## ABSTRACT

ROGERS, CONSTANCE. Protein Labeling Strategies for Improving the Efficiency of Structure Determination by NMR.

(Under the direction of Dr. John Cavanagh)

Rapid and efficient methods for preparing isotopically labeled recombinant proteins and refining solved structures via NMR are presented. The former approach was developed for $^{2}H/^{13}C/^{15}N$ isotopic labeling and specific protonation of the methyl groups of isoleucine, leucine, and valine (ILV) residues of rat brain calbindin $D_{28K}$, a calcium sensor and buffer. This protocol produces cell mass using unlabeled rich media followed by exchange into labeled media at high cell density. Allowing for a short period for growth recovery and unlabeled metabolite clearance, the cells were induced. Additionally, the solution structure of LuxU, a subunit of the quorum sensing circuit of *Vibrio harveyi*, has been refined using residual dipolar coupling (RDC). In slightly anisotropic environments, large one-bond internuclear dipolar interactions no longer average to zero, and therefore can provide information on the average orientation of the corresponding vectors relative to the magnetic field. This ordering was induced by introducing Pf1 filamentous bacteriophage into a solution of $^{13}C^{15}N$ isotopically labeled LuxU.

# Protein Labeling Strategies for Improving the Efficiency of Structure Determination by NMR

by

**Constance A. Rogers**

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Master of Science

Molecular and Structural Biochemistry

Raleigh

2003

APPROVED BY:

| | |
|---|---|
| _____ | _____ |
| John Cavanagh, Ph.D.<br>Chair of Advisory Committee | William L. Miller, Ph.D. |
| _____ | |
| Michael B. Goshe, Ph.D. | |

## DEDICATION

I dedicate this work to my grandmother Sally Williams who always believed in me and has provided perpetual love and personal guidance.

# BIOGRAPHY

Constance Rogers grew up in Bear Grass, North Carolina with her two parents, Henry and Cathy, and older brother, Jason. She joined Dr. John Cavanagh's protein structure research laboratory at the beginning of her senior year in college at North Carolina State University. Upon receiving her Bachelor's degree in Biochemistry in 2002, she chose to stay in Dr. Cavanagh's lab and further develop her skills on protein expression, purification, and labeling strategies. She will receive her Master's degree in Biochemistry in May 2004.

# ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. John Cavanagh for his guidance and advice. His confidence in me has greatly contributed to the success I have achieved at this institution and I am extremely grateful. I would also like to extend my gratitude to my committee members Dr. William Miller and Dr. Michael Goshe for their insights and helpful suggestions. Thanks to my colleagues Douglas Kojetin for providing me with LuxU and calbindin $D_{28K}$ images, Benjamin Bobay for giving me a table of NMR experiments, David Kordys, Patrick McLaughlin, Daniel Sullivan, Michael Deak, and Keriann Paul for their cooperation, constructive criticism, and comity. Finally, I would like to thank my lab supervisor Richele Thompson for helping me become the person I am today. Her guidance, friendship, and trust have extended beyond science and have left a permanent mark on my mind, heart, and soul. For that, I am eternally grateful.

# TABLE OF CONTENTS

Page

## LIST OF FIGURES

Page

## LIST OF TABLES

## BASIC NUCLEAR MAGNETIC RESONANCE THEORY

Determining the three dimensional structure of biological macromolecules, such as proteins, is critical in understanding their function. Presented here are some basic aspects of Nuclear Magnetic Resonance (NMR) that are essential for understanding this method as it is used for structure determination (1-4). The basis of the NMR phenomenon is a property of the nucleus: the spin. Nuclei that possess a spin of ½ are considered "NMR active". Spin ½ allows for two different nuclear spin states: spin up and spin down. For spin ½, there is a magnetic moment that can be interpreted as a magnetic dipole. When exposed to an external electromagnetic field, these dipoles can only orient parallel or antiparallel to the field. These two orientations correspond to slightly different energies. The spins are allowed to move from one energy level to the other, absorbing or emitting the energy difference in the form of electromagnetic radiation. Even in a strong magnetic field, the difference between the number of spins in the two energy states is extremely small. At the highest field strengths used for NMR, for 1 million hydrogen ($^1$H) nuclei (at room temperature), the difference in the number of atoms populating the two orientations is only by about 60 (5). This small difference between the number of parallel and antiparallel spins is what gives rise to the NMR signal and is precisely why NMR is an insensitive technique. The small imbalance of spin orientations produces a small polarization of the nuclear spins in the sample, resulting in a net macroscopic magnetization. By irradiating the electromagnetic field, this magnetization can be manipulated in a variety of ways (2,3,6,7). The frequency of the irradiation must match the energy difference between the two possible states of the spins and typically lies in the radio frequency (*rf*) range,

between 50 and 900 MHz. The irradiation is usually applied only for a few microseconds as *rf* pulses. After a pulse (or pulses) causes a perturbation of the spins equilibrium state an NMR signal can be observed. The signal consists of *rf* waves with frequencies that match the energy difference between the spin-up and spin-down states of the individual nuclei involved. Over time the system will to return to its equilibrium state, the signal will decay (relax) and typically vanishes within 100 ms. As the sample returns to its equilibrium state, the free induction decay (FID), is recorded. The FID consists of a sum of decaying cosine waves whose frequencies represent the resonance frequencies of the nuclei in the sample. From this data, the NMR frequency spectrum is then obtained by a Fourier transformation (Figure 1).

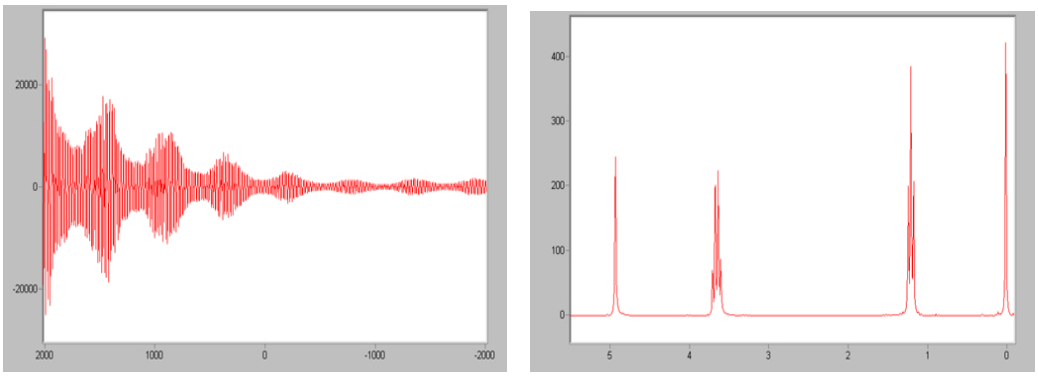



**Figure 1.** A) FID of ethanol. The signal oscillates due to the fact that the detector is set at a particular precession frequency and the actual frequency (or frequencies for a sample with more than one type of proton) of precession will differ somewhat. B) Fourier transformed FID of ethanol. Fourier transform converts the FID in the time domain to the conventional NMR spectrum in the frequency domain.

In an NMR spectrum, the nuclei are represented by characteristic resonance frequencies. For example, protons are represented in a proton resonance spectrum or carbon atoms in a carbon spectrum. The resonance frequencies of different types of nuclei are widely different. For example, protons ($^{1}H$) resonate at a ten times higher frequency than nitrogen nuclei ($^{15}N$) and four times higher than carbon nuclei ($^{13}C$).

The resonance frequencies of different nuclei of the same type lie in a much narrower frequency range. For example, the resonance lines for different protons in a molecule vary in a range of a few parts per million (ppm) around the average proton resonance frequency. This variation is due to the interaction between the nuclei and surrounding electrons, which affect the local magnetic field experienced by a particular nucleus and thus influence its resonance frequency. For example, the resonance frequencies of amide protons, α-protons, or methyls are quite distinct within the proton spectrum

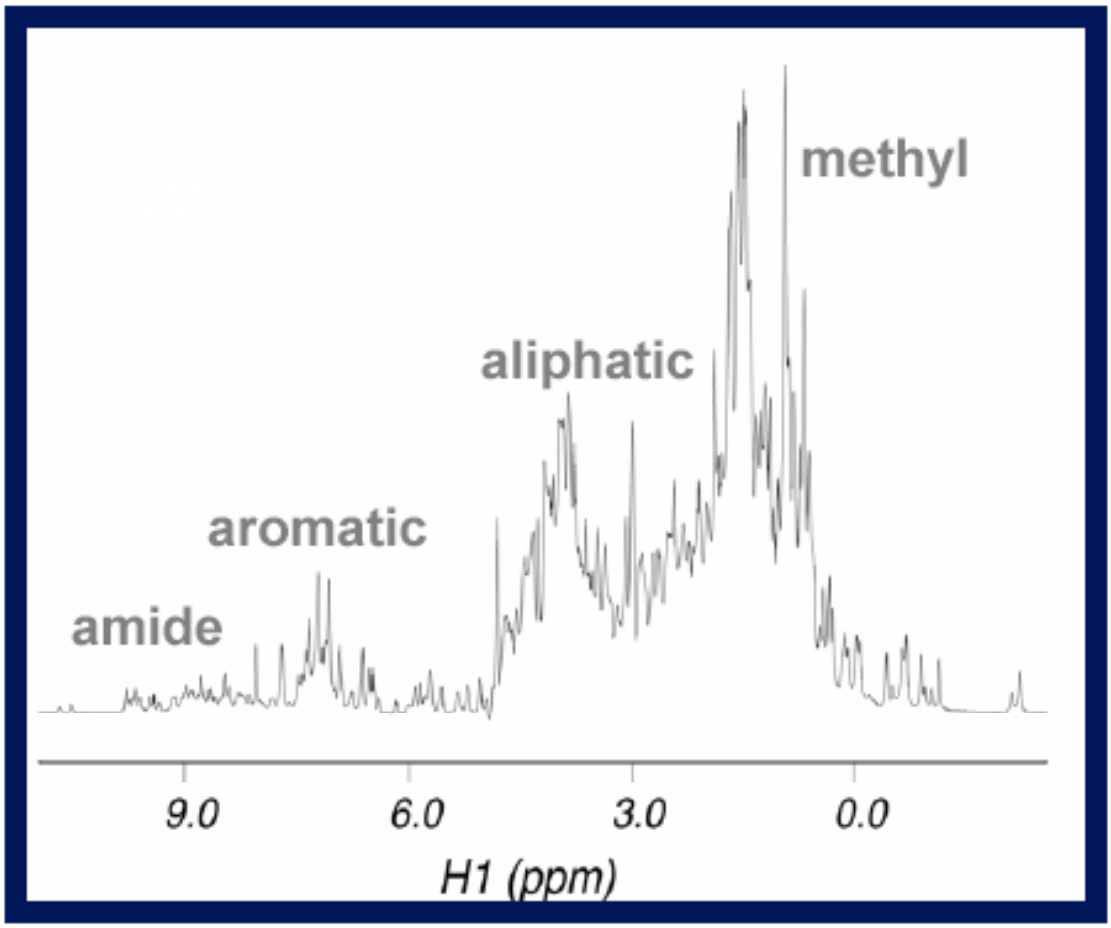



**Figure 2.** The 1-D NMR spectra of Ubiquitin contains numerous overlapping signals making it impossible to assign resonances uniquely. The local environment of a proton modulates its chemical shift: differences arise from both covalent and conformational effects.

and allow a tentative assignment to the different classes of protons (Figure 2). This figure shows the different classes of protons: amide, aromatic, aliphatic, and methyl. The chemical shift is very sensitive to many structural, electronic, magnetic, and dynamic variables and contains a wealth of information on the state of the system under investigation.

Correlations in NMR

A fundamental element employed in NMR spectroscopy is that the magnetic moments of the individual nuclei interact with the small magnetic fields created by the spins of nearby nuclei. This interaction can be used to correlate different atoms in a molecule with one another. Nuclei interact either “indirectly”, through a chemical bond, or “directly”, through space. Interactions between nuclei, transmitted through the electrons that participate in covalent bonds, are known as spin-spin coupling or *J* coupling. Interactions via electrons that “feel” each other through space are the basis for the nuclear Overhauser effect (NOE), which ultimately allows distance measurements between protons. The detailed analysis of protein spectra is based on these correlations.

Through-bond correlations group individual spins into spin systems. Nuclei between which spin-spin interactions exist are known as spin systems. These spin systems are used in the analysis of the spectra. A protein of N residues has N distinct backbone based spin systems. The N spin systems are assigned an amino acid type or one of several possibilities based on the coupling and chemical shift. In proteins, *J* couplings over more than three chemical bonds are usually too small to be observed. Consequently, only spin systems within neighboring atoms in the individual amino

acids can be obtained in proton spectra. In proteins that are isotopically labeled with $^{15}N$ and $^{13}C$, *J* couplings between $^{1}H$, $^{15}N$ and $^{13}C$ nuclei allow through bond correlations across the peptide bond. NMR experiments that correlate different nuclei via *J* coupling are often referred to as correlation spectroscopy (COSY)-type experiments (3,8,9). An important feature of COSY-type experiments is that they can transfer magnetization between different types of nuclei. This property makes it possible to start an experiment with one type of nucleus and to transfer the magnetization to another or several other types of nuclei in the course of the experiment. Such magnetization transfers are of great practical importance because widely different sensitivities are observed with different types of nuclei.

Many protein NMR experiments begin with proton magnetization and transfer the signal via so-called heteronuclei (carbon and/or nitrogen) back to protons for the recording of the FID. Through space correlations provide the basis for geometric information required to determine the structure of a macromolecule and are measured via the NOE (3,10). The NMR method for protein structure determination relies on a dense network of distance constraints derived from NOE's between nearby hydrogen atoms in the protein (4). NOE's connect pairs of hydrogen atoms separated by less than 5 Å. In contrast to COSY-type experiments, the nuclei involved in the NOE correlation may be far apart along the protein sequence but close in space. The NOE is the transfer of magnetization between nuclei interacting via their associated dipoles. For molecules with a molecular weight of more than 5 kDa, the intensity of an NOE is approximately proportional the molecular weight and to $r^{-6}$, where *r* is the distance between the two interacting spins. Dependence on the inverse sixth power of

the distance between the nuclei causes the NOE intensity to diminish rapidly with increasing distance. This is why NOE's between protons in a protein that are separated by more than 5 Å are typically not observed. NMR experiments that measure the NOE are referred to as NOE spectroscopy (NOESY) experiments (9,11).

Multidimensional NMR

NMR spectra of biological macromolecules may contain thousands of signals (resonance lines) that cannot be resolved in a conventional one dimensional (1-D) NMR experiment. However, for a detailed structural analysis, resolution of these signals is essential. Multidimensional NMR spectra provide both increased resolution and correlations that are easy to analyze. This requires more complex NMR experiments that consist of a series of *rf* pulses separated by short delays during which no pulses are applied. This pulse sequence is then followed by the recording of the resulting magnetization.

The process for generating multidimensional NMR experiments is the extension of the experiment from one to two dimensions and of a straightforward combination of two dimensional (2-D) experiments. All 2-D NMR experiments use the same basic scheme, which consists of the following four, consecutive time periods: excitation - evolution – mixing - detection.

During the excitation period, the spins are irradiated to obtain the desired state from which the chemical shifts of the individual nuclei will be observed during the evolution period, $t_1$. In the mixing period, the spins are correlated with each other, and the information on the chemical shift (unique frequency) of one nucleus ends up on another nucleus of which the frequency is measured during the detection period, $t_2$.

To obtain a 2-D data set, a number of experiments are recorded with progressive incrementations of the evolution period $t_1$ to generate a data matrix $S(t_1,t_2)$. A 2-D Fourier transformation of $S(t_1,t_2)$ yields the desired 2-D frequency spectrum $S(\omega_1,\omega_2)$. A resonance in the 2-D spectrum, a peak, represents a pair of nuclei that suitably interact during the mixing time. Figure 3 presents an example of a 2-D NMR spectrum. A 2-D [$^{15}$N,$^{1}$H] correlation spectrum (12) of a $^{15}$N-labeled protein in water solution is shown. The spectrum exhibits correlations between amide nitrogens and the attached protons. Correspondingly, there are two different chemical shift axes, $\omega_1$ and $\omega_2$: one for nitrogens ($^{15}$N) and the other one for protons ($^{1}$H). Each cross peak in

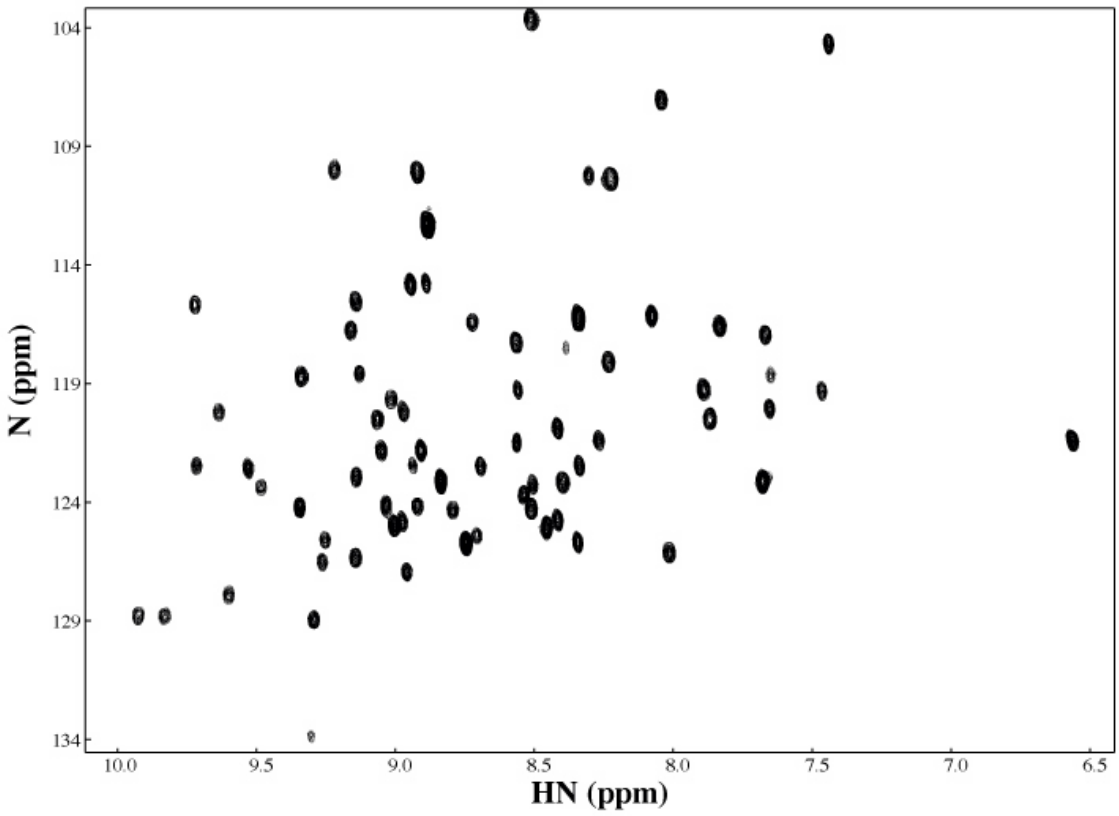



**Figure 3. $^{15}$N** HSQC spectrum of Ubiquitin. HSQC experiments correlate the $^{1}$H chemical shift to the chemical shift of $^{15}$N or $^{13}$C that is connected by a single bond by transferring magnetization through large scalar couplings.

the spectrum of Figure 3 represents one NH-group in the backbone of the polypeptide chain. All amino acids except proline have such an amide group.

The extension from a 2-D to an *n*-dimensional (*n*-D) NMR experiment consists of the combination of (*n*-1) 2-D experiments. These result in a pulse sequence comprising (*n*-1) independently incremented evolution periods and (*n*-1) mixing periods and one detection period. Only during the detection period the signal is physically measured, and this period is thus often referred to as the direct dimension in contrast to the evolution periods, which are referred to as indirect dimensions. To obtain an *n*-D data matrix, $S(t_1, t_2, ..., t_n)$, the pulse sequence is repeated many times with incremental increased or decreased lengths of evolution periods. An *n*-D Fourier transform of $S(t_1, t_2, ..., t_n)$ provides an *n*-D spectrum that depends on *n* frequency variables, $s(\omega_1, \omega_2, ..., \omega_n)$. If two nuclei suitably interact with each other in a mixing time, then this interaction will be shown by a resonance in the spectrum, a cross peak, at a position characterized by the resonance frequencies of the interacting nuclei. For example, in the schematic drawing of Figure 4, the nuclei $^1$H and NH interact in one mixing time and $^1$H and $^{15}$N in the other mixing time of a 3-D experiment. Correspondingly, one peak in the 3-D spectrum appears at the position indicated by the chemical shifts of the three nuclei.

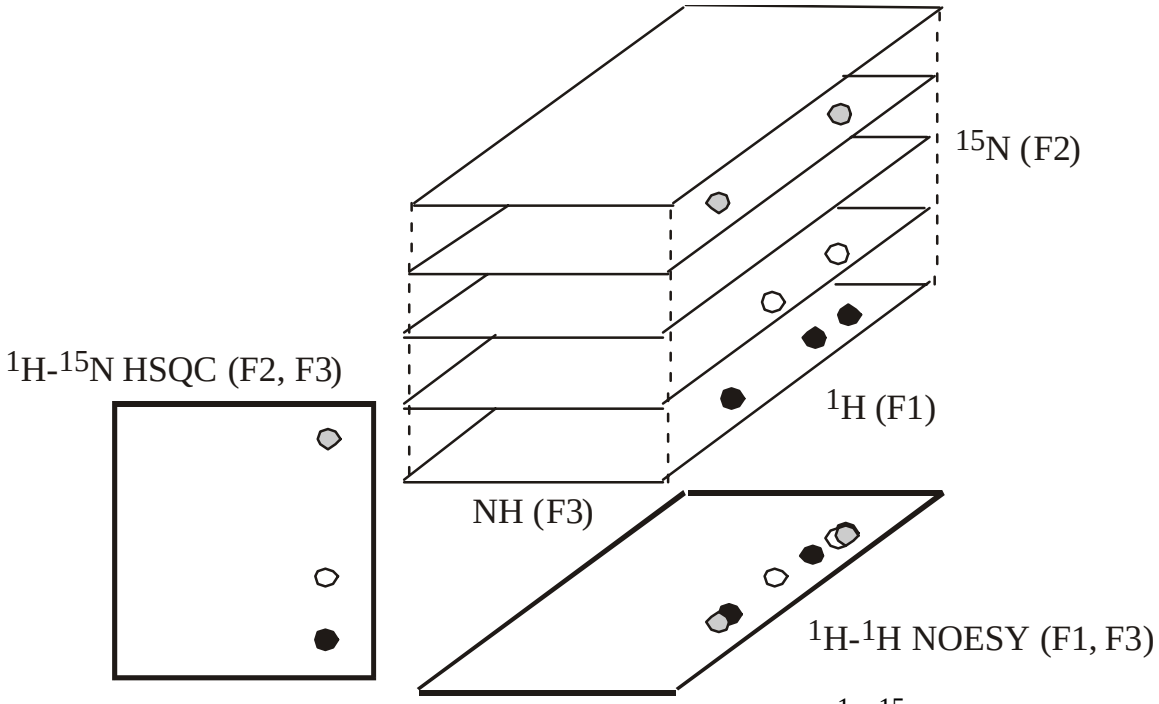



**Figure 4.** A schematic illustration showing the relation between a 3-D $^{1}$H-$^{15}$N NOESY-HSQC experiment and 2-D $^{1}$H-$^{1}$H NOESY (F1, F3) and 2-D $^{1}$H-$^{15}$N HSQC (F2, F3) spectrum.

Figure 4 illustrates the increasing resolution obtained when going from 2-D to 3-D spectra. A 3-D spectrum can be obtained by correlating the amide groups with the $^{15}$N nuclei. The chemical shifts of these $^{15}$N nuclei are used to spread the resonances from the 2-D plane into a third dimension. In Figure 4, the 3-D spectrum is represented by a stack of 2-D planes that corresponds to the actual digitization of the data in the $^{15}$N dimension. The same representation is used to analyze 3-D spectra. Two frequency axes are chosen to define 2-D planes that are stacked in the third dimension.

NMR spectroscopy with biological macromolecules

Most protein NMR experiments are measured with three dimensions. The measurements almost always use protons ($^{1}$H) and, depending on the isotope labeling, $^{13}$C and/or $^{15}$N nuclei. The sensitivity obtainable with these types of nuclei greatly varies even if the sample is fully isotope-labeled with $^{13}$C or $^{15}$N. The proton offers

the best sensitivity and hence constitutes the preferred nucleus for detection of the NMR signal. The other nuclei are usually measured during evolution periods of multidimensional NMR experiments, and their information is transferred to protons for detection.

For the necessary detailed analysis of NMR spectra, hundreds or even thousands of resonance lines must be observed as separate peaks. For protein spectra, this is only possible in multidimensional NMR spectra that are measured on NMR instruments operating with strong magnetic fields corresponding to a proton resonance frequency of 600 MHz or higher. Many NMR measurements with biological macromolecules rely on the presence of exchangeable protons in the molecules and must therefore be performed in water solution. There is a whole family of schemes that are used for the assignment of resonances to individual atoms in the protein. These are shown in Table 1 and the connections occur via the heteronuclear couplings shown in Fig. 5.

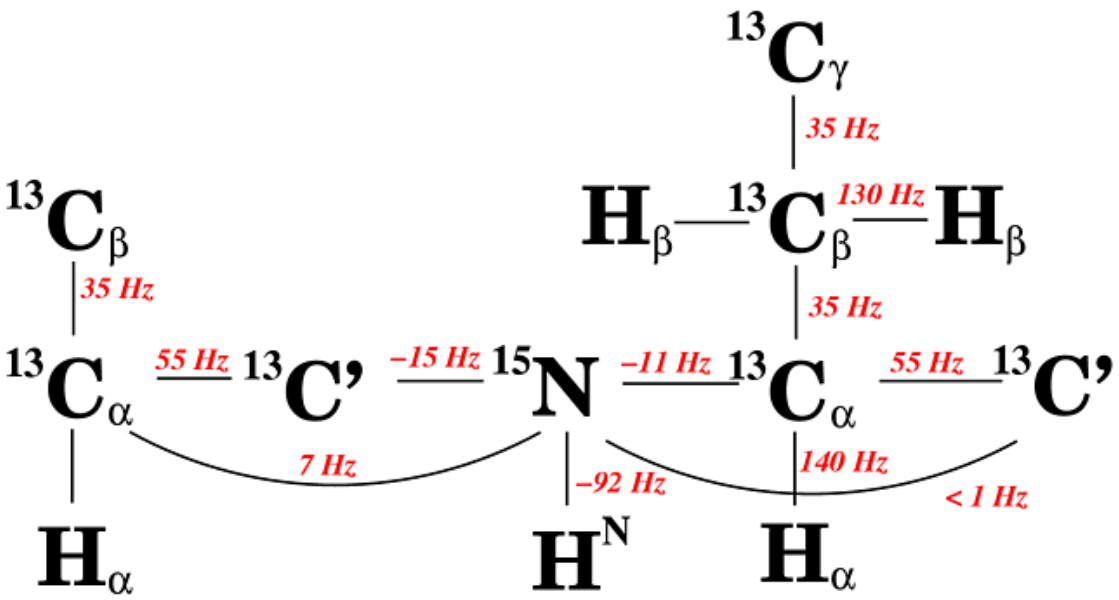
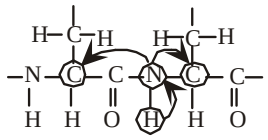
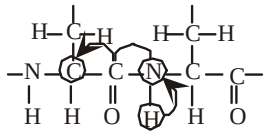
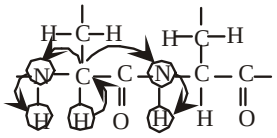
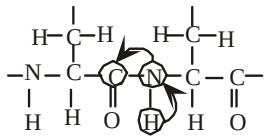
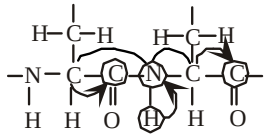
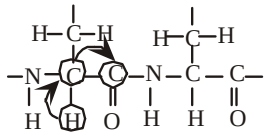
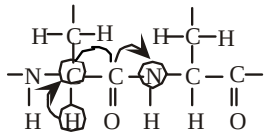
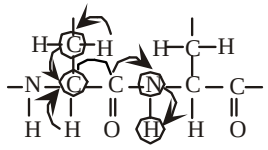
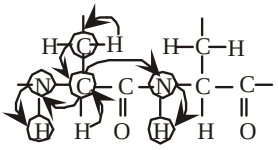



**Figure 5**. Scalar coupling constants for proteins. The distance of the lines in a multiplet (in Hz) is given by the coupling constant J which is independent of the magnetic field strength. Therefore, it is possible to derive parts of the molecular geometry from the knowledge of coupling constants.

**Table 1**: Triple resonance experiments for resonance assignment of $^{15}N$, $^{13}C$-labeled proteins.

| Experiment | Correlations observed | Magnetization transfer | J Couplings |
|---|---|---|---|
| HNCA | $^{1}H_i^N - {}^{15}N_i - {}^{13}C_i^{\alpha}$<br>$^{1}H_i^N - {}^{15}N_i - {}^{13}C_{i-1}^{\alpha}$ |  | $^{1}J_{NH}$<br>$^{1}J_{NC\alpha}$<br>$^{2}J_{NC\alpha}$ |
| HN(CO)CA | $^{1}H_i^N - {}^{15}N_i - {}^{13}C_{i-1}^{\alpha}$ |  | $^{1}J_{NH}$<br>$^{1}J_{NCO}$<br>$^{1}J_{C\alpha CO}$ |
| H(CA)NH | $^{1}H_i^{\alpha} - {}^{15}N_i - {}^{1}H_i^N$<br>$^{1}H_i^{\alpha} - {}^{15}N_{i+1} - {}^{1}H_{i+1}^N$ |  | $^{1}J_{C\alpha H\alpha}$<br>$^{1}J_{NC\alpha}$ $^{2}J_{NC\alpha}$<br>$^{1}J_{NH}$ |
| HNCO | $^{1}H_i^N - {}^{15}N_i - {}^{13}CO_{i-1}$ |  | $^{1}J_{NH}$<br>$^{1}J_{NCO}$ |
| HN(CA)CO | $^{1}H_i^N - {}^{15}N_i - {}^{13}CO_i$<br>$^{1}H_i^N - {}^{15}N_i - {}^{13}CO_{i-1}$ |  | $^{1}J_{NH}$<br>$^{1}J_{NC\alpha}$ $^{2}J_{NC\alpha}$<br>$^{1}J_{C\alpha CO}$ |
| HCACO | $^{1}H_i^{\alpha} - {}^{13}C_i^{\alpha} - {}^{13}CO_i$ |  | $^{1}J_{C\alpha H\alpha}$<br>$^{1}J_{C\alpha CO}$ |
| HCA(CO)N | $^{1}H_i^{\alpha} - {}^{13}C_i^{\alpha} - {}^{15}N_{i+1}$ |  | $^{1}J_{C\alpha H\alpha}$<br>$^{1}J_{C\alpha CO}$<br>$^{1}J_{NCO}$ |
| CBCA(CO)NH | $^{13}C_i^{\beta} - {}^{13}C_i^{\alpha} - {}^{15}N_{i+1} - {}^{1}H_{i+1}^N$ |  | $^{1}J_{CH}$ $^{1}J_{C\alpha C\beta}$<br>$^{1}J_{NCO}$ $^{1}J_{C\alpha CO}$<br>$^{1}J_{NH}$ |
| CBCANH | $^{13}C_i^{\beta}/{}^{13}C_i^{\alpha} - {}^{15}N_i - {}^{1}H_i^N$<br>$^{13}C_i^{\beta}/{}^{13}C_i^{\alpha} - {}^{15}N_{i+1} - {}^{1}H_{i+1}^N$ |  | $^{1}J_{CH}$ $^{1}J_{C\alpha C\beta}$<br>$^{1}J_{NC}$ $^{2}J_{NC\alpha}$<br>$^{1}J_{NH}$ |

The overall goal is to assign each individual resonance in each NMR spectrum collected to a specific atom in the protein. This is known as chemical shift or resonance assignment. All procedures start with the known amino acid sequence to sequentially connect nuclei of amino acid residues that are neighbors in the sequence. This is known as sequential assignment.

Conformational constraints

At this point, though we have assigned all resonances in the protein, no 3-D information has been generated. For use in structure calculation, geometric conformational information in the form of distances has to be derived from the NMR data. This crucial information comes from NOE measurements that provide distance information between pairs of protons (Figure 6).

The quality of the structure improves with the number of input constraints used in the structure calculation. For high-resolution structure determination, the maximum possible number of NOE constraints must be collected as input for the calculation of the complete 3-D protein structure. In a folded protein, there are many proton pairs with an internuclear distance up to 0.5 nm; thus, NOESY spectra tend to be very crowded with peaks, even if they are measured as 3-D or even 4-D spectra to resolve the proton signals by the chemical shift of the attached $^{13}C$ and/or $^{15}N$ nucleus. It is important to note that such NOE distance constraints between protons will only be accurate if the initial resonance assignments are completely accurate. In practice, as indicated in Figure 7, the determination of proton-proton distances is achieved in several iterations.

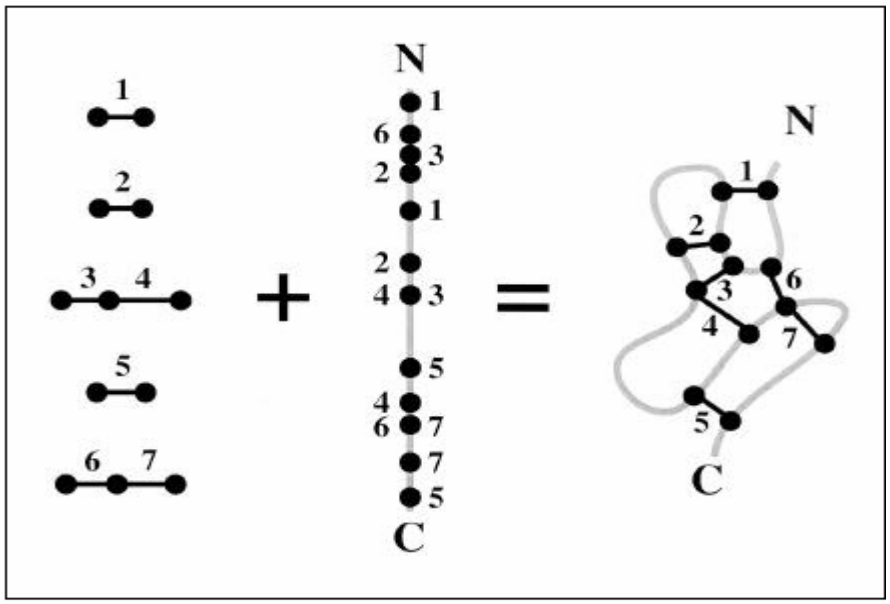



**Figure 6.** Schematic drawing of the use of distance constraints for the determination of the 3-D structure of biological macromolecules. On the left, distances obtained from NOESY spectra for seven pairs of protons are indicated and numbered from 1 to 7. In the middle, the gray vertical line represents the polypeptide chain, and the black dots represent the sequential positions of the individual protons that define the seven distances on the left-hand side of the figure. N and C mark the N and C terminus of the polypeptide chain. The protons are numbered according to the distances they define. Adding the information on the distance between pairs of protons and their position in the polypeptide sequence allows the construction of possible arrangements in which all distance constraints are fulfilled as indicated on the right-hand side of the figure.

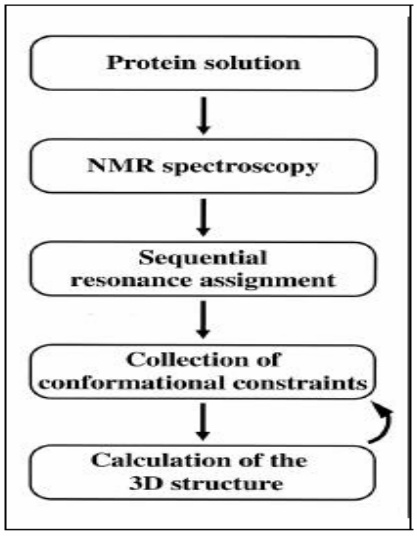



**Figure 7.** Strategy of structure determination by NMR. This is an outline of the general strategy used to solve the 3-D structure of biological macromolecules in solution by NMR.

Once a low-resolution structure has been calculated from an unambiguous subset of the NOE data, it is usually possible to discriminate between several structural possibilities, especially if some prospective proton pairs are not within 0.5 nm. In this way, successively more of the originally ambiguous distances can be assigned to specific pairs of protons, which leads to improved 3- D structures (Figure 8).

These are the traditional methods of acquiring three dimensional protein structures via NMR. However, despite their ubiquitous usage, these methods sometimes prove to be inadequate. This is especially true for large proteins and helical proteins, which

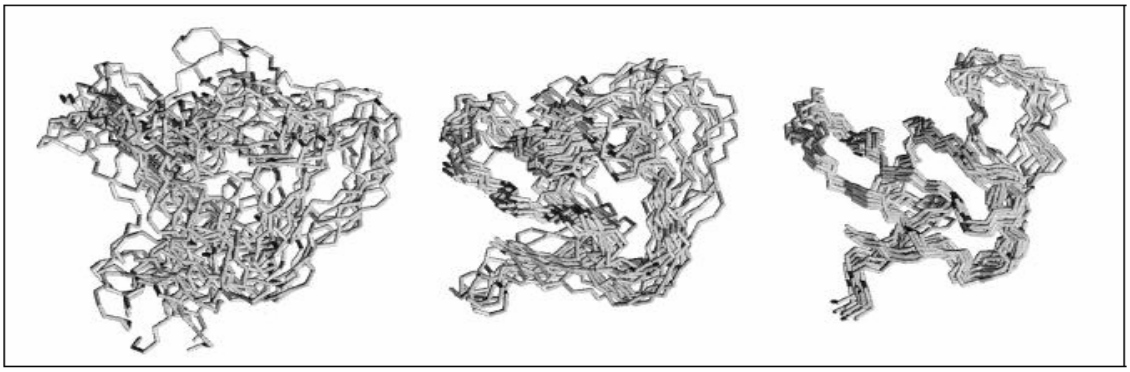


**Figure 8.** Protein backbone structures calculated with different numbers of NMR constraints. The structures show the $SH_3$ domain of human p56 Lck tyrosine kinase (26) at various stages of the assignment of additional distance constraints on the basis of preliminary structures. Backbone superpositions of ten conformers are shown with 1113 constraints (left), 1336 (middle) and 1687 constraints (right).

generate too many signals to be properly resolved in a single spectrum. Larger proteins tend to tumble more slowly in solution than small proteins, which results in poor sensitivity in correlation experiments due to shorter transverse relaxation times; in essence all the NMR signal disappears during the pulse sequence before it is able to be detected. To circumvent these problems, a method involving traditional $^{15}N$ and $^{13}C$ isotopic labeling and perdeuteration of an entire protein coupled with selective

protonation has been developed. Incorporating deuterium ions into the protein significantly slows transverse relaxation rates, but significantly reduces the number of signals generated in correlation spectra. Selective protonation of methyl groups increases the chemical shift dispersion and reintroduces signals in the spectra. The protonation of methyl groups is strategically significant because it ensures more connectivities will be present due to their abundance in the hydrophobic core of proteins. The most widely used residues for this purpose are isoleucine, leucine, and valine (ILV).

Helical proteins also generate overcrowded spectra due to the excessive number of signals present within α-helices, but the loops that connect α-helices generate obscure peaks in NOESY spectra because of a lack of interresidue contacts. This problem can be solved by obtaining residual dipolar couplings (RDC's), which takes advantage of slight molecular alignment present in anisotropic solutions. The resulting alignment imposes measurable perturbations to scalar couplings as measured in non-decoupled spectra. The couplings between dipolar and scalar couplings can be separated by measuring the splittings at different field strengths and the scalar couplings can be removed. The left-over dipolar couplings are angle dependent and in anisotropic media, they provide structural restraints that do not rely on neighboring nuclei to define their position and orientation (Figure 9). In the equation at the bottom, γ is the gyromagnetic constant, which is specific to each nucleus, *h* is Plank's constant, r is the distance between the interacting nuclei, θ measures the angle between the internuclear vector and the main field, $B_0$, and D is the dipolar splitting.

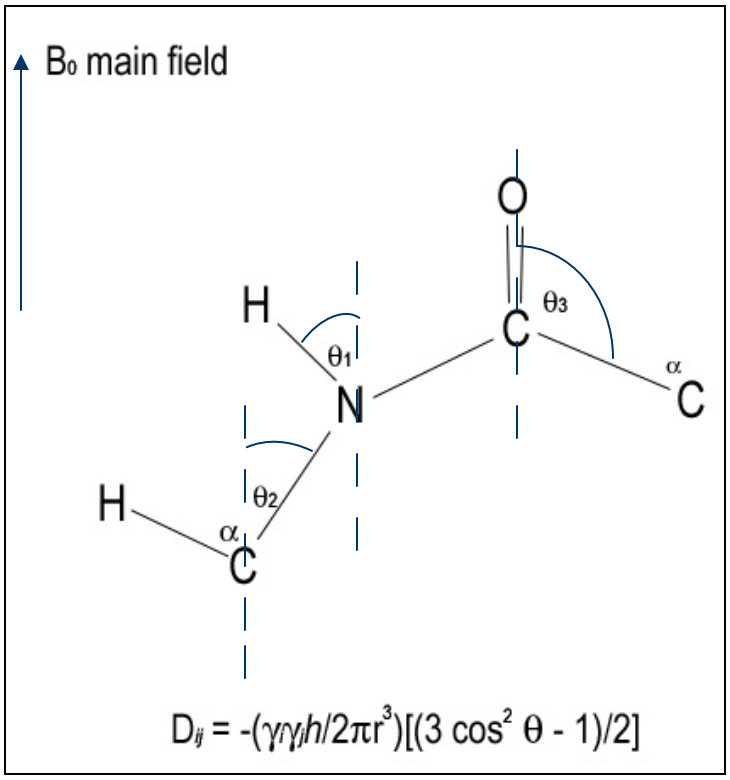



**Figure 9.** Residual dipolar coupling (RDC) angles. In the presence of a strong, static magnetic field, biomolecules in anisotropic solution will adopt a slight alignment with the main field. This alignment creates three Euler angles between the interacting nuclei and the static field (θ). Applying the value of this angle into the equation below the image provides information on biomolecular structure that cannot be gathered from molecules tumbling in isotropic solutions. However, the orientation of the tensor cannot be determined without structure calculation.

The protein sample plays an important role in overcoming some of the challenges faced by the NMR spectroscopist. In the Cavanagh lab, all of the isotopically labeled proteins are expressed in the bacterial host *Escherichia coli*. More often than not the expression of protein in adequate amounts and purity is difficult. NMR requires large amounts of isotopically labeled proteins, the production of which is often costly, time consuming, and, occasionally, impossible. Two proteins that failed to give samples using traditional methods were a large protein, calbindin $D_{28K}$, and a predominantly helical protein, LuxU.

## LARGE PROTEINS

Research in our lab required a 1 mM pure sample of rat brain calbindin $D_{28K}$, a relatively large protein which has a molecular weight just below 30 kDa. This protein has been the target of several other NMR labs for over a decade. Suitable samples and data have proven illusive.

Calbindin $D_{28K}$ background

Calbindin $D_{28K}$ is a protein containing 261 amino acids and has a molecular weight of 29,994 Da. It is a member of the calmodulin superfamily, which includes, among others, calretinin, troponin C, and parvalbumin. A common structural feature of this superfamily is the EF-hand motif, which adopts a helix-loop-helix conformation and has the ability to bind more than one calcium ion. Calbindin $D_{28K}$ is a calcium-binding protein comprising six putative EF-hands capable of binding a total of four calcium ions. It is expressed in many tissues, including brain, peripheral nervous system, kidney, and intestine.

The concentrations of calbindin $D_{28K}$ outside of the nervous system are enhanced by vitamin D and its derivatives. However, within the nervous system, this protein's concentrations are independent of vitamin D (13,14). The tissues in which calbindin $D_{28K}$ is found typically contain high amounts of the protein. For example, it constitutes between 0.1 and 1.5% of the total soluble protein in the brain, and calbindin $D_{28K}$ levels in auditory neurons are estimated to reach concentrations of up to 2 mM (15).

Calbindin $D_{28K}$ is involved in modulating calcium channel activity in neuronal cells, thus modifying intrinsic neuronal firing events, as well altering synaptic interaction in

hippocampal slices, and is implicated in the regulation of neuronal degeneration (16,17,18). It is also abundantly expressed in calcium-transporting tissues such as the intestine and kidney, and is localized in the β cells of the pancreas, where it enhances the synthesis and release of insulin from the islet cells (19,20-22). Despite the ubiquitous and critical nature of this protein, there is little known about its tertiary structure in either its calcium-free (apo) or calcium-loaded form. However, it has recently been demonstrated that apo-calbindin $D_{28K}$ is inherently less stable than the calcium-loaded form and that calcium-induced folding of the fragments may be important for successful reconstitution of the protein (23,24).

Other recent structural studies on calbindin $D_{28K}$ have focused on the specific effects of calcium ion loading. These investigations have primarily concerned the conformational state of apo-calbindin $D_{28K}$, the order and affinity in which calcium ions bind, and the specific identity of the EF-hands that are responsible for calcium ion binding. The most recent results of these studies concludes that apo-calbindin $D_{28K}$ adopts an ordered conformation and that calcium ion loading is not a parallel event, in which all metal-binding sites would pick up calcium ions at the same time, but rather a sequential event. One EF-hand loads calcium first (EF-hand 1), presumably the highest affinity site, then the next two sites load approximately simultaneously (EF-hands 4 and 5), while the final site lags behind slightly (EF-hand 3) (25). The EF-hands of calbindin $D_{28K}$ were determined by homology modeling.

Some members of the calmodulin superfamily are calcium sensors, which undergo a calcium-induced conformational change resulting in the exposure of a hydrophobic surface. The hydrophobic patch can then be used as a binding surface for target

molecules, which may become activated or attenuated after binding. Another class of calcium-binding proteins are calcium buffers or signal modulators. Unlike calcium sensors, the calcium-buffer proteins do not expose hydrophobic surfaces upon calcium binding and are believed to act to prevent sustained intracellular elevations of calcium. However, hydrophobic regions are exposed on the surface of calbindin $D_{28K}$ in both the calcium-free and calcium-loaded states. Therefore, calbindin $D_{28K}$ may possess the dual roles of buffer and sensor, or some as yet unidentified role. A protein designed only to bind calcium would not need exposed hydrophobic surfaces because such surfaces are thermodynamically unfavorable (26).

Large protein NMR: calbindin $D_{28K}$ challenges

The majority of protein structures that have been solved using NMR are below 30 kDa. The predominant reasons for this are signal overlap due to overcrowding of the NMR spectra and poor sensitivity. Overcrowding arises from the larger number of protons present in the protein or complex, and extensive resonance overlap is the result. A fully protonated sample of calbindin $D_{28K}$ will generate an excess of 10,000 NOE's.

Poor sensitivity is a result of slower tumbling times (i.e. longer rotational correlation times), which is directly linked to shorter transverse relaxation times. The rotational correlation time increases with molecular weight, resulting in large $^{1}H$ line widths and an associated decrease in the sensitivity of correlation experiments.

A traditional method for dealing with these problems is growing cells that produce our proteins in $D_2O$ and labeling them with $^{13}C/^{15}N$ isotopes. Because the gyromagnetic ratio of deuterium ($^{2}H$) is 6.5-fold smaller than that of hydrogen,

deuterons are 45-fold less effective than protons at causing dipolar relaxation of both the nucleus directly attached to a deuteron and the surrounding proton nuclei. Therefore, the incorporation of deuterium into proteins isotopically labeled with $^{13}C/^{15}N$ drastically increases the transverse relaxation ($T_2$) times of the directly attached carbon and $H_N T_2$. However, the introduction of $^{2}H$ into the proteins drastically reduces the number of NOE signals observed in NOESY spectra and inhibits structure calculations. The full perdeuteration of calbindin $D_{28K}$ would produce approximately 500 NOE's. This would not be enough to generate a high quality structure.

In an effort to optimize transverse relaxation times and the number of NOE signals in spectra, selective protonation of methyl groups, in addition to perdeuteration of the rest of the protein, is used. Combined with $^{13}C$ labeling, this technique retains excellent $T_2$ relaxation characteristics as well as increasing the available chemical-shift dispersion. The advantage of protonating methyl groups is that they are plentiful within the hydrophobic core of proteins. Their placement ensures considerably more NOE correlations than would be otherwise generated by surface hydrophilic residues and provides important restraints for structure calculations. ILV protonation produces spectra with an intermediate number of NOE's. Consequently, this process of specific protonation relieves spectral congestion and addresses the foremost problem associated with determining the structure of large proteins. Selectively protonating these methyl groups will ultimately include an extra 58 protonated amino acids that will generate additional NOE's.

To achieve this selective protonation, [3-$^{2}$H], $^{13}C_{\alpha}$-ketoisovalerate is used to supplement the $^{13}C/^{2}H$-glucose, $^{15}NH_4Cl$, $D_2O$ bacterial growth media, producing $^{13}C/^{2}H/^{15}N$ labeled proteins selectively protonated at the methyl groups of leucine (($^{1}$H-δ methyl)-leucine) and valine (($^{1}$H-γ methyl)-valine). Additionally, [3,3-$^{2}$H], $^{13}C_{\alpha}$-ketobutyrate is added to the media, producing ($^{1}$H-δ1 methyl)-isoleucine. These compounds are key members of the synthesis pathways these three amino acids. Their presence in the media ensures that their methyl groups will be retained in the final produce (i.e. the amino acid). Using these methods, more than 90% of the target residues acquired methyl protonation. Once the proteins have been protonated at the ILV residues, they can be subjected to through-space NOESY experiments ($H_N$-$H_N$, $H_N$-methyl, and methyl-methyl) to produce data that identifies ligand-binding sites and maps sites of interaction.

To generate protein samples suitable for such experiments, their expression in bacteria must be optimized. This is usually achieved by determining optimal growth conditions in rich media and then determining whether the same conditions are applicable to minimal media with similar results. However, this method does not always produce an adequate amount of the target protein for NMR experiments, especially when the cells are grown in deuterated media. The traditional methods employed for obtaining high expression of a target protein in deuterated media were described by Venters *et al* in 1995 (27). They outlined a method for "training" the cells to tolerate high levels of $D_2O$ in order to obtain high levels of deuterium incorporation into the protein. Their technique involved diluting cells directly from $H_2O$ LB media into LB media containing high levels of $D_2O$. The $D_2O$ tolerant cells

were then diluted 50-fold into a defined media, grown up to an inducible point, and incubated. This method was tested on calbindin $D_{28K}$, but the cells failed to produce adequate amounts of protein. Therefore, we developed an alternative method that did not involve $D_2O$ LB media. Our new technique involved growing cells in $H_2O$ LB media, pelleting the cells by gentle centrifugation once the culture reaches $A_{600nm}$=0.8-0.9 (for calbindin $D_{28K}$) or $A_{600nm}$=0.4-0.6 (for LuxU), and resuspending the pellet in defined $D_2O$ media at the same volume as the original culture. At that point, the culture is monitored to ensure cell growth. Once growth has been positively confirmed, protein production is induced with isopropyl-β-D-thiogalactopyranoside (IPTG -- a molecule that irreversibly binds the T7 promoter, which is upstream of the target gene, and induces constitutive production of the protein), and the culture is incubated at the appropriate temperature until harvest.

Calbindin $D_{28K}$ materials and methods

*Calbindin $D_{28K}$ protein expression and purification.* High-level expression of calbindin $D_{28K}$ in *E. coli* has been achieved by the construction of a vector containing the protein gene with a glutathione-S-transferase (GST) tag subcloned behind a phage T7 RNA polymerase promoter vector. One liter LB broth containing 100 μg/ml ampicillin was inoculated and grown at 34°C, 140 rpm to an optical density $A_{600nm}$ of between 0.8 – 0.9. The cultures were then centrifuged at 3,000 rpm to form soft pellets using a JLA 9.100 rotor in a Beckman Coulter Avanti J-25 centrifuge. The triply-labeled samples were made by resuspending a pellet in $D_2O$ M9T media with 1 g of $^{15}N$ ammonium chloride, 4 g of $^{13}C$ deuterated d-glucose, 50 mg of $^{13}C$ α-ketobutyric acid, and 85 mg of $^{13}C$ α-ketoisovaleric acid (all labeled chemicals came

from Cambridge Isotope Laboratories, Inc.). After waiting a brief period to ensure that the cells were recovering and growing in the new media (approximately 15 minutes), protein production was achieved by the addition of 1 mM IPTG, which induces a chromosomal copy of T7 polymerase in the cell line and in turn starts transcription of the protein gene. The cultures were harvested after 7 hours of growth after induction at 8,000 rpm for 15 minutes. The supernatant was discarded and the bacterial pellet was resuspended in 25 ml of lysis buffer (140 mM NaCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, 5 mM EDTA, 2 mM DTT, pH 7.5) and 40 μl 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF). The cells were lysed using a Vibra Cell sonicator (Sonics & Materials) approximately 10 times in 60 second bursts. The cell lysate was centrifuged at 18,000 rpm for 25 minutes, in a JA-20 rotor using a Beckman Coulter Avanti J-25 centrifuge. The supernatant was saved and applied to a glutathione-superflow resin (BD Biosciences Clontech). Calbindin $D_{28K}$ bound to the equilibrated glutathione resin and had to be eluted off the column using a salt bridge (gradient elution) between lysis buffer and elution buffer (140 mM NaCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, 5 mM EDTA, 2 mM DTT, 10 mM glutathione, pH 7.5). The fractions containing calbindin $D_{28K}$ were pooled and concentrated to 30 mls. One Taraunit of Thrombin (Plasminogen-Free, Bovine from Calbiochem) was used to cleave the GST tag from the protein. The protein was then loaded onto a Q-Sepharose High Performance agarose column (Amersham Pharmacia Biotech AB) and eluted using a salt bridge between column wash (50 mM Tris, 2 mM DTT, 5 mM EDTA, pH 7.5) and column elution (50 mM Tris, 2 mM DTT, 5 mM EDTA, 500 mM NaCl, pH 7.5). Fractions containing calbindin $D_{28K}$ were pooled,

concentrated and buffer-exchanged into lysis buffer. The protein was re-loaded onto glutathione-superflow resin. Calbindin $D_{28K}$ does not stick to the column at this stage, but comes off in the flow-thru, while the GST tag stuck to the column. The fractions containing purified calbindin $D_{28K}$ were pooled and concentrated to about 1 mM.

Calbindin $D_{28K}$ results: perdeuteration and site-specific protonation

Proteins grown for ILV protonation retain a high level of deuteration, which facilitates assignments of both backbone nuclei and side chain methyl groups. As the size of a protein increases, so do the difficulties associated with assigning its residues. Figure 10A shows a fully protonated TROSY spectra while Figure 10B is a 4-D NOESY plot of calbindin $D_{28K}$. The former image shows that there is poor resolution, even when the most crowded portion is magnified, and the later displays a tremendous amount of spectral overlap and poor resolution. Conversely, the Figures

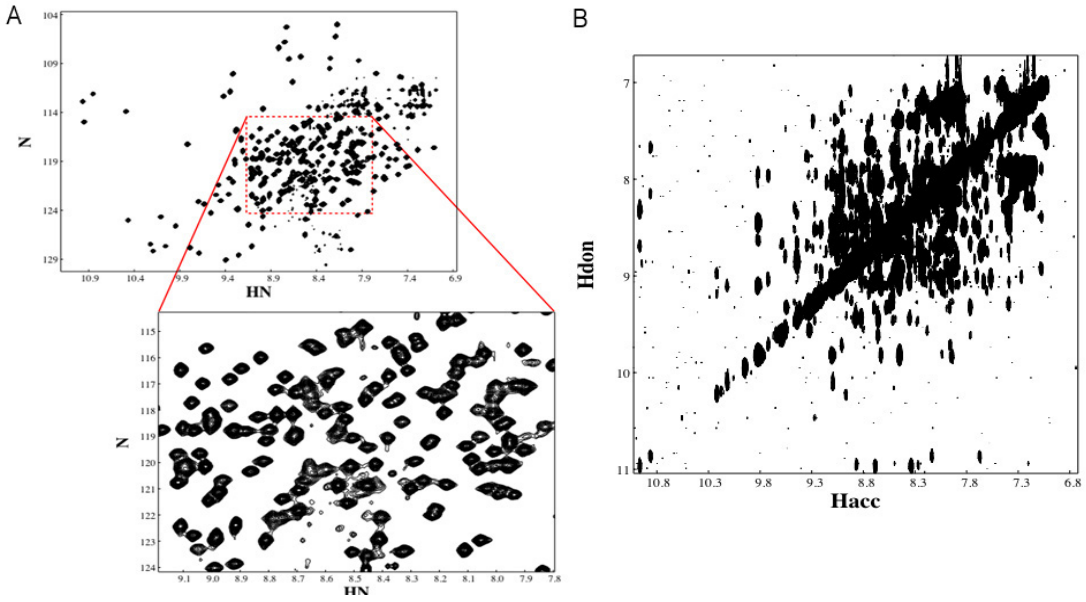



**Figure 10.** Calbindin $D_{28K}$ spectra. A) Fully protonated TROSY spectrum. B) 4D NOESY spectrum. Both spectra were taken in the following conditions: pH 7.2, 10 mM TRIS, 1 mM DTT, 0.02% $NaN_3$, $Ca^{2+}$ loaded to 6 molar equivalents.

11A and 11B contain significantly fewer peaks. These spectra represent a comfortable compromise involving perdeuteration and specific isotopic labeling of the ILV residues. Figure 11A shows NOESY spectra of triply labeled calbindin $D_{28K}$ and Figure 11B is a CHMQC spectrum of calbindin $D_{28K}$. The upper portion of the later spectrum corresponds to Ile residues while the bottom corresponds to Leu and Val residues. This spectrum is much less crowded than fully protonated spectra, yet contains enough constraints to generate a representative structure of the protein. The constraints corresponding to each residue are in Figure 11C. The residues with the most constraints are located in the hydrophobic core of the protein while those with few constraints correspond to amino acids situated at solvent-exposed edges of the protein where there are few interresidual contacts.

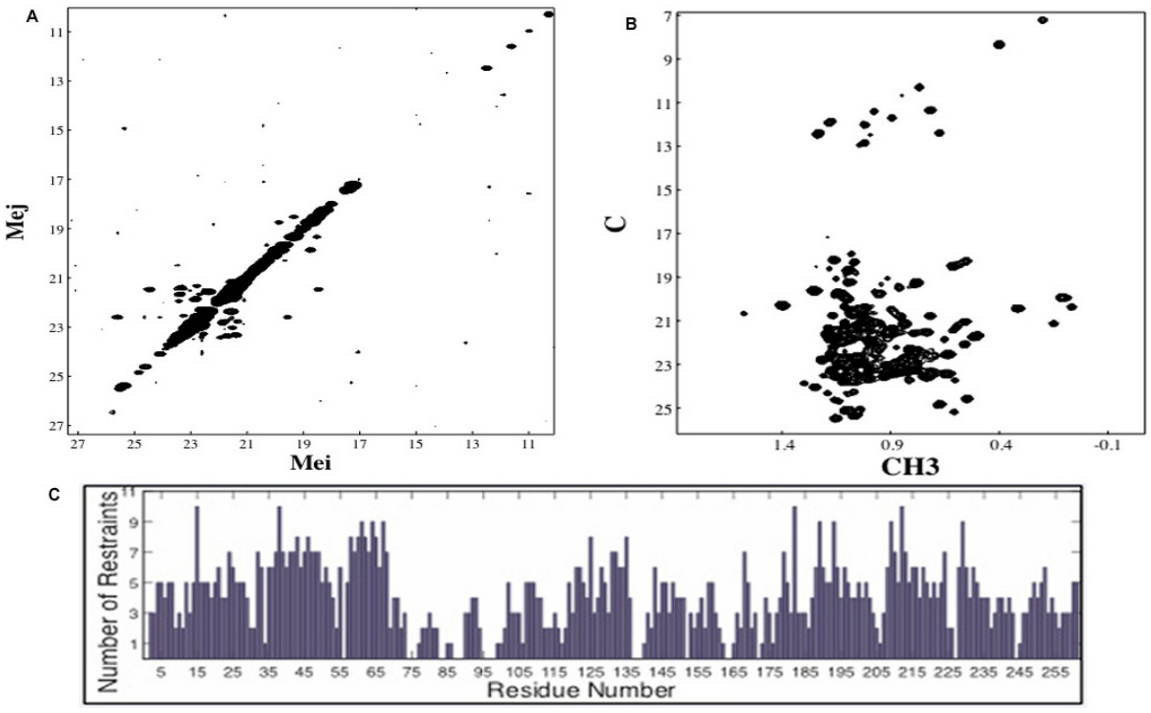



**Figure 11.** Calbindin $D_{28K}$ spectra. A) NOESY spectrum with $^{2}H^{13}C^{15}N$ labeling and ILV protonation. B) CHMQC spectrum with $^{2}H^{13}C^{15}N$ labeling and ILV protonation. C) Chart of the number of restraints that correspond to each individual residue. Spectra in A and B were taken in the following conditions: pH 7.2, 10 mM TRIS, 1 mM DTT, 0.02% $NaN_3$, $Ca^{2+}$ loaded to 6 molar equivalents.

The difference between the number of restraints for each residue is crucial to the identification of secondary structure. As the structure of calbindin $D_{28K}$ has evolved, this parameter has proven to be of the utmost importance. As illustrated by Figure 12A, 12B and 12C, the structure of calbindin $D_{28K}$ has begun to change as more data has been collected and more restraints have been employed to calculate its structure. (All structure calculations of calbindin $D_{28K}$ were performed by Doug Kojetin of the Cavanagh lab.) The most significant changes are on the outer edges of the protein where unstructured loops have been determined to have α-helical secondary structure.

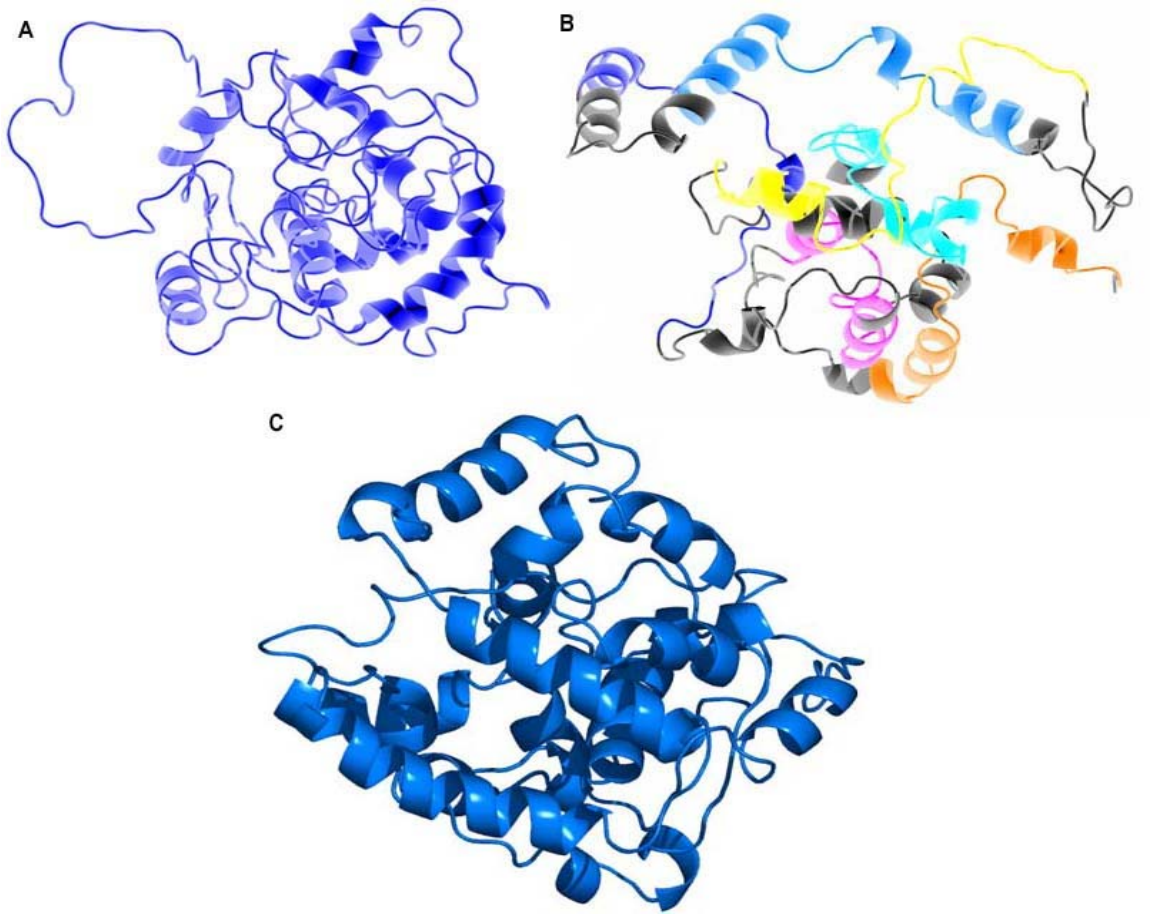



**Figure 12.** Calculated structures of calbindin $D_{28K}$. Image C is the most recent structure and contains more restraints per residue than images A and B. This is achieved by repeatedly analyzing the data and tightening the resulting structure from previous calculations.

## HELICAL PROTEINS

LuxU background

LuxU, a protein from *Vibrio harveyi*, contains 114 amino acids, weights 12,689 Da, and we believe forms a four helix bundle when it is properly folded. LuxU is present in bacteria that use chemical signals to coordinate group behavior and gene expression. They accomplish this by producing, releasing, detecting, and responding to molecules called autoinducers. The ability of bacteria to detect autoinducers allows them to differentiate between high and low cell density and to coordinately regulate gene expression of the community in a process called quorum sensing. This process was first described in the marine bioluminescent bacteria *Vibrio fischeri* and *V. harveyi*, but this process is known to exist in many species of bacteria. *V. harveyi* has evolved a quorum sensing pathway that has characteristics of both Gram-negative and Gram-positive bacterial quorum sensing systems. It uses an acyl homoserine lactone autoinducer like Gram-negative quorum sensing bacteria, but the signal detection and relay apparatus is comprised of two-component signaling proteins similar to those used by Gram-positive quorum sensing bacteria (28).

A model illustrating the *V. harveyi* hybrid quorum sensing two-component system is presented in Figure 13. As shown, *V. harveyi* produces two autoinducers, called AI-1 and AI-2. AI-1 is *N*-(3-hydroxybutanoyl)-homoserine lactone and AI-2 is a novel furanosyl borate diester molecule that has no similarity to other known autoinducers (Figure 14). AI-1 is produced by the protein LuxLM, yet this protein has no sequence homology to the LuxI family of molecules. The protein responsible for the production of AI-2 is LuxS (29).

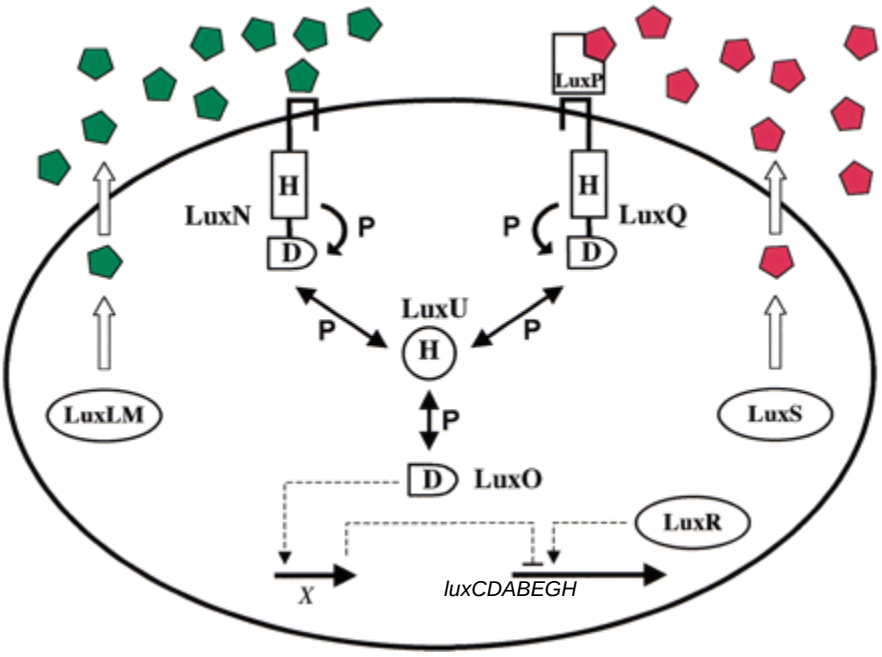



**Figure 13.** The hybrid quorum sensing circuit of *V. harveyi*. Elements characteristic of both Gram-negative and Gram-positive bacterial quorum sensing systems are combined to make up the *V. harveyi* circuit. An acyl-HSL autoinducer (AI-1, green pentagons) is produced by the activity of LuxLM. A second autoinducer (AI-2, red pentagons) is synthesized by the enzyme LuxS. AI-2 is proposed to be a furanone. Both autoinducers accumulate as a function of cell density. The sensor for AI-1 is LuxN, and two proteins, LuxP and LuxQ, function together to detect AI-2. LuxP is homologous to the periplasmic ribose binding protein of *E. coli*. LuxN and LuxQ are hybrid sensor kinase/response regulator proteins that transduce information to a shared integrator protein called LuxU. LuxU sends the signal to the response regulator protein LuxO. The mechanism of signal transduction is phosphorelay (denoted P). LuxO controls the transcription of a putative repressor protein (denoted X), and a transcriptional activator protein called LuxR is also required for expression of the luciferase structural operon (*luxCDABE*). The details of the signaling mechanism are given in the text. The conserved phosphorylation sites on the two-component proteins are indicated as H (histidine) and D (aspartate). (28)

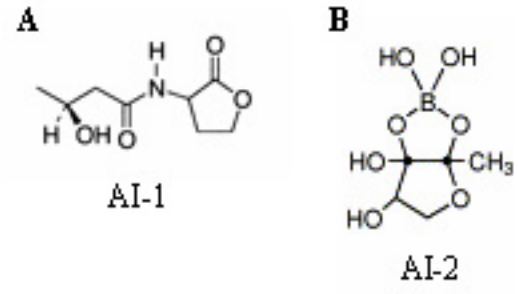



**Figure 14**. *Vibrio harveyi* autoinducer molecules. (A) Autoinducer-1: *N*-(3-hydroxybutanoyl)-homoserine lactone. (B) Autoinducer-2: furanosyl borate diester. (28,29)

As Figure 13 indicates, LuxN and LuxPQ recognize AI-1 and AI-2, respectively. LuxP is a soluble periplasmic protein that is assumed to be the primary receptor for AI-2 and interacts with LuxQ when LuxP and AI-2 form a complex (28,30-32). At low cell density, autoinducer concentration is low and LuxN and LuxQ autophosphorylate conserved histidine residues in their respective sensor kinase domains. They subsequently transfer phosphate to conserved aspartate residues in their adjacent response regulator domains. The phosphate groups from both LuxN and LuxQ are transferred to conserved histidine residues of LuxU, which then delivers the phosphate to the appropriate aspartate residue of LuxO. It is postulated that phosphorylated LuxO interacts with $\sigma^{54}$ and activates expression of an unknown downstream repressor X that turns off light production. Conversely, at high cell density, the pathway goes in reverse and the autoinducers LuxN and LuxQ function as phosphatases to deplete phosphate from LuxO via LuxU. Unphosphorylated LuxO is inactive and therefore the repressor X is not transcribed. Under these conditions, a transcriptional activator LuxR is activated, the *lux* operon is transcribed, and the cells produce light (33,34).

The dual activity of the AI-1 and AI-2 pathways in *V. harveyi* is rare and it is hypothesized that this bacterium uses AI-1 for intraspecies communication and AI-2 for interspecies communication (28,35,36). *V. fischeri* primarily exists in a pure environment within a specialized host light organ, whereas *V. harveyi* utilizes quorum sensing in environments that contain a diverse array of bacteria. Therefore, a single species-specific language is not adequate for density sensing in bacteria that commonly live in mixed-species environments. *V. harveyi* presumably uses its

specific AI-1 language to assess its own population density and uses the nonspecific AI-2 language to assess the population density of all the other species in the area.

Helical proteins: LuxU challenges

Proteins with high α-helical content generally prove to be more challenging to generate structures for via NMR than proteins with a high percentage of β-strands or a mix of β-strands and α-helices. Highly α-helical proteins yield spectra with limited resolution for assignments and NOE measurements because of the overlap generated by excessive proton coupling signals. Conversely, loop regions connecting α-helices tend to generate broad, indistinct peaks in NOESY spectra. This is due to an insufficient number of interresidual contacts present in these areas of the protein.

One alternative to improving NMR spectra of helical proteins is residual dipolar coupling. Effectively, at high magnetic field strengths a small amount of molecular alignment occurs for molecules that have a nonzero magnetic sensitivity. Of greatest importance is the angle between the vector connecting the interacting nuclei and the applied magnetic field of the spectrometer. Knowing the value of this angle for a bonded pair of nuclei could be very useful in defining a molecular structure. However, since the molecules are tumbling in an isotropic solution, they uniformly sample all directions in space so all the angles are time averaged and thus equal zero.

This small alignment results in measurable one-bond $^{1}H$-$^{15}N_{H}$ residual dipolar couplings. In directly bonded nuclei (like amide $^{1}H_{N}$-$^{15}N$ pairs in proteins), these couplings reveal themselves as small, field-dependent perturbations to the $^{1}H_{N}$-$^{15}N$ scalar coupling as measured in non-decoupled spectra. Because dipolar couplings are proportional to the square of the field strength and scalar coupling is not, the effects

of dipolar and scalar couplings can be separated by measuring the splittings that result from the combined couplings at several different field strengths. Once separated, the constant scalar coupling contribution can be removed. Since residual dipolar couplings are angle dependent, they can be used as structural restraints that do not rely on neighboring nuclei to define their position and orientation. Therefore, they can uniquely provide information that can orient structural elements located in the extremities of the tertiary fold of the protein.

Measurement of dipolar couplings is best realized by using anisotropic media for macromolecular alignment. Macromolecules dissolved in such media adopt a small degree of alignment due to steric effects. However, the resultant protein alignment must be properly adjusted to ensure that only closely spaced nuclei give measurable dipolar couplings. If there is too much protein alignment, there may be too many large dipolar coupling interactions and the resulting spectra will be impossible to interpret. Additionally, the spacing between the anisotropic media must be greater than the size of the target macromolecule to guarantee $T_2$ relaxation rates are not adversely affected.

Despite the fact that this sample was not being grown in $D_2O$ media, cells containing the LuxU plasmid failed to produce an adequate amount of the target protein in standard minimal media. As a result, to obtain a doubly labeled sample of LuxU ($^{1}H^{13}C^{15}N$), we employed the same technique developed for calbindin $D_{28K}$. This method successfully produced high quantities of LuxU.

LuxU materials and methods

*LuxU protein expression and purification.* High-level expression of LuxU was achieved by cloning the protein gene into a pRSET A vector subcloned behind a phage T7 RNA polymerase promoter vector using NheI at the 3' end of the insert, as well as a thrombin cleavable his-tag, and HindIII at the 5' end by Dagny Ulrich in Pat Loria's lab at Yale. DNA was isolated using a QIAprep Spin Miniprep Kit (250) from Qiagen. The plasmids were transformed into competent *E. coli* BL21(DE3) pLysS purchased from VWR. One liter LB broth containing 100 μg/ml ampicillin was inoculated and grown at 34$^{o}$C, 140 rpm to an optical density $A_{600nm}$ of between 0.4 – 0.6. The cultures were then centrifuged at 3,000 rpm to form soft pellets using a JA-10 rotor in a Beckman Coulter Avanti J-25 centrifuge. The doubly-labeled samples were made by resuspending a pellet in $H_2O$ M9T media with 1 g of $^{15}$N ammonium chloride and 1 g of $^{13}$C D-glucose supplements. One liter LB broth containing 100 μg/ml ampicillin was inoculated and grown at 34$^{o}$C, 140 rpm to an optical density $A_{600nm}$ of between 0.4 – 0.6. The cultures were then centrifuged at 3,000 rpm to form soft pellets using a JA-10 rotor in a Beckman Coulter Avanti J-25 centrifuge. The triply-labeled samples were made by resuspending a pellet in $D_2O$ M9T media with 1 g of $^{15}$N ammonium chloride, 2 g of $^{13}$C deuterated D-glucose, 50 mg of $^{13}$C α-ketobutyric acid, and 85 mg of $^{13}$C α-ketovaleric acid (all labeled chemicals came from Cambridge Isotope Laboratories, Inc.). After waiting a brief period to ensure that the cells were recovering and growing in the new media, protein production was achieved by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside, which induces a chromosomal copy of T7 polymerase in the cell line and in turn starts

transcription of the protein gene. The cultures were grown at 27$^{o}$C, 120 rpm and harvested after 7 hours of growth after induction at 8,000 rpm for 15 minutes. The supernatant was discarded and the bacterial pellet was resuspended in 25 ml of lysis buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 5 mM β-Mercaptoethanol, pH 8.0), 0.05% Triton X-100, and 40 μl of AEBSF. The cells were lysed using a Vibra Cell sonicator (Sonics & Materials) approximately 20 times in 60 second bursts. The cell lysate was centrifuged at 10,000 rpm for 60 minutes, in a JA-20 rotor using a Beckman Coulter Avanti J-25 centrifuge. The supernatant was saved and applied to a Ni-NTA Agarose column (Qiagen) to purify it. LuxU bound to the agarose and had to be eluted off the column using a salt bridge between wash buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 5 mM β-Mercaptoethanol, 20 mM Imidazole, pH 8.0) and elution buffer (50 mM $NaH_2PO_4$, 300 mM NaCl, 5 mM β-Mercaptoethanol, 500 mM Imidazole, pH 8.0). The fractions were pooled and dialyzed into column wash (50 mM Tris, 2 mM DTT, 5 mM EDTA, pH 8.0). The dialyzed protein was then loaded onto a Q-Sepharose High Performance agarose column (Amersham Pharmacia Biotech AB) and eluted using a salt bridge between column wash buffer and column elution buffer (50 mM Tris, 2 mM DTT, 5 mM EDTA, 500 mM NaCl, pH 8.0). Fractions containing LuxU were pooled and concentrated to two-thirds of its total volume. Approximately 180 μg of Thrombin (Plasminogen-Free, Bovine from Calbiochem) was used to cleave the N-terminal his-tag from the protein, concentrated to approximately 3 ml, and dialyzed into 50 mM $KH_2PO_4$, 50 mM NaCl, and 1 mM DTT at pH 6.4. Size-exclusion liquid chromatography at 4$^{o}$C was then performed on the sample and all fractions containing LuxU were pooled and concentrated to approximately 1 mM.

LuxU Results: monomer or dimer?

Initially, it was assumed that LuxU was a dimer and was often compared to Spo0B, a key protein in the sporulating pathway of *B. subtilis*, because of the strong similarity that they have with one another both in structure and function. Like LuxU, Spo0B is an entirely helical protein. Also, its primary role is to accept a phosphate group from an aspartate pocket of Spo0F onto histidine residues and then transfer the phosphate to an aspartate pocket of Spo0A. However, Spo0B only has two α-helices and is only functional when it is a dimer and there are therefore four helices involved in its function (Figure 15). Because LuxU independently forms a four helix bundle, it seemed irrational for it to assemble as a homodimer. To experimentally determine dimer formation, size exclusion chromatography (SEC) and dynamic light scattering (DLS) were performed on LuxU.

**Figure 15.** Spo0B (center) with Spo0F bound on either end. The four-helix bundle formed by Spo0B dimers strongly resembles the fully-folded structure of LuxU.

DLS was performed on only one of the LuxU constructs, whereas SEC was performed on both of them. Figure 16 contains SEC data on both LuxU and His-

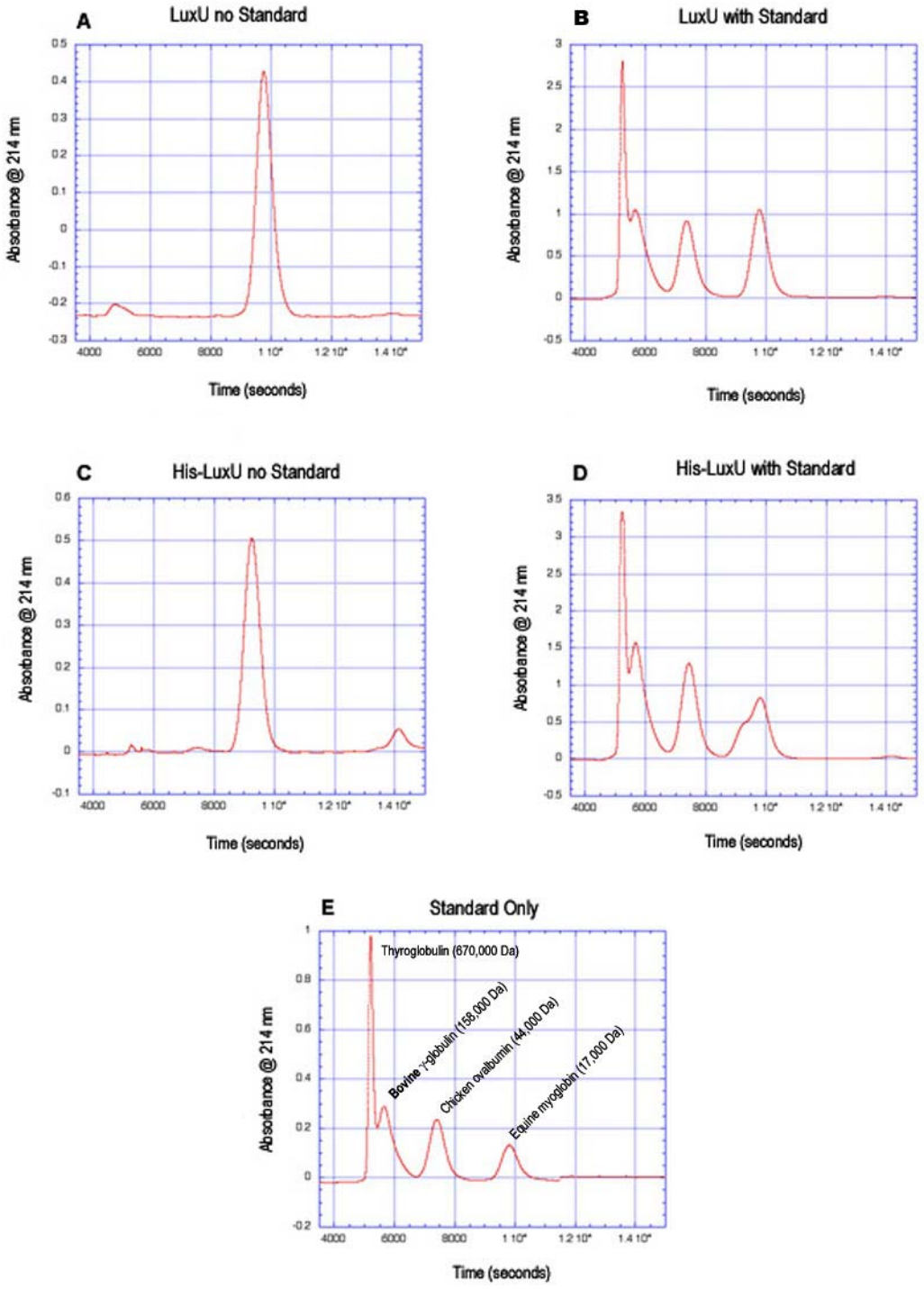



**Figure 16.** LuxU size exclusion chromatography data. Conditions for all are the same: 50 mM $KH_2PO_4$, 50 mM NaCl, 1 mM DTT, pH 6.4. LuxU elutes off the column at a molecular weight that corresponds to a monomer conformation. Data was collected using a HiPrep™ 26/60 Sephacryl™ S-100 High Resolution size exclusion column.

tagged LuxU. Both illustrate that LuxU elute off the column at a molecular weight that is consistent with a monomer conformation. DLS data acquired on LuxU also provided information that corroborated with that collected from SEC (data not shown). Together, these data suggested that LuxU does not adopt a dimer conformation; rather it functions as a monomer in the cell.

LuxU results: residual dipolar coupling

The structure of LuxU was primarily determined using the experiments described in the section on multidimensional NMR. The through space NOESY on $^{1}H$ $^{13}C$ $^{15}N$ LuxU was very important, but provided insufficient restraints to refine the structure (Figure 20). However, data acquired from this experiment allowed us to calculate a minimum energy structure of LuxU without violations. The RDC HSQC experiment was used on the same sample of $^{1}H$ $^{13}C$ $^{15}N$ LuxU with Pf1 filamentous bacteriophage titrated into it. Figure 17 shows the HSQC spectra of LuxU before the bacteriophage was titrated into the sample.

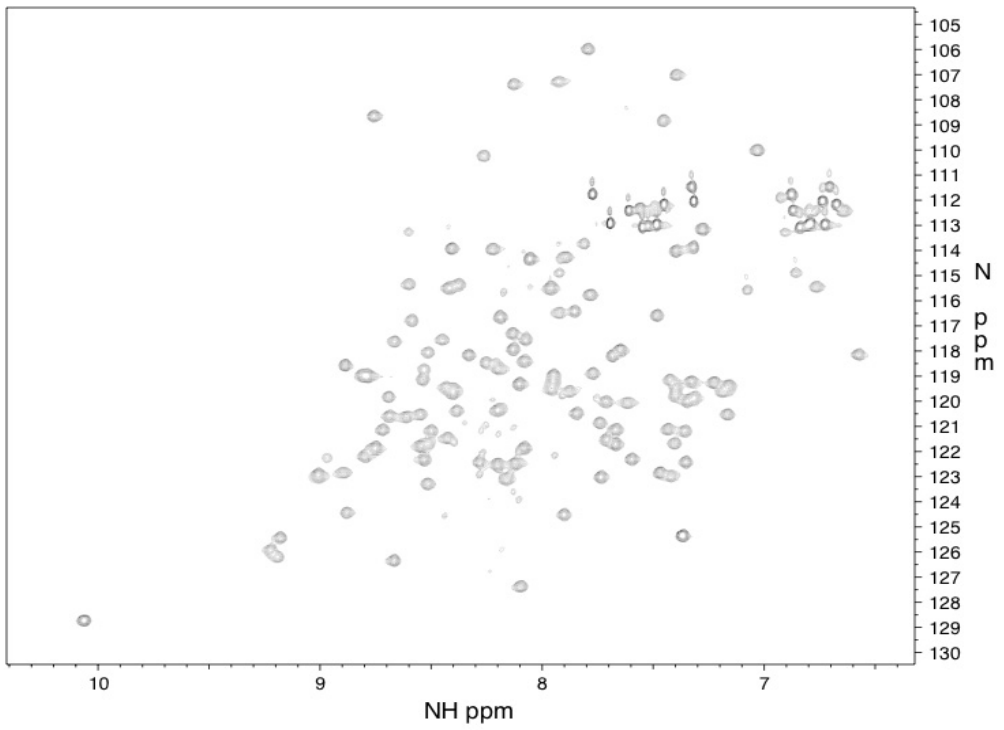



**Figure 17**. $^{1}H^{13}C^{15}N$ LuxU HSQC spectrum taken on the sample used for RDC before Pf1 filamentous bacteriophage was titrated into the sample.

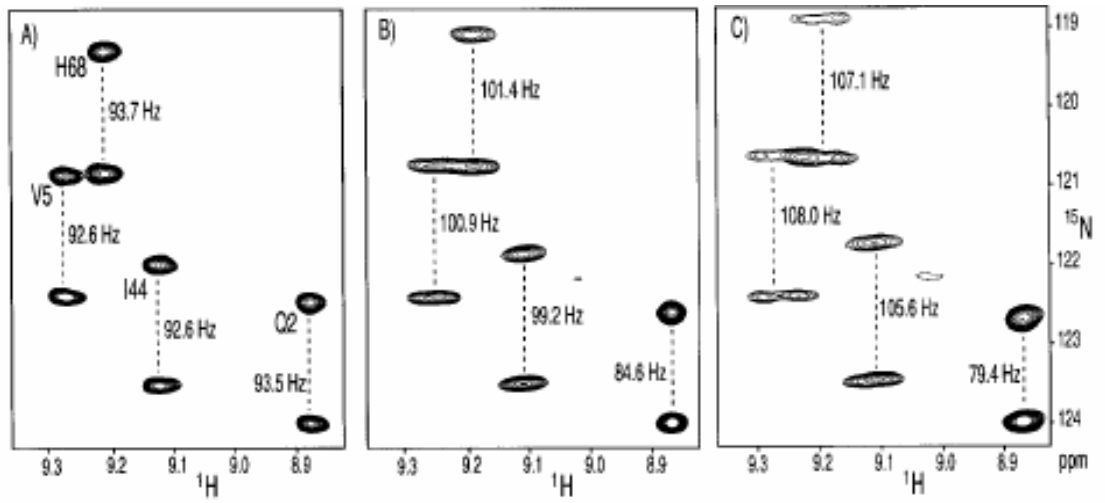



**Figure 18.** Small regions of the 600 MHz $^{15}N$-$^{1}H$ correlation spectra of ubiquitin, recorded in the absence of $^{1}H$ decoupling in the $^{15}N$ dimension, at three different levels of molecular alignment. (A) Isotropic spectrum, with t he marked splitting corresponding to $^{1}J_{NH}$. (B) Spectrum recorded in 4.5% (w/v) bicelles, consisting of a 30:10:1 molar ratio of DMPC, DHPC, and cetyl-trimethyl ammonium bromide (CTAB). (C) Spectrum recorded in 8% (w/v) bicelles. Marked splittings in panels B and C correspond to the sum of the $^{1}J_{NH}$ and dipolar coupling. The broadening in the $^{1}H$ dimension, observed in panels B and C relative to A is caused by $^{1}H$-$^{1}H$ dipolar couplings. (37)

The observed splittings with bacteriophage were compared to those without bacteriophage using MATLAB software. Figure 18 provides an example of the splitting observed in HSQC experiments on isotropic vs. anisotropic solutions. Just below it, Figure 19 also illustrates the difference between spectra before and after phage was titrated into the $^{1}H$ $^{13}C$ $^{15}N$ LuxU sample. Due to the area being observed, the splittings in Figure 19 are not as obvious as those seen in Figure 18, yet the information garnered from each data set provide identical data for the respective protein being analyzed.

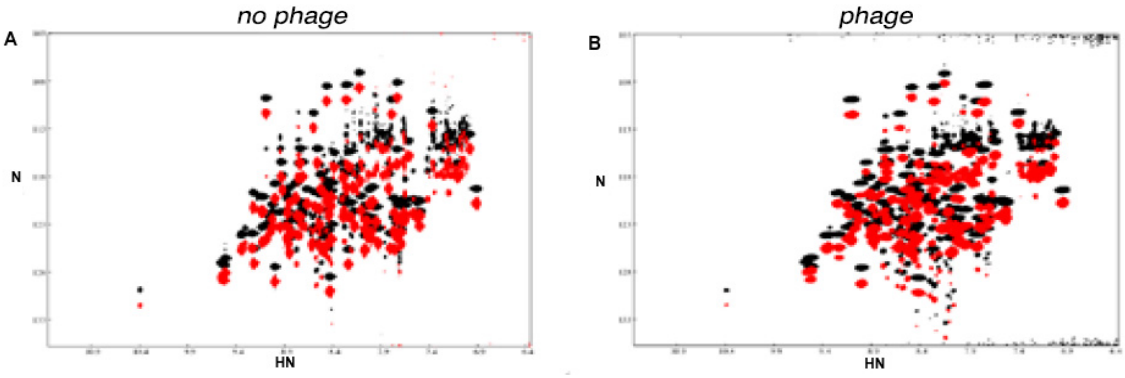



**Figure 19.** $^{1}H^{13}C^{15}N$ HSQC spectra of LuxU on a 600 MHz magnet. A) Isotropic spectrum. B) Anisotropic spectrum recorded in 5% (w/v) Pf1 filamentous bacteriophage.

The small differences observed between spectra on pure samples and those containing phage become critical when determining the three-dimensional structure of proteins. Figure 20 is the structure of LuxU as generated entirely from samples containing no phage. These structures contain a high degree of variation and it is difficult to accurately determine the true structure of the protein. However, spectra collected from no-phage and phage samples provided excellent RDC data, which helped generate the structures in Figure 21A and 21B. The figures containing an overlay of

the structures with and without RDC's clearly illustrate the advantage of using anisotropic solutions in the acquisition of spectra. The structure generated with RDC's shows that the helices are more parallel than otherwise shown in the structure generated without the assistance of RDC's. Table 2 provides the raw data used to generate the structures seen in Figure 21. (All structure calculations of LuxU were performed by Doug Kojetin of the Cavanagh lab.)

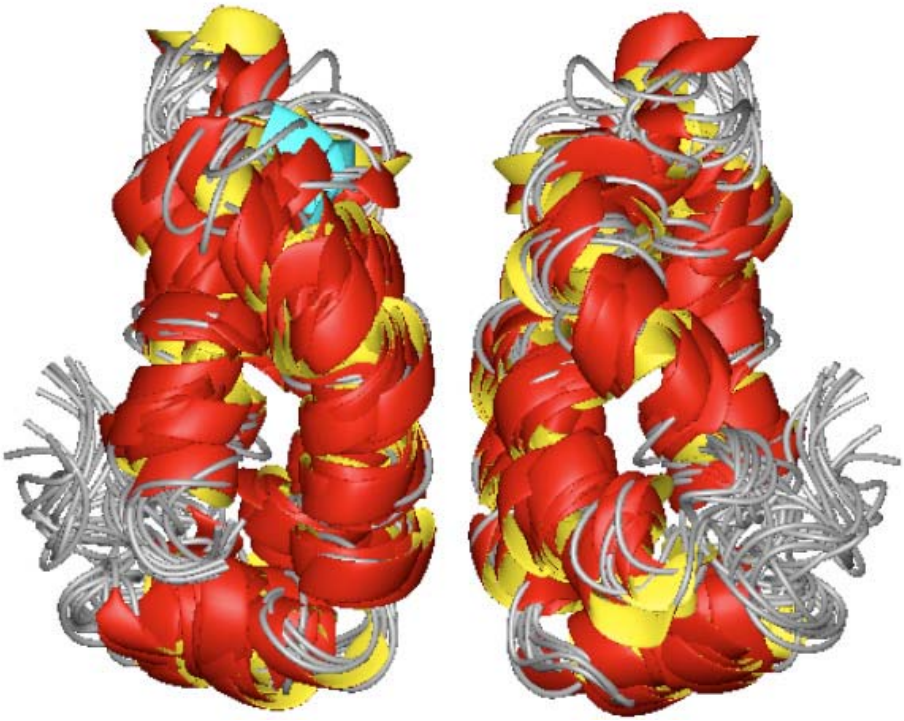


**Figure 20.** Structure generated from data before RDC was performed on the $^{1}H^{13}C^{15}N$ LuxU sample. The inaccuracy of the overlapping structures is indicative of an inadequate number of restraints per residue, especially for loop regions and the ends of the protein.

**Table 2.** Table of the root mean square deviation (RMSD) values that indicate structure quality.

| # | RDCs? | bb rmsd to avg | bb std dev | sc rmsd | sc std dev | notes |
|---|---|---|---|---|---|---|
| 1 | no | 0.758846 | 0.11303 | 1.5237 | 0.18658 | some NOEs commented out; dihedral angle ranges >=20 |
| 2 | no | 0.593148 | 0.15071 | 1.2384 | 0.16078 | some NOEs commented out; dihedral angle ranges at original values & edited |
| 3 | no | 0.539257 | 0.14868 | 1.271 | 0.13938 | same as 2, but calculated 200 structures |
| 4 | YES | 0.463705 | 0.14549 | 1.2641 | 0.14355 | same as 2, except input HN RDC data |

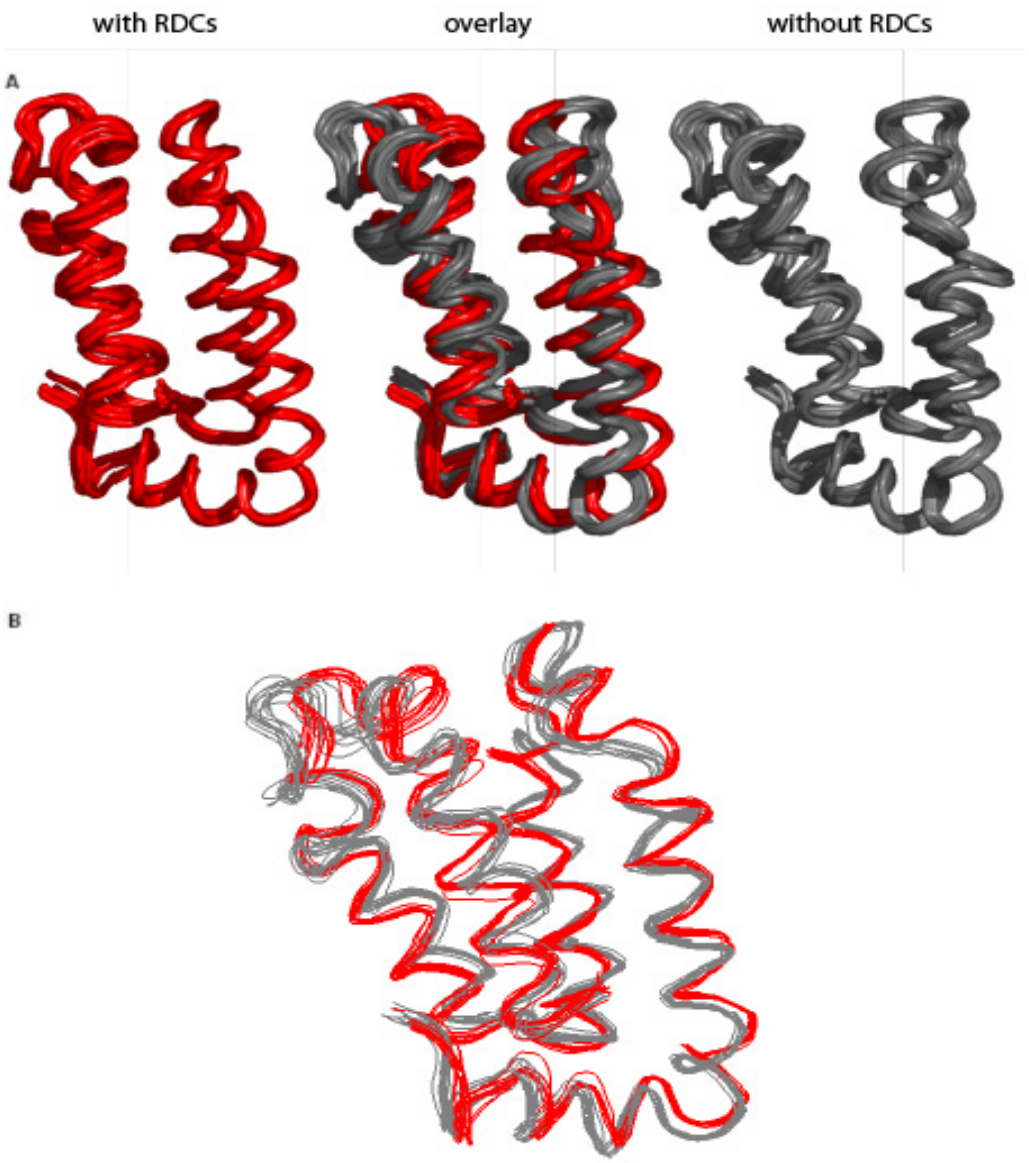



**Figure 21.** A) The right image is the structure of LuxU that was calculated without RDC, the left is LuxU calculated with RDC, and the center is of the two structures overlapped. B) Another image of the two LuxU structures calculated before and after RDC experiments.

## CONCLUSION

With the completion of multiple genomes and the concomitant rate at which the easiest protein structures are being elucidated, the larger, more complex ones left will be especially challenging. As a result, a majority of biomolecular NMR techniques will require specialized selective isotopic labeling ($^{2}H$, $^{13}C$, and $^{15}N$). The specialized labeling of these samples will be required to allow optimization of NMR solution conditions and the collection of high quality data allowing for sequential assignments and structure determination. Many of these investigations will require large quantities of pure isotopically labeled proteins, the production of which will often prove to be costly, time consuming, and difficult, if not impossible.

Currently, most isotopically labeled recombinant proteins are expressed in the bacterial host *E. coli*. Methods of generating heteronuclear labeled samples in *E. coli* commonly use standard versions of M9 minimal media (38). A variety of different strategies for enhancing *E. coli* growth and protein expression have been employed. Minimal media supplements such as trace metal mixtures, vitamin cocktails, as well as commercially available algal and microbial hydrolysates, have shown enhancements in growth and expression, but failed for the two proteins examined in this thesis. The method developed here employs cell mass that is predominately grown in unlabeled rich media which makes the cells healthy and viable as well as allowing for rapid growth to high cell densities. Following the growth in the unlabeled medium, cells are exchanged into the desired specialized isotopically defined media at high cell densities and are thereby optimized for maximal protein

expression. The isotopic incorporation rates obtained were indiscernible from proteins produced in the traditional manner.

Expression yields were increased, and cells containing plasmids that had formerly failed to thrive or express protein in specialized minimal media grew and expressed protein at levels acceptable for NMR. The expression protocol was developed based on contemplation of the following ideas. First, the growth rate for *E. coli* in minimal media is always slow. Longer growth rates often result in reduced expression due to cytotoxic effects associated with the plasmid gene products and/or natural by-products of the cell. In addition, it makes sense that a "huge" healthy viable culture of cells would dramatically increase the rate and amount of protein expression if they lived through the "shock" of the media exchange. Additionally, isotope consumption would be reduced by generating the majority of cell mass using unlabeled media. The only potential drawback was that it was necessary to have plasmids that didn't "leak", as they could potentially introduce a large amount of unlabeled protein into the NMR sample. This method offers an alternative to fermenter growths, which require additional setup time and specialized equipment which in an academic setting is cost prohibitive. The time involved in preparing isotopically labeled samples using this method is only slightly greater than growth in unlabeled rich media, saving time when compared to standard minimal media growth procedures. Additionally, higher expression levels seen in the media made with deuterium oxide make this method ideal for selective side-chain or amino acid labeling procedures that employ extremely expensive isotope-labeled amino acids or precursors.

This method was successfully applied to two proteins: calbindin $D_{28K}$ and LuxU. Protein obtained from both of these samples provided high quality NMR data, as seen in the results, which has allowed for three dimensional structure determination and structure refinement.

## REFERENCES


**1. Cavanagh, J., W. J. Fairbrother, A. G. Palmer III, and N. J. Skelton.** 1996. Protein NMR spectroscopy: principles and practice. Academic Press, San Diego.

**2. Ernst, R. R., G. Bodenhausen, and A. Wokaun.** 1987 and 1994. Principles of nuclear magnetic resonance in one and two dimensions. Clarendon Press, Oxford.

**3. Wider, G.** 1998. Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. Prog. NMR Spectrosc. **32:**193-275.

**4. Wüthrich, K.** 1986. NMR of proteins and nucleic acids. John Wiley & Sons, New York.

**5. Sanders, J. K. M., and B. K. Hunter.**1993. Modern NMR spectroscopy: a guide for chemists, Second Edition. Oxford Universiy Press, New York.

**6. Bax, A., and S. Grzesiek.** 1993. Methodological advances in protein NMR. Acc. Chem. Res. **26:**131-138.

**7. Sattler, M., J. Schleucher, and C. Griesinger.** 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Progr. NMR Specrosc. **34:**93-158.

**8. Aue, W. P., E. Bartholdi, and R. R. Ernst.** 1976. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. J. Chem. Phys. **64:**2229-2246.

**9. Wider, G., S. Macura, Anil. Kumar, R. R. Ernst, and K. Wüthrich.** 1984. Homonuclear two-dimensional $^{1}H$ NMR of proteins. Experimental Procedures. J. Magn. Reson. **56:**207-234.

**10. Solomon, I.** 1995. Relaxation processes in a system of two spins. Phys. Rev. **99:**559-565.

**11. Kumar, A., R. R. Ernst, and K. Wüthrich.** 1980. A 2D NOE experiment for the elucidation of complete proton-proton cross-relaxation networks in biological macromolecules. Biochem. Biophys. Res. Commun. **95:**1-6.

**12. Bodenhausen, G. and D. J. Ruben.** 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. Chem. Phys. Lett. **69:**185-189.

**13. Christakos, S., E. J. Friedlander, B. R. Frandsen, and A. W. Norman.** 1979. Studies on the mode of action of calciferol. XIII. Development of a radioimmunoassay for vitamin D-dependent chick intestinal calcium-binding protein and tissue distribution. Endocrinology. **104:**1495-1503.

**14. Christakos, S., and A. W. Norman.** 1980. Radioimmunoassay for chick intestinal calcium-binding protein. Methods Enzymol. **67:**500-503.

**15. Oberholtzer, J. C., Buettger, C. Summers, M. C., and F. M. Matschinsky.** 1988. The 28-kDa calbindin-D is a major calcium-binding protein in the basilar papilla of the chick. Proc. Natl. Acad. Sci. U. S. A. **85:**3387-3390.

**16. Kohr, G., and I. Mody.** 1991. Endogenous intracellular calcium buffering and the activation/inactivation of HVA calcium currents in rat dentate gyrus granule cells. J. Gen. Physiol. **98:**941–967.

**17. Kohr, G., C.E. Lambert, and I. Mody.** 1991. Calbindin $D_{28K}$ (CaBP) levels and calcium currents in acutely dissociated epileptic neurons. Exp. Brain Res. **3:**543–551.

**18. Mody, I., J.N. Reynolds, M.W. Salter, P.L. Carlen and J.F. MacDonald**. 1990. Kindling-induced epilepsy alters calcium currents in granule cells of rat hippocampal slices. Brain Res. **531:**88–94.

**19. Christakos, S., C. Gabrielides and W.B. Rhoten.** 1989. Vitamin D-dependent calcium binding proteins: chemistry, distribution, functional considerations, and molecular biology. Endocr. Rev. **10:**3–26.

**20. Johnson, J.A. and R. Kumar.** 1994. Vitamin D and renal calcium transport. Curr. Opin. Nephrol. Hypertens. **3:**424–429.

**21. Johnson, J.A. and R. Kumar.** 1994. Renal and intestinal calcium transport: roles of vitamin D and vitamin D-dependent calcium binding proteins. Semin. Nephrol. **14** (1994), pp. 119–128.

**22. Johnson, J.A., J.P. Grande, P.C. Roche and R. Kumar.** 1994. Immunohistochemical localization of the 1,25(OH)2D3 receptor and calbindin $D_{28k}$ in human and rat pancreas. Am. J. Physiol. **267:**356–360.

**23. Berggard,T., E. Thulin, K.S. Akerfeldt and S. Linse.** 2000. Fragment complementation of calbindin $D_{28K}$. Protein Sci. **9:**2094–2108.

**24. Berggard, T., M. Simona, P. Onnerfjord, E. Thulin, K.S. Akerfeldt, J.J. Enghild, M. Akke and S. Linse.** 2002. Calbindin $D_{28K}$ exhibits properties characteristic of a $Ca^{2+}$ sensor. J. Biol. Chem. **277:**16662–16672.

**25. Venters, R. A., L. M. Benson, T. A. Craig, K. H. Paul, D. R. Kordys, R. Thompson. S. Naylor, R. Kumar, and J. Cavanagh.** 2003. The effects of $Ca^{2+}$

binding on the conformation of calbindin $D_{28K}$: A nuclear magnetic resonance and microelectrospray mass spectrometry study. Anal. Biochem. **317:**59-66.

**26. Berggard, T., M. Silow, E. Thulin, and S. Linse.** 2000. $Ca^{2+}$- and $H^{+}$-dependent conformational changes of calbindin $D_{28K}$. Biochemistry. **39:**6864-6873.

**27. Venters, F. A., C. Huang, B. T. Farmer II, R. Trolard, L. D. Spicer, and C. A. Feirke.** 1995. High-level $^{2}H/^{13}C/^{15}N$ labeling of proteins for NMR studies. J. Biomol. NMR. **5:**339-344.

**28. Schauder, S. and B.L. Bassler.** 2001 The languages of bacteria. Genes and Dev. **15:**1468-1480.

**29. Xavier, K. B., and B. L. Bassler.** 2003. LuxS quorum sensing: more than just a numbers game. Curr Opin Microbiol. **6:**191-197.

**30. Bassler, B. L., M. Wright, R. E. Showalter, and M. R. Silverman.** 1993. Intercellular signaling in *Vibrio harveyi*: Sequence and function of genes regulating expression of luminescence. Mol. Microbio. **9:**773-786.

**31. Bassler, B. L, M. Wright, and M. R. Silverman.** 1994a. Multiple signaling systems controlling expression of luminescence in *Vibrio harveyi*: Sequence and function of genes encoding a second sensory pathway. Mol. Microbiol. **13:**273-286.

**32. Bassler, B. L, M. Wright, and M. R. Silverman.** 1994b. Sequence and function of LuxO, a negative regulator of luminescence in *Vibrio harveyi*. Mol. Microbiol. **12:**403-412.

**33. Martin, M., R. Showalter, and M. Silverman.** 1989. Identification of a locus controlling expression of luminescence genes in *Vibrio harveyi*. J. Bacteriol. **171:**2406-2414.

**34. Showalter, R. E., M. O. Martin, and M. R. Silverman.** 1990. Cloning and nucleotide sequence of *luxR*, a regulatory gene controlling bioluminescence in *Vibrio harveyi*. J. Bacteriol. **172:**2946-2954.

**35. Bassler, B.L.** 1999. How bacteria talk to teach other: Regulation of gene expression by quorum sensing. Curr. Opin. Microbiol. **2:** 582-587.

**36. Bassler, B. L., E. P. Greenberg, and A. M. Stevens.** 1997. Cross-species induction of luminescence in the quorum-sensing bacterium *Vibrio harveyi*. J. Bacteriol. **179:**4043-4045.

**37. Bax, A.** 2003. Weak alignment offers new NMR opportunities to study protein structure and dynamics. Pro. Sci. **12:**1-16.

**38. Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989 Molecular Cloning: A laboratory manual, Vol. 3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.