Investigating the Aggregation of the Basic Leucine Zipper (bZIP) Domain of Activating Transcription Factor 5 (ATF5)

By

© 2010
Natalie A. Ciaccio
B.S. Pharmacy, Purdue University 2001
M.S. Pharmaceutical Chemistry, The University of Kansas 2008

Submitted to the graduate degree program in Pharmaceutical Chemistry and the graduate faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Committee: _____
Chairperson

_____

_____

_____

_____

Date Defended: _____

The Dissertation Committee for Natalie A. Ciaccio certifies that this is the approved version of the following dissertation:

Investigating the Aggregation of the Basic Leucine Zipper (bZIP) Domain of Activating Transcription Factor 5 (ATF5)

Committee: _____
Chairperson

_____

_____

_____

_____

Date Defended: _____

I dedicate this dissertation to my parents, Nick and Nancy, whose unwavering love and support helped me to achieve this goal.

ABSTRACT

Protein aggregation is a major problem for biopharmaceuticals. Aggregate formation in a drug formulation can have serious health implications for the patient. While the control of protein aggregation is critically important for the future of protein pharmaceuticals, the mechanism is still poorly understood. In particular, the role that protein structure plays in aggregate assembly is not well understood. We have selected the basic leucine zipper (bZIP) domain of activating transcription factor 5 (ATF5) as model system with which to investigate the relationship between protein structure and aggregate assembly. This domain contains three regions with differing structural propensity: a disordered N-terminal polybasic region possessing transient helicity; a central, helical leucine zipper region; and a C-terminal, extended valine zipper region with a propensity to form beta structure. Additionally, a centrally positioned cysteine residue readily forms an intermolecular disulfide bond. We have modulated solution conditions and engineered mutations that affect the structure and aggregation of this domain in order to characterize how different structural elements participate in aggregate formation. Specifically, we have evaluated the impact of intermolecular disulfide bond formation on ATF5 structure and stability. We have also investigated how removal of the C-terminal valine zipper region affects ATF5 structure and aggregation.

Our results indicate that intermolecular disulfide bond formation facilitates the retention of helical structure and reduces the growth of thermally induced protein aggregates. Additionally, the C-terminal valine zipper region is critical for the

formation of α-helical structure. Removal of this region results in a change in protein structure and a change in the mechanism of protein self-association. The structure and stability of this truncated mutant are largely unaffected by intermolecular disulfide bond formation. When compared to the wild-type ATF5 protein, this mutant displays increased self-association at low temperature but improved resistance to aggregate growth upon temperature elevation.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, Dr. Jennifer Laurence, who has been a source of inspiration and encouragement throughout my graduate training. I also wish to thank past and present members of the Laurence lab for their support, especially Drew Vartia, Mary Krause, Talia Martin and Amanda Glass for their friendship and encouragement. I thank Dr. Russ Middaugh and Middaugh lab members for shared use of their instrumentation and discussion of the data. I especially thank Dr. Brooke Barrett, Dr. Reza Esfandiary, Dr. Aaron Markham, Dr. Tim Priddy and Dr. Chris Olsen for instrument training and technical assistance. I thank Dr. Todd D. Williams of the KU mass spectrometry laboratory for his efforts in acquiring the ESI spectra as well as Dr. David Moore and Heather Shinogle at the KU Microscopy and Imaging Laboratory for the use of their instrumentation and guidance in performing the densitometry analysis. Several undergraduate students assisted with the research presented here. I thank Matt Moreno, Rachel Bauer and T. Steele Reynolds for their efforts. I would like to thank the members of my qualifying exam committee, Dr. Jennifer Laurence, Dr. Jeff Krise, Dr. Christian Schoeneich, Dr. Valentino Stella and Dr. Mario Rivera, for their time and assistance. I acknowledge and thank the members of my dissertation committee, Dr. Jennifer Laurence, Dr. Cory Berkland, Dr. Christian Schoeneich, Dr. Russ Middaugh, Dr. Teruna Siahaan and Dr. Pauletter Spencer, for their guidance and assistance. I thank the KU Department of Pharmaceutical Chemistry and School of Pharmacy for their support.

I thank the Madison and Lila Self Graduate Fellowship fellows and staff for their support and encouragement. Lastly, I thank my family and friends for their love and support. This work was made possible by NIH Grant Number P20 RR-17708 from the National Center for Research Resources and the Kansas University Center for Research. Additional support was provided by the Takeru Higuchi Fellowship and the Madison and Lila Self Graduate Fellowship for N. Ciaccio, the Initiative for Maximizing Student Diversity (IMSD) at the University of Kansas (NIGMS, MORE, NIH R25 GM62232) for M. Moreno and the KU Undergraduate Research Assistantship Fund for R. Bauer and T.S. Reynolds.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBEVIATIONS

AMP	adenine monophosphate
AUC	analytical ultracentrifugation
ApoA-I	Apolipoprotein A-I
ATF4	Activating Transcription Factor 4
ATF5	Activating Transcription Factor 5
Asn	asparagines
bZIP	basic leucine zipper
CD	circular dichroism
C/EBPβ	CCAAT Enhancer Binding Protein-β
Cys	cysteine
CRE	Cyclic AMP Response Element
DLS	dynamic light scattering
DNA	deoxyribonucleic acid
dsCRE	double-stranded CRE
DTT	dithiothreitol
ESI	electrospray ionization
FTIR	Fourier transform infrared spectroscoy

GCN4	General Control Protein GCN4
Gln	glutamine
GSH	glutathione
GST	glutathione s-transferase
HCl	hydrochloric acid
HPLC	high-performance liquid chromatography
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl β-D-1 thiogalactopyranoside
LB	Luria-Bertani broth
LC/MS	liquid chromatography/mass spectrometry
Leu	leucine
NMR	nuclear magnetic resonance spectroscopy
MW	molecular weight
MS	mass spectrometry
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PRL-1	Phosphatase of Regenerating Liver-1
R_H	hydrodynamic radius

SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
SLS	static light scattering
TFE	trifluoroethanol
T_m	melting temperature
Val	valine

CHAPTER 1

INTRODUCTION

1.1 PROTEIN AGGREGATION

Achieving and maintaining product stability is a primary objective of the biopharmaceutical industry.[1] Protein drugs are labile compounds in comparison to traditional, small molecule therapeutics. Whereas the activity of small organic compounds is dependent solely on their chemical structure, proteins possess sophisticated higher order structure that must be maintained for activity. Physical degradation of the native protein conformation can result in a loss of drug potency. Furthermore, physical degradation often leads to aggregation or non-native self-association of protein molecules in solution. Aggregation is a particular concern for drug products because it increases the likelihood that a patient will develop an undesired immune response.[2]

While preventing or at least controlling aggregate formation in pharmaceuticals is critically important, the mechanisms of assembly are not well understood. Increasing evidence suggests that the structural features of partially folded intermediates play a key role in influencing aggregate assembly.[3] The majority of structural information available regarding aggregate assembly comes from studies examining fibrillar protein aggregates. This type of protein aggregate has been heavily researched because of its involvement in numerous different disease states including Alzheimer's disease (AD) and Parkinsons's disease (PD).[4] Fibrillar

aggregates, also called amlyoids, are distinguished by a characteristic cross -beta-sheet structure where the beta-sheet strands align in a highly organized format, perpendicular to the axis of the fibril.[4a-c, 5] While the mechanism for formation of these beta-rich fibrils is not understood, several studies have reported the identification of a helical protein intermediate.[6] These data suggest that a helix-to-beta structural transition might be important in facilitating fibril formation.

Aggregate species that do not display long-range order have been termed amorphous protein aggregates. Even though these aggregates lack long-range order they do possess structure and typically display a high beta-sheet content.[7] The propensity of a protein system to form either amorphous or fibrillar protein aggregates *in vitro* is dependent not only upon the protein sequence and fold but on solution conditions.[7a, b, 8] A number of proteins have been shown to selectively develop into either amyloid fibrils or amorphous aggregates depending on these conditions.[8a, 9] A critical determinant in this selection is the nature of the unfolded or partially unfolded state.[8a, 9] The structural features of the intermediate species present in pathways of amorphous aggregation are largely unknown, however it is clear that they play an important role in driving aggregate assembly.[8a, 9]

An important consideration when evaluating mechanisms of physical degradation is the relationship between protein chemical and physical structure. Numerous mechanisms for protein chemical degradation exist including deamidation, oxidation and disulfide scrambling.[1a, b, 10] Any chemical modification to the protein molecule can affect the stability of the native structure or fold, possibly resulting in

physical degradation. These affects are quite protein-specific and heavily dependent upon the type and location of the chemical modification.

1.2 ACTIVATING TRANSCRIPTION FACTOR 5 (ATF5)

Activating Transcription Factor 5 (ATF5) is a protein recognized for its importance in neural development and its contribution to brain cancer.[11] The protein has been shown to maintain neural cells in a proliferative state and must be down regulated for differentiation to occur.[11a, b] ATF5 expression is not detected in mature neurons, but is present at high levels in developing neural tissue.[11a, b] Elevated levels of ATF5 have been observed in human brain cancer, specifically glioblastoma.[11c, d] Additionally, increased ATF5 expression is also detected in a number of human and rat glioma cell lines.[11c] The absence of detectable ATF5 in mature neurons, but elevated levels in cancerous tissue suggests that ATF5 may prove an ideal anticancer drug target. An *in vivo* rat model demonstrated that interference with ATF5 expression caused glioma cells to undergo apopotosis without affecting normal brain tissue.[11c] These data support the ideal that ATF5 is a promising target for anticancer drug treatment.

ATF5 was previously identified as a binding partner of a protein tyrosine phosphatase named Phosphatase of Regenerating Liver-1 (PRL-1).[12] PRL-1 is the first isoform of the PRL family, which consists of three enzymes originally identified as up regulated during liver regeneration.[13 14] High levels of PRL-1 expression are not just seen in regenerating liver but in many other tissues as well, including neural

tissue.[15] The interaction between PRL-1 and ATF5 has been mapped to the protein tyrosine phosphatase or active site domain of PRL-1 and the basic leucine zipper (bZIP) domain of ATF5.[12] The biological significance of this interaction is unknown, however, increased levels of PRL-1 expression have been shown to induce cellular proliferation and tumor formation in several different systems.[13-14]

ATF5 belongs to the bZIP family of transcription factors, which are so named because all members possess a characteristic bZIP motif (Figure 1.1). This motif consists of a basic region required for DNA binding followed by a leucine zipper region that facilitates protein dimerization.[16] All members of the bZIP protein family function as dimers to bind DNA and regulate the process of transcription. These proteins either homo- or hetero-dimerize selectively via stabilizing interactions between their leucine zipper domains resulting in the formation of a two-stranded coiled-coil.[16] It is this dimerization between transcription factors that provides functional diversity and variable selectivity for DNA binding to the bZIP family.[17]

The leucine zipper motif consists of a repeating heptad unit (*a*-*g*), where the *d*-position is almost exclusively occupied by leucine (Figure 1.2).[18] The *a*-position is typically occupied by a hydrophobic or uncharged amino acid, while the *e*- and *g*-positions are usually occupied by charged amino acids. The *a*- and *d*-positions lie on one side of the alpha-helix and together comprise the hydrophobic interface of the coiled-coil. The *e*- and *g*-positions are adjacent to the hydrophobic interface and facilitate electrostatic interactions that further stabilize coiled-coil formation. A unique feature of the ATF5 protein is that three valine residues rather than leucine

occupy the *d*-positions at the C-terminal end of the leucine zipper doman.[12, 19] The biological significance of this substitution is unknown, but it likely affects the selectivity, affinity and structure of ATF5 dimerization.

Currently no high-resolution structural data is available for ATF5. A crystal structure has been reported for Activating Transcription Factor 4 (ATF4), a close homolog of ATF5 (Figure 1.3).[20] ATF4 does not possess the extensive valine repeat present in ATF5, but the two proteins do share 74% sequence similarity. ATF4 was reported to form a hetero-dimer with another bZIP protein, CCAAT Enhancer Binding Protein-β (CEBP-β). In this structure ATF4 behaves as a rigid rod, while CEBP-β bends to wrap around ATF4 and form the coiled-coil.[20] This behavior is unusual for a bZIP dimer, which typically consists of two monomers that both coil to wrap around each other. While some evidence for ATF4 homo-dimerization exists, it seems unlikely for dimerization to be facilitated in the typical manner, suggesting that an alternative mechanism might be employed.[20] Both ATF4 and ATF5 contain a cysteine residue in the *a*-position of the second heptad repeat of the zipper domain. While a number of other bZIP proteins possess cysteine residues within their zipper domain, only ATF4 and ATF5 possess a cysteine in this position. It has previously been hypothesized that this cysteine might play a role in facilitating ATF4 dimerization via disulfide bond formation, since it is positioned at the coiled-coil interface (Figure 1.4).[20] The function of other transcription factors, including the Fos/Jun bZIP heterodimer, is regulated by disulfide bond formation.[21] It is not known what effect disulfide bond formation has on ATF5 structure and function.

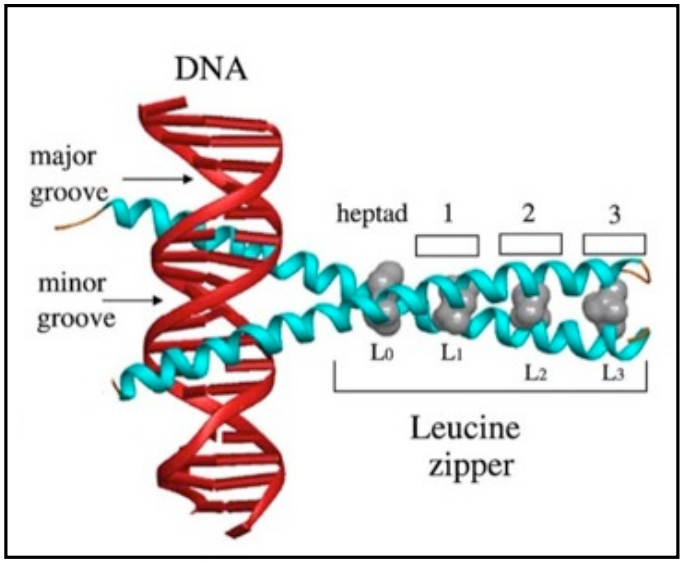

Figure 1.1 X-ray crystal structure of the GCN4 bZIP homodimer bound to DNA.[19a] DNA is shown in red. Each bZIP monomer is shown in blue. The leucine zipper domain comprised of the heptad repeat is depicted with the leucine residues shown in gray.

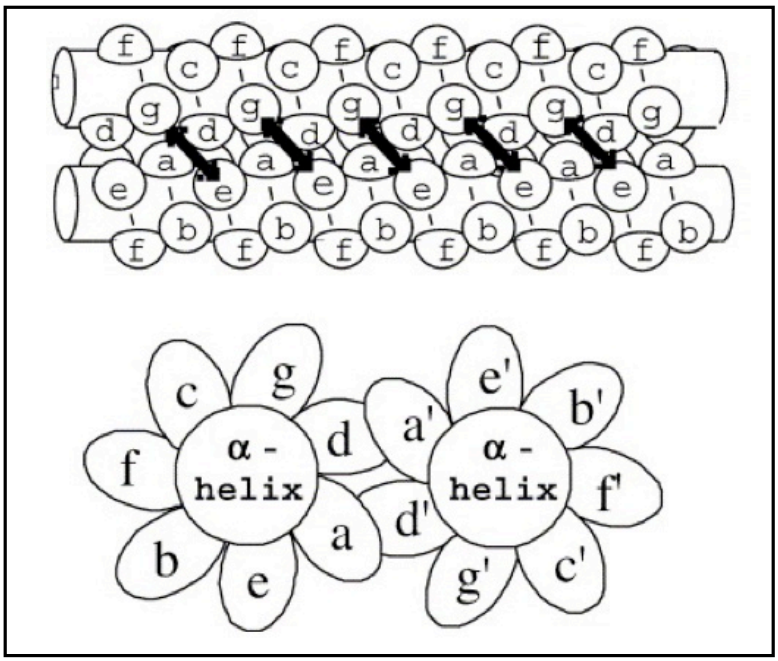

Figure 1.2 A side and end-view schematic illustrating the packing arrangement of a two-stranded coiled-coil.[19a] The heptad repeat (a-d) is shown with arrows highlighting the electrostatic interactions between the *e* and *g* positions.

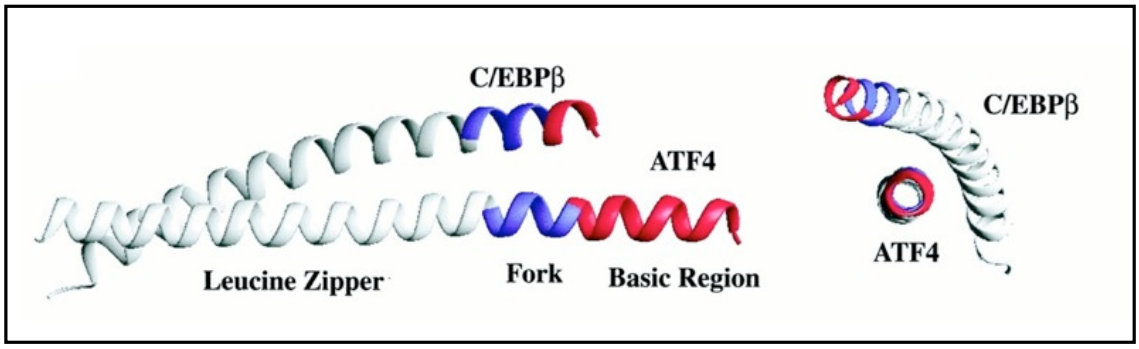

Figure 1.3 X-ray crystal structure of the ATF4-C/EBPβ heterodimer.[20] The basic region of each protein is shown in red. The fork region, which is a flexible region that connects the basic and leucine zipper domains, is shown in blue. The leucine zipper domain is shown in white.

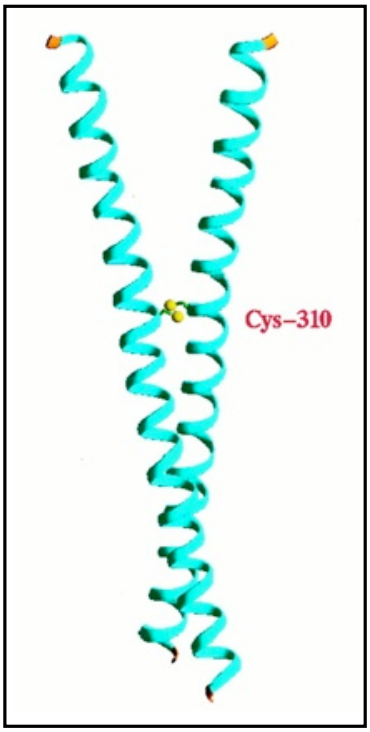

Figure 1.4 Theoretical structure of an ATF4 homodimer containing an intermolecular disulfide via C310.[20]

1.3 ATF5 AS A MODEL SYSTEM OF PROTEIN AGGREGATION

Typically, model systems used for studying protein aggregation consist of amyloid-forming proteins or peptides. The amorphous systems that have been investigated consist of molecules possessing sophisticated tertiary structure, often containing more than one intramolecular disulfide bond.[22] This complexity makes a mechanistic interpretation of amorphous aggregation more difficult. The bZIP domain of ATF5 possesses several characteristics that make it an interesting and unique model system for investigating amorphous protein aggregation. The protein is helical in solution, but only partially folded and highly dynamic.[23] This domain possesses only a single centrally-located cysteine residue that readily forms an intermolecular disulfide bond in solution and can be modulated by changing solution conditions.[23] One side of the ATF5 helix consists of a hydrophobic interface that contains a leucine repeat following by a valine repeat. An advantage of using ATF5 as a model is the simplicity of its sequence and structure. Single features affecting aggregation, such as helical content and disulfide cross-linking, can be modulated and evaluated individually. Additionally, the effects of modification at the hydrophobic interface of the leucine zipper region can be investigated.

The analysis presented herein consists of the application of site-directed mutagenesis and the alteration of solution conditions to modulate ATF5 sequence and structure and evaluate the impact on protein aggregation. A variety of complementary biophysical techniques were utilized to characterize the structural features of ATF5 and monitor the assembly of protein aggregates. The effects of intermolecular

disulfide bond formation and protein helicity on ATF5 structure and stability are presented in Chapter 3. The effects of removal of the valine zipper region on ATF5 structure and stability are presented in Chapter 4. This investigation comprises a unique contribution to the study of protein aggregation by furthering our understanding of the relationship between structure and amorphous aggregate assembly using a novel protein system.

1.4 REFERENCES

1. (a) Manning, M. C.; Chou, D. K.; Murphy, B. M.; Payne, R. W.; Katayama, D. S., Stability of Protein Pharmaceuticals: An Update. *Pharm Res*; (b) Manning, M. C.; Patel, K.; Borchardt, R. T., Stability of protein pharmaceuticals. *Pharm Res* **1989,** *6* (11), 903-18; (c) Volkin, D. B.; Sanyal, G.; Burke, C. J.; Middaugh, C. R., Preformulation studies as an essential guide to formulation development and manufacture of protein pharmaceuticals. *Pharm Biotechnol* **2002,** *14*, 1-46.
2. (a) Fradkin, A. H.; Carpenter, J. F.; Randolph, T. W., Immunogenicity of aggregates of recombinant human growth hormone in mouse models. *J Pharm Sci* **2009,** *98* (9), 3247-64; (b) Hermeling, S.; Crommelin, D. J.; Schellekens, H.; Jiskoot, W., Structure-immunogenicity relationships of therapeutic proteins. *Pharm Res* **2004,** *21* (6), 897-903; (c) Hermeling, S.; Schellekens, H.; Maas, C.; Gebbink, M. F.; Crommelin, D. J.; Jiskoot, W., Antibody response to aggregated human interferon alpha2b in wild-type and transgenic immune tolerant mice depends on type and level of aggregation. *J Pharm Sci* **2006,** *95* (5), 1084-96; (d) Rosenberg, A. S., Effects of protein aggregates: an immunologic perspective. *AAPS J* **2006,** *8* (3), E501-7.
3. (a) Bauer, R.; Carrotta, R.; Rischel, C.; Ogendal, L., Characterization and isolation of intermediates in beta-lactoglobulin heat aggregation at high pH. *Biophys J* **2000,** *79* (2), 1030-8; (b) Gomez-Orellana, I.; Variano, B.; Miura-Fraboni, J.; Milstein, S.; Paton, D. R., Thermodynamic characterization of an intermediate state of human growth hormone. *Protein Sci* **1998,** *7* (6), 1352-8; (c) Munishkina, L. A.; Fink, A. L.; Uversky, V. N., Accelerated fibrillation of alpha-synuclein induced by the combined action of macromolecular crowding and factors inducing partial folding. *Curr Alzheimer Res* **2009,** *6* (3), 252-60; (d) Frare, E.; Mossuto, M. F.; de Laureto, P. P.; Tolin, S.; Menzer, L.; Dumoulin, M.; Dobson, C. M.; Fontana, A., Characterization of oligomeric species on the aggregation pathway of human lysozyme. *J Mol Biol* **2009,** *387* (1), 17-27; (e) Jahn, T. R.; Parker, M. J.; Homans, S. W.; Radford, S. E., Amyloid formation under physiological conditions proceeds via a native-like folding intermediate. *Nat Struct Mol Biol* **2006,** *13* (3), 195-201; (f) Mach, H.; Ryan, J. A.; Burke, C. J.; Volkin, D. B.; Middaugh, C. R., Partially structured self-associating states of acidic fibroblast growth factor. *Biochemistry* **1993,** *32* (30), 7703-11.
4. (a) Maji, S. K.; Wang, L.; Greenwald, J.; Riek, R., Structure-activity relationship of amyloid fibrils. *FEBS Lett* **2009,** *583* (16), 2610-7; (b) Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Sambashivan, S.; Ivanova, M. I.; Madsen, A. O.; Riekel, C., The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Acc Chem Res* **2006,** *39* (9), 568-75; (c) Bellotti, V.; Nuvolone, M.; Giorgetti, S.; Obici, L.; Palladini, G.; Russo, P.; Lavatelli, F.; Perfetti, V.; Merlini, G., The workings of the amyloid diseases. *Ann Med* **2007,** *39* (3), 200-7; (d) Stefani, M.; Dobson, C. M., Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* **2003,** *81* (11), 678-99; (e) Fandrich, M.; Meinhardt, J.;

Grigorieff, N., Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. *Prion* **2009,** *3* (2), 89-93.
5. Squires, A. M.; Devlin, G. L.; Gras, S. L.; Tickler, A. K.; MacPhee, C. E.; Dobson, C. M., X-ray scattering study of the effect of hydration on the cross-beta structure of amyloid fibrils. *J Am Chem Soc* **2006,** *128* (36), 11738-9.
6. (a) Knowles, T. P.; Zahn, R., Enhanced stability of human prion proteins with two disulfide bridges. *Biophys J* **2006,** *91* (4), 1494-500; (b) Abedini, A.; Raleigh, D. P., A role for helical intermediates in amyloid formation by natively unfolded polypeptides? *Phys Biol* **2009,** *6* (1), 15005; (c) Harada, A.; Azakami, H.; Kato, A., Amyloid fibril formation of hen lysozyme depends on the instability of the C-helix (88-99). *Biosci Biotechnol Biochem* **2008,** *72* (6), 1523-30; (d) Mihara, H.; Takahashi, Y., Engineering peptides and proteins that undergo alpha-to-beta transitions. *Curr Opin Struct Biol* **1997,** *7* (4), 501-8; (e) Olofsson, A.; Borowik, T.; Grobner, G.; Sauer-Eriksson, A. E., Negatively charged phospholipid membranes induce amyloid formation of medin via an alpha-helical intermediate. *J Mol Biol* **2007,** *374* (1), 186-94; (f) Smirnovas, V.; Winter, R.; Funck, T.; Dzwolak, W., Thermodynamic properties underlying the alpha-helix-to-beta-sheet transition, aggregation, and amyloidogenesis of polylysine as probed by calorimetry, densimetry, and ultrasound velocimetry. *J Phys Chem B* **2005,** *109* (41), 19043-5; (g) Williamson, J. A.; Miranker, A. D., Direct detection of transient alpha-helical states in islet amyloid polypeptide. *Protein Sci* **2007,** *16* (1), 110-7; (h) Kunjithapatham, R.; Oliva, F. Y.; Doshi, U.; Perez, M.; Avila, J.; Munoz, V., Role for the alpha-helix in aberrant protein aggregation. *Biochemistry* **2005,** *44* (1), 149-56; (i) Thompson, A. J.; Barnham, K. J.; Norton, R. S.; Barrow, C. J., The Val-210-Ile pathogenic Creutzfeldt-Jakob disease mutation increases both the helical and aggregation propensities of a sequence corresponding to helix-3 of PrP(C). *Biochim Biophys Acta* **2001,** *1544* (1-2), 242-54.
7. (a) Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F., Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* **2003,** *20* (9), 1325-36; (b) Weiss, W. F. t.; Young, T. M.; Roberts, C. J., Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. *J Pharm Sci* **2009,** *98* (4), 1246-77; (c) Wang, L.; Maji, S. K.; Sawaya, M. R.; Eisenberg, D.; Riek, R., Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol* **2008,** *6* (8), e195.
8. (a) Lee, C. F., Self-assembly of protein amyloids: a competition between amorphous and ordered aggregation. *Phys Rev E Stat Nonlin Soft Matter Phys* **2009,** *80* (3 Pt 1), 031922; (b) Mahler, H. C.; Friess, W.; Grauschopf, U.; Kiese, S., Protein aggregation: pathways, induction factors and analysis. *J Pharm Sci* **2009,** *98* (9), 2909-34.
9. (a) Demeule, B.; Gurny, R.; Arvinte, T., Where disease pathogenesis meets protein formulation: renal deposition of immunoglobulin aggregates. *Eur J Pharm Biopharm* **2006,** *62* (2), 121-30; (b) Grudzielanek, S.; Jansen, R.; Winter, R., Solvational tuning of the unfolding, aggregation and amyloidogenesis of insulin. *J Mol Biol* **2005,** *351* (4), 879-94; (c) Kim, Y. S.; Randolph, T. W.; Stevens, F. J.;

Carpenter, J. F., Kinetics and energetics of assembly, nucleation, and growth of aggregates and fibrils for an amyloidogenic protein. Insights into transition states from pressure, temperature, and co-solute studies. *J Biol Chem* **2002,** *277* (30), 27240-6.
10. Volkin, D. B.; Mach, H.; Middaugh, C. R., Degradative covalent reactions important to protein stability. *Mol Biotechnol* **1997,** *8* (2), 105-22.
11. (a) Mason, J. L.; Angelastro, J. M.; Ignatova, T. N.; Kukekov, V. G.; Lin, G.; Greene, L. A.; Goldman, J. E., ATF5 regulates the proliferation and differentiation of oligodendrocytes. *Mol Cell Neurosci* **2005,** *29* (3), 372-80; (b) Angelastro, J. M.; Mason, J. L.; Ignatova, T. N.; Kukekov, V. G.; Stengren, G. B.; Goldman, J. E.; Greene, L. A., Downregulation of activating transcription factor 5 is required for differentiation of neural progenitor cells into astrocytes. *J Neurosci* **2005,** *25* (15), 3889-99; (c) Angelastro, J. M.; Canoll, P. D.; Kuo, J.; Weicker, M.; Costa, A.; Bruce, J. N.; Greene, L. A., Selective destruction of glioblastoma cells by interference with the activity or expression of ATF5. *Oncogene* **2006,** *25* (6), 907-16; (d) Monaco, S. E.; Angelastro, J. M.; Szabolcs, M.; Greene, L. A., The transcription factor ATF5 is widely expressed in carcinomas, and interference with its function selectively kills neoplastic, but not nontransformed, breast cell lines. *Int J Cancer* **2007,** *120* (9), 1883-90; (e) Greene, L. A.; Lee, H. Y.; Angelastro, J. M., The transcription factor ATF5: role in neurodevelopment and neural tumors. *J Neurochem* **2009,** *108* (1), 11-22.
12. Peters, C. S.; Liang, X.; Li, S.; Kannan, S.; Peng, Y.; Taub, R.; Diamond, R. H., ATF-7, a novel bZIP protein, interacts with the PRL-1 protein-tyrosine phosphatase. *J Biol Chem* **2001,** *276* (17), 13718-26.
13. (a) Stephens, B. J.; Han, H.; Gokhale, V.; Von Hoff, D. D., PRL phosphatases as potential molecular targets in cancer. *Mol Cancer Ther* **2005,** *4* (11), 1653-61; (b) Bessette, D. C.; Qiu, D.; Pallen, C. J., PRL PTPs: mediators and markers of cancer progression. *Cancer Metastasis Rev* **2008,** *27* (2), 231-52.
14. Sun, J. P.; Wang, W. Q.; Yang, H.; Liu, S.; Liang, F.; Fedorov, A. A.; Almo, S. C.; Zhang, Z. Y., Structure and biochemical properties of PRL-1, a phosphatase implicated in cell growth, differentiation, and tumor invasion. *Biochemistry* **2005,** *44* (36), 12009-21.
15. Takano, S.; Fukuyama, H.; Fukumoto, M.; Kimura, J.; Xue, J. H.; Ohashi, H.; Fujita, J., PRL-1, a protein tyrosine phosphatase, is expressed in neurons and oligodendrocytes in the brain and induced in the cerebral cortex following transient forebrain ischemia. *Brain Res Mol Brain Res* **1996,** *40* (1), 105-15.
16. (a) Landschulz, W. H.; Johnson, P. F.; McKnight, S. L., The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* **1988,** *240* (4860), 1759-64; (b) O'Shea, E. K.; Rutkowski, R.; Kim, P. S., Evidence that the leucine zipper is a coiled coil. *Science* **1989,** *243* (4890), 538-42; (c) Vinson, C. R.; Sigler, P. B.; McKnight, S. L., Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **1989,** *246* (4932), 911-6.

17. Kouzarides, T.; Ziff, E., Leucine zippers of fos, jun and GCN4 dictate dimerization specificity and thereby control DNA binding. *Nature* **1989,** *340* (6234), 568-71.
18. (a) Crick, F. H. C., The packing of [alpha]-helices: simple coiled-coils; (b) Gruber, M.; Lupas, A. N., Historical review: another 50th anniversary--new periodicities in coiled coils. *Trends Biochem Sci* **2003,** *28* (12), 679-85; (c) Lupas, A. N.; Gruber, M., The structure of alpha-helical coiled coils. *Adv Protein Chem* **2005,** *70*, 37-78.
19. (a) Vinson, C.; Acharya, A.; Taparowsky, E. J., Deciphering B-ZIP transcription factor interactions in vitro and in vivo. *Biochim Biophys Acta* **2006,** *1759* (1-2), 4-12; (b) Ciaccio, N. A.; Moreno, M. L.; Bauer, R. L.; Laurence, J. S., High-yield expression in E. coli and refolding of the bZIP domain of activating transcription factor 5. *Protein Expr Purif* **2008,** *62* (2), 235-43.
20. Podust, L. M.; Krezel, A. M.; Kim, Y., Crystal structure of the CCAAT box/enhancer-binding protein beta activating transcription factor-4 basic leucine zipper heterodimer in the absence of DNA. *J Biol Chem* **2001,** *276* (1), 505-13.
21. Abate, C.; Patel, L.; Rauscher, F. J., 3rd; Curran, T., Redox regulation of fos and jun DNA-binding activity in vitro. *Science* **1990,** *249* (4973), 1157-61.
22. (a) Alford, J. R.; Kwok, S. C.; Roberts, J. N.; Wuttke, D. S.; Kendrick, B. S.; Carpenter, J. F.; Randolph, T. W., High concentration formulations of recombinant human interleukin-1 receptor antagonist: I. Physical characterization. *J Pharm Sci* **2008,** *97* (8), 3035-50; (b) Lin, J. J.; Meyer, J. D.; Carpenter, J. F.; Manning, M. C., Aggregation of human serum albumin during a thermal viral inactivation step. *Int J Biol Macromol* **2009,** *45* (2), 91-6; (c) Roy, S.; Mason, B. D.; Schoneich, C. S.; Carpenter, J. F.; Boone, T. C.; Kerwin, B. A., Light-induced aggregation of type I soluble tumor necrosis factor receptor. *J Pharm Sci* **2009,** *98* (9), 3182-99; (d) Salinas, B. A.; Sathish, H. A.; Bishop, S. M.; Harn, N.; Carpenter, J. F.; Randolph, T. W., Understanding and modulating opalescence and viscosity in a monoclonal antibody formulation. *J Pharm Sci 99* (1), 82-93; (e) Thirumangalathu, R.; Krishnan, S.; Brems, D. N.; Randolph, T. W.; Carpenter, J. F., Effects of pH, temperature, and sucrose on benzyl alcohol-induced aggregation of recombinant human granulocyte colony stimulating factor. *J Pharm Sci* **2006,** *95* (7), 1480-97.
23. Ciaccio, N. A.; Laurence, J. S., Effects of disulfide bond formation and protein helicity on the aggregation of activating transcription factor 5. *Mol Pharm* **2009,** *6* (4), 1205-15.

CHAPTER 2

HIGH-YIELD EXPRESSION IN E. COLI AND REFOLDING OF THE BZIP DOMAIN OF ATF5

2.1 INTRODUCTION

Activating Transcription Factor 5 (ATF5) recently has been recognized for its importance in neurological development and contribution to brain cancer. This protein functions to maintain the cell in a proliferative state and must be down regulated in order for differentiation to occur in neural progenitor cells [1]. While ATF5 is highly expressed in developing neurons, it is not expressed to any detectable level in healthy mature neural tissue [1a, 1c]. Increased levels of ATF5 have, however, been observed in primary brain tumors, and expression is elevated to particularly high levels in human glioblastoma [2]. It also is overproduced in several human and rat glioma cell lines [2]. The absence of ATF5 expression in mature neurons and its prominence in brain tumors has made it an appealing target for anti-cancer therapy. Importantly, an *in vivo* rat model has demonstrated that interference with ATF5 function caused glioma cell death in primary tumors, while it did not affect the status of normal cells surrounding the tumor [2]. This data suggests ATF5 is a prime target for pharmaceutical intervention.

In general, there are two main approaches used to identify and develop lead pharmaceutical candidates. The first relies on empirical testing of chemical compounds and requires a screening assay that is capable of quantifying the effect

compounds have on a given pharmaceutical target. The second entails rational drug design, which is based upon high-resolution structural information about a pharmaceutical target. The first step that must be taken to begin the analysis using either approach is production of the pharmaceutical protein target. In many cases involving multi-domain proteins, it is acceptable or advantageous to simplify the analysis by focusing the characterization specifically on the domain(s) that confer activity. In the case of ATF5, the portion responsible for DNA binding is of interest for assay development and structural characterization. ATF5 belongs to the bZIP family of transcriptions factors, which are composed of an unstructured N-terminal activating domain followed by a C-terminal bZIP domain [3]. The bZIP domain consists of a basic sequence that directly binds the DNA. The DNA binding sequence is followed by a helical leucine zipper region, which facilitates dimerization, increasing the DNA binding affinity [4]. The large, unstructured N-terminal domain typically is involved in more complex aspects of transcriptional regulation [3, 5]. The bZIP domain from numerous transcription factors has been shown to be amenable to structural characterization, and there are many instances where study of a bZIP protein has been performed using only this C-terminal domain [4].

The goal of the research described here was to produce the bZIP domain of ATF5 in sufficient yield and quantity to make high-resolution structure analysis and assay development feasible. We have developed a process to generate and maintain a soluble and stable solution of protein at sufficiently high concentration to permit such analyses. This paper describes the procedure whereby the bZIP domain of ATF5 was

successfully expressed, purified and concentrated sufficiently to perform two-dimensional NMR analysis. SDS-PAGE and mass spectrometry were used to confirm that we successfully produced and isolated the intact bZIP domain of ATF5. The CD data indicate that ATF5 contains alpha helical content, showing that the protein obtained using our method has the expected secondary structure. 2D ^{1}H-^{15}N NMR data was acquired and the spectrum shows the protein is predominantly in the monomeric x-form state, indicating that structural analysis can be performed on ATF5 that has been produced using the method presented here [6]. The functionality of ATF5 prepared under these conditions was confirmed using an electromobililty shift assay where the protein specifically bound to the cyclic AMP response element (CRE), which is a known DNA binding site of ATF5 [7].

2.2 MATERIALS AND METHODS

2.2.1 Cloning and Construction of the Expression Plasmid

The cDNA of ATF5 was obtained through ATCC (**MGC-842**). The portion encoding the bZIP domain was amplified using PCR for insertion to a plasmid using the following primers: 5'GCGCGCCCATGGGCCCTGCCACCACCCGA3' (forward primer with NcoI restriction site), 5'GCGCGCCATATGCCTGCCACCACCCGAGGG3' (forward primer with NdeI restriction site), 5'CGCGCGGGATCCTCAGCTACGGGTCCTCTG3' (reverse primer with BamHI restriction site). The amplified fragments were digested with the corresponding

endonucleases, gel purified and ligated into the pET-42b vector (Novagen) by overnight incubation with T4 Ligase at 16°C. The NcoI site was used to insert the ATF5 gene following a Glutathione-S-Transferase (GST) tag, whereas insertion at the NdeI site generated a construct from which untagged ATF5 could be expressed. The ligation product was transformed into competent Novablue *E. coli* (Novagen) using the standard heat shock protocol. Transformed colonies were grown overnight at 37°C on LB agar plates containing 30 μg/ml kanamycin. Individual colonies were selected and used to inoculate 5 ml M9ZB containing 0.4% glucose, 1 mM $MgSO_4$ and 30 μg/ml kanamycin. Cultures were incubated with shaking overnight at 37°C. Plasmid DNA was isolated from the overnight cultures using a Qiagen miniprep kit. Purified DNA was verified to be correct by bidirectional sequencing (Northwoods DNA).

2.2.2 Protein expression

BL21(DE3) *E. coli* (Novagen) were transformed with the pET-42b-ATF5 bZIP plasmid using the standard heat shock protocol. Transformed colonies were selected and grown overnight at 37°C on LB agar plates containing 30 μg/mL kanamycin. Colonies were selected and used to inoculate 5 mL LB containing 0.4% glucose, 1 mM $MgSO_4$, and 30 μg/mL kanamycin. Cultures were grown with shaking overnight at 37°C. Two starter cultures were used to inoculate one 500 ml culture containing minimal media consisting of deionized water supplemented with 0.1% ammonium chloride, 1% glucose, 10 mM MgCl, 40 μg/mL thiamine HCl, 30 μg/mL kanamycin plus additional salts (100 mM KH_2PO_4, 57 mM K_2HPO_4, 63 mM

Na_2HPO_4, 14 mM K_2SO_4) and trace minerals (200 μM $CaCl_2 \bullet H_2O$, 100 μM $FeSO_4 \bullet H_2O$, 50 μM $MnCl_2 \bullet 6H_2O$, 15 μM $CoCl_2 \bullet 6H_20$, 10 μM $ZnSO_4 \bullet 7H_2O$, 9 μM $CuCl_2 \bullet 2H_2O$, 1.5 μM H_3BO_4, 1.2 μM $(NH_4)Mo_7O_{24} \bullet 4H_2O$, 65 μM Na_2EDTA). Isotopic labeling was accomplished via substitution with ^{15}N-labeled ammonium chloride (Spectra Isotopes). Cultures were grown to an OD_{550} of 0.6 to 0.8 before inducing protein expression with a final concentration of 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Cells were incubated with shaking at 37°C for 4 additional hours post induction, harvested by centrifugation at 3000 x g and stored at -80°C.

2.2.3 Purification of the GST-tagged ATF5 construct

Cell pellets were thawed on ice and resuspended in cold phosphate buffered saline (PBS), pH 7.3 (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) containing 5 mM dithiothreitol (DTT). Each pellet derived from 1 L culture was taken up in 25 mL solution and lysed via passage through a french pressure cell. Lysates were centrifuged for 1 hour at 21,000 x g and 4°C. The supernatant was decanted and passed through a 0.2 μm filter. The filtrate was loaded onto a 5 mL GSTrap column (GE Healthcare) at a flow rate of 1 mL/min. The column was rinsed with 10 column volumes (CV) of loading buffer or PBS, pH 7.3 containing 5 mM DTT at 3 mL/min. Elution was performed with 3 CV of elution buffer or 50 mM Tris HCl, pH 8.0 containing 10 mM reduced glutathione (GSH) and 5 mM DTT at 3 mL/min. Chromatographic separation was carried out at 4-8°C. Fractions were collected and stored cold prior to analysis by SDS-PAGE.

Eluted fractions containing ATF5 were concentrated by ultrafiltration using a 3500 MWCO Millipore ultrafilter with a Biomax membrane and centrifuged at 4000 x g and 4°C. Subsequently, Factor Xa (Novagen) was added to each sample at a concentration of 10 units/mL and incubated overnight (approximately 16 hours) at room temperature to cleave the GST tag. Cleavage progress was monitored by SDS-PAGE. The GST tag was removed by incubating the cleaved protein with glutathione-agarose beads (BD Biosciences) for 30 minutes at room temperature. The sample was centrifuged at 16,000 x g for 10 minutes to pellet the agarose beads and recover the ATF5 protein.

2.2.4 Protein extraction and refolding

Cell pellets were thawed, resuspended in PBS, pH 7.3 containing 5 mM DTT and lysed as described above. The insoluble pellet was resuspended by vortexing it in 4 mL of 50 mM sodium phosphate, pH 6.5 containing 100 mM NaCl, 6 M guanidine HCl and 5 mM DTT. The solution was incubated at room temperature for 3 hours to permit denaturation to occur before separating the soluble and insoluble material by centrifugation for 1 hour at 21,000 x g and 4°C. The solubilized protein was refolded by dilution into 40 mL of 50 mM sodium phosphate, pH 6.5 containing 100 mM NaCl and 5 mM DTT [8]. Insoluble components were pelleted by centrifugation for 1 hour at 21,000 x g and 4°C. The supernatant containing refolded ATF5 was dialyzed using 1000 MWCO dialysis tubing with a cellulose membrane (Spectra) overnight at 4°C into 2 L of phosphate buffer (50 mM sodium phosphate, pH 6.5, 100 mM NaCl, 5mM DTT) to remove the guandinium. The ratio of protein to dialysis solution was

0.04 L protein: 2 L dialysis buffer, which results in a 50-fold dilution of guanidinium to a final concentration of 12 mM. Any precipitation generated during dialysis was removed by centrifugation for 1 hour at 21,000 x g and 4°C. The supernatant containing soluble protein was collected and concentrated using a 15 mL 3500 MWCO Millipore ultrafilter with a Biomax membrane and exchanged into the same buffer solution (50 mM sodium phosphate, pH 6.5, 100 mM NaCl) without DTT. The solution was concentrated to approximately 1 mL and then exchanged into 20-fold excess buffer to further remove guanidinium, reducing the final concentration to less than 1 mM. Sample purity was assessed by SDS-PAGE analysis.

2.2.5 Denaturing Gel Electrophoresis

SDS-PAGE was performed according to the Laemmli procedure [9]. GST-tagged ATF5 was separated on a discontinuous system consisting of a 5% (v/v) stacking gel and a 12% (v/v) resolving gel, whereas untagged ATF5 was resolved using a 15% (v/v) gel. Samples were diluted with 2x Laemmli buffer and heated at 90°C for 10 minutes before being loaded onto the gel. To examine the presence of disulfide-linked species a non-reducing loading buffer was used, which did not contain any reducing agent. An unstained molecular weight marker was used for reference (Bio-Rad Precision Plus standards). Coomassie staining was performed to visualize protein content.

Densitometry was perfomed on Coomassie stained SDS-PAGE gels using a Typhoon Trio Variabel Mode Imager (Amersham Biosciences) and the amount of protein in each band was determined using ImageQuant TL software (Amersham Biosciences).

2.2.6 Protein Concentration Determination

The molar concentration of the purified ATF5 solution was determined using the Beer-Lambert Law ($A=\varepsilon lc$). A theoretical molar absorptivity value (ε) corresponding to 280 nm was calculated to be 4470 $M^{-1}cm^{-1}$ using the Expasy Proteomics Server [10]. A_{280} absorption readings were measured in a 1 cm, quartz cuvette using a Cary 100 UV-Visible spectrophotometer.

2.2.7 Liquid Chromatography and Mass Spectrometry

Non-reduced samples of ATF5 were prepared at a concentration of 0.2 mM in 100 mM ammonium bicarbonate buffer, pH 6.0. In order to prepare reduced samples, 5 mM DTT was added to the non-reduced stock and the solution was heated at 37°C for 20 minutes prior to injection on the column. Microbore HPLC/MS experiments were performed to determine the mass of the ATF5 product and investigate the existence of intermolecular disulfide bonds. The HPLC step was used to desalt and separate proteins chromatographically (Waters Acquity). Solvents A and B contained 99% H_2O, 1% CH_3CN and 99% CH_3CN, 1% H_2O, respectively, and both contained 0.08% formic acid. Separations were performed on a 1 mm ID x 5 cm long C4 reverse phase column (Vydac C4, 300Å pore size, 3.5 μM particles packed by MicroTech Scientific, Vista, CA) at 110 μl/min. The gradient was held at 1% B for 3 min, then ramped to 23% B by 4 min, 45% by 14 min, and 90% B by 15 min.

Partial resolution of the ATF5 monomer from dimer in the non-reduced sample was observed near 28 % B with this gradient.

Electrospray ionization (ESI) spectra were acquired on a Q-Tof-2 (Micromass Ltd, Manchester UK) hybrid mass spectrometer operated in MS mode and acquiring data with the time of flight analyzer. The instrument was operated with analyzer settings optimized for maximum sensitivity while infusing a 50% organic/water solution of lysozyme at the flow rate of the chromatography. The cone and collision cell voltages (35 and 20V, respectively) were a compromise to maximize sensitivity and keep the M-H_2O/NH_3 ion below 20% of the primary ion in each charge state cluster. Argon was admitted to the collision cell using 16 psi on the supply regulator or 5.3 x 10^{-5} mBar on a penning gauge near the collision cell. Spectra were acquired at an 11,364 Hz pusher frequency, covering the mass range 800 to 3000 u and accumulating data for 5 seconds per cycle. Time to mass calibration was made with CsI cluster ions acquired under the same conditions. The resulting suite of charge states in the ESI spectrum were subject to charge state deconvolution to present a "zero" charge mass spectrum using the Transform or MaxEnt1 routine in MassLynx software.

2.2.8 NMR Spectroscopy

Two-dimensional ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra were recorded at 25°C using a Bruker AVANCE 800 MHz spectrometer equipped with a triple-resonance CRYO-probe with pulse field gradients. Samples were prepared at a concentration of 0.2 mM in 50 mM sodium phosphate, pH 6.5,

containing 100 mM NaCl, 5 mM DTT and 5% D_2O. Water suppression was accomplished using flip-back pulses. Data were acquired in 128 scans with 2048 points in ^{1}H and 128* increments in ^{15}N. ^{1}H chemical shifts were referenced with respect to an external DSS standard in D_2O [11]. Indirect referencing relative to ^{1}H was determined for ^{15}N, assuming a ratio of $^{15}N/^{1}H = 0.101329118$ [11]. Data were processed using NMRPipe and Sparky software [12].

2.2.9 Circular Dichroism Spectroscopy

Data were collected on a Jasco J-720 spectrapolarimeter equipped with a Peltier temperature control unit. Samples consisted of 20 μM protein in 50 mM sodium phosphate buffer, pH 6.5, and 100 mM NaCl. Spectra were acquired at 10°C using a quartz cuvette with a 1 mm path length. Scans were collected in the range of 190 to 260 nm, a data pitch of 1 nm, a band width of 5 nm, a scanning speed of 50 nm/min and a response time of 8 seconds. Sample spectra were corrected by subtraction of the buffer spectrum, and molar ellipticity was calculated using a molecular weight of 9.33 kDa.

2.2.10 Electromobility Shift Assay

The DNA binding activity of the bZIP domain of ATF5 was determined using a 20 base-pair, double-stranded oligonucleotide probe containing the cyclic AMP response element (CRE) binding motif (5′-TAAAAATGACGTCATGGTAA-3′). The forward primer was modified at its 5’ end with a 6FAM fluorescent tag for detection (IDT). The forward and reverse strands were preannealed to generate double-stranded DNA by heating at 90°C for 5 minutes before cooling to room temperature over the

next 11 minutes. A 2 μL volume of a 1 mM stock of purified ATF5 was combined with 2 μL of a 1 μM stock CRE DNA containing the fluorescent tag in 12 μL of binding buffer (10 mM Tris, pH 7.5, 50 mM NaCl, 1 mM EDTA 1 mM DTT, 5 mM $MgCl_2$, 10% (v/v) glycerol) and incubated at room temperature for 15 minutes. As a control, 100-fold excess of preannealed unlabeled CRE DNA was added to ATF5 as a competitive inhibitor in the above reaction. After the addition of loading buffer, 10 μL of each sample was loaded onto a 20% nondenaturing polyacrylamide gel. The gel was run at 100 V for one hour and the 6FAM-labeled DNA was detected using a Typhoon Trio Variable Mode Imager (Amersham Biosciences). Excitation of the fluorophore was done at a wavelength of 488 nm and emission was detected at 520 nm.

2.3 RESULTS

2.3.1 Expression and purification of GST-tagged ATF5

The initial attempt to obtain the bZIP domain of ATF5 involved production of it in a fusion construct with a N-terminal GST-tag, as previously reported [7]. SDS-PAGE analysis of pre- and post-induction samples demonstrated successful expression of this protein, as detected by the presence of a band at 38 kDa, which is the expected molecular weight for this construct (Figure 2.1A). Although the majority of recombinant protein was insoluble following cell lysis (Figure 1B, lane 9), we isolated the soluble fraction (lane 1) by centrifugation, which contained approximately 9 mg GST-ATF5. Affinity chromatography then was performed on

the soluble fraction using a GSTrap column. Isolation of the fusion construct from other soluble proteins was achieved using this method; however, extensive secondary proteolysis occurred during the purification. The initial cell lysate contained a 38 kDa band (Figure 2.1B, lane 1), which corresponds to the full-length fusion protein, whereas the eluate (lane 5) lacked this band and instead contained a band at approximately 35 kDa. Based on the change in molecular weight of these observed bands, proteolytic cleavage does not correspond to release of the 10 kDa ATF5 bZIP protein. Subsequent attempts to specifically cleave the GST tag from ATF5 using Factor Xa were unfruitful, as demonstrated by the lack of a band near the expected molecular weight for the full-length ATF5 bZIP domain or a partially cleaved 7 kDa fragment (Figure 2.1B, lane 7). Moreover, the intensity of the band at 35 kDa did not decrease to an appreciable extent following incubation with Factor Xa (Figure 2.1B, lane 7), suggesting that access to the cleavage motif may be blocked. After completing the chromatographic separation, the GSTrap resin was washed with 1% SDS to remove any residual proteins from the column. This fraction contained a large amount of GST-ATF5 (Figure 2.1B, lane 8), which suggests the protein had aggregated on the column or formed large oligomers that were not effectively competed off the resin during the elution step.

Most of the expressed fusion protein was located in the insoluble cellular fraction (Figure 2.1B, lane 9), suggesting that the ATF5 bZIP protein may be poorly soluble or prone to aggregation. Aggregate formation in solution also could limit access to the Factor Xa cleavage site, resulting in retention of the apparent higher

molecular weight bands. Because much of the GST-tagged ATF5 was insoluble following cell lysis and efforts to cleave the tag from the soluble protein were unsuccessful, an alternative approach was devised to obtain ATF5. Because GST (not fused to ATF5) is present in the soluble portion of the cell lysate, we expected that untagged ATF5 would appear in the pellet. Many proteins that are directed to inclusion bodies can be recovered from the insoluble pellet using a denaturation and refolding procedure [8]. We decided to investigate whether improved yields of the ATF5 bZIP protein could be achieved by purifying an untagged version from this insoluble fraction.

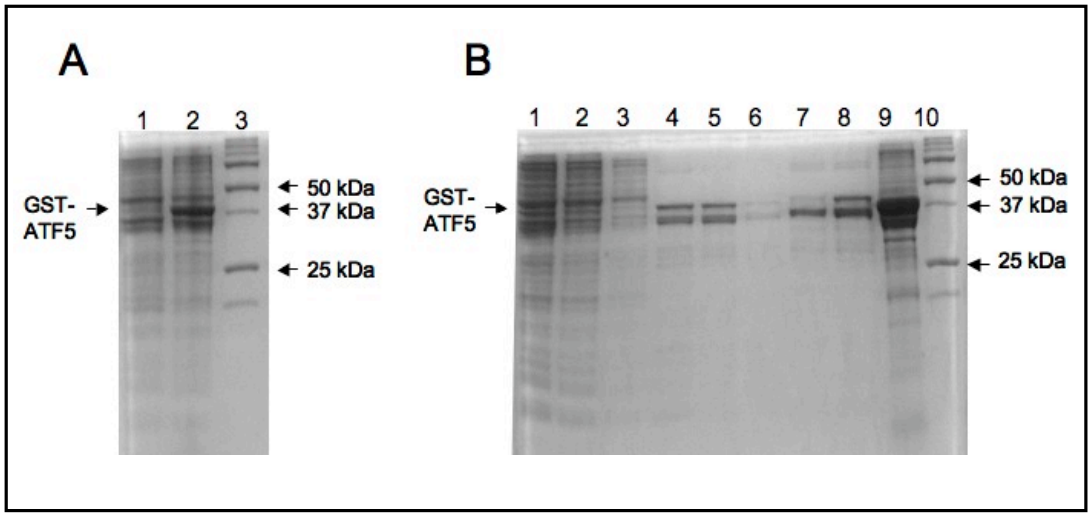

Figure 2.1 (a) Coomassie stained SDS-PAGE showing expression of GST-ATF5 bZIP. Lane 1: *E.Coli* lysate pre-IPTG induction, Lane 2: *E.Coli* lysate post-IPTG induction, Lane 3: molecular weight marker (Bio-Rad Precision Plus Unstained MW Standard). (b) Coomassie stained SDS-PAGE gel showing fractions from the purification of GST-ATF5 bZIP using GSTrap affinity chromatography. Lane 1: supernatant following cell lysis, Lane 2: column flow-through, Lane 3: rinse fraction, Lane 4: elution fraction #1, Lane 5: elution fraction #2, Lane 6: elution fraction #3, Lane 7: Eluate after dialysis, reaction with Factor Xa and incubation with GSTrap beads, Lane 8: GSTrap beads containing residual protein post GST removal, Fraction 9: insoluble pellet after cell lysis, Fraction 10: molecular weight marker (Bio-Rad Precision Plus Unstained MW Standard).

2.3.2 Expression and refolding of untagged ATF5

Analysis of the SDS-PAGE data revealed that a high level of expression of the untagged form of ATF5 was achieved from this construct (Figure 2.2A). Anomalous behavior was observed for this protein when evaluated using SDS-PAGE analysis. The protein runs more slowly than would be expected, likely due to its basicity and overall highly positive charge (pI=10.07). This behavior has been observed with ATF5 before and with other highly basic proteins [7, 13]. To confirm the band at approximately 12 kDa corresponds to ATF5, mass spectrometric data was collected (Figure 2.3). The MS spectrum shows a peak at 9333.7 u, which provides confirmation that the product obtained has the expected mass calculated for ATF5 bZIP (9333.4 Da).

The cells containing untagged ATF5 were lysed and centrifuged using the same conditions and procedure as described above. SDS-PAGE revealed that the majority of ATF5 protein is contained in the pellet following lysis. ATF5 was extracted efficiently from the insoluble pellet when incubated with a high concentration of the denaturant guanidine and the reducing agent DTT (data not shown). The protein refolded as the denatured solution was diluted drop wise into a large volume of non-denaturing buffer. Samples from each step in the purification process could not be analyzed by SDS-PAGE, because this method in incompatible with the use of guanidine HCl. There was some precipitation or turbidity observed during the refolding step, which could be due to the presence of lipids and/or membrane proteins that did not refold upon dilution. This insoluble material was

successfully pelleted by centrifugation following refolding, and the clarified supernatant was decanted and dialyzed. The dialysis step, which facilitated a 50-fold reduction in the guanidine concentration, resulted in some visible precipitation. The precipitate was removed by centrifugation prior to concentration of the final material. Initial attempts at refolding were conducted in Tris buffer at pH 7.5. Under these solutions conditions, the protein was found to be highly prone to aggregation as detected by visual observation of large amounts of precipitation. Performing the refolding step in phosphate buffer at pH 6.5, however, allowed for greater protein stability and increased solution concentrations to be achieved.

Using phosphate buffer, our scheme for obtaining untagged ATF5 from the insoluble cellular fraction yielded approximately 8 mg protein from a 1 L culture, as determined by an A_{280} absorption reading (Table 2.1). Densitometry was performed to quantify the total amount of ATF5 produced and the percent purity of the final product (Figure 2.2B). Approximately 45 mg of total protein were present in the whole cell lysate derived from 1 L culture, and 20% (or 9 mg) of this corresponds to ATF5 bZIP (Table 2.1). The final yield of purified ATF5 bZIP was 8 mg, which in relation to the total amount produced is 89%. This methodology not only allowed for a very high yield of purified product, but using these conditions concentrations of ATF5 in excess of 0.2 mM (or 2 mg/mL) could be achieved, which permits study of the protein using multi-dimensional NMR analysis.

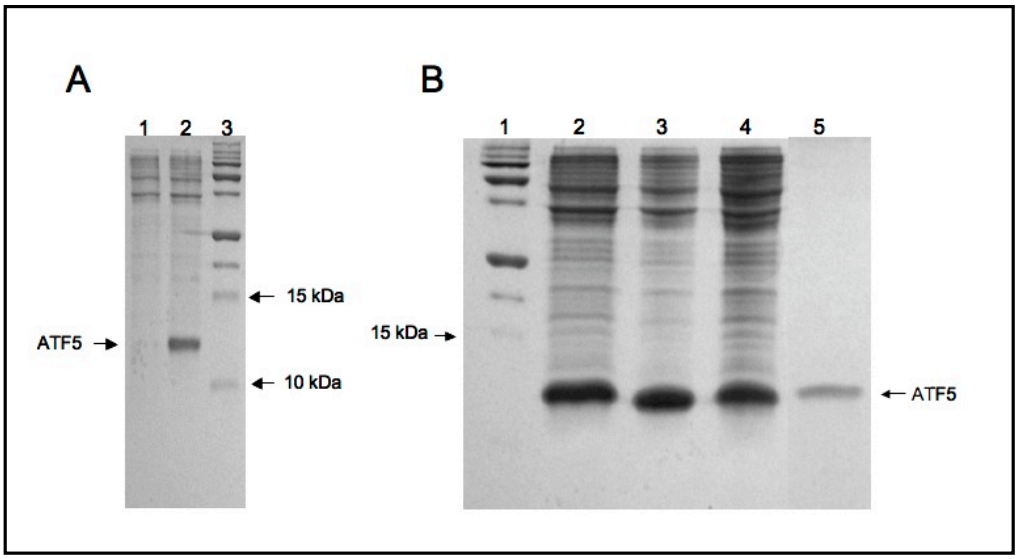

Figure 2.2 (a) Coomassie stained SDS-PAGE showing expression of ATF5 bZIP. Lane 1: *E.Coli* lysate pre-IPTG induction, Lane 2: *E.Coli* lysate post-IPTG induction, Lane 3: molecular weight marker (Bio-Rad Precision Plus Unstained MW Standard). (b) Coomassie stained SDS-PAGE gel showing lysis fractions and purified ATF5 bZIP. Lane 1: molecular weight marker (Bio-Rad Precision Plus Unstained MW Standard), Lane 2: whole cell lysate, Lane 3: insoluble pellet after lysis, Lane 4: lysis supernatant, Lane 5: purified ATF5 bZIP domain.

Total protein (mg) Whole Cell Lysate	Amount Expressed ATF5 (mg)	Final Amount Purified ATF5 (mg)	Final Yield Protein (%)
45	9	8	89

Table 2.1 ATF5 expression and final yield.

2.3.3 Mass Spectrometric Analyses

Analysis of a non-reduced sample of ATF5 indicated the presence of both monomeric (average theoretical mass = 9333.4) and covalently-linked dimeric (average theoretical mass = 16664.8) species (Figure 2.3A). Reduction of ATF5 by DTT resulted in loss of signal from the dimer and a concomitant increase in signal from the monomer (Figure 2.3B). SDS-PAGE performed in the absence of DTT or any reducing agent further confirmed the presence of a mixture of monomeric and dimeric species in the refolded ATF5 sample (Figure 2.3C). Addition of DTT to the sample followed by SDS-PAGE produced a single band corresponding exclusively to the monomer. The apparent molecular weight of the dimer is 24 kDa, which is consistent with the migration of the monomeric species at approximately 12 kDa. The SDS-PAGE results indicate an intermolecular disulfide bond forms readily in solution. Densitometry performed on the SDS-PAGE gel shown in Figure 3 reveals a ratio of 1.0:1.7 monomer:dimer.

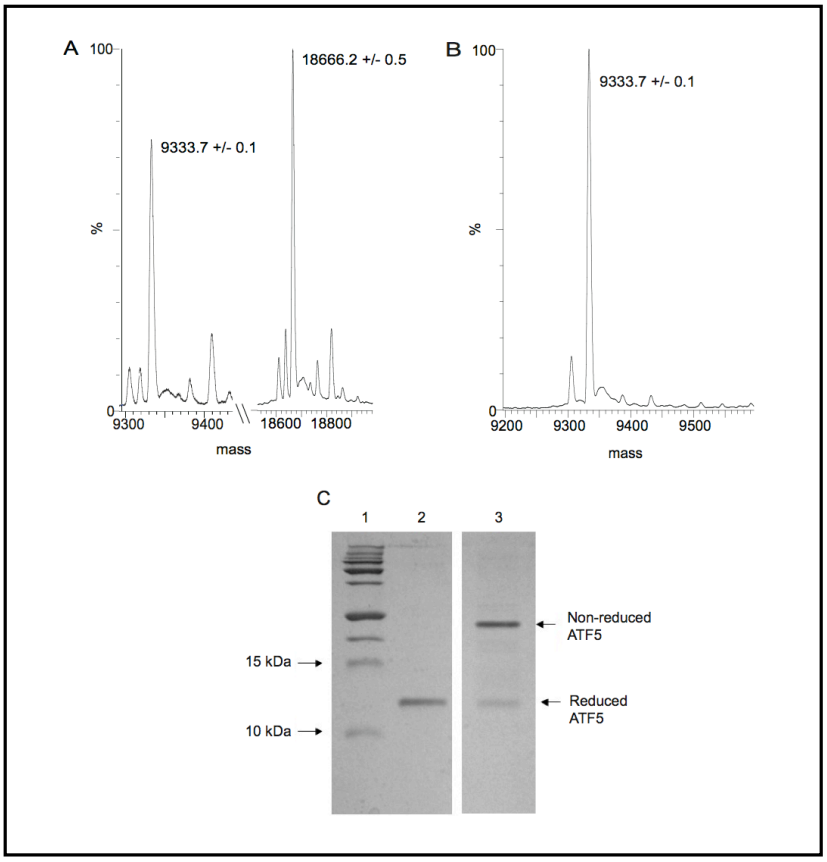

Figure 2.3 (a) Transformed mass spectrum of non-reduced ATF5 bZIP domain. (b) Transformed mass spectrum of reduced ATF5 bZIP domain. (c) Coomassie stained SDS-PAGE gel of purified ATF5 bZIP. Lane 1: molecular weight marker (Bio-Rad Precision Plus Unstained MW Standard), Lane 2: reduced sample of ATF5, Lane 3: Non-reduced sample of ATF5. Note: The non-reduced sample was run on the same gel as the reduced sample but with several lanes in between the two to prevent diffusion of the reducing agent into the non-reduced lane. The intermediate lanes were removed from the figure above and the two sample lanes were placed adjacent one another to allow for better comparison.

2.3.4 NMR Analysis

Purified ^{15}N-labeled ATF5 bZIP was examined using two-dimensional, heteronuclear NMR to verify that the protein can be detected by the spectrometer and is not aggregated when prepared at high concentration, as this would impede further structural investigations using solution NMR. A spectrum of the reduced ATF5 sample was acquired using the ^{15}N-HSQC experiment, because further evaluation of protein structure using NMR often relies on HSQC-based experiments. The ^{15}N-HSQC selectively detects NH pairs and correlates each ^{1}H to the directly attached ^{15}N. This spectrum is considered a "fingerprint" of the protein, because chemical shifts reflect the unique environment encountered by each nucleus in the backbone in an organized structure. A well-behaved globular protein will have good signal dispersion, typically ranging from 6 to 12 ppm on the ^{1}H axis and 100 to 140 ppm on the ^{15}N. This dispersion results from the presence of stable structural features that confine the amide groups of the constituent amino acids to unique chemical environments within the folded protein. As such, more disordered proteins typically have peaks clustered more closely around random coil values (~8.3 ppm in ^{1}H), because on average the nuclei experience more similar chemical environments in solution than in the structured protein [14]. The HSQC spectra of proteins composed of only a single helix, like the ATF5 bZIP monomer, are difficult to interpret without additional data, because in a standard alpha helix, all the amides are in similar environments. Their environment, however, can be altered upon dimerization, and the presence of peaks in more unusual positions can imply coiled-coil formation [6, 15].

The ^{1}H-^{15}N HSQC spectrum of reduced ATF5 suggests that the protein is predominantly in a monomeric, homogeneous state under the conditions used (Figure 2.4). Limited chemical shift dispersion is observed in the spectrum, as the majority of peaks appear between 7.9 and 8.5 ppm on the ^{1}H axis (see black peaks in Figure 2.4). The narrow range of peak positions is consistent with the presence of a species that is composed of a single helix and/or is unstructured [14]. Side chain NH_2 moieties from Asn and Gln residues produce a pair of ^{1}H peaks corresponding to a single ^{15}N value, which are observed at approximately 112 ppm. In the spectrum of ATF5, all the side chain amides have nearly equivalent chemical shifts to each other and the free amino acids, suggesting all are equally solution exposed. While this single NMR spectrum of ATF5 cannot be used to distinguish regions of α-helix from random coil, it does indicate that the protein exists largely in a homogeneous state. Increased signal dispersion is, however, observed closer to the noise (see red peaks in Figure 2.4). The peaks highlighted in red are visible only at a 10-fold lower signal-to-noise ratio than the peaks shown in black. The red peaks near 111 ppm on the ^{15}N axis and also those at larger ^{15}N values between 8.3 and 7.4 ppm on the ^{1}H axis are characteristic of higher order structure. Among these peaks is a pair that corresponds to an NH_2 side chain (111.1 ppm ^{15}N, 6.75 and 7.46 ppm ^{1}H). The presence of these peaks indicates that the environment around one of the Asn or Gln side chains is altered in the more structured state by participating in a protein-protein interaction. Overall, the red peaks constitute only a small percentage of the total peak intensity (less than 10%) in the

ATF5 spectrum, suggesting that even at 0.2 mM concentration a minimal amount of higher-order structure is present.

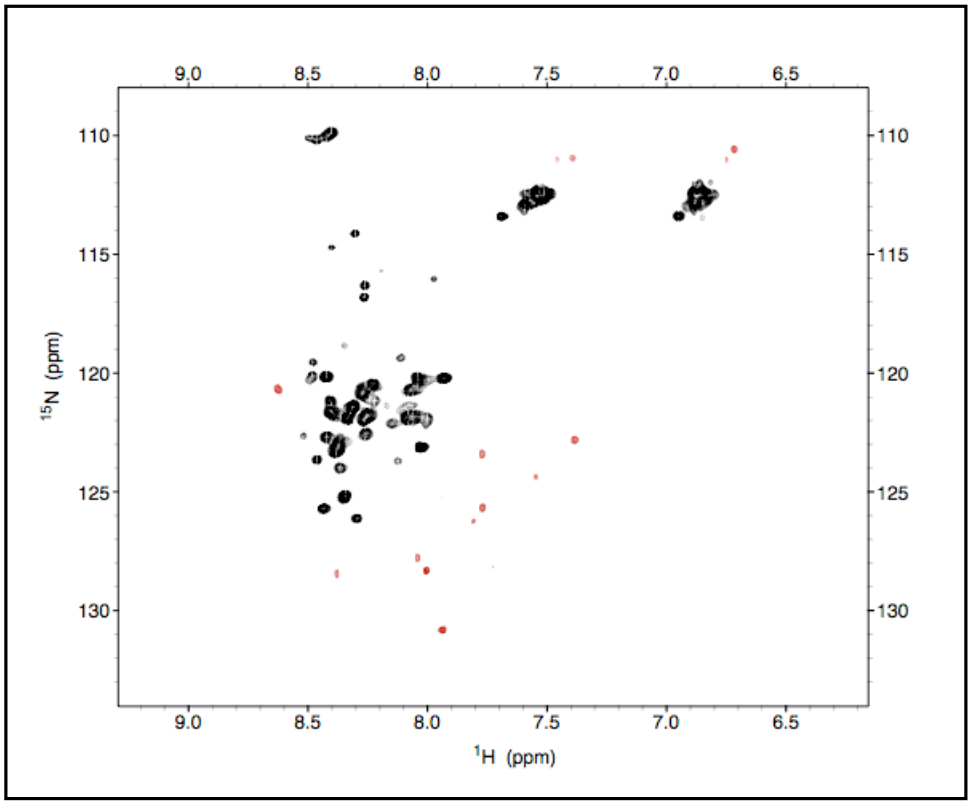

Figure 2.4 2D ^{1}H-^{15}N HSQC spectrum of 0.2 mM ^{15}N-labeled ATF5 in 50 mM sodium phosphate, pH 6.5, containing 100 mM NaCl and 5 mM DTT. Peaks in black are observed at S/N ratio = 1000:1. Peaks in red were overlaid onto the spectrum in black and are observed only at signal/noise ratio = 100:1.

2.3.5 Circular Dichroism Analysis

CD was performed on ATF5 bZIP to determine the secondary structure. Analysis of the CD spectrum of the purified bZIP domain of ATF5 shows the presence of α-helical structure, evident in the double minima absorption observed at both 208 and 222 nm (Figure 2.5). CD data can be used to quantify the relative amounts of secondary structure in a protein. A simple calculation using the following equation can be performed to estimate the α-helical content:

$$[\theta]_{222} = -30{,}300f_H - 2340,$$

where $[\theta]_{222}$ is the mean residue ellipticity at 222nm and f_H is the fraction of α-helical content in the protein [16]. This equation has been applied to other bZIP domains as a means of approximating the percent of alpha helix composing the structure [17]. Using this method, the calculated percent helix in ATF5 bZIP is 27%. The estimate equates to the involvement of 21 residues per monomer in helical structure. A more accurate quantitative analysis requires the inclusion of the data from 190 to 260 nm. Absorption data for this entire range could not be obtained for ATF5. Data between 190 and 200 nm is not shown because measurement in this region is unreliable under the conditions required for ATF5 analysis. Absorption from buffer components interferes with the data collection, and rigorous quantitative analysis cannot be performed because signal from the ATF5 protein cannot be deconvoluted.

2.3.6 DNA Binding Analysis

The DNA binding activity of ATF5 was examined using an electromobility shift assay (Figure 2.6). Because it was previously reported that ATF5 binds to the

CRE motif, we incubated the purified bZIP domain with fluorescently labeled double-stranded oligonucleotide DNA containing this motif to verify the protein functions as expected [7]. A shift in the electrophoretic mobility of the CRE DNA is observed in the presence of ATF5. As expected the DNA: protein complex migrates more slowly (Figure 2.6, lane 2) than that of the DNA alone (lane 1). Addition of 100-fold excess unlabeled CRE DNA results in restoration of the band corresponding to unbound labeled CRE (lane 3). This indicates that binding is reversible because unlabeled CRE effectively competes with labeled CRE for ATF5 binding.

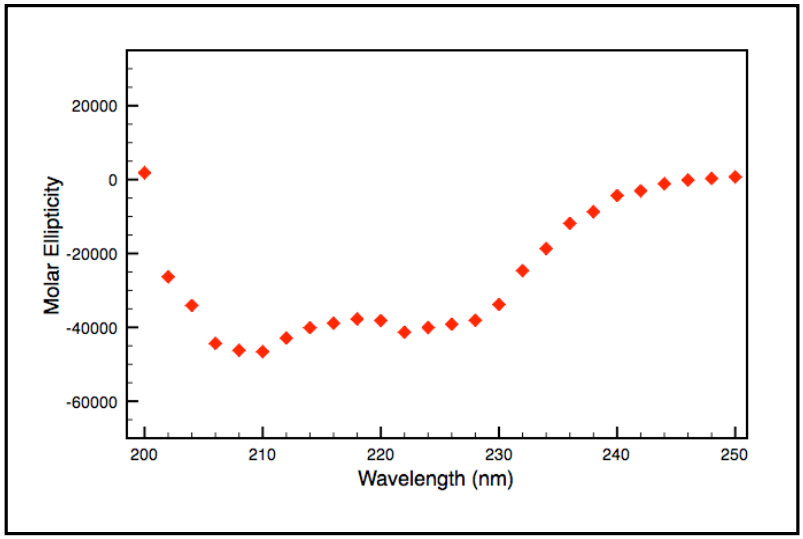

Figure 2.5 Circular dichroism absorption spectrum of the bZIP domain of ATF5 at 20 μM concentration in 50 mM sodium phosphate, pH 6.5, containing 100 mM NaCl and 5 mM DTT. The double minimum at 208 and 222 nm indicates α-helix is present in ATF5.

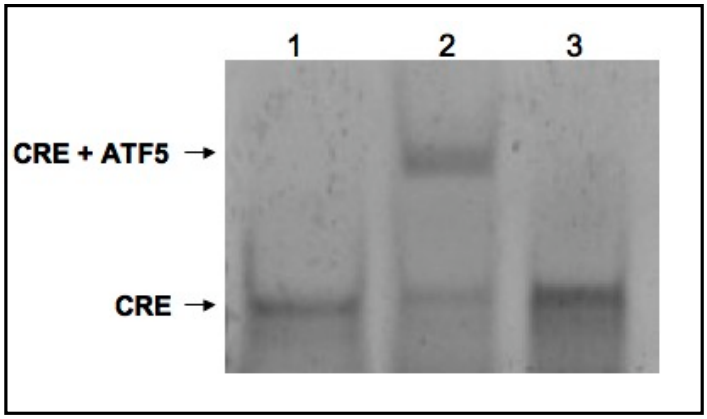

Figure 2.6 Native PAGE showing the electrophoretic mobility of fluorescently tagged CRE DNA in the absence and presence of ATF5. Lane 1: dsCRE DNA, Lane 2: dsCRE DNA in the presence of ATF5, Lane 3: dsCRE DNA in the presence of ATF5 and 100-fold molar excess of unlabeled dsCRE DNA.

2.4 DISCUSSION

The results presented here show for the first time that the bZIP domain of ATF5 can be expressed and purified in sufficient quantity, concentration and homogeneity to facilitate assay development and structural analyses. Our refolding procedure for isolating ATF5 is simple and involves limited introduction of surfaces that could initiate aggregation or adsorption events, which is important because the protein appears prone to aggregation. The procedure is efficient, resulting in a very high yield of soluble protein in which the final product is sufficiently pure to permit spectroscopic or structural analysis.

Most reports of bZIP proteins indicate that these domains are unstructured or weakly helical and monomeric in solution and their helicity is stabilized by DNA binding via coiled-coil formation. An exception to this was reported in two separate studies on GCN4 that provided evidence that monomeric and dimeric species can be helical in solution and both are able to bind DNA [18]. Our data most closely resembles that of GCN4. The CD results for reduced ATF5 indicate approximately one-third of the bZIP domain is alpha helical in solution in the absence of DNA, while the NMR spectrum lacks the dispersion typically associated with coiled-coil formation. In the expressed protein, approximately 45% of the sequence constitutes the basic DNA binding segment while the other 55% is expected to compose the leucine zipper. The DNA binding region is expected to be in a random coil conformation in the absence of DNA. The remaining residues typically are helical in the coiled-coil form, which equates to 43 residues in our construct. Our CD data shows that approximately half of

the predicted zipper region (21 residues) is helical in solution. The comparatively low helical content likely results because the last two typically conserved Leu positions in the zipper region are occupied by valine residues in ATF5. Approximately 90% of the peaks in the NMR spectrum (black in Figure 4) are within a narrow range, which is consistent with the presence of a monomer. If ATF5 exists predominantly as a coiled-coil dimer, the chemical shifts in the NMR spectrum would be expected to appear in a broader range than is observed. The fact that the NMR data was acquired at a 10-fold greater protein concentration negates the possibility that there is any coiled-coil in the CD sample and indicates that the monomer is helical. The subset of more disbursed peaks in the NMR spectrum, which have weaker intensity, suggest that only a minor amount of homodimer is present at the higher protein concentration. An additional recent study of GCN4 provides good evidence to support this conclusion, as it identified the presence of an intermediate form of the GCN4 protein (x-form) that is monomeric yet retains helical structure in the absence of DNA [6]. This study further describes a subset of more distributed NMR peaks, which reflect the coiled-coil structure. The black peaks in our spectrum of ATF5 are consistent with ATF5 being in the x-form, while the red peaks resemble those observed for the coiled-coil form in the GCN4 study. The pair of NH_2 peaks shown in red further lends support to the idea that these peaks reflect dimer formation, because in other bZIP domains a conserved Asn is embedded in the dimer interface [4d]. This residue is in the **a** position in the helix, which directly participates in the "knob and hole" packing arrangement composing the coiled-coil [19]. It is most likely that the shifted NH_2 peaks correspond

to the analogous position in ATF5, which is occupied by Asn245. Peak assignment and structure determination will be required to test this hypothesis.

The bZIP family proteins are known to homo- and heterodimerize selectively [4]. It is this selective dimerization that helps impart DNA binding specificity to these transcription factors. In many cases, bZIP proteins have been shown to be unstructured in the absence of DNA but adopt structure in the DNA-bound form [4a, 4g, 17, 20]. While ATF5 appears to lack higher order structure in solution, because ATF5 can bind specifically to the CRE DNA motif in the absence of another protein, DNA binding may induce formation of a coiled-coil homodimer. ATF4, the closest homolog of ATF5 at 74% sequence homology, has been shown to homo- and heterodimerize [21]. Currently, a high-resolution structure of a homodimer of ATF4 or ATF5 is not available, but a crystal structure of a heterodimer involving ATF4 has been reported [40]. Based on strong sequence identity, this structure can be used to inform our understanding until a structure of ATF5 is completed. In the reported structure ATF4 forms a heterodimeric coiled-coil with CCAAT Enhancer-binding Protein β (C/EBPβ). The ATF4 helix is unusually straight, and heterodimerization is accomplished because C/EBPβ is able to wrap around ATF4. The rod-like conformation observed for ATF4 in the heterodimeric complex seems incompatible with homodimer formation, suggesting an alternative structure must be employed by the homodimer to accomplish DNA binding.

Disulfide bond formation commonly is considered to be an artifact of *in vitro* experiments, and this may be the case in our studies. However, in recent years data

has been accumulating to support that oxidation and reduction of many proteins regulates their activity *in vivo* [22]. Furthermore, it has been previously reported that DNA binding of the Fos/Jun bZIP heterodimer is modulated by redox redulation of a conserved cysteine residue [23]. Examination of the ATF4 structure reveals the Cys involved in homodimeric cross-linking is positioned at the interface of the coiled-coil structure. A cysteine is present in the analogous position in the bZIP domain of ATF5, and this is the only Cys present in either protein as expressed. In the ATF4-C/EBPβ structure paper the authors noted the need to use a mutant of the ATF4 protein in which the cysteine residue in the center of the bZIP domain was mutated to alanine in order to prevent precipitation and obtain the crystal structure of the heterodimer. Likewise, inclusion of reducing agent during refolding helped minimize precipitation of ATF5. This difference in solubility indicates that cross-linking alters the structure of these transcription factors. Our solution studies show that the physical association of ATF5 monomers is weak, with little coiled-coil structure observed even at high protein concentration. If the homodimer has a function *in vivo*, then obviously cross-linking would greatly stabilize the interaction. In ATF5 the intermolecular disulfide bond forms readily in the absence of reducing agents and is slowly reversible with the addition of 1000-fold excess of the thiol-modulating compound DTT. Because ATF5 is upregulated in response to oxidative stress [24], it is tempting to speculate that this intermolecular disulfide bond may influence DNA binding and/or have functional significance in transcriptional regulation. The research presented here provides a basis for conducting a more extensive investigation into

how this disulfide linkage affects the structure of the ATF5 dimer and may influence its interaction with DNA.

2.5 REFERENCES

1. (a) Angelastro, J. M.; Ignatova, T. N.; Kukekov, V. G.; Steindler, D. A.; Stengren, G. B.; Mendelsohn, C.; Greene, L. A., Regulated expression of ATF5 is required for the progression of neural progenitor cells to neurons. *J Neurosci* **2003,** *23* (11), 4590-600; (b) Angelastro, J. M.; Mason, J. L.; Ignatova, T. N.; Kukekov, V. G.; Stengren, G. B.; Goldman, J. E.; Greene, L. A., Downregulation of activating transcription factor 5 is required for differentiation of neural progenitor cells into astrocytes. *J Neurosci* **2005,** *25* (15), 3889-99; (c) Mason, J. L.; Angelastro, J. M.; Ignatova, T. N.; Kukekov, V. G.; Lin, G.; Greene, L. A.; Goldman, J. E., ATF5 regulates the proliferation and differentiation of oligodendrocytes. *Mol Cell Neurosci* **2005,** *29* (3), 372-80.

2. Angelastro, J. M.; Canoll, P. D.; Kuo, J.; Weicker, M.; Costa, A.; Bruce, J. N.; Greene, L. A., Selective destruction of glioblastoma cells by interference with the activity or expression of ATF5. *Oncogene* **2006,** *25* (6), 907-16.

3. (a) Yoon, M. K.; Shin, J.; Choi, G.; Choi, B. S., Intrinsically unstructured N-terminal domain of bZIP transcription factor HY5. *Proteins* **2006,** *65* (4), 856-66; (b) Uesugi, M.; Nyanguile, O.; Lu, H.; Levine, A. J.; Verdine, G. L., Induced alpha helix in the VP16 activation domain upon binding to a human TAF. *Science* **1997,** *277* (5330), 1310-3.

4. (a) Santiago-Rivera, Z. I.; Williams, J. S.; Gorenstein, D. G.; Andrisani, O. M., Bacterial expression and characterization of the CREB bZip module: circular dichroism and 2D 1H-NMR studies. *Protein Sci* **1993,** *2* (9), 1461-71; (b) O'Neil, K. T.; Shuman, J. D.; Ampe, C.; DeGrado, W. F., DNA-induced increase in the alpha-helical content of C/EBP and GCN4. *Biochemistry* **1991,** *30* (37), 9030-4; (c) Patel, L.; Abate, C.; Curran, T., Altered protein conformation on DNA binding by Fos and Jun. *Nature* **1990,** *347* (6293), 572-5; (d) Junius, F. K.; O'Donoghue, S. I.; Nilges, M.; Weiss, A. S.; King, G. F., High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer. *J Biol Chem* **1996,** *271* (23), 13663-7; (e) Saudek, V.; Pasley, H. S.; Gibson, T.; Gausepohl, H.; Frank, R.; Pastore, A., Solution structure of the basic region from the transcriptional activator GCN4. *Biochemistry* **1991,** *30* (5), 1310-7; (f) Saudek, V.; Pastore, A.; Castiglione Morelli, M. A.; Frank, R.; Gausepohl, H.; Gibson, T.; Weih, F.; Roesch, P., Solution structure of the DNA-binding domain of the yeast transcriptional activator protein GCN4. *Protein Eng* **1990,** *4* (1), 3-10; (g) Weiss, M. A.; Ellenberger, T.; Wobbe, C. R.; Lee, J. P.; Harrison, S. C.; Struhl, K., Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* **1990,** *347* (6293), 575-8.

5. (a) Liang, G.; Hai, T., Characterization of human activating transcription factor 4, a transcriptional activator that interacts with multiple domains of cAMP-responsive element-binding protein (CREB)-binding protein. *J Biol Chem* **1997,** *272* (38), 24088-95; (b) Schoch, S.; Cibelli, G.; Magin, A.; Steinmuller, L.; Thiel, G., Modular structure of cAMP response element binding protein 2 (CREB2). *Neurochem Int* **2001,** *38* (7), 601-8.

6. Nikolaev, Y.; Pervushin, K., NMR spin state exchange spectroscopy reveals equilibrium of two distinct conformations of leucine zipper GCN4 in solution. *J Am Chem Soc* **2007,** *129* (20), 6461-9.
7. Peters, C. S.; Liang, X.; Li, S.; Kannan, S.; Peng, Y.; Taub, R.; Diamond, R. H., ATF-7, a novel bZIP protein, interacts with the PRL-1 protein-tyrosine phosphatase. *J Biol Chem* **2001,** *276* (17), 13718-26.
8. Kim, S.; Jao, S.; Laurence, J. S.; LiWang, P. J., Structural comparison of monomeric variants of the chemokine MIP-1beta having differing ability to bind the receptor CCR5. *Biochemistry* **2001,** *40* (36), 10782-91.
9. Laemmli, U. K., Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **1970,** *227* (5259), 680-5.
10. Gasteiger, E.; Gattiker, A.; Hoogland, C.; Ivanyi, I.; Appel, R. D.; Bairoch, A., ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* **2003,** *31* (13), 3784-8.
11. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **1995,** *6* (2), 135-40.
12. (a) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995,** *6* (3), 277-93; (b) Goddard, T. D. K., D.G.
13. (a) Lee, G.; Cowan, N.; Kirschner, M., The primary structure and heterogeneity of tau protein from mouse brain. *Science* **1988,** *239* (4837), 285-8; (b) Lewis, S. A.; Wang, D. H.; Cowan, N. J., Microtubule-associated protein MAP2 shares a microtubule binding motif with tau protein. *Science* **1988,** *242* (4880), 936-9; (c) West, R. R.; Tenbarge, K. M.; Olmsted, J. B., A model for microtubule-associated protein 4 structure. Domains defined by comparisons of human, mouse, and bovine sequences. *J Biol Chem* **1991,** *266* (32), 21886-96.
14. (a) Liu, J.; Song, J., A novel nucleolar transcriptional activator ApLLP for long-term memory formation is intrinsically unstructured but functionally active. *Biochem Biophys Res Commun* **2008,** *366* (2), 585-91; (b) Laguri, C.; Sadir, R.; Rueda, P.; Baleux, F.; Gans, P.; Arenzana-Seisdedos, F.; Lortat-Jacob, H., The Novel CXCL12gamma Isoform Encodes an Unstructured Cationic Domain Which Regulates Bioactivity and Interaction with Both Glycosaminoglycans and CXCR4. *PLoS ONE* **2007,** *2* (10), e1110; (c) Hayes, P. L.; Lytle, B. L.; Volkman, B. F.; Peterson, F. C., The solution structure of ZNF593 from Homo sapiens reveals a zinc finger in a predominantly unstructured protein. *Protein Sci* **2008,** *17* (3), 571-6; (d) Eliezer, D., Characterizing residual structure in disordered protein States using nuclear magnetic resonance. *Methods Mol Biol* **2007,** *350*, 49-67; (e) Bertoncini, C. W.; Rasia, R. M.; Lamberto, G. R.; Binolfi, A.; Zweckstetter, M.; Griesinger, C.; Fernandez, C. O., Structural characterization of the intrinsically unfolded protein beta-synuclein, a natural negative regulator of alpha-synuclein aggregation. *J Mol Biol* **2007,** *372* (3), 708-22.
15. (a) Matousek, W. M.; Ciani, B.; Fitch, C. A.; Garcia-Moreno, B.; Kammerer, R. A.; Alexandrescu, A. T., Electrostatic contributions to the stability of the GCN4

leucine zipper structure. *J Mol Biol* **2007,** *374* (1), 206-19; (b) Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Janicka, A.; Martinelli, M.; Kozlowski, H.; Palumaa, P., A structural-dynamical characterization of human cox17. *J Biol Chem* **2008,** *283* (12), 7912-20; (c) Gingras, A. R.; Bate, N.; Goult, B. T.; Hazelwood, L.; Canestrelli, I.; Grossmann, J. G.; Liu, H.; Putz, N. S.; Roberts, G. C.; Volkmann, N.; Hanein, D.; Barsukov, I. L.; Critchley, D. R., The structure of the C-terminal actin-binding domain of talin. *EMBO J* **2008,** *27* (2), 458-69; (d) Ishigaki, T.; Ohki, I.; Utsunomiya-Tate, N.; Tate, S. I., Chimeric structural stabilities in the coiled-coil structure of the NECK domain in human lectin-like oxidized low-density lipoprotein receptor 1 (LOX-1). *J Biochem* **2007,** *141* (6), 855-66; (e) Moreau, V. H.; da Silva, A. C.; Siloto, R. M.; Valente, A. P.; Leite, A.; Almeida, F. C., The bZIP region of the plant transcription factor opaque-2 forms stable homodimers in solution and retains its helical structure upon subunit dissociation. *Biochemistry* **2004,** *43* (16), 4862-8.

16. (a) Chen, Y. H.; Yang, J. T.; Martinez, H. M., Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry* **1972,** *11* (22), 4120-31; (b) Chen, Y. H.; Yang, J. T.; Chau, K. H., Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry* **1974,** *13* (16), 3350-9.

17. John, M.; Leppik, R.; Busch, S. J.; Granger-Schnarr, M.; Schnarr, M., DNA binding of Jun and Fos bZip domains: homodimers and heterodimers induce a DNA conformational change in solution. *Nucleic Acids Res* **1996,** *24* (22), 4487-94.

18. Chan, I. S.; Fedorova, A. V.; Shin, J. A., The GCN4 bZIP targets noncognate gene regulatory sequences: quantitative investigation of binding at full and half sites. *Biochemistry* **2007,** *46* (6), 1663-71.

19. Vinson, C.; Acharya, A.; Taparowsky, E. J., Deciphering B-ZIP transcription factor interactions in vitro and in vivo. *Biochim Biophys Acta* **2006,** *1759* (1-2), 4-12.

20. (a) Chen, L.; Glover, J. N.; Hogan, P. G.; Rao, A.; Harrison, S. C., Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **1998,** *392* (6671), 42-8; (b) Glover, J. N.; Harrison, S. C., Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **1995,** *373* (6511), 257-61; (c) Krebs, D.; Dahmani, B.; el Antri, S.; Monnot, M.; Convert, O.; Mauffret, O.; Troalen, F.; Fermandjian, S., The basic subdomain of the c-Jun oncoprotein. A joint CD, Fourier-transform infrared and NMR study. *Eur J Biochem* **1995,** *231* (2), 370-80.

21. Podust, L. M.; Krezel, A. M.; Kim, Y., Crystal structure of the CCAAT box/enhancer-binding protein beta activating transcription factor-4 basic leucine zipper heterodimer in the absence of DNA. *J Biol Chem* **2001,** *276* (1), 505-13.

22. (a) Biswas, S.; Chida, A. S.; Rahman, I., Redox modifications of protein-thiols: emerging roles in cell signaling. *Biochem Pharmacol* **2006,** *71* (5), 551-64; (b) Thannickal, V. J.; Fanburg, B. L., Reactive oxygen species in cell signaling. *Am J Physiol Lung Cell Mol Physiol* **2000,** *279* (6), L1005-28; (c) Rhee, S. G.; Bae, Y. S.; Lee, S. R.; Kwon, J., Hydrogen peroxide: a key messenger that modulates protein phosphorylation through cysteine oxidation. *Sci STKE* **2000,** *2000* (53), PE1; (d) Giannoni, E.; Buricchi, F.; Raugei, G.; Ramponi, G.; Chiarugi, P., Intracellular

reactive oxygen species activate Src tyrosine kinase during cell adhesion and anchorage-dependent cell growth. *Mol Cell Biol* **2005,** *25* (15), 6391-403; (e) Tonks, N. K., Redox redux: revisiting PTPs and the control of cell signaling. *Cell* **2005,** *121* (5), 667-70; (f) Wu, R. F.; Xu, Y. C.; Ma, Z.; Nwariaku, F. E.; Sarosi, G. A., Jr.; Terada, L. S., Subcellular targeting of oxidants during endothelial cell migration. *J Cell Biol* **2005,** *171* (5), 893-904; (g) Blankman, J. I.; Shahzad, N.; Miller, C. J.; Guiles, R. D., Direct voltammetric investigation of the electrochemical properties of human hemoglobin: relevance to physiological redox chemistry. *Biochemistry* **2000,** *39* (48), 14806-12.

23. Abate, C.; Patel, L.; Rauscher, F. J., 3rd; Curran, T., Redox regulation of fos and jun DNA-binding activity in vitro. *Science* **1990,** *249* (4973), 1157-61.

24. Watatani, Y.; Ichikawa, K.; Nakanishi, N.; Fujimoto, M.; Takeda, H.; Kimura, N.; Hirose, H.; Takahashi, S.; Takahashi, Y., Stress-induced translation of ATF5 mRNA is regulated by the 5'-untranslated region. *J Biol Chem* **2008,** *283* (5), 2543-53.

CHAPTER 3

EFFECTS OF DISULFIDE BOND FORMATION AND PROTEIN HELICITY ON THE AGGREGATION OF ATF5

3.1 INTRODUCTION

Amorphous aggregation is a common phenomenon and a major obstacle to handling proteins *in vitro*. The aggregation of proteins isolated for use as pharmaceutical products poses a serious concern, as aggregates can stimulate the immune system, which may have devastating affects on the patient.[1] Although the study of protein aggregation has wide-reaching implications, the process by which aggregation occurs is still difficult to predict. The study of amyloids has provided understanding about the structure and formation of fibrillar aggregates, but a lack of data regarding amorphous structure has prevented verifiable application to amorphous aggregation. In particular, the role that disulfide bonds play in this process is poorly understood, because it varies among different systems. Reports indicate in some instances that disulfide bond formation promotes protein aggregation, yet in others it dampens this process.[2] These effects appear to be largely protein specific and dependent on features of tertiary structure. Alpha-helices can also play an important role in aggregation. In most instances, retention of α-helical structure has been linked to improved protein stability; [3] however, several recent reports indicate that α-helical structure can facilitate aggregate formation.[4]

The lack of understanding surrounding non-fibrillar aggregation is largely due to the fact that the assembled species lacks a regular repeating pattern, making it is difficult to study. Well-structured and natively disordered proteins do not readily aggregate in solution, whereas partially unfolded proteins tend to aggregate rapidly, suggesting structured features facilitate aggregation. Here we present a unique and simple system, which is amenable to investigating the mechanism of protein aggregation involving a helix to coil transition. We have previously reported the isolation of the bZIP domain of activating transcription factor 5 (ATF5), a protein notable for its potential as a pharmaceutical target for treatment of glioblastoma multiforme.[5] This domain consists of a single α-helix and possesses a single cysteine residue. It is only partially structured (~25% helix) and displays marginal stability in solution under physiological conditions. We have tested solution conditions that modulate helical propensity and the oxidation state of the thiol, to successfully investigate the role that α-helical structure and disulfide bond formation play in protein stability. These studies have led to the identification of stabilizing conditions for ATF5, which will enable further study of the protein as a pharmaceutical target and provided a model system for further examination of residue-specific contributions to the formation of amorphous aggregates.

3.2 MATERIALS AND METHODS

3.2.1 Protein Expression and Purification

The cDNA of ATF5 was obtained through ATCC (MGC-842) and the bZIP domain was PCR amplified and inserted into a Novagen (San Diego, CA) pET-42b vector as described previously.[5] Site-directed mutagenesis was performed to create the C240A point mutation in the bZIP region using the following primers: 5'GAGGCCCTGGAGGGCGAGGCCCAGGGGCTGGAGGCACGG (forward primer), 5'-CCGTGCCTCCAGCCCCTGGGCCTCGCCCTCCAGGGCCTC (reverse primer). Plasmids were transformed into Novagen BL21(DE3) *E. Coli* for expression. Purification was done using a refolding procedure described previously.[5] Protein concentrations were determined using a standard Bradford assay. ATF5 purification and isolation was confirmed using SDS-PAGE and LC/MS.[5] These methods were also used to confirm intermolecular disulfide bond formation for wild-type ATF5.[5] Isotopic labeling for two-dimensional NMR was accomplished with the substitution of ^{15}N-ammonium chloride. The 15N-ammonium chloride (>99% N-15) was obtained from Spectra Stable Isotopes (Columbia, MD).

3.2.2 Circular Dichroism

Circular dichroism (CD) spectra were acquired with a Jasco-810 polarimeter (Tokyo, Japan) equipped with a Peltier-type temperature controller and a six-position sample holder. Samples were prepared at a final protein concentration of 100 μM in 20 mM MES buffer at pH 6.0 in the absence or presence of dithiothreitol (DTT) or trifluoroethanol (TFE). Scans were performed at 4°C from 260 to 190 nm in a 0.1-cm path length cell. A scanning speed of 50 nm/min and a response time of 8 sec were used. The selected bandwidth was 5 nm and the data pitch was 1 nm. Scans were run

in duplicate for each sample and two samples were prepared and analyzed for each set of solution conditions. Using the Jasco Spectra Manager v1.18.00 software the measured values were converted to molar ellipticity. The resulting average and standard deviations were plotted for each condition using Plot. Percent helicity was calculated using the mean residue ellipticity observed at 222 nm as previously described.[5-6]

To evaluate protein stability the absorbance at 222 nm was monitored at a function of temperature. The temperature was adjusted from 4 to 85°C at a ramping rate of 15°C/hour and the absorbance was monitored every 0.5°C. The results were converted to molar ellipticity using the instrument software. Two samples were prepared and analyzed for each set of solution conditions. The resulting average and standard deviations were plotted using Plot. Midpoints of thermal transition were calculated using Microcal Origin sigmoidal fit graphing tools. None of the solution conditions used produced signals above background at any temperature and therefore, no buffer subtraction manipulations were performed.

A single sample t-test was performed to determine the statistical significance in mean CD signal between two populations. In all cases the theoretical mean comprising the null hypothesis consisted of the CD signal observed for the ATF5 protein in the absence of DTT and/or TFE. A two-tail p-value is reported.

3.2.3 Static Light Scattering

Static light scattering (SLS) data were acquired with a Photon Technology International (PTI) spectrafluorometer (Lawrenceville, NJ) equipped with a Peltier-

type temperature controller and a four-position sample holder. Samples were prepared at a final protein concentration of 100 μM in 20 mM MES buffer at pH 6.0 in the absence or presence of dithiothreitol (DTT) or trifluoroethanol (TFE). The intensity of scattered light was measured as a function of temperature and detected at an angle of 90° to the light source by a photomultiplier tube. An arc lamp white light source was used in all cases. The excitation wavelength was set at 275 nm. The emission wavelength range was set from 250 nm to 350 nm. Spectra were obtained following a 5 min equilibration period at each temperature. Data were collected every 2.5°C between the temperature range of 4 to 81.5°C. The background was subtracted from each data point based on a blank containing the corresponding buffer solution. Two samples were prepared and analyzed for each set of solution conditions. The resulting average and standard deviations were plotted using Plot.

3.2.4 Dynamic light scattering

The average hydrodynamic diameter was monitored as a function of temperature using a Brookhaven (Holtsville, NY) BLS-9000 DLS instrument equipped with a 50 mW HeNe diode laser operating at 532 nm and a Brookhaven BI-200SM goniometer. Samples were prepared in 20 mM MES buffer pH 6.0 at a final concentration of 100 μM. Five consecutive 30 sec data collection intervals were taken every 2.5°C over the range of 10 to 90 °C. Light scattering was monitored 90° to the incident light and autocorrelation functions were generated using a Brookhaven BI-9000AT digital autocorrelator. Analysis was performed using the method of

cumulants. The hydrodynamic diameter was calculated from the translational diffusion coefficient using the Stokes-Einstein equation.[7]

3.2.5 Nuclear Magnetic Resonance

Two-dimensional ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra were recorded at 25°C using a Bruker AVANCE 800 MHz spectrometer equipped with a triple-resonance CRYO-probe with pulse field gradients. Samples were prepared at a concentration of 0.2 mM in 50 mM sodium phosphate, 100mM NaCl pH 6.5, containing 5% D_2O. Water suppression was accomplished using flip-back pulses. Data were acquired in 128 scans with 2048 points in ^{1}H and 128* increments in ^{15}N. ^{1}H chemical shifts were referenced with respect to an external DSS standard in D_2O.[8] Indirect referencing relative to ^{1}H was determined for ^{15}N, assuming a ratio of $^{15}N/^{1}H = 0.101329118$. Data were processed using NMRPipe and Sparky software.[9]

3.3 RESULTS

3.3.1 Effects of Disulfide Bond Formation on ATF5 Structure and Stability

CD spectroscopy was used to investigate the influence of a single intermolecular disulfide bond (C240-C240) on the structure of the bZIP domain of ATF5. The spectra reveal absorption characteristics consistent with the presence of α-helical structure for both the disulfide-bound dimer and the C240A monomer (Figure 3.1A). This is depicted by the double absorption minima pattern detected at 222 nm and 208 nm.[10] The ratio of 222 to 208 nm is 0.84 for wild type and 0.85 for

C240A, which is consistent with a-helical but not coiled-coil structure.[11] An increase in helical content is observed for the dimer in comparison to the monomer. This difference can be quantified for each species using the mean residue ellipticy at 222 nm.[5-6] The percent helicity calculated for the dimer under these conditions is 23.4% ± 0.3%, whereas the percent helicity of the C240A monomer is 19.0% ± 1.0% (p-value < 0.02).

Upon the addition of reducing agent to wild-type (WT) ATF5 (10 mM DTT), a decrease in the CD signal that corresponds to α-helix is observed (Figure 3.1B). The absorption intensity for the reduced WT more closely resembles that of the C240A monomer under reducing conditions with a calculated percent helicity of 17.9% (± 0.2%) and 16.7% (± 0.1%), respectively (p-value > 0.2). This suggests the intermolecular disulfide bond is responsible for the increased helicity originally observed for the dimeric species. It is important to note that the decrease in absorption intensity observed at 208 nm relative to 222 nm in the presence of DTT is due to substantially increased interference caused by DTT on the detection at shorter wavelengths. For this reason, absorption measurements taken at wavelengths below 210 nm in the presence of DTT should be considered less accurate. As such, the ratio of 222 nm to 208 nm was not assessed because the interpretation of this value is also unreliable for determining a-helical vs coiled-coil content.

The impact of the intermolecular disulfide bond on thermal stability was evaluated by monitoring the α-helix CD absorption signal at 222 nm as a function of temperature. A decrease in absorption signal at 222 nm indicates a loss of secondary

structure, which reflects protein unfolding. The disulfide-bound dimer is more stable than the C240A monomer. The calculated melting temperature (T_m) of the dimer is 55.5°C, whereas it is 15.9°C for the monomer (Figure 3.2A). In the presence of reducing agent (10 mM DTT) the T_m for the C240A mutant remains relatively unchanged (15.1°C), whereas the melting curve for the dimer shifts dramatically to more closely resemble that of the monomer (Figure 3.2B). The calculated transition temperature for the dimer under these conditions is 19.9°C. Notably, unfolding of all forms of ATF5 was found to be largely irreversible due to precipitation of aggregated species (data not shown).

The effect this intermolecular disulfide bond has on protein aggregation was then evaluated using a static light scattering assay. The intensity of scattered light was measured as a function of temperature for the two forms of ATF5. In this assay, an increase in the intensity of scattered light indicates increased aggregate formation. The data show that the onset of aggregation is more rapid, and the extent of aggregation is greater for the C240A mutant compared to the WT under non-reducing conditions (Figure 3.3A). The onset of aggregation is more gradual and the extent lessened for the disulfide-bound dimer (Figure 3.3A). Upon reduction of the disulfide bond via the addition of reducing agent (10 mM DTT), the scattering profile of the dimer more closely parallels that of the C240A monomer (Figure 3.3B). The decrease in signal seen at higher temperatures in Figure 3B indicates the formation of very large aggregates that have settled out of solution, resulting in diminished absorbance intensity. This was verified using dynamic light scattering analysis, which

indicated that the average hydrodynamic diameter of protein in solution was much greater in the presence of DTT (data not shown). Clearly, reduction of the disulfide bond decreases thermal stability, as the two proteins now behave similarly.

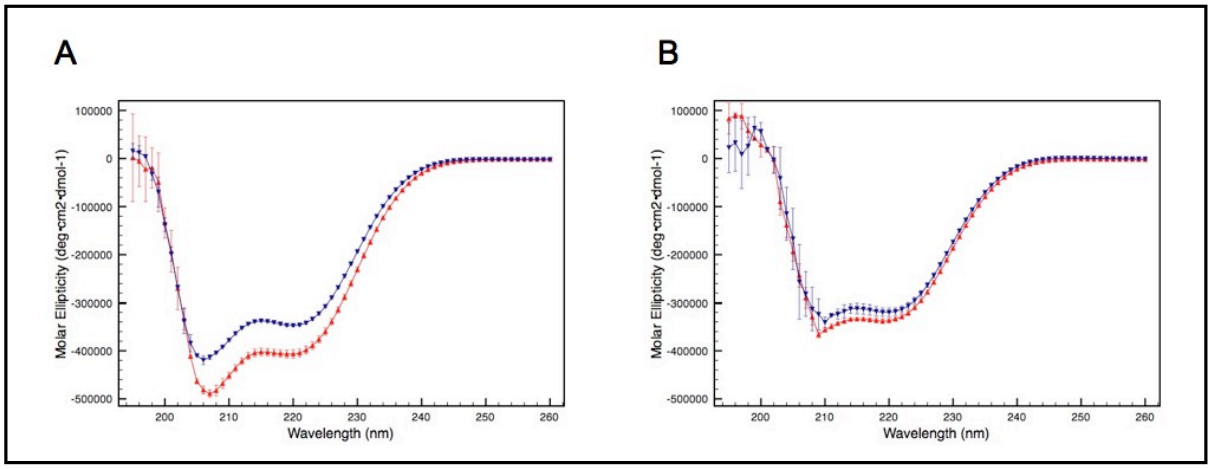

Figure 3.1 (A) CD absorption spectra of the disulfide-linked ATF5 dimer (red upright triangles) and the C240A monomer (blue downturned triangles) in the absence of 10 mM DTT. (B) CD absorption spectra of the CD absorption spectra of the ATF5 dimer (red upright triangles) and the C240A monomer (blue downturned triangles) in the presence of 10 mM DTT.

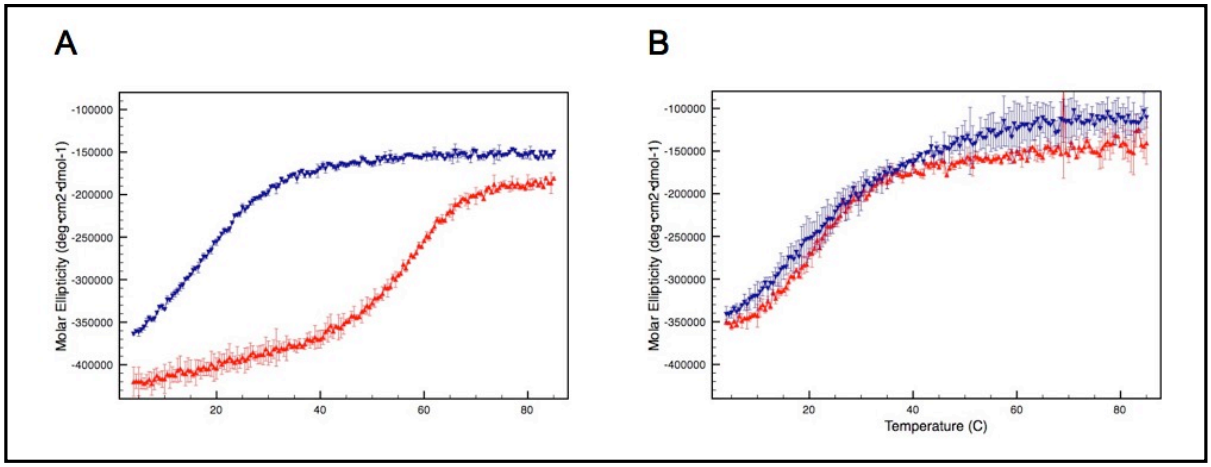

Figure 3.2 (A) CD absorption at 222 nm of the disulfide-bound ATF5 dimer (red upright triangles) and C240A monomer (blue downturned triangles) in the absence of 10 mM DTT as a function of temperature. (B) CD absorption at 222 nm of the disulfide-bound ATF5 dimer (red upright triangles) and C240A monomer (blue downturned triangles) in the prescence of 10 mM DTT as a function of temperature.

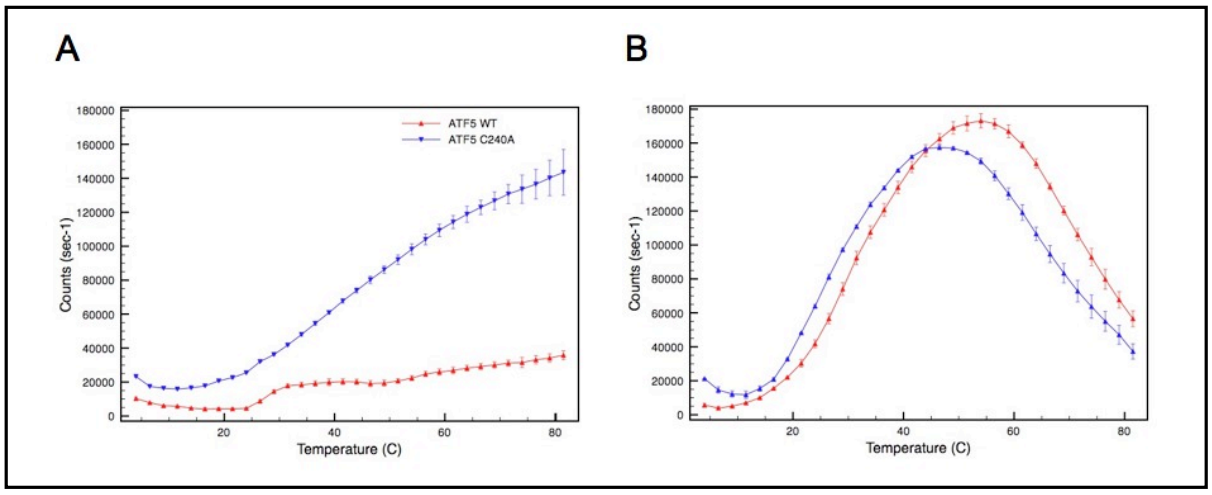

Figure 3.3 (A) SLS analysis of the disulfide bound ATF5 dimer (red upright triangles) and C240A monomer (blue downturned triangles) in the absence of 10 mM DTT as a function of temperature. (B) SLS analysis of the ATF5 dimer (red upright triangles) and C240A monomer (blue downturned triangles) in the presence of 10 mM DTT as a function of temperature.

	Percent Helicity (%)	Melting Temperature (°C)	Maximum Intensity of Scattered Light (CPS)
WT, no DTT	23.4 ± 0.3	55.5	35,815 ± 2,658
WT + DTT	17.9 ± 0.2	19.9	173,132 ± 4,151
C240A, no DTT	19.0 ± 1.0	15.9	143,560 ± 13,475
C240A + DTT	16.7 ± 0.1	15.1	175,415 ± 1,149

Table 3.1 Summary of the results from the CD and SLS experiments for the WT and C240A forms of ATF5 in the absence and presence of the 10 mM DTT.

3.3.2 Effects of TFE-induced Helicity on ATF5 Structure and Stability

The effect that α-helical structure has on protein stability was evaluated using the addition of the co-solvent trifluoroethanol (TFE) to induce helical structure. The use of TFE for the purpose of inducing protein structure is common and well cited.[12] TFE is added to peptides to enhance α-helicity, because the solubility of the backbone changes. Low percentages of TFE typically increase helicity, but at high concentrations (often >40%) β-sheets may form in some peptides.[12a] The impact of TFE on the structure of ATF5 was assessed using CD spectroscopy. The helical content of both the disulfide-bound dimer and the C240A monomer increased in a dose-dependent fashion upon the addition of TFE within the experimental range used (Figure 3.4). The increase in helicity observed for a given amount of TFE added was greater for the C240A mutant than for the dimer, but at 10% TFE and above both forms displayed equivalent amounts of helical structure. The percent helicity was quantified as described above for both forms of ATF5. The dimer displayed a percent helicity of 25.5% (± 0.6%), 30.8% (± 0.6%), and 39.3% (± 1.0%) in the presence of no TFE, 10% TFE (p-value < 0.1) and 20% TFE (p-value < 0.05), respectively. The comparative values for the C240A monomer are 19.5% (± 0.3%), 30.5% (± 0.3%), and 42.0% (± 2.2%) in the presence of no TFE, 10% TFE (p-value < 0.02) and 20% TFE (p-value < 0.1), respectively.

The thermal stability of both forms of ATF5 in the presence of increasing concentration of TFE was investigated by monitoring the α-helical CD absorption signal at 222 nm as a function of temperature. The melting profiles of both forms

changed in the presence of TFE (Figure 3.5). The data show that increasing the helicity of either form of ATF5 improves thermal stability, because greater α-helical structure is retained at all temperatures in the presence of TFE. A standard quantitative analysis comparing T_m values is not possible for these data, because the melting profiles in the presence of 10% and 20% TFE could not be fit to a sigmoidal curve. The melting temperature calculated for the WT form of ATF5 in the absence of TFE is 56.0°C. At that same temperature in the presence of 10% and 20% TFE, there is greater helical structure present. The molar ellipticity values for the WT dimer at 56°C in the presence of 0%, 10% and 20% TFE are -426,189, -581,870, and -630946, respectively. This equates to approximately 9% more α-helix at 10% TFE than at 0% TFE, and 3% more helix at 20% TFE than at 10%. The melting temperature calculated for the C240A monomer in the absence of TFE is 17.2°C. The molar ellipticity values for the C240A mutant at 17°C in the presence of 0%, 10% and 20% TFE are -428,822, -653,246 and -846,000, respectively, equating to 12% more helix at 10% TFE than 0% TFE, and 11% more helix at 20% TFE than 10% TFE.

Analysis of static light scattering data was used to evaluate the effect of TFE concentration on protein aggregation. Increasing concentrations of TFE diminished aggregation for both the cross-linked WT dimer and the C240A monomer (Figure 3.6). The maximum intensity of scattered light in counts per second detected for the WT form is 259,404, 150,731 and 139,743 in the presence of 0%, 10% and 20% TFE, respectively. Notably, a greater difference in the extent of aggregation is observed

between 0% and 10% TFE than between 10% and 20% TFE. This same trend is observed for the C240A monomer, where the maximum intensity of scattered light in counts per second detected in the presence of 0%, 10% and 20% is 779,386, 277,567, and 248,188, respectively. The greatest impact on aggregation in all cases is observed for the C240A mutant in the presence of 10% TFE, where the inclusion of the co-solvent dramatically reduces the onset and extent of aggregation. The impact on aggregation for the WT form at the same concentration of 10% TFE is much less dramatic, even though the percent helicity induced is roughly the same as that of the C240A monomer (approximately 30%). At low temperatures, the addition of TFE leads to a small amount of increased light scattering for both the WT and C240A. There is more scattering at 20% TFE compared to 10% TFE, suggesting that an alternative mode of association may be responsible for aggregation in the TFE-containing samples.

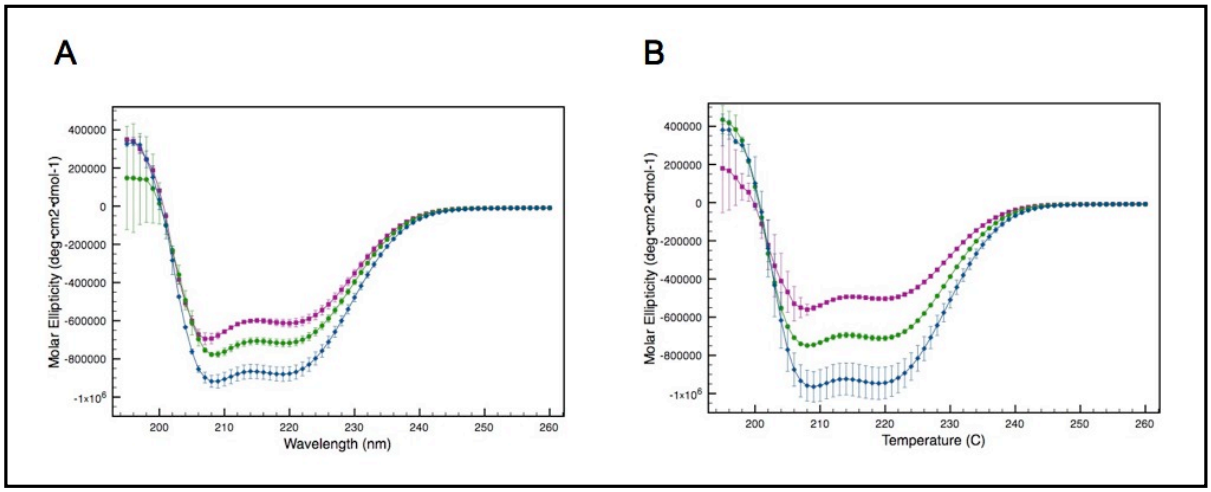

Figure 3.4 (A) CD absorption spectra of the disulfide-linked ATF5 dimer in the presence of increasing amounts of TFE: purple squares (no TFE), green circles (10% [v/v] TFE) and blue diamonds (20%[v/v] TFE). (B) CD absorption spectra of the ATF5 C240A monomer in the presence of increasing amounts of TFE: purple squares (no TFE), green circles (10% [v/v] TFE) and blue diamonds (20%[v/v] TFE).

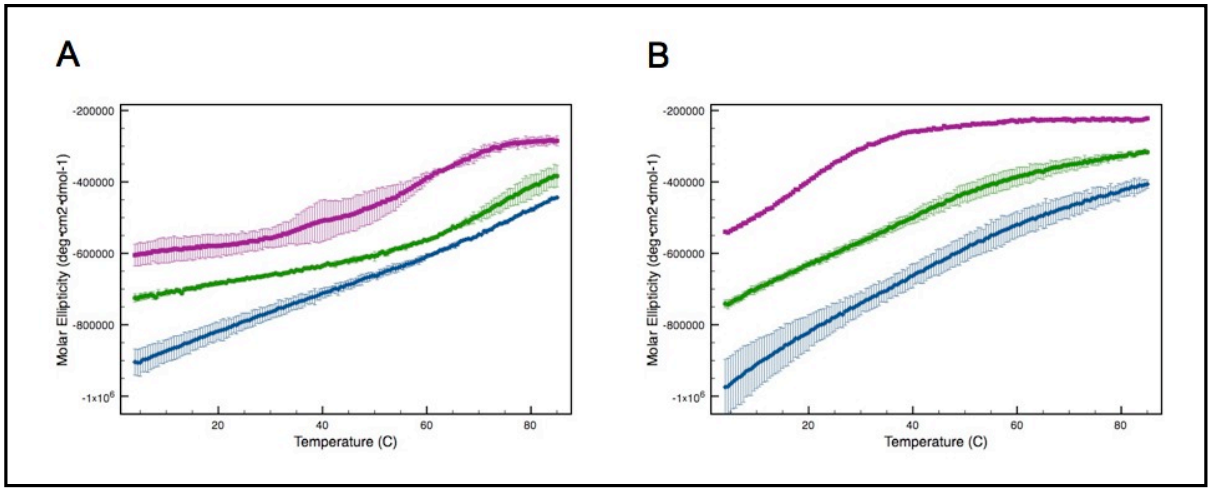

Figure 3.5 (A) CD absorption at 222 nm of the disulfide-linked ATF5 dimer as a function of temperature in the presence of increasing amounts of TFE: purple squares (no TFE), green circles (10%[v/v] TFE) and blue diamonds (20%[v/v] TFE). (B) CD absorption at 222 nm of the ATF5 C240A monomer as a function of temperature in the presence of increasing amounts of TFE: purple squares (no TFE), green circles (10%[v/v] TFE) and blue diamonds (20%[v/v] TFE).

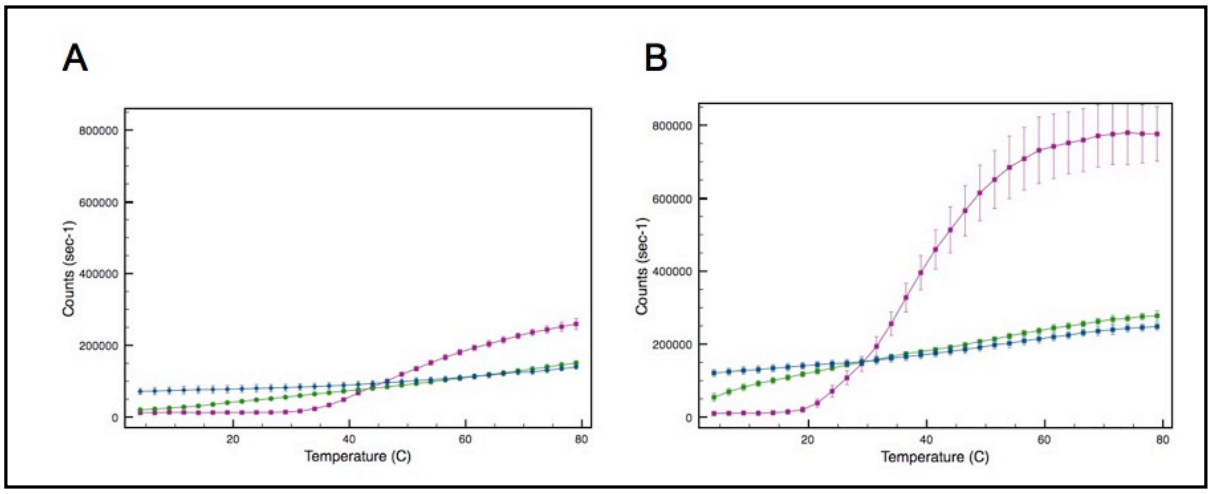

Figure 3.6 (A) SLS analysis of the disulfide-linked ATF5 dimer in the presence of increasing amounts of TFE: (purple squares) no TFE, (green circles) 10% TFE and (blue diamonds) 20% TFE. (B) SLS analysis of the ATF5 C240A monomer in the presence of increasing amounts of TFE: (purple squares) no TFE, (green circles) 10% TFE and (blue diamonds) 20% TFE.

	Percent Helicity (%)	Melting Temperature (°C)	Maximum Intensity of Scattered Light (CPS)
WT, no TFE	25.5 ± 0.6	56.0	259,404 ± 15,061
WT, 10% TFE	30.8 ± 0.6	NA	150,731 ± 6,096
WT, 20% TFE	39.3 ± 1.0	NA	139,743 ± 6,404
C240A, no TFE	19.5 ± 0.3	17.2	779,386 ± 87,881
C240A, 10% TFE	30.5 ± 0.3	NA	277,567 ± 13,598
C240A, 20% TFE	42.0 ± 2.2	NA	248,188 ± 9,972

Table 3.2 Summary of the results from the CD and SLS experiments for the WT and C240A forms of ATF5 in the presence of increasing amounts of the co-solvent TFE. A melting temperature could not be calculated for the protein in the presence of 10% and 20% TFE, because the data could not be fit to a sigmoidal curve. NA= not applicable.

3.3.3 NMR Analysis of the Effects of Disulfide Bond Formation on ATF5 Structure

NMR experiments were performed to evaluate general differences in structure between the WT and C240A forms of ATF5 at higher resolution. The 2D ^{1}H-^{15}N HSQC experiment used here selectively detects all NH pairs in the protein and correlates each ^{1}H to the directly attached ^{15}N atom. The resultant spectrum is considered a "fingerprint" of the protein, because the chemical shifts depicted reflect the conformationally-weighted average of the environment experienced by the nuclei, which is defined by the overall protein fold. Consequently, the ^{1}H-^{15}N HSQC experiment can provide a quick indication of protein structure. A well-behaved globular protein will display good signal dispersion, typically ranging from 6 to 12 ppm on the ^{1}H-axis and 100-140 ppm on the ^{15}N-axis. This dispersion results from the presence of stable, structural features that confine the individual nuclei to unique chemical environments. More disordered proteins typically have peaks clustered more closely around random coil values (~8.3 ppm in ^{1}H). This is because, on average, the nuclei experience more similar chemical environments in solution than in the structured protein. The HSQC spectra of proteins composed of only a single helix, like the bZIP of ATF5, are difficult to interpret without full assignments and additional data, because in a standard alpha helix, all the amides are in similar environments, which can result in overlapping NMR signals. Typically, when bZIP proteins functionally dimerize they interact to form a two-stranded coiled-coil. Higher-order protein structure, such as coiled-coil, would place the NH nuclei in more distinct chemical environments, resulting in greater dispersion of the NMR

signal, as has been demonstrated experimentally for a number of bZIP proteins examined structurally using solution NMR.[13]

An overlay of the 2D HSQC spectrum of the C240A mutant onto the WT spectrum shows that the chemical shifts for both forms are largely clustered around 8.3 ppm in ^{1}H and only a few differences in resonance positions between the spectra are apparent (Figure3.7). This indicates the average structure of the two forms is quite similar and both proteins are conformationally labile. The limited chemical shift dispersion observed for both forms of ATF5 represents α-helix and/or random coil structure. The lack of dispersion in the NMR spectra further indicates coiled-coil structure is not developed, even in the obligate wild-type dimer. This is consistent with the CD spectra in which the ratio of 222 to 208 nm indicate a-helix is present but coiled-coil structure is not. The unique red peaks in the WT dimer likely correspond to residues adjacent to C240 involved in the helical region that are stabilized by disulfide bond formation. The NH signal generated by the side chains of Asn and Gln residues cluster in a different chemical shift range and account for the signal observed between 6.8 and 7.6 ppm on the ^{1}H -axis. There is little evidence for the existence of a coiled-coil structure for either form of ATF5, even at the high concentrations required for these studies. These data suggest that despite the presence of 23% helical content in the covalent dimer and its resistance to aggregation when the helix is maintained, ATF5 does not develop well-defined tertiary structure. Interestingly, the addition of 20% TFE alters the spectrum to only a small extent (data not shown),

suggesting TFE may temporarily stabilize residues already prone to helix formation rather than propagating helicity to other portions of the sequence.

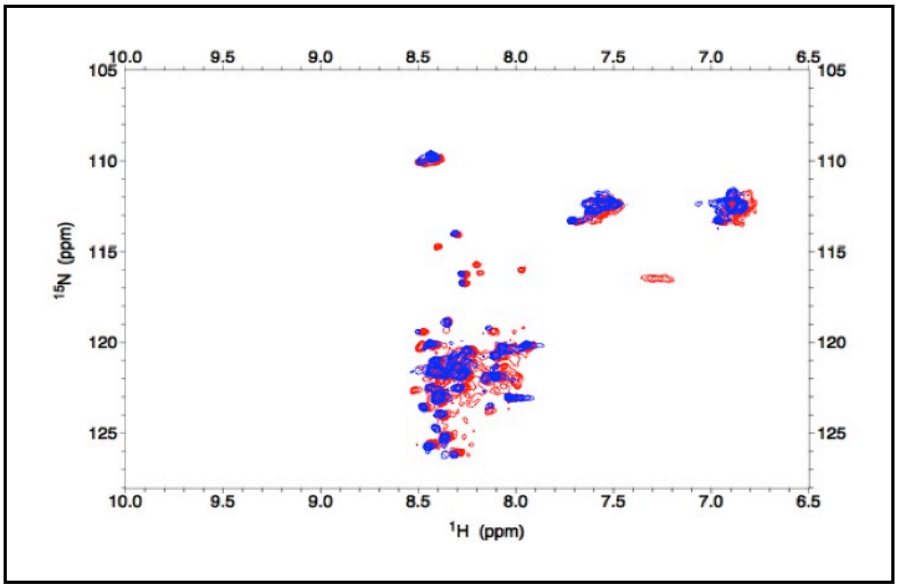

Figure 3.7 2D ^{1}H-^{15}N HSQC spectrum of 0.2 mM ^{15}N-labeled ATF5 WT dimer (red peaks) and C240A monomer (blue peaks) in 20 mM MES, pH 6.5.

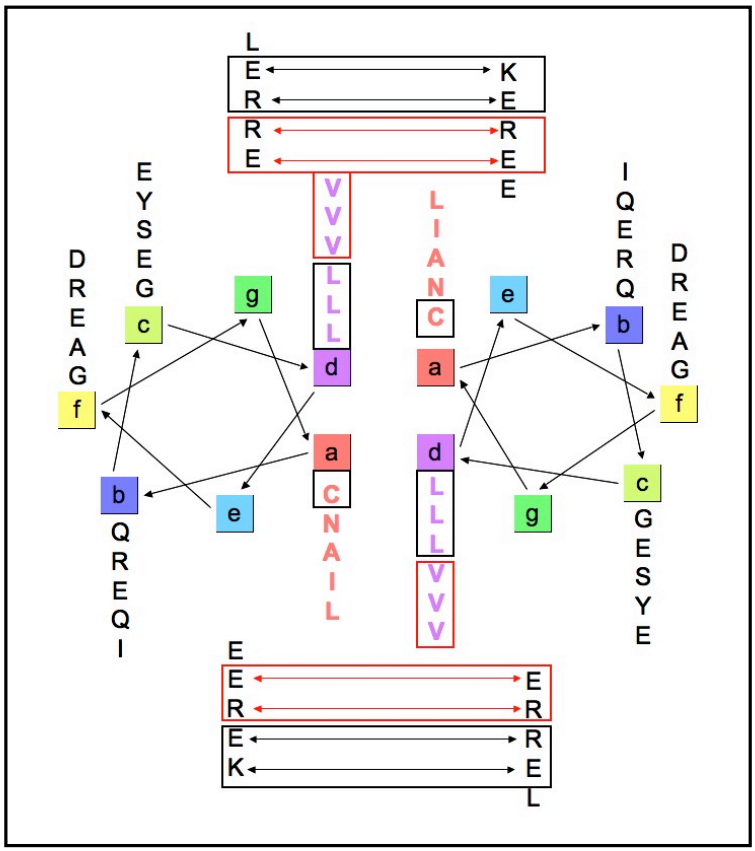

Figure 3.8 Helical wheel diagram of an ATF5 homodimer depicting standard coiled-coil interactions. Favorable interactions that would stabilize formation of a coiled-coil are delineated in black. Repulsive interactions that would be destabilizing are shown in red. The diagram reveals the lack of consensus among homotypic interactions to support coiled-coil stabilization.

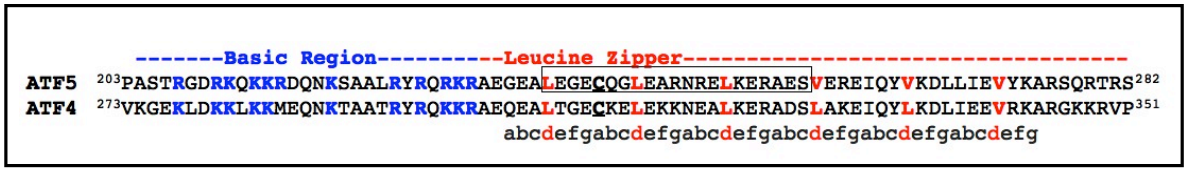

Figure 3.9 Sequence alignment of the bZIP domain of ATF4 and ATF5. The basic residues comprising the basic DNA binding domain are highlighted in blue. The leucine and valine residue located in the d-position of the leucine zipper heptad are highlighted in red. The unusual cysteine residue poised to form an intermolecular disulfide bond is underlined. The region of ATF5 most likely to form α-helical structure is boxed.

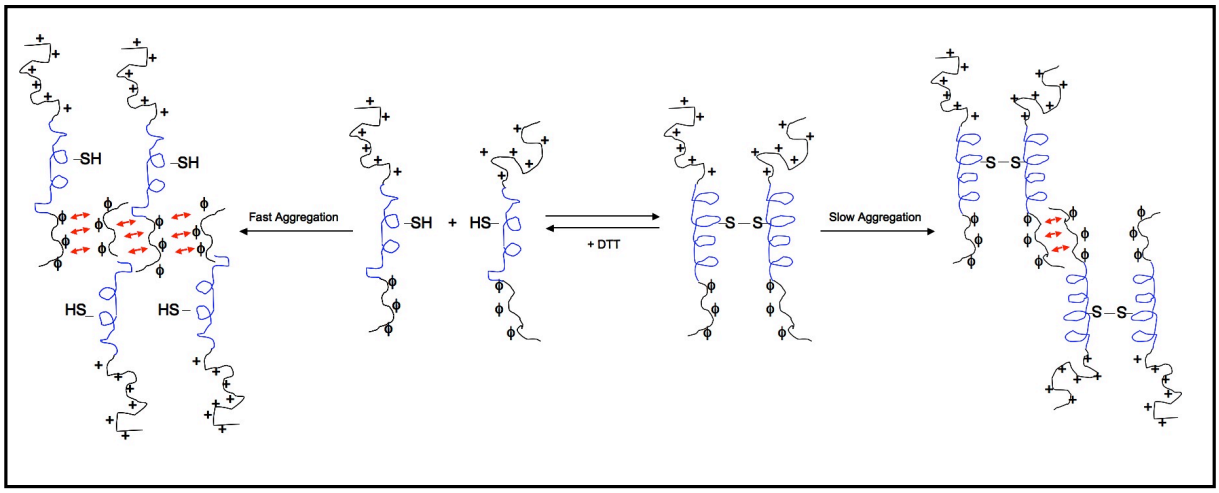

Figure 3.10 Aggregation model of ATF5. Slow aggregation of the cross-linked ATF5 dimer is facilitated by antiparallel association of the disordered valine-containing region (ϕ). Reduction of the disulfide bond creates the monomeric species, which displays decreased helicity and participates in fast aggregation through antiparallel alignment of the same valine-containing region. The N-terminal basic domain is illustrated by the addition of symbols indicating positive charge (+). The helical portion of the leucine zipper is depicted in blue. Note: A limited set of arrangements is shown here, but others are equally possible.

3.4 DISCUSSION

ATF5 belongs to the basic leucine zipper (bZIP) family of transcription factors. These proteins function as dimers that bind DNA and regulate transcription. The bZIP domain, present in all family members, contains a repeating pattern of leucine in every seventh position, which facilitates dimerization through formation of a parallel two-stranded coiled-coil.[14] Favorable hydrophobic interactions between residues in the *a* and *d* positions of the leucine zipper heptad in conjunction with attractive electrostatic interactions between residues in the *e* and *g* positions typically drive coiled-coil formation in bZIP proteins.[14] The basic regions of each monomer then bind DNA in what has been described as a "scissors-grip" model of binding.[15] bZIP proteins can either homodimerize or heterodimerize with other bZIP family members. Previous findings suggest that ATF5 forms a bZIP homodimer (Figure 3.8).[16] We have identified some favorable interactions that exist in our helical wheel diagram of the ATF5 homodimer that could assist in formation of a coiled-coil. For instance, the leucine residues in the *d* position (boxed in black) have a high propensity to adopt a helical conformation and are known to pack favorably in two-stranded coiled-coils.[17] We have shown that disulfide bond formation at the cysteine residue in the *a* position (boxed in black) retains helical structure, which could be a stabilizing force for a coiled-coil. Additionally, there are a few favorable electrostatic interactions that exist between residues in the *e* and *g* positions (depicted by black double-headed arrows and boxed) that could help stabilize a coiled-coil homodimer. However, there are also several unfavorable interactions present here that are not

found in other bZIP coiled-coils. The valine residues in the *d* position (outlined in red) have a much lower propensity to adopt helical structure.[17a] β-branched amino acids, like valine, are also not favored in the packing arrangement of two-stranded coiled-coils.[17b, c] A series of unfavorable electrostatic interactions exist between residues in the *e* and *g* positions (delineated by red double-headed arrows and boxed) that would further prevent formation of a stable coiled-coil. We have collected structural data on the bZIP domain of ATF5 at high concentrations using CD and solution NMR,[5] and our data indicate that the ATF5 protein does not form a stable coiled-coil in solution under physiological pH, ionic strength and reducing conditions.[5] The absence of valine residues and repulsive electrostatic interactions in the ATF4 and other bZIP coiled-coil structures implies that these features are responsible for the prevention of a stable coiled-coil for the ATF5 homodimer.

ATF5 shares 74% sequence homology with another bZIP protein family member, ATF4 (Figure 3.9). A crystal structure has been solved for ATF4, in which the protein participates in formation of a bZIP heterodimer with CCAAT enhancer binding protein β (C/EBPβ).[18] In this structure, ATF4 is shown to exist as a very straight α-helix, while C/EBPβ curves to wrap around ATF4 and create the coiled-coil. This is a unique feature of ATF4, as most bZIP proteins display more flexibility, with two bZIP monomers curving equally to wrap around each other.[19] The structure suggests that the rigidity of the ATF4 α-helix disfavors homodimerization, i.e. coiled-coil formation, and therefore must dimerize with a more flexible partner. The lack of stable homodimer formation under reducing conditions suggests ATF5 may

heterodimerize *in vivo* and that it is prone to aggregation *in vitro* because its hydrophobic side chains are exposed and unable to form native interactions without the alignment imposed by covalent attachment.

Alternatively, it has been suggested that intermolecular disulfide bond formation might facilitate formation of an ATF4 or ATF5 homodimer.[18] A unique cysteine residue is located at the center of the bZIP region of ATF4.[18] This analogous cysteine is positioned such that it would be at the interface of a coiled-coil dimer. ATF4 and ATF5 are the only bZIP proteins that contain a cysteine residue in this position, suggesting a functional role for this residue *in vivo*. We have collected NMR data in the presence and absence of this intermolecular disulfide bond for ATF5. The notable lack of dispersion in ^{1}H signal in the spectra of both the reduced monomer and oxidized dimer compared to spectra of other coiled-coil proteins indicates disulfide bond formation at this position does not confer formation of a stable coiled-coil (Figures 3.7 and 3.8). This disulfide bond does, however, play a role in retention of α-helical structure and physical stability of the protein to resist aggregation. If homodimerization has a biological function, the protein is likely to be redox regulated, as has been reported for other transcription factors.[20]

A naturally occurring mutant of apolipoprotein A-I (ApoA-I) known as the Milano variant (ApoA-I_M) was found to affect the stability of this protein in a parallel manner to that observed for ATF5. The Milano mutation introduces a non-native Cys residue that results in the formation of a disulfide-bonded homodimer. Cys substitution significantly increases both the percent helical content and the stability of

this protein at high concentration.[21] Cross-linking stabilizes a conformational change in ApoA-I_M that increases the helical content to match that of the native dimerized form, which is present at higher concentrations of protein. At lower protein concentration this mutation and other substitutions at the same site leads to diminished helicity and stability due to disruption of a salt bridge.[22] Addition of TFE to monomeric ApoA-I also increases its helical content to match that of the Milano form and concomitantly improves stability of the native protein against aggregation. ApoA-I is prone to form amyloid aggregates *in vivo*, whereas the disulfide-bound Milano form resists aggregation.[23] Interestingly, mutation at the same or adjacent positions in the protein that do not confer covalent attachment render ApoA-I more susceptible to aggregation.[23] Despite that ApoA-I and ATF5 are unrelated in sequence or function, these two systems display parallel behaviors with respect to the influence of helicity and disulfide cross-linked dimerization on stability and aggregation. Further investigation of ATF5 and comparison to ApoA-I may provide insights to better understand the basis of how helical structure and dynamics influence stability and aggregate formation.

The bZIP domain of ATF5 contains several valine residues instead of leucine residues at the C-terminal end of this region. While valine residues are commonly located in the *a*-position of the leucine zipper heptad, they rarely occupy the *d*-position. ATF5 is the only bZIP protein to possess three consecutive valines in this domain. ATF4 encodes only a single valine residue in the last *d*-position of its bZIP domain. It is not known what affect these β-branched amino acids would have on the

binding partner selectively of ATF5 *in vivo*, but they seem to prevent coiled-coil formation and are likely to be a key component of the protein's instability *in vitro*.

Our data show that the bZIP domain of ATF5 is only partially structured in solution, with its percent helicity increasing from 19.0% to 23.4% in the presence of the intermolecular disulfide bond. Using the DISEMBL^{TM} server we predicted the intrinsic disorder of this domain (data not shown).[24] The N-terminal basic domain (residues 203-238) and the end of the C-terminal leucine zipper domain (residues 275-282) are the most likely regions to be disordered. Additionally, a number of studies have shown that the helical propensity of leucine is much greater than valine.[17a] This suggests that the central portion of the leucine zipper of ATF5 is most likely to be helical. There are seventeen residues between the cysteine at position 240 and the valine at position 257, which make up the first half of the leucine zipper domain. If this region displayed 100% helicity and the rest of the protein was disordered, the predicted percent helicity would be 22.8%. This matches closely with the experimental values obtained for ATF5. An intermolecular disulfide bond positioned at cysteine 240 would likely help to stabilize the helical structure in this region by constraining the two chains and increasing intermolecular contact between them.

Computational methods for predicting protein aggregation have been developed, and we used several to examine ATF5. TANGO is a statistical mechanical model used to identify nucleation sites for aggregation based on the observations that aggregates often contain increased β-structure and the core regions of an aggregate

are completely buried, such that nucleating sequences will have their hydrogen bonding potential largely satisfied.[25] This program predicts that the region most likely to initiate aggregation in ATF5 lies between residues 260 and 276. This corresponds to the valine-containing portion of the leucine zipper, which is largely hydrophobic (Figure 3.9). AGGRESCAN is another sequence-based tool, which relies upon the aggregation-propensity for each of the individual amino acids to identify sequences likely to aggregate.[26]An evaluation of the bZIP domain of ATF5 identified the region contained between residues 265 and 273 as most likely to aggregate, which again, falls within the valine-containing portion of the leucine zipper. Based upon this information, it is likely that the structural features demonstrated to impart improved stability to ATF5 function protect the less-structured and aggregation-prone valine-containing portion of the protein from intermolecular association. A third program was used to evaluate the propensity of protein molecules to self-associate. Prediction of amyloid structure aggregation (PASTA) is based on the alignment observed in most cross-β structures.[27] PASTA predicted that it is unlikely for ATF5 to form cross-β fibrillar aggregates, which is consistent with our experimental observations of amorphous aggregates. The PASTA results, however, indicated that the antiparallel alignment of ATF5 in solution is energetically preferred over parallel alignment. This small preference could explain why aggregates form more rapidly between monomers. Covalent dimerization may not so much alter the structure of the monomer as constrain the Val-containing sequences in close proximity and provide an adjacent hydrophobic surface with which to interact.[28]

We have shown that the retention of α-helical structure through the formation of an intermolecular disulfide bond prevents rapid, thermally-induced protein aggregation. The addition of the co-solvent TFE also diminishes protein aggregation and induces α-helical structure, although the correlation between these two events is less clear since TFE affects multiple solution parameters that could also influence aggregation (i.e. dielectric constant, hydrogen-bonding pattern etc.). Based on the prediction results derived from first principles, it is most likely that the disordered regions of ATF5 play a role in initiating the process of aggregation rather than the uncoiled helical sequence. As electrostatic repulsion will dominate and prevent association between the basic regions, we propose that the hydrophobic residues present in the valine-containing portion of the protein are involved in initiating protein aggregation. Reduction of the disulfide bond at C240 would result in greater conformational freedom and exposure of hydrophobic residues in this region, which would increase protein aggregation. This explains why aggregation proceeds so much more quickly for the reduced WT and C240A monomer. The addition of TFE could potentially stabilize the helicity of the Leu-containing region, providing enhanced rigidity and reduced dynamic motion in the Val-containing region that would diminish solvent exposure and prevent self-association. Based upon our experimental evidence along with the results acquired using the computational tools identified above, we have proposed a model that depicts the aggregation mechanism of ATF5 (Figure 3.10). In using such a simple system to investigate protein aggregation, we have clear evidence to support the notion that increasing the

structural integrity of native α-helices improves protein physical stability. This conclusion has implications for handling other helical proteins *in vitro* and provides a basis for further investigating amorphous aggregation at higher resolution.

3.5 REFERENCES

1. (a) Hermeling, S.; Schellekens, H.; Maas, C.; Gebbink, M. F.; Crommelin, D. J.; Jiskoot, W., Antibody response to aggregated human interferon alpha2b in wild-type and transgenic immune tolerant mice depends on type and level of aggregation. *J Pharm Sci* **2006,** *95* (5), 1084-96; (b) Purohit, V. S.; Middaugh, C. R.; Balasubramanian, S. V., Influence of aggregation on immunogenicity of recombinant human Factor VIII in hemophilia A mice. *J Pharm Sci* **2006,** *95* (2), 358-71; (c) Schellekens, H., How to predict and prevent the immunogenicity of therapeutic proteins. *Biotechnol Annu Rev* **2008,** *14*, 191-202.
2. (a) Pecher, P.; Arnold, U., The effect of additional disulfide bonds on the stability and folding of ribonuclease A. *Biophys Chem* **2008**; (b) Kumar, S.; Ravi, V. K.; Swaminathan, R., How do surfactants and DTT affect the size, dynamics, activity and growth of soluble lysozyme aggregates? *Biochem J* **2008,** *415* (2), 275-88; (c) Knowles, T. P.; Zahn, R., Enhanced stability of human prion proteins with two disulfide bridges. *Biophys J* **2006,** *91* (4), 1494-500; (d) Lu, B. Y.; Chang, J. Y., A 3-disulfide mutant of mouse prion protein expression, oxidative folding, reductive unfolding, conformational stability, aggregation and isomerization. *Arch Biochem Biophys* **2007,** *460* (1), 75-84; (e) Huang, K.; Maiti, N. C.; Phillips, N. B.; Carey, P. R.; Weiss, M. A., Structure-specific effects of protein topology on cross-beta assembly: studies of insulin fibrillation. *Biochemistry* **2006,** *45* (34), 10278-93.
3. (a) Dima, R. I.; Thirumalai, D., Probing the instabilities in the dynamics of helical fragments from mouse PrPC. *Proc Natl Acad Sci U S A* **2004,** *101* (43), 15335-40; (b) Liu, W.; Prausnitz, J. M.; Blanch, H. W., Amyloid fibril formation by peptide LYS (11-36) in aqueous trifluoroethanol. *Biomacromolecules* **2004,** *5* (5), 1818-23.
4. (a) Kunjithapatham, R.; Oliva, F. Y.; Doshi, U.; Perez, M.; Avila, J.; Munoz, V., Role for the alpha-helix in aberrant protein aggregation. *Biochemistry* **2005,** *44* (1), 149-56; (b) Morgan, G. J.; Giannini, S.; Hounslow, A. M.; Craven, C. J.; Zerovnik, E.; Turk, V.; Waltho, J. P.; Staniforth, R. A., Exclusion of the native alpha-helix from the amyloid fibrils of a mixed alpha/beta protein. *J Mol Biol* **2008,** *375* (2), 487-98; (c) Watzlawik, J.; Skora, L.; Frense, D.; Griesinger, C.; Zweckstetter, M.; Schulz-Schaeffer, W. J.; Kramer, M. L., Prion protein helix1 promotes aggregation but is not converted into beta-sheet. *J Biol Chem* **2006,** *281* (40), 30242-50.
5. Ciaccio, N. A.; Moreno, M. L.; Bauer, R. L.; Laurence, J. S., High-yield expression in E. coli and refolding of the bZIP domain of activating transcription factor 5. *Protein Expr Purif* **2008,** *62* (2), 235-43.
6. Weiss, M. A., Thermal unfolding studies of a leucine zipper domain and its specific DNA complex: implications for scissor's grip recognition. *Biochemistry* **1990,** *29* (35), 8020-4.

7. Priddy, T. S.; Middaugh, C. R.; Carlson, G. M., Electrostatic changes in phosphorylase kinase induced by its obligatory allosteric activator Ca2+. *Protein Sci* **2007,** *16* (3), 517-27.
8. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **1995,** *6* (2), 135-40.
9. (a) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995,** *6* (3), 277-93; (b) T.D. Goddard, D. G. K.
10. Johnson, W. C., Analyzing protein circular dichroism spectra for accurate secondary structures. *Proteins* **1999,** *35* (3), 307-12.
11. (a) Manning, M. C.; Woody, R. W., Theoretical CD studies of polypeptide helices: examination of important electronic and geometric factors. *Biopolymers* **1991,** *31* (5), 569-86; (b) Zhou, N. E.; Kay, C. M.; Hodges, R. S., The role of interhelical ionic interactions in controlling protein folding and stability. De novo designed synthetic two-stranded alpha-helical coiled-coils. *J Mol Biol* **1994,** *237* (4), 500-12.
12. (a) Buck, M., Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q Rev Biophys* **1998,** *31* (3), 297-355; (b) Povey, J. F.; Smales, C. M.; Hassard, S. J.; Howard, M. J., Comparison of the effects of 2,2,2-trifluoroethanol on peptide and protein structure and function. *J Struct Biol* **2007,** *157* (2), 329-38.
13. (a) Atkinson, R. A.; Saudek, V.; Huggins, J. P.; Pelton, J. T., 1H NMR and circular dichroism studies of the N-terminal domain of cyclic GMP dependent protein kinase: a leucine/isoleucine zipper. *Biochemistry* **1991,** *30* (39), 9387-95; (b) Santiago-Rivera, Z. I.; Williams, J. S.; Gorenstein, D. G.; Andrisani, O. M., Bacterial expression and characterization of the CREB bZip module: circular dichroism and 2D 1H-NMR studies. *Protein Sci* **1993,** *2* (9), 1461-71; (c) Lumb, K. J.; Carr, C. M.; Kim, P. S., Subdomain folding of the coiled coil leucine zipper from the bZIP transcriptional activator GCN4. *Biochemistry* **1994,** *33* (23), 7361-7; (d) Ishigaki, T.; Ohki, I.; Utsunomiya-Tate, N.; Tate, S. I., Chimeric structural stabilities in the coiled-coil structure of the NECK domain in human lectin-like oxidized low-density lipoprotein receptor 1 (LOX-1). *J Biochem* **2007,** *141* (6), 855-66; (e) Nikolaev, Y.; Pervushin, K., NMR spin state exchange spectroscopy reveals equilibrium of two distinct conformations of leucine zipper GCN4 in solution. *J Am Chem Soc* **2007,** *129* (20), 6461-9.
14. O'Shea, E. K.; Rutkowski, R.; Kim, P. S., Evidence that the leucine zipper is a coiled coil. *Science* **1989,** *243* (4890), 538-42.
15. Vinson, C. R.; Sigler, P. B.; McKnight, S. L., Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* **1989,** *246* (4932), 911-6.
16. Peters, C. S.; Liang, X.; Li, S.; Kannan, S.; Peng, Y.; Taub, R.; Diamond, R. H., ATF-7, a novel bZIP protein, interacts with the PRL-1 protein-tyrosine phosphatase. *J Biol Chem* **2001,** *276* (17), 13718-26.

17. (a) Krittanai, C.; Johnson, W. C., Jr., The relative order of helical propensity of amino acids changes with solvent environment. *Proteins* **2000,** *39* (2), 132-41; (b) Moitra, J.; Szilak, L.; Krylov, D.; Vinson, C., Leucine is the most stabilizing aliphatic amino acid in the d position of a dimeric leucine zipper coiled coil. *Biochemistry* **1997,** *36* (41), 12567-73; (c) Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T., A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **1993,** *262* (5138), 1401-7.
18. Podust, L. M.; Krezel, A. M.; Kim, Y., Crystal structure of the CCAAT box/enhancer-binding protein beta activating transcription factor-4 basic leucine zipper heterodimer in the absence of DNA. *J Biol Chem* **2001,** *276* (1), 505-13.
19. (a) Junius, F. K.; O'Donoghue, S. I.; Nilges, M.; Weiss, A. S.; King, G. F., High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer. *J Biol Chem* **1996,** *271* (23), 13663-7; (b) Glover, J. N.; Harrison, S. C., Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **1995,** *373* (6511), 257-61; (c) Schumacher, M. A.; Goodman, R. H.; Brennan, R. G., The structure of a CREB bZIP.somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding. *J Biol Chem* **2000,** *275* (45), 35242-7; (d) Miller, M.; Shuman, J. D.; Sebastian, T.; Dauter, Z.; Johnson, P. F., Structural basis for DNA recognition by the basic region leucine zipper transcription factor CCAAT/enhancer-binding protein alpha. *J Biol Chem* **2003,** *278* (17), 15178-84; (e) Ellenberger, T. E.; Brandl, C. J.; Struhl, K.; Harrison, S. C., The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* **1992,** *71* (7), 1223-37.
20. Arrigo, A. P., Gene expression and the thiol redox state. *Free Radic Biol Med* **1999,** *27* (9-10), 936-44.
21. Calabresi, L.; Vecchio, G.; Longhi, R.; Gianazza, E.; Palm, G.; Wadensten, H.; Hammarstrom, A.; Olsson, A.; Karlstrom, A.; Sejlitz, T.; et al., Molecular characterization of native and recombinant apolipoprotein A-IMilano dimer. The introduction of an interchain disulfide bridge remarkably alters the physicochemical properties of apolipoprotein A-I. *J Biol Chem* **1994,** *269* (51), 32168-74.
22. Alexander, E. T.; Tanaka, M.; Kono, M.; Saito, H.; Rader, D. J.; Phillips, M. C., Structural and functional consequences of the milano mutation (R173C) in human apolipoprotein A-I. *J Lipid Res* **2009**.
23. Obici, L.; Franceschini, G.; Calabresi, L.; Giorgetti, S.; Stoppini, M.; Merlini, G.; Bellotti, V., Structure, function and amyloidogenic propensity of apolipoprotein A-I. *Amyloid* **2006,** *13* (4), 191-205.
24. (a) Linding, R.; Jensen, L. J.; Diella, F.; Bork, P.; Gibson, T. J.; Russell, R. B., Protein disorder prediction: implications for structural proteomics. *Structure* **2003,** *11* (11), 1453-9; (b) Iakoucheva, L. M.; Dunker, A. K., Order, disorder, and flexibility: prediction from protein sequence. *Structure* **2003,** *11* (11), 1316-7.
25. Fernandez-Escamilla, A. M.; Rousseau, F.; Schymkowitz, J.; Serrano, L., Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat Biotechnol* **2004,** *22* (10), 1302-6.

26. Conchillo-Sole, O.; de Groot, N. S.; Aviles, F. X.; Vendrell, J.; Daura, X.; Ventura, S., AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinformatics* **2007,** *8*, 65.
27. Trovato, A.; Seno, F.; Tosatto, S. C., The PASTA server for protein aggregation prediction. *Protein Eng Des Sel* **2007,** *20* (10), 521-3.
28. Ofran, Y.; Rost, B., Analysing six types of protein-protein interfaces. *J Mol Biol* **2003,** *325* (2), 377-87.

CHAPTER 4.

EFFECTS OF THE VALINE ZIPPER REGION ON THE AGGREGATION OF THE BZIP DOMAIN OF ACTIVATING TRANSCRIPTION FACTOR 5

4.1 INTRODUCTION

Protein aggregation is a major concern of the biopharmaceutical industry.[1] The aggregation of protein products may not only lead to a decrease in drug potency, but aggregated species can also display enhanced immunogenicity.[2] Consequently, controlling the formation of protein aggregates in a pharmaceutical formulation is of great importance. The difficulty in preventing aggregation partially stems from our lack of understanding of the mechanism by which aggregates form. In particular, we do not understand the role that protein structure plays in this process. Increasing evidence suggests that partially folded protein intermediates critically influence the aggregation pathway; however, this relationship remains ill-defined.[3]

Two general categories of protein aggregates have been reported in the literature: amyloid fibrils and amorphous aggregates.[4] Fibrillar aggregates have received much attention due to their association with a growing number of serious disease states for which limited treatment options exist.[4c, 5] They are characterized by a cross-beta structure, in which the beta strands align in a highly organized format along the long-axis of the fibril.[4c, 5a, 6] Amorphous aggregates are often associated with recombinant protein production and purification *in vitro;* however increasing evidence indicates that a number of proteins can develop into both fibrillar and

amorphous aggregates *in vitro*.[7] The nature of the aggregated species is highly dependent upon protein sequence and solution conditions that control protein conformation.[7] Amorphous aggregates are so named because unlike fibrillar aggregates, they do not display long-range order.[3h, 4a, 4d] However, these aggregates do possess structure and, typically, an increase in intermolecular beta-sheet content is observed compared to the native state.[3h, 4a, 8] This is not true in all cases, and the development of alternative structures, including alpha helix, has been reported in aggregates.[3g, 9] It is currently unclear how the amino acid sequence and higher order structure directs aggregate assembly and organization.

We have previously reported the production of the basic leucine zipper (bZIP) domain of Activating Transcription Factor 5 (ATF5), a notable protein target for treatment of glioblastoma.[10] Because of its propensity to aggregate *in vitro*, we have used ATF5 as a model system to investigate protein aggregation.[11] Our previous data indicate that disulfide bond formation prevents the formation of large amorphous protein aggregates through improved retention of alpha-helical structure.[11] Computational analysis predicted that the C-terminal region, which contains a series of three valine residues (V257, V264 and V271) in lieu of the conserved leucine residues typically observed in these positions, most likely facilitates aggregation.[11] We suggested that disulfide bond formation minimized solution exposure of this C-terminal region through improved structural retention, resulting in diminished protein aggregation.[11] The work presented here further characterizes the contribution of this C-terminal region to the aggregation of ATF5.

A mutant form of the ATF5 protein that lacks the C-terminal region (residues 257-282) was created using site-directed mutagenesis. This variant (V257STOP) was created via the insertion of a stop codon immediately preceding the first valine residue (V257) in this region. The structure of the purified mutant form of ATF5 was evaluated using circular dichroism spectroscopy (CD), Fourier transform infrared (FTIR) spectroscopy and two-dimensional nuclear magnetic resonance (NMR) spectroscopy. The results were compared with that of the wild-type (WT) form of ATF5 to determine the effect of C-terminal truncation on protein structure. The extent of thermal aggregation of both proteins was monitored using static light scattering (SLS) analyses and dynamic light scattering (DLS) analyses. Structural changes that occurred during the aggregation process were evaluated using FTIR analysis. These data were collectively used to compare the formation and growth of protein aggregates of both the WT and V257STOP forms of ATF5 and identify correlations between protein structure and aggregate development. Our data indicate that removal of the C-terminal region of ATF5 changes the structure of the protein and the mechanism of thermally induced protein self-association.

4.2 MATERIALS AND METHODS

4.2.1 Protein Expression and Purification

The cDNA of ATF5 was obtained through ATCC (MGC-842), and the bZIP domain was PCR amplified and inserted into a Novagen (San Diego, CA) pET-42b vector as described previously.[10a] The V257STOP truncated mutant was generated

using site-directed mutagenesis to insert a stop codon at the valine in position 257. Mutagenesis was performed by GenScript USA (Piscataway, NJ). Plasmids were transformed into Novagen BL21(DE3) *E. Coli* for expression. Cell growth, lysis and purification were performed using the procedure described previously with the following modification.[10a] Unlike wild type, the truncated mutant was present largely in the soluble fraction following cell lysis and centrifugation. This fraction was heated at 60 °C for 1 hour to precipitate the majority of cellular proteins. The mutant remained soluble under these conditions. To ensure that any aggregates formed during this purification step were eliminated the protein was denatured and subsequently subjected to the same refolding procedure as previously described for the WT.[10a] Protein concentrations were determined using a standard Bradford assay. The purity of ATF5 was determined using densitometry of Coomassie stained SDS-PAGE gels.[10a] This method was also used to confirm intermolecular disulfide bond formation using non-reducing conditions. Isotopic labeling for two-dimensional NMR was accomplished during cellular expression with the use of ^{15}N-ammonium chloride (>99% N-15; Isotec, St. Louis, MO).

4.2.2 Circular Dichroism Spectroscopy

Circular dichroism (CD) spectra were acquired with an Applied Photophysics Chirascan (Leatherhead, UK) equipped with a Peltier-type temperature controller and a four-position sample holder. ATF5 WT and V257STOP samples were prepared at a final protein concentration of 10 μM in 20 mM MES buffer at pH 6.0. Scans were performed at 10 °C in a 0.1-cm path length cell from 260 to 190 nm. The selected

bandwidth was 4 nm, the data pitch was 1 nm and the sampling time was 0.5 sec. Samples were prepared and analyzed in duplicate. Raw data were converted into units of molar ellipticity using the standard formula, $[\theta] = \theta*M*10/C*l$, where $[\theta]$ is the molar ellipticity, θ is the ellipticy in mdeg, M is the molecular weight in mg/mmol, C is the concentration in mg/mL and l is the path length in cm. Multiplying by 10 facilitates the conversion from millimoles to decimoles, resulting in the final units of $deg*cm^{2}*dmol^{-1}$. The resulting average and standard deviations were plotted.

4.2.3 Fourier Transform Infrared Spectroscopy

Absorption spectra (256 scans) were collected at a 5 mg/mL concentration from 900 to 4000 cm^{-1} at a resolution of 4 cm^{-1} on a Bruker (Billerica, MA) Tensor 27 Fourier transform infrared spectrophotometer instrument using a Harrick Scientific Products (Pleasentville, NY) BioATR Cell II sample apparatus and a Thermo Scientific (Waltham, MA) Haake DC30-K20 refrigerated circulator bath under constant nitrogen purge. The spectrophotometer was equipped with a KBr beam splitter and a liquid-nitrogen-cooled mercury cadmium telluride (MCT) detector. Reference spectra consisting of 20 mM MES buffer pH 6.0 with or without 10 mM dithiothreitol (DTT) were collected prior to each sample solution and subtracted as background. Resultant spectra were first processed using OPUS 6.5 software (Bruker Optics). A correction was performed to minimize absorption interference from atmospheric water and carbon dioxide. The minimum and maximum values within the amide I region (1600-1700 cm^{-1}) were normalized. Data were then further

processed using GRAMS/AI 8.0 spectroscopy software (Thermo Fisher Scientific). A baseline correction was performed, and the data were smoothed using a five-point Savitsky-Golay function. Fourier self-deconvolution was performed and the data were fitted by inserting individual Gausssian bands to match the second derivative peak position. Secondary structure assignments were made based upon reference spectra reported in the literature.[12]

The temperature was increased from 10 °C to 90 °C to facilitate thermal analysis of stability. Scans (256) were collected every 2.5 °C at a ramping rate of 2.5 °C per minute. A reference melt consisting of 20 mM MES buffer pH 6.0 with or without 10 mM dithiothreitol (DTT) was collected immediately prior to sample analysis and subtracted as background. The resultant spectra were processed using OPUS 6.5 software. A correction was performed to minimize absorption interference from atmospheric water and carbon dioxide. The minimum and maximum values within the amide I region (1600-1700 cm^{-1}) were normalized. The absorbance spectra in the Amide I region were plotted as a function of temperature.

4.2.4 Nuclear Magnetic Resonance Spectroscopy

Two-dimensional ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra were recorded at 25 °C using a Bruker AVANCE 800 MHz spectrometer equipped with a triple-resonance CRYO-probe with pulse field gradients. Samples were prepared at a concentration of 1 mM in 20 mM MES buffer pH 6.0, containing 5% D_2O. Water suppression was accomplished using flip-back pulses. Data were acquired in 8 scans with 1024 points in ^{1}H and 128* increments in ^{15}N. ^{1}H chemical

shifts were referenced with respect to an external DSS standard in D_2O.[13] Indirect referencing relative to ^1H was determined for ^{15}N, assuming a ratio of $^{15}N/^1H$ = 0.101329118. Data were processed using NMRPipe and Sparky software.[14]

4.2.5 Static Light Scattering

Static light scattering (SLS) data were acquired with a Photon Technology International (PTI) spectrofluorometer (Lawrenceville, NJ) equipped with a Peltier-type temperature controller and a four-position sample holder. Samples were prepared at a final protein concentration of 100 μM in 20 mM MES buffer at pH 6.0 in the presence or absence of 10 mM DTT. The intensity of scattered light was measured as a function of temperature and detected at an angle of 90° to the light source by a photomultiplier tube. An arc lamp white light source was used in all cases. The excitation wavelength was set at 350 nm. The emission wavelength range was set from 300 nm to 400 nm. Spectra were obtained following a 5-min equilibration period at each temperature. Data were collected every 2.5 °C between the temperature range of 4 to 81.5 °C. The background was subtracted from each data point based on a blank containing the corresponding buffer solution. Two samples were prepared and analyzed for each set of solution conditions. The resulting average and standard deviations were plotted.

4.2.6 Dynamic Light Scattering

The average hydrodynamic radius was monitored as a function of temperature using a Wyatt Technology Corporation (Santa Barbara, CA) DynaPro Plate Reader Plus instrument. ATF5 WT and V257STOP samples were prepared in 20 mM MES

buffer pH 6.0 at a final concentration of 100 μM in the presence or absence of 10 mM DTT. Samples containing the isolated V257STOP aggregate or monomer species were prepared at a 40 μM concentration in phosphate buffered saline (PBS) at pH 7.4. All samples were filtered before analysis using a 0.2 μm Anotop 10 Whatman (Piscataway, NJ) 10 mm inorganic syringe filter. A 384-well Corning (Corning, NY) polystyrene plate was used for these experiments. Data were obtained in three consecutive acquisition periods, each 15 seconds in length, and averaged. Acquisitions were obtained every 2.5 °C over a temperature range of 10 °C to 67.5 °C with a ramping rate of 1 °C/min. A constant nitrogen purge at 60 cc/min and 60 psi was performed to prevent condensation. Average hydrodynamic radii and polydispersity indices were calculated using cumulant analysis.[15] Two samples were prepared and analyzed for each set of solution conditions. The resulting average and standard deviations were plotted. Particle distribution was evaluated using a non-negative least squares regularization method of analysis.[16] A representative data set for each form of ATF5 was plotted.

4.3 RESULTS

4.3.1 Effects of C-terminal Truncation on ATF5 Structure

Circular dichroism (CD) spectroscopy was performed to evaluate differences in secondary structure between the wild-type (WT) and the truncated mutant (V257STOP) forms of ATF5. The mutant protein displays a CD absorption spectrum that is distinct from the WT form (Figure 4.1). The WT possesses a double-minima

absorption pattern with negative peaks at 222 nm and 208 nm. This pattern is indicative of α-helical protein structure.[17] Given that the extended C-terminal portion of the protein was removed to make the V257STOP protein, one would expect that the remaining leucine zipper region would retain helicity, resulting in a proportional increase in overall helical signal. In fact, the V257STOP mutant displays a single absorption minimum at 209 nm and a broad shoulder between 260 nm and 230 nm, which is not consistent with alpha-helical structure. Negative absorption at 209 nm has been observed for proteins possessing β-structure.[18] Additionally, it has been reported that certain types of β-bend can give negative absorption at 208 nm.[19] The absorption observed between 260 nm and 230 nm could be attributed to β-turn structure, which has been demonstrated to display CD signal in this region.[20]

Further experimental analysis of secondary structure was performed using Fourier transform infrared spectroscopy (FTIR) to better understand these structural differences. Deconvolution of FTIR absorption in the Amide I region for the WT and V257STOP forms of ATF5 also indicates that the two proteins possess distinctly different secondary structure (Figures 4.2 and 4.3). The absorption peaks observed here are broader than those typically observed for a well-structured system.[21] The data for both forms of ATF5 are suggestive of a dynamic or partially folded system. The WT form displays absorption peaks at 1621 cm^{-1}, 1650 cm^{-1}, 1676 cm^{-1} and 1688 cm^{-1}. The absorption band at 1621 cm^{-1} corresponds to β-sheet structure.[12a, 22] The dominant absorption peak observed at 1650 cm^{-1} corresponds to α-helical structure.[12a, 22] The absorption peak observed at 1676 cm^{-1} could represent β-sheet or

β-turn structure.[12a, 22] Strong β-sheet absorption at low wavenumbers in the Amide I region (1620 cm^{-1} to 1639 cm^{-1}) is often accompanied by weaker absorption at higher wavenumbers (1670 cm^{-1} to 1680 cm^{-1}).[21c, 23] It is likely, based upon the absorption observed at 1621 cm^{-1}, that the peak at 1676 cm^{-1} primarily represents β-sheet structure. Lastly, the absorption peak at 1688 cm^{-1} is reflective of β-turn structure.[12a, 22]

The V257STOP mutant form of ATF5 displays broad absorption peaks at 1644 cm^{-1} and 1682 cm^{-1}. While the fitting procedure resulted in the placement of two broad peaks for the mutant form, the second derivative data suggest the presence of two states, one of which may be α-helical (1650 cm^{-1}) and the other unordered structure (1641 cm^{-1}).[12a, 22, 24] The broad absorption peak at 1682 cm^{-1} could represent β-sheet and/or β-turn structure.[12a, 22] The absorption band observed at 1601 cm^{-1} likely corresponds to side chain vibrations and is not predictive of secondary structure.[25]

Additional studies were performed using two-dimensional NMR to compare chemical shift signal dispersion for the two forms of ATF5 (Figure 4.4). The ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) experiment used here gives an indication of overall protein fold and structure based upon signal dispersion. This experiment can also provide general insight into protein dynamics, which are reflected in and often dictate peak shape and intensity. Both forms of ATF5 display NH signals largely between 8.0 and 8.5 ppm on the ^{1}H-axis (Figure 5). The signal observed between 6.5 ppm and 7.5 ppm arises from NH_2 groups in the Asn and Gln

side chains, which typically have chemical shifts distinct from those of the NH groups in the amide backbone. While some overlap in backbone signal is observed between the two forms of ATF5, their chemical shifts are largely distinct, which reaffirms the substantial differences in their overall structure and/or conformational dynamics. A qualitative comparison of peak shape between the two spectra suggests that both forms of ATF5 possess well-defined structural regions as well as more conformationally labile regions. This is illustrated by the presence of both sharp, circular cross-peaks that correspond to well-defined structure and broad, misshapen cross-peaks, which are indicative of conformational exchange. The NMR signal intensity is proportional to the concentration of protein and approximately reflects the amount of protein that is present in the monomeric/dimeric state. Large molecular weight species are not observed in the standard HSQC experiment, in this case self-associated species composed of more than approximately 8-10 protomers. The signal to noise ratio observed in these experiments was 1.8-fold greater for the WT form than for the truncated mutant. This suggests that roughly 45% of the mutant signal was lost due to protein aggregation or precipitation. No visible precipitation was observed in the NMR sample tube, which indicates that the aggregate species must be soluble in nature.

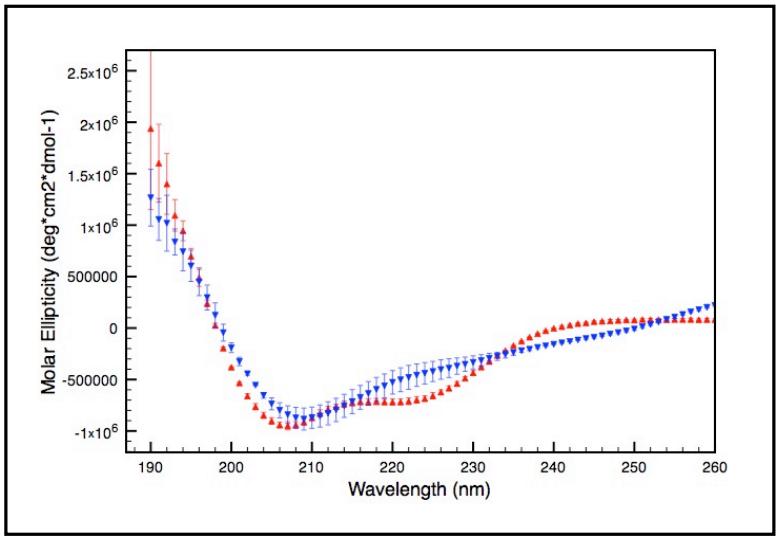

Figure 4.1 CD absorption spectra of the WT (red) and the V256STOP mutant (blue) forms of ATF5 in 20 mM MES pH 6.0 at 4 °C.

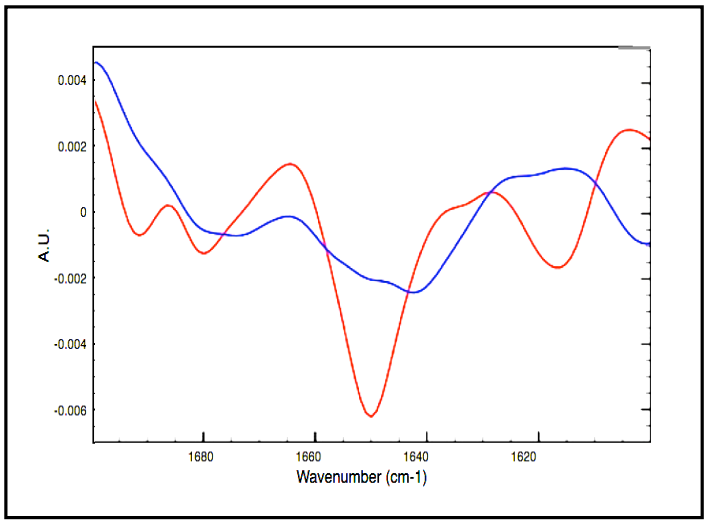

Figure 4.2 Second derivative of FTIR spectra for the WT (red) and V257STOP mutant (blue) forms of ATF5 at 10 °C.

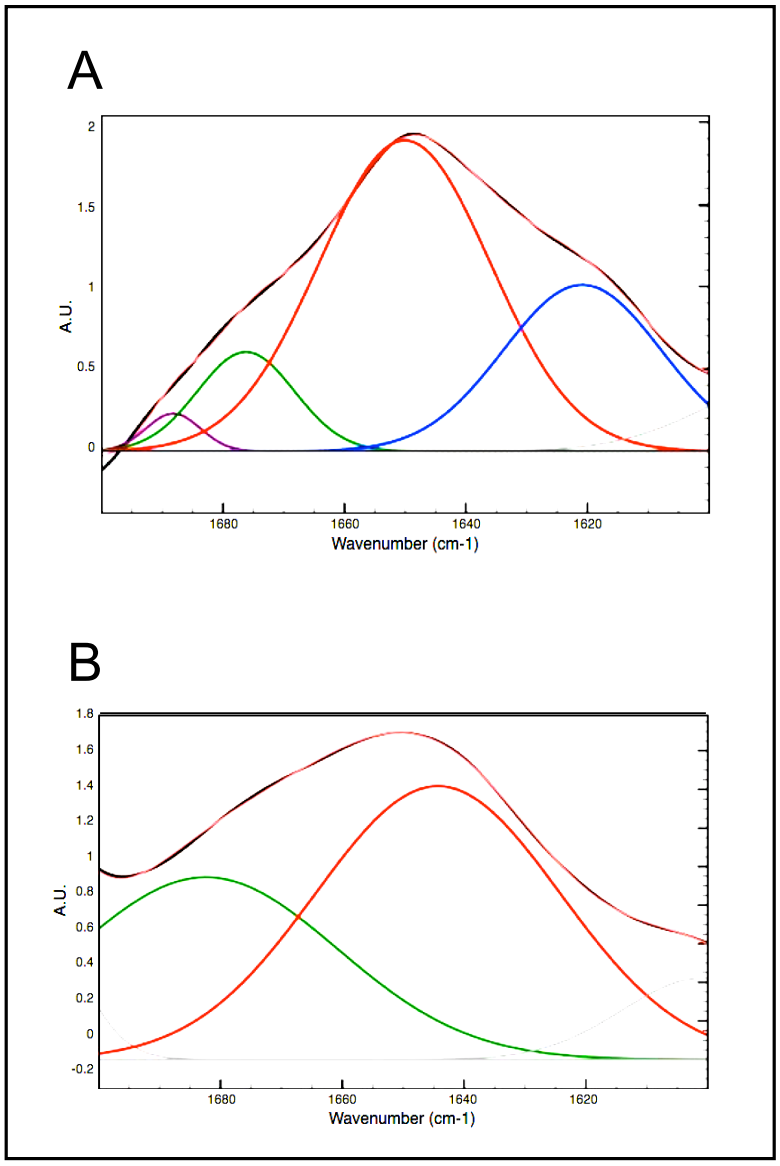

Figure 4.3 Deconvolution of the FTIR absorption spectra for the WT (A) and V257STOP mutant (B) forms of ATF5 at 10 °C. Absorption in the Amide I (1600 - 1700 cm^{-1}) and Amide II (1500 -1600 cm^{-1}) regions is plotted. The black outline denotes the parent absorption spectrum. The pink overlay indicates the fitted spectrum. Individual absorption peaks in the Amide I region are color-coded according the structure they represent (blue = extended structure; red = alpha helix/ unordered; green = extended structure/ beta turn; purple = beta turn).

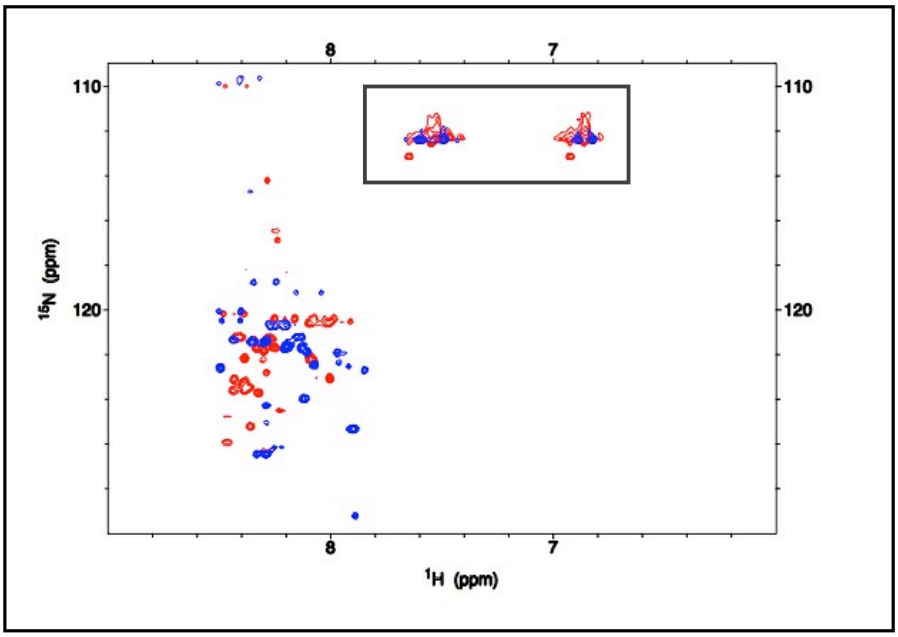

Figure 4.4 2D ^{1}H-^{15}N HSQC spectrum of 1 mM ^{15}N-labeled ATF5 WT (red peaks) and V257STOP mutant (blue peaks) in 20 mM MES buffer, pH 6.0. Peaks corresponding to NH_2 side chains rather than NH backbone groups are boxed.

4.3.2 Analysis of Change in Structure Observed During ATF5 Aggregation

In order to investigate the structural changes that accompany aggregate growth and development, FTIR analysis was employed. This technique permits the analysis of sample both in solution and in the solid state, thereby allowing us to monitor structural changes that occur as the protein precipitates. In these experiments, FTIR absorption was monitored for both forms of ATF5, each under reducing and non-reducing conditions, as a function of temperature (Figure 4.5). The WT form begins to change structure between 10 °C and 25 °C. This transition is characterized by a loss in α-helical structure at 1650 cm^{-1} and an increase in β-sheet structure at 1620 cm^{-1}. Increased absorption at ~1620 cm^{-1} is typically observed during protein aggregation and is indicative of intermolecular β-sheet structure.[21a, 21c, 23] In the presence of DTT, two isobestic points are observed for the WT melt (~1630 cm^{-1} and 1670 cm^{-1}). This suggests that the loss of helical structure is directly linked to a gain in structure at lower and higher wavenumber. In contrast, a more complex structural transition is observed for the WT form in the absence of DTT. In this case, an initial loss of signal is observed between 10 °C and 25 °C at lower wavenumber, followed by a subsequent increase in β-sheet signal at 1620 cm^{-1}.

Interestingly, the V257STOP mutant form of ATF5 does not change structure until the temperature rises above 40 °C. A decrease in unordered absorption at 1644 cm^{-1} and an increase in β-sheet absorption at 1620 cm^{-1} are then observed. A loss in absorption at higher wavenumber (between 1670 cm^{-1} to 1700 cm^{-1}) is also observed as a function of temperature. This suggests a reduction in β-sheet or β-turn structure.

This loss continues until the temperature exceeds 70 °C, where a slight increase in absorption in this region is then observed. The structural transitions observed for the V257STOP mutant are largely unaffected by DTT addition.

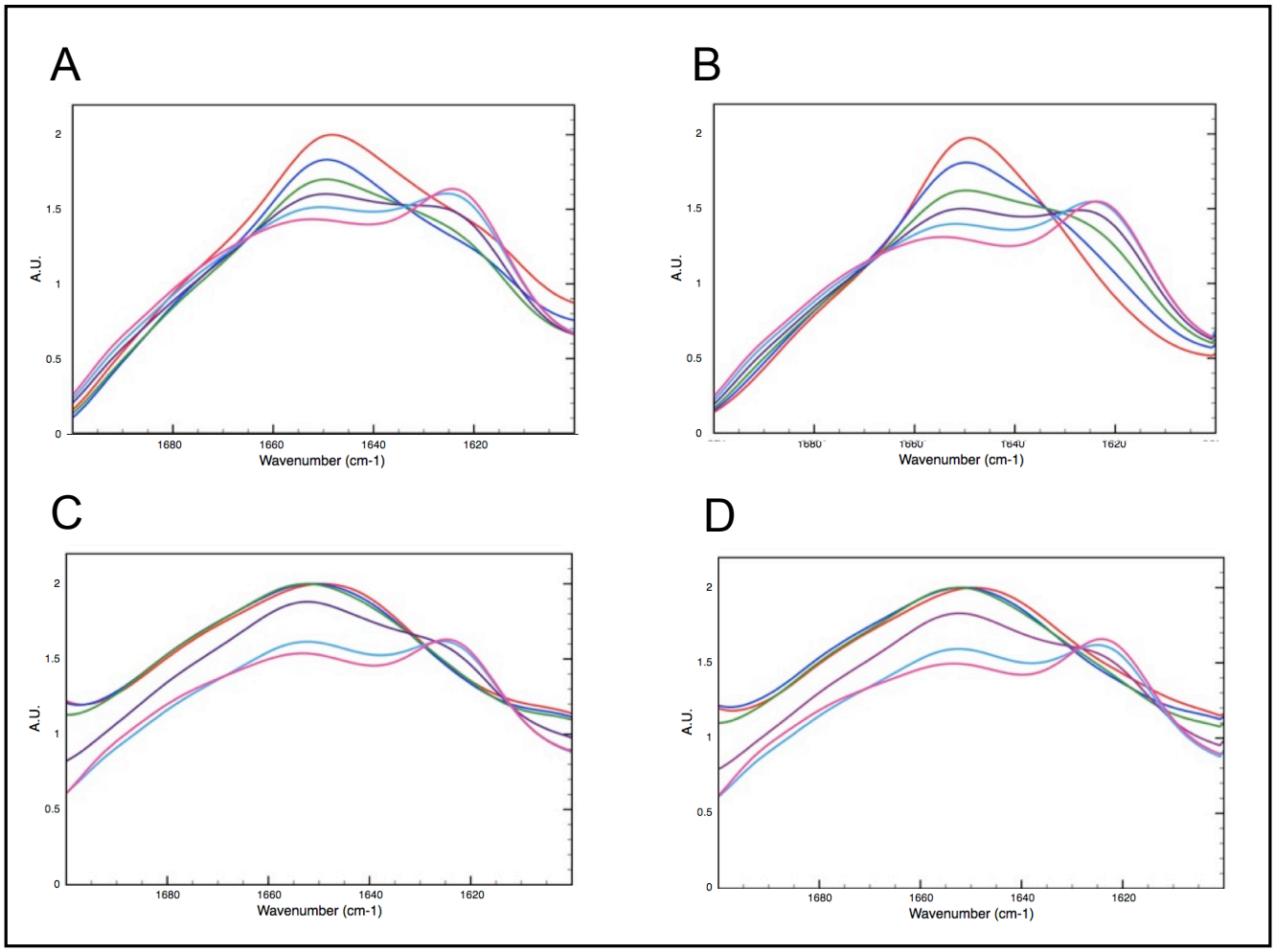

Figure 4.5 FTIR absorption in the Amide I region is shown for the WT and V257STOP forms both with and without 10 mM DTT at varying temperature. Panels A and B show data collected on the WT form in the absence (A) and presence (B) of DTT. Panels C and D show data collected on the V257STOP mutant form in the absence (C) and presence (D) of DTT. The data shown was collected at 10 °C (red), 25 °C (blue), 40 °C (green), 55 °C (purple), 70 °C (light blue) and 90 °C (pink).

4.3.3 Effects of C-terminal Truncation on ATF5 Aggregation

Static light scattering (SLS) analysis was performed to compare the growth of thermally induced protein aggregates between the WT and V257STOP forms of ATF5. In these experiments the average intensity of scattered light was measured in counts per second as a function of temperature under both reducing and non-reducing conditions (Figure 4.6). The onset of aggregation occurs much earlier for the WT form of ATF5 than for the V257STOP mutant. In the presence of DTT, WT aggregation is characterized by a sharp increase in scattering intensity between 25 °C and 30 °C. The scattering intensity eventually levels off and starts to decline at high temperature (between 70°C and 80 °C) due to the settling of large particles out of solution. This was confirmed by visual inspection of samples following the melt. In the absence of DTT, the WT onset of aggregation is shifted to higher temperature. An increase in scattering intensity is not observed until almost 40 °C. Additionally, particulate growth appears to be more gradual under these conditions.

Aggregation onset for the V257STOP mutant form of ATF5 is not observed until the temperature exceeds 50 °C. Aggregation of this species is depicted by a gradual increase in scattering intensity observed between 60 °C and 70 °C, followed by a sharper increase in scattering intensity between 70 °C and 80 °C. Unlike WT, the aggregation of this form of ATF5 is unaffected by the addition of DTT. Further analysis was performed using dynamic light scattering (DLS) analysis to evaluate particle size distribution as a function of temperature.

Dynamic light scattering (DLS) uses autocorrelation analysis to measure the time-dependent fluctuations in light scattering intensity. The autocorrelation function reflects particle diffusion in solution, which is directly related to particle size. In these experiments the autocorrelation function was acquired at increasing temperatures for the two forms of ATF5 under both reducing and non-reducing solution conditions. Cumulant analysis provided an average hydrodynamic radius (R_H) value and polydispersity index for the particles in solution (Figure 4.7).[15]

The WT form of ATF5 displays an average R_H of 10 nm at 10 °C under both reducing and non-reducing solution conditions. This average value is unchanged at low temperatures but begins to gradually increase in size between 25 °C and 30 °C. While this trend is the same for the WT protein under both reducing and non-reducing conditions, the average value becomes slightly larger for the reduced sample as a function of temperature. In contrast, the V257STOP mutant form of ATF5 maintains an average R_H of approximately 22 nm at nearly all temperatures measured. A slight increase in average R_H occurs for the V257STOP form as the temperature exceeds 60 °C. (Analysis above 65 °C was not performed, because the polystyrene plates used in these experiments cannot be used above this temperature.) Again, there is no distinction between data collected under reducing and non-reducing solution conditions for the V257STOP mutant for ATF5.

While this data provides useful information about trends associated with increasing temperature, the cumulant analysis provides an average R_H value. The polydispersity index reported with the average value provides an estimation of

distribution about the mean. The polydispersity values reported for both the WT and V257STOP systems suggest multimodal sample populations (Figure 4.8). Additionally, the average values reported here are much larger than what would be expected for ATF5 monomer or dimer. For example an IgG mononclonal antibody with an estimated molecular weight of 150 kDa typically displays an average hydrodynamic radius of approximately 5 nm to 8 nm[26]. This suggests that self-associated protein species are present in both the WT and V257STOP samples. Solution polydispersity decreases for both forms of ATF5 as particle size increases (Figure 8). This could be related to the development of micron-size protein aggregates that exceed the range of DLS detection (approximately 0.5 nm to 1 μm). As protein precipitates from solution, the polydispersity of the remaining soluble species may be lowered.

Regularization analysis is an alternative method for analyzing DLS data that makes no assumptions about sample homogeneity and permits characterization of multimodal sample populations.[16] This analysis was performed on the dynamic light scattering data collected for both forms of ATF5 between 10 °C and 65 °C. The data is plotted as a distribution illustrating the relative intensity of light scattered as a function of particle size in solution (Figure 4.9). The data for both forms of ATF5 indicate heterogeneous, multimodal sample populations. The WT sample contains two sample populations. A smaller species ranging from approximately 2 nm to 4 nm in radii that likely represents monomeric or dimeric protein, and a larger population that ranges from 8 nm to 35 nm in size and consists of aggregated protein species.

The V257STOP sample also contains two sample populations. A smaller species ranging again from approximately 2 nm to 4 nm in radii that reflects monomeric or dimeric protein and a larger population ranging from 10 nm to 70 nm in size that is consistent with aggregated species.

The range in size observed for protein monomer/dimer is consistent with the NMR data, which indicates a partially folded and highly dynamic system. The range observed for the aggregated species indicates the existence of particles of varying sizes. Interestingly, the V257STOP mutant form consistently displays a much broader size range of aggregated protein than the WT form. Moreover, this distribution remains relatively constant across the temperatures measured. This in contrast to the WT, whose aggregated population steadily increases in size as a function of temperature. This suggests that aggregation of the WT form occurs through the growth or increase in size of soluble protein aggregates in the 10 to 35 nanometer size range.

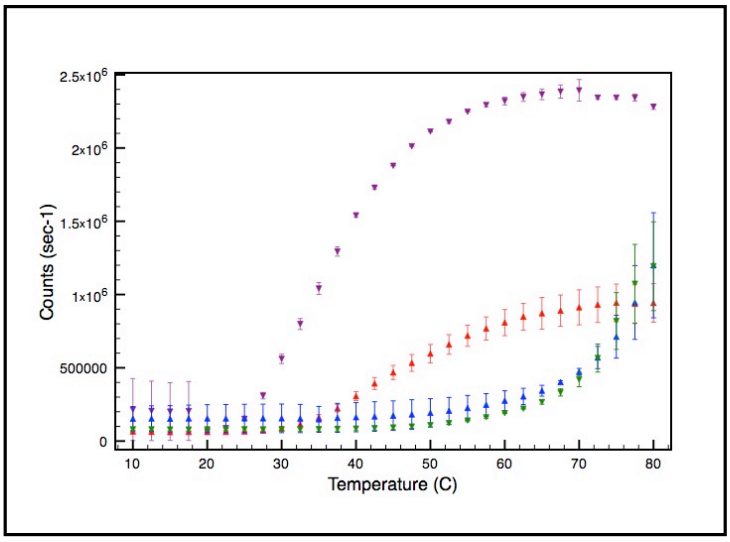

Figure 4.6 SLS analysis of the WT and V257STOP truncated forms of ATF5 in both the absence and presence of 10 mM DTT. The intensity of scattered light is plotted as a function of temperature. Data for the WT form in the absence and presence of DTT is illustrated in red and purple, respectively. Data for the V257STOP mutant in the absence and presence DTT is illustrated in blue and green, respectively.

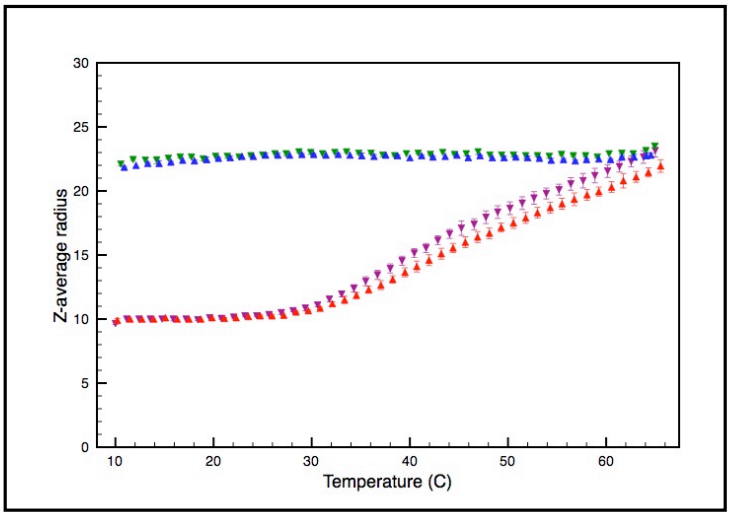

Figure 4.7 DLS analysis of the WT and V257STOP truncated forms of ATF5 in both the absence and presence of 10 mM DTT. The average hydrodynamic radii are plotted as a function of temperature. Data for the WT form in the absence and presence of DTT is illustrated in red and purple, respectively. Data for the V257STOP mutant in the absence and presence DTT is illustrated in blue and green, respectively.

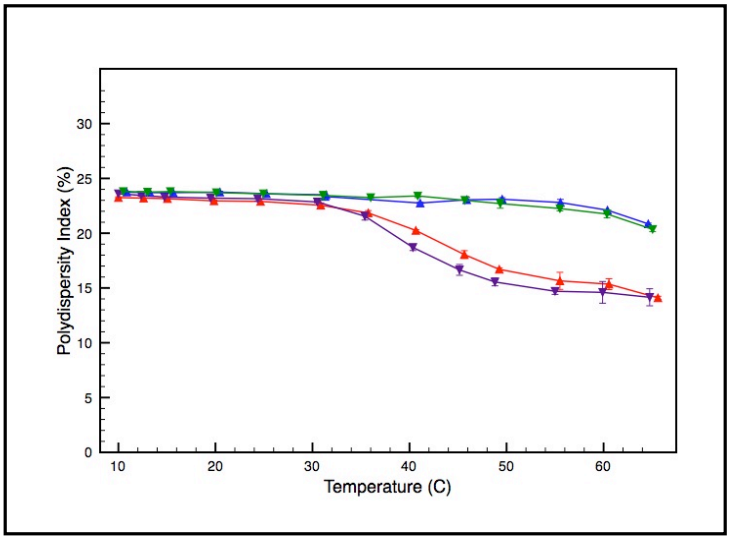

Figure 4.8 DLS analysis of the WT and V257STOP truncated forms of ATF5 in both the absence and presence of 10 mM DTT. The polydispersity index is plotted as a function of temperature. Data for the WT form in the absence and presence of DTT is illustrated in red and purple, respectively. Data for the V257STOP mutant in the absence and presence DTT is illustrated in blue and green, respectively.

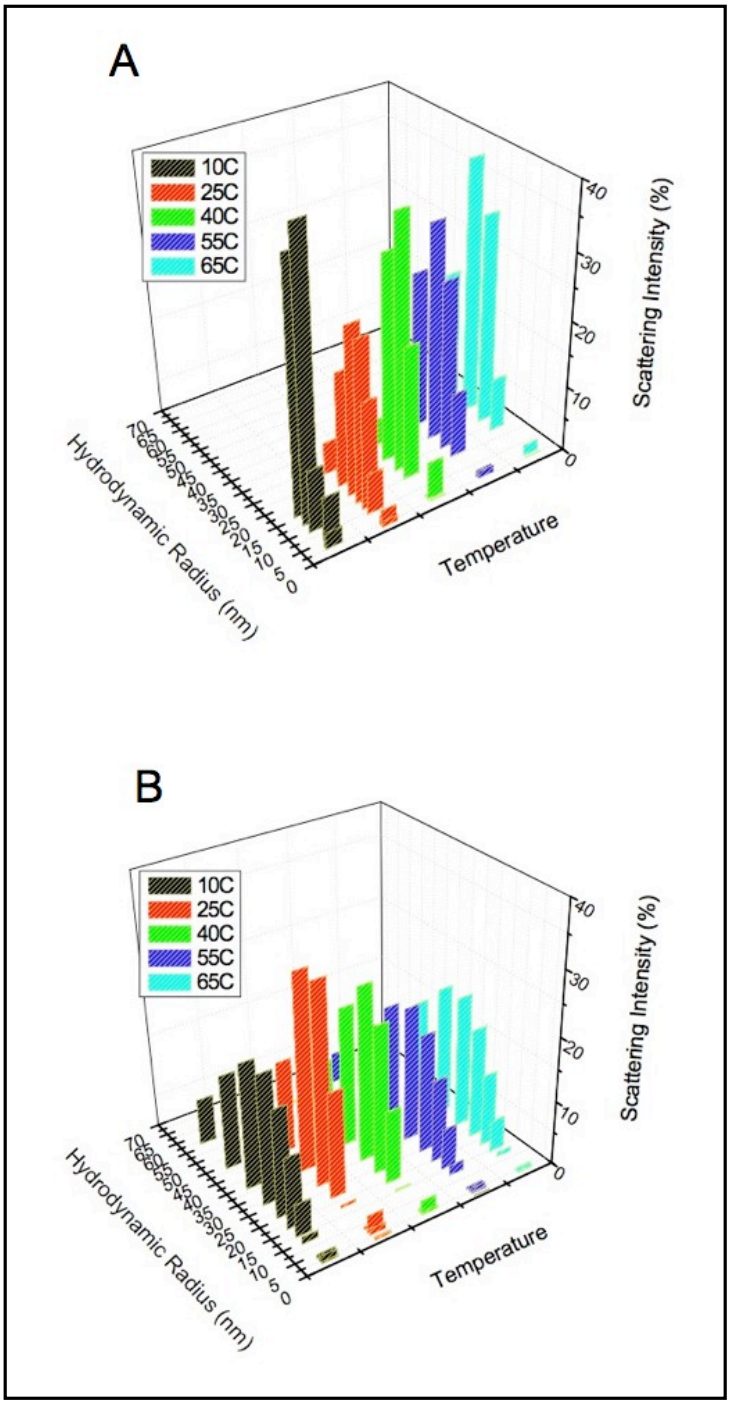

Figure 4.9 DLS distribution data for the (A) WT and (B) V257STOP truncated form of ATF5. The percent intensity of scattered light is displayed as a function of hydrodynamic radius.

4.4 Discussion

In this report we compare the structure and aggregation of the wild-type (WT) form of ATF5 with a truncated mutant form that lacks the valine zipper region (V257STOP). The truncated ATF5 displays a distinctly different structure than the WT form. Removal of the valine zipper region results in a loss of α-helical structure. This is interesting because the leucine zipper region of ATF5 is the region with the highest propensity to adopt α-helical structure, whereas the valine zipper region is predicted to be disordered but have beta propensity.[11] Based on this information, one would expect that removal of the valine zipper region would result in an increased overall helical structure for the mutant. Interestingly, the opposite effect is observed. This indicates that the valine zipper region is important for maintaining the α-helical structure of the adjacent helical region. If as predicted, the central leucine zipper region is the most structured region of ATF5, then the valine zipper region must help stabilize this helical structure. It is not clear whether structural stabilization is accomplished through backbone or side-chain interactions or both.

The dynamic light scattering data indicate that both forms of ATF5 tend to self-associate at low temperature (10 °C). It is difficult to quantify the extent of self-association, because both systems are highly dynamic and heterogeneous. However, based upon the NMR signal achieved using millimolar concentrations of protein, a large percentage of these samples must be present in their low molecular weight monomeric/dimeric form in order to be observed on the NMR timescale.[27] The structural data collected using CD and FTIR reflect the structure of both free and self-

associated protein species. As such, it is likely that the helical signal observed for the WT form of ATF5 is largely reflective of the base protomer structure and the β-structure observed reflects the presence of self-associated protein species. Stable beta structure would result in NH chemical shifts appearing further downfield, above 8.5 ppm, and no such signal is observed in the spectra.

The structural data collected on the mutant form of ATF5 indicate that the protein possesses β-sheet and β-turn structure. However, the data also suggest that the system contains non-standard structural features. Again, because ATF5 self-association occurs at low temperature, the CD and FTIR data may reflect signal from both monomeric or dimeric species and larger soluble aggregates. The complexity of this system precludes clear delineation of the structure of the V257STOP mutant. Extensive higher resolution studies using multi-dimensional NMR will be required to describe this system in further detail.

Thioflavin T fluorescence was measured in the presence of both forms of ATF5 as a function of temperature. Thioflavin T is a fluorescent dye commonly used for the detection of amyloid fibrils.[28] Interestingly, the mutant form of ATF5 was able to induce Thioflavin T fluorescence, but the WT did not induce fluorescence. The emission intensity at 480 nm for the mutant was 710,813 counts/second at 10 °C and the emission intensity for the WT at the same temperature was 875 counts/second. Despite the presence of aggregates in both samples a 1000-fold difference in emission intensity was observed. These data indicate that the initial aggregated or self-associated state of the V257STOP mutant possesses a structure or conformation that

is able to bind and induce Thioflavin T fluorescence. Given that Thioflavin T fluorescence is typically observed in the presence of fibrillar or amyloid aggregates, this suggests that the initial aggregated state of the mutant protein possesses structural features that are similar to those present in amyloid fibrils.[28] The fact that the WT form of ATF5 does not induce Thioflavin T fluorescence indicates that a structural difference exists between the soluble aggregated forms of the two proteins.

The FTIR data indicate that the final aggregated states of the two forms of ATF5 are structurally similar. The increased FTIR absorption signal at 1620 cm^{-1} is reflective of intermolecular β-structure and has been observed for numerous aggregated protein species.[21a, 21c, 23, 25] This suggests that while the pathway or mechanism leading to the formation of the final state may differ between proteins, the structure of the final state is sequence independent.[21a, 21c, 23, 25]

It appears that the WT form of ATF5 protein possesses a conformation that readily permits intermolecular β-sheet addition or elongation. Our initial hypothesis was that solution exposure of the valine zipper region facilitated protein self-association through the initiation of intermolecular β-structure. Therefore, under reducing conditions, where the valine zipper region is more exposed, protein aggregation occurs more readily. Our data show that upon removal of the valine zipper region, the bZIP domain still self-associates, but these soluble structured aggregates are thermally-stable thus reducing further aggregate growth. Moreover, no difference in aggregation is observed between the reduced and non-reduced forms of the ATF5 mutant. It appears that removal of the valine zipper region changes the

conformation of the alpha-helical region such that intermolecular disulfide bond formation is no longer a determining factor in aggregate formation, structure or thermal stability.

4.5 References

1. (a) Carpenter, J. F.; Randolph, T. W.; Jiskoot, W.; Crommelin, D. J.; Middaugh, C. R.; Winter, G.; Fan, Y. X.; Kirshner, S.; Verthelyi, D.; Kozlowski, S.; Clouse, K. A.; Swann, P. G.; Rosenberg, A.; Cherney, B., Overlooking subvisible particles in therapeutic protein products: gaps that may compromise product quality. *J Pharm Sci* **2009,** *98* (4), 1201-5; (b) Bee, J. S.; Davis, M.; Freund, E.; Carpenter, J. F.; Randolph, T. W., Aggregation of a monoclonal antibody induced by adsorption to stainless steel. *Biotechnol Bioeng* **2010,** *105* (1), 121-9; (c) Ludwig, D. B.; Carpenter, J. F.; Hamel, J. B.; Randolph, T. W., Protein adsorption and excipient effects on kinetic stability of silicone oil emulsions. *J Pharm Sci* **2010,** *99* (4), 1721-33; (d) Wang, W., Protein aggregation and its inhibition in biopharmaceutics. *Int J Pharm* **2005,** *289* (1-2), 1-30.

2. (a) Hermeling, S.; Crommelin, D. J.; Schellekens, H.; Jiskoot, W., Structure-immunogenicity relationships of therapeutic proteins. *Pharm Res* **2004,** *21* (6), 897-903; (b) Hermeling, S.; Schellekens, H.; Maas, C.; Gebbink, M. F.; Crommelin, D. J.; Jiskoot, W., Antibody response to aggregated human interferon alpha2b in wild-type and transgenic immune tolerant mice depends on type and level of aggregation. *J Pharm Sci* **2006,** *95* (5), 1084-96; (c) Rosenberg, A. S., Effects of protein aggregates: an immunologic perspective. *AAPS J* **2006,** *8* (3), E501-7; (d) Fradkin, A. H.; Carpenter, J. F.; Randolph, T. W., Immunogenicity of aggregates of recombinant human growth hormone in mouse models. *J Pharm Sci* **2009,** *98* (9), 3247-64; (e) Purohit, V. S.; Middaugh, C. R.; Balasubramanian, S. V., Influence of aggregation on immunogenicity of recombinant human Factor VIII in hemophilia A mice. *J Pharm Sci* **2006,** *95* (2), 358-71.

3. (a) Munishkina, L. A.; Fink, A. L.; Uversky, V. N., Accelerated fibrillation of alpha-synuclein induced by the combined action of macromolecular crowding and factors inducing partial folding. *Curr Alzheimer Res* **2009,** *6* (3), 252-60; (b) Bauer, R.; Carrotta, R.; Rischel, C.; Ogendal, L., Characterization and isolation of intermediates in beta-lactoglobulin heat aggregation at high pH. *Biophys J* **2000,** *79* (2), 1030-8; (c) Gomez-Orellana, I.; Variano, B.; Miura-Fraboni, J.; Milstein, S.; Paton, D. R., Thermodynamic characterization of an intermediate state of human growth hormone. *Protein Sci* **1998,** *7* (6), 1352-8; (d) Speed, M. A.; Morshead, T.; Wang, D. I.; King, J., Conformation of P22 tailspike folding and aggregation intermediates probed by monoclonal antibodies. *Protein Sci* **1997,** *6* (1), 99-108; (e) Calloni, G.; Lendel, C.; Campioni, S.; Giannini, S.; Gliozzi, A.; Relini, A.; Vendruscolo, M.; Dobson, C. M.; Salvatella, X.; Chiti, F., Structure and dynamics of a partially folded protein are decoupled from its mechanism of aggregation. *J Am Chem Soc* **2008,** *130* (39), 13040-50; (f) Khare, S. D.; Dokholyan, N. V., Molecular mechanisms of polypeptide aggregation in human diseases. *Curr Protein Pept Sci* **2007,** *8* (6), 573-9; (g) Watzlawik, J.; Skora, L.; Frense, D.; Griesinger, C.; Zweckstetter, M.; Schulz-Schaeffer, W. J.; Kramer, M. L., Prion protein helix1 promotes aggregation but is not converted into beta-sheet. *J Biol Chem* **2006,** *281* (40), 30242-50; (h) Chi, E. Y.; Krishnan, S.; Randolph, T. W.; Carpenter, J. F.,

Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* **2003,** *20* (9), 1325-36.
4. (a) Weiss, W. F. t.; Young, T. M.; Roberts, C. J., Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. *J Pharm Sci* **2009,** *98* (4), 1246-77; (b) Lee, C. F., Self-assembly of protein amyloids: a competition between amorphous and ordered aggregation. *Phys Rev E Stat Nonlin Soft Matter Phys* **2009,** *80* (3 Pt 1), 031922; (c) Maji, S. K.; Wang, L.; Greenwald, J.; Riek, R., Structure-activity relationship of amyloid fibrils. *FEBS Lett* **2009,** *583* (16), 2610-7; (d) Fawzi, N. L.; Yap, E. H.; Okabe, Y.; Kohlstedt, K. L.; Brown, S. P.; Head-Gordon, T., Contrasting disease and nondisease protein aggregation by molecular simulation. *Acc Chem Res* **2008,** *41* (8), 1037-47.
5. (a) Eisenberg, D.; Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Sambashivan, S.; Ivanova, M. I.; Madsen, A. O.; Riekel, C., The structural biology of protein aggregation diseases: Fundamental questions and some answers. *Acc Chem Res* **2006,** *39* (9), 568-75; (b) Stefani, M.; Dobson, C. M., Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* **2003,** *81* (11), 678-99; (c) Fandrich, M.; Meinhardt, J.; Grigorieff, N., Structural polymorphism of Alzheimer Abeta and other amyloid fibrils. *Prion* **2009,** *3* (2), 89-93; (d) Bellotti, V.; Nuvolone, M.; Giorgetti, S.; Obici, L.; Palladini, G.; Russo, P.; Lavatelli, F.; Perfetti, V.; Merlini, G., The workings of the amyloid diseases. *Ann Med* **2007,** *39* (3), 200-7.
6. (a) Fernandez-Busquets, X.; de Groot, N. S.; Fernandez, D.; Ventura, S., Recent structural and computational insights into conformational diseases. *Curr Med Chem* **2008,** *15* (13), 1336-49; (b) Squires, A. M.; Devlin, G. L.; Gras, S. L.; Tickler, A. K.; MacPhee, C. E.; Dobson, C. M., X-ray scattering study of the effect of hydration on the cross-beta structure of amyloid fibrils. *J Am Chem Soc* **2006,** *128* (36), 11738-9.
7. (a) Bellesia, G.; Shea, J. E., Effect of beta-sheet propensity on peptide aggregation. *J Chem Phys* **2009,** *130* (14), 145103; (b) Rezaei-Ghaleh, N.; Zweckstetter, M.; Morshedi, D.; Ebrahim-Habibi, A.; Nemat-Gorgani, M., Amyloidogenic potential of alpha-chymotrypsin in different conformational states. *Biopolymers* **2009,** *91* (1), 28-36; (c) Qin, Z.; Hu, D.; Zhu, M.; Fink, A. L., Structural characterization of the partially folded intermediates of an immunoglobulin light chain leading to amyloid fibrillation and amorphous aggregation. *Biochemistry* **2007,** *46* (11), 3521-31.
8. Wang, L.; Maji, S. K.; Sawaya, M. R.; Eisenberg, D.; Riek, R., Bacterial inclusion bodies contain amyloid-like structure. *PLoS Biol* **2008,** *6* (8), e195.
9. (a) Kunjithapatham, R.; Oliva, F. Y.; Doshi, U.; Perez, M.; Avila, J.; Munoz, V., Role for the alpha-helix in aberrant protein aggregation. *Biochemistry* **2005,** *44* (1), 149-56; (b) Thompson, A. J.; Barnham, K. J.; Norton, R. S.; Barrow, C. J., The Val-210-Ile pathogenic Creutzfeldt-Jakob disease mutation increases both the helical and aggregation propensities of a sequence corresponding to helix-3 of PrP(C). *Biochim Biophys Acta* **2001,** *1544* (1-2), 242-54.

10. (a) Ciaccio, N. A.; Moreno, M. L.; Bauer, R. L.; Laurence, J. S., High-yield expression in E. coli and refolding of the bZIP domain of activating transcription factor 5. *Protein Expr Purif* **2008,** *62* (2), 235-43; (b) Angelastro, J. M.; Canoll, P. D.; Kuo, J.; Weicker, M.; Costa, A.; Bruce, J. N.; Greene, L. A., Selective destruction of glioblastoma cells by interference with the activity or expression of ATF5. *Oncogene* **2006,** *25* (6), 907-16; (c) Monaco, S. E.; Angelastro, J. M.; Szabolcs, M.; Greene, L. A., The transcription factor ATF5 is widely expressed in carcinomas, and interference with its function selectively kills neoplastic, but not nontransformed, breast cell lines. *Int J Cancer* **2007,** *120* (9), 1883-90.
11. Ciaccio, N. A.; Laurence, J. S., Effects of disulfide bond formation and protein helicity on the aggregation of activating transcription factor 5. *Mol Pharm* **2009,** *6* (4), 1205-15.
12. (a) Byler, D. M.; Susi, H., Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **1986,** *25* (3), 469-87; (b) Krimm, S.; Bandekar, J., Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv Protein Chem* **1986,** *38*, 181-364.
13. Wishart, D. S.; Bigam, C. G.; Yao, J.; Abildgaard, F.; Dyson, H. J.; Oldfield, E.; Markley, J. L.; Sykes, B. D., 1H, 13C and 15N chemical shift referencing in biomolecular NMR. *J Biomol NMR* **1995,** *6* (2), 135-40.
14. (a) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A., NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* **1995,** *6* (3), 277-93; (b) Goddard, T.; Kneller, D. *Sparky 3*, University of California San Francisco.
15. Koppel, D. E., Analysis of Macromolecular Polydispersity in Intensity Correlation Spectroscopy: The Method of Cumulants. *J. Chem. Phys.* **1972,** *57* (11), 4814-20.
16. Provencher, S. W., Inverse problems in polymer characterization: Direct analysis of polydispersity with photon correlation spectroscopy. *Die Makromolekulare Chemie* **1979,** *180* (1), 201-209.
17. Chen, Y. H.; Yang, J. T.; Martinez, H. M., Determination of the secondary structures of proteins by circular dichroism and optical rotatory dispersion. *Biochemistry* **1972,** *11* (22), 4120-31.
18. (a) Fan, H.; Vitharana, S. N.; Chen, T.; O'Keefe, D.; Middaugh, C. R., Effects of pH and polyanions on the thermal stability of fibroblast growth factor 20. *Mol Pharm* **2007,** *4* (2), 232-40; (b) Liu, C.; Chu, D.; Wideman, R. D.; Houliston, R. S.; Wong, H. J.; Meiering, E. M., Thermodynamics of denaturation of hisactophilin, a beta-trefoil protein. *Biochemistry* **2001,** *40* (13), 3817-27; (c) Green, N. M.; Wrigley, N. G.; Russell, W. C.; Martin, S. R.; McLachlan, A. D., Evidence for a repeating cross-beta sheet structure in the adenovirus fibre. *EMBO J* **1983,** *2* (8), 1357-65.
19. Woody, R., Study of Theoretical Circular Dichrosim of Polypeptides: Contribution of Beta-turns In *Peptides, Polypeptides and Proteins* Blout, E.; Bovey, F.; Goodman, M.; Lotan, N., Eds. John Wiley: NY, 1974.
20. (a) Chang, C. T.; Wu, C. S.; Yang, J. T., Circular dichroic analysis of protein conformation: inclusion of the beta-turns. *Anal Biochem* **1978,** *91* (1), 13-31; (b)

Yang, J. T.; Wu, C. S.; Martinez, H. M., Calculation of protein conformation from circular dichroism. *Methods Enzymol* **1986,** *130*, 208-69.
21. (a) Dong, A.; Prestrelski, S. J.; Allison, S. D.; Carpenter, J. F., Infrared spectroscopic studies of lyophilization- and temperature-induced protein aggregation. *J Pharm Sci* **1995,** *84* (4), 415-24; (b) Prestrelski, S. J.; Tedeschi, N.; Arakawa, T.; Carpenter, J. F., Dehydration-induced conformational transitions in proteins and their inhibition by stabilizers. *Biophys J* **1993,** *65* (2), 661-71; (c) Clark, A. H.; Saunderson, D. H.; Suggett, A., Infrared and laser-Raman spectroscopic studies of thermally-induced globular protein gels. *Int J Pept Protein Res* **1981,** *17* (3), 353-64.
22. Susi, H.; Byler, D. M., Protein structure by Fourier transform infrared spectroscopy: second derivative spectra. *Biochem Biophys Res Commun* **1983,** *115* (1), 391-7.
23. Casal, H. L.; Kohler, U.; Mantsch, H. H., Structural and conformational changes of beta-lactoglobulin B: an infrared spectroscopic study of the effect of pH and temperature. *Biochim Biophys Acta* **1988,** *957* (1), 11-20.
24. Priddy, T. S.; Middaugh, C. R.; Carlson, G. M., Electrostatic changes in phosphorylase kinase induced by its obligatory allosteric activator Ca2+. *Protein Sci* **2007,** *16* (3), 517-27.
25. Schwegman, J. J.; Carpenter, J. F.; Nail, S. L., Evidence of partial unfolding of proteins at the ice/freeze-concentrate interface by infrared microscopy. *J Pharm Sci* **2009,** *98* (9), 3239-46.
26. Li, G.; Kasha, P. C.; Late, S.; Banga, A. K., Application of hanging drop technique to optimize human IgG formulations. *J Pharm Pharmacol* **2010,** *62* (1), 125-31.
27. (a) Sapienza, P. J.; Lee, A. L., Using NMR to study fast dynamics in proteins: methods and applications. *Curr Opin Pharmacol* **2010,** *10* (6), 723-30; (b) Spyracopoulos, L.; Sykes, B. D., Thermodynamic insights into proteins from NMR spin relaxation studies. *Curr Opin Struct Biol* **2001,** *11* (5), 555-9.
28. (a) Hawe, A.; Sutter, M.; Jiskoot, W., Extrinsic fluorescent dyes as tools for protein characterization. *Pharm Res* **2008,** *25* (7), 1487-99; (b) LeVine, H., 3rd, Thioflavine T interaction with synthetic Alzheimer's disease beta-amyloid peptides: detection of amyloid aggregation in solution. *Protein Sci* **1993,** *2* (3), 404-10.

CHAPTER 5

CONCLUSIONS AND FUTURE WORK

5.1 CONCLUSIONS

Intermolecular disulfide bond formation of the basic leucine zipper (bZIP) domain of Activating Transcription Factor 5 (ATF5) reduces the size of thermally induced protein aggregates through the retention of alpha-helical structure**.** Removal of the C-terminal valine zipper region results in a loss of alpha-helical structure and a change in the mechanism of protein self-association. The ATF5 protein displays improved thermal stability with regard to aggregate growth in the absence of the valine zipper region, despite an increased tendency to self-associate at lower temperatures.

5.2 FUTURE WORK

5.2.1 Aggregation Kinetics

Kinetic analyses are often employed when studying protein aggregation and can provide valuable mechanistic information. While various aggregation models have been reported in the literature, they can generally be classified into the following categories: unfolding-limited, aggregation-limited and nucleation/growth-limited.[1] Unfolding-limited kinetics describe a system in which the unfolding of the protein is slow and rate-limiting compared to aggregation.[1b, c] This system will display true first-order kinetics. In all other cases, where refolding of the

protein is fast compared to aggregate formation, a folding equilibrium will be maintained over the course of the experiment.[1b, c] In an aggregation-limited system the simplest model assumes that dimer formation is rate-limiting and equivalent to the rate of subsequent growth.[1b, c] This system will display second-order kinetics and the observed rate constant (k_{obs}) will be dependent upon the initial protein concentration (C_0). Lastly, a nucleation/growth driven system describes the case where protein self-association is thermodynamically unfavorable up to a certain oligomerization state and growth occurs only by aggregate elongation or polymerization.[1b, c] (Aggregate-aggregate association is negligible in this case.) This system will display first-order kinetics because nucleation is finished after only a small fraction of monomer is lost. The rate constant will also depend upon C_0. In cases where aggregate-aggregate association is observed, the overall aggregate concentration will decline in proportion with the rate of new aggregate formation.[1b, c] Given the simplest case provided above, where dimer formation is rate-limiting, the system would display second-order kinetics.

Some degree of ambiguity can clearly arise in discerning between the general aggregation mechanisms described above. Certainly, reaction order alone would not be sufficient to assign mechanism. Evaluating the dependence of k_{obs} on C_0 would also be important. Furthermore, an initial monomeric state would be preferred to perform these analyses. A more complicated starting point could limit the amount of information that could be obtained. The Activating Transcription Factor 5 (ATF5) system presented herein would contain monomeric protein, covalent dimer and

soluble aggregates at time zero. The interpretation of kinetic data from this system could be quite complex given the initial conditions. Moreover, given that nucleation has already occurred in this system, analysis might be limited to the evaluation of growth kinetics.

An important consideration when performing kinetic analysis is the methodology used to detect and quantitate aggregation.[1b, c] Often direct detection is performed using size exclusion chromatography (SEC).[1b, c] Unfortunately, SEC is not a viable option for the analysis of ATF5 aggregation, because the protein adsorbs to the stationary phase of the column. Other direct methods of analysis include native PAGE analysis and analytical ultracentrifugation (AUC).[1b, c] A disadvantage of native PAGE analysis is that it is only a semi-quantitative method. Additionally, AUC analysis typically requires substantial user training and can be quite time-consuming in comparison to SEC.

Certainly indirect methods of analysis are available and often used.[1b, c] Solution turbidity can be easily monitored over time, but detects only insoluble aggregates Alternatively, dynamic light scattering (DLS) analysis will monitor the development of soluble aggregates, but not larger insoluble species. Static light scattering (SLS) is another viable option for quantitative detection, but would require substantial experimental modification from the current set-up to accommodate multi-angle detection. Spectroscopic analysis including fluorescence, circular dichroism (CD) and Fuorier transform infrared spectroscopy (FTIR) have also been used to monitor structural changes that reflect aggregate formation over time.[1b, c] However,

the complete lack of tryptophan residues in ATF5 limits the usefulness of a fluorescence assay. Furthermore, the complexity of the ATF5 solution, which contains both monomer/dimer and aggregated species, limits the unambiguous assignment of specific spectral features for this purpose.

5.2.2 High-Resolution Structural Analysis using Nuclear Magnetic Resonance Spectroscopy (NMR)

NMR analysis could be used to obtain more detailed structural information about the WT and V257STOP forms of ATF5.[2] This analysis would require the execution of three-dimensional NMR experiments that would facilitate the assignment of spectral peaks to specific amino acid residues. However, this assignment may prove difficult given the complex nature of the two-dimensional ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectra already obtained. Typically, the time-consuming process of NMR peak assignment is not undertaken until a well-behaved protein system can be obtained. This means that peak overlap is minimal and protein conformational exchange is limited. In the case of ATF5, there are numerous overlapping peaks and clear evidence of conformational exchange. If peak assignments can be obtained, then there is the potential to obtain higher-resolution information about ATF5 structure. Moreover, protein dynamics and structural changes associated with increasing temperature might be investigated in more detail.

5.2.3 Additional Particle Characterization

While the static light scattering (SLS) and dynamic light scattering (DLS) analyses performed in these studies provide useful information about particle growth associated with ATF5 aggregation, the data still lacks the detail and quantitation required to fully characterize this system. Complementary techniques are available that would provide additional information about the aggregation of ATF5. Micro-Flow Imaging (Brightwell Technologies Inc.) uses digital microscopy and micro-fluidics to image particles in solution.[3] This technique provides information about the number, size and shape of particles within the micron size range (approximately 1 μm to 300 μm). This information would complement that obtained using DLS, which detects particles in the nanometer size range (approximately 1 nm to 1000 nm). It would permit a quantitative comparison of the size and morphology of insoluble protein aggregates between the WT and V257STOP forms of ATF5.

Another new technology that is used for particle analysis is NanoSight (NanoSight Ltd.), which uses optical microscopy to track the diffusion of nanoparticles in solution.[4] In this method each particle is measured separately, which permits better resolution and quantitation of different sized particles in solution than that obtained using DLS. The size range for nanoparticle detection using NanoSight is approximately 30 nm to 1000 nm for protein aggregates (depending on refractive index). Smaller species (below 30 nm) that can be detected using DLS will not be observed using this method. However, NanoSight analysis could provide a more quantitative comparison of the size and distribution of the soluble aggregates between the WT and V257STOP forms of ATF5.

5.2.4 Single Point Mutations in the Valine Zipper Region

Further investigation of the role of the valine zipper region in facilitating ATF5 aggregation could be performed by introducing point mutations into this C-terminal region. For example, we have suggested that the inherent propensity of valine residues to adopt beta conformation versus α-helical structure may explain the more facile aggregation of ATF5 observed with the valine zipper region intact.[5] This could be evaluated by mutating the individual valine residues in this region (V257, V264 and V271) to leucine, which has a lower propensity to adopt β-structure. In fact, mutagenesis has already been performed on the ATF5 cDNA in the pET-42b expression vector to create the following constructs toward this end: V257L, V264L and V271L. Additionally, the double mutant V257L/V264L and the triple mutant V257L/V264L/V271L have also been made. Preliminary data from the initial purification of these mutant forms of ATF5 indicates that their solubility following cell lysis differs from the WT form. It is likely that modified purification schemes will be required for mutant protein isolation.

5.2.5 Agitation-Induced Protein Aggregation Studies

Increased exposure to an air-water interface is known to induce protein aggregation.[6] It is believed that proteins can partially unfold at this interface and then more readily self-associate in solution. Sample agitation is one way in which increased exposure to this interface can occur.[6] Agitation studies are commonly employed in the biopharmaceutical industry to evaluate product stability. It is likely that the mechanism of protein aggregation is influenced by the type of stress applied

during a stability study. The development of thermally induced protein aggregates can therefore proceed through a different mechanism than that observed for agitation-induced aggregation. Therefore, studies comparing the aggregation of the WT and the V257STOP forms of ATF5 could be extended to evaluate the effects of agitation on ATF5 aggregation.

5.2.6 Impact of Excipient Addition on ATF5 Aggregation

Excipients are solution additives that are commonly used to improve protein stability.[7] Although there are existing theories about how these additives function to stabilize proteins, in many cases the exact mechanisms remain unclear. Given that the WT and V257STOP forms possess different structure and self-associate differently, it would be interesting to compare the effects of excipient addition on these two systems. Preliminary screening was performed using different types of excipients to evaluate their impact on ATF5 aggregation. The data indicate that sucrose, arginine and polysorbate 20 all have obvious, but unique effects on ATF5 aggregation. A more detailed analysis is currently underway to compare the effects of these three excipients at varying concentration on the stability of the WT and V257STOP forms of ATF5.

5.3 REFERENCES

1. (a) Morris, A. M.; Watzky, M. A.; Finke, R. G., Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. *Biochim Biophys Acta* **2009,** *1794* (3), 375-97; (b) Roberts, C. J., Non-native protein aggregation kinetics. *Biotechnol Bioeng* **2007,** *98* (5), 927-38; (c) Weiss, W. F. t.; Young, T. M.; Roberts, C. J., Principles, approaches, and challenges for predicting protein aggregation rates and shelf life. *J Pharm Sci* **2009,** *98* (4), 1246-77.
2. Wuthrich, K., NMR studies of structure and function of biological macromolecules. *Biosci Rep* **2003,** *23* (4), 119-68.
3. Wuchner, K.; Buchler, J.; Spycher, R.; Dalmonte, P.; Volkin, D. B., Development of a microflow digital imaging assay to characterize protein particulates during storage of a high concentration IgG1 monoclonal antibody formulation. *J Pharm Sci* **2010,** *99* (8), 3343-61.
4. Filipe, V.; Hawe, A.; Jiskoot, W., Critical evaluation of Nanoparticle Tracking Analysis (NTA) by NanoSight for the measurement of nanoparticles and protein aggregates. *Pharm Res* **2010,** *27* (5), 796-810.
5. Ciaccio, N. A.; Laurence, J. S., Effects of disulfide bond formation and protein helicity on the aggregation of activating transcription factor 5. *Mol Pharm* **2009,** *6* (4), 1205-15.
6. (a) Kiese, S.; Papppenberger, A.; Friess, W.; Mahler, H. C., Shaken, not stirred: mechanical stress testing of an IgG1 antibody. *J Pharm Sci* **2008,** *97* (10), 4347-66; (b) Serno, T.; Carpenter, J. F.; Randolph, T. W.; Winter, G., Inhibition of agitation-induced aggregation of an IgG-antibody by hydroxypropyl-beta-cyclodextrin. *J Pharm Sci* **2010,** *99* (3), 1193-206; (c) Sicorello, A.; Torrassa, S.; Soldi, G.; Gianni, S.; Travaglini-Allocatelli, C.; Taddei, N.; Relini, A.; Chiti, F., Agitation and high ionic strength induce amyloidogenesis of a folded PDZ domain in native conditions. *Biophys J* **2009,** *96* (6), 2289-98.
7. (a) Hamada, H.; Arakawa, T.; Shiraki, K., Effect of additives on protein aggregation. *Curr Pharm Biotechnol* **2009,** *10* (4), 400-7; (b) Manning, M. C.; Patel, K.; Borchardt, R. T., Stability of protein pharmaceuticals. *Pharm Res* **1989,** *6* (11), 903-18.