Production, Characterization and Structure Determination of the C-terminal Domain of Stt3p: the Catalytic Subunit of Yeast Oligosaccharyl Transferase

by

Chengdong Huang

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 9, 2010

Keywords: Oligosaccharyl Transferase, NMR, Integral Membrane Protein, Stt3p, Structure Determination

Approved by

Smita Mohanty, Chair, Associate Professor of Chemistry and Biochemistry
Doug Goodwin, Associate Professor of Chemistry and Biochemistry
Peter Livant, Associate Professor of Chemistry and Biochemistry
Orlando Acevedo, Assistant Professor of Chemistry and Biochemistry
Narendra Singh, Professor of Biological Sciences

Abstract

N-glycosylation, the most ubiquitous protein modification in eukaryotes, is catalyzed by the enzyme complex Oligosaccharyl Transferase (OT). This protein co-translational modification has been implicated in a multitude of cellular processes, and defects in the N-glycosylation cause a group of inherited human disorders known as Congenital Disorders of Glycosylation (CDG), while complete loss of N-linked glycosylation is lethal to all eukaryotic organisms.

In the key reaction of N-glycosylation, OT transfers preassembled oligosaccharide moieties from lipid-linked donors onto the asparagine residues in a consensus sequence of Asn-Xaa-Thr/Ser (where Xaa ≠ proline) on nascent polypeptides. For eukaryotes, OT is a remarkably complex multisubunit enzyme that, in the case of the yeast *Saccharomyces cerevisiae*, contains nine nonidentical integral membrane protein subunits, among which Wbp1, Swp1, Ost1, Ost2, and Stt3 proteins are essential for the viability of cells. Although the detailed enzymatic reaction mechanism and the roles of the other subunits are not yet fully understood, a multitude of experimental evidences show that the C-terminal domain of Stt3p is the catalytic domain of the OT complex.

My doctoral dissertation is primarily focused on the following three parts: (1) production, (2) biophysical characterization and (3) 3D structure determination of the C-terminal Stt3p by high-resolution solution NMR.

The C-terminal domain of Stt3p was expressed at 60~70 mg/L in *E. coli* and purified by a robust but novel method which has been developed by our lab, "SDS Elution". Circular Dichroism (CD) and NMR spectra indicate that the C-terminal Stt3p is highly helical and has a stable tertiary structure in SDS micelles. In addition, the comparative analysis of the CD, fluorescence and NMR data of the mutant and the wild-type protein revealed that the replacement of the key residue Asp^{518}, which is located within the $W^{516}WDYG^{520}$ signature motif, led to a distinct tertiary structure, even though both proteins have similar overall secondary structures. This observation strongly suggests that Asp^{518}, which was previously proposed to primarily function as a catalytic residue, also plays a critical structural role. Moreover, the activity of the protein was confirmed by saturation transfer difference (STD) and NMR titration studies.

For NMR structure determination, approximately 93% of the backbone resonances and most of the side-chain resonances have been assigned. To determine the atomic-resolution solution structure of the C-terminal domain of Stt3p, so far the largest α-helical integral membrane protein whose structure is to be determined by NMR, a combination of various constraints have used, including NOEs from {^{15}N, ^{13}C}-double-labeled, partially deuterated (50%) triple-labeled, uniformly {^{2}H, ^{13}C, ^{15}N}-triple-labeled, and ILV methyl protonated otherwise uniformly {^{2}H, ^{13}C, ^{15}N}-triple-labeled sample, together with backbone dihedral angles from chemical shift analysis (TALOS+), residual dipolar couplings (RDCs) and paramagnetic relaxation enhancement (PRE) measurements from 15 nitroxide labeled samples. At the end, we were able to determine the 3D structure of the C-terminal domain of Stt3p. To date, this is the first high-resolution structure of the catalytic domain of the eukaryotic OT complex. Considering

the high sequence homology among eukaryotic Stt3ps, we hope our results can provide a significant step toward the structural understanding of the mechanisms of the N-glycosylation in eukaryotes.

Acknowledgments

I would like to express my highest gratitude to the following people for help and advices throughout this project:

my advisor, Dr. Smita Mohanty, for her supervision and continuous support, her attention in my training as a scientist, her excellent introduction into the field of protein NMR, her general patience and attention to details in reviewing the draft thesis;

my committee members, Dr. Doug Goodwin, Dr. Peter Livant, Dr. Orlando Acevedo, and the outside reader Dr. Narendra Singh, for their encouragements, guidance and support;

Dr. Rajagopalan Bhaskaran for his fruitful discussion and kind helps in the structure calculations;

Dr. Tianzhi Wang for his constructive discussion and introduction into the manipulation of NMR instrument, NMR data processing and analysis.

I would also like to thank many outside professors for their scientific and technical advices: Dr. Chuck Sanders (Vanderbilt University) for advices in membrane protein, PRE and RDC studies, Dr. Ad Bax (NIH) and Dr. Frank Delaglio (NIH) for help in NMR 4D-data processing, Dr. Dr. James Prestegard (University of Georgia) for advice in RDC studies, Dr. Lewis Kay (University of Toronto) and Dr. Vitali Tugarinov (University of Maryland) for suggestions in ILV-sample preparation.

Moreover, I must thank my colleagues for their support, including Dr. Joshua Ring, Dr. Uma Katre (soon-to-be-mom), Dr. David Zoetewey, Dr. Shigeki Saito, Ms. Priscilla Ward (soon-to-be-mom), Mr. Monimoy Banerjee, Mr. Suman Mazumder, and Mr. Mohiuddin Ovee, as well as my friends Mr. Honglei Sun, Dr. Chao Xu, Mr. Yunfeng Li, Dr. Chong Liu, Ms. Qi Chen, Dr. Na yang, Dr. Mi Wang, and Ms. Changyun Zhu.

I would especially like to thanks Dr. Weiya Xu for her advice, encouragement, and moral support.

I am also grateful to the Department of Chemistry and Biochemistry, Auburn University, and ACHE-GRSP (Alabama Commission on Higher Education Graduate Research Scholars Program) Scholarship for their financial support during my research.

At last, I wish to thank my parents for their motivation, patience and constant moral support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

16-DSA	16-Doxyl-Stearic Acid
APS	Ammonium Persulfate
BMRB	Biological Magnetic Resonance Data Bank
CD	Circular Dichroism
CMC	Critical Micellar Concentration
COSY	Correlation Spectroscopy
CSA	Chemical Shift Anisotropy
CSI	Chemical Shift Index
CT	Constant Time
DADMAC	Diallyldimethylammonium Chloride
DDM	n-Dodecyl-β-D-Maltoside
dMTSL	(1-acetyl-2,2,5,5-tetramethyl-η3-pyrroline-3-methyl)-methanethiosulfonate
DPC	Dodecyl Phosphocholine
EDTA	Ethylenediaminetetraacetic Acid
EM	Electron Microscope
ER	Endoplasmic Reticulum
Gd-DTPA	Gd(III)-diethylenetriaminepentaacetic Acid
HSQC	Heteronuclear Single Quantum Coherence
IMP	Integral Membrane Protein
ILV	{Ile(δ_1 only), Leu($^{13}CH_3$, $^{12}CD_3$), Val($^{13}CH_3$, $^{12}CD_3$)} U-{^{15}N, ^{13}C, ^{2}H}

IPAP	In-Phase and Anti-Phase
IPTG	Isopropyl-β-D-Thiogalactopyranoside
LDAO	Lauryl Aimethylamine Oxide
MALDI	Matrix-Assisted Laser Desorption Ionization
MG	Molten Globule
MTSL	(1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-ethanethiosulfonate
MWCO	Molecular Weight Cut Off
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
PDB	Protein Data Bank
ppm	parts per million
PRE	Paramagnetic Relaxation Enhancement
OG	Octyl-β-Glucoside
OT	Oligosaccharyl Transferase
RDC	Residual Dipolar Coupling
RER	Rough Endoplasmic Reticulum
RMSD	Root Mean Square Deviation
SA	Sinapinic Acid
SAG	Strain-induced alignment in polyacrylamide gel
SAIL	Stereo-Array Isotope Labeling
STD	Saturation Transfer Difference
SDS	Sodium Dodecyl Sulfate

SDSL	Site-directed Spin Labeling
TALOS	Torsion Angle Likeness Obtained from Shift and Sequence Similarity
TEMED	N, N, N′,N′- Tetramethylethylene Diamine
TM	Transmembrane
TOCSY	Total Correlation Spectroscopy
TOF	Time of Flight
TROSY	Transverse Relaxation Optimized Spectroscopy

CHAPTER ONE

LITERATURE REVIEW

"*The world of the nuclear spins is a true paradise for theoretical and experimental physicists.*" – Richard R. Ernst, 1992.

1.1 NMR

1.1.1 Introduction

NMR began as a curiosity of physics.

In 1946, the phenomenon of NMR was discovered independently by two physicists, Felix Bloch and Edward M. Purcell, both of whom were awarded the Nobel Prize for this finding in 1952. NMR spectroscopy is based on the fact that some atomic nuclei in a magnetic field absorb radiation at characteristic frequencies. The scientific usefulness of NMR results largely from the fact that nuclei of the same element in different chemical environments give rise to distinct spectral signals. This makes NMR spectroscopy an important method for the observation of the structure and properties of even complex biological macromolecules.

Over the 60 years since its discovery, NMR spectroscopy has gone through two major theoretical breakthroughs, accompanied by plethora of technical improvements in our opinion. The first major theoretical breakthrough was the development of pulse

Fourier transform methods. Here, the radio frequency radiation is applied to the sample in the form of a single short pulse or a sequence of pulses, and the spectrum is obtained by Fourier transformation of the response of the nuclear spins to these pulse programs. This led to a major improvement in the signal-to-noise ratio of NMR. The conception of Fourier transform NMR spectroscopy was brought forward by Richard R. Ernst in 1964 and won him the Nobel Prize in 1992.

Another major theoretical breakthrough was the development of multi-dimensional NMR, in which resonance intensity is recorded as a function of multiple frequency variables. In fact, multiple dimensional NMR is the major conceptual advance in the application of NMR as a method of macromolecule structure determination. Spreading out the signals into multiple dimensions not only produces a tremendous increase in spectral resolution, but also much more correlation information which can be detected and interpreted.

NMR is a versatile technique. In addition to its well-known robust capability for atomic-resolution structure determination, NMR can provide detailed information on conformational dynamics, and both structural and kinetic aspects of interactions of a biomolecules with ligand molecules. For example, NMR can be utilized to characterize the charge state, conformation, and dissociation rates of bound ligands, and to identify contacts between atoms of the ligands and protein. The ability to combine structural and dynamic information is perhaps the most important attribute of NMR in the context of structural molecular biology (1).

Compared to other spectroscopic techniques, such as IR, UV-Vis and Raman,

NMR is rather insensitive. The low sensitivity is indeed the main drawback of NMR spectroscopy. As a result, for bio-NMR structural studies, milligram quantities of pure and homogenous protein are required to obtain sufficiently strong resonances. In addition to the strict requirement of sample concentration, the implementation of heteronuclear multiple dimensional NMR spectroscopy necessitates enriching some isotopes, such as ^{13}C and ^{15}N, which have very low natural abundance. Indeed, for modern NMR structural biologists, sample isotope labeling has evolved into a rather sophisticated technique. The most classic sample labeling approach is to uniformly or partially label the biological macromolecules by isotopes (^{13}C, ^{15}N and ^{2}H). For proteins that can be overexpressed in bacterial systems (especially in *E. coli*), such labeling usually can be readily achieved by growing the organism in minimal media supplemented by addition of $^{15}NH_4Cl$ and ^{13}C-glucose as the sole nitrogen and carbon sources respectively, and using D_2O (deuterium dioxide) as the aqueous medium.

The combination of multi-dimensional NMR and isotope labeling makes it possible for 3D structure determination of proteins of up to medium size ($\leq$ 25 kD). In the last decade, some new selective isotope labeling techniques have been developed, such as Stereo-Array Isotope Labeling (SAIL) (2) and protonation of only the methyl groups of some certain hydrophobic residues (3). The general idea behind these methods is to label certain groups on the side chain of some amino acid residues, and therefore to provide much simpler, which allows larger proteins to be examined.

1.1.2 Basics of NMR

Nuclei of certain isotopes possess intrinsic angular momentum, or spin.

According to a basic principle of quantum mechanics, the maximum experimentally observable component of the angular momentum of a nucleus possessing a spin is a half-integral or integral multiple of $h/2\pi$, where h is Plank's constant. This maximum component of the angular momentum is *I*, the spin quantum number, which is a constant characteristic of the isotope.

As a spinning charge generates a magnetic field, there is a magnetic moment associated with this angular momentum. If $I \neq 0$, the nucleus will possess a magnetic moment, μ, which is always taken as parallel or antiparallel to the angular momentum vector (Eq.1.1):

$$\mu = \gamma h[I(I+1)]^{1/2}/2\pi \qquad \text{(Eq.1.1)}$$

in which γ is the gyromagnetic ratio, a characteristic constant for a given nucleus. The properties of the most important magnetic isotopes for biological molecules are summarized in Table 1.1.

The permitted values of the vector moment along with any chosen axis are described by means of a set of magnetic quantum numbers *m*, which is given by the series (Eq. 1.2):

$$m = I, (I\text{-}1), (I\text{-}2), \ldots, -I. \qquad \text{(Eq. 1.2)}$$

As seen, altogether there are $2I + 1$ possible orientations or states of the nucleus equally spaced with spin quantum number *I*, and each state is associated with a different potential energy - the Zeeman splitting. In the absence of an external magnetic field, these states have the same energy level, zero-field splitting. However, if a uniform magnetic field B_0 is applied, they correspond to states of different

Table 1.1 Properties of some selected nuclei of biological NMR importance. Adapted from reference 4.

Nucleus	I	γ (Ts^{-1})	Natural abundance (%)
^{1}H	1/2	$2.6752 * 10^{8}$	99.99
^{2}H	1	$4.107 * 10^{7}$	0.012
^{13}C	1/2	$6.728 * 10^{7}$	1.07
^{14}N	1	$1.934* 10^{7}$	99.63
^{15}N	1/2	$-2.713 * 10^{7}$	0.37
^{17}O	1/2	$-3.628 * 10^{7}$	0.038
^{19}F	1/2	$2.518 * 10^{8}$	100
^{31}P	1/2	$1.0839 * 10^{8}$	100

potential energy， $m\mu B_0/I$. The energies are shown diagrammatically as a function of magnetic field strength in Figure 1.1, using a nucleus of I = ½ as an example:

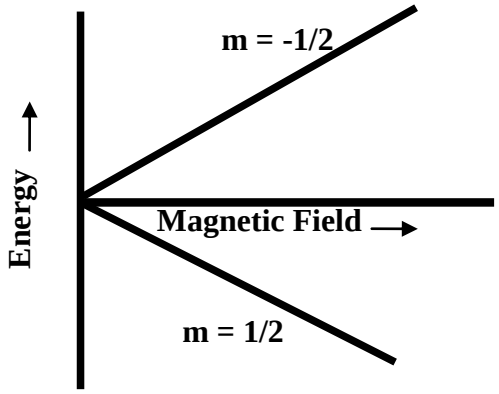


Figure 1.1 Energy splitting as a function of magnetic field strength. Adapted from reference 4.

As with other forms of spectroscopy, the presence of a series of different energy states provides a situation where interaction can take place with electromagnetic radiation of the correct frequency and cause transitions between the energy states. Not all transitions are allowed, while for NMR, the selection rule is $\Delta m = \pm 1$. Thus, the frequency of the electromagnetic radiation can be calculated using the followings Eq. 1.3:

$$\Delta E = h\nu = \mu B_0/I \qquad \text{(Eq. 1.3)}$$

According to the definition of magnetogyric ratio γ, the frequency relation can be written in terms of γ (Eq. 1.4):

$$\nu = \gamma B_0/2\pi \qquad \text{(Eq. 1.4)}$$

The radiation frequency of a nucleus, termed the Larmor frequency, depends only on the applied magnetic field and the nature of the nucleus.

There are several basic but important NMR terms which will occur frequently throughout this dissertation and thus are discussed in detail here.

Chemical Shift: Depending on the local chemical environment, different nuclei in a molecule resonate at slightly different frequencies. The frequency shift of a particular nucleus is called its chemical shift. Chemical shift is customarily given as a fraction of the applied magnetic field, in parts per million (ppm) and is measured relative to the chemical shift of a standard compound. For the nuclei ^{1}H and ^{13}C, tetramethylsilane (TMS) is commonly used as a reference.

Scalar Coupling (or J-coupling) and Dipolar Coupling: There are two important interactions between pairs of nuclei: the scalar through-bond electron-mediated spin-spin interaction, called scalar coupling or J-coupling, and the through-space magnetic dipolar interaction, called dipolar coupling.

Scalar coupling arises from the interaction of different spin states through the network of chemical bonds connecting the coupled nuclei and results in the splitting of NMR signals. Scalar coupling is propagated by the interaction of nuclear spin with the spins of bonding electrons. Consider a nucleus A with a spin $I = 1/2$. Nucleus A can occupy either of two spin states, $m = 1/2$ or $m = -1/2$. Electrons that reside in bonding orbitals overlapping with nuclear spin A will be affected by the spin state of A, and the electron spin states will change slightly in energy in response to the spin of the nucleus. This perturbation of electronic spin states can be propagated to another nucleus (nucleus B) if nucleus B also overlaps with the affected orbitals. This results in a slight change in the resonance frequency of nucleus B depending on whether

nucleus A is in the m = 1/2 or the m = -1/2 state, and nucleus A and B are said to be J-coupled. Coupling is a mutual interaction, *i.e.* if nucleus A is coupled to B, nucleus B is also coupled to A. The frequency difference between the split signal lines is called J-coupling constant and is usually designated as J_{AB}.

Scalar coupling is extremely useful for the NMR spectroscopist. For instance, the coupling pattern can be utilized to provide detailed insight into the connectivity of atoms in a molecule. Moreover, the three-bond J-couplings can be used as a measure for the dihedral angle about the central bond. A more important use for scalar coupling is that it makes possible for the coherence transfer for multi-dimensional NMR experiments. In NMR spectroscopy, the phenomenon of exchange of nuclear spins magnetization though direct and indirect spin-spin interactions are called coherence transfer (or magnetization transfer or polarization transfer). In fact, coherence transfer via J-couplings is a basic concept and routinely used in many multi-dimensional NMR experiments. The J-coupling constants of importance for protein NMR are listed in Figure 1.2.

Unlike chemical shielding, the magnitude of scalar coupling depends only on the interaction of the nuclear magnetic dipoles so it does not vary with the spectrometer field. J-coupling constant does not, therefore, vary from different instruments, and is a property of the molecular structure.

Dipolar Coupling: In addition to scalar coupling, a through-bond interaction, there is another important interaction, dipolar coupling, which is a direct through-space interaction between nuclear spins. In anisotropic media such as solids and

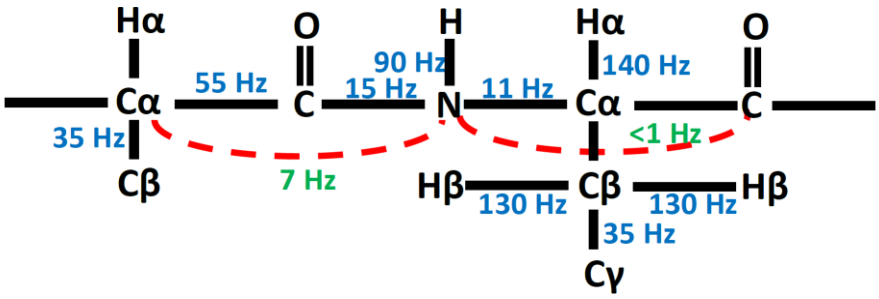


Figure 1.2 Typical J-coupling constants between ^{1}H, ^{15}N, and ^{13}C along a polypeptide chain. These J-coupling constants are very useful in multi-dimensional protein NMR studies. Aadapted from reference 5.

oriented phases (liquid crystals, bicelles, *etc.*), splitting caused by dipolar coupling can be observed directly, and can take a value as large as thousands of Hertz. Nonetheless, in isotropic liquids, in which motion of molecules allows vectors to sample directions uniformly in space, splitting due to dipolar coupling is not observable since it averages to zero with time (see Chapter 5 for theoretical details). Despite this, dipolar coupling is still important for solution NMR in a variety of phenomena and has a couple of significant consequences. The first is that most spin-spin relaxation is primarily mediated by dipolar coupling. As such, the magnitude of dipolar coupling will dictate a number of experimental parameters and dramatically affect resonance linewidth. The second important consequence of dipolar coupling is the nuclear Overhauser effect (NOE), which is observed experimentally as a change in intensity in the signal of one nucleus when the signal of a nearby nucleus (to which the first is dipolar coupled) is perturbed. This makes it possible to determine the molecular structures and to investigate many other phenomena involving interactions

between molecules. In addition, recently, the use of weakly orienting media and intrinsic magnetic anisotropies have allowed residual dipolar coupling (RDC) to be measured and hence provides an additional important source of structural and dynamic information (6).

Relaxation: In principle, NMR experiments begin from the equilibrium state, in which the populations of the energy levels of the system are defined by the Boltzmann distribution. When the equilibrium is perturbed and the perturbing source is then removed, the system will take a finite time to return to its original equilibrium condition. This returning process is called relaxation. The concept of relaxation with regard to assemblies of magnetically active nuclei is of high importance to understand a considerable number of NMR phenomena. In particular, dipolar cross-relaxation gives rise to the nuclear Overhauser effect (NOE) and makes possible the determination of three dimensional structures by NMR.

For isotropic systems, which are uniform in all directions such as solution, there are two components of the relaxation in the absence of chemical exchange: longitudinal or spin-lattice relaxation (T_1), and transverse or spin-spin relaxation (T_2). Here chemical exchange refers to any process in which a nucleus exchanges between two or more environments in which its NMR parameters (e.g. chemical shift scalar coupling or relaxation) differ. Longitudinal relaxation (T_1) is the mechanism by which the excited magnetization vector returns to its thermal equilibrium state (conventionally shown along the z axis, which is defined as the same direction as the direction of external applied magnetic field). The recovery of longitudinal

magnetization follows an exponential curve (Eq. 1.5):

$$M_t = M_0 [1\text{-}\exp (-t/T_1)] \qquad \text{(Eq. 1.5)}$$

Longitudinal relaxation is due to energy exchange between the spins and surroundings, the lattice, (that is why it is called spin-lattice relaxation) and involves re-distributing the populations of nuclear spin states in order to reach the thermal equilibrium distribution. Once it is complete, thermal equilibrium is re-established, and the energy absorbed from radio frequency (RF) irradiation is released back to the surrounding lattice. Thus, basically, spin-lattice relaxation does not involve change in entropy; rather it is an enthalpy-driven process. Rates of longitudinal relaxation are usually strongly dependent on the magnetic field and higher magnetic field generally leads to a slower T_1.

Transverse or spin-spin relaxation (T_2) is the mechanism by which the excited magnetization vector (conventionally shown in the x-y plane, which is defined as perpendicular to the direction of external applied magnetic field) decays. Similar to that of T_1, the magnitude decay of the magnetic moment in the x-y plane decay can also be described by an exponential curve, which is characterized by the time constant T_2 (Eq. 1.6):

$$M_t = M_0 \exp (-t/T_2). \qquad \text{(Eq. 1.6)}$$

Transverse relaxation, which is caused by spin-spin interaction, results in loss of coherence of the transverse nuclear spin magnetization. As spins move together, their magnetic fields interact, slightly modifying their local magnetic fields. These random fluctuations of the local magnetic field lead to random variations in the

instantaneous NMR precession frequency of the interacting spins. Consequently, transverse relaxation causes cumulative losses in phase and results in transverse magnetization decay. In contrary to longitudinal relaxation, which is an enthalpy-driven process, spin-spin relaxation leads to the loss of phase coherence (order), hence it can be considered as an entropy-driven process. Another distinction is that, unlike T_1, T_2 is generally unrelated to magnetic field.

Both T_1 and T_2 can be determined by NMR experiments and T_2 is always shorter than T_1.

Nuclear Overhauser Effect (NOE): When the resonance of a spin in an NMR spectrum is perturbed by radio frequency radiations, it may cause the spectral intensities of its neighboring spins in the spectrum to change. This phenomenon is called the nuclear Overhauser effect or NOE. The intensity change caused by NOE originates from the population changes of the Zeeman states of coupled spins after perturbation through the dipolar relaxation.

This can be clearly illustrated for a simplified two-spin-1/2 system, in which the two spins (I and S) are coupled only by dipolar interaction and there is no scalar coupling between the spins. As shown in Figure 1.3, the energy diagram for this two-spin system contains four energy states: αα (both spins in lower energy states), βα and αβ (spin I in higher energy state and S in the lower energy state, and vice versa), and ββ (both spins in higher energy states). Therefore, there are two transitions for spin I (αα → βα and αβ → ββ) and two transitions for spin S (αα → αβ and βα → ββ). Upon saturation of one spin, say spin I, the populations of αα is equal to βα, and αβ is equal

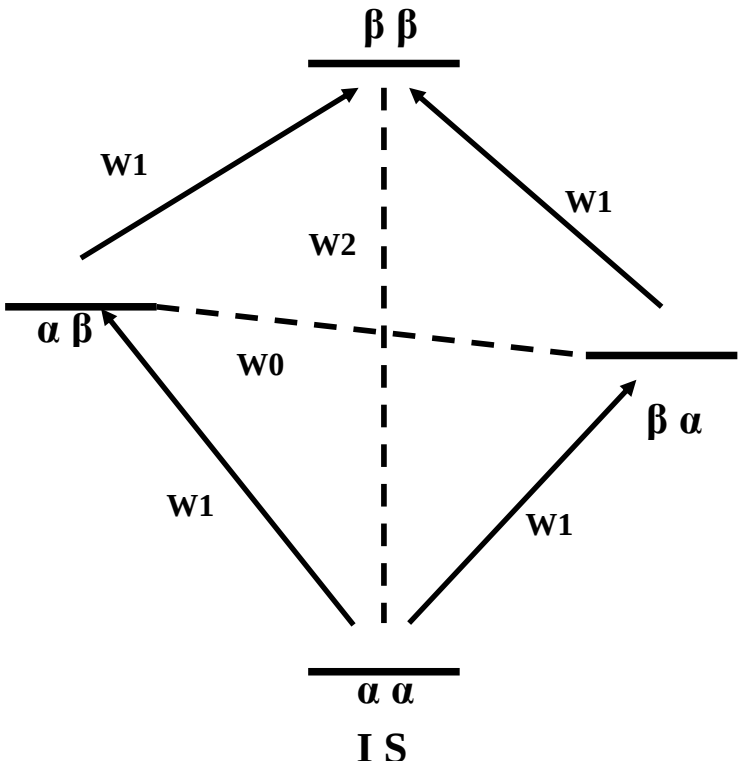


Figure 1.3 Energy diagram for a dipolar-coupled two-spin system. The four states are αα, αβ, βα, and ββ; the zero- single- and double-quantum transitions are represented by W_0, W_1 and W_2, respectively, Drawn according to reference 6.

to ββ. As a result, the population αα and αβ is decreased compared to the equilibrium whereas βα and ββ are more populated. When the radiation source is removed, the system will be recovered to its equilibrium state through all allowable relaxation processes. It is clear that spin-lattice relaxation, W_1 transition, or single quantum transition since $\Delta m = 1$, does not change the state population for spin S. Thus the W_1 can not change the intensity of spin S. However, in addition to the single quantum transition, there are two other relaxation pathways (spin-spin relaxations): W_0 (βα → αβ), or zero-quantum transition since $\Delta m = 0$; and W_2 (ββ → αα), or double-quantum transition since $\Delta m = 2$. If W_2 dominates, as for small molecules whose tumbling

times are short, the population differences between states αα and αβ, as well as βα and ββ, are increased. In other words, the NMR signals intensity for spin S is increased, namely, positive NOE. On the other hand, slow tumbling for large molecules favors W_0 transition, which leads to an intensity reduction for spin S, causing negative NOE. For medium sized molecules with molecular weights of ~1000-3000, the two relaxation pathways are competing in the system, and sometimes the NOE can be very weak or zero. NOE can only be detected when two spins are close in space (usually within 5Å), and its intensity is inversely proportional to r^6, where r is the distance between the two spins. Hence, it is clear that with the increasing of the distance, the intensity of NOE decreases sharply (4).

The fundamentals of the NOE were described very early in the history of NMR in a classic paper by Solomon published in 1955 (7). This paper included the first experimental demonstration of the NOE, which followed Overhauser's original prediction that saturation of electrons in a metal would produce a large polarization of the metal nuclear spins (8, 9). The first paper to demonstrate the power of the NOE in structural studies was demonstrated by Anet and Bourn in 1965 (10). Since then, a major advance of the application of NOE was the introduction of the two-dimensional NOE experiment, NOE spectroscopy (NOESY), which was largely achieved by Ernst's group in early 1980s (11, 12). Today, the NOE plays a central role in modern NMR structural biology.

1.1.3 Multidimensional NMR

The explosive growth in the application of NMR spectroscopy to biological

macromolecules in the past three decades may be attributed mainly to the success of multidimensional experiments.

The first two-dimensional (2D) NMR experiment was proposed by Jean Jeener at an Ampere Summer School in 1971 (13). This was regarded as the forefather of a whole class of 2D experiments. In general, all 2D NMR experiments can be reduced to the same basic conceptual scheme as shown in Figure 1.4. Compared to basic 1D NMR experiment, between the preparation and acquisition periods (t_2), two more elements are introduced for the 2D NMR: the evolution period (t_1), during which the spins are labeled according to their chemical shifts, and the mixing period (M_1), during which the spins are correlated with each other. The experiment is repeated many times with successively (usually linearly) incremented values of the evolution period t_1 to yield a data matrix S (t_1, t_2). Fourier transformation in the t_2 dimension yields a set of 1D spectra in which the intensities of the resonances are sinusoidally modulated as a function of the t_1 duration. Subsequent Fourier transformation in the t_1 dimension yields the desired 2D spectrum S (ω_1, ω_2).

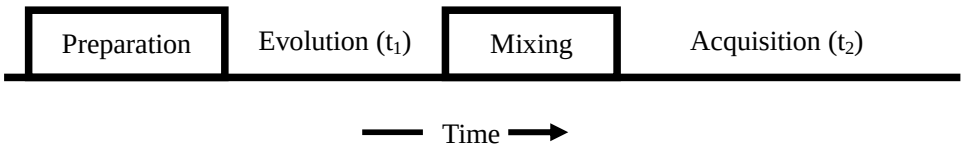


Figure 1.4 General scheme for two-dimensional NMR spectroscopy. Adapted from reference 4.

The 2D NMR experiments of most **important** use are COSY (Correlation spectroscopy), TOCSY (Total Correlation Spectroscopy), NOESY (Nuclear

Overhauser Effect Spectroscopy) and HSQC (Heteronuclear Single Quantum Coherence), among which COSY, TOCSY and NOESY are homonuclear experiments, while HSQC is a heteronuclear experiment.

COSY: COSY was one of the first and simplest multi-dimensional experiments (14). In a COSY experiment, magnetization is transferred through the chemical bonds between protons on adjacent atoms, and it thus provides information about protons connected by J-coupling (Figure 1.5 A).

TOCSY: In a TOCSY experiment, during this pulse sequence, after the evolution period, the magnetization is spin-locked, i.e. the magnetization is kept in the transverse plane for certain amount of time. During this mixing time (spin-lock period) the coherence is transferred through scalar coupling. Consequently, TOCSY creates correlations among all protons within a given spin system, not just between geminal or vicinal protons which are J-coupled to each other as in COSY (Figure 1.5 B).

HSQC: The HSQC experiment was proposed by Bodenhausen and Ruben about 30 years ago (15). In an HSQC experiment, magnetization is transferred from hydrogen nuclei to the directly attached heteronuclei via J-coupling. The chemical shift is evolved on the heteronuclei and the magnetization is then transferred back to the hydrogen nuclei for detection. Therefore, the HSQC experiment is in fact a double INEPT (Insensitive Nuclei Enhanced by Polarization Transfer) experiment, which correlates protons with their directly attached heteronuclei (single-bond correlations), and the resulting 2D spectrum has one axis for proton chemical shift and the other for a heteronucleus chemical shift (most often ^{13}C or ^{15}N) (See Chapter Four and Chapter

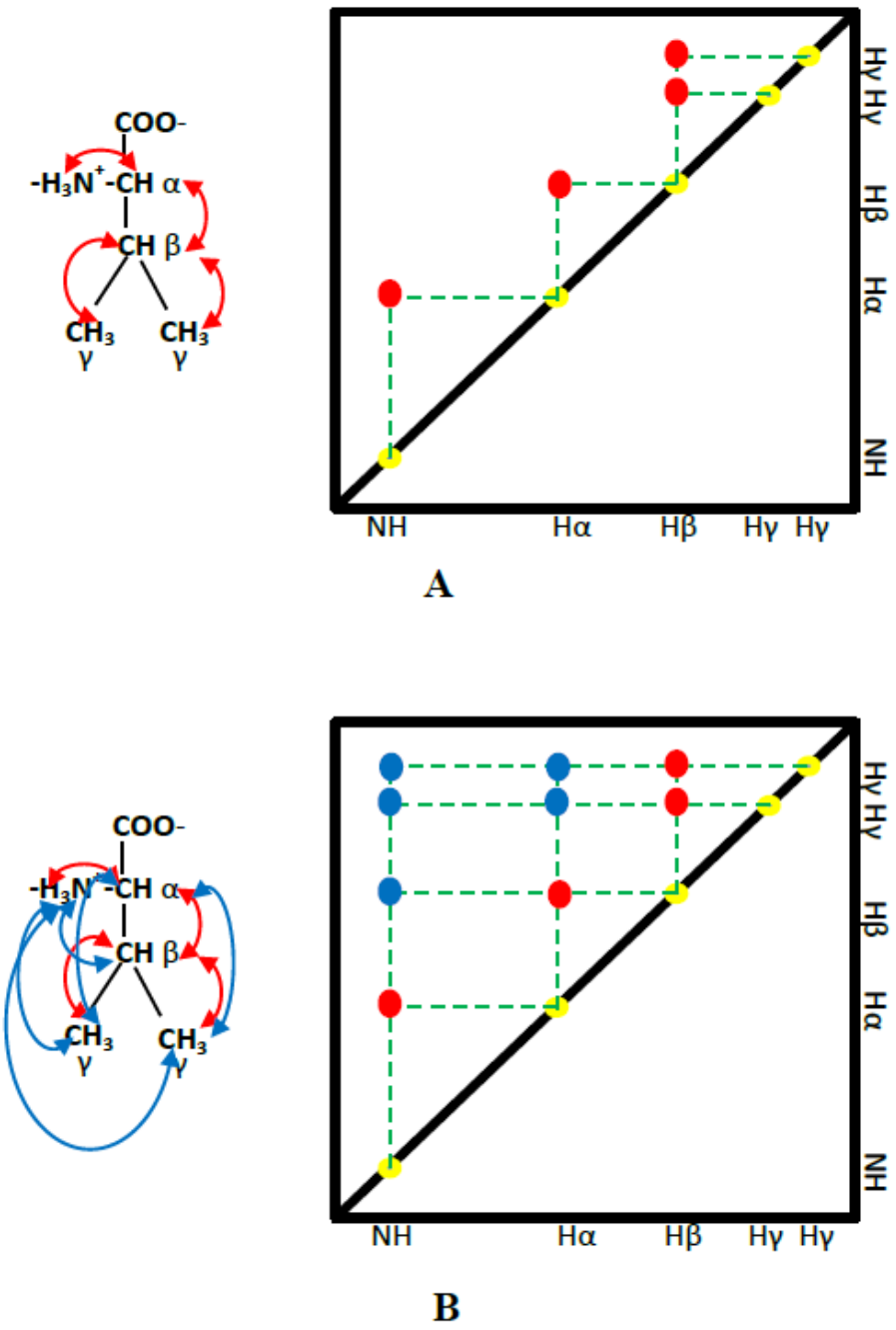


Figure 1.5 Using amino acid Val as an example to show 2D spectra of COSY and TOCSY. Note that both experiments provide diagonally symmetric spectra. Here, for simplicity, only half (upper right part) cross peaks are shown. A: the COSY spectrum shows correlations between protons on adjacent atoms, B: the TOCSY spectrum shows correlations between all protons in the spin system.

Five for Examples of Figures). Since each residue of a protein (except proline) has an amide proton attached to a nitrogen atom in the peptide bond, if no peak overlapping occurs, ideally, the number of peaks in the ^{15}N-HSQC spectrum should match the number of non-proline residues in the protein (though side chains with nitrogen-bound protons will add some additional peaks). Moreover, because {^{1}H, ^{15}N}-HSQC is extremely sensitive to changes (such as pH, temperature, chemical environments, etc.) to the protein sample, it is often called the "fingerprint" of a protein. As a result, ^{15}N-HSQC is one of the most frequently recorded experiments in protein NMR.

NOESY: Unlike the experiments above, which depend on through-bond J-couplings, a NOESY experiment depends only on the spatial proximity between protons. During the mixing time the magnetization is transfered through scalar coupling. As mentioned previously, NOESY is one of the most useful techniques as it allows correlating nuclei through space (distance smaller than 5Å). By measuring cross peak intensity, distance information can be extracted.

Although 2D NMR spectroscopy has proved to be one of the most important developments in modern high-resolution NMR, for macromolecules whose molecular weights are larger than 10 kDa, even the 2D spectra with the best resolution are often insufficient. This makes it necessary to further increase the number of frequency dimensions in the spectrum. In principle (although in practice, it is almost always inevitably much more complicated), 2D NMR experiments can easily be expanded to multidimensional spectroscopy by the appropriate combination of 2D NMR experiments. For example, as illustrated schematically in Figure 1.6, a 3D experiment

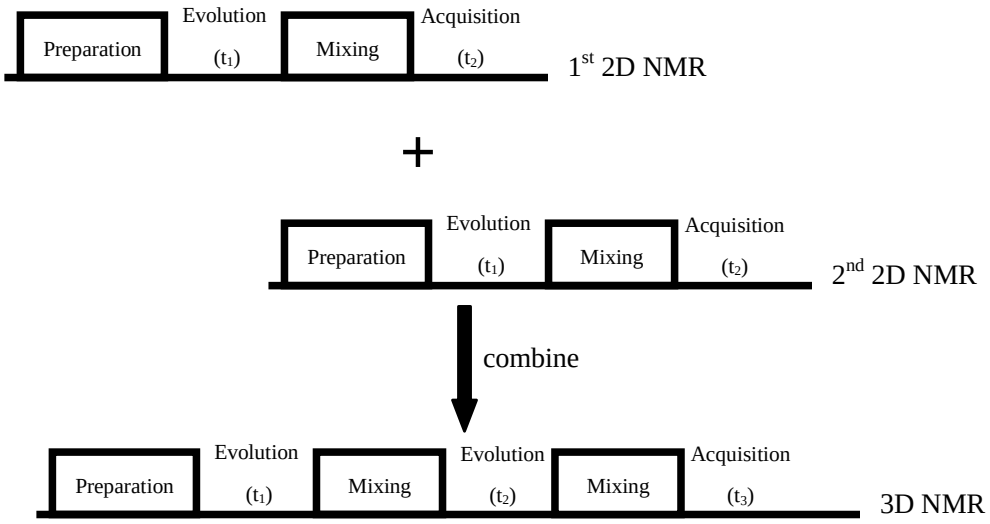


Figure 1.6 Schematic generation of a 3D NMR experiment from the combination of two 2D NMR experiments. The mixing period of the first 2D experiment and the preparation period of a second 2D experiment are combined. The 3D experiment contains three independent time periods. Adapted from reference 4.

can be constructed by two 2D pulse sequences by leaving out the detection period of the first experiment and the preparation pulse for the second. The resulting pulse program comprises two independently incremented evolution period t_1 and t_2. In the same way, a 4D experiment is obtained by combining three 2D experiments (or two 3D experiments) in an analogous fashion. Thus, at least conceptually, n-dimensional NMR experiments can be conceived as a straightforward extension of a series of appropriate 2D NMR experiments. In general, fewer overlaps (and hence fewer ambiguities in resonance interpretation) derive from increasingly higher dimensionality (Figure 1.7), but increasing of the dimensionality also leads to spectra of lower sensitivity and less digital resolution. Therefore, for large biological

molecules, so far the dimensionality of NMR experiments of the most practical use is limited to 3D or 4D.

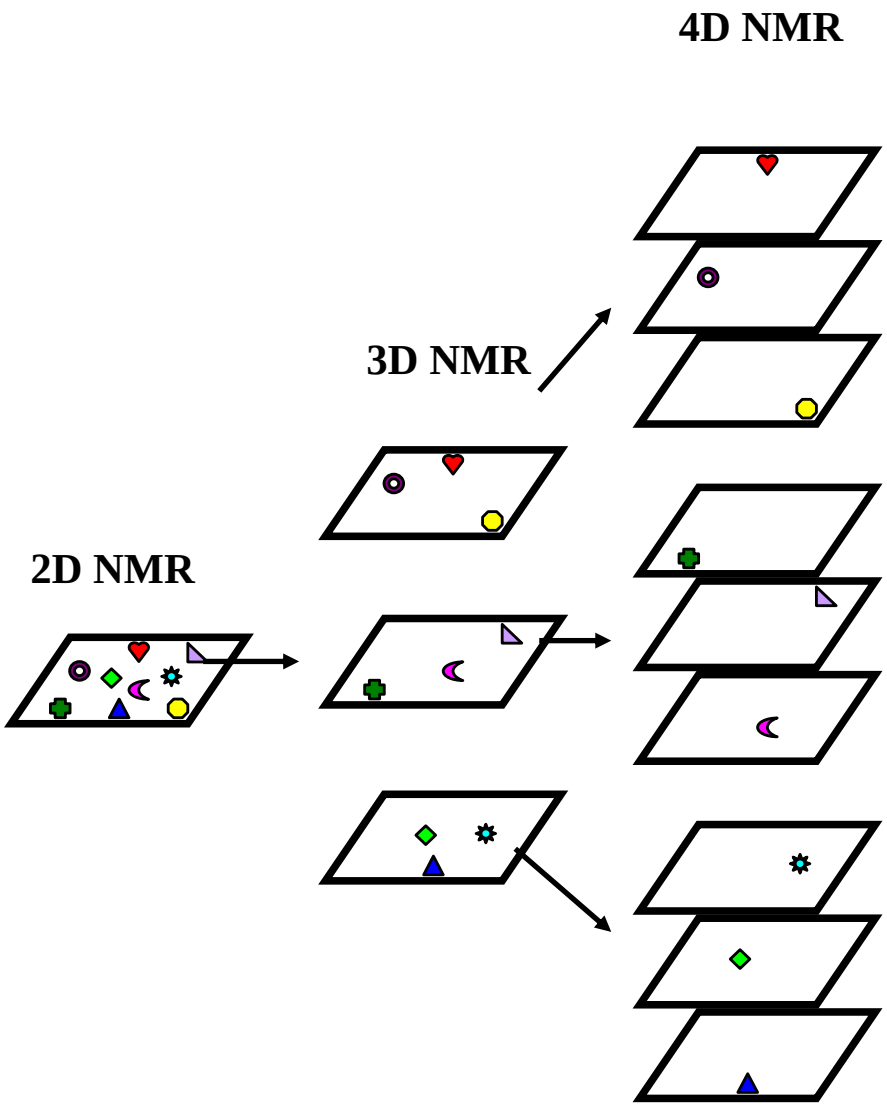


Figure 1.7 The development of a 4D NMR data set from 3D data set and 2D data set. The introduction of an additional evolution period generates a new frequency dimensions and therefore can greatly alleviate spectral ambiguities and overlapping.

1.1.4 Protein NMR

In general, the determination of a NMR solution structure of protein may be dissected into five major parts: (1) sample preparation, (2) recording and processing of NMR data, (3) sequential resonance assignment and side-chain assignments, (4) collection of structural restraints, and (5) NMR structure calculation and refinement. Among which the step (4) and (5) are iterative and may go many cycles before the final structure is determined.

Unlike X-ray crystallography, whose application is limited due to the stochastic nature of crystallization, the sample requirement for NMR spectroscopy is not as harsh. For example, solution NMR is performed on aqueous samples of purified protein, which contains ~ 300 to 600 microlitres of protein sample with a concentration in the range of 0.1 to 3 millimolar. However, due to its insensitivity, NMR also has its major limitations: the molecular size and time constraints. In the last decade, several exciting developments have emerged in the field of high resolution NMR spectroscopy both to extend significantly the molecular weight range and to improve the efficiency of structure determination and the quality of the resulting structures. The availability of cryo-probe (reducing the operating temperature of the NMR coil assembly and the preamplifier) and high magnetic field NMR devices significantly increased the spectral resolution and sensitivity. In 2009, the Bruker Company announced AVANCE 1000, the world's first 1 Gigahertz NMR spectrometer. In addition, transverse relaxation optimized spectroscopy (TROSY) type experiments serve as another milestone (16). In TROSY experiments, only the narrow component

of the ^{15}N-^{1}H or ^{13}C-^{1}H doublet is selected and sharp resonances can be observed for proteins of a molecular weight well beyond 100 kDa.

Before the step of structure calculation, each resonance must be assigned to an individual proton, and then through-space NOE interactions must be assigned (assignments of NOESY spectra). In principle, this can be achieved in a relatively straightforward manner, using correlation experiments to identify resonances belonging to different amino acid types via through-bond connectivities, and then linking these residues sequentially. However, in practice, it is difficult, especially for proteins whose molecular weights are larger than 20 kDa. The reasons are twofold. First, there is an extensive degree of resonance overlap and chemical shift degeneracy. Secondly, large proteins have much slower tumbling and correspondingly rapid transverse relaxation rates. These effects substantially broaden the resonances and make weak resonances harder to detect.

To overcome these problems, heteronuclear 3D NMR experiments are performed, which requires NMR samples to be enriched with ^{13}C and ^{15}N. Because the cost of ^{13}C, ^{15}N and ^{2}H nutrition sources is significantly higher than natural abundance sources, the isotopic labeling of the proteins is usually done in minimal growth media using bacterial expression systems.

In order to conduct the protein sequence-specific resonance assignments, quite a few triple-resonance NMR experiments have been designed, in which three different nuclei, such as ^{1}H, ^{13}C, and ^{15}N are correlated. For backbone assignment, the most commonly used 3D experiments are HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH

(or HN(CO)CACB for perdeuterated protein sample), HNCO, and HN(CA)CO. Because the one-bond (^{1}J) and two-bond (^{2}J) couplings are rather strong (Figure 1.2) and independent of protein conformation, the magnetization transfers through these couplings can efficiently compete with the loss of magnetization caused by short transverse relaxation times during the experiment.

All six experiments for protein backbone assignments mentioned above consist of an [^{15}N, ^{1}H]-HSQC 2D-plane expanded to a ^{13}C third dimension. Among these, the HNCA correlates each amide proton with the Cα chemical shift of its own residue (residue i) and of the residue proceeding in the sequence (residue i-1), while the HN(CO)CA correlates each amide proton only with the Cα chemical shift of the previous residue (residue i-1). Sequential assignment can then be undertaken by matching the shifts of each spin system's own and previous Cα carbons. The HNCO and HN(CA)CO work in a similar manner, just with the carbonyl carbons rather than alpha carbons, and the HNCACB and the CBCA(CO)NH contain the chemical shifts of both the alpha carbon and the beta carbon (see Chapter Four for experimental examples of Figures). For large proteins, all six experiments should be used interactively to confirm the assignments and to rule out the ambiguities resulting from the spectral overlapping. A summary of these six triple resonances experiments and the connectivities observed in them are provided in Table 1-2.

The starting point of protein backbone assignment can be readily made since the Cα and Cβ chemical shifts adopt characteristic values of the amino acid type. For example, for some certain residues such as alanine, serine, threonine and glycine, their

amino acid types are easy to be identified as their Cβ chemical shifts are very different from those of the other amino acids: alanine, serine and threonine have a Cβ of ~ 18 ppm, ~ 63 ppm and ~ 69 ppm, respectively, while glycine has no Cβ with a Cα of ~ 45 ppm.

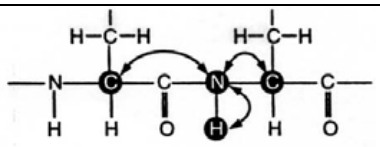
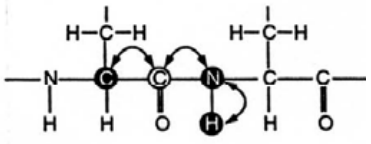
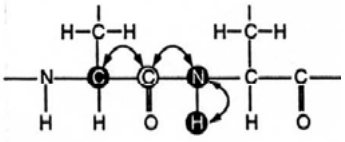
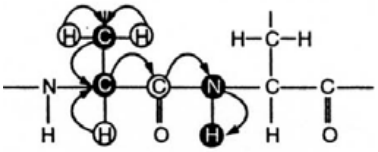
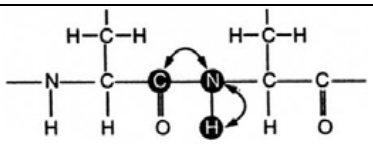
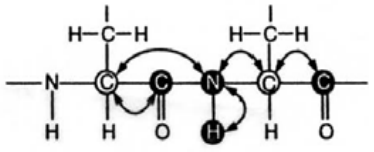
Once the backbone sequential assignment is made, it is rather straightforward to assign the side chains using 3D NMR experiments such as HCCH-TOCSY (Total Correlation Spectroscopy), ^{15}N-HSQC-TOCSY and HCC(CO)NH, etc.

In NMR protein structure determination, the principal source of geometric information lies in inter-proton distance restraints derived from NOE measurement, as well as angular constraints based on coupling constants. The physical basis for NOE has been described earlier. NOE assignment can be achieved by comparison of the chemical shifts of peaks in the NOESY spectrum with those of the backbone and side chains.

The structure calculation can then be performed by providing the experimentally determined distance constraints obtained from the NOESY, and dihedral angular constraints from coupling constants as "input" files to computer programs such as CYANA (17) or XPLOR-NIH (18, 19). The calculations result in an ensemble of structures which, if the data are sufficient to dictate a certain fold, will converge.

Although measurements of NOE will no doubt continue to play an essential role in protein structure determination, some new methods have been introduced recently. For example, for protein samples prepared in dilute, aqueous, liquid-crystal solutions,

Table 1.2 Most commonly used triple-resonance NMR experiments for protein backbone assignment.

Experiment	Correlation	Magnetization Transfer
HNCA	$H^N(i)$, $^HN(i)$, Cα(i) and Cα(i-1)	
HN(CO)CA	$H^N(i)$, $^HN(i)$, Cα(i-1)	
CBCANH (HNCACB)	Cβ(i-1), Cα(i-1), Cβ(i), Cα(i), $H^N(i)$, $^HN(i)$	
CBCA(CO)NH	Cβ(i-1), Cα(i-1), $H^N(i)$, $^HN(i)$	
HNCO	CO(i-1), $H^N(i)$, $^HN(i)$	
HN(CA)CO	CO(i), $H^N(i)$, $^HN(i)$	

their residual dipolar couplings (RDCs) can be used to directly measure the relative orientation of internuclear bond vectors (6). Moreover, by incorporation of the paramagnetic spin-label to the protein sample, the effects of induced Paramagnetic Relaxation Enhancement (PRE) can be converted to long-range distance constraints (20), thus improving the quality of the resulting protein. All of these techniques have been applied to the present project and will be discussed in depth in the following chapters.

1.2 Integral Membrane Proteins

1.2.1 Introduction

Proteins can be divided into two categories based on their solubility in water: globular or water-soluble proteins, and integral membrane proteins (IMPs), which are hydrophobic in nature and insoluble in water. Solublizing agents, such as detergents, are used to render IMPs water soluble (Figure 1.8). By this definition, peripheral membrane proteins don't belong to the category of IMPs because they either associate with the membrane through electrostatic interactions and hydrogen bonding with the hydrophilic domains of integral proteins, or with the polar head groups of membrane lipids. Once released from the biological membrane by relatively mild treatments that interfere with electrostatic interactions or break hydrogen bonds, such as carbonate at high pH, peripheral membrane proteins behave the same as water-soluble proteins.

In nature, native IMPs are embedded in the lipid bilayers of biological membranes. The firm attachment of IMPs to membranes results from hydrophobic interactions between lipid acyl chains and hydrophobic domains of the proteins.

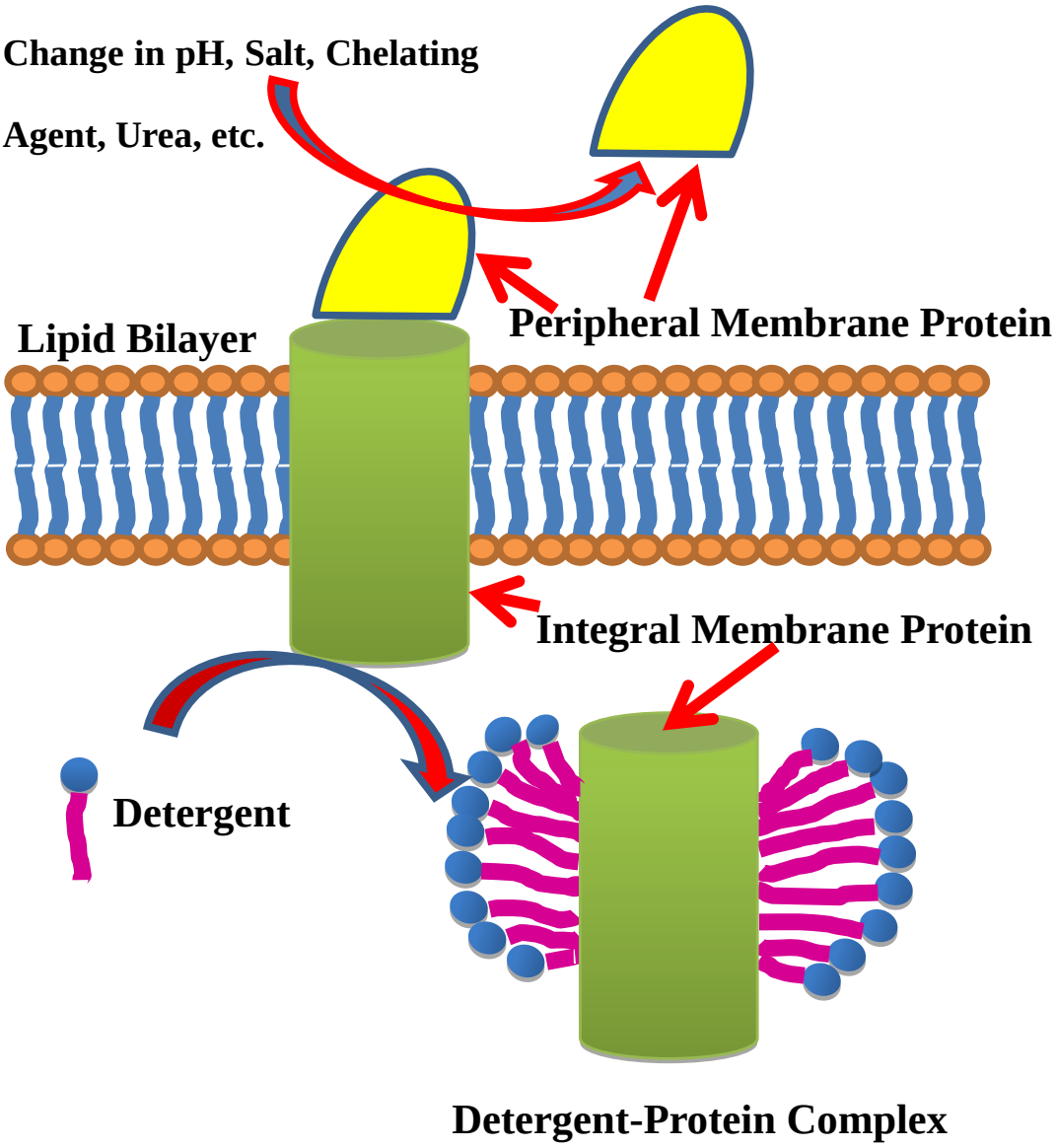


Figure 1.8 Peripheral and integral membrane proteins.

Based on the number of transmembrane (TM) domains, IMPs can be further categorized as single membrane-spanning proteins and multiple membrane-spanning proteins.

IMPs come in two basic architectures: the α-helix bundle and the β-barrel. As implied by their names, helix-bundle membrane proteins are built from long transmembrane α-helices consisting of between 18 – 24 amino acid residues packing together into more or less complicated bundles, whereas the transmembrane domain of β-barrel proteins are large antiparallel β-sheets rolled up into a barrel closed by the first and last strands in the sheet. A striking architectural characteristic of IMPs is, whether helix-bundle or β-barrel, the hydrophobic transmembrane domains are almost always flanked by two "aromatic girdles" composed of Trp and Tyr residues (15, 21–23). This mirrors the structure of the surrounding lipid bilayer, with the lipid headgroup regions contacting the aromatic girdles and the hydrocarbon tail region interacting with the hydrophobic transmembrane (TM) domain. This architectural pattern ensures a seamless fit of IMPs to the biological membrane.

Helix-bundle IMPs are found in all cellular membranes and represent the majority of IMPs, while β-barrel IMPs account for a much smaller percentage. So far, all identified β-barrel IMPs are limited to the outer membrane protein of gram-negative bacteria and roughly estimated to account for 10% of all *E. coli* IMPs (24).

Membrane proteins perform a staggering range of important biological functions, such as energy transduction, material (drugs and nutrients) transport, signal transduction, cell-cell communication, etc. Numerous heritable diseases are associated

with the misassembly of membrane proteins, including the common disorders cystic fibrosis, retinitis pigmentosa, Charcot-Marie-Tooth disease, and hereditary hearing loss (25).

The diverse functions of IMPs require a large variety of membrane proteins to be present in cells. According to the results of various genome projects, it has been estimated that membrane proteins account for between 25 and 30% of all encoded proteins (26), and approximately 70% of all current pharmaceutical targets are membrane proteins (27). Yet despite their importance, membrane proteins currently represent less than 1% of the >54000 structures deposited in the Protein Data Bank (PDB, http://www.rcsb.org/pdb/home/home.do), mainly because of the technical challenges associated with these highly hydrophobic molecules, such as their overexpression, purification, and subsequent structural characterization. As a result, membrane proteins are widely regarded as “the last frontier” or “the wild west” of structural biology (28).

1.2.2 3D Structure Determination of Integral Membrane Proteins

In comparison to soluble proteins, membrane proteins have unique structural and energetic properties as a consequence of their being embedded in lipid bilayer milieu. Distinct biological features are associated with membrane protein biogenesis and trafficking. As mentioned earlier, membrane proteins are involved in many essential cell functions including respiration, photosynthesis, signal transduction, molecular transport and motility. Consequently, membrane proteins are targets for a majority of the currently marketed drugs. Thus a detailed knowledge of their

structures and functions is essential to facilitate the rational design of effective drugs and to develop new therapies for genetic diseases.

In principle, the same techniques that are used to determine the three-dimensional structures of water-soluble proteins can be applied to membrane proteins as well. However, in practice, due to their insolubility in water, *in vitro* studies of membrane proteins are so complicated that in fact, just a few decades ago, conventional wisdom held that it was impossible to determine structures for integral membrane proteins (29). Today we know it is not impossible; it is simply very hard. Membrane protein structure determination is still in its infancy and remains quite an unexplored area in structural biology.

Two events define the beginning of the modern era of membrane-protein biophysics: the determination of the three-dimensional structure of bacteriorhodopsin at low resolution by Richard Henderson and Nigel Unwin in 1975 using electron microscopy (30), and the atomic-resolution structure of the *Rhodopseudomonas viridis* photosynthetic reaction center (at 2.3Å resolution) by Johann Deisenhofer and Hartmut Michel in 1985 (31), twenty seven years after the first water-soluble protein structure, myoglobin, was determined. This pioneering work won Deisenhofer and Michel the Nobel Prize in 1988. Now, over two decades later, the number of unique membrane proteins solved is less than 200. In contrast to the speed at which water-soluble protein structures are solved, the progress for membrane protein structure determination still seems abysmally slow.

By the end of 2009, there were a total of 54432 protein structures deposited to

Protein Data Bank (PDB, http://www.rcsb.org/pdb/home/home.do), out of which only 1057 structures are those of membrane proteins. However, after the removal of redundant protein structures, the comparison becomes even sharper: out of 33975 unique protein structures, only 197 unique structures represent membrane proteins. In another words, less than 0.6% of currently available protein structures belong to membrane proteins. Although it has been claimed that the progress in membrane protein structure determination has started to accelerate in the last decade (32), its future is far from optimistic. For example, the average number of unique membrane protein structures reported annually over the last three years (2006, 2007 and 2008) is only about 25. These numbers not only underscore the importance of membrane proteins, but also emphasize the enormous biochemical and structural work that remains to be done in the field of membrane proteins.

Two major bottlenecks account for this huge disparity: the difficulties in the production of homogeneous membrane protein samples in high yield and the difficulties associated with their structure determination.

High-resolution structural studies require milligram quantities of pure proteins and thus it is important to obtain a high-yield expression system for the production of desired protein. This is especially true with respect to structural studies of IMPs by NMR, since IMPs typically have to be labeled with the stable isotopes ^{2}H, ^{13}C and ^{15}N for multidimensional heteronuclear NMR experiments. Isotope labeling is intrinsically quite expensive, and deuteration often causes a drastic reduction in the yield of protein synthesis due to the negative influence of the deuterated medium on

the cell metabolism. Since the natural abundance of membrane proteins is usually too low to purify sufficient quantity of material for functional and structural studies, currently recombinant expression of membrane proteins in *E. coli* is the primary machine for large-scale protein production. However, even this is notoriously problematic, often resulting in little to no protein expression. In fact, the expression machinery for membrane proteins is so complicated that it is still unclear that why a particular membrane protein can be expressed by some cell lines while can not be expressed at all by the others. Consequently, the screening of high-yield systems for IMPs remains a process of "trial-and-error".

Another major obstacle to membrane proteins production arises from the need to solubilize these proteins in detergent solution and/or organic solvents for purification and further biophysical characterization. An ideal detergent will effectively solubilize and stabilize the membrane proteins in an unaggregated state without causing denaturation, and without interfering with purification or subsequent biophysical characterization. However, currently there is little basic understanding of the detailed interactions between proteins and detergents that could serve as the basis for rationally deciding which detergents would be suitable for use with a particular protein. Thus, the suitable detergent for a particular protein cannot be determined as *a priori*, rather it must be determined by screening a number of detergents and sample conditions. Unfortunately, there are dozens of different detergents commonly used in biochemistry, dozens more that are less well characterized but probably still useful, and many novel detergents currently under development. Moreover, mixtures of

detergents are also used, along with nondetergent additives, which serve very well for many membrane proteins (33). Therefore, the size of the detergent parameter space is very large indeed and the screening of detergents is almost always a lengthy process.

In addition to membrane protein expression and detergent screening, the purification of membrane proteins is much more complicated due to the presence of detergents. Although the methods for purification of water-soluble proteins are very well established, these methods cannot necessarily be applied in a straightforward manner to membrane proteins due to their hydrophobic nature.

The second bottleneck is the hardship of structure determination for membrane proteins. Currently, X-ray crystallography and NMR spectroscopy are the only two available techniques for atomic-resolution structure determination of proteins. As is the case for soluble proteins, most structures of membrane proteins have been solved by X-ray crystallography, which is still regarded as “the cornerstone of structural biology” (34). However, despite its relative success, structure determination of membrane proteins by X-ray crystallography must still be considered a high art and the preparation of diffraction-quality crystals remains the major bottleneck in the pursuit of high-resolution structures of membrane proteins. This is mainly due to the presence of essential lipids, or their mimic detergents, which dramatically complicates and thus makes it particularly difficult to prepare diffraction-quality crystals. As a result, practically, the search for appropriate crystallization conditions must sample a much larger space than a typical soluble protein crystallization screen. However, this cannot be merely reduced to the issue of which screening method or crystallization set

up is to be used. Rather, thorough biochemical and/or biophysical work and intensive protein characterization, in combination with comprehensive screening for the most suited detergent, may be the most efficient strategy to cope with the difficulties of membrane protein crystallization. The reason for this is our rather limited knowledge about manipulation of IMPs bearing hydrophobic/amphipathic surfaces which are usually enveloped with membrane lipid layer. More often than not, the membrane proteins get trapped as an intractable aggregate in micelles during crystallization, which makes it inherently resistant to forming ordered crystal lattices (35).

NMR offers an alternative method. Solution NMR spectroscopy has been a very successful method for determining structures of soluble proteins up to molecular weights of ~30 kDa and, in a few cases, beyond. The use of NMR as a tool to determine structures of membrane proteins, however, has been still in a developmental stage.

In principle, the structures of membrane proteins can be studied in different environments by NMR, such as lipid bilayers, bicelles and detergent micelles (Figure 1.9). Lipid bilayers are the natural environment of membrane proteins. Direct structure determination of membrane proteins embedded in lipid bilayers requires the approach of solid-state NMR. Membrane protein samples in lipid bilayers are too large to tumble with a short enough correlation time (the time it takes to rotate by one radian) to yield narrow and well-resolved resonance lines, as required for high-resolution NMR. Currently, this problem can be resolved and the individual peaks can be obtained for the samples which are either mechanically oriented in the magnetic

field or unoriented, but spun at the magic angle (the angle at dipolar coupling of the sample becomes zero) in the NMR spectrometer. However, so far, as the result of the fact that research on solid-state bioNMR has just started and thus is still far behind solution bioNMR, this approach has been successfully employed to determine the complete structures of only a few very short hydrophobic peptides (36-39). However, many researchers are currently putting strong efforts to extend these methods to larger proteins (40, 41).

Membrane proteins can also be studied in bicelles. Bicelles are disk-shaped aggregates of phospholipid and detergent that orient spontaneously perpendicular to an applied magnetic field owing to their diamagnetic moment. Several recipes to create bicelles of different sizes, shapes and orientation properties have been described (42). Although originally devised to orient membrane proteins in the magnetic field for solid-state NMR studies, they have recently gained more use in introducing small degrees of residual orientation for soluble proteins in order to determine dipolar couplings, which have proven extremely beneficial for structure determinations of soluble proteins by high-resolution NMR. Apart from studying the structures of small membrane-bound peptides (43, 44), bicelles have so far not found wide application in the structure determination of large membrane proteins (45).

In addition to lipid bilayers and bicelles, membrane proteins can be analyzed in detergent micelle systems by solution NMR techniques. Currently, detergent micelles are shown to be the most appropriate environments for studying membrane proteins by high-resolution NMR techniques, although it is also difficult. From the solution

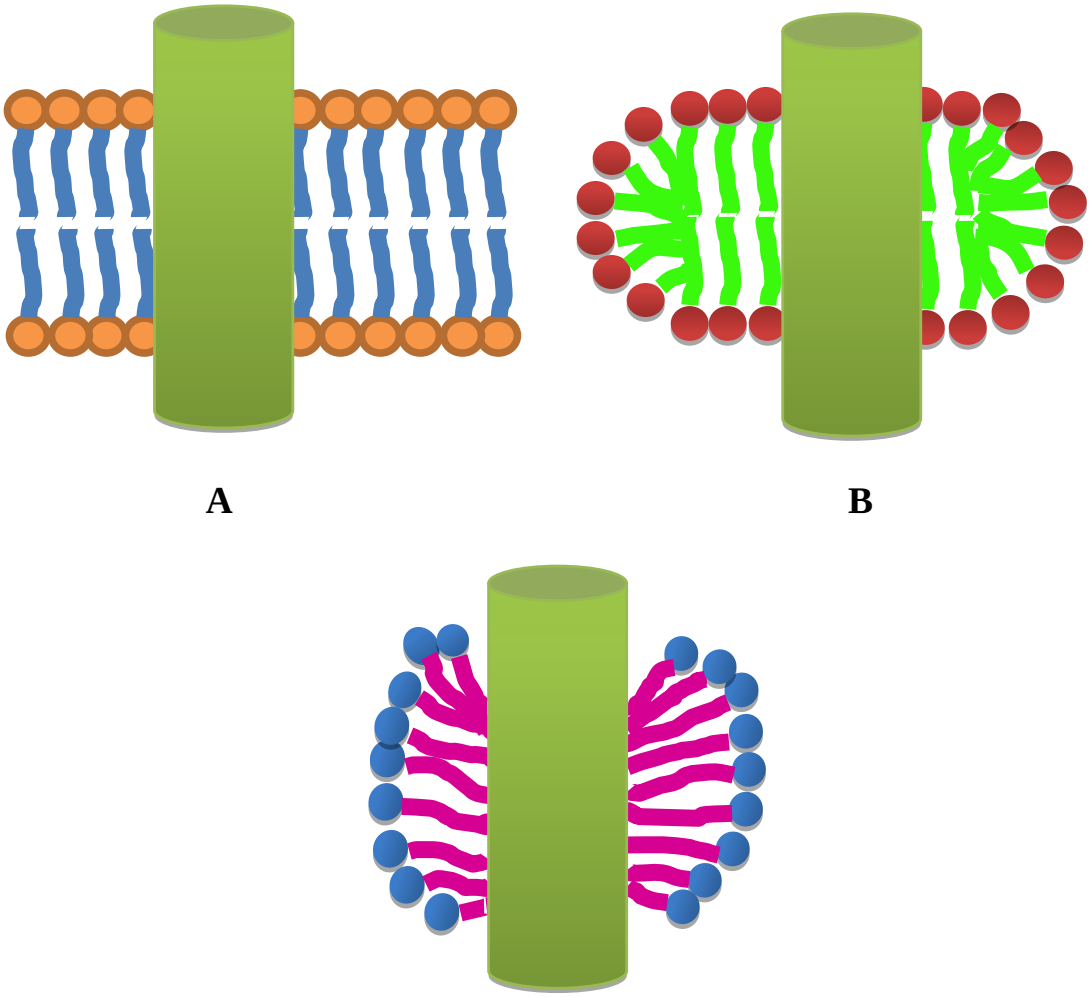


Figure 1.9 Membrane and membrane-like systems commonly used in biophysical studies of membrane proteins. A: Lipid bilayers, the natural environment of membrane proteins, are used in solid-state NMR studies of membrane proteins. B: Bicelles are disk-like structures composed of bilayer-forming lipids and detergents. They orient with their normal orthogonal to the magnetic field. Bicelles are used to orient soluble proteins in solution NMR studies and membrane-bound peptides in solid-state NMR studies. C: Detergent micelles are small, mostly spherical structures used in solution NMR studies of membrane proteins. Adapted from reference 45.

NMR perspective, a protein associated with detergent molecules tumble as part of a large complex, which leads to slower tumbling and rapid transverse relaxation rates, thus causing substantial signal broadening, poor sensitivity and reduced spectral resolution. This is even more problematic considering the fact that helical membrane proteins often have very narrow spectral dispersion due to the preponderance of similar amino acid types located in the hydrophobic domain/s (46).

A major advance in solution NMR spectroscopy that has had a significant impact on the determination of membrane protein structures in detergent micelles has been the development of TROSY (16). The problems associated with high magnetic fields (currently up to proton frequencies of 1000 MHz) are that transverse relaxation resulting from chemical shift anisotropy (CSA), and dipolar interactions causes significant line broadening, which offsets some of the high-field advantages for resolution and sensitivity. CSA is defined as the chemical shift difference between the isotropic and anisotropic states, However, in TROSY-type experiments, the scalar heteronuclear spin–spin couplings are not decoupled, and only one of the four peaks in the multiplet is retained and the chemical shift anisotropy relaxation (at high fields) is used to compensate dipolar relaxation (for theoretical details, see 16). This procedure results in improved sensitivities for proteins and complexes that are larger than ~20 kDa, which is almost always the case for membrane proteins in detergent micelles.

However, even with the development of TROSY type experiments, to date, successful examples for NMR structure determination of IMPs have been limited to

only very small, structurally simple IMPs (47-49) and for outer membrane bacterial proteins (50–52). For β-barrel membrane proteins, its β-barrel fold allows for large spectral dispersion (thus higher spectral resolution) and collection of ample interstrand long-range backbone-to-backbone NOEs, which are sufficient to determine the fold of the protein. Unfortunately, for the vast majority of the α-helical IMPs of medium to large sizes, the spectral resolution is narrow causing severe resonance overlap that affects both resonance assignment and structure calculation.

As a result, access to the structure, and thus function, of tens of thousands of α-helical membrane proteins remains very limited. NMR and other biophysical approaches for membrane protein structure determination need to be further developed in order to promote the field of structural biology of membrane proteins to a level that measures up to that of soluble proteins.

1.3 Oligosaccharyl Transferase (OT)

Many proteins of living organisms are modified in various ways during or after their expression process to be functional. These co- or post-translational modifications include protein phosphorylation, alkylation, acylation, glycosylation, etc. Among these, for eukaryotic cells, the most ubiquitous, and at the same time the most complex protein modification is N-linked glycosylation. N-linked glycosylation is catalyzed by oligosaccharyl transferase (OT, EC 2.4.1.119). OT is a remarkably complex multisubunit enzyme. In the case of baker's yeast, *Saccharomyces cerevisiae*, a frequently used eukaryotic model organism, OT contains nine non-identical integral membrane protein (IMP) subunits: Ost1p, Ost2p, Ost3p, Ost4p, Ost5p, Ost6p, Wbp1p,

Swp1p, and Stt3p (53). Among these, Ost3p and Ost6p are homologous, interchangeable subunits, while Stt3p, Wbp1p, Swp1p, Ost1p, and Ost2p are essential for the viability of the cell (54). Ost4p is essential for growth of the cell at 37 ℃, but not at 25 ℃. Ost3p/Ost6p and Ost5p subunits are not essential for the viability of the yeast cell but are required for maximal enzyme activity (53).

In the central reaction, OT transfers a preassembled oligosaccharide moiety from a dolichol pyrophosphate-linked (Dol-PP-oligosaccharide) donor onto the side chain of the Asn of the nascent polypeptide chain as it enters the lumen of the rough endoplasmic reticulum (RER). The glycosylated Asn residues are specified by the -N-X-T/S- consensus sequence, where X can be any amino acid except proline (53, 55), as shown in Figure 1.10. According to the statistics, which shows that only 66% of the signature sequons are glycosylated, further structural requirements have to be fulfilled for N-linked glycosylation to occur (56, 57). Hence, the amino acids within and around the sequon, the position of the sequon in the peptide chain, the rate of protein folding and the availability of the dolichol precursor saccharide, all influence the efficiency of N-glycosylation (58-60).

The N-linked oligosaccharide moieties of these proteins serve highly diverse functions, such as stabilizing the proteins against denaturation and proteolysis, enhancing solubility, modulating immune responses, facilitating orientation of proteins relative to a membrane, conferring structural rigidity to proteins, regulating protein turnover, fine-tuning the charge and isoelectric point of proteins, and

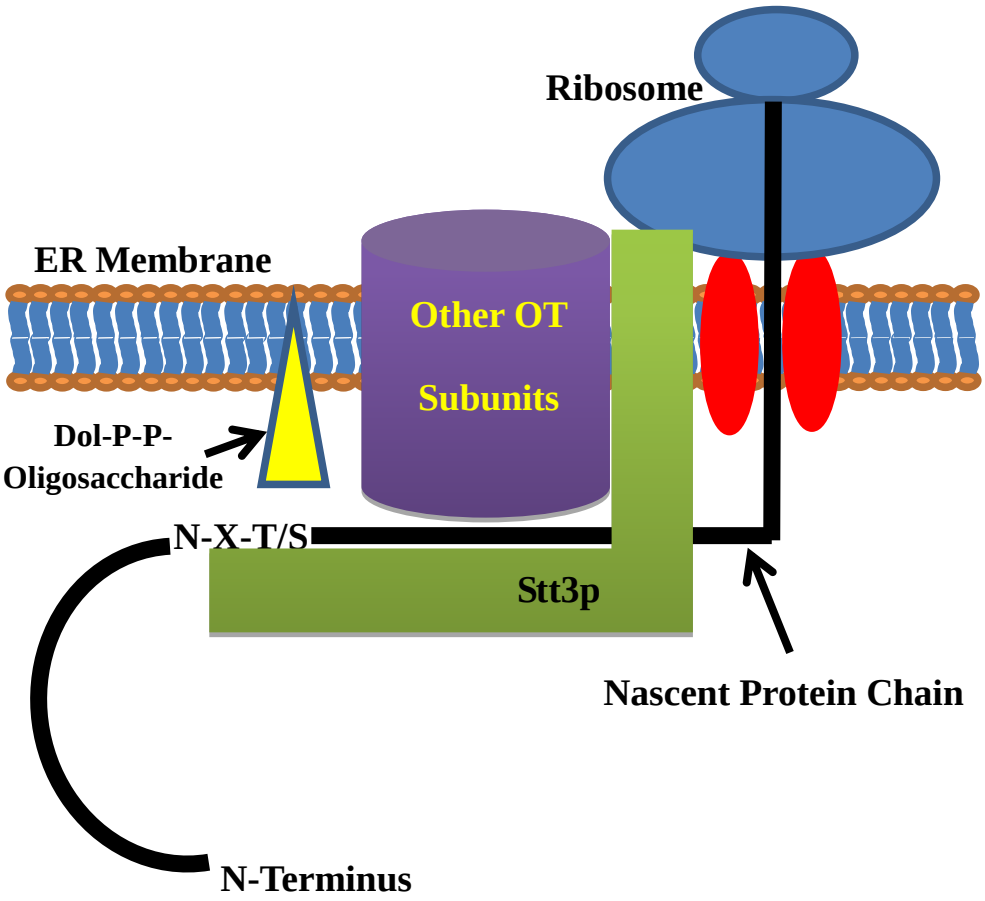


Figure 1.10 A cartoon model for OT catalytic reaction.

mediating interactions with pathogens (61-63). Actually, in eukaryotic cells, no other covalent protein modification is as common, as complex chemically, and as energetically costly as that of N-glycosylation. In fact, no other modification is employed for so many different purposes as N-glycosylation (63). Collaborative efforts between physicians and scientists have led to the discovery of 18 inherited human disorders known as Congenital Disorders of Glycosylation (CDG) that result from defects in protein N-linked glycosylation processes (64). These conditions affect multiple organs with severe clinical manifestations including mental retardation,

developmental delay, hypoglycemia, liver dysfunction, etc. Although the molecular details leading to these diseases are only vaguely understood, it seems clear that saccharide components of proteins play a major role in embryonic and post embryonic development of humans as well as of all higher eukaryotes (64). Complete loss of N-linked glycosylation is lethal to all eukaryotic organisms.

Given its extreme importance, much effort has been put into understanding the structure and mechanisms of this enzyme complex over the last few decades. Although many questions, including the most fundamental question, the enzymatic mechanism of N-linked glycosylation, have continued to be unanswered, investigators have provided some clues as to the possible functions of the OT subunits in this modification reaction.

Ost1p: Once it was proposed that Ost1p bear the peptide-binding site of the OT complex (65), but this proposal was later disproved by more extensive mutagenesis studies (66). To date, the function of Ost1p remains unclear although it was suggested that the luminal domain of Ost1p is involved in funneling the newly synthesized polypeptides into the active site on Stt3p for the glycosylation OT reaction.

Ost2p: Since it has been shown that Ost2p interacts strongly with Wbp1p, Ost2p may aid Wbp1p in recognition of the Dolichol-PP-oligosaccharide (67).

Ost3p and Ost6p: In yeast, Ost3p and Ost6p are products of paralogous genes, and have the same predicted topology of an N-terminal domain located in the ER lumen followed by 4 transmembrane helices. It is thus believed that Ost3p and Ost6p perform redundant function(s) in the OT reaction (68). Ost3p and Ost6p are suggested

to play a role in generating the two isoforms of the OT complex, which associate with the two structurally similar translocon complexes (67). Crystal structure of the ER luminal domain (N-terminal) of Ost3/6p has been reported (69). It reveals that the ER luminal domain of Ost3/6p, which contains a thioredoxin-like fold with a characteristic CxxC active-site motif, functions as an active oxidoreductase (Figure 1.11). Further studies show that the oxidoreductase and redoxdependent peptide binding activities of Ost3/6p increase the glycosylation efficiency of defined sites in protein substrates by OT (69).

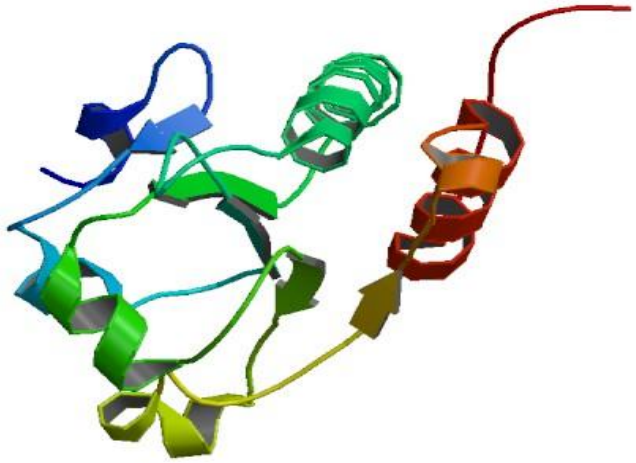


Figure 1.11 Crystal structure of the N-terminal soluble domain of Ost6p. Structure figure was obtained from PDB (Protein Data Bank, http://www.rcsb.org/pdb/home/home.do).

Ost4p: This 36-residue minimembrane protein, Ost4p, a non-essential yeast OT subunit, is the first yeast OT subunit whose structure has been determined (Figure 1.12) (70). It is proposed that Ost4p is involved in recruiting Ost3p or Ost6p into the OT complex (71).

Ost5p: Ost5p is another nonessential OT subunit, and its deletion only results

in a minor defect in OT activity (72). As for Swp1p, presently the function of Ost5p is still uncertain.

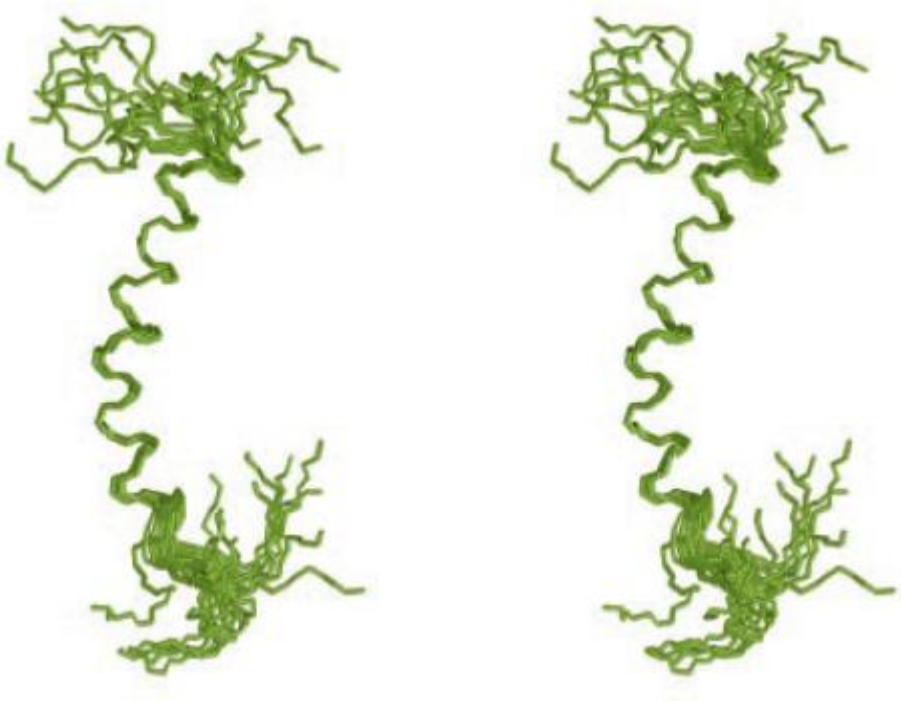


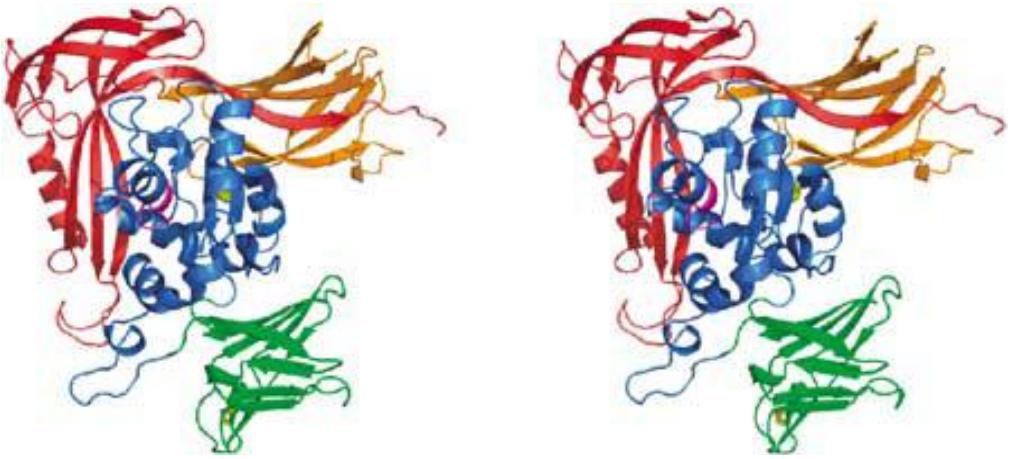
Figure 1.12 NMR solution structure of the mini-subunit of OT, Ost4p. Structure figure was obtained from reference 70 with permission. Copyright (2004) National Academy of Sciences, U.S.A.

Wbp1p: Based on the studies of chemical modification of cysteine residues, it was suggested that Wbp1p might contain the recognition site of the donor substrate of N-glycosylation, a dolichol pyrophosphate-linked oligosaccharide (Dol-PP-oligosaccharide) (73). Other evidence in favor of this proposal is that Wbp1p possesses a GIFT domain, which when present in other proteins is known to bind to oligosaccharides (74).

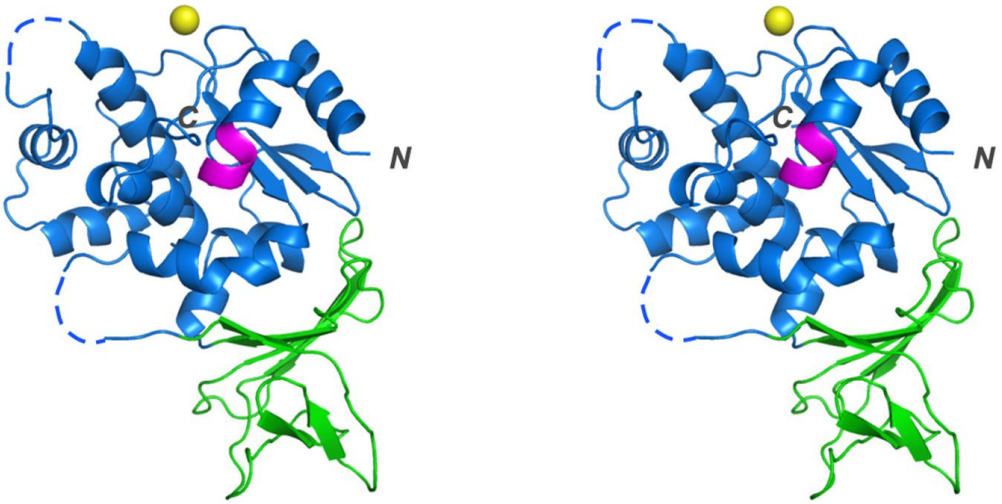
Swp1p: Although it is a product of an essential gene, the function of Swp1p remains unclear.

Stt3p: Among the nine subunits of OT complex, Stt3p is the largest and the only conserved subunit in all three domains of life (75). During the last several years, much evidence has been obtained indicating the direct involvement of C-terminal domain of Stt3p in the catalytic process of glycosylation in eukaryotes (66, 76, 77), the bacterium *Campylobacter jejuni* (78), and the archaea *Pyrococcus furiosus* (79). The most direct evidence demonstrating Stt3p as the catalytic subunit of the eukaryotic OT came from the finding that PglB, a Stt3p homolog in *Campylobacter jejuni* bacteria, catalyzes N-linked glycosylation activity by itself. Moreover, the expression of PglB in *E. coli* reconstituted N-linked glycosylation activity in the *E. coli* host unless point mutations were introduced into the WWDYG motif in PglB (80). Similarly, expression of the *Leishmania major* Stt3p homolog in yeast not only complements the yeast *STT3* deletion, but also is able to replace the whole OT complex of yeast (81, 82). Just recently, two crystal structures of the water-soluble C-terminal domain of Stt3p prokaryotic homologs were determined (79, 83). These two prokaryotic Stt3p homologs, *P. furiosus* AglB protein and *C. jejuni* PglB protein, have very little sequence similarities to the eukaryotic Stt3p (Figure 1.13). Their structures demonstrate that for homologs, the catalytic domains, the domain containing the well-conserved WWDYG motif, mainly consists of α-helices.

In many eukaryotic organisms, the subunit of Stt3p is present in multiple isoforms (STT3A and STT3B, STT3-A is shorter than STT3-B, and both forms share a 59% amino acid identity). This, together with the homologous Ost3/6p, results in the presence of the OT complexes with different protein isoform compositions, which in



A



B

Figure 1.13 Stereoscopic views of crystal structures of the C-terminal domain of Stt3p homolog from prokaryotic sources. A: AglB protein from *P. furiosus,* and B: PglB protein from *C. jejuni.* The catalytic domain, so called "center core domain", contains the WWDYG catalytic motif (blue color) and is mainly α-helical. Structure figures were obtained from reference 79 and 83, respectively for A and B with permission.

turn have been reported to affect OT glycan and protein substrate-specific activities. The mammalian STT3A/B isoforms are differentially expressed in various tissues, and result in the OT with altered kinetic characteristics (84) and preferences for co- or posttranslational glycosylation (85).

Structural Organization of OT: The structural organization of the OT complex also remains obscure. Based on genetic screens and biochemical co-immunoprecipitation experiments, it was once proposed that the OT is composed of three subcomplexes: Ost1p-Ost5p, Ost3p-Stt3p-Ost4p, and Ost2p-Wbp1p-Swp1p (86), as shown in Figure 1.14 Model A.

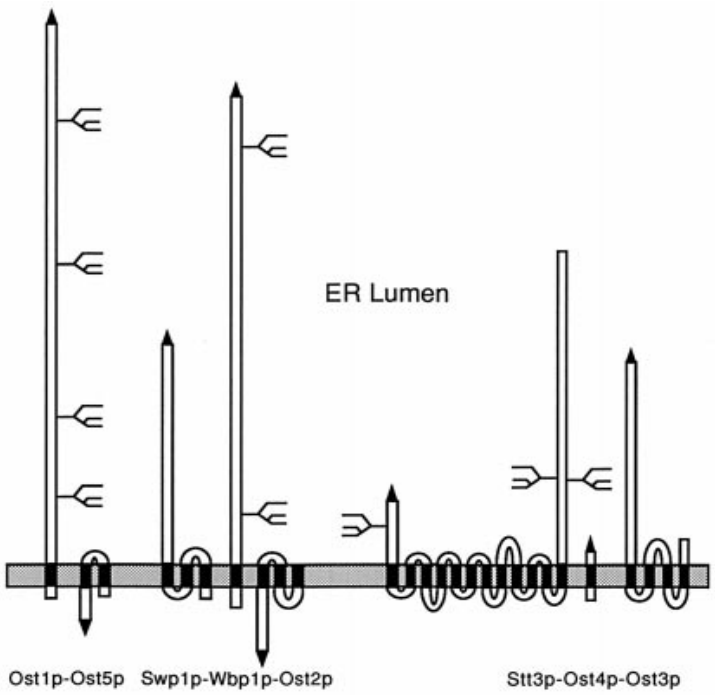


Figure 1.14 Model A for the structural organization of OT in the ER membrane. Three subcomplexes designated as the Stt3p-Ost4p-Ost3p subcomplex, the Swp1p-Wbp1p-Ost2p subcomplex, and the Ost5p-Ost1p subcomplex are proposed to be assembled into the octameric OT. Figure is obtained from reference 86.

However, further studies with application of a cross-linker yielded a different model, in which it is proposed that two isoforms of the OT complex exist in the ER membrane and differ only in the presence of either Ost3p or Ost6p. In this model, as shown in Figure 1.15, Model B, five essential gene products in the OT complex were found to be within a distance of 12 Å of each other. Two low molecular weight subunits, Ost4p and Ost5p, were shown to interact with only a restricted number of subunits and were proposed to locate closely to Stt3p and Ost1p, respectively, while Ost1p was found to be cross-linked to all of the other eight components and therefore was placed in the core of the OT complex (87).

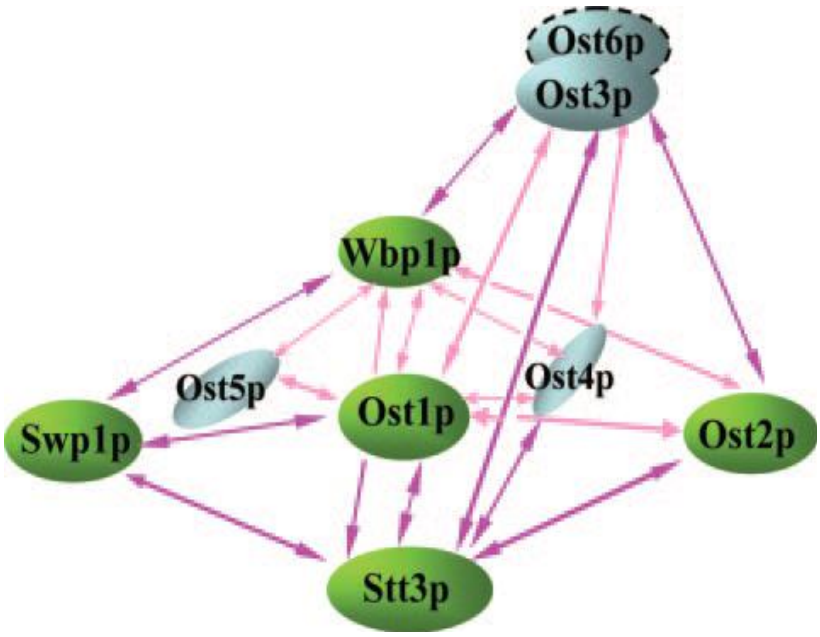


Figure 1.15 Model B for the interrelationship of yeast the OT subunits detected by cross-linking studies. Five essential gene products (shown as green) are located within 12 Å of each other. Ost4p and Ost5p (light blue) are found to interact only with a restricted number of subunits. Ost3p and Ost6p (blue) are present in the complex alternatively. Figure is obtained from reference (88).

Structure of the OT complex: As can be seen, our knowledge about the structure of OT remains very limited. So far, the only available atomic-resolution structures of OT are from the C-terminal water-soluble domain of two prokaryotic Stt3p homologs (79, 83), the C-terminal water-soluble domain of Ost6p (69), and the 36-residue Ost4p (70). Recently, a 12Å resolution cryo-electron microscopy structure of yeast OT from yeast was reported (89). From this rather low-resolution structure, it was found that the OT has a large luminal domain in endoplasmic reticulum where the catalysis occurs. The luminal domain mainly comprises Stt3p, Wbp1p, and Ost1p, and a prominent groove was observed between these three subunits (Figure 1.16). The authors proposed that the nascent polypeptide from the translocon threads through this groove while being scanned by OT for the presence of the glycosylation sequon.

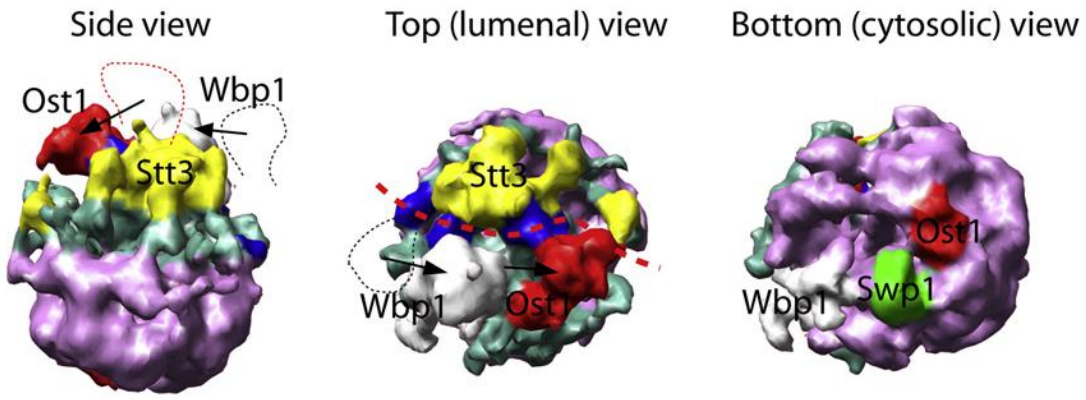


Figure 1.16 Low-resolution Cryo-EM structure of the yeast OT. There is a groove between the lumenal domains of Stt3p, Wbp1p, and Ost1p (dashed red curve), where the nascent proteins are proposed to be scanned and glycosylated by OT. Structure figures were obtained from reference 89 and used with permission.

The scarcity of its structural information inevitably leads to the lag of our knowledge about OT's functional mechanism. For example, it is still unclear why nature needs nine different subunits to catalyze the N-linked glycosylation, a relatively uncomplicated reaction. To ultimately answer this question and obtain the full understanding of the mechanism of the OT complex, the atomic-resolution structure of each subunit must be solved.

In this dissertation, the NMR structure of the C-terminal domain of Stt3p, the catalytic subunit of the OT complex, are to be determined by solution NMR.

CHAPTER 2

PRODUCTION OF THE C-TERMINAL DOMAIN OF STT3P

"The only thing we learn from history is that we learn nothing from history." -

Friedrich, Hegel.

2.1 Overexpression of the C-terminal Domain of Stt3p

2.1.1 Introduction

One major bottleneck preventing studies on the OT complex is due to the inherent difficulties associated with the preparation of milligram quantities of membrane proteins, which are necessary for both structural and functional studies.

Membrane proteins are fickle entities and repeatedly resist even the most determined efforts to overexpress and purify them for structural studies. In fact, it is not surprising that overexpression of membrane proteins is problematic. Unlike their cytoplasmic counterparts, membrane proteins must do far more than simply fall off the ribosome in order to achieve correct folding and targeting. Once synthesis of a membrane protein begins, either the secretary machinery is engaged and that expressed protein will be targeted to and inserted into the membrane; or it will lead to the formation of inclusion bodies, insoluble aggregates of misfolded proteins. In the latter case, while refolding proteins isolated from inclusion proteins is common

practice for soluble proteins, our knowledge of the *in vitro* renaturation/reconstitution of membrane proteins is considerably less advanced. Overexpression of membrane proteins, which is required to efficiently introduce appropriate isotopes essential for structure determination by heteronuclear NMR, often leads to cell toxicity and reduction in protein expression. Therefore, screening an optimal host cell line to express a particular membrane protein is, more often than not, a painstaking process.

As mentioned in Chapter one, several independent laboratories have each proposed that, among the nine subunits of OT, Stt3p is the catalytic subunit that is directly involved in the N-glycosylation reaction. Moreover, the absolutely conserved motif WWDYG (residues 516 to 520) in the C-terminal domain is believed to play a central role in the glycosylation process.

Stt3p is the most conserved of the known OT subunits. As shown in Figure 2.1, the C-terminal domain of Stt3p is highly conserved among different species, from yeast through human. The ultimate goal of this study is to determine the solution structure of the catalytic domain of the OT complex, the C-terminal domain of Stt3p (residues 466 to 718). The first step is, therefore, to clone the gene and find an efficient expression system to produce the protein of interest.

2.1.2 Methods and Materials of Subcloning and Overexpression of the C-terminal Domain of Stt3p

The subcloning and overexpression of the C-terminal domain of Stt3p in *E. coli* have been previously accomplished in our lab (90) and the optimized expression protocol was followed to produce the protein. Briefly, the C-terminal domain

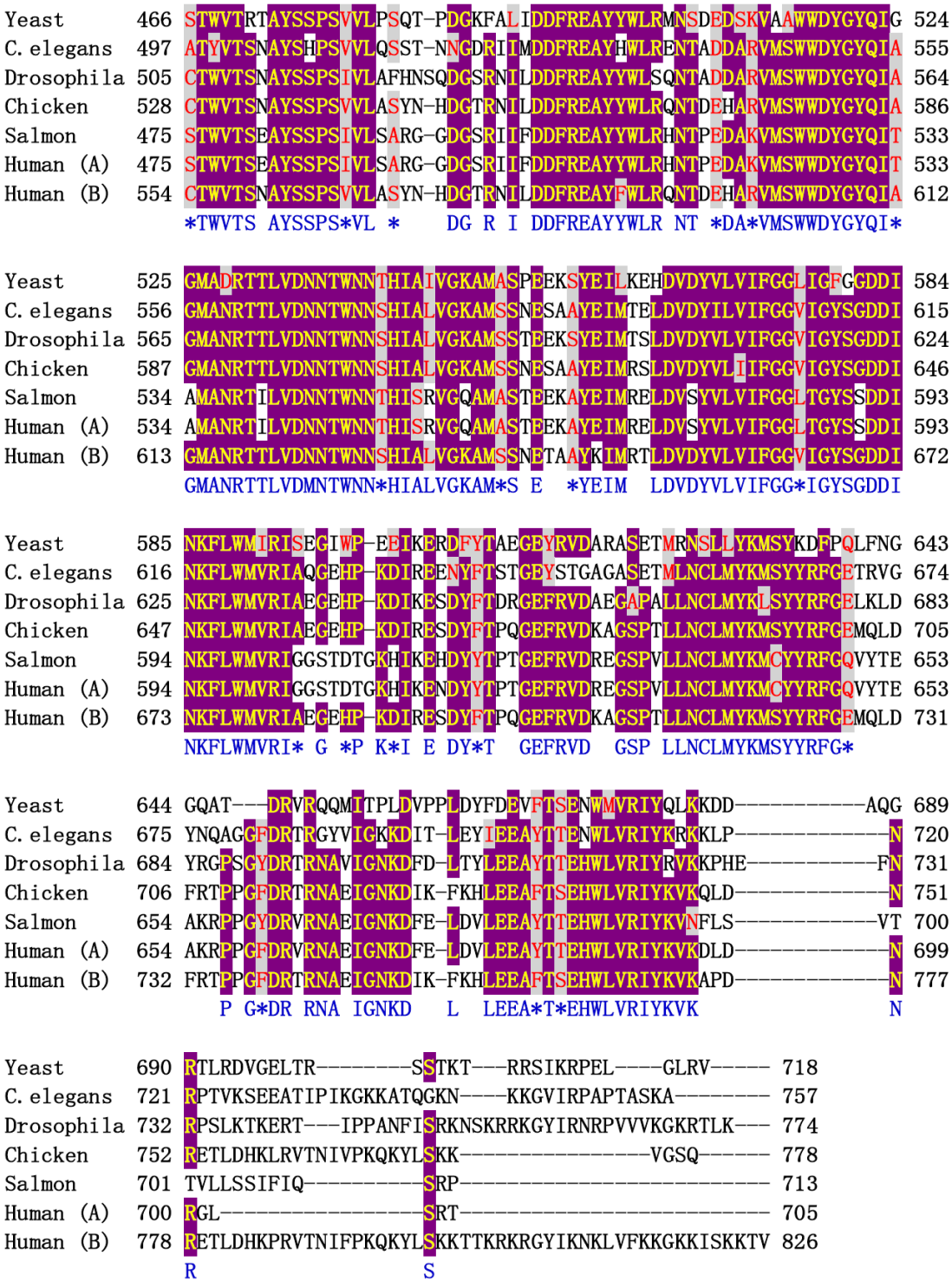


Figure 2.1 Sequence alignment of the C-terminal domain of Stt3p among different eukaryotic species, from yeast to human. The identical amino acid residues are shaded in purple, whereas the conservative replacements are shaded in gray. Gaps are indicated by dashes.

(residues 466-718) of the yeast Stt3p was produced in *Escherichia coli* BL21(DE3)-CodonPlus cells (Strategene) using pET-28c vector (Invitrogen), in which the promoter is T7 promoter. Expression of the N-terminal His_6-tagged Stt3p in the pET-28c vector was under the control of *lac* operator, an IPTG (isopropyl-β-D-thiogalactopyranoside) -inducible operator.

The overnight starter culture of the transformed cells was diluted to an OD_{600} of 0.1 in fresh LB media containing 50 μg/L kanamycin and were grown at 37 ℃ to an OD_{600} of 0.4-0.6. At that point, the temperature was reduced to 30 ℃ and protein production was induced by the addition of IPTG to a final concentration of 0.5 mM. After 4 hours, the cells were harvested by centrifugation (10,000×*g*) for 20 min at 4 ℃, and frozen at -80 ℃ until needed. The volume for a typical run for protein expression is 500 mL.

2.1.3 Results and Discussion

It is imperative to produce milligram quantities of pure protein for structural characterization of any protein. This requirement along with the necessity of a suitable membrane mimetic, are formidable obstacles to structural characterization of IMPs. Existing strategies for overexpression in *E. coli* have proven adequate for many prokaryotic proteins; eukaryotic membrane proteins, however, require significant technical developments before routine overexpression is a reality.

As a result, to date, there are no reports of recombinant overexpression of any eukaryotic OT subunit in *E. coli* or in any other heterologous system. This deficiency seriously impedes any biophysical and/or biochemical research on N-linked

glycosylation. In our lab, the expression level for C-terminal domain of Stt3p was excitingly high: ~65 mg/L in LB media for unlabeled protein or ~35mg/L in minimal media for uniform ^{15}N labeled protein. The target protein was expressed as inclusion bodies (Figure 2.2), which is quite common for eukaryotic membrane proteins since these proteins are often not incorporated well into the plasma membrane of *E. coli*. Indeed, to date, most heteronuclear NMR studies of membrane proteins have been carried out with proteins that were expressed in *Escherichia coli*, recovered from inclusion bodies and subsequently refolded.

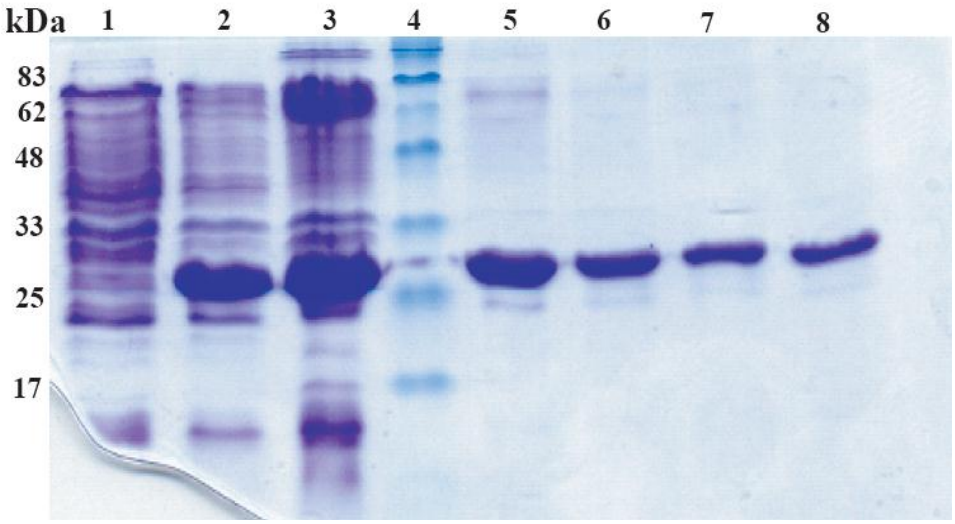


Figure 2.2 Coomassie-stained SDS-PAGE of samples from a typical C-terminal Stt3p expression and purification run. The mobility of the His-tagged C-terminal Stt3p in the SDS-PAGE gel is compatible with its molecular mass (31.4 kDa). Lane 1, before induction; Lane 2, 4 hours after induction with 0.5mM IPTG; Lane 3, inclusion body; Lane 4, protein molecular weight markers; Lane 5-8, protein purified by "SDS Elution" which will be described later.

2.2 Purification of the C-terminal Domain of Stt3p

2.2.1 Introduction

Purity and homogeneity are as critical in the structure determination of membrane proteins as they are for water-soluble proteins. Hence, regardless of the type of expression host employed, the protein target must be purified, using standard biochemical techniques. However, proteins are notoriously individualistic in their behavior. This individuality requires that purification protocols be tailored to suit particular molecules.

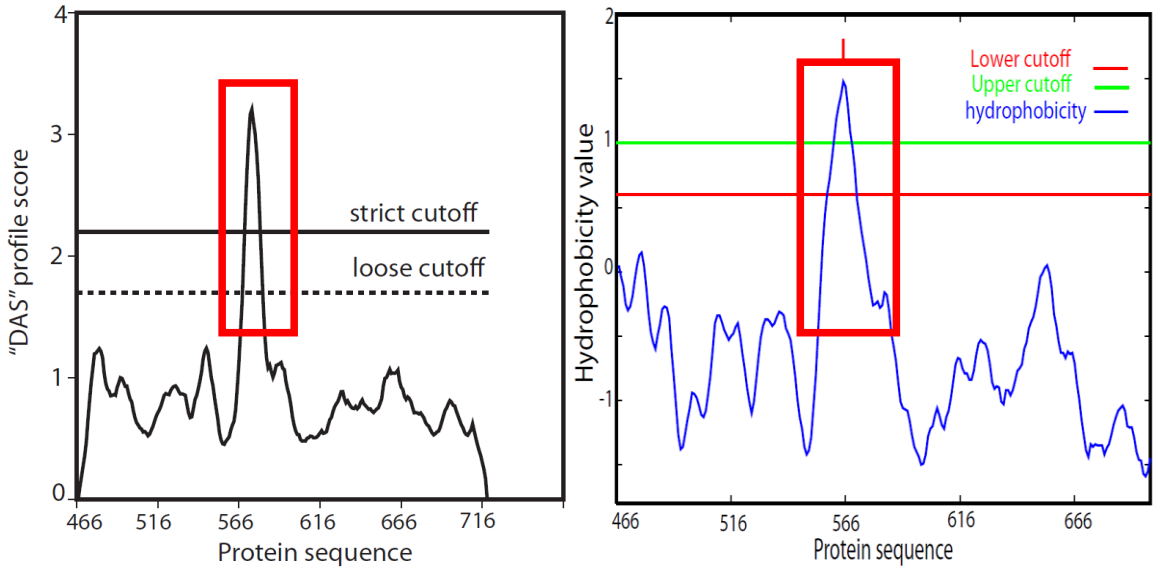
Membrane proteins are difficult to handle; the difficulties reside in the amphipathic nature of their surface. They possess a hydrophobic surface where they are in contact with the alkyl chains of the lipids, and they possess a polar surface where they are in contact with the aqueous phases on both sides of the membrane or with the polar head-groups of the lipids. In order to solubilize and to purify membrane proteins one has to add a vast excess of detergents (well above their critical micellar concentration (CMC)). The detergent micelles take up the membrane proteins and cover the hydrophobic surface of the membrane protein with their alkyl chains in a belt-like manner, while the polar head groups of the detergents face the aqueous environment. Compared to their water-soluble counterparts, more attention should be paid to membrane proteins due to the presence of detergents. In particular, since the detergent and proteins form a protein–detergent complex which is soluble in aqueous solvents and this complex contains comparable quantities of protein and detergent, the need for sample homogeneity extends to the nonprotein components of the complex.

It is also important that the structural integrity of the membrane protein be maintained (conformational homogeneity) and that nonspecific aggregates be avoided (aggregation state homogeneity). In some cases, these physical properties can be substantially more difficult to assess than mere protein purity.

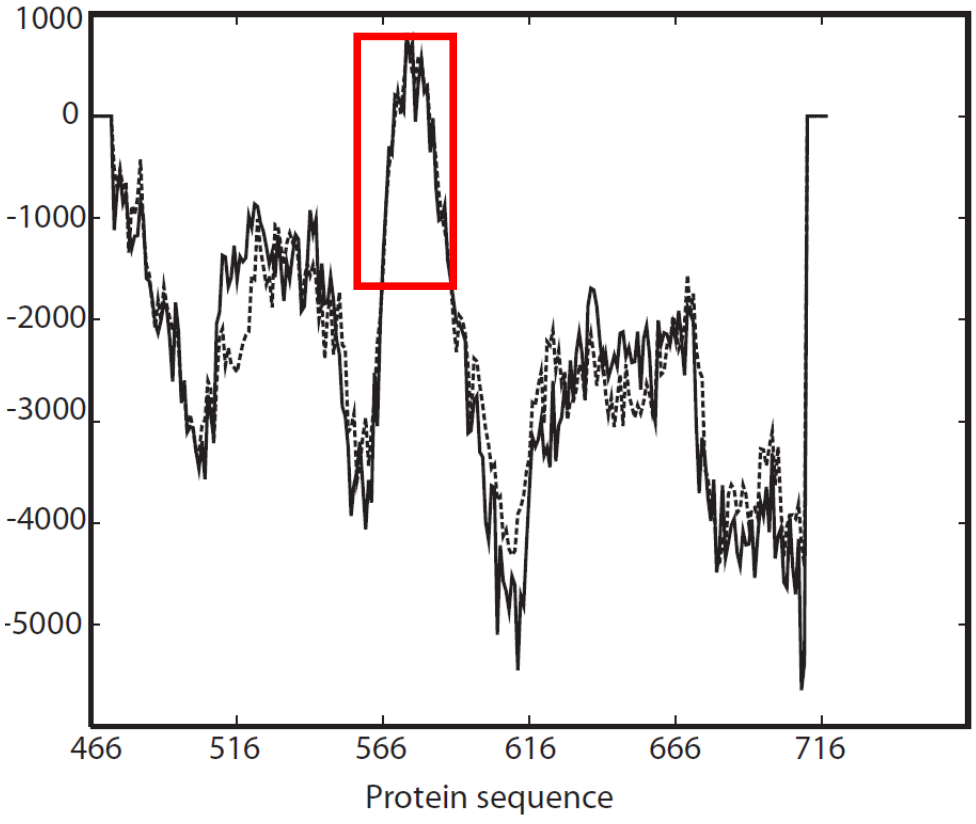
Since the C-terminal domain of Stt3p was expressed into the form of inclusion bodies, it had to be solublized first and then purified and refolded. A novel method for purification of the C-terminal domain of Stt3p was developed in our lab, involving His-Tag Nickel Affinity chromatography without the use of any imidazole.

Here, we would also like to emphasize that the C-terminal domain of Stt3p (466-718) is only soluble in detergent micelles and behaves like a membrane protein. It was previously reported that this domain is a hydrophilic luminal domain (75, 91) based on the results of topology reporter studies. However, we used several TM prediction programs such as DAS (92) (Figure 2.3 A), TopPred (93) (Figure 2.3 B), TMpred (94) (Figure 2.3 C), SPLIT 4.0 (95) (Figure 2.3 D) predict residues 564-584 to be a TM domain both for full length and the C-terminal dommina. Kyte-Doolittle hydropathy plot is a widely used method for delineating hydrophobic character of a protein (96). In this method, each amino acid is given a hydrophobicity score between 4.6 and -4.6. A score of 4.6 is the most hydrophobic and a score of -4.6 is the most hydrophilic. So in Kyte-Doolittle hydropathy plot, regions with values above zero are hydrophobic while below zero are hydrophilic. As shown in Figure 2.3 E, the Kyte-Doolittle hydropathy plot predicts the presence of at least one TM region (residues 564-584) located within the C-terminal domain of Stt3p. It is also very clear from our

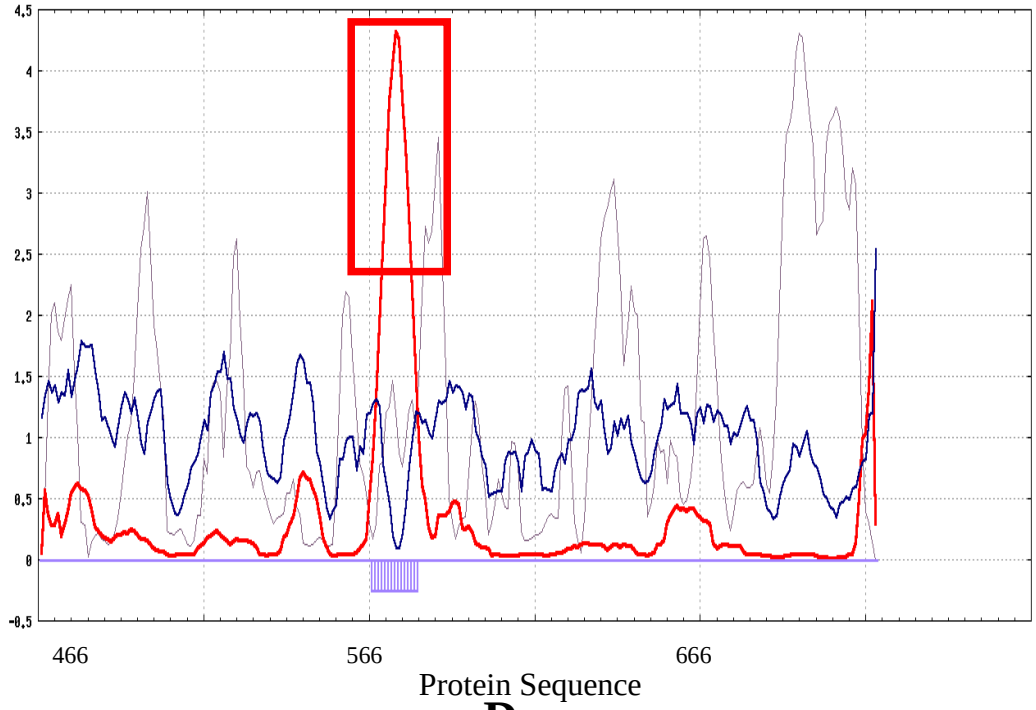
protein purification work that this domain (466-718) is unusually hydrophobic in character.



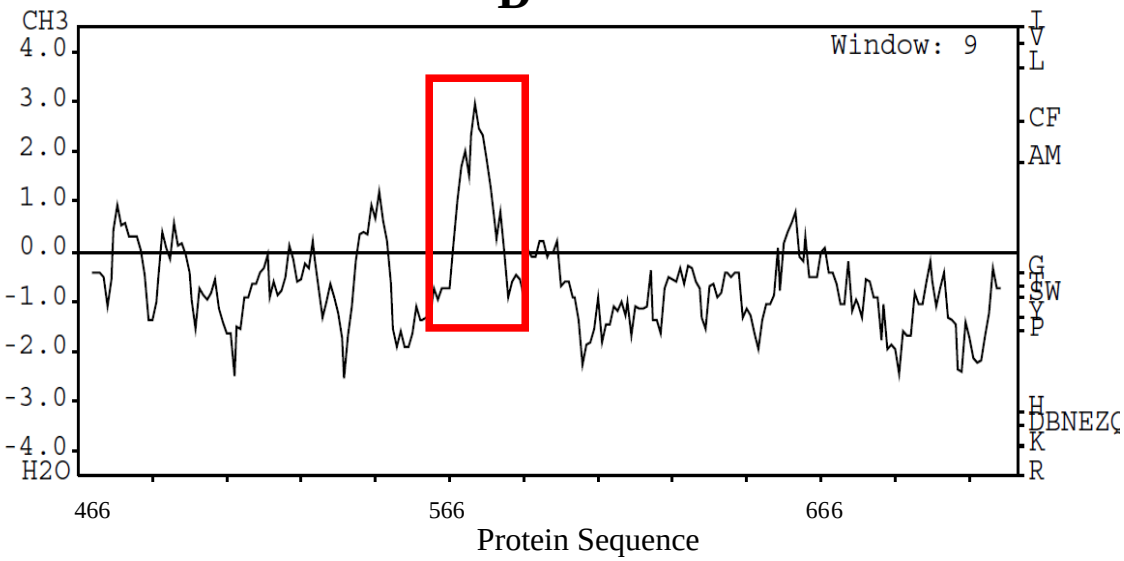
A B



C



D



E

Figure 2.3 TM domain predictions by various computer programs. The apparent positive domain (in red frame) is indicative of TM domain. All of these programs show a consensus TM domain, residue 564-584. A: prediction by DAS. B: Prediction by TopPred. C: prediction by TMpred. D: Prediction by SPLIT 4.0, and E: Prediction by Kyte-Doolittle Hydrophobicity Plot.

2.2.2 Methods and Materials

2.2.2.1 Preparation of inclusion bodies

The *E. coli* cells containing C-terminal domain of Stt3p wild-type (or D518E mutant) were passed through 4 cycles of freeze-thaw using liquid nitrogen and ice respectively before resuspension in B-PER solution (Pierce). The cells were then subjected to sonication (10×15s); the supernatant was removed after centrifugation at 10,000 × *g* for 30 min. The pellet was resuspended once with 10% B-PER solution, sonicated and centrifuged again as above. The inclusion bodies were stored at -20 ℃ until needed.

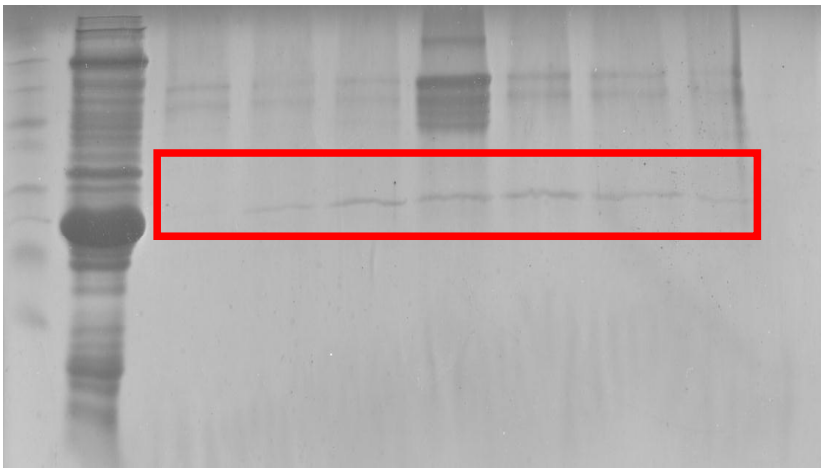
2.2.2.2 Purification and simultaneous refolding of the C-terminal domain of Stt3p

The C-terminal domain of inclusion bodies were dissolved in denaturing buffer containing 6 M guanidine hydrochloride (Gnd-HCl), 500 mM NaCl, 25 mM imidazole in 20 mM phosphate buffer at pH 7.4 and left at 42 ℃ overnight. The insoluble materials were removed by centrifugation. The supernatant containing solubilized C-terminal domain of Stt3p was loaded onto the Ni-NTA column (GE Healthcare) which was pre-equilibrated with binding buffer (500 mM NaCl, 25 mM imidazole, 20 mM phosphate buffer, pH 7.4). Impurities were removed using a washing buffer (20 mM phosphate buffer, pH 7.4, 500 mM NaCl, 200 mM imidazole, and 1% triton X-100 (v/v)) several times with shaking. In order to remove imidazole and NaCl from the protein sample before elution, a final wash was followed with 20 mM phosphate buffer, pH 6.5. The absorbance of the washing was monitored by

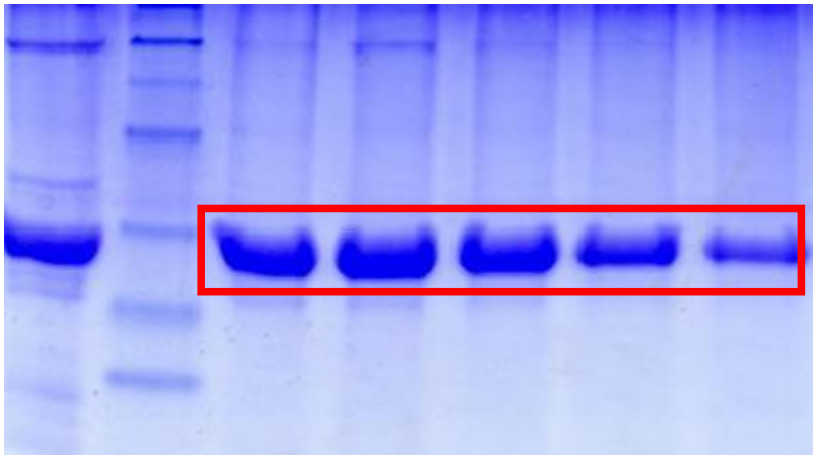
measuring OD_{280} until there was no apparent reading. The elution and simultaneous refolding were carried out by loading elution buffer (50 mM SDS, 1% glycerol, 20 mM phosphate buffer, pH 6.5) to the column followed by shaking for at least 2 hours. To keep the protein concentration high, the volume of elution buffer added was kept to a minimum (< 1 ml). The elution was continued until there was no more absorbance as monitored by OD_{280} readings. Protein concentration was calculated from the A_{280} using an extinction coefficient of 63083 M^{-1} cm^{-1} (97). The purity of the protein in each elution was assessed by SDS-PAGE analysis and it shows the target protein is >95% pure. The pure protein samples were kept at room temperature away from light.

2.2.3 Results and Discussion

Inclusion bodies were solubilized by using 6 M Gdn-HCl, followed by binding to Ni (II) metal ion affinity resin and washing off all of the impurities. Detergent was used during purification processes since the C-terminal Stt3p domain was found to be a water insoluble protein. The standard protocol is the use of imidazole to compete off the His-tagged protein from the Ni-NTA resin. However, this simple procedure did not work well for the C-terminal Stt3p domain. In fact, previous studies in our lab showed that most of the protein remained bound to the Ni-NTA resin even when the imidazole concentration was increased to ~ 2 M in the elution buffer containing digitonin (as analyzed by SDS PAGE, Figure 2.4 A). It was clear that there are other interactions most likely between the hydrophobic regions of the protein and the resin that are playing a major role in the binding. This observation was proved to be correct



A



B

Figure 2.4 SDS-PAGE analysis of the C-terminal Stt3p. A: purification by conventional method using imidazole, this gel picture was obtained from the previous work in our lab; B: sample purification using SDS elution.

when the His-tagged C-terminal domain was able to efficiently bind to even EDTA treated Ni^{2+} depleted resin. Thus, developing a new method for the elution of the protein off the resin was unavoidable in our case.

A novel, simple, yet robust purification protocol for the C-terminal Stt3p domain without using imidazole was developed in our laboratory (90). The protein bound to the Ni-NTA resin was efficiently eluted off of the column with buffer containing 50 mM SDS in 20 mM phosphate buffer at pH 6.5 after 2 hours of shaking at room temperature. Indeed, the first several elutions (500 µL of each eluted fraction) contained ~200 µM of pure protein (Figure 2.4 B). However, C-terminal Stt3p domain could also be eluted off the Ni-NTA column with SDS concentration as low as 10 mM. In fact, this method also worked for the elution of the protein from Ni^{2+} depleted resin. SDS was exchanged freely to any other detergent by following the protocol described under experimental procedures.

2.3 Conclusions

Although N-linked glycosylation is an essential, critical and highly conserved process in all eukaryotes, very little structural and functional information on the OT enzyme complex is known. Difficulties in the production of milligram quantities of integral membrane proteins (IMPs) for structural or functional characterization have hampered progress. Recombinant expression of IMPs in *E. coli*, the primary machine for large-scale protein production for structural studies, has had very limited success (97). As a result, as of now there are only a couple of examples of recombinant expression of C-terminal domain of Stt3p homolog from prokaryotic sources (85, 99),

together with one example of the N-terminal domain of Ost6p of yeast OT (69). For all of these examples, the domains chosen for expression are water-soluble.

We show here a high-level recombinant expression in *E. coli* and purification of the C-terminal domain of Stt3p from the yeast *Saccharomyces cerevisiae*. This is the first report of heterologous expression of a eukaryotic Stt3P subunit. This high level production of pure C-terminal domain of Stt3p makes isotopic labeling for structural characterization either by solution NMR or by X-ray crystallography straightforward, affordable and most importantly, possible.

After many unsuccessful attempts to refold denatured C-terminal domain of Stt3p in aqueous solution without the use of any detergents, we were convinced that a membrane mimetic environment is necessary for its purification and reconstitution. This evidence suggests that the C-terminal domain of Stt3p may contain at least one TM helix or several membrane embedded residues. Purification and reconstitution of membrane proteins are notoriously difficult tasks. Indeed, reports of successful isolation and refolding of IMPs from inclusion bodies have thus far been limited to a small number of proteins (100-102). In an elegant work, Page and co-workers have reported two methods for isolation and purification of helical integral membrane proteins: 'Detergent Exchange' and 'Reconstitution'. Both of these methods use standard protocols for detergent mediated purification via Ni^{2+} affinity chromatography (103). Here, we developed a novel method for the one-step purification and reconstitution of the C-terminal domain of Stt3p that we have named 'SDS Elution'. Using our method, we were able to obtain very high yield of purified

protein (60-70 mg of protein per liter of bacterial culture) in a single step. To evaluate the efficiency of our method, we also produced ^{15}N-labeled C-terminal Stt3p following the "reconstitution" method of Page *et al*. (104). After the precipitated proteins were reconstituted in 100 mM SDS and 300mM DPC micelles as per the "reconstitution" method, the HSQC spectra were collected and compared with the spectra of the samples prepared by our "SDS Elution" method. Although the HSQC spectra were similar overall (Figure 2.5), the quality of the spectra for the protein obtained by "reconstitution" method appeared to be deteriorated possibly due to protein aggregation. Indeed, in principle, "SDS Elution" combines both "Detergent Exchange" and "Reconstitution" together, but greatly simplifies the protocols.

This novel methodology has several advantages over the conventional methods. First and foremost, with our method, purification and reconstitution are achieved simultaneously, which dramatically shortens the sample preparation process. Secondly, since there is no imidazole in elution buffer, the conventional method of removal of imidazole through buffer exchange is not required, which avoids loss and dilution of the protein samples. Moreover, since the detergent is not introduced until the last step, it saves the amount of detergents used. This is especially true for NMR sample preparations where the use of deuterated detergents is necessary because deuterated detergents are generally very expensive. Investigation of the versatility of this method using for other integral membrane proteins needs further studies.

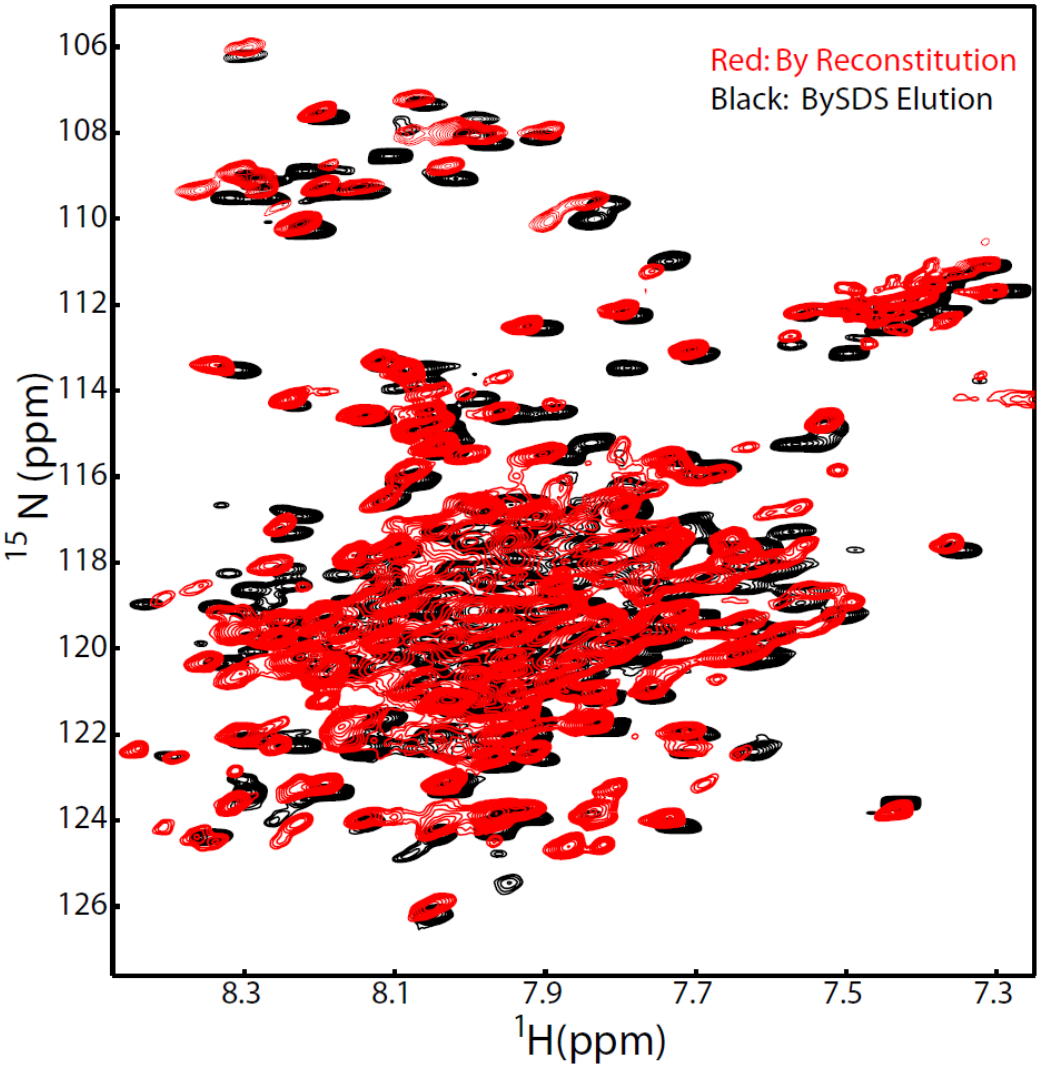


Figure 2.5 Comparison of 2D [^{1}H, ^{15}N]-HSQC spectra of the C-terminal domain of Stt3p prepared by different methods. Red: prepared by conventional "reconstitution" method, and Black: prepared by our "SDS Elution" method.

CHAPTER 3

BIOPHYSICAL CHARACTERIZATION AND FUNCTIONAL PROBING OF THE C-TERMINAL DOMAIN OF STT3P

"At the moment physics is again terribly confused. In any case, it is too difficult for me, and I wish I had been a movie comedian or something of the sort and had never heard of physics." - Wolfgang Pauli, 1925.

3.1 Introduction

Along with their expression, another barrier for membrane protein structure determination is to find a suitable detergent, because detergent micelles are usually used as mimics for lipid bilayers in structure/function studies of membrane proteins. Moreover, because membrane proteins are idiosyncratic in their interactions with detergents, there is no one detergent that can solublize every IMP and provide a stable environment for structure/function studies. As a result, finding a suitable detergent among the myriad detergents available is still very much a process of trial and error for an IMP.

After the C-terminal domain of Stt3p was purified by our "SDS elution", the questions that need to be investigated were:

(1) Is the C-terminal domain of Stt3p folded properly in SDS micelles, or is it just in the molten globule (MG) state, containing some native-like secondary structures but lacking a stable tertiary structure? Or is it simply denatured in sodium dodecyl sulfate (SDS) micelles and does not have any orderly structure at all?

(2) Is SDS the optimal detergent for its structure determination by solution NMR?

To answer these questions, in this chapter, a thorough biophysical characterization was carried out by using various biophysical techniques, including NMR, Circular Dichroism (CD) and Fluorescence spectroscopy. Furthermore, interaction of the C-terminal domain of Stt3p with an acceptor peptide containing the N-X-T/S consensus motif was also investigated by NMR to confirm the activity of the protein.

3.2 Methods and Materials

The detergents used in this study were sodium dodecyl sulfate (SDS) (Sigma), dodecylphosphocholine (DPC) (Anatrace), lauryl dimethylamine oxide (LDAO) (Anatrace), octyl-β-glucoside (OG) (Sigma), n-dodecyl-β-D-maltoside (DDM) (Anatrace) and digitonin (Calbiochem). For NMR studies, predeuterated detergents used were SDS (Sigma, 98% atom D) and DPC (Cambridge Isotope Laboratories, D38, >98%).

3.2.1 Mutagenesis

Oligonucleotides to introduce the D518E substitution were designed according to the QuickChange site-directed mutagenesis prodedure (Stratagene, USA). The

following sense and antisense primers were used wherein the sites of the mutation are italicized and underlined: 5'-GTTGCAGCGTGGTGG*GAA*TACGGTTACCAAATGG-3'(sense), 5'-CCAATTTGGTAACCGTA*TTC*CCACCACGCTGCAAC -3' (antisense). Incorporation of the mutation was verified by DNA sequencing.

3.2.2 Overexpression and Purification of the ^{15}N-labeled Proteins

For the production of ^{15}N-labeled C-terminal domain of Stt3p, cells were grown in M9 minimal media culture containing 0.12% $^{15}NH_4Cl$ (Cambridge Isotope Laboratories). All the rest of the procedures were same except that cells were grown for 8 hours after induction with IPTG before harvesting.

The same protocol was followed to overexpress the D518E mutant and the overexpression level was nearly identical to that of wild-type C-terminal domain of Stt3p. The purifications of ^{15}N-labeled Wide-type and D518E Mutant were achieved by following the same protocol, "SDS Elution", as described in chapter two.

3.2.3 Matrix-Assisted Laser Desorption Ionization (MALDI)-Time of Flight (TOF) Mass Spectrometry

The protein sample for MALDI-TOF measurement was in 10mM SDS, 20mM ammonium acetate. The target was spotted as follows: protein sample was mixed with the matrix (10 mg/ml sinapinic acid (SA) in 4:6 methanol:water, 0.1% trifluoroacetic acid) in a 1:4 ratio for a total of 1μL and the entire solution applied to target and allowed to air dry. MALDI mass spectra were acquired on an Autoflex II TOF mass spectrometer from Bruker. The spectra were acquired in linear mode using the

following settings: laser power 55%, ion source 1: 20 kV, ion source 2: 18 kV, lens: 6.50 kV, number of shots: 50, detection: 10,000 to 100000.

3.2.4 NMR Sample Preparation

Each sample for NMR measurement was concentrated to 0.2 mM using an Amicon Ultra-15 (MWCO, Molecular Weight Cut Off, = 5 kDa) centrifugal ultrafiltration cartridge. The final NMR sample was in 20 mM phosphate buffer, pH 6.5, 1 mM EDTA, 100 mM SDS and 5 % D_2O (v/v). In this study, besides SDS, five other detergents were screened to find out the most suitable membrane mimetic for C-terminal Stt3p domain. These detergents are: DPC, LDAO, OG, DDM, and Digitonin. The protein samples in the above detergents were prepared by buffer exchange of the protein in SDS detergent to the desired detergent by using Amicon ultrafiltration device with a MWCO of 5 kDa. Typically, 500 µL desired detergent solution was added to 500 µL SDS-containing protein sample in Amicon Ultra-15 tube and centrifuged until there was approximately 500 µL solution left. This process was repeated 10 times for complete detergent exchange.

3.2.5 NMR Measurement

[^{1}H, ^{15}N]-HSQC spectra were acquired for both wild-type and D518E mutant of the C-terminal domain of Stt3p. NMR measurements were conducted at 308 K. In this dissertation, except when specifically mentioned otherwise, all data were collected on a Bruker Avance 600 MHz spectrometer fitted with a cryogenic triple-resonance probe equipped with z-axis pulsed field gradients in Chemistry department at Auburn University. The data were acquired with 256 and 2048 complex points in the t1 time

domain (^{15}N dimension) and t2 time domain (^{1}H dimension) respectively. The data were zero-filled to 512 × 4096 and apodized using a Gaussian window function prior to Fourier transformation using NMRPipe (104).

3.2.6 Circular Dichroism (CD) Spectropolarimetry

All CD experiments were performed on a JASCO J-810 automatic recording spectropolarimeter using a 0.05 cm path length quartz cell at room temperature. Wild-type C-terminal domain of Stt3p was recorded in both SDS and DPC micelles, while D518E was recorded only in SDS micelles. The buffer used was 20 mM phosphate buffer (pH 6.5). The protein concentration was 10 μM for far-UV CD measurement and 89 μM for near-UV CD measurement. Data were averaged over 100 scans with a response time of 1s, and scan speed of 100 nm min^{-1}. CD data were converted to mean residual ellipticity (θ) by standard procedures.

3.2.7 Fluorescence

All fluorescence spectra were recorded on a Perkin Elmer Precisely LS 55 Luminescence spectrofluorometer. All experiments were carried out in 10 mM phosphate buffer, pH 6.5 containing 1 μM protein at 25 ^{o}C. The data were recorded by monitoring intrinsic tryptophan fluorescence (excitation at 280 nm and emission 300–500 nm).

3.2.8 Ligand Binding Studies by Saturation Transfer Difference (STD) NMR Spectroscopy

For STD studies, the methyl-protonated {Ile(δ_1 only), Leu($^{13}CH_3$, $^{12}CD_3$), Val($^{13}CH_3$, $^{12}CD_3$)} U-{^{15}N, ^{13}C, ^{2}H} labeled C-terminal domain of Stt3p was

overexpressed by using the same cell lines and vectors as described earlier. Briefly, the transformed cells picked from LB agar plate were grown in 3 mL of LB medium at 37 °C for 3 h, transferred to 25 mL of unlabeled minimal M9/H_2O medium, and grown until an OD_{600} of ~ 0.5. The cells were separated from the medium by centrifugation at 3,000 rpm for 15 minutes and transferred to 100 mL of M9/D_2O culture containing 0.12% (m/v) of $^{15}NH_4Cl$ as the sole nitrogen source and 0.4% (m/v) of ^{13}C, ^{2}H –glucose (Cambridge Isotope Laboratory, Andover, MA) as the sole carbon source. At OD_{600} ~ 0.5, the culture was diluted to 500 mL with M9/D_2O. One hour prior to induction, 35 mg of 2-keto-3,3-d_2-1,2,3,4-^{13}C-butyrate (Sigma Aldrich) and 60 mg of 2-keto-3-methyl-d_3-3-d_1-1,2,3,4-^{13}C-butyrate (Sigma Aldrich) were added to medium. The expression of the protein was induced at OD_{600} ~ 0.4 with 0.5 mM IPTG, and the culture was allowed to grow for an additional 11-12 h at 30 °C (final OD_{600} ~ 2.0), at which point the cells were harvested by centrifugation. The protocols for cell lysis and purification were the same as described previously.

The six-residue peptide (Tyr-Asn-Ser-Thr-Ser-Cys-Am, purity >99%) was custom synthesized by Biomatik USA, LLC. Protein NMR sample for STD experiment was prepared in 20 mM phosphate buffer (pH 6.5), containing 100 mM perdeuterated SDS and 10% D_2O. Protein and substrate peptide concentrations in the NMR sample were 30 μM and 300 μM respectively.

The STD measurements were performed at 308 K. The irradiation power was set to $(\gamma/2\pi)B_1$ = 20 Hz, which was applied on-resonance at 0.738 ppm where no peptide signals were present, or off-resonance at 100 ppm, where no protein signals were

present. In order to efficiently saturate the entire protein by spin diffusion, the saturation time was set to 10 s. A 50-ms spin-lock pulse ($T_{1\rho}$ filter) was used to eliminate the background protein resonances to facilitate analysis. The spectra were collected in an interleaved pseudo-2D fashion to reduce temporal fluctuations. AU program "stdsplit" from TOPSPIN 2.1 (Bruker) was used to subtract the unprocessed on- and off-resonance spectra.

3.2.9 Ligand Binding Studies by NMR HSQC Titrations

In order to measure dissociation constants (K_D), a series of 2D [^{1}H, ^{15}N]-HSQC spectra were collected with progressive additions of substrate peptide (Asn-Asp-Thr-NH_2) to ^{15}N-labled C-terminal Stt3p to attain molar ratios of protein to peptide of 1:0, 1:0.5, 1:1, 1:5, 1:10, 1:20, 1:35, 1:50, 1:75 and 1:100. The starting sample contained 170 μM protein in 20 mM phosphate buffer, pH 6.5, 100 mM SDS, 5% D_2O, 1% glycerol and 5 mM Mg^{2+}. The peptide with > 95 % purity was custom synthesized by Genemed Synthesis, Inc. (South San Francisco, CA, USA). NMR data collection and processing were the same as previously described. The chemical shift changes of the affected residues of the protein were plotted against the peptide concentration and fitted by Hill model in Origin 7.0 (Microcal). Chemical shift perturbations were calculated as $[(\delta^{1}H)^2 + (\delta^{15}N/5)^2]^{1/2}$, in which $\delta^{1}H$ and $\delta^{15}N$ are changes in chemical shift for ^{1}H and ^{15}N, respectively.

3.3 Results

3.3.1 Mass Determination by MALDI-TOF Mass Spectrometry

MALDI-TOF mass spectrometry was utilized to confirm the mass of the purified protein. However, this simple approach was complicated by the presence of SDS. Several reports have demonstrated that SDS is detrimental to MALDI-MS (105-107). After numerous trials, the appropriate conditions, including solvent system, matrix type and concentration of SDS, were determined to obtain reliable MALDI signals. Ammonium acetate buffer was found to be essential and the optimized SDS concentration was found to be 10 mM. The mass spectrum of the purified protein showed a molecular ion at *m/z* 31493.1 (Figure 3.1). This is in accordance with the calculated mass of 31553.4 Da for His-tagged C-terminal domain of Stt3p as calculated by the ProtParam tool of expert protein analysis system (ExPASy) (108). The error margin is only 0.19%, which is well within the acceptable error margin for MALDI-TOF data for biological molecules (109).

3.3.2 Detergent Screening by NMR Spectroscopy

The search for appropriate solution conditions for the NMR analysis requires the consideration of a larger number of variable parameters for membrane proteins than for water-soluble proteins. In addition to the temperature, the pH and the ionic strength, one has to consider the choice of the detergent, the detergent concentration, and the protein-to-detergent ratio. Moreover, membrane protein solutions tend to deteriorate in the NMR sample tubes, especially at the elevated temperatures, typically above 30 ℃, that are usually preferred for NMR spectroscopy. Long-term

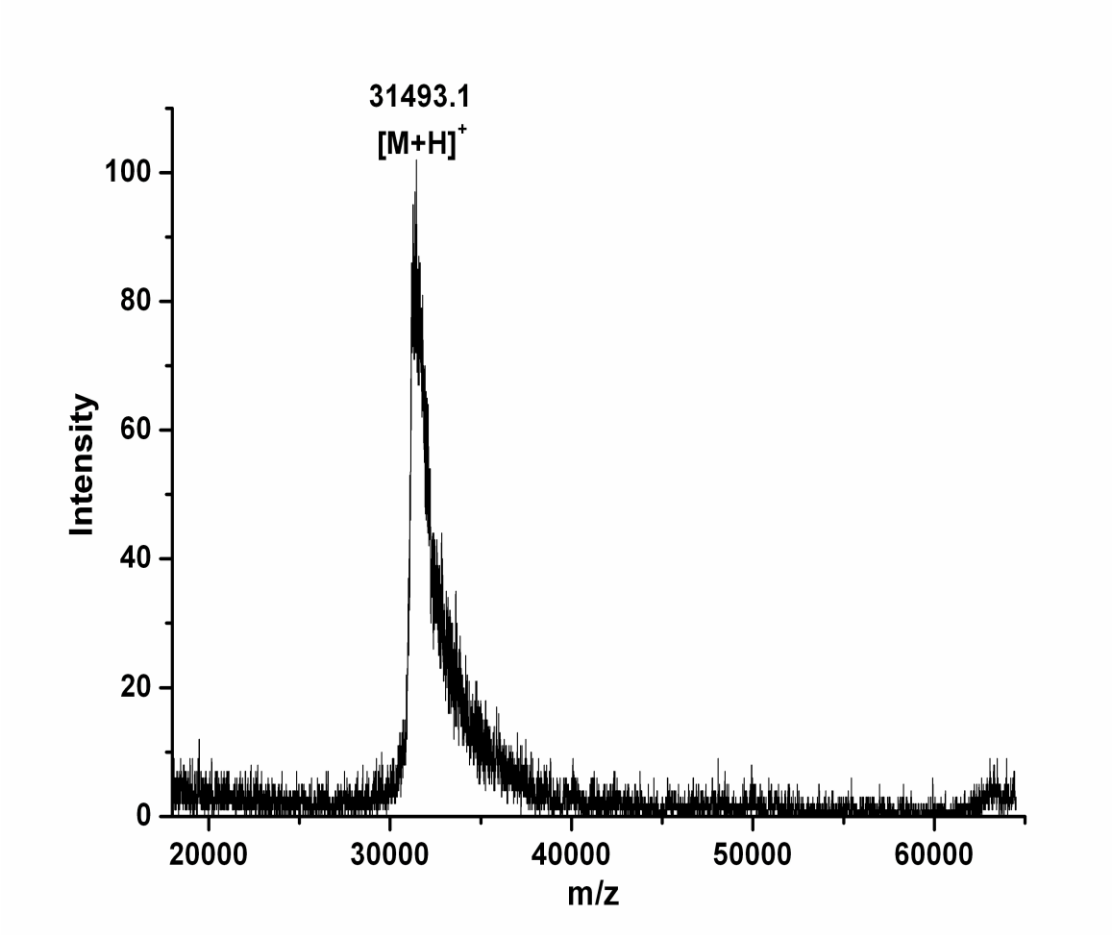


Figure 3.1 MALDI-TOF analysis of the molecular mass of the purified His-tagged C-terminal domain of Stt3p.

stability of the sample is thus an additional variable to take into account during the optimization process.

Since it is not currently possible to determine the best detergent *a priori*, in this study, a number of different detergents including SDS, DPC, LDAO, OG, Digitonin, and DDM were screened. During the detergent exchange process to LDAO, the protein precipitated indicating that LDAO is not a proper detergent to solubilize the C-terminal domain of Stt3p. The above-mentioned five detergents were screened by NMR spectroscopy to determine their suitability for reconstitution of the C-terminal domain of Stt3p. The 2D HSQC spectrum provides both qualitative and quantitative information for the evaluation of whether a protein is well folded and exists in a single conformation. The quality and the number of peaks present in 2D HSQC NMR spectrum reveals whether a protein is monomeric or exists in oligomeric forms. This information is vital to assess the feasibility of further solution NMR based structural characterizations. As shown in Figure 3.2 and Figure 3.3, the quality of HSQC spectra varies markedly as a function of detergent. The HSQC spectrum of the C-terminal domain of Stt3p in DPC micelles, a detergent often found to provide high quality NMR spectra for membrane proteins (103), showed very broad linewidths and a number of missing resonances. Digitonin and DDM micelles produced poorly resolved spectra (Figure 3.2). These observations clearly demonstrate that the C-terminal domain of Stt3p is oligomerized under the above micellar environments. Oligomerization leads to slower tumbling and rapid transverse relaxation rates, which substantially broaden and weaken the resonances, thus dramatically reducing spectral

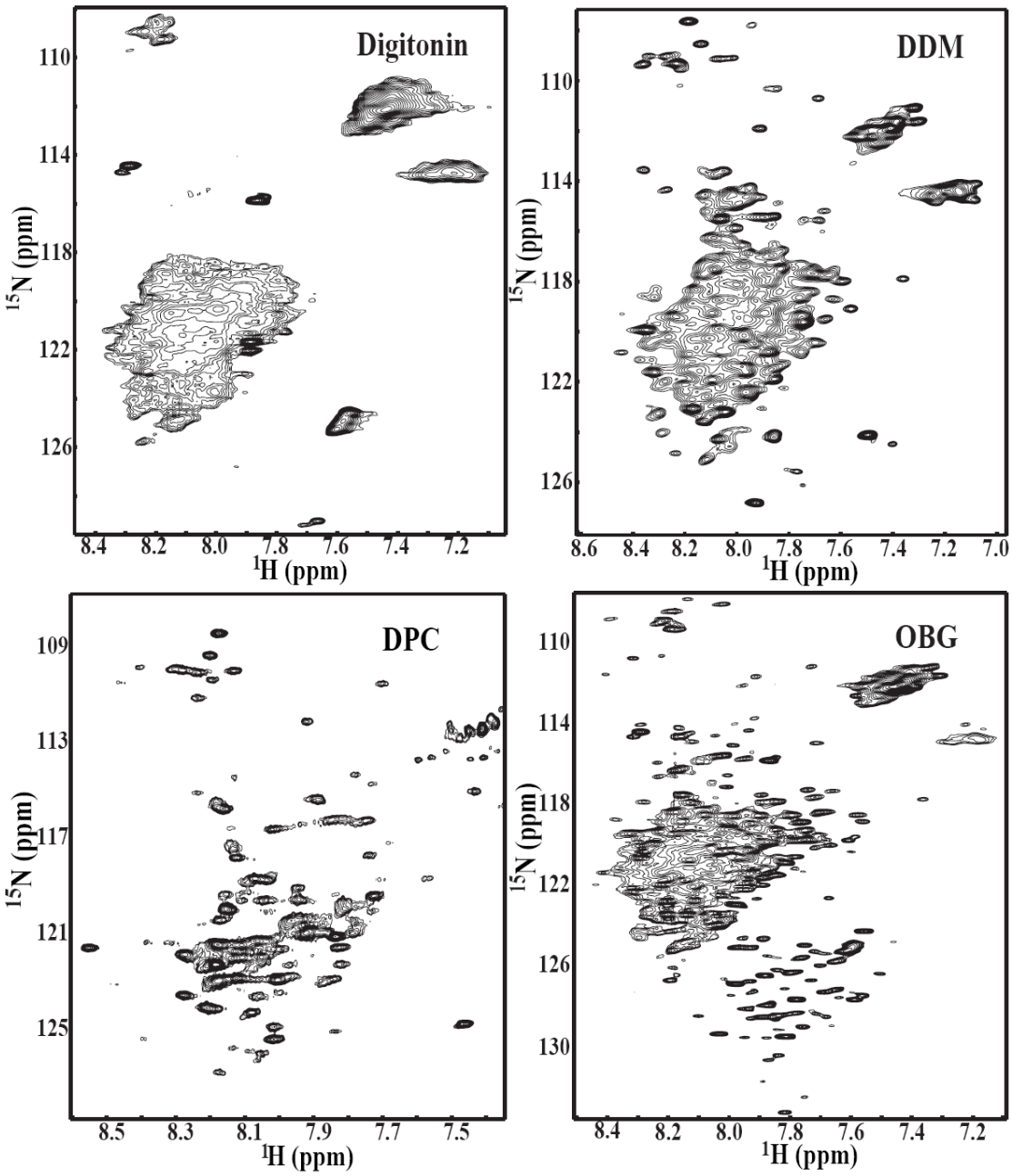


Figure 3.2 2D NMR [^{1}H, ^{15}N] HSQC spectra of the purified [U-^{15}N] His-tagged C-terminal domain of Stt3p in different detergent micelles. (A) 1.5% Digitonin, (B) 1% DDM, (C) 300 mM DPC, and (D) 150 mM OG.

resolution. In the case of the detergent OG, more peaks were observed than expected in the HSQC spectrum, indicating the presence of multiple conformations or oligomeric equilibria. Among all the five detergents tested, SDS was determined to be the best for further NMR based structural characterization. It produced a far superior spectrum (Figure 3.3) with favorable dispersion and narrow linewidths, which indicated that the C-terminal domain of Stt3p was folded into a single stable conformation under the experimental condition. Furthermore, out of 263 non-proline residues, 245 resolved peaks with relatively uniform intensity were counted.

The optimum concentration of SDS was determined by thoroughly investigating the effect of SDS on protein conformation by NMR. Our data indicated that the HSQC spectra were well resolved and closely resembled one another when SDS concentration is in the range of 50-200 mM (Figure 3.4 A-C). However, the spectra started to lose its resolution at a concentration above 250mM. When SDS concentration was increased to 400 mM or above the resonance dispersion became very narrow with many missing peaks indicating that the protein had partially denatured (Figure 3.4 D). Taken together, 100 mM SDS was chosen as the working condition for further NMR characterization.

3.3.3 Characterization by Far-UV and Near-UV CD Spectropolarimetry

Far-UV and near-UV CD spectroscopy were employed to probe the secondary and tertiary structure of the C-terminal domain of Stt3p in 100 mM SDS micelles and 400 mM DPC micelles for comparison. The far-UV CD spectra (Figure 3.5 A) both in SDS and DPC micelles had the characteristics of a typical α-helical protein with CD

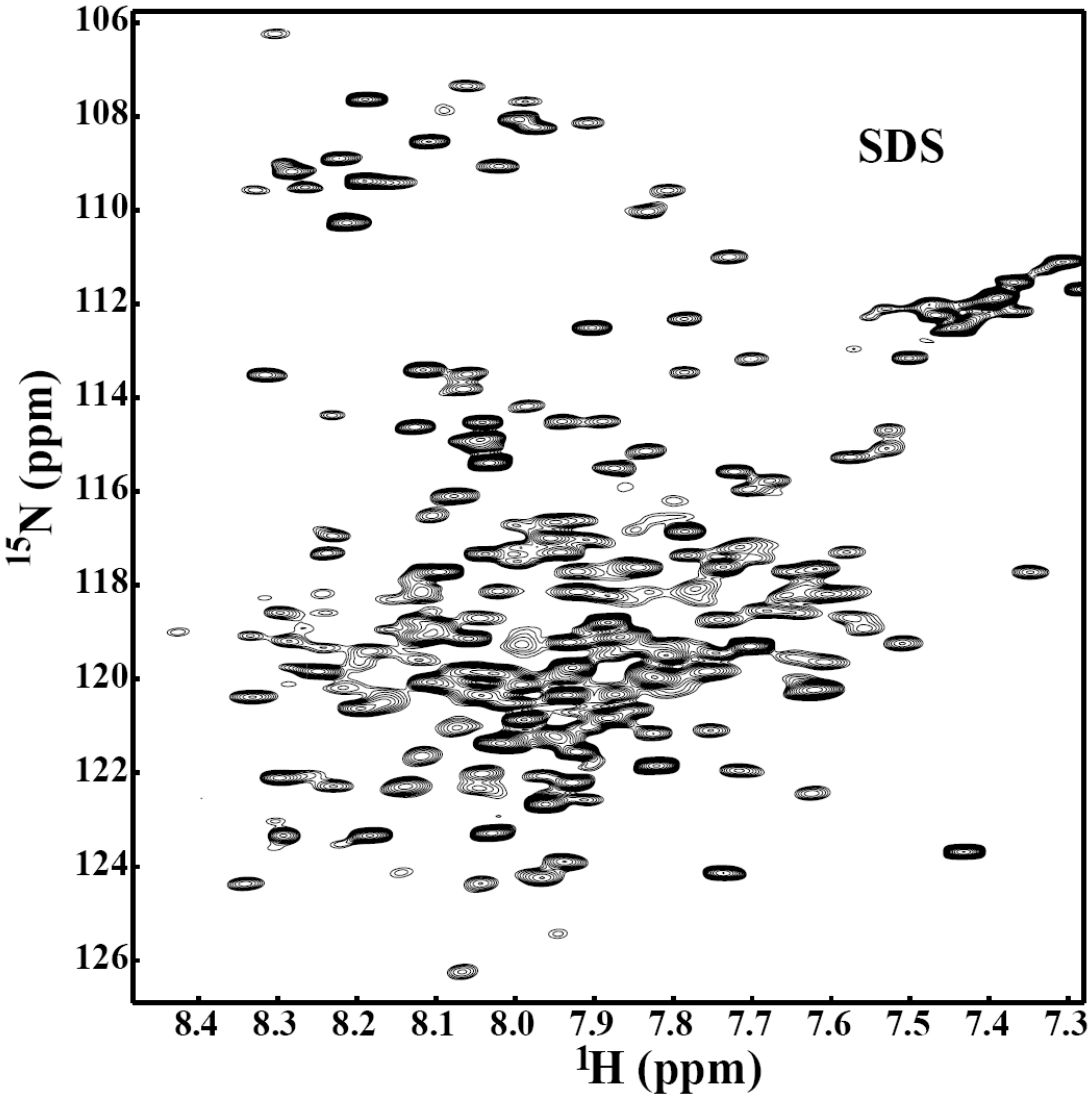


Figure 3.3 2D NMR [^{1}H, ^{15}N] HSQC spectrum of the purified [U-^{15}N] His-tagged C-terminal domain of Stt3p in SDS micelles.

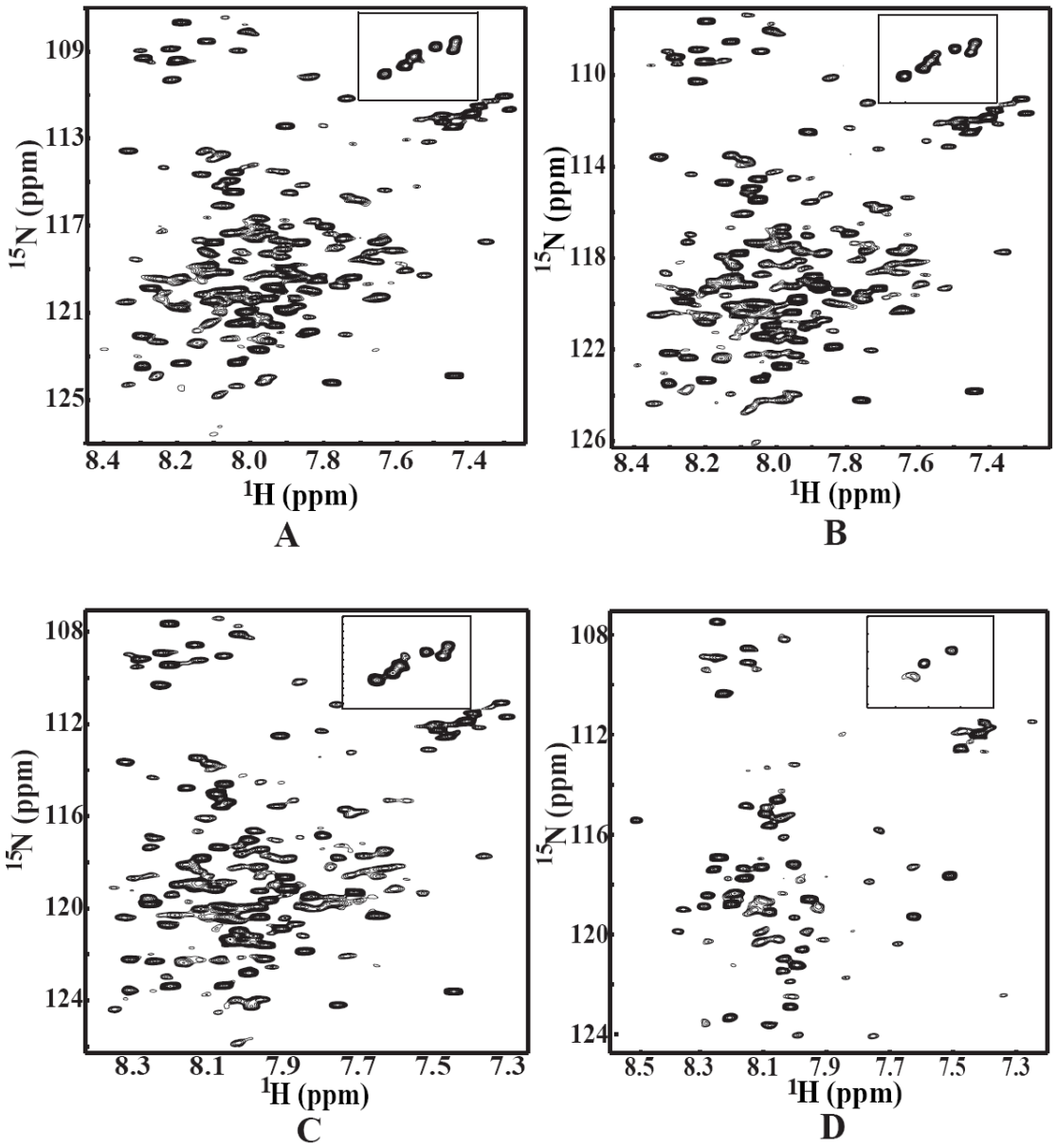
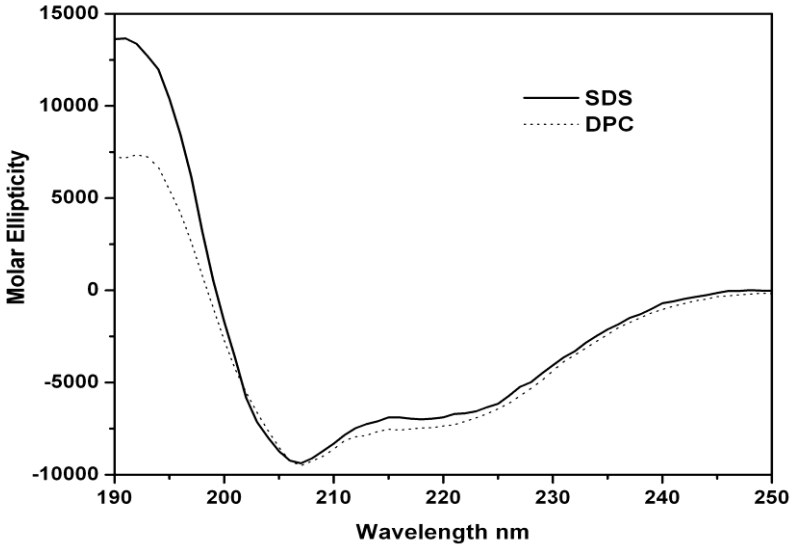
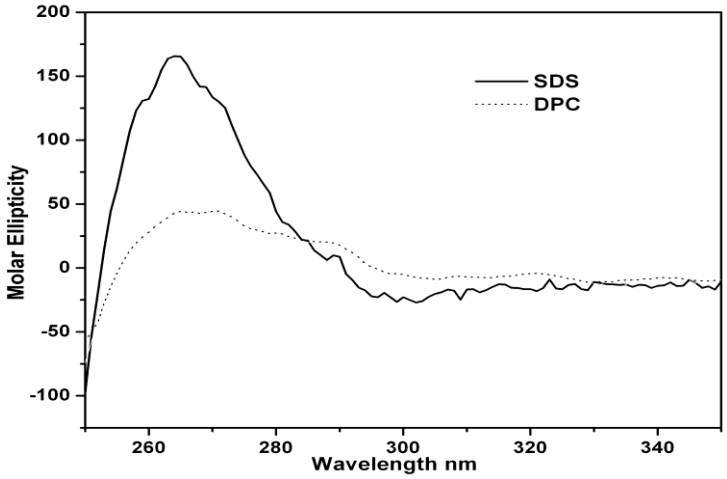


Figure 3.4 2D NMR [^{1}H, ^{15}N] HSQC spectra of the purified [U-^{15}N] His-tagged C-terminal domain of Stt3p as a function of SDS concentration. The inner figure is close-up view of the tryptophan indole amide proton region from the same spectrum. The concentrations of SDS were as follows: (A) 50 mM SDS; (B) 100 mM SDS; (C) 200 mM SDS and (D) 400 mM SDS.



A



B

Figure 3.5 CD spectroscopic analysis of the C-terminal domain of Stt3p. (A) far-UV CD spectra of the C-terminal domain of Stt3p in 300 mM DPC and 100 mM SDS detergent micelles. The protein concentration was 10 μM in 20 mM phosphate buffer, pH 6.5. The characteristic double minima at 208 and 222 nm are indicative of significant α-helical content. (B) near-UV CD spectra of the C-terminal domain of Stt3p in 300 mM DPC and 100 mM SDS detergent micelles. The protein concentration was 89 μM, and the buffer conditions were same as for A.

minima at 208 and 222 nm. This observation is consistent with what was seen in the 2D HSQC spectrum i.e. relatively narrow proton dispersion, which is another indication of a helical protein. These results are also consistent with the crystal structure of its archaea homolog reported recently (79), even though there is only very limited sequence similarities between these two proteins.

Near-UV (250-350 nm) CD spectrum are due to the dipole absorption of the aromatic residues and disulfide bonds (if present), which depends upon the orientation and nature of the surrounding environment of these chromophores, and is therefore sensitive to the overall tertiary structure of a protein. For a protein in an unfolded or molten globule state, one of the classical spectroscopic signatures is the absence of a near-UV signal (110). In other words, the presence of significant near-UV signals is a good indication that the protein is folded into a well-defined structure (111, 112). The presence of near-UV CD signal for the C-terminal domain of Stt3p in 100 mM SDS (Figure 3.5 B) indicates that the protein has a well-defined tertiary structure. Interestingly, the tertiary structure appears to be disrupted in DPC micelles (Figure 3.5 B), which is consistent with the NMR data (Figure 3.2 C) although it is widely believed that DPC is usually a "milder" detergent that generally doesn't denature proteins. Close inspection of the near-UV CD spectrum in SDS micelles reveals that there are three humps, from left to right, which can be attributed to the absorption of phenylalanine, tyrosine and tryptophan respectively.

3.3.4 Intrinsic Tryptophan Fluorescence

Intrinsic fluorescence, especially with tryptophan as a probe, provides a powerful analytical tool for membrane protein studies due to its sensitivity and simplicity (113). The fluorescence emission spectrum of the C-terminal domain of Stt3p (containing 8 tryptophan residues) upon excitation at 280 nm showed a broad emission spectrum with λ_{max} ranging from 330 nm to 350 nm (Figure 3.6). This result

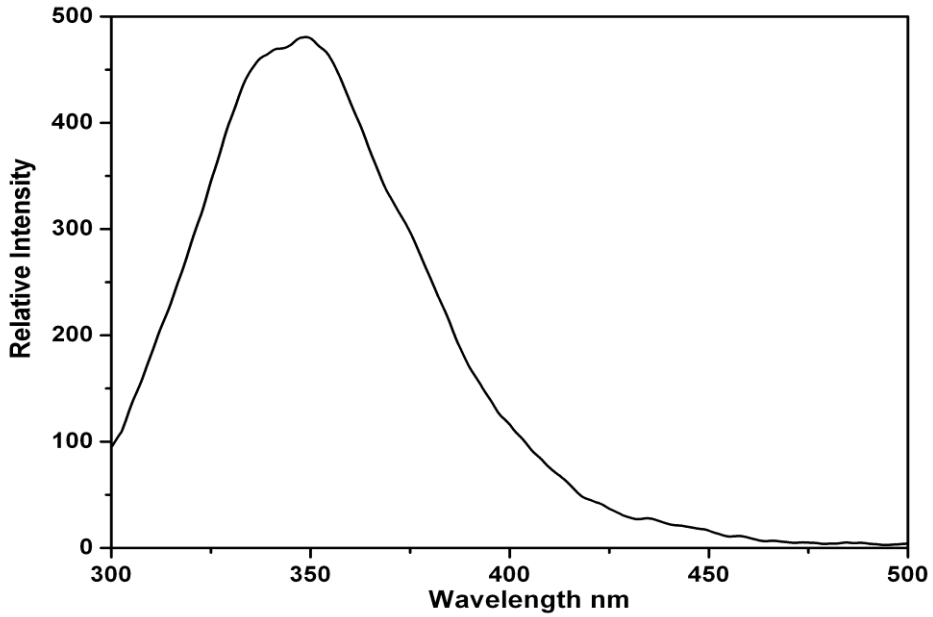


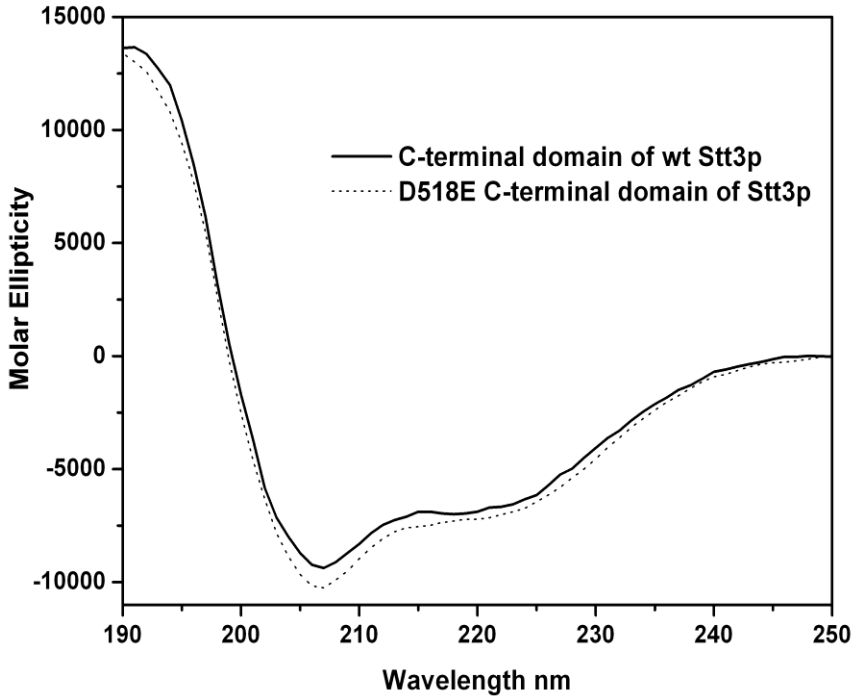
Figure 3.6 Fluorescence emission spectra for the His-tagged C-terminal domain of Stt3p. Protein is in 10 mM phosphate, 100 mM SDS, pH 6.5. Spectrum was recorded from 300 to 500 nm.

indicates that of the 8 tryptophan residues, some are totally buried into the hydrophobic core; some are partially exposed to water, while the rest are completely exposed to water. This result is not surprising at all since in integral membrane proteins, tryptophan residues have been found to show preferential clustering at the membrane interface (114-120). The exact location and orientation of each tryptophan

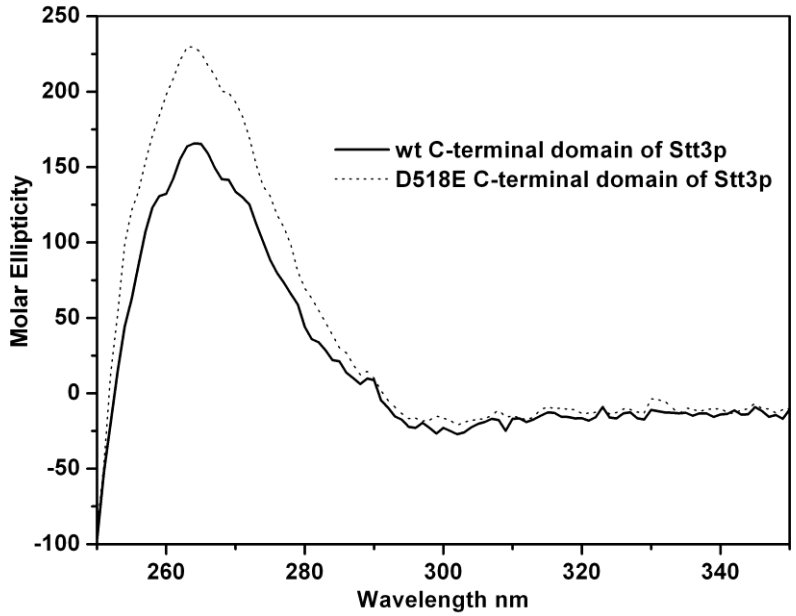
residue can only be clear once the high resolution 3D structure of the C-terminal domain of Stt3p is solved.

3.3.5 Comparison of Wild-type and Mutant Protein

In the C-terminal domain of Stt3p, residues 516-520, which make up the WWDYG motif, are highly conserved through several branches on the evolutionary tree. This motif has been proposed to be directly involved in the glycosylation site recognition and/or in the catalytic glycosylation process based on co-immunoprecipitation, photoaffinity labeling, and both block and single mutational analysis (66). Furthermore, a conservative mutation of a single residue such as Asp^{518} to Glu renders the enzyme completely inactive causing cell death in yeast, *Saccharomyces cerevisiae* (66, 78). This observation demonstrates that there may be a strict geometric or conformational requirement for the enzyme to catalyze the N-linked glycosylation reaction. To investigate whether Asp^{518} acts only as a catalytic base as previously proposed (66), or has any other role in the conformational geometry required for catalysis, we carried out a detailed biophysical characterization of both the wild-type and the D518E mutant under identical conditions. The far-UV CD spectra (Figure 3.7 A) of the wild-type Stt3p C-terminal domain and that of D518E mutant are very similar, suggesting that there is no significant change in the secondary structure upon point mutation. In contrast, there are significant differences in the near-UV CD spectra, which reveal that both proteins have distinct tertiary structures (Figure 3.7 B). This evidence is further supported by measurements of their tryptophan fluorescence spectra. The D518E mutation led to an apparent blue-shift as



A



B

Figure 3.7 CD spectra of the wild-type and D518E mutant (continued on following page).

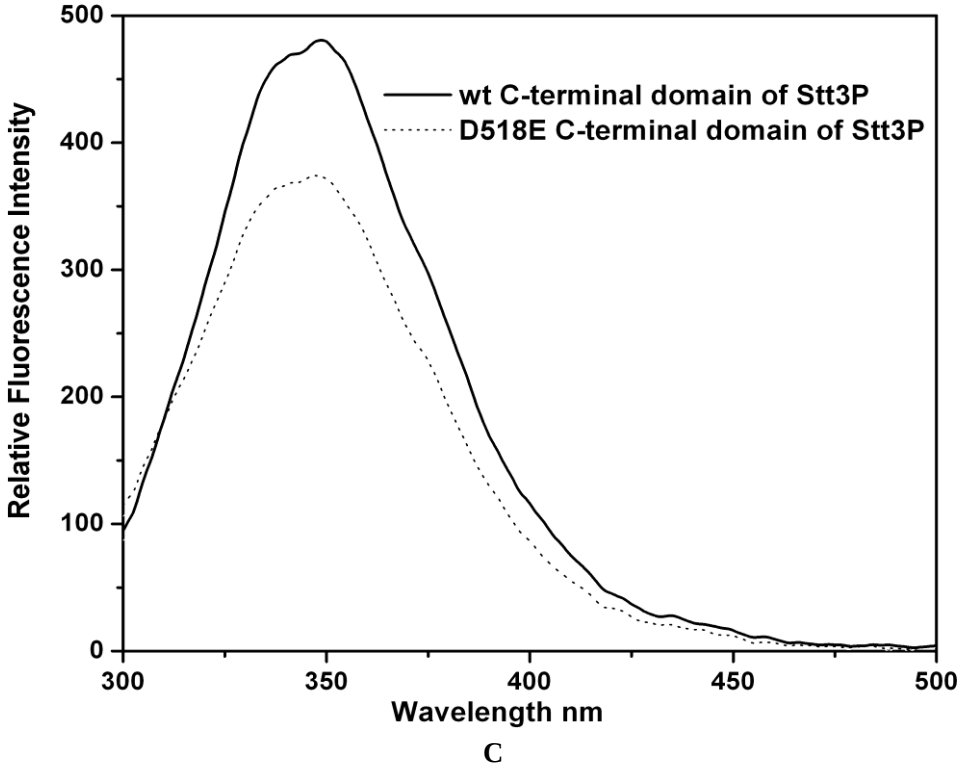


Figure 3.7 CD spectra of the wild-type and D518E mutant. The data were collected under the same conditions. A: far-UV CD spectra. The protein concentrations were 10 μM in 20 mM phosphate buffer, pH 6.5, 100 mM SDS. B: near-UV CD spectra. The protein concentrations were 89 μM in 20 mM phosphate buffer, pH 6.5, 100 mM SDS. C: intrinsic tryptophan fluorescence spectra. The protein concentrations were 1 μM in 10 mM phosphate buffer, pH 6.5, 100 mM SDS. The introduction of the mutation leads to an intensity quench and blue shift of the spectrum.

well as quenching of the actual intensity of the fluorescence emission of the wild-type protein (Figure 3.7 C), indicating the change in the microenvironments of the tryptophan residues. This observation demonstrates that the D518E mutation did change the structure of the C-terminal domain of Stt3p affecting the microenvironment and solvent exposure of some tryptophan residues, most likely neighboring W516 and W517 (105, 121).

NMR is an extremely powerful technique to monitor the changes in the conformation of a protein sample due to change in pH, temperature, salt or addition of a ligand. The 2D [^{1}H, ^{15}N]-HSQC spectrum represents the "fingerprint region" of a protein. This region is extremely sensitive and any dramatic perturbation in the chemical shifts or resonances from the original positions may suggest a change in the conformation of the protein. This change can be local, involving few residues or a global conformational change involving most of the residues in the protein. In the present study, HSQC spectra were collected to compare the fingerprint region of the wild-type and the D518E mutant in SDS micelles under identical conditions. It is clear from Figure 3.8 that the D518E point mutation induced drastic changes in the chemical shift positions of a number of peaks indicating that wild-type and the D518E mutant have distinctly different conformations. This conformational change cannot be attributed to the change of the local environment around Asp^{518} since chemical shift perturbation is dramatic for most of the peaks in the HSQC spectrum. In fact, some of the resonances observed in the wild-type HSQC spectrum did disappear in the spectrum of the mutated protein. This observation demonstrates that the D518E

mutation, indeed affects both conformation and dynamics of the wild-type protein, which may have bearing on OT function.

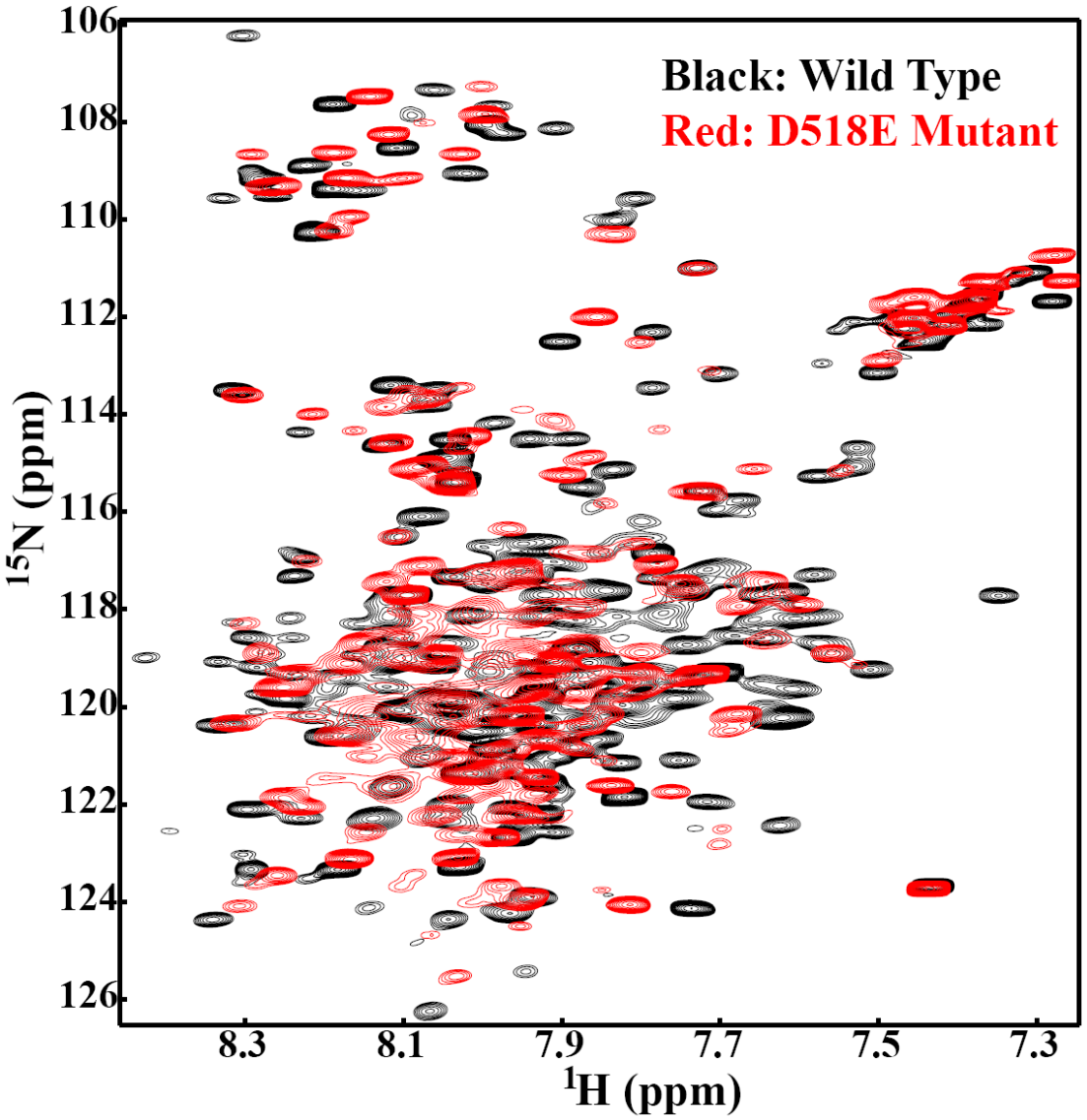


Figure 3.8 The impact of the D518E mutation on the 2D [^{1}H, ^{15}N] -HSQC spectrum. The black spectrum represents the wild-type, while the superimposed red spectrum is of the D518E mutant of the C-terminal domain of Stt3p.

3.3.6 Acceptor Substrate Binding Studies by STD Spectroscopy

To investigate the interactions of acceptor substrate of OT with the C-terminal domain of Stt3p, binding studies were carried out with a six-residue peptide containing the consensus N-linked glycosylation sequon by saturation transfer difference (STD) NMR spectroscopy. STD has been proven to be a powerful method to probe low affinity interactions ($K_D \approx 10^{-8}$ to 10^{-3} M) of small molecules with proteins (122-127). In the STD technique, selective saturation of a protein resonance leads to a rapid spread of the magnetization over the entire protein *via* spin diffusion, and intermolecular transfer of magnetization from protein to ligand leads to changes in NMR signal intensity of the ligand. However, for interaction studies involving proteins and peptides, attention should be paid to make sure a well separated peak in the protein is picked for STD experiment. Thus the saturation resonance must exclusively belong to protein. Moreover, the resulting signals in STD spectra must exclusively belong to peptide ligand. The latter is especially true if incomplete protein signal suppression occurs.

To overcome these, here, a methyl-protonated {Ile(δ_1 only), Leu($^{13}CH_3$, $^{12}CD_3$), Val($^{13}CH_3$, $^{12}CD_3$)} U-{^{15}N, ^{13}C, ^{2}H} labeled sample of the C-terminal domain of Stt3p was prepared by using biosynthetic precursors (128) . This labeling pattern is extremely desirable for STD studies since in these labeled proteins, except for water-exchangeable protons, only the methyl groups of the Ile (δ_1 only), Leu and Val residues are protonated. On one hand, the commonly used regions for irradiation of protein remain, such as the up-field region (at around 0 ppm) or down-field region

(about 10 ppm). On the other hand, the simplified protein spectrum facilitates the data analysis process significantly and reduces the risk of having a pseudo-positive effect resulting from incomplete elimination of background protein signals. The C-terminal domain of Stt3p and peptide ligand complex was irradiated at 0.738 ppm, where no peptide NMR signal was present. The peaks a, b, c, d, e, and f in STD spectrum exclusively correspond to the peaks 1, 2, 3, 4, 5 and 6 respectively in the NMR spectrum of acceptor peptide (Figure 3.9B and 3.9C). The appearance of the NMR peaks of the peptide ligand in the difference spectrum unequivocally indicates that the acceptor peptide ligand is bound to the C-terminal domain of Stt3p. More importantly, close inspection of the difference spectrum reveals that the amide protons (peak a, which has a chemical shift of 7.13 ppm) on the side chain of Asn residue, the N-glycosylation site, are significantly affected by the saturation pulse (Figure 3.9C), which strongly suggest that the side-chain of Asn residue is directly involved in the protein-substrate recognition process.

3.3.7 Acceptor Substrate Affinity Studies by NMR Titrations

To further determine the affinity of acceptor substrate (of OT) with the recombinant C-terminal domain of Stt3p, titration studies were carried out with Asn-Asp-Thr-NH_2 acceptor peptide containing the consensus N-linked glycosylation sequon. Substrate binding was followed by monitoring the changes in chemical shift positions in the fingerprint region of the protein in 2D HSQC spectra as shown in Figure 3.10. The chemical shift perturbation of four representative peaks were fitted to Hill model by using Origin® 7.0 software. As shown in Figure 3.11, upon addition

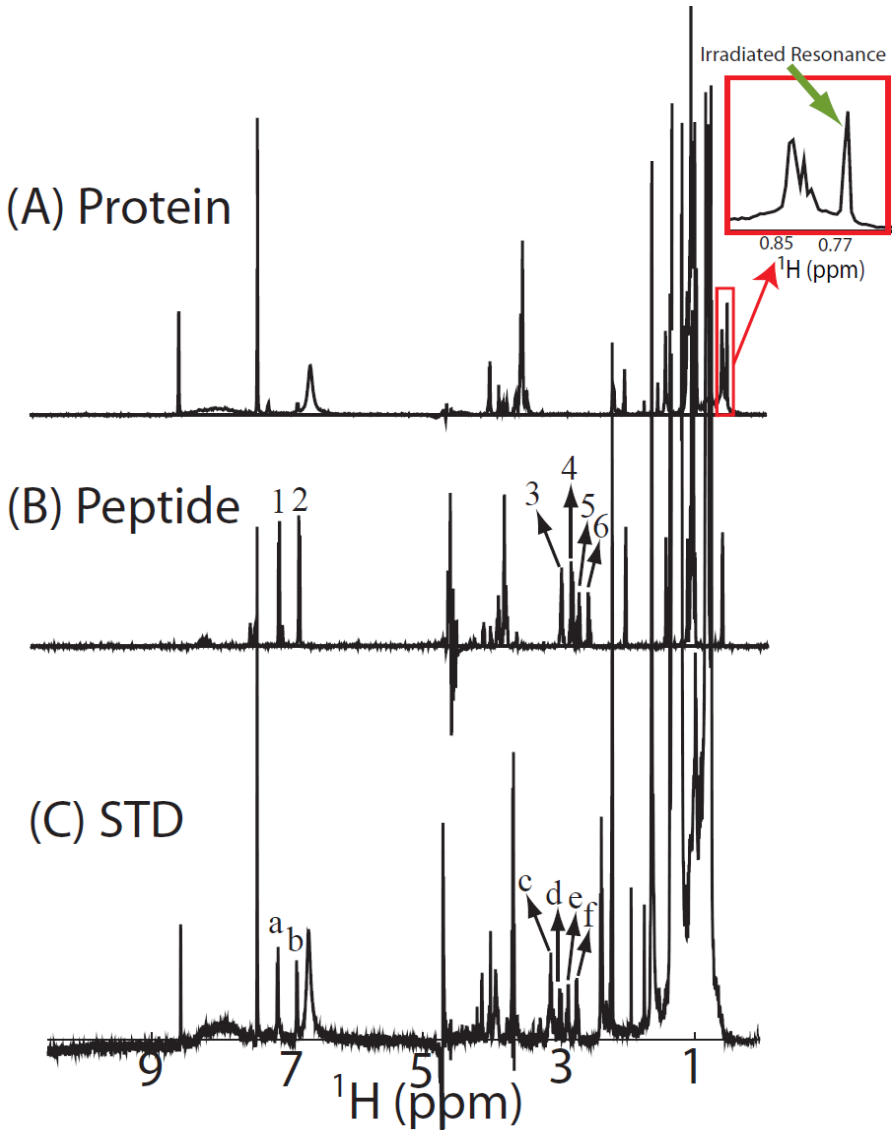


Figure 3.9 STD studies of substrate binding. A: 1D NMR spectrum of ILV-labeled sample of the C-terminal domain of Stt3p. Irradiated resonance is indicated by green arrow in the upper-left enlarged spectrum. B: 1D NMR spectrum of peptide ligand, Tyr-Asn-Ser-Thr-Ser-Cys-Am. C: STD NMR spectrum of the complex of ILV-labeled sample of the C-terminal domain of Stt3p and acceptor peptide substrate. The appearance of peaks a, b, c, d, e and f in STD spectrum, which correspond to the peaks 1, 2, 3, 4, 5 and 6 in the NMR spectrum of acceptor peptide, reveals the C-terminal domain of Stt3p binds to the acceptor substrate of OT.

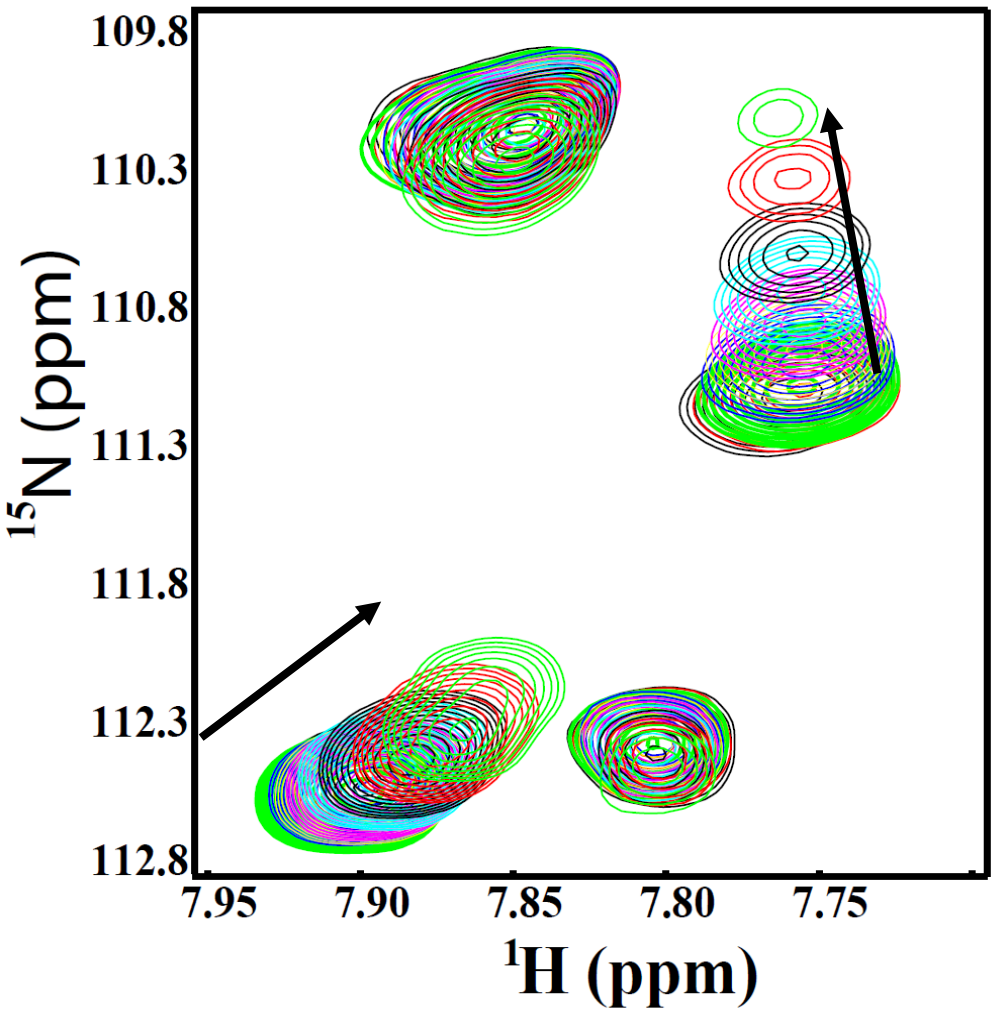


Figure 3.10 Substrate binding studies by HSQC titrations. An expanded region of the overlay of 2D [^{1}H, ^{15}N]-HSQC spectra of the [U-^{15}N]-labeled C-terminal domain of Stt3p (170 µM) showing changes in the chemical shift positions upon addition of increasing concentration of the substrate peptide. Ratios of protein to peptide are: 1:0 (black), 1:0.5 (red), 1:1 (green), 1:5 (blue), 1:10 (yellow), 1:20 (purple), 1:35 (cyan), 1:50 (black), 1:75 (red) and 1:100 (green).

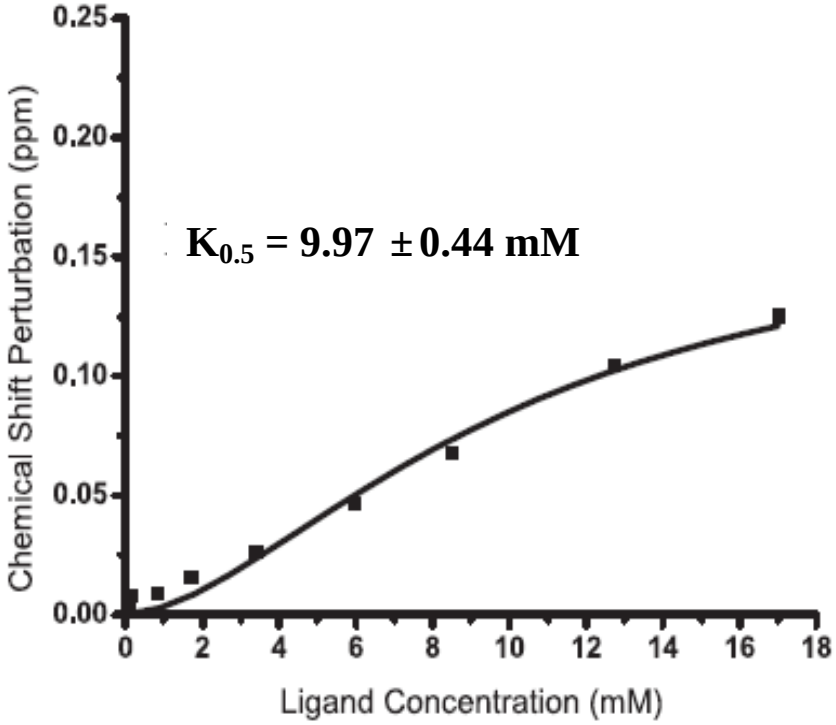


Figure 3.11 The chemical shift perturbations upon substrate addition. The chemical shift perturbation of average of four representative resonances are plotted as a function of the concentration of the substrate peptide and fitted using Hill model of Origin® 7.0 software.

of ligand substrate, the C-terminal domain of Stt3p exhibits a sigmoidal saturation curve, and the substrate peptide binds the protein with an apparent $K_{0.5}$ of 9.97 ± 0.44 mM and Hill coefficient n of 1.70. These results suggest it is a positively cooperative binding (n > 1) with relatively low affinity.

3.4 Discussion

3.4.1 Feasibility of Structure Determination by Solution NMR

To carry out structure determination of any membrane protein by solution NMR, detergent screening to find the suitable membrane mimetic is an essential prerequisite. The suitability of a detergent micelle is determined by taking into account the protein solubility and stability along with the quality of the 2D HSQC NMR spectrum. The 2D HSQC spectrum correlates the amide proton and the corresponding nitrogen pair for each residue within a protein and provides a map of the fingerprint region. It also serves as a building block for a multitude of multidimensional NMR experiments upon which the resonance assignments and the determination of the 3D structure of a protein rest. Thus, obtaining high quality, i.e. sufficiently resolved HSQC spectra is imperative for structural characterization by solution NMR.

To find a suitable detergent for obtaining a homogeneous sample of the C-terminal domain of Stt3p, six detergents were screened. These include digitonin, which has been successfully used to extract and reconstitute the OT complex in microsomes (129); SDS and DPC, which are commonly used for solution NMR; LDAO, DDM and OG, the common detergents for membrane protein crystallization.

As expected, digitonin gave unresolvable HSQC spectra due to its large micellar sizes (70 kDa). For all the rest of the detergents except OG and SDS, the protein appeared to be oligomerized leading to poorly resolved spectra with broad linewidths and missing resonances. For OG, while it has a small micellar size (25 kDa), it seems that its short alkyl chain (C8) jeopardizes the protein conformational stability since the number of HSQC peaks is much higher than expected. In contrast, SDS micelles yielded an HSQC spectrum that was far superior in quality in comparison to all the rest of the detergents that were screened in all of the aspects: number of resonances, signal to noise ratio (data not shown), dispersion, linewidths and uniformity of signal intensities. In fact, SDS has served as one of the most popular detergents for IMPs studies (130), and has been widely used as a membrane mimetic for membrane protein structural and functional studies (46, 130-135). The high quality HSQC spectra along with the CD, and fluorescence data suggest that the C-terminal domain of Stt3p in SDS micelles, is well-folded producing a homogeneous sample. The above observations support the feasibility of conducting solution NMR-based structural studies of the C-terminal domain of Stt3p.

In fact, the quality of the 2D HSQC spectrum is much better than what would be expected for such a large protein–detergent complex, implying relatively small relaxation rate. The ^{15}N T_1, T_2 relaxation measurements show that the rotational correlation time for the C-terminal domain of Stt3p in SDS micelles is surprisingly short- ~10 ns, a value expected for a 20 kDa protein tumbling isotropically in solution (data not shown). This, however, is consistent with the results reported by Krueger-

Koplin, *et al.* (46), where a survey of seven membrane proteins in different detergent micelles showed a rather short rotational correlation time ranging from 8 to 12 ns. According to these authors, this phenomenon can be attributed to the fluid property of detergents which allows rotation of the proteins within the confines of the micelle. In the case of the C-terminal domain of Stt3p, an alternative explanation is its flexible dynamic property. The high flexibility of the C-terminal domain of Stt3p is reasonable since N-linked glycosylation is co-translational. Therefore, only a flexible active site can recognize glycosylatable sequons rapidly and efficiently in all different types of growing polypeptide chains. This ensures the rapid product discharge from the active site. Furthermore, the flexibility of this domain is supported by the cryo-electron microscopy structure of the yeast OT, which shows a flexible groove formed between the luminal domains of Ost1p, Wbp1p, and Stt3p (89). This groove is proposed to thread and scan the unfolded nascent polypeptide chain (89).

3.4.2 Comparison of the Wild-type C-terminal Domain of Stt3p with D518E Mutant

One striking feature of the C-terminal domain of Stt3p is that it is highly conserved in eukaryotes. Actually, the sequence alignment shows that, from yeast to humans, the sequence identity is over 50% (see Figure 2.1 in Chapter two). The strictly conserved "WWDYG" motif is believed to be the catalytic and/or acceptor protein recognition site. The aspartate residue (Asp^{518}) of this conserved motif was thought to function as a catalytic base (66). However, it appears that the role of Asp^{518} is more than simply to act as a base in the catalysis since the D518E mutation results

in a complete loss of enzyme activity, even though both Asp and Glu residues have similarly charged side chains (side-chain pKa values for Asp and Glu are 3.9 and 4.1, respectively). If the role of Asp is just to act as a base, then how can the loss of activity for D518E be explained?

To address the above question, comprehensive biophysical characterizations of the D518E mutant and wild-type C-terminal domain of Stt3p were carried out. Interestingly, while both the wild-type and D518E mutant share nearly identical secondary structural contents, they have distinctly different tertiary structures as revealed by near-UV CD, fluorescence and NMR spectroscopies. The most direct evidence for this conclusion has come from the comparison of their 2D HSQC spectra. The replacement of Asp^{518} with the longer Glu side chain leads to large global changes in the structure involving nearly all of the amino acid residues (Figure 3.8). This observation led to the conclusion that the residue Asp^{518} is critical to maintain the catalytically active conformational geometry of the C-terminal domain of Stt3p.

Additionally, the apparent disruption of the active conformation after a point mutation strongly suggests that the C-terminal domain of Stt3p has folded into its native conformation in SDS micelles. This is based on the fact that it is very unlikely to change the 'structure' of a protein that is denatured or in a molten globule state by the replacement of one residue with a structurally similar residue.

The loss of enzyme activity by mutation of Asp→Glu is not that common, but OT is not unique in this regard (66). In fact, for the enzyme Ca^{2+}-ATPase, mutation of

D601E and D707E result in an inactive enzyme (136). More importantly, the residues $Asp^{601}$and Asp^{707} have been proposed to play structural but not catalytic or substrate recognition roles. It is therefore likely that OT and ATPase may have similar mechanisms of function. For example, both of these two enzyme complexes need metal ions to be active (137, 138); both the enzymes catalyze energy transfer from phosphate ester bond cleavage; and both the enzymes undergo allosteric transition upon substrate binding (139). It seems logical to compare these two enzyme complexes from an evolving view of enzymatic studies.

3.4.3 Functional Probing of the C-terminal Domain of Stt3p

We conducted STD experiment and HSQC titrations to probe the in vitro protein-substrate interaction. Our results demonstrate that the C-terminal domain of Stt3p interacts with the acceptor peptide substrate containing the N-linked glycosylation recognition motif. The strong signals that belong exclusively to the acceptor peptide were observed in the STD spectrum, while chemical shift perturbations were observed in the HSQC experiments upon addition of the substrate peptide. These observations provide direct experimental proof that the C-terminal domain of Stt3p contains the recognition site for the N-glycosylation acceptor substrate even though the affinity is relatively low (K_D~10 mM). One explanation could be that SDS micelles may not mimic the native lipid bilayer, which may impair the activity of the protein to some extent. Additionally, since the functional OT complex is composed of eight different subunits, it is more likely that while the C-

terminal domain of Stt3p possesses the substrate recognition site, the other subunit(s) may facilitate the binding process (89).

The C-terminal domain of Stt3p in SDS micelles has a short rotational correlation time of ~10 ns, suggesting that it is a monomer under the experimental conditions. However, the sigmoidal saturation curve observed upon acceptor substrate binding indicates that this monomeric protein is allosterically activated, suggesting that it may contain more than one binding site. The binding of a peptide substrate (allosteric activator) to the activator site results in an increased affinity in the second site (active site). The detailed regulatory mechanism can be addressed only by further structure-function studies.

CHAPTER 4

NMR ASSIGNMENTS OF THE C-TERMINAL DOMAIN OF STT3P

"Be practical as well as generous in your ideals. Keep your eyes on the stars, but remember to keep your feet on the ground!" Theodore Roosevelt, 1904

4.1 Introduction

As mentioned in Chapter one, atomic-resolution structures of membrane proteins are essential to a wide range of biomedical and biotechnological applications of IMPs. However, structural research on membrane proteins remains largely an unexplored area due to various technical problems.

Structure determination by NMR spectroscopy usually consists of several essential steps, each using a separate set of highly specialized techniques. These conventional steps include:

(1) Sample preparation, including the preparation of ^{13}C, ^{15}N-double labeled and/or ^{2}H, ^{13}C, ^{15}N-triple labeled protein samples. If necessary, a series of triple labeled protein samples with different deuteration levels must also be prepared.

(2) NMR data collection, including series of 2D, 3D and even 4D homonuclear and heteronuclear NMR experiments.

(3) Resonance assignments, including backbone assignment, side chain

assignment and NOE assignment.

(4) Restraints generation, including incorporation of NOE information, dihedral angles (derived from the coupling constants, or for big proteins, from the backbone and side-chain chemical shifts), and hydrogen bonds (H-bonds).

(5) Structure calculation by computer programs.

Among these, the task of resonance assignments is usually the most time-consuming step. In order to make the problem of resonance assignment more tractable, in the last several years, some powerful approaches have been developed and technical improvements in NMR instrumentation have been generated. These include the introduction of new NMR methods (such as TROSY), optimization of existing pulse sequences, new protein sample labeling strategies, higher field magnets (up to one Gigahertz), *etc*. It has long been recognized that the process of NMR spectral analysis could be accomplished by automated, computational approaches (140). However, to date, the success of that approach is limited to some small water soluble proteins. For membrane proteins and proteins of larger than medium size (MW > 25 kDa), resonance assignment remains a laborious, time-consuming, and daunting task.

The assignment of protein backbone is usually the first step of resonance assignments, and once achieved, it can be extended to the aliphatic side-chain carbons and protons in a straightforward manner using a set of TOCSY (Total Correlation Spectroscopy) and COSY (Correlation Spectroscopy) type experiments. NOE peak assignment is vital for structure determination, as it serves as the primary source of structural constraints for structural calculation. In theory, this step can be readily

accomplished by comparing the chemical shifts of the cross-peaks on NOESY with the previously completed backbone and side-chain assignment.

In this Chapter, the NMR assignments of the C-terminal domain of Stt3p are discussed.

4.2 Backbone Assignments and Chemical Sift Index (CSI) Analysis

4.2.1 Introduction

For structural investigation of proteins by NMR spectroscopy, the backbone assignment is the initial stage, and at the same time, an essential step. In this step, each nucleus on the protein backbone (such as backbone amide groups, C^{α}, C^{β} as well as carbonyl carbon atoms) must be associated with the resonances in the correlated NMR spectrum. Resonance assignments must be sequence specific, i.e., each resonance must be assigned to a spin in a particular amino acid residue within the protein sequence. Despite great progress toward automation of assignment, for large proteins and membrane proteins, most crucial analysis steps must be accomplished manually.

The critical strategy in the protein backbone resonance assignment, also known as the "sequential assignment" strategy, was first developed by Wüthrich and coworkers using a set of 2D NMR experiments on unlabeled protein samples about 26 years ago (141). Nowadays, the assignment strategy makes uses of uniformly isotopically enriched protein samples, and a series of well-constructed highly efficient 3D NMR experiments, which are based primarily on one-bond J-couplings between adjacent atoms.

As mentioned in Chapter One, the most common 3D NMR experiments used for protein backbone assignment are: HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCANH and CBCA(CO)NH. For highly deuterated {^{2}H, ^{13}C, ^{15}N}-triple-labeled protein samples, the last two experiments are replaced by HNCACB and HN(CO)CACB, respectively, due to the absence of the aliphatic protons. These 3D heteronuclear correlation experiments make use of one-bond ^{13}CO(i-1) – ^{15}N(i), ^{15}N(i)-^{13}Cα(i) and ^{13}Cα(i)-^{13}CO(i), as well as two-bond ^{13}Cα(i-1)-^{15}N(i) couplings (Figure 1.2). In this manner, the backbone resonances of both residue (i-1) and (i) or just residue (i) are correlated with the amide group of residue (i). Therefore, sequential assignment is achieved and confirmed by linking the resonances of one residue with those of its adjacent neighbor through multiple independent pathways (Cα, Cβ, and CO). The main reason for correlation of the backbone resonances with amide groups is that the amide groups are the usually best resolved set of signals.

Once the protein backbone assignments are achieved, secondary structure can be determined by a method called CSI (Chemical Shift Index), which was developed by Wishart *et al*. (142, 143). The CSI method uses backbone chemical shift data to identify protein secondary structure. This is based on the widely accepted notion that the chemical shifts of a protein contain its structural information (144-149). As reported, CSI can be used to identify and locate the protein secondary structure with a predictive accuracy in excess of 92% in absence of NOE data.

A hallmark of the historical development of biological NMR spectroscopy is the continued increase in the size of the molecular species amenable to investigation. For

water-soluble protein, the backbone resonances of a 723-residue protein were assigned successfully 8 years ago (150). However, for an α-helical membrane protein, it remains difficult. Until now, the largest helical membrane protein whose backbone assignment has been accomplished contained 241-residue (151).

In this section, the backbone assignment of the His-tagged C-terminal domain of Stt3p will be shown. To our knowledge, this is now the largest helical membrane protein (274 residues including the His-tag) for which backbone assignment has been achieved. Moreover, there were some interesting findings during this process worth sharing, such as the unambigious identification of Isoaspartyl linkage and Proline cis/trans isomerizational linkage by NMR.

4.2.2 Methods and Materials

4.2.2.1 Overexpression and purification of ^{2}H ^{13}C, ^{15}N-labeled C-terminal domain of Stt3p

[^{2}H, ^{13}C, ^{15}N]-labeled C-terminal domain of Stt3p was obtained by overexpression from cultures of *E. coli* BL21(DE3) codon plus cells transformed with the plasmid pET-28c containing an IPTG-inducible gene for C-terminal domain of Stt3p with a C-terminal hexa-histidine tag. The protocol used for expression of the uniformly triple-labeled C-terminal domain of Stt3p was shown in Figure 4.1. Briefly, the transformed cells picked from LB agar plate were grown in 3 mL of LB medium at 37 ℃ for 3 h, transferred to 12.5 mL of unlabeled minimal M9/H_2O medium, and grown until an OD_{600} of ~ 0.5. The cells were separated from the medium by centrifugation at 3,000 rpm for 15 minutes and transferred to 50 mL of M9/D_2O culture containing 0.12% (m/v) of $^{15}NH_4Cl$ as the sole nitrogen source and 0.4% (m/v)

of ^{13}C -glucose as the sole carbon source. At $OD_{600} \approx 0.5$, the culture was diluted to 300 mL with M9/D_2O. The expression of the His-tagged protein was induced at OD_{600} ~ 0.4 with 0.5 mM IPTG, and the culture was allowed to grow for an additional 11-12 h at 30 ℃ (final OD_{600} ~ 2.0), at which point the cells were harvested by centrifugation. The *E. coli* cell pellets were passed through 4 cycles of freeze-thaw using liquid nitrogen and ice respectively before resuspended in B-PER solution (Pierce). The cells were then subjected to sonication (10×15s); the supernatant was removed after centrifugation at 10,000 × g for 30min. The pellet was resuspended once with 10% BPER solution, sonicated and centrifuged again as above. The inclusion bodies were stored at -20 ℃ until needed.

The ^{2}H, ^{13}C, ^{15}N-labeled C-terminal domain of Stt3p was purified by following the protocols described previously.

4.2.2.2 NMR samples

NMR samples contain 20 mM sodium phosphate buffer (pH 6.5), 1% (v/v) glycerol, 100 mM sodium dodecyl-d25 sulfate (SDS, Aldrich), 1 mM EDTA, 10% D_2O. Optimal spectral were obtained with a protein concentration of ~0.6 mM. Deterioration of spectra was observed at higher concentrations, presumably because of protein aggregation.

4.2.2.3 NMR spectroscopy

All NMR experiments were carried out at 328 K using uniformly ^{2}H, ^{13}C, ^{15}N-labeled C-terminal domain Stt3p. HNCACB and HN(CA)CO experiments were collected both by a Bruker Avance 600 MHz spectrometer at our department and a

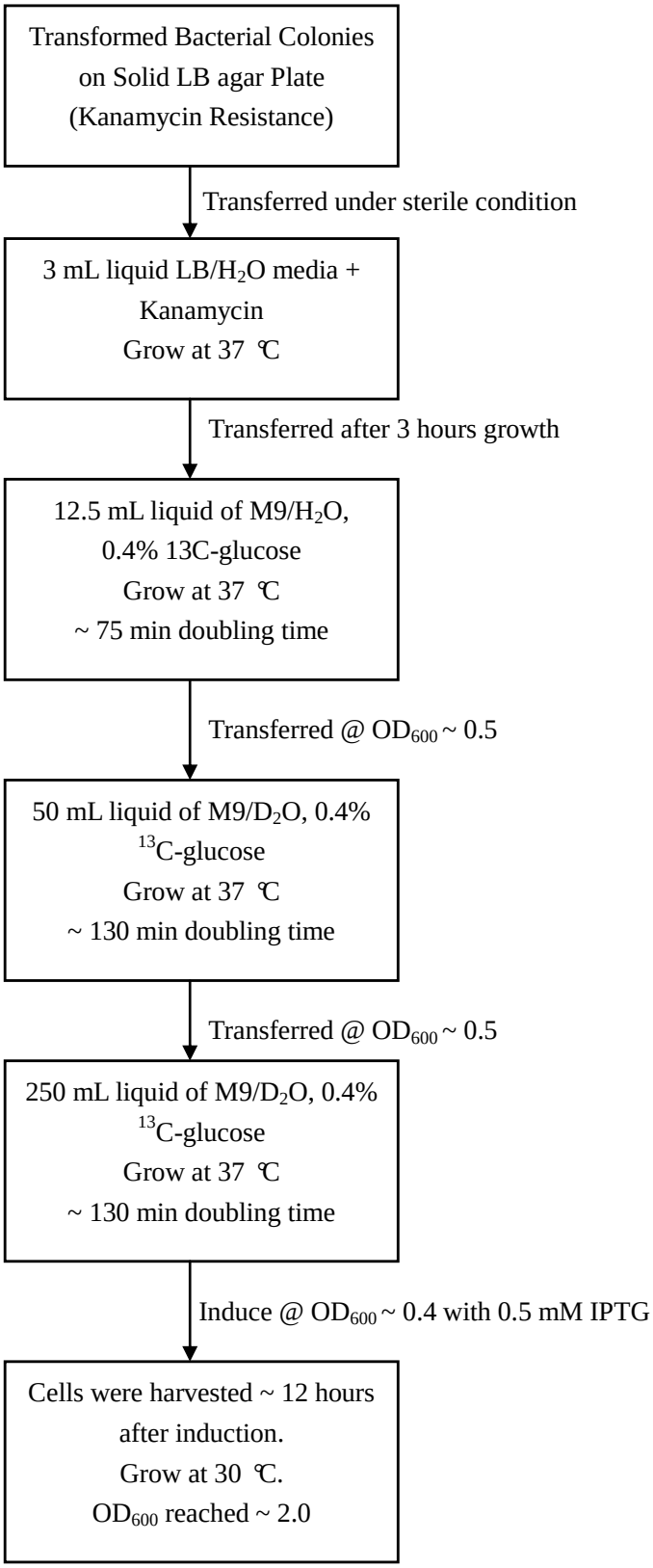


Figure 4.1 Expression protocols for producing the highly deuterated C-terminal domain of Stt3p in *E. coli.*

Varian Inova 900 MHz NMR spectrometer equipped with a cold probe at the University of Georgia. Except HN(CO)CACB which was acquired as Constant Time (CT) type experiment, all other experiments, including HNCACB, HNCA, HN(CO)CA, HNCO, HN(CA)CO, were collected as TROSY-based. Due to the high sensitivity of HNCA experiment which can provide abundant information on C^{α}, to obtain enough resonances of C^{β}, TROSY-HNCACB experiment was recorded with the $^{13}C^{\alpha}$-$^{13}C^{\beta}$ transfer times optimized for maximum sensitivity of $^{13}C^{\beta}$ peaks using delays which were less than $1/(2JC^{\alpha}C^{\beta})$. This led to the appearance of typically weak $^{13}C^{\alpha}$ correlations in these spectra, in addition to strong cross-peaks involving the $^{13}C^{\beta}$. In total, ~2 months of spectrometer time was required to record all of the data sets listed above.

All NMR experiments used for the backbone assignments are listed in Appendix Table A-2.

4.2.2.4 NMR data processing

All NMR spectra were processed and analyzed using the suite of programs provided in NMRPipe (104) and NMRView (152) software. Briefly, the residual water signal was minimized by time domain deconvolution. The ^{15}N time domain of all the spectra was doubled using mirror image linear prediction, before apodization with a cosine-bell window function and Fourier transformation. The ^{13}C time domains of all of the spectra were doubled using mirror-image linear prediction and apodized with cosine squared window functions. Linear prediction in a given dimension was performed only after all of the other spectral dimensions were transformed. The

frequency domain spectra acquired were recalibrated in ^{1}H and ^{15}N dimensions for consistency with the TROSY-based experimental data to account for differences in chemical shifts of the TROSY component (in ppm). The transformed data sets were reduced to include only the regions of interest and analyzed using the NMRView program (152).

4.2.3 Results and Discussion

4.2.3.1 Sequential assignments

All NMR experiments were carried out at 328 K using uniformly [^{2}H, ^{13}C, ^{15}N]-labeled C-terminal domain of Stt3p. The elevated experiment conducting temperature was found to be necessary to improve spectral resolution by increasing sample tumbling rate and thus reducing the resonance line width. Protein stability at elevated temperatures was verified by Circular Dichroism (CD) melting point measurement and NMR spectroscopy. It shows the melting point of the C-terminal domain of Stt3p in SDS micelles is above 348 K (data not shown). Protein sample has been shown to be stable at 328 K for at least one month. It is noteworthy that 100% perdeuteration was essential for assigning most resonances (Figure 4.2). It is also found that TROSY-based experiments offer significantly improvements in both resolution and sensitivity in these ^{1}HN-^{15}N correlation based experiments.

The high content of α-helical secondary structure in this protein, combined with the relatively large number of cross peaks (263 expected non-proline residues), results in severe overlap in the central part of the HSQC spectra, which makes the NMR assignment of this protein extremely challenging.

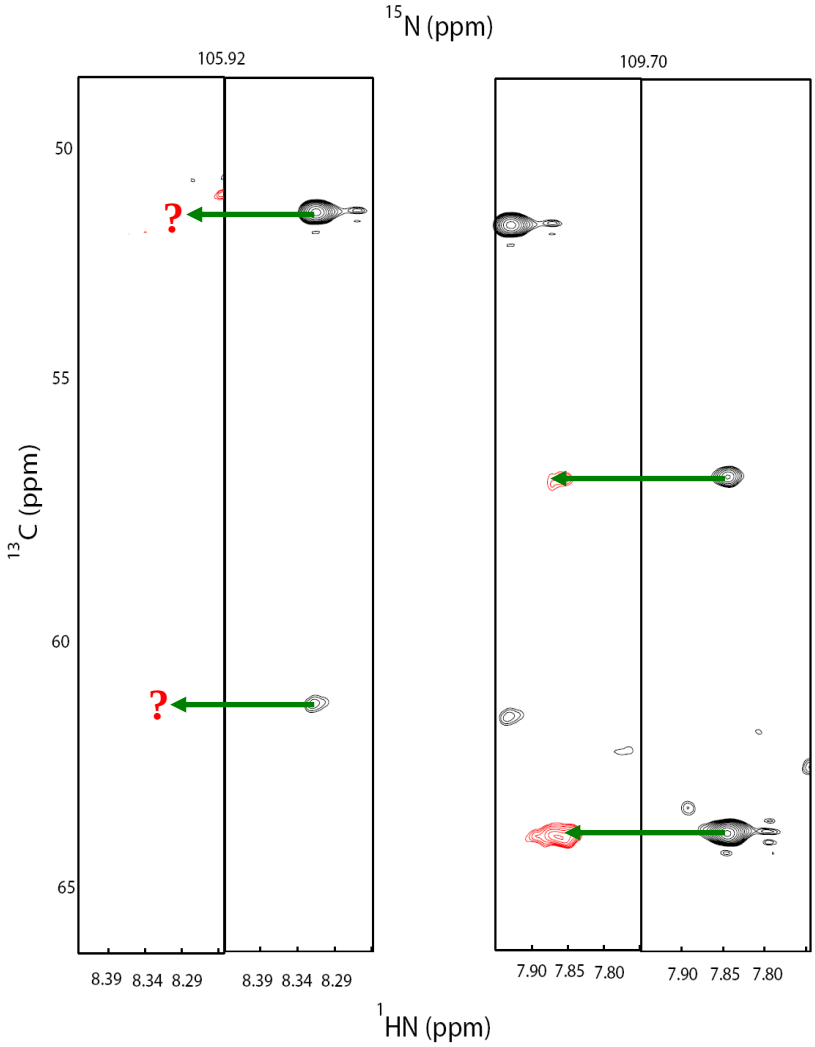
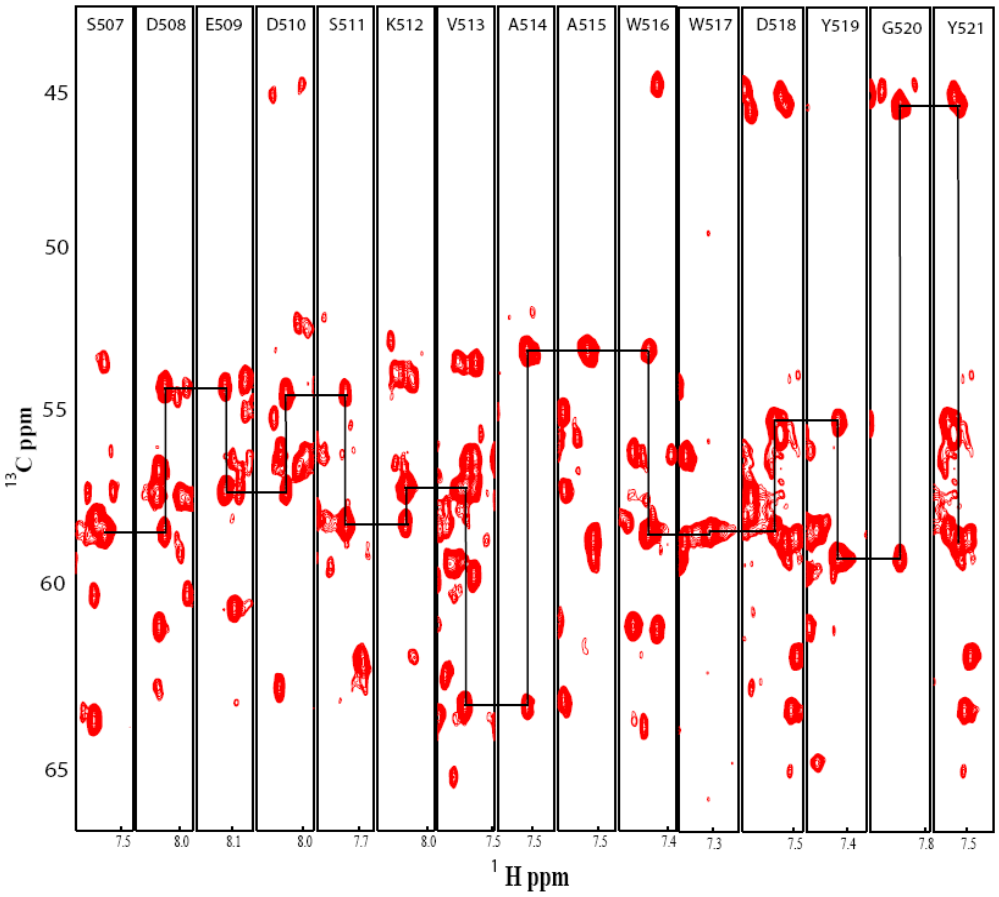


Figure 4.2 Comparison of [^{1}H, ^{13}C]-strips from 3D HNCA spectra using [^{15}N, ^{13}C]-double labeled and [^{2}H, ^{15}N, ^{13}C]-triple labeled samples. The cross peaks on the spectra of ^{15}N, ^{13}C-labeled sample are in red, while the corresponding ^{2}H, ^{15}N, ^{13}C-labeled sample resonances are in black. As shown, corresponding peaks using double-labeled sample either disappear or have very weak intensities comparing to the spectrum acquired using triple-labeled sample.

Sequential NMR spin system connectivities were established using {^{15}N-^{1}H}-TROSY-HNCACB and {^{15}N-^{1}H}-TROSY-HNCA, which provided intraresidual and sequential cross-peaks of C^{β} and C^{α} respectively (Figure 4.3 A). Ambiguities were resolved by {^{15}N-^{1}H}-TROSY-HN(CO)CA and CT-HN(CO)CACB (Figure 4.3 B), which provides only sequential cross-peaks (only from residues i-1). All the assignments were also confirmed by another complementary pair experiments of {^{15}N-^{1}H}-TROSY-HNCO and [^{15}N-^{1}H]-TROSY-HN(CA)CO (Figure 4.3 C).

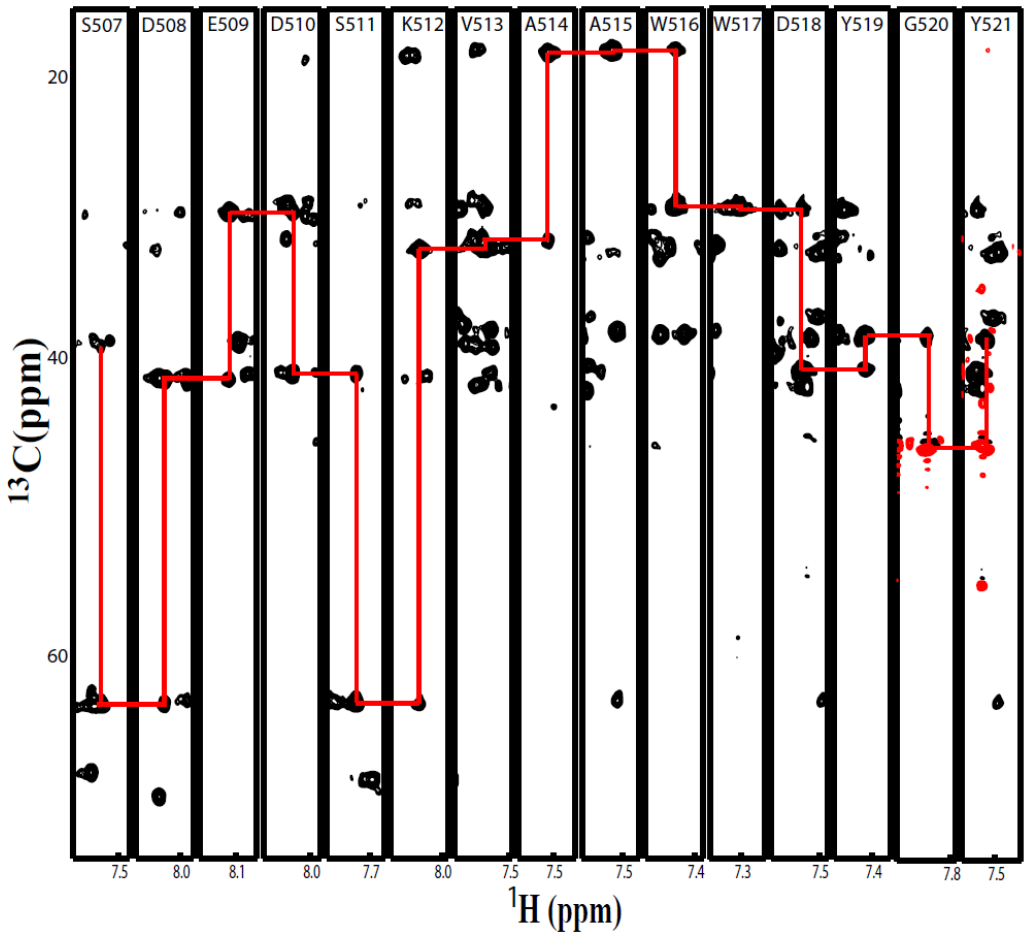
Briefly, the sequence specific assignments started from the amino acid residues which have characteristic chemical shifts for C^{α} or C^{β} (for example, for residues of Ala, Gly, Ser or Thr, as mentioned in Chapter One). Residues were connected sequentially until another residue of an unambiguous type (one of Ala, Gly, Ser, or Thr) was reached. The connected stretch, for example, Ala-$(X)_n$-Ala (X indicates any amino acid), was then positioned in the C-terminal domain of Stt3p primary structure taking into account residue-type information of all the intervening residues X from 3D HN(CA)CB and HN(CO)CACB spectra. During the process of assignment, the constantly updated table of Statistics Calculated for All Chemical Shifts from Atoms in the 20 Common Amino Acids (Biological Magnetic Resonance Data Bank, BMRB, http://www.bmrb.wisc.edu/) was frequently used as a reference to rule out some ambiguities or to confirm the amino acid type.

It is noteworthy that the use of CT-HN(CO)CACB proved to be extremely useful not only in the improvement of resolution in the carbon dimension, but also, and more importantly, in providing the phase information, which can be used in the



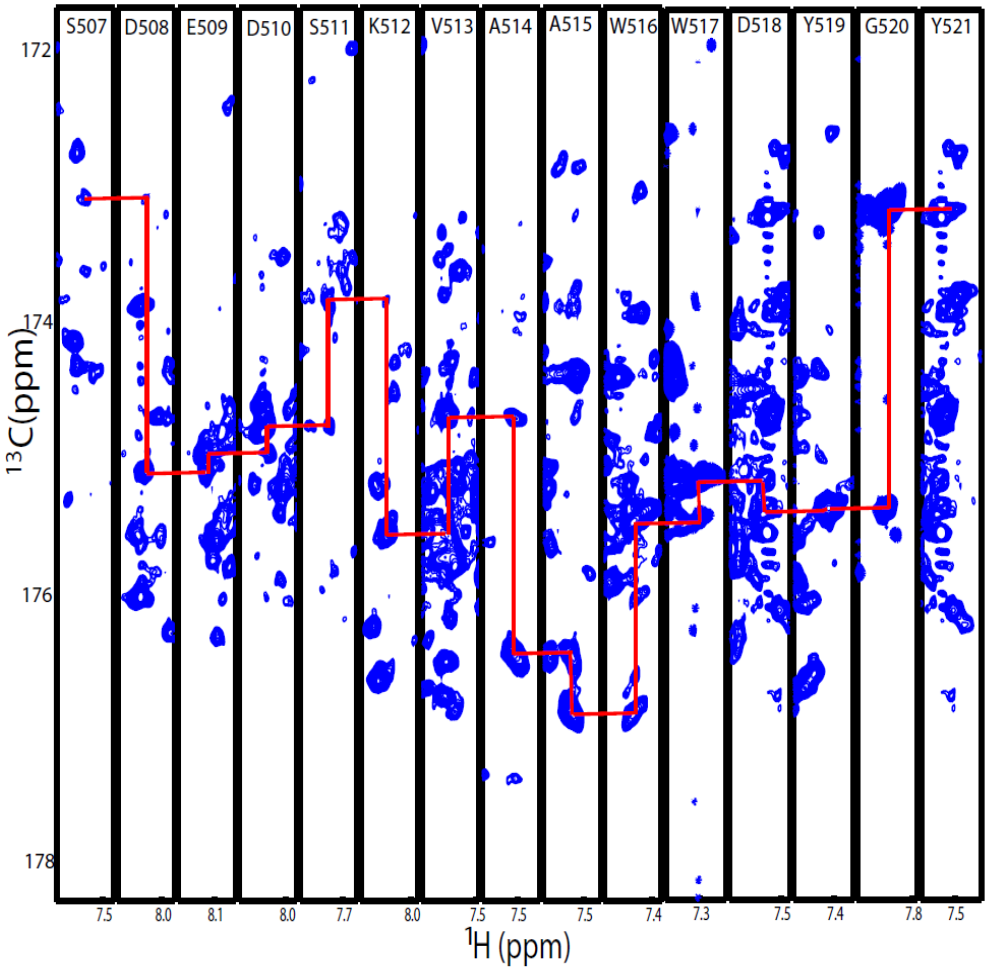
(A)

Figure 4.3 A: [^{1}H, ^{13}C]-strips from HNCA experiment showing sequential assignment for residues S507-Y521.



(B)

Figure 4.3 B: [^{1}H, ^{13}C]-strips from HNCACB experiment showing sequential assignment for residues S507-Y521.



(C)

Figure 4.3 C: [^{1}H, ^{13}C]-strips from HN(CA)CO experiments showing sequential assignment for residues S507-Y521.

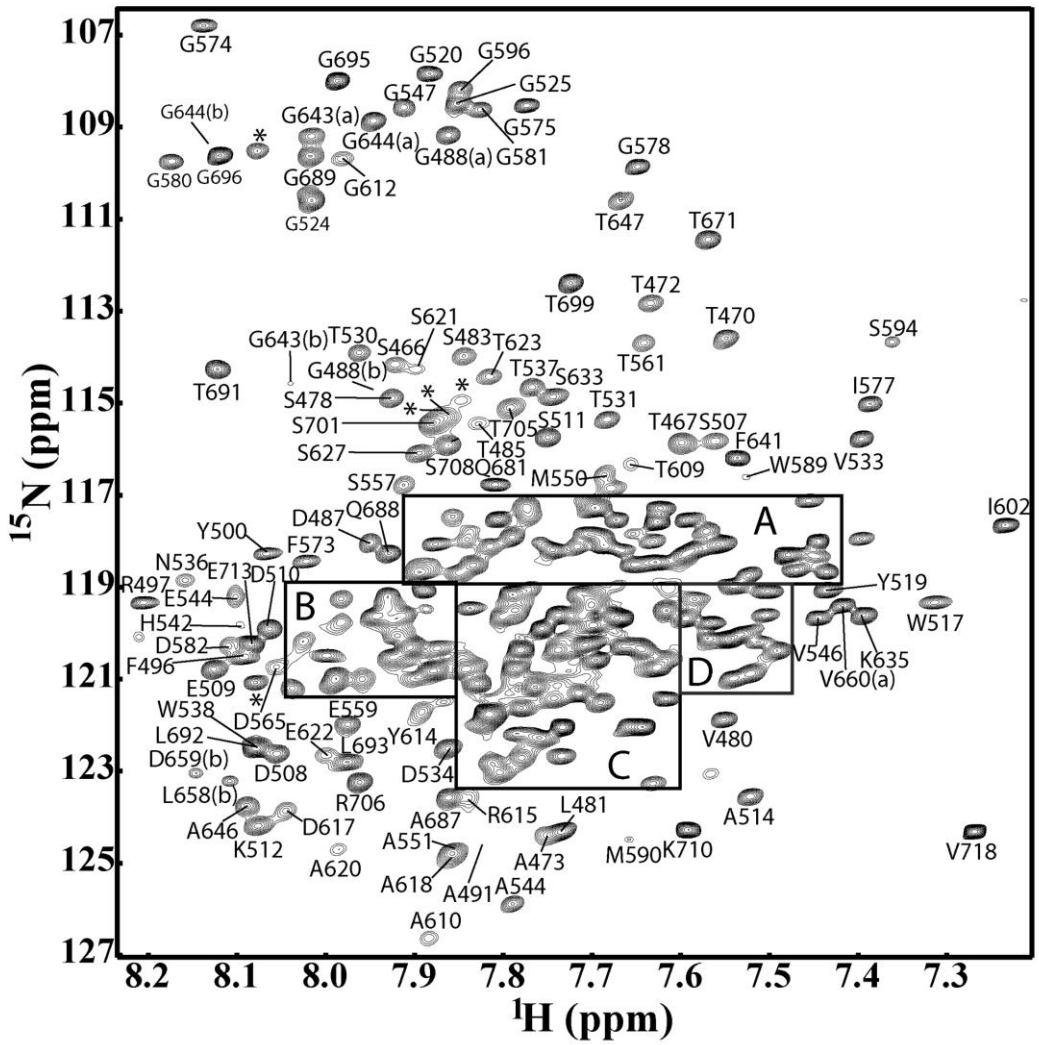


Figure 4.4 Continued on following page.

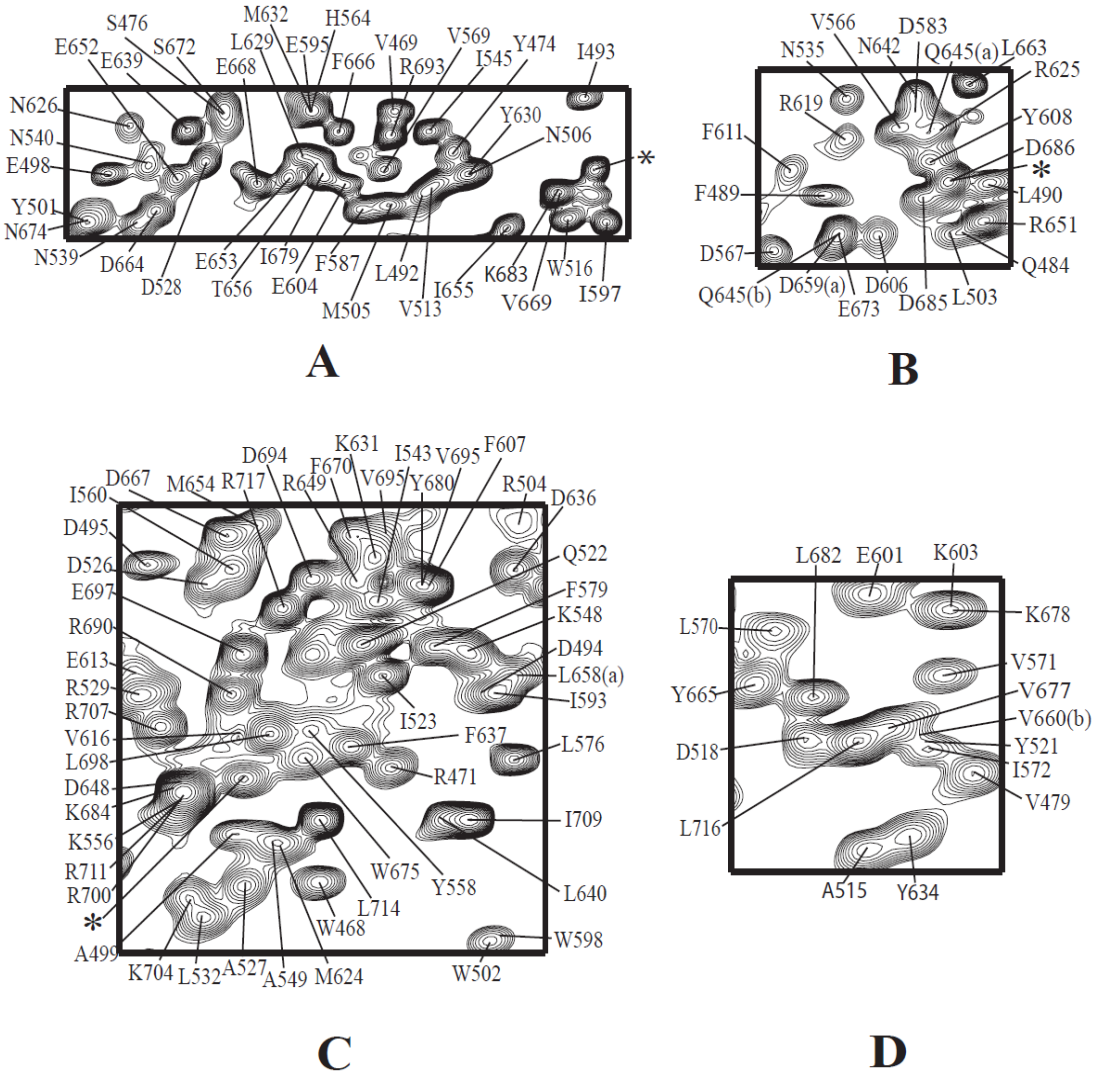


Figure 4.4 ^{15}N-^{1}HN TROSY-HSQC spectrum of U-{^{15}N, ^{13}C, ^{2}H}-labeled C-terminal domain of Stt3p. Four regions of the spectrum are enlarged and peaks are labeled with residue numbers. Due to isoaspartyl, isoasparaginyl and proline cis/trans isomerisational linkage, there are two sets of assignment for a few residues, which were labeled with (a) and (b) respectively. The His-tagged residues are labeled with the symbol of *.

assignment process. As reported, by using CT-HN(CO)CACB, residues with an odd number of aliphatic carbons attached to C^{β} will give rise to opposite sign of C^{β} peaks to that of those residues with C^{β} coupled to an even number of aliphatic carbons (153). Based on the fact the number of amino acids containing odd sets of aliphatic carbons attached to the C^{β} is approximately same as that of those amino acids with even sets of aliphatic carbons attached to the C^{β}, the sign of the cross peaks of C^{β} significantly facilitates resolution of ambiguities in assignments. After a few months of effort, 93% (255 of 274 residues) of backbone resonance assignments were completed (Figure 4.4). The assigned backbone chemical shifts have been deposited in the BioMagResBank (BMRB accession number 16701).

As shown in Table A-1 (Appendix Table), there are seventeen residues which are not assigned. Among the 17 unassignable residues, 11 residues are located on the N-terminal His-tag: M1, G2, S3, S4, H5, H6, H7, H8, H9, S19 and H20. The rest of unassignable residues are L561, K562, I584, N585, I591 and S702. Presumably, the unassignability of these residues can be attributed to the signal broadening effect resulted from local dynamic exchange, which leads to the appearance of very weak or undetectable NMR signals for these residues.

4.2.3.2 CSI analysis

The deviations of $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ chemical shifts from mean random coil values, which have been corrected for one-, two-, and three-bond deuterium isotope effects (154), were evaluated and the secondary structure of the C-terminal Stt3p was determined on the basis of the chemical shift index, CSI (142, 143). In Figure 4.5, the

parameter, $\Delta C^{\alpha} - \Delta C^{\beta}$, was plotted versus the protein sequence. ΔC^{α} and ΔC^{β} are chemical shift differences obtained by subtracting the C^{α} and C^{β} chemical shift values of the protein to that of the random-coil values respectively. Random coil chemical shifts were taken from the reduced database of protein chemical shifts in the BioMagResBank and corrected for one-, two-, and three-bond ^{2}H isotope effects. $\Delta C^{\alpha}-\Delta C^{\beta}$ is calculated and subsequently smoothed by averaging over three successive residues. $\Delta C^{\alpha}-\Delta C^{\beta}$ is a qualitative indicator of secondary structure in proteins with positive values being associated with the α-helix and negative values correlated with β-strands.

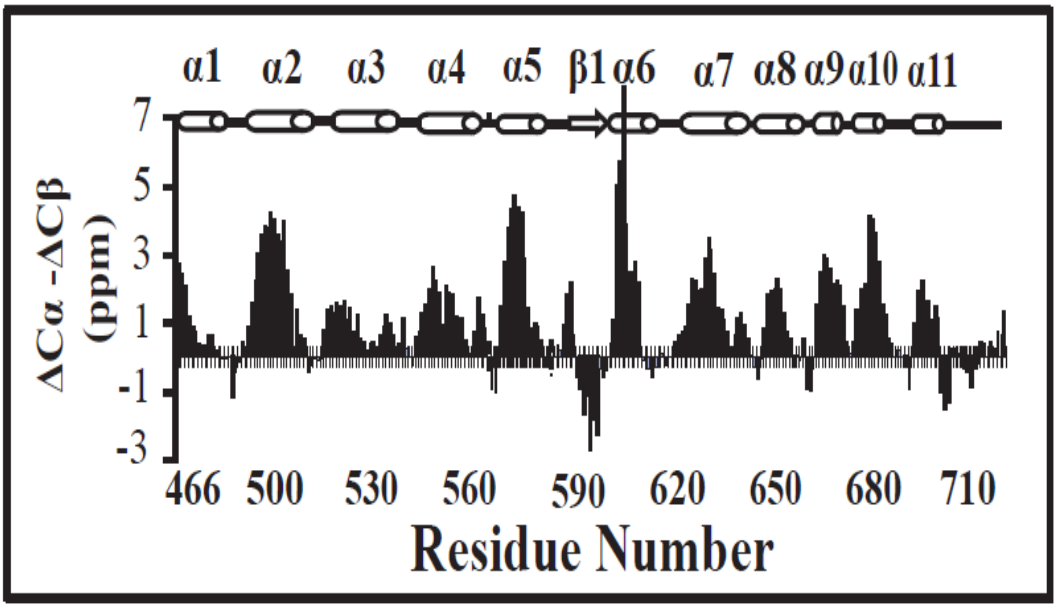


Figure 4.5 CSI analysis of the C-terminal domain of Stt3p.

The stretches of positive values clearly indicate the presence of eleven helical regions α1–α11, which are mostly separated by shorter stretches lacking well-defined secondary structure (most likely loops); while an apparent negative stretch located between α5 and α6, which is indicative of the presence of β sheet. This NMR-based

secondary structure result is consistent with the far-UV CD spectroscopy data, which also indicates that the C-terminal domain of Stt3p is highly helical.

4.2.3.3 Unambiguous identification of isoaspartyl linkage and proline cis/trans isomerizational linkage

Interestingly, during the course of the assignment, an isoaspartyl linkage in the protein sequence IsoAsp^{642}-Gly^{643}, which is an isomerized form of the deamidated Asn642-Gly643 connection (β-linked peptide), was unambiguously identified on the basis of the fact that in the CT-HN(CO)CACB spectrum, the cross-peaks involving C^{α} and C^{β}have opposite signs (Figure 4.6). Extensive studies of asparaginyl deamidation in proteins have shown that this nonenzymatic post-translational modification may play an important role in protein stability and have a significant impact on protein structure and/or function (155, 156).

Another interesting finding is the presence of a proline cis/trans isomerizational linkage in this protein. Proline is unique in the realm of amino acids because it can adopt completely distinct cis and trans conformations. It has been shown that proline cis/trans isomerization plays a key role in protein folding (157) and regulatory mechanism as a molecular timer (see review 158). During the backbone assignment, in fact, we assigned two sets of crosspeaks for residues between Leu^{658} and Val^{660} (Figure 4.7), suggesting that the peptide bond Val^{660}-Pro^{661} adopts both cis and trans conformations. The question as to whether this proline cis/trans isomerization plays an important role can be addressed only by further studies.

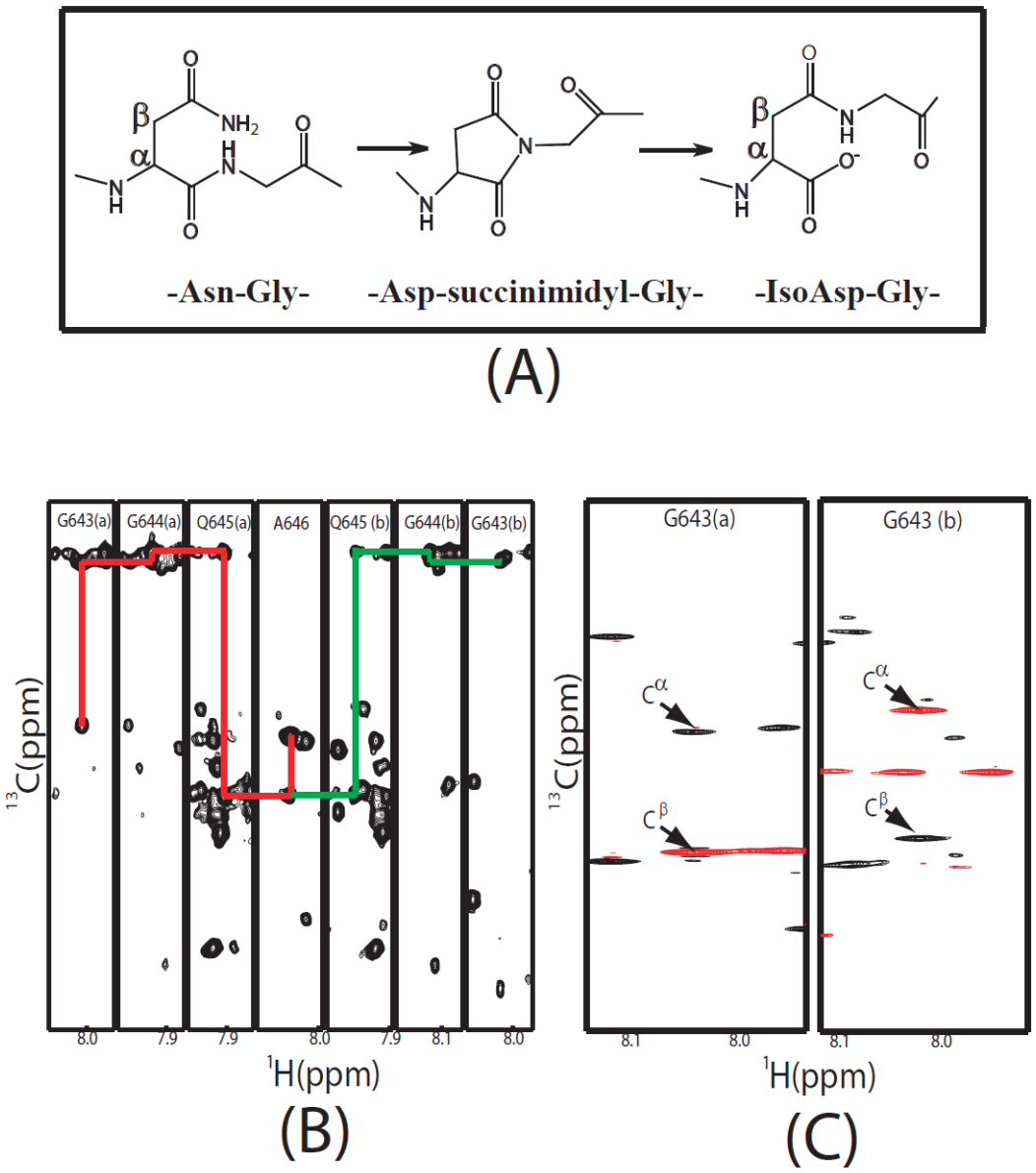


Figure 4.6 Unambiguous Identification of Isoaspartyl Linkage. A: The deamidation of the N642 side chain resulting in the isoasparaginyl linkage between the N642-G643 pair. B shows the co-existence of two sets of HNCA spectrum strip plots of assignments for residues G643-Q645 (shown as red and green line connectivity respectively). These two sets of assignment correspond to the residues following the –Asn-Gly- (residues connected by red lines) and –IsoAsp-Gly- linkages (residues connected by green lines) respectively. C: CT-HN(CO)CACB strip plot of the two assigned G643 residue. Negative peaks are shown in red while positive peaks are shown in black. Note in the left strip, the negative sign of C^{α} and positive sign of C^{β} for residue N642 are unambiguously indicative of –IsoAsp-Gly- linkage between the N642-G643 pair; while the right strip demonstrates a normal –Asn-Gly- linkage (positive sign for C^{α} and negative sign for C^{β}).

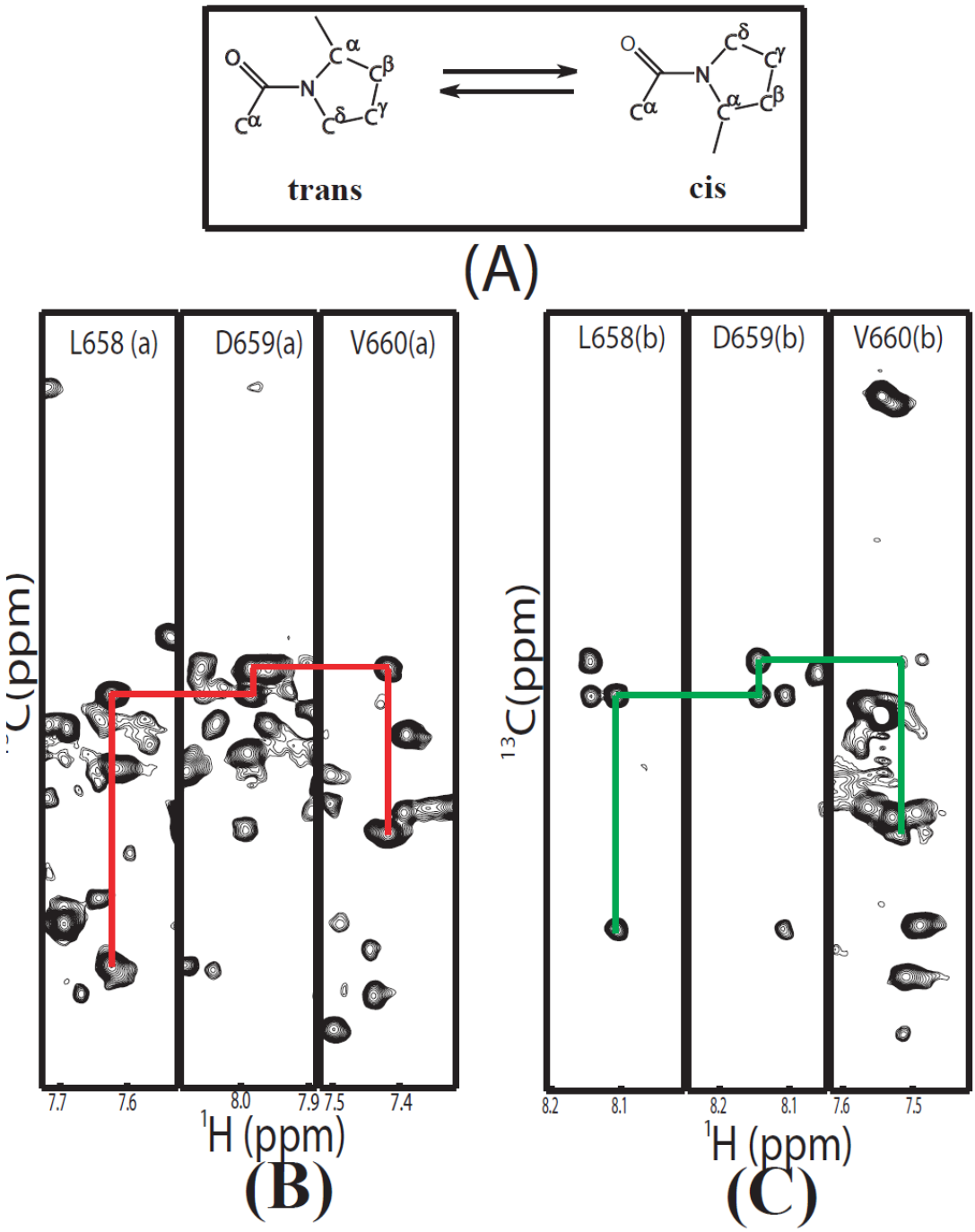


Figure 4.7 Identification of proline cis/trans isomerisational linkage. A: proline cis/trans isomerisation. B and C are two sets of NMR backbone assignment for these three residues, indicating there are two conformations involved: conformations of cis and trans.

In summary, near complete backbone (^{1}HN, ^{15}N, ^{13}CO and $^{13}C^{\alpha}$) and side-chain C^{β} chemical shift was accomplished for the His-tagged C-terminal domain of Stt3p, a 31.5 kDa, 274-residue helical integral membrane protein. The secondary structure was also determined based on the backbone nuclei chemical shifts and the method of Chemical Shift Index (CSI). The completion of the majority of NMR resonance assignment is prerequisite for any further studies by NMR, both for functional and structural studies.

4.3 SIDE-CHAIN ASSIGNMENTS OF THE C-TERMINAL DOMAIN OF STT3P

4.3.1 Introduction

Compared to the protein backbone assignment, the side-chain assignments are rather straightforward: after aligning the backbone sequential assignment with the protein amino acid sequence, side chain amino acid spin systems can be identified from HCCH-TOCSY (for proton resonances), or (H)CC(H)-TOCSY (for carbon resonances), ^{15}N-edited TOCSY, HBHA(CBCA)NH, etc. Each cross-peak in these TOCSY type experiments correlates the chemical shifts of H_i, H_j, … on the side-chain for each residue particular spin system. Magnetization coherence transfer schemes of four common used experiments for side-chain assignments are shown in Figure 4.8. In principle, the HCCH-TOCSY experiment is more sensitive than the ^{15}N-edited TOCSY experiment. This is because rather than relying on the relatively weak $^{3}J_{HH}$ coupling as in the case of ^{15}N-edited experiments, HCCH-TOCSY experiment utilizes the large $^{1}J_{CC}$ scalar coupling to transfer magnetization along the side-chain. However,

in the case of proteins of large size, such as the C-terminal domain of Stt3p, HCCH-TOCSY experiment is of limited utility due to their severely overlapped ^{13}C-edited HSQC spectrum.

In this section, the side-chain assignments of the C-terminal domain of Stt3p are described.

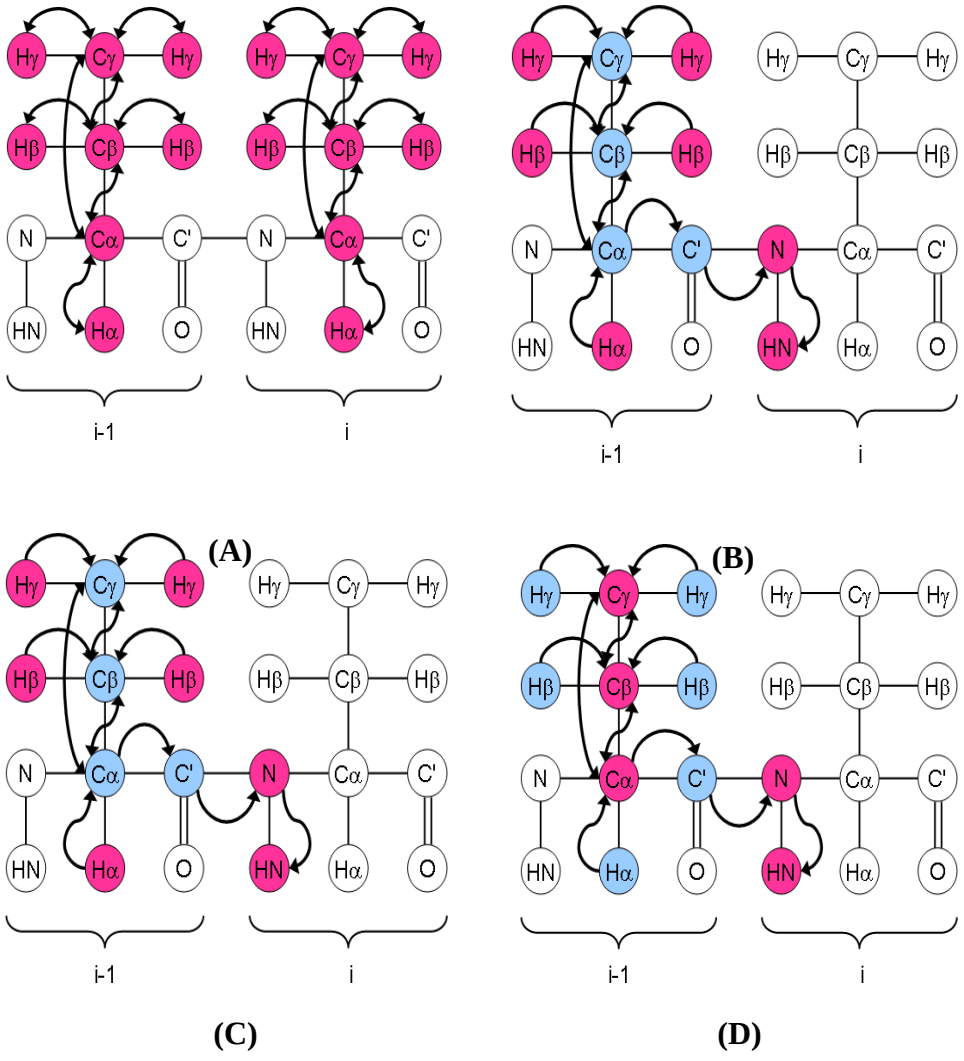


Figure 4.8 Magnetization coherence transfer schemes of some commonly used 3D NMR experiment for protein side-chain experiments. A: HCCH-TOCSY, B: ^{15}N-HSQC-TOCSY, C: H(CCCO)NH, and D: (H)CC(CO)NH.

4.3.2 Methods and Materials

4.3.2.1 NMR sample preparation

Both the uniformly {^{13}C, ^{15}N}-double labeled and {^{2}H, ^{13}C, ^{15}N}-triple labeled C-terminal domain of Stt3p were expressed and purified as described previously. The final NMR samples contain 20 mM sodium phosphate buffer (pH 6.5), 1% (v/v) glycerol, 100 mM sodium dodecyl-d_{25} sulfate (SDS, Aldrich), 1mM EDTA and 10% D_2O. Protein concentration was ~ 0.6 mM.

4.3.2.2 NMR spectroscopy

{^{13}C, ^{15}N}-double labeled and partially deuterated (50%) {^{2}H, ^{13}C, ^{15}N}-triple labeled protein samples of the C-terminal domain of Stt3p were prepared as the NMR samples for side-chain assignments. {^{13}C, ^{15}N}-double labeled protein was expressed by following the same protocols as the ^{15}N-labeled protein expression, except using the ^{13}C-glucose as the only carbon source. Partially triple-labeled protein expression was carried out by following the same protocol as the perdeuterated triple-labeled protein expression (Figure 4.1), except using the 50% D_2O M9 media.

All NMR experiments were carried out at 328 K. These experiments are listed in the Appendix Table A-2.

4.3.3 Results and Discussion

Two of the most commonly used experiments for the assignment of aliphatic side-chain chemical shifts in protonated, ^{15}N, ^{13}C labeled proteins are the (H)C(CO)NH-TOCSY and H(CCO)NH-TOCSY. In these pulse schemes, magnetization originating on side-chain protons is relayed via a carbon TOCSY step

to the backbone Cα position and finally transferred to the ^{15}N, ^{N}H spins of the subsequent residue (residue i-1). (H)C(CO)NH-TOCSY and H(CCO)NH-TOCSY provide correlations linking either aliphatic carbons and protons with backbone amide group chemical shifts, respectively. The large number of transfer steps involved in the relay of magnetization from side-chain to backbone sites in these experiments limits their utility to proteins or protein complexes with molecular weights on the order of 20 kDa or less. For large proteins, as in the case of the C-terminal domain of Stt3p, in order to overcome the problems resulting from proton cross relaxation and obtain more carbon assignments, the highly deuterated (≥50% deuterated) protein samples have to be prepared. In this case, a modification has to be made to the conventional (H)C(CO)NH-TOCSY pulse program to allow magnetization to originate on (deuterated) aliphatic carbon sites (159).

To acquire ample side-chain assignments of the C-terminal domain of Stt3p, both double labeled and perdeuterated triple labeled protein samples were prepared. This led to some practical difficulties regarding the isotope effect on the resonance chemical shifts. Figure 4.9 shows an example of side-chain assignment of the C-terminal domain of Stt3p by using different experiments. Moreover, due to the large size of the target protein and the limited spectral dispersion, after an effort of a few months, over 70% percent of the side-chain assignments were accomplished.

Compared to the backbone assignment, the percentage of side-chain assignment for the C-terminal domain of Stt3p is relatively low. In fact, it has been shown that extensive side chain NMR resonance assignments were not possible for large

membrane proteins (160). The completion of backbone and side-chain assignment of makes it possible for the subsequent NOE assignment, which will be discussed in the next section.

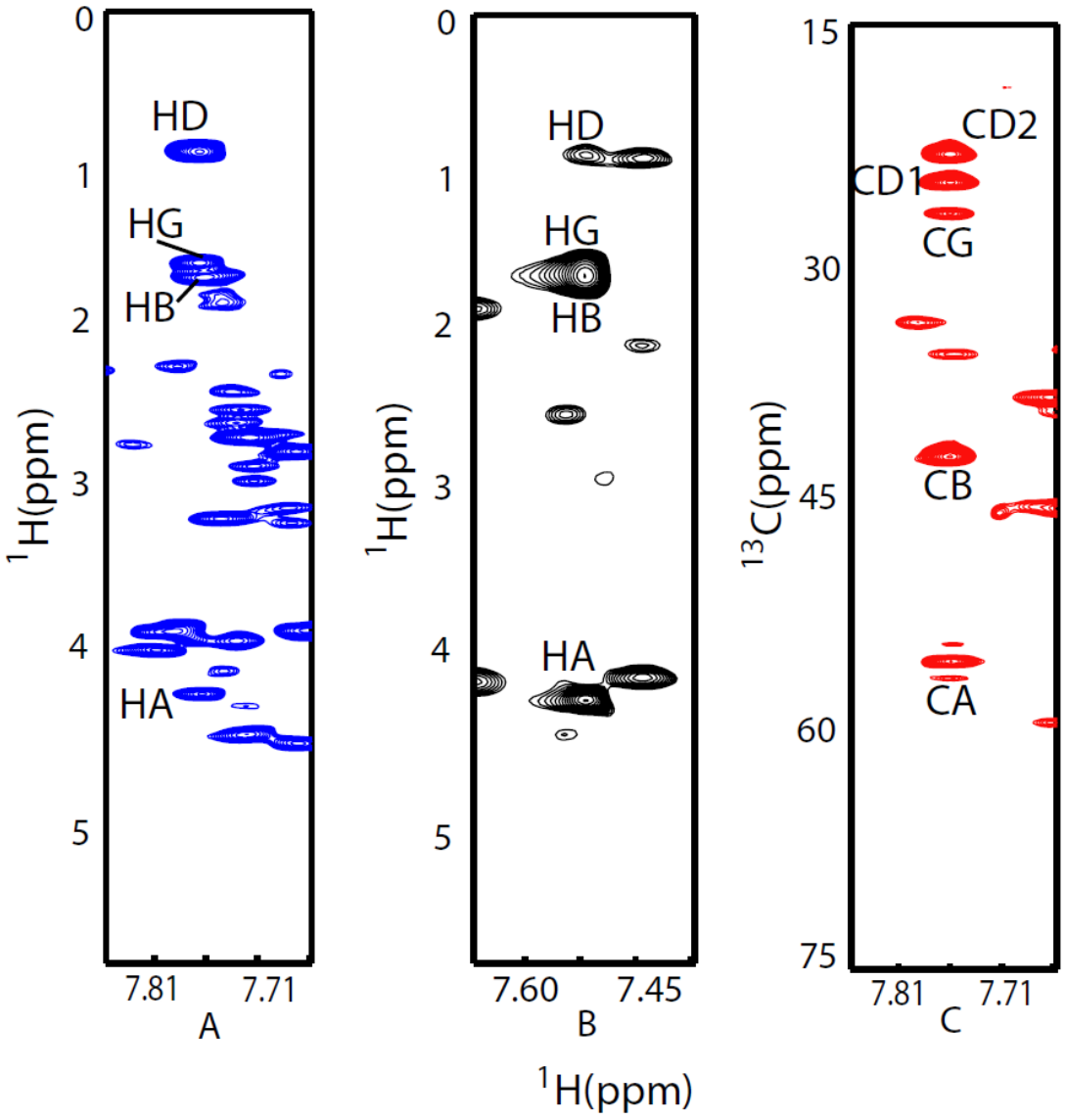


Figure 4.9 Take the residue L714 as an example to show the side-chain assignments of the C-terminal domain of Stt3p. The slices were taken from experiments, from left to right, A: HCCCONH, B: TOCSY-HSQC, and C: (H)CCCONH, respectively.

4.4 NOE Assignments of the C-terminal Domain of Stt3p

4.4.1 Introduction

In *de novo* 3D structure determination of proteins in solution by NMR spectroscopy, the key restraint data are upper distance limits derived from evaluation of proton-proton cross-relaxation (nuclear Overhauser effects or NOE) (7). To extract distance constraints from nuclear Overhauser effect spectroscopy (NOESY) spectra, the cross-peaks have to be assigned, i.e., the pairs of interacting hydrogen atoms have to be identified. In general, the NOESY assignment is based on previously determined chemical shift values which result from protein backbone assignments and side-chain assignments, and at least theoretically, these connectivities among the protons can be readily assigned. However, due to ambiguities caused mainly by the peak overlapping, spectral artifacts, noise, the absence of expected signal caused by fast relaxation or conformational exchange, and inconsistency to some extent of the NOESY cross-peak positions compared to those obtained by resonance assignment due to isotope effects, obtaining a comprehensive set of distance constrains from NOESY spectra is an iterative process. In the initial stage of the assignment, usually, only a fraction of the total NOESY cross-peaks can be assigned unambiguously. In this process, preliminary structures, which are calculated from limited unambiguous NOE assignments, serve to reduce the ambiguity of subsequent NOE assignments. Additional NOESY cross-peaks are then assigned during the iterative steps of the structure calculation.

In this section, the NOE assignments of the C-terminal domain of Stt3p ae discussed.

4.4.2 Methods and Data Collection

In order to obtain enough NOE information, three protein samples with different deuteration levels were prepared: {^{13}C, ^{15}N}- double labeled protein sample, {^{13}C, ^{15}N, ^{2}H} (50%)- partially deuterated triple labeled protein sample, and {^{13}C, ^{15}N, ^{2}H} (100%)- perdeuterated triple labeled protein sample. Protein samples were prepared as described previously and protein concentrations were ~ 0.6 mM.

All NOESY data were recorded at 328 K. For {^{13}C, ^{15}N}- double labeled protein samples, ^{15}N-edited 3D NOESY-HSQC, ^{13}C-edited 3D NOESY-HSQC (both aliphatic and aromatic ^{13}C), and [^{13}C, ^{15}N]- edited 4D HSQC-NOESY-HSQC data were collected; for {^{13}C, ^{15}N, ^{2}H}- perdueterated triple labeled protein sample, ^{15}N-edited 3D NOESY-HSQC data was collected; while for {^{13}C, ^{15}N, ^{2}H} (50%)-partially deuterated triple labeled protein sample, ^{15}N-edited 3D NOESY-HSQC data was collected by a NMR manager at the University of Georgia on a Varian Inova 900 MHz NMR spectrometer equipped with a cold probe. The mixing times for ^{15}N-edited 3D NOESY-HSQC, ^{13}C-edited 3D NOESY-HSQC and [^{15}N, ^{13}C] -edited 4D HSQC-NOESY-HSQC were set as 150 ms, 175 ms and 150 ms respectively. It is shown that no spin-diffusion occurred using these mixing times (data not shown).

All the NOESY experiments are listed in Appendix Table A-2. NMR data collected were subsequently processed by NMRPipe (104) program and analyzed by NMRView (152) software.

4.4.3 Results and Discussion

While other experiments may provide complementary information, the primary observables for structure determination are still cross-relaxation rates measured in a nuclear Overhauser effect spectroscopy (NOESY) experiment. The quality of any given structure is, therefore, heavily influenced by both the total number and the accuracy and precision of the input restraints.

As to the protein sample for NOESY measurement, there are several advantages of using highly deuterated (≥50% deterated) protein sample for NOESY experiment. Firstly, it can improve the accuracy of NOE-derived interproton distance measurements, achieved largely through a reduction of spin diffusion. Secondly, because the linewidths of the remaining protons in a deuterated molecule can be significantly narrowed, overlap is reduced and, in the case of NH-NH cross-peaks, in particular, sensitivity is appreciably increased. Moreover, high deuteration also facilitates the use of longer NOE mixing times, allowing the measurement of larger distances than would be possible in protonated systems.

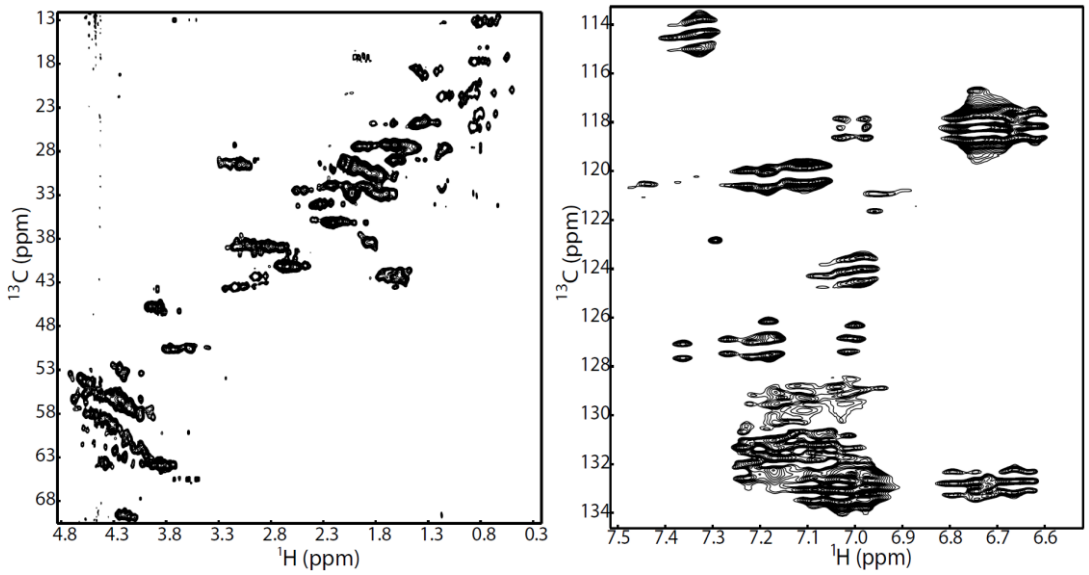
However, like almost anything else, benefit comes along with a cost. Deuteration reduces the concentration of protons that would normally be available for providing NOE-based distance restraints, decreasing the amount of structural information that can be used for analysis. Thus, it is necessary to prepare a series of protein samples with different deuteration levels to find a good balance, between the need for as many distance restraints as possible and the requirement of each cross peak being resolvable.

For this purpose, three protein samples were prepared in this study, and they

were uniformly {^{13}C, ^{15}N}-double labeled with a deuteration level of 0%, 50% and 100% by using M9 media containing 0%, 50% and 100% D_2O respectively. As expected, higher level of deuteration led to more resolvable NOESY spectra, but with less and weaker cross peaks (data not shown). In addition, the introduction of deuterium to the protein samples inevitably resulted in the isotope effect, causing shifts to different extents of the protons peaks depending on the deuteration level. Thus, it is necessary to properly re-reference all the resulted spectra accordingly before any further analysis.

For NOE assignment of the C-terminal domain of Stt3p, a helical membrane protein with a molecular weight of over 31 kDa, the main difficulty results from extensive chemical shift degeneracy, especially for ^{13}C nuclei. Consequently, although ^{13}C-edited NOESY-HSQC generally is more informative than ^{15}N-edited NOESY-HSQC, in the case of the C-terminal domain of Stt3p, its assignments met only very limited success due to severe resonance overlapping (Figures 4.10). Moreover, the relatively weak medium-range and long-range NOE peaks are masked by the strong NOE peaks from intra-residues or neighboring residues, thus making them extremely difficult to identify (Figure 4.11 (A)). Therefore, conventional 3D NOESY data, which are usually sufficient for structure determination of small proteins, are only of limited use once applied to the C-terminal domain of Stt3p. As a result, only very few long-range NOEs were identified using 3D NOESY data (Figure 4.12). In order to overcome this, 4D [^{15}N, ^{13}C]-edited HSQC-NOESY-HSQC spectrum was also collected. This 4D NOESY data proved very helpful in unambiguous identification of

many medium-range NOE peaks, a characteristic for α-helices (Figure 4.11 (B)), together with a few long-range NOEs.



(A) (B)

Figure 4.10 [^{1}H, ^{13}C]-HSQC spectra of the {^{13}C, ^{15}N}-double labeled C-terminal domain of Stt3p. A: Aliphatic region, and B: Aromatic region.

From the above NOESY data, about 2,000 NOE peaks were identified. However, it is sobering (although not surprising) to note that most of the NOE assignments belong to intra-residue, short-range (defined as NOE cross peaks correlating protons on adjacent residues) or medium-range (defined as NOE cross peaks correlating protons on residues separated by up to three residues) NOEs (Figure 4.12).

Although intra-residue, sequential, and medium-range NOEs and dihedral angles constraints (derived from the backbone assignments) enabled the assignment of secondary structure, without enough long-range restraints the global fold of the protein could not be determined. Indeed, the absence of sufficient long-range NOEs

represents one of the biggest challenges in structure determination of helical

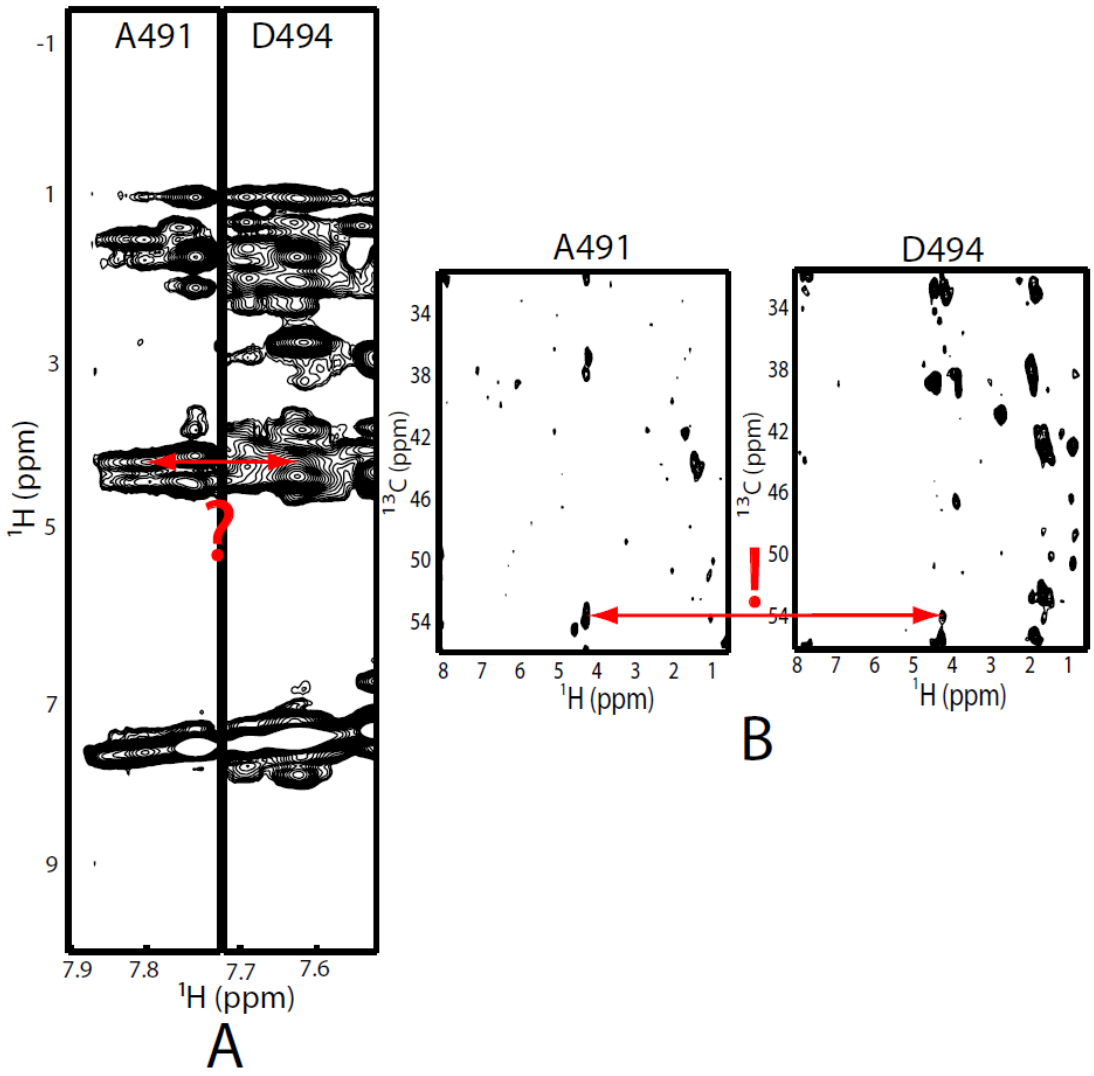


Figure 4.11 Using 4D [^{13}C, ^{15}N]-edited NOESY to identify some NOE peaks. A: by conventional 3D-NOESY, due to the severe spectral overlapping, it was impossible to find the NOE between Cα (i) to Cα (i+3), which is typical for α-helix. B: by the 4D NOESY, the NOE between Cα (i) to Cα (i+3) can be easily identified.

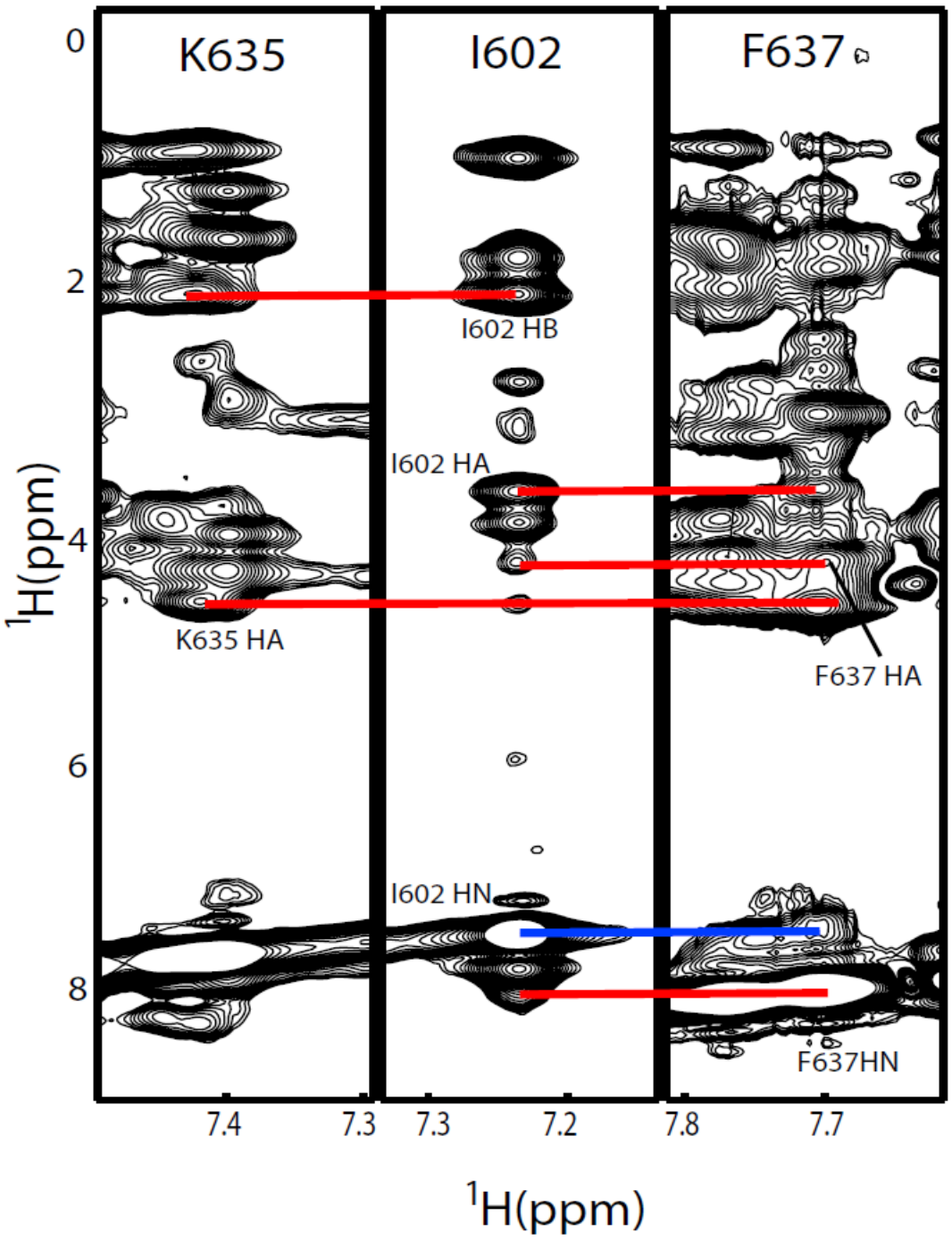


Figure 4.12 Strips from a 3D [^{1}H, ^{1}H]-NOESY-^{15}N-HSQC defining the closure between α6 (residue I602) and α7 (residues K635 and F637). Red and blue lines show the contacts for the depicted residues.

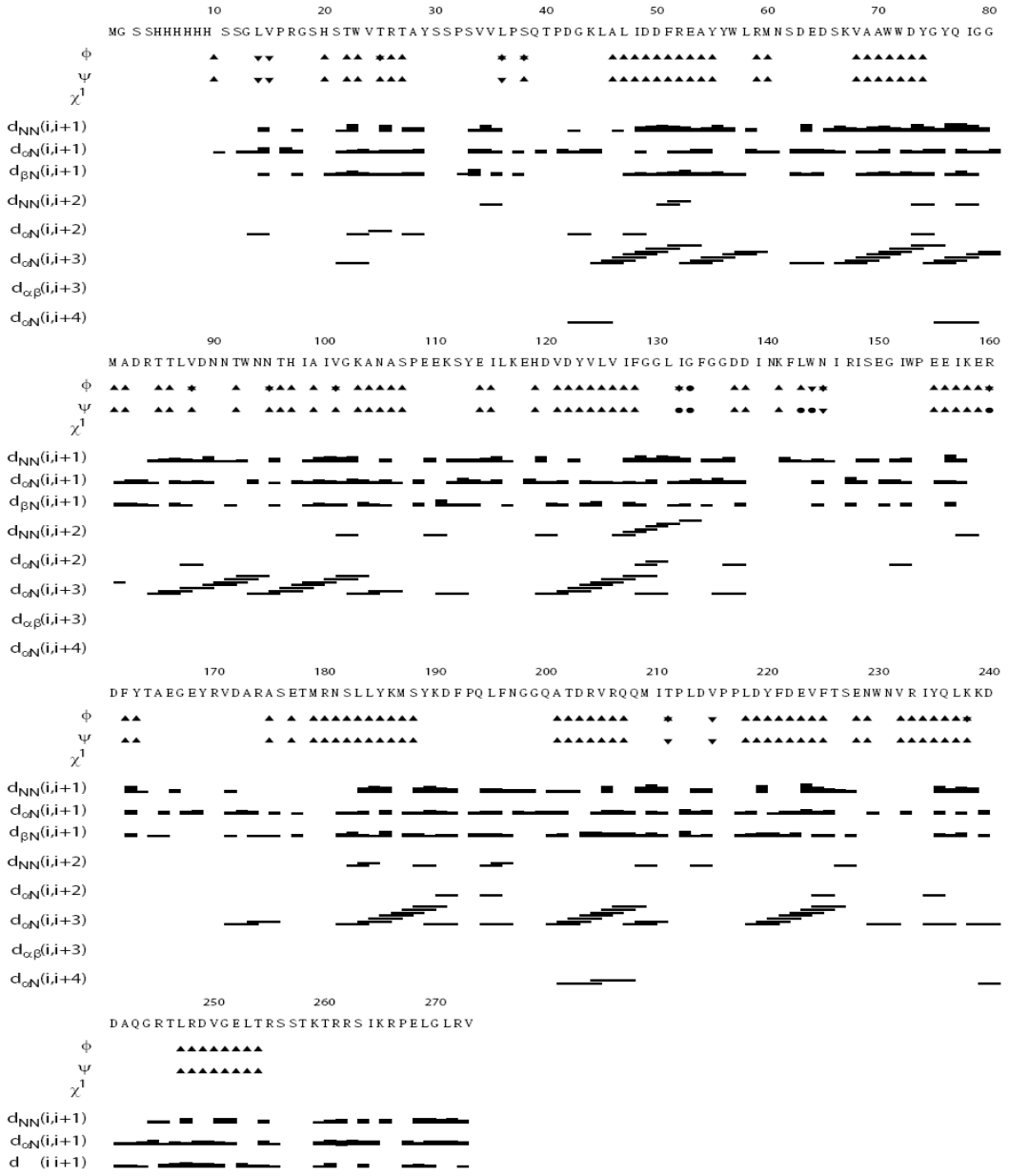


Figure 4.13 Summary of NOE assignments for the C-terminal domain of Stt3p.

It can be seen that most NOE assignments belong to medium range or short range, which are not enough to fold the protein.

membrane proteins by NMR. Therefore, it is clear that, besides the conventional approach, more restraints from other sources, such as PRE, RDC, etc, which are shown later in this thesis, have to be collected and incorporated to final structure calculation.

4.5 ILV-Methyl Protonated Sample Preparation and Assignments

4.5.1 Introduction

The development of isotope labeling methodology has had a significant impact on NMR studies of large proteins and macromolecular complexes. Of particular importance in this regard are methods developed for partial or uniform deuteration of large protein molecules with selective protonation at specific sites. Deuteration of a high molecular weight protein can significantly improve the relaxation properties of the remaining subset of protons that is detected in NMR experiments (161). Recently, an isotope labeling technique called Stereo-Array Isotope Labeling (SAIL) was reported. This techqique utilizes stereospecific protonation/deuteration of protein side-chains to achieve structure determination for the proteins in the 40-50 kDa molecular weight range (2). However, this technique has not been widely used due to the extremely high cost associated with sample preparation or lack of availability of the required isotopically labeled amino acids (162).

Another labeling method is to aim at selective protonation of only the methyl groups of some hydrophobic amino acid residues (Val, Ile and Leu) while leaving the other non-solvent exchangeable protons deuterated. This method takes advantage of the fact that certain α-ketoacids can serve as biosynthetic precursors for the

incorporation of desired isotope labeling pattern into the side-chains of Ile, Leu and Val residues in proteins expressed in minimal media (3). In addition, while remaining a critical number of protons for further structural and dynamics studies, this specific labeling pattern preserves many other important features of perdeuteration with respect to relaxation benefits.

The methyl groups of hydrophobic residues are of particular interest because:

(1) Empirically, for proteins of high molecular weight, structure calculation strategies based on the side-chains nuclei obtained from ^{13}C-NOESY, will fail due to issues involving poor sensitivity and severe resonance overlap. However, since each methyl group contains three magnetically equivalent protons, NMR spectra of methyl groups, even for proteins of high molecular weight, are expected to be of good quality due to their intense correlations and favorable relaxation properties.

(2) Compared to those water-exposed hydrophilic groups which are often located on structure-loose domains (such as loops), methyl groups occur frequently in the hydrophobic cores of proteins due to their hydrophobic nature. Therefore, the distance restraints measured from NOESY experiments between methyl groups can provide valuable information for structure determination.

(3) As mentioned above, relatively cost-effective approaches for the production of Ile (δ1), Leu, Val-methyl protonated, highly deuterated ^{15}N, ^{13}C-labeled proteins by *E. coli* by using deuterated minimal media supplemented with biosynthetic precursors, have already been well developed (128).

In this chapter, the production of selectively labeled protein, NMR spectral data

collection, NMR spectra assignments are described.

4.5.2 Methods and Materials

4.5.2.1 NMR sample preparation

The protocol of expression of methyl protonated {I(δ1 only), L($^{13}CH_3$,$^{12}CD_3$),V($^{13}CH_3$,$^{12}CD_3$)} U-[^{15}N,^{13}C,^{2}H] sample of C-terminal domain of Stt3p was the same as the expression of U-{^{15}N,^{13}C,^{2}H} sample except for the followings:

(1) For 500 mL growth medium, one hour prior to induction, 35 mg of 2-keto-3,3-d_2-1,2,3,4-^{13}C-butyrate sodium salt (Sigma Aldrich) and 60 mg of 2-keto-3-methyl-d_3-3-d_1-1,2,3,4-^{13}C-butyrate sodium salt (Sigma Aldrich) were added to the media;

(2) U-[^{13}C,^{2}H]-glucose (Cambridge Isotope Laboratory, Andover, MA) was used as the main carbon source. This can significantly increase the deuteration efficiency at the position of Cα.

The protein was purified by following the protocol of "SDS elution" as previously described. The final NMR samples contain 20 mM sodium phosphate buffer (pH 6.5), 1% (v/v) glycerol, 100 mM sodium dodecyl-d_{25} sulfate (SDS, Aldrich), 1mM EDTA and 10% D_2O. Protein concentration was ~ 0.6 mM.

4.5.2.2 NMR spectroscopy

Programs for NMR experiments for site-specific methyl assignment were written in our lab, mainly based on the pulse schemes reported by Kay's group (3). The HMCM[CG]CBCA experiment, providing correlations of the form [$\omega_{C\alpha}$(i), ω_{Cm}(i), ω_{Hm}(i)], was recorded with (160, 80, 2048) complex points in the ($^{13}C_\alpha$, $^{13}C_m$, $^{1}H_m$)

dimensions with corresponding acquisition times of (12.6, 13.2, and 142 ms). A relaxation delay of 1.0 s was used along with 16 scans/FID. The Ile,Leu-HMCM(CGCBCA)CO and Val-HMCM(CBCA)CO experiments, which correlates the form [ω_{CO}(i), ω_{Cm}(i), ω_{Hm}(i)], were both acquired with (80, 80, 2048) complex points in the (^{13}CO, $^{13}C_m$, $^{1}H_m$) dimensions with corresponding acquisition times of (29.4, 13.2, 142 ms). A relaxation delay of 1 s along with 32 scans/FID was given for both Ile, Leu-HMCM(CGCBCA)CO and Val-HMCM(CBCA)CO.

For NOESY data collections, both 3D ^{13}C-edited NOESY-HSQC and 4D [^{13}C, ^{13}C]-edited HSQC-NOESY-HSQC were recorded using {^{15}N, ^{13}C}-double labeled protein samples.

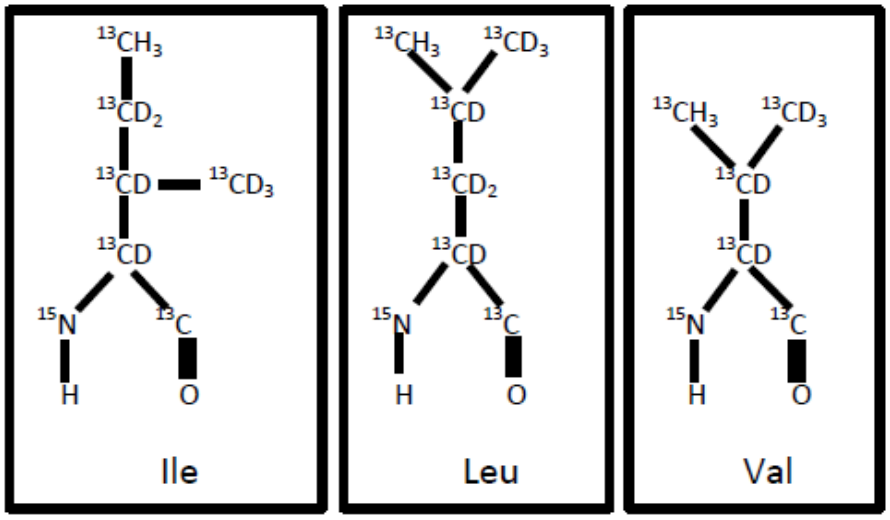
All NMR spectra were processed using the suite of programs provided in the NMRPipe/NMRDraw software package (104). Briefly, the ^{15}N time domains of all of the HN-detected spectra were doubled using mirror image linear prediction, before apodization with a squared cosine window function, and subsequent Fourier transformation. The methyl carbon ($^{13}C_m$) and proton ($^{1}H_m$) indirect-detected time domains of all of the spectra were doubled using mirror-image linear prediction and apodized with cosine window functions. Linear prediction in a given dimension was performed only after all of the other spectral dimensions were transformed (echo processing). The transformed data sets were reduced to include only the regions of interest and analyzed using the NMRView program (152).

4.5.3. Results and Discussion

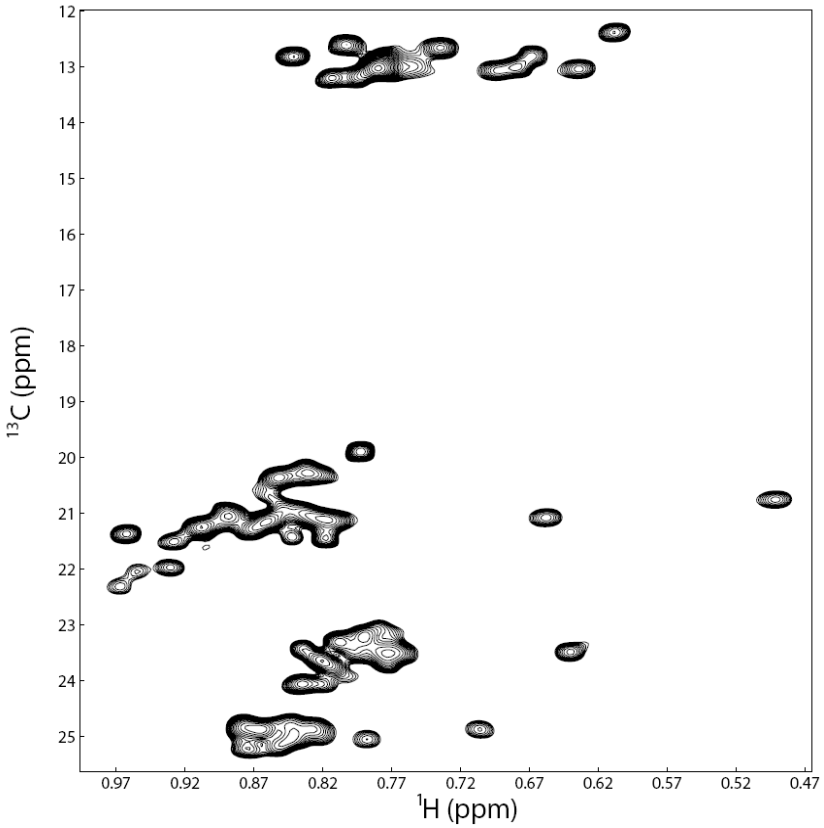
4.5.3.1. Assignments of Ile, Leu, and Val methyl groups of the C-terminal domain of Stt3p

Due to their often severely crowded ^{13}C-HSQC spectra, for proteins of large size, especially integral membrane proteins, conventional ^{13}C-edited NOESY spectra (both of aliphatic and aromatic regions) are only of very limited use. To circumvent this problem, the protein labeling scheme was introduced, which only produce the ^{13}C, ^{15}N-labeled samples with Ile (δ1), Leu, Val-methyl protonated in an otherwise highly deuterated sample. This labeling pattern (Figure 4.14A) provides the NMR spectra with a very "clean background", and the resulting spectral simplicity greatly facilitates the subsequent NOESY assignment.

Figure 4.14B shows the 600 MHz [^{1}H, ^{13}C]-HSQC spectrum of a methyl protonated {I(δ1 only), L($^{13}CH_3$, $^{12}CD_3$), V($^{13}CH_3$, $^{12}CD_3$)} U-[^{15}N, ^{13}C, ^{2}H] sample of the C-terminal domain of Stt3p prepared using the biosynthetic precursors mentioned in Methods and Materials. The 274-residue His-tagged C-terminal domain of Stt3p, as expected for an integral membrane protein, has a high content of Ile, Leu, and Val residues: 15 Ile (only $^{13}C_{\delta 1}$ protonated), 20 Val, and 17 Leu (89 methyl groups in total). Thus, as shown in Figure 4.14B the assignment for this system is quite challenging mainly because the large number of methyl groups are squeezed in NMR spectra of rather narrow dispersions.



A



B

Figure 4.14 Preparation of ILV-protonated sample. A: labeling strategy for Ile, Leu and Val residues. B: [^{1}H, ^{13}C]- HSQC of the ILV-methyl protonated sample of the C-terminal domain of Stt3p.

All the NMR data collected for methyl group assignment can be divided into two categories: (1) HN detected experiments, which correlate the methyl groups with the amide protons; and (2) out-and-back type methyl group detected experiments, which correlate the methyl groups with C_α, C_β or the carbonyl carbons. HN detected experiments include 3D Ile, Leu-(HM)CM(CGCBCA)NH, 3D Val-(HM)CM(CBCA)NH, 3D Ile,Leu-HM(CMCGCBCA)NH and 3D Val-HM(CMCBCA)NH. In principle, this set of experiments by itself is sufficient to make the assignments of methyl groups, since it correlates with the already assigned HN, which usually has an excellent dispersion of chemical shifts. Unfortunately, since it is impossible to transfer polarization from all three methyl protons up the side chain, the inherent low sensitivity of some methyl-HN correlations make it only possible to assign about 30% methyl groups by this type of experiment.

An alternative, more sensitive approach is the out-and-back type experiment, in which the total polarization from the methyl protons is utilized and magnetization both originates from and is transferred back to methyl protons. It has been shown that, compared to the TOCSY-based experiments, out-and-back type experiments can significantly improve the sensitivity and resolution of NMR spectra (3).

Figure 4.15 shows ^{1}Hm-^{13}Cα and ^{1}Hm-^{13}CO strips from HMCM[CG]CBCA and HMCM([CG]CBCA)CO data sets for selected Ile, Leu, and Val residues of the C-terminal domain of Stt3p. Four frequencies (^{13}Cα, ^{13}Cβ, ^{13}Cλ, and ^{13}CO) can be used for identification of methyl groups belonging to the same Leu residue, while ^{13}Cα, ^{13}Cβ, and ^{13}CO shifts are matched to obtain methyl pairs of Val residues. The

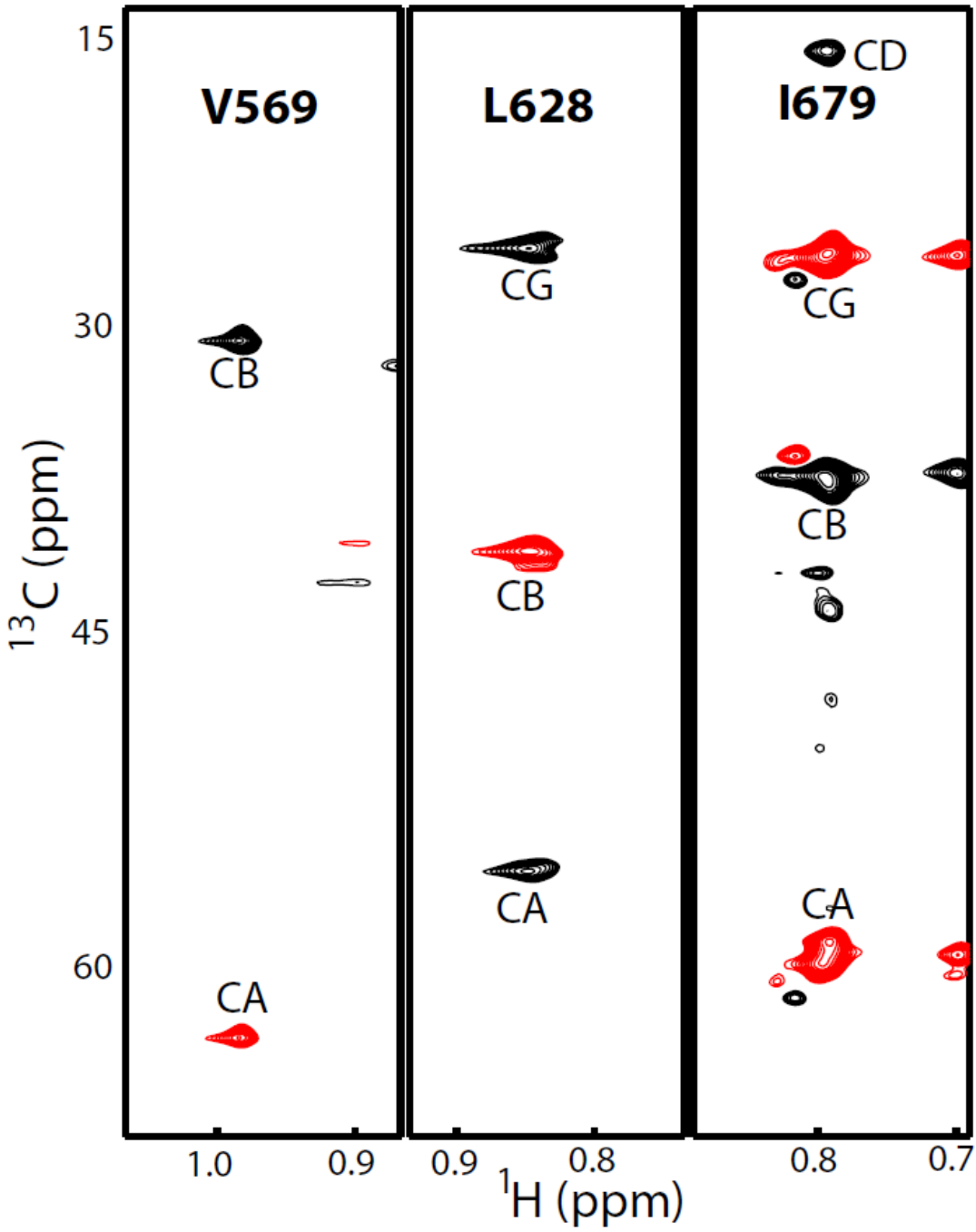


Figure 4.15 Examples of methyl group assignments for some selected residues. Peaks in black indicate they have positive signs, while peaks in red indicate they have negative signs.

sequence-specific assignments of Ile, Leu, and Val methyl groups were made by matching three ^{13}C frequencies - $^{13}C^{\alpha}$, $^{13}C^{\beta}$ (from the HMCM[CG]CBCA data set), and ^{13}CO (from either Ile,Leu-HMCM(CGCBCA)CO or Val-HMCM(CBCA)CO) to those available from earlier backbone and side-chain assignments, taking into account the two-bond deuterium isotope shift. It is of interest to point out that that $^{13}C^{\alpha}$ and $^{13}C^{\beta}$ cross-peaks of Val have opposite signs from those of Ile and Leu in the HMCM[CG]CBCA data set (Figure 4.15), making it straightforward to distinguish the different amino acid types.

Using all these experiments, as shown in Figure 4.16, the methyl groups assigned were 11 Ile δ1 methyl groups (11 out of 15, 73 %), 31 methyl groups of Leu (31 out of 40, 78 %), and 22 methyl groups of Val, (22 out of 34, 65%). The assignments open up the possibility of obtaining more long-range NOE information between methyl groups, which will be incorporated for the final structure determination. The missing assignments are primarily due to the relatively low sensitivity of the HMCM[CG]CBCA experiment. Another reason for incomplete methyl assignments is the extensive chemical shift degeneracy in α-helical membrane proteins. For example, all twenty leucine residues, except Leu^{640}, have ^{1}H chemical shifts within 0.15 ppm of one another (see Figure 4.14).

4.5.3.2. Assignment of NOE between methyl groups of Ile, Leu and Val

3D ^{13}C-edited HSQC-NOESY, 4D [^{13}C, ^{13}C]-edited and 4D [^{13}C, ^{15}N]-edited HSQC-NOESY-HSQC data were collected to obtain inter-residue methyl NOE correlations. Unfortunately, the limited spectral dispersion and spectral overlap, as

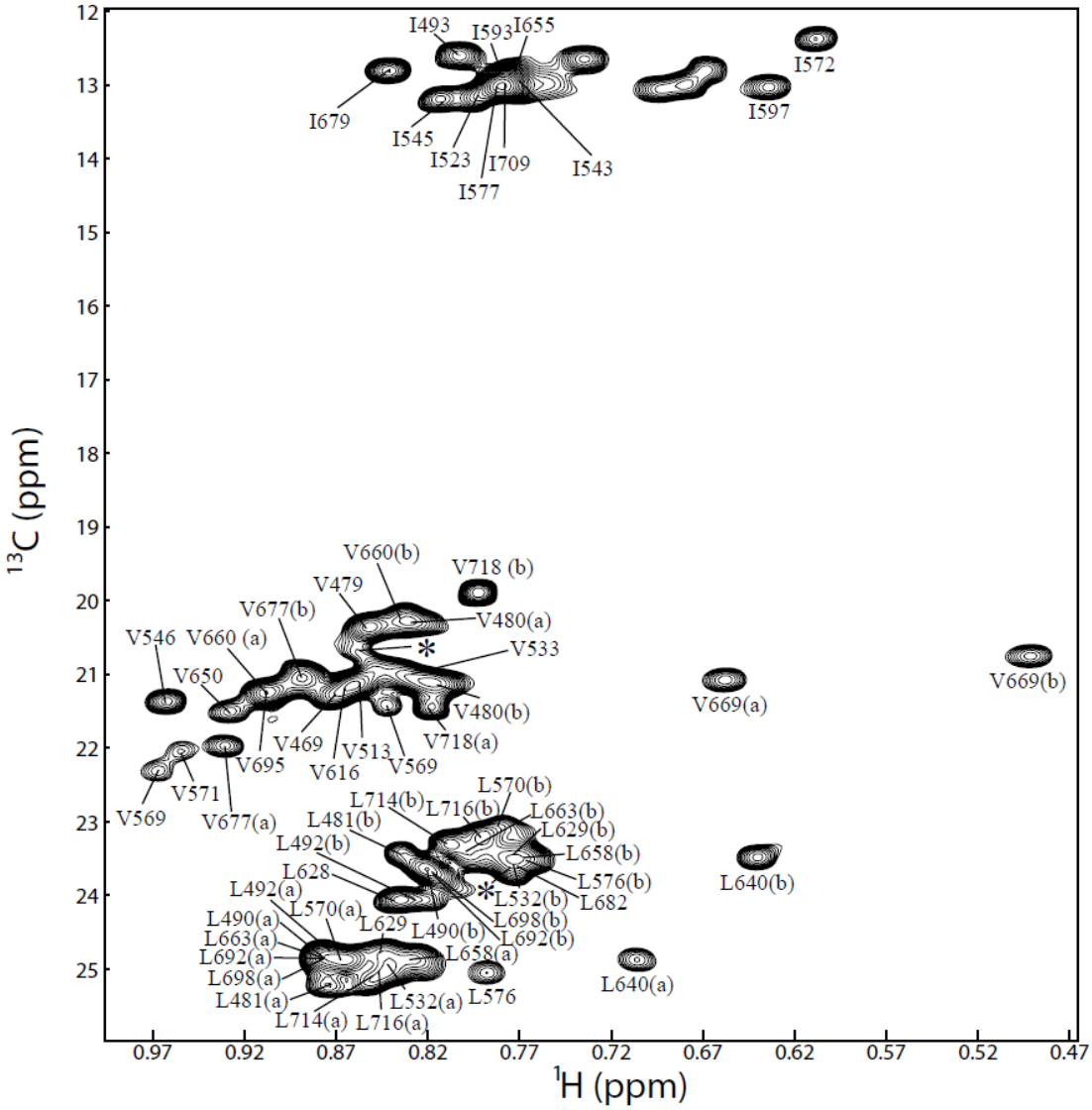


Figure 4.16 Methyl group assignments of the ILV-methyl protonated sample of the C-terminal domain of Stt3p.

well as the sensitivity limitation, precludes a detailed analysis. As a result, we only identified 16 unambiguous long-range NOEs (V669-V695, V669-I545, V669-V569, V669-V469, L640-V677, L640-L576, V677-L576, I545-V479, I493-I560, I543-I572, V480-V569, V469-I493, L481-I493, L490-I545, V480-V669 and I572-I523), together with 32 ambiguous long-range NOEs. These constraints will be employed in later structure calculation for the C-terminal domain of Stt3p in Chapter 5.

CHAPTER 5

STRUCTURE DETERMINATION OF THE C-TERMINAL DOMAIN OF STT3P BY NMR

"We choose to go to the moon. We choose to go to the moon in this decade and do the other things, not because they are easy, but because they are hard…" J. F. Kennedy, 1962

Today for many, if not most, NMR applications to proteins, the ultimate goal is to determine their 3D structures. However, NMR is not a microscope or scanner with atomic resolution that would directly produce an image of a protein. Instead, it contains a wealth of indirect structural information, which can only be converted into the "visible structure" by extensive calculations.

Traditionally, the most crucial structural information is a large number of semiquantitative local restraints, the ^{1}H-^{1}H NOE, which provides distance information for pairs of protons separated by less than ~5 Å. Another commonly used conventional structural information is three-bond J couplings, either homonuclear ^{1}H-^{1}H, ^{13}C-^{13}C, or heteronuclear ^{13}C-^{1}H, ^{13}C-^{15}N, or ^{15}N-^{1}H. By empirically parameterized Karplus relationships (163-165), these three-bond scalar coupling constants can be readily interpreted into the intervening dihedral angles. As mentioned

many times previously in this thesis, some lately emerged NMR measurements, such as the RDCs (6) and PRE-derived distance information (20), if incorporated properly, can also play a significant role in structure determination and refinement.

Nowadays, structure calculation of a protein by NMR method is usually performed by computer programs such as XPLOR-NIH (18, 19) or CYANA (17). These programs take different structural restraints in a particular format as input files and automatically calculate the ensemble of 3D structures. Structure calculation is an iterative process, during which many ambiguities and distance with structural violations are removed and corrected, and some new restraints are added, until an ensemble of structures with an acceptable RMSD (root mean square deviation) values are produced.

In this chapter, the final solution structure of the C-terminal domain of Stt3p is presented.

5.1 Incorporation of Distance Constraints from Paramagnetic Relaxation Enhancement (PRE)

As mentioned in Chapter four, although a large number of intraresidue, sequential, and medium-range NOEs have been assigned, without enough long-range restraints, the fold of the protein could not be determined. Hence, complementary methods are needed for obtaining restraints for structure calculations when long-range NOE data are not sufficient.

The utilization of paramagnetic relaxation enhancement (PRE) is one of the complementary methods which can provide long-range distance restraints. In fact, this

effect has long been recognized as a method for providing long-range distance information that can complement conventional NOE restraints, which are limited to distances of up to 5 Å (166). But until the end of last century, the PRE method has not been frequently used due to lack of paramagnetic centers in most proteins. Site-directed spin labeling (SDSL) offers a straightforward approach to introduce paramagnetic nitroxide centers into proteins (167). Thanks to the elegant work reported by Gerhard Wagner' group, the paramagnetic broadening effects can be readily converted into distance restraints from the measured PREs (20). Since that work, PRE method has been gaining increasing popularity particularly for α-helical membrane proteins, since there are not enough long-range NOEs that can be assigned unambiguously.

The distance calculation is to make use of the modified Solomon-Bloembergen equation for transverse relaxation (166, 168):

$$r = \left[\frac{K}{R_2^{sp}}\left(4\tau_c + \frac{3\tau_c}{1+\omega_h^2\tau_c^2}\right)\right]^{1/6} \qquad \text{Eq. 5.1}$$

where r is the distance between the unpaired electron (localized on the nitroxide spin label) and the nuclear spins (the amide protons); K is a constant, 1.23×10^{-32} cm^6 s^{-2}; τ_c, the correlation time for the electron-nuclear interaction; ω_h is the Larmor frequency of the proton nuclear spin; and R_2^{sp} is the transverse relaxation rate enhancement contributed by the paramagnetic spin-label, which can be determined by:

$$\frac{I_{para}}{I_{dia}} = \frac{R_2 \exp(-R_2^{sp} t)}{R_2 + R_2^{sp}} \qquad \text{Eq. 5.2}$$

where I_{para} and I_{dia} are peak heights of resonances in the paramagnetic- and diamagnetic-labeled protein spectra, respectively; R_2 is the transverse relaxation rate of the resonance in the diamagnetic sample; t is the total evolution time in the proton dimension.

In order to obtain valuable long-range distance restraints derived from PRE, a mono-cysteine mutant library of 16 mutants was prepared by mutagenesis protocol. The 16 mutants are S475C, S483C, S507C, W516C, G520C, T531C, A551C, G580C, S594C, G612C, S621C, S627C, T647C, F670C, G689C and S702C. The preparation, labeling of these mono-cysteine mutants, as well as PRE measurements, are shown in this section.

5.1.1 Methods and Materials

5.1.1.1 Mutagenesis and purification of mutant proteins

These 16 mono-cysteine mutants of C-terminal domain of Stt3p site-directed were made by a PCR-based method using pfuTurbo DNA polymerase according to a protocol developed by Stratagene® (QuickChange Site-Directed Mutagenesis Kit). The mutagenic primers for each mutant (custom made by Invitrogen®) are listed as follows, wherein the sites of the mutation are italicized and underlined:

For S575C: forward (5' GGGTAACAAGAACTGCATAC*TGT*TCTCCTTCT GTTGTTTTGCC 3'); reverse (5' GGCAAAACAACAGAAGGAGA*ACA*GTATGC AGTTCTTGTTACCC 3').

For S483C: forward (5' CTCCTTCTGTTGTTTTGCCA***TGT***CAAACCCCAGAT GGTAAATTG 3'); reverse (5' CAATTTACCATCTGGGGTTTG***ACA***TGGCAAA

ACAACAGAAGGAG 3').

For S507C: forward (5' CTATTGGTTAAGAATGAAC***TGT***GATGAGGACAGTA AGGTTGC 3'); reverse (5' GCAACCTTACTGTCCTCATC***ACA***GTTCATTCTTAA CCAATAG 3').

For W516C: forward (5' GTAAGGTTGCAGCG***TGT***TGGGATTACGGTTACC 3'); reverse (5' GGTAACCGTAATCCCA***ACA***CGCTGCAACCTTAC 3').

For G520C: forward (5' GCAGCGTGGTGGGATTAC***TGT***TACCAAATTGGT GGC 3'); reverse (5' GCCACCAATTTGGTA***ACA***GTAATCCCACCACGCTGC 3').

For T531C: forward (5' GTGGCATGGCAGACAGAACC***TGT***TTAGTCGATAA CAACACG 3'); reverse (5' CGTGTTGTTATCGACTAA***ACA***GGTTCTGTCTGCCA TGCCAC 3').

For A551C: forward (5' CATCGTTGGTAAAGCCATG***TGT***TCCCCTGAAGAGA AATC 3'); reverse (5' GATTTCTCTTCAGGGGA***ACA***CATGGCTTTACCAACGA TG 3').

For G580C: forward (5' GGTGGTCTAATTGGGTTT***TGT***GGTGATGACATCAAC 3'); reverse (5' GTTGATGTCATCACC***ACA***AAACCCAATTAGACCACC 3').

For S594C: forward (5' CTTGTGGATGATCAGAATT***TGT***GAGGGAATCTGGC CAGAAG 3'); reverse (5' CTTCTGGCCAGATTCCCTC***ACA***AATTCTGATCATCC ACAAG 3').

For G612C: forward (5' GTGATTTCTATACCGCAGAG***TGT***GAATACAGAGTA GATGCAAGG 3'); reverse (5' CCTTGCATCTACTCTGTATTC***ACA***CTCTGCGG TAT AGAAATCAC 3')

For S621C: forward (5' GAGTAGATGCAAGGGCT***TGT***GAGACCATGAGGAA CTCG 3'); reverse (5' CGAGTTCCTCATGGTCTC***ACA***AGCCCTTGCATCTACTC 3').

For S627C: forward (5' CTTCTGAGACCATGAGGAAC***TGT***CTACTTTACAAG ATGTCCTAC 3'); reverse (5' GTAGGACATCTTGTAAAGTAG***ACA***GTTCCTCA TGGTCTCAGAAG3').

For T647C: forward (5' CAATGGTGGCCAAGCC***TGT***GACAGAGTGCGTCAAC 3'); reverse (5' GTTGACGCACTCTGTCA***CAG***GCTTGGCCACCATTG 3').

For F670C: forward (5' GACTACTTCGACGAAGTT***TGT***ACTTCCGAAAAC TGGATGG 3'); reverse (5' CCATCCAGTTTTCGGAAGT***ACA***AACTTCGTCGAA GTAGTC 3').

For G689C: forward (5' GAAGAAGGATGATGCCCAA***T***GTAGAACTTTGAGG GACG 3'); reverse (5' CGTCCCTCAAAGTTCT***ACA***TTGGGCATCATCCTTCTTC 3').

For S702C: forward (5' GGTGAGTTAACCAGGTCT***TGT***ACGAAAACCAGAAG GTCC 3'); reverse (5' GGACCTTCTGGTTTTCGT***ACA***AGACCTGGTTAACTC ACC 3').

The mutation results were confirmed by DNA sequencing.

The expression of uniformly ^{15}N-labeled mono-cysteine mutants of the C-terminal domain of Stt3p were performed by following the same protocol as described previously. Except for the S475C mutant, which could not be expressed after several attempts, , all the other 15 mono-cysteine mutants were successfully overexpressed

and the protein yields were comparable to that of the wild type protein.

The purifications of all the 15 mono-cysteine mutants were achieved by following our "SDS elution" method, except 100 μM DTT (dithiothreitol, Sigma) was added to all the solutions (denaturing buffer, binding buffer, washing buffer and elution buffer). The presence of DTT is to ensure that the cysteine residues were kept in the reduced state to avoid the formation of inter-molecular disulfide bonds. It should be noted here that the concentration of DTT can not be too high; otherwise it will lead to reduction of Ni^{2+}, which is indicative by color change from light blue to dark yellow or brown.

5.1.1.2. MTSL and dMTSL spin labeling of mutant proteins

Uniformly ^{15}N-labeled C-terminal domain of Stt3p mono-cysteine mutants in SDS micelles were spin-labeled by using the paramagnetic spin label agents, (1-oxyl-2,2,5,5-tetramethyl-3-pyrroline-3-methyl)-ethanethiosulfonate, MTSL, Toronto Research Chemicals, Toronto), and a diamagnetic analogue of MTSL, (1-acetyl-2,2,5,5-tetramethyl-η3-pyrroline-3-methyl) methanethiosulfonate (dMTSL, Toronto Research Chemicals, Toronto), in which the oxygen on the nitroxide of MTSL was replaced with an acetyl group. Briefly, purified reduced uniformly ^{15}N-labled mono-cysteine mutant was split into two equal portions for parallel labeling with MTSL and dMTSL. Both labeling reagents were added from 75 mM stock solution in methanol at a 10-fold molar excess over protein. The solutions were mixed for 3 hours at 37 ℃ and then incubated overnight at room temperature. Excess MTSL was removed by using Amicon ultrafiltration device with a MWCO of 5 kDa. Typically, 500 μL of

phosphate buffer (20 mM, pH 6.5) was added to 350 μL spin-labeled protein sample in Amicon Ultra-15 tube and centrifuged until there was approximately 350 μL solution left. This process was repeated 9 times for complete removal of free spin labels. For the last time, 500 μL of buffer containing 50 mM SDS, 10% D_2O, 1% glycerol, 1 mM EDTA, 20 mM phosphate buffer, pH 6.5 was added to the 350 μL protein solution and centrifuged until 350 μL of solution left. The final samples had protein concentrations ranging from 0.2 to 0.3 mM and the resulting HSQC spectra can be overlapped with that of the unlabeled protein.

5.1.1.3 NMR spectroscopy

NMR measurements of both MTSL and dMTSL-labeled proteins were conducted at 328 K. The data were acquired with 256 and 2048 complex points in the t_1 time domain (^{15}N dimension) and t_2 time domain (^{1}H dimension) respectively. The data were zero-filled to 512×4096 and apodized using a Gaussian window function prior to Fourier transformation using NMRPipe (104). Peak assignments of spin-labeled mutants were based on comparison with the spectra of wild-type protein. For some mutants, the resonances of some residues close to the mutation site had disappeared or the chemical shifts had significantly changed, and such peaks were excluded from any further analysis.

5.1.1.4 PRE-based distance restraints analysis

Paramagnetic perturbation analysis was conducted in strict accordance with the protocols reported by Lukas K. Tamm's research group (169). In brief, 2D [^{15}N, ^{1}H]-HSQC spectra of otherwise identical samples were collected for all fifteen MTSL-

labeled modified samples, as well as dMTSL-labeled C-terminal domain of Stt3p. All spectra were calibrated against the dMTSL-labeled spectra using at least five peaks displaying the least relative signal decrease, to compensate for any global effects from variations in protein concentration or possible small fluctuations of the spectrometer response. Individual peaks were measured by peak intensities and compared to fifteen dMTSL-labeled corresponding reference spectra. Due to spectral crowding roughly 35% of the peaks for any given sample could be definitively assigned and measured.

According to the calculation making use of the modified Solomon-Bloembergen equation for transverse relaxation, residues were assigned into two qualitative categories: (1) protons with I_{para}/I_{dia} ratios between 15% and 85%; (2) protons whose I_{para}/I_{dia} ratios were less than 15%, including protons whose resonances were no longer observable in the paramagnetic spectra; and (3) protons whose I_{para}/I_{dia} ratios were over 85%, where I_{para} and I_{dia} are peak intensities of resonances in the MTSL and dMTSL-labeled protein spectra, respectively. These groupings were translated into three classes of distance restraints: > 15 Å but < 24 Å; < 15 Å; and > 25 Å, respectively (169). Although PRE distances are less precise than NOE distances, the larger number and the longer distance range of PREs compensate for the better precision of the NOEs.

5.1.2 Results and Discussion

In order to minimize possible interference of the spin labels with the structure of the C-terminal domain of Stt3p, the sites were selected for mutation and in such a way that they are not located in the middle of any secondary structures (by comparing

to the CSI result shown in Chapter 4), except W^{516} and G^{520}. W^{516} and G^{520} were selected deliberately to probe the structure of the proposed catalytic site. The results show that, comparing to the HSQC spectrum of wild-type protein, only some resonances of residues around the mutation sites exhibited rather large changes in chemical shifts (they actually seem to have "disappeared" since they shifted to other places), whereas resonances of all other residues exhibited only very small chemical shift changes. Moreover, there are almost no chemical shift differences after further introduction of spin label (either MTSL or dMTSL) to these mono-cysteine mutants. These results are similar to those previous reports (20, 169), in which the authors also concluded that the protein fold was not significantly perturbed by the introduction of spin labels.

Figure 5.1 shows comparisons of parts of the 2D [^{15}N, ^{1}H]-HSQC spectra in the presence of paramagnetic and diamagnetic spin labels. It is clear that the intensities of many residues were affected by PRE, and these intensity perturbations can be used to obtain long-range distance restraints. However, according to North *et al.* (170, 171), PREs to loop residues should be excluded for structure calculation because the observed distances are not the average but the "closest contact" distances due to the flexible nature of the loop residues.

From the fifteen spin-labeled samples, altogether 467 long-range upper-distance and 302 lower distance restraints were obtained. These extremely valuable distance restraints derived from PRE will be employed for later structure calculation.

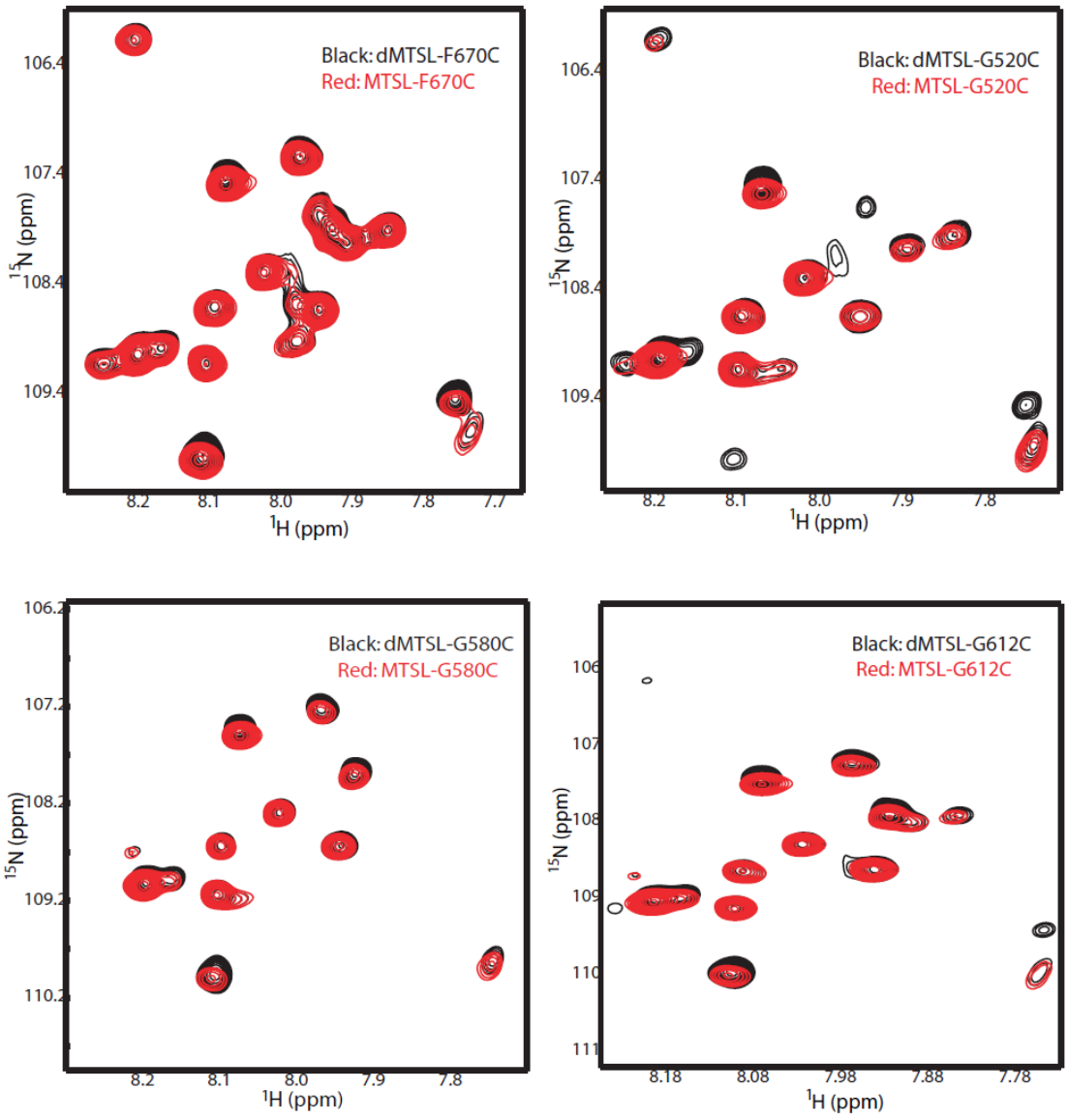


Figure caption is on page 157

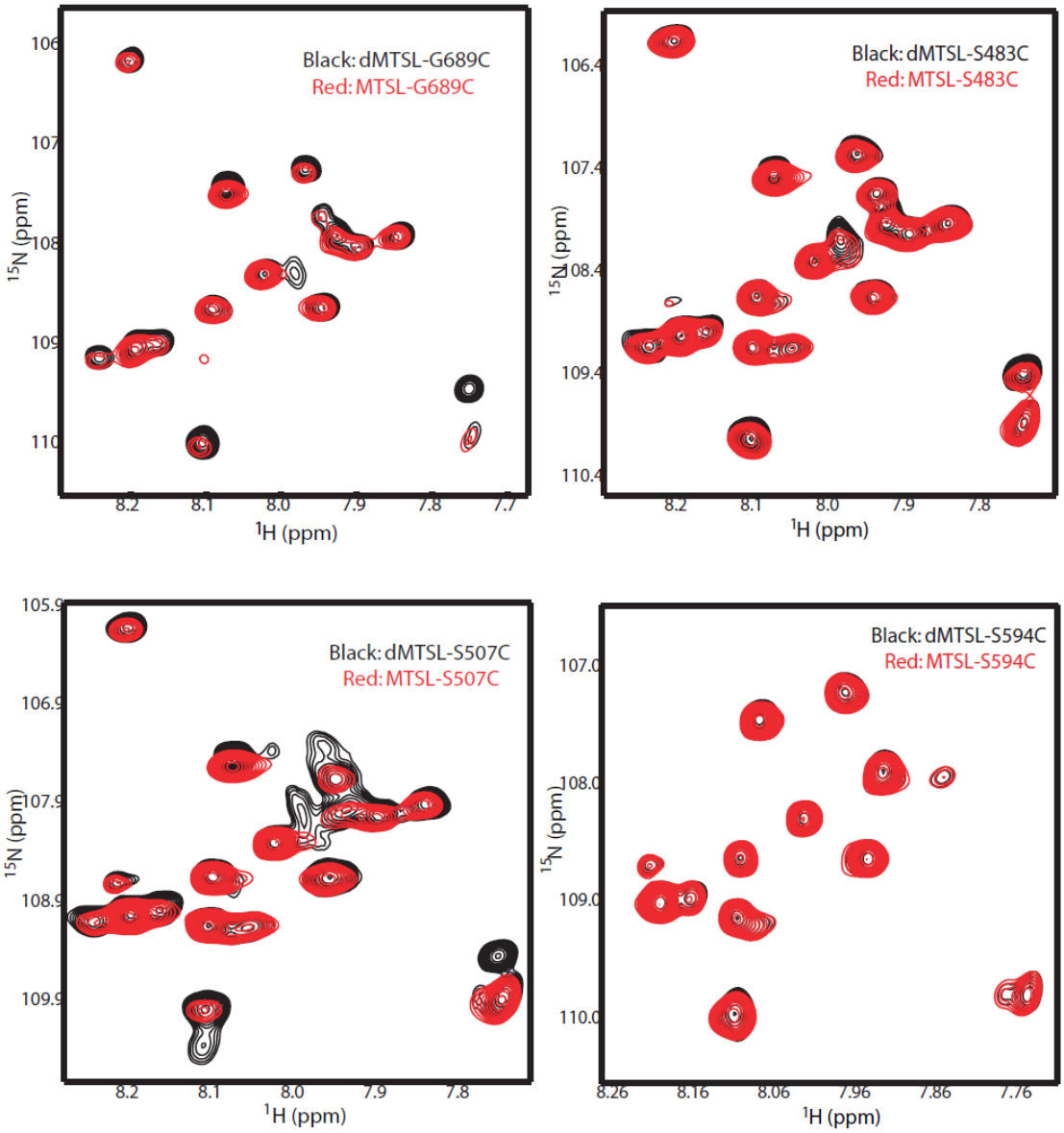


Figure caption is on page 157

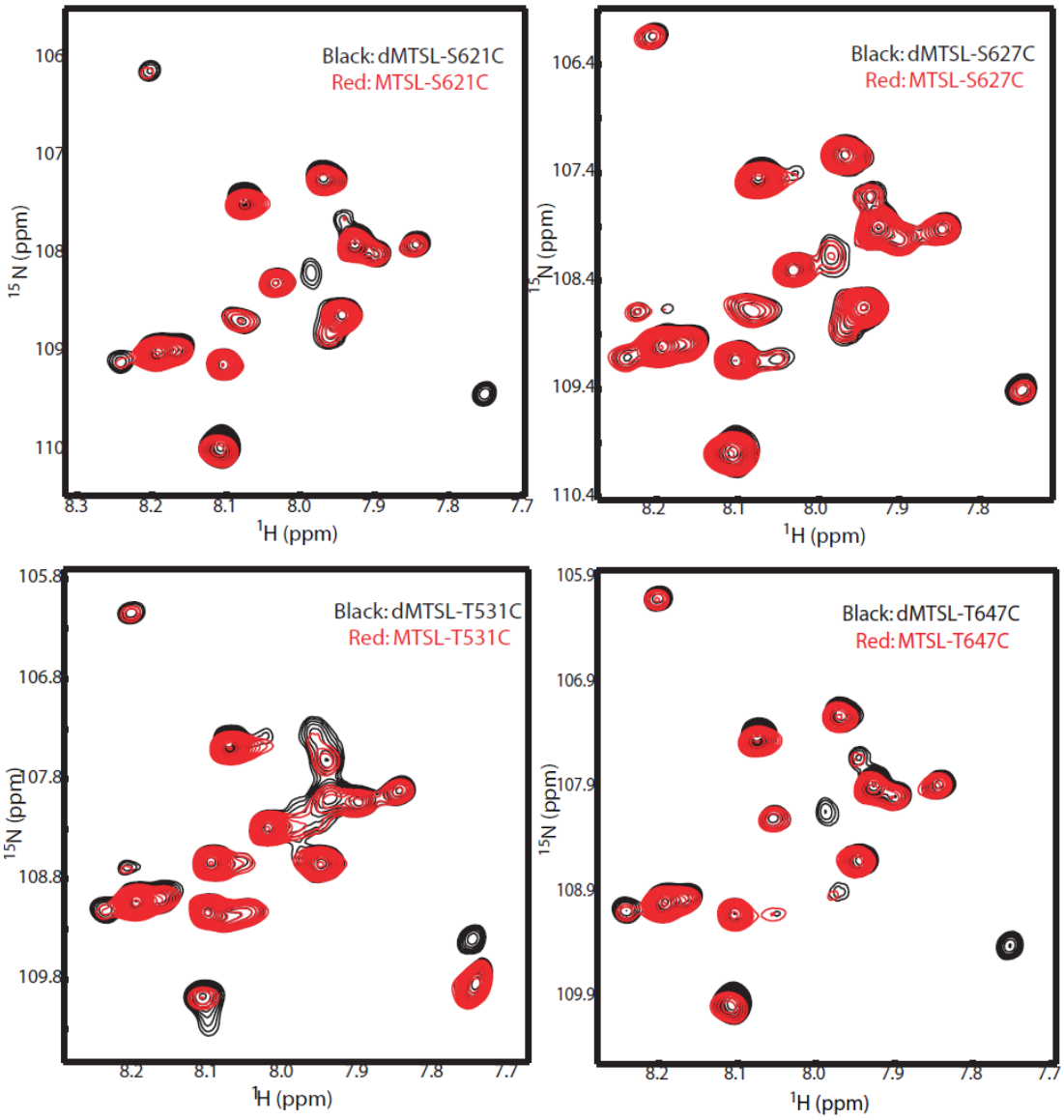


Figure caption is on page 157

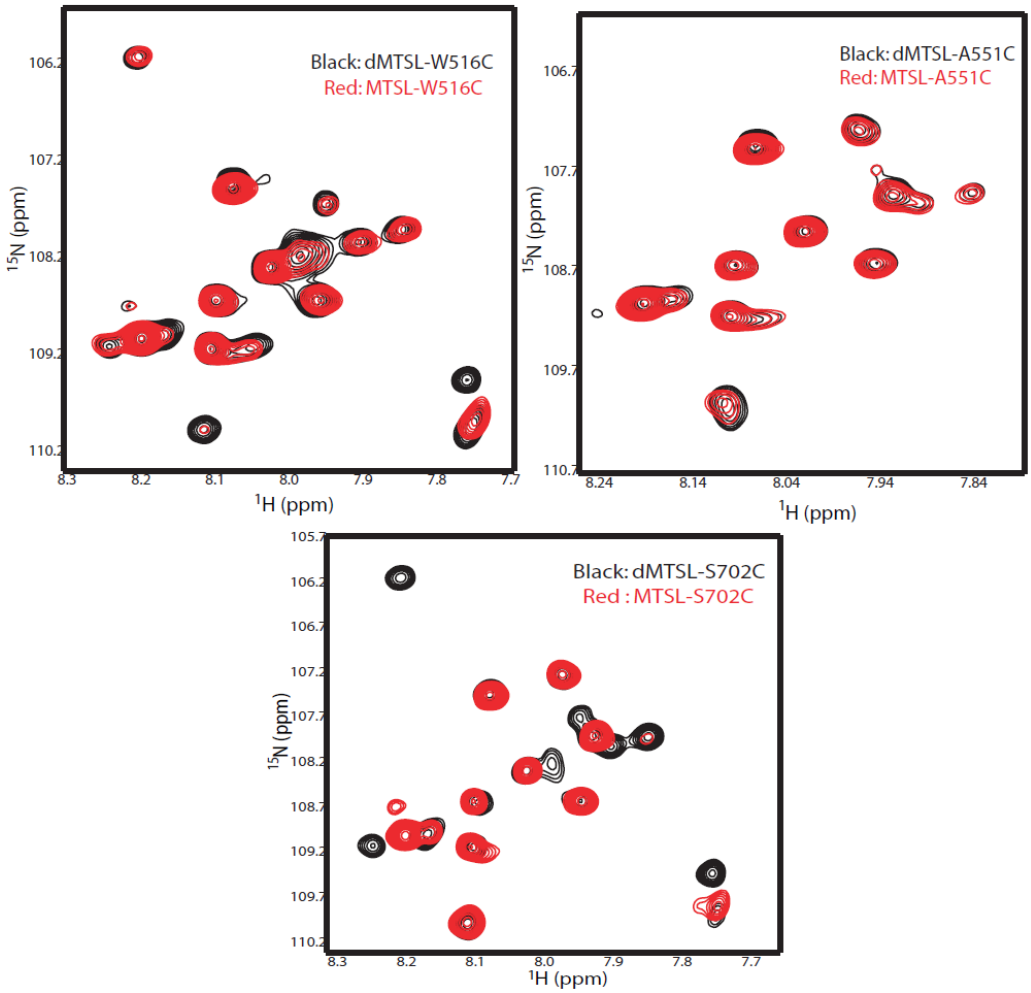


Figure 5.1 Overlay of part of [^{1}H, ^{15}N] - HSQC spectra of the MTSL-labeled and dMTSL-labeled mono-cysteine mutants of the C-terminal domain of Stt3p. The spectra in black are dMTSL-labeled protein while the spectra in red are MTSL-labeled. Comparisons of peaks intensities between MTSL-labeled and dMTSL-labeled were used to obtain long-range distance restraints.

5.2 Constraints from Residual Dipolar Couplings (RDC)

Residual dipolar couplings (RDCs) have recently emerged as a new tool in NMR with which to study macromolecular structure and function in a solution environment. RDCs are complementary to the more conventional use of NOEs to provide structural information: while NOEs provide local distance restraints in nature (typically within 5 Å), RDCs provide long-range orientational information and can be readily employed to improve the structural accuracy. Thus, RDCs are now widely utilized in protein structure calculations, especially for α-helical integral membrane proteins, which usually lack enough long range NOEs.

The underlying mechanism for RDC is the through-space dipole-dipole interactions. For a pair of spin 1/2 nuclei (i and j) in a magnetic field, such as ^{1}H, ^{13}C or ^{15}N, the observable dipolar coupling, D_{ij}, can be expressed as Eq. 5.3 (below). In equation 1, r is the distance between a specific pair of nuclei, γ_i and γ_j are the magnetogyric ratios for the nuclei, μ_0 is the permittivity of space, h is Planck's constant, and θ is the angle between the considered internuclear vector and the magnetic field. When all the parameters are given in SI units, the resulting D_{ij} is given in the unit of hertz. Many measurements of RDCs are made between pairs of bonded nuclei, so that r is fixed; RDCs have, thus, been used primarily to provide angular information.

$$D_{ij} = -\frac{\mu_0 \gamma_i \gamma_j h}{(2\pi r)^3}\left\langle \frac{3\cos^2\theta - 1}{2} \right\rangle \qquad \text{Eq. 5.3}$$

Note that in Eq. 5.3, the brackets around the angular term denote averaging over the fast molecular motion that occurs in solution or liquid crystal media. If motion allows vectors to sample directions uniformly in space as a result of the effects of Brownian motion, the expression reduces to zero. Hence, the introduction of partial alignment of samples is the key to the observation of RDCs.

To date, several different media for aligning samples have been reported including bicelles (172), bacteriophage (173) and polyacrylamide gels (174). However, for micelle-solubilized membrane proteins, preparation of aligned samples to obtain reliable RDC values remains technically challenging (175-177). Bicelles and bacteriophage are incompatible with membrane proteins because their accompanying lipids are destructive to the bacteriophage media, interfere with some of the mixtures, or merge with the lipids in the bicelles (175). The most successful approach has been the incorporation of membrane proteins into compressed polyacrylamide gels. Polyacrylamide gel is the only medium suitable for alignment of membrane proteins reconstituted in detergent micelles, because it is chemically inert, therefore, the samples are stable over a wide range of temperature, ionic strength, and pH (176, 178). Strain-induced alignment in polyacrylamide gel (SAG) employs either vertical or radial compression of the gel in order to alter the pore shape and induce preferential alignment of the protein. Furthermore, the extent and direction of alignment can be ‘‘tuned’’ by physically altering the mechanical compression or the gel composition, for example, by addition of a charged component to the gel (174, 176-179).

Assuming the principal alignment frame is known, equation 5.3 is usually

rewritten as (for theoretical details, see reference 6):

$$D_{ij} = D_a [(3\cos^2\theta - 1) + 3/2\ R \sin^2\theta\cos^2\phi] \qquad \text{Eq. 5.4}$$

where Da and R are the axial and rhombic components, respectively, of the molecular alignment tensor, **A**, in the principal coordinate frame. Molecular alignment tensor contains the principal components A_{xx}, A_{yy}, and A_{zz}. According to typical convention, the magnitudes of the principal components are $|A_{zz}| \geq |A_{yy}| \geq |A_{xx}|$. D_a is equal to $1/3[A_{zz} - (A_{xx} + A_{yy})/2]$ and R is equal to $2/3(A_{xx} - A_{yy})/A_{zz}$. D_a is in units of hertz and R is unitless and always positive.

For the purpose of incorporating RDCs to NMR structure calculations, RDCs are usually not used in initial structure calculations, but rather in a refinement stage of structure calculations. The reasons are that the potential energy surface is very rough and including RDCs initially may trap the structure into a false minimum, leading to convergence problems (180).

In this section, in order to induce alignment of structurally useful degrees and resulting resolvable RDCs of the C-terminal domain in SDS micelles, neutral polyacrylamide gels, together with a series of charged polyacrylamide gels (positively, negatively charged and zwitterionic) were prepared. The RDC values obtained were analyzed and used for final structure refinement.

5.2.1 Methods and Materials

5.2.1.1 Sample preparation

The expression, isolation, and purification of the uniformly ^{15}N-labeled C-terminal domain of Stt3p have been previously described in Chapter 2. Isotropic

samples for solution NMR spectroscopy consist of 150 μM protein, 100 mM sodium dodecyl-d_{25} sulfate (SDS, Aldrich), 1mM EDTA, 10% D_2O, 1% glycerol (v/v), 25 mM phosphate buffer, pH 6.5.

Polyacrylamide gel samples were prepared from a stock solution of 40% (w/v) acrylamide (Sigma) and 2% (w/v) N, N′-methylenebisacrylamide (EMD Chemicals Inc.). Samples were prepared by mixing the appropriate amount of acrylamide, N, N′-methylenebisacrylamide, and water to make the desired acrylamide concentration. To introduce charge to the gel, 5% acrylamide was replaced by an equimolar amount of acrylic acid (J. T. Baker) or diallyldimethylammonium chloride (DADMAC; Sigma-Aldrich, Inc.) to make gel negatively charged and positively charged respectively. Zwitterionic gel was prepared by adding equimolar amount of acrylic acid and DADMAC (both are 5% of acrylamide concentration). Chemical polymerization was initiated by the addition of 0.08% w/v ammonium persulfate (APS) and 0.6% N,N,N′,N′- tetramethylethylene diamine (TEMED). Right before the addition of TEMED, the mixture solution was filtered through a 0.2 μm syringe filter to remove any natural polymerized impurities. The gel sample was prepared by using the apparatus from New Era Enterprises. Briefly, after addition of all components, the mixture solution was immediately transferred to the gel chamber of 5.4 mm diameter and allowed to polymerize for at least 2 hours. Polymerized gels were first dialyzed overnight against deionized water to remove unreacted chemicals. Subsequently, the gels were dehydrated at 37 ℃ oven for at least 24 hours, prior to soaking for 24 hours in 400 μL protein sample in 5.4 mm diameter cylinder. Gels were forced into an open-

ended NMR tube through a connecting funnel (Figure 5.2).

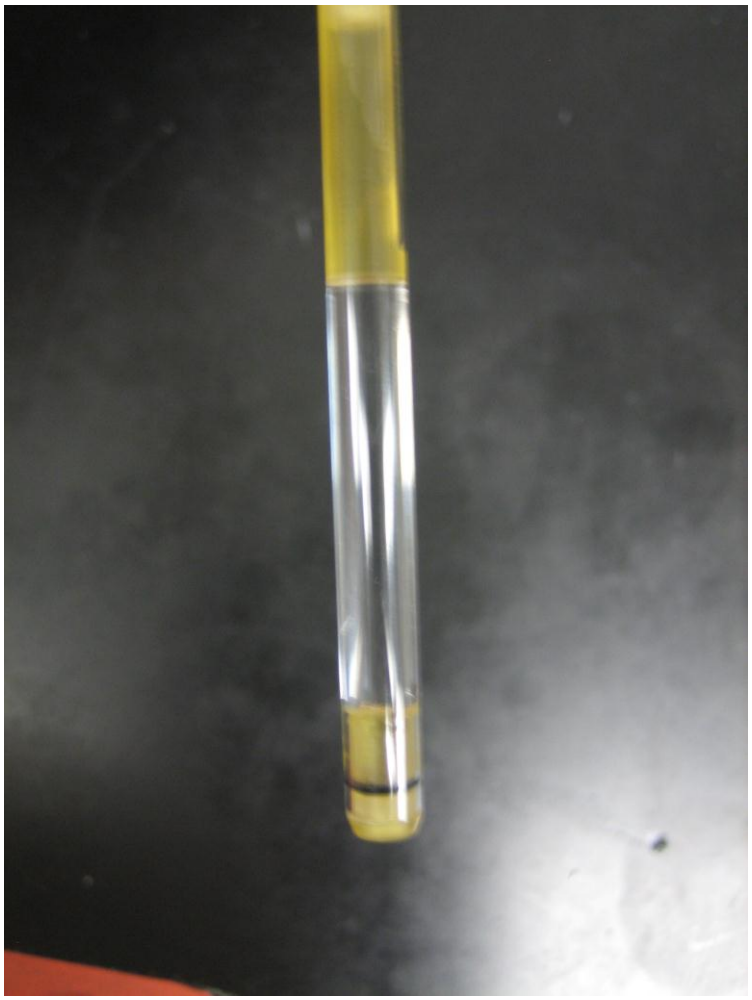


Figure 5.2 Picture of protein sample for RDC studies. The ^{15}N-labeled C-terminal domain of Stt3p is weakly aligned in 6% polyacrylamide gel in NMR tube as the sample for RDC study. Note the homogeneity of NMR sample is critical for good shimming results.

5.2.1.2 NMR data collection and analysis

NMR measurements were conducted at 328 K. 1D ^{2}H spectra were acquired with the deuterium field/frequency lock turned off. IPAP (in-phase anti-phase)-HSQC was employed to determine the $^{1}D_{HN}$ residual dipolar couplings (181). NMR raw data of IPAP-HSQC were split by the AU program, SPLITIPAP2, provided by Bruker

TOPSPIN 2.0 and the resulting data were then processed and analyzed by NMRPipe (104) and NMRView programs (152) respectively.

5.2.2 Results and Discussion

5.2.2.1 Alignment in 6% polyacrylamide gels

The requirement of detergent makes the studies of integral membrane proteins consistently more complicated. Large membrane protein-detergent complexes often require the preparation of gels at low concentration while retaining high homogeneity, which otherwise leads to NMR spectra of poor resolution. On the other hand, gels at too low concentration usually cannot provide resolvable RDCs. Therefore, although the methods for the alignment of soluble proteins have been well-established, the application of this technique for integral membrane proteins is significantly more demanding.

In order to prepare suitable samples for RDC studies of the C-terminal domain of Stt3p, a series of gels based on polyacrylamide with different charged properties were screened. The introduction of gel anisotropy was achieved by stretching of the gel in the axial direction of the NMR tube- forcing a cylindrical gel of bigger diameter (5.4 mm) into an open-ended NMR tube of smaller internal diameter (4.2 mm). Hence, after the gel was squeezed into the NMR tube, the pores within the gel on average were prolate-shaped with their long axis parallel to the NMR tube. It is worthy to point out that this method ensures samples of high homogeneity, making it possible to utilize automatic gradient shimming.

The deuterium solvent signal, normally used for field-frequency lock purposes,

can provide a very convenient probe for monitoring the weak alignment of protein sample in RDC studies (182). The rapid exchange of water molecules between the partially aligned hydration shell of oriented solutes and bulk solvent causes incomplete averaging of the ^{2}H quadrupole splitting (183). Thus, the presence of the alignment of RDC samples can be monitored by the observation of the deuterium splitting of the H_2O/D_2O solvent (typically value of several Hz).

Quadrupolar splitting of the ^{2}H NMR signal of the solvent was observed in the neutral polyacrylamide gels (6%), with the solvent ^{2}H splitting of 3.8 Hz, which suggests the presence of weak alignment (Figure 5.3). However, when the concentration of acrylamide was decreased to 4%, no obvious splitting was observed, indicating its concentration was too low to obtain the sample anisotropy (data not shown). Polyacrylamide gel with higher concentration than 6% was not tested since 6% gels are already able to provide sufficient alignments for RDC studies and usually higher concentration of gel only leads to poorer-quality NMR spectra. It is noteworthy that, for each polyacrylamide gel sample, there is no noticeable temperature effect on the quadrupolar ^{2}H splitting from 303 K to 328 K. Furthermore, the temperature stability was evaluated by recording replicates of spectra for the C-terminal of Stt3p at 328 K which also showed no changes in the measured values over the course of 2 weeks. These observations confirm the reported long-term thermal stability of polyacrylamide gel as the protein alignment media for RDC studies at elevated temperature, which is often employed for membrane proteins (184).

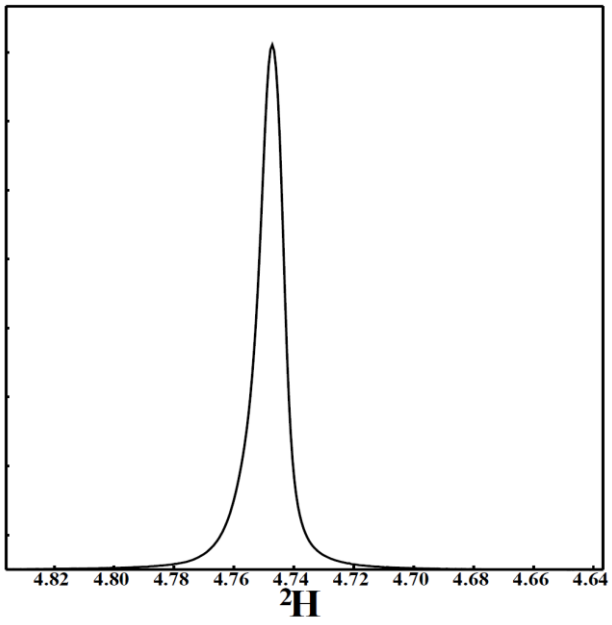


Figure 5.3 The solvent ^{2}H spectra of the C-terminal domain of Stt3p protein sample. A: in solution (no alignment); and B: in 6% polyacrylamide neutral gel. The presence of the ^{2}H splitting of 3.8 Hz indicates the protein is properly aligned in this medium.

5.2.2.2 Effects of charge of gel

In the anisotropically stretched charged polyacrylamide gels, protein orientation is determined by both steric effects and electrostatic interactions. It has been demonstrated that the protein alignments were significantly different in positively and negatively charged gels (185). One major advantage of application of different alignments is that it can reduce inherent degeneracy in dipolar couplings in terms of orientations (186) and dramatically improves the accuracy of calculated structures (187). For polyacrylamide gel, its different electrostatic environments can be readily achieved by adding different chemicals.

Here, negatively charged gels were prepared by copolymerization of acrylic acid with acrylamide, while positively charged gels were obtained by copolymerization with DADMAC (diallyldimethylammonium chloride). Zwitterionic polyacrylamide gels were generated by copolymerization with equal molar amount of DADMAC and acrylic acid. As expected, during dialysis, all of these gels underwent dramatically electroosmotic swelling in water.

As shown in Figure 5.3 and 5.4, quadrupolar splittings of the ^{2}H NMR signal of the solvent were observed in all of the four media: neutral, negatively-charged, positively charged and zwitterionic polyacrylamide gels (6%), with the solvent ^{2}H splitting of 3.8 Hz, 2.9 Hz, 3.8 Hz and 3.6 Hz, respectively. The similar values of ^{2}H splitting suggest comparable magnitudes of sample alignment in media of different charged properties (Figure 5.4).

One major drawback of using charged orienting media for RDC studies is the

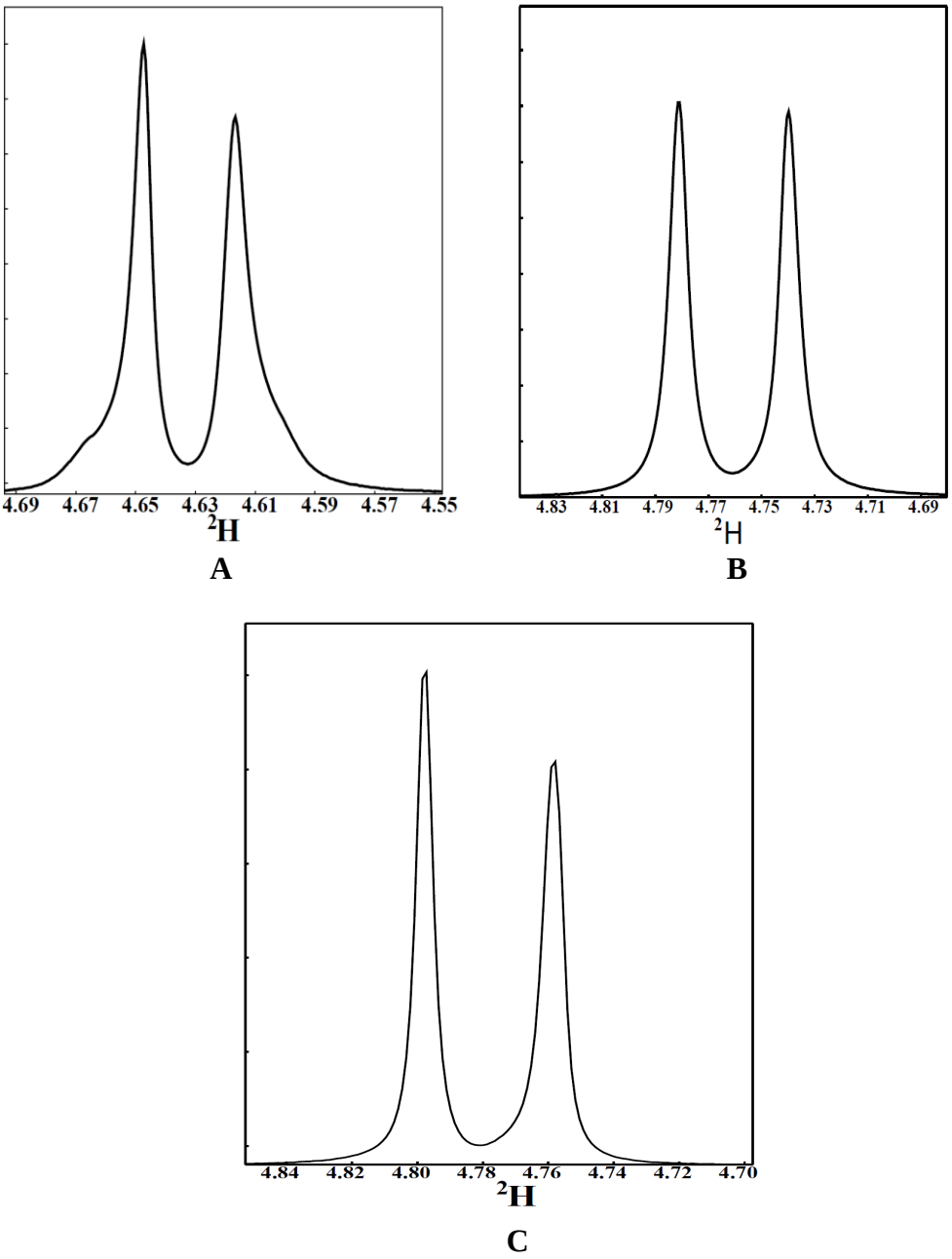
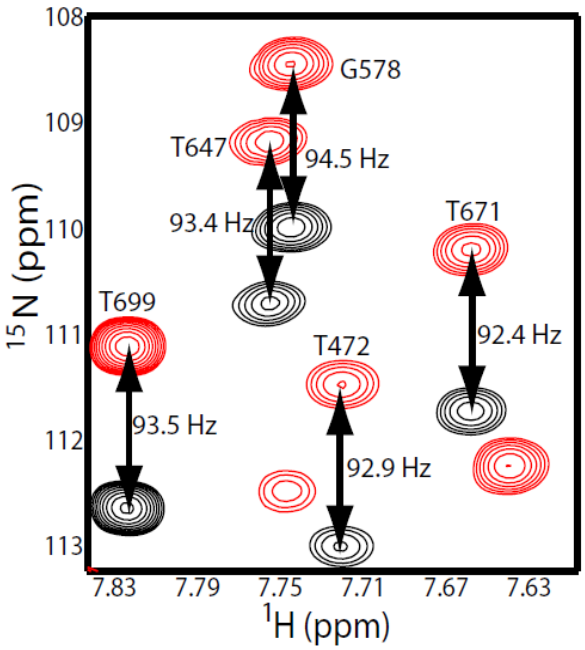


Figure 5.4 Quadrupolar splittings of the ^{2}H NMR spectra of the solvents for the C-terminal domain of Stt3p in polyacrylamide gels with different charges. A: negatively-charged gel; B: positively charged gel and C: zwitterionic gel, with the solvent ^{2}H splitting of 2.9 Hz, 3.8 Hz and 3.6 Hz, respectively.

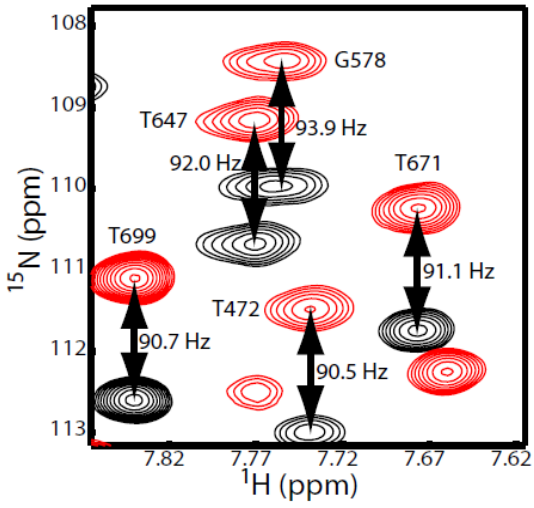
risk of the unfavorable electrostatic interactions between media and the protein, which usually results in degradation of the quality of NMR spectra (175). Our experimental data confirmed that the strong electrostatic interactions played an important role in determination of the resulting NMR spectral qualities. Compare to the relatively good quality NMR spectra obtained from neutral polyacrylamide gels, all charged gels, including zwitterionic gels, resulted in significant line-broadening or missing of many resonances in the NMR spectra (Figure 5.5). This is especially true for the negatively charged gels, which yielded a poorly-resolved spectrum with a very low signal/noise ratio even though the concentration of protein used was reasonably high. Presumably, the reason for this could be the electrostatic repulsive force formed between the fixed negative charges of the polyelectrolyte and the SDS micelles, which prevents the ready diffusion of the protein into the gel pores. In principle, such unfavorable electrostatic interactions may be partially quenched by addition of high concentration of inert salt (such as NaCl) as an intermediate. However, no further optimization was performed since the cryo-probe of the Bruker NMR instrument used here for RDC studies requires low-salt or no salt environments of the samples.

5.2.2.3 Analysis of RDC data

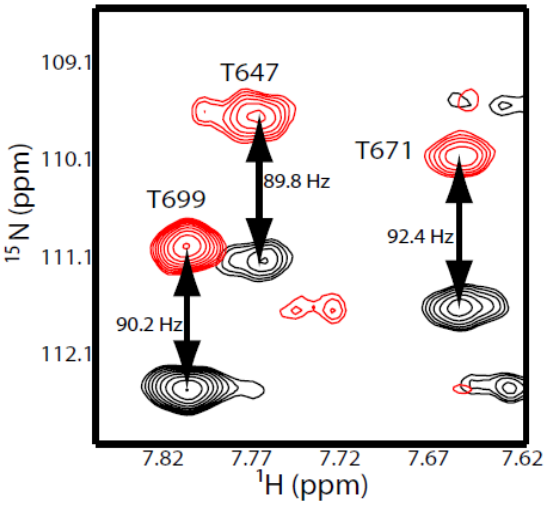
RDCs were measured for those well-resolved peaks with no overlap from nearby neighboring resonances. The measured RDCs of the C-terminal domain of Stt3p in different alignment media are listed in Appendix Table A-3. As mentioned above, due to resonance broadening effects, the numbers of RDCs obtained from charged media are very limited. In fact, the numbers of RDCs obtained from well-


108
109
110
111
112
113
^{15}N (ppm)
G578
T647
94.5 Hz
93.4 Hz
T671
T699
92.4 Hz
T472
93.5 Hz
92.9 Hz
7.83
7.79
7.75
7.71
7.67
7.63
^{1}H (ppm)

A


108
109
110
111
112
113
^{15}N (ppm)
G578
T647
93.9 Hz
92.0 Hz
T671
T699
91.1 Hz
T472
90.7 Hz
90.5 Hz
7.82
7.77
7.72
7.67
7.62
^{1}H (ppm)

B


109.1
110.1
111.1
112.1
^{15}N (ppm)
T647
89.8 Hz
T671
T699
92.4 Hz
90.2 Hz
7.82
7.77
7.72
7.67
7.62
^{1}H (ppm)

C

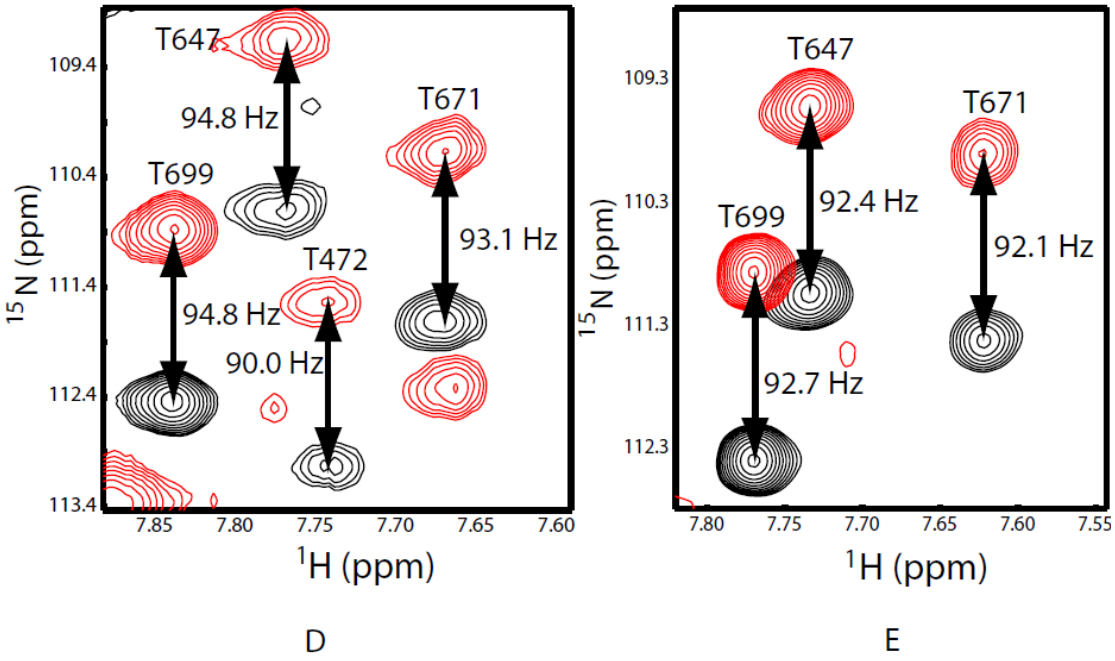


Figure 5.5 IPAP-HSQC spectra for the C-terminal domain of Stt3p showing values of $^{1}D_{HN}$ coupling constants in different media. A: no alignment, B: neutral 6% Polyacrylamide gel, C: negative charged polyacrylamide gel, D: positive charged polyacrylamide gel, and E: zwitterionic polyacrylamide gel.

resolve peaks are 51, 19, 28 and 20 for the alignment media of no charge (neutral), negative charge, positive charge and zwitterionic charge, respectively.

Before using RDCs in any type of structure refinement, as implied in Eq. 5.4, good estimates for D_a and R must be available. There are several methods for determining D_a and R, but if no structure information is available prior to the refinement, like the case for the C-terminal domain of Stt3p, one useful method is the "histogram method" demonstrated by Clore *et al.* (188). In this method, the RDCs are measured and plotted in a histogram. This histogram closely resembles a CSA

(Chemical Shift Anisotropy) powder pattern spectrum characteristic of solid-state NMR spectra, in which values for A_{zz}, A_{yy}, and A_{xx} are taken from the three extrema of the histogram: the high extreme values, the low extreme values and the most populated values, respectively. These values can be used with Eq. 5.5, 5.6, and 5.7 to solve for D_a and R.

. When $\theta = 0$, $A_{zz} = 2Da$ (Eq. 5.5)

When $\theta = \pi/2$, $\phi = \pi/2$, $A_{yy} = -Da\ (1+3R/2)$ (Eq. 5.6)

When $\theta = \pi/2$, $\phi = 0$, $A_{xx} = -Da\ (1-3R/2)$ (Eq. 5.7)

According to the "histogram method", experimentally, the values of A_{zz} and A_{yy} are obtained by taking the average of the high and low extreme values of the residual dipolar couplings, respectively. The value of A_{xx} corresponds to the most populated value in the histogram of the observed RDCs. With two unknowns and three observables (A_{xx}, A_{yy}, and A_{zz}), the values of D_a and R can then be calculated.

By taking the high and low extreme values of two residual dipolar couplings (for neutral gel, three extreme RDC values were taken for averaging), the values of D_a (R) were calculated as 2.61 (0.48), 3.47 (0.22), 3.42 (0.68), and 2.86 (0.16) for the protein sample in neutral, negatively charged, positively charged and zwittersionic gel, respectively. These values, together with the RDCs, will be incorporated into structural refinement of the C-terminal domain of Stt3p.

5.3 Topology Determination of the C-terminal domain of Stt3p

A fundamental aspect of the structure of membrane proteins is their membrane topology, i.e. the number of membrane-embedded segments and their orientations in

the membrane. Fortunately, despite many difficulties in obtaining high-resolution structures of an IMP, the topology of an IMP can be predicted rather accurately by using computer programs based on the hydrophobicity analyses of their amino acid sequences (96, 189-196). In general, membrane protein topology predictions are based on the observations that: (1) the transmembrane α-helices have an overall high hydrophobicity; and (2) the charge distribution of the hydrophilic loops that connect the transmembrane segments follows the "positive inside" rule, which states that nontranslocated loops are enriched in positively charged residues compared to translocated loops (93). The first observation is used to identify the membrane-embedded segments in the amino acid sequence by analyzing the hydropathic properties of the amino acid sequence, and the second observation is used to predict the overall orientation of the protein in the membrane. However, in some cases, there remain some ambiguities regarding the prediction results, i.e., different computer programs can give different results. This is especially true for the proteins of which no homologous protein structures are available.

The topology of an IMP can also be determined experimentally. Classical *in vivo* IMP topology determination methods include Enzyme Tags, Glycosylation Tags, Chemical Modification, and BAD (Biotin Acceptor Domain) Tags (for review, see reference 197). Lately, a new method for *in vitro* topology study was introduced by incorporation of paramagnetic spin reagents to the IMP/detergent complex and then measuring the effect of paramagnetic relaxation enhancement (PRE) by NMR (198).

In 2005, the *in vitro* membrane topology of the full-length Stt3p of the yeast,

Saccharomyces cerevisiae, was determined experimentally by Kim *et al.*, using C-terminal reporter fusions and insertion of glycosylation sites. It is shown that the full-length Stt3p has eleven trans-membrane domains, with the N-terminus located in cytosol and the C-terminus in ER lumen (91). However, this result didn't show a detailed membrane topology mapping of C-terminal domain of Stt3p (residues 466-718) because this part of protein was somehow neglected and no glycosylation site was chosen after the residue 440.

In order to address the question as to whether C-terminal domain of Stt3p has some membrane-embedded domains (or trans-membrane domain), here, a comprehensive study was carried out by using both hydrophobic and hydrophilic paramagnetic spin reagents as the membrane topology probes. Our experimental results, together with the program prediction results demonstrated in Chapter Two, show that the C-terminal domain of Stt3p contains at least one membrane-embedded domain. This result is consistent with the previous observations that the C-terminal domain of Stt3p is insoluble in water unless the proper detergent is added.

5.3.1 Methods and Materials

5.3.1.1 Determination of transmembrane (TM) domain by titration of C-terminal domain of Stt3p with paramagnetic relaxation enhancement agents- 16-doxyl-stearic acid (16-DSA)

16-Doxyl-stearic acid (16-DSA) was used as the hydrophobic paramagnetic spin probe to determine the trans-membrane domain of the C-terminal domain of Stt3p in SDS micelles. Titrations were performed by stepwise addition of 16-DSA to a

constant amount of protein. Briefly, 16-DSA was dissolved in methanol to make a stock solution of 50 mM. An appropriate amount of 16-DSA stock solution was transferred to an eppendorf tube, and the solvent was evaporated using a SpeedVac Concentrator (Thermo Electron Co.) without heating. The protein sample was then added to the dried aliquot to make the desired 16-DSA concentration. 16-DSA titrated over a concentration range of 0-2 mM to a 0.1 mM U-^{15}N-protein sample. [^{1}H, ^{15}N]-HSQC experiments were carried out using the same parameters except P_1 (the 90 degree hard pulse) and shimming values. The peak intensities were measured at each titration point to assess the amount of paramagnetic induced line broadening.

5.3.1.2 Determination of water-exposed domain by titration of C-terminal domain of Stt3p with hydrophilic paramagnetic relaxation enhancement agents-Gd-DTPA, (Gd(III)-diethylenetriaminepentaacetic acid)

It has been reported that the addition of Gd-DTPA can lead to NMR spectra that exhibit not only line broadening but also some peak shifting effects, most likely due to the transient coordination of Gd(III) ligand sites with negatively charged side chains on the protein. To remove the unwanted interactions, EDTA, an effective chelating agent, is suggested to add simultaneously with Gd-DTPA (199). The Gd-DTPA stock solution contained 150 mM Gd-DTPA, 250 mM EDTA, 100 mM SDS (deuterated), 1% glycerol and 10% D_2O, in 25 mM phosphate buffer, pH 6.5. Gd-DTPA was added over a concentration range of 0-10 mM to a 1 mM U-^{15}N- sample of the C-terminal domain of Stt3p from the stock solution. [^{1}H, ^{15}N]-HSQC spectra were collected using the same parameters except P_1 (the 90 degree hard pulse) and

shimming, and the peak intensities were measured at each titration point to assess the amount of paramagnet induced line broadening.

5.3.1.3 NMR experiments

The U-^{15}N-labeled C-terminal Stt3p was prepared as described previously. NMR measurements were conducted at 328 K. The data were acquired with 256 and 2048 complex points in the t_1 time domain (^{15}N dimension) and t_2 time domain (^{1}H dimension) respectively. The data were zero-filled to 512×4096 and apodized using a Gaussian window function prior to Fourier transformation using NMRPipe (104).

5.3.2 Results and Discussion

The detailed topology mapping of the C-terminal domain of Stt3p with respect to the micellar membrane was obtained by assessing backbone amide proton accessibility to the polar and nonpolar paramagnetic probes Gd-DTPA and 16-DSA. The reductions in peak intensities induced by paramagnetic electrons were recorded from HSQC spectra acquired in both the absence and presence of these probes (Figure 5.6 and 5.7).

The result of 16-DSA titration is shown in Figure 5.8. In this figure, only those peaks that are both assigned and sufficiently well resolved to allow unambiguous measurement of peak intensities were used. The experimental results indicate that the presence of 16-DSA results in significant reductions in peak intensities for the following segments: residues 488-504, 511-526, 539-551, and 705-718. Since segment of residues 566-582 is also predicted by computer programs as transmembrane or membrane embedded domain (see Chapter 3), we conclude that

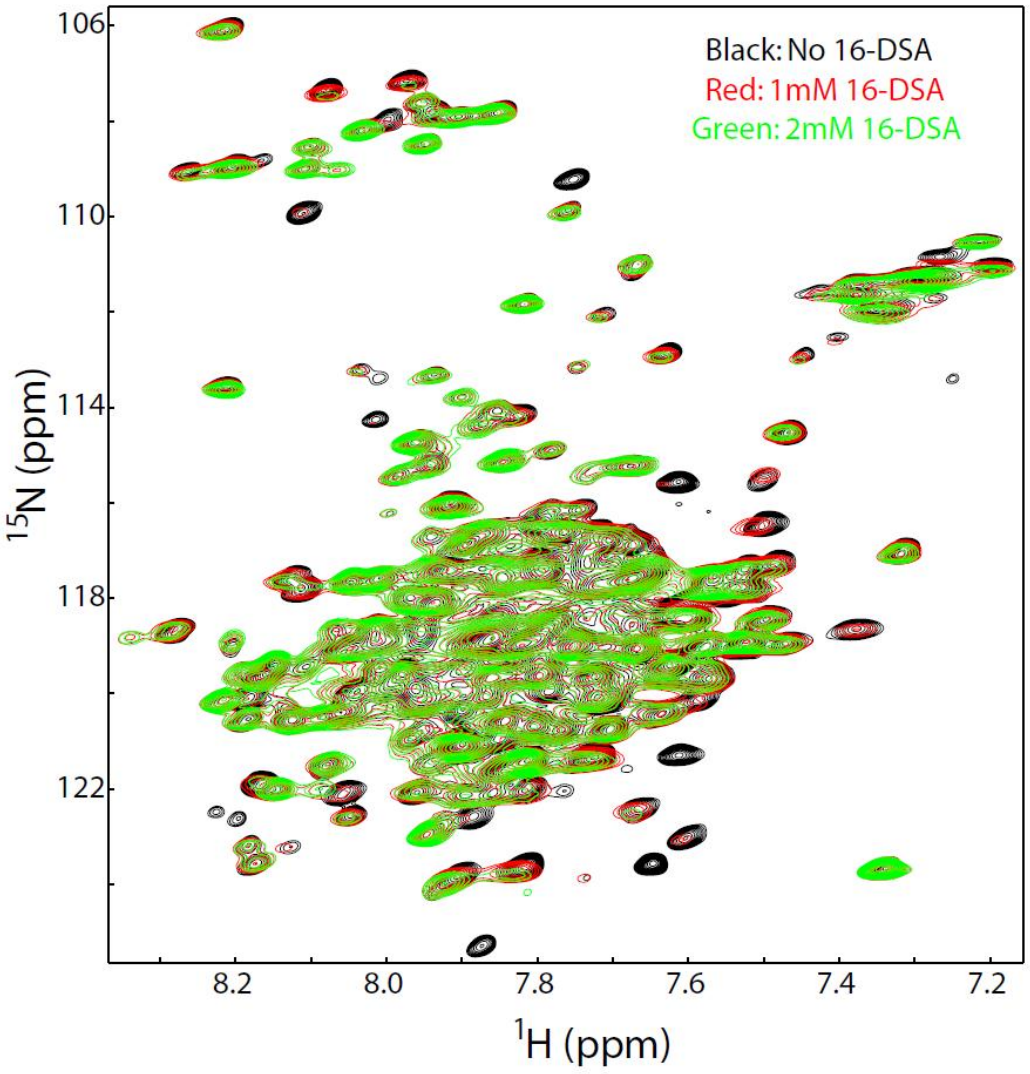


Figure 5.6 Effects of 16-DSA on [^{1}H, ^{15}N]-HSQC peak intensities for the U-^{15}N-labeled C-terminal domain of Stt3p in SDS micelles at pH 6.5 and 328 K. The black peaks represent the HSQC spectrum of the C-terminal domain of Stt3p in absence of a paramagnetic probe. The superimposed red and green spectra were acquired after addition of 16-DSA to a concentration of 1 mM and 2 mM respectively.

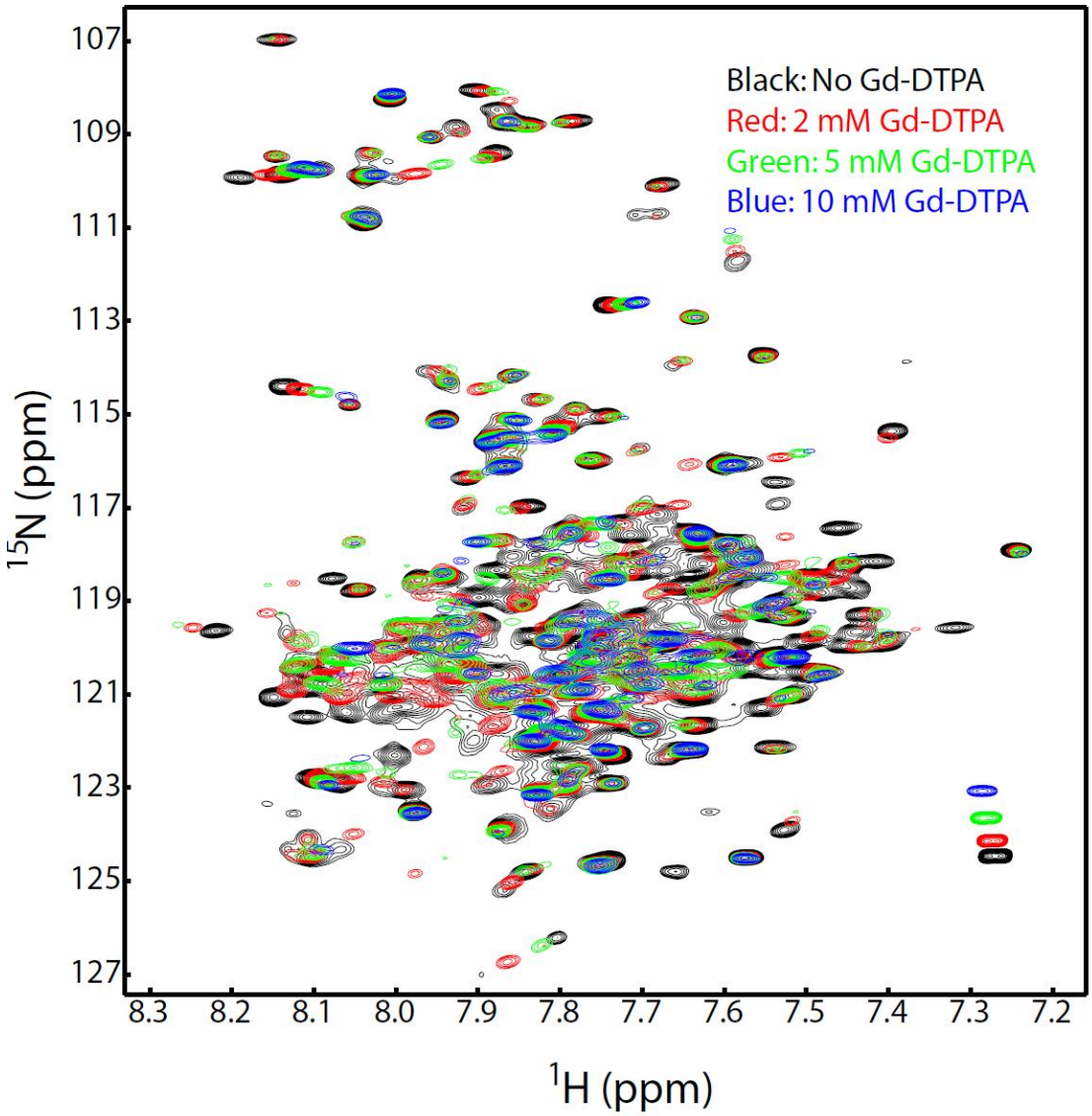


Figure 5.7 Effects of Gd-DTPA on [^{1}H, ^{15}N]-HSQC peak intensities for the U-^{15}N-labeled C-terminal domain of Stt3p. The black peaks represent the HSQC spectrum of the C-terminal domain of Stt3p in the absence of a paramagnetic probe. The superimposed red, green and blue spectra were acquired after addition of Gd-DTPA to a concentration of 2 mM, 5mM and 10 mM respectively.

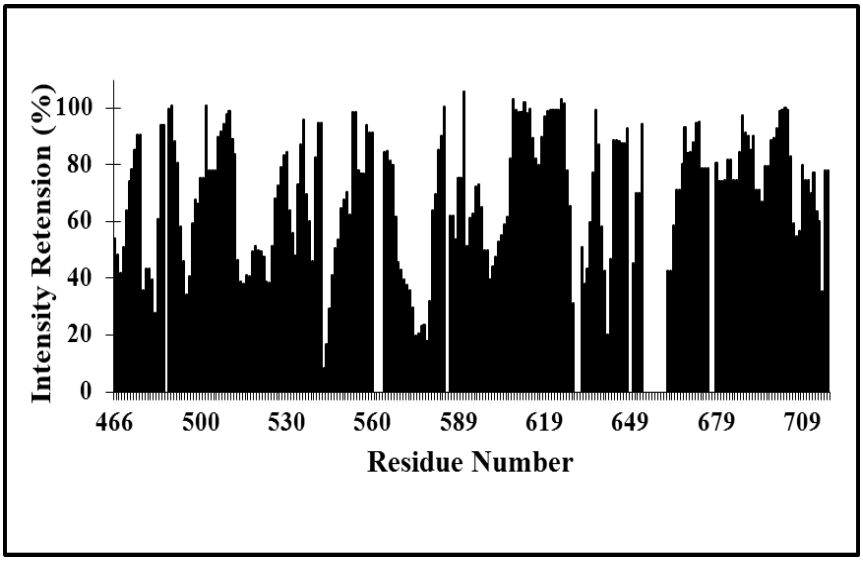


Figure 5.8 Site-specific reductions in ^{15}N-^{1}HN HSQC peak intensities as a result of adding 2mM 16-DSA to U-^{15}N labeled protein samples. The result was subsequently smoothed by averaging over three successive residues. The His-tagged residues were excluded from analysis.

this segment is the membrane embedded domain.

One possible explanation for the peak intensities reduction in the other segments is that those segments may form a hydrophobic pocket where 16-DSA can also be anchored. Actually, for a long time DSA has been used as a ligand to probe the interactions between fatty acid and hydrophobic pockets located on water soluble proteins (200-202).

As mentioned in Section 5.3.12, it has been shown that the presence of Gd-DTPA can lead to NMR spectra exhibiting not only line broadening but also the peak shifting effects (199). Beel *et al.* believe this is due to the specific interaction of Gd(III) with some negatively charged residues such as Asp, Glu, etc., therefore affecting the chemical environments around those residues. In their report, they

suggested using EDTA to remove or alleviate it (199). However, our results show that even in the presence of 250 mM EDTA, there were still many peaks shifted, making it very difficult to identify those residues, especially the residues located in the crowded center (Figure 5.7). Actually, in the case studied here, the presence of SDS, a negatively charged detergent, can make the resulting data more problematic than simple peak shifting. These is because positively charged Gd(III) can, at least in principle, can bind to negatively charged SDS micelles and affect those residues embedded in SDS micelles. So, we suggest that, once charged micelles are used, the data derived from using Gd-DTPA to probe the water-exposed domain of membrane protein need to be interpreted with extreme caution.

5.4 Structure Calculation of the C-terminal domain of Stt3p

In this section, the solution structure of this 31.5 kDa helical membrane protein in detergent micelles as determined by high-resolution NMR will be presented. This will be the first structural report of eukaryotic Stt3p, and to our knowledge, the largest membrane protein whose 3D structure has been determined by NMR.

5.4.1 Methods

Backbone dihedral angle restraints were obtained from the backbone chemical shifts using TALOS+ (203, 204). Backbone hydrogen-bond restraints were included only for residues in helices, as determined and verified by the presence of a series of characteristic short-range and medium-range NOEs, together with the results from chemical shift index analysis (CSI) (142, 143). It was observed that calculations performed without inclusion of hydrogen-bonds yielded essentially identical helices.

Structure calculation was carried out with CYANA 2.1 (17) using dihedral angle, NOE, hydrogen-bond and PRE restraints. From the 100 initial generated structures, 10 structures of lowest total energy were chosen to represent the ensemble conformation of the C-terminal domain of Stt3p. The small residual constraint violations in the 10 conformers and the good coincidence of experimental NOEs show that the input data represent a self-consistent set and that the restraints are well satisfied in the calculated conformers.

Structural figures were prepared with either PyMOL (available on the World Wide Web: http://www.pymol.org) or MOLMOL (205).

5.4.2 Results and Discussion

In the present study, we have used NOEs from various samples including double-labeled, partially deuterated triple-labeled, uniformly {^{2}H, ^{13}C, ^{15}N}- triple-labeled and {^{2}H, ^{13}C, ^{15}N} triple-labeled ILV methyl protonated sample, together with backbone dihedral angles from chemical shift analysis, residual dipolar couplings (RDC), and paramagnetic relaxation enhancement (PRE) measurements from 15 nitroxide labeled samples. A summary of constraints used for structure determination is given in Table 5.1.

TALOS+ Analysis- The backbone dihedral angles (phi and psi) of a protein can be predicted by a program called “TALOS” (Torsion Angle Likeness Obtained from Shift and Sequence Similarity), which utilize chemical shifts for the calculation of phi and psi angles (203). In principle, TALOS divides the sequence of a protein input data into a series of tripeptide sets (residues i-1, i, and i+1) and compares them to the ten

best tripeptides (j-1, j, and j+1) in its database for matches in terms of both chemical shift and residue types. The TALOS database is from those proteins with high resolution crystal structures, which serves as the source of the phi and psi angles. Once the psi and phi angles of the tripeptides match with at least nine out of ten database values and fall within the same cluster of the Ramachandran map, TALOS can essentially make an accurate prediction of the torsion angles for a residue. Lately, the updated version of TALOS, TALOS+, was reported, which significantly expanded its database, from a 20-protein database to a 200-protein database (204).

The dihedral angles of the C-terminal domain of Stt3p are predicted by the TALOS+ program and the values are shown in Appendix Table A-4. The data shows the protein under study contains eleven helices: residues S466-A473 (α1), L490-N506 (α2), S511-I523 (α3), N535-S552 (α4), E559-I577 (α5), E600-Y608 (α6), A618-K635 (α7), G644-M654 (α8), L663-F670 (α9), S672-A687 (α10), and L692-R700 (α11); and two β-strands: W589-E595 (β1) and S708-R711 (β2). These results are in a striking agreement with the CSI program prediction results. The only minor difference is that in the CSI analysis, the small β-strand near the C-terminus of protein, β2, was not identified. This is because although the negative sign of the chemical shift differences in CSI analysis is also indicative of a β-strand, the values of difference are not large enough to be conclusive. It is noteworthy that comparing to TALOS, TALOS+ gives more prediction results which are classified as “good” values (data not shown), presumably due to its larger database. Most areas where TALOS+ cannot give good predictions are located in loop regions.

Table 5.1 Summary of NMR restraints statistics for the structure calculation of the C-terminal domain of Stt3p used at the moment of writing.

Distance Constraints	NOE	Intraresidue NOEs 158
		Sequential (i – j = 1) NOEs 1432
		Medium-range (1 < i – j < 5) NOEs 563
		Long-range (i – j ≥ 5) NOEs 126
	Hydrogen Bond	132
	PRE	Upper Bound 245
		Lower Bound 768
Dihedral Angle Constraints	Φ	253
	Ψ	253
Residual Dipolar Coupling ($^1D_{HN}$)	Neutral Gel	51
	Positive Gel	38
	Negative Gel	19
	Zwitterionic Gel	20

Overall Structure and Topology of the C-terminal Domain of Stt3p- As presented in Figure 5.9 and 5.10, the structure of the C-terminal domain of Stt3p reveals an overall "oblate spheroid" shape structure, with a major axis of ~68 nm and minor axis of ~37 nm. We note that this model is compatible with the low-resolution structure of the luminal domain of Stt3p subunit determined by EM methods, in which it represents an overall "platform-shape" (Figure 5.9A). The C-terminal domain of Stt3p is primarily helical, containing eleven helices. Its high helicity is consistent with our previous experimental data from far-UV CD and CSI analysis (90). Although both TALOS+ program and CSI predict the formation of a β-strand encompassing residues R592-W598, we were not able to find supportive NOEs to confirm it, presumably due to peak overlapping. In fact, the absence of β-sheet has even been correctly predicted by some structure predictions programs (83). For later descriptive purpose, these eleven helices are named as α1-α11 from the N-terminal to C-terminal along with its sequence. Figure 5.11 shows surface electrostatic potential of the C-terminal Domain of Stt3p. As expected for a monotopic membrane protein with only a small domain embedded in membrane, it retains a large hydrophilic surface and is assembled internally with a typical hydrophobic core.

Our earlier topology studies show that the addition of hydrophobic probe, 16-DSA, resulted in significant reduction in peak intensities for the following segments: residues 488-504 (α2), 511-526 (α3), 539-551 (α4), 566-582 (α5), and 705-718. From the structure presented in Figure 5.10, we observe that except peptide segment 566-582, the α5 helix, the other peptide segments namely, α2, α3, α4, and α5 helices

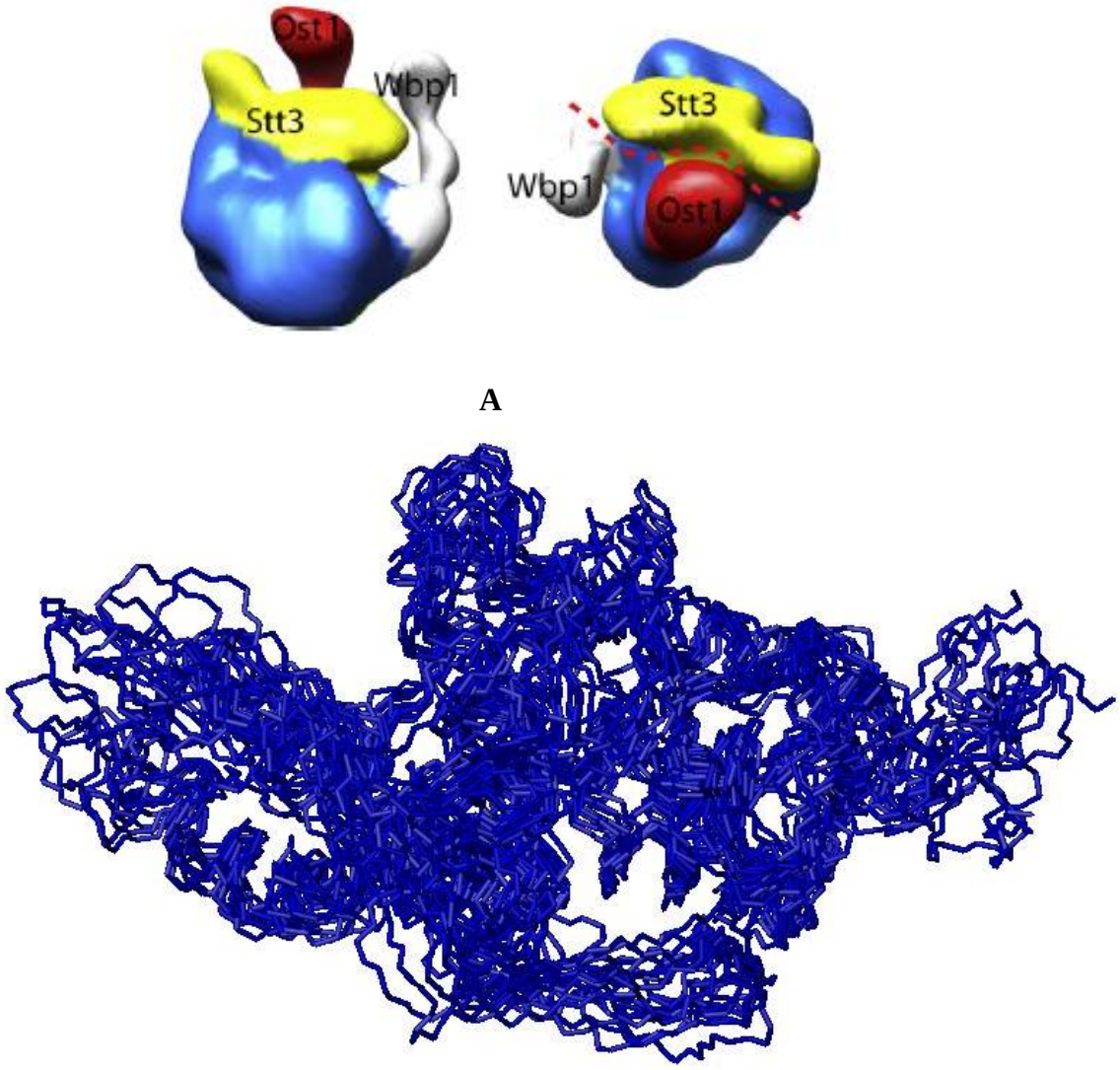


Figure 5.9 Solution structure of the C-terminal domain of Stt3p. A: Low-resolution Cryo-EM structure of the OT complex. Figure is obtained from reference 89 with permission. B: Superposition of 10 conformers representing the final NMR structure.

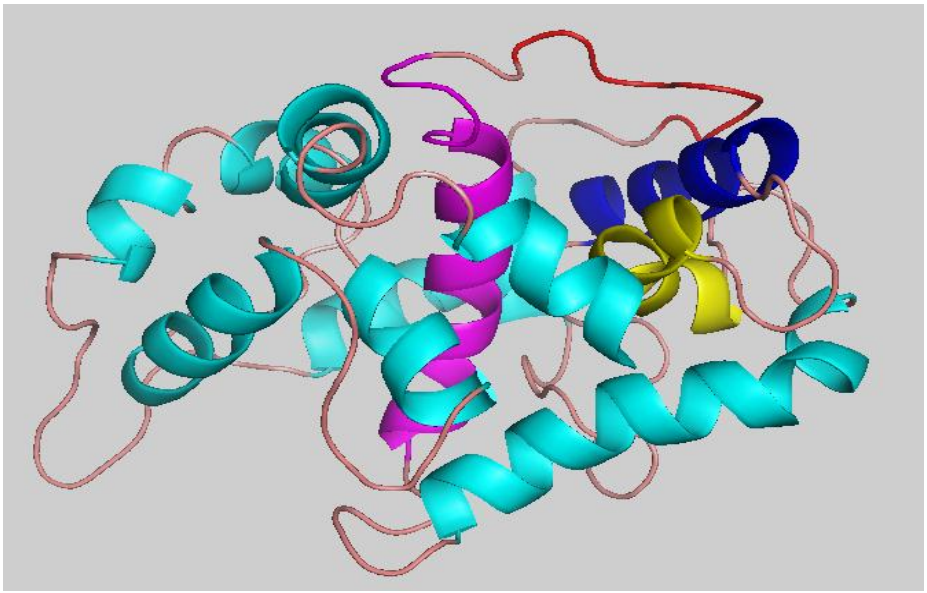


Figure 5.10 Ribbon structure of the lowest energy conformer.

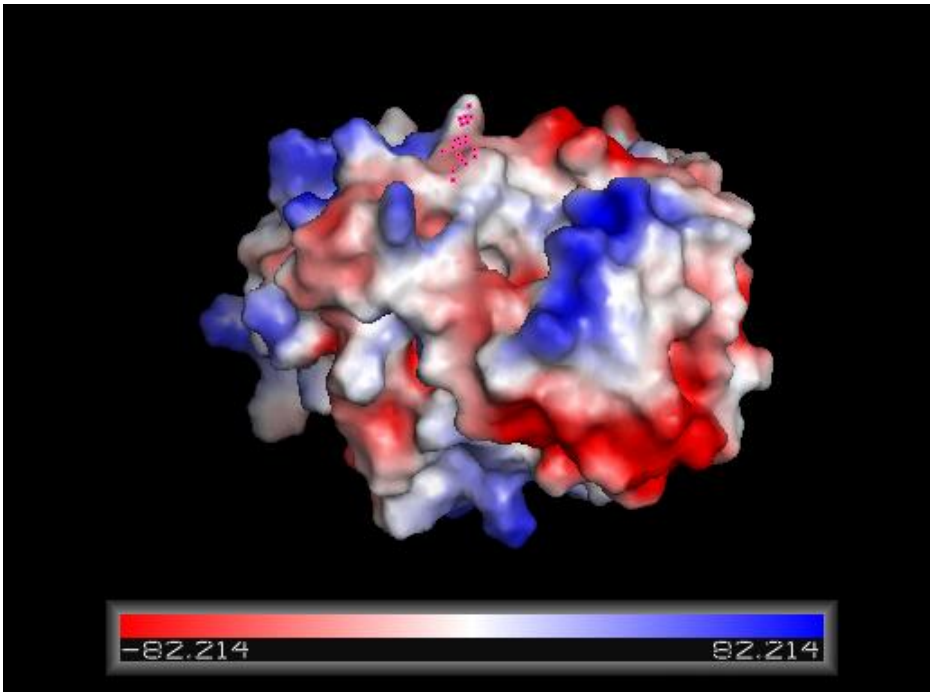


Figure 5.11 Electrostatic potential of the C-terminal domain of Stt3p. Negatively charged surface is in red, positively charged surface is in blue, while nonpolar surface is in white.

that show resonance intensity reduction, are located in the core of the protein, forming a hydrophobic pocket. This explains the reason for resonance intensity reduction in those segments since the nonpolar probe 16-DSA most likely anchors in the hydrophobic pocket. We predict that helix-5, encompassing residues 566 to 582, is a membrane-embedded domain. Since it is known that the C-terminal domain of Stt3p is located in the luminal side of ER, we conclude this membrane embedded segment must not be a transmembrane domain. According to Blobel, membrane proteins are categorized as monotopic, bitopic and polytopic, depending on the mode by which the protein interacts with the membrane (206). The monotopic proteins only interact with one of the monolayer leaflets of the bilayer, while bitopic and polytopic proteins have one or more segments spanning the full membrane bilayer, respectively. Therefore, we conclude the protein under study belongs to monotopic membrane protein. This topology model is also consistent with the structure shown in Figure 5.12, namely, the highly hydrophobic helix-5 penetrates into the lipid bilayer.

Moreover, according to Nilsson *et al.*, the catalytic site is 30-40 Å above the membrane in the EM structure of yeast OT and is oriented roughly parallel to the membrane surface (68, 207). As shown in Figure 5.13, our model is consistent with these results.

Two crystal structures of the C-terminal soluble domain of prokaryotic Stt3p homologs have been reported: *P. furiosus* AglB (79) and *C. jejuni* PglB (83), although both have very limited sequence similarity with eukaryotic Stt3p. The C-terminal AglB protein consists of four structural domains, one mainly α-helical “central core”

Figure 5.12 Ribbon structure of the lowest energy conformer to show the proposed membrane-embedded domain. Proposed membrane-embedded domain is shown in color of blue.

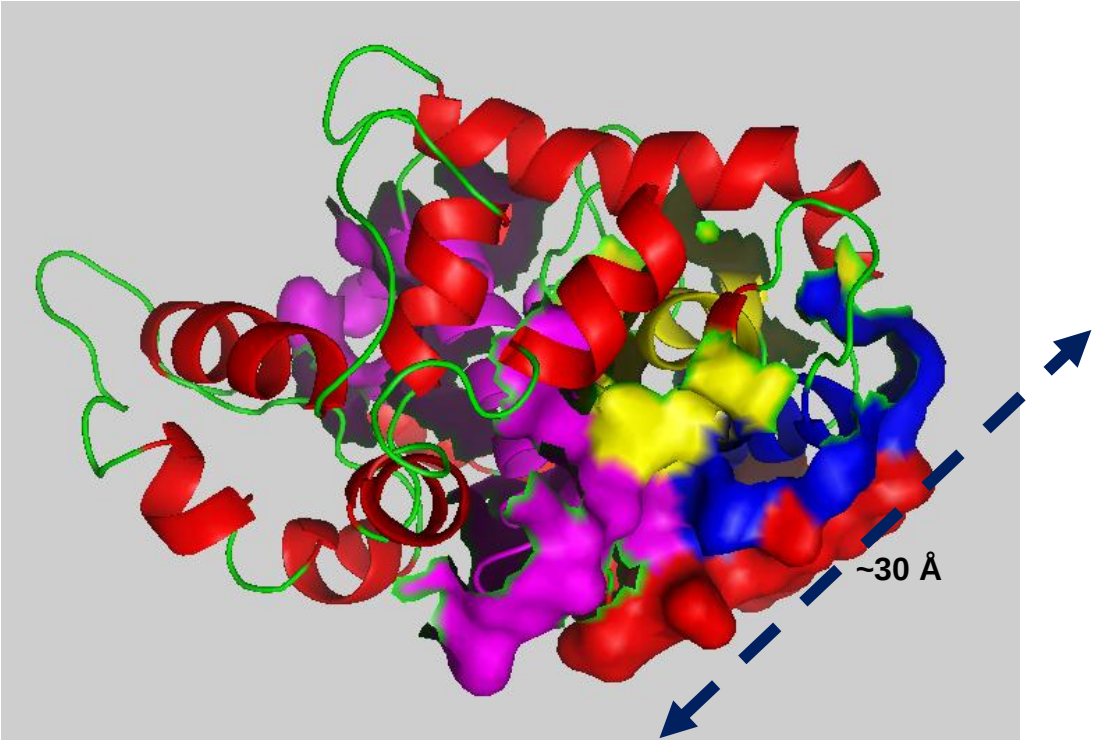


Figure 5.13 Distance Measurement from the proposed membrane embedded segment to the WWDYG motif.

domain located at the center, and three β-sheet-rich domains encircling the "central core" domain; whereas the much smaller C-terminal PglB protein contains only two structural domains, one mainly α-helical "central core" domain and one "inserted" β-sheet-rich domain. In the structures of both homologs, the "central core" domain, contains the well-conserved WWDYG motif and were therefore proposed as catalytic domain.

Comparison with the crystal structures of prokaryotic Stt3p homologs: The two crystal structures of the C-terminal soluble domain of prokaryotic Stt3p homologs

AglB of *P. furiosus* (79) and PglB of *C. jejuni* (83) comprise primarily an α-helical "central core" domain that is encircled by β-sheet-rich domains for AglB whereas for PglB, the β-sheet-rich domains insert into the α-helical "central core" domain. The "central core" domain containing the well-conserved WWDYG motif is therefore proposed to be the catalytic domain.

Comparison of the solution structure of the C-terminal domain of yeast Stt3p with its prokaryotic homolog structures reveals that there are two major differences. First, the C-terminal domains of both prokaryotic Stt3p homologs are water-soluble, lacking any membrane-embedded segment, while our data show that the C-terminal domain of yeast Stt3p is a monotopic membrane protein, with helix-5 embedded into the lipid bilayer. We postulate that the anchoring of the OT catalytic center to the ER membrane makes it close to its donor substrate, dolichol-linked oligosaccharide, which is also embedded in ER membrane. This can potentially increase the effective local concentration of the donor substrate, and hence facilitates the N-glycosylation process. A second striking difference is that the counterparts to the β-sheet-rich domains in AglB and PglB proteins are missing in the C-terminal domain of yeast Stt3p. It is thus reasonable to postulate that the C-terminal domain of Stt3p as a whole is corresponding to the "central core domain" in the above two homologs, although it has a larger size and contains more helical elements. The function of those β-sheet-rich domains might be fulfilled by the other subunit(s) in the case of yeast OT.

We focus our attention on the catalytic center of Stt3p. The highly conserved WWDYG motif (residues 516-520) in the C-terminal domain of Stt3p has been

reported to play a central role in the glycosylation process, and point mutations in this motif either eliminate or sharply reduce OT activity (66). Based on the two crystal structures of prokaryotic Stt3p homolog proteins and phylogenetics studies, Maita *et al.* proposed that the catalytic site of eukaryotic OT is formed between the WWDYG motif and the DK motif (83), a so-called "A-type catalytic center". Our solution structure of the C-terminal domain of yeast Stt3p reveals that despite their very limited amino acid sequence similarities, the structure of the catalytic center of yeast Stt3p is similar to its prokaryotic homologs to some extent, although constructed in a more sophisticated manner. As shown in Figure 5.13, we propose that the surface of the catalytic site of yeast Stt3p is formed by α2, α3, α4, α5 and the DK motif (, residues D582-E586), among which α3 contains the conserved WWDYG motif. It is noteworthy that we are uncertain about the local secondary structure of the DK motif of the C-terminal domain of Stt3p because we are missing the NMR assignments for the following residues, I584 and N585. However, from CSI analysis and TALOS+ prediction based on those assigned neighboring residues (data not shown), it is very likely there is a small helix formed between residues D582 to E587 (data not shown).

In summary, the high-resolution structures we present here comprise the first high-resolution structure of the catalytic domain of the eukaryotic OT complex. However, as shown below in Figure 5.14, Ramachandran plot demonstrate that some residues are still located in disallowed regions, even though most residues are in energy-favored regions. Therefore, right now we are still working on structural refinement, trying to reduce the root mean square deviation (RMSD) value and

structural violations. Considering the high sequence homology among eukaryotic Stt3p, we hope our results can provide a significant step toward the structural understanding of the mechanisms of the N-glycosylation in eukaryotes.

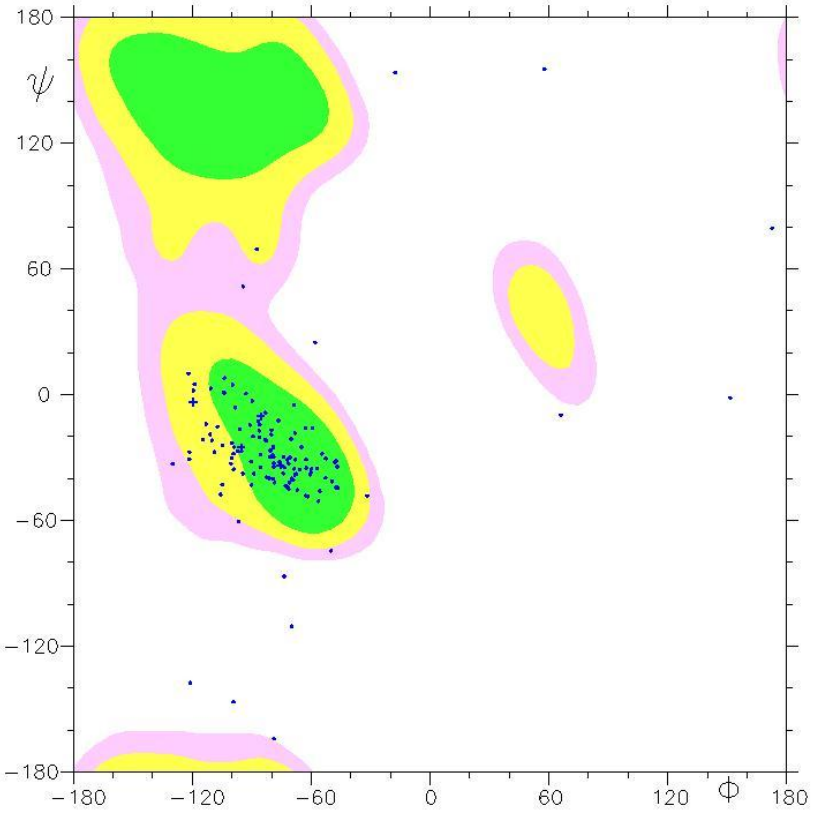


Figure 5.14 Ramachandran plot of the C-terminal domain of Stt3p.

REFERENCES

1. Roberts, G. C. K. (1993) NMR of macromolecules. A practical Approach. *Oxford University Press.*

2. Kainosho, M., Torizawa, T., Iwashita, Y., Terauchi, T., Ono, M. A., and Güntert, P. (2006) Optimal isotope labeling for NMR protein structure determinations. *Nature 440*, 52-57.

3. Tugarinov, V., and Kay, L. E. (2003) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods *J. Am. Chem. Soc. 125*, 13868-13878.

4. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Skelton, N. J., and Rance, M. (2006) Protein NMR spectroscopy: principles and practice. *Academic Press; 2nd ed.*

5. Reid D. G. (1997) Protein NMR techniques (methods in molecular biology). *Humana Press; 1st ed.*

6. Prestegard, J. H., Bougault, C. M., and Kishore, A. I. (2004) Residual dipolar couplings in structure determination of biomolecules, *Chem. Rev. 104*, 3519-3540.

7. Solomon, I. (1955) Relaxation processes in a system of two spins. *Phys. Rev. 99*, 559-565.

8. Overhauser, A. W. (1953) Paramagnetic relaxation in metals. *Phys. Rev. 89*,

689-700.

9. Overhauser, A. W. (1953) Polarization of nuclei in metals. *Phys. Rev. 92*, 411-415.
10. Anet, F. A. L., and Bourn, A. J. R. (1965) Nuclear magnetic resonance spectral assignments from nuclear Overhauser effects. *J. Am. Chem. Soc. 87*, 5250-5251.
11. Macura, S., and Ernst, R. R. (1980) Elucidation of cross relaxation in liquids by two-dimensional N.M.R. Spectroscopy. *Mol. Phys. 41*, 95-117.
12. Kumar, A., Ernst, R. R., and Wuthrich, K. (1980) A two-dimensional nuclear Overhauser enhancement (2D NOE) experiment for the elucidation of complete proton–proton cross-relaxation networks in biological macromolecules. *Biochem. Biophys. Res. Commun. 95*, 1-6.
13. Jeener, J. (1971) Ampere International Summer School, Basko Polje, Jugoslavia, (unpublished).
14. Aue, W. P.; Bartholdi, E.; and Ernst, R. R. (1976) Two dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* 64, 2229-2246.
15. Bodenhausen, G.; and Ruben, D. J. (1980) Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem. Phys. Lett., 69*, 185-189.
16. Pervushin, K., Riek, R., Wider, G., and Wuthrich, K., (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological

macromolecules in solution. *Proc. Natl. Acad. Sci. USA. 94*, 12366–12371.

17. Güntert, P., Mumenthaler, C., Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J. Mol. Biol. 273*, 283-298.
18. Schwieters, C. D., Kuszewski, J. J., Tjandra, N., and Clore, G. M. (2003) The Xplor-NIH NMR molecular structure determination package. *J. Magn. Res. 160*, 66-74.
19. Schwieters, C. D., Kuszewski, J. J. and Clore, G. M. (2006) Using Xplor-NIH for NMR molecular structure determination. *Progr. NMR Spectroscopy 48,* 47-62.
20. Battiste, J. L., and Wagner, G. (2000) Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear Overhauser effect data. *Biochemistry* 39, 5355-5365.
21. Wallin, E., Tsukihara, T., Yoshikawa, S., von Heijne, G., and Elofsson, A. (1997) Architecture of helix bundle membrane proteins: An analysis of cytochrome c oxidase from bovine mitochondria. *Prot. Sci. 6,* 808–815.
22. Seshadri, K, Garemyr, R., Wallin, E., von Heijne, G., and Elofsson, A. (1998) Architecture of beta-barrel membrane proteins: analysis of trimeric porins. *Prot. Sci. 7,* 2026–2032.
23. Ulmschneider, M. B., Sansom, M. S., and Di Nola, A. (2005) Properties of integral membrane protein structures: Derivation of an implicit membrane

potential. *Proteins: Structure, Function, and Bioinformatics. 59,* 252–265.

24. Elofsson, A., and von Heijne, G. (2007) Membrane protein structure: Prediction versus reality. *Annu. Rev. Biochem. 76,* 125–140.
25. Sanders, C. R. and Myers, J. K. (2004) Disease-related misassembly of membrane proteins. *Annu. Rev. Biophys. Biomol. Struct. 33,* 25–51.
26. Hopkins, A. L., and Groom, C. R. (2002) The druggable genome. *Nat. Rev. Drug Discov.* 1, 727-730.
27. Yildirim, M. A., Goh, K. I., Cusick, M. E., Barabasi, A. L., and Vidal, M. (2007) Drug-target network. *Nat. Biotechnol. 25,* 1119–1126.
28. Torres, J., Stevens, T. J., and Samso, M. (2003) Membrane proteins: the 'Wild West' of structural biology. *Trends. Biochem. Sci. 28,* 137-144.
29. Loll, P. J. (2003) Membrane protein structural biology: the high throughput challenge. *J. Struct. Biol. 142,* 144–153.
30. Henderson, R., and Unwin, P. N. T. (1975) Three-dimensional model of purple membrane obtained by electron microscopy. *Nature 257,* 28–32.
31. Deisenhofer, J., Epp, O., Miki, K., Huber, R., and Michel, H. (1985) Structure of the protein subunits in the photosynthetic reaction centre of *Rhodopseudomonas viridis* at 3Å resolution. *Nature 318,* 618–624.
32. White, S. H. (2009) Biophysical dissection of membrane proteins. *Nature 459,* 344-346.
33. Weiss, H. M., and Grisshammer, R. (2002) Purification and characterization of the human adenosine A(2a) receptor functionally expressed in *Escherichia coli*.

Eur. J. Biochem. 269, 82-92.

34. Rhodes, D. (2002). Climbing mountains. A profile of Max Perutz 1914–2002: a life in science. *EMBO Rep. 3*, 393–395
35. Caffrey, M. (2003) Membrane protein crystallization. *J. Struct. Biol. 142*, 108-132.
36. Jaroniec, C. P., MacPhee, C. E., Bajaj, V. S., McMahon, M. T., Dobson, C. M., and Griffin, R. G. (2004) High-resolution molecular structure of a peptide in an amyloid fibril determined by magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. USA, 101*, 711-716.
37. Rienstra, C. M., Tucker-Kellogg, L., Jaroniec, C. P., Hohwy, M., Reif, B., McMahon, M. T., Tidor, B., Lozano-Perez, T., and Griffin, R. G. (2002) *De novo* determination of peptide structure with solid-state magic angle spinning NMR spectroscopy. *Proc. Natl. Acad. Sci. USA, 99*, 10260-10265.
38. Cady, S. D., Mishanina, T. V., and Hong M. (2009) Structure of amantadine-bound M2 transmembrane peptide of influenza A in lipid bilayers from magic-angle-spinning solid-state NMR: the role of Ser31 in amantadine binding. *J. Mol. Biol. 385*, 1127-1141.
39. Wang, J., Kim, S., Kovacs, F., and Cross, T. A. (2001) Structure of the transmembrane region of the M2 protein H(+) channel. *Prot. Sci.* 10, 2241-2250.
40. Cady, S. D., Schmidt-Rohr, K., Wang, J., Soto, C. S., Degrado, W. F., and Hong, M. (2010) Structure of the amantadine binding site of influenza M2

proton channels in lipid bilayers. *Nature*. 463, 689-692.

41. Opella, S. J., Marassi, F. M., Gesell, J. J., Valente, A. P., Kim, Y., Oblatt-Montal, M., and Montal, M. (1999) Structures of the M2 channel-lining segments from nicotinic acetylcholine and NMDA receptors by NMR spectroscopy. *Nat. Struct. Biol. 6*, 374-379.
42. Sanders, C. R., Hare, B. J., Howard, K. P., and Prestegard, J. H. (1994) Magnetically oriented phospholipid micelles as a tool for the study of membrane associated molecules. *Prog. NMR Spectrosc.* 26, 421–444.
43. Whiles, J. A., Brasseur, R., Glover, K. J., Giuseppe, M., Komives, E. A., and Vold, R. R (2001) Orientation and effects of mastoparan X on phospholipid bicelles. *Biophys. J. 80*, 280–293.
44. Losonczi, J. A. and Prestegard, J. H. (1998) Nuclear magnetic resonance characterization of the myristoylated, N-terminal fragment of ADP-ribosylation factor 1 in a magnetically oriented membrane array. *Biochemistry 37*, 706–716.
45. Arora, A., and Tamm, L. K. (2001) Biophysical approaches to membrane protein structure determination. *Curr. Opin. Struct. Biol*. *11*, 540–547.
46. Krueger-Koplin, R., Sorgen, P., Krueger-Koplin, S., Rivera-Torres, I., Cahill, S., Hicks, D., Grinius, L., Krulwich, T., and Girvin, M. (2004) An evaluation of detergents for NMR structural studies of membrane proteins. *J. Biomol. NMR. 28*, 43-57.
47. Rastogi, V. K. and Girvin, M. E. (1999) Structural changes linked to proton

translocation by subunit c of the ATP synthase. *Nature 402,* 263-268.

48. MacKenzie, K. R., Prestegard, J. H., and Engelman, D. M. (1997) A transmembrane helix dimer: structure and implications. *Science, 276,* 131-133.
49. Roosild, T. P., Greenwald, J., Vega, M., Castronovo, S., Riek, R., and Choe, S. (2005) NMR structure of Mistic, a membrane-integrating protein for membrane protein expression. *Science, 307,* 1317-1321.
50. Arora, A. Abildgaard, F., Bushweller, J. H., and Tamm, L. K. (2001) Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat. Struct. Biol. 8,* 334-338.
51. Hwang, P. M., Choy, W. Y., Lo, E. L., Chen, L., Forman-Kay, J. D., Raetz, C. R. H., Prive, G. G, Bishop, R. E., and Kay, L. E. (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc. Natl. Acad. Sci. USA. 99,* 13560-13565.
52. Fernandez, C., Hilty, C., Wider, G., Guntert, P., and Wuthrich, K. (2004) NMR structure of the integral membrane protein OmpX. *J. Mol. Biol.* 336, 1211-1221.
53. Knauer, R., and Lehle, L. (1999) The Oligosaccharyltransferase complex from Saccharomyces cerevisiae. *Biochim. Biophys. Acta. 1426,* 259–273.
54. Silberstein, S., and Gilmore, R. (1996) Biochemistry, molecular biology, and genetics of the oligosaccharyltransferase. *FASEB J. 10,* 849-858.
55. Welply, J. K., Shenbagamurthi, P., Lennarz, W. J., and Naider, F. (1983) Substrate recognition by oligosaccharyltransferase. Studies on glycosylation of

modified Asn-X-Thr/Ser tripeptides. *J. Biol. Chem. 258*, 11856–11863.

56. Apweiler, R., Hermjakob, N., and Sharon, N. (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta., 1473*, 4-8.
57. Petrescu, A. J. Milac, A. L., Petrescu, S. M., Dwek, R. A., and Wormald, M., R. (2004) Statistical analysis of the protein environment of N-glycosylation sites: implications for occupancy, structure, and folding. Glycobiology, 14, 103-114.
58. Kasturi, L., Chen, H., and Shakin-Eshleman, S. H. (1997) Regulation of N-linked core glycosylation: use of a site-directed mutagenesis approach to identify Asn-Xaa-Ser/Thr sequons that are poor oligosaccharide acceptors. *Biochem. J. 323*, 415-419.
59. Mellquist, J. L., Kasturi, L., Spitalnik, S. L., and Shakin-Eshleman, S. H. (1998) The amino acid following an asn-X-Ser/Thr sequon is an important determinant of N-linked core glycosylation efficiency. *Biochemistry*, *37*, 6833-6837.
60. Roitsch, T., and Lehle, L. (1989) Structural requirements for protein N-glycosylation. *Eur. J. Biochem.* 181, 525-529.
61. Dempski, R. E., Jr., and Imperiali, B. (2002) Oligosaccharyl transferase: Gatekeeper to the secretory pathway. *Curr. Opin. Chem. Biol.* **6**, 844-850.
62. Helenius, A., and Aebi, M. (2004) Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem. 73*, 1019-1049.
63. Lehle,L. Strahl, S., and Tanner, W. (2006) Protein glycosylation, conserved

from yeast to man: a model organism helps elucidate congenital human diseases. *Angew. Chem. Int. Ed. 45,* 6802 – 6818

64. Marquardt, T., and Denecke, J. (2003) Prenatal cardiac ultrasound finding in congenital disorder of glycosylation type 1a. *Eur. J. Pediatr. 162,* 359-379.
65. Yan, Q., Prestwich, G.. D., and Lennarz, W. J. (1999) The Ost1p subunit of yeast oligosaccharyl transferase recognizes the peptide glycosylation site sequence, -Asn-X-Ser/Thr-. *J. Biol. Chem. 274,* 5021-5025.
66. Yan, Q., and Lennarz, W. J. (2002) Studies on the function of oligosaccharyl transferase subunits. Stt3p is directly involved in the glycosylation process. *J. Biol. Chem.* 277**,** 47692-47700.
67. Yan, A., and Lennarz, W. J. (2005) Two oligosaccharyl transferase complexes exist in yeast and associate with two different translocons. *Glycobiology 15,* 1407-1415.
68. Chavan, M., Yan, A., and Lennarz, W. J. (2005) Subunits of the translocon interact with components of the oligosaccharyl transferase complex. *J. Biol. Chem. 280,* 22917-22924.
69. Schulza, B. L., Stirnimann, C. U., Grimshawc, J. P., Brozzoc, M. S., Fritscha, F., Mohorkoc, E., Capitanib, G., Glockshuberc, R., Grütterb, M. G., and Aebi, M. (2009) Oxidoreductase activity of oligosaccharyltransferase subunits Ost3p and Ost6p defines site-specific glycosylation efficiency. *Proc. Natl. Acad. Sci. USA. 106,* 11061–11066**.**
70. Zubkov, S., Lennarz, W. J., and Mohanty, S. (2004). Structural basis for the

function of a minimembrane protein subunit of yeast oligosaccharyltransferase. *Proc. Natl. Acad. Sci. USA. 10,* 3821–3826.

71. Spirig, U., Bodmer, D., Wacker, M., Burda, P., and Aebi, M. (2005) The 3.4-kDa Ost4 protein is required for the assembly of two distinct oligosaccharyltransferase complexes in yeast. *Glycobiology 15,* 1396-1406.
72. Reiss, G., te Heesen, S., Gilmore, R., Zufferey, R., and Aebi, M. (1997) A specific screen for oligosaccharyltransferase mutations identifies the 9 kDa OST5 protein required for optimal activity *in vivo* and *in vitro*. *EMBO J 16***,** 1164-1172.
73. Pathak, R., Hendrickson T. L., and Imperiali, B. (1995) Sulfhydryl modification of the yeast Wbp1p inhibits oligosaccharyl transferase activity. *Biochemistry 34,* 4179-4185.
74. Beatson, S., and Ponting, C. P. (2004) GIFT domains: Linking eukaryotic intraflagellar transport and glycosylation to bacterial gliding. *Trends Biochem. Sci. 29,* 396-399.
75. Zufferey, R., Knauer, R., Burda, P., Stagljar, I., te Heesen, S., Lehle, L., and Aebi, M. (1995) Stepwise assembly of the lipid-linked oligosaccharide in the endoplasmic reticulum of Saccharomyces cerevisiae: identification of the ALG9 gene encoding a putative mannosyl transferase. *EMBO J. 14,* 4949–4960.
76. Kelleher, D. J., Karaoglu, D., Mandon, E. C., and Gilmore, R. (2003) Oligosaccharyltransferase isoforms that contain different catalytic STT3

subunits have distinct enzymatic properties. *Mol. Cell. 12*, 101–111.

77. Nilsson I., Kelleher, D. J., Miao, Y., Shao, Y., Kreibich, G., Gilmore, R., von Heijne, G., and Johnson, A. E. (2003) Photocross-linking of nascent chains to the STT3 subunit of the oligosaccharyltransferase complex. *J. Cell Biol., 161*, 715–725.
78. Glover, K. J., Weerapana, E., Numao, S., and Imperiali, B. (2005) Chemoenzymatic synthesis of glycopeptides with PglB, a bacterial oligosaccharyl transferase from *Campylobacter jejuni*. *Chem. Biol. 12*, 1311–1315.
79. Igura, M., Maita, N., Kamishikiryo, J., Yamada, M., Obita1, T., Maenaka, K., and Kohda, D. (2008) Structure-guided identification of a new catalytic motif of oligosaccharyltransferase. *EMBO J. 27*, 234–243.
80. Feldman, M. F., Wacker, M., Hernandez, M., Hitchen, P. G., Marolda, C. L., Kowarik, M., Morris, H. R., Dell, A., Valvano, M. A., and Aebi, M. (2005) Engineering N-linked protein glycosylation with diverse O antigen lipopolysaccharide structures in *Escherichia coli*. *Proc. Natl. Acad. Sci.* USA. *102*, 3016–3021.
81. Nasab, F. P., Schulz, B. L., Gamarro, F., Parodi, A. J., and Aebi, M. All in one: Leishmania major STT3 proteins substitute for the whole oligosaccharyltransferase complex in *Saccharomyces cerevisiae*. (2008) *Mol. Biol. Cell. 19*, 3758-3768.
82. Hese, K., Otto, C. Routier, F. H., and Lehle, L. (2009) The yeast

oligosaccharyltransferase complex can be replaced by STT3 from Leishmania major. *Glycobiololy. 19*, 160–171.

83. Maita, N., Nyirenda, J., Igura, M., Kamishikiryo, J., and Kohda, D. (2010) Comparative structural biology of eubacterial and archaeal oligosaccharyltransferases, *J. Biol. Chem. 285*, 4941–4950.
84. Kelleher, D. J., Karaoglu, D., Mandon, E. C., and Gilmore, R. (2003) Oligosaccharyltransferase isoforms that contain different catalytic STT3 subunits have distinct enzymatic properties. *Mol. Cell 12*, 101–111.
85. Ruiz-Canada, C., Kelleher, D. J., and Gilmore, R. (2009) Cotranslational and posttranslational *N*-glycosylation of polypeptides by distinct mammalian OST isoforms. *Cell 136*, 272– 283.
86. Karaoglu, D., Kelleher, D. J., and Gilmore, R. (1997) The highly conserved Stt3 protein is a subunit of the yeast oligosaccharyltransferase and forms a subcomplex with Ost3p and Ost4p. *J. Biol. Chem. 272*, 32513–32520.
87. Yan, A., Ahmed, E., Yan, Q., and Lennarz, W. J. (2003) New findings on interactions among the yeast oligosaccharyl transferase subunits using a chemical cross-linker**.** *J. Biol. Chem. 278*, 33078–33087.
88. Yan, A., and Lennarz, W. J. (2005) Unraveling the mechanism of protein N-glycosylation. *J. Biol. Chem. 280*, 3121-3124
89. Li, H., Chavan, M., Schindelin, H., Lennarz, W. J., and Li, H. (2008) Structure of the oligosaccharyl transferase complex at 12 Å resolution. *Structure 16*, 432-440.

90. Huang, C., Mohanty, M., and Banerjee, M. (2010) A novel method of production and biophysical characterization of the catalytic domain of yeast oligosaccharyl transferase. *Biochemistry, 49*, 1115–1126.

91. Kim, H., von Heijne, G., and Nilsson, I. (2005) Membrane topology of the STT3 subunit of the oligosaccharyl transferase complex. *J. Biol. Chem. 280*, 20261-20267.

92. Cserzo, M., Wallin, E., Simon, I., von Heijne G., and Elofsson, A. (1997) Prediction of transmembrane alpha-helices in procariotic membrane proteins: the Dense Alignment Surface method. *Prot. Eng.* 673-676.

93. von Heijne, G. (1992) Membrane protein structure prediction. *J. Mol. Biol. 225*, 487-494.

94. Hofmann, K., and Stoffel, W. (1993) TMBASE - A database of membrane spanning protein segments. *Biol. Chem. Hoppe-Seyler 374*,166.

95. Juretic, D., Zoranic, L., and Zucic, D. (2002) Basic charge clusters and predictions of membrane protein topology. *J. Chem. Inf. Comput. Sci. 42,* 620-632.

96. Kyte, J., and Doolittle, R. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol. 157*, 105-132.

97. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Prot. Sci. 4,* 2411-2423.

98. Tate, C. G. (2001) Overexpression of mammalian integral membrane proteins

for structural studies. *FEBS Lett.504*, 94-98.

99. Mayumi, I., Nobuo, M., Obita, T., Kamishikiryo, J., Maenaka, K., and Kohda, D. (2007) Purification, crystallization and the preliminary X-ray diffraction studies of the soluble domain of the oligosaccharyltransferase STT3 subunit from the thermophilic archaeon *Pyrococcus furiosus. Acta. Crystallogr. Sect. F. Struc.t Bio.l Crys.t Commun. 63*, 798-801.
100. Rogl, H., Kosemund, K., Kuhlbrandt, W., and Collinson, I. (1998) Refolding of *Escherichia coli* produced membrane protein inclusion bodies immobilised by nickel chelating chromatography. *FEBS Lett. 432*, 21–26.
101. Gorzelle, B. M., Nagy, J. K., Oxenoid, K., Lonzer, W. L., Cafiso, D. S., and Sanders, C. R. (1999) Reconstitutive refolding of diacylglycerol kinase, an integral membrane protein. *Biochemistry 38*, 16373–16382.
102. Baneres, J. L., Martin, A., Hullot, P., Girard, J. P., Rossi, J. C., and Parello, J. (2003) Structure-based Analysis of GPCR Function: Conformational Adaptation of both Agonist and Receptor upon Leukotriene B_4 Binding to Recombinant BLT1. *J. Mol. Biol. 329*, 801–814.
103. Page, R. C., Moore, J. D., Nguyen, H. B., Sharma, M., Chase, R., Gao, F. P., Mobley, C. K., Sanders, C. R., Ma, L., Sönnichsen, F. D., Lee, S., Howell, S. C., Opella, S. J., and Cross, T. A. (2006) Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. *J. Struct. Func. Genom. 7,* 51-64.
104. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A.

(1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR. 6,* 277-293.

105. Rosinke, B., Strupat, K., Hillenkamp, F., Rosenbusch, J., Dencher, N.; Kruger, U., and Galla, H. (1995) Matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) of membrane proteins and non-covalent complexes. *J. Mass Spectrom. 30*, 1462-1468.

106. Galvani, M., and Hamdan, M. (2000) Electroelution and passive elution of γ-globulins from sodium dodecyl sulphate polyacrylamide gel electrophoresis gels for matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom. 14*, 721-723.

107. Jeannot, M. A., Jing, Z., and Li, L. (1999) Observation of sodium gel-induced protein modifications in dodecylsulfate polyacrylamide gel electrophoresis and its implications for accurate molecular weight determination of gel-separated proteins by matrix-assisted laser desorption ionization time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom*. *10,* 512-520.

108. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., and Bairoch, A. (2003) ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res. 31*, 3784–3788.

109. Chatterjee, S., Schoepe, J., Lohmer S., and Schomburg, D. (2005) High level expression and single-step purification of hexahistidine-tagged L-2-hydroxyisocaproate dehydrogenase making use of a versatile expression

vector set. *Protein Expr. Purif., 39*, 137-143.

110. Demarest, S. J., Boice, J. A., Fairman, R., and Raleigh, D. P. (1999) Defining the core structure of the α-lactalbumin molten globule state. *J. Mol. Biol. 294,* 213-221.

111. Batenjany, M. M., Mizukami, H., and Salhany, J. M. (1993) Near-UV circular dichroism of band 3. Evidence for intradomain conformational changes and interdomain interactions. *Biochemistry. 32,* 663-668.

112. Taylor, R. M., Zakharov, S. D., Bernard, H. J., Girvin, M. E., and Cramer, W. A. (2000) Folded state of the integral membrane colicin E1 immunity protein in solvents of mixed polarity. *Biochemistry. 39,* 12131-12139.

113. Turk, E., Gasymov, O. K., Lanza, S., Horwitz, J., and Wright, E. M. (2006) A reinvestigation of the secondary structure of functionally active vSGLT, the vibrio sodium/galactose cotransporter. *Biochemistry. 45,* 1470-1479.

114. Ladokhin, A. S., Jaysainghe S., and White, S. H. (2000) How to Measure and Analyze Tryptophan Fluorescence in Membranes Properly, and Why Bother? *Anal. Biochem*. *285*, 235-245.

115. Reshetnyak, Y. K., Koshevnik, Y., and Burstein E. A. (2001) Decomposition of Protein Tryptophan Fluorescence Spectra into Log-Normal Components. III. Correlation between Fluorescence and Microenvironment Parameters of Individual Tryptophan Residues. *Biophys. J. 81*, 1735–1758.

116. Reithmeier R. A. (1995) Characterization and modeling of membrane sequence analysis. *Curr. Opin. Struct. Biol*. *5,* 491-500.

117. Deber, C. M., and Goto, N. K. (1996) Folding proteins into membranes. *Nat. Struc. Biol. 3,* 815-818.

118. Landolt-Marticorena, C., Williams, K. A., Deber, C. M., and Reithmeier, R. A. (1993) Non-random distribution of amino acids in the transmemrbane segments of human type I single span membrane proteins. *J. Mol. Biol. 229,* 602-608.

119. Eftink, M. R. (1991) Methods of Biochemical Analysis, *John Wiley, New York*, 127–205.

120. Lakowicz, J. R. (1999) Principles of Fluorescence Spectroscopy, *Kluwer-Plenum, New York.*

121. Vivian, J. T., and Callis, P. R. (2001) Mechanisms of Tryptophan Fluorescence Shifts in Protein. *Biophys. J. 80,* 2093-2109.

122. Mayer, M., and Meyer, B. (1999) Characterization of ligand binding by saturation transfer difference NMR spectroscopy. *Angew. Chem. Int. Ed. 38,* 1784–1788.

123. Mayer, M., and Meyer, B. (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor. *J. Am. Chem. Soc. 123,* 6108–6117.

124. Peng, J. W., Lepre, C. A., Fejzo, J., Abdul-Manan, N., and Moore, J. M. (2001) Nuclear magnetic resonance-based approaches for lead generation in drug discovery. *Meth. Enzymol. 338,* 202–230.

125. Stockman, B. J., and Dalvit, C. (2002) NMR screening techniques in drug

discovery and drug design. *Prog. Nucl. Magn. Reson. Spectrosc. 41,* 187–231.

126. Meinecke, R., and Meyer, B. (2001) Determination of the binding specificity of an integral membrane protein by saturation transfer difference NMR: RGD peptide ligands binding to integrin $\alpha_{IIb}\beta_3$. *J. Med. Chem. 44,* 3059–3065.

127. Streiff, J. H., Juranic, N. O., Macura, S. I., Warner, D. O., Jones, K. A., and Perkins, W. J. (2004) Saturation Transfer Difference Nuclear Magnetic Resonance Spectroscopy as a Method for Screening Proteins for Anesthetic Binding. *Mol. Pharmacol. 66,* 929–935.

128. Goto, N. K., Gardner, K. H., Mueller, G. A., Willis, R. C., and Kay, L. E. (1999) A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J. Biomol. NMR. 13,* 369-374.

129. Kelleher, D. J., Kreibich, G. and Gilmore, R. (1992) Oligosaccharyltransferase activity is associated with a protein complex composed of ribophorins I and II and a 48 kd protein. *Cell 69,* 55–65.

130. Baleja, J. D. (2001) Structure determination of membrane-associated proteins from NMR data. *Anal. Biochem. 288,* 1-15.

131. Chill, J. H., Louis, J. M., Miller, C., and Bax, A. (2006) NMR study of the tetrameric KcsA potassium channel in detergent micelles. *Prot. Sci. 15,* 684-698.

132. Jaroniec, C. P., Kaufman, J. D., Stahl, S. J., Viard, M., Blumenthal, R., Wingfield, P. T., and Bax, A. (2005) Structure and Dynamics of Micelle-

Associated Human Immunodeficiency Virus gp41 Fusion Domain. *Biochemistry 44*, 16167-16180.

133. Howell, S. C., Mesleh, M. F, and Opella S. J. (2005) NMR Structure Determination of a Membrane Protein with Two Transmembrane Helices in Micelles: MerF of the Bacterial Mercury Detoxification System. *Biochemistry 44*, 5196-5206.

134. Lee, S., Mesleh, M. F., and Opella S. J. (2003) Structure and dynamics of a membrane protein in micelles from three solution NMR experiments. J. *Biomol. NMR. 26*, 327–334.

135. Mascioni, A., Porcelli, F., Ilangovan, U., Ramamoorthy, A., and Veglia, G. (2003). Conformational preferences of the amylin nucleation site in SDS micelles: an NMR study. *Biopolymers 69*, 29-41.

136. Clarke, D. M., Loo, T. W., and MacLennan, D. H. (1990) Functional consequences of alterations to amino acids located in the nucleotide binding domain of the Ca^{2+}-ATPase of Sarcoplasmic Reticulum. *J. Biol. Chem. 265*, 22223-22227.

137. Jorgensen, P. L., Hakansson, K. O., and Karlish, S. J. (2003) Structure, Function and Regulation of Na, K-ATPase. *Annu. Rev. Physiol. 65*, 817-849.

138. Sharma, C. B., Lehele, L., and Tanner, W. (1981). N-Glycosylation of Yeast Proteins: Characterization of the Solubilized Oligosaccharyl Transferase. *Eur. J. Biochem. 116*, 101-108.

139. Karaoglu, D., Kelleher, D. J., and Gilmore, R. (2001) Allosteric regulation

provides a molecular mechanism for preferential utilization of the fully assembled dolichol-linked oligosaccharide by the yeast oligosaccharyltransferase. *Biochemistry 40*, 12193-12206.

140. Moseley, H. N., and Montelinone, G. T. (1999) Automated analysis of NMR assignments and structure for proteins. *Curr. Opin. Struct. Biol.* 9, 635-642.

141. Wüthrich, K. (1986) NMR of proteins and Nucleic Acids. *Wiley, New York*.

142. Wishart, D. S., Sykes, B. D., and Richards, F. M. (1992) The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry 31,* 1647-1651.

143. Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C Chemical-Shift Index: A simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. *J. Biomol. NMR. 4*, 171-80.

144. Wagner, G., Pardi, A., and Wüthrich, K. (1983) Hydrogen-bond length and H-1-NMR chemical-shifts in proteins. *J. Am. Chem. Soc. 105,* 5948–5949.

145. Williamson, M. P., and Asakura, T. (1993) Empirical comparisons of models for chemical-shift calculation in proteins. *J. Magn. Reson. B 101,* 63–71.

146. Case, D. A. (1995) Calibration of ring-current effects in proteins and nucleic acids. *J. Biomol. NMR. 6,* 341–346.

147. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR 13*, 289–302.

148. Xu, X. P., and Case, D. A. (2001) Automated prediction of N-15, C-13(alpha), C-13(beta) and C-13' chemical shifts in proteins using a density functional database. *J. Biomol. NMR 21*, 321–333.

149. Shen, Y., and Bax, A. (2007) Protein backbone chemical shifts predicted from searching a database for torsion angle and sequence homology. *J. Biomol. NMR 38*, 289–302.

150. Tugarinov, V., Muhandiram, R., Ayed, A., and Kay, L. E. (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase G. *J. Am. Chem. Soc. 124*, 10025-10035.

151. Gautier, A., Kirkpatrick, J. P., and Nietlispach, D. (2008) Solution-state NMR spectroscopy of a seven-helix transmembrane protein receptor: backbone assignment, secondary Structure, and dynamics. *Angew. Chem. Int. Ed. 47*, 7297 –7300.

152. Johnson, B. A., and Blevins, J. (1994) NMRVIEW: a computer program for the visualization and analysis of NMR data. *J. Biomol. NMR. 4*, 603-614.

153. Shan, X., Gardner, K. H., Muhandiram, D. R., Rao, N. S., Arrowsmith, C. H., and Kay, L. E. (1996) Assignment of ^{15}N, $^{13}C^{\alpha}$, $^{13}C^{\beta}$, and HN resonances in an ^{15}N, ^{13}C, ^{2}H labeled 64 kDa Trp repressor−operator complex using triple-resonance NMR spectroscopy and ^{2}H-decoupling. *J. Am. Chem. Soc. 118*, 6570-6579.

154. Venters, R. A., Farmer, B. T., Fierke, C. A., and Spicer, L. D. (1996)

Characterizing the use of perdeuteration in NMR studies of large proteins: ^{13}C, ^{15}N and ^{1}H assignments of human carbonic anhydrase II. *J. Mol. Biol. 264*, 1101-1116.

155. Manning, M. C.; Patel, K.; Borchard, R. T. (1989) Stability of protein pharmaceuticals. *Pharm. Res. 1989*, 6, 903-918.

156. Aswad, D. W., Paranandi, M. V., and Schuter, B. T. (2000) Isoaspartate in peptides and proteins: formation, significance, and analysis. *J. Pharm. Biomed. Anal.* 21, 1129-1136.

157. Brandts, J. F., Halvorson, H. R., and Brennan, M. (1975) Consideration of the Possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. *Biochemistry 14*, 4953−4963.

158. Lu, K. P., Finn, G., Lee, T. H., and Nicholson, L. K. (2007) Prolyl cis-trans isomerization as a molecular timer. *Nat. Chem. Biol. 3*, 619 – 629.

159. Farmer B. T., and Venters, R. A. (1995) Assignment of side-chain 13C resonances in perdeuterated proteins. *J. Am. Chem. Soc. 117*, 4187–4188.

160. Sanders, C. R., and Sönnichsen, F. (2006) Solution NMR of membrane proteins: practice and challenges. *Magn. Reson. Chem. 44*, 24–40.

161. Grzesiek, S., Anglister, J., Ren, H., and Bax, A. (1993) ^{13}C line narrowing by ^{2}H decoupling in $^{2}H/^{13}C/^{15}N$-enriched proteins. Applications to triple resonance 4D J-connectivity of sequential amides. *J. Am. Chem. Soc. 115*, 4369-4370.

162. Otten, R., Chu, B., Krewulak, K. D., Vogel, H. J., and Mulder, F. A. (2010)

Comprehensive and cost-effective NMR spectroscopy of methyl groups in large proteins. *J. Am. Chem. Soc. 132*, 2952-2960.

163. Karplus, M. (1959) Contact electron-spin coupling of nuclear magnetic moments. *J. Phys. Chem. 30*, 11-15.

164. Bystrov, V. F. (1976) Spin-spin couplings and the conformational states of peptide spin systems. *Prog. Nucl. Magn. Reson. Spectrosc. 10*, 41-81.

165. Hu, J. S., and Bax, A. (1997) Determination of phi and chi(1) angles in proteins from C-13-C-13 three bond J couplings measured by three-dimensional heteronuclear NMR. How planar is the peptide bond? *J. Am. Chem. Soc. 119*, 6360-6368.

166. Kosen, P. A. Spin labeling of proteins. (1989) *Meth. Enzymol. 177*, 86-121.

167. Hubbell, W. L., and Altenbach, C. (1994) Investigation of structure and dynamics in membrane proteins using site-directed spin labeling. *Curr. Opin. Struct. Biol. 4*, 566-573.

168. Solomon, I., and Bloembergen, N., (1956) Nuclear magnetic interactions in the HF molecule. *J. Chem. Phys. 25*, 261-266.

169. Liang, B., Bushweller, J. H., and Tamm, L. K. (2006) Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy, *J. Am. Chem. Soc. 128*, 4389-4397.

170. North, C. L., Franklin, J. C., Bryant, R. G., and Cafiso, D. S. (1994) Molecular flexibility demonstrated by paramagnetic enhancements of nuclear

relaxation. Application to alamethicin: a voltage-gated peptide channel. *Biophys. J.*, *67*, 1861-1866.

171. Shenkarev, Z. O., Paramonov, A. S., Balashova, T. A., Yakimenko, Z. A., Baru, M. B., Mustaeva, L. G., Raap, J., Ovchinnikova, T. V., and Arseniev, A. S. (2004) High stability of the hinge region in the membrane-active peptide helix of zervamicin: paramagnetic relaxation enhancement studies *Biochem. Biophys. Res. Commun. 325*, 1099-1105.

172. Bax, A., and Tjandra, N. (1997) High-resolution heteronuclear NMR of human ubiquitin in an aqueous liquid crysalline medium. *J. Biomol. NMR. 10*, 289-292.

173. Clore, G. M., Starich, M. R., and Gronenborn, A. N. (1998) Measurement of residual dipolar couplings of macromolecules aligned in the nematic phase of a colloidal suspension of rod-shaped viruses. *J. Am. Chem. Soc. 120*, 10571-10572.

174. Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T., and Grzesiek, S. (2000) Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR 18*, 303-309.

175. Jones, D. H., and Opella, S. J. (2004) Weak alignment of membrane proteins in stressed polyacrylamide gels. *J. Magn. Reson. 171*, 258–269.

176. Cierpicki, T., and Bushweller, J. H. (2004) Charged gels as oriented media for measurement of residue dipolar couplings in soluble and membrane

proteins. *J. Am. Chem. Soc. 126,* 16259-16266.

177. Chill, J. H., Louis, J. M., Delaglio, F., and Bax, A. (2007) Local and global structure of the monomeric subunit of the potassium channel KcsA probed by NMR. *Biochim. Biophys. Acta. 1768,* 3260–3270.

178. Tycko, R., Blanco, F. J., and Ishii, Y. (2000) Alignment of biopolymers in strained gels: a new way to create detectable dipole–dipole couplings in high-resolution biomolecular NMR, *J. Am. Chem. Soc. 122*, 9340–934.

179. Meier, S., Haussinger, D., and Grzesiek, S. (2002) Charged acrylamide copolymer gels as media for weak alignment. *J. Biomol. NMR 24*, 351-356.

180. Meiler, J., Blomberg, N., Nilges, and M., Griesinger, C. (2000). A new approach for applying residual dipolar couplings as restraints in structure elucidation. *J. Biomol. NMR 16,* 245–52.

181. Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson. 131*, 373-378.

182. Ottiger, M., and Bax, A. (1998) Characterization of magnetically oriented phospholipid micelles for measurement of dipolar couplings in macromolecules. *J. Biomol. NMR, 12,* 361–372.

183. Salsbury, N.J., Darke A. and Chapman, D. (1972) Deuteron magnetic resonance studies of water associated with phospholipids. *Chem. Phys. Lipids, 8,* 142–151.

184. Cierpicki, T., and Bushweller, J. H. (2004) Charged gels as orienting media

for measurement of residual dipolar couplings in soluble and integral membrane proteins. *J. Am. Chem. Soc. 126*, 16259-16266.

185. Ulmer, T. S., Ramirez, B. E., Delaglio, F., and Bax, A. (2003) Evaluation of backbone proton positions and dynamics in a small protein by liquid crystal NMR spectroscopy. *J. Am. Chem. Soc. 125*, 9179-9191.

186. Al-Hashimi, H. M., Valafar, H., Terrell, M., Zartler, E. R., Eidsness, M. K., and Prestegard, J. H. (2000) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *J. Magn. Reson. 143*, 402-406.

187. Clore, G. M., Starich, M. R., Bewlwy, C. A., Cai, M., and Kuszewski, J. (1999) Impact of residual dipolar couplings on the accuracy of NMR structures determined from a minimal number of NOE restraints. *J. Am. Chem. Soc. 121*, 6513-6514.

188. Clore, G. M, Gronenborn, A. M, and Bax, A. (1998) A robust method for determining the magnitude of the fully asymmetric alignment tensor of oriented macromolecules in the absence of structural information. *J. Magn. Reson. 133*, 216–21.

189. von Heijne, G. (1992) Membrane protein structure prediction. *J. Mol. Biol. 255*, 487–494.

190. Wallin, E. and von Heijne, G. (1998) Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Prot. Sci. 7*, 1029–1038.

191. Jones, D. T., Taylor, W. R. and Thornton, J. M. (1994) A model recognition approach to the prediction of all-helical membrane protein structure and topology. *Biochemistry 33*, 3038–3049.

192. Rost, B., Fariselli, P. and Casadio, R. (1996) Topology prediction for helical transmembrane proteins at 86% accuracy. *Prot. Sci. 4*, 521–533.

193. Sonnhammer, E., von Heijne, G. and Krogh, A. (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. *Intell. Syst. Mol. Biol. 6*, 175–182.

194. Nilsson, J., Persson, B., and von Heijne, G. (2002) Prediction of partial membrane protein topologies using a consensus approach. *Prot. Sci. 11*, 2974–2980

195. Amico, M., Finelli, M., Rossi, I., Zauli, A., Elofsson, A., Viklund, H., von Heijne, G., Jones, D., Krogh, A., Fariselli, P., Martelli, P. L., and CasadioAmico, R. (2006) PONGO: a web server for multiple predictions of all-alpha transmembrane proteinsNucl. *Acids Res. 34*, 169–172

196. Tusnady, G. E., and Simon, I. (1998) Principles governing amino acid composition of integral membrane proteins: application to topology prediction. *J. Mol. Bio. 283*, 489–506.

197. Van Geest, M., and Lolkema, S. J. (2000) Membrane topology and insertion of membrane proteins: search for topogenic Signals. *Microbiol. Mol. Biol. Rev*, *64*, 13–33.

198. Hilty, C., Wider, G., Fernandez, C., and Wuthrich, K. (2004) Membrane

protein-lipid interactions in mixed micelles studied by NMR spectroscopy with the use of paramagnetic reagents. *Chem. Bio. Chem. 5,* 467-473.

199. Beel, A. J., Mobley, C. K., Kim, H. J., Tian, F., Hadziselimovic, A., Jap, B., Prestegard, J. H., and Sanders, C. R. (2008) Structural studies of the transmembrane C-terminal domain of the amyloid precursor protein (APP): Does APP function as a cholesterol sensor? *Biochemistry 47,* 9428–9446.

200. Otero, C., Castro, R., and Soria, J. (1998) Electron paramagnetic resonance studies of spin-labeled fatty acid binding sites in *Candida Rugosa* lipases, *J. Phys. Chem. B. 102,* 8611-8618.

201. Narayan, M., and Berliner, L. J. (1997) Fatty acids and retinoids bind independently and simultaneously to β-Lactoglobulin, *Biochemistry 36,* 1906-1911.

202. Seeliger, M. A., Ranjitkar, P., Kasap, C., Shan, Y., Shaw, D. E., Shah, N. P., Kuriyan, J., and Maly, D. J. (2009) Equally potent inhibition of c-Src and Abl by compounds that recognize inactive kinase conformations, *Cancer Res. 69,* 2384-2392

203. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR. 13,* 289–302.

204. Shen, Y., Delaglio, F., Cornilescu, G., and Bax, A. (2009) TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR. 44,* 213-223.

205. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J. Mol. Graphics 14*, 51-55.

206. Blobel, G. (1980) Intracellular protein topogenesis. *Proc. Natl. Acad. Sci. USA. 77*, 1496–1500.

207. Nilsson, I. M., and von Heijne, G. (1993). Determination of the distance between the oligosaccharyltransferase active site and the endoplasmic reticulum membrane. *J. Biol. Chem. 268*, 5798–5801.

Appendix Table A-1 Backbone chemical shift assignments of the C-terminal domain of Stt3p

Residue Number	Amino Acid	HN	N	CA	CB	CO
	M					
	G					
	S					
	S					
	H					
	H					
	H					
	H					
	H					
	H			56.617	29.539	
His-Tag	S	7.873	115.242	58.603	63.761	174.97
	S	7.847	114.938	61.849	63.878	174.954
	G	8.02	110.528	45.12		173.853
	L	7.779	121.627	55.653	41.852	176.4
	V	7.446	118.014	59.409	31.783	173.9
	P			62.945	30.792	176.5
	R	7.906	120.317	55.822	29.871	177.1
	G	8.077	109.515	45.295		
	S					
	M					
	H			57.176	31.777	
466	S	7.922	114.156	60.069	63.093	175.787
467	T	7.599	115.858	63.716	68.685	175.352
468	W	7.734	122.672	59.018	29.028	177.4
469	V	7.623	117.286	63.726	31.224	176.6
470	T	7.549	113.586	62.767	68.942	175.2
471	R	7.69	121.512	56.719	29.467	176.8
472	T	7.634	112.828	57.991	68.4	174.5
473	A	7.752	124.395	52.59	18.076	177.3
474	Y	7.571	117.787	57.765	37.988	175.6
475	S	7.6	115.86	57.881	63.442	173.67
476	S	7.773	117.245	58.911	63.803	172.915
477	P			63.291	31.117	176.6
478	S	7.925	114.899	58.317	63.062	174.149
479	V	7.491	120.383	61.829	31.993	175
480	V	7.551	121.878	61.375	32.175	175
481	L	7.734	124.315	52.531	40.655	175.5
482	P			63.188	30.991	178.1
483	S	7.842	113.959	58.315	63.826	174.539
484	Q	7.891	120.893	55.373	28.939	176.8
485	T	7.858	117.002	59.317	69.285	176.4
486	P			63.754	30.92	176.8
487	D	7.95	118.035	54.168	40.395	176.764
488	G	7.862	109.173	45.258		174.407
489	K	7.999	120.488	56.888	32.109	174.684
490	L	7.874	120.335	55.544	40.671	177.5
491	A	7.804	124.531	53.318	17.623	178.5
492	L	7.596	118.288	56.941	41.39	178.2
493	I	7.456	117.119	63.479	36.528	177.2
494	D	7.622	120.663	57.005	39.961	178.2
495	D	7.837	119.458	56.073	38.762	178

496	F	8.092	120.518	60.432	38.079	177.1
497	R	8.205	119.337	59.775	29.167	178.1
498	E	7.876	118.065	58.478	28.371	178.6
499	A	7.784	122.184	54.515	17.696	178.7
500	Y	8.067	118.273	57.471	27.542	176.691
501	Y	7.888	118.625	60.44	36.905	177.689
502	W	7.631	123.27	56.2	38.629	175.5
503	L	7.904	121.27	55.521	40.65	176.504
504	R	7.614	119.031	57.142	31.662	176.55
505	M	7.628	118.456	56.912	31.692	176.603
506	N	7.552	118.066	53.277	38.64	174.714
507	S	7.561	115.823	58.215	64.01	174.065
508	D	8.055	122.62	54.049	40.947	176.512
509	E	8.126	120.784	57.023	29.279	176.347
510	D	8.063	119.906	54.221	40.543	176.116
511	S	7.75	115.744	57.983	63.926	174.986
512	K	8.076	124.193	56.887	31.774	177.079
513	V	7.59	118.226	63.322	31.239	176.029
514	A	7.522	123.552	52.974	17.914	178.138
515	A	7.541	120.984	52.968	17.646	178.687
516	W	7.469	118.618	58.261	28.845	176.97
517	W	7.314	119.333	58.165	37.475	128.786
518	D	7.569	120.117	54.992	40.25	176.899
519	Y	7.435	119.062	58.947	37.812	176.844
520	G	7.883	107.835	45.76		174.189
521	Y	7.512	120.192	58.462	37.944	176.024
522	Q	7.707	120.259	55.824	28.593	176.231
523	I	7.696	120.583	58.48	37.486	176.759
524	G	8.016	110.587	45.509		174.915
525	G	7.851	108.483	45.371		174.922
526	M	7.8	119.637	56.071	32.023	176.439
527	A	7.779	122.712	52.768	18.259	177.369
528	D	7.79	117.914	54.018	40.474	176.439
529	R	7.842	120.781	56.387	29.724	176.6
530	T	7.962	113.902	62.72	69.041	174.935
531	T	7.683	115.357	61.934	69.024	174.833
532	L	7.804	123.03	56.134	41.577	177.147
533	V	7.396	115.775	61.883	31.845	175.635
534	D	7.862	122.503	54.059	40.773	176.112
535	N	7.982	119.252	53.845	38.354	175.221
536	N	8.158	118.85	53.973	38.446	175.637
537	T	7.767	114.655	60.995	69.159	175.038
538	W	8.078	122.495	57.529	31.827	177.107
539	N	7.848	118.677	53.82	38.197	176.226
540	N	7.841	117.965	53.47	38.398	175.382
541	T	7.641	113.691	62.479	69.36	174.553
542	H	8.096	119.814	56.26	28.1	176.635
543	I	7.698	119.827	61.759	37.404	175.913
544	A	7.789	125.884	53.282	18.077	178.567
545	I	7.592	117.534	62.705	37.441	176.642
546	V	7.444	119.663	64.129	31.014	177.075
547	G	7.912	108.568	45.977		175.229

548	K	7.645	120.335	57.188	31.782	177.321
549	A	7.759	122.276	53.074	18.077	178.087
550	N	7.686	116.562	55.585	37.412	175.861
551	A	7.852	124.776	52.976	18.037	177.674
552	S	7.846	113.967	59.367	63.668	178.139
553	P			63.668	31.032	177.285
554	E	8.101	119.21	56.958	29.149	176.625
555	E	7.864	119.983	57.344	29.284	176.139
556	K	7.815	121.763	55.681	43.907	176.093
557	S	7.912	116.783	58.216	63.779	174.164
558	Y	7.739	121.134	57.768	37.718	175.923
559	E	7.975	121.984	57.231	28.953	177.296
560	I	7.786	119.5	62.614	37.469	177.141
561	L					
562	K					
563	E			56.229	28.95	178.764
564	H	7.695	117.27	54.716	31.439	177.821
565	D	8.055	120.736	53.815	40.533	176.232
566	V	7.943	119.631	63.32	31.524	175.844
567	D	8.038	121.199	55.535	39.533	177.227
568	Y			60.107	37.922	178.338
569	V	7.632	118.012	65.429	30.727	177.03
570	L	7.584	119.261	57.241	40.4	179.3
571	V	7.505	119.609	65.182	30.855	178.667
572	I	7.505	120.187	63.448	36.642	177.76
573	F	8.022	118.44	59.847	38.235	177.305
574	G	8.137	108.527	46.605		175.786
575	G	7.774	108.527	45.639		175.001
576	L	7.616	121.429	56.313	41.814	177.433
577	I	7.387	115.013	61.361	37.499	176.339
578	G	7.645	109.673	45.328		174.388
579	F	7.663	120.285	57.71	39.139	175.939
580	G	8.173	109.744	44.908		174.599
581	G	7.825	108.617	45.533		174.411
582	D	8.11	120.28	54.784	40.582	176.695
583	D	7.93	119.434	54.984	40.504	176.705
584	I					
585	N					
586	K			58.084	31.437	177.225
587	F	7.652	118.508	59.156	38.482	176.702
588	L			55.714	41.822	176.495
589	W	7.526	116.61	56.773	38.872	174.705
590	N	7.658	124.486	54.635	41.02	171.095
591	I					
592	R			54.6	43.894	175.444
593	I	7.645	120.796	61.029	37.864	174.646
594	S	7.362	113.669	57.082	69.022	172.91
595	E	7.691	117.27	56.487	30.654	173.285
596	G	7.848	108.17	45.249		174.012
597	I	7.436	118.67	61.124	37.877	175.098
598	W	7.631	123.27	54.71	29.067	174.519
599	P					

600	E			59.406	29.031	177.154
601	E	7.54	118.969	57.994	30.826	177.5
602	I	7.234	117.654	65.031	30.902	177.686
603	K	7.502	119.089	58.116	29.081	178.458
604	E	7.667	118.201	62.354	28.635	177.134
605	R			31.538	28.529	175.661
606	D	7.959	121.001	55.108	40.394	175.983
607	F	7.672	119.663	62.145	37.431	176.35
608	Y	7.919	120.057	57.673	38.305	177.079
609	T	7.656	116.327	61.148	69.475	173.844
610	A	7.884	126.63	52.148	18.407	177.464
611	E	8.025	120.181	56.356	29.434	176.895
612	G	7.981	109.687	45.034		173.797
613	E	7.84	120.78	55.989	29.617	175.833
614	Y	7.892	121.72	57.369	38.083	175.294
615	R	7.84	123.596	55.414	30.228	175.715
616	V	7.783	121.212	61.875	31.937	175.533
617	D	8.045	123.867	53.3	40	175.815
618	A	7.857	124.786	52.349	18.47	177.674
619	R	7.982	119.759	56.016	29.714	175.938
620	A	7.986	124.696	52.737	18.174	178.022
621	S	7.898	114.267	58.819	63.552	174.954
622	E	8	122.651	57.223	29.164	177.116
623	T	7.815	114.417	62.82	69.195	175.661
624	M	7.755	122.274	57.16	31.967	178.618
625	R	7.916	119.665	57.977	28.965	176.613
626	N	7.857	117.466	54.394	38.169	176.26
627	S	7.895	116.089	60.022	63.724	175.662
628	L	7.976	122.774	57.274	41.179	177.959
629	L	7.709	117.825	57.023	40.552	178.634
630	Y	7.557	118.065	59.52	37.45	177.343
631	K	7.7	119.387	57.695	30.5	177.555
632	M	7.698	117.275	55.5	31.775	177.545
633	S	7.744	114.866	58.939	63.661	174.845
634	Y	7.521	120.884	58.366	37.462	175.631
635	K	7.394	119.608	55.988	31.466	175.508
636	D	7.616	119.506	53.596	40.534	175.279
637	F	7.715	121.301	57.119	38.173	174.108
638	P			63.736	30.816	176.537
639	Q	7.806	117.523	55.936	27.925	176.594
640	L	7.66	122.061	55.88	41.808	176.676
641	F	7.536	116.192	57.077	38.255	175.663
642	N	7.93	119.217	53.083	38.668	175.967
643	G	8.016	109.196	45.684		174.983
644	G	7.946	108.864	45.256		174.499
645	Q	7.919	119.681	56.248	28.565	176.613
646	A	8.089	123.753	58.5	18.671	178.398
647	T	7.667	110.592	62.6	67.8	175.33
648	D	7.815	121.767	55.598	40.317	176.999
649	R	7.709	119.632	57.23	29.425	177.433
650	V	7.7	119.227	64.055	31.196	176.854
651	R	7.878	120.821	57.572	29.422	177.298

652	Q	7.815	118.114	56.557	28.259	176.444
653	Q	7.717	118.116	55.863	28.467	175.925
654	M	7.773	119.067	55.893	32.436	175.679
655	I	7.523	118.742	60.945	37.78	175.676
656	T	7.69	117.946	59.601	69.266	172.71
657	P			63.801	31.031	177.135
658	L	7.627	120.745	54.705	41.798	176.666
659	D	7.989	120.995	53.992	40.528	175.225
660	V	7.418	119.402	58.972	32.166	173.502
661	P					
662	P			63.44	30.817	178.1
663	L	7.889	119.065	56.274	40.607	178.098
664	D	7.834	118.524	55.528	39.604	177.447
665	Y	7.594	119.682	59.648	37.956	176.649
666	F	7.672	117.539	59.178	38.286	176.102
667	D	7.789	119.162	55.985	40.235	177.141
668	E	7.744	118.189	57.341	28.811	177.45
669	V	7.474	118.283	63.895	31.508	176.642
670	F	7.713	119.153	57.985	38.202	176.027
671	T	7.569	111.442	61.555	69.4	175.014
672	S	7.773	117.278	58.916	63.803	174.606
673	E	7.999	121.104	56.544	28.977	176.2
674	N	7.895	118.587	53.651	38.247	176.111
675	W	7.743	121.416	59.444	29.023	177.149
676	N			59.322	40.543	175.091
677	V	7.519	119.96	64	30.917	177.715
678	R	7.494	119.089	58.132	29.077	178.5
679	I	7.686	118.083	63.629	36.578	175.998
680	Y	7.679	119.659	60.258	37.457	177.3
681	Q	7.809	116.772	57.196	27.75	177.253
682	L	7.566	119.781	56.232	41.434	177.672
683	K	7.479	118.299	56.11	31.705	176.834
684	K	7.815	121.765	56.564	31.756	176.812
685	D	7.923	120.526	54.497	40.696	176.286
686	D	7.905	120.319	54.277	40.476	176.16
687	A	7.861	123.578	52.575	18.321	177.815
688	Q	7.929	118.265	55.89	28.539	176.5
689	G	8.017	109.629	45.195		173.985
690	R	7.787	120.771	55.671	30.323	175.833
691	T	8.122	114.259	60.995	70.4	175.025
692	L	8.078	122.498	56.39	40.899	177.639
693	R	7.627	117.579	57.078	29.375	176.775
694	D	7.738	119.605	55.108	40.394	177.352
695	V	7.672	119.665	64.016	31.002	177.226
696	G	8.119	109.613	46.015		174.653
697	E	7.78	120.366	57.154	29.048	177.564
698	L	7.764	121.171	56.026	41.271	177.893
699	T	7.724	112.386	62.767	69.106	175.316
700	R	7.816	121.764	55.234	29.062	176.812
701	S	7.88	115.439	59.276	68.942	174.968
702	S					
703	T			61.934	69.024	174.629

704	K	7.812	122.836	56.078	32.065	176.37
705	T	7.792	115.101	61.641	69.283	174.38
706	R	7.962	123.243	55.856	29.935	176.153
707	R	7.83	121.099	55.596	30.113	175.424
708	S	7.862	115.916	57.666	63.851	173.945
709	I	7.643	122.041	60.548	38.218	175
710	K	7.592	124.276	55.076	32.465	175.023
711	R	7.815	121.767	53.71	29.342	174.452
712	P			62.828	31.116	175.907
713	E	8.084	120.242	55.872	28.699	176.169
714	L	7.733	122.046	54.992	41.752	177.408
715	G	7.986	107.99	45.528		174.231
716	L	7.545	120.129	55.364	41.547	177.066
717	R	7.756	119.898	55.525	29.731	175.186
718	V	7.269	124.317	63.006	32.284	172.694

Appendix Table A-2 Summary of NMR experiments and protein samples prepared for the studies in this dissertation

	Experiment	Protein samples
Backbone Assignment	3D-HNCACB	{^{2}H, ^{13}C, ^{15}N}-triple labeled protein sample
	3D-HN(CO)CACB	
	3D-HNCO	
	3D-HN(CA)CO	
	3D-HNCA	
	3D-HN(CO)CA	
Side-chain Assignment	3D-HBHA(CO)NH	{^{13}C, ^{15}N}-double labeled protein sample
	3D-TOCSY-HSQC	
	3D-HCCH-TOCSY	
	3D-HNHA	
	3D-HCC(CO)NH	{^{2}H (50%), ^{13}C, ^{15}N}-partially triple labeled protein sample
	3D-(H)CC(CO)NH	
	3D-TOCSY-HSQC	
NOE Assignment	3D-^{15}N-NOESY-HSQC	{^{2}H, ^{13}C, ^{15}N}-triple labeled protein sample
		{^{13}C, ^{15}N}-double labeled protein sample
		{^{2}H (50%), ^{13}C, ^{15}N}-partially triple labeled protein sample
	3D-^{13}C-NOESY-HSQC (Aliphatic Region)	{^{13}C, ^{15}N}-double labeled protein sample
	3D-^{13}C-NOESY-HSQC (Aromatic Region)	
	4D-^{13}C, ^{15}N-HSQC-NOESY-HSQC	

Appendix Table A-2 Summary of NMR experiments and Protein Samples Prepared for the Studies in this Dissertation (Continued)

	Experiment	Protein samples
Acceptor Substrate Peptide Binding	[^{1}H, ^{15}N]-HSQC	^{15}N-single labeled protein sample
	STD	ILV-labeled sample
ILV-sample Study	Ile, Leu-(HM)CM(CGCBCA)NH	Methyl Protonated {I(δ_1 only), L($^{13}CH_3$, $^{12}CD_3$), V($^{13}CH_3$, $^{12}CD_3$)} U-[^{15}N, ^{13}C, ^{2}H] sample
	Val-(HM)CM(CBCA)NH	
	Ile,Leu-HM(CMCGCBCA)NH	
	Val-HM(CMCBCA)NH	
	HMCM[CG]CBCA	
	Ile,Leu-HMCM(CGCBCA)CO	
	Val-HMCM(CBCA)CO	
	3D-^{13}C-NOESY-HSQC	
	4D-^{13}C, ^{13}C-HSQC-NOESY-HSQC	
	4D-^{13}C, ^{15}N-HSQC-NOESY-HSQC	
RDC	IPAP-HSQC	^{15}N-single labeled protein sample in different polyacrylamide gel
Topology Study	16-DSA titration	^{15}N-single labeled protein sample
	Gd-DTPA titration	
PRE	[^{1}H, ^{15}N]-HSQC	^{15}N-single labeled mutant protein sample

Appendix Table A-3 RDCs of the C-terminal domain of Stt3p in different media

Negative Gel		Positive Gel		Neutral Gel		Zwitterionic Gel	
Residue	RDC (Hz)	Residue	RDC (Hz)	Residue	RDC (Hz)	Residue	RDC (Hz)
695	-5.23	509	-7.12	590	-6.26	641	-4.74
696	-4.01	533	-6.14	533	-4.32	695	-2.37
509	-3.71	520	-3.71	696	-2.80	512	-1.76
699	-3.04	470	-2.98	468	-2.13	531	-1.09
647	-2.74	644	-2.49	577	-2.13	696	-0.67
718	-2.07	472	-2.07	534	-2.07	718	-0.61
689	-1.70	531	-0.91	531	-2.07	671	-0.43
520	0.18	687	-0.24	509	-1.82	647	-0.30
644	0.85	524	-0.18	470	-1.70	520	0.00
710	0.91	695	0.06	699	-1.58	644	0.18
478	0.97	696	0.36	478	-1.58	699	0.24
687	1.34	671	1.09	695	-1.46	687	0.61
483	1.70	691	1.16	497	-1.40	689	1.28
488	1.82	689	1.22	671	-1.28	524	1.52
671	2.25	602	1.28	524	-1.22	643	2.19
524	2.31	718	1.64	574	-1.16	472	2.37
706	2.92	511	1.76	472	-1.16	691	3.10
547	5.47	699	2.25	580	-0.85	706	3.59
602	8.39	514	2.25	550	-0.73	710	4.20
		647	2.43	483	-0.49	602	7.24
		478	2.80	644	-0.43		
		559	3.10	647	-0.36		
		530	3.16	488	-0.30		
		488	4.07	480	-0.24		
		483	4.62	530	-0.18		
		708	4.93	687	-0.06		
		706	5.47	466	0.06		
		710	8.21	602	0.18		
				559	0.18		
				578	0.30		
				514	0.30		
				520	0.36		
				511	0.36		
				689	0.55		
				544	0.55		
				710	0.67		
				718	0.79		
				706	0.85		
				594	1.09		
				708	1.34		
				517	1.76		
				561	2.31		
				691	2.74		
				681	2.98		
				493	3.28		
				547	3.77		
				544	3.95		
				575	4.44		
				557	4.62		
				576	5.17		
				643	5.84		

Appendix Table A-4 TALOS+ dihedral angle predictions for the C-terminal domain of Stt3p

		Dihydral Angles		Standard Deviation (+/-)				
RESID		PHI	PSI	DPHI	DPSI	COUNT	CLASS	
466	S	-63.008	-33.299	6.052	16.687	10	Good	
467	T	-64.197	-44.081	3.038	5.842	10	Good	
468	W	-63.793	-38.974	7.093	11.291	10	Good	
469	V	-74.073	-30.105	19.243	30.128	10	Good	α1
470	T	-75.077	-33.867	24.912	23.973	10	Good	
471	R	-65.18	-38.809	8.211	10.517	10	Good	
472	T	-70.112	-34.411	19.452	24.348	10	Good	
473	A	-66.108	-33.864	9.314	23.117	10	Good	
474	Y	-96.91	-7.581	23.521	23.169	10	Good	
475	S	-108.76	4.915	16.49	20.237	8	Warn	
476	S	-93.678	127.192	36.46	32.481	10	Good	
477	P	-64.624	-21.387	9.479	14.256	10	Good	
478	S	-82.127	-9.904	15.491	20.215	10	Dyn	
479	V	-96.576	133.311	82.828	40.617	6	Dyn	
480	V	-128.3	137.552	55.886	25.575	10	Good	
481	L	-94.54	129.174	22.025	31.275	10	Good	
482	P	-63.158	151.196	4.614	11.28	6	Warn	
483	S	-85.058	-6.298	14.736	15.281	10	Good	
484	Q	-88.04	137.637	68.746	22.701	9	Warn	
485	T	-79.997	133.766	21.041	15.424	10	Good	
486	P	-59.103	-28.194	7.999	7.413	10	Good	
487	D	-89.141	-0.216	13.63	12.08	10	Good	
488	G	89.48	1.595	12.468	12.521	10	Good	
489	K	-77.435	127.887	56.992	36.363	9	Warn	
490	L	-72.312	-28.904	12.637	22.265	10	Good	
491	A	-60.576	-40.495	4.91	11.22	10	Good	
492	L	-62.623	-38.485	4.491	9.469	10	Good	
493	I	-65.706	-41.231	4.229	3.384	10	Good	
494	D	-62.176	-45.036	2.796	5.885	10	Good	
495	D	-69.889	-36.813	6.596	7.924	10	Good	
496	F	-69.472	-37.436	17.868	10.16	10	Good	
497	R	-59.727	-41.832	4.412	6.915	10	Good	
498	E	-62.39	-42.986	4.617	7.729	10	Good	α2
499	A	-63.718	-40.34	3.9	6.488	10	Good	
500	Y	-67.569	-41.119	7.629	4.391	10	Good	
501	Y	-71.729	-30.48	10.357	11.628	10	Good	
502	W	-94.747	-11.349	22.105	10.913	9	Good	
503	L	-80.082	-11.7	24.552	24.661	9	Good	
504	R	-66.124	-38.157	9.316	13.054	10	Good	
505	M	-70.208	-38.563	21.107	13.515	10	Good	
506	N	-88.833	-18.737	28.541	20.76	10	Good	
507	S	-100.42	144.953	84.994	32.285	6	Warn	
508	D	-70.95	-31.122	10.421	22.058	10	Good	
509	E	-69.17	-31.674	12.994	23.221	10	Good	
510	D	-78.435	-25.899	26.678	24.258	10	Good	
511	S	-69.495	-34.737	9.857	7.106	10	Warn	
512	K	-61.596	-39.637	10.664	14.215	10	Good	
513	V	-61.882	-41.642	6.264	8.415	10	Good	

514	A	-64.533	-43.063	3.489	6.094	10	Good	α3
515	A	-62.43	-41.922	4.233	5.203	10	Good	
516	W	-64.855	-38.838	7.514	10.148	10	Good	
517	W	-63.662	-43.668	4.808	4.688	10	Good	
518	D	-61.329	-41.344	5.808	11.523	10	Good	
519	Y	-78.146	-21.343	14.31	18.471	10	Good	
520	G	-82.165	-30.967	31.42	21.897	7	Warn	
521	Y	-72.527	-32.628	9.789	21.24	10	Dyn	
522	Q	-68.82	-33.593	9.54	12.458	10	Dyn	
523	I	-82.203	-26.176	19.159	32.322	10	Dyn	
524	G	90.026	8.559	12.183	20.873	10	Dyn	
525	G	-69.099	138.365	68.876	35.288	6	Warn	
526	M	-66.535	-38.681	10.382	12.13	10	Good	
527	A	-66.409	-35.965	9.279	23.803	10	Good	
528	D	-70.581	-30.249	10.328	20.707	10	Good	
529	R	-69.487	-33.475	11.139	23.197	10	Good	
530	T	-65.181	-37.557	8.971	22.947	10	Good	
531	T	-73.037	-32.114	17.012	19.36	10	Good	
532	L	-68.735	-29.665	9.389	9.651	10	Good	
533	V	-88.162	-16.535	15.618	19.062	10	Good	
534	D	-109.45	121.761	36.968	19.921	6	Warn	
535	N	-70.155	-29.773	12.179	22.639	10	Good	α4
536	N	-67.008	-40.239	11.721	8.811	10	Good	
537	T	-67.775	-38.483	7.991	7.25	10	Good	
538	W	-71.577	-29.26	8.225	22.061	10	Good	
539	N	-65.237	-38.742	9.487	10.482	10	Warn	
540	N	-67.604	-35.681	9.524	23.858	10	Good	
541	T	-67.418	-41.956	10.045	8.517	10	Good	
542	H	-71.346	-32.121	9.623	21.504	10	Good	
543	I	-65.266	-36.548	9.98	24.956	10	Good	
544	A	-61.841	-38.698	5.641	11.081	10	Good	
545	I	-66.001	-38.62	4.047	11.152	10	Good	
546	V	-76.624	-33.315	20.044	16.591	10	Good	
547	G	-67.421	-25.167	6.782	13.784	9	Warn	
548	K	-63.464	-40.617	5.721	10.195	10	Good	
549	A	-63.321	-43.471	3.708	6.482	10	Good	
550	N	-61.274	-44.112	4.508	5.35	10	Good	
551	A	-67.325	-33.802	6.689	11.837	10	Good	
552	S	-68.533	-38.066	6.146	12.14	9	Warn	
553	P	-60.721	-22.084	8.627	14.909	10	Good	
554	E	-73.923	-13.793	15.278	16.497	10	Good	
555	E	-71.963	-32.798	27.065	19.388	10	Good	
556	K	-81.594	-27.596	25.922	21.085	10	Dyn	
557	S	-87.758	-18.263	27.308	21.21	9	Dyn	
558	Y	-69.383	-32.577	10.189	2.764	5	Dyn	
559	E	-65.978	-32.321	11.488	13.737	10	Good	
560	I	-76.969	-38.026	19.542	19.315	10	Good	
561	L	9999	9999	0	0	0	None	
562	K	9999	9999	0	0	0	None	
563	E	-67.052	-34.812	9.967	33.124	10	Good	

564	H	-77.562	-26.955	18.487	21.904	10	Good	α5
565	D	-69.179	-35.36	9.955	10.869	10	Good	
566	V	-64.345	-41.045	9.014	23.265	10	Good	
567	D	-64.787	-35.477	5.64	10.404	10	Good	
568	Y	-69.764	-38.612	17.205	11.374	10	Good	
569	V	-64.461	-41.906	5.032	7.586	10	Good	
570	L	-63.676	-38.277	5.362	7.949	10	Good	
571	V	-65.394	-46.159	5.715	3.936	10	Good	
572	I	-64.533	-41.588	6.254	8.48	10	Good	
573	F	-67.398	-32.403	8.627	14.4	10	Good	
574	G	-77.747	-21.705	17.772	17.4	10	Good	
575	G	-66.881	-37.642	10.216	9.905	10	Warn	
576	L	-71.35	-31.13	12.381	13.97	10	Good	
577	I	-91.738	-2.053	14.732	23.043	10	Good	
578	G	92.204	7.628	15.756	11.139	10	Dyn	
579	F	-79.435	128.662	55.496	26.047	10	Dyn	
580	G	-116.55	155.925	79.011	39.055	9	Dyn	
581	G	-89.014	158.001	78.33	50.696	8	Warn	
582	D	-66.865	-34.917	10.916	22.432	10	Good	DK Motif
583	D	-73.831	-32.763	19.874	16.439	10	Good	
584	I	9999	9999	0	0	0	None	
585	N	9999	9999	0	0	0	None	
586	K	-63.703	-44.041	4.303	6.749	10	Good	
587	F	-68.793	-30.238	8.488	28.264	10	Good	
588	L	-97.343	-24.171	15.289	21.332	10	Good	
589	W	-139.74	141.307	36.328	11.308	10	Good	β1
590	N	-128.41	131.875	29	13.277	10	Good	
591	I	-118.27	133.763	24.644	15.835	10	Good	
592	R	-134.61	150.458	16.739	12.739	10	Good	
593	I	-103.52	-14.774	23.1	27.401	5	Warn	
594	S	-146.95	162.019	17.796	9.632	10	Good	
595	E	-150.96	160.338	15.061	13.334	10	Warn	
596	G	83.875	13.248	17.219	16.702	10	Good	
597	I	-97.595	129.374	23.631	34.684	9	Warn	
598	W	-96.07	123.082	24.225	36.37	10	Good	
599	P	-66.5	146.35	11.879	9.523	8	Warn	
600	E	-61.736	-37.646	4.869	6.761	10	Good	α6
601	E	-64.933	-46.628	5.609	7.6	10	Good	
602	I	-65.541	-38.729	4.276	6.893	10	Good	
603	K	-62.356	-37.908	5.496	6.083	10	Good	
604	E	-66.528	-31.431	5.361	18.714	10	Good	
605	R	-81.147	-28.992	16.107	18.13	10	Good	
606	D	-93.47	-29.352	28.144	24.446	10	Good	
607	F	-62.092	-38.605	6.711	13.071	10	Good	
608	Y	-83.52	-12.678	15.523	18.829	10	Good	
609	T	-111.27	139.656	32.588	18.832	7	Warn	
610	A	-95.791	134.184	69.586	44.967	9	Dyn	
611	E	-60.919	147.194	59.201	17.183	6	Dyn	
612	G	92.598	-2.766	13.946	16.375	10	Dyn	
613	E	-78.02	137.501	71.408	39.228	10	Dyn	

614	Y	-127.93	137.366	56.11	19.98	10	Dyn	
615	R	-103.95	132.113	63.659	19.421	10	Dyn	
616	V	-97.699	128.403	66.97	24.996	10	Good	
617	D	-101.55	123.179	64.336	27.316	10	Good	
618	A	-74.995	-28.204	18.013	28.198	10	Dyn	α7
619	R	-65.606	-37.504	10.073	7.449	10	Dyn	
620	A	-68.507	-34.73	12.406	23.312	10	Dyn	
621	S	-68.005	-32.948	9.348	8.124	10	Good	
622	E	-67.569	-37.019	10.145	11.766	10	Good	
623	T	-61.069	-40.029	3.765	6.145	10	Good	
624	M	-62.364	-39.439	5.186	10.511	10	Good	
625	R	-62.902	-38.121	4.239	9.064	10	Good	
626	N	-65.058	-39.733	3.901	12.904	10	Good	
627	S	-63.139	-38.844	5.405	12.888	10	Good	
628	L	-63.787	-42.078	4.936	6.926	10	Good	
629	L	-62.208	-37.93	4.233	10.061	10	Good	
630	Y	-66.905	-42.038	13.184	7.658	10	Good	
631	K	-62.785	-37.509	6.11	12.116	10	Good	
632	M	-68.698	-33.723	7.38	17.173	10	Good	
633	S	-65.126	-40.782	9.084	10.476	10	Good	
634	Y	-79.284	-28.024	25.933	26.021	10	Good	
635	K	-89.525	-17.896	23.495	23.766	10	Good	
636	D	-104.28	0.981	19.756	26.169	10	Good	
637	F	-75.772	131.598	17.58	25.332	10	Good	
638	P	-62.913	-26.559	10.498	11.442	5	Warn	
639	Q	-75.727	-18.705	12.869	15.228	10	Good	
640	L	-70.388	-29.988	11.001	11.462	10	Good	
641	F	-106.52	-3.312	16.094	9.551	8	Warn	
642	N	-83.156	136.056	32.141	19.443	9	Dyn	
643	G	93.118	-4.618	15.266	20.332	8	Dyn	
644	G	-68.695	136.418	62.052	36.759	7	Dyn	α8
645	Q	-65.81	-36.8	9.702	6.677	10	Good	
646	A	-61.005	-39.828	5.945	11.441	10	Good	
647	T	-62.681	-45.218	4.191	5.448	10	Good	
648	D	-62.006	-37.81	5.533	13.303	10	Good	
649	R	-61.212	-41.936	6.537	11.17	10	Good	
650	V	-62.857	-43.51	6.74	3.209	10	Good	
651	R	-63.877	-34.028	7.551	16.08	10	Good	
652	Q	-68.671	-38	9.984	12.888	10	Good	
653	Q	-67.927	-39.192	9.294	11.511	10	Good	
654	M	-82.202	-26.086	17.853	22.541	10	Good	
655	I	-107.35	130.696	61.081	29.09	10	Good	
656	T	-85.606	123.829	18.966	22.788	10	Good	
657	P	-63.884	149.776	4.741	11.719	7	Warn	
658	L	-83.867	-6.333	12.06	17.358	7	Warn	
659	D	-94.787	122.854	21.709	25.597	6	Warn	
660	V	-107.56	135.327	20.471	25.121	10	Good	
661	P	-68.265	145.811	7.386	15.625	10	Good	
662	P	-56.587	-31.674	4.197	4.372	10	Good	
663	L	-62.984	-36.164	6.376	14.223	10	Good	

664	D	-62.487	-41.265	8.311	5.64	10	Good	α9
665	Y	-63.433	-38.949	7.051	8.943	10	Good	
666	F	-63.907	-41.046	9.515	11.589	10	Good	
667	D	-60.317	-35.185	6.387	12.692	10	Good	
668	E	-67.778	-33.398	10.317	13.194	10	Good	
669	V	-70.334	-35.575	18.403	12.73	10	Good	
670	F	-78.731	-15.876	12.71	15.641	10	Good	
671	T	-89.331	-23.354	22.012	26.316	9	Warn	
672	S	-70.314	-34.91	18.12	21.101	10	Good	α10
673	E	-67.648	-37.478	9.051	9.067	10	Good	
674	N	-66.761	-40.444	11.399	9.805	10	Good	
675	W	-71.463	-37.2	7.91	10.255	10	Good	
676	N	-87.451	-23.568	21.61	22.367	10	Warn	
677	V	-60.041	-42.986	5.272	7.832	10	Good	
678	R	-62.539	-34.39	6.918	12.446	10	Good	
679	I	-64.889	-35.186	5.062	10.507	10	Good	
680	Y	-63.054	-38.588	8.291	6.237	10	Good	
681	Q	-67.661	-42.404	11.224	7.08	10	Good	
682	L	-74.902	-38.962	19.39	13.909	10	Good	
683	K	-75.988	-24.146	25.871	18.238	10	Good	
684	K	-68.456	-33.95	12.008	23.561	10	Good	
685	D	-70.008	-30.005	11.893	21.321	10	Good	
686	D	-69.047	-34.08	10.592	9.995	10	Good	
687	A	-71.105	-28.93	10.437	21.103	10	Good	
688	Q	-96.844	6.824	14.504	11.92	9	Dyn	
689	G	92.933	-0.82	15.548	19.941	10	Dyn	
690	R	-77.303	131.259	53.919	33.549	10	Dyn	
691	T	-87.553	157.726	15.271	24.033	10	Good	
692	L	-58.853	-36.925	4.779	7.536	10	Good	α11
693	R	-63.968	-40.858	6.017	6.774	10	Good	
694	D	-62.133	-42.77	4.493	11.86	10	Good	
695	V	-61.753	-39.157	5.686	11.363	10	Good	
696	G	-64.788	-37.131	7.791	11.116	10	Good	
697	E	-67.616	-39.595	9.302	9.204	10	Good	
698	L	-67.697	-36.414	7.326	12.877	10	Good	
699	T	-67.951	-29.082	11.996	19.717	10	Good	
700	R	-91.953	-11.872	16.93	16.717	10	Good	
701	S	-88.292	-20.01	20.514	21.745	5	Warn	
702	S	-78.694	-18.434	37.843	31.176	6	Warn	
703	T	-93.631	-30.871	26.058	27.003	10	Dyn	
704	K	-82.794	-20.544	22.928	26.949	8	Dyn	
705	T	-75.485	-37.787	20.998	15.752	7	Dyn	
706	R	-96.491	-5.987	27.717	19.496	7	Dyn	
707	R	-101.81	133.405	70.385	52.321	9	Dyn	
708	S	-102.51	141.7	66.983	28.759	10	Good	β2
709	I	-114.97	130.302	61.15	14.066	10	Good	
710	K	-122.38	159.993	30.628	32.381	10	Good	
711	R	-92.641	124.699	21.467	39.062	10	Good	
712	P	-66.699	151.447	8.836	8.765	8	Dyn	
713	E	-92.87	-0.995	23.831	22.138	7	Dyn	

714	L	-87.099	-2.278	27.753	23.475	9	Dyn	
715	G	77.763	11.514	11.7	20.023	5	Dyn	
716	L	-82.383	-8.634	15.62	18.489	10	Dyn	
717	R	-80.352	141.679	31.229	10.778	10	Dyn	
718	V	9999	9999	0	0	0	None	