

Georgia State University

ScholarWorks @ Georgia State University

---

Chemistry Dissertations

Department of Chemistry

---

8-4-2008

## Determining The Site Specific Metal Binding and Structural Properties of EF-Hand Protein Using Grafting Approach

Hsiau-Wei Lee

Follow this and additional works at: [https://scholarworks.gsu.edu/chemistry\\_diss](https://scholarworks.gsu.edu/chemistry_diss)

 Part of the [Chemistry Commons](#)

---

### Recommended Citation

Lee, Hsiau-Wei, "Determining The Site Specific Metal Binding and Structural Properties of EF-Hand Protein Using Grafting Approach." Dissertation, Georgia State University, 2008.  
doi: <https://doi.org/10.57709/1059269>

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](mailto:scholarworks@gsu.edu).

# DETERMINING THE SITE SPECIFIC METAL BINDING AND STRUCTURAL PROPERTIES OF EF-HAND PROTEIN USING GRAFTING APPROACH

by

HSIAU-WEI LEE

Under the Direction of Jenny J. Yang

## ABSTRACT

Calmodulin is an essential EF-hand protein with a helix-loop-helix calcium binding motif. Understanding Ca(II) dependent activation of calmodulin and other EF-hand proteins is limited by Ca(II)-induced conformational change, multiple and cooperative binding of Ca(II) ions, and interactions between the paired EF-hand motifs. The goal of this research project is to probe key determinants for calcium binding properties and pairing interactions at the site specific level using a grafting approach and high resolution NMR. An individual Ca(II) binding site of the EF-hand motifs of calmodulin was grafted into a non-calcium dependent protein, CD2, to bypass limitations associated with natural EF-hand proteins and peptide fragments. Using high resolution NMR, we have shown that the grafted EF-loop III of calmodulin in the host protein retains its native conformation with a strong loop and  $\beta$ -conformation preference. Grafted ligand residues in the engineered protein are directly involved in binding of Ca(II) and La(III). The NMR studies support our hypothesis that both ligand arrangement and dynamic properties play essential role in tuning Ca(II) binding affinities. Using pulse-field diffusion NMR and protein engineering, we further demonstrated that grafted EF-

loop remains as a monomer. Although the EF-loop with flanking helices dimerizes in the presence of Ca(II). Additionally, removal of conserved hydrophobic residues at the flanking helices of the EF-hand motif leads to be monomer in the absence and presence of metal ions. Our results suggest that conserved hydrophobic residues are essential for the pair-paired interaction in the coupled EF-hand protein. We have shown that our developed grafting approach can be applied to probe intrinsic Ca(II) binding affinities of different Ca(II) binding sites.

INDEX WORDS: Calcium, EF-hand, Calmodulin, CD2, NMR, Diffusion

DETERMINING THE SITE SPECIFIC METAL BINDING AND STRUCTURAL  
PROPERTIES OF EF-HAND PROTEIN USING GRAFTING APPROACH

by

HSIAU-WEI LEE

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

Doctor of Philosophy

In the College of Arts and Sciences

Georgia State University

2007



Copyright by  
Hsiau-Wei Lee  
2007

DETERMINING THE SITE SPECIFIC METAL BINDING AND STRUCTURAL  
PROPERTIES OF EF-HAND PROTEIN USING GRAFTING APPROACH

by

HSIAU-WEI LEE

Major Professor: Jenny J Yang  
Committee: David W Wilson  
James H Prestegard  
Jeffrey L Urbauer

Electronic Version Approved:

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

## DEDICATION

I am very grateful to have great family members and friends to support me during my Ph.D. studies. First, I would like to dedicate this dissertation to my father and mother, General Rong-Chang Lee and Ching-Huang Chen, for all of their supports and encouragements. I would also like to dedicate this work to my wife Lu Yin for the company, proof reading my dissertation, encouragements and help me becoming a better person. I would also like to thank my brother and his wife (Saul and Melody); you guys are the best sibling I could ever ask for. I dedicate this dissertation Dr. Anna Manicca for the last minute dissertation proof, all of the fights we had in the lab, all of your suggestion for my research. To John Manicca, thanks for listening to our boring science related conversation. To Dr. April Ellis, thank you for all of your advice in work and life. I would also like to thank Dr. Yang for making me as part of your family member and making sure that I would have a bright future.

I would also like to thank my host family: Ms Carter, Tommie, Stephen, Penny Clare, Courtney, Brad, Noah, Dr and Ms Counts, Mr and Mrs Gjerdes; the Lenox Oaks Buckhood members: Andy, Sallie, Zack, Jen, Betsy, Donna, Susan, Bill, Ben, Timmy, Steve; the kids: Abhay, Emil, Haval, Dougious, Misha, Mandy, Katherine, Bhavi, Keta, Mansi, Jen, Kimmie for all of the supports and encouragements.

## ACKNOWLEDGMENTS

All of the work in this dissertation is carried out under the direction of Dr. Jenny J Yang. I want to thank Dr. Yang helping me and guiding to reach my dream (to complete a Ph.D. degree). I would also like to thank Dr. Yang for giving me this great opportunity to learn and develop important biological work on calcium. I want to thank Dr. Wei Yang for teaching everything about Dezymer, NMR, structural calculation, and answering all of my questions. I want to thank Dr. Yiming Ye for engineering proteins and providing me with a lot of proteins and supports in the grafting projects. I want to thank Yubin Zhou and Nancy Huang for been great friends and all of your scientific supports including protein purification and labeling of proteins. I want to thank Dr. Shunyi Li for teaching the special tricks in protein purifications. I want to thank Johnny Chen for helping me preparing excellent protein sample. I would like thank Michael Kirberger for his support in the computational project and dissertation proof. I want to thank Lisa M Jones, Angela Holder, Dr. Jin Zou, Ning Chen, Jing Juan Qiao, Shen Tang, Jiang Jie, Julian Johnson, David T Mpofu, Adriana Castiblanco for helping me in the lab. I want to thank Dr. Don Harden for all of the computer help and knowledge. I want to thank Dr John Glushka, Dr. Fang Tian, Dr. Sonal Bansal, Dr. Catherine Bougault, Dr. Lianmei Feng, Dr. Thomas Weldegiorghis, Dr. Anita Kishore from UGA for helpful discussions and consultation on NMR. We thank Dr. Gary Pielak and his student at UNC for the analytical ultracentrifugation study.

Many thanks to my committee members, Drs. David D Wilson, James H Prestegard, Jeffrey L Urbauer for providing important guidance for my research work.

I would like to thank the Department of Chemistry and MBD pre-doctoral fellowship for their support. This work is also supported by funding from NIH and NSF. We also thank the support of the state NMR facility(800 and 900 MHz) hosted at UGA.

## Table of Contents

	Page
Dedication	iv
Acknowledge	v
List of Tables	xvii
List of Figures	xix
List of Abbreviations	xxvi
Chapter 1     Introduction	1
1.1     The Role of Calcium in Biological System	1
1.2     Structural Properties of Ca(II) Binding Proteins	2
1.2.1   Property of the EF-hand Motif	3
1.2.2   Pairing of the EF-hand Motifs and Cooperative Binding	4
1.3     Structural Studies of CaM	5
1.4     Understanding the Site Specific Metal Binding Properties of CaM	8
1.5     Our Research Approach	11
1.5.1   Choice of the Host Protein	12
1.6     Motivations and Overview of These Studies	13
Chapter 2     Methods and Material	25
2.1     Protein Engineering and Purification	25
2.1.1   Protein Expression	25
2.1.2   Protein Purification	26

2.1.3	Protein Stability	29
2.2	Metal Titration	30
2.2.1	1D $^1\text{H}$ NMR Titration	30
2.2.2	2D $^1\text{H}$ - $^{15}\text{N}$ HSQC Titration	31
2.2.3	Calculation of $K_d$	32
2.3	Sequential Assignment	33
2.3.1	2D NMR Experiments with Homonuclear Samples	33
2.3.2	Aromatic Ring Assignment in $\text{D}_2\text{O}$ Condition	33
2.3.3	3D TOCSYHSQC and NOESYHSQC	35
2.3.4	3D Triple Resonance Experiments	36
2.4	Backbone Dihedral Angles	37
2.4.1	$J_{\text{HNHA}}$ Coupling Constant	37
2.4.2.1	Dihedral Angle Prediction Using TALOS	38
2.4.2.2	Predicting Dihedral Angles for Wild Type CD2	38
2.5	Residual Dipolar Coupling	39
2.5.1	Residual Dipolar Coupling Using an External Alignment Medium	39
2.5.2	Field Induced Residual Dipolar Coupling	41
2.6	Structural Calculation	42
2.6.1	Structural Calculation Using CYANA	42
2.6.2	Structural Calculation Using CNS	43
2.7	Dynamic Studies	43
2.7.1	HX Sample Preparation and NMR Experimental Parameters	43

2.7.2	Relaxation Studies on CaM-CD2-III-5G and CaM-CD2-IV-5G	47
2.7.3	ModelFree Simulation Experimental Preparation	49
2.8	Gradient Diffusion Experiments	51
Chapter 3	Developing the Grafting Approach and Using the Grafting Approach to Study the Metal Binding Properties of Calcium Binding Proteins	53
3.1	Developing the Grafting Approach	53
3.1.1	Engineering an EF-hand Loop into CD2 with Optimized Linkers	53
3.1.1.1	CD2 Variants with Different Lengths of Glycine Linkers	53
3.1.1.2	Metal Binding Studies of the CD2 Variants	56
3.1.2.1	NMR Structural Studies on Engineered Ca(II) Binding Protein: CaM-CD2-III-9G and CaM-CD2-III-13G	58
3.1.2.4	Summary of the NMR Studies on CaM-CD2-III-9G & CaM-CD2-III-13G	60
3.1.3	Determining the Effect of Local Electrostatic Environment	61
3.1.3.1	CD2 Variants with Different Protein Environments	61
3.1.3.2	Conformational and Metal Binding Studies	62
3.1.4	Metal Binding Studies on the Four EF-hand Motifs in Calmodulin	63
3.2	NMR Structural Studies on of C-terminal Domain of CaM	67
3.2.1	NMR Structural Studies on CaM-CD2-III-5G and CaM-CD2-IV-5G	68
3.2.2	Metal Binding Studies	72



3.2.2.1	Preparing NMR Sample for Metal Binding Studies	73
3.2.2.2	Ca(II) Metal Titration with CD2 Variants	73
3.2.2.3	La(III) Metal Titration with CD2 Variants	75
3.2.2.4	La(III) Metal Studies using 3D $^{15}\text{N}$ Edited Experiments	78
3.2.2.5	Discussion for the NMR Studies of the C-terminal Domain of CaM	80
3.3	Application of the Grafting Approach to the Study of Metal Binding Properties of a Predicted EF-hand Motif from Rubella Virus	83
3.3.1	NMR Structural Studies on the Rub-CD2-5G	84
3.3.2	La(III) Binding Study on Rub-CD2-5G	85
3.4	Application of a Grafting Approach to a Study of Metal Binding Properties of a Predicted Non-EF-hand Ca(II) Binding Site from CaR	86
3.4.1	NMR Structural Studies on CaR-CD2-III-0G	87
3.5	Summary	89
Chapter 4	Structural Determination of Engineered Proteins Grafted with EF-hand Motifs of Calmodulin using Heteronuclear NMR	122
4.1	Assignment of CaM-CD2-III-5G Using Homonuclear and Heteronuclear NMR Experiments	125
4.1.1.1	Homonuclear Sequential Assignment of CaM-CD2-III-5G	125

4.1.1.2	Homonuclear Sequential Assignment for the Host Protein	126
4.1.1.3	Aromatic Ring Assignment under D <sub>2</sub> O Conditions	129
4.1.1.4	Discussion of the Homonuclear Assignment	130
4.1.1.5	The Effect of Temperature and pH on the NMR Spectra	131
4.1.2	Sequential Assignment of the Heteronuclear Experiments	133
4.1.2.1	Triple Resonances Sequential Assignment Strategy	134
4.1.2.2	Sequential Assignment for CaM-CD2-III-5G	136
4.1.2.3	Summary of the Heteronuclear Sequential Assignment	140
4.1.2.4	Chemical Shift Analysis	141
4.1.3	Backbone and Sidechain NOE Assignment for Structural	
	Calculation	142
4.1.3.1	Assigning NOE for the Unresolved Region using 3D <sup>15</sup> N- NOESYHSQC	142
4.1.3.2	Mainchain and Sidechain NOE Assignment on the 2D NOESY Spectrum for Structural Calculation	144
4.1.3.3	Backbone to Backbone NOE Assignment	145
4.1.3.4	Sidechain to Sidechain and Sidechain to Mainchain NOE Assignment	145
4.1.3.5	Summary of the Assignment for CaM-CD2-III-5G from both the Homonuclear and Heteronuclear Experiments	146
4.2	Structural Calculation	147
4.2.1	CYANA Calculation using Automatic Distance Calibration	149

4.2.2	CYANA Calculation using Manual Mode	150
4.2.3.1	NOE Calibration for 3D $^{15}\text{N}$ NOESYHSQC	150
4.2.3.2	NOE Calibration for 2D NOESY	152
4.2.4.1	Obtaining the Dihedral Angles using HNHA	153
4.2.4.2	Assignment of the 3D HNHA Spectrum	154
4.2.4.3	Calculating the Dihedral Angles using TALOS	155
4.2.4.4	Predicting Dihedral Angles for CaM-CD2-III-5G	156
4.2.4.5	Summary of the Dihedral Angle vs. HNHA	158
4.2.4.6	Comparison to Calmodulin	158
4.2.5.1	Structural Calculation with NOE Distance Restraints	160
4.2.5.2	Adding the Dihedral Angle Restraint Table	161
4.2.5.3	Summary of NOE used for Calculation	163
4.2.5.4	The Structure of CD2 Host Protein of CaM-CD2-III-5G	165
4.2.5.5	The Structure of the Inserted EF-loop III of CaM-CD2-III-5G	165
4.2.5.6	Comparing the EF-loop III of CaM-CD2-III-5G to Calmodulin	166
4.2.5.6.1	Formation of the $\beta$ -Strand at Position 7 to 9 of the EF-loop	167
4.2.5.6.2	Dynamic Properties of the EF-loop Residues	169
4.2.5.6.3	No $3_{10}$ Helix Formation and Exiting Helix Formation	
	Observed in the EF-loop III of CaM-CD2-III-5G	170
4.3	Residual Dipolar Coupling Studies	171
4.3.1	Residual Dipolar Coupling from External Medium Induced	
	Alignment	172

4.3.2	Calculated Order Parameters using REDCAT	172
4.3.3	Structural Refinement Using Residual Dipolar Couplings	173
4.4	Paramagnetic Induced Alignment Using Ln(III) Metal Ions	175
4.4.1	Tm(III) Metal Titration of CaM-CD2-III-5G	179
4.4.2	Residual Dipolar Coupling from Field Induced Alignment	181
4.4.3.1	Model One: The Inserted EF-loop Metal Binding Site	185
4.4.3.2	Model II: The EF-loop and Non-Specific-Binding Site Residues D25, D26, D28, and E29	187
4.4.4.1	What are the Reasons for Low Magnitude of PCS and RDC	187
4.4.4.2	Did the Tm(III) and Dy(III) Induce Alignment	190
Chapter 5.0	Dynamics Studies on the C-terminal Domain of Calmodulin Using the Grafting Approach	234
5.1	Hydrogen Exchange Studies on the CD2 Variants	236
5.1.1	1D HX Experiments on CD2, CaM-CD2-III-5G, and CaM-CD2-IV-5G	236
5.1.2	2D HX Experiments on CD2 and CaM-CD2-III-5G	237
5.1.3	Comparing the HX Studies of CD2 to CaM-CD2-III-5G and CaM-CD2-IV-5G	239
5.2	T <sub>1</sub> , T <sub>2</sub> , and NOE Relaxation Studies for CaM-CD2-III-5G	241
5.2.1	T <sub>1</sub> Relaxation Studies on CaM-CD2-III-5G	241
5.2.2	Transverse Relaxation Studies on CaM-CD2-III-5G	243

5.2.3	HN One Bond NOE Studies on CaM-CD2-III-5G	244
5.2.4	Summary of T1, T2 and NOE Studies on CaM-CD2-III-5G	245
5.2.5	Calculation of $S^2$ Values Using ModelFree Simulation	246
5.2.5.1	ModelFree Simulation for CaM-CD2-III-5G	246
5.2.5.2	The Calculated Order Parameters for CaM-CD2-III-5G	248
5.2.5.3	Comparison of the Order Parameters of CaM-CD2-III-5G to the Corresponding Residues in 6D15 and Wild Type CD2	249
5.2.5.4	Comparison of the Order Parameters of the Inserted EF-loop III in CaM-CD2-III-5G to the EF-loop III of Calcium Free CaM	250
5.3	T1 and T2 Relaxation Studies for CaM-CD2-IV-5G	252
5.3.1	T1 Relaxation Studies on CaM-CD2-IV-5G	252
5.3.2	Transverse Relaxation Studies on CaM-CD2-IV-5G	253
5.3.3	Comparing the T1 and T2 Relaxation Studies of CaM-CD2-IV-5G to the CaM-CD2-III-5G	254
Chapter 6.0	Determining the Oligomeric States of CD2 Variants	280
6.1	Introduction	280
6.2	Determining the Oligomeric State of an Isolated EF-hand Loop	283
6.2.1	The Grafted EF-loop III in CD2 Remains Unpaired in the Absence of Metal Ions	283
6.2.2	The Grafted EF-loop III Remains Unpaired Upon Metal Binding	286

6.2.3	Discussion	287
6.3	NMR Structural Studies on CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF	290
6.3.1	Conformational Analysis on Engineered Calcium Binding Protein, CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF	290
6.3.2	Oligomeric Studies with CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF	293
6.4	Understanding the Contribution of the Helices to Dimerization	296
6.4.1	Proposed Modification to the Flanking Helices of Site III in CaM	298
6.4.2	Conformational Analysis on SKEAA using 1D $^1\text{H}$ NMR	299
6.4.3	2D NMR Structural Studies on the Engineered Calcium Binding Protein, SKEAA	299
6.4.4	Studies on the Oligomeric State of SKEAA	301
6.5	Major Finding of this Chapter	302
Chapter 7.0	Conclusions and Major Findings	327
7.1	Major Findings in Establishing the Grafting Approach	327
7.2	Major Finding in Obtaining the Site Specific Ca(II) Binding Properties	330
7.3	Major Findings in Determining the Contribution of the Helices to Metal Binding Affinity and Pair-Pair Interactions	332
	References	334
	Appendix	

2.1	NMRPipe Processing Scripts for IPAPHSQC	348
2.2	Input files and Scripts for CYANA	350
2.3	Input files and Scripts for CNS	374
3.1	Chemical shifts of CaM-CD2-III-5G	391
3.2	Chemical shifts of CaM-CD2-IV-5G	405

## List of Tables

Table 1.1	Summary of structural studies on CaM	21
Table 1.2	Interhelical angle in apo-CaM and Ca(II)-CaM	22
Table 3.1	Summaries of the CD2 Variants	56
Table 3.2	Metal binding affinities of CD2 variants	95
Table 3.3	Chemical shifts of HE1 protons of W32 of CD2 variants	97
Table 3.4	Ca(II) binding affinities of the CD2 variants with the EF-loops of CaM	101
Table 4.1	Summary of the sequential Assignment for the EF-loop III of CaM-CD2-III-5G	142
Table 4.2	NOE distance calibration scale for 3D $^{15}\text{N}$ -NOESYHSQC	151
Table 4.3	NOE distance calibration scale for $^1\text{H}$ 2D NOESY	152
Table 4.4	HNHA J-couplings of CaM-CD2-III-5G	212
Table 4.5	Predicted Dihedral Angles of CD2 and CaM-CD-III-5G using TALOS	213
Table 4.6	Summary of the first 6 cycle calculations	160
Table 4.7	Summary of the cycle 7 to 15 calculations	162
Table 4.8	Summary of the NOE Restraints	165
Table 4.9	Residual dipolar couplings of CaM-CD2-III-5G	223
Table 4.10	Summary of the PCS and RDC from Dy(III) and Tm(III)	230
Table 4.11	Putative distances between the paramagnetic metal ion	



	and affected residues	232
Table 4.12	Putative distances between affected residues to D26 and D28	233
Table 5.1	Hydrogen exchange rate for CD2 variants	263
Table 5.2	T1, T2, and NOE relaxation values for CaM-CD2-III-5G And CaM-CD2-IV-5G	268
Table 5.3	Simulation Approaches for ModelFree	246
Table 5.4	Comparing the T1 and T2 of CaM-CD2-III-5G to CaM	273
Table 5.5	Segment Secondary Structure Comparison Between CaM-CD2-III-5G and CaM-CD2-IV-5G	279
Table 6.1	Diffusion constants of CD2 variants	308

## List of Figures

Figure 1.1	Biological functions of Ca(II)	17
Figure 1.2	Properties of an EF-hand motif	18
Figure 1.3	Coordination of an EF-hand Ca(II) binding site	19
Figure 1.4	Example of CaM structures	20
Figure 1.5	Structural difference of an EF-hand motif in the presence and absence of Ca(II)	23
Figure 1.6	Structures of cell adhesion molecules	24
Figure 3.1	The grafting approach	91
Figure 3.2	<sup>1</sup> H Spectra of CD2 Variants with different glycine linkers	92
Figure 3.3	Calcium titration of CaM-CD2-III-5G	93
Figure 3.4	TOCSY spectra of CD2 Variants with extended glycine Linkers	94
Figure 3.5	HN chemical shifts comparison between CD2 variants and wild type CD2	96
Figure 3.6	HN chemical shifts comparison between CD2 variants and CaM-CD2-III-5G	98
Figure 3.7	<sup>1</sup> H Spectra of EF-loop III inserted at Q22, S52, and T83 positions in CD2	99
Figure 3.8	<sup>1</sup> H spectra of CD2 variants with different EF-loops of CaM insertion	100

Figure 3.9	The charged-balanced ligand model	101
Figure 3.10	Assigned $^{15}\text{N}$ HSQC spectrum of CaM-CD2-III-5G	102
Figure 3.11a	Fingerprint region of CaM-CD2-IV-5G TOCSY Spectrum	103
Figure 3.11b	Assigned $^{15}\text{N}$ HSQC spectrum of CaM-CD2-IV-5G	104
Figure 3.12	Amide region of $^1\text{H}$ spectra of CD2, CaM-CD2-III-5G, and CaM-CD2-IV-5G	105
Figure 3.13	$^{15}\text{N}$ HSQC spectra comparison between CaM-CD2-III-5G and CaM-CD2-IV-5G	106
Figure 3.14	$^{15}\text{N}$ HSQC Ca(II) titration for CaM-CD2-III-5G	108
Figure 3.15	$^{15}\text{N}$ HSQC Ca(II) titration for CaM-CD2-IV-5G	109
Figure 3.16	$^{15}\text{N}$ HSQC La(III) titration for CaM-CD2-III-5G	110
Figure 3.17	$^{15}\text{N}$ HSQC La(III) titration for CaM-CD2-IV-5G	111
Figure 3.18	EF-loop III region of the La(III) Titration	112
Figure 3.19	EF-loop IV region of the La(III) titration	113
Figure 3.20	$^{15}\text{N}$ HSQC spectrum of CaM-CD2-III-5G in the presence of 2 mM La(III)	114
Figure 3.21	3D NOESYHSQC strip plot of CaM-CD2-III-5G	115
Figure 3.22	Structural comparison between Ca(II) free and Ca(II) loaded of CaM	116
Figure 3.23	Fingerprint region of the Rub-CD2-5G TOCSY spectrum	117
Figure 3.24	TOCSY spectra of Rub-CD2-5G in the presence of	

	La(III) and EGTA	118
Figure 3.25	Fingerprint region of the CaR-CD2-III-0G TOCSY Spectrum	119
Figure 3.26	$^{15}\text{N}$ HSQC spectrum of CaR-CD2-III-0G	120
Figure 3.27	Comparison between the $^{15}\text{N}$ HSQC spectra of CaR-CD2-III-0G to CaM-CD2-III-5G	121
Figure 4.1	TOCSY spectra of CD2 and CaM-CD2-III-5G	194
Figure 4.2	Example of sequential assignment of CaM-CD2-III-5G	195
Figure 4.3	Example of Asn sidechain assignment of CaM-CD2-III-5G	196
Figure 4.4	Aromatic ring protons assignment of CaM-CD2-III-5G	197
Figure 4.5	Homonuclear assignment of CaM-CD2-III-5G	198
Figure 4.6	Chemical exchange rates of backbone amides	199
Figure 4.7	TOCSY spectra of CaM-CD2-III-5G in pH 5.0 and 7.4	200
Figure 4.8	$^{13}\text{C}$ HSQC spectrum of CaM-CD2-III-5G	201
Figure 4.9	Example of triple resonance assignment procedure	202
Figure 4.10	HNCACB strip plot of CaM-CD2-III-5G	203
Figure 4.11	Assigned $^{15}\text{N}$ HSQC spectrum of CaM-CD2-III-5G	204
Figure 4.12	Assignment using HNCACB and CBCACONH spectra	205
Figure 4.13	Comparing HN and $\text{C}\alpha$ chemical shifts of CaM-CD2-III-5G to the same residues in wild type CD	206
Figure 4.14	CSI analysis for CaM-CD2-III-5G	207
Figure 4.15	Domain 1 of the cell adhesion molecule, CD2	208

Figure 4.16	Example of the NOESY spectrum of CaM-CD2-III-5G	209
Figure 4.17	The interaction between the W32 and residues that are within 5 Å	210
Figure 4.18	Solution structure of CaM-CD2-III-5G with no error Correction	211
Figure 4.19	Dihedral angle comparison between the HNHA and TALOS results to CD2	215
Figure 4.20	Dihedral angle comparisons on the EF-loop III of CaM-CD2-III-5G to CaM	216
Figure 4.21	Solution structure of CaM-CD2-III-5G without dihedral angle restraints	217
Figure 4.22	Final solution structure of CaM-CD2-III-5G	218
Figure 4.23	NOE pattern of the CaM-CD2-III-5G	219
Figure 4.24	PEG medium for the RDC studies	221
Figure 4.25	Example of an IPAPHSQC spectrum	222
Figure 4.26	Comparison of experiment RDC and calculated RDC	224
Figure 4.27	The detectable paramagnetic effect	225
Figure 4.28	Tm(III) titration on CaM-CD2-III-5G	226
Figure 4.29	HN chemical shifts change in Tm(III) titrations	227
Figure 4.30	<sup>15</sup> N HSQC spectra of CaM-CD2-III-5G in the presence of La(III) and Tm(III)	228
Figure 4.31	<sup>15</sup> N HSQC spectra of CaM-CD2-III-5G in the presence	

	of La(III) and Dy(III):	229
Figure 4.32	Simulated structure of metal bound CaM-CD2-III-5G	231
Figure 5.1	Exchangable protons in protein	257
Figure 5.2	Hydrogen exchange spectra of CD2	258
Figure 5.3	The fitting cures for the HX rates of CD2 variants	259
Figure 5.4	TOCSY spectra of CD2 in D2O	260
Figure 5.5	TOCSY spectra of CaM-CD2-III-5G in D2O	261
Figure 5.6	Location of the liable protons with HX rates	262
Figure 5.7	T1 relaxation spectra of CaM-CD2-III-5G	264
Figure 5.8	Intensities decay on T1 spectra	265
Figure 5.9	T1 data fitting curve for CaM-CD2-III-5G	266
Figure 5.10	Summaries of the T1, T2, and NOE values	267
Figure 5.11	T2 relaxation spectra of CaM-CD2-III-5G	269
Figure 5.12	NOE on and off spectra of CaM-CD2-III-5G	270
Figure 5.13	S <sup>2</sup> ordered parameters for CaM-CD2-III-5G	271
Figure 5.14	S <sup>2</sup> ordered parameters comparison between CaM-CD2-III-5G and 6D15	272
Figure 5.15	T1 relaxation spectra of CaM-CD2-IV-5G	274
Figure 5.16	Summaries of the T1 and T2 values for CaM-CD2-IV-5G	275
Figure 5.17	T2 relaxation spectra of CaM-CD2-IV-5G	276
Figure 5.18	T1 and T2 values of CD2 variants:	277
Figure 5.19	Section Comparison between the secondary structure	

	of CaM-CD2-III-5G and CaM-CD2-IV-5G	278
Figure 6.1	Sequence details of CD2 variants with helices	305
Figure 6.2	Diffusion spectra of CaM-CD2-III-5G, glycine and buffer	306
Figure 6.3	The data fitting curve for the diffusion constants	307
Figure 6.4	CD spectra of CD2 variants	309
Figure 6.5	$^1\text{H}$ NMR spectra of CD2, CaM-CD2-III-5G, CaM-CD2-III-5G-F, and CaM-CD2-III-5G-EF	310
Figure 6.6	Temperature studies on CaM-CD2-III-5G-EF	311
Figure 6.7	TFE titration on CaM-CD2-III-5G-F	312
Figure 6.8	TFE titration on CaM-CD2-III-5G-EF	313
Figure 6.9	Ca(II) titration on CaM-CD2-III-5G-F	314
Figure 6.10	Ca(II) titration on CaM-CD2-III-5G-EF	315
Figure 6.11	$^1\text{H}$ Spectra of CD2 variants in the presence of EDTA	316
Figure 6.12	PFG diffusion studies on the CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF	317
Figure 6.13	Sedimentation Studies on CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF	318
Figure 6.14	The CSU analysis on the C-terminal domain of CaM	319
Figure 6.15	Modification scheme for removing hydrophobic residues of the EF-helices	320
Figure 6.16	$^1\text{H}$ spectrum of SKEAA	321
Figure 6.17	Fingerprint region of the SKEAA TOCSY spectrum	322

Figure 6.18	HN chemical shifts comparison between CaM-CD2-III-5G and SKEAA	323
Figure 6.19	Comparison of the sidechain region of CaM-CD2-III-5G to SKEAA TOCSY spectra	324
Figure 6.20	Comparison between the fingerprint regions of CaM-CD2-III-5G to SKEAA	325
Figure 6.21	PFG diffusion studies on SKEAA	326



### List of Abbreviations

Ca(II)	Calcium
CaM	Calmodulin
CD	Circular Dichroism
CD2	Cluster of differentiation 2
DNA	Deoxynucleic acid
DTT	Dithiothreitol
EGTA	Ethylene Glycol-bis( $\beta$ -aminoethyl Ether)
FPLC	Fast performance liquid chromatography
GS4B	Glutathione sepharose 4B
GST	Glutathione-S-transferase
HSQC	Heteronuclear Single Quantum Correlation
IPTG	Isopropyl $\beta$ -D-thiogalactoside
K <sub>d</sub>	Dissociate constant
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Enhancement Spectroscopy
PBS	Phosphate buffer saline
PDB	Protein data bank
TFE	2,2-trifluoroethanol
TOCSY	Total Correction Spectroscopy

## **Chapter 1.0 Introduction**

### **1.1 The Role of Calcium in Biological Systems**

Calcium (Ca(II)) is an important element in biological systems. It is an essential component in the biomineralization of teeth, bones, and shells, as well as a second messenger regulating cellular processes such as cell division and growth, secretion, ion transport, and muscle contraction (1). The biological role of Ca(II) is dependent on the Ca(II) concentration in the intra- and extracellular compartments of the cell, which regulates cellular processes by mediating the activities of Ca(II) receptors and/or Ca(II) binding proteins (2-15). As shown in Figure 1.1, each compartment of the cell contains different types of Ca(II) binding proteins whose Ca(II) binding affinities vary by  $10^6$ -fold (2, 16-23). Ca(II) binding proteins are divided into several major types: the trigger/sensors, buffer proteins, and stabilizing proteins. The trigger/sensor Ca(II) binding proteins in the cytosol, such as calmodulin (CaM) and troponin C (TnC), have Ca(II) binding affinities in the sub-micromolar range (24, 25). Upon binding Ca(II), the trigger/sensor proteins, such as CaM, undergo conformational changes that allow them to interact with target molecules and regulate more than 100 different cellular events and processes (3-6). The buffer type Ca(II) binding proteins, such as calbindinD9k and parvalbumin, bind Ca(II) without major conformational changes and act as intracellular Ca(II) buffers to regulate cytosolic Ca(II) levels. The functions of buffer proteins are important to prevent cells from Ca(II) overload and apoptosis (14). The Ca(II) binding affinities for extracellular proteins such as

Ca(II)-sensing receptors and cadherins are between 0.1–10 mM, which corresponds to the lower extracellular Ca(II) concentration (26-33). CaR is a transmembrane G-protein coupled receptor that activates numerous intracellular processes upon Ca(II) binding (34). Other types of Ca(II) binding proteins, such as cadherins, have been shown to support the structural integrity of cells and tissues and also to mediate many signal transduction pathways (2, 35).

## **1.2 Structural Properties of Ca(II) Binding Proteins**

Due to the relatively large ionic radius of Ca(II) and positive (+2) charge, Ca(II) prefers a high coordination number and negatively-charged ligands. The Ca(II) binding sites are classified as either continuous or discontinuous based on the organization and the location of the Ca(II) binding ligands. For protein with continuous binding sites, the Ca(II) binding ligands are sequentially adjacent a small continuous segment of the primary protein sequence (ie, the EF-hand motif). These Ca(II) binding sites, that have highly conserved regions in the primary sequence, can be identified by computational pattern searches or structural determination (36).

The Ca(II) binding ligands of the discontinuous proteins, such as cadherins, C<sub>2</sub> domains, site I of thermitase, phospholipase A<sub>2</sub>, and the D-galactose binding protein, are formed by segments located remotely from one another in the protein sequence (37). It is difficult to identify the locations of such ligands by sequence information alone. Our laboratory has shown that both

continuous and discontinuous Ca(II) binding sites can be identified based on the local geometric description (38).

### 1.2.1 Properties of the EF-hand Motif

The EF-hand motif is one of the most common and important motifs for proteins with continuous Ca(II) binding sites. Genomic research has shown that one in five motifs found in animal cells is an EF-hand motif. Currently, there are more than 1000 EF-hand Ca(II) binding proteins found in prokaryotes and eukaryotes and the EF-hand proteins can be classified into 77 distinct sub-families (39-41). The term EF-hand motif was first used by Dr. Kretsinger to describe the Ca(II)-binding site in parvalbumin (42) (Figure 1.2).

The EF-hand motif can be further categorized into several different classes based on their ligand arrangements as canonical, pseudo EF-hand, essential light chain, or BM40. A canonical EF-hand motif spans 29 residues and consists of a highly-conserved Ca(II)-binding loop flanked by two helices (helix-loop-helix). A 12 residue loop (EF-hand loop) contains all of the Ca(II)-binding ligands. Although the ligand number of small compounds such as EGTA varies from 3 to 9, an EF-hand protein typically binds seven oxygen atoms. These ligand atoms from the sidechains of Asp, Asn, Glu, the mainchain, and water at the EF-loop positions 1, 3, 5, 7, 9, and 12 coordinate the Ca(II) ion in a pentagonal bipyramidal geometry (Figure 1.3) (21, 43). In a typical geometry, position 1 of the Ca(II)-binding loop (the side-chain of Asp) serves as the ligand

on the x-axis. The x-axis (position 9) is filled by a bridged water molecule connecting the sidechain of Asp, Ser and Asn (44-46). The z-axis is shared by the bidentate carboxyl oxygen atoms at position 12 (predominately Glu).

The central EF-loop of the pseudo EF-hand motif contains 14 residues rather than 12 residues (47, 48). All of the oxygen ligands originate from the mainchain except the residue at position 14, which normally comes from the sidechain of Glu. Site I of the proteins in the S100 family all contain psuedo EF-hand motifs. Also, the essential light chain and BM40 have several additional insertions on the central EF-loop.

### **1.2.2 Pairing of the EF-hand Motifs and Cooperative Binding**

The majority of EF-hand proteins contain multiple even numbered copies (in a range from two to twelve copies) of EF-hand motifs, where the two closely packed helix-loop-helix modules within a single globular domain constitute the basic Ca(II)-binding unit of EF-hand proteins. Two EF-hand motifs are arranged with respect to each other in a pseudo 2-fold symmetry in the same protein domain and yield highly cooperative Ca(II) binding systems (21, 49). The distance between two Ca(II) ions in two paired EF-hand motifs is usually about 11 Å, and the coordination shells of the paired EF-hand binding motifs in the same protein domain can completely overlap in most EF-hand proteins. Peptide fragment studies on the EF-hand motifs of troponin C and calbindinD9k have shown that the hydrophobic residues on the helices and at position 8 of the EF-

loop are forming hydrophobic interactions with another EF-hand peptide fragment. The hydrophobic interactions between the paired EF-hand motif is one of the major driving forces for pairing the EF-hands (50-52). It has been reported that EF-hand proteins utilize the paired EF-hand Ca(II) binding motifs to regulate many cellular functions such as muscle contraction, neuronal signaling, apoptosis, and cell cycle control (8, 21, 53-55). There are a few exceptions (e.g. parvalbumin and calpain) where the EF-hand proteins contain an odd number of EF-hand motifs. The hydrophobic residues on the helices of parvalbumin are buried by another helix element in the same protein. In the case of calpain, the fifth EF-hand motif forms an intermolecular dimer with another protein (12). It is important to understand the key determinants that contribute to the pairing of EF-hand Ca(II) binding proteins. This question will be addressed in chapter 6.

### **1.3 Structural and Dynamic Studies of CaM**

As shown in Figure 1.4, CaM is an  $\alpha$ -helical protein with 148 amino acids. CaM has two domains that are connected by a flexible central helix linker. The two domains are referred to as the N-terminal and C-terminal domains and each domain has a pair of canonical EF-hand motifs. The structure of the open form (Ca(II) bound) CaM was first solved by Babu and co-workers in 1985 using X-Ray crystallography (56). Since then, advances in X-Ray crystallography and NMR spectroscopy techniques have enabled researchers to determine the structures for the closed form (apo form), the open form (bound with Ca(II), Ce(II))

and Pb(II)), and the complex forms of CaM (complex with target peptides) (57, 58). Currently, there are more than 100 structures related to CaM deposited in the protein data bank (59). We will focus on the closed form and open form of CaM. A summary of structural studies for CaM are listed in Table 1.1.

In the absence of Ca(II), the structure is referred to as the closed form where the molecular recognition surface of CaM is buried inside the hydrophobic surface (60, 61). The structural studies on the Ca(II) free CaM by Bax, Ikura, and Forsen's research groups in 1995 using NMR revealed that the four EF-hand motifs still maintain the helix-loop-helix conformation (60, 62, 63). In 2004, Schumacher et al. determined the structure of the apo form CaM using X-Ray crystallography (64). Both the NMR and X-Ray structural studies suggested that the helices of the EF-hand motif possess a parallel orientation forming a four-helix bundle for each domain.

The dynamic studies of the Ca(II) free CaM indicated that the first six residues in all four EF-loops are dynamic and disordered due to charge repulsion from the charged sidechains of the ligand residues (65). The NMR studies on CaM further revealed that the conserved hydrophobic residues on the helices and at position 8 (a small  $\beta$ -strand formed from residues 7 to 9) of the EF-loop form non-covalent interactions between the paired EF-hand motifs to provide stability (60, 62). Additionally, mutation studies on position 8 of the EF-loop by Browne et al demonstrated that removal of the hydrophobic residues can destabilize CaM's structure both in the presence and absence of Ca(II) (66).

Four Ca(II) ions bind cooperatively to CaM, which leads to a conformational change in the protein and allowing for the interaction with target molecules. The bidentate ligand at position 12 of the EF-loop moves closer to position 1 of the EF-loop to form the Ca(II) binding pocket. The movement of the Ca(II) binding ligands causes the orientation of the entering and exiting helices of the EF-hand motif to become orthogonal to each other, and the changes in the interhelical angle between the two helices are shown in Figure 1.5 and Table 1.2. It is important to point out that the interhelical angles within a given motif ( $>25^\circ$ ) for all four EF-hand motifs, while the interhelical angles between the helices from two different EF-hand motifs does not change significantly. For example, the interhelical angle between the entering helix (HELIX I) of the EF-hand motif I and exiting helix (HELIX IV) of EF-hand motif II are  $127^\circ$  and  $110^\circ$  in the absence and presence of Ca(II), respectively. This observation suggests that the interactions between the helices of paired EF-hand motifs provide stability for the EF-hand protein in the presence and absence of metal. The interaction between the paired EF-hand motifs in the closed and open forms may be responsible for the conformational changes and cooperativity, which is observed in all proteins in the CaM superfamily (Ca(II)-dependent protein kinase, CaM, CaM-like protein, caltractin, squidulin, and troponin C).

CaM is found in most living organisms from yeast to humans, with highly conserved sequences and the tertiary Ca(II) bound structures (46, 56, 63, 67-77). The structural studies shows that all open forms of CaM contain two globular



domains with small structural deviations and a flexible central linker (Figure 1.5). The structural studies on the different species of Ca(II) free CaM have indicated that there are noticeable minor structural differences (60, 62, 63, 74, 78). It was speculated that the structural differences in the closed form may enable CaM to adapt to different cellular environments, which allows the open form of CaM to regulate more than 100 types of signaling processes (3-6).

#### **1.4 Understanding the Site Specific Metal Binding Properties of CaM**

To understand the conformational changes and the cooperative binding of Ca(II) to CaM, it is essential to know the site specific Ca(II) binding and their relative contribution to the overall Ca(II)-dependent conformational change process. To date, the studies on the site specific Ca(II) binding properties of CaM are based on site-directed mutagenesis on the full length protein, sub-domain by trypsin cleavage, or synthesized EF-loop peptide. One of the common approach is to make mutations of the key residues that are involved in the Ca(II) binding. More than 50 different mutations have been made on the Ca(II) binding loop, helices, and the loop region of CaM (66, 79-89). More than 50 different mutations have been made on the Ca(II) binding loop, helices, and the loop region of CaM. The mutation studies either enhanced or decreased the metal binding affinities of the native Ca(II) binding site (66, 79-89).

One of the major complications preventing us using mutagenesis to probe key factors that contribute to Ca(II) binding is due to the labile protein frame and

different level of cooperativity. As we mentioned briefly in section 1.3, the hydrophobic interactions between the paired EF-hand motifs play crucial roles in stability and cooperativity of the EF-hand protein. An interesting study was carried out by Fefeu and co-workers, which the Val residue at position 8 of the EF-loop IV of CaM was mutated to a Gly (V136G) (81). Position 8 of the EF-loop is highly conserved in the EF-hand superfamily, as it provides two important interactions. First, the sidechain at position 8 plays is involved in forming the hydrophobic core of each domain of CaM. Second, the backbone carbonyl (C=O) and amine (N-H) groups at position 8 form inter-strand hydrogen bonding between the paired EF-hand sites, which stabilizes the cycle of cooperative interactions between the paired sites (44, 81, 90). The V136G mutation will affect the first interaction because the hydrophobic sidechain is no longer present, but will not affect the second interaction described here because the non-covalent interaction between the backbone of the two motifs remains the same. The structural studies indicated that site IV of the V136G mutant is unfolded in the absence of Ca(II). The thermal stability studies on the V136 mutant show that the removal of the hydrophobic residue at position 8 of EF-loop IV decreased the stability of the C-terminal domain. Fefeu et al. reported that decrease in the stability of the C-terminal domain causes the Ca(II) binding affinities of Site III and Site IV (C-terminal domain) of CaM to decrease dramatically in comparison to the wild type CaM. Since the Ca(II) binding ligands of both sites III and IV were not mutated, thus the changes in the

hydrophobic interactions between the paired EF-hand motif have affected the magnitude of the Ca(II) binding affinity. By monitoring the NMR signals of I27 and I63 (position 8 of site I and site II CaM, respectively), the N-terminal domain of V136G mutant was observed to have stronger Ca(II) binding affinities in comparison to those of the wild type CaM. This further indicated that the Ca(II) ions occupied the N-terminal sites first. Fefe et al. concluded that the N-terminal domain interacts with the hydrophobic and acidic residues of the unfolded site IV of V136G, which leads to a stronger Ca(II) binding affinity for the N-terminal domain. Studies with the V136G mutant suggest there are inter-motif interactions between the paired sites and inter-domain interactions between the N-terminal and C-terminal domain and both interactions will also affect the metal binding affinity of CaM.

To study the Ca(II) binding properties on each individual domain of CaM, Linse and co-workers used trypsin to cleave CaM into two domains to study the metal binding properties of each domain (91). They were able to conclude that both the N-terminal and C-terminal domains have strong Ca(II) binding affinity without the presence of the other domain. The average binding affinities of the EF-loop I and II in N-terminus and EF-loop III and IV in C-terminus are 2.0 and 0.31 to 3.0  $\mu\text{M}$ , respectively. To further understand the contribution of the EF-loop to Ca(II) binding affinity, several synthesized peptides with the sequence of EF-loop III of CaM was used for metal binding studies (40, 92, 93). Unfortunately, the metal binding affinity of the synthesized peptide is 3 orders of

magnitude lower than the average metal binding affinity of the C-terminal domain of CaM. The low metal binding affinity observed in the synthesized peptide is likely due to the isolated EF-loop peptide existing in an unstructured state in solution. A model system is required to obtain the site specific Ca(II) binding properties of EF-hand protein, such as CaM.

### **1.5 Our Research Approach**

The goal of our research is to understand the site-specific properties of EF-hand Ca(II) binding proteins. There is a strong need to develop a model system to gain a clear understanding on the activation of CaM and other EF-hand proteins. The single site model system is designed to allow us to study the Ca(II) binding properties of EF-hand proteins without the complications associated with multiple site cooperativity (e.g. conformational changes, binding to multiple ions, EF-hand pairing interactions).

To study the individual EF-hand Ca(II) binding motifs without the complications of natural Ca(II) binding proteins, we endeavored to obtain site-specific Ca(II) binding properties by grafting a single Ca(II) binding site into a scaffold protein to evaluate the intrinsic binding affinity and the contribution of residue types on the EF-loop to Ca(II) binding. To achieve this goal, several criteria were considered. First, the host protein must be able to maintain its native structure upon insertion of the EF-loop. This feature would not only allow us to focus specifically on the intrinsic binding ability of an individual Ca(II)-

binding site without the limitations of peptide models, but also avoided the difficulty in determining the Ca(II) binding affinity of a single binding site due to the cooperative binding of multiple sites or conformational changes (21).

Second, the inserted EF-loop in the foreign host protein would have to be able to maintain the native conformation of wild type CaM, and the structure of the host protein should not restrict the motion of the inserted EF-loop. Third, the variables in the environment at the insertion location, such as the electrostatic potential, should not interfere with the ability of the EF-loop to bind metal ions.

### 1.5.1 Choice of the Host Protein

Domain 1 of cell adhesion protein CD2, comprised of nine anti-parallel  $\beta$ -strands, belongs to the immunoglobulin superfamily (IgSF). The Ig-fold architecture is similar to the Ca(II) dependent cell adhesion molecule, cadherin (Figure 1.6) (18, 94-97). The N-terminal domain CD2 was selected as the host protein for several reasons (94). First, the N-terminal domain of CD2 is small (99 amino acids) with several solved structures at high resolutions (98-102).

Second, CD2 is highly tolerant toward protein engineering. Previous studies have reported that single mutations at 40 different locations on CD2 have no apparent effects on the expression, solubility, and structural integrity of the protein (94, 103-105). Using computational design methods, our group has created *de novo* Ca(II) binding sites in CD2 that are shown to maintain their native-like conformations (106, 107). Third, high resolution NMR studies show

that CD2 is able to maintain its native structure and conformation in a pH range from 1 to 10, suggesting that electrostatic interactions play a minor role in the folding of CD2 (94). Fourth, the dynamic properties of CD2 are well studied by Driscoll and coworkers (102, 108). Our laboratory has previously reported the dynamic properties of CD2 with a designed Ca(II) binding site (109). These important features and knowledge of CD2 allow us to specifically explore the conformational and dynamic properties of the host protein with a grafted Ca(II) binding site.

## **1.6 Motivations and Overview of These Studies**

The focus of this study is to obtain the key factors that regulate Ca(II) binding and dimerization of the EF-hand proteins by a grafting approach. NMR spectroscopy is an important technique that can provide conformational and dynamic properties for molecules of different sizes in solution. In this dissertation, NMR was utilized to determine metal binding, structural, and dynamic properties of the engineered Ca(II) binding proteins. Using the grafting approach, the EF-hand sites of CaM are individually inserted into a scaffold protein to determine the specific Ca(II) binding properties. The application of the grafting approach to other proteins, such as CaR and Rubella is also discussed.

**Chapter 2** discusses the methods and materials used for all studies. This chapter includes detailed procedures for protein expression, purification, and

NMR experimental parameters for the CD2 variants. A summary of trial steps to improve the protein stability of the CD2 variants is also included in this chapter.

**Chapter 3** focuses on establishing the grafting approach and using the grafting approach to obtain site specific Ca(II) binding properties of CaM. CD2 variants with different numbers of glycine linkers and different numbers of glycine residues per linker were constructed to ensure that the EF-hand motif has enough flexibility to bind with Ca(II). The EF-hand motif was inserted into three locations of CD2 with different electrostatic potentials to determine effects of the local electrostatic environment on the metal binding affinity. The structural and metal binding properties of the CD2 variants were investigated using homonuclear and heteronuclear experiments. The metal binding affinities of all four EF-hand motifs of CaM are summarized. We have further proposed a working model to explain the magnitude of metal binding affinities in EF-hand proteins. Both the structural difference in apo form and dynamic properties are the key determinants for their metal binding affinity. Using the grafting approach, the metal binding properties of the continuous metal binding sites of rubella virus protease and CaR were also characterized.

**Chapter 4** discusses the structural determination for the engineered protein with the inserted CaM EF-loop III. The sequential assignments using homonuclear and heteronuclear strategies are discussed in this chapter. The

procedure for generating the NOE distance and dihedral angle restraints for structural calculations using CYANA are presented. The structural properties of the host protein and the inserted EF-loop are discussed in detail. The residual dipolar coupling studies for the engineered protein are also discussed in this chapter.

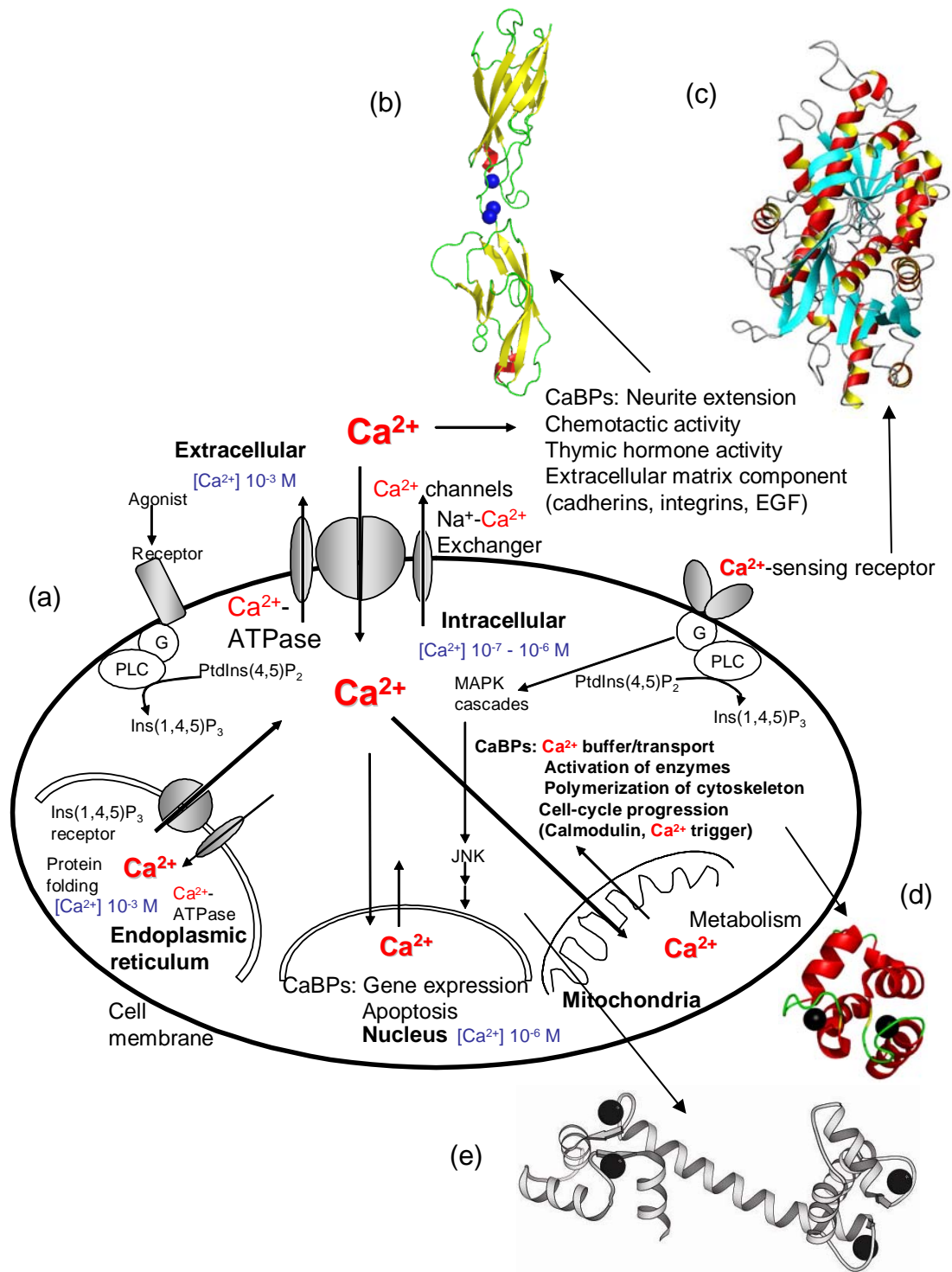
**Chapter 5** focuses on obtaining dynamic properties of the C-terminal domain of CaM. The T1, T2, and NOE relaxation rates were obtained using heteronuclear experiments. The simulation of the order parameter for the engineered protein using the program MODELFREE is presented. The hydrogen exchange studies for wild type CD2, CaM-CD2-III-5G (EF-loop III of CaM inserted into CD2), and CaM-CD2-IV-5G (EF-loop IV of CaM inserted into CD2) are discussed in detail.

**Chapter 6** focuses on the contribution of the EF-helices to metal binding affinity and cooperative binding. The oligomeric states of the CD2 variants in the presence and absence of metal were determined using PFG diffusion experiments. Mutagenesis studies on the hydrophobic residues of the EF-helices are also discussed in detail. The contributions of hydrophobic residues on the EF-helices for pairing the EF-hand motifs were further examined.

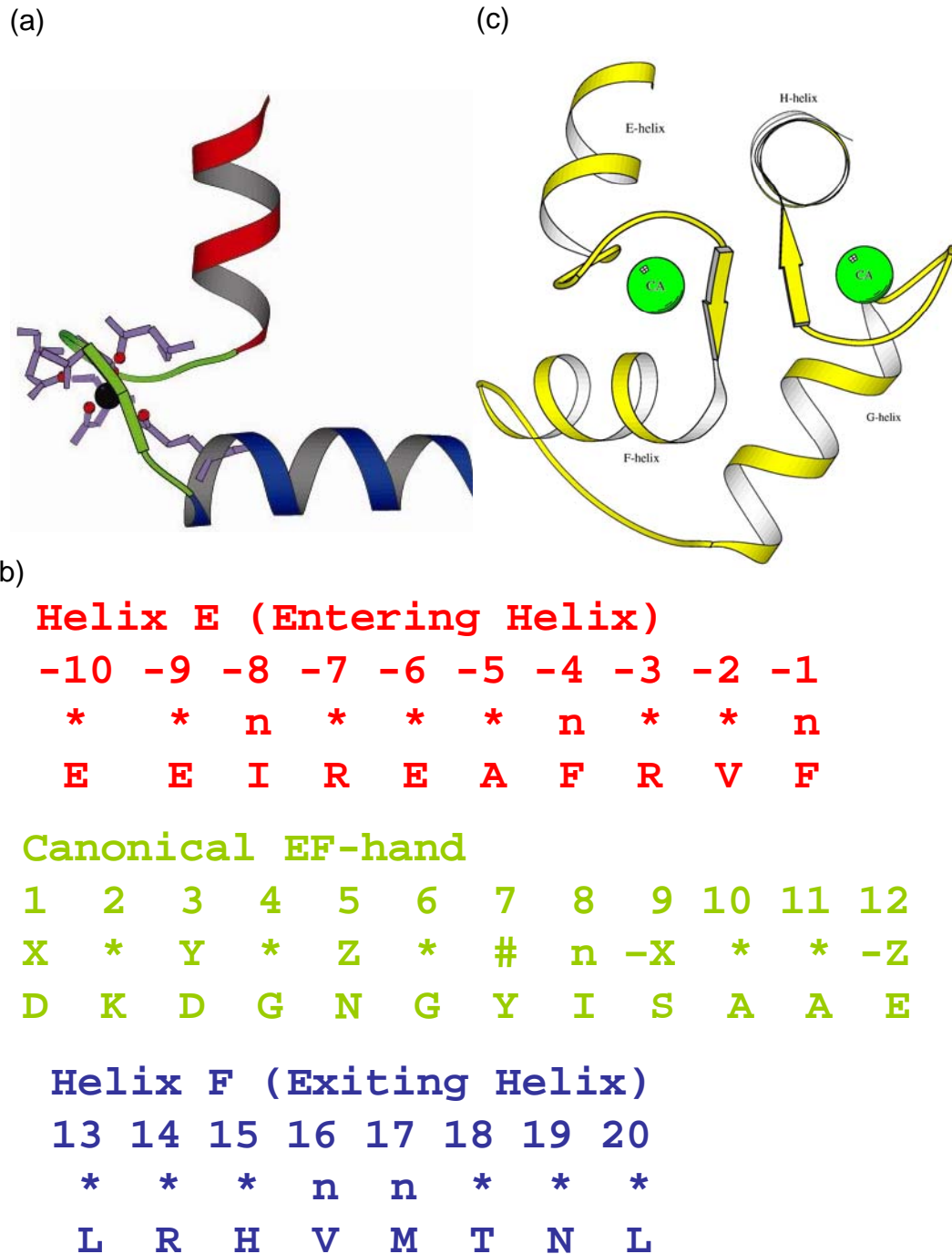
The information on the single EF-hand site will enable us to understand the intrinsic variability of the intact EF-hand protein. Knowledge regarding the



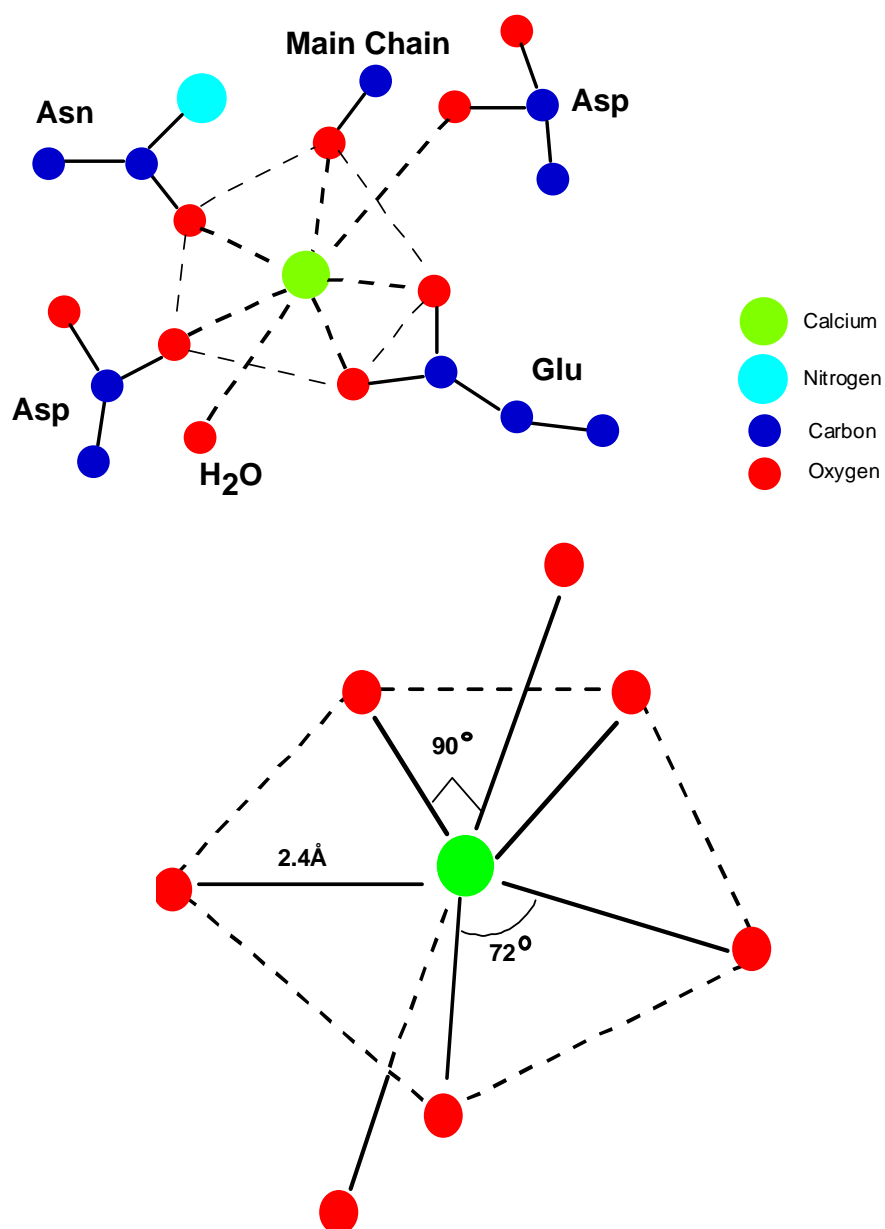
mechanisms of cooperative binding and conformational changes can provide inside views into how the EF-hand proteins control Ca(II) signaling and cellular Ca(II) levels (110)(111).



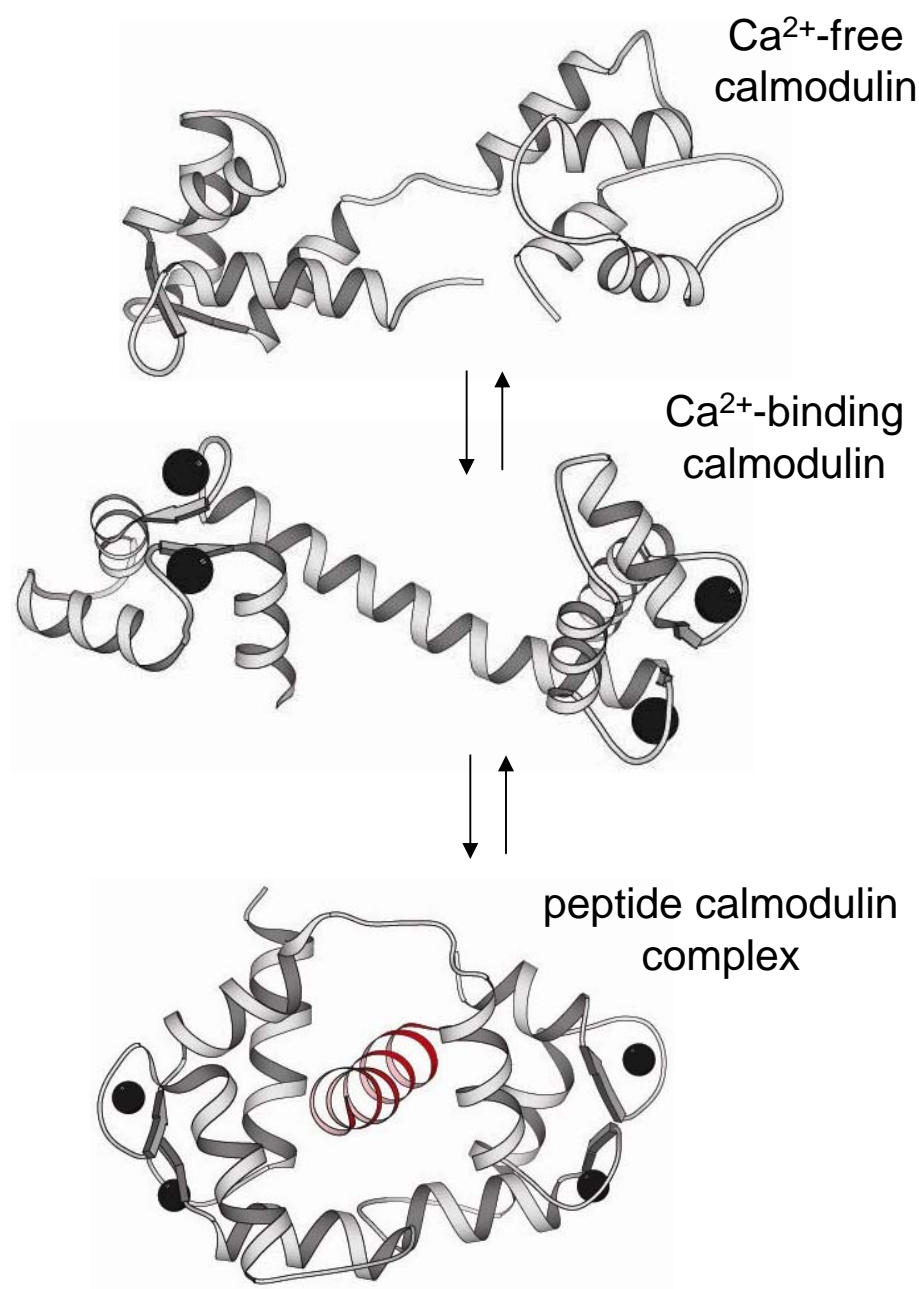
**Figure 1.1 Biological functions of Ca(II):** (a) Ca(II)-signaling and Ca(II)-binding proteins in a eukaryotic cell. (b) Cadherin (1EDH.pdb), (c) CaR (homology model based on 1EWT.pdb), (d) Calbindin D9k (4ICB.pdb), and (f) CaM (3CLN.pdb).



**Figure 1.2 Properties of an EF-hand motif:** (a) The entering helix is shown in blue and the exiting helix is shown in red. (b) The hydrophobic residues on the helices and EF-loop are label as n. The EF-loop of canonical EF-hand motif is 12 residues long and is shown in green. The Ca(II) binding ligands are located at positions 1(X), 3(Y), 5(Z), 7, and 12(-Z) of the EF-loop. (c) An example of the C-terminal domain of CaM forming pair-pair interaction.



**Figure 1.3 Coordination of an EF-hand Ca(II) binding site:** The oxygen atoms from sidechain, mainchain and water coordinate the calcium ion in a pentagonal bipyramidal geometry..



**Figure 1.4 Example of CaM structures:** (a) Ca(II) free from of CaM (1CFC.pdb), (b) Ca(II) loaded form of CaM (3CLN.pdb), (c) the CaM forms peptide complex with M13 peptide (1CDL.pdb).

Table 1.1 Summary of Structural Studies on CaM

PDB	Year	Length	pH	Exp	Res	Metal	Species	Authors	Correspond
1CFC	1995	Full	6.3	NMR	25	Apo	African frog	Kuboniwa, H	Bax, A
1CFD	1995	Full	6.3	NMR	1	Apo	African frog	Kuboniwa, H	Bax, A
1CMF	1995	C-terminal	6.0	NMR	20	Apo	Bovine	Finn, BE	Forsen, S
1DMO	1996	Full	7.5	NMR	30	Apo	African frog	Zhang, M	Ikura, M
1F70	2000	N-terminal	7.0	NMR	10	Apo	African frog	Chou, JJ	Bax, A
1F71	2000	C-terminal	7.0	NMR	10	Apo	African frog	Chou, JJ	Bax, A
1LKJ	2003	Full	7.0	NMR	31	Apo	Yeast	Ishida, H	Yazawa, m
1F54	2003	N-terminal	6.8	NMR	30	Apo	Yeast	Ishida, H	Yazawa, m
1QX5	2004	Full	6.5	X-Ray	2.54	Apo	Rat	Schumacher, MA	Miller, MC
3CLN	1988	Full		X-Ray	2.2	Ca(II)	Rat	Babu, YS	Cook, WJ
4CLN	1992	Full	4.0	X-Ray	2.2	Ca(II)	Fruit Fly	Taylor, DA	Quiococho, FA
1CLL	1993	Full		X-Ray	1.7	Ca(II)	Human	Chattopadhyaya, R	Quiococho, FA
1CLM	1993	Full	5.0	X-Ray	1.8	Ca(II)	Ciliate	Rao, ST	Sundaralingam, M
1OSA	1993	Full	5.0	X-Ray	1.68	Ca(II)	Ciliate	Ban, C	Sundaralingam, M
1DEG	1994	Full	6.1	X-Ray	2.9	Ca(II)	Bovine	Kretsinger, RH	Persechini, A
1CMG	1995	C-terminal	6.0	NMR	20	Ca(II)	Bovine	Finn, BE	Forsen, S
1AK8	1997	N-terminal	6.0	NMR	23	Ce(III)	Bovine	Bentrop, D	Malmendal, A
1EXR	2000	Full	5.0	X-Ray	1	Ca(II)	Ciliate	Wilson, MA	Brunger, AT
1J7O	2001	N-terminal	7.0	NMR	3	Ca(II)	Human	Chou, JJ	Bax, A
1J7P	2001	C-terminal	7.0	NMR	3	Ca(II)	Human	Chou, JJ	Bax, A
1FW4	2001	C-terminal	4.6	X-Ray	1.7	Ca(II)	Bovine	Olson, LL	Sjolin, L
1F55	2003	N-terminal	6.8	NMR	30	Ca(II)	Yeast	Ishida, H	Yazawa, m
1PRW	2003	Full	5.4	X-Ray	1.7	Ca(II)	Bovine	Fallon, JL	Quiococho, FA
1NOY	2003	Full	5.0	X-Ray	1.75	Pb(II)	Ciliate	Wilson, MA	Brunger, AT
1RFJ	2004	Full	3.9	X-Ray	2	Ca(II)	Potato	Yun, CH	Liang, DC
1UP5	2005	Full	4.0	X-Ray	1.9	Ca(II)	Chicken	Wilson, MA	Rupp, B

Table 1.2 Interhelical Angle in apo-CaM and Ca(II)-CaM

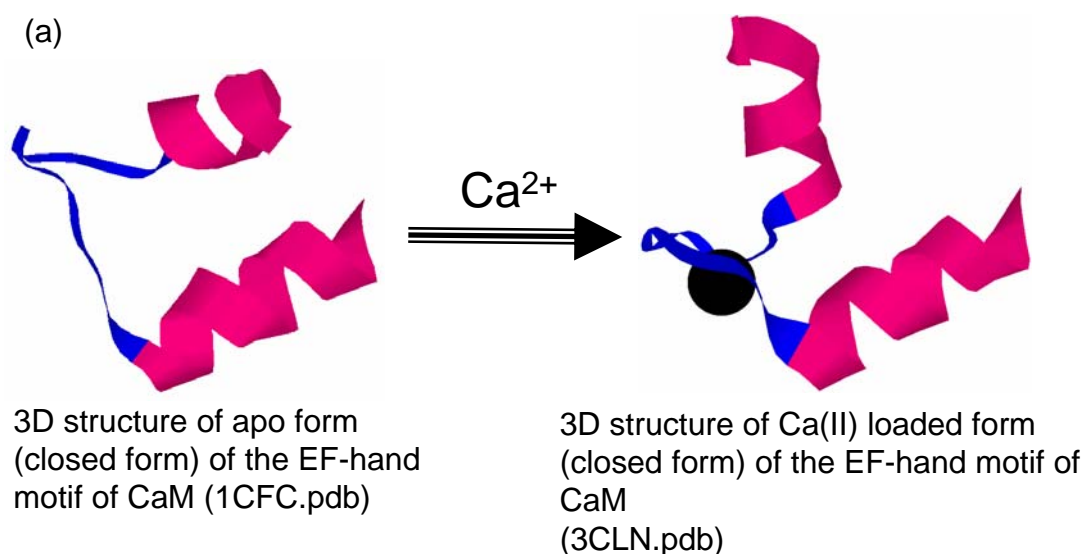
Helix	Apo 1CFD	Apo 1DMO	Apo 1CMF	Ca(II) 3CLN	87
A/B	138 <sup>+</sup> <sup>+</sup> <sup>+</sup> $\pm 2$		128 $\pm 3$		
A/C	88 $\pm 3$			160	
A/D	127 $\pm 2$	121 $\pm 2$		110	
B/C	126 $\pm 6$	130 $\pm 4$		113	
B/D	47 $\pm 2$			45	
C/D	130 $\pm 3$	130 $\pm 4$		84	
E/F	131 $\pm 4$	137 $\pm 3$	131 $\pm 4$	105	
E/G	81 $\pm 5$		108 $\pm 4$	142	
E/H	142 $\pm 5$	144 $\pm 3$	111 $\pm 4$	119	
F/G	141 $\pm 5$	144 $\pm 3$	117 $\pm 3$	113	
F/H	30 $\pm 5$		49 $\pm 3$	37	
G/H	133 $\pm 4$	132 $\pm 5$	131 $\pm 4$	96	

1CFD, Kuboniwa et al., *Nature Structural Biology* (1995), 2, 768

1DMO, Zhang et al., *Nature Structural Biology* (1995), 2, 758

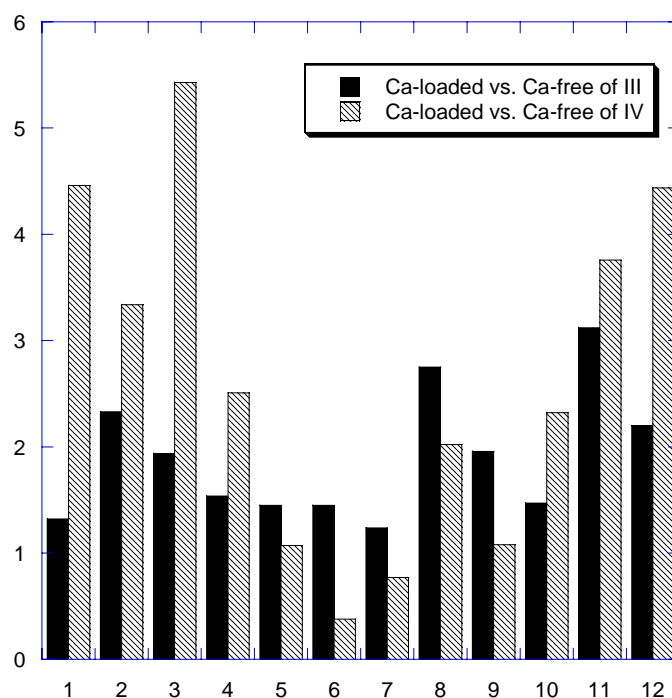
1CMF, Finn et al., *Nature Structural Biology* (1995), 2, 777

3CLN, Babu et al., *Journal of Molecular Biology* (1988), 204, 191.



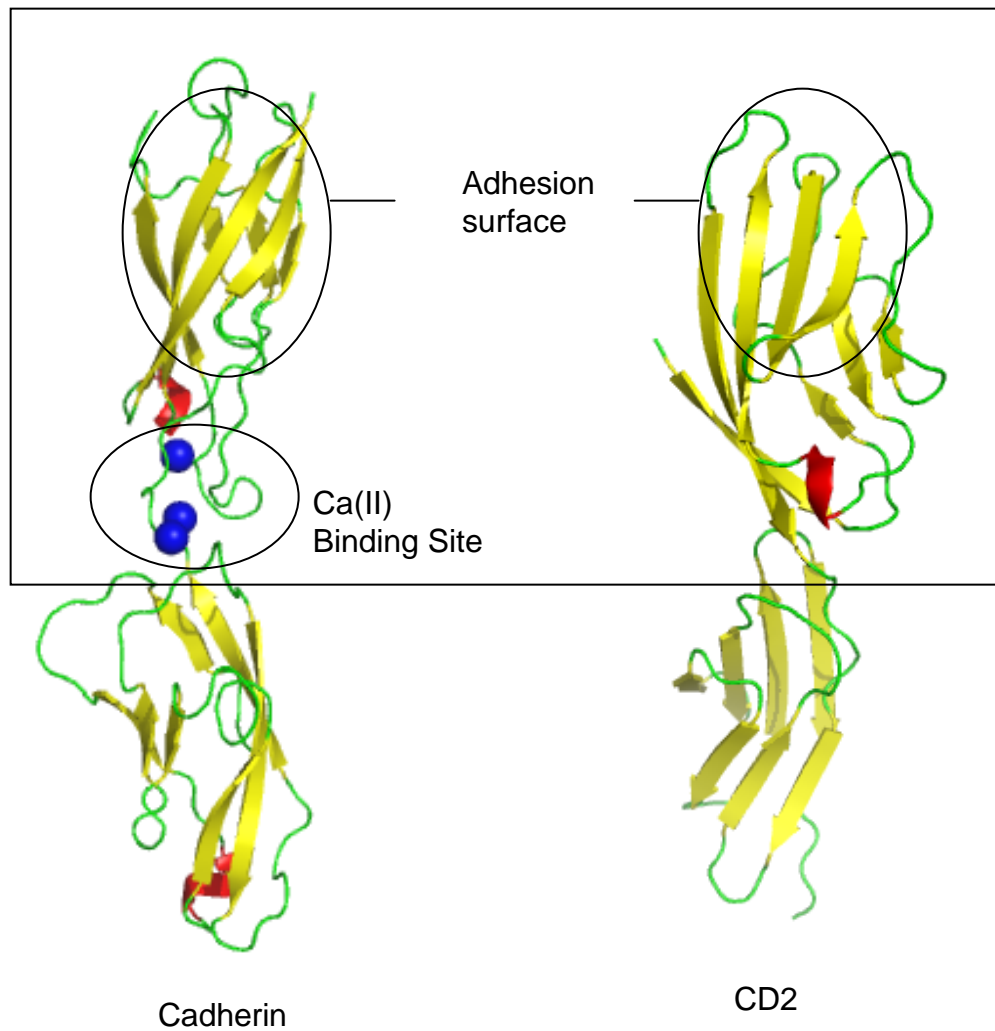
(b) Loop III: **D-K-D-G-N-G-Y-I-S-A-A-E**

Loop IV: **D-I-D-G-D-G-Q-V-N-Y-E-E**



**Figure 1.5 Structural difference of an EF-hand motif in the presence and absence of Ca(II):** (a) The EF-hand motif undergoes conformational change upon binding with calcium. (b) The RMS deviation between the Ca(II) free form (1CFC) and Ca(II) loaded form (3CLN) of C-terminal domain of CaM. The EF-loop IV has large structural deviation between the Ca(II) free and Ca(II) loaded form in comparison with EF-loop III.





**Figure 1.6 Structures of cell adhesion molecules: CD2** (1HNG.pdb) protein has a cell adhesion surface similar to the calcium binding protein cadherin (1EDH.pdb). These two proteins share similar cell adhesion surface.

## 2.0 Methods and Materials

### 2.1 Protein Engineering and Purification

The CD2 engineered proteins with metal binding site insertions were engineered as described in previous work by Ye *et al.* (112). The homonuclear,  $^{15}\text{N}$ -labeled, and  $^{13}\text{C}$ - $^{15}\text{N}$ -labeled proteins were expressed in SV medium (45.6 mM  $\text{K}_2\text{HPO}_4$ , 32.4 mM  $\text{KH}_2\text{PO}_4$ , 0.2 mM  $\text{MgSO}_4$ , and 0.018 mM  $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2$ ) with 0.5 g of ammonium chloride ( $^{14}\text{NH}_4\text{Cl}$  for homonuclear and  $^{15}\text{NH}_4\text{Cl}$  for  $^{15}\text{N}$ -labeled) and 5 g of glucose ( $^{12}\text{C}$  glucose for homonuclear and  $^{13}\text{C}$  glucose for  $^{13}\text{C}$ -labeled) per liter of medium.

#### 2.1.1 Protein Expression

1. The CD2 variant DNA was transformed using BL-21 competent cells and was plated in LB-amp (LB: Luria and Bertani) with 0.3 % glucose.
2. Multiple colonies were inoculated into 2 mL of SV media using the 15 mL culture tubes for 8 h. A minimum of 10 tubes were prepared to obtain 4 liters of expressed protein.. To the 2 mL SV medium were added 2  $\mu\text{L}$  100 mg/mL ampicillin, 50  $\mu\text{L}$  of 20 % glucose and 5  $\mu\text{L}$  of 20 % ammonium chloride.
3. Culture tubes with best observed growth were selected and transferred to 100 mL of medium (2 tubes for each 100 mL) for overnight growth before transferring to 1 L of medium for expression. The 100 mL of medium was placed in a 500 mL flask for overnight growth. Next, 100  $\mu\text{L}$  of 100 mg/mL

ampicillin, 2.5 mL of 20 % glucose and 250  $\mu$ L of 20 % ammonium chloride were added to each 100 mL of SV medium.

4. The overnight culture was then inoculated into 1 L SV medium. The 1 L culture was incubated at 37 °C with agitation until the OD<sub>600</sub> reached 0.6.

The following three items were added to each 1 L SV medium:

1 mL of 100 mg/mL ampicillin

25 mL of 20 % glucose

5  $\mu$ L of 20 % ammonium chloride

5. Once an OD<sub>600</sub> of 0.6 was reached (usually within 2.5 to 3.0 hours), protein expression was induced with IPTG (isopropyl- $\beta$ -thiogalactoside, 0.15 mM, 150  $\mu$ L of 1 M IPTG). An additional 4 h expression was allowed before harvest.
6. The cell pellet was centrifuged at 7 k rpm for 20 minutes. The supernatant was discarded and the cell pellet was then stored at -20 °C.

### **2.1.2 Protein Purification**

1. The cell pellet was defrosted and suspended in lysate buffer (30 mL of lysate buffer per each liter growth). DTT was added to the pellet mixture to reach a final DTT concentration of 5 mM. EDTA was added to the pellet mixture to reach final EDTA concentration of 10 mM. The mixture was blended for 30 seconds.

Lysate buffer, Preparation for 250 mL:

1% sarcosyl (N-lauroyl sarcosine)	2.5 g
10 mM DTT	0.385 g
1 mM EDTA	0.10 g
5 $\mu$ M inhibitor (AEBSF)	2.5 mL

Fill to 250 mL with 1x PBS pH 7.3

1x PBS, Preparation for 1 L:

NaCl	8.1750 g
KCl	0.2009 g
Na <sub>2</sub> HPO <sub>4</sub> *7H <sub>2</sub> O	2.7075 g
KH <sub>2</sub> PO <sub>4</sub>	0.2448 g

Add 1 L ddH<sub>2</sub>O

Verified pH at or near 7.3

2. The blended mixture was separated into multiple small plastic beakers (< 25 mL per beaker) and each beaker was sonicated 6 times. The duration of each sonication cycle is 10 seconds with a 10 minute interval between sonication cycles.
3. Once the sonication cycles were complete, the content was centrifuged for 30 minutes at 17 k rpm. The supernatant was filtered using 0.45  $\mu$ m syringe filter while the remaining pellet was re-suspended in lysate buffer to repeat purification steps 1 to 3.
4. After cell lysis, the fusion proteins in the supernatant were loaded onto a column of GS-4B resin (Pharmacia). After binding, each column was washed with 50 mL of 1x PBS buffer. The fusion proteins were cleaved by

thrombin (2.5 mL of 1x PBS buffer with 30  $\mu$ L of thrombin (1 unit/ $\mu$ L)) on the beads following elution of waste materials.

5. The column was first kept at 4 °C with agitation for 14 hours, then placed at room temperature for two hours before elution. The proteins were eluted using 20 mL of 1x PBS buffer per column.
6. The eluted proteins were further purified using a superdex 75 column (Pharmacia) and it was eluted out of FPLC using 10 mM Tris buffer at pH 7.4. The size exclusion purification step is followed by a cationic exchange column (Hitrap SP Sepharose). The protein was bound to the cationic exchange column using 20 mM acetate buffer at pH 3.5. The protein was eluted with an increasing pH gradient from 3.5 to 8.0 (50 mM Tris buffer).
7. The identities of CD2 variants were confirmed by SDS–PAGE and mass spectrometry. The protein concentration was measured with  $\epsilon_{280} = 11700 \text{ M}^{-1} \text{ cm}^{-1}$  for CD2 (98).

$\text{Ca}^{2+}$ -free proteins were prepared by separating the EGTA (ethylene glycol-bis( $\beta$ -aminoethylether)-*N,N,N',N'*-tetraacetic acid) from a protein–EGTA mixture using an SP column with a pH gradient from 4 to 8. All solution for metal-binding studies was pre-treated on a Chelex-100 (Bio-Rad) column.

### 2.1.3 Protein Stability

Initially, the CD2 variants were first purified using the GST affinity column. The GST-tag was cleaved using the thrombin protease. Once the cleavage process was completed, the GST protein and the CD2 variant was eluted using the elution buffer (20 mM Glutathione elution buffer). The protein mixtures were separated using the size exclusion column. The size exclusion column was able to separate the GST protein and the CD2 variant, but could not completely remove the residual thrombin protease from the CD2 variant. The viability of the CD2 variant was very short (less than three days), especially for the high concentration NMR sample.

To improve the protein stability during the NMR experiment, the following aspects of protein sample preparation were improved:

First, the GS4 resin used in the affinity column was cleaned using 20 mM glutathione elution buffer at pH 12.0. The GS4 resin was suspended in 5 mL of elution buffer for more than 24 hours with agitation. Each GST affinity column was then washed with 50 mL of elution buffer (this step is very important, if there are residual protein left in the column, as soon as the PBS buffer was poured into the column, the old GST protein would bind to the column again and will not be eluted out. It is better to wash the column with more elution buffer). Then the column was washed with 50 mL of 1X PBS. Second, once the thrombin cleavage process was completed; only the CD2 variant will be eluted from the column. The GST protein remains attached to the GS4 beads. This step

reduces problems with the GST protein mixing with the CD2 variant. Third, after the size exclusion column purification step, the CD2 variant is further purified using an anionic exchange column (SP). The CD2 variant was bound to an SP column in low pH. The protein was then eluted using a pH gradient (pH 3.5 to 8.0). Fourth, CD2 variants were dialyzed into the NMR experiment buffer at low protein concentration, which reduced the potential for protein precipitation and extended the sample shelf life. Next, the protein concentration was increased using the Amicon stirred ultrafiltration cell (mw cut off for the membrane is 3500 KDa) instead of the centrifuged filter devices. The concentration process can usually be completed within 1 hour, which can also extend the sample shelf life. Finally, two kinds of inhibitors, Hirudin and Sigma Protease Cocktail inhibitors, were added to the NMR sample. The improvements discussed in this section have already been incorporated into the purification procedure in Section 2.12.

## **2.2 Metal Titration**

### **2.2.1 1D $^1\text{H}$ NMR Titration**

For metal titration of the CD2 variants, NMR samples were prepared by diluting proteins in 10 mM Tris-HCl with 10% D<sub>2</sub>O at pH 7.4. Protein concentrations were varied from 150 to 300  $\mu\text{M}$ . Dioxane was used as an internal reference for the NMR spectra (3.743 ppm). All NMR spectra were recorded using Varian Inova 500 MHz and 600 MHz NMR spectrometers. Spectra widths of 6600 Hz and 8000 Hz were used at 500 MHz and 600 MHz,

respectively. A water suppression pulse sequence from Varian Biopack was used with 8 K complex data points at 25 °C. The stock (10-40 µl of 1 mM, 10 mM, and 100 mM metal ion stock solutions at pH 7.4) solutions were gradually injected into the NMR sample tube, and mixed thoroughly. A 30 minute equilibrium time was allowed for each titration point. All the 1D NMR experiments were collected with 1024 scans.

The data were processed with the program FELIX98 (MSI). After Fourier transformation, typically a squared sinebell window function shifted over 75° was used. Post acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with a function width of 20.

### **2.2.2 2D $^1\text{H}$ - $^{15}\text{N}$ HSQC Titration**

For the 2D metal titration of CD2 variants, NMR samples were prepared by diluting proteins in 20 mM PIPES and 20 mM KCl with 10% D<sub>2</sub>O at pH 6.8. Protein concentrations were varied from 250 to 400 µM. All  $^1\text{H}$ - $^{15}\text{N}$  HSQC spectra were recorded using Varian Inova 500 and 600 MHz spectrometers. Spectra widths of 6600 Hz for  $^1\text{H}$  and 1800 Hz for  $^{15}\text{N}$  were used at 500 MHz. Spectra widths of 8000 Hz for  $^1\text{H}$  and 2200 Hz for  $^{15}\text{N}$  were used at 600 MHz. The "gNhsqc" pulse sequence from Biopack was used for 2D NMR with 1 K complex data points for  $^1\text{H}$  and 64 complex data points for  $^{15}\text{N}$  at 25 °C. 50 µM EGTA was added to the protein first. The metal stock solutions (10 mM and 100 mM) were gradually injected into the NMR sample tube and mixed thoroughly. A



30 minute equilibrium time was allowed for each titration point. All spectra were collected with 16 scans for each FID.

The 2D HSQC data were processed with the program FELIX98 with sensitivity enhance option (MSI). After Fourier transformation, typically a squared sinebell window function shifted over 75° was used for both dimensions. Post-acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with a function width of 40. The assignment was performed with Sparky (113).

### 2.2.3 Calculation of $K_d$

Calcium-binding affinities of proteins were calculated using data obtained from calcium titrations by NMR. The chemical shift values of the resolved peaks in the absence of metal ions were used as  $S_0$  with 0% metal-binding and the chemical shift values of the same peaks in 10 mM  $\text{CaCl}_2$  or  $\text{LaCl}_3$  were used as  $S_{100}$  with 100% metal binding. The fractional change values,  $f$ , at different metal concentrations were calculated using the equation  $f = (S - S_0) / (S_{100} - S_0)$ .

$K_d$  values of the protein were calculated by fitting the titration curves from NMR chemical shift changes using the following equation:

$$F = \frac{([P]_T + [M]_T + K_d) - \sqrt{([P]_T + [M]_T + K_d)^2 - 4[P]_T[M]_T}}{2[P]_T} \quad \text{equation 2.1}$$

where  $[M]_T$  and  $[P]_T$  are the total concentration of  $\text{Ca(II)}$  and protein, respectively (114, 115).

## **2.3 Sequential Assignment**

### **2.3.1 2D NMR Experiments with Homonuclear Samples**

NMR samples were prepared by diluting proteins in 10 mM Tris-HCl, 10% D<sub>2</sub>O at pH 6.8 or 7.4. Protein concentrations were varied from 0.8 to 1.5 mM. Dioxane was used as an internal reference for the NMR spectra (3.743 ppm). All NMR spectra were recorded using Varian Inova 500 MHz and 600 MHz NMR spectrometers. Spectral widths of 6600 and 8000 Hz in each dimension were used at 500 and 600 MHz, respectively. WG-TOCSY (mixing time, 36-90 ms) and WG-NOESY (mixing time, 100-200 ms) spectra were collected at 25 °C. In the t<sub>2</sub> dimension 1 k complex data points were collected and 400 t<sub>1</sub> increments and 128 scans were collected for each t<sub>1</sub> increment were used.

The data were processed with the program FELIX98 (MSI). After Fourier transformation, typically a squared sinebell window function shifted over 75° was used for both dimensions. For 2D experiments, the data were zero-filled to yield 2K x 2K (t<sub>2</sub>,t<sub>1</sub>) data matrices. Post acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with a function width of 20.

### **2.3.2 Aromatic Ring Assignment in D<sub>2</sub>O Condition**

For the assignment of aromatic ring protons and the sidechain of Asn and Gln, one common practice is to substitute the exchangeable proton in the protein

by deuterium. The substituted proton would no longer be detected in the TOCSY and NOESY spectra. The protons that are attached to carbon are the non-exchangable protons while the protons that are attached to nitrogen are generally exchangeable protons. By dissolving the protein in 100% deuterium solution, sidechain  $\text{-NH}_2$  hydrogens of Asn and Gln would no longer be observed whereas the resonances of the aromatic ring remain.

***The Exchanging Process:***

1. The protein solution was diluted 10 x using an identical buffer solution made in  $\text{D}_2\text{O}$  and placed at 37 °C for 30 minutes.
2. The protein must be frozen below -80 °C before lyophilizing. The protein solution was frozen at -80 °C for 30 minutes to ensure that the solution will not melt during lyophilization.
3. The frozen sample was then lyophilized until the protein sample became a powder (usually 24 hours).
4. The protein powder was resuspended in  $\text{D}_2\text{O}$  solution. The exchange process steps 1 to 3 were repeated two more times.

NMR samples were prepared by diluting proteins in  $\text{D}_2\text{O}$ , and the protein concentrations were varied from 0.8 to 1.0 mM. All NMR spectra were recorded using a Varian Inova 600 MHz NMR spectrometer. A spectral width of 8000 Hz was used at 600 MHz. WG-TOCSY (mixing time, 60 ms) and WG-NOESY (mixing time, 100 ms) spectra were collected at 25 °C. In the  $t_2$  dimension, 1 k

complex data points were collected. In  $t_1$ , 400 increments were collected with 128 scans per increment.

The data were processed with the program FELIX98 (MSI). After Fourier transformation, typically a squared sinebell window function shifted over  $75^\circ$  was used for both  $t_1$  and  $t_2$  dimensions. For 2D experiments, the data were zero-filled to yield  $2K \times 2K$  ( $t_2, t_1$ ) data matrices. Post acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with a function width of 60.

### **2.3.3 3D TOCSYHSQC and NOESYHSQC**

The 3D  $^{15}\text{N}$  TOCSYHSQC and NOESYHSQC spectra were collected using an 800 MHz NMR equipped with a cold probe. A  $^{15}\text{N}$ -labelled CaM-CD2-III-5G sample at a concentration of 600  $\mu\text{M}$  with 2 mM EGTA in 20 mM PIPES, 20 mM KCl at pH 6.8, was used for the experiment. The TOCSYHSQC spectrum was collected using the gtocsyNhsqc.c (Biopack pulse sequence) at  $25^\circ\text{C}$  with an isotropic mixing time of 60 ms. The NOESYHSQC spectrum was collected using the gnoesyNhsqc.c (Biopack pulse sequence) at  $25^\circ\text{C}$  with an isotropic mixing time of 120 ms. Both spectra were collected with 2048 complex points ( $np=4096$ ) in  $D_1$ , 128 complex points in  $D_2$ , and 24 complex points in  $D_3$ .

These spectra were processed using FELIX98 with SINBELL square window function at a  $72^\circ$  shift.  $D_2$  and  $D_3$  dimension were processed with linear prediction. The first dimension was zero filled to 4096 points,  $D_2$  was zero filled

to 256 points, and D3 was zero filled to 128 points. All of the sequential assignments of 3D  $^{15}\text{N}$  TOCSYHSQC and NOESYHSQC were verified with the assignment from the triple resonances experiments. Because this experiment is  $^{15}\text{N}$ -edited, all cross peaks are separated according to the correlated nitrogen chemical shift in the third dimension.

### **2.3.4 3D Triple Resonance Experiments**

NMR samples were prepared by diluting proteins in 20 mM PIPES and 20 mM KCl with 10%  $\text{D}_2\text{O}$  at pH 6.8. Protein concentrations for the  $^{13}\text{C}$ - $^{15}\text{N}$  double-labeled samples were varied from 0.6 to 0.8 mM. All of the spectra used for assignment were collected in the absence of calcium at 25 °C. All of the 3D experiments for CaM-CD2-III-5G were collected using 24 complex points in the  $^{15}\text{N}$  dimension. The  $^{13}\text{C}$  dimensions of triple resonance experiments were collected with 47 complex points for constant time experiments and the rest were collected with between 64 and 128 complex points depending on available instrument time. The 3D experiments were collected using two different 600 MHz spectrometers.

The 2D HSQC data were processed with the program FELIX98 with sensitivity enhance option (MSI). After Fourier transformation, typically a squared sinebell window function shifted over 75° was used. Post acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with a function width of 40.

The 3D triple resonance data were processed with the program FELIX98 and NMRpipe with the sensitivity enhance option. The 3D process macro of FELIX98 was used to process the 3D data. Each dimension is processed in the order of D1 → D2 → D3. D1 is the  $^1\text{H}$  dimension. D2 is the  $^{13}\text{C}$  dimension. D3 is the  $^{15}\text{N}$  dimension and the squared sinebell window function shifted over  $75^\circ$  was used. Post acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with function width of 40. The assignment was performed with Sparky (113).

## **2.4 Backbone Dihedral Angles**

### **2.4.1 $J_{\text{HNHA}}$ Coupling Constant**

The NMR sample was prepared by diluting proteins in 10 mM TRIS and 10 mM KCl with 10%  $\text{D}_2\text{O}$  at pH 7.4. The protein concentration for the  $^{15}\text{N}$ -labeled sample was 0.88 mM. The HNHA experiment was collected in the absence of calcium at  $25^\circ\text{C}$ . The HNHA data was collected using the protein pack pulse sequence, ghnha.c, with 2 k complex points on D1, 72 complex points in D2, and 30 complex points in D3 at 800 MHz. Spectra widths of 10703 and 2200 Hz were used for  $^1\text{H}$  and  $^{15}\text{N}$  dimensions, respectively.

The HNHA data was processed with the program FELIX98 with the sensitivity enhance option. The 3D process macro of FELIX98 was used to process the 3D data. Each dimension is processed in the order of D1 → D2 → D3. D1 is the  $^1\text{H}$  dimension. D2 is the  $^1\text{H}$  dimension and D3 is the  $^{15}\text{N}$

dimension. The squared sinebell window function shifted over  $75^\circ$  was used.

Post acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with function width of 40. The assignment was performed with Sparky (113).

#### **2.4.2.1 Dihedral angle prediction using TALOS**

The  $\phi$  and  $\psi$  dihedral angles for CaM-CD2-III-5G were predicted using the TALOS program. The TALOS program is part of the NMRPipe distribution from NIH (116, 117). The chemical shift used for TALOS calculations were obtained from the following spectra;  $^1\text{H}$ - $^1\text{H}$  TOCSY with different mixing times,  $^1\text{H}$ - $^1\text{H}$  NOESY with different mixing times,  $^{15}\text{N}$  TOCSYHSQC,  $^{15}\text{N}$  NOESYHSQC, HNCO, HNCA, HNCACB, and CBCA(CO)NH. The chemical shift table was generated using the SPARKY program (113). TALOS will use  $\text{H}\alpha$ ,  $\text{C}\alpha$ ,  $\text{C}\beta$ , CO, and N chemical shifts and ignore the rest of the chemical shifts. The chemical shift table was modified to the TALOS input style using the KaleidaGraph program (Synergy Software). A total of 95, 109, 95, 98, and 104 chemical shifts of  $\text{H}\alpha$ ,  $^{13}\text{C}\alpha$ ,  $^{13}\text{C}\beta$ ,  $^{13}\text{CO}$ , and  $^{15}\text{N}$  were used for prediction, respectively.

#### **2.4.2.2 Predicting Dihedral Angles for Wild Type CD2**

The chemical shifts of the wild type CD2 were obtained from the Biological Magnetic Resonance Data Bank (<http://www.bmrb.wisc.edu>). The accession number for CD2 is 4109 and was deposited in 1998 by Chen and coworkers (118). Previous work in our laboratory involving temperature dependent studies

of the far UV CD spectra suggested that the secondary and tertiary structure of the wild type CD2 and CaM-CD2-III-5G are stable from pH 2.0 to pH 10.0. The chemical shifts for wild type CD2 at pH 5.0 and 7.0 were very similar, but at pH 5.0 the exchange rate is slower, so there are more assignments under the pH 5.0 condition. The chemical shifts of the wild type CD2 at pH 5.0 and 7.0 were used for prediction. For CD2 at pH 5.0, only 56 residues of the predictions are classified as good. For CD2 at pH 7.0, 59 residues of the predictions are classified as good.

## **2.5 Residual Dipolar Coupling**

### **2.5.1 Residual Dipolar Coupling Using an External Alignment Medium**

NMR samples were prepared by diluting proteins in 20 mM PIPES and 20 mM KCl with 10% D<sub>2</sub>O at pH 6.8. The protein concentration for the <sup>15</sup>N-labeled sample was 0.25 mM. The residual dipolar couplings for CaM-CD2-III-5G were obtained by aligning the protein using PEG (C12E5, pentaethylene glycol monododecyl ether) bicelle medium. 10 uL of C12E5 was first mixed with 30 uL of D<sub>2</sub>O and 270 uL of protein and the mixture was vortexed. 3.6 uL of hexanol was added in 1.8 uL increments to achieve a final concentration of 3.13 % (w/v) PEG. The <sup>2</sup>H splitting was verified using the s2pul pulse sequence with channel 1 set to "tn=1k". The <sup>2</sup>H splittings were verified before the IPAP-HSQC and after IPAPHSQC. Both were observed to be 15.4 Hz. The residual dipolar couplings were measured in the absence of calcium at 25 °C. The residual dipolar coupling



experiment was collected using protein pack pulse sequence IPAPHSQC with the following setting, IPAP='n','y' phase=1,2 array='IPAP,phase'. The experiment was collected with 2k complex points in D1 and 256 complex points in D2. Spectra widths of 9852 and 2344 Hz were used for  $^1\text{H}$  and  $^{15}\text{N}$  dimensions, respectively. Both the isotropic and aligned experiments were performed using a Varian INOVA 800 MHz spectrometer at UGA.

The IPAP-HSQC results were processed using NMRPipe (117). D2 dimension was processed with linear prediction. The first dimension was zero filled to 2048 points and D2 was zero filled to 2048 points. All of the assignments from this spectrum were verified with 2D  $^{15}\text{N}$  HSQC. The NMRPipe macros used to process the IPAP-HSQC spectrum are shown in Appendix 2.1.

To obtain the Euler angles, the principle order parameter ( $S_{zz}$ ), and the asymmetry parameter  $\eta$  [ $\eta = (S_{xx} - S_{yy})/S_{zz}$ ], the structure constructed by CYANA was used to calculate the order parameter using the REDCAT program (119-121). For the REDCAT calculation, the number of error space sampling was set at 10000 and residue with estimated error of greater than 5.50 Hz was discarded. After the error correction, a total of 57 residual dipolar couplings were used. The anisotropy parameter “Da” was calculated to be 6.77 [(Dmax x  $S_{zz}$ )/2] (120). The rhombicity parameter “R” [(2/3) $\eta$ ] was calculated to be 0.469.

### 2.5.2 Field Induced Residual Dipolar Coupling

All NMR samples were prepared by diluting proteins in 20 mM PIPES and 20 mM KCl with 10% D<sub>2</sub>O at pH 6.8. The protein concentration for the <sup>15</sup>N-labeled sample was 450 μM. The diamagnetic or paramagnetic metal ions concentration was 350 μM. The final metal ion to protein ratio is 0.88. The NMR sample content is listed below:

- 300 μL of 450 μM protein sample
- 30 μL of D<sub>2</sub>O
- 5 μL of Sigma cocktail inhibitors
- 5 μL of metal ion solution

The internal alignment method was performed using the lanthanide family metal ions. The RDC experiments were collected using 600 MHz and 800 MHz instruments. The following metal ions were used for RDC experiments, they are La(III), Tm(III), and Dy(III). The experiments were collected using the “gNhsqc\_IPAP.c” pulse sequence. The residual dipolar coupling experiment was collected using protein pack pulse sequence IPAPHSQC with the following setting, IPAP='n','y' phase=1,2 array='IPAP,phase'. The experiment was collected with 2k complex points in D1 and 256 complex points in D2. Spectra widths of 9852 and 2344 Hz were used for <sup>1</sup>H and <sup>15</sup>N dimensions at 800 MHz, respectively. Spectra widths of 8384 and 2200 Hz were used for <sup>1</sup>H and <sup>15</sup>N dimensions at 600 MHz, respectively.

The IPAP-HSQC results were processed using NMRPipe (117). D2 dimension was processed with linear prediction. The first dimension was zero

filled to 2048 points and D2 was zero filled to 2048 points. All of the assignments from this spectrum were verified with 2D  $^{15}\text{N}$  HSQC.

The Dy(III) paramagnetic experiment was performed using 800 MHz NMR. The diamagnetic contribution from metal binding was subtracted by comparing the PCS and RDC to the La(III) diamagnetic experiment. The Tm(III) paramagnetic experiment was performed using 600 MHz NMR. The diamagnetic contribution from metal binding was subtracted by comparing the PCS and RDC to the La(III) diamagnetic experiment.

## **2.6 Structural Calculation**

### **2.6.1 Structural Calculation Using CYANA**

The CYANA software was purchased from Dr. Guntert (122). The calculations were performed using computers running REDHAT 9.0 and OS X operating systems. For the manual mode, the NOE distance restraint list and the dihedral angle restraint list were generated manually using the standard UNIX editor. Since CYANA can directly read in the CNS style NOE distance restraint list, the list was constructed using the CNS style format so the same restraint data can be used by both programs.

Initially, the structural calculations were performed using CYANA version 1.5. The later calculations (including the completed CaM-CD2-III-5G structure) were performed using CYANA version 2.1. During the optimization stage for the NOE distance and dihedral angle restraints, 100 structures (5000 steps) were

generated using simulated annealing procedure. The NOE distance restraint corrections were based on NOE violations in the 20 lowest energy structures. For the final structure, 100 structures (10000 steps) were generated using simulated annealing procedure. The 20 lowest energy structures reported 2 NOE distance violations for all structures. The calculation scripts (CALC.cya and calc\_all.cya) and simulated annealing script (anneal.cya) are shown in Appendix 2.2.

## **2.6.2 Structural Calculation Using CNS**

The CNS (version 1.1) software was obtained from CNS website (<http://cns.csb.yale.edu/>) (123). Calculations were performed using computers running REDHAT 9.0 and OS X operating systems. The NOE distance restraints and dihedral angle restraints were transferred from CYANA calculations. All of the CNS calculations were performed using the simulated annealing procedure, and 100 structures were generated for each calculation. The 20 lowest energy structures were collected for additional analysis. The NOE distance restraints table, dihedral angle restraints table, orientation restraints (dipolar couplings), and simulated annealing script (anneal.inp) are shown in Appendix 2.3.

## **2.7 Dynamic Studies**

### **2.7.1 HX Sample Preparation and NMR Experimental Parameters**

The HX experiments were carried out using the Varian Inova 600 MHz housed at Georgia State University. A spectra width of 8000 Hz was used for all experiments. The watergatek pulse sequence was used for the 1D NMR experiments. The wgtocsy pulse sequence was used for the 2D TOCSY NMR experiments. All of the HX experiments were carried at 25 °C. The processes for the HX experiment are summarized in the following scheme:

***Protein Sample ( $H_2O$ ) → Lyophilization → Saturation in  $D_2O$  → Data Collection***

For the HX experiments, the protein concentrations were varied from 120 to 800  $\mu$ M. The protein samples were first diluted into the experimental concentrations using 10 mM Tris-HCl at pH 7.4, and then the protein samples were frozen at -80 °C for 30 minutes in either a cold box, or using the dry ice and acetone mixture. Previous work in our laboratory has produced better results using the cold box freezing method. The frozen protein sample was lyophilized for a minimum of 30 hours. The HX process occurs as soon as the protein sample is dissolved in the  $D_2O$  solution. In order to acquire the spectrum immediately, the probe, lock, and shim tuning processes on the NMR instrument were initially performed using a  $D_2O$  sample with the same volume as the HX sample. The NMR sample composition was comprised of 500  $\mu$ L of  $D_2O$ , 5  $\mu$ L of Dioxane and the Lyophilized Protein Sample.

The HX experiments for wild type CD2, CaM-CD2-III-5G, CaM-CD2-IV-5G were all conducted under the same conditions. A total of 43 spectra were collected for each set of HX experiments. The time between sample dissolution

in D<sub>2</sub>O and initiation of data acquisition was approximately 120 seconds. For each spectrum, 256 scans were used to collect the data. Different pre-acquisition delays were used before acquiring each of the 43 spectra (0.5, 300, 300, 300, 300, 300, 300, 300, 300, 300, 300, 600, 600, 600, 600, 600, 600, 600, 600, 600, 600, 1200, 1200, 1200, 1200, 1200, 1200, 1200, 1200, 1200, 1200, 1200, 2400, 2400, 2400, 2400, 2400, 2400, 2400, 2400, 2400, 2400, 3600, 3600 s). Each set of HX experiments took over 20 hours.

Data were processed with the program FELIX98 (MSI). The Varian instrument organized the 43 arrayed spectra onto the D1 axis. After Fourier transformation, typically a squared sinebell window function shifted over 75° was used. Post acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with a function width of 60.

The peak intensities of the spectra changed as a function of the exchanging process. The spectra were baseline-corrected using the following protocol (Figure 2.1). First, three areas (*a*, *b*, and *c*) were selected. Area *c* is the data that were used to calculate the HX rate. Areas *a*, *b*, and *c* were obtained for every spectra. Areas *a* and *b* were used for baseline correction with respect to area *c*. The area integration was performed using a macro for FELIX98 (Figure 2.1).

---

```

inter2.mac
def matpfx /pie/chehwlx/nmr/matrix/
def datpfx /pie/chehwlx/nmr/data/
get 'first value?' fv2
get 'last value?' lv2

```

```

get 'which region?' region
get 'name of matrix?' hsmatnam
mat &hsmatnam
cl
def int_value 0
set 0
set 1
for inte &fv2 &lv2
loa &inte 0
adb 1
nex
ldb 1
wr &region
cmx
end

```

---

Next the baseline correction using Equation 2.2 was performed using KaleidaGraph software.

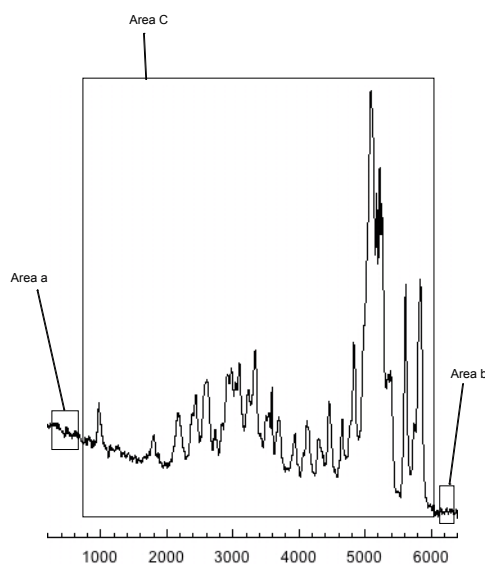


Figure 2.1 Example of Baseline correction

$$c' = c - \left( \frac{a}{n} + \frac{b}{m} \right) / 2 * L$$

Equation 2.2

In Equation 2.2,  $c'$  is the corrected area and  $c$  is area before correction.  $n$ ,  $m$  and  $L$  are the points used for integration of areas  $a$ ,  $b$ , and  $c$ , respectively. The resulting area  $c'$  was fitted using a double exponential decay equation as a function of time.

## 2.7.2 Relaxation Studies on CaM-CD2-III-5G and CaM-CD2-IV-5G

### NMR Experimental Parameters and Sample Preparations for the Relaxations

The protein concentrations of CaM-CD2-III-5G and CaM-CD2-IV-5G for the relaxation experiments were varied from 250  $\mu$ M to 400  $\mu$ M. The NMR samples were dialyzed against the 20 mM PIPES and 10 mM KCl buffer at pH 6.8 for 3 hours, during which time the buffer was changed 4 times. The T1, T2, and NOE relaxation experiments were carried out at 25 °C. The relaxation experiments were performed using the Varian Inova 600 MHz NMR housed at the University of Georgia (courtesy of Dr. Urbauer) and Georgia State University. The T1, T2, and NOE relaxation experiments are field strength dependent, so all of the experiments were performed using the 600 MHz NMR. The NMR sample content and experimental parameters are summarized below:

#### Sample Content (T1 and T2)

- 300  $\mu$ L of Protein (CaM-CD2-III-5G or CaM-CD2-IV-5G)
- 3.5  $\mu$ L of either 100 mM EGTA or 4  $\mu$ L of 100 mM Ca(II) stock solution
- 30  $\mu$ L of D<sub>2</sub>O
- 3.5  $\mu$ L of Sigma Cocktail Inhibitors

#### Longitudinal Relaxation, T1

The T1 spectra were collected using the gNhsqc.c pulse sequence (set T1='y', T2='n', and T1rho='n', Biopack pulse sequence) with a relaxation delay of 0, 10, 60, 130, 230, 340, 480, 740, 1000, 1500 ms. Each experiment was carried



out in an individual experiment instead of arrayed into one experiment. A T1 experiment with 130 ms relaxation delay was repeated after all of the T1 spectra were collected to ensure that there were no changes in the sample or the instrument during the experiments.

T1 and T2 relaxation times for CaM-CD2-III-5G and CaM-CD2-IV-5G were calculated using the following equation

$$I_t = I_0 \exp(-t/T_i) \quad \text{Equation 2.3}$$

where  $I_t$  is the observed magnetization at time  $t$ ,  $I_0$  is the initial magnetization, and  $T_i$  is the relaxation time.

### Transverse Relaxation, T2

The T2 spectra were collected using the gNhsqc.c pulse sequence (set T1='n', T2='y', and T1rho='n', Biopack pulse sequence) with relaxation delays of 10, 30, 50, 70, 90, 110, 130, and 150 ms. A T2 experiment with a 30 ms relaxation delay was repeated after all of the T2 spectra were collected to ensure there were no changes in the sample or the instrument during the experiments.

### NOE Relaxation

The NOE relaxation spectra were collected using the gNoe.c pulse sequence. One spectrum was collected with NOE saturation on and one with NOE saturation off. The NOE relaxation delay was set at 4 seconds.

### Sample Content (NOE)

300  $\mu$ L of 400  $\mu$ M CaM-CD2-III-5G  
3.5  $\mu$ L of either 100 mM EGTA  
30  $\mu$ L of D<sub>2</sub>O  
3.5  $\mu$ L of Sigma Cocktail Inhibitors

The data were processed with both FELIX98 and NMRPipe. After Fourier transformation, typically a squared sinebell window function shifted over 75° was used. Post acquisition suppression of the water signal was achieved by deconvolution of a Gaussian function with a functional width of 60°. The assignments on the T1, T2, and NOE experiments were carried out using the Sparky assignment software.

The assignments for the NOE spectra of CaM-CD2-III-5G were based on the homonuclear and heteronuclear strategies described in chapter 4.0. The NOE experiment (gNoe.c pulse sequence) does not have strong signal to noise ratio, and required over 20 hours for completion.

### **2.7.3 ModelFree Simulation Experimental Preparation**

To calculate the  $S^2$  value, the ModelFree simulation program was obtained from Dr. Palmer's website (124, 125). The software was installed on an Apple Xserver (aasgard.gsu.edu) and Apple Powerbook laptop (zonda.gsu.edu). The ModelFree program required the BLAS and LAPACK algebra package and a gfortran program. On a single processor computer, the ModelFree program was executed using the following script (run.mf):

```
#!/bin/sh
~/Program/modelfree4_mac/PowerPC/modelfree4 -i mfinput -p mfparam -d mfddata -m mfmodel -s mfpdb -o mfout
```

For a computer with multiple processors such as the Xserver, the following script was used:

```
# For the DYLD Library path
setenv DYLD_LIBRARY_PATH "${DYLD_LIBRARY_PATH}:/Users/chehwix/usr/local/lib"
# The working Directory
set Working = "/Users/chehwix/Calculation/Modelfree/Run3-rev_r2/Model1/G5"
set ModelFree4 = "/Users/chehwix/Program/modelfree4_mac/PowerPC/modelfree4"
$ModelFree4 -i $Working/mfinput -p $Working/mfparam -d $Working/mfddata -m $Working/mfmodel -s $Working/mfpdb -o $Working/mfout
```

To distribute the calculations to available processors, the program was submitted using the qsub command. The qsub command utilized the C-shell. The running script was modified to accommodate the requirement of the C-shell.

Additional information for running the ModelFree software on the Xserver can be found in `"/common/ModelFree4_Mac_Example/ReadMe"` of [aasgard.gsu.edu](http://aasgard.gsu.edu).

The program requires five input files: the simulation parameters (mfinput), HN or CH one bond information (mfparam), protein data (mfddata), model selections (mfmodel), and protein coordinate (mfpdb). Results from the simulation are then directed as output to mfout. The mfddata file contains the following information: the field strength of the magnet, R1, R2, and NOE relaxation data. The T1, T2, and NOE relaxation data collected using a 600 MHz NMR were used for the calculation. The R1 and R2 relaxation times are inversely proportion to the T1 and T2 relaxation times. The uncertainty error values for R1, R2, and NOE were set at 0.03, 0.3, and 0.04 for every residue, respectively. The model selection for each residue is discussed in chapter 5.0.

The calculated structure of CaM-CD2-III-5G using the CYANA program was used as the protein coordinate for ModelFree. All comments in the coordination file were removed.

## 2.8 Gradient Diffusion Experiments

Protein concentrations varied from 0.15 mM to 1.2 mM. Spectra widths of 6600 Hz and 8000 Hz were used for diffusion experiments at 500 MHz and 600 MHz NMR, respectively. The spectra were collected using a modified diffusion\_LED pulse sequence (126, 127) with 16K or 8K complex data points for each FID in 10 mM Tris-HCl pH 7.4 at 25 °C. The diffusion constants were obtained by fitting the desired integrated area of the resonances of each arrayed spectrum through equation 2.4 (128)

$$A = A_0 \exp [-(\gamma\delta G)^2 (\Delta - \delta/3)D] \quad \text{Equation 2.4}$$

where  $\gamma$  is the gyromagnetic ratio of proton. The time between PFG pulse ( $\Delta$ ) and the PFG duration time ( $d$ ) were 80.5 and 5 ms, respectively. The gradient strength ( $G$ ) was arrayed from 0.2 Gauss/cm to about 31.0 Gauss/cm using 64 or 50 steps.  $A$  is the integrated area of desired resonances at each array spectrum after subtraction of baselines.  $A_0$  is the integrated area of the desired resonances when the PFG strength is zero. Using an internal reference with the known diffusion constant and simplifying equation 2.4 to

$$A = A_0 \exp (-C G^2) \quad \text{Equation 2.5}$$

the unknown diffusion constants can then be measured by equation 2.6

$$D = D_0 C/C_0 \quad \text{Equation 2.6}$$

where  $C$  and  $C_0$  are the combination constants for the studied molecule and the internal reference. The data were processed following the steps for the process of dimension 1 in the 2D NMR spectra using Felix98. The  $A$  values used in the calculation were obtained by integrating NMR signals of all the resonances in the identical regions at each 1D spectrum following subtraction of the baseline. The diffusion constant  $D$  was obtained by fitting  $A$  as a function of gradient strength using equation 1 and KaleidaGraph 3.5 (Synergy). The baseline was corrected using a zero-order polynomial. Tris and dioxane were used as internal references and their signals were processed using the same procedures followed for the protein samples. All experiments were repeated three times to ensure the required accuracy.

The resolved peaks of 1D NMR spectra both with and without metal ions were fitted using the 'peak optimize' function of Felix98, which gave the linewidth, position and intensity of each optimized peak.

All of the protein images used in this dissertation were generated using Molscript (Avatar Software AB), MOLMOL (Koradi *et al.* 1996, J Mol Graphics, 14, 51), and Pymol (DeLano Scientific LLC).

## **Chapter 3.0 Developing the Grafting Approach and Using the Grafting Approach to Study the Metal Binding Properties of Calcium Binding Proteins**

In this chapter, we report our studies in developing the grafting approach in four parts. First, we evaluate the lengths of the glycine linkers and the contribution of the host protein environment on the Ca(II) binding properties of the grafted EF-loop. Second, we describe our investigation of Ca(II) binding and conformational properties of the CD2 variants with the inserted EF-hand loops from calmodulin. Third, we describe detailed metal binding studies and conformational analysis for the C-terminal domain of CaM, a predicted EF-hand calcium binding site from a Rubella protease, and a predicted continuous Ca(II) binding site (non-EF-hand) from the Ca(II) sensing receptor (CaR) using heteronuclear NMR.

### **3.1 Developing the Grafting Approach**

#### **3.1.1 Engineering an EF-hand Loop into CD2 with Optimized Linkers**

##### **3.1.1.1 CD2 Variants with Different Lengths of Glycine Linkers**

To identify the key determinants of calcium binding, a loop of the EF-hand calcium-binding site from calmodulin was inserted into the host protein (domain 1 of rat CD2). By inserting this loop into a non-calcium dependent protein, we were able to investigate the contribution of the EF-loop to calcium binding. A solvent-

exposed loop region of the CD2 protein between the C' and D strands (between residues Ser52 - Gly53) was selected for insertion of the EF-hand loop (Figure 3.1). This location was chosen for the insertion of the EF-hand loop to minimize alterations in the protein's conformation, preserve the hydrophobic pocket of the protein, and eliminate the possible interactions between the calcium-binding motif and the host protein. Previous studies have shown that mutations at Ser52 and Gly53 did not change the adhesion affinity of the host protein significantly, suggesting that the integrity of the protein is not disrupted by protein engineering at this location (105, 129).

One of the criteria for the grafting approach is that the grafted Ca(II) binding site should retain its native Ca(II) binding properties (Section 1.5). As shown in Figure 1.5, natural trigger-like-proteins, such as calmodulin and troponin C, undergo major conformational changes upon binding to calcium. The native EF-hand calcium-binding site has two helices to accommodate structural changes in the EF-loop, and the distance between the ends of the two helices changes up to 15 Å upon binding to calcium (Figure 1.5). In contrast, the host protein CD2 has a rigid  $\beta$ -sheet structure, which could restrict the movement of the inserted EF-loop (Figure 3.1). A series of CD2 variants with different lengths of glycine linkers were constructed in order to determine if the inserted EF-loop requires additional flexibility (Table 3.1). As described below, different lengths of glycine linkers were used to connect the EF-loop III of CaM at the CD loop of CD2.

**CaM-CD2-III-0G** "CaM" indicates the protein graft donor. "CD2" is the scaffold protein. "III" identifies the grafted structure. In the case of EF-loop III, "0G" specifies that no glycine (G) linkers were utilized. (Colors for the sequence below, CD2, EF-loop, and glycine linker)

R1....S52-D-K-D-G-N-G-Y-I-S-A-A-E-G53....E99

**CaM-CD2-III-3G** One glycine linker (3 glycine residues) was used to connect the N-terminus of the EF-loop to S52 of the CD2 host protein.

R1....S52-GGG-D-K-D-G-N-G-Y-I-S-A-A-E-G53....E99

**CaM-CD2-III-5G** Two glycine linkers were used to connect the EF-loop to CD2 host protein. Three glycine residues were used to connect the N-terminus of the EF-loop to S52, and due to the existence of G53, only two glycine residues were used to connect the C-terminus EF-loop to the host protein (this applied to all of the CD2 variants with an insertion at S52-G53).

R1....S52-GGG-D-K-D-G-N-G-Y-I-S-A-A-E-GG-G53....E99

**CaM-CD2-III-9G** Five glycine residues were used to connect the N-terminus of the EF-loop to S52 and four glycine residues were used to connect the C-terminus of the EF-loop to G53.

R1....S52-GGGGG-D-K-D-G-N-G-Y-I-S-A-A-E-GGGG-G53....E99



**CaM-CD2-III-13G** Seven glycine residues were used to connect the N-terminus of the EF-loop to S52 and six glycine residues were used to connect the C-terminus of the EF-loop to G53.

R1....S52-GGGGGGG-D-K-D-G-N-G-Y-I-S-A-A-E-GGGGGG-G53....E99

The full list of the CD2 variants is summarized in Table 3.1. The EF-loop III of calmodulin was chosen to investigate the size of the glycine linker and the location of the insertion because extensive studies on the calcium-binding affinity of this EF-loop using peptide models and mutagenesis studies are available (21, 40, 48, 83, 91, 92, 130, 131).

Table 3.1 Summaries of the CD2 Variants

Name	# of Gly in N-terminal linker	# of Gly in C-terminal linker	EF-loop of CaM	Insertion Location in CD2 Host Protein
CaM-CD2-III-0G	0	0	III	S52-G53
CaM-CD2-III-3G	3	0	III	S52-G53
CaM-CD2-III-5G	3	2	III	S52-G53
CaM-CD2-III-9G	5	4	III	S52-G53
CaM-CD2-III-13G	7	6	III	S52-G53
CaM-CD2-III-6G-22	3	3	III	Q22-M23
CaM-CD2-III-6G-83	3	3	III	T83-N84
CaM-CD2-I-5G	3	2	I	S52-G53
CaM-CD2-II-5G	3	2	II	S52-G53
CaM-CD2-IV-5G	3	2	IV	S52-G53

### 3.1.1.2 Metal Binding Studies of the CD2 Variants

To determine if the inserted EF-loop in CD2 still retains its metal binding properties, we carried out metal titration studies with Ca(II) and La(III) ions. The metal binding studies of the CD2 variants were carried out by titrating Ca(II) and

La(III) stock solution into the NMR tube. The protein concentrations of the CD2 variants varied from 150  $\mu$ M to 300  $\mu$ M. The EF-hand loop from site III of calmodulin has been inserted into CD2 with no glycine linker (CaM-CD2-III-0G), one (CaM-CD2-III-3G), two (CaM-CD2-III-5G), or two length-extended glycine linkers (CaM-CD2-III-9G and CaM-CD2-III-13G) by Dr. Yiming Ye (112). The secondary and tertiary structural conformations of the engineered proteins were investigated by NMR. As shown in Figure 3.2, the aromatic ring protons of Trp 32, Trp 7, and Tyr 76 and the methyl protons of the side chains of Val 78, Val 39, and Leu 16 of wild type CD2 show dispersed signals. The conformational analysis by 1D  $^1\text{H}$  NMR indicates that all of the engineered proteins maintain native-like secondary and tertiary structures in the presence and absence of metal ions (Figure 3.2). This result is consistent with CD and Trp fluorescence studies (112). To investigate if the isolated EF-hand motif in CD2 still has the ability to bind metal ions and to determine which residues are involved in the binding process, we have used NMR to monitor metal ion titration experiments. The chemical shift of a nucleus is sensitive to the local protein environments. Upon binding to calcium, the chemical shifts ( $^1\text{H}$ ,  $^{13}\text{C}$ , or  $^{15}\text{N}$ ) of the inserted EF-hand motif in CD2 are likely to shift. The changes observed in the intensity or chemical shift can be used to monitor the metal binding affinity of the protein and to identify residues involved in the binding. As shown in Figure 3.3, the addition of Ca(II) shifted several resonances at 6.96 and 7.89 ppm. On the other hand, the majority of the protein is not changed suggesting a specific binding (Figure

3.2). By monitoring chemical shift changes at 6.96 and 7.89 ppm as a function of Ca(II), we estimated the binding affinity at 180  $\mu$ M Ca(II) using a 1:1 binding model (the equation is listed in Chapter 2). 1D NMR indicates that for the rest of the engineered proteins a Ca(II) titration does not result in any significant changes that can be used to obtain the metal binding affinity using non-labeled protein. The Ca(II) binding affinities were obtained by Dr. Yiming Ye using CD and fluorescence (112). The calcium binding affinities of the CD2 variants are listed in Table 3.2 (112). The CaM-CD2-III-5G and CaM-CD2-III-3G have calcium-binding affinities 20-fold and 10-fold stronger than that of CaM-CD2-III-0G, respectively. The comparison between the calcium binding affinities of CaM-CD2-III-5G and CaM-CD2-III-13G indicates that the extended-length glycine linkers do not improve calcium binding affinity.

The low binding affinities of CaM-CD2-III-0G and CaM-CD2-III-3G could be due to the EF-loop being constrained by the host protein so that a proper calcium binding geometry is not formed. The addition of two glycine linkers (CaM-CD2-III-5G, CaM-CD2-III-9G, and CaM-CD2-III-13G) provides the necessary flexibility to insure the formation of a proper calcium binding geometry. For the remaining section of this dissertation, the discussions focus on the construct with two glycine linkers.

### **3.1.2.1 NMR Structural Studies on Engineered Ca(II) Binding Protein CaM-CD2-III-9G and CaM-CD2-III-13G**

### Structural Studies on the Host Protein

All of the NMR experiments were carried out in the presence of EGTA at 25 °C. The fingerprint regions of the TOCSY spectra of CaM-CD2-III-9G and CaM-CD2-III-13G are shown in Figures 3.4a and 3.4b, respectively. The majority of the backbone resonances from the host protein regions in CaM-CD2-III-9G (G4 to L50 and A75 to E120) and CaM-CD2-III-13G (G4 to L50 and A79 to E124) have been assigned. The chemical shifts of backbone HN, HE1 of Trp, and sidechain of V95 were compared to the same residues of CaM-CD2-III-5G and wild type CD2 to verify the host protein conformations of these two CD2 variants (Figure 3.5 and Table 3.3). These comparisons (see section 3.2.1 for detail on chemical shifts comparisons) indicate that the hydrophobic core and the tertiary structure of the host protein sections of CaM-CD2-III-9G and CaM-CD2-III-13G are not changed upon the insertion of 21 (CaM-CD2-III-9G) and 25 (CaM-CD2-III-13G) residues.

### Structural Studies on the EF-loop III

The numbering system is different for CaM-CD2-III-9G and CaM-CD2-III-13G because both have longer glycine linkers than CaM-CD2-III-5G. So the EF-loop III residues are referred to as positions 1 to 12 to simplify the comparison. The HN chemical shift comparison reveals that some of the HNHA crosspeaks of CaM-CD2-III-9G and CaM-CD2-III-13G are observed at similar locations to those of CaM-CD2-III-5G shown in Figure 3.6. The HNHA crosspeaks of the residues

at positions 3, 6, 7, 8, and 10 of EF-loop III are at similar locations on the TOCSY spectra for CaM-CD2-III-5G, CaM-CD2-III-9G and CaM-CD2-III-13G. In the absence of metal, the isolated EF-loop III has similar structural properties in these three different constructs. The HNHA crosspeaks for the remaining inserted sequences (glycine linker residues and residues at position 1, 2, 4, 9, 11, and 12) of CaM-CD2-III-9G and CaM-CD2-III-13G are also likely to be at similar locations as those of CaM-CD2-III-5G, but these residues were not assigned due to the signal overlap on the TOCSY spectrum of CaM-CD2-III-5G.

#### **3.1.2.4      Summary of the NMR Studies on CaM-CD2-III-9G & CaM-CD2-III-13G**

As shown in the 2D NMR analysis, there are no conformational changes in the host protein sections of CaM-CD2-III-9G and CaM-CD2-III-13G following the insertion of the EF-loop III with the appended glycine linkers. The sidechain assignment of the hydrophobic core residues indicates that the hydrogen bonding network of the host protein is still similar to the CaM-CD2-III-5G, which suggests that the packing of the host protein is also unchanged.

The metal binding abilities of the CaM-CD2-III-9G and CaM-CD2-III-13G are described in Section 3.1.1. The metal binding studies indicate that the metal binding properties of CaM-CD2-III-9G and CaM-CD2-III-13G are similar to that of CaM-CD2-III-5G. Therefore, in the following studies, the isolated calcium binding

sites were grafted into the CD2 protein using three glycine residues per glycine linker.

### **3.1.3 Determining the Effect of Local Electrostatic Environment**

#### **3.1.3.1 CD2 Variants with Different Protein Environments**

The protein surface charge around the metal binding site can directly affect the metal binding affinity, due to either charge attraction or repulsion. This is demonstrated by calbindinD9k, the metal binding affinity decreased up to 456 fold following removal of the negatively-charged residues (132, 133).

Furthermore, mutagenesis studies by Ababou et al indicated that replacing the polar residues at position 41 and 75 (outside of the EF-hand loop) of calmodulin to non-polar residues reduced the calcium binding affinity of EF-loop I and II of calmodulin (79).

To investigate the contribution of the local electrostatic environment of the protein to the metal binding affinity of the inserted EF-loop in CD2, the EF-loop III of CaM was inserted in Q22-M23 and T83-N84 positions in addition to the S52-G53 (Table 3.1). The estimated electrostatic potentials for these three insert locations (see Chapter 2 for details on the procedure for estimating the electrostatic potentials) were predicted using the AMMP (Another Molecular Mechanics Program), and were found to be +47, -42, and -21 kcal/mol for S52-G53, Q22-M23, and T83-N84, respectively (134). Since these three locations also have different hydrophobic environments, hydrogen bonding networks, and

are surrounded by different secondary structures, we can further test if the glycine linkers provide enough flexibility for metal binding.

**CaM-CD2-III-6G-22** The EF-loop III of CaM was inserted in between Q22 and M23, therefore the name of this protein has a number 22 attached at the end. This convention was not applied to the previous nomenclature because these insertions were all completed at the same point in the sequence between S52 and G53. Two glycine linkers were used to connect the EF-loop to the CD2 host protein.

R1....Q22-GGG-D-K-D-G-N-G-Y-I-S-A-A-E-GGG-M23....E99

**CaM-CD2-III-6G-83** The EF-loop III of CaM was inserted between T83 and N84. Two glycine linkers were used to connect the EF-loop to the CD2 host protein.

R1....T83-GGG-D-K-D-G-N-G-Y-I-S-A-A-E-GGG-N84....E99

### 3.1.3.2 Conformational and Metal Binding Studies

The conformational and metal-binding properties of the CD2 variants were investigated using NMR, CD, and fluorescence experiments. The conformational analysis by 1D  $^1\text{H}$  NMR indicated that CaM-CD2-III-6G-22 and CaM-CD2-III-6G-83 maintained native-like secondary and tertiary structures in the presence and absence of metal ions (Figure 3.7) (135). The La(III) binding affinities are 87, 64, and 55  $\mu\text{M}$  for CaM-CD2-III-5G, CaM-CD2-III-6G-22, and CaM-CD2-III-6G-83

using CD and fluorescence by Dr. Yiming Ye, respectively (Table 3.2) (135). The unaffected metal binding ability of the EF-loop III suggests that the glycine linkers provide sufficient flexibility so the EF-loop III is not affected by the hydrophobicity, hydrogen bond, and secondary structure of each insert location. It is also possible that the glycine linkers extended the inserted moiety to highly solvated region that is more than 10 Å away from the host protein, which minimized the effect of the electrostatic environment on the metal binding studies.

### 3.1.4 Metal Binding Studies on the Four EF-hand Motifs in Calmodulin

To investigate the site-specific metal binding properties of each of the four EF-hand motifs in CaM, each EF-loop of calmodulin was individually inserted into the protein between S52 and G53 with a glycine linker on either side of the loop (Table 3.1).

**CaM-CD2-I-5G** The EF-loop I of CaM was inserted between S52 and G53. Three glycine residues were used to connect the N-terminus of the EF-loop to S52 and two glycine residues were used to connect the C-terminus of the EF-loop to G53.

R1....S52-GGG-D-K-D-G-D-G-T-I-T-T-K-E-GG-G53....E99

**CaM-CD2-II-5G** The EF-loop II of CaM was inserted between S52 and G53. Three glycine residues were used to connect the N-terminus of the EF-loop to



S52 and two glycine residues were used to connect the C-terminus of the EF-loop to G53.

R1....S52-GGG-D-A-D-G-N-G-T-I-D-F-P-E-GG-G53....E99

**CaM-CD2-IV-5G** The EF-loop IV of CaM was inserted between S52 and G53.

Three glycine residues were used to connect the N-terminus of the EF-loop to S52 and two glycine residues were used to connect the C-terminus of the EF-loop to G53.

R1....S52-GGG-D-I-D-G-D-G-Q-V-N-Y-E-E-GG-G53....E99

The  $^1\text{H}$  spectra of the CD2 variants and wild type CD2 are shown in Figure 3.8. The well-resolved resonances of all four engineered proteins, observed at the downfield and upfield regions of the spectra, including the aromatic ring proton of W32, the backbone amide proton of the Y93, and the methyl protons of the side chains of V95, V39, and L16, are similar to wild type CD2. The chemical shifts are sensitive to the local environment. The majority of the resonances from the host protein region of CaM-CD2-III-5G do not change significantly, suggesting that the integrity and packing of the host protein frame is maintained after the insertion of the calcium binding loop from calmodulin.

The calcium binding affinities of the CD2 variants with different EF-loop insertions were determined by NMR, CD, and fluorescence (Table 3.4). The calcium binding affinities of the CD2 variants with EF-loops I to IV of CaM are 34,

245, 185, 814  $\mu\text{M}$ , respectively. EF-loop I of CaM has the strongest calcium binding affinity followed by EF-loop III and EF-loop II. EF-loop IV of CaM has the weakest metal binding affinity.

The calcium binding affinities of the four CD2 variants with different EF-loops of CaM, however, are not in good agreement with the acid-pair hypothesis postulated by Reid and co-workers (136). The acid-pair hypothesis stated that there are two criteria to determine the calcium binding affinity of an EF-loop: one is the total number of acidic residues, and the other is the arrangement of the acidic residues in the coordination sphere (Figure 3.9). The EF-loops I, II, and IV of CaM have four acidic residues in the coordination sphere whereas the EF-loop III of CaM only has three acidic residues, which suggests that the EF-loops I, II, and IV of CaM have more stable anionic arrangements (EF-loop III does not have paired axis charge) than EF-loop III. As long as the arrangement of the acidic residues is concerned, Reid and Hodges suggested that the ideal arrangement in the coordination sphere would be to have the acidic residues paired on the x- and z-axes to reduce the dentate-dentate repulsion (there is no paired acidic sidechain on the y-axis since the y position is chelated by a carbonyl oxygen at position 7 of the EF-loop). As shown in Figure 3.9, the coordination sphere of CaM EF-loop I has a pair of charged residues in the z-axis. Similarly, EF-loop IV also has an acid-pair in the z-axis. Taken these two criteria together, according to the acid-pair hypothesis, the calcium binding affinities for the EF-loops in CaM would be  $\text{I} \approx \text{IV} > \text{II} > \text{III}$ . Since EF-loop I and IV

of CaM have four acidic residues in the coordination sphere with paired acidic residues on the Z-axis, the metal binding affinities for these two EF-loops should be the strongest among the four EF-loops of CaM. Because EF-loop III of CaM has three acidic residues in the coordination sphere with no paired acidic axis, EF-loop III would have the weakest calcium binding affinity.

The calcium binding affinities for the CD2 variants with the CaM EF-loops insertion indicate the order of the calcium binding affinities as  $I > III \approx II > IV$ . According to the assumptions of the acid-pair hypothesis, the calcium binding affinity of EF-loop IV should display an affinity similar to EF loop I. However; the calcium binding affinity of EF-loop IV is weaker than expected. Our studies on the CD2 variants with CaM EF-loops insertion prompted us to propose a charge-ligand-balanced model that can well define this discrepancy. The charge ligand-balanced model agrees with the acid-pair hypothesis that the number of acidic residues in the coordination sphere affects the metal binding affinity of an EF-hand motif, but further suggests that the residues adjacent to the calcium binding ligand also affect the affinity. A positively charged residue adjacent to the calcium binding ligand balances the electron dentate-dentate repulsion in the presence or absence of calcium. The calcium binding ligands for the EF-hand motif are located at positions 1, 3, 5, 7, and 12 of the loop. Positions 4 and 6 of the EF-loop are generally Gly residues (21). So the residue types at position 2 and 11 affect the metal binding ability of the EF-loop more frequently. The EF-loop I of CaM has Lys at positions 2 and 11, which can balance out the charges

in the EF-loop. On the contrary, the EF-loop IV of CaM has a Glu at position 11. It is possible that the Glu at this position causes more repulsion in the absence of calcium. The structure comparisons between the calcium free and calcium loaded forms of the EF-hand motifs in the C-terminal domain of CaM have indicated that EF-loop IV exhibits a larger deviation between the apo and the loaded form than EF-loop III. EF-loop II of CaM has four acidic residues in the coordination sphere, but it does not have basic residues at position 2 and 11 of the EF-loop. EF-loop III only has three acidic residues in the coordination sphere, but it has a Lys at position 2 of the EF-loop to balance the charge repulsion in the N-terminus of the loop. The calcium binding affinity of the EF-loop III is similar to the EF-loop II because both of the EF-loops include one of the criteria described in the charge-ligand-balance model.

### **3.2 NMR Structural Studies on of C-terminal Domain of CaM**

To understand the difference in metal binding affinities of EF-loop III and IV of CaM, we have completed detailed NMR studies. The structural and dynamic properties of the C-terminal domain (CaM-CD2-III-5G and CaM-CD2-IV-5G) were determined using NMR. The sequential assignments of the CD2 variants were completed using the homonuclear and heteronuclear assignment strategies. The structures of CaM-CD2-III-5G and CaM-CD2-IV-5G were calculated using the NOE distance restraints, dihedral angles, and residual dipolar couplings to determine if their EF-loops still form native-like geometries in

the foreign host. Metal titration experiments were carried out to determine if the engineered proteins still use residues at position 1, 3, 5, 7, and 12 to coordinate the calcium in a pentagonal bipyramidal geometry. The dynamic properties of the engineered protein in the absence and presence of calcium were studied using  $^{15}\text{N}$  relaxation experiments to understand the difference in the metal binding affinities between the EF-loops.

### **3.2.1 NMR Structural Studies on CaM-CD2-III-5G and CaM-CD2-IV-5G**

#### **Structural Studies on the Host Protein**

The sequential assignment procedure for CaM-CD2-III-5G and CaM-CD2-IV-5G was completed using TOCSY, NOESY, and  $^{15}\text{N}$  HSQC experiments. Additional heteronuclear experiments were carried out to complete the assignment for CaM-CD2-III-5G. Assigned chemical shifts of CaM-CD2-III-5G and CaM-CD2-IV-5G are listed in Appendix 3.1 and 3.2, respectively. The majority of the backbone resonances from the host protein regions in CaM-CD2-III-5G and CaM-CD2-IV-5G (G4 to L50 and A70 to E116) were assigned. The conformations of the host protein regions were verified by comparison to wild type CD2. These comparisons were conducted in three areas: the HN chemical shift comparison, the chemical shift of the HE1 proton of the W32, and the chemical shifts of the methyl protons of L16 and V95. First, the HN chemical shifts of CaM-CD2-III-5G and CaM-CD2-IV-5G were plotted versus the HN chemical shifts in wild type CD2 (Figures 3.5). The differences between the HN

chemical shifts are less than 0.05 ppm, which indicates no major change being observed. Second, the HE1 protons of W32 (the aromatic ring is buried inside the hydrophobic core) for CaM-CD2-III-5G and CaM-CD2-IV-5G are at 10.31 ppm. The HE1 protons of W7 (the aromatic ring is exposed to the solvent) for CaM-CD2-III-5G and CaM-CD2-IV-5G are at 10.13 ppm. The lower field chemical shifts observed for the HE1 protons of W32 suggest that the HE1 protons of W32 are packed inside a hydrophobic environment, similar to those in wild type CD2 (Table 3.3). Third, the sidechain methyl protons from L16 and V95 of CaM-CD2-III-5G and CaM-CD2-IV-5G are at similar chemical shifts compared to the corresponding protons in the wild type CD2 (Table 3.3). These results indicate that the hydrophobic core and the tertiary structure of the host protein of CaM-CD2-III-5G and CaM-CD2-IV-5G are not changed upon the insertion of 17 residues.

### Structural Studies on the EF-loop III

A total of eleven residues of the EF-loop III were assigned using the backbone triple resonances experiments. The sequence of the EF-loop III insertion is shown below:

G53-G54-G55-**D56-K57-D58-G59-N60-G61-Y62-I63-S64-A65-A66-E67-G68**-G69-G70

The assigned region of the EF-loop III insertion is highlighted in bold and the  $^{15}\text{N}$  HSQC spectrum is shown in Figure 3.10. The explanation for assigning the EF-loop is discussed in Chapter 4.

### Structural Studies on the EF-loop IV

The partially-assigned TOCSY and HSQC spectra of CaM-CD2-IV-5G are shown in Figures 3.11a and 3.11b. The assignment for the EF-loop IV of CaM-CD2-IV-5G was relatively difficult using the 2D homonuclear and 2D  $^{15}\text{N}$  HSQC experiments. The resonances of the inserted EF-loop IV and the glycine linkers overlapped with the host protein resonances in the homonuclear spectra. Due to this reason, further sequential assignments on the inserted EF-hand residues will be completed using 3D  $^{15}\text{N}$  TOCSYHSQC and NOESYHSQC in the future. In the absence of the NMR assignment, some of the EF-loop IV crosspeaks were identified by comparing the spectrum of CaM-CD2-IV-5G to that of CaM-CD2-III-5G and wild type CD2. The 1D  $^1\text{H}$  spectrum of CaM-CD2-IV-5G is shown together with CaM-CD2-III-5G and wild type CD2 in Figure 3.12. The resonances in this region are from the backbone amide of Q22, T24, and S90 and sidechain of Asn, Gln, and Arg. The 2D  $^{15}\text{N}$  HSQC spectrum of CaM-CD2-IV-5G is overlaid with that of CaM-CD2-III-5G in Figure 3.13. The sidechain amide region of the CaM-CD2-III-5G HSQC spectrum has an additional pair of crosspeaks in comparison with the CD2 spectrum that were assigned as sidechain amide protons of N60 (position 8 of the loop). The sidechain amide region of the CaM-CD2-IV-5G HSQC spectrum has two additional pairs of crosspeaks (positions 7 and 9 of the loop). The HB\* protons of Asn are usually observed between 2.50 to 2.80 ppm while the HB\* protons of Gln are usually observed between 1.80 and

2.30 ppm. The identities of these four resonances were assigned based on the HD2\* to HB\* NOE connectivities for Asn and the HE\* to HB\* NOE connectivities for Gln. The resonances with chemical shifts of 7.57, 6.82, and 113.63 ppm were assigned as HE21, HE22, and NE2 of Q62, respectively. The resonances with chemical shifts of 7.58, 6.92, and 113.90 ppm were assigned as HD21, HD22, and ND2 of N64, respectively. The HD22 proton of N60 of the CaM-CD2-III-5G and the HE22 proton of the Q62 and HD22 of N64 of CaM-CD2-IV-5G are labeled in Figure 3.12. The sidechain amide regions of the two engineered proteins are different from the same region of wild type CD2. The La(III) titration studies presented in the next section indicate that these peaks change as a function of metal concentrations.

In the glycine region of the HSQC spectrum of CaM-CD2-IV-5G, the crosspeaks for the host protein glycine residues were all assigned. They are G4, G8, G11, G13, G35, G70, G78, G91, and G102. The linewidth of the G78 is very broad, which is due to the overlapping of between resonances of glycine linker residues and G78. There are three resolved newly appeared resonances in the glycine region of the CaM-CD2-IV-5G HSQC spectrum. Both the CaM-CD2-III-5G and CaM-CD2-IV-5G have conserved glycine residues at positions 4 (G59) and 6 (G61) of the EF-loop. The HN and N chemical shifts of both G59 and G61 in CaM-CD2-IV-5G are very similar to those of CaM-CD2-III-5G, where the two glycine crosspeaks of CaM-CD2-IV-5G located at the same positions with those of CaM-CD2-III-5G (Figure 3.13). The two extra crosspeaks observed in the



glycine region of the HSQC spectrum were assigned as G59 and G61 for CaM-CD2-IV-5G. The remaining resolved unassigned resonances of CaM-CD2-IV-5G overlap with G54 of the CaM-CD2-III-5G. Since the glycine linker system used in CaM-CD2-IV-5G is the same as CaM-CD2-III-5G, this resonance was assigned as G54 for CaM-CD2-IV-5G.

The majority of the host protein resonances (>93%) on the HSQC spectrum of CaM-CD2-IV-5G were assigned. The remaining unassigned resonances are classified as part of the inserted EF-loop IV. The sequence composition of the EF-loop IV is different from that of the EF-loop III (7 out of 12 residues are different), and the unassigned resonances on the CaM-CD2-IV-5G spectrum do not overlap with the EF-loop III resonances on the CaM-CD2-III-5G spectrum. The results from the 1D  $^1\text{H}$  spectrum and 2D  $^{15}\text{N}$  HSQC spectrum have indicated that the CaM-CD2-IV-5G protein has maintained the native CD2 structure, and the inserted EF-loop IV has different structural properties from the EF-loop III inserted in CaM-CD2-III-5G.

### **3.2.2 Metal Binding Studies**

There are two questions that need to be answered with respect to EF-loop III and IV insertion into CD2. First, does the EF-loop grafted in the host protein still have the metal binding ability? Second, does the grafted EF-loop in CD2 still utilize the same residues as in the wild type calmodulin to bind with metal? To answer the first question, engineered CaM-CD2-III-5G and CaM-CD2-IV-5G

were titrated with Ca(II) and La(III). Since calcium is a spectroscopically silent metal ion, metal ions in the Ln(III) family are generally used as probes for calcium (137). The metals in the Ln(III) family have similar ionic radius and binding geometry as calcium but normally have stronger binding affinities due to the extra charge. In addition, the Mn(II) has unpaired electrons that broaden the linewidth of the surrounding nuclei and can be used to identify the location of the metal binding pocket (138).

### **3.2.2.1 Preparing NMR Sample for Metal Binding Studies**

Metal ions, especially calcium, can be found in deionized water, on stirring bars, and within the glass in NMR tubes. During the last purification step, CaM-CD2-III-5G was bound to a cation exchange column at a pH <4.0. The undesired proteins were washed away with an acetate buffer at pH 3.5. Then the protein was eluted using the chelex-100 treated Tris buffer. The majority of the metal contamination should have been removed at this point. All of the metal titrations in this chapter contain 50  $\mu$ M EGTA as the first point.

### **3.2.2.2 Ca(II) Metal Titration with CD2 Variants**

#### **Host Protein Conformation in the Presence of Ca(II)**

The  $^{15}\text{N}$  HSQC titration spectra of the CaM-CD2-III-5G and CaM-CD2-IV-5G are shown in Figures 3.14 and 3.15. To verify the host protein conformations of CaM-CD2-III-5G and CaM-CD2-IV-5G in the presence of Ca(II), the HN

chemical shifts of CaM-CD2-III-5G and CaM-CD2-IV-5G in the presence of 1 and 10 mM Ca(II) were compared with wild type CD2, respectively (Figure 3.5). The HN chemical shifts comparison indicates that the host protein regions of the engineered proteins remain unchanged in the presence of Ca(II).

#### EF-loop III of CaM-CD2-III-5G in the Presence of Ca(II)

The Ca(II) titration for CaM-CD2-III-5G was carried out by titrating the protein with 10 mM La(III) stock. The titrations were monitored using  $^{15}\text{N}$  HSQC experiments. During the titration, the crosspeaks from the EF-loop III of CaM-CD2-III-5G did not change as a function of Ca(II) concentration. The final point of the calcium titration was 1026  $\mu\text{M}$ . It is likely that the protein was not saturated with Ca(II), therefore no changes were observed in the HSQC spectra. During the 1D  $^1\text{H}$  calcium titration, the chemical shifts of the HD22 of N60 and the HN of E67 were used to calculate the binding affinity of CaM-CD2-III-5G. However, no significant changes were observed for these two resonances in the  $^{15}\text{N}$  HSQC calcium titration. The 1D experiment was performed at pH 7.4. The 2D titration was performed at pH 6.8. The lower pH may weaken the calcium binding affinity of CaM-CD2-III-5G, hence no change was observed in the  $^{15}\text{N}$  HSQC experiments.

#### EF-loop IV of CaM-CD2-IV-5G in the Presence of Ca(II)

The Ca(II) titration for CaM-CD2-IV-5G was carried out by titrating the protein with a 10 mM Ca(II) stock solution. The titration was monitored with  $^{15}\text{N}$  HSQC NMR experiments. The assigned crosspeaks of EF-loop IV, such as G59, G61, the sidechain resonances of Q62, and the sidechain resonances of N64, change as a function of Ca(II) calcium concentration (Figure 3.15). The chemical shifts of these crosspeaks changed as a function of Ca(II) concentrations, and the changes were monitored up to 10.32 mM Ca(II). The chemical shift changes during the Ca(II) titration are smaller than the chemical shift changes observed in the La(III) titration. The crosspeaks in the HSQC spectrum that were tentatively assigned as putative EF-loop IV crosspeaks also changed as a function of Ca(II) concentration, which further indicates that these crosspeaks belong to the inserted EF-loop IV.

### **3.2.2.3 La(III) Metal Titration with CD2 Variants**

#### **Host Protein Conformation in the Presence of La(III)**

La(III) has a similar ionic radius as Ca(II); it is commonly used as a probe for calcium binding proteins. The  $^{15}\text{N}$  HSQC titration spectra acquired during titration of the CaM-CD2-III-5G and CaM-CD2-IV-5G complexes with La(III) are shown in Figures 3.16 and 3.17, respectively. The La(III) titration for both CaM-CD2-III-5G and CaM-CD2-IV-5G were carried out using a protein concentration of 250  $\mu\text{M}$ . To verify the host protein conformations of CaM-CD2-III-5G and CaM-CD2-IV-5G in the presence of La(III), the HN chemical shifts of CaM-CD2-

III-5G and CaM-CD2-IV-5G in the presence of 230 and 378  $\mu\text{M}$  La(III) were compared with wild type CD2, respectively (Figure 3.5). The HN chemical shifts comparison indicates that the host protein region of the engineered proteins remained unchanged in the presence of La(III).

#### EF-loop III of CaM-CD2-III-5G in the Presence of La(III)

The La(III) titration for CaM-CD2-III-5G was carried out by titrating the protein with 10 mM La(III) stock solution. The titration was monitored using  $^{15}\text{N}$  HSQC experiments. These titration spectra of CaM-CD2-III-5G are shown in Figure 3.16. The resonances from the inserted EF-hand loops and the glycine residues from the glycine linkers changed as a function of La(III) concentration. The crosspeaks for D56 and S64 overlapped with the crosspeak of E67 in the 2D  $^{15}\text{N}$  HSQC spectrum. The crosspeak for E67 was very broad. In the apo form, the edge of the crosspeak of E67 overlapped with the crosspeak of I114. During the La(III) titration, E67 moved upfield (from 8.32 ppm to 8.20 ppm) as a function of the La(III) concentration (Figure 3.18a). At the end of the titration, the edge of the E67 crosspeak overlapped with the T37 crosspeak. The K57 residue is located at position 2 of the EF-loop III. Residue K57 is located at the random coil region of the spectrum where the HN- $^{15}\text{N}$  crosspeak overlaps with other residues. Although both the HN and  $^{15}\text{N}$  chemical shifts of K57 did not show any changes during the La(III) titration, the peak intensity of the K57 decreased significantly in comparison with the resonances of the CD2 host protein. This observation is likely due to the crosspeak of K57 overlapping with the crosspeak of a residue

that belongs to the CD2 host protein, which does not change in the presence of La(III). The D58 residue is located at position 3 of the EF-loop III and the sidechain of D58 is a putative calcium binding ligand. The D58 crosspeak disappeared at a La(III) concentration of 167  $\mu$ M (Figure 3.18b). The G59 crosspeak started to shift upfield at a La(III) concentration of 167  $\mu$ M and disappeared at a La(III) concentration of 230  $\mu$ M (Figure 3.18c). The backbone crosspeaks of N60, G61, Y62, and I63 (Figure 3.18b) and the sidechain of N60 did not change at La(III) concentrations up to 140  $\mu$ M, but disappeared when the La(III) concentration was higher than 167  $\mu$ M. The A65 crosspeak started to move down field at a La(III) concentration of 167  $\mu$ M and the crosspeak disappeared at a La(III) concentration of 230  $\mu$ M (Figure 3.18b). The A66 crosspeak was not observed in the  $^{15}$ N HSQC spectrum. G68 is the first glycine residue after E67 at the C-terminal end of the EF-loop. The G68 crosspeak started to move upfield at a La(III) concentration of 140  $\mu$ M and disappeared at a La(III) concentration of 230  $\mu$ M (Figure 3.18c). During the La(III) titration, the crosspeaks from D58, G59, N60, and N60 sidechains and G61, Y62, I63, A65, E67, and G68 all shifted in the early stages and then disappeared at the higher La(III) concentrations. This collection of La(III) titration data clearly indicate that the inserted EF-loop III in CaM-CD2-III-5G retains the ability to bind metal ions. The change and then disappearance of the NMR signals during La(III) titration indicated that the EF-loop III binds to La(III) in a fast to medium exchange time

regime. The metal binding properties of EF-loop III is further discussed in Section 3.2.4.

#### *EF-loop IV of CaM-CD2-IV-5G in the Presence of La(III)*

The La(III) titration for CaM-CD2-IV-5G was carried out by titrating the protein with 10 mM La(III) stock using  $^{15}\text{N}$  HSQC experiments. The crosspeaks that were tentatively assigned for EF-loop IV, such as G59, G61, sidechains of Q62 and N64, and the additional unassigned crosspeaks, all changed as a function of La(III) concentration (Figure 3.19). The chemical shifts of these crosspeaks shifted after addition of 54  $\mu\text{M}$  La(III) up to 216  $\mu\text{M}$  La(III). The chemical shift changes for the La(III) titration are larger than the chemical shift changes observed in the Ca(II) titration. Unlike the La(III) titration on CaM-CD2-III-5G, the crosspeaks for EF-loop IV remained visible in the presence of 1.46 mM La(III). This data indicates that the metal binding and dynamic properties of the isolated EF-loop IV are different than the EF-loop III in the scaffold protein.

#### **3.2.2.4 La(III) Metal Studies using 3D $^{15}\text{N}$ Edited Experiments**

The La(III) ion is classified as a diamagnetic ion because it does not have unpaired electrons and produces no paramagnetic effects. Therefore, the chemical shift and the linewidth changes during the La(III) titration are due to protein binding to the metal. So during the La(III) titration, the EF-loop III, the glycine linkers, and some of the crosspeaks from CD2 should not disappear from

the spectrum. The 3D  $^{15}\text{N}$  edit experiments were performed to determine if the reason that some resonances disappear in the presence of La(III) is simply due to signal overlap.

The 2D  $^{15}\text{N}$  HSQC spectrum of CaM-CD2-III-5G in the presence of 2 mM La(III) is shown in Figure 3.20. 524 crosspeaks were assigned in the  $^{15}\text{N}$  NOESYHSQC spectrum. No resonances of the inserted EF-loop III are observed in the La(III) loaded spectra. The strip plot of the La(III) free NOESYHSQC spectrum of CaM-CD2-III-5G is shown in Figure 3.21 along with the spectrum of La(III) loaded form. The NOE crosspeaks of N60HN-N60HA and N60HN-G59HA2 are observed in the La(III) free spectrum and these two crosspeaks are not present in the 2 mM La(III) spectrum (see blue box). In the La(III) free spectrum, an inter-residue and an intra-residue NOE crosspeak are observed for both K45 and F49 (K45HN-K45HA and K45HN-R44HA for K45, F49HN-F49HA and F49HN-P48HA). The same crosspeaks have also been observed in the presence of La(III). Since residues K45 and F49 are close to the EF-loop insertion, these data indicate that the sequential NOE interactions of the host protein are not affected by metal binding. In the La(III) free spectrum, there are three NOE crosspeaks for A40: they are A40HN-A40HA, A40HN-V39HA, and A40HN-E33HA. Residue E33 is part of  $\beta$ -strand C, and A40 is part of  $\beta$ -strand C'. The NOE crosspeak A40HN-E33HA is the inter-strand NOE interaction between the  $\beta$ -strands C and C'. All three of the NOE crosspeaks of A40 are in the same location in the La(III) loaded spectrum. In the La(III) free spectrum,



there are four NOE crosspeaks for G3, and they are G35HN-G35HA1, G35HN-G35HA2, G35HN-R34HA, and G35HN-Y93HA. Residue G35 is part of  $\beta$ -strand C, while Y93 is part of  $\beta$ -strand F. The NOE crosspeak G35HN-Y93HA is the inter-strand NOE interaction between  $\beta$ -strands C and F. All four NOE crosspeaks are at the same location on the La(III) loaded spectrum. The nitrogen chemical shift of E41 is shifted slightly by the La(III). The NOE interactions for residues E33, G35, K43, and F49 are observed in both conditions, which suggest that the overall structure of the CD2 host protein is not changed in the presence of metal. The arrangement of the  $\beta$ -strands on the GFCC'C' face of the CaM-CD2-III-5G is similar to that of wild type CD2. The La(III) free spectrum was collected using the Varian Inova 800 MHz NMR housed at the University of Georgia. The 2 mM La(III) spectrum was collected using the Varian Inova 600 MHz NMR housed at Georgia State University. The extra resonances observed at 4.8 ppm in the D2 dimension are the result of differences in water suppression from the different instruments. The comparison between the La(III) free and La(III) loaded spectra indicates that the resonances for EF-loop III all disappeared in the La(III) spectrum.

#### **3.2.2.5 Discussion for the NMR Studies of the C-term-Domain of CaM**

As stated in one of the criteria in section 1.5, it is important that the host protein can provide a stable environment for the inserted EF-loop to interact with metal ions. For both CaM-CD2-III-5G and CaM-CD2-IV-5G, the HN chemical

shifts of the La(III) or Ca(II) loaded forms are comparable to the wild type CD2 (Figure 3.5). There are no major changes observed for the host protein residues, suggesting that the host protein has maintained its native conformation in the presence of La(III) or Ca(II). The goal of the grafting approach was to bypass the conformational change of the native calcium binding protein but also provide a stable environment that allows the EF-loop to form a proper calcium binding geometry. Based on the results of these titration studies, we are able to further confirm that the host protein has also maintained the native conformation in the presence of La(III) or Ca(II). In addition, the direct observation of residues from grafted EF-loop III and IV of CaM involved in Ca(II) binding suggest that they maintain the original Ca(II) binding properties. These results are in good agreement with the original design plan.

The metal binding studies in section 3.1.4 have shown that the EF-loop III of CaM-CD2-III-5G has a stronger metal binding affinity than the EF-loop IV of CaM-CD2-IV-5G. The charge-ligand-balance model suggests that the inclusion of positively charged residues at position 2 of the loop enables EF-loop III to have a stronger metal binding affinity than EF-loop IV. Our metal binding studies led us to propose the following working models with two determinants that contribute to the different site specific calcium binding affinities of the C-terminal domain of calmodulin.

First, it is possible that the conformational differences between the apo and Ca(II) loaded forms of site IV are larger than site III of calmodulin.

Therefore, more energy is required for the formation of the metal binding pocket. In the absence of metal, we have observed that the resonances of the calcium binding ligands at position 1, 3, 5, 7, and 12 of EF-loop III in CaM-CD2-III-5G have different chemical shifts than the calcium binding ligands at the corresponding positions of EF-loop IV in CaM-CD2-IV-5G (Figure 3.13). These data indicate that EF-loop III and IV of calmodulin have different structural properties in the apo form. This possibility is further supported by the structural comparison between the apo and loaded forms of the EF-loops from CaM. Figure 3.22 shows that the RMS deviation between the  $C\alpha$  atoms of the apo form to the calcium loaded form for site IV is significantly greater than that of site III of calmodulin (1CFD compared to 3CLN) (46, 62). On the other hand, both sites are very similar with small RMS deviations in the presence of Ca(II).

Second, the dynamic properties of the two EF-loops are different, which can also affect the metal binding affinity due to a difference in conformational entropy. While it is widely accepted that dynamic properties will contribute to the Ca(II) binding, the direct demonstration of such an effect is complicated by the coupled EF-hand motifs and domain-domain interaction in the intact Ca(II) binding proteins. Elegant studies on the dynamic properties of the EF-hand proteins were previously studied by several research groups, such as Bax, Ikura, and Linse, using the intact calmodulin and calbindinD9k (60, 65, 139). The binding of the first EF-hand motif in the same domain was shown to affect the dynamic properties of the second EF-hand motif and vice versa. The advantage

of our grafting approach to engineer a single calcium binding site is that the dynamic properties can be obtained without the interference of other metal binding site. Our La(III) titration studies on both CaM-CD2-III-5G and CaM-CD2-IV-5G reveal that the binding processes of these two EF-loops are different. The resonances for the EF-loop III disappeared from the spectrum after the addition of 230  $\mu$ M La(III). The addition of La(III) to CaM-CD2-IV-5G caused the resonances of the EF-loop IV to shift to different regions but remained visible in the spectrum. The results from the NMR spectra suggest that the residues for EF-loop IV are in a fast exchange time scale, while the residues for EF-loop III are in the medium exchange time scale. The disappearance of crosspeaks observed in EF-loop III at higher La(III) concentrations is a result of extensive line broadening. Since, EF-loop III and EF-loop IV were grafted into CD2 with two glycine linkers, it is reasonable to state that the changes observed in the NMR spectra are solely from the inserted EF-loop.

To understand the dynamic properties of CaM-CD2-III-5G and CaM-CD2-IV-5G, we have carried out detailed structural studies on CaM-CD2-III-5G using heteronuclear NMR experiment in chapter 4. In addition, we have also employed hydrogen exchange and  $^{15}\text{N}$  relaxation methods to study the dynamic properties of CaM-CD2-III-5G and CaM-CD2-IV-5G as is shown in chapter 5.

### **3.3 Application of the Grafting Approach to the Study of Metal Binding Properties of a Predicted EF-hand Motif from Rubella Virus**

In the previous sections, we have demonstrated that the grafting system can be used to obtain site-specific properties of each individual EF-loop from CaM. In this section, use of the grafting system to investigate the metal-binding properties and conformation of the EF-loop from Rubella virus is discussed. The EF-loop was inserted between S52 and G53 with two glycine linkers, the insertion sequence is shown below:

R1....S52-GGG-DASPDGTGDPLD-GG-G53....E99

The structural studies on Rub-CD2-5G were carried out using homonuclear experiments. All of the NMR experiments were carried out in 20 mM PIPES and 10 mM KCl buffers at 25 °C. The fingerprint region of the TOCSY spectrum of Rub-CD2-5G is shown in Figure 3.23.

### 3.3.1 NMR Structural Studies on the Rub-CD2-5G

#### Structural Studies on the Host Protein

The sequential assignment for Rub-CD2-5G was carried out using TOCSY and NOESY spectra. The HN and sidechain chemical shifts were analogous to those observed for CaM-CD2-III-5G and the other CD2 engineered proteins (Figures 3.5 and Table 3.3). No major changes in the backbone chemical shifts were observed in the NMR spectra, which suggest that the glycine linkers allow the CD2 host protein to tolerate the EF-loop insertion from a different protein.

#### Structural Studies on the inserted EF-loop

The assignment for the EF-loop of Rub was more challenging than the EF-loop III of CaM because of the two Pro residues in the EF-loop. A Pro residue does not have an HN proton. No HNHA crosspeaks are observed in the fingerprint region of the TOCSY or NOESY spectra, which causes a break in the "NOESY walk" process. A total of five residues from the inserted EF-loop were assigned (D56, A57, G61, T62, and L66).

A57 has a distinct HB# crosspeak at 1.36 ppm with NOE crosspeaks to the H $\beta$  protons of D56. An additional Gly-Thr pattern has been observed in the spectrum. There were three such patterns in the engineered protein (G74-T75, G102-T103, and G61-T62). Residues G74, T75, G102, and T103 were already assigned in the host protein. Thus, the additional pattern was assigned to G61-T62. There was no NOE link between L66 and P65. All of the Leu residues for the CD2 host protein were all assigned. The remaining unassigned residues in the EF-loop were Leu, Ser, two Pro, three Asp, and Gly. Leu is the only residue that would have more than two sidechain protons. So, the HN, HA, HB1, HB2, HD1#, and HD2# protons of L66 were assigned at 8.46, 4.29, 1.68, 1.59, 0.91, and 0.85 ppm, respectively.

### **3.3.2 La(III) Binding Study on Rub-CD2-5G**

To understand the metal binding properties of Rub-CD2-5G, an NMR structural study was carried out in the presence of 1 mM La(III). The La(III) TOCSY spectrum of Rub-CD2-5G was overlaid with the spectrum that had

excess EGTA added to the sample (Figure 3.24). The host protein resonances do not change significantly, so the assignment for the La(III) loaded form was carried over from the EGTA spectrum of Rub-CD2-5G. A comparison of the HN chemical shifts between the host protein of Rub-CD2-5G and La(III)-Rub-CD2-5G is shown in Figure 3.5. The resonances of D56, A57, G61, T62, and L66 all shifted in the presence of La(III). This result suggests that the residues on the EF-loop of Rub-CD2-5G bind to La(III). The chemical shifts of the CD2 host protein of Rub-CD2-5G remained unchanged, which indicates that the host protein structure is not affected by the metal binding process of the inserted EF-loop.

### **3.4 Application of a Grafting Approach to a Study of Metal Binding Properties of a Predicted Non-EF-hand Ca(II) Binding Site from CaR**

The extracellular calcium binding sites of the calcium sensing receptor (CaR) contain five putative calcium binding sites that are of either the continuous or non-continuous types. The arrangements of the continuous type of calcium binding ligands in CaR are different from the EF-hand motif. The identification of the continuous and non-continuous calcium binding sites in CaR was based on the geometric description established in our lab using the Dezymer algorithm.

The calcium binding sites of CaR were predicted based on the computational model of CaR based on a geometric description. A total of five calcium binding sites were predicted. In this part of the dissertation, the third potential calcium binding site (based on the sequence order) was inserted between S52 and G53 for metal binding and conformational analysis. The insertion sequence of the CaR-CD2-III-0G is shown below:

**R1....S52-GIEKFREEAEERDI-G53....E99**

The main difference between the grafting approach for this CaR-CD2-III-0G protein and the other grafting approach for the other CD2 variants discussed in the previous sections is that the calcium binding site III of the CaR is not an EF-hand calcium binding site. The calcium binding loop of CaR is fourteen-residues long instead of twelve (the canonical EF-hand loop). The arrangement of the calcium binding ligands is different in comparison to those of the EF-hand motif.

The structural studies on CaR-CD2-III-0G were carried out using homonuclear and heteronuclear experiments. All of the NMR experiments were carried out in 20 mM PIPES and 10 mM KCl buffers at 25 °C. The fingerprint region of the TOCSY spectrum of CaR-CD2-III-0G is shown in Figure 3.25. The <sup>15</sup>N HSQC spectrum of CaR-CD2-III-0G is shown in Figure 3.26.

### **3.4.1 NMR Structural Studies on CaR-CD2-III-0G**

#### **Structural Study on the Host Protein**



The sequential assignments for CaR-CD2-III-0G were carried out using the TOCSY and  $^{15}\text{N}$  HSQC experiments. First, the spin systems in the TOCSY spectrum were compared to the TOCSY spectrum of CaM-CD2-III-5G and assigned. Apart from the signal overlap observed in the TOCSY spectra for the CD2 variants, there were several residues identified with similar HN chemical shifts. The assignment was further verified with the  $^{15}\text{N}$  HSQC spectrum. Similar to all the CD2 variants studied in the previous sections, the proton chemical shifts for the host protein section of the CaR-CD2-III-0G was very similar to those of CaM-CD2-III-5G ( $< 0.05$  ppm difference, Figure 3.5). The calcium binding loop of CaR is two residues longer than the EF-loop of the canonical EF-hand motif and it was inserted between S52 and G53 without using glycine linkers. Based on the HN chemical shift comparisons, the CD2 host protein retains its native conformation following the insertion of a non-EF-hand I loop.

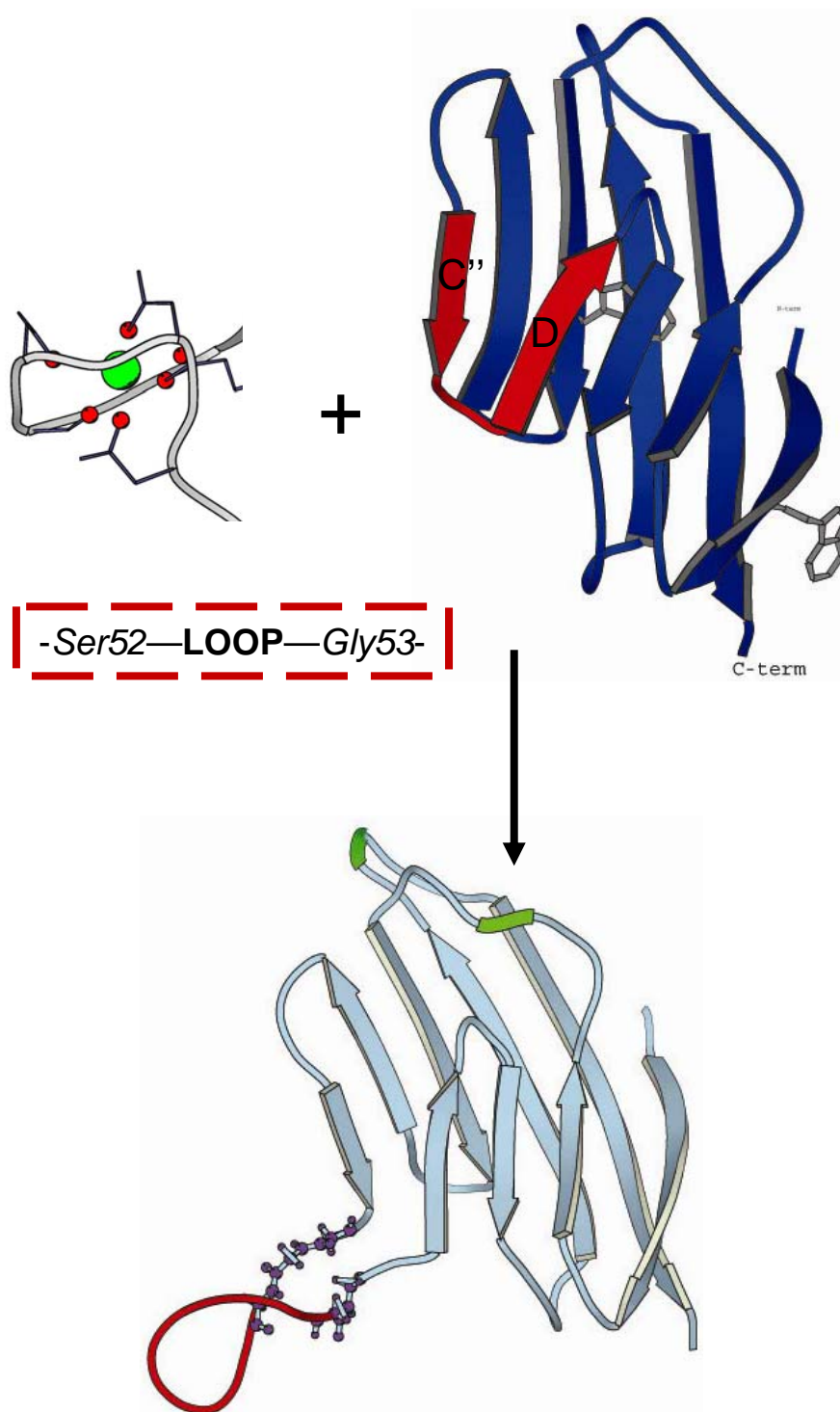
#### Structural Study on the Predicted Calcium Binding Loop

The  $^{15}\text{N}$  HSQC spectrum of the CaR-CD2-III-0G is overlaid on top of the spectrum of CaM-CD2-III-5G in Figure 3.27. The resonances of the calcium binding loop residues of CaR-CD2-III-0G were located at different locations in comparison to the EF-loop III resonances of CaM-CD2-III-5G. The calcium binding loop III of the CaR is composed of 1 Gly, 1 Phe, 1 Ala, 1 Asp, 2 Ile, 5 Glu, 2 Arg, and 1 Lys residues. Because the NOESY spectrum is not available, the assignment of the calcium binding loop was based on the spin system

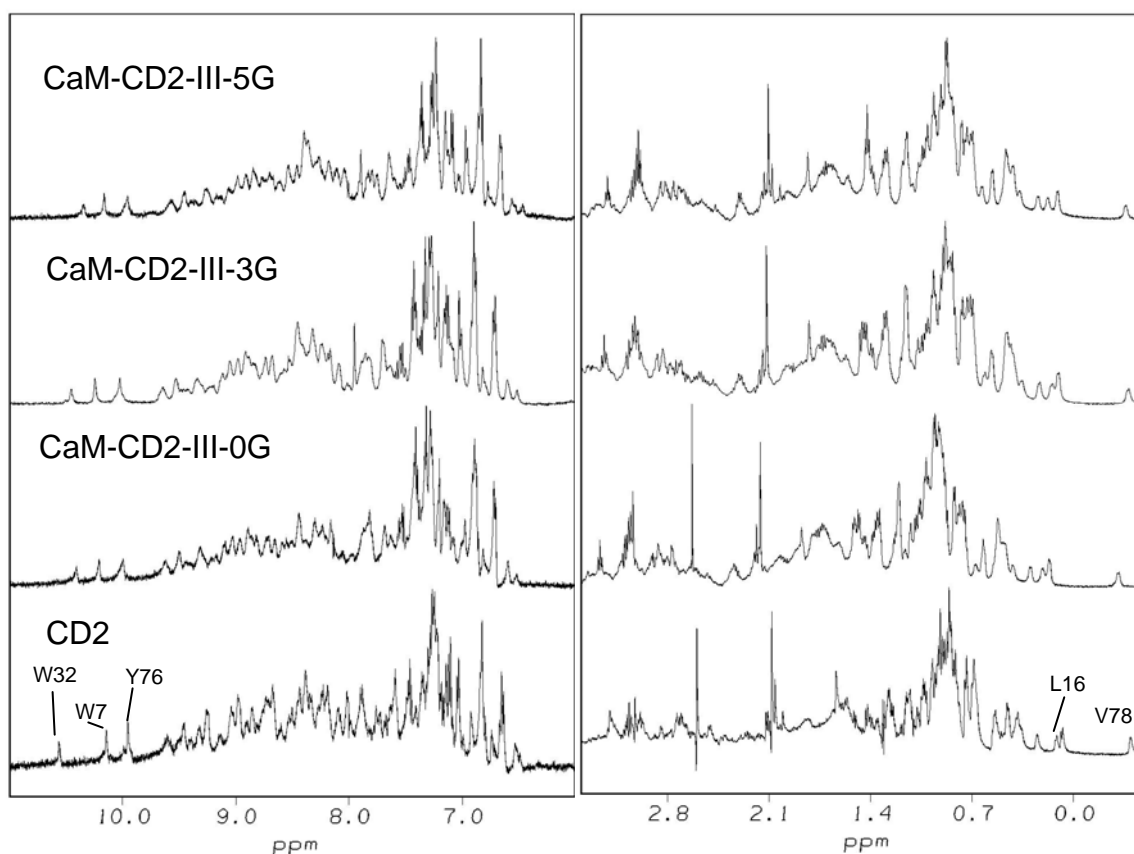
characteristics of each spin type. Using the TOCSY and HSQC spectra, residues G53, F57, A61, and D65 were assigned. Since all of the glycine residues in the CD2 host protein have been assigned, the remaining pair of crosspeaks belonging to Gly were assigned as the HA1 and HA2 protons for G61. There were two type-J spin systems observed in the TOCSY spectrum. Residues F57 and D65 were the only residues in the calcium binding loop that have the type-J spin systems. The H $\beta$  protons of Phe usually display more downfield chemical shift values than the H $\beta$  protons of Asp. Thus, the crosspeaks observed at 3.14 and 3.05 ppm at D1 = 8.20 ppm were assigned as H $\beta$ 1 and H $\beta$ 2 of F57. The H $\alpha$  proton of F57 was at 4.58 ppm. The crosspeaks observed at 2.74 and 2.60 ppm were assigned as H $\beta$ 1 and H $\beta$ 2 of D65. The H $\alpha$  proton of D65 was at 4.64 ppm. The Ala residue only has 1 crosspeak between 0.9 to 1.40 ppm with a strong intensity for the sidechain H $\beta$  protons because the signals for the three H $\beta$  protons are usually averaged into a single crosspeak. Two crosspeaks observed at 4.27 and 1.38 ppm appeared at D1 = 8.25 ppm, and these two protons were assigned as H $\alpha$  and H $\beta$  protons of A61. The remaining 10 residues on the calcium binding loop of CaR-CD2-III-0G were not assigned because there were multiple occurrences of the same type of residues, and without the NOESY spectrum, the order of the residues could not be established.

### 3.5 Summary

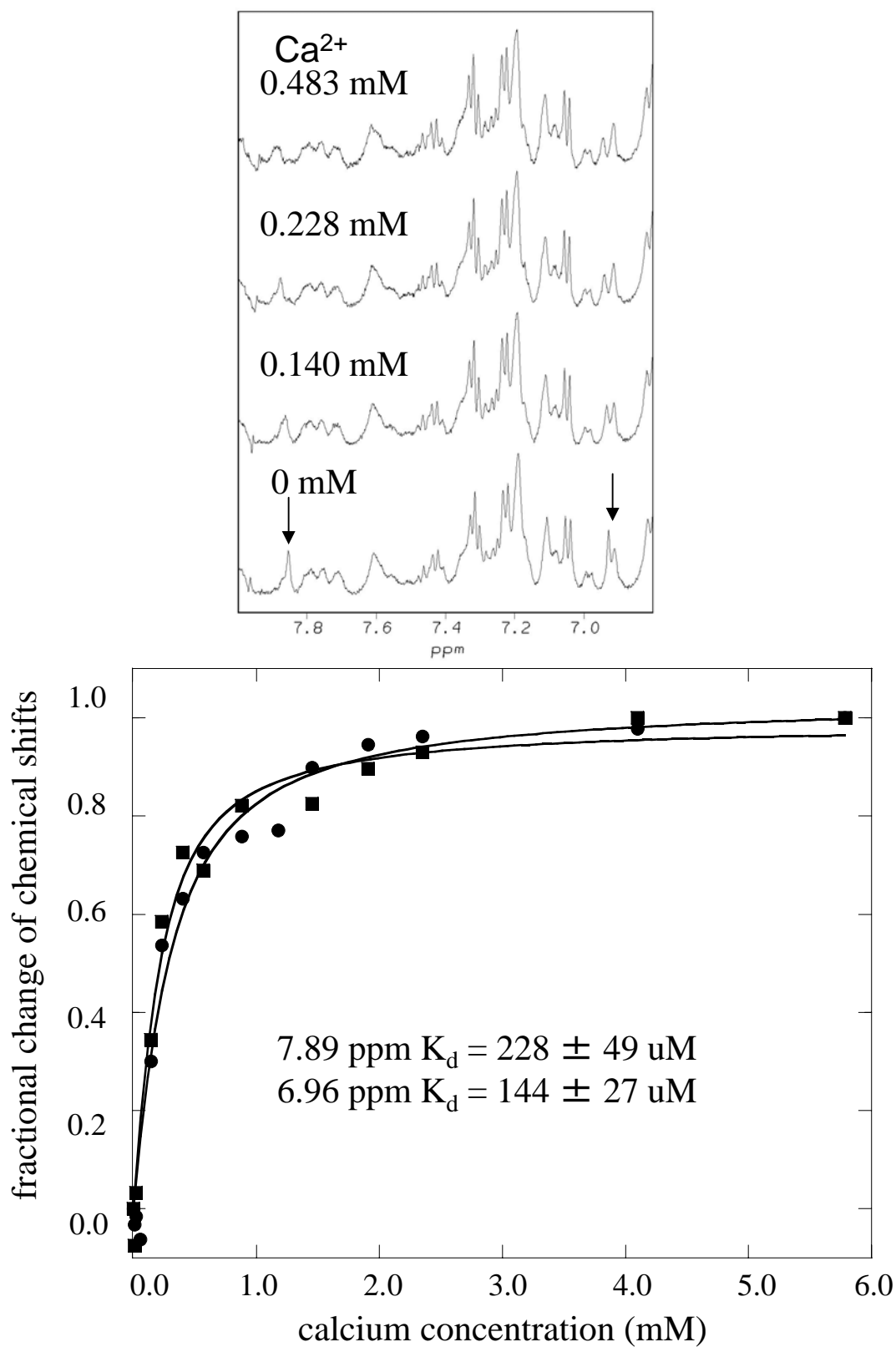
In summary, we have demonstrated the advantages of using the grafting system to study the calcium binding sites from CaM, Rub, and CaR. The insertion moieties vary with different lengths of glycine linkers, canonical EF-loops with different sequences, and non-EF-hand type calcium binding loop. The NMR structural studies on these CD2 variants have shown that the CD2 host protein retains a native-like conformation and does not exhibit non-covalent interactions with the inserted EF-loop. The calcium binding site is able to bind metals without the complication of conformational change and pair-pair interaction. A working model with regards to the different structural and dynamic properties has been proposed to understand the different Ca(II) binding affinities between the EF-loop III and EF-loop IV.



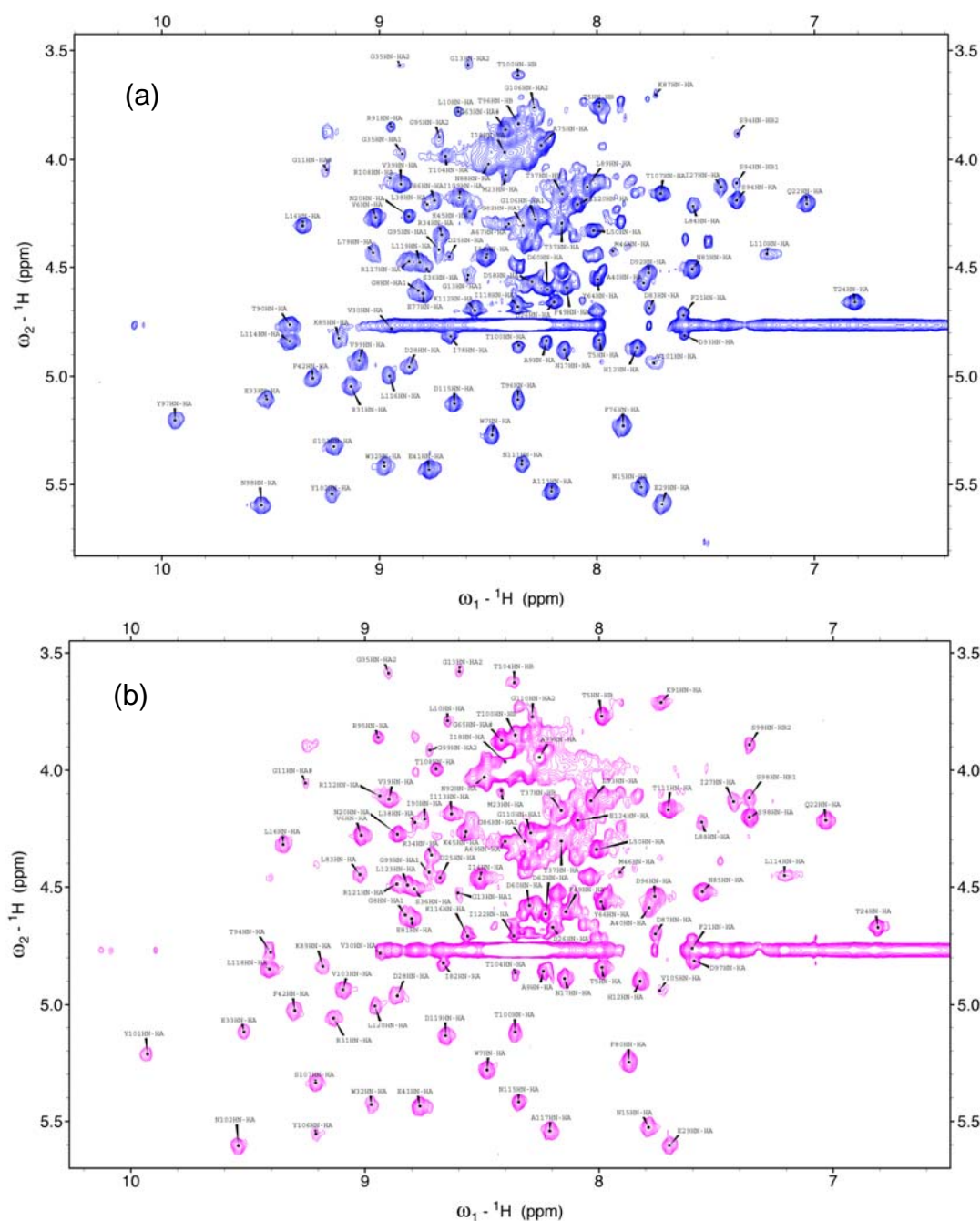
**Figure 3.1 The grafting approach:** The EF-loop III of calmodulin inserted in between S52 and G53 of CD2. The computational model of CaM-CD2-III-5G was generated using AMMP.



**Figure 3.2  $^1\text{H}$  spectra of CD2 variants with different glycine linkers:** 1D spectra of CaM-CD2-III-0G, CaM-CD2-III-3G, and CaM-CD2-III-5G stack in the same spectra as wild type CD2. All of the 1D spectra were collected in 10 mM Tris 10 mM KCl (pH 7.4) buffer at 25 °C using 500 MHz NMR.



**Figure 3.3 Calcium titration of CaM-CD2-III-5G:** The calcium binding affinity of CaM-CD2-III-5G was calculated using the chemical shift change as a function of calcium concentration. The titration experiments were performed in 10 mM Tris 10 mM KCl at pH 7.4 using 600 MHz NMR.

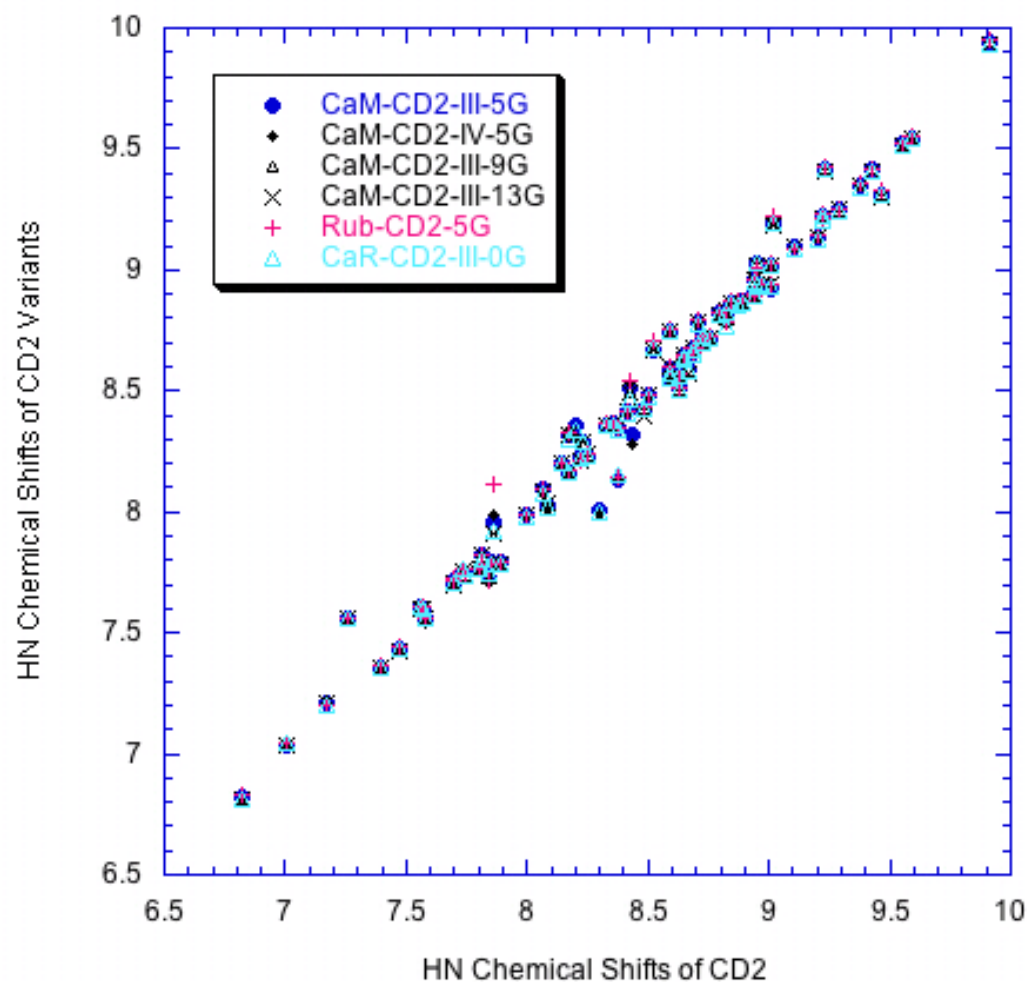


**Figure 3.4 TOCSY spectra of CD2 variants with extended glycine linkers:** TOCSY spectra of CaM-CD2-III-9G (a) and CaM-CD2-III-13G (b) were recorded using a Varian 600 MHz NMR instrument. Both of the TOCSY spectra were collected in the presence of 1 mM EGTA in 10 mM Tris 10 mM KCl (pH 7.4) buffer at 25 °C.

Table 3.2 Metal Binding Affinities of CD2 Variants

<b>Name</b>	<b><math>K_d</math> <math>\text{Ca}^{2+}</math> (<math>\mu\text{M}</math>)</b>	<b><math>K_d</math> <math>\text{La}^{3+}</math> (<math>\mu\text{M}</math>)</b>
CaM-CD2-III	$\sim 3000$	570 - 1500
CaM-CD2-III-3G	$> 3000$	$420 \pm 5$
CaM-CD2-III-5G	$186 \pm 40$	$87 \pm 9$
CaM-CD2-III-6G-22		$64 \pm 4$
CaM-CD2-III-6G-83		$55 \pm 4$
CaM-CD2-III-9G	n/a	
CaM-CD2-III-13G	$195 \pm 30$	





**Figure 3.5 HN chemical shifts comparison between CD2 variants and wild type CD2:** The HN chemical shifts comparisons with the same residues in wild type CD2.

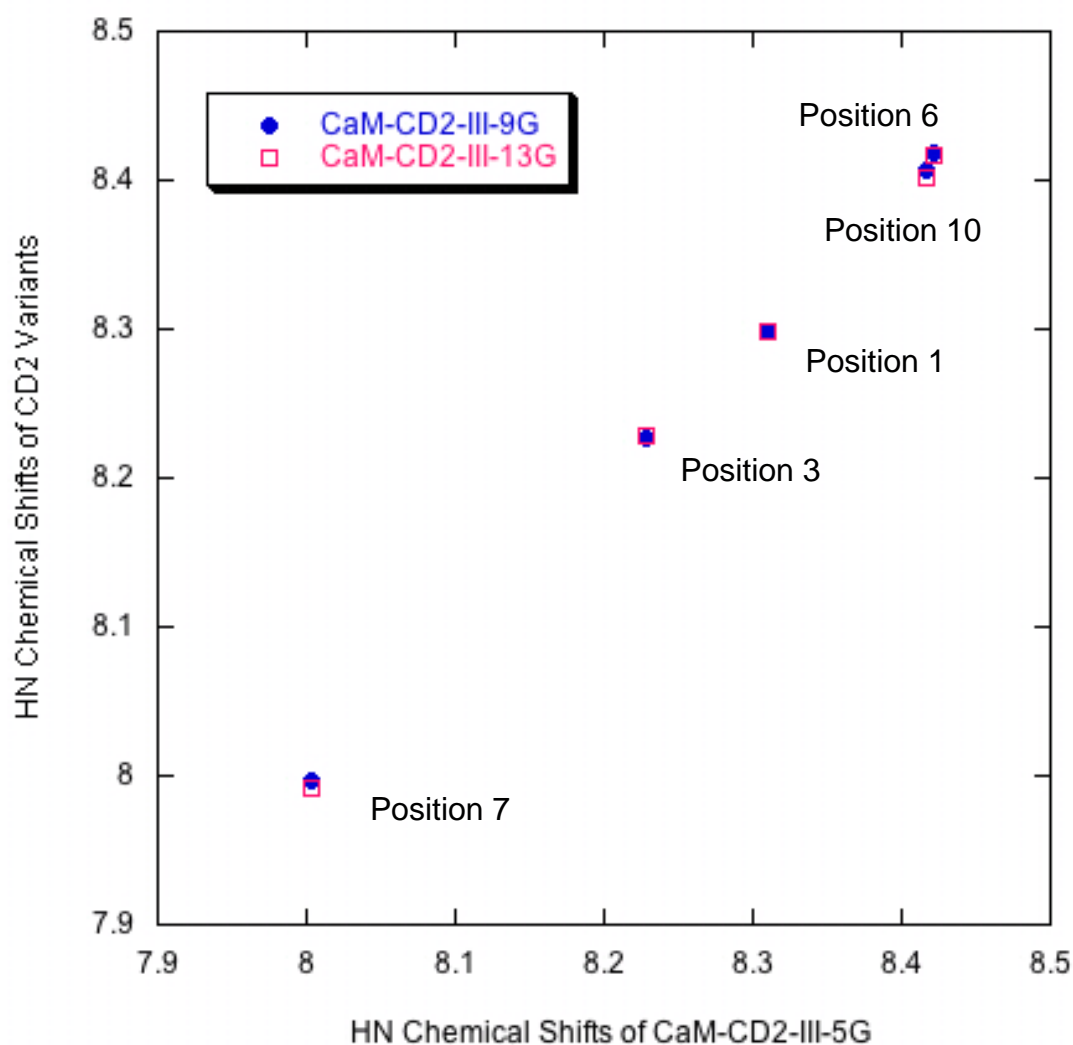
Table 3.3 Chemical Shifts of HE1 Protons of W32 of CD2 Variants

Name	HE1 Proton of W32 (ppm)
CD2	10.57
CaM-CD2-III-5G	10.32
CaM-CD2-III-9G	10.31
CaM-CD2-III-13G	10.31
CaM-CD2-IV-5G	10.29
CaR-CD2-III-0G	10.12
Rub-CD2-5G	10.29

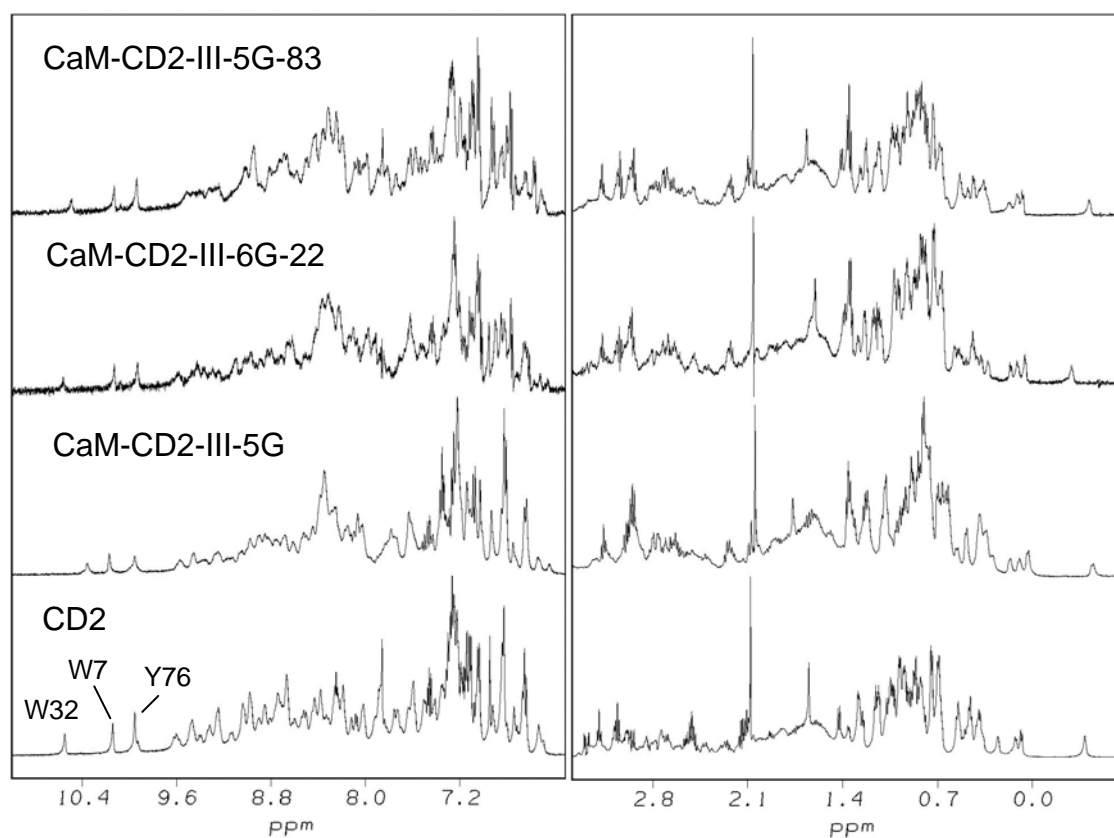
Assignment of CD2 is from BMRB 4109, Chen et al (1998) J Biomol NMR, 457

The CD2 assignment was obtained in 20 mM phosphate buffer at pH 5.0.

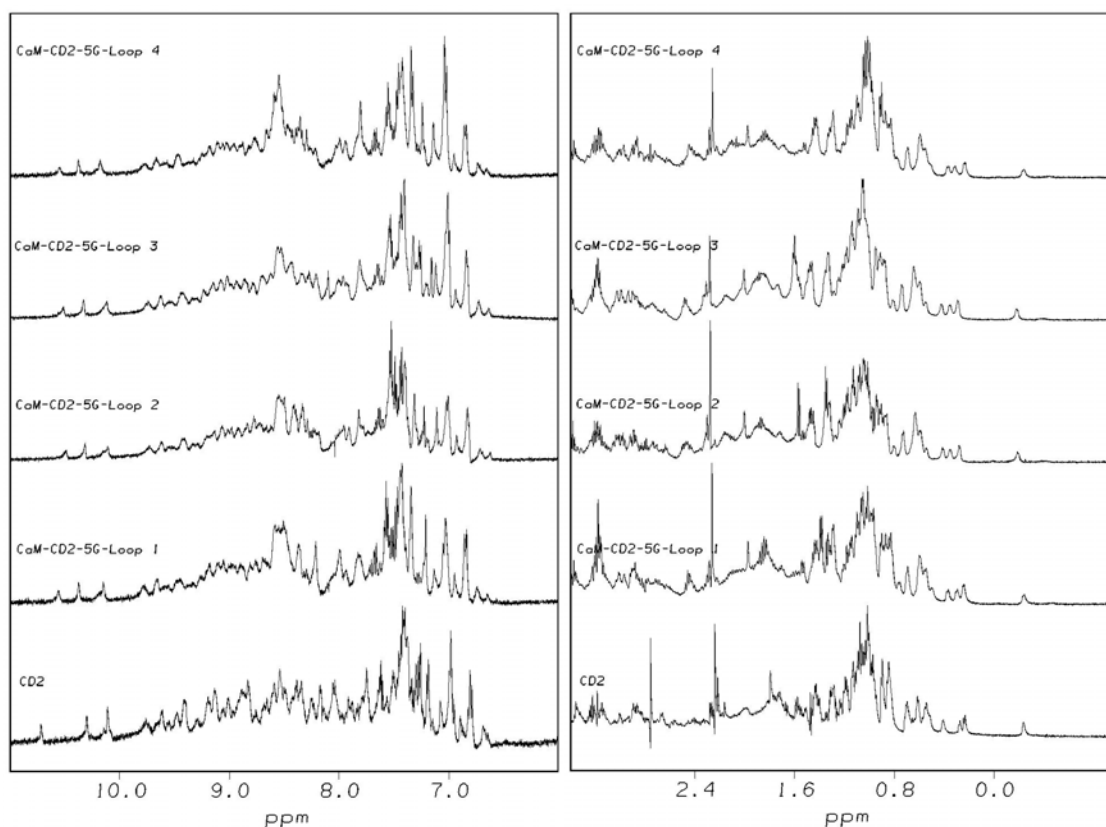
All of the assignment of engineered Ca(II) binding proteins were obtained in 10 mM Tris 10 mM KCl at pH 7.4.



**Figure 3.6 HN chemical shifts comparison between CD2 variants and CaM-CD2-III-5G:** HN chemical shifts comparison on the EF-loop residues of CaM-CD2-III-5G, CaM-CD2-III-9G, and CaM-CD2-III-13G.



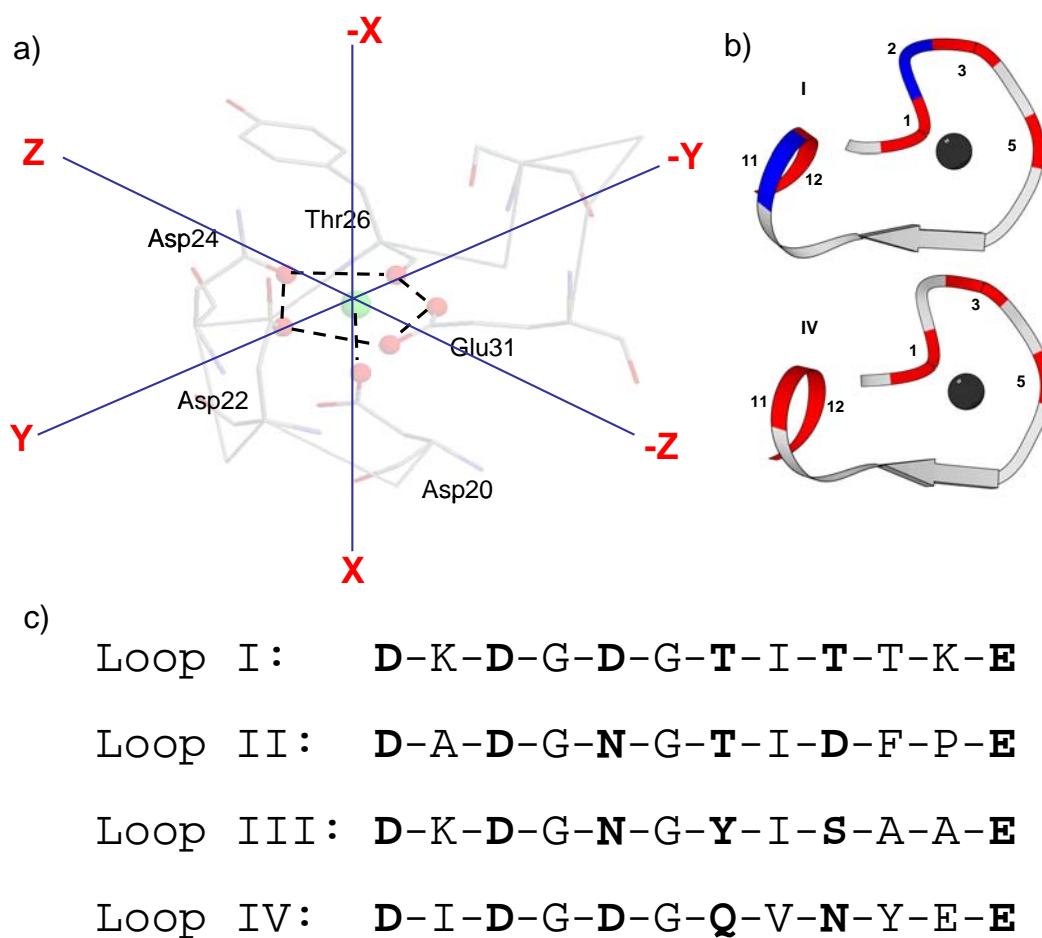
**Figure 3.7  $^1\text{H}$  spectra of EF-loop III inserted at Q22, S52, and T83 positions in CD2:** 1D spectra of CaM-CD2-III-6G-83, CaM-CD2-III-6G-22, and CaM-CD2-III-5G stack in the same spectra as wild type CD2. All of the 1D spectra were collected in 10 mM Tris 10 mM KCl (pH 7.4) buffer at 25 °C.



**Figure 3.8  $^1\text{H}$  spectra of CD2 Variants with different EF-loops of CaM insertion:** 1D  $^1\text{H}$  spectra of CD2 variants with EF-loops I, II, III, and IV insertion stacked on the same spectra with the wild type CD2. All of the 1D spectra were collected in 10 mM Tris 10 mM KCl (pH 7.4) buffer at 25 °C..

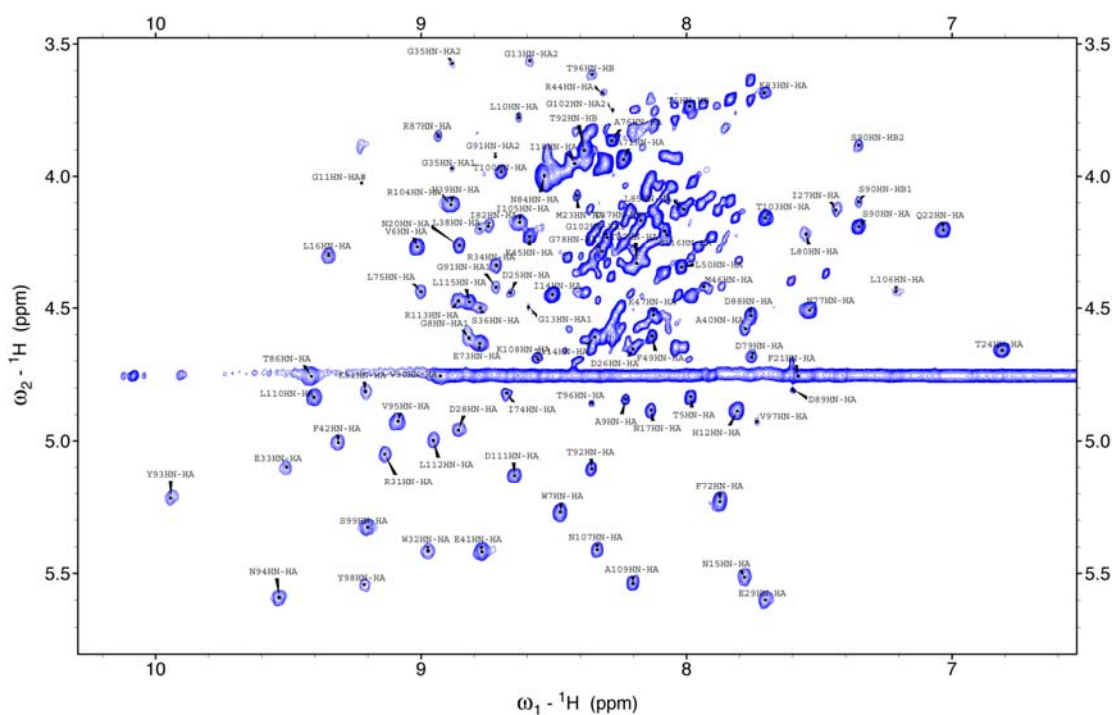
Table 3.4 Ca(II) Binding Affinities of the CD2 Variants With the EF-loops of CaM.

Name	Charge	Axis	P2	P11	$K_d \text{ Ca}^{2+} (\mu\text{M})$
CaM-CD2-I-5G	4	Z	<b>K</b>	<b>K</b>	$34 \pm 9$
CaM-CD2-III-5G	3	-	<b>K</b>	<i>A</i>	$185 \pm 30$
CaM-CD2-II-5G	4	X	<i>A</i>	<i>P</i>	$245 \pm 6$
CaM-CD2-IV-5G	4	Z	<i>I</i>	<i>E</i>	$814 \pm 38$



**Figure 3.9 The charged-balanced ligand model:** a) The coordination geometry of EF-hand Site I of calmodulin. b) The EF-loop I (top) and IV (below) of calmodulin. The negative residues are labeled in red and the positive residues are labeled in blue. c) The sequence of the EF-loops in calmodulin.

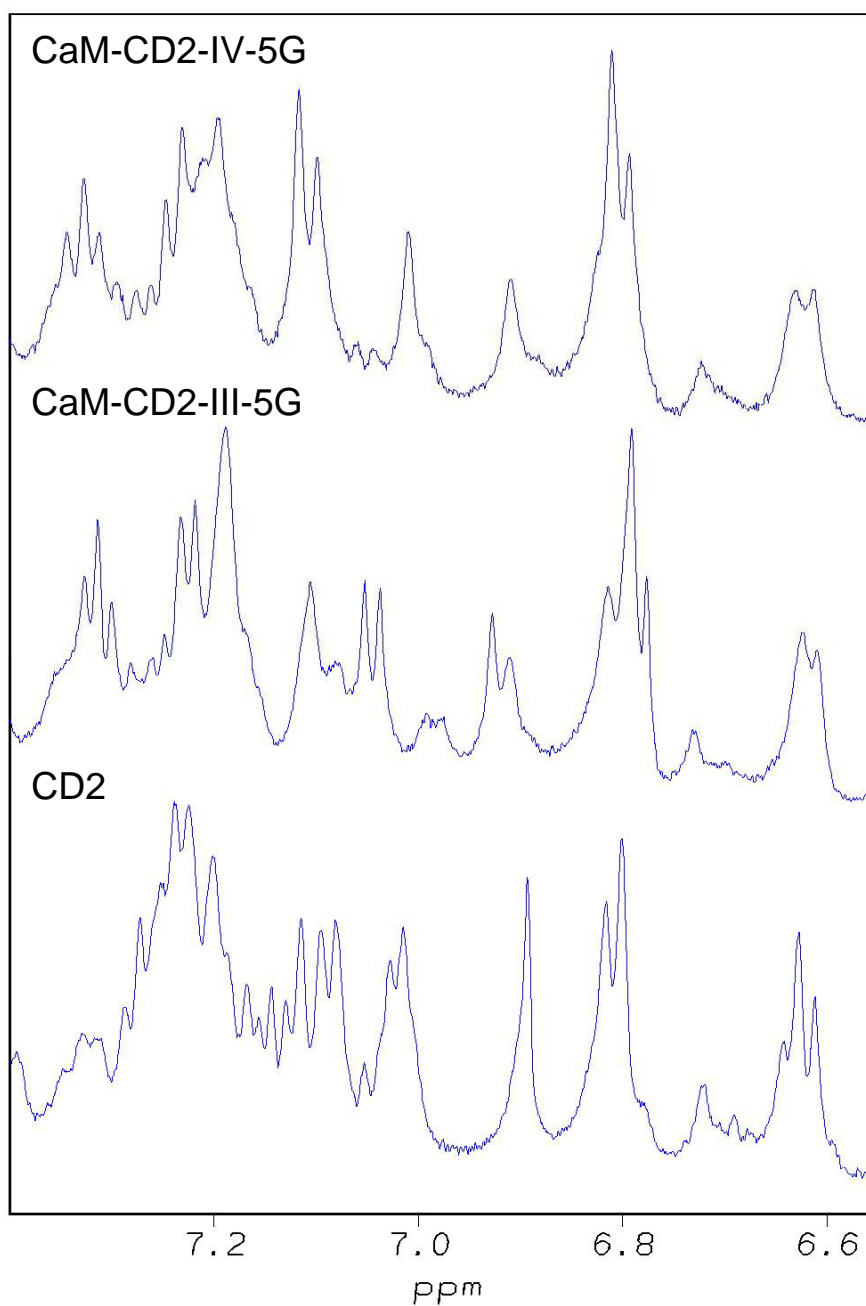
**Figure 3.10 Assigned  $^{15}\text{N}$  HSQC spectrum of CaM-CD2-III-5G:** The HSQC spectrum was collected in the presence of 1 mM EGTA in 20 mM PIPES 10 mM KCl in pH 6.8 buffer at 25 °C.



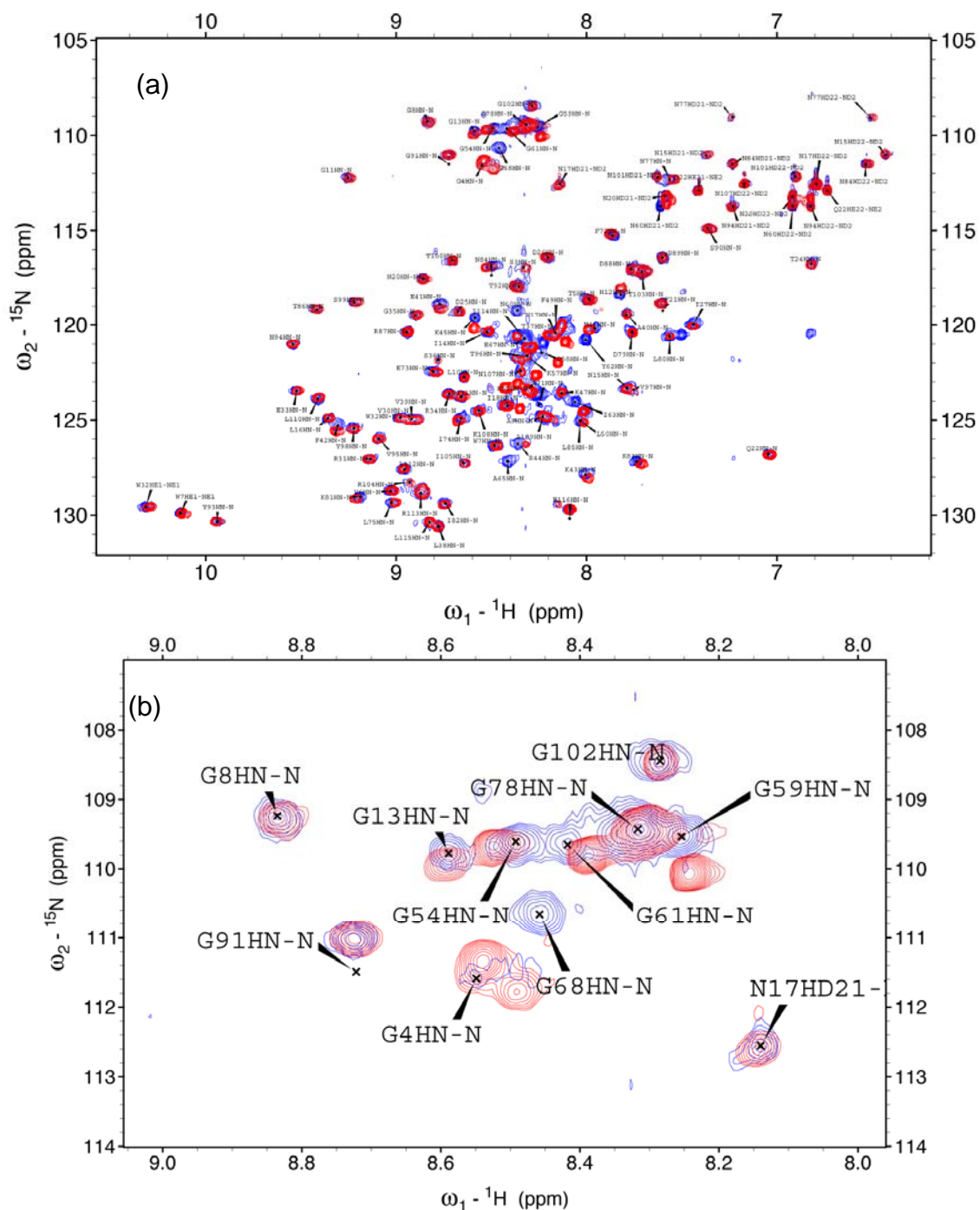
**Figure 3.11a Fingerprint region of CaM-CD2-IV-5G TOCSY spectrum:** TOCSY spectrum of CaM-CD2-IV-5G was recorded using Varian 600 MHz NMR. The spectrum was collected in 10 mM Tris 10 mM KCl (pH 7.4) buffer at 25 °C.



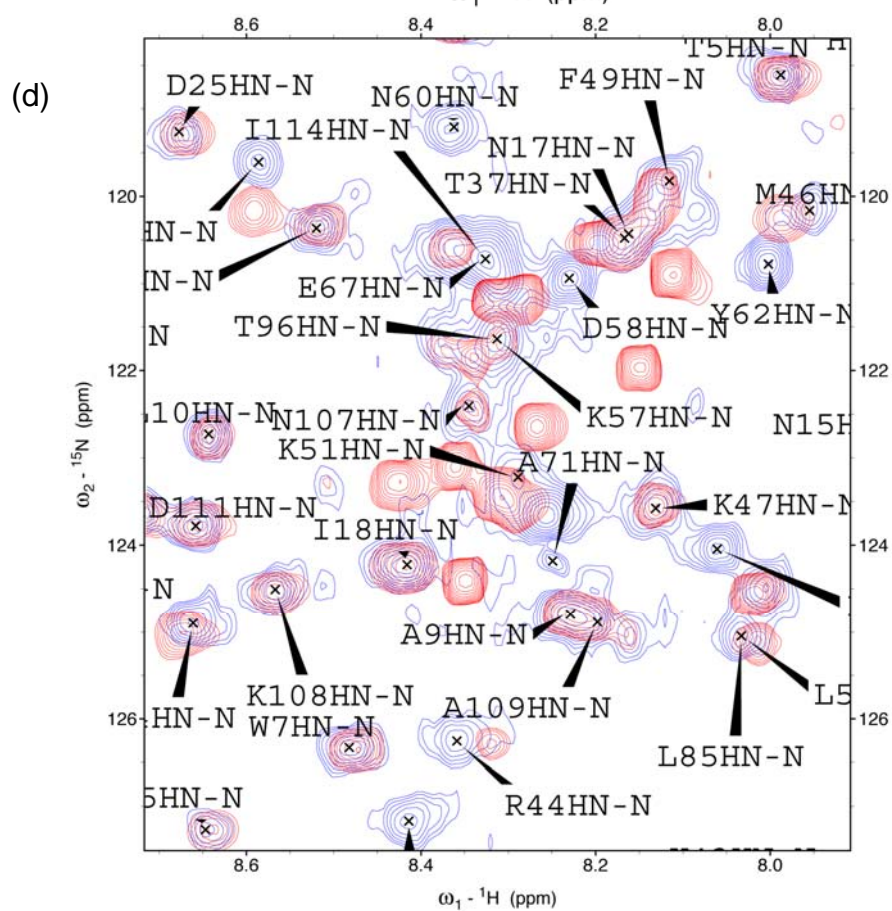
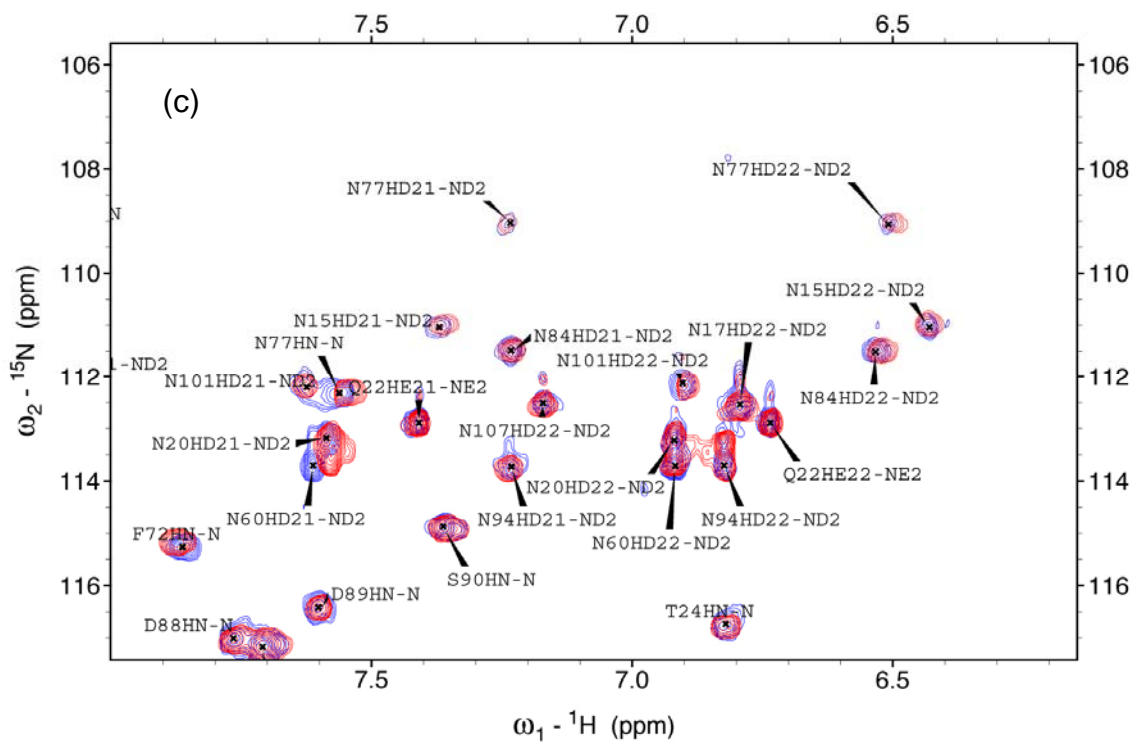
**Figure 3.11b Assigned  $^{15}\text{N}$  HSQC spectrum of CaM-CD2-IV-5G:** HSQC spectrum of CaM-CD2-IV-5G was recorded using Varian 600 MHz. The experiments were performed using 500 MHz NMR at 25 ° C. The protein sample was prepared in 20 mM PIPES 20 mM KCl (pH 6.8).



**Figure 3.12 Amide region of  $^1\text{H}$  spectra of CD2, CaM-CD2-III-5G, and CaM-CD2-IV-5G:**  $^1\text{H}$  spectra of CD2, CaM-CD2-III-5G, and CaM-CD2-IV-5G which indicated that the EF-loop III has different structural properties than EF-loop IV.

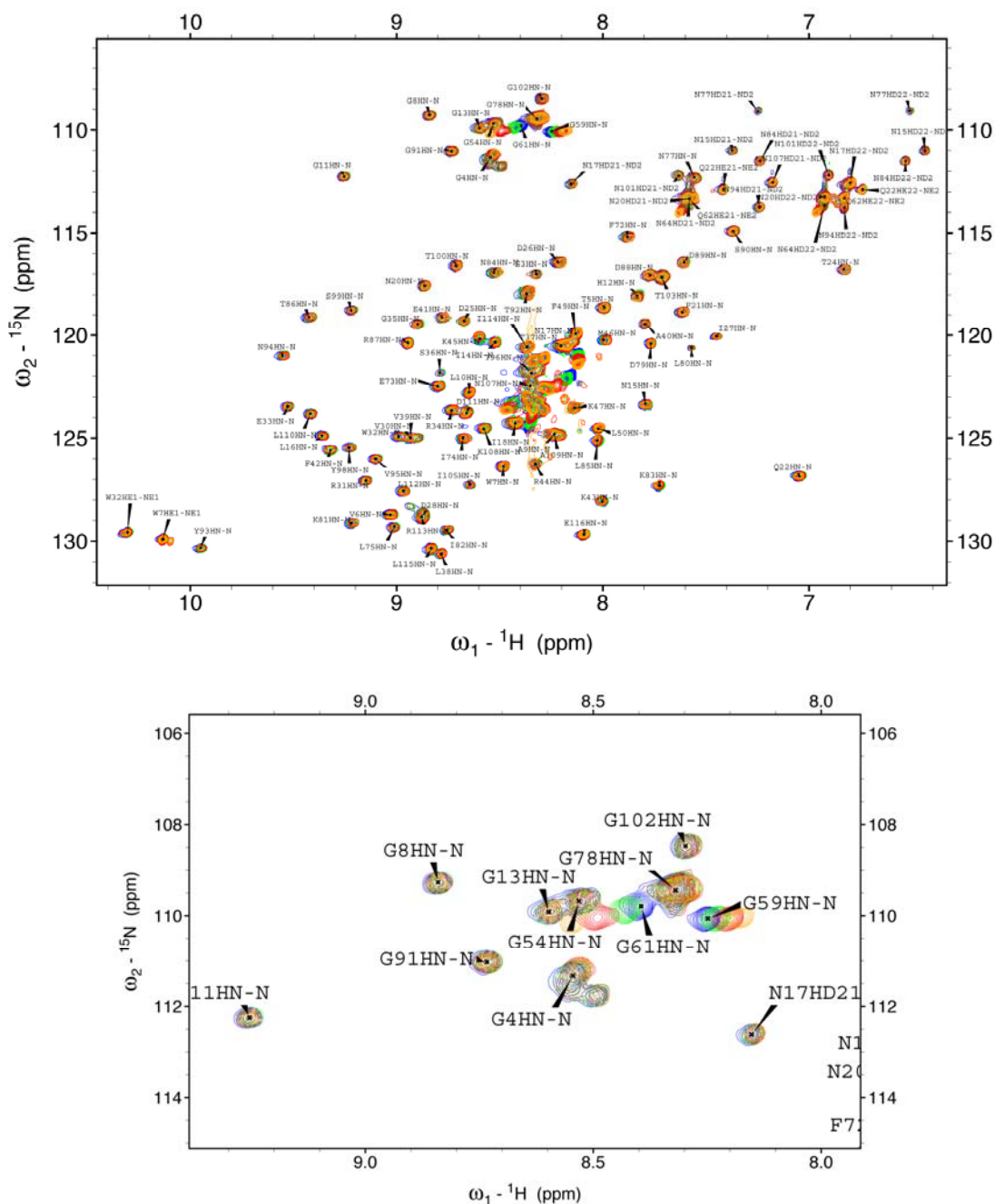


**Figure 3.13  $^{15}\text{N}$  HSQC spectra comparison between CaM-CD2-III-5G and CaM-CD2-IV-5G:** (a) The HSQC spectra of the CaM-CD2-IV-5G (red) is overlay on top of CaM-CD2-III-5G (blue). The HSQC spectrum was collected in the presence of 1 mM EGTA in 20 mM PIPES 10 mM KCl (pH 6.8) buffer at 25 °C. (b) Gly region of HSQC spectra. (c) Asn and Gln sidechain region of HSQC spectra. (d) Location of the Ca(II) binding ligand for both CaM-CD2-III-5G and CaM-CD2-IV-5G.



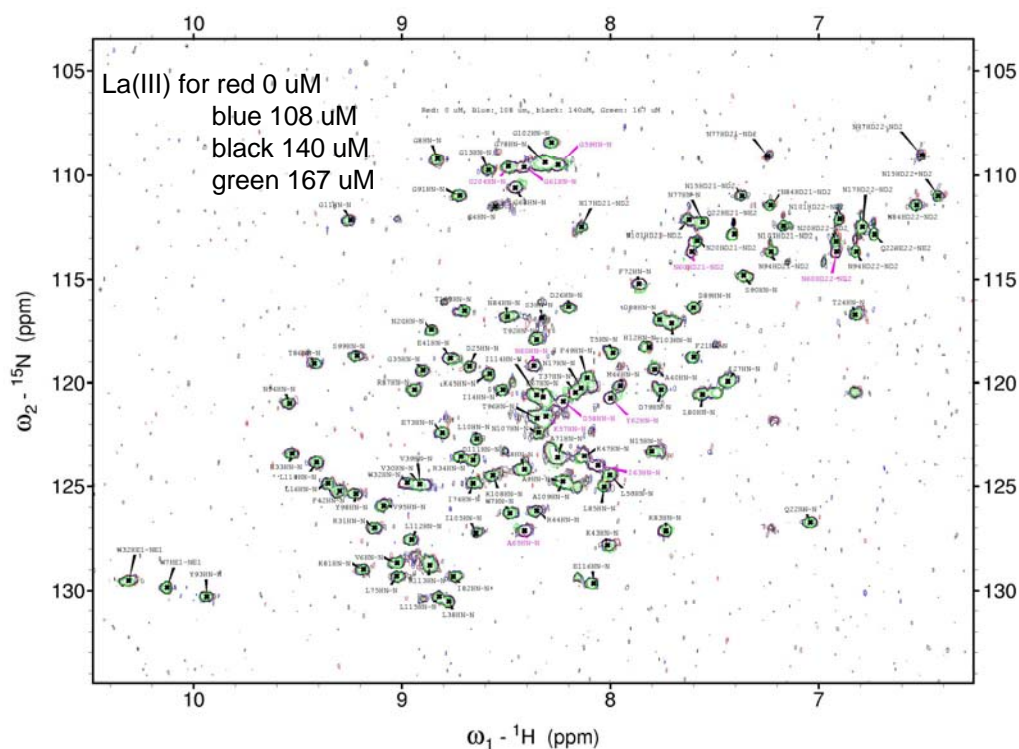
See previous page for explanation.

**Figure 3.14  $^{15}\text{N}$  HSQC Ca(II) titration for CaM-CD2-III-5G:** The  $^{15}\text{N}$  HSQC spectra with 0, 140, 194, 302 and 1026  $\mu\text{M}$  of calcium are shown as blue, cyan, magenta, yellow, and red color, respectively. The  $^{15}\text{N}$  HSQC spectra are stacked onto the first point of the titration. The experiments were performed using 500 MHz NMR at  $25^\circ\text{C}$ . The protein sample was prepared in 20 mM PIPES 20 mM KCl (pH 6.8).

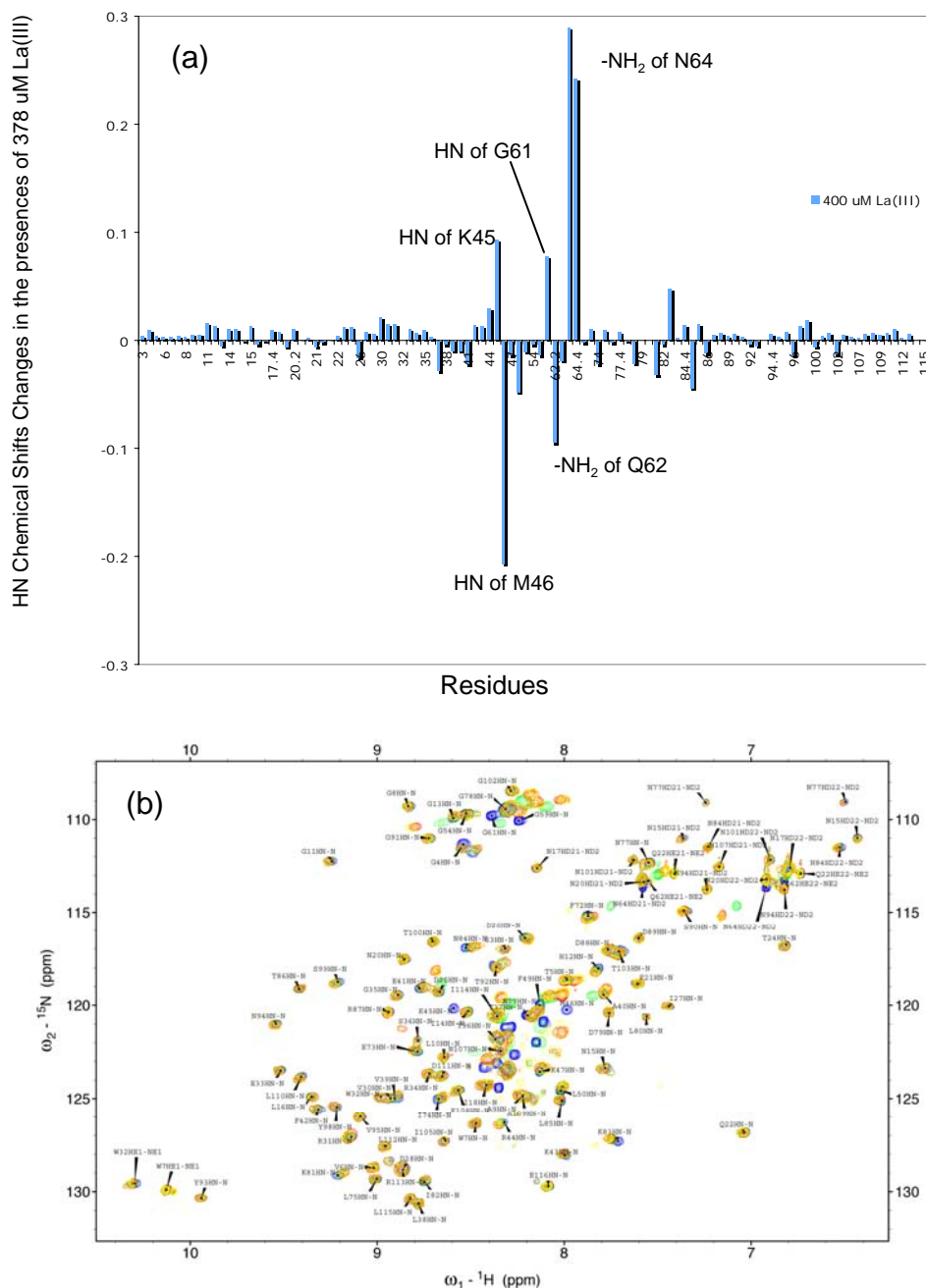


**Figure 3.15**  $^{15}\text{N}$  HSQC Ca(II) titration for CaM-CD2-IV-5G: The  $^{15}\text{N}$  HSQC spectra with 0, 0.51, 4.36, and 10.32 mM of calcium are shown as blue, green, red, and yellow color, respectively. The  $^{15}\text{N}$  HSQC spectra are stacked onto the first point of the titration. The experiments were performed using 500 MHz NMR at 25 ° C. The protein sample was prepared in 20 mM PIPES 20 mM KCl (pH 6.8).



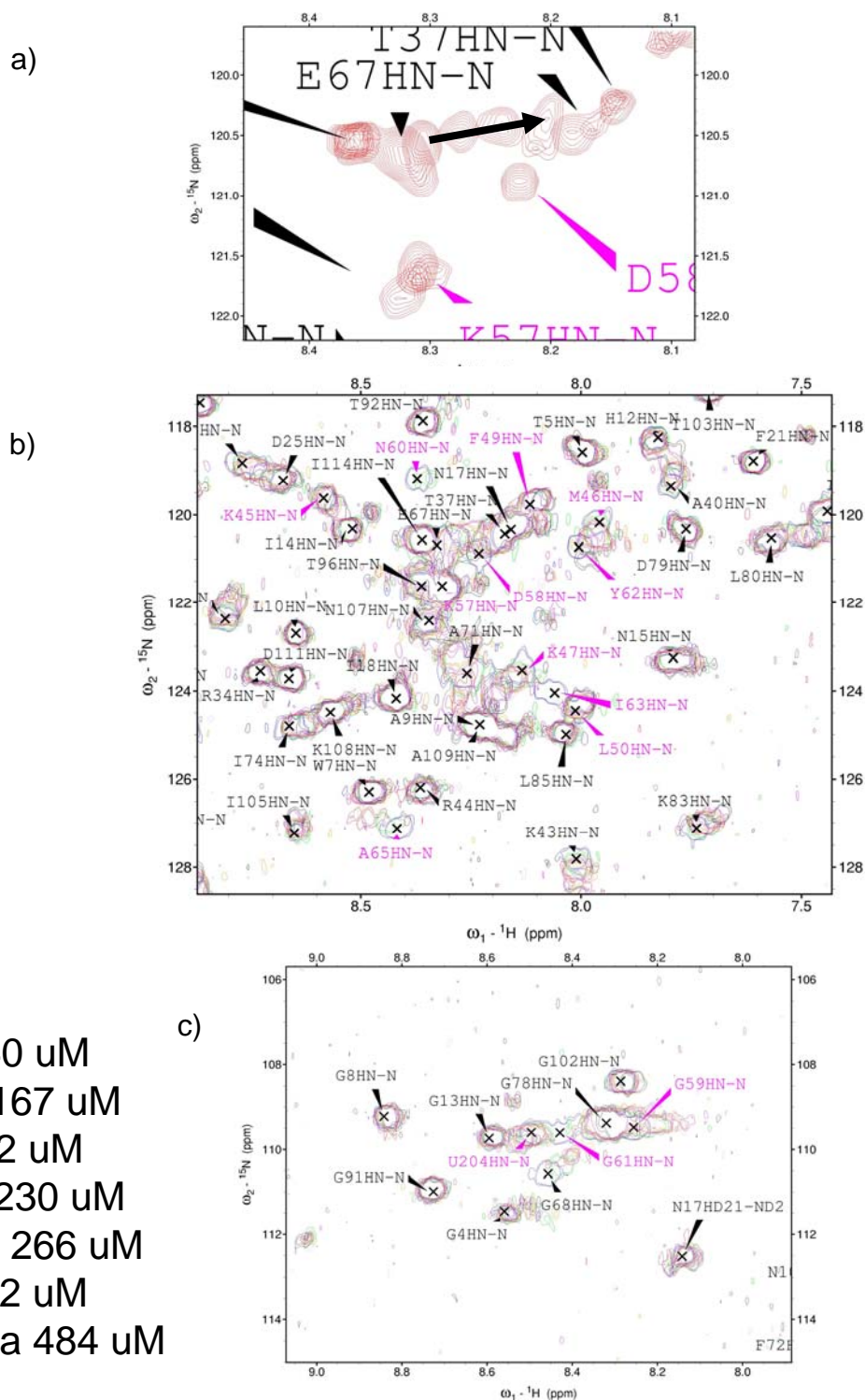


**Figure 3.16  ${}^{15}\text{N}$  HSQC La(III) titration for CaM-CD2-III-5G:** The HSQC spectra with La(III) concentration of 0, 108, 140, and 167  $\mu\text{M}$  are shown in red, blue, black, and green respectively. The experiments were performed using 500 MHz NMR at 25  $^\circ\text{C}$ . The protein sample was prepared in 20 mM PIPES 20 mM KCl (pH 6.8).

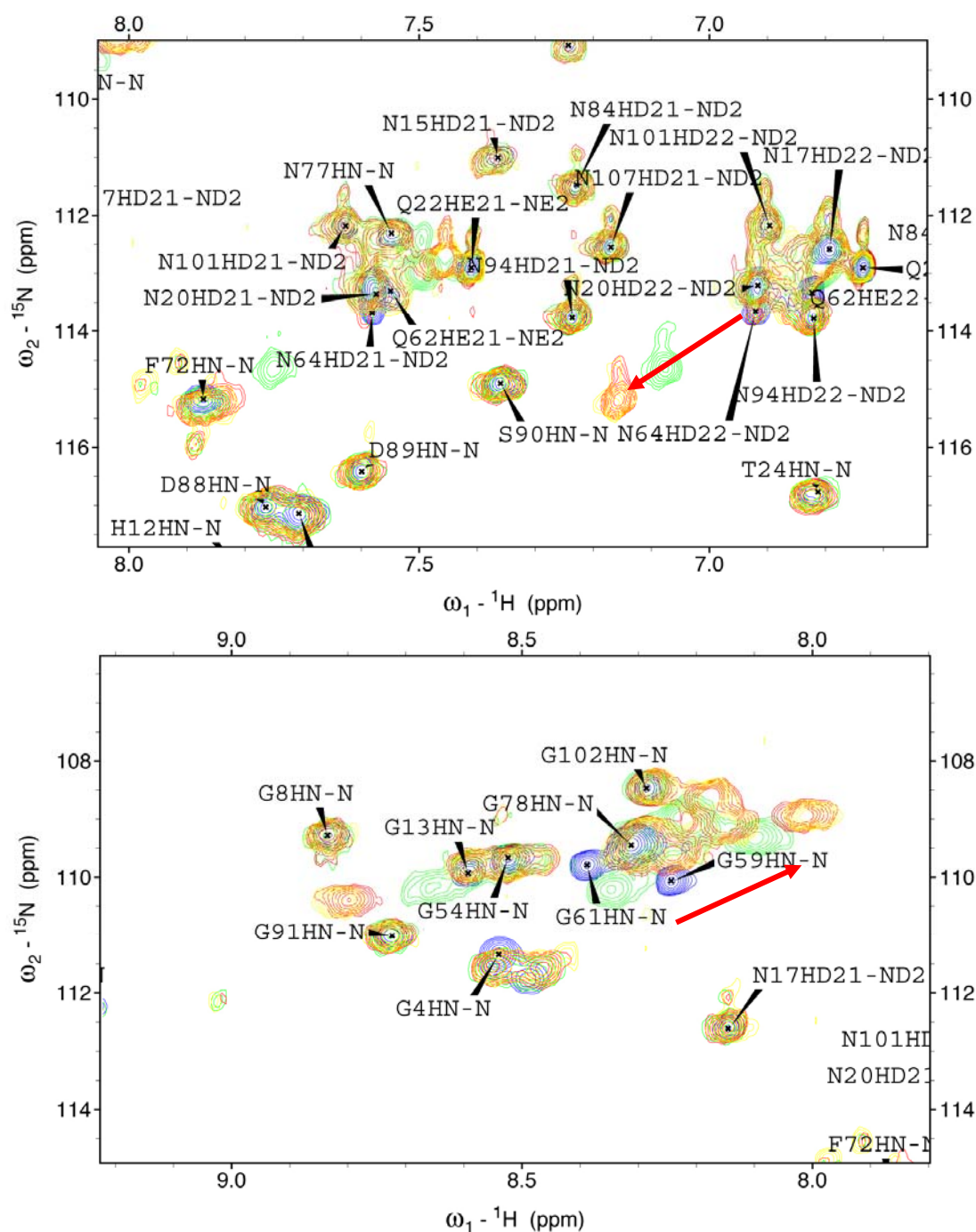


**Figure 3.17 <sup>15</sup>N HSQC La(III) titration for CaM-CD2-IV-5G:** (a) The HN chemical shifts changes in the presence of 378 uM La(III). (b) The <sup>15</sup>N HSQC spectra with 0, 162, 216, and 378 uM of La(III) are shown as blue, green, red, and yellow color, respectively. The <sup>15</sup>N HSQC spectra are stacked onto the first point of the titration. The experiments were performed using 500 MHz NMR at 25 ° C. The protein sample was prepared in 20 mM PIPES 20 mM KCl (pH 6.8).

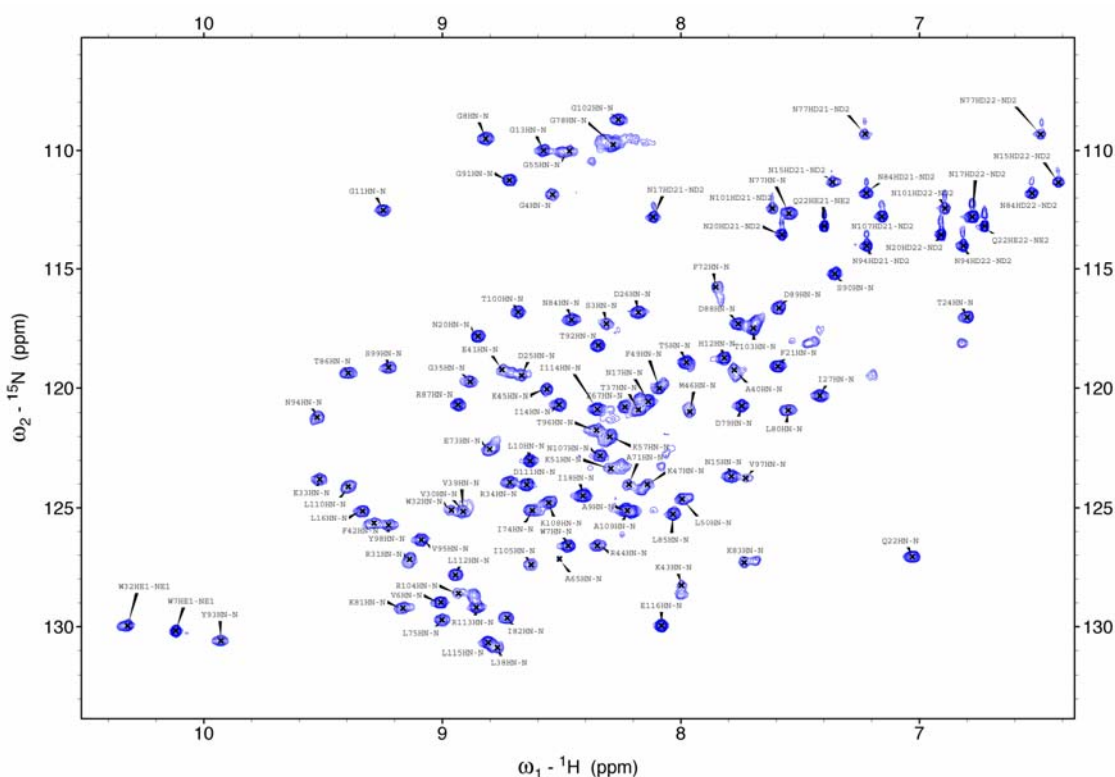




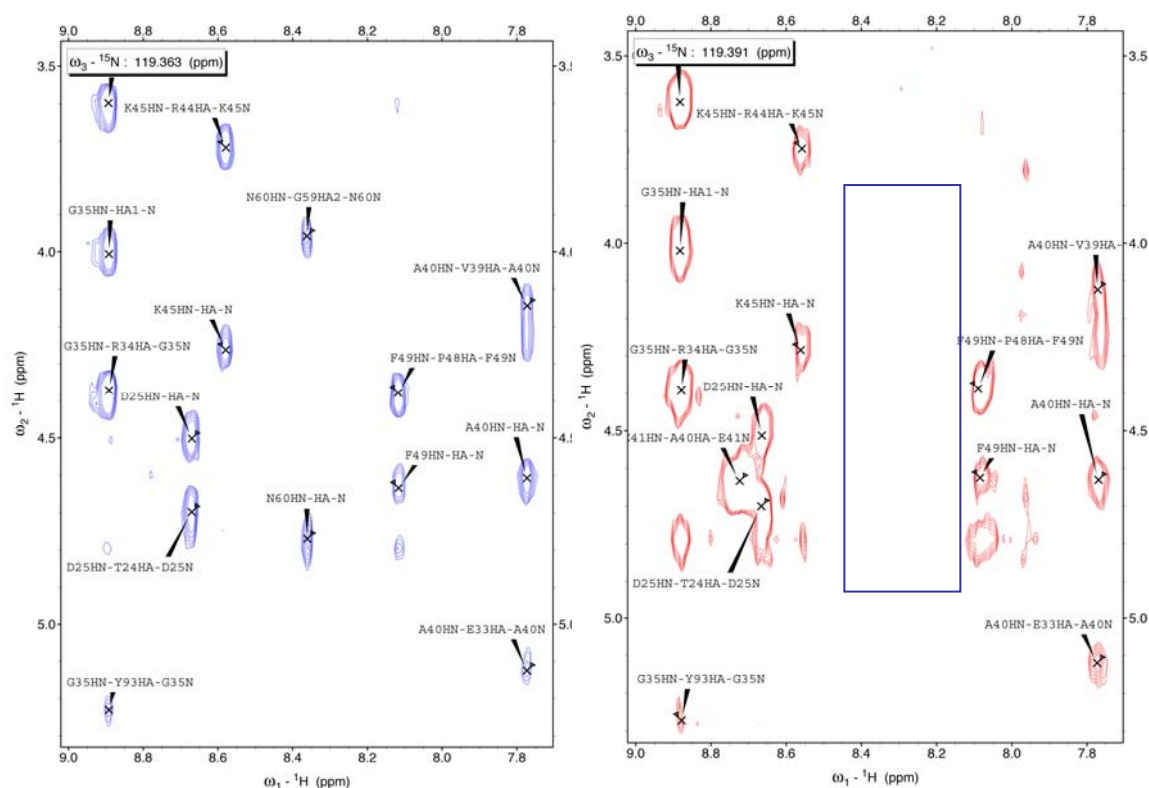
**Figure 3.18 EF-loop III region of the La(III) Titration:** The HSQC spectra with 140, 167, 192, 230, 266, 302, and 484  $\mu\text{M}$  are shown in blue, green, red, purple, orange, pink, and magenta, respectively. The experiments were performed using 500 MHz NMR at 25 ° C. The protein sample was prepared in 20 mM PIPES 20 mM KCl (pH 6.8).



**Figure 3.19 EF-loop IV region of the La(III) titration:** The  $^{15}\text{N}$  HSQC spectra with 0, 162, 216, and 378  $\mu\text{M}$  of La(III) are shown as blue, green, red, and yellow color, respectively. The  $^{15}\text{N}$  HSQC spectra are stacked onto the first point of the titration. The experiments were performed using 500 MHz NMR at 25 ° C. The protein sample was prepared in 20 mM PIPES 20 mM KCl (pH 6.8).



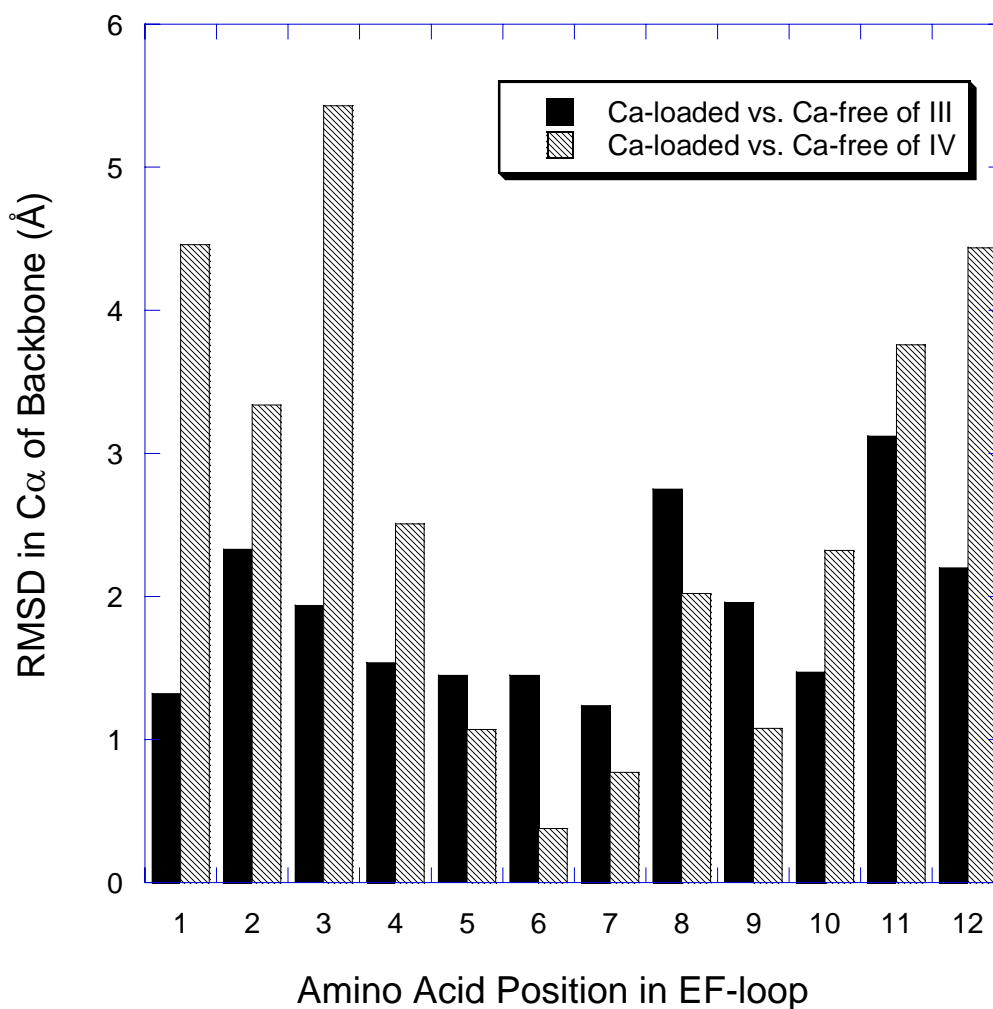
**Figure 3.20**  $^{15}\text{N}$  HSQC spectrum of CaM-CD2-III-5G in the presence of 2 mM La(III): This spectrum was recorded using Varian 600 MHz in 20 mM PIPES 20 mM KCl (pH 6.8) at 25 °C.



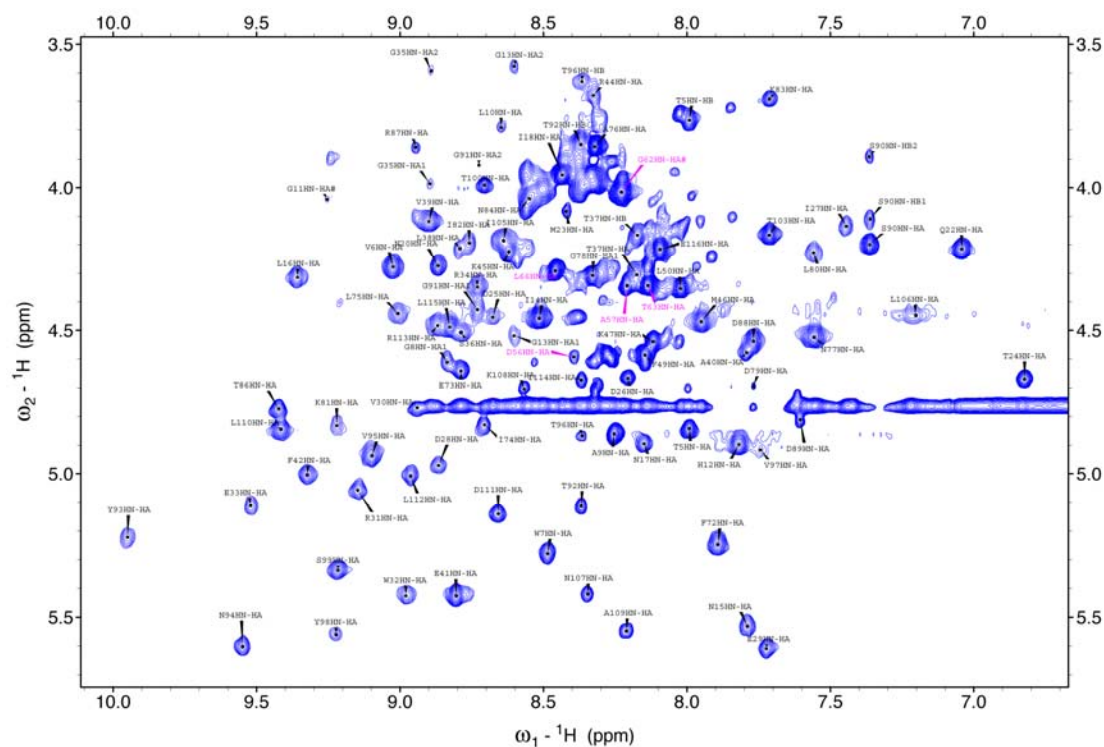
**Figure 3.21 3D NOESYHSQC strip plot of CaM-CD2-III-5G:** The spectrum collected in the presence of 1 mM EGTA is shown in blue and the spectrum was collected using Varian 800 MHz NMR at 25 °C. The spectrum collected in the presence of 2 mM La(III) is shown in red and the spectrum was collected using Varian 600 MHz NMR at 25 °C. Both samples were prepared in 20 mM PIPES 20 mM KCl (pH 6.8). The crosspeaks for N60 disappeared in the presence of 2 mM La(III).

Loop III: **D-K-D-G-N-G-Y-I-S-A-A-E**

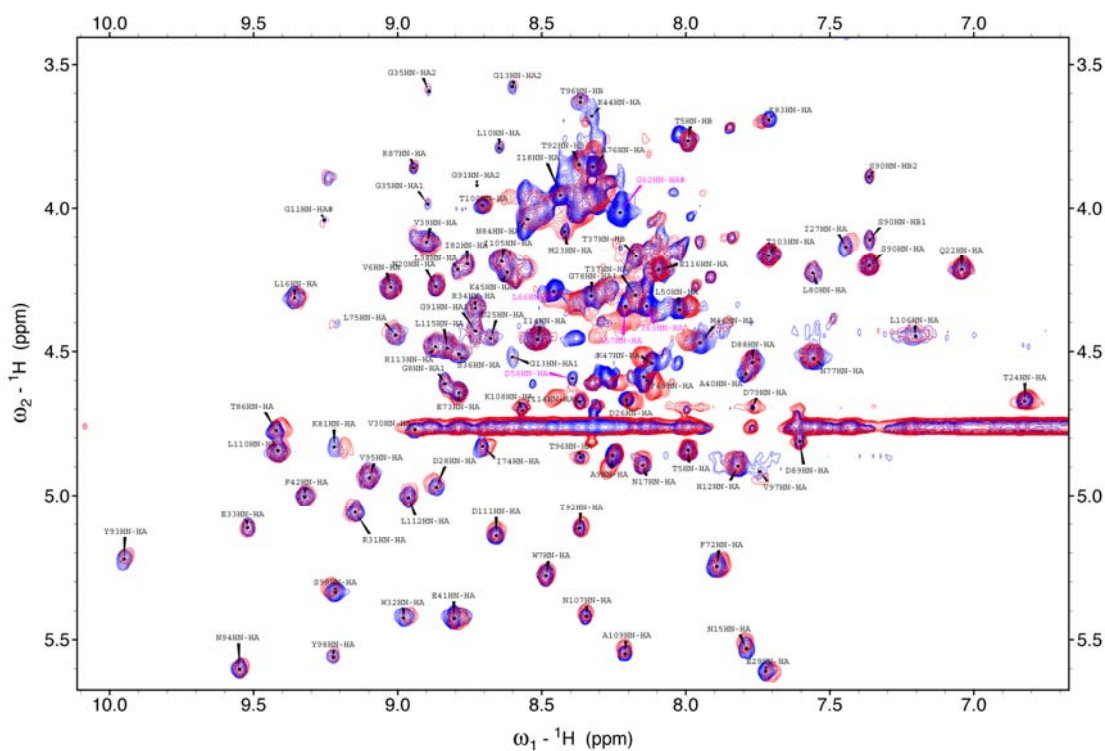
Loop IV: **D-I-D-G-D-G-Q-V-N-Y-E-E**



**Figure 3.22 Structural comparison between Ca(II) free and Ca(II) loaded of CaM:** The RMS deviation between the Ca(II) free form (1CFC) and Ca(II) loaded form (3CLN) of C-terminal domain of CaM. The EF-loop IV has large structural deviation between the Ca(II) free and Ca(II) loaded form in comparison with EF-loop III.

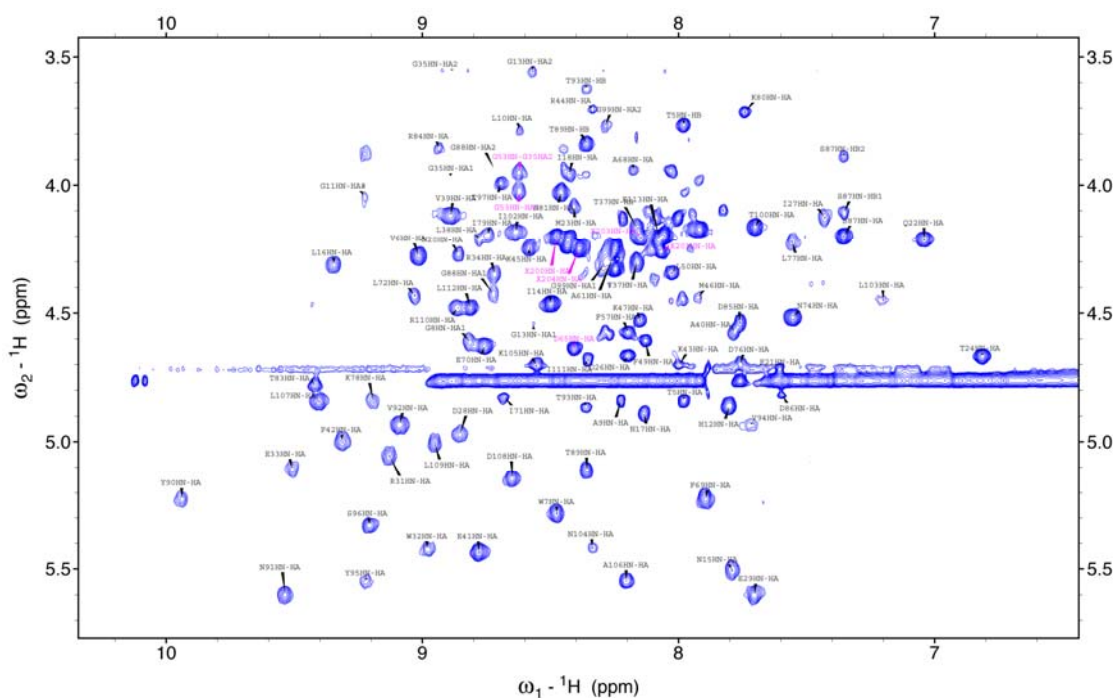


**Figure 3.23 Fingerprint region of the Rub-CD2-5G TOCSY spectrum:** The spectrum was recorded using Varian 600 MHz at 25 °C. The protein was prepared in the presence of 1 mM EGTA in 20 mM PIPES 20 mM KCl at pH 6.8.



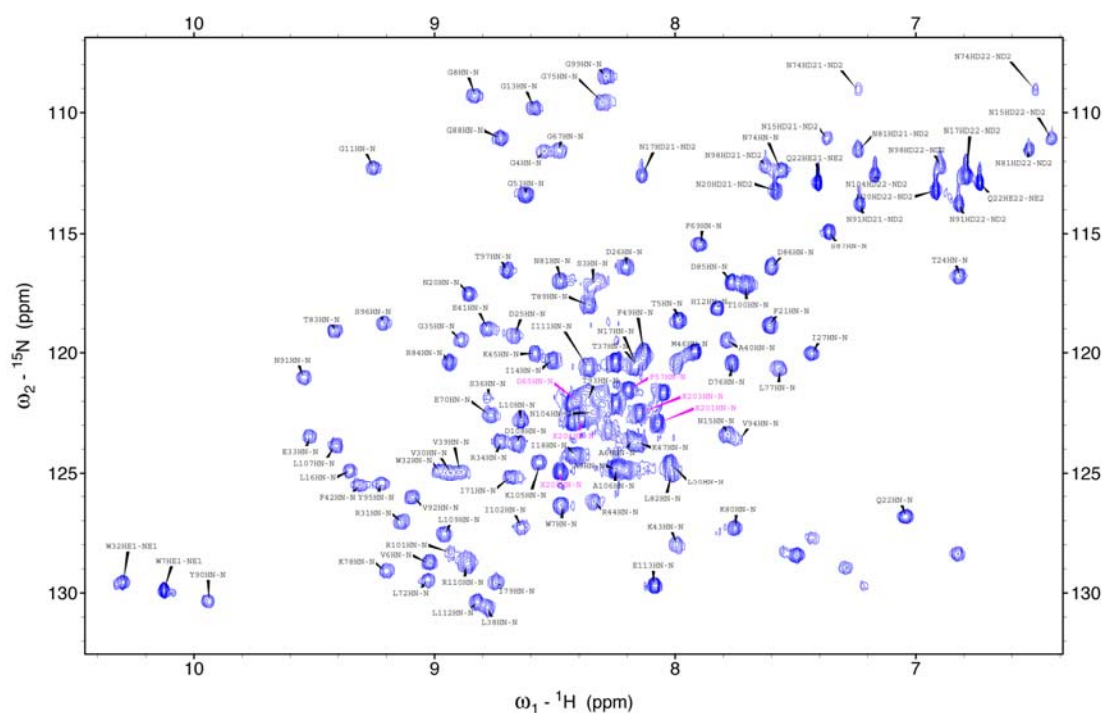
**Figure 3.24 TOCSY spectra of Rub-CD2-5G in the presence of La(III) and EGTA:** The crosspeaks of Rub-CD2-5G in 1 mM La(III) is showing in red. The crosspeaks Rub-CD2-5G in 1 mM EGTA is showing in blue.





**Figure 3.25 Fingerprint region of the CaR-CD2-III-0G TOCSY spectrum:** This spectrum was collected using Varian 600 MHz at 25 °C. The protein sample was prepared in the presence of 1 mM EGTA in 20 mM PIPES 20 mM KCl at pH 6.8.





**Figure 3.26  $^{15}\text{N}$  HSQC spectrum of CaR-CD2-III-0G:** This spectrum was recorded using Varian 500 MHz NMR at 25 °C. The protein was prepared in the presence of 1 mM EGTA in 20 mM PIPES 20 mM KCl at pH 6.8.

**Figure 3.27 Comparison between the  $^{15}\text{N}$  HSQC spectra of CaR-CD2-III-0G to CaM-CD2-III-5G:** The crosspeaks of CaM-CD2-III-5G are showing in red. The crosspeaks of CaR-CD2-III-0G are showing in blue. Both spectra were recorded using 600 MHz NMR at 25 ° C. Both samples were prepared in 20 mM PIPES 20 mM KCl at pH 6.8.

## **Chapter 4.0 Structural Determination of Engineered Proteins Grafted with EF-hand Motifs of Calmodulin using Heteronuclear NMR**

To determine whether or not the isolated EF-hand motif in CD2 still forms the native pentagonal bipyramidal geometry observed in the intact EF-hand protein, we used NMR to determine the solution structure of the engineered protein. The traditional method of solving protein structure relies on NOE distance constraints and dihedral angle constraints obtained from *J*-coupling or chemical shift analysis. In order to obtain the NOE distance constraints, the chemical shift assignment of the protein is required. The homonuclear assignment process usually begins by identification of the spin systems in COSY or TOCSY spectra and then this information plus the connectivity information from NOESY spectra are used to perform a sequential assignment. The spin system identification and sequential assignment can also be completed using triple resonance experiments with  $^{13}\text{C}$  and  $^{15}\text{N}$ -labeled protein. The triple resonance assignment strategy has fewer problems with crosspeak overlap so the assignment can be achieved at a much faster pace and the carbon chemical shifts can be used to estimate  $\phi$  and  $\psi$  dihedral angle constraints. The NOE peak intensity is correlated to the distance between two nuclei. The distance restraints are calculated based on the peak intensities of two and three dimensional NOESY experiments. The dihedral angles restraints are constructed using the experimental data from the HNHA experiment and

predicted dihedral angles from TALOS. The solution structure of the engineered protein is then calculated using CYANA and CNS algorithms.

In recent years, residual dipolar couplings have become an important tool for structure calculation. Residual dipolar coupling can be translated into an angle  $\theta$ , which describes the orientation of an intermolecular vector (for example,  $^1\text{H}$ - $^{15}\text{N}$ ) to the applied magnetic field in a partially-ordered system. For solution NMR in an isotropic condition, the residual dipolar coupling is usually averaged to zero due to the molecular tumbling. The residual dipolar coupling can be obtained by partially aligning the protein to a specific orientation to the magnetic field. This alignment effect can be achieved by using an internal alignment with a paramagnetic moiety or an external alignment to restrict the orientation of a protein by charge or space.

The field-induced alignment method using lanthanide ions has been pioneered by Dr. Bertini (140). Paramagnetic lanthanides are capable of orienting the protein in high magnetic fields to an extent similar to that obtained by using external orienting media, and each lanthanide orients according to its magnetic susceptibility tensor. Calcium analogs (Ce(III), Tb(III), Ho(III), Er(III), Tm(III), and Yb(III)), have been used to substitute for calcium in calcium-binding proteins, such as the EF-hand proteins calbindinD9K and calmodulin, to obtain  $^1\text{H}$ - $^{15}\text{N}$  residual dipolar couplings for solution structural determinations (141-144). Furthermore, the paramagnetic properties of the lanthanides also produce additional useful information, such as the pseudocontact shifts (142). The affect

of the pseudocontact shifts is distance-dependent between the metal ions and the observed nuclei. Here, the field-induced alignment using the lanthanides methods were used to obtain the dynamics and structural properties of CaM-CD2-III-5G.

Residual dipolar coupling using external orientation media such as bicelles or phages can be easily measured from NMR spectra of the isotropically labeled protein. This method was pioneered by Drs. James Prestegard and Ad Bax. The structural information can be obtained with or without a limited number of NOE constraints (69, 121, 145-147). In this study, residual dipolar couplings for CaM-CD2-III-5G were measured using PEG-bicelles (3.3% (w/v) C12E5-hexanol bicelles resulting in 11-15 Hz of residual  $^2\text{H}$  water splitting) to provide additional orientation restraints.

In Chapter 3.0, we reported the development of the grafting approach and investigation of metal binding and conformational properties of EF-hand motifs grafted in CD2 using homonuclear NMR. In this chapter, first, the complete assignment of CaM-CD2-III-5G using homonuclear and heteronuclear NMR will be reported. Second, our studies in structural determination of the protein are shown. Third, RDC studies of the engineered protein using external (PEG alignment medium) and internal (paramagnetic) alignment methods are discussed. Finally, the host protein region and the EF-loop region of the solved structure of CaM-CD2-III-5G will be compared in detail.

## **4.1 Assignment of CaM-CD2-III-5G Using Homonuclear and Heteronuclear NMR Experiments**

### **4.1.1.1 Homonuclear Sequential Assignment of CaM-CD2-III-5G**

The sequential assignment for CaM-CD2-III-5G was initiated using homonuclear experiments such as TOCSY and NOESY. The fingerprint region (6.5 to 10.5 ppm in the D1 and 3.0 to 6.0 ppm in the D2 dimension) of the TOCSY spectrum is shown in Figure 4.1 along with the TOCSY spectrum of the wild type CD2. The fingerprint region shows the correlations between the backbone amide protons and the backbone  $\alpha$  protons. Under ideal conditions, each amino acid in the protein would show one crosspeak in this region (except for Proline (Pro), Glycine (Gly), Serine (Ser), and Threonine (Thr)). Pro does not have crosspeaks at the fingerprint region and Gly has two  $\alpha$  protons ( $H_{\alpha 1}$  and  $H_{\alpha 2}$ ), so there will be two crosspeaks observed in the fingerprint region. Ser and Thr have hydroxyl group sidechains; the chemical shifts for the  $\beta$  proton resonances generally move downfield into the fingerprint region. Ser has 3 resonances ( $H_{\alpha}$ ,  $H_{\beta 1}$ , and  $H_{\beta 2}$ ), while Thr has 2 ( $H_{\alpha}$  and  $H_{\beta}$ ). Among the 116 residues of CaM-CD2-III-5G, there are 14 Gly, 5 Ser, 8 Thr, and 2 Pro. In principle, the fingerprint region of CaM-CD2-III-5G would have 146 crosspeaks. However, the terminal residues and residues with fast exchange rates are sometimes not observed. Thus 130 crosspeaks are observed in the fingerprint region of CaM-CD2-III-5G. The well-resolved crosspeaks in the CaM-CD2-III-5G spectrum are very similar to those in the wild type CD2 spectrum.

The chemical shifts of residues with high flexibility are usually very similar to those of the residues in random coil structures observed on short peptides in solution (148). The random coil shift for HN is between 7.5 ppm to 8.5 ppm and for H $\alpha$  is between 3.0 to 4.6 ppm. The majority of the residues in the inserted glycine linkers and the inserted EF-hand loop III of calmodulin as well as the flexible areas of CD2 have chemical shifts in this region.

#### 4.1.1.2 Homonuclear Sequential Assignment for the Host Protein

The sequential assignment for CaM-CD2-III-5G was performed using the assignment protocol by Redfield *et al* (a non-published heteronuclear assignment protocol was given by Dr. Christina Redfield at Oxford University) and Wuthrich *et al* (149, 150). The identification process began by classifying the crosspeaks into different categories using the TOCSY spectrum. The spin systems were then sequentially identified by assigning the NOE crosspeak between the sequential residues in the fingerprint region (connectivity between HN<sub>*i*</sub> to H $\alpha$ <sub>*i-1*</sub>). Once all of the mainchain assignments were completed, the sidechain inter-residue interactions were assigned.

Figure 4.2 shows an example for the sequential assignment using the residues G91 to N94 (G74 to N77 of CD2) of CaM-CD2-III-5G.

----- G91 – T92 – Y93 – N94 -----

Asp, Asn, Trp, Tyr, Phe, and His are usually referred to as type J residues in which the two H $\beta$  crosspeaks are observed in the range 2.5 to 3.2 ppm. Asn,

Trp, Tyr, Phe, and His can be identified by looking for the NOE connectivity between H $\beta$  protons to each amino acid's sidechain proton (NH<sub>2</sub> for the Asn and aromatic ring proton for the others). This sidechain assignment process utilizing the hydrogen exchange properties of each amino acid will be discussed in the following section. The chemical shift of the HN and H $\alpha$  protons for N77 are 9.54 and 4.60 ppm, respectively. The H $\beta$ 1 and H $\beta$ 2 of the N94 merge into a single crosspeak at 2.51 ppm (H $\beta^*$  will be used for sidechains that merge to one crosspeak). The chemical shifts for HD21 and HD22 protons of N94 are at 7.23 and 6.82 ppm, respectively. The identity of N94 was further verified by the NOE crosspeak observed between HD22 and H $\beta^*$ , and the backbone sequential NOE showed an interaction between the HN<sub>*i*</sub> and H $\alpha_{i-1}$  protons with a mixing time of 150 ms. The inter- and intra- NOE interactions can be verified by comparing them to the TOCSY spectra. TOCSY identifies connectivity through bonds while the NOESY experiment detects connectivity through space. Both the TOCSY and NOESY spectra show the crosspeaks for intra-residue connectivity while the inter-residue connectivity only appears in NOESY spectra. The crosspeaks for intra-residue connectivity may not always be recorded in one TOCSY spectrum. To ensure maximum intra-residue connectivity, TOCSY spectra with various mixing times were used (from 36 to 90 ms). By following the HN chemical shift of N94, a crosspeak was found at 5.21 ppm on the D2 dimension, which is the H $\alpha$  chemical shift of Y93. This H $\alpha$  chemical shift has a crosspeak at 9.94 ppm, which is the HN chemical shift of Y93. At D1 = 9.94 ppm, a pair of side chain



crosspeaks are observed at D2 = 3.12 and 3.00 ppm, which correspond to H $\beta$ 1 and H $\beta$ 2, respectively. The HN proton of N94 also has NOE crosspeaks with these two sidechain protons. This is a type J spin system. Among the 9 Asn residues in CaM-CD2-III-5G (I14-N15, L16-N17, P19-N20, G59-N60, A76-N77, Y93-N94, T100-N101, and L106-N107), only N94 has a type J residue (Y93) in the previous position of the sequence. Therefore, the crosspeaks with HN chemical shifts of 9.94 correspond to Y93. The sidechain NOE crosspeaks from HD# protons of the aromatic ring also indicates that the residue is Tyr. The residue preceding Y93 is T92 and the sequential NOE (NOESY walk) points to a crosspeak with HN and H $\alpha$  chemical shifts of 8.36 and 5.12 ppm, respectively. The spin system at 8.36 ppm has sidechain crosspeaks at 3.84 and 1.15 ppm. Thr has a hydroxyl group in the HG position, so the H $\beta$  shifts approximately 3.50 to 4.20 ppm. The HG sidechain protons of Thr are incorporated into one crosspeak that is usually observed around 1.50 to 0.80 ppm. The spin system for the residue preceding Y93 has the crosspeak pattern of Thr. The inter-residue NOE between the HN of Y93 to the sidechain protons of T92 are also observed. Therefore, the three crosspeaks observed for D1 = 8.36 ppm, H $\alpha$  = 5.12, H $\beta$  = 3.84, and HG = 1.15 ppm, were assigned to T92. The HN proton of T92 has two crosspeaks at 4.44 and 3.92 ppm. These two crosspeaks also link to two crosspeaks observed at D1 = 8.73 ppm. No sidechain crosspeak has been found; therefore this spin system was assigned as G91. Since the NOESY experiment indicates that the distance between two nuclei is within 5 Å, the

information for the tertiary structure of the protein can be obtained from the non-sequential NOE crosspeak and will be discussed in section 4.2.6.

#### **4.1.1.3 Aromatic Ring Assignment under D<sub>2</sub>O Conditions**

The aromatic ring protons in Tyr, Phe, and Trp (except the HE1 proton) are non-exchangeable protons while the NH<sub>2</sub> group of Asn and Gln are exchangeable protons (Figure 5.1). In the NOESY spectrum, the assignment for the crosspeaks between the H $\beta$  proton and the aromatic ring protons of the aromatic residues can be obtained without the interference of HN-HN, HN-H $\alpha$ , and HN-H $\beta$  crosspeaks.

The resonances for the aromatic proton and the sidechain NH<sub>2</sub> groups of Asn and Gln are observed around 6.5 to 8.0 ppm in both D1 and D2 dimension of the TOCSY and NOESY spectra. An example of the NOESY spectrum with Asn sidechain assignments is shown in Figure 4.3. The chemical shifts of the HN and H $\alpha$  protons for N15 are 7.79 and 5.52 ppm, respectively. The chemical shifts for H $\beta$ 1 and H $\beta$ 2 of N15 are 2.55 and 2.15 ppm, respectively. The chemical shifts for HD21 and HD22 protons of N15 are at 7.36 and 6.43 ppm, respectively. The HD21 and HD22 protons of N15 both show intra-residue NOE crosspeaks with H $\beta$ 1 and H $\beta$ 2. The HD21 and HD22 protons of N15 also have inter-residue NOE crosspeaks to the sidechain of I14 that is further confirmed by the assignment of N15. The aromatic ring protons of Tyr 93, Tyr 98, Phe21, Phe 42, Phe 49, Phe 72, Trp 7, and Trp 32 were all assigned. The assigned aromatic region of the

TOCSY spectra recorded in H<sub>2</sub>O and D<sub>2</sub>O is shown in Figure 4.4. The TOCSY spectrum recorded in H<sub>2</sub>O also includes the assignments for the exchangeable protons of Asn and Gln.

#### **4.1.1.4 Discussion of the Homonuclear Assignment**

The resonances for 80 residues have been assigned based on the homonuclear assignment strategy, which is only 69% of the total number of residues in CaM-CD2-III-5G. The assigned region of CaM-CD2-III-5G using the homonuclear assignment strategy is highlighted on the calculated structure (Figure 4.5). The assigned regions are from T5 to F42 and F72 to E116 except R1, D2, S3, P19, K43 to A71, A76, K83, and N101. The first assigned stretch (T5 to F42) contains the residues in the  $\beta$ -strands A, B, and C. The second assigned stretch (F72 to E116) contains the residues in the  $\beta$ -strands E, F, and G. The first three residues at the N-terminus (R1, D2, and S3) were not assigned. The crosspeak for N101 has not been observed in the fingerprint region. Among the unassigned residues, K43 to S52 are part of the host protein. In wild type CD2, this region is a flexible loop with a short  $\beta$ -strand formed at F49 and L50. The chemical shifts of crosspeaks for K43 to S52 are all between 7.9 to 8.6 ppm in the D1 and 3.7 to 4.8 in the D2 dimension of the fingerprint region. In the spectrum of CaM-CD2-III-5G, the crosspeaks for the inserted glycine linkers and the EF-loop III (G53 to G70) observed in this region overlap with the crosspeaks from K43 to S52. There are 27 unassigned residues in this region.

The sequential assignment of the inserted EF-loop III of calmodulin in CD2 is almost impossible using only homonuclear experiments due to the chemical shift degeneracy. The resolution may be improved by either changing the pH or temperature or by isotopically labeling the protein.

#### **4.1.1.5 The Effect of Temperature and pH on the NMR Spectra**

The chemical shift is sensitive to the protein environment. By changing the temperature, the protein environment will change, especially in the solvent exposed regions. The resonances of CaM-CD2-III-5G are shifted at different temperatures, which assists the assignment process. Reducing the temperature to 4 °C, however, decreases the S/N ratio of the experiment because the effective rotational correlation time of the protein increases. The linewidth of the NMR resonance is dependent on the  $T_2^*$  relaxation time of the nuclei, which is dependent on the correlation time. If the protein rotational rate decreases, the linewidth of the peak should increase, hence leading to the loss of sensitivity. Increasing the temperature would increase the exchange rate of HN protons with solvent, so it would also lead to the loss of sensitivity and at the same time decrease the shelf life of the protein.

Protein NMR spectroscopy generally suffers from low sensitivity because the concentration of the protein sample usually cannot be above 1 mM (indicated in the case of CaM-CD2-III-5G). At higher concentration, macromolecules such as proteins have an increasing tendency to oligomerize or precipitate.

Furthermore, concentrating the protein also concentrates proteases in the solution, which may lead to more rapid protein degradation. Since increasing the concentration of the protein is not feasible, the loss of sensitivity cannot be overcome; therefore, lowering the temperature is not possible. So, improving the assignment procedure by altering temperature is not practical either.

The homonuclear NMR experiments were carried out in 10 mM Tris and 10 mM KCl at pH 7.4. The proton exchange with water is very fast at this pH (Figure 4.6). The ideal pH for metal binding studies would be near physiological pH, approximately 7.0. Lowering the pH to 5.0 will slow down the exchange rates for the solvent-exchangeable hydrogens of the protein and more resonances can be observed; however, at this low pH proteins might lose time metal binding ability. For the purpose of improving the assignment, a CaM-CD2-III-5G protein sample with a concentration of 1 mM was prepared in 20 mM phosphate buffer at pH 5.0. The TOCSY spectrum shows a mixture of resonances from folded and unfolded CaM-CD2-III-5G (Figure 4.7). Only one TOCSY spectrum was collected before the protein sample degraded. Fast degradation was possibly due to the decreased proteolytic stability of the protein at this pH due to partial unfolding.

Therefore, the ideal approach is to use a heteronuclear assignment strategy for the assignment of the unresolved region. An alternative approach is to perform the heteronuclear experiment using the natural abundance of  $^{13}\text{C}$  nuclei from a homonuclear sample. The natural abundance for  $^{13}\text{C}$  is 1.5 %. The

2D  $^{13}\text{C}$  HSQC spectrum can be acquired from a high protein concentration sample using an NMR instrument with high S/N ratio. The Varian Inova 600 MHz NMR housed at Georgia State University has a signal to noise ratio of 1200:1 (based on the standard sample), which is suitable for natural abundance experiments. The  $^{13}\text{C}$  HSQC spectrum will provide the one bond correlation between the carbon and its attached hydrogen. Although the CaM-CD2-III-5G protein has 554 carbon atoms and it cannot be fully assigned solely based on the natural abundance  $^{13}\text{C}$  HSQC, the aromatic region and the  $\text{C}\alpha\text{-H}\alpha$  region of the spectrum are still useful for assignment.

The double label of the protein with both  $^{15}\text{N}$  and  $^{13}\text{C}$  was achieved by ...The HSQC experiment was tested using a 5 mM lysozyme double-labeled sample and the experimental details are listed in Methods and Materials. A double-labeled CaM-CD2-III-5G protein with a concentration of 0.75 mM in 10 mM Tris, 10 mM KCl was used for the carbon HSQC experiment. The S/N ratio for the CaM-CD2-III-5G HSQC spectrum was lower than expected. No resonances have been observed in the  $\text{C}\alpha\text{-H}\alpha$  region for the proton dimension (3.6 to 6.0 ppm) and for the carbon dimension (44 to 65 ppm). Therefore the experiment did not provide additional information for the sequential assignment. The detailed information on trouble shooting of the instrument is shown in Methods and Materials.

#### **4.1.2 Sequential Assignment of the Heteronuclear Experiments**

#### 4.1.2.1 Triple Resonances Sequential Assignment Strategy

The advantages of the backbone triple resonance experiment are as follows:

1. Overlapping resonances are separated according to the backbone nitrogen chemical shift of each amino acid.
2. The magnetization can be transferred from  $^1\text{H}$  to  $^{13}\text{C}$  and to  $^{15}\text{N}$  and vice-versa.
3. The chemical shifts of HN proton,  $\text{H}\alpha$  proton, and  $^{13}\text{C}\alpha$  carbon sometimes overlap between different residues. Using the experiments such as HNCACB and CBCA(CO)NH, the chemical shifts of the  $\text{C}\beta$  proton can be used to distinguish different types of residues.
4. The observed HN chemical shifts are from 6.5 to 11 ppm for our protein. This range is well-removed from the chemical shift of the large water resonance. Thus slight imperfections in the water suppression will not affect the heteronuclear amide-resolved spectra as much as the homonuclear experiments are affected.

The triple resonance backbone experiments that we used are listed below:

	<b>D1</b>	<b>D2</b>	<b>D3</b>
HNCO:	HN	CO (i-1)	N
HNCA:	HN	$\text{C}\alpha$ , $\text{C}\alpha$ (i-1)	N
HNCACB:	HN	$\text{C}\alpha$ , $\text{C}\alpha$ (i-1), $\text{C}\beta$ , $\text{C}\beta$ (i-1)	N
HNCOCA:	HN	$\text{C}\alpha$ (i-1)	N
HNCACO:	HN	CO	N
CBCA(CO)NH:	HN	$\text{C}\alpha$ (i-1), $\text{C}\beta$ (i-1)	N

The HNCO experiment correlates the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for a given amino acid residue to the  $^{13}\text{CO}$  chemical shift of the preceding residue in the amino acid sequence. The HNCACO experiment correlates the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for a given amino acid residue to the  $^{13}\text{CO}$  chemical shift of the same residue in the amino sequence. The HNCACO experiment has a low sensitivity that requires a very high concentration sample and long experimental acquisition. Since the chemical shift range for  $^{13}\text{CO}$  is small (only 10 ppm), there is substantial degeneracy that limits unambiguous links between residues to be established. The  $^{13}\text{CO}$  chemical shift, however, is important for structural calculation; it can be used to predict the dihedral angles using TALOS (116). Data from an HNCO experiment was collected to evaluate this. The HNCA experiment correlates the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for a given amino acid residue to the  $^{13}\text{C}\alpha$  chemical shift of the current and preceding residues in the amino acid sequence. The HNCOCA experiment correlates the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for a given amino acid residue to the  $^{13}\text{C}\alpha$  chemical shift of the preceding residue in the amino acid sequence. HNCA and HNCOCA spectra are also able to provide sequential connectivity information (Figure 4.9). For the CD2 protein and the inserted EF-loop, there are several overlapping  $\text{C}\alpha$  chemical shifts for several sequential residues. Therefore, the  $\text{C}\beta$  chemical shift is required to unambiguously link these residues. The HNCACB experiment correlates the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for a given amino acid residue to the  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts of the current and preceding residues in the amino



acid sequence. The HNCACB is also a low sensitive experiment that requires longer experimental acquisition. A higher sensitivity experiment CBCA(CO)NH is used in conjunction with HNCACB. The CBCA(CO)NH experiment correlates the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for a given amino acid residue to the  $^{13}\text{C}_\alpha$  and  $^{13}\text{C}_\beta$  chemical shifts of the preceding residue in the amino acid sequence.

#### 4.1.2.2 Sequential Assignment for CaM-CD2-III-5G

The HNCO, HNCA, HNCACB, and CBCA(CO)NH are all  $^{15}\text{N}$  edited experiments where the carbon chemical shift is correlated to the HN proton chemical shift at D1 and the  $^{15}\text{N}$  chemical shift at D3 of each amino acid. The HN proton chemical shifts of the amino acids assigned (69%) from the homonuclear experiments were used as starting points, and the triple resonance experiments were used to verify the homonuclear assignments. The heteronuclear assignment procedure was mainly carried out based on the HNCACB and CBCA(CO)NH spectra. The Sparky assignment software was used to visualize the data and assist in the assignment procedure (Methods and Materials).

Eleven residues of the EF-loop and the first glycine residue of the C-terminal linkers were assigned using the backbone triple resonances experiments. The sequence of the insertion is shown below:

G53-G54-G55-**D56-K57-D58-G59-N60-G61-Y62-I63-S64-A65**-A66-**E67-G68**-G69-G70

The assigned region of the insertion is highlighted in bold and the assigned strip plot is shown in Figure 4.10. The  $^{15}\text{N}$  HSQC spectrum is shown in Figure 4.11. Assignments for the EF-loop were determined based on interpretations of the  $^{15}\text{N}$  and  $^{13}\text{C}$  chemical shifts.

The  $^{15}\text{N}$  chemical shift for G68 is located at 110.9 ppm. Since glycine does not have a sidechain, no  $\text{C}\beta$  resonance is observed. The  $\text{C}\alpha$  of the previous residue has a  $^{13}\text{C}$  chemical shift of 56.84 ppm and the  $\text{C}\beta$  of the previous residue has a  $^{13}\text{C}$  chemical of 30.38 ppm. The magnitude of these observed  $^{13}\text{C}$  chemical shifts are only consistent with Glu, Lys, or Arg. The CD2 host protein has 8 Gly (G8, G11, G13, G35, G53, G78, G91, and G102). Only G35 is preceded by a Glu, Lys or Arg, which is R34. The resonances for R34 and G35 were assigned. The G35 has a nitrogen chemical shift of 119 ppm. Thus, since G68 is the only remaining glycine preceded in amino sequence by Glu, Lys, or Arg, the assignment for G68 is verified.

The  $^{15}\text{N}$  chemical shift for E67 is observed at 121 ppm. The  $\text{C}\alpha$  and  $\text{C}\beta$  resonances correlating to these chemical shifts are in good agreement with the values observed on the G68 plane of the CBCA(CO)NH experiment.

The HN- $\text{C}\alpha$ -N crosspeaks for A66 (position 11 of the 12 residue EF-loop) are shown to overlap with A65. The crosspeaks of A65 are in the  $^{15}\text{N} = 127.46$  ppm with  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts of 52.75 and 19.34 ppm, respectively. On the nitrogen plane corresponding to  $^{15}\text{N} = 127.46$  ppm, the CBCA(CO)NH

and HNCACB spectra both indicate that the previous residue is S64 with  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts of 58.22 and 64.04 ppm, respectively. At the backbone  $^{15}\text{N}$  chemical shift of 121.01 ppm for E67, the CBCA(CO)NH and HNCACB spectra both indicate that the previous residue has  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts of 52.75 and 19.34 ppm, respectively. No resonances have been found at 52.75 and 19.34 ppm for any other nitrogen planes, therefore we conclude that the resonances of A66 overlap with A65.

The  $^{15}\text{N}$  chemical shift for Ser occurs at 121.9 ppm. The  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts for S64 are observed at 54.28 and 64.09 ppm, respectively. This is consistent with the observation describe above. For A65, the CBCA(CO)NH data shows  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts corresponding to S64. The  $\text{C}\alpha$  and  $\text{C}\beta$  chemical shifts of I63 are equivalent to the result observed on the S64 plane of CBCA(CO)NH.

The  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts of Y62 agree with the characteristic chemical shifts of Tyr. For Y62, the CBCA(CO)NH experiment shows a  $^{13}\text{C}\alpha$  chemical shift of 45.52 ppm, which suggests that the preceding residue is a Gly.

The  $^{15}\text{N}$  chemical shifts of G61 and G59 are both observed at approximately 109.9 ppm. G61 is preceded by N60 and G59 is preceded by D58 in the primary sequence. Asp usually shows a more downfield  $^{13}\text{C}\beta$  chemical shift than Asn. So the Gly residue with a HN chemical shift at 8.423 ppm was assigned to G61. The CBCA(CO)NH spectrum for G61 showed an upfield  $^{13}\text{C}\beta$

that corresponds to N60. The  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts for D58 are equal to the ones observed for G59 in the CBCA(CO)NH experiment.

The  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts for K57 agree well with the usual values for Lys. The CBCA(CO)NH shows that the previous residue is Asp. For D56 ( $^{15}\text{N} = 121.2$  ppm) the CBCA(CO)NH spectrum shows that there is a Gly in front of it.

There is a Gly observed at  $^{15}\text{N} = 109.9$  ppm. The HNCACB spectrum shows only one peak. The CBCA(CO)NH and HNCOCA spectra also show a peak at the same position. This glycine is located after a glycine, which is part of the glycine linker. Since the glycine linker is a fairly flexible region of the protein, it is highly possible that several glycine residues overlap at the same location.

In the ideal situation, the sequential assignment proceeds from the C-terminal to the N-terminal end of the protein. However, when a Pro residue, chemical shift degeneracies, or unknown insertion regions are present, it is impossible to complete the process. Figure 4.12 shows an example for assigning from the N-terminal to the C-terminal direction. The  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  resonances of L38 are at 56.48 and 42.07 ppm, respectively. The related HN chemical shift is at 8.78 ppm and the  $^{15}\text{N}$  chemical shift is 130.9 ppm. At this point in the assignment procedure, the  $^{15}\text{N}$  dimension of the NHCACB dataset is searched to find the plane (or planes) that show assignment of L38. The third dimension (the nitrogen) is then flipped through the  $\text{C}\alpha$  and  $\text{C}\beta$  crosspeaks that match the chemical shift of L38. The next residue in the sequence is V39.

Valine typically displays  $^{13}\text{C}\alpha$  chemical shifts from 60 to 65 ppm and  $^{13}\text{C}\beta$  chemical shifts from 30 to 35 ppm. The CBCA(CO)NH experiment is much more sensitive than the HNCACB experiment. When the crosspeaks are too weak in the HNCACB spectrum, the CBCA(CO)NH can be used. In this case, the next residue, V39, is observed at  $^{15}\text{N} = 123.2$  ppm with a HN chemical shift of 8.90 ppm. The  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts of V39 are 62.68 and 32.47 ppm, respectively. Both the HNCACB and CBCA(CO)NH spectra also show the  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  crosspeaks corresponding to the chemical shift values of L38.

The majority of the  $^{13}\text{C}\alpha$  resonances have distinct chemical shifts. But the  $\text{C}\alpha$  chemical shifts for seven pairs of sequential residues (28-29, 33-34, 72-73, 75-76, 110-111, 111-112, and 112-113) differ by less than 0.5 ppm and the corresponding signals overlap or are indistinguishable from one other in the  $^{13}\text{C}$  dimension. The sequential assignments for these resonances were mainly accomplished using the  $^{13}\text{C}\beta$  chemical shifts. Using the triple resonance assignment strategy, the previously unassigned residues S3, K43 to K51, G70, A71, A76, and K83 (resonances overlapped in homonuclear experiments) were assigned. No resonances are found for S52. No HN correlated crosspeaks have been shown for N101, but its  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts were assigned based on residue G102.

#### **4.1.2.3 Summary of the Heteronuclear Sequential Assignment**

Overall, 88  $^{13}\text{CO}$ , 102  $^{13}\text{C}\alpha$ , 88  $^{13}\text{C}\beta$ , 104  $^{15}\text{N}$ , and 107 HN chemical shifts were assigned using triple resonance heteronuclear spectra. The sequential assignment for wild type CD2 was previously completed by Chen et al. (118). The HN and  $^{13}\text{C}\alpha$  chemical shift differences between CaM-CD2-III-5G and wild type CD2 are shown in Figure 4.13 (118). Most of the HN chemical shifts do not change, whereas the residues that are close to the inserted EF-loop show minor differences. This difference is due to expected local environmental changes. These data suggest that the host protein has retained its native fold after insertion of the EF-loop.

#### 4.1.2.4 Chemical Shift Analysis

The Chemical Shift Index (CSI) analysis was performed to determine the secondary structure of CaM-CD2-III-5G (148). The random coil chemical shift of each residue type was subtracted from the observed values for each amino acid in CaM-CD2-III-5G (148). A positive CSI value for the HN chemical shift suggests that the residues have  $\beta$ -strand propensity, while a negative CSI value suggests that the residues have  $\alpha$ -helix propensity (148). The positive CSI value for  $^{13}\text{C}\alpha$  suggests  $\alpha$ -helical propensity, while the negative CSI value indicates  $\beta$ -strand propensity. The secondary structure prediction is based on CSI value of three or more consecutive residues. For example, three consecutive residues with positive CSI values of HN are indications of  $\beta$ -strand secondary structure.

A CSI value close to zero for both HN and  $^{13}\text{C}\alpha$  is an indication of a flexible, less defined structure.

The HN CSI analyses for CaM-CD2-III-5G are shown in Figure 4.14. The HN CSI index analysis indicates that there are several  $\beta$ -strand secondary structure elements in the CaM-CD2-III-5G, and the host protein maintains its  $\beta$ -sheet secondary structure. The CSI analysis for the EF-loop III suggests that most of the EF-loop possesses a flexible loop conformation whereas the residues from positions 7-9 of the EF-loop show  $\beta$ -sheet conformation.

### 4.1.3 Backbone and Sidechain NOE Assignment for Structural Calculation

#### 4.1.3.1 Assigning NOE for the Unresolved Region using 3D $^{15}\text{N}$ -NOESYHSQC

There are no mainchain sequential NOE assignments for residues from R44 to G68 because the fingerprint region of the 2D NOESY spectrum was heavily overlapped. The sequential assignment for this stretch of residues in CaM-CD2-III-5G was achieved by using the 3D  $^{15}\text{N}$  NOESYHSQC and 3D  $^{15}\text{N}$  TOCSYHSQC. The information obtained from the 3D  $^{15}\text{N}$  NOESYHSQC and 3D  $^{15}\text{N}$  TOCSYHSQC was further verified with the triple resonance experiments discussed in the previous section.

**Table 4.1 Summary of the sequential NOE Assignment for the EF-loop III of CaM-CD2-III-5G.**

Residue	$\text{HN}_i - \text{HA}_i$	$\text{HN}_i - \text{H}^*_i$	$\text{HN}_i - \text{HA}_{i-1}$	$\text{HN}_i - \text{H}^*_{i-1}$	Other
D56 HN	Overlap	Overlap	Overlap	Overlap	
K57 HN	HA	HB*, HG*	HA	HB*	
D58 HN	HA	HB*	HA	HB*, HG*	

G59 HN	HA2	N/A	HA	HB*	G61HN
N60 HN	HA	HB*	HA2	N/A	
G61 HN	HA*	N/A	HA	HB*	Y62HN
Y62 HN	HA	HB*, HD*	HA*	N/A	
I63 HN	HA	HB, HG2*	HA	HB*	
S64 HN	HA	Not assign	HA	HB, HG2*	
A65 HN	HA	HB*	HA	Not assign	
A66 HN	Not assign	Not assign	Not assign	Not assign	
E67 HN	HA	HB3, HG*	Not assign	HB*	

\*protons with equivalent chemical shifts.

The summary of the sequential NOE assignment for the inserted EF-loop of CaM-CD2-III-5G is listed in Table 4.1. The triple resonance experiments have shown that D56, S64, and E67 have HN chemical shifts of 8.31, 8.32, and 8.33 ppm, respectively. The  $^{15}\text{N}$  chemical shifts for D56, S64, and E67 are 121.2, 121.1, and 121.0 ppm, respectively. The  $^{13}\text{C}\alpha$  and  $^{13}\text{C}\beta$  chemical shifts of these three residues are significantly different. The  $^{13}\text{C}\alpha$  chemical shifts for D56, S64, and E67 are 54.22, 58.31, and 56.84 ppm, respectively. The  $^{13}\text{C}\beta$  chemical shifts for D56, S64, and E67 are 41.35, 64.10, and 30.38 ppm, respectively. The backbone  $^{15}\text{N}$  and HN chemical shifts are very similar for these three residues. In the triple resonance experiments, the signals for D56, S64, and E67 are clearly identified because of the large difference in the  $^{13}\text{C}$  chemical shifts. In the  $^{15}\text{N}$  edited NOESYHSQC and TOCSYHSQC experiments, however, their HA resonances are very similar. Three crosspeaks arise from interactions between A65 and S64 proton nuclei. The  $\text{H}\beta$  chemical shift of A65 is observed at 1.423 ppm. The resonances for A66 are not observed. There is one additional crosspeak observed in the glycine region of the  $^{15}\text{N}$  HSQC spectrum. The triple resonance experiments suggest that this is a glycine residue and the residue in



front of it is also a glycine. There were three possible assignments for this: G55, G69 or G70. Since the HA chemical shift did not match the HA shift of G68 or the NOE crosspeak from A71, this crosspeak was assigned as G55.

Four additional residues have been assigned using the three dimensional experiments. The  $^{15}\text{N}$  chemical shift of K45 is 120.1 ppm, and its HN chemical shift is 8.579 ppm. The  $^{15}\text{N}$  chemical shift M46 is 120.6 ppm, and its HN chemical shift is 7.953 ppm. The  $^{15}\text{N}$  chemical shift K47 is 124.0 ppm, and its HN chemical shift is 8.133 ppm. The  $^{15}\text{N}$  chemical shift K51 is 123.6 ppm, and its HN chemical shift is 8.284 ppm.

Using the three dimensional  $^{15}\text{N}$ -NOESYHSQC experiment, 22 crosspeaks were assigned from R44 to K51 (except for P48 since proline does not have backbone HN). 33 crosspeaks have been assigned for the inserted sequences.

#### **4.1.3.2 Mainchain and Sidechain NOE Assignment on the 2D NOESY Spectrum for Structural Calculation**

The NMR solution structures are generally constructed using the distance information from the NOE distance restraints and the backbone dihedral angles from either coupling constant experiments or prediction based on the chemical shifts. CaM-CD2-III-5G contains 9 anti-parallel  $\beta$ -strands that fold up into two  $\beta$ -

sheet layers (Figure 4.15). One layer is formed by  $\beta$ -strands GFCC'C" while the other is formed by  $\beta$ -strands ABED. The  $\beta$ -strands are connected by several loops. The sequential NOE (residue<sub>*i*</sub> to residue<sub>*i-1*</sub>) and intra- $\beta$ -strand NOE (residue<sub>*i*</sub> to residue<sub>*i-2*</sub>) define the twist and the bend of the  $\beta$ -strands. The inter-strand NOE defines how the anti-parallel- $\beta$ -strands line up together. The inter- $\beta$ -sheet NOE (also inter- $\beta$ -strand) defines how the  $\beta$ -sheet is formed.

#### **4.1.3.3 Backbone to Backbone NOE Assignment**

The backbone NOE interactions in proteins are the HN – HN\*, HN - H $\alpha$ \*, and H $\alpha$  - H $\alpha$ \* (the \* refer to any residues that are within 5 Å in distance). The sequential HN<sub>*i*</sub> – HN<sub>*i-1*</sub> NOE in a  $\beta$ -strand typically has very weak peak intensity because the two protons point in different directions. The distance is typically longer than 4.0 Å. The sequential HN<sub>*i*</sub> - H $\alpha$ <sub>*i-1*</sub> peak in a  $\beta$ -strand typically has very strong intensity because the two protons point in the same direction. The peak intensities for HN to HN\* or H $\alpha$ \* interactions depend on the distance between the two nuclei and indicate which  $\beta$ -strands are lined up together. The HN-HN\* NOE assignment is shown in Figure 4.16.

#### **4.1.3.4 Sidechain to Sidechain and Sidechain to Mainchain NOE Assignment**

The other types of NOE interactions in proteins are the sidechain to sidechain and sidechain to mainchain interactions. In CaM-CD2-III-5G, there are intra-strand, inter-strand, and inter- $\beta$ -sheet sidechain interactions. For example, the aromatic ring of W32 has more than 40 NOE interactions with the sidechains from different  $\beta$ -strands or other  $\beta$ -sheet layers. The sidechain of W32 also has NOE interactions with the mainchain HN proton of I18, which is on the other face of the protein. The calculated structure of CaM-CD2-III-5G is shown in Figure 4.17. The W32 residue is shown in red and the sidechains with which W32 interacts are shown in orange.

#### **4.1.3.5 Summary of the Assignment for CaM-CD2-III-5G from both the Homonuclear and Heteronuclear Experiments**

The assigned  $^{15}\text{N}$  HSQC spectrum for CaM-CD2-III-5G is shown in Figure 4.11. There are a total of 117 crosspeaks in the  $^{15}\text{N}$  HSQC spectrum: 96 from the mainchain, with 18 crosspeaks from the sidechains of W7, W32, N15, N17, N20, N60, N77, N84, N101, and N107. Based on the 2D and 3D heteronuclear experiments, 854 chemical shifts have been assigned. A total of 546 crosspeaks were assigned in the  $^{15}\text{N}$  NOESYHSQC spectrum including 209 intra, 190 sequential, and 147 medium to long range NOE interactions. A total of 1361 crosspeaks were assigned in the 2D NOESY spectrum with a mixing time of 150 ms including 589 intra, 428 sequential, and 344 medium to long range NOE interactions. There are a total of 926 crosspeaks assigned in the 2D NOESY

spectrum with a mixing time of 100 ms including 461 intra, 212 sequential, and 253 medium to long NOE interactions. Not all of the NOE crosspeaks in the NOESY spectrum with 100 ms mixing time have been assigned.

Using the homonuclear and heteronuclear experiments, assignments for 109 amino acids were achieved. The sequential assignment for CaM-CD2-III-5G is 94% complete (Appendix 3.1). Like many other proteins, the N-terminal residues R1 and D2 are very flexible and no corresponding signals were observed. S52 was not observed in any of the NMR experiments at pH 6.8 or pH 7.4 in 25 °C. G54 and G69 were either not observed in the  $^{15}\text{N}$  HSQC spectrum or overlapped with other glycine residues. L89 was only observed in the homonuclear TOCSY and NOESY experiments. Only the sidechains of P19 and P48 were observed since they do not have HN protons. For N101, only the sidechain resonances were observed in both homonuclear and heteronuclear experiments.

## **4.2 Structural Calculation**

To determine if the inserted EF-loop III in CaM-CD2-III-5G still maintains similar structural and metal binding properties as in calmodulin, the NMR solution structure for CaM-CD2-III-5G was calculated. NOE distance restraints, dihedral angle restraints, and hydrogen bond restraints are traditional methods to

determine the NMR solution structure. More recently, the residual dipolar coupling has been used to refine the NMR structure. The structural calculation process usually starts by assigning as many NOE crosspeaks as possible. The NOEs between the sequential residues define the primary and secondary structures of proteins. The mid-range and long-range NOEs can define the folding and tertiary structure of the protein. The peak height or the peak volume of the NOE crosspeak is then converted into the distance restraint. The chemical shift of the protein can also be used during the calculation to assist the prediction of the secondary structure of the protein. The dihedral angle restraints for the backbone and the sidechains of the protein can be obtained from various types of NMR experiments, such as the HNHA experiment, which gives the  $\phi$  backbone dihedral angles. The dihedral angles can also be calculated using the TALOS program based on the chemical shifts of the backbone atoms. The hydrogen bond restraints can also be used to define the folding of the protein.

The solution structure for CaM-CD2-III-5G was calculated using both the CYANA and CNS programs. The CYANA program has a very simple and detailed output summary for NOE restraint violations and was chosen as the first calculation step in order to reduce the time for NOE restraint violation correction. The structural calculation for CaM-CD2-III-5G by CYANA was conducted using two different methods. First was the automatic distance calibration mode, where the CYANA program reads in the peak height or peak volume of the NOE crosspeak list and performs the distance calibration. Second was the manual

mode in which the NOE restraint and dihedral angle restraints were manually prepared. The structure of both methods was derived by simulated annealing combined with torsion angle dynamics (122).

#### **4.2.1 CYANA Calculation using Automatic Distance Calibration**

The structural calculation for CaM-CD2-III-5G was carried out using CYANA 1.1 with three types of information: the chemical shifts, NOE peak intensities, and coupling constants. The NOE distance restraint list was generated using CYANA's auto calibration script "CALIBA.cya". The program automatically calibrates and converts the peak intensity to NOE distance for every assignment. But the NOE distance restraints for CaM-CD2-III-5G were not calibrated correctly. The calculated CaM-CD2-III-5G structure has a high target function energy and large number of NOE violations. The 20 lowest energy structures of the calculated CaM-CD2-III-5G are shown in Figure 4.18. The secondary structure of the host protein does not resemble the wild type CD2. The accuracy of the NOE crosspeak intensity is very important for the automatic calibration method; however, the intensity of the NOE crosspeak does not always correspond to the true distance. The exchange rate of the protein, secondary structure, crosspeak overlap, and systematic errors can all contribute to the inaccuracy of the peak height. Therefore, the NOE crosspeaks for CaM-CD2-III-5G were manually calibrated.

#### 4.2.2 CYANA Calculation using Manual Mode

The NOE distance restraint list was generated based on the NOE crosspeaks from 2D and 3D NOESY experiments. The lower limit and upper limit restraints are both stated in the list. The  $\phi$  backbone dihedral angle list was generated based on the HNHA experiment and the prediction from TALOS. The  $\psi$  backbone dihedral angle list was generated based on predictions obtained from TALOS. Since CYANA version 2.1 was released while the study was ongoing, the earliest calculations were obtained using version 1.1 and the finished structure was calculated using version 2.1. The improvement and the changes in the atom notation are listed in the Methods and Material section. Since the dihedral angle restraint list was provided to CYANA, the GRIDSEARCH (to predict backbone and mainchain dihedral angles) function in CYANA was not used. The chemical shift list was not used during the calculation.

##### 4.2.3.1 NOE Calibration for 3D $^{15}\text{N}$ NOESYHSQC

A total of 542 inter- and intra-crosspeaks have been assigned in this spectrum. The spectrum was assigned using the SPARKY 3.110 assignment software (113). The S/N ratio is easy to read and compare, so the calibration of the NOE distance was based on the values of the S/N ratios. The range of the

NOE distance is between 1.80 Å and up to 5.00 Å. The calibration scale for this spectrum is shown below in Table 4.2.

**Table 4.2 NOE Distance Calibration Scale for 3D  $^{15}\text{N}$ -NOESYHSQC**

S/N Ratio	Assigned Distance	Lower Limit	Upper Limit
$x > 200$	2.00	0.20	0.40
$200 > x > 150$	2.20	0.40	0.40
$150 > x > 100$	2.40	0.40	0.40
$100 > x > 50$	2.80	0.40	0.40
$x < 50$	3.20	0.60	0.80

$x$  = peak intensity

The calibration scale was created based on the analysis of NOE cross peaks from different regions of the protein.

*Example 1,  $\beta$ -strand NOE:*

The HN protons of residues  $i$  and  $i+1$  are in opposite directions in the  $\beta$ -strand type secondary structure, while the HN proton ( $i$ ) is in the same direction as the HA proton ( $i-1$ ). Therefore, the NOE crosspeak for  $\text{HN}_i - \text{H}\alpha_{i-1}$  has a stronger intensity because the  $\text{HN}_i$  and  $\text{H}\alpha_{i-1}$  are typically close in the  $\beta$ -strand arrangement and the distance is usually around 2.20 Å (151). For the G8HN-W7HA-G8N crosspeak, the S/N ratio is 170. The S/N ratios for W7HN-V6HA-W7N and V6HN-T5HA-V6N are 253 and 246, respectively. Based on the secondary structure prediction and the wild type CD2, this segment of the protein contains the anti-parallel  $\beta$ -strand. The expected distances for these restraints are around 2.20 Å. According to the calibration scale, the distances corresponding to the three cross peaks are within the upper and lower limit of  $\beta$ -strand secondary structure.



*Example 2, 3<sub>10</sub>-helix:*

For the 3<sub>10</sub>-helix, the distance for HN<sub>i</sub> to H $\alpha_{i-1}$  is approximately 3.40 Å (151). Residues R87, D88, and D89 (R70, D71, and D72 for wild type CD2) for CaM-CD2-III-5G exhibit 3<sub>10</sub>-helix conformation. The S/N ratios for S90HN-D89HA-S90N, D89HN-D88HA-D89N, and D88HN-R87HA-D88N are 33, 38, and 33, respectively. According to the calibration scale, all four distance restraints are within the average values for 3<sub>10</sub>-helix of 3.40 Å.

#### 4.2.3.2 NOE Calibration for 2D NOESY

A total of 1150 crosspeaks have been assigned for this spectrum. The range of the NOE distance is between 1.80 Å up to 4.00 Å. The calibration scale for this spectrum is shown in Table 4.3.

**Table 4.3 NOE Distance Calibration Scale for 2D <sup>1</sup>H NOESY**

S/N Ratio	Assigned Distance	Lower Limit	Upper Limit
x > 150	2.20	0.40	0.40
150 > x > 80	2.40	0.40	0.40
80 > x > 60	2.60	0.40	0.40
60 > x > 40	3.00	0.40	0.40
40 > x > 20	3.20	0.40	0.40
x < 20	3.40	0.40	0.40

x = peak intensity

The calibration scale was created based on the analysis of NOE crosspeaks from different regions of the protein.

*Example 1, H $\delta$  - H $\epsilon$  Distance of Tyr:*

The aromatic ring of the Tyr is very rigid compared to the other parts of the proteins, so the distance between  $H\delta$  and  $H\epsilon$  can be used to calibrate the NOE distance. The distance between  $H\delta$  and  $H\epsilon$  is approximately 2.50 Å. For the Y93 $H\delta$ -Y93 $H\epsilon$  crosspeak, the S/N ratio is 91. The S/N ratio for Y98 $H\delta$ -Y98 $H\epsilon$  is 134. According to the calibration scale, the distances of both crosspeak assignments are within the upper and lower limit of the expected distances for these restraints around 2.50 Å.

*Example 2,  $3_{10}$ -helix:*

For the  $3_{10}$ -helix, the distance for  $HN_i$  to  $HA_{i-1}$  is approximately 3.40 Å (151). Residues R87, D88, and D89 (R70, D71, and D72 for wild type CD2) for CaM-CD2-III-5G exhibit  $3_{10}$ -helix conformation. The S/N ratios for S90HN-D89HA, D89HN-D88HA, and D88HN-R87HA are 16, 11, and 11, respectively. According to the calibration scale, all four distances restraints are within the average values for  $3_{10}$ -helix of 3.40 Å.

#### 4.2.4.1 Obtaining the Dihedral Angles using HNHA

The HNHA experiment measures the homonuclear HN- $H\alpha$   $J$ -coupling and each  $J$ -correlation plane is separated by the chemical shift of the backbone nitrogen. The  $J$ -coupling between HN and  $H\alpha$  can be used to calculate the  $\phi$

angle of the peptide bond, which is an important parameter for structural calculation. The HNHA observes a set of peaks for every residue, a crosspeak (HN-H $\alpha$ ) and a diagonal peak (HN-HN). The crosspeak of the spectrum is positive and the diagonal peak is negative. The  $J$ -coupling can be calculated using equation 4.1 (152):

$$S_{\text{cross}}/S_{\text{diag}} = -\tan^2(2\pi J_{\text{HH}}\zeta) \quad 4.1$$

$$J(\phi) = A \cos^2(\phi-60) + B \cos(\phi-60) + C \quad 4.2$$

The peak intensities of both the crosspeak and the diagonal peak are integrated using the Gaussian fit method (113). The ratio of the two peaks is then used in equation 4.1 to obtain the  $J$  value. Once the  $J$  value is obtained, the  $\phi$  angle for the backbone of the protein can be calculated using the Karplus equation (equation 4.2).

#### 4.2.4.2 Assignment of the 3D HNHA Spectrum

The HNHA experiment observed HN protons on the D1 dimension, HN and H $\alpha$  proton on the D2 dimension. Each resonance was located at the backbone nitrogen plane of each amino acid. The resonance identified in the HNHA spectrum was the same as the 3D  $^{15}\text{N}$  TOCSYHSQC. Residues R1, D2, S3, G4, P19, D25, S36, R44, P48, K51, S52, D56, N60, S64, A66, A76, V97, N101, R104, I105, and L106 (not including Gly) were not observed in the HNHA experiments. The resonances for 95 residues are assigned, with 17 unidentified due to overlapping HN resonances. Even though the H $\alpha$  resonance was clearly

defined, the coupling constant is dependent on the ratio between the  $H\alpha$  resonances and the HN resonances. Therefore, the residues with overlapping HN-HN resonances were not used for the dihedral angle calculation. CaM-CD2-III-5G has a total of 15 Gly and the observed  $J$ -coupling values for these residues were not used for  $\phi$  dihedral angle calculation. A total of 63 residues from the HNHA experiment were used for calculations. The assignment and the coupling constant are shown in Table 4.4. The HNHA results were combined with the prediction from TALOS (described below) for structural calculation.

#### **4.2.4.3 Calculating the Dihedral Angles using TALOS**

The full name for the TALOS dihedral angle prediction program is Torsion Angle Likelihood Obtained from Shift and Sequence Similarity (116). TALOS predicts  $\phi$  and  $\psi$  backbone torsion angles of the protein by comparing the backbone chemical shifts of the protein against the TALOS database. The TALOS database contains the chemical shift and coordinates of twenty well known proteins (116). These twenty proteins have distinct secondary structures, which are used to compare to the target protein to predict backbone torsion angles. The chemical shift of the HN proton is sensitive to the secondary structure and hydrogen bonding (116). The  $C\alpha$  and  $C\beta$  chemical shift also have direct correlation with both  $\phi$  and  $\psi$  angles (116). For the torsion angle prediction, five types of chemical shifts are used ( $H\alpha$ ,  $C\alpha$ ,  $C\beta$ , CO, and N). TALOS uses the chemical shifts of three consecutive residues to predict  $\phi$  and  $\psi$

torsion angle for the central residue in a triplet. The program compares the chemical shifts of the triplet with the chemical shifts of the known secondary structure protein in the database. The comparison gives 10 of the closest chemical shift matches, and then predicts the  $\phi$  angle and the  $\psi$  angle from the database.

#### **4.2.4.4 Predicting Dihedral Angles for CaM-CD2-III-5G**

The chemical shifts used for the TALOS calculation were obtained from the following spectra: TOCSY with different mixing times, NOESY with different mixing times,  $^{15}\text{N}$  TOCSYHSQC,  $^{15}\text{N}$  NOESYHSQC, HNCO, HNCA, HNCACB, and CBCA(CO)NH. CaM-CD2-III-5G has 116 amino acids. TALOS has predicted backbone torsion angles for 111 residues. The results are shown in Table 4.5 along with the dihedral angles of wild type CD2 for comparison. The torsion angles for R1 at the N-terminal and E116 at the C-terminal were not predicted. There are no assignments for S52, G53, and G69, so no predictions for those residues were made.

TALOS predicts the dihedral angle by comparing the  $\text{H}\alpha$ ,  $\text{C}\alpha$ ,  $\text{C}\beta$ , CO, and N chemical shifts of each amino acid to the database. The predicted dihedral angle is the average of the 10 best matches, which TALOS found in the database. If  $\phi$  and  $\psi$  dihedral angles of the 10 best matches are all in the same favorable region of the ramachandran map, TALOS will classify the prediction as good. However, if one of the matches falls outside the favorable region of the

Ramachandran map, TALOS will classify the prediction as new. For CaM-CD2-III-5G, only 58 of the predictions are classified as good and the rest were classified as new. To ensure an accurate prediction, the assignment of the CaM-CD2-III-5G was carefully verified. After extensive review with the sequential assignment, no errors in assignment were found. The chemical shifts of the wild type CD2 were used to predict torsion angles and TALOS classified 59 of the predictions as good.

There are two possible explanations for the results of TALOS on CaM-CD2-III-5G and CD2. First, the dihedral angle prediction is incorrect, because the TALOS database does not have enough information to predict the dihedral angle correctly. Second, the dihedral angle prediction is correct. The TALOS rating function, however, does not have enough knowledge to evaluate the quality of the prediction. To evaluate these two possibilities, the PROCHECK software was used to extract the backbone dihedral angles for the crystalline structure of wild type CD2 (1HNG.pdb) (100). The dihedral angles of wild type CD2 from PROCHECK are summarized in Table 4.5.

TALOS considers the prediction for 58 residues as good, and the majority of these angles are within  $\pm 15^\circ$  of the corresponding angle in the crystal structure of wild type CD2(1HNG). The dihedral angles for 18 residues that are considered as new predictions by TALOS are also within  $\pm 15^\circ$  of the corresponding angles of the crystal structure. The dihedral angles for 8 residues are within  $\pm 20^\circ$  and 2 residues are within  $\pm 30^\circ$ .

The comparison of dihedral angles of CaM-CD2-III-5G to 1HNG.pdb of wild type CD2 has suggested that the majority of the TALOS prediction closely resemble the dihedral angles of wild type CD2. In conclusion to the data analysis the of TALOS prediction, the TALOS program has the ability to predict useful backbone dihedral angles for structure calculation; yet, the evaluation function of the TALOS does not give conclusive analysis for the output dihedral angles.

#### **4.2.4.5 Summary of the Dihedral Angle vs. HNHA**

The HNHA experiment has yielded  $\phi$  dihedral angles for 63 residues of CaM-CD2-III-5G. The comparison between the experimental HNHA  $\phi$  dihedral angles and the TALOS prediction is shown in Figure 4.19. The angles for 25 residues are within  $\pm 20^\circ$  from the two methods. The angles for 19 residues are within  $\pm 30^\circ$  and 8 residues are within  $\pm 40^\circ$ . Eleven of the HNHA experimental  $\phi$  dihedral angles have deviations  $\geq 40^\circ$  in comparison with the corresponding residues in the TALOS prediction and the results of these eleven residues were not used. Using the results from HNHA and TALOS, the dihedral angle restraint list for CYANA has been constructed for 86 residues from the host protein and 12 residues from the inserted EF-loop. The calculated dihedral angles of the glycine linker were not used. The dihedral angle table for CYANA can be found in Appendix 2.2.

#### **4.2.4.6 Comparison to Calmodulin**

The TALOS program predicted  $\phi$  and  $\psi$  angles for all 12 residues of the inserted loop. The  $\phi$  and  $\psi$  backbone dihedral angles for calcium free calmodulin were obtained from 1CFD.pdb using PROCHECK (Figure 4.20) (62). Most of the predicted  $\phi$  and  $\psi$  dihedral of CaM-CD2-III-5G are within  $35^\circ$  of the corresponding ones in the calcium free calmodulin, especially loop residues at positions 7 and 8. The  $\psi$  dihedral angle at the N-terminal and both  $\phi$  and  $\psi$  dihedral angles at the C-terminal of the inserted EF-loop III in CD2 have large deviations from calmodulin. This is likely due to the fact that the terminal ends of the inserted EF-loop III are attached to the CD2 host protein by two glycine linkers, which exhibit more flexible conformations. The  $\phi$  and  $\psi$  dihedral angles only at position 4 of the inserted EF-loop III has a large deviation from calmodulin, which may be due to a relatively larger allowed conformation with the sidechain constraints. The larger deviation observed in the  $\psi$  dihedral angle at position 9 and  $\phi$  dihedral at position 10 might be caused by the following reasons. In the native calmodulin structure, the EF-loop positions 7 to 9 usually form a small  $\beta$ -strand. The exiting helix of the EF-hand motif usually starts from position 10 of the EF-loop. In our design, there are no entering and exiting helices, so the conformations on both end of the EF-loop are not restricted. In addition, a grafted single EF-loop insertion results in some destabilization due to the absence of hydrogen bonds between the two  $\beta$ -strands as in the native calmodulin. Positions 9 and 10 therefore form a different geometry. Positions 10



to 12 are at the beginning of the exiting helix. However, there is no exiting helix in CaM-CD2-III-5G, so the conformations of the residues are less defined.

#### 4.2.5.1 Structural Calculation with NOE Distance Restraints

The structural calculation for CaM-CD2-III-5G was started with 1091 NOE distance restraints from the 2D and 3D NOESY experiments. The NOE distance restraints were modified or removed based on the NOE violation from CYANA. The final structure of CaM-CD2-III-5G took over 10 cycles to complete. The summary of each calculation cycle is shown in the following section.

The first 6 cycles of calculation were carried out without the dihedral angle restraint. The summary of the first 6 cycle calculations is shown in Table 4.6.

Table 4.6 Summary of the First 6 Cycle Calculations

Structure	Average Target Function	# of NOE Violations
1	80.59	285
2	66.94	205
3	65.25	199
4	42.64	126
5	14.21	69
6	8.41	56

The structure calculations were performed using CYANA 1.1. The first cycle had the worst target function energy and the highest NOE distance violations with the most errors. The cycle 1 structure with the lowest energy is shown in Figure 4.18. The structure does not have the distinct  $\beta$ -sheet structure, yet the protein appears to be folded. The NOE correction is based on the NOE distance violation from the overview file of each calculation cycle. The overview

file was generated based on 20 calculated structures with the lowest energies. From cycle 1 to cycle 6, only restraints with violation in more than 10 structures were corrected. In cycles 1 and 2, 1091 NOE distance restraints were used for structure calculation. The calculation yields 285 and 205 NOE violations for cycles 1 and 2, respectively. Only the lower limit NOE violations were corrected in this step. The incorrect lower limit NOE distance restraints may cause the surrounding residues to have local NOE violation, which may not be true.

In cycle 3, 1091 NOE distance restraints were used. The calculation yields 199 NOE violations. From cycle 3, the NOE violation correction includes both the upper and lower limit violations. Any NOE distance restraint that is either mis-assigned or possibly incorrect was removed. In cycle 6, 1056 of the original 1091 NOE distance restraints were used for the calculations. The calculation yields 56 NOE violations with the average target function energy down to 8.41. The structure with the lowest energy is shown in Figure 4.21. There are four visible  $\beta$ -strands in the structure calculated from cycle 6 with two  $\beta$ -strands on GFCC'C" side of the protein and two  $\beta$ -strands on the other face of the protein.

#### **4.2.5.2 Adding the Dihedral Angle Restraint Table**

The calculation in cycle 1 to cycle 6 was focused on sorting out the major incorrect NOE distance restraints. In cycle 7 to cycle 11, the structures were calculated with the dihedral angle restraint using CYANA 1.1. In cycle 11 to

cycle 15, the structures were calculated with dihedral angle restraints using CYANA 2.1. The summary of cycle 7 to 15 is shown in Table 4.7.

Table 4.7 Summary of the Cycle 7 to 15 Calculations

Cycle	CYANA	Average Target Function	# of NOE Violations
7	1.1	19.17	79
8	1.1	15.02	57
9	1.1	12.12	38
10	1.1	11.08	43
11	1.1	9.87	30
11	2.1	23.40	87
12	2.1	19.32	66
13	2.1	11.73	40
14	2.1	10.51	27
15	2.1	11.16	25

In cycle 7, the average target function increased to 19.17 with NOE violations up to 79. In cycle 6, the average target function was 8.41 with 56 NOE violations. The increases in the target function and NOE violations in cycle 7 are likely due to the dihedral angle restraints restricting the structure to a certain conformation. In cycle 6, without the dihedral angle restraint, the CYANA 1.1 program can accommodate the incorrect NOE distance restraints by orientating the structure to the lowest energy conformation without the backbone conformation restraint. By continuing the calculation along with NOE distance restraint corrections, the average target function is reduced to 9.87 with 30 NOE violations at cycle 11 using CYANA 1.1

The CYANA 2.1 program was released in 2005. There are several new features for the program. The program uses the standard IUPAC format for the nomenclature, more refined automated NOE assignment, and larger van der

Waals radii (for more detail on CYANA 2.1, please visit the author's website at <http://quentert.gsc.riken.go.jp/Software/Cyana/NewFeatures.html>). The larger van der Waal radii for the repulsive term are used to obtain a better energetically-favorable structure. With this improvement in mind, the structure calculation for CaM-CD2-III-5G was migrated to CYANA 2.1. For CYANA 2.1, the naming of the HN proton is changed to H. For the sequence file (5g3-52.seq) and the dihedral angle restraint file (05nov-talos.aco), the plus and minus signs on the charged amino acid, ARG+, LYS+, ASP-, and GLU-, were removed.

The parameters used for cycle 11 calculation on CYANA 1.1 were used to calculate a structure for CaM-CD2-III-5G using CYANA 2.1. The target function energy increased to 23.40 from 9.87. The NOE violations increased from 30 up to 87. The increased target function energy was likely due to the improved van der Waal library and the larger van der Waal radius function setting in the new program. As the NOE correction process continued to cycle 24, both the target function energy and NOE violations decreased.

#### **4.2.5.3 Summary of NOE used for Calculation**

The solution structure of apo CaM-CD2-III-5G was calculated using 1059 NOE restraints and 189 dihedral angle restraints (a combination of  $\phi$  and  $\psi$  dihedral angles). The obtained structure is shown in Figure 4.22. The NOE distances table, dihedral angle restraints table, and all of the scripts used for

CYANA are listed in appendix 2.2. As we mentioned in the previous section about the NOE assignment, the inserted EF-loop is more than 10 Å away from the host protein. There are no NOEs between the inserted EF-loop III and the host protein. The glycine linkers were designed to allow the EF-loop to have enough flexibility to accommodate the metal ion. But this increases the distance between the residues at the terminal ends (D56, K57, A66, and E67) and the CD2 portion of CaM-CD2-III-5G. No NOE connectivities have been observed between the inserted EF-loop III and the CD2 portion of CaM-CD2-III-5G. The overlay of the twenty lowest energy structures based on the host protein makes the conformation of the inserted EF-loop III to look like mushroom shape (same applied when overlaying based on the EF-loop). So the overlaid structures are shown in both ways.

The NOE distance restraints for the inserted EF-loop III were obtained mainly from the 3D  $^{15}\text{N}$  NOESYHSQC due to the signal overlapping in the 2D NOESY spectrum. A total of 59 distance restraints have been assigned related to the inserted EF-loop. There are 31 intra-residues, 23 sequential, 4  $i$  to  $i+2$ , and 1 restraint between the N-terminus and C-terminus end of the EF-loop (K57/E67, position 2 to position 12 of the loop) NOE restraints assigned.

For the host protein portions of CaM-CD2-III-5G, 992 NOE distances have been assigned. There are 466 intra-residues, 255 sequential, 38  $i$  to  $i+2$ , and 21 within  $i$  to  $i+5$  NOEs (Table 4.8). There are 215 NOE assignments for residues  $i$  to  $> i+5$  and most of these NOE are from inter- $\beta$ -strands connectivity.

**Table 4.8: Summary of the NOE Restraints**

	Loop+Host	Loop	Host
Total	1051	59	992
Intraresidue ( <i>i</i> to <i>i</i> )	497	31	466
Sequential ( <i>i</i> to <i>i</i> +1)	278	23	255
<i>i</i> to <i>i</i> +2	39	1	38
<i>i</i> to <i>i</i> +3 and <i>i</i> to <i>i</i> +4	21		21
<i>i</i> to $\geq i$ +5	216	1	215

#### 4.2.5.4 The Structure of CD2 Host Protein of CaM-CD2-III-5G

The secondary and tertiary structures of the host protein are similar to the wild type CD2. It does have 9 anti-parallel  $\beta$ -strands and folds into an IgG fold similar to domain 1 of wild type CD2. The GFCC'C"  $\beta$ -strands constitute one face of the  $\beta$ -sheet while the  $\beta$ -strands ABDE make up the  $\beta$ -sheet on the opposite face. In the wild type CD2 structure, the short segment of the  $3_{10}$ -helix is made up by residues R70, D71, and D72 next to  $\beta$ -strand G. In CaM-CD2-III-5G, the  $3_{10}$ -helix is also formed next to  $\beta$ -strand G by the same residues as the wild type CD2. The RMSD for the host protein can only be compared as two sections due to the flexible glycine linkers. The RMSD for the 20 lowest structures of the first section of the host protein (G4-L50) is 0.378 Å. The RMSD for the 20 lowest structures of the second half of the host protein (F55-R96) is 0.225 Å.

#### 4.2.5.5 The Structure of the Inserted EF-loop III of CaM-CD2-III-5G

The overlay for the 20 lowest structures of the inserted EF-loop III is shown in Figure 4.22. The RMSD of the entire EF-loop is 1.761 Å. The residues

at both terminal ends of the loop are more flexible. This is likely due to the glycine linker not providing a rigid platform. So, when the EF-loop binds with metal it is able to re-orientate itself to accommodate the metal ion. The RMSD for the 8 center residues (DGNGYISA) decreases from 1.761 to 0.993 Å.

#### **4.2.5.6 Comparing the EF-loop III of CaM-CD2-III-5G to Calmodulin**

The NMR solution structure of calcium free calmodulin was previously determined by Bax, Ikura, and Forsen (C-terminal domain) (60, 62, 63). The previous work has suggested that in the calcium free form, calmodulin still maintains similar secondary structure as the calcium loaded form, while all of the EF-hand motifs still maintain the helix-loop-helix conformation. To determine if the calcium free EF-loop III in CaM-CD2-III-5G retains a conformation similar to the calcium free EF-loop III in calmodulin, the following areas were compared:

First, residues at positions 7 to 9 exhibit strong  $\beta$ -strand conformation preferences, similar to apo-calmodulin. On the other hand NOE interactions between the  $\beta$ -strands of the two paired EF-hand motifs are more observed in the intact protein.

Second, the  $^{15}\text{N}$  relaxation studies and hydrogen exchange studies by Bax and Ikura suggested that the first six residues of the EF-loop of calcium free calmodulin are dynamic disordered and only a very few NOE restraints were observed (60, 62).

Third, the last residue of the entering helix and the first residue of the EF-loop form a  $3_{10}$  helix conformation, and this conformation was only observed for the first EF-hand motif in both the N-terminal and the C-terminal domains (Site I and Site III). Further, the end of the EF-loop in calcium free calmodulin (residues 10 to 12) forms the beginning section of the exiting helix and also provides N-capping for the exiting helix.

#### **4.2.5.6.1 Strong $\beta$ -Strand Conformation Preference at Position 7 to 9 of the EF-loop**

All four EF-hand motifs in calcium free calmodulin have the helix-loop-helix secondary structure with a small  $\beta$ -strand formed by residues 7 through 9. The  $\beta$ -strand formation can be verified by the NOE pattern and the coupling constant. The NOE pattern for the inserted EF-loop III of CaM-CD2-III-5G is shown in Figure 4.23. One of the main characteristics of the  $\beta$ -strand formation is the strong NOE  $d_{\alpha N}(i, i+1)$ . The strong NOE crosspeak  $d_{\alpha N}(i, i+1)$  for residues YIS (positions 7 to 9) indicates  $\beta$ -strand propensity. The  $J_{\text{HNH}\alpha}$  backbone coupling constants for the residues at positions 7 (Y62) and 8 (I63) of the inserted EF-loop III in CaM-CD2-III-5G are 6.74 and 8.29 Hz, respectively. The coupling constant for S64 (position 9 of the EF-loop) was not calculated due to signal overlap. The coupling constants for  $\beta$ -strand secondary structure are usually between 8 to 9 Hz. The experimental coupling constants suggest that positions 7 and 8 of the EF-loop III in CaM-CD2-III-5G have strong  $\beta$ -strand



propensity. The  $\phi$  and  $\psi$  dihedral angles of an amino acid in a  $\beta$ -strand secondary structure are typically close to  $-120^\circ$  and  $+120^\circ$ , respectively. Based on the HN, N, C,  $C\alpha$ , and  $C\beta$  chemical shifts, the TALOS program predicts  $\phi$  angles as  $-113^\circ$  and  $-108^\circ$  for Y62 and I63, respectively. The TALOS program predicts  $\psi$  angles as  $144^\circ$  and  $133^\circ$  for Y62 and I63, respectively. The NOE pattern,  $J_{\text{HNH}\alpha}$  backbone coupling constants, and the predicted dihedral angles suggest that residues Y62, I63, and S64 (positions 7 to 9) have strong  $\beta$ -strand propensity.

The EF-loop III in the calculated structures for CaM-CD2-III-5G does not form the  $\beta$ -strand secondary structure and as a result of limited NOE distance restraints and the isolated EF-loop does not have another pair of EF-hand motif to form pair-pair interactions. Due to the resonances overlapping in the homonuclear NOESY spectrum, all of the NOE distance restraints used in the calculation were correlated to the HN proton using the  $^{15}\text{N}$  editing experiment. No sidechain NOE distance restraints were used. In both calcium free and calcium loaded calmodulin, there are hydrogen bonding networks between the paired EF-hand motifs within each domain. The NMR studies by Bax and Ikura observed several NOE interactions between the  $\beta$ -strands of the paired EF-hand motifs (site I and II for N-terminal, site III and IV for C-terminal domain) (60, 62). The rigidity in this section of the EF-loop likely results from the contribution of the  $\beta$ -strand forming a small  $\beta$ -sheet secondary structure with the  $\beta$ -strand on the paired EF-hand motifs. Furthermore, studies by Forsen and coworkers have

shown that there are hydrophobic interactions between the two helices of the EF-hand motif in both the calcium free and calcium loaded forms (63). In the calcium free form, the helix has a more helix bundle orientation. In the presence of calcium, the helices re-orientate themselves to accept the calcium ion. For both the calcium free or calcium loaded form, the entering and exiting helices provide stable support to the EF-loop. Hence, the  $\beta$ -strand secondary structure is observed in the calmodulin structure. In our design, we are interested in the metal binding properties of the EF-loop, so we did not include the helices in the CaM-CD2-III-5G. In the native calmodulin, each domain has two EF-hand motifs. Our model protein only includes a single EF-hand loop. The stability of the helices will be evaluated with engineered variants that include flanking helices in future studies.

#### **4.2.5.6.2 Dynamic Properties of the EF-loop Residues**

Previous observations by Bax and Ikura all suggested that the 6 residues at the N-terminal of the EF-loop are dynamically disordered in the calcium free form (65). Their hydrogen exchange experiments have shown that the backbone amide exchange rate for the 6 N-terminal residues is less than 200s. Previous studies also indicated that few NOE restraints are observed for the first half of the EF-loop. The hydrogen exchange experiments conducted on the inserted EF-

loop III in our lab have produced similar results. The  $S^2$  order parameters for the calcium free EF-loop III have lower values than those of the same residues in calmodulin (see chapter 5.0) (65). These data indicate that the two glycine linkers have provided flexibility to the inserted EF-loop III, and its motion is not restricted by the CD2 host protein. In the absence of the calcium ion, the negative charges from the charged sidechain residues are not balanced. Therefore, the loop conformation becomes dynamically disordered.

#### **4.2.5.6.3 No $3_{10}$ Helix Formation and Exiting Helix Formation Observed in the EF-loop III of CaM-CD2-III-5G**

The  $3_{10}$ -type hydrogen bonds are shown for site I and site III of the calcium free calmodulin where NOE interactions can be observed between the last residue of the entering helix and the first residue of the EF-loop as well as between the third or fourth residue to the last residue of the entering helix. The resonances for Asp at position 1 of the inserted EF-loop III heavily overlaps at N chemical shifts with Ser and HN with Glu at positions 9 and 12 of the inserted EF-loop III. Nevertheless, no additional NOEs have been observed from Asp to the glycine linker. These results are consistent with the secondary structure preferences exhibited by Gly residues, where at least six consecutive residues are required for formation of a helical secondary structure. Consequently, the glycine residue is usually considered one of the least helix forming residues.

Based on the observation from NMR spectra, no  $3_{10}$  helix is formed at the beginning of the inserted EF-loop III of CaM-CD2-III-5G.

#### **4.3 Residual Dipolar Coupling Studies**

The residual dipolar couplings have become an important tool for NMR structural calculations, where the additional orientation restraints provide complimentary structural information to support the NOE distance restraints. Orientation restraints are especially useful when the NOE experiments show limited NOE distance restraints for certain areas of the protein (for example, the inserted EF-loop III of CaM-CD2-III-5G). There are several applications for the orientation restraints in structural determination, and they can be used to refine the distance-restraint-based NMR structure, generate a new structure from a molecular fragment, or they can generate target structures based on a structural template that has a similar global fold (similar to homology modeling except the structure is generated based on the residual dipolar coupling experimental data) (120). For example, Bax and co-workers determined the solution structure of CaM using the residual dipolar couplings, which provide additional structural information on the domain swapping of CaM (measured in liquid crystalline medium) (69). Using the  $^1\text{H}$ - $^{15}\text{N}$  residual dipolar couplings, Chou and Bax also demonstrated that the structure of apo CaM can be built using structures of Ca(II)-loaded CaM, Ca(II) free recoverin, or Ca(II) loaded parvalbumin as templates (153).

The EF-loop III region of the CaM-CD2-III-5G structure was calculated using limited NOE restraints. In this study, residual dipolar couplings were measured using a PEG-bicelles (3.3% w/v) medium to further elucidate the conformation of the inserted EF-loop (Figure 4.24). The orientation restraints were applied in the structural determination using the CNS program.

#### **4.3.1 Residual Dipolar Coupling from External Medium Induced Alignment**

Data from experiments on residual dipolar coupling were collected at UGA using the Inova 800 MHz NMR with a cryogenic probe. The sum and the difference spectra of the IPAP-HSQC experiment for anisotropic and isotropic conditions are all shown in Figure 4.25. A total of 80 residual dipolar couplings were observed from the experiment (Table 4.9). Among the 36 missing couplings, there are two prolines, five glycines from the linker, and the remaining residues either had signal overlap or not were observed at this condition. The RDC values for the host protein ranged from -14.66 to 14.84 Hz. The RDC values for the inserted sequence ranged from -0.73 to 2.11 Hz, which is much lower than the RDC values of the host protein. The EF-loop III was inserted in the solvent-exposed region with flexible linkers. **It is possible that the EF-loop III did not align with the field like the host protein.** The EF-loop III may have remained in the isotropic condition.

#### **4.3.2 Calculated Order Parameters using REDCAT**

To obtain the Euler angles, the principle order parameter ( $S_{zz}$ ), and the asymmetry parameter  $\eta$ , the structure constructed by CYANA was used to calculate the order parameter using the REDCAT program (119-121). For the REDCAT calculation, the number of error space sampling was set at 10,000 and the residues with an estimated error of  $\geq 5.50$  Hz were discarded. After the error correction, a total of 57 residual dipolar couplings were used (5 to 7, 12 to 17, 20, 21, 24, 26, 27, 30 to 32, 35, 36, 38 to 44, 49, 57 to 59, 62, 67, 68, 71, 73, 74, 75, 77, 79, 86 to 90, 94 to 96, 98, 102, 107, 108, 110, 111, 114, and 115).

The Euler angles for  $\alpha$ ,  $\beta$ , and  $\gamma$  are -36.63, 159.85, and 64.49, respectively. The principal order parameter  $S_{zz}$ ,  $S_{xx}$ , and  $S_{yy}$  are  $-7.11 \times 10^{-5}$ ,  $-4.11 \times 10^{-4}$ , and  $4.82 \times 10^{-4}$ , respectively. The value for  $\eta$   $[(S_{xx} - S_{yy})/S_{zz}]$  and GDO parameters are  $7.04 \times 10^{-1}$  and  $5.20 \times 10^{-4}$ , respectively (120). The order parameters obtained above were then used to back calculate the residual dipolar coupling, and the comparison between the experimental and calculated dipolar coupling is shown in Figure 4.26. The anisotropy parameter “Da” was calculated to be 6.77  $[(D_{max} \times S_{zz})/2]$  (120). The rhombicity parameter “R” was calculated to be 0.705.

### 4.3.3 Structural Refinement Using Residual Dipolar Couplings

Structural refinement using residual dipolar couplings for CaM-CD2-III-5G was performed with CNS (123). The NOE distance restraints and dihedral angle restraints tables were transferred from the CYANA calculation. CYANA can read

CNS style NOE restraint tables with minor modifications (for example, CYANA uses H instead of HN for backbone amide protons), so it is more beneficial to generate the NOE restraint table in CNS format. Similar to the NOE distance restraint optimization, the orientation restraint that caused the largest violation was removed from the calculation. A total of 57 orientation restraints were used for structural refinement. The structures calculated both without and with the orientation are shown in Figure 4.27a and 4.27b, respectively. The RMSD (using residues G4 to F42) among the 20 lowest energy structures are 0.303 and 0.404 Å for structures that are calculated without and with orientation restraints, respectively. The orientation restraints should improve the quality of the solution structure. However, the structure that is calculated with orientation restraints shows no improvement compared to the structure without the orientation restraints. Several possible reasons are offered to explain this. First the Dmax (24350 Hz) was calculated assuming the  $^1\text{H}$ - $^{15}\text{N}$  bond length was 1.02 Å. The order parameters were simulated using the CYANA calculated structure, which has an average bond length of approximately 0.99 Å. Second, the order parameters will be simulated based on different fragments of CaM-CD2-III-5G. For example, the calculations will be carried out on the whole, first half (G4 to L50), second half (F72 to R113, since the EF-loop was inserted half way into the protein sequence), GFCC'C" face (CD2 is a  $\beta$ -sheet protein with two faces), and ABDE face of the protein. The order parameters were calculated using different fragments of CaM-CD2-III-5G and it will be used to refine the structure of CaM-

CD2-III-5G in the future. Third, the residual dipolar couplings for CaM-CD2-III-5G will be measured using a different medium to verify the alignment. The residual dipolar couplings can also be measured using different PEG-bicelle concentrations. Different bicelle concentrations will yield different magnitudes of dipolar coupling but it will not prevent problems such as non-specific binding between the PEG-bicelle and protein. Fourth, the residual dipolar couplings for  $^1\text{H}$  to  $^{13}\text{C}\alpha$  or  $^{13}\text{CO}$  will also be measured to be use in conjunction with the  $^1\text{H}$ - $^{15}\text{N}$  dipolar couplings to improve the structural refinement.

#### 4.4 Paramagnetic Induced Alignment Using Ln(III) Metal Ions

The paramagnetic effect arises from metalloprotein binding to a paramagnetic metal ion with unpaired electrons, which induces an anisotropic magnetic moment. The paramagnetism restraints are very useful for structural refinement; they are pseudo contact shift (PCS), paramagnetic relaxation enhancement (PRE), cross-correlated relaxation (CCR) between the Curie-spin and  $^1\text{H}$ - $^{15}\text{N}$  dipole-dipole relaxation, and residual dipolar coupling (RDC).

PCS is distance dependent and subject to the same paramagnetic susceptibility tensor as the residual dipolar coupling (Equation 4.3).

$$\delta^{PCS} = \frac{1}{12\pi r^3} \left[ \Delta\chi_{ax}(3\cos^2\theta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\theta\cos 2\varphi \right] \quad 4.3$$



$\Delta\chi_{ax}$  and  $\Delta\chi_{rh}$  are the axial and rhombic magnetic susceptibility tensors.  $\theta$  and  $\phi$  are the polar coordinates of the nucleus with respect to the magnetic susceptibility tensor. The  $r$  is the distance between the metal ion and the paramagnetic affected nucleus.

The PRE is also a distance restraint parameter similar to the NOE distance restraint and the distance restraint occurs between the paramagnetic metal ion and the paramagnetically-affected nuclei.

$$\mathcal{A}^{PRE} = \frac{k}{r^6} \left( 4\tau_r + \frac{3\tau_r}{1 + \omega_H^2 \tau_r^2} \right) \quad 4.4$$

The CCR paramagnetic restraint is a result of the Curie relaxation cross-correlated with the  $^1\text{H}$ - $^{15}\text{N}$ -dipole-dipole relaxation. CCR provides the distance information from the paramagnetic metal to the paramagnetic affected nuclei. CCR also provides the angle between the coupled nuclei ( $^1\text{H}$ - $^{15}\text{N}$ ) dipole with respect to the metal ion.

$$\eta^{CCR} = k \frac{3\cos^2 \nu - 1}{r^3} \left( 4\tau_r + \frac{3\tau_r}{1 + \omega_H^2 \tau_r^2} \right) \quad 4.5$$

The paramagnetic-induced RDC is similar to the external alignment media induced RDC which can be translated into polar angles  $\theta$  and  $\phi$ , which describe

the orientation of an intermolecular vector ( $^1\text{H}$ - $^{15}\text{N}$ ) to the applied magnetic field as a function of the paramagnetic susceptibility tensor.

$$\Delta\nu^{RDC} = k \left[ \Delta\chi_{ax} (3\cos^2\Theta - 1) + \frac{3}{2} \Delta\chi_{rh} \sin^2\Theta \cos 2\Phi \right] \quad 4.6$$

The paramagnetic induced RDC is not distance dependent with respect to the metal ion. The main difference between the paramagnetic induced RDC and the external alignment media induced RDC is that the external alignment orients the molecule to the magnetic field by external force such as space restriction or ionic interaction. The paramagnetic alignment is induced because the bound metal ion is aligned to the magnetic field. The direction and the magnitude of the magnetic anisotropy tensor are different between the two alignment methods. However, the two alignment methods can be applied to the same molecule simultaneously and the contributions of each alignment occur together, as it has been demonstrated using calbindin<sub>D9k</sub> with either Tb(III) or Dy(III) and dissolved in nonionic liquid crystalline phase alignment medium (154).

The PCS is a contribution to the hyperfine shift. The other contribution to the hyperfine shift is the contact shift from the process of metal binding. The PCS can be obtained by removing the diamagnetic effect in which the chemical shift of the nuclei in diamagnetic condition is subtracted from the chemical shift of the corresponding nuclei in the paramagnetic condition. Paramagnetism studies have been performed on calmodulin, calbindin<sub>D9k</sub>, and troponin C, parvalbumin,

and other metal binding proteins that bind lanthanides (141, 143, 155, 156). For proteins without metal binding sites, a lanthanide(III) tag can be chemically added to the protein for paramagnetism studies (157, 158).

The paramagnetic studies on calbindin<sub>D9k</sub> using different lanthanides by Bertini et al. have indicated that the direction of the principle axes of the magnetic tensor are very similar but different in magnitude. The lanthanides with larger magnetic anisotropy have the ability to relax nuclei at greater distances, so the detectable paramagnetic zone is farther away from the metal ion. At the same time, the signal of nuclei that are closer to the metal are broadened beyond detection, which is referred to as a blind zone (Figure 4.27) (159). The lanthanide with smaller magnetic anisotropy has a smaller detectable paramagnetic zone and a smaller blind zone in comparison with the lanthanides with larger magnetic anisotropy. This can provide more structural information for the nuclei that are closer to the metal. This was demonstrated by Bertini et al using the PCS restraints from Ce(III), Yb(III), and Dy(III) to calculate the structure calbindin<sub>D9k</sub> using PSEUDODYANA (159, 160). Ce(III) affect nuclei that are between 5 to 15 Å away from the metal. Dy(III) affect nuclei that are between 13 to 40 Å away from the metal. The structure calculated using the NOE restraint was used for comparison. The structure calculated using both the NOE and Ce(III) paramagnetic restraints showed improvement in the RMSD value for residues that are closer to the metal ion in comparison with the NOE restraint only structure. The structure using NOE and Dy(III) paramagnetic restraints

revealed lower RMSD values for residues that are further away from the metal ion, but higher RMSD values for residues that are close to the metal ion due to the larger blind zone of Dy(III). Since the paramagnetic tensor for lanthanides has similar direction, the PCS restraints from different lanthanides can be used together to refine protein structure (141). The PCS is a useful restraint for structural calculation, but alone is not sufficient. PCS can define the distance of the paramagnetic affected nucleus to the metal with respect to the magnetic anisotropic tensor, but still requires a few NOEs to define the folding of the protein and the orientation of the secondary structure (142). The majority of paramagnetic metal ions in the Ln(III) family have a similar ionic radius to Ca(II). CaM-CD2-III-5G is a good candidate for paramagnetism studies since the inserted calcium binding site is from an EF-hand protein, which can substitute the Ca(II) for Ln(III). The paramagnetism studies for CaM-CD2-III-5G were carried out using Tm(III) and Dy(III). Tm(III) has similar orbital properties as Yb(III), so the size of the blind zone and detectable paramagnetic zone are similar. The blind zone for Yb(III) was reported as 9 Å and the detectable paramagnetic zone was reported up to 25 Å. The blind zone and the detectable paramagnetic effected zone for Dy(III) are 13 and 40 Å, respectively.

#### **4.4.1 Tm(III) Metal Titration of CaM-CD2-III-5G**

The Tm(III) titration was completed to determine if the isolated EF-loop in CD2 will have metal binding affinity for Tm(III). The HSQC spectra of the CaM-

CD2-III-5G Tm(III) titration are shown in Figure 4.28. At the Tm(III) concentration of 235  $\mu$ M, the crosspeaks for the EF-loop residues D58, A59, N60 (also the sidechain), G61, Y62, and E67 disappeared. The peak intensity for A65 was very weak at 235  $\mu$ M of Tm(III) and disappeared at 289  $\mu$ M of Tm(III). The crosspeaks for the glycine linker residues G55 and G68 disappeared at the Tm(III) concentration of 289  $\mu$ M. The peak intensities for the host protein residues decreased at a Tm(III) concentration of 166  $\mu$ M. The crosspeak for residues D25 disappeared at a Tm(III) concentration of 235  $\mu$ M. The crosspeaks for host protein residues D26, R31, L38, E41, F42, and K43 disappeared at the Tm(III) concentration of 289  $\mu$ M. At the Tm(III) concentration of 500  $\mu$ M, the peak intensities for residues 49, 50, 71, 72, 102, 27, 34, 45, 47, 51, 78, 86, 88, 100, 103, 35, 57, 73, 94, 95, 24, 46, 87, 90, 92, 99, 108, 110, 116, 80, 81, 84, 85, 107, 111, 112, 77, 79, 83, and 93 (residues are listed based on the intensity in decreasing order) decreased more than 60%. The HN chemical shifts of the CaM-CD2-III-5G in the 500  $\mu$ M were plotted against the chemical shift of the metal free protein (Figure 4.29). Since all of the resonances for the inserted EF-loop disappeared at 289  $\mu$ M of Tm(III), the remaining resonances of the CD2 host protein were used for chemical shift comparison. The HN chemical shift of residues T24, M46, F72, T86, D88, S99, G102, and T103 changed more than 0.05 ppm in the presence of 500  $\mu$ M Tm(III), while the chemical shift of the rest of the CD2 host protein changed less than 0.05 ppm.

The crosspeaks for the inserted EF-loop and some of the CD2 host protein disappeared during the Tm(III) titration. The crosspeaks for the inserted EF-loop disappeared at lower Tm(III) concentrations in comparison to the CD2 host protein, which indicate that the EF-loop has a specific binding pocket for Tm(III). The observed changes in the peak intensities and the HN chemical shifts on the CD2 host protein are due to two reasons. First, some residues are close to the inserted EF-loop, such as residues M46, F72, T86, and D88. The peak intensities and the HN chemical shifts of these four residues changed as a function of the Tm(III). The distance between these residues were measured using the NMR structure of CaM-CD2-III-5G with a simulated metal binding site (later in this chapter). The residues M46, F72, T86, and D88 are all within 25 Å of the metal ion. Second, the changes of some residues are due to non-specific Tm(III) binding. The non-specific binding effects are also observed in the La(III) titrations.

#### **4.4.2 Residual Dipolar Coupling from Field Induced Alignment**

The paramagnetic CaM-CD2-III-5G sample was prepared by adding the lanthanide stock solution to the calcium saturated CaM-CD2-III-5G to reach a 0.88:1 metal and protein ratio. The paramagnetic restraints, PCS and RDC, were obtained from the  $^{15}\text{N}$  HSQC and  $^{15}\text{N}$  HSQC-IPAP spectra, respectively. The assignment of lanthanide bound CaM-CD2-III-5G was achieved by comparing the spectrum to the metal free spectrum of CaM-CD2-III-5G.

The Tm(III) paramagnetic study was carried out using the Varian Inova 800 MHz and the Dy(III) paramagnetic study was carried out using the Varian Inova 600 MHz (Figures 4.30 and 4.31 for Tm(III) and Dy(III), respectively). The La(III) diamagnetic experiments were carried out in both field strengths as the diamagnetic reference. In the presence of 350  $\mu$ M La(III), the resonances for D58, G59, N60, G61, Y62, I63, and E67 broadened beyond detection. The chemical shifts and peak intensities of resonances G55, A65, and G68 shifted and decreased, respectively. The chemical shifts of the host protein residues remained unchanged.

The HSQC experiment observed 67 crosspeaks for the backbone of Tm(III)-CaM-CD2-III-5G (Figure 4.30). In the presence of 350  $\mu$ M Tm(III), the crosspeaks of residues D25, D26, D28, E29, V30, R31, W32, E33, L38, V39, A40, E41, F42, K43, R44, G55, K57, D58, G59, N60, G61, Y62, I63, A65, E67, G68, I74, Y98, G102, and T103 broadened beyond detection. The PCS and RDC of CaM-CD2-III-5G induced by Tm(III) on 800 MHz NMR are shown in Table 4.10. All of the crosspeaks from the inserted EF-loop and the glycine linkers broadened beyond detection. The HN PCS of residues T24, M46, D89, and S90 were -0.04, -0.03, 0.04 and 0.03 ppm, respectively. The PCS for the remaining 63 resonances were less than 0.020 ppm for Tm(III)-CaM-CD2-III-5G. The RDC of CaM-CD2-III-5G induced by Tm(III) on 800 MHz NMR ranged from -2.43 up to 1.53 Hz. The previous work by Bertini et al reported the PCS of calbindinD9k in the presence of Tm(III) between 0.76 to 4.09 ppm using the 800

MHz NMR (141). The Tm(III) induced RDC for calbindinD9k were reported between -17.8 to 14.4 Hz (141). The PCS and RDC observed for Tm(III)-CaM-CD2-III-5G are very small in comparison to the calbindinD9k in the presence of Tm(III). It is possible that the flexible glycine linkers have reduced the paramagnetic alignment for the host protein region of CaM-CD2-III-5G (see section 4.4.4.2 for additional information).

The HSQC experiment observed 50 crosspeaks for the backbone of Dy(III)-CaM-CD2-III-5G. Dy(III) has a strong magnetic susceptibility anisotropy compared to Tm(III), so Dy(III) will induce higher magnitude of PCS and RDC. All of the crosspeaks that were broaden beyond detection in the Tm(III) experiment were also broaden beyond detection in the Dy(III) paramagnetic experiment. In addition, the crosspeaks of residues I27, R34, T37, T45, K47, K51, A71, F72, L75, T86, D88, D89, S99, T100, I105, and N107 broadened beyond detection in the Dy(III) experiment. The PCS and RDC of CaM-CD2-III-5G induced by Dy(III) on 600 MHz NMR is shown in Table 4.10. All of the resonances from the inserted EF-loop and the glycine linkers broadened beyond detection. Only 50% of the residues in the CD2 host protein were observed, while the rest also broadened beyond detection. The PCS of residues E33, L80, S90, and E116 in the presence of Dy(III) were -0.03, -0.04, -0.03, and -0.06 ppm, respectively. The PCS for the remaining 46 resonances were less than 0.020 ppm for Dy(III)-CaM-CD2-III-5G. Residue L80 had the largest RDC value of -12.89 Hz. The RDC of the remaining residues were between -3.31 to 3.51 Hz.



The previous work by Bertini *et al* reported the PCS of calbindinD9k in the presence of Dy(III) between 0.81 to -4.38 ppm using the 800 MHz NMR (141). The Tm(III) induced RDC for calbindinD9k were reported from -26.7 to 19.1 Hz (141). The PCS and RDC observed for Dy(III)-CaM-CD2-III-5G were very small in comparison to the calbindinD9k in the presence of Dy(III). It is also possible that the flexible glycine linkers reduced the paramagnetic alignment for the host protein region of CaM-CD2-III-5G, which was observed in the Tm(III) experiment (see section 4.4.4.2 for additional information).

The Tm(III) and Dy(III) paramagnetic studies indicated that the coverage of the blind zone reaches the opposite end of the host protein. The expected blind zone for both metals should be less than 13 Å away from the metal center. The paramagnetic affected nuclei in CaM-CD2-III-5G are classified into two classes. Class I are the residues from the inserted EF-loop, the glycine linkers, K51, A71, F72, I74, L75, T86, D88, and D89. Class II are the residues that are on the opposite end of the CD2 host protein (if the EF-loop was inserted at the south end of the protein, then class II residues are located at north end of the protein). And they are D25 to R34, T37 to K47, Y98 to T100, G102, T103, I105, N107, and K108. To verify the distances between the paramagnetic affected nucleus and the metal ion, a simulated metal containing CaM-CD2-III-5G was used for measurement. The metal containing CaM-CD2-III-5G structure was calculated using a combination of the experimental NOE restraints and the metal binding restraints based on the geometric description from calmodulin

(3CLN.pdb) using CYANA (Figure 4.32). The distances between the metal ions to the paramagnetic affected nuclei were measured for residues with resonance broadened beyond detection (Table 4.11). Table 4.11 contains the metal to nucleus distance for the average structure and lowest energy structure of CaM-CD2-III-5G. Because the conformation of the inserted EF-loop is very flexible, Table 4.11 also includes the lowest and highest metal to nucleus distance among the 20 lowest energy structures. Two binding models were proposed to analyze the paramagnetic studies on CaM-CD2-III-5G. In model one, CaM-CD2-III-5G only has one metal binding site, the inserted EF-loop III. In model two, CaM-CD2-III-5G includes the inserted EF-loop as one binding site as well as a charged location in CD2 with binding capabilities.

#### **4.4.3.1 Model One: The Inserted EF-loop Metal Binding Site**

##### **Class I**

The EF-loop was inserted into CD2 with two glycine linkers and each glycine linker contains three glycine residues. Based on the average simulated metal-CaM-CD2-III-5G structure, the distance between the metal ion to the host protein residues K51 (S52 was not observable) and A71 are more than 12 Å away. The expected blind zone for Tm(III) and Dy(III) are approximately 9 and 13 Å, respectively. The resonances for the inserted EF-loop and glycine linkers broadened beyond detection in both the Tm(III) and Dy(III) experiments. These results are expected since the HN protons of these residues are less than 10 Å

away from the metal ion. The resonance for I74 is broadened beyond detection in the Tm(III) experiment. The resonances for K51, A71, F72, I74, L75, T86, D88, and D89 are broadened beyond detection in the Dy(III) experiment. The additional paramagnetic effects observed with the Dy(III) experiment indicate that the Dy(III) metal ion has a larger paramagnetic coverage than the Tm(III) metal ion. The average structure suggests that the HN proton of the A71 is less than 13 Å away from the metal, while the HN protons of K51, F72, I74, L75, T86, D88, and D89 are more than 14 Å away. Among the twenty structures with the lowest target function energies, the lowest distance between the metal, and the HN proton of residues K51, A71, F72, I74, L75, T86, D88, D89 are 14.63, 11.06, 12.95, 14.28, 18.81, 13.11, 11.96, and 13.73 Å, respectively. The average structure indicates that only the HN proton of residue A71 is less than 13 Å away from the metal, while the rest of the residues are outside the suggested blind zone for Dy(III) but within the detectable paramagnetic affected zone. The paramagnetic effects observed on residues K51, F72, I74, L75, T86, D88, and D89 are due to the conformational flexibility of the inserted EF-loop. The resonances for the surrounding residues such as E73, L85, and R87 are not broadened beyond detection, but the peak intensities decreased significantly in comparison with the same residues in the diamagnetic condition.

Class II

The crosspeaks for D25 to W32, L38 to R44, Y98, G102, and T103 are broadened beyond detection in the presence of Tm(III) and Dy(III). The resonances for R34, T37, K45, K47, S99, T100, I105, N107, and K108 are broadened beyond detection in the presence of Dy(III). The HN protons of these residues are between 18.7 to 38.9 Å away from the metal ion based on the average CaM-CD2-III-5G structure. The lowest possible distance between the metal ion and the HN protons of these residues are between 16.8 to 33.8 Å. These residues are all outside the blind zone for both metal ions and their signal should not be broadened beyond detection. The paramagnetic effect observed in this region of the CD2 host protein is not a result of the inserted EF-loop binding with paramagnetic metal ion.

#### **4.4.3.2 Model II: The EF-loop and a Metal-sticking Site of CD2**

##### **Residues D25, D26, D28, and E29**

CD2 contains 8 Asp, 5 Glu, 8 Asn, and 1 Gln with sidechains exposed on the protein surface and these residues are potential metal binding ligands. Residues D25, D26, D28, and E29 are four negatively charged residues that are in close proximity and these residues are part of the class II residues, defined as a metal-sticking site. The distance between the paramagnetic nuclei were measured between the HN protons of D26 and D28 to estimate the distance to the metal-sticking site (Table 4.12). The RMSD of the twenty lowest energy structures for the host protein section of CaM-CD2-III-5G are 0.378 (G4 to L50)

and 0.225 (F72 to R113), which indicates that the conformation of the host protein is not dynamically disordered and the distance measurement was performed on the average structure.

The class I residues are more than 20 Å away from the HN proton of D26 except I74 and L75. Residues I74 and L75 are 13.8 and 15.2 Å away from the HN proton of D28, respectively. While the rest of the class I residues are more than 18 Å away. The paramagnetic effect observed in class I residues is a result of the EF-loop binding with paramagnetic metal ions, which is in good agreement with the one metal binding site model. The residues in class II are all within 20 Å to either D26 or D28 except T37 and L38. Residues V30, R31, F42, K43, R44, K45, M46, Y98, S99, T100, G102, T103, and I105 are less than 10 Å away from the HN proton of D28 and these residues are within the blind zone of both Tm(III) and Dy(III). The distance between the HN protons of class II residues to the HN proton of either D26 or D28 indicate that the paramagnetic effects observed for the class II residues are a result of the metal-sticking sites at D25, D26, D28, and E29.

The non-specific, or sticking interaction observed in CaM-CD2-III-5G was verified by performing the paramagnetic experiment with the wild type CD2. The non-specific interactions between the metal and the surface residues of CD2 can be reduced by increasing the salt concentration.

#### **4.4.4.1 What are the Reasons for Low Magnitude of PCS and RDC?**

### Multiple Metal Binding Sites

The PCS and RDC for Tm(III)-CaM-CD2-III-5G and Dy(III)-CaM-CD2-III-5G are small in comparison to that for calbindinD9k. The presence of a true metal binding site (the inserted EF-loop) and non-specific metal -sticking site does not reduce the paramagnetic effect of the lanthanides as was demonstrated in the paramagnetic studies of calmodulin by Biekofsky *et al* (144). The two EF-hand motifs at the N-terminal substituted Ca(II) for Tb(III). The RDC were determined with different Tb(III) to protein ratios. The observed RDC for CaM1 (calmodulin-EF1(Tb<sup>3</sup>)-EF2(Ca<sup>2</sup>)-EF3(Ca<sup>2</sup>)-EF4(Ca<sup>2</sup>)), CaM2 (calmodulin-EF1(Ca<sup>2</sup>)-EF2(Tb<sup>3</sup>)-EF3(Ca<sup>2</sup>)-EF4(Ca<sup>2</sup>)), and CaM3 (calmodulin-EF1(Tb<sup>3</sup>)-EF2(Tb<sup>3</sup>)-EF3(Ca<sup>2</sup>)-EF4(Ca<sup>2</sup>)) are all different. CaM1 and CaM2 only have Tb(III) ions occupied in one of the two EF-hand sites in the N-terminal domain. Each metal induced different alignments to the magnetic field. In CaM3, both of the EF-hand sites in the N-terminal bound Tb(III) and the measured RDC reflects the contribution of both paramagnetic metal ions. It is difficult to determine the contribution of the metal-sticking site to the PCS and RDC of Tm(III)- and Dy(III)-CaM-CD2-III-5G. If the metal-sticking sites also induce alignment to the magnetic field, the paramagnetic effect observed for CaM-CD2-III-5G would be a sum from both the inserted EF-loop and the non-specific metal binding site. The small paramagnetic effect observed on CaM-CD2-III-5G is not due to the influence of multiple metal ions.

#### 4.4.4.2 Did the Tm(III) and Dy(III) Induce Alignment?

The resonances for residues in class I and II disappeared in Tm(III)-CaM-CD2-III-5G and Dy(III)-CaM-CD2-III-5G indicating the presence of paramagnetic metal ions. The detectable paramagnetic zone for Tm(III) and Dy(III) are approximately 25 and 40 Å, respectively. If Dy(III) binds to the inserted EF-loop, the detectable Dy(III) paramagnetic zone would cover the entire dimension of the CD2 host protein. The Tm(III)-CaM-CD2-III-5G spectrum indicates that only 4 residues out of the 67 observable residues have chemical shift changes larger than 0.03 ppm. The Dy(III)-CaM-CD2-III-5G spectrum indicates that only 4 residues out of the 50 observable residues have chemical shift changes larger than 0.03 ppm. The largest PCS values for Tm(III)- and Dy(III)-CaM-CD2-III-5G are -0.04 (T24) and -0.06 (E116), respectively. In Tm(III)-CaM-CD2-III-5G, residues T24, M46, D89, and S90 have PCS values greater than 0.03 ppm, but the PCS for the surrounding residues such as F21, Q22, K45, K47, D88, and G91 are less than 0.02 ppm. Dy(III) has a stronger magnetic moment than the Tm(III) and 17 additional resonances are broadened beyond detection in the presence of the Dy(III). But the remaining detectable crosspeaks have very small chemical shift changes for both Tm(III)- and Dy(III)-CaM-CD2-III-5G. The missing crosspeaks in the HSQC spectra for both metals indicate the occupancy of metal ion in the binding pocket. The paramagnetic metal will induce the inserted EF-loop to align with the field. Because the inserted EF-loop is inside the blind zone for both Tm(III) and Dy(III), the resonances are broadened beyond

detection. The glycine linkers allow the EF-loop to bind with a metal ion without restriction from the host protein. As a result, the host protein is flexible in the solution and does not align to the field following the aligned loop. The conformational average results in low PCS values for Tm(III)- and Dy(III)-CaM-CD2-III-5G.

RDC shares the same paramagnetic anisotropy tensor as the PCS, but RDC is not distance dependent. The RDC provides information on how the  $^1\text{H}$ - $^{15}\text{N}$  vector is aligned to the magnetic field regardless of the location of the metal ion. The magnitudes of the RDC for Tm(III)- and Dy(III)-CaM-CD2-III-5G are lower than the Tm(III)- and Dy(III)-calbindinD9k, respectively. Calbindin has two EF-hand motifs packed close to each other, which induces higher magnitude RDC. The Ln(III)-tag approach usually attaches the Ln(III)-binding loop to the N-terminal domain of the protein to induce paramagnetic alignment for proteins without natural metal binding sites. The Ln(III)-tag has less interaction with the protein in comparison to the natural EF-hand calcium binding protein, so the magnitude of the RDC will be lower. The RDC values for Ln(III)-tag-ubiquitin with Dy(III) are between -6.6 to 6.1 Hz (157). The RDC values for Ln(III)-tag-ubiquitin with Tm(III) are between -2.9 to 4.5 Hz (157). It is important to point out that the magnitude of the RDC for different systems are not directly comparable since the orientation parameters are different, but the RDC values observed in the calbindinD9k and Ln(III)-tag studies can be used as a guideline for the magnitude of RDC values that we observed in our system. The magnitude of the RDC for

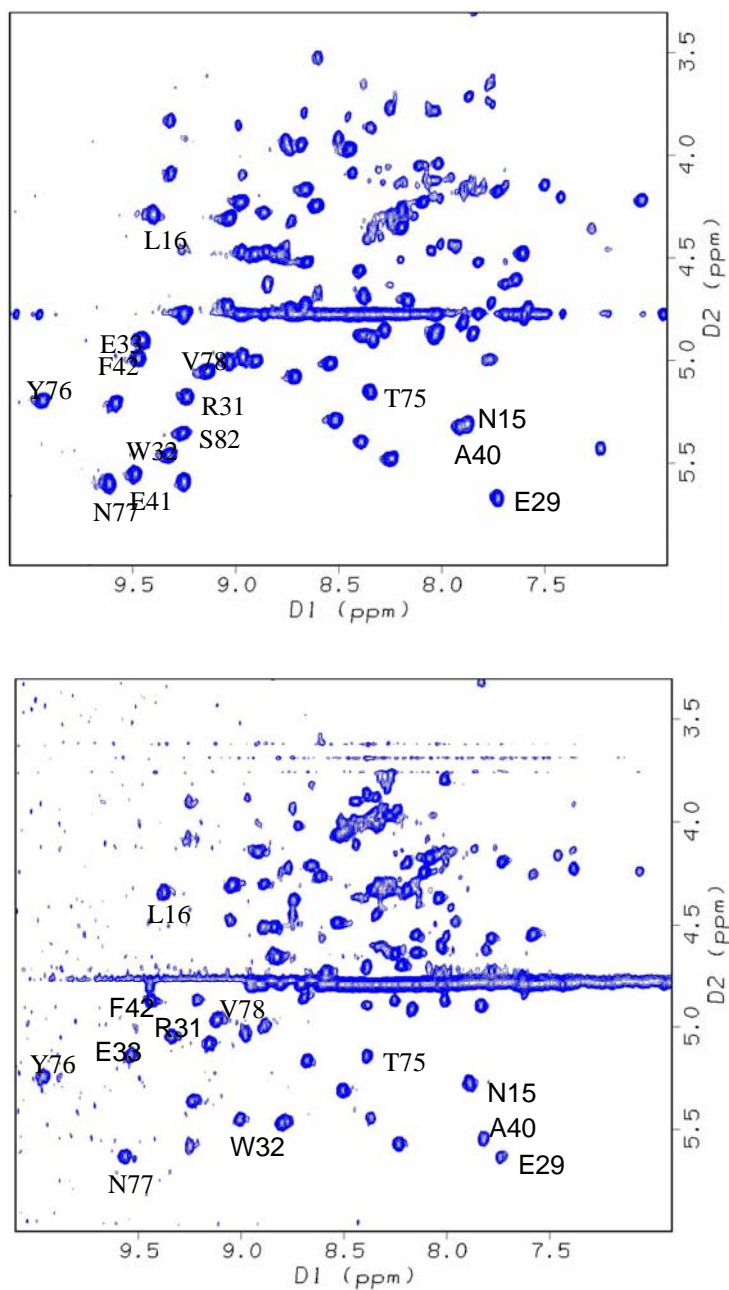


Ln(III)-tag is lower than intact EF-hand proteins such as calbindin D9k, but the magnitude of the RDC for Ln(III)-tag is still higher than CaM-CD2-III-5G.

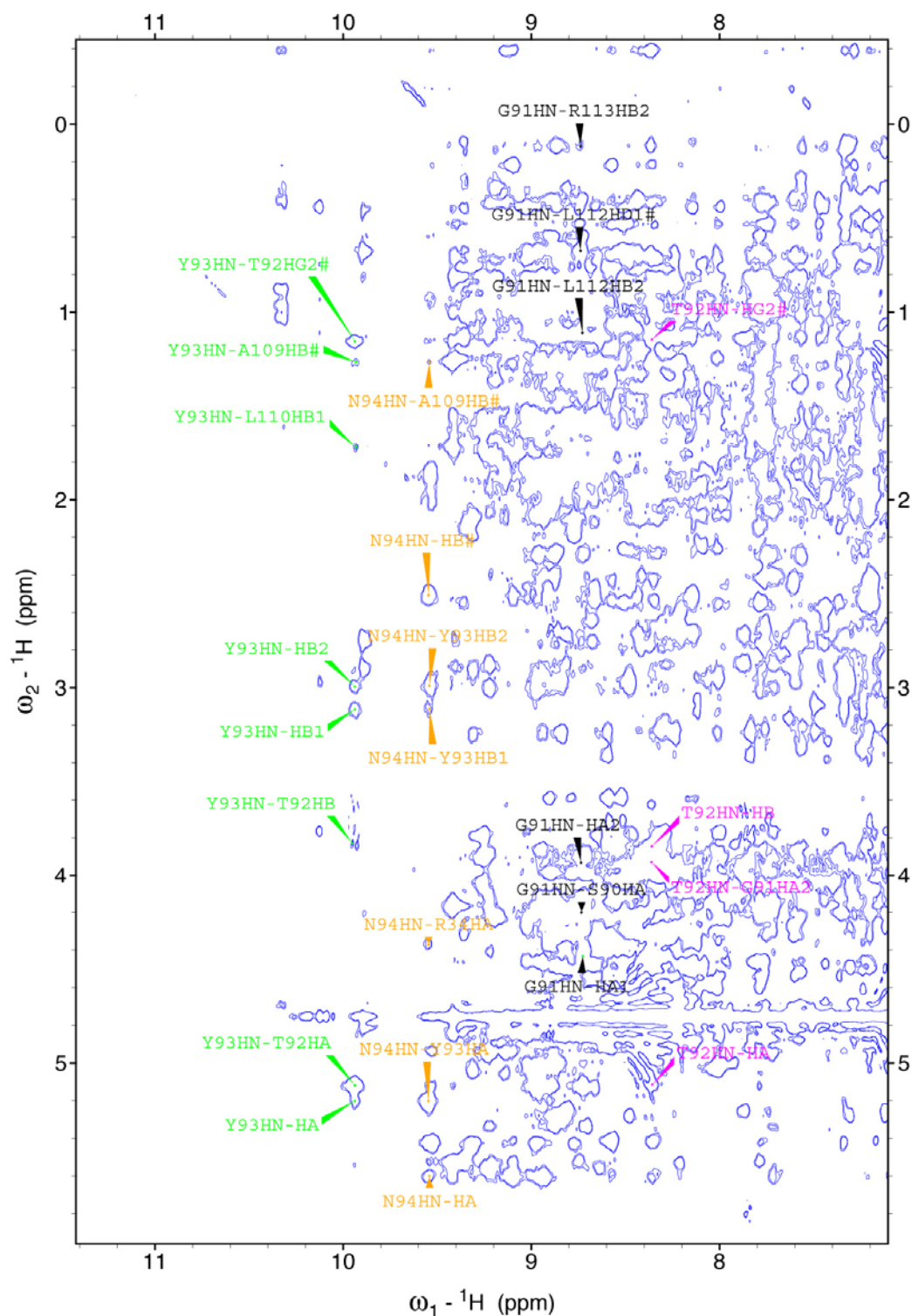
The smaller magnitude of the RDC for CaM-CD2-III-5G is due to the conformational flexibility between the EF-loop and the host protein, which causes the host protein to be isotropic, while the inserted EF-loop is anisotropic. Since the RDC and PCS shares the same paramagnetic anisotropy tensor, the RDC results are in good agreement with the PCS results. The RDC values are affected by the conformational average. It is interesting to point out that the external alignment media has the opposite effect on CaM-CD2-III-5G. In the external alignment method, the CaM-CD2-III-5G was aligned to the magnetic field using PEG-bicelles, which form disk layers to restrict the orientation of the protein. The RDC values for the host protein region of CaM-CD2-III-5G are between -14.0 to 14.8 Hz. But the inserted EF-loop still retains the isotropic condition because the glycine linkers allow the EF-loop to move freely. The inserted EF-loop is too small, so the PEG-bicelles are not able to restrict the orientation of the EF-loop. Therefore, the RDC for the inserted EF-loop is close to 0 Hz.

The conformational effects on the PCS and RDC values were previously studied for calmodulin. The two EF-hand motifs in the N-terminal domain of calmodulin substituted Tb(III) for Ca(II), while the two EF-hand motifs in the C-terminal domain retained bound Ca(II) ions. The two domains of calmodulin were connected by a flexible central helix. In the absence of the peptide of

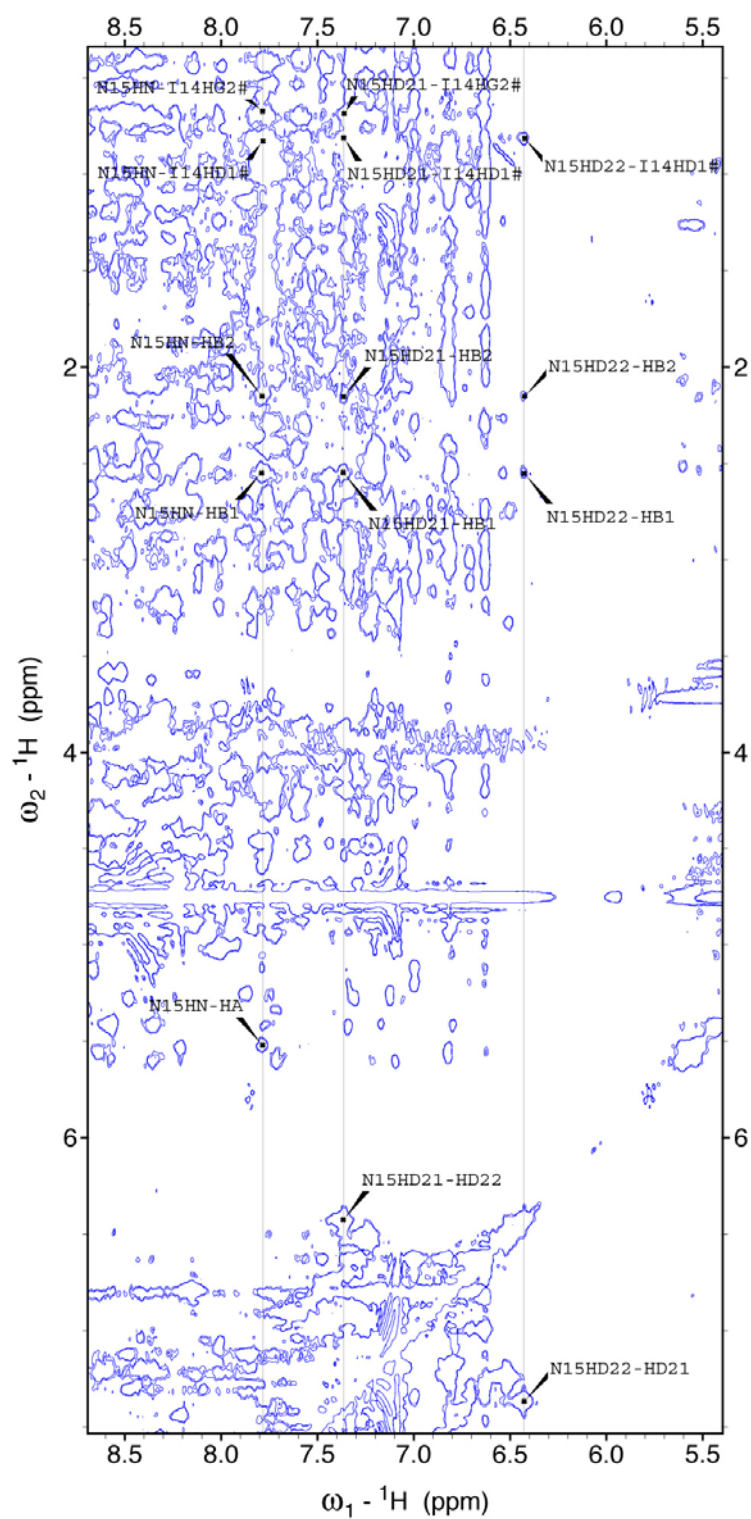
skeletal muscle myosin light chain kinase, the conformation of the C-terminal was independent of the N-terminal (144). The PCS and RDC for the C-terminal were averaged close to zero. In the presence of the peptide, the entire calmodulin protein was observed to wrap around the peptide and became one rigid sub-unit. The magnitude for the PCS and RDC of the C-terminal domain increased. Therefore, it is possible that the small PCS and RDC values observed in the paramagnetic studies are a result of the flexible glycine linkers.



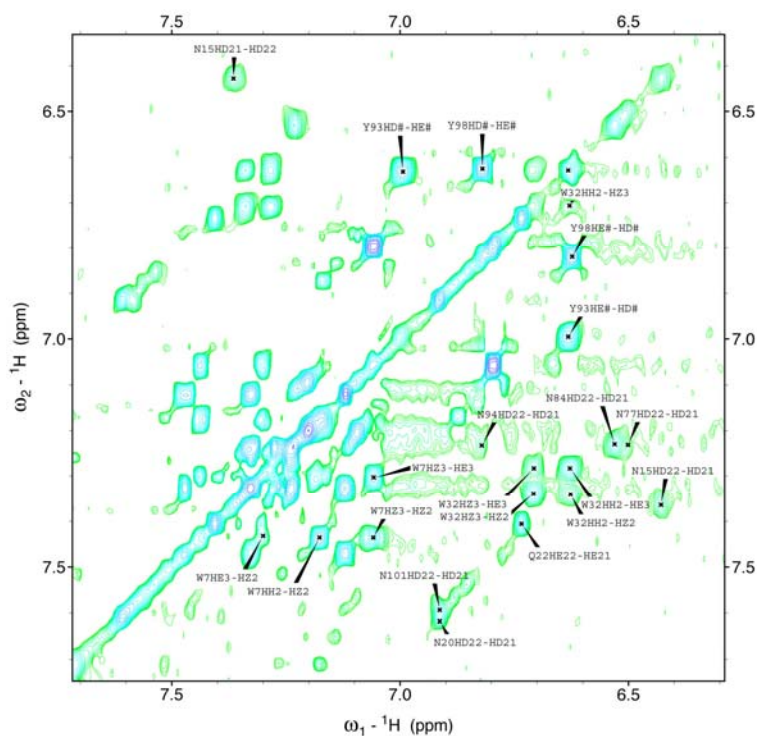
**Figure 4.1 TOCSY spectrum of CD2 and CaM-CD2-III-5G:** The fingerprint region of the TOCSY spectra for CaM-CD2-III-5G (bottom) and wild type CD2 (top).



**Figure 4.2 Example of sequential assignment of CaM-CD2-III-5G:** The sequential assignment in the fingerprint region of CaM-CD2-III-5G

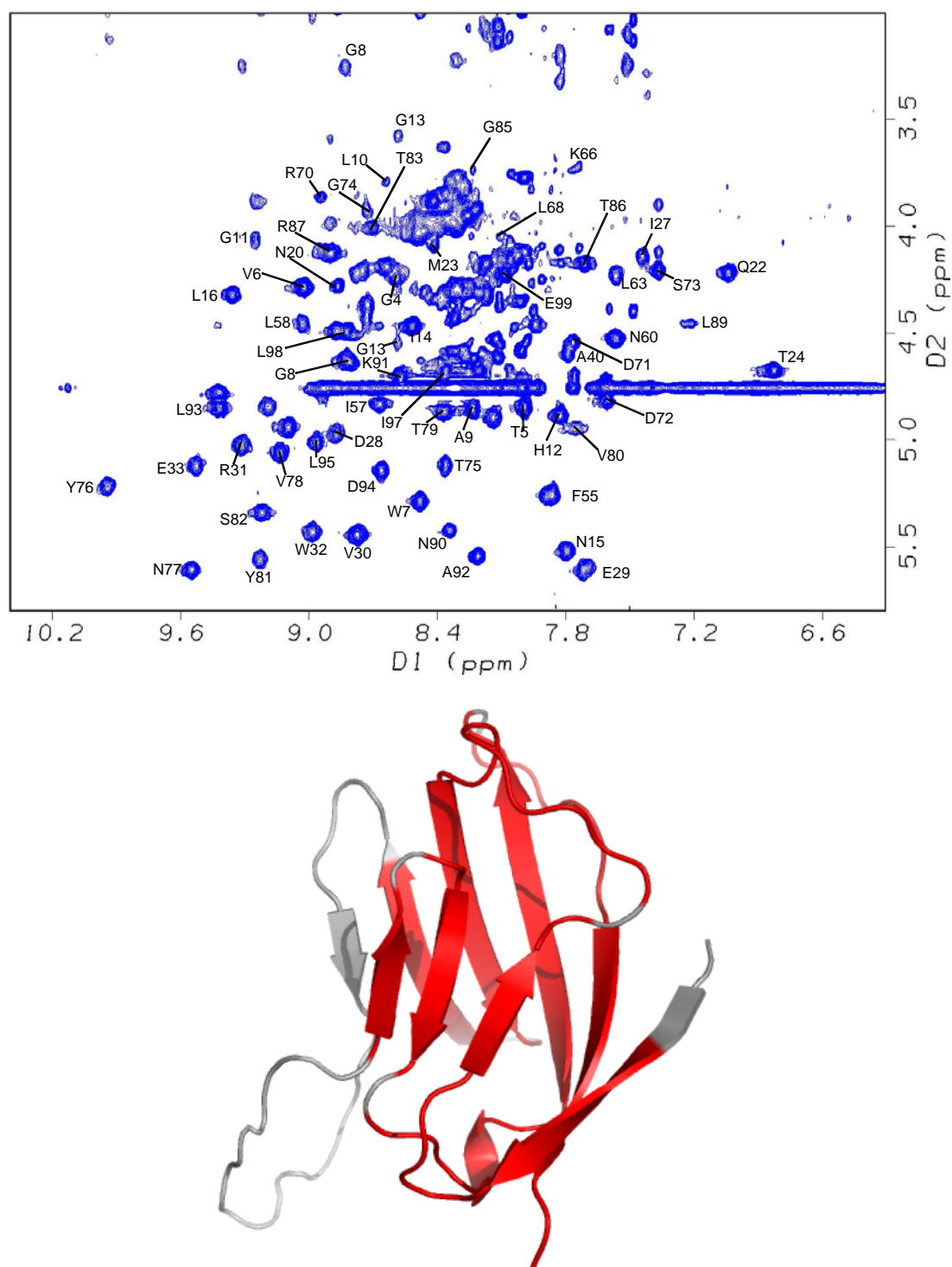


**Figure 4.3 Example of Asn sidechain assignment of CaM-CD2-III-5G:** An example of Asn sidechain NOE assignment for N15. The assignment for the other residues were removed for this example.

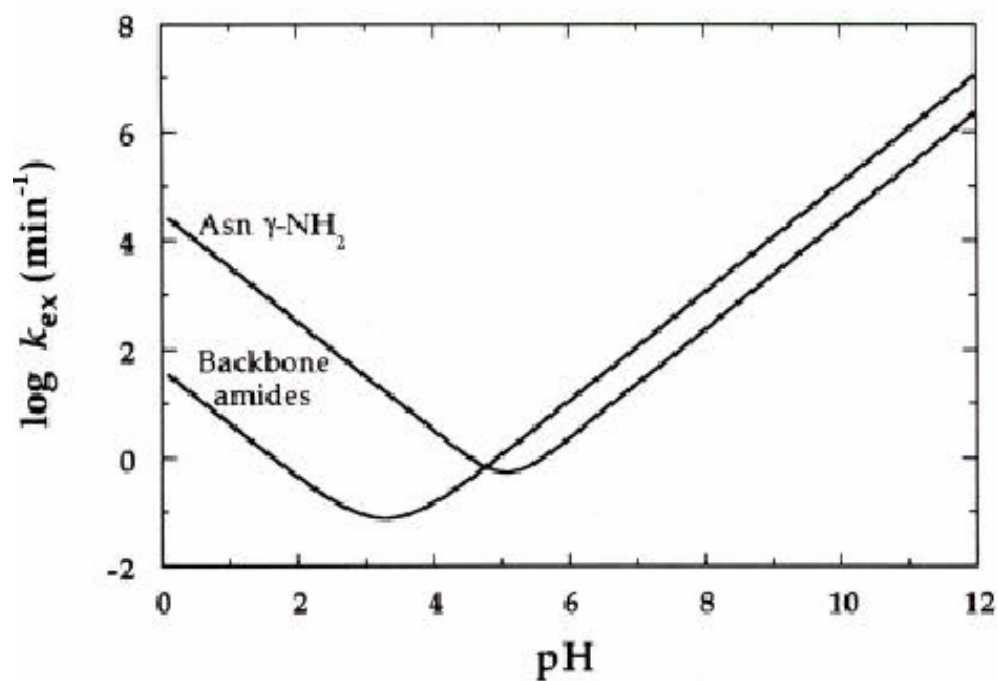


**Figure 4.4 Aromatic ring protons assignment of CaM-CD2-III-5G:** Aromatic region of the TOCSY spectra. CaM-CD2-III-5G in H<sub>2</sub>O is shown in green while the D<sub>2</sub>O form is shown in magenta.



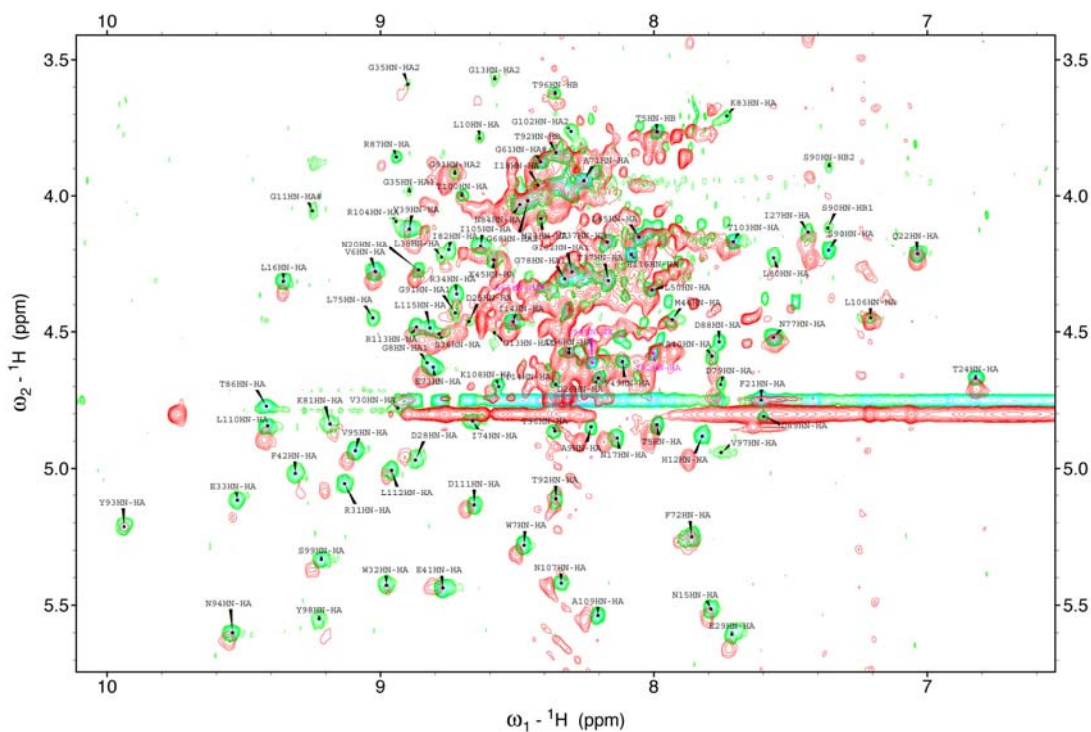


**Figure 4.5 Homonuclear assignment of CaM-CD2-III-5G:**  
 The fingerprint region of the CaM-CD2-III-5G TOCSY spectrum is shown on top. The assigned regions are shown in red while the areas that are not assigned are shown in grey.

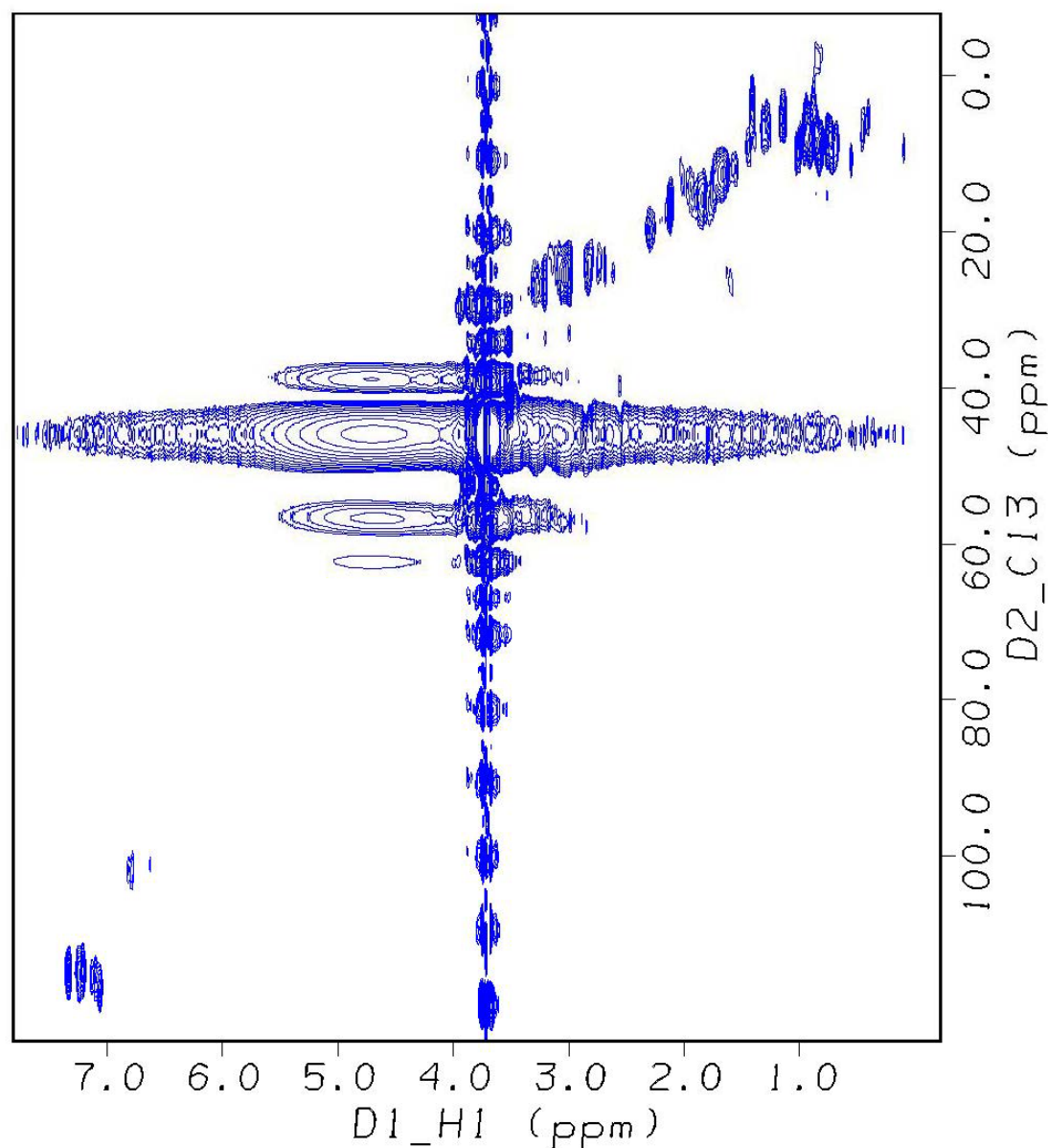


**Figure 4.6 Chemical exchange rates of backbone amides:** The chemical exchange rates of backbone amides and sidechain of Asn changes as a function of pH. By adjusting the pH of the protein to alter the exchange rate may induce some resonance to be observable or caused chemical shift change in NMR spectrum.

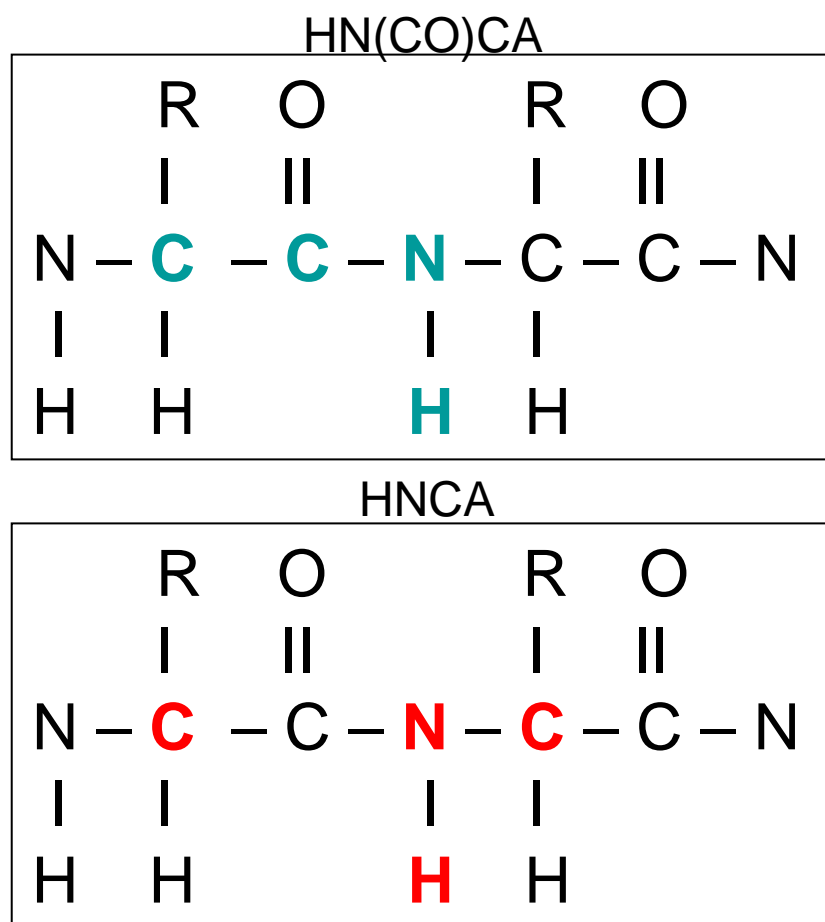




**Figure 4.7 TOCSY spectra of CaM-CD2-III-5G pH 5.0 and 7.4:** The spectrum collected in 10 mM Tris 10 mM KCl in pH 7.4 is shown in green. The spectrum collected in 20 mM Phosphate buffer in pH 5.0 is shown in red. Both spectra were collected using Varian Inova 600 MHz NMR at 25° C.

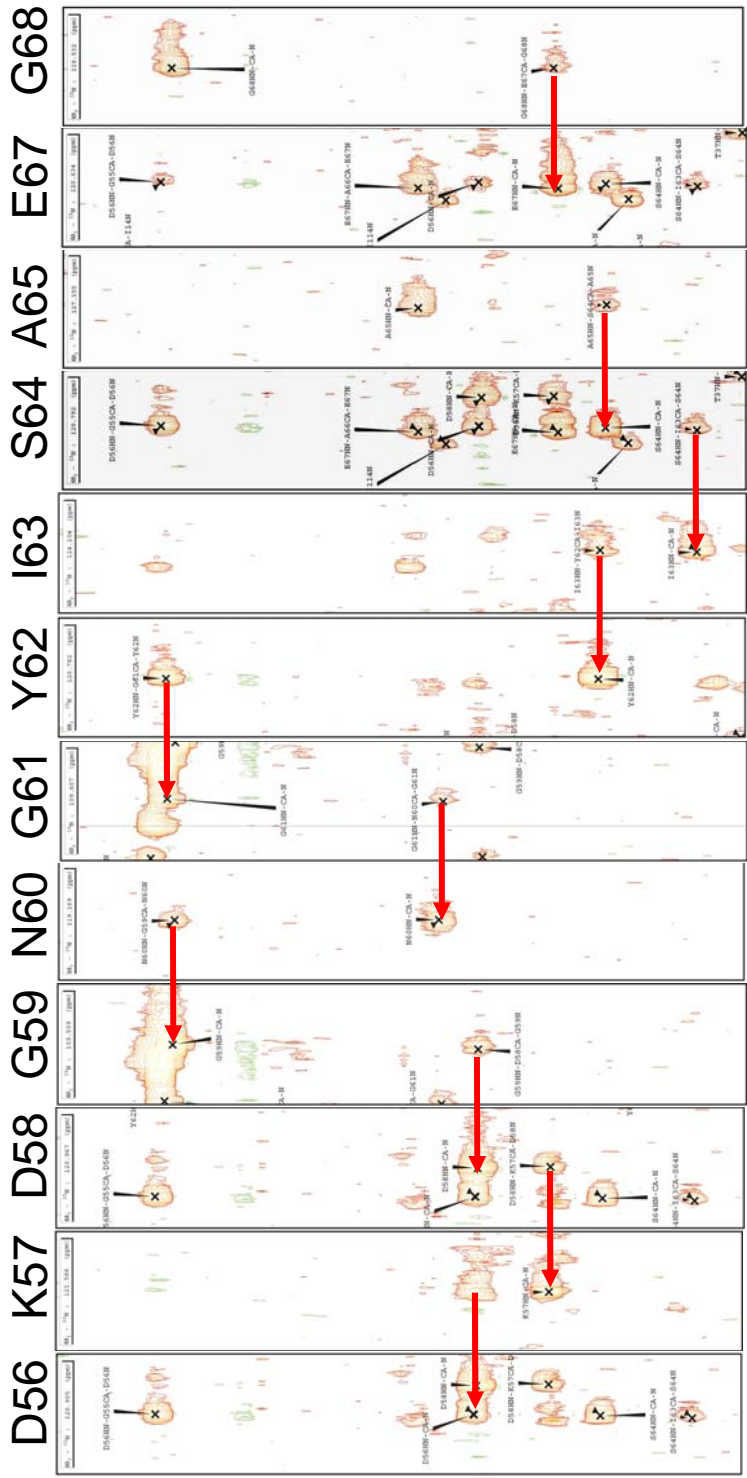


**Figure 4.8** <sup>13</sup>C HSQC spectrum of CaM-CD2-III-5G: The natural abundance carbon HSQC spectrum was collected using 1 mM homonuclear labeled CaM-CD2-III-5G in 100% D<sub>2</sub>O.

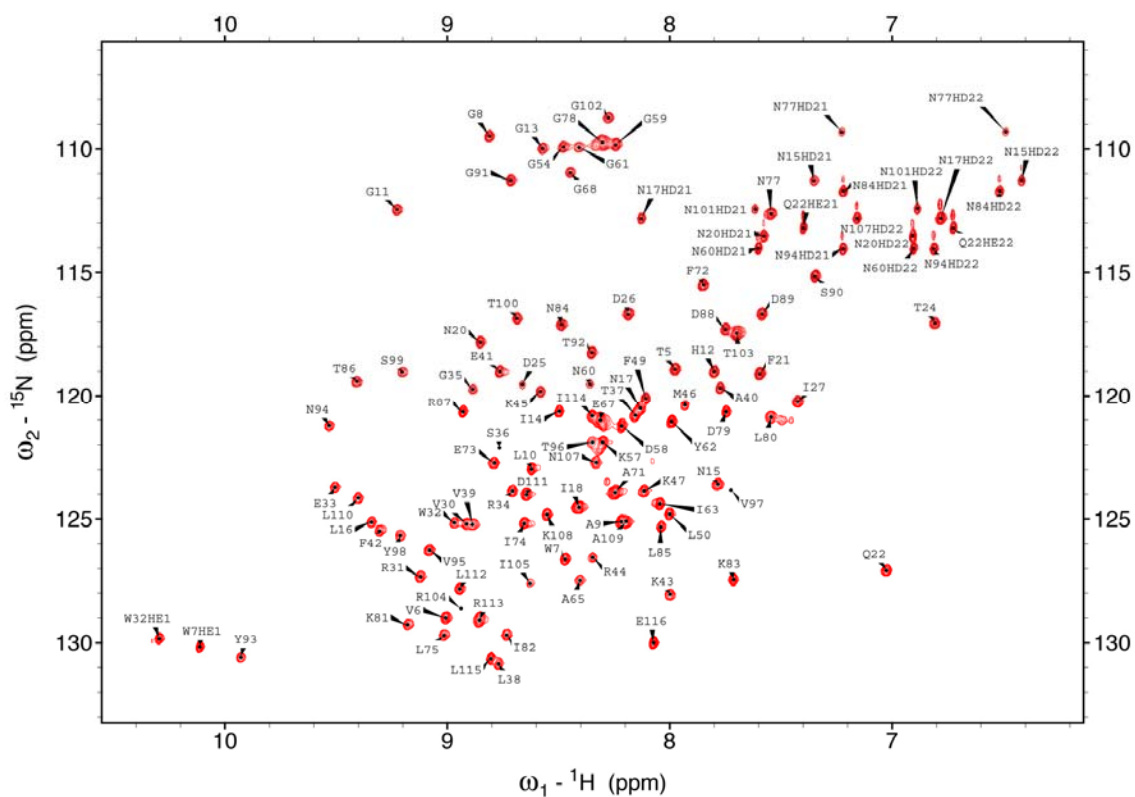


**Figure 4.9 Example of triple resonance assignment**

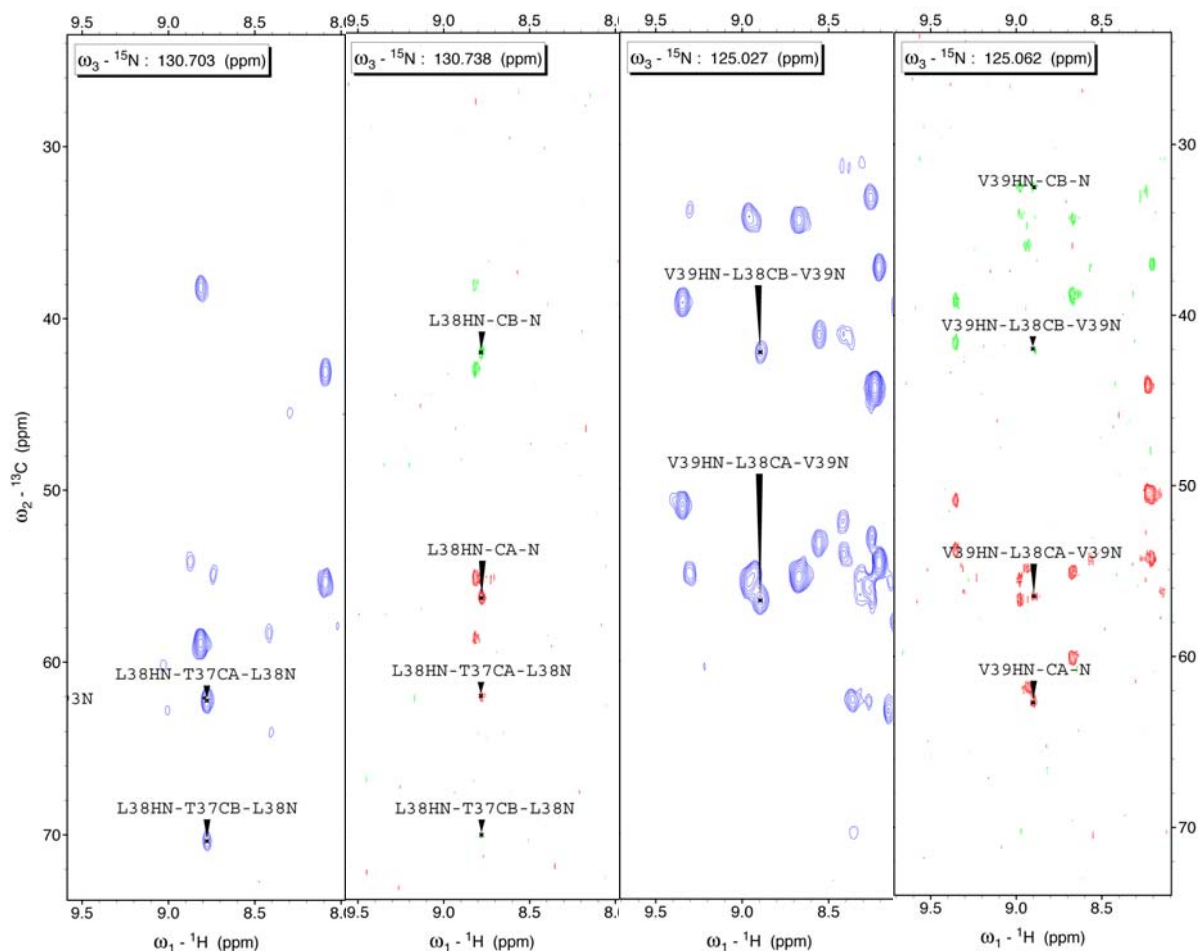
**procedure:** The HNCA experiment correlates the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for a given amino acid residue to the  $^{13}\text{C}_\alpha$  chemical shift of the current and preceding residues in the amino acid sequence. The HNCOCA experiment correlates the  $^1\text{H}$  and  $^{15}\text{N}$  chemical shifts for a given amino acid residue to the  $^{13}\text{C}_\alpha$  chemical shift of the preceding residue in the amino acid sequence. So HNCA and HNCOCA spectra are also able to provide sequential connectivity information (Figure 4.9).



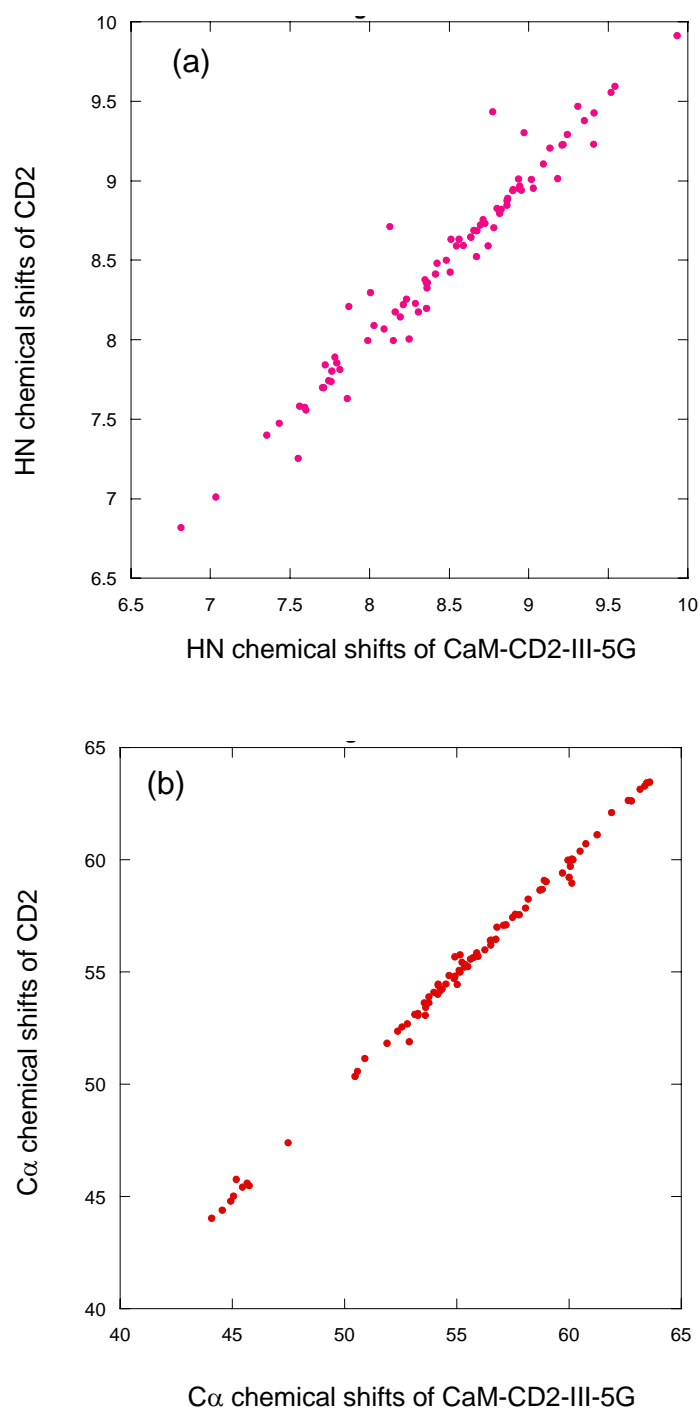
**Figure 4.10 HNCACB strip plot of CaM-CD2-III-5G:** The assigned HNCACB strip plot of the inserted EF-loop III of CaM-CD2-III-5G.



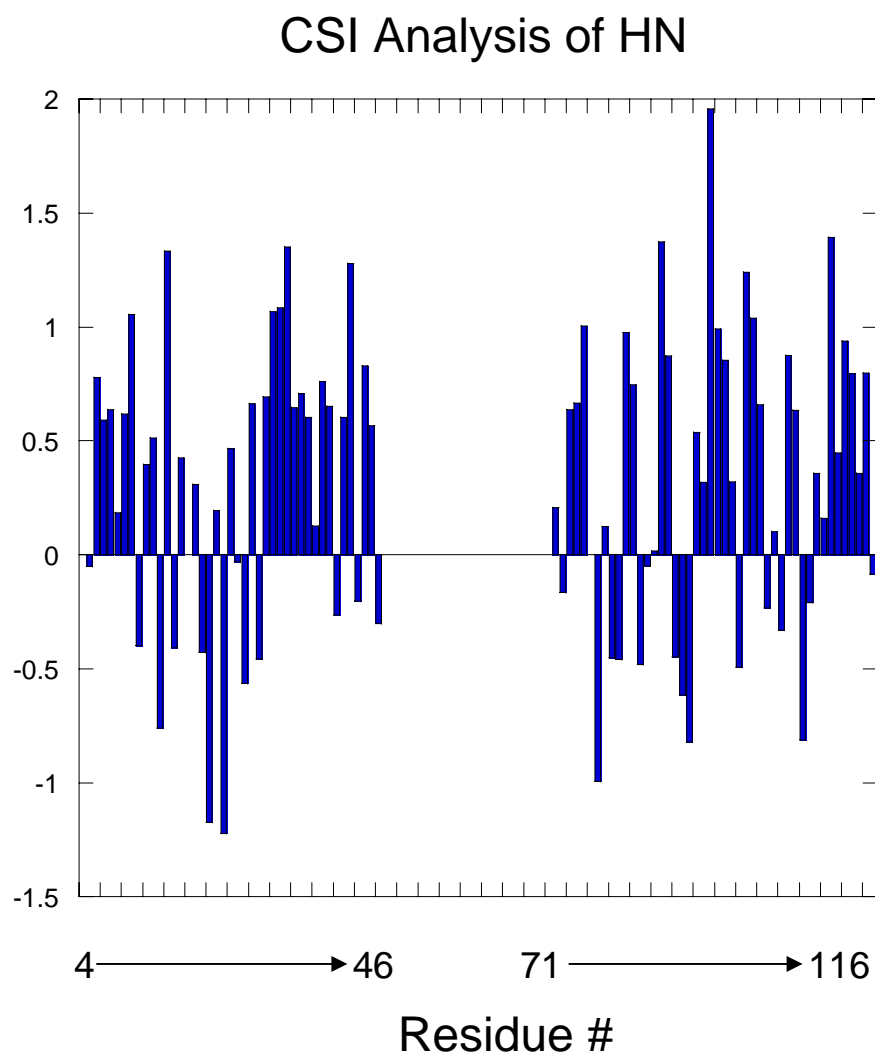
**Figure 4.11 Assigned  ${}^{15}\text{N}$  HSQC spectrum of CaM-CD2-III-5G:** The assigned  ${}^{15}\text{N}$  HSQC spectrum of CaM-CD2-III-5G was collected using Varian Inova 600 MHz NMR at 25 ° C. The sample was prepared in 20 mM PIPES 20 mM KCl at pH 6.8.



**Figure 4.12 Assignment using HNCACB and CBCACONH spectra:** The CBCA(CO)NH spectra are shown in red. The HNCACB spectra are shown in green/red. Both HNCACB and CBCACONH spectra were collected using Varian Inova 600 MHz NMR at 25 ° C. The CaM-CD2-III-5G samples were prepared in 20 mM PIPES 20 mM KCl at pH 6.8.

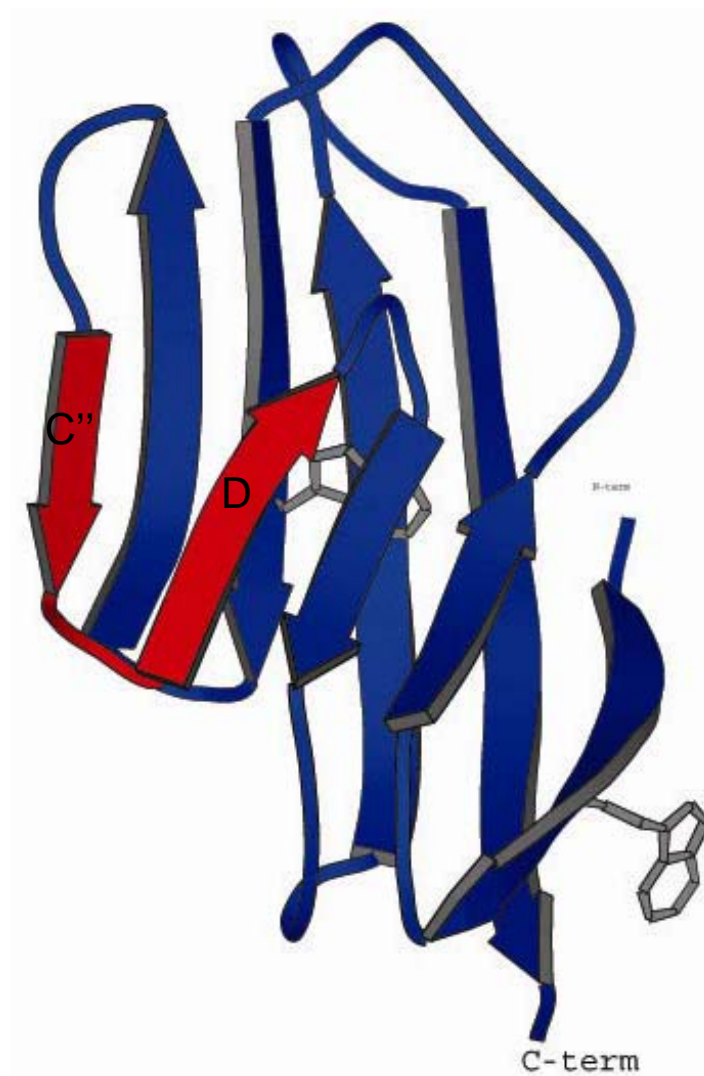


**Figure 4.13 Comparing the HN and C $\alpha$  chemical shifts of CaM-CD2-III-5G to the same residues in wild type CD2:** (a) The HN chemical shifts comparison between the CD2 and CaM-CD2-III-5G. (b) The C $\alpha$  chemical shifts comparison between CD2 and CaM-CD2-III-5G.



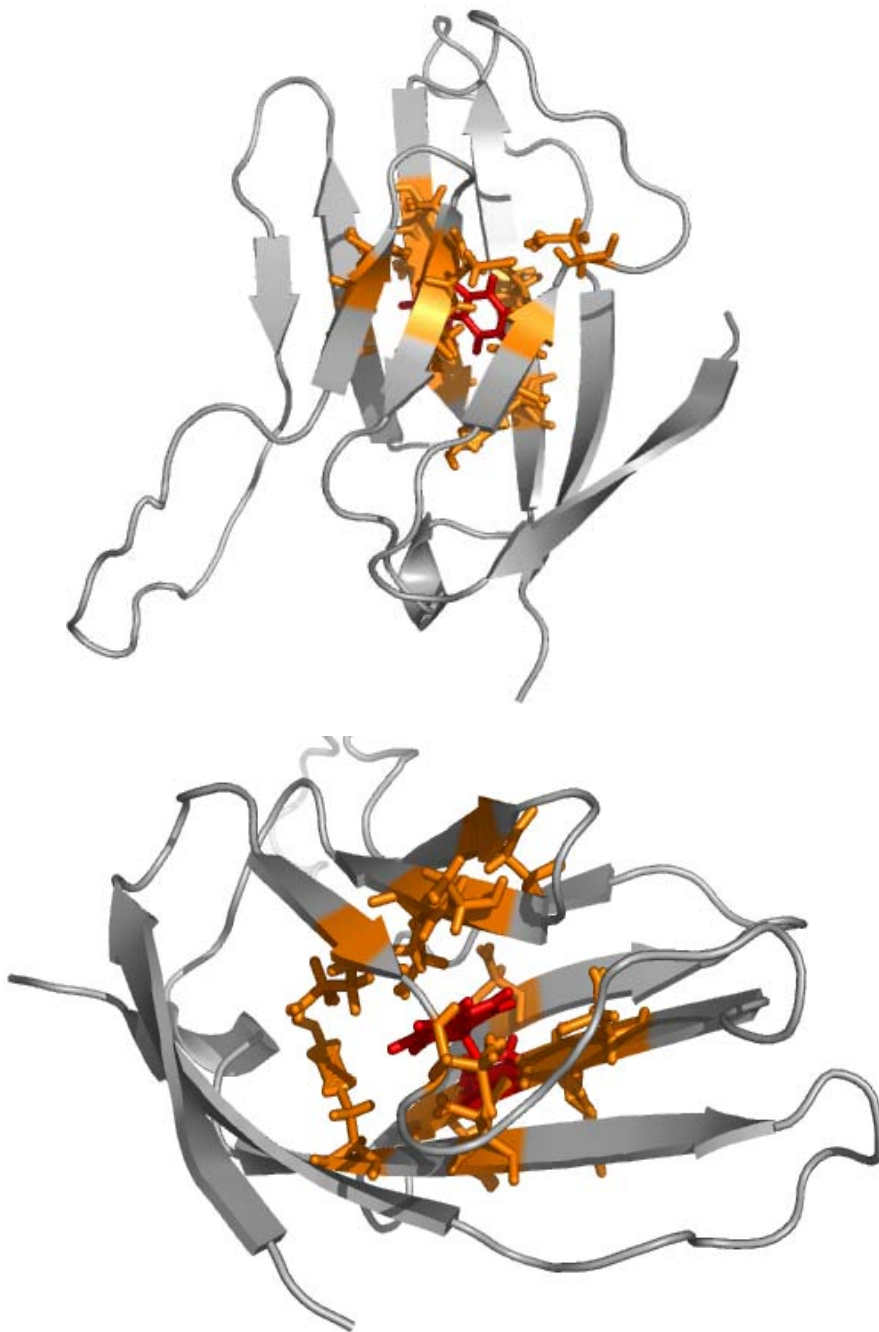
**Figure 4.14 CSI Analysis for CaM-CD2-III-5G:** The CSI analysis was performed by comparing the HN chemical shifts of CaM-CD2-III-5G to the random chemical shifts.



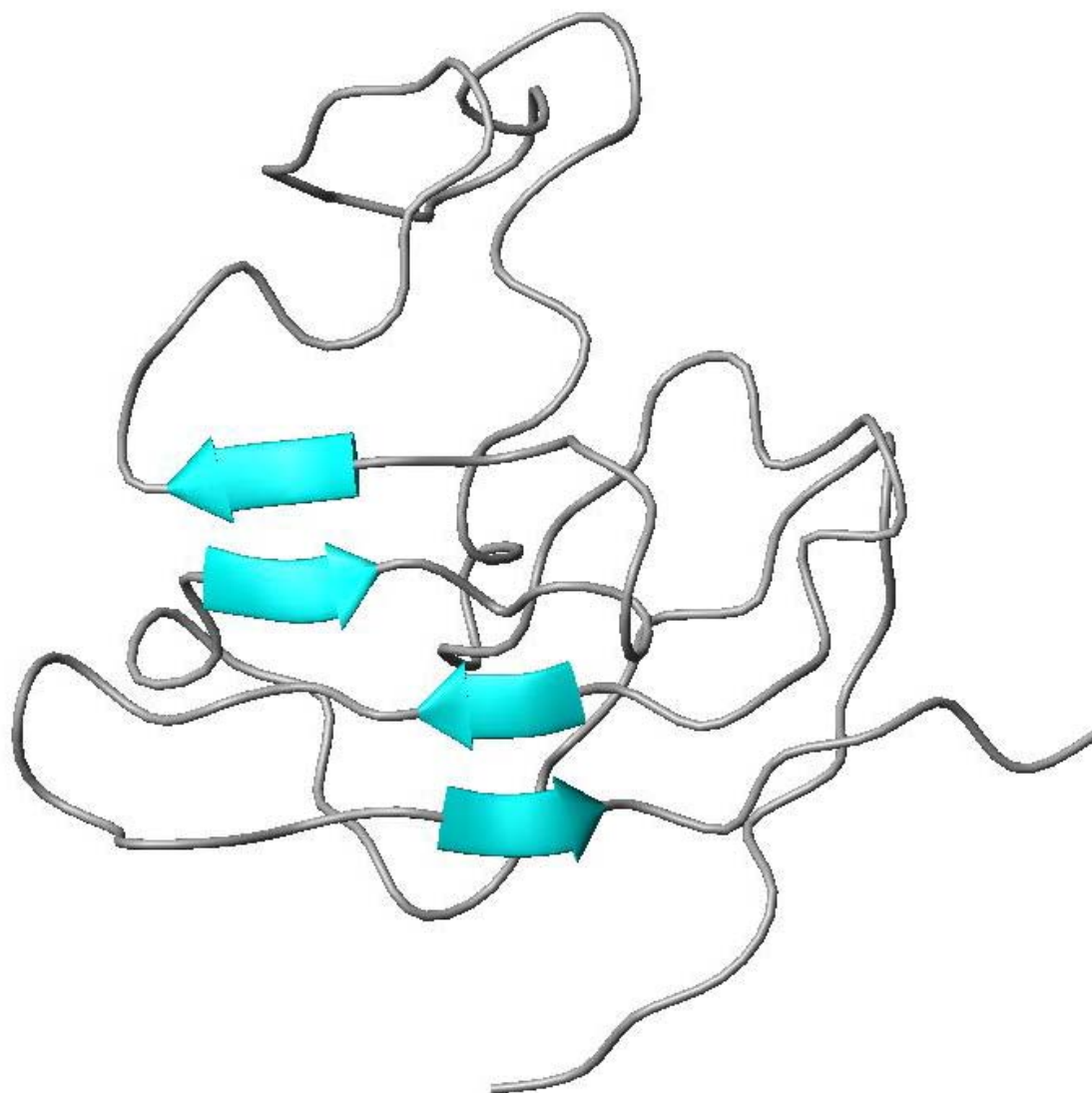


**Figure 4.15 Domain 1 of the cell adhesion molecule, CD2:** CD2 contains two layers of  $\beta$ -sheet. One of the layer is consist of  $\beta$ -strands GFCC'C'' while the other layer is consist of  $\beta$ -strands ABDE. The EF-loop is inserted in between  $\beta$ -strands C'' and D.

**Figure 4.16 Example of the NOESY spectrum of CaM-CD2-III-5G:** This region of the spectrum display NOE connectivities for backbone HN protons, sidechain aromatic ring protons, sidechain HD protons of Asn, and sidechain HE protons of Gln.



**Figure 4.17** The interaction between the W32 and residues that are within 5 Å: The CaM-CD2-III-5G structure is shown in grey. W32 is shown in red and the residues contains protons that are within 5 Å of W32 is shown in yellow. These residues reside in the hydrophobic core of CD2.



**Figure 4.18 Solution structure of CaM-CD2-III-5G with no error correction:** This is the cycle 1 structure of CaM-CD2-III-5G using CYANA 1.1.

Table 4.4 HNHA J-Couplings of CaM-CD2-III-5G

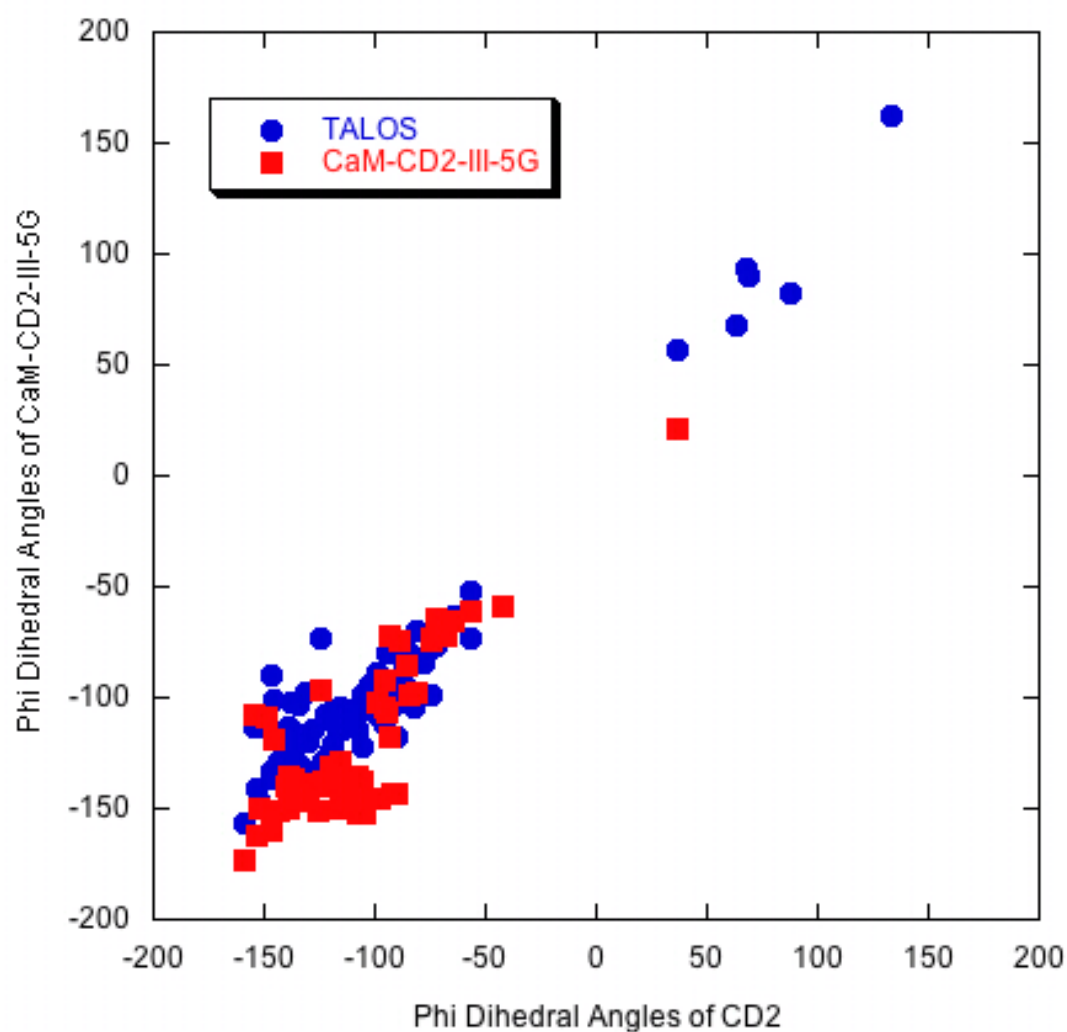
Residue Number	JHNHA coupling	Calculated Phi Angles
5	7.3935	-155.75
6	9.3586	-136.17
7	9.1333	-139.19
8	8.0523	-150.23
10	2.9066	-61.27
12	9.7514	-128.96
14	6.8095	-107.89
15	9.2871	-137.18
16	8.1303	-149.54
17	6.6642	-104.81
18	4.0887	-74.18
20	6.9918	21.079
21	8.0123	-150.58
22	9.0459	-140.25
24	9.4145	-135.34
26	7.7863	-152.52
27	7.2826	-156.64
28	5.8217	-92.57
29	6.5798	-162.08
30	6.8945	-159.67
31	8.6396	-144.7
32	9.0458	-140.25
33	9.0661	-140.25
34	8.6445	-144.65
37	7.9315	-151.29
38	4.1217	-74.52
40	5.1565	-172.76
41	7.9622	-151.02
42	8.7866	-143.17
43	8.1238	-149.6
45	6.7698	-106.98
47	6.4827	-101.62
50	7.1263	-157.87
57	5.9127	-93.68
58	7.4648	-155.18
62	6.5966	-103.56
63	8.1057	-149.76
65	6.1689	-96.99
71	3.1307	-63.932
72	9.2479	-137.71
73	8.1262	-149.57
74	7.8536	-151.95
75	6.2712	-98.408
77	8.3329	-147.69
79	3.9831	-73.103
80	8.7256	-143.81
81	7.8315	-152.14
82	9.2497	-137.68
83	3.919	-72.442
84	8.6013	-145.09
86	8.2993	-148
87	2.7519	-59.319
88	6.1868	-97.23
89	7.0562	-117.23
90	5.1522	-85.085
93	9.3928	-135.66
94	9.6852	-130.51
95	8.7204	-143.87
98	9.1805	-138.59
100	3.2261	-65.021
103	3.8715	-71.953
107	7.0651	-118.36
108	8.005	-150.65
110	8.4347	-146.72
111	8.5829	-145.27
112	8.7752	-143.29
113	8.0535	-150.22
114	7.7516	-152.82
115	6.1668	-96.96
116	7.8117	-152.31

Table 4.5 Dihedral Angles of CD2 and CaM-CD-III-5G from TALOS

#	CD2 pH 5.0			CD2 pH 7.4			CaM-CD2-III-5G		
	$\phi$	$\psi$		$\phi$	$\psi$		$\phi$	$\psi$	
1	9999	9999	None	9999	9999.00	None	9999	9999	None
2	-111.17	147.69	New	-77.69	133.28	New	-91.15	143.84	New
3	-113.08	136.81	New	-86.23	-10.92	New	-83.12	-21.74	New
4	-171.73	-154.14	New	138.77	-39.64	New	160.07	-148.82	New
5	-106.66	131.51	Good	-105.09	131.72	Good	-105.7	130.93	Good
6	-125.48	132.37	Good	-122.51	132.03	Good	-125.79	132.85	Good
7	-107.11	136.19	Good	-108.93	137.31	Good	-108.21	134.53	Good
8	-123.13	149.28	New	-122.65	145.43	New	-119.72	149.1	New
9	-107.53	136.37	New	-99.48	139.46	Good	-105.56	134.91	New
10	-52.72	132.84	New	-67.69	139.89	Good	-52.72	132.84	New
11	95.47	-13.95	New	93.53	-15.17	New	92.92	-13.22	New
12	-103.33	136.94	New	-100.34	139.74	New	-104.53	144.64	Good
13	-84.32	136.08	New	-170.89	-135.57	New	80.35	8.65	New
14	-111.74	145.97	Good	-101.44	137.86	Good	-113.16	149.05	Good
15	-111.08	126.2	Good	-100.61	120.26	Good	-111.87	126.37	Good
16	-96.54	115.23	Good	-89.95	125.13	Good	-91.48	119.84	Good
17	-110.61	124.89	Good	-113.15	134.48	New	-111.91	131.54	New
18	-100.19	130.65	Good	-76.44	126.72	New	-77.21	137.17	Good
19	-74.89	147.71	New	-56.54	145.14	New	-56.6	144.18	New
20	-96.7	142.93	New	-64.21	-30.48	New	57.14	32.45	New
21	-128.74	139.22	New	-97.84	126.87	New	-128.74	139.22	New
22	-113.33	131.72	Good	-106.21	142.47	Good	-116.45	132.11	Good
23	-89.59	125.88	Good	-87.79	140.99	Good	-80.27	131.63	Good
24	-114.54	158.87	Good	-108.22	164.98	Good	-112.92	162.33	Good
25	-61.64	-22.69	Good	-59.13	-28.97	Good	-61.64	-22.69	Good
26	-96.11	-9.29	New	-94.52	-12.41	New	-84.36	-14.53	New
27	-95.29	114.16	New	-86.57	122.00	New	-101.19	110.05	Good
28	-118.34	148.65	New	-108.38	143.43	New	-112.65	146.25	New
29	-140.34	148.19	Good	-145.31	151.81	Good	-140.74	148	Good
30	-133.04	129.09	Good	-126.04	125.22	Good	-133.45	135.9	Good
31	-113.99	133.53	Good	-111.3	134.47	Good	-119.63	133.26	Good
32	-118.78	128.9	Good	-116.4	131.06	Good	-120.63	134.29	Good
33	-131.11	142.3	Good	-131.25	139.21	Good	-134.8	139.17	Good
34	-109.7	120.18	Good	-111.17	119.31	Good	-102.8	121.92	Good
35	67.55	-119.71	New	67.99	-112.14	New	67.55	-119.71	New
36	-97.38	1.9	New	-98.12	-2.01	Good	-97.38	1.9	New
37	-106.93	131	Good	-100.06	128.21	Good	-97.56	129.73	Good
38	-85.84	132.36	New	-85.04	129.64	Good	-76.24	131.19	Good
39	-99.74	-33.54	New	-105.62	-27.51	New	-97.37	-29.11	New
40	-141.84	145.67	Good	-151.41	150.82	Good	-157.14	150.74	Good
41	-138.29	134.03	Good	-133.93	135.68	Good	-136.12	145.6	Good
42	-115.08	125.79	Good	-108.55	127.09	Good	-122.39	129.35	Good
43	-124.05	129.06	Good	-93.67	140.60	Good	-115.89	124.87	Good
44	54.41	43.91	New	-60.67	-37.96	New	-64.22	-23.7	New
45	60.03	24.91	New	-68.14	-24.53	New	-79.59	-17.43	New
46	-98.41	128.62	New	-105.87	122.30	New	-112.99	145.31	New
47	-91.91	128.07	Good	-84.67	146.71	New	-89.21	149.57	Good
48	-60.86	141.82	Good	-59.75	143.82	Good	-64.3	147.56	New
49	-111.78	120.41	Good	-114.17	123.92	Good	-102.76	136.22	New
50	-112.6	136.18	Good	-97.62	125.61	Good	-96.52	127.39	New
51	-71.16	-25.09	New	-74.55	-31.10	Good	-86.18	137.07	New
52	-102.24	6.77	New	-103.5	-2.39	Good	9999	9999	None
53							9999	9999	None
54							-176.31	-21.89	New
55							-159.31	-147.32	New
56							-68.14	-27.83	New
57							-77.45	-21.07	New
58							-70.43	-17.63	New
59							-6.94	-47.28	New
60							-87.21	-0.22	New

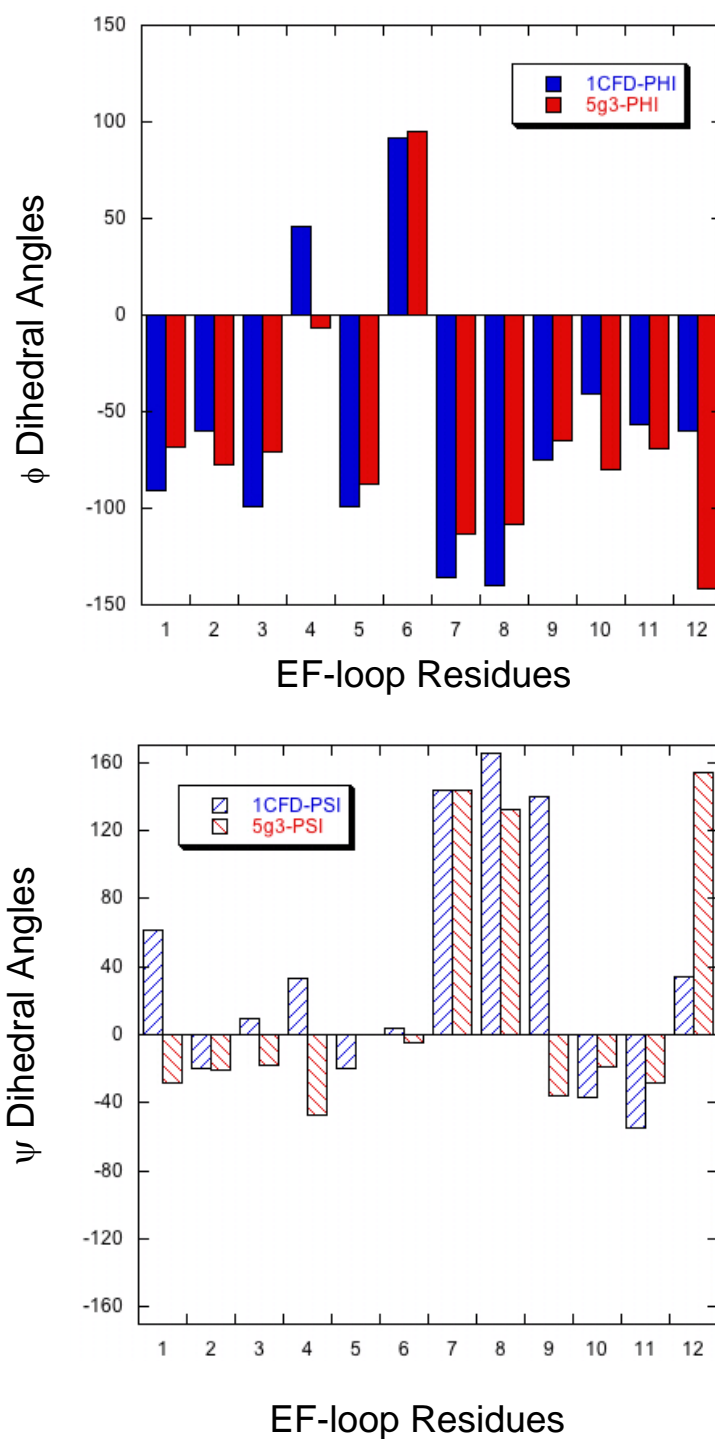
Table 4.5 Dihedral Angles of CD2 and CaM-CD-III-5G from TALOS

#	CD2 pH 5.0			CD2 pH 7.4			CaM-CD2-III-5G		
	$\phi$	$\psi$		$\phi$	$\psi$		$\phi$	$\psi$	
61							94.82	-5.09	New
62							-113.16	143.56	New
63							-108.48	132.68	New
64							-64.6	-36.02	New
65							-80.19	-18.52	Good
66							-69.08	-27.88	New
67							-141.27	154.19	New
68							-166.94	-134.01	New
69							9999	9999	None
70	81.13	8.81	New	86.99	3.86	New	92.67	-1.23	New
71	-85.75	140.52	New	-65.4	143.43	New	-76.93	135.58	New
72	-138.78	149.83	Good	-117.45	137.55	New	-129.12	150.3	Good
73	-145.78	144.97	Good	-139.5	147.31	Good	-146.77	143.82	Good
74	-119.87	131.24	Good	-126.29	129.08	Good	-108.98	122.87	Good
75	-80.1	153.94	New	-85.45	144.60	New	-80	161.44	New
76	-67.59	-19.91	Good	-65.2	-27.17	Good	-61.49	-26.21	Good
77	-94.69	1.08	Good	-85.77	0.42	Good	-94.69	1.08	Good
78	86	4.08	New	91.85	-5.69	Good	82.32	6.39	New
79	-95.51	129.35	New	-96.83	130.17	New	-98.8	151.66	New
80	-109.76	112.61	New	-98.06	118.52	Good	-117.6	111.12	New
81	-113.33	129.31	Good	-110.48	131.07	Good	-114.56	126.26	Good
82	-96.8	130.59	Good	-109.24	125.66	New	-98.72	128.3	Good
83	-71.01	-25.96	New	-61.69	-33.51	New	-66.43	-33.17	New
84	-89.61	-4.48	New	-68.99	-29.09	Good	-89.66	-5.84	New
85	-76.99	132.03	Good	-72.7	130.59	Good	-73.46	136.96	New
86	-103.54	169.37	Good	-116.95	167.59	Good	-102.16	171.79	Good
87	-58.24	-35.74	New	-59.79	-37.49	Good	-58.91	-35.57	New
88	-69.2	-29.64	Good	-66.36	-32.68	Good	-69.7	-29.23	Good
89	-85.39	-18.66	Good	-84.31	-26.20	Good	-78.14	-26.52	New
90	-88.32	-14.03	New	-96.63	153.85	New	-94.99	157.89	New
91	167.28	12.55	New	142.21	12.32	New	162.63	132.63	New
92	-103.91	126.67	Good	-108.81	131.06	Good	-103.91	126.67	Good
93	-105.71	132.98	Good	-106.12	131.21	Good	-106.67	130.71	Good
94	-121.77	131.02	Good	-123.88	135.43	Good	-124.34	133.36	Good
95	-116.96	128.59	Good	-111.14	122.69	Good	-114.02	129.21	Good
96	-114.33	128.1	Good	-112.99	127.21	Good	-114.96	132.07	Good
97	-120.9	133.11	Good	-112.82	135.51	Good	-113.93	133.55	Good
98	-119.27	131.33	Good	-117.34	127.31	Good	-117.12	128	Good
99	-86.73	160.34	Good	-92.09	152.19	Good	-84.36	158.45	Good
100	-61.47	-22.05	Good	-60.9	-24.77	New	-63.02	-22.35	Good
101	-88.26	-0.56	New	-84.48	-3.24	New	-86.87	-0.29	New
102	87.96	-1.12	New	87.69	-1.85	New	89.73	-3.76	New
103	-93.35	126.16	Good	-85.24	123.81	Good	-93.35	126.16	Good
104	-86	134.94	Good	-78.76	124.96	New	-77.6	131.67	Good
105	-99.64	-28.04	New	-102.16	-26.79	New	-93.99	127.23	New
106	-156.33	139.03	Good	-155.21	142.34	Good	-139.29	136.14	Good
107	-111.05	136.51	Good	-112.28	135.44	New	-101.29	134.93	Good
108	-129.95	138.71	Good	-127.65	151.27	Good	-137.38	144.26	Good
109	-116.54	138.13	Good	-108.81	137.33	Good	-112.52	137.62	Good
110	-124.84	138.49	Good	-123	133.84	Good	-130.98	139.4	Good
111	-106.49	123.53	Good	-99.73	120.83	Good	-108.82	124.8	Good
112	-96.31	124.6	Good	-102.02	125.03	Good	-104.09	125.38	Good
113	-115.67	139.13	Good	-115.27	127.96	Good	-113.51	135.85	Good
114	-106.4	121.78	Good	-105	121.48	Good	-105.05	121.98	Good
115	-91.85	131.51	New	-87.25	126.43	New	-72.79	136.62	New
116	9999	9999	None	9999	9999.00	None	9999	9999	None

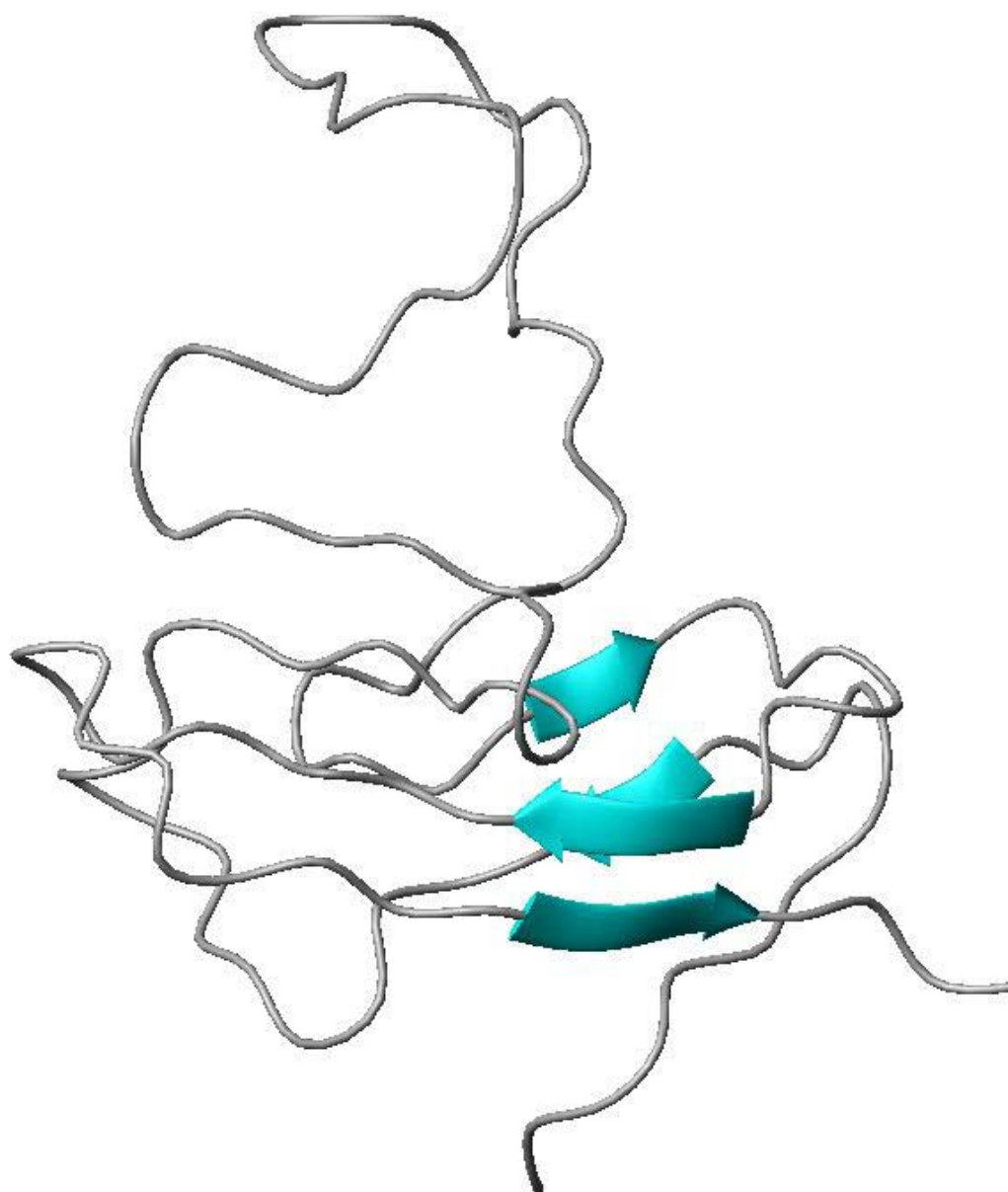


**Figure 4.19 Dihedral angle comparison between the HNHA and TALOS results to CD2:** The  $\phi$  dihedral angles of CaM-CD2-III-5G (calculated from TALOS and  $J_{\text{HNHA}}$  couplings) are plotted as function of the same residues in wild type CD2.

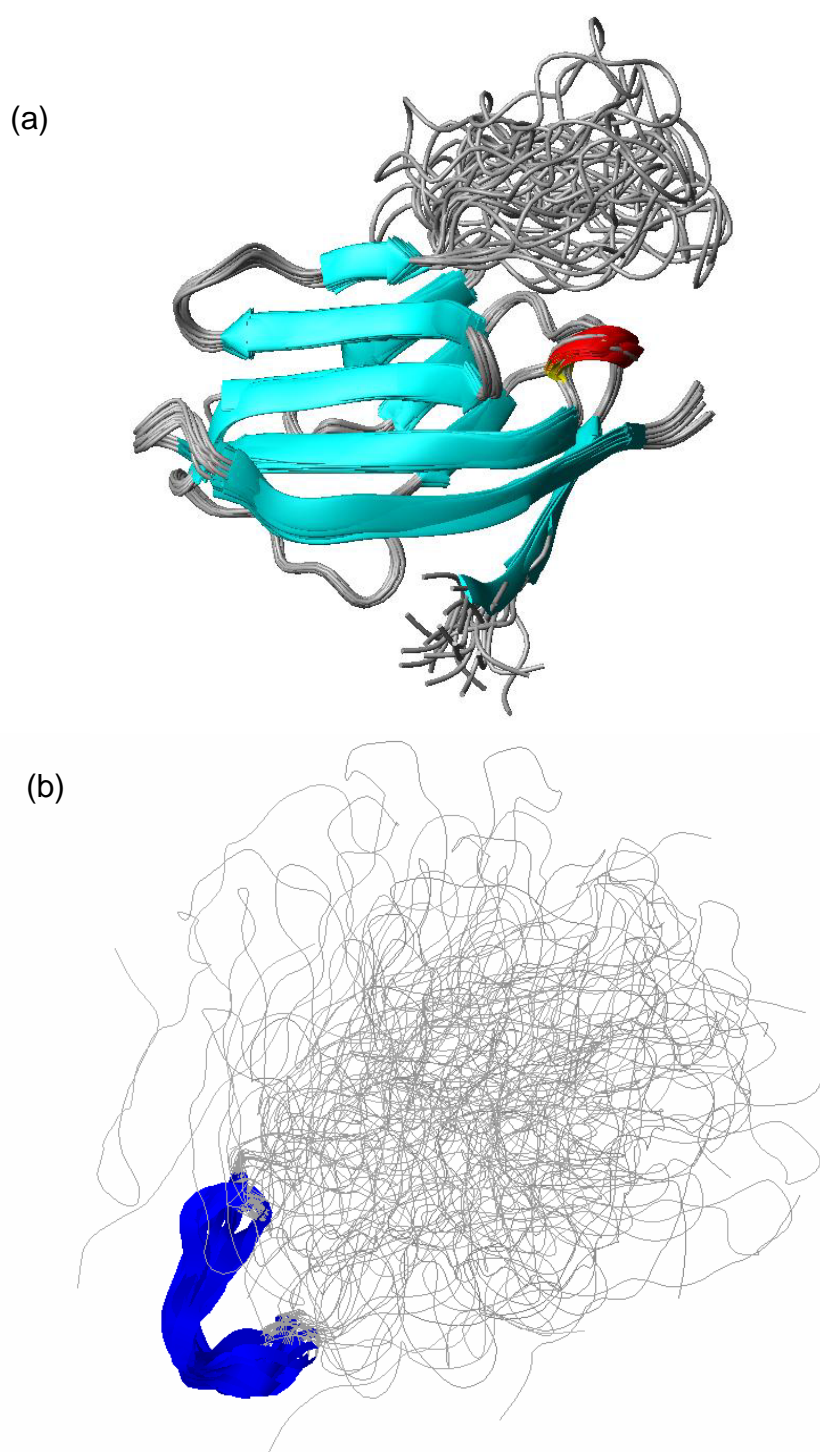




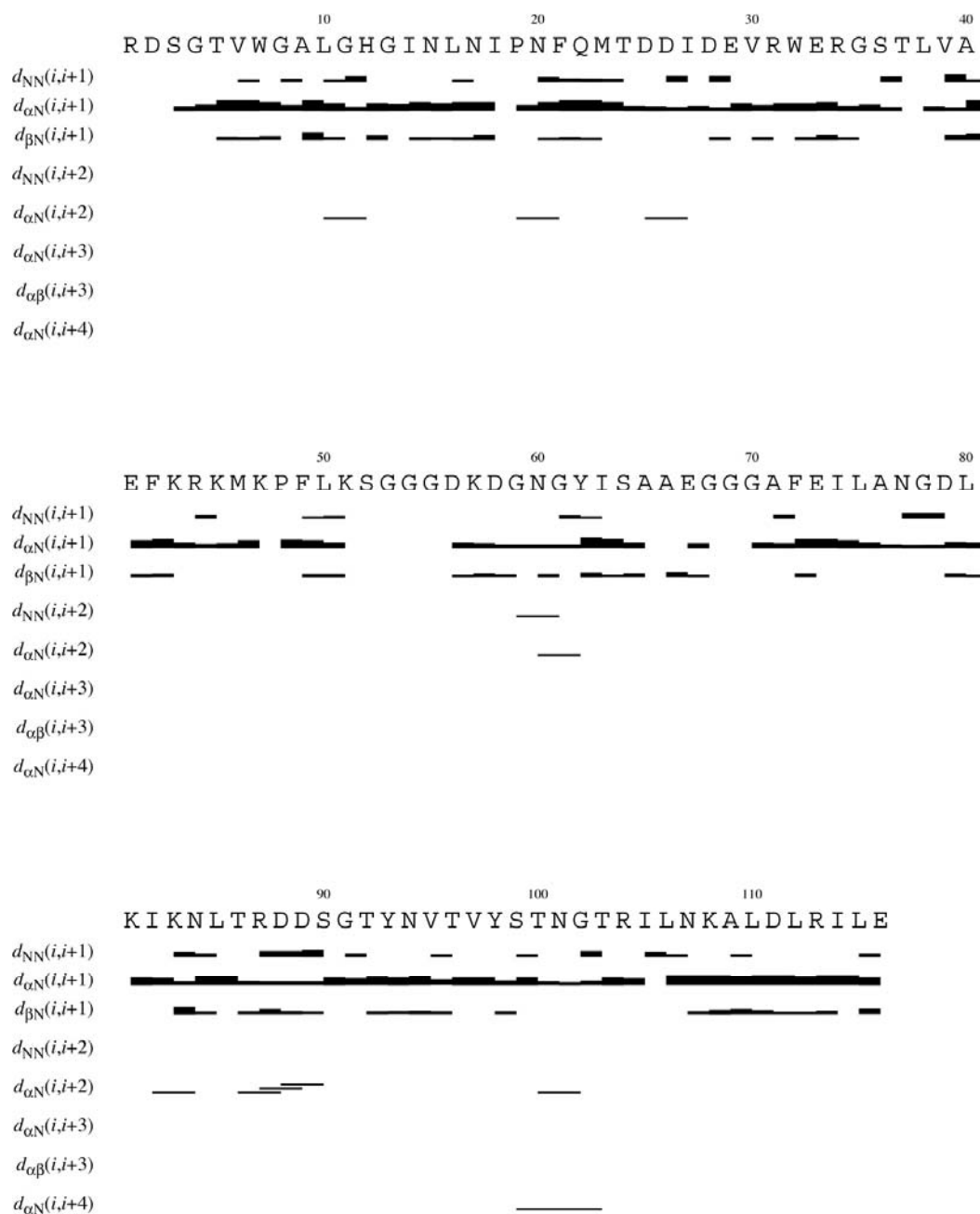
**Figure 4.20 Dihedral angle comparisons on the EF-loop III of CaM-CD2-III-5G to CaM:** The dihedral angles of CaM-CD2-III-5G were calculated using TALOS. The dihedral angles of calmodulin were obtained from pdb 1CFD.pdb using procheck.



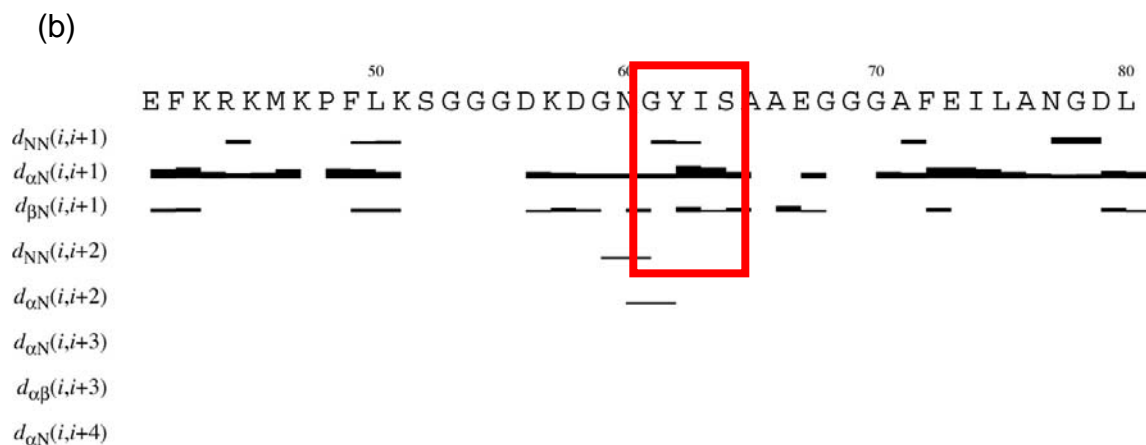
**Figure 4.21** Solution structure of CaM-CD2-III-5G without dihedral angle restraints: This is the cycle 6 structure of CaM-CD2-III-5G using CYANA 1.1.



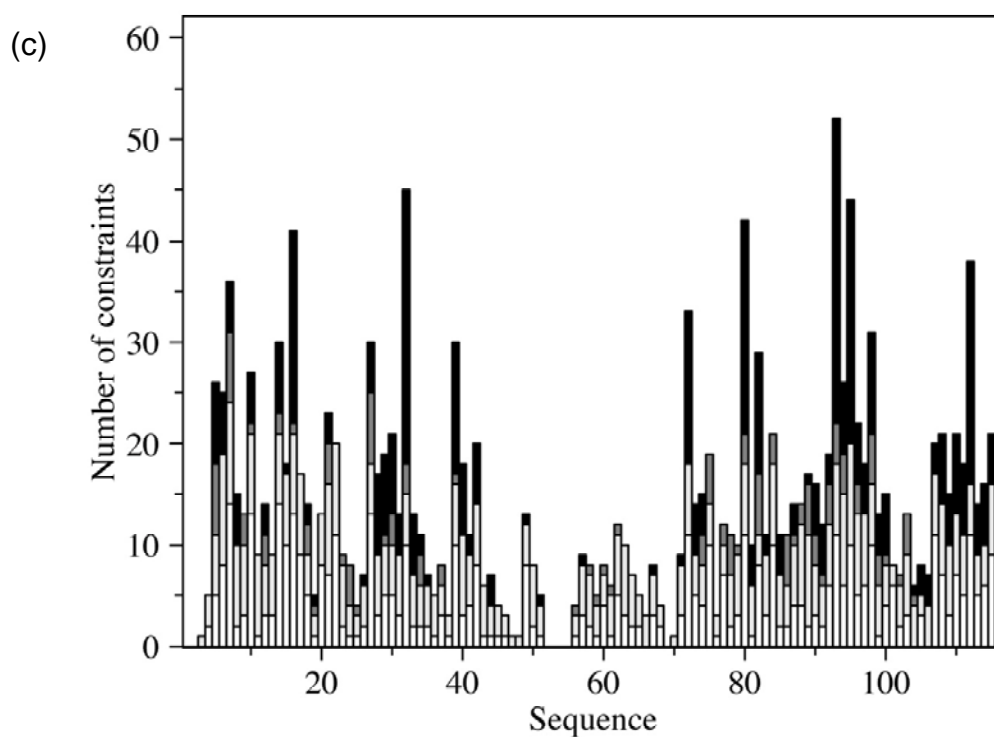
**Figure 4.22 Final solution structure of CaM-CD2-III-5G:** This is the cycle 23 structure of CaM-CD2-III-5G using CYANA 2.1. The twenty lowest energies structure are overlaid using (a) residues G4 to F42 and (b) the inserted EF-loop residues.



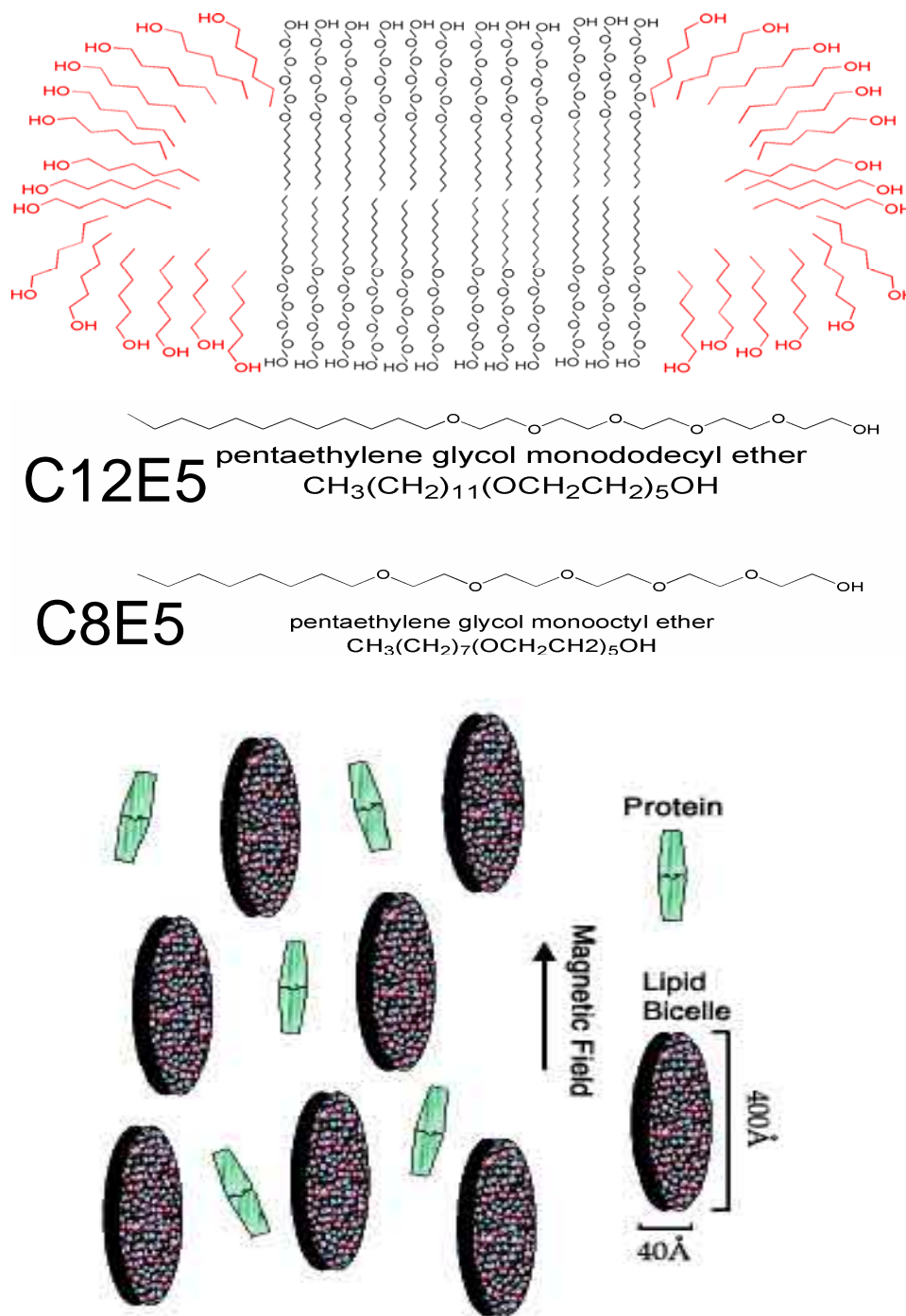
**Figure 4.23 NOE pattern of the CaM-CD2-III-5G:** (a) NOE patterns of CaM-CD2-III-5G. (b) Positions 7 to 9 of the EF-loops usually forms a short  $\beta$ -strand. The  $\beta$ -strand secondary structure usually have strong  $H_{\alpha}$  to  $HN_{(i+1)}$ . Position 7 to 9 of the inserted EF-loop III in CaM-CD2-III-5G has NOE patterns of small  $\beta$ -strand character. (c) Number of NOE restraints per residue.



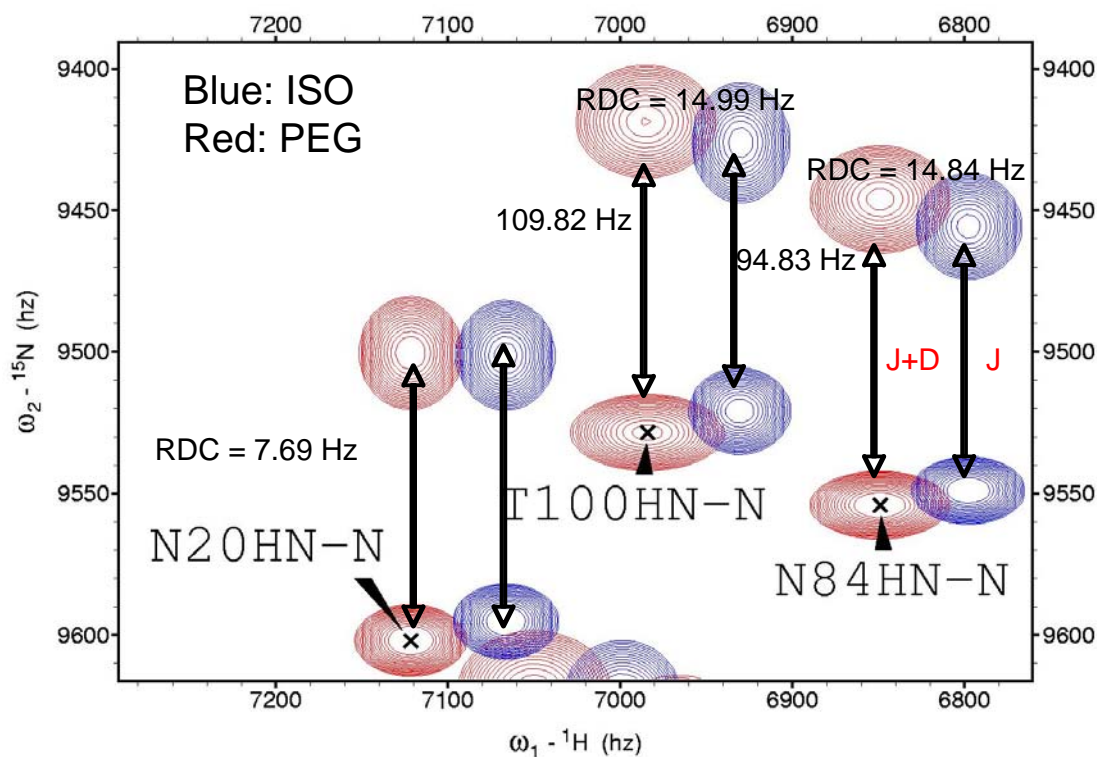
D-K-D-G-N-G-Y-I-S-A-A-E



**Figure 4.23 NOE pattern of the CaM-CD2-III-5G:** (a) NOE patterns of CaM-CD2-III-5G. (b) Positions 7 to 9 of the EF-loops usually forms a short  $\beta$ -strand. The  $\beta$ -strand secondary structure usually have strong  $H\alpha$  to  $HN_{(i+1)}$ . Position 7 to 9 of the inserted EF-loop III in CaM-CD2-III-5G has NOE patterns of small  $\beta$ -strand character. (c) Number of NOE restraints per residue.



**Figure 4.24 PEG medium for the RDC studies:** The residual dipolar couplings for CaM-CD2-III-5G were performed in 3.3 % PEG-bicelle (C12E5) medium (These figures are courtesy of Dr. Anita Kishore).

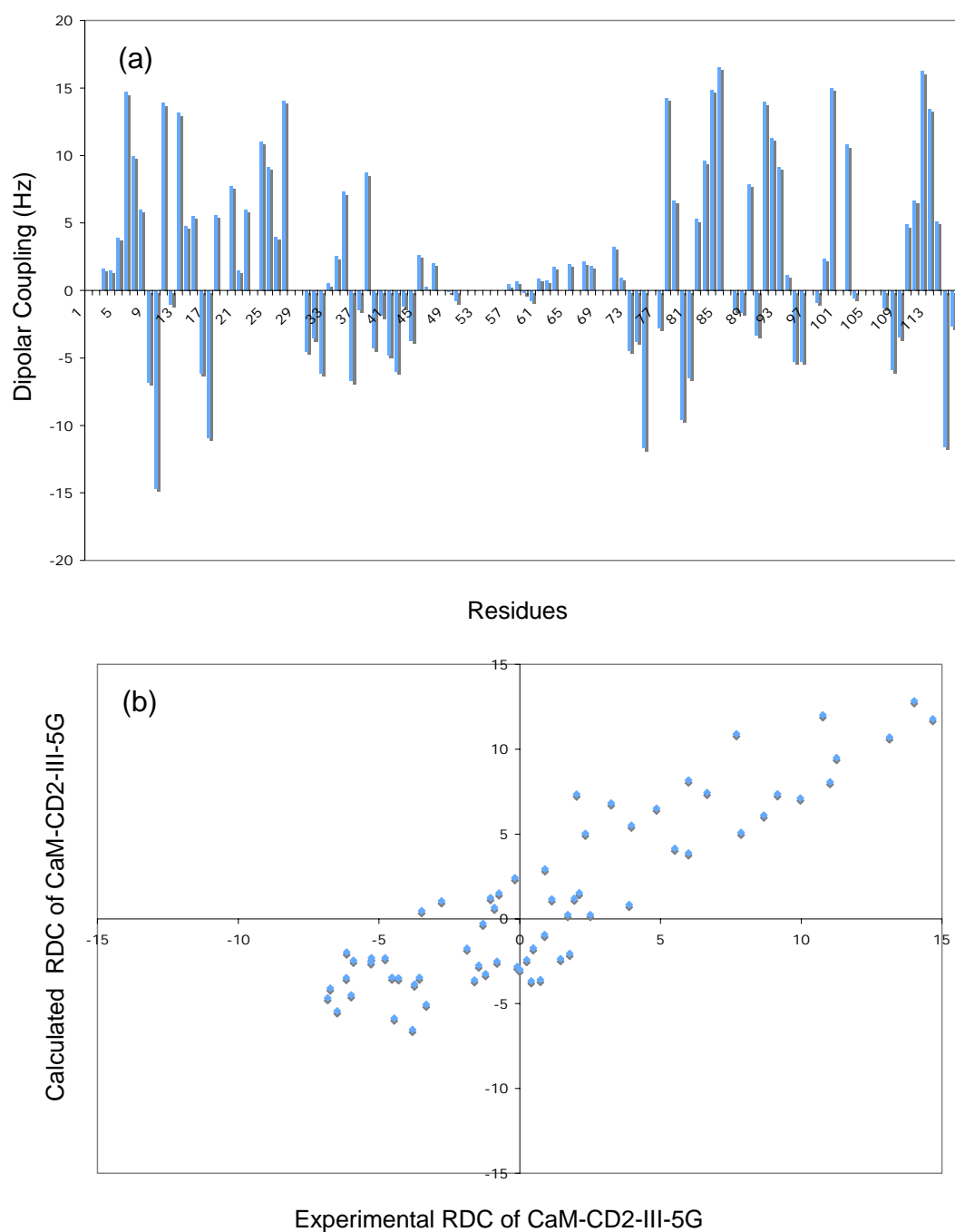


**Figure 4.25 Example of an IPAPHSQC spectrum.** The IPAPHSQC spectrum of CaM-CD2-III-5G collected in isotropic conditions is shown in blue. The IPAPHSQC spectrum of CaM-CD2-III-5G collected in the presence of 3.3 PEG-bicelle is shown in red. Both samples were prepared in 20 mM PIPES and 20 mM KCl at pH 6.8. Both IPAPHSQC spectra were collected using 800 MHz NMR at 25 °C.

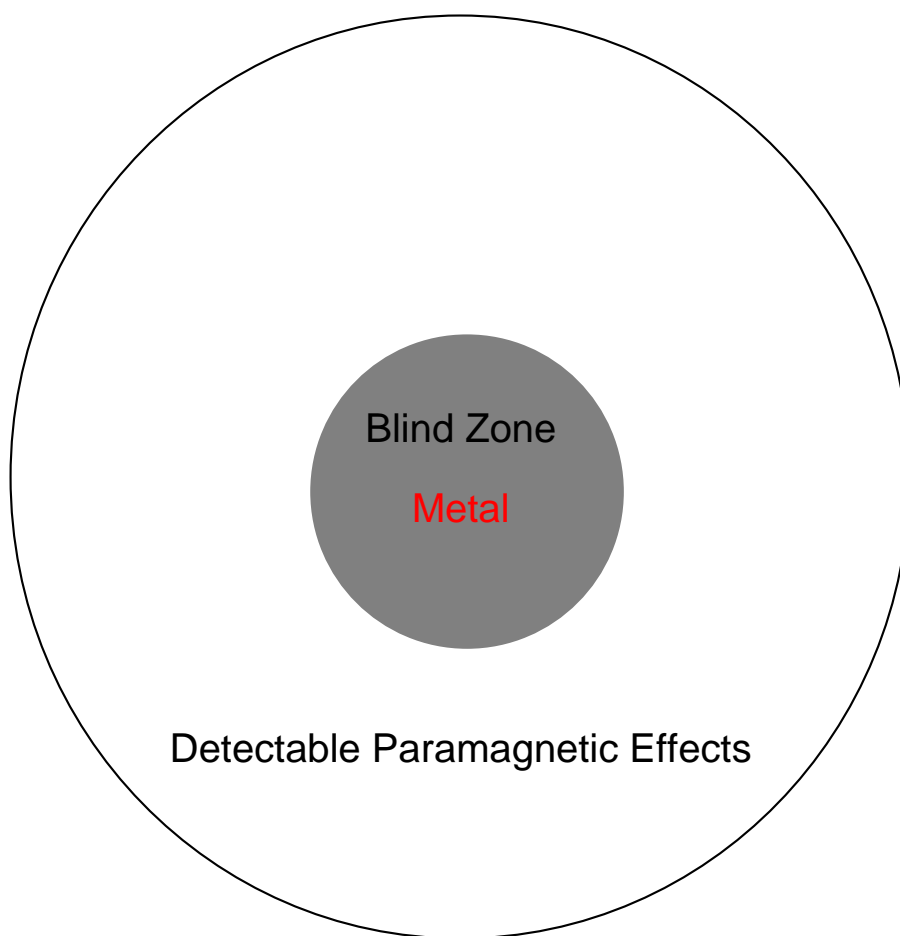
Table 4.9 Residual dipolar couplings of CaM-CD2-III-5G

Residue Number	Experimental RDC (Hz)	Back-calculated RDC (Hz)
3	1.6201	11.6969
4	1.46	12.1174
5	3.8896	3.11242
6	14.67	12.1158
7	9.9697	7.83711
8	6	2.67728
9	-6.8203	-3.13492
10	-14.66	-9.68585
11	13.86	7.61131
12	-1.04	-0.8915
13	13.13	11.5203
14	4.79	7.16611
15	5.5098	3.27061
16	-6.1494	-5.43505
17	-10.931	-8.31587
18	5.5908	-9.4407
20	7.6992	8.96165
21	1.4492	3.76274
22	6	2.5591
24	11.02	11.0759
25	9.1602	6.13086
26	3.9707	4.48755
27	14.01	11.9629
30	-4.54	-5.64655
31	-3.5605	-5.12572
32	-6.1592	-8.27246
33	0.48047	-3.18994
34	2.5107	-2.36099
35	7.29	4.51721
36	-6.7295	-5.51737
37	-1.46	5.6294
38	8.6787	9.80154
39	-4.3008	-1.86786
40	-1.8701	-0.874285
41	-4.7803	-4.45035
42	-5.9902	-5.22168
43	-1.2109	-3.4877
44	-3.7402	-6.66066
45	2.5996	-10.2434
46	0.25	11.0704
47	2.0195	12.4047
49	-0.09082	-0.287335
50	-0.81055	6.32295
57	0.41016	2.83548
58	0.63965	-0.209442
59	-0.16992	-2.32232
60	-0.73047	-4.47368
61	0.88965	11.6277
62	0.73926	1.61634
63	1.71	11.37
65	1.9404	0.0971286
67	2.1104	4.75284
68	1.7793	4.21326
71	3.25	6.31271
72	0.90039	5.86679
73	-4.459	-4.67695
74	-3.8096	-0.543479
75	-11.671	-9.95907
77	-2.7705	-2.14876
78	14.26	-3.01287
79	6.6396	6.79472
80	-9.5596	-5.38971
81	-6.4805	-9.69799
82	5.2705	-2.97711
83	9.5693	2.00801
84	14.84	6.50135
85	16.54	11.909
86	0.00097656	2.13921
87	-1.4502	-1.95458



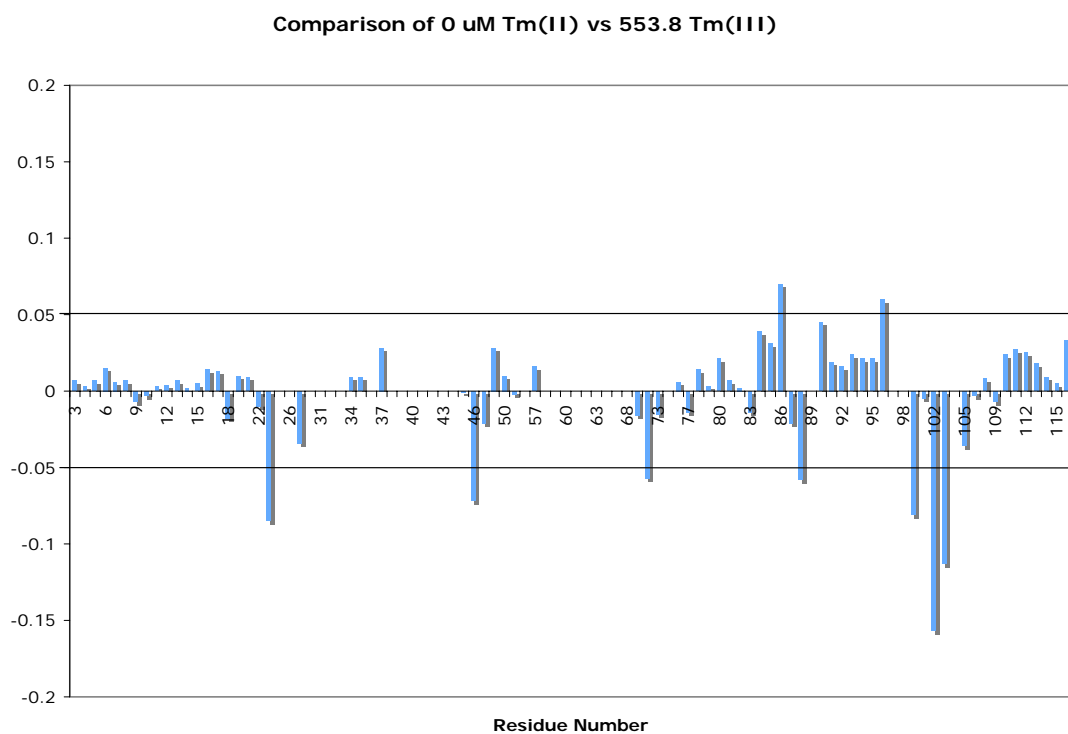


**Figure 4.26 Comparison of experiment RDC and calculated RDC:** (a) The experimental RDC of CaM-CD2-III-5G were collected in 3.3 % PEG-bicelle medium. (b) The order parameters were obtained from REDCAT calculation using the solution structure of CaM-CD2-III-5G and dipolar coupling. Based on the calculated order parameters, a new set of RDC was back calculated based on the structure of CaM-CD2-III-5G. The residual dipolar couplings that yields high errors during order parameter calculation were removed from plot (b).

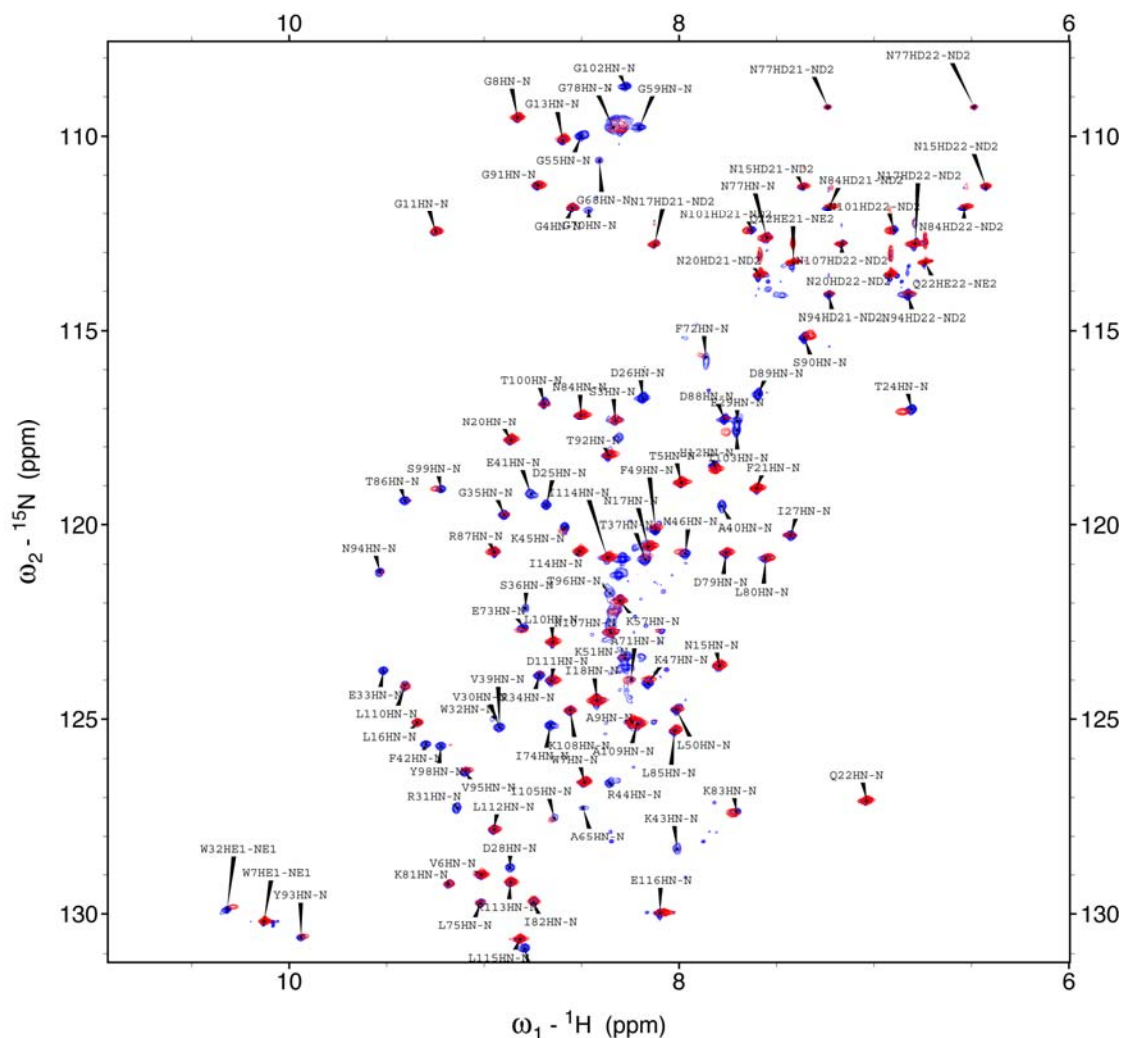


**Figure 4.27 The detectable paramagnetic effect:** The use of paramagnetic metal ions will cause the signals of nuclei that are in close proximity to the metal ion to be broaden beyond detection and is usually refer to as blind zone. The ranges (effected distance) of the blind zone and detectable paramagnetic effects is depending on the orbital properties of the metal ions.

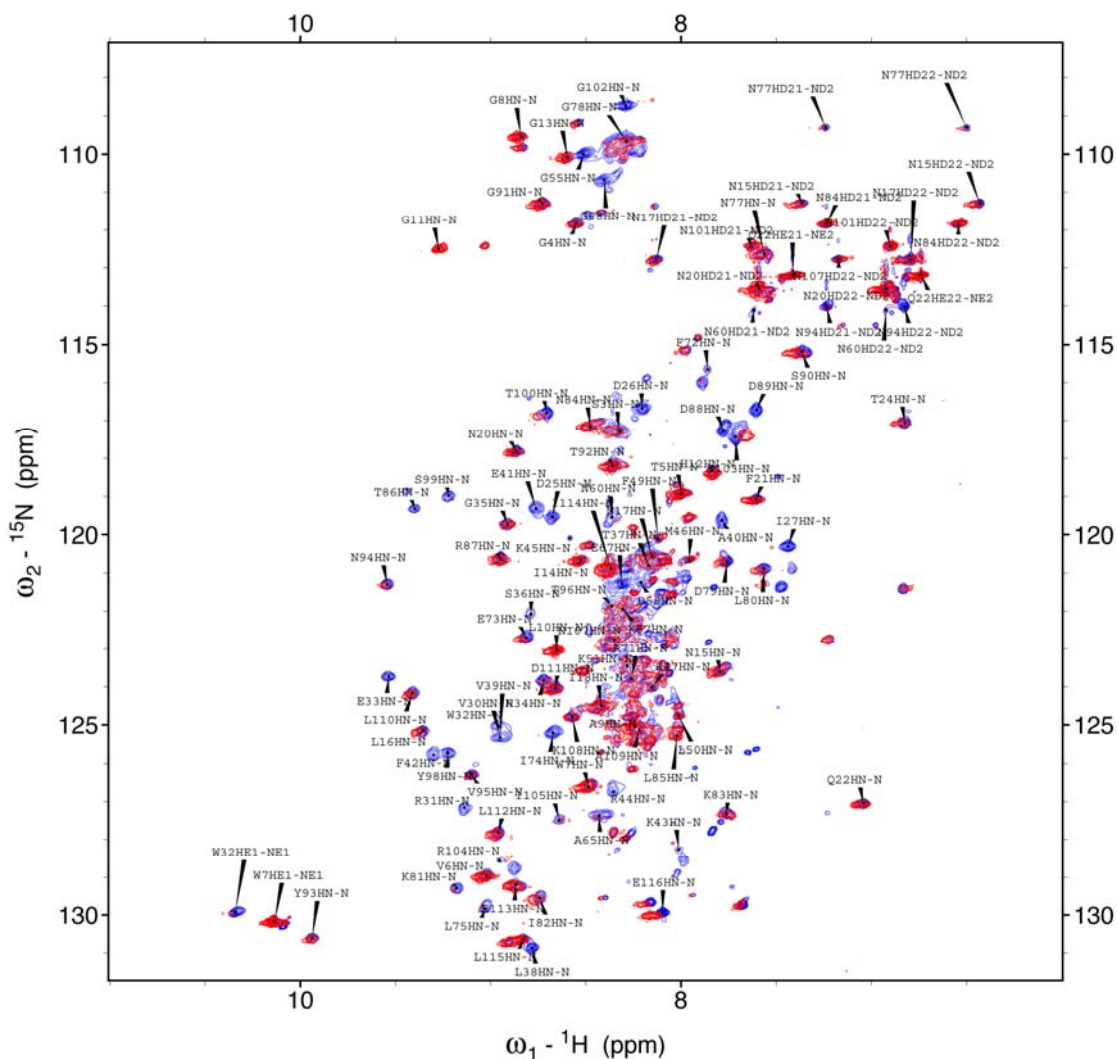
**Figure 4.28 Tm(III) titration on CaM-CD2-III-5G:** The  $^{15}\text{N}$  HSQC spectra with 0, 235, and 500  $\mu\text{M}$  of Tm(III) are shown as blue, red and green color, respectively. The  $^{15}\text{N}$  HSQC spectra are stacked onto the first point of the titration. These spectra were collected using 600 MHz NMR at 25  $^{\circ}\text{C}$ . The protein sample was prepared in 20 mM PIPES and 20 mM KCl at pH 6.8.



**Figure 4.29 HN chemical shifts change in Tm(III) titrations:**  
 Comparison of HN chemical shifts of CaM-CD2-III-5G in the presence 553.8  $\mu$ M of Tm(III) and absence of Tm(III).




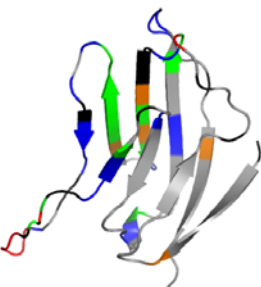
**Figure 4.30**  $^{15}\text{N}$  HSQC spectra of CaM-CD2-III-5G in the presence of La(III) and Tm(III): The Tm(III) HSQC spectrum (blue) is stacked on top of the La(III) spectrum (red) of CaM-CD2-III-5G. These spectrum were collected using Varian 800 MHz NMR at 25 C. Both sample were prepared in 20 mM PIPES 20 mM KCl at pH 6.8.



**Figure 4.31  $^{15}\text{N}$  HSQC spectra of CaM-CD2-III-5G in the presence of La(III) and Dy(III):** The Dy(III) HSQC spectrum (red) is stacked on top of the La(III) spectrum (blue) of CaM-CD2-III-5G. These spectrum were collected using Varian 800 MHz NMR at 25 C. Both sample were prepared in 20 mM PIPES 20 mM KCl at pH 6.8.

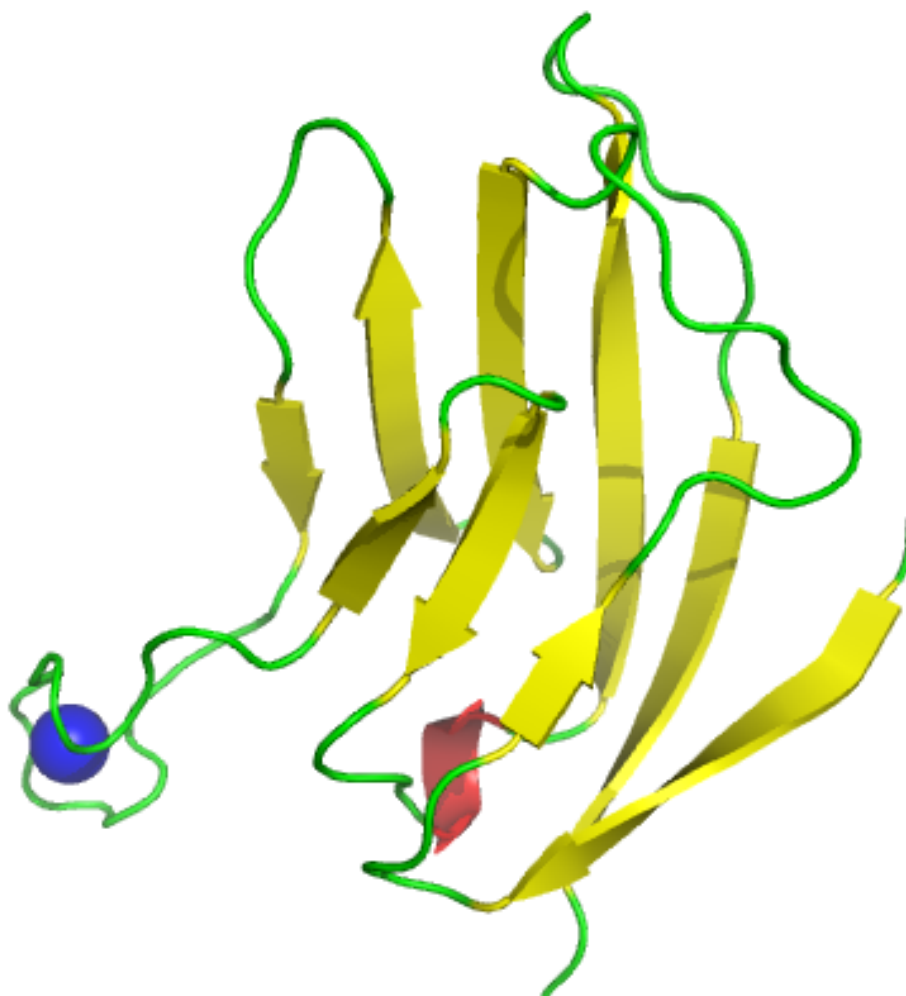
Table 4.10 Summary of the PCS and RDC from Dy(III) and Tm(III)

HN800		HN600		RDC800		RDC-Fit		No-fit			
Res	Tm-La	Dy-La	Res	Tm-La	Dy-La	Dy-La	Dy-La				
1			1								
2			2								
3	0.004	-0.016	3	-0.49	-1.24	2.68					
4	0.003	-0.003	4	0.89	-1.01	-1.16					
5	0.006	-0.010	5	0.16	-0.20	-0.89					
6	0.007	-0.018	6	-0.16	-0.26	-0.29					
7	0.007	-0.016	7	0.17	-0.05	-0.76					
8	0.007	-0.015	8	0.16	0.10	-0.12					
9	0.007	-0.011	9		0.43	0.45					
10	0.005	-0.008	10	0.08	0.01	-0.44					
11	0.008	0.008	11	-0.08	-0.33	-0.39					
12	0.009	0.001	12	-0.49	0.36	-0.48					
13	0.009	-0.013	13	0.16	2.90	0.99					
14	0.008	-0.011	14	0.25	-0.40	1.48					
15	0.006	-0.022	15	0.32	1.52	3.85					
16	0.004	-0.015	16	0.08	-1.33	-3.75					
17	0.014	-0.010	17	-0.08							
18	-0.003	-0.013	18	-0.08	-0.44	1.66					
19			19								
20	0.007	-0.013	20	-0.24	0.66	1.77					
21	0.006	-0.015	21	-0.08	-0.21	-0.19					
22	0.002	-0.012	22	-0.09	0.14	1.17					
23			23								
24	-0.043	-0.014	24	0.49	0.32	2.49					
25			25								
26			26	1.30							
27	-0.002		27								
28			28								
29			29								
30			30								
31			31								
32			32								
33		-0.032	33	0.40							
34	0.003		34	-0.97							
35	0.002	0.020	35	1.39	-0.15	4.08					
36			36								
37	0.009		37	-2.10							
38			38								
39			39								
40			40								
41			41								
42			42								
43			43								
44			44								
45	-0.012		45	0.41							
46	-0.027	0.015	46	0.98							
47	0.006		47	0.08							
48			48								



49	0.006	0.016	49	0.17				
50	0.010	0.001	50	-0.40				
51	-0.007		51					
52			52					
53			53					
54			54					
55			55					
56			56					
57	0.004		57	1.46				
58			58					
59			59					
60			60					
61			61					
62			62					
63			63					
64			64					
65			65					
66			66					
67			67					
68			68					
69			69					
70			70					
71	-0.014		71					
72	-0.023		72					
73	-0.006	-0.021	73	-0.57	2.21	-3.35		
74			74					
75	-0.001		75	0.58				
76			76					
77	-0.002	-0.022	77		-3.37	-1.06		
78	0.042	-0.014	78	-0.48	-0.94	0.07		
79	0.010	-0.020	79	0.25	0.66	6.54		
80	0.017	-0.036	80	-0.57	-12.89	-14.99		
81	0.003	-0.019	81	0.25				
82	0.003	-0.020	82	0.00	3.51	6.84		
83	-0.014	-0.005	83	-0.07	-0.79	-0.94		
84	0.011	-0.007	84	0.16	1.18	2.48		
85	0.010	-0.004	85	-0.16	-2.23	0.04		
86	0.018		86	0.65				
87	-0.004	-0.008	87	0.00	-0.13	0.78		
88	0.006		88					
89	0.043		89					
90	0.027	-0.032	90	-0.72	-0.70	-1.70		
91	0.012	-0.015	91	0.08	-0.27	1.53		
92	0.011	-0.006	92	0.25	0.19	-0.69		
93	0.011	-0.002	93	0.89	-0.63	-2.04		
94	0.006	-0.016	94	0.72				
95	0.006	0.005	95	0.41	0.74	2.21		
96			96	-2.19	-12.89	-11.54		
97			97					
98			98	1.13				
99	-0.023		99	0.24				
100	0.001		100	-1.45				
101			101					
102			102					
103			103		0.20	1.77		
104			104					
105	-0.010		105	1.53				
106			106					
107	0.002		107	-2.43				
108	-0.001		108	-0.25				
109	0.000	-0.007	109					
110	0.004	-0.013	110	-1.12	-1.03	-2.76		
111	0.015	-0.018	111	-0.24	-0.26	-0.27		
112	0.009	-0.017	112	-0.01	0.65	0.54		
113	0.006	-0.020	113	-0.33	0.47	-1.64		
114	0.011	-0.017	114	-0.24	0.19	0.31		
115	0.004	-0.017	115	-0.23	1.02	0.24		
116	0.023	-0.057	116	0.25	1.55	0.46		





**Figure 4.32 Simulated structure of metal bound CaM-CD2-III-5G:**

This structure was generated using the metal to oxygen distance (Ca(II) to Ca(II) binding ligands at positions 1, 3, 5, 7, and 12 of the EF-loop) obtained from crystal structure of CaM (3CLN.pdb). The structure was calculated using CYANA.



Table 4.11 Putative Distances between the Paramagnetic Metal Ion and Affected Residues

Residue	Mean	Lowest	Range
25	38.91	40.14	29.95(16) - 42.02(7)
26	38.44	39.70	30.07(16) - 41.66(9)
27	36.06	37.39	28.07(16) - 39.33(9)
28	34.04	35.53	26.65(16) - 37.69(7)
29	32.01	33.39	25.17(16) - 35.65(7)
30	27.90	29.07	22.01(16) - 31.60(7)
31	27.20	27.75	23.62(16) - 30.87(9)
32	22.94	23.35	20.89(16) - 26.60(9)
33	22.16	23.74	22.34(6) - 27.70(9)
34	21.45	20.41	19.63(6) - 24.90(9)
37	21.40	20.37	19.03(6) - 26.08(16)
38	18.71	18.91	16.84(15) - 22.51(9)
39	19.70	19.86	17.63(6) - 23.40(9)
40	19.89	20.35	18.05(6) - 23.57(9)
41	20.32	21.77	16.67(16) - 24.13(7)
42	24.51	25.89	19.52(16) - 28.28(7)
43	25.74	27.97	18.49(16) - 30.08(7)
44	30.05	32.18	22.16(16) - 34.35(7)
45	30.75	33.29	22.59(16) - 35.76(7)
46	28.63	30.97	19.16(16) - 33.87(7)
47	24.49	27.30	17.20(16) - 29.99(7)
49	20.51	22.33	15.07(16) - 26.12(7)
50	16.12	18.27	13.59(16) - 20.94(7)
51	14.36	16.14	14.63(15) - 19.20(9)
71	12.32	13.61	11.06(16) - 14.85(9)
72	14.36	15.19	12.95(16) - 17.17(9)
73	18.77	19.15	15.05(16) - 21.36(9)
74	21.26	22.64	14.28(16) - 24.29(7)
75	24.91	25.70	18.81(16) - 27.73(9)
85	16.19	13.77	12.96(10) - 23.68(19)
86	16.21	13.47	13.11(10) - 23.76(19)
87	15.07	11.46	11.46(1) - 25.16(19)
88	16.03	11.96	11.96(1) - 24.60(16)
89	17.19	13.73	13.73(1) - 24.62(19)
90	19.57	16.05	16.05(1) - 26.45(19)
98	32.29	32.95	26.94(16) - 36.01(9)
99	36.40	37.26	30.73(16) - 40.20(9)
100	38.10	39.58	30.55(16) - 41.86(7)
102	37.22	38.51	30.55(16) - 41.16(7)
103	37.81	38.74	32.02(16) - 41.58(7)
105	35.59	36.19	31.61(16) - 39.45(9)
107	35.87	35.59	33.76(6) - 39.59(9)
108	31.96	31.36	30.01(6) - 35.55(9)

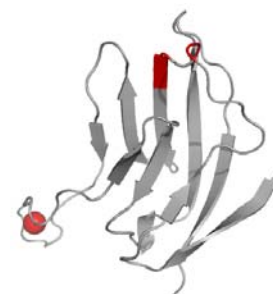
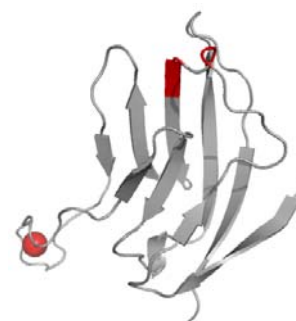


Table 4.12 Putative Distances between Affected Residues to D26 and D28

HN of Residue	To D26HN of the Avg Str	To D28HN of Avg Str
V30	11.76	6.26
R31	13.71	8.83
W32	17.51	12.78
E33	19.41	15.11
R34	23.77	19.23
T37	26.62	21.72
L38	25.84	20.51
V39	22.00	17.15
A40	20.48	15.74
E41	18.81	13.82
F42	14.91	9.64
K43	14.53	9.31
R44	11.44	6.01
K45	13.33	8.16
M46	15.44	9.97
K47	18.94	13.51
F49	18.93	13.84
L50	23.21	18.22
K51	23.74	19.00
A71	26.98	23.51
F72	24.69	21.17
E73	21.00	18.22
I74	17.93	15.24
L75	15.29	13.86
L85	29.53	27.15
T86	29.56	26.54
R87	33.84	30.63
D88	33.36	29.83
D89	31.02	27.61
S90	30.48	27.22
Y98	9.41	4.22
S99	9.09	5.07
T100	6.12	4.19
G102	10.18	5.79
T103	11.12	7.00
I105	11.42	7.97
N107	15.83	13.56
N108	16.18	13.59



## Chapter 5.0 Dynamics Studies on the C-terminal Domain of Calmodulin

### Using the Grafting Approach

The dynamic properties of proteins have been shown to play important roles in their biological functions, ligand binding, and Ca(II) dependent conformational changes. The molecular motion and dynamic properties of a protein can be described by the intrinsic hydrogen exchange (HX) rate or characterized by the  $S^2$  generalized order parameters. The HX rate determines how fast the exchangeable proton on the amide can be replaced with deuterium. Once exchanged, the substituted resonance will no longer be detected in classical NMR experiments (all of the homonuclear and some  $^{15}\text{N}$  edited experiments that transfer the signal back to  $^1\text{H}$  channel). The exchangeable protons in proteins include backbone NH protons and sidechain amides of Trp, Asn, Gln, Arg, Lys, and His (Figure 5.1). The exchangeable protons on the surface of proteins can be easily substituted, unlike those buried inside the hydrophobic core. The HX rate for an exchangeable proton is dependent on solvent exposure or accessibility of the proton, the hydrogen bonding, and the tertiary environments. This method has been applied to determine the protein folding pathway, packing of a protein with and without mutations, and protein-protein and protein-ligand interactions (161-166).

The  $S^2$  generalized order parameters are used to characterize the local and global dynamics on the backbone of proteins. The  $S^2$  value for an ordered

structure, such as residues included within an  $\alpha$ -helix or  $\beta$ -sheet, is generally greater than 0.85. The  $S^2$  values for flexible regions of proteins, such as loops between different secondary structure elements, are normally less than 0.85. The  $S^2$  order parameters cannot be directly measured by NMR. It is simulated using the  $^{15}\text{N}$  or  $^{13}\text{C}$  relaxation parameters including the longitude and transverse relaxation times and the single bond NOE between the H and N of the backbone. The ModelFree method is most popularly applied for the simulation (125).

As discussed in chapter 3, using the grafting approach, we have determined the site-specific calcium binding affinities of each individual EF-hand loop for CaM. Their relative binding affinities follow the order of I > III > II > IV. This order is very different from the proposed order of the acid-pair hypothesis (167). According to the acid-pair hypothesis, site IV has the strongest calcium binding affinity. There are two possible explanations for the different Ca(II) binding affinities between EF-loop III and IV. It is possible that the structural difference between the apo and calcium loaded form is smaller for EF-loop III. Thus, less energy is required for the formation of the calcium binding geometry, which enables the EF-loop III to have stronger calcium binding affinity. Alternatively, it is also possible that different dynamic ensembles of the apo forms might be responsible for their different Ca(II) binding affinities. To understand the origin of the differential Ca(II) binding affinities of each EF-hand site, in chapter 4, detailed analyses comparing not only the structures of the calcium free and calcium loaded forms of the EF-loops in the intact calmodulin,

but also the RMSD deviation are shown. In this chapter, we will describe our studies of the dynamic properties of CaM-CD2-III-5G and CaM-CD2-IV-5G. We first examine the intrinsic amide HX for the engineered calcium binding proteins. We then determine the relaxation properties for the engineered calcium binding proteins. Finally, we compare the dynamic studies of the engineered calcium binding proteins to other EF-hand proteins. The HX and  $^{15}\text{N}$  relaxation experiments were applied to understand the factors that affect the metal binding properties of the EF-loop and the molecular motions of the host protein and the inserted EF-loop.

## **5.1 Hydrogen Exchange Studies on the CD2 Variants**

### **5.1.1 1D HX Experiments on CD2, CaM-CD2-III-5G, and CaM-CD2-IV-5G**

Sample preparation and the detailed process of the HX experiment are described in chapter 2. The HX experiment for CD2 is shown in Figure 5.2. The example shows the amide region of the HX spectra of CD2 collected at 0.1, 5.5, and 20.6 hours. Some of the solvent exposed regions, such as the protein surface, were substituted to deuterium before the first 1D spectrum was completed. For example, the HE1 proton of W7 that is on the protein surface was not observed at the first spectra. On the other hand, the HE1 proton of W32 that is buried inside the hydrophobic core was still observable at the end of the HX experiment. The intensity of the HE1 resonance of W32 at 20.6 hours was

95% of that at 0.1 hours, which indicates that no deuterium exchange process occurs inside the hydrophobic core.

The amide region of the 1D spectrum shows the backbone amide protons, sidechain amide protons, and non-exchangeable protons of the aromatic ring. The majority of the resonances in the amide region overlap one another. Since it is very difficult to measure the intrinsic HX rate for each exchangeable proton, by integrating the bulk exchange rates of the amide regions from 6.8 to 10.1 ppm were measured. The data were fit as two exponential decay components. Therefore, two rates, one fast and one slow, were obtained (Table 5.1). Table 5.1 summarizes the HX data for the CD2 variants in the absence of calcium (1 mM EGTA) and presence of 1 mM Ca(II). In the presence of 1 mM Ca(II), the overall HX rate for the amide region of wild type CD2 is the slowest among the three proteins, followed by CaM-CD2-III-5G (Figure 5.3). The HX rate for CaM-CD2-IV-5G is the fastest. In the presence of 500  $\mu$ M EDTA, the exchange rate of CD2 is also slower compared with CaM-CD2-III-5G. The exchange rate for CaM-CD2-IV-5G in 500  $\mu$ M EDTA was not measured since the signal to noise ratio was too low.

### **5.1.2 2D HX Experiments on CD2 and CaM-CD2-III-5G**

The 1D HX experiments of CD2 and CaM-CD2-III-5G suggest that some of the available protons that are buried inside the hydrophobic core were not deuterated even after incubation in D<sub>2</sub>O for 20 hr. To determine the identity of

those residues, 2D HX experiments were performed using the TOCSY experiment and a homonuclear sample. The experimental protein concentration for the CD2 variants was limited below 1 mM to ensure the protein remained soluble (see chapter 3.0). Each TOCSY experiment required a minimum of 10 hours to complete. Therefore, only the residues with very slow HX rate were observed, and the data points were not sufficient for the HX rate calculation.

Four TOCSY spectra with mixing times of 60 ms were collected for both CD2 and CaM-CD2-III-5G. The experimental times for the initial two TOCSY spectra were 10 hours each, and the experimental times for the latter two TOCSY spectra were 20 hours each. The fingerprint region, HE1 proton of Trp region, and sidechain region (the aliphatic protons) of the TOCSY spectra of CD2 are shown in Figures 5.4a, 5.4b and 5.4c, respectively. The majority of the HN-HA crosspeaks for CD2 were not observed in the first TOCSY spectrum, which indicates that those backbone amide protons have already been replaced with deuterium. The HN-HA crosspeaks for I14, D62, T69, and L80 disappeared after 20 hours of experimental time. The HN-HA crosspeaks for G4, L16, E29, V30, R31, W32, E33, L63, K64, T76, N77, T79, Y81, L93, and L95 were observed in the 4<sup>th</sup> TOCSY spectrum, and the protein sample had been incubated in the D<sub>2</sub>O solution for more than 60 hours. These HN protons are buried inside the hydrophobic core of the CD2 hence no deuterium exchange occurs. Consistent with the previous 1D study, the HE1 proton of W32 remained unchanged after 60 hours in D<sub>2</sub>O solution, while the HE1 proton of W7 was not observed in any of

the HX spectra. The aliphatic sidechain proton is non-exchangeable. The example sidechain region of the CD2 spectrum (Figure 5.4c) shows that the sidechains of T5, V80, and R96 of CD2 were not exchanged. The crosspeak intensities ensure the spectra were compared at the correct ratio and the crosspeaks that disappeared were due to deuterium exchange.

The fingerprint region, HE1 proton of the Trp region, and the sidechain region of the TOCSY spectra of CaM-CD2-III-5G are shown in Figures 5.5a, 5.5b, 5.5c, respectively. The majority of the HN-HA crosspeaks for CaM-CD2-III-5G were not observed during the initial TOCSY spectrum like those for CD2. However, the HN-HA crosspeaks for residues I14, L16, E29, V30, R31, W32, L80 (63), T86 (69), Y93 (76), N94 (77), V95 (78), T96 (79), L110 (93), and L112 (95) were observed in the second TOCSY spectrum but were not observed in the third spectrum, which indicates that the HX rate of these HN protons are faster than those of wild type CD2. Similarly, the HE1 proton of W32 in CaM-CD2-III-5G were observed in the second TOCSY spectrum but were not observed in the third spectrum. On the sidechain region of the CaM-CD2-III-5G, the crosspeaks for T5, V97 (80), and R113 (R96) were located in similar positions as in the wild type CD2 spectrum. No major changes were observed in the sidechain region for CD2 compared to CaM-CD2-III-5G.

### **5.1.3 Comparing the HX Studies of CD2 to CaM-CD2-III-5G and CaM-CD2-IV-5G**



The 1D HX experiments indicate that the HX rates for the CD2 variants are as follow, CaM-CD2-IV-5G  $\rightarrow$  CaM-CD2-III-5G  $\rightarrow$  wild type CD2 (fastest to the slowest). The 2D HX experiments also indicate that the HX for the CaM-CD2-III-5G is on a faster time scale than the wild type CD2. There are several factors that can contribute to the faster HX exchange rates observed in the CaM-CD2-III-5G and CaM-CD2-IV-5G.

First, what is the HX Rate of the EF-loop and glycine linker residues? The EF-loop was inserted between S52 and G53 to allow the EF-loop to be solvent exposed with a fast HX rate. It is possible that the inserted EF-loop has a faster exchange rate, while the HX rates observed on the engineered protein are the result of using the bulk area. The homonuclear resonances for the inserted EF-loop III and IV overlapped with the resonances of the CD2 host protein. Such overlap prevents us from answering this question using homonuclear NMR. Detailed studies using  $^{15}\text{N}$  labeled proteins will be carried out in the future. The HX rate of CD2 variants will further measured using the  $^{15}\text{N}$  HSQC with the CLEANEX filter, which does not require the addition of  $\text{D}_2\text{O}$ .

Second, how much does the EF-loop insertion affect the hydrophobic core? Did the inserted EF-loop affect the dynamic properties of the host protein, CD2? The 2D HX experiments of wild type CD2 show that some HN protons were not exchanged after 60 hours of experimental time. On the other hand, most of these HN resonances were exchanged after 20 hours for CaM-CD2-III-5G. As shown in Figures 5.6a and 5.6b, the HN protons of these residues are

located in the center of the hydrophobic core. The chemical shift values, NOE patterns, and backbone coupling constants of the host protein portion of CaM-CD2-III-5G indicate that the tertiary structure and the packing of the protein are similar to that of wild type CD2. The resonances for the aromatic ring protons and the sidechain protons of CaM-CD2-III-5G are at the same location in both H<sub>2</sub>O and D<sub>2</sub>O form. The HX properties of CD2 were previously studied by Parker and coworkers (161-166). Their observations indicated that the conformation and packing of CD2 is not dependent on pH. In the presence of D<sub>2</sub>O, the pH of CaM-CD2-III-5G was measured at 7.36, which will not affect the conformation of the host protein of CaM-CD2-III-5G. Even though the inserted EF-loop does not cause a large change in the host protein, it may have changed the dynamic properties of the CD2 host protein. The HX rate from the 1D NMR experiments suggest that the EF-loop IV might cause more changes to the dynamic properties on the CD2 host protein than the EF-loop III.

## **5.2 T<sub>1</sub>, T<sub>2</sub>, and NOE Relaxation Studies for CaM-CD2-III-5G**

### **5.2.1 T<sub>1</sub> Relaxation Studies on CaM-CD2-III-5G**

To further examine the dynamic properties, we have carried out extensive studies of T<sub>1</sub>, T<sub>2</sub>, and NOE relaxation studies using <sup>15</sup>N-labeled protein and a dynamic heteronuclear experiment. The HSQC spectra with 130 ms delay were collected at the start and end of the experiments and were compared; there was less than a 3% peak intensity difference. Therefore, the sample condition

remains the same throughout the experiment. The assignments for the T1 spectra of CaM-CD2-III-5G were based on the homonuclear and heteronuclear experiments described in chapter 3.0. Figure 5.7 shows the T1 relaxation spectra for CaM-CD2-III-5G. CaM-CD2-III-5G is composed of 116 residues. A total of 82 dispersive resonances were assigned, and only 25 residues were not distinguished in the T1  $^{15}\text{N}$  HSQC spectrum of CaM-CD2-III-5G. Figure 5.8 shows an area of relaxation decay at different relaxation times for CaM-CD2-III-5G. The spectra recorded using 480 (yellow) and 1500 (red) ms relaxation delays are stacked on top of the spectrum recorded using 0 ms relaxation delay (blue). As shown in Figure 5.8, residues K81 and L112, located at the  $\beta$ -strands D and G, have stronger peak intensities in the 1500 ms spectrum. On the other hand, A65 at the grafted loop has a much weaker intensity. These results suggest that residues K81 and L112 have longer T1 relaxation times in comparison to A65.

An example of the data fit for the T1 relaxation time decay using equation 4.1 is shown in Figure 5.9. The resonances for 9 residues overlapped with on another (residues A9, N17, V30, T37, V39, G78, T96, A109, and I114), and were not used in the calculation. The T1 relaxation times for CaM-CD2-III-5G are summarized in Figure 5.10a and Table 5.2. As shown in Table 5.3 and Figure 5.12, the average T1 relaxation time for the host protein region of CaM-CD2-III-5G is 735 ms. Residues 3, 4, 5, 25, 46, 47, 49, 50, 57, 58, 59, 62, 63, 65, 67, 68, 71, and 100 have T1 relaxation lower than 650 ms. The rest of the host protein

residues have T1 relaxation time greater than 650 ms. The average T1 relaxation time for the inserted sequences (G53 to G69) of the CaM-CD2-III-5G is 568 ms. The results indicate that the inserted region has a more flexible conformation than the host protein region. Residues M46, K47, F49, L50, and A71 close to the insertion region have the T1 values different from the other residues of the host protein, but similar to the inserted residues.

### 5.2.2 Transverse Relaxation Studies on CaM-CD2-III-5G

Similarly, the assignments for the T2 spectra of CaM-CD2-III-5G were based on the homonuclear and heteronuclear strategies described in chapter 3.0. The lowest T2 relaxation delay was set at 10 ms because the gNhsqc.c pulse sequence would not allow the T2 relaxation delay to be set at 0 ms. The T2 relaxation spectra for CaM-CD2-III-5G are shown in Figure 5.11. The HSQC spectra with 30 ms delays collected at the start and end of the experiments were compared, which indicate less than a 3% difference in the peak intensities. This suggests that the sample conditions remain unchanged through out the experiment.

The T2 relaxation times for CaM-CD2-III-5G were calculated using Equation 4.1. The T2 relaxation times for CaM-CD2-III-5G are summarized in Figure 5.10b and Table 5.2. The average T2 relaxation time for the host protein without the inserted residues is 100 ms. The average T2 relaxation time for the inserted sequences is 171 ms. The residues of the N- and C-terminal end and

the inserted EF-hand moiety have T2 relaxation times greater than 110 ms (residues 3, 4, 5, 7, 46, 47, 49, 50, 57, 58, 59, 62, 65, 67, 70, 71, 114, and 116). The rest of the host protein residues have T2 relaxation times less than 110 ms. Unlike the T1 relaxation, the flexible residues that are expected to have shorter T1 values usually have longer T2 values than the rigid ones. The T2 relaxation times for CaM-CD2-III-5G are in good agreement with the T1 relaxations (Figure 5.10)

### 5.2.3 HN One Bond NOE Studies on CaM-CD2-III-5G

For a molecule without internal motion or dynamics, the residues will have an NOE of 1. If the residue is experiencing faster motion, the NOE will be close to 0. The NOE "on" and "off" spectra for CaM-CD2-III-5G are shown in Figure 5.12. The NOE ratios for CaM-CD2-III-5G are summarized in Figure 5.10c and Table 5.2. The dispersive resonances for 81 of the 116 residues were assigned for the backbone of CaM-CD2-III-5G. The resonances for 26 residues were not observed in the  $^{15}\text{N}$  HSQC spectrum. The resonances for 9 residues overlapped with other resonances, and the intensities of these residues were not used for the NOE calculations. The NOEs for 13 of the 81 residues showed large errors and were excluded from the calculations. As shown in Figure 5.10, the 68 calculated NOEs range from 0 to 0.95. The NOEs of residues S3, G4, T5, I14, I18, M46, K47, F49, L50, K57, D58, G59, Y62, A65, E67, G68, and A71 are below 0.60, while the rest of the values are greater than 0.60.

#### 5.2.4 Summary of T1, T2 and NOE Studies on CaM-CD2-III-5G

The EF-loop III of CaM was inserted between S52 and G53 of CD2 protein. Residues M46, K47, F49, L50, and A71 (A54 in CD2) of the CD2 protein are close to the insertion location, and the T1, T2, and NOE relaxation values of these residues all changed significantly in comparison with other host protein residues (Table 5.2). This indicates that the internal dynamics of these residues changed after insertion of the EF-loop. The T1, T2, and NOE values of residues P48, K51, S52, and G70 (G53 in CD2) of the host protein were not determined since the resonances were not observed in the T1, T2, and NOE spectra. The peak intensities of residues A9, N17, V30, W32, T37, V39, G78, T96, A109, and I114 were not measurable, so these residues were excluded from the following calculations.

The relaxation data for the inserted residues K57, D58, G59, Y62, A65, E67 and G68 were obtained, while the values for the rest of the residues were not available due to either signal overlap or no observable peak. The average T1 relaxation time of EF-loop III and the glycine linkers is 23% shorter than that of the host protein in CaM-CD2-III-5G. The average T2 relaxation time of the EF-loop III and the glycine linkers is 40% longer than that of the host protein. The average NOE of the EF-loop III and the glycine linkers is more than 80% smaller than that of the host protein. These T1, T2, and NOE differences indicate that the conformation for the EF-loop III is more dynamic than the host protein and is

in good agreement with the structural calculations and residual dipolar coupling studies discussed in chapter 3.0.

## 5.2.5 Calculation of $S^2$ Values Using ModelFree Simulation

### 5.2.5.1 ModelFree Simulation for CaM-CD2-III-5G

There are total of five different simulating models for ModelFree using the experimental R1, R2, and NOE relaxation data. The explanation for each model is shown in Table 5.3.

Table 5.3 Simulation Approaches for ModelFree

Model	Simulating Parameters
Model 1	$S^2$
Model 2	$S^2, \tau_e$
Model 3	$S^2, R_{ex}$
Model 4	$S^2, \tau_e, R_{ex}$
Model 5	$S^2_f, S^2_s, \tau_e$

Using the experimental relaxation data and the protein coordination file, ModelFree will back calculate the R1, R2, and NOE values for each resonance and also calculate the  $S^2$ ,  $\tau_e$ , and  $R_{ex}$  according to the simulation model procedure. The R1, R2, and NOE values were calculated for 68 residues. The simulation was performed in three stages. In stage 1, the simulation was performed based on the Model1 where only the  $S^2$  order parameters were calculated for each resonance. To verify the accuracy of the simulation, the sum

of experimental errors ( $\Gamma$ ) for each resonances were compared with the critical values ( $\alpha$ ) of the corresponding resonances.

$$\Gamma_{sum} = \left( \frac{(\text{exp}-\text{pred})^2}{(\text{uncert})^2} \right)_{R1} + \left( \frac{(\text{exp}-\text{pred})^2}{(\text{uncert})^2} \right)_{R2} + \left( \frac{(\text{exp}-\text{pred})^2}{(\text{uncert})^2} \right)_{NOE}$$

$$\alpha_{sum} = \left( \frac{(0.1 \times \text{pred})^2}{(\text{uncert})^2} \right)_{R1} + \left( \frac{(0.1 \times \text{pred})^2}{(\text{uncert})^2} \right)_{R2} + \left( \frac{(0.1 \times \text{pred})^2}{(\text{uncert})^2} \right)_{NOE}$$

Equation 4.2

exp = experimental value

pred = predicated value from ModelFree

uncert = uncertainty value used for each type of relaxation parameters

If the sum of experimental value is smaller than the critical values (preferably more than 50% smaller), this simulation model is suitable for this resonance. If the sum of the experimental value is larger than the critical value, this simulation model is not sufficient for this resonance and the simulation should be performed using a different model. The model1 was suitable for 21 residues. Then the remaining 47 residues were simulated using model2 and model3 in a parallel fashion during stage 2. The same error verification protocol used in model1 is applied here for model2 and model3. The model that yields the lower percentage error ratios ( $\Gamma/\alpha$ ) would be considered suitable for that particular resonance.

Model2 simulates the  $S^2$  and  $T_e$ ; model3 simulates the  $S^2$  and  $R_{ex}$ . The model2 simulation was chosen for 6 residues, while the model3 simulation was chosen for 23 residues. The other resonances (S3, G4, T5, G8, I18, Q22, K47, F49, L50, K57, L58, G59, Y62, A65, E67, G68, A71, and E116) have large errors and are subject to additional simulations. These resonances are mainly from the



terminal end of the host protein and also the inserted EF-loop III. They were simulated using model4 and model 5 in a parallel fashion during the stage 3 calculation. Model4 simulates the  $S^2$ ,  $T_e$ , and  $R_{ex}$  and model5 simulates  $S^2_s$ ,  $S^2_f$ , and  $T_e$ . The model that yields the lower percentage error ratios ( $\Gamma/\alpha$ ) between model4 and model5 would be considered suitable for that particular resonance. The model4 simulation was chosen for 13 residues, while the model5 simulation was chosen for 5 residues.

#### **5.2.5.2 The Calculated Order Parameters for CaM-CD2-III-5G**

The final simulation was performed using the best model combination for each resonance. All of the resonances except M46 have the sum of the experimental errors smaller than the critical values. The calculated  $S^2$  values for the 68 residues were graphed as a function of residues in Figure 5.13.

As we mentioned previously, the resonances from the flexible region would have smaller  $S^2$ , while the rigid region of the protein would have greater  $S^2$  values (close to 1). The end of the host protein has lower  $S^2$  values, especially the N-terminal. The C-terminal region of the protein is more ordered than the N-terminal, which is observed in the T1, T2, and NOE relaxation experiments for CaM-CD2-III-5G. The  $S^2$  values for the inserted glycine linkers and EF-loop III residues were much smaller than the host protein residues. The average  $S^2$  value for the entire protein is 0.74, while that for the host protein is 0.83, and for the inserted glycine linkers and EF-loop is 0.35. The average  $S^2$  value for the

insertion region is more than 60% lower than the host protein. The T1 and T2 experiments were more sensitive than the NOE experiment. T1 and T2 experiments observed more residues than the NOE experiment. The NOE experiment only yielded the relaxation data for K57, D58, G59, Y62, A65, E67, and G68 in the inserted sequences. The T1 and T2 experiments observed three additional residues (N60, G61, and I63), and their T1 and T2 values are similar to the other residues in the inserted sequences. These data indicate that the residues in the glycine linkers and the EF-loop III have similar dynamic properties, so the average  $S^2$  order parameters based on the limited number of residues suggest the dynamic behavior of the whole inserted moiety.

#### **5.2.5.3 Comparison of the Order Parameters of CaM-CD2-III-5G to the Corresponding Residues in 6D15 and Wild Type CD2**

The dynamic properties for CD2 and 6D15 reported previously were conducted using the 500 MHz, while the dynamic properties for CaM-CD2-III-5G were conducted using a 600 MHz NMR. Therefore, the field dependent parameters (T1, T2, and NOE) are incomparable. However, the  $S^2$  order parameter is not field dependent, and can be directly compared. The average  $S^2$  values of wild type CD2 and 6D15 was reported as  $0.81 \pm 0.07$  and 0.85, respectively. The average  $S^2$  value for the host protein region of the CaM-CD2-III-5G is 0.83. This suggests that the host protein region of CaM-CD2-III-5G maintain the dynamic properties after the insertion of the EF-loop. The  $S^2$  order

parameter comparison between 6D15 and CaM-CD2-III-5G is shown in Figure 5.14. The C-terminal half of the host protein region of CaM-CD2-III-5G has similar  $S^2$  to the 6D15, while the first 30 residues of the N-terminal of CaM-CD2-III-5G have lower  $S^2$  values. Residues M46, K47, F49, L50, A71, F72, and E73 are close to the EF-loop insertions and the  $S^2$  order parameters for these residues are lower than 6D15. The local dynamic properties around the S52-G53 of CD2 are affected by the EF-loop insertion, but it is not critical to the folding of the host protein.

#### **5.2.5.4 Comparison of the Order Parameters of the Inserted EF-loop III in CaM-CD2-III-5G to the EF-loop III of Calcium Free CaM**

To understand the intrinsic variability of different forms of CaM (see Introduction for more details on different forms of CaM), such as calcium free, calcium loaded, or bound with target peptide, previous studies have been carried out using the HX and relaxation experiments. The research groups of Bax, Ikura, and Forsen used the HX experiments to characterize the dynamics of calcium free calmodulin. All three research groups concluded that in the absence of calcium, the conformation of the EF-loops are very dynamic mainly due to the charge repulsion. The structural studies by Ikura and co-workers on the calcium free calmodulin indicated that all of the Met sidechains are buried inside the hydrophobic surface, which is the reason why CaM has lower affinity for the target peptide in the resting state. Once the CaM is activated by calcium, the

conformation rearrangement will expose the molecular recognition surface to the solvent and allow CaM to bind to target peptide(ref).

The dynamic properties of the EF-loop III in CaM-CD2-III-5G were compared to that of calcium free CaM obtained by Bax and co-workers. The results have revealed that the isolated EF-loop III in an engineered protein is less ordered than the intact CaM (Table 5.4). The average order parameter of the EF-loop III in calcium free CaM is almost two-fold greater than that in CaM-CD2-III-5G. This is likely due to the following two reasons: First, the glycine linkers in CaM-CD2-III-5G allow the inserted EF-loop III to be more flexible. In the absence of metal ion, the repulsion between the charged residues causes the EF-loop to be dynamically disordered. Second, there are no paired-paired interactions with another EF-hand motif. For calcium free CaM, there are several none covalent interactions, such as a) intra-helix motif interactions between the entering and exiting helices of an EF-hand motif, b) inter-helix motif interactions between the paired EF-hand motifs (such as site III and IV of CaM), and c) inter-strand interactions between the small  $\beta$ -strand on the EF-loop of the paired EF-hand motifs. In CaM-CD2-III-5G, there are no entering and exiting helices so there are no intra- and inter-helix motif interactions. The diffusion studies suggest that the CaM-CD2-III-5G remains as a monomer in the absence and presence of La(III), suggesting that there is no interaction between the EF-loop III of one molecule to another in solution. Therefore the dynamic properties of the

inserted EF-loop are less ordered in comparison with the EF-loop III of intact CaM.

In future studies, the contribution of the glycine linker to the conformation flexibility can be further investigated using the CaM-CD2-III-0G construction that does not use the glycine linkers to attach the EF-loop to a host protein. The dynamic studies for the calcium loaded CaM-CD2-III-5G will also be studied.

### **5.3 T1 and T2 Relaxation Studies for CaM-CD2-IV-5G**

#### **5.3.1 T1 Relaxation Studies on CaM-CD2-IV-5G**

The sample preparation and NMR experimental parameters for the T1 relaxation studies on CaM-CD2-IV-5G were carried out in a similar fashion as the CaM-CD2-III-5G studies described in section 4.2.1. The T1 spectra were carried out using the following relaxation delays, 0, 10, 60, 130, 230, 340, 480, 740, 1000, and 1500 ms. The T1 relaxation times for CaM-CD2-IV-5G were calculated using equation 4.1. The T1 relaxation spectra for CaM-CD2-IV-5G are shown in Figure 5.15. T1 relaxation times for CaM-CD2-IV-5G are summarized in Figure 5.16a and Table 5.2.

CaM-CD2-IV-5G is composed of 116 residues. The sequential assignment of EF-loop IV and the glycine linkers for CaM-CD2-IV-5G are incomplete. The host protein region of CaM-CD2-IV-5G consists of 99 residues, 15 of them were not observed in the  $^{15}\text{N}$  HSQC spectrum of CaM-CD2-IV-5G. The resonances for 10 residues overlap with each other (A9, N17, V30, W32,

T37, V39, F49, G78, T96, and A109). The peak intensities of the HSQC spectra with 130 ms delay collected at the start and end of the experiments were less than 3% different, suggesting that the sample conditions remained the same through out the entire experiment set. Residues 3, 4, 35, 37, 47, 50, 91, and 92 have T1 relaxation times lower than 710 ms. The rest of the host protein residues have T1 relaxation times greater than 710 ms. The average T1 relaxation time for the host protein region of CaM-CD2-IV-5G is 782 ms.

### 5.3.2 Transverse Relaxation Studies on CaM-CD2-IV-5G

The sample preparation and NMR experimental parameters for T2 relaxation studies on CaM-CD2-IV-5G were carried out in a similar fashion as the CaM-CD2-III-5G studies described in section 4.2.1. The T2 relaxation spectra were carried out using the following relaxation delays, 10, 30, 50, 70, 90, 110, 130, and 150 ms. The T2 relaxation times for CaM-CD2-IV-5G were calculated using equation 4.1. The T2 relaxation spectra for CaM-CD2-IV-5G are shown in Figure 5.17. The T2 relaxation times for CaM-CD2-IV-5G are summarized in Figure 5.16b and Table 5.2.

Among the 99 residues of the host protein region, a total of 72 dispersive resonances were assigned for the backbone of CaM-CD2-IV-5G. The resonances of 12 residues overlap with each other (A9, N17, V30, W32, T37, V39, M46, F49, G78, T96, N107, and A109), and the decays of these residues were calculated and used with caution. The resonances for the inserted EF-loop

IV and glycine linker residues were not assigned. The peak intensities of the HSQC spectra with 30 ms delays collected at the start and end of the experiments were less than 3% different, suggesting that the sample conditions remained the same throughout the entire experimental set. The flexible region of the CaM-CD2-IV-5G has longer T<sub>2</sub> values than the rigid region of the protein, which is similar to the observation for CaM-CD2-III-5G. Residues 3, 4, 35, 46, 47, 49, 50, 78, 92, and 116 have T<sub>2</sub> relaxation times greater than 110 ms. The rest of the host protein residues have T<sub>2</sub> relaxation times lower than 110 ms. The average T<sub>2</sub> relaxation times for the host protein region of CaM-CD2-IV-5G is 104 ms.

### **5.3.3 Comparing the T<sub>1</sub> and T<sub>2</sub> Relaxation Studies of CaM-CD2-IV-5G to the CaM-CD2-III-5G**

Comparison of the dynamic properties between CaM-CD2-III-5G and CaM-CD2-IV-5G were focused on the host protein region of these two-engineered proteins. The T<sub>1</sub> relaxation times of CaM-CD2-III-5G and CaM-CD2-IV-5G are shown in Figure 5.18a. The T<sub>2</sub> relaxation times of CaM-CD2-III-5G and CaM-CD2-IV-5G are shown in Figure 5.18b. The flexible regions of CaM-CD2-III-5G and CaM-CD2-IV-5G, such as the terminal ends or the loops that are in between  $\beta$ -strands, all have smaller T<sub>1</sub> values and greater T<sub>2</sub> values. The average T<sub>1</sub> relaxation times of CaM-CD2-III-5G and CaM-CD2-IV-5G are 735 and 782 ms, respectively. The average relaxation time of CaM-CD2-IV-5G is

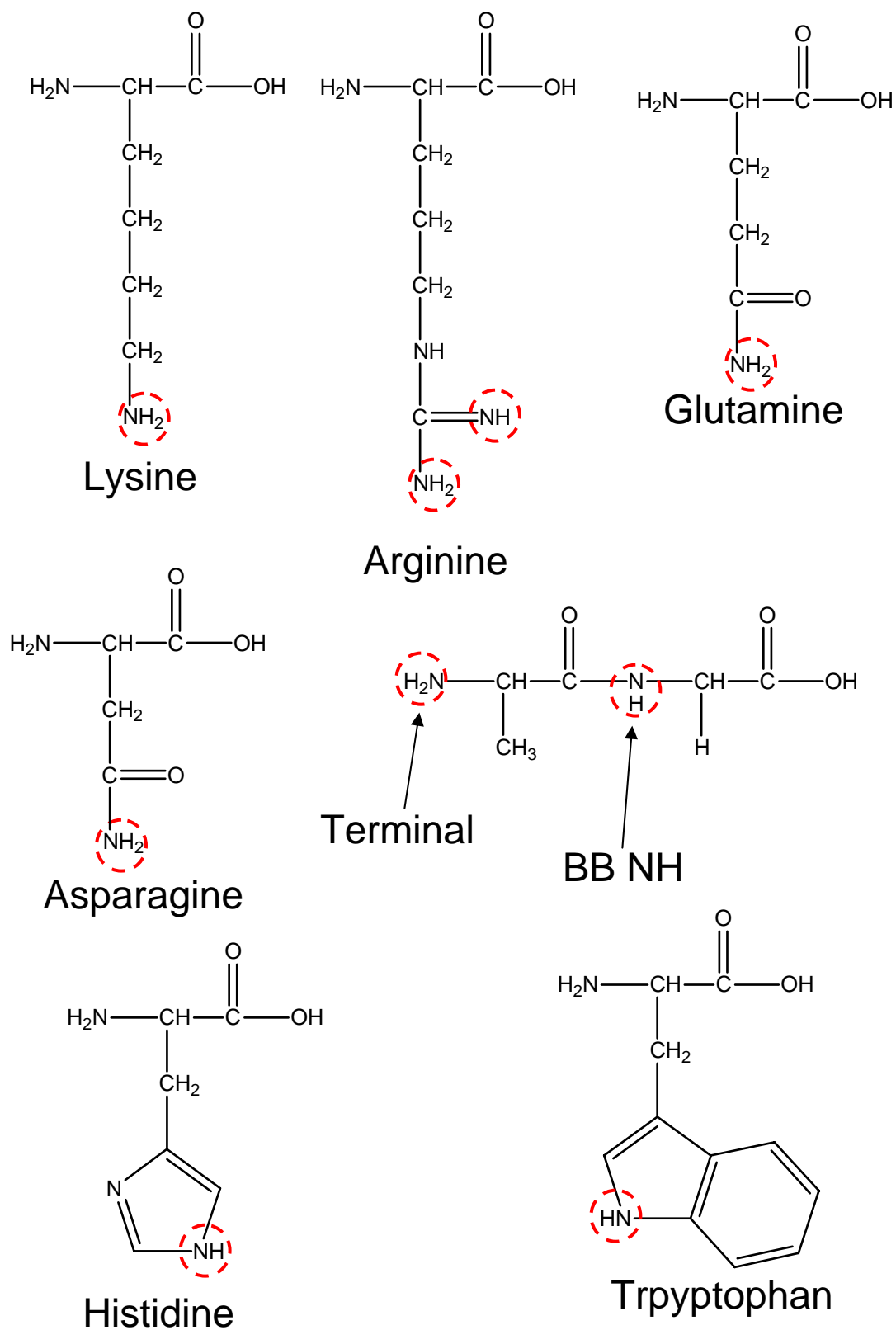
slightly longer than CaM-CD2-III-5G, but the overall patterns for T1 relaxation times are similar. The average T2 relaxation times of CaM-CD2-III-5G and CaM-CD2-IV-5G are 100 and 104 ms, respectively. The average T2 relaxation time for CaM-CD2-IV-5G is in good agreement with that of CaM-CD2-III-5G.

For a more detailed comparison, the engineered protein was dissected into 21 sections. The average T1 and T2 values of each section were based on the observable residues within each section (Figure 5.18 and Table 5.6). The average T1 relaxation times of CaM-CD2-IV-5G for the majority of  $\beta$ -strands and turns are higher than that of CaM-CD2-III-5G, but they are within the calculated errors. The average T2 relaxation times for the majority of the sections in CaM-CD2-IV-5G are within 5 ms of CaM-CD2-III-5G. There are four sections that exhibit a larger difference between CaM-CD2-III-5G and CaM-CD2-IV-5G. The first section is the N-terminal end of the engineered protein (residues 3 and 4), which is very flexible. The average T1 relaxation of both CaM-CD2-III-5G and CaM-CD2-IV-5G are lower than 650 ms, while the average T2 relaxation times are both greater than 110 ms. The terminal ends of the protein usually exhibit rapid conformational change; therefore, it is difficult to characterize the dynamic properties in this region.

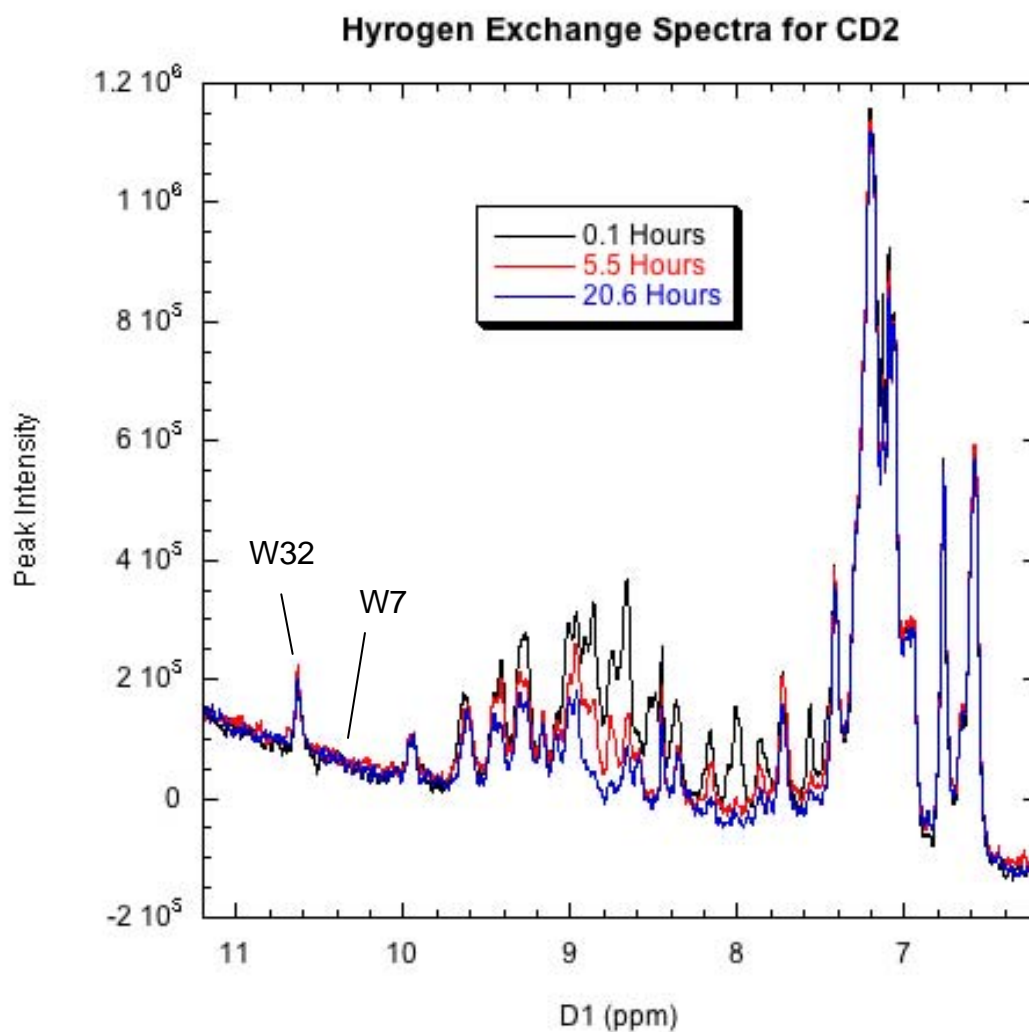
Residues 44 to 50 are close to the EF-loop insertion and were divided into three sections for analysis. The T1 values of CaM-CD2-III-5G begin to decrease from residues 44 and 45 (part 1 of the  $\beta$ -turn between strands C' and C"). The T1 values for CaM-CD2-III-5G dropped below 600 ms from residues 46 to 50



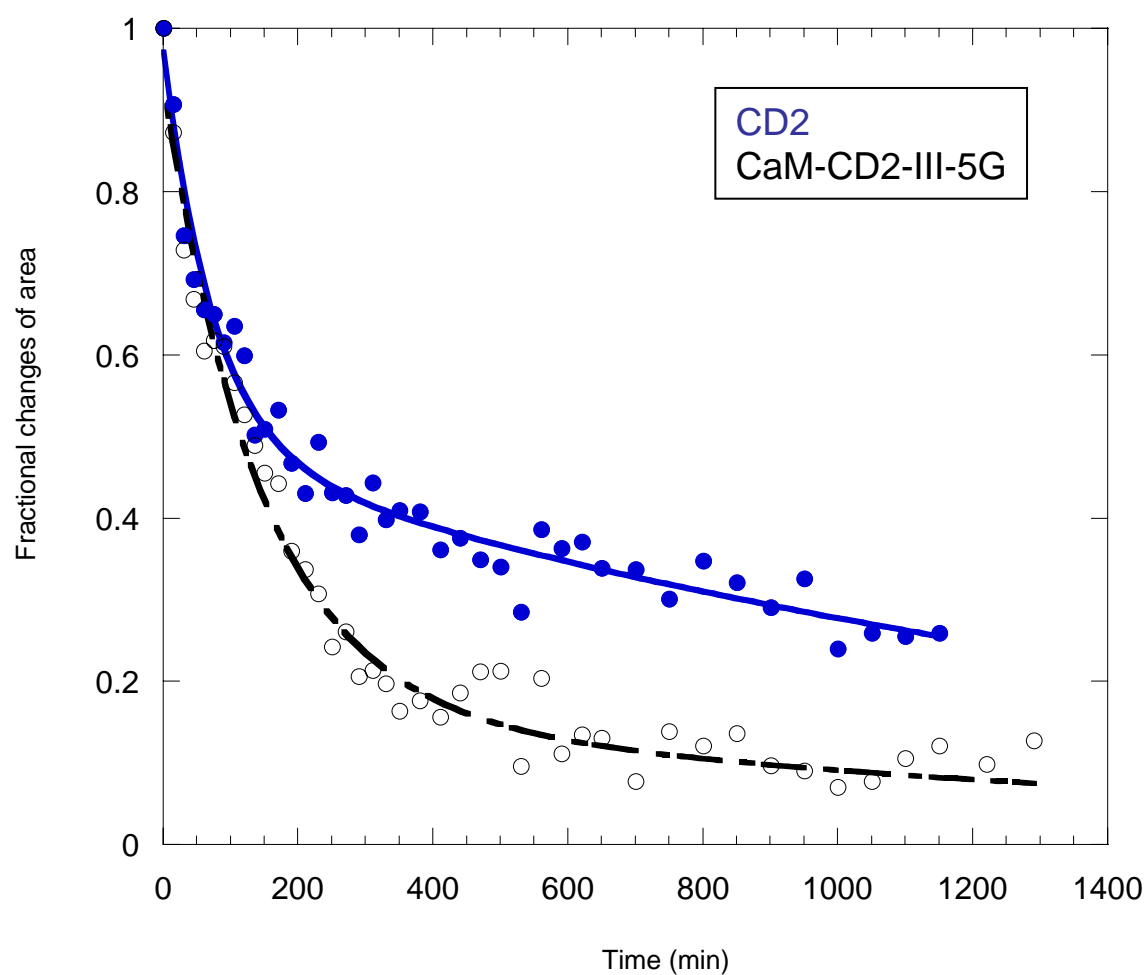
(part 2 of  $\beta$ -turn between strands C' and C'' and strand C''). On the other hand, the T1 values for residues 44 to 49 of CaM-CD2-IV-5G were all longer than 700 ms. The longer T1 relaxation time is an indication that this region of CaM-CD2-IV-5G is more rigid than that of CaM-CD2-III-5G. The T2 values for residues 44 and 45 are less than 110 ms for both CaM-CD2-III-5G and CaM-CD2-IV-5G. The T2 values for residues 46 to 50 are all longer than 110 ms for both proteins. Even though there are significant differences between the T2 values of CaM-CD2-III-5G and CaM-CD2-IV-5G in these sections, the T2 values for both proteins indicate that residues 46 to 50 possess more flexible properties than the other regions of the proteins. The T2 values of CaM-CD2-IV-5G are in good agreement with the T2 values of CaM-CD2-III-5G, but the T1 values of residues 44 to 50 of CaM-CD2-IV-5G do not agree with CaM-CD2-III-5G. It is difficult to conclude the differences in the dynamic properties of the proteins without calculating  $S^2$  order parameters. The current T1 and T2 results indicate that in the absence of calcium, the overall dynamics of the CD2 host protein for CaM-CD2-III-5G and CaM-CD2-IV-5G are similar. In the future, the assignment will be completed in order to compare the  $S^2$  order parameters.



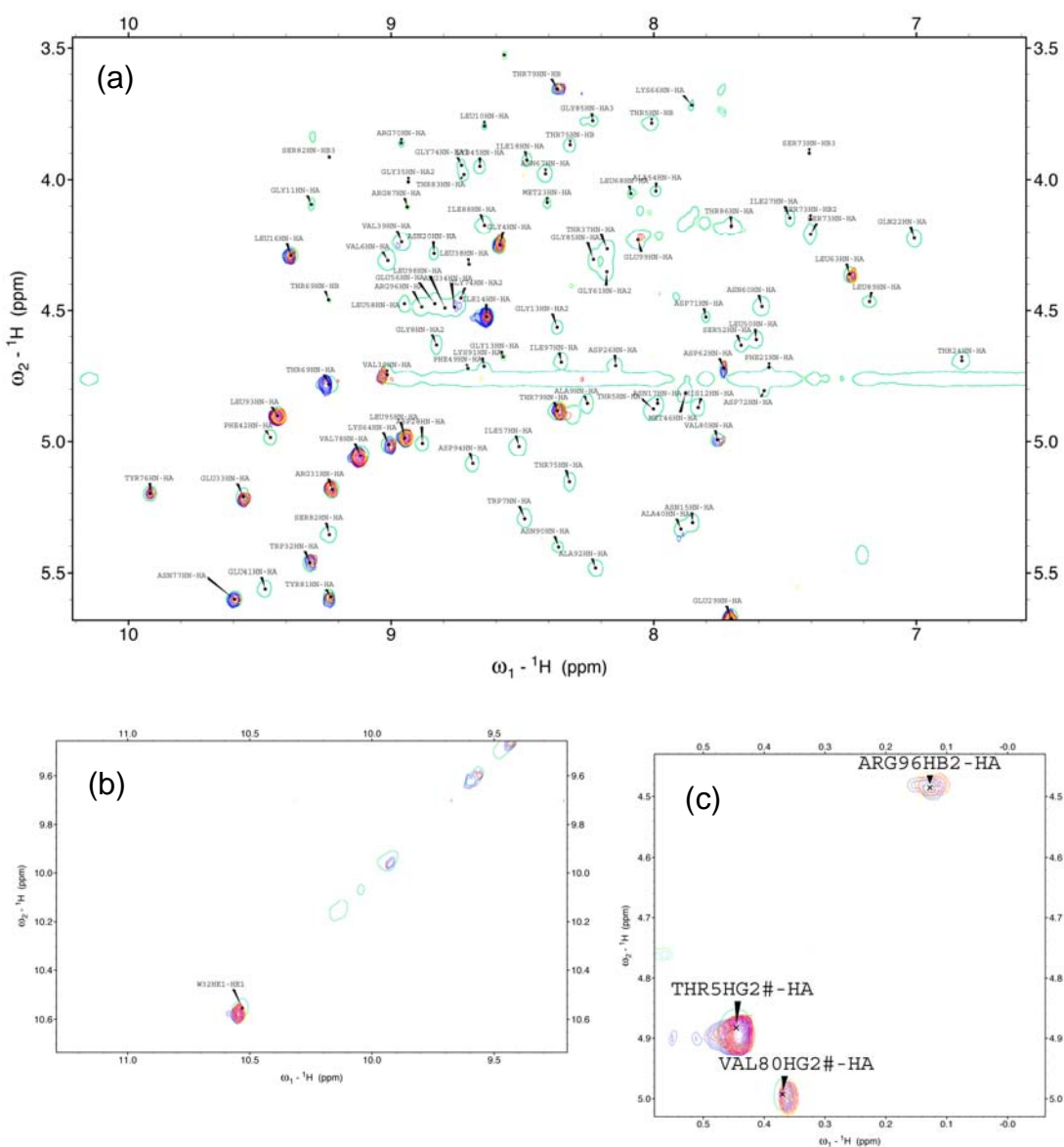
**Figure 5.1 Exchangable protons in protein:** They are the sidechain of Lys, Arg, Gln, Asn, His, and Trp. The backbone amide proton is liable proton.



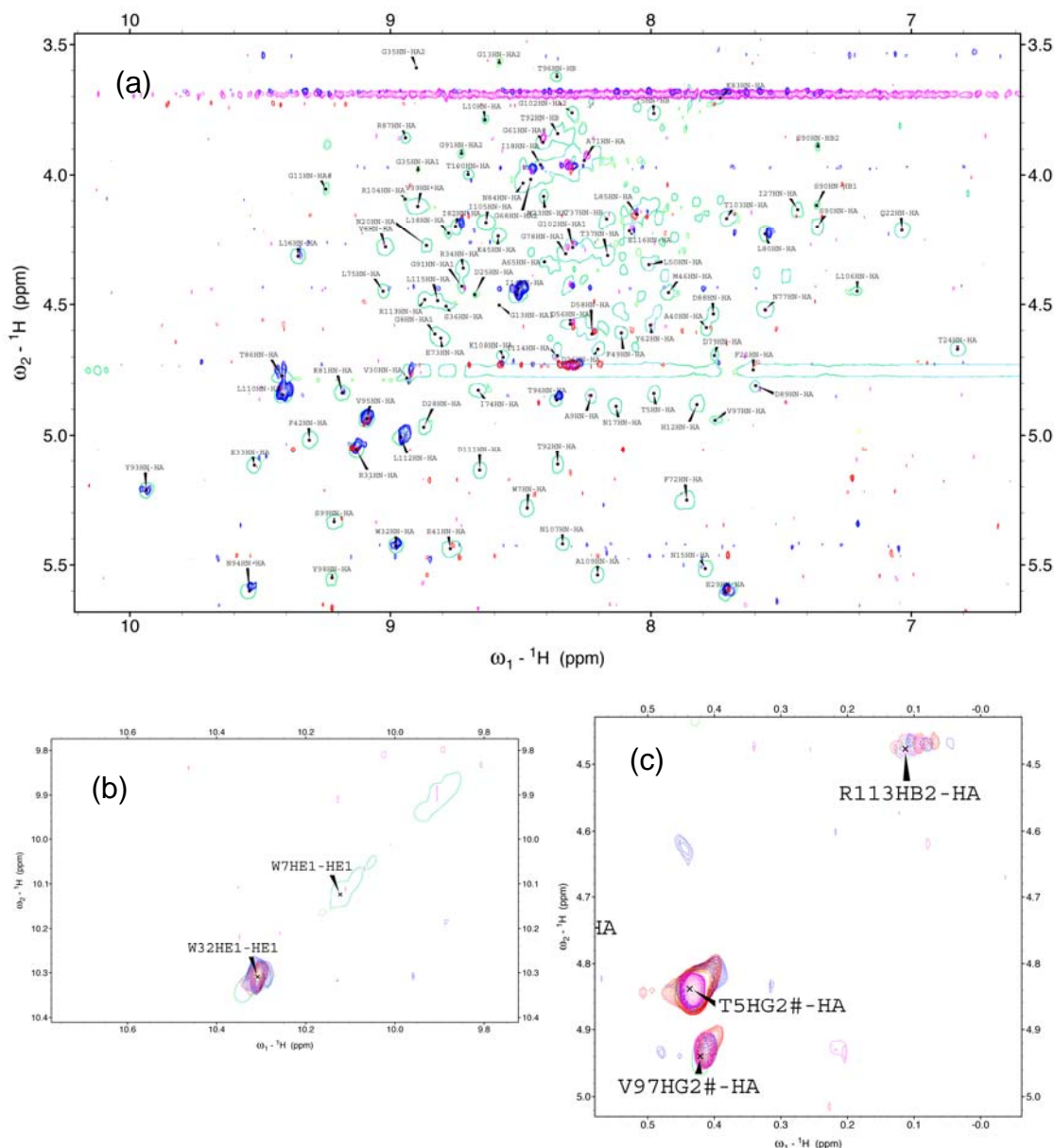
**Figure 5.2 Hydrogen exchange spectra of CD2:** The experiment was conducted by dissolving the CD2 protein (powder form) in 500  $\mu\text{L}$   $\text{D}_2\text{O}$  solution. The hydrogen exchange spectra collected at 0.1, 5.5, and 20.6 hours are shown in black, red, and blue color, respectively.



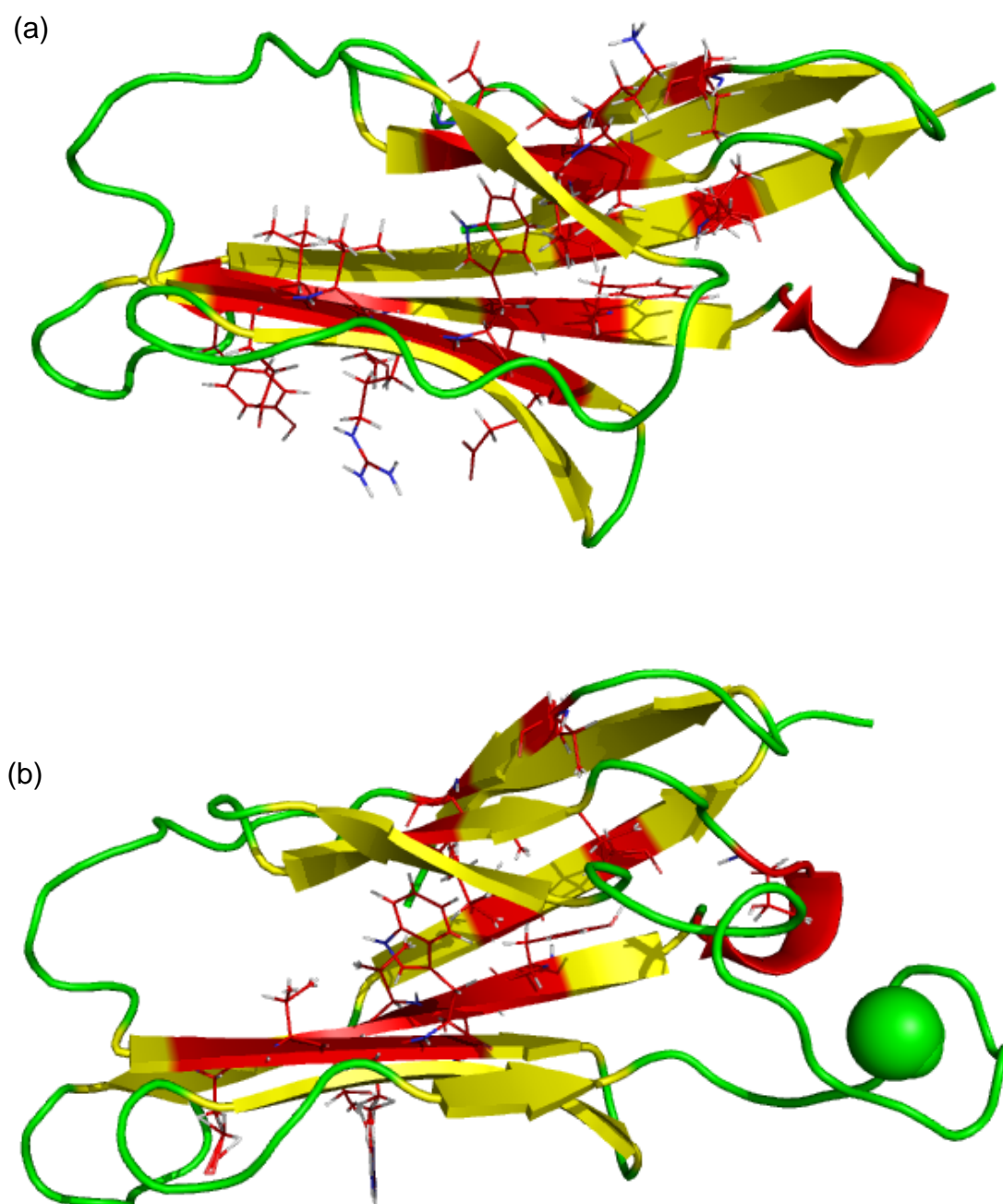
**Figure 5.3 The fitting curves for the HX rates of CD2 variants:** Wild type CD2 has slower HX rate in comparison to CaM-CD2-III-5G.



**Figure 5.4 TOCSY spectra of CD2 in D<sub>2</sub>O:** (a) Fingerprint region, (b) HE1 proton of Trp, (c) sidechain region of the TOCSY spectra of wild type CD2. The CD2 protein in 95 % H<sub>2</sub>O is shown in green color. The spectra collected at 10, 20, 40, and 60 hours are shown as blue, red, yellow, and magenta color, respectively.



**Figure 5.5 TOCSY spectra of CaM-CD2-III-5G in D<sub>2</sub>O:** (a) Fingerprint region, (b) HE1 proton of Trp, (c) sidechain region of the TOCSY spectra of CaM-CD2-III-5G. The CaM-CD2-III-5G protein in 95 % H<sub>2</sub>O is shown in green color. The spectra collected at 10, 20, 40, and 60 hours are shown as blue, red, yellow, and magenta color, respectively.

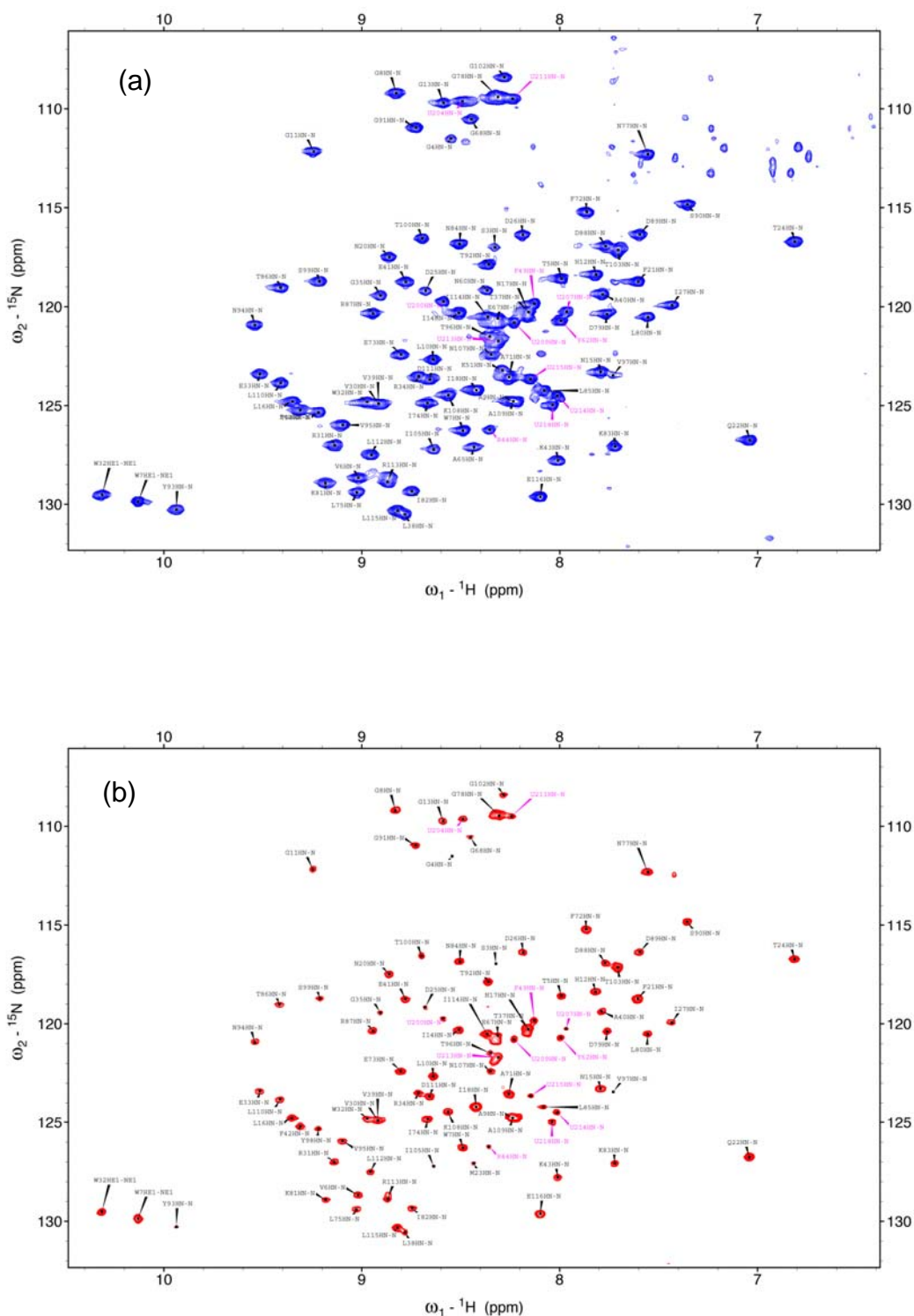


**Figure 5.6 Location of the liable protons with HX rates:** The HN protons that were observed at 20 hours are highlighted in the CD2 (a) and CaM-CD2-III-5G (b) structure.

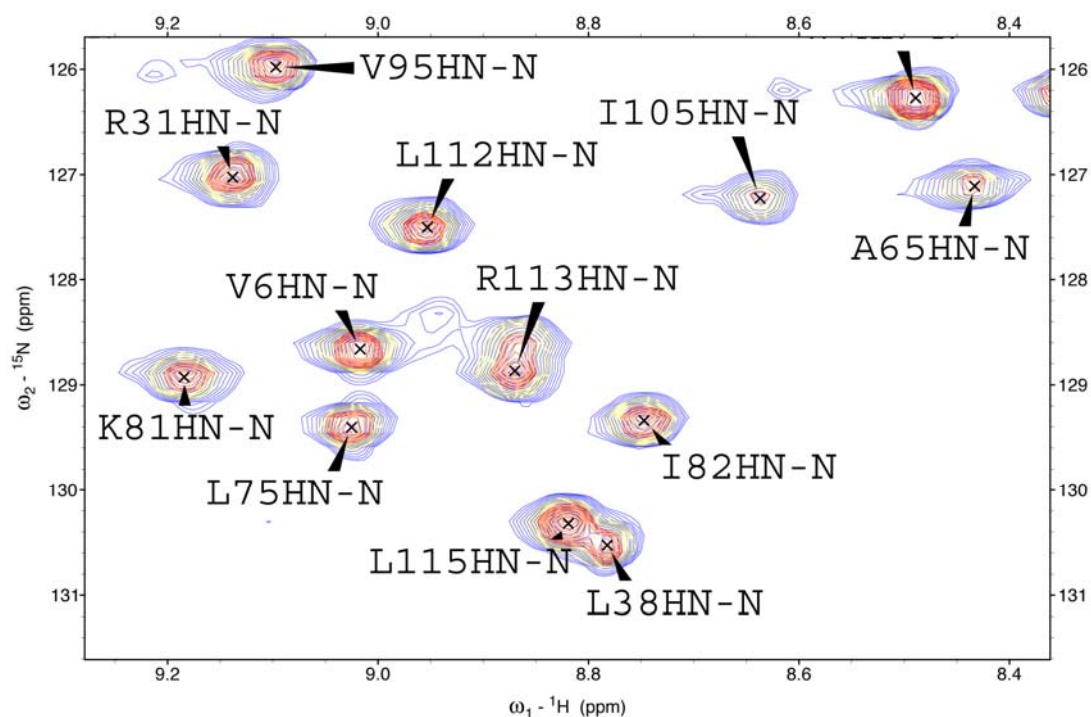
Table 5.1, Hydrogen Exchange Rate for CD2 Variants

Protein	1 mM Ca(II) (minute)		500 uM EDTA (minute)	
	Fast	Slow	Fast	Slow
CD2	$170 \pm 50$	$3800 \pm 1000$	$75 \pm 10$	$1800 \pm 250$
CaM-CD2-III-5G	$100 \pm 20$	$20000 \pm 20000$	$140 \pm 20$	$1600 \pm 1000$
CaM-CD2-IV-5G	$12 \pm 12$	$190 \pm 10$	n/a	n/a

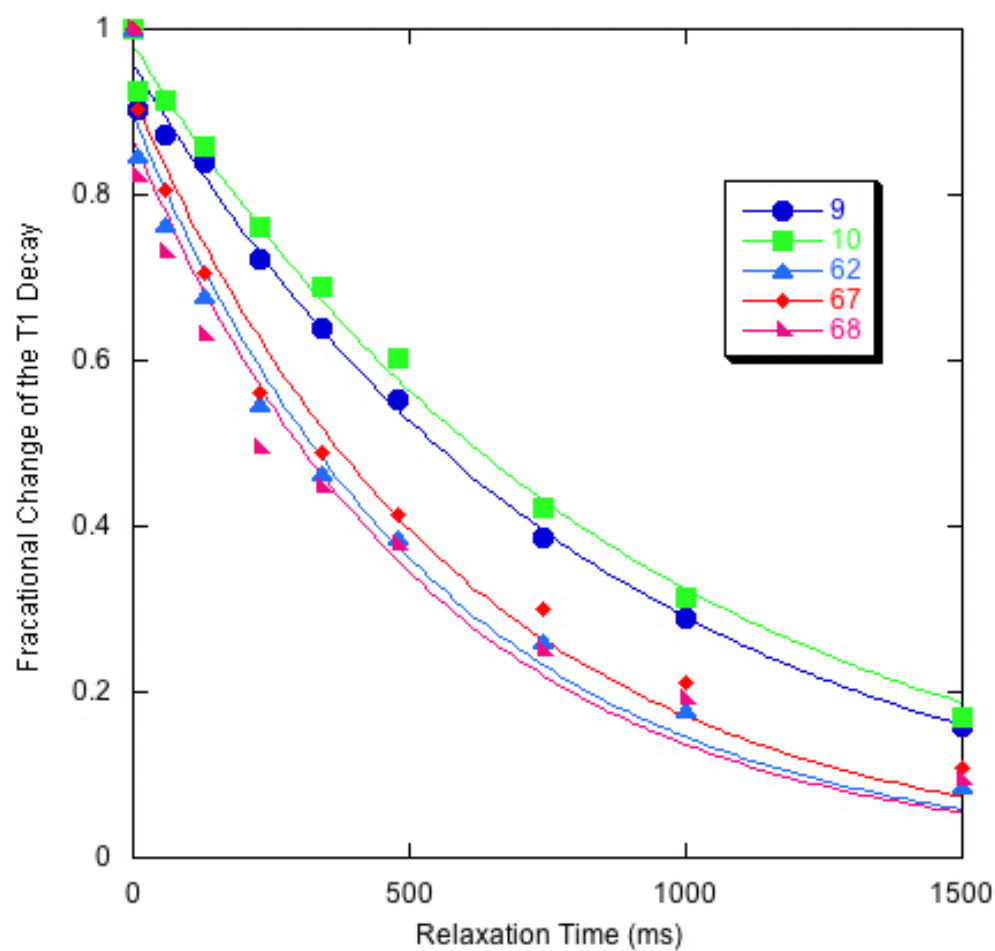




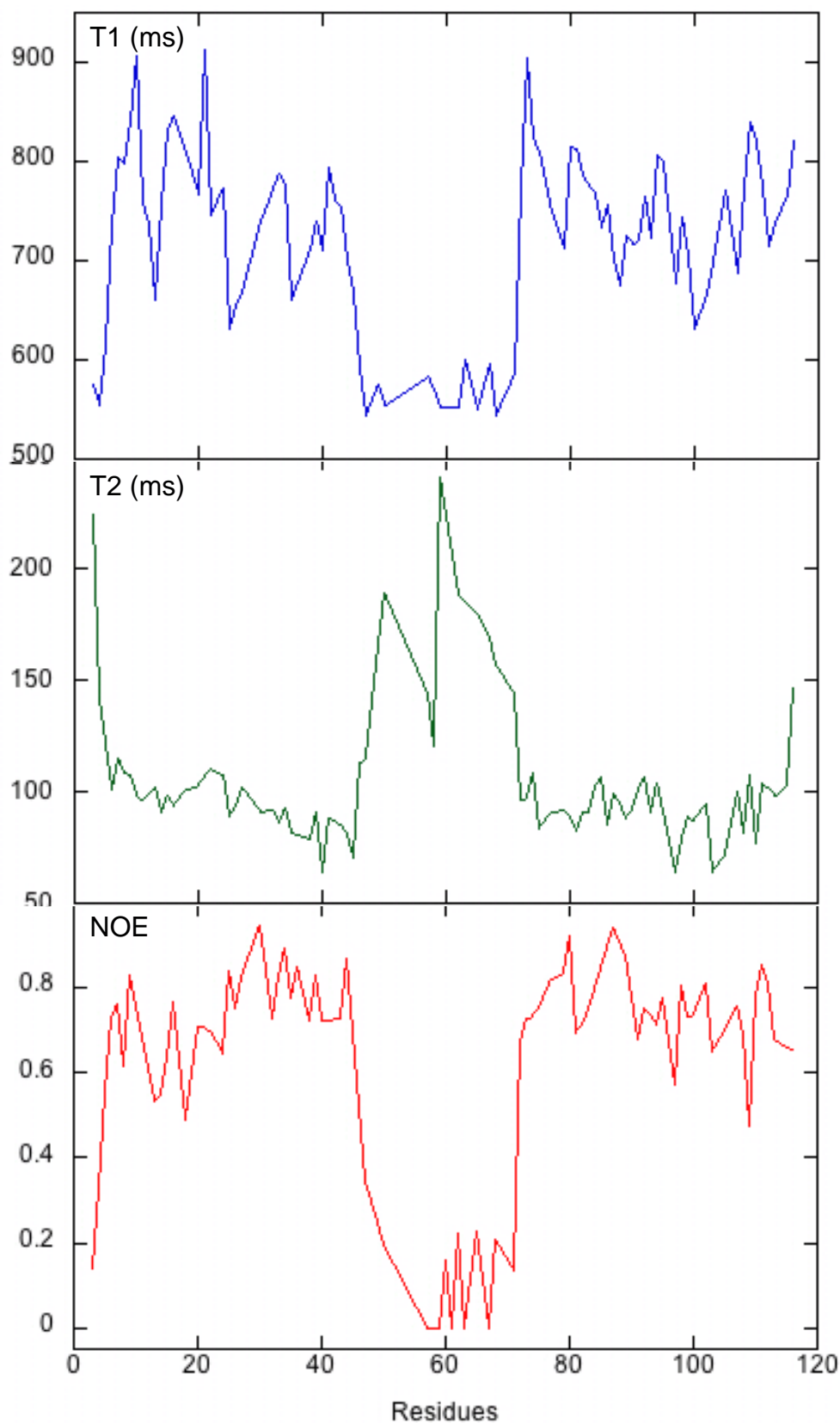
**Figure 5.7 T1 relaxation spectra of CaM-CD2-III-5G:** (a) The spectrum was collected with T1 relaxation delay of 0 ms. (b) the spectrum was collected with T1 relaxation delay of 1500 ms. These spectra were collected using 600 MHz NMR at 25 °C. The protein sample was prepared in 20 mM PIPES 20 mM KCl at pH 6.8.



**Figure 5.8 Intensities decay on T1 spectra:** The spectrum with T1 delay of 0 ms and 1500 ms are shown in blue and red, respectively. Residue A65 and I105 have short T1 relaxation times than L38 and L115 so the intensities of the A65 and I105 crosspeaks are lower. These spectra were collected using 600 MHz NMR at 25 °C. The protein sample was prepared in 20 mM PIPES 20 mM KCl at pH 6.8



**Figure 5.9 T1 data fitting curve for CaM-CD2-III-5G:**  
Residues A9 and L10 (host protein residues) have longer T1 relaxation times than the inserted residues (Y62, E67, and G68).

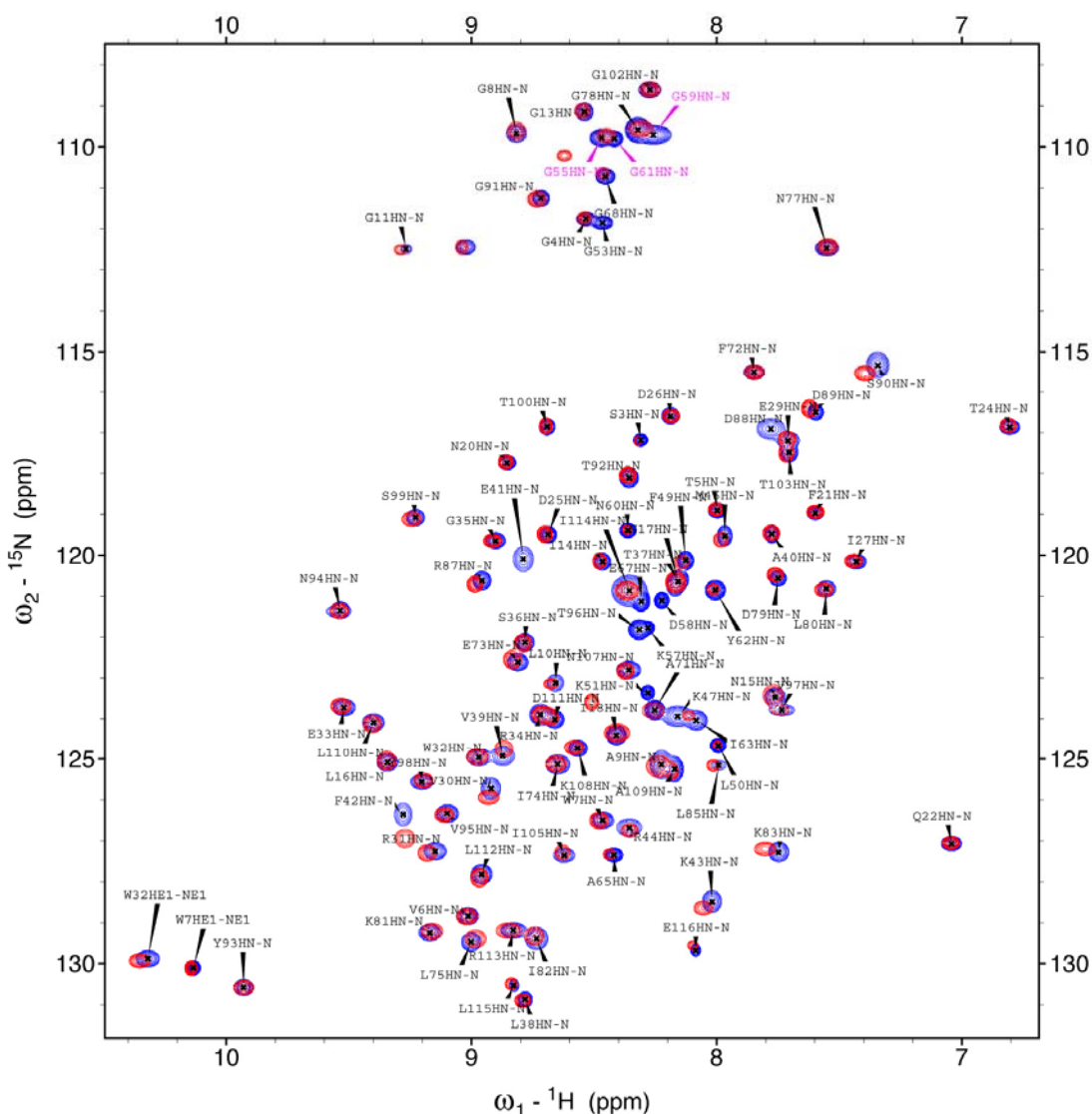


**Figure 5.10 Summaries of the T1, T2, and NOE values:** From the top to bottom, T1, T2, and NOE of CaM-CD2-III-5G. The T1, T2, and NOE values indicated that the inserted EF-loop III has greater mobility than the host protein.

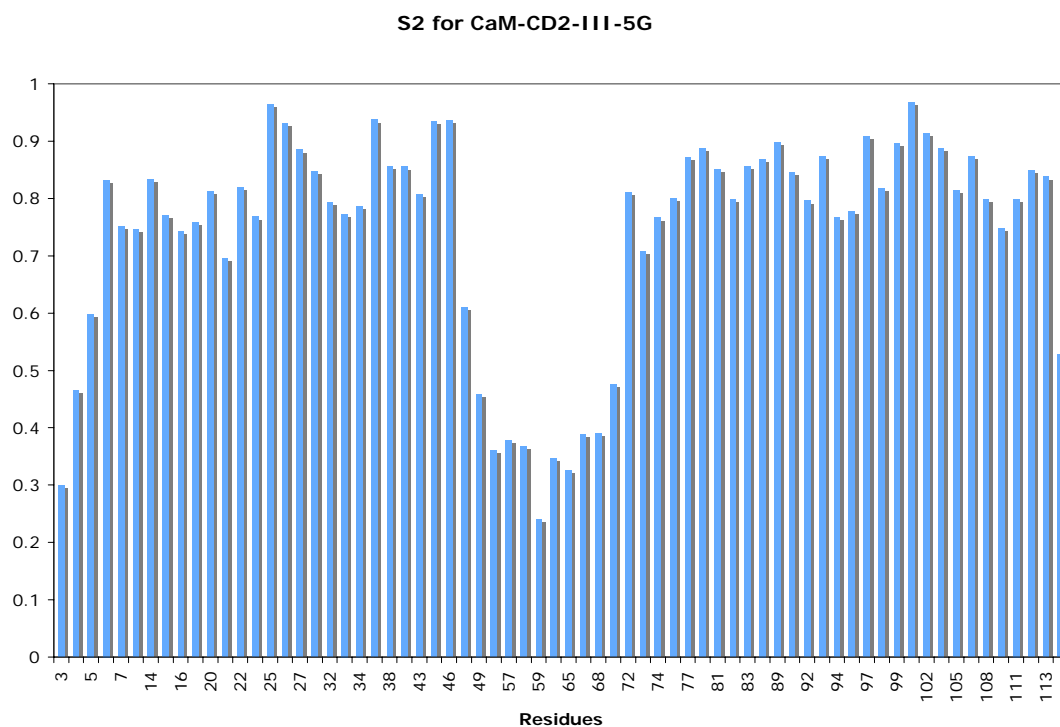
Table 5.2 T1, T2, and NOE Relaxation Values for CaM-CD2-III-5G and CaM-CD2-IV-5G

Residues	Loop III T1	Loop III T2	Loop IV T1	Loop IV T2	Assignment	NOE Values of CaM-CD2-III-5G
3	576.04	223.71	631.96	183.12	S3HN-N	0.299
4	553.58	141.62	662.49	219.18	G4HN-N	0.335
5	614.6	121.32	730.43	122.84	T5HN-N	0.453
6	735.37	101.03	754	100.87	V6HN-N	0.724
7	805.28	115.3	799.66	103.43	W7HN-N	0.774
8	797.89	108.01	812.73	98.907	G8HN-N	0.624
9	840	107.16	904.23	102.44	A9HN-N	0.924
10	906.8	98.313	901.55	86.561	L10HN-N	1.294
11	758.9	96.773	770.21	104.54	G11HN-N	0.238
12	736.52	98.916	750.96	92.393	G13HN-N	0.438
13	661.52	102.22	756.55	98.528	I14HN-N	0.579
14	755.68	90.543	754.85	86.246	N15HN-N	0.643
15	831.86	98.118	806.28	95.081	L16HN-N	0.667
16	845.02	93.532	862.47	100.41	I18HN-N	0.473
17	912.89	87.048	763.67	124.68	N20HN-N	0.646
18	809.09	100.94	839.6	103.49	F21HN-N	0.903
20	767.03	103.07	868.33	104.27	Q22HN-N	0.609
21	912.06	106.46	975.94	102.69	T24HN-N	0.553
22	746.22	110.02	852.16	109.43	D25HN-N	0.930
23					D26HN-N	0.747
24	773.36	107.54	817.98	102.53	I27HN-N	0.811
25	632.23	89.258	730.8	93.224	R31HN-N	0.831
26	653.9	93.646	754.8	95.548	W32HN-N	0.814
27	667.48	102.14	718.82	100.8	G35HN-N	0.766
			773.89	86.677	S36HN-N	0.816
30	739.63	90.932	817.46	85.97	L38HN-N	0.595
31	754.94	91.86	747.43	88.048	V39HN-N	0.714
32	772.86	91.521	781.82	81.301	A40HN-N	0.879
33	788.17	85.718	812.94	80.731	E41HN-N	0.423
34	775.46	92.259	770.42	93.942	F42HN-N	0.080
35	661.2	81.183	702.92	81.003	K43HN-N	0.857
			761.58	75.197	R44HN-N	0.935
37	912.89	87.048	688.58	263.84	K47HN-N	0.951
38	711.52	78.481	792.4	82.901	F49HN-N	0.369
39	739.63	90.932	764.73	90.086	L50HN-N	0.189
40	711.17	64.313	748.73	66.51	K57HN-N	0.000
41	794.74	88.001	872.68	90.404	D58HN-N	0.000
42	759.73	87.034	807.98	88.667	G59HN-N	0.000
43	754.21	84.843	796.34	78.719	N60HN-N	0.103
44	699.42	80.302	813.87	54.386	G61HN-N	0.000
45	666.43	70.818	777.93	83.557	Y62HN-N	0.222
46	592.29	112.92	724.64	121.9	I63HN-N	0.000
47	542.85	144.7	702.34	161.38	A65HN-N	0.203
49	574.21	165.32	763.67	124.68	E67HN-N	0.000
50	554.64	189.03	683.56	207.89	G68HN-N	0.183
57	583.57	143.72			A71HN-N	0.112
58	569.15	120.27			F72HN-N	0.632
59	551.54	240.42			E73HN-N	0.801
62	552.05	188.03			I74HN-N	0.768
63	599.13	102.68			L75HN-N	0.447
65	550.05	179.95			N77HN-N	0.655
67	596.1	169			G78HN-N	0.223
68	543.18	156.77			D79HN-N	1.010
71	586.07	143.88			L80HN-N	0.846
72	772.37	96.578	798.5	91.757	K81HN-N	0.942
73	904.61	97.059	896.51	95.325	I82HN-N	0.725
74	822.23	107.96	896.95	99.943	K83HN-N	0.706
75	807.63	83.208	838.95	80.024	D88HN-N	0.000
77	748.91	91.008	832.5	98.364	D89HN-N	0.743
78	601.27	172.85	727.7	334.88	G91HN-N	0.788
79	711.97	92.082	726.03	87.034	T92HN-N	0.896
80	815.64	88.174	821.27	91.125	Y93HN-N	0.713
81	811.99	82.155	838.47	79.13	N94HN-N	0.782
82	785.53	90.846	783.98	88.65	V95HN-N	0.810
83	774.2	91.032	748.82	90.938	T96HN-N	0.000
84	767.26	102.13	785.23	92.922	V97HN-N	0.405
85	733.75	106.14	733.61	95.926	Y98HN-N	0.790
86	756.17	85.056	787.59	81.813	S99HN-N	0.584
87	699.82	98.824	805.67	90.804	T100HN-N	0.761
88	674.2	94.265	782.42	98.01	G102HN-N	0.864
89	725.81	87.958	764.47	86.67	T103HN-N	0.739
90	716.84	91.595	778.76	85.989	I105HN-N	0.820
91	719.97	100.58	699.15	99.245	N107HN-N	0.604
92	765.25	106.77	703.31	154.11	K108HN-N	0.751
93	724	90.667	768.04	95.75	A109HN-N	0.535
94	806.56	103.34	809.52	94.853	L110HN-N	0.825
95	798.3	91.049	798.36	85.639	D111HN-N	0.960
96	724.64	62.368	760.31		L112HN-N	0.912
97	677.85	63.61			R113HN-N	0.676
98	744.37	78.992	758.56	85.269	I114HN-N	0.699
99	704.65	88.718	719.82	89.665	E116HN-N	0.542
100	630.8	87.177	730.56	89.64		
102	665.39	94.255	721.68	85.861		
103	698.62	64.005	764.71	65.41		
105	770.59	71.733	803.06	64.337		
107	687.64	100.12	789.58			
108	772.38	81.534	821.33	74.826		
109	840	107.16	851.56	99.362		
110	819.64	77.175	808.67	74.755		
111	772.16	103.39	796.26	109.39		
112	714.61	101.9	738.18	95.764		
113	738.05	98.581	768.59	92.436		
114	755.57	120.33	780.21	103.52		
115	765.51	103.07	790.87	101.3		
116	820.31	145.98	830.57	147.3		

**Figure 5.11 T2 relaxation spectra of CaM-CD2-III-5G:** (a) The spectrum was collected with T1 relaxation delay of 10 ms. (b) the spectrum was collected with T1 relaxation delay of 130 ms. These spectra were collected using 600 MHz NMR at 25 °C. The protein sample was prepared in 20 mM PIPES 20 mM KCl at pH 6.8.

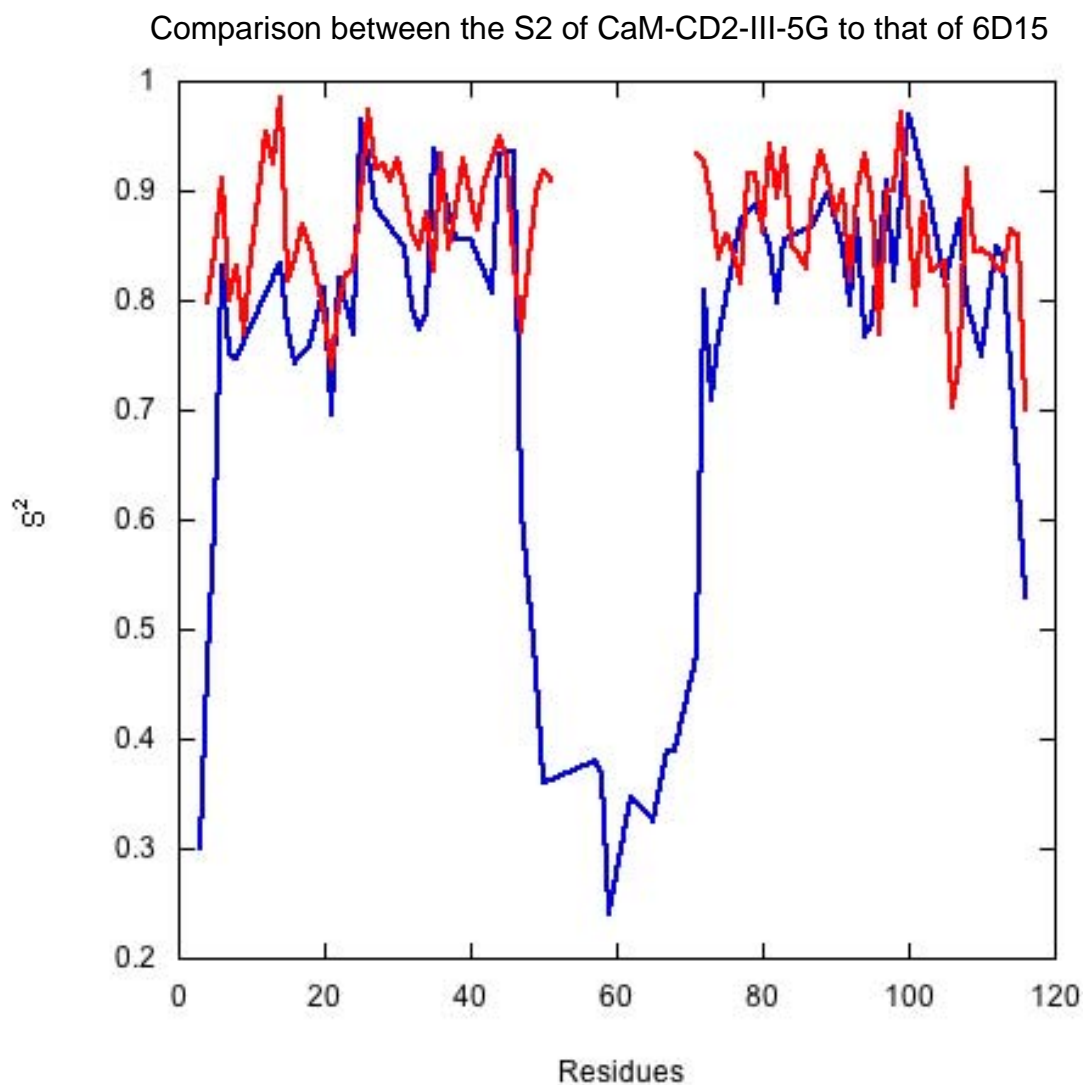


**Figure 5.12 NOE on and off spectra of CaM-CD2-III-5G:** The NOE off and on spectra are shown in blue and red, respectively. Some of the HN-N crosspeaks are not observed in the off spectrum. These conformation of these residues are likely to be very dynamic and the intensities of these residues are consider as value of 0.



**Figure 5.13 S<sup>2</sup> ordered parameters for CaM-CD2-III-5G:** The S<sup>2</sup> ordered parameters for CaM-CD2-III-5G were calculated using ModelFree software. The inserted EF-loop III has smaller S<sup>2</sup> values than the CD2 host protein. This data is indicating that the inserted EF-loop has greater mobility than the host protein.



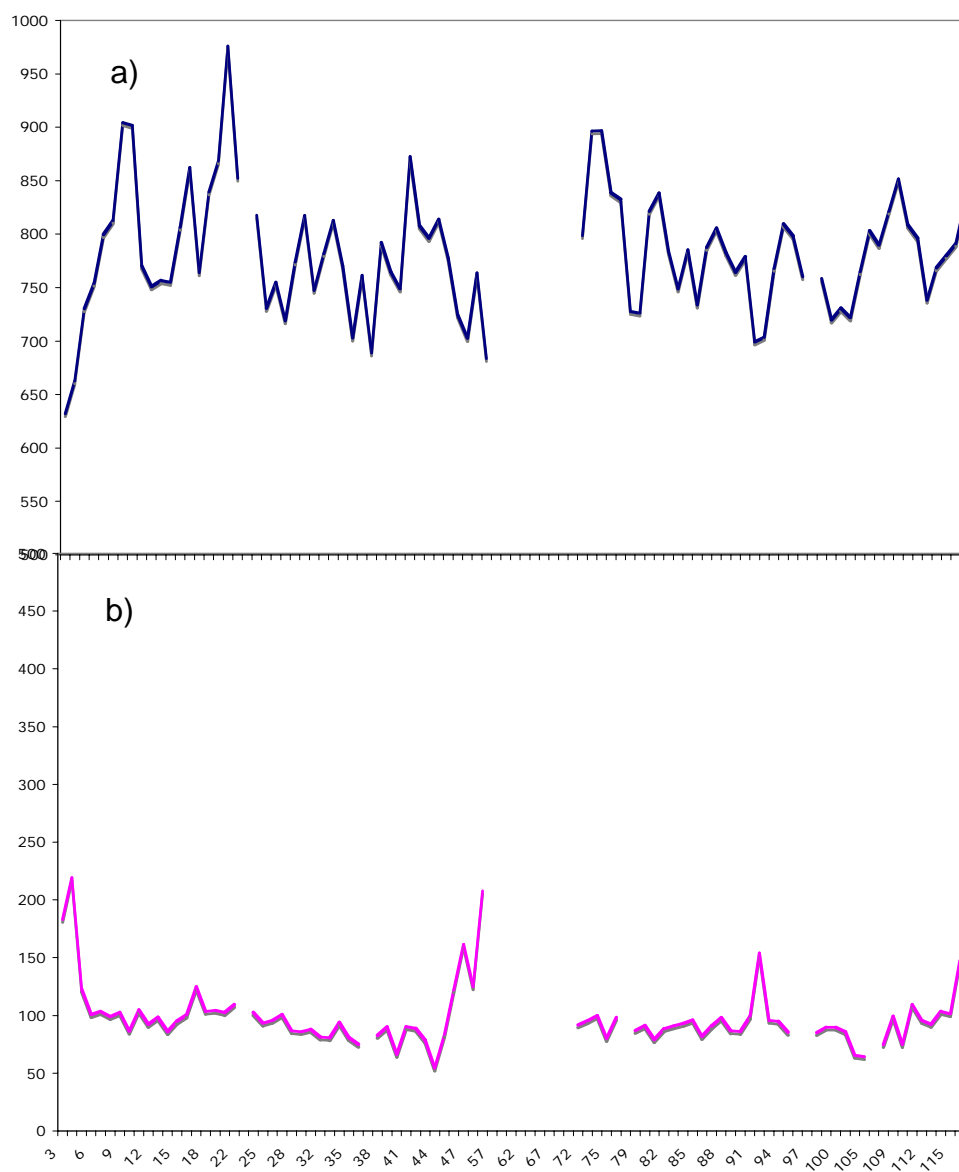


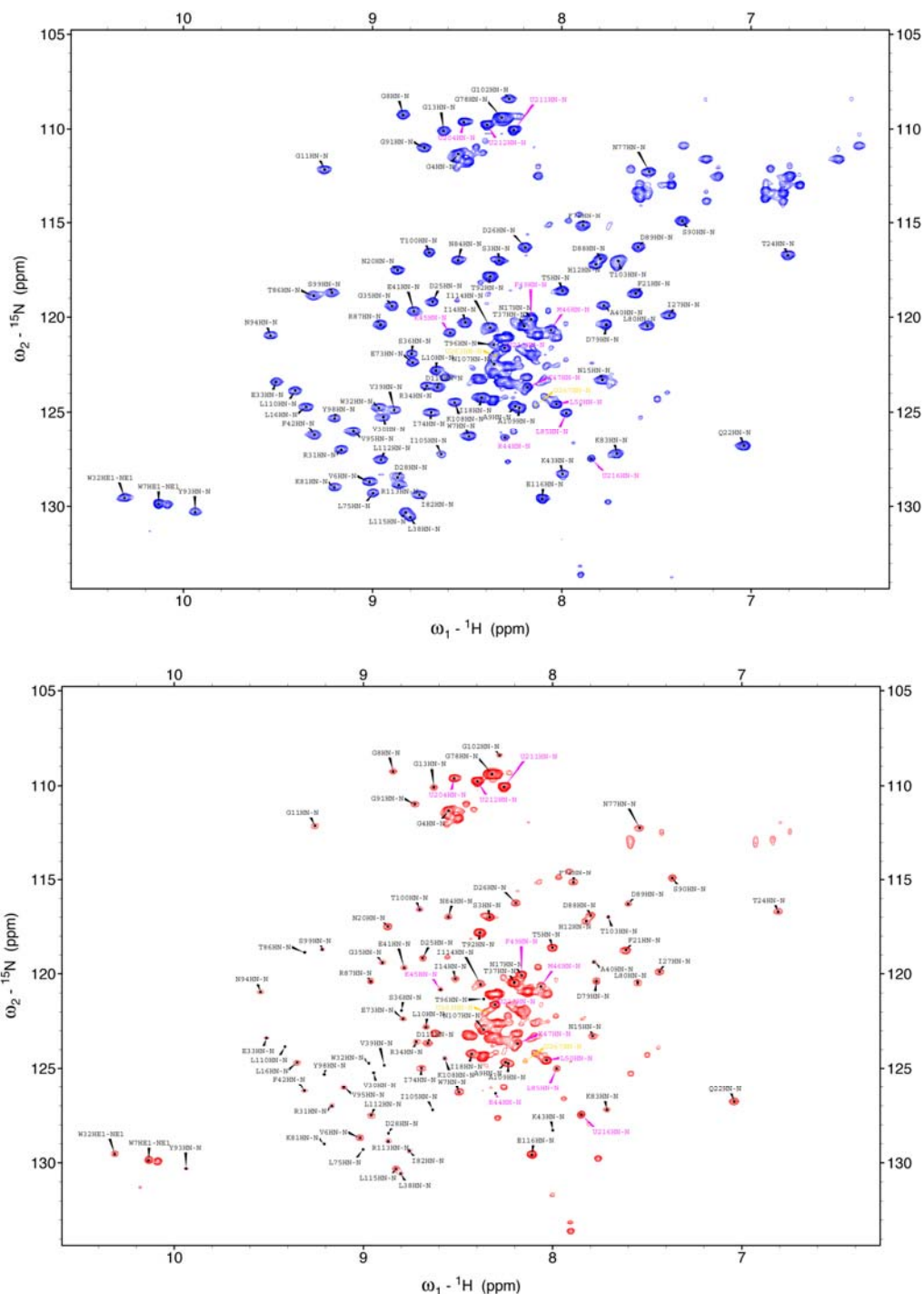
**Figure 5.14 S<sup>2</sup> ordered parameters comparison between CaM-CD2-III-5G and 6D15:** Comparing the S<sup>2</sup> ordered parameters of CaM-CD2-III-5G to between the S<sup>2</sup> of CaM-CD2-III-5G to the corresponding residues in 6D15.

Table 5.4 Comparing the T1 and T2 of CaM-CD2-III-5G to CaM

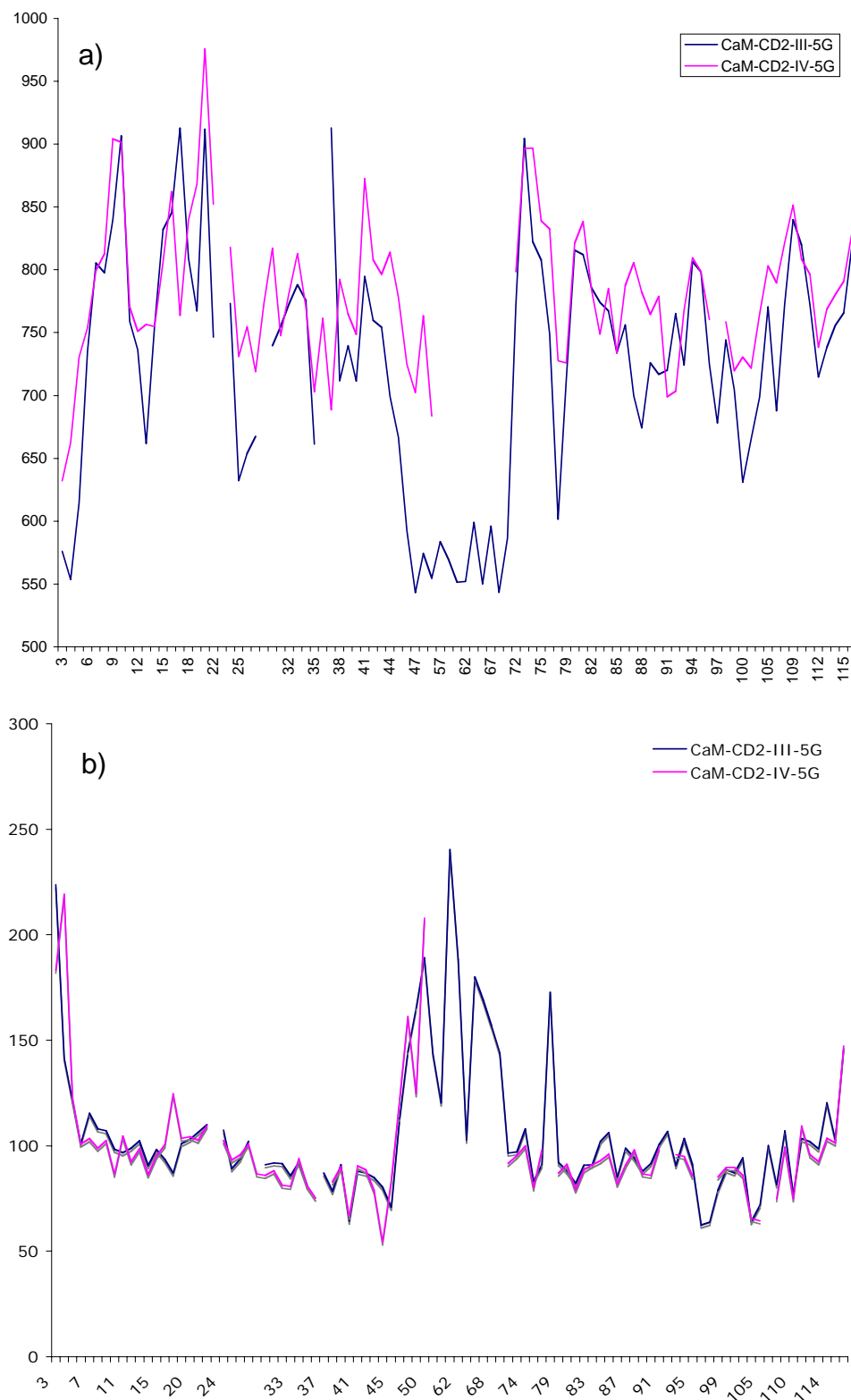
Pos	T <sub>1</sub> CaM	T <sub>1</sub> 5g3	T <sub>2</sub> CaM	T <sub>2</sub> 5g3
1 D	750		77	143
2 K	719	584	73	120
3 D	645	569	76	240
4 G	648	551	99	
5 N	656	552	107	
6 G				
7 Y	631	552	68	188
8 I		599		
9 S	640		79	
10 A	669	550	80	180
11 A	662		78	
12 E	670	596	81	169

**Figure 5.15 T1 relaxation spectra of CaM-CD2-IV-5G:** (a) The spectrum was collected with T1 relaxation delay of 0 ms. (b) the spectrum was collected with T1 relaxation delay of 1500 ms. These spectra were collected using 600 MHz NMR at 25 °C. The protein sample was prepared in 20 mM PIPES 20 mM KCl at pH 6.8.

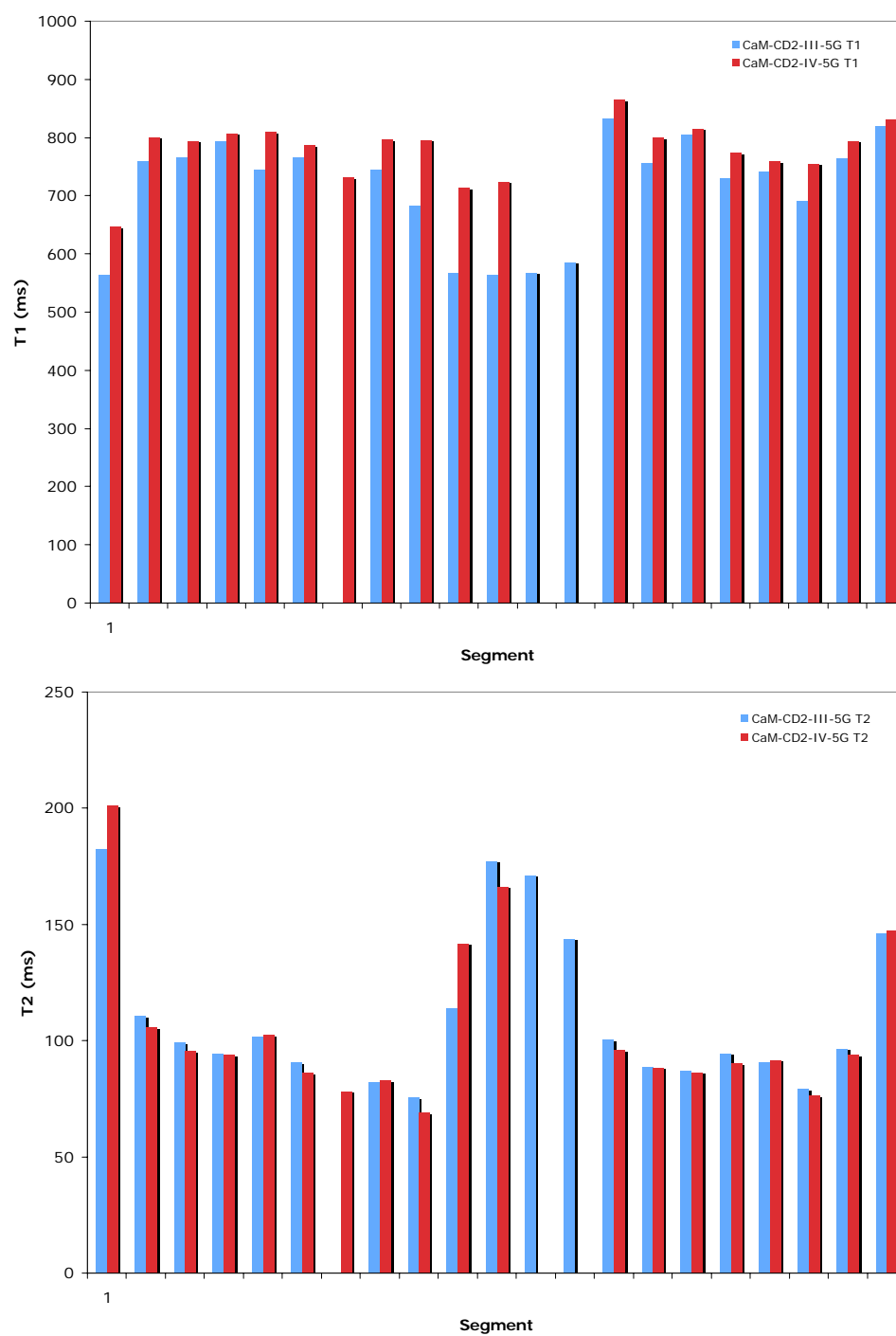




**Figure 5.17 T2 relaxation spectra of CaM-CD2-IV-5G:** (a) The spectrum was collected with T1 relaxation delay of 10 ms. (b) the spectrum was collected with T1 relaxation delay of 130 ms. These spectra were collected using 600 MHz NMR at 25 °C. The protein sample was prepared in 20 mM PIPES 20 mM KCl at pH 6.8.



**Figure 5.18 T1 and T2 values of CD2 variants:** a) T1 of CaM-CD2-III-5G and CaM-CD2-IV-5G. b) T2 of CaM-CD2-III-5G and CaM-CD2-IV-5G.



**Figure 5.19 Section Comparison between the secondary structure of CaM-CD2-III-5G and CaM-CD2-IV-5G.**

Table 5.5 Segment Secondary Structure Comparison Between  
CaM-CD2-III-5G and CaM-CD2-IV-5G

Structure	Residue #	Loop III T1 (ms)	Loop III T2 (ms)	Loop IV T1 (ms)	Loop IV T2 (ms)
N-ter	3 to 4	564	182.67	647.23	201.15
A	5 to 9	758.62	110.56	800.21	105.7
T-AB	10 to 13	765.94	99.06	794.82	95.51
B	14-16	793.77	94.33	807.87	93.91
T-BC	17-28	745.17	101.63	809.6	102.33
C	29-34	766.21	90.46	786.01	86
T-CC'	35-36			732.25	78.1
C'	37-43	745.17	82.27	797.14	82.881
T-C'C"1	44-45	682.92	75.56	795.9	68.97
T-C'C"2	46-48	567.57	113.81	713.49	141.64
C"	49-50	564.43	177.18	723.61	166.28
loop3	57-68	568.1	171.17		
T-G_D	70-71	586.07	143.88		
D	72-74	833.07	100.53	863.99	95.68
T-DE	75-79	756.17	88.77	799.16	88.47
E	80-82	804.39	87.06	814.57	86.3
T-EF	83-90	731.01	94.63	773.32	90.38
F	91-99	742.62	90.47	759.11	91.74
T-FG	100-105	691.35	79.29	755	76.31
G	106-115	763.75	96.62	793.92	93.92
C-ter	116	820.31	145.98	830.57	147.3



## Chapter 6.0 Determining the Oligomeric States of CD2 Variants

### 6.1 Introduction

The majority of EF-hand protein structures deposited in the protein data bank have at least paired EF-hand motifs. Two closely packed helix-loop-helix modules formed within a single globular domain constitute the basic calcium-binding unit of EF-hand proteins. Two EF-hand motifs arranged with respect to each other in a pseudo-2-fold symmetry in the same protein domain yield highly cooperative calcium binding systems (21). The distance between two calcium ions in two paired EF-hand motifs is usually about 11 Å, and the coordination shells of two calcium binding motifs can completely overlap in almost all EF-hand proteins. It is believed that EF-hand proteins use the paired EF-hand calcium binding motifs to regulate many cellular functions such as muscle contraction, neuronal signaling, apoptosis, and cell cycle control (8, 21, 53-55). Whether an isolated EF-hand motif can function as an individual unit has been a hot debate for several decades. Peptide models have been used extensively to understand the mechanism of metal binding and calcium-induced conformational change of EF-hand proteins. Previous work carried out by Sykes and colleagues has shown that peptide fragments, encompassing the EF-hand motifs III and IV with flanking helices from troponin C (TnC), associate in solution with a native like structure of the C-terminal domain of the protein (50). Linse's group have also shown that the peptide fragments from calbindinD9k encompassing flanking helices dimerized in

the presence of Ca(II) (52). Hydrophobic residues on the flanking helices of the EF-hand motif were shown to be essential for the dimerization (168). In contrast, the shorter peptide fragment composed of only the 12-residue EF-loop III of TnC remained monomeric both in the presence and absence of metal ions (169).

Bierzynski and colleagues have shown that the isolated 12-residue peptide with both ends blocked did not dimerize in the presence of calcium but dimerized to form a native-like structure in the presence of Ln(III), which is a metal with similar ionic radius and coordination properties as Ca(II) (40, 170). Based on their observation, they concluded that local interactions between the EF-hand calcium binding loops alone could be responsible for the observed cooperativity of calcium binding to the EF-hand protein domains. Whether an isolated EF-loop in solution and at physiological conditions can form an unpaired monomer remains questionable. A mechanism for calcium-induced conformational change is still not clear.

Pulsed-field-gradient NMR (PFG NMR) has been proven to be a valuable technique for the study of molecular motions and the effective dimensions of native, unfolded, and oligomeric states of proteins in solution (126, 171-177). The translational motion of well-packed spherical-like molecules in solution has a direct correlation with the hydrodynamic radius and molecular size according to the equation:  $D = K_B T / 6\pi a \eta$  (127, 128). At a given temperature (T) and solvent viscosity ( $\eta$ ), the diffusion constant (D) of this particular molecule decreases as the radius (a) of the molecule increases. This method is extremely useful for the

characterization of the oligomeric states of proteins in solution. Since most well packed globular proteins usually can be well described as a spherical shape, the molecular size of a protein and the oligomeric state of the protein can be estimated by measuring the diffusion constant (128, 178). The PFG NMR technique is advantageous for our studies since the diffusion constant can be measured using the same conditions as those for monitoring metal binding or structural determination (179-181).

In this chapter, we would like to establish whether an isolated EF-hand motif can stabilize as a monomer in solution. First, we will describe our investigations of the oligomeric states of the CD2 variants grafted with the EF-loop III of calmodulin in the presence and absence of Ca(II) and La(III) using pulsed-field gradient (PFG) diffusion NMR (section 6.3). Second, we will introduce our studies on the contribution of flanking F and EF helices to the dimerization of the EF-hand motif using CD2 variants grafted with the EF-loop flanked by F and EF helices. Third, we will report our efforts in examining the contribution of the conserved hydrophobic residues on the flanking helices to the conformation and dimerization of the EF-hand motif (Figure 6.1).

## 6.2 Determining the Oligomeric State of an Isolated EF-hand Loop

### 6.2.1 The Grafted EF-loop III in CD2 Remains Unpaired in the Absence of Metal Ions

By monitoring the fractional changes of chemical shifts at several positions in 1D  $^1\text{H}$  NMR and Trp fluorescence signal changes at several different wavelengths, the metal binding affinities for CD2 variants with the inserted EF-loop III of calmodulin were determined (112). CaM-CD2-III-5G, with two linkers, has the highest calcium binding affinity with a  $K_d$  of  $1.86 \times 10^{-4}$  M, while CaM-CD2-III and CaM-CD2-III-3G have one order of magnitude weaker calcium binding affinities with  $K_d$ s  $\geq 1$  mM. CaM-CD2-III-5G also has the highest La(III) binding affinity with a  $K_d$  of 58  $\mu\text{M}$  that is about 3-fold stronger than the calcium binding, which is possibly due to lanthanum's higher charge number. CaM-CD2-III and CaM-CD2-III-3G again exhibit one order lower binding affinities with  $K_d$   $\geq 400$  mM. All of the affinity values are in good agreement with those obtained by CD and fluorescence studies (112). That our measured calcium affinity of CaM-CD2-III-5G is weaker than the average value of the C-terminal domain of calmodulin is likely due to the lack of cooperative interactions between paired EF-hand motifs. As pointed out in several papers, the formation of pairs of EF hand motifs will lower the energy state of the calcium-binding form and increase the stability resulting from a mutual polarization of the associated groups. The binding of the first metal ion will enhance the binding of the second metal ion (48, 182). Isolated EF-hand motifs lack such stabilization, and, hence, its calcium

binding affinity is not increased. Our recent work also shows that the measured metal affinity is not affected by differing protein environments (135). This suggests that the calcium binding affinity we measured for the isolated EF-hand motif reflects its intrinsic calcium affinity.

The oligomeric states of the CD2 variants with the inserted EF-loop in the absence of metal ions were investigated by the pulsed-field-gradient diffusion NMR experiment. Amino acid Gly and proteins such as lysozyme (14.3 KDa), trypsinogen (24 KDa), and carbonic anhydrase (28 KDa) with different molecular weights were used as molecular standards to evaluate the accuracy of the PFG NMR method and to optimize experimental conditions. To minimize the contribution of other factors such as salt concentrations, protein concentrations, and temperature, all of the PFG-gradient diffusion experiments were carried out at identical conditions in 10 mM Tris-10 mM KCl, pH 7.4 at 25 °C. Signals from Tris buffer or dioxane were used as internal references to eliminate viscosity effects and to normalize diffusion constants of different NMR samples. Figure 6.2a shows the 1D  $^1\text{H}$  NMR signal decay of CaM-CD2-III-5G as a function of pulsed field gradient strength. Figure 6.2b shows the internal standards, dioxane, glycine and Tris buffer. All of the amplitudes of the NMR resonances gradually decrease with the increase of the gradient strength. NMR signals from the protein have a significantly slower decay than that of the small molecules dioxane and Tris buffer. The integrated area of NMR signals of the protein, dioxane and Tris buffer fit well to equations 1 and 2 with  $R > 0.999$  (Figure 6.3).

The obtained diffusion constants of Tris and dioxane are  $67.6$  and  $98.9 \times 10^{-7} \text{ cm}^2/\text{s}$ , respectively. Our measured diffusion constant for hen egg-white lysozyme is  $10.6 \times 10^{-7} \text{ cm}^2/\text{s}$ , which is almost identical to that ( $10.8 \times 10^{-7} \text{ cm}^2/\text{s}$ ) reported by Altieri *et al.* (128). Under identical condition, the diffusion constant of wild type CD2 is  $11.0 \times 10^{-7} \text{ cm}^2/\text{s}$ , which is very close to the value of lysozyme that has a similar molecular weight (14.3 KDa). As expected from the Stokes-Einstein equation, larger proteins, such as trypsinogen (24 KDa) and carbonic anhydrase (28 KDa) have significantly smaller diffusion constants of  $9.5$  and  $9.1 \times 10^{-7} \text{ cm}^2/\text{s}$ , respectively (Table 6.1). Wilkins *et al.* reported that the hydrodynamic radius of dioxane is  $2.12 \text{ \AA}$  (171). Using dioxane as a reference of size, the effective hydrodynamic radii of CD2, lysozyme, trypsinogen, and carbonic anhydrase are  $39.2$ ,  $40.2$ ,  $45.4$ , and  $47.4 \text{ \AA}$ , respectively. As shown in Table 6.1, the relative sizes of these proteins are consistent with their relative molecular weights.

The oligomeric states of three CD2 variants with the insertion of EF-hand loop III from calmodulin were investigated using NMR under the same conditions as for wild type CD2. As shown in Figure 6.3, increasing the gradient strength gradually attenuates NMR signals from three CD2 variants in a manner similar to that of wild type at pH 7.4 in 10 mM Tris, 10 mM KCl. The obtained diffusion constants of three proteins, CaM-CD2-III-0G, CaM-CD2-III-3G and CaM-CD2-III-5G, are  $10.9$ ,  $11.1$ , and  $11.1 \times 10^{-7} \text{ cm}^2/\text{s}$ , respectively, which are very close to that of wild type CD2 (Table 6.1). In addition, the linewidth of the resolved 1D

NMR resonances from aromatic residues W7 at 10.5 ppm and V95 at 0.3 ppm for the CD2 variants with the inserted EF-loop do not broaden noticeably compared to those of wild type CD2 and lysozyme. Further, increasing the CD2 variants' concentrations from 0.11 to 0.80 mM does not result in a detectable increase in their linewidth at pH 7.4 in 10 mM Tris-10 mM KCl. The strong similarities in the diffusion constants and the linewidth of the CD2 variants to that of wild type CD2 indicate that the loop insertion into CD2 does not cause dimerization of the protein in the absence of metal ions.

### **6.2.2 The Grafted EF-loop III Remains Unpaired Upon Metal Binding**

To examine whether the isolated EF-loop III of calmodulin in CD2 dimerizes upon metal binding, the diffusion constants of the engineered proteins CaM-CD2-III-3G and CaM-CD2-III-5G have been measured in the presence of La(III) under physiological conditions using the pulsed-field-gradient diffusion NMR experiment. At 1 mM La(III) concentration, the CD2 variants are expected to be predominantly metal-loaded. The obtained diffusion constants for CaM-CD2-III-3G and CaM-CD2-III-5G are  $11.4 \times 10^{-7} \text{ cm}^2/\text{s}$ . The corresponding hydrodynamic radius is 37.8 Å. These results are very surprising, since they are close to the corresponding values in the absence of metal ions with a slight increase of about 3%, which is within experimental error.

The NMR linewidth of the resolved NMR resonances from aromatic residues W7 at 10.5 ppm and V95 at 0.3 ppm of wild type CD2 and its variants

were further measured in the presence of 1 mM La(III). No dramatic increase in the linewidth was observed with the addition of metal ions. All linewidths are in the range from 9.3 Hz to 14.8 Hz, which is similar to those of metal-free proteins. Given the diffusion constant results, we conclude that the engineered proteins with the insertion of an EF-loop from calmodulin are monomeric in the absence and presence of metal ions.

### 6.2.3 Discussion

As shown in the Stokes-Einstein equation, any factors that affect the viscosity and temperature will introduce variations in the diffusion constants. The measured diffusion constants are likely affected by sample conditions, such as salt concentrations, protein concentrations, and/or buffer conditions. We have shown that these effects can be normalized/eliminated by using the relative values to the internal references dioxane and Tris. For example, lysozyme is known to be a monomer at protein concentrations up to 1.5 mM (177). Baseline correction was applied to further overcome the baseline distortion. Using the two internal references, the measured diffusion constants and hydrodynamic radii of this protein remain the same from protein concentrations 0.1 to 1.2 mM (data not shown). The measured diffusion constants for the protein lysozyme ( $10.6 \times 10^{-7} \text{ cm}^2/\text{s}$ ) agree with the reported value ( $10.8 \times 10^{-7} \text{ cm}^2/\text{s}$ ) (128). Chen et al. reported that chemical exchange of the “bound” and “free” molecules could sometimes lead to a strong signal distortion (178). In our cases, we have not



observed any signal distortions due to chemical exchanges of calcium binding and calcium free forms. The error values in Table 6.1 are based on three different measurements of the same sample. We have shown that the error values of all the experiments with CD2 variants are  $< 3\%$ . These data clearly demonstrate that the PFG NMR method with the internal reference has the accuracy and reproducibility to distinguish oligomeric states of proteins under physiological conditions.

The radius and diffusion constant of a dimerized protein can be estimated by treating it as a protein with doubled molecular weight. Teller *et al.* used the Stokes-Einstein equation to calculate the diffusion constant ( $D$ ) after oligomerization for various geometries (183). They suggested that the expected ratio of  $D_{\text{dimer}} : D_{\text{monomer}}$  is 0.75 upon dimerization (183). A further study completed by Altieri *et al.* on the dimerization of ubiquitin has shown that the ratio of  $D_{\text{dimer}} : D_{\text{monomer}}$  is 0.72 (128), which agrees with the monomer-dimer described above. Therefore, a globular well-folded protein with doubled molecular weight will increase its radius by 27% if the protein is assumed to be a well-packed hard sphere. Thus, the diffusion constant ( $D$ ) is expected to be reduced by about 23%. The diffusion constants for all of the CD2 variants with the inserted EF-loop III of calmodulin and with different numbers of Gly linkers are the same as that of wild type within experimental error ( $< 3\%$ ). Further, these proteins do not have any decreased diffusion constants upon binding of La(III) (Table 6.1). This

suggests that the engineered proteins do not dimerize either in the absence or presence of metal ions.

Bierzynski *et al.* have shown that the isolated 12-residue peptide encompassing EF-loop III of calmodulin exhibits line broadening for some NMR resonances at 50 °C (40, 170) in the presence of high concentrations of La(III). Such broadening of NMR peaks can be fit by assuming the peptide dimerized at high temperature. To support their hypothesis that the EF-loop dimerizes at high temperature and retains native EF-loop structure upon La(III) binding, they carried out another experiment using a frozen solution mixed with the EF-loops loaded with La(III) and Ho(III). Based on the observation of the energy transfer from Ho(III) to La (III) and the line broadening of some of the resonances at 50 °C, they concluded that local hydrophobic interactions between the EF-hand calcium binding loops at loop positions 7 and 8 alone could be responsible for the observed cooperativity of calcium binding to EF-hand protein domains. The relatively strong La(III) binding affinity (6.7 mM) was attributed to the cooperative binding of the formed dimer.

Our PFG-NMR experiments carried out at physiological conditions provide direct evidence of the size of the protein upon metal binding. All three proteins containing isolated EF-loop III of calmodulin remain monomeric with and without the bound metal. In contrast to their conclusion, our data unambiguously support our conclusion that the isolated EF-loop III is stable in solution without interacting with another EF-hand motif. Our observed monomeric form of EF-loop III of

calmodulin in an isolated environment agrees with previous work of the 12-residue peptide of this loop EF-loop by Borin *et al.* (62, 130). They did not observe any dimerization of the 12-residue peptide fragment of the calcium binding loop III of calmodulin both in the presence and absence of Ca(III) or Tb(III). Therefore, additional factors that reside outside of the EF-loop III may contribute to the pairing of EF-hand motifs of calmodulin.

### **6.3 NMR Structural Studies on CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF**

To investigate the contribution of the helices of the EF-hand motif to the metal binding and oligomerization, two engineered proteins with different number of helices were generated. CaM-CD2-III-5G-F contains a helix at C-terminal end of the EF-loop and CaM-CD2-III-5G-EF contains a helix at both N-terminal and C-terminal end of the EF-loop (Figure 6.1). The insertion procedure for these two engineered proteins was similar to CaM-CD2-III-5G. Two glycine linkers were used to connect the host protein with the EF-hand element.

#### **6.3.1 Conformational Analysis on Engineered Calcium Binding Protein**

##### **CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF**

The secondary structures of the CD2 variants, monitored by far UV CD, are shown in Figure 6.4 (These studies were carried out by Dr. Yiming Ye). The addition of EF-helices into an all  $\beta$ -sheet protein results in a far UV CD spectrum with two negative maximums at 208 and 222 nm, which is a CD

spectrum similar to that of an  $\alpha$ -helical structure. To investigate the conformation of the host protein with the insertion of different EF-hand elements, we have used 1D  $^1\text{H}$  NMR to identify the secondary and tertiary structures of the CD2 variants (Figure 6.5). The amide regions and the sidechain regions of CaM-CD2-III-5G-F have a close similarity with CaM-CD2-III-5G suggesting that these modified proteins are well folded and the native structure of CD2 is retained after the insertion of the helix and the loop. The resonances of CaM-CD2-III-5G-EF in the amide region and side chain region have similar chemical shift positions as CaM-CD2-III-F. It is interesting to point out, while the chemical shift values of the Trp aromatic proton are different than wild type, but they are similar to the chemical shift values of CaM-CD2-III-5G. Since the loop-helix motif and helix-loop-helix motifs were inserted into CD2 at same location as CaM-CD2-III-5G, the chemical shift values suggest the integrity and packing of CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF are similar to CaM-CD2-III-5G but less compact. Further, the linewidth of the protein EF is significant greater than that of wild type CD2 and CaM-CD2-III-5G.

To examine whether lower temperatures will stabilize conformation, we have collected  $^1\text{H}$  NMR spectra of CaM-CD2-III-5G-EF at 10, 15, 20, and 25  $^{\circ}\text{C}$  as shown in Figure 6.6. The dispersive signal of the methyl proton for V114 in Figure 6.6 remains unaltered in all temperatures. This data suggest that the folding and packing of CaM-CD2-III-5G-EF is stable in the temperature range from 10 to 25  $^{\circ}\text{C}$ . The signal to noise ratio is slightly higher at 25  $^{\circ}\text{C}$  in

comparison to the lower temperature spectra. At temperatures  $\geq 25$  °C the NMR sample becomes less stable and soluble.

The organic solvent, 2,2,2-trifluoroethanol (TFE), has the ability to induce secondary structure formation for peptide fragments that have high secondary structure propensity. It also destabilizes proteins by altering none covalent interactions. Therefore, we examined the effect of TFE on the conformation of CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF. The TFE titration experiments were carried out by gradually adding 0, 5, 10, 15, 18, and 20 % of solution into the NMR sample tube. As shown in Figures 6.7 and 6.8 (for CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF, respectively); the addition of 5% TFE did not alter the dispersive resonances in the amide and sidechain region of the spectra. Unfortunately, increased TFE above 5% (v/v) resulted in excessive precipitates to appear in the NMR sample, which caused the majority of dispersive resonances to diminish. The limited solubility of our NMR samples prevented further detailed NMR study.

The effect of Ca(II) on the conformation of CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF were examined by NMR in the same manner as CaM-CD2-III-5G. As shown in Figures 6.9 and 6.10, 1D spectra of CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF do not exhibit detectable changes upon addition of Ca(II) from 0 to 5.72 mM Ca(II). The sidechain region of the Asn and Gln (6.8 to 7.2 ppm) showed the largest changes during the calcium titration and the resonances at the other regions remained similar to the calcium free form spectrum. EDTA has

been shown to have strong affinities for metal ions, and it is commonly used as a chelating agent to decrease background interference from metal ions. Figure 6.11 shows that addition of 1 mM EDTA initially to the samples resulted in small changes of the resonances between 6.85 to 7.05 ppm and 7.50 to 8.00 ppm for CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF. These changes suggest that the residues at these location are binding with Ca(II) from the buffer solutions. Nevertheless, Ca(II) induced changes were largely not observed for both CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF partly due to the limited resolution of 1D NMR.

### 6.3.2 Oligomeric Studies with CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF

The oligomeric state of the engineered calcium binding proteins were first investigated using the PFG Diffusion NMR experiments under the same conditions for CaM-CD2-III-5G in the presence and absence of Ca(II). The diffusion experiments were carried out using 200  $\mu$ M of CaM-CD2-III-5G-F and 130  $\mu$ M of CaM-CD2-III-5G-EF in the absence of Ca(II) (1 mM EGTA) in pH 7.4 Tris buffer at 25 °C. The integrated areas of NMR signals of the protein (CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF), dioxane, and Tris buffer were fitted to equations 6.1 and 6.2 (Figure 6.12). In the absence of metal ions, the diffusion coefficient for CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF were calculated as  $11.7 \pm 0.3$  and  $11.1 \pm 0.2 \times 10^{-7} \text{ cm}^2/\text{s}$ , respectively. The differences in the diffusion constants of the CD2 variants with the EF-helices inserted were within 5% of CaM-CD2-III-5G ( $11.1 \pm 0.2 \times 10^{-7} \text{ cm}^2/\text{s}$ ). As shown in Table 6.1, the

molecular weight of the CaM-CD2-III-5G, CaM-CD2-III-5G-F, and CaM-CD2-III-5G-EF were 12.5, 13.5, and 15.1 KDa, respectively. The molecular weight of CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF are 8 and 18 % larger than CaM-CD2-III-5G, respectively. The diffusion coefficients of the CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF are in good agreement with CaM-CD2-III-5G. These results suggest that the CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF remain in monomeric form in the absence of metal ions.

To further investigate the oligomeric state of the grafted EF-hand motif in CD2, we have carried out analytical ultracentrifugation analyses at different protein concentrations with different rotor speeds in collaboration with Dr. Gary Pielak at UNC. Figure 6.13a shows that the EF-loop III grafted into CD2 (CaM-CD2-III-5G) has a molecular weight of 14-15 KDa in the presence of 1 mM EGTA at three protein concentrations (50, 65, and 104  $\mu$ M) at 21,000, 25,000, and 37,000 rpm in 20 mM PIPES, pH 6.8 at 4 C° for 20 hours. The addition of calcium (10 mM) does not change the apparent molecular weight, which corresponds to the calculated molecular weight in the monomeric state. This result agrees with our previously published conclusion using pulsed-field NMR (184). On the other hand, the molecular weight of CaM-CD2-III-5G-EF at 23-65  $\mu$ M protein concentrations in the presence of 1 mM EGTA is 18.18 and 15.8 KDa at 21,000, 25,000, and 37,000 rpm respectively. With increased rotor speed, a slight decrease in the apparent molecular weight was observed, which suggests the existence of multiple species. In the absence of Ca(II), this protein has an

apparent molecular weight that is greater than the calculated molecular weight of 15.2 KDa similar to the peptide fragment of EF2 encompassing the EF-hand motif 2 of calbindin<sub>D9k</sub>. It was attributed to be the extended conformation due to electrostatic repulsion or a mixed monomer and dimer conformation. In the presence of 10 mM Ca(II), the measured apparent molecular weight of this protein increases 26-27 KDa when the protein concentration increases from 10 to 26  $\mu$ M. These results clearly demonstrate that the attachment of flanking helices to the EF-loop results in dimerization in the presence of Ca(II). Table 6.1 summarized the single species analysis of sedimentation equilibrium data and PFG of CD2 variants

The dimer formation of CaM-CD2-III-5G-EF in the presence of Ca(II) is consistent with the linewidth of the protein observed in NMR. As shown in Figure 6.13b, the linewidth of 1D  $^1\text{H}$  NMR signal of CaM-CD2-III-5G-EF is significantly increased upon addition of Ca(II) while the linewidth of neither CD2 nor CaM-CD2-III-5G experiences a noticeable change.

The conformation analysis indicates that the integrity and packing of the host protein frame is maintained after the insertion of the helix F and helices EF from calmodulin for CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF, respectively. The EDTA studies with CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF show that both of the engineered proteins interact with the background calcium in the buffer solutions. This is an indication that CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF have strong metal binding affinities for calcium. The PFG diffusion



analysis suggests that CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF remain in monomeric forms in the absence of calcium. However, the line broadening effect on the  $^1\text{H}$  NMR spectrum and the analytical ultracentrifugation results suggest that CaM-CD2-III-5G-EF forms a mixture of monomer and dimer in the presence of 10 mM calcium.

#### 6.4 Understanding the Contribution of the Helices to Dimerization

To determine the possible intra-motif and inter-motif helices interactions, the CSU (Contacts of Structural Units) software was used to analyze the interatomic contacts in the crystal structure of calcium bound calmodulin (3CLN) (46) (Figure 6.14). The x-axis in Figure 6.14 represents the sequence of helix-loop-helix of site III of CaM. The y-axis is representing the sequence of site III and site IV of CaM. The value inside the intercept cell between the two axes is the lowest contact distance between two atoms from two different residues. Since crystal structure does not contain coordinates for the hydrogen atoms, the majority of the distances from CSU are between two different carbon atoms. Some of the atoms from residues at positions -1, -4, and -8 of the entering helix of site III are less than 5 Å away from the exiting helix of site IV (especially position 17). Some of the atoms from residues at positions 13, 16, and 17 of the exiting helix of site III are less than 5 Å away from the exiting helix of site III. There are some atoms from residues at positions 13, 14, and 17 of the entering helix of site III that are less than 5 Å away from the exiting helix of site IV. The

close proximity between the residues described here are likely the candidates for hydrophobic interaction or hydrogen bonding.

Many peptide fragments encompassing EF-hand motifs with different lengths were shown to be dimers and this dimerization was proposed to enhance Ca(II) binding affinity (185). Sykes and colleagues have shown that a 34 residue synthetic peptide encompassing EF-hand motif III of troponin C was first shown to form a dimer in the presence of Ca(II) with a head-to-tail structure that resembles closely the native pairing of two coupled EF-hand motifs (50). The shorter peptide fragment composed of only the 12 residue EF-loop III of troponin C remains in monomeric form both in the presence and absence of metal ions.

Hydrophobic residues on the flanking helices of the EF-hand motif were shown to be essential for the dimerization. For TnC, the replacement of hydrophobic residues L98 and F102 in the E-helix or I121 and L122 (position xx and yy, respectively) at the F-helix by Ala has little effect on the formation of helix but resulted in 100 and 300 fold decrease in Ca(II) affinity (168). Simultaneously replacing hydrophobic residues at both flanking helices decreases the Ca(II) binding affinity by 1000-fold (168). This suggests that the hydrophobic residues on the flanking helices are essential for the stabilization and dimerization of the Ca(II) binding loop and its Ca(II) binding affinity. Linse's group have shown that peptide fragments containing EF-hand motifs 1 and 2 of calbindinD9k have different abilities to dimerize in the absence of calcium (52). Both motifs dimerize upon Ca(II) binding with a strong preference for heterodimer formation over

homodimer formation. Both peptide fragments of the EF-hand motifs have similar Ca(II) binding affinities in the range of  $\lg K=4.6-5.3$  (186). Using a phage display library, they further showed that EF-hand motif III prefers heterodimer formation with EF-hand motif IV of CaM in the presence of Ca(II) (187).

Based on our studies and those of others, we hypothesize that the hydrophobic residues at the flanking helices of the EF-hand motif play an essential role for the dimerization and pairing of the EF-hand motifs.

#### **6.4.1 Proposed Modification to the Flanking Helices of Site III in CaM**

To test our hypothesis, the hydrophobic residues on the entering and exiting flanking helices were removed. The CSU analysis on the site III of CaM indicated that residues I(-8), F(-4), V(16), and M(17) are possibly involved in interaction with the other EF-hand motif (site IV). The modification on the CaM-CD2-III-5G-EF protein is shown in Figure 6.15. First, Ile at position -8 was changed to a Lys and Phe at position -4 was changed to Glu. Lys and Glu residues have the ability to form salt bridges. Second, a Ser was added to the beginning of the entering helix to provide N-capping. Third, Val and Met at positions 16 and 17 were both changed to Ala. Ala has a small sidechain with high helix forming propensity. The modified engineered protein is denoted as CaM-CD2-III-5G-EF-SKEAA (SKEAA). These mutations were made by Dr. Yiming Ye. The proteins were expressed and purified by Yubin Zhou.

#### **6.4.2 Conformational Analysis on SKEAA using 1D $^1\text{H}$ NMR**

The 1D  $^1\text{H}$  NMR spectrum of SKEAA is shown in Figure 6.16 along with the  $^1\text{H}$  spectra of CaM-CD2-III-5G and CaM-CD2-III-5G-EF. The comparison between the spectra of SKEAA and CaM-CD2-III-5G-EF has shown that the resonances observed between 6.8 to 8.8 ppm exhibit large changes. The resonances at this region are usually from the backbone amide protons and the sidechain of Asn and Gln. This difference between SKEAA and CaM-CD2-III-5G-EF likely due to the mutations on the helices, which causes changes to the local environment of the inserted EF-hand motif and the insertion location of the host protein. On the other hand, the aromatic ring protons of Trp and the methyl protons of Val and Leu for SKEAA were observed at similar chemical shifts as those of CaM-CD2-III-5G and CaM-CD2-III-5G-EF, which indicates that the mutations on the helices of the engineered protein do not alter the conformation of the CD2 host protein.

#### **6.4.3 2D NMR Structural Studies on the Engineered Calcium Binding Protein, SKEAA**

The sequential assignment for the host protein region of SKEAA was completed using TOCSY and NOESY experiments (Figure 6.17). The assignment was verified and compared with CaM-CD2-III-5G. Most of the backbone resonances from T5 to F42 and I94 to E136 were assigned for SKEAA. The SKEAA construct has 20 more residues than CaM-CD2-III-5G.

There are more crosspeaks on the fingerprint region of the TOCSY spectrum of SKEAA. The conformation of the host protein region for the engineered protein was verified by comparing its properties to the host protein of CaM-CD2-III-5G (described in Section 5.1.1). The comparisons were carried out in three areas, the HN chemical shift comparison, the chemical shifts of the HE1 proton of W32, and the chemical shifts of the methyl protons of L16 and V115. First, the HN chemical shifts of SKEAA were plotted as a function of the same host protein residues in the CaM-CD2-III-5G (Figure 6.18). The differences between the HN chemical shifts of the host protein region of SKEAA to that of the CaM-CD2-III-5G are less than 0.50 ppm, which indicate no major change is observed between T5 to F42 and I94 to E136.

Second, the HE1 protons of W7 and W32 for SKEAA are observed at 10.13 and 10.34 ppm, respectively. The lower field chemical shifts observed for HE1 proton of W32 for SKEAA suggests that this proton is packed inside of the hydrophobic environment. These observations in SKEAA are similar to those in CaM-CD2-III-5G. Third, the sidechain methyl protons of L16 and V95 of SKEAA are observed at similar chemical shifts as the corresponding protons in CaM-CD2-III-5G (Figure 6.19). The mainchain and sidechain protons of SKEAA are in good agreement with the same host proton residues of CaM-CD2-III-5G. These results indicate that the hydrophobic core and the tertiary structure of the host protein of SKEAA are not changed after the insertion of 37 residues.

The sequential assignment on the EF-hand motif III of SKEAA was very difficult due to the signal overlaps in the fingerprint region of the homonuclear TOCSY and NOESY spectra. The comparison between the TOCSY spectra of CaM-CD2-III-5G and SKEAA is shown in Figure 6.20. The HN-HA crosspeaks for D56, D58, G61, Y62, and A65 (positions 1, 3, 6, 7, and 10 of the EF-loop III, respectively) were assigned in the TOCSY spectrum of CaM-CD2-III-5G. The comparison between the fingerprint region of SKEAA to CaM-CD2-III-5G indicates that the HN-HA crosspeaks of EF-loop III residues at positions 1, 3, 6, 7, and 10 of SKEAA are shifted to different locations. This observation indicates that the presence of the EF helices has changed the structural properties of the EF-loop III of SKEAA.

#### **6.4.4 Studies on the Oligomeric State of SKEAA**

To determine if SKEAA will form the native pair-pair interactions after removal of the hydrophobic residues on the EF-helices, the oligomeric state of the engineered calcium binding protein, SKEAA, was investigated using the PFG diffusion NMR experiment. The diffusion experiments were performed using the same conditions as those for wild type CD2, CaM-CD2-III-5G, and CaM-CD2-III-5G-EF in the presence and absence metal ions. The diffusion experiments were carried out using 380  $\mu$ M of SKEAA in the presence of 1 mM EGTA, 1 mM Ca(II), and 1 mM (La(III) in pH 7.4 10 mM Tris 10 mM KCl buffer at 25 °C. The integrated areas of the NMR signals of SKEAA and Tris buffer were fit to

equations 2.4 and 2.5 (Figure 6.21). The diffusion coefficients of SKEAA in the presence of 1 mM EGTA, 1 mM Ca(II) and 1 mM La(III) were calculated as  $12.30 \pm 0.3$ ,  $12.36 \pm 0.2$ , and  $12.01 \pm 0.2 \times 10^{-7} \text{ cm}^2/\text{s}$ , respectively (Table 6.1). In the absence of metal ions, the diffusion coefficient for CaM-CD2-III-5G-EF was calculated as  $11.1 \pm 0.2 \times 10^{-7} \text{ cm}^2/\text{s}$ . The diffusion constants of SKEAA in all three conditions are slightly higher than the diffusion constant of CaM-CD2-III-5G-EF in the presence of 1 mM EGTA (within 8%). This observation is likely due to the existence of small portion of cleaved fragment. These results suggest that SKEAA remained in monomeric in the presence of 1 mM EGTA, 1 mM Ca(II), and 1 mM La(III). The mutations on the hydrophobic residues at positions -8, -4, 16, and 17 have prevented SKEAA to form pair-pair interactions with another molecule in the presence and absence of metal ions as observed in the intact calmodulin.

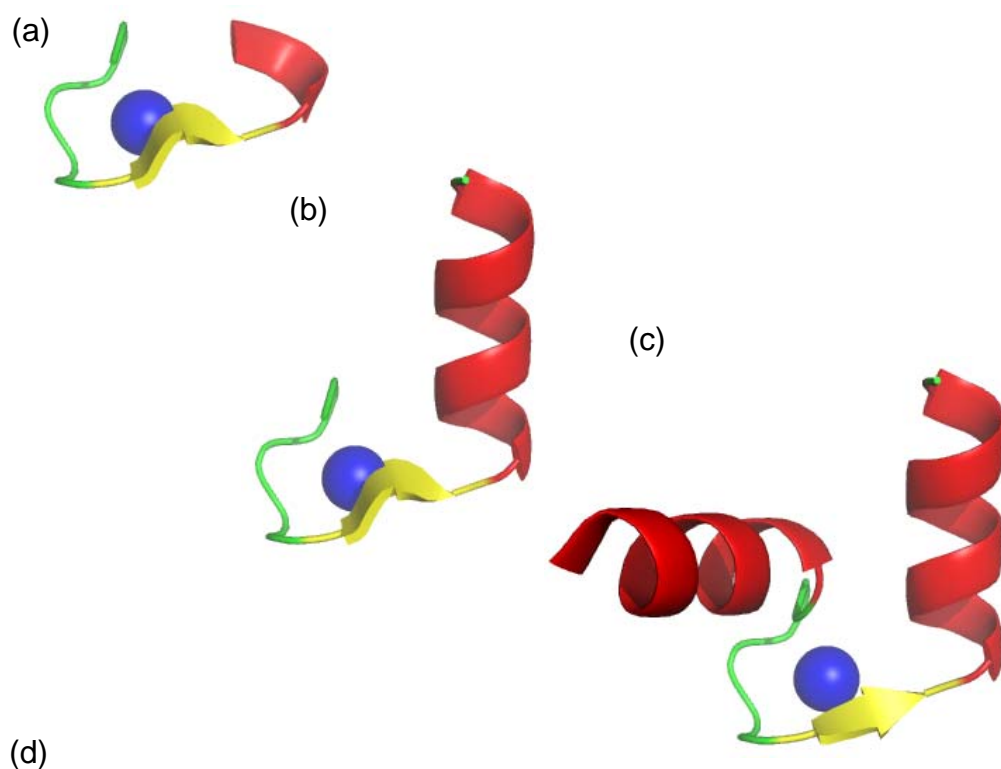
## 6.5 Major Finding of this Chapter

In contrast to the previously reported dimeric coupling of the 12-residue peptide encompassing the EF-loop III of calmodulin in the presence of La(III), our lab has shown that the same EF-loop grafted into CD2 remains monomeric both in the absence and presence of Ca(II) and La(III) using PFG diffusion NMR. The PFG diffusion studies were carried out using the same experimental conditions as those during the metal binding studies (184). The diffusion coefficient values of the CD2 variants are approximately  $11.1 \times 10^{-7} \text{ cm}^2/\text{s}$  both in the presence and

absence of metal ions, which are equivalent to those of wild type CD2. This suggests that the isolated EF-loop III of calmodulin inserted in the scaffold protein is able to bind calcium and lanthanum as a monomer, which contradicts the previous observation for the EF-hand motif. Our results imply that additional factors that reside outside of the EF-loop III may contribute to the pairing in EF-hand motifs of calmodulin. Also, using ultracentrifugation, we have further demonstrated that the isolated EF-loop in CD2 is monomeric, with or without calcium. The addition of flanking helices to the EF-loop III of CaM does not lead to a significant dimerization in the absence of Ca(II). Its apparent molecular weight is greater than the calculated molecular weight and the apparent molecular weight slightly increases with the protein concentration, which is an indication of low degree self-associating interaction. In addition, the apparent molecular weight of 18 KDa at a rotor speed of 21,000 rpm is decreased to 16 KDa at 37,000 rpm. All of these results suggest that CaM-CD2-III-5G-EF is primarily in monomeric form, but a small portion is in dimeric form in the absence of calcium. Strikingly, calcium binding converts CaM-CD2-III-5G-EF completely to the dimer form at low protein concentrations. The molar ellipticity of this protein at several wavelengths is unchanged from protein concentrations of 0.1 to 100  $\mu$ M suggesting that it remains as a dimer in this range. Removal of hydrophobic residues on the flanking helice leads to the protein as a monomer in the absence and presence of Ca(II) and La(III). Our systematic investigation using a grafting approach clearly suggests that the flanking helices play a pivotal



role in dimerization and enhances calcium binding affinity, suggesting that there is a strong coupling between calcium binding, folding, and association of the EF-hand motif.



CaM-CD2-III-0G

R1...S52-D-K-D-G-N-G-Y-I-S-A-A-E-G53...E99

CaM-CD2-III-3G

R1...S52-GGG-D-K-D-G-N-G-Y-I-S-A-A-E-G53...E99

CaM-CD2-III-5G

R1...S52-GGG-D-K-D-G-N-G-Y-I-S-A-A-E-GG-G53...E99

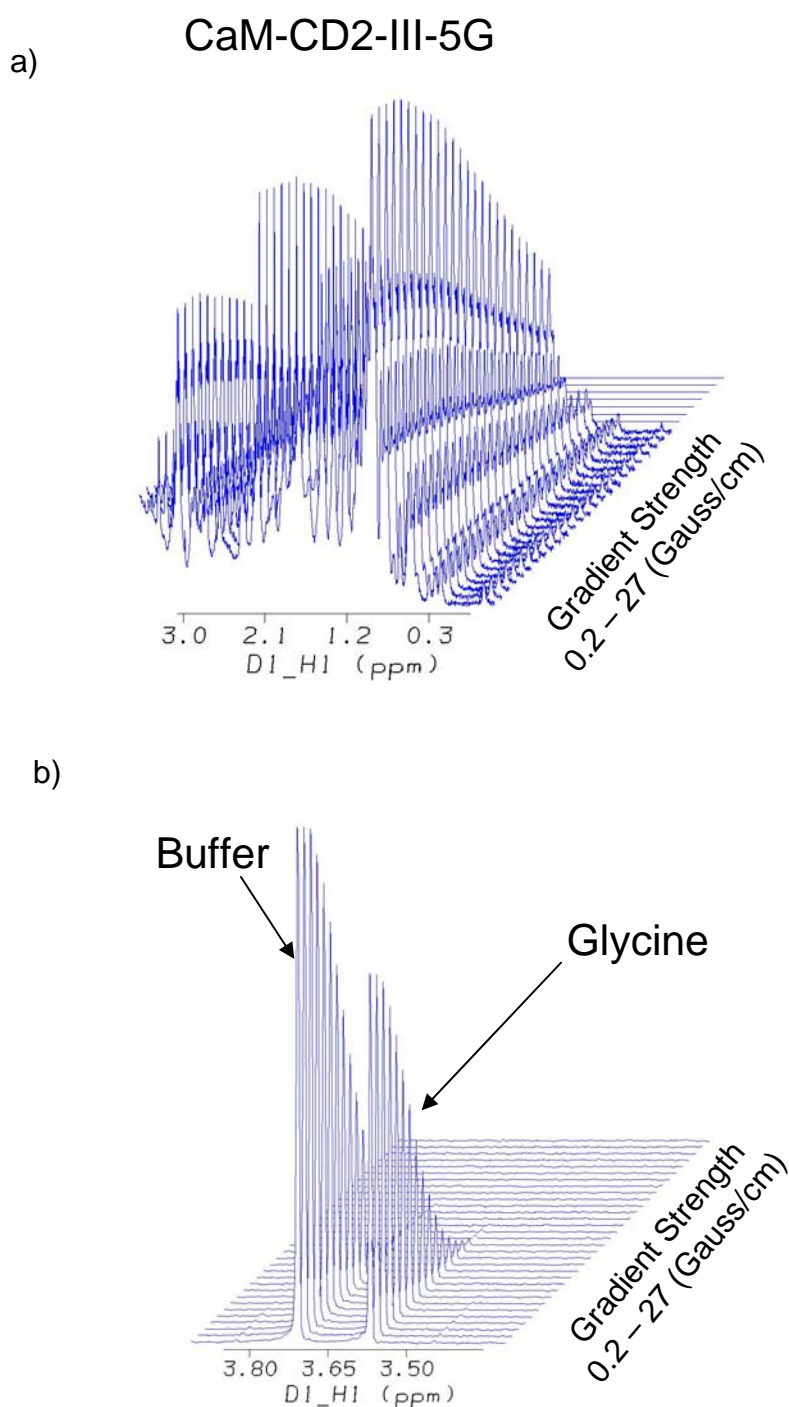
CaM-CD2-III-5G-F

R1...S52-GGG-D-K-D-G-N-G-Y-I-S-A-A-E-L-R-H-V-M-T-N-L-GG-G53...E99

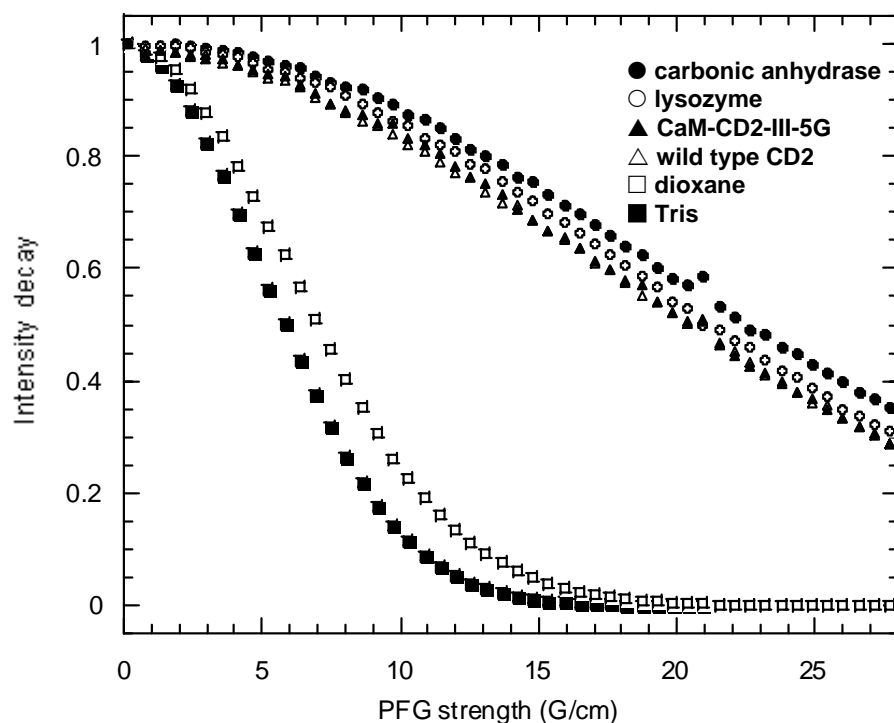
CaM-CD2-III-5g-EF

S52-GGG-E-E-E-I-R-E-A-F-R-V-F-D-K-D-G-N-G-Y-I-S-A-A-E-L-R-H-V-M-T-N-L-GG-G53

**Figure 6.1 Sequence details of CD2 variants with helices:** (a) The EF-loop III of calmodulin. (b) the EF-loop III with F helix of calmodulin. (c) the EF-loop III with E and F helices. (d) the insertion sequences of CD2 variants.



**Figure 6.2 Diffusion spectra of CaM-CD2-III-5G, glycine and buffer:** a) Diffusion spectra of CaM-CD2-III-5G. b) Diffusion spectrum of Tris buffer and glycine. The buffer and glycine resonances has smaller molecular weight, hence the signal decay faster than the protein as a function of gradient strength. These spectra were collected using 500 MHz NMR at 25 °C. The samples were prepared in 10 mM Tris 10 mM KCl buffer at pH7.4.



$A$  = intensity of chemical shifts

$\gamma$  = gyromagnetic ratio of proton

$\delta$  = PFG duration time (sec)

$G$  = gradient strength (Gauss/cm<sup>2</sup>)

$\Delta$  = time between PFG pulse (sec)

$D$  = diffusion constant

$$^*A = A_0 \exp [-(\gamma\delta G)^2 (\Delta - \delta/3)D]$$

**Figure 6.3 The data fitting curve for the diffusion constants:**

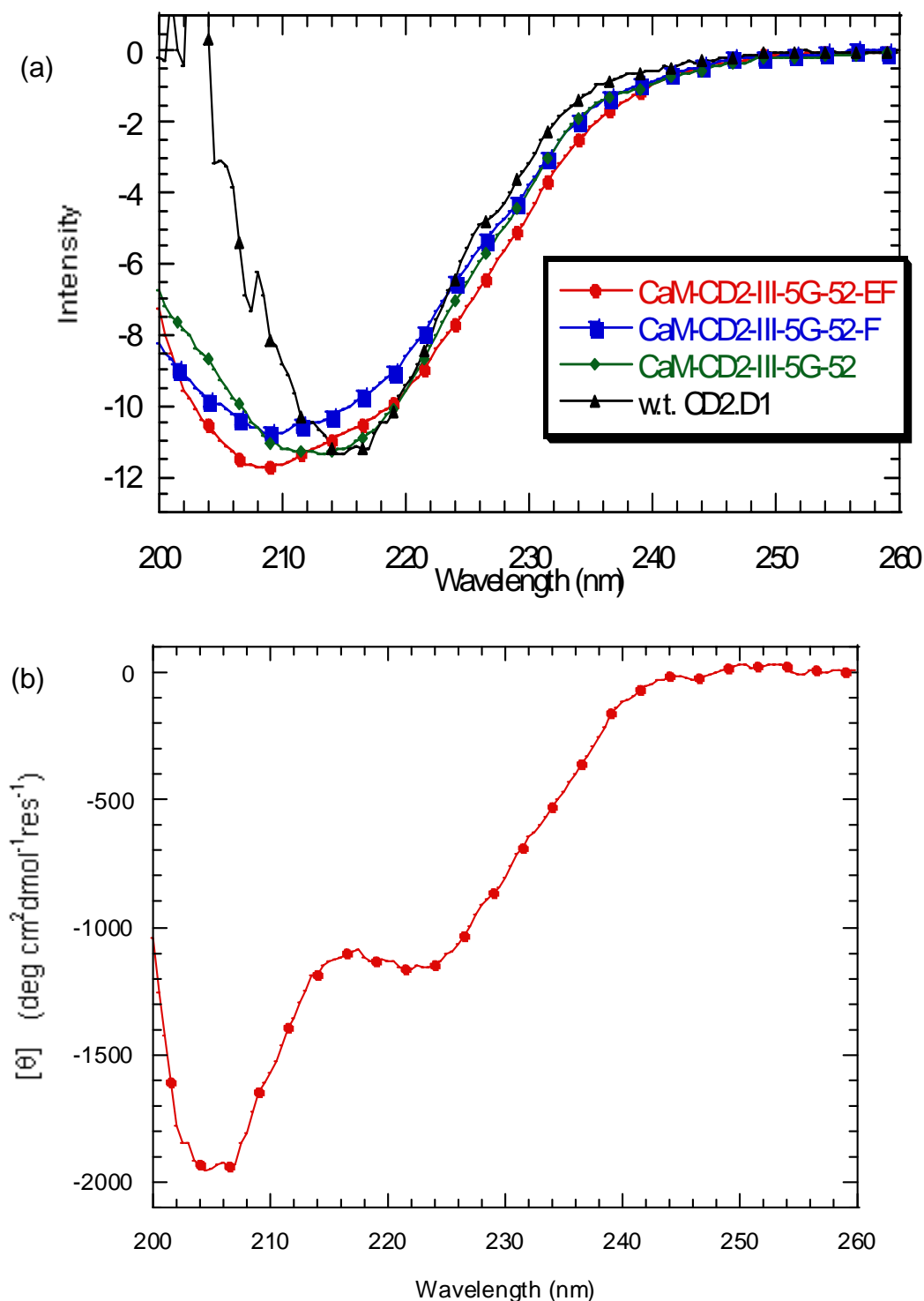
These diffusion experiments were all collected using 500 MHz NMR at 25 °C. The samples were prepared in 10 mM Tris 10 mM KCl buffer at pH7.4.

Table 6.1 Diffusion Constants of CD2 Variants

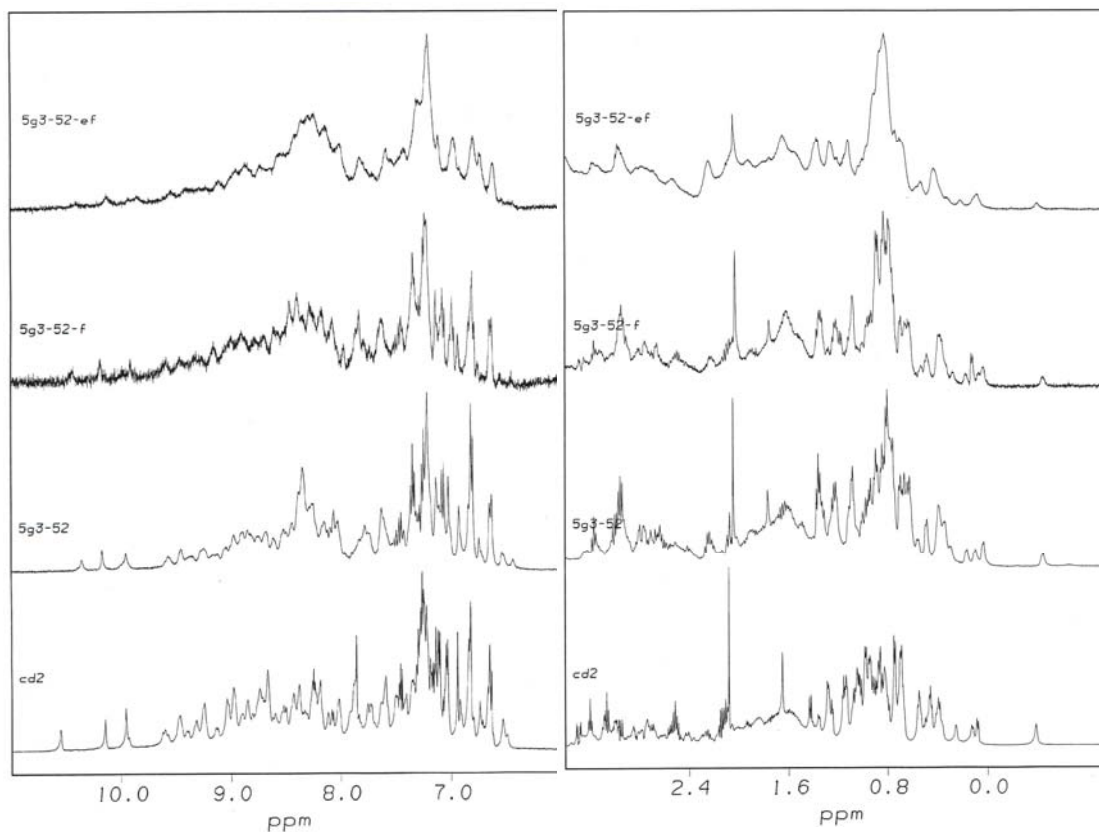
Molecule	D (x10 <sup>7</sup> cm <sup>2</sup> /s)
Dioxane	98.9 ± 0.9
Tris	67.6 ± 1.3
glycine	93.3 <sup>b</sup>
lysozyme	10.6 ± 0.2
CD2	11.0 ± 0.2
CaM-CD2-III-5G	11.1 ± 0.2
CaM-CD2-III-5G 1 mM La <sup>3+</sup>	11.4 ± 0.2
CaM-CD2-III-3G	11.1 ± 0.3
CaM-CD2-III-3G 1 mM La <sup>3+</sup>	11.4 ± 0.2
CaM-CD2-III	10.9 ± 0.2
CaM-CD2-III-5G-F	11.7 ± 0.2
CaM-CD2-III-5G-EF	11.1 ± 0.2
SKEAA	12.3 ± 0.2
SKEAA 1 mM Ca <sup>2+</sup>	12.4 ± 0.2
SKEAA 1 mM La <sup>3+</sup>	12.0 ± 0.2

<sup>a</sup>Jones et al. (1998), *J. of Biomol. NMR*, **10**, 199-203

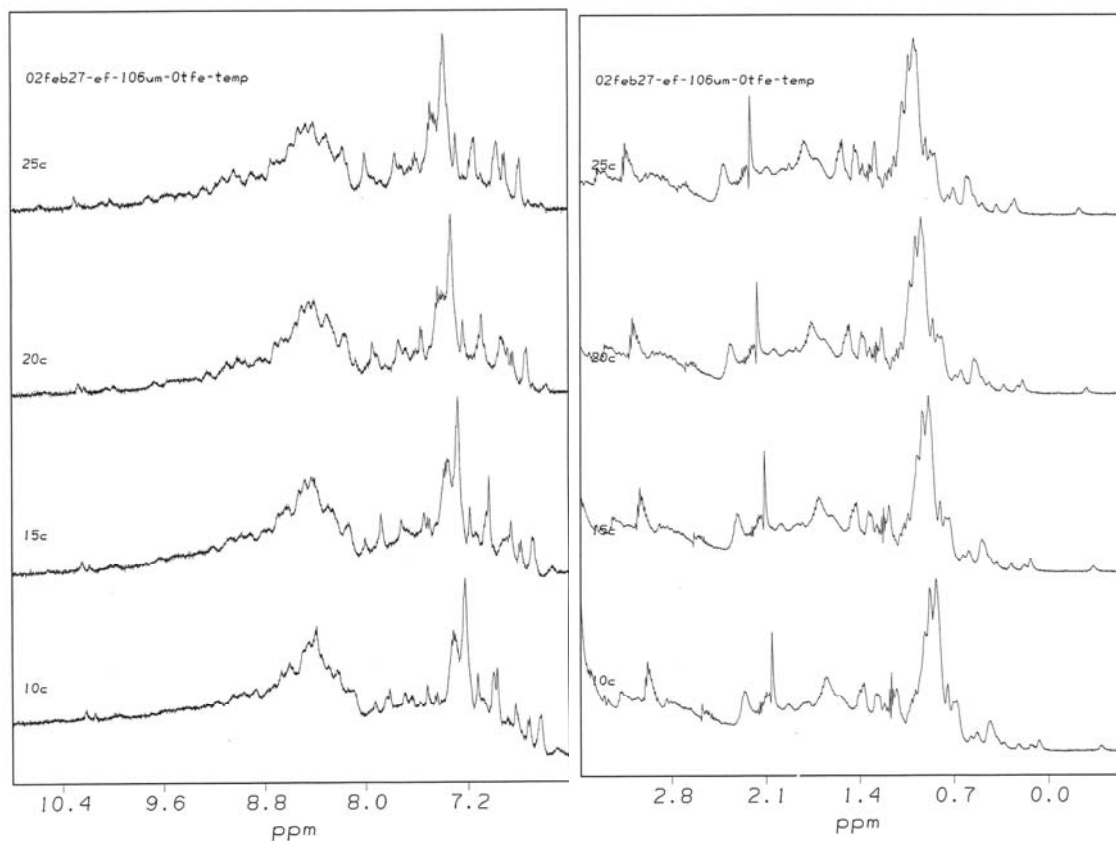
<sup>b</sup>20 °C, Altieri et al. (1995), *J. Am. Chem. Soc.*, **117**, 7566-7567



**Figure 6.4 CD spectra of CD2 variants:** (a) Far UV CD of 2.3 mM CaM-CD2-III-5G-52-F (■), CaM-CD2-III-5G-52-EF (●), CaM-CD2-III-5G-52 (◆) and w.t. CD2.D1 (▲) in 10 mM Tris-HCl-KCl, pH 6.9. (b) Far UV CD spectra of CaM-CD2-III-5G-52-EF substrate by CaM-CD2-III-5G-52 in 10 mM Tris-HCl-KCl, pH 6.9

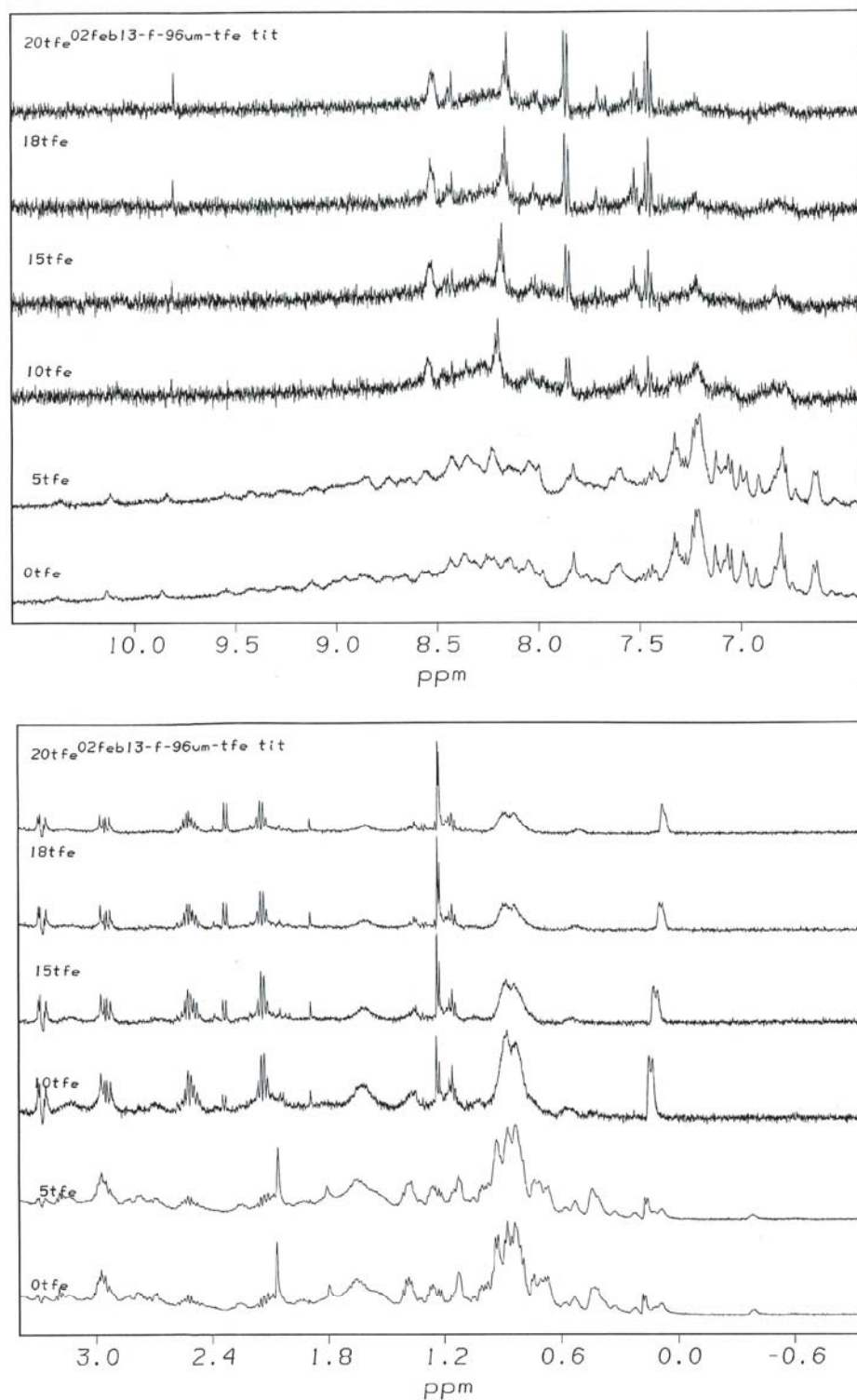


**Figure 6.5  $^1\text{H}$  NMR spectra of CD2, CaM-CD2-III-5G, CaM-CD2-III-5G-F, and CaM-CD2-III-5G-EF:** These spectra for CD2 variants were all recorded in pH 7.4 10 mM Tris 10 mM KCl at 25 °C using 600 MHz NMR.

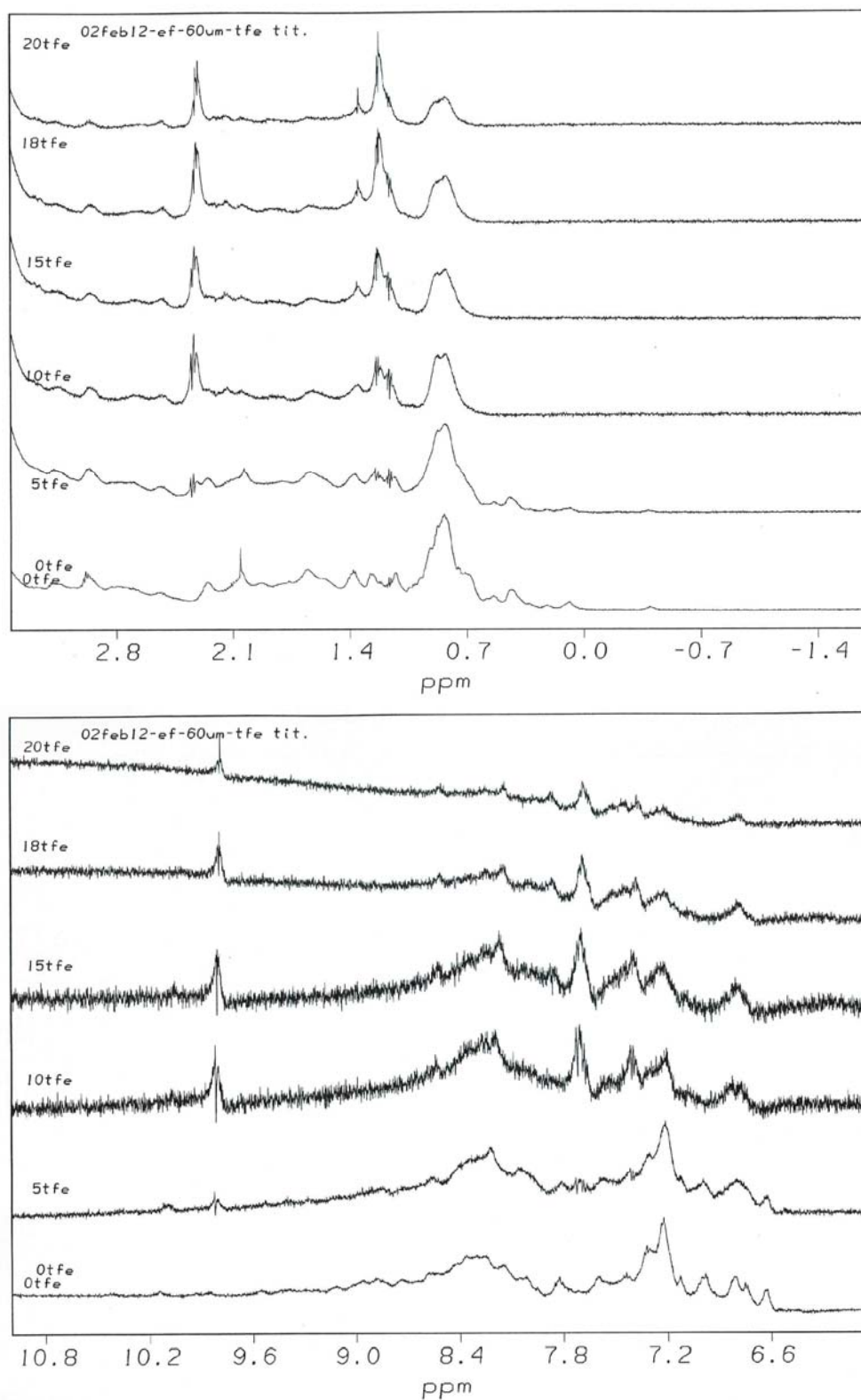


**Figure 6.6 Temperature studies on CaM-CD2-III-5G-EF:** The spectra collected at different temperature studies for CaM-CD2-III-5G-F were all recorded in pH 7.4 10 mM Tris 10 mM KCl using 600 MHz NMR.

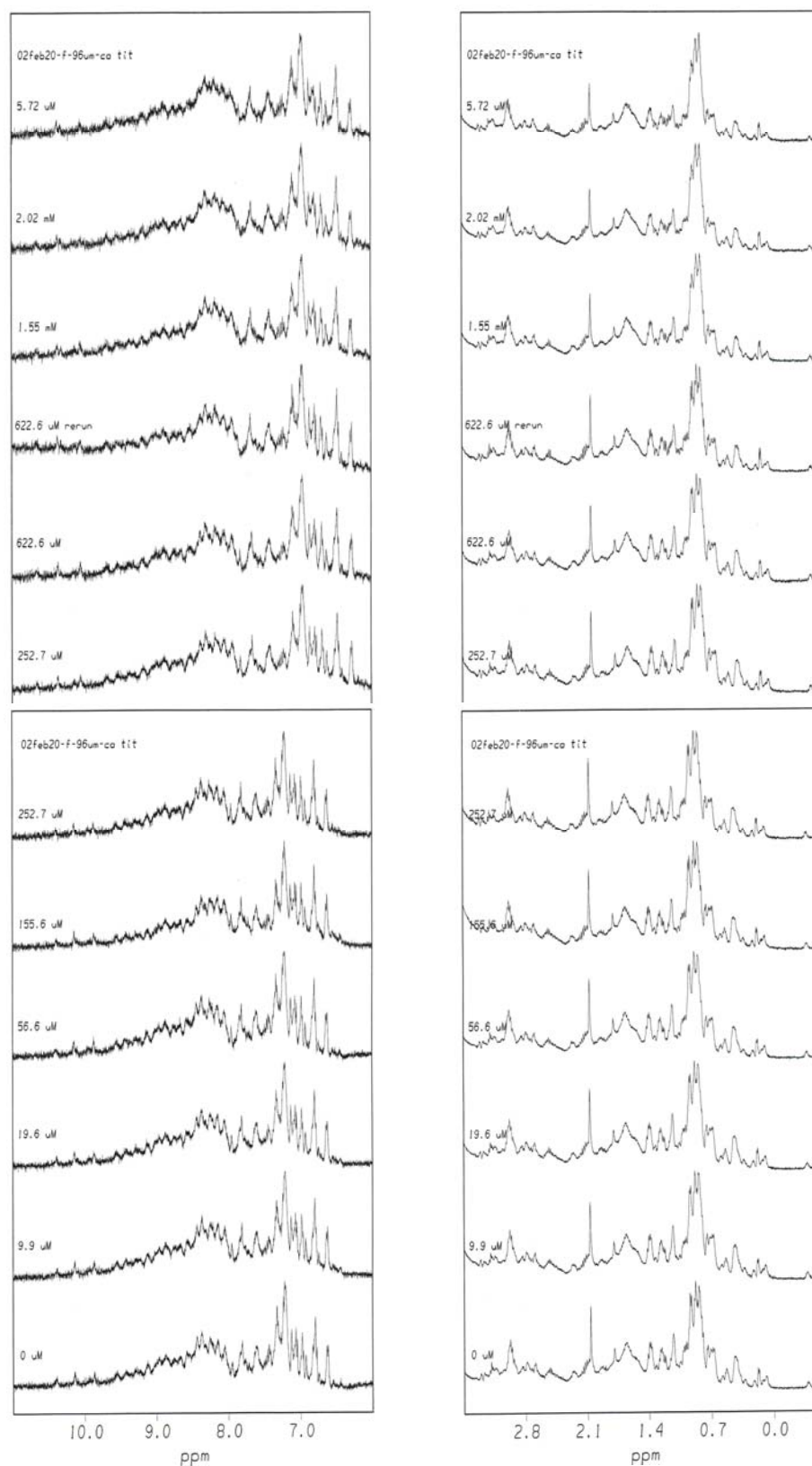




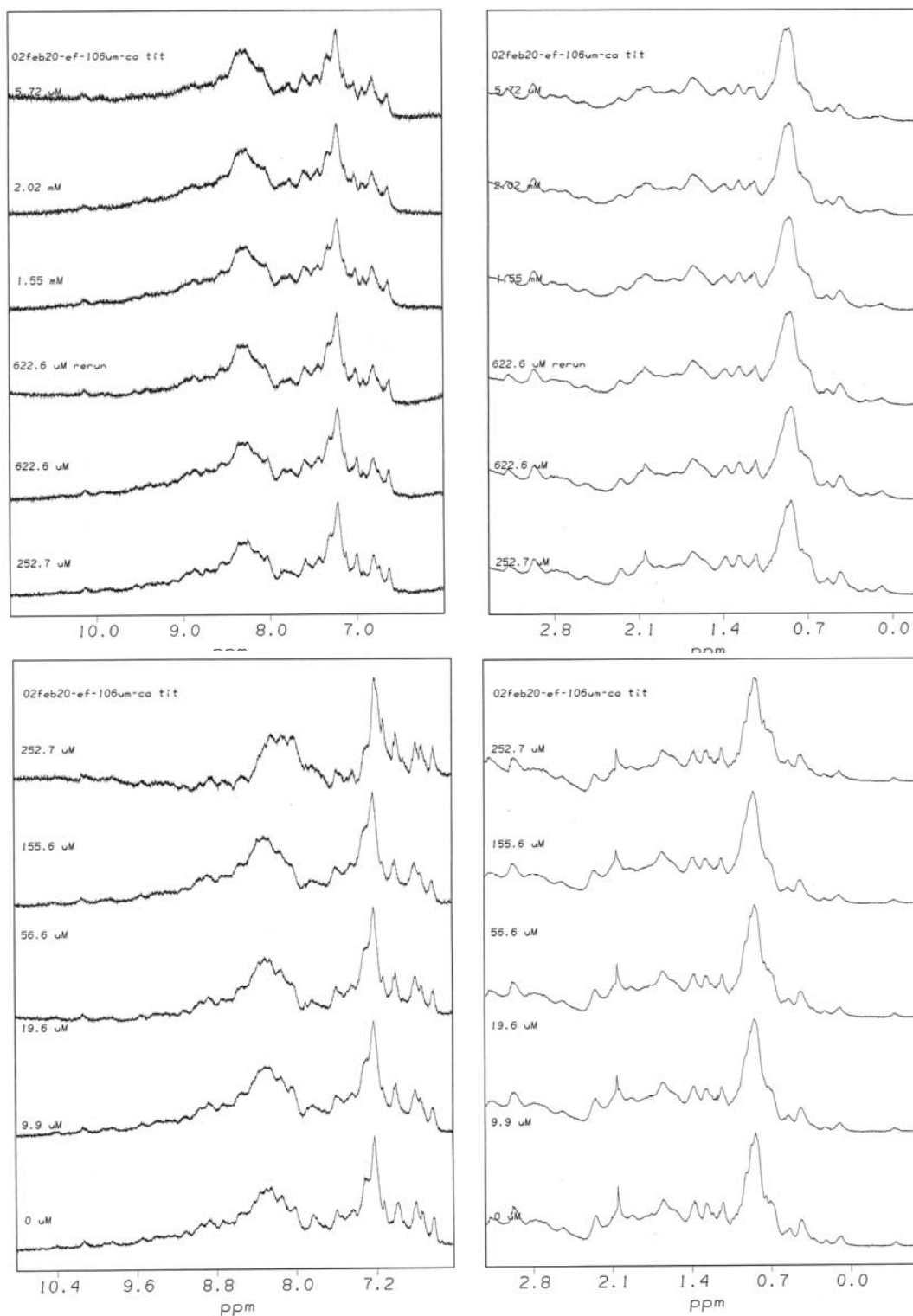
**Figure 6.7 TFE titration on CaM-CD2-III-5G-F:** The TFE titration spectra for CaM-CD2-III-5G-F were all recorded in 10 mM Tris 10 mM KCl (pH 7.4) at 25 °C using 600 MHz NMR.



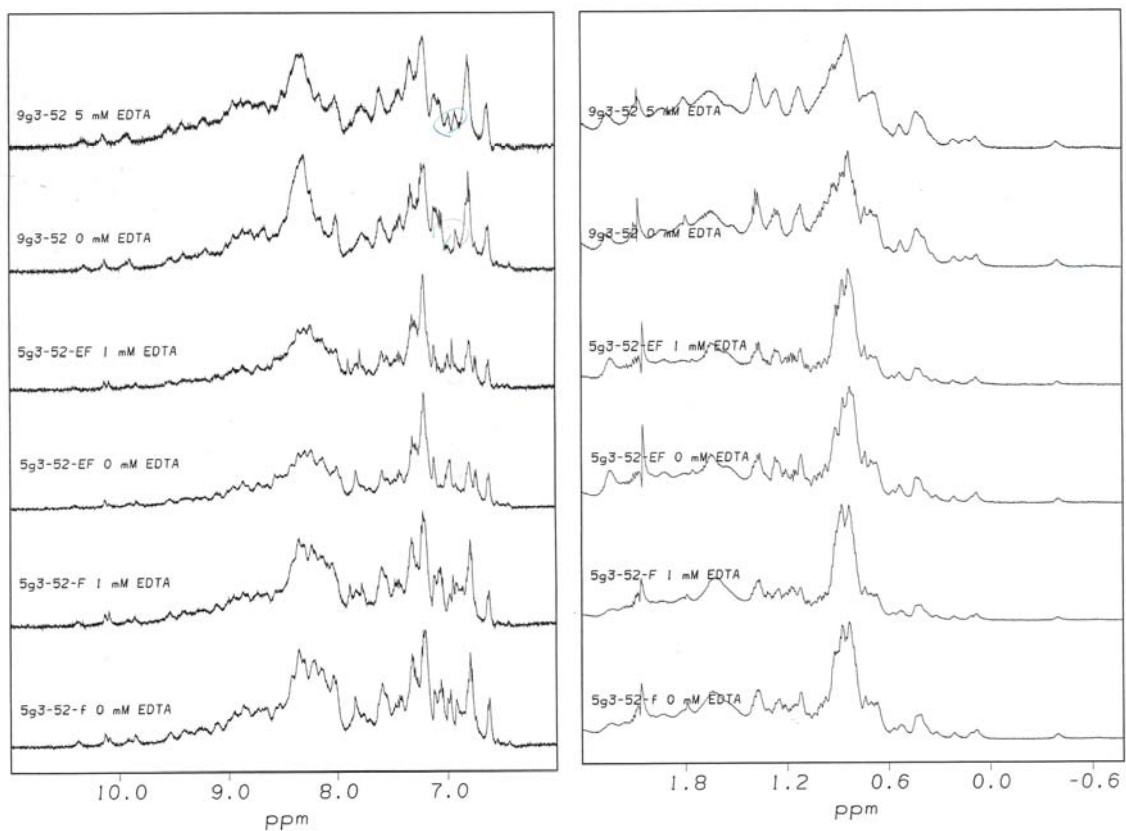
**Figure 6.8 TFE titration on CaM-CD2-III-5G-EF:** The TFE titration spectra for CaM-CD2-III-5G-EF were all recorded in 10 mM Tris 10 mM KCl (pH 7.4) at 25 °C using 600 MHz NMR.



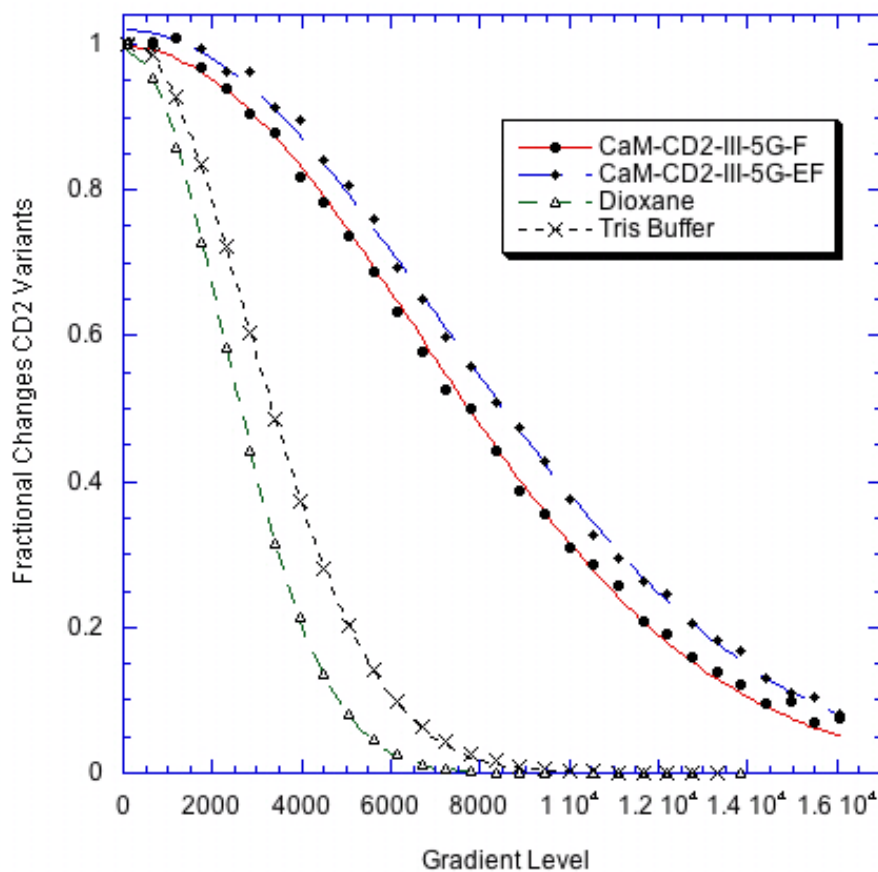
**Figure 6.9 Ca(II) titration on CaM-CD2-III-5G-F:** The Ca(II) titration spectra for CaM-CD2-III-5G-F were all recorded in 10 mM Tris 10 mM KCl (pH 7.4) at 25 °C using 600 MHz NMR.



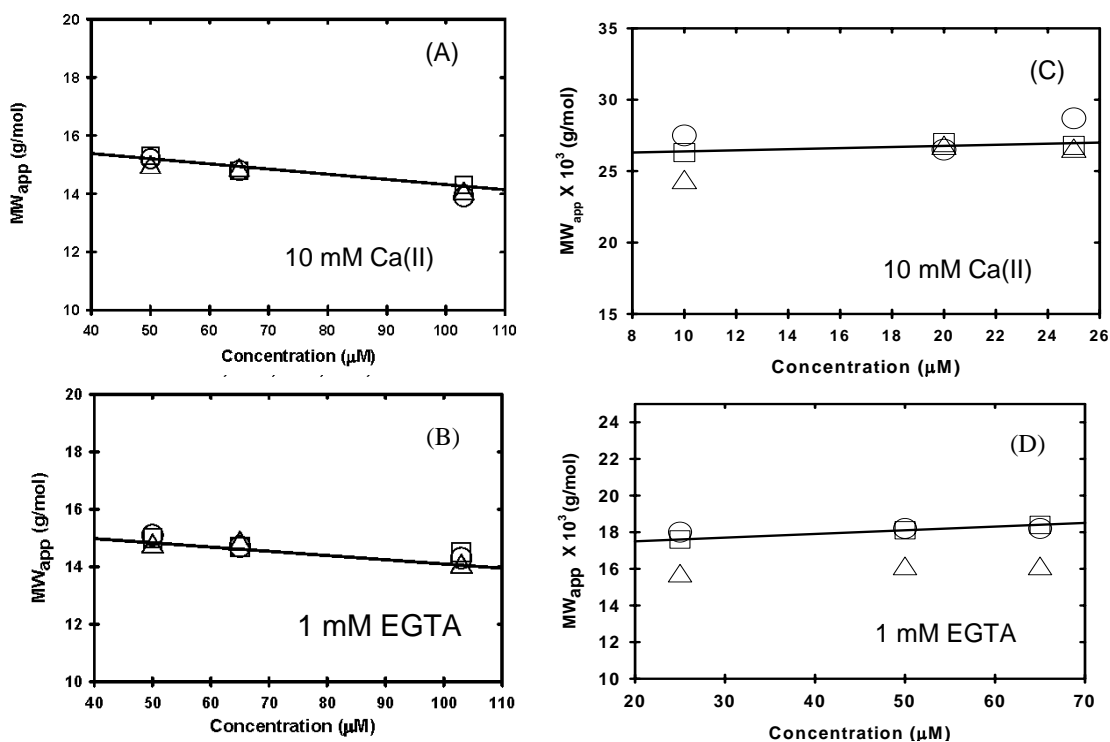
**Figure 6.10 Ca(II) titration on CaM-CD2-III-5G-EF:** The Ca(II) titration spectra for CaM-CD2-III-5G-EF were all recorded in 10 mM Tris 10 mM KCl (pH 7.4) at 25 °C using 600 MHz NMR.



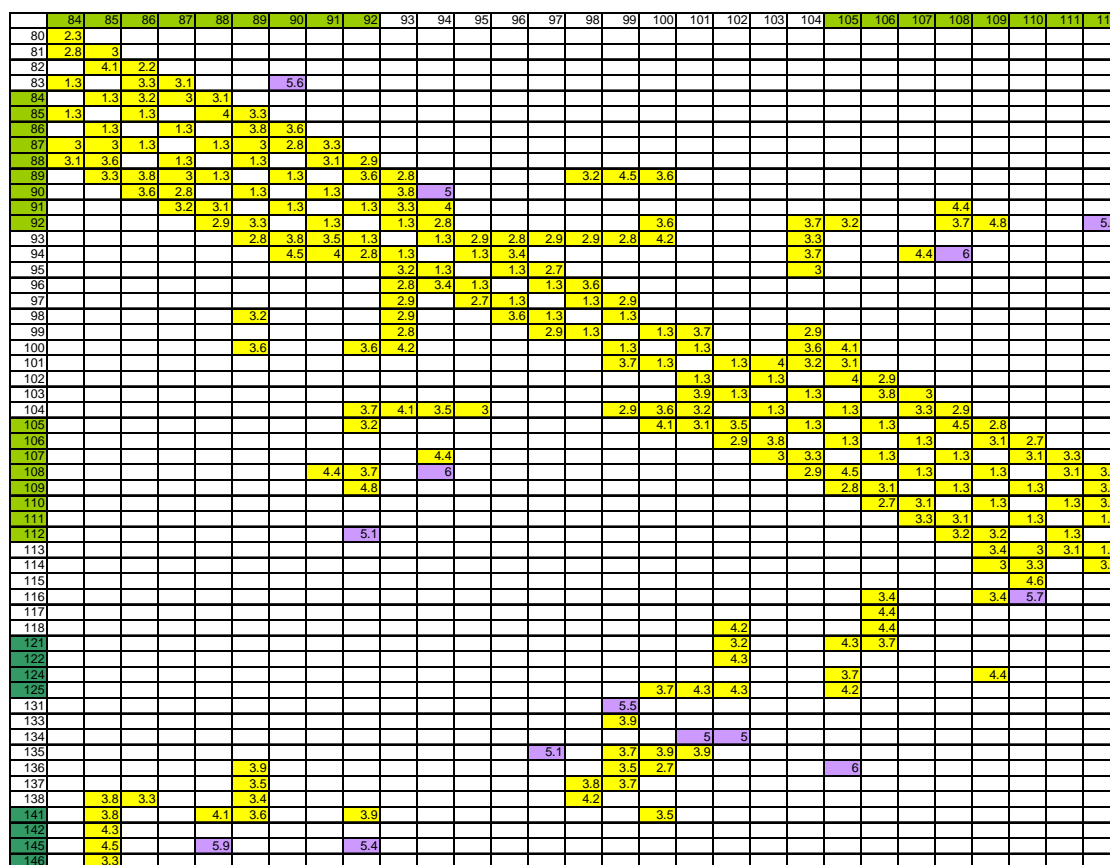
**Figure 6.11  $^1\text{H}$  Spectra of CD2 variants in the presence of EDTA:**  $^1\text{H}$  spectra of CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF in the presence of 1 mM EDTA. These spectra for CD2 variants were all recorded in pH 7.4 10 mM Tris 10 mM KCl at 25 °C using 600 MHz NMR.



**Figure 6.12 PFG diffusion studies on the CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF:** The diffusion studies of CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF in the presence of 1 mM EGTA were performed on 500 MHz at 25 °C. The proteins samples were prepared in 10 mM Tris 10 mM KCl at pH 7.4.



**Figure 6.13 Sedimentation studies on CaM-CD2-III-5G-F and CaM-CD2-III-5G-EF:** Sedimentation Equilibrium Analysis of CaM-CD2-III-5G-52 (A,B) and CaM-CD2-III-EF-5G-52 (C,D) at 21,000 rpm ( $\square$ ), 25,000 rpm (O), 37,000 rpm ( $\Delta$ ) in 20 mM PIPS, pH 6.8 at 4  $^{\circ}C$  for 20 hours



**Figure 6.14 The CSU analysis on the C-terminal domain of CaM:**

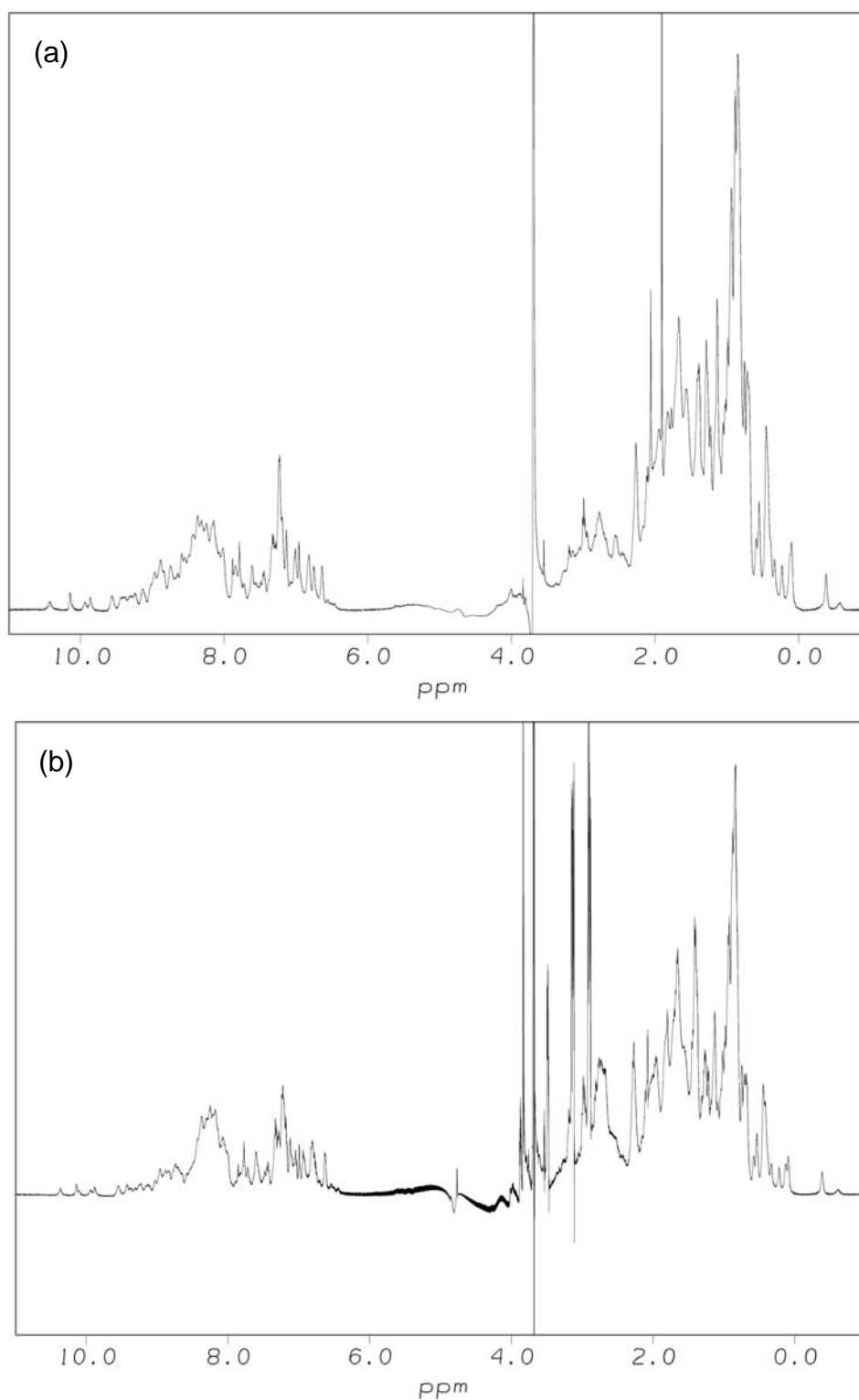
The CSU analysis was carried out using the structure of Ca(II) loaded form of CaM (3CLN.pdb). The intercept cell between the axis is the shortest distance between atoms (mainly carbon atoms) from two different residues. The distances that are within 5 Å are shown in yellow while the distances that are longer than 5 Å are shown in purple. This plot indicate there are possible non-covalent interactions between the intra- and inter-motif helices, which may be responsible for pairing of the EF-hand motifs.



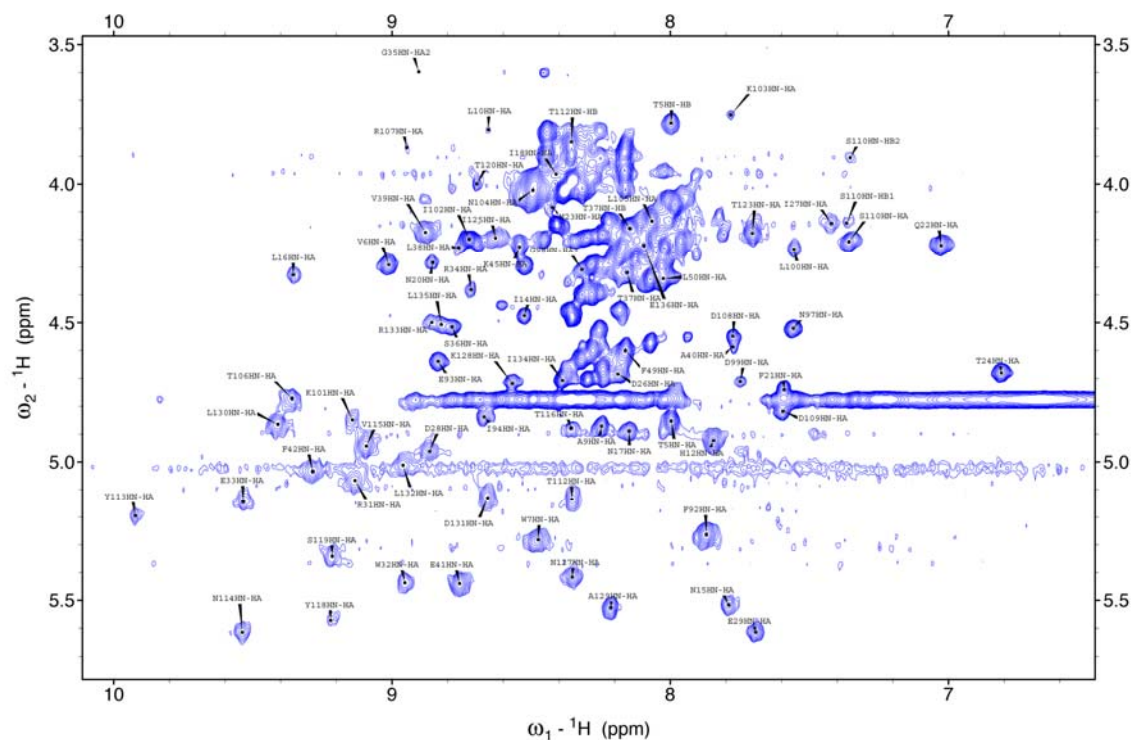
New insertion will be denoted as CaM-CD2-III-5G-EF-SKEAA



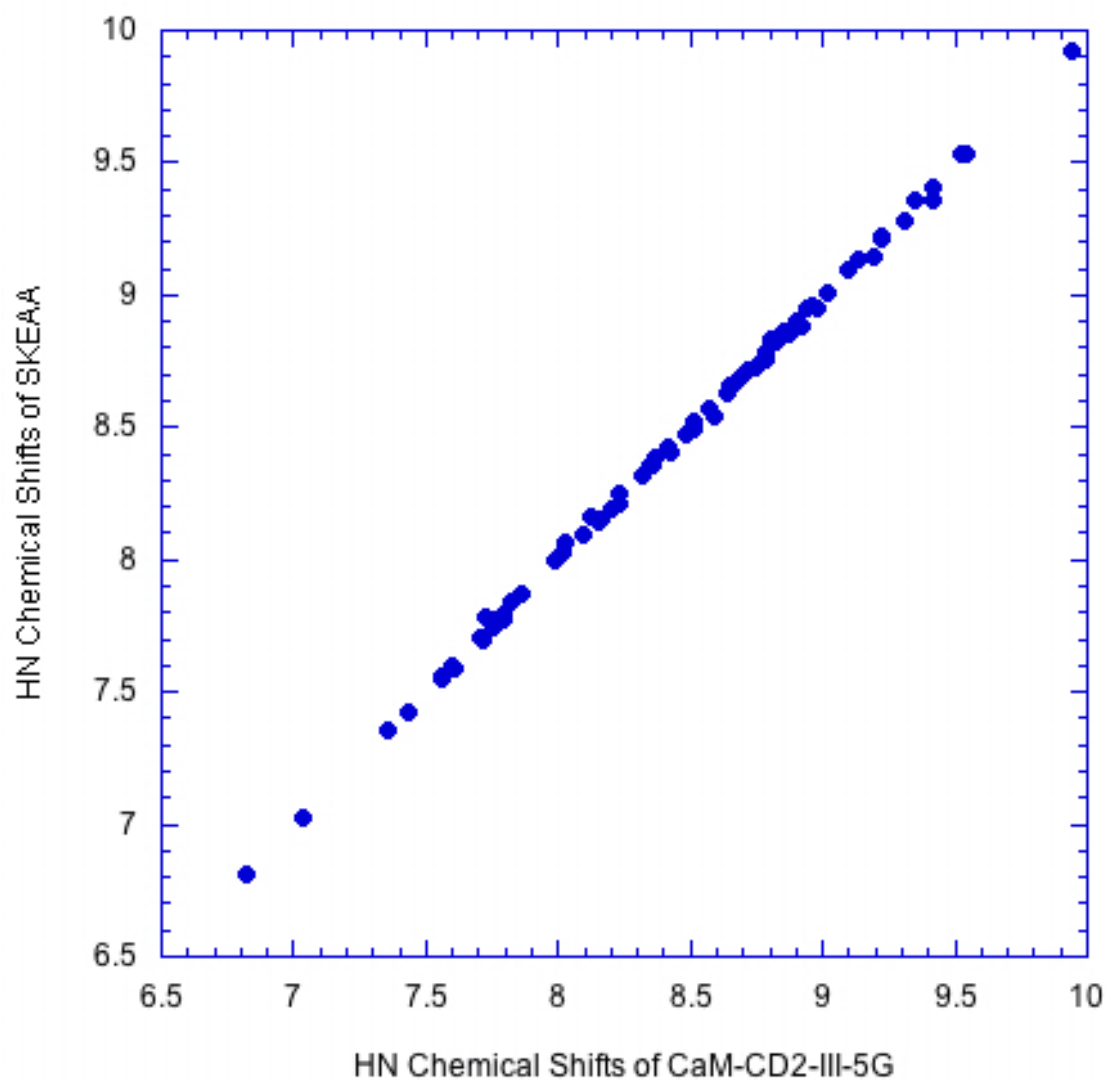
**Figure 6.15 Modification scheme for removing hydrophobic residues of the EF-helices:** Residues at positions -8,-4, 16, and 17 of the EF-hand are likely to be responsible of pairing of EF-hand motifs, thus hydrophobic residues on the E helix were replaced with K and E to form salt bridge. The hydrophobic residues on the F helices were replaced with Ala, which has high helix propensity. A Ser is added to the beginning of the EF-hand motif to function as N-capping. The new protein is denoted as SKEAA.



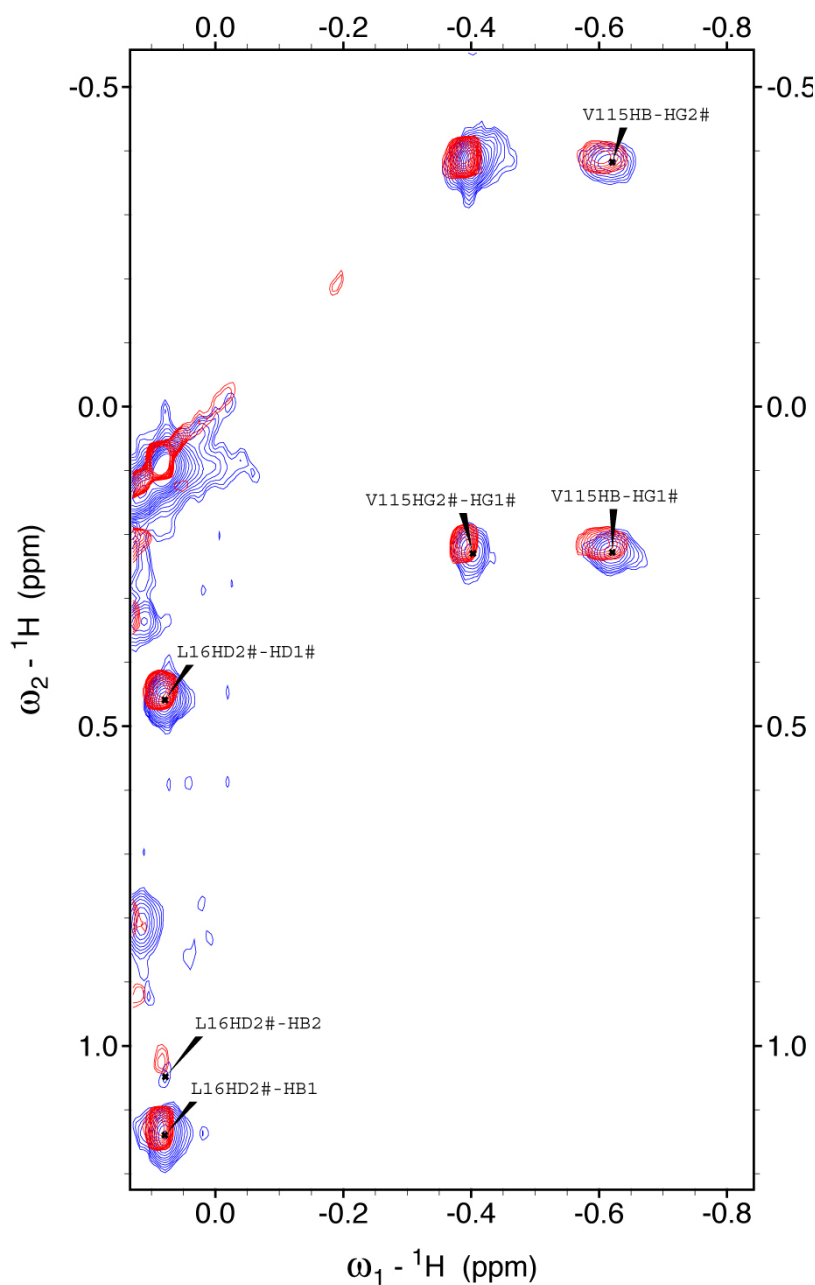
**Figure 6.16  $^1\text{H}$  spectrum of SKEAA:** The  $^1\text{H}$  spectra of CaM-CD2-III-5G-EF (a) and SKEAA (b). Both spectra were collected using 600 MHz NMR at 25 °C. The protein was prepared in 10 mM Tris 10 mM KCl at pH 7.4.



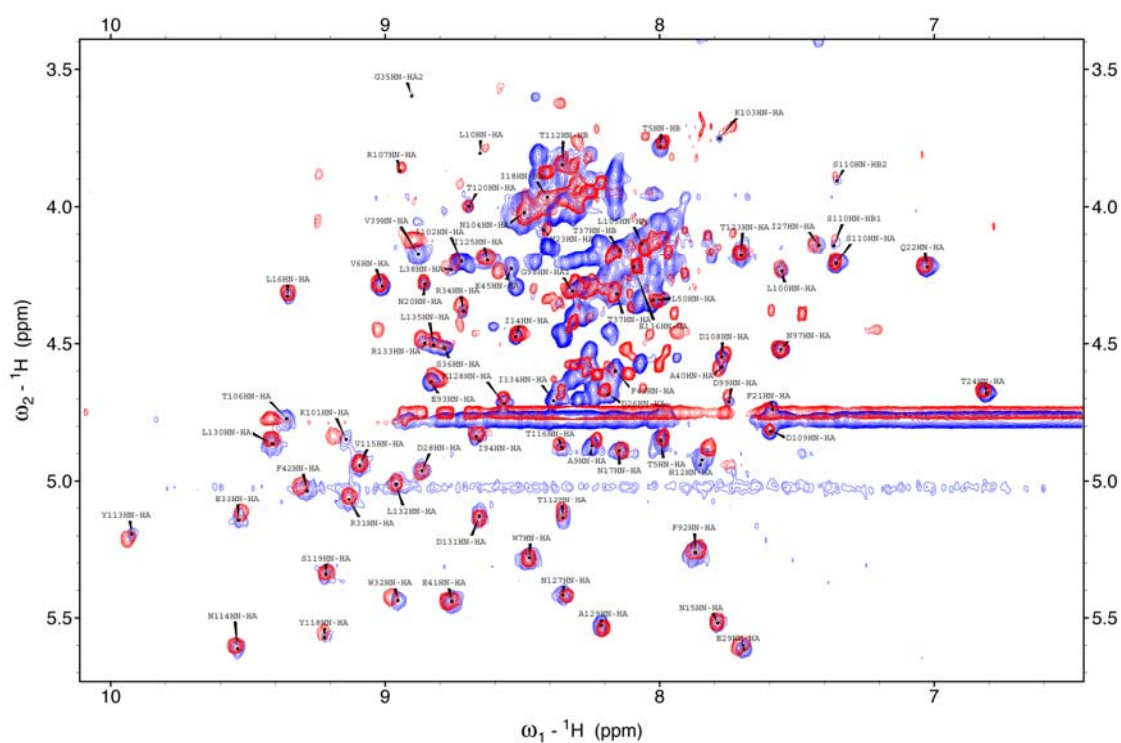
**Figure 6.17 Fingerprint region of the SKEAA TOCSY spectrum:** The assignment for majority of the host protein residues were observed at similar locations. The TOCSY spectrum was collected using 600 MHz NMR at 25 °C. The protein sample was prepared in 10 mM Tris 10 mM KCl at pH 7.4.



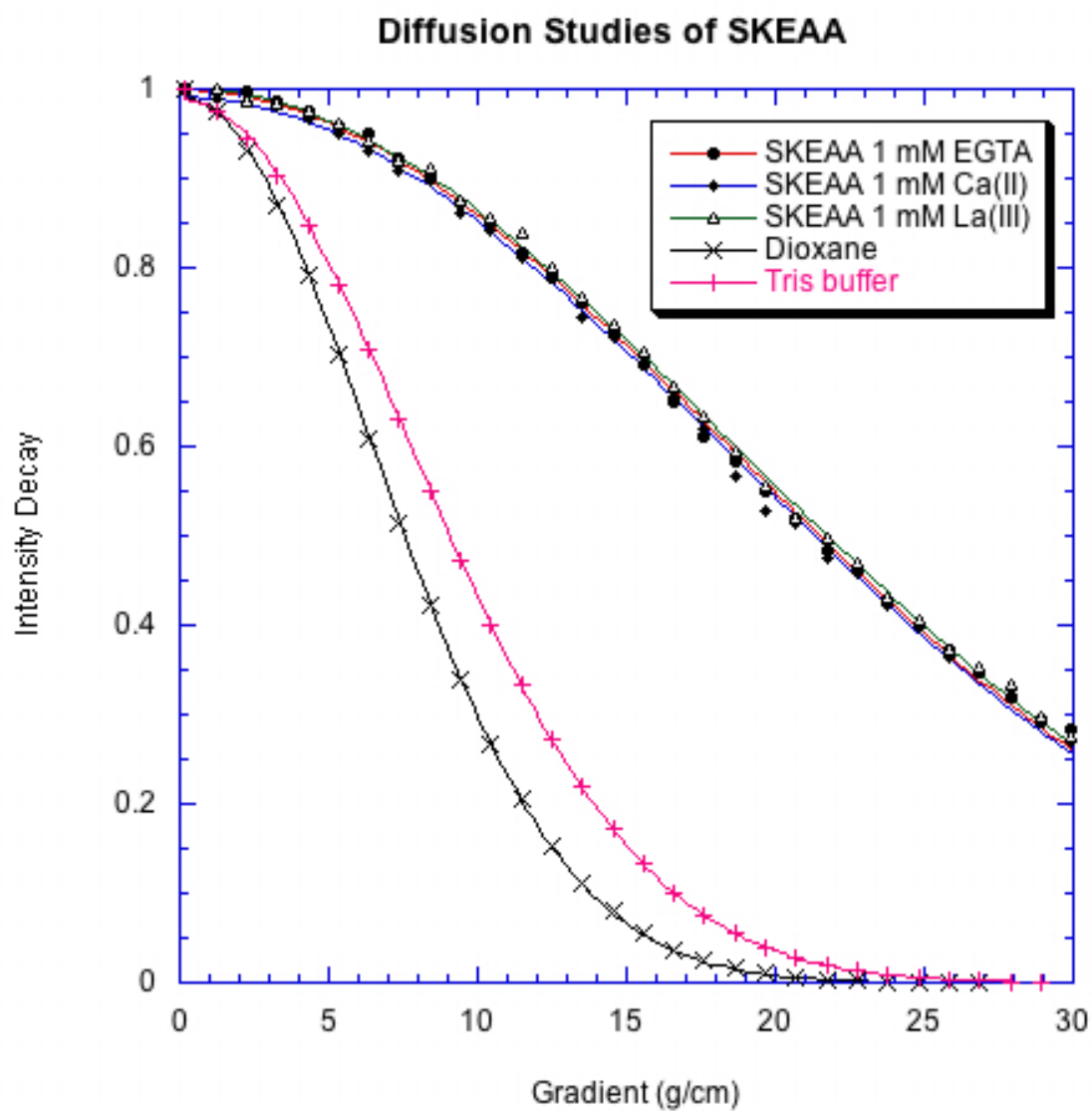
**Figure 6.18 HN chemical shifts comparison between CaM-CD2-III-5G and SKEA:** The HN chemical shifts comparison indicated that the host protein conformation of SKEA remain similar to the CaM-CD2-III-5G.



**Figure 6.19 Comparison of the sidechain region of CaM-CD2-III-5G to SKEAA TOCSY spectra:** The crosspeaks of CaM-CD2-III-5G TOCSY spectrum is showing in red. The crosspeaks of SKEAA TOCSY spectrum is showing in blue. The crosspeaks displayed in this spectra are from sidechain protons that are buried inside the hydrophobic core of the host protein. These TOCSY spectrum were collected using 600 MHz NMR at 25 °C. The protein samples were prepared in 10 mM Tris 10 mM KCl at pH 7.4.



**Figure 6.20 Comparison between the fingerprint regions of CaM-CD2-III-5G to SKEAA:** The crosspeaks of CaM-CD2-III-5G TOCSY spectrum is showing in red. The crosspeaks of SKEAA TOCSY spectrum is showing in blue. These TOCSY spectrum were collected using 600 MHz NMR at 25 °C. The protein samples were prepared in 10 mM Tris 10 mM KCl at pH 7.4.



**Figure 6.21 PFG diffusion studies on SKEAA:** The diffusion studies of SKEAA in the presence of 1 mM EGTA, 1 mM Ca(II), and 1 mM La(III) were performed on 500 MHz at 25 °C. The proteins samples were prepared in 10 mM Tris 10 mM KCl at pH 7.4.

## Chapter 7.0 Conclusions and Major Findings

### 7.0 Conclusions of This Dissertation

The goal of our research is to understand the site specific structural and metal-binding properties of EF-hand proteins. To accomplish this, it was necessary to isolate a single EF-hand Ca(II) binding motif to minimize or eliminate cooperative effects related to the multiple binding sites in a natural Ca(II) binding protein. These effects would include conformational changes upon binding with metal ions, multiple binding of Ca(II) ions, and interactions between the paired EF-hand motifs. We endeavored to obtain site-specific Ca(II) binding properties by grafting the Ca(II) binding site into a scaffold protein to evaluate its intrinsic binding affinity and the contribution of residue types in the EF-loop to Ca(II) binding. Grafted binding sites were inserted into the scaffold with flanking linker sequences comprised of differing numbers of glycine residues. The number of glycine linkers and the length of each linker for connecting the EF-loop were optimized. Further, the conformation of the host protein is not altered in the presence or absence of the metal ion. Using the dihedral angle, NOE, and residual dipolar coupling constraints, the structure of CaM-CD2-III-5G (CD2 host protein with insertion from EF-loop III of CaM) was calculated using CYANA, and it was shown that the insertion of the EF-loop into CD2 did not alter the conformation of the host protein. Moreover, the inserted EF-loop exhibited the same native conformational properties as CaM, especially the strong  $\beta$ -



conformation preferences at position 7 and 8 of the EF-loop. More importantly, we have shown that grafted ligand residues of loop III and IV are directly involved in the binding of Ca(II) and La(III) using high resolution NMR. The diffusion NMR studies suggested that the grafted EF-loop functions as a monomer in solution. Engineered variants with the grafted EF-loop and flanking helices are generally dimerized in the presence of Ca(II). However, removal of hydrophobic residues on the flanking helices results in monomerization of the engineered protein in the absence and presence of Ca(II) or La(III). Therefore, our studies suggest that the flanking helices are responsible for the dimerization.

The grafting approach has also been used to probe the site specific calcium binding properties of calcium binding sites from different types of calcium binding proteins such as CaR and the Rubella virus protein.

The site specific calcium binding affinities of calmodulin determined by the grafting approach follows the order  $I > III \approx II > IV$ . The coupling of the EF-hand motif contributes more than 100-fold to the calcium binding affinity. Our structural and metal binding studies on the C-terminal domain of CaM have indicated that the sequentially different EF-loops III and IV have different structural, metal-binding, and dynamic properties, which leads to different Ca(II) binding affinities. These findings suggest that the magnitude of the Ca(II) binding affinity for an EF-hand site is dependent on the number of charged Ca(II) binding ligands in the coordination sphere, the type of residue that is adjacent to the Ca(II) binding ligand, and the dynamic properties.

## **7.1 Major Findings in Establishing the Grafting Approach**

The purpose of the grafting approach is to be able to study the site specific properties of an isolated EF-hand motif without the complication of multiple binding sites and conformational changes that often occur in the natural EF-hand proteins. The structural properties of the host protein, CD2, were extensively studied using NMR, fluorescence, and CD to ensure that interactions between the EF-loop and the host protein were minimal or nonexistent. It is vital that the host protein can tolerate different sizes and types of calcium binding sites, because instability in the host protein structure will interfere with structural and metal binding properties of the inserted calcium binding moiety. The homonuclear and heteronuclear NMR studies on the 13 engineered proteins discussed in this dissertation indicated that the secondary structures and the packing of the host proteins remained similar to that of wild type CD2, and the inserted calcium binding moieties all retained their metal binding capabilities.

The results of these reported studies suggest several advantages or merits to the grafting approach method. First, the grafting approach can be used to obtain the structural and metal binding and dynamic properties of an individual Ca(II) binding site without the associated complications of proteins with multiple binding sites or conformational changes. We are able to obtain the dynamic properties of EF-loop III and EF-loop IV, which are highly coupled in the native protein. Second, the CD2 host system with a protein frame does not exhibit the

conformational instability due to water solvation that is often associated with small peptide models. Also, the bacterial expression system is very cost effective. Third, the grafting approach can be used to probe Ca(II) binding sites in proteins with biochemical problems or those that are experimentally challenging, such as those exhibiting low solubility, low availability, oligomerization, or if the molecular size of the protein is larger than the experimental limitations. Fourth, since the host protein has a high tolerance for different types of insertions, the grafting approach can be applied to rapidly screen the site-specific calcium binding properties of different EF-hand proteins. We have applied this approach to probe intrinsic Ca(II) binding properties of predicted Ca(II) binding site in CaR and Rubella virus.

## **7.2 Major Finding in Obtaining the Site Specific Ca(II) Binding Properties of EF-hand Protein and the Influence of Charge Arrangement in an EF-hand Ca(II) Binding Site**

The calcium binding affinities of EF-loops I, II, III, and IV of CaM were obtained using the grafting approach and they are 34, 245, 185, and 814  $\mu\text{M}$ , respectively. The EF-loop I of CaM has the strongest Ca(II) binding affinity followed by EF-loop III and EF-loop II, while EF-loop IV of CaM exhibited the weakest metal binding affinity. The Ca(II) binding affinities of the four CD2 variants with different EF-loops of CaM contradict the acid-pair hypothesis postulated by Reid and co-workers who predicted Ca(II) binding affinities for the

EF-loops of CaM in the order  $I \approx IV > II > III$ . Our studies on the CD2 variants with CaM EF-loops prompted us to devise a working model to define this discrepancy. We propose that both the differences in the apo form of the protein and their dynamic properties of the EF-hand motif are the key determinants for the observed differential metal binding affinity.

As discussed in chapters 4 and 5, the NMR studies on EF-loop III and EF-loop IV clearly showed that these two EF-loops have different structural and dynamic properties in the absence of Ca(II). The results in this dissertation enable us to state that the Ca(II) binding affinity for an EF-loop is dependent on the number of the charged residues in the coordination sphere, the type of residues adjacent to the ligands, and the conformational entropy of the EF-loop. The knowledge of the site-specific Ca(II) binding affinity for EF-hand proteins will allow us to understand the ability of EF-hand proteins to regulate different biological functions at different locations and with different Ca(II) concentrations. Previous studies have shown that mutations in Ca(II) binding sites can cause human disorders (188, 189). Using the developed grafting system, we can probe the effect of disease-associated mutations on the intrinsic Ca(II) binding capability. The mutation studies using our model system to gain a better understanding for the molecular basis of diseases related to Ca(II) binding and the mechanisms of calcium signaling (111).

### **7.3 Major Findings in Determining the Contribution of the Helices to Metal Binding Affinity and Pair-Pair Interactions**

In contrast to the previously reported dimeric coupling of the 12-residue peptide encompassing EF-loop III of calmodulin in the presence of La(III), our laboratory has shown that the same EF-loop grafted into CD2 remains monomeric both in the absence and presence of Ca(II) and La(III) using PFG diffusion NMR. The PFG diffusion studies were carried out under the same experimental conditions as used during the metal binding studies (184). The diffusion coefficient values for the CD2 variants are approximately  $11.1 \times 10^{-7} \text{ cm}^2/\text{s}$  both in the presence and absence of metal ions, which are equivalent to those of wild type CD2. This suggests that the isolated EF-loop III of calmodulin inserted in the scaffold protein is able to bind calcium and lanthanum as a monomer, which is contradictory to previous observations for the EF-hand motif (40, 170). Our results imply that additional factors that reside outside of EF-loop III may contribute to the pairing in EF-hand motifs of calmodulin. Here, we have further demonstrated that the isolated EF-loop in CD2 is monomeric, both with or without calcium, using ultracentrifugation. The addition of flanking helices to the EF-loop III of CaM does not lead to a significant dimerization in the absence of Ca(II). Since its apparent molecular weight is greater than that of the calculated molecular weight and the apparent molecular weight slightly increases with the protein concentration, this is an indication of low degree self-associating interactions. In addition, the apparent molecular weight of 18 KDa at a rotor

speed of 21,000 rpm is decreased to 16 KDa at 37,000 rpm. All of these results suggest that CaM-CD2-III-5G-EF is primarily in the monomer formation with a minor fraction in the dimer formation in the absence of calcium. Strikingly, calcium binding appears to dimerize all CaM-CD2-III-5G-EF at low protein concentrations. The molar ellipticity of this protein at several wavelengths is unchanged from protein concentrations ranging from 0.1 to 100  $\mu$ M, suggesting that it remains as a dimer in this range. Furthermore, removal of the hydrophobic residues on the flanking helices completely eliminated dimer formation in the presence and absence of Ca(II). Taken together, our systematic investigation using a grafting approach clearly suggests that the flanking helices play a pivotal role in dimerization and enhances calcium binding affinity, suggesting that there is a strong coupling between calcium binding, folding, and association of the EF-hand motif.

## Reference:

1. Da Silva, J. J. R. F., and Williams, R. J. P. (1991) The biological chemistry of the elements The Inorganic chemistry of life .
2. Maurer, P., Hohenester, E., and Engel, J. (1996) Extracellular calcium-binding proteins, *Curr Opin Cell Biol* 8, 609-617.
3. Zhang, X., and Joseph, S. K. (2001) Effect of mutation of a calmodulin binding site on Ca<sup>2+</sup> regulation of inositol trisphosphate receptors, *Biochem J* 360, 395-400.
4. Lin, X., Sikkink, R. A., Rusnak, F., and Barber, D. L. (1999) Inhibition of calcineurin phosphatase activity by a calcineurin B homologous protein, *J Biol Chem* 274, 36125-36131.
5. Klee, C. B., and Vanaman, T. C. (1982) Calmodulin, *Adv Protein Chem* 35, 213-321.
6. Ermak, G., Morgan, T. E., and Davies, K. J. (2001) Chronic overexpression of the calcineurin inhibitory gene DSCR1 (Adapt78) is associated with Alzheimer's disease, *J Biol Chem* 276, 38787-38794.
7. Berridge, M. J., Bootman, M. D., and Lipp, P. (1998) Calcium--a life and death signal, *Nature* 395, 645-648.
8. Clapham, D. E. (1995) Calcium signaling, *Cell* 80, 259-268.
9. Carafoli, E. (2002) Calcium signaling: A tale for all seasons, *PNAS* 99, 1115-1122.
10. Carafoli, E., Nicotera, P., and Santella, L. (1997) Calcium signalling in the cell nucleus, *Cell Calcium* 22, 313-319.
11. Hu, J., Jia, X., Li, Q., Yang, X., and Wang, K. (2004) Binding of La<sup>3+</sup> to calmodulin and its effects on the interaction between calmodulin and calmodulin binding peptide, polistes mastoparan, *Biochemistry* 43, 2688-2698.
12. Yap, K. L., Ames, J. B., Swindells, M. B., and Ikura, M. (1999) Diversity of conformational states and changes within the EF-hand protein superfamily, *Proteins* 37, 499-507.
13. Wall, M. E., Clarage, J. B., and Phillips, G. N. (1997) Motions of calmodulin characterized using both Bragg and diffuse X-ray scattering, *Structure* 5, 1599-1612.
14. Dowd, D. R. (1995) Calcium regulation of apoptosis, *Adv Second Messenger Phosphoprotein Res* 30, 255-280.
15. Berridge, M. J. (1998) Neuronal calcium signaling, *Neuron* 21, 13-26.
16. Ikura, M., Osawa, M., and Ames, J. B. (2002) The role of calcium-binding proteins in the control of transcription: structure to function, *Bioessays* 24, 625-636.
17. Weis, W. I., and Drickamer, K. (1994) Trimeric structure of a C-type mannose-binding protein, *Structure* 2, 1227-1240.

18. 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.
19. Maler, L., Blankenship, J., Rance, M., and Chazin, W. J. (2000) Site-site communication in the EF-hand  $\text{Ca}^{2+}$ -binding protein calbindin D9k, *Nat Struct Biol* 7, 245-250.
20. Spyrapoulos, L., Gagne, S. M., Li, M. X., and Sykes, B. D. (1998) Dynamics and thermodynamics of the regulatory domain of human cardiac troponin C in the apo- and calcium-saturated states, *Biochemistry* 37, 18032-18044.
21. 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.
22. Chen, W., Steenbergen, C., Levy, L. A., Vance, J., London, R. E., and Murphy, E. (1996) Measurement of free calcium in sarcoplasmic reticulum in perfused rabbit heart loaded with 1,2-bis(2-amino-5,6-difluorophenoxy)ethane-N,N,N',N'-tetraacetic acid by  $^{19}\text{F}$  NMR, *J Biol Chem* 271, 7398-7403.
23. Glusker, J. P. (1991) Structural aspects of metal liganding to functional groups in proteins, *Adv Protein Chem* 42, 1-76.
24. Goll, D. E., Thompson, V. F., Li, H., Wei, W., and Cong, J. (2003) The calpain system, *Physiol Rev* 83, 731-801.
25. 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.
26. Ward, D. T. (2004) Calcium receptor-mediated intracellular signalling, *Cell Calcium* 35, 217-228.
27. Mosher, D. F. (1993) Assembly of fibronectin into extracellular matrix, *Current Opinion in Structural Biology* 3, 214-222.
28. Busch, E., Hohenester, E., Timpl, R., Paulsson, M., and Maurer, P. (2000) Calcium affinity, cooperativity and domain interactions of extracellular EF-hands present in BM-40, *J Biol Chem*.
29. Evenas, J., Malmendal, A., and Forsen, S. (1998) Calcium, *Curr. Opin. Chem. Biol.* 2, 293-302.
30. Lewis, J. E., Jensen, P. J., Johnson, K. R., and Wheelock, M. J. (1994) E-cadherin mediates adherens junction organization through protein kinase C, *J Cell Sci* 107, 3615-3621.
31. Brown, E. M., Pollak, M., and Hebert, S. C. (1998) The extracellular calcium-sensing receptor: its role in health and disease, *Annu Rev Med* 49, 15-29.
32. Shapiro, L., and Colman, D. R. (1998) Structural biology of cadherins in the nervous system., *Curr Opin Neurobiol* 8, 593-599.
33. Brown, E. M., and MacLeod, R. J. (2001) Extracellular calcium sensing and extracellular calcium signaling, *Physiol Rev* 81, 239-297.



34. Bai, M. (2004) Structure-function relationship of the extracellular calcium-sensing receptor, *Cell Calcium* 35, 197-207.
35. Downing, A. K., Knott, V., Werner, J. M., Cardy, C. M., Campbell, I. D., and Handford, P. A. (1996) Solution structure of a pair of calcium-binding epidermal growth factor-like domains: Implications for the Marfan Syndrome and other genetic disorder, *Cell* 85, 597-605.
36. Zhou, Y., Yang, W., Kirberger, M., Lee, H. W., Ayalasomayajula, G., and Yang, J. J. (2006) Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins, *Proteins* 65, 643-655.
37. Yang, J. J., and Yang, W. (2005) Calcium-binding Proteins, *Encyclopedia of Inorganic Chemistry, Second Edition* 2, 630-666.
38. Yang, W., Jones, L. M., Isley, L., Ye, Y., Lee, H. W., Wilkins, A., Liu, Z. R., Hellinga, H. W., Malchow, R., Ghazi, M., and Yang, J. J. (2003) Rational Design of a Calcium-binding Protein, *J Am Chem Soc* 125, 6165-6171.
39. Henikoff, S., Greene, E. A., Pietrokovski, S., Bork, P., Attwood, T. K., and Hood, L. (1997) Gene families: the taxonomy of protein paralogs and chimeras, *Science* 278, 609-614.
40. Wojcik, J., Goral, J., Pawlowski, K., and Bierzynski, A. (1997) Isolated calcium-binding loops of EF-hand proteins can dimerize to form a native-like structure, *Biochemistry* 36, 680-687.
41. Kawasaki, H., Nakayama, S., and Kretsinger, R. H. (1998) Classification and evolution of EF-hand proteins, *Biomaterials* 11, 277-295.
42. 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.
43. Kawasaki, H., and Kretsinger, R. H. (1995) Calcium-binding proteins 1: EF-hands, *Protein Profile* 2, 297-490.
44. Strynadka, N. C., and James, M. N. (1989) Crystal structures of the helix-loop-helix calcium-binding proteins, *Annu Rev Biochem* 58, 951-998.
45. Herzberg, O., Moulton, J., and James, M. N. (1986) A model for the Ca<sup>2+</sup>-induced conformational transition of troponin C. A trigger for muscle contraction, *J Biol Chem* 261, 2638-2644.
46. Babu, Y. S., Bugg, C. E., and Cook, W. J. (1988) Structure of calmodulin refined at 2.2 Å resolution, *J Mol Biol* 204, 191-204.
47. Svensson, L. A., Thulin, E., and Forsen, S. (1992) Proline cis-trans isomers in calbindin D9k observed by X-ray crystallography, *J Mol Biol* 223, 601-606.
48. Linse, S., and Forsen, S. (1995) Determinants that govern high-affinity calcium binding, *Adv Second Messenger Phosphoprotein Res* 30, 89-151.
49. Kretsinger, R. H. (1980) Structure and evolution of calcium-modulated proteins, *CRC Crit Rev Biochem* 8, 119-174.
50. Shaw, G. S., Hodges, R. S., and Sykes, B. D. (1990) Calcium-induced peptide association to form an intact protein domain: <sup>1</sup>H NMR structural evidence, *Science* 249, 280-283.

51. Monera, O. D., Kay, C. M., and Hodges, R. S. (1991) Protein denaturation with guanidine hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions, *Protein Science* 3, 1984-1991.
52. Julenius, K., Robblee, J., Thulin, E., Finn, B. E., Fairman, R., and Linse, S. (2002) Coupling of ligand binding and dimerization of helix-loop-helix peptides: spectroscopic and sedimentation analyses of calbindin D9k EF-hands, *Proteins* 47, 323-333.
53. Ikura, M. (1996) Calcium binding and conformational response in EF-hand proteins, *Trends Biochem Sci* 21, 14-17.
54. Ghosh, A., and Greenberg, M. E. (1995) Calcium signaling in neurons: molecular mechanisms and cellular consequences, *Science* 268, 239-247.
55. Bootman, M. D., and Berridge, M. J. (1995) The elemental principles of calcium signaling, *Cell* 83, 675-678.
56. Babu, Y. S., Sack, J. S., Greenhough, T. J., Bugg, C. E., Means, A. R., and Cook, W. J. (1985) Three-dimensional structure of calmodulin, *Nature* 315, 37-40.
57. Bentrop, D., Bertini, I., Cremonini, M. A., Forsen, S., Luchinat, C., and Malmendal, A. (1997) Solution structure of the paramagnetic complex of the N-terminal domain of calmodulin with two Ce<sup>3+</sup> ions by <sup>1</sup>H NMR, *Biochemistry* 36, 11605-11618.
58. Wilson, M. A., and Brunger, A. T. (2003) Domain flexibility in the 1.75 Å resolution structure of Pb<sup>2+</sup>-calmodulin, *Acta Crystallogr D Biol Crystallogr* 59, 1782-1792.
59. Amir, D., Krausz, S., and Haas, E. (1992) Detection of local structures in reduced unfolded bovine pancreatic trypsin inhibitor., *Proteins: Structure, Function and Genetics* 13, 162-173.
60. Zhang, M., Tanaka, T., and Ikura, M. (1995) Calcium-induced conformational transition revealed by the solution structure of apo calmodulin, *Nature Structural Biology* 2, 758-767.
61. Nelson, M. R., and Chazin, W. J. (1998) An interaction-based analysis of calcium-induced conformational changes in Ca<sup>2+</sup> sensor proteins, *Protein Sci* 7, 270-282.
62. Kuboniwa, H., Tjandra, N., Grzesiek, S., Ren, H., Klee, C. B., and Bax, A. (1995) Solution structure of calcium-free calmodulin, *Nature structure biology* 2, 768-776.
63. Finn, B. E., Evenas, J., Drakenberg, T., Waltho, J. P., Thulin, E., and Forsen, S. (1995) Calcium-induced structural changes and domain autonomy in calmodulin, *Nat Struct Biol* 2, 777-783.
64. Schumacher, M. A., Crum, M., and Miller, M. C. (2004) Crystal structures of apocalmodulin and an apocalmodulin/SK potassium channel gating domain complex, *Structure* 12, 849-860.

65. Tjandra, N., Kuboniwa, H., Ren, H., and Bax, A. (1995) Rotational dynamics of calcium-free calmodulin studied by  $^{15}\text{N}$ -NMR relaxation measurements, *Eur J Biochem* 230, 1014-1024.
66. Browne, J. P., Strom, M., Martin, S. R., and Bayley, P. M. (1997) The role of beta-sheet interactions in domain stability, folding, and target recognition reactions of calmodulin, *Biochemistry* 36, 9550-9561.
67. Chattopadhyaya, R., Meador, W. E., Means, A. R., and Quiocho, F. A. (1992) Calmodulin structure refined at 1.7 Å resolution, *J Mol Biol* 228, 1177-1192.
68. Taylor, D. A., Sack, J. S., Maune, J. F., Beckingham, K., and Quiocho, F. A. (1991) Structure of a recombinant calmodulin from *Drosophila melanogaster* refined at 2.2-Å resolution, *J Biol Chem* 266, 21375-21380.
69. Chou, J. J., Li, S., Klee, C. B., and Bax, A. (2001) Solution structure of  $\text{Ca}(2+)$ -calmodulin reveals flexible hand-like properties of its domains, *Nat Struct Biol* 8, 990-997.
70. Wilson, M. A., and Brunger, A. T. (2000) The 1.0 Å crystal structure of  $\text{Ca}(2+)$ -bound calmodulin: an analysis of disorder and implications for functionally relevant plasticity, *J Mol Biol* 301, 1237-1256.
71. Rupp, B., Marshak, D. R., and Parkin, S. (1996) Crystallization and preliminary X-ray analysis of two new crystal forms of calmodulin, *Acta Crystallogr D Biol Crystallogr* 52, 411-413.
72. Yun, C. H., Bai, J., Sun, D. Y., Cui, D. F., Chang, W. R., and Liang, D. C. (2004) Structure of potato calmodulin PCM6: the first report of the three-dimensional structure of a plant calmodulin, *Acta Crystallogr D Biol Crystallogr* 60, 1214-1219.
73. Fallon, J. L., and Quiocho, F. A. (2003) A closed compact structure of native  $\text{Ca}(2+)$ -calmodulin, *Structure* 11, 1303-1307.
74. Ishida, H., Takahashi, K., Nakashima, K., Kumaki, Y., Nakata, M., Hikichi, K., and Yazawa, M. (2000) Solution structures of the N-terminal domain of yeast calmodulin:  $\text{Ca}^{2+}$ -dependent conformational change and its functional implication, *Biochemistry* 39, 13660-13668.
75. Olsson, L. L., and Sjölin, L. (2001) Structure of *Escherichia coli* fragment TR2C from calmodulin to 1.7 Å resolution, *Acta Crystallogr D Biol Crystallogr* 57, 664-669.
76. Ban, C., Ramakrishnan, B., Ling, K. Y., Kung, C., and Sundaralingam, M. (1994) Structure of the recombinant *Paramecium tetraurelia* calmodulin at 1.68 Å resolution, *Acta Crystallogr D Biol Crystallogr* 50, 50-63.
77. Rao, S. T., Wu, S., Satyshur, K. A., Ling, K. Y., Kung, C., and Sundaralingam, M. (1993) Structure of *Paramecium tetraurelia* calmodulin at 1.8 Å resolution, *Protein Sci* 2, 436-447.
78. Ishida, H., Nakashima, K., Kumaki, Y., Nakata, M., Hikichi, K., and Yazawa, M. (2002) The solution structure of apocalmodulin from *Saccharomyces cerevisiae* implies a mechanism for its unique  $\text{Ca}^{2+}$ -binding property, *Biochemistry* 41, 15536-15542.

79. Ababou, A., and Desjarlais, J. R. (2001) Solvation energetics and conformational change in EF-hand proteins, *Protein Sci* 10, 301-312.
80. Sorensen, B. R., and Shea, M. A. (1998) Interactions between domains of apo calmodulin alter calcium binding and stability, *Biochemistry* 37, 4244-4253.
81. Fefeu, S., Biekofsky, R. R., McCormick, J. E., Martin, S. R., Bayley, P. M., and Feeney, J. (2000) Calcium-induced refolding of the calmodulin V136G mutant studied by NMR spectroscopy: evidence for interaction between the two globular domains, *Biochemistry* 39, 15920-15931.
82. Waltersson, Y., Linse, S., Brodin, P., and Grundstrom, T. (1993) Mutational effects on the cooperativity of Ca<sup>2+</sup> binding in calmodulin, *Biochemistry* 32, 7866-7871.
83. Wu, X., and Reid, R. E. (1997) Structure/calcium affinity relationships of site III of calmodulin: testing the acid pair hypothesis using calmodulin mutants, *Biochemistry* 36, 8649-8656.
84. Wu, X., and Reid, R. E. (1997) Conservative D133E mutation of calmodulin site IV drastically alters calcium binding and phosphodiesterase regulation, *Biochemistry* 36, 3608-3616.
85. Mukherjea, P., Maune, J. F., and Beckingham, K. (1996) Interlobe communication in multiple calcium-binding site mutants of Drosophila calmodulin, *Protein Sci* 5, 468-477.
86. Beckingham, K. (1991) Use of site-directed mutations in the individual Ca<sup>2+</sup>(+)-binding sites of calmodulin to examine Ca<sup>2+</sup>(+)-induced conformational changes, *J Biol Chem* 266, 6027-6030.
87. Haiech, J., Kilhoffer, M. C., Lukas, T. J., Craig, T. A., Roberts, D. M., and Watterson, D. M. (1991) Restoration of the calcium binding activity of mutant calmodulins toward normal by the presence of a calmodulin binding structure, *J Biol Chem* 266, 3427-3431.
88. Starovasnik, M. A., Su, D. R., Beckingham, K., and Klevit, R. E. (1992) A series of point mutations reveal interactions between the calcium-binding sites of calmodulin, *Protein Sci* 1, 245-253.
89. Maune, J. F., Beckingham, K., Martin, S. R., and Bayley, P. M. (1992) Circular dichroism studies on calcium binding to two series of Ca<sup>2+</sup> binding site mutants of Drosophila melanogaster calmodulin, *Biochemistry* 31, 7779-7786.
90. Biekofsky, R. R., and Feeney, J. (1998) Cooperative cyclic interactions involved in metal binding to pairs of sites in EF-hand proteins, *FEBS Lett* 439, 101-106.
91. Linse, S., Helmersson, A., and Forsen, S. (1991) Calcium binding to calmodulin and its globular domains, *J Biol Chem* 266, 8050-8054.
92. Dadlez, M., Goral, J., and Bierzynski, A. (1991) Luminescence of peptide-bound terbium ions. Determination of binding constants, *FEBS Lett* 282, 143-146.

93. Borin, G., Calderan, A., Ruzza, P., Pezzoli, A., Marchiori, F., and Peggion, E. (1989) Synthesis of the dodecapeptide corresponding to domain III of bovine brain calmodulin: alpha-beta shift side reactions during the synthesis by the classical method in solution, *Biopolymers* 28, 333-352.
94. Yang, J. J., Carroll, A. R., Yang, W., Ye, Y. M., and Nguyen, C. (2000) Nonnative intermediate state of acid-stable beta-sheet protein, *Cell Biochemistry and Biophysics* 33, 253-273.
95. Arulanandam, A. R., Moingeon, P., Concino, M. F., Recny, M. A., Kato, K., Yagita, H., Koyasu, S., and Reinherz, E. L. (1993) 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.
96. Davis, S. J., Ikemizu, S., Wild, M. K., and van der Merwe, P. A. (1998) CD2 and the nature of protein interactions mediating cell-cell recognition, *Immunol Rev* 163, 217-236.
97. Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Grubel, G., Legrand, J. F., Als-Nielsen, J., Colman, D. R., and Hendrickson, W. A. (1995) Structural Basis of Cell-Cell Adhesion by Cadherins, *Nature* 374, 327-337.
98. 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.
99. Bodian, D. L., Jones, E. Y., Harlos, K., Stuart, D. I., and Davis, S. (1994) Crystal structure of the extracellular region of the human cell adhesion molecule CD2 at 2.5 Å resolution, *Structure* 2, 755-766.
100. 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.
101. Withka, J. M., Wyss, D. F., Wagner, G., Arulanandam, A. R., Reinherz, E. L., and Recny, M. A. (1993) Structure of the glycosylated adhesion domain of human T lymphocyte glycoprotein CD2, *Structure* 1, 69-81.
102. Wyss, D. F., Choi, J. S., Li, J., Knoppers, M. H., Willis, K. J., Arulanandam, A. R., Smolyar, A., Reinherz, E. L., and Wagner, G. (1995) Conformation and function of the N-linked glycan in the adhesion domain of human CD2 [see comments], *Science* 269, 1273-1278.
103. Arulanandam, A. R., Withka, J. M., Wyss, D. F., Wagner, G., Kister, A., Pallai, P., Recny, M. A., and Reinherz, E. L. (1993) 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.
104. Yang, J. J., Yang, H., Ye, Y., Hopkins, H., Jr., and Hastings, G. (2002) Temperature-induced formation of a non-native intermediate state of the all beta-sheet protein CD2, *Cell Biochem Biophys* 36, 1-18.

105. Davis, S. J., Davies, E. A., Tucknott, M. G., Jones, E. Y., and van der Merwe, P. A. (1998) 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.
106. 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.
107. Maniccia, A. W., Yang, W., Li, S. Y., Johnson, J. A., and Yang, J. J. (2006) Using protein design to dissect the effect of charged residues on metal binding and protein stability, *Biochemistry* 45, 5848-5856.
108. Pfuhl, M., Improta, S., Politou, A. S., and Pastore, A. (1997) When a module is also a domain: the role of the N terminus in the stability and the dynamics of immunoglobulin domains from titin, *J Mol Biol* 265, 242-256.
109. Yang, W., Wilkins, A. L., Li, S., Ye, Y., and Yang, J. J. (2005) The effects of Ca<sup>2+</sup> binding on the dynamic properties of a designed Ca<sup>2+</sup>-binding protein, *Biochemistry* 44, 8267-8273.
110. Rebeor, J. F., and Senior, A. E. (1998) Effects of cardiovascular drugs on ATPase activity of P-glycoprotein in plasma membranes and in purified reconstituted form, *Biochim Biophys Acta* 1369, 85-93.
111. Farley, R. A. (1983) Identification of hydrophobic regions of the calcium-transport ATPase from sarcoplasmic reticulum after photochemical labeling with adamantane diazirine, *Int J Biochem* 15, 1423-1427.
112. Ye, Y., Lee, H. W., Yang, W., Shealy, S. J., Wilkins, A. L., Liu, Z. R., Torshin, I., Harrison, R., Wohlhueter, R., and Yang, J. J. (2001) Metal binding affinity and structural properties of an isolated EF-loop in a scaffold protein, *Protein Eng* 14, 1001-1013.
113. Goddard, T., Kneller, D., Sparky 3, *University of California, San Francisco*.
114. Bagshaw, C. R., and Harris, D. A. (1987) in *Spectrophotometry and spectrofluorimetry: A practical approach* (Bagshaw, C. R., and Harris, D. A., Eds.), pp 91-113, IRL Press, Washington.
115. Yang, W., Tsai, T., Kats, M., and Yang, J. J. (2000) Peptide analogs from E-cadherin with different calcium-binding affinities, *J Pept Res* 55, 203-215.
116. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J Biomol NMR* 13, 289-302.
117. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6, 277-293.
118. Chen, H. A., Pfuhl, M., Davis, B., and Driscoll, P. C. (1998) Sequence specific <sup>1</sup>H, <sup>13</sup>C and <sup>15</sup>N resonance assignment of rat CD2 domain 1, *J Biomol NMR* 12, 457-458.

119. Valafar, H., and Prestegard, J. H. (2004) REDCAT: a residual dipolar coupling analysis tool, *J Magn Reson* 167, 228-241.
120. Prestegard, J. H., Bougault, C. M., and Kishore, A. I. (2004) Residual dipolar couplings in structure determination of biomolecules, *Chem Rev* 104, 3519-3540.
121. Prestegard, J. H., Mayer, K. L., Valafar, H., and Benison, G. C. (2005) Determination of protein backbone structures from residual dipolar couplings, *Methods Enzymol* 394, 175-209.
122. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J Mol Biol* 273, 283-298.
123. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr D Biol Crystallogr* 54, 905-921.
124. Palmer III, A. G., Rance, M., and Wright, P. E. (1991) Intramolecular Motions of a Zinc Finger DNA-Binding Domain from Xfin Characterized by Proton-Detected Natural Abundance  $^{13}\text{C}$  Heteronuclear NMR Spectroscopy, *J. Am. Chem Soc.* 113, 4371-4380.
125. Mandel, A. M., Akke, M., Palmer, A. G., and III. (1995) Backbone Dynamics of Escherichia coli Ribonuclease HI: Correlations with Structure and Function in an Active Enzyme, *J. Mol. Biol.* 246, 144-163.
126. Gibbs, S. J., and Johnson, C. S. (1991) A PFG NMR experiment for accurate diffusion and flow studies in the presence of eddy currents, *Journal of Magnetic Resonance* 93, 395-402.
127. Jones, J. A., Wilikins, D. K., Smith, L. J., and Dobson, C. M. (1997) Characterisation of protein unfolding by NMR diffusion measurements, *Journal of Biomolecular NMR* 10, 199-203.
128. Altieri, A. S., Hinton, D. P., and Byrd, R. A. (1995) Association of biomolecular systems via pulsed field gradient NMR self-diffusion measurements, *J. Am. Chem. Soc.* 117, 7566-7567.
129. Recny, M. A., Neidhardt, E. A., Sayre, P. H., Ciardelli, T. L., and Reinherz, E. L. (1990) Structural and functional characterization of the CD2 immunoadhesion domain. Evidence for inclusion of CD2 in an alpha-beta protein folding class, *J Biol Chem* 265, 8542-8549.
130. Borin, G., Ruzza, P., Rossi, M., Calderan, A., Marchiori, F., and Peggion, E. (1989) Conformation and ion binding properties of peptides related to calcium binding domain III of bovine brain calmodulin, *Biopolymers* 28, 353-369.
131. 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.

132. 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.
133. Linse, S., Johansson, C., Brodin, P., Grundstrom, T., Drakenberg, T., and Forsen, S. (1991) Electrostatic contributions to the binding of  $\text{Ca}^{2+}$  in calbindin D9k, *Biochemistry* 30, 154-162.
134. Harrison, R. W. (1999) Integrating quantum and molecular mechanics, *Journal of Computational Chemistry* 20, 1618-1633.
135. Ye, Y., Shealy, S., Lee, H. W., Torshin, I., Harrison, R., and Yang, J. J. (2003) A grafting approach to obtain site-specific metal-binding properties of EF-hand proteins, *Protein Eng* 16, 429-434.
136. 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.
137. Bertini, I., Lee, Y. M., Luchinat, C., Piccioli, M., and Poggi, L. (2001) Locating the metal ion in calcium-binding proteins by using cerium(III) as a probe, *Chembiochem* 2, 550-558.
138. Mildvan, A. S., and Cohn, M. (1970) Aspects of enzyme mechanisms studies by nuclear spin relaxation induced by paramagnetic probes, *Advances in enzymology and related areas of molecular biology* 33, 1-70.
139. Linse, S., Teleman, O., and Drakenberg, T. (1990)  $\text{Ca}^{2+}$  binding to calbindin D9k strongly affects backbone dynamics: measurements of exchange rates of individual amide protons using  $^1\text{H}$  NMR, *Biochemistry* 29, 5925-5934.
140. Banci, L., Bertini, I., Savellini, G. G., Romagnoli, A., Turano, P., Cremonini, M. A., Luchinat, C., and Gray, H. B. (1997) Pseudocontact shifts as constraints for energy minimization and molecular dynamics calculations on solution structures of paramagnetic metalloproteins, *Proteins* 29, 68-76.
141. Bertini, I., Janik, M. B., Lee, Y. M., Luchinat, C., and Rosato, A. (2001) Magnetic susceptibility tensor anisotropies for a lanthanide ion series in a fixed protein matrix, *J Am Chem Soc* 123, 4181-4188.
142. Bertini, I., Donaire, A., Jimenez, B., Luchinat, C., Parigi, G., Piccioli, M., and Poggi, L. (2001) Paramagnetism-based versus classical constraints: an analysis of the solution structure of Ca Ln calbindin D9k, *J Biomol NMR* 21, 85-98.
143. Feeny, J., Birdsall, B., Bradbury, A. F., Biekofsky, R. R., and Bayley, P. M. (2001) Calmodulin tagging provides a general method of using lanthanide induced magnetic field orientation to observe residual dipolar couplings in proteins in solution, *J Biomol NMR* 21, 41-48.
144. Biekofsky, R. R., Muskett, F. W., Schmidt, J. M., Martin, S. R., Browne, J. P., Bayley, P. M., and Feeney, J. (1999) NMR approaches for monitoring domain orientations in calcium-binding proteins in solution using partial replacement of  $\text{Ca}^{2+}$  by  $\text{Tb}^{3+}$ , *FEBS Lett* 460, 519-526.



145. Tolman, J. R., Flanagan, J. M., Kennedy, M. A., and Prestegard, J. H. (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution, *Proc Natl Acad Sci U S A* 92, 9279-9283.
146. Tjandra, N., and Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science* 278, 1111-1114.
147. Jain, N. U., Noble, S., and Prestegard, J. H. (2003) Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings, *J Mol Biol* 328, 451-462.
148. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy, *Biochemistry* 31, 1647-1651.
149. Wuthrich, K. (1986) *NMR of Proteins and Nucleic acids*, John Wiley & Sons, New York.
150. Redfield, C. (1993) Resonance assignment strategies for small proteins, *NMR of Macromolecules, A Practical Approach*.
151. Wüthrich, K. (1986) *NMR of proteins and nucleic acids*, John Wiley & Sons, New York Chichester Brisbane Toronto Singapore.
152. Vuister, G., and Bax, A. (1993) Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J(HNHA) Coupling Constants in <sup>15</sup>N-Enriched Proteins, *JACS* 115, 7772-7777.
153. Chou, J. J., Li, S., and Bax, A. (2000) Study of conformational rearrangement and refinement of structural homology models by the use of heteronuclear dipolar couplings, *J Biomol NMR* 18, 217-227.
154. Barbieri, R., Bertini, I., Lee, Y. M., Luchinat, C., and Velders, A. H. (2002) Structure-independent cross-validation between residual dipolar couplings originating from internal and external orienting media, *J Biomol NMR* 22, 365-368.
155. Baig, I., Bertini, I., Del Bianco, C., Gupta, Y. K., Lee, Y. M., Luchinat, C., and Quattrone, A. (2004) Paramagnetism-based refinement strategy for the solution structure of human alpha-parvalbumin, *Biochemistry* 43, 5562-5573.
156. Gay, G. L., Lindhout, D. A., and Sykes, B. D. (2004) Using lanthanide ions to align troponin complexes in solution: order of lanthanide occupancy in cardiac troponin C, *Protein Sci* 13, 640-651.
157. Wohnert, J., Franz, K. J., Nitz, M., Imperiali, B., and Schwalbe, H. (2003) Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings, *J Am Chem Soc* 125, 13338-13339.
158. Ma, C., and Opella, S. J. (2000) Lanthanide ions bind specifically to an added "EF-hand" and orient a membrane protein in micelles for solution NMR spectroscopy, *J Magn Reson* 146, 381-384.

159. Allegrozzi, M., Bertini, I., Janik, M. B. L., Lee, Y.-M., Liu, G., and Luchinat, C. (2000) Lanthanide-Induced Pseudocontact Shifts for Solution Structure Refinements of Macromolecules in Shells up to 40 Å from the Metal Ion *J Am Chem Soc* 122.
160. Arnesano, F., Banci, L., Bertini, I., and Felli, I. C. (1998) The solution structure of oxidized rat microsomal cytochrome b5, *Biochemistry* 37, 173-184.
161. Hosszu, L. L., Craven, C. J., Parker, M. J., Lorch, M., Spencer, J., Clarke, A. R., and Waltho, J. P. (1997) Structure of a kinetic protein folding intermediate by equilibrium amide exchange, *Nat Struct Biol* 4, 801-804.
162. 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.
163. Parker, M. J., and Clarke, A. R. (1997) Amide backbone and water-related H/D isotope effects on the dynamics of a protein folding reaction, *Biochemistry* 36, 5786-5794.
164. Parker, M. J., Dempsey, C. E., Hosszu, L. L., Waltho, J. P., and Clarke, A. R. (1998) Topology, sequence evolution and folding dynamics of an immunoglobulin domain, *Nat Struct Biol* 5, 194-198.
165. Parker, M. J., Dempsey, C. E., Lorch, M., and Clarke, A. R. (1997) Acquisition of native beta-strand topology during the rapid collapse phase of protein folding, *Biochemistry* 36, 13396-13405.
166. Parker, M. J., Lorch, M., Sessions, R. B., and Clarke, A. R. (1998) Thermodynamic properties of transient intermediates and transition states in the folding of two contrasting protein structures, *Biochemistry* 37, 2538-2545.
167. 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.
168. Monera, O. D., Shaw, G. S., Zhu, B. Y., Sykes, B. D., Kay, C. M., and Hodges, R. S. (1992) Role of interchain alpha-helical hydrophobic interactions in Ca<sup>2+</sup> affinity, formation, and stability of a two-site domain in troponin C, *Protein Sci* 1, 945-955.
169. Marsden, B. J., Hodges, R. S., and Sykes, B. D. (1988) <sup>1</sup>H-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.
170. Siedlecka, M., Goch, G., Ejchart, A., Sticht, H., and Bierzyski, A. (1999) Alpha-helix nucleation by a calcium-binding peptide loop, *Proc Natl Acad Sci U S A* 96, 903-908.
171. Wilkins, D. K., Grimshaw, S. B., Receveur, V., Dobson, C. M., Jones, J. A., and Smith, L. J. (1999) Hydrodynamic radii of native and denatured

- proteins measured by pulse field gradient NMR techniques, *Biochemistry* 38, 16424-16431.
172. Mistry, N., Ismail, I. M., Farrant, R. D., Liu, M., Nicholson, J. K., and Lindon, J. C. (1999) Impurity profiling in bulk pharmaceutical batches using  $^{19}\text{F}$  NMR spectroscopy and distinction between monomeric and dimeric impurities by NMR-based diffusion measurements, *J Pharm Biomed Anal* 19, 511-517.
  173. Mayo, K. H., and Ilyina, E. (1998) A folding pathway for betapep-4 peptide 33mer: from unfolded monomers, *Protein Sci* 7, 358-368.
  174. Nyden, M., and Soderman, O. (1998) An NMR Self-Diffusion Investigation of Aggregation Phenomena in Solutions of Ethyl(hydroxyethyl)cellulose, *Macromolecules* 31, 4990-5002.
  175. Dingley, A. J., Mackay, J. P., Chapman, B. E., Morris, M. B., Kuchel, P. W., Hambly, B. D., and King, G. F. (1995) Measuring protein self-association using pulsed-field-gradient NMR spectroscopy: application to myosin light chain 2, *J Biomol NMR* 6, 321-328.
  176. Lin, M., and Larive, C. K. (1995) Detection of insulin aggregates with pulsed-field gradient nuclear magnetic resonance spectroscopy, *Anal Biochem* 229, 214-220.
  177. Nesmelova, I. V., and Fedotov, V. D. (1998) Self-diffusion and self-association of lysozyme molecules in solution, *Biochim Biophys Acta* 1383, 311-316.
  178. Chen, A., Johnson, C. S., Lin, M., and Shapiro, M. J. (1998) Chemical Exchange in Diffusion NMR Experiments, *J. Am. Chem. Soc.* 120, 9094-9095.
  179. Pfeuffer, J., Broer, S., Broer, A., Lechte, M., Flogel, U., and Leibfritz, D. (1998) Expression of aquaporins in *Xenopus laevis* oocytes and glial cells as detected by diffusion-weighted  $^1\text{H}$  NMR spectroscopy and photometric swelling assay, *Biochim Biophys Acta* 1448, 27-36.
  180. Buchko, G. W., Rozek, A., D, W. H., Cushley, R. J., and Kennedy, M. A. (1998) The use of sodium dodecyl sulfate to model the apolipoprotein environment. Evidence for peptide-SDS complexes using pulsed-field-gradient NMR spectroscopy, *Biochim Biophys Acta* 1392, 101-108.
  181. Mansfield, S. L., Jayawickrama, D. A., Timmons, J. S., and Larive, C. K. (1998) Measurement of peptide aggregation with pulsed-field gradient nuclear magnetic resonance spectroscopy, *Biochim Biophys Acta* 1382, 257-265.
  182. Biekofsky, R. R., Martin, S. R., Browne, J. P., Bayley, P. M., and Feeney, J. (1998)  $\text{Ca}^{2+}$  coordination to backbone carbonyl oxygen atoms in calmodulin and other EF-hand proteins:  $^{15}\text{N}$  chemical shifts as probes for monitoring individual-site  $\text{Ca}^{2+}$  coordination, *Biochemistry* 37, 7617-7629.
  183. Teller, D. C., Swanson, E., and Dehaen, C. (1979) *Methods Enzymol.* 61, 103-124.

184. Lee, H. W., Yang, W., Ye, Y., Liu, Z., Glushka, J., and Yang, J. J. (2002) Isolated EF-loop III of calmodulin in a scaffold protein remains unpaired in solution using pulsed-field-gradient NMR spectroscopy, *Biochim Biophys Acta* 1598, 80-87.
185. Franchini, P. L., and Reid, R. E. (1999) A model for circular dichroism monitored dimerization and calcium binding in an EF-hand synthetic peptide, *J Theor Biol* 199, 199-211.
186. Berggard, T., Julenius, K., Ogard, A., Drakenberg, T., and Linse, S. (2001) Fragment complementation studies of protein stabilization by hydrophobic core residues, *Biochemistry* 40, 1257-1264.
187. Linse, S., Voorhies, M., Norstrom, E., and Schultz, D. A. (2000) An EF-hand phage display study of calmodulin subdomain pairing, *J Mol Biol* 296, 473-486.
188. Seamon, K. B., and Kretsinger, R. H. (1983) in *Calcium in Biology*. ( Spiro, T. S. ed.), *John Wiley & Sons*, 1-51.
189. Heizmann, C. W., and Cox, J. A. (1998) New perspectives on S100 proteins: a multi-functional Ca(2+)-, Zn(2+)- and Cu(2+)-binding protein family, *Biometals* 11, 383-397.

## Appendix 2.1 NMRPipe Processing Scripts for IPAPHSQC

### Convert Varian data format to NMRPipe

```
#!/bin/csh
var2pipe -in ../fid \
  -xN 2048 -yN 1024 \
  -xT 1024 -yT 256 \
  -xMODE Complex -yMODE Complex \
  -xSW 9852.217 -ySW 2344.00 \
  -xOBS 799.742 -yOBS 81.045 \
  -xCAR 4.766 -yCAR 119.828 \
  -xLAB HN -yLAB N \
  -ndim 2 -aq2D States \
  | nmrPipe -fn QMIX -ic 4 -oc 2 -cList 1.0 0 0 1.0 0 0 0 0 \
  -out gnhsqipap_inphase.fid -verb -ov \

var2pipe -in ../fid \
  -xN 2048 -yN 1024 \
  -xT 1024 -yT 256 \
  -xMODE Complex -yMODE Complex \
  -xSW 9852.217 -ySW 2344.00 \
  -xOBS 799.742 -yOBS 81.045 \
  -xCAR 4.766 -yCAR 119.828 \
  -xLAB HN -yLAB N \
  -ndim 2 -aq2D States \
  | nmrPipe -fn QMIX -ic 4 -oc 2 -cList 0 0 0 0 1.0 0 0 1.0 \
  -out gnhsqipap_antiphase.fid -verb -ov \

var2pipe -in ../fid \
  -xN 2048 -yN 1024 \
  -xT 1024 -yT 256 \
  -xMODE Complex -yMODE Complex \
  -xSW 9852.217 -ySW 2344.00 \
  -xOBS 799.742 -yOBS 81.045 \
  -xCAR 4.766 -yCAR 119.828 \
  -xLAB HN -yLAB N \
  -ndim 2 -aq2D States \
  | nmrPipe -fn QMIX -ic 4 -oc 2 -cList 1.0 0 0 1.0 0 -1.1 1.1 0 \
  -out gnhsqipap_sum.fid -verb -ov \

var2pipe -in ../fid \
  -xN 2048 -yN 1024 \
  -xT 1024 -yT 256 \
  -xMODE Complex -yMODE Complex \
  -xSW 9852.217 -ySW 2344.00 \
  -xOBS 799.742 -yOBS 81.045 \
  -xCAR 4.766 -yCAR 119.828 \
  -xLAB HN -yLAB N \
  -ndim 2 -aq2D States \
  | nmrPipe -fn QMIX -ic 4 -oc 2 -cList 1.0 0 0 1.0 0 1.1 -1.1 0 \
  -out gnhsqipap_dif.fid -verb -ov \
```

### Fourier Transformation and Data Processing

```
#!/bin/csh
nmrPipe -in gnhsqipap_inphase.fid \
  | nmrPipe -fn SOL \
  | nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 0.5 \
  | nmrPipe -fn ZF -size 2048 \
  | nmrPipe -fn FT \
  | nmrPipe -fn PS -p0 180.0 -p1 0.0 -di \
  | nmrPipe -fn EXT -x1 6.00ppm -xn 11.00ppm -verb 2 -sw \
  | nmrPipe -fn TP \
  | nmrPipe -fn LP -x1 1 -xn 256 -pred 256 -after -verb \
  | nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 1.0 \
  | nmrPipe -fn ZF -size 2048 \
  | nmrPipe -fn PS -p0 0 -p1 0 \
  | nmrPipe -fn FT \
  | nmrPipe -fn PS -p0 0.0 -p1 0.0 -di \
```

```

nmrPipe -fn TP \
nmrPipe -fn EXT -y1 100.00ppm -yn 132.00ppm -sw \
nmrPipe -out gnhsqciapap_inphase.ft2 -verb 2 -ov

nmrPipe -in gnhsqciapap_antiphase.fid \
nmrPipe -fn SOL \
nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 0.5 \
nmrPipe -fn ZF -size 2048 \
nmrPipe -fn FT \
nmrPipe -fn PS -p0 180.0 -p1 0.0 -di \
nmrPipe -fn EXT -x1 6.00ppm -xn 11.00ppm -verb 2 -sw \
nmrPipe -fn TP \
nmrPipe -fn LP -x1 1 -xn 256 -pred 256 -after -verb \
nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 1.0 \
nmrPipe -fn ZF -size 2048 \
nmrPipe -fn PS -p0 0 -p1 0 \
nmrPipe -fn FT \
nmrPipe -fn PS -p0 -0.0 -p1 0.0 -di \
nmrPipe -fn TP \
nmrPipe -fn EXT -y1 100.00ppm -yn 132.00ppm -sw \
nmrPipe -out gnhsqciapap_antiphase.ft2 -verb 2 -ov

nmrPipe -in gnhsqciapap_sum.fid \
nmrPipe -fn SOL \
nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 0.5 \
nmrPipe -fn ZF -size 2048 \
nmrPipe -fn FT \
nmrPipe -fn PS -p0 180.0 -p1 0.0 -di \
nmrPipe -fn EXT -x1 6.00ppm -xn 11.00ppm -verb 2 -sw \
nmrPipe -fn TP \
nmrPipe -fn LP -x1 1 -xn 256 -pred 256 -after -verb \
nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 1.0 \
nmrPipe -fn ZF -size 2048 \
nmrPipe -fn PS -p0 0 -p1 0 \
nmrPipe -fn FT \
nmrPipe -fn PS -p0 0.0 -p1 0.0 -di \
nmrPipe -fn TP \
nmrPipe -fn EXT -y1 100.00ppm -yn 132.00ppm -sw \
nmrPipe -out gnhsqciapap_sum.ft2 -verb 2 -ov

nmrPipe -in gnhsqciapap_dif.fid \
nmrPipe -fn SOL \
nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 0.5 \
nmrPipe -fn ZF -size 2048 \
nmrPipe -fn FT \
nmrPipe -fn PS -p0 180.0 -p1 0.0 -di \
nmrPipe -fn EXT -x1 6.00ppm -xn 11.00ppm -verb 2 -sw \
nmrPipe -fn TP \
nmrPipe -fn LP -x1 1 -xn 256 -pred 256 -after -verb \
nmrPipe -fn SP -off 0.5 -end 0.99 -pow 2 -c 1.0 \
nmrPipe -fn ZF -size 2048 \
nmrPipe -fn PS -p0 0 -p1 0 \
nmrPipe -fn FT \
nmrPipe -fn PS -p0 0.0 -p1 0.0 -di \
nmrPipe -fn TP \
nmrPipe -fn EXT -y1 100.00ppm -yn 132.00ppm -sw \
nmrPipe -out gnhsqciapap_dif.ft2 -verb 2 -ov

```

## Appendix 2.2 Input files and Scripts for CYANA

### The Run Script for CYANA

#### CALC.cya

```
# ----- read input files -----

read xplor 5g3-52-noe-cyana2-cns-cyana2.tbl
# read upper distance bounds
read aco 06apr-talos-cyana2.aco
# read dihedral angle restraints

# ----- structure calculation -----

calc_all 100 steps=10000
# calculate conformers

# ----- structure calculation -----

overview 5g3-52.ovw structures=20 pdb cor
# write overview file and 20 best conformers
```

#### calc\_all.cya

```
# Copyright (c) 2002-05 Peter Guntert. All rights reserved.
## 7MACROS: calc_all - CYANA macro
##
## Parameters: structures=n                (default: all selected structures)
##              command=command            (default: anneal)
##              parameters
##
## Calculates a group of structures using the given command (with optional
## parameters) for each individual conformer. If the number of structures n
## is specified, the calculation will be performed starting from n random
## start conformers; otherwise the calculation is performed for all selected
## structures. Structure calculations are performed in parallel, if
## possible.

var i info echo t params name

syntax structures=0<=@i=0 command=*anneal ** serial broadcast

t = walltime
echo := off
info := minimal

if (structures.gt.0) random_all $structures
if (nstruct.eq.0) error "No structures selected."

params:=
do i 1 nparam
  params:=$params $p$i
end do

if (serial) then
  forall save skip $command $params
else
  if (broadcast) then
    name := d$getpid
    if (master) write ang $name.ang all info=none
    synchronize
    read ang $name.ang info=none
    synchronize
    if (master) remove $name.ang
  end if
  forall parallel save skip $command $params
end if
```

```

if (structures.gt.0 .and. nstruct.lt.structures) then
  do i 1 nstruct
    structure copy istrict(i) i
  end do
  structures select 1..$nstruct info=full
end if

t=walltime-t
print "      $nstruct structures finished in $t s (${t/nstruct} s/structure)."
```

## anneal.cya

```

# Copyright (c) 2002-05 Peter Guntert. All rights reserved.
## 7MACROS: calc_all - CYANA macro
##
## Parameters: structures=n           (default: all selected structures)
##             command=command       (default: anneal)
##             parameters
##
## Calculates a group of structures using the given command (with optional
## parameters) for each individual conformer. If the number of structures n
## is specified, the calculation will be performed starting from n random
## start conformers; otherwise the calculation is performed for all selected
## structures. Structure calculations are performed in parallel, if
## possible.

var i info echo t params name
syntax structures=0<=@i=0 command=*=anneal ** serial broadcast
t = walltime
echo := off
info := minimal
if (structures.gt.0) random_all $structures
if (nstruct.eq.0) error "No structures selected."
params:=
do i 1 nparam
  params:=$params $p$i
end do

if (serial) then
  forall save skip $command $params
else
  if (broadcast) then
    name := d$getpid
    if (master) write ang $name.ang all info=none
    synchronize
    read ang $name.ang info=none
    synchronize
    if (master) remove $name.ang
  end if
  forall parallel save skip $command $params
end if
if (structures.gt.0 .and. nstruct.lt.structures) then
  do i 1 nstruct
    structure copy istrict(i) i
  end do
  structures select 1..$nstruct info=full
end if

t=walltime-t
print "      $nstruct structures finished in $t s (${t/nstruct} s/structure)."
```

## Dihedral angle restraints

```

# 1 ARG PSI -105.0 225.0
# 2 ASP PHI -72.3 -52.3      # If talos is good, +- 15 of talos
# 2 ASP PSI -72.0 -12.0# If talos is new, but procheck is OK
# 3 SER PHI -137.3 -77.3     # also +- 15 of talso, or SPECIFIC EACH
# 3 SER PSI 106.0 166.0
5 THR PHI -120.0 -90.0 # talos good, procheck ok
5 THR PSI 115.0 150.0 # talos good, procheck 145 to 150
```



6 VAL	PHI	-140.0	-110.0	# talos good, procheck ok
6 VAL	PSI	118.0	148.0	# talos good, procheck ok
7 TRP	PHI	-123.0	-93.0	# talos good, procheck ok
7 TRP	PSI	120.0	150.0	# talos good, procheck ok
8 GLY	PHI	-140.0	-100.0	# talos new, procheck +-20
8 GLY	PSI	135.0	165.0	# talos new, procheck ok
9 ALA	PHI	-120.0	-90.0	# talos new, procheck ok
9 ALA	PSI	120.0	150.0	# talos new, procheck ok
10 LEU	PHI	-68.0	-38.0	# talos new, procheck ok
10 LEU	PSI	118.0	148.0	# talos new, procheck ok
11 GLY	PHI	62.0	122.0	# talos new, +-30
11 GLY	PSI	-43.0	17.0	# talos new, +-30
12 HIS	PHI	-120.0	-90.0	# talos good, procheck ok
12 HIS	PSI	114.0	174.0	# talos good, +-30
# 13 GLY	PHI	60.0	80.0	
# 13 GLY	PSI			
14 ILE	PHI	-158.0	-68.0	# talos good, +-45
14 ILE	PSI	135.0	165.0	# talos good, procheck ok
15 ASN	PHI	-127.0	-97.0	# talos good, procheck ok
15 ASN	PSI	111.0	141.0	# talos good, procheck ok
16 LEU	PHI	-106.0	-76.0	# talos good, procheck ok
16 LEU	PSI	105.0	135.0	# talos good, procheck ok
17 ASN	PHI	-132.0	-92.0	# talos new, procheck +-20
17 ASN	PSI	112.0	152.0	# talos new, procheck +-20
18 ILE	PHI	-92.0	-62.0	# talos good, procheck ok
18 ILE	PSI	117.0	157.0	# talos good, procheck +-20
# 19 PRO	PHI	-72.0	-42.0	# talos new, procheck ok
19 PRO	PSI	159.0	129.0	# talos new, procheck ok
20 ASN	PHI	37.0	77.0	
20 ASN	PSI	12.0	52.0	
21 PHE	PHI	-148.0	-108.0	# talos new, procheck +-20
21 PHE	PSI	115.0	145.0	# talos new, procheck ok
22 GLN	PHI	-146.0	-86.0	# talos good, +-30
22 GLN	PSI	117.0	147.0	# talos good, procheck ok
23 MET	PHI	-95.0	-65.0	# talos good, procheck ok
23 MET	PSI	117.0	147.0	# talos good, procheck ok
24 THR	PHI	-142.0	-82.0	# talos good, +-30
24 THR	PSI	157.0	187.0	# talos good, procheck ok
25 ASP	PHI	-76.0	-46.0	# talos good, procheck ok
25 ASP	PSI	-38.0	-8.0	# talos good, procheck ok
26 ASP	PHI	-99.0	-69.0	# talos new, procheck ok
26 ASP	PSI	-30.0	-0.1	# talos new, procheck ok
27 ILE	PHI	-116.0	-86.0	# talos good, procheck ok
27 ILE	PSI	90.0	130.0	# talos good, procheck +-20
28 ASP	PHI	-132.0	-92.0	# talos new, procheck +-20
# 28 ASP	PSI	116.0	166.0	
29 GLU	PHI	-160.0	-120.0	# talos good, procheck +-20
29 GLU	PSI	133.0	163.0	# talos good, procheck ok
30 VAL	PHI	-148.0	-118.0	# talos good, procheck ok
30 VAL	PSI	121.0	151.0	# talos good, procheck ok
31 ARG	PHI	-135.0	-105.0	# talos good, procheck ok
31 ARG	PSI	118.0	148.0	# talos good, procheck ok
32 TRP	PHI	-135.0	-105.0	# talos good, procheck ok
32 TRP	PSI	120.0	150.0	# talos good, procheck ok
33 GLU	PHI	-150.0	-120.0	# talos good, procheck ok
33 GLU	PSI	125.0	155.0	# talos good, procheck ok
34 ARG	PHI	-142.0	-62.0	# talos good, +-40
34 ARG	PSI	106.0	136.0	# talos good, procheck ok
35 GLY	PHI	53.0	83.0	# talos new, procheck ok
35 GLY	PSI	-135.0	-105.0	# talos new, procheck ok
36 SER	PHI	-112.0	-82.0	# talos new, procheck ok
36 SER	PSI	-33.0	37.0	# talos new, +-35
37 THR	PHI	-137.0	-57.0	# talos good, +-40
37 THR	PSI	115.0	145.0	# talos good, procheck ok
38 LEU	PHI	-91.0	-61.0	# talos good, procheck ok
38 LEU	PSI	116.0	146.0	# talos good, procheck ok

```

39 VAL PHI -112.0 -82.0 # talos new, procheck ok
39 VAL PSI -45.0 -15.0 # talos new, procheck ok
40 ALA PHI -172.0 -142.0 # talos good, procheck ok
40 ALA PSI 136.0 166.0 # talos good, procheck ok
41 GLU PHI -151.0 -121.0 # talos good, procheck ok
41 GLU PSI 131.0 161.0 # talos good, procheck ok
42 PHE PHI -142.0 -102.0 # talos good, procheck +-20
42 PHE PSI 116.0 146.0 # talos good, procheck ok
43 LYS PHI -145.0 -85.0 # talos good, +-30
43 LYS PSI 110.0 140.0 # talos good, procheck ok
44 ARG PHI -84.0 -44.0 # talos new, procheck +-20
# 44 ARG PSI -43.0 -3.0
45 LYS PHI -95.0 -65.0 # talos new, procheck ok
# 45 LYS PSI -36.0 4.0
46 MET PHI -152.0 -72.0 # talos new, +-40
46 MET PSI 110.0 180.0 # talos new, +-35
47 LYS PHI -105.0 -75.0 # talos good, procheck ok
47 LYS PSI 135.0 165.0 # talos good, procheck ok
# 48 PRO PHI -84.0 -44.0 # talos new, procheck +-20
48 PRO PSI 133.0 163.0 # talos new, procheck ok
49 PHE PHI -118.0 -88.0 # talos new, procheck ok
49 PHE PSI 125.0 151.0 # talos new, procheck ok
50 LEU PHI -131.0 -61.0 # talos new, +-35
50 LEU PSI 112.0 142.0 # talos new, procheck ok
51 LYS PHI -126.0 -46.0 # talos new, +-40
# 51 LYS PHI -205.0 95.0
# 52 SER PHI -215.0 95.0
# 52 SER PSI -115.0 235.0
56 ASP PHI -96.0 -56.0
# 56 ASP PSI 104.0 144.0
57 LYS PHI -98.0 -58.0
57 LYS PSI -42.0 -2.0
58 ASP PHI -100.0 -50.0 #
58 ASP PSI -37.0 10.0 #
# 59 GLY PHI -166.0 -126.0
# 59 GLY PSI -106.0 -66.0
60 ASN PHI -107.0 -67.0
60 ASN PSI -25.0 20.0 #
61 GLY PHI 74.0 114.0
61 GLY PSI -25.0 15.0
62 TYR PHI -140.0 -84.0 #
62 TYR PSI 120.0 160.0
63 ILE PHI -145.0 -93.0 #
63 ILE PSI 113.0 165.0 #
64 SER PHI -107.0 -67.0
# 64 SER PSI 109.0 149.0
# 65 ALA PHI -100.0 -60.0
65 ALA PSI -40.0 4.0 #
66 ALA PHI -89.0 -49.0
66 ALA PHI -57.0 -7.0 #
67 GLU PHI -81.0 -41.0
67 GLU PSI -59.0 -19.0
71 ALA PHI -92.0 -62.0 # talos new, procheck ok
# 71 ALA PSI 114.0 154.0
72 PHE PHI -144.0 -114.0 # talos good, procheck ok
72 PHE PSI 135.0 165.0 # talos good, procheck ok
73 GLU PHI -162.0 -132.0 # talos good, procheck ok
73 GLU PSI 129.0 159.0 # talos good, procheck ok
74 ILE PHI -124.0 -94.0 # talos good, procheck ok
74 ILE PSI 108.0 138.0 # talos good, procheck ok
75 LEU PHI -95.0 -65.0 # talos new, procheck ok
75 LEU PSI 146.0 176.0 # talos new, procheck ok
76 ALA PHI -81.0 -41.0 # talos good, procheck +-20
76 ALA PSI -41.0 -11.0 # talos good, procheck ok
77 ASN PHI -110.0 -80.0 # talos good, procheck ok
77 ASN PSI -14.0 16.0 # talos good, procheck ok

```

78 GLY	PHI	67.0	97.0	# talos new, procheck ok
78 GLY	PSI	-9.0	21.0	# talos new, procheck ok
79 ASP	PHI	-114.0	-84.0	# talos new, procheck ok
79 ASP	PSI	137.0	167.0	# talos new, procheck ok
80 LEU	PHI	-152.0	-82.0	# talos new, +-35
80 LEU	PSI	91.0	131.0	# talos new, procheck +-20
81 LYS	PHI	-130.0	-100.0	# talos good, procheck ok
81 LYS	PSI	111.0	141.0	# talos good, procheck ok
82 ILE	PHI	-120.0	-80.0	# talos good, procheck +-20
82 ILE	PSI	113.0	143.0	# talos good, procheck ok
83 LYS	PHI	-81.0	-51.0	# talos new, procheck ok
83 LYS	PSI	-48.0	-18.0	# talos new, procheck ok
# 84 ASN	PHI	-109.0	-69.0	
# 84 ASN	PSI	-25.0	15.0	
85 LEU	PHI	-93.0	-53.0	# talos new, procheck +-20
85 LEU	PSI	122.0	152.0	# talos new, procheck ok
86 THR	PHI	-142.0	-62.0	# talos good, +-40
86 THR	PSI	135.0	180.0	# talos good, +- ??
87 ARG	PHI	-74.0	-44.0	# talos new, procheck ok
87 ARG	PSI	-51.0	-21.0	# talos new, procheck ok
88 ASP	PHI	-85.0	-55.0	# talos good, procheck ok
88 ASP	PSI	-60.0	0.0	# talos good, procheck +-30
89 ASP	PHI	-98.0	-58.0	# talos new, procheck +-20
89 ASP	PSI	-57.0	3.0	# talos new, procheck +-30
90 SER	PHI	-110.0	-80.0	# talos new, procheck ok
90 SER	PSI	143.0	173.0	# talos new, procheck ok
# 91 GLY	PHI			
# 91 GLY	PSI	112.0	152.0	
92 THR	PHI	-134.0	-74.0	# talos good, procheck +-30
92 THR	PSI	113.0	143.0	# talos good, procheck ok
93 TYR	PHI	-122.0	-92.0	# talos good, procheck ok
93 TYR	PSI	110.0	150.0	# talos good, procheck +-20
94 ASN	PHI	-139.0	-109.0	# talos good, procheck ok
94 ASN	PSI	118.0	148.0	# talos good, procheck ok
95 VAL	PHI	-129.0	-99.0	# talos good, procheck ok
95 VAL	PSI	114.0	144.0	# talos good, procheck ok
96 THR	PHI	-145.0	-85.0	# talos good, procheck +-30
96 THR	PSI	117.0	147.0	# talos good, procheck ok
97 VAL	PHI	-133.0	-93.0	# talos good, procheck +-20
97 VAL	PSI	119.0	149.0	# talos good, procheck ok
98 TYR	PHI	-132.0	-102.0	# talos good, procheck ok
98 TYR	PSI	103.0	143.0	# talos good, procheck ok
99 SER	PHI	-99.0	-69.0	# talos good, procheck ok
99 SER	PSI	140.0	175.0	# talos good, procheck +25
100 THR	PHI	-78.0	-48.0	# talos new, procheck ok
100 THR	PSI	-37.0	-7.0	# talos new, procheck ok
101 ASN	PHI	-102.0	-72.0	# talos new, procheck ok
101 ASN	PSI	-25.0	25.0	# talos new, procheck +-25
102 GLY	PHI	60.0	120.0	# talos new, procheck +-30
102 GLY	PSI	-30.0	30.0	# talos new, procheck +-30
103 THR	PHI	-108.0	-78.0	# talos good, procheck ok
103 THR	PSI	111.0	141.0	# talos good, procheck ok
104 ARG	PHI	-93.0	-63.0	# talos good, procheck ok
104 ARG	PSI	91.0	171.0	# talos good, +-40
105 ILE	PHI	-138.0	-48.0	# talos new, +-45
# 105 ILE	PSI	107.0	147.0	
106 LEU	PHI	-160.0	-120.0	# talos good, procheck +-20
106 LEU	PSI	121.0	151.0	# talos good, procheck ok
107 ASN	PHI	-150.0	-50.0	# talos good, +-50
107 ASN	PSI	100.0	170.0	# talos good, +-35
108 LYS	PHI	-152.0	-122.0	# talos good, procheck ok
108 LYS	PSI	129.0	159.0	# talos good, procheck ok
109 ALA	PHI	-128.0	-98.0	# talos good, procheck ok
109 ALA	PSI	122.0	153.0	# talos good, procheck ok
110 LEU	PHI	-146.0	-116.0	# talos good, procheck ok
110 LEU	PSI	124.0	154.0	# talos good, procheck ok

```

111 ASP PHI -124.0 -94.0      # talos good, procheck ok
111 ASP PSI 110.0 140.0      # talos good, procheck ok
112 LEU PHI -119.0 -89.0     # talso good, procheck ok
112 LEU PSI 110.0 140.0     # talos good, procheck ok
113 ARG PHI -129.0 -99.0     # talos good, procheck ok
113 ARG PSI 120.0 151.0     # talos good, procheck ok
114 ILE PHI -120.0 -90.0 # talos good, procheck ok
114 ILE PSI 100.0 142.0 # talos good, procheck +20
# 115 LEU PHI -92.0 -52.0
# 115 LEU PSI 116.0 156.0
# 116 GLU PHI 35.0 285.0
# 116 GLU PSI -165.0 85.0

```

## NOE distance restraints

```

! G4
assign (residue 4 and name H) (residue 3 and name HA) 3.00 0.60 0.60
assign (residue 4 and name H) (residue 4 and name HA2) 3.20 0.60 0.60
assign (residue 4 and name H) (residue 4 and name HA3) 3.00 0.60 0.60

! T5
assign (residue 5 and name QG2) (residue 5 and name HA) 2.40 0.40 0.40
assign (residue 5 and name QG2) (residue 5 and name HB) 2.30 0.40 0.40
! 2D S/N 6, sparky
assign (residue 5 and name QG2) (residue 7 and name HA) 4.00 0.40 1.00
assign (residue 5 and name QG2) (residue 7 and name HD1) 3.80 0.40 0.40
! 2D S/N 6, sparky
assign (residue 5 and name QG2) (residue 111 and name HA) 4.20 0.40 1.00
assign (residue 5 and name QG2) (residue 111 and name HB2) 3.60 0.60 0.60
assign (residue 5 and name QG2) (residue 111 and name HB3) 2.40 0.40 0.40
! NOE error
assign (residue 5 and name QG2) (residue 113 and name HD2) 2.60 0.40 0.40
assign (residue 5 and name QG2) (residue 113 and name HD3) 3.20 0.40 0.40
assign (residue 5 and name H) (residue 4 and name HA2) 2.40 0.40 0.40
assign (residue 5 and name H) (residue 4 and name HA3) 2.80 0.60 0.60
assign (residue 5 and name H) (residue 5 and name HA) 2.80 0.60 0.60
assign (residue 5 and name H) (residue 5 and name HB) 2.20 0.20 0.40
assign (residue 5 and name H) (residue 5 and name QG2) 4.00 0.60 0.60

! V6
assign (residue 6 and name HB) (residue 5 and name HA) 3.80 0.40 0.80
assign (residue 6 and name HB) (residue 6 and name HA) 3.00 0.40 0.40
! NOE error
assign (residue 6 and name HB) (residue 112 and name HA) 2.60 0.40 0.40
assign (residue 6 and name QG1) (residue 6 and name HA) 2.60 0.40 0.40
assign (residue 6 and name QG2) (residue 6 and name HA) 2.40 0.40 0.40
assign (residue 6 and name QG2) (residue 6 and name HB) 2.40 0.40 0.40
assign (residue 6 and name QG2) (residue 6 and name H) 3.40 0.40 0.40
assign (residue 6 and name H) (residue 5 and name HA) 2.00 0.20 0.40
assign (residue 6 and name H) (residue 5 and name HB) 3.20 0.60 0.80
assign (residue 6 and name H) (residue 5 and name QG2) 2.40 0.40 0.40
assign (residue 6 and name H) (residue 6 and name HA) 3.20 0.40 0.40
assign (residue 6 and name H) (residue 6 and name HB) 2.80 0.40 0.60
assign (residue 6 and name H) (residue 6 and name QG1) 4.00 0.60 0.60
! 2D, S/N 3
assign (residue 6 and name H) (residue 7 and name HD1) 4.20 0.40 1.00
! 2D, S/N 5
assign (residue 6 and name H) (residue 7 and name H) 4.00 0.40 1.00
assign (residue 6 and name H) (residue 111 and name HB3) 3.20 0.60 0.80
assign (residue 6 and name H) (residue 112 and name HA) 3.00 0.40 0.40
! Too far, its probably D28HN-HB3
! assign (residue 6 and name H) (residue 112 and name HB3) 3.40 0.60 0.80
! Flip QD1 to QD2
assign (residue 6 and name H) (residue 112 and name QD2) 4.20 0.60 0.60

! W7
assign (residue 7 and name QB) (residue 7 and name HA) 2.60 0.40 0.40
assign (residue 7 and name HD1) (residue 5 and name HB) 4.50 0.40 0.40
assign (residue 7 and name HD1) (residue 6 and name HA) 3.20 0.40 0.40
assign (residue 7 and name HD1) (residue 7 and name QB) 3.00 0.40 0.40
assign (residue 7 and name HE1) (residue 5 and name HB) 3.20 0.40 0.40
assign (residue 7 and name HE1) (residue 5 and name QG2) 3.00 0.40 0.40
assign (residue 7 and name HE1) (residue 7 and name QB) 4.00 0.40 1.00
assign (residue 7 and name HE1) (residue 7 and name HD1) 2.40 0.40 0.40
! It's 5.0, may be missed assign with other ring proton
! assign (residue 7 and name HE1) (residue 7 and name HH2) 3.40 0.40 0.40
assign (residue 7 and name HE1) (residue 7 and name HZ2) 3.00 0.40 0.40
assign (residue 7 and name HE3) (residue 7 and name HA) 3.20 0.40 0.40
assign (residue 7 and name HE3) (residue 7 and name QB) 2.60 0.40 0.40
assign (residue 7 and name HE3) (residue 7 and name HZ3) 2.60 0.40 0.40
! R113 is around 5.1

```

```

! assign (residue 7 and name HE3) (residue 115 and name HA) 3.20 0.40 0.40
assign (residue 7 and name HE3) (residue 115 and name HB2) 4.40 0.40 0.60
assign (residue 7 and name HE3) (residue 115 and name QD1) 2.50 0.40 1.00
! Kind of far, need to DCA
assign (residue 7 and name H) (residue 5 and name QG2) 4.00 0.60 1.00
assign (residue 7 and name H) (residue 6 and name HA) 2.00 0.20 0.40
assign (residue 7 and name H) (residue 6 and name HB) 3.60 0.40 0.60
assign (residue 7 and name H) (residue 6 and name QG1) 3.00 0.60 0.40
assign (residue 7 and name H) (residue 6 and name QG2) 4.00 0.40 0.40
assign (residue 7 and name H) (residue 7 and name HA) 3.20 0.60 0.80
assign (residue 7 and name H) (residue 7 and name QB) 2.40 0.40 0.40
assign (residue 7 and name H) (residue 7 and name HD1) 2.60 0.60 0.60
! It's right at limit of detection
! assign (residue 7 and name H) (residue 7 and name HE1) 3.60 0.40 0.40
! assign (residue 7 and name H) (residue 7 and name HE3) 3.40 0.40 0.40
! Its close to L112QD2
! assign (residue 7 and name H) (residue 115 and name QD1) 3.60 0.40 0.40
assign (residue 7 and name HZ2) (residue 5 and name QG2) 2.80 0.40 0.40
assign (residue 7 and name HZ2) (residue 7 and name HH2) 2.40 0.40 0.40
assign (residue 7 and name HZ2) (residue 7 and name HZ3) 4.20 0.40 0.40
assign (residue 7 and name HZ3) (residue 7 and name QB) 4.60 0.40 0.40
! assign (residue 7 and name HZ3) (residue 115 and name HA) 3.40 0.40 0.40
assign (residue 7 and name HZ3) (residue 115 and name QD1) 3.40 0.40 0.40

! G8
assign (residue 8 and name H) (residue 7 and name HA) 2.20 0.40 0.40
assign (residue 8 and name H) (residue 7 and name QB) 2.80 0.40 1.00
! assign (residue 8 and name H) (residue 7 and name HD1) 3.40 0.30 0.30
assign (residue 8 and name H) (residue 7 and name HE3) 3.60 0.40 0.40
! CYANA always has error on this NOE, lowered to 3.50
assign (residue 8 and name H) (residue 8 and name HA2) 3.50 0.60 0.60
assign (residue 8 and name H) (residue 8 and name HA3) 2.80 0.60 0.60
assign (residue 8 and name H) (residue 9 and name H) 4.00 0.40 0.60
! Close to 5.0
assign (residue 8 and name H) (residue 14 and name QD1) 3.60 0.40 1.40
! Flip QD1 with QG2
assign (residue 8 and name H) (residue 114 and name QG2) 3.20 0.60 1.00
assign (residue 8 and name H) (residue 115 and name QB) 4.60 0.40 0.50

! A9
assign (residue 9 and name QB) (residue 8 and name HA2) 4.00 0.40 0.50
! Kind of far, DCA
assign (residue 9 and name QB) (residue 8 and name HA3) 4.50 0.40 0.60
assign (residue 9 and name QB) (residue 9 and name HA) 2.60 0.40 0.40
assign (residue 9 and name QB) (residue 12 and name HB2) 2.40 0.40 0.40
assign (residue 9 and name H) (residue 8 and name HA2) 2.20 0.40 1.00
assign (residue 9 and name H) (residue 8 and name HA3) 2.20 0.40 1.00
assign (residue 9 and name H) (residue 9 and name HA) 2.80 0.60 0.60
assign (residue 9 and name H) (residue 9 and name QB) 2.40 0.40 0.40
assign (residue 9 and name H) (residue 14 and name QD1) 4.20 0.40 0.40

! L10
! This is a repeat
! assign (residue 10 and name HA) (residue 11 and name H) 2.40 0.80 0.80
assign (residue 10 and name HB2) (residue 10 and name HA) 2.60 0.40 0.40
assign (residue 10 and name HB2) (residue 10 and name H) 2.40 0.40 0.40
! This is wrong and is a repeat
! assign (residue 10 and name HB2) (residue 11 and name H) 2.40 0.80 0.80
assign (residue 10 and name HB3) (residue 10 and name HA) 3.20 0.40 0.40
assign (residue 10 and name HB3) (residue 10 and name H) 2.60 0.40 0.40
assign (residue 10 and name QD1) (residue 10 and name HA) 2.20 0.40 0.40
assign (residue 10 and name QD1) (residue 10 and name HG) 2.60 0.40 0.40
assign (residue 10 and name QD2) (residue 10 and name HA) 4.20 4.00 0.40
assign (residue 10 and name QD2) (residue 10 and name QD1) 2.00 0.20 0.40
assign (residue 10 and name QD2) (residue 87 and name HA) 4.50 0.60 0.60
assign (residue 10 and name QD2) (residue 87 and name H) 4.00 0.60 0.60
assign (residue 10 and name HG) (residue 10 and name HA) 3.00 0.80 0.80
! Increase the distance, NOE error
assign (residue 10 and name HG) (residue 10 and name H) 4.20 0.60 0.40
assign (residue 10 and name H) (residue 9 and name HA) 2.00 0.20 0.40
assign (residue 10 and name H) (residue 9 and name QB) 2.40 0.40 0.40
assign (residue 10 and name H) (residue 10 and name HA) 3.20 0.60 0.80
! These are repeats, see above
! assign (residue 10 and name H) (residue 10 and name HB2) 3.00 0.60 0.40
! assign (residue 10 and name H) (residue 10 and name HB3) 2.20 0.40 0.40
assign (residue 10 and name H) (residue 10 and name QD1) 4.20 0.40 0.40
assign (residue 10 and name H) (residue 10 and name QD2) 4.20 0.40 0.60

! G11
assign (residue 11 and name H) (residue 10 and name HA) 2.40 0.40 0.40
! Change according to NOE error
assign (residue 11 and name H) (residue 10 and name HB2) 3.60 0.60 0.60
assign (residue 11 and name H) (residue 10 and name HB3) 4.00 0.40 0.40
assign (residue 11 and name H) (residue 10 and name QD1) 2.40 0.40 0.40
! At the limit of detection, DCA
! assign (residue 11 and name H) (residue 10 and name QD2) 4.00 0.60 1.00
assign (residue 11 and name H) (residue 10 and name HG) 3.20 0.60 0.80
assign (residue 11 and name H) (residue 10 and name H) 4.60 0.40 0.40

```

assign (residue 11 and name H) (residue 11 and name QA) 2.80 0.60 0.60

! H12

assign (residue 12 and name H) (residue 9 and name QB) 3.60 0.60 0.80  
 assign (residue 12 and name H) (residue 10 and name HA) 3.60 0.60 0.80  
 assign (residue 12 and name H) (residue 11 and name QA) 3.00 0.60 0.80  
 assign (residue 12 and name H) (residue 11 and name H) 2.85 0.40 0.40  
 assign (residue 12 and name H) (residue 12 and name HA) 2.80 0.60 0.60  
 assign (residue 12 and name H) (residue 12 and name HB2) 2.60 0.60 0.60  
 ! Increase according to the NOE error  
 assign (residue 12 and name H) (residue 12 and name HB3) 3.40 0.60 0.60  
 assign (residue 12 and name H) (residue 85 and name HB2) 2.60 0.40 0.40  
 assign (residue 12 and name H) (residue 85 and name HB3) 4.40 0.60 0.60  
 assign (residue 12 and name H) (residue 85 and name H) 3.40 0.40 0.40

! G13

assign (residue 13 and name HA3) (residue 13 and name HA2) 2.20 0.40 0.40  
 assign (residue 13 and name H) (residue 12 and name HA) 2.40 0.40 0.40  
 assign (residue 13 and name H) (residue 12 and name HB2) 3.80 0.60 0.60  
 assign (residue 13 and name H) (residue 12 and name HB3) 2.20 0.40 0.40  
 assign (residue 13 and name H) (residue 13 and name HA2) 2.80 0.60 0.60  
 assign (residue 13 and name H) (residue 13 and name HA3) 2.60 0.60 0.60

! I14

assign (residue 14 and name HB) (residue 14 and name HA) 2.80 0.40 0.40  
 assign (residue 14 and name HB) (residue 14 and name H) 2.80 0.40 0.80  
 assign (residue 14 and name HB) (residue 15 and name H) 4.00 0.40 0.40  
 assign (residue 14 and name QD1) (residue 8 and name HA2) 2.55 0.40 0.40  
 assign (residue 14 and name QD1) (residue 8 and name HA3) 3.20 0.40 0.40  
 assign (residue 14 and name QD1) (residue 14 and name HA) 2.25 0.40 0.50  
 assign (residue 14 and name QD1) (residue 14 and name HB) 2.40 0.40 0.40  
 assign (residue 14 and name QD1) (residue 14 and name H) 4.00 0.40 1.00  
 assign (residue 14 and name HG12) (residue 14 and name HA) 3.80 0.40 0.40  
 assign (residue 14 and name HG13) (residue 14 and name HA) 2.80 0.40 0.40  
 assign (residue 14 and name QG2) (residue 13 and name HA3) 4.60 0.40 0.40  
 assign (residue 14 and name QG2) (residue 14 and name HA) 3.40 0.40 0.40  
 assign (residue 14 and name QG2) (residue 14 and name HB) 2.40 0.40 0.40  
 assign (residue 14 and name QG2) (residue 14 and name HG12) 2.40 0.40 0.40  
 ! Too far from L16, DCA  
 ! assign (residue 14 and name QG2) (residue 16 and name HB3) 2.60 0.40 0.40  
 assign (residue 14 and name QG2) (residue 82 and name HB) 2.60 0.70 0.40  
 assign (residue 14 and name QG2) (residue 82 and name H) 3.20 0.40 0.40  
 assign (residue 14 and name H) (residue 13 and name HA2) 2.20 0.40 0.60  
 assign (residue 14 and name H) (residue 13 and name HA3) 2.80 0.40 0.40  
 assign (residue 14 and name H) (residue 14 and name HA) 2.60 0.60 0.60  
 assign (residue 14 and name H) (residue 14 and name HB) 3.00 0.40 0.40  
 ! Increase QD1, it is at the limit of detection  
 assign (residue 14 and name H) (residue 14 and name QD1) 4.20 0.40 0.80  
 assign (residue 14 and name H) (residue 14 and name QG2) 2.60 0.40 0.40

! N15

assign (residue 15 and name HB2) (residue 15 and name HA) 2.70 0.40 0.40  
 assign (residue 15 and name HB2) (residue 15 and name H) 2.40 0.80 0.80  
 assign (residue 15 and name HD21) (residue 15 and name HB2) 2.80 0.40 0.60  
 assign (residue 15 and name HD21) (residue 15 and name HB3) 2.80 0.40 0.80  
 assign (residue 15 and name HD21) (residue 15 and name HD22) 2.20 0.40 0.40  
 assign (residue 15 and name HD22) (residue 15 and name HB2) 3.80 0.80 0.80  
 assign (residue 15 and name HD22) (residue 15 and name HB3) 3.80 0.80 0.80  
 assign (residue 15 and name H) (residue 14 and name HA) 2.20 0.40 0.40  
 ! NOE error suggest 2.58  
 assign (residue 15 and name H) (residue 14 and name QD1) 2.60 0.40 0.40  
 assign (residue 15 and name H) (residue 14 and name QG2) 3.20 0.60 1.00  
 assign (residue 15 and name H) (residue 15 and name HA) 3.20 0.60 0.80  
 assign (residue 15 and name H) (residue 15 and name HB2) 2.80 0.60 0.60  
 assign (residue 15 and name H) (residue 15 and name HB3) 3.20 0.60 0.80

! L16

assign (residue 16 and name HB3) (residue 16 and name HA) 2.60 0.40 0.40  
 ! NOE error, change to HD2#  
 assign (residue 16 and name QD2) (residue 6 and name HB) 3.40 0.40 0.40  
 ! assign (residue 16 and name QD1) (residue 15 and name HA) 3.60 0.40 0.80  
 ! Change to QD2 from QD1  
 assign (residue 16 and name QD2) (residue 16 and name HA) 2.15 0.40 0.40  
 assign (residue 16 and name QD1) (residue 16 and name HB2) 2.40 0.40 0.40  
 assign (residue 16 and name QD1) (residue 16 and name HB3) 2.60 0.40 0.40  
 ! NOE error  
 assign (residue 16 and name QD1) (residue 32 and name HZ3) 3.20 0.40 0.40  
 assign (residue 16 and name QD1) (residue 93 and name QD) 3.40 0.40 0.40  
 assign (residue 16 and name QD1) (residue 93 and name QE) 4.00 0.40 0.60  
 ! Flip QD1 to QD2  
 assign (residue 16 and name QD2) (residue 112 and name HA) 3.80 0.40 0.40  
 assign (residue 16 and name QD1) (residue 112 and name QD1) 3.20 0.40 0.40  
 ! assign (residue 16 and name QD2) (residue 14 and name HB) 3.60 0.40 0.80  
 ! assign (residue 16 and name QD2) (residue 15 and name HA) 3.60 0.40 0.80  
 assign (residue 16 and name QD2) (residue 16 and name HA) 2.20 0.40 0.40  
 ! Flip to QD1 for the next two lines  
 assign (residue 16 and name QD1) (residue 16 and name HB2) 2.40 0.40 0.40  
 assign (residue 16 and name QD1) (residue 16 and name HB3) 2.30 0.40 0.40

```

assign (residue 16 and name QD2) (residue 16 and name QD1) 2.20 0.40 0.40
! Flip to QD1 for the next two lines
assign (residue 16 and name QD1) (residue 93 and name HB2) 3.40 0.40 0.60
assign (residue 16 and name QD1) (residue 93 and name HB3) 3.40 0.40 0.60
! Too far, should be QD1, see above
! assign (residue 16 and name QD2) (residue 93 and name QE) 2.90 0.40 0.40
! Change according to NOE error
assign (residue 16 and name QD2) (residue 110 and name HB3) 3.60 0.40 0.40
assign (residue 16 and name QD2) (residue 112 and name QD1) 4.00 0.40 0.40
! assign (residue 16 and name H) (residue 14 and name QD1) 3.40 0.80 0.80
assign (residue 16 and name H) (residue 14 and name QG2) 4.00 0.40 0.60
assign (residue 16 and name H) (residue 15 and name HA) 2.40 0.40 0.40
assign (residue 16 and name H) (residue 15 and name HB2) 4.40 0.40 0.60
assign (residue 16 and name H) (residue 15 and name HB3) 3.40 0.80 0.80
assign (residue 16 and name H) (residue 16 and name HA) 3.20 0.40 0.40
assign (residue 16 and name H) (residue 16 and name HB2) 2.80 0.60 0.60
assign (residue 16 and name H) (residue 16 and name HB3) 3.60 0.40 0.40
! Possible error
assign (residue 16 and name H) (residue 16 and name QD1) 4.00 0.40 0.40
assign (residue 16 and name H) (residue 16 and name QD2) 3.60 0.40 0.40
assign (residue 16 and name H) (residue 80 and name HB2) 3.60 0.60 0.40
assign (residue 16 and name H) (residue 80 and name HB3) 3.00 0.40 0.40
assign (residue 16 and name H) (residue 81 and name HA) 3.40 0.80 0.80

! N17
assign (residue 17 and name HB2) (residue 17 and name HA) 2.40 0.40 0.40
assign (residue 17 and name HB2) (residue 18 and name H) 2.40 0.40 0.40
! Temp take out, may be too far
! assign (residue 17 and name HD21) (residue 17 and name HA) 3.60 0.80 0.80
! Open up, CYANA error
assign (residue 17 and name HD22) (residue 17 and name HB2) 2.65 0.50 0.80
assign (residue 17 and name HD21) (residue 17 and name HB3) 4.00 0.40 0.60
assign (residue 17 and name HD21) (residue 17 and name HD22) 2.20 0.40 0.40
! Open up 1.00, CYANA error
assign (residue 17 and name HD22) (residue 17 and name HA) 4.00 0.40 1.00
! assign (residue 17 and name HD21) (residue 17 and name HB2) 3.80 0.80 0.80
assign (residue 17 and name HD22) (residue 17 and name HB3) 3.80 0.40 0.40
assign (residue 17 and name H) (residue 16 and name HA) 2.20 0.40 0.40
assign (residue 17 and name H) (residue 16 and name HB2) 4.40 0.40 0.40
assign (residue 17 and name H) (residue 16 and name HB3) 3.80 0.40 0.40
! Flip QD1 to QD2
! assign (residue 17 and name H) (residue 16 and name QD1) 3.00 0.60 0.80
assign (residue 17 and name H) (residue 16 and name QD2) 4.40 0.40 0.60
assign (residue 17 and name H) (residue 16 and name H) 4.20 0.40 0.80
assign (residue 17 and name H) (residue 17 and name HA) 3.00 0.60 0.60
assign (residue 17 and name H) (residue 17 and name HB2) 3.60 0.40 0.40
assign (residue 17 and name H) (residue 17 and name HB3) 2.80 0.40 0.40

! I18
assign (residue 18 and name HB) (residue 21 and name QD) 2.80 0.60 0.60
assign (residue 18 and name QG2) (residue 18 and name HA) 2.60 0.40 0.60
assign (residue 18 and name QG2) (residue 18 and name HB) 2.60 0.40 0.40
! NOE error suggest 2.66
assign (residue 18 and name QG2) (residue 19 and name QD) 2.70 0.40 0.40
assign (residue 18 and name QG2) (residue 21 and name HB2) 2.80 0.40 0.40
assign (residue 18 and name QG2) (residue 21 and name HB3) 2.80 0.40 0.40
assign (residue 18 and name H) (residue 17 and name HA) 2.20 0.40 0.40
! Repeat and this one is not correct distance
! assign (residue 18 and name H) (residue 17 and name HB2) 3.40 0.40 0.40
assign (residue 18 and name H) (residue 17 and name HB3) 3.40 0.60 0.60
assign (residue 18 and name H) (residue 18 and name HA) 3.20 0.60 0.80
assign (residue 18 and name H) (residue 18 and name HG13) 2.60 0.60 0.90
! NOE error suggest 3.8
assign (residue 18 and name H) (residue 18 and name QG2) 4.00 0.40 0.60

! P19
assign (residue 19 and name QB) (residue 19 and name HA) 2.60 0.40 0.40

! N20
! use a different for right side portion of spectrum
assign (residue 20 and name HB2) (residue 20 and name HA) 2.80 0.40 0.40
assign (residue 20 and name HB3) (residue 20 and name HA) 2.60 0.40 0.40
assign (residue 20 and name HD21) (residue 20 and name HB2) 3.80 0.50 0.80
assign (residue 20 and name HD21) (residue 20 and name HB3) 3.40 0.60 0.40
assign (residue 20 and name HD22) (residue 20 and name HB3) 3.00 1.00 1.00
assign (residue 20 and name H) (residue 19 and name HA) 2.40 0.80 0.80
assign (residue 20 and name H) (residue 20 and name HA) 2.00 0.20 0.40
assign (residue 20 and name H) (residue 20 and name HB2) 3.60 0.80 0.80
assign (residue 20 and name H) (residue 20 and name HB3) 3.60 0.80 0.80

! F21
assign (residue 21 and name QD) (residue 21 and name HB2) 2.60 0.40 0.40
assign (residue 21 and name QD) (residue 21 and name HB3) 2.35 0.40 0.40
assign (residue 21 and name QD) (residue 21 and name QE) 2.2 0.20 0.40
! Adjusted
assign (residue 21 and name QD) (residue 22 and name HA) 4.40 0.40 0.40
! assign (residue 21 and name QD) (residue 23 and name HA) 3.20 0.40 0.40
assign (residue 21 and name QD) (residue 27 and name QD1) 4.00 0.40 0.60

```

```

! Too far
! assign (residue 21 and name QD) (residue 95 and name QG2) 3.60 0.40 0.80
! NOE error suggest 4.09
assign (residue 21 and name H) (residue 19 and name HA) 4.10 0.40 0.40
assign (residue 21 and name H) (residue 20 and name HA) 2.20 0.40 0.40
assign (residue 21 and name H) (residue 20 and name HB2) 4.20 0.40 0.40
assign (residue 21 and name H) (residue 20 and name HB3) 4.40 0.40 0.40
assign (residue 21 and name H) (residue 20 and name H) 3.20 0.40 0.40
assign (residue 21 and name H) (residue 21 and name HA) 2.80 0.60 0.60
assign (residue 21 and name H) (residue 21 and name HB2) 3.00 0.40 0.60
assign (residue 21 and name H) (residue 21 and name HB3) 2.40 0.40 0.40
assign (residue 21 and name H) (residue 21 and name QD) 4.40 0.40 0.40
assign (residue 21 and name H) (residue 22 and name H) 3.60 0.40 0.70

! Q22
! assign (residue 22 and name HB2) (residue 21 and name QD) 3.20 0.40 0.40
! Adjusted
assign (residue 22 and name HB2) (residue 22 and name HA) 2.60 0.40 0.40
! Adjusted
assign (residue 22 and name HB2) (residue 22 and name QG) 2.60 0.40 0.40
assign (residue 22 and name HB3) (residue 22 and name HA) 2.80 0.50 0.40
assign (residue 22 and name HB3) (residue 22 and name QG) 2.50 0.40 0.40
assign (residue 22 and name HE21) (residue 22 and name HE22) 2.00 0.40 0.40
! Adjusted
assign (residue 22 and name HE22) (residue 22 and name QG) 2.60 0.40 0.80
assign (residue 22 and name QG) (residue 22 and name HA) 3.00 0.40 0.40
assign (residue 22 and name H) (residue 21 and name HA) 2.00 0.20 0.40
assign (residue 22 and name H) (residue 21 and name HB2) 3.40 0.80 0.80
assign (residue 22 and name H) (residue 21 and name HB3) 3.40 0.80 0.80
assign (residue 22 and name H) (residue 22 and name HA) 3.20 0.60 0.80
! Flip HB2 with HB3
assign (residue 22 and name H) (residue 22 and name HB3) 2.60 0.40 0.40
assign (residue 22 and name H) (residue 22 and name HB2) 3.20 0.40 0.40
! Problem, increase
assign (residue 22 and name H) (residue 22 and name QG) 4.00 0.80 1.00
assign (residue 22 and name H) (residue 23 and name H) 3.60 0.40 0.80

! M23
! Take out, too far
! assign (residue 23 and name H) (residue 21 and name HD#) 3.20 0.60 0.80
assign (residue 23 and name H) (residue 22 and name HA) 2.00 0.20 0.40
assign (residue 23 and name H) (residue 22 and name HB2) 4.40 0.40 0.40
! Check NOE data height
! assign (residue 23 and name H) (residue 22 and name HB3) 3.40 0.40 0.40
assign (residue 23 and name H) (residue 22 and name QG) 3.40 0.40 0.40
assign (residue 23 and name H) (residue 23 and name HA) 3.20 0.60 0.80
assign (residue 23 and name H) (residue 23 and name HB2) 2.20 0.20 0.40

! T24
assign (residue 24 and name H) (residue 23 and name HA) 2.20 0.40 0.40
! DCA
assign (residue 24 and name H) (residue 23 and name H) 3.60 0.40 1.00
assign (residue 24 and name H) (residue 24 and name HA) 2.80 0.60 0.60
assign (residue 24 and name H) (residue 27 and name HB) 3.40 0.60 0.80
assign (residue 24 and name H) (residue 27 and name QD1) 3.60 0.60 0.80
assign (residue 24 and name H) (residue 27 and name HG12) 2.60 0.60 0.60
assign (residue 24 and name H) (residue 27 and name HG13) 2.80 0.60 0.60

! D25
assign (residue 25 and name H) (residue 24 and name HA) 2.80 0.60 0.60
assign (residue 25 and name H) (residue 25 and name HA) 2.40 0.40 0.60

! D26
assign (residue 26 and name H) (residue 25 and name HA) 3.00 0.60 0.60
assign (residue 26 and name H) (residue 26 and name HA) 2.60 0.60 0.60
assign (residue 26 and name H) (residue 26 and name QB) 2.40 0.40 0.60
assign (residue 26 and name H) (residue 27 and name HG12) 3.40 0.60 0.80

! I27
assign (residue 27 and name HB) (residue 27 and name HA) 3.20 0.40 0.60
assign (residue 27 and name QD1) (residue 27 and name HB) 3.80 0.40 0.40
assign (residue 27 and name QD1) (residue 99 and name HA) 3.20 0.40 0.40
assign (residue 27 and name HG12) (residue 23 and name HA) 3.20 0.40 0.60
assign (residue 27 and name HG12) (residue 27 and name HB) 2.60 0.40 0.40
assign (residue 27 and name HG13) (residue 27 and name HB) 2.60 0.40 0.40
! May this is I27QD1 to F21QD, 4.35
! assign (residue 27 and name QG2) (residue 21 and name QD) 3.40 0.40 0.40
assign (residue 27 and name QG2) (residue 27 and name HA) 2.50 0.40 0.40
assign (residue 27 and name QG2) (residue 27 and name HB) 2.40 0.40 0.40
assign (residue 27 and name QG2) (residue 27 and name HG13) 2.60 0.40 0.40
assign (residue 27 and name QG2) (residue 99 and name HA) 3.80 0.40 0.80
assign (residue 27 and name H) (residue 25 and name HA) 3.40 0.60 0.80
assign (residue 27 and name H) (residue 26 and name HA) 3.20 0.60 0.80
assign (residue 27 and name H) (residue 26 and name H) 2.80 0.40 0.40
assign (residue 27 and name H) (residue 27 and name HA) 3.20 0.60 0.80
assign (residue 27 and name H) (residue 27 and name HB) 2.40 0.40 0.40
assign (residue 27 and name H) (residue 27 and name QD1) 3.20 0.60 0.80
assign (residue 27 and name H) (residue 27 and name HG12) 2.40 0.40 0.40

```



```

! Increased to from 2.80 to 3.80
assign (residue 27 and name H) (residue 27 and name HG13) 3.80 0.60 0.60
assign (residue 27 and name H) (residue 27 and name QG2) 4.00 0.60 0.80

! D28
assign (residue 28 and name HB2) (residue 28 and name HA) 2.80 0.40 0.40
assign (residue 28 and name HB2) (residue 100 and name HA) 2.20 0.40 0.60
assign (residue 28 and name HB3) (residue 28 and name HA) 3.20 0.40 0.40
assign (residue 28 and name HB3) (residue 100 and name HA) 3.00 0.40 0.40
assign (residue 28 and name H) (residue 27 and name HA) 2.80 0.60 0.60
assign (residue 28 and name H) (residue 27 and name QG2) 3.40 0.40 0.40
assign (residue 28 and name H) (residue 28 and name HB2) 3.00 0.60 0.80
assign (residue 28 and name H) (residue 99 and name HA) 3.20 0.60 1.00

! E29
assign (residue 29 and name HB2) (residue 29 and name HA) 2.40 0.40 0.40
assign (residue 29 and name HB3) (residue 29 and name HA) 3.00 0.40 0.40
assign (residue 29 and name H) (residue 27 and name QG2) 2.80 0.60 0.60
assign (residue 29 and name H) (residue 28 and name HA) 3.40 0.60 0.80
assign (residue 29 and name H) (residue 28 and name HB2) 3.40 0.60 0.80
assign (residue 29 and name H) (residue 28 and name HB3) 3.40 0.60 0.80
assign (residue 29 and name H) (residue 28 and name H) 2.80 0.60 0.40
assign (residue 29 and name H) (residue 29 and name HA) 3.20 0.60 0.80
assign (residue 29 and name H) (residue 29 and name HB2) 3.20 0.60 0.80
assign (residue 29 and name H) (residue 29 and name HB3) 3.00 0.60 0.80
assign (residue 29 and name H) (residue 98 and name HB2) 3.40 0.40 0.40
assign (residue 29 and name H) (residue 98 and name HB3) 3.80 0.80 0.80
! 2D S/N 5
assign (residue 29 and name H) (residue 99 and name HA) 4.20 0.40 0.80

! V30
assign (residue 30 and name QG1) (residue 30 and name HB) 2.60 0.40 0.40
! May be miss assignment, but still use it
assign (residue 30 and name QG1) (residue 97 and name HA) 4.00 0.40 1.00
assign (residue 30 and name QG2) (residue 30 and name HB) 2.40 0.40 0.40
assign (residue 30 and name H) (residue 29 and name HA) 2.30 0.40 0.50
assign (residue 30 and name H) (residue 30 and name HA) 3.40 0.60 0.80
assign (residue 30 and name H) (residue 30 and name HB) 3.80 0.40 0.40
assign (residue 30 and name H) (residue 30 and name QG1) 2.60 0.40 0.40
! assign (residue 30 and name H) (residue 42 and name HB2) 3.60 0.60 0.80
assign (residue 30 and name H) (residue 42 and name HB3) 3.40 0.60 0.80
assign (residue 30 and name H) (residue 42 and name HD#) 3.40 0.60 0.80

! R31
! Lower the distance, NOE error
assign (residue 31 and name H) (residue 30 and name HA) 2.60 0.60 0.60
! Increase distance up 0.20, NOE error
assign (residue 31 and name H) (residue 30 and name HB) 3.80 0.40 0.40
assign (residue 31 and name H) (residue 30 and name QG1) 4.20 0.40 0.80
assign (residue 31 and name H) (residue 30 and name QG2) 2.80 0.40 0.40
assign (residue 31 and name H) (residue 31 and name HA) 3.40 0.60 0.80
assign (residue 31 and name H) (residue 31 and name HB2) 3.40 0.40 0.40
assign (residue 31 and name H) (residue 31 and name HB3) 3.40 0.40 0.40
! 2D S/N 4
assign (residue 31 and name H) (residue 41 and name HA) 4.20 0.40 0.80
! Increase distance, because R31HN-V95HG1# keep getting error
assign (residue 31 and name H) (residue 95 and name HA) 4.00 0.60 0.80
assign (residue 31 and name H) (residue 95 and name QG1) 3.40 0.40 0.40
assign (residue 31 and name H) (residue 96 and name H) 2.60 0.40 0.80
! Too far, DCA
! assign (residue 31 and name H) (residue 97 and name QG2) 3.60 0.60 0.80

! W32
assign (residue 32 and name HA) (residue 33 and name H) 2.40 0.40 0.40
! Change V30HG1# to V30HG2#; 2.40 to 2.60, NOE error
assign (residue 32 and name HD1) (residue 30 and name QG2) 2.60 0.45 0.40
assign (residue 32 and name HD1) (residue 31 and name HA) 3.60 0.40 0.80
assign (residue 32 and name HD1) (residue 32 and name HA) 4.00 0.40 0.40
assign (residue 32 and name HD1) (residue 74 and name QD1) 3.00 0.40 0.40
! HB2 is too far, change to HB1, NOE error
assign (residue 32 and name HD1) (residue 80 and name HB2) 4.60 0.40 0.40
assign (residue 32 and name HD1) (residue 80 and name QD2) 4.20 0.40 0.60
assign (residue 32 and name HD1) (residue 95 and name QG1) 3.80 0.40 0.80
assign (residue 32 and name HE1) (residue 30 and name QG1) 2.90 0.40 0.40
assign (residue 32 and name HE1) (residue 30 and name QG2) 2.80 0.40 0.40
assign (residue 32 and name HE1) (residue 32 and name HD1) 2.60 0.40 0.40
! assign (residue 32 and name HE1) (residue 32 and name HE3) 3.40 0.40 0.40
assign (residue 32 and name HE1) (residue 74 and name QD1) 2.70 0.40 0.40
! HB2 may be too far, change to HB1, NOE error
assign (residue 32 and name HE1) (residue 80 and name HB2) 3.60 0.40 0.80
! Too far, NOE error
! assign (residue 32 and name HE1) (residue 80 and name QD1) 3.60 0.40 0.80
! Change to HG1#
assign (residue 32 and name HE1) (residue 95 and name QG1) 3.80 0.60 0.80
! assign (residue 32 and name HE3) (residue 16 and name HB3) 3.20 0.40 0.40
! assign (residue 32 and name HE3) (residue 18 and name H) 3.20 0.40 0.40
assign (residue 32 and name HE3) (residue 32 and name HH2) 4.40 0.40 0.40
assign (residue 32 and name HE3) (residue 32 and name HZ3) 2.60 0.40 0.40

```

```

! assign (residue 32 and name HE3) (residue 79 and name HA) 3.20 0.40 0.40
! Change to HB1
assign (residue 32 and name HE3) (residue 80 and name HB2) 3.60 0.40 0.80
! assign (residue 32 and name HE3) (residue 95 and name HB) 3.60 0.40 0.80
! Temp take out
! assign (residue 32 and name HE3) (residue 95 and name QG2) 4.00 0.40 0.40
! assign (residue 32 and name H) (residue 30 and name HB) 3.40 0.60 0.80
assign (residue 32 and name H) (residue 31 and name HA) 2.40 0.40 0.40
assign (residue 32 and name H) (residue 32 and name HA) 3.20 0.60 0.80
assign (residue 32 and name H) (residue 32 and name QB) 2.40 0.40 0.40
assign (residue 32 and name H) (residue 32 and name HD1) 3.60 0.80 0.80
assign (residue 32 and name H) (residue 40 and name QB) 3.50 0.60 0.80
! assign (residue 32 and name H) (residue 96 and name QG2) 3.20 0.60 0.80
! assign (residue 32 and name HZ2) (residue 32 and name HA) 3.40 0.40 0.40
assign (residue 32 and name HZ2) (residue 32 and name HH2) 2.60 0.40 0.40
assign (residue 32 and name HZ2) (residue 32 and name HZ3) 4.40 0.40 0.40
! assign (residue 32 and name HZ2) (residue 95 and name HB) 3.60 0.40 0.40
! Change to HG1#
assign (residue 32 and name HZ2) (residue 95 and name QG1) 3.40 0.60 0.40
assign (residue 32 and name HZ3) (residue 16 and name HB3) 3.20 0.40 0.40
assign (residue 32 and name HZ3) (residue 16 and name QD2) 4.20 0.40 0.40
assign (residue 32 and name HZ3) (residue 32 and name HH2) 2.40 0.40 0.40
! assign (residue 32 and name HZ3) (residue 80 and name QD2) 3.60 0.40 0.80
assign (residue 32 and name HZ3) (residue 93 and name HB2) 3.60 0.40 0.80
! Lower it, NOE error
assign (residue 32 and name HZ3) (residue 93 and name HB3) 3.00 0.40 0.40
! Too far, NOE error
! assign (residue 32 and name HZ3) (residue 95 and name HB) 3.60 0.40 0.80
! Lower to 3.40 from 3.80, NOE error
assign (residue 32 and name HZ3) (residue 95 and name QG2) 3.40 0.60 0.60

! E33
assign (residue 33 and name H) (residue 32 and name HA) 2.40 0.40 0.40
assign (residue 33 and name H) (residue 32 and name QB) 3.40 0.60 0.80
assign (residue 33 and name H) (residue 33 and name HA) 3.60 0.80 0.80
assign (residue 33 and name H) (residue 33 and name QB) 3.20 0.60 0.80
assign (residue 33 and name H) (residue 94 and name QB) 3.40 0.40 0.80
! NOE error
assign (residue 33 and name H) (residue 95 and name HA) 4.00 0.40 0.40

! R34
assign (residue 34 and name H) (residue 33 and name HA) 2.20 0.40 0.40
assign (residue 34 and name H) (residue 33 and name QB) 2.80 0.60 0.60
assign (residue 34 and name H) (residue 34 and name HA) 3.20 0.60 0.80
assign (residue 34 and name H) (residue 34 and name QB) 3.10 0.60 0.80
assign (residue 34 and name H) (residue 37 and name HB) 3.20 0.60 0.80
! Too far, DCA
! assign (residue 34 and name H) (residue 39 and name QG1) 3.20 0.60 0.80
assign (residue 34 and name H) (residue 39 and name QG2) 3.40 0.40 0.40

! G35
assign (residue 35 and name H) (residue 34 and name HA) 2.80 0.60 0.60
assign (residue 35 and name H) (residue 34 and name QB) 3.80 0.60 0.80
assign (residue 35 and name H) (residue 35 and name HA2) 2.20 0.40 0.40
assign (residue 35 and name H) (residue 35 and name HA3) 2.80 0.40 0.40
assign (residue 35 and name H) (residue 93 and name HA) 4.50 0.40 0.50

! S36
assign (residue 36 and name H) (residue 35 and name HA2) 2.80 0.40 0.80
assign (residue 36 and name H) (residue 35 and name HA3) 2.20 0.40 0.40
assign (residue 36 and name H) (residue 36 and name HA) 3.40 0.60 0.60

! T37
! It should be T37QG2 to L38HN, see L38 below
! assign (residue 37 and name QG2) (residue 36 and name H) 3.20 0.40 0.40
assign (residue 37 and name QG2) (residue 37 and name HA) 2.60 0.40 0.40
assign (residue 37 and name H) (residue 34 and name H) 3.20 0.40 0.40
assign (residue 37 and name H) (residue 36 and name HA) 3.20 0.60 0.80
assign (residue 37 and name H) (residue 36 and name H) 3.00 0.40 0.40
assign (residue 37 and name H) (residue 37 and name HB) 2.30 0.60 0.80
assign (residue 37 and name H) (residue 37 and name QG2) 3.60 0.60 1.00

! L38
assign (residue 38 and name H) (residue 37 and name QG2) 3.20 0.60 0.40
assign (residue 38 and name H) (residue 38 and name HA) 2.60 0.60 0.80

! V39
assign (residue 39 and name HB) (residue 39 and name HA) 2.80 0.40 0.40
assign (residue 39 and name QG1) (residue 32 and name QB) 3.40 0.60 0.60
! Wrong assignment DCA
! assign (residue 39 and name QG1) (residue 33 and name HA) 2.60 0.40 0.40
! NOE error suggested 2.50
assign (residue 39 and name QG1) (residue 39 and name HA) 2.40 0.40 0.40
assign (residue 39 and name QG1) (residue 39 and name HB) 2.40 0.40 0.40
assign (residue 39 and name QG1) (residue 93 and name QE) 2.60 0.40 0.40
assign (residue 39 and name QG2) (residue 33 and name HA) 3.60 0.40 0.60
assign (residue 39 and name QG2) (residue 39 and name HA) 2.60 0.60 0.40
assign (residue 39 and name QG2) (residue 39 and name HB) 2.40 0.40 0.40

```

```

assign (residue 39 and name QG2) (residue 39 and name QG1) 2.40 0.40 0.40
! NOE error 4.78
assign (residue 39 and name QG2) (residue 93 and name HD#) 4.50 0.40 0.40
assign (residue 39 and name QG2) (residue 93 and name QE) 3.00 0.40 0.40
assign (residue 39 and name H) (residue 33 and name HA) 2.80 0.60 0.80
assign (residue 39 and name H) (residue 38 and name HA) 2.80 0.60 0.80
assign (residue 39 and name H) (residue 39 and name HA) 3.60 0.70 1.00
! NOE error, 2.38
assign (residue 39 and name H) (residue 39 and name HB) 2.40 0.60 0.60
! NOE error 3.33
assign (residue 39 and name H) (residue 39 and name QG1) 3.50 0.60 0.60
assign (residue 39 and name H) (residue 39 and name QG2) 3.20 0.40 0.40

! A40
assign (residue 40 and name QB) (residue 32 and name HD1) 3.40 0.40 0.40
assign (residue 40 and name QB) (residue 40 and name HA) 2.40 0.40 0.60
assign (residue 40 and name QB) (residue 72 and name HB3) 3.40 0.40 0.40
! NOE error
assign (residue 40 and name H) (residue 32 and name QB) 2.60 0.40 0.40
! 2D S/N 6
assign (residue 40 and name H) (residue 32 and name HA) 4.60 0.40 0.60
assign (residue 40 and name H) (residue 33 and name HA) 4.40 0.40 0.60
assign (residue 40 and name H) (residue 39 and name HA) 3.40 0.60 0.80
! NOE error
assign (residue 40 and name H) (residue 39 and name HB) 3.00 0.40 0.40
assign (residue 40 and name H) (residue 39 and name QG1) 3.00 0.40 0.40
! May be too far
assign (residue 40 and name H) (residue 39 and name QG2) 4.00 0.40 1.00
assign (residue 40 and name H) (residue 39 and name H) 2.40 0.40 0.40
assign (residue 40 and name H) (residue 40 and name HA) 2.80 0.60 0.80
assign (residue 40 and name H) (residue 40 and name QB) 2.60 0.40 0.60

! E41
assign (residue 41 and name QB) (residue 41 and name HA) 2.60 0.40 0.40
assign (residue 41 and name QB) (residue 41 and name H) 2.40 0.80 0.80
assign (residue 41 and name H) (residue 40 and name HA) 2.00 0.20 0.40
assign (residue 41 and name H) (residue 40 and name QB) 2.40 0.40 0.60
assign (residue 41 and name H) (residue 41 and name HA) 3.20 0.60 0.80
assign (residue 41 and name H) (residue 41 and name QB) 2.80 0.60 0.60
! 2D S/N 8
assign (residue 41 and name H) (residue 40 and name H) 4.20 0.40 0.80

! F42
assign (residue 42 and name HB2) (residue 42 and name HA) 2.40 0.40 0.40
assign (residue 42 and name QD) (residue 29 and name HA) 3.40 0.60 0.40
assign (residue 42 and name QD) (residue 30 and name QG1) 2.40 0.60 0.80
assign (residue 42 and name QD) (residue 42 and name HA) 3.20 0.40 0.40
assign (residue 42 and name QD) (residue 42 and name HB2) 2.50 0.40 0.40
assign (residue 42 and name QD) (residue 42 and name HB3) 2.60 0.40 0.40
assign (residue 42 and name H) (residue 30 and name QG1) 2.60 0.40 0.60
! NOE Error
assign (residue 42 and name H) (residue 30 and name H) 3.00 0.40 0.40
assign (residue 42 and name H) (residue 41 and name HA) 2.40 0.40 0.40
! Adjusted
assign (residue 42 and name H) (residue 41 and name QB) 3.80 0.40 0.60
assign (residue 42 and name H) (residue 42 and name HA) 3.20 0.60 0.80
assign (residue 42 and name H) (residue 42 and name HB2) 3.20 0.60 0.80
! Adjusted
assign (residue 42 and name H) (residue 42 and name HB3) 2.80 0.60 0.80
assign (residue 42 and name H) (residue 42 and name QD) 3.60 0.60 0.80

! K43
assign (residue 43 and name H) (residue 42 and name HA) 2.20 0.40 0.40
assign (residue 43 and name H) (residue 42 and name HB2) 3.40 0.60 0.80
assign (residue 43 and name H) (residue 42 and name HB3) 3.40 0.60 0.80
! New, need to update sparky, similar data height to K43H-F42HB2
assign (residue 43 and name H) (residue 42 and name QD) 3.40 0.60 0.80
NOE error 2.80
assign (residue 43 and name H) (residue 43 and name HA) 2.80 0.40 0.40

! R44
assign (residue 44 and name H) (residue 28 and name HA) 3.40 0.60 0.80
! New, need to update sparky, and possible for E29QB DCA
assign (residue 44 and name H) (residue 28 and name HB3) 4.00 0.40 1.00
assign (residue 44 and name H) (residue 29 and name HA) 3.80 0.60 0.80
assign (residue 44 and name H) (residue 43 and name HA) 2.80 0.60 0.60
assign (residue 44 and name H) (residue 44 and name HA) 3.40 0.60 0.80

! K45
assign (residue 45 and name H) (residue 44 and name HA) 3.20 0.60 0.60
assign (residue 45 and name H) (residue 45 and name HA) 2.80 0.60 0.80
! New, sparky
assign (residue 45 and name H) (residue 44 and name H) 3.40 0.40 0.60

! M46
assign (residue 46 and name H) (residue 45 and name HA) 2.80 0.60 0.80
assign (residue 46 and name H) (residue 46 and name HA) 3.20 0.60 0.80

```

```

! K47
assign (residue 47 and name H) (residue 46 and name HA) 2.40 0.40 0.40

! F49
assign (residue 49 and name HB2) (residue 49 and name HA) 3.00 0.60 0.60
assign (residue 49 and name HB3) (residue 49 and name HA) 3.00 0.40 0.60
assign (residue 49 and name QD) (residue 49 and name HB2) 3.00 0.60 0.40
assign (residue 49 and name QD) (residue 49 and name HB3) 2.60 0.40 0.40
assign (residue 49 and name H) (residue 41 and name QB) 3.40 0.60 0.80
assign (residue 49 and name H) (residue 48 and name HA) 2.30 0.40 0.40
assign (residue 49 and name H) (residue 49 and name HA) 2.80 0.60 0.80
assign (residue 49 and name H) (residue 49 and name HB2) 2.80 0.60 0.60
assign (residue 49 and name H) (residue 49 and name HB3) 2.40 0.80 0.80
assign (residue 49 and name H) (residue 49 and name QD) 3.60 0.60 0.80

! L50
assign (residue 50 and name H) (residue 49 and name HA) 2.40 0.40 0.40
assign (residue 50 and name H) (residue 49 and name HB2) 3.60 0.60 0.80
! 2D S/N 14
assign (residue 50 and name H) (residue 49 and name H) 4.20 0.40 0.80
assign (residue 50 and name H) (residue 50 and name HA) 3.20 0.60 0.80
assign (residue 50 and name H) (residue 50 and name QB) 2.30 0.30 0.40

! K51
! 2D S/N 4
assign (residue 51 and name H) (residue 39 and name QG2) 4.00 0.40 1.00
assign (residue 51 and name H) (residue 50 and name HA) 2.80 0.60 0.60
assign (residue 51 and name H) (residue 50 and name QB) 3.60 0.60 0.80
! New, sparky
assign (residue 51 and name H) (residue 50 and name H) 4.00 0.40 0.60
! NOE error suggest lower
assign (residue 51 and name H) (residue 51 and name HA) 3.00 0.60 0.60

! D56
! 2D S/N 18
assign (residue 56 and name QB) (residue 56 and name HA) 3.00 0.65 1.00

! K57
! 2D S/N 12
assign (residue 57 and name QG) (residue 67 and name QG) 5.00 1.00 1.20
! 3D S/N 60
assign (residue 57 and name H) (residue 56 and name HA) 2.80 0.60 0.60
! 3D S/N 33
assign (residue 57 and name H) (residue 56 and name QB) 4.00 0.80 1.00
assign (residue 57 and name H) (residue 57 and name HA) 3.20 0.60 0.80
! 3D S/N 36
assign (residue 57 and name H) (residue 57 and name QB) 3.00 0.65 0.80
! 3D S/N 49
assign (residue 57 and name H) (residue 57 and name QG) 2.80 0.60 0.60

! D58
! 2D S/N 21
assign (residue 58 and name H) (residue 56 and name QB) 4.20 0.60 1.00
! 3D S/N 81
assign (residue 58 and name H) (residue 57 and name HA) 3.00 0.60 0.60
! 3D S/N 28
assign (residue 58 and name H) (residue 57 and name QB) 3.20 0.60 0.80
! 3D S/N 14
assign (residue 58 and name H) (residue 57 and name QG) 4.00 0.60 0.80
assign (residue 58 and name H) (residue 58 and name HA) 3.10 0.50 0.80
assign (residue 58 and name H) (residue 58 and name QB) 2.80 0.60 0.60

! G59
assign (residue 59 and name H) (residue 58 and name HA) 3.20 0.60 0.80
! 3D S/N 22
assign (residue 59 and name H) (residue 58 and name QB) 4.00 0.60 0.80
assign (residue 59 and name H) (residue 59 and name HA3) 2.40 0.40 0.40
! 3D S/N 15
assign (residue 59 and name H) (residue 61 and name H) 4.00 0.60 0.80

! N60
! 3D S/N 48, use NOE constraint above 50
assign (residue 60 and name HD21) (residue 60 and name QB) 2.80 0.60 0.60
assign (residue 60 and name HD22) (residue 60 and name QB) 3.60 0.60 0.80
assign (residue 60 and name H) (residue 59 and name HA3) 3.20 0.60 0.80
assign (residue 60 and name H) (residue 60 and name HA) 3.20 0.60 0.80
assign (residue 60 and name H) (residue 60 and name QB) 3.20 0.60 0.80

! G61
assign (residue 61 and name H) (residue 60 and name HA) 3.20 0.60 0.80
assign (residue 61 and name H) (residue 60 and name QB) 3.60 0.60 0.80
assign (residue 61 and name H) (residue 61 and name QA) 2.80 0.60 0.80
! 3D S/N 12
assign (residue 61 and name H) (residue 62 and name H) 3.20 1.00 1.00

! Y62
! 2D S/N 13
assign (residue 62 and name QD) (residue 62 and name HA) 3.40 0.40 0.40

```

```

! 2D S/N 44
assign (residue 62 and name QD) (residue 62 and name QB) 2.80 0.80 0.60
! 3D S/N 29
assign (residue 62 and name H) (residue 60 and name HA) 4.00 0.60 1.00
! 2D S/N 34, 3D S/N is 162, change to 3.20 and open lower limit
assign (residue 62 and name H) (residue 61 and name QA) 3.20 1.00 0.80
assign (residue 62 and name H) (residue 62 and name HA) 2.80 0.60 0.60
! 3D S/N 150
assign (residue 62 and name H) (residue 62 and name QB) 3.00 0.80 0.60
assign (residue 62 and name H) (residue 62 and name HD#) 3.20 0.60 0.80

! I63
! 3D S/N 242
assign (residue 63 and name H) (residue 62 and name HA) 2.00 0.20 0.40
! 3D S/N 55
assign (residue 63 and name H) (residue 62 and name QB) 2.80 0.60 0.80
! 2D S/N 4
assign (residue 63 and name H) (residue 62 and name QD) 4.00 0.60 1.00
! 2D S/N 26
assign (residue 63 and name H) (residue 62 and name H) 4.00 0.60 1.00
assign (residue 63 and name H) (residue 63 and name HA) 2.80 0.60 0.60
! 3D S/N 79
assign (residue 63 and name H) (residue 63 and name HB) 2.80 0.60 0.60
! 3D S/N 34
assign (residue 63 and name H) (residue 63 and name QG2) 3.20 0.60 0.80

! S64
! S/N 3D 181
assign (residue 64 and name H) (residue 63 and name HA) 2.20 0.40 0.40
! S/N 3D 11
assign (residue 64 and name H) (residue 63 and name HB) 4.00 2.00 1.20
! S/N 3D 42
assign (residue 64 and name H) (residue 63 and name QG2) 3.20 0.60 1.20
assign (residue 64 and name H) (residue 64 and name HA) 3.20 0.60 0.80
! S/N 3D 51
assign (residue 64 and name H) (residue 64 and name QB) 2.80 0.60 0.60

! A65
! 2D S/N 44 Temp take out
! assign (residue 65 and name QB) (residue 64 and name HA)
! 3D S/N 68
assign (residue 65 and name H) (residue 64 and name HA) 2.80 0.80 0.60
! 3D S/N 16
assign (residue 65 and name H) (residue 64 and name QB) 3.20 0.60 0.80
assign (residue 65 and name H) (residue 65 and name HA) 3.20 0.60 0.80
! 3D S/N 72
assign (residue 65 and name H) (residue 65 and name QB) 2.80 0.60 0.60
! 3D S/N 79
assign (residue 65 and name H) (residue 66 and name QB) 2.80 0.60 0.60

! E67
! 2D S/N 13 Flip this with A66 from A65
assign (residue 67 and name QB) (residue 66 and name H) 4.20 0.60 1.20
! 3D S/N 79
assign (residue 67 and name H) (residue 66 and name QB) 2.80 0.60 0.60
assign (residue 67 and name H) (residue 67 and name HA) 2.40 0.60 0.40
! 3D S/N 99
assign (residue 67 and name H) (residue 67 and name QB) 2.80 0.80 0.80
! 3D S/N 57
assign (residue 67 and name H) (residue 67 and name QG) 2.80 0.60 0.60

! G68
assign (residue 68 and name H) (residue 67 and name HA) 2.80 0.60 0.80
! 3D S/N 12
assign (residue 68 and name H) (residue 67 and name QB) 4.00 1.20 1.00
assign (residue 68 and name H) (residue 68 and name HA2) 3.20 0.60 0.80
assign (residue 68 and name H) (residue 68 and name HA3) 2.40 0.40 0.40

! A71
assign (residue 71 and name QB) (residue 71 and name HA) 2.60 0.40 0.40
assign (residue 71 and name H) (residue 70 and name QA) 3.00 0.40 0.40
assign (residue 71 and name H) (residue 71 and name HA) 2.40 0.40 0.60
assign (residue 71 and name H) (residue 71 and name QB) 2.20 0.40 0.40
assign (residue 71 and name H) (residue 72 and name H) 3.40 0.60 0.40

! F72
assign (residue 72 and name HB2) (residue 72 and name HA) 3.00 0.40 0.40
assign (residue 72 and name HB3) (residue 72 and name HA) 2.60 0.40 0.40
assign (residue 72 and name QD) (residue 39 and name QG1) 3.00 0.40 0.60
assign (residue 72 and name QD) (residue 71 and name QB) 2.70 0.40 0.60
assign (residue 72 and name QD) (residue 72 and name HA) 3.20 0.40 0.40
assign (residue 72 and name QD) (residue 72 and name HB2) 2.60 0.40 0.40
assign (residue 72 and name QD) (residue 72 and name HB3) 2.30 0.40 0.40
assign (residue 72 and name QD) (residue 80 and name QD2) 2.75 0.80 0.60
assign (residue 72 and name QD) (residue 82 and name HA) 3.00 0.40 0.40
assign (residue 72 and name QE) (residue 39 and name QG1) 3.60 0.40 0.40
assign (residue 72 and name QE) (residue 39 and name QG2) 3.80 0.40 0.40
assign (residue 72 and name QE) (residue 71 and name QB) 3.20 0.40 0.40

```

```

assign (residue 72 and name QE) (residue 72 and name HB3) 4.40 0.40 0.40
assign (residue 72 and name QE) (residue 72 and name QD) 2.00 0.40 0.40
! It should be at limit of the detection
! assign (residue 72 and name QE) (residue 80 and name QD2) 3.20 0.40 0.40
assign (residue 72 and name QE) (residue 82 and name HB) 4.80 0.40 0.40
! Wrong assignment DCA
! assign (residue 72 and name H) (residue 40 and name QB) 4.60 0.80 0.50
assign (residue 72 and name H) (residue 71 and name HA) 2.80 0.60 0.80
! Need to DCA
! assign (residue 72 and name H) (residue 71 and name QB) 2.40 0.40 0.40
assign (residue 72 and name H) (residue 72 and name HA) 2.80 0.60 0.60
assign (residue 72 and name H) (residue 72 and name HB2) 2.80 0.60 0.60
assign (residue 72 and name H) (residue 72 and name HB3) 3.00 0.60 0.80
assign (residue 72 and name H) (residue 72 and name QD) 3.20 0.60 0.80
! NOE is very weak, set to the max distance
assign (residue 72 and name H) (residue 82 and name HA) 4.20 0.40 0.90

! E73
assign (residue 73 and name HB2) (residue 73 and name HA) 3.00 0.40 0.60
assign (residue 73 and name HB3) (residue 73 and name HA) 3.40 0.40 0.40
! S/N 9
assign (residue 73 and name HB3) (residue 81 and name H) 4.00 0.60 0.80
assign (residue 73 and name H) (residue 72 and name HA) 2.20 0.40 0.40
assign (residue 73 and name H) (residue 72 and name HB2) 3.40 0.40 0.80
assign (residue 73 and name H) (residue 72 and name HB3) 3.20 0.60 0.80
assign (residue 73 and name H) (residue 73 and name HA) 3.10 0.60 0.80
assign (residue 73 and name H) (residue 73 and name HB2) 3.40 0.50 0.40
assign (residue 73 and name H) (residue 73 and name HB3) 3.10 0.60 0.80
! Flip to QD2 from QD1
assign (residue 73 and name H) (residue 80 and name QD2) 3.20 0.60 0.80
! Missed assigned, change from I82HA to L80HA
assign (residue 73 and name H) (residue 80 and name HA) 4.00 0.40 0.60
! New, sparky
assign (residue 73 and name H) (residue 81 and name H) 3.20 0.40 0.40

! I74
! May be QD1 of I74
! assign (residue 74 and name QG2) (residue 74 and name HA) 2.60 0.40 0.40
assign (residue 74 and name QG2) (residue 74 and name HB) 2.40 0.40 0.40
assign (residue 74 and name QG2) (residue 75 and name H) 3.40 0.40 0.60
! NOE error suggest 5.1
assign (residue 74 and name H) (residue 40 and name QB) 4.90 0.40 0.40
assign (residue 74 and name H) (residue 73 and name HA) 2.20 0.40 0.40
assign (residue 74 and name H) (residue 74 and name HA) 3.20 0.60 0.80
assign (residue 74 and name H) (residue 74 and name HB) 2.30 0.40 0.80
assign (residue 74 and name H) (residue 74 and name QG2) 2.80 0.60 0.60

! L75
assign (residue 75 and name HB2) (residue 75 and name HA) 3.20 0.40 0.40
assign (residue 75 and name HB3) (residue 75 and name HA) 2.40 0.80 0.80
assign (residue 75 and name HB3) (residue 75 and name H) 3.00 0.40 0.40
assign (residue 75 and name QD1) (residue 75 and name HA) 2.40 0.40 0.40
assign (residue 75 and name QD1) (residue 75 and name H) 3.40 0.40 0.40
assign (residue 75 and name HG) (residue 75 and name HA) 3.20 0.40 0.40
assign (residue 75 and name HG) (residue 75 and name HB2) 2.60 0.40 0.60
assign (residue 75 and name H) (residue 74 and name HA) 2.40 0.40 0.40
assign (residue 75 and name H) (residue 74 and name QG2) 3.20 0.60 0.80
assign (residue 75 and name H) (residue 75 and name HA) 3.00 0.60 0.80
assign (residue 75 and name H) (residue 75 and name HB2) 2.20 0.40 0.40
assign (residue 75 and name H) (residue 75 and name HG) 4.20 0.40 0.40
assign (residue 75 and name H) (residue 80 and name HA) 3.60 0.40 0.80

! A76
assign (residue 76 and name H) (residue 75 and name HA) 2.80 0.60 0.60
assign (residue 76 and name H) (residue 76 and name HA) 3.00 0.60 0.80

! N77
assign (residue 77 and name HD21) (residue 77 and name HB2) 3.00 0.60 1.00
assign (residue 77 and name HD21) (residue 77 and name HB3) 3.40 0.60 0.80
assign (residue 77 and name HD22) (residue 77 and name HB2) 3.60 0.60 0.80
assign (residue 77 and name HD22) (residue 77 and name HB3) 3.60 0.60 0.80
assign (residue 77 and name H) (residue 75 and name HB2) 3.60 0.40 0.60
assign (residue 77 and name H) (residue 75 and name HB3) 2.80 0.40 0.60
assign (residue 77 and name H) (residue 76 and name HA) 3.30 0.60 0.80
assign (residue 77 and name H) (residue 77 and name HA) 3.20 0.60 0.80
assign (residue 77 and name H) (residue 77 and name HB2) 3.20 0.60 0.80
assign (residue 77 and name H) (residue 77 and name HB3) 2.80 0.60 0.60

! G78
assign (residue 78 and name H) (residue 74 and name QG2) 4.00 0.60 1.00
assign (residue 78 and name H) (residue 75 and name HB2) 3.60 0.60 0.80
assign (residue 78 and name H) (residue 75 and name HB3) 3.80 0.60 0.80
assign (residue 78 and name H) (residue 77 and name HA) 3.60 0.60 0.80
assign (residue 78 and name H) (residue 77 and name H) 2.80 0.40 0.40
assign (residue 78 and name H) (residue 78 and name HA2) 3.00 0.60 0.80
assign (residue 78 and name H) (residue 78 and name HA3) 2.80 0.60 0.80
assign (residue 78 and name H) (residue 74 and name QD1) 3.60 0.40 0.60
! Repeat, see above

```

```

! assign (residue 78 and name H) (residue 75 and name HB2) 3.20 0.60 0.80

! D79
! 3D, S/N 21, sparky, assignment is QG2 but I think is QD1
assign (residue 79 and name H) (residue 74 and name QD1) 4.00 0.40 1.00
assign (residue 79 and name H) (residue 78 and name HA2) 3.40 0.60 0.80
assign (residue 79 and name H) (residue 78 and name HA3) 3.40 0.60 0.80
assign (residue 79 and name H) (residue 78 and name H) 2.80 0.40 0.40
assign (residue 79 and name H) (residue 79 and name HA) 3.20 0.60 0.80
assign (residue 79 and name H) (residue 79 and name HB2) 2.40 0.40 0.40
! Adjusted
assign (residue 79 and name H) (residue 79 and name HB3) 3.00 0.60 0.60

! L80
assign (residue 80 and name HB3) (residue 16 and name HB2) 2.60 0.40 0.60
! assign (residue 80 and name HB3) (residue 74 and name QG2) 3.60 0.40 0.60
assign (residue 80 and name HB3) (residue 80 and name HA) 2.60 0.40 0.60
assign (residue 80 and name HB3) (residue 80 and name QD1) 2.40 0.60 0.40
! Up the distance
assign (residue 80 and name HB3) (residue 80 and name QD2) 3.00 0.40 0.60
assign (residue 80 and name HB3) (residue 82 and name QG1) 3.60 0.40 0.60
! Too far, should be L80HD2#-W32HD1 with 4.55
! assign (residue 80 and name QD1) (residue 32 and name HD1) 3.00 0.40 0.40
! Too far, miss assignment
! assign (residue 80 and name QD1) (residue 32 and name HZ2) 3.20 0.40 0.40
! Too far, may be L80HD2#-A40HN with 4.72
! assign (residue 80 and name QD1) (residue 40 and name H) 3.60 0.40 0.60
! Too far, see HD2# below
! assign (residue 80 and name QD1) (residue 72 and name HA) 3.60 0.40 0.60
assign (residue 80 and name QD1) (residue 72 and name HB3) 4.00 0.40 0.60
assign (residue 80 and name QD1) (residue 72 and name HD#) 3.20 0.40 0.40
! Too far, see HD2# below
! assign (residue 80 and name QD1) (residue 74 and name HA) 3.40 0.40 0.40
! Flip value with HD2#
assign (residue 80 and name QD1) (residue 80 and name HA) 4.00 0.40 0.40
assign (residue 80 and name QD1) (residue 80 and name HG) 2.20 0.40 0.60
assign (residue 80 and name QD1) (residue 80 and name H) 4.60 0.40 0.40
! Raise distance
assign (residue 80 and name QD2) (residue 39 and name HB) 4.60 0.40 0.40
assign (residue 80 and name QD2) (residue 39 and name QG1) 2.30 0.40 0.40
assign (residue 80 and name QD2) (residue 72 and name HA) 3.60 0.40 0.60
assign (residue 80 and name QD2) (residue 74 and name HA) 3.40 0.40 0.40
! Flip value with HD1# above
assign (residue 80 and name QD2) (residue 80 and name HA) 2.40 0.40 0.60
assign (residue 80 and name QD2) (residue 80 and name QD1) 2.20 0.40 0.40
assign (residue 80 and name QD2) (residue 80 and name HG) 2.20 0.40 0.40
! 4.76, temp take out
! assign (residue 80 and name QD2) (residue 80 and name H) 3.60 0.40 0.60
! Change to HD1#
assign (residue 80 and name QD1) (residue 82 and name QD1) 3.00 0.40 0.40
! Change to HD1#
assign (residue 80 and name QD1) (residue 93 and name QE) 3.20 0.60 0.40
assign (residue 80 and name H) (residue 16 and name HB2) 3.00 0.40 0.60
assign (residue 80 and name H) (residue 16 and name HB3) 3.60 0.60 0.80
assign (residue 80 and name H) (residue 16 and name H) 3.20 0.40 0.60
assign (residue 80 and name H) (residue 79 and name HA) 2.60 0.60 0.60
assign (residue 80 and name H) (residue 79 and name HB2) 3.60 0.60 0.80
assign (residue 80 and name H) (residue 79 and name HB3) 3.00 0.60 0.80
assign (residue 80 and name H) (residue 80 and name HA) 3.00 0.60 0.80
! Temp take out
! assign (residue 80 and name H) (residue 80 and name HB2) 3.20 0.60 0.80
assign (residue 80 and name H) (residue 80 and name HB3) 2.80 0.40 0.40
! assign (residue 80 and name H) (residue 73 and name H) 3.60 0.40 0.80

! K81
! New, sparky
assign (residue 81 and name H) (residue 73 and name HB3) 3.40 0.40 0.40
assign (residue 81 and name H) (residue 80 and name HA) 2.80 0.60 0.60
! Flip HB3 to HB2
assign (residue 81 and name H) (residue 80 and name HB2) 4.60 0.40 0.40
! New, sparky
assign (residue 81 and name H) (residue 80 and name HG) 4.40 0.40 0.40
! Flip to QD2
assign (residue 81 and name H) (residue 80 and name QD2) 3.20 0.60 0.80
assign (residue 81 and name H) (residue 81 and name HA) 3.20 0.60 0.80

! I82
assign (residue 82 and name QD1) (residue 72 and name QE) 3.20 0.40 0.40
! NOE error suggest 3.9
assign (residue 82 and name QD1) (residue 82 and name HA) 4.00 0.40 0.60
assign (residue 82 and name QD1) (residue 82 and name HB) 2.40 0.40 0.40
assign (residue 82 and name QD1) (residue 93 and name HH) 3.60 0.40 0.60
assign (residue 82 and name QG1) (residue 82 and name HB) 2.40 0.40 0.40
assign (residue 82 and name QG2) (residue 72 and name QD) 3.00 0.40 0.40
assign (residue 82 and name QG2) (residue 72 and name QE) 2.80 0.40 0.40
assign (residue 82 and name QG2) (residue 82 and name HA) 2.60 0.40 0.40
assign (residue 82 and name QG2) (residue 82 and name HB) 2.20 0.40 0.40
assign (residue 82 and name QG2) (residue 83 and name H) 3.20 0.40 0.40

```

```

assign (residue 82 and name QG2) (residue 84 and name H) 2.40 0.80 0.80
assign (residue 82 and name H) (residue 14 and name QG2) 3.20 0.60 0.80
assign (residue 82 and name H) (residue 15 and name HA) 3.40 0.60 0.80
assign (residue 82 and name H) (residue 81 and name HA) 2.40 0.40 0.40
assign (residue 82 and name H) (residue 82 and name HA) 3.40 0.60 0.80
assign (residue 82 and name H) (residue 82 and name HB) 2.80 0.60 0.80
assign (residue 82 and name H) (residue 82 and name QG2) 3.60 0.60 0.80

! K83
assign (residue 83 and name H) (residue 71 and name HA) 3.60 0.40 0.80
assign (residue 83 and name H) (residue 72 and name HA) 3.60 0.60 0.80
assign (residue 83 and name H) (residue 82 and name HA) 2.50 0.40 0.40
assign (residue 83 and name H) (residue 83 and name HA) 2.80 0.60 0.60
assign (residue 83 and name H) (residue 83 and name QB) 2.40 0.40 0.40
assign (residue 83 and name H) (residue 83 and name QG) 2.80 0.60 0.60

! N84
assign (residue 84 and name HB2) (residue 84 and name HA) 2.80 0.40 0.40
assign (residue 84 and name HB3) (residue 84 and name HA) 3.20 0.40 0.40
assign (residue 84 and name HB3) (residue 84 and name HB2) 2.00 0.40 0.40
assign (residue 84 and name HB3) (residue 84 and name H) 2.40 0.80 0.80
assign (residue 84 and name HD21) (residue 84 and name HB2) 3.40 0.60 0.80
assign (residue 84 and name HD21) (residue 84 and name HB3) 3.60 0.60 0.80
assign (residue 84 and name HD21) (residue 84 and name HD22) 2.40 0.80 0.80
assign (residue 84 and name HD22) (residue 84 and name HB2) 3.40 0.60 0.80
! Temp take out
! assign (residue 84 and name HD22) (residue 84 and name HB3) 3.60 0.60 0.80
! Repeat
! assign (residue 84 and name HD22) (residue 84 and name HD21) 2.40 0.80 0.80
! Adjusted
assign (residue 84 and name H) (residue 82 and name HA) 4.00 0.40 0.80
assign (residue 84 and name H) (residue 82 and name QG2) 3.20 0.40 0.40
assign (residue 84 and name H) (residue 83 and name HA) 3.20 0.60 0.80
assign (residue 84 and name H) (residue 83 and name QB) 2.40 0.40 0.40
! Adjust from 3.40 to 4.00
assign (residue 84 and name H) (residue 83 and name QG) 4.00 0.40 0.80
assign (residue 84 and name H) (residue 83 and name H) 3.00 0.40 0.40
assign (residue 84 and name H) (residue 84 and name HA) 2.80 0.60 0.80
assign (residue 84 and name H) (residue 84 and name HB2) 3.20 0.60 0.80
! Repeat
! assign (residue 84 and name H) (residue 84 and name HB3) 3.20 0.60 0.80
assign (residue 84 and name H) (residue 85 and name H) 3.60 0.40 0.80

! L85
assign (residue 85 and name QD1) (residue 10 and name HA) 4.20 0.40 0.40
! Too far, DCA
! assign (residue 85 and name QD2) (residue 75 and name HB2) 3.20 0.40 0.40
! The NOE height is very weak, should be stronger, DCA
! assign (residue 85 and name QD2) (residue 85 and name HA) 3.20 0.40 0.40
assign (residue 85 and name QD2) (residue 85 and name HG) 2.40 0.40 0.40
assign (residue 85 and name H) (residue 84 and name HA) 2.20 0.40 0.40
assign (residue 85 and name H) (residue 84 and name HB2) 3.80 0.60 0.80
assign (residue 85 and name H) (residue 84 and name HB3) 3.60 0.60 0.80
assign (residue 85 and name H) (residue 85 and name HA) 3.20 0.60 0.80

! T86
assign (residue 86 and name H) (residue 82 and name QG2) 4.20 0.60 1.00
assign (residue 86 and name H) (residue 85 and name HA) 2.20 0.40 0.40
! NOE error 2.71
assign (residue 86 and name H) (residue 86 and name HA) 2.80 0.40 0.40
! Adjusted
assign (residue 86 and name H) (residue 86 and name QG2) 3.20 0.60 1.00
assign (residue 86 and name H) (residue 89 and name HB2) 2.40 0.40 0.40
assign (residue 86 and name H) (residue 89 and name HB3) 3.60 0.60 0.80

! R87
assign (residue 87 and name QB) (residue 87 and name HA) 2.60 0.40 0.40
assign (residue 87 and name QG) (residue 87 and name HA) 2.60 0.40 0.40
assign (residue 87 and name H) (residue 10 and name QD1) 2.60 0.40 0.60
assign (residue 87 and name H) (residue 86 and name HA) 3.00 0.60 0.80
! Adjusted, DCA
assign (residue 87 and name H) (residue 86 and name HB) 4.00 0.40 0.40
assign (residue 87 and name H) (residue 86 and name QG2) 2.80 0.60 0.60
assign (residue 87 and name H) (residue 87 and name HA) 2.80 0.60 0.60
assign (residue 87 and name H) (residue 87 and name QB) 2.20 0.40 0.60

! D88
assign (residue 88 and name HB3) (residue 88 and name HA) 3.20 0.40 0.40
assign (residue 88 and name H) (residue 86 and name HA) 3.80 0.40 0.80
! DCA
! assign (residue 88 and name H) (residue 86 and name QG2) 3.40 0.60 0.80
assign (residue 88 and name H) (residue 87 and name HA) 3.20 0.60 0.80
assign (residue 88 and name H) (residue 87 and name QB) 2.80 0.60 0.60
assign (residue 88 and name H) (residue 87 and name H) 2.60 0.40 0.60
assign (residue 88 and name H) (residue 88 and name HA) 2.80 0.60 0.60
assign (residue 88 and name H) (residue 88 and name HB2) 2.80 0.60 0.60
assign (residue 88 and name H) (residue 88 and name HB3) 3.20 0.40 0.40
assign (residue 88 and name H) (residue 89 and name H) 2.40 0.40 0.80

```



```

! D89
assign (residue 89 and name H) (residue 86 and name QG2) 3.60 0.60 0.80
! NOE error 4.20
assign (residue 89 and name H) (residue 87 and name HA) 4.00 0.40 0.40
! DCA, may be too far
! assign (residue 89 and name H) (residue 87 and name QB) 3.60 0.60 0.80
assign (residue 89 and name H) (residue 88 and name HA) 3.20 0.60 0.80
assign (residue 89 and name H) (residue 88 and name HB2) 3.80 0.40 0.40
! DCA, check NOE height 4.20
assign (residue 89 and name H) (residue 88 and name HB3) 3.40 0.60 0.80
assign (residue 89 and name H) (residue 88 and name H) 2.40 0.40 0.80
assign (residue 89 and name H) (residue 89 and name HA) 3.00 0.60 0.80
assign (residue 89 and name H) (residue 89 and name HB2) 2.60 0.60 0.60
! DCA
! assign (residue 89 and name H) (residue 89 and name HB3) 2.60 0.60 0.60
! Miss assignment, flip QD1 to QG2
assign (residue 89 and name H) (residue 114 and name QG2) 3.40 0.40 0.80

! S90
assign (residue 90 and name H) (residue 88 and name HA) 3.20 0.60 0.80
assign (residue 90 and name H) (residue 89 and name HA) 3.20 0.60 0.80
! Increase from 3.20 to 3.60
assign (residue 90 and name H) (residue 89 and name HB2) 3.60 0.60 0.80
! NOE error suggest 4.26
assign (residue 90 and name H) (residue 89 and name HB3) 4.30 0.40 0.40
! NOE error suggest 2.60
assign (residue 90 and name H) (residue 89 and name H) 2.60 0.40 0.40
assign (residue 90 and name H) (residue 90 and name HA) 2.80 0.60 0.60
assign (residue 90 and name H) (residue 90 and name HB2) 2.80 0.60 0.60
! NOE error suggest 3.20
assign (residue 90 and name H) (residue 90 and name HB3) 3.30 0.40 0.40
! NOE error suggest 4.00, but should be HG2# instead of HD1#
assign (residue 90 and name H) (residue 114 and name QG2) 4.00 0.40 0.40
assign (residue 90 and name H) (residue 114 and name HG12) 3.80 0.60 0.80

! G91
assign (residue 91 and name H) (residue 90 and name HA) 2.40 0.40 0.40
assign (residue 91 and name H) (residue 91 and name HA2) 3.00 0.60 0.60
assign (residue 91 and name H) (residue 91 and name HA3) 2.40 0.40 0.40
assign (residue 91 and name H) (residue 92 and name H) 3.60 0.40 0.80
assign (residue 91 and name H) (residue 93 and name QE) 3.40 0.60 0.80
! Flip HB1 and HB2
assign (residue 91 and name H) (residue 112 and name HB2) 3.30 0.60 0.60
assign (residue 91 and name H) (residue 112 and name HB3) 2.80 0.60 0.80
! Intensity on the 3D is very weak, try the 3D
assign (residue 91 and name H) (residue 112 and name QD1) 4.20 0.40 0.80
assign (residue 91 and name H) (residue 112 and name H) 3.00 0.40 0.40
! Intensity is also very weak, change HB3 to HG1, DCA
assign (residue 91 and name H) (residue 113 and name HG2) 4.40 0.40 0.80

! T92
assign (residue 92 and name HB) (residue 92 and name H) 2.60 0.40 0.40
assign (residue 92 and name QG2) (residue 92 and name HA) 3.00 0.40 0.60
assign (residue 92 and name QG2) (residue 92 and name HB) 2.40 0.40 0.40
assign (residue 92 and name QG2) (residue 94 and name HA) 3.60 0.40 0.80
! Intensity is very weak
assign (residue 92 and name QG2) (residue 94 and name QB) 4.00 0.40 1.00
! May be too far, DCA
! assign (residue 92 and name QG2) (residue 94 and name HD22) 3.40 0.40 0.40
! Intensity is very weak S/N is 6 in 2D
assign (residue 92 and name QG2) (residue 94 and name H) 4.00 0.40 1.00
assign (residue 92 and name QG2) (residue 110 and name H) 3.40 0.40 0.40
assign (residue 92 and name H) (residue 91 and name HA2) 2.60 0.40 0.40
assign (residue 92 and name H) (residue 91 and name HA3) 2.60 0.40 0.40
assign (residue 92 and name H) (residue 92 and name HA) 3.00 0.60 0.60
! Go with the NOE in 2D, S/N 55, DCA 3D
assign (residue 92 and name H) (residue 92 and name HB) 3.00 0.40 0.40
! Adjusted
assign (residue 92 and name H) (residue 92 and name QG2) 3.40 0.60 1.00

! Y93
! Too far, should be Y93HD# to L16HD1# which is previous assigned in L16
! assign (residue 93 and name QD) (residue 16 and name QD2) 2.80 0.40 0.40
assign (residue 93 and name QD) (residue 32 and name HZ3) 4.00 0.40 0.40
! NOE intensity S/N 14 in 2D
assign (residue 93 and name QD) (residue 33 and name HA) 4.20 0.40 1.00
NOE error suggest 4.1
assign (residue 93 and name QD) (residue 39 and name QG1) 4.00 0.40 0.40
! Too far for QD2, change to L80QD1 is closer
assign (residue 93 and name QD) (residue 80 and name QD1) 3.20 0.40 0.40
! Too far, DCA
! assign (residue 93 and name QD) (residue 92 and name QG2) 4.00 0.40 0.80
assign (residue 93 and name QD) (residue 93 and name HA) 2.80 0.40 0.40
assign (residue 93 and name QD) (residue 93 and name HB2) 2.50 0.40 0.40
assign (residue 93 and name QD) (residue 93 and name HB3) 2.50 0.40 0.40
assign (residue 93 and name QD) (residue 93 and name QE) 2.40 0.40 0.40
assign (residue 93 and name QD) (residue 93 and name H) 2.40 0.80 0.80

```

```

assign (residue 93 and name QE) (residue 82 and name QG1) 4.20 0.40 0.40
assign (residue 93 and name QE) (residue 90 and name HA) 3.40 0.40 0.40
assign (residue 93 and name HH) (residue 34 and name QD) 3.80 0.40 0.80
assign (residue 93 and name HH) (residue 89 and name HB3) 2.80 0.40 0.40
assign (residue 93 and name HH) (residue 90 and name HA) 3.60 0.40 0.80
assign (residue 93 and name HH) (residue 93 and name QD) 3.60 0.40 0.80
assign (residue 93 and name HH) (residue 93 and name QE) 2.50 0.40 0.40
assign (residue 93 and name H) (residue 92 and name HA) 2.30 0.40 0.40
assign (residue 93 and name H) (residue 92 and name HB) 3.60 0.40 0.80
assign (residue 93 and name H) (residue 92 and name QG2) 3.20 0.60 0.80
assign (residue 93 and name H) (residue 93 and name HA) 3.00 0.40 0.40
assign (residue 93 and name H) (residue 93 and name HB2) 2.60 0.60 0.80
assign (residue 93 and name H) (residue 93 and name HB3) 3.00 0.60 0.80
assign (residue 93 and name H) (residue 93 and name QD) 3.20 0.60 0.80
assign (residue 93 and name H) (residue 109 and name QB) 4.20 0.40 1.00
assign (residue 93 and name H) (residue 110 and name HB2) 3.60 0.60 0.80

! N94
assign (residue 94 and name QB) (residue 94 and name HA) 2.50 0.40 0.40
assign (residue 94 and name HD22) (residue 94 and name QB) 3.00 0.60 0.80
assign (residue 94 and name HD21) (residue 94 and name HD22) 2.20 0.40 0.40
! Weak NOE
assign (residue 94 and name HD22) (residue 92 and name QG2) 4.20 0.60 1.00
assign (residue 94 and name HD21) (residue 94 and name QB) 3.20 0.90 0.80
! Weak NOE, S/N 9
assign (residue 94 and name HD22) (residue 109 and name QB) 4.00 0.40 0.60
assign (residue 94 and name H) (residue 34 and name HA) 3.80 0.40 0.80
assign (residue 94 and name H) (residue 93 and name HA) 2.40 0.40 0.40
assign (residue 94 and name H) (residue 93 and name HB2) 3.60 0.60 0.80
assign (residue 94 and name H) (residue 93 and name HB3) 3.40 0.40 0.80
assign (residue 94 and name H) (residue 93 and name HD#) 3.60 0.40 0.80
assign (residue 94 and name H) (residue 94 and name HA) 3.40 0.60 0.80
assign (residue 94 and name H) (residue 94 and name QB) 2.80 0.60 0.60
! Too far, DCA
! assign (residue 94 and name H) (residue 109 and name QB) 3.80 0.60 0.80

! V95
assign (residue 95 and name HB) (residue 95 and name HA) 3.00 0.40 0.40
assign (residue 95 and name HB) (residue 95 and name QG1) 2.20 0.40 0.40
assign (residue 95 and name HB) (residue 95 and name QG2) 2.20 0.40 0.40
assign (residue 95 and name QG1) (residue 18 and name HA) 4.00 0.40 0.40
! Change to V95HG2#
assign (residue 95 and name QG2) (residue 19 and name HD#) 3.60 0.40 0.40
assign (residue 95 and name QG1) (residue 32 and name HA) 3.40 0.40 0.40
assign (residue 95 and name QG1) (residue 32 and name HE3) 3.60 0.40 0.60
assign (residue 95 and name QG1) (residue 32 and name HH2) 4.40 0.60 0.60
assign (residue 95 and name QG1) (residue 32 and name HZ3) 4.00 0.60 0.60
! Change to V95HG2#
assign (residue 95 and name QG2) (residue 94 and name HA) 3.80 0.40 0.60
assign (residue 95 and name QG1) (residue 95 and name HA) 2.40 0.40 0.40
! assign (residue 95 and name QG1) (residue 108 and name HB2) 3.20 0.40 0.40
! Change to V95HG2#
assign (residue 95 and name QG2) (residue 108 and name HB3) 3.60 0.40 0.40
! Change to V95HG2#
assign (residue 95 and name QG2) (residue 108 and name HG2) 4.20 0.40 0.40
! Change to V95HG2#
assign (residue 95 and name QG2) (residue 110 and name QD2) 4.20 0.40 0.40
assign (residue 95 and name QG2) (residue 18 and name HA) 3.40 0.60 0.80
assign (residue 95 and name QG1) (residue 30 and name QG2) 2.40 0.40 0.40
! assign (residue 95 and name QG2) (residue 30 and name QG2) 2.40 0.40 0.40
! Change to V95HG1#
assign (residue 95 and name QG1) (residue 32 and name HA) 3.40 0.40 0.40
assign (residue 95 and name QG2) (residue 94 and name HA) 4.00 0.40 0.40
assign (residue 95 and name QG2) (residue 95 and name HA) 2.90 0.40 0.40
assign (residue 95 and name QG2) (residue 95 and name QG1) 2.60 0.40 0.40
! Chage, NOE error
assign (residue 95 and name QG2) (residue 108 and name HB2) 2.20 0.40 0.40
assign (residue 95 and name H) (residue 94 and name HA) 2.20 0.40 0.40
assign (residue 95 and name H) (residue 94 and name QB) 3.20 0.60 0.80
! 2D S/N 4
assign (residue 95 and name H) (residue 94 and name HD21) 4.40 0.40 0.80
assign (residue 95 and name H) (residue 95 and name HA) 3.20 0.60 0.80
assign (residue 95 and name H) (residue 95 and name HB) 3.40 0.60 0.80
! Flip the next two line, HG1# to HG2#
assign (residue 95 and name H) (residue 95 and name QG2) 2.40 0.40 0.40
assign (residue 95 and name H) (residue 95 and name QG1) 3.60 0.60 0.80
! 2D S/N 5
assign (residue 95 and name H) (residue 96 and name H) 4.40 0.40 0.60
assign (residue 95 and name H) (residue 107 and name HA) 4.20 0.60 0.60
! assign (residue 95 and name H) (residue 108 and name HG2) 3.60 0.60 0.80
assign (residue 95 and name H) (residue 109 and name HA) 3.40 0.40 0.40

! T96
assign (residue 96 and name QG2) (residue 96 and name HA) 2.50 0.40 0.40
assign (residue 96 and name QG2) (residue 96 and name HB) 2.20 0.40 0.40
! Weak NOE, S/N 15
assign (residue 96 and name QG2) (residue 97 and name H) 4.00 0.40 0.40
assign (residue 96 and name QG2) (residue 98 and name QE) 3.60 0.60 0.80

```

```

assign (residue 96 and name QG2) (residue 107 and name HB2) 3.00 0.60 0.60
assign (residue 96 and name QG2) (residue 107 and name HB3) 3.20 0.40 0.60
assign (residue 96 and name H) (residue 32 and name HA) 3.80 0.60 1.00
assign (residue 96 and name H) (residue 95 and name HA) 2.60 0.40 0.60
assign (residue 96 and name H) (residue 95 and name HB) 3.60 0.60 0.80
! NOE error suggested 2.80
assign (residue 96 and name H) (residue 95 and name QG1) 2.70 0.40 0.40
! NOE error suggested 4.30
assign (residue 96 and name H) (residue 95 and name QG2) 4.50 0.40 0.60
assign (residue 96 and name H) (residue 96 and name HA) 2.80 0.60 0.60
assign (residue 96 and name H) (residue 96 and name HB) 3.40 0.60 0.80
! Weak NOE, S/N 18
assign (residue 96 and name H) (residue 96 and name QG2) 3.40 0.40 0.40

! V97
assign (residue 97 and name QG1) (residue 21 and name HD#) 3.40 0.40 0.40
assign (residue 97 and name QG1) (residue 97 and name HA) 3.60 0.40 0.40
assign (residue 97 and name QG1) (residue 97 and name HB) 2.20 0.40 0.40
assign (residue 97 and name QG2) (residue 30 and name QG1) 4.40 0.40 0.40
! DCA
! assign (residue 97 and name QG2) (residue 30 and name QG2) 2.60 0.60 0.40
assign (residue 97 and name QG2) (residue 97 and name HA) 2.40 0.40 0.40
assign (residue 97 and name QG2) (residue 97 and name HB) 2.20 0.40 0.40
assign (residue 97 and name QG2) (residue 97 and name QG1) 2.40 0.40 0.40
assign (residue 97 and name QG2) (residue 97 and name H) 3.80 0.40 0.40
assign (residue 97 and name H) (residue 96 and name HA) 2.40 0.40 0.40
assign (residue 97 and name H) (residue 96 and name QG2) 3.80 0.60 1.00
assign (residue 97 and name H) (residue 106 and name H) 3.00 0.40 0.60

! Y98
assign (residue 98 and name HB2) (residue 98 and name HA) 2.80 0.60 0.80
assign (residue 98 and name HB3) (residue 98 and name HA) 2.60 0.40 0.40
assign (residue 98 and name QD) (residue 29 and name HB2) 3.80 0.40 0.80
assign (residue 98 and name QD) (residue 96 and name HB) 3.60 0.60 0.80
assign (residue 98 and name QD) (residue 97 and name HA) 3.40 0.40 0.40
assign (residue 98 and name QD) (residue 98 and name HA) 3.40 0.40 0.40
assign (residue 98 and name QD) (residue 98 and name HB2) 2.40 0.40 0.40
assign (residue 98 and name QD) (residue 98 and name HB3) 2.40 0.40 0.40
! It farther than the NOE suggested, temp, DCA
! assign (residue 98 and name QD) (residue 99 and name H) 3.20 0.40 0.40
assign (residue 98 and name QD) (residue 103 and name H) 3.60 0.60 0.80
assign (residue 98 and name QE) (residue 96 and name HB) 3.40 0.40 0.40
! Very weak NOE, 2D S/N 4
assign (residue 98 and name QE) (residue 98 and name HA) 4.20 0.60 1.00
assign (residue 98 and name QE) (residue 98 and name QD) 2.40 0.40 0.40
! DCA
! assign (residue 98 and name QE) (residue 99 and name H) 3.60 0.60 0.80
assign (residue 98 and name QE) (residue 104 and name HA) 3.60 0.60 0.80
! 2D, S/N 15
assign (residue 98 and name H) (residue 27 and name QD1) 4.00 0.40 0.80
! 2D, S/N 5
assign (residue 98 and name H) (residue 28 and name H) 4.00 0.40 0.80
! DCA, the NOE is very weak, 2D S/N 7
assign (residue 98 and name H) (residue 29 and name HB2) 4.20 0.60 1.00
assign (residue 98 and name H) (residue 29 and name H) 2.80 0.40 0.40
assign (residue 98 and name H) (residue 97 and name HA) 2.40 0.40 0.40
assign (residue 98 and name H) (residue 97 and name QG1) 3.60 0.40 0.80
! NOE error suggest 2.55
assign (residue 98 and name H) (residue 97 and name QG2) 2.60 0.40 0.40
assign (residue 98 and name H) (residue 98 and name HA) 3.60 0.60 0.80
assign (residue 98 and name H) (residue 98 and name HB2) 2.40 0.40 0.40
! Temp take out
! assign (residue 98 and name H) (residue 98 and name HB3) 2.60 0.60 0.80
assign (residue 98 and name H) (residue 98 and name QD) 3.20 0.60 0.80

! S99
assign (residue 99 and name H) (residue 98 and name HA) 2.80 0.60 0.80
assign (residue 99 and name H) (residue 98 and name HB2) 4.00 0.40 0.40
assign (residue 99 and name H) (residue 99 and name HA) 3.60 0.60 0.80
! 3D S/N 10
assign (residue 99 and name H) (residue 100 and name H) 4.00 0.40 1.00
! 2D S/N 9
assign (residue 99 and name H) (residue 103 and name HA) 4.40 0.40 0.80
! 2D, S/N 9
assign (residue 99 and name H) (residue 104 and name HA) 4.00 0.40 0.60
! More peaks on the 3D, DCA

! T100
assign (residue 100 and name QG2) (residue 26 and name HA) 3.00 0.40 0.40
! 2D, S/N 4
assign (residue 100 and name QG2) (residue 27 and name H) 4.00 0.40 0.80
assign (residue 100 and name QG2) (residue 28 and name HA) 3.40 0.40 0.40
assign (residue 100 and name QG2) (residue 28 and name QB) 3.00 0.60 0.40
! 2D, S/N 10
assign (residue 100 and name QG2) (residue 99 and name HA) 4.00 0.40 0.80
assign (residue 100 and name QG2) (residue 100 and name HA) 2.60 0.40 0.40
assign (residue 100 and name QG2) (residue 100 and name HB) 2.60 0.40 0.40
! Weak NOE, 3D S/N 17

```

```

! assign (residue 100 and name H) (residue 27 and name QD1) 4.20 0.40 1.00
assign (residue 100 and name H) (residue 99 and name HA) 2.40 0.40 0.40
assign (residue 100 and name H) (residue 100 and name HA) 2.80 0.60 0.60
assign (residue 100 and name H) (residue 100 and name QG2) 2.40 0.40 0.40
! Too far, DCA
! assign (residue 100 and name H) (residue 102 and name HA2) 2.80 0.60 0.80

! N101
assign (residue 101 and name HD21) (residue 101 and name HA) 3.80 0.60 1.00
assign (residue 101 and name HD21) (residue 101 and name HB2) 2.80 0.40 0.40
assign (residue 101 and name HD21) (residue 101 and name HB3) 2.80 0.40 0.40
assign (residue 101 and name HD22) (residue 101 and name HA) 3.60 0.60 0.80
assign (residue 101 and name HD22) (residue 101 and name HB2) 3.60 0.60 0.80
assign (residue 101 and name HD22) (residue 101 and name HB3) 3.40 0.60 0.80
assign (residue 101 and name H) (residue 100 and name HA) 3.60 0.40 0.40

! G102
assign (residue 102 and name H) (residue 100 and name HA) 3.80 0.80 0.80
assign (residue 102 and name H) (residue 101 and name HA) 3.60 0.60 0.80
assign (residue 102 and name H) (residue 102 and name HA2) 2.80 0.60 0.60
assign (residue 102 and name H) (residue 102 and name HA3) 2.60 0.40 0.60

! T103
assign (residue 103 and name QG2) (residue 103 and name HA) 2.40 0.40 0.40
assign (residue 103 and name QG2) (residue 104 and name H) 3.40 0.40 0.40
! 3D S/N 9
assign (residue 103 and name H) (residue 99 and name HA) 4.20 0.40 1.00
! 3D S/N 5
assign (residue 103 and name H) (residue 98 and name HA) 4.20 0.40 0.80
assign (residue 103 and name H) (residue 102 and name HA2) 2.80 0.40 0.40
assign (residue 103 and name H) (residue 102 and name HA3) 3.40 0.60 0.80
assign (residue 103 and name H) (residue 102 and name H) 2.80 0.40 0.40
assign (residue 103 and name H) (residue 103 and name HA) 3.00 0.40 0.40
assign (residue 103 and name H) (residue 103 and name QG2) 2.80 0.60 0.80

! R104
assign (residue 104 and name H) (residue 103 and name HA) 2.40 0.40 0.40
assign (residue 104 and name H) (residue 103 and name QG2) 3.80 0.60 0.80

! I 105
assign (residue 105 and name HB) (residue 105 and name HA) 2.90 0.60 0.50
! NOE error suggested 2.48
assign (residue 105 and name QG2) (residue 105 and name HA) 2.50 0.40 0.40
! 3D S/N 7
assign (residue 105 and name H) (residue 97 and name H) 4.00 0.40 0.80
assign (residue 105 and name H) (residue 98 and name HA) 2.40 0.40 0.60
! 3D S/N 9
assign (residue 105 and name H) (residue 98 and name H) 4.20 0.40 1.00
assign (residue 105 and name H) (residue 104 and name HA) 2.60 0.50 0.40
assign (residue 105 and name H) (residue 105 and name HA) 3.20 0.60 0.80
! Lower to 3.00 from 3.40
assign (residue 105 and name H) (residue 106 and name H) 3.00 0.40 0.40

! L106
assign (residue 106 and name QD1) (residue 21 and name QD) 3.20 0.40 0.40
! 2D S/N 7
assign (residue 106 and name H) (residue 96 and name HA) 4.00 0.40 0.80

! N107
assign (residue 107 and name HB2) (residue 107 and name HA) 2.80 0.40 0.40
assign (residue 107 and name HB3) (residue 107 and name HA) 3.00 0.40 0.40
! Weak NOE, 3D S/N 15
assign (residue 107 and name HD21) (residue 107 and name HA) 4.00 0.60 1.00
assign (residue 107 and name HD21) (residue 107 and name HB2) 3.30 1.00 0.60
assign (residue 107 and name HD21) (residue 107 and name HB3) 3.40 0.60 0.80
assign (residue 107 and name HD21) (residue 107 and name HD22) 2.20 0.40 0.40
assign (residue 107 and name HD22) (residue 107 and name HA) 4.40 0.80 0.80
assign (residue 107 and name HD22) (residue 107 and name HB2) 2.40 0.40 0.80
assign (residue 107 and name H) (residue 106 and name HA) 2.00 0.20 0.40
assign (residue 107 and name H) (residue 106 and name QD2) 3.60 0.60 0.80
! 2D and 3D S/N 9
assign (residue 107 and name H) (residue 106 and name H) 4.20 0.40 0.80
assign (residue 107 and name H) (residue 107 and name HA) 3.40 0.60 0.80
assign (residue 107 and name H) (residue 107 and name HB2) 3.40 0.60 0.80
! NOE error suggested 2.60
assign (residue 107 and name H) (residue 107 and name HB3) 2.60 0.40 0.40

! K108
! NOE error suggested 2.50
assign (residue 108 and name HB3) (residue 108 and name HA) 2.50 0.40 0.40
assign (residue 108 and name HD3) (residue 108 and name HD2) 2.00 0.40 0.40
assign (residue 108 and name HG2) (residue 108 and name HE2) 3.30 0.40 0.40
assign (residue 108 and name HG2) (residue 108 and name HE3) 3.20 0.40 0.40
! NOE error suggested 4.65
assign (residue 108 and name H) (residue 94 and name HA) 4.70 0.40 0.40
! QG1 may be too far
! assign (residue 108 and name H) (residue 95 and name QG1) 2.80 0.60 0.80
! QG2 is closer

```

```

assign (residue 108 and name H) (residue 95 and name QG2) 3.40 0.60 0.80
assign (residue 108 and name H) (residue 95 and name H) 3.40 0.40 0.40
assign (residue 108 and name H) (residue 96 and name HA) 3.40 0.60 0.80
assign (residue 108 and name H) (residue 107 and name HA) 2.00 0.20 0.40
assign (residue 108 and name H) (residue 107 and name HB2) 3.60 0.60 0.80
assign (residue 108 and name H) (residue 107 and name HB3) 3.60 0.60 0.80
assign (residue 108 and name H) (residue 108 and name HA) 3.00 0.60 0.80
assign (residue 108 and name H) (residue 108 and name HB2) 3.20 0.60 0.80
assign (residue 108 and name H) (residue 108 and name HB3) 3.00 0.60 0.60

! A109
assign (residue 109 and name HA) (residue 109 and name QB) 2.20 0.40 0.40
! NOE error 3.95
assign (residue 109 and name QB) (residue 94 and name HA) 4.00 0.40 0.40
! Repeat, see above
! assign (residue 109 and name QB) (residue 109 and name HA) 2.40 0.40 0.40
! 2D S/N 4
assign (residue 109 and name H) (residue 94 and name HA) 4.40 0.40 0.80
assign (residue 109 and name H) (residue 108 and name HA) 2.00 0.20 0.40
assign (residue 109 and name H) (residue 108 and name HB2) 3.80 0.40 0.40
assign (residue 109 and name H) (residue 108 and name HB3) 2.60 0.40 0.40
assign (residue 109 and name H) (residue 108 and name HD2) 4.00 0.40 0.60
assign (residue 109 and name H) (residue 109 and name HA) 3.20 0.60 0.80
assign (residue 109 and name H) (residue 109 and name QB) 2.00 0.20 0.60
! Weak NOE, 2D S/N 4
assign (residue 109 and name H) (residue 110 and name H) 4.20 0.60 1.00

! L110
assign (residue 110 and name HB2) (residue 110 and name HA) 2.70 0.40 0.40
assign (residue 110 and name HB3) (residue 110 and name HA) 3.00 0.40 0.40
assign (residue 110 and name HB3) (residue 110 and name HB2) 2.00 0.40 0.40
assign (residue 110 and name H) (residue 92 and name HA) 3.60 0.60 0.80
assign (residue 110 and name H) (residue 92 and name QG2) 2.60 0.40 0.80
assign (residue 110 and name H) (residue 93 and name H) 2.80 0.40 0.40
assign (residue 110 and name H) (residue 94 and name HA) 3.40 0.40 0.40
assign (residue 110 and name H) (residue 109 and name HA) 2.20 0.40 0.40
assign (residue 110 and name H) (residue 109 and name QB) 2.80 0.60 0.60
assign (residue 110 and name H) (residue 110 and name HA) 3.20 0.60 0.80
assign (residue 110 and name H) (residue 110 and name HB2) 2.80 0.60 0.60
assign (residue 110 and name H) (residue 110 and name HB3) 3.40 0.60 0.80
assign (residue 110 and name H) (residue 110 and name QD2) 3.60 0.60 0.80

! D111
assign (residue 111 and name HB3) (residue 6 and name H) 3.60 0.60 0.80
assign (residue 111 and name HB2) (residue 111 and name HA) 2.60 0.40 0.40
! Weak NOE, 2D S/N 7
assign (residue 111 and name HB2) (residue 112 and name H) 4.00 0.60 1.00
assign (residue 111 and name HB3) (residue 5 and name HA) 2.60 0.40 0.40
assign (residue 111 and name HB3) (residue 111 and name HA) 3.20 0.40 0.40
! Weak NOE, 2D S/N 8
assign (residue 111 and name H) (residue 5 and name QG2) 4.00 0.60 1.00
assign (residue 111 and name H) (residue 110 and name HA) 2.00 0.20 0.40
assign (residue 111 and name H) (residue 110 and name HB2) 3.20 0.60 0.80
assign (residue 111 and name H) (residue 110 and name HB3) 2.80 0.60 0.60
assign (residue 111 and name H) (residue 111 and name HA) 3.20 0.60 0.80
assign (residue 111 and name H) (residue 111 and name HB2) 2.40 0.40 0.40
assign (residue 111 and name H) (residue 111 and name HB3) 2.40 0.40 0.40

! L112
! NOE error suggested 2.55
assign (residue 112 and name HB2) (residue 93 and name QD) 2.55 0.40 0.40
assign (residue 112 and name HB2) (residue 93 and name QE) 2.80 0.40 0.40
! CYANA error
assign (residue 112 and name HB3) (residue 93 and name QE) 3.00 0.70 0.40
assign (residue 112 and name HB3) (residue 112 and name HA) 2.60 0.40 0.40
assign (residue 112 and name HB3) (residue 112 and name HB2) 2.00 0.30 0.40
! Flip QD1 to QD2
assign (residue 112 and name QD1) (residue 7 and name HA) 3.40 0.40 0.40
assign (residue 112 and name QD2) (residue 14 and name QG1) 2.40 0.40 0.40
assign (residue 112 and name QD1) (residue 90 and name HA) 3.00 0.60 0.60
assign (residue 112 and name QD1) (residue 93 and name QD) 3.80 0.40 0.40
assign (residue 112 and name QD1) (residue 93 and name QE) 2.60 0.40 0.40
assign (residue 112 and name QD1) (residue 93 and name HH) 3.00 0.40 0.40
assign (residue 112 and name QD1) (residue 112 and name HA) 3.00 0.40 0.40
assign (residue 112 and name QD1) (residue 112 and name HB2) 2.40 0.40 0.70
assign (residue 112 and name QD1) (residue 112 and name HB3) 2.20 0.40 0.40
! Change according to NOE error 2.73 avg
assign (residue 112 and name HG) (residue 112 and name HA) 2.80 0.40 0.70
assign (residue 112 and name HG) (residue 112 and name HB2) 2.50 0.40 0.40
! Too far
! assign (residue 112 and name H) (residue 5 and name QG2) 3.80 0.60 0.80
assign (residue 112 and name H) (residue 93 and name QD) 3.40 0.60 0.80
assign (residue 112 and name H) (residue 93 and name QE) 3.60 0.60 0.80
assign (residue 112 and name H) (residue 93 and name H) 3.80 0.60 0.80
assign (residue 112 and name H) (residue 111 and name HA) 2.00 0.20 0.40
assign (residue 112 and name H) (residue 111 and name HB3) 4.00 0.40 0.40
assign (residue 112 and name H) (residue 112 and name HA) 3.40 0.60 0.80
assign (residue 112 and name H) (residue 112 and name HB2) 2.40 0.40 0.40

```

```

assign (residue 112 and name H) (residue 112 and name HB3) 2.90 0.40 0.40
! NOE error suggested 4.50
assign (residue 112 and name H) (residue 112 and name QD1) 4.60 0.40 0.60

! R113
! NOE error suggested 3.02
assign (residue 113 and name HB2) (residue 113 and name HA) 3.00 0.40 0.40
assign (residue 113 and name HB2) (residue 113 and name HD3) 3.20 0.60 0.40
assign (residue 113 and name HB3) (residue 113 and name HA) 2.60 0.45 0.45
! Weak NOE, 2D S/N 11
assign (residue 113 and name H) (residue 5 and name QG2) 4.00 0.60 0.80
! Weak NOE, 2D S/N 5
assign (residue 113 and name H) (residue 90 and name HB3) 4.20 0.40 1.00
assign (residue 113 and name H) (residue 112 and name HA) 2.30 0.20 0.40
assign (residue 113 and name H) (residue 112 and name HB2) 4.40 0.40 0.40
assign (residue 113 and name H) (residue 113 and name HA) 3.40 0.60 0.80
assign (residue 113 and name H) (residue 113 and name HB3) 3.20 0.60 0.80

! I114
! NOE error suggested 2.67
assign (residue 114 and name HB) (residue 114 and name HA) 2.50 0.40 0.40
! Below two line should be QG2 instead of QD1
! assign (residue 114 and name QG2) (residue 90 and name HB2) 2.20 0.40 0.40
assign (residue 114 and name QG2) (residue 90 and name HB3) 3.20 0.40 0.40
! DCA
! assign (residue 114 and name QD1) (residue 114 and name HA) 2.40 0.60 0.60
! DCA
! assign (residue 114 and name QD1) (residue 114 and name HB) 3.00 0.40 0.40
! assign (residue 114 and name QG2) (residue 9 and name HA) 3.20 0.40 0.40
assign (residue 114 and name QD1) (residue 10 and name H) 3.40 0.60 0.40
! NOE error suggested 3.12
assign (residue 114 and name QG2) (residue 114 and name HA) 3.00 0.60 0.40
assign (residue 114 and name QG2) (residue 114 and name HB) 2.40 0.40 0.40
assign (residue 114 and name H) (residue 113 and name HA) 2.00 0.20 0.40
assign (residue 114 and name H) (residue 113 and name HB3) 3.20 0.60 0.80
assign (residue 114 and name H) (residue 114 and name HA) 3.20 0.80 0.80
! DCA
! assign (residue 114 and name H) (residue 114 and name HB) 2.30 0.30 0.40
! NOE error suggested 4.24
assign (residue 114 and name H) (residue 114 and name QD1) 4.20 0.40 0.40
! Increased base on NOE error
assign (residue 114 and name H) (residue 114 and name HG12) 3.60 0.60 0.80
! DCA
! assign (residue 114 and name H) (residue 114 and name QG2) 3.20 0.60 0.80

! L115
assign (residue 115 and name HB2) (residue 115 and name HA) 2.60 0.40 0.40
assign (residue 115 and name HB3) (residue 115 and name HA) 2.80 0.40 0.40
assign (residue 115 and name QD1) (residue 7 and name QB) 2.80 0.70 1.00
! assign (residue 115 and name QD1) (residue 7 and name HD1) 3.40 0.40 0.40
! assign (residue 115 and name QD1) (residue 9 and name HA) 3.20 0.40 0.40
assign (residue 115 and name QD1) (residue 115 and name HA) 2.40 0.40 0.40
assign (residue 115 and name QD1) (residue 115 and name HB2) 2.20 0.40 0.40
assign (residue 115 and name QD1) (residue 115 and name HB3) 2.80 0.60 0.80
assign (residue 115 and name H) (residue 114 and name HA) 2.00 0.20 0.40
assign (residue 115 and name H) (residue 114 and name QG2) 2.80 0.60 0.80
assign (residue 115 and name H) (residue 115 and name HA) 3.20 0.60 0.80
assign (residue 115 and name H) (residue 115 and name HB2) 2.40 0.40 0.40
! NOE error on both HB1 and HB2, N150 shows HB2 is weaker, 2.8 to 3.4
assign (residue 115 and name H) (residue 115 and name HB3) 3.40 0.40 0.60
assign (residue 115 and name H) (residue 115 and name QD1) 2.30 0.40 0.60

! E116
assign (residue 116 and name HB2) (residue 116 and name QG) 2.50 0.40 0.40
assign (residue 116 and name HB3) (residue 116 and name HA) 3.20 0.40 0.40
assign (residue 116 and name HB3) (residue 116 and name QG) 2.30 0.40 0.40
assign (residue 116 and name QG) (residue 116 and name HA) 3.20 0.40 0.40
! assign (residue 116 and name H) (residue 114 and name QG2) 3.60 0.60 0.80
assign (residue 116 and name H) (residue 115 and name HA) 2.40 0.00 0.40
assign (residue 116 and name H) (residue 115 and name HB2) 2.80 0.60 0.80
assign (residue 116 and name H) (residue 115 and name HB3) 2.40 0.60 0.60
assign (residue 116 and name H) (residue 115 and name QD1) 2.80 0.60 0.60
assign (residue 116 and name H) (residue 115 and name H) 3.60 0.60 0.80
assign (residue 116 and name H) (residue 116 and name HA) 2.80 0.60 0.60
assign (residue 116 and name H) (residue 116 and name HB2) 3.20 0.60 0.80
assign (residue 116 and name H) (residue 116 and name HB3) 2.80 0.60 0.60
assign (residue 116 and name H) (residue 116 and name QG) 2.80 0.60 0.60

```

## Appendix 2.3 Input files and Scripts for CNS

### The Run Script for CNS

#### Generate the topology file (generate\_seq.inp)

```
{+ file: generate_seq.inp +}
{+ directory: general +}
{+ description: Generate structure file for protein, dna/rna, water,
               ligands and/or carbohydrate from sequence information only +}
{+ comment: modified by Brian Smith (Edinburgh University) to allow protein
           residue renumbering +}
{+ authors: Paul Adams, and Axel Brunger +}
{+ copyright: Yale University +}

{- Guidelines for using this file:
  - all strings must be quoted by double-quotes
  - logical variables (true/false) are not quoted
  - do not remove any evaluate statements from the file -}

{- Special patches will have to be entered manually at the relevant points
   in the file - see comments throughout the file -}

{- begin block parameter definition -} define(
===== input files =====}

{* multiple sequence files of the same type can be defined by duplicating
   the entries below and incrementing the file number *}

{* protein sequence file *}
====> prot_sequence_infile_1="sequence-5g3-52.seq";
{* segid *}
====> prot_segid_1="N122";
{* start residue numbering at *}
====> renumber_1=1;

{* nucleic acid sequence file *}
====> nucl_sequence_infile_1="";
{* segid *}
====> nucl_segid_1="";

{* water sequence file *}
====> water_sequence_infile_1="";
{* segid *}
====> water_segid_1="";

{* carbohydrate sequence file *}
====> carbo_sequence_infile_1="";
{* segid *}
====> carbo_segid_1="";

{* prosthetic group sequence file *}
====> prost_sequence_infile_1="";
{* segid *}
====> prost_segid_1="";

{* ligand sequence file *}
====> lig_sequence_infile_1="";
{* segid *}
====> lig_segid_1="";

{* ion sequence file *}
====> ion_sequence_infile_1="";
{* segid *}
====> ion_segid_1="";

===== output files =====}

{* output structure file *}
====> structure_outfile="5g3-52.mtf";
```

```

{===== disulphide bonds =====}

{* Select pairs of cysteine residues that form disulphide bonds *}
{* First 2 entries are the segid and resid of the first cysteine (CYS A). *}
{* Second 2 entries are the segid and resid of the second cysteine (CYS B). *}
+ table: rows=8 numbered
  cols=5 "use" "segid CYS A" "resid CYS A" "segid CYS B" "resid CYS B" +}

{+ choice: true false +}
====> ss_use_1=true;
====> ss_i_segid_1=""; ss_i_resid_1=11;
====> ss_j_segid_1=""; ss_j_resid_1=27;

{+ choice: true false +}
====> ss_use_2=true;
====> ss_i_segid_2=""; ss_i_resid_2=45;
====> ss_j_segid_2=""; ss_j_resid_2=73;

{+ choice: true false +}
====> ss_use_3=false;
====> ss_i_segid_3=""; ss_i_resid_3=0;
====> ss_j_segid_3=""; ss_j_resid_3=0;

{+ choice: true false +}
====> ss_use_4=false;
====> ss_i_segid_4=""; ss_i_resid_4=0;
====> ss_j_segid_4=""; ss_j_resid_4=0;

{+ choice: true false +}
====> ss_use_5=false;
====> ss_i_segid_5=""; ss_i_resid_5=0;
====> ss_j_segid_5=""; ss_j_resid_5=0;

{+ choice: true false +}
====> ss_use_6=false;
====> ss_i_segid_6=""; ss_i_resid_6=0;
====> ss_j_segid_6=""; ss_j_resid_6=0;

{+ choice: true false +}
====> ss_use_7=false;
====> ss_i_segid_7=""; ss_i_resid_7=0;
====> ss_j_segid_7=""; ss_j_resid_7=0;

{+ choice: true false +}
====> ss_use_8=false;
====> ss_i_segid_8=""; ss_i_resid_8=0;
====> ss_j_segid_8=""; ss_j_resid_8=0;

{===== carbohydrate links =====}

{* Select pairs of residues that are linked *}
{* First entry is the name of the patch residue. *}
{* Second and third entries are the resid and segid for the atoms
referenced by "-" in the patch. *}
{* Fourth and fifth entries are the resid and segid for the atoms
referenced by "+" in the patch *}
+ table: rows=6 numbered
  cols=6 "use" "patch name" "segid -" "resid -" "segid +" "resid +" +}

{+ choice: true false +}
====> carbo_use_1=false;
====> carbo_patch_1="B1N";
====> carbo_i_segid_1="BBBB"; carbo_i_resid_1=401;
====> carbo_j_segid_1="AAAA"; carbo_j_resid_1=56;

{+ choice: true false +}
====> carbo_use_2=false;
====> carbo_patch_2="B1N";
====> carbo_i_segid_2="BBBB"; carbo_i_resid_2=402;
====> carbo_j_segid_2="AAAA"; carbo_j_resid_2=182;

{+ choice: true false +}
====> carbo_use_3=false;
====> carbo_patch_3="";
====> carbo_i_segid_3=""; carbo_i_resid_3=0;
====> carbo_j_segid_3=""; carbo_j_resid_3=0;

```



```

{+ choice: true false +}
====> carbo_use 4=false;
====> carbo_patch 4="";
====> carbo_i_segid 4=""; carbo_i_resid 4=0;
====> carbo_j_segid 4=""; carbo_j_resid 4=0;

{+ choice: true false +}
====> carbo_use 5=false;
====> carbo_patch 5="";
====> carbo_i_segid 5=""; carbo_i_resid 5=0;
====> carbo_j_segid 5=""; carbo_j_resid 5=0;

{+ choice: true false +}
====> carbo_use 6=false;
====> carbo_patch 6="";
====> carbo_i_segid 6=""; carbo_i_resid 6=0;
====> carbo_j_segid 6=""; carbo_j_resid 6=0;

{===== generate parameters =====}

{* hydrogen flag - determines whether hydrogens will be retained *}
{* must be true for NMR, atomic resolution X-ray crystallography
or modelling. Set to false for most X-ray crystallographic
applications at resolution > 1Å *}
{+ choice: true false +}
====> hydrogen_flag=true;

{* set bfactor flag *}
{+ choice: true false +}
====> set_bfactor=false;

{* set bfactor value *}
====> bfactor=15.0;

{* set occupancy flag *}
{+ choice: true false +}
====> set_occupancy=false;

{* set occupancy value *}
====> occupancy=1.0;

{===== protein topology and parameter files =====}

{* protein topology file *}
====> prot_topology_infile="CNS_TOPPAR:protein-allhdg.top";

{* protein linkage file *}
====> prot_link_infile="CNS_TOPPAR:protein.link";

{* protein parameter file *}
====> prot_parameter_infile="CNS_TOPPAR:protein-allhdg.param";

{=====nucleic acid topology and parameter files =====}

{* nucleic acid topology file *}
====> nucl_topology_infile="CNS_TOPPAR:dna-rna-allatom.top";

{* nucleic acid linkage file *}
====> nucl_link_infile="CNS_TOPPAR:dna-rna.link";

{* nucleic acid parameter file *}
====> nucl_parameter_infile="CNS_TOPPAR:dna-rna-allatom.param";

{===== water topology and parameter files =====}

{* water topology file *}
====> water_topology_infile="CNS_TOPPAR:water.top";

{* water parameter file *}
====> water_parameter_infile="CNS_TOPPAR:water.param";

{===== carbohydrate topology and parameter files =====}

{* carbohydrate topology file *}
====> carbo_topology_infile="CNS_TOPPAR:carbohydrate.top";

```

```

{ * carbohydrate parameter file *}
{==>} carbo_parameter_infile="CNS_TOPPAR:carbohydrate.param";

{===== prosthetic group topology and parameter files =====}

{ * prosthetic group topology file *}
{==>} prost_topology_infile="";

{ * prosthetic group parameter file *}
{==>} prost_parameter_infile="";

{===== ligand topology and parameter files =====}

{ * ligand topology file *}
{==>} lig_topology_infile="";

{ * ligand parameter file *}
{==>} lig_parameter_infile="";

{===== ion topology and parameter files =====}

{ * ion topology file *}
{==>} ion_topology_infile="CNS_TOPPAR:ion.top";

{ * ion parameter file *}
{==>} ion_parameter_infile="CNS_TOPPAR:ion.param";

{=====
  things below this line do not need to be changed unless
  you need to apply patches - at the appropriate places marked
=====}

) {- end block parameter definition -}

checkversion 1.1

evaluate ($log_level=quiet)

topology
  if ( &BLANK%prot_topology_infile = false ) then
    @@&prot_topology_infile
  end if
  if ( &BLANK%nucl_topology_infile = false ) then
    @@&nucl_topology_infile
  end if
  if ( &BLANK%water_topology_infile = false ) then
    @@&water_topology_infile
  end if
  if ( &BLANK%carbo_topology_infile = false ) then
    @@&carbo_topology_infile
  end if
  if ( &BLANK%prost_topology_infile = false ) then
    @@&prost_topology_infile
  end if
  if ( &BLANK%lig_topology_infile = false ) then
    @@&lig_topology_infile
  end if
  if ( &BLANK%ion_topology_infile = false ) then
    @@&ion_topology_infile
  end if
end

parameter
  if ( &BLANK%prot_parameter_infile = false ) then
    @@&prot_parameter_infile
  end if
  if ( &BLANK%nucl_parameter_infile = false ) then
    @@&nucl_parameter_infile
  end if
  if ( &BLANK%water_parameter_infile = false ) then
    @@&water_parameter_infile
  end if
  if ( &BLANK%carbo_parameter_infile = false ) then
    @@&carbo_parameter_infile
  end if

```

```

if ( &BLANK%prost_parameter_infile = false ) then
  @@&prost_parameter_infile
end if
if ( &BLANK%lig_parameter_infile = false ) then
  @@&lig_parameter_infile
end if
if ( &BLANK%ion_parameter_infile = false ) then
  @@&ion_parameter_infile
end if
end

evaluate ($counter=1)
evaluate ($done=false)
while ( $done = false ) loop prot
  if ( &exist_prot_sequence_infile $counter = true ) then
    if ( &BLANK%prot_sequence_infile_$counter = false ) then
      do (refx=0) (all)
        segment
          chain
            @@&prot_link_infile
            sequence @@&prot_sequence_infile_$counter end
          end
        end
        do (segid="T^" + encode($counter)) (attr refx=9999)
      end if
      if ( &exist_renumber_$counter = true ) then
        if ( &BLANK%renumber_$counter = false ) then
          evaluate ($segid="T^" + encode($counter))
          do ( resid = encode(decode(resid) + &renumber_$counter - 1))
            ( attr refx=9999) and segid $segid )
          end if
        end if
        evaluate ($counter=$counter+1)
      else
        evaluate ($done=true)
      end if
    end loop prot

    evaluate ($counter=1)
    evaluate ($done=false)
    while ( $done = false ) loop nseg
      if ( &exist_prot_sequence_infile $counter = true ) then
        if ( &BLANK%prot_sequence_infile_$counter = false ) then
          evaluate ($segtmp="T^" + encode($counter))
          do (segid=capitalize(&prot_segid_$counter)) (segid $segtmp)
          end if
          evaluate ($counter=$counter+1)
        else
          evaluate ($done=true)
        end if
      end loop nseg

      evaluate ($ssc=1)
      evaluate ($done=false)
      while ( $done = false ) loop ssbr
        if ( &exist_ss_use_$ssc = true ) then
          if ( &ss_use_$ssc = true ) then
            evaluate ($segidtmp1=capitalize(&ss_i_segid_$ssc))
            evaluate ($segidtmp2=capitalize(&ss_j_segid_$ssc))
            patch disu
              reference=1=(segid $QUOTE%segidtmp1 and resid &ss_i_resid_$ssc)
              reference=2=(segid $QUOTE%segidtmp2 and resid &ss_j_resid_$ssc)
            end
          end if
          evaluate ($ssc=$ssc+1)
        else
          evaluate ($done=true)
        end if
      end loop ssbr

      { * any special protein patches can be applied here * }
      {====>}

      {<===}

      evaluate ($counter=1)

```

```

evaluate ($done=false)
while ( $done = false ) loop nucl
  if ( &exist nucl_sequence_infile_$counter = true ) then
    if ( &BLANK%nucl_sequence_infile_$counter = false ) then
      do (refx=0) (all)
        segment
        chain
          @@&nucl_link_infile
          sequence @@&nucl_sequence_infile_$counter end
        end
      end
      do (segid=capitalize(&nucl_segid_$counter)) (attr refx=9999)
    end if
    evaluate ($counter=$counter+1)
  else
    evaluate ($done=true)
  end if
end loop nucl

{* patch rna sugars to dna here if needed - select the residues *}
{====>}
  for $resid in ( ) loop dna
    patch deox reference=nil=(resid $resid) end
  end loop dna
{<===}

{* any special nucleic acid patches can be applied here *}
{====>}
{<===}

evaluate ($counter=1)
evaluate ($done=false)
while ( $done = false ) loop carbo
  if ( &exist_carbo_sequence_infile_$counter = true ) then
    if ( &BLANK%carbo_sequence_infile_$counter = false ) then
      do (refx=0) (all)
        segment
        chain
          sequence @@&carbo_sequence_infile_$counter end
        end
      end
      do (segid=capitalize(&carbo_segid_$counter)) (attr refx=9999)
    end if
    evaluate ($counter=$counter+1)
  else
    evaluate ($done=true)
  end if
end loop carbo

evaluate ($carc=1)
evaluate ($done=false)
while ( $done = false ) loop cabr
  if ( &exist_carbo_use_$carc = true ) then
    if ( &carbo_use_$carc = true ) then
      evaluate ($segidtmp1=capitalize(&carbo_i_segid_$carc))
      evaluate ($segidtmp2=capitalize(&carbo_j_segid_$carc))
      patch &carbo_patch_$carc
      reference--=(segid $QUOTE%segidtmp1 and
                    resid &carbo_i_resid_$carc)
      reference+=(segid $QUOTE%segidtmp2 and
                    resid &carbo_j_resid_$carc)
    end
  end if
  evaluate ($carc=$carc+1)
else
  evaluate ($done=true)
end if
end loop cabr

{* any special carbohydrate patches can be applied here *}
{====>}
{<===}

evaluate ($counter=1)

```

```

evaluate ($done=false)
while ( $done = false ) loop prost
  if ( &exist prost_sequence_infile_$counter = true ) then
    if ( &BLANK%prost_sequence_infile_$counter = false ) then
      do (refx=0) (all)
        segment
        chain
        sequence @@&prost_sequence_infile_$counter end
      end
    end
    do (segid=capitalize(&prost_segid_$counter)) (attr refx=9999)
  end if
  evaluate ($counter=$counter+1)
else
  evaluate ($done=true)
end if
end loop prost

{* any special prosthetic group patches can be applied here *}
{===>}

{<===}

evaluate ($counter=1)
evaluate ($done=false)
while ( $done = false ) loop liga
  if ( &exist lig_sequence_infile_$counter = true ) then
    if ( &BLANK%lig_sequence_infile_$counter = false ) then
      do (refx=0) (all)
        segment
        chain
        sequence @@&lig_sequence_infile_$counter end
      end
    end
    do (segid=capitalize(&lig_segid_$counter)) (attr refx=9999)
  end if
  evaluate ($counter=$counter+1)
else
  evaluate ($done=true)
end if
end loop liga

{* any special ligand patches can be applied here *}
{===>}

{<===}

evaluate ($counter=1)
evaluate ($done=false)
while ( $done = false ) loop ion
  if ( &exist ion_sequence_infile_$counter = true ) then
    if ( &BLANK%ion_sequence_infile_$counter = false ) then
      do (refx=0) (all)
        segment
        chain
        sequence @@&ion_sequence_infile_$counter end
      end
    end
    do (segid=capitalize(&ion_segid_$counter)) (attr refx=9999)
  end if
  evaluate ($counter=$counter+1)
else
  evaluate ($done=true)
end if
end loop ion

{* any special ion patches can be applied here *}
{===>}

{<===}

evaluate ($counter=1)
evaluate ($done=false)
while ( $done = false ) loop water
  if ( &exist water_sequence_infile_$counter = true ) then
    if ( &BLANK%water_sequence_infile_$counter = false ) then

```

```

do (refx=0) (all)
segment
chain
sequence @@&water_sequence_infile_$counter end
end
end
do (segid=capitalize(&water_segid_$counter)) (attr refx=9999)
end if
evaluate ($counter=$counter+1)
else
evaluate ($done=true)
end if
end loop water

{* any special water patches can be applied here *}
{====>}

{<====}

{* any final patches can be applied here *}
{====>}

{<====}

if (&hydrogen_flag=false) then
delete selection=( hydrogen ) end
end if

if (&set_bfactor=true) then
do (b=&bfactor) ( all )
end if

if (&set_occupancy=true) then
do (q=&occupancy) ( all )
end if

write structure output=&structure_outfile end

stop

```

## Generated Extended PDB file for CaM-CD2-III-5G (generate\_extended.inp)

```

{+ file: generate_extended.inp +}
{+ directory: nmr_calc +}
{+ description: Generates an extended strand with ideal geometry
for each connected polymer.
The molecular structure file must not contain any
closed loops except disulfide bonds which are automatically
excluded from the generation of the strand conformation. +}
{+ authors: Axel T. Brunger +}
{+ copyright: Yale University +}

{- begin block parameter definition -} define(
{===== molecular structure =====}

{* structure file(s) *}
{====>} structure_file="5g3-52.mtf";

{* parameter file(s) *}
{====>} par_1="CNS_TOPPAR:protein-allhdg.param";
{====>} par_2="";
{====>} par_3="";
{====>} par_4="";
{====>} par_5="";

{===== input parameters =====}

{* maximum number of trials to generate an acceptable structure *}
{====>} max_trial=10;

{===== output files =====}

```

```

{ * output coordinates *}
{====> } output_coor="5g3-52_extended.pdb";

{=====
  things below this line do not normally need to be changed
=====}

) {- end block parameter definition -}

checkversion 1.1

evaluate ($log_level=quiet)

structure @&structure_file end

parameter
  if (&par_1 # " ") then
    @@&par_1
  end if
  if (&par_2 # " ") then
    @@&par_2
  end if
  if (&par_3 # " ") then
    @@&par_3
  end if
  if (&par_4 # " ") then
    @@&par_4
  end if
  if (&par_5 # " ") then
    @@&par_5
  end if
end

{ Set force constants for S-S bond lengths and angles to zero }
parameter
  bonds ( name SG ) ( name SG ) 0. 1.
end

igroup interaction=(all) (all) end

ident (x) ( all )
do (x=x/5.) ( all )
do (y=random(0.5) ) ( all )
do (z=random(0.5) ) ( all )

flags exclude * include bond angle impr dihe vdw end
parameter
  nbonds
    rcon=50. nbxmod=-3 repel=0.8 cutnb=6.
    rexp=2 irexp=2 inhibit=0.0 wmin=0.1 tolerance=0.5
  end
end

evaluate ($count=1)
while ($count < 10 ) loop l1
  do (x=x+gauss(0.1)) ( all )
  do (y=y+gauss(0.1)) ( all )
  do (z=z+gauss(0.1)) ( all )
  minimize powell nstep=200 nprint=10 end
  evaluate ($count=$count+1)
end loop l1

evaluate ($accept=false)
evaluate ($trial=1)
while ($accept=false) loop accp
  for $1 in id ( tag ) loop resi
    igroup
      interaction=( byresidue (id $1 ) and not name SG )
      ( not name SG )
    end
    evaluate ($accept=true)
    print thres=0.1 bonds
  end
end

```

```

        if ($violations > 0) then
            evaluate ($accept=false)
        end if
        print thres=10. angles
        evaluate ($angles=$result)
        if ($violations > 0) then
            evaluate ($accept=false)
        end if
        print thres=10. improper
        if ($violations > 0) then
            evaluate ($accept=false)
        end if
        if ($accept=false) then
            do (x=x+gauss(0.3)) ( byresidue (id $1 ) )
            do (y=y+gauss(0.3)) ( byresidue (id $1 ) )
            do (z=z+gauss(0.3)) ( byresidue (id $1 ) )
        end if
    end loop resi
igroup interaction=( all ) ( all ) end

parameter
    nbonds
        rcon=50. nbxmod=-3 repel=3. cutnb=10.
    end
end
flags exclude angle improper end

minimize powell nstep=200 nprint=10 end

parameter
    nbonds
        rcon=50. nbxmod=-3 repel=0.8 cutnb=6.
    end
end
flags include angle improper end

evaluate ($count=1)
while ($count < 5 ) loop 12
    do (x=x+gauss(0.05)) ( all )
    do (y=y+gauss(0.05)) ( all )
    do (z=z+gauss(0.05)) ( all )
    minimize powell nstep=200 nprint=10 end
    evaluate ($count=$count+1)
end loop 12

parameter
    nbonds
        rcon=50. nbxmod=3 repel=0.8 cutnb=6.
    end
end

minimize powell nstep=300 nprint=10 end
minimize powell nstep=300 nprint=10 end

igroup interaction=( not name SG ) ( not name SG ) end
energy end

evaluate ($accept=true)

print thres=0.05 bonds
evaluate ($bonds=$result)
if ($violations > 0) then
    evaluate ($accept=false)
end if

print thres=10. angles
evaluate ($angles=$result)
if ($violations > 0) then
    evaluate ($accept=false)
end if

print thres=10. improper
evaluate ($impr=$result)
if ($violations > 0) then
    evaluate ($accept=false)
end if

```



```

end if

print thres=180. dihedral
evaluate ($dihe=$result)

evaluate ($trial=$trial + 1)
if ($trial > &max_trial ) then
    exit loop accp
end if

end loop accp

remarks extended strand(s) generation
remarks input molecular structure file=&structure_file
remarks final rms deviations (excluding disulfide bonds):
remarks    bonds=          $bonds[F8.4] A
remarks    angles=        $angles[F8.4] degrees
remarks    impropers= $impr[F8.4] degrees
remarks    dihedrals= $dihe[F8.4] degrees (not used in some parameter sets!)
remarks final van der Waals (repel) energy=$vdw kcal/mole

write coordinates output=&output_coor end

stop

```

### anneal.inp (This is only the input portion of the anneal.inp)

```

{+ file: anneal.inp +}
{+ directory: nmr_calc +}
{+ description: dynamical annealing with NOEs, coupling constants,
               chemical shift restraints starting from extended
               strands or pre-folded structures. +}
{+ authors: Gregory Warren, Michael Nilges, John Kuszewski,
             Marius Clore and Axel Brunger +}
{+ copyright: Yale University +}

{+ reference: Clore GM, Gronenborn AM, Tjandra N, Direct structure refinement
               against residual dipolar couplings in the presence of rhombicity
               of unknown magnitude., J. Magn. Reson., 131, In press, (1998) +}
{+ reference: Clore GM, Gronenborn AM, Bax A, A robust method for determining
               the magnitude of the fully asymmetric alignment tensor of
               oriented macromolecules in the absence of structural
               information., J. Magn. Reson., In press (1998) +}
{+ reference: Garrett DS, Kuszewski J, Hancock TJ, Lodi PJ, Vuister GW,
               Gronenborn AM, Clore GM, The impact of direct refinement against
               three-bond HN-C alpha H coupling constants on protein structure
               determination by NMR., J. Magn. Reson. Ser. B, 104(1),
               99-103, (1994) May +}
{+ reference: Kuszewski J, Qin J, Gronenborn AM, Clore GM, The impact of direct
               refinement against 13C alpha and 13C beta chemical shifts on
               protein structure determination by NMR., J. Magn. Reson. Ser. B,
               106(1), 92-6, (1995) Jan +}
{+ reference: Kuszewski J, Gronenborn AM, Clore GM, The impact of direct
               refinement against proton chemical shifts on protein structure
               determination by NMR., J. Magn. Reson. Ser. B, 107(3), 293-7,
               (1995) Jun +}
{+ reference: Kuszewski J, Gronenborn AM, Clore GM, A potential involving
               multiple proton chemical-shift restraints for
               nonstereospecifically assigned methyl and methylene protons.
               J. Magn. Reson. Ser. B, 112(1), 79-81, (1996) Jul. +}
{+ reference: Nilges M, Gronenborn AM, Brunger AT, Clore GM, Determination
               of three-dimensional structures of proteins by simulated
               annealing with interproton distance restraints: application
               to crambin, potato carboxypeptidase inhibitor and barley
               serine proteinase inhibitor 2. Protein Engineering 2,
               27-38, (1988) +}
{+ reference: Nilges M, Clore GM, Gronenborn AM, Determination of
               three-dimensional structures of proteins from interproton
               distance data by dynamical simulated annealing from a random
               array of atoms. FEBS Lett. 239, 129-136. (1988) +}
{+ reference: Rice LM, Brunger AT, Torsion Angle Dynamics: Reduced Variable
               Conformational Sampling Enhances Crystallographic Structure
               Refinement., Proteins, 19, 277-290 (1994) +}

```

```

{+ reference: Stein EG, Rice LM, Brunger AT, Torsion angle molecular
dynamics: a new efficient tool for NMR structure calculation.,
J. Mag. Res. Ser. B 124, 154-164 (1997) +}
{+ reference: Tjandra N, Garrett DS, Gronenborn AM, Bax A, Clore GM, Defining
long range order in NMR structure determination from the
dependence of heteronuclear relaxation times on rotational
diffusion anisotropy. Nature Struct. Biol., 4(6), 443-9,
(1997) June +}
{+ reference: Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A, Use of
dipolar 1H-15N and 1H-13C couplings in the structure
determination of magnetically oriented macromolecules in
solution. Nature Struct. Biol., 4(9), 732-8, (1997) Sept +}

! Data taken from: Qin J, Clore GM, Kennedy WP, Kuszewski J, Gronenborn AM,
!                 The solution structure of human thioredoxin complexed with
!                 its target from Ref-1 reveals peptide chain reversal.,
!                 Structure, 4(5), 613-620, 1996 May 15.

{- Guidelines for using this file:
  - all strings must be quoted by double-quotes
  - logical variables (true/false) are not quoted
  - do not remove any evaluate statements from the file -}

{- begin block parameter definition -} define(

{===== molecular structure =====}

{* parameter file(s) *}
{==>} par.1="CNS_TOPPAR:protein-allhdg.param";
{==>} par.2="CNS_TOPPAR:ion.param";
{==>} par.3="";
{==>} par.4="";
{==>} par.5="";

{* structure file(s) *}
{==>} struct.1="5g3-52.mtf";
{==>} struct.2="";
{==>} struct.3="";
{==>} struct.4="";
{==>} struct.5="";

{* input coordinate file(s) *}
{==>} pdb.in.file.1="5g3-52_extended.pdb";
{==>} pdb.in.file.2="";
{==>} pdb.in.file.3="";

{===== atom selection =====}

{* input "backbone" selection criteria for average structure generation *}
{* for protein      (name n or name ca or name c)
   for nucleic acid (name O5' or name C5' or name C4' or name C3'
                        or name O3' or name P) *}
{==>} pdb.atom.select=(name n or name ca or name c);

{===== refinement parameters =====}

{* type of molecular dynamics for hot phase *}
{+ choice: "torsion" "cartesian" +}
{==>} md.type.hot="torsion";

{* type of molecular dynamics for cool phase *}
{+ choice: "torsion" "cartesian" +}
{==>} md.type.cool="torsion";

{* refine using different initial velocities or coordinates
   (enter base name in "input coordinate files" field) *}
{+ choice: "veloc" "coord" +}
{==>} md.type.initial="veloc";

{* seed for random number generator *}
{* change to get different initial velocities *}
{==>} md.seed=82364;

{* select whether the number of structures will be either trial or
   accepted structures and whether to print only the trial, accepted,

```

```

    both sets of structures. *}
{+ list: The printing format is as follows:
    trial = pdb.out.name + _#.pdb , accepted = pdb.out.name + a_#.pdb +}

{* are the number of structures to be trials or accepted? *}
+ choice: "trial" "accept" +}
==> flg.trial.struc="trial";
{* number of trial or accepted structures *}
==> pdb.end.count=100;

{* print accepted structures *}
+ choice: true false +}
==> flg.print.accept=false;
{* print trial structures *}
+ choice: true false +}
==> flg.print.trial=true;

{* calculate an average structure for either the trial or
  accepted structure. If calculate accepted average is false then
  an average for the trial structures will be calculated. *}

{* calculate an average structure? *}
+ choice: true false +}
==> flg.calc.ave.struct=true;
{* calculate an average structure for the accepted structures? *}
+ choice: true false +}
==> flg.calc.ave.acpt=false;
{* minimize average coordinates? *}
+ choice: true false +}
==> flg.min.ave.coor=false;

{===== torsion dynamics parameters =====}

{* maximum unbranched chain length *}
{* increase for long stretches of polyalanine or for nucleic acids *}
==> md.torsion.maxlength=50;

{* maximum number of distinct bodies *}
==> md.torsion.maxtree=4;

{* maximum number of bonds to an atom *}
==> md.torsion.maxbond=6;

{===== parameters for high temperature annealing stage =====}

{* temperature (proteins: 50000, dna/rna: 20000) *}
==> md.hot.temp=50000;
{* number of steps (proteins: 1000, dna/rna: 4000) *}
==> md.hot.step=8000;
{* scale factor to reduce van der Waals (repel) energy term *}
==> md.hot.vdw=0.1;
{* scale factor for NOE energy term *}
==> md.hot.noe=150;
{* scale factor for dihedral angle energy term (proteins: 100, dna/rna: 5) *}
==> md.hot.cdih=100;
{* molecular dynamics timestep *}
==> md.hot.ss=0.015;

{===== parameters for the first slow-cool annealing stage =====}

{* temperature (cartesian: 1000, torsion: [proteins: 50000, dna/rna: 20000]) *}
==> md.cool.temp=50000;
{* number of steps *}
==> md.cool.step=6000;
{* scale factor for final van der Waals (repel) energy term
  (cartesian: 4.0, torsion: 1.0) *}
==> md.cool.vdw=1.0;
{* scale factor for NOE energy term *}
==> md.cool.noe=150;
{* scale factor for dihedral angle energy term *}
==> md.cool.cdih=200;
{* molecular dynamics timestep (cartesian: 0.005, torsion: 0.015) *}
==> md.cool.ss=0.015;
{* slow-cool annealing temperature step (cartesian: 25, torsion: 250) *}
==> md.cool.tmpstp=250;

```

```

{===== parameters for a second slow-cool annealing stage =====}

{* cartesian slow-cooling annealing stage to be used only with torsion
slow-cool annealing stage *}
{* this stage is only necessary when the macromolecule is a protein
greater than 160 residues or in some cases for nucleic acids *}

{* use cartesian cooling stage? *}
+ choice: true false +}
==> md.cart.flag=true;
{* temperature *}
==> md.cart.temp=2000;
{* number of steps *}
==> md.cart.step=5000;
{* scale factor for initial van der Waals (repel) energy term *}
==> md.cart.vdw.init=1.0;
{* scale factor for final van der Waals (repel) energy term *}
==> md.cart.vdw.finl=4.0;
{* scale factor for NOE energy term *}
==> md.cart.noe=150;
{* scale factor for dihedral angle energy term *}
==> md.cart.cdih=400;
{* molecular dynamics timestep *}
==> md.cart.ss=0.005;
{* slow-cool annealing temperature step *}
==> md.cart.tmpstp=25;

{===== parameters for final minimization stage =====}

{* scale factor for NOE energy term *}
==> md.pow.noe=75;
{* scale factor for dihedral angle energy term *}
==> md.pow.cdih=1000;
{* number of minimization steps *}
==> md.pow.step=2000;
{* number of cycles of minimization *}
==> md.pow.cycl=10;

{===== noe data =====}

{- Important - if you do not have a particular data set then
set the file name to null ("") -}

{* NOE distance restraints files. *}

{* restraint set 1 file *}
==> nmr.noe.file.1="5g3-52-noe-cyana2-cy24r3-to-cns.tbl";
{* restraint set 2 file *}
==> nmr.noe.file.2="";
{* restraint set 3 file *}
==> nmr.noe.file.3="";
{* restraint set 4 file *}
==> nmr.noe.file.4="";
{* restraint set 5 file *}
==> nmr.noe.file.5="";

{* NOE averaging modes *}

{* restraint set 1 *}
+ choice: "sum" "cent" "R-6" "R-3" "symm" +}
==> nmr.noe.ave.mode.1="sum";
{* restraint set 2 *}
+ choice: "sum" "cent" "R-6" "R-3" "symm" +}
==> nmr.noe.ave.mode.2="sum";
{* restraint set 3 *}
+ choice: "sum" "cent" "R-6" "R-3" "symm" +}
==> nmr.noe.ave.mode.3="R-6";
{* restraint set 4 *}
+ choice: "sum" "cent" "R-6" "R-3" "symm" +}
==> nmr.noe.ave.mode.4="";
{* restraint set 5 *}
+ choice: "sum" "cent" "R-6" "R-3" "symm" +}
==> nmr.noe.ave.mode.5="";

{===== hydrogen bond data =====}

```

```

{ * hydrogen-bond distance restraints file. *}
{==>} nmr.noe.hbnd.file="";

{ * enter hydrogen-bond distance averaging mode *}
{ + choice: "sum" "cent" "R-6" "R-3" "symm" +}
{==>} nmr.noe.ave.mode.hbnd="sum";

{===== 3-bond J-coupling data =====}
{ * the default setup is for the phi dihedral *}

{ * Class 1 *}

{ * 3-bond J-coupling non-glycine restraints file *}
{==>} nmr.jcoup.file.1="";
{ * 3-bond J-coupling non-glycine potential *}
{ + choice: "harmonic" "square" "multiple" +}
{==>} nmr.jcoup.pot.1="harmonic";
{ * 3-bond J-coupling non-glycine force value *}
{==>} nmr.jcoup.force.1.1=1;
{ * 3-bond j-coupling multiple class force second value *}
{==>} nmr.jcoup.force.2.1=0;
{ * 3-bond j-coupling Karplus coefficients *}
{ * the default values are for phi *}
{==>} nmr.jcoup.coef.1.1=6.98;
{==>} nmr.jcoup.coef.2.1=-1.38;
{==>} nmr.jcoup.coef.3.1=1.72;
{==>} nmr.jcoup.coef.4.1=-60.0;

{ * Class 2 *}

{ * 3-bond j-coupling glycine restraints files *}
{==>} nmr.jcoup.file.2="";
{ * 3-bond J-coupling glycine potential *}
{ * The potential for the glycine class must be multiple *}
{ + choice: "harmonic" "square" "multiple" +}
{==>} nmr.jcoup.pot.2="multiple";
{ * 3-bond J-coupling first force value *}
{==>} nmr.jcoup.force.1.2=1;
{ * 3-bond j-coupling glycine or multiple force second value *}
{==>} nmr.jcoup.force.2.2=0;
{ * 3-bond j-coupling Karplus coefficients *}
{ * the default values are for glycine phi *}
{==>} nmr.jcoup.coef.1.2=6.98;
{==>} nmr.jcoup.coef.2.2=-1.38;
{==>} nmr.jcoup.coef.3.2=1.72;
{==>} nmr.jcoup.coef.4.2=0.0;

{===== 1-bond heteronuclear J-coupling data =====}

{ * Class 1 *}

{ * 1-bond heteronuclear j-coupling file *}
{==>} nmr.oneb.file.1="";
{ * 1-bond heteronuclear j-coupling potential *}
{ + choice: "harmonic" "square" +}
{==>} nmr.oneb.pot.1="harmonic";
{ * 1-bond heteronuclear j-coupling force value *}
{==>} nmr.oneb.force.1=1.0;

{===== alpha/beta carbon chemical shift data =====}

{ * Class 1 *}

{ * carbon, alpha and beta, chemical shift restraints file *}
{==>} nmr.carb.file.1="";
{ * carbon, alpha and beta, chemical shift restraint potential *}
{ + choice: "harmonic" "square" +}
{==>} nmr.carb.pot.1="harmonic";
{ * carbon, alpha and beta, chemical shift restraint force value *}
{==>} nmr.carb.force.1=0.5;

{===== proton chemical shift data =====}

{ * Class 1 *}

{ * class 1 proton chemical shift restraints file *}

```

```

{====> nmr.prot.file.1="";
* class 1 proton chemical shift potential *}
+ choice: "harmonic" "square" "multiple" +}
{====> nmr.prot.pot.1="harmonic";
* class 1 proton chemical shift force value *}
{====> nmr.prot.force.1.1=7.5;
* 2nd class 1 proton chemical shift force value for multi *}
{====> nmr.prot.force.2.1=0;
* class 1 proton chemical shift violation cutoff threshold *}
{====> nmr.prot.thresh.1=0.3;

{* Class 2 *}

{* class 2 proton chemical shift restraints file *}
{====> nmr.prot.file.2="";
* class 2 proton chemical shift potential *}
+ choice: "harmonic" "square" "multiple" +}
{====> nmr.prot.pot.2="harmonic";
* class 2 proton chemical shift force value *}
{====> nmr.prot.force.1.2=7.5;
* 2nd class 2 proton chemical shift force value for multi *}
{====> nmr.prot.force.2.2=0;
* class 2 proton chemical shift violation cutoff threshold *}
{====> nmr.prot.thresh.2=0.3;

{* Class 3 *}

{* class 3 proton chemical shift restraints file *}
{====> nmr.prot.file.3="";
* class 3 proton chemical shift potential *}
+ choice: "harmonic" "square" "multiple" +}
{====> nmr.prot.pot.3="harmonic";
* class 3 proton chemical shift force value *}
{====> nmr.prot.force.1.3=7.5;
* 2nd class 3 proton chemical shift force value for multi *}
{====> nmr.prot.force.2.3=0;
* class 3 proton chemical shift violation cutoff threshold *}
{====> nmr.prot.thresh.3=0.3;

{* Class 4 *}

{* class 4 proton chemical shift restraints file *}
{====> nmr.prot.file.4="";
* class 4 proton chemical shift potential *}
+ choice: "harmonic" "square" "multiple" +}
{====> nmr.prot.pot.4="multiple";
* class 4 proton chemical shift force value *}
{====> nmr.prot.force.1.4=7.5;
* 2nd class 4 proton chemical shift force value for multi *}
{====> nmr.prot.force.2.4=0;
* class 4 proton chemical shift violation cutoff threshold *}
{====> nmr.prot.thresh.4=0.3;

{===== diffusion anisotropy restraint data =====}

{* fixed or harmonically restrained external axis *}
+ choice: "fixed" "harm" +}
{====> nmr.dani.axis="harm";

{* Class 1 *}

{* diffusion anisotropy restraints file *}
{====> nmr.dani.file.1="";
* diffusion anisotropy potential *}
+ choice: "harmonic" "square" +}
{====> nmr.dani.pot.1="harmonic";
* diffusion anisotropy initial force value *}
{====> nmr.dani.force.init.1=0.01;
* diffusion anisotropy final force value *}
{====> nmr.dani.force.finl.1=1.0;
* diffusion anisotropy coefficients *}
* coef: <Tc> <anis> <rhomnicity> <wh> <wn> *}

{* Tc = 1/2 (Dx+Dy+Dz) in <ns> *}
{====> nmr.dani.coef.1.1=13.1;
* anis = Dz/0.5*(Dx+Dy) *}

```

```

{==>} nmr.dani.coef.2.1=2.1;
* rhombicity = 1.5*(Dy-Dx)/(Dz-0.5*(Dy+Dx)) *}
{==>} nmr.dani.coef.3.1=0.0;
* wH in <MHz> *}
{==>} nmr.dani.coef.4.1=600.13;
* wN in <MHz> *}
{==>} nmr.dani.coef.5.1=60.82;

{===== susceptibility anisotropy restraint data =====}

{ * fixed or harmonically restrained external axis *}
+ choice: "fixed" "harm" +}
{==>} nmr.sani.axis="harm";

{ * Class 1 *}

{ * susceptibility anisotropy restraints file *}
{==>} nmr.sani.file.1="";
* susceptibility anisotropy potential *}
+ choice: "harmonic" "square" +}
{==>} nmr.sani.pot.1="harmonic";
* susceptibility anisotropy initial force value *}
{==>} nmr.sani.force.init.1=0.01;
* susceptibility anisotropy final force value *}
{==>} nmr.sani.force.finl.1=50.0;
* susceptibility anisotropy coefficients *}
* coef: <DFS> <axial > <rhombicity>;
a0+a1*(3*cos(theta)^2-1)+a2*(3/2)*sin(theta)^2*cos(2*phi) *}

{ * DFS = a0 *}
{==>} nmr.sani.coef.1.1=-0.0601;
* axial = a0-a1-3/2*a2 *}
{==>} nmr.sani.coef.2.1=-8.02;
* rhombicity = a2/a1 *}
{==>} nmr.sani.coef.3.1=0.4;

{===== other restraint data =====}

{ * dihedral angle restraints file *}
{ * Note: the restraint file MUST NOT contain restraints
dihedral or end *}
{==>} nmr.cdih.file="06apr-talos-cns.tbl";

{ * DNA-RNA base planarity restraints file *}
{ * Note: include weights as $pscale in the restraint file *}
{==>} nmr.plan.file="";
* input planarity scale factor - this will be written into $pscale *}
{==>} nmr.plan.scale=150;

{ * NCS-restraints file *}
{ * example is in inputs/xtal_data/egl_ncs_restrain.dat *}
{==>} nmr.ncs.file="";

{===== input/output files =====}

{ * base name for input coordinate files *}
{==>} pdb.in.name="";

{ * base name for output coordinate files *}
{==>} pdb.out.name="str_5g3-52";

```

### Appendix 3.1      Chemical Shifts of CaM-CD2-III-5G

1	54.653	0.013	CA 2
2	41.423	0.000	CB 2
3	175.347	0.000	C 3
4	58.812	0.128	CA 3
5	64.159	0.053	CB 3
6	4.638	0.007	HA 3
7	8.323	0.005	HN 3
8	117.210	0.124	N 3
9	173.653	0.000	C 4
10	45.670	0.034	CA 4
11	4.223	0.009	HA1 4
12	4.046	0.014	HA2 4
13	8.542	0.005	HN 4
14	111.759	0.153	N 4
15	173.441	0.000	C 5
16	62.819	0.123	CA 5
17	70.825	0.038	CB 5
18	4.838	0.007	HA 5
19	3.769	0.008	HB 5
20	0.441	0.008	HG2# 5
21	7.990	0.006	HN 5
22	118.917	0.105	N 5
23	173.211	0.000	C 6
24	61.280	0.079	CA 6
25	34.541	0.012	CB 6
26	4.280	0.010	HA 6
27	1.821	0.008	HB 6
28	0.928	0.004	HG1# 6
29	0.879	0.007	HG2# 6
30	9.019	0.007	HN 6
31	129.014	0.094	N 6
32	176.717	0.000	C 7
33	55.885	0.027	CA 7
34	31.068	0.008	CB 7
35	5.278	0.008	HA 7
36	2.961	0.008	HB# 7
37	7.120	0.004	HD1 7
38	10.129	0.005	HE1 7
39	7.301	0.006	HE3 7
40	7.176	0.006	HH2 7
41	8.482	0.006	HN 7
42	7.436	0.004	HZ2 7
43	7.056	0.003	HZ3 7
44	126.599	0.083	N 7
45	130.014	0.167	NE1 7
46	171.052	0.000	C 8
47	44.090	0.046	CA 8
48	4.621	0.017	HA1 8
49	3.249	0.008	HA2 8
50	8.829	0.006	HN 8
51	109.483	0.173	N 8
52	177.055	0.000	C 9
53	50.474	0.062	CA 9
54	20.616	0.215	CB 9
55	4.851	0.012	HA 9
56	1.129	0.008	HB# 9
57	8.238	0.017	HN 9
58	125.060	0.088	N 9
59	179.290	0.000	C 10
60	56.279	0.171	CA 10
61	42.108	0.014	CB 10
62	3.787	0.008	HA 10
63	1.606	0.006	HB1 10
64	1.560	0.024	HB2 10



65	0.944	0.006	HD1#	10
66	0.797	0.006	HD2#	10
67	1.717	0.006	HG	10
68	8.644	0.007	HN	10
69	123.004	0.109	N	10
70	173.942	0.000	C	11
71	45.762	0.051	CA	11
72	4.050	0.012	HA#	11
73	9.249	0.012	HN	11
74	112.425	0.159	N	11
75	175.477	0.000	C	12
76	54.606	0.051	CA	12
77	32.881	0.120	CB	12
78	4.887	0.021	HA	12
79	3.253	0.000	HB#	12
80	3.333	0.028	HB1	12
81	3.214	0.024	HB2	12
82	7.000	0.000	HD1	12
83	7.820	0.011	HN	12
84	118.587	0.187	N	12
85	174.366	0.000	C	13
86	45.063	0.023	CA	13
87	4.533	0.015	HA1	13
88	3.580	0.012	HA2	13
89	8.591	0.006	HN	13
90	110.003	0.212	N	13
91	171.227	0.000	C	14
92	60.008	0.075	CA	14
93	41.737	0.003	CB	14
94	4.466	0.021	HA	14
95	1.835	0.005	HB	14
96	0.826	0.012	HD1#	14
97	1.244	0.005	HG11	14
98	0.674	0.009	HG2#	14
99	8.512	0.007	HN	14
100	120.650	0.099	N	14
101	174.570	0.000	C	15
102	50.922	0.068	CA	15
103	39.099	0.013	CB	15
104	5.520	0.009	HA	15
105	2.552	0.011	HB1	15
106	2.151	0.010	HB2	15
107	7.363	0.005	HD21	15
108	6.425	0.006	HD22	15
109	7.790	0.008	HN	15
110	123.617	0.102	N	15
111	111.214	0.140	ND2	15
112	174.527	0.000	C	16
113	53.761	0.049	CA	16
114	41.606	0.043	CB	16
115	4.318	0.007	HA	16
116	1.128	0.009	HB1	16
117	1.011	0.008	HB2	16
118	0.446	0.005	HD1#	16
119	0.085	0.009	HD2#	16
120	9.351	0.007	HN	16
121	125.109	0.084	N	16
122	175.068	0.000	C	17
123	51.888	0.031	CA	17
124	40.961	0.024	CB	17
125	4.888	0.007	HA	17
126	2.715	0.008	HB1	17
127	2.515	0.012	HB2	17
128	8.130	0.009	HD21	17
129	6.787	0.005	HD22	17
130	8.147	0.007	HN	17

131	120.546	0.105	N 17
132	112.726	0.137	ND2 17
133	60.810	0.218	CA 18
134	37.392	0.000	CB 18
135	3.959	0.006	HA 18
136	1.920	0.004	HB 18
137	0.694	0.012	HG12 18
138	0.391	0.010	HG2# 18
139	8.427	0.008	HN 18
140	124.524	0.094	N 18
141	176.769	0.000	C 19
142	63.388	0.110	CA 19
143	32.360	0.002	CB 19
144	4.329	0.004	HA 19
145	2.282	0.002	HB# 19
146	3.882	0.461	HD# 19
147	173.639	0.000	C 20
148	54.348	0.076	CA 20
149	38.052	0.018	CB 20
150	4.277	0.015	HA 20
151	2.971	0.010	HB1 20
152	2.700	0.014	HB2 20
153	7.591	0.015	HD21 20
154	6.914	0.005	HD22 20
155	8.863	0.007	HN 20
156	117.765	0.119	N 20
157	113.503	0.149	ND2 20
158	173.713	0.000	C 21
159	57.637	0.121	CA 21
160	43.394	0.046	CB 21
161	4.739	0.016	HA 21
162	2.774	0.012	HB1 21
163	2.450	0.010	HB2 21
164	7.198	0.006	HD# 21
165	7.087	0.006	HE# 21
166	7.603	0.006	HN 21
167	119.059	0.112	N 21
168	173.004	0.000	C 22
169	53.763	0.038	CA 22
170	31.313	0.029	CB 22
171	4.213	0.007	HA 22
172	1.688	0.008	HB# 22
173	1.711	0.004	HB1 22
174	1.629	0.011	HB2 22
175	7.408	0.006	HE21 22
176	6.734	0.004	HE22 22
177	2.105	0.009	HG# 22
178	7.035	0.006	HN 22
179	127.084	0.099	N 22
180	113.076	0.122	NE2 22
181	175.725	0.000	C 23
182	55.631	0.080	CA 23
183	31.143	0.006	CB 23
184	4.087	0.010	HA 23
185	1.969	0.010	HB1 23
186	1.901	0.000	HB2 23
187	2.802	0.000	HG# 23
188	8.417	0.007	HN 23
189	124.521	0.058	N 23
190	175.675	0.000	C 24
191	60.125	0.060	CA 24
192	72.191	0.031	CB 24
193	4.681	0.013	HA 24
194	2.914	0.000	HB 24
195	2.726	0.000	HG2# 24
196	6.814	0.009	HN 24

197	117.028	0.113	N 24
198	175.886	0.000	C 25
199	56.538	0.113	CA 25
200	40.666	0.000	CB 25
201	4.479	0.013	HA 25
202	8.675	0.005	HN 25
203	119.490	0.117	N 25
204	175.265	0.000	C 26
205	54.172	0.032	CA 26
206	40.945	0.043	CB 26
207	4.680	0.011	HA 26
208	2.643	0.005	HB# 26
209	2.802	0.000	HB1 26
210	2.598	0.009	HB2 26
211	8.194	0.010	HN 26
212	116.658	0.141	N 26
213	176.097	0.000	C 27
214	59.741	0.103	CA 27
215	35.849	0.040	CB 27
216	4.136	0.010	HA 27
217	2.581	0.011	HB 27
218	0.852	0.007	HD1# 27
219	1.979	0.011	HG11 27
220	1.711	0.009	HG12 27
221	0.986	0.006	HG2# 27
222	7.434	0.008	HN 27
223	120.239	0.102	N 27
224	175.334	0.000	C 28
225	55.173	0.191	CA 28
226	45.800	0.000	CB 28
227	4.968	0.012	HA 28
228	2.412	0.009	HB1 28
229	2.295	0.007	HB2 28
230	8.870	0.009	HN 28
231	128.991	0.200	N 28
232	175.033	0.000	C 29
233	54.894	0.044	CA 29
234	34.297	0.094	CB 29
235	5.607	0.007	HA 29
236	2.025	0.000	HB# 29
237	2.060	0.010	HB1 29
238	1.941	0.007	HB2 29
239	7.713	0.008	HN 29
240	117.446	0.057	N 29
241	174.226	0.000	C 30
242	61.908	0.133	CA 30
243	35.932	0.065	CB 30
244	4.774	0.012	HA 30
245	2.102	0.013	HB 30
246	0.965	0.007	HG1# 30
247	0.881	0.009	HG2# 30
248	8.929	0.010	HN 30
249	125.225	0.117	N 30
250	174.413	0.000	C 31
251	55.505	0.091	CA 31
252	34.045	0.000	CB 31
253	5.052	0.012	HA 31
254	1.894	0.000	HB# 31
255	1.891	0.008	HB1 31
256	1.827	0.004	HB2 31
257	9.136	0.007	HN 31
258	127.310	0.105	N 31
259	175.878	0.000	C 32
260	56.783	0.145	CA 32
261	32.489	0.039	CB 32
262	5.427	0.013	HA 32

263	3.007	0.015	HB# 32
264	6.799	0.004	HD1 32
265	10.317	0.005	HE1 32
266	7.280	0.004	HE3 32
267	6.627	0.003	HH2 32
268	8.972	0.009	HN 32
269	7.334	0.004	HZ2 32
270	6.707	0.003	HZ3 32
271	125.128	0.090	N 32
272	129.742	0.213	NE1 32
273	174.964	0.000	C 33
274	54.505	0.023	CA 33
275	35.578	0.035	CB 33
276	5.120	0.009	HA 33
277	1.860	0.009	HB# 33
278	9.519	0.011	HN 33
279	123.730	0.104	N 33
280	176.992	0.000	C 34
281	55.186	0.179	CA 34
282	31.187	0.021	CB 34
283	4.366	0.007	HA 34
284	1.451	0.013	HB# 34
285	2.747	0.004	HD# 34
286	8.717	0.008	HN 34
287	123.863	0.103	N 34
288	47.494	0.058	CA 35
289	3.994	0.010	HA1 35
290	3.590	0.008	HA2 35
291	8.897	0.005	HN 35
292	119.715	0.113	N 35
293	174.201	0.000	C 36
294	58.182	0.038	CA 36
295	63.486	0.096	CB 36
296	4.513	0.008	HA 36
297	8.786	0.005	HN 36
298	122.042	0.147	N 36
299	173.375	0.000	C 37
300	62.290	0.144	CA 37
301	70.315	0.040	CB 37
302	4.310	0.006	HA 37
303	4.178	0.018	HB 37
304	1.113	0.002	HG2# 37
305	8.165	0.006	HN 37
306	120.751	0.163	N 37
307	176.442	0.000	C 38
308	56.528	0.126	CA 38
309	42.071	0.036	CB 38
310	4.260	0.041	HA 38
311	8.781	0.005	HN 38
312	130.845	0.107	N 38
313	175.671	0.000	C 39
314	62.636	0.077	CA 39
315	32.471	0.012	CB 39
316	4.123	0.011	HA 39
317	1.373	0.009	HB 39
318	0.598	0.007	HG1# 39
319	0.393	0.007	HG2# 39
320	8.897	0.010	HN 39
321	125.197	0.102	N 39
322	174.261	0.000	C 40
323	52.913	0.148	CA 40
324	22.661	0.016	CB 40
325	4.596	0.010	HA 40
326	1.310	0.013	HB# 40
327	7.783	0.006	HN 40
328	119.659	0.148	N 40

329	174.315	0.000	C 41
330	54.938	0.128	CA 41
331	33.791	0.045	CB 41
332	5.438	0.014	HA 41
333	2.198	0.037	HB# 41
334	8.774	0.007	HN 41
335	119.099	0.148	N 41
336	173.479	0.000	C 42
337	56.285	0.191	CA 42
338	42.652	0.039	CB 42
339	5.026	0.012	HA 42
340	3.247	0.007	HB1 42
341	2.912	0.012	HB2 42
342	7.094	0.005	HD# 42
343	9.308	0.007	HN 42
344	125.581	0.112	N 42
345	175.667	0.000	C 43
346	54.251	0.043	CA 43
347	34.933	0.047	CB 43
348	4.744	0.002	HA 43
349	8.004	0.010	HN 43
350	128.166	0.136	N 43
351	177.415	0.000	C 44
352	58.688	0.007	CA 44
353	30.607	0.000	CB 44
354	3.715	0.010	HA 44
355	8.350	0.009	HN 44
356	126.593	0.118	N 44
357	176.732	0.000	C 45
358	56.749	0.003	CA 45
359	32.627	0.011	CB 45
360	4.252	0.010	HA 45
361	8.584	0.008	HN 45
362	119.902	0.162	N 45
363	176.271	0.000	C 46
364	55.416	0.028	CA 46
365	32.963	0.011	CB 46
366	4.469	0.009	HA 46
367	7.950	0.009	HN 46
368	120.466	0.216	N 46
369	8.136	0.006	HN 47
370	123.877	0.132	N 47
371	176.853	0.000	C 48
372	63.616	0.124	CA 48
373	32.072	0.000	CB 48
374	4.379	0.000	HA 48
375	175.854	0.000	C 49
376	57.905	0.134	CA 49
377	39.313	0.053	CB 49
378	4.614	0.010	HA 49
379	3.122	0.010	HB1 49
380	3.078	0.025	HB2 49
381	7.239	0.004	HD# 49
382	8.121	0.008	HN 49
383	120.163	0.108	N 49
384	177.099	0.000	C 50
385	55.024	0.035	CA 50
386	42.739	0.010	CB 50
387	4.353	0.005	HA 50
388	1.574	0.001	HB# 50
389	1.591	0.000	HB1 50
390	1.531	0.002	HB2 50
391	8.008	0.006	HN 50
392	124.804	0.102	N 50
393	56.522	0.000	CA 51
394	33.125	0.000	CB 51

395	4.795	0.000	HA 51
396	8.291	0.005	HN 51
397	123.449	0.130	N 51
398	45.521	0.000	CA 54
399	45.327	0.027	CA 55
400	4.014	0.000	HA# 55
401	8.486	0.012	HN 55
402	109.779	0.219	N 55
403	176.935	0.000	C 56
404	54.523	0.002	CA 56
405	41.347	0.006	CB 56
406	4.579	0.012	HA 56
407	2.668	0.009	HB# 56
408	8.310	0.003	HN 56
409	121.208	0.072	N 56
410	176.764	0.000	C 57
411	56.720	0.027	CA 57
412	32.888	0.007	CB 57
413	4.289	0.005	HA 57
414	1.802	0.008	HB# 57
415	1.397	0.007	HG# 57
416	8.309	0.004	HN 57
417	121.898	0.106	N 57
418	177.028	0.000	C 58
419	54.574	0.033	CA 58
420	41.444	0.003	CB 58
421	4.630	0.010	HA 58
422	2.692	0.007	HB# 58
423	8.225	0.004	HN 58
424	121.279	0.103	N 58
425	174.650	0.000	C 59
426	45.806	0.042	CA 59
427	3.963	0.004	HA2 59
428	8.254	0.009	HN 59
429	109.807	0.175	N 59
430	176.025	0.000	C 60
431	53.484	0.024	CA 60
432	39.153	0.015	CB 60
433	4.775	0.006	HA 60
434	2.783	0.009	HB# 60
435	7.610	0.005	HD21 60
436	6.914	0.003	HD22 60
437	8.366	0.004	HN 60
438	119.501	0.111	N 60
439	113.900	0.172	ND2 60
440	173.956	0.000	C 61
441	45.534	0.012	CA 61
442	3.884	0.008	HA# 61
443	8.419	0.004	HN 61
444	109.904	0.169	N 61
445	175.795	0.000	C 62
446	58.153	0.139	CA 62
447	38.990	0.000	CB 62
448	4.591	0.009	HA 62
449	2.963	0.007	HB# 62
450	7.059	0.005	HD# 62
451	6.797	0.000	HE# 62
452	7.999	0.008	HN 62
453	121.087	0.090	N 62
454	175.976	0.000	C 63
455	60.916	0.105	CA 63
456	39.067	0.081	CB 63
457	4.165	0.000	HA 63
458	1.780	0.012	HB 63
459	0.859	0.006	HG2# 63
460	8.056	0.005	HN 63

461	124.421	0.094	N 63
462	174.676	0.000	C 64
463	58.322	0.136	CA 64
464	64.079	0.035	CB 64
465	3.925	0.012	HA 64
466	8.317	0.003	HN 64
467	121.105	0.097	N 64
468	52.820	0.073	CA 65
469	19.341	0.000	CB 65
470	4.338	0.011	HA 65
471	1.422	0.001	HB# 65
472	8.418	0.021	HN 65
473	127.474	0.139	N 65
474	178.228	0.000	C 66
475	52.770	0.017	CA 66
476	19.339	0.000	CB 66
477	1.401	0.006	HB# 66
478	177.432	0.000	C 67
479	56.832	0.051	CA 67
480	30.385	0.043	CB 67
481	4.319	0.000	HA 67
482	2.113	0.016	HB# 67
483	2.002	0.000	HB3 67
484	2.295	0.003	HG# 67
485	8.317	0.024	HN 67
486	120.945	0.082	N 67
487	45.602	0.000	CA 68
488	4.794	0.002	HA1 68
489	4.026	0.007	HA2 68
490	8.455	0.007	HN 68
491	110.890	0.189	N 68
492	45.176	0.000	CA 70
493	4.347	0.000	HA# 70
494	8.479	0.000	HN 70
495	177.293	0.000	C 71
496	53.601	0.786	CA 71
497	19.465	0.100	CB 71
498	3.943	0.017	HA 71
499	1.144	0.009	HB# 71
500	8.255	0.012	HN 71
501	123.929	0.123	N 71
502	175.373	0.000	C 72
503	55.956	0.117	CA 72
504	42.655	0.002	CB 72
505	5.253	0.010	HA 72
506	2.873	0.007	HB1 72
507	2.716	0.009	HB2 72
508	7.121	0.005	HD# 72
509	7.327	0.004	HE# 72
510	7.864	0.007	HN 72
511	115.556	0.138	N 72
512	174.007	0.000	C 73
513	55.167	0.151	CA 73
514	34.379	0.024	CB 73
515	4.633	0.011	HA 73
516	1.866	0.001	HB1 73
517	1.773	0.013	HB2 73
518	8.804	0.008	HN 73
519	122.724	0.105	N 73
520	176.173	0.000	C 74
521	60.121	0.077	CA 74
522	38.861	0.060	CB 74
523	4.832	0.010	HA 74
524	1.618	0.012	HB 74
525	1.030	0.007	HG2# 74
526	8.667	0.011	HN 74

527	125.202	0.098	N 74
528	178.838	0.000	C 75
529	54.207	0.100	CA 75
530	42.116	0.000	CB 75
531	4.452	0.009	HA 75
532	1.951	0.011	HB1 75
533	1.620	0.010	HB2 75
534	0.716	0.005	HD1# 75
535	1.532	0.006	HG 75
536	9.026	0.008	HN 75
537	129.727	0.185	N 75
538	177.806	0.000	C 76
539	54.919	0.017	CA 76
540	18.437	0.000	CB 76
541	3.829	0.011	HA 76
542	1.350	0.000	HB# 76
543	8.322	0.004	HN 76
544	122.316	0.124	N 76
545	176.153	0.000	C 77
546	52.375	0.016	CA 77
547	37.863	0.000	CB 77
548	4.521	0.007	HA 77
549	2.746	0.000	HB# 77
550	3.338	0.013	HB1 77
551	2.756	0.005	HB2 77
552	7.234	0.005	HD21 77
553	6.492	0.011	HD22 77
554	7.557	0.010	HN 77
555	112.600	0.118	N 77
556	109.186	0.145	ND2 77
557	171.862	0.000	C 78
558	44.940	0.023	CA 78
559	4.324	0.017	HA1 78
560	3.228	0.012	HA2 78
561	8.307	0.010	HN 78
562	109.835	0.137	N 78
563	174.468	0.000	C 79
564	55.248	0.104	CA 79
565	41.710	0.000	CB 79
566	4.700	0.013	HA 79
567	2.693	0.001	HB1 79
568	2.317	0.003	HB2 79
569	7.756	0.006	HN 79
570	120.685	0.091	N 79
571	174.479	0.000	C 80
572	53.268	0.020	CA 80
573	41.754	0.020	CB 80
574	4.227	0.011	HA 80
575	0.295	0.013	HB1 80
576	-1.352	0.007	HB2 80
577	0.338	0.007	HD1# 80
578	0.144	0.004	HD2# 80
579	0.804	0.007	HG 80
580	7.557	0.008	HN 80
581	120.826	0.095	N 80
582	175.703	0.000	C 81
583	55.072	0.185	CA 81
584	34.503	0.011	CB 81
585	4.837	0.008	HA 81
586	9.186	0.008	HN 81
587	129.264	0.106	N 81
588	176.422	0.000	C 82
589	60.880	0.093	CA 82
590	38.099	0.000	CB 82
591	4.201	0.009	HA 82
592	1.717	0.008	HB 82



593	0.455	0.005	HD1# 82
594	0.801	0.005	HG1# 82
595	0.527	0.005	HG2# 82
596	8.744	0.006	HN 82
597	129.675	0.098	N 82
598	176.577	0.000	C 83
599	60.091	0.132	CA 83
600	33.149	0.062	CB 83
601	3.708	0.013	HA 83
602	1.660	0.012	HB# 83
603	1.219	0.011	HG# 83
604	7.727	0.011	HN 83
605	127.425	0.099	N 83
606	172.735	0.000	C 84
607	53.258	0.021	CA 84
608	40.277	0.000	CB 84
609	4.029	0.008	HA 84
610	2.548	0.011	HB1 84
611	2.263	0.010	HB2 84
612	7.230	0.006	HD21 84
613	6.532	0.006	HD22 84
614	8.508	0.010	HN 84
615	117.154	0.141	N 84
616	111.779	0.143	ND2 84
617	177.378	0.000	C 85
618	56.843	0.173	CA 85
619	42.845	0.027	CB 85
620	4.156	0.012	HA 85
621	1.720	0.000	HB# 85
622	1.756	0.016	HB1 85
623	1.664	0.010	HB2 85
624	0.718	0.002	HD2# 85
625	1.564	0.002	HG 85
626	8.031	0.020	HN 85
627	125.289	0.122	N 85
628	175.774	0.000	C 86
629	60.182	0.053	CA 86
630	72.927	0.062	CB 86
631	4.779	0.013	HA 86
632	4.459	0.005	HB 86
633	1.259	0.013	HG2# 86
634	9.407	0.009	HN 86
635	119.356	0.116	N 86
636	178.612	0.000	C 87
637	59.975	0.076	CA 87
638	29.595	0.048	CB 87
639	3.863	0.011	HA 87
640	1.843	0.010	HB# 87
641	3.264	0.002	HD# 87
642	1.671	0.004	HG# 87
643	8.946	0.006	HN 87
644	120.646	0.104	N 87
645	176.195	0.000	C 88
646	55.738	0.034	CA 88
647	40.961	0.039	CB 88
648	4.537	0.008	HA 88
649	2.774	0.010	HB1 88
650	2.422	0.011	HB2 88
651	7.765	0.007	HN 88
652	117.272	0.114	N 88
653	177.496	0.000	C 89
654	55.370	0.084	CA 89
655	42.709	0.018	CB 89
656	4.813	0.008	HA 89
657	2.895	0.015	HB1 89
658	2.808	0.060	HB2 89

659	7.595	0.007	HN 89
660	116.656	0.121	N 89
661	173.402	0.000	C 90
662	59.012	0.091	CA 90
663	64.069	0.032	CB 90
664	4.199	0.009	HA 90
665	4.130	0.037	HB1 90
666	3.882	0.014	HB2 90
667	7.357	0.008	HN 90
668	115.123	0.129	N 90
669	172.354	0.000	C 91
670	44.560	0.024	CA 91
671	4.440	0.016	HA1 91
672	3.915	0.014	HA2 91
673	8.727	0.008	HN 91
674	111.260	0.143	N 91
675	174.810	0.000	C 92
676	63.197	0.125	CA 92
677	69.987	0.040	CB 92
678	5.122	0.010	HA 92
679	3.845	0.016	HB 92
680	1.153	0.011	HG2# 92
681	8.359	0.006	HN 92
682	118.203	0.108	N 92
683	174.592	0.000	C 93
684	57.238	0.144	CA 93
685	41.028	0.094	CB 93
686	5.212	0.010	HA 93
687	3.117	0.010	HB1 93
688	2.999	0.010	HB2 93
689	6.995	0.004	HD# 93
690	6.630	0.007	HE# 93
691	9.900	0.016	HH 93
692	9.937	0.006	HN 93
693	130.628	0.106	N 93
694	173.889	0.000	C 94
695	52.573	0.099	CA 94
696	43.098	0.006	CB 94
697	5.602	0.008	HA 94
698	2.516	0.009	HB# 94
699	7.230	0.006	HD21 94
700	6.821	0.005	HD22 94
701	9.542	0.006	HN 94
702	121.251	0.102	N 94
703	113.968	0.133	ND2 94
704	173.258	0.000	C 95
705	57.525	0.168	CA 95
706	32.185	0.000	CB 95
707	4.939	0.007	HA 95
708	-0.609	0.013	HB 95
709	0.216	0.004	HG1# 95
710	-0.389	0.008	HG2# 95
711	9.095	0.005	HN 95
712	126.281	0.095	N 95
713	4.872	0.024	HA 96
714	3.625	0.009	HB 96
715	0.943	0.005	HG2# 96
716	8.359	0.007	HN 96
717	121.795	0.098	N 96
718	174.813	0.000	C 97
719	60.504	0.085	CA 97
720	35.368	0.000	CB 97
721	4.943	0.005	HA 97
722	1.672	0.083	HB 97
723	1.082	0.005	HG1# 97
724	0.419	0.005	HG2# 97

725	7.737	0.014	HN 97
726	123.750	0.199	N 97
727	177.058	0.000	C 98
728	56.569	0.140	CA 98
729	42.428	0.013	CB 98
730	5.559	0.013	HA 98
731	2.999	0.011	HB1 98
732	2.873	0.006	HB2 98
733	6.818	0.005	HD# 98
734	6.624	0.002	HE# 98
735	9.220	0.007	HN 98
736	125.694	0.092	N 98
737	176.861	0.000	C 99
738	57.132	0.127	CA 99
739	65.506	0.010	CB 99
740	5.337	0.010	HA 99
741	9.215	0.007	HN 99
742	119.000	0.144	N 99
743	65.179	0.226	CA 100
744	69.040	0.000	CB 100
745	3.992	0.008	HA 100
746	4.286	0.006	HB 100
747	1.280	0.003	HG2# 100
748	8.697	0.008	HN 100
749	116.810	0.138	N 100
750	176.425	0.000	C 101
751	52.779	0.030	CA 101
752	37.779	0.082	CB 101
753	4.796	0.003	HA 101
754	3.173	0.008	HB1 101
755	2.892	0.004	HB2 101
756	7.618	0.011	HD21 101
757	6.895	0.010	HD22 101
758	8.320	0.000	HN 101
759	112.389	0.122	ND2 101
760	174.770	0.000	C 102
761	45.459	0.035	CA 102
762	4.299	0.011	HA1 102
763	3.769	0.010	HA2 102
764	8.286	0.010	HN 102
765	108.680	0.187	N 102
766	173.417	0.000	C 103
767	63.462	0.107	CA 103
768	69.600	0.048	CB 103
769	4.170	0.005	HA 103
770	1.127	0.006	HG2# 103
771	7.708	0.008	HN 103
772	117.491	0.146	N 103
773	175.904	0.000	C 104
774	56.296	0.000	CA 104
775	30.209	0.000	CB 104
776	4.097	0.007	HA 104
777	8.938	0.010	HN 104
778	128.739	0.186	N 104
779	61.545	0.000	CA 105
780	4.178	0.019	HA 105
781	1.795	0.006	HB 105
782	0.930	0.003	HG2# 105
783	8.634	0.008	HN 105
784	127.551	0.116	N 105
785	174.444	0.000	C 106
786	4.453	0.007	HA 106
787	0.540	0.008	HD2# 106
788	7.213	0.006	HN 106
789	173.632	0.000	C 107
790	53.128	0.049	CA 107

791	41.060	0.086	CB 107
792	5.418	0.006	HA 107
793	2.528	0.000	HB# 107
794	2.536	0.009	HB1 107
795	2.435	0.007	HB2 107
796	7.083	0.146	HD21 107
797	7.169	0.001	HD22 107
798	8.346	0.008	HN 107
799	122.757	0.141	N 107
800	112.712	0.111	ND2 107
801	173.263	0.000	C 108
802	54.270	0.030	CA 108
803	37.088	0.001	CB 108
804	4.707	0.018	HA 108
805	1.673	0.011	HB1 108
806	1.356	0.006	HB2 108
807	1.817	0.006	HD1 108
808	1.647	0.009	HD2 108
809	3.040	0.005	HE1 108
810	2.937	0.002	HE2 108
811	1.286	0.008	HG1 108
812	8.564	0.009	HN 108
813	124.789	0.088	N 108
814	177.250	0.000	C 109
815	50.619	0.068	CA 109
816	21.878	0.039	CB 109
817	5.540	0.013	HA 109
818	1.264	0.008	HB# 109
819	8.211	0.006	HN 109
820	125.124	0.104	N 109
821	174.707	0.000	C 110
822	54.155	0.022	CA 110
823	46.108	0.030	CB 110
824	4.845	0.008	HA 110
825	1.719	0.011	HB1 110
826	1.547	0.009	HB2 110
827	1.002	0.004	HD2# 110
828	9.414	0.007	HN 110
829	124.166	0.084	N 110
830	173.684	0.000	C 111
831	53.635	0.090	CA 111
832	42.487	0.014	CB 111
833	5.130	0.007	HA 111
834	2.522	0.003	HB1 111
835	2.480	0.020	HB2 111
836	8.656	0.006	HN 111
837	124.024	0.095	N 111
838	175.343	0.000	C 112
839	53.996	0.065	CA 112
840	43.813	0.074	CB 112
841	5.006	0.011	HA 112
842	2.291	0.007	HB1 112
843	1.093	0.010	HB2 112
844	0.669	0.007	HD1# 112
845	1.232	0.008	HG 112
846	8.958	0.007	HN 112
847	127.853	0.093	N 112
848	174.169	0.000	C 113
849	53.542	0.084	CA 113
850	32.430	0.010	CB 113
851	4.486	0.010	HA 113
852	1.049	0.001	HB1 113
853	0.112	0.010	HB2 113
854	2.767	0.006	HD1 113
855	2.664	0.007	HD2 113
856	0.919	0.000	HG# 113

857	1.107	0.000	HG1 113
858	0.996	0.000	HG2 113
859	8.866	0.006	HN 113
860	129.170	0.131	N 113
861	176.931	0.000	C 114
862	58.921	0.108	CA 114
863	38.130	0.043	CB 114
864	4.686	0.008	HA 114
865	1.934	0.008	HB 114
866	0.687	0.008	HD1# 114
867	1.278	0.007	HG11 114
868	0.862	0.007	HG2# 114
869	8.362	0.005	HN 114
870	120.835	0.110	N 114
871	175.671	0.000	C 115
872	55.386	0.126	CA 115
873	43.056	0.005	CB 115
874	4.488	0.010	HA 115
875	1.652	0.010	HB1 115
876	1.587	0.031	HB2 115
877	0.743	0.004	HD1# 115
878	8.821	0.007	HN 115
879	130.659	0.089	N 115
880	58.148	0.214	CA 116
881	31.341	0.000	CB 116
882	4.216	0.006	HA 116
883	2.085	0.009	HB1 116
884	1.921	0.009	HB2 116
885	2.253	0.010	HG# 116
886	8.091	0.005	HN 116
887	129.997	0.075	N 116

## Appendix 3.2 Assigned chemical Shifts of CaM-CD2-IV-5G

1	4.616	0.000	HA 3
2	8.316	0.000	HN 3
3	116.971	0.000	N 3
4	4.241	0.012	HA1 4
5	4.030	0.013	HA2 4
6	8.534	0.004	HN 4
7	111.355	0.053	N 4
8	4.844	0.027	HA 5
9	3.776	0.008	HB 5
10	0.435	0.014	HG2# 5
11	7.982	0.002	HN 5
12	118.665	0.023	N 5
13	4.300	0.015	HA 6
14	1.835	0.020	HB 6
15	0.922	0.016	HG1# 6
16	0.878	0.014	HG2# 6
17	9.011	0.003	HN 6
18	128.711	0.062	N 6
19	5.295	0.005	HA 7
20	2.962	0.013	HB# 7
21	10.132	0.014	HE1 7
22	7.308	0.017	HE3 7
23	7.167	0.000	HH2 7
24	8.471	0.003	HN 7
25	7.457	0.001	HZ2 7
26	7.045	0.000	HZ3 7
27	126.384	0.014	N 7
28	129.892	0.000	NE1 7
29	4.639	0.014	HA1 8
30	3.258	0.009	HA2 8
31	8.822	0.005	HN 8
32	109.234	0.039	N 8
33	4.866	0.005	HA 9
34	1.131	0.009	HB# 9
35	8.229	0.004	HN 9
36	124.780	0.027	N 9
37	3.808	0.008	HA 10
38	1.595	0.000	HB1 10
39	1.562	0.016	HB2 10
40	0.907	0.071	HD1# 10
41	0.797	0.007	HD2# 10
42	1.715	0.004	HG 10
43	8.632	0.003	HN 10
44	122.776	0.033	N 10
45	4.060	0.004	HA# 11
46	9.234	0.005	HN 11
47	112.253	0.010	N 11
48	4.910	0.007	HA 12
49	3.307	0.056	HB1 12
50	3.218	0.006	HB2 12
51	7.810	0.003	HN 12
52	118.086	0.023	N 12
53	4.530	0.009	HA1 13
54	3.582	0.007	HA2 13
55	8.586	0.003	HN 13
56	109.921	0.018	N 13
57	4.492	0.021	HA 14
58	1.857	0.011	HB 14
59	0.839	0.010	HD1# 14
60	0.663	0.008	HG2# 14
61	8.504	0.003	HN 14
62	120.307	0.013	N 14
63	5.539	0.004	HA 15
64	2.545	0.012	HB1 15

65	2.141	0.013	HB2 15
66	7.363	0.012	HD21 15
67	6.431	0.013	HD22 15
68	7.779	0.007	HN 15
69	123.365	0.010	N 15
70	110.947	0.081	ND2 15
71	4.328	0.011	HA 16
72	1.132	0.014	HB1 16
73	1.052	0.069	HB2 16
74	0.435	0.013	HD1# 16
75	0.075	0.009	HD2# 16
76	9.343	0.004	HN 16
77	124.883	0.022	N 16
78	4.908	0.007	HA 17
79	2.718	0.006	HB1 17
80	2.510	0.012	HB2 17
81	8.143	0.012	HD21 17
82	6.793	0.013	HD22 17
83	8.132	0.007	HN 17
84	120.273	0.139	N 17
85	112.572	0.030	ND2 17
86	3.978	0.003	HA 18
87	1.930	0.013	HB 18
88	0.687	0.000	HG12 18
89	0.382	0.012	HG2# 18
90	8.421	0.003	HN 18
91	124.241	0.070	N 18
92	4.301	0.011	HA 20
93	2.981	0.011	HB1 20
94	2.686	0.013	HB2 20
95	7.583	0.015	HD21 20
96	6.915	0.002	HD22 20
97	8.853	0.002	HN 20
98	117.586	0.026	N 20
99	113.381	0.102	ND2 20
100	4.762	0.015	HA 21
101	2.774	0.007	HB1 21
102	2.456	0.009	HB2 21
103	7.234	0.000	HD# 21
104	7.596	0.007	HN 21
105	118.828	0.017	N 21
106	4.231	0.002	HA 22
107	1.654	0.000	HB# 22
108	1.703	0.019	HB1 22
109	1.639	0.017	HB2 22
110	7.423	0.013	HE21 22
111	6.732	0.003	HE22 22
112	2.114	0.013	HG# 22
113	7.033	0.003	HN 22
114	126.838	0.028	N 22
115	112.898	0.000	NE2 22
116	4.106	0.029	HA 23
117	1.966	0.020	HB1 23
118	1.928	0.000	HB2 23
119	2.857	0.000	HG# 23
120	8.407	0.002	HN 23
121	124.276	0.006	N 23
122	4.686	0.003	HA 24
123	2.934	0.000	HB 24
124	2.753	0.000	HG2# 24
125	6.810	0.005	HN 24
126	116.743	0.018	N 24
127	4.481	0.007	HA 25
128	8.663	0.002	HN 25
129	119.279	0.013	N 25
130	4.687	0.003	HA 26

131	2.618	0.000	HB# 26
132	2.854	0.000	HB1 26
133	2.615	0.011	HB2 26
134	8.196	0.007	HN 26
135	116.406	0.012	N 26
136	4.146	0.008	HA 27
137	2.587	0.014	HB 27
138	0.846	0.001	HD1# 27
139	1.966	0.007	HG11 27
140	1.695	0.000	HG12 27
141	0.983	0.015	HG2# 27
142	7.429	0.004	HN 27
143	120.023	0.020	N 27
144	4.983	0.010	HA 28
145	2.413	0.014	HB1 28
146	2.298	0.015	HB2 28
147	8.856	0.005	HN 28
148	128.683	0.084	N 28
149	5.617	0.008	HA 29
150	2.059	0.013	HB1 29
151	1.941	0.015	HB2 29
152	7.698	0.003	HN 29
153	117.113	0.020	N 29
154	4.796	0.023	HA 30
155	2.117	0.024	HB 30
156	0.958	0.015	HG1# 30
157	0.894	0.011	HG2# 30
158	8.920	0.005	HN 30
159	125.010	0.052	N 30
160	5.085	0.012	HA 31
161	1.854	0.000	HB# 31
162	9.131	0.004	HN 31
163	127.052	0.055	N 31
164	5.443	0.001	HA 32
165	2.983	0.003	HB# 32
166	6.817	0.000	HD1 32
167	10.301	0.014	HE1 32
168	7.303	0.000	HE3 32
169	6.615	0.003	HH2 32
170	8.974	0.004	HN 32
171	7.363	0.000	HZ2 32
172	6.710	0.019	HZ3 32
173	124.917	0.020	N 32
174	129.571	0.000	NE1 32
175	5.121	0.008	HA 33
176	1.863	0.023	HB# 33
177	9.510	0.003	HN 33
178	123.483	0.055	N 33
179	4.366	0.004	HA 34
180	1.426	0.009	HB# 34
181	8.714	0.003	HN 34
182	123.620	0.015	N 34
183	3.998	0.007	HA1 35
184	3.593	0.004	HA2 35
185	8.887	0.004	HN 35
186	119.515	0.082	N 35
187	4.520	0.005	HA 36
188	8.777	0.005	HN 36
189	121.848	0.046	N 36
190	4.342	0.019	HA 37
191	4.191	0.005	HB 37
192	1.117	0.017	HG2# 37
193	8.170	0.015	HN 37
194	120.549	0.043	N 37
195	4.253	0.042	HA 38
196	8.776	0.002	HN 38



197	130.619	0.002	N 38
198	4.136	0.004	HA 39
199	1.382	0.019	HB 39
200	0.606	0.013	HG1# 39
201	0.387	0.012	HG2# 39
202	8.883	0.007	HN 39
203	124.986	0.010	N 39
204	4.601	0.005	HA 40
205	1.310	0.014	HB# 40
206	7.778	0.004	HN 40
207	119.443	0.027	N 40
208	5.447	0.002	HA 41
209	2.170	0.039	HB# 41
210	8.767	0.003	HN 41
211	119.110	0.038	N 41
212	5.044	0.018	HA 42
213	3.268	0.019	HB1 42
214	2.919	0.012	HB2 42
215	7.134	0.003	HD# 42
216	9.306	0.006	HN 42
217	125.552	0.027	N 42
218	4.735	0.000	HA 43
219	7.988	0.004	HN 43
220	128.080	0.039	N 43
221	3.714	0.005	HA 44
222	8.315	0.004	HN 44
223	126.295	0.031	N 44
224	4.253	0.003	HA 45
225	8.586	0.004	HN 45
226	120.198	0.024	N 45
227	4.470	0.013	HA 46
228	7.968	0.024	HN 46
229	120.328	0.099	N 46
230	4.559	0.000	HA 47
231	8.125	0.003	HN 47
232	123.564	0.003	N 47
233	4.338	0.000	HA 48
234	4.631	0.007	HA 49
235	3.110	0.026	HB1 49
236	3.081	0.026	HB2 49
237	7.267	0.000	HD# 49
238	8.124	0.004	HN 49
239	119.974	0.110	N 49
240	4.332	0.049	HA 50
241	1.557	0.004	HB# 50
242	1.598	0.000	HB1 50
243	1.555	0.000	HB2 50
244	8.012	0.005	HN 50
245	124.535	0.023	N 50
246	4.780	0.000	HA 51
247	8.302	0.003	HN 51
248	123.331	0.073	N 51
249	8.524	0.000	HN 54
250	109.668	0.000	N 54
251	3.974	0.000	HA# 55
252	8.519	0.000	HN 55
253	109.681	0.000	N 55
254	4.631	0.000	HA 56
255	8.291	0.013	HN 56
256	121.651	0.001	N 56
257	4.184	0.002	HA 57
258	1.869	0.011	HB 57
259	8.107	0.003	HN 57
260	120.895	0.027	N 57
261	4.628	0.001	HA 58
262	2.657	0.000	HB# 58

263	8.343	0.003	HN 58
264	124.417	0.008	N 58
265	3.956	0.001	HA2 59
266	8.238	0.003	HN 59
267	110.040	0.016	N 59
268	4.627	0.002	HA 60
269	2.698	0.000	HB# 60
270	8.279	0.005	HN 60
271	121.160	0.020	N 60
272	3.929	0.006	HA# 61
273	8.382	0.003	HN 61
274	109.784	0.004	N 61
275	4.356	0.000	HA 62
276	7.549	0.000	HE21 62
277	6.823	0.000	HE22 62
278	113.315	0.013	NE2 62
279	4.076	0.003	HA 63
280	1.942	0.000	HB 63
281	0.728	0.010	HG1# 63
282	8.141	0.004	HN 63
283	121.962	0.025	N 63
284	4.254	0.000	HA 64
285	2.678	0.000	HB# 64
286	7.581	0.000	HD21 64
287	6.920	0.000	HD22 64
288	8.419	0.004	HN 64
289	123.287	0.025	N 64
290	113.672	0.018	ND2 64
291	4.559	0.002	HA 65
292	3.083	0.004	HB1 65
293	2.912	0.004	HB2 65
294	7.140	0.003	HD# 65
295	8.260	0.003	HN 65
296	122.667	0.023	N 65
297	1.902	0.000	HB# 66
298	2.213	0.000	HG# 66
299	8.298	0.002	HN 66
300	123.394	0.072	N 66
301	2.269	0.000	HG# 67
302	8.356	0.003	HN 67
303	123.099	0.036	N 67
304	3.938	0.019	HA 71
305	1.149	0.001	HB# 71
306	8.258	0.027	HN 71
307	123.623	0.008	N 71
308	5.254	0.006	HA 72
309	2.876	0.014	HB1 72
310	2.704	0.014	HB2 72
311	7.148	0.000	HD# 72
312	7.866	0.005	HN 72
313	115.218	0.060	N 72
314	4.665	0.005	HA 73
315	1.877	0.012	HB1 73
316	1.763	0.023	HB2 73
317	8.777	0.003	HN 73
318	122.468	0.036	N 73
319	4.846	0.004	HA 74
320	1.624	0.021	HB 74
321	1.010	0.005	HG2# 74
322	8.666	0.006	HN 74
323	125.043	0.061	N 74
324	4.466	0.002	HA 75
325	1.951	0.004	HB1 75
326	1.603	0.005	HB2 75
327	8.996	0.004	HN 75
328	129.334	0.035	N 75

329	3.847	0.033	HA 76
330	8.322	0.024	HN 76
331	121.861	0.016	N 76
332	4.526	0.008	HA 77
333	3.322	0.002	HB1 77
334	2.742	0.009	HB2 77
335	7.243	0.008	HD21 77
336	6.500	0.009	HD22 77
337	7.537	0.004	HN 77
338	112.278	0.025	N 77
339	109.037	0.055	ND2 77
340	4.320	0.006	HA1 78
341	3.216	0.006	HA2 78
342	8.307	0.004	HN 78
343	109.574	0.063	N 78
344	4.713	0.004	HA 79
345	2.669	0.001	HB1 79
346	2.288	0.003	HB2 79
347	7.752	0.003	HN 79
348	120.420	0.032	N 79
349	4.240	0.007	HA 80
350	0.281	0.009	HB1 80
351	-1.355	0.005	HB2 80
352	0.332	0.010	HD1# 80
353	0.180	0.000	HD2# 80
354	7.547	0.004	HN 80
355	120.569	0.039	N 80
356	4.841	0.002	HA 81
357	9.201	0.006	HN 81
358	129.138	0.007	N 81
359	4.210	0.010	HA 82
360	1.736	0.006	HB 82
361	0.448	0.002	HD1# 82
362	0.801	0.019	HG1# 82
363	0.510	0.020	HG2# 82
364	8.738	0.005	HN 82
365	129.463	0.048	N 82
366	3.711	0.004	HA 83
367	1.648	0.008	HB# 83
368	1.188	0.002	HG# 83
369	7.704	0.004	HN 83
370	127.319	0.035	N 83
371	4.039	0.010	HA 84
372	2.545	0.009	HB1 84
373	2.250	0.011	HB2 84
374	7.233	0.014	HD21 84
375	6.518	0.003	HD22 84
376	8.516	0.007	HN 84
377	116.915	0.018	N 84
378	111.454	0.028	ND2 84
379	4.179	0.020	HA 85
380	1.605	0.000	HB# 85
381	1.863	0.014	HB1 85
382	1.641	0.000	HB2 85
383	0.720	0.011	HD2# 85
384	1.590	0.029	HG 85
385	8.034	0.036	HN 85
386	125.070	0.125	N 85
387	4.786	0.003	HA 86
388	1.241	0.005	HG2# 86
389	9.409	0.004	HN 86
390	119.114	0.023	N 86
391	3.872	0.004	HA 87
392	1.834	0.004	HB# 87
393	3.284	0.002	HD# 87
394	1.657	0.000	HG# 87

395	8.936	0.001	HN 87
396	120.395	0.014	N 87
397	4.536	0.016	HA 88
398	2.767	0.020	HB1 88
399	2.422	0.013	HB2 88
400	7.758	0.002	HN 88
401	117.015	0.020	N 88
402	4.801	0.023	HA 89
403	2.904	0.005	HB1 89
404	2.756	0.006	HB2 89
405	7.590	0.003	HN 89
406	116.440	0.015	N 89
407	4.198	0.019	HA 90
408	4.163	0.035	HB1 90
409	3.907	0.004	HB2 90
410	7.350	0.003	HN 90
411	114.920	0.019	N 90
412	4.445	0.004	HA1 91
413	3.931	0.021	HA2 91
414	8.715	0.003	HN 91
415	111.013	0.016	N 91
416	5.136	0.004	HA 92
417	3.888	0.032	HB 92
418	1.151	0.011	HG2# 92
419	8.358	0.010	HN 92
420	117.993	0.020	N 92
421	5.247	0.019	HA 93
422	3.127	0.012	HB1 93
423	3.012	0.014	HB2 93
424	7.017	0.016	HD# 93
425	6.648	0.015	HE# 93
426	9.934	0.006	HN 93
427	130.347	0.024	N 93
428	5.616	0.006	HA 94
429	2.506	0.009	HB# 94
430	7.241	0.014	HD21 94
431	6.820	0.002	HD22 94
432	9.532	0.004	HN 94
433	121.018	0.023	N 94
434	113.698	0.079	ND2 94
435	4.952	0.006	HA 95
436	-0.616	0.004	HB 95
437	0.215	0.015	HG1# 95
438	-0.396	0.012	HG2# 95
439	9.085	0.002	HN 95
440	126.007	0.034	N 95
441	4.900	0.022	HA 96
442	3.637	0.011	HB 96
443	0.929	0.016	HG2# 96
444	8.351	0.008	HN 96
445	121.829	0.292	N 96
446	4.952	0.003	HA 97
447	1.627	0.011	HB 97
448	1.074	0.016	HG1# 97
449	0.413	0.002	HG2# 97
450	7.724	0.007	HN 97
451	123.516	0.066	N 97
452	5.568	0.003	HA 98
453	2.875	0.000	HB# 98
454	3.004	0.014	HB1 98
455	2.888	0.010	HB2 98
456	6.832	0.017	HD# 98
457	6.629	0.017	HE# 98
458	9.214	0.012	HN 98
459	125.409	0.038	N 98
460	5.352	0.012	HA 99

461	9.203	0.004	HN 99
462	118.774	0.027	N 99
463	3.991	0.021	HA 100
464	4.306	0.000	HB 100
465	1.272	0.006	HG2# 100
466	8.696	0.004	HN 100
467	116.588	0.043	N 100
468	4.787	0.007	HA 101
469	3.156	0.000	HB1 101
470	2.876	0.001	HB2 101
471	7.619	0.006	HD21 101
472	6.893	0.012	HD22 101
473	112.274	0.437	ND2 101
474	4.305	0.032	HA1 102
475	3.773	0.009	HA2 102
476	8.280	0.009	HN 102
477	108.427	0.029	N 102
478	4.182	0.005	HA 103
479	1.129	0.012	HG2# 103
480	7.698	0.004	HN 103
481	117.166	0.044	N 103
482	4.136	0.000	HA 104
483	8.912	0.019	HN 104
484	128.343	0.037	N 104
485	4.182	0.035	HA 105
486	1.828	0.021	HB 105
487	0.932	0.017	HG2# 105
488	8.625	0.005	HN 105
489	127.228	0.027	N 105
490	4.466	0.003	HA 106
491	0.530	0.000	HD2# 106
492	7.220	0.013	HN 106
493	5.440	0.005	HA 107
494	2.446	0.000	HB# 107
495	2.574	0.081	HB1 107
496	2.446	0.016	HB2 107
497	7.162	0.005	HD21 107
498	8.339	0.004	HN 107
499	122.425	0.061	N 107
500	112.492	0.075	ND2 107
501	4.742	0.023	HA 108
502	1.684	0.023	HB1 108
503	1.341	0.026	HB2 108
504	1.842	0.007	HD1 108
505	1.676	0.016	HD2 108
506	3.066	0.000	HE1 108
507	3.000	0.000	HE2 108
508	1.319	0.059	HG1 108
509	8.555	0.004	HN 108
510	124.511	0.043	N 108
511	5.568	0.003	HA 109
512	1.266	0.013	HB# 109
513	8.205	0.007	HN 109
514	124.856	0.018	N 109
515	4.859	0.004	HA 110
516	1.724	0.012	HB1 110
517	1.555	0.016	HB2 110
518	0.974	0.000	HD2# 110
519	9.397	0.004	HN 110
520	123.816	0.018	N 110
521	5.157	0.008	HA 111
522	2.530	0.016	HB1 111
523	2.487	0.014	HB2 111
524	8.647	0.006	HN 111
525	123.625	0.345	N 111
526	5.031	0.023	HA 112

527	2.292	0.011	HB1 112
528	1.079	0.002	HB2 112
529	0.655	0.002	HD1# 112
530	1.226	0.000	HG 112
531	8.949	0.004	HN 112
532	127.580	0.029	N 112
533	4.497	0.007	HA 113
534	1.035	0.000	HB1 113
535	0.100	0.008	HB2 113
536	2.790	0.000	HD1 113
537	2.686	0.000	HD2 113
538	0.950	0.000	HG# 113
539	8.857	0.003	HN 113
540	128.837	0.066	N 113
541	4.687	0.022	HA 114
542	1.944	0.018	HB 114
543	0.681	0.020	HD1# 114
544	1.270	0.024	HG11 114
545	0.870	0.033	HG2# 114
546	8.352	0.004	HN 114
547	120.579	0.035	N 114
548	4.510	0.002	HA 115
549	1.623	0.000	HB# 115
550	1.651	0.023	HB1 115
551	1.596	0.024	HB2 115
552	0.728	0.015	HD1# 115
553	8.818	0.003	HN 115
554	130.347	0.016	N 115
555	4.237	0.003	HA 116
556	2.091	0.018	HB1 116
557	1.920	0.014	HB2 116
558	2.262	0.016	HG# 116
559	8.078	0.004	HN 116
560	129.667	0.016	N 116