar but may be attributed to the increased side chain length of Lys as compared to Asp. Lys may have a more significant effect because it is closer to the Ca^{2+} binding site than Asp.

The effect of removal of a positive charge around the Ca^{2+} binding site can be observed for the K91 and R34 variants. The differences between these two variants are the distance of each residue from the Ca^{2+} binding site, 11.4 Å for K91 and 9.5 Å for R34, and the position of the residues. K91 is positioned beneath the Ca^{2+} binding site

(Figure 3.9) and therefore if a negative charge is in this position it may act as an anchor to attract the positively charged Ca^{2+} . On the other hand, the positive charged Lys may act to repel the Ca^{2+} thus a lower binding affinity is observed. Upon the removal of the positive charge of both of these residues varying results are observed for the $\Delta G_{\text{binding}}$. For R34I, a 1.5-fold more negative binding energy compared to 6D79 is observed, while for K91I, the binding energy is actually more positive indicating weaker Ca^{2+} binding. The introduction of a negative charge at these positions has a more considerable effect on the binding energy. Both R34E and K91E have a more negative binding energy (2.3-fold for R34E and 2-fold for K91E). The calculated binding affinities are also stronger compared to 6D79 with R34E having a slightly more significant effect

Both of the D94 mutants, D94K and D94A, display more negative binding energies and stronger binding affinities than 6D79. The binding affinity differences are 2 and 3-fold stronger for D94K and D94A, respectively. This trend is contradictory to the phenomenon observed in calbindin D_{9k} , where the removal of negative surface charges around the Ca^{2+} binding site weakened the Ca^{2+} binding affinity (Linse et al., 1991).

The results of our free energy calculations are flawed due to the limitations of continuum models in the case of metal binding. This may be indicated by the high predicted binding affinities of the variants. The extracellular Ca^{2+} binding protein cadherin has a Ca^{2+} binding affinity of 0.16 mM (Yang et al., 2000d). The predicted binding affinity of 6D79 is 300-fold weaker (51.02 mM). There are several reasons continuum models are not accurate in measuring metal binding. First, for measuring binding energies, these continuum models rely on large structural changes between the

apo and bound forms of the components. Because of the electrostatic nature of protein-metal interactions only small differences in positions occur, which result in large errors on the estimation of the interaction energies (Schymkowitz et al., 2005). The effect of these small changes is exacerbated in the 6D79 protein because all of the Ca^{2+} binding ligands are found in β -sheet secondary structure of the protein. The rigidity of β -sheets severely limits local conformational change upon Ca^{2+} binding. Second, there has been a lag in the advancement of development of force fields that are sensitive enough to detect small metals and metal induced changes. For this reason, the grid energy calculated by DelPhi is not accurate in calculating the binding energy of metal binding. Due to the small size and charge on the Ca^{2+} , a large error is introduced in the calculation when measuring Ca^{2+} alone. The use of grid energy has been successful in previous analysis of protein-protein interactions (Nicholls A and Honig, 1991) but to date has not been reported for use in metal binding studies. This study would greatly benefit from advances in the development of force fields that can accurately model metals and metal binding. In the case of the charge distribution variants of 6D79 other methods for predicting binding affinity were explored.

3.6.2 Predicting Ca^{2+} binding affinity through Debye-Hückel methods

The next approach to predicting the binding affinities of the charge distribution variants was through the use of Debye-Hückel methods. This model is based on the work of Lee et al. that showed that the complex pair-wise interactions of the deprotonation of a charge residue on a protein surface can be treated approximately using simple Debye-Hückel potential arguments (Lee et al., 2002). This model (Equation 3.3) has been

extended to include the dissociation of divalent metal complexes by Dr. Stuart Allison (Chemistry Department, Georgia State University, Atlanta, GA). A schematic of the model is represented in Figure 3.10. The model takes into account the effect of a single alteration in the charge on site j , by $q\Delta z_j$, located a distance r_{ij} from the site, on the K_d of binding site i . That K_d is compared to a reference K_d^o that does not have a charge alteration (in this case 6D79). The K_d can be calculated with Equation 3.8

$$\ln K_d - \ln K_d^o = (2q^2 \Delta z_j^2 / k_B T \epsilon_r r_{ij}) e^{-\kappa r_{ij}} \quad \text{Eq. 3.8}$$

where $q\Delta z$ is the charge change, k_B is Boltzmann's constant, T is temperature, ϵ is the dielectric constant of water, r is the distance between site i and j , and κ is the Debye-Hückel screening length. To test the validity of this model for our host system, equation 3.3 was used to predict the pK_a values of CD2. Chen et al. had previously determined the pK_a values of all of the negatively charged residues in CD2 by monitoring the chemical shift changes using ^{15}N -labelled protein (Chen et al., 2000). The program surfprot, written by Dr. Stuart Allison, utilizes this Debye-Hückel model to calculate pK_a values. Surfprot was employed to calculate the pK_a values of the negatively charged amino acid residues in CD2 and compare them with the experimentally determined values. The experimental buffer conditions were used (20 mM potassium phosphate) to calculate the ionic strength for the calculations and a dielectric constant of 4 was used. The comparison of the calculated and experimental results is shown in Table 3.7 and in Figure 3.11. Although the calculated pK_a values follow the general trends of the experimental values there are some glaring differences between the two sets of values.

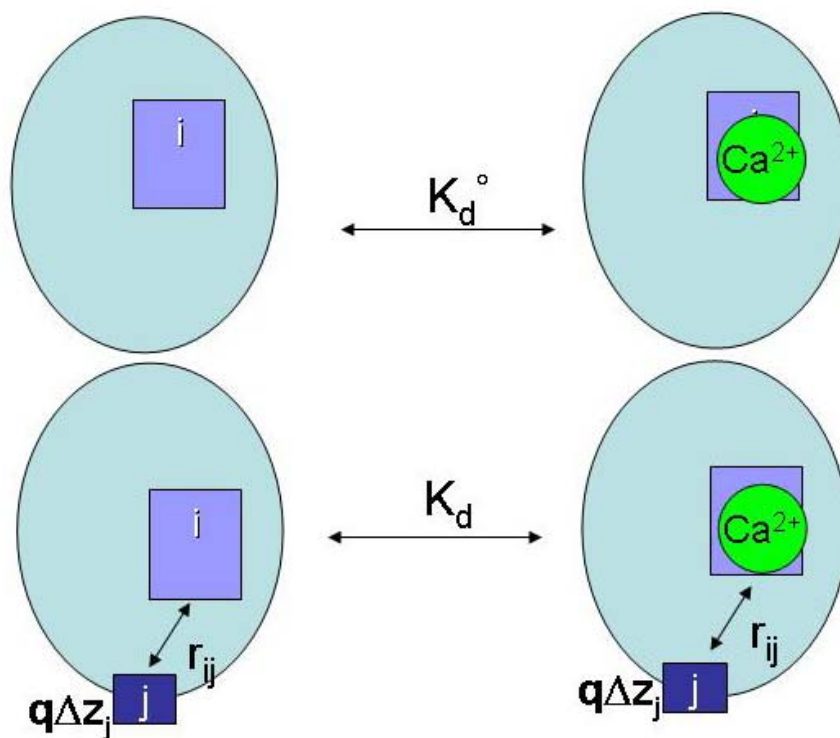


Figure 3.10. Schematic diagram of the Debye-Hückel model. The effect of a single alteration in the charge on site j, by $q\Delta z_j$, located a distance r_{ij} from the site, on the K_d of binding site i is measured using equation 3.4.

Residue	pK _a experimental	pK _a calculated surfprot
Asp2	3.55	3.5
Asp25	3.53	3.51
Asp26	3.58	3.76
Asp28	3.57	3.73
Asp62	4.15	3.73
Asp71	3.18	3.43
Asp72	4.14	3.68
Asp94	3.87	3.72
Glu29	4.42	4.86
Glu33	4.16	4.38
Glu41	6.73	4.89
Glu56	3.92	4.41
Glu99	4.25	4.31

Table 3.7. Comparison of pK_a values of CD2 calculated with surfprot and obtained experimentally.

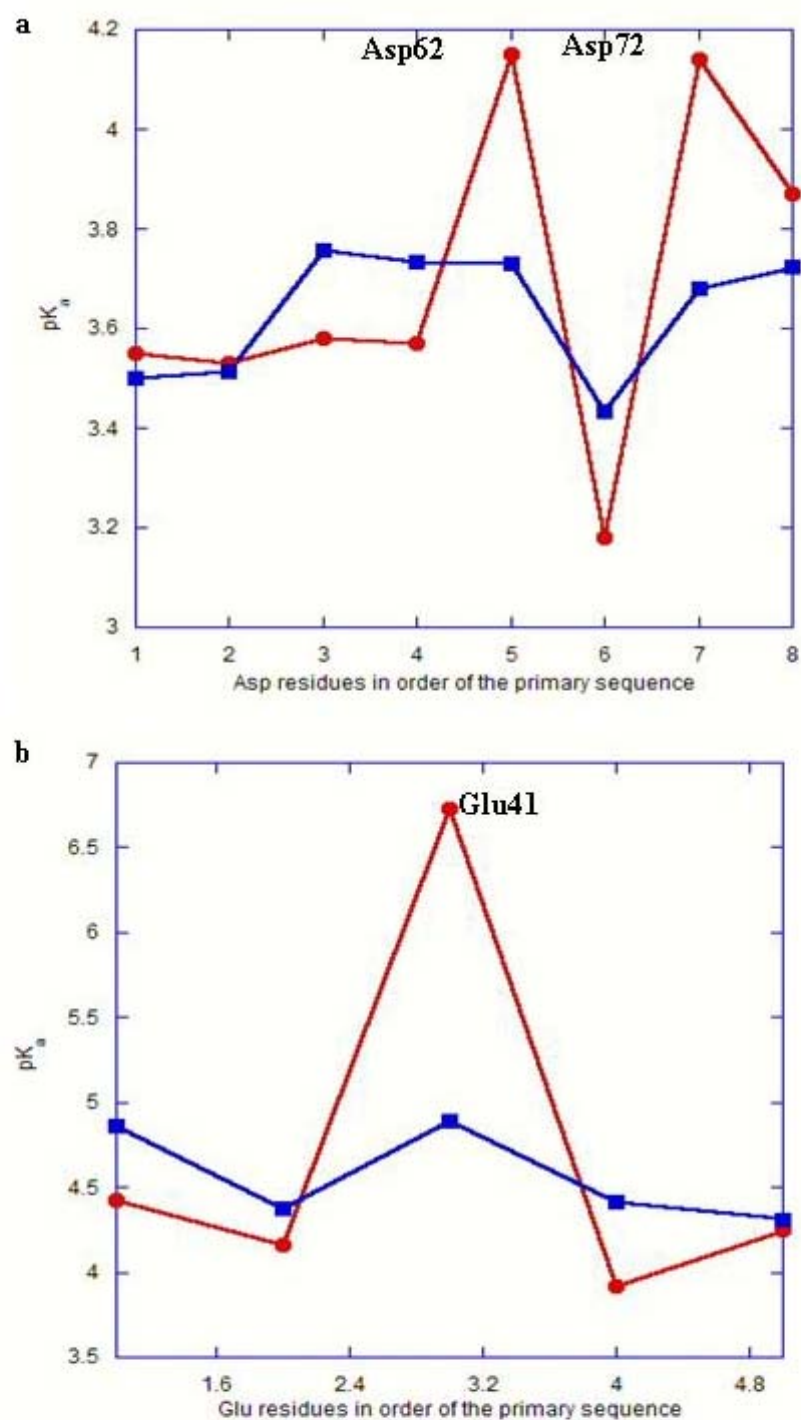


Figure 3.11. pK_a values of CD2. Comparison of pK_a values between the experimental data (red) and the data calculated using surfprot (blue) surfprot for the Asp (a) and Glu (b) residues of CD2.

The most significant difference is the pK_a of E41. The experimentally determined value of pK_a is 6.73, which is unusually high. The typical pK_a value of Glu in proteins is 4.4 (Stryer, 1995).

Examination of the crystal structure of CD2 reveals that the side chain methylene groups of E41 and E29 are adjacent, and the carboxyl groups are positioned such that charge-charge repulsion would be very strong (Chen et al., 2000). These two residues are within 5.6 Å of one another. The two Glu residues are prevented from rotating away from one another under the force of electrostatic repulsion due to the limited mobility of these and adjacent residues as a result of a complex electrostatic network. Interestingly, two basic groups (R31 and K43) located immediately adjacent to E41 do not aid in reducing the electrostatic repulsion. Rather they themselves are influenced and dominated by the effects of the proximity of E41 and E29 (Chen et al., 2000). This is evident in the electrostatic calculations of the charged cluster KRREEED (K43, R87, R31, E33, E28, and E29) from **section 3.3**, which has a significantly negative potential of -41.86 kcal/mol, even though it has a mixture of positive and negative charged amino acids. The greater influence is seen on residue 31, which has a potential of 4.8 kcal/mol. This is at least three-fold lower compared to the other basic amino acids K43, R44, K47, K51, K66, and R87, which have potentials of 14.9, 18.2, 19.2, 22.2, 22.1, 17.2 kcal/mol, respectively. The closest approach of the carboxyl O^\ominus atoms of E41 and E29 is 2.9 Å, which is sufficiently close for a hydrogen bond to be formed between a protonated E41 and a deprotonated E29. Chen et al. hypothesize that this scenario is consistent with the elevation of the pK_a of E41 and that the lack of perturbation of the pK_a of E29 is due to

the balance of additional side chains with positive and negative charges (R31, K43, and D28) in the immediate vicinity (Chen et al., 2000). Additionally, surfprot failed to accurately measure the pK_a values of D62 and D72, which have slightly elevated pK_a values. These residues had the lowest degree of solvent accessibility compared to other acidic residues, which could explain their elevated pK_a values.

The Debye-Huckel model does not take any structural information into account except for the distance, r_{ij} . For pK_a calculations, the model assumes that the full pK_a shift is attributable to a pairwise Coulombic interactions between charged groups. Hence this model was not precise in calculating the pK_a of E41 due to the structural origins of the elevated pK_a value. To evaluate whether another model would be more precise in predicting pK_a values, MCCE was used.

3.7 Calculating pK_a's with MCCE

The Multi-conformation Continuum Electrostatics (MCCE) method calculates the equilibrium conformation and ionization state of protein side chains, buried waters, ions, and ligands at a defined pH (Alexov and Gunner, 1999); (Alexov and Gunner, 1997). This program relies on conformers in order to take into account the relaxation effects of the molecule in solvent. Conformers are calculated from the input pdb file to take into account different atomic positions and ionization states. Look-up tables of electrostatic and non electrostatic conformer self- and pair-wise interactions are generated. Monte Carlo sampling then establishes each conformer's probability in the Boltzmann distribution at a given pH. In Monte Carlo sampling, after a conformer is generated, its energy is calculated and compared to the original conformation. If the energy of the new

conformer is less than the original, the new conformer is kept and once again slightly perturbed to a new conformation. If the energy of the new conformer is greater than the old, the new conformation is discarded randomly. A Monte Carlo sampling repeated many times will lead toward a minimum free energy. The MCCE procedure is divided into 3 stages: (1) selection of conformers; (2) calculation of energy look up tables; (3) Monte Carlo sampling and calculation of pK_a 's. This model incorporates both PB and DH theory. The calculation of energies is performed using the PB equation as calculated by DelPhi which has been incorporated into the program. Pairwise interactions are also calculated, which is the basis of the DH theory.

MCCE was first used to calculate the pK_a values of the Ca^{2+} binding protein calbindin D_{9K} to determine the accuracy of the program. Previous studies by Kesvatera et al. have determined experimentally the pK_a 's of the charged residues of calbindin D_{9K} using multi-dimensional NMR (Kesvatera et al., 2001b). In addition the pK_a 's were calculated using Monte-Carlo simulations and the values were compared to the experimental values to determine the accuracy of their simulation method.

For the calculations by Kesvatera et al., each protein atom was represented as a hard sphere, either neutral or charged and impenetrable to any solvent ions (Kesvatera et al., 2001b). Glutamates, aspartates, tyrosine, lysines and the C and N-termini were allowed to change their charge state according to the solution pH. The protein was kept fixed at the center of a spherical cell, whose radius was determined by the protein concentration. Equation 3.6 was used for the calculations.

$$u(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon r_{ij}} \quad \text{Eq. 3.9}$$

where q is the partial charge, ϵ_0 is the permittivity of free space, and r_{ij} is the distance between the two particles. A dielectric constant of 77.8 was used at the experimental temperature of 300 K. Salt ions were added to the cell in accordance with experimental conditions. Thus, this model accounts for the screening effects of both salt and protein which is in contrast to the PB and DH methods. The simulation was performed using a standard Monte Carlo sampling of Equation 3.6 (Kesvatera et al., 2001b).

Figure 3.12 and Table 3.8 show the comparison of the experimental and Monte-Carlo simulation calculated, and MCCE calculated pK_a values. The Kesvatera method agrees very well with the experimental data. pK_a values for several groups are strongly shifted up or down the pH scale, as compared with standard pK_a values characteristic to unstructured peptides. This is caused by electrostatic interactions between ionic groups in the folded native protein (Kesvatera et al., 2001b). The MCCE method does accurately predict the pK_a values of several of the residues (82%) but it does not properly predict the pK_a values of some of the residues. The most dramatic differences are observed for D54, E60, and E65, where MCCE calculates strongly basic pK_a values. Even though Monte Carlo simulations are used in both models there is a difference in data accuracy. This may be due to the Kesvatera model using a large dielectric constant of 77.8 as compared to MCCE which utilizes two dielectric constants, 4 and 80, for the interior and exterior, respectively. Even with the discrepancies between the experimental and calculated pK_a values of calbindin D_{9k}, MCCE was also used to calculate the pK_a values of CD2.

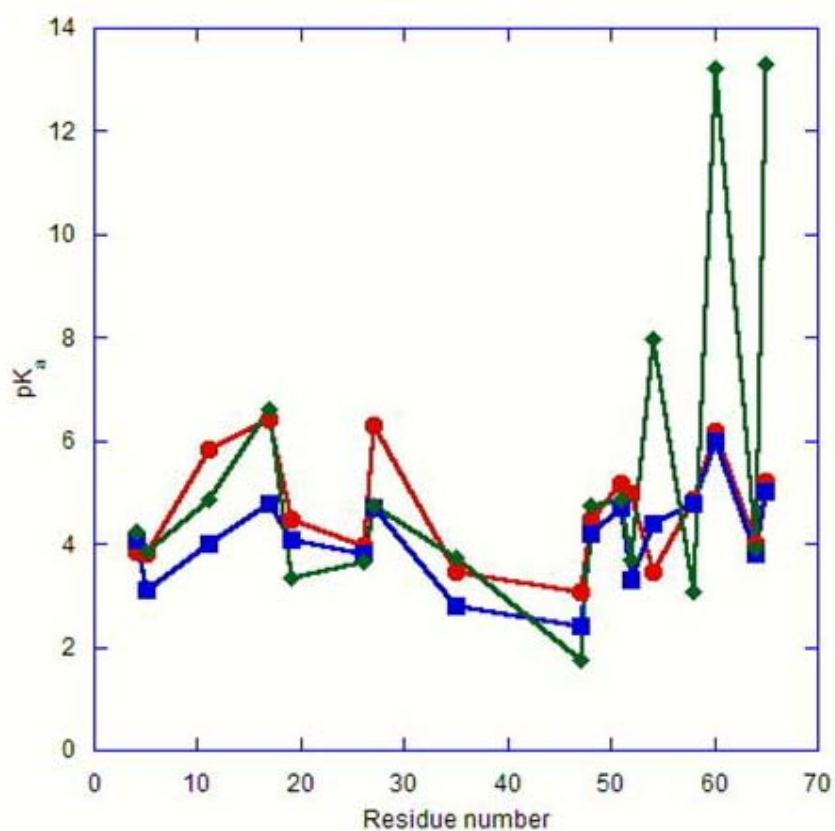


Figure 3.12. pK_a values of calbindin D_{9k} . Comparison of the pK_a values of the charged residues of calbindin D_{9k} obtained experimentally (red), through Monte Carlo simulations (blue), and by MCCE (green).

Residue	Experimental pK_a values	Monte Carlo calculated pK_a values	MCCE calculated pK_a values
Glu4	3.85	4.10	4.24
Glu5	3.80	3.10	3.87
Glu11	5.83	4.00	4.86
Glu17	6.41	4.80	6.62
Asp19	4.49	4.10	3.33
Glu26	3.97	3.80	3.67
Glu27	6.31	4.70	4.75
Glu35	3.45	2.80	3.73
Asp47	3.07	2.40	1.76
Glu48	4.49	4.20	4.73
Glu51	5.16	4.70	4.86
Glu52	4.97	3.30	3.68
Asp54	3.45	4.40	7.98
Asp58	4.85	4.80	3.09
Glu60	6.20	6.00	13.2
Glu64	4.00	3.80	3.93
Glu65	5.22	5.00	13.3

Table 3.8. Comparison of pK_a values of calbindin D_{9k} obtained experimentally, through Monte Carlo simulations by Kesvatera et al., and calculated by MCCE by us.

An interior dielectric constant of 4 was applied to calculate the pK_a values of the acidic residues of CD2. Figure 3.13 and Table 3.9 compare the calculated and experimental data. Similar to the surfprot data, the trend between the data has been replicated. However, MCCE does predict an elevated pK_a for E41 although it exceeds the experimental data. It also incorrectly predicts an elevated pK_a of E29, probably taking into account the hydrogen bonding between the two residues but not the stabilization of E29 by its adjacent charged residues.

Previous studies have shown the calculation of pK_a values by continuum methods are more accurate if an elevated dielectric constant of 20 is used (Antosiewicz et al., 1996). Calculations on folded proteins yield dielectric constants in the range of 2-5 (Gilson and Honig, 1986) except when ionized groups are considered to contribute to the fluctuations in dipole moment of the molecule (King, 1991). It has been argued that the lower values of 2-4 are the most appropriate for use in PB calculations, because the contribution of ionizable groups to the fluctuations in dipole moment should not be included in the dielectric constant (Antosiewicz et al., 1994). It is somewhat surprising that pK_a values calculated with an assumed protein dielectric constant of 20 agree with measured pK_a values substantially better than those calculated with a dielectric constant of 4 (Antosiewicz et al., 1994). Although the reason for this is unknown, there are several possibilities. One possibility is that the charges and radii that are typically assigned to the atoms of the protein are not optimally parameterized. Another is that crystallization of the proteins for determination of their structures leads to conformational changes. Studies have shown that when calculations were carried out on conformations

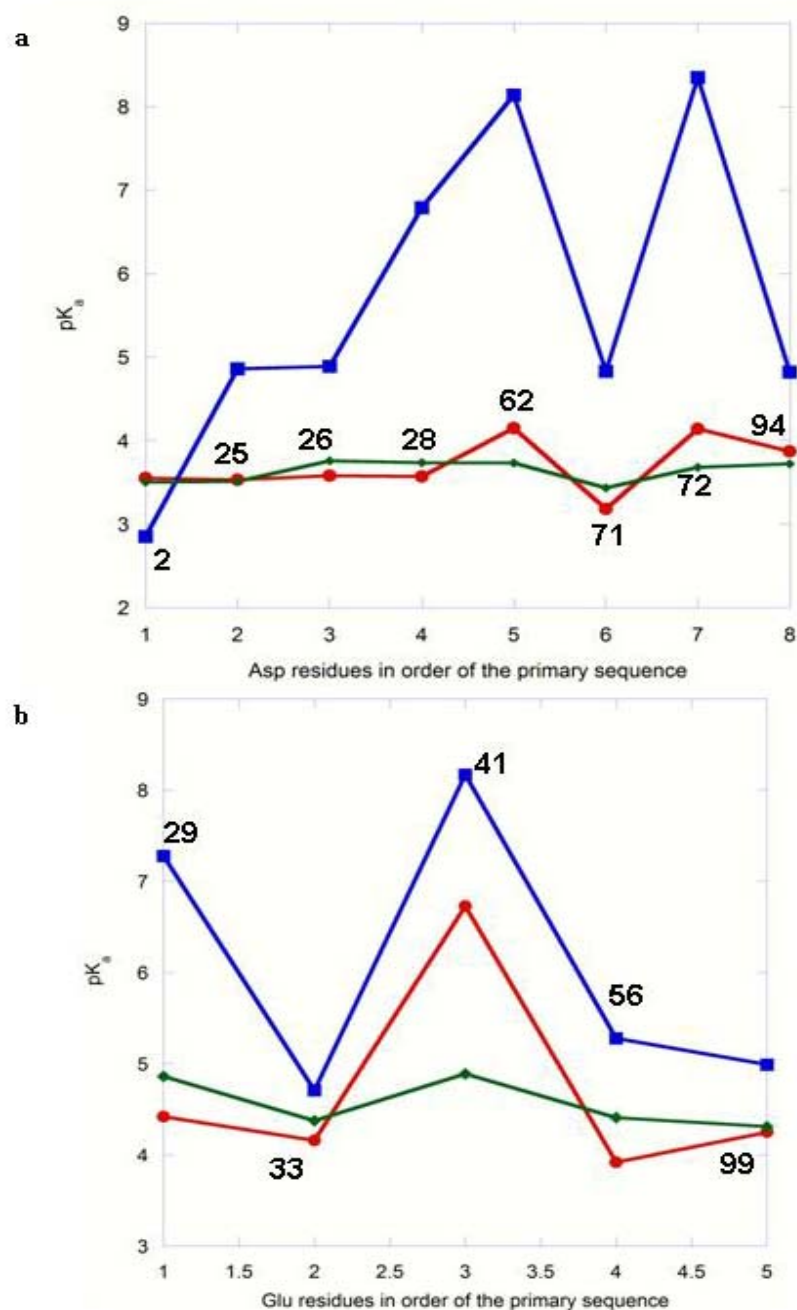


Figure 3.13. pK_a values of CD2 calculated by MCCE. Comparison of pK_a values between the experimental data (red), the data calculated using MCCE (blue), and calculated using surfprot (green) for the Asp (a) and Glu (b) residues of CD2.

Residue	pK _a experimental	pK _a calculated MCCE
Asp2	3.55	2.85
Asp25	3.53	4.86
Asp26	3.58	4.89
Asp28	3.57	6.79
Asp62	4.15	8.14
Asp71	3.18	4.83
Asp72	4.14	8.35
Asp94	3.87	4.82
Glu29	4.42	7.28
Glu33	4.16	4.71
Glu41	6.73	8.17
Glu56	3.92	5.28
Glu99	4.25	4.99

Table 3.9. Comparison of pK_a values of CD2 calculated with MCCE and obtained experimentally.

more appropriate to the solution conditions under which the pK_a measurement was made, accurate pK_a values would be obtained with a protein dielectric constant of 4 (Antosiewicz et al., 1996).

Even though MCCE aims to minimize the hindrance of the conformational change through the calculation of conformers, the calculation was repeated with a dielectric constant of 20 to determine if the data were more accurate. Figure 3.14 and Table 3.10 show the comparison of the data. With a dielectric constant of 20, not only is the trend the same between the experimental and calculated data but the accuracy of the calculation is increased. The predicted pK_a of E41 is very close the actual value as well as for E29. This data suggest that conformational changes are not the only basis for the increased accuracy of a dielectric constant of 20. And the optimization of charge and radii parameters may play a significant role.

The Debye-Hückel arguments only take into account one dielectric constant, the dielectric constant of water (80), and therefore it would not be useful to test this method with a dielectric constant of 20. Since accurate pK_a data could not be calculated using surfprot there was no attempt at using the model for predicting binding affinity.

3.8 Predicting stability using the Fold-X force field

Since previous methods to predict binding affinity were unsuccessful, the focus was shifted to calculating the stability of the mutants using the force field Fold-X. The empirical force field Fold-X was initially developed to allow rapid free energy calculations (Guerois et al., 2002). The capabilities of Fold-X have recently been extended for the prediction of water and metal binding sites and their affinities

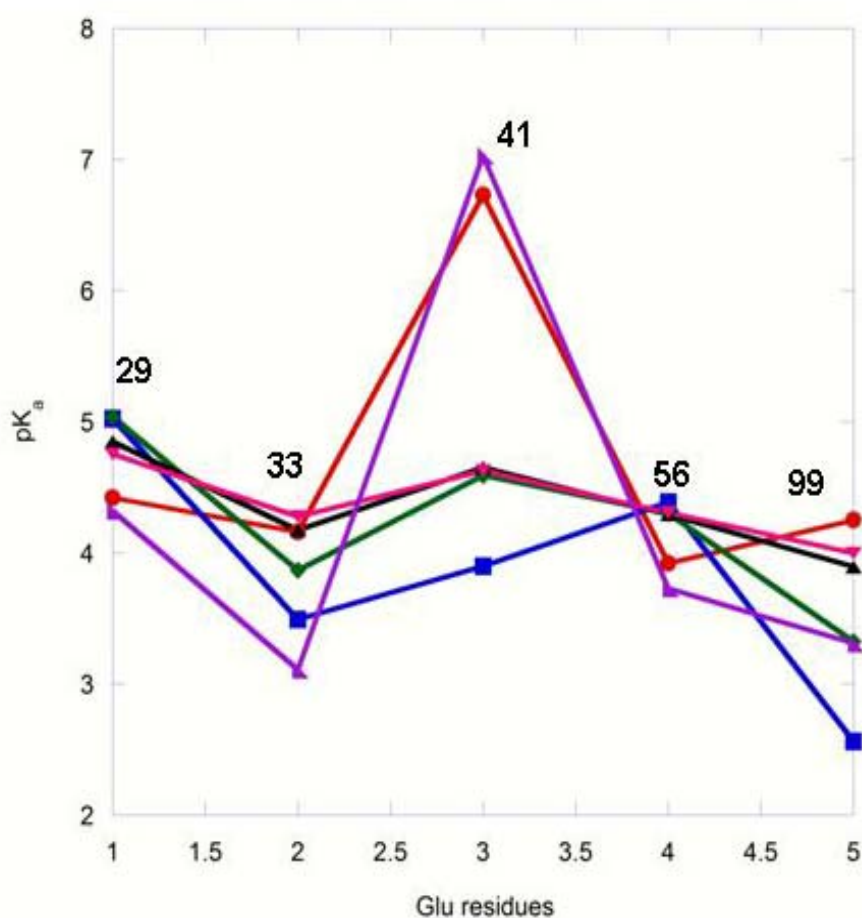


Figure 3.14. pK_a calculations with an increased dielectric constant. Comparison of pK_a values between the experimental data (red), the surfprot data with dielectric constants of 4 (blue), 20 (green), 50 (black), and 80 (pink) and the data calculated using MCCE with a dielectric constant of 20 (purple) of the Glu residues of CD2.

Residue	pK _a experim ental	pK _a calculat ed MCCE	pK _a Surfprot E=4	pK _a Surfprot E=20	pK _a Surfprot E=50	pK _a Surfprot E=80
Asp2	3.55	2.51				
Asp25	3.53	4.15				
Asp26	3.58	3.44				
Asp28	3.57	1.97				
Asp62	4.15	6.25				
Asp71	3.18	2.92				
Asp72	4.14	5.78				
Asp94	3.87	2.42				
Glu29	4.42	4.32	5.0227	5.0365	4.8499	4.7535
Glu33	4.16	3.11	3.4923	3.8679	4.1730	4.2743
Glu41	6.73	7.03	3.8987	4.5894	4.6510	4.6257
Glu56	3.92	3.73	4.3850	4.3007	4.2918	4.3049
Glu99	4.25	3.31	2.5630	3.3228	3.8910	3.9945

Table 3.10. Comparison of pKa values of CD2 calculated with MCCE with a dielectric constant of 20, surfprot at various dielectric constants, and obtained experimentally.

(Schymkowitz et al., 2005). Fold-X uses a continuum solvation model based on experimental transfer energies of model compound representing amino acids from water to vapor and to cyclohexane extended with the explicit consideration of water bridges (Schymkowitz et al., 2005). This approach allows for a more accurate energy evaluation. This force field has been successful in predicting the position of binding sites for Mn^{2+} , Mg^{2+} , Ca^{2+} , Cu^{2+} and Zn^{2+} , as well as accurately predicting binding affinities of known metal binding sites. However, the extended application model is currently not available for public use. Thus, the original model of Fold-X was used to calculate the stability of the designed protein 6D79 and its charge distribution variants. The pdb file utilized had the hydrogens added and was minimized as discussed in **section 2.11**. Each pdb file was loaded onto the Fold-X server. The calculation was done with 10 mM salt at 298 K.

The results indicate that compared to CD2, the designed protein 6D79 is somewhat more stable (Table 3.11). In regards to the charge distribution variants, the removal of the negative charge at D94 increases the stability compared to 6D79. An increase in stability is also seen in the T75K and the K91I mutant proteins. The increased stability, however, may not translate in increased binding affinity because this is a highly charged surface. The increase in stability may be due to increased charge balance on the surface. Both R34 mutants have a decreased stability with the effect more pronounced with R34E. Interestingly for the K91 and T75 mutants one of the mutations has a stabilizing effect (K91I and T75K), and the other has a destabilizing effect (K91E and T75D). The addition of negative charges at these two positions may decrease the charge balance whereas in the case of the neutral residue I for K91 and the basic residue K for

Protein	Conformational stability free energy kcal/mol
CD2	-9.50
6D79	-10.04
K91E	-7.25
K91I	-14.10
D94K	-13.22
D94A	-15.36
R34E	-1.90
R34I	-6.47
T75K	-11.70
T75D	-8.54

Table 3.11. Predicted conformational stabilities of CD2, designed protein 6D79, and its charge distribution variants.

T75 the charge balance is increased. The experimental data for stability and binding affinity of 6D79 and its charge distribution variants will be discussed in chapters 5, 6, and 7.

Chapter 4. Conformational Analyses and Metal Binding of Designed Ca^{2+} Binding Proteins without a Stable Structure

4.1 The rationale for studying the first generation proteins

As discussed in **section 1.5** electrostatic interactions around the Ca^{2+} coordination shell play a role in the binding affinity. The electrostatic environment has been shown to be important for Ca^{2+} binding as is the case for the binding of Ca^{2+} to the protein α -lactalbumin and not the structurally similar protein lysozyme. The goal with the first generation proteins is to develop criteria for the design and selection of Ca^{2+} binding proteins and to examine the role of electrostatic environment on site-specific Ca^{2+} binding affinity. Each of the first generation proteins are in different electrostatic environments (**section 3.4**) and with different deviations from the target design site with a pentagonal bipyramidal geometry and therefore the comparison of their metal binding properties will help gain insight into the precise role of the electrostatic environment. In this chapter, first, the design of the first generation variants will be discussed. Second, the optimization of expression of these proteins will be addressed. Third, the conformational and metal binding properties of these proteins will be analyzed. Fourth, the attempt to generate stable variants of these proteins is addressed. Fifth, the design of second generation proteins will be discussed.

We have chosen CD2, a non Ca^{2+} dependent cell-cell adhesion molecule on T-lymphocyte cells, as a host protein. Previous work by our lab and others has shown that CD2 is an excellent host protein (**section 1.9.2**). It retains a stable β -sandwich structure over a large range of pH, salt concentration, and organic solvent concentration (Driscoll

et al., 1991; Erickson, 1994; Williams et al., 1987). To achieve the task of designing de novo calcium-binding sites in proteins, we have carried out detailed structural analysis of more than 500 calcium-binding sites in small chelators and natural calcium-binding proteins as well as extensive folding studies of the host protein CD2. After surveying several classes of naturally evolved calcium-binding proteins, we chose the most common geometry of pentagonal bipyramid for the coordination shell of the target calcium-binding site in CD2 (CD2.Ca1). Parameters describing the pentagonal bipyramidal geometry of calcium-binding sites such as the Ca-O bond lengths (2.0-3.0 Å), O-Ca-O angles (ideal angle $\pm 45^\circ$), and O-Ca-O-C dihedral angles in proteins were established and tested on naturally evolved calcium-binding proteins (Hellinga and Richards, 1991; Wilkins et al., 2002; Yang et al., 2002b) (Figure 1.6). Analogous to the natural sites, one of the seven coordination sites was left open to allow solvent to access and to reduce steric crowding in our description. All first generation of Ca^{2+} binding sites are designed with two oxygen atoms from the bidentate Glu. The remaining four positions were occupied by unidentate ligand oxygen atoms contributed by the carboxyl groups of Asp, Asn, Glu, and/or main-chain carbonyls. Approximately 7000 potential sites were generated by the program. The designed proteins were filtered using several criteria. First, the Ca^{2+} binding site should have a native geometry with small $U(p)$ number since all of the natural binding sites in 13 proteins analyzed by Yang et al. have a small geometric deviation ($U(p) < 20$) (Yang et al., 2000c). Second, the calcium should be solvent accessible. Third, the mutations should introduce little or no side-chain steric conflicts with the preexisting atoms. Fourth, minimal disruption of hydrogen bonding and

hydrophobic interactions is required. Fifth, three or four negatively charged residues at the primary coordination shell are preferred (Marsden et al., 1990). On the basis of these criteria, we identified four sites (designated 206, 2780, 5606, and 6775) involving five mutations to construct the primary coordination sphere.

Based on these criteria four designed proteins were selected for engineering. Proteins 206, 2780, 5606, and 6775 (Figure 4.1), named for the order they were predicted by the program Dezymer, were chosen for the study of the effect of the electrostatic environment on Ca^{2+} binding affinity. The mutations required to engineer the Ca^{2+} binding site in each of these proteins is shown in Table 4.1. For proteins 206, 2780, and 6775, five mutations were engineered while for 5606 only four mutations were required and the native D72 was used as a ligand. All four sites have a -4 charge in the coordination shell. Sites 206 and 5606 have U(p) numbers of ~4. Site 2780 has a U(p) of 19 and 6775 has a U(p) of ~25. This site was also chosen to test whether altering the geometry has any effect on Ca^{2+} binding. The site was also called CD2.Ca1 and published in *JACS* (Yang et al., 2003).

Designed site 206 is made up of ligands from the C', and E strands of the protein (39, 63, and 65, respectively). The site also includes residue 67 from the EF loop and residue 72 from the 3_{10} helix. Site 5606, which is in close proximity to site 206, contains one residue, 68, from the EF loop. The two sites share one residue, 72, though in 5606 the natural Asp was used and for 206 the residue was mutated to Glu. Site 5606 also contains residues from the B, E, and G strands of the protein, residues, 14, 65, and

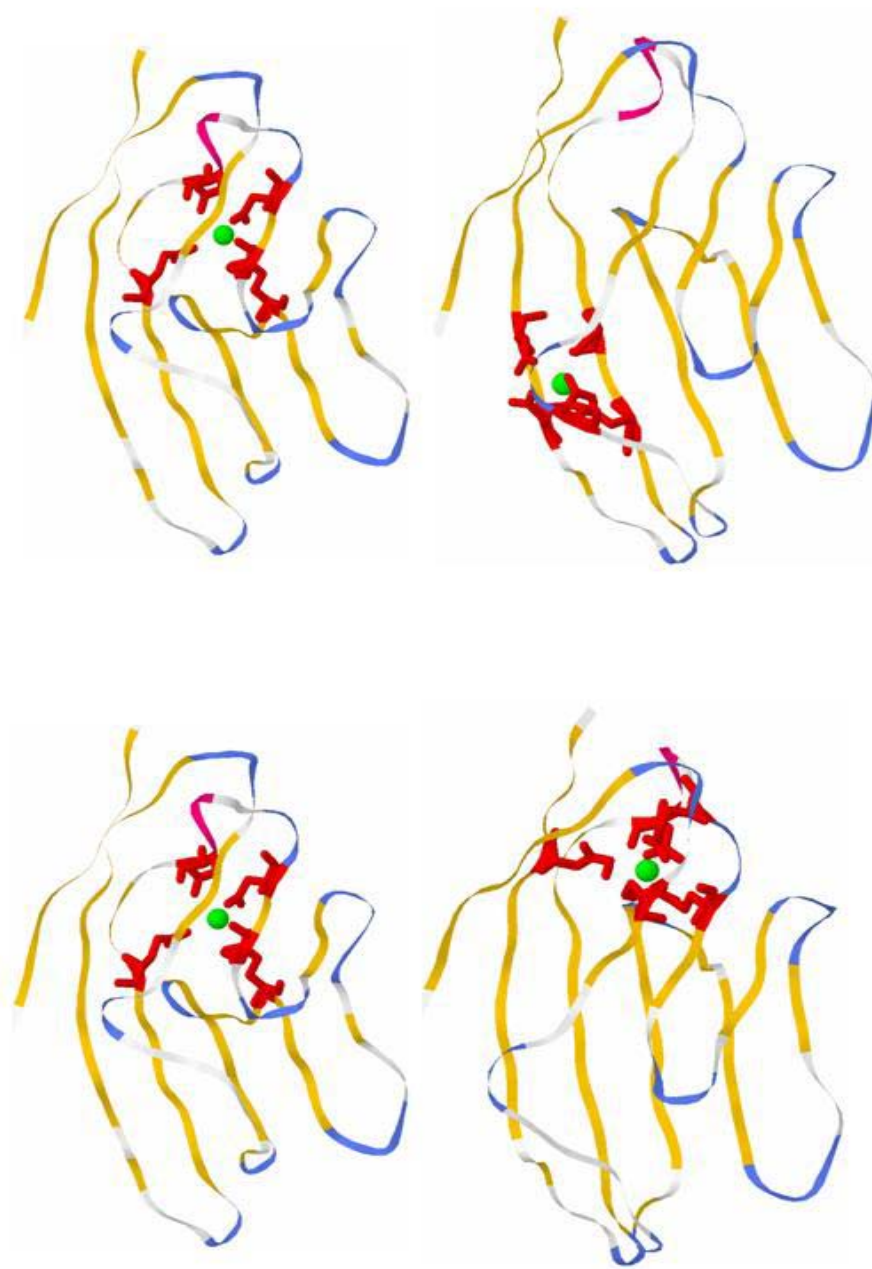


Figure 4.1. The first generation designed proteins. First generation proteins 206 (a), 2780 (b) 5606 (c), an 6775 (d) are in different regions of CD2

Designed Protein	U(p)	Ca ²⁺ Binding Ligands
206	3.15	V39E, I65E, D72E, N67E, L63Q
5606	4	L68E, D72, I65E, I14Q, L95D
6775	25	V80E, F21E, K91D, L89D, V78N

Table 4.1. Ca²⁺ binding ligands of the first generation proteins. The engineered mutations to form the Ca²⁺ binding sites in each of the first generation designed proteins

95, respectively. For site 6775 the bidentate oxygen from Glu (V80E) is located at the end of the F strand, while the four side-chain oxygens come from strand F (V78N), the G strand (I89D and K91D), and the flexible BC loop (F21E). As with many natural calcium-binding proteins, these designed calcium-binding site is partially exposed to solvent (Yang et al., 2003). This designed calcium binding sites are in close proximity with several aromatic residues (Tyr-81 and Trp-32) that allow us to use fluorescence resonance energy transfer to monitor calcium binding.

As discussed in **section 3.4**, each designed site is in a different electrostatic environment. Site 206 is within 4 Å of the charged cluster KRRDD and 9 Å of the charged cluster DREKKK, which have electrostatic potentials of -10.21 and 49.44 kcal/mol, respectively. The effect of long range electrostatic interactions in proteins has been observed at distances up to 15 Å (Gilson and Honig, 1988a), therefore, both clusters should have an effect on the electrostatic environment of 206. The electrostatic environment of site 5606 is highly positive. The site is 11 Å from the charged cluster DREKKK. Site 6775 is 11 Å from the cluster KREEED which has an electrostatic potential of -41.86 kcal/mol. This site is the only site in a solely negatively charged electrostatic environment.

4.1.1 Cloning, expression and purification of the first generation designed proteins

The first generation designed proteins were engineered and cloned into the pGEX-2T vector for expression by Dr. Wei Yang, Dr. Zhi-ren Liu (Yang, 2001), and Leanne Isley (Isley, 2001). Site 2780 was described in detail in the dissertations by Drs. Wei Yang and Anna Wilkins. In this chapter, I will focus on sites 206, 5606, and 6775 of

the first generation designed proteins. The pGEX-2T vector incorporates an affinity tail protein, glutathione s-transferase (GST), fused to the protein of interest. The GST fusion partner provides a convenient and efficient method for the purification of the expressed fusion proteins using affinity chromatography and is one of the most commonly used fusion protein expression systems (Leong, 1999). The vector contains a sequence between GST and the protein of interest that encodes the thrombin recognition sequence Leu, Val, Pro, Arg, Gly, Ser. This will allow for cleavage of the fusion protein to yield the protein of interest. The GST fusion system has been successful in purifying wild type CD2 in significant quantities (Driscoll et al., 1991; Parker and Clarke, 1997a).

Protein expression was optimized comparing four different media types for growth. Terrific Broth (TB), Luria-Bertani (LB), 2x YT, and SOB media were all analyzed to see which media generated the largest amount of protein. The difference in the media recipes is shown in Table 4.2. The tryptone, yeast, and salt concentrations are varied amongst the four different media. The designed protein 6775 was used for the optimization. The protein was expressed in each of the four media (1 L each medium). Figure 4.2 shows the sodium dodecyl polyacrylamide gel electrophoresis (SDS-PAGE) of the expression. After 3 hr of induction with IPTG expression of the fusion protein was observed in all four media. However, larger bands were observed in the 2x YT and the LB media although the bands for 2x YT media were somewhat larger than for LB. The increased concentrations for tryptone and yeast for 2x YT as compared to LB media is the most suitable for expression of these proteins. On the other hand, the increased salt concentrations found in SOB and TB media are inappropriate and possibly damaging to

Medium	Ingredients for 1L preparation
LB	Bacto-tryptone 10 g Bacto-yeast extract 5 g NaCl 10 g
TB	Bacto-tryptone 12 g Bacto-yeast extract 24 g Glycerol 4 mL KH ₂ PO ₄ 17 mM K ₂ HPO ₄ 72 mM
SOB	Bacto-tryptone 20 g Bacto-yeast extract 5 g NaCl 0.5 g KCl 2.5 mM MgCl ₂ 10 mM
2x YT	Bacto-tryptone 16 g Bacto-yeast extract 10 g NaCl 5 g

Table 4.2. Bacterial expression media recipes. Recipe for four different media used to optimize the expression of the designed protein 6775.

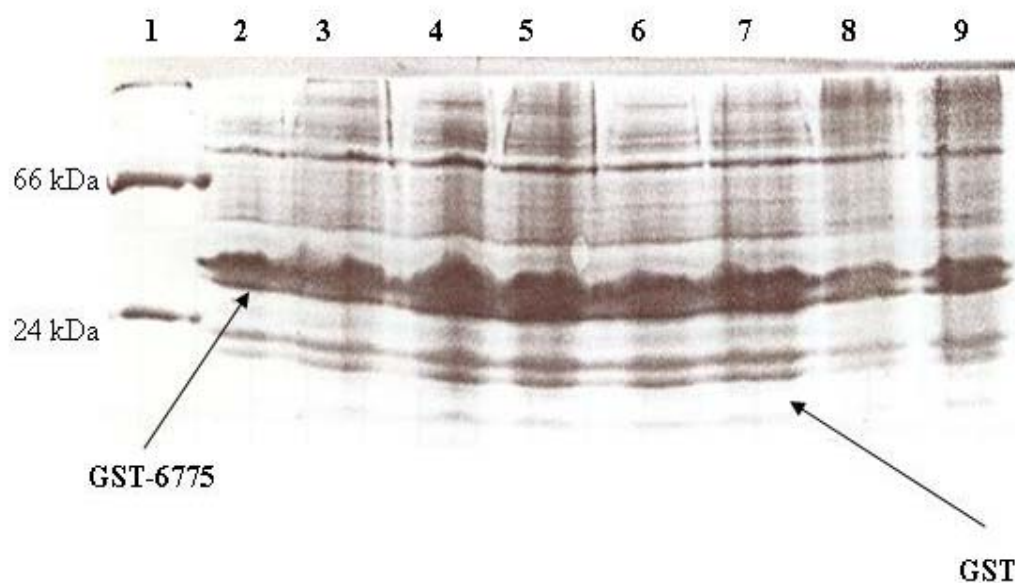


Figure 4.2. Optimization of expression of the first generation proteins. SDS-PAGE gel of expression of 6775 in four different media after 3 hr of induction. For each medium type samples were recorded in duplicate.

- Lane 1: marker
- Lane 2: SOB medium 1
- Lane 3: SOB medium 2
- Lane 4: 2x YT medium 1
- Lane 5: 2x YT medium 2
- Lane 6: LB medium 1
- Lane 7: LB medium 2
- Lane 8: TB medium 1
- Lane 9: TB medium 2

the protein. Therefore, for subsequent expressions of 6775, 206, and 5606, 2x YT media was used (Figure 4.3).

The proteins were purified using affinity chromatography. After the cell pellets were lysed through sonication and centrifuged, the filtered supernatant was decanted over glutathione sepharose beads (GS4B). GST binds to glutathione sepharose via disulfide bonds. Any excess protein and cell particulate was washed from the column with PBS buffer. The fusion protein is then cleaved with thrombin for 16 hr at 4 °C. The protein was then further purified using gel filtration. The protein was run on a Sephadex 75 column using an FPLC. The purity of the designed proteins was verified by electrospray mass spectrometry analysis (Georgia Tech University mass spectrometry facility) (Figure 4.4). The concentrations of the designed proteins were calculated using a molar extinction coefficient of $11,700 \text{ M}^{-1} \text{ cm}^{-1}$ which was experimentally determined for CD2 (Driscoll et al., 1991). The yields recovered for 206, 6775, and 5606 was 6.1, 9.1, and 1.6 mg, respectively for 4 L of expression.

4.2 Conformational analyses of the first generation proteins

4.2.1 Circular dichroism (CD)

CD is a valuable technique for examining the structure of proteins in solution. This technique has allowed detailed insights into the function of systems of ever-increasing size, including complex cellular assemblies (Kelly et al., 2005). CD refers to the differential absorption of left (rotating counter-clockwise) and right (rotating clockwise) polarized light (Kelly et al., 2005). All proteins have intrinsic optical activity making CD an excellent technique to monitor their conformational properties.

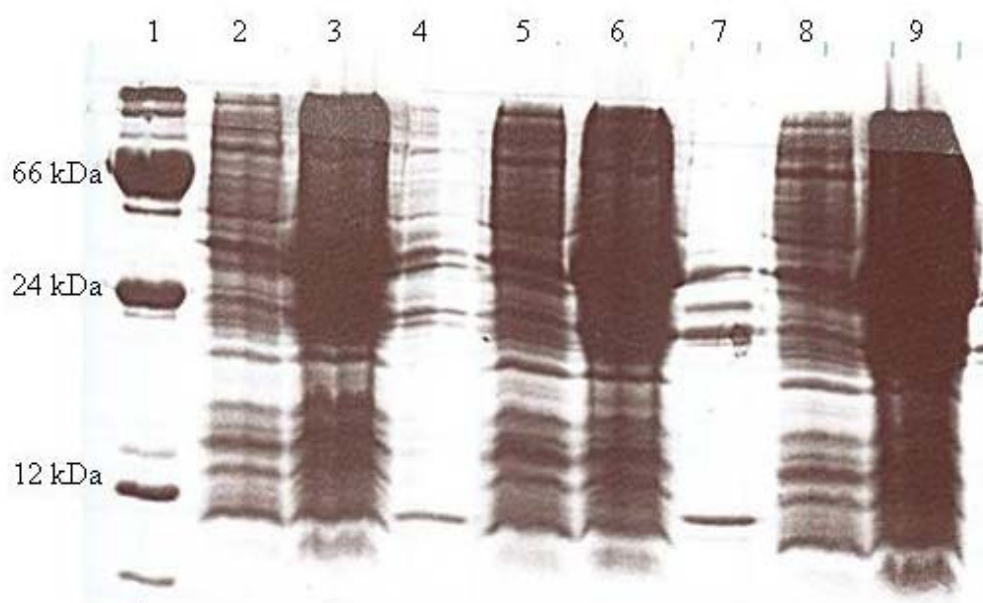


Figure 4.3. Expression of the first generation proteins. SDS-PAGE gel of expression of the first generation proteins in 2x YT medium.

Lane 1: marker
Lane 2: 206 before induction
Lane 3: 206 3 hr after induction
Lane 4: 6775 before induction
Lane 5: 6775 1 hr after induction
Lane 6: 6775 3 hr after induction
Lane 7: 5606 before induction
Lane 8: 5606 1 hr after induction
Lane 9: 5606 3 hr after induction

In the far UV region of the CD instrument (190-240 nm), absorption is due to the backbone peptide bond of proteins. The different types of regular secondary structure found in proteins, α -helix, β -sheet, and random coil, give rise to characteristic CD spectra in the far UV (Woody, 1985). Far UV is sensitive to structural changes in proteins caused by the binding of ligands including Ca^{2+} . The conformational changes of Ca^{2+} binding proteins have previously been examined using far UV CD where the binding of Ca^{2+} produces a decrease in CD signal (Venyaminov et al., 2004).

4.2.2 Fluorescence

When a molecule absorbs light, it goes from the ground state to the excited state. After a brief period, the excited species relaxes to its ground state, transferring its excess energy to other atoms or molecules. Fluorescence is the phenomenon where molecules in the excited state relaxes to the ground state (Skoog, 1996).

Compounds containing aromatic rings give the most intense and most useful molecular fluorescence emission (Skoog, 1996). The aromatic amino acids Trp and Tyr have fluorescence signals though Trp generally is the dominating residue that produces fluorescence signal in proteins. CD2 contains a buried Trp at position 32 an exposed Trp at position 7. Consequently, intrinsic Trp fluorescence can be used to probe the tertiary packing of the designed proteins. CD2 has a Trp fluorescence spectra that is blue shifted with respect to free Trp. An alteration in the tertiary packing of the protein would be characterized by a shift in the fluorescence spectra similar to free Trp.

4.2.3 The conformation of the apo form of the first generation proteins

The first step in the study of the designed proteins was to examine their conformational properties. The goal of the design was for the proteins to retain a wild type like conformation upon the addition of the Ca^{2+} binding site. Both far UV CD and Trp fluorescence were used to analyze the conformation of the designed proteins.

CD2 has a β -sheet architecture and the far UV CD spectrum of this protein exhibits a typical pattern for β -sheet proteins with a negative maxima at 216 nm. The far UV spectra of designed proteins 206, 5606, and 6775 with 1 mM EGTA compared to CD2 is shown in Figure 4.5a. Compared to wild type CD2, the conformation of the designed proteins is altered. The negative maximum of the spectra of the designed proteins is ~ 205 nm. The spectra appear to be a mixture of random coil and some β -sheet secondary structure.

The tertiary packing of the protein was examined using Trp fluorescence. CD2 has two Trp residues in different environments, Trp 32 is buried in the hydrophobic core while Trp 7 is exposed to solvent. The fluorescence spectrum of CD2 has an emission maximum at 326 nm. The spectra of the first generation designed sites have emission maxima of 356 nm similar to that of free Trp (Figure 4.5b) in 10 mM Tris pH 7.4. This indicates the tertiary packing of the protein has been perturbed because both Trp residues are now in the same environment, presumably they are exposed to solvent.

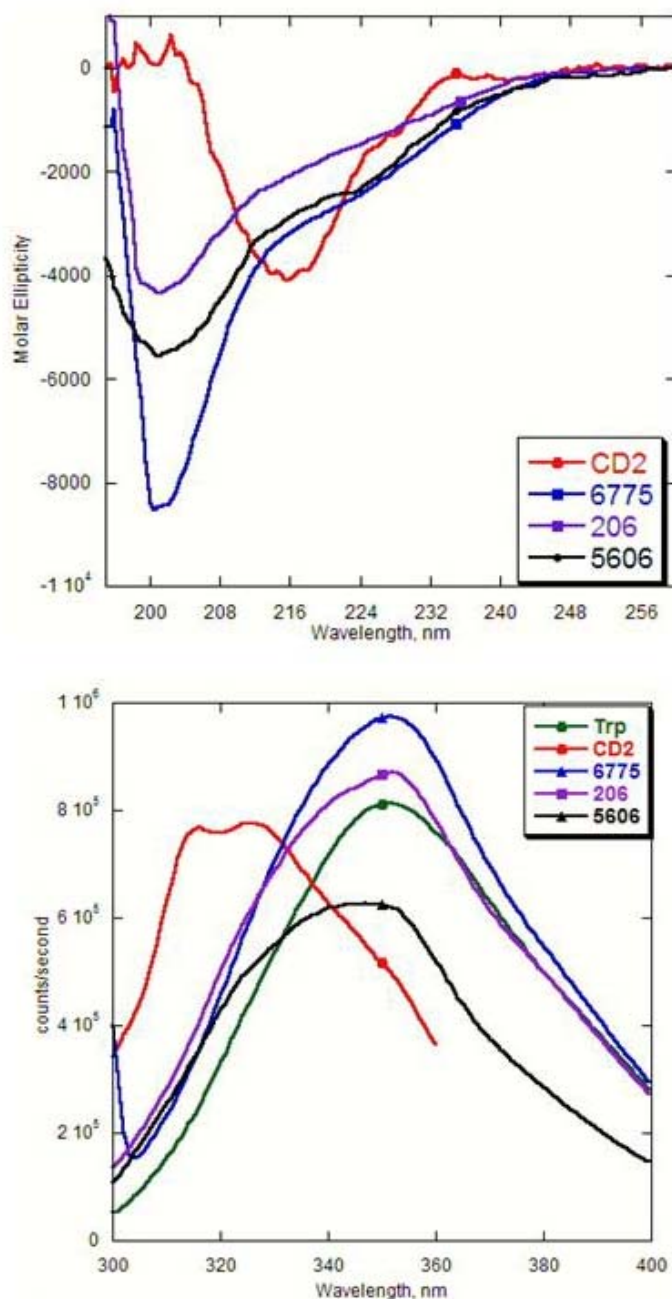


Figure 4.5. Conformational analyses of the first generation proteins. (a) Far UV CD of CD2 and the first generation proteins (3 μ M) (b) Trp fluorescence of CD2, the first generation proteins, and free Trp. The emission spectra excited at 283 nm show that the designed proteins have an alternate conformation compared to wild type CD2 in 10 mM Tris pH 7.4.

4.2.4 The effect of metals on the conformation of the first generation proteins

The conformation of the proteins was examined in the presence of Ca^{2+} and its analogues Tb^{3+} and La^{3+} . Several natural Ca^{2+} binding proteins including calmodulin undergo a global conformational change upon the addition of Ca^{2+} . Since our first generation proteins has an altered, unstable conformation we wanted to determine whether metal binding would induce the protein to undergo a conformational change to its native state. While a 10 mM concentration of Ca^{2+} was used for the study only 100 μM of Tb^{3+} and La^{3+} was used due to the precipitation of the metal bound protein at higher concentrations of Tb^{3+} and La^{3+} . Since the binding affinities of these metals are strong (**section 4.3**) this concentration was sufficient for saturation. Figures 4.6-4.7 displays the conformation of proteins 206 and 6775 upon the addition of metal. Neither protein undergoes a global conformation change with the addition of metal. However, local changes can be observed, through the change in intensities of the spectra, indicating metal binding. These changes are more profound with La^{3+} and Tb^{3+} binding which agrees with the metal binding data where Tb^{3+} and La^{3+} bind stronger than Ca^{2+} .

4.3 Metal binding studies of the first generation proteins

To determine the metal binding affinity of all three designed proteins three different methods were used. Far UV CD, FRET, and ICP-MS were used to measure the binding affinities of the three proteins to Ca^{2+} and its analogs Tb^{3+} and La^{3+} . All of the binding affinities for 6775 were obtained by Dr. Wei Yang using fluorescence resonance energy transfer and ICP-MS (Yang et al., 2003).

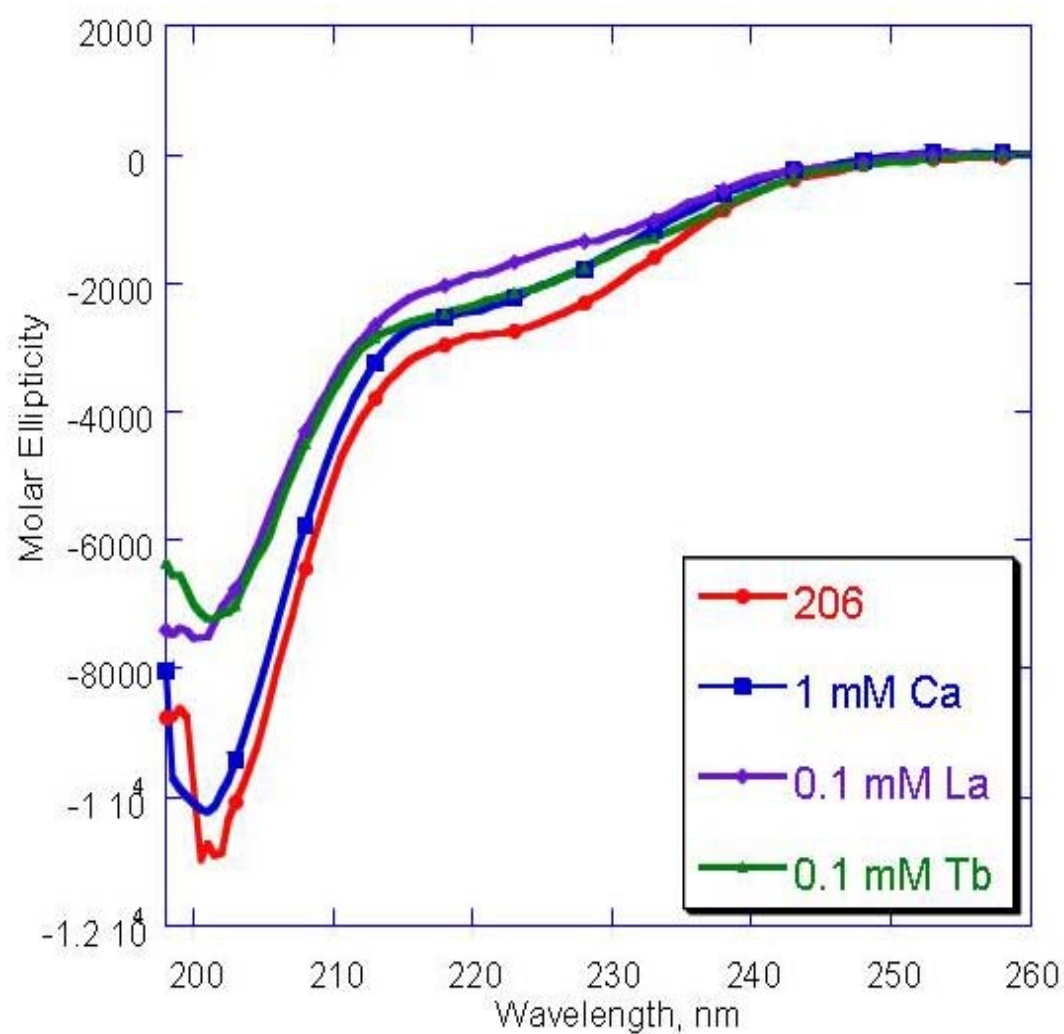


Figure 4.6. Conformational analysis of 206 in the presence of different metal ions. Far UV CD spectra of the first generation protein 206 (3 μ M in 10 mM Tris pH 7.4) upon metal binding does not display a global conformational change with different metals.

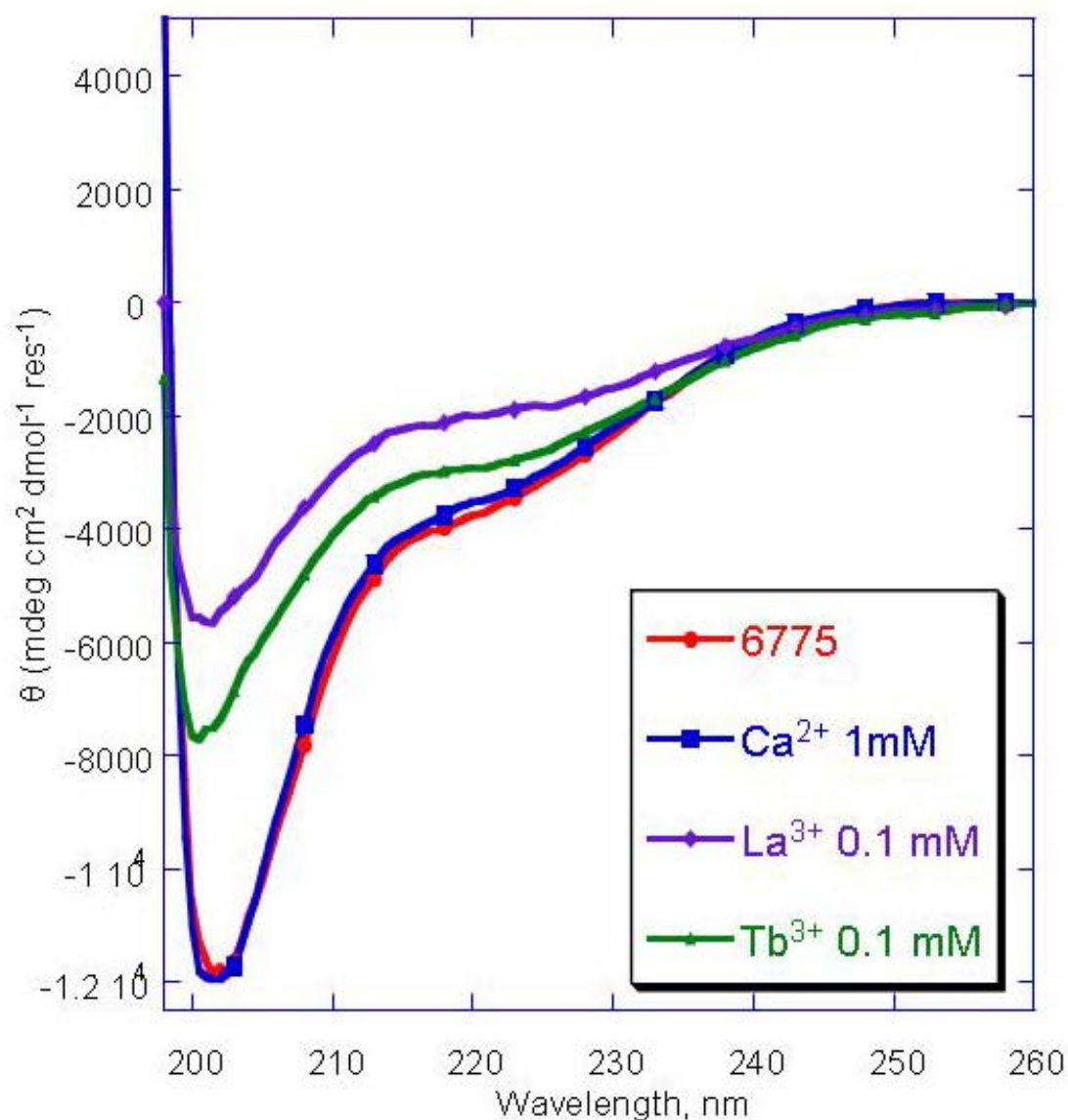


Figure 4.7. Conformational analysis of 6775 in the presence of different metal ions. Far UV CD spectra of the first generation protein 6775 (3 μ M in 10 mM Tris pH 7.4) upon metal binding does not display a global conformational change with different metals.

4.3.1 Far UV CD

The CD signal change was used to determine the binding of Ca^{2+} , Tb^{3+} and La^{3+} to the designed proteins. The experiments were performed in 1 mM MOPS pH 6.9. The protein concentration used for the titration was 3 μM . Stock solutions of metal also contained 3 μM of protein so that the protein concentration remained constant throughout the experiment. In this manner the only variable that changed was the metal concentration. This made it possible to calculate the binding affinity using equation 3.2. The addition of metal to the protein results in a decrease in the signal intensity of the CD signal at 205 nm. The change in signal intensity for the designed protein 5606 binding to Ca^{2+} is shown in Figure 4.8. Even though the change in signal is minimal, a time averaging method used (section 2.6) allows for data resolution. To increase the sensitivity, 10-min collection of 1000 points at 205 nm was averaged at each metal concentration. This time averaging method gives the resolution up to 0.01 mdeg. Binding constants were obtained for three metals by measuring the fractional change of the CD signal at 205 nm. The total protein and metal concentration was taken into account in calculating the binding affinity. The metal binding affinities obtained for the three proteins, 206, 5606, and 6775 were 53, 33, and 55 μM , respectively.

4.3.2 Fluorescence resonance energy transfer (FRET)

Fluorescence resonance energy transfer (FRET) is a quantum mechanical phenomenon that occurs between a fluorescence donor and a fluorescence acceptor in close proximity if the emission spectrum of the donor overlaps with the excitation spectrum of the acceptor (De Angelis, 1999). In FRET, a donor fluorophor is excited by

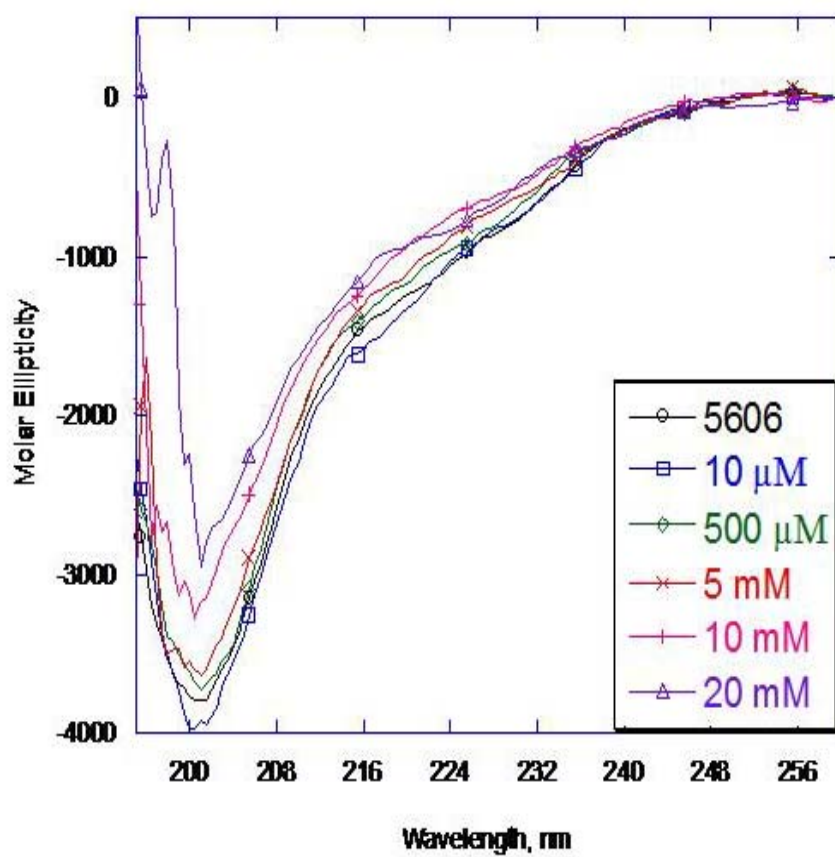


Figure 4.8 Far UV CD of 5606 with increasing additions of Ca^{2+} . Upon the addition of metal to 3 μM of the protein in 1 mM MOPS pH 6.8, the CD signal decreases.

incident light, and if an acceptor is in close proximity, the excited state energy from the donor can be transferred (Selvin, 2000). The efficiency of this process depends on the inverse sixth-distance between donor and acceptor:

$$E = 1/(1+(R/R_0)^6) \quad \text{Eq. 4.1}$$

where R_0 , typically 20 -60 Å, is the distance at which half of the energy is transferred (Forster, 1959). Although FRET is typically used for distance measurements it has been a valuable tool in monitoring metal binding using lanthanides (Snyder et al., 1990). CD2 has two Trp residues and two Tyr residues which can be used as donors. Each designed protein used for this study is within the acceptable distance range to one of these aromatic residues.

Tb^{3+} , a natural fluorophor, has similar ionic radii and metal coordination chemistry with Ca^{2+} , 1.00 and 0.92 Å, respectively, and has been commonly used to probe Ca^{2+} binding (Drake et al., 1997; Horrocks, 1993). Tb^{3+} bound to the protein was selectively excited by energy transfer from protein tryptophans and emission of Tb^{3+} was observed at 545 nm. This emission is further enhanced as increasing amounts of Tb^{3+} are added. Alternatively, the metal binding affinity of the Ca^{2+} analogue La^{3+} was also monitored by FRET using a competition method. The protein was first bound to Tb^{3+} and increasing concentrations of La^{3+} was added. As La^{3+} competed with Tb^{3+} for the binding pocket the Tb^{3+} emission at 545 nm decreased. The direct binding of Tb^{3+} was established by demonstrating FRET between Trp emission and Tb^{3+} fluorescence. Close proximity of a Trp residue to a metal binding site allows for the observation of Forster energy transfer from aromatic residues to bound Tb^{3+} ions. To minimize metal

precipitation a higher concentration of MOPS was used than for CD. The increase in buffer concentration to 100 mM MOPS better maintains the pH at 6.9 which is an optimal pH for Tb^{3+} studies. At this pH the precipitation of metal is minimal at high concentrations of Tb^{3+} .

First, to examine whether our designed proteins bound Tb^{3+} , an increasing protein titration was performed. The addition of protein to the Tb^{3+} solution results in a large enhancement of Tb^{3+} fluorescence at 545 nm with excitation at 282 nm (Figure 4.9). This phenomenon is not observed with wild type CD2 (Yang et al., 2003). Next, a Tb^{3+} titration was performed so as to obtain the binding constants for the three designed proteins. An increase in the fluorescence intensity was observed at 545 nm indicating Tb^{3+} binding to the protein (Figure 4.10a). The K_d 's attained for Tb^{3+} were 11.6, 2.8, and 2.3 μM for proteins 206, 5606, and 6775, respectively, using freshly purified proteins (Figure 4.10b and summarized in Table 4.3). These experiments were performed in triplicate. These metal binding affinities were determined by measuring the fractional change of the fluorescence signal from 530 - 560 nm. Additionally the FRET method can be used as a competition method to determine the binding affinity of other metals that can bind to the Ca^{2+} binding site. The protein is first bound to Tb^{3+} and then increasing amounts of the metal of interest is added to the sample. If the metal competes with Tb^{3+} for the binding pocket, a decrease in the fluorescence emission of Tb^{3+} at 545 nm is observed. This method was utilized to determine the La^{3+} binding affinity of sites 206 and 5606 which is 9 ± 8 and 3 ± 0.6 μM , respectively in 10 mM MOPS pH 6.8 (Figure 4.11).

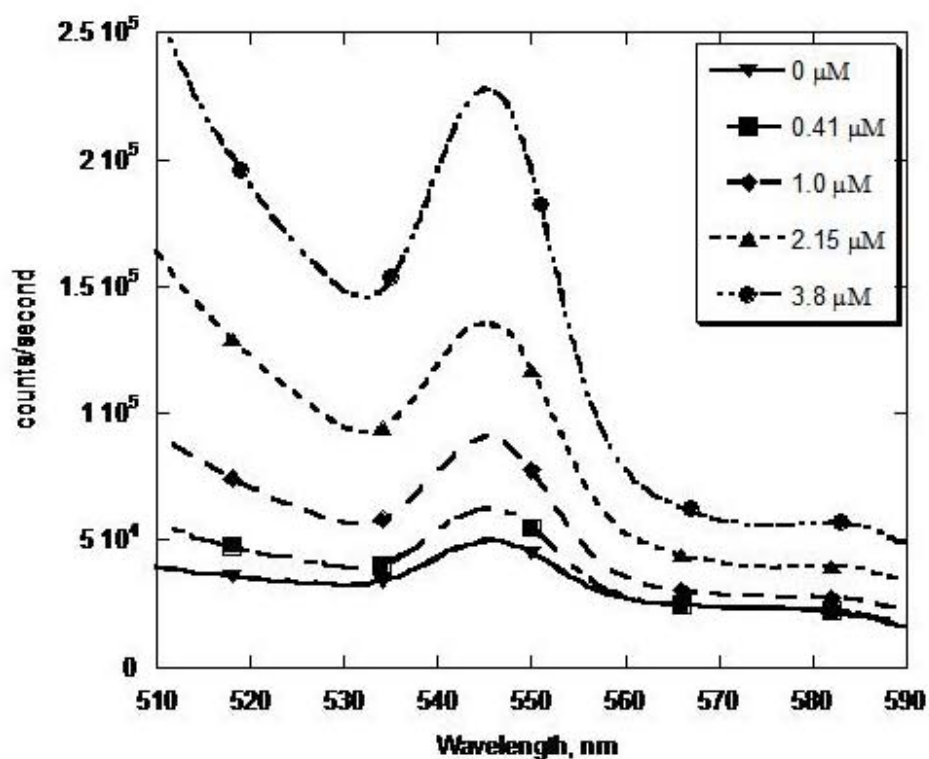


Figure 4.9. Enhancement of Tb^{3+} fluorescence emission upon the addition of protein. The fluorescence emission of Tb^{3+} at 545 nm increases as increasing amounts of the protein 206 is added indicating the metal is binding to the protein. Excitation was at 283 nm.

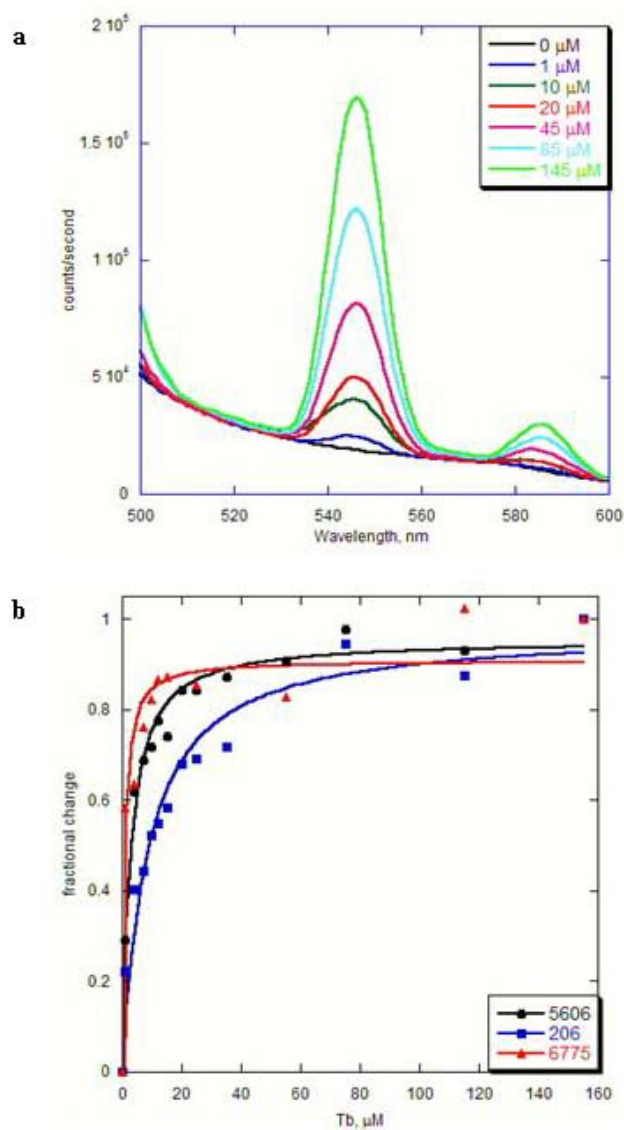


Figure 4.10. Tb^{3+} binding affinity of the first generation designed proteins. (a) Enhancement of the fluorescence intensity of Tb^{3+} at 545 nm as increasing amounts of Tb^{3+} are added to the designed protein 5606. (b) The fit of the fractional change of the three designed proteins 5606, 206, and 6775 gives K_d 's of 3 ± 0.8 , 12 ± 2 , and $2 \pm 0.3 \mu\text{M}$, respectively in 10 mM MOPS pH 6.8 in triplicate.

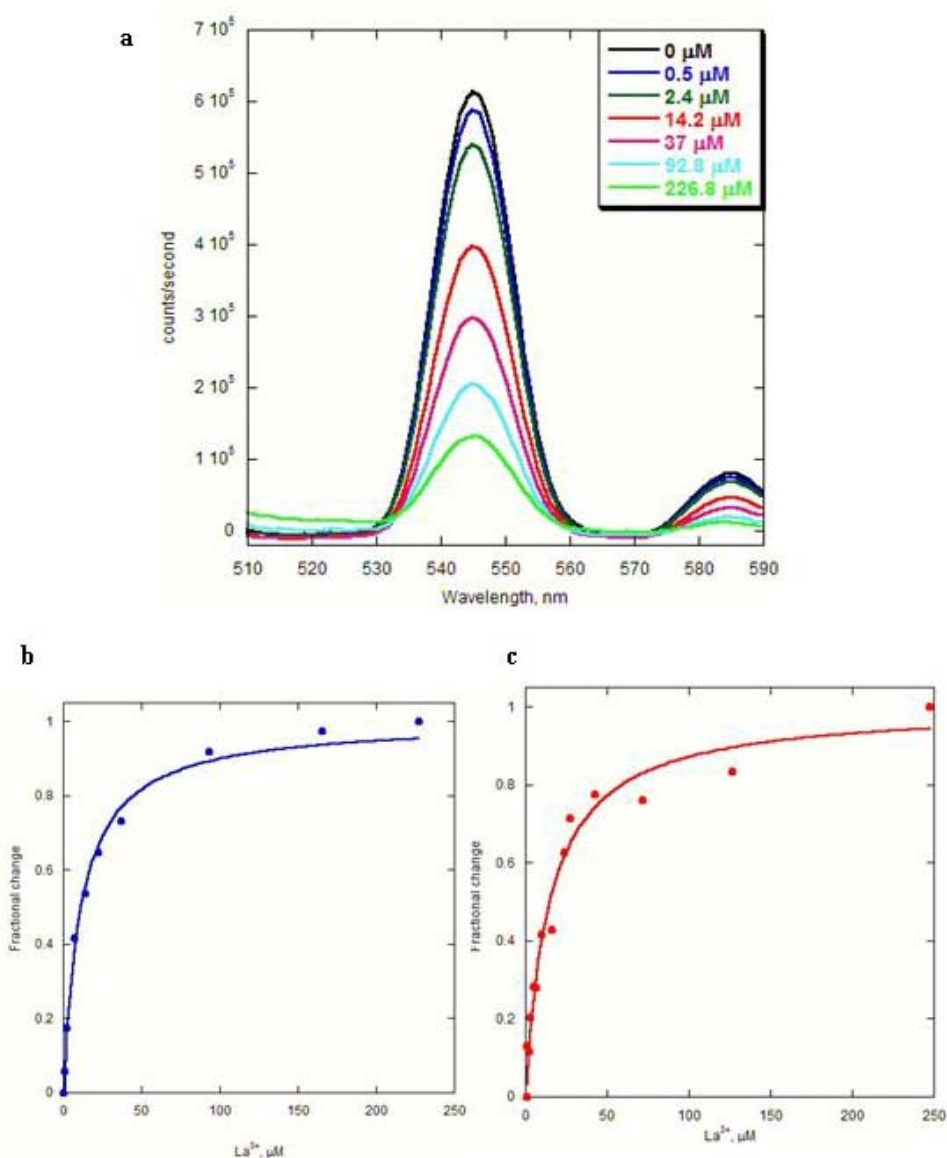


Figure 4.11. La^{3+} binding affinity of the first generation variants. (a) Fluorescence emission decrease of Tb^{3+} emission at 545 nm as increasing additions of La^{3+} are added to designed protein 206. (b) Fractional change of the fluorescence decrease for protein 206 yields a K_d of $9 \pm 8 \mu\text{M}$. (c) Fractional change of the fluorescence decrease for protein 5606 yields a K_d of $3 \pm 0.6 \mu\text{M}$.

4.3.3 Inductively coupled plasma mass spectrometry (ICP-MS)

ICP-MS is a highly sensitive technique for the fast, precise and accurate determination of trace element concentrations in a wide variety of liquid and solid materials (Frei, 2005). For ICP-MS analysis, samples are dispersed into a stream of argon gas and carried to an inductively coupled plasma where they are ionized. The ions are then separated according to their mass and charge ratio using a mass spectrometer and subsequently counted. For liquids, for example water samples and solid samples brought into solution, this is achieved by spraying the sample into the argon carrier gas using a nebulizer. ICP-MS is a very powerful tool for trace (ppb-ppm) elemental analysis.

ICP-MS has become the technique of choice for the determination of elements in a wide range of samples at concentrations in the ng l^{-1} to mg l^{-1} range. The versatility and reliability of the technique in terms of element specificity and sensitivity make it ideally suitable use as a chromatographic detector. The detection technique has several inherent advantages over other possible methods of detection including a wide linear dynamic range, low limits of detection, high speed of analysis, multi-element capability, simple spectra and the ability to perform isotopic analysis.

An equilibrium dialysis method was applied for the metal binding studies of the first generation proteins. All dialyses bags used were boiled in 1 mM EGTA to remove any background Ca^{2+} . The samples were dialyzed against 1 mM Tris, 5 μM Ca^{2+} , pH 6.8 for 36 hr with 3 changes of buffer. The concentrations of proteins 206, 5606, and 6775 were 14.8, 10.0, and 31.4 μM , respectively. The free Ca^{2+} , free protein, and bound Ca^{2+} -P concentrations were measured using ICP-MS by Dr. Russell Malchow in the Department

of Geology at Georgia State University. The calculated Ca^{2+} binding affinities of the three designed proteins, 206, 5606, and 6775 was 34 ± 3 , 29 ± 3 , and 76 ± 7 μM , respectively.

4.3.4 Metal selectivity

In vivo Ca^{2+} binding proteins have selectivity for Ca^{2+} over other physiologically relevant metals including Mg^{2+} and Na^+ (Williams, 2002). To examine whether the designed proteins exhibited the same selectivity FRET experiments were employed. Each protein (3 μM) was first incubated with 30 μM of Tb^{3+} for 15 min to ensure metal binding. From this single stock, four different samples (2 mL) were made up with varying concentrations of metal, La^{3+} (100 μM), Ca^{2+} (10 mM), Mg^{2+} (10 mM), and Na^+ (10 mM). In addition a sample of each protein with Tb^{3+} alone was also used. The fluorescence emission was observed for each metal sample and compared to the emission of the protein sample with Tb^{3+} alone. If the metal were competing with Tb^{3+} for the binding pocket a decrease in the fluorescence emission would be observed compared to the protein- Tb^{3+} sample. If the metal does not bind to the protein then no change in the fluorescence emission would be detected compared to the protein- Tb^{3+} sample. Both 206 and 6775 exhibited significant selectivity for Ca^{2+} and La^{3+} compared to Mg^{2+} and Na^+ (Figure 4.12a, b). For both proteins 100 μM La^{3+} competed better for the binding pocket than 10 mM Ca^{2+} , which is indicative of the stronger binding affinity of the proteins for La^{3+} as compared to Ca^{2+} . Protein 5606 also exhibited selectivity but to a lesser extent (Figure 4.12c). This data shows the designed proteins resemble natural Ca^{2+} binding proteins with selectivity for Ca^{2+} over Mg^{2+} and Na^+ .

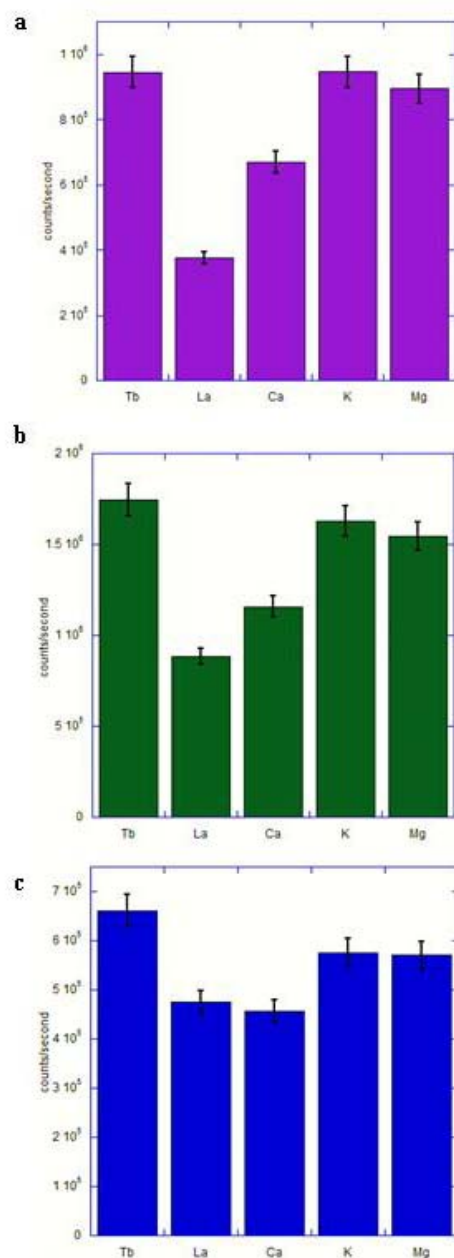


Figure 4.12. Selectivity of the first generation designed proteins. The changes in Tb³⁺ fluorescence emission at 545 nm of proteins 206 (a), 6775 (b), and 5606 (c) all with 3 μ M concentrations upon the addition of metals 100 μ M La³⁺, 10 mM Ca²⁺, 10 mM K⁺, and 10 mM Mg²⁺ in 10 mM MOPS pH 6.8.

4.4 Comparison of metal binding affinities of the three designed proteins

Table 4.3 shows the metal binding affinities obtained for all 3 first generation proteins. For the determination of Ca^{2+} binding affinities, two methods, far UV CD and ICP-MS were applied. There is good correlation between the two methods. Tb^{3+} and La^{3+} bound to all three proteins with a stronger affinity than Ca^{2+} . This was expected to the +3 charge of those metals compare to the +2 charge of calcium. The Ca^{2+} binding affinities of the three proteins were similar to and in some case stronger than natural extracellular Ca^{2+} binding proteins. Cadherin, which has a similar β -sheet architecture as our host protein CD2, has a Ca^{2+} binding affinity of 160 μM (Yang et al., 2000d). In addition, the designed proteins have stronger affinities than other extracellular proteins such as integrin, which has several Ca^{2+} binding sites with average affinities of 200 – 300 μM (Rivas and Gonzalez-Rodriguez, 1991). The extracellular galactose binding protein has a stronger affinity than our designed proteins (2 μM) (Vyas et al., 1989). In contrast to the intracellular Ca^{2+} binding protein parvalbumin has a strong Ca^{2+} binding affinity of 11.0 nM (Eberhard and Erne, 1994). The concentration of Ca^{2+} in the extracellular matrix is higher than in the cytoplasm (2 mM versus 2 μM) (Vogel, 2002), therefore the binding affinities of extracellular proteins do not have to be as strong as for intracellular proteins. However, our designed proteins do have Ca^{2+} affinities that are useful for studies performed in the extracellular matrix.

The binding affinities of all three proteins are within the same order of magnitude. This is puzzling because each site is in a different electrostatic environment. The electrostatic environments around sites 206, 5606, and 6775 are 68.15, 49.44, and -41.86

Designed Protein	Ca ²⁺ K _d (μ M)	Tb ³⁺ K _d (μ M)	La ³⁺ K _d (μ M)
206	34 \pm 3 ^a 52 \pm 9 ^b	12 \pm 2 ^c	9 \pm 8 ^c
5606	29 \pm 3 ^a 33 \pm 3 ^b	3 \pm 0.8 ^c	3 \pm 0.6 ^c
6775	76 \pm 7 ^a 55 \pm 7 ^b	2 \pm 0.3 ^c	3 \pm 0.2 ^c

^a1 mM MOPS, pH 6.8, CD

^b10 mM TRIS, pH 7.4, ICP-MS

^c10 mM MOPS, pH 6.8, FRET

Table 4.3 Metal binding affinities of the first generation designed proteins.

kcal/mol, respectively. The similarities in binding affinity may be due to the altered conformation of the proteins. The hydrophobic surface of the proteins may be exposed to solvent and the metals may be sticking to this surface as well as binding to the designed ligands.

4.5 Reducing protein degradation by protease inhibitors

Mass spectrometry analysis of the designed proteins reveals the proteins undergo significant degradation into small fragments even when stored at low temperatures (Figure 4.13). The partially unfolded conformation of the proteins may make them susceptible to protease cleavage. In the protein purification process the serine protease thrombin is used to cleave the fusion protein. Even though the designed proteins are purified using gel filtration chromatography thrombin and other proteases may still be present and is continually cleaving the protein. In order to reduce the degradation of protein protease inhibitors had to be utilized. There are several inhibitors available and therefore the use of inhibitors had to be optimized. The most successful concentration of inhibitors as well as the effectiveness of mixtures of different inhibitors had to be established.

The protease inhibitors studied were AEBSF (Calbiochem) and D-Phe-Pro-Arg-chloromethylketone, HCl (PPACK) (Calbiochem). AEBSF is an irreversible inhibitor of serine proteases (Markwardt et al., 1973). It reacts covalently with a component of the active site (Lawson et al., 1982). The suggested working concentration of AEBSF is < 1 mM. Hirudin is a potent non-covalent inhibitor of thrombin (Cannon, 1995). PPACK is a potent and selective inhibitor of thrombin (Kettner and Shaw, 1979).

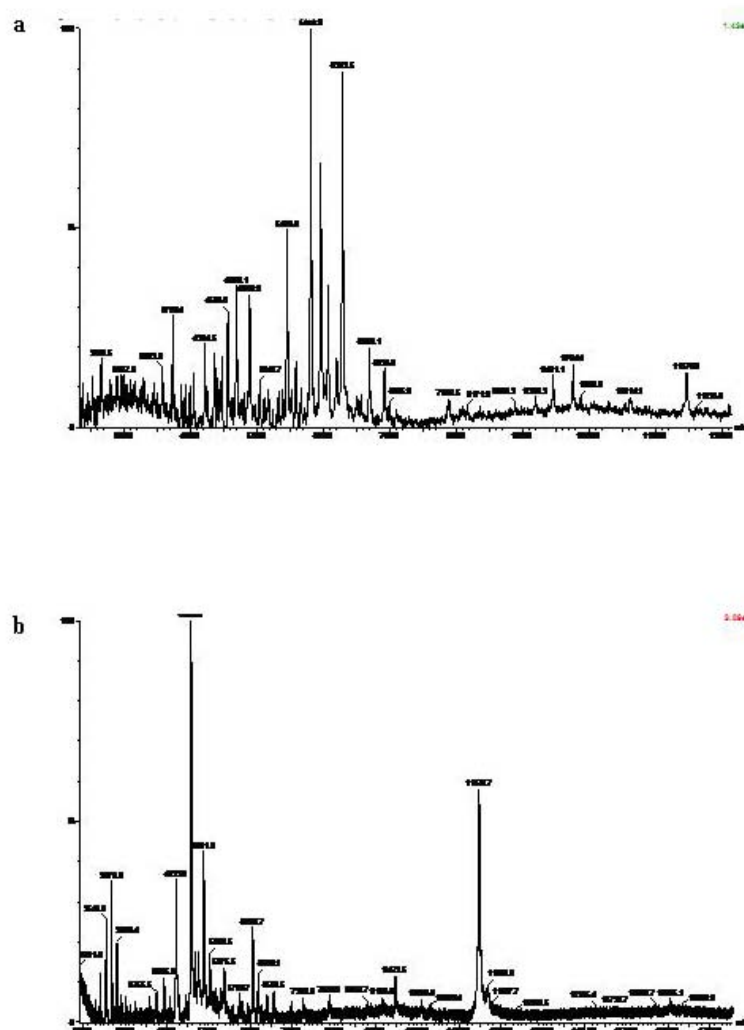


Figure 4.13. Cleavage of the first generation proteins. MALDI mass spectrometry analysis of designed proteins 206 (a) and 5606 (b) show the degradation of the proteins with small molecular weight fragments through proteolytic cleavage.

The protein chosen to do this study with was 6775 because of the ability to obtain the purified protein by FPLC in high yields. In addition previous work with this protein has shown it to be more stable than the other designed proteins. The metal chelators EDTA and EGTA were also examined because the presence of metal ions has been shown to activate proteases. Both EDTA and EGTA are reversible metalloprotease inhibitors (Powers, 1986; Scaloni et al., 1992).

In the first experiment the levels of protein degradation were monitored over a 48 hr period. The 6775 sample (20 μ M) contained \sim 50 μ M of the inhibitor AEBSF. As shown in Figure 4.14a, the AEBSF inhibitor was sufficient to prevent significant degradation in the samples without thrombin. The other inhibitors were adequate in prevent complete degradation of the protein. A more significant degradation was seen in the sample containing thrombin, Figure 4.14b, indicating that possibly a higher concentration ($> 100 \mu$ M) of inhibitor is needed to inhibit thrombin cleavage. A literature search of AEBSF showed that its working concentration is at least 100 μ M. The experiment was repeated with higher concentrations of AEBSF and with the inhibitor PPACK. This time the experiment was monitored for 72 hr with aliquots taken at 48 and 72 hr. The resulting gels, Figures 4.15 and 4.16, show that PPACK is not effective in inhibiting the degradation of our protein. Once again AEBSF was very effective in stabilizing our protein indicating that higher yields of AEBSF are required for successful inhibition of degradation for our designed proteins. In addition, EDTA was shown to be a good inhibitor while EGTA was not effective. EDTA has a stronger affinity for the

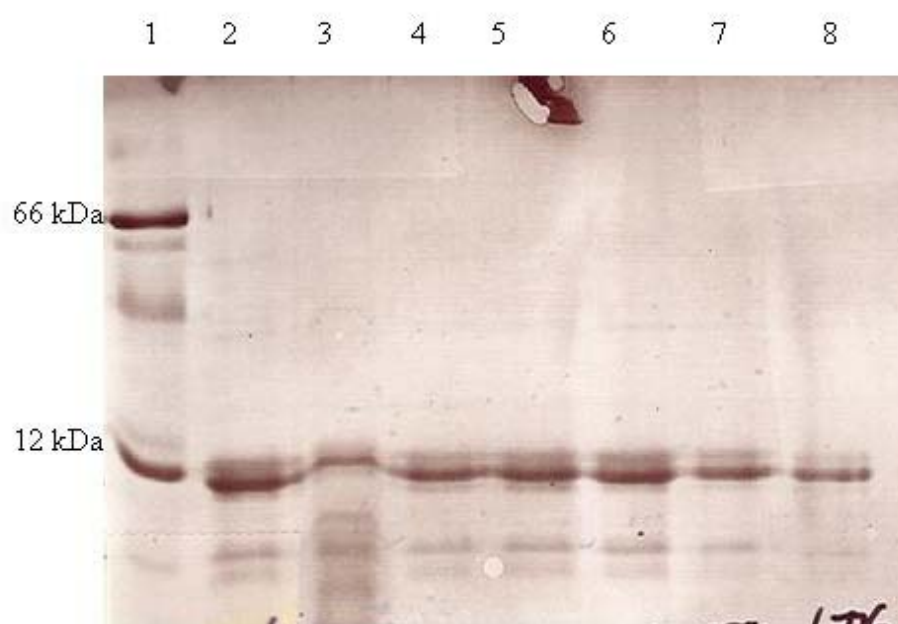


Figure 4.14a. Optimization of protease inhibitors. SDS-PAGE gel analysis of 6775 with different inhibitors after 48 hr.

Lane 1: marker
Lane 2: 6775
Lane 3: PPACK inhibitor
Lane 4: 50 μ M AEBSF
Lane 5: 100 μ M AEBSF
Lane 6: 1 mM EDTA
Lane 7: 1 mM EGTA
Lane 8: 0 hr (starting point)

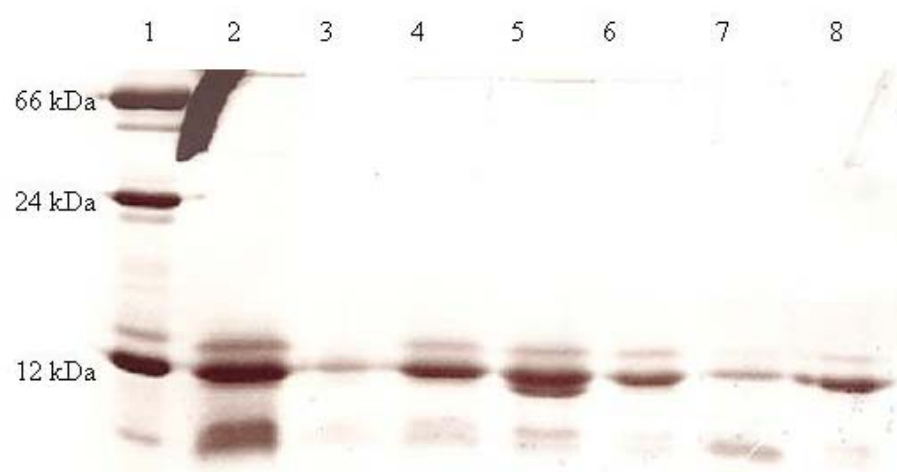


Figure 4.14b. Optimization of protease inhibitors with thrombin. SDS-PAGE gel analysis of 6775 with thrombin and different inhibitors after 48 hr.

Lane 1: marker
Lane 2: 6775
Lane 3: PPACK
Lane 4: 50 μ M AEBSF
Lane 5: 100 μ M AEBSF
Lane 6: 1 mM EDTA
Lane 7: 1 mM EGTA
Lane 8: 6775 0 hr (starting point)

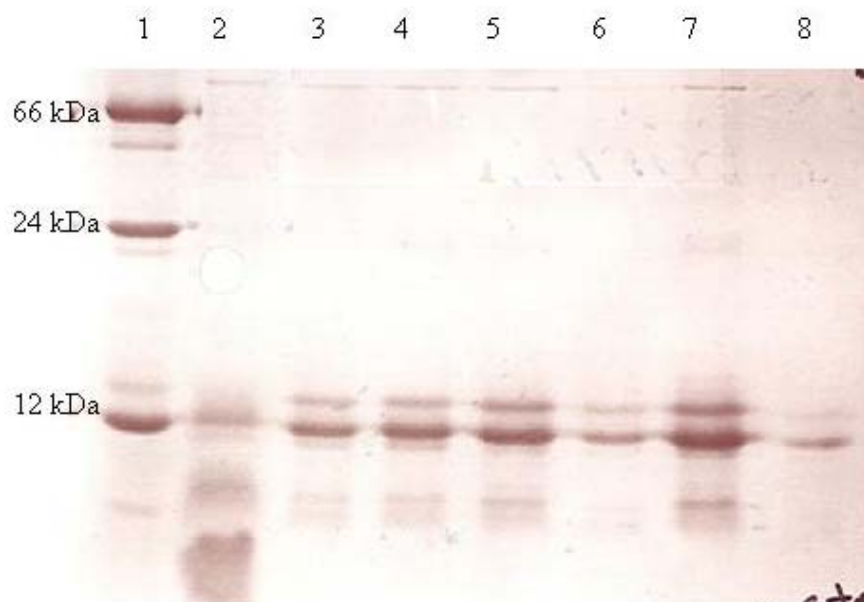


Figure 4.15. SDS-PAGE analysis of 6775 with different inhibitors at 72 hr.
 Lane assignments are as follows: (1) marker (2) PPAK (3) 50 μ M AEBSF (4) 100 μ M AEBSF (5) EDTA (6) EGTA (7) 6775 (8) 6775 0 hr.

Lane 1: marker
 Lane 2: PPAK
 Lane 3: 50 μ M AEBSF
 Lane 4: 100 μ M AEBSF
 Lane 5: 1 mM EDTA
 Lane 6: 1 mM EGTA
 Lane 7: 6775 72 hr
 Lane 8: 6775 0 hr

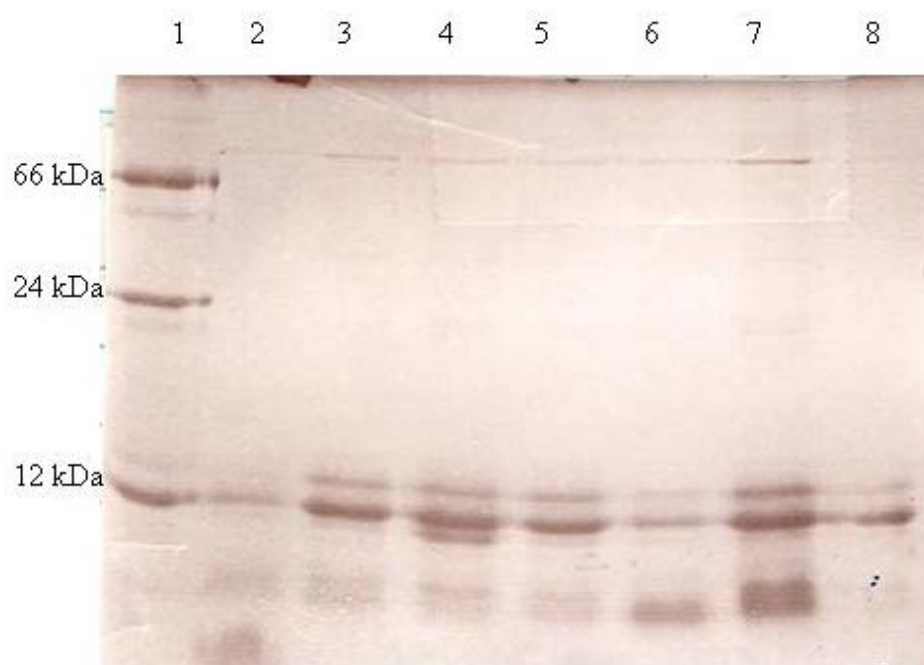


Figure 4.16. SDS-PAGE analysis of 6775 with thrombin and different inhibitors at 72 hr.

Lane 1: marker
Lane 2: PPACK
Lane 3: 50 μ M AEBSF
Lane 4: 100 μ M AEBSF
Lane 5: 1 mM EDTA
Lane 6: 1 mM EGTA
Lane 7: 6775 72 hr
Lane 8: 6775 0 hr

metal ions Zn^{2+} and Mg^{2+} than does EGTA. The specific ions may be responsible for activation of proteases which makes EDTA a better inhibitor than EGTA.

4.6 Rationale for the N80/89 mutant

To examine whether less aggressive mutations would allow the protein to fold properly a variant of 6775 was engineered. The designed protein 6775 mutates two hydrophobic residues V80 and L89 to the charged residues, Glu and Asp, respectively. These are aggressive mutations to introduce charge residues into the interior of the protein and may be the origin of the altered conformation of this protein. In order to examine the role of these two mutations in the conformation of 6775, a variant of this protein, N80/89 was engineered. For this protein, residues 80 and 89 were mutated to the neutral residue Asn which has the capability to bind Ca^{2+} . All of the other mutations that form the Ca^{2+} binding site, F21E, K91D, and V78N remained intact.

The protein was engineered using PCR with the *Pfu* DNA polymerase. The PCR was performed in two separate rounds with the N80 and N89 mutation introduced separately. The protein was expressed in the pGEX-2T vector and purified using affinity chromatography. The purity of the protein was confirmed with mass spectrometry analysis. From a 4 L expression, 1.0 mg of protein was recovered, which is a much lower yield compared to wild type CD2.

4.7 Conformational analyses of N80/89

The conformation of variant N80/89 was examined using far UV CD and Trp fluorescence. The CD spectra indicate the protein has an alternate conformation with respect to wild type CD2 with a negative maximum at 205 nm (Figure 4.17a). This conformation is similar to the 6775 and therefore the introduction of the less aggressive mutations has not altered the conformation. The conformational analysis using Trp fluorescence shows the emission of this protein is 356 nm (Figure 4.17b) similar to free Trp indicating the tertiary packing of the protein has been perturbed. Therefore, either the polar residue Asn in the interior of the proteins is generating conformational changes as well or the 80 and 89 residues are not contributing to the altered conformation.

4.8 Folding Properties of CD2

In an effort to understand the origin of the altered conformation of the designed proteins previous research on the structural properties of CD2 were examined. In a study by Lorch et al. a series of core mutations were introduced into CD2 to examine their influence on the rapidly formed intermediate state (I-state) (Lorch et al., 1999). This I-state transiently accumulates in the folding reaction of CD2 (Parker and Clarke, 1997b). The study showed the rate limiting transition state of the folding reaction is attained by the formation of a tightly localized hydrophobic nucleus which includes residues V30, I18, and V78. For designed protein 6775, V78 has been mutated to N, consequently disrupting the hydrophobic core of the protein. This mutation is the basis for the altered conformation of this protein. For this reason the mutation of residues of 80 and 89 to Asn has no effect on the conformation. Other residues involved in the folding nucleus in the

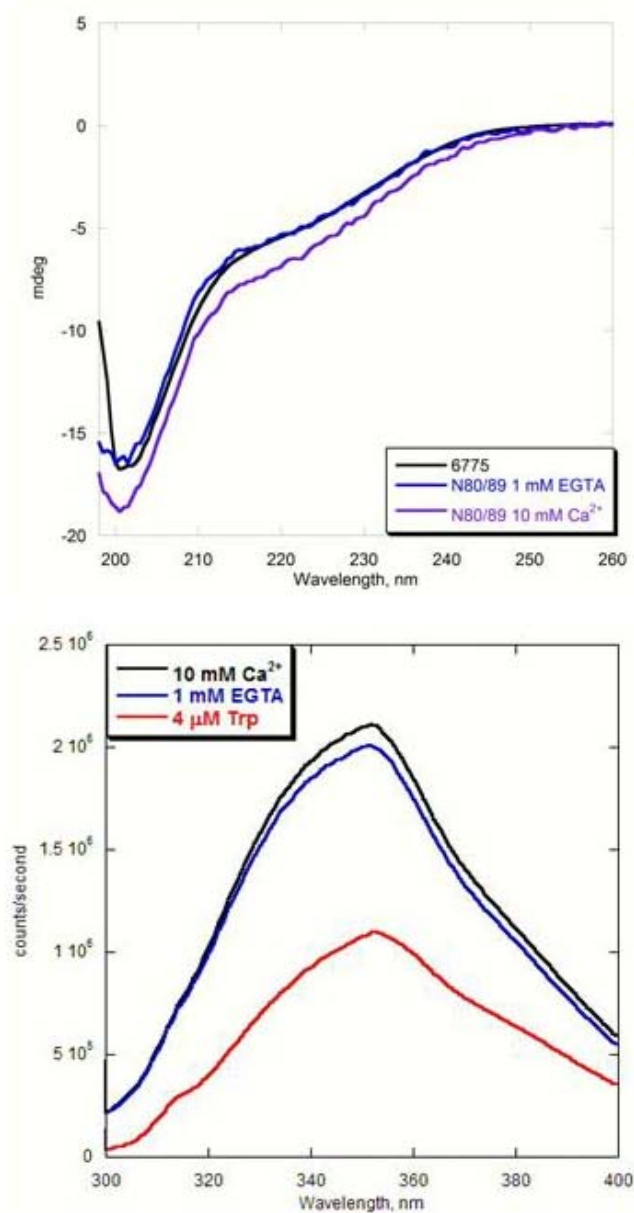


Figure 4.17. Conformational analyses of N80/89. Conformational analyses of N80/89 (3 μM) using far UV CD (a) and Trp fluorescence (excitation 283 nm) (b) in 10 mM Tris pH 7.4.

transition state include L16, A40, W32, and L95. In the designed protein 5606, L95 was mutated to Asp, accounting for the altered conformation of this protein. Even though no residues that were mutated to form the binding site in protein 206 are part of the folding nucleus, both I65 and L63 are in close proximity to the hydrophobic core (Figure 4.18). The mutation of these hydrophobic residues to polar residues may account for the altered conformation of this protein.

4.9 Rationale for using the pET28aM1 vector to express and purify proteins

Since the first generation proteins were not very stable and degrade easily, a rapid method for purification was desirable. The purification process for the pGEX-2T vector was lengthy (> than 24 hr). During this time the designed proteins would be degraded. The use of thrombin to cleave the GST fusion protein might be too non-specific for the partially folded proteins.

The pET28aM1 vector is a His tag vector with DNA encoding for an N- and C-terminal His tag. The purification process involves a nickel chelating column where the His tag will bind to nickel. Once excess cell particulate are washed out the column the His tag protein is eluted with an imidazole gradient. The imidazole competes with the nickel for the His tag binding.

Purification of proteins expressed in the pET28aM1 vector is much more rapid. Protein 6775 was sub-cloned into the pET28aM1 vector and the expression was assessed using several different competent cells. Three different strains of competent cells, BL-21, HMS, JM109, were tested for their expression yields in pET28aM1. Expression in the pET28aM1 vector is similar to the pGEX-2T vector except that pET28aM1 is

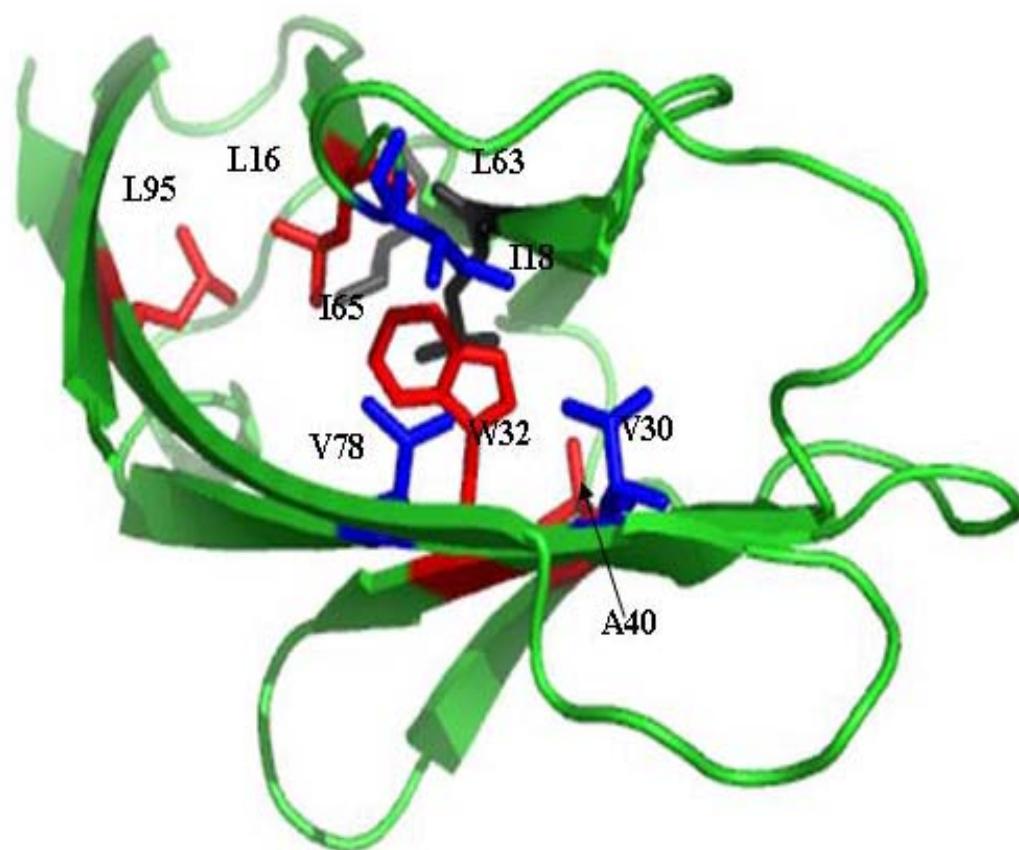


Figure 4.18. Hydrophobic interior of CD2. The hydrophobic nucleus residues are in blue, the folding nucleus residues are in red, and two Ca²⁺ binding ligands of 206 are shown in black.

selective for the antibiotic kanamycin instead of ampicillin. After the OD_{600} reached 0.6 the growth was induced with IPTG for 3 hr. An SDS-PAGE gel of expression of 6775 with different competent cells is shown in Figure 4.19. The expression of 6775 was not very successful. There was no protein expression in JM109 cells and only limited expression was observed in the HMS and BL-21 competent cells.

To further optimize expression, different expression media was utilized. Four different kinds of media, SB, SOB, LB, and SV media were employed for protein expression. The protein was expressed using BL-21 competent cells since there was a limited amount of expression with that competent cell. The expression in all four media was limited (Figure 4.20). For SOB media, a strong expression band was observed at 0.5 hr but the band decreased as time progressed indicating protein death. This may be due to rapid protein expression because of high concentrations of IPTG (1 mM). Therefore further optimization of protein expression needed to be carried out.

Two different IPTG concentrations, 0.1 and 1 mM, were applied for protein expression of 6775 in pET28aM1 vector with BL-21 competent cells in SOB media. The protein did not express well under either condition (Figure 4.21). This can be contributed to the lack of a well folded fusion protein like GST that may aid in stabilizing the protein in a sufficient amount for expression. The presence of a fusion partner for expression in a vector that permits rapid purification may be the criteria for good protein yields of the first generation designed proteins.

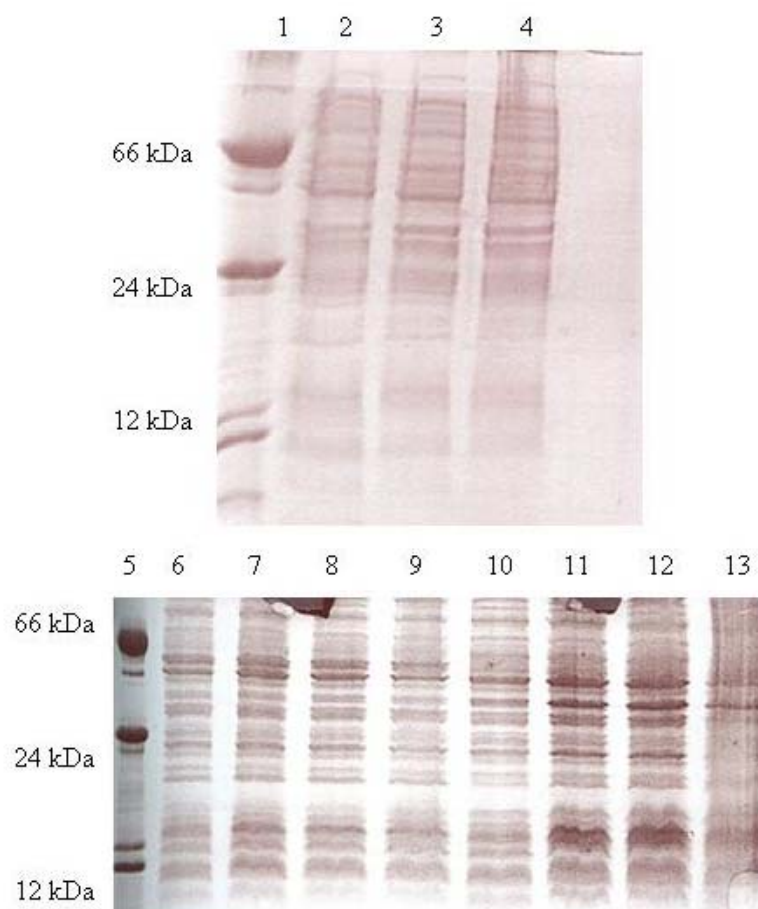


Figure 4.19. Optimization of competent cells for expression in pET28aM1. SDS-PAGE analysis of the expression of 6775 in the pET28aM1 vector using the JM109 and competent cells.

Lane 1: marker
 Lane 2: before induction
 Lane 3: JM109 1 hr after induction
 Lane 4: JM109 2 hr after induction
 Lane 5: marker
 Lane 6: HMS before induction
 Lane 7: HMS 1 hr after induction

Lane 8: HMS 2 hr after induction
 Lane 9: HMS 3 hr after induction
 Lane 10: BL-21 before induction
 Lane 11: BL-21 1 hr after induction
 Lane 12: BL-21 2 hr after induction
 Lane 13: BL-21 3 hr after induction

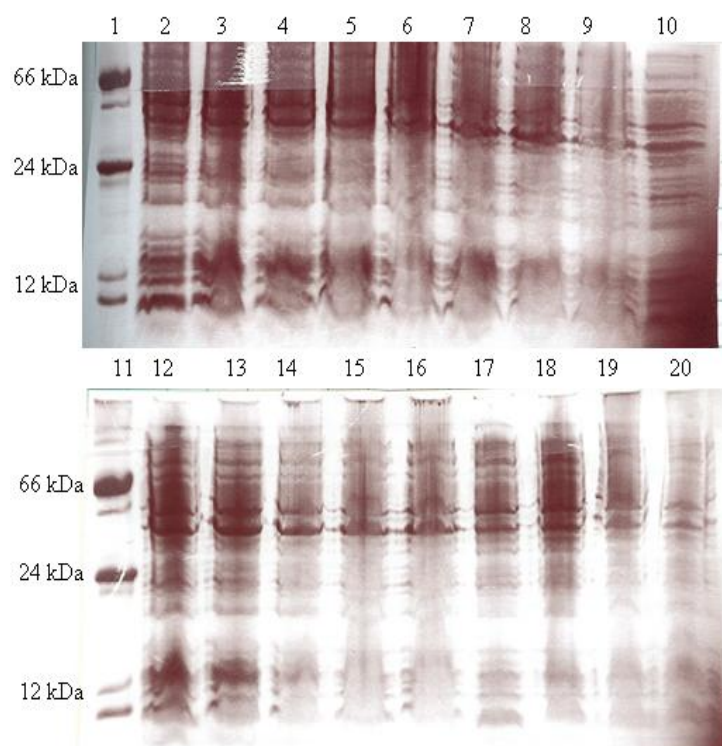


Figure 4.20. Optimization of expression media in the pET28aM1 vector. SDS-PAGE gel of expression of 6775 in the pET28aM1 vector using four different media.

Lane 1: marker	Lane 11: marker
SB medium	Lane 12: 0.5 hr after induction
Lane 2: before induction	Lane 13: 1 hr after induction
Lane 3: 0.5 hr after induction	Lane 14: 2 hr after induction
Lane 4: 1 hr after induction	Lane 15: 2.5 hr after induction
Lane 5: 1.5 hr after induction	Lane 16: 3 hr after induction
Lane 6: 2 hr after induction	LB medium
Lane 7: 3 hr after induction	Lane 17: before induction
Lane 8: 3.5 hr after induction	Lane 18: 0.5 hr after induction
Lane 9: 4 hr of induction	Lane 19: 1 hr after induction
SOB medium	Lane 20: 2 hr after induction
Lane 10: before induction	

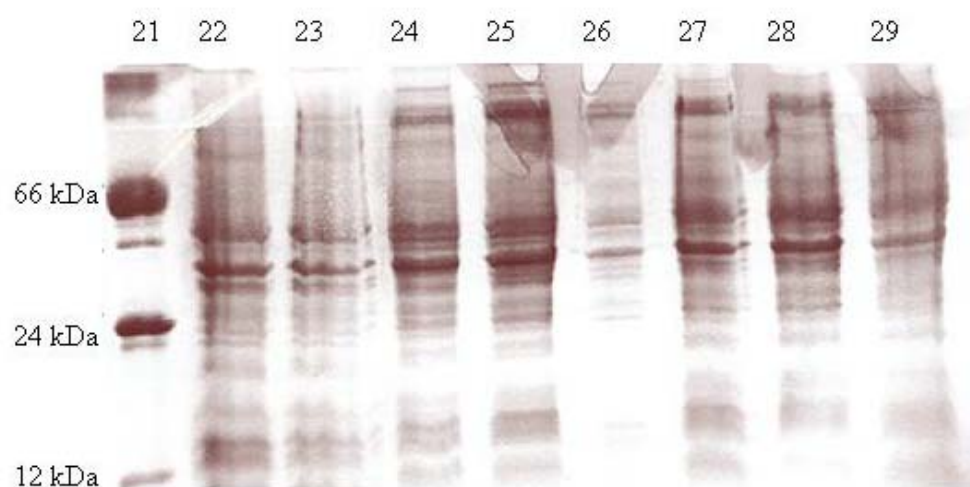


Figure 4.20. continued

Lane 21: marker	Lane 25: 0.5 hr after induction
Lane 22: 2.5 hr after induction	Lane 26: 1.5 hr after induction
Lane 23: 3 hr after induction	Lane 27: 2 hr after induction
SV medium	Lane 28: 2.5 hr after induction
Lane 24: before induction	Lane 29: 3 hr after induction

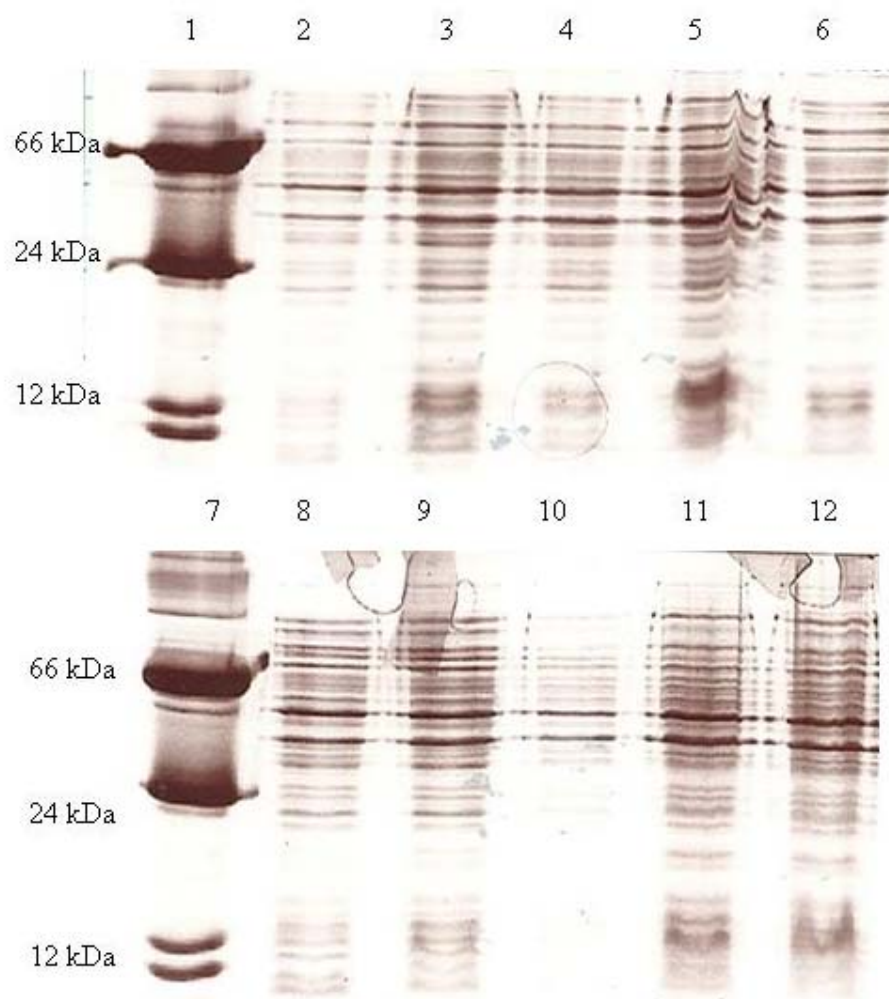


Figure 4.21. Optimization of IPTG concentrations for expression in the pET28aM1 vector. SDS-PAGE gel analysis of the expression of 6775 in the pET28aM1 vector with varying IPTG concentrations.

1 mM IPTG

Lane 1: marker

Lane 2: before induction

Lane 3: 1 hr after induction

Lane 4: 2 hr after induction

Lane 5: 3 hr after induction

Lane 6: 4 hr after induction

0.1 mM IPTG

Lane 7: marker

Lane 8: before induction

Lane 9: 1 hr after induction

Lane 10: 2 hr after induction

Lane 11: 3 hr after induction

Lane 12: 4 hr after induction

4.10 The pET32a vector

As shown in Figure 2.2, the pET32a vector is designed for high level expression of proteins fused with thioredoxin. Thioredoxin is a 109 amino acid protein that aids in folding of its fusion partner during expression. The pET32a vector also contains the encoding for a C-terminal His tag so the nickel chelating column can be used for protein purification. The vector has a T7 promoter and a lac operator. Unlike pET28aM1, pET32a is selective for the antibiotic ampicillin.

4.11 Expression and purification of N80/89 in pET32a

The designed proteins 206, 5606, and N80/89 were sub-cloned into the pET32a vector and their expression was examined. The proteins were expressed using BL-21 competent cells, in LB media, and induced at an OD₆₀₀ of 0.6 with 0.1 mM IPTG. All three proteins expressed well in this vector (Figure 4.22) compared to expression in the pET28aM1 vector. To further optimize the expression of the proteins the IPTG concentration was increased to 0.2 mM. At this IPTG concentration the expression was substantial based on the large bands observed at 32 kDa in the SDS-PAGE gel (Figure 4.23).

The N80/89 protein was purified using the nickel chelating column. The cell pellet lysed using sonication and centrifuged. The filtered supernatant was loaded onto a nickel chelating column using an FPLC. Any excess proteins and cell particulate were washed out of the column. An imidazole gradient was run and the designed protein was eluted out. Figure 4.24 shows the fractions of the protein that were eluted out of the column. Fractions 4-11 contained large amounts of protein. Fraction 6 which contained

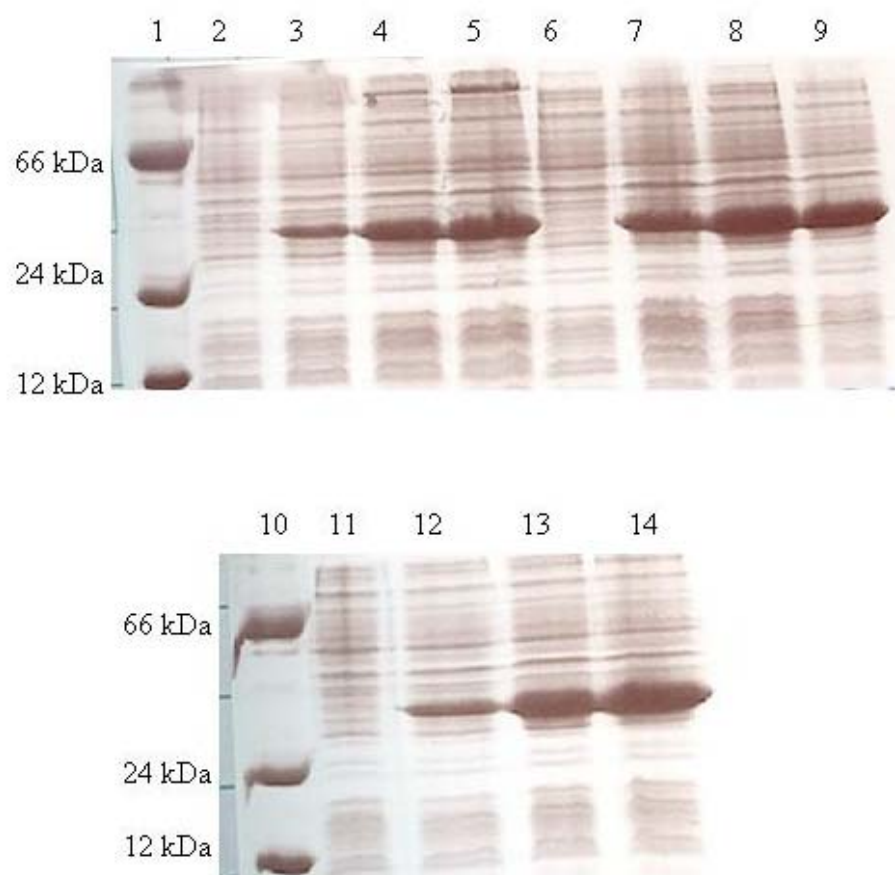


Figure 4.22. Expression in the pET32a vector. SDS-PAGE gel analysis of expression of proteins 206, 5606, and N80/89 in the pET32a vector.

Lane 1: marker	Lane 8: 5606 2 hr after induction
Lane 2: 206 before induction	Lane 9: 5606 3 hr after induction
Lane 3: 206 1 hr after induction	Lane 10: marker
Lane 4: 206 2 hr after induction	Lane 11: N80/89 before induction
Lane 5: 206 3 hr after induction	Lane 12: N80/89 1 hr after induction
Lane 6: 5606 before induction	Lane 13: N80/89 2 hr after induction
Lane 7: 5606 1 hr after induction	Lane 14: N80/89 3 hr after induction

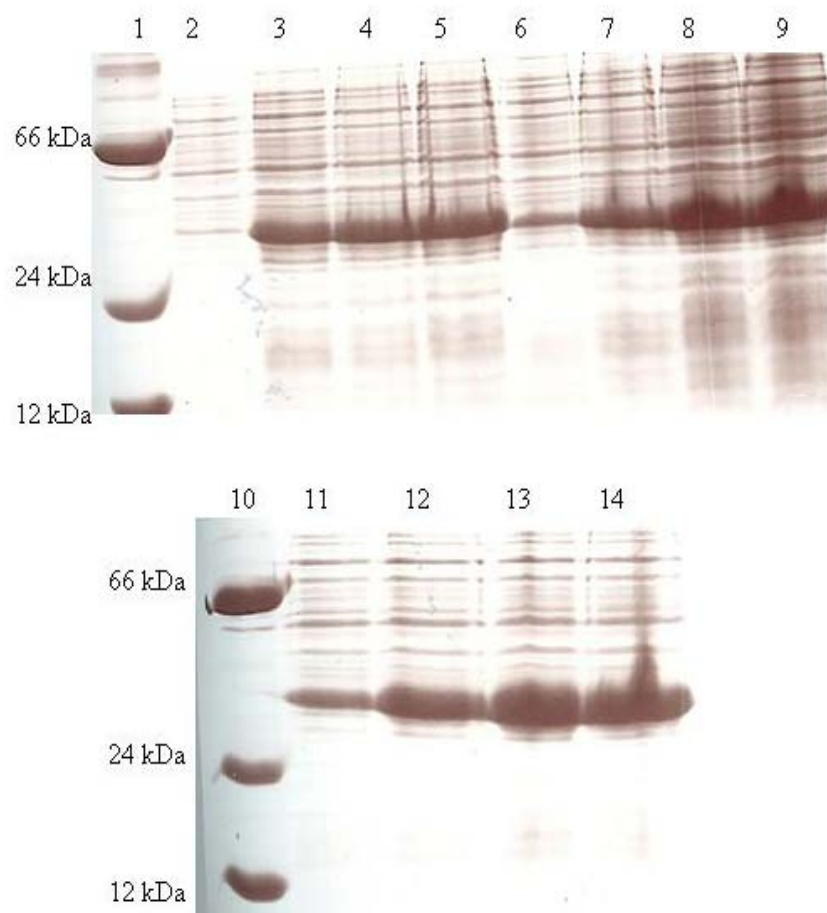


Figure 4.23. Expression in the pET32a vector using 0.2 mM IPTG. SDS-PAGE gel analysis of proteins 205, 2506, and N80/89 in the pET32a vector using 0.2 M IPTG.

Lane 1: marker
 Lane 2: 206 before induction
 Lane 3: 206 1 hr after induction
 Lane 4: 206 2 hr after induction
 Lane 5: 206 3 hr after induction
 Lane 6: 5606 before induction
 Lane 7: 5606 1 hr after induction

Lane 8: 5606 2 hr after induction
 Lane 9: 5606 3 hr after induction
 Lane 10: marker
 Lane 11: N80/89 before induction
 Lane 12: N80/89 1 hr after induction
 Lane 13: N80/89 2 hr after induction
 Lane 14: N80/89 3 hr after induction

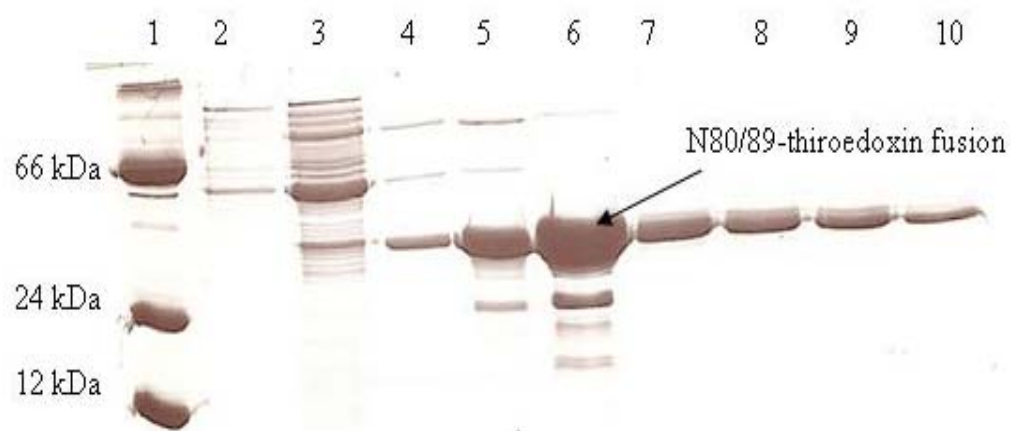


Figure 4.24. Purification on N80/89 expressed in the pET32a vector. SDS-PAGE analysis of the purification of N80/89 with a thioredoxin tag. The fractions have been recovered from a nickel chelating column.

Lane 1; marker	Lane 6: fraction 6
Lane 2: fraction 2	Lane 7: fraction 8
Lane 3: fraction 3	Lane 8: fraction 9
Lane 4: fraction 4	Lane 9: fraction 10
Lane 5: fraction 5	Lane 10: fraction 11

most significant amount of the fusion protein also contained other proteins indicated by the presence of other bands on the gel. These bands may be attributed to protein degradation so further purification was necessary. Fraction 6 was dialyzed against 20 mM Tris-Cl pH 8.0 to remove the imidazole. The protein was then loaded onto a Hi-trap Q HP column, which is an anion exchange column, for further purification. Three injections were required to load all of the protein onto the column. The protein was eluted using a salt gradient. The anion exchange column was successful in purifying fraction 6 (Figure 4.25).

To facilitate the calculation of protein concentration the extinction coefficient had to be estimated. The protein calculator website was used to determine the extinction coefficient of the thioredoxin-N80/89 (Trx-N80/89) fusion protein. The program estimated the molar extinction coefficient based on a method by Gill and von Hippel (Gill and von Hippel, 1989) where lyophilized proteins were used to establish an absorbance curve based on the number of tryptophans, tyrosines, and disulfide bonds. A molar extinction coefficient of $20,910 \text{ M}^{-1}\text{cm}^{-1}$ was estimated for Trx-N80/89. The protein concentration was calculated for the purification of fraction 6 of injection 2 from Figure 4.23. From that fraction, 5 mg of 6 mL of protein was recovered.

4.12 Tb^{3+} binding of N80/89 fusion protein.

A Tb^{3+} titration was carried out on Trx-N80/89 in 100 mM MOPS. However, saturation could not be reached for this experiment presumably due to non-specific interactions with thioredoxin. The buffer was changed to 20 mM PIPES, 10 mM KCl pH 6.8. The salt will limit the non-specific binding by interacting with charges on the

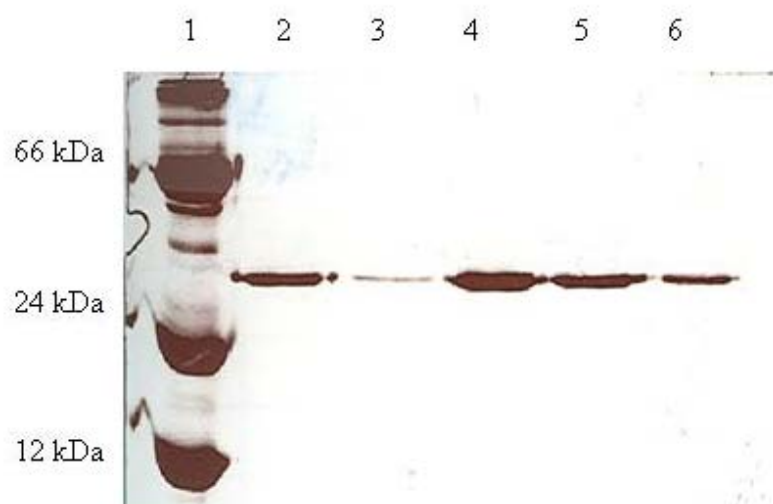


Figure 4.25. Further purification of N80/89-thioredoxin. SDS-PAGE analysis of fractions recovered from an anion exchange column to further purify N80/89-thioredoxin fusion protein.

Lane 1: marker
Lane 2: fraction 6 (injection 1)
Lane 3: fraction 5 (injection 2)
Lane 4: fraction 6 (injection 2)
Lane 5: fraction 6 (injection 3)
Lane 6: fraction 5 (injection 3)

surface of the protein. The titration reached saturation and yielded a Tb^{3+} binding affinity of $6.6 \pm 0.5 \mu\text{M}$ (Figure 4.26). This affinity is similar to the K_d of the N80/89 protein without the fusion tag, however, the addition of salt to the buffer, shielding the surface charges, decreases the binding affinity of the protein. Thus the two experiments are not comparable. Presumably the K_d of Trx-N80/89 is stronger than for N80/89 since the titration with salt gives a K_d within the same magnitude.

4.13 Second generation variants

The altered conformation and rapid degradation of the first generation proteins renders further analysis of these proteins very difficult. Even with the thioredoxin tag, the designed protein N80/89 is not folded and binding affinities obtained with a fusion tag are not as reliable as the protein alone. Therefore, it became necessary to design second generation proteins gleaned from the knowledge obtained from the study of the first generation proteins.

For the design of the second generation proteins several new criteria were added based on the first generation sites. First, sites were chosen based on the number of mutations that had to be engineered. For the first generation sites five mutations were engineered which was very aggressive. For the second generation proteins the sites with the least number of mutations to be engineered were selected. Second, residues that were exposed to solvent and not involved with the hydrophobic core were selected. Third, the loops of the protein are the preferred region of the sites.

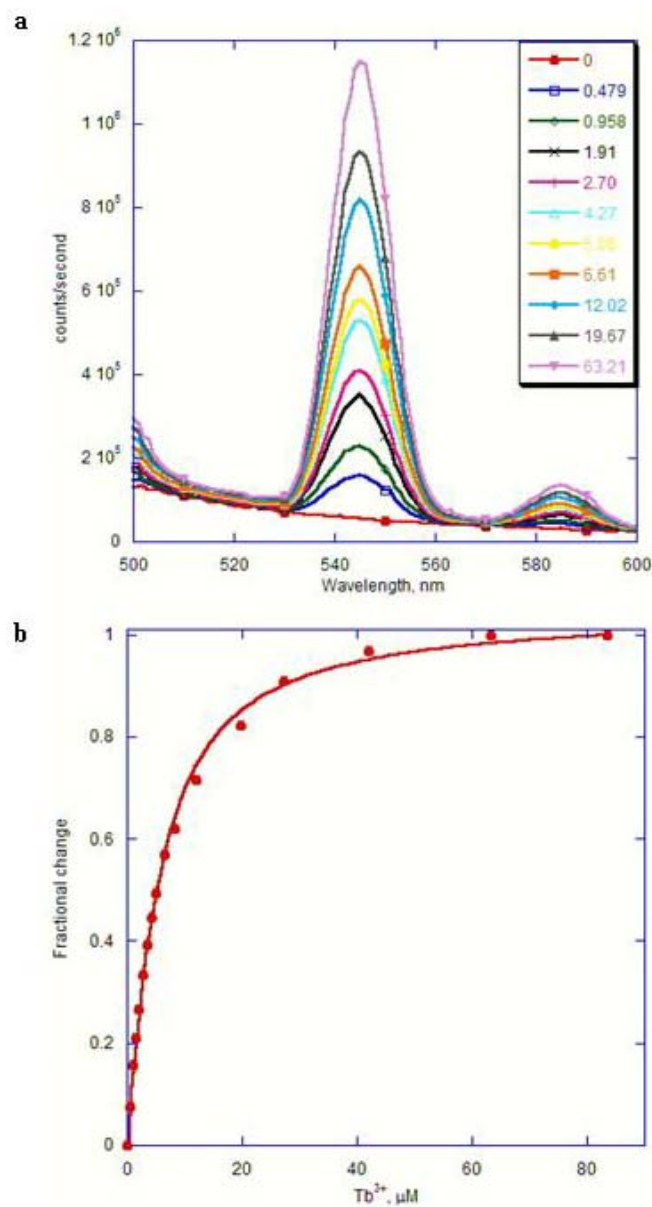


Figure 4.26. Tb^{3+} binding affinity of N80/89-thioredoxin. (a) Enhancement of fluorescence emission at 545 nm at increasing additions of Tb^{3+} . (b) The fit of the fractional change of emission yields a Tb binding affinity of $6.6 \pm 0.5 \mu\text{M}$.

4.13.1 Second generation designed protein 7E54

Two second generation designed proteins, 7E54 and 6D79, were engineered as discussed in **section 3.4**. The 7E54 protein required three mutations, A54E, I65D, T69D, and also utilizes the natural ligands N67 and D72 (Figure 4.27). The protein is named 7E54 because according to the design it has a coordination number of seven with E54 as the bidentate ligand.

The protein was expressed in the pGEX-2T vector in LB media and purified using the established GST fusion protocol. After purification 2.6 mg of protein was recovered. This yield is smaller than yields recovered for wild type CD2.

The expression of 7E54 in LB buffer was successful (Figure 4.28). However, the purification yields were very low. The majority of the protein formed inclusion bodies in the cell pellet after sonication (Figure 4.28). The conformational analysis of the protein was examined using Trp fluorescence. The spectra reveal the tertiary packing of the protein has been perturbed indicated by a shifted emission of 356 nm, which is similar to free Trp (Figure 4.29). At this point it was assumed that since this protein had an altered conformation it would behave similarly as the first generation proteins. The protein would not be as stable as wild type CD2 and would degrade rapidly. In fact the low expression yields of the protein could be indicative of the instability of the protein. Therefore, no further studies were done with designed protein 7E54.

The second generation protein 6D79 will be discussed in detail in Chapter 5 of this dissertation.

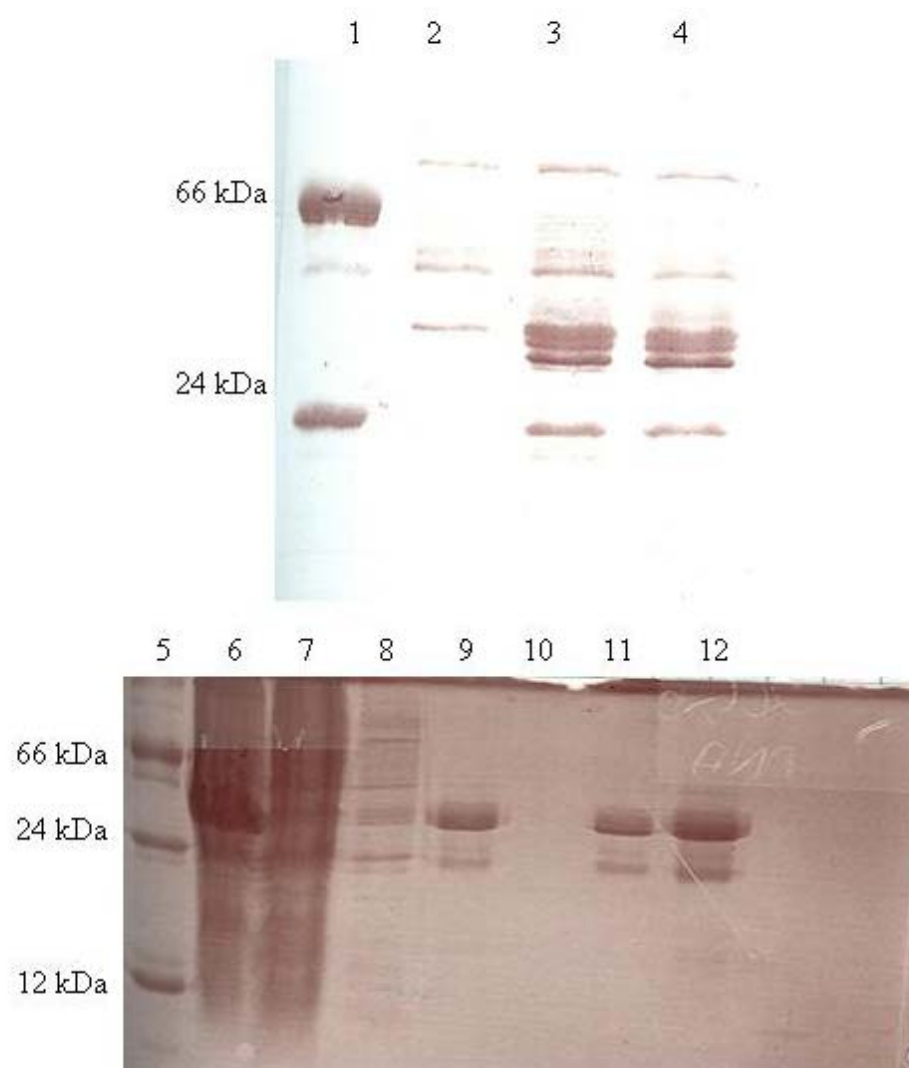


Figure 4.28. Expression and purification of 7E54. SDS-PAGE gel analysis of the expression and purification of the second generation designed protein 7E54.

- | | |
|--------------------------------------|-------------------------------|
| Lane 1: marker | Lane 7: waste |
| Lane 2: before induction | Lane 8: wash |
| Lane 3: 1 hr after induction | Lane 9: beads before binding |
| Lane 4: 2 hr after induction | Lane 10: beads before binding |
| Lane 5: marker | Lane 11: beads after binding |
| Lane 6: cell pellet after sonication | Lane 12: beads after binding |
| | Lane 13: elution |

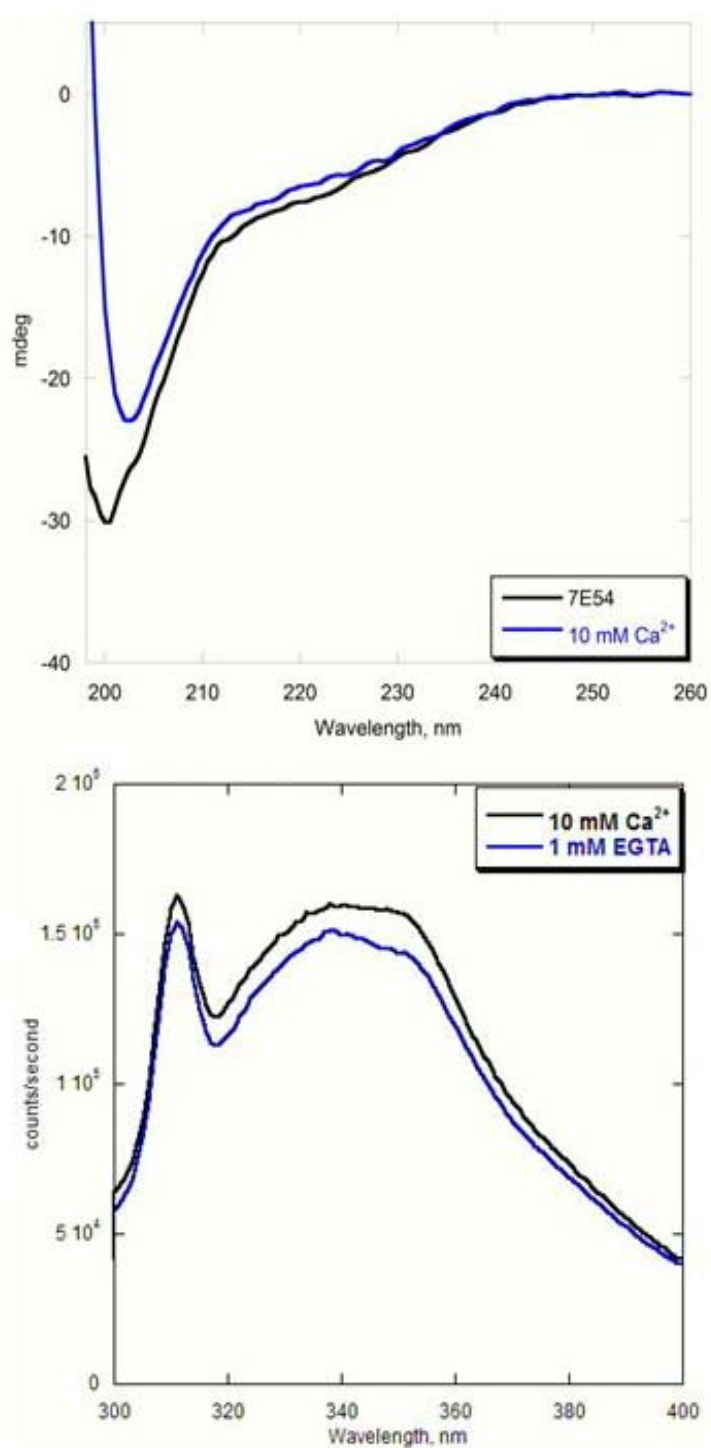


Figure 4.29. Conformational analyses of 7E54 in the presence and absence of Ca^{2+} . The far UV CD spectra (a) and fluorescence emission (excited at 283 nm) of 7E54 (3 μM) in 10 mM Tris pH 7.4.

Chapter 5. Rational design of a novel calcium binding site adjacent to the ligand binding site on CD2 increases its adhesion function

5.1 Cell adhesion proteins

Cell adhesion molecules are a large group of proteins involved in diverse cell-to-cell and cell-to-extra-cellular matrix interactions (Figure 1.19). The Ca^{2+} -dependent cell adhesion proteins cadherins form homophilic dimers from different cells and are important for the development of multicellular organisms, and a prerequisite to maintain tissue integrity by stabilizing cell-cell contacts (Herrenknecht, 1996). Non- Ca^{2+} -dependent cell adhesion proteins such as CD2, that is expressed on most human T cells and natural killer (NK) cells, initiate and maintain contact with antigen-presenting or target cells and play an important role in signaling (Blattman and Greenberg, 2004; Driscoll et al., 1991; van der Merwe et al., 1995). The extracellular domain 1 of CD2 that is involved in molecular recognition has a similar topology to that of the Ca^{2+} -dependent cell adhesion molecule cadherin (Figure 1.19). A relatively large surface area is reported for CD2 protein-protein interactions, which is different from the strong and focused enzyme/inhibitor interactions (Nooren and Thornton, 2003). In addition to their low binding affinities ($K_d > \mu\text{M}$), the highly dynamic, transient and reversible interactions of CD2 between T-lymphocytes and antigen-presenting cells (Chen et al., 2000; van der Merwe and Barclay, 1994) make understanding the key factors that contribute to the molecular recognition challenging.

To date, CD2 is one of the most extensively studied cell adhesion proteins and has served as a model system to identify the key determinants involved in molecular recognition. CD2 consists of two extracellular domains and an intracellular C-terminal domain linked by a single transmembrane region. The physiological binding partner of CD2 has been shown to be CD48 in rodents and CD58 in humans. Heterophilic protein/protein interaction at the cell surface has been identified using mutagenesis and structural studies (Selvaraj et al., 1987). The GFCC'C'' surfaces of the N-terminal IgSF domains for both rat and human of CD2 are relatively flat and highly charged. Compared with ~19% for the average solvent-exposed protein surface, about 35 and 70% of the surface residues are charged in the rat and human CD2, respectively (Chen et al., 2000). Seven charged residues (D28, E29, R31, E33, E41, K43, and R87) and three aromatic or hydrophobic residues (Y81, F49 and L38) located on GFCC'C'' surface contribute to the recognition of CD48 (Davis et al., 1998a). Mutations with reversed surface charges have been shown to result in weak binding or loss of binding to CD48 (van der Merwe et al., 1995) (**section 1.9**). The ligand-binding ability is restored if there is a complementary charge reversal on the CD48 surface, which suggests the importance of the electrostatic complementarity for the binding interaction (Davis et al., 1998a).

5.2 Electrostatic interactions and Ca^{2+} binding affinity

As discussed in **section 1.6**, electrostatic interactions are also important for Ca^{2+} binding. Ca^{2+} is predominantly chelated by oxygen atoms from charged carboxyl side chains of Asp and Glu (Yang et al., 2002b), hydroxyl groups of Ser and Thr, carbonyl Asn, Gln, and main chain groups, and oxygen from the solvent. A high percentage of

charged ligand residues have been observed in naturally occurring Ca^{2+} binding sites (Marsden et al., 1990) and clustering of charged ligand residues contributes to strong Ca^{2+} binding affinity. In addition, analysis of the structural characteristics of Ca^{2+} binding sites in naturally occurring proteins have shown that these sites are often positioned within loop or turn regions on or near the protein surface (Strynadka, 1993). To date, to our knowledge, there have been no reports of a naturally occurring Ca^{2+} site that possesses ligand residues all from the β -sheets (**section 1.2**).

We have previously reported our success in the design *de novo* of Ca^{2+} binding sites in non- Ca^{2+} -binding proteins without splicing in an entire, known metal-binding domain (Yang et al., 2003; Yang et al., 2005). The designed protein Ca.CD2 binds Ca^{2+} at the intended site with the designed arrangement, which validates our general strategy for designing *de novo* Ca^{2+} -binding proteins. In addition, the designed protein Ca.CD2 retains the ability of associating with the native partner CD48.

To understand electrostatic interactions, in particular, in molecular interaction and Ca^{2+} binding, in this chapter, we report the success in engineering a novel Ca^{2+} binding site into CD2 using our developed design approach. NMR studies indicate that Ca^{2+} specifically binds to this all β -sheet ligands site without a significant conformational change. Interestingly, the introduction of negatively charged Ca^{2+} ligands increased the CD2-CD48 binding affinity, suggesting that the molecular recognition capability of CD2 can be modulated by altering the local charge environment. This first success in designing a novel Ca^{2+} -binding site with enhanced cell adhesion function represents a major achievement toward designing functional proteins and unveiling key determinants

to molecular recognition. This designed protein also served as a model protein to further understand non-local factors that contribute to protein stability (chapter 6), calcium binding (chapter 7), and protein folding (chapter 8).

5.3 Rational design and selection of the novel Ca^{2+} binding site

Protein design is an important tool that allows for the study of the structural and functional properties of proteins. The design of Ca^{2+} binding proteins provides insight into the biological role of Ca^{2+} ions without the difficulties encountered upon a global conformational change as observed in many natural Ca^{2+} binding proteins. This study endeavors to elucidate the role of electrostatic interactions on Ca^{2+} binding affinity and molecular recognition using protein design and structural analysis.

To achieve our objective of creating a Ca^{2+} binding site formed by all β -sheet ligand residues without the disruption of cell adhesion function, we have designed potential Ca^{2+} binding sites within the backbone of domain 1 of CD2 (1hng) using our established design strategy. In addition to the use of the prevalent pentagonal-bipyramidal geometry and a library of side chain rotamers of ligand residues (or atoms from the main chain) with strong Ca^{2+} binding preferences (Davis et al., 1998a; Wilkins et al., 2002; Yang et al., 2003), several new criteria were applied for the design and selection processes. First, the bidentate ligand Asp (Pidcock and Moore, 2001) with a branched side chain has a strong β -sheet propensity, which is important for stabilization of the native conformation. Second, the site was engineered at the edge of the CD48 binding site in order to modulate the adhesion. Third, residues known to be essential for CD48 binding were not mutated (Figure 5.1). Fourth, solvent accessibility, side chain steric

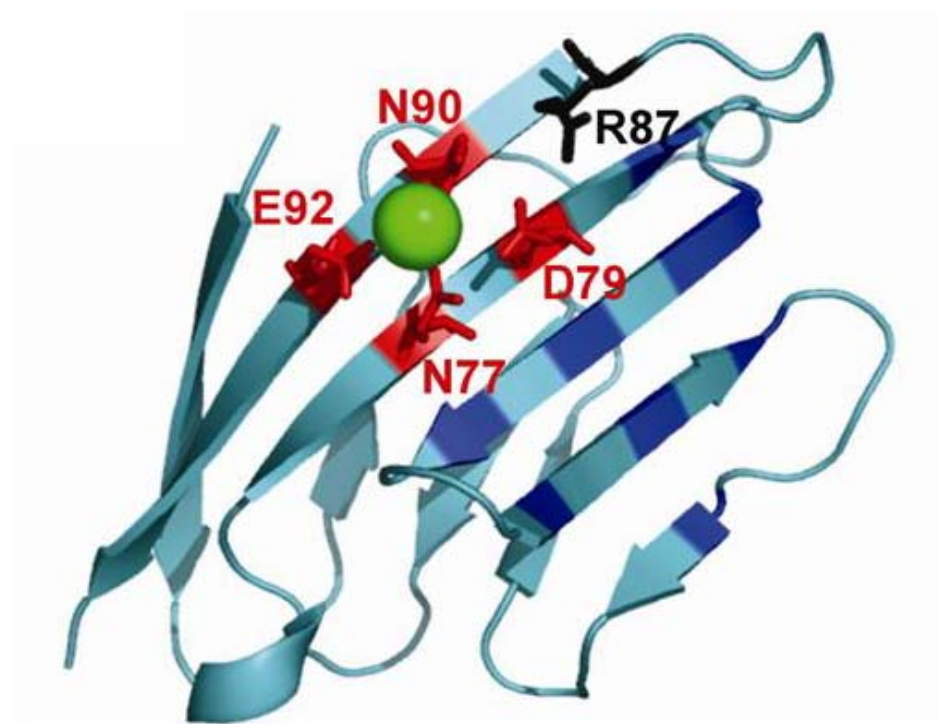


Figure 5.1. Structure of CD2.6D79. The model structure of the N-terminal domain of rat CD2 with a designed calcium binding site (CD2.6D79) based on 1hng generated by Molscrip. Calcium is coordinated by the sidechain oxygen and nitrogen of residues N77, D79 (bidentate), N90, and E92 along with two oxygen atoms from solvent water in a pentagonal bipyramidal geometry. Original residues and the designed calcium binding residues are shown in ball-and-stick representation. Essential residues involved in the recognition of CD48: D28, E29, R31, E33, L38, E41, K43, F49, Y81 and R87 are also highlighted in blue.

conflicts with existing atoms, and disruption of hydrogen bonding, conserved folding residues, and hydrophobic interactions (Yang et al., 2005) have also been considered.

Figure 5.1 shows the model structure of the selected Ca^{2+} binding site CD2.6D79. The proposed Ca^{2+} binding pocket is formed by the side chain oxygen atoms of two charge mutations T79D and A92E and two natural amino acids, N77 and N90. Residues T79D and N77 are from the F β -sheet while A92E and N90 are from the G β -sheet of the protein (Figure 5.1). This site is adjacent to the essential cell adhesion residues E33, R31, and Y81. Electrostatic calculations show that the surface potential of CD2.6D79 compared to CD2 has decreased by 74.1 kcal/mol. The mutation T79D contributes 1.8 fold greater (63%) to the perturbation of the electrostatic surface than that does A92E (37%).

5.4.1.1 Conformational analyses of CD2.6D79

Several spectroscopic techniques were employed to analyze the secondary and tertiary structure of CD2.6D79. The far UV CD and the Trp fluorescence spectra of CD2.6D79 are almost identical to that of wild type CD2 (Figure 5.2 a, b). Therefore, the introduction of the Ca^{2+} binding site does not alter the secondary and tertiary structure of CD2. In addition, no detectable change in the CD and fluorescence spectra of CD2.6D79 was observed upon the addition of Ca^{2+} (Figure 5.2a, b).

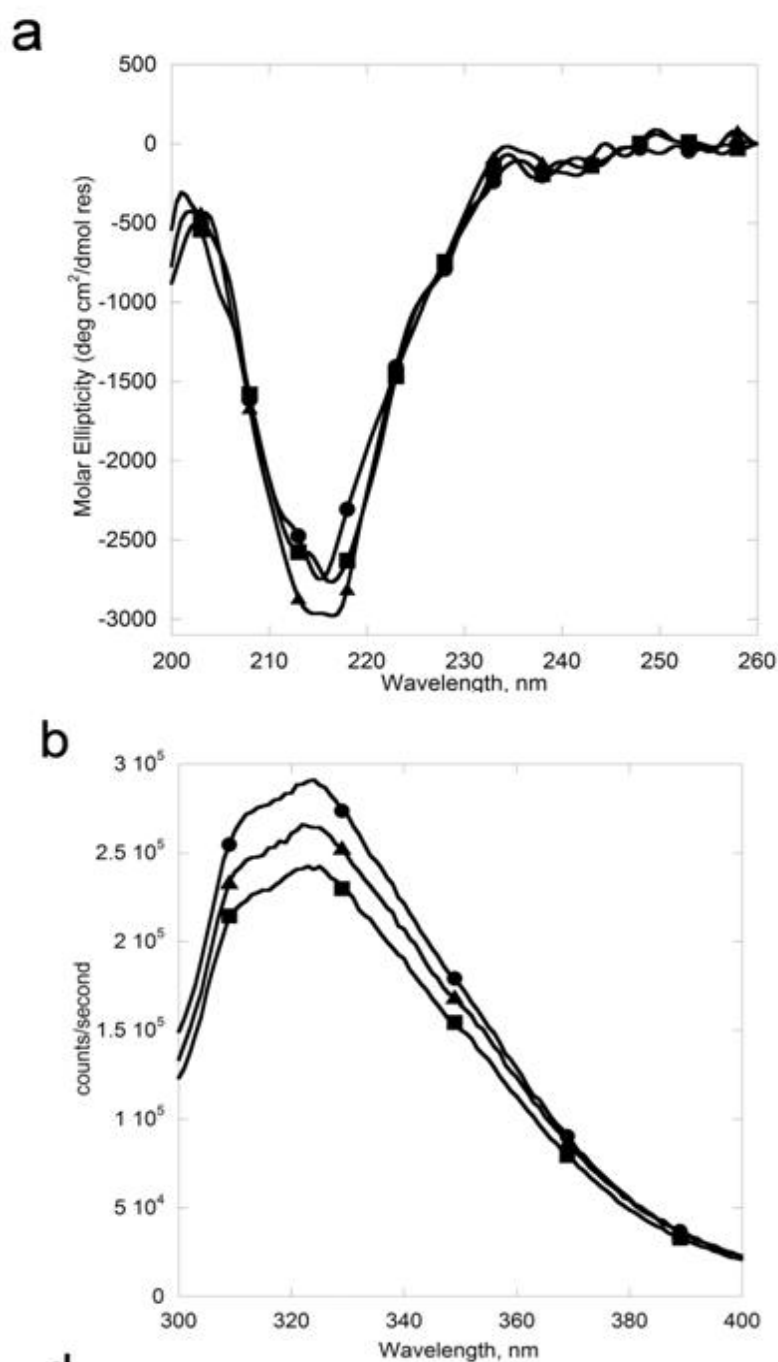


Figure 5.2. Conformational analyses of CD2.6D79. (a) The far UV CD spectra of CD2.6D79 in the absence (triangles) and presence (squares) of 10 mM Ca²⁺ and the spectra of CD2 (circles). (b) The Trp fluorescence spectra of CD2.6D79 in the absence (triangles) and presence (squares) of 10 mM Ca²⁺ and the spectra of CD2 (circles).

5.4.1.2 High resolution NMR

NMR is one of the principle techniques used to obtain physical, chemical, electronic and structural information about a molecule. It is the only technique that can provide detailed information on the exact three-dimensional structure of biological molecules in solution (Wuthrich, 1986).

NMR is a physical phenomenon based upon the magnetic property of an atom's nucleus. NMR studies a magnetic nucleus by aligning it with an external magnetic field and perturbing this alignment using an electromagnetic field (Hornak, 1997-2006). The response to the field is what is exploited in NMR spectroscopy. Some nuclei experience this phenomenon, and others do not, dependent upon whether they possess certain spin properties. Nuclei are surrounded by orbiting electron that partially shields the nuclei. The amount of shielding depends on the exact local environment. For example, a hydrogen bonded to an oxygen will be shielded differently than a hydrogen bonded to a carbon atom. In addition, two hydrogen nuclei can interact via a process known as spin-spin coupling if they are on the same molecule, which will split the lines of the spectra in a recognizable way. By studying the peaks of nuclear magnetic resonance spectra the structure of many compounds can be determined. It can be a very selective technique, distinguishing among many atoms within a molecule or collection of molecules of the same type, but which differ only in terms of their local chemical environment (Wuthrich, 1986). This makes NMR a powerful tool to study proteins and their interactions.

One-dimensional ^1H NMR is a commonly used technique because of the natural abundance of ^1H and the large occurrence of ^1H in proteins. The chemical shifts of the

varying protons in the protein will be different based on whether they are a side chain proton, an α carbon proton, or a backbone proton. In addition chemical shifts of protons can be visualized upon ligand binding.

The introduction of 2D NMR techniques dramatically increased the resolution of protein NMR spectra compared to 1D spectra. This led to the development of a systematic method for the assignment of the 2D NMR spectra of proteins that relied only on information about the amino acid sequence of the proteins (Redfield, 1998). This method is the sequential assignment procedure. The assignment of resonances in the complex NMR spectrum of a protein is the first step in any study of protein structure, function, or dynamics (Redfield, 1998). Homonuclear TOCSY and NOESY experiments are frequently used 2D spectra (Wuthrich, 1986). TOCSY experiments are based on through bond coupling while NOESY experiments are based on through space coupling. Increased resolution of spectra was seen further with the advancement of heteronuclear labeling ^{15}N -HSQC experiments. Although Ca^{2+} is not directly measured in NMR techniques chemical shift comparisons between the Ca^{2+} free and Ca^{2+} bound samples will be very useful in examining Ca^{2+} binding

High resolution NMR was utilized to reveal the detailed structural change of CD2.6D79. Homonuclear protein was expressed and purified as detailed in section 2.1. Figure 5.3 shows the 1D NMR of 6D79 vs wild type CD2. There are no significant differences between the two spectra indicating that CD2.6D79 has not undergone any significant conformational change upon the introduction of the calcium binding site.

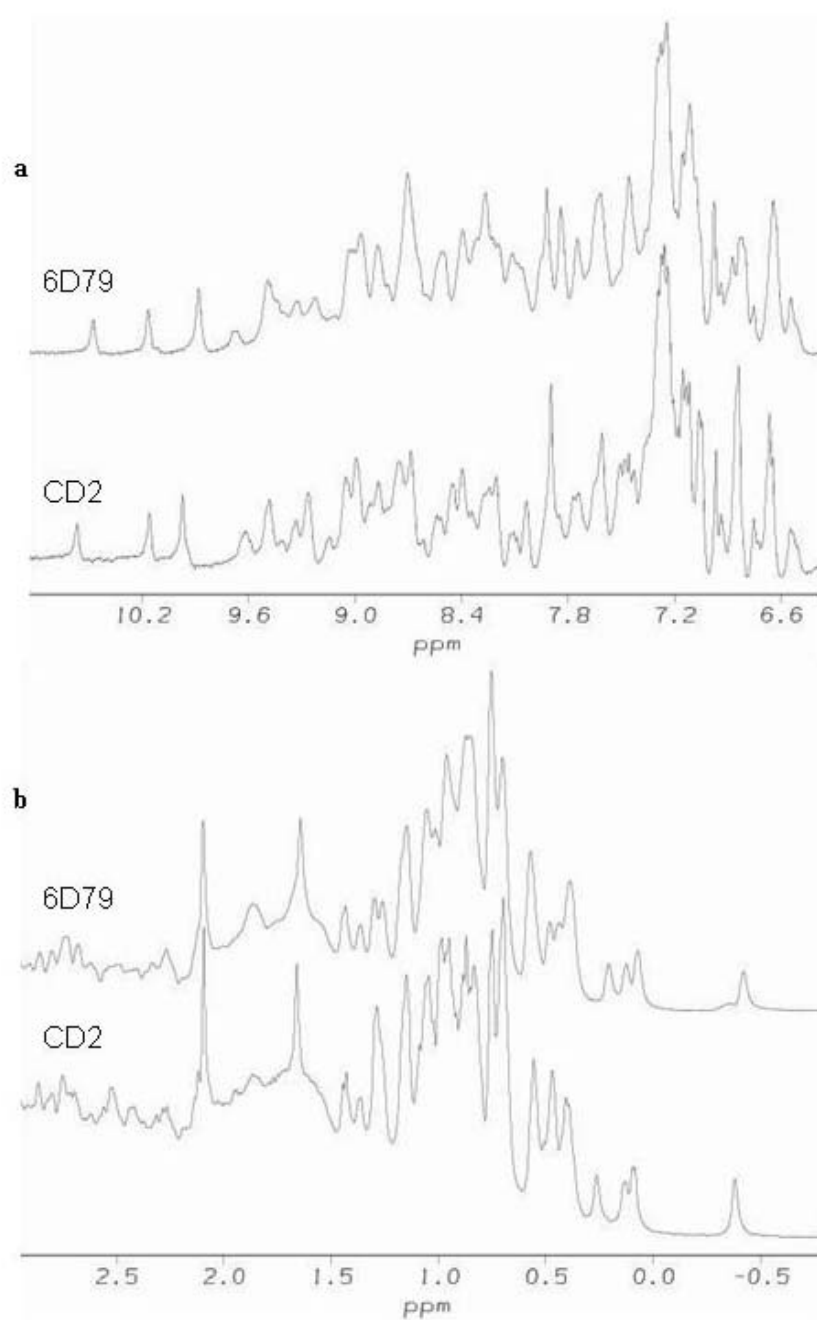


Figure 5.3. 1D NMR of CD2.6D79. The amide (a) and side chain (b) regions of the NMR spectra of CD2.6D79 compared to CD2 indicate the conformation of CD2.6D79 is similar to CD2.

Homonuclear TOCSY and NOESY experiments were performed on the protein to further analyze its conformation. Freshly purified proteins were prepared as detailed in **section 2.8.2**. The fingerprint region of the TOCSY spectra is shown in Figure 5.4. The chemical shift changes between the apo-form (1 mM EGTA) and the loaded form (10 mM Ca^{2+}) in homonuclear 2D total correlation spectra (TOCSY) of the majority of the residues at the β -sheets including those from the residues around the Ca^{2+} binding site such as V78, Y81, and K91 are very small (< 0.05 ppm) (Figure 5.5a). In contrast, residues such as E41, K43, and K45 at the flexible loop regions exhibits more pronounced changes, similar to that of CD2 (Chen et al., 2002). On the other hand the more positively charged La^{3+} displays larger chemical shifts when bound to the protein (Figure 5.5b). The lack of overall perturbation in the chemical shifts by Ca^{2+} binding indicates a limited flexibility of the Ca^{2+} binding site formed by a less dynamic secondary structure (β -sheets). According to our and Driscoll's studies, the residues from the β -strands F and G are much more rigid than those from the loops of the protein.

5.4.2 CD2.6D79 has a reduced thermal stability

We have further investigated the thermal stability of CD2.6D79. The CD and fluorescence spectra of CD2.6D79 at 25 and 90 °C are very similar to that of w.t. CD2 suggesting that both proteins have similar native and thermal denaturation states (Figure 5.6a). The thermal transition curves of CD2.6D79 monitored by CD signal at 218 and 222 nm in the presence and absence of Ca^{2+} are consistent with a two-state model (Figure 5.6b) as is w.t. CD2 (Yang et al., 2002a). The measured T_m values for the Ca^{2+} -free and Ca^{2+} -bound forms of the protein are 49 ± 1 and 53 ± 2 °C, respectively, while that for

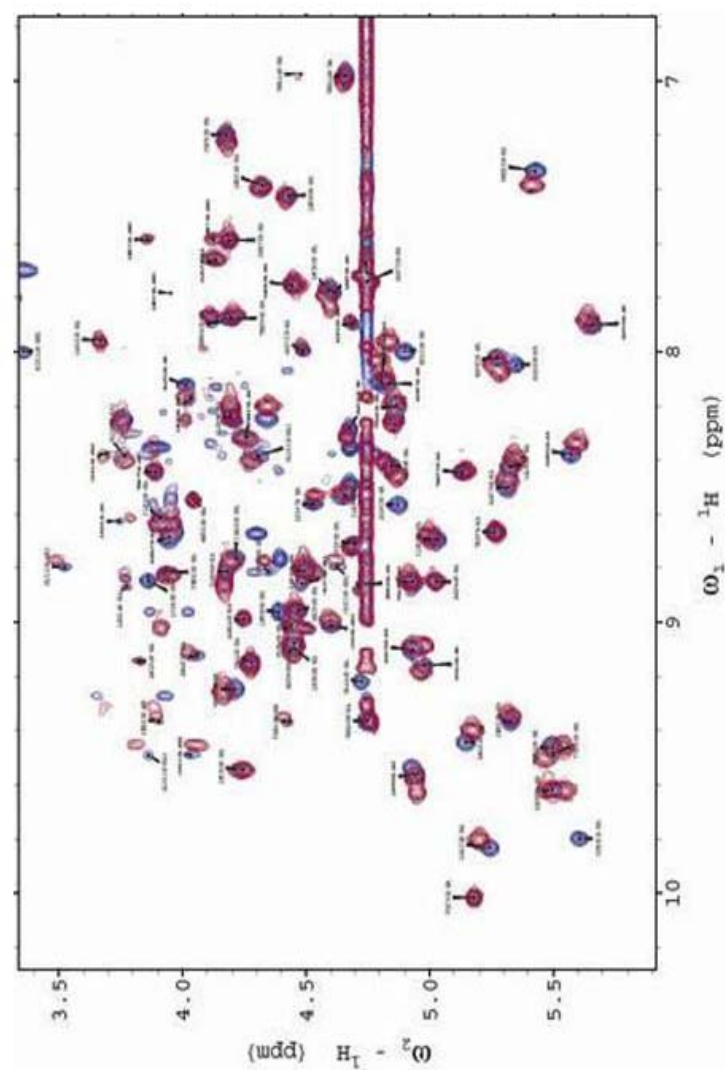


Figure 5.4. TOCSY of CD2.6D79. The comparison of the TOCSY spectra of the Ca^{2+} free and Ca^{2+} bound forms of CD2.6D79 in 20 mM PIPES, 10 mM KCl pH 6.8 indicate changes have occurred when the protein binds metal.

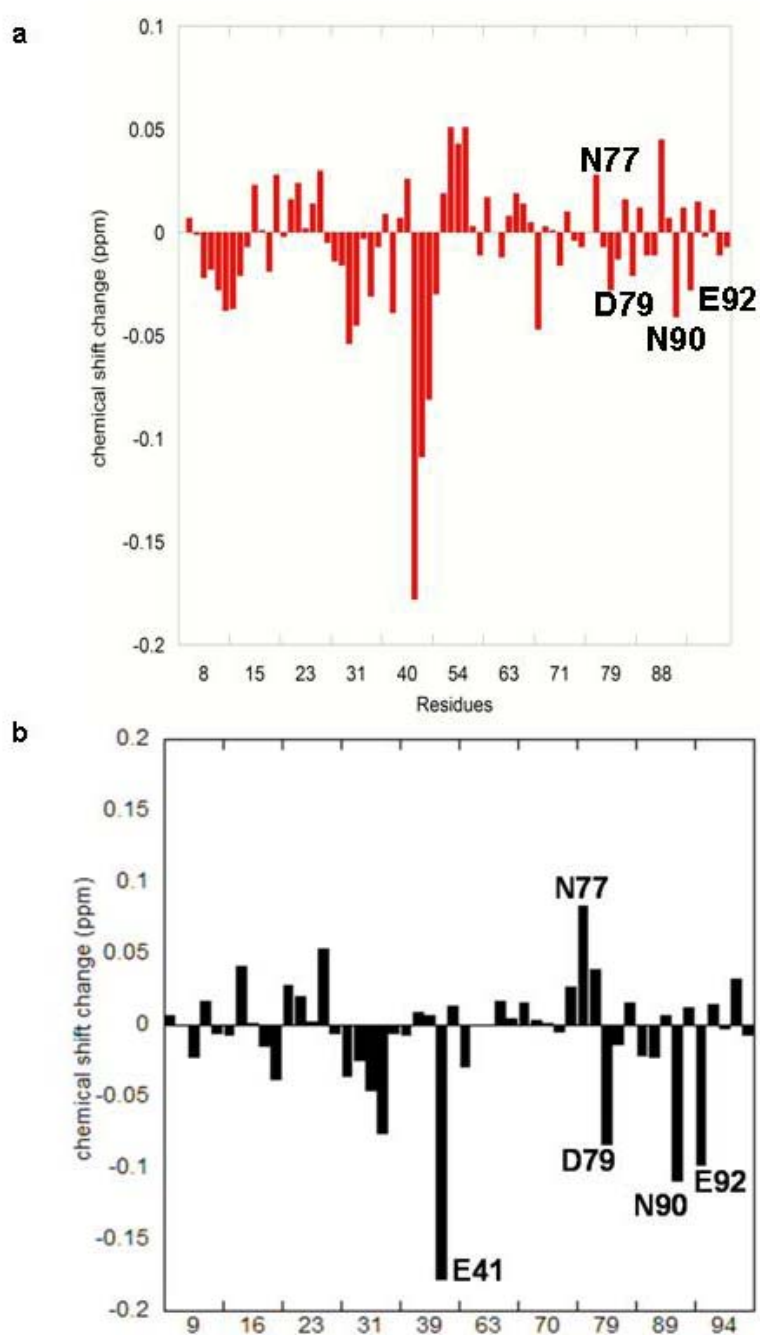


Figure 5.5. Chemical shift changes upon metal binding. (a) Amide proton chemical shift differences of the Ca^{2+} bound and Ca^{2+} free forms of CD2.6D79. (b) Amide proton chemical shift differences of the La^{3+} bound and La^{3+} free forms of CD2.6D79.

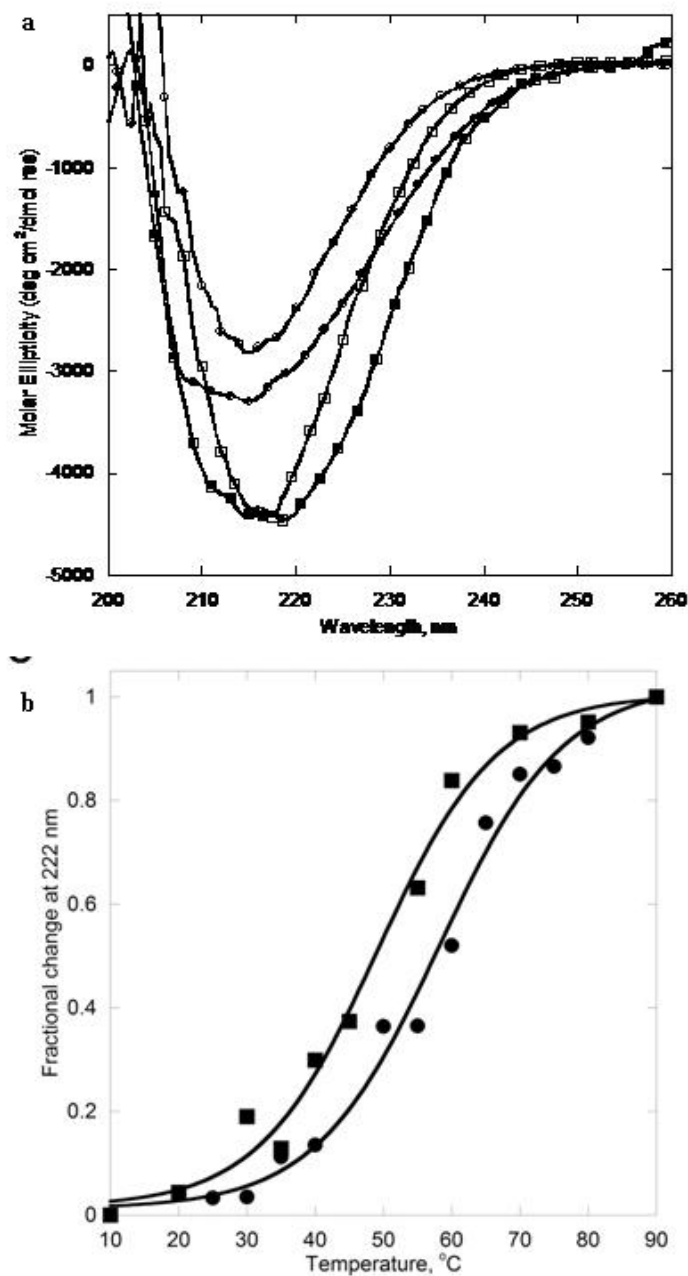


Figure 5.6. Thermal denaturation of CD2.6D79. (a) far UV CD spectra of CD2.6D79 at 25 (open squares) and 90 °C (filled squares) compared to CD2 at 25 (open circles) and 90 °C (filled circles). (b) Thermal stability of CD2.6D79 (20 μ M) in 10 mM Tris, 10 mM KCl pH 7.0 in the absence (filled squares) and presence of calcium (filled circles) give T_m values of 49 and 53 °C, respectively.

CD2 is 62 °C. This thermal stability decrease is likely due to electrostatic repulsion effects. In addition, the binding of the trivalent Ca^{2+} analogues Tb^{3+} and La^{3+} further increase the thermal stability of the protein, 58 ± 2 and 56 ± 1 °C, respectively.

5.5 CD2.6D79 binds to Ca^{2+} and Ln^{3+}

Ln^{3+} has similar ionic radii and metal coordination chemistry with Ca^{2+} and is commonly used to probe Ca^{2+} binding (Drake *et al.*, 1997; Horrocks, 1993) due to the silent spectroscopic properties of Ca^{2+} , especially when the protein lacks conformational change (Table 1.1). Although the majority of residues do not show a significant change in their chemical shifts, addition of Ca^{2+} to CD2.6D79 specifically results in a modest resonance change at 9.7 ppm in 1D ^1H NMR, which is from the designed Ca^{2+} ligand residue Asn-77 (Figure 5.7). Also, the addition of the Ca^{2+} analogue La^{3+} results in a considerable resonance change of residue Asn-77 (Figure 5.7). In addition, significant chemical shift changes in the amide region corresponding to the residues in proximity, such as W32 (11.8 Å) and Y76 (11.6 Å), are observed with La^{3+} binding. This is consistent to the greater effect of La^{3+} than Ca^{2+} on the Ca^{2+} ligand residues, possibly due to the greater charge number of the La^{3+} ion and the increased metal binding affinity (Table 5.1) (Drake *et al.*, 1997; Horrocks, 1993).

To further analyze its Ca^{2+} binding properties and the effect of Ca^{2+} binding on the conformation of the protein, we have applied ^1H - ^{15}N heteronuclear single quantum correlation spectroscopy (HSQC). The ^{15}N -labeled protein was expressed and purified as detailed in **section 2.4.1**. First, the HSQC spectrum of the Ca^{2+} free form of the protein was assigned (Figure 5.8). Figure 5.9 shows the various regions of the HSQC spectra of

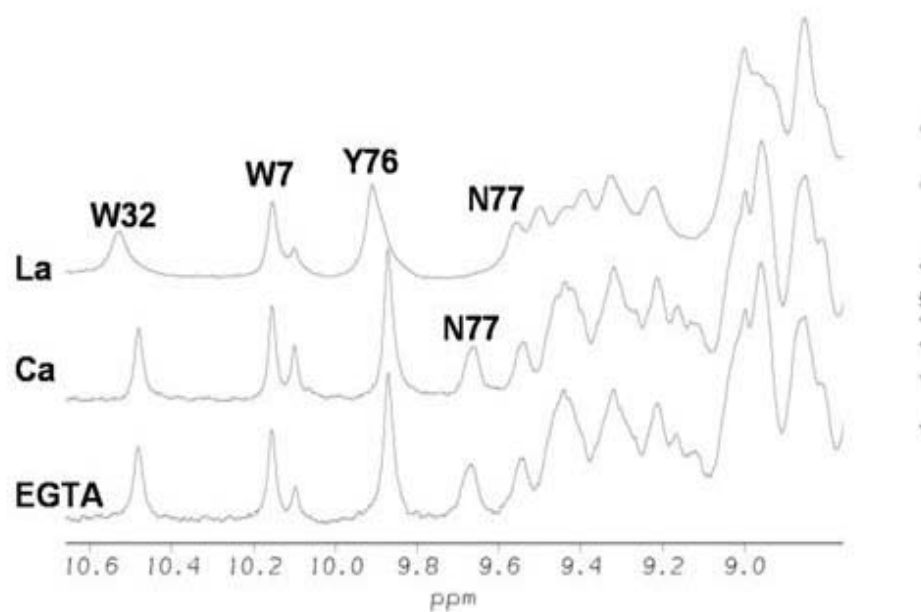


Figure 5.7. Metal induced chemical shift changes CD2.6D79. 1D ^1H -NMR of the amide region of CD2.6D79 in the absence and presence of $600\ \mu\text{M}\ \text{Ca}^{2+}$ or $600\ \mu\text{M}\ \text{La}^{3+}$ in 20 mM PIPES, 10 mM KCl pH 6.8.

6D79	K_d (μ M)	T_m ($^{\circ}$ C) ^c
EGTA	N/A	49 ± 1
Ca ²⁺	630 ± 50^a	53 ± 2
Tb ³⁺	15.3 ± 1.5^b	58 ± 2
La ³⁺	7.4 ± 1.3^b	56 ± 1

^aNMR-20 mM PIPES, 10 mM KCl pH 6.8

^bFRET-10 mM Tris pH 7.4

^cCD-10 mM Tris, 10 mM KCl pH 7.0

Table 5.1. Summary of metal binding affinity and thermal stability of CD2.6D79 to different metal ions.

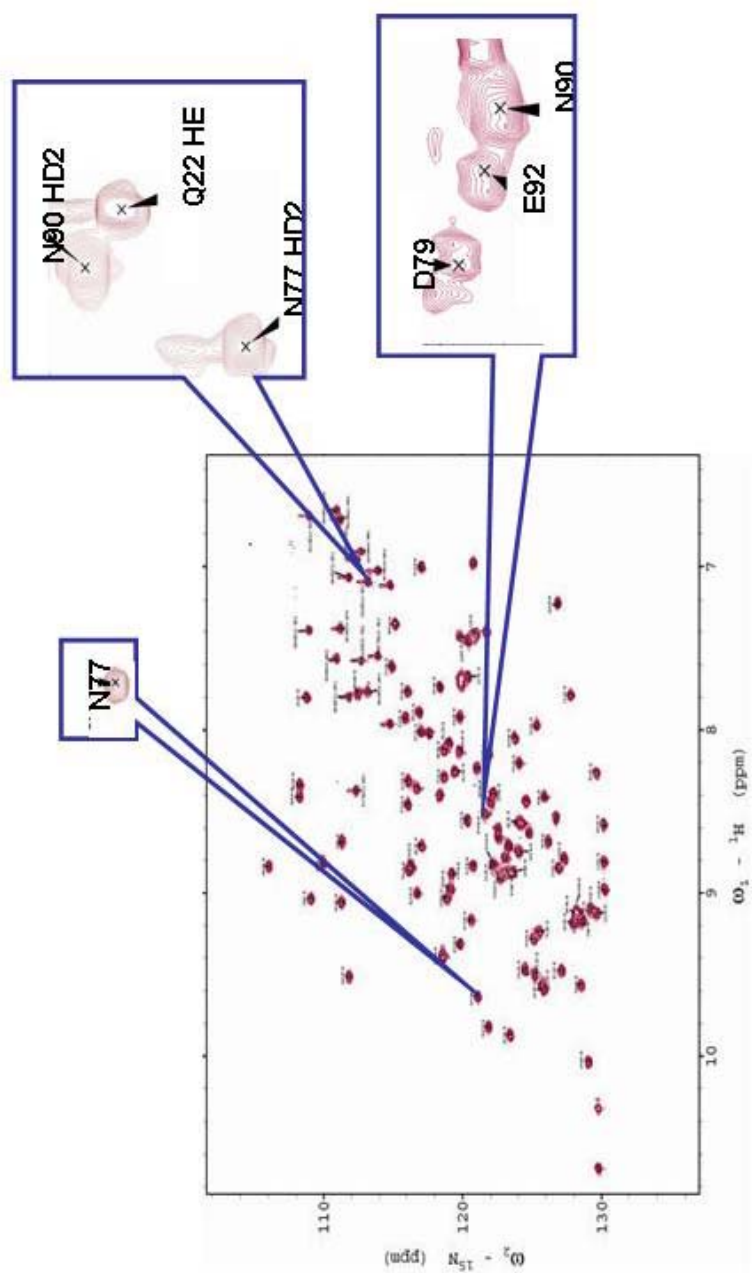


Figure 5.8. HSQC assignment of CD2.6D79. The Ca^{2+} free form of the HSQC spectrum of CD2.6D79 in 20 mM PIPES, 10 mM KCl pH 6.8 was assigned.

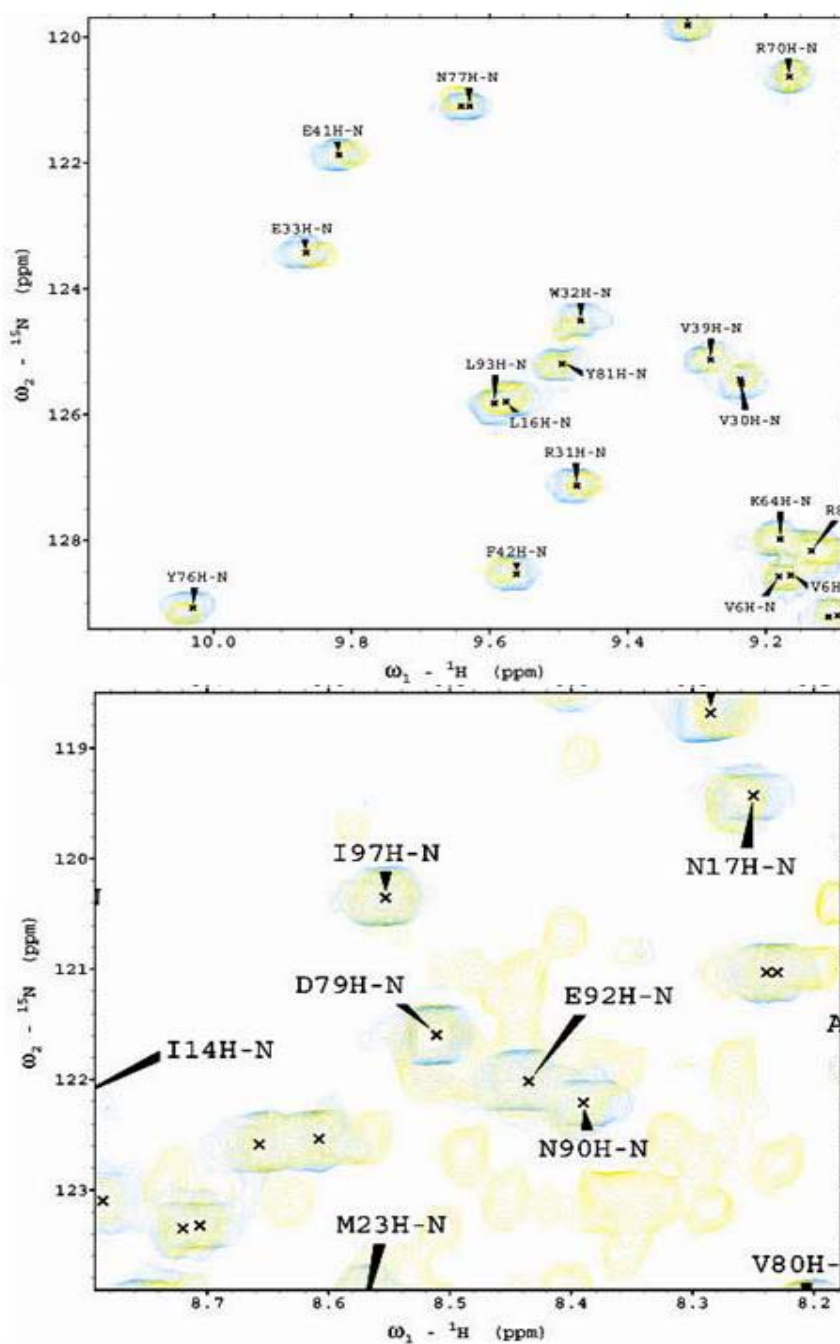


Figure 5.9 HSQC Ca^{2+} titration. Various regions of the HSQC spectra of CD2.6D79 in 20 mM PIPES, 10 mM KCl pH 6.8 in the Ca^{2+} free form (cyan) and the Ca^{2+} bound form (yellow).

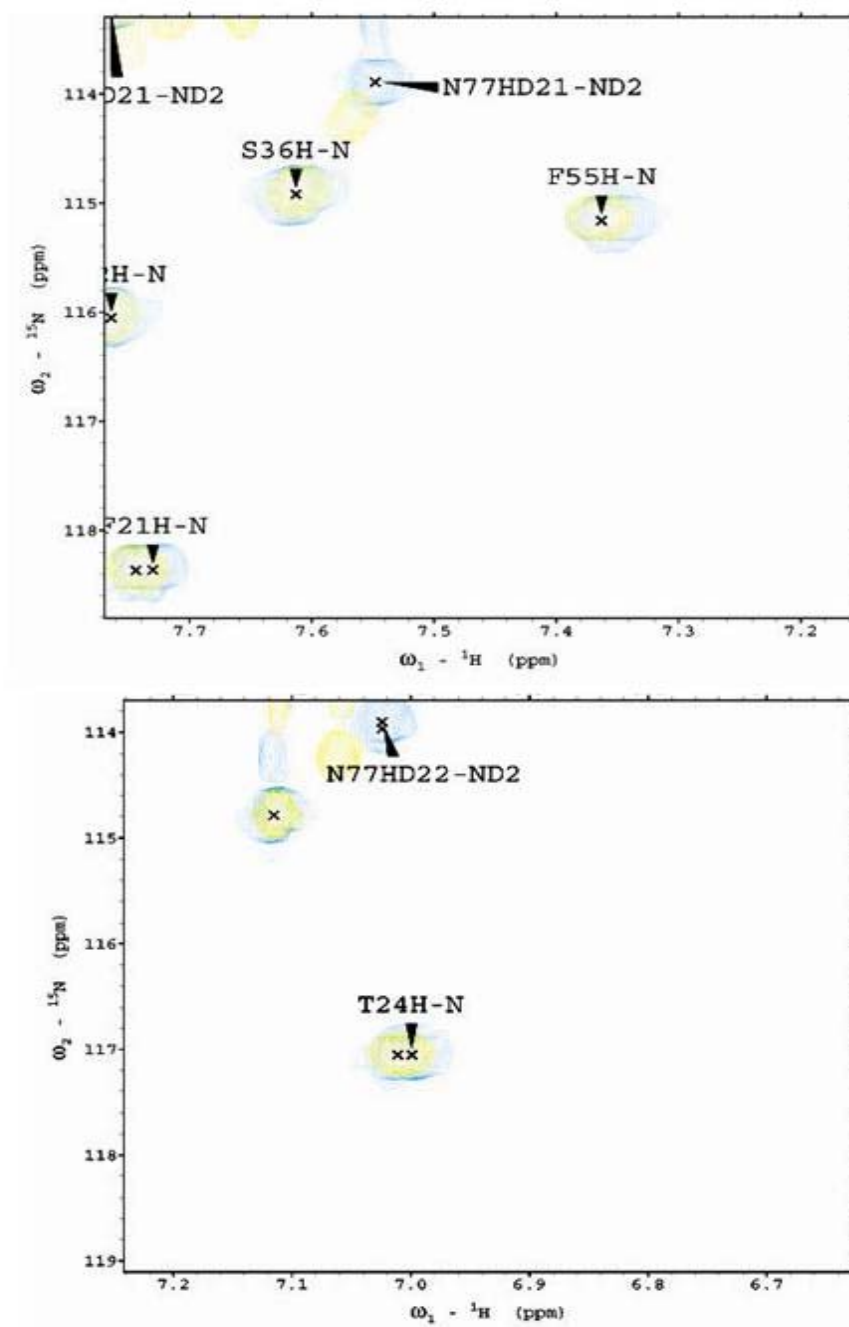


Figure 5.9 continued

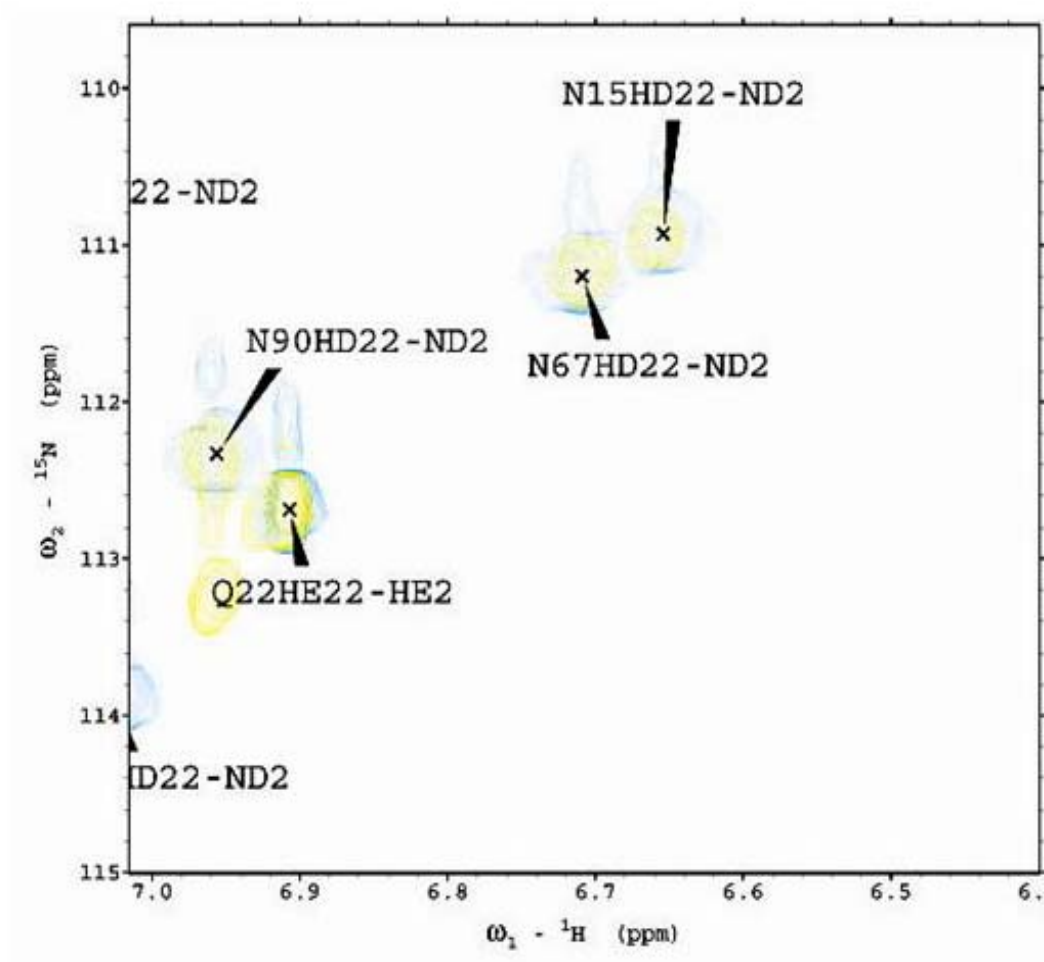


Figure 5.9 continued

6D79 in the absence (1 mM EGTA) and presence of 10 mM Ca^{2+} . The most significant peak shifting upon the addition of Ca^{2+} occurred in the side chain protons of N77 (Figure 5.10). By monitoring this resonance change as a function of Ca^{2+} concentration, a dissociation constant for Ca^{2+} of 0.63 ± 0.05 mM was obtained fitting the fractional change of the chemical shift changes to equation 2.2, a value comparable to those of naturally occurring extracellular Ca^{2+} binding proteins such as cadherins (Yang et al., 2000d). These results suggest that Ca^{2+} and La^{3+} specifically interact with the designed Ca^{2+} ligand residues without imposing a significant conformational change of CD2.

Direct binding of Ln^{3+} was investigated using FRET between aromatic residues and Tb^{3+} . As shown in Figure 5.11a, when excited at 280 nm, fluorescence at 545 nm gradually increases with the addition of Tb^{3+} , giving a K_d of 15.3 ± 1.5 μM to CD2.6D79 obtained by fitting the fractional change of the fluorescence enhancement to equation 2.2 (Figure 5.11b). In addition, this Tb^{3+} fluorescence signal was significantly decreased upon the addition of 10 mM Ca^{2+} or 0.1 mM La^{3+} to the Tb^{3+} -protein complex, suggesting that Ca^{2+} and La^{3+} competitively displaced the bound Tb^{3+} ion. The metal binding affinity of 7.4 ± 1.3 μM for La^{3+} was obtained using this competitive binding (Figure 5.12). The greater affinities for Tb^{3+} and La^{3+} over Ca^{2+} are likely a result of high positive charges for these metal ions. In contrast, addition of 10 mM Mg^{2+} , 10 mM K^+ , 0.1 mM Ni^{2+} , and 0.1 mM Co^{2+} did not significantly decrease the fluorescence intensity, indicating that these concentrations of metals ions cannot compete with Tb^{3+} (Figure 5.13). The limited solubility of Ni^{2+} , Co^{2+} , and Mn^{2+} under the experimental conditions prevents further quantitative competition study at higher metal concentrations but it is

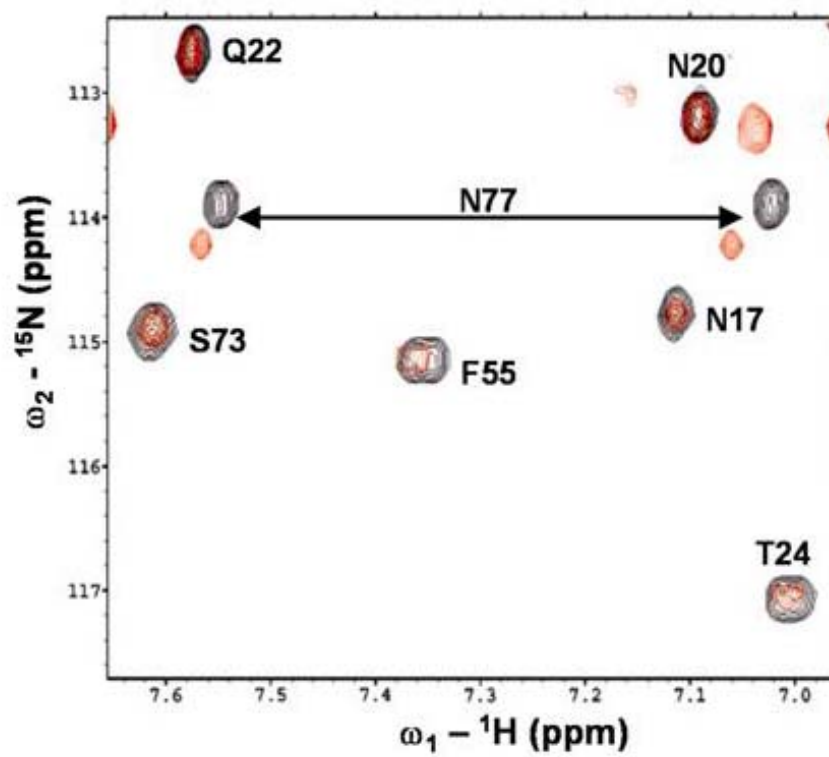


Figure 5.10. Ca^{2+} binding of CD2.6D79. The changes in the ${}^1\text{H}$ - ${}^{15}\text{N}$ HSQC spectra upon the gradual addition of Ca^{2+} (Ca^{2+} free, black; Ca^{2+} bound, red).

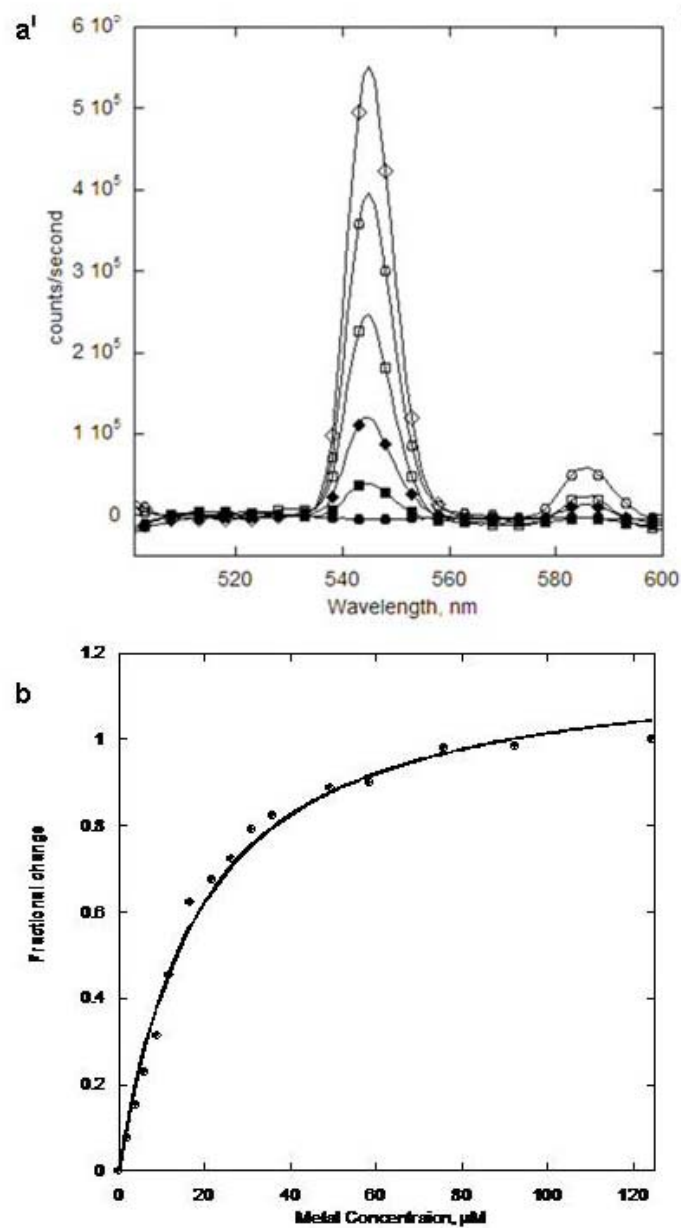


Figure 5.11. Lanthanide binding to CD2.6D79. (a) Enhancement of fluorescence signal at 545 nm as Tb^{3+} concentration is increased (from bottom to top: 0, 2.0, 6.0, 11.9, 26.3, and 124.3 μM) (b) Fractional change of fluorescence enhancement at 545 nm as a function of Tb^{3+} .

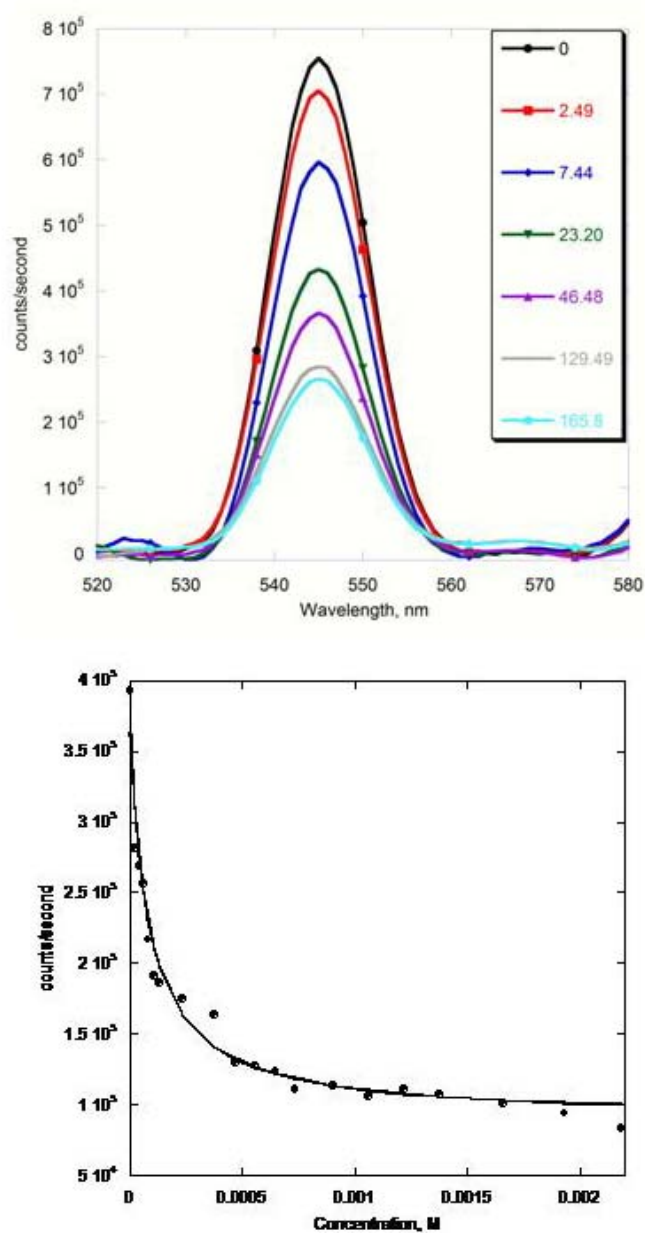


Figure 5.12. La^{3+} binding to CD2.6D79. (a) Decrease of fluorescence emission at 545 nm as La^{3+} concentration (b) Fractional change of fluorescence emission decrease at 545 nm as a function of La^{3+} .

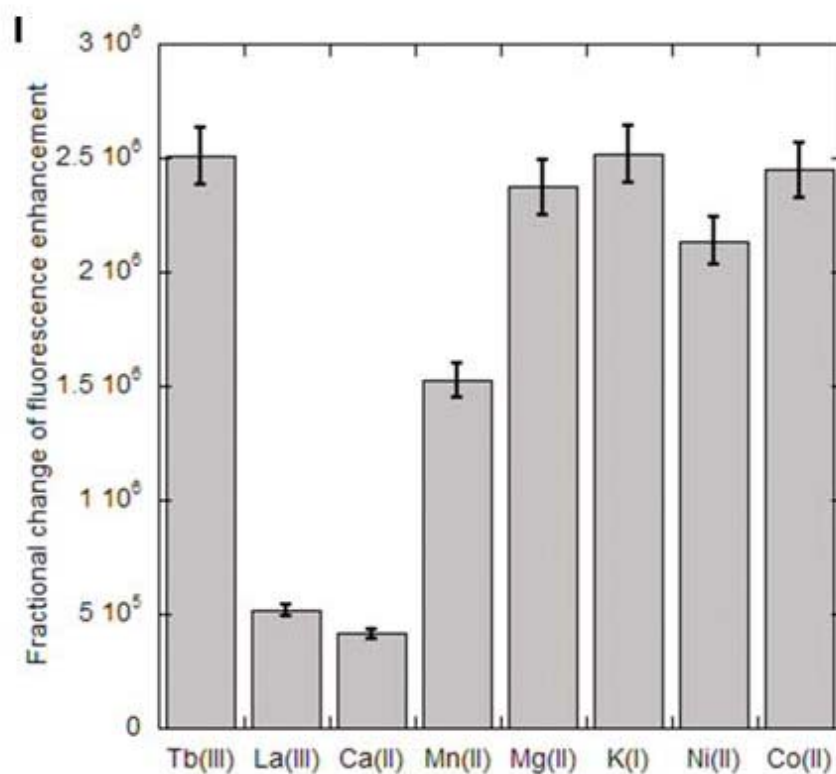


Figure 5.13. Selectivity of CD2.6D79. Tb^{3+} fluorescence decrease in the presence of Ca^{2+} (10 mM), La^{3+} (100 μM), Mn^{2+} (100 μM), Mg^{2+} (10 mM), K^+ (10 mM), Ni^{2+} (100 μM), and Co^{2+} (100 μM)

noteworthy that the physiological concentrations of these metal ions are also very low. The results show that CD2.6D79 exhibits the selective binding of Ca^{2+} over K^{+} and Mg^{2+} and the selective binding of Ln^{3+} over Ni^{2+} and Co^{2+} . Like other natural Ca^{2+} binding proteins, the same concentration of the divalent metal ion Mn^{2+} is able to compete with Tb^{3+} for the binding pocket as the trivalent ion La^{3+} although to a less degree extent (Ubach *et al.*, 1998).

5.6 Further structural validation of our designed Ca^{2+} binding protein

We have applied a paramagnetic relaxation NMR method to further test the validity of our design approach. The Mn^{2+} displays strong relaxation effects on nuclei in proportional to r^{-6} , where r is the distance between the proton and the Mn^{2+} ion, and therefore can be used to predict the location of bound metal ions (Ubach *et al.*, 1998; Yang *et al.*, 2005). Figure 5.14a-b shows the ^1H - ^{15}N HSQC spectra of CD2.6D79 with 1 mM Ca^{2+} in the absence and presence of varying concentrations of Mn^{2+} . The strongest broadening effects at low concentrations of Mn^{2+} were observed in the backbone of the designed ligand residues N77, N90, E92, and D79. Extensive broadening was also observed in the side chain of N77 and N90. Additionally, broadening effects were also detected but to a lesser extent for residues around the metal binding site including E33, R31, and K91. The resonance intensity decreases and the HN-Ca^{2+} distances calculated from the model structure (figure 5.14c) satisfy the theoretical relationship (equation 2.3), validating that the metal ions bind to the designed ligands.

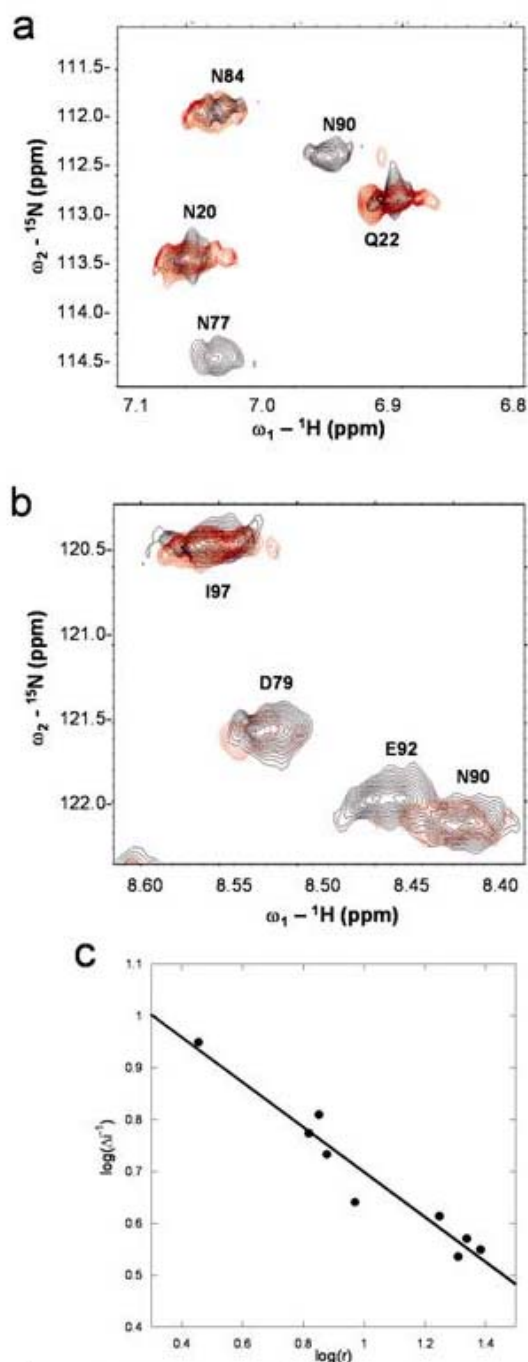


Figure 5.14. Mn^{2+} relaxation of the HSQC spectra. (a-b) HSQC spectra of the protein in the absence (black) and presence (red) of $10 \mu\text{M}$ Mn^{2+} . (c) Correlation of relaxation rate to amide distance of residues A9, F21, E5, D71, D79, V80, N90, E92, and K91 at $10 \mu\text{M}$ Mn^{2+} .

5.7 CD2.6D79 binds CD48 with an enhanced affinity to the target protein

The biological function of CD2.6D79 was examined using surface plasmon resonance (SPR) to measure binding to its target molecule CD48 using previously reported methods (Davis et al., 1998a; Yang et al., 2005). As shown in Figure 5.15, a three-fold decrease in the dissociation constant for CD48 binding was observed for CD2.6D79 compared with wild type CD2 in the absence of metal (1 mM EGTA). Addition of 10 mM Ca^{2+} or 0.10 mM La^{3+} did not significantly alter the affinity of CD48 binding to CD2.6D79. Our results suggest that CD2.6D79 with two additional charged residues enhances its binding to its target molecule CD48. This three-fold increase in binding affinity is significant, since the majority of charged ligand mutants reported to date, except R87A (5- fold increase), show either a decrease or less pronounced affinity change (Davis et al., 1998a; Kim et al., 2001).

5.8 Common metal binding properties for naturally occurring calcium binding proteins

Calcium regulates many biological processes through interacting with proteins with different conformational, dynamic and metal binding affinity. According to the survey of more than 500 naturally occurring calcium binding proteins and small chelators, Ca^{2+} binding sites have the most common coordination of 6-7 with the pentagonal bipyramidal geometry (Pidcock and Moore, 2001; Yang et al., 2002b). Ca^{2+} ligand residues arise from the turn/loop, helix, and β -sheet secondary structural elements in 72.5, 13.9, and 13.5% of cases, respectively (Pidcock and Moore, 2001). Our lab has previously reported that different classes of calcium binding sites in naturally occurring

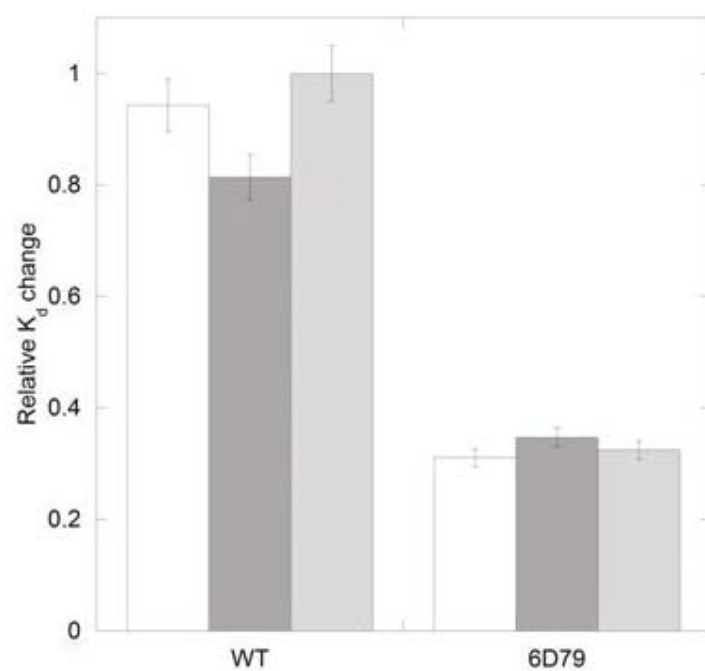


Figure 5.15. CD48 binding of CD2.6D79. The relative binding of CD2.6D79 and CD2 to CD48 in the absence of metal (1 mM EGTA, white bars) as well as in the presence of 100 μM La^{3+} (grey bars) or 10 mM Ca^{2+} (light grey bars).

calcium binding proteins can be identified and described using the popular pentagonal bipyramidal geometry in combination with the side chains of Asp, Glu, Asn, Gln and main chain oxygen in the protein. In principle, residues located at the adjacent β -sheets are capable of satisfying the geometrical requirement for a Ca^{2+} -binding site (Figure 5.1). It was quite puzzling why nature does not use this form of pocket for Ca^{2+} binding given the large range of affinities and the diversified roles of Ca^{2+} in regulating numerous biological functions. Even for the naturally occurring Ca^{2+} binding proteins with a weak affinity (K_d at 0.1 to 3 mM), Ca^{2+} ions are largely chelated by the residues at the loop regions. For example, the Ca^{2+} -binding site in concanavalin A, which also has a net charge of -2 and a Ca^{2+} affinity similar to CD2.6D79, is formed by residues D19, D10, N14 and Y12 from the loop regions (Bouckaert et al., 2000).

5.9 Implications of the designed novel Ca^{2+} binding protein

The design study of CD2.6D79, whose Ca^{2+} binding site is composed of ligand residues all from the β -sheets, reveals several important features for Ca^{2+} binding. First, introducing the clustered charged ligand residues for Ca^{2+} binding results in a large decrease in thermal stability although the overall structural, conformational and folding properties of the host protein are largely retained (Figure 5.2 a,b). As revealed by electrostatic calculations, the charge repulsions between the ligands and the neighboring residues such as E33 significantly decreased the electrostatic potential. Further, in the absence of Ca^{2+} , the fixed backbone conformation of the β -sheets does not allow for a decrease in the charge repulsion through conformational change. It also limits the use of local conformational change to complete neutralization of CD2.6D79 by Ca^{2+} binding.

Hence, calcium binding only leads to a small increase in the thermal stability. Consistently, a greater gain of the protein stability was achieved by binding of La^{3+} and Tb^{3+} (Table 5.1) with a higher positive charge.

Second, Ca^{2+} binding to the all- β -sheet location exhibits a less profound effect on the conformation of the protein than that of previously investigated Ca^{2+} binding proteins (Alattia et al., 1999; Chazin, 1995; Garcia et al., 2004; Ikura, 1996; Ubach et al., 1998; Yang et al., 2005). Ca^{2+} binding to the designed metal binding site in CD2.6D79 was validated by Mn^{2+} relaxation methods. However, the Ca^{2+} or La^{3+} -induced chemical shift changes for the residues at and around the Ca^{2+} binding pocket are very limited, < 0.05 ppm and 0.10 ppm, respectively, and even smaller than residues at the flexible loop regions such as E41 in the same protein (Figure 5.5), suggesting that their shielding environments were not perturbed by Ca^{2+} binding. This is significantly different from the commonly observed global or local Ca^{2+} -induced conformational changes for the intracellular (CaM and TnC) (Chazin, 1995; Ikura, 1996) or extracellular proteins (cadherin and C2 domain) (Alattia et al., 1999; Garcia et al., 2004). Relatively large conformational changes in the Ca^{2+} binding loop regions were still observed in the buffer proteins calbindin D_{9k} and parvalbumin (Chazin, 1995; Ikura, 1996). Similarly, in our previously designed Ca^{2+} binding protein Ca.CD2, significant chemical shift changes (> 0.1 ppm) were observed for the two ligand residues in the loop regions (N60 and D62) while no significant change was detected for the ligand residue Asp-15 at the β -strand B (Figure 5.16). This observed differential response to Ca^{2+} binding is structural specific since Ca.CD2 has a 2.5 fold weaker Ca^{2+} binding affinity than that of CD2.6D79.

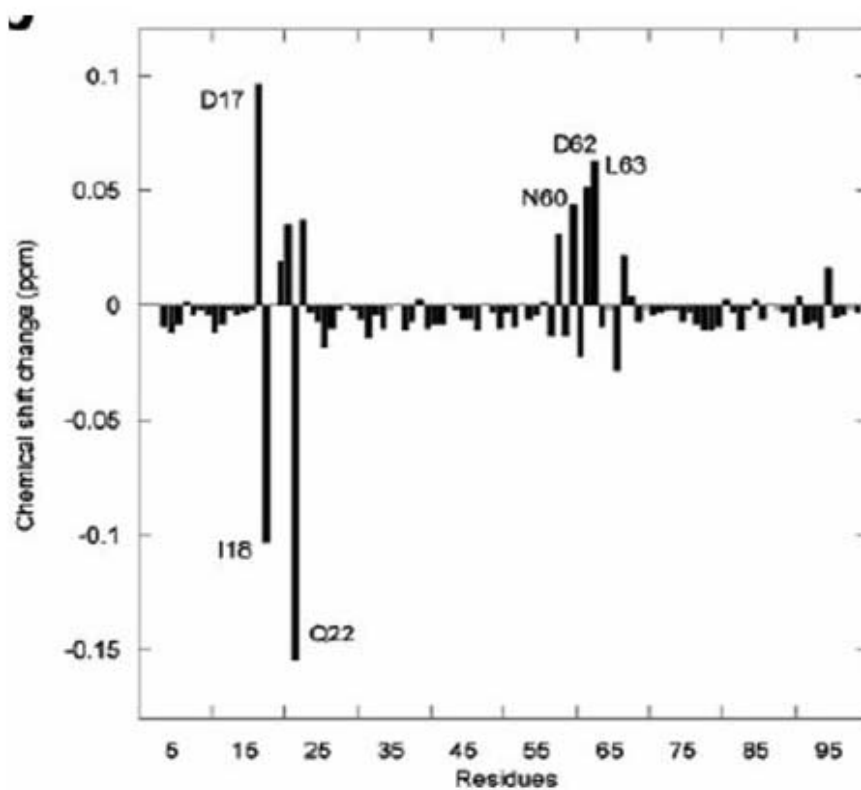


Figure 5.16. Chemical shift changes of Ca.CD2. The limited chemical shift changes of the ligand residues of Ca.CD2 found in β -sheet secondary structure compared to the loop regions

The large energetic cost in destabilizing the protein both in apo- and loaded forms and the lack of Ca^{2+} -induced conformational change explain why nature does not use all β -sheet ligand residues for Ca^{2+} binding. The observed strong preference of Asp and Glu in the naturally occurring Ca^{2+} -binding proteins is consistent with the electrostatic nature of Ca^{2+} binding since almost all the binding sites observed so far have at least two negatively charged ligand residues. The use of the ligand residues predominantly from the loops and flexible regions of the proteins serves to minimize the repulsion between charged ligand residues in the apo-form and maximize the electrostatic interaction of the loaded form by conformational change. Hence, our design studies and structural analysis provide a close view for the structural determinants that are necessary for biological functions of Ca^{2+} binding in regulation and stability.

5.10 Implication for the electrostatic contribution on the cell adhesion

The functional roles of CD2 in initiating cell immune response (Blattman and Greenberg, 2004) require a quick, reversible, and specific interaction with its target (CD48 for rat and CD58 for human CD2). Hence, differing from the strong interactions often observed in enzyme-inhibitor and antibody binding (nM), CD2 binds to CD48 with a fast off rate that limits its affinity to the μM range. The large charged residues in the cell adhesion surface in addition to shape complimentary were proposed to be essential for the target specificity and less or so for the affinity (Davis et al., 1998a). Most single mutations of charged residues on CD2 (to Ala or an opposite charge) decreases (or has no effect) on ligand binding except R87A (CD2.R87A), which results in a 5-fold increase (Davis et al., 1998a). Interestingly, a three-fold decrease in the dissociation constant for

CD48 binding was observed for CD2.6D79 compared with CD2 in the absence and presence of Ca^{2+} or La^{3+} (Figure 5.15). The results suggest that the CD48-binding ability of CD2 is enhanced by altering the electrostatic environment with two charged mutations for the Ca^{2+} binding.

Figures 5.17 and 5.18 show the electrostatic potential of residue R87 on the ligand binding surface of CD2, CD2.R87A, and CD2.6D79. The calculation was performed using DelPhi and visualized using GRASP as described in **section 2.11**. The mutation in CD2.R87A reverses the potential of residue 87 from positive in CD2 to negative. A similar decrease in potential is observed at residue R87 in CD2.6D79 with the introduction of two negatively charged Ca^{2+} binding residues. Calculations on the single mutants indicate that this change is due to the T79D rather than the A92E mutation, which are 3.2 and 10.7 Å away from R87, respectively. Therefore, long range electrostatic steering introduced by the charged mutations for the Ca^{2+} binding is responsible or at least part to the increased binding affinity of CD2.6D79 to CD48. Long range electrostatic steering has been shown to be important in molecular recognition systems including the interaction of the extracellular ribonuclease barnase and its natural inhibitor barstar. The binding sites of barnase and barstar have complementary charges and the charge distributions are optimized for tight binding (Lee and Tidor, 2001). Charged residues close to the binding site enhance the associate rate of the barnase-barstar binding.

Electrostatic switches have often been observed or proposed to be a major player in regulating ligand binding, including phosphorylation/dephosphorylation, lipid and

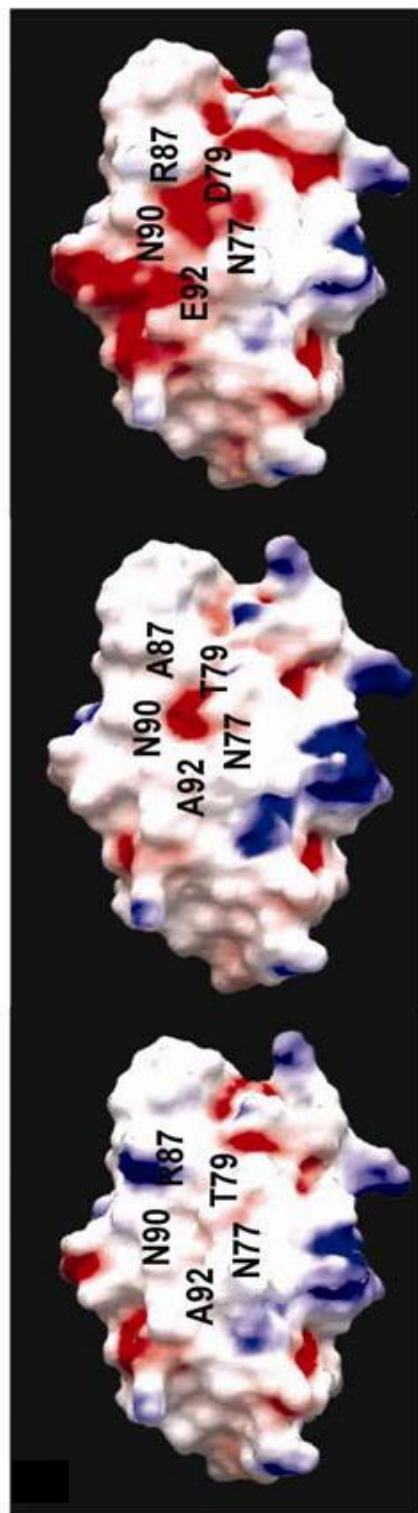


Figure 5.17. Surface potential of the CD48 binding surface. The electrostatic potential was displayed with GRASP, of CD2 (left), CD2.R87 (middle) and CD2.6D79 (right) show the changes in potential that occur at residue 87 in each protein. Potentials less than -7kT are red, those greater than 7kT are blue, and neutral potentials are white.

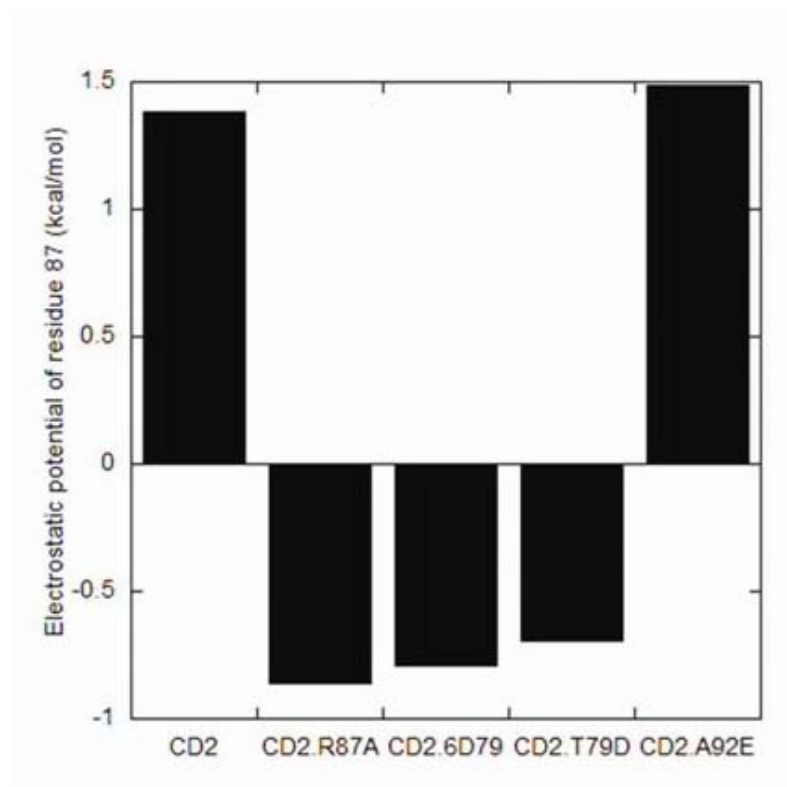


Figure 5.18. Electrostatic potential of R87. Electrostatic potential calculations of residue 87 in the proteins CD2, CD2.R87A, CD2.6D79, CD2.T79D, and CD2.A92E with a grid size of 165 and an interior dielectric constant of 4 revealing the negative potential of that residue in CD2.6D79 and CD2.R87A, which accounts for increased charge complementarity between these two proteins and CD48.

membrane binding, and protein-protein interactions. For example, the dynamic and weak binding to lipid membranes by C2 domain proteins, which also have a similar Ig fold to CD2, was made possible due to the neutralization of the negatively charged surface by Ca^{2+} binding. The binding of CD2 to CD48 has been characterized by a relatively low binding affinity with a very fast off rate which allows for reversible cell-cell interactions (McAlister et al., 1996). Dynamic and weak protein-protein interactions (K_d at μM instead of nM range) usually involve interaction surfaces of 500 -1000 Å (Nooren and Thornton, 2003). Calcium binding does not have a strong effect on the ligand binding capability to CD48 by CD2.6D79 which might imply that subtle environmental effects of adjacent residues exhibit a contribution to the molecular recognition. The detailed examination of the binding characteristics of CD2 and CD48 will provide insight into the properties of cell to cell interactions since this binding mode is a distinguishing characteristic of many types of molecular recognitions especially for cell adhesion proteins. Further analysis of how calcium binding relates to biological function can certainly be benefited from the development of force fields that accurately measure metals in continuum methods (Schymkowitz et al., 2005). Currently, we are investigating the contribution of protein environment to calcium binding and molecular recognition using the reported design approach.

In summary, we have successfully introduced a novel Ca^{2+} binding site at an all β -sheet regions in the molecular-recognition surface of CD2. The protein, CD2.6D79, selectively binds to Ca^{2+} as well as its analogues Tb^{3+} and La^{3+} and maintains a wild type like structure. The small Ca^{2+} -induced conformational changes and the decreased thermal

stability caused by the all β -sheet Ca^{2+} site reveal the structural requirement of Ca^{2+} binding in regulating biological processes. In addition, the CD48-binding affinity of CD2 was increased with the introduction of the negatively charged Ca^{2+} ligands, elucidating the role of electrostatics in modulating cell adhesion and molecular recognition. Furthermore it demonstrates the power and impact of protein design providing insight into the structure, folding, and function of proteins as well as the potential in developing new biomaterials, sensors, and pharmaceutical drugs as well as in intervening biological systems with tailored functions (Benson et al., 2001; Bolon and Mayo, 2001; Corey and Corey, 1996; Di Costanzo et al., 2001; Lu et al., 2001; Pinto et al., 1997; Wang et al., 1999).

Chapter 6. Conformational analyses and stability of the charge distribution variants of CD2.6D79

6.1 The rationale for studying the charge distribution variants

As discussed in chapters 1 and 5, the designed protein CD2.6D79 exhibits strong metal binding affinities and metal selectivity while retaining a CD2-like conformation. Therefore this protein can be used as a model system for investigating the role of electrostatic interactions on Ca^{2+} binding affinity without the condition of conformational entropy. As shown in Figure 3.7 the surface where the metal binding site of CD2.6D79 is located is highly charged. These charged residues have been mutated in the past to understand their role in biological function (Davis et al., 1998a). The electrostatic environment around the metal binding site can be altered through site-directed mutagenesis thereby providing a different approach compared to the study of the first generation proteins. In this chapter, first, we will discuss our criteria in selection of the charge variants to understand the role of electrostatic interactions. Second, the stability of the different charged variants will be examined and compared to the calculation discussed in **section 3.8** using Fold-X and DelPhi.

Protein stability is an important factor for the understanding of the protein folding process. The net stability of a protein is defined by the difference in free energy between the native and denatured state. The stabilization of proteins is the result of a delicate balance among countervailing forces including hydrogen and disulfide bonds and electrostatic and hydrophobic interactions (Li et al., 2005). A more rigid, highly efficiently packed conformation and metal binding has also been shown to contribute to

protein stability (Leone et al., 2004). In recent years, progress has been made in understanding the forces responsible for protein stability (Lee et al., 2005; Strickler et al., 2006). The role of surface electrostatic interactions in protein stability has been somewhat controversial. A case can be made that the contribution of surface charged residues to protein stability may be reduced due to their exposure to a high dielectric solvent (Schueler-Furman et al., 2005). However, several studies have shown that surface charge interactions do play a role in protein stability (Kumar et al., 2000; Loladze et al., 1999; Permyakov et al., 2005; Spector et al., 2000; Xiao and Honig, 1999).

The electrostatic description of proteins consists of short-range and long-range electrostatic interactions both of which play an important role in the stability of proteins. Short-range electrostatic interactions are characterized by salt bridges. The requirement of salt bridge formation is two-fold. First, the centers of the side-chain charged group atoms in the oppositely charged residues must lie within 4.0 Å of one another. Second at least one pair of side-chain carbonyl oxygen and nitrogen atoms of these residues also must lie within this distance (Barlow and Thornton, 1983; Kumar and Nussinov, 1999). Long range electrostatic interactions involve charged groups with greater distances between one another.

Optimized electrostatic interactions through increasing the number of salt bridges have been shown to contribute to the enhanced thermotolerance of proteins in hyperthermophilic microorganisms (Karshikoff and Ladenstein, 2001). The amounts of Glu, Lys, and Arg residues are higher in thermophilic proteins than in mesophilic proteins suggesting that there is a higher number of salt bridges present in thermophilic

proteins than in equivalent mesophilic proteins. However, the presence of electrostatic interactions can also be destabilizing. Studies have shown the majority of salt bridges in proteins are stabilizing and the majority of long range electrostatic interactions are destabilizing (Kumar and Nussinov, 2002b). Salt bridges stabilize proteins through increasing the rigidity of local regions for better fits for the functional requirements of the proteins (Sinha and Smith-Gill, 2002).

The role of Ca^{2+} binding in protein stability has been well characterized. Several proteolytic enzymes, including some members of the subtilisin and trypsin family of serine proteinases, the heat stable metalloproteinase thermolysin, DNase I, amylase, and dihydrofolate reductase, require specific Ca^{2+} concentrations for the maintenance of structural integrity and/or protection against proteolytic digestion (McPhalen et al., 1991). In the majority of cases the Ca^{2+} binding sites are found within the flexible loops of the proteins at least 10 Å away from the catalytic residues and are not directly involved in catalysis (Leszczynski and Rose, 1986). It is presumed the role of Ca^{2+} is to order the residues in the floppy loops (Strynadka, 1993), aid in the maintenance of secondary structure, and reduce the susceptibility to proteolytic digestion at these exposed areas of the enzyme. This phenomenon is observed in the metalloproteinase thermolysin which has a thermostability and the four bound Ca^{2+} ions are 17.8, 19.1, 19.4, and 31.4 Å from the catalytic site of the enzyme (English et al., 2001).

Previous studies have shown that our host protein CD2 retains its native structure at pH from 1-10 which is unique for proteins sharing its Ig-fold (Yang et al., 2000a). It has been proposed that this strong stability against pH is attributed to the large solvent

accessibility of the charged residues of this protein, which are clustered on the GFCC'C" surface, and the lack of an intra-chain disulfide bonding (Yang et al., 2001a). Most Ig-like proteins contain a conserved disulfide bond buried in their hydrophobic core (Sali et al., 1991). Additionally, CD2 is stable with increasing salt concentrations up to 4 M (Yang et al., 2002a). However, upon the increase in temperature from 10 – 90 °C the protein monitored by CD signal at 222 nm undergoes a melting transition that is consistent with a two-state model with a T_m of 66 ± 1 °C in PBS buffer (Yang et al., 2001a). These studies suggest that the native β -sheet structure of CD2 is mainly stabilized by side chain interactions due to the tight packing of the protein.

We have previously reported our success in designing the novel Ca^{2+} binding protein CD2.6D79 (chapter 5) (Jones, 2006). This protein has a decreased thermal stability compared to the wild type protein CD2, 49 ± 1 °C for CD2.6D79 and 62 ± 1 °C for CD2 in Tris buffer (Table 5.1, Figure 5.6). However, upon the addition of Ca^{2+} the stability of the protein increased (53 ± 1 °C). Further, the binding of the trivalent Ca^{2+} analogue Tb^{3+} produced a greater stabilization effect (58 ± 2 °C). The binding of Ca^{2+} and Tb^{3+} was a stabilizing factor through a decrease in the electrostatic repulsion of residues directly involved in Ca^{2+} binding. Here we report the examination of the effect short- and long-range effect on protein stability using this model system. The surface charge-charge interactions of the protein were altered through single mutations. This work is important for the engineering of thermostable enzymes which will be useful in the areas of biotechnology.

6.2 Selection of the charge distribution variants

As discussed in **section 3.5**, the selection of the residues to be mutated was based on several criteria including distance to the metal binding site and solvent accessibility. First, the electrostatic surface was visualized using GRASP to identify charged residues near the metal binding site which could be mutated. Second, the distance of each residue was ascertained using the model structure of 6D79, a distance requirement of 15 Å was employed. Third, the electrostatic potential of the metal binding site was calculated with the introduction of charge reversal mutations at the potential residues. Fourth, the solvent accessibility of each potential residue was calculated using the program NACCESS (Hubbard et al., 1991; Hubbard and Thornton, 1993) with a cutoff value of 20%. Fifth, the residues directly involved in CD48 binding, such as D28, D29, R31, E33, and R87 (Davis et al., 1998a) were eliminated so the functional properties of the protein would not be altered.

Table 6.1 shows the residues chosen for mutation as well as their distance to the metal binding site, solvent accessibilities, and electrostatic potential changes. Figure 6.1 shows the model structure of 6D79 with all of the chosen residues highlighted. For each residue selected, two mutants were engineered, one to the opposite charge residue and one to a neutral residue (Table 6.2). All charge distribution variants were engineered by PCR, expressed in *E. coli*, and purified by affinity chromatography using the protocols established in **sections 2.2-2.5**. This chapter will focus on the K91, D94, and T75K variants due to their stable β -sheet conformation. The R34 variants will be discussed in chapter 8.

Residue	Distance from Ca ²⁺ binding site ^a Å	Δelectrostatic potential ^b kcal/mol	% solvent accessibility ^c
K91	11.4	-41.6	36.1
D94	14.1	30.3	43.4
T75	7.0	--	71.0
R34	9.5	-39.7	62.3

^aMeasured from the model structure of 6D79 using Pymol

^bCalculated using DelPhi (condition see section 3.5)

^cCalculated by Naccess (Hubbard and Thornton, 1993; Hubbard et al., 1991)

Table 6.1 Selection of the charge distribution variants. Charged residues on the surface of site 6D79 which have been identified for mutation

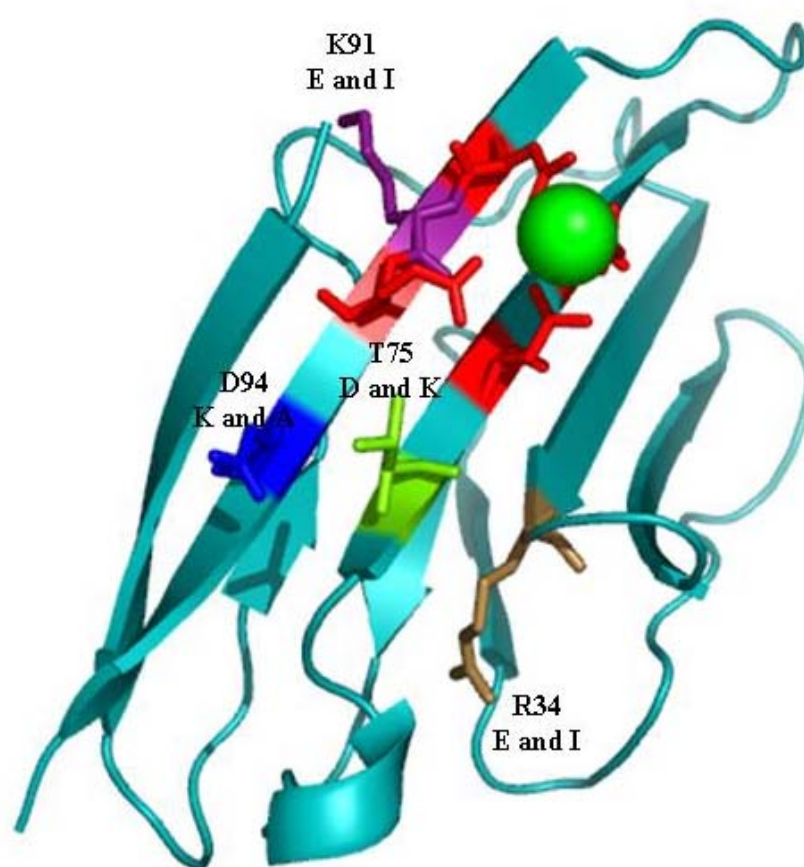


Figure 6.1. Positions of the charge distribution variant. The residues chosen for mutation to study the role of electrostatic interactions in Ca^{2+} binding affinity include R34, T75, K91, and D94.

Protein	34	75	79	91	92	94
wild type	R	T	T	K	A	D
CD2.6D79	R	T	D	K	E	D
6D79.K91E	R	T	D	E	E	D
6D79.K91I	R	T	D	I	E	D
6D79.D94K	R	T	D	K	E	K
6D79.D94A	R	T	D	K	E	A
6D79.T75K	R	K	D	K	E	D
6D79.T75D	R	D	D	K	E	D
6D79.R34E	E	T	D	K	E	D
6D79.R34I	I	T	D	K	E	D

Table 6.2. The primary sequence of the charge distribution variants. The difference in protein sequences of the charge distribution variants compared to 6D79 and CD2.

6.3 The 6D79.K91 Variants

Residue K91 is located 7.7 Å from the metal binding site of 6D79 and its side chain faces the opposite direction (Figure 6.1). This residue was mutated to the negatively charged residue Glu (6D79.K91E) and to the neutral residue Ile (6D79.K91I). The side chain of the ligand is positioned towards the interior of the protein even though it is a polar residue. Additionally, the residue is in close proximity to Asp2, therefore mutating to a negative residue may cause a conformational change of the protein base on electrostatic repulsion (Figure 6.2).

6.3.1 Conformational analyses of the 6D79.K91 variants

The conformation of the K91 variants was examined using far UV CD, Trp fluorescence, and 1D proton NMR. For the far UV CD analysis the protein was diluted to 3 µM in 10 mM Tris pH 7.4. The CD spectra indicate that both K91E and K91I have a negative maximum at 216 nm which is indicative of a β -sheet protein (Figure 6.3a). This conformation is similar to that of wild type CD2 and 6D79. When excited at 283 nm, the Trp emission fluorescence measurements reveal an emission maximum of 326 nm, which signifies the tertiary packing of both of the variants is intact (Figure 6.3b). The differences in intensity of the spectra are indicative of varying protein concentrations. The far UV CD samples were used for the fluorescence measurements and the inadvertent dilution of samples by improper drying of the CD cell may attribute to the differences in the intensity of the fluorescence spectra.

One dimensional ^1H NMR using a 500 MHz Varian instrument was employed to examine the detailed conformation of the 6D79.K91 variants in 20 mM PIPES, 10 mM

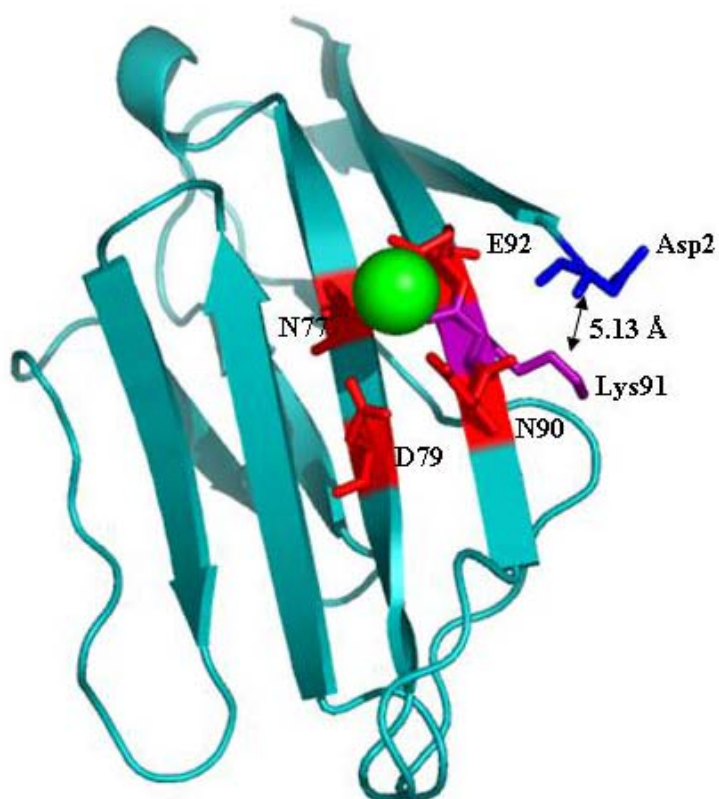


Figure 6.2. Proximity of Asp2 to Lys91. The close proximity of Asp2 and Lys91 may cause electrostatic repulsion when the K91E mutation is engineered.

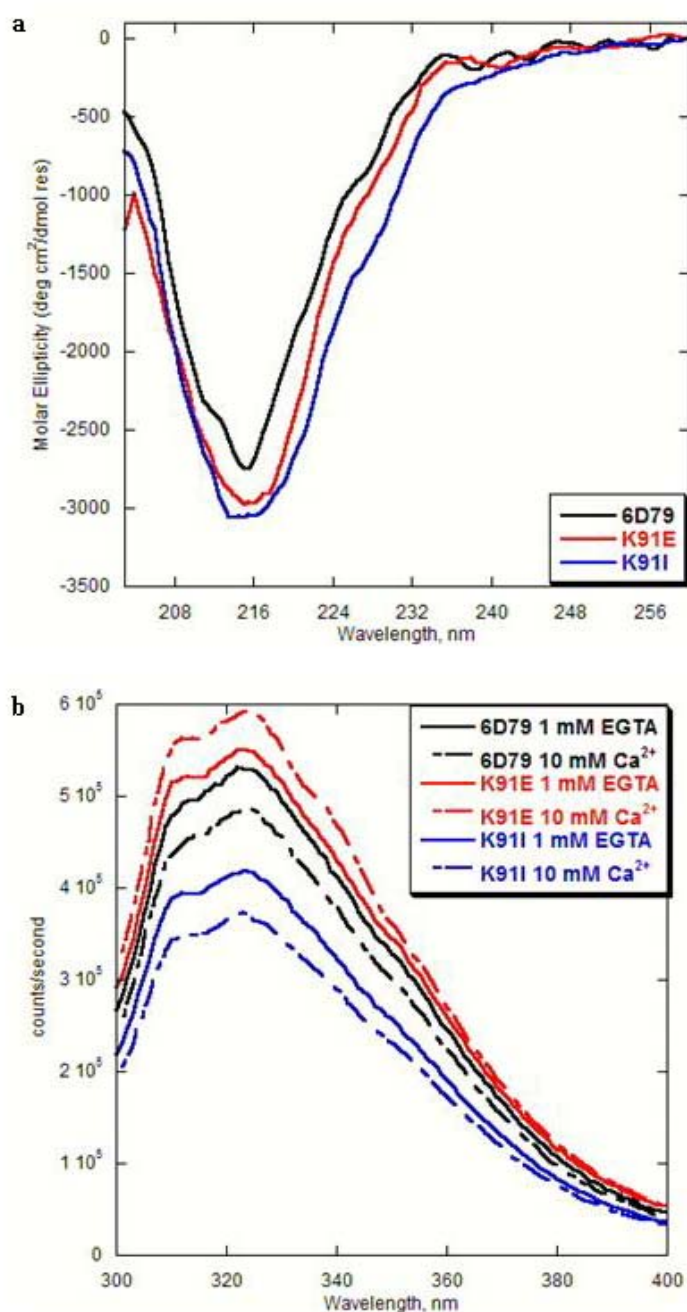


Figure 6.3. Conformation of the 6D79.K91 variants. Conformational analyses of 6D79.K91E and 6D79.K91I compared to 6D79 using (a) far UV CD and (b) Trp fluorescence emission spectra (excitation 283 nm) in 10 mM Tris pH 7.4.

KCl pH 6.8 at 25 °C in comparison to CD2.6D79 (Figure 6.4). The concentration of protein was 120 and 89 μ M, respectively, for 6D79.K91E and 6D79.K91I in 20 mM PIPES, 10 mM KCl pH 6.8. The amide region of the spectra indicates the difference in the tertiary packing of the three proteins. The aromatic residues W32 and Y76 are indicators of the tertiary packing of the protein. W32 is buried in the hydrophobic core and therefore shielded from solvent. This is observed in the comparison of the NMR spectra of CD2 and 6D79 where the W32 has a higher chemical shift (10.53 for CD2 and 10.47 for 6D79) and therefore a tighter tertiary packing (Figure 6.4, Table 6.3). Y76 is hydrogen bonded to W32 so changes in the chemical shift of this residue also provide insight into the tertiary packing of this protein.

Figure 6.4 shows the tertiary packing of 6D79.K91E is similar to that of 6D79 with no significant shifting of the W32 peaks of the two proteins. However, the W32 peak for 6D79.K91I is shifted downfield indicating the tertiary packing of the protein has been altered. An examination of the structure of the crystal structure of CD2, modified to introduce the binding site of 6D79, reveals that K91 is pointed inwards towards the hydrophobic core even though it is a charged residue (Figure 6.5). When K91 is mutated to Ile, the hydrophobic residue, in an attempt to remove itself from solvent, may shift to a more buried position in the protein. This may be the cause of the reduced tertiary packing of 6D79.K91I. Further evidence of this is observed in the side chain region of the NMR spectra (Figure 6.6). In the 6D79 and 6D79.K91E spectra, three distinct peaks are observed at \sim 0.1 ppm, corresponding to the hydrophobic residues V78, V39, and L16.

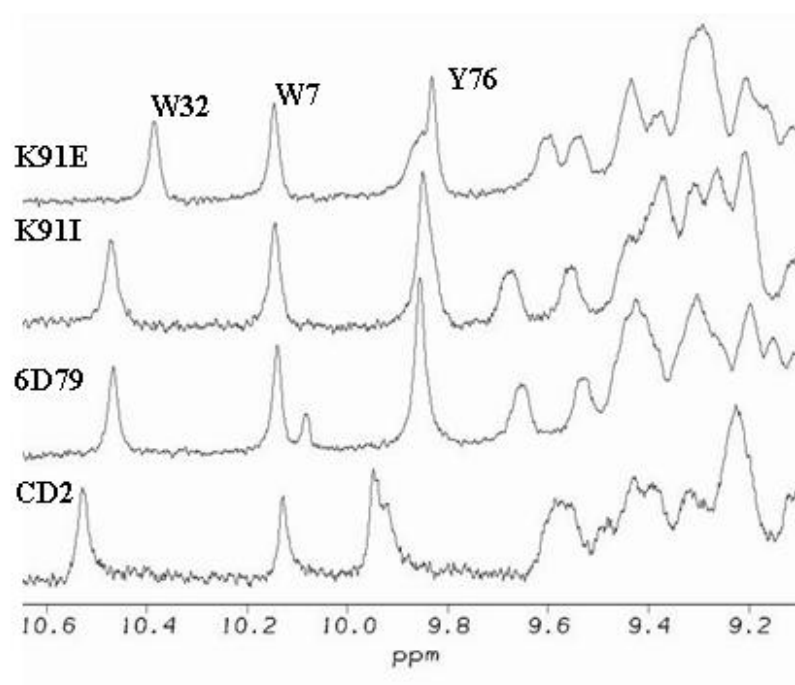


Figure 6.4. NMR of the 6D79.K91 variants amide region. 1D ^1H NMR of CD2, 6D79, K91E, and K91I in 20 mM PIPES, 10 mM KCl pH 6.8 at 25 °C. The chemical shift of W32 and Y76 in the amide region of the proteins shows the tertiary packing of K91I is different than 6D79, K91E, and CD2.

Protein	W32 chemical shift (ppm)	Y76 chemical shift (ppm)
CD2	10.53	9.93
CD2.6D79	10.47	9.86
6D79.K91E	10.47	9.85
6D79.K91I	10.39	9.83

Table 6.3. Comparison of the chemicals shifts of the K91 variants amide region. The chemical shifts of residues W32 and Y76 in CD2, CD2.6D79, 6D79.K91E, and 6D79.K91I indicate the tertiary packing of the proteins is different.

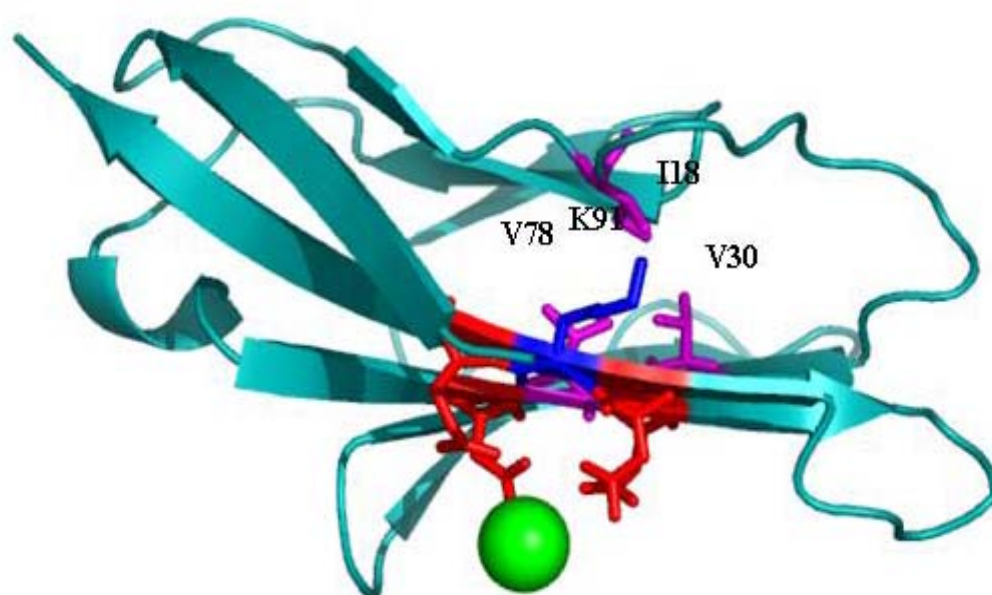


Figure 6.5. The side chain of Lys91 is near the hydrophobic core of the protein. Residue K91 points in towards the hydrophobic core of the protein even though it is a charged residue. This may account for the perturbed tertiary packing of K91I as compared to 6D79 and K91E

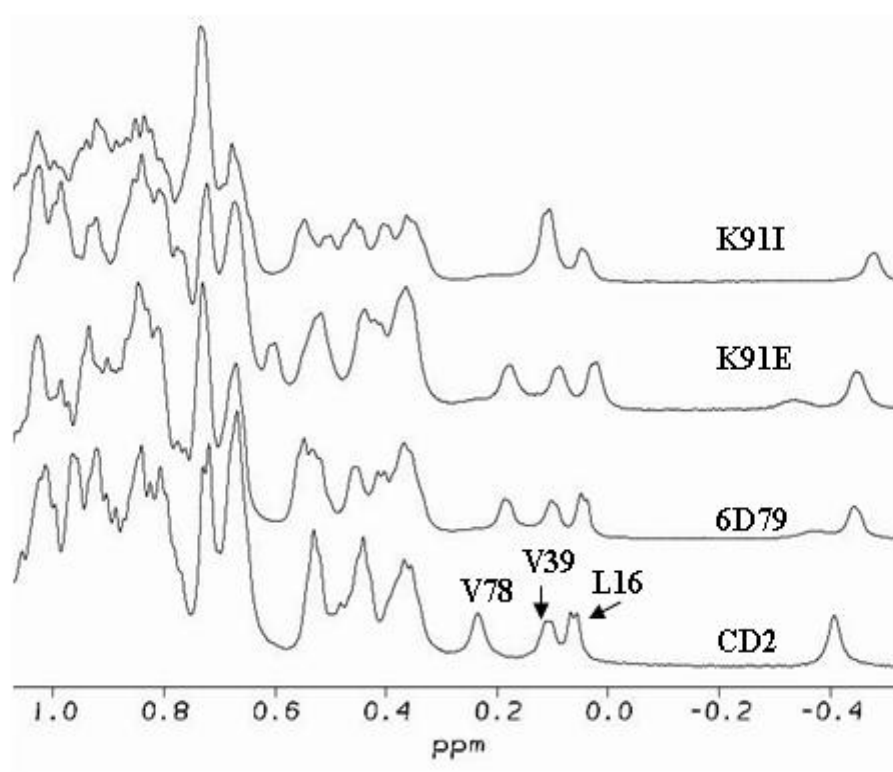


Figure 6.6. NMR of the K91 variants side chain region. Side chain region of the NMR spectra of 6D79, K91E, and K91I in 20 mM PIPES, 10 mM KCl pH 6.8 at 25 °C displays the peak differences in the K91 spectra at 0.1 ppm.

However in the 6D79.K91I spectra only two peaks are visible, one larger than the other. This may indicate that I91 is overlapping with the other hydrophobic residues generating the large peak due to these residues being in the same environment.

The 2D spectra of 6D79 indicate the chemical shift of the amide proton of K91 is 8.7 ppm. This region of the 1D spectrum was examined to determine whether any significant chemical shift changes have occurred in this region (Figure 6.7). There are several chemical shift changes in this region, but there is also significant spectral overlap, hence distinguishing any individual residue by 1D NMR shifting is difficult.

Further assessment of the conformation of 6D79.K91E was accomplished using 2D ^{15}N HSQC experiments using ^{15}N -labelled protein in 20 mM PIPES, 10 mM KCl pH 6.8. The 6D79.K91E spectrum was assigned and compared to the 6D79 spectra under the same condition. Several changes were observed between the two proteins. The peak where residue K91 was in the CD2.6D79 spectra disappeared and there was one new peak that appeared in the 6D79.K91E spectra indicating significant shifting of residue 91 upon mutation. Residue K91 in 6D79 had a chemical shift of 8.7 and 124 in the ^1H and ^{15}N dimensions, respectively, while residue E91 in K91E had a chemical shift of 7.8 and 123.7 ppm (Figure 6.8). Overall, taken together with far UV CD and fluorescence these mutations at positions 91 in K91E and K91I do not cause any significant change in the secondary and tertiary structure of these proteins.

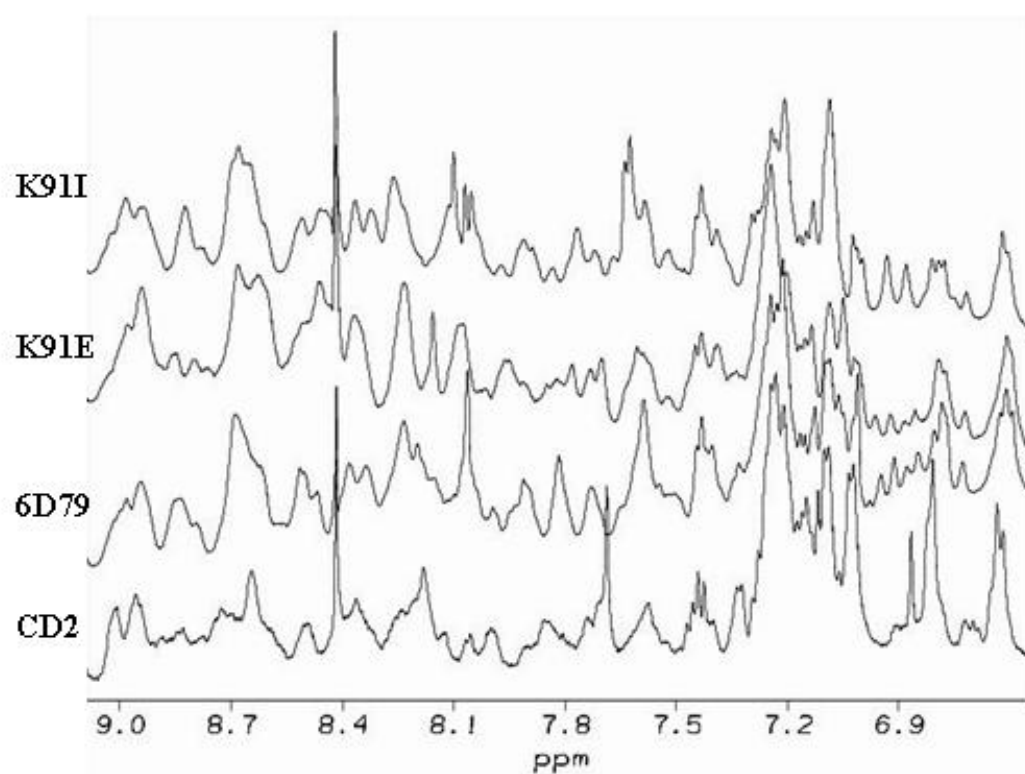


Figure 6.7. Amide region of the K91 variants. 1D ^1H NMR spectra of CD2, 6D79, K91E, and K91I in 20 mM PIPES, 10 mM KCl pH 6.8 at 25 °C.

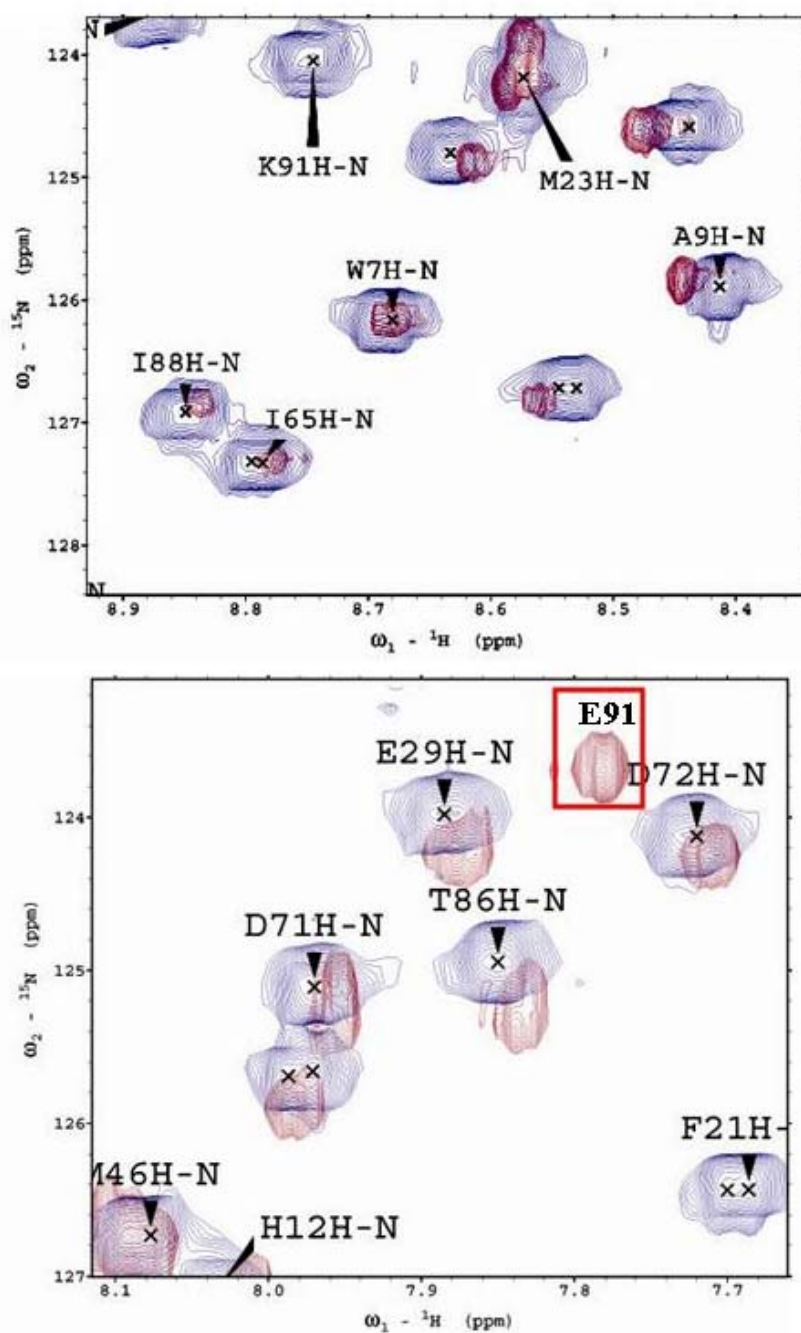


Figure 6.8. HSQC spectra of 6D79 and 6D79.K91E. 2D ^{15}N HSQC spectra of K91E (purple) compared to 6D79 (blue) shows the chemical shift changes of the K91 residue upon mutation in 20 mM PIPES, 10 mM KCl pH 6.8 at 25 °C.

6.4 The 6D79.D94 Variants

Residue D94 is located 14 Å from the Ca^{2+} binding site. Unlike K91, residue D94 is facing the same direction of as the Ca^{2+} binding site. This residue was mutated to the positively charged residue Lys (6D79.D94K) and to the neutral residue Ala (6D79.D94A). D94 is in close proximity to and forms a salt bridge with R96 (Figure 6.9), therefore, any mutation at this position may alter the stability of the protein.

6.4.1 Conformational analyses of the 6D79.D94 variants

The conformational analyses of these variants were examined using far UV CD, Trp fluorescence, and 1D ^1H NMR. Both the far UV CD and Trp fluorescence spectra were obtained with 3 μM protein in 10 mM Tris buffer with a pH of 7.4. The NMR spectra were obtained with 275 and 243 μM of 6D79.D94K and 6D79.D94A, respectively, in 20 mM PIPES, 10 mM KCl pH 6.8. The far UV CD spectra indicate that both D94K and D94A have negative maxima at 216 nm which is indicative of a β -sheet protein (Figure 6.10). This conformation is similar to that of wild type CD2 and 6D79.

One dimensional ^1H NMR collected by 500 MHz Varian instrument in 20 mM PIPES, 10 mM KCl pH 6.8 at 25 °C was employed to examine the conformation of the 6D79.D94 variants in comparison to 6D79 (Figure 6.11). No significant changes are observed for the aromatic residues W32 and Y76 (chemical shift changes < 0.002 ppm) indicating the tertiary packing of the protein has not been altered. Chemical shift changes are seen in the amide region around 9.2 and 7.6 ppm's which are in different regions than the chemical shift of D94 (8.825 ppm) observed in the 2D spectra of 6D79. In addition, the side chain region for the dispersed peaks is not changed. However, unlike the K91I

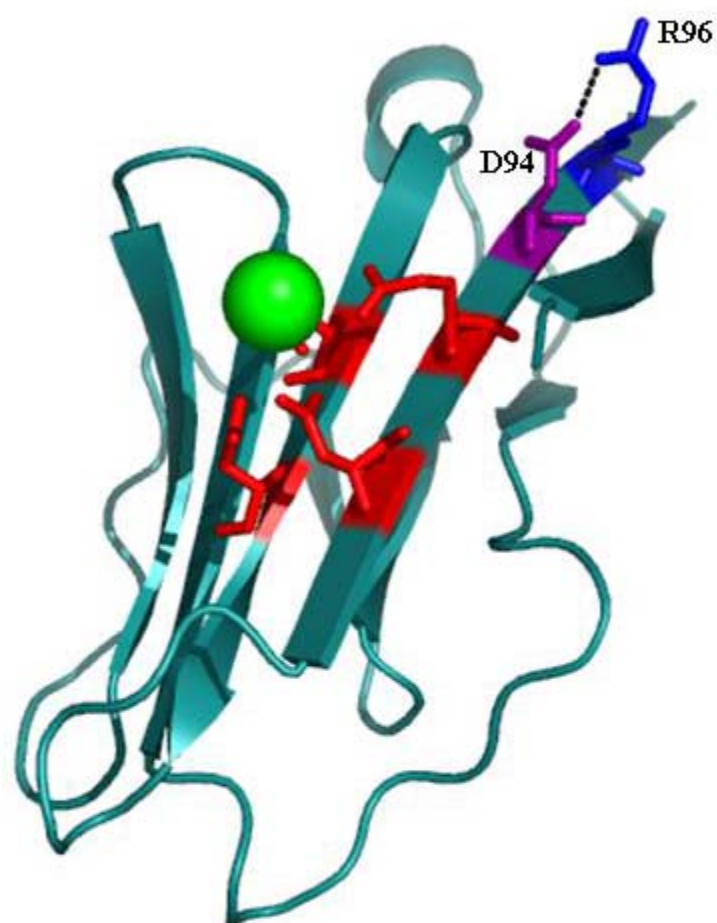


Figure 6.9. Hydrogen bonding of D94 to R96. The model structure of 6D79 using Pymol reveals the side chains of D94 and R96 are hydrogen bonded to one another. The hydrogen bond is indicated by the dashed black line.

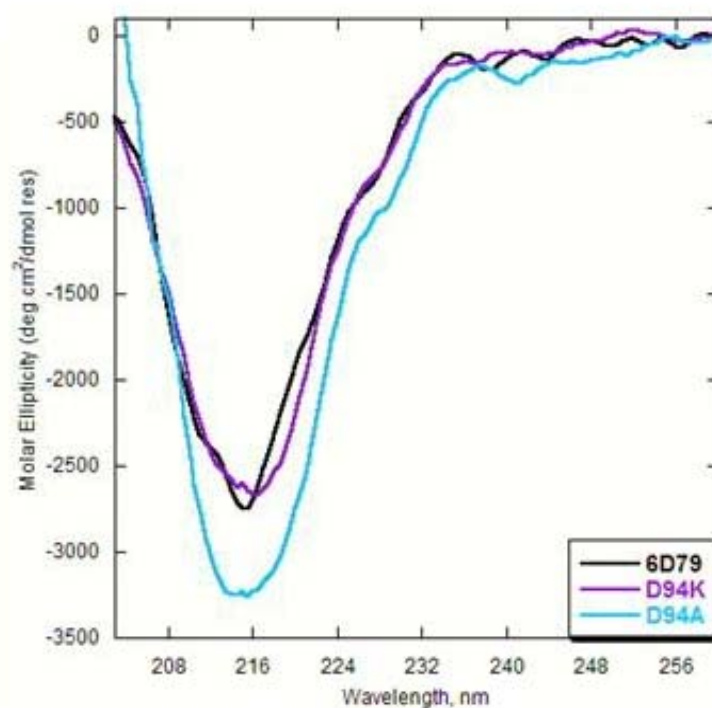


Figure 6.10. Conformation of the 6D79.D94 variants. Conformational analyses of (3 μ M) of both D94K and D94A compared to 6D79 using far UV CD in 10 mM Tris pH 7.4.

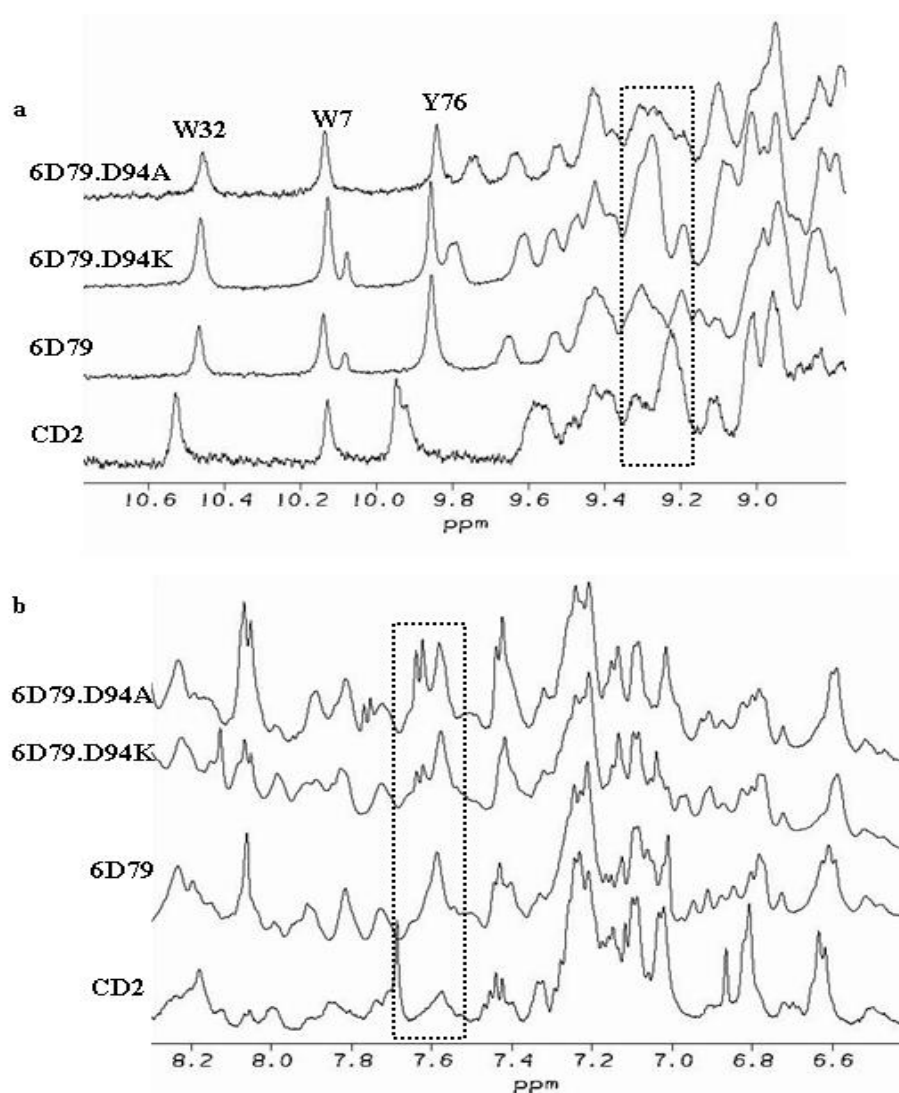


Figure 6.11. NMR of the 6D79.D94 variants. 1D ^1H NMR of CD2, 6D79, K91E, and K91I in 20 mM PIPES, 10 mM KCl pH 6.8. (a) The chemical shift of W32 and Y76 in the amide region of 6D79.D94K and 6D79.D94A are similar to 6D79 indicating the tertiary packing of the protein has not been disturbed. (b) Overlapping in this region of the spectra inhibits the assignment of residue 94 with the engineered mutation.

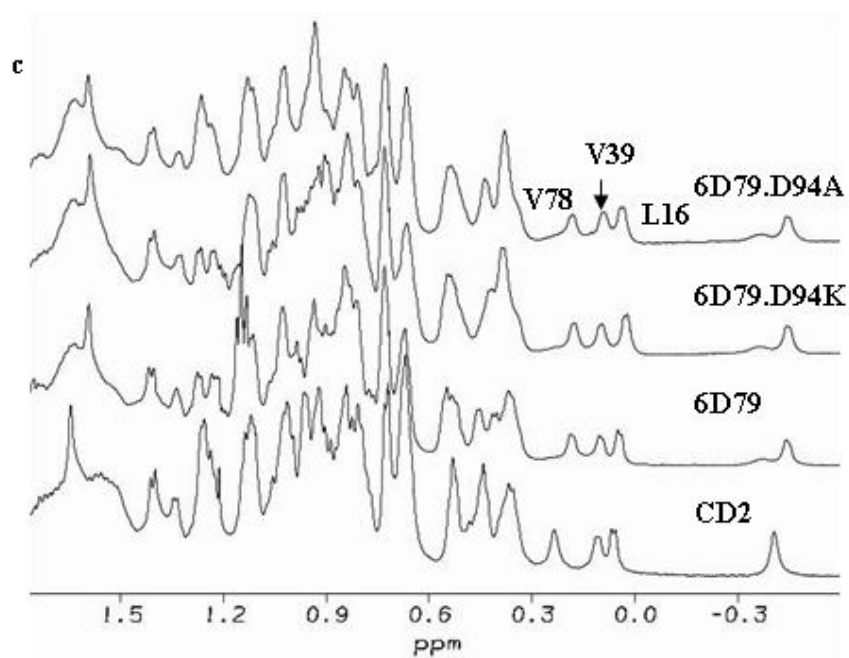


Figure 6.11 cont. (c) The side chain region of all four spectra is very similar.

variants, which have dispersed peaks in this region, the peak shifting of the mutated residues cannot be distinguished due to significant overlapping of peaks in the 1D spectrum especially at δ -CH₂ (1-1.4 ppm) for CD2.D94A. In summary the 6D79.D94 variants do not display a significant conformational change compared to CD2.6D79 with a mutation at position 94.

6.5 The 6D79.T75 Variants

T75 is a neutral residue 8.5 Å from the metal binding site. The side chain of T75 is 61% exposed to solvent. This residue was mutated to the positively charged Lys residue (6D79.T75K) and also to the negatively charged Asp residue (6D79.T75D). Residue T75 is in close proximity to D94 (4.06 Å as calculated by Pymol using the model structure) (Figure 6.12). The mutation of T75D may alter the conformation and/or stability of the protein due to electrostatic repulsion with D94.

6.5.1 Conformational analyses of the 6D79.T75 variants

The conformation of the 6D79.T75 variants was examined using far UV CD, Trp fluorescence, and NMR. The far UV CD and Trp fluorescence measurements were done in 10 mM Tris buffer pH 7.4 (3 µM protein). The NMR spectra of T75K were recorded with a protein concentration of 127 µM in 20 mM PIPES, 10 mM KCl pH 6.8. The far UV CD spectra show that 6D79.T75K has β -sheet architecture similar to CD2 and 6D79 (Figure 6.13) while 6D79.T75D has an alternate, partially folded conformation. Since this altered conformation is only observed in the 6D79.T75D variant and not the 6D79.T75K variant the change can be attributed to the repulsion between residue D75 and the wild type residue D94. In an attempt to reduce the repulsion the D94 residue

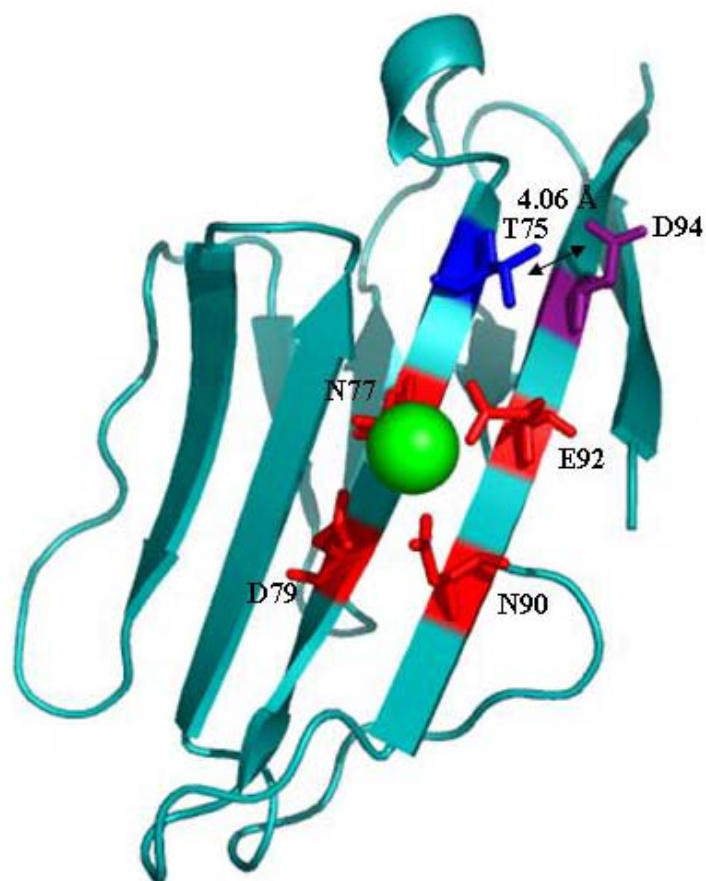


Figure 6.12. Proximity of T75 to D94. The model structure of 6D79 reveals that residues D94 and T75 are within 4 Å of one another

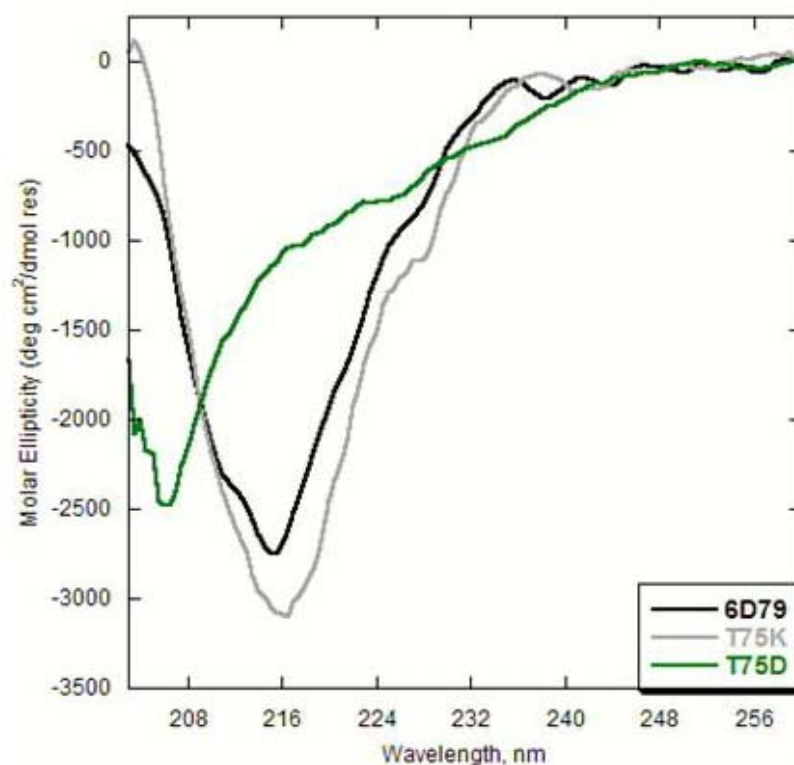


Figure 6.13. Conformation of the 6D79.T75 variants. Conformational analysis of 6D79.T75K and 6D79.T75D compared to 6D79 using far UV CD in 10 mM Tris pH 7.4 with 1 mM EGTA indicate that 6D79.T75D has a partially folded conformation.

could be mutated to a neutral or opposite charged residue. However, D94 is also one of the residues chosen as a charge distribution variant and calculations have shown that its mutation has an effect on Ca^{2+} binding affinity. Any mutation of residue D94 will have an effect on the metal binding site and therefore will not be useful in understanding the T75D mutation. Therefore, the protein stability and metal binding affinities of the variant 6D79.T75K only was characterized.

Further conformational analysis of variant 6D79.T75K was performed using 1D ^1H NMR. The NMR spectra of 6D79.T75K displays a change in signal intensity of the N77 residue peak as compared to 6D79 and CD2 (Figure 6.14). This intensity change may signify an environment change for the N77 residue with the introduction of the positive Lys at T75, which is 7.3 Å away.

6.6 Stability of the charge distribution variants

6.6.1 Calculation of stability using DelPhi

The electrostatic free energy of each of the charge distribution variants was calculated using DelPhi to predict which proteins would be most stable in the apo form. Calculations of the metal bound form of the proteins were not conducted because DelPhi does not predict metal binding accurately. For each calculation, a salt concentration of 10 mM was used to mimic the experimental buffer conditions (10 mM Tris, 10 mM KCl pH 7.0). An interior dielectric constant of 4 was used and the non-linear solution to the PB equation was imposed.

Table 6.4 shows the calculated electrostatic free energies of each of the variants. Variant 6D79.D94A (-240 kcal/mol) has a less negative free energy compared to

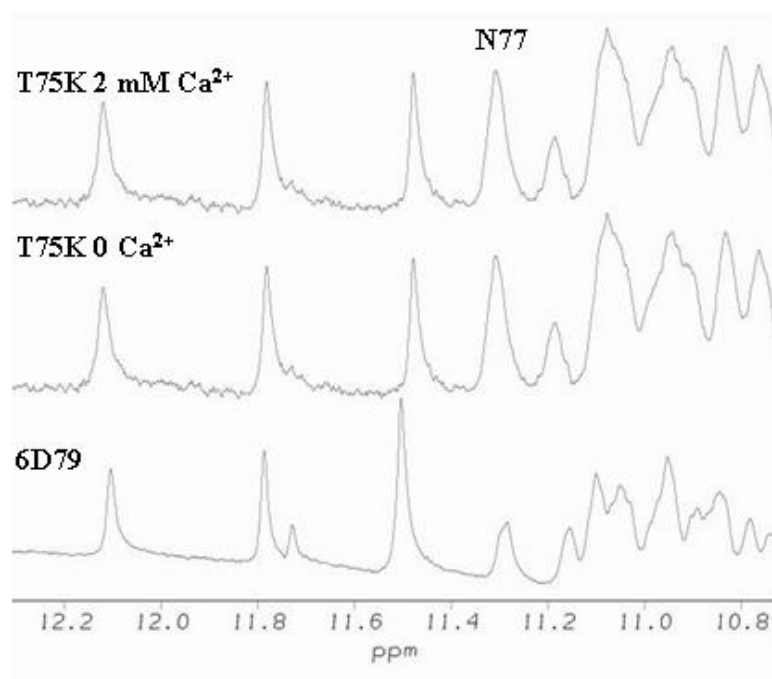


Figure 6.14. NMR of 6D79.T75K. 1D ^1H NMR of T75K compared with 6D79 reveals the environmental difference of residue N77 which is indicated by a peak shape change at 11.3 ppm's.

CD2.6D79 (-253 kcal/mol) indicating the protein is less stable. Variant 6D79, D94K on the other hand has the same free energy as CD2.6D79 signifying the stability of these two proteins is similar. Both 6D79.K91E (-320 kcal/mol) and 6D79.K91I (-275 kcal/mol) have increased negative free energies compared to CD2.6D79 with 6D79.K91E displaying a larger increase. Variant 6D79.T75K has a slightly less negative free energy (-247 kcal/mol) compared to CD2.6D79. The stability of these charge variants was also calculated using the Fold-X server (section 3.8, Table 3.11).

6.6.2 Measurement of thermal stability in the apo form of the CD2.6D79 variants

The T_m of five of the CD2.6D79 variants was measured using far UV CD (Table 6.4). The melting transition of the proteins was observed from 10 – 90 °C. The CD spectra of the CD2.6D79 variants at 25 and 90 °C are very similar to those of CD2.6D79 and CD2 (data not shown) suggesting that both proteins have similar native and thermal denaturation states. The thermal transition curves of the CD2.6D79 variants monitored by CD signal at 230 nm in the presence and absence of Ca^{2+} are consistent with a two-state model as is CD2.6D79 (Jones, 2006) and CD2 (Yang et al., 2002a). The measured T_m value for the Ca^{2+} -free form of CD2.6D79 is 49 ± 1 °C while that for CD2 is 62 °C.

Figure 6.15a and 6.16a shows the change in spectra of the apo form (1 mM EGTA) of K91E and K91I, respectively, at 10 and 90 °C. As the temperature is increased the protein changes conformation and becomes more unstructured. The fractional change of the spectra displays a T_m value of 55.3 ± 1.1 °C (Figure 6.15b) for K91E and 53 ± 2 °C (Figure 6.16b) for K91I in the absence of Ca^{2+} . Both 6D79.K91E and 6D79.K91I have an increased thermal stability compared to CD2.6D79. According to the crystal structure

Protein	Charge	T _m °C EGTA ^a	T _m °C Ca ²⁺ ^b	T _m °C Tb ³⁺ ^c	Electrostatic free energy (kcal/mol) ^d
CD2	-1	62	N/A	N/A	
CD2.6D79	-3	49 ± 1	53 ± 1	58 ± 2	-253
6D79.D94A	-2	47 ± 2	48 ± 2	48 ± 1	-240
6D79.D94K	-1	56 ± 0.1	53 ± 1	55 ± 1	-253
6D79.K91E	-5	55 ± 1	53 ± 1	43 ± 1	-320
6D79.K91I	-4	53 ± 2	46 ± 2	30 ± 1	-275
6D79.T75K	-2	60 ± 1	55 ± 1	53 ± 1	-247

^a10 mM Tris, 10 mM KCl, 1 mM EGTA pH 7.1

^b10 mM Tris, 10 mM KCl, 10 mM Ca²⁺ pH 7.1

^c10 mM Tris, 10 mM KCl, 50 μM Tb³⁺ pH 7.1

^dCalculated with 10 mM monovalent salt in the apo form

Table 6.4. Melting temperature and electrostatic free energy of the CD2.6D79 variants

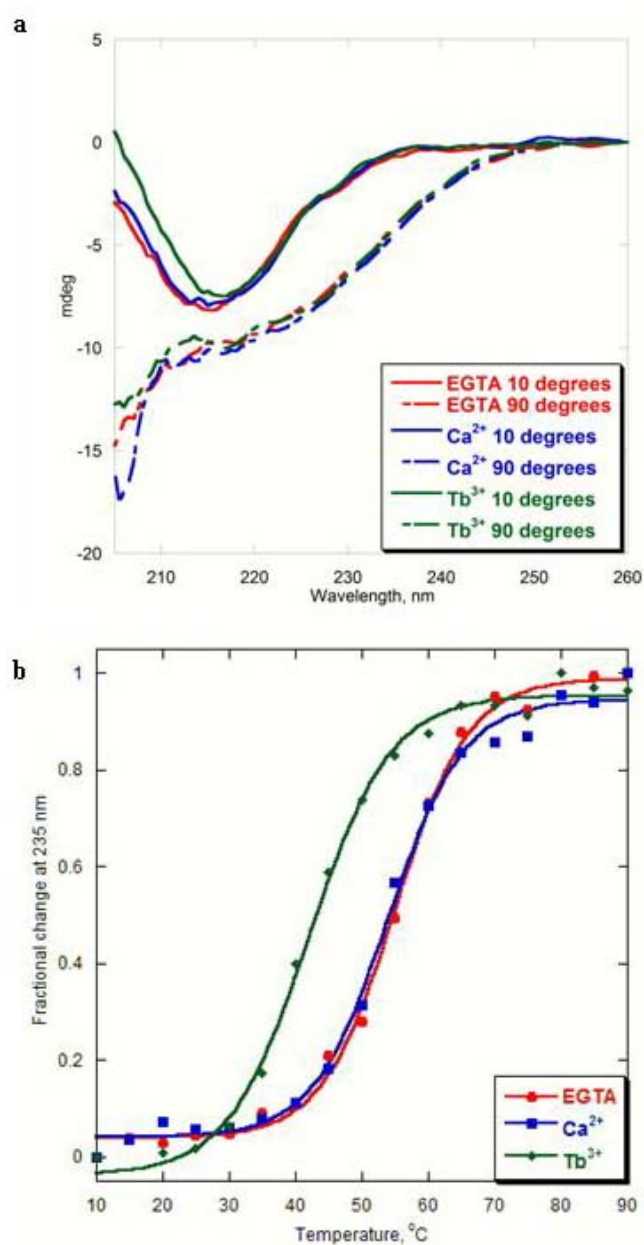


Figure 6.15. Thermal denaturation of 6D79.K91E. (a) Far UV CD spectra of 6D79.K91E (20 μ M) with 1 mM EGTA, 10 mM Ca^{2+} , and 50 μ M Tb^{3+} at 10 and 90 degrees in 10 mM Tris, 10 mM KCl pH 7.0. (b) Fractional change of the far UV signal at 235 nm as a function of temperature.

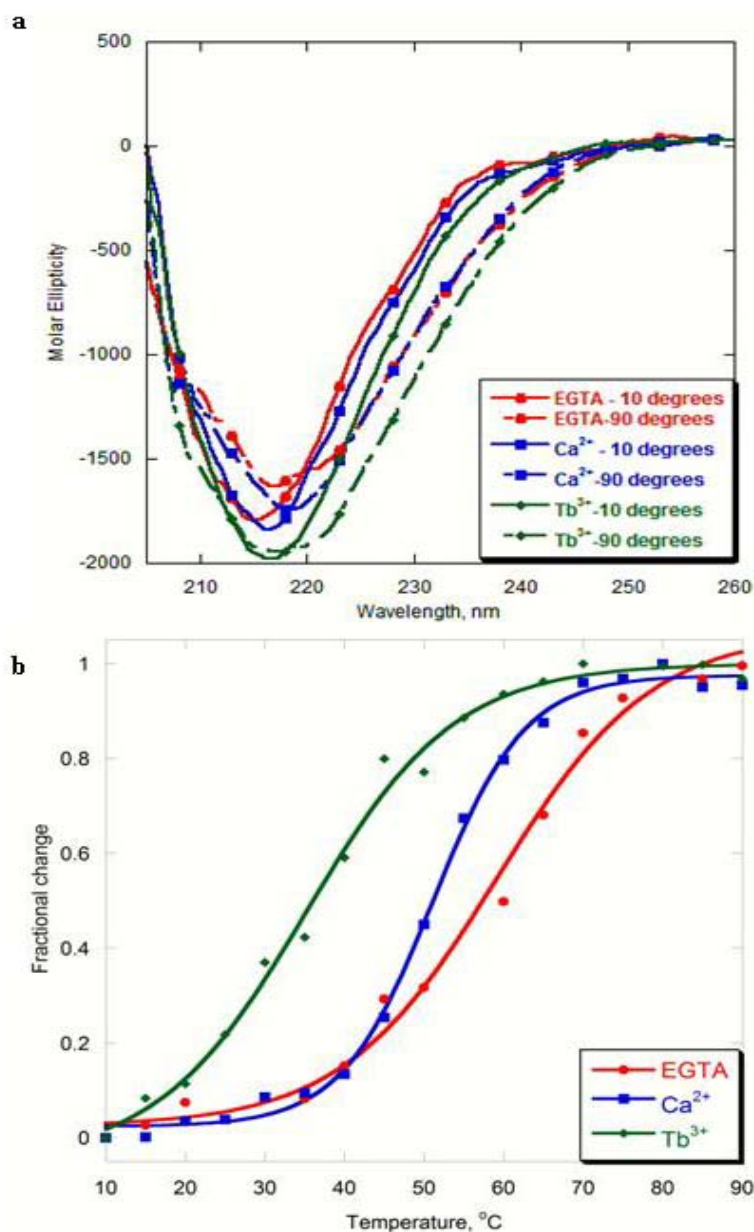


Figure 6.16. Thermal denaturation of 6D79.K91I. (a) Far UV CD spectra of 6D79.K91I (20 μM) with 1 mM EGTA, 10 mM Ca^{2+} , and 50 μM Tb^{3+} at 10 and 90 degrees in 10 mM Tris, 10 mM KCl pH 7.0. (b) Fractional change of the far UV signal at 235 nm as a function of temperature.

of CD2, residue 91 is pointed towards the opposite direction of the metal binding surface of CD2.6D79. Therefore, this residue interacts with other charge residues on the opposite surface of the protein. The removal of the positive charge at position 91 has reduced charged repulsion in the microenvironment of this residue, therefore, increasing the protein stability of the two variants compared to CD2.6D79.

For the D94 variants, the mutation of D94 to Ala has not altered the thermal stability, 47 ± 2 °C (Figure 6.17) compared to CD2.6D79, while for the mutation to Lys the stability of the protein increased, 56 ± 1 °C (Figure 6.18), compared to CD2.6D79. The introduction of a Lys residue at position 94 allows for the formation of a salt bridge with the Ca^{2+} binding ligand E92 (Figure 6.19). The formation of this salt bridge reduces the electrostatic repulsion between the negatively charged residues around the Ca^{2+} binding site including D2 and D79 (Figure 6.19). The electrostatic repulsion of these residues along with D29 and E33 account for the decreased stability of CD2.6D79 compared to CD2 (Figure 6.19). The 6D79.D94A variant does not reduce these electrostatic repulsive effects through formation of a salt bridge and therefore does not increase the stability of the protein.

The variant 6D79.T75K has a thermal stability of 60 ± 1 °C which is similar to wild type CD2 (Figure 6.20). This increased thermal stability compared to CD2.6D79 can be attributed to the formation of a salt bridge between T75K and D94 (Figure 6.12), which are 5.3 Å from one another. The formation of a salt bridge between K75 and D94 increases the protein stability similar to that of wild type CD2. The proximity of D94 also accounts for the altered conformation of 6D79.T75D. The electrostatic repulsion

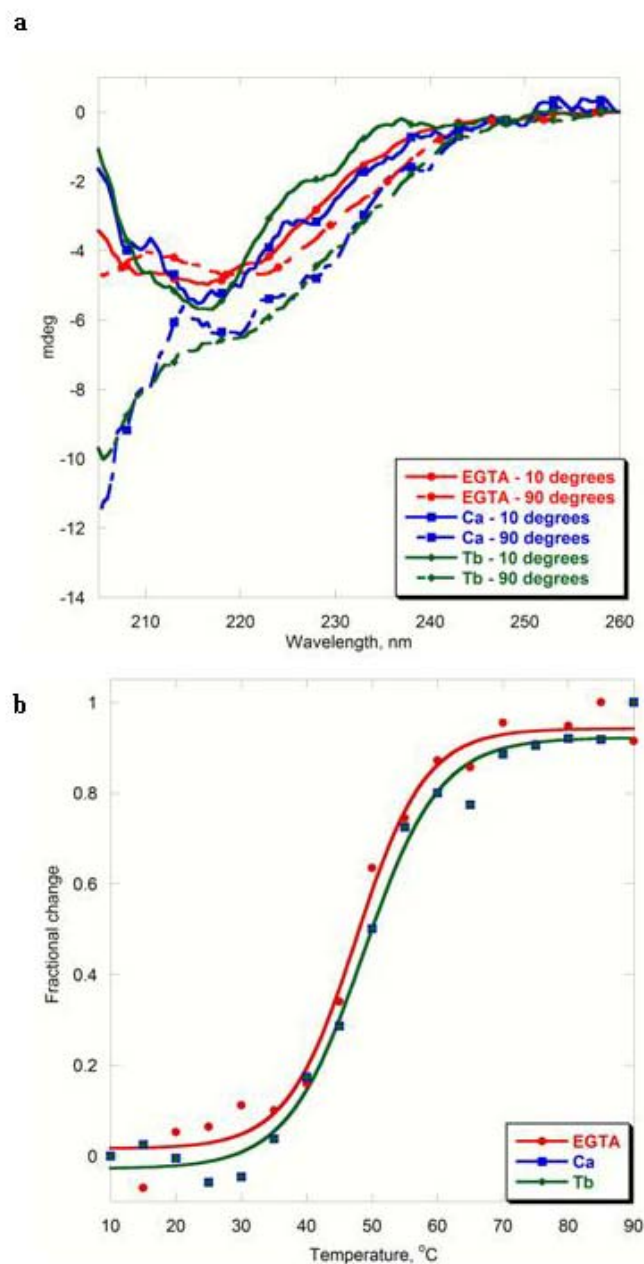


Figure 6.17 Thermal denaturation of 6D79.D94A. (a) Far UV CD spectra of 6D79.D94A (20 μ M) with 1 mM EGTA, 10 mM Ca^{2+} , and 50 μ M Tb^{3+} at 10 and 90 degrees in 10 mM Tris, 10 mM KCl pH 7.0. (b) Fractional change of the far UV signal at 235 nm as a function of temperature.

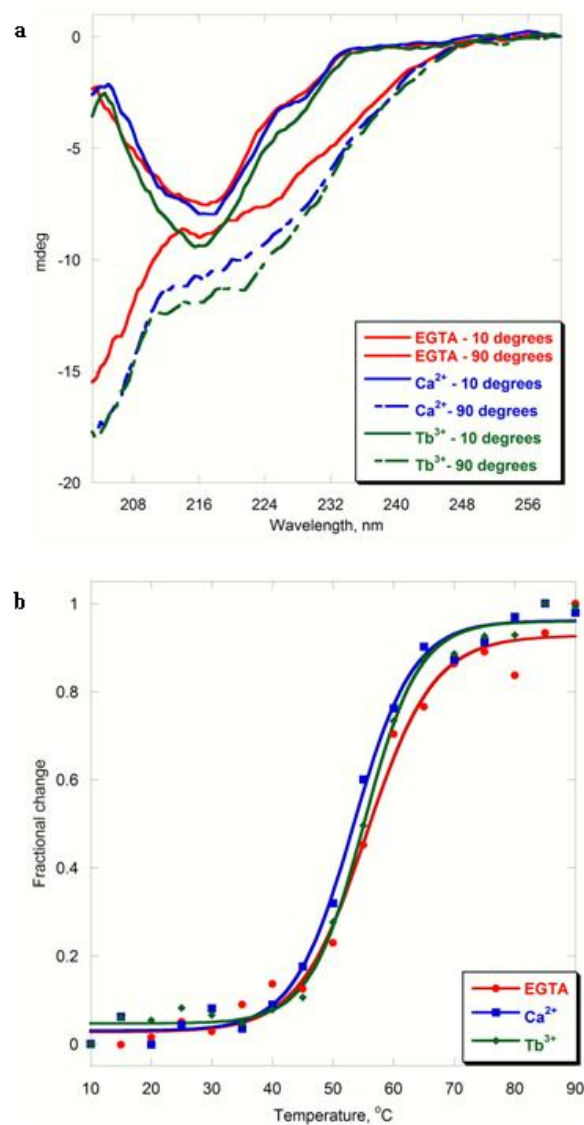


Figure 6.18. Thermal denaturation of 6D79.D94K. (a) Far UV CD spectra of 6D79.D94K (20 μ M) with 1 mM EGTA, 10 mM Ca^{2+} , and 50 μ M Tb^{3+} at 10 and 90 degrees in 10 mM Tris, 10 mM KCl pH 7.0. (b) Fractional change of the far UV signal at 235 nm as a function of temperature.

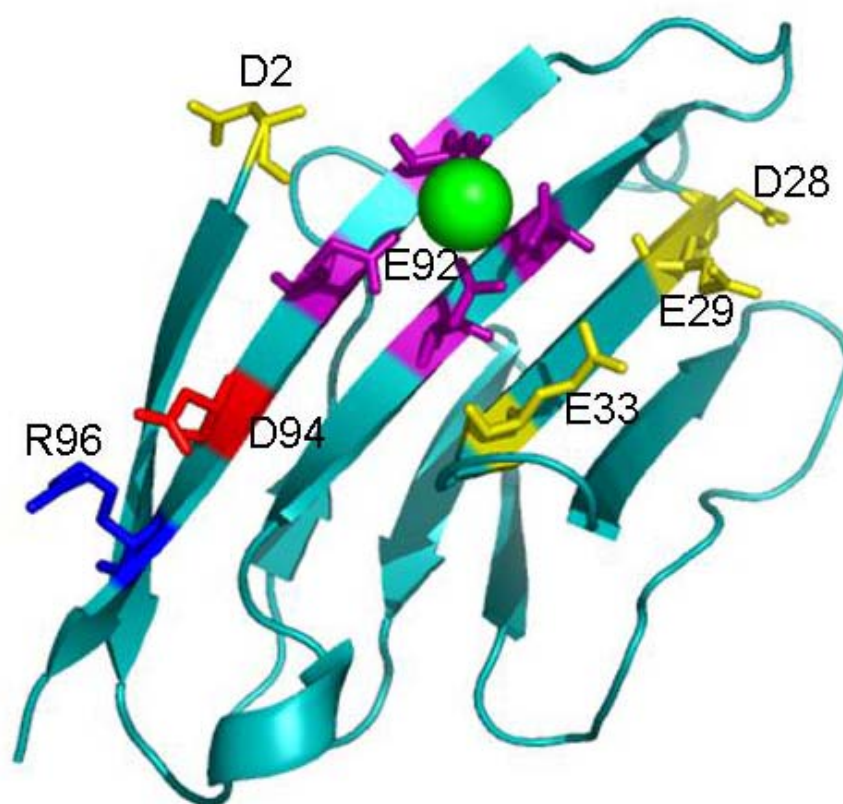


Figure 6.19. Formation of a salt bridge between R96 and E92. The mutation of D94 to Lys removes the salt bridge to R96 but forms one with E92 reducing the electrostatic repulsion of E92 with other negatively charged residues in the vicinity.

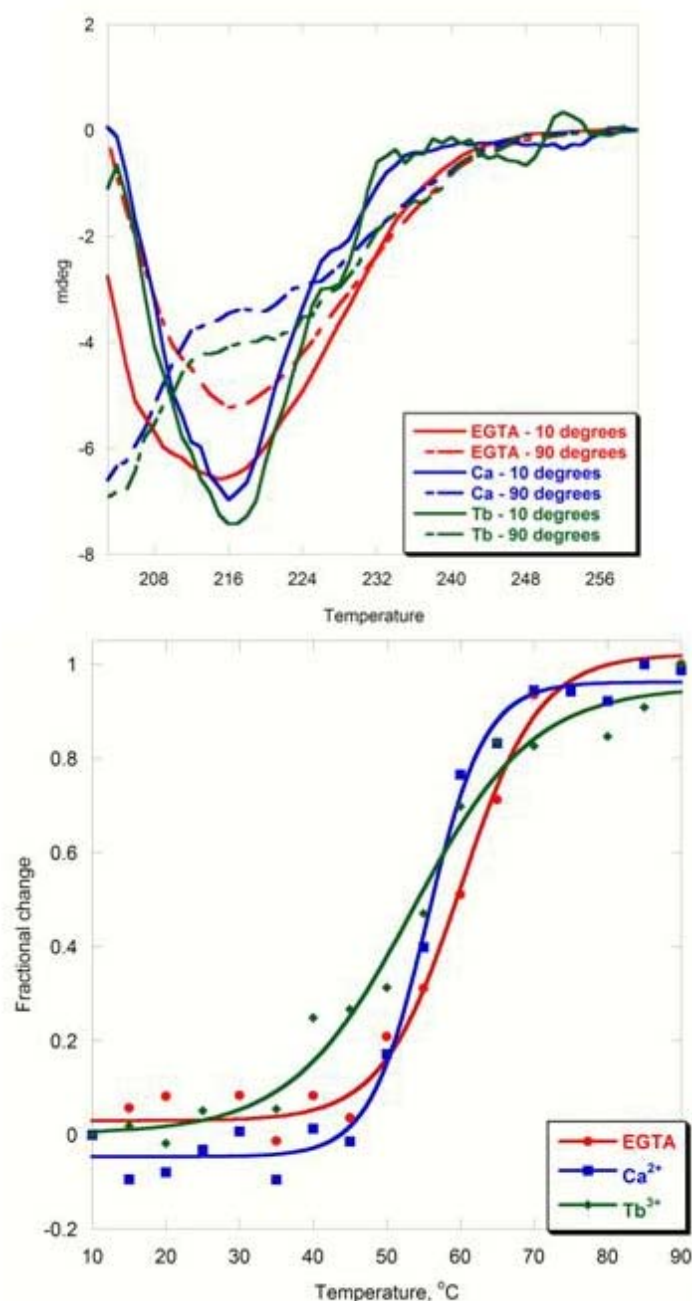


Figure 6.20. Thermal denaturation of 6D79.T75K. (a) Far UV CD spectra of 6D79.T75K (20 μ M) with 1 mM EGTA, 10 mM Ca^{2+} , and 50 μ M Tb^{3+} at 10 and 90 degrees in 10 mM Tris, 10 mM KCl pH 7.0. (b) Fractional change of the far UV signal at 235 nm as a function of temperature.

between D94 and D75 prevents the protein from properly folding.

6.6.3 Comparison of protein yields to stability

The yield of purified protein was compared to the protein stability to determine if there was a correlation between the two properties. Table 6.5 details the yield of purified protein obtained for each of the charge variants. For most of the proteins the purification yields were very similar. However, the purification yields of CD2.D94A were significantly lower than the other variants. In addition, this protein degraded rapidly at room temperature within a 36 hr period making it difficult to complete an entire NMR titration. The low purification yields of this protein as well as its rapid degradation may be linked to the low stability of this protein.

6.6.4 Comparison of calculated and experimental stabilities of the CD2.6D79 variants.

The calculated and experimental stability data were compared to one another to determine the validity of the free energy calculations. The experimental data shows that the 6D79.K91 variants are both more stable than CD2.6D79 in the apo form (Table 6.4). The free energy of these proteins corresponds with the same trend with both variants having a more negative free energy than that of CD2.6D79. Interestingly, the calculated data accurately predict the increased stability of 6D79.K91E over 6D79.K91I. The calculated data for the 6D79.D94 variants do not correlate with the experimental data. The thermal stability of 6D79.D94A is essentially, with the error bar, the same as CD2.6D79. However, the calculated data predict this protein would be less stable than CD2.6D79. The calculation also predicts 6D79.D94K to have a similar stability as

Protein	Purification yields (mg/L)
CD2.6D79	3
6D79.D94A	0.8
6D79.D94K	2.2
6D79.K91E	3.4
6D79.K91I	2.1
6D79.T75K	2.9

Table 6.5 Purification yields of charge distribution variants.

CD2.6D79 but the experimental data indicates the protein is much more stable. Additionally, the calculated data inaccurately predicts a decreased stability for 6D79.T75K whereas the experimental data indicates the protein has a large increased stability compared to CD2.6D79. This data suggests the limited use of the free energy calculations in predicting affinity.

6.6.5 Measurement of thermal stability in the metal bound form of the CD2.6D79 variants

The effect of metal binding on the protein stability was measured through observing the transition curve in the presence of 10 mM Ca^{2+} as well as the trivalent Ca^{2+} analogue Tb^{3+} (50 μM). The stability of CD2.6D79 is increased in the presence of metal due to the positive charge decreasing the charge repulsion of the clustered negative charges. For the 6D79.K91 variants a decrease in protein stability, 53 ± 1 and 46 ± 2 °C for K91E (Figure 6.15) and K91I (Figure 6.16), respectively, with the addition of Ca^{2+} is observed. Further, the stability decrease is more acute with the binding of the trivalent metal Tb^{3+} (43 ± 1 for K91E and 30 ± 1 for K91I). There are several possible explanations for the decrease of thermal stability for the metal loaded form of the K91 variants. It is possible that the increased stability in the apo form of the protein is due to the removal of the positive charge, which must be interacting and repulsing with other like charged residues, because the addition of positive charge through the binding of Ca^{2+} and the trivalent Tb^{3+} , which produces a more dramatic effect decreases the stability of the protein. The protein destabilization is more significant in the 6D79.K91I variant compared to 6D79.K91E. This is due to the presence of the negative charge in

6D79.K91E which is balancing the electrostatic interactions within the microenvironment around residue 91. The lack of the balance in the presence of the neutral residue Ile renders the protein more susceptible to destabilization through metal binding.

The 6D79.D94 variants do not display a significant change in protein stability when bound to Ca^{2+} or Tb^{3+} (Table 6.4, Figures 6.17-6.18). The limited effect of metal on protein stability can be attributed to the weak binding affinity of the 6D79.D94 variants (chapter 7). The 6D79.T75K variant also displays a decrease in stability upon binding metal (55 ± 1 for Ca^{2+} and 53 ± 1 °C for Tb^{3+} , Figure 6.20) indicating the positively charge Lys which is 7 Å from the metal binding site is interacting with and repulsing the bound metal ion.

6.7 Conclusions

Both short- and long-range electrostatic interactions have been shown to play a role in protein stability. The examination of the thermal stability of several charge distribution variants of CD2.6D79 elucidates the role of these interactions. In the apo form of the protein, the removal of an Asp residue at position 94 does not alter the thermal stability of the protein. Furthermore, a charge reversal at this same position (D94K) increases the thermal stability by 9 °C. The increased stability of the 6D79.D94K variant can be attributed to its long range interactions on the Ca^{2+} binding surface where 70% of the residues are charged. Statistical analysis has shown that when multiple charged residues are present, more complex salt bridges are often formed (Musafia et al., 1995). Further, charged residues have a tendency to form cooperative networks (Kumar and Nussinov, 2002a). The large number of charged residues on the

Ca^{2+} binding surface of CD2.6D79 may form a cooperative network of electrostatic interactions. The introduction of two mutations, T79D and A92E, to form the Ca^{2+} binding site of CD2.6D79 has reduced the stability of this protein compared to CD2 by greater than 10 °C (Jones, 2006). This decreased stability is attributed to the electrostatic repulsion between these introduced charges and the negative charges already present on the surface of the protein. However, the introduction of Lys at this position appears to balance the charges in this network sufficiently enough to increase the thermal stability. Nevertheless, the thermal stability of this variant is still lower than CD2 indicating the charge repulsion is not completely neutralized. Moreover, the 6D79.D94A variant does not alter the thermal stability indicating that simply removing the negative charge at position 94 is not adequate to decrease the electrostatic repulsion. The addition of metal in both 6D79.D94A and 6D79.D94K does not have a significant effect on the thermal stability even though charge interactions have been altered at position 94. This indicates the distance between the Ca^{2+} binding site and residue 94, 14.1 Å, is too substantial for charge-charge interactions to occur.

The increased stability of the 6D79.K91 variants can also be attributed to a complex network of charge interactions. However, this network may be different than the one that residue 94 is a part of. According to the crystal structure of CD2, residue 91 is pointed towards the opposite direction of the metal binding surface of CD2.6D79. The ligands that form the Ca^{2+} binding site are located on two adjacent β -strands. In β -sheet secondary structure, side-chains project alternately above and below the plane of the sheet. Therefore, even in the absence of a structure of CD2.6D79 and the 6D79.K91

variants, one can assume the direction of the residue K91 is not altered. Residue K91 is likely interacting with other like-charge residues on the opposite surface of the protein. The removal of the positive charge at position 91 has reduced charged repulsion in the environment around this residue thereby increasing the protein stability of the two variants compared to CD2.6D79. Ca^{2+} binding decreases the T_m of both of these variants indicating the introduction of the +2 charge re-introduces the charge repulsion. This is also indicated by the fact that the binding of the trivalent Ca^{2+} analogue Tb^{3+} decreases the thermal stability even further. The protein destabilization is more significant in the 6D79.K91I variant compared to 6D79.K91E. This is due to the presence of the negative charge in 6D79.K91E which is balancing the electrostatic interactions around residue 91. The lack of the balance in the presence of the neutral residue Ile renders the protein more susceptible to destabilization through metal binding.

The largest increase in stability was observed for the 6D79.T75K variants. The T_m has increased 11 °C over CD2.6D79. This increase can be ascribed to the formation of a salt bridge between K75 and D94 and is an example stabilizing short-range electrostatic interactions. Residues 75 and 94 are located on β -strands that are adjacent to one another and are cross-pairs of one another. In both parallel and antiparallel β -strands the distance between the side-chains of cross-strand residues are close whereas the side-chains of same-strand residues have a greater distance thereby reducing the potential of side-chain interaction within a strand (Cootes et al., 1998). Cross-strand side-chain interactions modulate the stability of β -sheets with the greatest stabilizations effects observed for charge-charge interactions (Merkel et al., 1999). The distance of T75 to

D94 in CD2 is 5.4 Å, which is larger than the requirement for salt bridge formation. However, Lys has a longer side-chain than Thr which may reduce the distance between the two residues allowing for salt bridge formation. The proximity of D94 also accounts for the altered conformation of 6D79.T75D. The electrostatic repulsion between D94 and D75 prevents the protein from properly folding. The binding of Ca^{2+} and Tb^{3+} reduces the thermal stability of the protein demonstrating the destabilizing effects of long range electrostatic interactions.

Although the stabilizing properties of Ca^{2+} are generally attributed to rigidifying the flexible regions of protein we have shown that the long range electrostatic properties of Ca^{2+} can be a stabilizing or a destabilizing force. In CD2.6D79 Ca^{2+} has a stabilizing effect through the reduction of charge repulsion between negative charges at the binding site. On the other hand, for several of the CD2.6D79 variants, Ca^{2+} is a destabilizing force due to the increased positive charge around the metal binding site. However, there is a limit to the range of these interactions. The binding of metal 14.1 Å from a charge alteration at position 94 does not cause a significant change in the stability. The inability of metal binding to have an effect on the stability indicates the distance between residue 94 and the metal binding site is beyond the range of the electrostatic interactions. The limit of the range of the electrostatic properties of Ca^{2+} is also evident with the 6D79.T75D variant. The apo form of the protein has an altered conformation due to electrostatic repulsion with D94. The binding of Ca^{2+} to this variant does not cause the protein to undergo a conformational change indicating the range of the electrostatic properties of Ca^{2+} is not sufficient to reduce the repulsion of these residues. This is

contrary to what we have observed in another designed calcium binding protein of CD2, CD2.6D31 where electrostatic repulsion between the Ca^{2+} binding ligands induces an alternate fold of the protein. However, upon metal binding, the protein undergoes a conformational change to β -sheet architecture similar to CD2. In summary, we have been successful in using protein design as a tool to investigate the role of short- and long-range electrostatic interactions on protein stability. We have shown the stabilizing and destabilizing effects of electrostatic interactions using our model study. This study will provide insight into the design of thermostable proteins which will be invaluable in the development of new biomaterials, sensors, and pharmaceutical drugs.

Chapter 7. Metal binding properties of the 6D79 charge distribution variants

7.1 Rationale for studying the metal binding affinity of the charge distribution variants

As discussed in **section 1.5** electrostatic interactions play an important role in calcium binding. The electrostatic distribution outside of the coordination shell of the metal binding site is considered to be a key determinant that governs high affinity Ca^{2+} binding (Linse and Forsen, 1995). The contribution of individual charges not directly involved in Ca^{2+} binding to the free energy of calcium binding to calbindin $\text{D}_{9\text{K}}$ was investigated by Linse and coworkers. They have shown that removal of three negative surface charges Glu, Asp, and Glu at positions 17, 19, and 26 (Figure 1.16) in the vicinity of the calcium binding sites of this protein leads to a 45 fold decrease in average affinity per site at low ionic strength and a five-fold reduction at 0.13 M KCl indicating the importance of electrostatic distribution (Linse et al., 1988; Linse et al., 1991). These studies were based on macroscopic binding affinities due to the fact that calbindin $\text{D}_{9\text{K}}$ bind two Ca^{2+} ions cooperatively. We wanted to study the role of long range electrostatic interactions on site specific Ca^{2+} binding affinity using various spectroscopic methods.

The study of the first generation proteins attempted to investigate this role through the analysis of three different proteins in varying electrostatic environments. The approach discussed in this chapter involves altering the electrostatic potential of a single Ca^{2+} binding site through mutation of charged residues around the Ca^{2+} binding site. Since the designed protein CD2.6D79 exhibits strong metal binding affinities and metal selectivity while retaining a CD2-like conformation it can be used as a model system for

investigating the role of electrostatic interactions on Ca^{2+} binding affinity. As discussed in **sections 3.5** and **6.3**, the selection of the residues to be mutated was based on several criteria including distance to the metal binding site and solvent accessibility. The effect of mutations on the conformation and thermal stability were discussed in chapter 6. In this chapter the Tb^{3+} and Ca^{2+} binding affinities of each of the charge distribution variants will be discussed.

7.2. The 6D79.K91 variants

Based on the calculation of the electrostatic potential, we hypothesize that removing the positive charge at this position would increase the Ca^{2+} binding affinity of 6D79. Further, if a negative charge was added at this position a more dramatic increase in the binding affinity would be observed. As shown in Figure 6.3 and Table 6.1 residue K91 is 11 Å from the Ca^{2+} binding site. This residue was mutated to the negatively charged residue Glu (K91E) and to the neutral residue Ile (K91I) to test our hypothesis. This set of mutants will provide insight into how the removal of a positive charge around the Ca^{2+} coordination shell impacts Ca^{2+} binding. K91E and K91I both expressed well in LB media and purified with ease yielding enough protein for conformational analyses and metal binding studies (Table 6.5).

7.2.1 Tb³⁺ binding affinity of the K91 variants

Fluorescent resonance energy transfer was used to measure the Tb³⁺ binding affinity of the K91 variants. The metal binding site is in suitable distance to three aromatic residues, W32, Y76, and Y81 (11.8, 11.6, and, 8.3 Å respectively), which will allow for energy transfer to occur. The protein concentration was kept constant (2 µM) while the Tb³⁺ concentration was increased. The buffer used, 20 mM PIPES, 10 mM KCl pH 6.8, helps minimize the precipitation of the protein as Tb³⁺ is added. Each experiment was performed in triplicate. As shown in Figure 7.1, the addition of Tb³⁺ increases the fluorescence emission at 545 and 585 nm when excited at 283 nm. For variant K91E the replacement of the positive charge with the negative charge has increased the binding affinity by several-fold, 6 ± 4 and 74 ± 9 µM for K91E and 6D79, respectively (Figure 7.1), indicating that the introduction of negatively charged residue within 8 Å of the Ca²⁺ coordination shell has a significant effect on Ca²⁺ binding affinity. The titration with K91I gives a K_d of 41 ± 7 µM (Figure 7.2). The removal of the positive residue at position 91 does increase the binding affinity for Tb³⁺ under the same conditions, but the increase is not as significant as replacing the residue with a negatively charged one.

7.2.2 Ca²⁺ binding affinity of the K91 variants

Ca²⁺ titrations were performed on K91E using 2D ¹⁵N HSQC and 1D ¹H NMR to determine their respective binding affinities. For the HSQC titration the protein concentration was 80 µM in 20 mM PIPES, 10 mM KCl pH 6.8. The initial titration point contained 25 µM EGTA to ensure a Ca²⁺ free sample. Upon the addition of Ca²⁺ several chemical shift changes were observed. The most significant peak shifting was

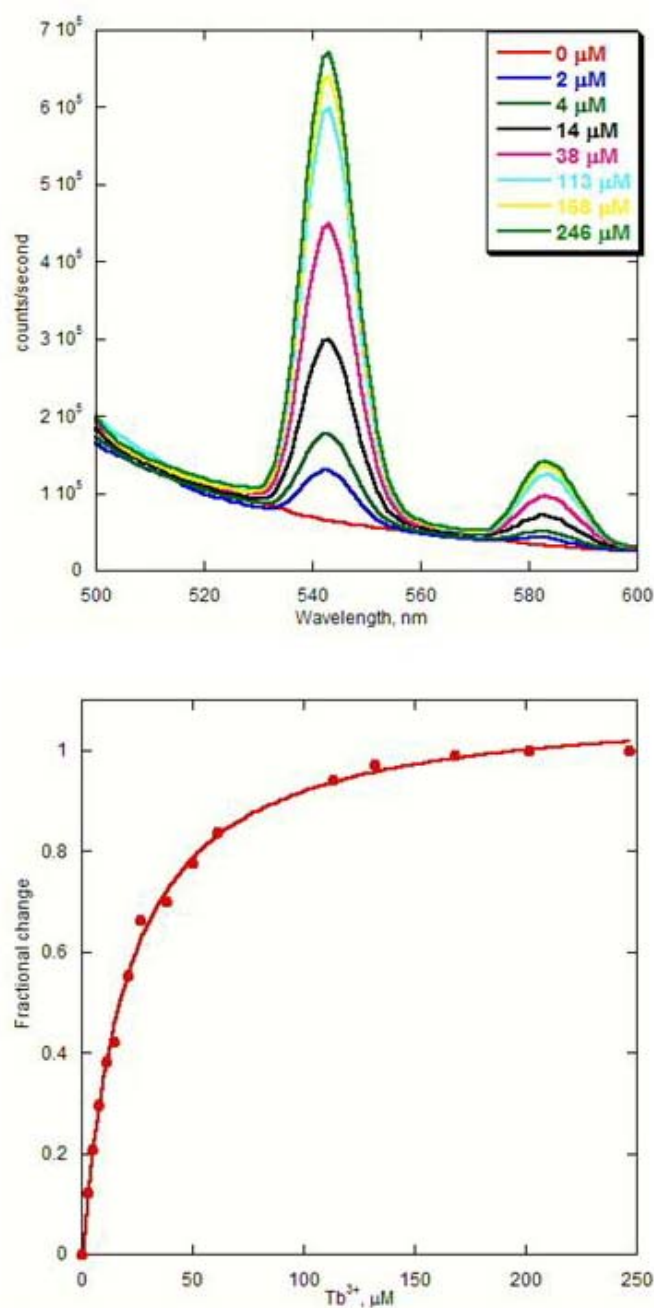


Figure 7.1. Tb^{3+} binding of 6D79.K91E. (a) Fluorescence emission increase of (3 μM) 6D79.K91E in 20 mM PIPES, 10 mM KCl pH 6.8 at 545 nm when excited at 283 nm upon Tb^{3+} binding (b) Fit of the fractional change of the fluorescence intensity yields a K_d of 6 ± 4 μM using equation 2.2.

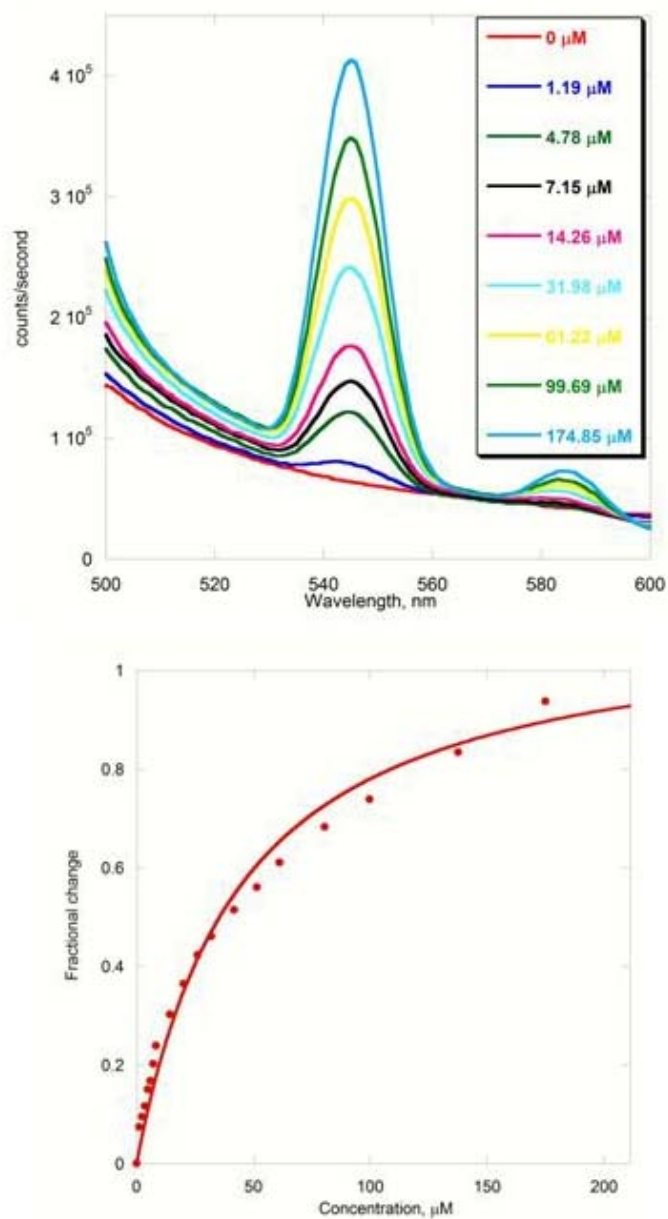


Figure 7.2. Tb³⁺ binding of 6D79.K91I. (a) Fluorescence emission increase of 6D79.K91I (3 μM) in 20 mM PIPES, 10 mM KCl pH 6.8 at 545 nm when excited at 283 nm upon Tb³⁺ binding (b) Fit of the fractional change of the fluorescence intensity yields a K_d of 41 ± 7 μM using equation 2.2.

seen for the side chain of N77 (Figure 7.3a). Small changes were also observed for E91 and the backbone of N77. The Ca^{2+} binding affinity determined from the fractional change of the chemical shift was 25 μM (Figure 7.3b).

One dimensional ^1H NMR Ca^{2+} titrations were also employed due to limited amounts of the hetero-labeled protein. The protein concentration for the 1D titration was 150 μM in 20 mM PIPES, 10 mM KCl pH 6.8. EGTA (25 μM) was added to the sample to ensure it was Ca^{2+} free. Then several additions of Ca^{2+} were added to a final concentration of 3.2 mM. K91E had chemical shift changes for three residues in the amide region at 11.5, 11.3, and 11.2 ppm's (Figure 7.4a). These chemical shift changes were small (Figure 7.5) compared to other Ca^{2+} binding proteins (Linse and Chazin, 1995). These changes were fit to a binding curve which yielded a K_d of $39 \pm 7 \mu\text{M}$ (Figure 7.4b).

For K91I, the protein concentration was 94 μM . Upon addition of Ca^{2+} chemical shift changes were observed for 2 peaks in the amide region of the protein at 11.3 and 11.2 ppm's (Figure 7.6a and 7.7). The fit of these chemical shift changes yields a K_d of $85 \pm 13 \mu\text{M}$ (Figure 7.6b). Similar to the Tb^{3+} data both variants have stronger binding affinities than 6D79 ($197 \pm 30 \mu\text{M}$), with the K91E mutant having stronger binding affinity. The data is consistent with our hypothesis that the removal of the positive charge around the Ca^{2+} binding site will increase the Ca^{2+} binding affinity and the mutation of the residue to a negatively charged residue will have a more significant effect.

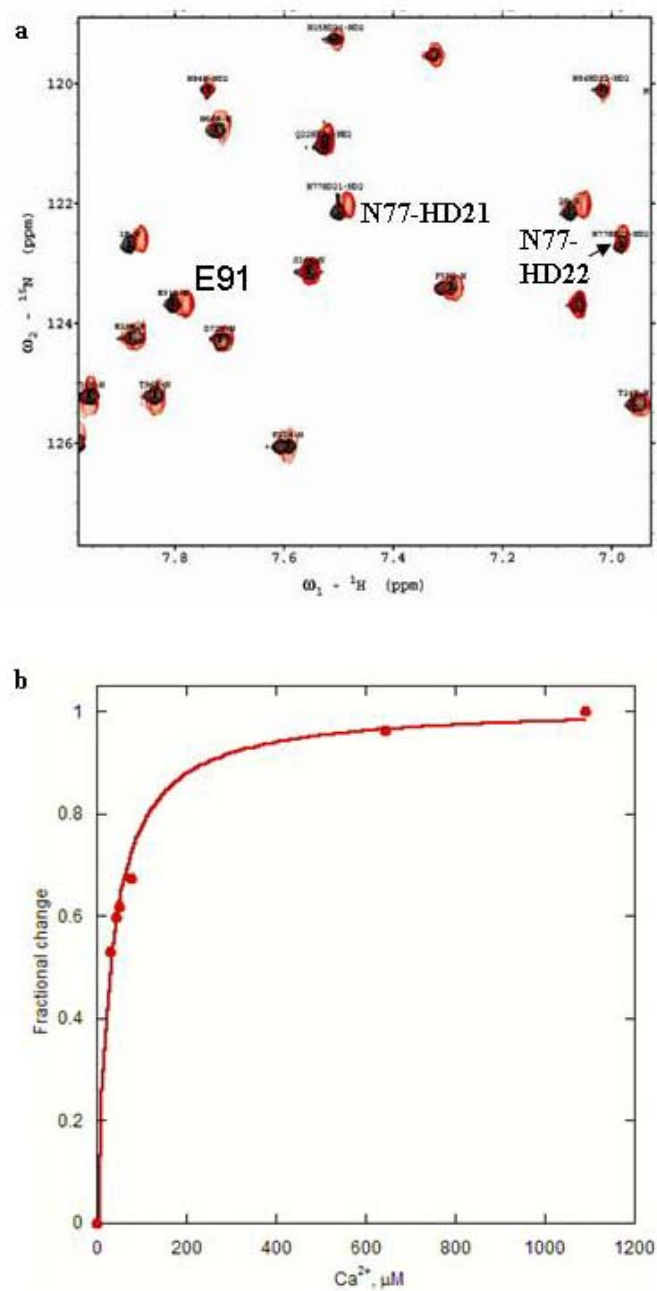


Figure 7.3. HSQC titration of 6D79.K91E (a) The chemical shift changes of 6D79.K91E (80 μM) in 20 mM PIPES, 10 mM KCl pH 6.8 in the Ca^{2+} free form (black) and with 3 mM Ca^{2+} (red). (b) The fit of the chemical shift changes of 6D79.K91E using equation 2.2.

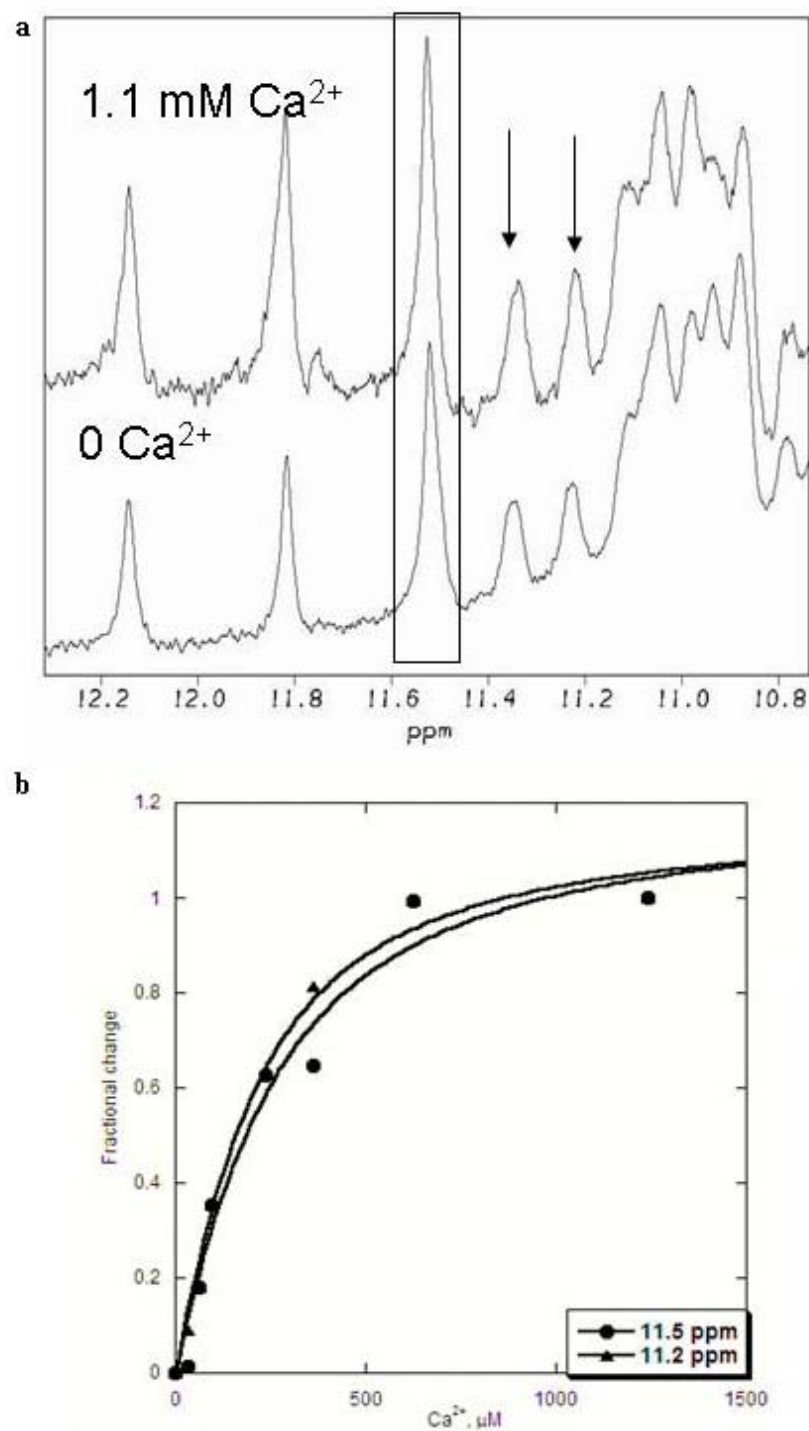


Figure 7.4. 1D ^1H NMR titration of 6D79.K91E. (a) 1D ^1H NMR Ca^{2+} titration of 6D79.K91E. Chemical shift changes are observed at 11.5 and 11.2 ppm, corresponding to Y76, N77, and E41, respectively. (b) The fit of the fractional change of the chemical shift yields a K_d of $39 \pm 7 \mu\text{M}$.

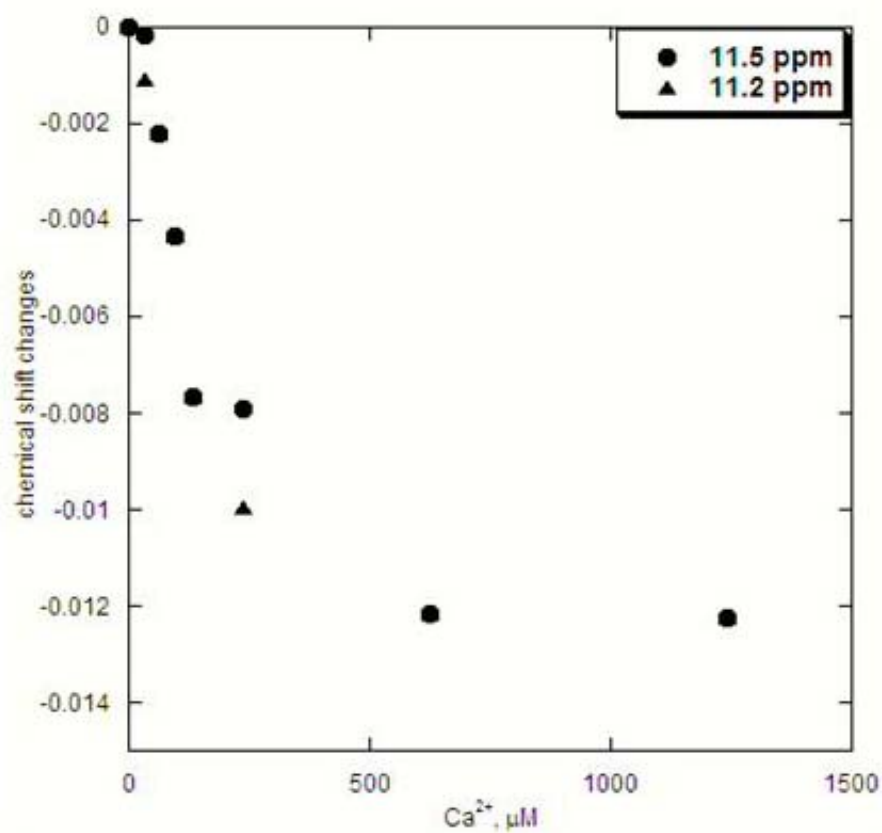


Figure 7.5. Chemical shift changes of 6D79.K91E. Chemical shift changes of peaks at 11.5 and 11.2 ppm of 6D79.K91E upon the addition of Ca^{2+} .

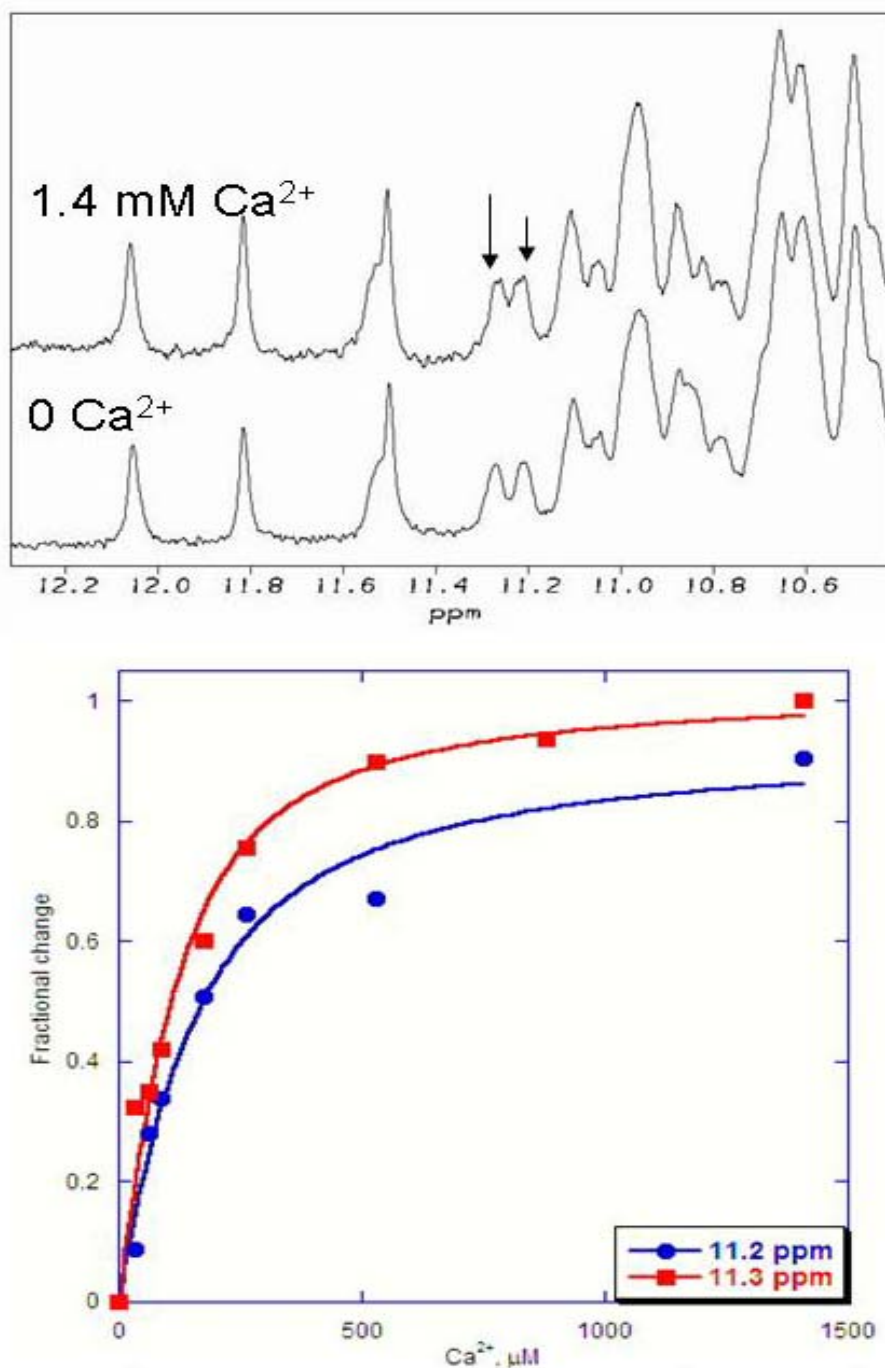


Figure 7.6. 1D ^1H NMR titration of 6D79.K91I. 1D ^1H NMR Ca^{2+} titration of 6D79.K91I. Chemical shift changes are observed at 11.3 and 11.2 ppm's, which correspond to N77 and E41, respectively.

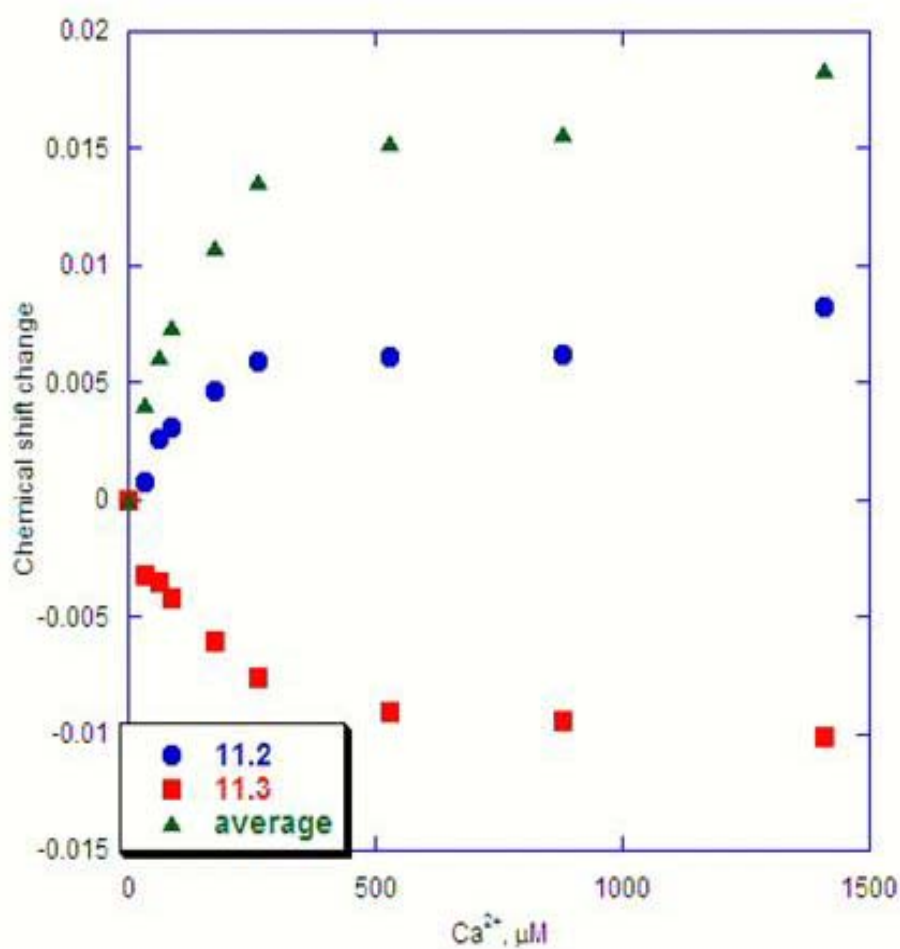


Figure 7.7. Chemical shift changes of 6D79.K91I. Chemical shift changes of peak N77 (11.3 ppm) and E41 (11.2 ppm) of 6D79.K91I upon the addition of Ca^{2+} .

7.3 The 6D79.D94 variants

This set of mutants will provide insight into how the removal of a negative charge around the Ca^{2+} coordination shell impacts Ca^{2+} binding. We hypothesize that removing the negative charge at this position would decrease the Ca^{2+} binding affinity of 6D79. Further, if a positive charge was added at this position a more dramatic decrease in the binding affinity would be observed.

7.3.1 Tb^{3+} binding affinity of the 6D79.D94 variants

FRET was employed to determine the Tb^{3+} binding affinities of the 6D79.D94 variants. Expecting the binding affinity to be rather weak, the titration was carried out with very small increments of Tb^{3+} being added (Figure 7.8). These minor additions reduce the amount of precipitation at the high metal concentrations required for the titration to reach saturation. The Tb^{3+} binding affinities of D94K and D94A were 139 ± 5 and $102 \pm 11 \mu\text{M}$, respectively (Figures 7.8-7.9). The experiments were performed in triplicate and the error is based on the differences in these three experiments. These affinities are slightly weaker than 6D79 ($74 \pm 9 \mu\text{M}$).

7.3.2 Ca^{2+} binding affinity of the 6D79.D94 variants

One-dimensional ^1H NMR was carried out to determine the Ca^{2+} binding affinities of the D94 variants. For both proteins chemical shift changes occurred at 11.3 ppm (Figure 7.10a). These two peaks correspond to the Ca^{2+} binding ligand N77 (11.3 ppm) (Figure 7.11). The Ca^{2+} binding affinity of D94K was determined to be $303 \pm 60 \mu\text{M}$ (Figure 7.10b). The Ca^{2+} binding affinity of D94A was not determined due to the instability of the protein. The protein degraded before the titration could be completed.

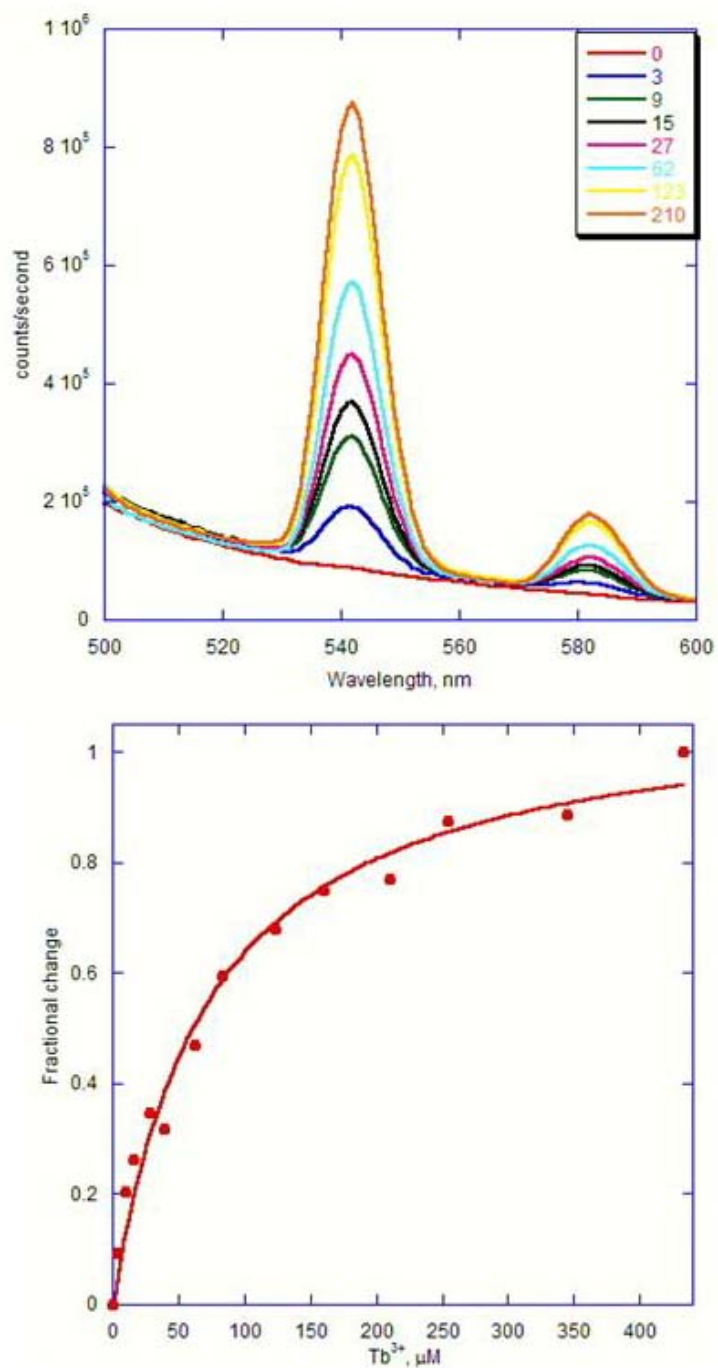


Figure 7.8. Tb^{3+} binding of 6D79.D94K. (a) Fluorescence intensity increase of 6D79.D94K at 545 nm upon Tb^{3+} binding (b) Fit of the fractional change of the fluorescence intensity yields a K_d of $139 \pm 5 \mu M$

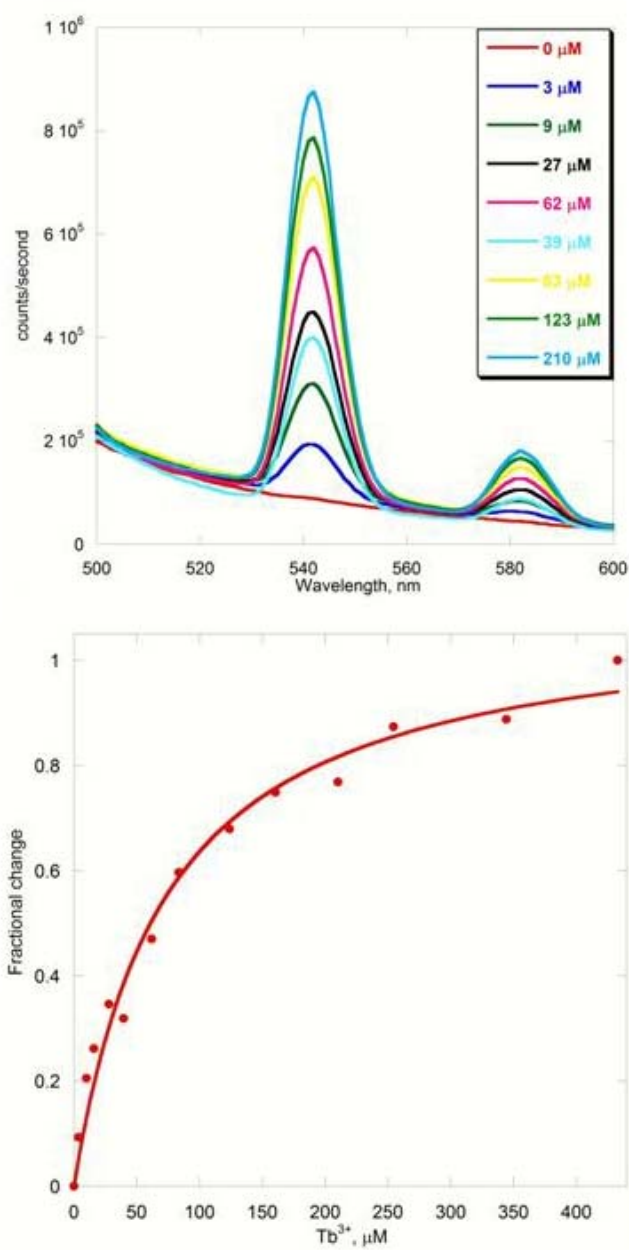


Figure 7.9. Tb^{3+} binding of 6D79.D94A. (a) Fluorescence intensity increase of 6D79.D94A at 545 nm upon Tb^{3+} binding (b) Fit of the fractional change of the fluorescence intensity yields a K_d of $102 \pm 11 \mu M$

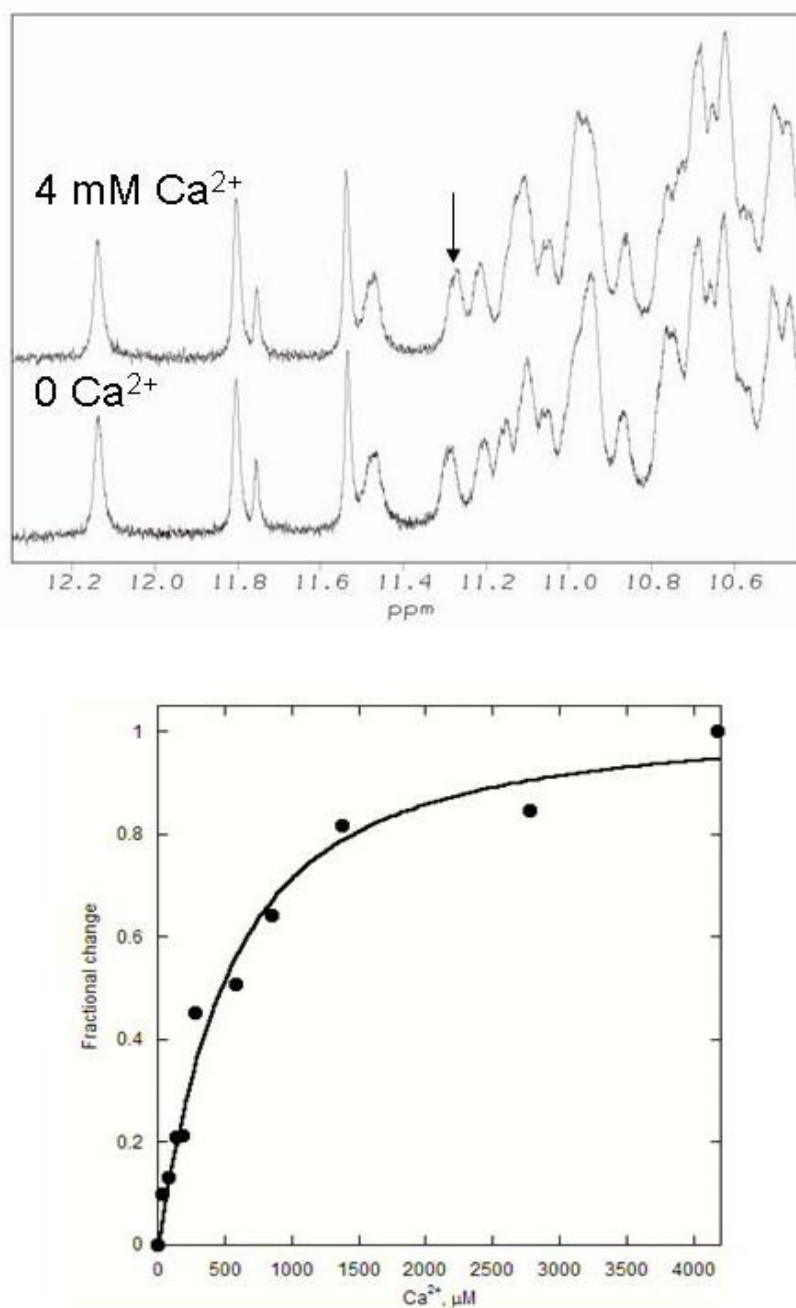


Figure 7.10. 1D ^1H NMR titration of 6D79.D94K. 1D ^1H NMR Ca^{2+} titration of 6D79.D94K in 20 mM PIPES, 10 mM KCl pH 6.8. Chemical shift changes are observed at 11.3 ppm which correspond to N77.

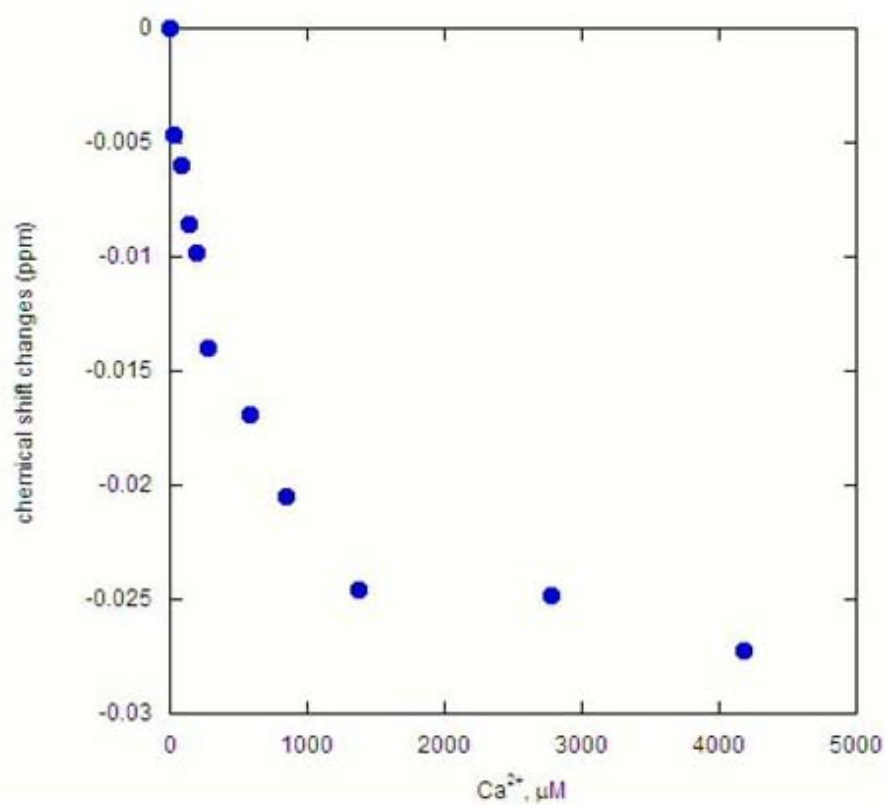


Figure 7.11. Chemical shift changes of D94K. Chemical shift changes of 3 peaks of D94K upon the addition of Ca^{2+} .

7.4 The 6D79. T75 variants

Since the 6D79.T75D variant has an altered conformation compared to CD2.6D79 the metal binding properties of this variant were not examined. For the 6D79.T75K variant we hypothesize that the addition of the Lys residue will decrease the metal binding affinity of the protein while the introduction of the Asp residue will increase the metal binding affinity.

7.4.1 Tb³⁺ binding affinity of 6D79.T75K

Once again FRET was employed to determine the Tb³⁺ binding affinity of T75K. Preliminary titration experiments indicated the Tb³⁺ binding was rather weak and the titration had difficulty reaching saturation. Therefore, to reduce precipitation, which occurs at high concentrations of Tb³⁺, small increments of Tb³⁺ was added to the protein. A Tb³⁺ binding affinity of $278 \pm 16 \mu\text{M}$ was determined (Figure 7.12). This K_d is almost 4-fold weaker than that of 6D79 indicating the presence of the positive charge is playing a significant role in the binding affinity.

7.4.2 Ca²⁺ binding affinity of T75K

The Ca²⁺ binding affinity of T75K was determined using 1D ¹H NMR. Similar to 6D79 and other charge distribution variants chemical shift changes were seen for the N77 residue (Figure 7.13-7.14). The Ca²⁺ binding affinity was revealed to be $444 \pm 60 \mu\text{M}$. This is almost 3-fold weaker than the Ca²⁺ binding affinity of 6D79.

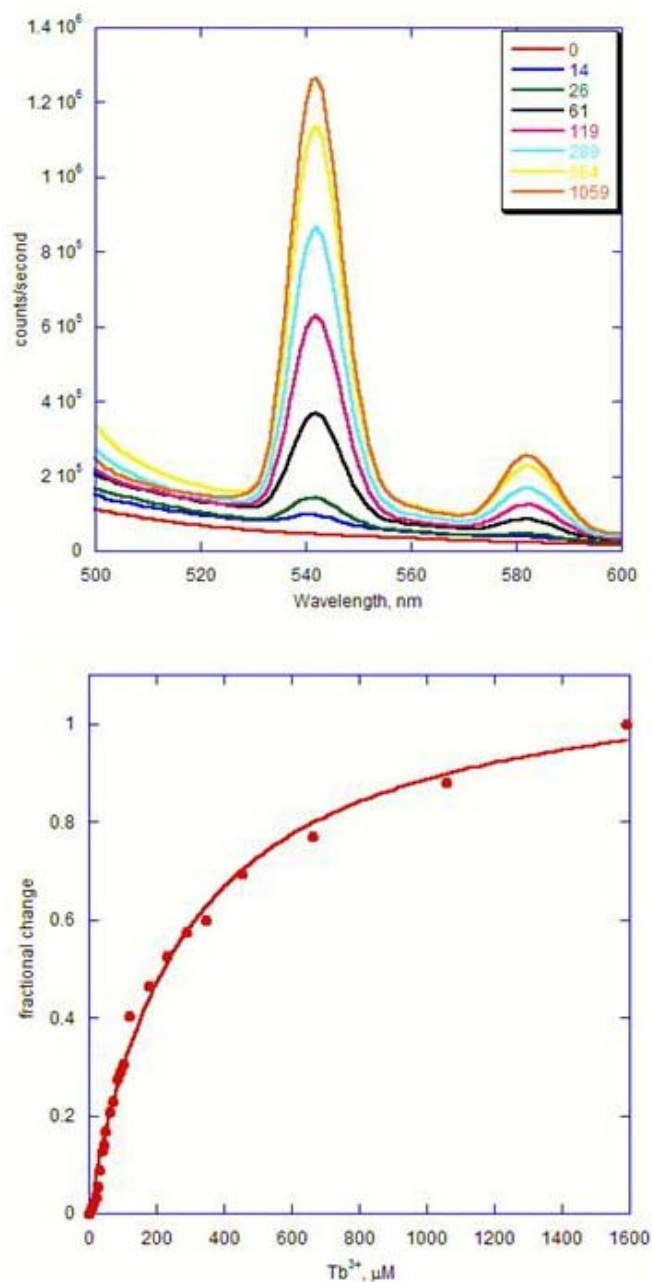


Figure 7.12. Tb³⁺ binding of 6D79.T75K. (a) Fluorescence emission increase of 6D79.T75K (3 μM) in 20 mM PIPES, 10 mM KCl pH 6.8 at 545 nm when excited at 283 nm upon Tb³⁺ binding (b) Fit of the fractional change of the fluorescence emission using equation 2.2 yields a K_d of 278 ± 16 μM

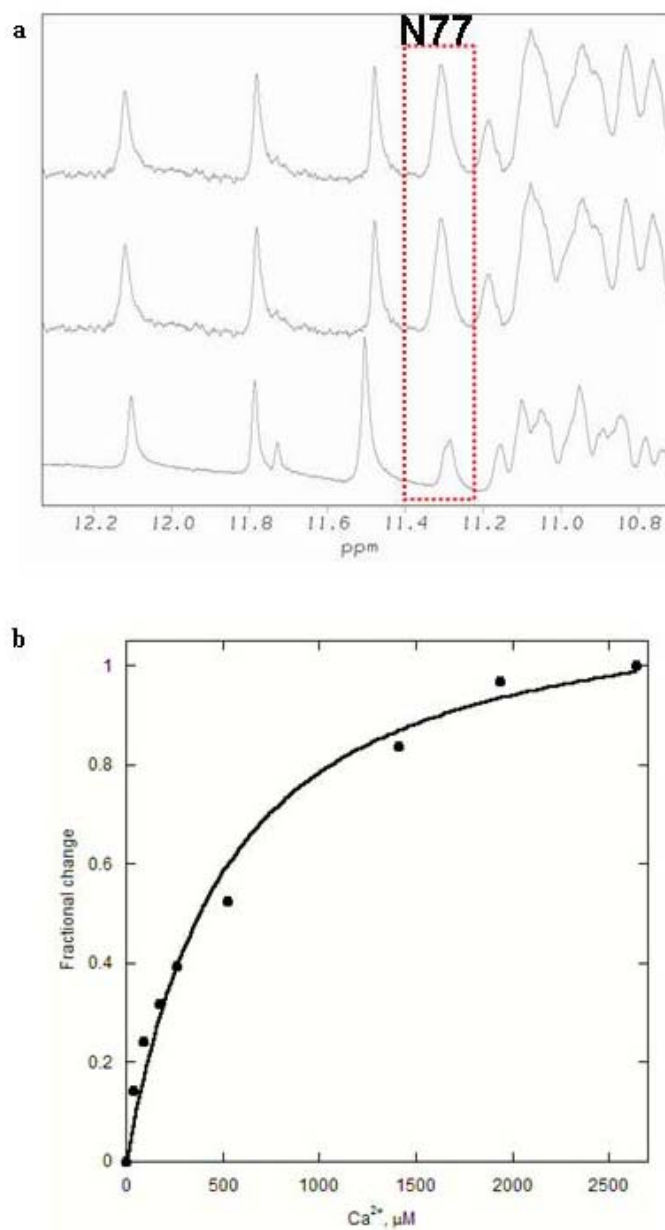


Figure 7.13. 1D ^1H NMR Ca^{2+} titration of 6D79.T75K. (a) Chemical shift changes are observed at 11.3 ppm, which correspond to N77. (b) The fit of the fractional change of the chemical shift yields a K_d of $444 \pm 60 \mu\text{M}$ in 20 mM PIPES, 10 mM KCl pH 6.8.

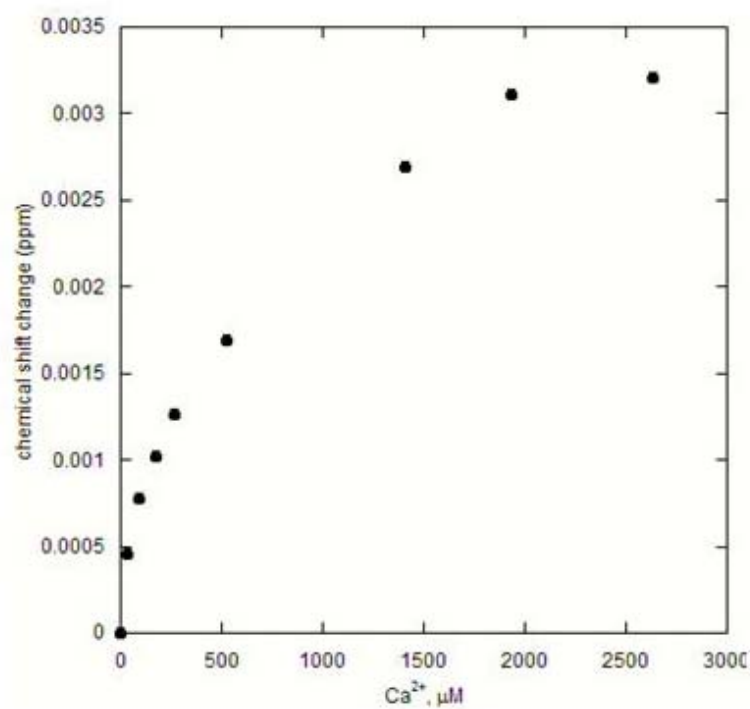


Figure 7.14. Chemical shift changes of 6D79.T75K. Chemical shift changes of peaks N77 of 6D79.T75K upon the addition of Ca^{2+} .

7.5 The role of electrostatic interactions around the coordination shell on Ca^{2+} binding affinity

Table 7.1 shows all of the Tb^{3+} and Ca^{2+} binding affinities determined for 6D79 and its charge distribution variants. The data reveals the removal of a positive charge around the Ca^{2+} binding site (K91I) increases the binding affinity (2-fold increase) while introducing a negative charge at that position (K91E) has a more dramatic effect (5-fold increase) on the binding affinity. A metal binding affinity decrease is observed when a negatively charged residue (D94) is removed from around the Ca^{2+} binding site similar to the trend observed with calbindin $\text{D}_{9\text{K}}$ (Linse et al., 1991). However, the introduction of the positive Lys at position D94 does not have as dramatic effect on the Ca^{2+} binding affinity as the introduction of Lys at position T75, 1.5-fold and 2.2 fold, respectively. This is due to the distance dependence of electrostatic interactions. D94 is 14 Å from the metal binding site while T75 is only 8.5 Å away. From this study we can conclude that Ca^{2+} binding affinity can be modulated by the surrounding electrostatic environment. This ability to modulate Ca^{2+} binding affinity of proteins will be valuable in developing Ca^{2+} sensors.

7.6 The comparison of binding affinities to protein stability

In comparison with the protein stability data, the protein with the highest stability is 6D79.T75K (Table 6.4). This variant has the weakest metal binding affinity due to the presence of the positive charge at position 75. Even though the metal binding is weak it still has an effect on the protein stability. Variant 6D79.T75K has decreased protein stability when it binds metal. The K91 variants have the strongest metal binding

Protein	Distance to Ca ²⁺ (Å)	Tb ³⁺ μM ^a	Ca ²⁺ μM ^b
6D79	---	74 ± 9	197 ± 30
K91E	11.4	6 ± 4	39 ± 7
K91I	11.4	41 ± 7	85 ± 13
D94K	14.1	139 ± 5	303 ± 60
D94A	14.1	102 ± 11	---
T75K	8.5	278 ± 16	444 ± 60

^aFRET-20 mM PIPES, 10 mM KCl pH 6.8

^bNMR-20 mM PIPES, 10 mM KCl pH 6.8

Table 7.1 Binding affinities of the charge distribution variants.

affinities even though metal binding decreases the stability of these proteins. This indicates electrostatic interactions play a role in both Ca^{2+} binding affinity and protein stability but these interactions can be separate from one another. The role that electrostatic interactions play in metal binding is independent of their role in protein stability.

Chapter 8. Conformational Properties of the 6D79.R34 Variants

8.1 Introduction

In addition to the roles in protein stability (chapter 6) and metal binding (chapter 7), electrostatic interactions also play a major role in protein folding. We hypothesize that when the destabilization of the folded state by the addition of charged residues results in unfolding of the native state the balance of charged repulsion by removing the charged residue will refold the protein. R34 is 9.4 Å from the Ca^{2+} binding site of 6D79 (Figure 8.1). The side chain of the residue faces towards the interior of the protein away from the binding site. Nonetheless, electrostatic calculations established the removal of the positive charge has a significant effect on the potential at the Ca^{2+} binding site ($\Delta\epsilon_p = -39.7$ kcal/mol). Therefore, R34 was chosen to be mutated as one of the charge distribution variants. Arg 34 was mutated to the negatively charged Glu and to the neutral Ile to determine their effects on Ca^{2+} binding affinity. As discussed in chapter 3 and 6, residue R34 was selected as a charge distribution mutant of CD2.6D79. However, the conformation of this protein is altered compared to CD2 and CD2.6D79, therefore, the metal binding properties of this protein was not analyzed. In this chapter, first, the conformational properties of these variants will be discussed. Second, the effect of pH on the conformation of 6D79.R34E will be discussed. Third, the attempt to recover the native conformation of the protein will be discussed. Fourth, the stability of the 6D79.R34I variant will be discussed. These studies of the conformational properties of these variants will be useful in elucidating the folding properties of CD2 and other

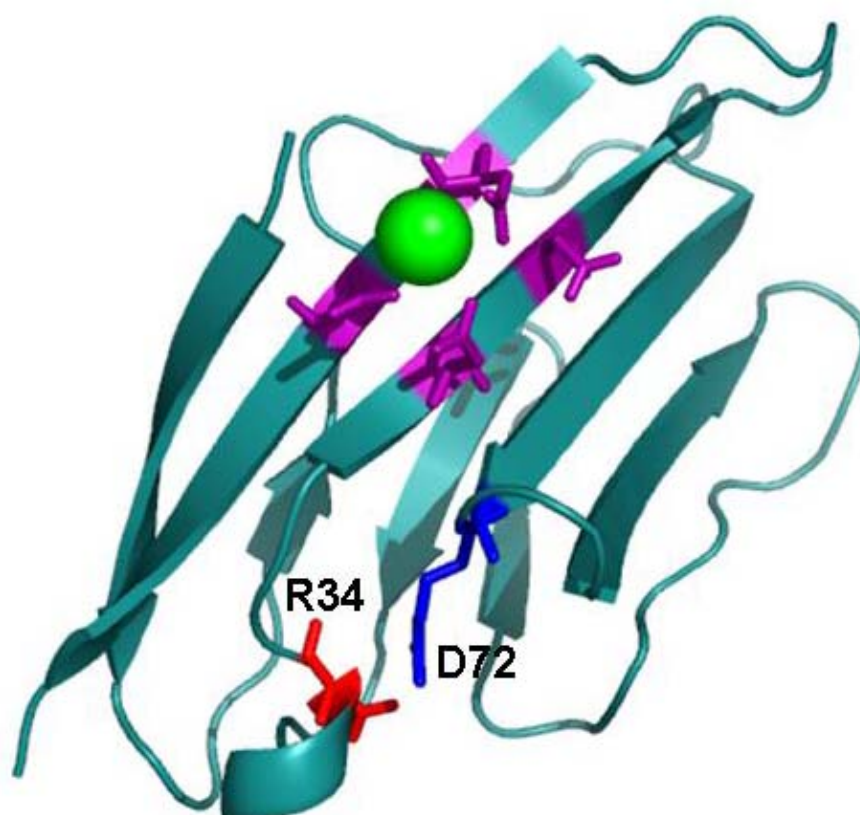


Figure 8.1. Position of residue R34. Residue R34 (red) is located 9.4 Å from the Ca²⁺ binding site. The close proximity of residues R34 and D72 (blue) to one another may be allowing for the formation of salt bridge between the two residues.

proteins with an Ig-fold mediated by electrostatic interactions. Table 8.1 summarizes all of the mutations discussed in this chapter.

8.2 Conformational analysis of the R34 mutants

The conformation of the R34 mutants was examined using far UV CD. The spectrum indicates 6D79.R34E has an alternate conformation with a small shoulder at 222 nm and negative maximum at 205 nm (Figure 8.2). This altered conformation is similar to that observed for the first generation proteins. The spectrum appears to be a combination of native β -sheet and random coil conformation.

The crystal structure of CD2 modified by Dezymer to include the 6D79 binding site was examined to determine the origins of the altered conformation. The structural analysis revealed that R34 is in close proximity to residue D72 which is on the opposite surface of the protein (Figure 8.1). The two residues are presumably forming a salt bridge with one another. The mutation of Arg to Ile does not alter the conformation of the protein but the introduction of the negatively charged Glu at this position instigates electrostatic repulsion which triggers protein unfolding.

8.3 pH titration using far UV CD and fluorescence

To ensure the hypothesis of charge repulsion induced unfolding was correct, a pH titration was performed using far UV CD. If the altered conformation was in fact due to an electrostatic effect you would expect the protein to be correctly folded at low pH. Residue D72 in CD2 has a pKa value of 4.14 (Chen et al., 2000) and unfolded random coil have a value of 4.5 (Stryer, 1995). In principle, at a pH lower than the pKa, both the E34 and D72 residues will be protonated and therefore the charge repulsion between the

Protein	The effect for protein folding
6D79.R34E	Charge reversal
6D79.R34I	Charge removal
6D79.R34E-D72K	Charge reversal
6D79.R34D-D72K	Size

Table 8.1. Proteins studied for protein folding

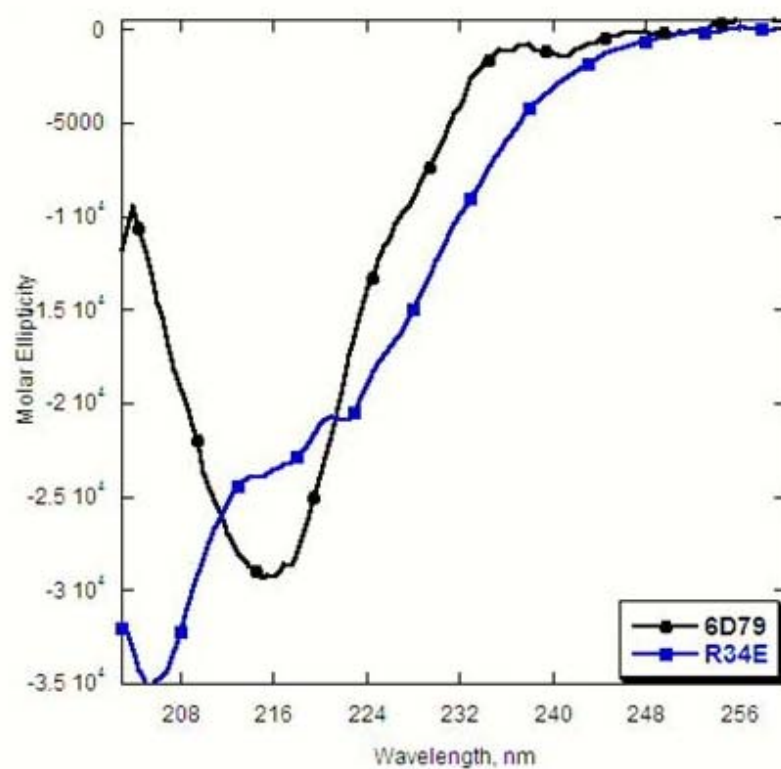


Figure 8.2. Conformational analysis of the R34 mutants. Conformational analysis of 6D79.R34E (3 μ M) compared to CD2.6D79 using far UV CD in 20 mM PIPES, 10 mM KCl pH 6.8 reveal an alternate conformation of the 6D79.R34E variant

two residues would be minimized inducing proper protein folding. On the other hand, as the pH is raised to and above the pK_a of these residues, they will be de-protonated and the negative charge repulsion would initiate protein unfolding.

A pH titration of the 6D79.R34E protein was carried out from pH 4-9 using far UV CD and Trp fluorescence. The titration was done under two conditions, one with 1 mM EGTA and the other with 10 mM Ca^{2+} , in order to determine if Ca^{2+} binding will have an effect on protein folding. At a pH of 7.4 the protein is not folded in the apo form (Figure 8.3a-8.4a). However, when the pH is decreased to 4 the protein is properly folded with a negative maximum of 216 nm in the CD spectra and a fluorescence emission of 320 nm. At pH 6.5, the apo form of the protein is unfolded with the negative maxima shifting to 205 nm. In the Trp fluorescence spectra, at pH 6.5, the fluorescence emission is shifted to 350 nm. At all pH values above 6.5 the protein in the apo form is partially folded.

As discussed in chapters 6 and 7, Ca^{2+} binding may interact with charged residues around the metal binding site. Residue R34 is within a suitable distance to the metal binding site where Ca^{2+} binding may reduce the electrostatic repulsion between E34 and D72. The far UV CD and Trp fluorescence spectra for the Ca^{2+} bound form of the protein at pH 7.4 indicate the protein is partially folded (Figure 8.3b-8.4b). The negative maximum of the CD spectra is 205 nm while the fluorescence emission is at 350 nm. When the pH is lowered to 4 the protein refolds to its proper conformation. Similar to the apo form of the protein, the protein begins to unfold at pH 6.5. This indicates that Ca^{2+} is not playing a role in the folding of this protein. Since D72 is on the opposite surface of

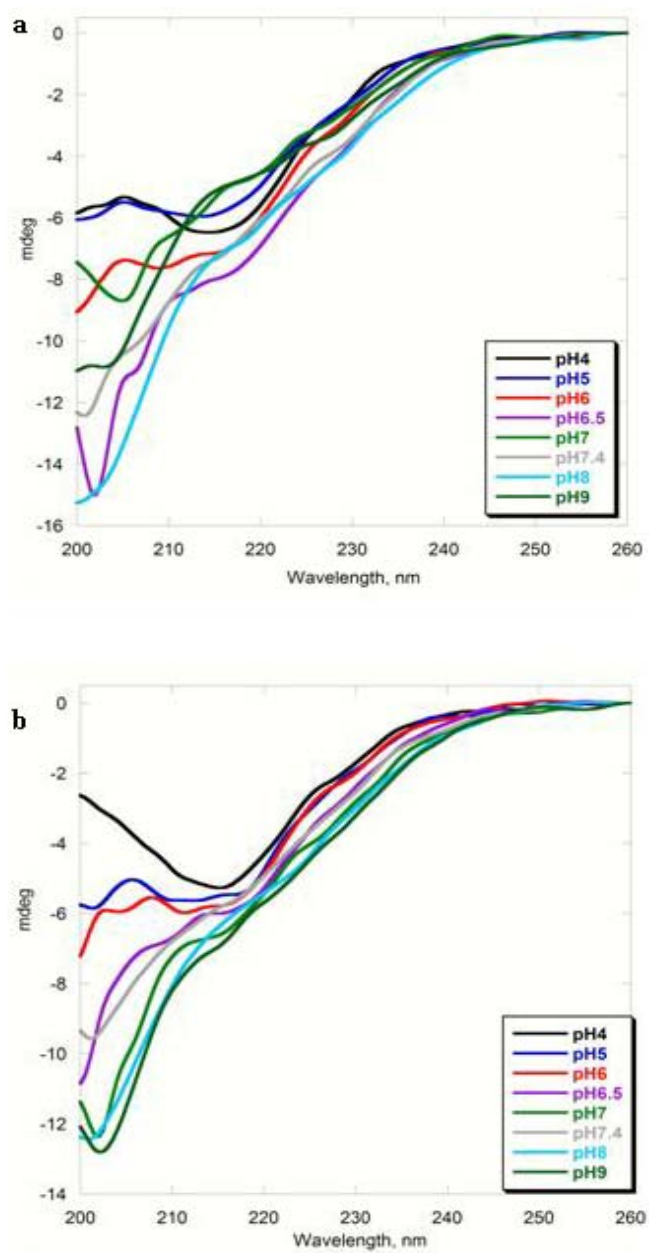


Figure 8.3. pH titration of 6D79.R34E using far UV CD. pH titration of R34E with 1 mM EGTA (a) and 10 mM Ca^{2+} (b)

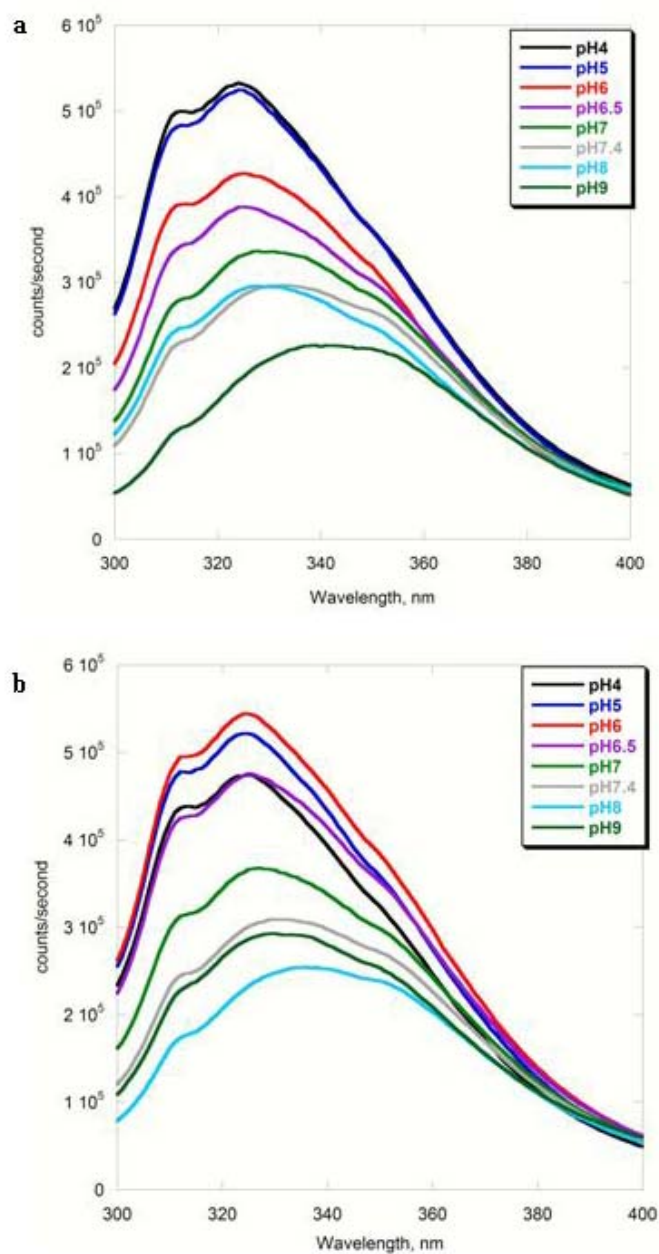


Figure 8.4. pH titration of 6D79.R34E using Trp fluorescence. pH titration of R34E with 1 mM EGTA (a) and 10 mM Ca^{2+} (b) excited at 283 nm at 25 °C.

the protein than the metal binding site, the distance may too great for Ca^{2+} to reduce the electrostatic repulsion.

8.4 Removing the charge by R34I

In an effort to reduce the charge repulsion between E34 and D72, the variant R34I was engineered. The protein was engineered into CD2.6D79 using *Pfu* DNA polymerase. Conformational analysis of 6D79.R34I was performed using far UV CD. The spectrum indicates the protein has a similar conformation to CD2.6D79 (Figure 8.5). The negative maximum of the spectrum is 216 nm indicating the protein has β -sheet architecture. The removal of the negative charge at position 34 does trigger the protein to refold.

The stability of the R34I mutant was determined through thermal denaturation studies. The apo form of the protein had a T_m of 43 °C (Figure 8.6). This is a large decrease in stability compared to wild type CD2 which has a T_m of 62°. The removal of the salt bridge between R34 and D72 caused a substantial decrease in the stability of the protein. Previous studies on the heat stable thermozymes indicate that even small changes in structure can lead to large changes in thermostability (Vieille et al., 2001). For example the presence of a few additional hydrophobic bonds in a thermozyme will increased the free energy by 5-7 kcal/mol. Free energy differences as small as these can account for thermostability increases of up to 12 °C. This decreased protein stability may lead to reduced purification yields due to the formation of inclusion bodies. Additionally, upon Ca^{2+} binding a slight increase in the stability of the protein is observed, 47 °C (Figure 8.6).

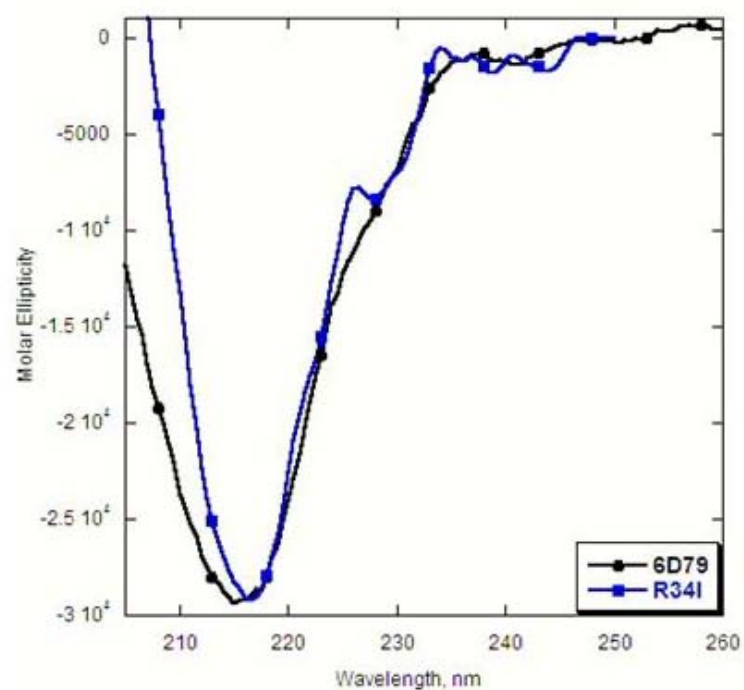


Figure 8.5. Conformational analysis of the 6D79.R34I variant. Conformational analysis of 6D79.R34I (3 μ M) compared to CD2.6D79 using far UV CD in 20 mM PIPES, 10 mM KCl pH 6.8.

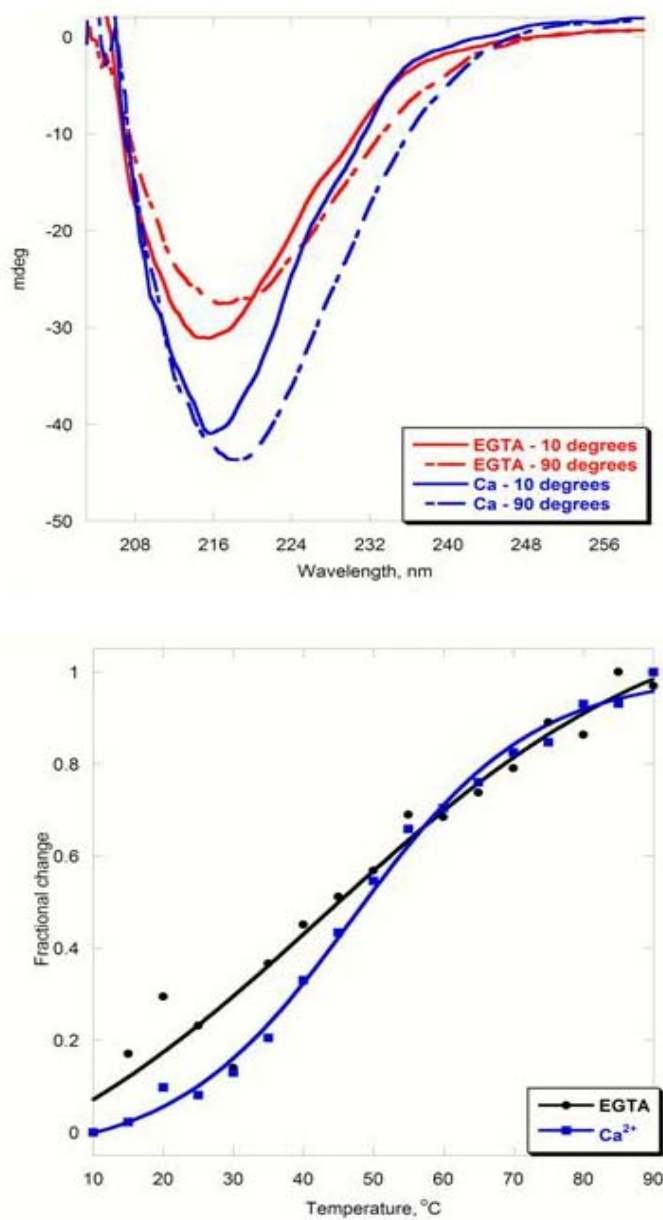


Figure 8.6. Thermal denaturation of 6D79.R34I. Thermal denaturation of the 6D79.R34I variant with 1 mM EGTA and 10 mM Ca²⁺ in 10 mM Tris, 10 mM KCl pH 7.0.

8.5 Reducing charge repulsion by reversing charge interactions (the R34E-D72K mutant)

In order to test the hypothesis that the regain of native conformation can be achieved by neutralizing introduced charged ligand residues, residue D72 was mutated to Lys in 6D79.R34E. As shown in Figure 8.7, the D72 residue is on the opposite face of the metal binding site and even though it interacts with R34 does not have any significant contribution to the electrostatic environment of the binding site. This was established through electrostatic potential calculations where the residue was mutated and the electrostatic potential of the Ca^{2+} binding site was examined.

The D72K mutation was further engineered using *Pfu* DNA polymerase. The conformation of the protein was examined using far UV CD. The spectra indicate the 6D79.R34E-D72K variant is largely unfolded (Figure 8.8). The negative maximum of the spectrum is shifted to 205 nm and the binding of Ca^{2+} has no effect on the folding of the protein. This was puzzling since the pH dependent conformational analyses indicated that the unfolding of the R34E originated from electrostatic repulsion. The structure of the protein was analyzed to ascertain the origin of the altered conformation of the R34E-D72K variant. The structure reveals that both R34 and D72 are facing the inside core of the protein (Figure 8.9). Residues Glu and Lys both have a larger radius than Asp and Arg. The volumes of Arg and Asp are 173.4 and 111.1 Å³, respectively. The volumes of Glu and Lys are 143.8 and 168.6 Å³, respectively (Stryer, 1995). The steric hindrance of these bulkier residues in the interior of the protein is initiating the altered conformation of the R34E-D72K protein.

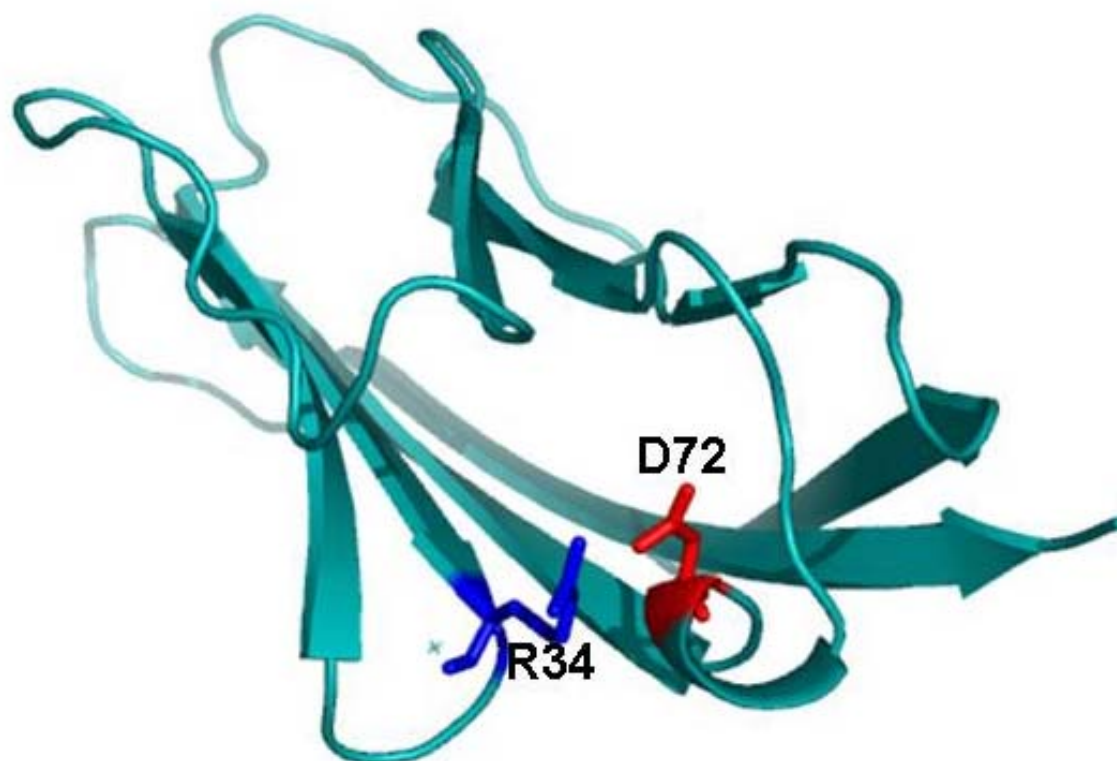


Figure 8.7. Residues R34 and D72 face the interior of the protein.
Residues R34 and D72 face the interior of the protein

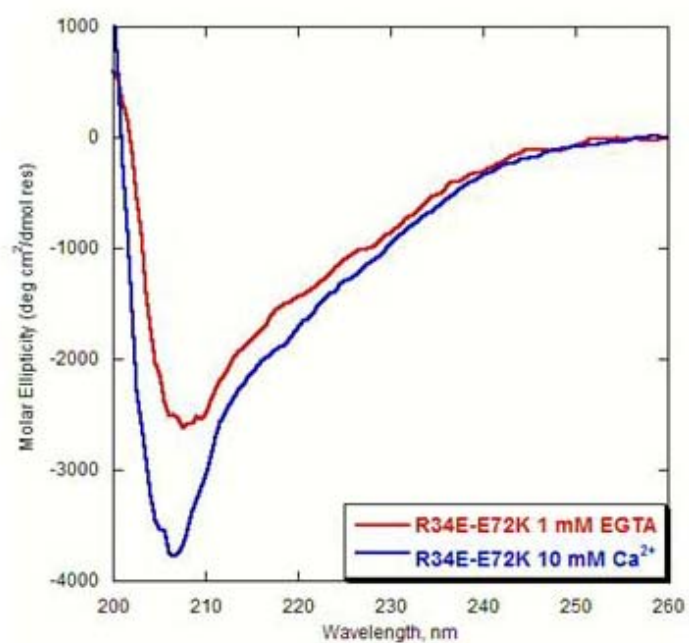


Figure 8.8. Conformational analysis of 6D79.R34E-D72K Conformational analyses of 6D79.R34E-D72K (3 μ M) in 10 mM Tris pH 7.4 using far UV CD.

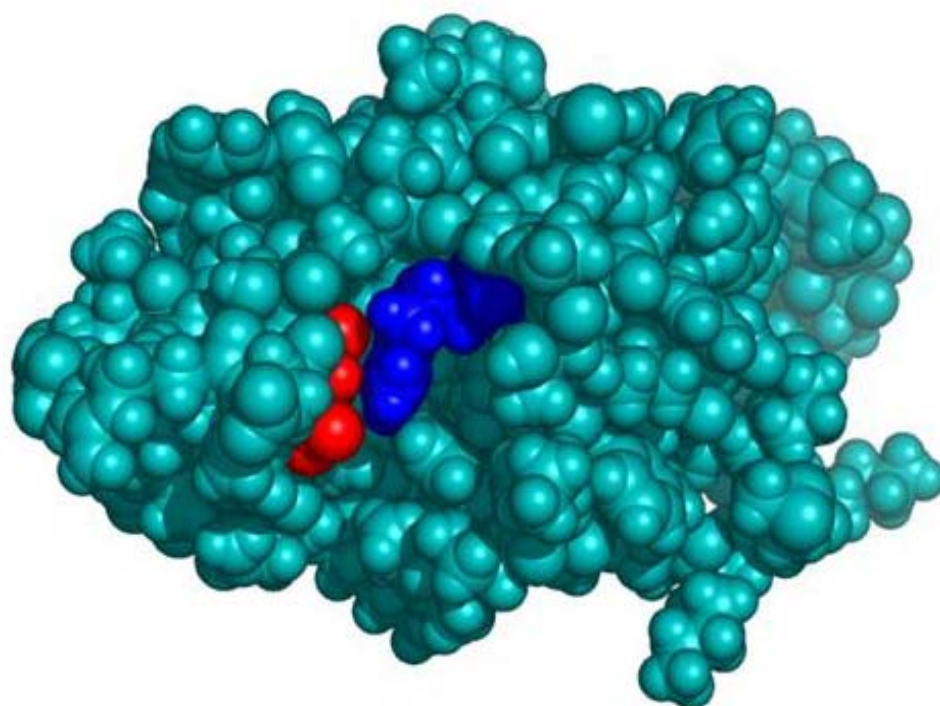


Figure 8.9. Space-filling model of CD2.6D79. The space-filling model of CD2.6D79 shows that residues R34 (blue) and D72 (red) are located towards the interior of the protein.

To further establish whether the R34 variant can be properly folded R34 was mutated to the less bulky Asp. The far UV CD and Trp fluorescence of the R34D-D72K variant reveals the protein exhibits a β -sheet conformation with proper tertiary packing (Figure 8.10). The far UV spectrum has a negative maximum at 216 nm, typical of a β -sheet protein. The Trp fluorescence spectrum reveals the protein has an emission maximum of 326 nm, which is similar or close to wild type CD2 and 6D79. The less bulky residue Asp reduces the steric hindrance between the 34 and 72 residues and the D72K mutation reduces the electrostatic repulsion between the two residues, therefore the protein, R34D-D72K is properly folded.

8.6 Comparison of purification yields of the R34 mutants

The purification yields of the R34 mutants have been poor in comparison with other charge distribution variants. For example the purification yields of the folded R34 variants, R34I and R34D-D72K, are only 1 and 1.3 mg, respectively for a 4 L expression. In comparison, the purification yield of the charge distribution variant D94K, is 10 mg for a 4 L expression. For the R34 variants large amounts of protein are trapped in the cell pellet after sonication as inclusion bodies. The formation of inclusion bodies is generally attributed to less stable proteins and is observed with the first generation proteins whose instability is due to their altered conformation.

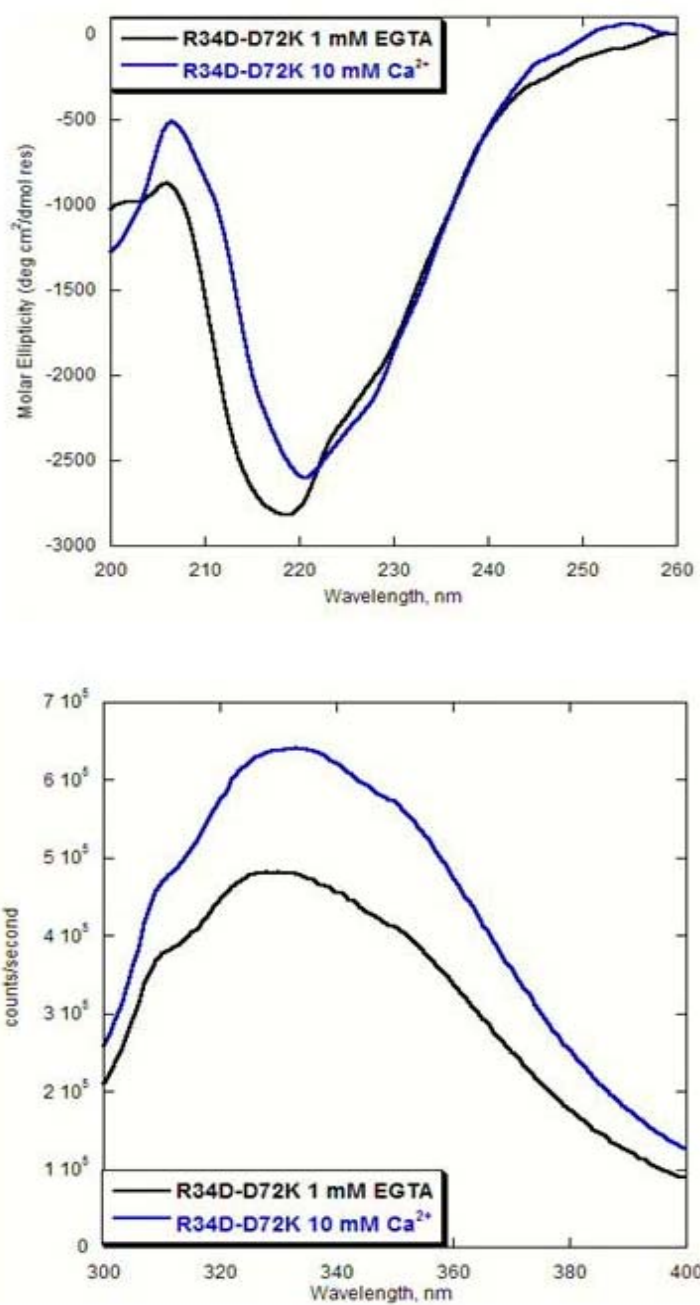


Figure 8.10. Conformational analyses of 6D79.R34D-D72K. Conformational analyses of 6D79.R34D-D72K (3 μ M) using far UV CD (a) and Trp fluorescence (b) in 10 mM Tris pH 7.4

8.6.1 Disruption of the native non-covalent interactions of CD2

The formation of a tight folding nucleus that includes three residues, I18, V39, and V78, is essential for the native fold of CD2 (Lorch et al., 1999). In addition non-covalent interactions are also important. However, previous studies of shown the formation of a complex hydrogen bonding network is also important for retaining the protein secondary structure (Parker and Clarke, 1997b). Our studies about the R34 variants give insight into one of the hot spots of this hydrogen bonding network and electrostatic interactions. Both R34 and D72 are involved in hydrogen bonding with other residues in close proximity including T37 and T69 (figure 8.11). The side chain of R34 is hydrogen bonded to the backbone of D72. This lost hydrogen bond may also play a role in the decreased stability of the R34 variants. Hydrogen bonds between buried charged residues have been shown to be favored over salt bridges and neutral-neutral hydrogen bonds in thermozymes for two reasons: 1. the entropy cost for burying a charged-neutral H-bond is lower than for burying a salt bridge; and 2. the entropy gained by burying a charged-neutral hydrogen bond is greater than for burying a neutral-neutral hydrogen bond (Li et al., 2005). We have already discussed in **section 7.3** how residues 34 and 72 face the interior of the protein. In fact residue 72 has a solvent accessibility of only 16%. The mutation of the R34 residue to Ile and to Asp, as well as the mutation of D72 to Lys, disrupts the hydrogen bonding network of the protein and therefore, although the protein conformation is similar to wild type CD2, the proteins are less stable. The experimental data correlate with the stability calculations by the Fold-X server in Table 3.11 where both 6D79.R34E and 6D79.R34I are less stable than CD2.6D79.

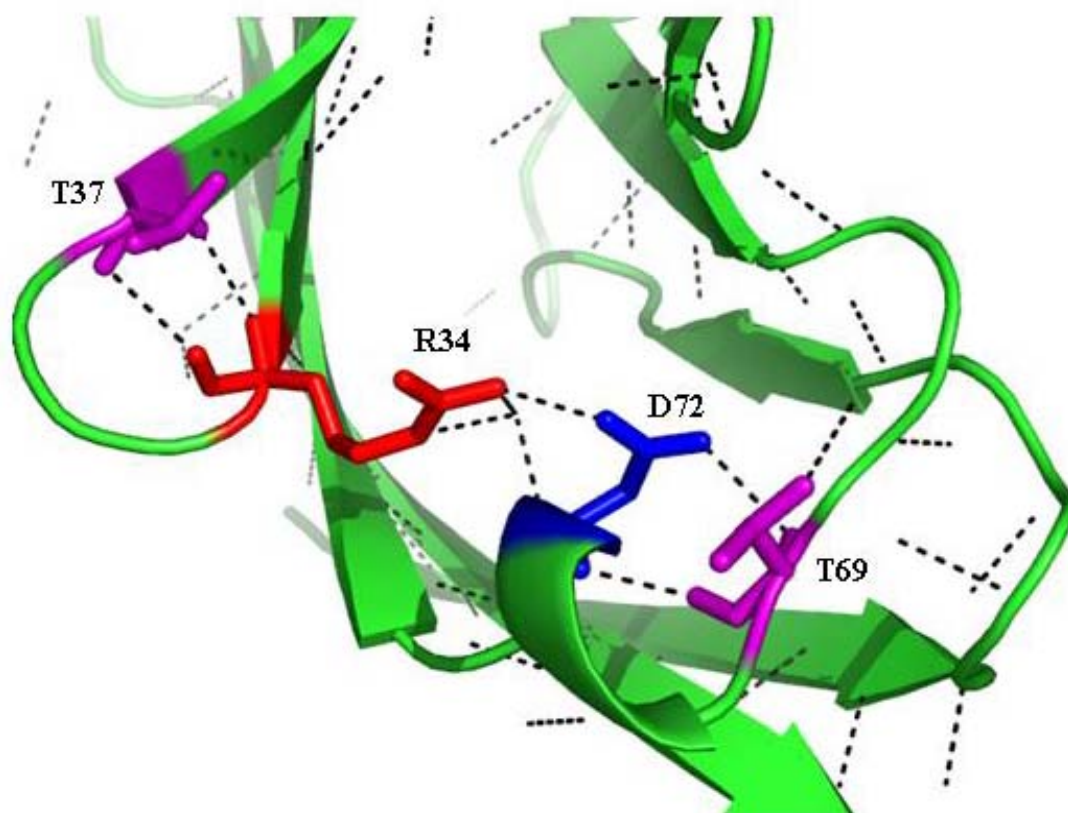


Figure 8.11. The hydrogen bonding network of CD2. The role of R34 and D72 in the hydrogen bonding network of CD2

This work on the 6D79.R34 variants has provided insight into the role of electrostatic interactions and hydrogen bonding in protein folding. This study supports our hypothesis that charge reversal can trigger protein re-folding. This work can be extrapolated to elucidate the folding properties of other proteins.

Chapter 9. Significance of this study

9.1 Review

Calcium binding proteins play a role in numerous biological functions including neuron transmission, muscle contraction, and cell adhesion. A common feature of Ca^{2+} binding proteins is multiple Ca^{2+} binding sites, which make it difficult to study Ca^{2+} binding affinity due to cooperativity between these sites. However, there is still numerous information available about the binding affinities of different Ca^{2+} binding proteins. Ca^{2+} binding proteins have various binding affinities based on their function. Ca^{2+} binding proteins have not evolved to the strongest possible affinity but rather to the most suitable for their function. This raises the question as to what are the key factors responsible for the varying binding affinities of these proteins. The study of the factors involved in Ca^{2+} binding affinity is important because of the numerous diseases involved in improper Ca^{2+} binding affinity. Ca^{2+} binding plays a role in many diseases including Alzheimer's disease, diabetes, and cardiovascular disease.

Local factors such as charge in the coordination shell and ligand type have been shown to be important for Ca^{2+} binding affinity. Additionally, non-local factors including an extensive hydrogen bonding network and electrostatic interactions have also been shown to be important. This study aims to investigate the role of electrostatic interactions around the coordination shell, not directly involved in binding, in Ca^{2+} binding affinity. This study attempts to answer four questions: 1. What are the structural requirements of Ca^{2+} binding sites? 2. What effect does the

electrostatic environment have on molecular recognition? 3. What effect do electrostatic interactions around the coordination shell have on thermal stability and Ca^{2+} binding affinity? 4. What is the effect of electrostatic interactions on protein folding?

9.2 The structural requirements of Ca^{2+} binding sites

Chapter 5 details the study of the structural requirements of Ca^{2+} binding sites through the elucidation of the conformational and metal binding properties of the designed Ca^{2+} binding protein CD2.6D79. This protein utilizes ligands all from β -sheet secondary structure. The protein binds Ca^{2+} and its analogues Tb^{3+} and La^{3+} with selectivity over other physiologically relevant metals. However, the designed protein is less stable than wild type CD2. This decrease in stability is due to the electrostatic repulsion between clustered negative charge residues at the binding site. Many natural Ca^{2+} binding proteins undergo a global conformational change upon Ca^{2+} binding. For CD2.6D79, in the absence of Ca^{2+} , the fixed backbone conformation of the β -sheets does not allow for a decrease in the charge repulsion through conformational change. It also limits the use of local conformational change to complete neutralization of CD2.6D79 by Ca^{2+} binding. Hence, calcium binding only leads to a small increase in the thermal stability. Consistently, a greater gain of the protein stability was achieved by binding of La^{3+} and Tb^{3+} with a higher positive charge.

Further, Ca^{2+} binding to the all- β -sheet location exhibits a less profound effect on the conformation of the protein than that of previously investigated Ca^{2+} binding proteins (Alattia et al., 1999; Chazin, 1995; Garcia et al., 2004; Ikura, 1996; Ubach et al., 1998; Yang et al., 2005). The Ca^{2+} or La^{3+} -induced chemical shift changes for the residues at

and around the Ca^{2+} binding pocket are very limited, < 0.05 ppm and 0.10 ppm, respectively, and even smaller than residues at the flexible loop regions in the same protein, suggesting that their shielding environments were not perturbed by Ca^{2+} binding. In our previously designed Ca^{2+} binding protein Ca.CD2, significant chemical shift changes (> 0.1 ppm) were observed for the two ligand residues in the loop regions (N60 and D62) while no significant change was detected for the ligand residue Asp-15 at the β -strand B. This observed differential response to Ca^{2+} binding is structural specific since Ca.CD2 has a 2.5 fold weaker Ca^{2+} binding affinity than that of CD2.6D79.

The large energetic cost in destabilizing the protein both in apo- and loaded forms and the lack of Ca^{2+} -induced conformational change explain why nature does not use all β -sheet ligand residues for Ca^{2+} binding. The observed strong preference of Asp and Glu in the naturally occurring Ca^{2+} -binding proteins is consistent with the electrostatic nature of Ca^{2+} binding since almost all the binding sites observed so far have at least two negatively charged ligand residues. The use of the ligand residues predominantly from the loops and flexible regions of the proteins serves to minimize the repulsion between charged ligand residues in the apo-form and maximize the electrostatic interaction of the loaded form by conformational change. Therefore, our design studies and structural analysis provide a close view for the structural determinants that are necessary for biological functions of Ca^{2+} binding in regulation and stability.

9.3 The effect of the electrostatic environment on molecular recognition

CD2.6D79 was also a useful model for understanding the role of electrostatic interactions on molecular recognition. The metal binding site is in close proximity to the CD48 binding surface of CD2. Studies have shown the large number of charges on this surface were responsible for target specificity rather than affinity (Davis et al., 1998a). Most single mutations of charged residues on CD2 (to Ala or an opposite charge) decreases (or has no effect) on ligand binding except R87A (CD2.R87A), which results in a 5-fold increase (Davis et al., 1998a). A three-fold decrease in the dissociation constant for CD48 binding was observed for CD2.6D79 compared with CD2 in the absence and presence of Ca^{2+} or La^{3+} suggesting that the CD48-binding ability of CD2 is enhanced by altering the electrostatic environment with two charged mutations for the Ca^{2+} binding. Further, electrostatic calculations reveal a potential decrease has occurred in residue 87. A similar potential decrease was observed in residue 87 in CD2.R87A which displays a five-fold increase in CD48 binding. The mutation in CD2.R87A reverses the potential of residue 87 from positive in CD2 to negative. Long range electrostatic steering introduced by the charged mutations for the Ca^{2+} binding is responsible for the increased binding affinity of CD2.6D79 to CD48. This finding is significant because it was widely understood the electrostatic interactions in CD2-CD48 binding were for target specificity rather than affinity (Davis et al., 1998a). CD2 is expressed on most human T cells and natural killer (NK) cells and initiates and maintains contact with antigen-presenting or target cells and play an important role in signaling (Blattman and Greenberg, 2004; Driscoll et al., 1991; van der Merwe et al., 1995). It is

one of the most extensively studied cell adhesion proteins. The detailed examination of the binding characteristics of CD2 and CD48 will provide insight into the properties of cell to cell interactions since this binding mode is a distinguishing characteristic of many types of molecular recognitions especially for cell adhesion proteins. In addition, the extracellular domain of CD2 that is involved in molecular recognition has a similar topology to that of the Ca^{2+} -dependent cell adhesion molecule cadherin. Therefore, the information for this model system can be applied to the understanding of Ca^{2+} dependent molecular recognition. Further mutations can be introduced to design Ca^{2+} dependent cell adhesion proteins.

9.4 The role of electrostatic interactions around the coordination shell on thermal stability and Ca^{2+} binding affinity

In order to study the role of long range electrostatic interactions on thermal stability and Ca^{2+} binding affinity charge distribution variants of CD2.6D79 were designed and engineered (**chapters 6-7**). The thermal denaturation data indicate that long range electrostatic interactions do play a role in the stability of the protein. In addition, Ca^{2+} , through its electrostatic properties can alter the stability of the protein. Previous studies have shown the protein stabilization of Ca^{2+} was due to its binding to the flexible loop regions of proteins. Ca^{2+} serves to order and rigidify these regions of the proteins making them more stable. Increasing the thermal stability of proteins is useful in the area of enzymes. The increased stability of enzymes could lead to more efficient catalysis by these proteins. In addition, in the area of pharmaceutical design the data we have collected based on the role of long range electrostatic interactions on stability will

be invaluable in designing stable drugs. Increased protein stability may be of multi-billion dollar value in food and drug processing, and in biotechnology, and protein drugs.

The charge distribution variants of CD2.6D79 were also useful for the study of the role of long range electrostatic interactions on Ca^{2+} binding affinity. Electrostatic calculations reveal that the mutations of several residues around the Ca^{2+} binding site will markedly alter the potential of the binding site (**section 3.5**). The metal binding affinities of these variants with altered electrostatic potentials at the Ca^{2+} binding site were measured. The data revealed that long range electrostatic do play a role in metal binding (**chapter 7**) and this role is distant dependent. Since Ca binding proteins are involved in many biological systems and improper Ca affinities of these proteins can lead to several disease states, the understanding of the key factors that govern Ca^{2+} binding affinity is important. Several studies have shown the importance of electrostatic interactions around the coordination on Ca^{2+} binding affinity on a macroscopic level (Linse and Forsen, 1995). This work looks at the site specific effects of electrostatic interactions in a static system. This work can help in establishing a set of rules for Ca^{2+} binding affinity and lead to an increased understanding of the improper Ca^{2+} binding affinities of proteins. This can lead to advancement in the area of several diseases related to improper Ca^{2+} binding affinities.

9.5 The role of electrostatic interactions for protein folding

Several factors are important for protein folding including hydrophobic and electrostatic interactions. As discussed in chapter 8, the R34 variants of CD2.6D79 were utilized to study the effect of electrostatic interactions on protein folding. The variant 6D79.R34E has a partially folded conformation due to electrostatic repulsion. However, the mutation to Ile triggers the protein to re-fold to a β -sheet conformation. This indicates that charge removal can instigate re-folding of proteins. The folding of proteins is an important step in understanding the functional roles of proteins. An understanding of the different interactions involved in protein folding is imperative to understand the folding process. Through our model system we have been able to elucidate the role of electrostatic interactions in protein folding. This work will provide will be useful in engineering proteins with desired structures.

9.6 Protein design as an important tool for studying the properties of proteins

Protein design is an invaluable tool in understanding the fundamental principles of proteins. We have successfully used protein design to answer significant questions about the metal binding and conformational properties of proteins. This work demonstrates the power and impact of protein design providing insight into the structure, folding, and function of proteins. In addition this work substantiates protein design as a suitable method in the development of new biomaterials, sensors, and pharmaceutical drugs as well as in intervening biological systems with tailored functions

Publications

1. **Lisa M. Jones**, Wei Yang, Anna L. Wilkins, Alice Kearney, P. Anton van der Merwe, and Jenny J. Yang. Rational design of a novel calcium binding site adjacent to the ligand binding site on CD2 increases its adhesion function. *Protein Engineering Design and Selection* submitted.
2. Wei Yang, **Lisa M. Jones**, Leanne Isley, Yiming Ye, Hsui-we Lee, Anna L. Wilkins, Zhi-ren Liu, Homme W. Hellinga, Russell Malchow, Mohammed Ghazi, and Jenny J. Yang. Rational design of a calcium-binding protein. *JACS* 2003; 125:6165-6171.

Appendix

Assignments of the TOCSY spectra of CD2.6D79 with 5 mM EGTA

Assignment	w1	w2
THR5H-HA	8.263	4.85
THR5H-HB	8.262	3.75
THR5H-HG	8.261	0.445
VAL6H-HA	9.155	4.274
VAL6H-HB	9.154	1.787
VAL6H-HG	9.158	0.885
TRP7H-HA	8.665	5.267
TRP7H-HB	8.664	2.94
GLY8H-HA2	8.992	4.597
GLY8H-HA3	8.992	3.251
ALA9H-HA	8.406	4.817
ALA9H-HB	8.408	1.106
LEU10H-HA	8.838	3.77
GLY11H-HA2	9.452	4.053
GLY11H-HA3	9.453	3.802
HIS12H-HA	7.962	4.833
HIS12H-HB2	7.963	3.245

HIS12H-HB3	7.963	3.146
GLY13H-HA2	8.773	3.487
GLY13H-HA2	8.77	4.614
ILE14H-HA	8.801	4.491
ILE14H-HB	8.798	1.825
ASN15H-HA	8.046	5.264
ASN15H-HB2	8.043	2.546
ASN15H-HB3	8.041	2.17
LEU16H-HA	9.546	4.24
ASN17H-HA	8.098	4.82
ASN17H-HB2	8.097	2.666
ASN17H-HB3	8.099	2.483
ILE18H-HA	8.638	3.897
ASN20H-HA	8.987	4.245
ASN20H-HB2	8.985	2.953
ASN20H-HB3	8.986	2.698
PHE21H-HA	7.723	4.733
PHE21H-HA	8.884	4.69
PHE21H-HB2	8.885	3.233
PHE21H-HB2	7.723	2.752

PHE21H-HB3	8.884	2.931
PHE21H-HB3	7.721	2.428
GLN22H-HA	7.222	4.178
GLN22H-HB2	7.224	1.688
GLN22H-HB3	7.224	1.606
GLN22H-HG	7.224	2.08
MET23H-HA	8.548	4.044
THR24H-HA	6.994	4.654
THR24H-HB	6.988	4.466
ASP26H-HA	8.313	4.666
ILE27H-HA	7.655	4.126
ILE27H-HB	7.656	2.613
ILE27H-HG	7.656	0.95
ASP28H-HA	9.088	4.932
ASP28H-HB2	9.089	2.36
ASP28H-HB3	9.088	2.243
ASP29H-HA	7.885	5.639
ASP29H-HB2	7.886	2.065
ASP29H-HB3	7.885	1.973
VAL30H-HA	9.164	4.742
VAL30H-HB	9.171	2.11
VAL30H-HG	9.168	0.799

ARG31H-HA	9.398	5.173
ARG31H-HB2	9.395	1.933
ARG31H-HB3	9.401	1.861
TRP32H-HA	9.459	5.537
TRP32H-HB2	9.455	3.137
TRP32H-HB3	9.464	2.968
GLU33H-HA	9.802	5.202
GLU33H-HB2	9.803	1.961
GLU33H-HB3	9.798	1.821
GLU33H-HG	9.806	2.181
ARG34H-HA	8.948	4.458
ARG34H-HB2	8.948	1.618
ARG34H-HB3	8.949	1.525
THR37H-HA	8.325	4.252
LEU38H-HA	8.773	4.33
VAL39H-HA	9.253	4.167
VAL39H-HB	9.249	1.432
VAL39H-HG	9.254	0.341
ALA40H-HA	8.075	5.283
ALA40H-HB	8.076	1.325
GLU41H-HA	9.621	5.541
GLU41H-HB2	9.621	2.279

GLU41H-HB3	9.623	2.09
PHE42H-HA	9.628	4.949
PHE42H-HB2	9.632	3.149
PHE42H-HB3	9.631	3.048
LYS43H-HA	8.458	4.867
LYS43H-HB2	8.459	1.636
LYS43H-HB3	8.459	1.431
LYS43H-HG	8.459	1.258
MET46H-HA	8.031	4.78
MET46H-HB2	8.033	2.065
MET46H-HB3	8.031	1.838
MET46H-HG	8.032	2.46
LYS47H-HA	8.533	4.533
LYS47H-HB2	8.535	1.728
LYS47H-HB3	8.533	1.559
LYS47H-HG	8.533	1.808
LEU50H-HA	7.791	4.583
LEU50H-HB2	7.795	1.59
LEU50H-HB3	7.791	1.49
LEU50H-HG	7.794	1.676
SER52H-HA	7.831	4.598
ALA54H-HA	8.166	4.01

ALA54H-HB	8.168	0.839
PHE55H-HA	7.383	5.411
PHE55H-HB3	7.386	2.718
PHE55H-HB3	7.382	2.937
GLU56H-HA	9.011	4.446
GLU56H-HB2	9.011	1.834
GLU56H-HB3	9.012	1.739
GLU56H-HG	9.01	2.072
ILE57H-HA	8.682	4.998
LEU58H-HA	9.115	4.442
LEU58H-HB2	9.115	1.97
LEU58H-HB3	9.115	1.641
ASN60H-HA	7.752	4.454
GLY61H-HA2	8.404	4.279
GLY61H-HA3	8.403	3.772
ASP62H-HA	7.887	4.681
ASP62H-HB2	8.317	2.675
ASP62H-HB3	8.313	2.538
LEU63H-HA	7.394	4.316
LEU63H-HB	7.394	0.232
LYS64H-HA	9.178	4.966
LYS64H-HB2	9.174	1.591

LYS64H-HB3	9.188	1.387
ILE65H-HA	8.778	4.197
ILE65H-HB	8.776	1.611
ILE65H-HG	8.779	0.528
LYS66H-HA	7.963	3.663
ASN67H-HA	8.638	3.951
ASN67H-HB2	8.636	2.487
ASN67H-HB3	8.634	2.255
LEU68H-HA	8.254	4.008
THR69H-HA	9.367	4.756
THR69H-HB	9.373	4.42
ARG70H-HA	9.144	3.825
ASP71H-HA	7.981	4.48
ASP71H-HB2	7.98	2.7
ASP71H-HB3	7.983	2.412
ASP72H-HA	7.75	4.744
ASP72H-HB2	7.74	2.837
ASP72H-HB3	7.741	2.665
SER73H-HA	7.584	4.187
SER73H-HB2	7.582	4.123
SER73H-HB3	7.584	3.854
THR75H-HA	8.438	5.142

THR75H-HB	8.438	3.885
TYR76H-HA	10.017	5.178
TYR76H-HB2	10.014	3.107
TYR76H-HB3	10.017	3.014
ASN77H-HA	9.618	5.487
ASN77H-HB2	9.619	2.615
ASN77H-HB3	9.618	2.501
VAL78H-HA	8.843	4.918
ASP79H-HA	8.474	5.312
ASP79H-HB2	8.476	2.461
ASP79H-HB3	8.476	2.274
VAL80H-HA	8.185	4.857
VAL80H-HB	8.184	1.647
VAL80H-HG	8.185	0.352
TYR81H-HA	9.5	5.464
TYR81H-HB	9.499	2.941
SER82H-HA	9.346	5.323
SER82H-HB	9.344	3.894
T83H-HA	8.826	3.939
THR86H-HA	7.866	4.208
THR86H-HB	7.866	4.111
ARG87H-HA	9.112	4.03

ILE88H-HA	8.856	4.169
LEU89H-HA	7.432	4.418
LEU89H-HB2	7.432	1.472
LEU89H-HB3	7.432	1.32
ASN90H-HA	8.339	5.592
ASN90H-HB3	8.34	2.302
LYS91H-HA	8.723	4.691
LYS91H-HB2	8.724	1.632
LYS91H-HB3	8.724	1.365
GLU92H-HA	8.396	5.343
GLU92H-HB2	8.399	2.072
GLU92H-HB3	8.396	1.86
GLU92H-HG	8.394	2.193
LEU93H-HA	9.577	4.939
LEU93H-HB2	9.577	1.683
LEU93H-HB3	9.576	1.46
ASP94H-HA	8.847	5.021
ASP94H-HB2	8.843	2.544
ASP94H-HB3	8.843	2.47
ARG96H-HA	9.089	4.448
ARG96H-HB2	9.091	1.113
ARG96H-HB3	9.084	0.132

ILE97H-HA	8.521	4.657
ILE97H-HB	8.517	0.846
ILE97H-HG	8.522	1.913
GLU99H-HA	8.233	4.197
GLU99H-HB2	8.233	2.06
GLU99H-HB3	8.233	1.899
GLU99H-HG	8.233	2.225

Assignments of the TOCSY spectra of CD2.6D79 with 10 mM Ca²⁺

VAL6H-HA	9.148	4.279
VAL6H-HB	9.147	1.787
VAL6H-HG	9.148	0.881
TRP7H-HA	8.666	5.26
TRP7H-HB	8.665	2.943
GLY8H-HA2	9.014	4.602
GLY8H-HA3	9.012	3.229
ALA9H-HA	8.424	4.845
ALA9H-HB	8.426	1.105
LEU10H-HA	8.866	3.774
GLY11H-HA2	9.49	4.031
GLY11H-HA3	9.492	3.862

HIS12H-HA	7.999	4.899
HIS12H-HB2	8	3.356
HIS12H-HB3	8.002	3.243
GLY13H-HA2	8.794	4.626
GLY13H-HA3	8.796	3.523
ILE14H-HA	8.808	4.496
ASN15H-HA	8.023	5.273
ASN15H-HB2	8.023	2.553
ASN15H-HB3	8.021	2.179
LEU16H-HA	9.545	4.245
LEU16H-HB2	9.543	1.205
LEU16H-HB3	9.547	0.983
ASN17H-HA	8.117	4.836
ASN17H-HB2	8.113	2.665
ASN17H-HB3	8.118	2.495
ILE18H-HA	8.61	3.907
ASN20H-HA	8.989	4.248
PHE21H-HA	7.707	4.744
PHE21H-HB2	7.72	2.758
PHE21H-HB3	7.72	2.429
GLN22H-HA	7.198	4.176
GLN22H-HB2	7.196	1.69

GLN22H-HB3	7.199	1.603
GLN22H-HG	7.199	2.08
MET23H-HA	8.546	4.045
THR24H-HA	6.98	4.659
THR24H-HB	6.973	4.478
ASP26H-HA	8.283	4.676
ILE27H-HA	7.66	4.134
ILE27H-HG	7.661	0.96
ILE27HN-HB	7.654	2.631
ASP28H-HA	9.102	4.931
ASP28H-HB2	9.102	2.319
ASP28H-HB3	9.095	1.968
ASP29H-HA	7.901	5.658
ASP29H-HB2	7.901	2.07
ASP29H-HB3	7.903	1.926
VAL30H-HA	9.218	4.723
VAL30H-HB	9.219	2.127
VAL30H-HG2	9.219	0.893
VAL30H-HG3	9.22	0.794
ARG31H-HA	9.443	5.145
TRP32H-HA	9.462	5.495
TRP32H-HB2	9.452	3.151

TRP32H-HB3	9.468	2.951
GLU33H-HA	9.833	5.243
GLU33H-HB2	9.838	1.959
GLU33HA-HB3	9.832	1.84
ARG34H-HA	8.955	4.47
THR37H-HA	8.316	4.257
LEU38H-HA	8.812	4.347
VAL39H-HA	9.246	4.211
VAL39H-HB	9.243	1.418
VAL39H-HG2	9.249	0.653
VAL39H-HG3	9.245	0.379
ALA40H-HA	8.049	5.348
ALA40H-HB	8.049	1.328
GLU41H-HA	9.799	5.605
GLU41H-HB2	9.802	2.385
GLU41HA-HB3	9.795	2.274
PHE42H-HB2	9.536	3.186
PHE42HA-HB3	9.537	2.999
LYS43H-HA	8.567	4.872
LYS43H-HB2	8.564	1.591
LYS43H-HB3	8.569	1.436
LYS43H-HG	8.568	1.278

LYS45H-HA	8.846	3.862
MET46H-HA	8.112	4.805
MET46H-HB2	8.112	2.058
MET46H-HB3	8.11	1.79
MET46H-HG	8.1	2.449
LYS47H-HA	8.563	4.527
PHE49H-HA	8.861	4.71
PHE49H-HB2	8.862	3.228
PHE49H-HB3	8.862	2.872
LEU50H-HA	7.772	4.592
LEU50H-HB2	7.766	1.588
LEU50H-HB3	7.769	1.495
LYS51H-HA	8.626	3.743
SER52H-HA	7.78	4.612
SER52H-HB3	7.783	3.952
ALA54H-HA	8.123	4.013
ALA54H-HB	8.125	0.813
PHE55H-HA	7.332	5.427
PHE55H-HB2	7.332	2.962
PHE55H-HB3	7.328	2.715
GLU56H-HA	9.008	4.439
GLU56H-HB2	9.007	1.84

GLU56H-HB3	9.006	1.74
GLU56H-HG	9.008	2.082
ILE57H-HA	8.693	5.025
LEU58H-HA	9.098	4.456
ASN60H-HA	7.752	4.442
GLY61H-HA2	8.385	4.303
GLY61H-HA3	8.374	3.694
ASP62H-HA	7.899	4.688
ASP62H-HB3	7.901	2.674
ASP62H-HB3	7.902	2.495
LEU63H-HA	7.386	4.323
LEU63H-HB	7.386	0.22
LYS64H-HA	9.159	4.976
LYS64H-HB2	9.159	1.587
LYS64H-HB3	9.156	1.399
ILE65H-HA	8.764	4.215
LYS66H-HA	7.958	3.668
ASN67H-HA	8.685	3.947
LEU68H-HA	8.171	4.033
THR69H-HA	9.364	4.749
THR69H-HB	9.366	4.434
ARG70H-HA	9.143	3.824

ASP71H-HA	7.997	4.488
ASP71H-HB2	7.997	2.695
ASP71H-HB3	7.997	2.42
ASP72H-HA	7.74	4.747
ASP72H-HB2	7.744	2.842
ASP72H-HB3	7.738	2.659
SER73H-HA	7.588	4.186
SER73H-HB2	7.584	4.125
SER73H-HB3	7.585	3.861
GLY74H-HA2	8.857	4.483
THR75H-HA	8.445	3.889
THR75H-HA	8.446	5.132
TYR76H-HA	10.017	5.178
TYR76H-HB2	10.011	3.108
TYR76H-HB3	10.012	3.004
ASN77H-HA	9.618	5.496
ASN77H-HB2	9.617	2.611
ASN77HA-HB3	9.617	2.508
VAL78H-HA	8.85	4.929
ASP79H-HA	8.502	5.311
ASP79H-HB2	8.504	2.44
ASP79H-HB3	8.501	2.279

VAL80H-HA	8.198	4.867
VAL80H-HB	8.195	1.651
TYR81H-HA	9.484	5.499
TYR81H-HB	9.47	2.863
SER82H-HA	9.367	5.327
SER82H-HB	9.368	3.906
THR83H-HA	8.814	3.927
GLY85H-HA3	8.385	3.764
THR86H-HA	7.877	4.195
THR86H-HB	7.877	4.195
ARG87H-HA	9.123	4.058
ILE88H-HA	8.811	4.179
LEU89H-HA	7.425	4.43
LEU89H-HB2	7.422	1.476
LEU89H-HB3	7.426	1.308
ASN90H-HA	8.38	5.571
ASN90H-HB2	8.382	2.647
ASN90H-HB3	8.38	2.319
LYS91H-HA	8.711	4.694
LYS91H-HB2	8.711	1.633
LYS91H-HB2	8.711	1.372
LYS91H-HG	8.712	1.217

GLU92H-HA	8.424	5.325
GLU92H-HB2	8.426	2.093
GLU92H-HB3	8.426	1.872
GLU92H-HG	8.427	2.211
LEU93H-HA	9.562	4.942
LEU93H-HB2	9.561	1.682
LEU93H-HB3	9.556	1.46
ASP94H-HA	8.849	5.026
ASP94H-HB2	8.843	2.546
ASP94H-HB3	8.841	2.486
ARG96HA-H	9.078	4.452
ILE97H-HA	8.532	4.655
LEU98H-HA	8.962	4.389
GLU99H-HA	8.24	4.193
GLU99H-HB2	8.242	2.065
GLU99H-HB3	8.242	1.9
GLU99H-HG	8.237	2.236

Assignments of HSQC spectra with EGTA

Assignment	W1	w2
G4H-N	8.712	111.325
T5H-N	8.244	118.58
V6H-N	9.187	128.591
W7H-N	8.68	126.162
G8H-N	9.039	109.069
A9H-N	8.403	125.683
L10H-N	8.844	122.215
G11H-N	9.514	111.814
H12H-N	8.074	119.03
G13H-N	8.804	109.978
I14H-N	8.828	122.167
N15H-N	8.045	123.701
N15HD21-ND2	7.562	110.921
N15HD22-ND2	6.674	110.974
N17H-N	8.271	119.576
N17HD21-ND2	8.36	112.309
N20H-N	9.009	116.747
N20HD21-ND2	7.772	113.3
N20HD22-ND2	7.101	113.297

F21H-N	7.745	118.349
Q22H-N	7.218	126.79
Q22HE21-NE2	7.583	112.716
Q22HE22-NE2	6.916	112.747
M23H-N	8.573	124.172
T24H-N	7.006	116.982
D25H-N	8.847	120.714
D26H-N	8.316	116.077
I27H-N	7.728	120.491
E29H-N	7.913	115.852
V30H-N	9.233	125.34
R31H-N	9.458	127.101
W32H-N	9.481	124.671
E33H-N	9.834	123.417
R34H-N	8.872	123.244
G35H-N	9.026	118.893
S36H-N	7.61	114.864
L38H-N	8.819	130.162
V39H-N	9.271	124.983
A40H-N	8.131	119.746
E41H-N	9.792	121.834
F42H-N	9.587	128.496

K43H-N	8.564	129.922
M46H-N	8.118	118.576
F49H-N	8.884	122.608
L50H-N	7.793	127.708
S52H-N	7.83	108.898
G53H-N	8.836	106.192
A54H-N	8.157	121.936
F55H-N	7.381	115.108
E56H-N	8.987	119.416
L58H-N	9.118	129.608
N60H-N	7.782	112.433
N60HD21-ND2	7.388	108.931
N60HD22-ND2	6.699	108.936
G61H-N	8.348	108.304
D62H-N	7.922	119.782
L63H-N	7.404	120.88
K64H-N	9.183	127.999
I65H-N	8.784	127.297
K66H-N	7.999	125.33
N67H-N	8.689	117.007
N67H-N	7.391	111.26
N67HD22-ND2	6.715	111.221

T69H-N	9.393	118.504
R70H-N	9.166	120.601
D71H-N	8.022	117.035
G74H-N	9.035	111.19
T75H-N	8.469	116.158
Y76H-N	10.053	129.175
N77H-N	9.65	120.949
N77HD21-ND2	7.564	114.239
N77HD22-ND2	7.074	114.241
V78H-N	8.873	119.138
D79H-N	8.515	121.535
V80H-N	8.211	123.941
Y81H-N	9.501	125.147
S82H-N	9.286	119.747
T83H-N	8.862	116.344
N84HD21-ND2	7.803	111.901
N84HD22-ND2	7.073	111.88
G85H-N	8.414	108.237
T86H-N	7.894	116.83
R87H-N	9.133	128.124
I88H-N	8.864	126.992
L89H-N	7.461	120.334

N90H-N	8.395	122.167
N90HD22-ND2	6.968	112.333
K91H-N	8.745	124.055
E92H-N	8.433	121.985
L93H-N	9.578	125.71
R96H-N	9.107	129.135
I97H-N	8.548	120.342
L98H-N	8.998	130.214
E99H-N	8.265	129.544

Assignment of HSQC with 10 mM Ca^{2+}

Assignment	HN	HA	Peak
			intensity
G4H-N	8.712	111.325	2475637
T5H-N	8.244	118.58	1316606
V6H-N	9.187	128.591	410015
W7H-N	8.68	126.162	523927
G8H-N	9.039	109.069	861990
A9H-N	8.403	125.683	997117
L10H-N	8.844	122.215	311134
G11H-N	9.514	111.814	385247

H12H-N	8.074	119.03	596465
G13H-N	8.804	109.978	669372
I14H-N	8.828	122.167	312228
N15H-N	8.045	123.701	691520
N15HD21-ND2	7.562	110.921	460981
N15HD22-ND2	6.674	110.974	346120
N17H-N	8.271	119.576	1103132
N17HD21-ND2	8.36	112.309	469553
N20H-N	9.009	116.747	573012
N20HD21-ND2	7.772	113.3	936118
N20HD22-ND2	7.101	113.297	996013
F21H-N	7.745	118.349	798603
Q22H-N	7.218	126.79	2282441
Q22HE21-NE2	7.583	112.716	2605259
Q22HE22-NE2	6.916	112.747	2602019
M23H-N	8.573	124.172	651220
T24H-N	7.006	116.982	781738
D25H-N	8.847	120.714	745030
D26H-N	8.316	116.077	693661
I27H-N	7.728	120.491	447661
E29H-N	7.913	115.852	622636
V30H-N	9.233	125.34	383032

R31H-N	9.458	127.101	565889
W32H-N	9.481	124.671	364765
E33H-N	9.834	123.417	291104
R34H-N	8.872	123.244	242058
G35H-N	9.026	118.893	445673
S36H-N	7.61	114.864	1031586
L38H-N	8.819	130.162	822721
V39H-N	9.271	124.983	309244
A40H-N	8.131	119.746	620293
E41H-N	9.792	121.834	527137
F42H-N	9.587	128.496	524970
K43H-N	8.564	129.922	634673
M46H-N	8.118	118.576	952523
F49H-N	8.884	122.608	1783823
L50H-N	7.793	127.708	724445
S52H-N	7.83	108.898	976199
G53H-N	8.836	106.192	1290352
A54H-N	8.157	121.936	925127
F55H-N	7.381	115.108	1121838
E56H-N	8.987	119.416	225526
L58H-N	9.118	129.608	417695
N60H-N	7.782	112.433	860438

N60HD21-ND2	7.388	108.931	255322
N60HD22-ND2	6.699	108.936	263368
G61H-N	8.348	108.304	893416
D62H-N	7.922	119.782	995873
L63H-N	7.404	120.88	380254
K64H-N	9.183	127.999	305707
I65H-N	8.784	127.297	521642
K66H-N	7.999	125.33	1018893
N67H-N	8.689	117.007	961306
N67H-N	7.391	111.26	477391
N67HD22-ND2	6.715	111.221	544533
T69H-N	9.393	118.504	1113858
R70H-N	9.166	120.601	794402
D71H-N	8.022	117.035	951556
G74H-N	9.035	111.19	913443
T75H-N	8.469	116.158	718606
Y76H-N	10.053	129.175	508680
N77H-N	9.65	120.949	641608
N77HD21-ND2	7.564	114.239	617467
N77HD22-ND2	7.074	114.241	442100
V78H-N	8.873	119.138	658903
D79H-N	8.515	121.535	636460

V80H-N	8.211	123.941	1058324
Y81H-N	9.501	125.147	387636
S82H-N	9.286	119.747	507123
T83H-N	8.862	116.344	1421226
N84HD21-ND2	7.803	111.901	222519
N84HD22-ND2	7.073	111.88	618326
G85H-N	8.414	108.237	1091834
T86H-N	7.894	116.83	614913
R87H-N	9.133	128.124	1151942
I88H-N	8.864	126.992	316246
L89H-N	7.461	120.334	554681
N90H-N	8.395	122.167	610749
N90HD22-ND2	6.968	112.333	572175
K91H-N	8.745	124.055	492025
E92H-N	8.433	121.985	676148
L93H-N	9.578	125.71	991419
R96H-N	9.107	129.135	286060
I97H-N	8.548	120.342	804597
L98H-N	8.998	130.214	795748
E99H-N	8.265	129.544	1262501

Assignments of HSQC spectra with 10 μM Mn^{2+}

Assignment	w1	w2	Peak intensity
G4H-N	8.705	111.315	406434
T5H-N	8.246	118.646	510790
V6H-N	9.171	128.535	446708
W7H-N	8.68	126.162	507841
G8H-N	9.036	109.092	479516
A9H-N	8.384	125.651	196031
L10H-N	8.841	122.275	143053
G11H-N	9.527	111.864	314399
H12H-N	8.079	119.078	464203
G13H-N	8.809	110.053	318577
I14H-N	8.833	122.162	135090
N15H-N	8.04	123.705	304578
N15HD21-ND2	7.552	110.994	219322
N15HD22-ND2	6.667	111.035	296770
L16H-N	9.562	125.731	537287
N17H-N	8.27	119.625	222973
N17HD21-ND2	8.361	112.405	202842
N20H-N	9	116.735	415025
N20HD21-ND2	7.771	113.301	547388

N20HD22-ND2	7.099	113.345	533147
F21H-N	7.741	118.383	316585
Q22H-N	7.217	126.831	452783
Q22HE21-NE2	7.582	112.728	811152
Q22HE22-NE2	6.912	112.731	737656
M23H-N	8.574	124.217	693517
T24H-N	7.007	117	559180
D25H-N	8.838	120.74	555487
D26H-N	8.315	116.149	145655
E29H-N	7.912	115.889	199214
V30H-N	9.229	125.352	155102
R31H-N	9.453	127.121	237733
W32H-N	9.485	124.683	188706
G35H-N	9.03	118.943	468583
S36H-N	7.61	114.913	732884
L38H-N	8.805	130.175	244565
V39H-N	9.268	125.027	181453
A40H-N	8.127	119.763	555407
E41H-N	9.797	121.858	399149
F42H-N	9.576	128.51	365249
K43H-N	8.574	129.912	385119
M46H-N	8.12	118.648	595367

F49H-N	8.879	122.673	740506
L50H-N	7.791	127.719	342307
S52H-N	7.824	108.95	665970
G53H-N	8.839	106.179	253843
A54H-N	8.155	121.98	601153
F55H-N	7.363	115.154	378038
E56H-N	8.978	119.473	97940
L58H-N	9.128	129.63	337670
N60H-N	7.781	112.438	551525
N60H-ND2	7.389	108.947	136717
N60HD22-ND2	6.694	109.053	111211
G61H-N	8.348	108.363	343244
D62H-N	7.92	119.783	486946
L63H-N	7.405	120.964	298689
K64H-N	9.184	128.04	277129
I65H-N	8.784	127.279	317548
K66H-N	8	125.325	554778
N67H-N	8.689	117.006	621220
N67H-N	7.391	111.297	343975
N67HD22-ND2	6.709	111.285	313075
T69H-N	9.39	118.525	543576
R70H-N	9.16	120.613	499908

D71H-N	8.018	117.082	177591
G74H-N	9.03	111.186	445066
T75H-N	8.463	116.171	617312
Y76H-N	10.038	129.204	299238
N77H-N	9.659	121.006	75421
V78H-N	8.871	119.192	101756
D79H-N	8.518	121.578	405307
V80H-N	8.205	123.945	582899
Y81H-N	9.504	125.212	189715
S82H-N	9.29	119.777	138354
T83H-N	8.866	116.348	503822
N84HD21-ND2	7.805	111.888	240831
N84HD22-ND2	7.071	111.901	255220
G85H-N	8.402	108.249	128649
T86H-N	7.889	116.86	566622
R87H-N	9.13	128.119	722173
I88H-N	8.864	126.991	330295
L89H-N	7.451	120.329	430346
N90H-N	8.385	122.14	316172
N90HD22-ND2	6.967	112.39	125387
K91H-N	8.719	123.984	79571
E92H-N	8.427	122.076	172974

L93H-N	9.584	125.737	624878
D94H-N	8.866	123.582	158158
R96H-N	9.091	129.173	310432
I97H-N	8.559	120.347	418921
L98H-N	8.986	130.204	433720
E99H-N	8.259	129.616	707203

Assignments of HSQC spectra with 20 μM Mn^{2+}

Assignment	w1	w2	Peak intensity
G4H-N	8.708	111.359	269594
T5H-N	8.244	118.608	689344
V6H-N	9.194	128.575	364971
W7H-N	8.68	126.162	558724
G8H-N	9.034	109.084	465894
A9H-N	8.397	125.663	281020
L10H-N	8.848	122.173	71758
G11H-N	9.509	111.836	605575
H12H-N	8.073	119.061	406533
G13H-N	8.801	110.028	258239
I14H-N	8.835	122.213	57724
N15H-N	8.049	123.716	573668

N15HD21-ND2	7.559	110.945	262259
N17H-N	8.292	119.677	92872
N17HD21-ND2	8.367	112.357	245946
N20HD21-ND2	7.78	113.323	636577
N20HD22-ND2	7.091	113.315	342119
F21H-N	7.75	118.371	193989
Q22H-N	7.216	126.839	377400
Q22HE21-NE2	7.592	112.764	901969
Q22HE22-NE2	6.918	112.741	1382654
M23H-N	8.583	124.263	613812
T24H-N	7.009	116.989	568391
D25H-N	8.844	120.734	629182
D26H-N	8.309	116.091	69190
V30H-N	9.232	125.402	144136
R31H-N	9.462	127.055	247088
W32H-N	9.482	124.639	130578
G35H-N	9.027	118.91	818918
S36H-N	7.617	114.881	719433
L38H-N	8.828	130.211	631585
V39H-N	9.267	125.066	149658
A40H-N	8.126	119.689	331527
E41H-N	9.79	121.856	508314

F42H-N	9.589	128.48	407731
K43H-N	8.56	129.955	339060
M46H-N	8.118	118.6	673083
F49H-N	8.897	122.652	767195
L50H-N	7.791	127.716	608505
S52H-N	7.818	108.896	447864
G53H-N	8.831	106.208	465099
A54H-N	8.159	121.959	839408
F55H-N	7.374	115.128	607527
E56H-N	8.981	119.443	65646
L58H-N	9.131	129.646	368680
N60H-N	7.785	112.462	512509
N60HD21-ND2	7.392	109.024	96164
G61H-N	8.347	108.308	467309
D62H-N	7.931	119.795	477457
L63H-N	7.408	120.913	376913
K64H-N	9.179	127.97	204241
I65H-N	8.795	127.306	352254
K66H-N	8.015	125.346	682296
N67H-N	8.681	117.001	690007
N67H-N	7.402	111.283	360822
T69H-N	9.382	118.519	435256

R70H-N	9.169	120.615	469970
D71H-N	8.023	116.962	137013
G74H-N	9.029	111.179	596409
T75H-N	8.471	116.193	565966
Y76H-N	10.062	129.168	334811
D79H-N	8.51	121.606	117592
V80H-N	8.198	123.969	241860
Y81H-N	9.5	125.149	115195
T83H-N	8.865	116.358	577950
N84HD21-ND2	7.807	111.902	204614
N84HD22-ND2	7.068	111.886	220181
G85H-N	8.421	108.216	99386
T86H-N	7.891	116.87	364257
R87H-N	9.129	128.121	656346
I88H-N	8.86	127.021	568552
L89H-N	7.461	120.323	349086
N90H-N	8.405	122.142	212144
N90HD22-ND2	6.97	112.41	64211
E92H-N	8.448	122.127	104738
L93H-N	9.584	125.737	681956
R96H-N	9.102	129.201	285560
I97H-N	8.552	120.365	781165

L98H-N	8.997	130.241	570056
E99H-N	8.265	129.588	853050

DelPhi run file for CD2.6D79

gsize=165

perfil=20

Acenter(0,0,0)

in(pdb,file="6d79.pdb")

in(siz,file="default-cal.siz")

in(crg,file="default-cal.crg")

out(phi,file="6d79.phi")

indi=4

exdi=80

prbrad=1.4

salt=0.01

bndcon=2

maxc=0.0001

autoc=t

nonit=800

energy (s,c,g)

DelPhi run file for R87 calculation in 6D79

gsize=165

perfil=20

Acenter(0,0,0)

in(pdb,file="6d79.pdb")

in(siz,file="default-cal.siz")

in(crg,file="default-cal.crg")

in(frc,file="6d79-87.frc")

out(frc,file="6d79-87-out.frc")

indi=4

exdi=80

prbrad=1.4

salt=0.01

bndcon=2

maxc=0.0001

autoc=t

nonit=800

energy (s,c,g)

Output frc file for residue R87 in CD2.6D79

DELPHI SITE POTENTIAL FILE

grid size,percent fill: 65 20

0

outer diel. and first one assigned : 8

0

4

ionic strength (M): 9.9999998E-03

ion excl., probe radii: 2.000000

1.4

1.4

0

linear, nolinear iterations: 10

800

boundary condition: 2

Data Output: COORDINATES CHARGE

POTENTIAL

LS FIELDS

title: qdiffxas: qdiffxs4 with an improv

ed ing
surfac routine

ATOM COORDINATES (X,Y,Z)	CHARGE	GRID PT.	GRID FIEL	DS: (Ex, Ez)	Ey, Ez)
-9.2880 10.5830 -9.2630 0.0000		-0.7468	-0.121	-0.175	0.0869
-8.1810 10.3110 -8.3580 0.0000		-0.9032	-0.1956	-0.358	0.108
-8.4400 11.0620 -7.0510 0.0000		-0.9752	-0.1962	-0.417	0.0875
-8.4420 12.2910 -7.0390 0.0000		-0.8197	-0.1396	-0.27	0.0609
-6.8510 10.7410 -8.9910 0.0000		-0.5409	-0.2552	-0.252	0.0771
-5.6920 10.5220 -8.0060 0.0000		-0.0836	-0.1676	-0.545	0.1265
-4.3350 10.8850 -8.6220 0.0000		0.5408	-0.0066	-0.382	0.0724
-3.9930 10.0510 -9.7820 0.0000		0.7749	0.0654	-0.152	0.0842
-3.6630 8.7490 -9.7410 0.0000		0.4608	0.1784	-0.726	0.5513
-3.6500 8.0830 -8.5810 0.5000		-2.3079	0.2346	-1.803	0.8075
-3.3490 8.1130 -10.8740 0.5000		1.5271	0.2387	-0.48	0.6465
-8.1330 9.2380 -8.1730 0.0000		-1.6034	-0.042	-0.618	0.2098
-6.6890 10.1480 -9.8910 0.0000		-0.4439	-0.2307	-0.174	0.0991
-6.8950 11.7970 -9.2650 0.0000		-0.5101	-0.1858	-0.143	0.0503
-5.8260 11.1620 -7.1330 0.0000		-0.2769	-0.1655	-0.602	0.1461
-5.7020 9.4860 -7.6680 0.0000		-1.4709	0.007	-1.057	0.3087
-4.3660 11.9270 -8.9430 0.0000		0.2545	0.0107	-0.12	0.0257
-3.5580 10.7970 -7.8630 0.0000		0.8779	0.0592	-0.745	0.1642
-4.0020 10.5150 -10.6800 0.0000		0.8612	0.0525	0.2874	-0.143
-3.3800 7.1010 -8.5450 0.0000		-3.8589	0.3034	-2.385	1.1121
-3.8700 8.5490 -7.7140 0.0000		-3.1511	0.2035	-2.03	0.6727
-3.2250 7.1040 -10.8550 0.0000		1.6278	0.3085	-0.814	0.993

-3.3260	8.5980	-11.7580	0.0000	0.8814	0.1842	-0.269	0.2458
-9.1250	11.2070	-10.0390	0.0000	-0.5838	-0.1035	-0.072	0.075
total energy = -0.1951905				kt			

DelPhi run file for R87 calculation in CD2

gsize=165

perfil=20

Acenter(0,0,0)

in(pdb,file="cd2.pdb")

in(siz,file="default-cal.siz")

in(crg,file="default-cal.crg")

in(frc,file="cd2-87.frc")

out(frc,file="cd2-87-out.frc")

indi=4

exdi=80

prbrad=1.4

salt=0.01

bndcon=2

maxc=0.0001

autoc=t

nonit=800

energy (s,c,g)

Output frc file for residue R87 in CD2

DELPHI SITE POTENTIAL FILE

grid size,percent fill: 65 20

0

outer diel. and first one assigned : 8

0

2

ionic strength (M): 0.0000000E+00

ion excl., probe radii: 2.000000

1.4

1.4

0

linear, nolinear iterations: 340

30

boundary condition: 2

Data Output: COORDINATES CHARGE

POTENTIAL

LS FIELDS

title: qdiffxas: qdiffxs4 with an improv

ed ing
surfac routine

ATOM COORDINATES (X,Y,Z)	CHARGE	GRID PT.	GRID FIEL	DS: (Ex,	Ey, Ez)
-9.2500 10.5830 -9.2870 0.0000		-2.5052	-0.0163	-0.208	0.279
-8.1150 10.3280 -8.4120 0.0000		-3.2492	-0.0434	-0.594	0.425
-8.3480 11.0810 -7.1010 0.0000		-3.5766	-0.0177	-0.762	0.374
-8.3510 12.3100 -7.0920 0.0000		-3.1246	0.0136	-0.526	0.275
-6.8170 10.7760 -9.0970 0.0000		-2.6979	-0.1998	-0.482	0.446
-5.6040 10.5640 -8.1790 0.0000		-2.6112	-0.1396	-1.271	0.781
-4.2890 10.9520 -8.8720 0.0000		-1.5552	-0.0264	-0.89	0.697
-4.0200 10.1510 -10.0770 0.0000		-0.625	0.0454	0.0317	0.554
-3.6900 8.8480 -10.0820 0.0000		-0.4217	0.1088	-0.651	1.722
-3.5640 8.1820 -8.9310 0.5000		-6.9997	-0.0246	-2.869	2.386
-3.4980 8.2090 -11.2400 0.5000		1.1884	0.1873	-0.348	1.722
-8.0470 9.2560 -8.2230 0.0000		-4.4741	0.4295	-0.863	0.613
-6.6930 10.1920 -10.0090 0.0000		-2.1458	-0.2204	-0.201	0.477
-6.8850 11.8330 -9.3610 0.0000		-2.4573	-0.1323	-0.273	0.325
-5.7010 11.1920 -7.2920 0.0000		-3.1069	-0.107	-1.501	0.821
-5.5810 9.5240 -7.8530 0.0000		-5.6649	0.2212	-2.096	1.162
-4.3460 12.0020 -9.1610 0.0000		-1.7401	0.0231	-0.356	0.45
-3.4620 10.8490 -8.1680 0.0000		-1.4772	-0.0195	-1.794	1.105
-4.1010 10.6310 -10.9610 0.0000		-0.651	0.0561	0.5121	0.078
-3.3220 7.1920 -8.9070 0.0000		-9.6627	-0.0962	-3.582	2.983
-3.6970 8.6630 -8.0550 0.0000		-9.7059	-0.0878	-3.728	2.111
-3.3850 7.1990 -11.2290 0.0000		1.759	0.2415	-0.758	2.444

-3.5570 8.6950 -12.1230 0.0000
-9.1090 11.1920 -10.0790 0.0000
total energy = -1.452820 kt

-0.0248 0.1444 -0.149 0.753
-2.1311 -0.0454 -0.049 0.249

References

- Alattia, J.R., Kurokawa, H. and Ikura, M. (1999) Structural view of cadherin-mediated cell-cell adhesion. *Cell Mol Life Sci*, **55**, 359-367.
- Alexov, E.G. and Gunner, M.R. (1997) Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys J*, **72**, 2075-2093.
- Alexov, E.G. and Gunner, M.R. (1999) Calculated protein and proton motions coupled to electron transfer: electron transfer from QA- to QB in bacterial photosynthetic reaction centers. *Biochemistry*, **38**, 8253-8270.
- Ames, J.B., Ishima, R., Tanaka, T., Gordon, J.I., Stryer, L. and Ikura, M. (1997) Molecular mechanics of calcium-myristoyl switches. *Nature*, **389**, 198-202.
- Antosiewicz, J., McCammon, J.A. and Gilson, M.K. (1994) Prediction of pH-dependent Properties of Proteins. *J. Mol. Biol.*, **238**, 415-436.
- Antosiewicz, J., McCammon, J.A. and Gilson, M.K. (1996) The determinants of pKas in proteins. *Biochemistry*, **35**, 7819-7833.
- Arulanandam, A.R., Moingeon, P., Concino, M.F., Recny, M.A., Kato, K., Yagita, H., Koyasu, S. and Reinherz, E.L. (1993a) A soluble multimeric recombinant CD2 protein identifies CD48 as a low affinity ligand for human CD2: divergence of CD2 ligands during the evolution of humans and mice. *J Exp Med*, **177**, 1439-1450.
- Arulanandam, A.R., Withka, J.M., Wyss, D.F., Wagner, G., Kister, A., Pallai, P., Recny, M.A. and Reinherz, E.L. (1993b) The CD58 (LFA-3) binding site is a localized

- and highly charged surface area on the AGFCC'C" face of the human CD2 adhesion domain. *Proc Natl Acad Sci U S A*, **90**, 11613-11617.
- Baker, E.N. and Hubbard, R.E. (1984) Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol*, **44**, 97-179.
- Barlow, D.J. and Thornton, J.M. (1983) Ion-pairs in proteins. *J Mol Biol*, **168**, 867-885.
- Benson, D.E., Conrad, D.W., de Lorimier, R.M., Trammell, S.A. and Hellinga, H.W. (2001) Design of bioelectronic interfaces by exploiting hinge-bending motions in proteins. *Science*, **293**, 1641-1644.
- Binstead, R.A., Zuberbuhler, A.D. (1998) SPECFIT Global Analysis System. Spectrum Software Associates, Chapel Hill, NC 27515.
- Blattman, J.N. and Greenberg, P.D. (2004) Cancer immunotherapy: a treatment for the masses. *Science*, **305**, 200-205.
- Bodian, D.L., Jones, E.Y., Harlos, K., Stuart, D.I. and Davis, S.J. (1994) Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5 Å resolution. *Structure*, **2**, 755-766.
- Bolon, D.N. and Mayo, S.L. (2001) Enzyme-like proteins by computational design. *Proc Natl Acad Sci U S A*, **98**, 14274-14279.
- Bouckaert, J., Dewallef, Y., Poortmans, F., Wyns, L. and Loris, R. (2000) The structural features of concanavalin A governing non-proline peptide isomerization. *J Biol Chem*, **275**, 19778-19787.
- Brown, I.D. (1988) What Factors Determine Cation Coordination Numbers? *Acta Crystallographica*, **B44**, 545-553.

- Buckle, A.M., Schreiber, G. and Fersht, A.R. (1994) Protein-Protein Recognition: Crystal Structural Analysis of a Barnase-Barstar Complex at 2.0-Å Resolution. *Biochemistry*, **33**, 8878-8889.
- Cannon, C.P. (1995) Thrombin inhibitors in acute myocardial infarction. *Cardiol Clin*, **13**, 421-433.
- Chattopadhyaya, R., Meador, W.E., Means, A.R. and Quijcho, F.A. (1992) Calmodulin structure refined at 1.7 Å resolution. *J Mol Biol*, **228**, 1177-1192.
- Chazin, W.J. (1995) Releasing the calcium trigger. *Nature Structural Biology*, **2**, 707-710.
- Chen, H.A., Pfuhl, M. and Driscoll, P.C. (2002) The pH dependence of CD2 domain 1 self-association and ¹⁵N chemical exchange broadening is correlated with the anomalous pK_a of Glu41. *Biochemistry*, **41**, 14680-14688.
- Chen, H.A., Pfuhl, M., McAlister, M.S. and Driscoll, P.C. (2000) Determination of pK(a) values of carboxyl groups in the N-terminal domain of rat CD2: anomalous pK(a) of a glutamate on the ligand-binding surface. *Biochemistry*, **39**, 6814-6824.
- Chin, D. and Means, A.R. (2000) Calmodulin: a prototypical calcium sensor. *Trends Cell Biol*, **10**, 322-328.
- Chothia, C. and Jones, E.Y. (1997) The molecular structure of cell adhesion molecules. *Annu Rev Biochem*, **66**, 823-862.
- Chou, J.J., Li, S., Klee, C.B. and Bax, A. (2001) Solution structure of Ca(2+)-calmodulin reveals flexible hand-like properties of its domains. *Nat Struct Biol*, **8**, 990-997.

- Chrysina, E.D., Brew, K. and Acharya, K.R. (2000) Crystal structures of apo- and holo-bovine alpha-lactalbumin at 2. 2-A resolution reveal an effect of calcium on inter-lobe interactions. *J Biol Chem*, **275**, 37021-37029.
- Cootes, A.P., Curmi, P.M., Cunningham, R., Donnelly, C. and Torda, A.E. (1998) The dependence of amino acid pair correlations on structural environment. *Proteins*, **32**, 175-189.
- Corey, M.J. and Corey, E. (1996) On the failure of de novo-designed peptides as biocatalysts. *Proc Natl Acad Sci U S A*, **93**, 11428-11434.
- Da Silva, F.L., Jonsson, B. and Penfold, R. (2001) A critical investigation of the Tanford-Kirkwood scheme by means of Monte Carlo simulations. *Protein Sci*, **10**, 1415-1425.
- Davis, S.J., Davies, E.A., Tucknott, M.G., Jones, E.Y. and van der Merwe, P.A. (1998a) The role of charged residues mediating low affinity protein-protein recognition at the cell surface by CD2. *Proc Natl Acad Sci U S A*, **95**, 5490-5494.
- Davis, S.J., Ikemizu, S., Wild, M.K. and van der Merwe, P.A. (1998b) CD2 and the nature of protein interactions mediating cell-cell recognition. *Immunol Rev*, **163**, 217-236.
- De Angelis, D.A. (1999) Why FRET over genomics? *Physiol Genomics*, **1**, 93-99.
- Di Costanzo, L., Wade, H., Geremia, S., Randaccio, L., Pavone, V., DeGrado, W.F. and Lombardi, A. (2001) Toward the de novo design of a catalytically active helix bundle: a substrate-accessible carboxylate-bridged dinuclear metal center. *J Am Chem Soc*, **123**, 12749-12757.

- Drake, S.K., Zimmer, M.A., Kundrot, C. and Falke, J.J. (1997) Molecular tuning of an EF-hand-like calcium binding loop. Contributions of the coordinating side chain at loop position 3. *J Gen Physiol*, **110**, 173-184.
- Driscoll, P.C., Cyster, J.G., Campbell, I.D. and Williams, A.F. (1991) Structure of domain 1 of rat T lymphocyte CD2 antigen. *Nature*, **353**, 762-765.
- Eberhard, M. and Erne, P. (1994) Calcium and magnesium binding to rat parvalbumin. *Eur J Biochem*, **222**, 21-26.
- Einspahr, H., Bugg, C.E. (1984) Crystal Structure Studies of Calcium Complexes. In Sigel, H. (ed.), *Metal ions in biological systems*. Dekker, Basil, Vol. 17, pp. 51-71.
- English, A.C., Groom, C.R. and Hubbard, R.E. (2001) Experimental and computational mapping of the binding surface of a crystalline protein. *Protein Eng*, **14**, 47-59.
- Erickson, H.P. (1994) Reversible unfolding of fibronectin type III and immunoglobulin domains provides the structural basis for stretch and elasticity of titin and fibronectin. *Proc. Natl. Acad.*, **91**, 10114-10118.
- Falke, J.J., Drake, S.K., Hazard, A.L. and Peersen, O.B. (1994) Molecular tuning of ion binding to calcium signaling proteins. *Q Rev Biophys*, **27**, 219-290.
- Forster, T. (1959) Transfer mechanisms of electronic excitation. *Discussions of the Faraday Society*, **27**, 7-17.
- Frausto da Silva, J.J.R.a.W., R.J.P. (1991) *The Biological Chemistry of the Elements*. Oxford University Press, Toronto.
- Frei, D., Kystol, J. (2005) Inductively Coupled Plasma - Mass Spectrometry Laboratory.

- Fushiki, M., Svensson, B., Jonsson, B. and Woodward, C.E. (1991) Electrostatic interactions in protein solution--a comparison between Poisson-Boltzmann and Monte Carlo calculations. *Biopolymers*, **31**, 1149-1158.
- Garcia, J., Gerber, S.H., Sugita, S., Sudhof, T.C. and Rizo, J. (2004) A conformational switch in the Piccolo C2A domain regulated by alternative splicing. *Nat Struct Mol Biol*, **11**, 45-53.
- Gilboa, R., Spungin-Bialik, A., Wohlfahrt, G., Schomburg, D., Blumberg, S. and Shoham, G. (2001) Interactions of *Streptomyces griseus* aminopeptidase with amino acid reaction products and their implications toward a catalytic mechanism. *Proteins*, **44**, 490-504.
- Gill, S.C. and von Hippel, P.H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem*, **182**, 319-326.
- Gilson, M.K. and Honig, B. (1988a) Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins*, **4**, 7-18.
- Gilson, M.K. and Honig, B.H. (1986) The dielectric constant of a folded protein. *Biopolymers*, **25**, 2097-2119.
- Gilson, M.K. and Honig, B.H. (1988b) Energetics of charge-charge interactions in proteins. *Proteins*, **3**, 32-52.
- Golinelli, M.P., Chatelet, C., Duin, E.C., Johnson, M.K. and Meyer, J. (1998) Extensive ligand rearrangements around the [2Fe-2S] cluster of *Clostridium pasteurianum* ferredoxin. *Biochemistry*, **37**, 10429-10437.

- Gros, P., Kalk, K.H. and Hol, W.G. (1991) Calcium binding to thermitase.
Crystallographic studies of thermitase at 0, 5, and 100 mM calcium. *J Biol Chem*, **266**, 2953-2961.
- Gros, P., Fujinaga, M., Dijkstra, B.W., Kalk, K.H. and Hol, W.G. (1989)
Crystallographic refinement by incorporation of molecular dynamics:
thermostable serine protease thermitase complexed with eglin c. *Acta Crystallogr B*, **45 (Pt 5)**, 488-499.
- Guerois, R., Nielsen, J.E. and Serrano, L. (2002) Predicting changes in the stability of
proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol*, **320**, 369-387.
- Handford, P., Downing, A.K., Rao, Z., Hewett, D.R., Sykes, B.C. and Kielty, C.M.
(1995) The calcium binding properties and molecular organization of epidermal
growth factor-like domains in human fibrillin-1. *J Biol Chem*, **270**, 6751-6756.
- Harvey, S.C. (1989) Treatment of electrostatic effects in macromolecular modeling.
Proteins, **5**, 78-92.
- Haussinger, D., Ahrens, T., Aberle, T., Engel, J., Stetefeld, J. and Grzesiek, S. (2004)
Proteolytic E-cadherin activation followed by solution NMR and X-ray
crystallography. *Embo J*, **23**, 1699-1708.
- Hellinga, H.W. and Richards, F.M. (1991) Construction of new ligand binding sites in
proteins of known structure I. Computer-aided modeling of sites with pre-defined
geometry. *J. Mol. Biol.*, **222**, 763-785.

- Henzl, M.T., Hapak, R.C. and Goodpasture, E.A. (1996) Introduction of a fifth carboxylate ligand heightens the affinity of the oncomodulin CD and EF sites for Ca^{2+} . *Biochemistry*, **35**, 5856-5869.
- Herrenknecht, K. (1996) Cadherins. In Horton, M.A. (ed.), *Molecular Biology of Cell Adhesion Molecules*. John Wiley & Sons Ltd., pp. 45-69.
- Herzberg, O. and James, M.N. (1988) Refined crystal structure of troponin C from turkey skeletal muscle at 2.0 Å resolution. *J Mol Biol*, **203**, 761-779.
- Honig, B. and Nicholls, A. (1995) Classical electrostatics in biology and chemistry. *Science*, **268**, 1144-1149.
- Hopper, K.E. and McKenzie, H.A. (1974) Comparative studies of alpha-lactalbumin and lysozyme: Echidna lysozyme. *Mol Cell Biochem*, **3**, 93-108.
- Hornak, J.P. (1997-2006) Basics of NMR.
- Horrocks, W.D., Jr. (1993) Luminescence spectroscopy. In *Methods in Enzymology*. Academic Press Inc., Vol. 226.
- Hubbard, S.J., Campbell, S.F. and Thornton, J.M. (1991) Molecular recognition. Conformational analysis of limited proteolytic sites and serine proteinase protein inhibitors. *J Mol Biol*, **220**, 507-530.
- Hubbard, S.J. and Thornton, J.M. (1993) 'NACCESS' Computer Program, Department of Biochemistry and Molecular Biology, University College London.
- Ikura, M. (1996) Calcium binding and conformational response in EF-hand proteins. *Trends Biochem Sci*, **21**, 14-17.

- Ikura, M., Clore, G.M., Gronenborn, A.M., Zhu, G., Klee, C.B. and Bax, A. (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science*, **256**, 632-638.
- Isley, L.J. (2001) Protein Engineering and Metal-Binding Ability of de novo Designed Calcium-Binding Proteins. *Chemistry*. Georgia State University, Atlanta.
- Jones, E.Y., Davis, S.J., Williams, A.F., Harlos, K. and Stuart, D.I. (1992) Crystal structure at 2.8 Å resolution of a soluble form of the cell adhesion molecule CD2. *Nature*, **360**, 232-239.
- Jones, L.M., Yang, W., Wilkins, A.L., Kearney, A., Van der Merwe, P.A., and Yang, J.J. (2006) Rational design of a novel calcium binding site adjacent to the ligand binding site of CD2 increases its adhesion function. *Protein Eng*, **submitted**.
- Jones, S. and Thornton, J.M. (1996) Principles of protein-protein interactions. *Proc Natl Acad Sci U S A*, **93**, 13-20.
- Kapila, Y., Doan, D., Tafolla, E. and Fletterick, R. (2001) Three-dimensional structural analysis of fibronectin heparin-binding domain mutations. *J Cell Biochem*, **81**, 156-161.
- Karshikoff, A. and Ladenstein, R. (2001) Ion pairs and the thermotolerance of proteins from hyperthermophiles: a "traffic rule" for hot roads. *Trends Biochem Sci*, **26**, 550-556.
- Kawasaki, H., Nakayama, S. and Kretsinger, R.H. (1998) Classification and evolution of EF-hand proteins. *Biometals*, **11**, 277-295.

- Kelly, C.A., Nishiyama, M., Ohnishi, Y., Beppu, T. and Birktoft, J.J. (1993)
Determinants of protein thermostability observed in the 1.9-Å crystal structure of
malate dehydrogenase from the thermophilic bacterium *Thermus flavus*.
Biochemistry, **32**, 3913-3922.
- Kelly, S.M., Jess, T.J. and Price, N.C. (2005) How to study proteins by circular
dichroism. *Biochim Biophys Acta*, **1751**, 119-139.
- Kesvatera, T., Jonsson, B., Telling, A., Tougu, V., Vija, H., Thulin, E. and Linse, S.
(2001a) Calbindin D(9k): a protein optimized for calcium binding at neutral pH.
Biochemistry, **40**, 15334-15340.
- Kesvatera, T., Jonsson, B., Thulin, E. and Linse, S. (2001b) Focusing of the electrostatic
potential at EF-hands of calbindin D(9k): titration of acidic residues. *Proteins*, **45**,
129-135.
- Kettner, C. and Shaw, E. (1979) D-Phe-Pro-ArgCH₂C1-A selective affinity label for
thrombin. *Thromb Res*, **14**, 969-973.
- Kim, M., Sun, Z.Y., Byron, O., Campbell, G., Wagner, G., Wang, J. and Reinherz, E.L.
(2001) Molecular dissection of the CD2-CD58 counter-receptor interface
identifies CD2 Tyr86 and CD58 Lys34 residues as the functional "hot spot". *J*
Mol Biol, **312**, 711-720.
- King, G., Lee, F.S., Warshel, A. (1991) Microscopic simulations of macroscopic
dielectric constants of solvated proteins. *Journal of Chemical Physics*, **95**, 4366-
4377.

- Knapp, S., de Vos, W.M., Rice, D. and Ladenstein, R. (1997) Crystal structure of glutamate dehydrogenase from the hyperthermophilic eubacterium *Thermotoga maritima* at 3.0 Å resolution. *J Mol Biol*, **267**, 916-932.
- Kordel, J., Skelton, N.J., Akke, M. and Chazin, W.J. (1993) High-resolution structure of calcium-loaded calbindin D9k. *J Mol Biol*, **231**, 711-734.
- Kretsinger, R.H. and Nockolds, C.E. (1973) Carp muscle calcium-binding protein. II. Structure determination and general description. *J Biol Chem*, **248**, 3313-3326.
- Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C.B. and Bax, A. (1995) Solution structure of calcium-free calmodulin. *Nat Struct Biol*, **2**, 768-776.
- Kumar, S. and Nussinov, R. (1999) Salt bridge stability in monomeric proteins. *J Mol Biol*, **293**, 1241-1255.
- Kumar, S. and Nussinov, R. (2002a) Close-range electrostatic interactions in proteins. *Chembiochem*, **3**, 604-617.
- Kumar, S. and Nussinov, R. (2002b) Relationship between ion pair geometries and electrostatic strengths in proteins. *Biophys J*, **83**, 1595-1612.
- Kumar, S., Tsai, C.J. and Nussinov, R. (2000) Factors enhancing protein thermostability. *Protein Eng*, **13**, 179-191.
- Lawson, W.B., Valenty, V.B., Wos, J.D. and Lobo, A.P. (1982) Studies on the inhibition of human thrombin: effects of plasma and plasma constituents. *Folia Haematol Int Mag Klin Morphol Blutforsch*, **109**, 52-60.
- Lee, B., Godfrey, M., Vitale, E., Hori, H., Mattei, M.G., Sarfarazi, M., Tsipouras, P., Ramirez, F. and Hollister, D.W. (1991) Linkage of Marfan syndrome and a

- phenotypically related disorder to two different fibrillin genes. *Nature*, **352**, 330-334.
- Lee, B. and Richards, F.M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.*, **55**, 379-400.
- Lee, C.F., Makhatadze, G.I. and Wong, K.B. (2005) Effects of charge-to-alanine substitutions on the stability of ribosomal protein L30e from *Thermococcus celer*. *Biochemistry*, **44**, 16817-16825.
- Lee, K.K., Fitch, C.A. and Garcia-Moreno, E.B. (2002) Distance dependence and salt sensitivity of pairwise, coulombic interactions in a protein. *Protein Sci*, **11**, 1004-1016.
- Lee, L.P. and Tidor, B. (2001) Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein Sci*, **10**, 362-377.
- Leone, M., Di Lello, P., Ohlenschlager, O., Pedone, E.M., Bartolucci, S., Rossi, M., Di Blasio, B., Pedone, C., Saviano, M., Isernia, C. and Fattorusso, R. (2004) Solution structure and backbone dynamics of the K18G/R82E *Alicyclobacillus acidocaldarius* thioredoxin mutant: a molecular analysis of its reduced thermal stability. *Biochemistry*, **43**, 6043-6058.
- Leong, L.E. (1999) The use of recombinant fusion proteases in the affinity purification of recombinant proteins. *Mol Biotechnol*, **12**, 269-274.
- Leszczynski, J.F. and Rose, G.D. (1986) Loops in globular proteins: a novel category of secondary structure. *Science*, **234**, 849-855.

- Li, W.F., Zhou, X.X. and Lu, P. (2005) Structural features of thermozymes. *Biotechnol Adv*, **23**, 271-281.
- Linse, S., Brodin, P., Johansson, C., Thulin, E., Grundstrom, T. and Forsen, S. (1988) The role of protein surface charges in ion binding. *Nature*, **335**, 651-652.
- Linse, S. and Chazin, W.J. (1995) Quantitative measurements of the cooperativity in an EF-hand protein with sequential calcium binding. *Protein Sci*, **4**, 1038-1044.
- Linse, S. and Forsen, S. (1995) Determinants that govern high-affinity calcium binding. *Adv Second Messenger Phosphoprotein Res*, **30**, 89-151.
- Linse, S., Johansson, C., Brodin, P., Grundstrom, T., Drakenberg, T. and Forsen, S. (1991) Electrostatic contributions to the binding of calcium in calbindin D9k. *Biochemistry*, **30**, 154-162.
- Loladze, V.V., Ibarra-Molero, B., Sanchez-Ruiz, J.M. and Makhatadze, G.I. (1999) Engineering a thermostable protein via optimization of charge-charge interactions on the protein surface. *Biochemistry*, **38**, 16419-16423.
- Lorch, M., Mason, J.M., Clarke, A.R. and Parker, M.J. (1999) Effects of core mutations on the folding of a beta-sheet protein: implications for backbone organization in the I-state. *Biochemistry*, **38**, 1377-1385.
- Lu, Y., Berry, S.M. and Pfister, T.D. (2001) Engineering novel metalloproteins: design of metal-binding sites into native protein scaffolds. *Chem Rev*, **101**, 3047-3080.
- Malik, N.A., Anantharamaiah, G.M., Gawish, A. and Cheung, H.C. (1987) Structural and biological studies on synthetic peptide analogues of a low-affinity calcium-binding site of skeletal troponin C. *Biochim Biophys Acta*, **911**, 221-230.

- Maniccia, A.W., Yang, W., Li, S., Johnson, J.A., and Yang, J.J. (2006) Using Protein Design to Dissect the Effect of Chare Residues on Metal Binding and Protein Stability. *Biochemistry*, **in Press**.
- Markwardt, F., Walsmann, P., Sturzebecher, J., Landmann, H. and Wagner, G. (1973) [Synthetic serin protease inhibitors. 1. Inhibition of trypsin, plasmin and thrombin by amidino- and guanidinobenzoic acid esters]. *Pharmazie*, **28**, 326-330.
- Marsden, B.J., Hodges, R.S. and Sykes, B.D. (1988) ¹H-NMR studies of synthetic peptide analogues of calcium-binding site III of rabbit skeletal troponin C: effect on the lanthanum affinity of the interchange of aspartic acid and asparagine residues at the metal ion coordinating positions. *Biochemistry*, **27**, 4198-4206.
- Marsden, B.J., Shaw, G.S. and Sykes, B.D. (1990) Calcium binding proteins. Elucidating the contributions to calcium affinity from an analysis of species variants and peptide fragments. *Biochem Cell Biol*, **68**, 587-601.
- McAlister, M.S., Mott, H.R., van der Merwe, P.A., Campbell, I.D., Davis, S.J. and Driscoll, P.C. (1996) NMR analysis of interacting soluble forms of the cell-cell recognition molecules CD2 and CD48. *Biochemistry*, **35**, 5982-5991.
- McKenzie, H.A. and White, F.H., Jr. (1987) Studies on a trace cell lytic activity associated with alpha-lactalbumin. *Biochem Int*, **14**, 347-356.
- McPhalen, C.A., Strynadka, N.C. and James, M.N. (1991) Calcium-binding sites in proteins: a structural perspective. *Adv Protein Chem*, **42**, 77-144.
- Mercier, P., Ferguson, R.E., Irving, M., Corrie, J.E., Trentham, D.R. and Sykes, B.D. (2003) NMR structure of a bifunctional rhodamine labeled N-domain of troponin

- C complexed with the regulatory "switch" peptide from troponin I: implications for in situ fluorescence studies in muscle fibers. *Biochemistry*, **42**, 4333-4348.
- Merkel, J.S., Sturtevant, J.M. and Regan, L. (1999) Side chain interactions in parallel beta sheets: the energetics of cross- strand pairings. *Structure Fold Des*, **7**, 1333-1343.
- Moncrief, N.D., Kretsinger, R.H. and Goodman, M. (1990) Evolution of EF-hand calcium-modulated proteins. I. Relationships based on amino acid sequences. *J Mol Evol*, **30**, 522-562.
- Musafia, B., Buchner, V. and Arad, D. (1995) Complex salt bridges in proteins: statistical analysis of structure and function. *J Mol Biol*, **254**, 761-770.
- Nagar, B., Overduin, M., Ikura, M. and Rini, J.M. (1996) Structural Basis of Calcium-Induced E-Cadherin Rigidification and Dimerization. *Nature*, **380**, 360-365.
- Nakamura, H. (1996) Roles of electrostatic interaction in proteins. *Q Rev Biophys*, **29**, 1-90.
- Nakayama, S., Moncrief, N.D. and Kretsinger, R.H. (1992) Evolution of EF-hand calcium-modulated proteins. II. Domains of several subfamilies have diverse evolutionary histories. *J Mol Evol*, **34**, 416-448.
- Nash, P.D., Opas, M. and Michalak, M. (1994) Calreticulin: not just another calcium-binding protein. *Mol Cell Biochem*, **135**, 71-78.
- Nelson, M.R. and Chazin, W.J. (1998) An interaction-based analysis of calcium-induced conformational changes in Ca²⁺ sensor proteins. *Protein Sci*, **7**, 270-282.

- Nemeth, E.F., Steffey, M.E., Hammerland, L.G., Hung, B.C.P., Van Wagenen, B.C., DelMar, E.G. and Balandrin, M.F. (1998) Calcimimetics with potent and selective activity on the parathyroid calcium receptor. *PNAS*, **95**, 4040-4045.
- Nicholls A and Honig, B. (1991) A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J Comp Chem*, **12**, 435-445.
- Nicholls, A., Sharp, K.A. and Honig, B. (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct Funct Gen*, **11**, 282.
- Niki, I., Yokokura, H., Sudo, T., Kato, M. and Hidaka, H. (1996) Ca²⁺ signaling and intracellular Ca²⁺ binding proteins. *J Biochem (Tokyo)*, **120**, 685-698.
- Nitta, K. and Sugai, S. (1989) The evolution of lysozyme and alpha-lactalbumin. *Eur J Biochem*, **182**, 111-118.
- Nooren, I.M. and Thornton, J.M. (2003) Structural characterisation and functional significance of transient protein-protein interactions. *J Mol Biol*, **325**, 991-1018.
- Parker, M.J. and Clarke, A.R. (1997a) Amide backbone and water-related H/D isotope effects on the dynamics of. *Biochemistry*, **36**, 5786-5794.
- Parker, M.J. and Clarke, A.R. (1997b) Amide backbone and water-related H/D isotope effects on the dynamics of a protein folding reaction. *Biochemistry*, **36**, 5786-5794.
- Permyakov, S.E., Makhatadze, G.I., Owenius, R., Uversky, V.N., Brooks, C.L., Permyakov, E.A. and Berliner, L.J. (2005) How to improve nature: study of the

- electrostatic properties of the surface of alpha-lactalbumin. *Protein Eng Des Sel*, **18**, 425-433.
- Perutz, M. (1994) Polar zippers: Their role in human disease. *Protein Science*, **3**, 1629-1637.
- Pidcock, E. and Moore, G.R. (2001) Structural characteristics of protein binding sites for calcium and lanthanide ions. *J Biol Inorg Chem*, **6**, 479-489.
- Pinto, A.L., Hellinga, H.W. and Caradonna, J.P. (1997) Construction of a catalytically active iron superoxide dismutase by rational protein design. *Proc Natl Acad Sci U S A*, **94**, 5562-5567.
- Ploegman, J.H., Drent, G., Kalk, K.H. and Hol, W.G. (1979) The structure of bovine liver rhodanese. II. The active site in the sulfur-substituted and the sulfur-free enzyme. *J Mol Biol*, **127**, 149-162.
- Powers, J.C., Harper, J.W. (1986). Elsevier, Amsterdam.
- Redfield, C. (1998) Resonance assignments in proteins. University of Oxford.
- Reid, R.E. (1990) Synthetic fragments of calmodulin calcium-binding site III. A test of the acid pair hypothesis. *J Biol Chem*, **265**, 5971-5976.
- Reid, R.E., Clare, D.M. and Hodges, R.S. (1980) Synthetic analog of a high affinity calcium binding site in rabbit skeletal troponin C. *J Biol Chem*, **255**, 3642-3646.
- Reid, R.E. and Hodges, R.S. (1980) Co-operativity and calcium/magnesium binding to troponin C and muscle calcium binding parvalbumin: an hypothesis. *J Theor Biol*, **84**, 401-444.

- Richards, F.M. (1977) Areas, Volumes, Packing, and Protein Structure. *Ann Rev Biophys Bioeng*, **6**, 151-176.
- Richardson, J.S. (1981) The anatomy and taxonomy of protein structure. *Adv Protein Chem*, **34**, 167-339.
- Rivas, G.A. and Gonzalez-Rodriguez, J. (1991) Calcium binding to human platelet integrin GPIIb/IIIa and to its constituent glycoproteins. Effects of lipids and temperature. *Biochem J*, **276 (Pt 1)**, 35-40.
- Rocchia, W., Alexov, E. and Honig, B. (2001) Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *J Phys Chem B*, **105**, 6507-6514.
- Sakai, L.Y., Keene, D.R. and Engvall, E. (1986) Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. *J Cell Biol*, **103**, 2499-2509.
- Sali, D., Bycroft, M. and Fersht, A.R. (1991) Surface Electrostatic Interactions Contribute Little to Stability of Barnase. *J. Mol. Biol.*, **220**, 779-788.
- Scaloni, A., Jones, W.M., Barra, D., Pospischil, M., Sassa, S., Popowicz, A., Manning, L.R., Schneewind, O. and Manning, J.M. (1992) Acylpeptide hydrolase: inhibitors and some active site residues of the human enzyme. *J Biol Chem*, **267**, 3811-3818.
- Schafer, B.W. and Heizmann, C.W. (1996) The S100 family of EF-hand calcium-binding proteins: functions and pathology. *Trends in Biochemical Sciences*, **21**, 134-140.
- Schueler-Furman, O., Wang, C., Bradley, P., Misura, K. and Baker, D. (2005) Progress in modeling of protein structures and interactions. *Science*, **310**, 638-642.

- Schwalbe, H., Grimshaw, S.B., Spencer, A., Buck, M., Boyd, J., Dobson, C.M., Redfield, C. and Smith, L.J. (2001) A refined solution structure of hen lysozyme determined using residual dipolar coupling data. *Protein Sci*, **10**, 677-688.
- Schymkowitz, J.W., Rousseau, F., Martins, I.C., Ferkinghoff-Borg, J., Stricher, F. and Serrano, L. (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. *Proc Natl Acad Sci U S A*, **102**, 10147-10152.
- Scott, D.L., White, S.P., Otwinowski, Z., Yuan, W., Gelb, M.H. and Sigler, P.B. (1990) Interfacial catalysis: the mechanism of phospholipase A2. *Science*, **250**, 1541-1546.
- Selander-Sunnerhagen, M., Ullner, M., Persson, E., Teleman, O., Stenflo, J. and Drakenberg, T. (1992) How an epidermal growth factor (EGF)-like domain binds calcium. High resolution NMR structure of the calcium form of the NH2-terminal EGF-like domain in coagulation factor X. *J Biol Chem*, **267**, 19642-19649.
- Selvaraj, P., Plunkett, M.L., Dustin, M., Sanders, M.E., Shaw, S. and Springer, T.A. (1987) The T lymphocyte glycoprotein CD2 binds the cell surface ligand LFA-3. *Nature*, **326**, 400-403.
- Selvin, P.R. (2000) The renaissance of fluorescence resonance energy transfer. *Nat Struct Biol*, **7**, 730-734.
- Sharp, K.A. and Honig, B. (1990) Electrostatic interactions in macromolecules: theory and applications. *Annu Rev Biophys Biophys Chem*, **19**, 301-332.

- She, M., Xing, J., Dong, W.J., Umeda, P.K. and Cheung, H.C. (1998) Calcium binding to the regulatory domain of skeletal muscle troponin C induces a highly constrained open conformation. *J Mol Biol*, **281**, 445-452.
- Sinha, N. and Smith-Gill, S.J. (2002) Electrostatics in protein binding and function. *Curr Protein Pept Sci*, **3**, 601-614.
- Skelton, N.J., Forsen, S. and Chazin, W.J. (1990) ¹H NMR resonance assignments, secondary structure, and global fold of Apo bovine calbindin D9k. *Biochemistry*, **29**, 5752-5761.
- Skoog, D.A., West, D.M., Holler, F.J. (1996) *Fundamentals of Analytical Chemistry*. Saunders College Publishing, Philadelphia.
- Snyder, E.E., Buoscio, B.W. and Falke, J.J. (1990) Calcium(II) site specificity: effect of size and charge on metal ion binding to an EF-hand-like site. *Biochemistry*, **29**, 3937-3943.
- Soderling, T.R. and Stull, J.T. (2001) Structure and regulation of calcium/calmodulin-dependent protein kinases. *Chem Rev*, **101**, 2341-2352.
- Spasov, V.Z., Karshikoff, A.D. and Ladenstein, R. (1995) The optimization of protein-solvent interactions: thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci*, **4**, 1516-1527.
- Spector, S., Wang, M., Carp, S.A., Robblee, J., Hendsch, Z.S., Fairman, R., Tidor, B. and Raleigh, D.P. (2000) Rational modification of protein stability by the mutation of charged surface residues. *Biochemistry*, **39**, 872-879.

- Stabler, S.M., Ostrowski, L.L., Janicki, S.M. and Monteiro, M.J. (1999) A Myristoylated Calcium-binding Protein that Preferentially Interacts with the Alzheimer's Disease Presenilin 2 Protein. *J. Cell Biol.*, **145**, 1277-1292.
- Strickler, S.S., Gribenko, A.V., Gribenko, A.V., Keiffer, T.R., Tomlinson, J., Reihle, T., Loladze, V.V. and Makhatadze, G.I. (2006) Protein stability and surface electrostatics: a charged relationship. *Biochemistry*, **45**, 2761-2766.
- Stryer, L. (1995) *Biochemistry*. W.H. Freeman and Company, NY.
- Strynadka, N.C. and James, M.N. (1989) Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu Rev Biochem*, **58**, 951-998.
- Strynadka, N.C.J., James, Michael N. G. (1993) Calcium-Binding Proteins. In King, R.B. (ed.), *Encyclopedia of Inorganic Chemistry*. John Wiley and Sons Ltd., pp. 477-507.
- Stuart, D.I., Acharya, K.R., Walker, N.P., Smith, S.G., Lewis, M. and Phillips, D.C. (1986) Alpha-lactalbumin possesses a novel calcium binding loop. *Nature*, **324**, 84-87.
- Sunnerhagen, M., Forsen, S., Hoffren, A.-M., Drakenberg, T., Teleman, O. and Stenflo, J. (1995) Structure of the Ca^{2+} -free GLA domain sheds light on membrane binding of blood coagulation proteins. *Nature Structural Biology*, **2**, 504-509.
- Tanaka, T., Ames, J.B., Harvey, T.S., Stryer, L. and Ikura, M. (1995) Sequestration of the membrane-targeting myristoyl group of recoverin in the calcium-free state. *Nature*, **376**, 444-447.

- Teplyakov, A.V., Kuranova, I.P., Harutyunyan, E.H., Vainshtein, B.K., Frommel, C., Hohne, W.E. and Wilson, K.S. (1990) Crystal structure of thermitase at 1.4 Å resolution. *J Mol Biol*, **214**, 261-279.
- Ubach, J., Zhang, X., Shao, X., Sudhof, T.C. and Rizo, J. (1998) Ca²⁺ binding to synaptotagmin: how many Ca²⁺ ions bind to the tip of a C2-domain? *Embo J*, **17**, 3921-3930.
- van der Merwe, P.A. and Barclay, A.N. (1994) Transient intercellular adhesion: the importance of weak protein- protein interactions. *Trends Biochem Sci*, **19**, 354-358.
- van der Merwe, P.A., McNamee, P.N., Davies, E.A., Barclay, A.N. and Davis, S.J. (1995) Topology of the CD2-CD48 cell-adhesion molecule complex: implications for antigen recognition by T cells. *Curr Biol*, **5**, 74-84.
- van Holde, K.E., Johnson, W.C., and Ho, P.S. (1998) *Principles of Physical Biochemistry*. Prentice Hall, Upper Saddle River.
- Venjaminov, S.Y., Klimtchuk, E.S., Bajzer, Z. and Craig, T.A. (2004) Changes in structure and stability of calbindin-D(28K) upon calcium binding. *Anal Biochem*, **334**, 97-105.
- Vieille, C., Epting, K.L., Kelly, R.M. and Zeikus, J.G. (2001) Bivalent cations and amino-acid composition contribute to the thermostability of *Bacillus licheniformis* xylose isomerase. *Eur J Biochem*, **268**, 6291-6301.
- Vogel, H.J., Brokx, R.D., and Ouyang, H. (2002) *Calcium-Binding Proteins*. Humana Press Inc., Totowa.

- Vyas, M.N., Jacobson, B.L. and Quioco, F.A. (1989) The calcium-binding site in the galactose chemoreceptor protein. Crystallographic and metal-binding studies. *J Biol Chem*, **264**, 20817-20821.
- Vyas, N.K., Vyas, M.N., and Quioco, F.A. (1994) In Heizmann, C. (ed.), *Novel Calcium Binding Proteins*. Springer Verlag, Berlin, pp. 403-423.
- Vyas, N.K., Vyas, M.N. and Quioco, F.A. (1987) A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. *Nature*, **327**, 635-638.
- Wang, Q., Buckle, A.M., Foster, N.W., Johnson, C.M. and Fersht, A.R. (1999) Design of highly stable functional GroEL minichaperones. *Protein Sci*, **8**, 2186-2193.
- Weis, W.I., Drickamer, K. and Hendrickson, W.A. (1992) Structure of a C-type mannose-binding protein complexed with an oligosaccharide. *Nature*, **360**, 127-134.
- Wilkins, A.L., Ye, Y., Yang, W., Lee, H.W., Liu, Z.R. and Yang, J.J. (2002) Metal-binding studies for a de novo designed calcium-binding protein. *Protein Eng*, **15**, 571-574.
- Williams, A.F., Barclay, A.N., Clark, S.J., Paterson, D.J. and Willis, A.C. (1987) Similarities in sequences and cellular expression between rat CD2 and CD4 antigens. *J Exp Med*, **165**, 368-380.
- Williams, R.J.P. (2002) Calcium. In Vogel, H.J. (ed.), *Calcium-Binding Protein Protocols, Vol 1: Reviews and Case Studies*. Humana Press Inc., Totowa.

- Wisz, M.S., Garrett, C.Z. and Hellinga, H.W. (1998) Construction of a family of Cys2His2 zinc binding sites in the hydrophobic core of thioredoxin by structure-based design. *Biochemistry*, **37**, 8269-8277.
- Wisz, M.S. and Hellinga, H.W. (2003) An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. *Proteins*, **51**, 360-377.
- Woody, R.W. (1985) Circular Dichroism of Peptides. *The Peptides*, **7**, 15-114.
- Wuthrich, K. (1986) *NMR of Proteins and Nucleic acids*. John Wiley & Sons, New York.
- Xiao, L. and Honig, B. (1999) Electrostatic contributions to the stability of hyperthermophilic proteins. *J Mol Biol*, **289**, 1435-1444.
- Yang, A.S., Gunner, M.R., Sampogna, R., Sharp, K. and Honig, B. (1993) On the calculation of pKas in proteins. *Proteins*, **15**, 252-265.
- Yang, J.J., Carroll, A.R., Yang, W., Ye, Y.M. and Nguyen, C. (2000a) An alternatively folded state of an acid-stable beta-sheet protein. *Cell Biochemistry and Biophysics*, **33**, 253-273.
- Yang, J.J., Carroll, A.R., Yang, W., Ye, Y. and Nguyen, C.N. (2000b) Nonnative intermediate state of acid-stable beta-sheet protein. *Cell Biochem Biophys*, **33**, 253-273.
- Yang, J.J., W. Yang. (2005) *Calcium-Binding Proteins*. John Wiley & Sons, New York.
- Yang, J.J., Yang, H., Ye, Y., Hopkins, H., Jr. and Hastings, G. (2002a) Temperature-induced formation of a non-native intermediate state of the all beta-sheet protein CD2. *Cell Biochem Biophys*, **36**, 1-18.

- Yang, J.J., Ye, Y., Carroll, A., Yang, W. and Lee, H. (2001a) Structural biology of the cell adhesion protein CD2. *Current Protein and Peptide Science*, **2**, 1-17.
- Yang, W. (2001) Rational Design of Calcium-binding Proteins. *Chemistry*. Georgia State University, Atlanta.
- Yang, W., Jones, L.M., Isley, L., Ye, Y., Lee, H.W., Wilkins, A., Liu, Z., Hellinga, H.W., Malcow, R., Ghazi, M. and Yang, J.J. (2003) Rational design of a calcium-binding protein. *JACS*, **125**, 6165-6171.
- Yang, W., Lee, H., Liu, Z., Hellinga, H.W. and Yang, J.J. (2001b) *Criteria for Designing a Calcium Binding Protein*. Kluwer, Norwell.
- Yang, W., Lee, H.W., Hellinga, H. and Yang, J.J. (2002b) Structural analysis, identification, and design of calcium-binding sites in proteins. *Proteins*, **47**, 344-356.
- Yang, W., Lee, H.W., Pu, M., Hellinga, H. and Yang, J.J. (2000c) Identifying and designing of calcium binding sites in proteins by computational algorithm. In Dadmun, M.D., Hook, W.A.V., Noid, D.W., Melnichenko, Y.B. and Sumpter, R.G. (eds.), *Computational Studies, Nanotechnology, and Solution Thermodynamics of Polymer Systems*. Kluwer Academic/Plenum Publishers, New York, pp. 127-138.
- Yang, W., Tsai, T., Kats, M. and Yang, J.J. (2000d) Peptide analogs from E-cadherin with different calcium-binding affinities. *J Pept Res*, **55**, 203-215.
- Yang, W., Wilkins, A.L., Ye, Y., Liu, Z.R., Li, S.Y., Urbauer, J.L., Hellinga, H.W., Kearney, A., van der Merwe, P.A. and Yang, J.J. (2005) Design of a Calcium-

Binding Protein with Desired Structure in a Cell Adhesion Molecule. *J Am Chem Soc*, **127**, 2085-2093.