Inhibition of TDP-43 Aggregation using Native-State Binding Ligands

by

Yulong Sun

A thesis submitted in conformity with the requirements for the degree of Master of Science
Graduate Department of Biochemistry
University of Toronto

© Copyright by Yulong Sun 2014
Inhibition of TDP-43 Aggregation using Native-State Binding Ligands

Yulong Sun

Master of Science

Department of Biochemistry
University of Toronto

2014

Abstract

TAR DNA binding protein of 43 kDa (TDP-43) has been implicated in the pathogenesis of amyotrophic lateral sclerosis and frontotemporal lobar degeneration. Pathologically misfolded and aggregated forms of TDP-43 are found in cytoplasmic inclusion bodies of affected neurons in these diseases. The mechanism by which TDP-43 misfolding causes disease is not well understood. We postulate that the aggregation process plays a major role in pathogenesis, and we hypothesize that oligonucleotide ligands of TDP-43 can stabilize the native functional state of the protein and ameliorate aggregation of this aggregation-prone protein. Using recombinant TDP-43 we were able to examine the extent to which various oligonucleotide molecules affects its aggregation in vitro. We have found that certain natural sequence and de novo designed oligonucleotides bind TDP-43 and prevent its natural tendency to aggregate. The clinical and therapeutic implications of these findings are discussed.
Acknowledgments

This study was made possible by a number of individuals. I thank past and present members of the Chakrabarty Lab - Dr. Aaron Kerman, Dr. Pharhad Arslan, Dr. Vikram Mulligan, Kevin Hadley, Philbert Ip, Qasim Khan, Priya Sharda and Antoinette Bugyei-Twum - for insightful discussions. The vector containing human TDP-43 sequence was a kind gift from Dr. Janice Robertson. The recombinant vYFP-TDP-43 plasmid was constructed by Dr. Pharhad Arslan. Atomic force microscopy experiments were performed by Dr. Amy Won and Dr. Christopher Yip. I thank my supervisory committee members Dr. Emil Pai and Dr. Janice Robertson for their guidance and critical evaluation of this work. I thank Dr. Avi Chakrabarty for his supervision and the opportunity to partake in this study. I also thank the Alzheimer’s Society of Canada, the Ontario Government, and Canadian Institute of Health Research for their funding.
Table of Contents

Acknowledgments iii

Table of Contents iv

List of Figures vi

Chapter 1: Introduction

1.1 Background 1

1.2 Genetic Factors Implicated in ALS and Related Neurodegenerative Diseases 1

1.2.1 Role of SOD1 in ALS 1

1.2.2 TDP-43 in ALS and Related Neurological Diseases 5

1.2.3 FUS/TLS in FTLD/ALS Related Diseases 9

1.2.4 C9orf72 Repeat Expansions in FTLD/ALS 11

1.3 Experimental Rationale 12

Chapter 2: Material and Methods 16

2.1 Protein Expression and Purification 16

2.2 Dynamic Light Scattering Measurements 17

2.3 Urea Denaturation Experiment 17

2.4 Circular Dichroism Spectroscopy 17

2.5 In vitro aggregation of vYFP-TDP-43 18

2.6 Right Angle Light Scattering 18

2.7 Fluorescence Microscopy 18
2.8 Atomic Force Microscopy

Chapter 3: Results

3.1 Sequence Selection of Inhibitors of TDP-43 Aggregation
3.2 Recombinant Expression of TDP-43
3.3 Aggregation and Aggregation Inhibition of TDP-43 by Oligonucleotides
3.4 Irreversible Nature of vYFP-TDP-43 Aggregation
3.5 Morphology of TDP-43 Aggregates

Chapter 4: Discussion

4.1 TDP-43 Aggregation is Inhibited by Native State Binding Ligands
4.2 Applications of Phosphorothioate DNA Derivatives
4.3 Mechanisms of TDP-43 Aggregation
4.4 Implications of TDP-43 Aggregation Inhibition by Naturally Occurring Targets

Chapter 5: Future Directions

References
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Proposed hypothesis of aggregation inhibition through the binding of a DNA target</td>
<td>15</td>
</tr>
<tr>
<td>2</td>
<td>Names and sequences of nucleotides used for native-state stabilization</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>SDS-PAGE of purified vYFP-TDP-43</td>
<td>23</td>
</tr>
<tr>
<td>4</td>
<td>Monitoring vYFP-TDP-43 expression in <em>E. coli</em></td>
<td>24</td>
</tr>
<tr>
<td>5</td>
<td>Circular dichroism spectrum of vYFP-TDP-43</td>
<td>26</td>
</tr>
<tr>
<td>6</td>
<td>Size distribution by mass of purified vYFP-TDP-43</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>Urea denaturation of vYFP-TDP-43</td>
<td>29</td>
</tr>
<tr>
<td>8</td>
<td>Size distribution by mass of vYFP-TDP-43 upon aggregation determined by dynamic light scattering</td>
<td>31</td>
</tr>
<tr>
<td>9</td>
<td>Circular dichroism spectrum of TG$_{12}$, vYFP-TDP-43, and their mixture</td>
<td>32</td>
</tr>
<tr>
<td>10</td>
<td>Inhibition of vYFP-TDP-43 aggregation using TG$_{12}$ monitored by right angle light scattering</td>
<td>34</td>
</tr>
<tr>
<td>11</td>
<td>Inhibition of vYFP-TDP-43 aggregation using varying lengths of TG-repeats</td>
<td>36</td>
</tr>
<tr>
<td>12</td>
<td>Inhibition of vYFP-TDP-43 aggregation using various natural oligonucleotide binding targets</td>
<td>37</td>
</tr>
<tr>
<td>13</td>
<td>Fluorescence microscopy of vYFP-TDP-43 aggregates</td>
<td>39</td>
</tr>
<tr>
<td>14</td>
<td>Atomic force microscopy of vYFP-TDP-43 aggregates</td>
<td>41</td>
</tr>
</tbody>
</table>
Chapter 1

1 Introduction

1.1 Background

Amyotrophic lateral sclerosis (ALS), also known as Lou Gehrig’s disease or motor neuron disease, is a fatal neuromuscular disease presenting as weakness, muscle atrophy, and spasticity, which is caused by selective degeneration of motor neurons in the brain, brainstem, and spinal cord (Tandan & Bradley 1985). Patients with ALS usually die within 5 years of disease onset due to respiratory failure (Majoor-Krakauer et al. 2003). While ALS presents mostly as a sporadic disease of unknown cause, familial cases of ALS (fALS) account for 10-15% of total cases. In most instances, sporadic and familial forms of the disease are clinically similar. The causes of sporadic cases of ALS (sALS) remain a mystery, but causative factors have been identified in fALS, including SOD1, TDP-43, and FUS/TLS mutations, as well as the most recently identified C9orf72 hexanucleotide repeat expansions (DeJesus-Hernandez, Ian R. Mackenzie, et al. 2011).

1.2 Genetic Factors Implicated in ALS and Related Neurodegenerative Diseases

1.2.1 Role of SOD1 in ALS

In 1993, a landmark study initiated the molecular era of ALS research with the identification of mutations in the gene encoding Cu, Zn superoxide dismutase (SOD1) as a causative factor for 20% of cases of fALS (Rosen et al. 1993). SOD1 is a homodimeric enzyme that converts the superoxide anion, a harmful byproduct of oxidative metabolism, into hydrogen peroxide and molecular oxygen. Each SOD1 monomer binds two metal ions, one copper atom and one zinc
atom. The zinc serves a structural role, while the copper cycles between the Cu(I) and Cu(II) states to convert one superoxide molecule into either hydrogen peroxide or molecular oxygen with each oxidation-state transition (Rakhit & Chakrabartty 2006). SOD1 contains a beta barrel fold in its native folded state (Deng et al. 1993). Over 100 different ALS-associated SOD1 mutations have been identified, and mutations are located across the entirety of the SOD1 sequence rather than just in the active site (Cleveland & Rothstein 2001; Majoor-Krakauer et al. 2003). In mouse models expressing ALS-related SOD1 mutants, the disease phenotype is recapitulated without observation of reduced SOD1 enzymatic function (Gurney et al. 1994). These findings suggest that the pathology of SOD1 mutation-induced ALS is triggered by a toxic gain of function of mutant SOD1.

Despite the early discovery of the SOD1 gene as a contributing factor to ALS, the mechanism by which SOD1 mutants generate toxicity remains a subject of debate and several mechanisms have been proposed. Generally, two mechanisms exist for the toxicity of SOD1 (Rakhit & Chakrabartty 2006). The first mechanism involves conformational changes that occur due to these mutations in SOD1, which allow new substrates to access SOD1’s copper atom at its active site to allow for new cytosolic functions of the protein. This may lead to aberrant protein-protein interactions which lead to a cascade of events that contribute to ALS pathology (Ilieva et al. 2009). This mechanism may explain the excessive synaptic firing of motor neurons due to the inability to remove glutamate at the synapse observed in SOD1 mutant mouse models and in familial and sporadic ALS patients (Rothstein et al. 1995). This over-stimulation by glutamate consequently elicits toxic events in the postsynaptic motor neuron, but direct interactions involving mutant SOD1 are not yet clearly defined. Another example of SOD1 mutants gaining aberrant protein-protein interactions is the observation of multiple SOD1 mutants, but not wild type protein, binding to the transmembrane protein Derlin-1, which is required for the
disassociation of misfolded protein from the ER to the cytosol in the ER-associated degradation (ERAD) pathway (Nishitoh et al. 2008). Binding of mutant SOD1 to Derlin-1 inhibits the ERAD pathway and generates ER stress, which may lead to ALS pathology. Mutant SOD1 also associates with components of neurosecretory vesicles chromogranin A (CgA) and chromogranin B (CgB), leading to extracellular secretion of mutant SOD1 which causes immune responses within the spinal cord and the induction of neuron death (Urushitani et al. 2006). Mutant SOD1 can also cause the generation of extracellular superoxide by associating with Rac1, a GTPase that affects the secretion of extracellular superoxide by phagocytic cells as a mechanism of pathogen defense. By interacting with higher affinity than wild type SOD1, presumably by conformational changes, SOD1 mutants may promote the excessive secretion of superoxide, leading to cellular damage (Harraz et al. 2008).

The second proposed mechanism through which mutant SOD1 causes ALS is that the mutations decrease the overall stability of SOD1, which increases the propensity of SOD1 to misfold and aggregate. SOD1 aggregates can accumulate in the ER lumen and bind BiP, a protein required for ER stress transducer regulation, causing ER stress (Kikuchi et al. 2006). In fact, any aberrant interaction with ER or ER-associated proteins may interfere with protein production and lead to cell toxicity. SOD1 misfolding and aggregation can overwhelm the endogenous proteosome activity in the cell, leading to sequestration of endogenous proteins into the SOD1-positive aggregates, furthering disease progression (Kabashi et al. 2004). Misfolded SOD1 can also deposit into spinal cord mitochondria causing structural damage (Vande Velde et al. 2008), and this is observed presymptomatically, suggesting a role in the initiation of pathology (Ilieva et al. 2009). It is still unclear whether the association of mutant SOD1 aggregates to mitochondria affect their endogenous functions.
Pathological pathways involving both mechanisms have also been proposed. For instance, SOD1 mutants can interfere with axonal cytoskeletal organization and/or inhibit axonal transport through aberrant interactions with proteins involved in such processes, or disrupt such processes by forming aggregates (Ilieva et al. 2009). Microhemorrhages can also form due to SOD1 mutants interacting with tight junction proteins or forming aggregates to cause damage to the blood-spinal barrier, leading to leakage of toxic products such as iron complexes from hemoglobin in the spinal cord, leading to disease (Zhong et al. 2008).

These proposed pathways collectively allow cells exhibiting toxicity to cause surrounding cells to die. Although motor neurons are the primary target of mutant SOD1 toxicity, neighbouring cells such as astrocytes and microglia are also affected by mutant damage, which leads to further progression of disease (Ilieva et al. 2009).

Although diverse in nature, these mechanisms are proposed to explain at the cellular level how mutant SOD1 generates toxicity. However, at a biophysical level, a key step of toxicity is the misfolding of SOD1 and the generation of misfolded protein species that cause cellular damage, either through conformational changes to give new functions such as aberrant protein-protein interactions or through misfolding and aggregation (Mulligan et al. 2008). In fact, SOD1 has an occupational hazard of oxidative damage from scavenging free radicals as its substrates, and oxidative damage to wild type SOD1 can trigger misfolding and aggregation (Rakhit & Chakrabarty 2006). In either case, the conformation flexibility of SOD1 is increased. Studies have shown, in a six-state unfolding model of SOD1, that wild-type SOD1 unfolds in a major pathway beginning with the concomitant loss of zinc and the dissociation of the dimer, followed by the conformational change of the beta barrel secondary structure, and ends with the loss of the copper atom. A minority copper-loss pathway involves firstly the unfolding of beta barrel
structure, followed by copper loss, then dimer dissociation and loss of zinc. It has been established that zinc-deficient monomeric SOD1 is a highly populated unfolding intermediate in the wild-type SOD1 unfolding pathway (Mulligan et al. 2008). Studies in the unfolding pathways of some fALS mutants show a modified 9-state model where in some mutants, zinc loss and dimer dissociation are observed as separate processes (Ip et al. 2011). Overall the SOD1 mutants investigated show a propensity to favour the copper loss pathway. These transiently populated, partially metalated unfolding intermediates generated by ALS-causing SOD1 mutations may play a role in the toxicity of mutant SOD1. If true for all SOD1 mutants, the Cu-deficient misfolded SOD1 species can be potential therapeutic targets. By blocking the conversion of these intermediates and thereby drive the folding pathway towards that of the native SOD1, toxic effects may be blocked in order to reduce mutant toxicity.

SOD1 mutations have been studied extensively and many mechanisms have been proposed for its involvement in disease. However, SOD1 mutations account for only 20% of fALS cases. Since the discovery of SOD1 mutations as a contributor to ALS, other genetic factors have now been identified.

1.2.2 TDP-43 in ALS and Related Neurological Diseases

In 2006, landmark studies discovered that inclusion bodies within neurons of patients with sporadic ALS or frontotemporal lobar degeneration with ubiquitin positive inclusions (FTLD-U) contain the TAR DNA binding protein-43 (TDP-43) (Neumann et al. 2006; Arai et al. 2006). Subsequently, mutations in the TDP-43 gene (TARDBP) shared by familial and sporadic forms of ALS were identified (Kabashi et al. 2008). FTLD is another neurodegenerative disease that fatally affects the central nervous system (CNS). While muscle paralysis in ALS is caused by degeneration of upper motor neurons in the cerebral cortex and lower motor neurons of the
brainstem and anterior horn of the spinal cord (Tandan & Bradley 1985), the site of degeneration in FTLD is located at frontal and temporal lobes of the cerebral cortex, presented as cognitive deficits and characterized by behavioral and language disorders (Neary et al. 1998). Symptoms of patients with FTLD have some pathological overlap with those suffering from ALS, and the difference in pathology is essentially caused by a difference in the type of affected neuron, suggesting a common disease mechanism. A total of 38 mutations have been found in ALS patients and correspond to 4% of cases of fALS and <1% of cases of sALS (Lagier-Tourenne et al. 2010). The common presence of TDP-43 inclusions in both ALS and FTLD cases suggests both diseases may be considered a part of a broad spectrum of diseases categorized as TDP-43 proteinopathies (Neumann et al. 2006).

TDP-43 is a 414 amino acid nuclear protein composed of two RNA recognition motifs (RRM1 and RRM2) and a C-terminal glycine-rich region. The structure of full length TDP-43 is not yet known, but crystallographic studies have provided data on the RRM2 domain of the protein (Kuo et al. 2009). The RRM2 domain consists of five antiparallel beta strand structure with two short alpha helices. This domain mediates the dimerization of TDP-43 \textit{in vivo} (Kuo et al. 2009). TDP-43 is a RNA and DNA binding protein that appears to mediate many different levels of RNA processing \textit{in vivo} (Lagier-Tourenne et al. 2010). RRM1 is essential for the binding of RNA/DNA, and preferentially binds poly-UG RNA and poly-TG repeats of single stranded DNA, while RRM2 mediates the dimerization of the protein. The C-terminal region is of particular interest in terms of the involvement of the protein in disease, since almost all ALS-associated mutations discovered in \textit{TARDBP} gene are located on the C-terminal glycine-rich region.
Normally, TDP-43 is localized to the nucleus and carries out a variety of RNA processing functions such as transcription regulation, as it has been found to interact with euchromatin (transcriptionally active genes) through RRM2. TDP-43 also associates with splicing factors and is implicated in splicing regulation. It is involved in micro-RNA processing in nucleocytoplasmic mRNA shuttling. In ALS and FTLD-U patients, TDP-43 is mislocalized to the cytoplasm due to nuclear clearance of the protein. TDP-43 subsequently aggregates and forms cytosolic inclusion bodies, where it is hyperphosphorylated, ubiquitinated or cleaved into 25 kDa and 35 kDa C-terminal fragments (Neumann et al. 2006; Arai et al. 2006). It has been shown that the expression of these fragments in cell culture can recapitulate the pathological features of TDP-43 proteinopathies (Igaz et al. 2009) and that exogenous expression of the 25 kDa fragment cause sequestration of wild type TDP-43 from the nucleus and can induce cell death through toxic gain of function (Zhang et al. 2009), suggesting that C-terminal fragments play a major role in the aggregation of the protein and is an important step in the pathogenesis of ALS and FTLD-U. Regions on the C-terminus also contain yeast prion-like motifs that are rich in Gln and Asn amino acids (Gitler & Shorter 2011). When 12 tandem repeats of one of the prion-like motifs are expressed in cell culture, the disease phenotypes of ubiquitinated and hyperphosphorylated inclusions are recapitulated (Budini et al. 2012). Misfolded TDP-43 aggregates also show prion-like behavior in that inclusions from diseased patient tissue can be used to induce the formation of phosphorylated and ubiquitinated inclusions in cell culture expressing native TDP-43 (Nonaka et al. 2013). It has also been shown that the C-terminus engages in self-interactions when TDP-43 is in its native dimer form and is prevented from aberrant interactions (Wang et al. 2012). Together, these findings suggest that a possible mechanism for TDP-43 misfolding involves the misfolding of the C-terminal domain into an aberrant structure that can act as a template for sequestration of other TDP-43 molecules into this misfolded form. The initiation of the process
is not known, but likely involves deviation of TDP-43 from its native structure, whether through caspase cleavage, nuclear clearance, or monomerization, causing the C-terminal domain to become prone to misfolding.

It is still currently under debate whether mutations in TDP-43 cause disease through a gain of toxic function or loss of function of TDP-43, or whether TDP-43 aggregates observed in neurons are the cause or the result of disease progression. It still remains unclear whether the phosphorylation and fragmentation of TDP-43 into 20-25 kDa C-terminal fragments is a primary event or epiphenomenon following earlier pathological events such as misfolding. Cytoplasmic localization of TDP-43 was proposed as an early event in TDP-43 proteinopathies (Giordana et al. 2010) and expression of mutant TDP-43 was found to be toxic to rat cortical neurons, associating with increased cytoplasmic distribution. The amount of cytoplasmic TDP-43 but not presence or number of inclusions associated with neuronal death (Barmada et al. 2010). Other studies have also confirmed the association of cytoplasmic localization of TDP-43 and the formation of inclusions, but these two events still suffer from the chicken and egg ambiguity (Lagier-Tourenne et al. 2010).

TDP-43 is also involved in stress granule (SG) assembly and recruitment, which are cytoplasmic, microscopically visible foci consisting of mRNA and RNA-binding proteins that stall translation during cellular stress. However, TDP-43 does not appear essential for SG formation and needs to be present in the cytoplasm in order for SG recruitment. Once formed, TDP-43 can subsequently form cytoplasmic aggregates (Parker et al. 2012) or the TDP-43 positive SGs can disassemble after cellular stress has passed. Whether TDP-43 or its disease-associated mutants can nucleate SGs has been studied intensely but no clear answers have emerged (Wolozin 2012).
Due to the diverse role of TDP-43 in the cell, its mislocalization, misfolding, and aggregation are of considerable interest in further investigating its roles in TDP-43 proteinopathies.

### 1.2.3 FUS/TLS in FTLD/ALS Related Diseases

Shortly after the discovery of TDP-43, another RNA/DNA-binding protein, fused in sarcoma/translocated in liposarcoma (FUS/TLS or FUS) was identified as a primary cause of fALS (Kwiatkowski et al. 2009; Vance et al. 2009) and FUS/TLS was found in inclusion bodies of non-SOD1 ALS and FTLD patients. 30 mutations have been identified in 4% of fALS patients. It is currently unknown whether FUS/TLS is involved in the pathogenesis of sALS and other forms of fALS without FUS/TLS mutations (Deng et al. 2010). In some cases, patients with FUS/TLS mutations show development of ALS phenotype concurrently with FTLD, while others showed FTLD in the absence of ALS, further suggesting that ALS and FTLD are part of a broad spectrum of diseases that can be categorized as TDP-43 and FUS/TLS proteinopathies.

FUS/TLS is a 526 amino acid protein consisting of an N-terminal domain enriched in glutamine, glycine, serine and tyrosine (QGSY region), a glycine-rich region, a RNA recognition motif, a zinc finger motif flanked by two arginine- and glycine-rich regions (RGGs) and an unstructured C-terminal region where the nuclear localization signal (NLS) resides. Like TDP-43, FUS/TLS is a nuclear protein in normal cells, but in patients with FUS/TLS mutations, cytoplasmic inclusions are found in neurons and glial cells (Kwiatkowski et al. 2009; Vance et al. 2009). In most FTLD-U patients, FUS/TLS inclusions do not include TDP-43, suggesting that the two proteins likely aggregate independently of each other, while in some non-SOD1 fALS, FUS/TLS aggregates seem to co-localize with TDP-43 and other proteins such as ubiquitin (Deng et al. 2010). ALS-associated mutations are largely found in the C-terminal NLS region responsible for the binding of transportin which allows nuclear-cytoplasmic shuttling of FUS/TLS (Dormann et
al. 2010). Disruption of binding through mutations or other unidentified means may promote nuclear depletion and cytoplasmic mislocalization and accumulation of the protein, a phenotype observed in brain tissue of ALS and FTLD patients (Kwiatkowski et al. 2009; Vance et al. 2009). A recent model suggests a two-hit mechanism for the involvement of FUS/TLS in disease pathogenesis (Dormann & Haass 2011). The disruption of transportin binding and cytoplasmic accumulation provides the first hit but is not sufficient to cause disease. The second hit involves induced cellular stress such as heat shock, oxidation or viral infections, causing the recruitment of FUS/TLS into stress granules. This association of mutant FUS/TLS to stress granules may cause impairment to the stress response in cells, leading to disease.

Very similar to TDP-43, FUS/TLS is a transcriptional regulator, involved in RNA splicing, miRNA processing, and RNA subcellular localization, translation, and decay. Both proteins are found in aggregates and associated with stress granules under conditions of cellular stress. The similarities of these proteins in term of function and the disease phenotype caused by their mutations suggests that the pathology associated with these proteinopathies likely involve the disruption of the many stages of mRNA processing. In fact, TDP-43 inclusions have been found in various forms of dementia, including 30% of Alzheimer’s patients and various forms of Parkinson’s disease (PD), where TDP-43 has been reported to partially co-localizes with tau or alpha-synuclein (proteins that form fibrillar aggregates in various neurodegenerative diseases) (Uryu et al. 2008).

The discovery of FUS/TLS as a component of inclusion bodies of FTLD is fairly recent and the presence of FUS/TLS in other neurodegenerative diseases has not yet been tested. However, given the similar roles of FUS/TLS and TDP-43, it is likely that both proteins are contributors to a broad range of neurodegenerative disease that share common phenotypes such as FTLD and
ALS. The mechanisms by which these proteins generate disease (whether through a gain of toxic function or loss of native function) is not well understood. Current studies are striving to determine the cellular functions of these proteins and interactions and commonalities between TDP-43, FUS/TLS and SOD1 have been identified.

1.2.4 C9orf72 Repeat Expansions in FTLD/ALS Disease Spectrum

The most recent factor contributing to FTLD/ALS was reported initially in 2006, where linkage of FTLD/ALS with human chromosome 9p21.3-13.3 was discovered (Morita et al. 2006; Vance et al. 2006). Recently, in 2011, the cause of this linkage was identified as a pathological expansion of a non-coding region on chromosome 9’s open reading frame 72 (C9orf72), which produces hexanucleotide repeat expansions consisting of GGGGCC (Renton et al. 2011). Abnormal expansion of c9orf72 is currently the most common genetic cause of diseases of the FTLD/ALS spectrum (Reviewed by Cruts et al. 2013). Individuals carrying mutations in this region can develop either FTLD, ALS, or both diseases with a wide range of age of onset (Majounie et al. 2012; Sabatelli et al. 2012). Age of onset has also been reported to inversely correlate with generation number, suggesting genetic anticipation (Arighi et al. 2012; Boeve et al. 2012; Stewart et al. 2012; Van Langenhove et al. 2013). In normal populations, “non-repeat expansion” (<30 repeats) size ranges from 2-24 GGGGCC repeats. While pathological expansion sizes are not well defined, they generally exceed 60 repeats, and cases of 750 and 4400 repeats have been reported (DeJesus-Hernandez, Ian R Mackenzie, et al. 2011; Buchman et al. 2013; Beck et al. 2013). The cause of these expansions, like other repeat expansion diseases, is likely strand slippage during DNA replication.

A few mechanisms of disease pathogenesis have emerged. The C9orf72 gene is transcribed into 3 different mRNA variants, but the GGGGCC repeat mutations are likely located to the
regulatory region of all three variants (Cruts et al. 2013). Thus disruption of the regulatory region achieved through distortion of the regulatory factor binding, altered DNA methylation or distance effects, may cause a loss of expression of the native gene encoded in this region, leading to disease. Another mechanism suggests that RNA splicing of expanded repeats may cause sequestration of RNA binding proteins or other factors into nuclear foci, and impairing their normal functions (Echeverria & Cooper 2012). This mechanism is most intriguing as previously identified proteins associated with FTLD/ALS such as FUS/TLS and TDP-43 are RNA-binding proteins sequestered into inclusion with compromised functions. A third mechanism is that the expanded repeats are abnormally transcribed followed by repeat-associated non-AUG-initiated (RAN) translation into aggregation-prone dipeptide sequences that cause aggregation and toxicity (Mori et al. 2013).

These proposed mechanisms may contribute to disease pathology to different extents in a non-mutually exclusive manner. For instance, GGGGCC has also been shown to form stable RNA G-quadruplexes, which may play a role in all three mechanisms listed above (Fratta et al. 2012). The discovery of this hexanucleotide repeat contributes to a significant portion of familial FTLD/ALS disease, but the mechanism of disease is still under investigation. A large percentage of genetic factors causing FT/ALS currently remain unidentified.

1.3 Experimental Rationale

Neurodegenerative diseases have a wide range of clinical symptoms; ranging from cognitive defects seen in AD and FTLD to motor neuron dysfunction present in patients with ALS. However a common hallmark in a large majority of these diseases is the aggregation of misfolded proteins inside the affected neurons, exemplified by the amyloid peptide Aβ in AD (Schenk et al. 1999), α-synuclein in Parkinson’s disease (PD) (Polymeropoulos 1997), and SOD1
in familial ALS (Majoor-Krakauer et al. 2003). These protein aggregates may play significant roles in these diseases, as they have been shown to have intrinsic toxicity in cells (Bucciantini et al. 2002). Some propose that their presence can also overwhelm the chaperone system by blocking proteasomes (Bence et al. 2001). In fALS, it is currently unresolved whether the presence of aggregates is the cause or byproduct of the disease. In certain cases, the soluble, misfolded, oligomeric form of the protein may be more toxic than the large aggregates end-products (Lambert et al. 1998). It is also unknown whether the formed aggregates serve a neuroprotective purpose (Caughey & Lansbury 2003). Recent studies have identified TDP-43 as a major component of inclusion bodies within neurons of patients with familial ALS (Neumann et al. 2006; Arai et al. 2006), and that mutations in the TDP-43 gene TARDBP are associated with the prevalence of familial ALS (Kabashi et al. 2008). In fact, TDP-43 is implicated in a wide range of neurodegenerative diseases now coined as TDP-43 proteinopathies (Lagier-Tourenne et al. 2010). In this study, we propose that stabilization of the native state of the protein should be beneficial, and that preventing the protein from entering a misfolded state would ameliorate toxicity.

The rationale of our hypothesis is adapted from work done on the protein Transthyretin (TTR), a natively homo-tetrameric protein involved in thyroid hormone (T4) transport. The aggregation of this protein leads to fibrillar deposits in peripheral nerves and non-CNS tissues including the heart, liver, and kidneys, contributing to a group of disease known as transthyretin amyloidosis (Rochet & Lansbury 2000). Studies by Kelly and co-workers have shown that under destabilizing conditions in vitro, tetrameric TTR dissociates, and undergoes a conformational change that potentially exposes certain hydrophobic regions of the protein, forming an aggregation-prone monomer, leading to fibrillar aggregates (Jiang et al. 2001). Giving that monomerization was the first step in the aggregation pathway, Kelly and co-workers reasoned
that the stabilization of native, tetrameric TTR should reduce the propensity of TTR to enter the aggregation-prone states. A screening of numerous small molecules that bind to the T4 binding site of tetrameric TTR generated candidate compounds that achieved native-state stabilization (Connelly et al. 2010). These compounds successfully reduced aggregation \textit{in vitro}, and one particular compound, tafamidis, has been approved for treatment (Nencetti et al. 2013).

We propose a similar strategy to stabilize the native, dimeric state of TDP-43 using native-state binding ligands, namely RNA and DNA sequences (Figure 1). The sequences we selected consist of both natural targets of TDP-43 and artificially constructed \textit{de novo} sequences based on consensus binding sites. Using this hypothesis-based approach, we aim to identify small molecules capable of inhibiting TDP-43 aggregation.
Figure 1: Proposed hypothesis of aggregation inhibition through the binding of a DNA target. We propose that the presence of a binding target (blue circle) to TDP-43 will stabilize its native state (green boxes) and reduce its propensity to enter the aggregation-prone pathway (red ovals).
Chapter 2

2 Materials and Methods

2.1 Protein Expression and Purification

The pET-30 vector modified with kanamycin resistance gene containing the recombinant vYFP-TDP43 protein sequence was transformed into BL21-AI competent cells (Invitrogen), and plated onto LB plates containing 35 µg/ml Kanamycin (BioBasic) and incubated at 37 °C for 12 h. Colonies were inoculated into 2 mL of LB medium, incubated at 37 °C for 5 h and diluted into 300 mL of LB medium. When the culture reached $A_{600}$ of 0.4 the temperature was reduced to 19 °C and induced using 1 mM IPTG and 0.2% Arabinose. Cells were harvested 6 h after induction by centrifugation at 3000 rpm (Sorvall SLA3000 rotor) for 20 min at 4 °C. The cell pellet was resuspended in 50 mL Lysis Buffer (Johnson et al. 2009) (40 mM HEPES-KCl, 500 mM KCl, 20 mM imidazole, 20 mM MgCl$_2$, 2 mM βME, 10% glycerol, cOmplete EDTA-free protease inhibitor (Roche), pH 7.4). Cells were lysed by sonication in 10 ml fractions at 4 °C for 5 min (5 s pulse, 5 s stop) using a Vibracell sonicator (Sonics). The lysate was centrifuged for 30 min at 4 °C at 15,000 rpm (Sorvall SS-34 rotor) to isolate soluble protein fraction (supernatant).

Purification of the His-tagged recombinant protein was achieved by Ni-NTA affinity chromatography using Ni-NTA beads (Sigma) at 4 °C. 10 mL of supernatant was applied to 0.5 mL bead volume of Ni-NTA beads. The beads were washed twice with 1 mL Lysis Buffer and eluted using modified Lysis Buffer containing 250 mM imidazole and no protease inhibitors to obtain purified protein.
2.2 Dynamic Light Scattering Measurements

Dynamic light scattering measurements of hydrodynamic radius ($R_h$) were made at 20°C with a DynaPro DLS instrument (Protein Solutions) using a 12 µL quartz cuvette. Samples were measured using 10 s averaging time. Ten or more consecutive measurements were used for regularization analysis using DYNAMICS software. Particle translational diffusion coefficients were calculated from decay curves of autocorrelation of light scattering data and converted to hydrodynamic radius ($R_h$) with the Stokes-Einstein equation. Histograms of mass versus $R_h$ were calculated using Dynamics data analysis software.

2.3 Urea Denaturation Experiment

Purified protein samples (2 µM) were unfolded in 0 M – 7.2 M urea and incubated for 30 minutes at room temperature. Trp fluorescence of the samples was measured in 1 cm cuvette using a Photon Technology International QM-1 fluorescence spectrophotometer using an excitation wavelength of 283 nm and an emission wavelength range of 315 nm to 335 nm using 2 nm bandpass. Integrated fluorescence was normalized and converted to % folded. The melting curve was fitted using non-linear least square fitting to the equation $y = \frac{y_{max} - y_{min}}{1 + (\frac{x}{EC50})^n} + y_{min}$ (Equation 1) using OriginPro 8.5 software.

2.4 Circular Dichroism Spectroscopy

Purified protein was dialyzed against buffer containing 20 mM MgCl$_2$, 500 mM KCl, 40 mM HEPES, pH 7.4 to remove imidazole for CD spectroscopy. Far UV CD spectra were acquired with an Aviv CD spectrometer model 62DS at 25 °C. Spectra were obtained from 209 nm to 320 nm (1 mm pathlength, 1 nm step sizes, 1 nm bandwidth and 16 s averaging time).
2.5  *In vitro* Aggregation of vYFP-TDP-43

Aggregation was induced by a 10-fold dilution of stock purified protein (20 µM) in elution buffer (Johnson et al. 2009) (40 mM HEPES-KCl, 500 mM KCl, 250 mM imidazole, 20 mM MgCl₂, 2 mM βME, 10% glycerol, pH 7.4) into final concentrations of 2 µM protein in 170 mM KCl, 36 mM HEPES, 25 mM Imidazole, 18 mM MgCl₂, 1% glycerol, 1.8 mM β ME, pH 7.4. The sample was mixed and incubated at room temperature. For right angle light scattering measurements, samples were incubated for 4 hr. For fluorescent microscopy and atomic force microscopy experiments, samples were incubated for 20 minutes.

2.6  Right Angle Light Scattering

Right angle light scattering of aggregated samples were measured in 1 cm quartz cuvette using a Photon Technology International QM-1 fluorescence spectrophotometer using an excitation wavelength of 400 nm and emission wavelength of 400 nm with 2 nm bandpass at room temperature. 200 µL sample volumes were mixed before measurements. 20 measurements were taken for each ssDNA concentration and averaged. Inhibition curves were fitted by non-linear least square fitting to the equation

\[
\frac{y}{y_{max} - y_{min}} = \frac{y_{max} - y_{min}}{1 + \left(\frac{E}{E_{50}}\right)^n} + y_{min}
\]

using OriginPro 8.5 software.

2.7  Fluorescence Microscopy

9.25 µL protein samples induced to aggregate in the presence or absence of oligonucleotides were inserted into a compartment constructed using Secure-Seal Imaging Spacers (Grace) and two no. 2 18 mm circular micro cover glass (VWR) for imaging. Fluorescence images were acquired on an Olympus IX70 (Olympus, Inc.) inverted microscope with a 40× TIRF objective (N.A. 1.45) (Olympus, Inc.), illuminated by a Xenon light (125 W; λₐₓ ~ 515 nm). Excitation light was reflected by a 485-555-650TBDR dichroic, and emitted light was passed through a
515-600-730TBEM filter (Omega Optical, Inc.). Images were digitized with a cooled Evolution QEi CCD camera (Media Cybernetics, Inc.).

2.8 Atomic Force Microscopy

All atomic force microscopy (AFM) images were acquired using tapping mode with a Digital Instruments Nanoscope IIIA Multimode scanning probe microscope (Digital Instruments, Santa Barbara, CA). The AFM images were collected using Nanoscope software (version 5.12) and the J scanning head that has a maximum lateral scanning area of 116 μm × 116 μm. V-shaped silicon nitride probes (SNL-10 cantilever D, Bruker AFM Probes, Camarillo, CA) were irradiated under UV for 30 min to remove organic contaminants. 10 μL of protein samples induced to aggregate in the presence or absence of oligonucleotides was transferred onto freshly cleaved mica and sealed in the liquid cell. All imaging were captured as 512 × 512 scans at tip scan rate between 0.7 and 1.2 Hz with cantilever drive frequency of ~ 8.5 kHz.
Chapter 3

3 Results

3.1 Sequence Selection of Inhibitors of TDP-43 Aggregation

In our hypothesis-driven approach to find compounds capable of inhibiting TDP-43 aggregation, we examined a number of potential TDP-43 binding targets, consisting of both artificial and naturally occurring sequences. The first target we selected is the transactive response (TAR) element of Human Immunodeficiency Virus-1 (HIV-1), the first binding target of TDP-43 through which the protein was discovered (Ou et al. 1995). HIV-1 is a retrovirus that infects the host immune cells and causes acquired immunodeficiency syndrome (AIDS) through the killing and depletion of cluster of differentiation 4 (CD4) T helper cells. The virus enters the cell by membrane fusion through various glycoproteins on the host cell surface and converts its single stranded RNA into double stranded cDNA using reverse transcriptase and host transcription factors. The viral cDNA is then incorporated into the host genome by the viral protein integrase, where the virus may lie dormant until expression of the viral gene occurs for viral propagation. Many factors, viral and host cellular, modulate the expression of this cDNA. A region in the integrated viral cDNA called the long-terminal repeat (LTR) contains many elements critical for the control of viral gene expression by serving as binding sites for cellular transcription factors. These elements include NF-κB, Sp11, TATA, and TAR, and are directly or indirectly responsible for levels of the viral protein trans-activator of transcription (tat) (Ou et al. 1995). Tat is a small protein of 86 to 101 residues depending on viral isolate and drastically enhances viral transcription and replication upon its activation (Debaisieux et al. 2012). TAR is essential for tat activation; it is transcribed into TAR RNA that contains a stem-loop structure that binds and activates tat. TDP-43 was found to bind to the double stranded TAR cDNA element, hence its
The binding sequence contains 46 nucleotides and two pyrimidine-regions, which are consensus sequences for TDP-43 binding (Xiao et al. 2011). TDP-43 not only binds to the double stranded TAR cDNA, but also has minor affinity for the coding single strand sequence, and no affinity for the transcribed TAR RNA (Ou et al. 1995). We decided to include both the double stranded and single stranded HIV1LTR in our TDP-43 aggregation assays.

Another consensus binding sequence for TDP-43 is a poly thymine-guanine repeat denoted by TG\textsubscript{n} (Buratti & Baralle 2001; Xiao et al. 2011). Single stranded DNA and RNA containing more than 3 repeats of TG are capable of binding TDP-43 with increased affinities for higher repeat numbers (Buratti & Baralle 2001). Given the extensive data on the binding of poly-TG to TDP-43, we include poly-TG single strand DNA as candidates for TDP-43 aggregation inhibiting compounds, denoted by TG\textsubscript{3}, TG\textsubscript{6}, TG\textsubscript{9} and TG\textsubscript{12} (Figure 2).

TDP-43 has many roles in mRNA processing such as splicing and targeting mRNA for degradation. There has been evidence for TDP-43 autoregulation of its own mRNA transcript levels through a negative feedback loop. TDP-43’s RRM and C-terminal domains are required for the binding interactions with the 3’ UTR of TDP-43 mRNA and promoting mRNA instability (Polymenidou & Lagier-Tourenne 2011; Ayala et al. 2011). The disruption of the ability to autoregulate TDP-43 levels may have many implications in disease pathogenesis (Budini & Buratti 2011). For instance, reduction of nuclear TDP-43 levels through aggregation or mislocalization to the cytosol may increase TDP-43 mRNA levels, and promote TDP-43 production, which may lead to further aggregation. The strongest binding sequence in TDP-43 3’UTR was found to be a stretch of guanine-rich sequence of 34 nucleotides by UV crosslinking immunoprecipitation (UV-CLIP) assay, denoted by CLIP34nt (Bhardwaj et al. 2013) (Figure 2).
We chose this sequence as a candidate RNA sequence for inhibiting TDP-43 aggregation, acknowledging its role in regulation of TDP-43 levels and its implications in disease.

HIV1LTR ssDNA:
5’ CTGCTTTTTGCTTCGTAGCTGGTCTCTCCTGTTAGACCAGATCTGAG 3’

HIV1LTR dsDNA:
5’ CTGCTTTTTGCTTCGTAGCTGGTCTCTCCTGTTAGACCAGATCTGAG 3’
3’ GACGAAAAACGGACATGACCCAGAGAGAGACCAATCTGGTCTAGACTC 5’

TG₃:
5’ TGTGTG 3’

TG₆:
5’ TGTGTGTGTGTG 3’

TG₉:
5’ TGTGTGTGTGTGTGTGTGTG 3’

TG₁₂:
5’ TGTGTGTGTGTGTGTGTGTG 3’

CLIP34nt RNA:
5’ GAGAGAGCGCGUGUCACAGACUUGGUGGGAGCAAA 3’

Figure 2: Names and sequences of nucleotides used for native-state stabilization. Poly-pyrimidine regions are underlined for HIV1LTR ssDNA and dsDNA.
3.2 Recombinant Expression of TDP-43

TDP-43 is an intrinsically aggregation-prone protein (Johnson et al. 2009). In order to facilitate its purification, a construct consisting of a His₆-tag followed by the Venus yellow fluorescent protein (vYFP), a tobacco etch virus (TEV) protease cleavable linker and the human TDP-43 sequence was utilized. The vYFP is derived from the well-known green fluorescent protein with mutations to increase its folding rate and the brightness of the fluorophore (Nagai et al. 2002). It serves as a visual indicator of the protein and increases the overall folding and solubility of the construct. The construct was expressed in E. coli cells (see Methods) using buffers adapted from previous studies (Johnson et al. 2009). TEV cleavage of the recombinant protein (vYFP-TDP-43) resulted in immediate precipitation of full length TDP-43, while C-terminally degraded TDP-43 remained in solution, suggesting that the C-terminus is required for aggregation to occur. Due to difficulties in the cleavage of the protein from vYFP, we carried out the aggregation assays using the uncleaved, recombinant vYFP-TDP-43.

The quality of the expressed protein was assessed using a variety of assays. Sodium dodecyl sulphate polyacrylamide gel electrophoresis of samples at various stages of purification is shown in Figure 3. The single band at 76 kDa in the eluate lane corresponds to the expected size of the recombinant vYFP-TDP-43. Gel analysis of total cell lysate during the expression of the protein also determined the 76 kDa band as the enriched species after induction of protein production in E. coli (Figure 4). Circular dichroism spectroscopy on the eluted protein produced a spectrum indicative of β-structure with a characteristic β peak at 218 nm (Figure 5). This was expected, as vYFP contains a beta barrel structure and contributes to this CD signal. The
Figure 3: SDS-PAGE of purified vYFP-TDP-43. Lanes containing crude cell lysate after sonication (lysate), soluble fraction of cell lysate (sup), fraction collected after washing with Lysis buffer (wash) and final eluted fractions (eluate) are shown. Approximately 10 μg of samples were applied to a 12% polyacrylamide gel for electrophoresis. The band corresponding to the size of vYFP-TDP-43 is indicated by the arrow.
Figure 4: Monitoring vYFP-TDP-43 expression in *E. coli*. A) vYFP-TDP-43 production of *E. coli* was monitored using fluorescence of vYFP (510 nm – 520 nm) excited at 487 nm after induction. Induced cells are shown in black, while the non-induced control is shown in white. B) SDS-PAGE of crude cell lysates of induced or non-induced samples at 400 minutes after induction time. Enriched band is indicated by the arrow.
Figure 5: Circular dichroism spectrum of vYFP-TDP-43. Circular dichroism was measured at 25 °C with 16 s averaging times. Signal for 8.8 μM vYFP-TDP-43 is shown with characteristic β signal.
absence of significant α-helix and random coil signals also indicate that TDP-43 is mostly β-structured, which agrees with recent crystallographic data on its RRM2 domain (Kuo et al. 2009). The sample was also subjected to dynamic light scattering (DLS) assay to assess particle size under purification conditions. DLS results indicate a single species of hydrodynamic radius of 4.86 nm, which constitutes 95% of the sample by mass (Figure 6). Conversion of radius to molecular weight yields a weight of 136 kDa, consistent with dimer configuration, which was also observed in recent studies (Kuo et al. 2009). Urea denaturation of the sample monitored by tryptophan fluorescence shows co-operative unfolding of the protein with a denaturation midpoint of 4.2 M urea (Figure 7). It is noteworthy that the green fluorescent protein family is resistant to urea denaturation under our experimental conditions based on literature (Alkaabi et al. 2005) and visual observation of the bright yellow fluorescence that is retained even at 7.2 M urea. This suggests that TDP-43 segment of the recombinant protein is also folded. Collectively, these assays indicate that the protein produced was 95% pure, dimeric, and folded, and that the vYFP does not interfere with these intrinsic properties of TDP-43.
**Figure 6: Size distribution by mass of purified vYFP-TDP-43.** 20 μM protein are eluted from column purification under non-aggregating conditions specified in purification methods. Measurements of $R_h$ and MW were calculated using appropriate software (Methods).
Figure 7: Urea denaturation of vYFP-TDP-43. Integrated Trp fluorescence from 315 nm to 335 nm using excitation wavelength of 283 nm was measured for samples containing 2.1 µM vYFP-TDP43 incubated with 0.0 – 7.2 M urea for 12 hours. Trp fluorescence was normalized and converted to % folded. Curve was fitted by non-linear least square fitting using Equation 1.
3.3 Aggregation and Aggregation Inhibition of TDP-43 by Oligonucleotides

Reduction of ionic strength of the buffer from eluting conditions (40 mM HEPES-KCl, 500 mM KCl, 20 mM imidazole, 20 mM MgCl₂, 2 mM βME, 10% glycerol, pH 7.4) to reduced salt conditions (170 mM KCl, 36 mM HEPES, 25 mM Imidazole, 18 mM MgCl₂, 1% glycerol, 1.8 mM βME, pH 7.4) reliably and reproducibly induced aggregation of the recombinant protein. Using this method, the effectiveness of various nucleotides to inhibit TDP-43 aggregation was tested.

The first small molecule tested was the TG₁₂ single strand DNA. 2 µM vYFP-TDP-43 was placed under aggregation conditions in the presence and absence of 5-fold molar excess of TG₁₂ and the sizes of the particles were monitored using dynamic light scattering (Figure 8). In the absence of any compounds under aggregation conditions after 4 hours, the majority of aggregates had hydrodynamic radii of 105.6 nm, with a significant secondary population up to 500-700 nm in radii. The estimated molecular weight of the average particle is approximately 18300 kDa (Figure 8A). In contrast, in the presence of 10 µM TG₁₂ ssDNA, the average particle size was 5.34 nm in hydrodynamic radius, corresponding to a molecular weight 170 kDa, close to the theoretical size of the dimer 154 kDa with two strands of TG₁₂ (7.6 kDa) bound (Figure 8B).

CD-spectroscopy was performed on TG₁₂ ssDNA and the mixture of vYFP-TDP-43 and TG₁₂ bound complex (Figure 9). Characteristic signals of ssDNA was observed at 248 and 280 nm for TG₁₂, while the mixture of TG₁₂ and vYFP-TDP-43 generated a CD spectrum that appears to be the sum of the two individual spectra. This finding suggests that binding of TG₁₂ to vYFP-TDP-43 does not greatly alter the secondary structure of the protein. These results indicate that TG₁₂ is an effective inhibitor of aggregation under these experimental conditions, and that inhibition
Figure 8: Size distribution by mass of vYFP-TDP-43 upon aggregation determined by dynamic light scattering. A) 2 µM vYFP-TDP-43 were placed under aggregation conditions for 4 hr at 20 ºC. B) Same conditions were applied to the sample in the presence of 10 µM TG12. Measurements of $R_h$ and MW were calculated using appropriate software (see Methods).
Figure 9: Circular dichroism spectrum of TG$_{12}$, vYFP-TDP-43, and their mixture. Circular dichroism was measured at 25 °C with 16s averaging times. Signals for 8.8 µM vYFP-TDP-43 (blue), 14.9 µM TG12 (red) and their mixture (black) are shown. Characteristic peaks for ssDNA and β-signals are noted.
occurs by preservation of the protein’s native state, and does not alter its secondary structure.

To quantify the potency of aggregation inhibition, a concentration dependency assay was carried out using TG₁₂ ssDNA monitored by solution turbidity (right angle light scattering at 400 nm). Under the same conditions as the dynamic light scattering assays, vYFP-TDP-43 was induced to aggregate in the presence varying concentrations of TG₁₂ ssDNA (0-2 μM) and AC₁₂ ssDNA as negative control (0-10 μM) (Figure 10). A concentration dependence of inhibition was observed with the TG₁₂ ssDNA and the data was fitted using non-linear least square fitting into the equation

\[ y = \frac{y_{\text{max}} - y_{\text{min}}}{1 + \left(\frac{x}{EC_{50}}\right)^n} + y_{\text{min}} \]  

(Equation 1), where \( y \) is the observed normalized scattering signal, and \( x \) is the compound concentration. An effective concentration (EC₅₀) of 134 nM TG₁₂ was observed for the aggregation inhibition of a sample containing 2 μM vYFP-TDP-43. The maximum effect of inhibition appears to have been achieved at 1 μM TG₁₂, a one to two molar ratio of compound and recombinant protein. For samples treated with AC₁₂, no effect was observed even at 5-fold molar excess of oligonucleotide.

Previous studies have indicated that the number of TG repeats correlate to the binding affinity of ssDNA to TDP-43, and that more than three TG repeats are required for binding (Bhardwaj et al. 2013). To evaluate whether the number of TG repeats also influence the potency of inhibition of aggregation, we repeated the concentration dependence experiment using ssDNA containing 9, 6, and 3 TG repeat lengths. Additionally, phosphorothioate derivatives of these ssDNA candidates were used in the experiment, denoted by sTG₉, sTG₆ and sTG₃. These compounds are derived by replacing the oxygen atom on the phosphate backbone of DNA with a sulfur atom, rendering the phosphate backbone resistant to DNAse digestion. This particular modification was used as it confers greater \textit{in vivo} stability to the compounds if future experiments warrant their
Figure 10: Inhibition of vYFP-TDP-43 aggregation using TG_{12} monitored by right angle light scattering. 2 \mu M vYFP-TDP-43 was incubated with varying concentrations of TG_{12} (black) and AC_{12} (white) ssDNA under aggregation conditions for 4 hours. Turbidity was determined by right angle light scattering at 400 nm. The inhibition curve of TG_{12} was fitted by non-linear least square fitting using Equation 1.

\[
EC_{50} = 134 \pm 9 \text{ nM}
\]
introduction into cell culture or *in vivo* applications. In our experimental conditions we find that sTG12 and TG12 had similar EC50 values and both potently inhibited vYFP-TDP-43 aggregation. Using phosphorothioated derivatives of these poly-TG oligonucleotides, the chain length dependence of aggregation inhibition was tested (Figure 11). Fitting each inhibition curve to Equation 1, the EC50 values for each compound was obtained. The sTG3 compound, similar to the sAC12 control, did not confer any aggregation inhibition, but sTG6, sTG9 and sTG12 all had significant effects. Their EC50 values were 365 nM, 262 nM and 201 nM respectively. These results suggest that inhibition of aggregation correlates with the strength of binding between the oligonucleotide and the vYFP-TDP-43, and may suggest that the inhibition is achieved through binding.

Although TG12 repeats show strong binding to vYFP-TDP-43, the majority of targets of TDP-43 identified by UV-CLIP assays do not have long stretches of consecutive UG residues (Xiao et al. 2011). To assess whether oligonucleotides already present in the cell are capable of inhibiting TDP-43 aggregation, biologically relevant targets of TDP-43 are examined as potential inhibitors.

Using TG12 and AC12 as positive and negative controls respectively, the aggregation inhibition of vYFP-TDP-43 by CLIP34nt RNA, HIV1LTR ssDNA, and HIV1LTR dsDNA were examined using right angle light scattering at 400 nm using previously described methods and conditions (Figure 12). EC50 values for CLIP34nt RNA and HIV1LTR ssDNA are around 2 μM. While those values are higher than that of the TG12 control (205 nM) they demonstrate that natural ligands of TDP-43 can inhibit aggregation of this protein. The HIV1LTR dsDNA did not have an effect on the aggregation of vYFP-TDP43 despite its reported ability to bind to this protein determined by previous studies (Ou et al. 1995).
Figure 11: Inhibition of vYFP-TDP-43 aggregation using varying lengths of TG-repeats. 2 µM protein was incubated under aggregation conditions with varying concentrations of phosphorothioated DNA for 4 hours. Turbidity was determined by right angle light scattering at 400 nm. EC$_{50}$ values were obtained for sTG$_{12}$, sTG$_{9}$, and sTG$_{6}$ by non-linear least square fitting using Equation 1.
Figure 12: Inhibition of vYFP-TDP-43 aggregation using various natural oligonucleotide binding targets. 2 µM protein was incubated under aggregating conditions with varying concentrations of DNA for 4 hr. Turbidity was determined by right angle light scattering at 400 nm. EC₅₀ values were obtained by non-linear least square fitting using Equation 1.
3.4 Irreversible Nature of vYFP-TDP-43 Aggregation

Our hypothesis suggests that by binding to the native, dimeric state of vYFP-TDP-43, native-state binding molecules can prevent the protein from entering the aggregation pathway. The following experiments assess whether these compounds have disaggregation properties once the protein has already entered the aggregated state. Samples were observed under a fluorescence microscope after 20 minutes under aggregating conditions in the presence or absence of varied length TG compounds (Figure 13A). By visual inspection, samples induced to aggregate in the presence of TG12, TG9, and TG6 showed largely diffuse distributions, with few aggregates. Untreated samples and those treated with TG3 or AC12 exhibited large aggregates of up to 100 μm in diameter. This result is in accordance with the right angle light scattering assays previously performed. When samples were subjected to aggregation for 10 minutes, followed by addition of various ssDNA molecules for 10 more minutes, no changes in aggregation were observed (Figure 13B). All samples contained aggregates of similar size to untreated samples and those pre-treated with AC12 or TG3. These results indicate that although native state binding molecules can reduce the propensity for entry into the aggregated state, pre-formed aggregates cannot be dissolved by the addition of the compound, and that the aggregation process is irreversible.
Figure 13: Fluorescence microscopy of vYFP-TDP-43 aggregates. Images were taken at 20°C targeting the vYFP tag (λ<sub>ex</sub> = 515 nm, λ<sub>em</sub> = 528) using 10× objective and 40× lens. A) 2 μM vYFP-TDP-43 were incubated for 20 minutes under aggregating conditions in the presence of various ssDNA. B) Various ssDNA were added to pre-formed aggregates incubating for 15 minutes and images were taken 5 minutes after ssDNA addition.
3.5 Morphology of TDP-43 Aggregates

Solutions of vYFP-TDP-43 incubated for 20 minutes at room temperature in the presence of CLIP34nt RNA, TG12 ssDNA, and in the absence of oligonucleotides were mounted onto a mica surface and observed using atomic force microscopy. The resulting images show formations of small clusters of globular structures ranging from 10 to 100 nm in diameter (Figure 14). The aggregates are non-fibrillar and show no apparent order.
Figure 14: Atomic force microscopy of vYFP-TDP43 aggregates. Samples induced to aggregate for 20 minutes in the presence of No DNA, CLIP34nt RNA and TG12 ssDNA are shown. Samples are placed on mica surface and measured at 0.5 – 5 µm scale.
Chapter 4

4 Discussion

4.1 TDP-43 Aggregation is Inhibited by Native State Binding Ligands

Aggregation of TDP-43 is a common hallmark of many neurodegenerative diseases such as ALS and FTLD. The mechanistic details of aggregation is largely unknown. Evidence has suggested many factors influence the solubility of the protein such as heat shock (Zhang et al. 2011; Udan-Johns et al. 2013), recruitment into stress granules (Parker et al. 2012), and cleavage and phosphorylation of the C-terminal region (Liachko et al. 2010). A recent study has also independently demonstrated the effectiveness of ssDNA and RNA containing TG$_{12}$ or UG$_{12}$ as effective inhibitors of TDP-43 aggregation using refolded TDP-43 (Huang et al. 2013). In this study, we demonstrate that naturally occurring sequences present in cells can also increase the solubility of TDP-43, and provide measurements of inhibition potency of naturally occurring and artificial sequences using soluble, vYFP-tagged TDP-43. We also demonstrate the irreversible nature of TDP-43 aggregation and offer insight into the morphology of said aggregates using microscopy methods.

The results suggest that native-state binding molecules such as poly-TG ssDNA and CLIP34nt are able to reduce the aggregation propensity of recombinant TDP-43 in vitro, as hypothesized. Furthermore, the binding of these compounds allow TDP-43 to retain its native, dimeric state, under aggregating conditions. In the assays for aggregation inhibition using different TG-repeats, we demonstrated that inhibition potency likely correlates with binding affinity of the oligonucleotide. The exact molecular mechanism of binding, however, is subject of speculation. In our aggregation inhibition assays, we observed that a 1:2 molar ratio of TG$_{12}$ to protein was
sufficient to achieve maximum inhibition (Figure 10). This substoichiometric inhibition suggests the possibility a single TG₁₂ molecule binding to multiple vYFP-TDP-43 molecules, possibly functioning as a tether for their native dimer configuration to stabilize this native state and reducing its propensity for monomerization under aggregation conditions.

4.2 Applications of Phosphorothioate DNA Derivatives

Our experiments demonstrated that phosphorothioated derivatives of TGₙ compounds are also effective in inhibiting TDP-43 aggregation *in vitro* (Figure 11). Specifically, the EC₅₀ value of sTG₁₂ (201 nM) is comparable to that of the unmodified counterpart, TG₁₂ (134 nM). S-oligonucleotides have successfully been used in antisense nucleotide treatment to deliver siRNA resistant to nucleases into targeted cellular compartments in the treatment of cancer and other diseases (Wang et al. 1995; Wang et al. 1999; Tian et al. 2011). Studies of these compounds have led to a variety of systems for cell delivery (Detzer et al. 2008). The vast number of applications of these compounds are found in several reviews (Stein & Cohen 1988; Dias & Stein 2002). With the advancement in delivery systems of S-oligonucleotides in patients, the sTGₙ compounds we have shown to inhibit TDP-43 aggregation may serve as a potential starting compound towards the design of chemical chaperones with clinical applications. An example of a successful nucleotide-derived drug is the deoxythymidine derivative 3'-Azido-3'-deoxythymidine (AZT) designed for the treatment of AIDS (Mitsuya et al. 1985).

4.3 Mechanisms of TDP-43 Aggregation

There is a general consensus that the C-terminus of TDP-43 is a key player to this aggregation pathway (Igaz et al. 2009; Budini et al. 2012). The C-terminus of TDP-43 is a GQN-rich prion-like domain and has been shown to have prion-like properties (Gitler & Shorter 2011; Budini et al. 2012; Nonaka et al. 2013). Specifically, expression of 12 tandem repeats of the residues 331-
369 of the C-terminus containing Q/N rich sequences can recruit native TDP-43 into aggregates (Budini et al. 2012), and seeding of C-terminal fragments from diseased brains into cell culture can recapitulate disease phenotypes (Nonaka et al. 2013). Morphologically, we find that the aggregates formed do not share characteristic fibrillar structure of prion aggregates (Figure 14), but are instead. However, this unstructured C-terminal region is not required for oligonucleotide interactions (Kuo et al. 2009) and is not expected to change conformations upon substrate binding. Yet, we find that native-state binding molecules effectively reduce aggregation caused by this C-terminal region without physical interactions. This suggests that there are likely functional differences with the C-terminal region of TDP-43 when the protein is dimeric versus monomeric, and that, the retention of TDP-43 in its native state by the binding of native-state binding compounds prevents the C-terminus from engaging in aggregation-prone interactions with C-termini of non-native TDP-43 in vitro. Native state stabilization via DNA binding has also been observed in other proteins, namely the tumor suppressor p53, where binding of a consensus DNA sequence to p53’s core domain (p53C) can stabilize the full length protein (Ishimaru et al. 2009).

In our proposed hypothesis, we also speculated the monomerization of TDP-43 is likely to initiate the aggregation pathway, and that once aggregates have formed, the process becomes irreversible. This was supported by the microscopy results, where pre-formed aggregates were not rescued by the addition of oligonucleotides.
4.4 Implications of TDP-43 Aggregation Inhibition by Naturally Occurring Targets

TDP-43 has a variety of cellular roles and the C-terminus has been implicated in many protein-protein interactions such as binding to Heterogeneous nuclear ribonucleoproteins (hnRNPs) (Buratti et al. 2005; McDonald et al. 2011) to participate in mRNA splicing, degradation, stabilization and other various functions. In cells where these protein-protein interactions are prevalent, many factors likely contribute to TDP-43 solubility. From our findings, it is likely that cognate binding to its mRNA targets is an additional factor to maintaining TDP-43 solubility under normal conditions. A recent study has identified that under native conditions, TDP-43 C-terminus likely self-interacts, but upon destabilization of this self-interaction, the C-terminus can undergo misfolding, resulting in non-functional aggregates (Wang et al. 2012). It is likely that under pathological conditions, various factors contribute to the disruption of TDP-43’s native state, which may include disruption of binding of TDP-43 to its mRNA targets, allowing the C-terminus to become pathogenic.

TDP-43 levels in cells are regulated through a negative feedback loop where TDP-43 binds to the 3’UTR of its own mRNA to target it for degradation (Ayala et al. 2011). We find that binding of CLIP34nt (a sequence on the 3’UTR of TDP-43 mRNA) to TDP-43 can potentially increase TDP-43 stability. Interestingly, this mRNA is likely to be easily accessible by nascent TDP-43 polypeptide synthesizes from the ribosome. Binding of TDP-43 to its own mRNA may have a two-fold effect; it may increase the solubility of the newly synthesized protein, while simultaneously targeting the mRNA for degradation. Once degraded, however, the protein loses this native-state binding molecule and becomes more aggregation-prone. Such a finding suggests additional layers of complexity in the spatial-temporal regulation of TDP-43 that previously may not have been considered.
5 Future Directions

In this study, we have demonstrated that a number of oligonucleotide compounds, synthetic and naturally occurring, can improve TDP-43 solubility under aggregating conditions by increasing the population of the protein’s native state \textit{in vitro} and the potency of aggregation inhibition likely correlates with binding affinity between the compound and protein. In accordance to our hypothesized model, aggregation can be prevented but once formed, cannot be reversed by the addition of these native-state binding molecules.

The next step in the study would assess the viability of these compounds in cell culture or animal models. To do so, a cell culture model for TDP-43 aggregation is required. A recent study has led to the development of a cell culture model of TDP-43 that recapitulates aggregation (Budini et al. 2012). In this model, tandem repeats of residue 331-369 on the C-terminal region of TDP-43 containing QN-repeats is expressed in cell culture. The expression of this region is able to sequester native TDP-43 to aggregate in inclusion bodies, consistent with the hypothesis that exposed, misfolded C-termini of TDP-43 can act as a seed to pull the folding equilibrium of TDP-43 towards the aggregation-prone states. Addition of phosphorothioated versions of CLIP43nt or poly-TG oligonucleotides used in our assays into this system may produce significant results by counteracting the effect of expressed tandem repeats and ameliorate aggregation through native-state binding.

In the long term, these native-state binding molecules may serve as a starting compound for drug development against diseases of the FTLD/ALS spectrum.
References


characteristics of C9orf72 expansion carriers compared with GRN, MAPT, and nonmutation carriers in a Flanders-Belgian FTLD cohort. *JAMA Neurol.* 70:365–73.


