2014

A Structure-Function Analysis of the Arabidopsis Chloroplast Import Protein atTic20

James H. Campbell
Wilfrid Laurier University, camp0410@mylaurier.ca

A Structure-Function Analysis of the *Arabidopsis* Chloroplast Import Protein atTic20

by

James Hugh Campbell

THESIS

Submitted to the Department of Biology, Faculty of Science

in partial fulfillment of the requirements for the Master of Science in Integrative Biology

Wilfrid Laurier University

2014

I hereby declare that I am the only author of this thesis.

I understand that my thesis may be made electronically available to the public.

Abstract

Arabidopsis thaliana protein Tic20 (atTic20) is a member of the translocon at the inner envelope membrane of chloroplasts. Evidence to date suggests it is part of the main preprotein conducting aperture in the complex, but its exact role is still debated. To help characterize its role, a protocol optimizing yield and purity of recombinantly expressed atTic20 was developed, and a series of experiments was performed to examine its secondary structure and its ability to interact with chloroplast transit peptides. The attempt to increase protein yield was successful, with growth at 20^{o}C in the auto-inducing media ZYP-5052 showing the greatest potential for recombinant protein expression. Interestingly, expression under these conditions resulted in the targeting of recombinant atTic20 to the bacterial membrane, which necessitated the adoption of a membrane isolation and solubilization procedure. This resulted in an almost eight-fold increase in protein yield per litre of culture compared to an inclusion body extraction procedure. Circular dichroism showed that atTic20 has largely α-helical conformation in liposomes, with a short, intrinsically disordered domain at its N-terminus. With regards to transit peptide interaction, solid phase binding assays and circular dichroism spectroscopy of atTic20 and putative binding partners indicated that atTic20 does not recognize chloroplastic targeting sequences. This strengthens the case for atTic20 as a primarily structural protein.

Acknowledgement

I would like to thank Patrick Hoang and Dr. Masoud Jelokhani for all their assistance, Dr. Matthew Smith and all current and former members of his lab, particularly Siddhartha Dutta, Kyle Weston, Howard Terenski and Spence Macdonald.Drs. R. Michael Garavito and John Froehlich of Michigan State University have my gratitude for their efforts in attempting to crystalize atTic20. I would also like to thank Dr. Joel Weadge, Tu Hoang for her help in native PAGE experiments and Dr. Marc Gibson who first suggested auto-induction. Finally I would like to thank my friends and family, particularly my father.

Table of Contents

Abstract II
Acknowledgements III
Table of Contents IV
List of Figures VI
List of Commonly Used Abbreviations VII

1.0 Introduction 1
1.1 Chloroplasts 1
1.2 Overview of Preprotein Transport into the Chloroplast 1
1.3 The TOC complex 4

1.4 TIC Complex overview 6
1.4.1 Overview 6
1.4.2 Tic20 7
1.4.3 Tic20, Tic110 and the Preprotein Pore 8

1.5 Intrinsic Disorder in Proteins 10
1.6 Summary 11
1.7 Hypothesis, Objectives and Rationale 12

2.0 Experimental Methods 14

2.1 Cloning and Transformation 14
2.1.1 atTic20 truncated mutation 14

2.2 SDS-PAGE & Western blotting general procedures 16

2.3 Expression in BL21 17
2.3.1 IPTG induced expression 17
2.3.2 Autoinduction expression 18
2.3.3 Expression test parameters 19

2.4 Membrane Isolation 19
2.4.1 Differential density centrifugation 19
2.4.2 Detergent Screening protocol 21
2.4.3 NADH oxidase 22

2.5 Purification 22
2.5.1 Immobilized metal affinity chromatography 22
2.5.2 Desalting / Concentrating procedures 23

2.6 Circular Dichroism 24
2.6.1 Range and Machine settings 24
2.6.2 Sample Preparation 24

2.7 Binding Assays 25
2.7.1 In-vitro Translation 25
2.7.2 Solid Phase Binding Assay 26
2.7.3 35S Detection 26

2.8 *In Silico* Methods 27

3.0 Results 28

3.1 *In Silico* Modelling 28

3.2 Optimization of Expression of atTic20 31

3.2.1 Autoinduction 31
3.2.2 Membrane isolation 33
3.2.3 Detergent Screening 37

3.3 Transit Peptide Interaction Study 45

3.3.1 Solid Phase Binding Assay 45
3.3.2 Circular Dichroism 47

3.4 Circular Dichroism of N-terminal Peptide 52

3.5 Crystallization Trials 55

4.0 Discussion 58
4.1 Autoinduction 58
4.2 Membrane Isolation and Detergent Selection 61
4.3 Transit Peptide Interaction Studies 62
4.4 Structure of atTic20, Intrinsic Disorder and the Role of the N-terminal Peptide 64
4.5 Integration of Approaches 67

References 68

Appendix I: atTic20 Clone and Vector 76

Appendix II: Schematics of atTic20 constructs 77

List of Figures

Figure 1. Basic Schematic of Plastid Membranes and Compartments. 2
Figure 2. TIC and TOC complexes 5
Figure 3. Schematic of Membrane Isolation Procedure. 20
Figure 4. Amino Acid Sequence and Models of atTic20 and Truncated atTic20 29
Figure 5. Hydrophobicity Plot of atTic20 30
Figure 6. Autoinduction Expression Test of Tic20 in *E. coli*. 32
Figure 7. Extraction of Tic20 from *E. coli*. 34
Figure 8. SDS PAGE analysis of the membrane fraction isolated from *E. coli*... 35
Figure 9. NADH oxidase assay. 36
Figure 10 CD Spectra of atTic20 in TX-100 and Liposomes. 38
Figure 11. SDS PAGE analysis of solibilization of membrane localized Tic20... 39
Figure 12. CD Spectra of Tic20 in different detergents. 41
Figure 13. IMAC purified atTic20 and truncated atTic20. 42
Figure 14. CD spectra of full length and truncated atTic20 in detergent. 44
Figure 15. Solid Phase Binding of ^{35}S Labelled Preproteins 46
Figure 16. CD spectra of atTic20, DHFR and combined. 48
Figure 17. CD spectra of atTic20, LHC4 and combined. 49
Figure 18. CD spectra of truncated atTic20, DHFR and combined. 50
Figure 19. CD spectra of truncated atTic20, LHC4 and combined. 51
Figure 20. Plot of disorder tendency for NTP. 53
Figure 21. CD spectra of N-terminal peptide in different TFE concentrations. 54
Figure 22. CD spectra of N-terminal peptide at different temperatures. 54
Figure 23. Refolded Tic20 chromatographed on a Superdex-200 column. 57

List of Commonly Used Abreviations

CD - Circular dichroism

CHAPS - 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate

DDM - Dodecyl Maltoside

IDP - Intrinsically Disordered Protein

IPTG - Isopropyl β-D-1-thiogalactopyranoside

LDAO - lauryldimethylamine oxide

LHCA4 - Light Harvesting Complex A4

NTP - N-terminal Peptide

SDS - Sodium dodecyl sulphate

SPP - Signal Processing Peptidase

SSU - Small Subunit of Rubisco

TFE - Trifluoroethanol

THP - Tris(hydroxypropyl)phosphine

TIC - Translocon of the Inner Chloroplast Membrane

TOC - Translocon of the Outer Chloroplast Membrane

TP - Transit Peptide

1. Introduction

1.1 Chloroplasts

Chloroplasts are double membraned organelles present in all photosynthetic plants (Fig. 1). They are one type of a set of organelles called plastids, present in plants and the protist class apicomplexa (Keeling, 2010). Plastids fulfill a wide range of cellular functions: they are the locus of photosynthesis (chloroplasts), the site of non-anthocyanin pigment accumulation (chromoplasts), biomolecule storage containers (elaioplasts, amyloplasts, other leucoplasts), gravisensitive organelles (amyloplasts), and are involved in nucleic acid and amino acid biosynthesis (Wise, 2006). Like mitochondria, plastids are believed to be present in eukaryotes as the result of an ancestral endosymbiotic event. Unlike mitochondria, however, chloroplasts have a tertiary internal membrane known as the thylakoid, which coils throughout the stromal space into stacks known as grana (Fig. 1; Wise, 2006). Chloroplasts have their own circular chromosome and synthesize a number of chloroplastic specific proteins; however, the vast majority of chloroplast proteins are encoded in the nuclear genome, translated in the cytosol and post translationally imported into the organelle (Cline and Henry, 1996).

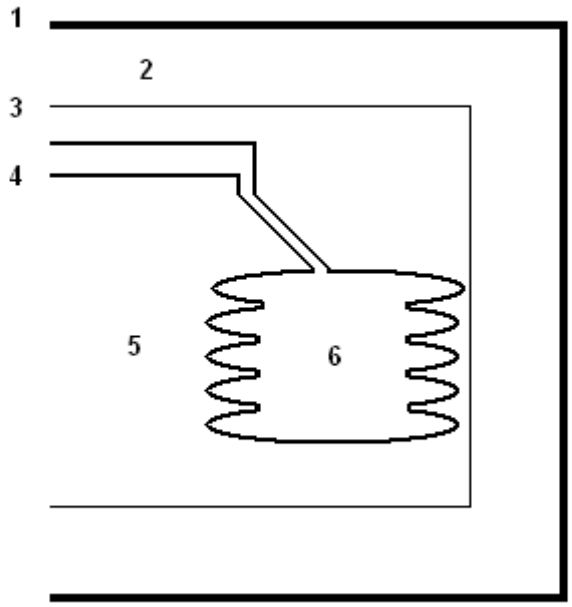


Figure 1. Basic Schematic of Chloroplast Membranes and Compartments. Diagram shows various membrane and soluble compartment regions as follows: 1. Outer envelope membrane 2. Intermembrane space 3. Inner envelope membrane 4. Thylakoid membrane 5. Stroma 6. Thylakoid lumen (in grana stacks)

1.2 Overview of Preprotein Transport into the Chloroplast

Any protein destined for the chloroplast interior has to pass two barriers: the inner and outer chloroplast membranes. It has been shown that preproteins are targeted to the chloroplast by a signalling sequence of amino acids at the N-terminus called a transit peptide, and that they transit the membranes in an unfolded state guided by a set of membrane proteins (Cline and Henry, 1996). Transit peptides have been shown to be both necessary and sufficient for preprotein import into chloroplasts (Karlin-Neumann and Tobin, 1986). This transition is mediated by two protein complexes, the TIC and the TOC complexes (*t*ranslocon at the *i*nner -or *o*uter, respectively- membrane of the *c*hloroplast; Arronson and Jarvis, 2009). Preprotein import into chloroplasts is an energy-dependent process, requiring GTP and ATP at the outer membrane and ATP at the inner membrane (Arronson and Jarvis, 2009). The transit peptide is then cleaved by a signal processing peptidase upon reaching the stroma, resulting in a shorter, transit peptide cleaved or "mature" protein (Arronson and Jarvis, 2009). There are, however, some exceptions to this transit peptide dependent pathway. The chloroplast envelope quinone oxidoreductase homologue lacks either an N- or C-terminal transit peptide, and yet is located in the chloroplast inner membrane (Miras, 2002). It appears to have an internal transit peptide that mediates its import to the inner membrane (Miras,2002) by an unknown mechanism. Likewise, the *Arabidopsis* carbonic anhydrase 1 enzyme is believed to be imported via an endoplasmic reticulum mediated system (Villarejo et al, 2005). Nevertheless, alternative transport across the double-membrane envelope is limited in its occurrence, with few known examples and many putative candidates being

shown either to have terminal targeting sequences or not to localize to the chloroplast at all (Armbruster et al, 2009).

1.3 The TOC complex

The TIC-TOC complexes (Fig. 2) are believed to form a scaffold and gating system by which the majority of chloroplast-targeted preproteins can be recognized and translocated, with translocation being completed by stromal molecular chaperones (Arronson and Jarvis, 2009). The three major components of the TOC complex are Toc34, Toc159 and Toc75 (Arronsson and Jarvis, 2009). Of these components, Toc159 and Toc34 are in fact represented by multi-gene families in most plant species (Kubis et al, 2003; Bauer et al, 2001; Yan et al, 2014). These include the epynomous Toc159 and Toc34, as well as Toc33 (in the latter family) and Toc132, Toc120 and Toc90 (in the former) in *Arabidopsis.* Toc159 and Toc34 both have GTPase domains, and are jointly involved in recognizing the transit peptide (and by extension translocating preproteins) (Arronsson and Jarvis, 2009). The mechanics of this system are unclear, with the two prevailing models each indicating a different member of the pair as the primary receptor of the translocating preprotein (Becker et al, 2004, Perry and Keegstra, 1994). What is more clear is that the different homologues of these two proteins distinguish between the functional purpose of translocating preproteins; with some preferring photosynthetic partners over housekeeping ones, or *vice versa* (Ivanova et al, 2004; Dutta et al, 2014). Toc75 is a β-barrel embedded in the outer membrane, and serves as a pore or conduit through which translocating preproteins are funnelled (Arronson and Jarvis, 2009).

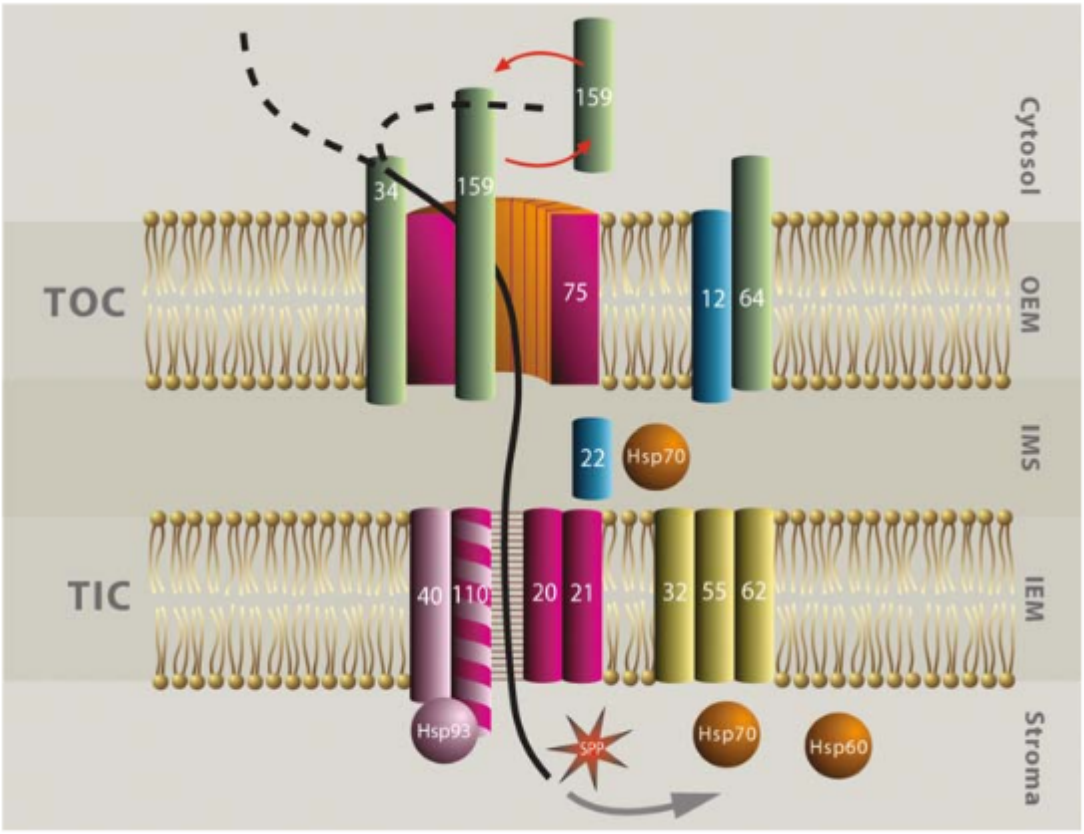


Figure 2. TIC and TOC complexes. Image from Aronsson and Jarvis, 2009, reproduced with permission. Image shows a preprotein (black line) passing through TOC complex at the outer membrane (OEM), through the inter membrane space (IMS) and through the TIC complex at the inner membrane (IEM). Dashed lines indicate the two possible recognition pathways, one with initial docking at Toc34, the other with Toc159. The image reflects the uncertainty of the identity of the inner envelope channel, with Tic110 and Tic20 both shown as possible candidates.

1.4 The TIC Complex

1.4.1 Overview

In contrast to the TOC complex, the TIC complex is less well characterized. Many significant questions remain about what role individual proteins perform in preprotein translocation. Though speculated, preferences for specific preprotein "types" (i.e. photosynthetic v. non-photosynthetic) have not been established (Kikuchi et al, 2009). More certain is the idea that preprotein translocation is physically driven by a molecular motor in the TIC complex, but the identity of this motor is unknown, with Tic110 and Hsp93 suggested to be prime candidates (Jarvis and Soll, 2001). More recently, it has been shown that chloroplastic Hsp90 (Hsp90C) co-precipitates with Tic110 and is likely part of a motor complex with Hsp93 and Tic110, but direct interaction between the three has yet to be demonstrated (Inoue and Schnell, 2012). There are some broad generalizations which can be made, however. Tic110, Tic40, Hsp93 and Hsp90 are believed to form a scaffold-chaperone unit involved in guiding and refolding preproteins on the stromal side of the inner membrane (Arronson and Jarvis, 2009; Inoue and Schnell, 2012). As discussed, Tic110 and the heat shock proteins are the believed to be the primary components of this unit, while Tic40 either serves as an additional chaperone or a regulator of the ATPase activity of Tic110 (Chou et al, 2003; Inoue and Schnell, 2012). The events facilitating preprotein movement through the intermembrane space and across the inner membrane itself are less well established, but Tic20, Tic22, Tic21 and Hsp70 are all believed to be involved (Arronson and Jarvis, 2009). A recent

publication by Kikuchi et al(2013) has uncovered several new members of the complex including Tic100, Tic214 and Tic56. These components, along with Tic20, are part of a one megadalton complex embedded in the inner membrane that serves as the locus of translocation (Kikuchi et al, 2013). Finally, the signal processing peptidase (SPP), which cleaves transit peptides post import, is located in association with the TIC complex on the stromal side of the inner envelope membrane (Arronson and Jarvis, 2009).

1.4.2 Tic20

Tic20 - so named for its association with the inner membrane peptide translocation machinery and its molecular weight (20 kDa) - is a protein that was first identified in *Pisum sativum* and has four known isoforms in *Arabidopsis spp* (I,II, IV and V, atTic20-I hereafter referred to as atTic20; Hirabayashi et al, 2011). The *P. sativum* homologue of Tic20 (psTic20) was shown by Kouranov and Schnell (1997) to associate directly with transiting preproteins using chemical cross-linking followed by chloroplast isolation. They also showed by crosslinking experiments that psTic20 was most often associated with peptides that were in the process of transitioning through the inner membrane (Kouranov and Schnell, 1997). Further work by these researchers demonstrated that psTic20 transiently associates with the outer membrane translocation complex during preprotein translocation (Kouranov et al, 1998). Further evidence for Tic20 being the channel forming protein emerged from studies using *Arabidopsis thaliana* plants that were transformed with DNA for antisense atTic20 mRNA (Chen et al, 2002). These plants produced low levels of atTic20, exhibited a pale phenotype, stunted growth, and had plastids arrested at a pre-chloroplastic developmental state (Chen

et al, 2002). Moreover, the chloroplasts in these plants were shown specifically to be impaired in transport of peptides across the inner membrane (Chen et al, 2002). Finally, atTic20 is also believed to be a preprotein aperture due to its similarity to cyanobacterial amino acid transporters and mitochondrial Tim23 (Kalanon et al, 2008).

Research into other components of the TIC complex shows that they largely play a supporting role in inner membrane translocation. A mutant of a structurally similar protein, atTic21 that is highly expressed in germinating Arabidopsis produced a similarly pale phenotype that was also deficient in protein transport (Teng et al, 2006). This led some researchers to speculate that atTic21 performed a role similar to atTic20 in developing plants (Teng et al, 2006). It was later found that atTic21 was present in mature plants, but only transiently and in a 1 MDa complex with atTic20 (Kikuchi et al, 2009). In both cases, atTic20 was an essential component of the complex, suggesting it may play a central role in the complex, such as serving as a preprotein-conducting pore.

1.4.3 Tic20, Tic110 and the Preprotein Pore

Whether or not Tic20 is "the" channel-forming protein of the TIC complex has been the subject of some controversy. Another TIC protein, Tic110, is also a potential candidate, as the protein has been predicted by some researchers to have a set of alpha-helical domains that will insert into the inner chloroplast membrane, and has been demonstrated to have channel activity by electrophysiological experiments (Heins et al, 2003; Balsera et al, 2009). Kovacs-Bogdan and her colleagues (2011) contended that the high abundance of Tic110 relative to Tic20 in the chloroplast membrane means that

Tic110 must be the main channel protein, while Tic20 forms an alternative conduit for a subset of proteins. However, experiments or analysis to relate transport kinetics to protein quantities to support this assertion were not performed. Conclusive evidence for the role of Tic110 remains elusive, because full length Tic110 resists expression in bacterial systems: each of the experiments hereafter mentioned were done on individual domains of recombinant Tic110 (Heins et al, 2003; Balsera et al, 2009). The possible role of Tic110 as a pore has been further called into question by other studies, which have demonstrated that the majority of the Tic110 protein extends into the stroma (Jackson et al, 1998). More recent evidence indicates that the domain of Tic110 that has been proposed to act as a channel (Balsera et al, 2009) behaves as a soluble protein when expressed in *E. coli* or in plants as a recombinant protein (Inaba et al, 2005). Based on these data, an alternative model for Tic110 suggests that it instead has a role in recruiting stromal factors, such as Hsp93 and the SPP to the TIC complex (Jarvis and Soll, 2001; Inaba et al, 2005). To further complicate matters, variants of atTic20 (atTic20-IV and atTic20-V) have been found localized to the thylakoid membrane and mitochondria (Machettira et al, 2011). While this finding does not rule out the possibility that atTic20 is the primary chloroplast import channel *per se*, the fact that some of its isoforms may be tied to functions in other organelles does raise questions when considering the specificity of preprotein transit.

One possibility raised by Kikuchi and colleagues (2009) is that atTic20 (or rather, one of its isoforms) preferentially imports photosynthetic preproteins. This speculation was predicated on the pale phenotype characteristics of the atTic20-I knock-out mutant (Kikuchi et al, 2009). Specificity for import of photosynthetic preproteins is believed to

be a property of some members of the Toc159 family of proteins, as evidenced by studies on transgenic plants and the predominance of certain variants in green tissue (Bauer, 2000; Smith et al, 2004; Ivanova et al 2004; Kubis et al, 2004). The similarity between the pale phenotypes of Toc159 and atTic20-I mutants, as well as the hypothesized presence of a photosynthetic preprotein-preferring translocation apparatus at the outer membrane (or alternatively, negative regulation against photosynthetic preproteins for some Toc159 isoforms (Dutta et al, 2014)) spurred investigation into the possibility that some atTic20 isoforms may also serve as transporters specifically for photosynthetic preproteins (Kikuchi et al, 2009). Further investigation revealed that atTic20-I mutants are specifically impaired in photosynthetic preprotein transport, and are primarily localized in leaves and shoots (Hirabayashi, 2011). The mechanism of selectivity for photosynthetic preproteins, and more particularly if it is atTic20 itself that provides the selectivity, is still unknown.

1.5 Intrinsicly Disordered Proteins

Intrinsically disordered or unstructured proteins (IDPs) are proteins that lack stable secondary structure when in isolation or in physiological conditions (Fink, 2005; Dyson and Wright, 2005). Typically, these proteins (or protein domains, in some cases) are thought to act as loci of protein-protein or protein-nucleic acid interactions, or as flexible linker regions (Dyson and Wright, 2005). In the case of protein or nucleic acid binding motifs, “unstructured” is something of a misnomer, as these domains often adopt secondary and tertiary structure upon binding with their ligand or binding partner (Fink,

2005; Dyson and Wright, 2005). In the TOC complex, the acidic domains of the Toc159 family have been shown to be intrinsically disordered, and may serve as the site of preprotein transit peptide recognition (Richardson et al, 2009; Inoue et al, 2010; Dutta et al, 2014). It is unclear whether or not there is a TIC complex protein with an analogous IDP region, though the N-terminal region of mature atTic20 is a possible candidate. Assessing whether or not a protein qualifies as being intrinsically disordered can be ascertained using far UV circular dichroism spectroscopy (CD): IDPs (at least in the case of those involved in protein-protein binding) show an extreme minima at 200 nm, but gain structure under certain circumstances, such as addition of trifluoroethanol (TFE), sample heating and extreme pH conditions (Uversky et al, 2000).

1.6 Summary

Chloroplasts are double membraned organelles of endosymbiotic origin. The transfer of many genes from the original endosymbiont to the genome of the host allows for higher fidelity of replication by eukaryotic polymerases, as well as concentrating most of the transcriptional and translational machinery in the same place, but it comes with the cost of having to transport targeted protein products to the plastid. The machinery necessary for the majority of this transport - the TIC-TOC complex - has been well-studied over the last 20 years, but many aspects of its structure and function remain unknown, with the TIC complex in particular requiring much investigation. Tic20 is a particularly tempting candidate from which to launch an investigation of the TIC complex. Previous work has shown that it is essential for the import of preproteins and, taking cues from Toc159, its suspected intrinsically disordered domain raises the

possibility of a role in transit peptide recognition and interaction. To further examine its role, I pursued both structural and protein-protein interaction studies to determine if atTic20 interacts directly with chloroplast transit peptides, and whether it does so preferentially with photosynthetic transit peptides.

1.7 Hypothesis, Objectives and Rationale

The primary purpose of this project was to characterize the structure and function of atTic20 using an *E. coli* expressed recombinant version of the atTic20 protein. Specifically, I wished to test whether atTic20 is capable of functioning as a selective preprotein conducting pore. If atTic20 does have such a role, then it would be predicted to:

1. Adopt a conformation that permits protein translocation and
2. Be capable of recognizing the translocating chloroplast preproteins either by interacting directly with preprotein transit peptides or indirectly through recognition by another member of the TIC complex (e.g. Tic21, Tic50, Tic100 or Tic110)

I attempted to address part 1 of the hypothesis using structural analysis with the aid of collaborators by performing x-ray crystallography. This was also to unequivocally establish the 3D structure of the protein and confirm the predicted 3 alpha-helical domain structure (Kouranov and Schnell, 1998). Our attempt to crystallize atTic20 was

unsuccessful, but I was able to complete some CD work that provided insight with respect to this first hypothesis. A solid phase binding assay was developed to determine if there is a physical interaction between transit peptides and atTic20. CD was also used to test if there was a conformational change of atTic20 in the presence of preprotein transit peptides, which can be indicative of an interaction. Finally, the twenty amino acid N-terminal region of atTic20 was examined by CD to determine if it is intrinsically disordered.

Elucidation of the function of the components of the chloroplast import apparatus, be it atTic20 or other components, could yield a number of benefits in the production of transgenic plants. Scientists working with proton transporters in *Arabidopsis* have shown that anthropogenic modification of H+ pumps can yield plants that are resistant to drought and tolerant of high concentrations of inorganic contaminants (Gaxiola et al, 2002). Theoretically, plants that are genetically modified with respect to chloroplast proteins could be produced that are shade tolerant or that are capable of increased levels of photosynthesis, leading to higher crop yields. However, any protein produced by a genetically modified plant intended for localization to the chloroplast would have to be compatible with the chloroplast import apparatus. These proteins therefore need to be fully characterized before technological applications can be considered. But more importantly, understanding the process of protein translocation in the chloroplast is an achievement that is long overdue. It has been over forty years since Lynn Margulis posited endosymbiont theory (Sagan, 1967). We know, and have known for a long time, more about an event that happened hundreds of millions of years ago than we do about a fundamental process happening in every living photosynthetic plant cell. And this is in

spite of having sequenced the entire genome of *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000). This is not meant as an admonishment – the research being proposed here could not be done without these other monumental accomplishments – but rather as an illustration of the large gaps still remaining in basic plant research.

2. Methods

2.1 Cloning and Transformation

2.1.1 atTic20

A cDNA clone coding for the mature (lacking the transit peptide; mature protein starts at residue 103) *Arabidopsis thaliana* atTic20 protein (see Figure 2 for amino acid sequence) incorporating a C-terminal hexahistidine tag and lacking its transit peptide in the pET21a vector (atTic20-pET21a) was previously prepared by Spence MacDonald (2009) (See appendix I). This clone encodes a recombinant version of mature Arabidopsis atTic20 including a C-terminal His-tag, denoted atTic20. Two *E. coli* stocks - a cloning stock in strain DH5α and an expression stock in BL21 RIPL CodonPlus - were stored at -80ºC. DNA isolation was achieved using the Qiagen QIAprep Spin Miniprep system following the manufacturer's instructions. Transformation of chemically-competent *E. coli* was achieved by the heat shock method (90 seconds at 42ºC) and cells were rendered competent using the standard method of treating with $CaCl_2$. Transformed cells were selected by growth on LB agar plates containing ampicillin (50 μg/ml), streptomycin (25 μg/ml) and chloramphenicol (25 μg/ml).

2.1.2. atTic20 Truncation Mutant

A mutant version of atTic20 lacking 20 amino acids from the N-terminus of the mature protein that are computationally predicted to correspond to a soluble domain was produced using a primer-adapter. The 5' primer adapter had the following sequence: 5'-CCCCCATATGTGGAGACTTTGGCTTGC-3'. The eighteen 3' nucleotides in the primer complement the (mature) atTic20 cDNA sixty nucleotides downstream of the mature atTic20 clone's start codon. To the 5' side of the complementary sequence is a NdeI recognition site (which includes a start codon, and is underlined in the primer-adapter sequence above), as well as four additional cytosines added to facilitate NdeI cutting of the final PCR product for direct ligation into pET21a. The primer-adapter was ordered from Invitrogen. An existing reverse primer containing a XhoI site used previously for cloning the cDNA of mature atTic20 was also used for PCR (5'-CTCGAGATCCGGATATAGTTCCTCCT-3'). PCR amplification of the truncated atTic20 cDNA was achieved using New England Biolabs' (NEB) ThermoPol Taq polymerase and associated buffers. PCR amplification (using atTic20 clone as template) was done for 40 cycles of 30 seconds at 94ºC (melting), 30 seconds at 65ºC (annealing) and 45 seconds at 72ºC (extension), with an additional 3 min of extension at 72ºC at the reaction's end. The PCR product was run on a 1% (w/v) agarose gel for 45 min at 100 V and gel-purified using Promega's Wizard® SV Gel and PCR Clean-Up System. The PCR product and pET21a vector were double digested with XhoI and NdeI restriction enzymes, both from NEB, for 3 h at 37ºC. Digested product and plasmid were gel-purified. Ligation with T4 DNA ligase (NEB) was done using a 4:1 insert to vector ratio at 16ºC overnight according to manufacturer's instructions. Ligated plasmid was transformed into chemically-competent *E. coli* DH5α by heat shock as previously

described (section 2.1.1). Transformed colonies were selected for on ampicillin (50 μg/ml) LB agar plates and the presence of the truncated atTic20 cDNA was detected using colony PCR (using previously mentioned conditions), and confirmed by DNA sequencing performed by Sick Kids Hospital Centre for Applied Genomics. The resulting plasmid, matTic20ΔN-pET21a, encodes a truncated recombinant version of Arabidopsis atTic20 lacking 20 amino acids from the N-terminus of atTic20.

2.2 SDS PAGE and Western Blotting

Polyacrylamide gels were cast in-house at a concentration of 12% acrylamide (w/v), excepting some gels for binding assays, which were cast at a concentration of 15% acrylamide (w/v), as noted in figure legends. Protein samples were mixed 1:1 with 2x sample buffer (5%w/v) SDS, 10%(v/v) glycerol, 100 mM TrisHCl pH 6.0, 10 mM dithiothreitol and 0.2% (w/v) bromophenol blue) and loaded onto gels. Two types of commercial molecular weight ladders were used: a Bio-rad broad range marker and a pre-stained Frogga Bio marker. Gels were run at 120 V for 45 min. Gels were stained and fixed in a solution of 0.05%(w/v) Coomassie R-250, 10%(v/v) acetic acid and 45%(v/v) methanol; they were then destained in a solution of 10%(v/v) acetic acid and 45%(v/v) methanol until contrast was achieved. Transfers for Western blot analysis were performed by the semi-dry method at 15 V for 105 min onto a nitrocellulose membrane (Transfer buffer contained 92mM glycine, 0.05% SDS, 12.5mM Tris, 10% methanol). Washes and incubations were performed with Tris buffered saline (10 mM Tris-HCl pH 7.5, 100 mM NaCl). The blots were probed exclusively with an anti-hexahistidine mouse antibody from Rockland Immunochemicals, as a specific antibody to atTic20 was unavailable.

Blots were blocked with either 1%(w/v) BSA or 5%(w/v) skim milk. They were then probed with a goat anti-mouse antibody conjugated to horseradish peroxidase fusion protein as a secondary antibody (Rockland). Blots were washed twice between steps with Tris-buffered saline (TBS) or TBS containing 0.1%(v/v) Tween-20 (Bioshop). The Western blot was visualized using a chemiluminescent horseradish peroxidase conjugated goat antimouse antibody (Rockland). Coomassie-stained gels and blots were imaged using a VersaDoc-4000 imager (Biorad).

2.3 Recombinant Protein Expression

Expression of recombinant atTic20 and truncated atTic20 was achieved using both standard isopropyl β-D-1-thiogalactopyranoside (IPTG)-induced expression in LB broth and the auto-induction system of F. William Studier (2005). Autoinducing media relies on using a ratio of glucose to lactose that permits growth without induction until the logarithmic phase of growth is reached at which point the lactose in the media induces expression of the transgene under the control of the T7/lac promoter-operator on the pET21 vector without addition of IPTG (Studier, 2005). Recombinant protein yield is generally higher when using auto-inducing conditions as compared to those that rely on the addition of IPTG, and recombinant protein toxicity to the host is often mitigated (Studier, 2005).

2.3.1 IPTG Induction

IPTG-induced expression was achieved by growing cells in LB-Miller media (LB; 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl) until an OD_{600} of 0.6 was reached. IPTG was then added to a final concentration of 0.8 mM. Cells were then incubated overnight (generally 18 h, though 16 h-24 h expressions were occasionally performed) at room temperature (hereby considered to be 22ºC). Cells were collected by centrifugation at 6,000 x *g* for 15 min and stored at -20ºC prior to protein isolation.

2.3.2 Autoinduction

Autoinduction of recombinant protein expression in *E. coli* was achieved according to the protocol described by Studier (2005), detailed as follows. Media components were prepared separately and then combined immediately prior to inoculation. These components were ZY media (5 g/L yeast extract and 10 g/L tryptone), a 50x concentrated solution of 5052 (25%(v/v) glycerol, 2.5%(w/v) glucose and 10% (w/v) α lactose), 1 mM magnesium sulphate and a 20x concentration solution of NPS (500 mM ammonium sulphate, 1 M monobasic potassium phosphate and 1 M dibasic sodium phosphate). This division of media into components is done to prevent reactions between ingredients during autoclaving (i.e. Maillard reaction). These components were therefore autoclaved separately and combined aseptically to generate the 1X ZYP autoinducing media. After media was prepared, it was inoculated with an overnight culture from bacterial stocks in LB and left to grow at room temperature on a shaker-incubator set to 240 rpm for 24 hours (expression time varied between 20 and 26 hours, with little observable effect). Cells were collected and frozen as stated for the IPTG-based expression.

2.3.3 Expression Test Parameters

To test expression of recombinant atTic20, the protein was expressed under both IPTG and autoinducing conditions either at room temperature for 18 h or 37ºC for 3 h. Cells were then resuspended in extraction buffer (10 mM Tris-HCl pH 8.0, 300 mM NaCl) and lysed using lysozyme treatment (0.2 mg/mL for 1 h at 4ºC) followed by sonication using a probe-tip sonicator (on ice at an output of 4 mW continuously for 10 min). Cells were then spun at 10,000 x *g* for 15 min to separate the soluble and membrane fractions (supernatant) from the insoluble aggregates (including inclusion bodies and unlysed cells) in the pellet. Extraction buffer was added to the insoluble pellet fraction up to the pre-separation volume and vortexed and probe sonicated until optically homogenous. Samples were grouped according to expression conditions and run on SDS-PAGE gels. Expression was judged qualitatively based on resulting band intensities. Conditions tested were 37ºC for 3 h with 0.8mM IPTG, 37ºC for 3 h with ZYP-5052 media, room temperature for 18 h with 0.8 mM IPTG, room temperature for 18 h with 0.4mM IPTG and room temperature for 18 h with ZYP-5052.

2.4 Membrane Isolation

2.4.1 Differential Density Centrifugation

Collecting bacterial membranes following recombinant protein expression was performed using a centrifugation based protocol described by Zoonens and Miroux (2004). Figure 3 details the centrifugation procedures and resultant fractions. First, cell lysis was achieved as described in the expression test section (2.3.3). Lysed cells were spun at 15,000 x *g* for

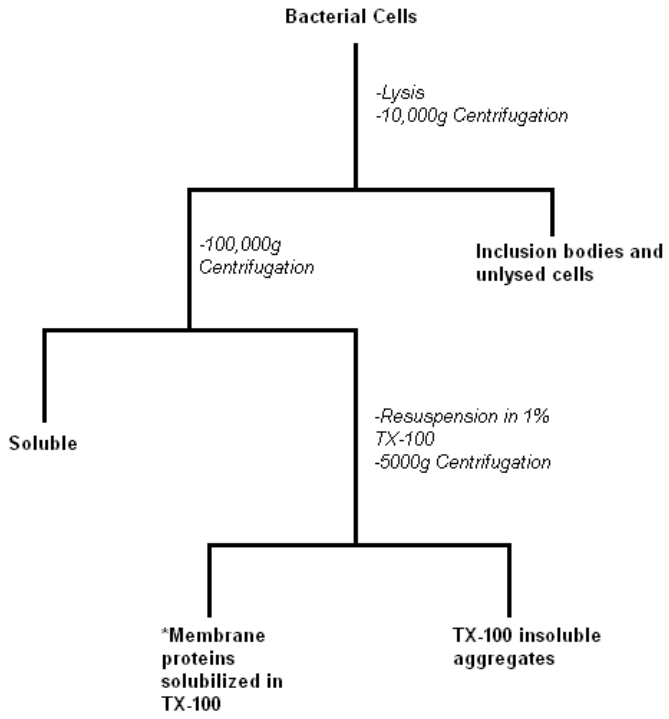


Figure 3. Schematic of Membrane Isolation Procedure. This flow chart shows the centrifugation conditions and resulting fractions from the membrane isolation procedure used in this study. This protocol is based on work by Zoonens and Miroux, 2004.

10 min at 4°C to separate the soluble and membrane fractions from the insoluble fraction; a higher centrifugation speed of 20,000 x *g* was used than during the protein expression test to ensure that no insoluble remnants remained in the membrane-containing fraction. The cleared lysate was then spun at 100,000 x *g* for 45 min at 4°C in a TL-100 rotor (Beckman) using an ultracentrifuge (Beckman Coulter Optima Max Ultracentrifuge) to separate the membrane fraction from the soluble components of the cleared lysate. The supernatant was discarded and the membrane pellet was resuspended in extraction buffer containing either 1%(v/v) Triton-X 100 (Bioshop) or 1%(w/v) Zwittergent 3-14 (EMD Millipore) by pipetting with a 5-mL syringe. Proteins in the pellet were left to solubilize while rotating overnight at 4°C. Following solubilization, the solution was spun at 10,000 x *g* for 10 min at 4°C to remove any remaining insoluble membrane components.

2.4.2 Detergent Screening Protocol

A series of detergents were screened to determine which was most effective at solubilizing recombinant atTic20 from the isolated bacterial membrane. Candidate detergents were selected based on their compatibility with nickel based immobilized metal affinity chromatography (IMAC; compatibility determined by examining manufacturer's (Qiagen) instructions) and their availability. Dodecylmaltoside (DDM), lauryldimethylamine oxide (LDAO), octylglucoside (OG), TX-100, Zwittergent 3-14 and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) were selected as candidates. Equal quantities (4 ml) of cleared lysate were distributed among six ultra-centrifuge tubes, which were spun at 100,000 x *g* to collect the membrane proteins in the pellet fraction. The membrane-protein containing pellet was solubilized as described

above (2.4.1), save that the detergent used was one of the six previously mentioned. Detergents were compared at consistent absolute concentrations (1% w/v). The solubilized protein was then cleared as described above, and levels of solubilized protein in the supernatant were assessed using SDS-PAGE.

2.4.3 NADH Oxidase Assay

To verify that the centrifugation protocol was effective for isolating bacterial membranes, an NADH oxidase assay was performed based on methods described by Arrecubieta (2000). Reactions were prepared in duplicate for the membrane, soluble and cleared lysate fractions. Each 0.35 mL reaction contained 50 mM Tris-HCl pH 7.5, 0.2 mM DTT and 0.12 mM beta-NADH along with 15 mg of sample protein. Blanks were also prepared lacking beta-NADH. Absorbance at 340 nm was then read over a period of five min using a microplate reader (BioTek Synergy HT). Specific activity, measured in units per milligram (U/mg), is defined as nanomoles of NADH oxidized per milligram of protein.

2.5 Recombinant atTic20 purification

2.5.1 IMAC

Gravity column Ni-IMAC was used to purify recombinant atTic20 from the solubilized bacterial membrane. Gravity column work was done in a 4^{o}C cold room. A glass, 50 mL column was attached to a retort stand and the column bed was packed by applying 4 mL of NTA nickel-agarose slurry from Qiagen, with a resulting bed volume of 2 ml. The column was equilibrated with 12 mL of binding buffer containing 10 mM

Tris-HCl pH 8.0, 300 mM NaCl, 1% (w/v) Zwittergent 3-14, 10 mM imidazole and 0.5 mM Tris(hydroxypropyl)phosphine (THP; an IMAC-compatible reducing agent supplied by VWR). The solubilized bacterial membrane was then applied and allowed to flow through. Typically, the flow through was re-applied once and the final flow-through was discarded. The column was then washed with 20 column volumes of wash buffer (10 mM Tris-HCl pH 8.0, 300 mM NaCl, 0.05% (w/v) Zwittergent 3-14, 40 mM imidazole and 0.5 mM THP). The protein was then eluted using 6 mL of elution buffer, collected either in whole or in 1 mL fractions. Elution buffer originally was as washing buffer but with 500 mM imidazole, however the imidazole concentration was exchanged to 300 mM to generate a more normally distributed elution profile. Detergent exchanges were also performed on the IMAC column: this was done by adding the appropriate concentration of the final detergent (e.g. 0.03%(w/v) DDM) to the solubilized membrane, binding and washing buffers and then eluting in a buffer lacking Zwittergent 3-14 but containing the desired detergent. Columns were stored in binding buffer at 4^{o}C and re-used up to five times prior to regeneration according to manufacturer's instructions. Effectiveness of IMAC protocols were monitored by SDS PAGE and verified by Western blot.

2.5.2 Desalting

Desalting of purified proteins in preparation for circular dichroism was achieved using Bio-rad 10DG gravity flow gel filtration columns. These columns have a MWCO of 6 kDa (Bio-rad, 2006). Desalting was performed to remove imidazole and/or to reduce NaCl concentration from 300 mM to 150 mM. Columns were used following manufacturer’s instructions.

2.6 Circular Dichroism (CD) Spectroscopy

2.6.1 Range and Machine Settings

CD was performed using an Aviv 215 CD Spectrometer. Due to the presence of chloride in buffers, wavelength scanning range was limited to 195 nm on the lower end, with the scan starting at 260 nm and a resolution of 1nm. For buffers containing <75 mM NaCl, a 0.5 mm path length cuvette was used. All other scans were done with a 0.2 mm path length cuvette. Four scans for each sample were done in succession. The cuvette was washed with methanol and water between samples and dried with nitrogen. Signal noise was monitored throughout the experiment; in the event that a dynode reading in excess of 500 was detected the scan was terminated. Samples were kept at room temperature. Buffer subtraction was performed for each scan. Buffer subtraction and mathematical addition of spectra was performed using Aviv CD software.

2.6.2 Sample Preparation for CD

Protein concentration was measured by either the Lowry or Bradford method (based on manufacturer's detergent compatibility assessment; Biorad Bradford Reagent and Biorad Detergent Compatible assay kits were used) for each sample run on the CD spectrophotometer. For protein interaction studies, protein samples and protein-protein combination samples were diluted to provide 1:1 molar equivalencies. In such cases where protein concentration was too low to reasonably achieve such a ratio, a 2:1 or 3:1 ratio of molar concentration was used based of the least concentrated sample. To minimize chloride concentration, samples were desalted to a NaCl concentration of 75

mM or 100 mM. Reducing agents (such as THP), denaturants (such as urea and guanidine HCl) were also removed by desalting prior to CD scanning. Glycerol was present as a stabilizer in experiments using dihydrofolate reductase fusion proteins; since scans using glycerol containing blanks as well as a cursory search of the literature indicated that glycerol does not interfere directly with data accumulation (Corrêa, 2009). A typical buffer used for CD analysis was as follows: 10 mM Tris-HCl pH 8.0, 100 mM NaCl, 0.05% (w/v) Zwittergent 3-14 unless otherwise stated in the figure legend. The synthetic peptide used in the intrinsic disorder studies was produced by Miljan Kuljanin of the Jelokhani lab at Wilfrid Laurier University.

2.7 Protein-protein binding Assays

2.7.1 *In vitro* Translation

In vitro translation of proteins was performed using Promega's T7 rabbit reticulocyte coupled TNT system following the manufacturer's instructions, including addition of optional Rnasin (an RNAse inhibitor). Reactions were done in 50 µl volumes. Kit components and *in vitro* products were stored at -80ºC. The ^{35}S-labelled methionine used in translations was provided by Perkin-Elmer (Easy tag express kit). Work was done in the radioactive fume hood, save for the 30ºC incubation, centrifugation and SDS PAGE work. An aliquot of 900 ng of pET21a-LHCA4 plasmid was used per reaction.

2.7.2 Solid Phase Binding Assay

Solid phase binding assays were performed essentially as described by Smith et al (2004). For each reaction, 20 µl of NTA nickel-agarose slurry was transferred to a

microcentrifuge tube. Beads were equilibrated with 100 µl of binding buffer (see 2.5.1 for components). The microcentrifuge tube was spun at maximum speed (21000 x *g*) in a microcentrifuge and the supernatant was aspirated. Various amounts of bait protein - either atTic20 or truncated atTic20 – in binding buffer (see 2.5.1) were then applied to the beads and incubated on a rotator for 10 min. Typically, ranges of 0 to 80 mg in 20 mg increments of protein were studied in an assay. Eight microlitres of *in vitro* translation products were then added to each reaction tube and left to incubate for 45 min. The tubes were again spun (21000 x *g* at room temperature for 2 min) and washed with 600 µl of wash buffer. After spinning (21000 x *g* at room temperature for 2 min) and removing the wash buffer, 20 µl of SDS-PAGE sample buffer containing 300 mM imidazole was applied to the beads. Samples were then used for electrophoresis on 15% SDS-PAGE gels.

2.7.3 Detection of radiolabelled proteins

SDS-PAGE gels of binding assay results were dried to filter paper for one hour using a heated vacuum gel drier. A phosphor screen that had previously been blanked for one hour on a light box was placed along with the dried gels in a cassette to expose the radiation to the phosphor screen. Exposure time ranged from 48 h to 4 days. After exposure, the phosphor screen was imaged using a Personal FX Phosphorimager (Bio-rad)

2.8 *In Silico* Methods

A variety of online tools were used for this project. Translation of nucleotide sequences was done using the ExPASy translate tool and hydrophobicity plots by the ExPASy Protscale tool (http://web.expasy.org/translate/; web.expasy.org/cgi-bin/protscale/protscale.pl). Molarity calculations were done using Promega's BioMath calculator tools (http://www.promega.ca/resources/tools/biomath-calculators). Three dimensional prediction of atTic20 and truncated atTic20 was done using the Distill Server suite and visualized using UCSF Chimera (http://distill.ucd.ie/distill/explanation.html; www.cgl.ucsf.edu/chimera/). Prediction of intrinsic disorder was done using IUPred (http://iupred.enzim.hu/)

3.0 Results

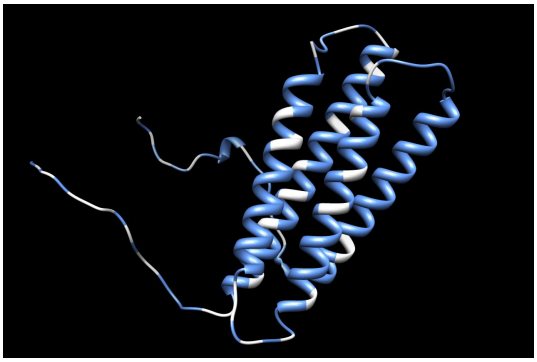
This study was undertaken to examine structural and functional characteristics of atTic20. To this end, a number of studies were performed. Structurally, *in silico* modelling and CD were done to attempt to verify atTic20's predicted alpha-helical structure in an amphiphatic environment, as well as to characterize the disorder of its small, soluble domain. Functionally, solid phase binding assays and further CD were done to detect interaction between the protein and chloroplastic targeting amino acid sequences.

3.1 *In Silico* modelling

The amino acid sequence of mature full-length atTic20 and that of the truncated mutant (Fig 4, left column) were inputted in to Distill 3D modelling software to generate three dimensional models (Fig 4, right column). Distill 3D uses secondary structure homology modelling, solvent accessibility prediction, and residue contact prediction to generate models (Pollastri and McLysaught, 2005). As hypothesized from earlier work (Kouranov and Schnell, 1998) atTic20 is predicted to have four hydrophobic transmembrane alpha-helices with a short, twenty residue N-terminal soluble segment in the mature protein that is predicted to be disordered (IUPred). Some hydrophilic residues are present in the helices; however, in the 3D prediction these residues do not align to form an obvious hydrophilic core. A hydrophobicity plot (Fig 5, from ExPASy) shows that the transmembrane segments are consistently hydrophobic, without regular hydrophilic intervals correlating with turns in

A

MASKDVPSSFRFPPMTKKPQWWWRTLACLPYLM
PLHETWMYAETAYHLHPFLEDFEFLTYPFLGAIGRL
PSWFLMAYFFVAYLGIVRRKEWPHFFRFHVVMGM
LLEIALQVIGTVSKWMPLGVYWGKFGMHFWTAVA
FAYLFTVLESIRCALAGMYADIPFVCDAAYIQIPYDLE



B

MWWWRTLACLPYLMPLHETWMYAETAYHLHPFL
EDFEFLTYPFLGAIGRLPSWFLMAYFFVAYLGIVRR
KEWPHFFRFHVVMGMLLEIALQVIGTVSKWMPLG
VYWGKFGMHFWTAVAFAYLFTVLESIRCALAGMYA
DIPFVCDAAYIQIPYDLE

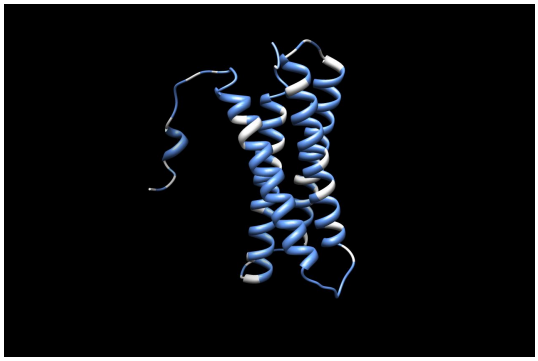


Figure 4: Amino acid sequence of mature TIC20 (A) and N-terminally truncated mutant (B). Modeled helices are underlined. Corresponding predicted 3D structure, with blue signifying hydrophobic residues, are shown on the right. 3D distill libraries were used to generate models (Gianluca and McLysaght, 2005) Max identity value for both predictions was 13.5%

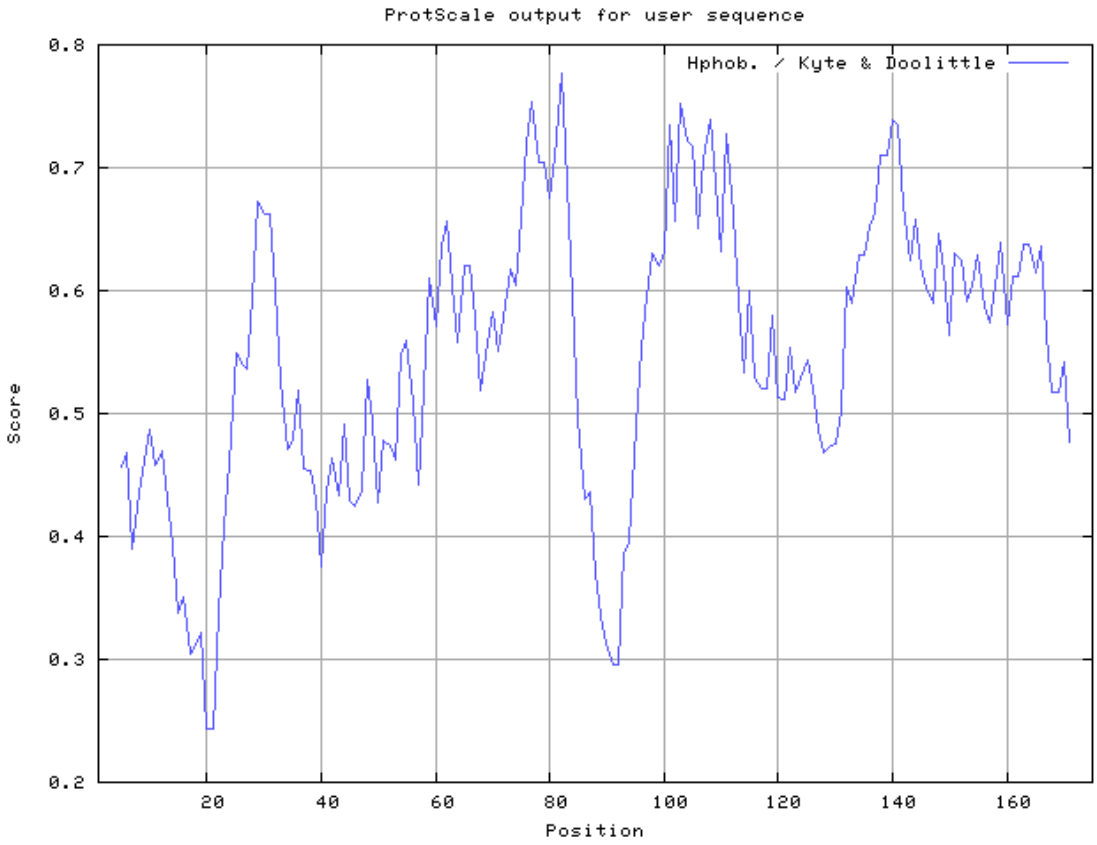


Figure 5. Hydrophobicity plot of atTic20. Generated using Expasy ProtScale tool. Hydrophobicity is standardized to a zero value, and has a window size of nine. Plot uses Kyte and Doolittle method.

the helix; a profile one would expect if the helices formed a soluble central aperture (a common expectation for transmembrane helices with the axis perpendicular to the plane of the membrane). Deletion of the N-terminal segment to produce the truncation mutant resulted in no predicted disruption of the transmembrane helices (Fig. 4).

3.2 Optimization of recombinant atTic20 expression

3.2.1 Autoinduction

In order to optimize recombinant protein yield, a variety of expression conditions were tested. Expression levels were qualitatively compared using coomassie-stained SDS-PAGE gels (Fig. 6). Prior to this test, expression was achieved using IPTG while shaking at RT for approximately 18h (Fig 6 lane 5). Attempts to shorten the expression duration to 3 h by increasing to 37^{o}C resulted in a failure to produce any recombinant protein, regardless of the media used (Fig. 6, Lanes 2 & 3). Changing IPTG concentration likewise had no significant effect on induction (Fig. 6, Lanes 4 & 5). Use of ZYP-5052 autoinducing media, however, resulted in a marked increase in atTic20 expression. Expression in ZYP-5052 at both RT and 37^{o}C resulted in enriched growth of the cultures, with optical density at 600 nm (OD_{600}) peaking at 2.0 and 2.4, respectively, while the uninduced culture in LB peaked at 1.8, of the 6h measured. Of the IPTG induced cultures, OD_{600} did not exceed 0.8 following induction.

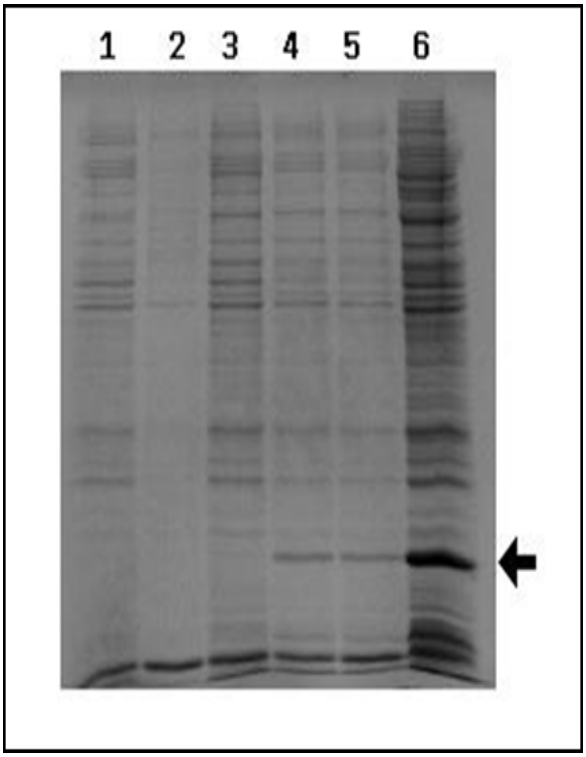


Figure 6. Autoinduction Expression Test of Tic20 in *E. coli*. This gel shows that expression of recombinant atTic20 was highest in autoinducing media. Bacterial lystates were resolved on an SDS-PAGE gel and stained with Coosmassie. Lane identities are as follows: Lane 1: Uninduced culture grown at 37ºC in LB; Lane 2: Culture grown for 3h after induction with 0.4 mM IPTG at 37ºC in LB; Lane 3: Culture grown for 3h after autoinduction at 37ºC in ZYP-5052 autoinducing media; Lane 4: Culture grown for 18h after induction with 0.4 mM IPTG at RT in LB; Lane 5: Culture grown for 18h after induction with 0.8 mM IPTG at RT in LB; Lane 6: Culture grown for 18h after autoinduction at RT in ZYP-5052 autoinducing media. Arrow indicates expected molecular weight of Tic20.

3.2.2 Membrane isolation

Initial attempts to isolate recombinant atTic20 from the inclusion bodies of *E. coli* cells following auto-induction resulted in lower than expected yields. Further investigation revealed that much of the protein in the "inclusion body" fraction was lost during washes with TX-100 (Fig. 7); this led to speculation that the protein was accumulating in the bacterial membrane rather than inclusion bodies. To test this hypothesis, a bacterial membrane isolation protocol using differential velocity centrifugation was adopted. The protein content of various subcellular fractions was compared using SDS-PAGE (Fig. 8). A significant quantity of recombinant atTic20 remained in the unlysed and insoluble fraction (Fig. 8, compare lanes 1 and 4). Although these differences are partly due to inefficient cell lysis, there is likely still some inclusion body formation during auto-induction.This is also evident by the presence of TX-100 insoluble atTic20 remnants. Nevertheless, a substantial portion of the recombinant atTic20 remained in the membrane fraction, though some was lost following a clearing spin following detergent solubilization (Fig. 8, lane 3). To verify that the membrane fraction was in fact being isolated, a bacterial membrane marker NADH oxidase assay was performed. Specific activity for NADH oxidase was highest in the putative membrane fraction. (Fig. 9).

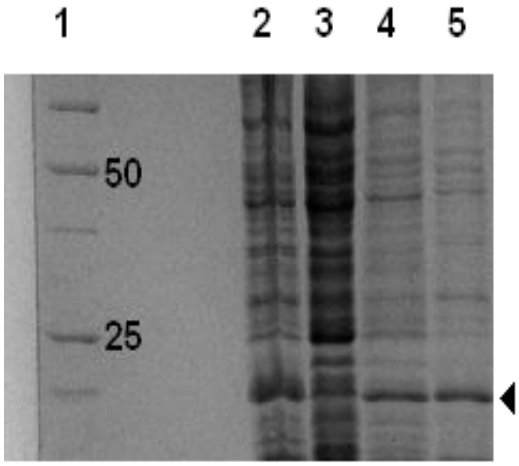


Figure 7. Extraction of Tic20 from *E. coli*. This gel was the first evidence that atTic20 was being directed to the bacterial membrane, as it shows the recombinant protein in the soluble fraction of TX-100 washes. Lane identities are as follows: Lane 1: MW Marker in kDa. Lane 2: Inclusion bodies and unlysed cells. Lane 3: Isolated unlysed cells. Lanes 4-5: Supernatant from successive 1%(w/v) TX-100 washes of inclusion bodies. Arrow indicates expected molecular weight of Tic20.

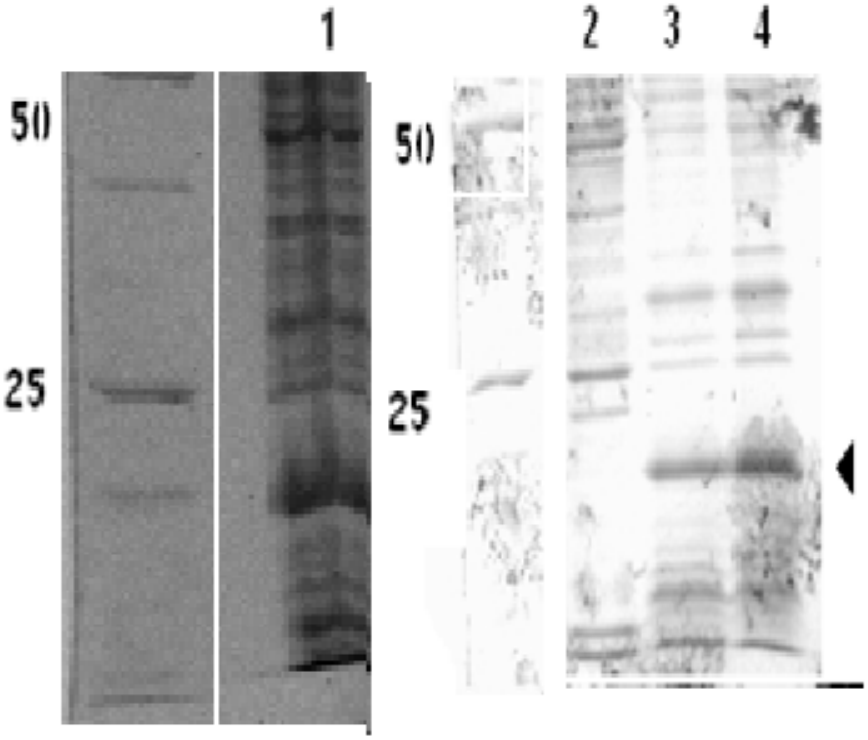


Figure 8: SDS PAGE analysis of the membrane fraction isolated from *E. coli* over-expressing recombinant Tic20 using an auto-induction protocol. This image shows the presence of atTic20 in isolated bacterial membranes. Fractions were isolated according to the protocol described in section 2.4.1, separated on an SDS-PAGE gel, and visualized using Coomassie blue. Lane identities are as follows: Lane 1: Whole cell extract (including inclusion bodies). Lane 2: Soluble protein extract. Lane 3: Triton X-100-insoluble and membrane aggregates. Lane 4: Triton X-100-soluble proteins extracted from the bacterial membrane fraction. Arrow indicates Tic20. Molecular weight markers are indicated to the left (kDa).

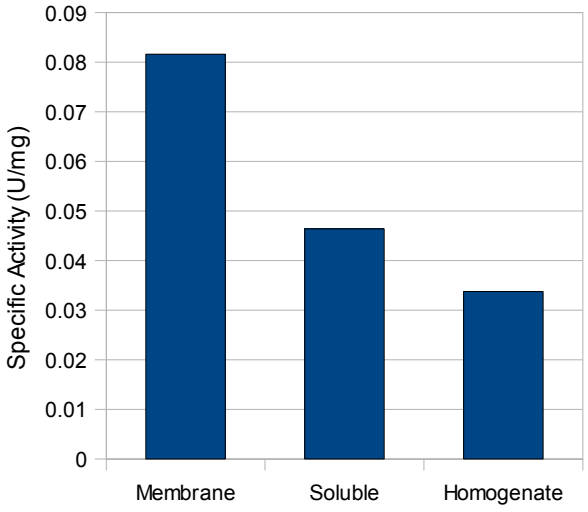


Figure 9: NADH oxidase assay. NADH oxidase activity was used to verify that fractions isolated according to methods described in 2.4.1 were in fact bacterial membranes. Putative fraction identities are labeled on the X-axis, while the Y axis shows specific activity in units (nMoles NADH oxidized) per milligram of protein.

3.2.3 Detergent Screening

Initial solubilization of the bacterial membranes following expression of recombinant atTic20 was done using Triton-X 100 (TX-100). However, as shown in Fig. 6, a large quantity of the protein remained in insoluble aggregates even after overnight solubilization with this detergent. Furthermore, comparison of the CD spectra for TX-100-solubilized atTic20 and atTic20 reconstituted into liposomes (Fig. 10) as well as spectra of atTic20 in DDM from earlier work (MacDonald and Smith, unpublished) showed that the TX-100 solubilized atTic20 contained less alpha-helical character than expected. To enhance recovery of atTic20 from the bacterial membrane fraction and to preserve its secondary (i.e. alpha-helical) structure, a series of mild detergents were screened for their ability to extract atTic20 from the bacterial membranes and maintain the α-helical character of the protein. Fig. 11 shows recombinant atTic20 preparations solubilized from the bacterial membrane fraction using a series of detergents. Assessment of the solubilization efficacy was done qualitatively by comparing band intensity at 20 kDa on an SDS-PAGE gel (Fig. 11); and the band intensity relative to TX-100 was calculated. Of the detergents tested, Zwittergent 3-14, CHAPS, LDAO and DDM all resulted in greater yields of solubilized atTic20 than TX-100 (Fig. 11). Band intensities relative to TX-100 (the initial detergent used for membrane isolation) were 1.6x for DDM, 1.4x for LDAO, 1.6x for CHAPS and 2x for Zwittergent, suggesting Zwittergent solubilized twice as much atTic20 than TX-100 (band intensities from Fig. 11) Insoluble aggregates remaining after solubilization were correspondingly lower when using these detergents. Of the detergents with greater solubilizing capacity than TX-100, all but DDM were selected to have their dissolved protein examined by CD (DDM having been

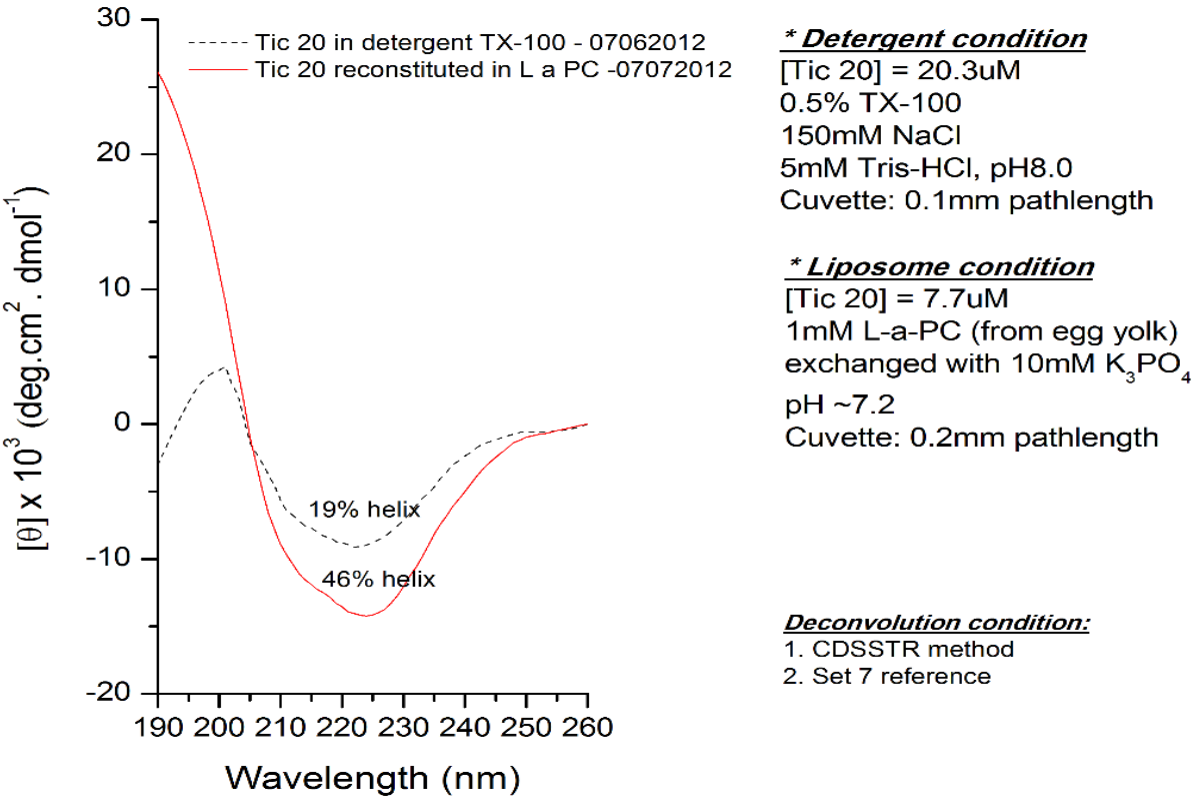


Figure 10: CD Spectra of atTic20 in TX-100 and Liposomes. This graph shows CD spectra of atTic20 solubilized in TX-100 and reconstituted into liposomes. CD conditions are listed in the right hand column. Figure was prepared with assistance from Tuan Hoang.

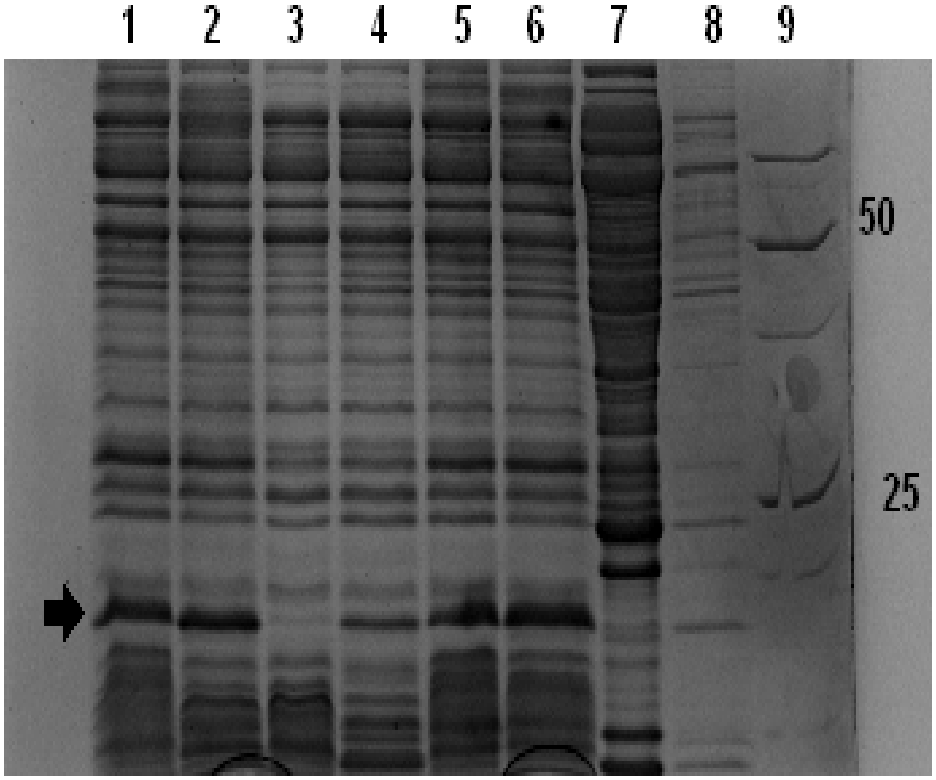


Figure 11: SDS PAGE analysis of solibilization of membrane localized Tic20 using different detergents. Following expresion using the auto-induction protocol, bacterial membranes were isolated, and protein extracted using different detergents at 1% (w/v). The detergent-soluble fractions were analyszed using SDS-PAGE. Lane identities are as follows: Lane 1: DDM. Lane 2: LDAO. Lane 3: OG. Lane 4: TX-100. Lane 5: CHAPS. Lane 6: Zwittegent 3-14. Lane 7: Water soluble protein fraction. Lane 8: Pre-centrifugation (loaded at a lower concentration than other lanes). Lane 9: MW Marker in kDa. Arrow indicates expected MW of atTic20.

previously examined.) CD spectra of recombinant full length atTic20 solubilized in these detergents is shown in Figure 12. The spectra of the protein in both CHAPS and Zwittergent had minima at 208 nm and 222 nm approaching that of atTic20 reconstituted in liposomes, though in CHAPS the minima were slightly blue shifted. atTic20 solubilized in LDAO had a much weaker CD signal than that in either CHAPS or Zwittergent. Deconvolution (using SELCON method) yielded helical content values of 35% for Zwittergent, 9% for LDAO and 29% for CHAPS, although this is somewhat unconvincing from a qualitative assessment, as the minima at 222 nm and 208 nm for CHAPS are lower than Zwittergent and have a near 1:1 ratio in left-handed intensity. Ultimately, Zwittergent 3-14 was selected as the detergent of choice for handling atTic20, due to its ability to readily solubilize the membrane fraction and the quality of the CD spectra of atTic20 in the detergent. While the protein in CHAPS had comparable spectra, it wasn't selected due to its unsuitability for crystallization work and complications when preparing samples for CD (the detergent sticks to the cuvette). In addition, Zwittergent 3-14 was the preferred detergent of our crystallization collaborators.

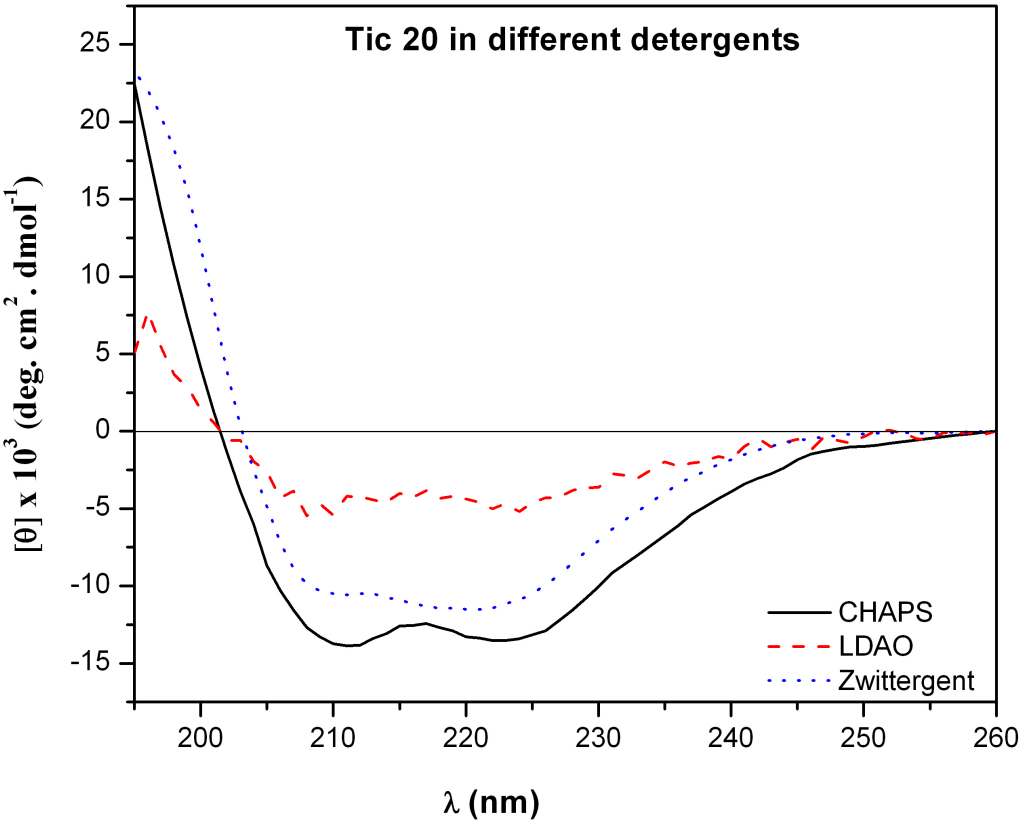


Figure 12:CD Spectra of Tic20 in different detergents. CHAPS and Zwittergent 3-14 both show strong alpha helical character, but not LDAO. Deconvolution using the SELCON method yielded percent helical values of 35% for Zwittergent 3-14, 29% for CHAPS and 9% for LDAO. Standard buffer for all samples contained 150 mM NaCl and 10 mM Tris-HCl pH 8. Detergent concentrations were 12 mM for CHAPS, 2mM for LDAO and 0.8mM for Zwittergent 3-14. The cuvette had a pathlength of 0.2mm.

In summary, the protocol developed before this study for expression of atTic20 using IPTG induction provided an average yield of less than 0.2 mg of protein per litre of culture after IMAC-purification of protein solubilized under harsh (i.e. 6M Guanidine HCl) conditions from inclusion bodies. Adoption of the auto-induction method not only provided higher protein yields after IMAC purification (1.5 mg/$L_{Culture}$ on average), but also resulted in the protein being targeted to the bacterial membranes (by an unknown mechanism), which allowed for a more mild extraction/solubilization of the protein using mild detergents, thereby eliminating the need to solubilize from inclusion bodies.

3.2.4 Truncated mutant expression, purification and structure.

The truncated mutant of atTic20 was expressed and purified under the same conditions as the full-length protein. Figure 13 shows IMAC purified atTic20 and truncated atTic20. The truncated atTic20 sample shown on the gel is from an isolated bacterial membrane, suggesting that the mechanism that targets recombinant atTic20 to the membrane is not dependent on the presence of the N-terminal portion of atTic20. This also suggests that the targeting factor is possibly a property of one of atTic20's transmembrane helices. Figure 14 shows CD spectra of both full length and truncated atTic20. Of particular note is the loss of intensity of the 208 nm minima in truncated atTic20 compared to full length atTic20.

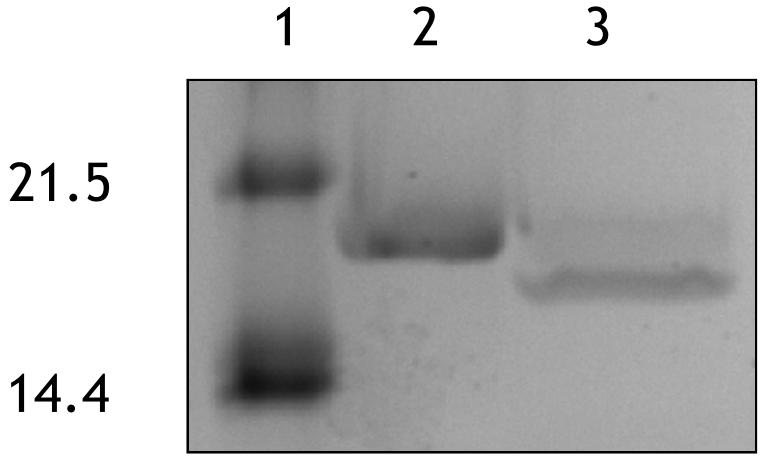


Figure 13. IMAC purified atTic20 and truncated atTic20. Lane identies are as follows: Lane 1: Molecular weight marker in kDa; Lane 2: Purified recombinant atTic20, Lane 3: Purified recombinant truncated atTic20. Both purified proteins were produced from isolated *E. coli* membranes, suggesting targeting in recombinant bacteria is not dependent on the N-terminal domain.

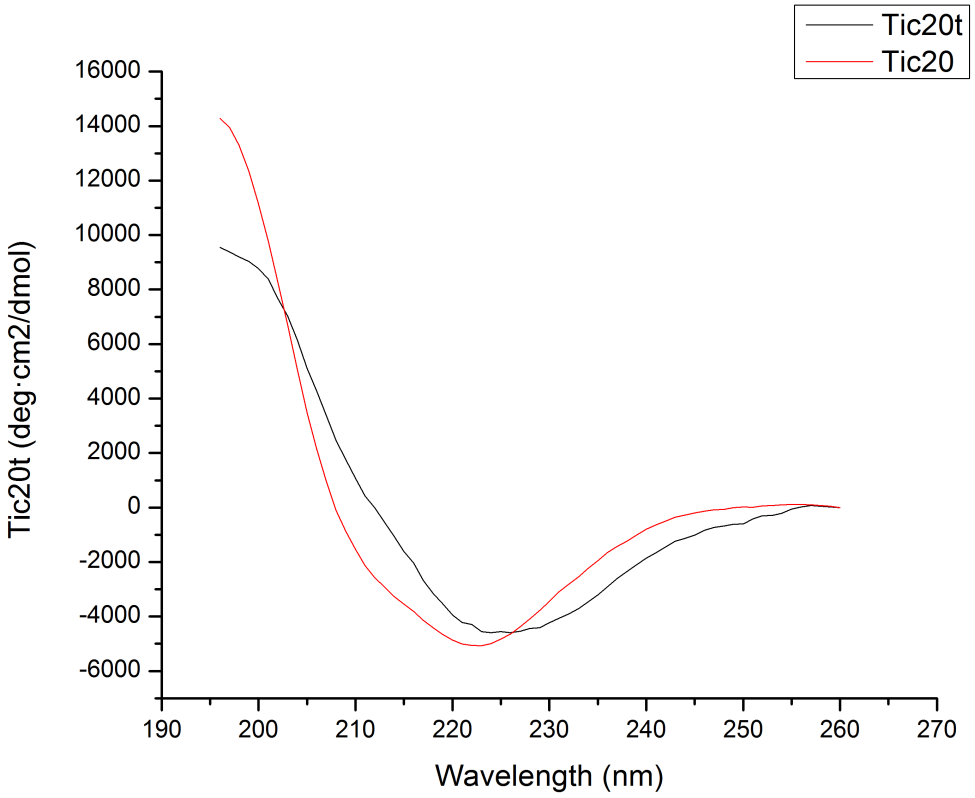


Figure 14. CD spectra of full length and truncated atTic20 in detergent. This graph shows the loss in intensity of the 208 minima in truncated atTic20. Standard buffer for all samples contained 150 mM NaCl and 10 mM Tris-HCl pH 8. Detergent concentration was 0.8mM of Zwittergent 3-14. The cuvette had a pathlength of 0.2mm.

3.3 Transit Peptide Interaction Study

3.3.1 Solid Phase Binding Assay

As mentioned in section 1.3, components of the TOC complex recognize preprotein transit peptides and possibly differentiate between photosynthetic and non-photosynthetic preproteins. It is possible that Tic20 serves a similar purpose at the inner membrane, however, it could also act as passive pore with other factors mediating protein cargo selection. To address this, a protein-protein binding assay was attempted to check for physical interactions between atTic20 and several transit peptides. Pilot studies using a methotrexate-agarose based system were initially suggestive of pull down of atTic20 by transit peptide containing fusion proteins (data not shown). However, despite several attempts at optimization, this system did not provide unambiguous evidence of an interaction between atTic20 and a transit peptide. To further test the possibility that Tic20 interacts directly with preprotein transit peptides, a switch was made to a more traditional solid phase protein-protein interaction assay using nickel-NTA agarose for immobilization of a bait protein and prey proteins radiolabelled using ^{35}S-Methionine to facilitate their detection. Figure 15 shows the results of experiments in which different amounts of bait were use to pull down two different transit peptide containing prey proteins. For these experiments, 50 μg of atTic20, or 50 μg of truncated atTic20, were used as the bait proteins immobilized on Ni-NTA resin. Figure 15 (top panel) shows binding of the small subunit of rubisco to the baits, while the bottom row shows binding of light harvesting complex component A4 (LHCA4; first lane shows *in vitro* translated prey protein). No binding was detected in any of the trials for any amount of bait tested.

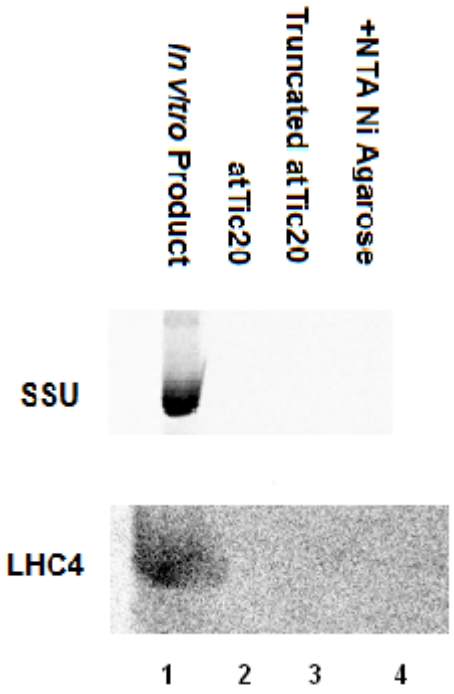


Figure 15. Solid phase binding of ^{35}S labelled transit peptide containing prey proteins. ^{35}S labelled TP containing prey proteins (SSU, top panel; LHC4-DHFR fusion protein, bottom panel) were incubated with Ni-NTA agarose with or without 50µg of atTic20 or truncated atTic20. Lane 1 contains 5% of the labelled prey protein that was added to each reaction.

3.3.2 Circular Dichroism

In addition to the binding assays, the potential interaction of transit peptides with atTic20 was studied using CD. A series of constructs consisting of dihydrofolate reductase (DHFR) fused to transit peptides had previously been generated in our lab for protein-protein studies. For this study the transit peptide from the LHCA4 protein fused to DHFR (DHFR-LHCA4TP) was used as the substrate and potential interaction partner for full-length mature atTic20 or truncated atTic20. First, the individual CD spectra were observed for the two atTic20 proteins, DHFR-LHCA4TP and DHFR alone. The spectra of atTic20 mixed in a 1:1 molar ratio with DHFR-LHCA4TP and DHFR were then collected and compared to the individual spectra. The same was done for truncated atTic20. By comparing the observed spectra of the protein mixtures with the mathematical sum of their individual spectra we would be able to detect conformational change(s) in one or both of the two proteins that were induced due to interaction between the two proteins, as a deviation from said mathematical sum (Wild et al, 1995). Figures 16 through 19 show the individual, mathematically summed and combined spectra of atTic20, truncated atTic20 and the DHFR-LHCA4TP fusion protein. In general, the experimentally observed combined spectra did not differ dramatically from the calculated sum of

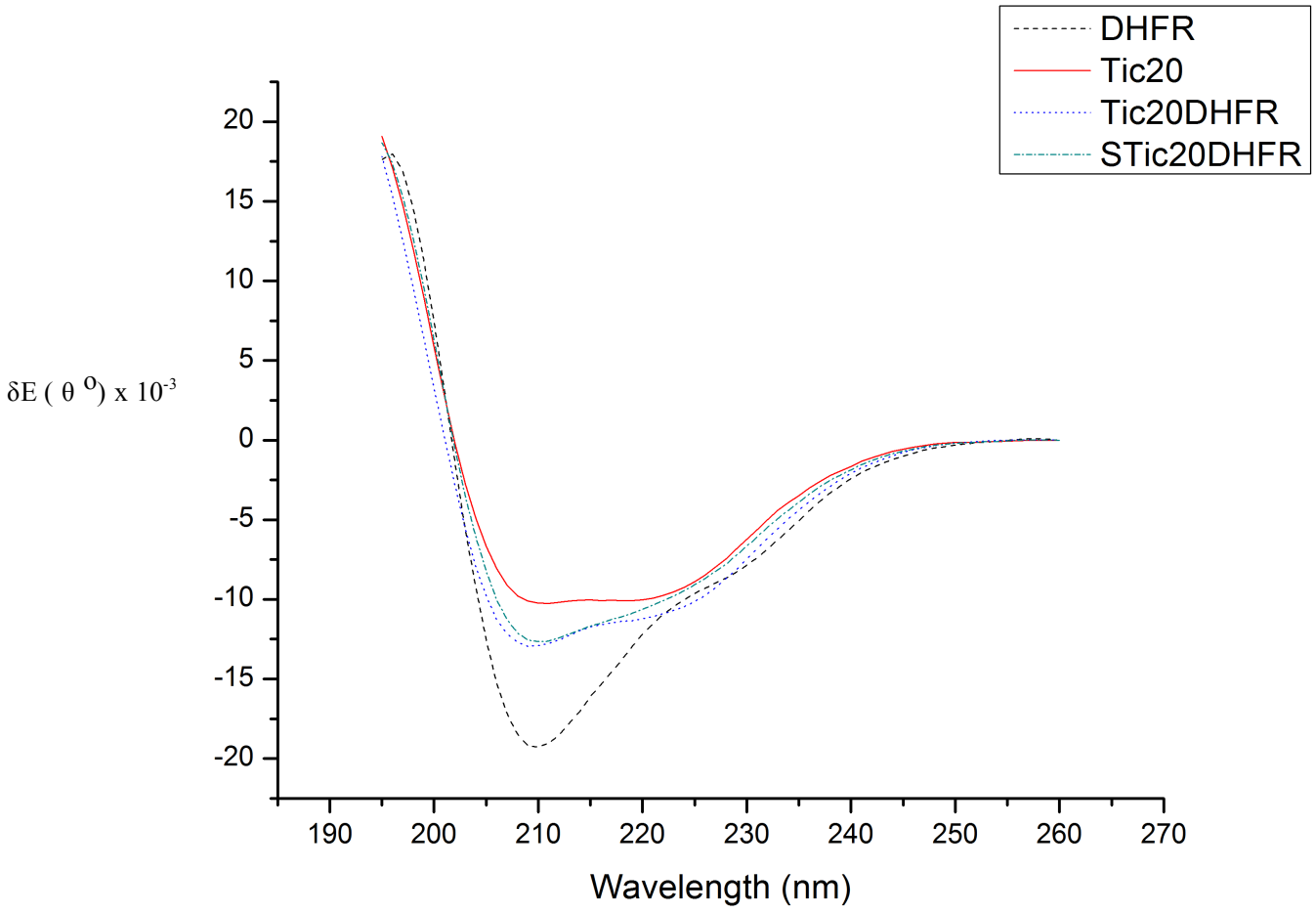


Figure 16:CD spectra of atTic20 and DHFR in combination and summed. Calculated as opposed to observed spectra is denoted by the addition of S to the line description. Standard buffer for all samples contained 2% (v/v) glycerol, 100 mM NaCl and 10 mM Tris-HCl pH 8. Detergent concentration was 0.8mM of Zwittergent 3-14. The cuvette had a pathlength of 0.2mm.

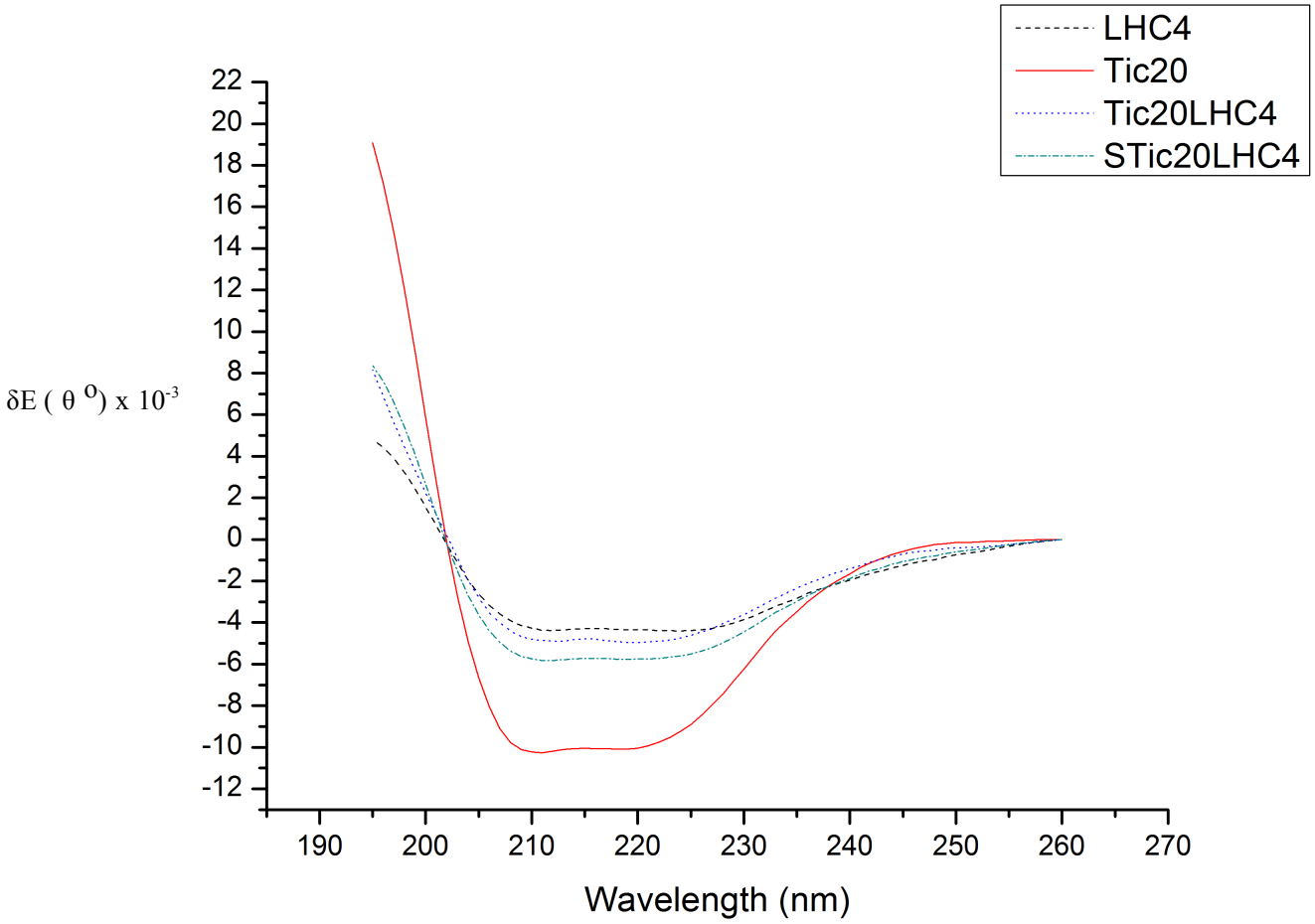


Figure 17:CD spectra of atTic20 and DHFR-LHC4 in combination and summed. Calculated as opposed to observed spectra is denoted by the addition of S to the line description. Standard buffer for all samples contained 2% (v/v) glycerol, 100 mM NaCl and 10 mM Tris-HCl pH 8. Detergent concentration was 0.8mM of Zwittergent 3-14. The cuvette had a pathlength of 0.2mm.

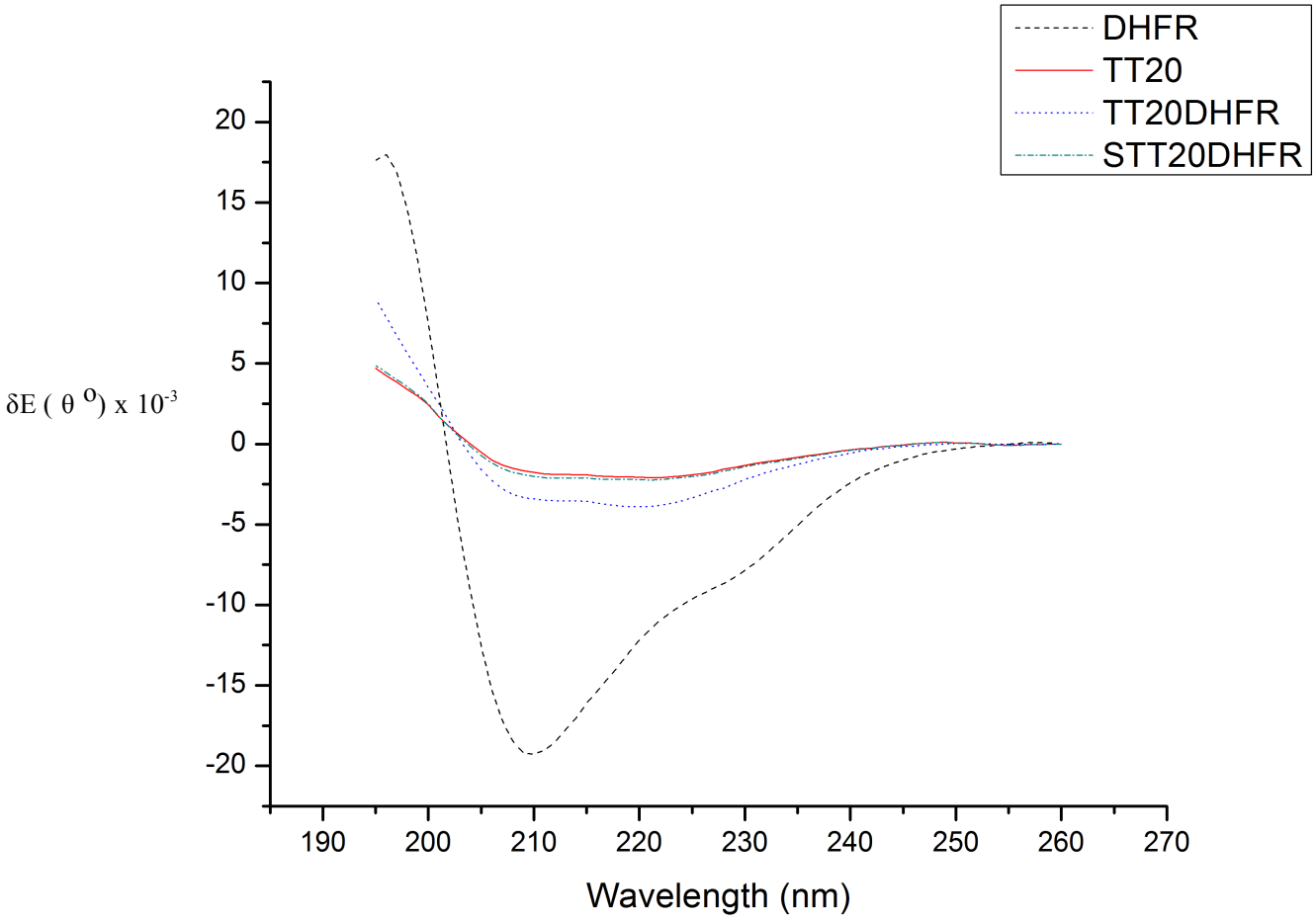


Figure 18:CD spectra of truncated atTic20 and DHFR in combination and summed. Calculated as opposed to observed spectra is denoted by the addition of S to the line description. Standard buffer for all samples contained 2% (v/v) glycerol, 100 mM NaCl and 10 mM Tris-HCl pH 8. Detergent concentration was 0.8mM of Zwittergent 3-14. The cuvette had a pathlength of 0.2mm.

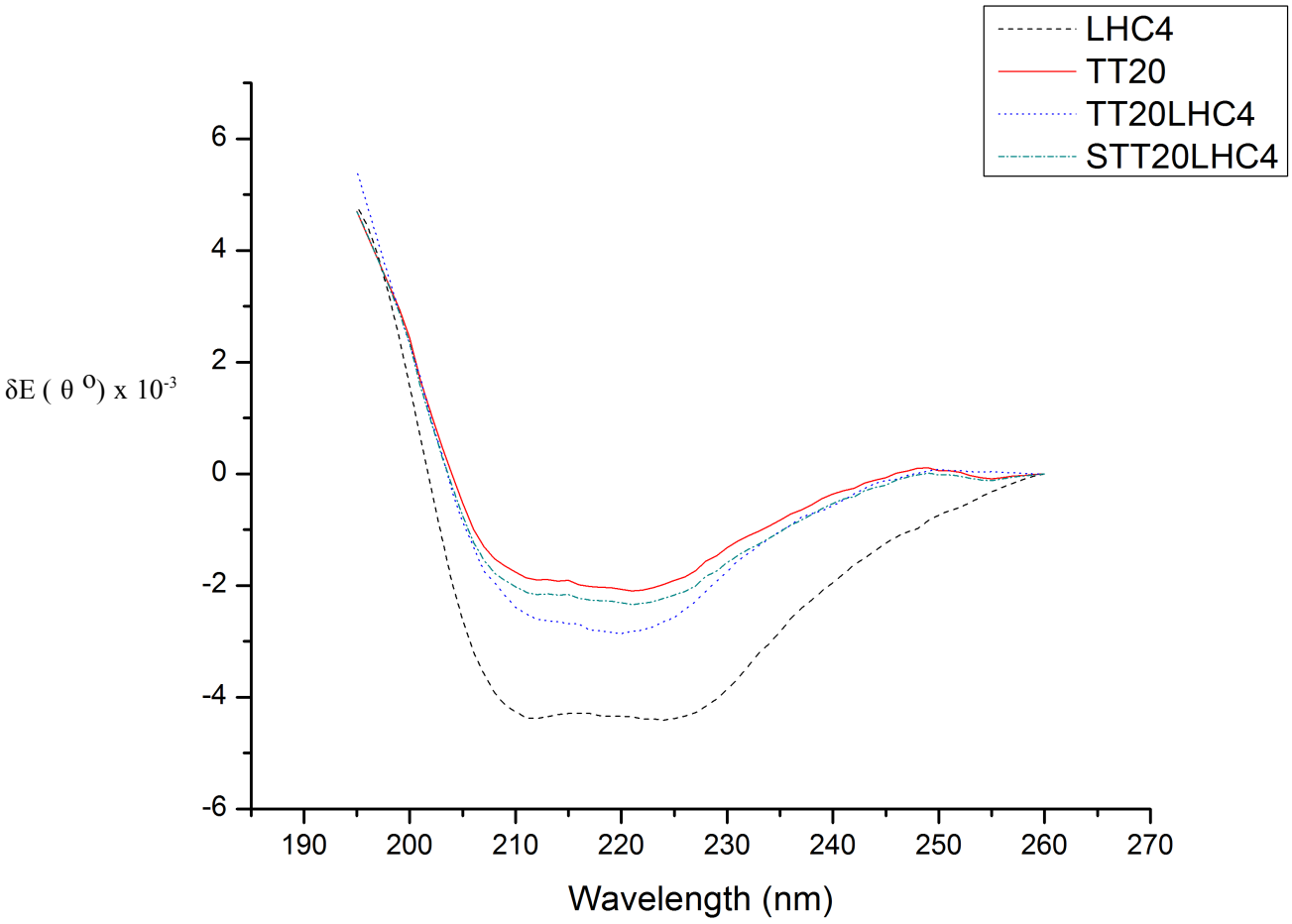


Figure 19:CD spectra of truncated atTic20 and DHFR-LHC4 in combination and summed. Calculated as opposed to observed spectra is denoted by the addition of S to the line description. Standard buffer for all samples contained 2% (v/v) glycerol, 100 mM NaCl and 10 mM Tris-HCl pH 8. Detergent concentration was 0.8mM of Zwittergent 3-14. The cuvette had a pathlength of 0.2mm.

the individual spectra; however, there are a few details that deserve note. While closely following the shape of the calculated spectra, the mixture of atTic20 and DHFR-LHCA4 (Fig. 17) shows some anomalous behaviour between 225 and 240 nm. In this range, the signal has a greater positive value (in millidegrees) than DHFR-LHCA4TP alone, let alone a combination of the two. This is also observed in the atTic20/DHFR mixture (Fig. 16), though in this case it has a greater negative value. This aberration is not observed in either of the truncation mutant treatments (Figs. 19 and 18). In the case of the atTic20/DHFR mixture, this deviation has the added effect of creating a local near-minima in the range of 220-230 nm, where calculations would otherwise suggest that the trace should have a consistent positive slope.

3.4 Circular Dichroism of N-terminal Peptide (NTP)

As mentioned in section 3.1, the 20 residue N-terminal segment of atTic20 is predicted to be intrinsically disordered, although it is short for an intrinsically disordered region (Fig. 20). To experimentally validate this prediction, we used CD on a chemically synthesized peptide corresponding to this segment of the protein to test three criteria characteristic of intrinsically disordered proteins. These criteria are that the peptide would gain structure under increasing trifluoroethanol (TFE) concentrations, with decreasing pH and with increasing temperature (Richardson et al, 2009; Smith and Jelokhani-Niaraki, 2012). This would indicate that under physiological conditions the domain is unstructured, but has the potential to gain structure in the presence of extrinsic factors (such as when associating with translocating preproteins). Figures 21 and 22 show the

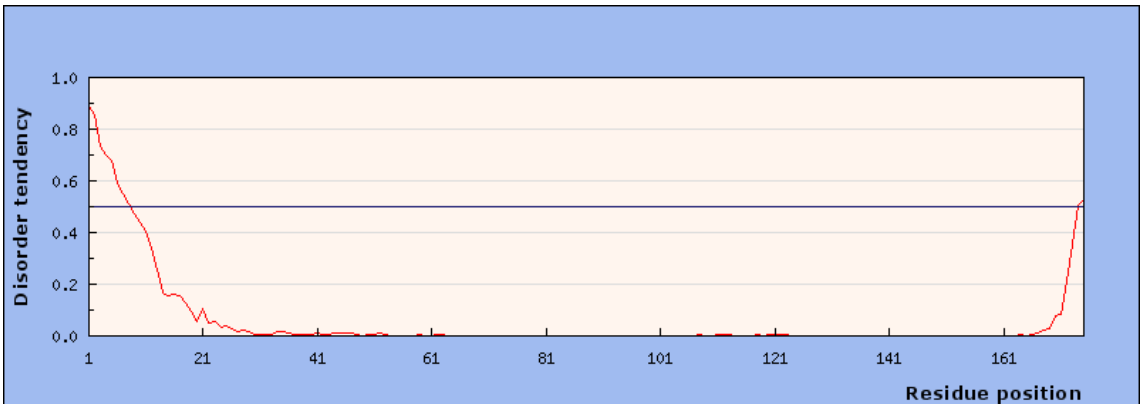


Figure 20. Plot of disorder tendency for the twenty amino acid N-teminal domain. Generated from IUPred using short disorder criteria: http://iupred.enzim.hu/. This tool generates a "disorder" score from zero to one based on homology to other disordered proteins.

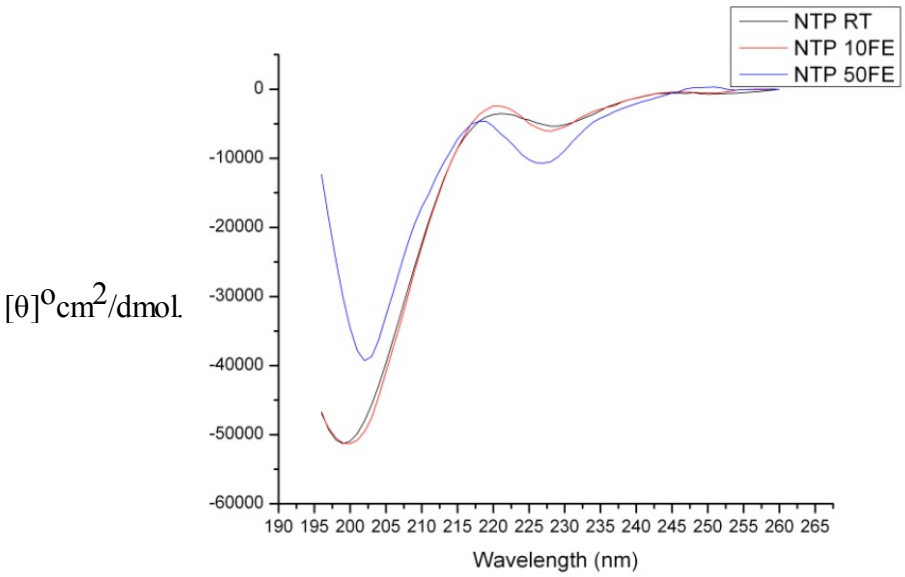


Figure 21:CD spectra of N-terminal peptide in different TFE concentrations. Loss of minima at 200 nm suggests gain of ordered secondary structure. Standard buffer for all samples contained 100 mM NaCl and 10 mM Tris-HCl pH 8. Detergent concentration was 0.8mM of Zwittergent 3-14. The cuvette had a pathlength of 0.2mm. cm2/dmol)

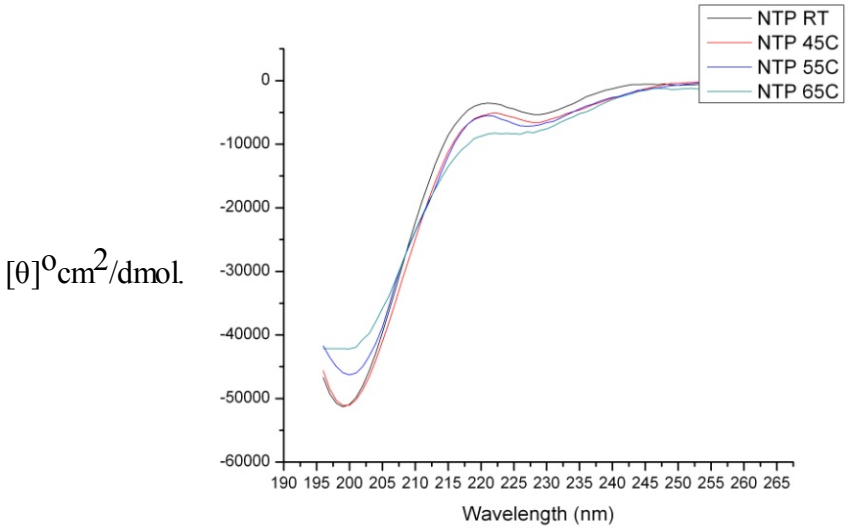


Figure 22:CD spectra of N-terminal peptide at different temperatures. Loss of minima at 200 nm suggests gain of ordered secondary structure. Standard buffer for all samples contained 100 mM NaCl and 10 mM Tris-HCl pH 8. Detergent concentration was 0.8mM of Zwittergent 3-14. The cuvette had a pathlength of 0.2mm. cm2/dmol)

results of our trials with these conditions. Gain of structure (or rather, loss of disorder) was observed in both the TFE and heat treatments, with it being more pronounced in increasing TFE concentrations (Figs. 21 and 22). In the TFE trials, the ellipticity at 200 nm changed from about -50,000^{o}cm^{2}/dmol to -35,000^{o}cm^{2}/dmol and at 230 nm from about -5,000^{o}cm^{2}/dmol to -10,000^{o}cm^{2}/dmol. In the elevated temperature trials, the same ellipticities changed to approximately -42,000^{o}cm^{2}/dmol and -5,000^{o}cm^{2}/dmol for 200 nm and 230 nm respectively at 65^{o}C. I was unable to be as comprehensive with our pH trials, however I was able to test very low pH (pH 2.7; data not shown). Decreasing pH did not result in the loss of disorder as seen in the other experiments, though we did observe a slight rightward shift of the spectra. The calculated pI of the N-terminal segment is high (~10) and therefore induction of structure was not predicted in the case of low pH. Despite multiple attempts, I was unable to produce a high pH Tris buffer that was compatible with CD.

3.5 Crystallization Trials

Initial crystallization trials had promising results. Our collaborator reported some positive hits during crystallization trials, with some microcrystal formation observed under some conditions. The conditions under which this formation occurred is not presently available to be reported in this document, however. Interestingly, when our collaborator attempted to further purify DDM solubilized samples we had sent him, gel filtration yielded a product containing a 60 kDa peak (Fig. 23). This could be a result of atTic20 forming complexes with detergent molecules, however it is also possible that Tic20 undergoes some form of oligerimization. An attempt to verify this using semi-

native and blue native PAGE was made, however it did not demonstrate oligomerization (data not shown), however earlier work by Spence MacDonald (unpublished) did show evidence of dimerization in blue native PAGE. Subsequent to adopting Zwittergent 3-14 as a working detergent, positive results in crystallization trials ceased. Our collaborator observed turbity in solubilized samples and precipitation of protein matter. I have also noticed that the protein tends to precipitate out of solution if left unfrozen for greater than one day when solubilized in Zwittergent 3-14, suggesting that atTic20 may not be stable in this detergent.

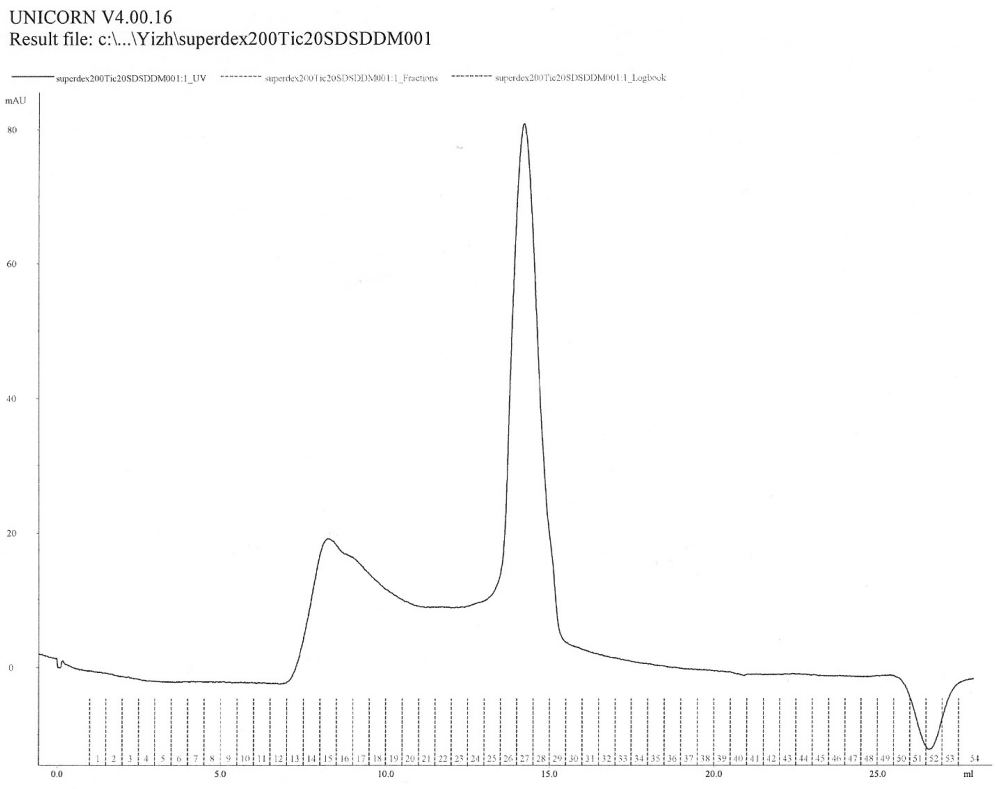


Figure 23. Refolded Tic20 chromatographed on a Superdex-200 column. Figure shows peak at 60 kDa. Prepared by R.M. Garavito.

4.0 Discussion

This study was undertaken to answer questions about atTic20; a protein that is essential for preprotein import into the chloroplast. Specifically, I wanted to determine if atTic20 interacted directly with the transit peptides of chloroplast preproteins, if this interaction occurred preferentially with photosynthetic preproteins and if the interaction was localized at the N-terminal segment of atTic20. This was assayed using purified, recombinant protein in solid phase protein binding trials and using circular dichroism spectroscopy. I also wanted to examine the structure of atTic20 using CD to see if it would provide insights into its hypothesized role as a translocation aperture. As the study evolved, it also became an opportunity to explore certain technical aspects of molecular biology and biochemistry, and allowed for some findings to be made both about the general practice of producing recombinant proteins in *E. coli*, and about the specific conditions for the production and purification of atTic20.

4.1 Autoinduction

When in its conceptual phase, this project seemed like it would be a fairly straightforward series of structural and functional analyses of atTic20; however, it quickly became apparent that generating a large enough yield of recombinant protein for these tests would be a major challenge. Much of the early work was therefore dedicated to optimizing protein yield and purity. A significant success in this area came with the switch to autoinducing media. Using the ZYP-5052 autoinducing media formulation resulted in consistently higher yields of atTic20 as compared to the more traditionally

used LB media combined with IPTG-based induction from the T7/lac promoter/operator. There are several possible explanations for this. The first and most obvious is that autoinducing medias are highly enriched in nutrients (in particular in terms of saccharides and additional minerals not present in LB - though it does have a comparable amount of amino acids and sodium); this would consequently allow for cultures to grow to very high densities. It is not atypical for a culture of *E. coli* to reach OD600 as high as 10 in this media (Studier, 2005). This explanation is at least partly consistent with our results, as the protein content in our ZYP-5052 cultures were higher in all conditions (Fig 6). It would not explain, however, why the high temperature LB cultures failed to produce any recombinant protein, despite having similar ODs and protein profiles to the cultures induced at room temperature. When he introduced the technique in his 2005 paper, Studier argued that autoinduction was particularly suited to the expression of proteins toxic to *E. coli* because the presence of glucose in the media represses leaky expression (already a strong feature of the T7 system). I favour this explanation for the increase in yield of recombinant atTic20. Not only are membrane proteins considered *a priori* more toxic to hosts for transgenic expression, but a number of TIC and TOC components are known to be either toxic to *E. coli* or are presumed to be toxic due to the inability to express them in an *E. coli* system (Tic110; Soll 2012). There is an interesting (if highly speculative) third possibility that may account for the change in the localization of atTic20 from inclusion bodies to the bacterial membrane that accompanies the switch to autoinducing media. It has been noted (Arechaga et al, 2000) that overexpression of the b subunit of the *E. coli* F1F0-ATP (ATPF) synthase complex causes a proliferation of intracellular membrane vesicles and accompanying, adjoining membrane threads in *E.*

coli K-12. This phenomenon has been harnessed as a tool for membrane protein expression where the protein of interest is co-expressed with a plasmid containing the ATPF gene (Zoonens and Miroux, 2010). In this system, the protein of interest expresses to the intracellular and cytosolic membrane, but not into inclusion bodies (Zoonens and Miroux, 2010). While parity in expression localization may be a fairly superficial line of evidence, it has been found that the presence of glycerol in the growth media results in an upregulation of the operon governing translation of ATPF in *E. coli* (Kasimoglu et al, 1996), which could lead in turn to proliferation of the intracellular vesicles observed by Arechaga et al (2000) . Whether or not this is in fact occurring in autoinducing media could be easily verified by microscopy, as the membrane proliferation described by Arechaga et al(2000) was clearly visible in electron micrographs.

While the succes of autoinduction with regards toTic20 is significant on its own, the mechanism behind the increase in yield may have some important ramifications. With regards to the chloroplastic protein import apparatus, atTic20 is far from the only member of the family to exhibit some recalcitrance in expressing in *E. coli* systems. Tic110 and full length Toc159 for example all resist expression to some degree in bacterial systems (Soll, 2012, Richardson, 2009). If in fact autoinduction does mitigate toxicity by preventing leaky expression, it may be possible to re-visit other members of the complex and devise a protocol that allows for high density culture growth prior to induction and permit a type of “one-off” transgenic protein expression. This would allow for a small amount of these proteins to be produced for study in scenarios where cell-free or synthetic systems are either unsuitable or cost-prohibitive. More generally, understanding in detail the mechanistic properties of varied expression strategies aids in standardizing

protein expression and purification. Of all the biomolecules produced by molecular techniques, proteins have the least generalized procedures for production and isolation to the point where a new protocol has to be devised and optimized for individual proteins. Understanding why a given expression or purification strategy succeeded or failed for a particular protein helps lay the foundation for a systematic approach for protein methods

4.2 Membrane Isolation and Detergent Selection

After optimizing our expression strategy, we moved on to choosing an appropriate detergent for extracting recombinant atTic20 from the bacterial membrane. Our criteria for the detergent screen was to select for candidates that solubilized large amounts of protein while not dramatically altering the character of the protein's CD spectra; in summary to maximize yield. Based strictly on this criteria, Zwittergent 3-14 was the optimal detergent of the six that were compared as it solubilized as much if not more recombinant atTic20 as the other detergents tested, and the isolated protein had CD spectra of good character (that is, spectra that shows alpha-helical content and of a strength comparable to that exhibited by reconstituted atTic20). Zwittergent 3-14 also was the only detergent to completely solubilize the isolated membrane, leaving no visible pellet after the centrifugation following solubilization (data not shown). Unfortunately, further work has shown that atTic20 may not be entirely stable in Zwittergent 3-14. In his report to us, our crystallization collaborator noticed that the Zwittergent 3-14-solubilized recombinant atTic20 that we sent him slowly precipitated out of solution when left unfrozen. This was also noted recently by Tuan Hoang (personal communication), who has been doing CD work on the protein in a neighbouring lab. This may be a

consequence of lower salt concentration in the buffer, since in the latter case chloride needs to be removed for CD, and in the former the protein needs to be concentrated - effectively lowering the salt concentration relative to that of atTic20. Furthermore, this precipitation has not been observed in samples stored in elution buffer, which has high concentrations of both NaCl (300 mM) and imidazole (300 mM). Since the salt concentration needs to be lowered regardless of whether or not it is responsible for maintaining stability of atTic20 in Zwittergent 3-14, it will be necessary to look for an alternative detergent to use for storage and working conditions. DDM may be a good candidate, as previous work with atTic20 in inclusion bodies showed that atTic20 was stable in that detergent (MacDonald and Smith, unpublished). Since it has been shown to be effective for purification, Zwittergent 3-14 should continue to be used for membrane solubilization and then be exchanged for a new detergent on the IMAC column. If further detergent screens are conducted, it would be productive to test against detergents at concentrations standardized by their critical micellar concentrations (CMCs) rather than absolute concentrations as used in our assay. This would allow us to make more meaningful comparisons across different detergents.

4.3 Transit Peptide Interaction Studies

This study has found no compelling evidence that atTic20 interacts with transit peptides, either at its N-terminus or at the exposed portions of its helical domains. This seems to be in line with existing arguments that atTic20 has a largely structural role in the complex that translocates preproteins across the inner membrane (Chen et al, 2002). It would still be wise however to seek further evidence as confirmation. In our study, we

used transit peptides from SSU and LHCA4, proteins that are considered photosynthetic. It has been shown that atToc159 and atToc132 can interact with slightly different, but largely overlapping groups of preprotein substrates, and that the relative affinities for some preproteins are different (Dutta et al, 2014) A solid phase binding assay in the vein of what was conducted in this study should be attempted for a housekeeping preprotein, in order to determine whether atTic20 might be specific for one functional group of proteins. It's also possible that the pull-down technique used to test for an interaction between atTic20 and preprotein TPs was not sensitive enough to detect what is presumably a transient interaction. There are a number of factors that could have confounded interactions. In the solid phase binding assay, the high ionic strength of the buffer could have weakened electrostatic interactions between bait and prey in the assay. In both the binding assays and the additive CD study, the proteins studied were in detergent. Since these proteins were in detergent micelles, they may not have been sufficiently exposed to interact. The solid phase binding assay could have been improved by adding a cross-linking agent to overcome transience of interactions. Both the solid phase binding assay and the additive CD would be more effectively done with bait proteins in liposomes and prey protein soluble in buffer. Furthermore, there are other ways to test this interaction. Since atTic20 is enriched in tryptophan residues, whereas the transit peptide of SSU lacks tryptophan residues, using changes in intrinsic tryptophan fluorescence might provide a more sensitive method for detecting an interaction (Vivian and Callis, 2001). A shift in fluorescence would indicate a change in exposure of atTic20's tryptophan residues to solution, which would indirectly suggest an interaction between the peptide and the protein. Because of its length (58 aa; Reiss,

1987), acquiring a synthetic version of the SSU transit peptide would not be trivial, however recent work by Lee et al (2009) has found that certain regions on the order of less than 20 aa in size have been implicated in interactions with the import machinery. This would allow us to test interactions with smaller, more easily synthesized fragments of the transit peptide. Another possibility is the use of a yeast two hybrid system to look for an interaction – this would allow for testing against a much wider range of transit peptide containing prey proteins.

4.4 Structure of atTic20, Intrinsic Disorder and the Role of the N-terminal Peptide

CD on atTic20 shows that it has significant α-helical characteristics in detergent. Deconvolution of CD of atTic20 reconstituted into liposomes suggests atTic20 is at least 46% α-helical. Though not as high, a plurality of the protein remains α-helical in detergent. Qualitatively, the CD spectra of the protein (both truncated and full length) has the hallmark 208 nm and 222 nm minima of an α-helical protein. This compares favourably with the *in silico* models, although deconvolution gives percentages that are less than 50% while in contrast the models show the protein is at most 69% α-helical. This may be due to the lack of sufficiently homologous proteins in the deconvolution reference sets, however. Both the model and the generally insoluble nature of the protein suggests that these helices are membrane bound, supporting existing studies placing atTic20 at the inner envelope membrane.

My preliminary work indicates that the N-terminal segment of atTic20 is intrinsically disordered. CD spectra of full length atTic20 generally shows a strong minima at 222 nm and a slightly less intense minima at 208 nm. This 208 nm “shoulder”

may be indicative of oligomerization of atTic20 (Hoang et al, 2013). However, deletion of the N-terminal segment in the truncation mutant consistently shows a stronger reduction of the intensity of this minima. This could be an indication that the N-terminal region plays a role in stabilizing the helices or that it inhibits oligomerization of the molecule. Because the N-terminal segment shows such strong disorder on its own, it could also simply be that its presence produces a strong signal that otherwise masks this shoulder in the context of the full-length protein.Since it shows strong disorder under physiological pH and temperature conditions, but is capable of assuming some ordered structure when exposed to TFE or under elevated temperatures, it meets the criteria of an intrisically disordered protein domain.

Because intrinsically disordered proteins are believed to be involved in dynamic protein-protein interactions, and because of its presence in a preprotein translocating complex, it is reasonable to hypothesize that this portion of the protein is involved in an interaction of atTic20 with another protein partner. Which protein, however, is open for debate. As mentioned previously (section 4.3), we were not able to provide evidence that it serves as a mediator of interaction with translocating preproteins. One possibility is that it serves as some sort of anchor to another member of the complex. The recent discovery of three new members of the TIC complex (Kikuchi et al, 2013) implies there are a host of candidates that need to coordinate their activity with atTic20 in order to function in preprotein translocation. Determining which of these directly interact with atTic20, and whether the N-terminal peptide is significant for this interaction, is a daunting task. A “brute force” approach which would involve performing solid phase binding assays using both atTic20 and truncated atTic20 with each partner from the complex, would be

labourious but could yield significant insights. A more ambitious approach would be to engineer transgenic *Arabidopsis thaliana* containing the truncated version of atTic20 in place of the full-length protein. This would not be easy, as deletion of atTic20 is lethal (Chen et al, 2002). It would require producing a plant that is hemizygous for atTic20 and then transforming with the truncation mutant. If successful, however, these plants could be very informative. Not only would it be possible to determine if the N-terminal disordered region is required for association with other complex members (by way of comparison with earlier studies; Kikuchi et al, 2013) it could also provide clues by way of phenotypic changes that may not involve its association with other complex members at all. This would complement earlier studies that found the deletion of atTic20 to be lethal and its knock down by anti-sense RNA leading to a pale phenotype commonly seen amongst studies compromising some element of the import apparatus (Chen et al, 2002). It could also address questions about the function of IDP domains in general as producing a mutant lacking only the supposed disordered region of atTic20 would allow one to directly correlate physiological effects with the disordered region. It would also be a way of providing material (i.e. the isolated but assembled TIC complex) for studies regarding the role the disordered domain has in assembling the complex. Furthermore, chloroplasts could be isolated from the mutant plants and import assays could be conducted to determine if the region is directly involved in import by examining if preproteins are conducted into the chloroplast, or to see if the region has a role in recruiting stromal side processing factors by examining if in fact post import preproteins have their transit peptide sequence cleaved. Finally, studies conducted on material from transformed plants would have the advantage of having sample materials under “physiologically relevant”

conditions - proteins for example would have any necessary post-translational modifications and membranes would have native lipid components.

4.5 Integration of Approaches

Molecular biology is, in some ways, a necessarily integrative type of biology. It draws heavily on chemistry for most technical considerations and could be thought of as a subset of biochemistry. But the questions it asks are rarely chemical ones, rather it is more often used to explore hypotheses about evolution or physiology. It is in this tradition that I pursued my study, using chemical tools to answer questions about chloroplast form and function. But this study also required the use of biophysical approaches, namely CD and (unsuccessfully) X-ray crystallography. For the portions of the study involving the examination of intrinsic disorder, this was almost a foregone conclusion - the entire concept of intrinsic disorder owes its discovery to CD techniques (Uversky et al, 2000). But it also yielded insights in areas where it is not typically used. For months, the solid phase binding assays failed to produce clear results. A second avenue of investigation was required to confirm what data I had at the time. Circular dichroism provided that avenue. If there is a lesson here to be learned about the practice of science, it is that it is in the liminal spaces between disciplines that fosters the most potential for new techniques to be developed and new discoveries to be made.

References

Arabidopsis Genome Initiative. 2000. Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. Nature 408:796-815

Arechaga I, Miroux B, Karrasch S, Huijbregts R, de Kruijff B, Runswick MJ, Walker JE. 2000. Characterisation of new intracellular membranes in Escherichia coli accompanying large scale over-production of the b subunit of F1Fo ATP synthase. FEBS Letters 482:215-219

Aronsson H,. and Jarvis P, (2009) The chloroplast protein import apparatus, its components, and their roles. In: Plant Cell Monographs, Vol. 13: The Chloroplast – Interactions With The Environment (A.S. Sandelius and H. Aronsson, eds.) Springer-Verlag, Heidelberg, Germany, pp. 89-123.

Balsera M, Goetze TA, Kovács-Bogdán E, Schurmann P, Wagner R, Buchanan BB, Soll J, Bölter B. 2009. Characterization of Tic110, a channel-forming protein at the inner envelope membrane of chloroplasts, unveils a response to Ca2+ and a stromal regulatory disulfide bridge. The Journal of Biological Chemistry 284(5): 2603-2616.

Bauer J, Chen K, Hiltbunner A, Wehrli E, Eugster M, Schnell D, Kessler F. 2000. The major protein import receptor of plastids is essential for chloroplast biogenesis. Nature 403(13):203-207

Becker T , Jelic M , Vojta A , Radunz A , Soll J , Schleiff E. 2004.Preprotein recognition by the Toc complex. The EMBO Journal 23:520-530

Chen X, Smith MD, Fitzpatrick L, Schnell D. 2002. In vivo analysis of the role of atTic20 in protein import into chloroplasts. The Plant Cell 14: 641-654.

Cline K, Henry R. 1996. Import and routing of nucleus-encoded chloroplast proteins. Annual Reviews of Cell Developmental Biology 12:1-26.

Correa DHA, Ramos CHI. 2009. The use of circular dichroism spectroscopy to study protein folding, form and function. African Journal of Biochemistry Research 3(5):164-173

Dyson HJ, Wright PE. 2005. Intrinsically Unstructured Proteins and their Functions. Nature Reviews Molecular and Cell Biology 6(3):197-208

Fink AL. 2005. Natively Unfolded Proteins. Current Opinion in Structural Biology 15(1):35-41

Gaxiola RA, Fink GR, Hirschi KD. 2002. Genetic manipulation of vacuolar proton pumps and transporters. Plant Physiology 129: 967-973.

Heins L, Mehrle A, Hemmler R, Wagner R, Kuchler M, Hormann F, Sveshnikov D, Soll J. 2002. The preprotein conducting channel at the inner envelope membrane of plastids. The EMBO Journal 21(11):2616-2625

Hirabayashi Y, Kikuchi S, Oishi M, Nakai M. 2011. In vivo studies on the roles of two closely related Arabadopsis Tic20 proteins, atTic20-I and atTic20-IV. Plant Cell Physiology 52(3): 469-478

Hoang T, Smith MD, Jelokhani-Niaraki M. 2013. Expression, folding, and proton transport activity of human uncoupling protein-1 (UCP1) in lipid membranes: Evidence for associated functional forms. The Journal of Biological Chemistry 288(51), 36244-36258.

Inaba T, Alvarez-Huerta M, Li M, Bauer J, Ewers C, Kessler F, Schnell DJ. 2005. Arabidopsis Tic110 is essential for the assembly and function of the protein import machinery of plastids. The Plant Cell. 17(5):1482-1496.

Ivanova Y, Smith MD, Chen K, Schnell D. 2004. Members of the Toc159 import receptor family represent distinct pathways for protein targeting to plastids. Molecular Biology of the Cell 15:3379–3392

Jarvis P, Soll, J. 2001. Toc, Tic, and chloroplast protein import. Biochimica et Biophysica Acta 1541: 64-97.

Jackson DT, Froehlich JE, Keegstra K. 1998. The hydrophilic domain of Tic110, an inner envelope membrane component of the chloroplastic protein translocation apparatus, faces the stromal compartment. The Journal of Biological Chemistry 273:16583-16588.

Jelokhani-Niaraki M, Ivanova MV, McIntyre BL, Newman CL, McSorley FR, Young EK, Smith MD. 2008. A CD study of uncoupling protein-1 and its transmembrane and matrix-loop domains. Biochemical Journal 411:593-603.

Kasimoglu, E., S.J. Park, J. Malek, C.P. Tseng, R.P. Gunsalus 1996. Transcriptional regulation of the proton-translocating ATPase (atpIBEFHAGDC) operon of Escherichia coli: control by cell growth rate. Journal of Bacteriology 178:5563-5567

Kasmati AR, Topel M, Patel R, Murtaza G, Jarvis P. 2011. Molecular and genetic analyses of Tic20 homologues in Arabidopsis thaliana chloroplasts. The Plant Journal 66:877-889.

Keeling P. 2010. The endosymbiotic origin, diversification and fate of plastids. Philosophical Transactions of the Royal Society B. 365: 729-748

Kikuchi S, Oishi M, Hirabayashi Y, Lee DW, Hwang I, Nakai M. 2009. A 1-megadalton translocation complex containing Tic20 and Tic21 mediates chloroplast protein import at the inner envelope membrane. The Plant Cell 21(6): 1781-1797

Kikuchi S, Bédard J, Mirano H, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, and Nakai M. 2013. Uncovering the Protein Translocon at the Chloroplast Inner Envelope Membrane. Science 339 (6119), 571-574.

Kouranov A, Schnell D. 1997. Analysis of the interactions of preproteins with the import machinery over the course of protein import into chloroplasts. The Journal of Cell Biology 139(7):1677-1685.

Kouranov A, Chen X, Fuks B, Schnell D. 1998. Tic20 and Tic22 are new components of the protein import apparatus at the chloroplast inner envelope membrane. The Journal of Cell Biology 143(16):991-1002.

Kovács-Bogdán E, Benz JP, Soll J, Bölter B. 2011. Tic20 forms a channel independent of Tic110 in chloroplasts. BMC Plant Biology. 11(133): 1-40.

Kubis S, Baldwin A, Patel R, Razzaq A, Dupree P, Lilley K, Kurth J, Leister D, Jarvis P. 2003. The Arabidopsis ppi1 mutant is specifically defective in the expression, chloroplast import and accumulation of photosynthetic proteins. The Plant Cell 15(8):1859-1871

Lee DW, Lee S, Oh YJ, Hwang I. 2009. Multiple Sequence Motifs in the Rubisco Small Subunit Transit Peptide Independently Contribute to Toc159-Dependent Import of Proteins into Chloroplasts. Plant Physiology 151(1): 129–141.

Machettira AB, Gross LE, Sommer MS, Weis BL, Englich G, Tripp J, Schleiff E. 2011. The localization of Tic20 proteins in Arabidopsis thaliana is not restricted to the inner envelope membrane of chloroplasts. Plant Molecular Biology 77:381-390.

Pollastri G, McLysaght A. 2005. Porter: a new, accurate server for protein secondary structure prediction. Bioinformatics 21(8):1719-1720

Perry SE , Keegstra K. 1994. Envelope membrane proteins that interact with chloroplastic precursor proteins . Plant Cell 6:93-105

Richardson LGL, Jelokhani-Niaraki M, Smith MD. 2009. The acidic domains of the Toc159 chloroplast preprotein receptor family are intrinsically disordered protein domains. BMC Biochemistry 10(35)

Sagan L. 1967. On the origin of mitosing cells. Journal of Theoretical Biology 14(3): 255-74

Smith MD, Hiltbrunner A, Kessler F, Schnell D. 2002. The targeting of the atToc159 preprotein receptor to the chloroplast outer membrane is mediated by its GTPase domain and is regulated by GTP. The Journal of Cell Biology 159(5):833-843.

Smith MD, Jelokhani-Niaraki M. 2012. pH Induced Changes in Intrinsically Disordered Proteins. In: Uversky V, Dunker AK editors. Intrinsically Disordered Protein Analysis. Humana Press 223-233.

Studier FW. 2005. Protein production by auto-induction in high-density shaking cultures. Protein Expression & Purification 41: 207-234

Uversky V N, Gillespie J R, Fink A L. Why are “natively unfolded” proteins unstructured under physiologic conditions?Proteins: Structure, Function, and Bioinformatics 41(3): 415-427

Vivian JT, Callis PR. 2001.Mechanisms of tryptophan fluorescence shifts in proteins.Biophysical Journal 80(5):2093-109.

Vullo A, Walsh I, Pollastri G. 2006. A two-stage approach for improved prediction of residue contact maps. BMC Bioinformatics, 7:180

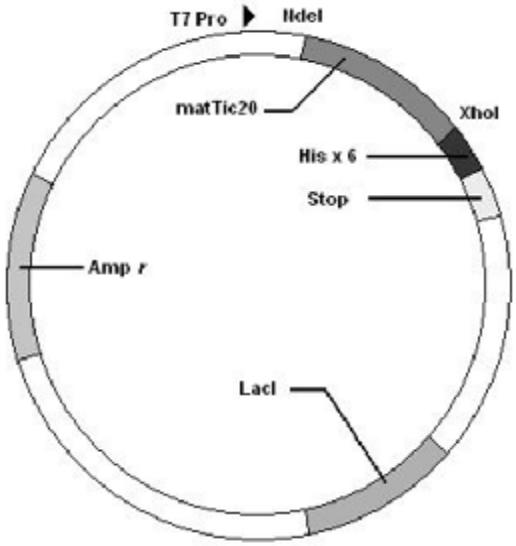
Wild C, Greenwell T, Shugars D, Rimsky-Clarke L, Matthews T. 1995. The Inhibitory Activity of an HIV Type 1 Peptide Correlates with Its Ability to Interact with a Leucine Zipper Structure. AIDS Research and Human Retroviruses 11(3):323-325

Wise R. 2006. The Structure and Function of Plastids. In: Advances in Photosynthesis and Respiration 23:3-26

Yan J, Campbell JH, Glick BR, Smith MD, Liang Y. 2014. Molecular Characterization and Expression Analysis of Chloroplast Protein Import Components in Tomato (Solanum lycopersicum).PLoS One 9(4):e95088

Zoonens M, Miroux B. Expression of membrane proteins at the E. coli membrane for structural studies. In: Veteau IM, editors. Methods in Molecular Biology #601: Heterologous Expression of Membrane Proteins: Methods and Protocols. Humana Press. 2010.

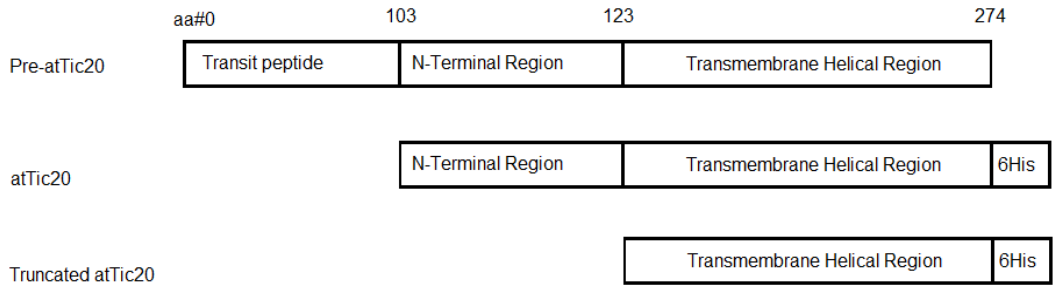
Appendix I: atTic20 Clone and Vector



Appended Figure 1. pET21a Plasmid containing matTic20his cDNA. Plasmid prepared by Spence MacDonald.

atTic20 is located at locus AT1G04940 and has TAIR accession #2010617. Data provided by The Arabidopsis Information resource – see http://www.arabidopsis.org/servlets/TairObject?id=27271&type=locus for reference and full length genomic sequece.

Appendix II: Schematics of atTic20 Constructs



This schematic shows the atTic20 constructs used in this study. Pre-atTic20 is the orignal template cDNA; it contains the Tic20 transit peptide. AtTic20 lacks the transit peptide and starts at amino acid 103 of the original protein, as well as having a hexahistidine tag attached at the C-terminus. Truncated atTic20 has an additional 20 amino acids removed from the N-terminus that are predicted to form an intrinsically disordered segment.