SOLUTION STRUCTURE OF THE TARGET RECOGNITION DOMAIN OF ZOOCIN A, AN ANTIBACTERIAL ENZYME

AND

THE METAL BINDING SITE OF ZOOCIN A

by

YINGHUA CHEN

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Chemistry
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2009

ABSTRACT

A high-resolution structure has been obtained by three-dimensional NMR spectroscopy for the recombinant target recognition domain (rTRD) of zoocin A. The rTRD is a 128-residue protein responsible for targeting the cell wall of sensitive bacteria. It has a globular domain consisting of an unstructured N-terminal region, two pairs of short anti-parallel beta sheets, two sets of three-strand β-sheet and an α-helix at the c-terminal. A search for the similar fold with DALI server and SWISS-MODEL server yields no significant match, suggesting the novel folding of the rTRD. A hypothesis for the location of the binding site was proposed after observing that EDTA bound specifically to the rTRD. Chemical shifts of backbone amides of T64, G67, T70, G79, Y80 and V82 are affected by EDTA binding, and these residues are close to the poorly-structured region (residues 71-74) of the rTRD, which could provide flexibility for the rTRD to bind to its natural substrate.

Additionally, the metal binding site of zoocin A has been studied by ^{113}Cd-NMR and ^{15}N-HSQC experiments. The two ^{113}Cd resonances at 113.6 ppm and 107.2 ppm suggested that two nitrogen and two oxygen atoms ligate to metal center according to the correlation between the chemical shifts of Cd resonances and coordination environment of ^{113}Cd complexes. As the metal binding site is only located at the catalytic domain (CAT) of zoocin A, two-dimensional ^{15}N HSQC experiments for the recombinant catalytic domain (rCAT), metal-free rCAT and reconstituted Zn-rCAT provide evidence for Zn^{2+} as the metal cofactor. Therefore, zoocin A is also a Zn metalloprotein like its homologous proteins lysostaphy and LytM.

DEDICATION

This dissertation is dedicated to everyone who helped me and guided me through the trials and tribulations of creating this manuscript.

LIST OF ABBREVIATIONS AND SYMBOLS

°C	Degrees Celsius
g	Microgram
µL	Microliter
µM	Micromolar
1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
4D	Four-dimensional
A	Alanine
Arg	Arginine
ARIA	Ambiguous restraints for iterative assignment
Asn	Asparagine
Asp	Aspartic acid
BLIS	Bacteriocin-like inhibitory substance
BME	Basal media eagle
CAT	Catalytic domain
cm	Centimeter
CM	Carboxymethyl
cm^{-1}	Reciprocal centimeter
CNS	Crystallography and NMR system
COSY	Correlation spectroscopy
Cys	Cysteine
Da	Dalton
DG	Distance geometry
DNA	Deoxyribonucleic acid
DSS	2,2-Dimethyl-2-silapentane-5-sulfonic acid

DTT	Dithiothreitol
E. coli	*Escherichia coli*
EXSY	Exchange spectroscopy
F	Phenylalanine
FT	Flow through
g	Gram
G	Glycine
GHz	Gigahertz
Gln	Glutamine
Gly	Glycine
H	Histidine
His	Histidine
HMQC	Homonuclear single quantum correlation
HSQC	Heteronuclear single quantum correlation
I	Isoleucine
Ile	Isoleucine
IPTG	Isopropyl-D-thiogalactoside
IUPAC	International Union of Pure and Applied Chemistry
K	Kelvin
KDa	Kilo-dalton
L	Liter(s)
LB	Luria-Bertani broth
Leu	Leucine
Lys	Lysine
M	Molar
M	Methionine
MDSA	Molecular dynamics-based simulated annealing
Met	Methionine
min	Minutes
mL	Milliliter
mM	Millimolar
mM^{-1}	Reciprocal millimolar

ms	Milliseconds
mV	Millivolts
Ni-NTA	Nickel-nitrilotriacetic acid
nm	Nanometer
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect correlation
NOESY	Nuclear Overhauser effect spectroscopy
P	Proline
pdb	Protein data bank
Phe	Phenylalanine
ppm	Parts per million
Pro	Proline
rCAT	Recombinant catalytic domain
RMSD	Root-mean-square distances
rpm	Rotation per minute
rTRD	Recombinant target recognition domain
rZOO	Recombinant zoocin A
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
Ser	Serine
Thr	Threonine
TP	Threonine-proline
TRD	Target recognition domain
Trp	Tryptophan
TSP	3-trimethylsilyl-propionic acid-d4, sodium salt
Tyr	Tyrosine
TYS	Trypticase-yeast extract-starch
V	Valine
Val	Valine
W	Tryptophan
Y	Tyrosine
ZOO	Zoocin A

ACKNOWLEDGMENTS

I have benefited from the help of many people throughout the years. First of all, I am very grateful to my advisor, Dr. Russell Timkovich for encouraging me into NMR field and teaching me an enormous amount with great patience. His guidance and encouragement during all stages of this dissertation were invaluable.

I would like to thank my committee members, Dr. John Vincent, Dr. Paul LeBlanc, Dr. Silas Blackstock and Dr. Stephen Woski, for their helpful advice and assistance on this dissertation. Also I thank Dr. Robin Simmonds for providing us engineered cells with generosity, and Dr. Sloan's group for their help with Cd-rZOO activity analysis.

I am thankful to my former group member, Dr. Qiaoli Liang, who did the preliminary work on zoocin A and gave me many helpful suggestions during my research work.

Dr. Ken Belmore provided me with invaluable assistance in obtaining NMR spectra. He was always ready to answer questions and generous with his time.

Last, but by no means least, I would like to thank my husband and my parents, without whose love and support, none of this would have ever been achieved.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

OVERALL INTRODUCTION

Some bacteria are capable of secreting certain proteinaceous antibiotics called bacteriocins, which exhibit inhibitory activity against some other bacteria and yet little toxicity to humans. However, certain bacteria have developed mechanisms for resisting antibiotics. Therefore, seeking new antibiotics becomes a way to keep pace with the development of bacterial resistance (James, 1988).

Zoocin A, an extracellular enzyme secreted by *Streptococcus* equi subsp. *zooepidemiccus* 4881, is a bacteriocin-like inhibitory substance (BLIS) (Simmonds, 1995). This is capable of disrupting the cell wall of most closely-related streptococcal bacteria, including all *S. pyogenes* strains, all *S. mutans* strains except *S.rattus*, and all of *S. zooepidemicus* strains except strain 4881 itself. In artificial plaque, zoocin A reduces the level of *mutans streptococcus*, that causes dental caries in humans and animals, and selectively eliminates it. Thus, it is a potential antibiotic.

The amino acid sequence of zoocin A has been determined (Figure 1) (Simmonds, 1997). Gene *zooA* can be encoded into a 30 kDa (285-amino acid residues) zoocin A pre-peptide, in which four regions have been identified from an amino acid sequence homology data. 1) The leader sequence (Met -23 to Ala -1) is a signal peptide, which is responsible for the transportation of zoocin A from the cell to its extra-cellular environment. It will be cleaved before a biologically active form of 262 residues zoocin A is released. 2) The N-terminal region

(Pro +20 to Pro +115) contains the catalytic domain (CAT), which starts from residue 24 and is highly homologous to the N-terminal sequence of several endopeptidases, including lysostaphin (about 40% identities, 63% positive match), a zinc endopeptidase produced by *Staphylococcus simulans* biovar *staphylolyticus*, and LytM (about 35% identities, 54% positive matches), an autolysin from *Staphylococcus aureus*. The above homology analysis was obtained through basic local alignment search tool (BLAST) II, which is a similarity search program designed to explore the entire available sequence database (Tatusova and Madden, 1999). "Identities" indicates how many amino acids in the query sequence exactly match the database's found sequence, while "positive matches" indicates how many amino acids in the query sequence are at least similar to the database's found amino acid sequence based on their size, acidity and polarity. 3) The C-terminal region (Ile +189 to Asp +206) contains the target recognition domain (TRD) which is able to specifically bind the cell walls of some bacteria. However, matches to the C-terminal region are generally weak. 4) A threonine-proline-linker (Thr 143 to Thr 153) joins the N-terminal domain to the C-terminal domain. This short TP-rich sequence also has homology to linkers of several domain-type proteins including lysostaphin.

The sequence of the the catalytic domain of zoocin A shows similarity to the endopeptidase lysostaphin (Simmonds, 1995), suggesting that zoocin A may also be an endopeptidase. Previous studies have demonstrated that zoocin A is capable of cleaving a hexaglycine peptide with

Figure 1. Amino acid sequence of zoocin A (Simmonds, 1997). The highlighted region is the TP-linker domain.

similar efficiency to that of lysostaphin (Browder, 1965; Heath, 2004). A recent study showed that zoocin A is a D-alanyl-L-alanine endopeptidase (Figure 2) (Gargis, 2009). Lysostaphin-producing bacteria protect themselves from the lytic reaction of their own product through expressing a gene, *epr* or *lif*, which specifies for changes in the peptidoglycan cross bridges of its cell wall. Likewise, zoocin A-producing bacteria have an immunity gene *zif* responsible for protecting the producer. The gene product, Zif immunity protein, provides resistance to peptidoglycan cross bridge hydrolases by introducing an additional L-alanine in the peptidoglycan cross bridges (Gargis, 2009).

The three-dimensional structure of zoocin A will provide an essential framework for a molecular level understanding of specific action mode of its anti-bacterial activity. NMR has become a major technique for determining protein structure since the 1990s. However, three-dimensional NMR experiments require $^{13}C/^{15}N$ labeled protein samples, which can become expensive. The target recognition domain (rTRD) and the catalytic domain (rCAT) have been expressed separately in *E.coli* (Simmonds, 1997). Previous studies showed that rTRD is more soluble and more stable than rCAT (Liang, 2004). Therefore, the first part of this dissertation focused on solving the solution structure of the target recognition domain of zoocin A by 3D NMR. The second part used 2D NMR and ^{113}Cd-NMR technology to probe the metal binding site of zoocin A.

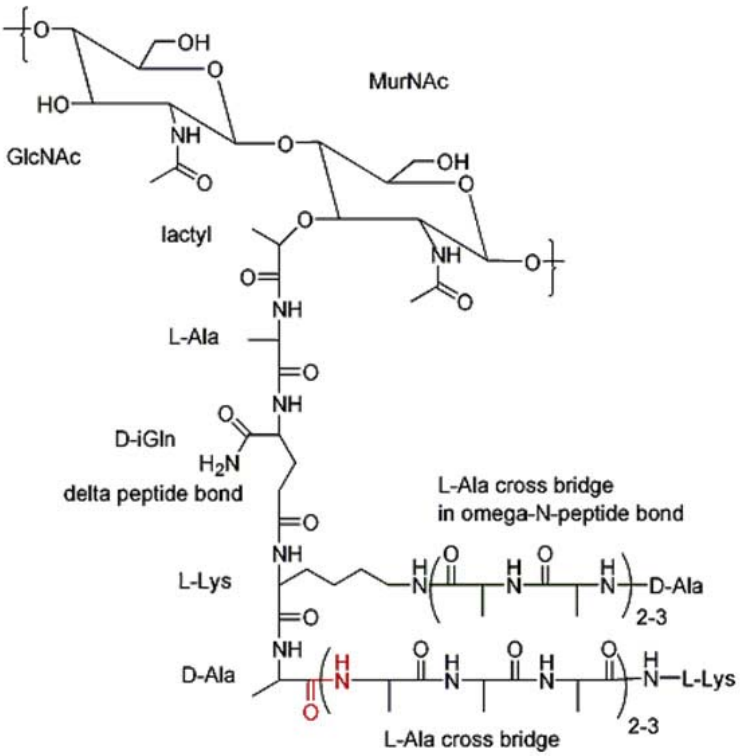


Figure 2. The action site of zoocin A on the peptidoglycan of the cell wall of its sensitive bacteria.

PART I

SOLUTION STRUCTURE OF THE TARGET RECOGNITION DAOMAIN OF ZOOCIN A, AN ANTIBACTERIAL ENZYME

1 Introduction

1.1 General Approach to Protein Structure: X-ray Crystallography vs. NMR Spectroscopy

X-ray crystallography and NMR spectroscopy are the two main techniques that are capable of providing a complete description for protein structure at atomic detail. In the database of known protein structures, 85% of the protein structures were solved by X-ray crystallography, and approximately 13% of the protein structures were solved by NMR with the remainder solved by electron microscopy (according to PDB staticstics at www.rcsb.org, accessed on Aug. 2009). NMR lags behind X-ray crystallography on the size of protein studied. X-ray crystallography has been applied to proteins with molecular weights up to 2,500 KDa, while the upper limit of NMR is typically less than 60 kDa (Machalek, 2007). The size limitation of NMR technology is mainly caused by the following two aspects. (1) Larger molecules have extensive resonance overlap, which makes spectra analysis complicated. (2) The transverse relaxation time T2, which

represents the lifetime of a resonance signal in the transverse XY plane, is inversely proportional to the half line width of the peak ($\Delta\upsilon_{1/2} = \frac{1}{\pi T_2}$). Larger molecules tumble more slowly in solution, and therefore have shorter T2, which leads to broader peaks and decreases resolution of NMR spectra. Despite this size limitation, one of the advantages of NMR over X-ray is that no crystal is needed. Obtaining crystals that diffract to sufficient resolution can be difficult. It can require large amounts of sample to test crystallization conditions, which is time-consuming and not always productive. Additionally, appropriate heavy metal derivatives are needed to determine accurately the phases of diffraction. Another advantage of NMR is that proteins can be characterized in solution under a physiologically relevant solution environment. Therefore, NMR is a powerful complimentary technique for protein structure determination. In this project, the molecular weight of rTRD is 14 kDa, which is well within the size limit of NMR technology.

1.2 One-dimensional to Multidimensional NMR

NMR technology has undergone the development from one-dimensional to multiple-dimensional experiments during the last 50 years. One-dimensional NMR has been widely used for organic molecules since the 1960s. The basic pulse sequence for any 1D experiment includes a preparation period that leads to the production of transverse magnetization to start the experiment (e.g. a 90º pulse to excite the nuclei) and a detection period where the signal is collected along the time axis (Figure 2). The application of 1D spectroscopy to biological systems was very limited due to its poor sensitivity and very low resolution at that time. In the 1970s, two-dimensional NMR spectroscopy increased the spectral resolution. The first generation of solution structures of biomacromolecules was reported in the 1980s using 2D

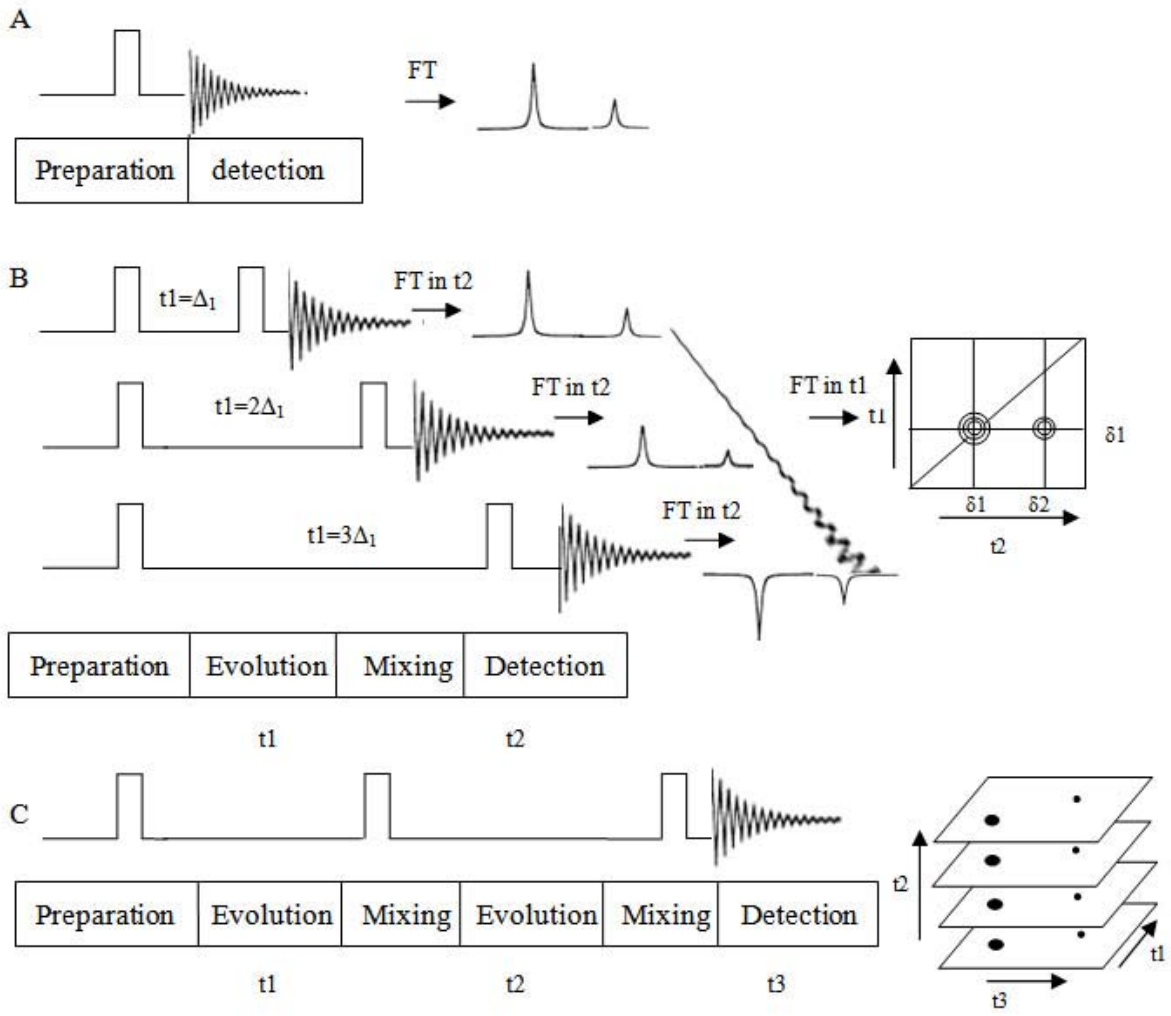


Figure 2. General schematic building blocks and pulse sequence for 1D and multi-dimensional NMR experiments. A) 1D, B) 2D, C) 3D

NMR (Williamson, 1985). In 2D NMR, an evolution time t1 (which allows the magnetization to precess under the influence of chemical shifts to create one of the frequency axes in a multidimensional experiment) and a mixing period (which allows magnetization or coherence to transfer from one spin to another) are inserted between the preparation period and detection periods (Figure 2B). The resultant 2D spectrum consists of diagonal peaks that contain the same information as in its 1D spectrum and cross-peaks that carry the frequencies from both evolution period t1 and detection period t2. The t2 dimension has better digital resolution than the t1 dimension because the detection period t2 is always longer than the evolution period t1. When the size of protein molecular weight increases above 10 kDa, spectral overlap in 2D spectra becomes a major hurdle for extracting chemical shift information. An additional frequency dimension can be added by inserting another one or two evolutions and mixing period building blocks into a 2D pulse sequence scheme (Figure 2C) such that the peaks crowded in a 2D spectrum can be spread along the third dimension or the fourth dimension according to the chemical shifts of another nucleus, which results in the multi-dimensional heteronuclear NMR spectra. Similar to 2D NMR spectra, the detection period t3 in 3D NMR is always longer than the evolution period t1 or t2 thus providing better resolution in the t3 dimension, which is always ^{1}H due to its highest sensitivity compared to other nuclei.

Two types of nuclear interactions are used in 2D or multidimensional NMR spectra. The first are experiments involving coherence transfer (through-bond interaction−scalar coupling) include total correlation spectroscopy (TOCSY), correlation spectroscopy (COSY), heteronuclear multiple quantum coherence (HMQC) and heteronuclear single quantum coherence (HSQC). The

second are experiments involving incoherent transfer (through-space interaction−dipolar coupling via NOE or chemical exchange) include nuclear Overhauser effect spectroscopy (NOESY) and chemical exchange experiment (EXSY).

1.3 General Procedure of Determining Solution Structure for Large Proteins (M.W. >10 kDa)

A general procedure can be given for the NMR structure determination of proteins with molecular weights higher than 10 kDa (Figure 4). Prior to structure calculation, chemical shifts for ^{1}H, ^{13}C, and ^{15}N nuclei within the protein are obtained through resonance assignments. Next, J coupling constants are important for geometric angles and must be determined or estimated. Finally, ^{1}H-^{1}H distances within 5.5Å, which are the most important restraints for protein folding, must be measured from 3D NOESY spectra. After structure calculation, hydrogen bonding and residual dipolar coupling constants (RDC) information can be used to refine the structures. H-bonds provide short distances between the carbonyl oxygen and the amide proton. RDC is complementary to NOE data since it provides long range distance (>5.5Å) information.

1.3.1 Protein Sample Requirement

Structure determination for a protein with molecular weight above 10 kDa may require proper collection and analysis of 10-20 individual two dimensional and three-dimensional NMR spectra, and most experiments will take several days of data acquisition. In order to collect all required data using a minimum number of $^{13}C/^{15}N$ labeled samples, the investigated protein must be soluble and stable without aggregation at room temperature for at least several weeks. In contrast to the 2D ^{15}N HSQC experiment where a sample of 0.3 mM is enough to produce a spectrum

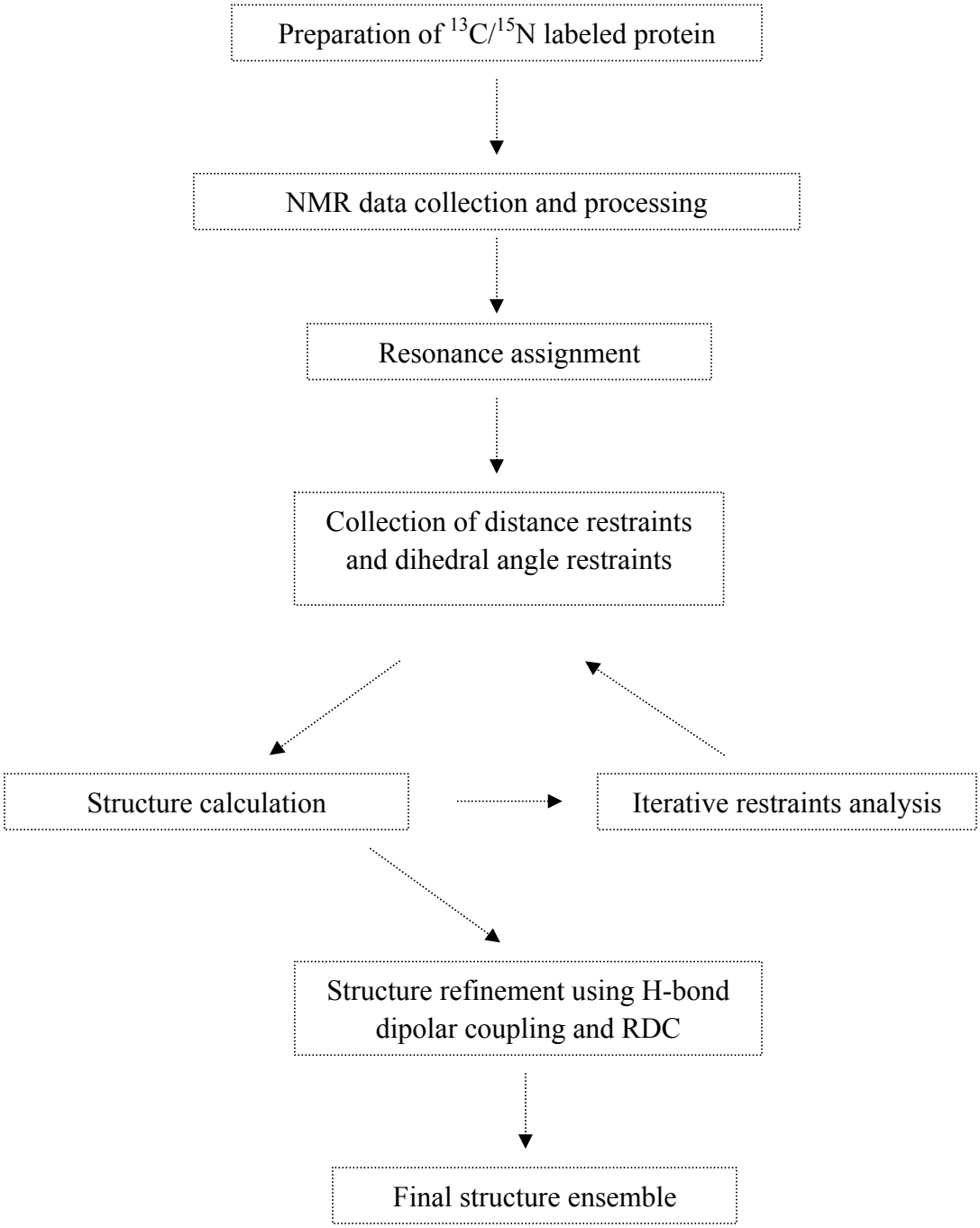


Figure 4. Strategy for NMR-based structure determination.

with adequate signal-to-noise ratio, good quality 3D spectra, as a rule of thumb, require higher concentrations.

1.3.2 Data Collection and Processing

3D NMR spectra require NMR spectrometers with triple resonance probes and gradient fields so that the desired coherence transfer pathway can be selected. Commonly used triple resonance experiments will be discussed under sequential resonance assignment in the next section. Peak analysis, such as peak picking and editing, can be performed by common software applications such as NMRView (Johnson and Blevins, 1994), CCPNmr analysis (integration of Sparky and Ansig) (Vranken et al., 2005), XEasy (Bartel et al., 1995), and PIPP (Garrett et al., 1991). NMRView and CCPNmr analysis are still being actively developed and are distributed for free to academic users. 3D spectra can be processed using a variety of software applications, such as NMRPipe (Delaglio et al., 1995), TopSpin (Bruker), and Azara (Boucher, 2002).

1.3.3 Sequential Resonance Assignment

Sequential resonance assignment is the process of assigning each observed resonance (chemical shift) to a nuclear spin within the protein. It is a prerequisite for structure interpretation because the approximate distance between protons and backbone dihedral angle restraints, which will be used in various type of calculation methods such as restrained molecular dynamics by simulated annealing (MDSA) (van Gunsteren and Dolenc, 2008) and distance geometry (DG) (Crippen, 1981), are generated based on the chemical shift information. Thus the completeness and uniqueness of resonance assignment are directly associated with the resolution and accuracy of protein structure.

Resonance assignment starts with ^{15}N HSQC analysis, followed by backbone assignment, and ends up with side-chain assignment.

1.3.3.1 Backbone Assignment

^{15}N HSQC is the preliminary and the most frequently recorded experiment in protein NMR (Bodenhausen and Ruben, 1980; Serber and Dötsch, 2001). It resolves individual amide proton-nitrogen correlations by exploiting the strong J coupling (~92 Hz) between the amide proton and its directly attached nitrogen in a ^{15}N-enriched sample (boxed in Figure 5A). The number of resolved resonances in this spectrum should basically agree with the number of amide protons, and the peak intensities should be uniform. These two features can be utilized to test the quality of isotopically labeled samples. Good spectral dispersion in both proton and nitrogen dimensions and proper number of resonances indicate whether a protein is folded or not. It should be noted that the ^{15}N HSQC experiment is not very useful in resonance assignment on its own because no assignment can be made solely based on this spectrum. However, it is a core part of multidimensional NMR experiments as almost all the 3D experiments for backbone resonance assignment are essentially spreading the cross-peaks in a 2D ^{15}N HSQC spectrum along some third dimension according to the chemical shifts of ^{13}C nuclei (C_α, C_β or C') (Figure 5B).

A typical standard routine for backbone resonance assignment requires a series of ^{1}H, ^{13}C, ^{15}N triple resonance experiments, HNCACB (Wittekind et al., 1993; Wittekind and Mueller, 1993; Yamazaki et al., 1994), HNCA (Grzesiek, 1992; Kay, 1994; Muhandiram, 1994), HN(CO)CA, HN(CA)CO and HNCO (Yamazaki et al., 1994), which correlate backbone amide

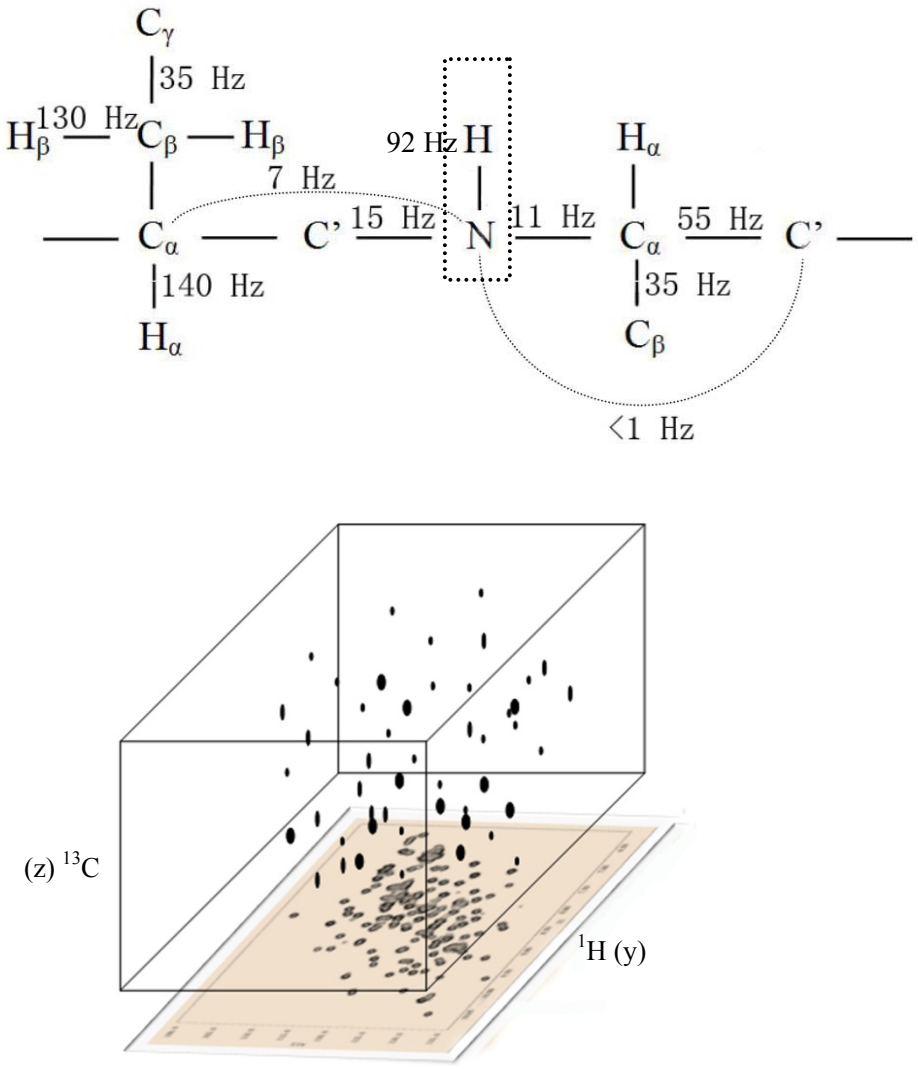


Figure 5. Basic information for 3D NMR. A) J coupling constants between ^{1}H, ^{15}N and ^{13}C along a polypeptide chain as used in triple resonance NMR experiments for resonance assignment (redraw from Teng, 2005). B) Diagram of a 3D HSQC-TOCSY spectrum. The projected spectrum in xy plan is a 2D ^{1}H-^{15}N HSQC spectra.

H and N, C^{α}, CO and H_{α} atoms from neighboring amino acids using one bond and two bond J-couplings (Figure 5A). The coupling constant between the hetero-nuclear spins is much larger than that between protons, so that the magnetization transfer will be more efficient resulting in more intense cross-peaks. For each of the above experiments, the magnetization of the proton spin is transferred to the neighboring spins and then back to the proton for detection by the same pathway because of the high sensitivity of ^{1}H nuclei. This is referred to as an "out-and-back" coherence transfer pathway. Experiment names follow a standard naming convention. All nuclei that are used for short distance coherence transfer by J-couplings are listed in the name. An atom in parenthesis is a "relay" station for the coherence transfer and does not appear in its own dimension in the spectra. For example, in the HN(CA)CO experiment, the magnetization transfers from amide H to amide N, then to the carbonyl carbon (C') through C_{α}, and the resultant 3D spectrum consists of axes for amide H, amide N and C', but not C_{α}. HNCACB assigns C_{α} and C_{β} resonances of a given residue and of the residue before it in the sequence. The correlations of this experiment are indicated schematically in Figure 6A. In most cases, the intensity of the resonance from residue *i* is stronger than that from *i*-1 due to the stronger $^{1}J_{CN}$ (~12Hz) coupling than $^{2}J_{CN}$ (~7Hz). But sometimes they are too close to distinguish, and this results in ambiguity. Also, missing resonances, degenerated resonances, and artifacts associated with large peaks are probable sources of ambiguity. Most of these ambiguities can be solved by other 3D experiments since there is a small chance that the same chemical shift degeneracy or ambiguity occurs in all other spectra. The HNCA experiment (Figure 6B) correlates the amide

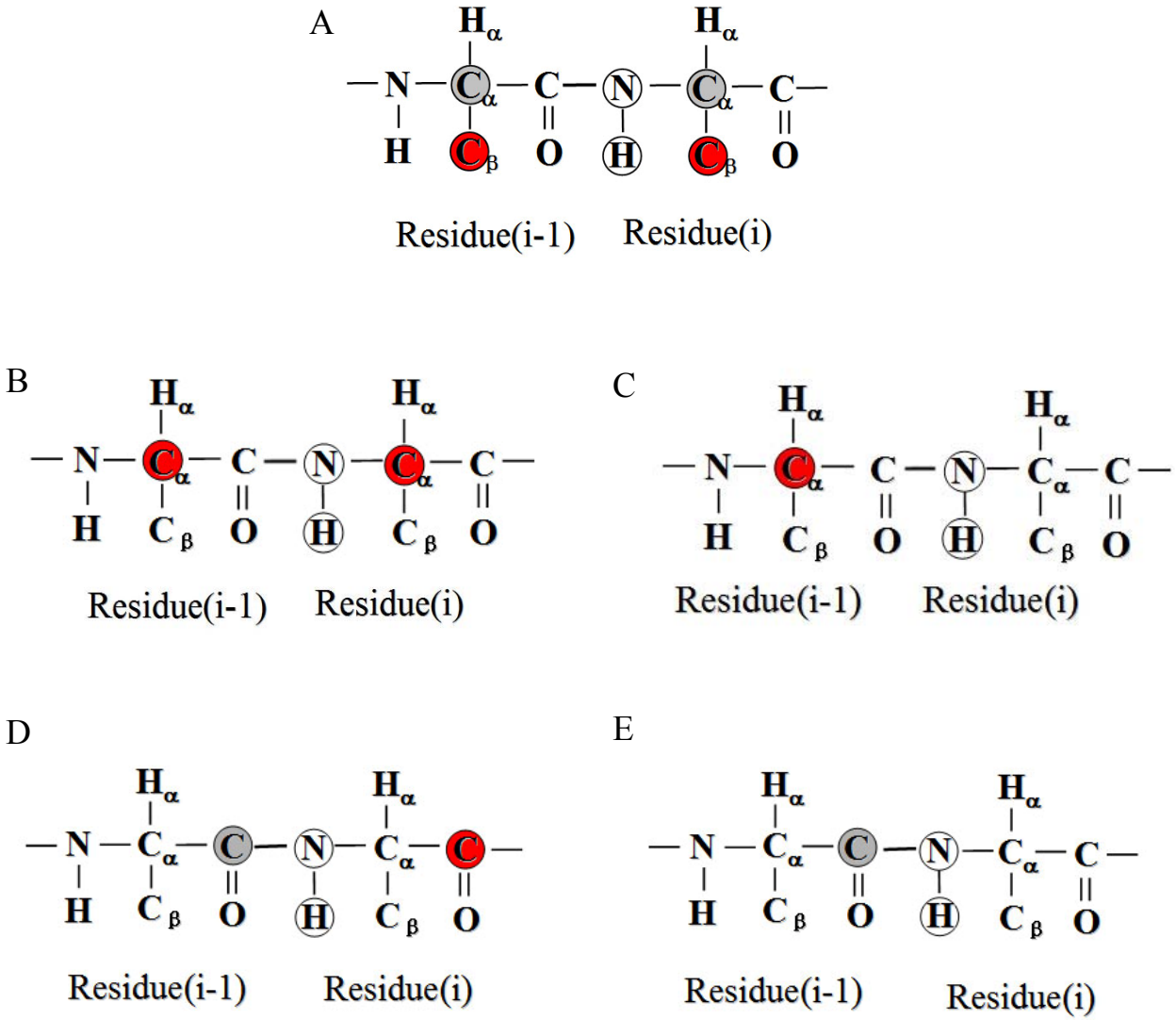


Figure 6. Three-dimensional experiments for Backbone resonance assignments. Involved nuclei are shown in closed or open circles. A) HNCACB; B) HNCA; C) HN(CO)CA; D) HNCO; E) HN(CA)CO.

proton of residue *i* to its attached N (N^i), to the attached C_α (C_α^i) and the C_α of the previous residue *i*-1 (C_α^{i-1}). The chemical shifts for C_α^i and C_α^{i-1} shown in this spectrum clarify the resonance assignments of C_α and C_β from the HNCACB experiment. The HN(CO)CA experiment (Figure 6C) gives the correlation of the amide H and amide N to the C_α of the previous residue (C_α^{i-1}) via the C_{i-1}=O (C') atom, thus only interresidual cross signal can be observed. This experiment helps the HNCA experiment distinguish C_α^i from C_α^{i-1} unambiguously. The above three experiments allow the correct order of chemical shifts of backbone NH, C_α and C_β to be deduced, following which, the characteristic chemical shift of C_α and C_β for some amino acid residues (highlighted in Table 1) (Schwarzinger, 2000) can then be used to map the correct order of chemical shifts onto the amino acid sequence. For example, chemical shift for C_β of Ala is lower than 25 ppm, the chemical shift for C_β of serine is around 62 ppm, and the chemical shift of C_β of Thr is around 70 ppm. Missing resonances may still leave gaps in the sequential assignment. The HN(CA)CO (Figure 6D) and HNCO (Figure 6E) experiments are used to fill in some of these gaps by showing carbonyl carbon chemical shifts. HN(CA)CO gives the correlation of the amide H and N to the carbonyl resonance as well as to that of the preceding residue by means of the $^1J_{HN}$, $^1J_{NC'}$ and $^2J_{NC'}$ coupling constants. The HNCO experiment correlates backbone connectivities between the amide N-H with the carbonyl resonance of the preceding residue, which can be used to confirm the carbonyl assignment made in HN(CA)CO.

1.3.3.2. Side-chain Assignment

Aliphatic side-chain resonance assignment requires CCC-TOCSY (Montelione, 1992), HCC-TOCSY (Serber, 2001) and HCCH-TOCSY (Sattler, 1995) experiments to provide through-bond

Residue	NH	H	$H_{\beta 1}$	$H_{\beta 2}$	C_{α}	C_{β}	CO
Ala	123.8	4.32	1.39		52.3	19.0	177.8
Cys	118.6	4.55	3.28	2.96	56.0	28.9	174.0
Asp	120.4	4.64	2.84	2.75	54.0	40.8	176.3
Glu	120.2	4.35	2.09	1.97	56.4	29.7	175.8
Phe	120.3	4.62	3.22	2.99	58.0	39.0	174.9
Gly	108.9	3.96			45.1		173.3
His	119.1	4.73	3.26	3.20	54.5	27.9	173.3
Ile	119.9	4.17	1.90		61.3	38.0	176.4
Lys	120.4	4.32	1.85	1.76	56.5	32.5	176.6
Leu	121.8	4.34	1.65		55.1	42.3	177.6
Met	120.5	4.48	2.15	2.01	55.3	32.6	176.3
Asn	118.7	4.74	2.83	2.75	52.8	37.9	175.2
Pro-trans		4.42	2.28	2.02	63.1	31.7	177.3
Gln	119.8	4.34	2.13	2.01	56.1	28.4	176.0
Arg	120.5	4.34	1.89	1.79	56.1	30.3	176.3
Ser	115.7	4.47	3.88		58.2	63.2	174.6
Thr	113.6	4.35	4.22		62.1	69.2	174.7
Val	119.2	4.12	2.13		62.3	32.1	176.3
Trp	121.3	4.66	3.32	3.19	57.7	30.3	176.1
Tyr	122.3	4.55	3.13	2.92	58.1	38.8	175.9

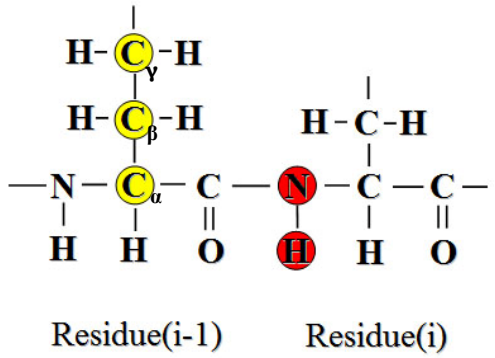
Table 1 ^{1}H, ^{13}C and ^{15}N chemical shifts (in ppm) for the 20 common amino acids in random coil referenced to TSP (3-trimethylsilyl-propionic acid-d4, sodium salt), TSP and liquid ammonia respectively. Characteristic chemical shifts are highlighted in yellow.

connectivities. The first two experiments correlate side chain C's or side-chain ^{1}H's with the backbone amide group of the next residue. CCC-TOCSY is designed to obtain the chemical shifts of all C's (Figure 7A). The magnetization is expected to be transferred from amide H to N then to the C_{α}, C_{β}, C_{γ} … of preceding amino acid residue. Therefore, only intraresidue cross-peaks can be possibly observed. Likewise, the chemical shift for all aliphatic ^{1}H's within the amino acid residue can then be obtained by the HCC-TOCSY experiment, which correlates all aliphatic ^{1}H with amide N-H of the preceding amino acid residue (Figure 7B). It should be pointed out that the number of observed cross-peaks in the above 3D TOCSY experiments varied with the isotropic mixing time. The longer the mixing time, the further the magnetization will be transferred through the bonds, but some of the other cross-peaks become weaker. Hence some resonances will be missing due to weak signals or short magnetization transfer. This problem can be solved to some extent by repeating experiments with different isotropic mixing times. Another experiment, HCCH-TOCSY (Figure 7C), correlates all aliphatic ^{1}H and ^{13}C within a residue and can fill in some of the missing resonances. In this experiment, the magnetization transfers from aliphatic proton to the attached ^{13}C spins within residues via $^{1}J_{HC}$ (140 Hz) and $^{1}J_{CC}$ (30-40 Hz) coupling constants and finally to their attached protons. The type of amino acid can be determined or confirmed by the pattern of the ^{13}C or ^{1}H chemical shifts of a spin system using the random coil chemical shift table (Table 1).

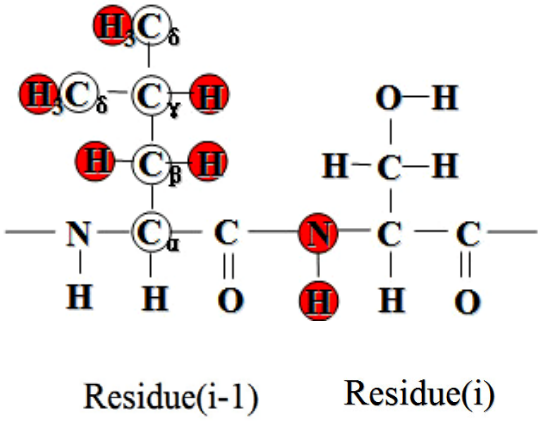
1.3.4 Resonance Assignments for Aromatic Ring Protons

Long range NOE's provide crucial distance restraints to define the tertiary structure of protein by NMR. Aromatic resonances are normally located at the hydrophobic core of protein and thus

A



B



C

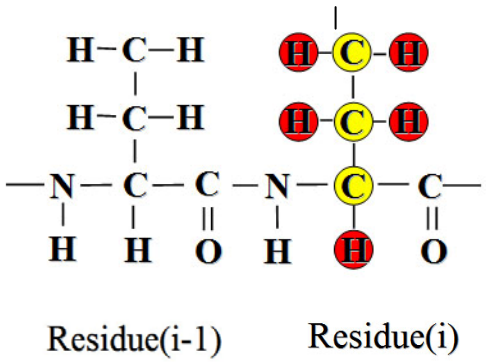


Figure 7. Three-dimensional experiments for side-chain resonance assignments. A) HCC-TOCSY ; B) HCCH-COSY and C) HCCH-TOCSY.

involve many long-range NOE contacts. Unlike its aliphatic region, the sequence specific assignment of aromatic side chains can not be achieved by ^{1}H-^{1}H TOCSY experiments as the scalar coupling is interrupted by the quaternary carbon at the γ position. Other 3D experiments such as HCCH-TOCSY also fail due to the large offset between aliphatic and aromatic carbon resonances. If the chemical shifts of H_β's are obtained, the Hβ(CβCγCδ)Hδ experiment can afford the desired H_δ assignments for all aromatic amino acid types, and the Hβ(CβCγCδCε)Hε experiments introduces the C_δ to C_ε transfer to give the chemical shift of Hε, which is especially important for Phe residue due to its small chemical shift differences between aromatic ^{13}C nuclei and the relatively large $^{1}J_{CC}$ couplings (Yamazaki, 1993). Similarly, (H)CB(CGCC-TOCSY)H^{ar} experiment proposed by Löhr Frank's group, relying on the side-chain to side-chain correlations involving β-carbons, is an alternative way to assign the aromatic proton resonaces. They also proposed an innovative experiment called H^{ar}(CC-TOCSY-CGCBCACO)NH, which is designed for sequential assignment of all aromatic protons in uniformly ^{13}C/^{15}N labeled proteins using standard spectrometer hardware without the prior knowledge of β-carbon and proton chemical shifts (Löhr, 2007).

1.3.5. The Role of Chemical Shifts

The chemical shift of a nucleus contains structural information as it is sensitive to the environment. NMR spectroscopists have generalized the correlation between chemical shifts of amide H, amide N, H_α, C_α and C=O (C') and the secondary structure of a protein from a large number of accurately determined protein structures, and sophisticated software such as random coil index (RCI) (Berjanskii, 2005) and CS-ROSETTA (Shen, 2009), have been used to predict

secondary and tertiary structure respectively for proteins by using chemical shifts of amide H, amide N, C_α. C_β, H_α, and C'. Once the sequential assignment has been made, the chemical shift can then be used as a "label" for each ^1H, ^{15}N and ^{13}C nucleus in the amino acid sequence so that the ^1H chemical shifts can be used to assign the NOE crosspeaks, which defines the spatial proximity between a pair of protons, and thus defines the global folding of a protein.

The secondary structure predicted by RCI may help NOE assignments in the regular secondary structure regions. For example, most of the regular through-space interactions between H_α^i and H_N^{i+3}, H_α^i and H_N^{i+4}, H_α^i and H_β^{i+3}, and H_N^i to H_N^{i+2} are expected to be observed in an alpha helix (Figure 8A). In a β-sheet, H_α^i shows strong interaction (NOE cross-peak) to H_N^{i+1}, H_α^i shows strong interaction to H_α^j, and H_N^i shows very weak interaction to H_N^{i+1} (Figure 8B). The tertiary structure predicted by CS-ROSETTA can be used as an initial structure for structure calculation.

1.3.6 Restraints for Structure Calculation

A polypeptide chain can adopt a huge number of conformations due to the free rotation around its chemical bonds (except for the peptide bond). Primary restraints including distance and torsion angle restraints are essentially the forces (or "glue") to fold the protein into a tertiary structure, all of which can be obtained experimentally. If enough restraints can be obtained, the polypeptide chain can be locked into a unique structure. As a rule of thumb, a good structure requires about 10 distance restraints per residue. Other restraints such as J-couplings, dipolar couplings and H-bonding constraints are also necessary to obtain a high resolution structure.

1.3.6.1 Distance Restraints and NOE Assignments

Distance restraints are derived from 2D or 3D NOESY experiments. Each NOE crosspeak represents a pair of protons that are within 6Å apart. Intra-residue NOEs only contain the distance information for protons within an amino acid residue ($i=j$, i and j are residue #'s). Sequential NOEs shows the distance information for protons that are located at neighbouring residues in sequence ($|i-j|=1$). In most cases, these short range NOEs are not crucial for protein folding but affect sequential side-chain packing. If an NOE cross-peak involves two protons no more than five amino acids apart in sequence, this cross-peak is defined as a medium range NOE ($2<|i-j|<5$), which is valuable for secondary structure determination. For example, in an α-helix, NOESY cross peaks are limited to a maximal of four residues ahead or behind. If an NOE cross-peak involves two protons more than five amino acids apart, this NOE is called a long range NOE ($|i-j|\geq 5$), which connects distant amino acids in sequence and is crucial for the overall protein folding. This type of NOE is common in the antiparallel or parallel β-sheet.

NOE assignment is the most laborious work in protein structure determination. The typical situation is that there are thousands of NOE cross-peaks in a NOESY spectrum, and usually only a very small portion of them can be unambiguously assigned solely based on the chemical shift agreement due to chemical shift degeneracy and inconsistency of the NOESY cross-peak positions. These unambiguously assigned cross-peaks, or seed assignment, are then converted to distance restraints based on the peak intensity or volume according to the following equation (Ernst, 1991):

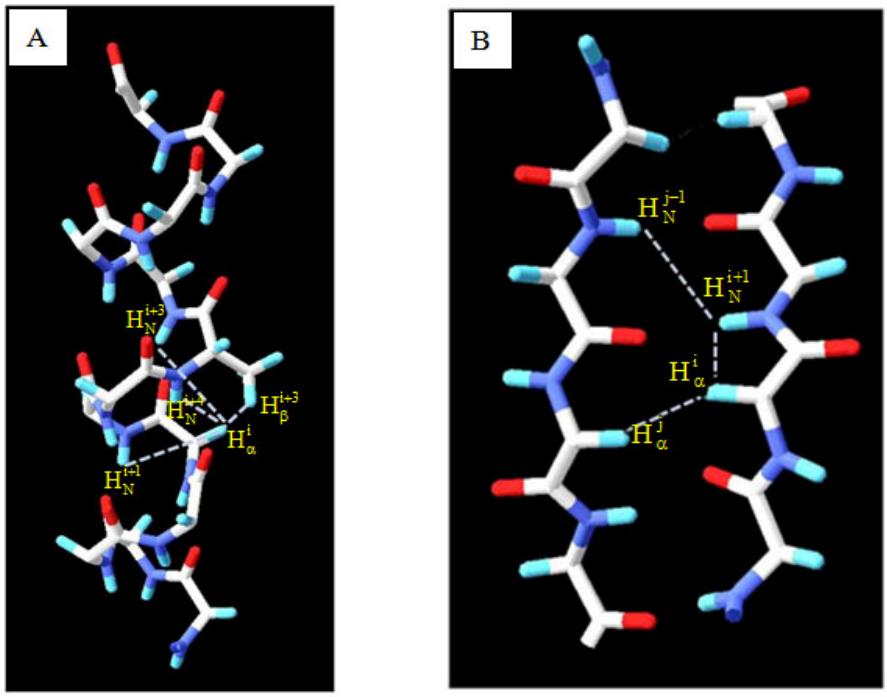


Figure 8. The typical NOE interactions for A) α-helix and B) β-sheet.

$$D_{ij} = (\alpha^{-1} V)^{-\frac{1}{6}}$$

where D_{ij} is the distance between two nuclei *i* and *j*. V is the volume or intensity of the peak. The calibration factor, α, can be computed according to the volume or intensity of a peak assigned to a pair of protons with fixed distance, i.e. two β protons in Leu, Ile, Lys, or Pro can be used to calibrate ^{13}C-edited NOESY spectrum, while the fixed distance between $H_{\delta 1}$ and $H_{\varepsilon 1}$ in Trp ring can be used to calibrate an ^{15}N-edited NOESY spectrum. The nomenclature for amino acids follows the IUPAC naming system. However, some cross-peaks may arise as a sum of two or more distinct NOE's, and some cross-peaks probably involve degenerate proton chemical shifts, both of which mean there are multiple assignment possibilities for one peak. To assign these ambiguous NOEs, an initial structural model with best agreement with the conformational restraints must be obtained from seed assignment, and then some of the ambiguous NOEs can be interpreted based on this initial model. After a second round of structure calculations, a new structure model can be generated and additional ambiguous NOEs can then be assigned. This process iterates until most of the ambiguous NOEs are assigned. This is also the principle used by some automated NOESY assignment programs such as ARIA (ambiguous restraints for iterative assignment) (Nilges et al., 1997; Nilges, 1997). In this study, automated ARIA did not lead to a consistent structural model, and the iteration process was performed manually.

1.3.6.2 Dihedral Angle Restraints

Dihedral angle restraints can be estimated from backbone chemical shifts using a program called "torsion angle likelihood obtained from shift and sequence similarity" (TALOS) (Cornilescu, 1999). TALOS predicts protein backbone angles φ and ψ (Figure 9) by evaluating

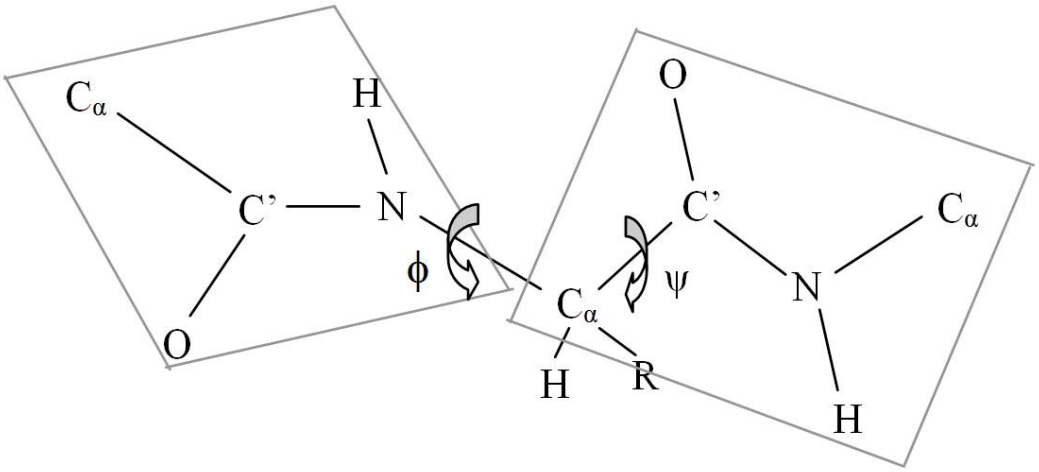


Figure 9. Dihedral ϕ and ψ angles of the protein backbone.

the sequence similarity and secondary chemical shifts (the difference between the chemical shift in the target protein and that in random coil) with those of well-defined high-resolution (2.2Å or better) crystal structures from a protein database. Chemical shift assignments for all amide N, H_α, C_α, C_β and C' are input for TALOS program, and this program not only examines the secondary chemical shifts and amino acid type of a given residue *j* but also those directly before and after it. Subsequently, it searches its database for triplets with similar sequence and residue type, and the ten best matches will be identified. If nine out of ten φ and ψ values of these matches are consistent with each other within certain standard deviations, then their averages and standard deviations will be used as a "good" prediction. Otherwise the matches are deemed "ambiguous", and no prediction can be made for the center residue *j* in a triplet. Further inspection of other data, such as $^1J(C_\alpha H_\alpha)$ (primarily for positive φ angle) or $^3J_{HNHA}$ (primarily for φ angle) is needed to evaluate those ambiguous predictions. Typically, TALOS alone allows a definitive prediction of the φ and ψ angles to be made for about two thirds of the residues.

It should be noted that both dihedral angle restraints and distance restraints are given in a range instead of a fixed number, because chemical exchange or protein motions may diminish the NOE intensities. The dihedral angle restraints derived from TALOS is simply based on a statistical method. The lower limit of any distance is 1.5Å, which corresponds to the van der Waals repulsion range. The intensity of a peak is proportional to the inverse sixth of the separation distance and the mixing time in the preparation period. To get the best information out of an NOESY experiment, a proper mixing time should be chosen. Short mixing times may lead to the loss of some weak NOE's which are mostly long-range NOEs and thus, important for

protein folding, whereas too long mixing times may cause the "spin-diffusion" effect such that NOEs are indirectly generated by spins in the vicinity > 6 Å (Stejskal, 1965). For a protein with the size of rTRD, appropriate mixing times, as a rule of thumb, are in the range of 100-125 msec.

1.3.7 Structure Calculation by Simulated Annealing

Structure calculation is a process of twisting and turning the polypeptide chain until all the distances are satisfied by the resulting model. This process is done by the program suite CNS (crystal and NMR system) (Brunger, 1998). The first step to generate a protein model utilizes the distance and angle restraints in an algorithm known as simulated annealing (Jeremy, 1999). An initial model of the protein, which consists of the correct sequence in a strictly linear polypeptide, was constructed by CNS (see section 2.4.1 for detail). To this, all atoms are assigned random initial trajectories (velocities and directions) according to a Maxwell distribution at a given temperature,

$$\frac{3N}{2} k_B T = \sum_{i=1}^{N} \frac{1}{2} m_i v_i^2$$

where k_B is the Boltzmann's constant, m_i and v_i are the mass and velocity of atom *i*, N is the number of atoms, and 3N is the number of degrees of freedom. The initial temperature is very high, 1,000-3,000 K, at which a large number of conformations of polypeptide interconvert very rapidly just by rotating backbones and side-chains to satisfy most of the restraints, and those conformations that better agree with NMR distance restraints are preferred and will go on to the next step. Thus, the global folding of a protein is identified at the high temperature stage of calculation. Subsequently, many discrete cooling stages are introduced to overcome local minima

and locate the global minimum region of the target function which represents the energy of the protein. The slower the cooling step, the smaller chance the calculated structure ensemble will get trapped in local minima. The final cooling steps place the protein at 0 K, where motions now effectively stop. The potential energy of each structure is evaluated by the following equation (van Gunsteren, 1993):

$E_{total} = E_{bond} + E_{angle} + E_{dihe} + E_{vdw} + E_{coulomb} + E_{experimental}$

where the first 5 terms are empirical energy terms describing the physical interactions between the atoms. They are defined by the following equations:

$$E_{bond} = \sum_{bond} \frac{1}{2} K_b (b - b_0)^2$$

$$E_{angles} = \sum_{angles} \frac{1}{2} K_\theta (\theta - \theta_0)^2$$

$$E_{dihe} = \sum_{dihe} \frac{1}{2} K_\phi [1 + \cos(n\phi - \delta)]$$

$$E_{vdw} = \sum_{pairs(i,j)} [C_{12} | r_{ij}^{12} - C_6 | r_{ij}^6]$$

$$E_{coulomb} = \sum_{pairs(i,j)} q_i q_j / 4\pi\varepsilon_0 \varepsilon_r r_{ij}$$

where b, θ, ϕ, δ are the bond length, bond angle and rotamer angles, b_0, θ_0 are the ideal bond length and angle respectively. K_b, K_θ, K_φ are all force constants (penalty factors). q_i, q_j are the charges of two particles at a distance r_{ij} apart in a dielectric medium described by the ε_0, ε_r terms. $E_{experimental}$ is the sum of potential energy derived from experimental restraints, which initially include the distance restraints and torsion angle restraints. Distance restraints have the effect of pulling the protons involved in an NOE cross-peak within the measured distance range. The

energy potential for this term is a pseudoharmonic function defined as follows (Clore et al., 1986):

$$E_{NOE} = \begin{cases} K_{NOE}(r_{ij} - r_{ij}^{u})^2 & \text{if } r_{ij} > r_{ij}^{u} \\ 0 & \text{if } r_{ij}^{l} \le r_{ij} \le r_{ij}^{u} \\ K_{NOE}(r_{ij} - r_{ij}^{l})^2 & \text{if } r_{ij} < r_{ij}^{u} \end{cases}$$

where r_{ij}^{u} and r_{ij}^{l} are the upper and lower limits of atom *i* to atom *j* obtained from experiment and the van der Waal's limit, respectively. r_{ij} is the actual distance in the current structure model. K_{NOE} is the force constant (penalty factor) of 30-50 KJ/(mol·Å^2)

Similarly, the potential for tortion angle restraints is defined by:

$$E_{tor} = \begin{cases} K_{tor}(\phi_i - \phi_i^{u})^2 & \text{if } \phi_i > \phi_i^{u} \\ 0 & \text{if } \phi_i^{l} \le \phi_i \le \phi_i^{u} \\ K_{tor}(\phi_i - \phi_i^{l})^2 & \text{if } \phi_i < \phi_i^{l} \end{cases}$$

Where ϕ_i^{u} and ϕ_i^{l} are the upper and lower limits of the atom *i* and *j* obtained from experiment, ϕ_i is its actual angle in the current structure model. K_{tol} is the force constant (penalty factor), which is typically on the order of 5-10 kcal/(mol·rad^2). The NMR experimental restraints are expected to be the dominating factor in determining the folding conformation of the molecule; therefore the force constants for the restraints are normally set sufficiently high such that the experimental data determines the final rotations about freely rotating bonds, but are less than the force constants for ideal geometry so that the structure model finally calculated will be composed of standard amino acid residues.

As the process of finding the best fit to experimental data is purely based on random movements, any given structure calculated may be trapped in a local rather than global energy

minimum out of which the random movements were insufficient to escape. Therefore the simulated annealing calculation is repeated several times each with new random trajectories. For example, a total of 30-100 possible structures may be calculated. To eliminate distortions by outliers, some subset of structures (one third to one half) with the best match to the constraints and the lowest energies is selected. These are averaged to obtain the best model at this stage, and then the rmsd's (root-mean-squared-deviations) are calculated for all atoms in the subset according to the following equation:

$$rmsd_{atom\ j} = \sqrt{\frac{1}{n}\sum_{family\ i} |\vec{r_i} - \overrightarrow{R_{av}}|^2}$$

where $\vec{r_i}$ is the actual coordinate vector in one family member and $\overrightarrow{R_{av}}$ is the corresponding vector in the averaged coordinate system. If the rmsd's are sufficiently low, the structure has converged, which means that the current set of restraints is sufficient to uniquely determine a consistent model.

Initially, the models calculated had high rmsd's because there were too few experimental restraints to determine a tight, consistent conformation. But, the overall global folding was approximately correct. The current model was then used to resolve NOE ambiguities and increase the set of restraints. As an example, suppose Val XX NH showed an inter-residue NOE to a methyl group at 1.25 ppm, but there are multiple methyl groups in many residues at this chemical shift. In the current model, the only one close to the Val NH was the methyl of Ala ZZ. This is now so assigned and a new restraint is added. Other ambiguities are similarly resolved as

much as possible and a new family of structures is computed. This is then used to add more constraints and the process is iterated.

The quality of the NMR structure compared to other known proteins is checked by programs called PROCHECK (Laskowski, 1996) and WHATIF (Hooft, 1996), which investigate if local geometries of residues match expected norms. For a consistent and converged structure, the rmsd's for backbone atoms should come out at least 1 Å, and the lower, the better. Values of 0.5 Å or less indicates an excellent result. There should be minimal violation of the experimental constraints. Non-zero but small violation values can be tolerated because of the statistical manner in which the structures are calculated. Standard bond distances, angles, and improper angles (which enforce planarity of aromatic rings and peptide bonds) should be preserved.

2 Materials and Methods

2.1 Preparation of ^{15}N-rTRD and ^{13}C/^{15}N-rTRD Sample

Protein NMR starts with the expression and purification of target protein followed by optimization of its yield and solubility at room temperature so that it will have the highest possible concentration during the entire data acquisition process. The expression and purification procedures for the unlabeled and ^{15}N labeled rTRD were originally designed by Dr. Simmonds' group (Simmonds, 1996). Their procedure for the purification of ^{15}N-labeled protein was modified in this work in order to maximize the yield of ^{15}N-rTRD and ^{13}C/^{15}N-rTRD.

2.1.1 Expression and Purification of ^{15}N labeled and ^{13}C/^{15}N Labeled rTRD

2.1.1.1 Original Method for Expression and Purification of ^{15}N Labeled rTRD

^{15}N-labeled rTRD was produced by growing 1 L of *E. coli* M15 (containing the pQETRD plasmid) cells to an OD_{600} of approximately 0.7 at 37 °C in Luria Broth (LB) medium (Table A1 of Appendix I). Collected cells were washed with basic M9 minimal medium (Table A2 of Appendix I) for three times and resuspended into a 250 mL of ^{15}N-labeled basic M9 minimal medium (with 0.33 g of 86% ^{15}N-enriched $^{15}(NH_4)_2SO_4$ added), followed by adding sterile ingredients including 25 µL of 1M $CaCl_2$, 0.5 mL of 1M $MgSO_4$, 12.5 mL of 20% glucose, 250 µL of 100 mg/mL of ampicillin and 250 µL of 25 mg/mL of kanamycin. The final concentrations

of the above sterile ingredients were 0.1 mM, 2 mM, 5 g/L M9, 1 mM and 1 mM respectively. The cells were cultured for an additional one hour before adding isopropyl-beta-D-thiogalacto-pyranside (IPTG) to induce the expression of rTRD protein. After IPTG was added, the cells were cultured for another 4 hours and then collected by centrifuge at 6000 rpm in GSA rotor for 30 min and resuspended in 40 mL of lysis buffer (Table A3, Appendix I). Then the cells were disrupted by sonication followed by lysing with 1 mg/mL lysozyme for 2 hr. The lysate was centrifuged at 12,500 rpm in SS-34 rotor for 20 min, and the supernatants were filtered through a Millipore Express® 0.22 μm polyethersulfone (PES) filter and loaded onto a Qiagen nickel nitrilotriacetic acid (Ni-NTA) agrose metal affinity column. Proteins without hexa-His tag could not bind to the resin and were then eluted away with 45 mL of lysis buffer (Table 3, Appendix I) and 60 mL of medium ionic strength wash buffer (50 mM imidazole); the hexa-His tagged rTRD was eluted from the beads with 60 mL of elution buffer (Table A3, Appendix I). Flow-through was collected in five 15 mL sterile tubes. The purity of all flow-through fractions was analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) using previous protocol (Laemmli, 1970). Fractions containing rTRD were combined and dialyzed against 10 mM potassium phosphate buffer at pH 7.0 three times at 4-hour interval. After dialysis, the concentration of the protein was determined based on the calculated extinction coefficient at 280 nm, which is 25 $mM^{-1}.cm^{-1}$ for rTRD (Liang, 2004). Finally, rTRD was concentrated by lyophilization or ammonium precipitation.

2.1.1.2 Optimization of Expression and Purification of $^{13}C/^{15}N$-Labeled rTRD

$^{13}C/^{15}N$-labeled rTRD can be prepared in the same way as ^{15}N-labeled rTRD except that ^{13}C-

glucose is used instead of natural abundance glucose. With the original method, the yield for rTRD was in the range of 6-10 mg per liter of M9 medium using 5 g of ^{13}C-glucose, which would result in a maximum sample concentration of 1.4 mM ^{13}C/^{15}N-TRD sample if there is no protein loss during protein concentration. However, protein loss during concentration was unavoidable and non-negligible, and ^{13}C-glucose was expensive when these experiments were carried out. Therefore, efforts were made to prepare an optimized NMR sample with low cost and high concentration.

First, protein yield was maximized. A series of experiments were carried out to optimize the ingredients of M9 medium (Table 2). In each experiment, 10 mL of basic M9 medium with 10 mM of ^{15}N labeled $(NH_4)_2SO_4$ was added into a 18 mm diameter test tube and autoclaved for 25 min. Other sterile ingredients were added to each tube according to Table 1, followed by the inoculation of freshly-plated rTRD cells into each test tube. Glucose is a necessary nutrient for *E.coli*, thus each test tube contains the same amount of glucose (0.1g). The first 5 batches were used to test the effect of Cai trace elements (Cai, 1991) on cell production. Batches 6 & 7 were used to compare the effect of Basal Medium Eagle (BME) with thiamine (vitamin B1). Batches 8 & 9 were used to investigate the effect of BME or thiamine on cell production with Ca^{2+} and Mg^{2+} present. Batches 10 and 11 were designed to test the effect of different amounts of Ca^{2+} and Mg^{2+} on cell production. The cell densities of the above media were monitored by measuring the optical density at 600 nm after 12 hrs culture.

No.	**Glucose**	**$CaCl_2$ (1M)**	**$MgSO_4$ (1M)**	**Cai trace element**	**BME**	**Thiamine (6g/L)**
1	0.1g			5μL		
2	0.1g			10μL		
3	0.1g			20μL		
4	0.1g			40μL		
5	0.1g			80μL		
6	0.1g			20μL		20μL
7	0.1g			20μL	0.1mL	
8	0.1g	1μL	20μL	20μL		20μL
9	0.1g	1μL	20μL	20μL	0.1mL	
10	0.1g	1μL	20μL	20μL	0.1mL	20μL
11	0.1g	2μL	40μL	20μL	0.1mL	20μL

Table 2. The optimization of the M9 medium compositions for ^{15}N-rTRD production.The ingredients of Cai trace element solution are listed in the table 5 of appendix I. Each ingredient must dissolve one by one, and the normal color of this solution is golden brown.

The length of post-induction time during over-expression was varied to maximize the protein yield. Cells were prepared in the same way as in the original protocol for ^{15}N-TRD except using an optimized M9 minimal medium. After IPTG was added to induce the protein expression, 25 mL of the cell suspension was sampled after 4, 5, 6, 7, 8, 9, 19, 20, 23, 24, 25 hours of post-induction-time respectively, and the corresponding cell density was measured.

Second, the cost of sample preparation was reduced by minimizing the usage of ^{13}C-glucose. With all other conditions being the same as the optimized protocol (table A4 of appendix I), the amount of glucose was varied (5 g, 4 g, 3 g, 2 g and 1 g per 500 mL of M9 medium), and the protein yields were compared.

Finally, the protein loss during concentration process was minimized. Even if an appropriate yield of target protein can be obtained by the optimization of the expression and purification procedure, losses during concentration can limit the NMR sample concentration. Three methods are generally used to concentrate protein: precipitation with ammonium sulfate, lyophilization, and concentration by ultrafiltration. The protein yields using these three methods were compared and will be discussed under results and discussion.

2.1.2 NMR Data Acquisition and Processing

All NMR experiments were performed on Bruker AV600 MHz NMR spectrometer equipped with an inverse ^{1}H/^{13}C/^{15}N triple resonance probe with pulsed field gradient capabilities. The following experiments were performed: ^{1}H-^{15}N HSQC, HNCACB, HNCA, HN(CA)CO, HNCO, HN(CO)CA, HNHA, CCC-TOCSY, HCCH-TOCSY, ^{15}N-edited TOCSY-HSQC, 3D ^{15}N-edited NOESY-HSQC (120 ms mixing time), and ^{13}C-edited NOESY-HSQC (120 ms mixing time).

Two-dimensional Hβ(CβCγCδCε)Hε and Hβ(CβCγCδ)Hδ experiments were performed on an unlabeled 2.1 mM rTRD sample. Amide exchange experiments were performed with a 1.4 mM ^{15}N-labeled TRD in 10 mM phosphate buffer, pH 7.0. Amide protons that were still detectable after 4 weeks were assumed to be involved in hydrogen bonding. The ^{15}N-labeled TRD was 2.1 mM in 10 mM phosphate buffer, pH7.0. Spectra were processed with NMRPipe (Delaglio, 1995) using the following procedure: 1) raw FID data were processed by solvent suppression, multiplied by window functions to enhance resolution and avoid truncation, zero filling, and linear prediction; 2) after Fourier transformation, phase and baseline corrections were used. Processed spectra were analyzed using the free software NMRView (Johnson, 1994), PIPP (Garrett, 1991), or NMRDraw (Delaglio, 1995). ^{1}H chemical shifts were referenced nominally with respect to external DSS (2,2-Dimethyl-2-silapentane-5-sulfonic acid) in D_2O, and ^{13}C and ^{15}N chemical shifts were referenced indirectly by means of the absolute proton transmitter frequency by the automatic chemical shift referencing algorithm of NMRPipe.

2.2 Example of Resonance Assignments and Structure Determination for rTRD

2.2.1 Sequential and Side-chain assignments

An ^{15}N-HSQC spectrum of rTRD was examined first as it shows proton resonances of every amide group in the protein (Figure 10). Besides the backbone amide proton resonances, the aromatic side-chain amide proton of Trp residues (circled in red), and the side-chain protons of NH_2 group in Asn and Gln (connected by blue green dotted horizontal lines) also appear in the spectrum. However, it should be noted that none of these could be specifically assigned at this stage. Backbone resonance assignments start with a HNCACB spectrum. Figure 11 shows strip

plots of 5 consecutive residues in the HNCACB spectrum of rTRD. Each strip shows the C_α and C_β of the i^{th} amino acid residue and those of the $(i\text{-}1)^{th}$ amino acid residue at the slice where amide H and N chemical shifts of the i^{th} residue are located. In most cases, cross-peaks corresponding to C_α and C_β of residue *i* are more intense than those of residue *i*-1. C_α and C_β always have the opposite signs due to the way the magnetization is transferred, and they are indicated as black and red contour lines, respectively. Aligning the weak peaks in one strip with the strong peaks from another establishes that the corresponding backbone N-H pairs are adjacent in the sequence. According to Table 1, the characteristic chemical shift of C_β at 64 ppm in strip #2 (circled in red) suggests that the residue *i* on this strip is a Ser (S) residue. The characteristic chemical shift of C_β at 19 ppm in strip #4 (circled in red) suggests that the residue *i* on this strip is an Ala (A). Thus a short sequence with "X S X A X" pattern (X is a residue type that cannot be predicted from this data only) can be mapped onto the sequence of TRD, and the sequence from residues 101 to 105 matches this pattern. In cases where the intensities of C_α from residue *i* and that from residue *i*-1 in HNCACB were close, HNCA and HN(CO)CA spectra were used to resolve the ambiguity.

As for the missing or overlapped resonances in the HNCACB and HNCA spectra, correlations via other types of nuclei, such as HN(CA)CO and the HNCO experiments were used to solve the ambiguitiy. These two experiments are independent alternatives for checking the sequential connectivities by correlating the frequency of an amide proton with the frequencies of the intra and/or interresidual carbonyl resonances. Figure 12 shows strip plots of 5 consecutive residues in the HN(CA)CO experiment. The more intense carbonyl resonances correspond to residue *i*.

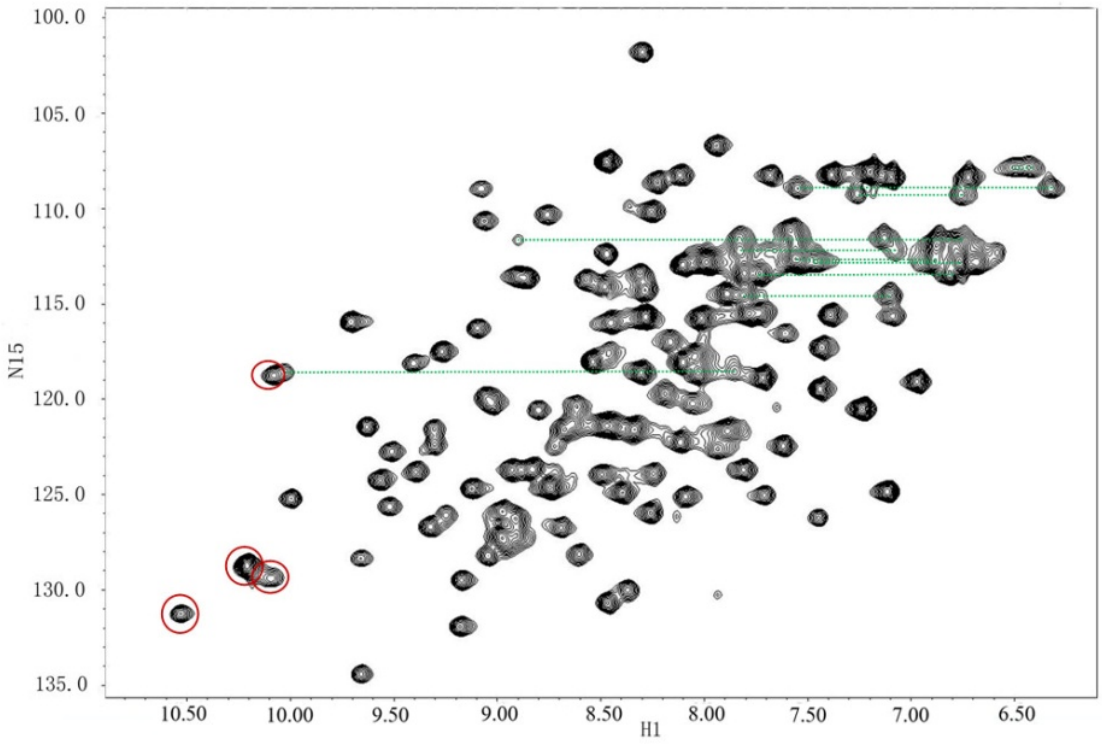


Figure 10. ^{1}H-^{15}N HSQC spectrum of 2.4 mM $^{13}C/^{15}N$ labeled rTRD in 10 mM potassium phosphate/5% D_2O, pH 7.0 acquired at 600 MHz and 298K. Peaks in red circles were assigned to the aromatic NH of Trp residues. Peaks corresponding to side-chains of the same Asn and Gln residue are connected by horizontal lines.

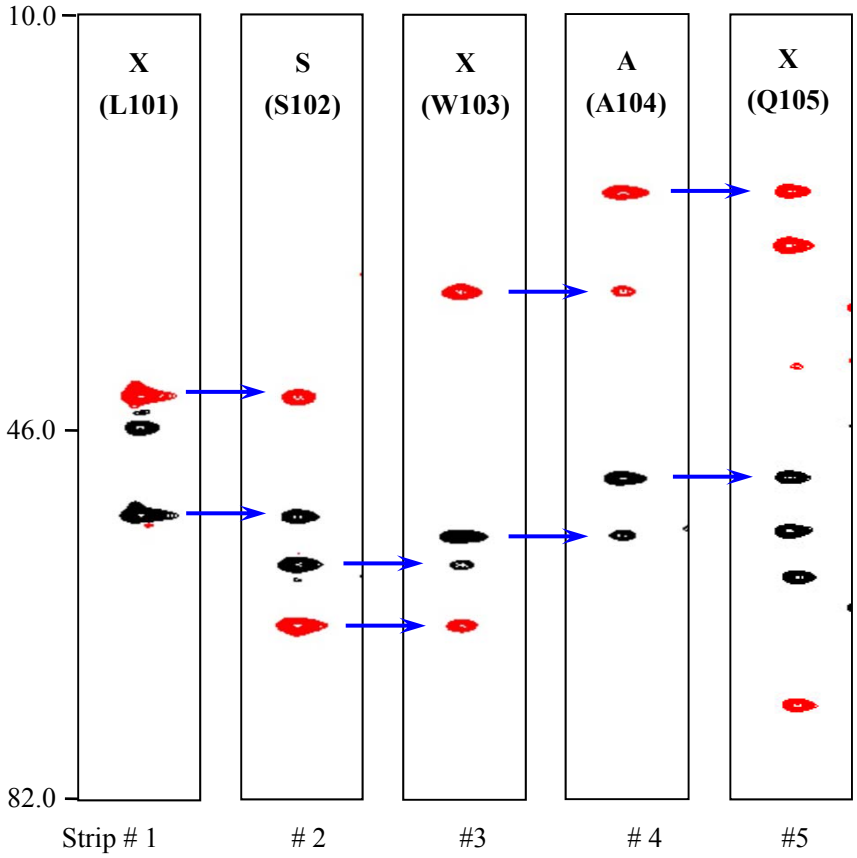


Figure 11. Strip plots of residues 101-105 of rTRD in a HNCACB spectrum. C_α resonances are black contours, and C_β resonances are red contours. The intense set of C_a and C_β belongs to residue *i* and the weaker set of C_α and C_β belongs to residue *i*-1. The more intense red peak in strip #2 is the C_β of a serine, while the more intense red peak in strip #4 is the C_β of an alanine. X represents an amino acid residue that cannot be assigned to a specific type from this data alone.

When intensities of carbonyl resonances from residue *i* and *i*-1 were close, the HNCO experiment, which only correlates amide N and H of residue *i* with C' of residue *i*-1, was used to distinguish the C' of residue *i* from that of residue *i*-1.

Aliphatic side chain resonances were assigned by CCC-TOCSY, HCC-TOCSY and HCCH-TOCSY spectra. CCC-TOCSY experiments were carried out with mixing time of 15 ms, 21 ms, and 25 ms. An example of the assignment of the CCC-TOCSY is shown in Figure 13A. For residue L102, the amide N and amide H chemical shifts had been identified from backbone assignment, which were 115.61 ppm and 7.98 ppm, respectively. According to the correlations described in section 1.3.3.2, C_α, C_β, C_γ, and C_δ of residue 101 were identified at the slice where the chemical shift of amide N and H of residue L102 are located, and the characteristic chemical shift table (Table 1) were used to assign these carbons specifically.

An example of HCC-TOCSY spectra is shown in Figure 13B, which displays all H_α, $H_{\beta1}$, $H_{\beta2}$, H_γ, $H_{\delta1}$ and $H_{\delta2}$ crosspeaks. The same chemical shift assignment strategy as in CCC-TOCSY was utilized except that the chemical shifts for side-chain ^{1}H's were detected. A HCCH-TOCSY spectrum (Figure 13C) was also recorded to assign side-chain protons. For example, with the C_α chemical shifts obtained from CCC-TOCSY, the directly attached ^{1}H should show up at its corresponding carbon slice as diagonal peak, and all other coupled protons as cross-peaks.

Two-dimensional Hβ(CβCγCδ)Hδ and Hβ(CβCγCδCε)Hε experiments were used to assign the aromatic ^{1}H's of Phe, Tyr and Trp. The observed H_δ and H_ε chemical shifts were also listed in Appendix II.

2.2.2 $^3J_{HNHA}$

The $^3J_{HNHA}$ coupling constants were derived from the intensity ratio between the red HA cross-peak and the black diagonal HN peak (Figure 14) using the following equation (Geerten and Bax, 1993):

$$\frac{I_{cross}}{I_{diag}} = -\tan^2(2\pi^3 J_{HNHA}\zeta)$$

where I_{cross} and I_{diag} are the intensities for the cross-peaks and diagonal peaks, respectively, and ζ is a delay constant, which was 13.05 ms. $^3J_{HNHA}$ coupling constants of rTRD were computed automatically by NMRView using the calibration factor 0.9 and noise background $7.4X10^6$. These $^3J_{HNHA}$ coupling constants can be used as conformational restraints directly by CNS software, and they can also be mapped into the following Karplus equation to compute the dihedral ϕ angle (Karplus, 1959, 1963):

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

where A, B and C are constants and the optimized values for proteins are 6.98,-1.38, and 1.72, respectively.

2.2.3 Stereo-Specific Assignments for Methyl and Methylene Groups

Stereo-specific assignments for most of the Glu and Asn's NH_2 groups were made based on the observation that the chemical shift of $H_{\varepsilon 21}$ in Gln (or $H_{\delta 21}$ in Asn) is typically higher than than that of $H_{\varepsilon 22}$ (or $H_{\delta 22}$). The nomenclature of amino acid is based on the IUPAC recommendation. The stereo-specific assignment for Val's methyl groups and Leu's methylene groups were made using methods previously proposed (Cai, 1991).

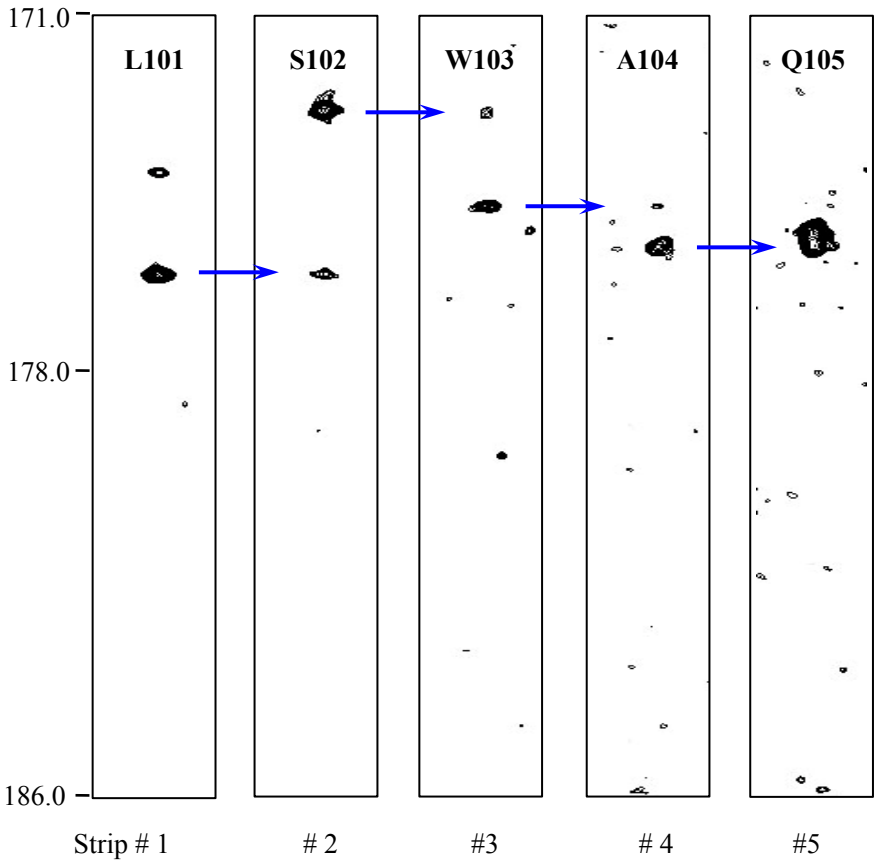


Figure 12. Strip plots of residues 101-104 in HNCACO experiment. The intense C' peak in each strip belongs to residue *i* and the weak peak belongs to residue *i*-1.

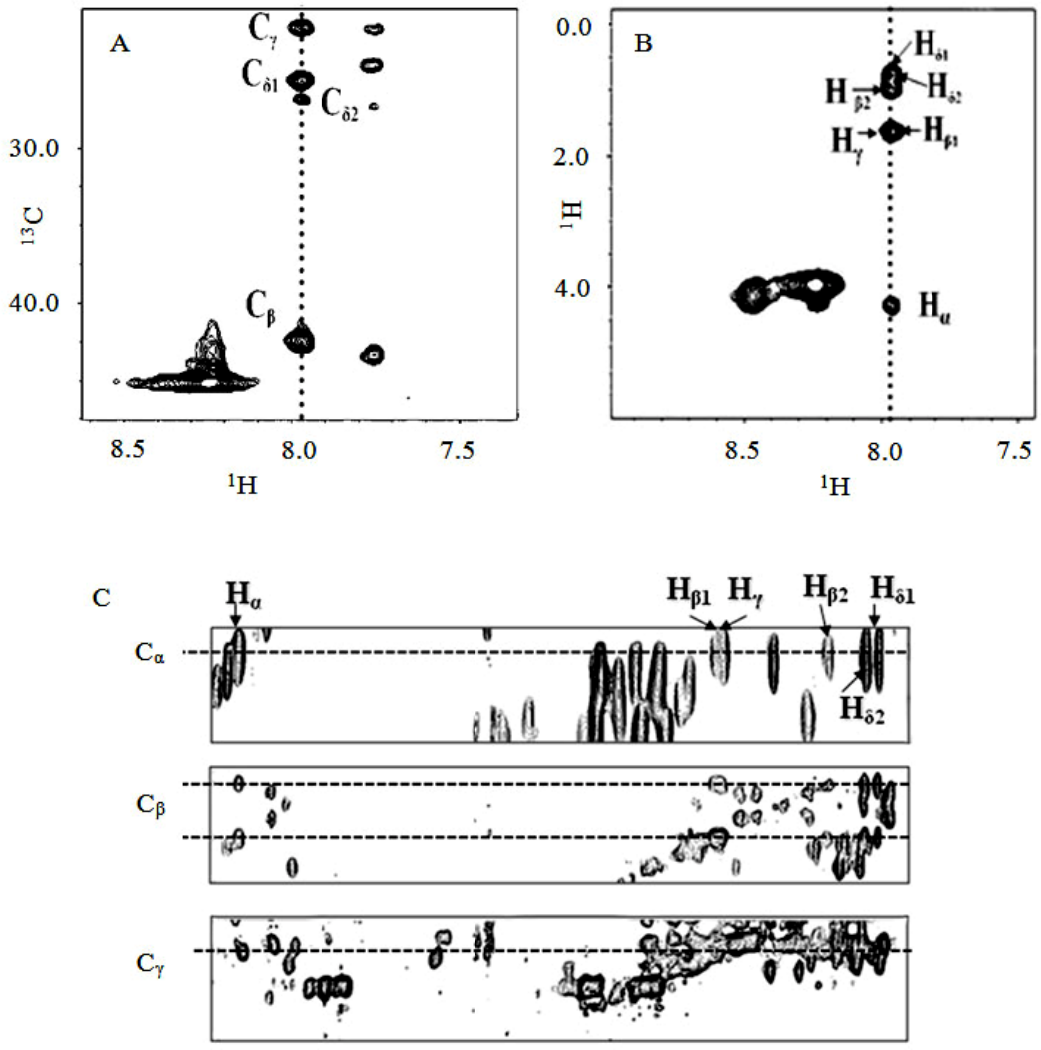


Figure 13. Experiments for side-chain assignments. A) A selected ^{13}C-^1H slice from CCC-TOCSY spectrum of rTRD. The side-chain carbon chemical shifts of L101 are shown at the slice where the chemical shifts of amide H and amide N are 7.98 ppm and 115.61 ppm, respectively B) A slice of HCC-TOCSY spectrum of rTRD. The side-chain carbon chemical shifts of L101 are shown at the slice where the chemical shifts of amide H and amide N are 7.98 ppm and 115.61 ppm respectively. C) Selected regions from ^{13}C slices of a 22.5 ms mixing time 3D HCCH-TOCSY of rTRD, showing correlations originating from the H_α, H_β, and H_γ of Leu 101.

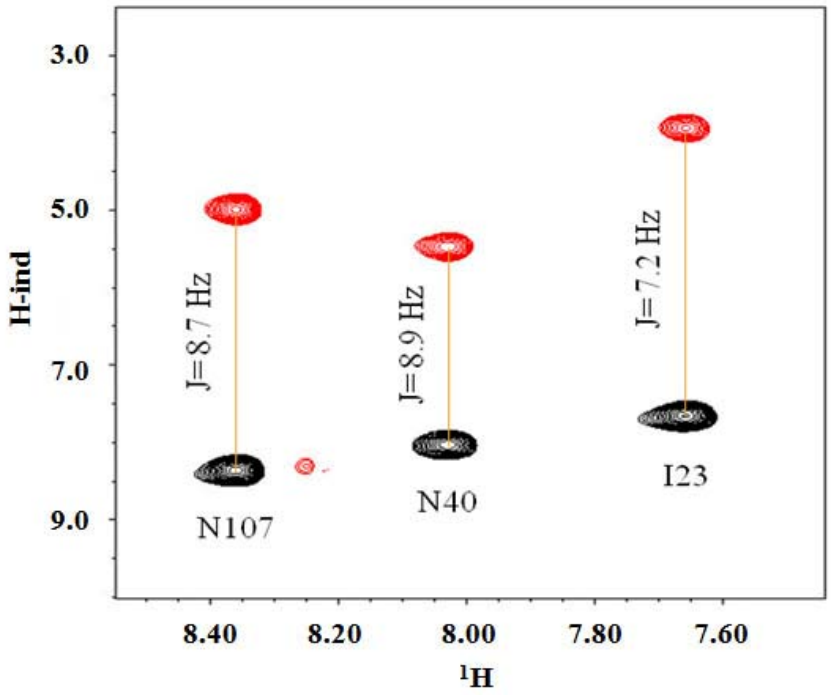


Figure 14. A selected $^{1}H\text{-}H^{N}$ slice, at a ^{15}N chemical shift of 118.9 ppm , from a HNHA spectrum of ^{15}N-labeled TRD acquired with δ=4.5 ms and $2\delta_2$=26.1 ms. The intraresidue peaks are labeled with their residue names. The cross peaks have phase opposite to that of the diagonal peaks. From the cross-peak to diagonal peak intensity ratios, the following J coupling constant values were obtained by NMRView: N107 8.7 Hz; N40 8.9 Hz; I23 7.2 Hz.

2.2.4 Preliminary Structure Calculation by CNS

The structure calculation was performed on a LINUX-based computer using the CentOS operating system using CNS 2.1. Generally, default CNS parameter values were used, but the simulated annealing protocol was the one used previously for the solution structure of cytochromes c-551 (Timkovich, 1998). Simulated annealing calculations were initially performed by CNSsolve software suite using 935 manually assigned and unambiguous NOEs involving amide proton to all other protons and 189 TALOS-derived dihedral angle restraints as conformational restraints. Multiple structures were computed starting with an extended protein chain. The average of the 5 lowest energy structures resulting from the calculations was used as an initial model to solve NOE restraints ambiguities.

2.2.5 NOE Violations Analysis and Removal of Incorrect Distance Restraints

In the preliminary structure calculation, the global folding of the protein was modeled although the resolution was low, and the statistics were not good. The generated structures contained violations of distance and torsion angles. A distance violation occurs when the difference between the inter-proton distance in the calculated structures and the distance bounds estimated from the observed NOE intensity is greater than the upper limit or less than the lower limit by 0.5 Å. Some violations in a few of the calculated models may be caused by a lack of restraints in parts of the structure. Consistent violations (violated in all calculated structures) can be caused by mis-assignment of NOEs or incorrect NOE intensity measurements, often caused by overlapping peaks. By removing the incorrect or mis-assigned NOE restraints, the quality of the structure was improved.

2.2.6 Iterative Structure Calculations for Model Refinement

After the global folding was obtained, NOE restraints were added iteratively. NOE's involving H_α to H_α were added first as they are always strong and diagnostic of β-sheet secondary structure. The next type of NOE involves H_α to all other protons, followed by NOE's involving methyl protons to all other protons, and finally NOE's involving protons with unusual chemical shifts. New structures were calculated at each step, and the number of violated NOE's decreased. Previously ambiguous NOE restraints were assigned based upon the improving structural model. A total of 100 structures were calculated, and the 10 structures with lowest energies were selected and superimposed for a visual display. The quality of resulting structure ensemble was evaluated using PROCHECK (Laskowski et al., 1996). Structure statistics were based upon the 50 lowest energy structures out of 100 calculated structures.

3 Results and Discussion

3.1 Expression and Purification of $^{13}C/^{15}N$ Labeled rTRD

The protein yield of unlabeled rTRD could reach 30 mg per liter of LB medium. However, the yield of ^{15}N- or $^{13}C/^{15}N$-labeled TRD decreased substantially to 10 mg per liter of M9 medium due to the poor nutrition source. This could result in a high cost of preparing $^{13}C/^{15}N$-labeled rTRD samples, since ^{13}C-enriched glucose is expensive. To reduce cost, experiments were carried out to optimize the M9 minimal medium composition for ^{15}N-labeled rTRD so that the protein yield could be maximized. The cell densities for each experiment were shown in Table 3. Results showed as follows. 1) For media 1-5, increasing the amount of Cai trace elements improved the cell production to some extent, but there was no significant improvement on the cell density when the amount of Cai trace elements was larger than 20 μL. 2) The cell density increased from ~0.08 (Table 3, medium 3) to 0.586 (Table 3, medium 6) and 0.734 (Table 3, medium 7) with the addition of thiamine and BME, respectively, therefore, the addition of BME gave a better yield than the addition of thiamine. 3) Ca^{2+} and Mg^{2+} gave better growth compared to the media without them (Table 3, media 8 and 9), and the cell density was further increased when both BME and thiamine were added (Table 3, media 10). 4) Doubling the concentration of Ca^{2+} and Mg^{2+} led to the highest cell density (Table 3, medium 11). The M9

medium used for growing $^{13}C/^{15}N$-TRD was prepared according to medium 11 (Table 3). Figure 15 showed the effect of post-induction time on the cell density. The highest cell density appeared at 21 hrs post-induction time, and the cell density did not change significantly beyond this post-induction time. In the ^{13}C-glucose usage minimization experiment, 2 g of glucose could yield 24-38 mg per liter of M9 medium, which was about the same amount of protein produced from 5 g of glucose with the original, unoptimized medium.

Protein concentration is an important step in preparing NMR samples. Even if high protein yield can be obtained through the optimization of protein expression and purification, the final concentration of an NMR sample can be low if improper concentration methods had been chosen. Here, three protein concentration methods were compared, and it was found that a 25% recovery rate was obtained by using ammonium sulfate precipitation, 87% by lyophilization, and 88% by ultrafiltration using an Amicon pressure device followed by a Centricon-3 centrifugal concentrator. The last two methods showed comparable recovery rate, but lyophilized TRD powder could not be completely redissolved. Therefore, $^{13}C/^{15}N$-labeled TRD was concentrated using ultrafiltration.

3.2 Sequential Resonance Assignment

With the assignment strategy discussed in "Materials and Methods", more than 90% of the backbone ^{1}H, ^{13}C and ^{15}N resonances were assigned sequentially. The resonances for six histidine residues H5-H10 in the N-terminal were unresolved and degenerate, which is common in NMR structure determinations for hexa-His tagged proteins (Gaspar, 2005, Singh, 2002). Therefore, the following discussion will not include this N-terminal hexa-His-tag. Backbone

No	Glucose	$CaCl_2$ (1M)	$MgSO_4$ (1M)	Cai trace element solution	BME	Thiamine (6g/L)	OD_{600nm}
1	0.1g			5μL			0.032
2	0.1g			10μL			0.040
3	0.1g			20μL			0.078
4	0.1g			40μL			0.077
5	0.1g			80μL			0.080
6	0.1g			20μL		20μL	0.582
7	0.1g			20μL	0.1mL		0.734
8	0.1g	1μL	20μL	20μL		20μL	0.596
9	0.1g	1μL	20μL	20μL	0.1mL		0.770
10	0.1g	1μL	20μL	20μL	0.1mL	20μL	0.815
11	0.1g	2μL	40μL	20μL	0.1mL	20μL	0.874

Table 3. The optimization of the M9 minimal medium compositions for ^{15}N-rTRD production (result).

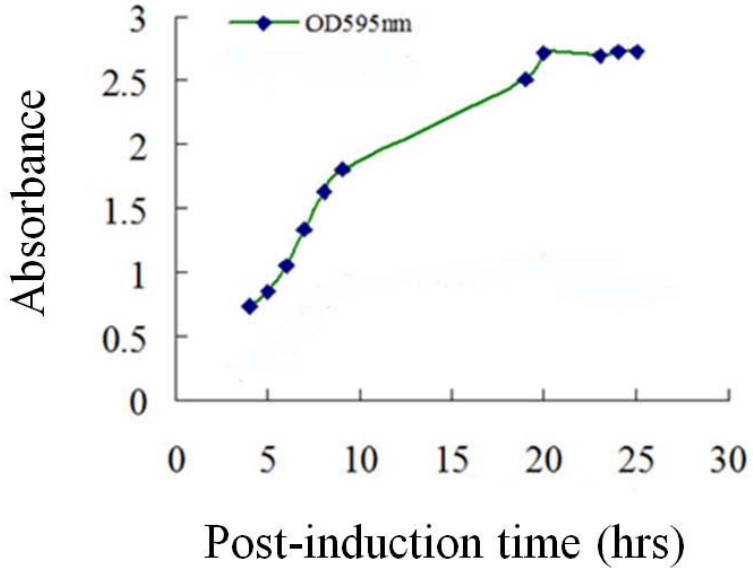


Figure 15. Cell density of ^{15}N-TRD measured at 595 nm versus the post-induction time.

amide NH assignments were made for all the non-proline residues except for residues M1, R2, N33, S71, and G99, which were insufficiently separated from other resonances for identification. All these residues are located at the N-terminus or in turn regions.

The CCC-TOCSY spectrum resolved almost all of the side-chain carbon chemical shifts except for the C_γ of T15, T16, T18 and T19, all of which are located in the N-terminal region. All the side-chain proton chemical shifts except for the H_α's of G11, H_δ of K25, and $H_{\varepsilon 2\#}$ of Q73 were obtained from HCC-TOCSY and HCCH-TOCSY spectra. After backbone and sidechain assignment, chemical shifts for the aromatic ring ^1H's, including all H_δ and H_ε of Trp, Tyr, and Phe residues, were obtained from 2D Hβ(CβCγCδ)Hδ and Hβ(CβCγCδCε)Hε experiments.

Interestingly, some unusual chemical shifts have been observed. First, in the HNCACB experiment, the chemical shifts of the sidechain $N_\varepsilon H_\varepsilon$ group of R39 were observed (circled in Figure 16A), whereas in most proteins this exchanges too rapidly with solvent to be observed. Additionally, the HNCA experiment provided chemical shifts for the side-chain NH and $C_{\varepsilon 1}$, and the HNHA experiment gave chemical shifts for the side-chain NH, and $H_{\varepsilon 1}$ of each Trp residue (circled in Figure 16B).

3.3 Flexibility of N-terminal Tail

No interresidue NOE's were identified for the N-terminal residues from 1 to 20, indicating that the N-terminal is probably poorly-structured in solution. Random coil index software (RCI) was used for N-terminal flexibility analysis (Berjanskii, 2005). It is the inverse weighted sum of the observed C_α, CO, C_β, N and H_α secondary chemical shifts (the chemical shift difference between measured and that of random coil). The RCI web server is available at http://wishart.

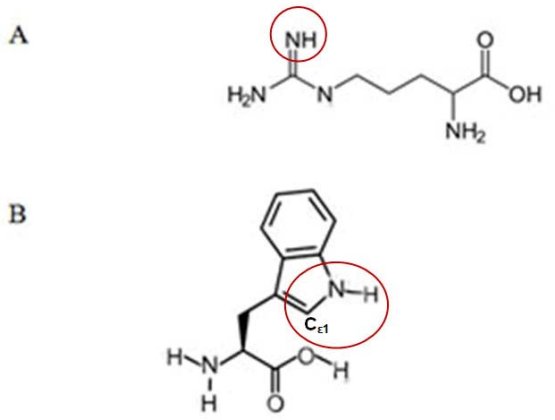


Figure 16. Special chemical shifts observed during the sequential assignment process.. A) N_ε and H_ε of Arg (circled) were observed in the HNCACB experiment. B) N_ε, $C_{\varepsilon 1}$ H_ε and H_δ of Trp amino acid (circled) were observed in HNCA and/or HNHA experiments.

biology.ualberta.ca/rci. It allows site-specific mapping of protein backbone flexibility based on an empirically derived relationship between secondary chemical shifts (see definition in section 1.3.6.2) and protein flexibility. In TRD, predictions couldn't be made for residues H6-H10 and G11 due to the lack of chemical shift information. RCI values of the rest of the N-terminal residues are much higher than the rest of the amino acid residues (Figure 17), indicating that the N-terminal of rTRD is a disordered tail.

3.4 Structural Restraints

3.4.1 Dihedral Angle Restraints

The backbone dihedral angle restraints were generated by the software TALOS (Cornilescu, 1999). Only "good" predictions (see definition in section 1.3.6.2) were used in structure calculations with the errors assigned to them by TALOS itself. In this protein, a total of 185 such dihedral angle restraints were used, and "ambiguous" dihedral φ angles (see definition in section 1.3.6.2) were re-evaluated by analysis of the 3D HNHA spectrum.

3.4.2 Hydrogen Bond Restraints

Since hydrogen bond networks stabilize protein secondary and tertiary structures, they are important for protein folding. NMR experiments have been developed to directly detect hydrogen bonds (Banci, 2005), but most of the H-bond restraints are still indirectly derived from amide exchange experiments, which are simple and convenient to perform. These H-bond restraints are normally used for structure refinement after the initial structure is calculated. Figure 18 is the amino acid sequence of rTRD showing residues with slowly exchanging backbone amide protons. Resonances still present in the TOCSY spectrum one month after the

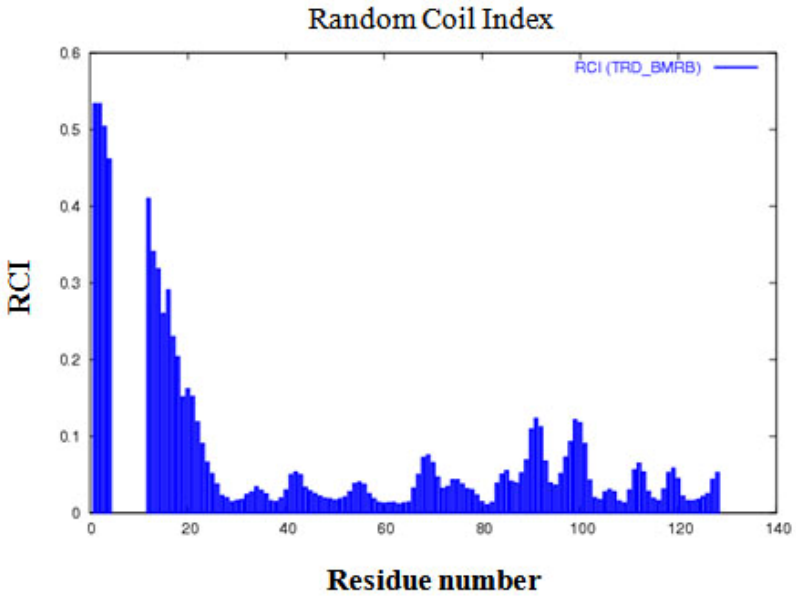


Figure 17. The predictions made by RCI. The gaps between residues 5 and 11 contain H6-H10 and G11 in the N-terminal sequence.

protein was dissolved in 99.8% D_2O are indicated with an asterisk (*). Secondary structure elements were mapped onto the primary sequence. With the exchange information, hydrogen bond donors and acceptors were identified in the secondary structure elements and NOE distance restraints were manually assigned to impose these hydrogen bonds. Each hydrogen bond in beta sheet was assigned a donor hydrogen and acceptor oxygen distance (1.7-2.4 Å), and a donor nitrogen and acceptor oxygen distance (2.7-3.2 Å). Hydrogen bonds in alpha helices were assigned donor hydrogen to acceptor oxygen distance (1.7-2.4 Å) only (Turner and Moore, 2004).

3.4.3 Distance Restraints

Obtaining distance restraints from NOESY assignments is the most laborious and complicated step in structure determination. This step can sometimes be automated by software as with ARIA. This was successfully used for the structure determination of *Pseudomonas* cytochrome mutant H47A in our group (Liang et al., 2007), but failed to generate a consistent structure model for rTRD. Thus the entire assignment process was done manually. It must be stressed that conceptually, what ARIA tries to do is the same as the following manual process, but without human error checking and decision making in the process. The process started with ^{15}N-edited NOESY spectra, which contains NOE's involving backbone amide protons only. Unambiguous NOE's are those between protons with distinct chemical shifts or between intra-residue protons which are expected to show the NOE effect because of a standard local geometry. For example, the backbone NH proton frequently shows an NOE to the α proton of the same residue with only the intensity of the effect depending on the separation distance. Unambiguous NOE's were used in simulated annealing computations to obtain an initial structural model. From this, previously

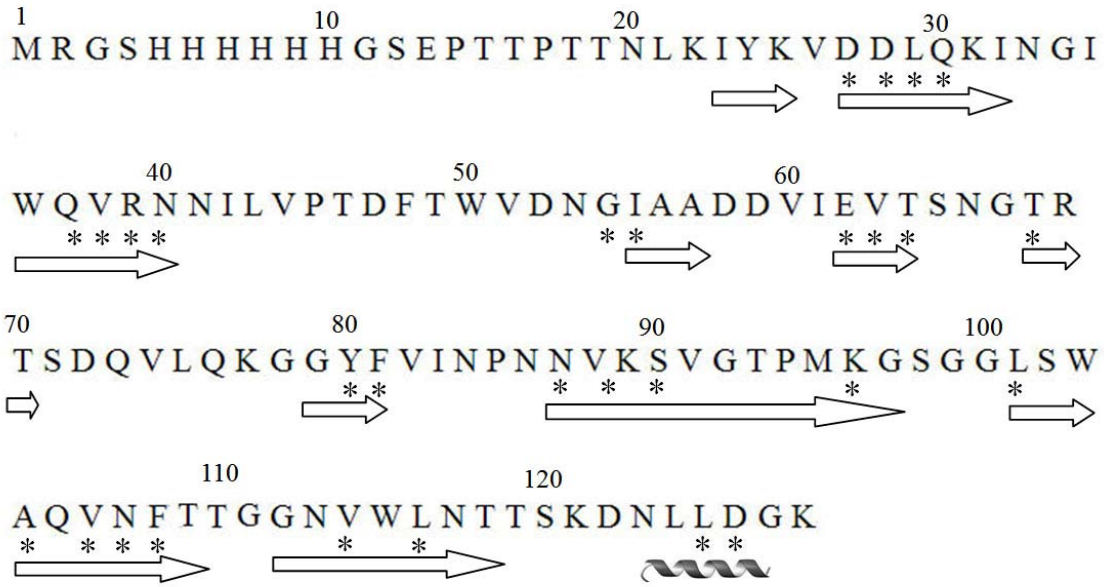


Figure 18. The amino acid sequence showing the the residues with slowly exchanging backbone amide protons. Resonances still present in the amide exchange experiment are indicated with asterisk (*). Secondary structure elements (β-sheets shown as arrow, α-helix shown as curved lines) are mapped onto the primary sequence as determined by the data in the final structures.

ambiguous NOE's could be assigned based on the current model. With the expanded restraint set a better model was calculated. Here "better" means no systematic violations of restraints, and the family of calculated structures shows better convergence. The ^{13}C-edited NOESY spectrum was examined next, and cross-peaks were assigned based upon their uniquesness and consistency with the previous model. Then another round of simulated annealing calculations produced a more refined model which was used to assign additional cross-peaks, and so on. Through this iterative procedure, 1823 non-redundant distance constraints were obtained. For all proline residues, a *trans* conformation was identified by the strong NOE's between the H_δ of Pro to the H_α of the previous residue.

3.4.4 Other Restraints

Additional experimental restraints were used to produce a more converged model. The three bond coupling constant $^3J_{HNHA}$ associated with the backbone phi angle can be measured from the HNHA experiment and incorporated as an additional class of experimental restraints (Garrett, 1994). C_α and C_β chemical shifts are sensitive to the folding environment, and can be matched against a database of such shifts in other proteins to provide another class of restraints (Kuszewski and Gronenborn, 1995). Proton chemical shifts can be treated the same way to add another class of experimental restraints (Kuszewski and Gronenborn, 1995). These are usually insufficient early in the analysis to determine the global folding, but when added in the later stages, they contribute toward convergence of the model.

3.5 Structure Calculations

Global folding of the rTRD was identified by using a set of 935 NOEs and 185 TALOS

derived dihedral angles. The average of the 5 lowest energy structures out of 50 calculated structures is shown in Figure 19. The current best models were calculated based on 1823 NOE distance restraints with 147 medium-range and 336 long-range restraints (see definition in section 1.3.6.1), 185 backbone dihedral angle restraints, 106 Hα chemical shifts restraints, 121 C_α/C_β chemical shift restraints, 93 J (phi) restraints, and 68 hydrogen bonds. The total number of conformational restraints is 2464, resulting in 20 constraints per amino acid residue, which is enough for a good quality structure. One hundred structures were calculated from the above restraint package, and 50 with the lowest total energy were further analyzed.

3.6 Presentation and Evaluation of the rTRD Structure

Violations and root mean squared displacements of atomic positions among the family (rmsd's) are the most important criteria for a good converged structure ensemble. Both must be small for a well converged structure. If the structure ensemble has few violations but the rmsd's are high, there are not enough constraints to restrain the structure. If the rmsd's of a structure ensemble are low but with a large number of violations, there must be some wrong assignments, most likely in the NOE restraints. Statistics for the rTRD determination are given in Table 4. The disordered random coil tail of residues 1-20 was excluded from the atomic coordinate rmsd calculation. Residues 71-74 were also excluded. Very few NOE's were observed for these residues suggesting, but not proving, that they also might be disordered in solution. Ramachandran plot of the average TRD structure was shown in Figure 20.

A stereo-view of the superposition of the 10 lowest energy structures out of 100 calculated structures is shown in Figure 21. Residues 71-74 show large variations in the structure ensemble.

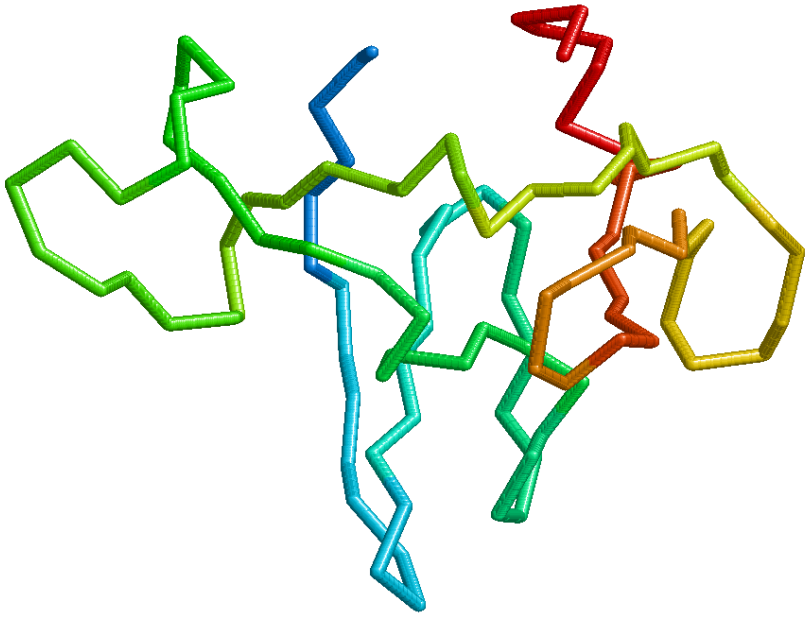


Figure 19. The globular domain of the average of the 5 lowest energy structures calculated with CNS using 935 manually assigned NOEs and 185 TALOS predicted dihedral angles. The first 20 residues are excluded. This presents the initial folding model.

A. Experimental restraints:

No. of distance restraints (total):	1823
Intraresidue NOEs:	873
Sequential NOEs:	467
Medium range NOEs:	147
Long range NOEs:	336
Hydrogen bond restraints:	68
Dihedral angle restraints:	185
$^{3}J_{HNHA}$ coupling constants:	93
C_α/C_β chemical shifts:	121
H_α chemical shifts:	106

B. Mean rms deviations from the experimental restraints:

Distance restraints (Å):	0.067±0.0013
Dihedral angle (deg):	0.960±0.083

C. Mean rms deviations from idealized covalent geometry:

Bond (Å):	0.0082±0.000
Angle (deg):	0.833±0.011
Improper (deg):	0.731±0.035

D. Ramanchandran analysis (residues 21-128):

Residues in favored regions:	84%
Residues in additional allowed regions:	16%
Residues in generously allowed regions:	0%

E. Atomic rms deviations (residues 21-128 with 71-74 excluded):

Backbone atoms (Å):	0.32
Heavy atoms (Å):	0.81

Table 4. Restraints summary and structural statistics

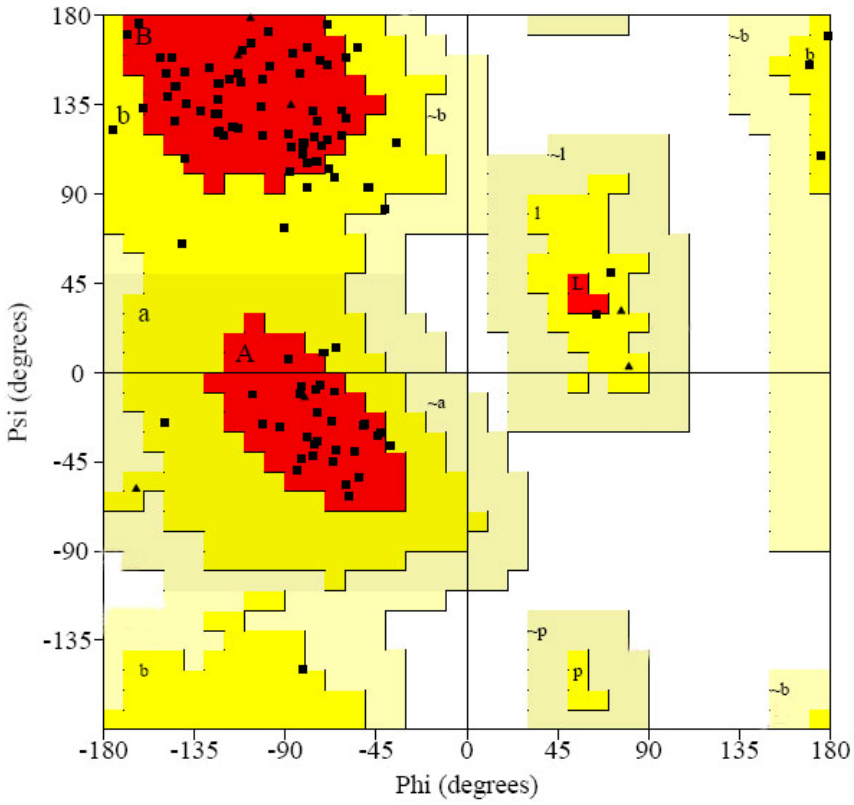


Figure 20. Ramachandran plot of the average TRD structure with the disordered N-terminal 20 residues excluded. 84% of residues are in the most favored regions, and 16% are in the additional allowed regions. The most favored regions are in the red area. Additional allowed regions are in dark yellow area. Generously-allowed regions are in the light yellow area, and the not- allowed regions are in the white area.

Figure 22 is a ribbon diagram of the average of the 10 lowest energy structures. Residues 24-26, 28-32, 35-39, 55-57, 62-64, 68-70, 79-81, 87-97, 101-109, and 112-118 make up β-strands numbered 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, respectively. Residues 122-126 make up the only α-helix in the protein. It should be pointed out that the β-sheets do not fall into any common folding superfamily such as a β-barrel or β-saddle. As mentioned in the introduction, rTRD shows no sequence homology to any known protein. When the structure was submitted to SWISS-MODEL (Schwede, 2003) or the DALI server (Holm, 2008), both of which are web-based protein structure homology model searches, no significant matches were found.

3.7 Hypothesis for the Cell Binding Site of rTRD

The novel fold of rTRD makes it difficult to predict an obvious binding site to the cell wall of sensitive bacterial strains without further chemical shift purturbation experiments. Future work will involve preparing soluble peptidoglycan fragments from sensitive bacteria strain ranging in molecular weights from 500-3000 Da, and complexing these with rTRD to establish the interface site. Before doing this, a hypothesis for the location of the interface was made by the discovery that EDTA binds to rTRD specifically. Figure 23 shows the chemical shift perturbation for rTRD backbone amides of T64, G67, T70, G79, Y80 and V82 in the 2D ^{15}N-HSQC spectra before and after the addition of 4 mM EDTA. These affected residues outline a surface cleft on the left upper corner of Figure 22. It is conceivable that the conformation of this portion of rTRD adjusts when it interacts with EDTA. These residues are also close to the surface of 71-74, which was poorly constrained in the structure calculations. Residues 71-74 may be disordered so as to provide flexibility to accommodate binding. The binding of EDTA to TRD may be an artifact, but, it is

specific and the functional groups of EDTA bear similarity to the peptide functionality of the peptidoglycan.

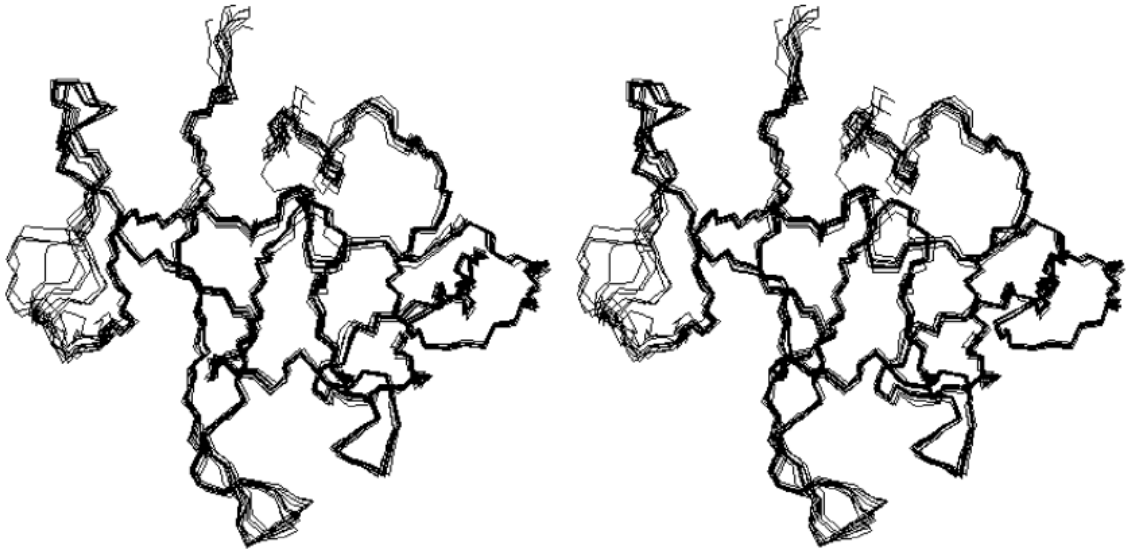


Figure 21. Stereo view showing the backbone superposition (residue 21-128) of a family of the 10 lowest energy structures of TRD. The backbone rmsd is 0.32 Å (first 20 residues and residues 71-74 were excluded). The poorly converged loop 71-74 is to the far left of the figure.

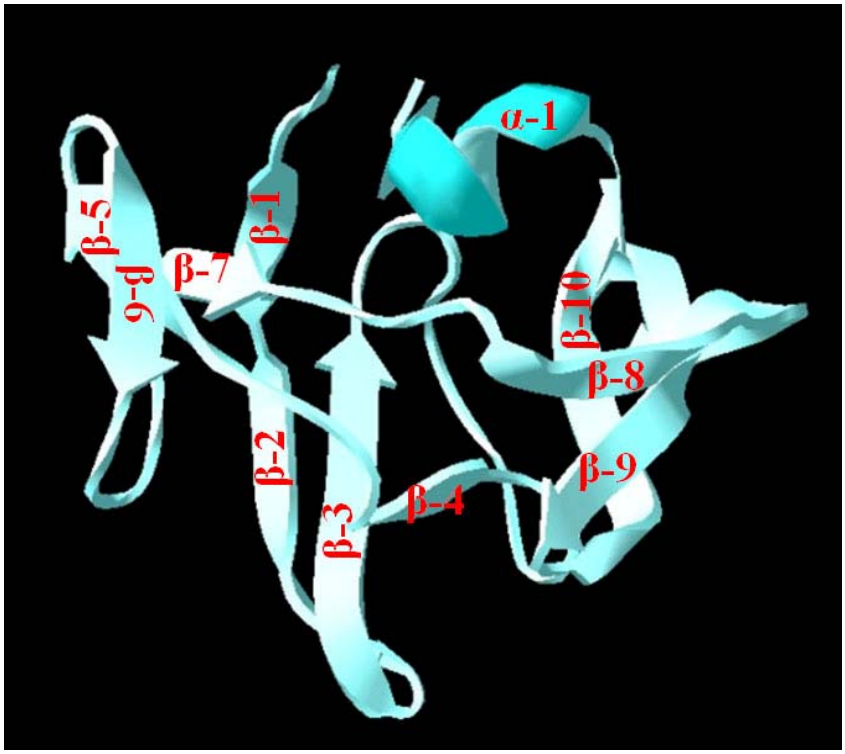


Figure 22. The average of the 10 lowest energy structures of TRD.

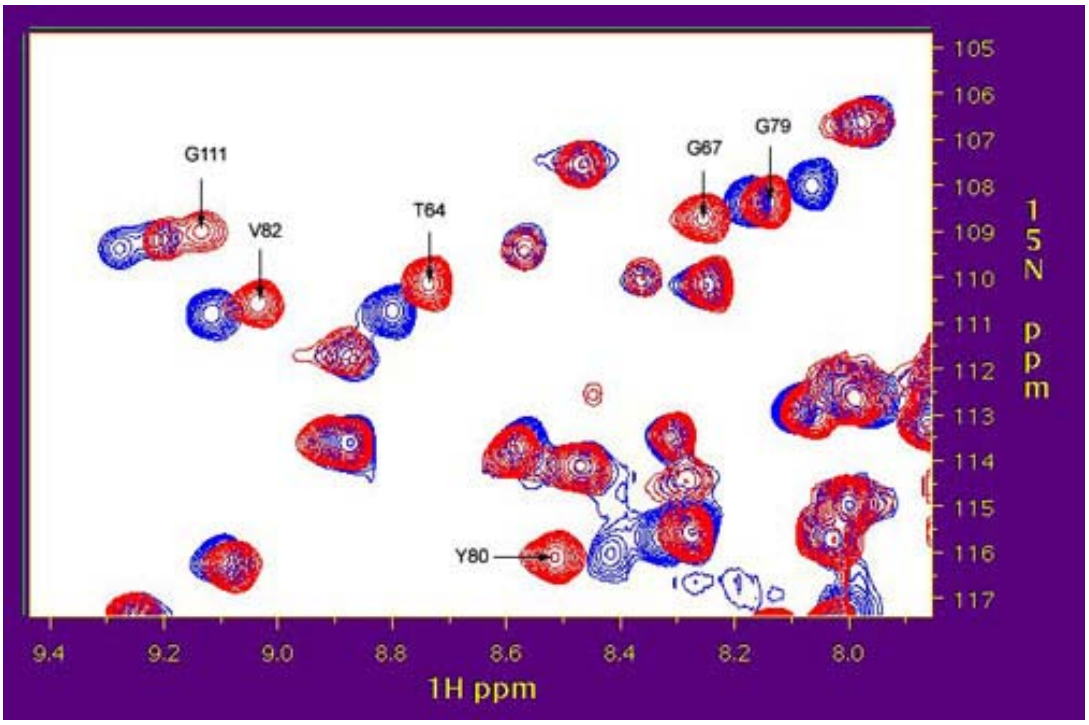


Figure 23. Superimposed select regions of the ^{15}N-hsqc spectra of rTRD before (blue contours) and after the addition of 4 mM sodium EDTA (red contours).

4. Conclusion

The protocol for the expression and purification of $^{13}C/^{15}N$-labeled rTRD was optimized in this work, and a 500 μL of 2.4 mM $^{13}C/^{15}N$-labeled rTRD NMR sample was obtained by using 2 g of expensive ^{13}C-glucose in the optimized M9 medium. The rTRD showed exceptional stability at room temperature, and only one sample was used to collect all required 3D NMR data.

The high resolution structure of the rTRD domain was solved by 3D NMR technology. A search for folds similar to that of TRD in the protein structure database yielded no significant matches, indicating a novel fold of the rTRD structure.

To understand the binding mechanism of the rTRD to the cell wall of other sensitive bacteria, further NMR titration experiments using soluble peptidoglycan fragments ranging in molecular weights from 500-3000 Da are required. A hypothesis for the location of the surface cleft was proposed after observing that EDTA bound specifically to rTRD. Chemical shifts of backbone amides of T64, G67, T70, G79, Y80 and V82 are affected by EDTA binding, and these residues are close to the poorly-structured region (residues 71-74) of the rTRD, which could provide flexibility for rTRD to bind to its natural substrate.

PART II

THE METAL BINDING SITE OF ZOOCIN A

1 Introduction

In many hydrolytic enzymes, the active site of a protein involves metal ions such as Zn, Mg, Cu or Ca, and many of the proteins become inactive once the metal ion is removed. The hypothesis that zoocin A is a zinc metalloproteinase was initially based on the fact that the N-terminal catalytic domain of zoocin A showed high homology (circa 35-40% identity, 50% positive match) to the active domain of zinc metalloproteinases lysostaphin and LytM (Ramadurai, 1999). Another important clue supporting this hypothesis was that negligible catalytic activity could be detected when the recombinant zoocin (rZOO) and the recombinant catalytic domain (rCAT) were treated with disodium ethylenediaminetetraacetic acid (EDTA) solution. However, direct metal analyses on the metal binding site failed (Chen, 2008). Inductively coupled plasma atomic emission spectrometry (ICP-AES) showed elevated levels of zinc in rZOO and rCAT, but the stoichiometry of metal to protein is less than 0.3:1, and other metals such as K, Na, P, Ni, B, Ca were also detected at comparable amounts. Fourier transform mass spectrometry (FT-MS) analysis also could not prove the presence of zinc due to the large

amounts of K^+ and Na^+ non-specifically bound to the protein. Therefore, indirect analysis based on solution NMR was employed to probe whether zoocin A is a zinc metalloproteinase and identify the ligand environment around the metal center.

^{113}Cd NMR has been used as a valuable probe for metal ion environment. On one hand, Cd^{2+} appears to bind remarkably well to a great variety of sites in proteins including those normally occupied by Mg^{2+} as well as those normally occupied by Zn^{2+} or Ca^{2+} due to their similar electron configuration or ionic radii. On the other hand, ^{113}Cd has good NMR sensitivity for a metal. The ^{113}Cd resonance in 0.01 M natural abundant $Cd(ClO_4)_2$ solution can be observed at an adequate signal-to noise ratio on a 500 MHz NMR instrument in a reasonable amount of time. Previous studies on approximately 20 different metalloproteinases using ^{113}Cd-NMR established a correlation between Cd resonance chemical shifts and the coordination environment of ^{113}Cd complexes (Figure 24) (Coleman, 1993). For example, coordination complexes with oxygen donors were observed to be the most shielding, whereas sulfur donors were the most deshielding, and proteins containing various mixtures of N and O donors at the Cd^{2+} binding sites were observed at a range of chemical shifts between 40 to 320 ppm. If zoocin A is a metalloprotein, the Cd-substituted rZOO should show resonance around this region with 0.1 M natural abundance Cd $(ClO_4)_2$ as chemical shift standard.

The 2D ^{15}N-homonuclear single quantum correlation (HSQC) experiment (Brodenhausen, 1980), which shows all amide groups in the protein polypeptide, is very sensitive to 3D structure and can detect conformation changes. If the metal ion were removed from zoocin A, changes would take place in the HSQC spectrum.

In this part of the dissertation, the isolation and purification of ^{113}Cd-rZOO and its ^{113}Cd-NMR will also be described.

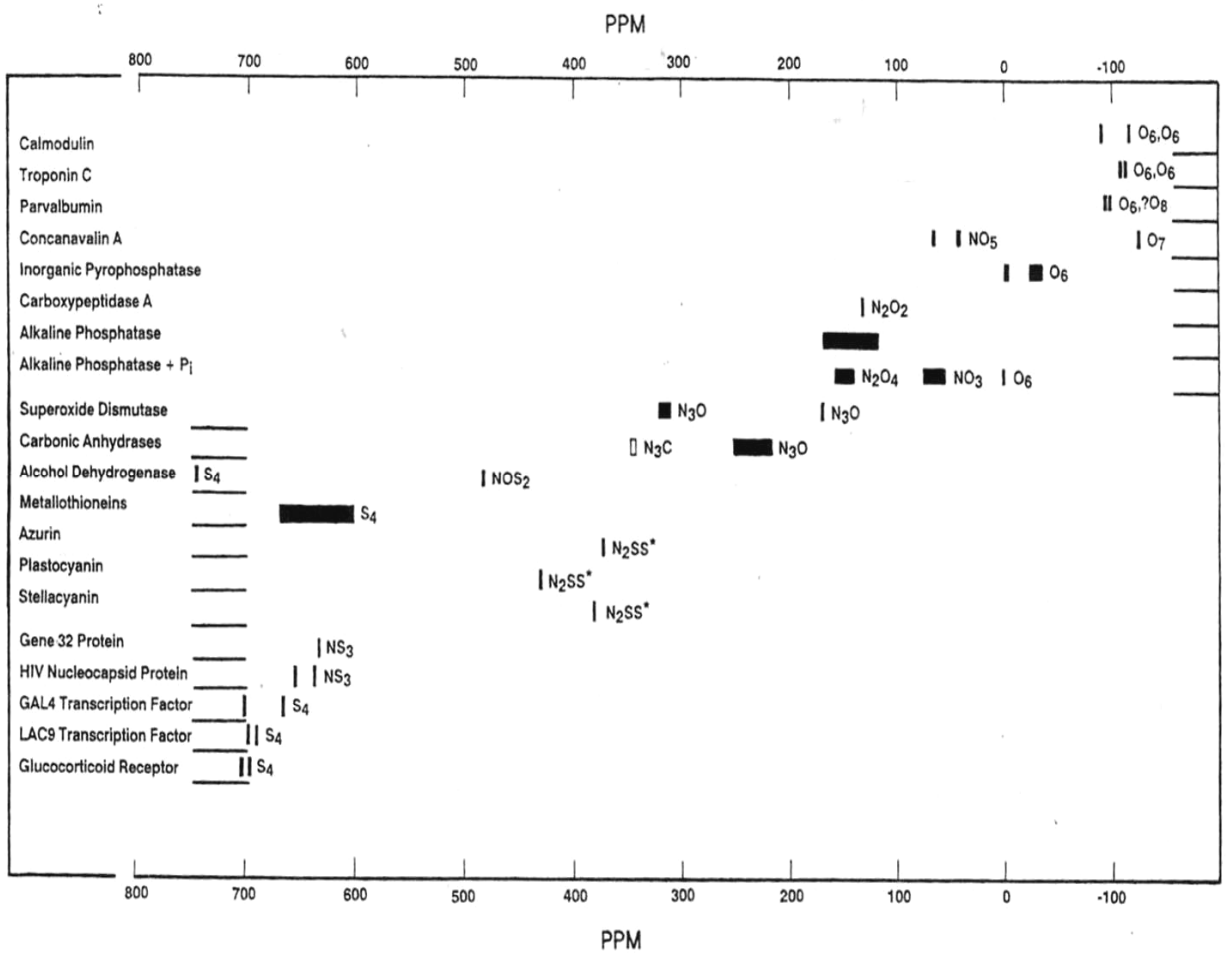


Figure 24. Chemical shifts of Cd protein complexes (Coleman, 1993)

2 Materials and Methods

2.1 Preparation of Cd-rZOO, Cd-rTRD, Cd-^{15}N-rCAT and Cd-azurin

Natural aboundance rZOO and ^{15}N-labeled rTRD were prepared according to a previously published protocol (Simmonds, 1997; Liang, 2004). Apo-rZOO was prepared by treating purified natural abundance rZOO with 5 mM disodium EDTA (or 10 mM 1,10-phennanthroline) in 10 mM potassium phosphate buffer, pH 6.8, for 6-8 hr, followed by removing EDTA-metal ion complexes by dialyzing against 10 mM sodium acetate buffer，pH 5.6, in which the solubility of rZOO is highest, four times with at least 5 hr between buffer changes. The volume of Apo-rZOO was reduced to 0.45 mL using a Centricon-3 ultrafiltration, and 10% D_2O was added to a final sample volume of 0.50 mL and a final sample concentration of 2.1 mM. $^{113}CdCl_2$ (94% enriched, Cambridge Isotope labs) was added directly into this sample to a final concentration of 5 mM, followed by incubation for 1 hr. Another ^{113}Cd-rZOO sample was prepared in parallel, but the final concentration of the $^{113}CdCl_2$ was sub-stoichiometric at 90% of the protein concentration. Apo-^{15}N-rTRD and apo-15N-rCAT were prepared in the same way as apo-rZOO. Apo-^{15}N-rCAT was unstable after EDTA treatment and considerable amounts of protein was lost. Apo-azurin, which was used to calculate the metal protein ratio of rZOO, was prepared by dialyzing Cu(I)-

azurin against 50 mM potassium cyanide in 25 mM tris-acetate, pH 8.0, followed by dialyzing versus tris-acetate, pH 8.0.

2.2 NMR Experiments

^{113}Cd-NMR spectrum was recorded at 302 K on a 500 MHz Bruker AVANCE spectrometer. Initial experiments used a very large spectral window so that resonances of interest appeared without folding-in. After correct resonace frequencies were found, the spectral window was narrowed to enhance spectral resolution. Figure 25 shows the resonances of a 2.1 mM ^{113}Cd-rZOO sample using 8196 data points over a spectral width of 198 ppm, a 45 transmitter pulse, an acquisition time of 0.19s, a recycle delay of 1s, and 57224 transients at 302 K. ^{113}Cd chemical shifts were reported using the shift of $Cd(ClO)_4$ as 0 ppm. The spectra were processed by the Bruker software package Topspin 1.3, and exponential line broadening of 100 Hz was applied. 2D ^{15}N HSQC spectra were recorded at 299K on a 600 MHz Bruker AVANCE spectrometer with a standard Bruker pulse program, hsqcetgpsi with Z-gradients (Brodenhausen, 1980). 1024 data points over a spectral window of 16 ppm were sampled in the proton dimension and 128 data points over a 35.3 ppm window in the ^{15}N dimension. A broadband GARP ^{15}N decoupling was employed during acquisition with a recycle delay of 1s, and the acquisition time was 0.053 s. The entire experimental time for one data set was 1.25 hr. The spectra were processed using NMRPipe software and apodized with two sine squared functions and zero-filled to 2048 and 1024 points. The same parameters were applied to the spectra when metal was removed and after Zn^{2+} reconstitution as well, except the number of scan was increased, which extended the

experimental time to 24 hr due to the severe protein losses during preparation of the metal-free samples.

The catalytic activities of rZOO, apo-rZOO and Cd-rZOO were determined by a dye-release assay with Remazol brilliant blue-labeled, heat-killed cells of *Streptococcus pyogenes* FF22 as a substrate as previously described (Lai, 2002).

3 Results and Discussion

Apo-rZOO could be prepared by treatment with either EDTA or 1,10-phenanthroline as described in the section of ‘‘Materials and Methods’’. Compared with 1,10-phenanthroline, EDTA proved to be more efficient at removing metal using relatively low levels and short exposure time, suggesting that the binding affinity of metal to rZOO is not extraordinarily high. This might explain why the previous metal analyses on the exhaustively dialyzed zoocin A sample showed less than a stiochiometric amount of zinc present.

Apo-rZOO can be prepared by either using 1, 10-phenanthronine or EDTA, and the latter removed the metal ion of zoocin A with relatively short exposure time, which suggested that the metal-protein interaction of zoocin A is not very strong. Catalytic activity analysis showed almost no streptococcolytic activity could be observed on apo-rZOO, while Cd-rZOO showed 80-85% of the native streptococcolytic activity, which suggested that the conformation of Cd^{2+} reconstituted rZOO is similar to that of the native rZOO.

^{113}Cd -NMR of Cd-substituted rZOO confirmed that Cd^{2+} was ligated to this protein. As shown in Figure 25A, a major resonance was observed at 113.6 ppm, with a minor peak at 87.3 ppm relative to $Cd(ClO_4)_2$. As it was well established that the chemical shift of the ^{113}Cd nucleus

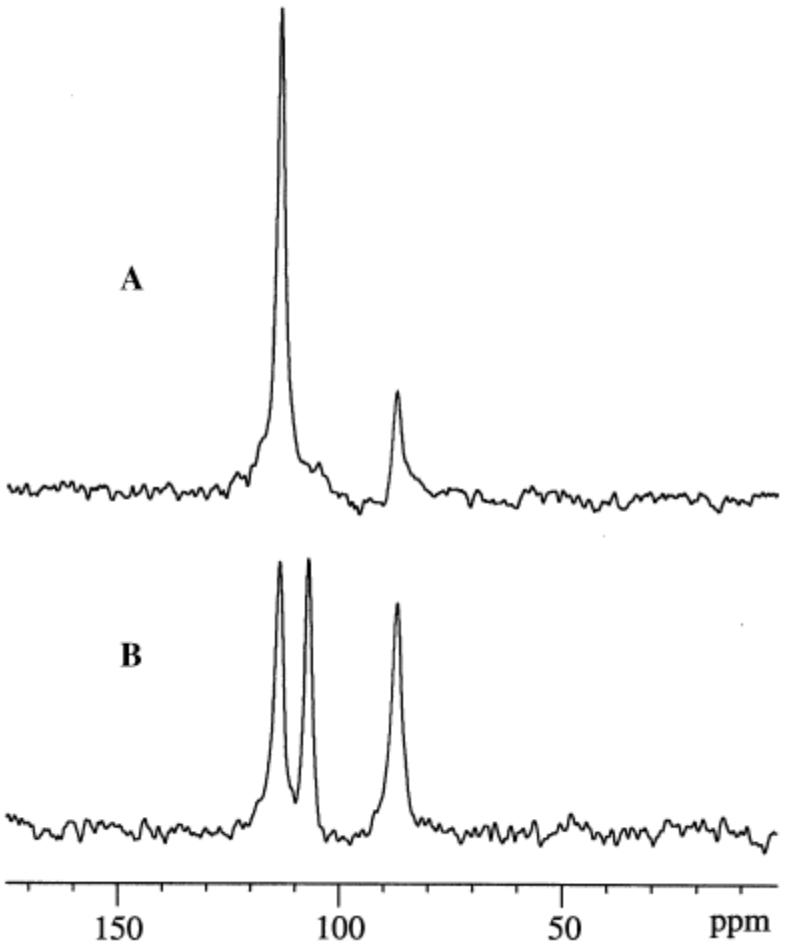


Figure 25. A) ^{113}Cd-NMR spectrum of ^{113}Cd-reconstituted zoocin A in 10 mM sodium acetate buffer, pH 5.6, at 302 K in the presence of excess ^{113}Cd. The relative intensity of the major peak at 113.6 ppm to the minor peak at 87.3 ppm is 0.84 to 0.16. B) Sample as in a, but after storage for 1 week. The relative intensities for the peaks at 113.6, 107.2, and 87.3 ppm are 0.39, 0.30, and 0.31. The excess ^{113}Cd appeared as a very broad resonance at 15.5 ppm.

was very sensitive to the nature and the number of donor atoms of the protein, the chemical shifts of Cd-substituted rZOO fell well within the window where two nitrogens and two oxygens bind to the metal.

Zoocin A is homologous to LytM, which is an autolysin secreted by *Staphylococcus aureas*. The crystal structures of inactive and active forms of LytM have been determined (Odintsov, 2004; Firczuk, 2005). Two side-chain nitrogens from H210 and H293, and two side-chain oxygens from D214 and N117 are ligated to the zinc metal in the inactive form of LytM. For the active form as studied in different crystals, water or a buffer anion replaces N117 as the fourth ligand. The type of the fourth ligand depended on how long they soaked the crystals in different buffers. Homologs for residue H210, D214 and H219 can be found in zoocin A, and they are H45, D49 and H133 respectively. However, no homolog for N117 can be found due to the lack of activation sequence in zoocin A. The minor peak of ^{113}Cd-rZOO could represent an alteration in the identity of the donor, in that the ^{113}Cd-rZOO sample was exposed to phosphate, acetate and chloride buffer during sample preparation, and Cd may vary the fourth ligand between those buffer ions and water. This situation became more complicated when sample was stored for 1 week. The major peak at 113.6 ppm split into two peaks at 113.6 and 107.2 ppm as shown in Figure 25B, and the intensity of these two peaks decreased, while the peak intensity at 87.3 ppm increased. This heterogeneity is probably caused by the displacement among buffer ions or some other slow conformational change at the active site on NMR time scale.

One concern about rZOO site was that the six-histidine tag at the N-terminus might constitute a physiologically meaningless metal binding site. This is unlikely because histidine rings are

protonated at pH 5.6, and hence will be poor ligands for Cd^{2+}. In a control experiment to substantiate this, Cd^{2+} was added to the recombinant target recognition domain rTRD which contains no metals, but has the hexa-His tag. ^{113}Cd-NMR only showed the very broad peak at 15.5 ppm due to the complex of Cd^{2+} with buffer ions. Therefore, it was concluded that the nitrogens from six-his tag will not bind to Cd^{2+} ion.

The multiple ^{113}Cd resonances observed conceivably could reflect multiple binding sites in the protein. A sample with a ratio of 0.90 equivalent of $^{113}Cd^{2+}$ to one equivalent of protein also showed the three-peak pattern, and the total resonance area was constant when more $^{113}Cd^{2+}$ was added.

The stoichiometry of Cd binding sites was experimentally determined by using a test protein with one metal binding site. Azurin is a blue copper electron transport protein found in *Pseudomonas* species with a single copper binding site, which can also be replaced by cadmium. More than 95% of the Cu^{2+} was removed by 1,10-phenanthronine, monitored the disappearance of the intense visible band at 625 nm due to charge transfer from the Cu^{2+}. Reconstituted Cd-azurin showed a ^{113}Cd resonance as reported (Engeseth, 1984). The intensity for a known quantity of Cd-azurin was used as an intensity standard for ^{113}Cd-rZOO, and the protein to metal ratio was 1:1.2±0.2.

The ^{1}H-^{15}N HSQC experiment shows all the amide proton and nitrogen in a protein, and since each amino acid residue has a unique environment for its amide group, this experiment is referred to as the fingerprint region for all proteins. It is very sensitive to 3D structure and conformational change. ^{15}N HSQC spectra were recorded for native rCAT, and after EDTA

removal of metal. Changes were evident in several regions, and the affected peaks were interpreted as peptide amides at or near the metal binding site. Figure 26 only shows this comparison for one selected region. The identities of the affected are not available at this time and will require complete chemical shifts assignments for rCAT. When the apo-rCAT was reconstituted with Zn, the spectrum changed back to that of rCAT. Therefore, zoocin A is a zinc metalloprotein to the extent that ^{15}N-HSQC experiment is diagnostic for a specific metal.

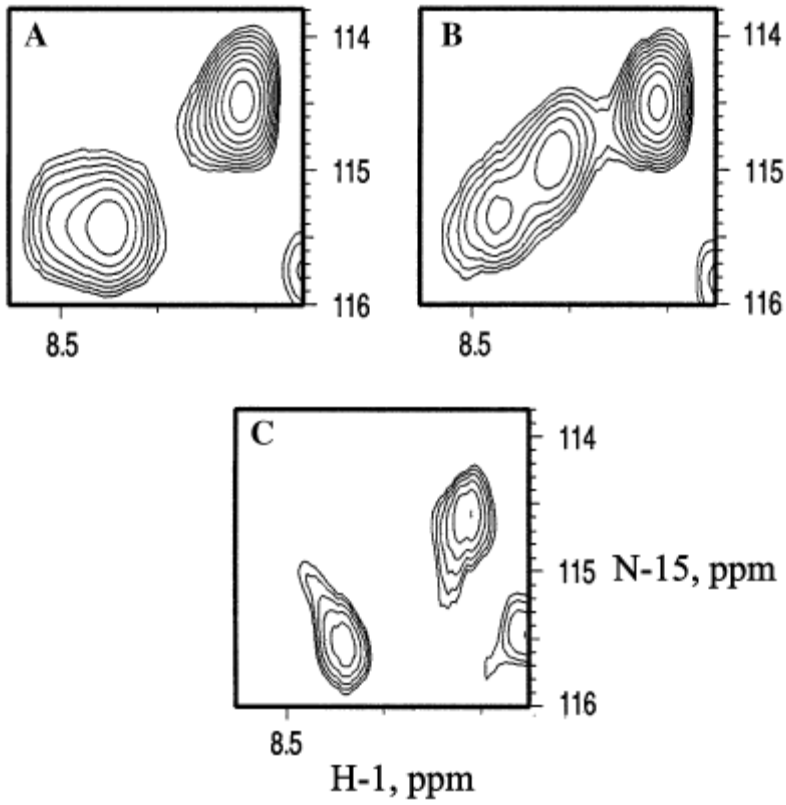


Figure 26. A select region of the ^{15}N–^{1}H homonuclear single quantum correlation spectrum of the ^{15}N recombinant catalytic domain of zoocin A (rCAT) illustrating how metal removal perturbed select fingerprint NH cross peaks, while reconstitution with Zn^{2+} restored a native like spectrum. A) Native ^{15}N-rCAT. B) After incubation with EDTA. C) After reconstitution with Zn^{2+} and sample cleanup to remove excess reagents and precipitated protein. The spectrum represented in C was weaker and had considerably more noise than those in A or B, because of massive protein loss due to the instability of apo-rCAT. So the contour level for the two-dimensional plot had to be raised to avoid the noise level and this makes the cross peaks in C appear artificially thinner.

4 Conclusion

^{13}Cd-rZOO showed 80-85% of native streptococcolytic activity, suggesting that the conformation of Cd-rZOO is similar to that of the native rZOO. Two resonances at 113.6 ppm and 87.3 ppm were observed one hour after Cd was incubated with apo-rZOO. After storage, the 113.6 ppm resonance split into two resonances at 113.6 ppm and 107.2 ppm, and the peak at 87.3 ppm increased intensity. The Cd chemical shifts indicate that two nitrogen atoms and two oxygen atoms are bound to the metal site.

The crystal structure of both active and inactive LytM has been reported. Based on the sequence similarity to LytM, the ligands in zoocin A were tentatively assigned to H45, D49, H133, and some combination of water or buffer ions as the fourth oxygen donor.

The protein binds a single metal atom, using a known quantity of Cd-azurin as an intensity standard for the amount of cadmium that can be bound to CAT.

Zinc removal and reconstitution experiment provide evidence for Zn^{2+} as the metal cofactor to the extent that ^{15}N-HSQC spectrum is diagnostic for a specific metal. Therefore, all evidences supports zoocin A as a single Zn^{2+} binding metalloprotein.

REFERENCES

Banci, L.; Felli, I. C.; Kummerle, R. *Biochemistry* **2002,** *41*, 2913-2920.

Berjanskii, M. V.; Wishart, D. S. A. *J. Am. Chem. Soc.* **2005,** *127*, 14970-14971.

Bartels, C.; Xia, T. H.; Billeter, M.; Guntert, P.; Wuthrich, K. *J. Biomol. NMR* **1995,** *6*, 1-10.

Boucher, W.A. 2002. http://www.ccpn.ac.uk/azara/.

Brodenhausen, G., Ruben, D. J. *Chem. Phys. Lett.* **1980**, *69*, 185-189.

Browder, H. P., Zygmunt w. A., Young, J. R., Tavormina, P. A. *Biochem. Biophys. Res. Commun.* **1965**, *19*, 383-389.

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.; Warren, G. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998,** *54*, 905-921.

Chen, Y.; Simmonds, R. S.; Sloan, G. L.; Timkovich, R. *J. Biol. Inorg. Chem.* **2008,** *13*, 855-860.

Clore, G.M. ; Brunger, A.T.; Karplus, M.; Gronenborn, A.M. *J. Mol. Biol.* **1986**, *191*, 523-551.

Coleman, J.E. *Methods Enzymol.* **1993**, *227*, 16-43.

Cornilescu, G.; Delaglio, F.; Bax, A. *J. Biomol. NMR* **1999,** *13*, 289-302.

Crippen, G.M. *Distance Geometry and Conformational Calculations,* Research Studies Press, **1981**.

Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995,** *6*, 277-293.

Engeseth, H.R.; McMillin, D.R.; Otvos, J.D. *J. Biol. Chem.* **1984**, *259*, 4822-4826.

Ernst, R. R., Bodenhausen, G., Wokaun, A. *Principles of Nuclear Magnetic Resonance in One and two Dimentions*, Clarendon Press, Oxford. **1987.**

Firczuk, M.; Mucha, A.; Bochtler, M. *J. Mol. Biol.* **2005,** *354*, 578-590.

Garrett, D. S.; Powers, R.; Gronenborn, A. M.; Clore, G. M. *J. Magn. Reson.* **1991,** *95*, 214-220.

Garrett, D.S.; Kuszewski, J.; Hancock, T.J.; Lodi, P.J.; Vuister, G.W.; Gronenborn, A.M.; Clore, G.M. *J. Magn. Reson. Ser. B*, **1994**, *104*, 99-103.

Gargis, S. R.; Heath, H. E.; Heath, L. S.; LeBlanc, P. A.; Simmonds, R. S.; Abbott, B. D.; Timkovich, R.; Sloan, G. L. *Appl. Environ. Microbiol*. **2009**, *1*, 72-77.

Gaspar, J. A.; Liu, C. S.; Vassall, K. A.; Meglei, G.; Stephen, R.; Stathopulos, P. B.; Pineda-Lucena, A.; Wu, B.; Yee, A.; Arrowsmith, C. H.; Meiering, E. M. *Protein Science* **2005,** *14*, 216-223.

Geerten, W.V.; Bax, A. *J. Am. Chem. Soc.* **1993**, *115*, 7772-7777.

Grzesiek, S.; Bax, A. *J. Magn. Reson.* **1992,** *96*, 432-440.

Heath, L. S.; Heath, H. E.; LeBlanc, P. A.; Smithberg, S. R.; Dufour, M.; Simmonds, R. S.; Sloan, G. L. *FEMS Microbiol. Lett.* **2004,** *236*, 205-211.

Holm, L.; Kaariainen, S.; Rosenstrom, P.; Schenkel, A. *Bioinformatics*, **2008**, *24*, 2780-2781.

Hooft, R. W. W.; Vriend, G.; Sander, C.; Abola, E. E. *Nature* **1996,** *381*, 272-274.

James, S. M.; Tagg, J. R. *Microb. Ecol. Hlth. Dis*. **1988**, *1*, 153-162.

Jeremy N.S.E. *Biomolecular NMR spectroscopy*. Springer, **1999**.

Johnson, B. A.; Blevins, R. A. *J. Biomol. NMR* **1994,** *4*, 603-614.

Karplus, K. *J. Phys. Chem.* **1959**, *30*, 11-15.

Karplus, K. *J. Am. Chem. Soc.* **1963**, *85*, 2870-2871.

Kay, L. E.; Xu, G. Y.; Yamazaki, T. *J. Magn. Reson., Ser. A* **1994,** *109*, 129-133.

Kuszewski, J.; Gronenborn, A.M.; Clore, G.M. *J. Magn. Reson. Ser. B* **1995**, *106*, 92-96.

Kuszewski, J.; Gronenborn, A.M.; Clore, G.M. *J. Magn. Reson. Ser. B* **1995**, 107, 293-297.

Laemmli, U. K. *Nature* **1970**, *227*, 680-685.

Lai, A.C.; Tran, S.; Simmonds, R.S. *FEMS Microbiol. Lett.* **2002**, *215*, 133-138.

Laskowski, R. A.; Thornton, J. M.; Humblet, C.; Singh, J. *J. Mol. Biol.* **1996,** *259*, 175-201.

Liang, Q.; Miller, G.T.; Beeghley, C.A.; Graf, C.B.; Timkovich, R. *Biophys. J.* **2007**, *93*, 1700-

1706.

Liang, Q.; Simmonds, R. S.; Timkovich, R. *Biochem. Biophys. Res. Commun.* **2004,** *317*, 527-530.

Löhr, F.; Hansel, R.; Rogov, V. V.; Dotsch, V. *J. Biomol. NMR* **2007,** *37*, 205-224.

Machalek, A.Z. *The structure of life,* NIH Pub, **2007**.

Montelione, G. T.; Emerson, S. D.; Lyons, B. A. *Biopolymers* **1992,** *32*, 327-334.

Muhandiram, D. R.; Kay, L. E. *J. Magn. Reson., Ser. B* **1994,** *103*, 203-216.

Nilges, M.; Macias, M. J.; ODonoghue, S. I.; Oschkinat, H. *J. Mol. Biol.* **1997,** *269*, 408-422.

Nilges, M. *Fold Des.* **1997,** *2*, S53-S57.

Odintsov, S. G.; Sabala, I.; Marcyjaniak, M.; Bochtler, M. *J. Mol. Biol.* **2004,** *335*, 775-785.

Ramadurai, L.; Lockwood, K.J.; Nadakavukaren, M.J.; Jayaswal, R.K. *Microbiol.* **1999**, *145*, 801-808.

Schwede, T.; Kopp, J.; Guex, N.; Peitsch, M.C. *Nucleic Acids Res.* **2003**, *31*, 3381-3385.

Simmonds, R. S.; Naidoo, J.; Jones, C. L.; Tagg, J. R. *Microb. Ecol. Hlth. Dis*. **1995**, *8*, 281-292.

Simmonds, R. S.; Simpson, W. J.; Tagg, J. R. *Gene* **1997,** *189*, 255-261.

Simmonds, R. S.; Pearson, L.; Kennedy, R. C.; Tagg, J. R. *Appl. Environ. Microbiol.* **1996,** *62*, 4536-4541.

Singh, S.; Folkers, G. E.; Bonvin, A. M. J. J.; Boelens, R.; Wechselberger, R.; Niztayev, A.; Kaptein, R. *Embo Journal* **2002,** *21*, 6257-6266.

Sattler, M.; Schmidt, P.; Schleucher, J.; Schedletzky, O.; Glaser, S. J.; Griesinger, C. *J. Magn. Reson., Ser. B* **1995,** *108*, 235-242.

Serber, Z., Dötsch, V. *Biochemistry* **2001**, *40*, 14317-14323.

Serber, Z.; Richter, C.; Dotsch, V. *Chem. Biochem.* **2001,** *2*, 247-251.

Schwarzinger, S.; Kroon, G. J. A.; Foss, T. R.; Wright, P. E.; Dyson, H. J. *J. Biomol. NMR* **2000,** *18*, 43-48.

Shen, Y.; Vernon, R.; Baker, D.; Bax, A. *J. Biomol. NMR* **2009,** *43*, 63-78.

Stejskal, E.O.; Tanner, J.E. *J. Chem. Phys.* **1965**, *42*, 288-292.

Tatusova, T.A.; Madden, T.L. *FEMS Microbiol. Lett.* **1999**, *177*, 187-188.

Teng, Q. *Structural Biology: Practical NMR Applications*. Springer, **2005**.

Timkovich, R.; Bergmann, D.; Arciero, D. M.; Hooper, A. B. *Biophysi.* **1998,** *75*, 1964-1972.

Turner, C. F.; Moore, P. B. *J. Mol. Biol.* **2004,** *335*, 679-684.

van Gunsteren, W. F.; Dolenc, J. Biomolecular simulation: historical picture and future perspectives. *Biochem. Soc. Trans.* **2008,** *36*, 11-15.

Vranken, W. F.; Boucher, W.; Stevens, T. J.; Fogh, R. H.; Pajon, A.; Llinas, P.; Ulrich, E. L.; Markley, J. L.; Ionides, J.; Laue, E. D. *Proteins: Struct., Funct., Bioinf.* **2005,** *59*, 687-696.

Williamson, M. P.; Havel, T. F.; Wuthrich, K. *J. Mol. Biol.* **1985,** *182*, 295-315.

Wittekind, M.; Friedrichs, M. S.; Constantine, K. L.; Metzler, W. J.; Bassolino, D.; Mueller, L. *J. Cell. Biochem.* **1993,** 258.

Wittekind, M.; Mueller, L. *J. Magn. Reson. Ser. B* **1993**, *101*, 210-205.

Yamazaki, T.; Lee, W.; Arrowsmith, C. H.; Muhandiram, D. R.; Kay, L. E. *J. Am. Chem. Soc.* **1994,** *116*, 11655-11666.

Yamazaki, T.; Formankay, J. D.; Kay, L. E. *J. Am. Chem. Soc.* **1993,** *115*, 11054-11055.

APPENDIX I

Protocol for isolation and purification of ^{13}C, ^{15}N labeled rTRD

Day 1

1. Prepare LB media:

1) Get two 2 L flasks and three 100 mL flasks.

2) Fill up each 2 L flask with 1 L of distilled or Milli-Q de-ionized water. Dissolve the following chemicals listed in Table A1 in water to make two flasks of 1 L LB media and one flask of 70 mL LB Agar medium.

Table A1: Ingredients of LB and agar LB media

Chemicals or solvent	**1 L LB**	**70 mL LB agar**
H_2O	1 L	70 mL
Bacto-tryptone	10 g	0.7 g
Bacto-yeast extract	5 g	0.35 g
NaCl	10 g	0.7 g
Agar	0 g	1.05 g

3) Transfer 50 mL of LB from each 1L LB medium to a 100 mL flask, resulting in two 950 mL and two 50 mL portions of LB media and one 70 mL portion of agar LB. The pH for each

4) solution is ~7.2. Cap with aluminum foil and stick autoclave tape onto them to indicate if they are sterilized or not.

2. Prepare 1 L of 10X M9 minimal stock solution in a 1 L flask according to Table A2:

Table A2: Ingredients of 10X M9 minimal stock solution

KH_2PO_4	30 g
Na_2HPO_4	66.52 g
NaCl	5 g

3. Dilute 50 mL of above 10X M9 stock solution into 1 L with distilled water in another 1 L flask.

4. Prepare buffers for running the column according to Table A3:

Table A3: Ingredients of lysis buffer and elution buffer

	Lysis buffer (1 L, adjust pH to 8.0 with NaOH)	Elution buffer (1 L, pH is ~8.0)
50 mM $NaH_2PO_4 \cdot H_2O$	6.9 g	6.9 g
300 mM NaCl	17.54 g	17.54 g
Imidazole	0.68 g (for 10mM imidazole)	17 g (for 250 mM imidazole)

5. Autoclave above LB media, M9 medium, buffers, together with two boxes of pipette tips (one large and one small size), tens of micro-centrifuge tubes (1.5 mL), and a 2 L fermentation jar using cycle P14 or P17. This step normally takes approximately 90 min.

6. While waiting, prepare 10 mL of ampicillin (A_m, 100 mg/mL) stock solution and 10 mL of kanamycin (K_m, 25 mg/mL) stock solution using 15 mL disposable tubes. Sterilize them with Millipore express (0.22 μm in pore size) polyether-sulfone (PES) filters instead of

autoclaving because antibiotics are heat-labile. Divide them into nine 1.5 mL micro-centrifuge tubes and store in the freezer.

7. Pour plates: Light the burner followed by wearing gloves, and then sterilize the working area of the bench with 75% of ethanol. Take out two disposable plates (Φ 90 mm) and label. Cool the 70 mL of agar LB to ~50 °C before adding 70 μL of Km and 70μL of Am. Remove the aluminum foil after the addition of antibiotics, and sterilize the bottle top with the flame quickly. Grab a disposable plate using your left hand, open it with your left thumb and move close to the flame, pour LB agar (~30 mL) into it and close the lid. Repeat this process for the second plate and let them sit at room temperature for 3-5 min to solidify. Invert the plates and incubate them at incubator for at least 6 hrs at 37°C, which ensures no contamination occurs in the plate and at the same time allowing the LB agar to "sweat" so that no extra moisture can interfere with the colonies later. This step should begin late in the afternoon to allow the overnight sweating.

Day 2

8. Streak plates: Get stock E.coli M15 cell line containing pQETRD gene (stored in -70°C refrigerator at Dr.B's lab, sample label is TRD dated 05-14-07, box label is Dr. Timkovich). Keep bacteria cells cold on dry ice. Light the burner followed by wearing gloves, and then sterilize the working area of the bench with 75% of ethanol 3 times. Thaw cells and pick up bacteria using a sterile wire loop, followed by spreading cells over the agar. Incubate the plates overnight (12-24 hrs).
9. Prepare 1 mL 100 mg/mL lysozyme stock solution, filter through PES filter and freeze.

10. Prepare 100 mM IPTG stock solution by dissolving 1.19 g of IPTG powder into 50 mL of milli-Q water. Filter through PEP membrane and freeze.

Day 3

11. Small culture (50 mL): This step is recommended to start after 6 pm of day 3. Light the burner, put on gloves, and then sterilize the working area of the bench with 75% of ethanol 3 times. Thaw antibiotics Km and Am and add 50 μL of Km and 50 μL of Am to each 50 mL portion of LB medium. Pick up single colony using sterile wire loop and scratch on the inside of the bottle. Incubate them overnight (12-14 hrs) at incubator shaker in room 137 or Dr. B's lab (reservation is required), 37°C. The plate can be stored in the refrigerator up to 2 weeks after overnight incubation.

Day 4

12. Large scale culture: this step is recommended to start before 10 am of day 4. Thaw antibiotics and add 950 μL of K_m and 950 μL of A_m to each 950 mL of LB medium prepared on day 1. Swirl them, followed by leaving out 1 mL in sterile micro-centrifuge tube as a blank solution for UV-Vis purposes. Transfer 50 mL cell suspensions to each 950 mL LB portion (with antibiotics) and incubate at 37°C. Measure OD_{600nm} every 40 min until it reaches 0.7 (middle of the log phase). This takes about two hours. (Note: After this step, 1 mM IPTG was added to induce the protein when preparing natural abundant rZOO, rCAT or rTRD. The cells were incubated for additional 5 hours, and then continue with step 17)
13. Spin down 2 L cells at 6000 rpm (GSA rotor), 20 min using four 500 mL centrifuge tubes. Decant the supernatants as completely as possible to remove residual LB medium in the cells.

14. Three-time washing: resuspend cell pellets in approximately 20 mL of M9 minimal solution (see day 1, dilute the 10X stock M9 solution first), this step requires approximately 15 min shaking. Once all cells are resuspended, add 100-150 mL of M9 minimal wash solution to each bottle, followed by balancing and re-centrifuging at 6000 rpm (GSA rotor) for 20 min. Decant the supernatants, and repeat the washing process for two more times
15. Do the following two things while waiting for the centrifuge: 1) thaw the ingredients for M9 medium (see Table A4) and add them to the 500 mL portion of sterile M9 minimal medium. Store in 37°C incubator. 2) Setup for fermentation using a circulating water bath and the special fermentation jar made by Mr. Rick Smith in the UA Glass Shop. Connect the jar to the water bath apparatus with tubing (pay attention to the direction of the water flow). Add distilled water into the water bath and turn on the power to heat the water up to 37°C and maintain the temperature.

Table A4: nutrients for 500 mL M9 minimal medium

20% ^{13}C-Glucose	10 mL	$^{15}NH_4Cl$	0.66 g
BVME	5 mL	6 mg/mL Thiamine	2 mL
1M $MgSO_4$	2 mL	K_m	500 µL
1M $CaCl_2$	0.1 mL	A_m	500 µL
Cai trace element*	1 mL		

The ingredients for the Cai trace element solution are listed in Table A5. Normal color for this solution is golden brown if sufficiently dissolved. Sterilize afterwards by filteration through a 0.22 µM filter.

Table A5: Cai trace elements (100 mL)

$CaCl_2$	60 mg	$ZnSO_4 \cdot 7H_2O$	7 mg
$FeSO_4 \cdot H2O$	60 mg	$CuCl_2 \cdot 2H_2O$	3 mg
EDTA	50 mg	$(NH_4)_6Mo_7O_{24} \cdot 4H_2O$	2.5 mg
$MnCl_2 \cdot 4H_2O$	11.5 mg	H_3BO_3	0.2 mg
$CoCl_2 \cdot 6H_2O$	8 mg		

16. Go back to the centrifuge, decant the supernatant and collect cells. Use the M9 medium with nutrients to resuspend those cells and transfer them to the fermentation jar. Adjust the air flow and turn on the stirrer to ensure sufficient surface area contact with air/O_2. Add 3-4 drops of antifoam A reagent, followed by incubating at 37 °C for one hour to "wake up" the cells. Finally, add 5 mL of IPTG stock solution (100 mM) to induce the protein expression and let it grow for 21 hrs.

Day 5

17. Spin down the cells at 6000 rpm, 4°C for 30 min. Keep all the supernatant, label and store it in the freezer.
18. Resuspend combined cell pellets in 40 mL of lysis buffer.
19. Sonicate cells for ~3 min (30 s burst with 30 s rest repeated 4X) on ice and incubate on ice with the addition of 400 µL of 1 mg/mL lysozyme for 2 hrs.
20. Centrifuge at 12,500 rpm for 20 min at 4°C and keep supernatant for purification.
21. Add small amount of DNaseI to digest the gluey DNA followed by filtering through PES membrane to remove the DNA.

Day 6

22. Prepare column: equilibrate 10 mL packed Ni-NTA column with 100 mL of lysis buffer for at least 45 min.
23. Load protein onto column and collect flow-through in a plastic tube (This fraction will be called LOAD FT)
24. Wash the column with 45 mL of lysis buffer and collect flow-through in a plastic tube (called WASH FT1)
25. Wash the column with 60 mL of medium ionic strength buffer (50 mM Imidazole, obtained by mixing 50 mL of lysis buffer with 10 mL of elution buffer), and collect flow-through in a plastic tube (called WASH FT2 and 3)
26. Elute protein using a 60 mL of elution buffer and collect flow-through in five 15 mL plastic tubes, 12mL each (called WASH FT4, 5, 6,7, and 8). Store them in the refrigerator temporarily.
27. While waiting for the column purification, wash dialysis bag (size exclusion 6000-8000 Da) with running water for 1-2 hours to remove residual glycerol.
28. Prepare dialysis buffer stock solution (100 mM KPi buffer, pH 7.0, 1mM NaN_3) (Note: for unlabeled or isotopically labeled rCAT and rZOO, dialysis buffer stock solution is 100 mM sodium acetate buffer, pH 5.6, 1 mM NaN_3).
29. Transfer proteins into the dialysis bag with pipette and dialyze them in the cold room with stirring using 1 L dialysis buffer three times to remove imidazole which interferes the

absorbance at 280 nm of protein. (Note: for rCAT and rZOO, 1 mM DTT was needed in the dialysis buffer so that disulfide bond cannot be formed)

Day 7

30. After dialysis (normally done by 4 pm of the day), Measure the A280 nm to determine the concentration or the mass of the protein, the volume should be around 60 mL, use the following equation to get the mass of the protein.

    Mass=(A*dilution factor)*V*M_w/ε

    Where A is absorbance, V is the total volume, M_w is the molecular weight for TRD, which is 14 kDa, ε is the extinction coefficient, which is 25 mM^{-1}.

Day 8

31. Concentrate the 60 mL protein solution to about 20 mL using a large size Amicon Ultrafiltration unit with a 43 mm diameter membrane, and then using small size amicon to concentrate it into 3-4 mL, finally using centricon-3 concentrator to reduce the volume to ~0.5 mL.
32. NMR sample was prepared using a quality NMR tube (5 mm in diameter), the final ratio for H_2O to D_2O is 95%:5%.

Preparation of Weakly Aligned ^{15}N-TRD Sample

1. Liquid crystalline medium preparation: 2X liquid crystalline medium solution contains 10% $C_{12}E_5$ (wt%), hexanol (molar ratio of $C_{12}E_5$: hexanol=0.85:1), and 10% D_2O in 10 mM phosphate buffer with 1mM NaN_3, pH 7.0. The detailed ingredients for 500 μL of 2X liquid crystalline medium is listed as following:
    1) 51.9 μL 10% C12E5 (Mw=406.61, ρ=0.963g/mL);
    2) 50 μL 10% D2O;
    3) 380 μL 10 mM sodium phosphate buffer, pH 7.0 (0.0584 g of NaH_2PO_4 + 0.1547 g of Na_2HPO_4 to make 100 mL of NaPi buffer, pH 7.0);
    4) 18 μL hexanol (Mw=102.17, ρ=0.82g/mL);
    5) Mix 2) and 3) first;
    6) Dissolve $C_{12}E_5$ into the mixture of 5), centrifuge at high speed for 1 min to remove bubbles;
    7) Add hexanol in microliters steps under vigorous shaking, if successful, light blue but clear solution can be formed. If it turns into a creamy solution, the preparation was failed. If it turns into something between creamy solution and clear light blue solution, leave it overnight or sonication may help;
    8) Centrifuge for 30 s~1 min is needed to remove bubbles.
2. Mix the ^{15}N-labeled TRD with the liquid crystalline medium in a 1:1 ratio. Avoid generating air bubbles and high salt concentration if possible. The sample should look translucent (slightly bluish) at room temperature, and clear at higher temperature.

APPENDIX II

Chemical Shift Assignments

1.CA	55.8	4.CA	58.36	8.HN	7.94
1.HA	3.8	4.HA	4.47	8.CA	54.12
1.CB	32.82	4.CB	63.59	8.HA	4.42
1.HB2	1.96	4.HB2	3.84	8.CB	43.24
1.HB1	1.86	4.C	174.19	8.HB2	1.28
1.CG	31.2	5.N	123.78	9.N	127.6
1.HG2	2.25	5.HN	8.45	9.HN	8.83
2.CA	56.49	5.CA	54.66	9.CA	58.6
2.HA	4.57	5.HA	4.54	9.CB	29.56
2.CB	31.02	5.CB	29.55	10.N	127.6
2.HB2	1.67	5.HB2	2.02	10.HN	8.83
2.HB1	1.75	5.HB1	1.98	10.HA	3.8
2.CG	23.46	5.C	174.71	10.CB	26.5
2.HG2	1.51	6.N	8.13	11.N	109.9
2.CD	43.39	6.HN	119.89	11.HN	8.33
2.HD2	3.11	6.CA	55.41	11.CA	43.19
2.NE	127.34	6.HA	4.57	11.HA2	4.3
2.HE	7.9	6.CB	30.95	12.N	107.11
2.C	175.84	6.HB2	2.27	12.HN	7.37
3.N	110.53	6.HB1	1.88	12.CA	58.21
3.HN	8.4	6.C	172.65	12.HA	4.45
3.CA	45.31	7.N	128.89	12.CB	64.11
3.HA2	3.95	7.HN	8.57	12.HB2	3.8
3.HA1	3.83	7.CA	59.23	12.C	173.4
3.C	174	7.HA	4.04	13.N	121.58
4.N	115.58	7.CB	38.45	13.HN	8.17
4.HN	8.24	8.N	130.46	13.CA	54.24

13.HA	4.39	17.HD1	3.79	21.C	176.75
13.CB	30.67	17.C	177.42	22.N	123.61
13.HB2	2.27	18.N	112.34	22.HN	8.79
13.HB1	1.95	18.HN	8.42	22.CA	56.6
13.C	174.77	18.CA	62.01	22.HA	3.81
14.CA	63.15	18.HA	4.35	22.CB	32.55
14.HA	4.45	18.CB	69.67	22.HB2	0.96
14.CB	32.21	18.HB	3.13	22.HB1	1.36
14.HB2	2.3	18.HG21	1.34	22.CG	24.77
14.HB1	1.95	18.C	174.9	22.HG2	0.17
14.CG	27.56	19.N	115.3	22.CD	29.76
14.HG2	1.98	19.HN	7.69	22.HD2	0.78
14.CD	50.75	19.CA	63.02	22.CE	42.38
14.HD2	3.92	19.HA	4.09	22.HE2	2.77
14.HD1	3.75	19.CB	69.88	22.HE1	2.61
14.C	177.02	19.HB	4.09	22.C	173.74
15.N	114.11	19.CG2	21.65	23.N	124.91
15.HN	8.23	19.HG21	1.13	23.HN	7.66
15.CA	62.28	19.C	173.91	23.CA	61.48
15.HA	4.35	20.N	120.26	23.HA	3.96
15.CB	70.29	20.HN	8.57	23.CB	39.24
15.HB	4.17	20.CA	53.51	23.HB	1.33
15.HG21	1.17	20.HA	4.72	23.CG1	19.49
15.C	174.55	20.CB	38.6	23.HG12	0.66
16.N	119.46	20.HB2	2.88	23.HG11	1.03
16.HN	8.15	20.CG	177.28	23.CD1	14.2
16.CA	60.21	20.ND2	112.07	23.HD11	0.57
16.HA	4.59	20.HD21	7.52	23.CG2	28.14
16.CB	69.98	20.HD22	6.86	23.HG21	1.64
16.HB	4.07	20.C	175.03	23.C	175.29
16.HG21	1.26	21.N	122.51	24.N	129.35
16.C	172.65	21.HN	7.91	24.HN	9.16
17.CA	63.1	21.CA	55.6	24.CA	56.88
17.HA	4.54	21.CB	42.16	24.HA	4.41
17.CB	32.44	21.HB2	1.73	24.CB	41.53
17.HB2	1.99	21.HB1	2.64	24.HB2	2.51
17.HB1	2.35	21.CG	27.64	24.HB1	0.58
17.CG	27.54	21.HG	1.47	24.HD1	6.95
17.HG2	2.08	21.CD1	26.05	24.HE1	6.61
17.CD	51.27	21.HD11	1.1	24.HE2	6.61
17.HD2	3.91	21.HD21	0.84	24.CD2	6.95

24.C	173.86	29.N	127.16	31.CE	41.4
25.N	120.45	29.HN	8.95	31.HE2	2.32
25.HN	8.75	29.CA	54.52	31.C	174.78
25.CA	55.59	29.HA	5.6	32.N	129.88
25.HA	4.53	29.CB	46.29	32.HN	8.57
25.CB	34.41	29.HB2	1.16	32.CA	58.68
25.HB2	1.69	29.HB1	1.69	32.HA	4.05
25.HB1	1.8	29.CG	27.17	32.CB	38.26
25.CG	24.98	29.HG	1.49	32.HB	1.61
25.HG2	1.14	29.CD1	24.36	32.CG1	25.61
25.CD	29.24	29.HD11	0.74	32.HG12	0.51
25.CE	42.16	29.CD2	27	32.HG11	0.7
25.HE2	2.85	29.HD21	0.45	32.CD1	10.8
25.C	177.5	29.C	174.78	32.HD11	-0.06
26.N	123.58	30.N	124.51	32.CG2	16.6
26.HN	8.82	30.HN	8.72	32.HG21	0.64
26.CA	61.59	30.CA	54.07	32.C	176
26.HA	4.07	30.HA	4.52	33.CA	53.94
26.CB	30.31	30.CB	32.96	33.HA	4.25
26.HB	2.3	30.HB2	0.37	33.CB	37.29
26.CG2	21.98	30.HB1	1.22	33.HB2	2.68
26.HG21	0.94	30.CG	34.43	33.HB1	2.98
26.CG1	24.3	30.HG2	2.03	33.CG	175.15
26.HG11	0.79	30.HG1	1.87	33.ND2	113.01
26.C	175.1	30.CD	179.6	33.HD21	7.75
27.N	130.62	30.NE2	112.7	33.HD22	6.77
27.HN	8.46	30.HE21	7.98	33.C	175.15
27.CA	58.32	30.HE22	6.75	34.N	101.69
27.HA	4.62	30.C	172.45	34.HN	8.27
27.CB	41.76	31.N	125.88	34.CA	45.19
27.HB2	2.27	31.HN	8.1	34.HA2	4.09
27.HB1	2.36	31.CA	54.82	34.HA1	3.25
27.C	177.13	31.HA	3.67	34.C	173.19
28.N	112.9	31.CB	30.4	35.N	119.02
28.HN	8.07	31.HB2	-0.85	35.HN	7.42
28.CA	53.9	31.HB1	0.77	35.CA	58.83
28.HA	4.69	31.CG	24.72	35.HA	4.36
28.CB	43.55	31.HG2	-0.38	35.CB	41.42
28.HB2	2.55	31.HG1	0.28	35.HB	1.79
28.HB1	2.82	31.CD	28.8	35.CG1	26.34
28.C	173.04	31.HD2	0.99	35.HG12	1.36

35.HG11	1.03	38.C	175.82	42.HA	4.09
35.CD1	12.32	39.N	127.97	42.CB	36.47
35.HD11	0.92	39.HN	8.59	42.HB	2.63
35.CG2	17.07	39.CA	53.81	42.CG1	28.33
35.HG21	0.85	39.HA	4.51	42.HG12	1.64
35.C	175.31	39.CB	32.81	42.HG11	1.72
36.N	126.17	39.HB2	1.14	42.CD1	11.15
36.HN	8.88	39.HB1	0.61	42.HD11	0.94
36.CA	58.32	39.CG	26.68	42.CG2	17.15
36.HA	4.52	39.HG2	-0.01	42.HG21	1.07
36.CB	29.52	39.CD	43.12	42.C	178.5
36.HB2	2.77	39.HD2	2.16	43.N	119.7
36.HB1	3.08	39.HD1	1.08	43.HN	9.02
36.HD1	7.29	39.NE	126.15	43.CA	56.47
36.NE1	129.2	39.HE	7.41	43.HA	4.4
36.HE1	10.15	39.C	173.89	43.CB	43.51
36.HZ2	7.2	40.N	124.98	43.HB2	1.7
36.C	176.3	40.HN	8.05	43.HB1	1.95
37.N	123.14	40.CA	53.13	43.CG	27.49
37.HN	9.37	40.HA	5.48	43.HG	1.62
37.CA	55.15	40.CB	42.89	43.CD1	22.54
37.HA	5.28	40.HB2	2.53	43.HD11	0.91
37.CB	34.73	40.HB1	2.85	43.CD2	24.85
37.HB2	2.02	40.ND2	118	43.HD21	0.23
37.HB1	1.97	40.HD21	9.98	43.C	177.46
37.CG	34.8	40.HD22	7.82	44.N	115.44
37.HG2	2.29	40.C	173.6	44.HN	7.76
37.NE2	107.5	41.N	125.11	44.CA	59.45
37.HE21	6.54	41.HN	9.97	44.HA	4.85
37.HE22	6.21	41.CA	54.28	44.CB	31.06
37.C	173.78	41.HA	4.57	44.HB	2.71
38.N	118.65	41.CB	37.66	44.CG2	21.02
38.HN	10.06	41.HB2	2.86	44.HG21	1.48
38.CA	60.83	41.HB1	2.6	44.CG1	23.3
38.HA	4.83	41.ND2	175.8	44.HG11	1.28
38.CB	35.39	41.HD21	7.24	44.C	172.45
38.HB	1.75	41.HD22	6.74	45.CA	64.67
38.CG2	21.86	41.C	176.62	45.HA	4.79
38.HG21	0.5	42.N	122.32	45.CB	33.31
38.CG1	22.98	42.HN	7.59	45.HB2	2.43
38.HG11	0.68	42.CA	62.65	45.HB1	1.91

45.CG	27.98	49.CB	71.77	53.HA	5.31
45.HG2	2.2	49.HB	4.09	53.CB	43.41
45.HG1	1.78	49.CG2	21.69	53.HB2	3.54
45.CD	50.78	49.HG21	1.12	53.HB1	2.19
45.HD2	3.36	49.C	172.95	53.ND2	108.6
45.HD1	3.48	50.N	123.66	53.HD21	8.51
45.C	176.85	50.HN	7.8	53.HD22	6.64
46.N	108.03	50.CA	60.44	53.C	173.58
46.HN	7.18	50.HA	3.7	54.N	106.53
46.CA	60.22	50.CB	28.89	54.HN	7.97
46.HA	4.51	50.HB2	2.99	54.CA	45.59
46.CB	70	50.HB1	3.36	54.HA2	4.07
46.HB	4.11	50.HD1	7.45	54.C	172.38
46.CG2	19.55	50.HE1	10.18	55.N	123.85
46.HG21	0.83	50.HZ2	7.43	55.HN	8.27
46.C	171.2	50.HH2	6.68	55.CA	60.12
47.N	118.48	50.HZ3	6.75	55.HA	3.79
47.HN	8.26	50.HE3	7.53	55.CB	41.4
47.CA	55.23	50.C	179.18	55.HB	1.27
47.HA	4.13	51.N	114.51	55.CG1	28.71
47.CB	39.68	51.HN	7.99	55.HG12	0.59
47.HB2	2.45	51.CA	65.08	55.HG11	1.92
47.HB1	2.7	51.HA	3.94	55.CD1	16.35
47.C	173.89	51.CB	31.6	55.HD11	0.24
48.N	117.98	51.HB	2.01	55.CG2	17.15
48.HN	8.05	51.CG2	21.73	55.HG21	0.77
48.CA	59.79	51.HG21	1.01	55.C	175.66
48.HA	3.99	51.CG1	22.01	56.N	131.75
48.CB	39.89	51.HG11	0.97	56.HN	9.15
48.HB2	2.66	51.C	176.53	56.CA	53.45
48.HB1	1.26	52.N	115.61	56.HA	4.3
48.HD1	7.43	52.HN	7.06	56.CB	16.86
48.HE1	7.26	52.CA	55.8	56.HB1	1.31
48.HZ	7.28	52.HA	5.01	56.C	178.06
48.HE2	7.26	52.CB	42.06	57.N	126.71
48.HD2	7.43	52.HB2	2.56	57.HN	8.75
48.C	176.88	52.HB1	2.34	57.CA	54.15
49.N	121.12	52.C	176.58	57.HA	3.63
49.HN	8.62	53.N	113.04	57.CB	18.81
49.CA	59.59	53.HN	7.86	57.HB1	1.19
49.HA	4.11	53.CA	53.6	57.C	179.27

58.N	113.61	62.N	128.49	66.HA	4.72
58.HN	8.55	62.HN	9.62	66.CB	38.64
58.CA	55.87	62.CA	57.51	66.HB2	2.92
58.HA	4.49	62.HA	4.64	66.HB1	2.69
58.CB	41.08	62.CB	28.23	66.CG	177.24
58.HB2	2.76	62.HB2	1.57	66.ND2	110
58.HB1	2.35	62.HB1	2.12	66.HD21	7.52
58.C	177.15	62.CG	36.82	66.HD22	6.82
59.N	117.64	62.HG2	2.87	66.C	176.61
59.HN	8.02	62.C	176.78	67.N	108.58
59.CA	55.14	63.N	119.97	67.HN	8.21
59.HA	5.1	63.HN	8.07	67.CA	45.54
59.CB	42.18	63.CA	58.65	67.HA2	4.22
59.HB2	2.52	63.HA	5.4	67.HA1	3.21
59.HB1	3.12	63.CB	37.77	67.C	173.18
59.C	174.65	63.HB	1.89	68.N	117.05
60.N	111.1	63.CG2	22.58	68.HN	7.41
60.HN	7.56	63.HG21	1.1	68.CA	62.76
60.CA	58.45	63.CG1	21.7	68.HA	4.15
60.HA	5.19	63.HG11	0.77	68.CB	69.38
60.CB	32.91	63.C	176.2	68.HB	4.34
60.HB	2.48	64.N	110.1	68.CG2	21.39
60.CG2	18.55	64.HN	8.71	68.HG21	1.18
60.HG21	0.58	64.CA	60.82	68.C	174.05
60.CG1	21.43	64.HA	4.57	69.N	126.83
60.HG11	0.37	64.CB	71.53	69.HN	8.84
60.C	174.65	64.HB	4.09	69.CA	57.35
61.N	121.46	64.CG2	22.15	69.HA	4.56
61.HN	9.27	64.HG21	1.25	69.CB	30.63
61.CA	60.09	64.C	176.51	69.HB1	0.84
61.HA	4.38	65.N	113.36	69.CG	27.94
61.CB	41.48	65.HN	8.87	69.HG2	1.71
61.HB	1.77	65.CA	60.74	69.CD	43.59
61.CG1	27.63	65.HA	2.74	69.HD2	3.18
61.HG12	1.02	65.CB	62.19	69.HD1	2.74
61.HG11	1.36	65.HB2	3.48	69.C	177.27
61.CD1	12.37	65.HB1	3.54	70.N	116.65
61.HD11	0.77	65.C	178.83	70.HN	7.63
61.CG2	17.44	66.N	115.5	70.CA	61.01
61.HG21	0.9	66.HN	7.37	70.HA	4.43
61.C	175.71	66.CA	51.85	70.CB	68.98

70.HB	4.54	75.HN	8.09	78.N	116.16
70.CG2	21.5	75.CA	55.41	78.HN	9.06
70.HG21	1.02	75.HA	4.1	78.CA	45.26
70.C	175.99	75.CB	42.33	78.HA2	4.34
71.CA	60.45	75.HB2	1.38	78.HA1	3.48
71.HA	4.32	75.HB1	1.1	78.C	174.25
71.CB	63.25	75.CG	27.07	79.N	108.19
71.HB2	3.86	75.HG	1.47	79.HN	8.11
71.C	175.84	75.CD1	24.2	79.CA	43.68
72.N	118.01	75.HD11	0.65	79.HA2	4.2
72.HN	8.44	75.CD2	25.76	79.HA1	4.06
72.CA	55.19	75.HD21	0.64	79.C	171.14
72.HA	4.47	75.C	175.49	80.N	116.01
72.CB	40.34	76.N	117.49	80.HN	8.47
72.HB2	2.99	76.HN	8.09	80.CA	56.97
72.HB1	2.63	76.CA	53.6	80.HA	5.6
72.C	175.75	76.HA	4.38	80.CB	42.28
73.N	113.13	76.CB	33.12	80.HB2	2.84
73.HN	7.67	76.HB2	2.16	80.HB1	2.68
73.CA	55.51	76.HB1	2.28	80.HD1	6.99
73.HA	3.67	76.CG	32.3	80.HE1	7.24
73.CB	27.66	76.HG2	1.87	80.HE2	7.24
73.HB2	2.2	76.CD	180.89	80.HD2	6.79
73.HB1	2.16	76.NE2	112.04	80.C	174.89
73.CG	33.65	76.HE21	7.54	81.N	115.77
73.HG2	1.85	76.HE22	6.74	81.HN	9.68
73.NE2	111.6	76.C	175.05	81.CA	56.34
73.HE21	6.84	77.N	123.33	81.HA	5.88
73.HE22	6.76	77.HN	8.87	81.CB	42.13
73.C	174.18	77.CA	58.84	81.HB2	2.98
74.N	121.78	77.HA	3.77	81.HB1	2.73
74.HN	7.79	77.CB	32.13	81.HD1	6.89
74.CA	61.21	77.HB2	1.78	81.HE1	6.79
74.HA	4.02	77.CG	26.02	81.HE2	6.79
74.CB	32.29	77.HG2	1.38	81.HD2	6.89
74.HB	1.87	77.CD	29.87	81.C	173.01
74.HG21	0.7	77.HD2	1.95	82.N	110.5
74.CG1	21.14	77.HD1	1.89	82.HN	9.07
74.HG11	1.1	77.CE	41.97	82.CA	58.54
74.C	174.74	77.HE2	3.14	82.HA	5.06
75.N	125.8	77.C	177.54	82.CB	33.4

82.HB	2.44	85.HD2	3.82	89.HG2	1.34
82.CG2	22.14	86.N	112.68	89.CD	28.63
82.HG21	0.84	86.HN	7.97	89.HD2	1.58
82.CG1	20.36	86.CA	54.51	89.CE	41.88
82.HG11	1.14	86.HA	4.49	89.HE2	2.88
82.C	176.21	86.CB	38.14	89.HE1	2.95
83.N	125.53	86.HB2	2.82	89.C	177.74
83.HN	9.48	86.HB1	2.67	90.N	108.12
83.CA	61.27	86.ND2	113.9	90.HN	7.36
83.HA	4.08	86.HD21	7.78	90.CA	58.1
83.CB	38.22	86.HD22	7.08	90.HA	4.39
83.HB	1.89	86.C	175.95	90.CB	64.41
83.CG1	29.05	87.N	118.58	90.HB2	3.72
83.HG12	1.92	87.HN	7.99	90.HB1	3.85
83.HG11	0.75	87.CA	52.74	90.C	171.84
83.CD1	11.88	87.HA	5.24	91.N	121.34
83.HD11	0.89	87.CB	37.35	91.HN	8.44
83.CG2	15.98	87.HB2	3.66	91.CA	62.37
83.HG21	1.43	87.HB1	2.9	91.HA	4.2
83.C	174.02	87.ND2	111.6	91.CB	34.58
84.N	121.67	87.HD21	8.89	91.HB	1.8
84.HN	7.96	87.HD22	6.71	91.CG2	22.08
84.CA	50.57	87.C	174.41	91.HG21	0.64
84.HA	4.58	88.N	119	91.CG1	23.42
84.CB	39.03	88.HN	6.95	91.HG11	0.7
84.HB2	2.38	88.CA	61.72	91.C	175.37
84.HB1	3.08	88.HA	3.93	92.N	114.06
84.CG	177.25	88.CB	31.93	92.HN	8.46
84.ND2	111.7	88.HB	1.9	92.CA	44.79
84.HD21	8.01	88.CG2	24.57	92.HA2	3.95
84.HD22	7.03	88.HG21	0.75	92.HA1	4.17
84.C	178.64	88.CG1	21.5	92.C	172.44
85.CA	64.96	88.HG11	0.41	93.N	115.52
85.HA	4.14	88.C	174.95	93.HN	8.24
85.CB	32.81	89.N	128.06	93.CA	61.19
85.HB2	2.33	89.HN	9.02	93.HA	4.4
85.HB1	0.75	89.CA	57.84	93.CB	70.02
85.CG	28.12	89.HA	4.12	93.HB	4.07
85.HG2	2.03	89.CB	32.23	93.CG2	21.58
85.HG1	2.15	89.HB2	1.73	93.HG21	1.38
85.CD	50.81	89.CG	24.77	93.C	180.58

94.CA	62.68	98.HN	8.7	103.CB	32.7
94.HA	4.89	98.CA	60.26	103.HB2	3.34
94.CB	32.46	98.HA	4.25	103.HB1	2.99
94.HB2	2.17	98.CB	62.84	103.HD1	7.06
94.HB1	1.94	98.HB2	3.63	103.HE1	10.5
94.CG	28.34	98.HB1	3.73	103.HZ2	7.25
94.HG2	1.1	98.C	176.21	103.HH2	7.49
94.CD	51.32	99.CA	45.89	103.HZ3	6.96
94.HD2	3.82	99.HA2	3.97	103.HE3	7.25
94.HD1	4.11	99.C	174.9	103.C	174.48
94.C	176.34	100.N	107.44	104.N	124.21
95.N	121.39	100.HN	8.45	104.HN	9.54
95.HN	9.59	100.CA	45.53	104.CA	50.22
95.CA	53.5	100.HA2	4.16	104.HA	4.94
95.HA	4.85	100.HA1	3.82	104.CB	23.37
95.CB	35.85	100.C	173.86	104.HB1	1.11
95.HB2	2.17	101.N	120.44	104.C	175.36
95.CG	31.24	101.HN	7.21	105.N	121.64
95.HG1	2.73	101.CA	53.9	105.HN	8.65
95.SD	2.78	101.HA	4.27	105.CA	55.99
95.C	175.41	101.CB	42.7	105.HA	3.54
96.N	124.57	101.HB2	0.98	105.CB	28.44
96.HN	8.69	101.HB1	1.61	105.HB2	0.56
96.CA	56.94	101.CG	26.89	105.HB1	1.37
96.HA	4.53	101.HG	1.63	105.CG	34.04
96.CB	32.36	101.CD1	22.33	105.HG2	0.67
96.HB2	1.75	101.HD11	0.72	105.HG1	0.99
96.CG	24.8	101.CD2	25.83	105.NE2	111.2
96.HG2	1.38	101.HD21	0.8	105.HE21	6.76
96.CD	29.1	101.C	175.93	105.HE22	6.84
96.HD2	1.6	102.N	115.61	105.C	175.06
96.CE	41.97	102.HN	7.98	106.N	134.33
96.HE2	2.93	102.CA	58.38	106.HN	9.64
96.C	177.48	102.HA	4.86	106.CA	61.4
97.N	113.59	102.CB	63.83	106.HA	4.17
97.HN	8.83	102.HB2	3.43	106.CB	32.65
97.CA	44.78	102.C	172.65	106.HB	2.03
97.HA2	4.45	103.N	124.52	106.CG2	21.26
97.HA1	3.71	103.HN	9.08	106.HG21	0.94
97.C	173.93	103.CA	55.6	106.CG1	21.96
98.N	122.45	103.HA	5.24	106.HG11	-0.02

106.C	173.43	111.N	108.83	115.HD1	7.17
107.N	124.88	111.HN	9.04	115.NE1	129.23
107.HN	8.37	111.CA	45.89	115.HZ2	7.36
107.CA	52.67	111.HA2	4.14	115.HE3	7.41
107.HA	5	111.HA1	3.79	115.C	176.02
107.CB	40.07	111.C	173.94	116.N	120.19
107.HB2	2.48	112.N	108.12	116.HN	8.99
107.HB1	2.7	112.HN	7.64	116.CA	53.43
107.CG	175.01	112.CA	45.39	116.HA	5.42
107.ND2	107.8	112.HA2	4.04	116.CB	47.79
107.HD21	7.06	112.HA1	3.83	116.HB2	1.7
107.HD22	6.69	112.C	174.32	116.HB1	1.52
107.C	173.74	113.N	121.33	116.CG	27.04
108.N	125.8	113.HN	8.44	116.HG	1.58
108.HN	8.96	113.CA	53.4	116.CD1	23.7
108.CA	56.39	113.HA	5.29	116.HD11	1.15
108.HA	5.64	113.CB	39.29	116.HD21	0.76
108.CB	42.97	113.HB2	2.68	116.C	176.02
108.HB2	2.96	113.HB1	2.13	117.N	122.69
108.HB1	3.67	113.CG	176.1	117.HN	9.48
108.HD1	7.45	113.ND2	112.1	117.CA	52.03
108.HE1	7.2	113.HD21	7.41	117.HA	4.78
108.C	178.77	113.HD22	6.65	117.CB	37.67
109.N	117.36	113.C	175.37	117.HB2	1.13
109.HN	9.24	114.N	126.64	117.HB1	2.5
109.CA	67.82	114.HN	9.38	117.CG	175.03
109.HA	3.79	114.CA	61.61	117.ND2	111
109.CB	68.94	114.HA	4.35	117.HD21	7.52
109.HB	4.14	114.CB	37.57	117.HD22	6.29
109.CG2	17.2	114.HB	2.16	117.C	172.11
109.HG21	1.23	114.CG2	21.48	118.N	118.76
109.C	175.7	114.HG21	1.02	118.HN	7.69
110.N	118.05	114.CG1	26.5	118.CA	60.3
110.HN	9.38	114.HG11	0.97	118.HA	4.62
110.CA	65.77	114.C	175.3	118.CB	68.99
110.HA	4.34	115.N	127.62	118.HB	3.94
110.CB	67.45	115.HN	8.96	118.CG2	18.7
110.HB	4.16	115.CA	57.27	118.HG21	0.83
110.CG2	21.98	115.HA	5.28	118.C	171.21
110.HG21	1.18	115.CB	30.37	119.N	107.87
110.C	177.84	115.HB2	2.92	119.HN	7.16

119.CA	61.62
119.HA	4.52
119.CB	70.3
119.HB	4.38
119.HG21	1.18
119.C	175
120.N	113.35
120.HN	7.78
120.CA	57.94
120.HA	4.52
120.CB	64.76
120.HB2	3.83
120.HB1	4.22
120.C	173.07
121.N	122.25
121.HN	9.27
121.CA	60.99
121.HA	3.81
121.CB	32.33
121.HB2	1.72
121.CG	26
121.HG2	1.27
121.HG1	1.05
121.CD	29.96
121.HD2	1.59
121.CE	42.25
121.HE2	2.91
121.C	176.94
122.N	117.93
122.HN	8.49
122.CA	57.3
122.HA	4.65
122.CB	40.44
122.HB2	2.62
122.C	179.24
123.N	121.46
123.HN	8.3
123.CA	57.81
123.HA	4.5
123.CB	40.88
123.HB2	2.9
123.HB1	2.6
123.CG	177.21
123.ND2	111
123.HD21	7.82
123.HD22	7.1
123.C	175.71
124.N	122.16
124.HN	8.08
124.CA	58.29
124.HA	3.99
124.CB	42.38
124.HB2	1.94
124.CG	28.26
124.HG	0.84
124.CD1	27
124.HD11	1.49
124.CD2	24.89
124.HD21	0.92
124.C	178.06
125.N	113.3
125.HN	8.28
125.CA	57.41
125.HA	4
125.CB	42.15
125.HB2	1.22
125.HB1	0.43
125.CG	26.57
125.HG	1.71
125.CD1	23.97
125.HD11	0.79
125.CD2	23.15
125.C	178.38
126.N	112.13
126.HN	7.8
126.CA	57.91
126.HA	4.88
126.CB	40.42
126.HB2	2.89
126.HB1	3.4
126.C	176.94
127.N	110.04
127.HN	8.23
127.CA	46.11
127.HA2	4.06
127.HA1	3.93
127.C	173.23
128.N	124.77
128.HN	7.09
128.CA	57.64
128.HA	4.12
128.CB	34.07
128.HB2	1.5
128.HB1	1.82
128.CG	24.9
128.HG2	1.35
128.CD	28.87
128.HD2	1.58
128.CE	42.15
128.HE2	2.88
128.C	180.59