Protein Behavior in Crowded Environments

Lisa M. Charlton

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctorate of Philosophy in the Department of Chemistry.

Chapel Hill
2008

Approved by

Advisor: Professor Gary J. Pielak, Ph.D.

Reader: Professor Linda L. Spremulli, Ph.D.

Reader: Professor Matthew R. Redinbo, Ph.D.

Abstract

Lisa M. Charlton: Protein behavior in crowded environments
(Under the direction of Professor Gary J. Pielak, Ph.D.)

The cell's interior is a complex milieu where proteins exist in an environment crowded with other macromolecules. The influence of macromolecular crowding is prevalent in every cellular function—from metabolism to signal transduction to protein folding. It is even hypothesized that macromolecular crowding dictates the organization of the intracellular environment and the evolution of a single-cell species. Even though proteins are prevalent in every biochemical process, we lack fundamental knowledge about how crowding affects proteins. When studying proteins in a crowded environment, one can either develop a system that mimics the cellular interior or develop a technique that observes proteins inside cells. Herein, I report a residue-level interrogation on the stability of chymotrypsin inhibitor 2 under macromolecular crowded conditions. This is the first study that offers comprehensive information about the effects of crowding on the stability of the native state of a globular protein. I also present observations about the behavior of globular, partially-folded, and natively-disordered proteins inside living *E. coli* and the consequences for in-cell NMR. Lastly, I discuss our progress in transitioning from in-cell NMR for living *E. coli* to the yeast, *Pichia pastoris*.

Dedication

To my husband Matt…. I could not have done this without you.

Acknowledgements

Graduate school is not a road that is taken alone. I would like to thank those who have helped and supported me over the years. To Gary Pielak, my advisor and an excellent scientist, thank you for all the opportunities and for training me to think like a scientist. To my committee members, Drs. Matt Redinbo, Linda Spremulli, Susan Lord, and Ashutosh Tripathy, thank you for your advice and time. To past and present Pielak lab members, you have made this time both memorable and endearing and for this, thank you. To my undergraduate assistants, Hayley Fischer, Christopher Barnes, and Kristen Black, I cannot give them enough praise. These are exceptional students and friends who will excel in any career choice. Thank you for everything. To Sarah Kennedy, Erin Milner, Candi Cunningham, and Brian McNulty, you have been pillars of strength and encouragement during those moments in time when science doesn't work, thank you. To Jillian NMI Orans, a great friend and scientist, together, we make one damn good scientist. I would not have survived without you, thank you. To my family, thank you for your support. And lastly, thank you to Linda, I am not sure who you are, but you make a damn tasty drink.

Table of Contents

List of Tables

List of Illustrations

List of Abbreviations and Symbols

1D	one-dimensional
2D	two-dimensional
3D	three-dimensional
$1/K_{op}$	equilibrium constant for closing
AOX1	primary promoter for alcohol oxidase
amp	ampicillin
BMGY	buffered complex minimal glycerol media
BMMY	buffered complex minimal methanol media
CaS	1 M sorbitol, 10 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$
CD	circular dichroism
CI2	chymotrypsin inhibitor 2
CIP	calf intestinal phosphatase
cm	centimeter
Da	Dalton
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
EtBr	ethidium bromide
g	standard gravity
GdmCl	guanidinium hydrochloride
h	hour
His	histidine
his^-	lacking histidine

HSQC	heteronuclear single quantum correlation
Hz	hertz
IPTG	isopropyl β-D-1-thiogalactopyranoside
kan	kanamycin
kcal	kilocalorie
k_{ex}	intrinsic 1st order rate constant
k_{obs}	observed 1st order rate constant
K_{op}	equilibrium constant for opening
kb	kilobase
kDa	kilodalton
kV	kilovolt
LB	Luria broth
MD	minimal dextrose media
MHz	megahertz
min	minute
mg	milligram
mL	milliliter
mm	millimeter
MM	minimal methanol media
Mut^+	methanol utilization fast
Mut^S	methanol utilization slow
nm	nanometer
NEB	New England Biolabs

NEP	N-ethyl pyrrolidone
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
pDNA	plasmid DNA
PEG	polyethylene glycol
PEG/CaT	40% PEG 3350 (w/v); 20 mM Tris-HCl, pH 7.5, 20 mM $CaCl_2$
PMSF	phenylmethylsulphonyl fluoride
PVP	poly(vinyl pyrrolidone)
R	gas constant
RD	regeneration dextrose
rpm	revolutions per minute
R-T	room temperature
SCE	1M sorbitol, 1 mM EDTA, 10 mM sodium citrate, pH 5.8
sec	seconds
SDS	sodium dodecyl sulfate
SED	1M sorbitol, 25 mM EDTA, 1M DTT, pH 8.0
STAM2	signal transducing adaptor molecule
T	temperature
TE	Tris-EDTA
Tris	trishydroxymethylaminomethane
TSP	3-(trimethylsily)propionic-2,2,3,3-d_4 acid sodium salt

v/v	volume/volume
w/v	weight/volume
YFG	your favorite gene
YNB	yeast nitrogen base
YPD	yeast extract, peptone, dextrose

Greek Based Symbols

α-SN	alpha-synuclein
ΔG_{den}	free energy of denaturation
ΔG_{op}	free energy of opening
µg	microgram
µL	microliter
µm	micrometer

1 Introduction

1.1 Macromolecular crowding and the cell

As early as the late 1950s, science was interested in studying proteins and polysaccharides in concentrated systems (*1*). At that time, a highly characterized polysaccharide, hyaluronan, was shown to have an unexpected concentration dependent sedimentation rate. At infinite dilution the sedimentation coefficient depended on the molecular weight, but at higher concentrations, there was no molecular weight dependence. This behavior was also observed for other polymers (*1*). It was determined that hyaluronan formed a 3-dimensional, randomly-distributed chain network at high concentrations (>100 mg/ml) but behaved as a non-interacting, expanded, random coil in dilute solution (*1*).

The ability of hyaluronan to form a mesh-like network at high concentrations prompted further studies. Ogston and Phelps, in 1960, made equilibrium measurements of the partitioning of proteins between hyaluronan solution and buffer and observed that proteins favored the buffer. From this observation, they concluded that hyaluronan excluded proteins and solutes from part of the space, and this excluded volume increased with increasing hyaluronan concentration (*1, 2*). Not only was this phenomenon observed with other components, such as collagen and polyethylene glycol, but the results also agreed with theoretically calculated values

(*1, 3, 4*). These results were some of the earliest demonstrations of excluded volume theory.

Excluded volume theory states that two molecules cannot occupy the same space at the same time. Therefore, the presence of cosolutes, particularly in high concentrations, excludes volume to another solute molecule. In general, a cell's interior is 20-30% volume-occupied, thus, having 20-30% of its volume excluded to other molecules. These macromolecules have partial specific volumes close to 1 ml mg^{-1}, resulting in a crowded interior with macromolecules achieving concentrations of 200-300 mg/ml (*5*). The steric exclusion that results from macromolecular crowding has considerable consequences for a cell's function.

The high protein concentration in the cellular environment affects diffusion of both small and large molecules; however, the cell has evolved to overcome these diffusive barriers. With respect to metabolism, enzymes organize to form static, multi-enzyme complexes of related enzymes that co-localize with the respective small molecules (*6*). The formation of such complexes promotes metabolite channeling, the direct transfer of a small molecule from one enzyme to an adjacent enzyme without the need for aqueous phase diffusion (*7, 8*). This process helps metabolism proceed efficiently and overcome diffusive barriers. Macromolecular crowding is also a force for protein association (*9*). It plays a role in the formation of the multi-enzyme complexes. Westerhoff and colleagues showed that crowding reduces dissociation constants of enzyme complexes (*10, 11*), and Srere *et al.* linked high macromolecular concentrations to the organization of the cytoplasm into microenvironments for substrate channeling (*8, 11*). When considering both

conclusions, one may argue that macromolecular crowding promotes metabolite channeling, which is considered to be an important evolutionary driving force behind cellular function (*11*).

The mechanism by which macromolecules form discrete compartments in the nucleus can also be explained by macromolecular crowding. The concentration of macromolecules in the nucleus approximates 100 mg/mL (*12*). Within this environment, non-membrane bound, nuclear compartments, such as the nucleoli, form from the compartmentalization of their macromolecular components. Macromolecular crowding plays an essential role in the assembly and function of this compartment. For instance, nucleoli disassembled and nucleolar transcription decreased by 85% when nuclei isolated from human erythroleukemia (K562) cells were expanded in medium of low cation concentration. The macromolecular concentration decreased by approximately 2-fold upon nuclear expansion. After addition of inert, penetrating macromolecules (8 kDa polyethylene glycol or 10.5 kDa dextran) to a homeostatic concentration, the nucleoli reassembled and nucleolar transcription resumed (*12*). DNA replication in the nucleus also depends on a crowded environment. A nucleus must form for DNA replication to initiate and progress (*13*) and has been demonstrated *in vitro* that crowding is critical to assembly of replication machinery (*14*).

The influence of macromolecular crowding is prevalent in every cellular function—from metabolism to signal transduction to protein folding. It is even hypothesized that macromolecular crowding dictates the organization of the intracellular environment and the evolution of a single-cell species (*11*). The earliest

studies mimicked the cellular environment by crowding with inert macromolecules. These experiments are inherently flawed because the concentration of macromolecules inside cells is neither homogenous nor inert. Experiments evolved to look at small molecules inside living cells, providing extraordinary detail on the function of the metabolome (*15*). Now, the interest lies in monitoring the behavior of macromolecules inside living cells. This can be accomplished using a novel technique, in-cell NMR.

1.2 In-cell NMR

In-cell protein NMR is the first technique to give atomic-level resolution inside living cells. An in-cell NMR spectrum can be collected on a ^{15}N,^{13}C, or ^{19}F-enriched, overexpressed, soluble protein within the cell's interior (*16*). The first in-cell experiment on a protein was from the Brindle lab in 1997 (*17*). These investigations quantified the rotational mobility of three glycolytic enzymes in *Saccharomyces cerevisia*e by observing an unnaturally incorporated 5-fluorotryptophan. Four years passed before the technique was adapted by Serber *et al.* to heteronuclear single quantum correlation (HSQC) experiments, which detects ^{15}N with a bound proton (*18*). Serber's paper showed proof of concept. With the advent of this technique, not only could one begin to answer the important question—what is the behavior of macromolecules inside living cells?—but also could provide an answer with atomic-level resolution. Only a few groups, however, have attempted to incorporate this technique into their research.

The Pielak lab published the first paper using in-cell NMR to investigate a natively-disordered protein, FlgM (*19*). A ^{15}N HSQC spectrum was collected for FlgM inside living *E. coli*, and crosspeaks belonging to the C-terminal residues disappeared, whereas those belonging to the N-terminal residues remained. The data suggest that the C-terminal portion becomes structured inside cells. Exchange is occurring between a structured and unstructured C-terminal which leads to line-broadening and disappearance of crosspeaks. The Pielak lab then went on to study another natively-disordered protein, α-synuclein (α-SN) (*20*). McNulty *et al.* determined that the *E. coli* periplasm prevented a temperature induced conformational change that occurs at 35 ^{o}C in dilute solution (*20*). The crowded periplasm stabilized the disordered α-SN monomer. These two studies emphasized the need to examine proteins in their natural environment, particularly since many disordered proteins are involved in neurodegenerative diseases.

Hubbard *et al.* reported the first in-cell NMR results relating the structure of a regulatory protein to its interaction with a potential drug (*21*). The regulatory protein, CheY, is involved in chemotaxis in bacteria and binds both Mg^{2+} and Ca^{2+}. By comparing chemical shifts between spectra acquired in dilute solution and inside cells, Hubbard and colleagues concluded that CheY exists predominantly in the Mg^{2+}-bound conformation inside cells. They identified the residues that interact with the potential drug inside cells by observing chemical shift changes.

The Shekhtman group provided the first glimpse of protein-protein interactions from studies of ubiquitin and signal transducing adaptor molecule (STAM2) inside *E. coli* (*22*). This study resulted in residue-specific information about

the binding interface between STAM2 and ubiquitin, and the authors hypothesized an increase in binding affinity supported by the appearance of an additional binding interface inside cells. Shekhtman and colleagues observed chemical shift changes, line-broadening, and crosspeaks disappearing in regions of binding inside cells when compared to data obtained in dilute solution.

The Dötsch lab was the first to move away from detecting ^{15}N nuclei to observing ^{13}C inside *E. coli (23, 24)*. Their 2003 paper incorporated $^{15}N/^{13}C$ enriched histidine in the protein NmerA to measure the intracellular pH and the tautomeric state of the histidine by observing the chemical shift of histidine $^{13}C^{\varepsilon 1}$ or $^{13}C^{\delta 2}$ resonances (*24*). Their second paper in this area described a method for the selective observation of ^{13}C-enriched methyl groups of methionine and alanine residues in proteins (*23*). This method not only increases the sensitivity 3-fold compared to the observation of amide proton resonances but also provides an attractive alternative to monitoring large proteins and complexes intracellularly.

Over the span of one decade, in-cell NMR has been used to observe structural changes in individual proteins, protein-ligand interactions, and protein-protein interactions inside living *E. coli*. The next frontier is to develop in-cell NMR for proteins in eukaryotic cells.

The first description of an in-cell experiment on a eukaryotic cell can once again be attributed to the Brindle lab who used *S. cervesiae* (*17*). The Wagner lab, however, is credited with collecting the first ^{15}N-HSQC spectrum inside a higher eukaryotic cell, namely *Xenopus laevis* oocytes (*25*). Instead of endogenously expressing the protein and incorporating an NMR active nuclei, the Wagner lab

microinjected the purified, recombinant, ^{15}N-enriched GB1 domain protein from *E. coli* into the cytoplasm of frog oocytes. Microinjecting proteins into the cellular environment prevents nonspecific incorporation of NMR-active nuclei into background metabolites. This elimination of background signal permits collection of high resolution NMR data at low intracellular protein concentrations (~10 μM) (*26*). The limitations of using this technique are then only size and solubility of the protein.

Few groups have attempted to use in-cell NMR to monitor basic biophysical characteristics of a protein such as stability, dynamics, and diffusion. I focused on protein stability. There were two questions I attempted to answer with my research. Does the intracellular environment of *E. coli* stabilize the protein chymotrypsin inhibitor 2 as predicted by others? And does this stability correlate with the stability measured in the presence of macromolecular crowding agents? Chapter 2 presents the results of measuring the stability of variant chymotrypsin inhibitor 2 in 300 g/L polyvinyl pyrrolidone (PVP) using NMR-detected amide proton exchange. Chapter 3 addresses several problems we encountered with in-cell NMR as well as the development of an encapsulation method as a solution to these problems. Chapter 4 reports the progress in developing in-cell NMR for monitoring proteins in the yeast *Pichia pastoris.*

2 Measuring the stability of chymotrypsin inhibitor 2 under macromolecular crowded conditions

2.1 Introduction

2.1.1 Implications of a crowded environment

Macromolecular crowding is predicted to stabilize proteins (*27*). Even though proteins participate in every biochemical process and are used as therapeutics, we lack fundamental knowledge about crowding and protein stability. The consequences of macromolecular crowding are expected to be profound (*27*), yet, surprisingly, most biophysical studies are conducted only in dilute solution. Quantitative information about the effects of crowding on globular proteins would not only provide new insights into intracellular constraints on protein stability, but also help find formulations that increase the stability and facilitate the long-term storage of protein pharmaceuticals.

As protein function in the cell occurs at macromolecular concentrations of 300 g/L or greater (*28*), relevant experiments must attempt to mimic this environment. Sequestering a protein in a biocompatible, protein-like polymer provides a simplified representation of the cellular interior that focuses on the potential stabilizing effect of steric repulsion (*29, 30*). A study conducted by Minton and coworkers showed that crowding with dextran (mw 35,000 Da) increases the equilibrium constant for folding of the molten-globule state of cytochrome *c* by 100-fold at pH 2.0 (*30*). This observation

shows crowding can promote stabilization. These data, however, were acquired under extreme conditions, did not involve the native state of the protein, and focused on bulk rather than residue-specific measurements.

Other studies of protein stability in the presence of macromolecules have reported minimal increases in stability (*31-34*). These reports may underestimate the increase in stability because such systems often ignore aggregation and do not focus on stable native states (*29, 35*). We have developed a simple system that provides information about the effects of macromolecular crowding on protein stability. Our system gives unambiguous data about the effects of a polymeric crowding agent on the stability of each residue in a globular protein. This quantitative information offers new insights into macromolecular stabilization of proteins.

2.1.2 Chymotrypsin inhibitor 2

We chose chymotrypsin inhibitor 2 (CI2) as our model globular protein because this protease inhibitor is small (7.5 kDa), amenable to biophysical study, and well-studied in terms of its structure, folding, and stability (*36-38*). More specifically, we used the I29A:I37H variant because the destabilizing isoleucine-to-alanine change (*39*) allows stability measurements to be completed in 24 h, and the histidine side chain provides an internal pH probe. We chose 40 kDa polyvinyl(pyrrolidone) (PVP40) as the macromolecular crowding agent because this random-coil polymer is extremely soluble, has protein-like solution properties, and is metabolically inert (*40, 41*). PVP has been widely used in pharmaceutical applications (*42, 43*), and it retards both protein aggregation and proteolytic

degradation (*44-46*), but it's usefulness as a stabilizing agent for small proteins has never been investigated.

2.2 Materials and Methods

2.2.1 Purification of chymotrypsin inhibitor 2

The pet28a plasmid (Novagen) containing the gene for truncated wildtype CI2 was a gift from Dr. Andrew Lee (UNC). The first 19 residues of full length CI2 are disordered and are not included in this construct. Residue 20 of full-length CI2 is referred to here as residue 1. Site-directed mutagenesis was used to incorporate the I29A and I37H mutations (Stratagene kit). The forward primer for I29A was 5'GAA GCG CAG **GCA** ATC GTG C 3', where the changed codon is bolded. The forward primer for I37H was 5' CT GCC GGT GGG **CAC** CCA TGT GAC CAT GGA ATA TC 3'. The mutated plasmid was transformed into DH5α *E. coli* cells. DNA sequence analysis (Genome Analysis Facility, UNC) with the T7 forward sequencing primer was used to confirm the sequence.

After sequence analysis, the plasmid was transformed into BL-21(DE3-Gold) competent *Escherichia coli* cells (Stratagene). The transformants were spread onto Luria broth agar plates containing 60 µg/mL kanamycin (LB_{kan}). Twenty-five mL of ^{15}N-enriched Spectra 9 media (Spectra Stable Isotopes) were inoculated with a single colony and incubated overnight at 37 ^{o}C with shaking at 200 rpm. This overnight culture was diluted into 1 L of ^{15}N-enriched Spectra 9 media in a 6 L autoclaved flask. The culture was grown at 37 ^{o}C with shaking at 200 rpm to an

absorbance at 600 nm of 1.0. The cells were induced with isopropyl β-D-1-thiogalactopyranoside (Sigma) at a final concentration of 1 mM. The cells were harvested after 5 h by centrifugation at 7300 x g in a swinging bucket centrifuge for 20 min.

The cell pellet was resuspended in 25 mL of 50-mM Tris buffer, pH 8.0. The slurry was sonicated (Fisher Scientific, Sonic Dismembrator Model 500) on ice (4 cycles, 4 min each, 30% duty cycle). The lysate was centrifuged at 20,000 x g for 30 min. The pellet was discarded, and the supernatant treated with a 10% solution of polyethyleneimine (final concentration 0.2%). After incubation on ice with stirring for 30 min, the precipitated nucleic acids were removed by centrifugation at 20,000 x g for 30 min.

The clear lysate was loaded onto an anion-exchange column (Q Sepharose® high performance resin, GE Healthcare, 1.6 cm x 10 cm) connected to an AKTA FPLC (GE Healthcare) at 4 ^{o}C. The column was equilibrated with 25-mM Tris buffer, pH 8.0, before loading the sample. After loading, the column was thoroughly washed with 100 mL of 25-mM Tris, pH 8.0. Cl2 does not interact strongly with the column material and was found in the flow-through and wash. Purity was assessed by using SDS-PAGE on an 18% gel with Coomassie staining. If needed, the protein was further purified in H_2O by size-exclusion chromatography (HiLoad 16/60 Superdex 75 column, 1.6 cm x 60 cm, GE Healthcare) at 4 ^{o}C. The protein eluted between 80 and 100 mL. Protein concentration was determined by using the absorbance at 280 nm and an extinction coefficient of 7.04 x 10^3 $M^{-1}cm^{-1}$ (*47*). Immediately following the purification, the protein was lyophilized and checked for

structural changes by comparing chemical shifts with nonlyophilized protein. No significant change was observed. Typical yield of pure protein is 100 mg per 1 L culture.

2.2.2 Sample preparation for nonexchange experiments

Lyophilized protein was resuspended in 1 mL of 50-mM sodium acetate buffer (90% H_2O/ 10% D_2O, v/v), pH 5.4, containing 50, 100, 200, or 300-g/L 40 kDa poly(vinylpyrrolidone) (PVP40, Fisher). The final protein concentration was 800 μM. Consecutive, 1 h ^{15}N HSQC spectra were collected at 37 ^{o}C. The data were analyzed for changes in peak volumes, line-widths, and chemical shifts.

2.2.3 NMR conditions for nonexchange experiments

Two-dimensional, gradient-enhanced ^{15}N HSQC spectra (*48, 49*) were acquired in dilute solution and in 300-g/L PVP40. Lyophilized protein was dissolved in 50-mM sodium acetate buffer, pH 5.4 to a final concentration of 700 μM for dilute solution and in 50-mM sodium acetate, 300-g/L PVP40, pH 5.4 to a final concentration of 800 μM. The data were collected on a Varian Inova 700-MHz spectrometer equipped with a coldprobe. The ^{1}H dimension was acquired with a sweep width of 11499 Hz and consisted of 1024 complex points. The ^{15}N dimension was acquired with a sweep width of 2188 Hz and consisted of 60 complex increments. Twenty-four transients were collected for each spectrum. The data were processed with NMRPipe and NMRDraw (*50, 51*).

^{1}H and ^{15}N resonance assignments were obtained from hncacb and cbca(co)nh spectra (*52, 53*) on a 1-mM sample of ^{13}C-^{15}N enriched protein (Table 1).

2.2.4 NMR

For exchange experiments, the spectrometer was first shimmed by using a protein sample in 50-mM sodium acetate buffer, pH 5.4, at 37 ^{o}C. The exchange reaction mixture was prepared by dissolving lyophilized protein in 1 mL of 50-mM sodium acetate buffer, pH 5.4, to a final concentration of 700 μM, followed by centrifugation and transfer of the supernatant to a 5-mm NMR tube. This buffer and those described below were made with 99.9% D_2O (Acros Organics). The sample was equilibrated in the spectrometer at 37 ^{o}C for 15 min before collecting the first spectrum. Consecutive HSQC spectra were collected for 24 h. The dilute-solution samples were saved for further experiments with GdmCl-induced denaturation. For the crowding study, lyophilized protein was resuspended in 50-mM sodium acetate buffer, 300-g/L, 100-g/L, or 50-g/L PVP40, pH 5.4 to a final concentration of 800 μM. Shimming and data acquisition were performed as described above. The pH of each sample was checked before and after data collection using a Fisher brand pH meter and electrode. The reported pH value for deuterated solutions is uncorrected for the isotope effect. Peak volumes were obtained by using NMRview (*51*).

2.2.5 Data analysis

Exchange data were processed in SigmaPlot. For the majority of residues, the decay in crosspeak volume versus time was fit to a three-parameter exponential decay:

$$I = Aexp(-k_{obs}t) + C,$$

where *I* represents the normalized peak volume, *t* is time in min, *A* is the amplitude of the exchange curve, k_{obs} is the observed exchange rate determined from the decrease in signal *versus* time, and *C* is the constant taking into account the residual water. Due to crosspeak overlap, residues 47/27, 8/11, 55/56 were fit a five-parameter exponential decay

$$I = Aexp(-k_{obs}t) + B(-k_{obs2}t) + C$$

The intrinsic rate constants for exchange, k_{ex}, were determined for each residue at the designated pH and temperature using the program Sphere (*54*).

2.2.6 Measuring k_{ex} using CLEANEX-PM

The intrinsic rate of exchange (k_{ex}) for amide protons in the unprotected loop of the I29A:I37H variant was measured by using the water saturation transfer experiment, CLEANEX-PM (*55*). Samples were prepared to a final concentration of 800-μM protein in 50-mM sodium acetate buffer, pH 5.4 alone or in the presence of 300-g/L PVP40, or 100-g/L N-ethylpyrrolidone (NEP). The pH of each sample was determined before and after data collection using a pH meter. The data were collected on a Varian Inova 600-MHz spectrometer at 37 ^{o}C. The mixing times were 0, 10, 20, 40 and 53 ms for dilute solution, 0, 10, 19, 25, and 53 ms for PVP40, and 0, 10, 19, 25, and 35 ms for NEP. NMR data were processed with NMRPipe (*50*) and analyzed with NMRview (*51*). Data were fit as described by Hwang *et al.* (*55*)

2.2.7 GdmCl equilibrium denaturation

Denaturation of the I29A:I37H variant of CI2 was measured by using circular dichroism (CD) spectrapolarimetry (Aviv Model 62DS) at 37 ^{o}C and 234 nm. The NMR samples were diluted to 20 μM with 50-mM sodium acetate buffer made with D_2O, pH 5.4. One mL of the sample was transferred to a 1-cm pathlength quartz cuvette. A solution of 20 μM protein in ~7 M GdmCl, 50-mM sodium acetate in D_2O was automatically titrated into the sample. The refractive index was measured at the end of each trial with a Palm Abbe refractometer (Misco). The GdmCl concentration was determined by using the equation:(*56*)

$$[\text{GdmCl}] = [(n-1.3344)/0.0161] \text{ mol/L},$$

where n is the refractive index.

The CD signal was plotted *versus* GdmCl concentration and fitted to the following equation by using SigmaPlot (Systat Software):

$$[\Theta] = \frac{m_n * [\text{GdmCl}] + b_n + (m_d * [\text{GdmCl}] + b_d)e^{-m*([\text{GdmCl}]-C_m)/R*T}}{1 + e^{-m*([\text{GdmCl}]-C_m)/R*T}},$$

where $[\Theta]$ is the ellipticity at a particular [GdmCl], m_n and b_n are the slope and intercept of the native baseline, m_d and b_d are the slope and intercept of the denatured baseline, C_m is the [GdmCl] at the transition midpoint, R*T is the product of the gas constant and absolute temperature. The free energy of denaturation $\Delta G^{\circ'}_{den}$ is calculated from:

$$\Delta G^{\circ'}_{den} = C_m * m,$$

where m is $\Delta\Delta G^{\circ'}_{den}/\Delta[\text{GdmCl}]$.

2.2.8 Measuring the diffusion coefficient

NMR experiments were performed on a 600-MHz Varian Inova spectrometer equipped with a triple resonance probehead and xyz gradients at 37 °C on a 1 mM sample of CI2 in 50-mM acetate buffer pH 5.4, 10% D_2O. An stimulated echo (STE) pulse sequence (*57*) with an x-gradient was employed to measure the diffusion of the small molecule reference compound, the sodium salt of trimethylsilyl-1-propane-sulfonic acid (TSP). The water signal was suppressed by using a selective pulse (sel-STE). Convection was suppressed by application of an x-gradient (*57*). The methyl proton signal of TSP was selectively excited in the diffusion measurements. Due to the short proton T_1 (~30 ms) and T_2 (~6 ms) values for CI2 in PVP40 (fast decay gave a low signal-to-noise ratio) a spin echo or stimulated echo was not suitable for assessing the diffusion of CI2 in PVP40. A long ^{15}N T_1 (~ 1.4 s) was utilized to determine the diffusion of CI2 in PVP40 by a heteronuclear stimulated echo (X-STE) as described by Ferrage *et al.* (*58*). Several pulse sequences were used to validate these measurements. [These pulse sequences included sel-STE (*57*), X-STE (*58*) and WIF-PGSTE (*59*).] The measurements agreed within 5%.

Ten 1D HSQC spectra were collected as a function of gradient strength. The logarithm of the peak intensities were plotted *versus* gradient strength, and fitted to the following equation

$$\ln(I_t) = \ln(I_o) - \gamma^2 G^2 \delta^2 (\Delta - \delta/3),$$

where I_o is the initial intensity, γ is the gyromagnetic constant, G and *δ* are the magnitude and duration of the field-gradient pulses, and Δ is the delay between gradient pulses (*57*). The slope of the line is the self-diffusion coefficient, D.

A 50 ms delay between gradient pulses was used for TSP, and 200 ms and 800 ms were used for CI2 in buffer and PVP40, respectively. Gradient strengths ranged from 480 G/m to 5800 G/m. The integrated area between 8 and10 ppm was used to monitor the signal decay as a function of gradient strength. The intensities were normalized to the first experiment with weakest gradient strength. Data were processed and analyzed by using VNMR (Varian) and NMRpipe.(*50*) Origin lab software was used to fit the NMR diffusion data.

2.3 Results and Discussion

High concentrations of crowding agents might alter a globular protein's structure and aggregation state (*60*), thereby confounding interpretation of their effects on stability. NMR chemical shifts are sensitive, empirical indicators of structure. We found PVP40-induced changes in ^{1}H and ^{15}N backbone chemical shifts only in the loop and turn regions of CI2 (Figure 2.1). Small changes in loops and turns are expected because crowding causes compaction and, unlike the protein core, loops are not by definition maximally compact (*61, 62*). These chemical shift changes might also reflect weak interactions between PVP and these surface regions. We assessed the aggregation state of CI2 by quantifying the effect of PVP40 on the protein's extinction coefficient, ^{1}H and ^{15}N resonance line-widths, and self-diffusion (Figures 2.2 and 2.3). A chemical reaction, such as dimerization, should alter the protein's molar extinction coefficient. The extinction coefficient, however, does not change in the 300-g/L PVP40 solution over the 24 h required to conduct a stability measurement.

An increase in line-widths would signify an increase in the molecular weight of the protein, particularly aggregation. Small increases and decreases in line-widths were observed in dilute solution and in 300-g/L PVP40 (Figure 2.2B). Since the changes were both positive and negative, instead of solely positive, these changes are attributed to random error introduced during the experiment.

Lastly, the diffusion data are consistent with the monomeric nature of the protein in both dilute and crowded solutions. The ratio of the self-diffusion coefficient (D) of CI2 in buffer to CI2 in 300-g/L PVP is 12; whereas, the ratio of D for the small molecule standard, trimethylsilyl-1-propane-sulfonic acid sodium salt (TSP), in buffer compared to 300-g/L PVP is 6 (Figure 2.3). The ratio of these two values is the oligomeric state of CI2 in PVP, which suggests CI2 is a dimer. Diffusion measurements are interpreted using the Stokes-Einstein equation which takes into consideration the viscosity of the solution. By comparing D_{TSP} to D_{CI2}, one can eliminate the need to measure the viscosity of the solution. However, TSP was determined to be an inappropriate standard for this experiment because of the large size difference between PVP40 and the small molecule. Therefore, we can conclude that CI2 is no more than a dimer in 300-g/L PVP as measured by the self-diffusion.

We conclude CI2 does not aggregate in the presence of 300-g/L PVP. These results indicate that the combination of PVP40 and CI2 is amenable to stability studies.

We used amide-proton exchange to quantify stability. In our experiments, the protein is lyophilized and then re-dissolved in a D_2O-containing solution (buffer alone

or buffer plus PVP40). The exchange rate of backbone amide protons for deuterons, k_{obs}, is quantified by measuring the decrease in the signal from individual amide protons in serially-acquired NMR spectra over 24 h. The exchange reaction is described by equation 1 (*63*).

$$\text{closed} \underset{k_{cl}}{\overset{k_{op}}{\rightleftharpoons}} \text{open states} \xrightarrow{k_{ex}} \text{exchange} \qquad \textbf{(1)}$$

The native state of the protein opens and closes with rate constants k_{op} and k_{cl}. Exchange occurs in the solvent-exposed open state with the rate constant k_{ex}. The generally accepted view is that the open states are ensembles whose subpopulations range from small, low amplitude fluctuations of the native state to rare, globally unfolded forms (*64, 65*).

As shown in Figure 2.4, the 300-g/L PVP40 solution exhibits considerably slower exchange compared to dilute solution. This result is consistent with crowding-induced stabilization, but quantitative analysis of protein stability requires information about the rate constants in equation 1.

The observed first-order rate constants for exchange at individual residues, k_{obs}, can be used to assess stability provided $k_{op}/k_{cl} << 1$ and $k_{cl} >> k_{ex}$. Under these conditions,

$$k_{obs} = (k_{op} / k_{cl})k_{ex} = K_{op}^{\circ'} k_{ex} \qquad \textbf{(2)}$$

where $K_{op}^{\circ'}$ is the equilibrium constant for opening.(*66*) Stability is measured as the free energy of opening, $\Delta G_{op}^{\circ'}$, where

$$\Delta G_{op}^{\circ'} = -RT \ln K_{op}^{\circ'} . \qquad \textbf{(3)}$$

and R is the gas constant and T is the absolute temperature. We know that k_{op}/k_{cl}

<<1 for the slowly-exchanging protons because independent unfolding experiments give $k_{op}/k_{cl} \sim 10^{-3}$ (Figure 2.5). To show that $k_{cl}>>k_{ex}$, we measured the pH dependence of k_{obs} (*66*). If $k_{cl}>>k_{ex}$ is true, then k_{obs} is dependent on k_{ex} which is acid/base catalyzed (*66*). Therefore, if $k_{cl} >> k_{ex}$, plotting k_{obs} values at one pH against values at another pH will give a line of unitary slope (*36*). As anticipated, plots for experiments conducted in dilute solution and in 300-g/L PVP40 at two pH values (Figure 2.6) exhibit unitary slopes. The data show that k_{op}/k_{cl} will equal K_{op} and can be used to assess stability in dilute solution and in 300-g/L PVP40, but using equations 2 and 3 to assess the effect of crowding requires knowledge of k_{ex} in dilute and crowded conditions.

Values of k_{ex} in dilute solution are computed from data on unstructured peptides (*54, 67, 68*). To assess the effect of 300-g/L PVP40 on k_{ex}, we used residue 37 in CI2's unprotected, extended loop (residues 35-45) as a mimic for an unstructured peptide. Exchange in the loop occurs too quickly to use the method described above, so we turned to the saturation transfer experiment, CLEANEX-PM (*55*). Those data indicate that k_{ex} in 300-g/L PVP40 equals k_{ex} in dilute solution (Figure 2.7). With these data in hand, we calculated $\Delta G^{\circ'}_{op}$ values at each residue of CI2 and assessed the effects of crowding on stability.

A histogram of $\Delta G^{\circ'}_{op}$ values in dilute solution and 300-g/L PVP40 as a function of CI2 sequence position is shown in Figure 2.8 ,and $\Delta G^{\circ'}_{op}$ values are superimposed on the protein structure (PDB 2CI2) in Figures 2.9A and 2.9B. PVP40 increases $\Delta G^{\circ'}_{op}$ for all measurable residues with a maximum increase of ~3 kcal/mol, which corresponds to a 100-fold increase in the equilibrium constant for folding

($1/K_{op}^{\circ'}$). This agrees with the predicted increases of 10- to 100-fold (*27*). We attempted to confirm the maximum $\Delta G_{op}^{\circ'}$ values by performing a circular-dichroism detected denaturation in PVP solutions supplemented with the denaturants guanidinium chloride or urea (Figure 2.10). Unfortunately, PVP interacts strongly with the cosolutes. Such behavior has been observed previously (*41*). Nevertheless, we feel our results are valid because the PVP concentration dependence of $\Delta G_{op}^{\circ'}$ values extrapolates smoothly to zero PVP concentration (*vide infra*), and many previous studies have shown correspondence between maximum $\Delta G_{op}^{\circ'}$ and free energies derived from global denaturation experiments (*64*).

To show the increases in $\Delta G_{op}^{\circ'}$ arise from the macromolecular nature of the crowding agent, we repeated the experiment in a solution of a model monomer for PVP, N-ethylpyrrolidone (NEP). CI2 precipitates in 300-g/L NEP, but is soluble in 100-g/L NEP. At this concentration, k_{ex}, as measured with the CLEANEX-PM experiment, increases four-fold compared to dilute solution (Figure 2.7). We used this NEP-observed value to scale the calculated k_{ex} values. Values of $\Delta G_{op}^{\circ'}$ in 100-g/L NEP are overlaid on the protein structure in Figure 2.9D. NEP has a small effect, with most changes being destabilizing compared to dilute solution. Higher NEP concentrations are expected to be even more destabilizing. These observations show that the stabilizing effect of PVP arises from its macromolecular nature and agrees with data showing that the effect of polyethylene glycol on the overall stability of lysozyme exceeds the effect of PEG's monomer, ethylene glycol (*69*).

Measuring $\Delta G_{op}^{\circ'}$ on a residue-by-residue basis provides insight into the native and open states of CI2. Since the magnitude of $\Delta G_{op}^{\circ'}$ in dilute solution depends on the amount of surface exposed upon unfolding (*70*), it is expected, and observed (Figure 2.9A), that the smallest $\Delta G_{op}^{\circ'}$ values tend to correspond to surface-exposed residues, and the largest values correlate with residues that become exposed only upon extensive unfolding. Identical patterns of local and global unfolding are observed in PVP40 and NEP compared to dilute solution (Figures 2.9B and 2.9C). These data show that PVP40 significantly increases the stability of globally-unfolding residues compared to dilute solution and NEP while leaving the stability of locally-unfolding residues nearly unchanged. From these observations, we conclude that crowding with PVP40 does not alter the exchange-competent forms (i.e., the open states) but decreases their probability.

Theories used to explain the effects of stabilizing (or destabilizing) cosolutes on protein stability predict that $\Delta G_{op}^{\circ'}$ should depend not only on differences in surface area but also on cosolute concentration (*27, 70-72*). We expect a positive correlation between $\Delta G_{op}^{\circ'}$ values and the differential quantity, $\partial \Delta G_{op}^{\circ'} / \partial$[PVP], which describes the sensitivity of $\Delta G_{op}^{\circ'}$ to PVP40 concentration. To test this idea we examined the effect of PVP40 concentration on $\Delta G_{op}^{\circ'}$ and mapped values of $\partial \Delta G_{op}^{\circ'} / \partial$[PVP] onto the structure (Figure 2.9D). Comparing Figure 2.9B to Figure 2.9D shows exactly the correlation predicted by theory; residues with the largest $\Delta G_{op}^{\circ'}$ values have the largest $\partial \Delta G_{op}^{\circ'} / \partial$[PVP] values.

2.4 Conclusions

We used a simple system comprising a globular protein, a random-coil polymer, and an established equilibrium thermodynamic method to quantify the stabilizing effects of macromolecular crowding at the level of individual residues. The maximum stabilizing effect is sizable, up to ~3 kcal/mol, which represents an increase in stability of ~50% compared to dilute solution. Our residue-by-residue measurements show that crowding has the largest effects on opening reactions that expose the most surface area but has little or no effect on regions that are solvent exposed in both the native and open states. This observation reinforces the assertion that macromolecular crowding stabilizes the native state by destabilizing the unfolded states. This observation also suggests that information about turns and unstructured loops gained under equilibrium conditions in dilute solution are valid under crowded conditions. Furthermore, the significant increase in protein stability as a result of crowding with a biocompatible polymer could lead to new approaches in the design and storage of protein pharmaceuticals.

2.5 Figures

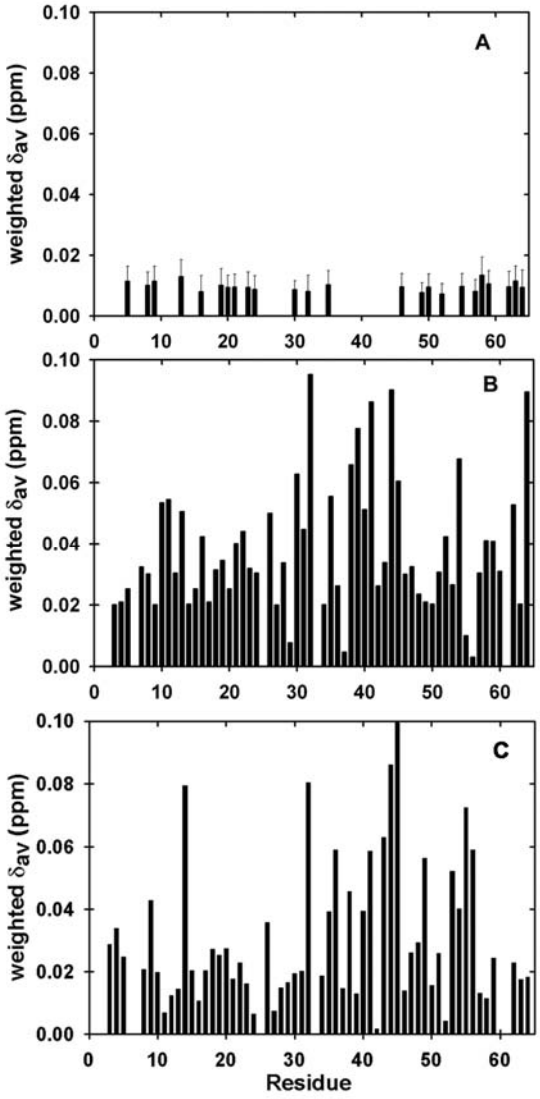

Figure 2.1. Plots of chemical shift disturbances in 300 g/L PVP and 100 g/L NEP.
Panel A: The average weighted chemical shift difference, with error bars representing the standard deviations from three trials, for the I29A:I37H variant of CI2 in 300 g/L PVP40 under exchange conditions. Weighted chemical shift differences were calculated by using the formula $[(\Delta^{1}H\ ppm)^2 + (\Delta^{15}N\ ppm \times 0.154)^2/2]^{1/2}$(*73*). These data provide an estimate of the precision for measuring chemical shifts. Panel B: The weighted chemical shift difference between HSQC spectra acquired in 300 g/L PVP40 and dilute solution. Weighted differences greater than 0.02 ppm are considered significant. Panel C: The weighted chemical shift difference between HSQC spectra acquired 100 g/L NEP and dilute solution.

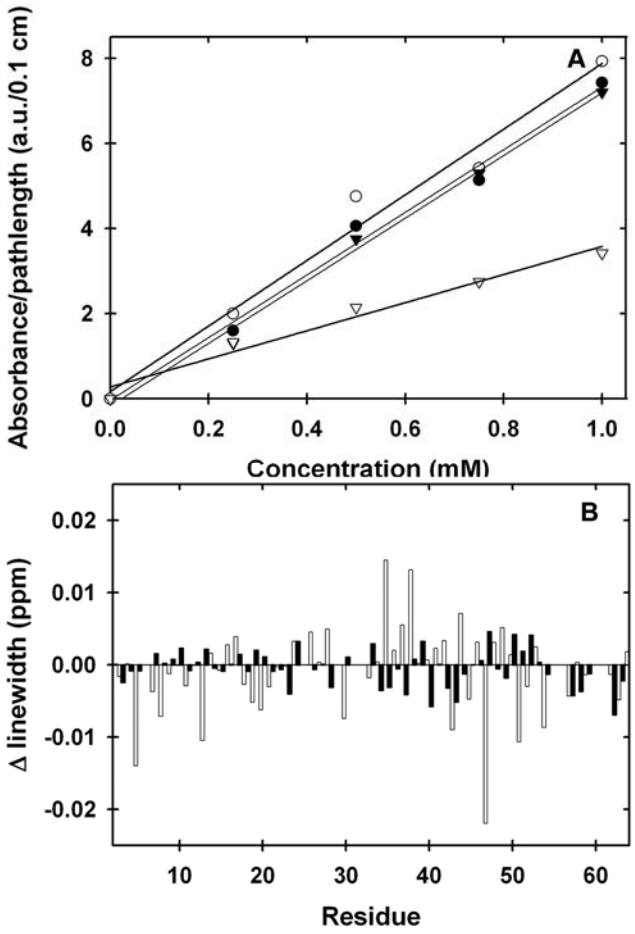

Figure 2.2. Plots of absorbance at 280 nm and linewidth changes in 300 g/L PVP.

Panel A: Dependence of the 280-nm absorbance on protein concentration in dilute solution, in 300 g/L PVP40, and in 400 g/L dextran at 0 and 24 h for the for the I29A:I37H variant of CI2. The extinction coefficient is 7.4 ± 0.3 $mM^{-1}cm^{-1}$for dilute solution (▼), 7.4 ± 0.4 $mM^{-1}cm^{-1}$ for 300 g/L PVP40 at 0 h (●), 7.7 ± 0.6 $mM^{-1}cm^{-1}$ for 300 g/L PVP40 at 24 h (o), and 3.3 ± 0.3 $mM^{-1}cm^{-1}$ for 400 g/L dextran (∇). The dextran data show that this technique is sensitive to aggregation. The published coefficient for CI2 is 7.0 $mM^{-1}cm^{-1}$(*47*). The similar values for dilute solution and PVP40 are consistent with the protein being a monomer in PVP40. **Panel B**: The change in linewidths at half-height (ppm) for the ^{1}H (white bars) and ^{15}N (black bars) backbone amide resonances at 0 and 24 h in 300 g/L PVP40 under nonexchange conditions. The observation of both small positive and small negative changes suggests the differences are not the result of protein aggregation. Residues 55/56 and 29/31 were not well enough resolved to determine linewidths.

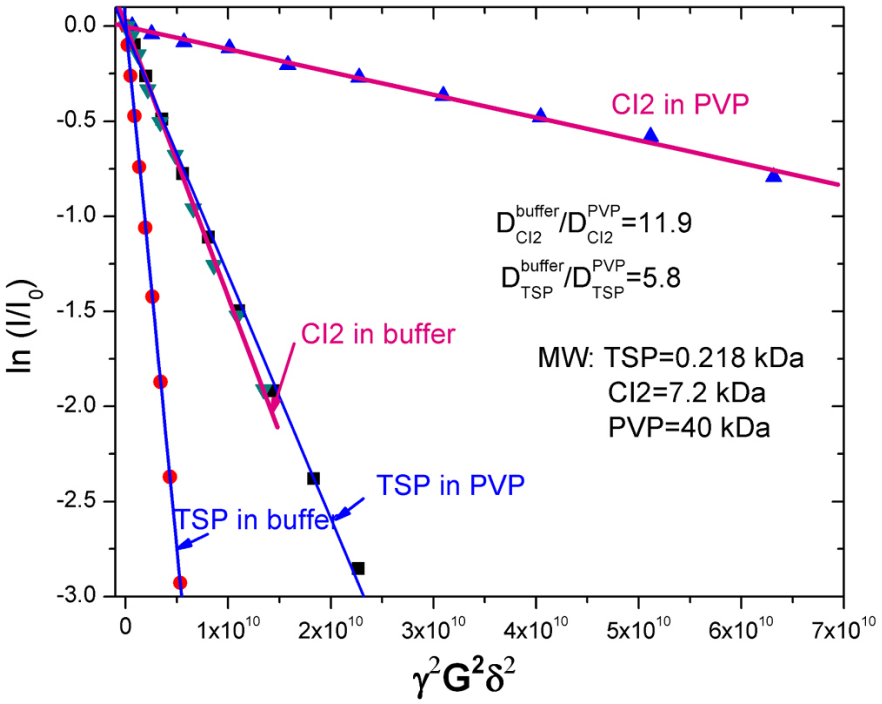

Figure 2.3. Molecular diffusion of the I29A:I37H variant in dilute solution and in 300 g/L PVP40, 50 mM acetate buffer, pH 5.4, 37 °C.
Using trimethylsilyl-1-propane-sulfonic acid sodium salt (TSP) as a reference small molecule, the diffusion coefficient of CI2 was 2 times greater in 300 g/L PVP40 than dilute solution suggesting dimerization. However, the measured diffusion of TSP in PVP40 results is an *apparent* diffusion coefficient ratio that is most likely an overestimation of the true value (*74*). We can presume that CI2 is a monomer in crowded solution, but can conclusively state that it is no more than a dimer.

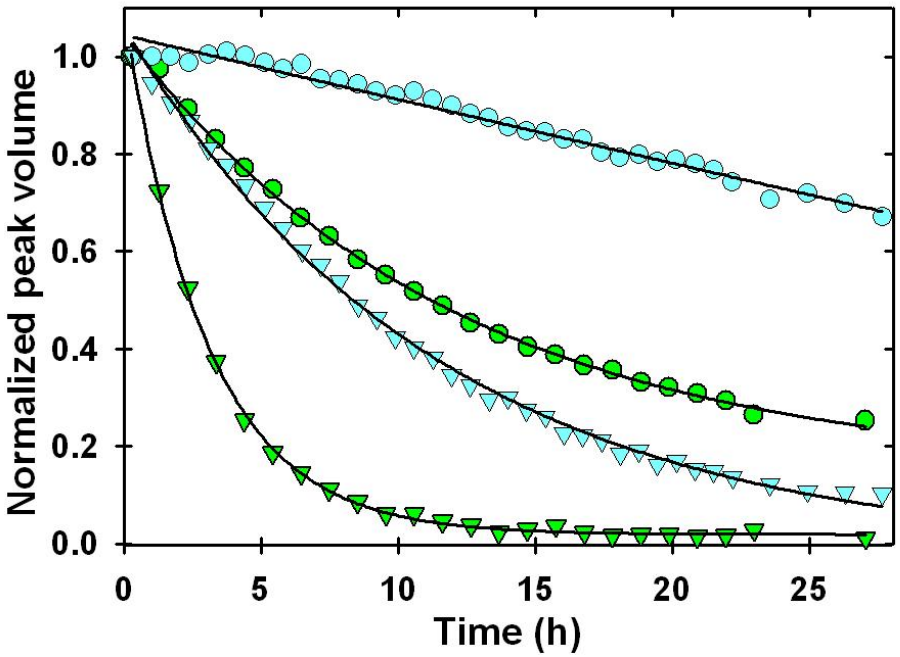

Figure 2.4. Exchange curves for Leu 32 (●) and Asp 52 (▼) in dilute solution (blue) and 300 g/L PVP (green).
PVP40 slows amide proton exchange. Exchange curves for the amide protons of Leu 32 (●) and Asp 52 (▼) in dilute solution (green) and 300 g/L PVP40 (cyan). Conditions: 50 mM d_3-acetate buffer, pH 5.4, 37 ^{o}C.

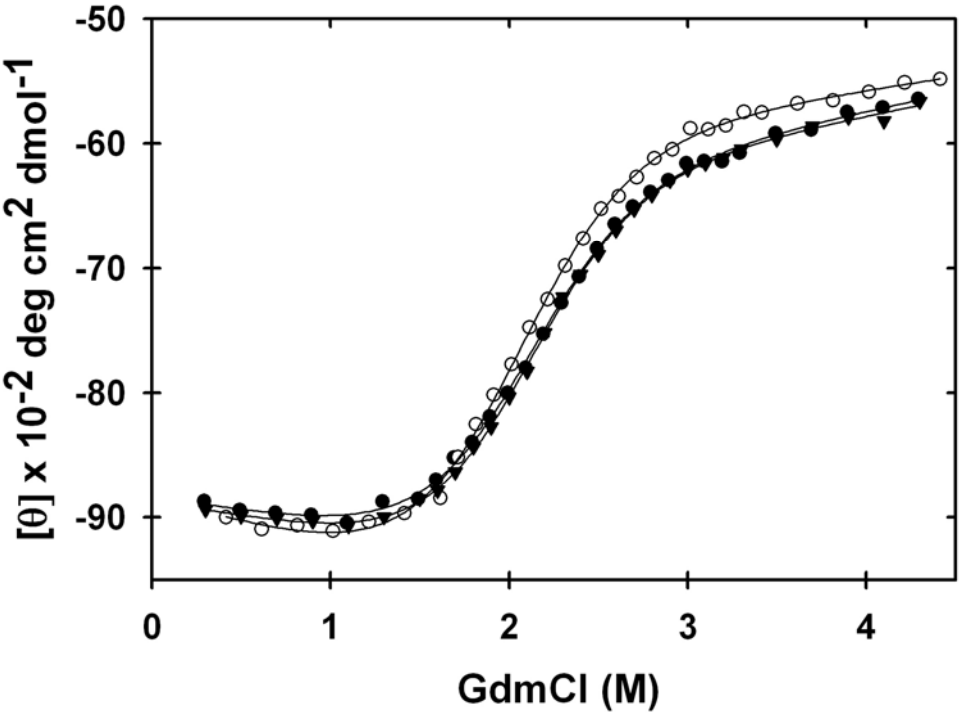

Figure 2.5. GdmCl denaturation curves of the I29A:I37H variant at 37 oC for three trials measured by circular dichroism at 234 nm.
The sample was 20 μM protein in 50 mM acetate buffer made with D_2O. The average m value is 1.97 ± 0.04, the average $[GdmCl]^{50\%}$ is 2.05 ± 0.04, and the average ΔG_{den} is 4.0 ± 0.2 kcal/mol.

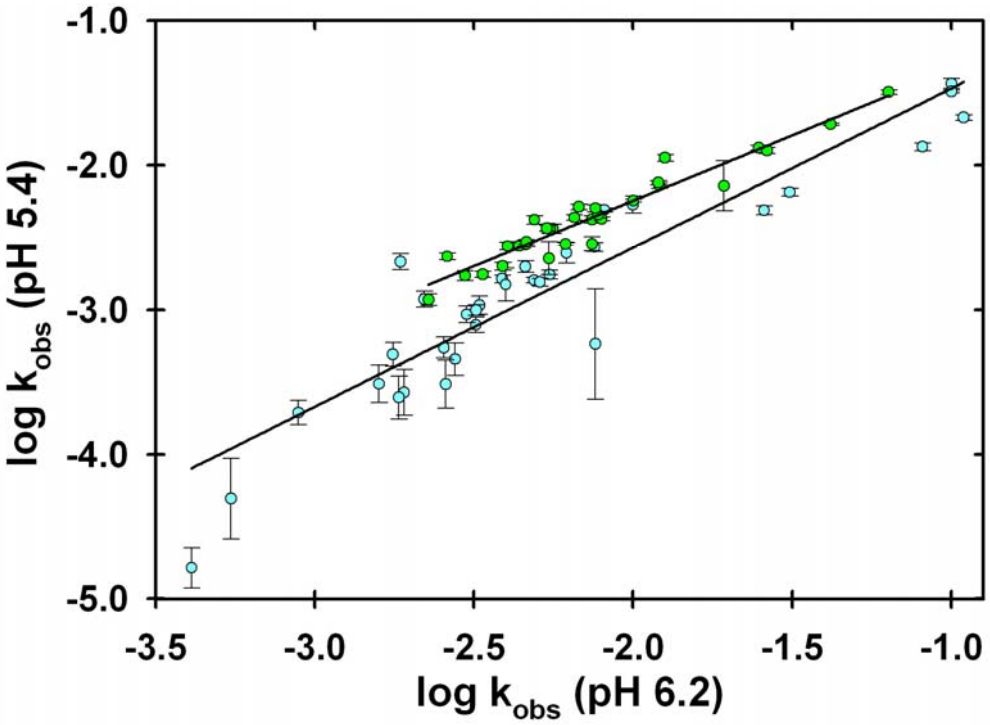

Figure 2.6. Plot of log kobs (pH 5.4) vs. log kobs (pH 6.2) to determine EX2 mechanism.

The I29A:I37H variant exchanges according to the EX2 mechanism. Linear regression yields slopes and R^2 values of 0.91 ± 0.5 and 0.92 in dilute solution (green) and 1.10 ± 0.08 and 0.86 in 300 g/L PVP40 (cyan), respectively. The error bars represent the standard errors for the averages from three trials collected at pH 5.4.

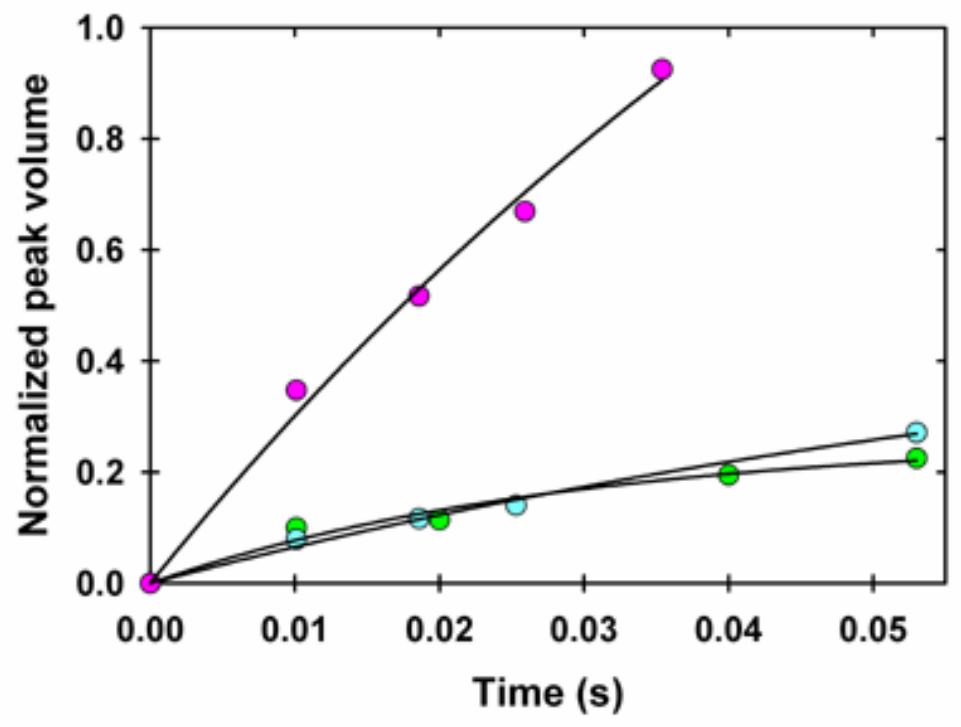

Figure 2.7. Exchange of Residue 37 using CLEANEX-PM
The exchange of residue 37 was measured in 50 mM acetate, pH 5.4 at 37°C in 300 g/L PVP40 (cyan), in dilute solution (green), and in100 g/L NEP (magenta) using the CLEANEX-PM experiment (*55*). The rate is 9 ± 2 s^{-1} in dilute solution, 7 ± 1 s^{-1} in PVP40, and 32 ± 3 s^{-1} in NEP. The smooth curve is determined as described by Hwang *et al.*(*55*) Crowding with 300 g/L PVP40 does not change the rate of exchange for loop residue 37, which is essentially unprotected, but adding 100 g/L NEP increases the rate approximately 4-fold.

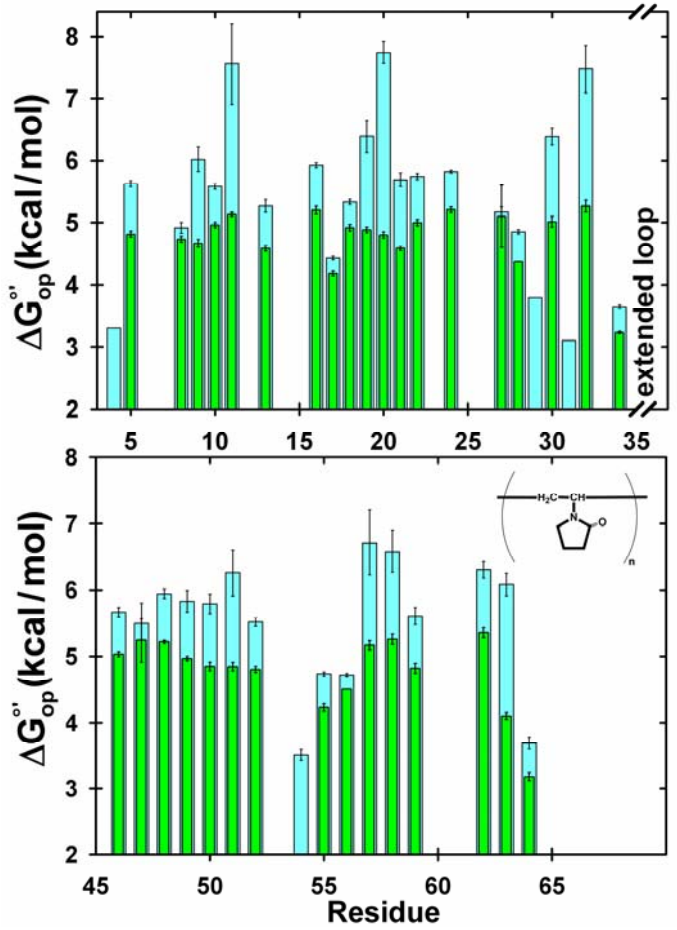

Figure 2.8. Histogram of for chymotrypsin inhibitor 2 in 300 g/L PVP and dilute solution.

Macromolecular crowding with PVP40 stabilizes the I29A:I37H variant of CI2 relative to dilute solution. Values of $\Delta G^{\circ'}_{op}$, in 300 g/L PVP40 (cyan) and dilute solution (green) are plotted *versus* residue number. The height of each bar represents the average from three trials. The error bars represent the standard deviation. Conditions: 700 - 800 μM variant protein, 50 mM acetate buffer made with D_2O, pH 5.4. The inset shows the backbone structure of PVP.

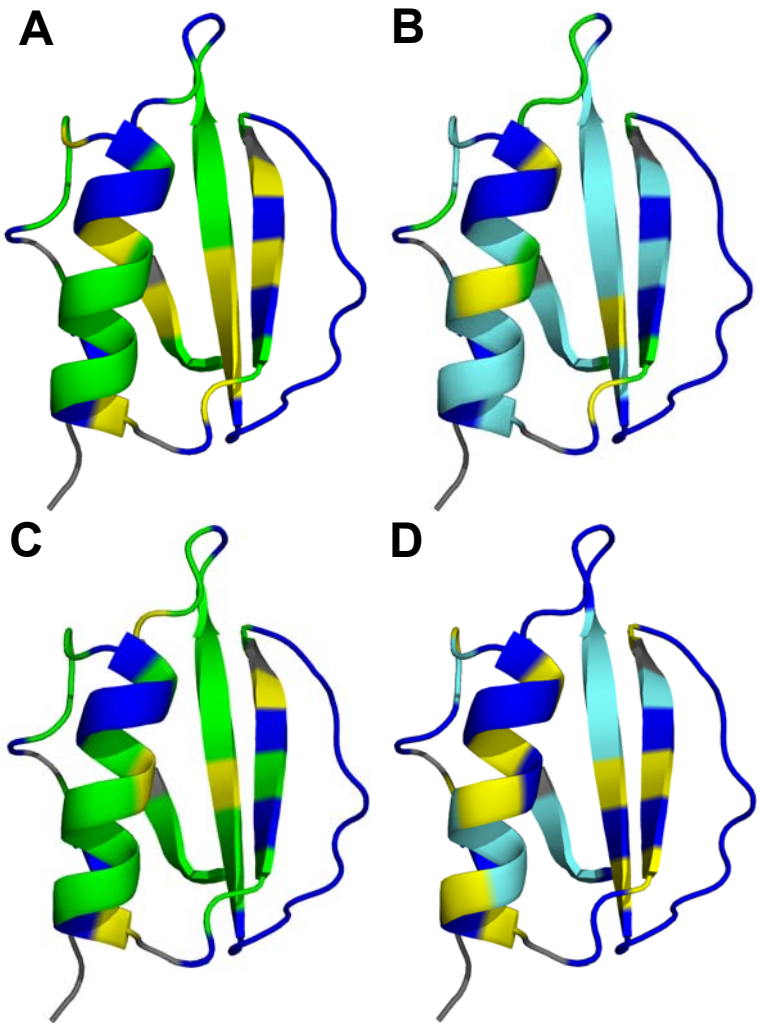

Figure 2.9. Ribbon structure of wild-type chymotrypsin inhibitor 2 (PDB, 2CI2) colored by $\Delta G^{\circ'}_{op}$.

(A) dilute solution (B) 300-g/L PVP40, and (C) 100-g/L NEP [blue: $\Delta G^{\circ'}_{op} < 3.0$ kcal/mol; green: 3.0 kcal/mol $\leq \Delta G^{\circ'}_{op} \leq$ 5.0 kcal/mol; yellow: 5.0 kcal/mol $< \Delta G^{\circ'}_{op} \leq$ 5.5 kcal/mol; cyan: 5.5 kcal/mol $< \Delta G^{\circ'}_{op} <$ 8.0 kcal/mol]. (D) $\Delta G^{\circ'}_{op}$ was measured in 0, 50, 100, and 300-g/L PVP40, and $\delta \Delta G^{\circ'}_{op} / \delta$[PVP40] values were superimposed onto the structure [blue: $\delta \Delta G^{\circ'}_{op} / \delta$[PVP40] = 0, yellow: $0 < \delta \Delta G^{\circ'}_{op} / \delta$[PVP40] $\leq 2.5 \times 10^{-3}$ (kcal/mol)/M, cyan: 2.5×10^{-3} (kcal/mol)/M $< \delta \Delta G^{\circ'}_{op} / \delta$[PVP40] $< 1.0 \times 10^{-2}$ (kcal/mol)/M]. Residues for which there are no data are colored grey.

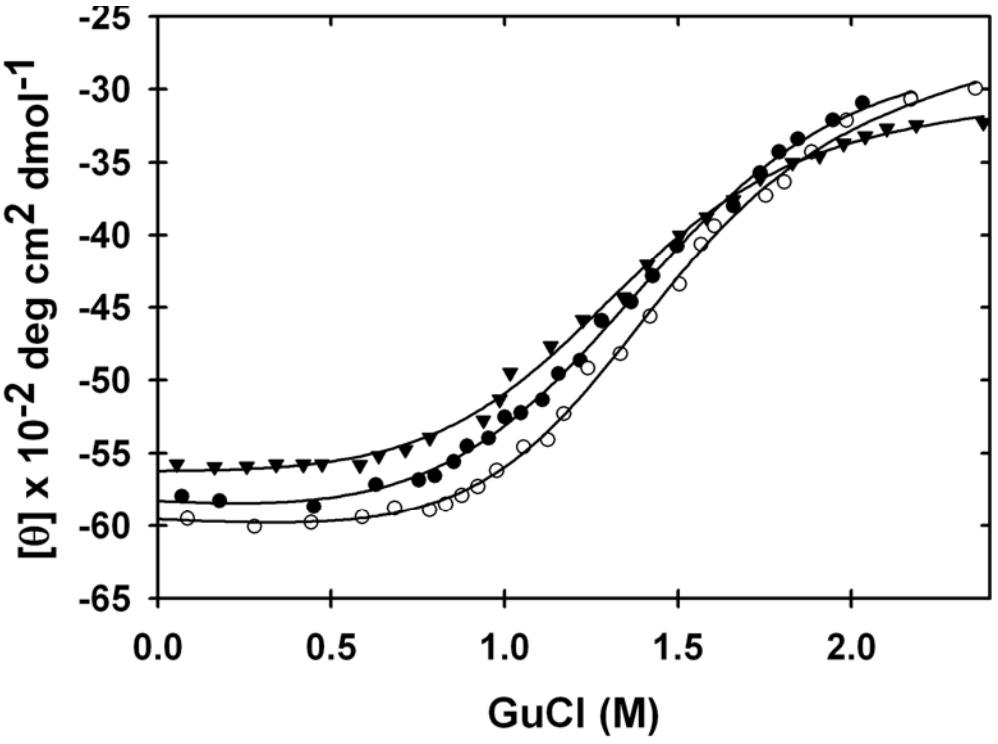

Figure 2.10. GdmCl denaturation of the I29A:I37H variant at 37 oC in 300 g/L PVP for three trials measured by circular dichroism at 242 nm.
The sample was 800 μM in 50 mM acetate buffer made with D_2O, 300 g/L PVP. The average m value is 2.3 ± 0.2 , the average $[GuCl]^{50\%}$ is 1.31 ± 0.04, and the average ΔG_{den} is 3.0 ± 0.3 kcal/mol.

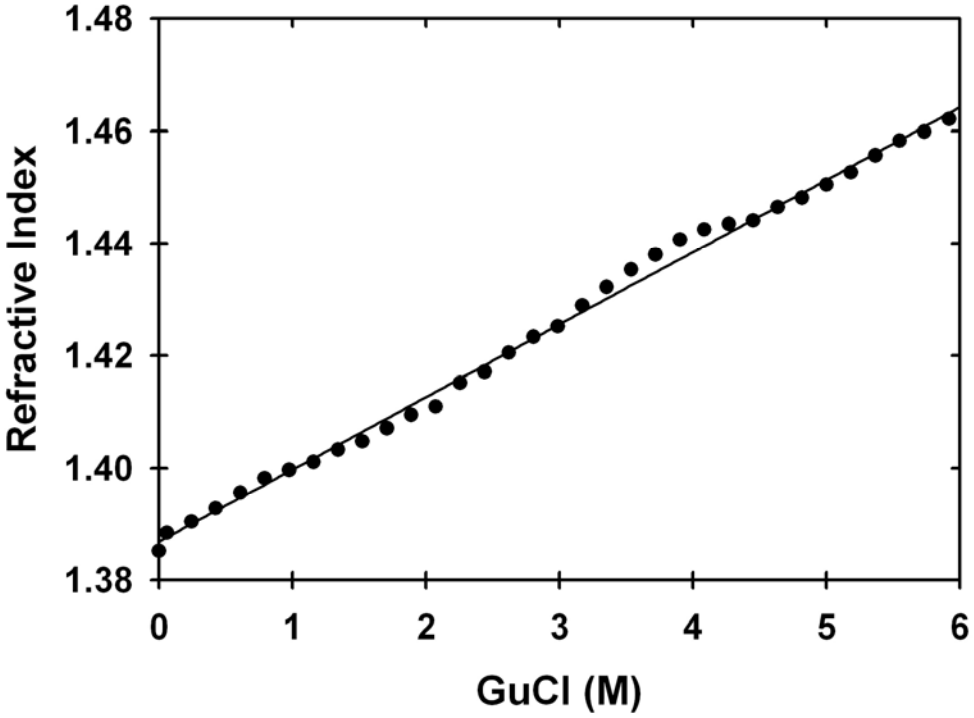

Figure 2.11. Standard curve relating refractive index to [GdmCl] in 300 g/L PVP.

Table 2.1. ^{1}H, ^{15}N assignments and k_{ex} for the I29A:I37H variant of chymotrypsin inhibitor 2

Table 1. ^{1}H,^{15}N assignments and k_{ex} for the I29A:I37H variant of chymotrypsin inhibitor 2

	Dilute Solution				300 g/L PVP			
Residue	^{1}H (ppm)	^{15}N (ppm)	k_{ex} (min^{-1})	St. Deviation [a] (min^{-1})	^{1}H (ppm)	^{15}N (ppm)	k_{ex} (min^{-1})	St. Deviation (min^{-1})
Met1	nv [b]	nv	na [c]	na	nv	nv	na	na
Lys2	nv	nv	na	na	nv	nv	na	na
Thr3	8.44	113.55	tf [d]	na	8.42	113.54	tf	na
Glu4	6.73	118.03	tf	na	6.71	117.99	tf	na
Trp5	8.34	120.17	3.66E-03	2.81E-04	8.32	120.07	9.91E-04	1.39E-04
Pro6	na	na	na	na	na	na	na	na
Glu7	10.75	121.34	tf	na	10.76	121.14	tf	na
Leu8	7.91	120.27	2.85E-03	2.19E-04	7.88	120.28	2.15E-03	4.86E-04
Val9	7.17	120.57	1.99E-03	2.17E-04	7.15	120.56	2.47E-04	1.47E-04
Gly10	9.11	117.65	1.32E-02	9.05E-04	9.07	117.42	4.87E-03	6.05E-04
Lys11	7.81	119.97	1.01E-02	5.55E-04	7.77	119.73	5.80E-04	8.82E-04
Ser12	8.60	117.05	tf	na	8.57	117.02	tf	na
Val13	8.58	123.79	7.24E-03	5.29E-04	8.53	123.74	2.47E-03	7.12E-04
Glu14	8.21	120.66	tf	na	8.23	120.64	tf	na
Glu15	7.56	121.60	tf	na	7.54	121.50	tf	na
Ala16	8.47	121.61	5.03E-03	5.16E-04	8.43	121.52	1.55E-03	1.65E-04
Lys17	8.51	117.44	3.19E-02	2.07E-03	8.49	117.40	2.13E-02	1.75E-03
Lys18	7.05	117.62	1.27E-02	1.03E-03	7.02	117.56	6.51E-03	7.23E-04
Val19	7.46	121.12	2.94E-03	2.16E-04	7.43	121.01	3.05E-04	2.00E-04
Ile20	8.00	120.78	1.75E-03	1.49E-04	7.98	120.68	1.63E-05	9.07E-06
Leu21	7.76	118.16	2.85E-03	1.60E-04	7.72	118.14	4.90E-04	1.58E-04
Gln22	7.35	119.17	6.56E-03	5.36E-04	7.32	118.96	1.97E-03	2.75E-04
Asp23	7.80	120.25	tf	na	7.77	120.18	tf	na
Lys24	9.00	124.73	4.35E-03	3.19E-04	8.97	124.70	1.60E-03	1.10E-04
Pro25	na	na	na	na	na	na	na	na
Glu26	9.51	119.36	tf	na	9.56	119.35	tf	na
Ala27	8.06	124.60	7.20E-03	5.00E-03	8.04	124.60	5.35E-03	1.25E-03
Gln28	8.68	122.70	2.96E-02	3.61E-05	8.65	122.60	1.34E-02	1.37E-03
Ala29	8.42	130.08	tf	na	8.42	130.03	tf	na
Ile30	7.98	125.83	1.72E-03	2.40E-04	7.92	125.71	1.94E-04	6.51E-05
Val31	8.41	129.90	tf	na	8.37	130.03	tf	na
Leu32	8.98	131.13	1.17E-03	1.89E-04	8.92	130.65	4.91E-05	5.50E-05
Pro33	na	na	na	na	na	na	na	na

Val34	8.87	126.79	1.92E-02	5.85E-04	8.85	126.80	9.74E-03	8.85E-04
Gly35	8.95	117.02	tf	na	8.91	116.77	tf	na
Thr36	7.59	118.84	tf	na	7.57	118.73	tf	na
His37	8.91	127.67	tf	na	8.91	127.64	tf	na
Val38	8.11	120.06	tf	na	8.08	119.68	tf	na
Thr39	7.90	113.88	tf	na	7.84	113.56	tf	na
Met40	8.48	120.02	tf	na	8.51	119.75	tf	na
Glu41	8.50	125.43	tf	na	8.50	124.87	tf	na
Tyr42	8.57	126.92	tf	na	8.59	126.81	tf	na
Arg43	8.80	132.46	tf	na	8.81	132.25	tf	na
Ile44	7.98	121.89	tf	na	8.06	121.62	tf	na
Asp45	8.32	117.74	tf	na	8.38	117.69	tf	na
Arg46	7.20	123.90	7.57E-03	4.64E-04	7.17	123.90	2.72E-03	3.22E-04
Val47	8.03	124.11	2.26E-03	1.03E-03	8.00	124.03	1.49E-03	6.86E-04
Arg48	9.07	128.11	5.67E-03	2.37E-04	9.05	128.03	1.75E-03	2.09E-04
Leu49	9.01	127.01	4.26E-03	2.32E-04	8.99	127.05	1.07E-03	2.72E-04
Phe50	9.16	124.54	4.22E-03	4.35E-04	9.14	124.52	9.28E-04	2.16E-04
Val51	8.74	116.53	2.75E-03	2.85E-04	8.71	116.49	3.07E-04	1.60E-04
Asp52	8.82	125.22	5.15E-03	3.88E-04	8.78	125.13	1.63E-03	1.69E-04
Lys53	8.18	117.39	tf	na	8.19	117.23	tf	na
Leu54	8.12	122.81	tf	na	8.10	122.39	3.67E-02	4.95E-03
Asp55	8.17	116.07	1.13E-02	9.91E-04	8.18	116.07	4.90E-03	2.35E-04
Asn56	8.22	116.31	tf	na	8.22	116.33	3.22E-02	1.20E-03
Ile57	9.10	121.32	2.77E-03	3.03E-04	9.07	121.29	2.68E-04	1.68E-04
Ala58	9.59	133.70	3.63E-03	4.49E-04	9.55	133.64	4.54E-04	2.07E-04
Glu59	7.25	116.36	4.19E-03	4.95E-04	7.22	116.18	1.18E-03	2.66E-04
Val60	8.96	127.05	tf	na	8.93	127.10	tf	na
Pro61	na	na	tf	na	na	na	na	na
Arg62	8.48	123.78	3.63E-03	4.31E-04	8.43	123.67	7.84E-04	1.63E-04
Val63	9.40	125.09	2.33E-03	1.92E-04	9.38	125.11	5.46E-04	1.56E-04
Gly64	8.73	123.91	4.54E-03	4.22E-04	8.65	123.65	1.99E-03	3.10E-04

[a] Standard deviation from three trials

[b] nv, not visible

[c] na, not applicable

[d] tf, too fast to measure

3 Caveats of in-cell NMR in *Escherichia coli*

3.1 Introduction to in-cell NMR of globular, partially folded, and natively disordered proteins

In-cell NMR is quite advantageous for studying proteins in a near physiologically relevant environment, such as inside living *E. coli*. The technique gives atomic-level information about a protein in a noninvasive manner. The primary disadvantage, however, is that the technique depends tremendously on the characteristics of the protein. In general, NMR is limited by the molecular weight of a protein. A large protein will slowly tumble causing broadening of linewidths, which can lead to complete disappearance of usable data. This tumbling rate depends upon the viscosity of the solution, and the viscosity of the *E. coli* cytoplasm is 9 times that of water (*75*). The viscosity of other cells types is 2 times greater than that of water (*76*). This increase in viscosity leads to an apparent increase in the molecular weight of a macromolecule inside the cytoplasm of all cell types, but most drastically for *E. coli*.

For a globular protein, signals from the amide backbone depend on the tumbling rate; therefore, the protein must be small, not bound to other macromolecules, and soluble. The relaxation of a natively-disordered protein, however, depends, not only on the overall tumbling, but also on rotation of the

backbone since it does not have a rigid structure. The increased rotational freedom reduces T2 relaxation before severe line-broadening occurs. This allows one to observe a greater molecular weight range for natively-disordered proteins inside the cell compared to globular or partially-folded proteins. For instance, the Pielak lab has observed the natively disordered proteins FlgM (14 kDa) (*19*), α-SN (14 kDa) (*20, 77*), and human tau protein (45 kDa) [Brian McNulty, unpublished], and the Dötsch lab has observed the globular protein NMerA (7 kDa) (*18*).

The most common application of in-cell NMR is the acquisition of a ^{15}N HSQC spectrum on the protein of interest. Incorporation of ^{15}N into cellular components creates background signal. Serber *et al.* have determined that the amount of background signal arising from incorporation of ^{15}N into other cell constituents is independent of both media and induction protocol (*16*). For instance, rifampicin was used to suppress synthesis of bacterial proteins; however, the background level was not altered by this addition (*16*). Since incorporation of ^{15}N in background molecules occurs with consistency, the expression level of the protein of interest is a relevant parameter when attempting to observe protein spectra inside cells. It is estimated that a protein must be overexpressed to greater than 5% of the total soluble protein inside living *E. coli* to be observed (*16*). Other protein specific factors, such as protein-protein interactions and stability, will affect the spectrum, but none will have as great of an influence on the quality of the spectrum as the expression level and the correlation time.

3.2 Materials and Methods

3.2.1 In-cell NMR preparation of chymotrypsin inhibitor 2 (CI2) and apocytochrome b_5

The following protocol is for in-cell NMR of CI2 and apocytochrome b_5. The plasmid was transformed into BL21 (DE3) Gold cells (Stratagene), plated on LB agar plates, and grown overnight at 37 ^{o}C. A single colony was used to inoculate 20 mL ^{15}N-enriched Martek media (Spectra Stable Isotopes). The culture was grown overnight at 37 ^{o}C with shaking (225 rpm). Fifteen mL of this starter culture was added to 250 mL of media in a 1 L Erlenmeyer flask. The culture was grown to an optical density of 1.0 (Abs_{600}) at 37 ^{o}C with shaking (225 rpm). The cells were induced with isopropyl thiogalactoside (IPTG, final concentration, 1 mM), and expression was allowed to occur at 37 ^{o}C for 5 h. Chloroamphenicol was then added to a final concentration of 50 μg/mL, and the culture was shaken for another 15 min. The cells were harvested at 1600 g for 30 min. The supernatant was decanted, and 2 mL were reserved for resuspension and spectrometer shimming. The wet pellet typically weighs 1.2 -1.5 g. The variation arises from the liquid remaining in the pellet. The pellet was resuspended in 200 μL of spent media (or buffer of interest). The final volume of the sample was approximately 1.2 mL. A 90:10 mixture of cell slurry:D_2O was added to the NMR tube.

The cell slurry was subjected to dilution plating before and after the NMR experiment to determine the fraction of cells that survived. The cell pellet and

supernatant were subjected to analysis by SDS-PAGE to determine the relative amount of protein found in the supernatant.

3.2.2 In-cell NMR preparation of α-synuclein and FlgM

The preparation for in-cell NMR of α-SN and FlgM was adapted from McNulty *et al.* (*20*) and Dedmon *et al.* (*19*), respectively. The pT7-7 vector containing the α-SN gene was transformed into *E. coli* BL21 Gold cells and selected with ampicillin. The pMCA4 vector containing the FlgM gene was also transformed into BL21 cells and selected with ampicillin (100 µg/mL). A 15 mL overnight culture was grown from a single colony and used to inoculate 250 mL of LB_{amp}. The culture was grown to an $Abs_{600}$of 1.0 at 37 ^{o}C with shaking at 225 rpm. The cells were gently centrifuged at 1600 g for 30 min at 4 ^{o}C. The cell pellet was resuspended in 250 mL of M9 minimal media (Table 2.1), induced with IPTG to a final concentration of 1 mM, and grown in a rotary shaker (225 rpm) at 37 ^{o}C for 4 h. Following expression, the cells were harvested by centrifugation at 1600 g for 30 min at 4 ^{o}C. The cell pellet was resuspended in spent media to a final volume of 1.5 mL. A 90:10 mixture of sample:D_20 was placed in a 5 mm NMR tube for data collection.

3.2.3 NMR analysis of extracellular protein

The following method was used to determine the amount of protein NMR signal arising from the extracellular supernatant during an in-cell NMR experiment for CI2, apocytochrome b_5, α-SN, and FlgM. In-cell NMR samples were prepared in

duplicate (referred to as Sample 1 and Sample 2). Sample 1 was immediately centrifuged. The supernatant from sample 1 was reserved for an NMR experiment and for analysis by SDS-PAGE. A ^{15}N HSQC was collected on a cell slurry from sample 2 before the supernatant was harvested by centrifugation at 775 x g for 30 min. The supernatant was then removed and ^{15}N HSQC spectra were collected on the supernatants from sample 1 and sample 2. The peak intensities were used to quantify the extracellular signal using NMRview (*51*).

3.2.4 Post-spectrometer SDS-PAGE analysis

The volume of the post-spectrometer sample was determined to two significant figures. The volume was approximately 1.5 mL. The sample was then centrifuged at 775 x g for 45 min, and the supernatant was removed. Both the supernatant and the cells were again centrifuged at 16,000 x g for 30 min. The second centrifugation frees the supernatant of cellular debris and removes most of the external buffer from the pelleted cells. All supernatants were combined and referred to as the extracellular supernatant since this contains any protein that has leaked from the cells during the experiment.

The volume of extracellular supernatant was subtracted from the total volume of sample to determine the volume of the cells. The pellet was resuspended in a volume of lysis buffer (50-mM acetate, pH 5.5)[1] that was equivalent to the pellet volume. Lysozyme was not used because it overlaps on SDS-PAGE with the

[1] The lysis buffer will differ per protein. Use the lysis buffer that is in your purification.

proteins investigated. The resuspended cells were subjected to ten freeze/thaw cycles by using a dry ice/ethanol bath and room temperature water bath. The slurry was then centrifuged at 16,000 g for 30 min. The supernatant was removed, and the pellet was centrifuged again for 15 min. The two supernatants were combined, and referred to as the cellular supernatant. Finally, the ratio of the combined supernatant volumes to the original volume of the pellet was determined.

The extracellular supernatant volume was adjusted according to the increase in volume in the previous step. For instance, if the original pellet volume is 600 μL and the supernatant volume from the pellet is 700 μL, the volume of the extracellular supernatant is increased by 15% with the lysis buffer.

Ten μL of extracellular supernatant was diluted to 40 μL with lysis buffer, and 8 μL of SDS loading dye. This regimen was repeated with the cellular supernatant. The samples were then placed at 90 ^{o}C for 5 min. Three μL of each sample were loaded into an 18% SDS gel (75 mm x 8 cm). Gels were electrophoresed at 200 V for 50 min. The gel was stained with Coomassie Brilliant Blue and destained. The destained gel was scanned and analyzed using the freeware program ImageJ (NIH, http://rsb.info.nih.gov/ij/). The pixel density from the extracellular and cellular supernatants was compared directly to quantify the relative amount of protein that is in the extracellular supernatant.

3.2.5 Encapsulating *E. coli* cells

A solution of 2% w/v alginate (Sigma) was prepared by dissolving solid alginate in LB. The solution was heated until all the alginate dissolved and then

filtered with a 0.22 filter. *E. coli* cells expressing ^{15}N-enriched protein were grown and prepared as previously described for in-cell experiments. A 50% v/v solution was made from the 2% alginate stock solution and the cell slurry.

The electrostatic encapsulation device comprised a 1-mL syringe, a 24-gauge angiocatheter (24 gauge, 0.7 x 19 mm tip, winged, Braun), a 23-gauge needle, a syringe pump, and an adjustable power supply ranging between 3.2 and 5.5 kV. The cell slurry/alginate mixture was loaded into the 1 mL syringe, the angiocatheter was attached, and the syringe was connected to the syringe pump in a vertical position (Fig. 3.1). The needle was inserted through the angiocatheter in a horizontal orientation. A 250-mL beaker with a stir bar and 150 mL of 150-mM $CaCl_2$ was placed below the syringe as a receiving bath. The positive voltage clamp was connected to the needle, and a copper-tip grounding wire was submerged in the receiving bath. The power supply was set to 3.7 kV, and the syringe pump was set to a rate of 0.71 mL/min. The cell slurry/alginate mixture was forced through the tip of the angiocatheter and streamed into the receiving bath below. The Ca^{2+}-ion induced polymerization of the alginate formin the encapsulated bead of cells. The encapsulated cells were retrieved with suction and placed in solution of 90:10 LB/D_2O for NMR studies.

3.2.6 NMR experiments of encapsulated cells

Data were collected on encapsulated cells expressing α-SN and CI2 for 16 h. The ^{15}N HSQC spectra consisted of 128 transients and 100 increments. Following the data collection, the LB/D_2O solution was removed from the encapsulated *E. coli*

cells and replaced with 1 mL of 10 mM ethylenediaminetetraacetic acid (EDTA). The EDTA dissolves the polymer by binding Ca^{2+}. An ^{15}N HSQC spectrum was collected on the unencapsulated cells.

3.3 Results

3.3.1 In-cell NMR of chymotrypsin inhibitor 2

It was initially believed that the CI2 was the perfect molecule for in-cell NMR. The protein is small, soluble, and overexpressed. Initial experiments resulted in well-resolved, low noise HSQC spectra of CI2 as seen in Fig. 3.2. The sample was analyzed for extracellular protein by SDS-PAGE, and this amount was determined to be minimal, ~100 μM (Fig. 3.3). However, the initial conclusions were incorrect. Although the amount of extracellular protein was small relative to the amounts of cytoplasmic protein (Fig 3.3), the NMR signal detected was from the extracellular protein. Figure 3.2, panel A shows a spectrum of the supernatant that has been immediately harvested from cells expressing CI2; panel B is a spectrum of CI2 "inside cells"; and panel C is a spectrum of supernatant that has been harvested after an in-cell experiment. It was initially thought that signal from inside the cells was seen "on top" of the signal from the extracellular protein since the concentration of cytoplasmic protein was at least 10-fold higher.

When consecutive HSQC spectra were collected on *E. coli* cells expressing CI2, an increase in both peak intensity and volume occured with time. An increase in peak volume indicates an increase in protein concentration; whereas, an increase

in peak intensity can be either the result of a decrease in linewidth, caused by, for instance, a decrease in viscosity, or an increase in protein concentration. A decrease in viscosity would be expected if the protein is being released from the cytoplasm into the solution. The widths of CI2 crosspeaks were measured over a period of 24 h, and no change in peak linewidth was detected. This observation suggested that the increasing peak intensity was due solely to a change in concentration. This change was attributed to an increase in cell density in the detector region of the spectrometer. This was referred to as “cell settling”.

Cell settling was an inaccurate interpretation of the data. Figure 3.2 shows that CI2 was found in the supernatant at the beginning and the end of the experiment. Since there was also no observable change in linewidths, we conclude that the CI2 signal originated from protein in the supernatant and not the cell. The increase in peak intensity was not the result cell settling but protein leaking from the cells into the supernatant. This CI2 result also implies that the CI2 in the cells is not observable, although, the concentration inside cells far exceeds the concentration of protein in the supernatant (Figure 3.3). The Almeida lab also concluded an increase in peak intensity was the result of increasing the concentration of extracellular protein (*78*).

3.3.2 In-cell NMR and α-synuclein

Proteins previously used for in-cell NMR were reinvestigated to ensure that protein signal was coming from inside the cells, not from the extracellular supernatant. The signal from α-SN for an in-cell experiment is from inside the cells

as shown in Figure 3.4. Figure 3.4 panels A and C show that there is no α-SN signal from the supernatant at either the beginning or the end of the experiment. The crosspeaks present in panels A and C belong to small molecules that incorporated ^{15}N during expression. The spectrum in panel B is identical to the original work of McNulty *et al.* (*20*). Also, SDS-PAGE analysis shows no protein detected in the extracellular supernatant (Figure 3.5). These data imply that α-SN does not leak from the cells during the length of an in-cell NMR experiment.

3.3.3 In-cell NMR and FlgM

FlgM was also investigated, and it was determined that all signal was from inside living *E. coli*. Figure 3.6 panels A and C shows that no FlgM is detectable in the extracellular supernatant before or after the in-cell NMR experiment. Panel B shows the spectrum of FlgM from the in-cell experiment inside living cells. The spectrum in panel B is identical to the original work of Dedmon *et al.* (*19*). The crosspeaks that are present in panel A and C belong to small molecules that incorporated ^{15}N during protein expression.

3.3.4 In-cell NMR and apocytochrome b_5

A ^{15}N HSQC spectrum of the supernatant immediately harvested after sample preparation is shown in Fig 3.7, panel A; the spectrum from “inside cells” is shown in panel B; and the spectrum from the supernatant collected after the in-cell experiment is shown in panel C. Apocytochrome b_5 behaves similarly to CI2 for in-cell

experiments. According to SDS-PAGE, the extracellular protein is minimal (Fig. 3.8); however, the signal is only observed from the extracellular protein (Fig 3.7). I concluded the signal from apocytochrome b_5 during in-cell experiments arises from the extracellular protein.

3.3.5 In-cell NMR of encapsulated cells

Panels A and B in Figure 3.9 show spectra from encapsulated cells expressing α-SN and CI2, respectively. α-SN can be detected inside encapsulated cells proving we can still observe protein from inside cells even when encapsulated (Fig. 3.9A). On the other hand, CI2 was not observed inside alginate-encapsulated cells (Fig. 3.9B). CI2, however, was observed after dissolution of the encapsulates (Fig. 3.10). This proves encapsulation prevents protein leakage into the surrounding media. The observation also shows that CI2 cannot be observed in the *E. coli* cytoplasm, suggesting the increased viscosity of the cytoplasm compared to dilute solution affects globular proteins differently than natively-disordered proteins.

3.4 Conclusions

The interpretation of in-cell NMR data must be pursued with caution. As determined above, SDS-PAGE analysis of the extracellular supernatant is not sufficient to support the interpretation that the spectra of CI2 and apocytochrome b_5 from “in-cell” experiments comes from protein in the cell. The protein signals never originated from inside cells, although the concentration of protein found in the supernatant was minimal. The two papers from our group on in-cell NMR of

apocytochrome b_5 have been retracted (*79*). The definitive control for in-cell NMR is to measure the signal in the supernatant at the beginning and end of the experiment.

Both apocytochrome b_5 and CI2 are ideal candidates for in-cell NMR since they are small, soluble, and overexpressed; yet, it appears that no signal arises from inside cells. Recent results measuring protein dynamics show that globular proteins are more affected by the intracellular viscosity than natively-disordered proteins (Li, *et al. in press*, JACS). The increased cytoplasmic viscosity increases the transverse relaxation rates of CI2, leading to line-broadening and no observable signal. The encapsulation study supports this reasoning because it proves that CI2 cannot be observed inside *E. coli*. The data also suggest that natively-disordered proteins are better suited for in-cell NMR on cells with high cytoplasmic viscosity such as *E. coli* and yeast.

3.5 Figures

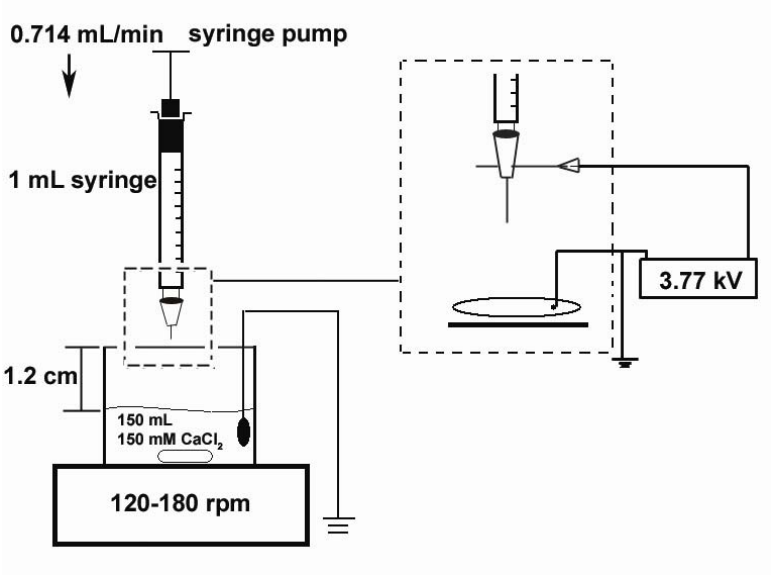

Figure 3.1. Schematic of encapsulation device.

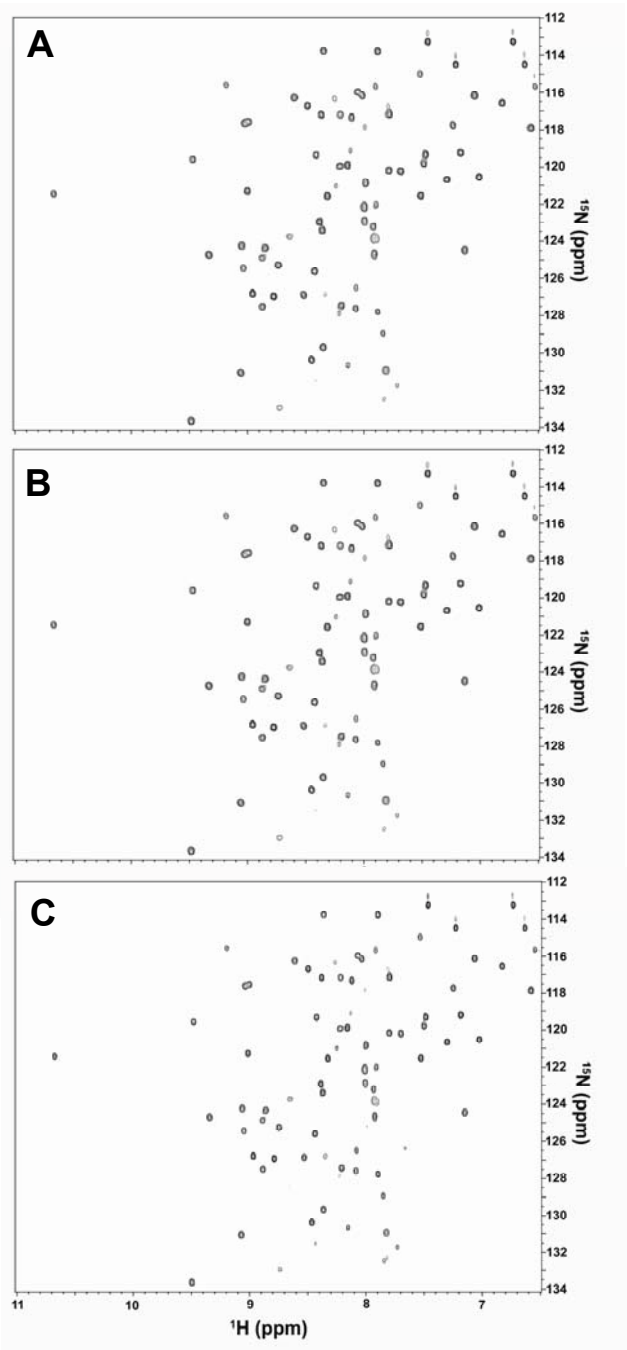

Figure 3.2. Spectra of chymotrypsin inhibitor 2 from in-cell NMR experiment. **Panel A:** Spectrum collected on the supernatant from cells expressing CI2. The supernatant was collected from immediate centrifugation of Sample 1. **Panel B:** Spectrum collected on an in-cell sample of CI2 (Sample 2). **Panel C:** Spectrum collected on the supernatant collected from the in-cell sample of CI2 (from panel B) after 2 h incubation at room temperature under anaerobic conditions.

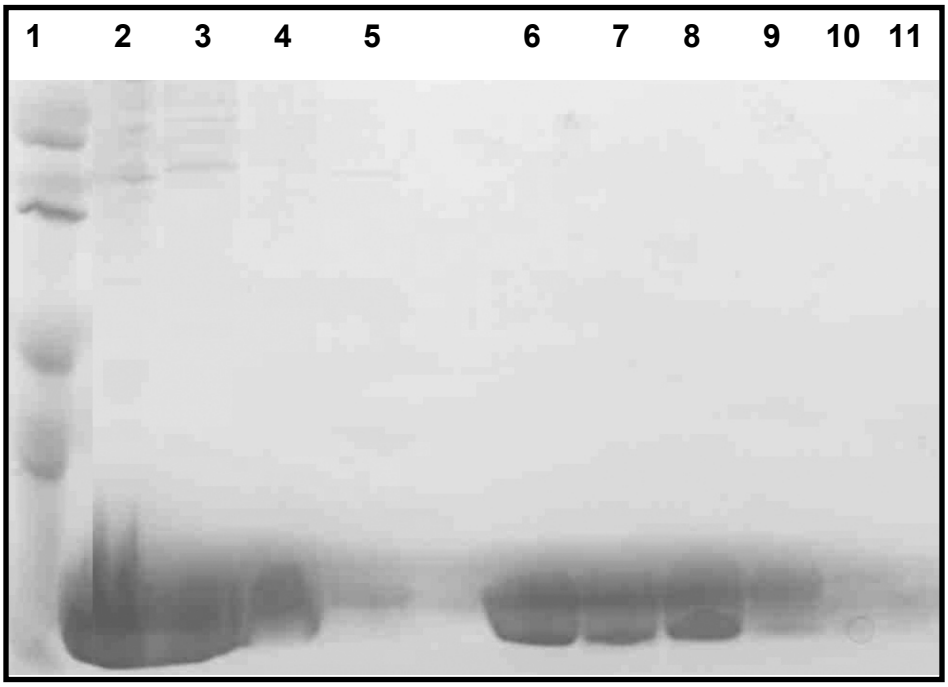

Figure 3.3. Commassie-stained, SDS-PAGE of chymotrypsin inhibitor 2 from the cytoplasm and the supernatant of an in-cell NMR experiment. **Lanes: 1,** Marker; **2,** CI2 recovered from cytoplasm; **3,** 2-fold dilution of lane 2; **4,** 4-fold dilution of lane 2; **5,** CI2 recovered from supernatant; **6,** 1 mM CI2 standard ; **7,** 750 µM CI2 standard**; 8,** 500 µM CI2 standard; **9,** 250 µM CI2 standard; **10,** 100 µM CI2 standard; **11,** 50 µM CI2 standard.

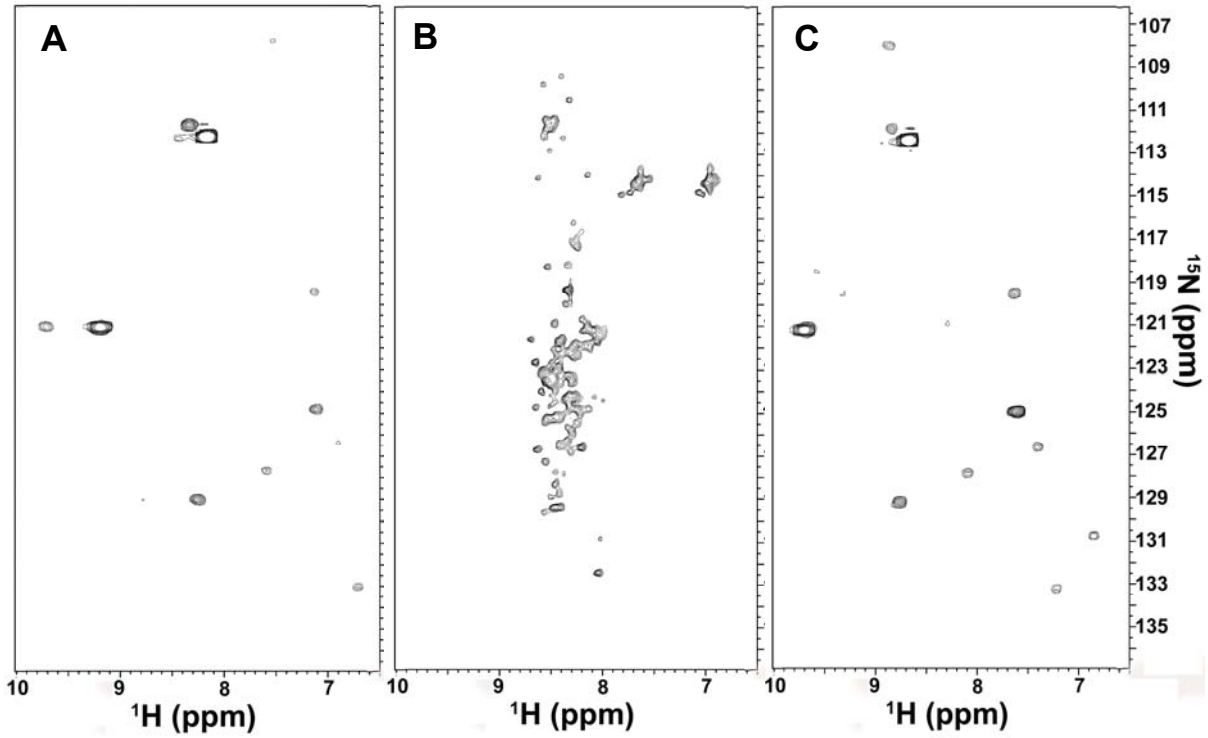

Figure 3.4. Spectra of α-synuclein from an in-cell NMR experiment. **Panel A:** Spectrum collected on the supernatant collected immediately after preparation from cells expressing α-SN (Sample 1). **Panel B:** Spectrum collected on an in-cell sample of α-SN (Sample 2). **Panel C:** Spectrum collected on the supernatant collected from the in-cell sample of α-synuclein (from Panel B) after 2 h incubation at room temperature under anaerobic conditions.

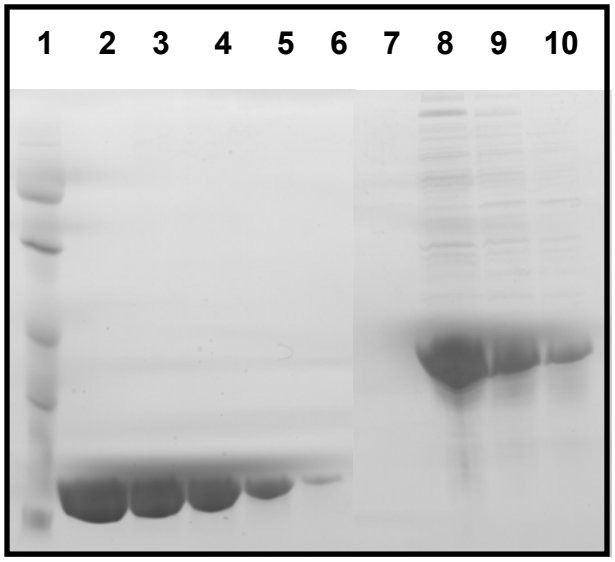

Figure 3.5. Commassie-stained, SDS-PAGE of α-synuclein recovered from the cytoplasm and supernatant of an in-cell NMR experiment.
Lanes: 1, Marker; **2,** 1 mM CI2 standard; **3,** 750 μM CI2 standard; **4,** 500 μM CI2 standard; **5,** 250 μM CI2 standard; **6,** 100 μM CI2 standard; **7,** extraceullar α-SN from in-cell NMR sample; **8,** cytoplasmic α-SN from in-cell NMR sample; **9,** 2-fold dilution of lane 8; **10,** 4-fold dilution of lane 8.

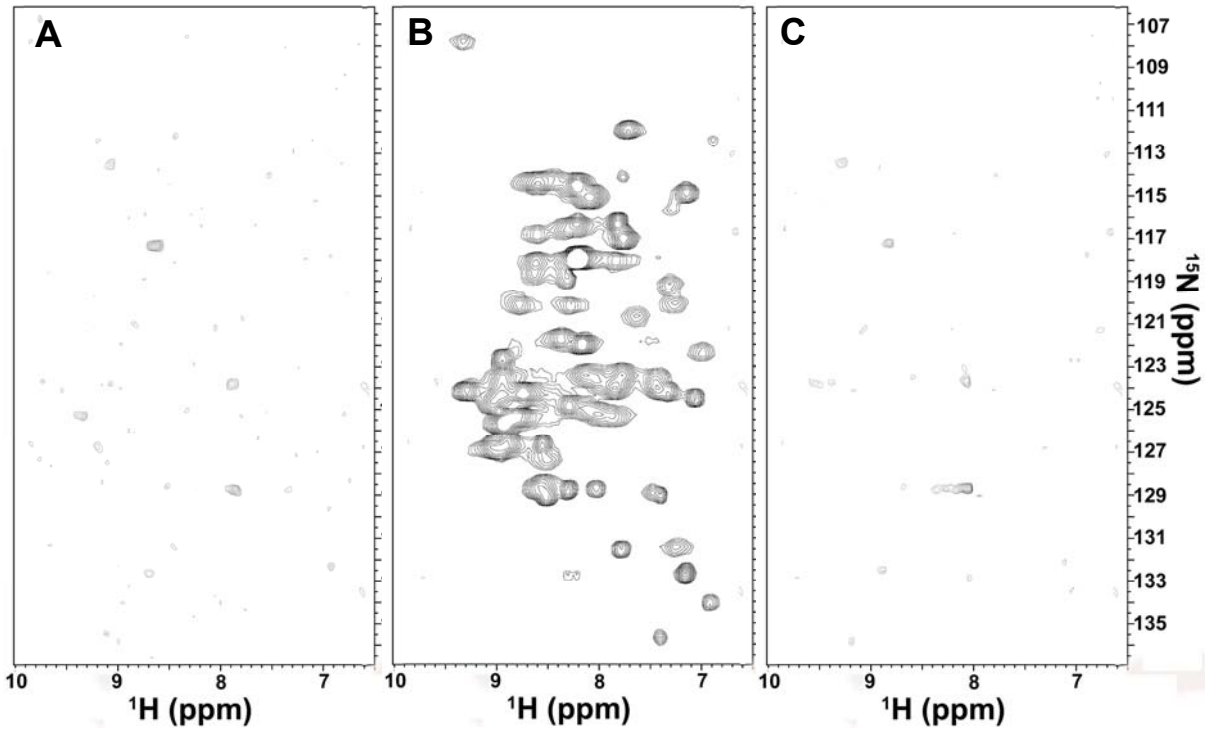

Figure 3.6. Spectra of FlgM from an in-cell NMR experiment
Panel A: Spectrum collected on the supernatant from cells expressing FlgM. The supernatant was collected from immediate centrifugation of the sample (Sample 1). **Panel B:** Spectrum collected on an in-cell sample of FlgM (Sample 2). **Panel C:** Spectrum collected on the supernatant collected from an in-cell sample of FlgM (from Panel B) after 2 h incubation at room temperature under anaerobic conditions.

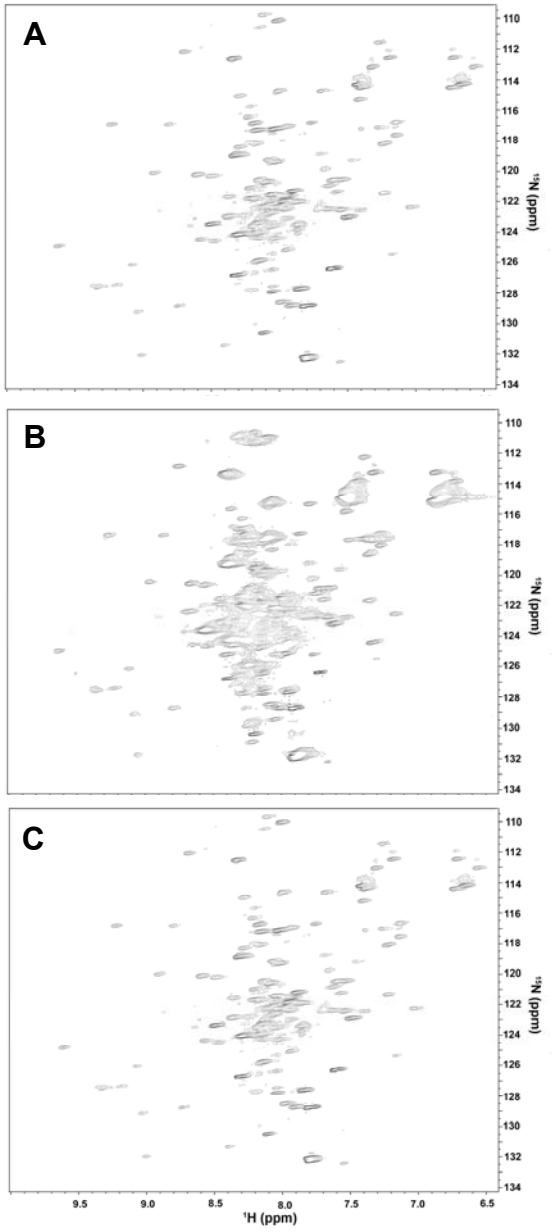

Figure 3.7. Spectra of apocytochrome b_5 from in-cell NMR experiment. **Panel A:** Spectrum collected on the supernatant from cells expressing apocytochrome b_5. The supernatant was collected from immediate centrifugation of the sample (Sample 1). **Panel B:** Spectrum collected on an in-cell sample of apo-b_5 (Sample 2). **Panel C:** Spectrum collected on the supernatant collected from an in-cell sample of apo-b_5 (from Panel B) after 2 h incubation at room temperature under anaerobic conditions.

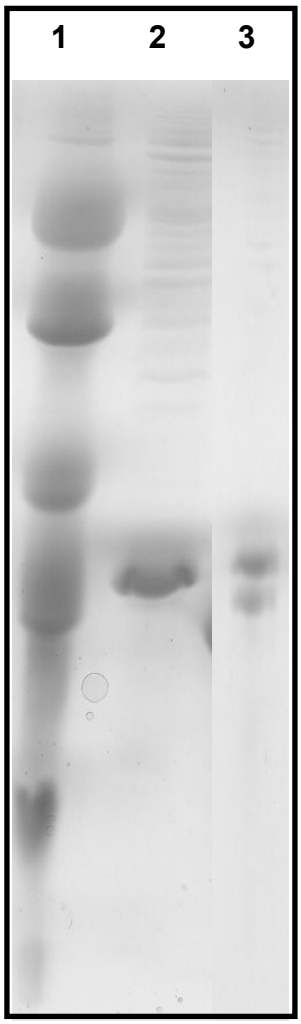

Figure 3.8. Commassie-stained, SDS-PAGE of apocytochrome b_5 recovered from the cytoplasm and supernatant of an in-cell experiment.
Lane 1: Marker, **Lane 2:** cytoplasmic apocytochrome b_5, **Lane 3:** apocytochrome b_5 recovered from the extracellular supernatant.

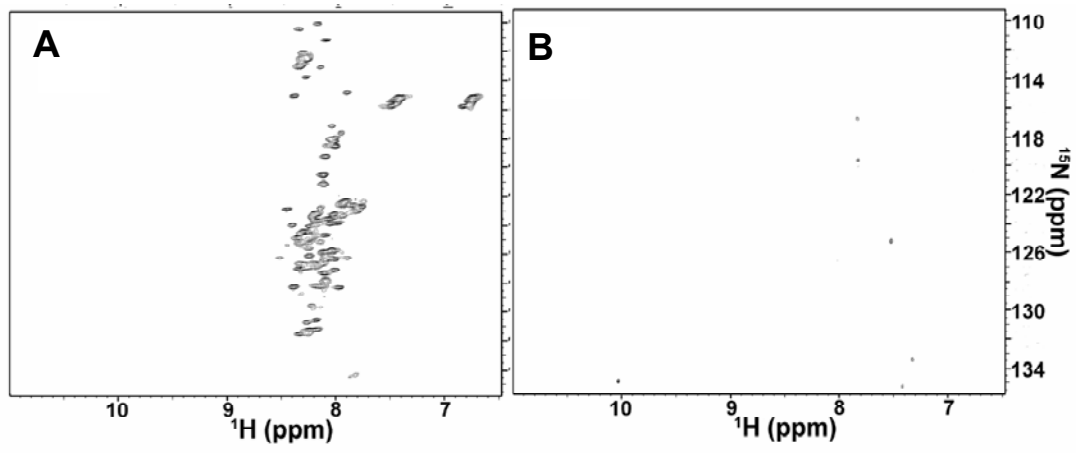

Figure 3.9. Spectra from α-synuclein and chymotrypsin inhibitor 2 inside cells, supernatant, and encapsulated cells.
Right panel: α-SN, Left panel: CI2, from alginate encapsulated cells. (Li, *et al.*, in press, JACS).

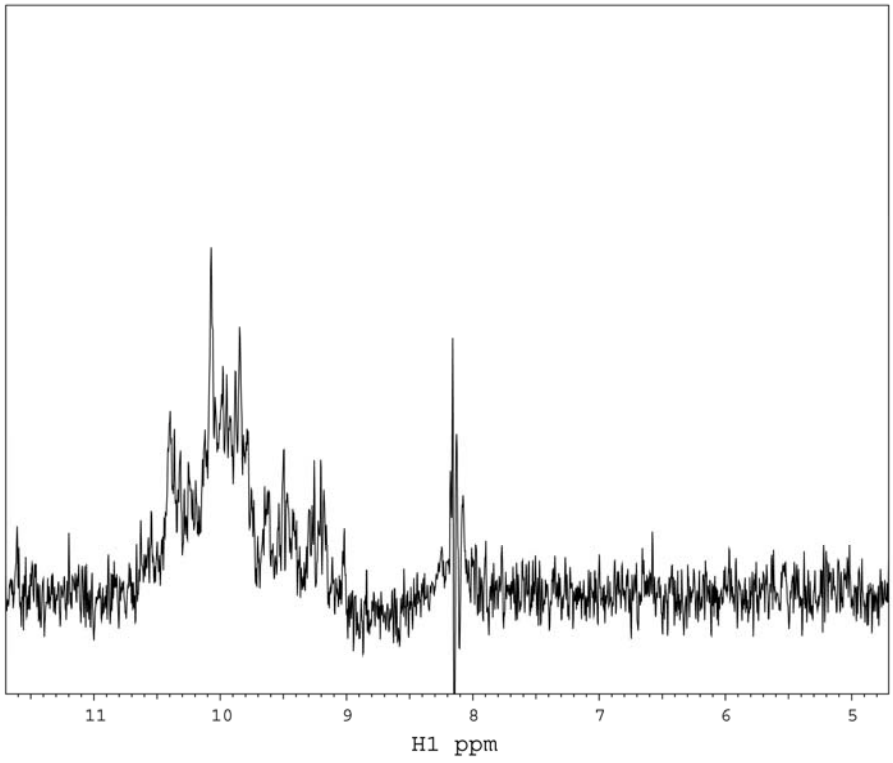

Figure 3.10. 1D ^{15}N -filtered spectrum of chymotrypsin inhibitor 2 after dissolving alginate encapsulate.

4 In-cell NMR of *Pichia pastoris*

4.1 Introduction

4.1.1 *Pichia pastoris*

The yeast, *Pichia pastoris*, serves as a eukaryotic expression system capable of producing milligram-to-gram quantities of recombinant protein. The system has the benefits of *E. coli* –high expression levels in a simple, inexpensive growth media and easy control over parameters such as pH, aeration, and carbon source feed rate. It also has the advantages of a eukaryotic synthesis pathway (*e.g.*, posttranslational modifications). These properties make it an interesting target for in-cell NMR studies.

P. pastoris is one of four strains belonging to the genera of methylotrophic yeast (*80*), so named because they utilize methanol as a carbon source in the absence of a repressing carbon source, such as glucose. The enzyme, alcohol oxidase, catalyzes the first step in methanol catabolism and is under the control of AOX1 promoter (*81-83*). This promoter is found in the expression vector to drive transcription of your favorite gene (YFG). Regulation of the AOX1 promoter involves repression/derepression and induction mechanisms (*83*). For cells grown in the presence of methanol, AOX1 mRNA constitutes ~5% of poly(A)+ RNA; whereas, in

the presence of other carbon sources, AOX1 mRNA is undetectable (*83*). The presence of methanol is necessary to induce transcription from the AOX1 promoter (*81*); the absence of a repressing carbon source alone does not result in transcription.

The *P. pastoris* expression vectors (Invitrogen) have been designed as *E. coli/ P. pastoris* shuttle vectors, containing an origin of replication for use in *E. coli*. The expression vector contains a 0.9 kb fragment from the AOX1 operon composed of a 5' promoter sequence and an AOX1-derived transcription termination sequence (Figures 4.1 and 4.2).

The *P. pastoris* expression vectors exploit the propensity for *P. pastoris* cells to undergo homologous recombination between genomic and artificially introduced DNAs. The vector contains a sequence that is shared by the host genome to promote recombination events that result in the integration of the vector into the target genome (*83*). The recombination creates a genetically stable recombinant element that is not expunged by the cell in the absence of antibiotic resistance. The recombination can occur as either a single or double crossover event (Figs. 4.3 and 4.4). A single crossover yields genomic DNA with YFG under the control of the AOX1 promoter and an intact, functional, AOX1 promoter specifically for methanol metabolism (Figure 4.3). The phenotype of these cells is methanol utilization plus (Mut^+). The Mut^+ cells grow on methanol at the wild-type rate and require high feeding rates of methanol (*80*). A double crossover replaces the AOX1 promoter involved in methanol metabolism with YFG under the control of another AOX1 promoter, resulting in cells with the phenotype methanol utilization slow (Mut^S)

(Figure 4.4). The disrupted AOX1 locus forces the cells to metabolize methanol using the weaker AOX2 gene (*84*).

Selectable markers functional in one or both organisms are present on the vector. In *P. pastoris HIS4* is one example of a selectable marker gene. After transformation of the shuttle vector into His^- auxotrophic yeast cells and upon homologous recombination with the genomic DNA, the *HIS4* gene is reformed and provides for a selectable marker when the transformants are grown on His^- media.

4.1.1 Why yeast?

Recent results from in-cell NMR experiments suggest that the most promising proteins for in cell studies belong to the group of natively-disordered proteins (Li, *et al.*, *in press*, JACS). Among this group, three proteins, α-SN, FlgM, and tau-40, have yielded in-cell information from *E. coli* (*19, 20, 77*). The next logical step was to study these proteins in a eukaryotic system. Proteins purified from *E. coli* have been microinjected into the cytoplasm of *Xenopus laevis* oocytes and monitored with NMR (*25, 85*). These proteins, however, lack the posttranslational modifications that occur as part of eukaryotic synthesis pathway. *P. pastoris* provides the simplest, most inexpensive eukaryotic system for endogenous expression of recombinant proteins and monitoring with NMR. The transition from *E. coli* to *P. pastoris* for in-cell NMR should have occurred with relative ease due to the similarities between systems.

4.2 Materials and Methods

4.2.1 Ligation

Genes for the recombinant proteins chymotrypsin inhibitor 2 (CI2), human α-SN, and FlgM were cloned into the pAO815 yeast/ *E. coli* shuttle vector (Figure 4.2). The pAO815 plasmid was initially transformed into *E. coli* DH5α cells and the transformants plated on LB_{amp} agar plates. Five-mL cultures of LB were inoculated with single colonies and grown overnight at 37 ^{o}C with shaking at 225 rpm. Plasmid DNA (pDNA) was retrieved using a Quiagen mini-prep kit. The pAO815 plasmid was linearized with EcoRI at 37 ^{o}C, and the enzyme was heat inactivated by incubation at 65 ^{o}C for 20 min. The plasmid was then dephosphorylated with calf-intestinal phosphatase (CIP) for 1 h at 37 ^{o}C. The mixture was electrophoresed on a 1% agarose gel with 0.1% ethidium bromide (EtBr) at 100 V. The digested, dephosphorylated plasmid was identified by comparison with undigested plasmid, cut out of the gel, and recovered in water using a Quiagen gel extraction kit.

The genes were PCR-amplified with flanking EcoRI restriction sites. The PCR product was electrophoresed on a 1% agarose gel (0.1 % EtBr, 100 V). The product was identified by size comparison with a 1 kb standard DNA marker (NEB), removed from the gel, and recovered in water using a gel extraction kit (Quiagen). The product was digested with EcoRI at 37 ^{o}C for 1 h and then the enzyme was heat inactivated at 65 ^{o}C for 20 min. This created sticky ends.

The ligation reaction mixture consisted of digested plasmid and PCR-amplified gene product in ratios varying from 1:1 to 1:10, T4 ligase (NEB), ligase

buffer, and water. The ratios were determined by comparing the intensities of bands on a UV-illuminated agarose gel. The ligation reaction was carried out at 16 ^{o}C overnight. Four μL of each reaction mixture were transformed into 50 μL of competent DH5α cells. The cells were allowed to recover in 350 μL of SOC media for 1 h. Two hundred μL were plated on LB_{amp} agar plates. Five-mL cultures of LB_{amp} were inoculated with individual colonies and grown overnight at 37 ^{o}C with shaking at 225 rpm. Plasmid DNA was extracted from the overnight cultures by using a miniprep kit (Quiagen), and small aliquots of pDNA were subsequently digested with EcoRI at 37^{o}C for 1 h and ran on a 1% agarose gel (0.1% EtBr, 100 V). The plasmids containing inserts were identified by size comparison with a 1 kb standard DNA marker (NEB) and the sequence of the insert was confirmed by DNA sequencing (Genome Analysis Facility, UNC).

4.2.2 Preparation of DNA for transformation into *P. pastoris*

The following protocol was used to prepare the pAO815 vector containing αSN, CI2, or FlgM for transformation. The recombinant pAO815 vector was linearized in the presence of Sal I (NEB) for 1 h at 37 ^{o}C. The digested plasmid was analyzed by agarose gel electrophoresis (1% agarose, 0.1% EtBr, 100 V) to check for complete digestion. The digested DNA was extracted by adding equal parts of a phenol: chloroform: isoamyl alcohol mixture (25:24:1), vortexing, and centrifuging at 12000 x g for 20 min. The top layer was recovered. The process was repeated, and the two extracts combined. The DNA was ethanol precipitated by adding 2 volumes of 100% ethanol and mixing. The mixture was centrifuged at 12000 x g for 10 min at

4 °C. The supernatant was discarded, and the pellet was washed with 1 volume of 70% ethanol and centrifuged at 12000 x g for 2 min. The supernatant was discarded. The DNA pellet was resuspended in 10 μL of TE buffer, and the concentration determined by measuring absorbance at 260 nm.

4.2.3 Spheroplasting

To enable yeast to incorporate DNA, it is necessary to partially remove the cell wall. The enzyme zymolyase partially digests the cell wall by hydrolyzing its β-1, 3 glucose polymers.

Cells were prepared for spheroplasting by streaking the GS115 strain of *P. pastoris* onto a yeast extract, peptone, dextrose (YPD) agar plate to isolate single colonies. The plate was incubated at 30 °C for 2 days. Ten mL of YPD media in a 50-mL culture tube was inoculated with a single colony and grown overnight at 30 °C with shaking at 250 rpm. Two-hundred mL of YPD media was placed in three 500-mL baffled culture flasks, inoculated with 5, 10, and 20 μL of the overnight culture, and incubated at 30 °C overnight with shaking at 250 rpm. After checking the optical density at 600 nm (1-cm cuvette), cells were harvested by centrifugation (1500 x g, 10 min, room temperature;R-T) from a culture with an OD_{600} between 0.2 and 0.3. The supernatant was discarded, and the cell pellet used for spheroplasting.

The cell pellet was resuspended in 20 mL of sterile water and transferred to a 50-mL conical tube. The suspension was centrifuged at 1500 x g for 5 min at R-T, and the supernatant was discarded. The cell pellet was further washed with 20 mL of fresh SED [1M sorbitol, 25 mM EDTA, 1M DTT, pH 8.0] and again centrifuged at

1500 x g for 5 min. The cells were then washed with 20 mL of 1 M sorbitol and centrifuged as previously described. The pellet was resuspended in 20 mL of SCE buffer [1M sorbitol, 1 mM EDTA, 10 mM sodium citrate, pH 5.8], and the suspension was divided between two 50-mL conical tubes. One tube was used to determine the optimal time of digestion with zymolyase to make spheroplasts.

Digestion of the cell wall was initiated by adding 7.5 μL (11 units) of enzyme to one conical tube of cells and incubating at 30 ^{o}C. Two-hundred μL aliquots of the digestion mixture were quenched with 800 μL of a 5% (w/v) SDS solution at two minute intervals for 40 min. Absorbance was monitored at 800 nm (1-cm pathlength) for each sample and used to determine the percent spheroplasting from the following equation:

% Spheroplasting = 100 – [(OD_{800} at time t / OD_{800} at time 0) x 100]

An optimal incubation period results in approximately 70% spheroplasting. After the optimal incubation period was determined (~20 min), 7.5 μL of zymolyase was added to the second conical of cells and incubated at 30 ^{o}C for the predetermined time period. The cells were harvested at 750 x g for 10 min at R-T, washed with 10 mL of CaS buffer [1 M sorbitol, 10 mM Tris-HCl, pH 7.5, 10 mM $CaCl_2$], harvested once more, and then resuspended in 0.6 mL of CaS buffer. Samples were used within 30 min.

4.2.4 Transformation of *P. pastoris*

Each transformation reaction comprised 100 μL of spheroplast mixture and 10 μg of linearized DNA. After 10 min of incubation at R-T, 1.0 mL of fresh PEG/ CaT

[40% PEG 3350 (w/v); 20 mM Tris-HCl, pH 7.5, 20 mM $CaCl_2$] solution was added to the cells. The sample was mixed gently and incubated again for 10 min at R-T. The cells were harvested (750 x g, 10 min, R-T), and the excess PEG/ CaT solution was carefully removed. The pellet was resuspended in 150 μL of SOS media (Invitrogen), incubated at R-T for 20 min. Sorbitol (850 μL of 1 M solution) was then added.

Between 100- and 300- μL of the spheroplast-DNA mixture was added to 10 mL of molten regeneration dextrose media [RD; 1 M sorbitol, 2% dextrose, 1.3% YNB, 4 x 10^{-5}% biotin, 0.005% amino acids] and poured onto RD agar plates. After the top agar hardened, the plates were inverted and incubated at 30 ^{o}C for 4 to 6 days.

4.2.5 Test expression of recombinant proteins

A successful transformation results in the reformation of *HIS4* and colony growth on his^- plates. The colonies produced are His^+ transformants that are either methanol utilization plus (Mut^+) or methanol utilization slow (Mut^S). Transformants resulting from a Sal I digest and transformation into GS115 should be Mut^+; however, there is a chance that recombination will disrupt the wild-type AOX1 gene (Fig. 4.4) and create a His^+ Mut^S transformant. The transformants, therefore, were screened to determine the phenotype.

Transformants were streaked onto minimal dextrose, [MD, 1.3% (w/v) yeast nitrogen base (YNB), 4 x 10^{-5}% (w/v) biotin, 2% (w/v) dextrose], and minimal methanol [MM, 1.3% (w/v) YNB, 4 x 10^{-5} % (w/v) biotin, 0.5% (w/v) methanol] agar

plates to isolate single colonies. The plates were incubated at 30 ^{o}C for 2 days. A His^{+} Mut^{+} phenotype results in faster growth on MM plates, resulting in larger colonies on MM plates compared to MD plates. A His^{+} Mut^{S} phenotype results in slower growth in the presence of methanol as manifested by smaller colonies on MM plates compared to MD plates. His^{+} Mut^{+} transformants were selected for test expression.

Recombination can occur in different ways that affects protein expression. It is necessary, therefore, to screen many transformants. Instead of conducting a test expression in liquid media on 10-20 colonies which would take approximately 10 days for growth and test expression, we developed a high-throughput screening method which is done in two days. The His^{+} Mut^{+} transformants were screened for protein expression by SDS-PAGE analysis on colonies from the MM plates.

The colonies were scraped off the plates with and resuspended in 30 μL of breaking buffer [50 mM sodium phosphate, pH 7.4, 1 mM phenylmethylsulfonyl fluoride (PMSF), 1 mM EDTA, and 5% (w/v) glycerol]. An equal volume of 0.5 mm acid-washed glass beads was added, and the samples were subjected to 8 cycles of 30 sec vortexing followed by 30 sec incubating on ice. The samples were then centrifuged at 12000 x g for 10 min at 4 ^{o}C. The supernatant was collected, mixed with 30 μL of 2X SDS-PAGE loading buffer [100 mM Tris-HCl, pH 6.8, 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (w/v) glycerol, 200 mM DTT], and boiled for 10 min. Fifteen μL were loaded into an 18% gel (Biorad) and electrphoresed for 50 min at 200 V. The gel was Coomassie-stained and then destained. α-SN gels

were also subjected to Western blot analysis. Plates with colonies producing the respective protein were saved for test expressions in liquid media.

The MD plate corresponding to the MM plate containing protein-expressing colonies was identified. A single colony from the MD plate was used to inoculate 25 mL of buffered glycerol-complex media [BMGY; 1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.3% (w/v) YNB, 4 x 10^{-5}% (w/v) biotin, 1% (w/v) glycerol] in a 250 mL baffled flask. This culture was grown for 16-18 h at 30 ^{o}C with shaking at 250 rpm. The cells were harvested in log phase (OD_{600} 2-6) by centrifugation at 1500 x g for 5 min at R-T. The pellet was resuspended in a buffered methanol-complex media [BMMY, 1% (w/v) yeast extract, 2% (w/v) peptone, 100 mM potassium phosphate, pH 6.0, 1.3% (w/v) YNB, 4 x 10^{-5}% (w/v) biotin, 0.5% (w/v) methanol] to an OD_{600} of 1.0. This solution was placed in a 1-L baffled flask, covered with two layers of cheesecloth, and returned to the incubator (250 rpm, 30 ^{o}C). Methanol was added to a final concentration of 0.5% twice a day. One mL aliquots were taken at 0, 6, 12, 24, 36, 48, 60, 72, 84, and 96 h. The aliquots were centrifuged at 12000 x g for 5 min at R-T. The supernatant was separated from the cell pellet, and both were kept for SDS-PAGE analysis.

For SDS-PAGE analysis, the cell pellets were thawed on ice and resuspended in 100 μL of breaking buffer. An equal volume of acid-washed glass beads was added to the solution. The solution was then subjected to 8 cycles of 30 sec vortexing followed by a 30 sec incubation on ice. The sample was centrifuged at 12000 x g for 10 min, and the clear cell lysate and pellet were separated. Fifty mL of 2X SDS-PAGE loading buffer was added to the clear cell lysate. Also, 50 μL of 2X

SDS loading buffer was added to 50 μL of the cellular supernatant from each time point to test for protein secretion. The samples were boiled for 10 min, and 15 μL from each sample was loaded onto an 18% polyacrylamide SDS gel and electrophoresed for 50 min at 200 V. The gel was Coomassie-stained and destained. Gels for α-SN were also subjected to Western blot analysis (Invitrogen primary antibody, Pierce secondary antibody).

4.3 Results

4.3.1 Expression and in-cell NMR of CI2

Expression of CI2 in *P. pastoris* was tested in two different media, buffered methanol-complex media (BMMY) and unenriched Yeastone® media. (Spectra Stable Isotopes). The complementary enriched Yeastone® media was used for in-cell NMR experiments with *P. pastoris.* Detectable levels of expression in BMMY media appeared after 16 h on a Coomassie-stained gel (Fig. 4.5). Similar results were obtained for expression in the Yeastone® media (Fig. 4.6).

Favorable expression levels in unenriched Yeastone® prompted in-cell NMR experiments. *P. pastoris* cells expressing ^{15}N-enriched CI2 were harvested after 24 h and resuspended in unenriched Yeastone®. An ^{15}N HSQC spectrum was collected on these cells for 16 h (Fig. 4.7). The resulting spectrum was of background molecules. CI2 was not detected inside cells by NMR but was observed on a Coomassie-stained SDS gel (Fig. 4.8).

4.3.2 Expression and in-cell NMR of α-SN

The initial test expression was conducted without screening for colonies producing α-SN on MM plates. The test expression in BMMY media resulted in no detectable level of expression on a Coomassie-stained SDS gel (Fig. 4.9) or a Western blot with an anti-synuclein antibody (Invitrogen) (data not shown). Next, we screened colonies grown on minimal methanol (MM) agar plates for the presence of α-SN and observed protein production in 2 out of 15 colonies (Fig. 4.10).

A test expression with these colonies in BMMY and Yeastone® was unsuccessful. Western blot analysis showed no detectable level of expression (Fig. 4.11) from the clear cell lysate. α-SN, however, was found in the insoluble cell lysate. Insoluble α-SN accumulated after 48 h of expressing in both Yeastone and BMMY media (Fig. 4.12 and Fig. 4.13) leading us to limit further experiments to less than 48 h.

An in-cell NMR experiment was conducted on *P. pastoris* cells expressing ^{15}N-enriched α-SN. A 16 h ^{15}N HSQC spectrum from cells expressing for 24 h resulted in signal from only background molecules (Fig. 4.14). A Coomassie-stained SDS gel (Fig. 4.15) and a Western blot (figure not shown) of lysed cells revealed that α-SN was present in the soluble and insoluble fractions.

4.3.3 Expression of FlgM

The initial test expression of FlgM in BMMY media resulted in no distinguishable band on a Coomassie-stained SDS gel (Fig. 4.16). Since anti-FlgM

antibody is not available, Western blot analysis could not be done. To determine if FlgM was expressed, we inserted a 6x histidine tag, for which an antibody is available, transformed, and tested for expression.

After successful mutagenesis and transformation reactions, colonies grown on MM agar plates were tested for presence of FlgM. The initial screen revealed one colony producing His-FlgM (Fig. 4.17). The test expression in BMMY and Yeastone media, however, showed no detectable level of His-FlgM (Fig. 4.18). A secondary screen for His-FlgM, along with α-SN and CI2, resulted in no colonies producing His-FlgM, including the initial colony (Fig. 4.19). CI2 was detected in a colony grown on an MM agar plate, thereby validating the method (Fig. 4.19). Further progress in expressing His-FlgM in *P. pastoris* is needed before proceeding with in-cell NMR experiments.

4.4 Conclusions

The high viscosity of the cytoplasm is a deterrent to measuring proteins inside living cells by NMR. We were unable to observe the globular protein CI2 inside *E. coli*, where the viscosity at least 9 times greater than that of water (*16*). The cytoplasmic viscosity of yeast cells is greater than that of *E. coli* (*86-88*); therefore, the lack of observable CI2 inside *P. pastoris* was not unexpected.

Recent results suggest that natively-disordered proteins are the most promising for in-cell NMR in eukaryotic cells (Li, *et al*., *in press*, JACS). The flexible backbone of disordered proteins is affected less by the increased viscosity of the

cytoplasm than a globular protein. Within this group, we have observed FlgM, α-SN, and tau inside living *E. coli* but did not detect α-SN inside *P. pastoris*.

Another necessary component for observing proteins inside cells by NMR is high level of expression. α-SN, however, forms cytoplasmic inclusion bodies in a concentration-dependent manner in *Saccharomyces cervasiae* (*89*) and associates with the plasma membrane (*89*). These findings are in agreement with our observations of α-SN separating into the insoluble fraction. In both instances, α-SN would behave like a larger, globular protein and be affected by the increased viscosity of the yeast cytoplasm making it difficult, if not impossible, to observe inside *P. pastoris.*

The *P. pastoris* system promotes high expression levels along with posttranslational modifications, making it an interesting target for in-cell NMR. The main advantage of the system is a significant disadvantage for in-cell NMR. Instead of significantly increasing protein concentration per cell, however, *P. pastoris* yields gram quantities of protein by increasing cell density (*90, 91*). Not only is it then difficult to observe the protein of interest over the background molecules but also the spectral quality deteriorates with increasing cell density in the NMR tube (*16*).

4.5 Figures

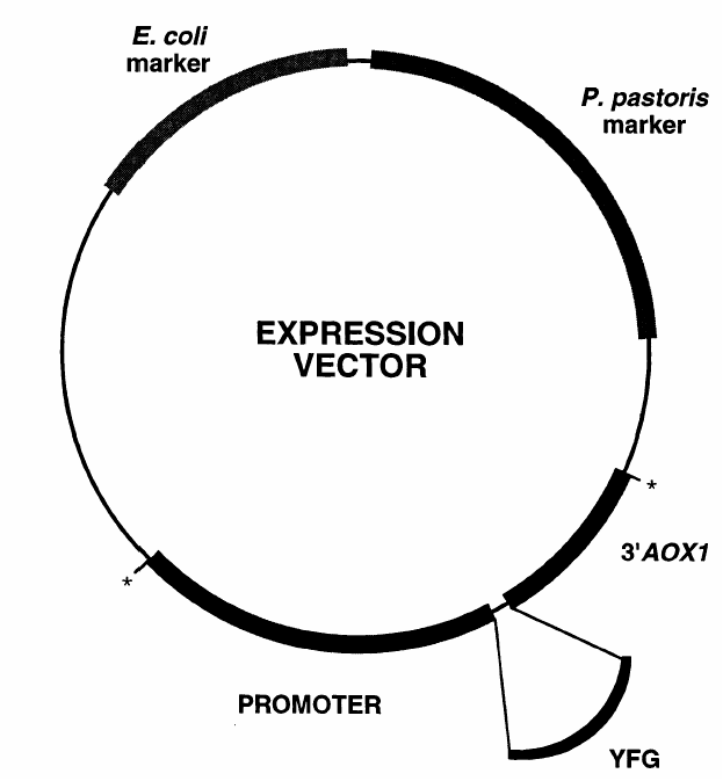

Figure 4.1. General diagram of *P. pastoris* vector

"YFG" is your favorite gene (*83*). The stars enclose the recombination sequence.

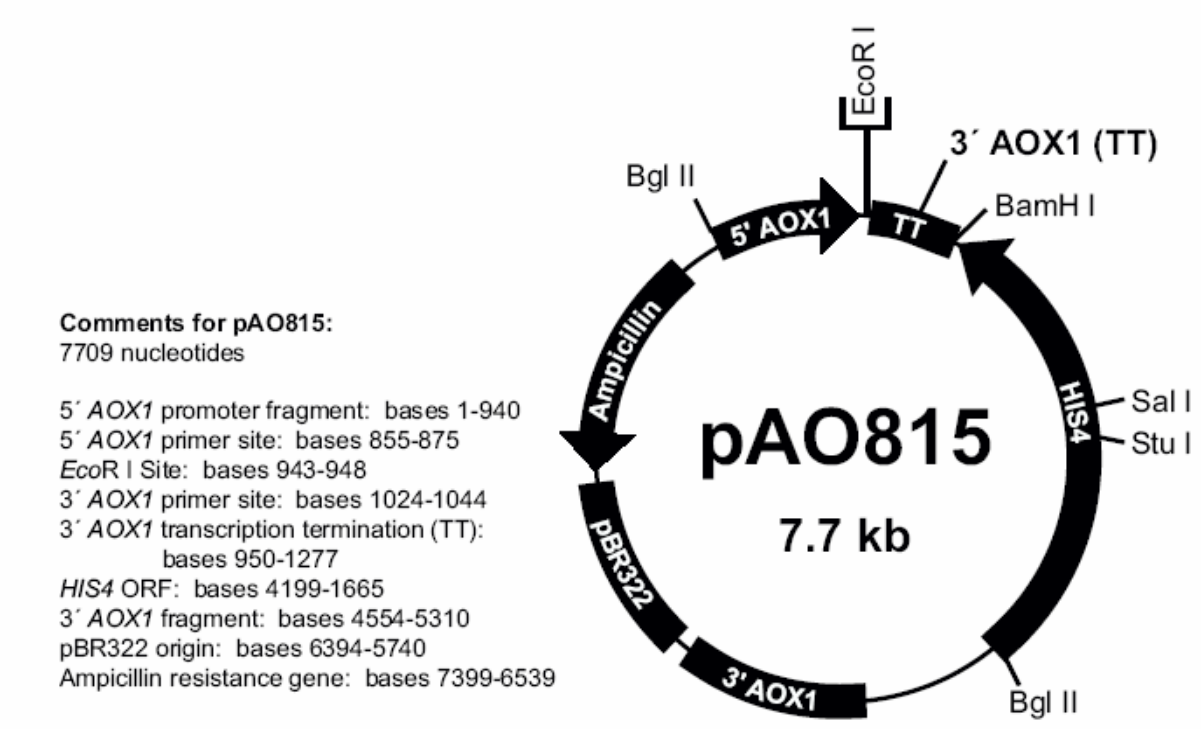

Figure 4.2. Yeast plasmid map used in present studies
(Invitrogen)

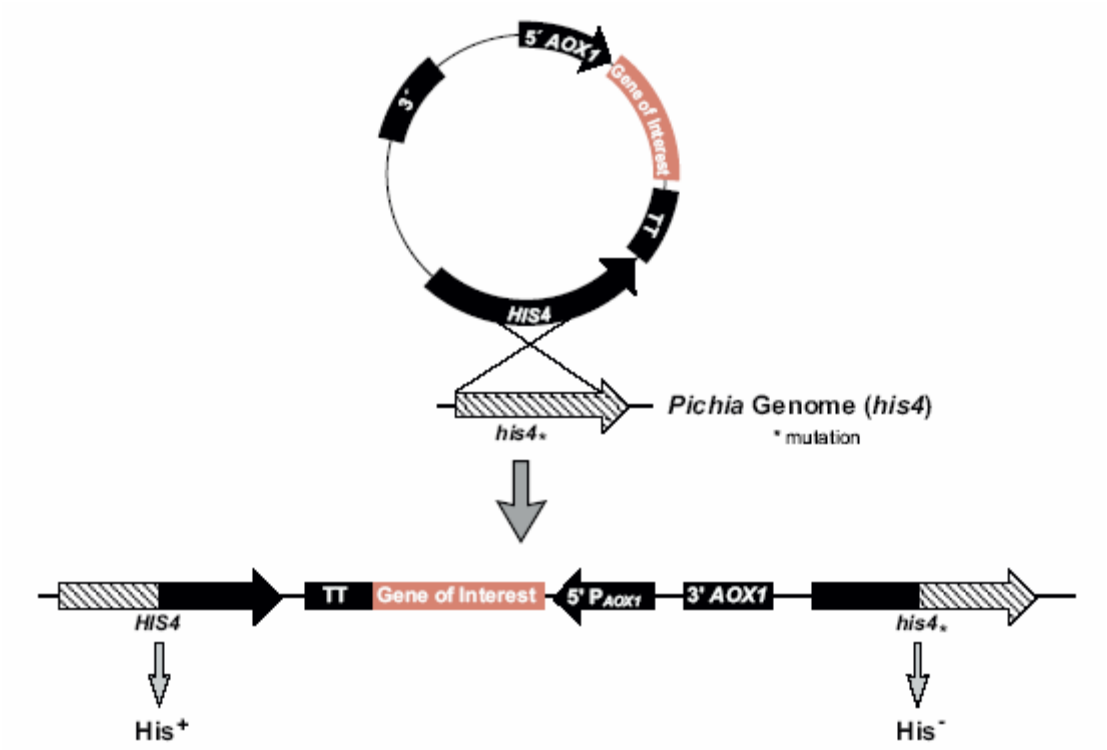

Figure 4.3. Single crossover during a homologous recombination event with *HIS4* selectable marker
(Invitrogen)

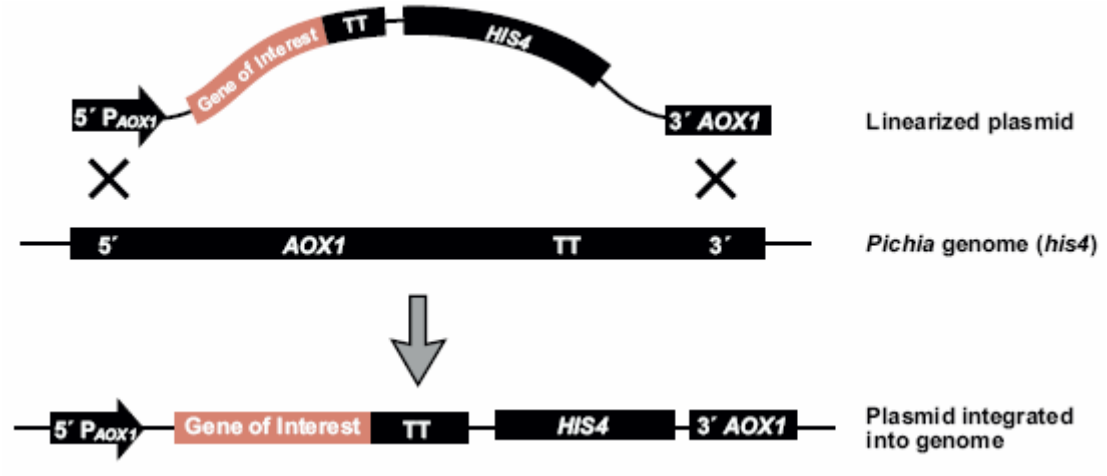

Figure 4.4. Double crossover event during homologous recombination
A double crossover event between the AOX1 promoter and 3' AOX1 regions of the vector and genomic DNA results in the complete removal of the AOX1 coding region. The resulting phenotype is $His^{+}Mut^{S}$. (Invitrogen)

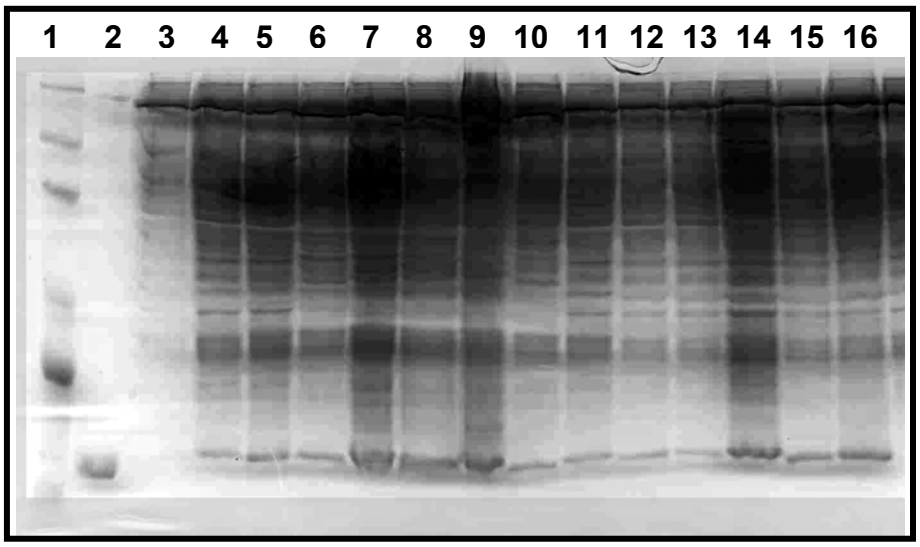

Figure 4.5. Coomassie-stained gel of test expression for chymotrypsin inhibitor 2 in *P. pastoris*
Representative SDS-PAGE gel of a test expression for CI2 in *P. pastoris* in BMMY media. **Lanes: 1,** 755 Marker; **2,** CI2 standard; **3,** Test expression 1, 9 h; **4,** Test expression 1, 16 h; **5,** Test expression 1, 24 h; **6,** Test expression 1, 40 h; **7,** Test expression 1, 48 h; **8,** Test expression 1, 72 h; **9,** Test expression 1, 96 h; **10,** Test expression 2, 9 h; **11,** Test expression 2, 16 h; **12,** Test expression 2, 24 h; **13,** Test expression 2, 40 h; **14,** Test expression 2, 48 h; **15,** Test expression 2, 72 h; **16,** Test expression 2, 96 h. Test expression 1 and 2 indicate two different colonies.

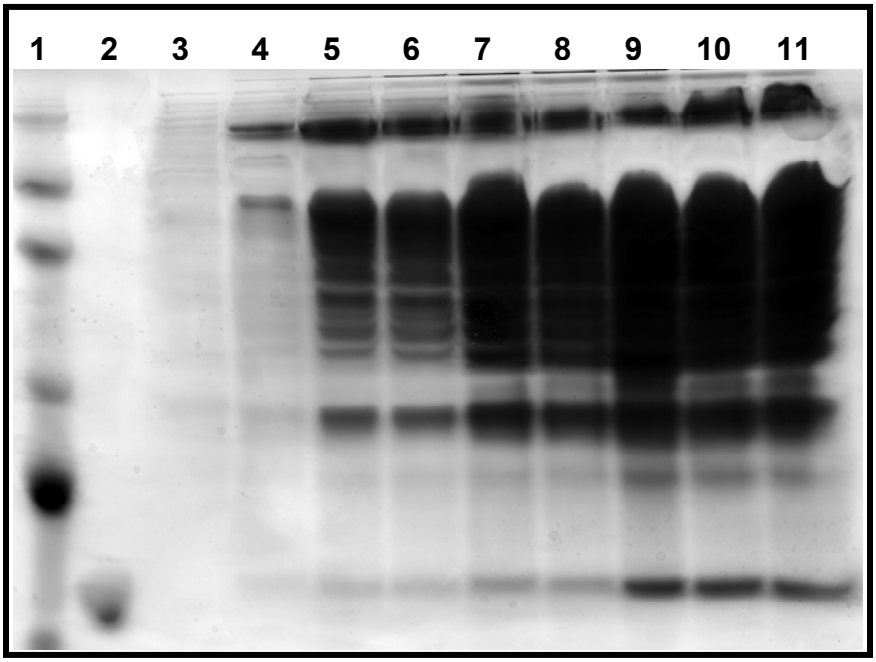

Figure 4.6. Coomassie-stained SDS gel of chymotrypsin inhibitor 2 test expression in *P. pastoris* in Yeastone media
Lanes: 1, Marker; **2,** CI2 standard; **3,** 0 h; **4,** 6 h **5,** 12 h; **6,** 24 h; **7,** 36 h; **8,** 48 h; **9,** 60 h; **10,** 72 h; **11,** 96 h.

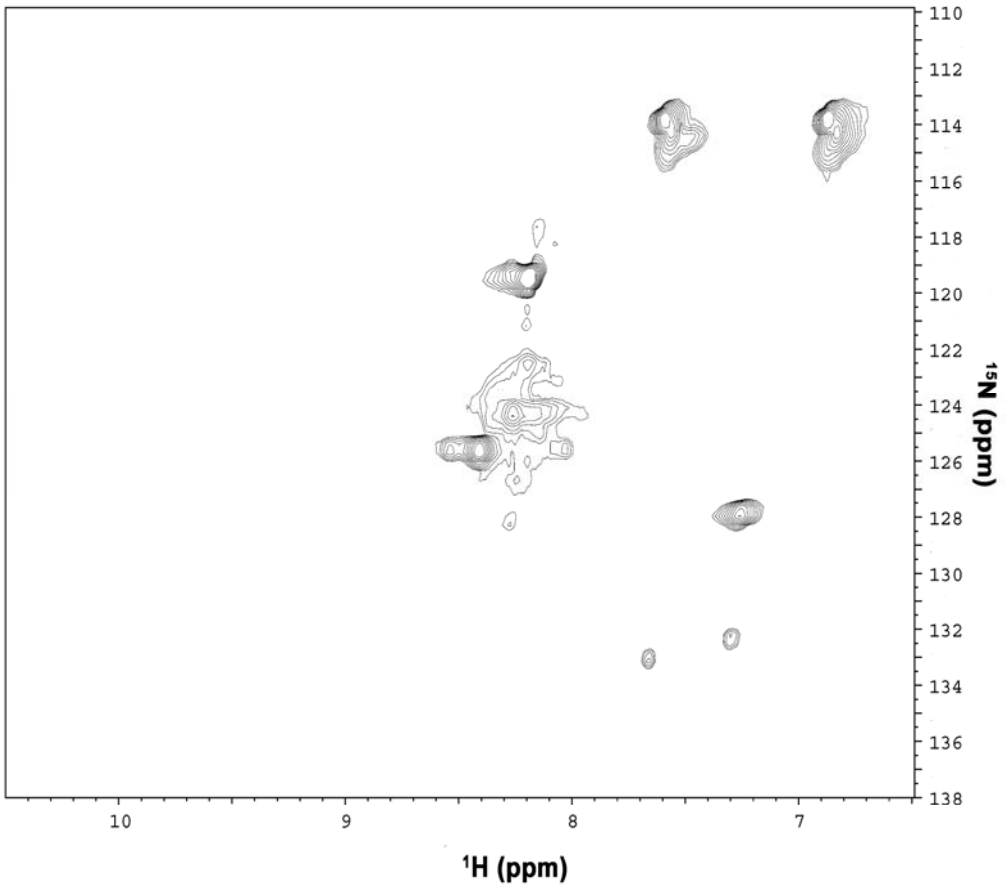

Figure 4.7. NMR spectrum *P. pastoris* cells expressing chymotrypsin inhibitor 2

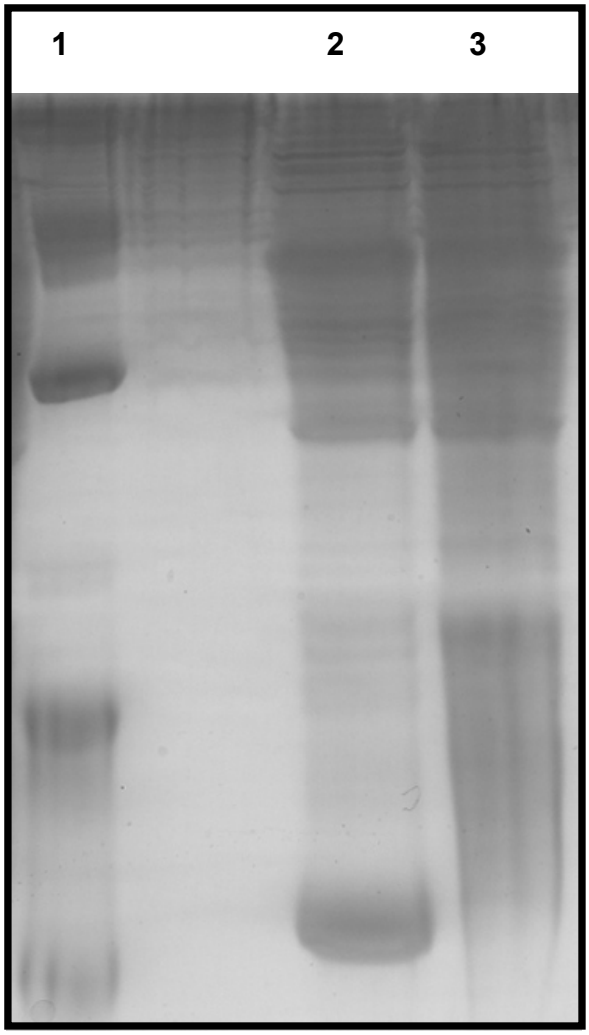

Figure 4.8. Coomassie-stained SDS gel of chymotrypsin inhibitor 2 after in-cell NMR experiment
Lanes: 1, 755 Marker; **2,** clear cell lysate; **3,** insoluble material

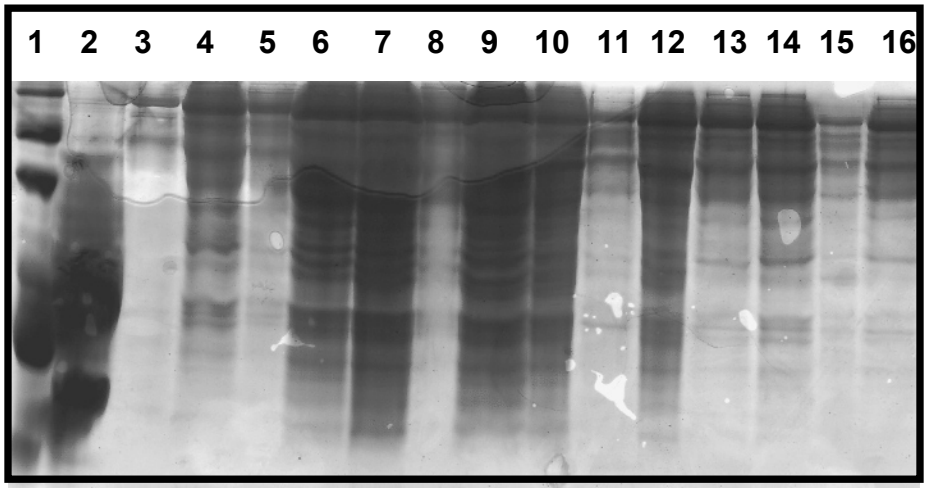

Figure 4.9. Coomassie-stained SDS gel for test expression of α-synuclein in *P. pastoris* in BMMY media

Lanes: 1, 755 Marker; **2,** α-SN standard; **3,** Test expression 1, 9 h; **4,** Test expression 1, 16 h; **5,** Test expression 1, 24 h; **6,** Test expression 1, 40 h; **7,** Test expression 1, 48 h; **8,** Test expression 1, 72 h; **9,** Test expression 1, 96 h; **10,** Test expression 2, 9 h; **11,** Test expression 2, 16 h; **12,** Test expression 2, 24 h; **13,** Test expression 2, 40 h; **14,** Test expression 2, 48 h; **15,** Test expression 2, 72 h; **16,** Test expression 2, 96 h. Test expression 1 and 2 indicate two different colonies.

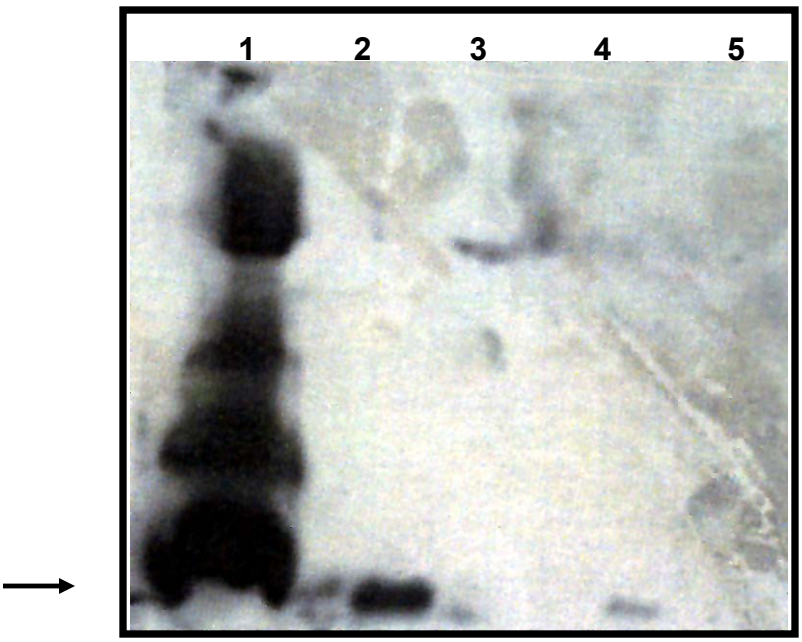

Figure 4.10. Representative Western blot analysis for presence of α-synuclein from screening colonies grown on MM agar plates. Lanes: 1, α-SN standard; **2,** colony 1; **3,** colony 2; **4,** colony 3; **5,** colony 4 **.** The arrow indicates monomeric α-synuclein.

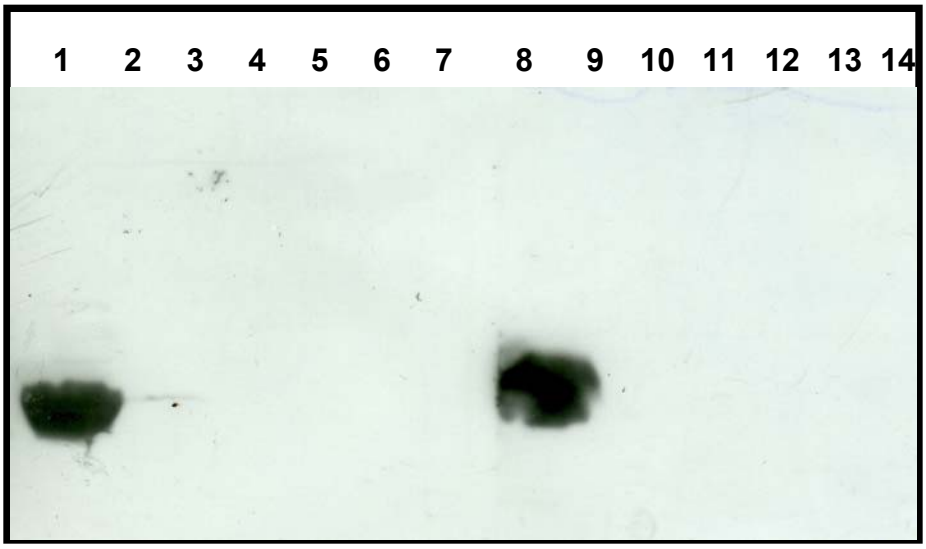

Figure 4.11. Western blot analysis of α-synuclein test expression in BMMY and Yeastone medias.
Lanes: 1, α-SN standard; **2,** BMMY media, 6 h; **3,** BMMY media, 12 h; **4,** BMMY media, 24 h; **5,** BMMY media, 36 h; **6,** BMMY media, 48 h; **7,** BMMY media, 72 h; **8,** α-SN standard; **9,** Yeastone media, 6 h; **10,** Yeastone media, 12 h; **11,** Yeastone media, 24 h; **12,** Yeastone media, 36 h; **13,** Yeastone media, 48 h; **14,** Yeastone media 72 h

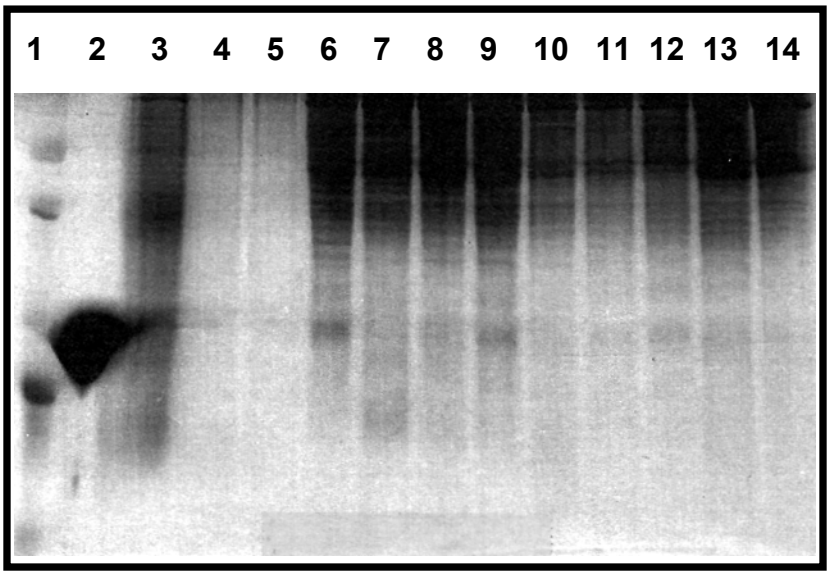

Figure 4.12. Coomassie-stained SDS gel of α-synuclein insoluble material from test expression with *P. pastoris.*
Lanes: 1, Marker; 2, □-SN standard; 3, Test expression 1, BMMY media, 24 h; 4, Test expression 1, BMMY media, 48 h; 5, Test expression 1, BMMY media, 72 h; 6, Test expression 2, BMMY media, 24 h; 7, Test expression 2, BMMY media, 48 h; 8, Test expression 2, BMMY media, 72 h; 9, Test expression 1, Yeastone, 24 h; 10, Test expression 1, Yeastone, 48 h; 11, Test expression 1, Yeastone, 72 h; 12, Test expression 2, Yeastone, 24 h; 13, Test expression 2, Yeastone, 48 h; 14, Test expression 2, Yeastone, 72 h. Test expression 1 and 2 indicate two different colonies.

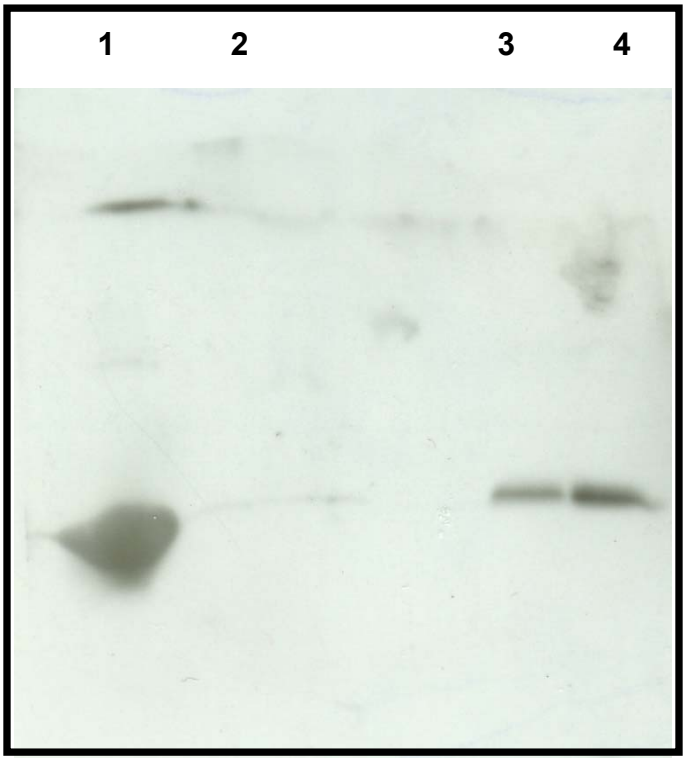

Figure 4.13. Representative gel of Western blot analysis of insoluble α-synuclein from test expression in BMMY and Yeastone.
Lanes: 1, α-SN standard; **2,** Test expression 1, Yeastone media, 24 h; **3,** Test expression 1, Yeastone media, 48 h; **4,** Test expression 1, Yeastone media, 72 h

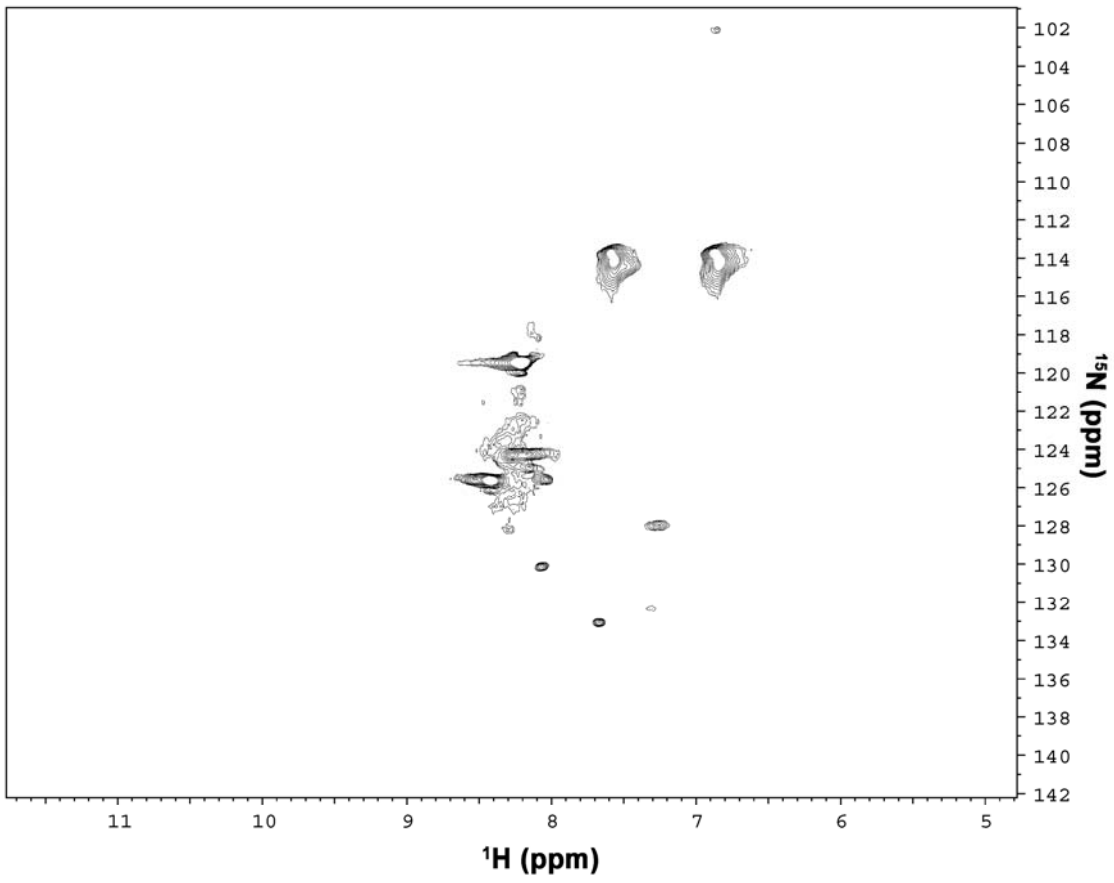

Figure 4.14. NMR spectrum of in-cell experiment for α-synuclein

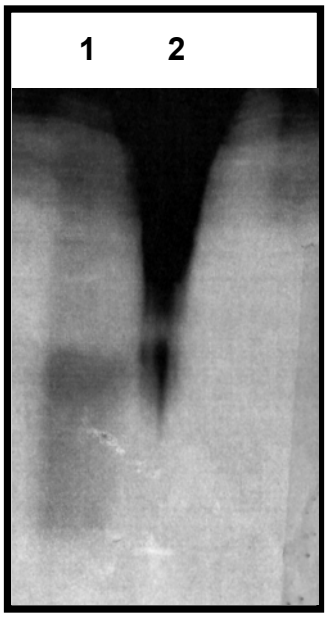

Figure 4.15. Western blot of α-synuclein samples after an in-cell NMR experiment with *P. pastoris*.
Lanes: 1, clear cell lysate; **2,** insoluble material

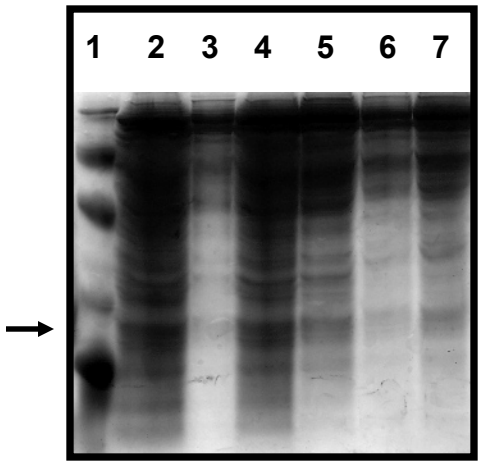

Figure 4.16. Coomassie-stained SDS gel of His-FlgM test expression in *P. pastoris* in BMMY media.
Lanes: 1, Marker; **2,** 6 h; **3,** 12 h; **4,** 24h, **5,** 36 h; **6,** 48 h, **7,** 72 h. The arrow dictates the probable location of FlgM.

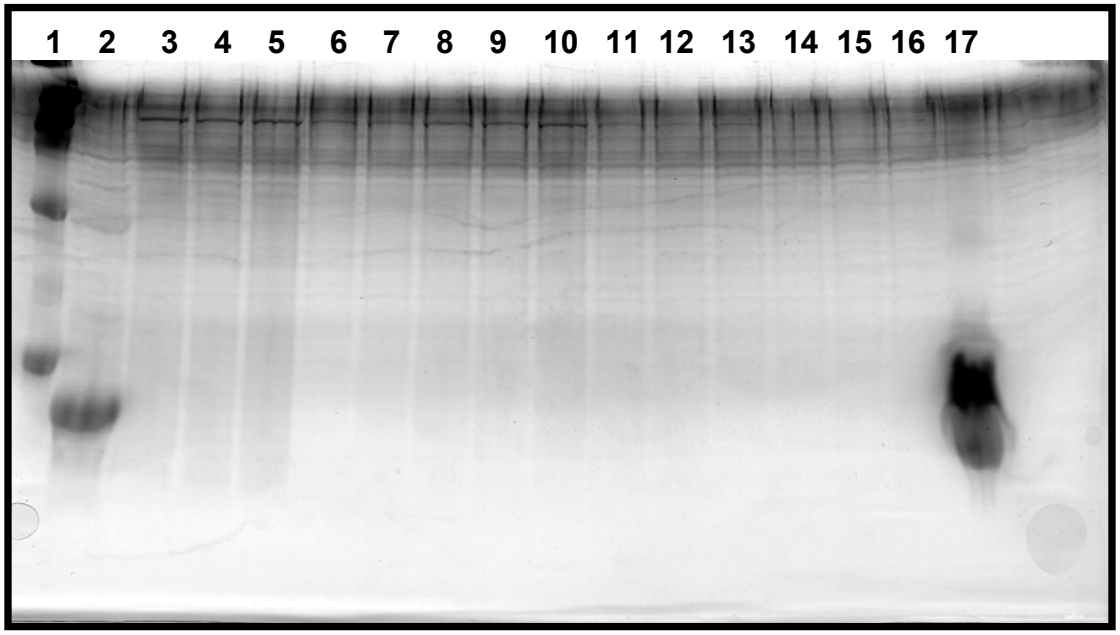

Figure 4.17. Coomassie-stained SDS-PAGE analysis for presence of His-FlgM after screening colonies grown on MM agar plates.
Lanes: 1, FlgM standard; **2,** colony 1; **3,** colony 2; **4,** colony 3; **5,** colony 4; **6,** colony 5; **7,** colony 6; **8,** colony 7; **9,** colony 8; **10,** colony 9; **11,** colony 10; **12,** colony 11; **13,** colony 12; **14,** colony 13; **15,** colony 14; **16,** colony 15; **17,** colony 16

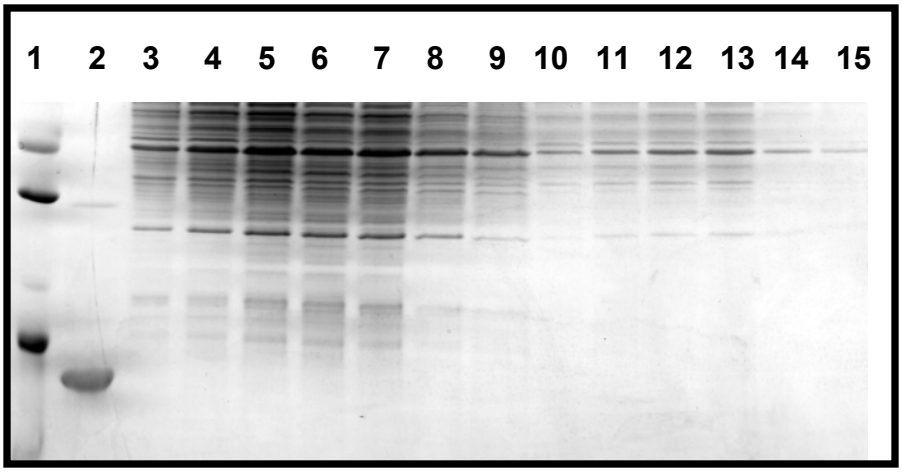

Figure 4.18. Coomassie-stained SDS gel of His-FlgM from test expression of *P. pastoris* in BMMY and Yeastone media.
Lanes: 1, Marker; **2,** FlgM standard; **3,** BMMY media, 6h; **4,** BMMY media, 12h; **5,** BMMY media, 24 h; **6,** BMMY media, 36 h; **7,** BMMY media, 48 h; **8,** BMMY media, 72 h; **9,** BMMY media, 96 h; **10,** Yeastone media, 6h; **11,** Yeastone media, 12h; **12,** Yeastone media, 24 h; **13,** Yeastone media, 36 h; **14,** Yeastone media, 48 h; **15,** Yeastone media 72 h

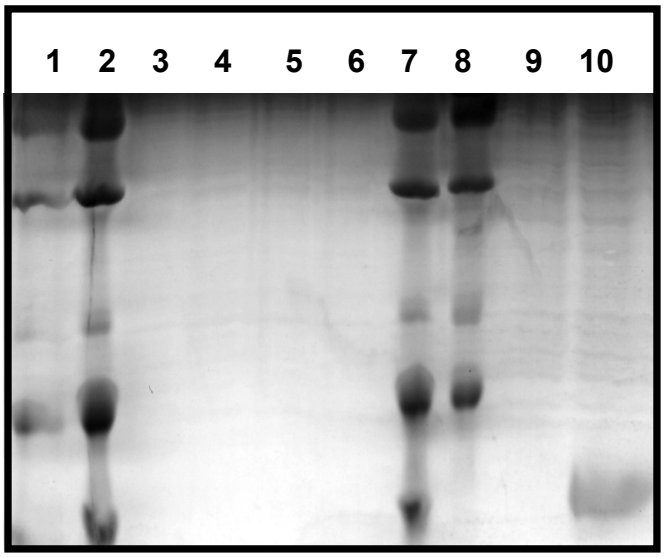

Figure 4.19. Coomassie-stained SDS-PAGE analysis for screening the presence of His-FlgM, α-synculein, or chymotrypsin inhibitor 2 in *P. pastoris* colonies grown on MM plates.
Lanes: 1, 756 Marker 1; **2,** 755 Marker; **3,** colony 1; **4,** colony 2; **5,** colony 3; **6,** colony 4; **7,** 755 Marker; **8,** 756 Marker; **9,** α-SN, **10,** CI2

References

1. T. C. Laurent, *Biophys. Chem.* **57**, 7-14 (1995).

2. A. G. Ogston, C. F. Phelps, *Biochem. J.* **78**, 827-833 (1961).

3. A. G. Ogston, *J. Chem. Soc., Faraday Trans.* **54**, 1754-1757 (1958).

4. R. H. Pearce, T. C. Laurent, *Biochem. J.* **163**, 617-625 (1977).

5. R. J. Ellis, *Trends Biochem. Sci.* **26**, 597-604 (2001).

6. F. H. Gaertner, *Trends Biochem. Sci.* **3**, 63-65 (1978).

7. J. Ovadi, V. Saks, *Mol. Cell Biochem.* **256/257**, 5-12 (2004).

8. P. A. Srere, J. Ovadi, *FEBS Lett.* **268**, 360-364 (1990).

9. A. P. Minton, *Curr. Opin. Struct. Biol.* **10**, 34-39 (2000).

10. J. M. Rohwer, P. W. Postma, B. N. Kholodenko, H. V. Westerhoff, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 10547-10552 (1998).

11. L. Acerenza, M. Grana, *J. Mol. Evol.* **63**, 583-590 (2006).

12. R. Hancock, *J. Struct. Biol.* **146**, 281-290 (2003).

13. J. Walter, L. Sun, J. Newport, *Mol. Cell* **1**, (1998).

14. T. C. Jarvis, D. M. Ring, S. S. Daube, P. H. v. Hippel, *J. Biol. Chem.* **265**, 15160-15167 (1990).

15. M. Oldiges *et al.*, *Appl. Microbiol. Biotechnol.* **76**, 495-511 (2007).

16. Z. Serber, R. Ledwidge, S. M. Miller, V. Dotsch, *J. Am. Chem. Soc.* **123**, 8895-8901 (2001).

17. S. P. Williams, P. M. Haggie, K. M. Brindle, *Biophys. J.* **72**, 490-498 (1997).

18. Z. Serber *et al.*, *J. Am. Chem. Soc.* **123**, 2446-2447 (2001).

19. M. M. Dedmon, C. N. Patel, G. B. Young, G. J. Pielak, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12681-12684 (2002).

20. B. C. McNulty, G. B. Young, G. J. Pielak, *J. Mol. Biol.* **355**, 893-897 (2006).

21. J. A. Hubbard, L. K. MacLachlan, G. W. King, J. J. Jones, A. P. Fosberry, *Mol. Microbiol.* **49**, 1191-1200 (2003).

22. D. S. Burz, K. Dutta, D. Cowburn, A. Shekhtman, *Nat. Methods* **3**, 91-93 (2006).

23. Z. Serber *et al.*, *J. Am. Chem. Soc.* **126**, 7119-7125 (2004).

24. N. Shimba *et al.*, *Biochemistry* **42**, 9227-9234 (2003).

25. P. Selenko, Z. Serber, B. Gade, J. Ruderman, G. Wagner, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 11904-11909 (2006).

26. P. Selenko, G. Wagner, *J. Struct. Biol.* **158**, 244-253 (2007).

27. A. P. Minton, *Biophys. J.* **88**, 971-985 (2005).

28. K. J. Luby-Phelps, *Int. Rev. Cytol.* **192**, 189-221 (2000).

29. A. P. Minton, *J. Pharm. Sci.* **94**, 1668-1675 (2005).

30. K. Sasahara, P. McPhie, A. P. Minton, *J. Mol. Biol.* **326**, 1227-1237 (2003).

31. X. Ai, Z. Zhou, Y. Bai, Wing-Yiu-Choy, *J. Am. Chem. Soc.* **128**, 3916-3917 (2006).

32. Y. Qu, D. W. Bolen, *Biophys. Chem.* **101-102**, 155-165 (2002).

33. D. S. Spencer, K. Xu, T. M. Logan, H. X. Zhou, *J. Mol. Biol.* **351**, 219-232 (2005).

34. B. van den Berg, R. Wain, C. M. Dobson, R. J. Ellis, *EMBO J.* **19**, 3870-3875 (2000).

35. B. Monterroso, A. P. Minton, *J. Biol. Chem.* **282**, 33452-33458 (2007).

36. J. L. Neira, L. S. Itzhaki, D. E. Otzen, B. Davis, A. R. Fersht, *J. Mol. Biol.* **270**, 99-110 (1997).

37. S. E. Jackson, A. R. Fersht, *Biochemistry* **30**, 10428-10435 (1991).

38. L. S. Itzhaki, J. L. Neira, A. R. Fersht, *J. Mol. Biol.* **270**, 89-98 (1997).

39. S. E. Jackson, M. Moracci, N. elMasry, C. M. Johnson, A. R. Fersht, *Biochemistry* **33**, 11259-11269 (1993).

40. H. A. Ravin, A. M. Seligman, J. Fine, *New Engl. J. Med.* **247**, 921-929 (1952).

41. C. M. Klech, A. E. C. III, A. B. S. III, *Colloid and Polym. Sci.* **269**, 643-649 (1991).

42. A. N. Lukyanov, V. P. Torchilin, *Adv. Drug. Deliv. Res.* **56**, 1273-1289 (2004).

43. S. O. Rogero *et al.*, *Artif. Organs* **27**, 424-427 (2003).

44. J. J. O'Malley, R. W. Ulmer, *Biotechnol. Bioeng.* **25**, 917-925 (1973).

45. W. R. Gombotz *et al.*, *Pharm. Res.* **11**, 624-632 (1994).

46. H. Kamada *et al.*, *Cancer Res.* **60**, 6416-6420 (2000).

47. K. R. Roesler, A. G. Rao, *Protein Eng.* **12**, 967-973 (1999).

48. L. E. Kay, P. Keifer, T. Saarinen, *J. Am. Chem. Soc.* **114**, 10663-10665 (1992).

49. G. Bodenhausen, D. J. Ruben, *Chem. Phys. Lett.* **69**, 185-189 (1980).

50. F. Delaglio *et al.*, *J. Biomol. NMR* **6**, 277-293 (1995).

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

52. M. Wittekind, L. Mueller, *J. Magn. Reson., Ser. B* **101**, 201-205 (1993).

53. S. Grzesiek, A. Bax, *J. Am. Chem. Soc.* **114**, 6291-6293 (1992).

54. Y.-Z. Zhang, thesis, University of Pennsylvania (1995).

55. T.-L. Hwang, P. C. M. van Zijl, S. Mori, *J. Biomol. NMR* **11**, 221-226 (1998).

56. Y. Nozaki, *Methods Enzymol.* **26**, 43-50 (1972).

57. M. L. Tillett, L.-Y. Lian, T. J. Norwood, *J. Magn. Res.* **133**, 379-384 (1998).

58. F. Ferrage, M. Zoonens, D. E. Warschawski, J.-L. Popot, G. Bodenhausen, *J. Am. Chem. Soc.* **125**, 2541-2545 (2003).

59. I. V. Nesmelova, D. Idiyatullin, K. H. Mayo, *J. Magn. Res.* **166**, 129-133 (2004).

60. A. P. Minton, *Mol. Cell Biochem.* **55**, 119-140 (1983).

61. M. Perham, L. Stagg, P. Wittung-Stafshede, *FEBS Lett.* **581**, 5065-5069 (2007).

62. L. Stagg, S.-Q. Zhang, M. S. Cheung, P. Wittung-Stafshede, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 189671-118981 (2007).

63. A. Hvidt, S. O. Nielsen, *Advan. Protein Chem.* **21**, 187-386 (1966).

64. J. Clarke, L. S. Itzhaki, *Curr. Opin. Struct. Biol.* **8**, 112-118 (1998).

65. C. K. Woodward, B. D. Hilton, *Biophys. J.* **32**, 561-575 (1980).

66. S. W. Englander, N. R. Kallenbach, *Q. Rev. Biophys.* **16**, 522-655 (1983).

67. Sphere, *http://www.fccc.edu/research/labs/roder/sphere/sphere.html*.

68. Y. Bai, J. S. Milne, L. Mayne, S. W. Englander, *Proteins* **17**, 75-86 (1993).

69. N. Tokuriki *et al.*, *Prot. Sci.* **13**, 125-133 (2004).

70. Y. Bai, T. R. Sosnick, L. Mayne, S. W. Englander, *Science* **269**, 192-196 (1995).

71. S. N. Timasheff, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 9721-9726 (2002).

72. R. L. Baldwin, *J. Mol. Biol.* **371**, 283-301 (2007).

73. T. S. Davison *et al.*, *J. Mol. Biol.* **307**, 605-617 (2001).

74. A. Tuteja, M. E. Mackay, *Nano Lett.* **7**, 1276-1281 (2007).

75. C. W. Mullineaux, A. Nenniger, N. Ray, C. Robinson, *J. Bacteriol.* **188**, 3442-3448 (2006).

76. A. M. Mastro, M. A. Babic, W. D. Taylor, A. D. Keith, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3414-3418 (1984).

77. B. C. McNulty *et al.*, *Prot. Sci.* **15**, 602-608 (2006).

78. C. Cruzeiro-Silva, F. P. Albernaz, A. P. Valente, F. C. L. Almeida, *Cell Biochem. Biophys.* **44**, 497-502 (2006).

79. J. E. Bryant, J. T. J. Lecomte, A. L. Lee, G. B. Young, G. J. Pielak, *Biochemistry* **46**, 27 (2007).

80. J. L. Cereghino, J. M. Cregg, *FEMS Microbiol. Lett.* **24**, 45-66 (2000).

81. J. F. Tschopp, P. F. Brust, J. M. Cregg, C. A. Stillman, T. R. Gingeras, *Nucleic Acids Res.* **15**, 3859-3876 (1987).

82. S. B. Ellis *et al.*, *Mol. Cell Biol.* **5**, (1985).

83. J. M. Cregg, K. R. Madden, *Development Industrial Microbiology* **29**, 33-41 (1988).

84. S. Macauley-Patrick, M. L. Fazenda, B. McNeil, L. M. Harvey, *Yeast* **22**, 249-270 (2005).

85. T. Sakai *et al.*, *J. Biomol. NMR* **36**, 179-188 (2006).

86. R. Swaminathan, C. P. Hoang, A. S. Verkman, *Biophys. J.* **72**, 1900-1907 (1997).

87. S. Nolan, A. E. Cowan, D. E. Koppel, E. Grote, *Molecular Biology of the Cell* **17**, 2439-2450 (2006).

88. M. B. Elowitz, M. G. Surette, P. E. Wolf, J. B. Stock, S. Leibler, *J Bacteriol* **181**, 197-203 (1999).

89. T. F. Outeiro, S. Lindquist, *Science* **302**, 1772-1775 (2003).

90. M. A. Romanos, C. A. Scorer, J. J. Clare, *Yeast* **8**, 423-488 (1992).

91. D. R. Higgins, J. M. Cregg, *Methods in Molecular Biology: Pichia Protocols* (Humana Press, Totowa, NJ, 1998)