The structural characterisation of two DNA protectants during stress; the tandem RRM domains of mouse TDP-43 and *E.coli* DPS

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy

By

James A. Austin

November 2013

Abstract

TAR DNA Binding protein (TDP-43) is a member of the heterogeneous nuclear ribonucleoprotein family with crucial splicing, transport and regulatory function of genetic material inside mammalian cells. Unfortunately, TDP-43 positive cytoplasmic aggregates occurring with post-translational modifications are a common hallmark in neurodegenerative diseases observed in Alzheimer's, Parkinson's, amyotrophic lateral sclerosis (ALS) and fronto-temporal lobar degeneration (FTLD) diseases. Mutations in the *TARDBP* gene responsible for encoding TDP-43, have been directly correlated with onset of ALS and FLTD. Disease models describing TDP-43 proteinopathy suggests onset may derive through either cytoplasmic mis-localisation or a loss of nuclear function but it is unclear if or how disease associated point-mutations contribute to these observations. In order to determine the effects these mutations have on the protein, a fragment containing the tandem RRM domains (residues 101-265), responsible for the proteins nucleic acid binding function was tested. Using small angle X-ray scattering, circular dichroism, isothermal titration calorimetry and thermal assay methodology it was demonstrated that initial structures of all variants are similar but mutations (D169G and K263E) confer resistance to thermal denaturation by up to 4.9 ± 0.6˚C. This stability positively correlated with an increase in half-life when tested in the full-length variant using a neuron cell model suggesting that protein turn-over is a contributing disease factor.

This study was also concerned with solving an X-ray crystallographic DNA binding complex structure for *E.coli* DPS and mapping interactions with neighbouring DPS complexes. These mechanisms are important in DPS function to protect nucleic acids during prokaryote stress. DPS is conserved in almost all prokaryotes however not all species can interact with DNA. Using X-ray crystallography, a model of *E.coli* DPS was built to 2.8 Å resolution from DNA containing samples showing both DNA and N-terminal residues were absent. Stabilising polar interactions were shown to form between neighbouring dodecamer structures involving T12, R18, D20, N99, S100, S106 and K134. Polar contacts are observed in all compared crystallographic structures from different species but the residues involved are poorly conserved, despite strong similarities between sequence and structure. This suggests that these contacts may contribute to stabilising the DNA-DPS complexes but form indiscriminately between exposed polar residues available on the dodecamer surface. These interactions are likely to contribute to the thermal stability of DNA-DPS complexes to aid in the proteins protective function.

Table of Contents

Acknowledgements

Thank you to my supervisors Prof. Samar Hasnain, Dr. Svetlana Antonyuk and Dr Guenter Grossmann for support, direction and encouragement.

My thanks and gratitude also go to Dr Gareth Wright for the advice and numerous paper revisions.

I'm grateful to Dr Mark Wilkinson and Mr Mark Prescott for performing the Mass Spectrometry and Dr Hyun Lee for help with setting up the SAXS experiments.

A big thank you to all in Lab D and the Molecular biophysics group from 2009-2014 at the University of Liverpool.

Finally a huge thanks to Laura, my family and friends who have provided support and distraction when needed over the years.

Table of Figures

Appendices

Tables

Abbreviations

ALS – Amyotrophic lateral sclerosis

BSA – Bovine serum albumin

Cα – Alpha carbon

CD – Circular dichroism

CPL – Circular polarised light

DPS – DNA binding protein from starved cells

DREF - DNA replication-related element binding factor

DSF – Differential scanning fluorimetry

EDTA - Ethylenediaminetetraacetic acid

ESCRT - Endosomal sorting complex required for transport

fALS – Familial amyotrophic lateral sclerosis

FTLD – Fronto-temporal lobar degeneration

FTLD-U – Front temporal lobar degeneration with ubiquitinated inclusions

GST – Glutathione-s-transferase

HIV-1 – Human immunodeficiency virus strain type-1

HnRNP – Heterogeneous ribonucleoprotein particle

IPTG - Isopropyl β-D-1-thiogalactopyranoside

MBG – Molecular biophysics group

MS – Mass spectrometry

mRNA – Messenger RNA

NES – Nuclear export sequence motif

NLS – Nuclear localisation sequence motif

NMR – Nuclear magnetic resonance

PTB – Polypyramidine tract binding protein

PX – X-ray crystallography

RMSD – Root mean squared deviation

RNP – Ribonucleo protein

RRM – RNA recognition motif

sALS – Sporadic amyotrophic lateral sclerosis

SAXS – Small angle X-ray solution scattering

SOD1 –Superoxide dismutase type-1

SPR – Surface plasmon resonance

TAR – Trans active response element

TARDBP – TDP-43 gene

TDP-43 – TAR DNA binding protein

UTR – Untranslated region

Chapter I

TDP-43 proteinopathy

1.1 TDP-43 proteinopathies

Mis-localised TAR DNA binding protein with a molecular weight of 43 kDa (TDP-43) is a common hallmark in many human neurodegenerative diseases. During TDP-43 related onset, the protein commonly forms cytoplasmic aggregates containing C-terminal truncates that are post-translationally modified by hyper-phosphorylation and ubiquitination in neurons and surrounding glial cells [1,2]. Mutations of the *TARDBP* gene located on the 1p36 locus [3] responsible for encoding TDP-43 have been directly associated with onset of amyotrophic lateral sclerosis (ALS) and fronto-temporal lobar degeneration (FTLD), both regarded as a spectrum of TDP-43 proteinopathy [4]. It is not uncommon for both diseases to manifest within the same individual to support this classification [5].

ALS, also referred to as common motor neuron disease or Lou Gehrig's disease in the United States, is associated with neuronal loss in the spinal cord leading to muscle weakness and atrophy [6]. The disease is progressive and initial symptoms can include dysarthria, dysphagia, sialorrhoea, dysphonia, limb weakness, cramps and paralysis [7]. As the disease progresses, reduced activity and inability of limb movement, loss of speech and ability to swallow lead to a loss of independence and reliance on full-time care [8]. Death typically occurs within five years after the onset of symptoms for the majority of patients, often from respiratory failure [9]. Around 10% of ALS is familial (fALS), arising from hereditary phenotypes, whereas the remaining 90% is described as sporadic (sALS) with unknown cause [10].

Similarly, FTLD is characterised by neuronal loss in the brain's front temporal lobes. Symptoms present as behavioural changes, including emotional outbursts, progressive non-fluent aphasia and dementia [11]. FTLD accounts for ~20% of pre-

senile dementia, manifesting between 45-65 years of age and is the second largest cause of global dementia after Alzheimer's disease [12,13]. These diseases differ by the level of a patient's preserved cognitive ability. FTLD patients are likely to retain memory, perception, spatial awareness and praxis [13] which are more often disrupted during the onset of Alzheimers'. FTLD is progressive and prognosis for survival is typically 8 years after onset [13]. There are no effective cures for ALS or FTLD, thus, both cause terminal illness.

TDP-43 aggregates are present in ~95% of ALS, ~50% of FTLD patients and in ~15% of patients manifesting both [14]. Globally, the prevalence of ALS is ~4 per 100,000 individuals [15] where mutations in the TDP-43 gene are associated with ~1.5% sALS [16] and 1-5% of fALS [17,18]. *TARDBP* mutations have an estimated prevalence within FTLD patients of 1.9-2.6%, as determined through a small Italian patient population of 252 individuals [19]. As a proportion of global dementia (estimated as ~4700 per 100,000 [20]) FTLD contributes to ~10-15% of cases, in which TDP-43 proteinopathy is present in ~55% known as FTLD with ubiquitinated inclusions (FTLD-U) [21]. Considering these high prevalence rates, rare *TARDBP* mutations could affect a significant number of individuals globally to warrant their study and expose potential causes for TDP-43 mis-localisation and disease onset. TDP-43 inclusions are also present in 20-25% of Alzheimer's [21] and Parkinson's disease cases [22] suggesting that the cause of onset in some of these cases may also derive from TDP-43 dysfunction.

1.2 TDP-43 function

TDP-43 is ubiquitously expressed in mammalian cell types and is a member of the heterogeneous ribonucleoprotein (hnRNP) family which have an integral role in manipulating a wide range of nucleic acid targets. The protein is a dimer [22] and performs a variety of functions including; splicing [23], transport [24], transcription [25,26], translation [27] and stress granule formation [28]. The first role described was the inhibition of the transcription of HIV-1 DNA through binding the trans active response (TAR) sequence motif [26] for which it is named thereafter. Thousands of nucleic acid targets have been predicted to bind TDP-43, including

both single and double stranded DNA and RNA ligands. These include the cyclin dependent kinase-6 gene [29], the pre mRNA of the cystic fibrosis transmembrane conductance regulator [30] and histone deacetylase 6 [31], where TDP-43 has direct splicing and regulatory functions.

In the adult brains of mice, TDP-43 is estimated to bind a third of all mRNA [32] with a similar number predicted in humans [33]. Hence TDP-43 is likely to have an integral role in the processing of genetic material, exemplified during TDP-43 knock-down that coincides with developmental defects and shortened life-span of eukaryotic organisms including mice [34–36], drosophila [37,38,31] and zebra fish [39]. Consequently, removal of TDP-43 in eukaryote models lead to lethal phenotypes emphasising the proteins importance during development and normal cell function.

TDP-43 is localised inside the nucleus. However, its natural function as an mRNA transporter requires that it fluxes between the cytoplasmic and nuclear compartments as an mRNA transport granule [40]. In this role, it is able to co-localise with fragile X mental retardation protein and staufen-1. These two proteins are involved in local translation at synapses in rat hippocampal neurons, specifically relating some of TDP-43's function as a neuronal activity responsive factor in neuronal cells [27].

The protein is capable of regulating its own expression by binding to the 3' UTR of its mRNA as part of a negative feedback loop [41]. It is postulated that this may be important during TDP-43 mediated disease models [42] discussed in the subsequent section. During stress TDP-43 can be recruited into stress granules [28,43] to protect mRNA which requires cytosolic redistribution from the nucleus which can then return once the stress-response is resolved [44,45].

1.3 TDP-43 disease models

TDP-43 mediated disease onset has been postulated by two different models: the loss or gain of TDP-43 auto-regulation. Both models are based on TDP-43's ability to

inhibit its own protein expression by binding to the 3' UTR of its mRNA (Figure 1.1) [41]. Each model describes TDP-43 cytoplasmic mis-localisation as a feature of neurodegeneration but the causes of toxicity are attributed to which cellular compartment auto-regulation occurs.

The 'loss of autoregulation' model assumes this function is performed within the nucleus where a reduction in nuclear TDP-43 up regulates nuclear expression. Here, an unknown stress causes the cytoplasmic mis-localisation of TDP-43, perhaps by stress granule recruitment that redistributes the protein from the nucleus which in-turn up-regulates TDP-43 nuclear expression. If the stress remains, newly synthesized protein will accumulate in the cytoplasm in a cycling manner. Increasing TDP-43 cytoplasmic concentrations undergo post translational modification and form inclusions that exacerbate the stress leading to a gain of cyto-toxicity and neuron degeneration.

Alternatively, TDP-43 is regulated in the cytoplasm as shown in the 'gain of auto-regulation' model (Figure 1.1B). Here elevated cytosolic TDP-43 levels down-regulate expression. Through an unknown stress, nuclear TDP-43 is again redistributed to increase cytosolic concentrations, though initially it cannot regulate its mRNA due to participation in other functions e.g. stress granule formation. The total TDP-43 concentration is initially increased, however over-time cytoplasmic species restore this auto-regulatory function leading to a reduction of newly synthesized protein, reduction in nuclear TDP-43 concentrations and a disruption of normal TDP-43 nuclear function, that causes neurodegeneration. The removal of TDP-43's nuclear localisation sequence by N-terminal cleavage may worsen the effect in this model since truncates cannot relocate back to the nucleus [42].

Evidence of disease associated mis-localisation is observed in animal models that support both 'auto-regulatory' models described for TDP-43 proteinopathy. For example, increasing levels of cytosolic TDP-43 have been correlated with neurodegeneration in primary rat cortical neurons, independent of inclusions and nuclear clearance [46], with similar observations in mice [47]. This exposes TDP-43 mis-localisation as a key disease feature. The formation of aggregates have been

shown to cause toxicity in yeast models [48,49] however symptoms of disease without inclusions are observed in neurons of rodent models [50,51] that implies aggregation need not be necessary but may increase the severity of toxicity. It is worth note that knock-down of TDP-43 leads to developmental defects and a reduction in survival in eukaryotic model organisms, to also support a loss of normal TDP-43 function as a viable cause of TDP-43 mediated neurotoxicity [31,34–39].

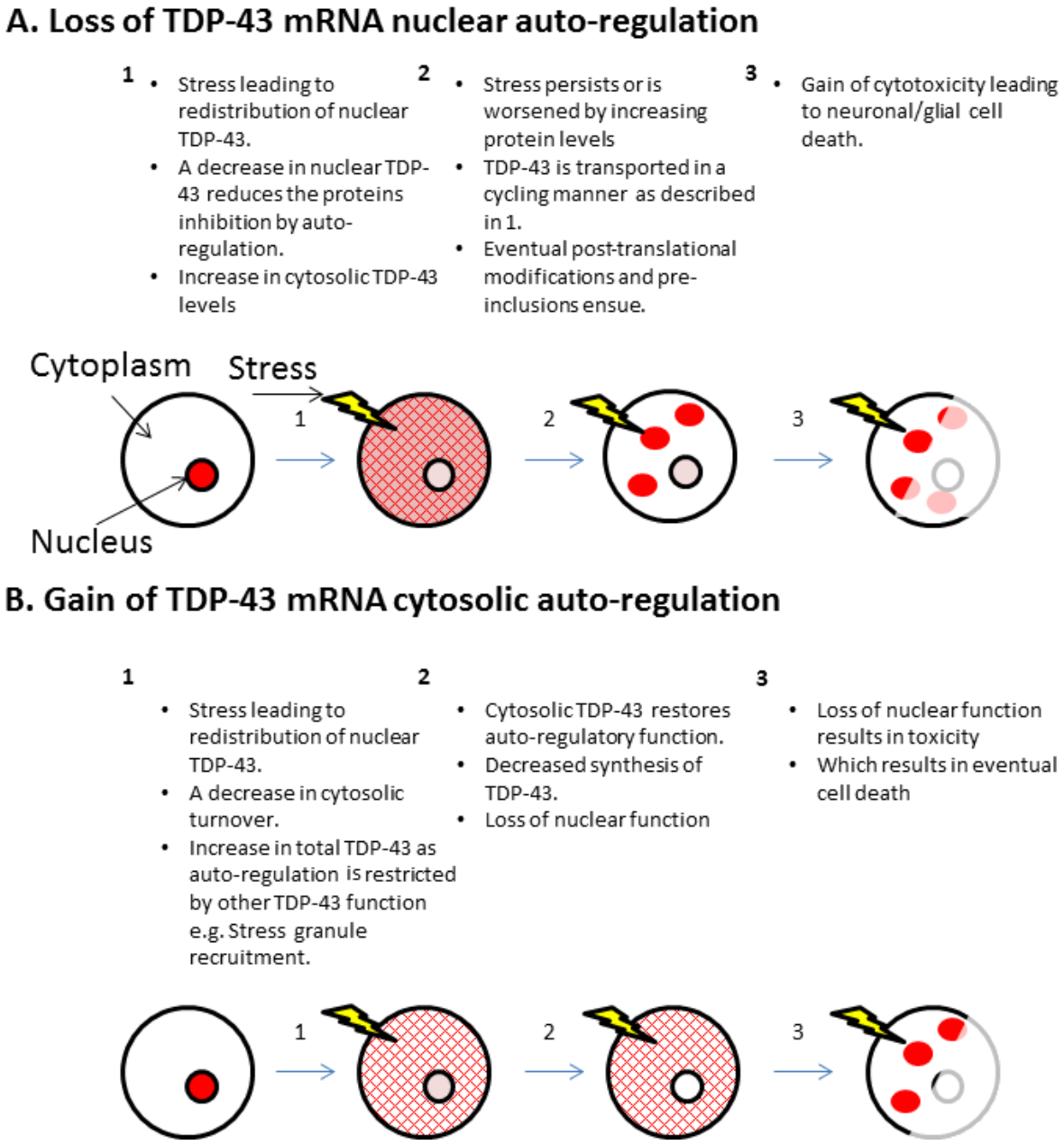


Figure 1.1 TDP-43 disease models. A, Loss of auto-regulation model and **B,** Gain of auto-regulation model, adapted from Lee *et. al.* 2011 [42].

1.4 TDP-43 mutations

There are forty identified ALS and FTLD-U disease related mutations situated predominately on the C-terminal domain of TDP-43 (Figure 1.2). Only three are situated elsewhere: the A90V mutation in the N-terminus, D169G within RNA recognition motif (RRM) 1 and K263E situated immediately after RRM2. At present only one C-terminal disease related mutation R361S has been shown to change a function of TDP-43, manifesting as an impairment to stress granule formation, whilst the D169G mutation had no effect to function [52]. However, other non-disease related C-terminal mutations have been shown to impair the self-aggregation potential of TDP-43 through a hydrophobic patch between residues 318-343. Here two mutations: A324E and M337E could individually reduce TDP-43 self-aggregation and completely perturb this function *in vitro* as a double mutation [53]. Interestingly, M337 is an ALS disease related locus and therefore the ALS associated M377V mutation may be important in the recruitment of TDP-43 into stress granules.

Unfortunately there are a limited amount of published data available exploring the relationship between C-terminal mutations and function to accurately determine whether mutant mediated disease onset arises through changes to TDP-43 cellular activity. Furthermore only 32% of the C-terminal domain has any function assigned, limited to residues ranging between 318-366 [53,54] meaning it is still unknown whether at least 62% of disease related loci cause a change to function [54] (Figure 1.3). The annotation of the C-terminal domain is further complicated by additional functions, as identified in TDP-43's interaction with hnRNPA2 mediated by a C-terminal binding motif that overlaps with the previously described hydrophobic patch. This region between residues 321-366 was shown as important in the joint splicing function involving hnRNPA2 and TDP-43. In addition, disease related mutations: Q331K, M377 and G348C do not perturb this binding or activity [55]. As such, there is no typical trend that links TDP-43 C-terminal disease mutations with a loss of function due in part by a lack of functional annotation and proportion of mutations tested.

A growing number of C-terminal domain disease mutations are being identified with a tendency to increase cellular longevity *in vivo*. This has been observed in neural and isogenic cell-line models and include a significant number of disease related mutants. In Neuro-2A cells a correlation was observed between cellular half-life and accelerated disease onset through several of these mutations [56]. These included: A315T, Q343R, N352S, G298S, M337V, G348C and A382T. Each mutant had a longer half-life than wild-type and a slight tendency to localise into the cytoplasm. This is in agreement with disease correlations in primary rat cortical neuron models [46]. This finding was also supported by ALS related mutants G298S, Q331K and M337V enhancing longevity in isogenic cell-lines and G298S even degraded a third slower than wild-type in primary human fibroblast cells [57]. A number of similar studies conclude that mutant turnover can increase half-life 3-fold more than wild-type to suggest that cellular mis-localisation and protein turnover are important disease determinants.

Furthermore, mutants that increased cellular stability also positively correlates with accelerated disease onset in Neuro-2A cell lines [56]. This implies that TDP-43 stability could have a direct effect on clinical prognosis. The frequency of mutations associated with ALS can be observed in Figure 1.3 adapted from Pesiridis *et. al* 2009, looking into ALS mutant frequency in patients [54]. Note that most of these mutations are situated on the C-terminal domain, however the most frequent; A382T has not been attributed to any change in function.

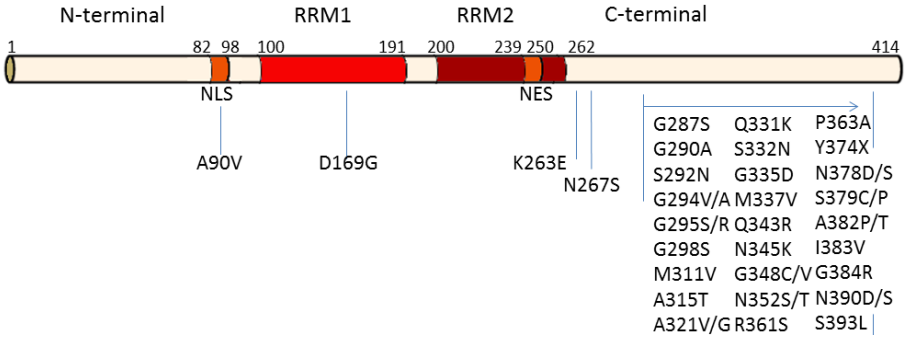


Figure 1.2 TDP-43 domain dissection and neurodegenerative disease point mutations. TDP-43s' functional domains containing; the N-terminal with the nuclear localisation signal (NLS), the RNA recognition motifs (RRM) 1 and 2 with the nuclear export signal (NES) and C-terminal. ALS and FTLD-U mutations are situated throughout the protein with the majority localised on the C-terminal domain.

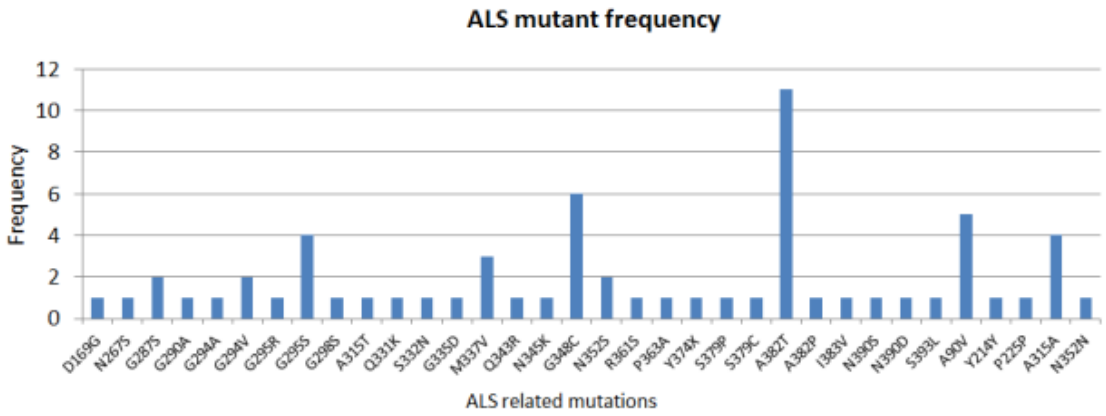


Figure 1.3 Frequency of TDP-43 ALS disease mutations. Summary of mutation frequencies identified in patients with ALS, adapted from Pesiridis *et. al.* 2009 [54].

1.5 Post-translational modifications

Post-translational modifications associated with TDP-43 inclusions are consistently found within the brains and spinal cords of patients [1]. These modifications include hyperphosphorylation, ubiquitination and protein cleavage. The truncation of TDP-43 results in 25 kDa [1,2], 27 kDa [58] and 35 kDa C-terminal fragments [59] found with full length within cytosolic inclusions. This has led to the hypothesis that the C-terminal domain with the aid of the RRM domains mediate the aggregation during disease onset [59–61]. Dissection of the C-terminal domain has determined regions that readily form amyloid-like fibrils [62] that may be significant in propagating disease, similar to seeding in prion mediated neurodegeneration [63].

Interestingly the ~25 kDa C-terminal fragments lacking RRM1 are readily degraded with a half-life of ~4 hours [64] compared to ~12 hours of full length TDP-43 in mammalian cells [57]. This implies that individually these fragments are non-toxic. However, when restricting caspase cleavage at D169, responsible for producing non-toxic 27 kDa species, neuro-toxic 35 kDa fragments are generated [58] to suggest that RRM1 residues have an active involvement in cyto-toxicity in cultured neurons.

The fact that full length TDP-43 with aid of the N-terminus can also facilitate aggregation [65] and that the N-terminal ALS associated A90V mutation increases inclusion formation, suggests that cleavage is not necessary for aggregation [66]. However, truncated fragments missing the NLS cannot return to the nucleus and may linger in the cytoplasm to enhance recruitment to pre-existing aggregates and contribute to their cytotoxicity. Although TDP-43 neurodegeneration is observed without inclusions in rodent models [50,51], fragments from inclusions can seed aggregation in healthy neurons [63] to provide a convenient explanation to describe disease progression and warrants further investigation.

The observed hyper-phosphorylation associated with TDP-43 proteinopathies occurs before ubiquitination [67,68]. This is not required for TDP-43 aggregation, truncation or toxicity in cultured cells [69,70]. Notably phosphorylation does decrease the turnover of TDP-43, at least for C-terminal fragments in human neuroblastoma cells from ~14 to ~22 hours [71]. This is similar to the enhanced half-life observed for TDP-43 C-terminal domain disease related mutations [56]. This begs the question of whether phosphorylation exacerbates toxicity through increasing TDP-43 cellular stability, via a similar mechanism as disease related point-mutations. TDP-43 hyper-phosphorylation may also confer a resistance to the ubiquitin degradation pathway, employed to remove misfolded protein [71]. This hypothesis is supported by rare functional null mutations in the *UBQLN*2 gene responsible for encoding ubiquilin-2, a protein involved in protein degradation that is associated to X-linked ALS [72].

Mutations within TDP-43 have yet to be associated with a modified phosphorylation profile to suggest they contribute to disease independently from post-translational modification. This is also likely to be true of the ubiquitination of TDP-43, which has not been associated with any disease related feature and is probably a consequence of degradation resistance.

1.6 TDP-43 toxicity: mis-localisation or loss of function ?

Current disease models using auto-regulation to describe disease onset (Figure 1.1) suggest that increased levels of cytoplasmic TDP-43 are a common factor in TDP-43 proteinopathy. The problem is that this observation is consistent through either a gain of toxicity associated with elevated cytosolic TDP-43 or as a loss of normal nuclear TDP-43 function by two exclusive paths as described in the auto-regulation models. Since disease onset is based on the location of auto-regulation, even associating cellular longevity with C-terminal mutations doesn't definitively explain how toxicity is mediated.

In support of a gain in cytosolic toxicity, disease symptoms observed in rat neurons correlated best with levels of mis-localised TDP-43 independent from inclusions or normal endogeneous levels of nuclear TDP-43 [46]. This suggests that normal nuclear function is still maintained during disease and that mis-localisation is in itself toxic. The disease mechanism in this explanation may arise through a resistance to ubiquitin mediated degradation. Problems associated with the machinery necessary to remove misfolded protein is a well-documented explanation of neurodegeneration, observed in X-linked ALS mutations in Ubiquilin-2 [72] and dysfunction within the 'Endosomal Sorting Complex required for transport' (ESCRT) pathway [73]. Though it is still important to consider that elevated levels within the cytoplasm could also indicate toxicity through TDP-43 nuclear evacuation, which may mirror toxicity observed in TDP-43 knock-down studies in mice [34–36], drosophila [37,38,31] and zebra fish [39].

Therefore, when relating disease associated mutations to disease onset, domain function should also be considered. Unfortunately, evidence to suggest that C-terminal domain mutations disrupt the function of TDP-43 is limited, as many of this domain's functional motifs are unknown or missing a large subset of mutations. Therefore, it is difficult to determine whether C-terminal mutations perturb a role of TDP-43 during disease. Thus, it is useful to pursue mutations that exist on a fully functionally characterised region with known structures and mutations to distinguish whether functional knock-down can explain common trends for disease

onset. Fortunately the RRM domains of TDP-43 are well annotated in regards to function and structure with each containing at least one disease mutation. Therefore, this study aimed to compare these domains against disease related mutations to examine whether loss of function or cytotoxicity could provide a common explanation and to substantiate further the auto-regulatory models proposed for disease onset.

Chapter II

TDP-43 structure, homology and RRM-RRM domain insights

2.1 TDP-43 structure

TDP-43 homologs in humans and mice consist of 414 amino acids. The protein can be dissected into four functional domains to include the N-terminal (residues 1-100), the RRM1 (residues 101-191), RRM2 (residues 200-262) and the C-terminal domains (residues 262-414) (Figure 1.2). The N- and C-terminal domains have no deposited structures, though they do display functional motifs involved in dimerization [65], nuclear import (residues 82-98) [74] and protein interactions [55,75]. The N-terminal is predicted to contain secondary structure by the PONDR prediction software (http://pondr.com) [62] and by computer assisted modelling [67]. The C-terminal structure is largely unknown, though a section (residues 311-360) has been shown to form an 'α-helix turn α-helix' structure shown by NMR methodology including the residues involved in functional aggregation (residues 318-343) [53]. There are no described structures for residues between 266-311 and 361-414, containing 16/27 of the identified disease related positions.

The TDP-43 nucleic acid binding region is well characterised and has three deposited models describing human RRM1 (PDB id: 2CQG), RRM2 (PDB id: 1WF0) by NMR and an X-ray crystallography model of mouse RRM2, bound to single stranded DNA, resolved at 1.65Å (PDB id: 3D2W) (Figure 2.1). The RRM1 adheres to the typical αββββα secondary structural RRM fold [76] whereas RRM2 contains an additional β-sheet arranged as αββββββα [77]. Both RRM domains contain two ribonuclear protein (RNP) variant motifs conserved in all RRM domains to orchestrate nucleic acid binding [76]. These are described as an eight (RNP1) or six residue (RNP2) consensus region described as K/R-G-F/Y-G/A-F/Y-VI/L-X-F/Y [78,79] or I/V/L-F/Y-I/V/L-X-N-L [80] respectively where X is any amino acid.

The RRM1 secondary structures from the N' to C' terminals are arranged as $\beta^1\alpha^1\beta^2\beta^3\alpha^2\beta^4$. The RNP2 spans between residues 106-111 involving the β^1 sheet whereas RNP1 is situated between residues 145-151 on the β^3 sheet anti-parallel to RNP2 (Figure 2.1). Residues W113, F147, F149 and R151 have been specifically shown to bind nucleic acids in RRM1 [81] with W133 being situated on a flexible loop between β^1 and α^1 to show that the unstructured regions within the RRM are involved in ligand interactions.

The RRM2 secondary structure is arranged as $\beta^1\alpha^1\beta^2\beta^3\alpha^2\beta^4\beta^5$ and contains the nuclear export sequence (NES) (residues 239-250) spanning between the α^2-helix, β^4-sheet and connecting loop, conserved in both human and mouse structures (Figure 2.1). This is critical in the transportation of TDP-43 by exportin-5 across the nuclear membrane [82]. Interestingly this β^4 sheet is oriented away from the remaining anti-parallel β-sheet platform and RNPs, suggesting that it is less likely to participate in nucleic acid ligand interactions. The RNP2 (residues 193-198) and RNP1 (residues 217-223) are situated on the β^1 and β^3 sheets respectively similar to RRM1. The high resolution ssDNA bound crystal structure of mouse identified residues K192, F194, R197, F221, R227, F231, S258, N259, A260 and E261 as capable of interacting with nucleic acid targets with high thymine-guanine content (TG) [77]. Given that the majority of nucleic acid contacts are orchestrated by several residues spanning the proteins anti-parallel β-barrel core. Domain unfolding is likely to change amino acid positions and possibly TDP-43's nucleic acid binding affinity leading to changes in cellular function.

The thermal stability of TDP-43's RRM domains has previously been tested by CD, to indicate that RRM2 is structurally more stable [77]. The melting temperatures using the RRM domains of mouse demonstrated that RRM2 was stable up to at least 85˚C in contrast to 49.5 ± 0.7˚C for RRM1. Sequence homology between human RRM1 and mouse RRM2 is 22.7% identical, however many similar propertied residues are conserved, especially surrounding the RNP regions described in further detail in the subsequent section (Figure 2.3). A couple of key structural differences exist between TDP-43 RRM domains that could explain this observation in stability. Firstly, the PDB structures of human RRM1 (PDB ID: 2CQG) and mouse RRM2 (PDB

id: 3D2W) can superimpose (RMSD of 0.9 Å) (Figure 2.2) to show that RRM1 has an extended loop region that does not form any additional contact with the main RRM1 body (Figure 2.2C). Without additional contact this extended loop is likely to be more flexible than its RRM2 counterpart which may lend to the unfolding of RRM1 during temperature denaturation. Secondly, the RRM2 contains more secondary structure namely from its additional β^4 sheet (Figure 2.2E) which likely conveys stability through the β-sheet order and additional contact between neighbouring residues. The aliphatic index has been shown as a good indicator of thermal stability in globular bacterial proteins, where the amount of hydrophobic aliphatic side chained residues is positively correlated with protein thermal stability [83]. The alphatic index calculated from the domain sequences (Figure 2.2,A) are 72.5 and 87.3 for human RRM1 and mouse RRM2, respectively (Expasy:Protparam), to suggest that aliphatic side-chains within these domains could mediate and substantiate the differences between domain stability.

Although RRM1 is structurally more unstable than RRM2 it is important to note that it is stable under physiological temperatures and orchestrates the tighter affinity with DNA/RNA, explained in section 2.3, which may be aided by the additional positive residues contained in its extended loop region (Figure 2.2D).

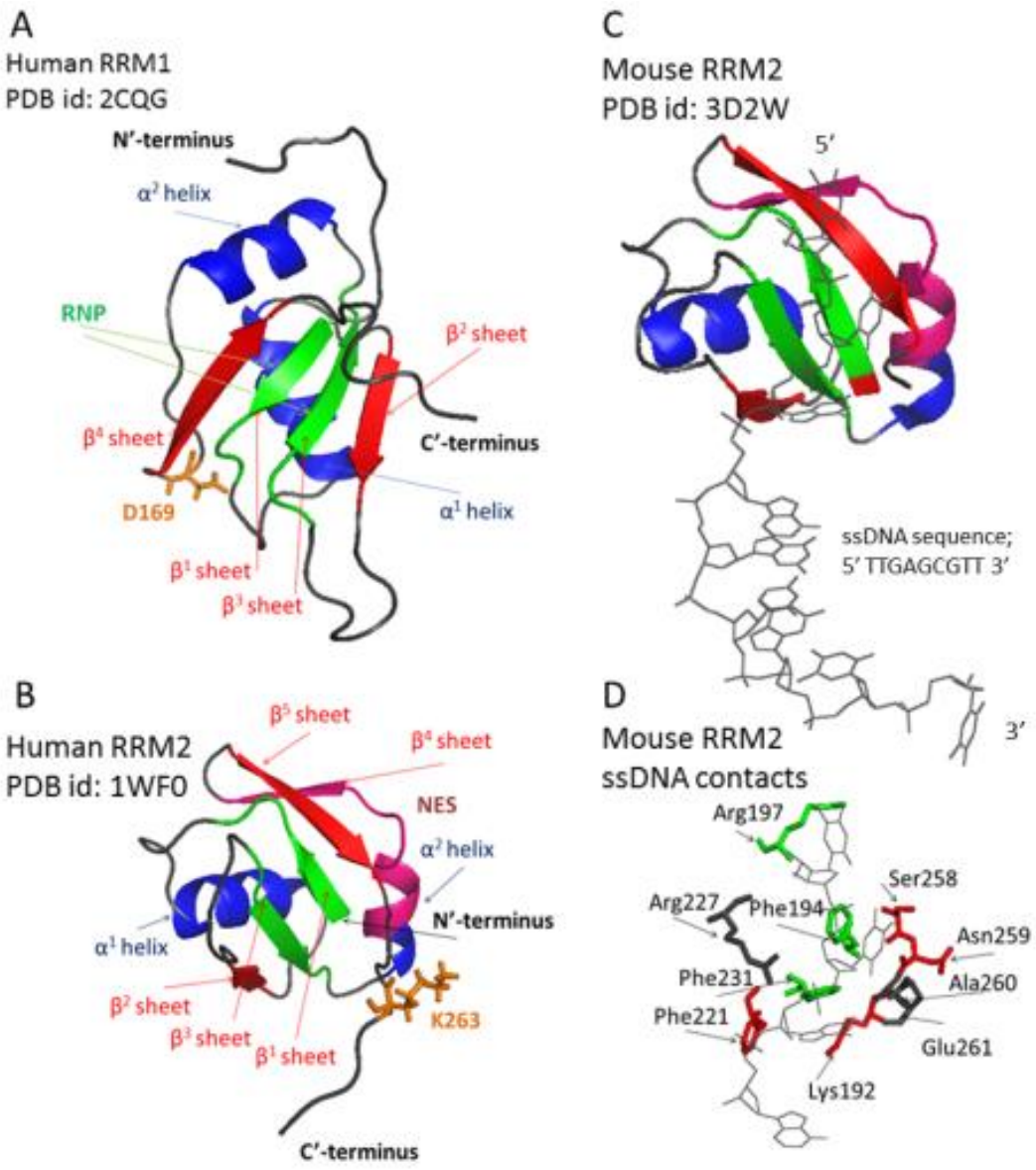


Figure 2.1 Deposited structures of TDP-43 RRM domains. The structural annotation of TDP-43 PDB deposited structures , where α-helices are shown in blue, β-sheets in red, loci of disease mutations as orange sticks, conserved RNP regions as green and NES in magenta**.** **A**, The human RRM1 NMR structure of TDP-43. **B,** The human NMR structure of RRM2. **C,** The ssDNA bound RRM2 crystal structure of mouse RRM2. **D,** The ssDNA contacts made in the crystal model with the majority of contacts involving the anti-parallel β-barrel core.

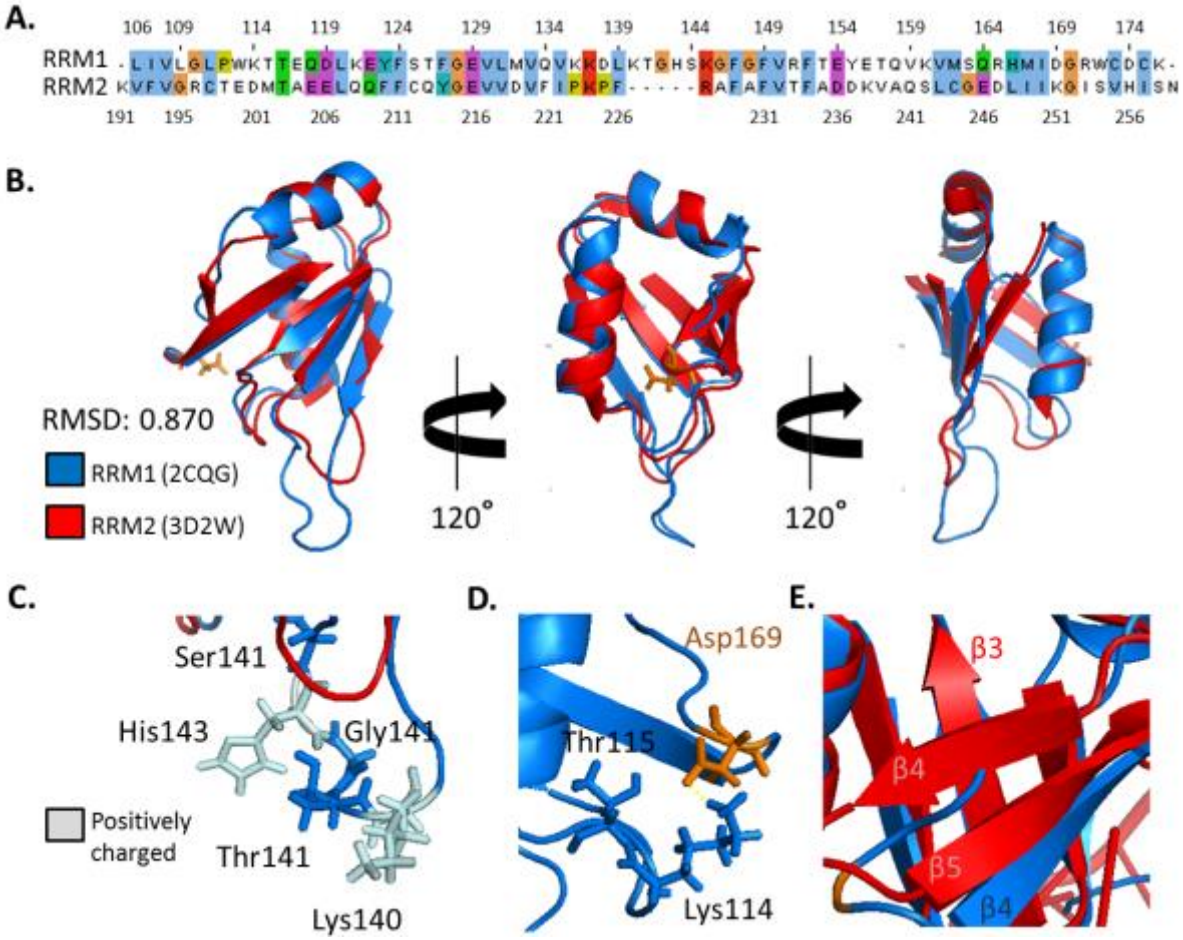


Figure 2.2 Structural comparison between human RRM1 and mouse RRM2. A, Sequence homology between human RRM1 and mouse RRM2 performed by ClustalOmega [84]. The alphiphatic index was calculated as 72.54 and 87.35 for RRM1 and RRM2 respectively. **B,** The PDB structures of these regions from PDB deposition 2CQG and 3D2W aligned with an RMSD of 0.9 Å. **C,** Detailed view of unaligned residues in the extended loop region of RRM1. **D,** Possible polar contacts forming between the disease loci Asp169 (orange), Lys114 and Thr 115. E, The extra β-sheet of RRM2 not conserved in the RRM1 structure.

2.2 RRM homology

The RRM containing proteins are the most abundant in humans, expected to express in ~2% of all gene products [76]. RRM structures are conserved in eukaryotes, though sequences can vary, as illustrated using human proteins containing two tandem RRM domains with known structure (Figure 2.3). The

majority of conserved residues are situated within the RNP motifs that orchestrate nucleic acid binding in all RRM species, including TDP-43. Interestingly other than the RNP regions, the residues in the crystal structure of hnRNP A1 (PDB id: 1HA1) which form inter-RRM interactions are also predominately conserved in TDP-43's tandem RRM region (residues 101-265) (Figure 2.3, red circles and Figure 2.4).

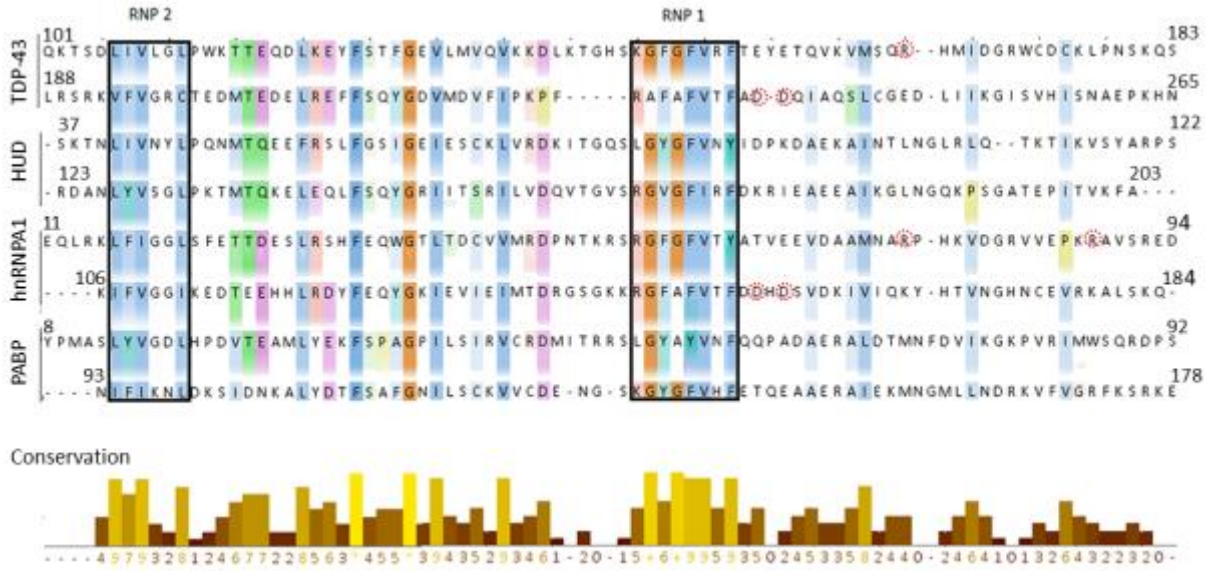


Figure 2.3 Homology between human proteins containing multiple RRM domains. The sequence alignments performed by ClustalOmega [84] between human proteins containing multiple RRM domains. The majority of conservation is between the RNP regions responsible for orchestrating nucleic acid binding. Residues in dashed red circles were shown to form inter-RRM contacts in hnRNP A1 crystal structure (1HA1) [85] the majority of which are conserved in TDP-43S. The proteins selected have following abbreviations: TAR DNA Binding protein (TDP-43, PDB id: 2CQG, 1WF0), ELAV-like protein 4 (HUD, PDB id: 1FXL), heterogeneous ribo nuclear protein A1 (hnRNP A1, PDB id: 1HA1) and Poly(A)-binding protein (PABP, PDB id: 1CVJ). Note that sequences do not vary significantly between RNP motifs as shown by the conservation bar in yellow. (Colours indicate ClustalX scheme, Appendix 1)

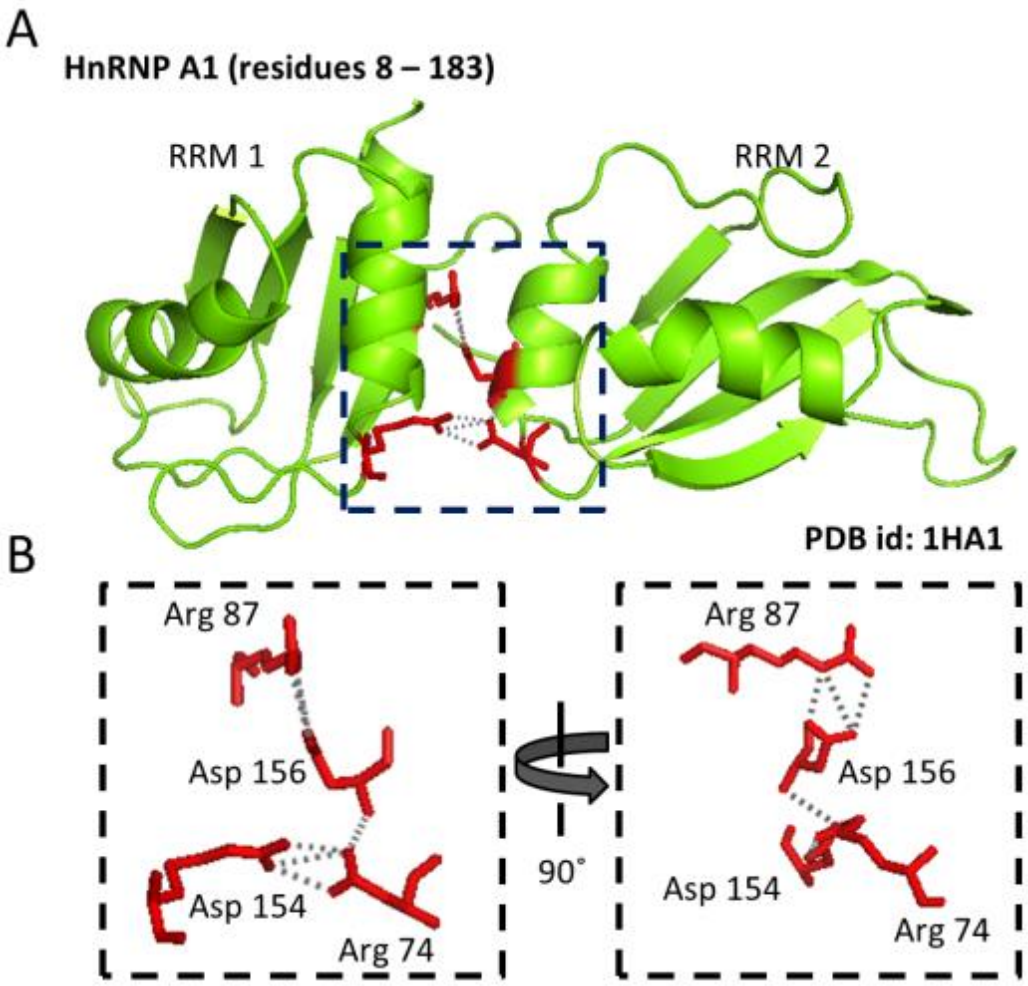


Figure 2.4 Salt bridging between the RRM domains of HnRNP A1. A, The crystal structure of the tandem RRM regions of HnRNP A1 (UP1). **B**, The residues involved in the six salt-bridges stabilising the RRM domains of UP1, which is similar in sequence to residues 101-265 in TDP-43 (see Figure 2.3).

RRM containing proteins usually comprise multiple RRM domains to increase specificity, affinity and selectivity that are important during function [76], as shown by ligand bound and unbound structures. For example, the tandem RRM domains of sex-lethal and nucleolin proteins tumble independently when not bound [86,87], whereas the RRMs of hnRNP A1 may interact to restrict movement and create a larger platform as shown in its crystal structure [85] (PDB id: 1HA1). When bound, the RRM-RRM regions of sex-lethal form a cleft in which RNA binds, whereas ligand interaction in nucleolin also involves the linker [76]. The residue variations in-between RNP regions therefore play an important function in how the RRM domains position for selectivity and binding.

The closest protein match to the TDP-43 RRM-RRM region is from hnRNP A1, known as UP1, which also contains two tandem RRM domains (PDB id: 1HA1). The identity between residues 101-265 of TDP-43 to UP1 is 36% and shares 54% similarity. Given that UP1 has a high resolution crystal structure, it is plausible that the RRM-RRM tandem domains of TDP-43 could also be resolved by X-ray crystallography.

2.3 TDP-43S and its conservation between human and mouse

Residues 101-265 containing the two tandem RRM domains is known as TDP-43S [77]. The TDP-43S structure and sequence are homologous between human, mouse and drosophila. The most similar to human is mouse, which is 93% identical and 98% similar in sequence (Figure 2.5). In this comparison, RRM1 differs by one amino acid from Glu to Asp at position 122, therefore the structure of RRM1 is likely to be virtually identical between species. Changes between human and mouse RRM2 are minimal: the structure is conserved, as shown with the alignment between the NMR human and X-ray crystallographic RRM2 structures, which aligned with an RMSD of 0.66 Å (Chapter 5, Figure 5.5). The binding affinities are also comparable between species, binding $(UG)_6$ with a K_d (nM) of 8.0 ± 0.7 and 14.2 ± 1.4, respectively for human and mouse, confirming that they are also functionally conserved [77,81].

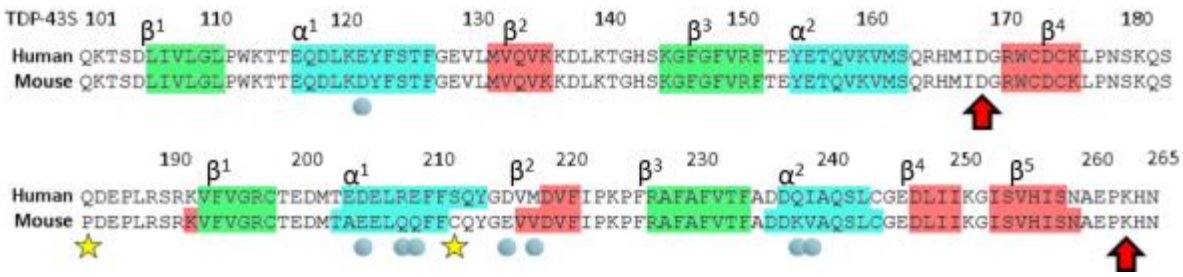


Figure 2.5 TDP-43S sequence alignment between human and mouse. ClustalOmega [84] alignment of human and mouse TDP-43S sequences. Alignment shows the sequnces are 93% identical, or 98% considering similar amino acid substitutions (blue spheres). Only two dissimilar residues exist (yellow stars), which aren't anticipated to change function as they do not occur within the RNP β-barrel core (green). Disease mutation loci are shown (red arrow) and secondary structure showing helices (blue) and sheets (red).

Study of the TDP-43S region for mouse has shown that the RRMs work in tandem to bind nucleic acid targets [77]. Importantly, the individual RRM domains have different ligand preferences, reflected by opposite trends of affinity to different lengths of single stranded uracil-guanine repeats (UG). RRM2 binds more strongly to $(UG)_3$ with an affinity of 379 ± 123 compared to 4690 ± 629 K_d (nM) in RRM1. However, RRM2 binds longer $(UG)_6$ with lower affinity at 4240 ± 381, which was tighter in RRM1, calculated as 65.2 ± 1.8 K_d (nM). This opposite trend in binding may reflect some specificity required by TDP-43 in selecting ligand targets for transport, splicing or sequestering nucleic acid targets.

TDP-43S can bind $(UG)_6$ tighter than both individual domains with a K_d of 14.2 ± 1.4 (nM) suggesting that both RRM domains are required in orchestrating the binding. This synergistic behaviour suggests that the binding stoichiometry with the ligand is likely to be 1:1, which has been postulated using non-radiometric DNA binding assays with TDP-43 using single stranded $(TG)_6$ [88].

2.4 Mouse 'TDP-43S' as a model for disease related disruption

Because mouse TDP-43S protein is homologous and includes a high resolution ssDNA bound crystal structure of RRM2 with data describing affinity [77] it is well suited to model the human version of the protein. An advantage of using the mouse counterpart is that binding affinities and structural annotation regarding RRM1 and RRM2 can be incorporated to provide a comparative foundation during analysis [77]. The two mutant loci are conserved between species to describe sALS (D169G) [89] and FTLD-U (K263E) [90] to enable experimentation of how disease related mutants may contribute structurally and functionally to disease onset in the mouse species. The RRM homology between eukaryotes has established that the domain secondary structure arrangement is vital for function and therefore determining whether these mutations perturb TDP-43's function or cause domain unfolding is an important step in understanding their significance during neurodegenerative diseases.

Therefore given the strong RRM structure to function relationship [76], TDP-43S is a suitable platform to establish whether mutations can cause changes to affinity. The C-terminus only has ~30% of its sequence annotated for function [55], to cover ~40% of mutant loci not containing the most prevalent mutation A382T [54] (Chapter I, Figure 1.3). Therefore, at present the annotation of the C-terminal region is too complicated to clearly establish whether mutations affect TDP-43's function. In contrast the RRM domains perform a single predominant role and therefore changes can be easily assessed by alterations to nucleic acid binding affinity and protein structure. Every function described for TDP-43 is reliant on its interaction with nucleic acids and therefore is a logical construct to test whether onset arises through a loss of function as similarly described in the 'gain of auto-regulation' model (Chapter I, Figure 1.1).

The overall structure of TDP-43S will be assessed using small angle X-ray scattering (SAXS) in combination with existing high resolution atomic models for RRM1 and RRM2, to determine TDP-43S' global shape. The differences in shape deduced from SAXS may then be compared to mutated proteins in order to explain whether disease mutations cause global structural changes e.g. RRM unfolding or inter domain interaction. Alongside this, the high resolution crystal structure of TDP-43S was sought to gain atomic details of the mutations, as similarly performed with UP1 (PDB id: 1HA1).

Protein structural stability tested by a resistance to thermal denaturation has been shown to correlate with increases to half-life *in vivo* [91–93]. This has been shown with mutations of inactive barnase (pseudo-wt barnase) isolated from *E.coli.* where an increase in thermal stability of 2.4 °C resulted in 14% more protein in the periplasm, providing a correlation between increased stability with reduced turnover [93]. Disease locations D169 and K263 exist on flexible regions within TDP-43S and may move to orchestrate some additional stability by enhancing interactions. This will be assessed as previously performed with bacterial pseudo-barnase protein [93] by comparing the mutant construct resistances to thermal denaturation. It is hoped that we can deduce whether RRM related disease mutations cause an increase to *in vivo* longevity as described for C-terminal

mutations. This would associate *TARDBP* disease related mutations occurring in two functionally separate domains (RRM and C-terminal regions) with a common mechanism of disease onset, independent from nucleic acid binding function.

Chapter III

Cloning, expression and the purification of mouse TDP-43S

3.1 TDP-43S cloning rationale

A prerequisite for the examination of the TDP-43S construct utilising the bio-physical techniques available was the development of an expression system capable of producing a high yield of >95% pure protein *in vitro*. The homogeneity is specifically important for small angle X-ray solution scattering (SAXS), circular dichroism (CD), differential scanning fluorimetry (DSF) and aromatic residue fluorescent tracking, since contaminants cannot be differentiated during data analysis.

To this end, *E.coli* was used to over-express tagged mouse TDP-43S protein from recombinant plasmid DNA. The benefits of this system meant a high yield of protein could be produced in large volumes using Luria-Bertani broth in a relatively short time frame. *E.coli* has been shown to express sufficient quantities of N-terminally histidine tagged mouse TDP-43S [77] using the M15 strain. However, this study opted for the basic DE3 strain (Appendix 2) which lacks the lac repressor plasmid 'pREP4', responsible for avoiding protein leaking in the absence of galactose or Isopropyl β-D-1-thiogalactopyranoside (IPTG). This change did not significantly affect the expression levels of TDP-43S protein providing a yield of ~8 mg/L of culture.

3.2 GST solubility tag

A Glutathione-S-Transferase (GST) tag was incorporated onto the N-terminal of TDP-43S. GST has a MW of 25 kDa and readily forms a stable 50 kDa homo-dimer. It is highly soluble and is commonly utilised to increase the likelihood of increasing

the solubility of the fusion protein [94,95]. This was exploited to obtain good yields of soluble GST-TDP-43S during early expression, similar to other GST tagged TDP-43 constructs [23,49]. In addition to this, GST can be purified by affinity chromatography to maximise the yields of TDP-43S and reduce the amount of purification steps. Exploiting the GST interaction to glutathione, commercial columns are readily available (GE healthcare) to facilitate the fusion proteins efficient purification by affinity chromatography. The incorporation of a GST tag was also used to maximise the likelihood of obtaining GST-TDP-43S crystals through increasing the number of crystal contacts [96,97]. As shown in the crystallisation of GST-fused 'DNA replication-related element binding factor' (DREF) a DNA binding protein that failed to crystallise prior to this tagging [98].

The pGEX3X expression vector encodes a tetrapeptide factor Xa protease cleavage site with the sequence; Ile-Glu-Gly-Arg situated between GST and TDP-43S to facilitate tag removal. Due to the plasmid vector's design, three non-native amino acids, Gly, Ile and Leu, are retained on the TDP-43S N-terminal end after cleavage. However, they did not appear to affect the construct's DNA binding affinity when compared against previously published K_d values for TDP-43S [77] or its oligomeric state (discussed in Chapter VII).

3.3 TDP-43S protein expression

The pGEX3X TDP-43S plasmid, (mTDP-43SpGEX3X, Appendix 3) contained a *tac* promoter [99] activated to signal protein over-expression in the presence of IPTG. The amount of IPTG, temperature and incubation time needed was determined by expression trials set within the recommendations described in the manufacturers protocol (GE Healthcare) (data not shown).

GST-TDP-43S over-expression from *E.coli* was most efficient using 5 mM IPTG and a 5 hour incubation at 30˚C. By minimising the amount of purification steps, this procedure produced high yields of >95% pure GST-TDP-43S protein. The purity could still be maintained even when cleaving the GST-TDP-43S protein whilst bound on the glutathione column, although some yield was compromised by the

efficiency of factor Xa cleavage, meaning some fusion protein remained bound. The purity could be improved to ~99% by passing down a gel filtration chromatography column (Superdex 200 16/60) (Figure 3.1). The yield of TDP-43S was estimated as 8 mg per litre of culture.

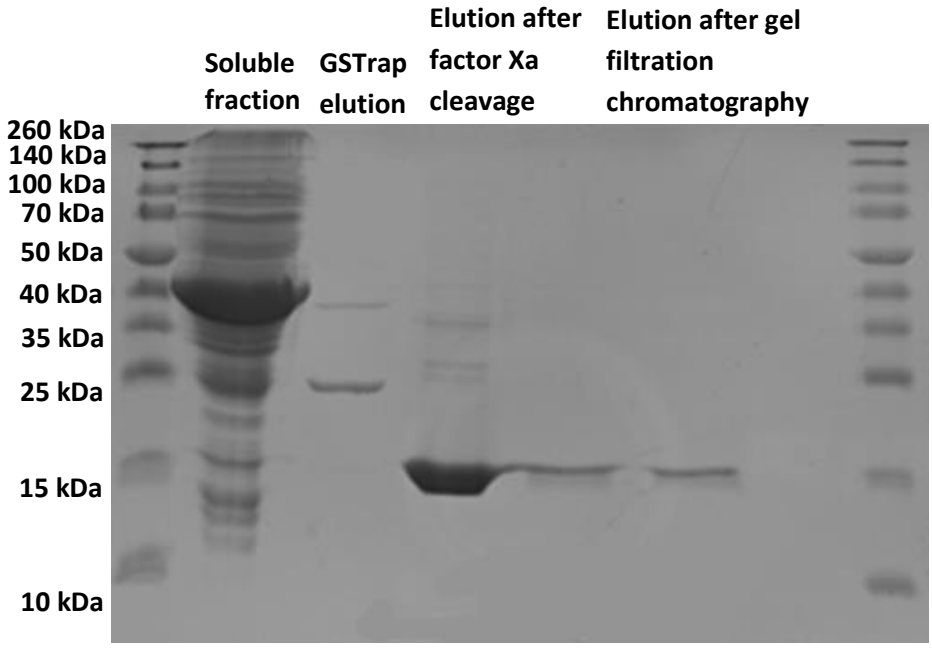


Figure 3.1 Purification of GST-TDP-43S and TDP-43S. The purification of cleaved TDP-43S was achieved in three purification steps. The soluble fraction shows over-expression at ~40 kDa representing GST-TDP-43S. After GSTrap elution, purity of TDP-43S is ~95%. The remaining contaminants can be removed by gel filtration to produce 99% pure TDP-43S protein.

3.4 TDP-43S identification by mass spectrometry

The GST-TDP-43S mass was calculated at 45 kDa by sequence analysis (EXPASY: protparam), but ran as a 40 kDa protein on a 12% acrylamide SDS-PAGE (Figure 3.1). To rule out truncation the purified samples were analysed by electrospray ionisation mass spectrometry (ESI-MS) to empirically determine its mass. Using this technique, the correct GST-TDP-43S protein size was verified as 45 kDa (Figure 3.2) but factor Xa cleaved TDP-43S could not be detected, presumably caused through problems regarding either protein protonation or aggregation.

TDP-43S has a theoretical weight of 19.1 kDa but also ran slower on a gel at ~15 kDa. Therefore to identify and account for more of its mass its gel band was cut and amended to a trypsin digest and matrix assisted laser desorption/ionisation time of flight (MALDI-TOF) mass spectrometry (MS) (Figure 3.2). The resulting MS fragment fingerprint could account for 17 kDa and correctly matched several theoretical TDP-43S fragment masses to correctly identify the protein (EXPASY:PeptideMass).

Since GST-TDP-43S was expressed, bacterial cleavage was not thought to occur. The tendency of GST-TDP-43S to behave as a smaller protein on an SDS-PAGE was also assumed for GST cleaved TDP-43S so it is likely to be the full sized protein.

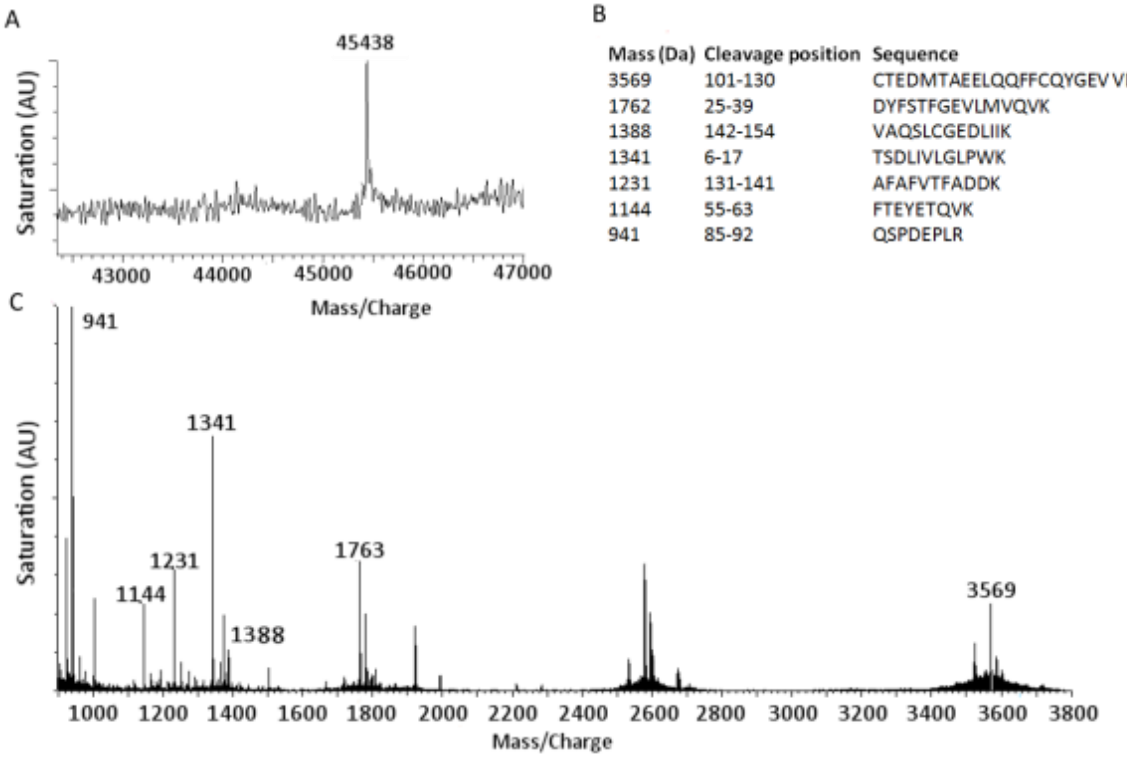


Figure 3.2 MS-MS analysis of full length GST-TDP-43S and trypsin digest of TDP-43S. The mass spectrometry results of expressed TDP-43S constructs. **A**, Full length GST-TDP-43S 45 kDa mass (EXPASY:protparam) was confirmed by ESI-MS/MS. **B,** The expected trypsin fragments of TDP-43S from the sequence (EXPASY: PeptideMass). **C**, The MALDI-TOF MS of trypsin digested TDP-43S fragments. These account for 17 kDa of TDP-43S, which verifies that the correct protein was being expressed.

3.5 Retention and removal of RNA/DNA bound to TDP-43S

During the expression of TDP-43S, the protein bound bacterial nucleic acid which were retained throughout purification. This was revealed through monitoring the ratio between the 260 and 280nm wavelengths. This ratio is commonly used for tracing contaminant nucleic acids in a protein sample [100]. A 260:280 nm ratio between 0.6-0.8 has been previously shown to be sufficient in removing nucleic acid contaminants in a control protein; Enolase and in established biochemical methodology [100,101]. The addition of 500 mM NaCl was sufficient in removing bound DNA/RNA (Table 3.1) and was included as an additional wash step whilst the protein was bound to the GSTrap column.

Table 3.1 Removing TDP-43S bound nucleic acids by increasing salt concentration

	NaCl (5mM)			**NaCl (500mM)**		
Repeats	260nm	280nm	260/280	260	280	260/280
1	0.360	0.360	1.00	0.214	0.296	0.72
2	0.087	0.075	1.16	0.079	0.103	0.76
3	0.225	0.274	0.84	0.189	0.276	0.68

3.6 The oligomeric state of TDP-43S determined by gel filtration chromatography

The TDP-43S protein has been shown to elute as either a dimer [77] or monomer [102]. During this study, mouse TDP-43S was found to be monomeric as compared against a calibrated gel filtration profile using myoglobin (17 kDa), superoxide dismutase-1 (SOD1) dimer (32 kDa) and bovine serum albumin (BSA) (67 kDa) even

in the absence of reducing agents. TDP-43S eluted between the SOD1 dimer (32 kDa) and myoglobin (17 kDa) which coincides well with the expected 19.1 kDa mass of TDP-43S (Figure 3.3). This supports previous findings that show that the N-terminal is important in orchestrating the contacts necessary for dimerization [60,102].

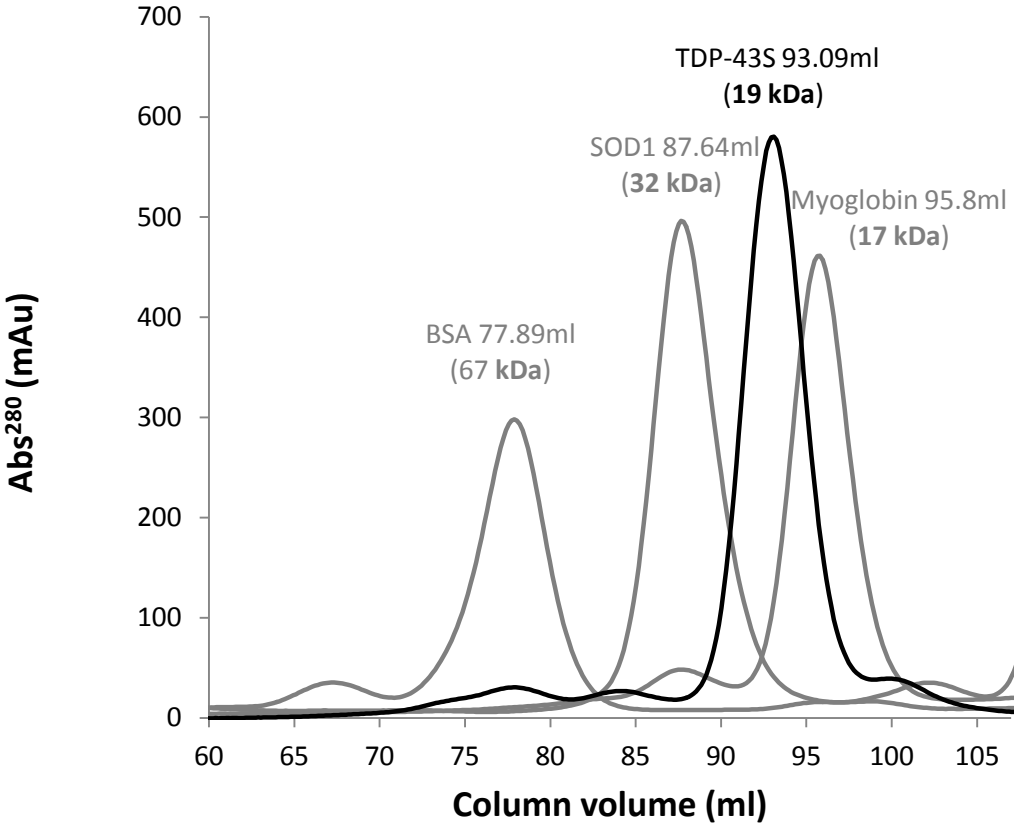


Figure 3.3. Gel filtration chromatography purification and oligomeric state of TDP-43S. TDP-43S Superdex 200 16/60 elution profile calibrated against protein standards; myoglobin (17 kDa), superoxide dismutase-1 (SOD1 dimer) (32 kDa) and bovine serum albumin (67 kDa). The TDP-43S protein eluted predominately as one peak between Myoglobin and SOD1, which fits well with a 19 kDa TDP-43S monomer prediction

3.7 Expression and the gel-filtration of mutant TDP-43S

Disease associated mutations were generated from the same mTDP-43SpGEX3X plasmid template by mutagenesis. As such, the same protocol was repeated for both D169G and K263E disease mutants. Mutant plasmids (Appendix 3) expressed and purified no different as described for wtTDP-43S and the gel filtration profiles were identical (Figure 3.4).

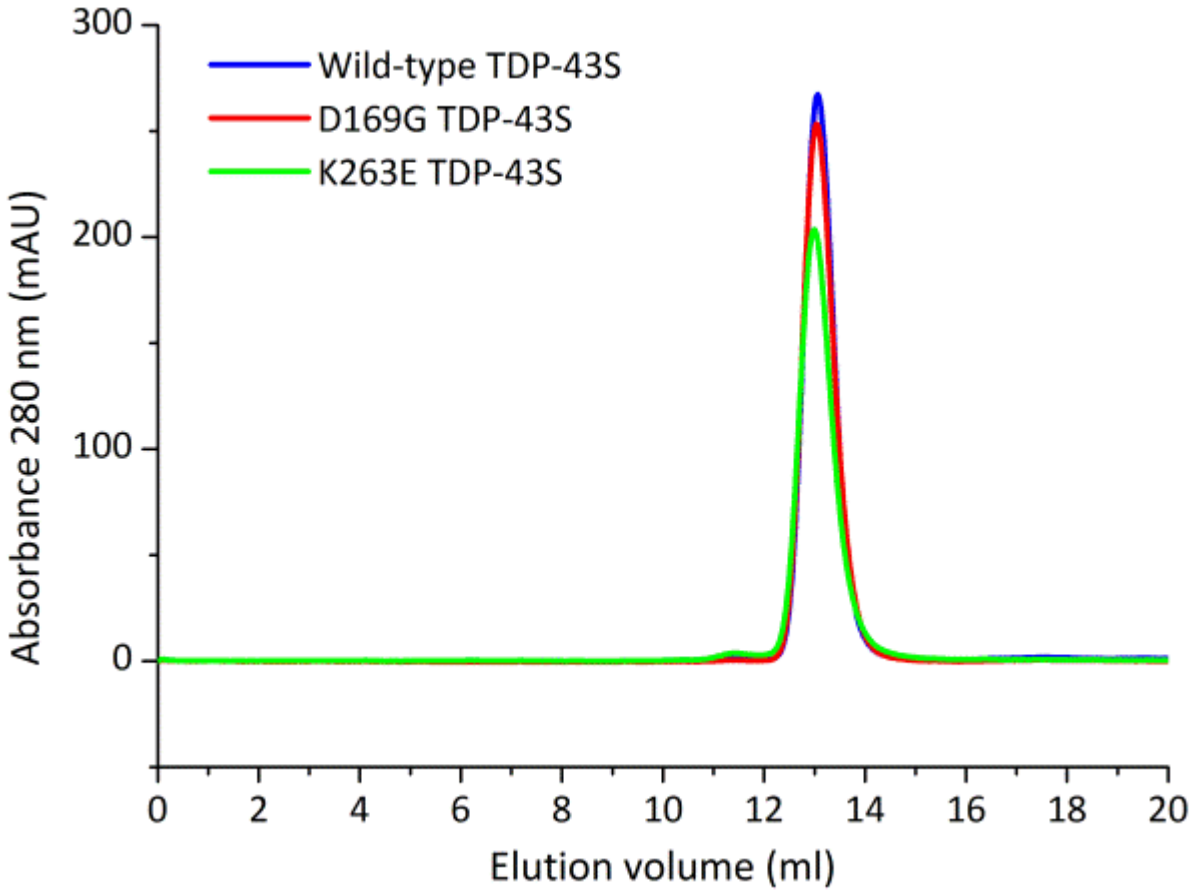


Figure 3.4 Mutant and wild-type gel filtration profiles. Wild-type and mutant constructs eluted at the same volume down a Superose 12 300/10 column, indicating that their mass and oligomeric state are similar, the mutants do not enhance TDP-43S oligomerisation.

3.8 TDP-43S crystallisation trials

To generate crystals for X-ray crystallography wild-type and mutant TDP-43S were screened using commercial crystallisation conditions. The constructs were setup using vapour diffusion trays either using hanging or sitting drop methodology depending on access to the high throughput robot (Figure 3.5). Protein samples and the reservoir solution used to constitute the sample drops were mixed in either a 1:1 or to increase protein concentration a 2:1 protein reservoir ratio. In each case the protein concentration was within a final 5-10 mg/ml concentration, suitable for super saturation of a small protein (~19kDa) [103]. The crystallisation of TDP-43S was anticipated based on the crystal structure of the tandem RRM-RRM domains of hnRNP A1 [85] and the TDP-43s' RRM2 [77].

A significant effort was put into the crystallisation of TDP-43S as a high resolution structure of the mutations and RRM-RRM arangement could determine whether mutations disrupt the structure. Despite testing over a 1000 different conditions, using both GST-tagged and untagged TDP-43 variants no protein crystals were observed. This effort is summarised in the table below (Table 3.2)

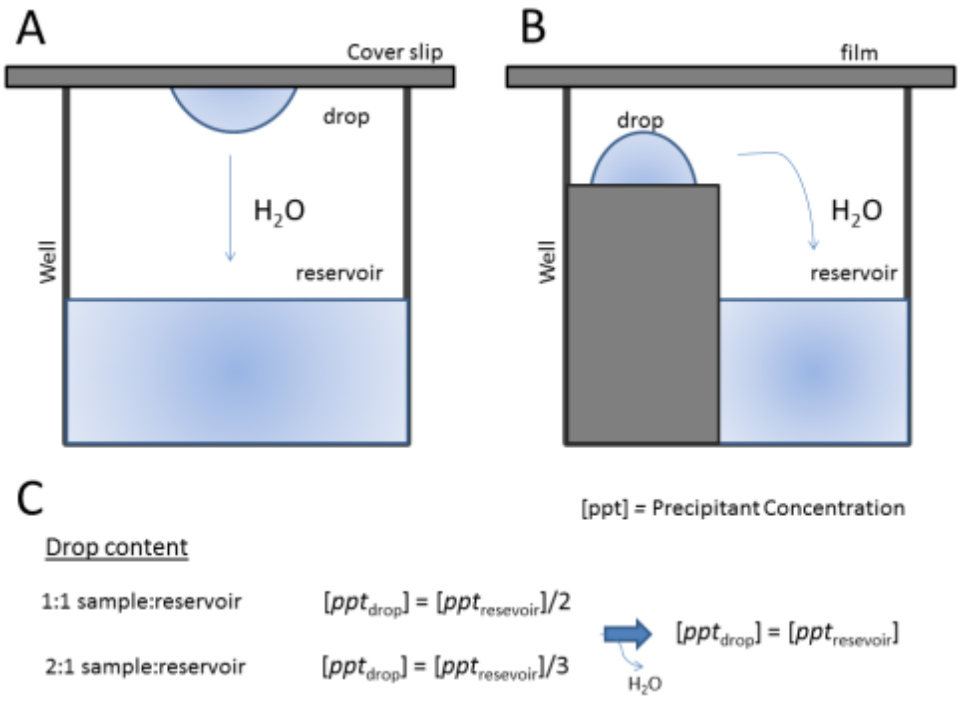


Figure 3.5 Crystallisation trial setups. Protein supersaturation is a prerequisite for crystallisation and is typically achieved by removing water and increasing precipitating agent (ppt) concentration. This is commonly achieved by via vapour diffusion methodology. The drop can be placed on **A,** the roof of the well, used in trials setup manually or **B,** on a shelf typical for automated trials setup by robot. **C,** Both setups remove water by diffusion from the higher water content of the drop, containing the protein of interest, to the lower content of the reservoir in a closed environment until ppt concentration is equal.

Table 3.2 TDP-43S crystallisation conditions

Crystal screen	Nº of conditions	Crystals	Diffraction	Temp (°C)	TDP-43S variants	WT GST-TDP-43S	Concentration (mg/ml)	Vapour diffusion
Morpheus (Molecular Dimensions)	96	3	Salt	4/RT	WT, KE, DG	N	5-10	Sitting/Hanging
Natrix (Hampton Research)	96	4	Salt	4/RT	WT, KE, DG	N	5-10	Sitting/Hanging
JCSG Plus (Molecular Dimensions)	96	7	Salt	4/RT	WT, KE, DG	Y	5-10	Sitting/Hanging
PACT premier (Qiagen)	96	6	Salt	4/RT	WT, KE, DG	Y	5-10	Sitting/Hanging
PEG/ION HT (Hampton Research)	96	3	Salt	4/RT	WT, KE, DG	N	5-10	Sitting/Hanging
HnRNP A1 (Custom)	48	0	N/A	4	WT	N	5-10	Sitting/Hanging
RRM2 *mouse* TDP-43 (Custom)	48	0	N/A	4	WT	N	5-10	Sitting/Hanging
Index HT (Hampton Research)	96	0	N/A	4	WT	N	5-10	Sitting/Hanging
Salt Rx 1 & 2 (Hampton Research)	96	0	N/A	4	WT	N	5-10	Sitting/Hanging
Wizard 1 & 2 (Emerald Biosystems)	96	0	N/A	4	WT	N	5-10	Sitting/Hanging
Classic Suite (Qiagen)	96	0	N/A	4	WT	N	5-10	Sitting
MPD Suite (Qiagen)	96	0	N/A	4/RT	WT	Y	5-10	Sitting
$AmSO_4$ Suite (Qiagen)	96	0	N/A	4/RT	WT	N	5-10	Sitting
pH Clear 2 Suite (Qiagen)	96	0	N/A	4/RT	WT	N	5-10	Sitting
PEGs Suite (Qiagen)	96	0	N/A	4/RT	WT	N	5-10	Sitting
MbClass Suite (Qiagen)	96	0	N/A	4/RT	WT	N	5-10	Sitting

Chapter IV

Techniques used to characterise the structure of TDP-43S

4.1 Small angle X-ray solution scattering (SAXS)

Small angle X-ray scattering (SAXS) is a technique often used to deduce the overall shape and size of a macromolecule in solution. The RRM domains of TDP-43 have well defined structures that may interact with each other in order to orchestrate the protein's interaction with nucleic acids, as observed in the hnRNPA1 crystal structure [85]. To determine whether this interaction is observed in solution for TDP-43S and whether mutations influence inter RRM interactions the solution based SAXS techniques were used. The SAXS analysis suite ATSAS [104] contains several programs developed to generate the average 3D shape, linker flexibility and RRM-RRM interactions of TDP-43S in solution. Through establishing a 3D structural model of wild-type TDP-43S, models of disease mutants can be built and compared to determine global structural differences.

SAXS is a low resolution technique typically capable of a resolution of up to 10-20 Å, limited by the rotational averaging of the protein's shape and structural parameters in solution [105]. It is one of two versatile techniques developed since the discovery of X-rays, the other being X-ray crystallography. Both techniques exploit the plane wave properties of X-rays which can be described by its amplitude and wavelength and its interaction with electrons. It is from these scattered secondary wavelets that constructively or destructively interfere with each other and their angles that are exploited during data interpretation.

The scattering from macromolecules at low angles in solution generates an isotropic profile on the detector with a characteristic intensity decreasing from low to high angles. From the 1D scattering profile one cannot discriminate between scatterers of different size and shape, therefore it is important that the sample is monodispersed and free from oligomeric or aggregated states. The smallest angle

which can be measured in SAXS defines the longest distance of the macromolecule, thus the region close to the beam-stop is most sensitive to aggregates and unwanted scatterers.

To this end the SAXS beam-line at the Soleil synchrotron, St Aubin, France, named SWING, has a coupled high performance liquid chromatography (HPLC) column to increase the likelihood that the desired proteins are monodisperse immediately prior to X-ray exposure [106]. The protein's elution through the column can be calibrated so that SAXS data can be recorded once the elution volume containing the protein of interest is known (Figure. 4.1).

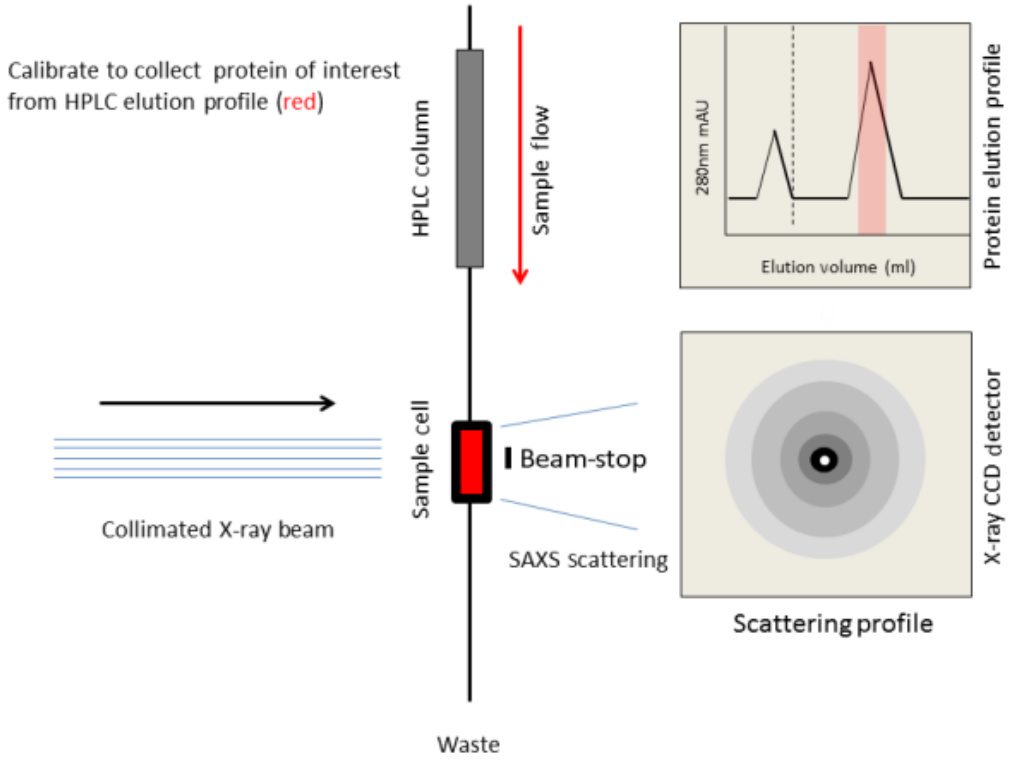


Figure 4.1 Schematic of HPLC coupled SAXS data collection. Data collection setup at Soleil, St Aubin, France on the SWING SAXS beam-line. A small sample volume ~40ul is automatically injected whilst in the experimental hutch. The sample is passed down a HPLC size exclusion column which ensures the sample is homogenous (red in the elution profile) in the absence of aggregates depicted in the void volume (left of the dashed line). The X-ray is calibrated to measure the elution of the sample as it runs through the X-ray permeable cell, since the sample is consistently moving X-ray damage is also minimised compared to a stationary sample. The centre of the isotropic profile has no scattering due to the beam stop, present to avoid permanent damage to the detector. Some information within this region can be calculated by extrapolating the data from the remaining low angle signal.

4.2 SAXS theory

The scattering pattern of a macromolecule in solution can be described by the intensity (I) as a function of the scattering angle θ. Generally the scattering angle alone is not disclosed, instead it is usually quoted as the length of the scattering vector $\vec{q}$ related to θ, known as momentum transfer *(q)* measured in $Å^{-1}$ (Equation 4.1);

$$q = \frac{4\pi sin\theta}{\lambda} \qquad \textbf{(Equation 4.1)}$$

Where λ denotes the wavelength and θ is half the angle between the incident and scattered radiation. Non-crystalline particles scatter as described by Debye *et. al* 1915 [107]. The contribution of each scattered particle to the detectable signal can be explained by the following relationship [108];

$$I(q) = \left\langle \int \left| \left(\rho(\vec{r}) - \bar{\rho}_s\right) e^{I\vec{q}\cdot\vec{r}} d\vec{r} \right|^2 \right\rangle \qquad \text{(Equation 4.2)}$$

Where the brackets ⟨ ⟩ denote the rotational averaging of the sample and $\rho(\vec{r}) - \bar{\rho}_s$ is the difference in the scattering densities between the measured scattering particle at the volume element position ($\vec{r}$) and that of the solvent (s)[108]. Essentially the signal used during analysis is the scattering of the protein of interest with the scattering of the solvent subtracted. The integration of scattering data over all directions yields the one-dimensional isotropic I(q) Vs q plot, from which all valuable geometric information of the protein can be calculated such as the radius of gyration.

4.3 Radius of gyration (Rg)

The radius of gyration (Rg) is the first piece of information regarding the shape of the macromolecule defined as the root-mean-squared distance of scattering particles weighted about its centre of mass, usually quoted in Å. Since

macromolecules of the same mass but different shapes yield different values it can be exploited to detect initial differences between mass distributions.

The Rg can be determined by two methods. Either by Guinier approximation (Equation 4. 3) [105] or by the interatomic distance distribution profile (P(r)) obtained by a modified Fourier transform of the I(q) Vs q plot (Equation 4.4).

In the Guinier approximation the Rg is calculated from the linear region at the smallest experimentally available angles in a Guiner plot (In[I(q)] Vs q^2) (Equation 4.3). This profile is only linear over a limited region, typically for globular proteins a qRg <1.3 but is dependent on the structural characteristics of the macromolecule under study. For intrinsically disordered proteins this range can be reduced to qRg <0.8 reflecting the complexity arising from flexibility and different conformations in solution [109,110].

Calculating the Rg from the P(r) profile utilises the entire scattering curve (Equation 4.5) and is not influenced by the restrictions associated with Guinier approximation. However, the errors associated with assigning an accurate maximum intra-atomic distance known as Dmax can be influenced by flexibility within a species. For purposes of accuracy, both methods are used for comparison.

Guinier approximation as determined by the straight line equation at small angle in a In[*I(q)] Vs* q^2 plot;

$$In\, I(q) = In\, I(0) - \frac{Rg^2}{3} \cdot q^2 \qquad \textbf{(Equation 4.3)}$$

y = c - m x

I(0) is the extrapolated intensity at q=0. The straight line components, depicted in red are used to generate the Rg from the gradient;

$$Rg = \sqrt{3\,m} \qquad \textbf{(Equation 4.4)}$$

Where m is the line gradient.

Equations for the Rg as determined automatically from e.g. Gnom [111] using the P(r) profile [110], equation (4.5);

$$Rg = \sqrt{\frac{\int_0^{Dmax} P(r) r^2 dr}{2\int_0^{Dmax} P(r) dr}} \qquad \textbf{(Equation 4.5)}$$

Where P(r) is related to the frequency of distances at a particular inter-atomic distance (r).

4.4 Interatomic distance / P(r) profile

The P(r) profile is the summation of all distances between scattering centres within the protein, depicted by their frequency at each vector length (r) to Dmax (Figure 4.2). This essentially describes the mass distribution of the macromolecule and can be utilised as an input for *ab initio* modelling.

The P(r) profile can indicate flexibility or intrinsically disordered proteins which is suggestive of an elongated tail before an appropriate Dmax can be assigned. This reflects the different sized populations used to generate the scattering data. Examples of this can be observed in the work of Receveur-Brechot and Durand, 2012 [110].

4.5 Kratky plots

Plotted as $I(q)q^2$ Vs q, the Kratky plot can be used to differentiate between folded globular proteins and those that are completely disordered. This plot is a measure of surface interfaces as described by Porods law [112] .

In a globular species, not all residues are accessible to solvent which tend to be masked in the core of the protein. Therefore, the ratio between protein mass and solvent surface accessibility is reduced. This results in a bell curve returning to zero as q increases in the plot [110]. Conversely in a completely unfolded species no bell curve is observed and tends to increase in $I(q)q^2$ as q increases. A flexible protein

containing structure could straddle both profiles, depending on the extent of the protein's flexibility [110]. This plot is able to show folding differences between mutant profiles but lacks the detail necessary to deduce how this change is derived.

An adaptation to this plot, a dimensionless Kratky, removes the size and shape contribution to allow for direct protein comparisons [113]. This plot is performed as above but the y and x axis are divided by the I(0) (the I(q) when $2\theta=0$) and multiplied by the protein's Rg respectively to remove molecular mass and size contributions [110]. This adaptation fulfils two roles. Firstly, it allows the direct comparison of proteins through scaling and secondly, proteins can be compared against known annotation whose trend for a well folded globular protein peaks at 1.104 $I(q)q^2/I(0)$ at $\sqrt{3}$ the qRg. Deviations from this represent varying degrees of flexibility, disorder and unfolding as commonly seen in intrinsically disordered proteins [110].

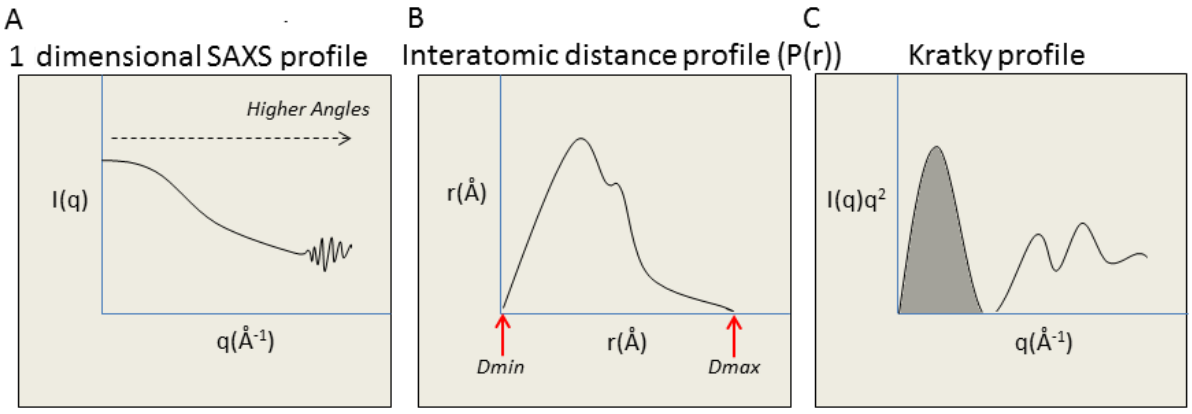


Figure 4.2 Primary SAXS profiles. Examples of **A**, the 1 dimensional SAXS profile (1D profile). **B**, Interatomic distance (P(r)) profile and **C**, the kratky plot, depicting the integral expected in a globular protein (dark grey).

4.6 *Ab Initio* modelling

From a 1D profile it is possible to model the 3D shape of the protein through *ab initio* modelling. Several *ab initio* programs are incorporated within the ATSAS SAXS interpretation software suite, of these, GASBOR [114] was employed within this study.

In this method the model of the macromolecule is constructed from dummy residues resembling Cα of each amino acid within the protein. The algorithms employed by GASBOR aim to compact these residues from a dispersed gas-like volume to agree with the dimensions and mass distributions of the SAXS profiles. The constraints require the model to adhere to the 'chain-compatibility' spatial arrangement meaning Cα are situated 0.38 nm from each neighbour representing the spatial separation of a peptide bond [114]. The dummy residues are scaled by the average amino acid mass in order to represent the correct I(q). Monte Carlo-like algorithms are employed to search for better fits to the 1D scattering data and the goodness of fit is assessed by χ^2minimisation [111]. It is usually appropriate to generate a number of GASBOR repeats for 3D superimposition, then averaging the models to establish the average model. Average models are also fitted to the 1D data to assess whether the model is in agreement with experimental scattering.

4.7 Rigid-body modelling

The high resolution PDB structures describing TDP-43 RRM domains can also be considered when fitting to the 1D SAXS curve. By incorporating known structural features during *ab initio* modelling we derive further evidence for the shape and a more realistic domain arrangement in solution.

A protein can be explained by any number of atomic structures assuming they describe a unique section of the macromolecule. The relationship for rigid-body structures (*k)* and its scattering contribution can be explained with the equation below (Equation 4.6)[115].

$$I(q) = \left\langle \left| \sum_{k=1}^{k} A^{(k)}(q) \right|^2 \right\rangle_{\Omega} \qquad \textbf{(Equation 4.6)}$$

Where $A^{(k)}(q)$ is the scattering amplitude of the k^{th} rigid-body and $\langle\,\rangle_{\Omega}$ is the rotational spherical averaging in reciprocal space. For regions that lack structural annotation such as the linker and terminal regions in TDP-43S an *ab initio* approach can be combined during analysis. This is performed by the ATSAS program BUNCH

to describe the entire I(q) signal including the rigid-bodies and dummy residue traces which considers rotations, distances and orientations of sub-structures [115], described as (Equation 4.7);

$$I(q) = 2\pi^2 \sum_{l=0}^{\infty} \sum_{m=-1}^{1} \left| \sum_{k} A_{lm}^{(k)}(q) + \sum_{i} D_{lm}^{(i)}(q) \right|^2 \qquad \textbf{(Equation 4.7)}$$

Where $A_{lm}^{(k)}$(q) and $D_{lm}^{(i)}$ (q) denote the partial scattering amplitudes of the kth rigid body and dummy residues respectively.

The closeness of fit to the experimental data is calculated again by minimising χ^2 values to indicate improved fits when dummy residues and rigid-bodies are orientated into a more favourable conformation [115]. The software does not allow spatial overlap between structures or infeasible rotation angles. Side-chains attached to Cα-atoms can be reconstituted using the online software SABBAC [116]. It is usually appropriate to test the fit of *ab initio* models by fitting the final reconstituted structures back to the original data by CRYSOL [117] to determine if the model correctly represents the data.

4.8 χ^2 minimisation

Caution is required when assessing χ^2values since this value is also influenced by the scale and quality of the SAXS data. For example χ^2 fits match noisy data easier since the error is increased so values are more likely to stray within its bounds. Obviously large χ^2values indicate poor fits but one should also assess the actual plot to see if it can accurately match the data up to a sensible q(Å^{-1}) value. They can therefore be used to choose the best fit between similar profiles of the same species. $\boldsymbol{\chi^2}$ minimisation is calculated as shown in Equation 4.8 (using BUNCH) [115].

$$\chi^2 = \frac{1}{\mathrm{N}-1} \sum_{j} \left[\frac{\mathrm{Iexp(qj)} - \mathrm{cI(qj)}}{\sigma(\mathrm{qj})} \right]^2 \qquad \textbf{(Equation 4.8)}$$

Where N is the number of experimental points, c is the scaling factor and $\sigma(qj)$ the experimental error at momentum transfer (qj).

4.9 Ensemble optimisation methodology (EOM)

In a flexible species the experimental SAXS curves can be considered to be made from the sum of populations characterised by different conformations. The ensemble optimisation method (EOM) [118] attempts to determine the frequency of these populations using a pool of random theoretical models limited by the sequence length and any user defined PDB depositions describing parts of the protein. This approach can be used to estimate the extent of protein flexibility in solution [118] as shown in other RRM-RRM species [119].

EOM generates 10,000 random models testing the entire spatial configuration space of the protein. The theoretical scattering of these models are compared and changed in proportion according to the best fit to the 1D experimental curve by χ^2 minimisation. The output of EOM yields the ratio of protein sizes displayed as a histogram in regards to both Dmax and Rg needed to fit the observed curve. The models used during this fit are available so conformation or spatial distances can be compared.

4.10 Circular dichroism (CD)

Circular dichroism (CD) is a biophysical technique which exploits absorbance differences between left and right circular polarised light (CPL) typically between the 178 – 260 nm wavelength ranges. Throughout this spectrum CPL is absorbed differently depending on the composition and type of protein secondary structure and can be used to predict the relative proportion within a protein of interest [120]. Its use during this project was to characterise any initial structural differences between native and disease related mutations of TDP-43S which could be used to support the global observations extracted by SAXS.

Differences of protein absorption between left and right circular polarised light (CPL) plotted as a function of wavelength is proportional to the content of

secondary structure, as described in Figure 4.3. This data is typically analysed through structural algorithms that compare the spectrum against an annotated reference set of proteins and assigns based on this the relative secondary structure content.

As part of the dichroweb on-line CD analysis suite [121] there are several algorithms that can analyse CD spectra. However, for reasons of relevance only the CDSSTR algorithm will be discussed, since it was used during this study.

CDSSTR uses singular value decomposition techniques to compare 5 features over the 178-260 nm range between sample and reference spectra [122]. This avoids over interpretation and reduces any anomalous errors introduced through contaminant absorbers, noise or the instrumentation introduced during analysis [122,123]. This technique has been described as one of the most thorough analytic packages for CD analysis and has been found to perform best when using longer wavelength ranges with smaller reference sets, as compared to other commonly used algorithm [124] such as CONTIN [125] and SELCON3 [126].

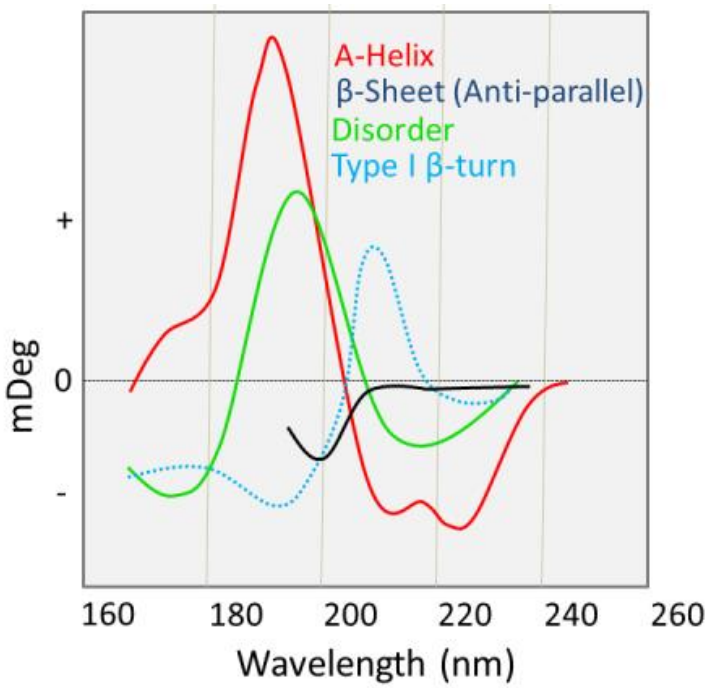


Figure 4.3. Contribution of secondary structure on a CD spectrum. The circular dichroism absorbances of secondary structure. Showing the difference between absorbance of left and right circularised light, shown in millidegrees. The schematic is adapted from *Brahms & Brahms, 1980* [127].

Chapter V

Assessing the effect TDP-43 RRM disease mutations have on global structure by small angle X-ray scattering and circular dichroism

5.1 Exploring the properties of RRM-RRM domains in solution

The TDP-43S construct is known to contain two structured RRM domains separated by a linker of unknown structure. Through comparisons made from other tandem RRM proteins, discussed in chapter II, similar linkers are often flexible, enabling individual RRMs to move independently. This could also be true of TDP-43S given that it also failed to crystallise, which is likely due to flexibility (Chapter IV, Table 3.2). Therefore, to draw meaningful conclusions about a protein's dynamic structural behaviour, describing e.g. protein unfolding, flexibility and inter-domain interactions, one should draw conclusions from solution based methodology. These findings also relate better to TDP-43's natural behaviour *in vivo*, compared to a protein crystal for example, which may restrict movement by crystal packing.

The technique used during this study was small angle X-ray solution scattering (SAXS), discussed in further detail in chapter IV. The main limitation of SAXS is its low resolution, caused by averaging different protein orientations within the sample. However, this is also beneficial, since the data collected provide information regarding the protein's average shape. Therefore, structural changes that infer a change to the maximum length or remove electron density should be reflected in the deduced SAXS models. This could occur through TDP-43S mutations by increased inter-domain interaction or domain unfolding, which would be anticipated to cause a more condensed or expanded structure respectively.

SAXS models are also complementary to high resolution techniques used to superimpose high resolution structures to assess their conservation in solution [128,129]. NMR and crystallographic structures can also be useful during the

generation of SAXS models to account for domain electron density. The two functional RRM domain structures previously solved by NMR and X-ray crystallography (PDB deposits; 2CQG & 3D2W respectively) for TDP-43 and can therefore be incorporated by SAXS methodology to provide constraints when determining the spatial arrangements and flexibility of the tandem RRMs in solution.

5.2 Suitability of SAXS to assess differences between TDP-43S mutants

The growing sophistication of analytical programs available to analyse scattering data means that SAXS can now provide information regarding the average *ab initio* shape, the radius of gyration, the particle's maximum dimension and an estimate of the different conformations within the sample. SAXS has been widely used to investigate structural changes in both globular and flexible protein species [119,130].

SAXS is sensitive enough to differentiate between the spatial differences of two RRM-RRM proteins during ligand binding, despite sharing a similar domain organisation [119]. The ensemble optimisation SAXS methodology combined with surface plasmon resonance (SPR) could determine differences of RRM positioning during ligand binding between TIAR and HuR, two proteins which contain two tandem RRM domains. This demonstrates that data derived through SAXS can be used to successfully analyse similarly organised proteins to TDP-43S.

SAXS with use of partially known structure has also been used to describe the structure of a full length RRM containing protein, as shown for polypyramidine tract binding protein (PTB). Using a combination of scattering curves for different RRM pairings and individual rigid-bodies the results produced just one valid conformation to describe the arrangement of the four RRM domains of the PTB structure which has eluded crystallography and NMR approaches [130].

SAXS is therefore an appropriate technique to explore differences between the global structure of TDP-43S and disease mutants. As such RRM-RRM spatial

arrangements, structural unfolding and flexibility of different mutants were assessed and compared.

5.3 TDP-43S SAXS data collection

SAXS data were collected at the Soleil synchrotron, taking advantage of the on-line HPLC coupling and purification immediately prior to data collection. This ensured that protein samples used for subsequent analysis were homogeneous and aggregate free, as highlighted in Chapter IV. Datasets were collected using 1 second exposures, with 0.5 second dead-time over 255 frames. These frames were calibrated to collect the proteins elution from size exclusion chromatography. The high intensity associated with synchrotron X-ray radiation and collection time per frame meant the radius of gyration (Rg) could be estimated to monitor the Rg and I(0) during the dataset. The sample in solution was in constant motion during the dataset to limit the amount of radiation damage.

The components used to make up the Rg Vs I(0) plot (Figure 5.1) are related to protein size and concentration respectively so that data from the highest protein concentration could be pooled. Limiting the data to the more intense scattering of higher protein concentrations meant that noise was reduced and improved signal could be measured at higher q($Å^{-1}$) ranges. Frames between 220 – 231 corresponding to a steady Rg (~22Å) at the highest I(0) were averaged and used to generate the 1 dimensional plot (1D) (Figure. 5.2).

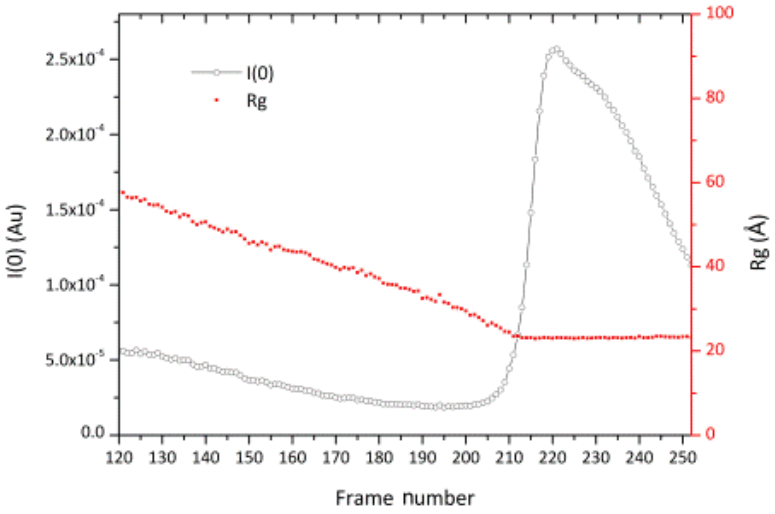


Figure 5.1. The Rg vs I(0) plot of wtTDP-43S HPLC elution. The elution of TDP-43S through the size exclusion column were analysed by SAXS to yield the scattering intensity I(0) and radius of gyration (Rg) used to identify frames 220-231 for pooling.

5.4 Approximation of Dmax and the radius of gyration

The pooled data resulted in a low error in the 1D profile up to a q($Å^{-1}$) of 0.5 (Figure 5.2) with an Rg estimate of 21.8 ± 0.1 Å (data points 1-101), determined by the Guinier approximation (Figure. 5.3). The data within the qRg <1.3 limit compared reasonably well with other RRM-RRM structures, with an Rg of 20 Å calculated for UP1 (PDB id: 1UP1 by CRYSOL [117](data not shown).

The 1D profile was re-plotted to produce the pair-wise distribution function (P(r)) to assess the distribution of mass (Figure 5.4). The maximum interatomic distance (Dmax) in the P(r) was calculated as 85.5 Å the distance frequencies clustered at ~20 r(Å) and at ~40 r(Å) to a lesser extent with a couple of smaller shoulders at higher distances. These two larger distance frequencies likely reflect the masses of the two RRM domains. The profile immediately prior to the Dmax is elongated similar to other flexible multi-domained proteins to suggest that TDP-43 was also flexible [110,131]. The Rg calculated from this Guinier approximation plot was larger at 23.3 ± 0.1 Å probably attributed to having to assign a large Dmax to allow for a reasonable fit of data, shown in Appendix 4. The regularised fit to the data is

based on the 'perceptual criterion' method established by Svergun (1992), implemented and performed automatically in the GNOM program [111], to establish a regularised fit to the experimental data.

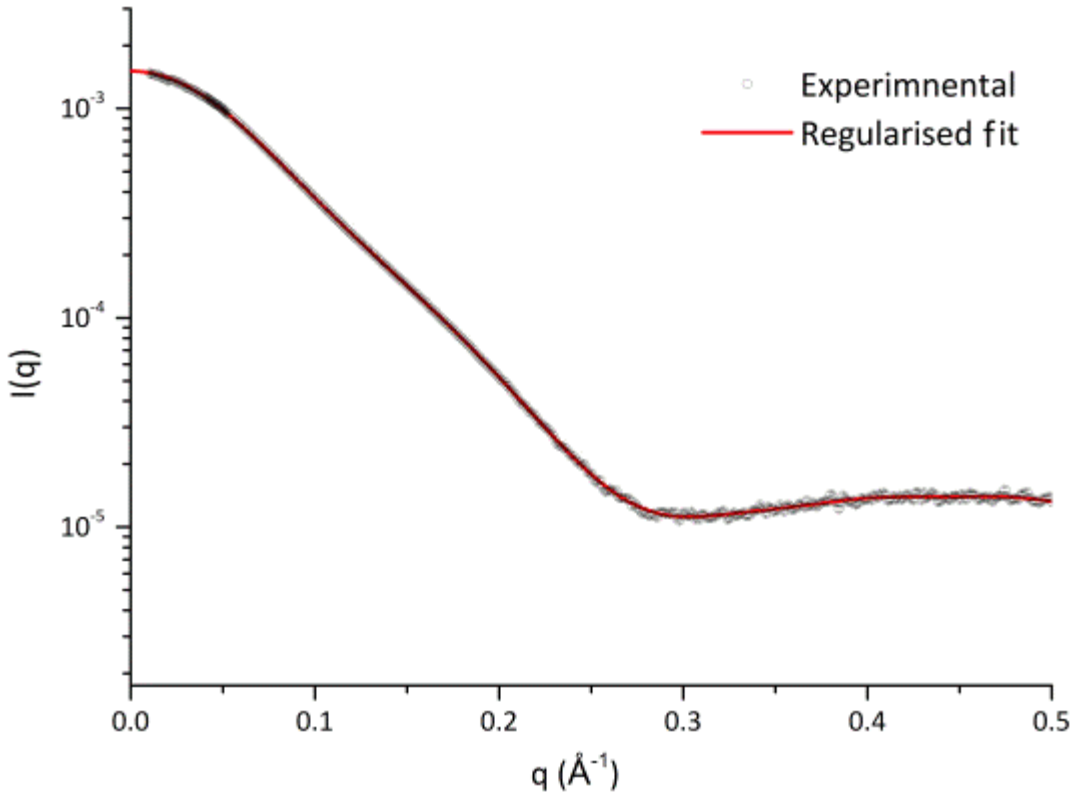


Figure 5.2 1 dimensional profile of wtTDP-43S. The pooled frames 220-231 gave rise to the 1D profile for wtTDP-43S.

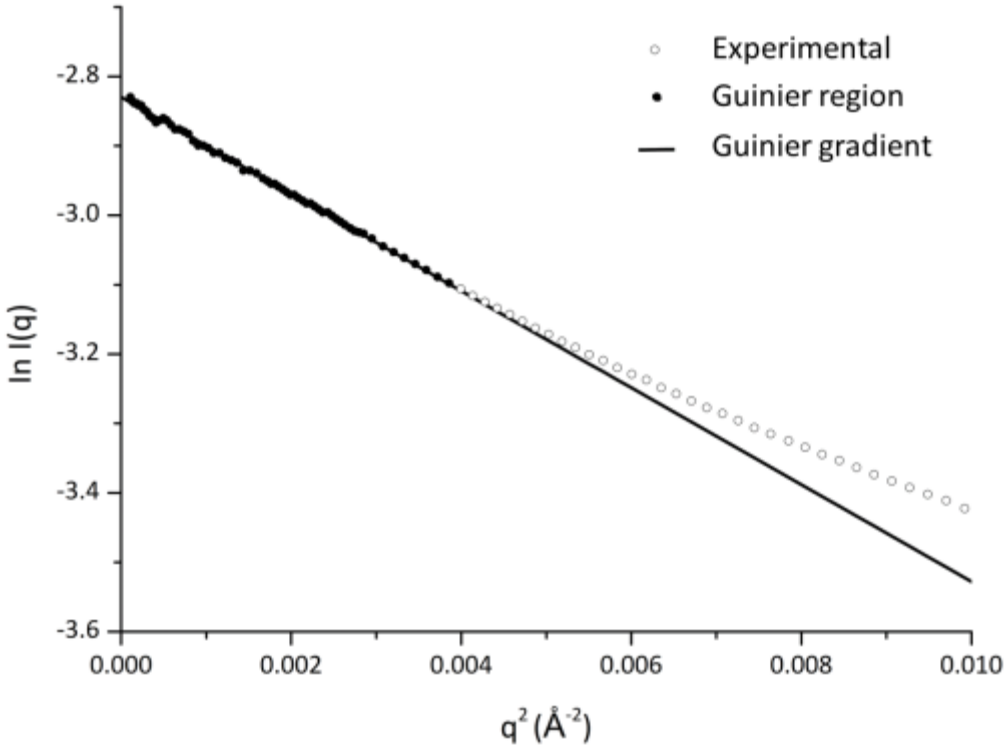


Figure 5.3 Guinier region and approximation of wtTDP-43S. The Guinier approximation of wtTDP-43S estimated the Rg as 21.8 ± 0.1 Å with qRg limits < 1.3.

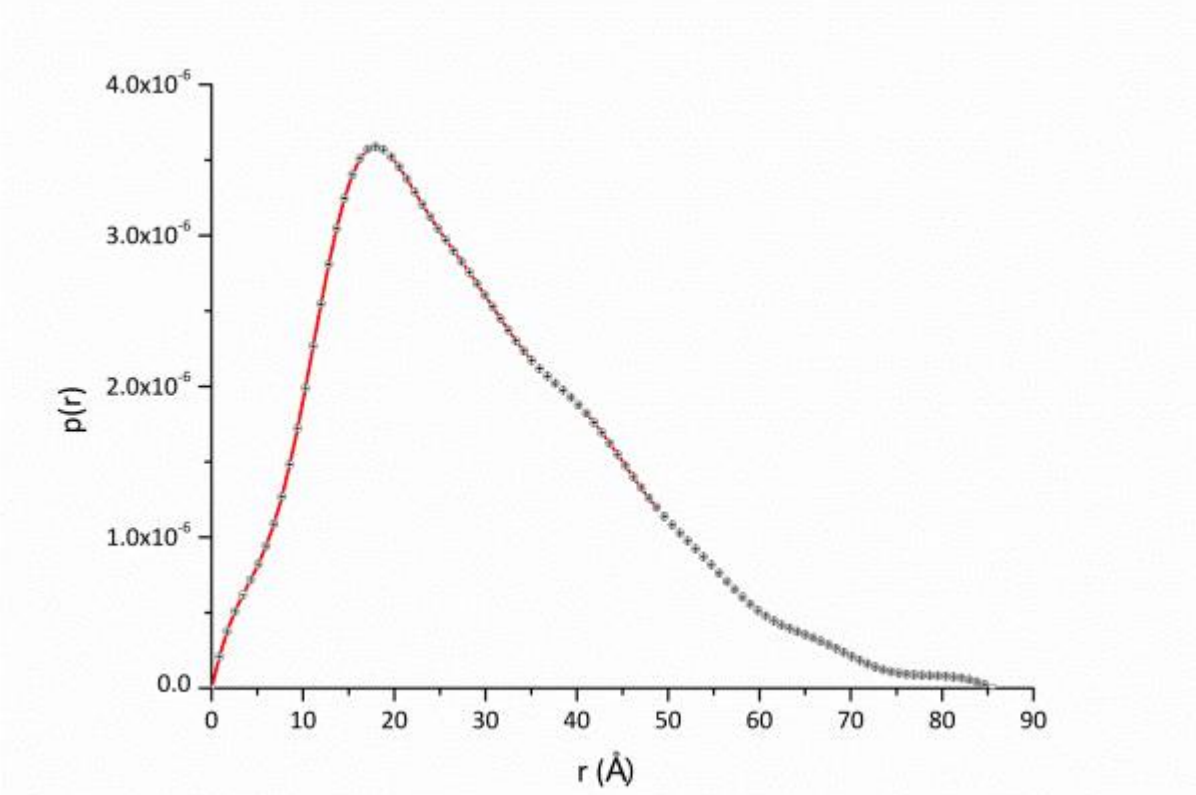


Figure 5.4 Pair-wise distribution profile wtTDP-43S. The P(r) profile of wtTDP-43S estimated a maximum interatomic distance of 85.5 ± 5Å and Rg of 23.3 ± 0.1 Å. The p(r) axis represents the relative frequency of inter-atomic distances and r(Å) the distance.

5.5 Kratky plot and molecular weight estimations.

The 1D curve conversion to the $I(q)q^2$ Vs q ($Å^{-1}$) Kratky plot produced a profile containing characteristics observed for disorder and folded protein [110]. An incomplete bell curve suggestive of globular structure is present between the 0.00 - 0.25 q($Å^{-1}$) range however was incomplete and increased in signal at higher q($Å^{-1}$) angles. This is a trend observed from flexible proteins [132] to imply that TDP-43S contains both regions of globularity and disorder. This is likely to be associated with the flexibility of TDP-43's linker, as shown in the Kratky analysis of other multi-domained flexible proteins [132]. The integral of the initial bell curve (up to 2.5 q($Å^{-1}$)) represents folding with a comparable mass prediction by SAXSMoW [8] of 22.4 ± 2.24 kDa compared to 19.1 kDa calculated from its sequence (Figure 5.6).

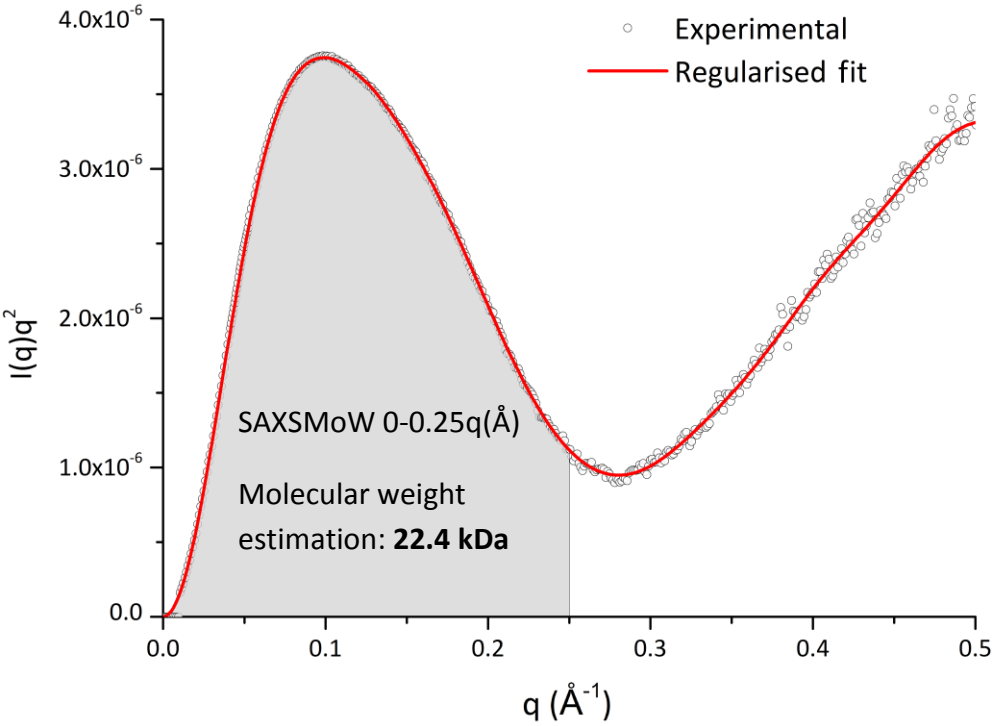


Figure 5.6 Folding of wtTDP-43S by Kratky plot analysis. The Kratky plot of the construct implies that the majority of protein in solution is structured as depicted by a bell curve between 0.0-0.25 q($Å^{-1}$). At higher q($Å^{-1}$) the $I(q)q^2$ increases after q($Å^{-1}$) ~0.28 to suggest the presence of unstructured regions. The SAXSMoW molecular weight estimate, depicted by the shadowed area, estimated the molecular weight of wtTDP-43S to be 22.4 ± 2.4 kDa.

5.6 Wild-type TDP-43S *ab initio* modelling

The wtTDP-43S SAXS curves were used as an input for the *ab initio* reconstruction program GASBOR [114]. A total of ten GASBOR models were produced, to provide a selection of possible solutions describing the TDP-43S structure which all could superimpose the wtTDP-43S 1D curve (Figure 5.7). Each model could fit the data to produce a valid model describing the shape of wtTDP-43S, ranging between $2.86<\chi^2<5.44$ by CRYSOL [117].

The individual models were pooled to incorporate all the unique structural features from the individual GASBOR runs to produce an average 3D model for wtTDP-43S (Figure 5.7).

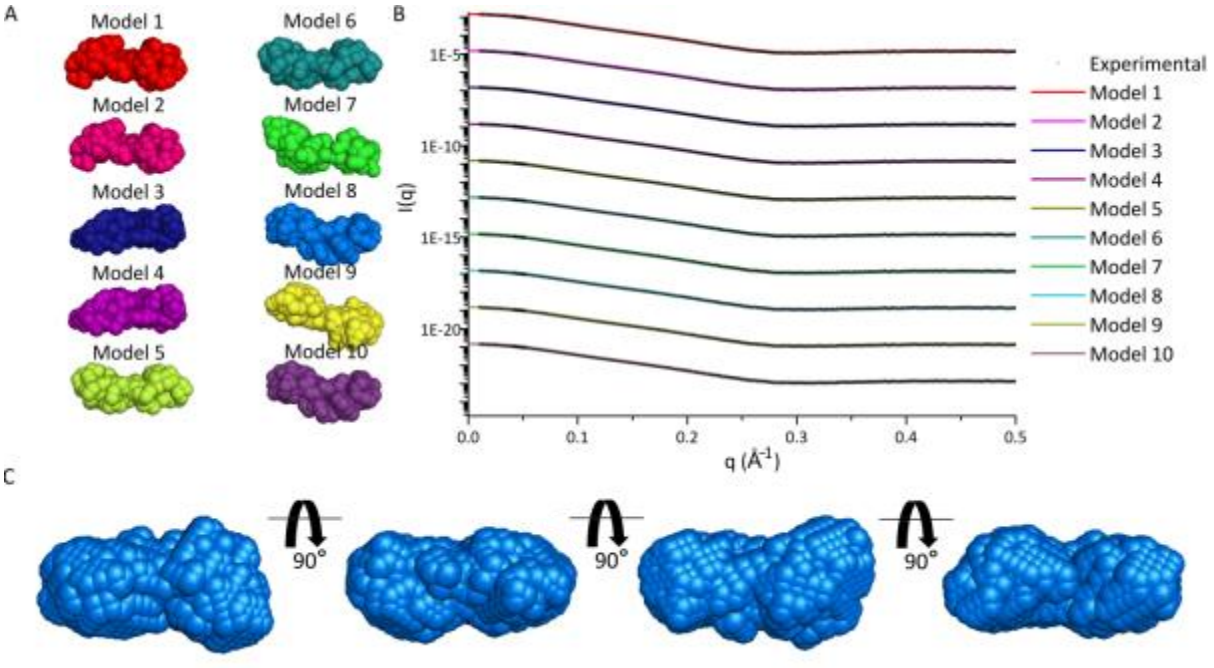


Figure 5.7 wtTDP-43S *ab initio* modelling. A, GASBOR was utilised to produce 10 individual models to represent wtTDP-43S in solution. **B**, Each model's theoretical scattering was compared against the original 1D wtTDP-43S data and goodness of fit ranged between $2.86<\chi^2<5.44$. **C,** The final model was produced by averaging the individual models using DAMAVER [133] to produce a representation with all spatial features from the individual models incorporated.

5.7 The incorporation of known structure (Rigid-bodies)

The two independent domains of wtTDP-43S are known in the form of an NMR structure describing human RRM1 (PDB: 2CQG) and X-ray crystallographic model of mouse RRM2 (PDB id: 3D2W) [77]. These structures can be used as rigid-bodies, described as static structures that are considered when generating SAXS models from the data. The incorporation of rigid-bodies can be used to describe regions of known TDP-43 structure and spatial positions within the 3D solution model. Therefore the spatial compactness of these domains could determine whether it is likely for the tandem RRM domains of TDP-43S to form inter-domain interactions as shown in the structure of hnRNP A1 (UP1) [85]. These bridges in UP1 were orchestrated by Arg75 and Arg88 in RRM1 to Asp155 and Asp157 of RRM2 and are also predominantly conserved in the TDP-43S sequence (Chapter II, Figure 2.3 & 2.4). Although it is worth noting that the UP1 structure is derived from X-ray crystallography so it may be possible that these contacts are forced by crystal packing.

TDP-43S rigid-body structures were derived from the protein data bank but truncated to contain the residues found in the wtTDP-43 construct. The structures of human RRM1 (PDB id: 2CQG, residues 106-177) and mouse RRM2 (PDB id: 3D2W, residues 193-259) without ssDNA were used as the rigid-bodies during SAXS analysis by BUNCH [115]. To ensure that the structure of bound RRM2 did not significantly alter when unbound it was superimposed against the *apo* structure of human RRM2 (PDB id: 1WF0) (Figure. 5.8). The structures superimpose well with an RMSD difference of 0.66 Å differing only at flexible loop regions. The rigid-body structures used for modelling are depicted in figure 5.9.

5.8 TDP-43S rigid-body modelling

The ATSAS software BUNCH [115] was employed to incorporate the RRM structures into a SAXS model. Amino acids not described in these structures were generated as a carbon alpha (Cα) trace by the simulated annealing protocol employed by the

software used to assign their most probable conformation. These Cα traces were subsequently rebuilt to include sidechains by the online program SABBAC [116] based on energy based scoring, whilst still preserving Cα positions. This was performed so that all unique residue features are considered when relating the model back to the experimetal data. A total of eight PDB rigid-body models were generated to provide a consensus of spatial arrangement that could convincingly fit the original 1D data by CRYSOL (Figure 5.10). Fits ranged between χ^2 values of $2.90<\chi^2<4.03$ which could also be matched by their respective theoretical curve upto the q($Å^{-1}$) ~0.47 angular range to show that the 3D models are a good represention of the data. A spatial consensus was achieved with a similar degree of domain seperation observed in all models to demonstrate that the SAXS analysis was concise and repeatable.

From the average spatial arrangement RRM domains exceeded the distance viable for salt-bridges, electrostatic, di-sulphide, hydrogen or Van der Waal bonds to form showing that the RRM domains do not always interact in solution. Whether inter-domain interaction occasionally occur between the RRM-RRM domains cannot be determined by an average model so was investigated further by ensemble optimisation methodology (EOM), described in this chapter, section 5.14.

The best model, defined as having the lowest χ^2 value (χ^2=2.90), could be superimposed into the *ab initio* GASBOR model to support the average solution structure of wtTDP-43S by two independent analytical methods (Figure 5.11). Interestingly the Dmax in the rigid-body models was ~75.5Å which is ~10Å below the Dmax estimated in the P(r) profile. The distance frequency after 75.5Å in the P(r) plot is negligible (Figure 5.4) to suggest that very little density exists at the maximum dimensions which is likely due to flailing end terminals. Note that trying to fit the P(r) to this cut-off results in a uninterpretable profile (Appendix 4) to suggest that some inter-domain distances do extend further.

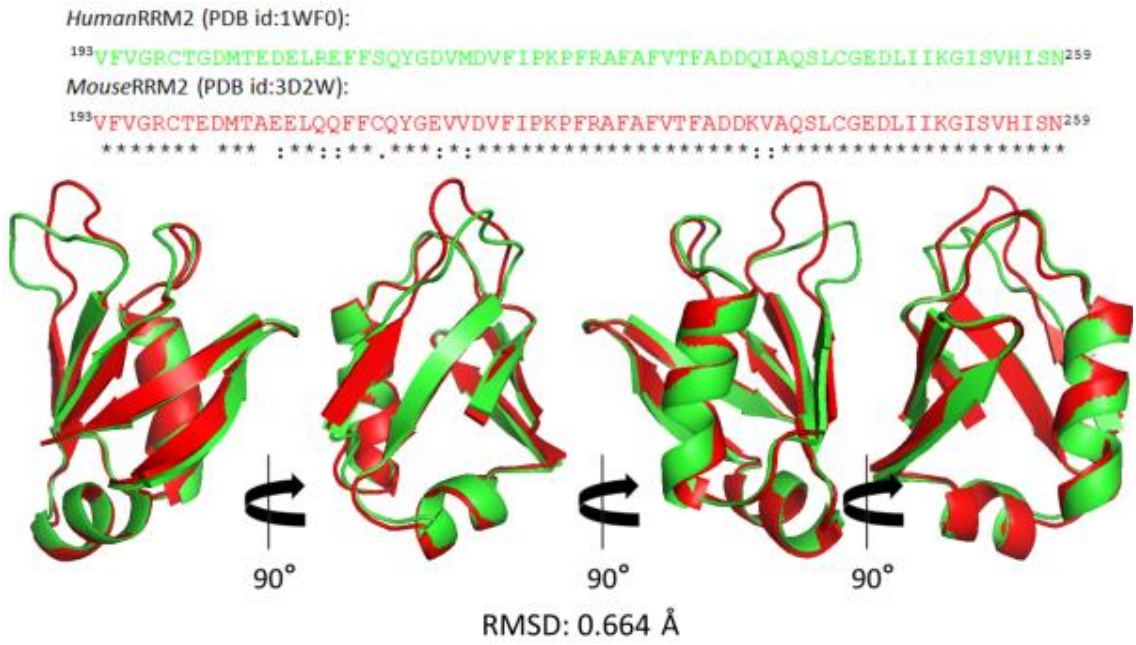


Figure 5.8 Comparison of solution and crystal RRM2 structures from human & mouse. The RRM2 rigid-body from mouse (Red) was ssDNA bound but superimposed against unbound human RRM2 (Green) with an RMSD difference of 0.66 Å.

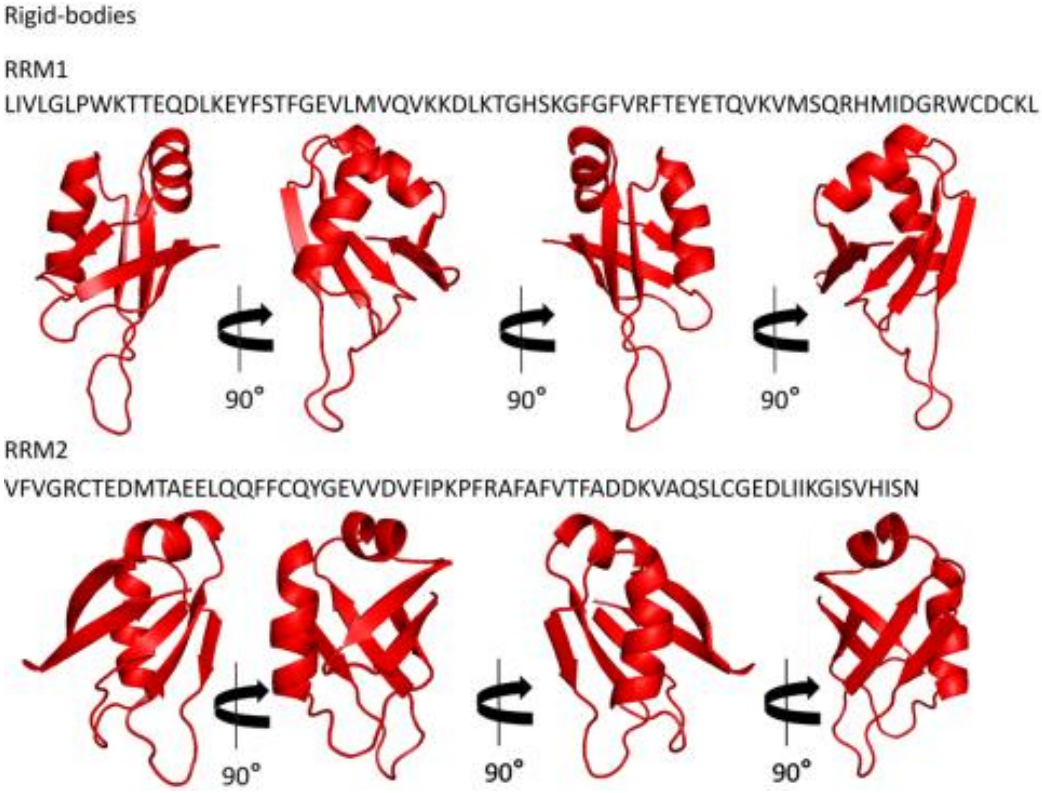


Figure 5.9 Rigid-bodies used with TDP-43S variant modelling. The rigid-bodies used during modelling depicting RRM1 (PDB id: 2CQG) truncated to contain residues 106-177 and RRM2 (PDB id: 3D2W) truncated to residues 193-259.

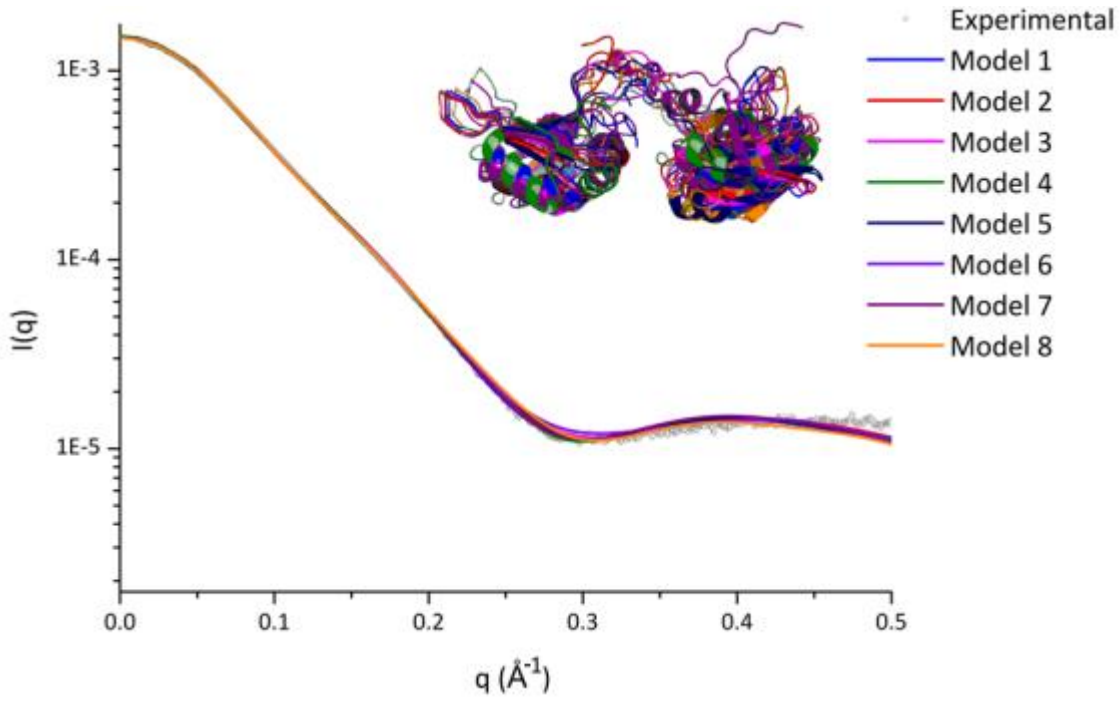


Figure 5.10 Ten individual rigid-body models fit to the 1D wtTDP-43S data . The ten rigid-body models fitted the data up to the ~0.47 q(Å^{-1}) range with χ^2values of between 2.90-4.1. The rigid-body models agreed on the spatial arrangement of the RRM domains. The average spatial arrangement suggests that no inter-domain interactions occur between RRM1 and 2.

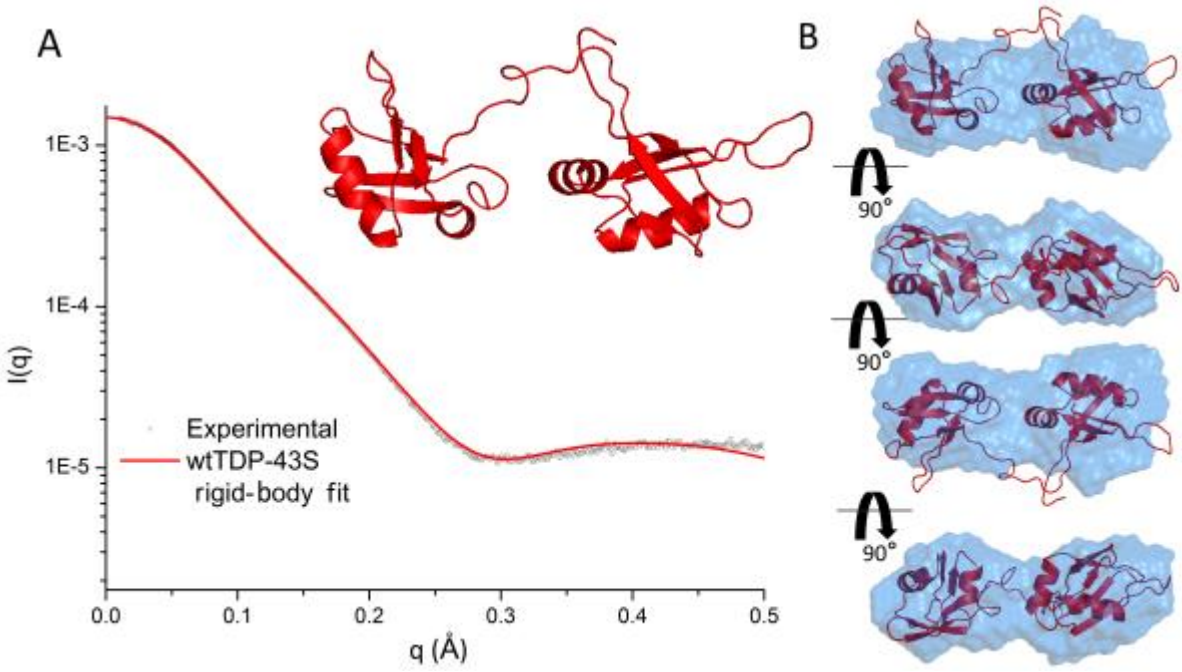


Figure 5.11 WtTDP-43S *ab initio* and rigid-body model comparison. A, The rigid-body model which best matched the data (χ^2: 2.90) fit the 1D curve. **B,** mouse wtTDP-43S could superimpose within GASBORs' *ab initio* model.

5.9 Rg comparisons between mutant TDP-43S

Building on the success of wtTDP-43S SAXS modelling, mutant models were generated in order to compare structural parameters. Two disease associated mutations were tested, D169G and K263E, using the same data collection methodology and data pooling criteria as previously described for wtTDP-43S (Figure 5.12). The mutant 1D profiles were similar to wild-type, however, the concentration of the K263E variant was lower which contributed to a slight increase in noise at elevated q($Å^{-1}$) >0.25 ranges (Figure 5.13).

The Guinier analysis of the data gave comparable Rg's of 22.6 ± 0.1Å, 22.4 ± 0.1Å and 21.8 ± 0.1Å for D169G, K263E and wild-type TDP-43S respectively providing an early indication that construct sizes don't significantly vary. The Rg calculated from the P(r) profile (Figure 5.14) were slightly increased calculated as 23.5 ± 0.1Å, 22.8 ± 0.1Å and 23.3 ± 0.1Å respectively for D169G, K263E and wtTDP-43S.

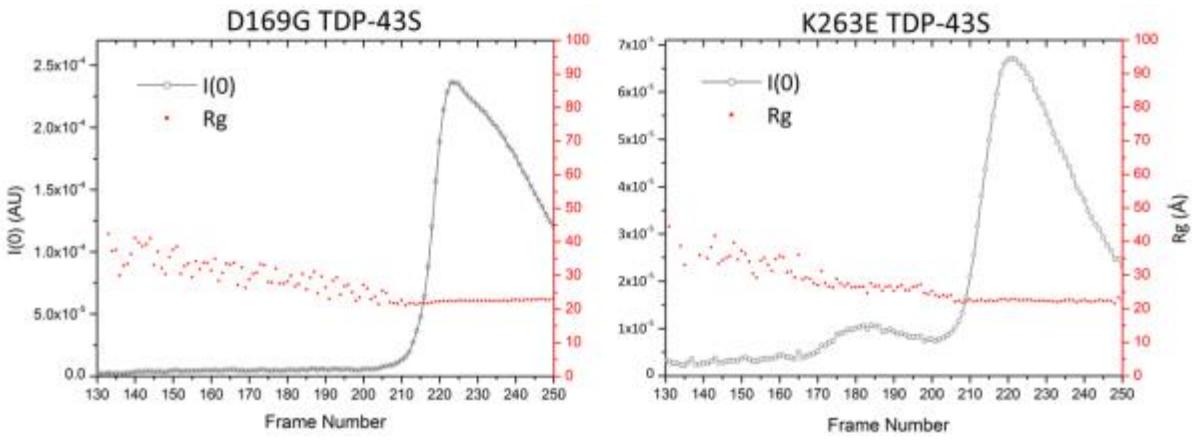


Figure 5.12 Frames used to describe TDP-43S mutants. The elution of TDP-43S mutants through the size exclusion column were analysed by SAXS to yield the scattering intensity I(0) and radius of gyration (Rg) used to identify frames 220-231 for pooling (as described for wild-type TDP-43S).

5.10 Comparison between variant TDP-43S P(r) profiles

The maximum distances were calculated at 85.5 Å for each variant though a higher distance frequency was located in the first peak between 20-35 Å for K263E and

slightly less when approaching Dmax >60 Å (Figure 5.15). This could indicate extra density within one of the globular regions and a reduction in the protein population that extend to these lengths. A possible explanation could be that the K263E mutation inhibits extra flailing of the C-terminal end by forming contacts with RRM2, to both increase electron density at one RRM domain and reduce the number of interatomic distances capable of extending to the longest dimensions (Dmax).

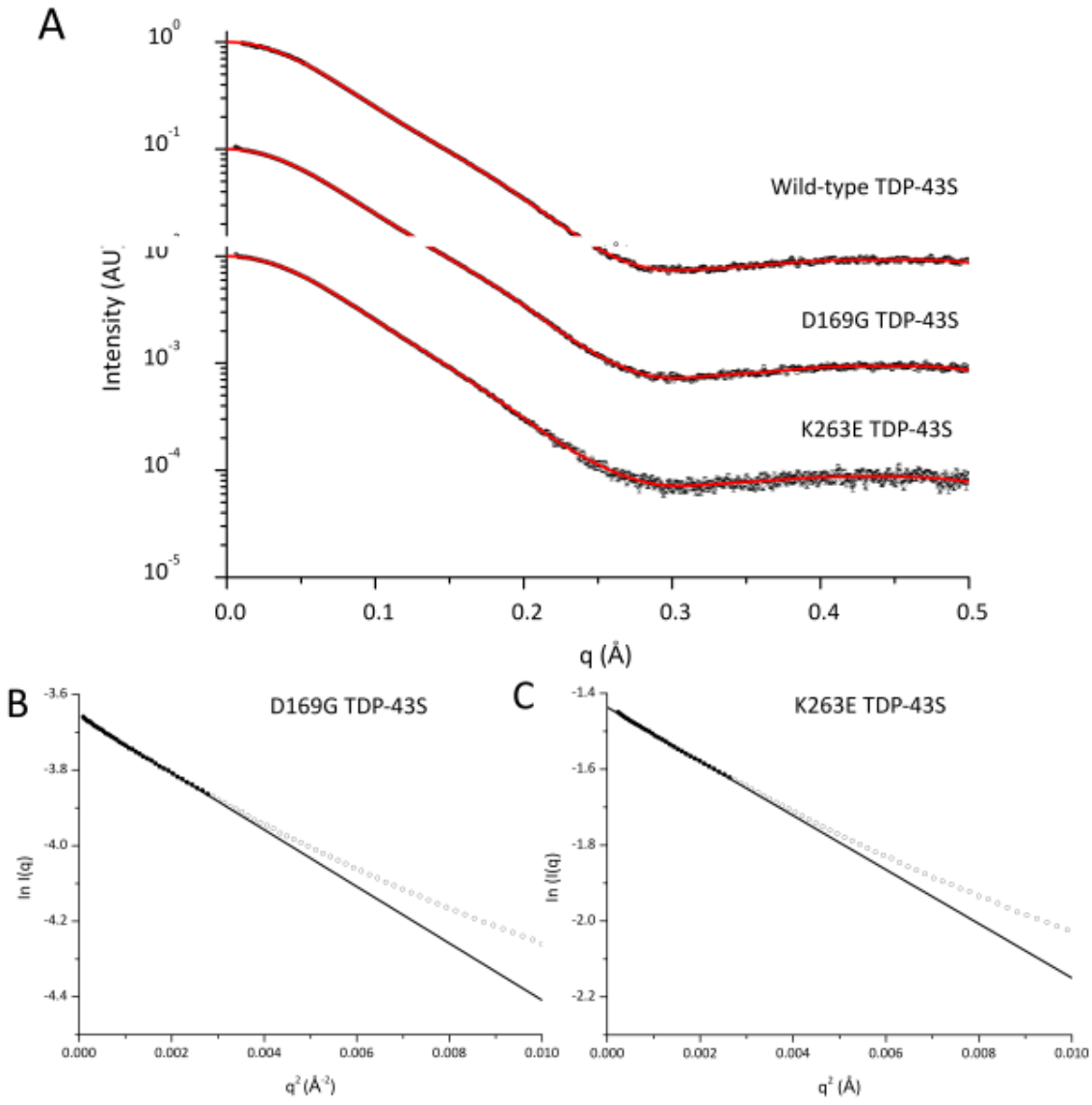


Figure 5.13 1D scattering profiles of TDP-43S variants and mutant Guinier approximation. A, The 1D scattering plots for wild-type, D169G and K263E TDP-43S variants. The 1D curves match well however the K263E mutation had a larger background contrast which increased its error at the higher angles q($Å^{-1}$) >0.25. **B,** The Guinier approximation of D169G resulted in an Rg of 22.6 ± 0.1Å **C,** for K263E TDP-43S an Rg of 22.3 ± 0.1 Å with qRg limits below 1.3.

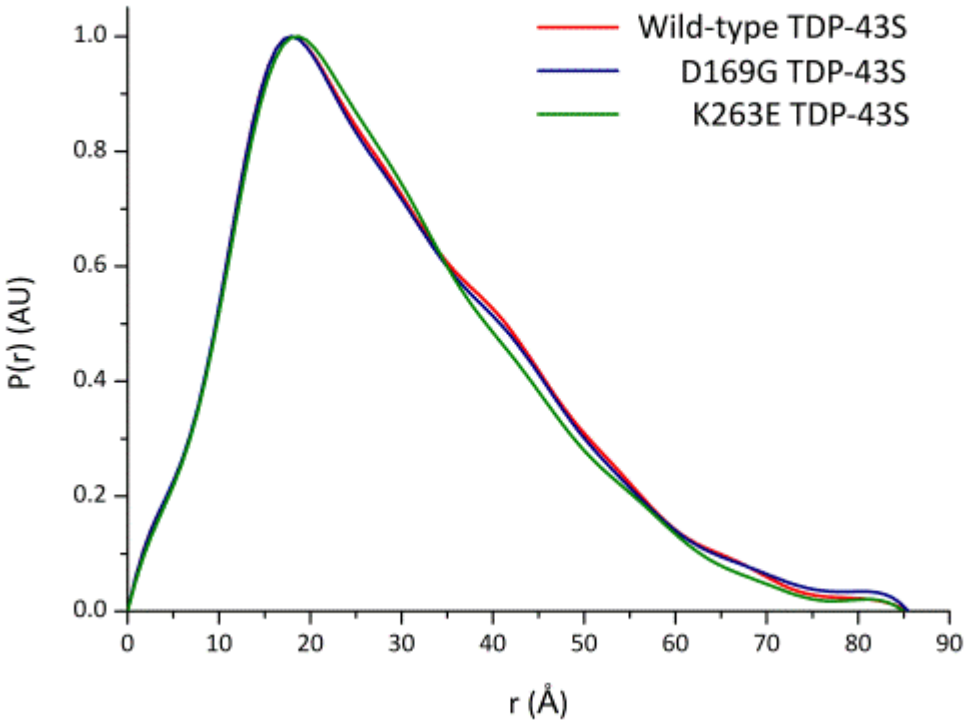


Figure 5.14 Pair-wise distribution of TDP-43S variants. The P(r) profiles of the TDP-43S variants were normalised and plotted against each other. No significant differences were observed between the wild-type (red) and D169G mutated TDP-43S (blue) profiles. The K263E TDP-43S mutant (green) profile varied in distance frequency at ~25 r(Å) peak and nearing Dmax.

5.11 Variant folding using Kratky plots

To assess protein folding differences Kratky plots were compared against variant and wild-type TDP-43S. The molecular weight estimates were within the 10 % experimental error of SAXSMoW [134] calculated at 20.6 ± 0.1, 21.4 ± 0.1 and 22.4 ± 0.1 kDa for D169G, K263E and wild-type TDP-43S respectively to suggest a similar structure (Figure 5.15).

However TDP-43S constructs were all shown to deviate from the criteria described for globular protein when compared through a dimensionless Kratky [110] (described in chapter IV). This was performed by normalising each construct to the I(0) and Rg of wild-type TDP-43S. The maximum peak height surpassed

$(qRg)^2I(q)/I(0)$ of 1.1 and didn't peak until after the qRg range, deviating from what is expected for a perfectly globular species. This deviation was similar for the multi-domained p47 phagocytic oxidase (PHOX) protein, which was shown to be separated by a flexible linker [110] and which is also likely to be the case for all TDP-43S variants.

The D169G mutant and wild-type TDP-43S had an identical overlay to indicate no changes. However, a narrower curve between qRg 0.0-5.5 was seen for the K263E mutant (Figure 5.15). This narrowing suggests that less surface area is accessible to the surrounding solvent, supporting the hypotheses that its C-terminal region is less exposed, perhaps interacting with RRM2.

5.12 Variant *ab initio* modelling and comparison

In order to compare structural differences conveyed by TDP-43S mutations, *ab initio* models were constructed as previously described for D169G and K263E mutated TDP-43S. All 10 models could re-fit the 1D data by CRYSOL for each construct with χ^2 minimisation ranges of between $5.77<\chi^2<8.27$ and $1.74<\chi^2<2.21$ respectively for D169G and K263E (Figure 5.16, 5.17). The averaged structures as expected produced comparable models to the wtTDP-43S protein with no differences to the maximum dimensions or signs that the domains are more compact or elongated (Figure 5.18).

The observed similarities between mutant and wild-type models suggest that mutations do not cause a global change to TDP-43S structure. The observed subtle differences regarding the K263E mutant P(r) profile did not seem to significantly affect the distribution of the average electron density of its 3D *ab initio* model. Therefore it is difficult to ascertain whether the mutation can mediate an interaction with the C-terminal to restrict flailing or whether noise attributed by a lower concentration allows the fit to deviate slightly more at higher q($Å^{-1}$) ranges during *ab initio* modelling. Due to the low resolution of SAXS, this observation would require further analysis by either NMR or X-ray crystallography.

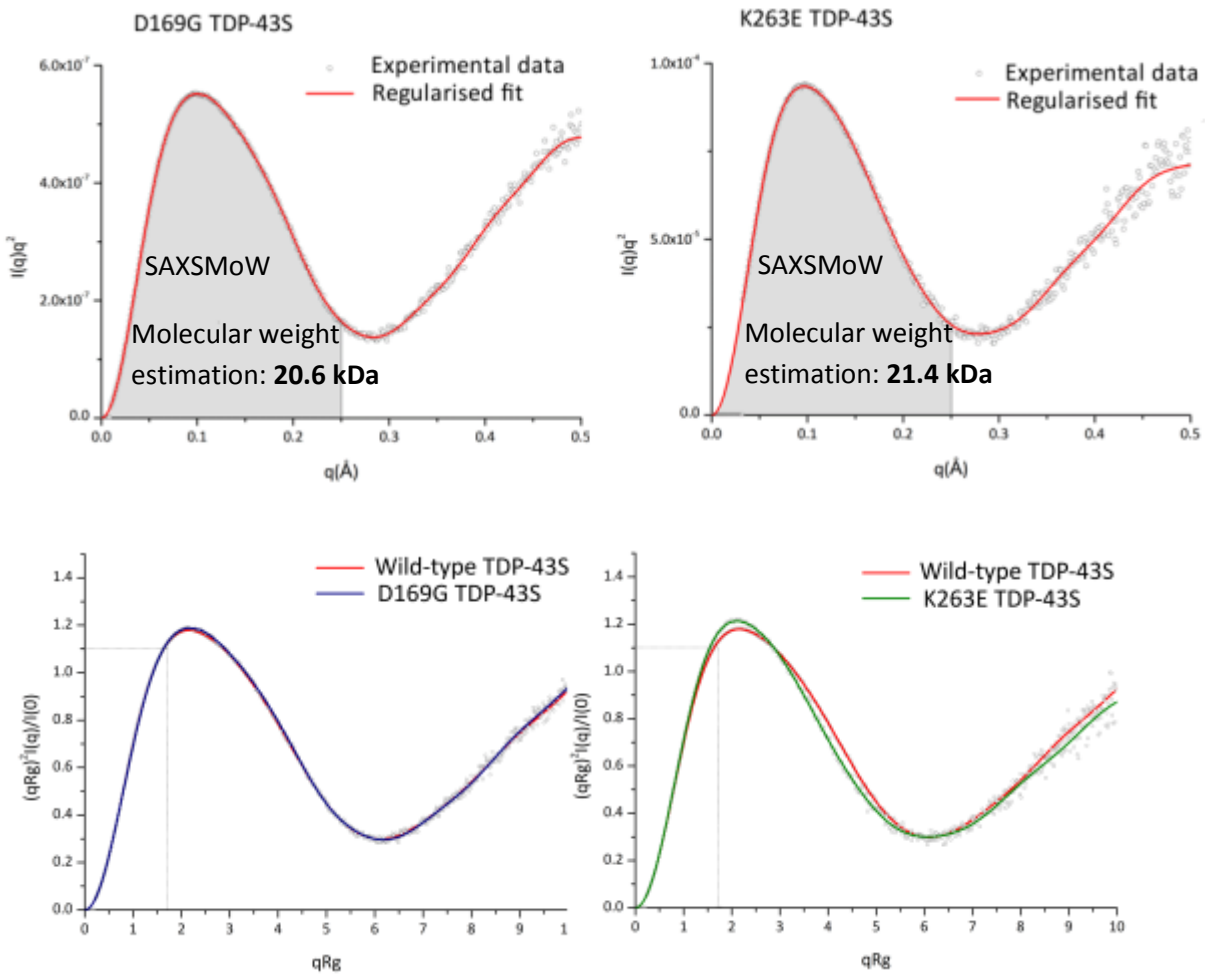


Figure 5.15 Dimensionless Kratky plots of the TDP-43S variants. The wtTDP-43S profile was used as a reference for the folding of both mutant variants shown in red. The peak height as described for ideal globular structures are depicted as a dashed line [110]. Profiles on the left show the D169G mutant which match the wtTDP-43S profile to indicate no structural changes. The right profiles depict K263E which contains a higher and narrower peak compared to wild-type, indicative of lower solvent exposure.

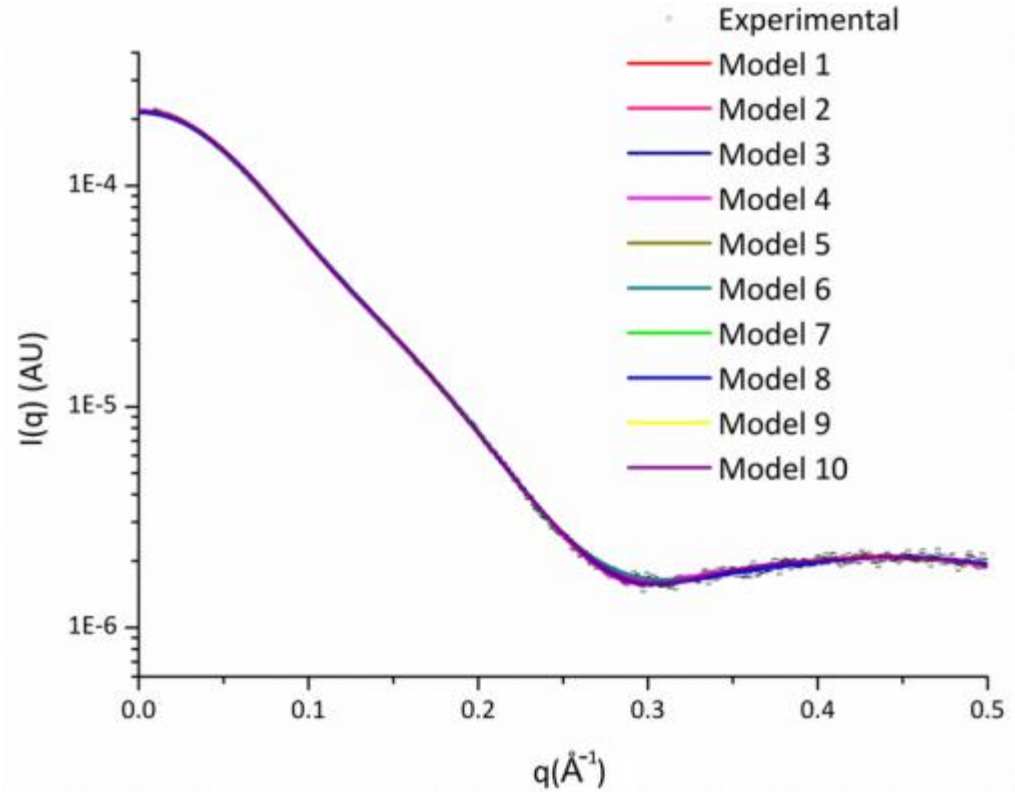


Figure 5.16 *Ab initio* model fits of D169G TDP-43S. Each of the individual models derived from GASBOR (Figure 5.18) fitted the entire experimental scattering profile.

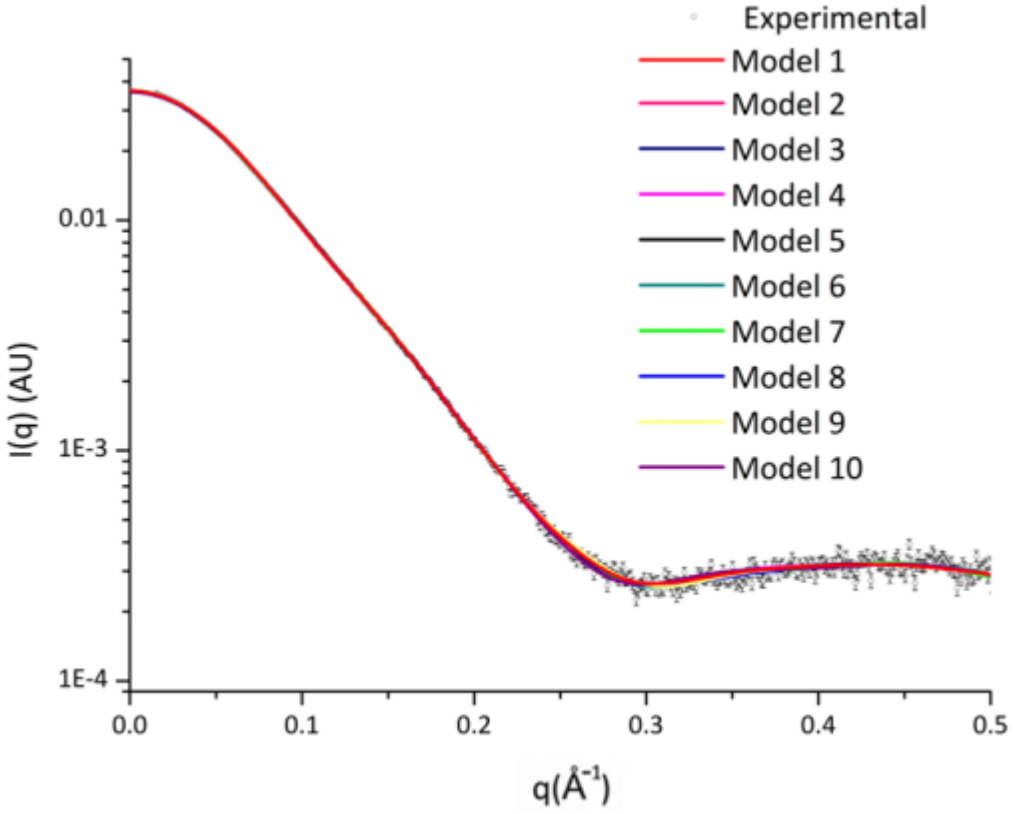


Figure 5.17 *Ab initio* model fits of K263E TDP-43S. Each of the individual models derived from GASBOR (Figure 5.18) fitted the entire experimental scattering profile.

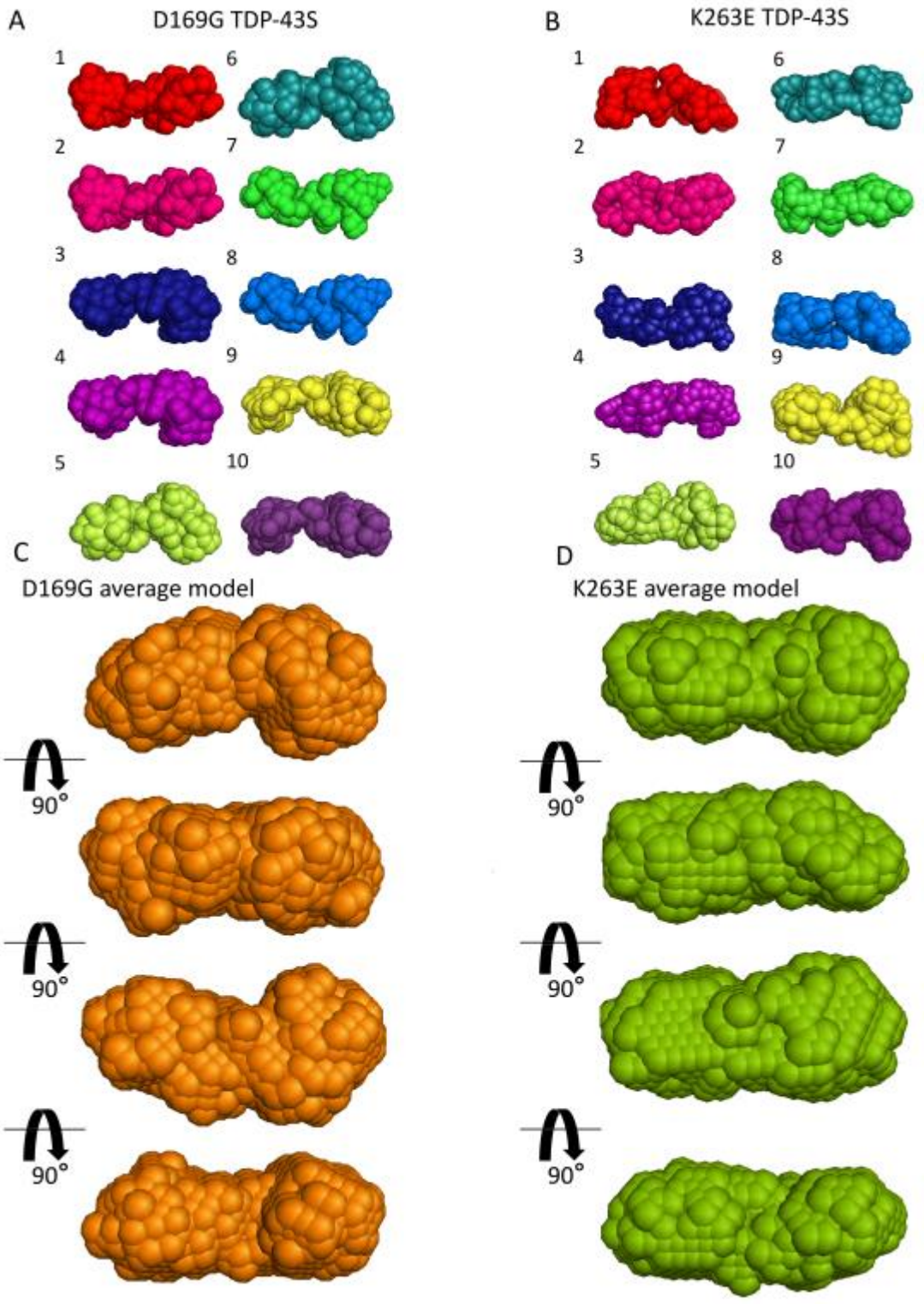


Figure 5.18 *Ab initio* models of D169G and K263E TDP-43S. A total ten *ab initio* GASBOR models were generated for each TDP-43S variant protein **A,** Showing the individual *ab initio* models of the D169G and **B,** K263E TDP-43S mutations. The averaged structures performed by DAMAVER produced **C,** The average electron distribution model of D169G and **D,** K263E TDP-43S.

5.13 Rigid-body modelling of TDP-43S variants

The same rigid-bodies and modelling procedures previously described for wtTDP-43S were performed for D169G and K263E TDP-43S to determine whether mutations disrupt the RRM-RRM spatial separation. The best fits based on the lowest χ^2 were χ^2=3.272 (Figure 5.19) and χ^2=2.059 (Figure 5.20) for D169G and K263E mutations respectively and in both models theoretical curves could re-fit the experimental 1D data well, up to at least the q(Å^{-1}) 0.45 range.

All TDP-43S variants tested could match the RRM spatial separation of wtTDP-43S with the same maximum distance of 75.5Å (Figure 5.21). Revealing that RRM disease mutations are unlikely to induce abnormal RRM inter-domain interactions and that these structures are likely conserved in solution. This implies that neurodegenerative disease mutations do not cause large areas of domain unfolding or conformational changes. Both *ab initio* and rigid-body models for all variants can superimpose to show constructs essentially share an identical global structure at this resolution (Figure 5.22).

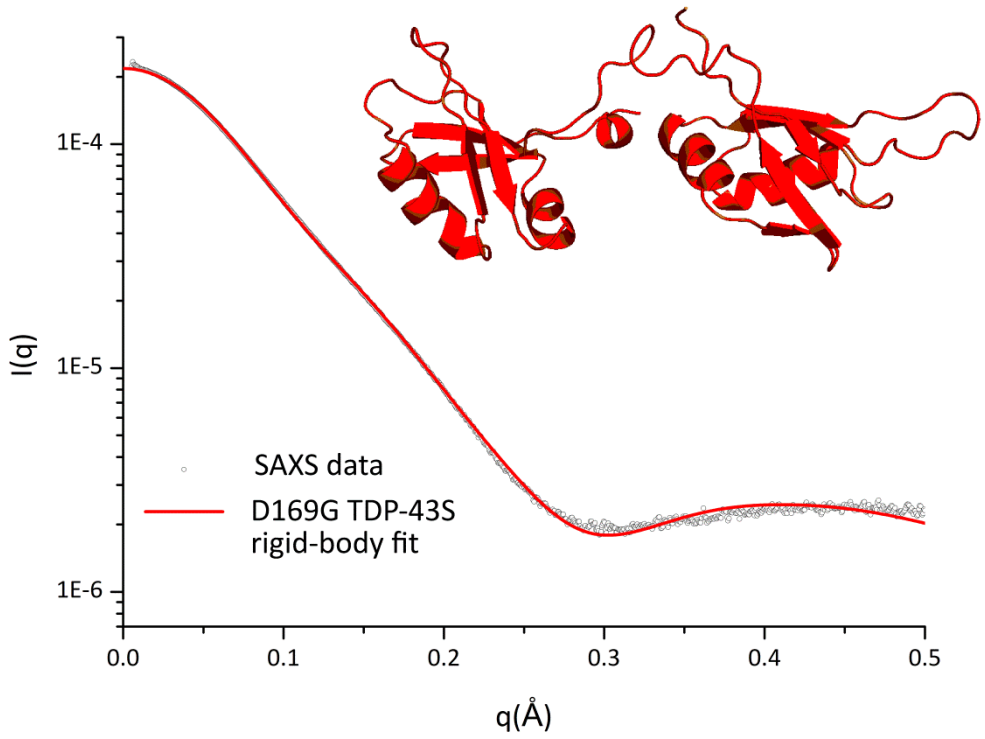


Figure 5.19 Rigid-body model of D169G TDP-43S. The fit of the D169 TDP-43S rigid-body against the experimental 1D curve.

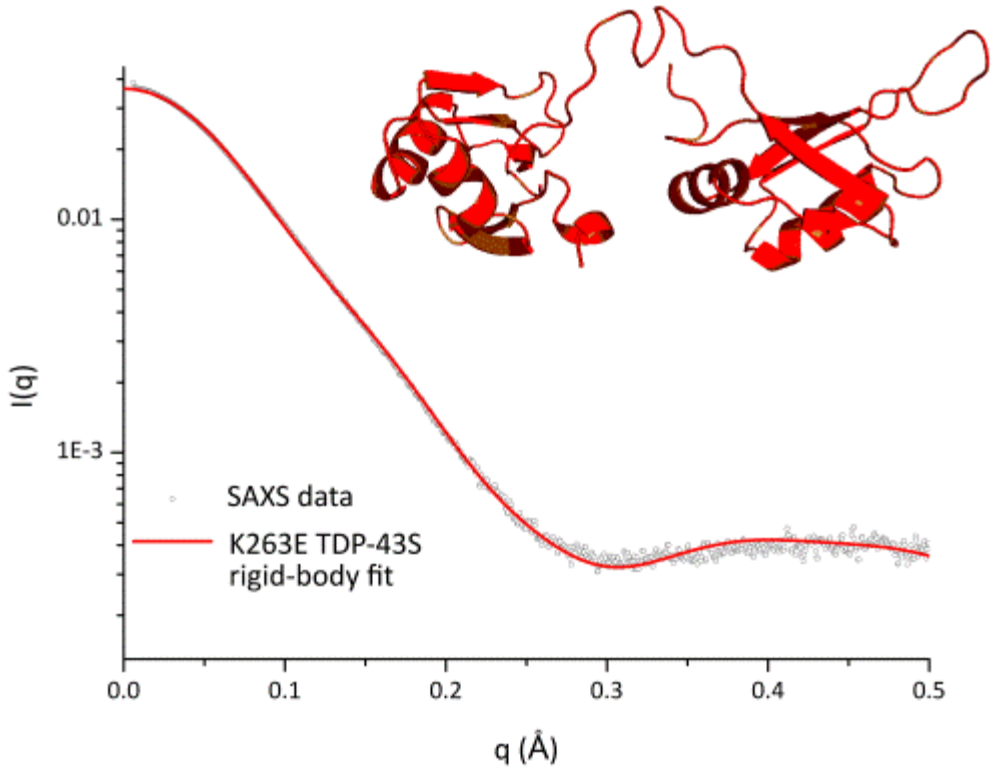


Figure 5.20 Rigid-body model of K263E TDP-43S. The fit of the K263E TDP-43S rigid-body against the experimental 1D curve.

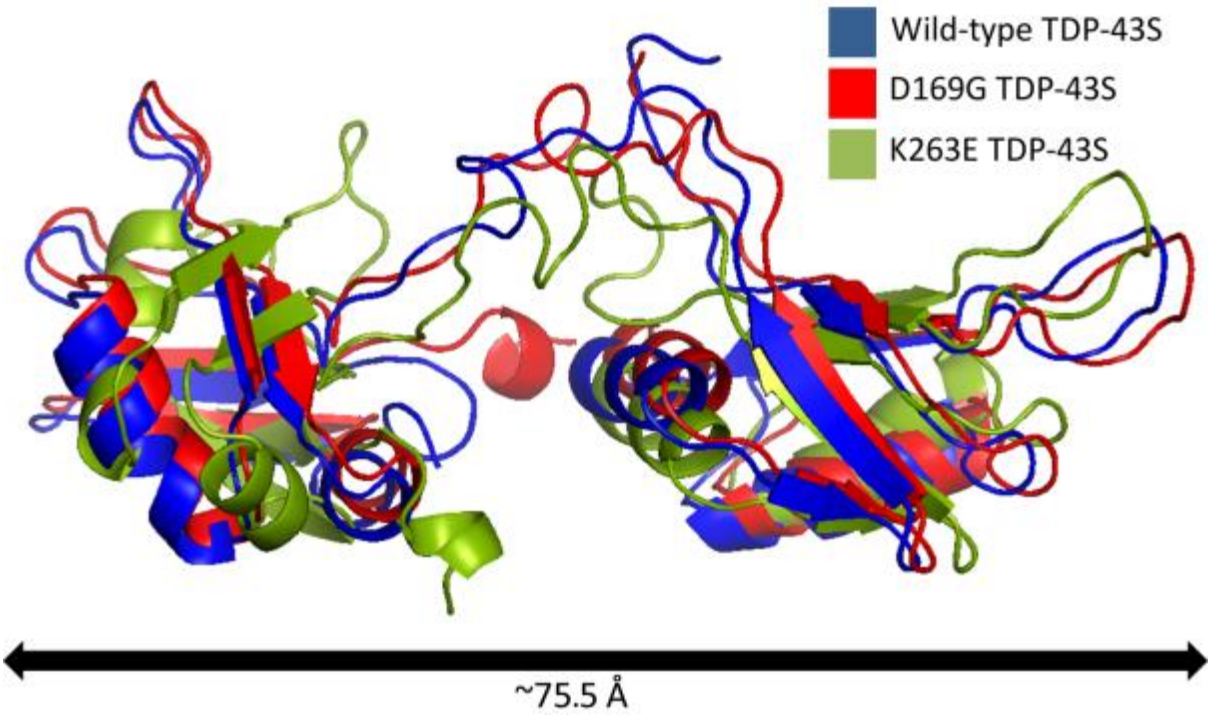


Figure 5.21 Spatial separation of the tandem RRM domains in the TDP-43S constructs. The rigid-body models for TDP-43S variants consistently separate the RRM domains the same spatial distance apart in the average model.

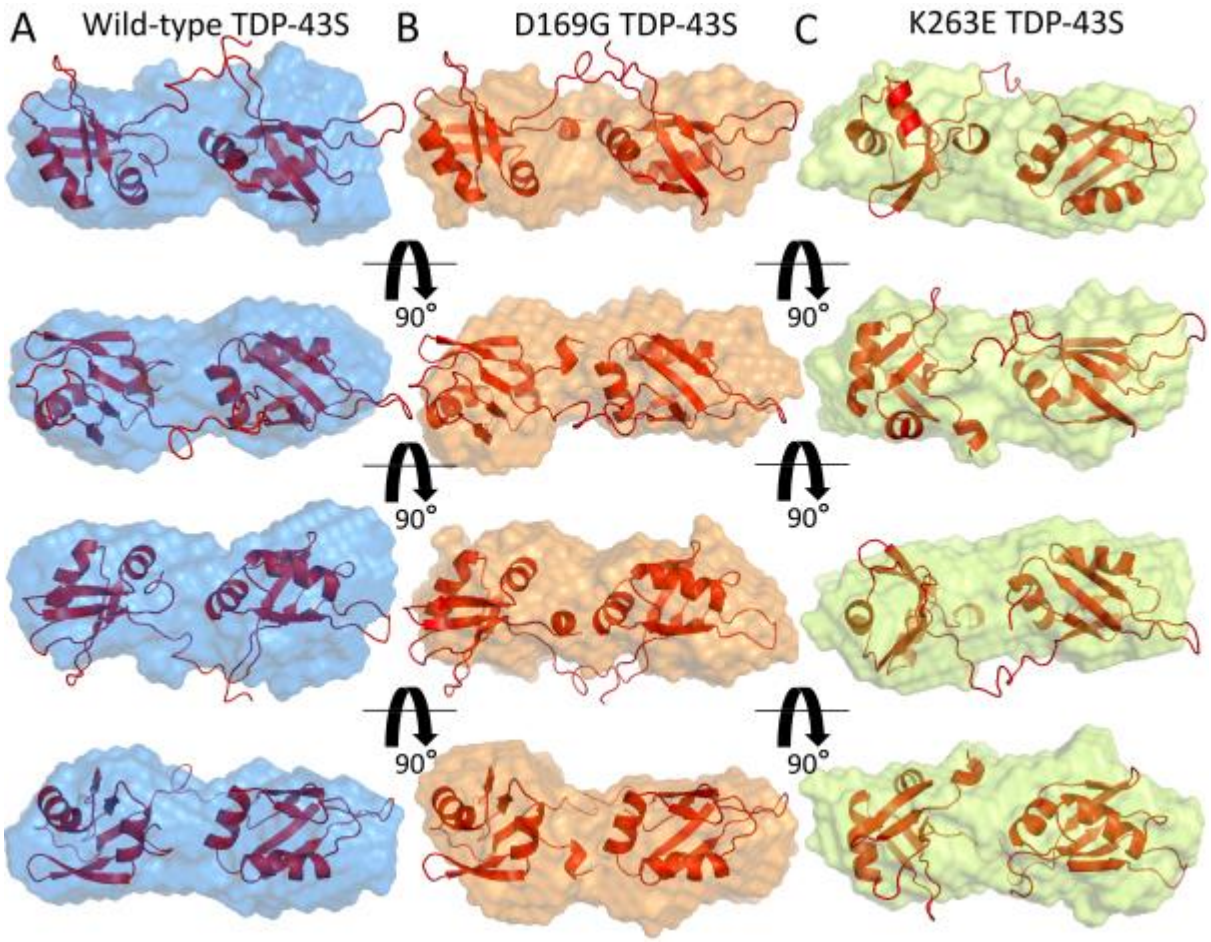


Figure 5.22 *Ab initio* and rigid-body fits of the TDP-43S variants. Comparisons between the SAXS models for **A,** wild-type **B,** D169G and **C,** K263E TDP-43S . All models support the *ab initio* reconstruction and spatially arrange the RRM domains the same distance apart.

5.14 Determining the flexibility of TDP-43 by EOM

To determine the extent of TDP-43S' flexibility as suggested by the P(r) and Kratky plots, EOM was employed. This technique is adapted for flexible systems and correctly considers that the 1D SAXS curve is comprised of different sized TDP-43S subpopulations. EOM attempts to match the experimental curve by changing the ratio of different sized models probing the entire spatial configuration until the theoretical scattering converges with the experimental data, as explained in chapter IV.

An ensemble consisting of 10,000 TDP-43S models with rigid-body constraints were generated for each variant. The results could fit the curves using two sub-populations of structures which peaked with an Rg at ~20 Å and ~27 Å for each construct (Figure 5.23). The curves converged with χ^2 values of 1.92, 2.76 and 1.81 for wild-type, D169G and K263E TDP-43S respectively indicating a good fit to the data as indicated (Figure 2.24).

The results suggest that TDP-43S and mutant variants are able to elongate from ~50 to ~90 Å in length. The fact that the histograms are predominantly generated from two populations suggests that there is a transition, perhaps limited by some rigidity in the linker to restrict the amount of intermediate protein lengths. Interestingly the largest contribution is from the shorter population with an Rg of ~20 Å, similar to UP1. From comparing the lengths of the mean structures in each population (used in the ensemble) the full extent of TDP-43S flexibility in solution can be appreciated (Figure 5.25). This shows that the linker is able to fully elongate but rarely do the two domains come into contact (from models used in the converging populations).

Interestingly, even the RRMs in the mean structure describing the shortest Dmax (~57Å) exceed the distance for inter-domain interactions to occur. Taken together with the features of the Kratky and P(r) plots the TDP-43S construct is flexible, but instead of oscillating in a smooth action the RRMs movement may be restricted by some rigidity of the linker, as suggested by two clear subpopulations in EOM. More importantly disease related mutations occurring in the TDP-43S region do not affect the protein's flexibility, inter-domain interactions or global unfolding, which should all be reflected in the experimental curves, fits and structures of each TDP-43S variant.

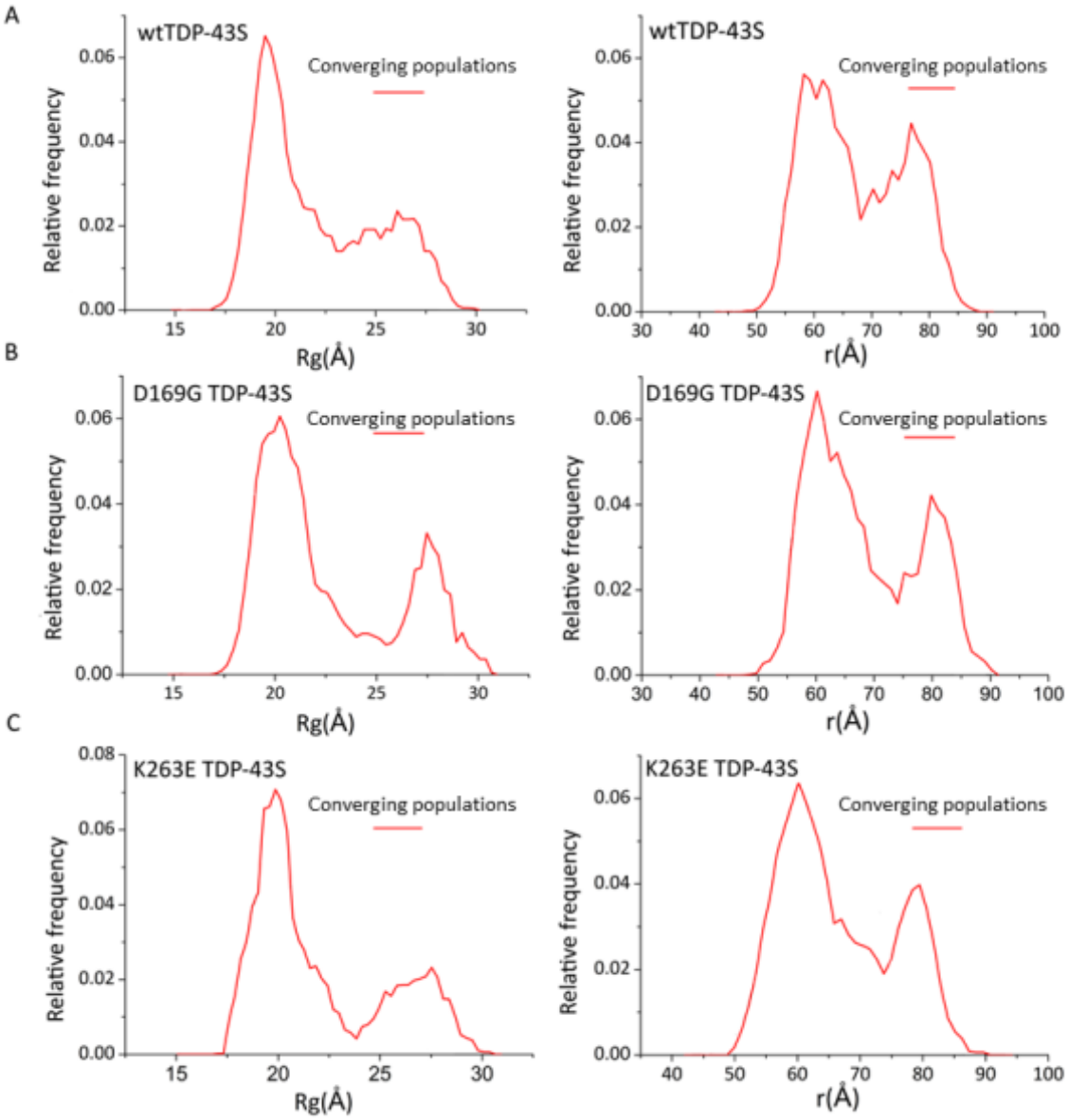


Figure 5.23 Ensemble optimisation methodology between TDP-43S variants. From The ensemble pool of structure (10,000 models), variant TDP-43S 1D profiles could converge using subset of these structures. **A,** wild-type **B,** D169G and **C,** K263E TDP-43S variants. The fit comprised of two extreme populations in regards to Rg and Dmax from ~20 to ~27 Å and from a maximum length of ~58 to 80 Å. Indicating a flexible protein with no significant differences between mutations. The converging fit can be seen Figure 5.24.

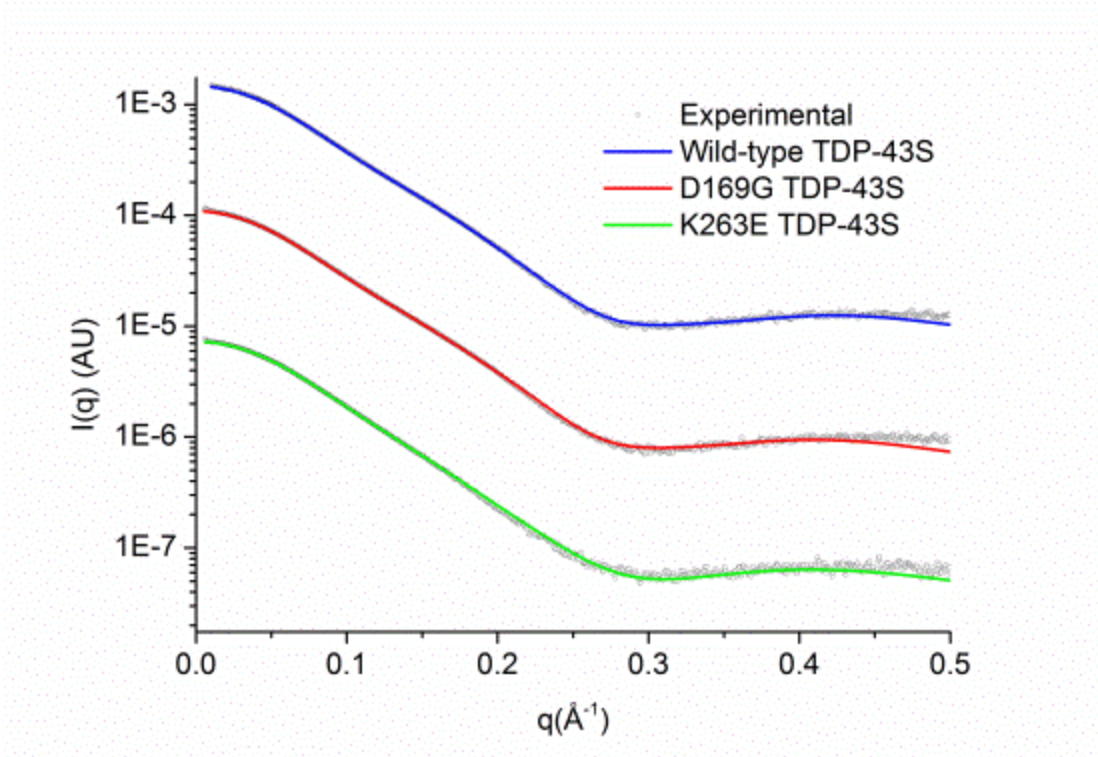


Figure 5.24 EOM converging population fit to the original TDP-43S variants 1D data. The fits of the converging populations shown in Figure 5.23 back to the original 1D data. The fits superimpose the experiemtal data well with low χ^2 values of 1.92, 2.76 and 1.81 for wild-type, D169G and K263E respectively.

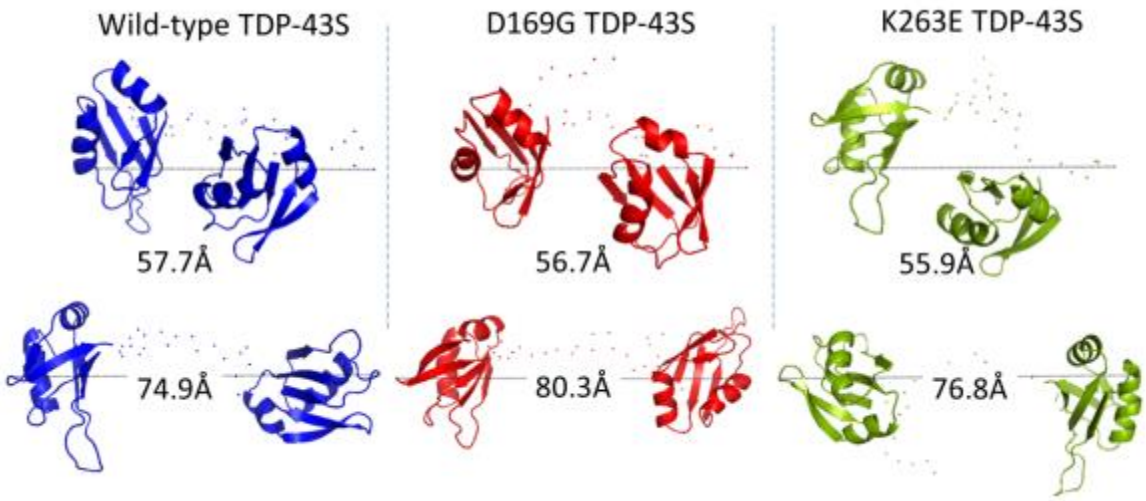


Figure 5.25 The elongation of TDP-43S variants. Spatial comparison between two manually chosen structures from the ensemble pool but included in the final converging fit for all TDP-43S variants. These describe the mean population sizes at Dmax ~57 and ~77 Å. The compact population still orientates the RRMs of TDP-43 outside the range for interactions to occur and at ~77 Å shows that the linker is able

to fully extend. Linker and terminal regions were not reconstituted and are shown as Cα trace.

5.15 The secondary structure content of TDP-43S predicted by circular dichroism

TDP-43S SAXS models were shown to be similar to suggest neurodegenerative disease mutations do not cause onset through large structural disruption. However, at this resolution we cannot exclude structural changes occurring at the local environment surrounding the mutation. Some proteins have shown that a single point mutation can change secondary structure with impacts on function and turn-over. Consider the G1015V point-mutation in human collagen alpha5(IV) chain, where a significant amount of secondary structure content is changed which is associated with disease severity of Alport syndrome [135].

Using CD, disease associated mutations in TDP-43S were analysed to detect whether there is any obvious change to their secondary structure content. CD measurements are most effective when measuring at the far-UV regions, 170-180nm, to uncover unique structural features of the protein (explained in Chapter IV, Figure 4.3). Chlorine is a large absorber at this range so HCL and NaCl were replaced by phosphate buffer and NaF to maintain protein solubility, maximise CD signal from protein and reduce background noise.

CD profiles were measured using protein concentrations of 0.81 mg/ml scanning between the 174-260 nm wavelengths (Figure 5.26). The spectra from all variants were comparable with errors that overlay to indicate that secondary structure is conserved in all constructs (Figure 5.26). To determine the secondary structure composition within TDP-43S the CDSSTR variable selection algorithm was utilised [122,124] as described in chapter IV.

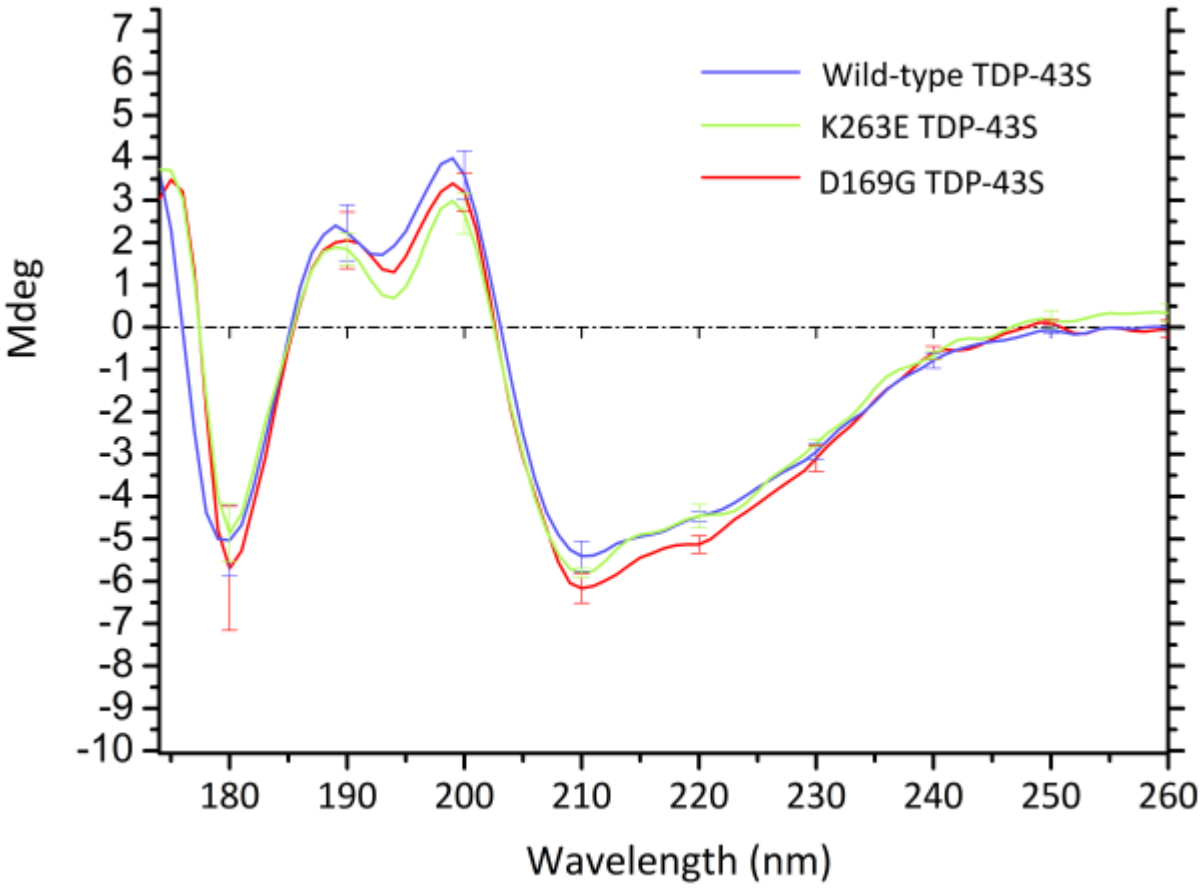


Figure 5.26 CD profile overlay of TDP-43S profiles. The profiles overlay between wild-type, D169G and K263E TDP-43S. Showing errors overlap to support similar secondary structure content.

The reference set of proteins were derived from the SP175 dataset adapted for use with synchrotron CD data. The SP175 reference set is comprised of 71 annotated proteins each with a CD profile measured from 175-240 nm and a high resolution structure available to increase the accuracy of relating structural features back to the experimental profile [136].

Secondary structure content estimates from CD by analytical algorithms should fit the experimental data with a normalised root mean squared deviation (NRMSD) of <0.1 [121]. The CDSSTR algorithm fit the full range from 175-240 nm with an NRMSD range of 0.11-0.14 (Table 5.1) which was improved to 0.091-0.097 when excluding data <190 nm (Figure 5.27). Unfortunately, secondary structures in proteins absorb

strongly between 175-200nm illustrated in chapter IV, figure 4.3 therefore both ranges were used to determine the proportion of secondary structure (Table 5.1).

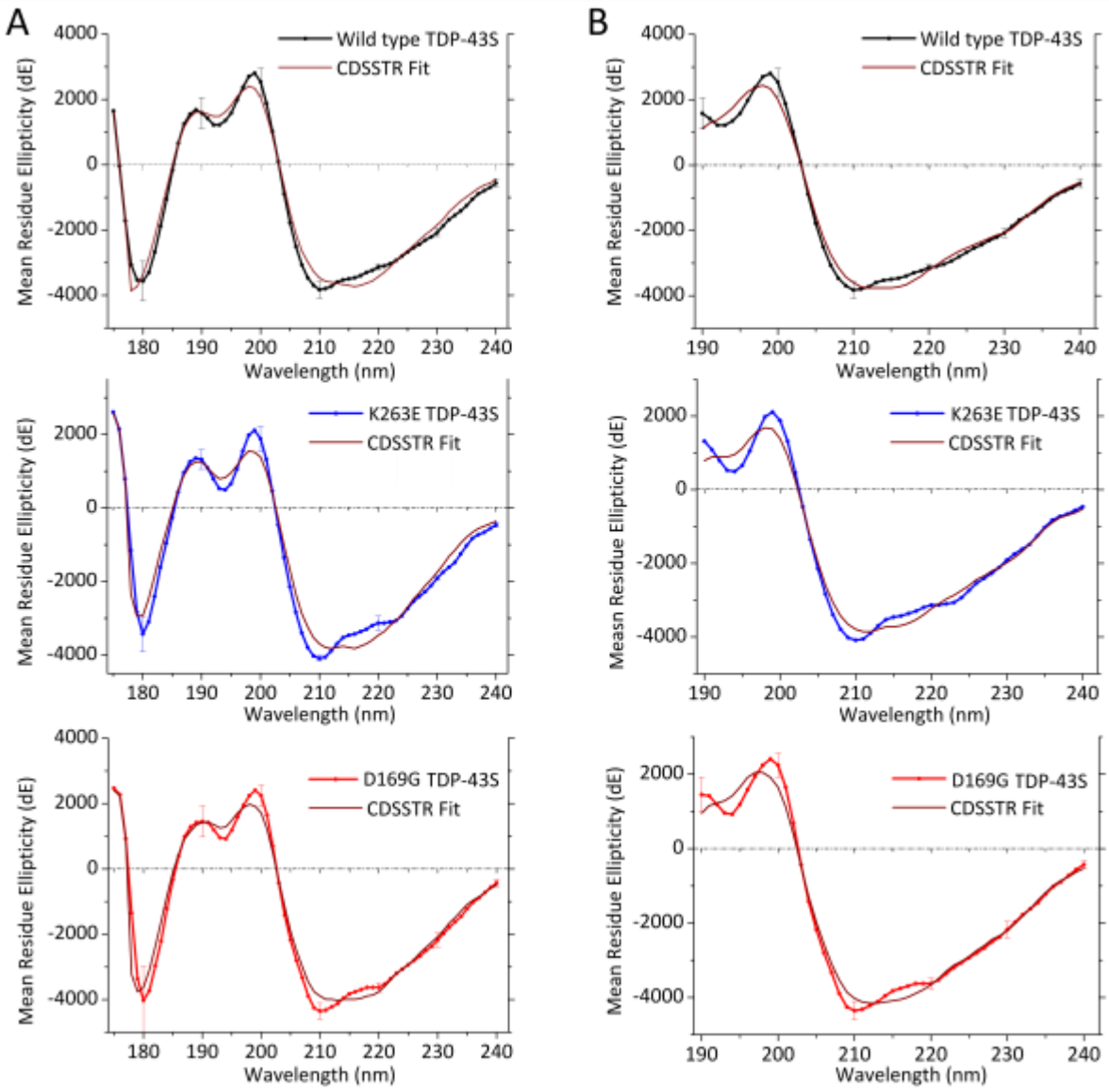


Figure 5.27 CDSSTR fits using full and truncated profiles. The CDSSTR secondary structure estimation software was used to assign secondary structure content to TDP-43S profiles. The algorithm could fit the majority of the profiles even at the far UV range. The fits were judged by the normalised root mean squared deviation (NRMSD) (as shown in Table 5.1).

CDSSTR predicted a β-sheet rich structure in all variants ranging from 38-41% and a low α-helix content of 6%. This is lower than expected by comparing the deposited RRM1 and 2 (PDB id; 2CQG and 3D2W) structures assuming no secondary structure

in the terminal or linker regions, which estimates 21% α-helix, 34% β-sheet, 8% turns and 37% disorder (Table 5.1). The CDSSTR prediction therefore underestimated the α-helix content by ~15% which

Table 5.1. Secondary structure content of TDP-43S profiles.

TDP-43S Variant	NRMSD		α-Helix		β-Sheet		Turn		Unordered	
+/-	+	-	+	-	+	-	+	-	+	-
Wild-type	0.11	0.09	6%	6%	41%	39%	11%	13%	41%	41%
K263E	0.14	0.10	6%	6%	38%	39%	13%	13%	41%	40%
D169G	0.13	0.09	6%	6%	39%	40%	13%	13%	40%	41%
PDB files;										
2CQG, 3D2W	N/A	N/A	21%		34%		8%		37.5%	

Two profiles were calculated using the full data range from 175 nm (+) and truncated before 190 nm (-) to 240nm.

was also observed by Greenfield using this method and data <208 nm [137,138]. A similar trend is observed in a fragment of the periplasmic receptor protein, TolA where CDSSTRprovided the most accurate fit but also underestimated the α-helix content. The authors suggest this may be due to the flexible extensions of the protein absent in the TolA crystal structure [139] which may also be true of TDP-43S. Other CD algorithms e.g. CONTIN [125], SELCON3 [126] were also used but could not accurately fit the data with NRMSD values in excess of 0.2 (*data not shown*).

However the CD profiles for each TDP-43S variant could be superimposed to within the errors of the data, to show that the secondary structure is likely to be similar between all TDP-43S variants.

5.16 The disruption of RRM structure is an unlikely cause for TDP-43S mediated disease

Susceptibility to thermal and chemical unfolding is a common characteristic of proteins involved in neurodegeneration [140–143] exemplified by ALS mutations in superoxide dismutase-1 (SOD1). Many of the ALS associated mutations in SOD1, situated throughout the structure are believed to cause structural instability by exposing hydrophobic regions and increasing the protein's aggregation potential [143,144]. Similar aggregation is associated with protein mis-folding exemplified by amyloid-β In Alzheimers' and α-synuclein in Parkinsons' diseases [145] to suggest that neurodegeneration is largely associated with protein instability.

In contrast, here using two structural solution techniques, SAXS and CD, show that disease mutations associated with TDP-43 RRM domains cause no change to structure. The RRM domains are predicted to contain most of TDP-43's ordered structure but the majority of mutations occur within the disordered C-terminal domain [62]. It is therefore unlikely that conformational changes or unfolding within the structure of TDP-43 is the main cause of disease onset. This is in direct contrast to the mechanism described in other proteins involved in neurodegenerative toxicity [140–143] to suggest that other protein properties such as function and stability could be important disease determinants. Therefore the structural stability of TDP-43S was tested using thermal denaturating assays to substantiate this further.

Chapter VI

Approaches used to determine structural stability and DNA affinity of TDP-43S constructs

6.1 Differential scanning fluorimetry (DSF)

This technique was employed to assess the melting transitions (T_m) between TDP-43S and its disease related mutations. This method uses a fluorescent dye which can be tracked by its fluorescent emission when interacting with non-polar environments like that of exposed hydrophobic residues [32].

During this study the Sypro Orange™ dye was used as similarly described for other DSF protein unfolding experiments [33]. This is adapted for protein unfolding since its wavelength emission (~610 nm) when excited (~470 nm) is sequestered by aqueous environments so is specific to exposed hydrophobic amino acids of proteins in a suitable buffer. The strength of the Sypro Orange™ fluorescent signal is proportional to the amount of dye molecules bound to a protein, which increases as the protein unfolds (Figure 6.1) [34]. Since this dye binds indiscriminately it is essential that samples are homogeneous to avoid measuring contaminant proteins.

Assays can be developed to test protein unfolding as a function of several parameters including e.g. urea, DTT, ligands, temperature or pH. Each are designed to unfold protein amino acid tertiary structure by disrupting bonds to cause unfolding. Protein stability is conferred by hydrogen, Van der Waals, disulphide and polar interactions which differ in number between proteins. Therefore, one is able to determine the stability of a protein by slowly incrementing the stress (i.e. via increasing concentration, acidity, temperature). Temperature was used during this study as it is readily repeatable and modern PCR equipment can carefully regulate small ˚C increments.

The experiment required a fluorimeter capable of emitting and measuring the required wavelengths through a temperature range that will cause the protein to

unfold, typically up to ~95˚C. Experiments are generally performed using a real-time PCR machine since sample volumes are small and modern instruments are capable of slow temperature increments and tracking at these wavelengths.

The resulting curves monitor the extent of unfolding by increases in fluorescence signal. It is not uncommon to observe a decline in signal after the flouresence peaks. This can be attributed to the removal of the dye by protein aggregation as shown in the DSF schematic [32] (Figure 6.1). The melting temperature from this curve can be obtained via the parameters of a sigmoidal Boltzmann fit or the plotting of its first derivatives [146].

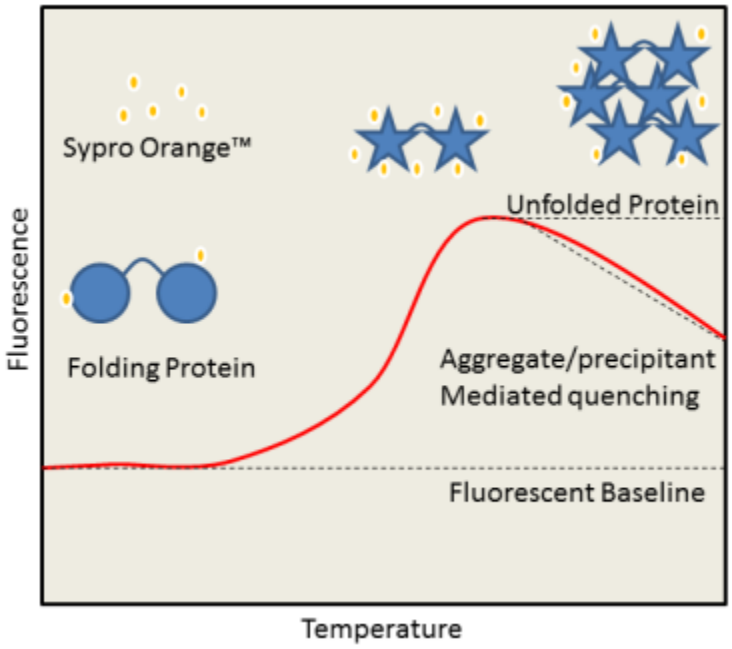


Figure 6.1 Typical features of a DSF curve. Folded protein is incubated with Sypro orange at low temperature and exposed hydrophobic regions dequench the dye to become the baseline. During unfolding a higher proportion of these residues will become exposed leading to increased signal and eventually diminishes as the dye is removed by protein aggregation.

6.2 Aromatic tracking

Aromatic tracking monitors the natural fluorescence of aromatic amino acids local environment and does not require the use of a secondary signalling molecule. Of the four aromatic residues only tryptophan and tyrosine and in rare occasions phenylalanine emit a signal strong enough to be used during analysis. The intrinsic fluorescent properties of aromatic residues have been used since the mid 1900s to track changes between aromatic local environments during protein unfolding [147,148].

During this study only the signal from tryptophan and tyrosines were measured since phenylalanine requires a different excitation wavelength (~240 nm) [149]. The excitation and emission wavelengths of tryptophan and tyrosine are ~280 and 308-355 nm respectively depending on the local environment [150]. This is usually empirically derived through scanning between these wavelengths.

The experimental requirements are the same as DSF, although many commercial real-time PCR machines do not contain the correct fluorescent filters, so cannot be used to achieve an increasing temperature range. To overcome this, protein samples can be pre-heated at various temperatures, spanning the proteins unfolding transition which can then be cooled and tested in a standard fluorimeter capable of measuring at these wavelengths.

Unlike DSF the fluorescent signal need not necessarily increase, as aromatics are sensitive and dependent on the local environment as well as aqueous sequestering [150]. However, one would expect a significant signal change when the protein unfolds, or changes conformation involving segments that contain one or several of these residues.

6.3 Isothermal Titration Calorimetry (ITC)

This technique is regarded as the 'gold-standard' in determining ligand affinity (K_d), entropy ($\triangle S$) and enthalphy ($\triangle H$) values [151–153].

ITC exploits the exchange of heat between the ligand, protein and local environment generated during an interaction. ITC specifically measures the temperature of the local closed environment and determines how much energy (calories) is needed to return the cell's temperature to the original base value. An accurate measurement of this interaction is related to the kinetic parameters in the following relationship (Equation 6.1);

$$-\boldsymbol{RTlnK_d} = \Delta\boldsymbol{H}° - \boldsymbol{T}\Delta\boldsymbol{S}° \qquad \text{(Equation 6.1)}$$

Where $\boldsymbol{R}$ is the universal gas constant (8.314 x 10^{-3} kJ $mol^{-1}K^{-1}$), $\boldsymbol{T}$ represents the time, K_d the single-site binding constant (M), $\Delta H°$ the heat of binding (K) and $\Delta S°$ the entropy (J/K). The K_d parameter is dependent upon the ligand and active macromolecule concentration, which in a 1:1 stoichiometry can be explained by (Equation 6.2);

$$\boldsymbol{K_d} = \mathbf{1}/\frac{[\boldsymbol{MX}]}{[\boldsymbol{M}][\boldsymbol{X}]} \qquad \text{(Equation 6.2)}$$

Where M is the concentration of macromolecule and X the ligand.

A carefully recorded heat exchange through ligand titration can be plotted, typically as calories/S^{-1} plotted Vs time. These peak values can be normalised against the protein concentration to generate an integrated heats plot (kCal/mole of ligand Vs protein concentration). A least squares regression analysis of this data with known molar ratios can give rise to the components required to determine the enthalpy, stoichiometry and binding constants as described in the schematic (Figure 6.2)

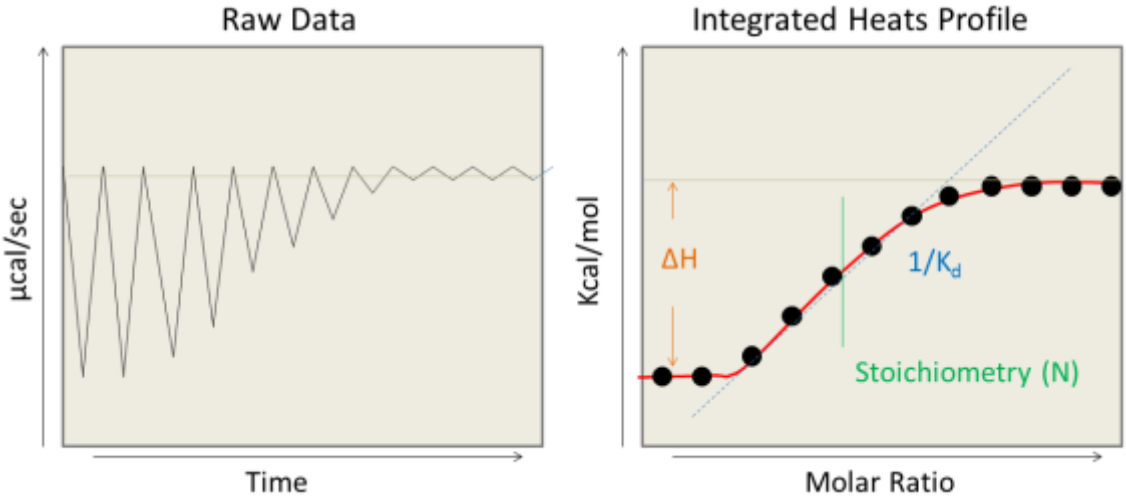


Figure 6.2 ITC heat exchange plots. The left plot depicts the raw data representing a series of individual ligand-protein heat exchanges. The curve measuring this series diminishing signal due to protein saturation is plotted as the integrated heat plot (right) which with known ligand and macromolecular concentrations can yield the ΔH, K_d and stoichiometry from the curves parameters.

Chapter VII

The thermal-stability, half-life and single stranded DNA affinity of TDP-43S mutations

7.1 Rationale for testing mutant mediated stability

The conclusions made from SAXS modelling and CD spectra suggest disease-associated mutations in TDP-43S do not cause a significant change to structure in terms of RRM spatial arrangements, secondary structure or unfolding. Therefore RRM mutant mediated toxicity is unlikely to arise through large structural changes as observed in other neurodegenerative disease related proteins [141,143]. However the nucleic acid binding function was tested to ensure that RRM function is still intact.

Auto-regulatory disease models show onset can occur through an increase in cytosolic TDP-43 and/or loss of nuclear function [42]. The fact many C-terminal disease related mutations correlate with increasing cytosolic levels and half-life supports that *in vivo* protein stability is a real indicator of neurodegenerative toxicity. Interestingly this feature may relate to TDP-43 RRM disease mutations via a gain in structural stability. This has been correlated with an increase in thermal-stability by mutations in pseudo-barnase and lysozyme proteins with enhanced cellular longevity [91–93].

7.2 Thermal-stability of TDP-43S variants by DSF

The thermal-stability of TDP-43S and mutant variants were tested using differential scanning flourimetry (DSF) scanning temperatures between 15-95˚C to cover both folded and unfolded states. Temperature increases were set at a steady 1.2˚C/min^{-1} increments so that a careful recording of temperature related fluorescence could be monitored during protein unfolding. Preliminary experiments found that a final protein concentration of 0.6mg/ml and 10x concentration of Sypro Orange™ dye

were sufficient to monitor TDP-43S unfolding. The blank without protein shows that the dye and buffer doesn't contribute to the fluorescent signal throughout this temperature range (Appendix 5) to show that the fluorescent signal is dependent on protein unfolding.

To increase accuracy, seven repeats were measured for each TDP-43S variant and averaged to include the standard deviation of the data (Figure 7.1). The unfolding transition monitored by the fluorescent signal of TDP-43S represents that of a typical folded protein [154] where a low fluorescent baseline is maintained until the protein begins to expose previously inaccessible hydrophobic residues during thermal-denaturation. The difference in peak heights between TDP-43S variants represent varying degrees of aggregation that typically tend to increase at elevated temperatures, as described [154].

The melting temperatures obtained from the first derivative plot (Figure 7.2) calculated for wtTDP-43S was 50.9 ± 0.2˚C and compared well to similar measurements previously determined by circular dichroism (49.7 ± 0.9˚C) [77]. Interestingly melting temperatures for both disease related mutations increased the T_m to 55.8 ± 0.2˚C and 51.9 ± 0.4˚C respectively for the D169g and K263E mutations (Figure 7.2) showing a gain in thermal-stability.

7.3 Thermal-stability of TDP-43S variants by aromatics fluorescent tracking

To verify this finding melting temperatures were again calculated using an independent method, relying on the natural fluorescence of TDP-43's aromatic residues. The fluorescent emissions from these residues are sensitive to their surrounding environment and as such can be used to track changes to their local environment during thermal-unfolding. The most intense fluorescent signals are achieved through tryptophan and tyrosine as described in chapter VI which are in sufficient quantities to enable the tracking of fluorescent signal (Figure 7.3). These aromatics are predominately located on RRM1 including W113, W173, Y122 and Y155 though RRM2 can contribute through Y214. The excitation and emission

wavelengths used were calculated by scanning the light spectrum between 260-400nm, which peaked at 280 and 340 nm respectively.

The fluorimeter capable of tracking these wavelengths was un-heated, therefore samples were pre-treated for 5 minutes at 2˚C intervals between the 37-65˚C temperature range to cover the unfolding transitions identified by DSF. A total of six repeats were performed for each TDP-43S variant and averaged to obtain the standard deviations (Figure 7.4). From the curve, two features concerning fluorescent signal strength were observed in the form of an initial decline at 45˚C and a rise to above the baseline at 50˚C suggesting that unfolding involves a two-state mechanism. The first state of quenching at ~45°C indicates a change in the aromatics local environment likely due to unfolding or a conformational change. This feature occurs at the same temperature in all profiles to suggest that at least some folding is similar initially.

Because of these two features an accurate sigmoidal curve could not be assigned to the raw data therefore curves were normalised to the trough after the initial decline in signal (Figure 7.5). The data were fitted using a Boltzmann curve to calculate the T_m's as 52.5 ± 0.9°C, 55.8 ± 0.8°C and 58.7 ± 2.8°C with standard deviation respectively for wild-type, K263E and D169G to coincide with the observations from DSF. The finding that both methods independently arrived at the same conclusions provide convincing evidence that TDP-43S disease associated mutants mediate a gain in thermal-stability. Interestingly mutant mediated stability increases of just 2.4˚C in pseudo-barnase led to 14% more protein in the periplasm in prokaryotes [93] to suggest that an increase of up to ~5˚C may have a significant effect on cytosolic levels of TDP-43, assuming a similar trend.

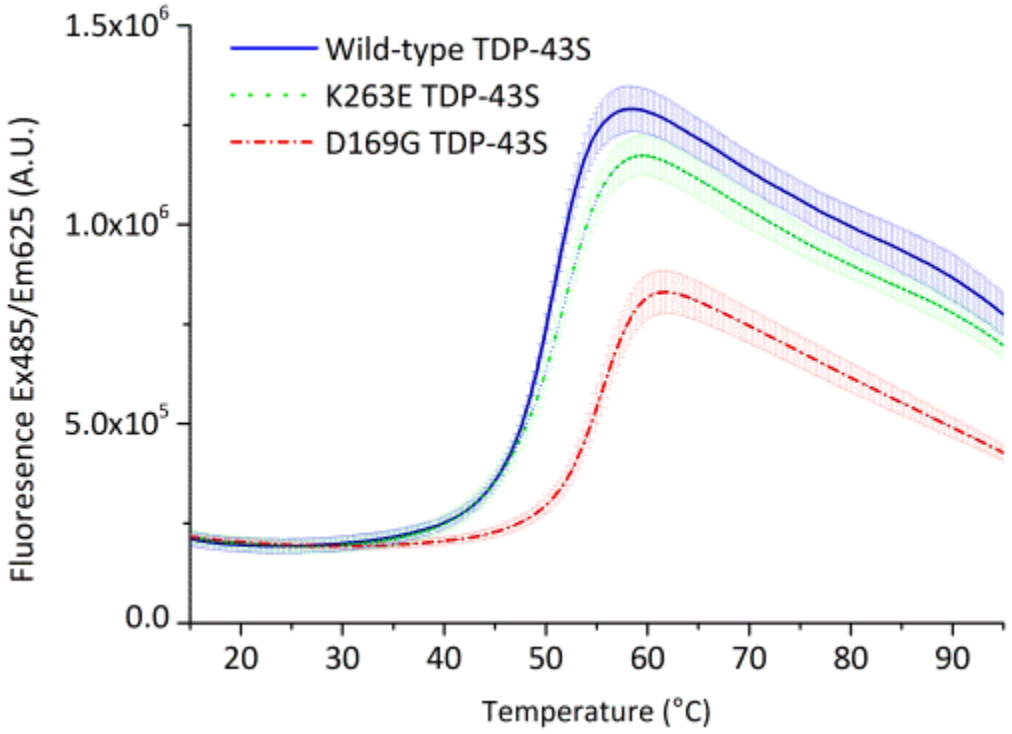


Figure 7.1 Differential scanning fluorimetry of mouse wtTDP-43S and mutant variants. Each construct had an average melting curve plotted from 7 individual runs showing standard deviation.

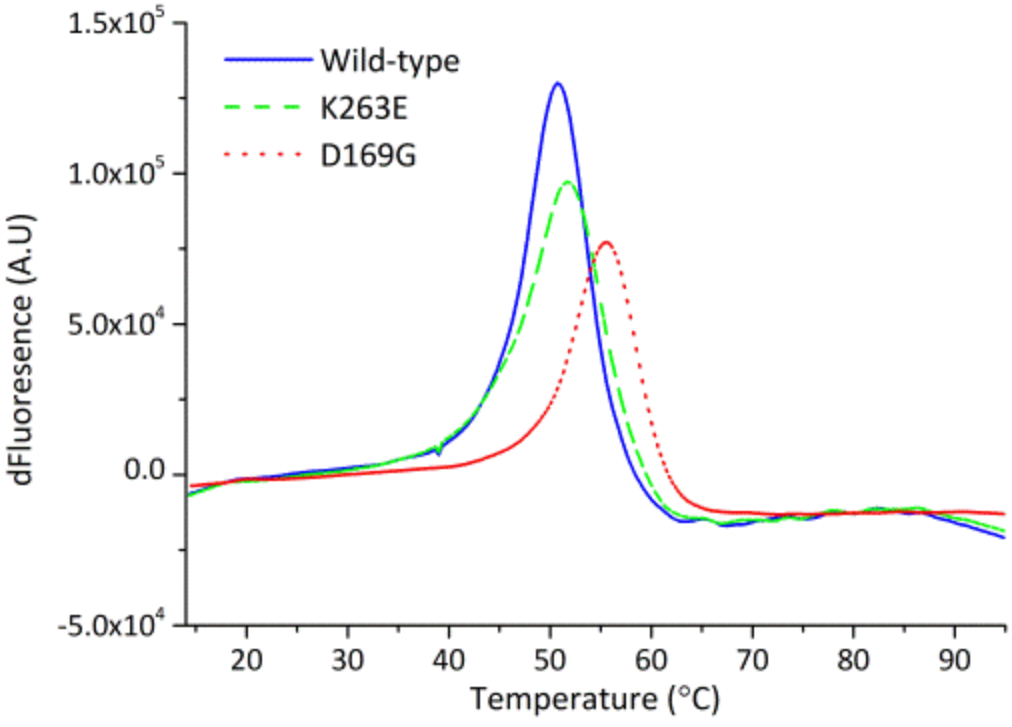


Figure 7.2 TDP-43S variant melting temperatures calculated from first derivatives. The melting temperatures were calculated from the first derivative of the plots of the average melt curves (Figure 7.1). The peak corresponds to the melting temperature of the variant, which calculated a melting temperature of 50.9 ± 0.2

°C, 51.9 ± 0.9 °C and 55.8 ± 0.8 °C for wild-type, K263E and D169G TDP-43S respectively.

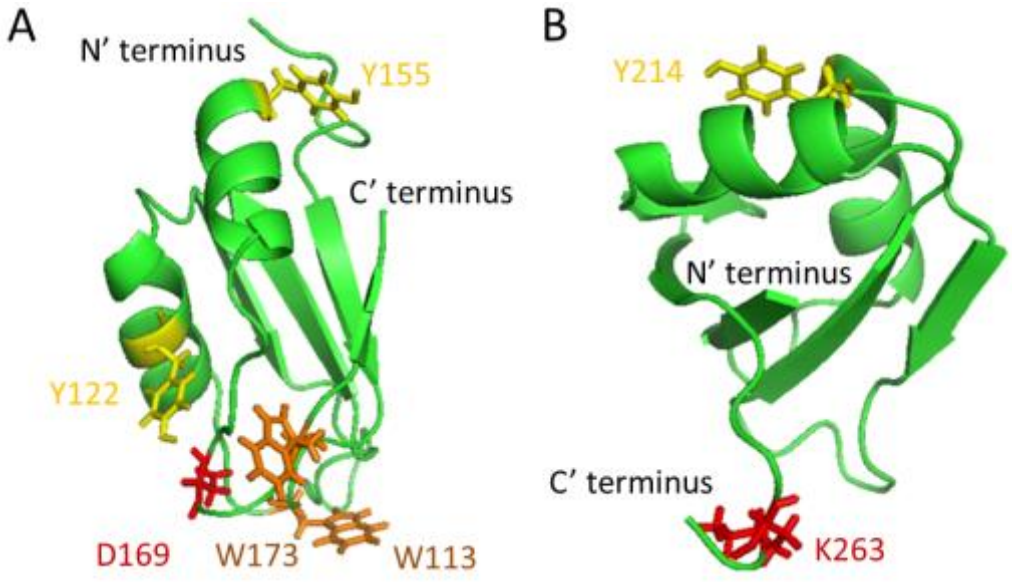


Figure 7.3 Location of aromatic residues within the RRM structures. The SAXS rigid-bodies of **A,** RRM1 and **B,** RRM2 with structural positions of aromatic residues (Orange; Tryptophan and Yellow; Tyrosine) and mutation locations (Red).

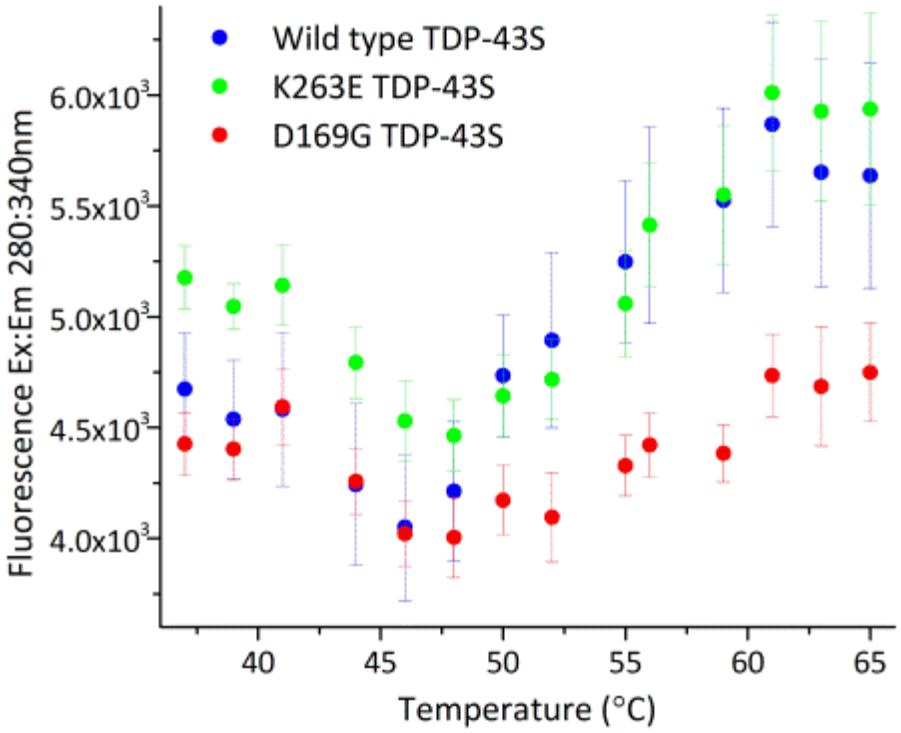


Figure 7.4 Aromatic fluorescent tracking of TDP-43S variants during thermal degradation. The thermal unfolding of TDP-43S variants were traced using the fluorescent signal of aromatic residues, depicted in Figure 7.3. The unfolding

indicated a two state unfolding reaction where aromatics are quenched at 45°C then dequench ~55°C.

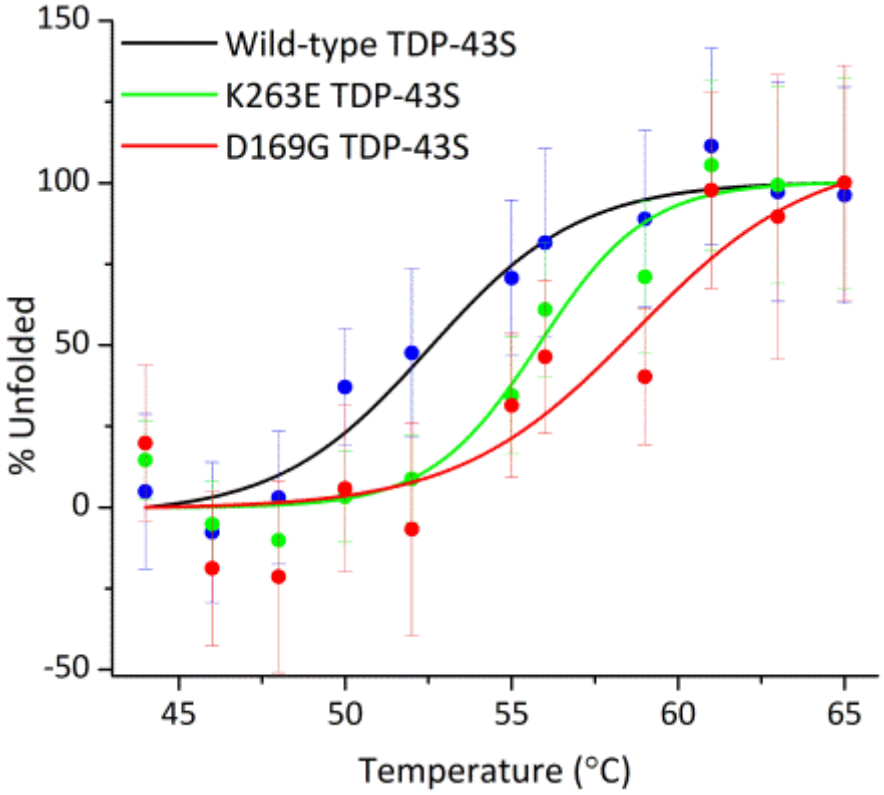


Figure 7.5 Normalised aromatic dequenching profile of TDP-43S variants. The aromatic dequenching curve was normalised to % unfolding, assuming 65°C is 100% and 45°C was 0%, the melting transitions were determined by a Boltzmann fit which calculated 52.5 ± 0.9°C, 55.8 ± 0.8°C and 58.7 ± 2.8°C for wild-type, K263E and D169G respectively.

7.4 *In vivo* half-life of TDP-43 mutations in a Neuro-2a cell-line

Since it was shown that disease related mutations occurring within the TDP-43S region increase thermal-stability, it was important to directly relate this to *in vivo* half-life as suggested in other proteins. This was critical to show that RRM related mutations also mimic the cellular longevity inferred from C-terminal disease mutations [56,57] adding to the growing trend of mutant mediated cellular stability.

To this end a collaboration between members of Yamanaka's group in Japan tested the half-life of native, D169G and K263E within the mouse neuroblastoma Neuro2a

cell-line using full length human TDP-43 (plasmids provided by the MBG, Liverpool). This cell-line was chosen since it has been used previously to model disease related proteins to human neurotoxicity [155–158]. Protein half-lifes were tested using the established pulse-chase methodology as described [56].

Levels of newly synthesised TDP-43 were non-invasively labelled with L-azidohomoalanine (AHA) (the pulse) and diminishing levels of the label were recorded over a 24 hour period using the natural expression of un-labelled TDP-43S (the chase).

Human TDP-43 variant half-lifes were determined as 9.7 ± 0.38, 22.1 ± 1.36 and 18.4 ± 0.93 hours respectively for wild-type, D169G and K263E mutations. The most thermally stable, D169G increased half-life by ~12.4˚C and K263E by ~8.7˚C above the wild-type species. The extent of thermal-stability established in DSF and aromatic fluorescent techniques therefore positively correlated with the longevity in these cells. This is the first example of thermal-stability increases being associated with increased cellular half-life of TDP-43. Thus establishing a possible cause and effect for disease onset.

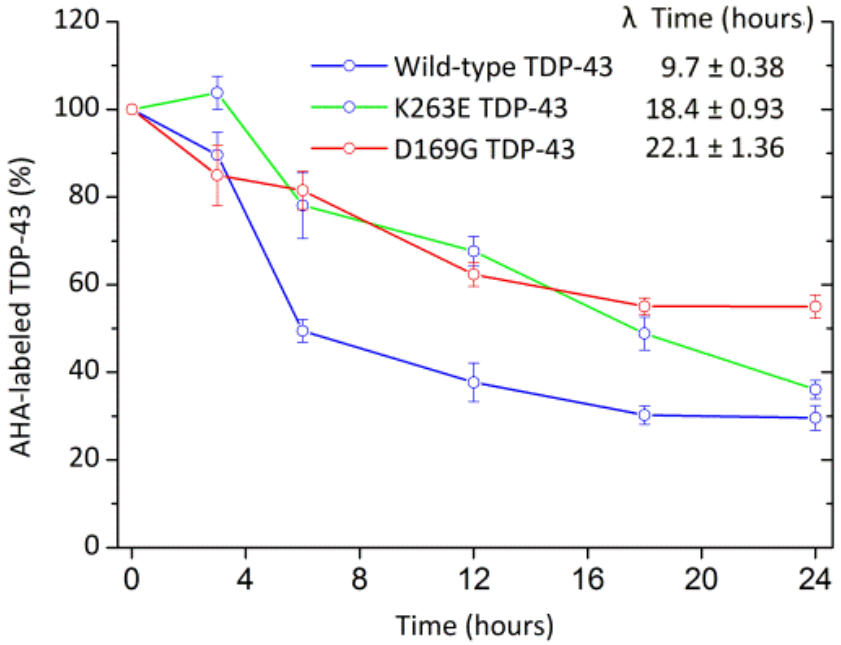


Figure 7.6 Half-life increases of TDP-43S mutations in cultured cells. The half-life of TDP-43 proteins were measured by Professor Yamanka's group, Japan using cultured Neuro-2a cells. The drop in the level of protein (%) correlates to the half-life inside the cell as previously performed [56].

7.5 Is TDP-3S stability associated with aggregation propensity?

The increase in stability identified by DSF coincides with aggregation observed by a reduction in fluorescent peak height (Figure 7.1). This reduction coincided with the extent of thermal stability, with the largest decline in signal occurring from D169G (a mutation which has not been shown to change a function or property of TDP-43). To establish whether mutations affect the aggregation potential of TDP-43S, SAXS methodology was used to determine the radius of gyration by static SAXS measurements at two temperatures. This data was collected at the Barkla laboratory in the University of Liverpool, measurements were averaged from 3 x 20 minute exposures using 3mg/ml protein concentration at a sample-detector distance of 125 cm. The custom made cell for these experiments used 30um Kapton windows and a cell thickness of 1.2 mm. As these were static measurements, protein aggregation cannot be removed before analysis by HPLC and therefore contribute to the analysis, especially in the calculated Rg value. Using the less intense home X-ray source meant proteins had to be exposed for at least 60 minutes to collect a sufficient amount of data. This caused radiation damage and precipitation even at 20˚C. Therefore the addition of 20mM DTT was included to minimise oxidative damage and improve solubility. Data from three 20 minute sessions were pooled and averaged for TDP-43S variants at 20˚C and 40 ˚C and produced the 1D curves observed in Figure 7.7.

The Guinier estimates using the same regions produced Rg values with standard error at 20˚C of 25.8 ± 5.3, 23.3 ± 5.2 and 25.0 ± 3.9Å (Figure 7.8) which rose to 40.1 ± 1.9Å, 25.7 ± 5.5Å and 24.5 ± 5.1Å at 40°C (Figure 7.9) respectively for wild-type, K263E and D169G. The initial static measurement may have included some aggregation leading to a slightly higher Rg compared to the HPLC-coupled data obtained at Soleil. The Rg between mutations are comparable at both temperatures but wild-type increased significantly by ~15 Å which was attributed with an increase in aggregation.

This suggests that stability increases conferred by the disease associated mutations are independent from aggregate formation and are therefore likely mediated by

interactions within the monomeric structure. This finding suggests that mutations do not increase the aggregation potential of TDP-43 and may in fact decrease its likelihood. This supports previous observations shown in rodent models where neuron degeneration is observed without the formation of aggregates or inclusions [50,51] though further work is needed to assess whether this is true of the full length protein in a suitable *in vivo* model.

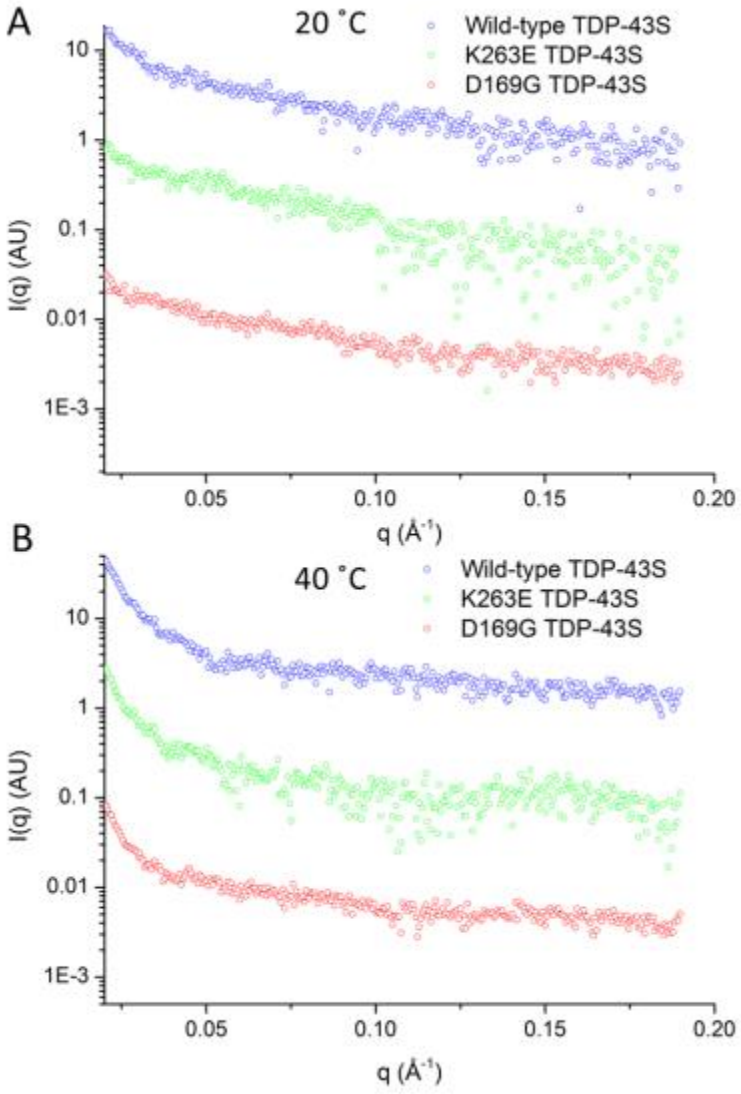


Figure 7.7 Raw 1D SAXS data at 20˚C and 40˚C. The 1D SAXS curves of Wild-type, K263E and D169G. A, Tested at 20 ˚C and B, 40 ˚C using the in-house X-ray source in the Barkla laboratory at the University of Liverpool.

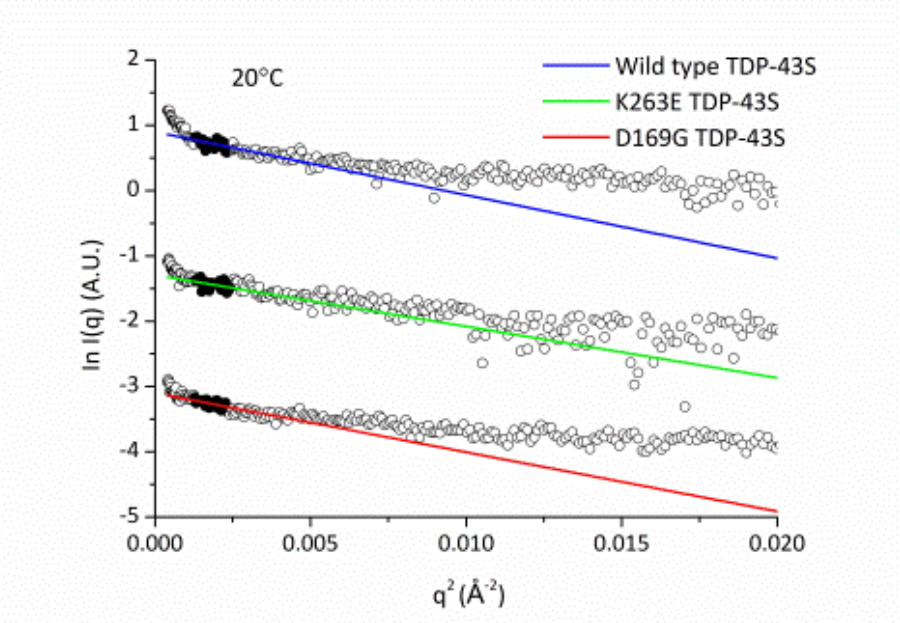


Figure 7.8 Guinier regions of wild-type and mutant TDP-43S constructs at 20°C. The linear Guinier regions were calculated from the filled circles, which were within the recommended qRg region limits <1.3. The Rg was calculated as 25.8 ± 5.3, 23.3 ± 5.2 and 25.0 ± 3.9Å respectfully for wild-type, K263E and D169G. Points used for Guinier region (filled black) were 33-58, 33-61 and 33-58 for wild-type, K263E and D169G respectively.

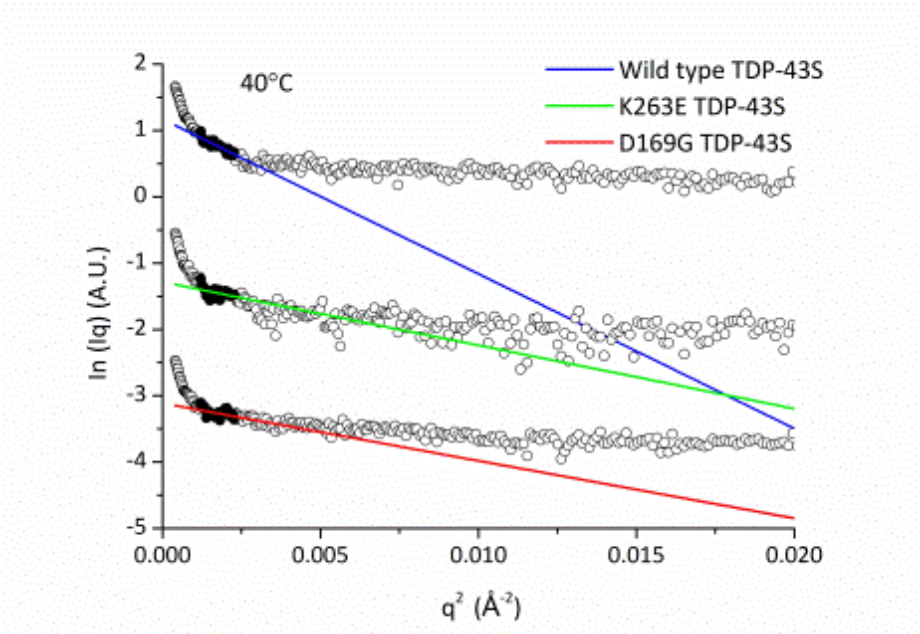


Figure 7.9 Guinier regions of wild-type and mutant TDP-43S constructs at 40°C. The linear Guinier regions were calculated from the filled circles, which were within the recommended qRg region of <1.3, except wild-type which was <2.0. The Rg was calculated as 40.1 ± 1.9Å, 25.7 ± 5.5Å and 24.5 ± 5.1Å at 40°C respectfully for wild-type, K263E and D169G. Points used for Guinier region (filled black) were 31-58, 31-58 and 31-58 for wild-type, K263E and D169G respectively.

7.6 Disease mutations and DNA affinity

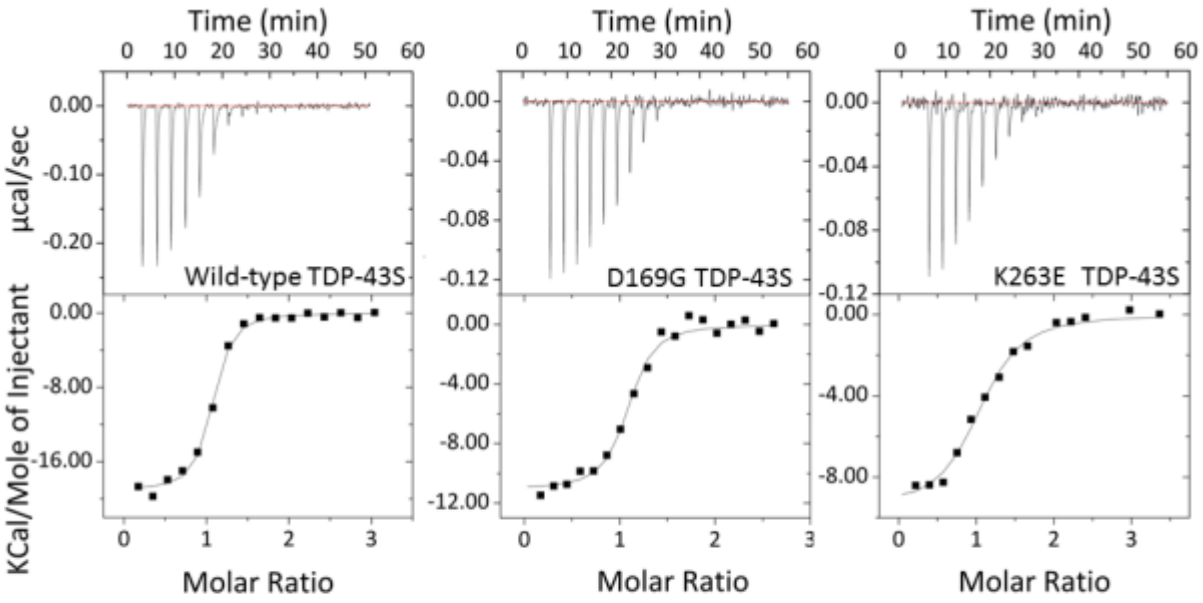
In the proposed 'gain of auto-regulation' model by Lees *et. al.* 2012, loss of nuclear function is attributed to neurotoxicity. a contributor to disease onset is caused through a loss of nuclear TDP-43 function. Since each described role for TDP-43 requires nucleic acid binding, a reduction in RNA/DNA binding affinity is likely to affect its regulatory, splicing and transport function which could relate to neuro degeneration by inhibiting function [42]. Therefore to determine whether TDP-43S mutations can affect TDP-43 nucleic acid binding its affinity to single stranded DNA was calculated and compared against disease mutations by isothermal titration calorimetry as described in chapter VI.

The stoichiometry of TDP-43 has already been suggested as 1:1 with biotinylated $(TG)_6$ using a high through-put non radiometric assay [88]. Gel filtration with wtTDP-43S and excess $ss(TG)_8$ also only eluted as one protein peak to suggest that a single TDP-43S protein is involved in the interaction (Appendix 6). Because functional TDP-43 levels could be lost due to aggregation (observed in the cell after experimentation) this stoichiometry was used to determine the amount of functional protein and calculate its affinity. These values could also be compared for accuracy against previously determined K_d affinity values for TDP-43S determined for similar $ss(TG)_6$ using nitrocellulose membranes [77]. Since single stranded DNA is highly soluble its concentration was kept constant throughout the analysis

7.7 TDP-43S ITC affinity parameters

The ITC experiment was analysed using Origin 8.6 software integrated with the MicroCal^{200} ITC instrument. Curve fitting could only converge with the integrated heats plot when using a single binding site curve (Figure 7.10) to indicate a single binding site for TDP-43S. The concentration of ligand was 60 μM and the estimated protein concentrations for wild-type, D169G and K263E were calculated as 4.2 μM, 2.6 μM and 2.0 μM respectively to reflect the 1:1 stoichiometry.

The ITC results calculated the K_d affinity to $(TG)_8$ ssDNA as 76.5 ± 14 nM, 105 ± 29 nM and 467 ± 98 nM for wild-type, D169G and K263E TDP-43S respectively. The values for wild-type and D169G were comparable to the affinity previously calculated using nitro cellulose assays at K_d 89.3 ± 19.3 nM to $(TG)_6$ ssDNA [77]. In contrast, the K263E mutant bound fourfold weaker to suggest that this mutation has a detrimental effect on the binding to ssDNA. This may be due to an interaction occuring at or near to the nucleic acid binding site, as proposed during SAXS analysis of the protein (chapter V).



TG8 Ligand binding parameters determined for TDP-43S variants

TDP-43S Variant	Stoichiometry	Kd (nM)	△H (cal/mol)	△S (cal/mol/deg)
Wild-type TDP-43S	1.00 ± 0.019	76.5 ± 14	$-1.908E^4$	-31.4
D169G TDP-43S	1.04 ± 0.019	105 ± 29	$-1.109E^4$	-5.23
K263E TDP-43S	1.03 ± 0.029	467 ± 98	$-9.554E^3$	-3.07

Figure 7.10 Isothermal titration calorimetry results of variant TDP-43S. The ITC results show the values of K_d affinity for wtTDP-43S (left), d169gTDP-43S (middle) and k263eTDP-43S (right). The K_d values illustrated in the table below show that 'wtTDP-43S' and 'd169GTDP-43s' affinities are comparable whilst the 'k263eTDP-43S' mutation is considerably higher which may highlight ligand binding dysfunction.

7.8 Neuro-degenerative RRM disease mutations increase structural stability and cellular half-life

Two independent thermal stability methods demonstrate TDP-43 RRM associated disease mutations increase TDP-43S thermal stability by up to 5°C that coincide with cellular longevity similar to observations determined in other protein species [91–93]. This was shown to be mediated without enhancing the constructs aggregation propensity suggesting that the stability is inferred from interactions occurring within the monomeric subunit. The FTLD-U mutation K263E also perturbed TDP-43S' affinity to single stranded TG repeated DNA when a 1:1 stoichiometry was assumed provided two properties that may enhance toxicity during disease.

The ALS associated mutant D169G had the largest resistance to thermal unfolding. Asp169 is located on loop 6 of the RRM1 structure separating α-helix 2 and β-strand 4 with potential hydrogen bond interactions with the side chains of Lys114 and Thr115 as represented in Chapter II, Figure 2.2. The substitution of acidic aspartate for non-polar glycine at this position would abrogate these bonds but may facilitate a tighter association of loop 6 with the hydrophobic inner core and orchestrate the observed increase in stability. This may be orchestrated by the closest hydrophobic residue, L120, situated within the first α-helix, which is separated by 8.8 Å from G169 in the human NMR structure of RRM1 (Figure 7.11). This residue is poised towards the hydrophobic core of RRM1 and may provide an anchor to stabilise the flexible loop where G169 is situated to confer the stability increase. This change is similar to the substitution occurring within the human Pin1 WW domain where the N30G mutant situated on a connective loop increased the thermal stability of the domain by a comparable 6.4 °C. This was thought to be due to an improved loop formation [159] which is similar to our explanation for D169G stability increases in TDP-43. It is possible that this interaction is formed during partial unfolding shown by the aromatic trace at 45˚C which may arise through movement of W173 and W113 that are close to the D169G position in the NMR RRM1 structure (PDB id; 2CQG) (Figure 7.3).

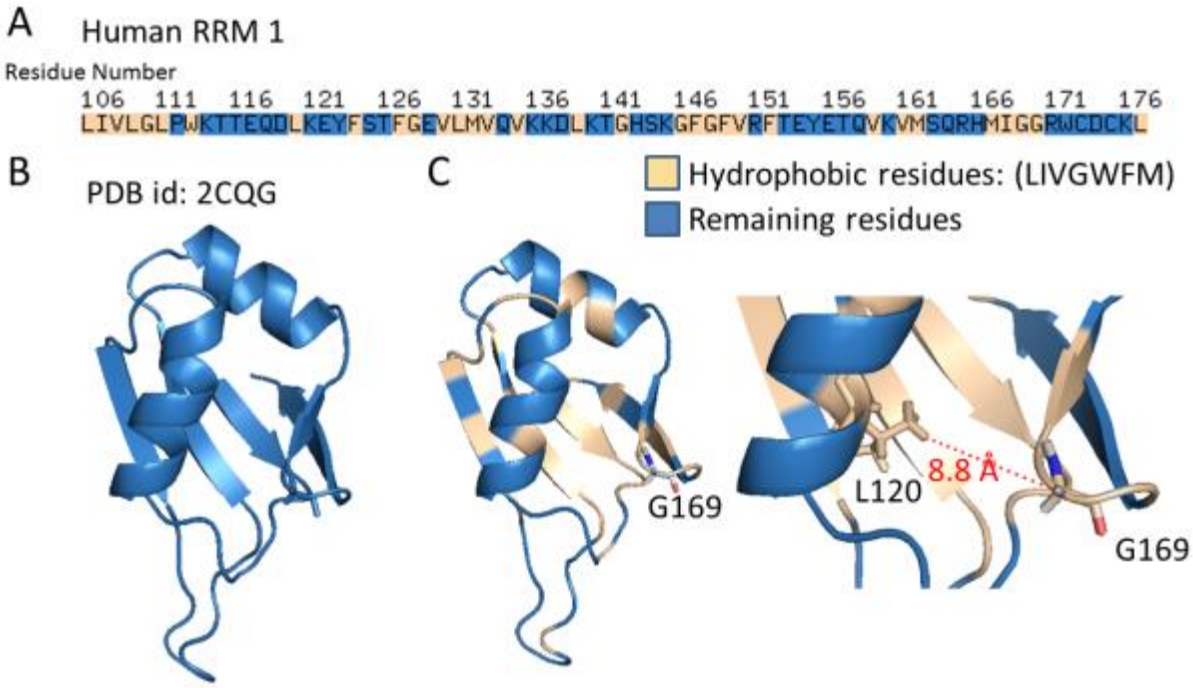


Figure 7.11 Additional interactions which may form due to the D169G mutation. The D169G mutation and other hydrophobic residues within the RRM1 structure. **A,** Schematic of the hydrophobic residues within RRM1. **B,** The PDB structure of Human RRM1 solved by NMR and **C,** the positions of the hydrophobic residues and the closest residue situated in the core of RRM1 to G169. Performed using PyMol (http://pymol.com).

The K263E mutant increased thermal stability by ~1-3°C and is close to the nucleic acid binding site. The polarity change from positive Lys to negative Glu may aid in forming electrostatic interactions to local positive residues located near to the nucleic acid binding interface restricting nucleic acid access and increasing stability by increasing the number of bonds (Figure 7.12). The K263 residue is also involved with forming direct interactions when both RRM domains interact with ssRNA as shown in a bound NMR structure of human TDP-43S [160]. This suggests that mutating this residue is therefore likely to affect nucleic acid affinity by reducing the amount of bonds able to form between TDP-43S and nucleic acids. However as this residue is only 1/25 of the total residues involved in this interaction and 1/9 of those occuring in the RRM2. It seems strange that removing this one interaction would cause such a large change (four-fold difference) to affinity. It was

demonstrated that mutating Lys 263 to Ala only caused a minor change to affinity (using GUGUGAAUGAAU RNA) from a K_d of 22.6 ± 4.3 to 35.6 ± 7.9 nm [160]. Therefore a mechanism whereby the binding pocket is partially blocked would provide a better explanation for these differences.

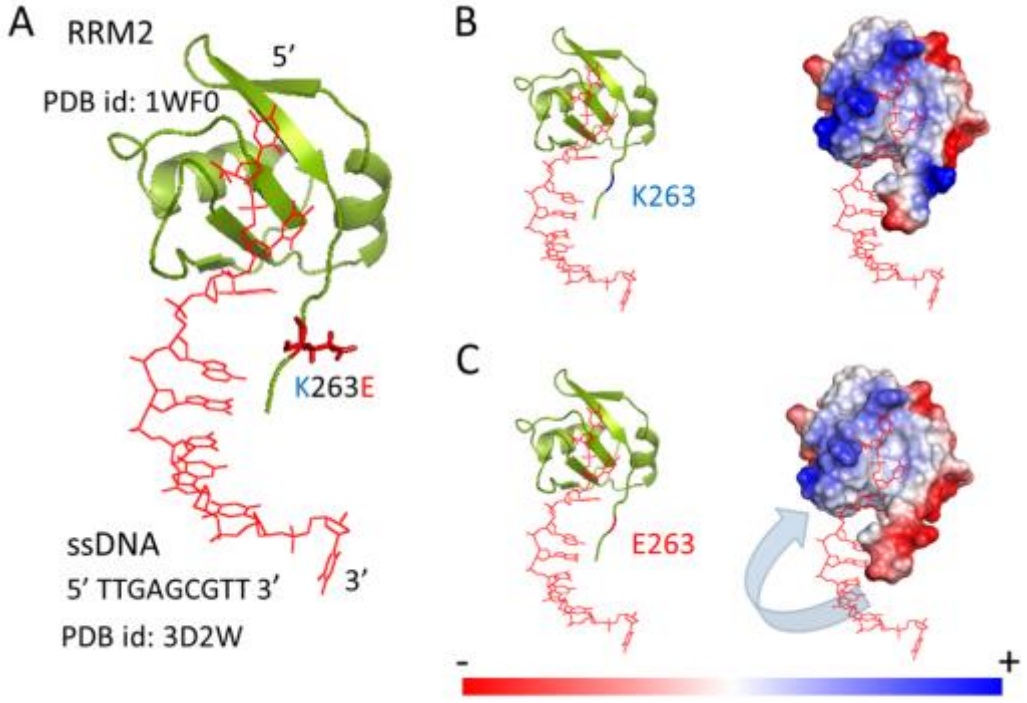


Figure 7.12 Location of K263E mutation and possible charge effects. The NMR structure of RRM2, truncated to residues 193-265 (PDB id: 1WF0), was aligned with the ssDNA from mouse RRM2 (PDB id: 3D2W). This shows **A,** the location of E263, which is situated on a flexible region close to the DNA binding region. The charge affects using vacuum electrostatics shows that **B,** the K263 residue confers a positive charge that reverses to negative when mutated to E263 as shown in **C**. This may form polar interactions across the binding pocket to restrict DNA access.

This is likely due to E263 forming an interaction directly with the RNP regions of the β-core or surrounding residues to abrogate DNA contacts or restrict ligand access. This extra interaction may also confer this slight increase in thermal-stability observed during the thermal-stability assays. However high resolution techniques

are required to accurately locate this interaction and deduce how unfolding and DNA binding are restricted.

Recent studies using primary rat neurons correlated disease symptoms best with an increase in cytoplasmic TDP-43 [46]. The results here suggest that D169G and K263E like C-terminal mutations may cause an increase in TDP-43 cytosolic levels that are correlated with disease onset. Here it is shown for the first time that structural stability concerning the RRM-RRM region could be a potential cause for increasing cellular half-life. The finding that mutations outside of the C-terminal domain confer cellular stability suggests that onset can occur independent of function. This is supported by the limited amount of RRM-RRM mutations related to disease and the finding that D169G did not cause a change to structure or binding affinity. The K263E mutation may cause onset through a perturbed DNA/RNA binding function however this is likely to be the only mutation that this applies to, since negative residue substitutions are rare, only observed in 4/40 identified mutant point-mutation locations and all are situated away from the RRM regions to suggest that their involvement with RRM function is unlikely.

Stability increases have been associated with a gain in resistance to ubiquitinated degradation. The fact neurodegeneration can be associated with impaired components of the ubiquination degradation pathways [72,73] supports this hypothesis. Therefore a build-up of cytosolic protein could be commensurate of toxicity. Simple stability increases may provide added pressure and overwhelm the degradation machinery causing disease onset perhaps as described by the loss of auto-regulation model (chapter I, Figure 1.1). Providing a common path to describe disease onset conferred by both RRM and C-terminal disease related mutations. The fact that these regions perform different functions in the protein also detract from a loss of function as a common trend to describe disease onset.

Chapter VIII

DNA Binding Protein in Starved Cells (DPS) and DNA protection during stress.

8.1 Introduction to DPS

The *E.coli* stress response is reliant on protein cascades that are regulated by a few stress related signalling factors e.g. RpoS and OxyR that regulate multi-functional proteins to sequester potentially harmful substances during stress. Prokaryotes dedicate a significant amount of energy into protecting susceptible biologically important molecules such as DNA by restricting access to mutagens, nuclease cleavages and oxidative damage. In prokaryotes this is commonly facilitated by the 'DNA binding protein in starved cells' commonly known as DPS identified in *E.coli* by Almiron in 1992 [161]. The protein is over-expressed during a variety of stress related pathways and forms large dodecameric complexes, able to sequester reactive oxygen species using catalase and ferritin-like activity.

8.2 RpoS and OxyR *E.coli* stress response

Prokaryotes are well adapted for efficient stress responses that are often encompassing to confer resistance to a variety of stress causes. For example, the stress response to starvation also provides resistance to oxidative, elevated temperature and pH stressors [162]. This is partly due to the multi-functional proteins that are expressed during stress and the specificity of the stress specific sigma factors that regulate them. The stress responses in *E.coli* are predominately orchestrated by the stationary/stress related sigma factor 'RNA polymerase sigma S' (RpoS) or 'oxidative stress regulator' (OxyR) responsible for DPS expression during starvation and oxidative damage respectively [163]. These paths are mutually exclusive [163,164] and regulate by direct and indirect methods a large proportion of the *E.*coli's genome with RpoS alone estimated to regulate ~10% [165]. The expression of these activators are raised in short regulatory paths [166]

to respond rapidly to environmental changes by stress sensing signalling molecules for example guanosine tetraphosphate ((p)ppGpp) responsible for activating RpoS expression during carbon shortages [162,165].

8.3 Importance of DNA binding protein in starved cells

E.coli DPS has homologues in many prokaryotic systems. The closest conserved to *E.coli* with known structure are depicted in Figure 8.1. DPS is integral to the stress response as demonstrated by the correlation between cellular survival and the knock-down of functional DPS during oxidative insult [161]. During stress DPS is the most abundantly expressed nucleoid associated protein, which reflects the extent prokaryotes rely on the protein during potentially toxic conditions [167]. This is unsurprising, considering the range of stress limiting functions and the conservation of DPS between prokaryote and archaea species [168]. These functions aim to remove potentially dangerous species, shield vulnerable nucleic acid targets and may regulate some protein expression.

8.4 DPS function during DNA protection

DPS is one of the main driving forces of DNA protection in *E. coli* and indiscriminately protects DNA from nuclease access by binding and condensing DNA within DPS aggregates [161]. DPS-DNA complexes are highly stable and able to resist heat-shock of >100°C and prolonged heating at ~65°C [161]. The mode of this protection is orchestrated by the self-aggregation propensity of DPS enhanced in the presence of DNA [169,170]. The mode of DPS self-aggregation is thought to be largely mediated by these interactions, since DNA is bound before DPS dodecamers aggregate, shown by atomic force microscopy [169]. This self-aggregation eventually form crystal plates [170] which has led to several solved DPS crystallographic structures in a variety of prokaryote species (Figure 8.2).

The DNA binding is largely dependent on the positively charged N-terminal lysines (Figure. 8.1) [169]. The removal of lys5, lys8 and lys10 by truncating the first 18

residues abolishes the DNA binding capacity, as shown in *E.coli*. The absence of this extended region generally coincides with a loss of DNA binding as observed for DPS homologs: Dlp-1, Dlp-2 and Ferritin from *B. anthraci and L. innocua,* respectively [171,172]. The structure of non-DNA binding DPS are structurally conserved to that in *E.coli,* except for a truncated N-terminal (Figure 8.2). This suggests that binding is not facilitated by surface structures or charge, as previously described for non-DNA binding DPS in *A. tumefaciens* [173].

Interestingly DNA binding can occur in rare cases without the N-terminal as shown through *M. smegmatis* and *H. pylori* where DPS homologues bind DNA through different regions. The *M. smegmatis* DPS binds DNA using similarly spaced lysines on its extended C-terminal [174] whereas the DPS homolog 'Neutrophil-activating protein' from *H. pylori* (HP-NAP) has adapted to bind DNA through rare positively charged regions on the surface of its dodecamer [175].

Though the N-terminal is commonly required to bind DNA it isn't necessary for cell survival, as shown through oxidative toxicity between native and N-terminally cleaved *E. coli* DPS in both an *in vitro* and *in vivo* setting [169]. The catalase and ferroxidase like roles of DPS are therefore sufficient in alleviating the danger from reactive oxygen. It is worth note that shielding DNA may protect the cell from prolonged stress and other stress inducing factors such as temperature, since the DNA:DPS complex is especially stable (up to 100 ˚C) [161].

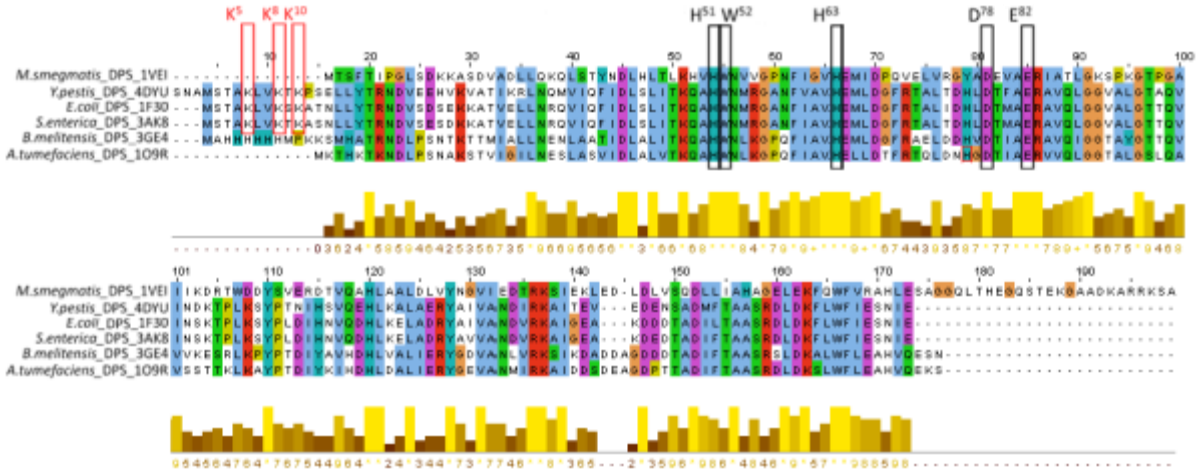


Figure 8.1 DPS sequence homology alignment between prokaryotes. A selection of PDB deposited DPS proteins with highest sequence similarity to *E. coli.* The alignment was performed by ClustalOmega [84] using its colour code (Appendix 4). The level of conservation is also illustrated in the bar ranging from fully conserved (*) to unconserved (-). The three N-terminal lysines important in DPS:DNA complex assembly are shown in the red box [169]. The black box shows residues important in the ferroxidase-like centres [176]. Prokaryotes sequences used are from the following organisms; *Myobacterium smegmatis, Yersinia pestis, Escherichia coli, Salmonella enterica, Brucella melitensis* and *Agrobacterium tumefaciens.*

8.5 Iron store & ferritin-like activity

DPS is adapted to facilitate an alternate pathway involving iron and ferroxidase-like sites to avoid the accumulation of mutagenic reactive oxygen species. Iron is an important element in prokaryotes used by proteins to fulfil a variety of important biological processes e.g. for generating energy in the Tricarboxylic Acid Cycle (TCA) [177] and DNA synthesis [178]. Unfortunately, excess free iron can undergo uncontrolled oxidation, when coupled to elevated levels of H_2O_2 iron can produce $OH\cdot$ radicals through the fenton reaction (Equation 8.1) to cause oxidative damage to susceptible DNA and proteins.

To avoid this, organisms produce ferritin which through ferroxidase activity facilitates the safe oxidation of Fe^{2+} to Fe^{3+} by molecular oxygen to avoid the accumulation of free radicals as described (equation 8.2) [179]. The DPS dodecamer facilitates this role but favours H_2O_2 instead of O_2 which not only removes oxidised

iron but also diminishes total H_2O_2 levels (Equation 8.3.) [180,181]. The preference of H_2O_2 by DPS as an oxidant owes to its ferritin-like description.

The 24 Fe^{3+} loaded DPS can also sequester H_2O_2 into harmless hydrogen and oxygen meaning that DPS is also described as having a catalase-like activity [180] (Equation 8.4). To provide a dual-function to avoid the accumulation of reactive oxygen species. *E.coli* DPS can store up to ~400 iron ions within its cavity which can be mobilised when reduced like iron in ferritin [182]. Iron storage is largely fulfilled by ferritins with bacterial ferritinA alone storing 50% of total iron during logarithmic and early stationary phase [183]. However it is worth note that given DPS' abundance during stress a significant amount of iron could be stored by DPS and under prolonged stress this responsibility may be more reliant in cells.

Fenton Reaction

$Fe^{2+}+H_2O_2 \rightarrow Fe^{3+}+OH^- + OH\cdot$ (Equation 8.1)

Common ferritin ferroxidase activity

$4Fe^{2+} + 4\,H^+ + O_2 \rightarrow 4Fe3 + 2H_2O$ (Equation 8.2)

DPS ferritin-like activity

$2Fe^{2+} + H_2O_2 + 2H_2O \rightarrow 2FeOOH\text{-}P + 4H^+$ (Equation 8.3)

DPS catalase-like activity

$2H_2O_2 \rightarrow 2H_2O + O_2$ (Equation 8.4)

8.6 Gene regulation

DPS has also been postulated to control protein expression. Evidence of this has not advanced much since DPSs' first discovery in *E.coli* [161] where null mutants had an altered 2D gel electrophoretic expressed protein profile compared to the wild type

strain. Since the only difference was the knock-down of DPS the authors suggest that DPS may in-directly regulate other proteins. In a 2011 review, authors suggest that DPS may regulate expression given its similar pleiotropic profile to histone-like proteins when knocked down [184] which have been attributed with a regulatory role in gene expression. However without a direct link this role for DPS is still under debate.

8.7 DPS structure

The multi-functions of DPS are orchestrated by its structural adaptions. The *dps* gene encodes a protein of 167 amino acids (18.7 kDa) in *E.coli*. The protein monomer consists of four anti-parallel α-helices that interact to form a functional dodecamer structure consisting of 12 DPS monomers arranged as a hollow sphere spanning ~90Å in diameter and ~40Å at its cavity. This arrangement including secondary structure is conserved between prokaryotes (Figure 8.2). The cavity acts as the iron store and contains the ferroxidase centres which are generated through interacting monomers (Figure 8.3) [180,181]. This cavity is accessible to the surroundings by two pore types each with three-fold symmetry. The N-terminal interface created by DPS monomers form the ferritin-like pore used for iron access and lined with negative residues to facilitate iron transfer [176]. The other pore is generated from monomeric C-terminals and is unique to DPS, this pore has not yet been attributed to any function and is highly variable between different organisms. The internal cavity has conserved residues at H51, H63, D78, E82 and W52 near the ferroxidase centre that are important in forming interactions with ferrous iron [176]. A total of 24 ferrous iron sites exist within the cavity constituting 12 dinuclear ferroxidase centres, containing both high and low affinity iron sites (Figure 8.2).

Except for the DPS homolog neutrophil-activating protein from *H. pylori* (HP-NAP), the DPS DNA binding function is also unique as the surface of the complex bears no resemblance to any known DNA/RNA binding motifs and is predominantly negative or neutral, apart from the flexible N-terminal regions, as judged through vacuum electrostatics (Figure 8.3). The current hypothesis is that DNA is threaded through holes generated between interacting DPS protein complexes during self-

aggregation [185]. These gaps are lined with the extended N-terminal ends from interacting dodecamers each contributing nine functionally important lysines to create a positive pocket to interact with DNA [169,185]. The indiscriminate interaction of DNA is thought to be achieved through Mg^{2+} bridging coupled with the N-terminal lysines, as judged through co-crystallisation studies with varying concentrations of Mg^{2+} and EDTA [169,186]. It is worth note that the N-terminals in all DNA binding DPS complexes are non-visible in the electronic density which suggests that they are flexible.

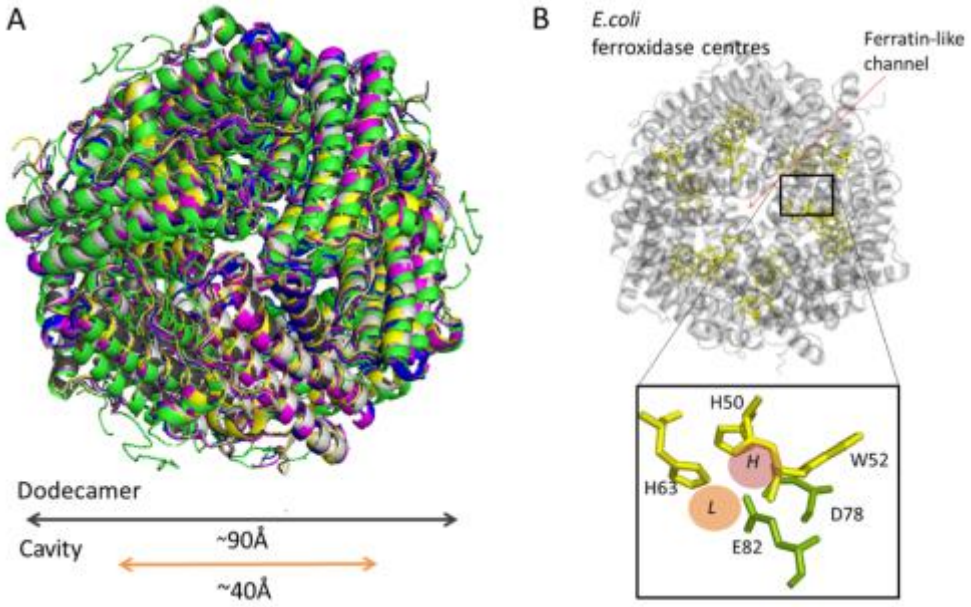


Figure 8.2 Structual conservation of DPS and ferroxidase-like sites from *E.coli* DPS. A, The prokaryote DPS dodecamers are structurally conserved for; *M. smegmatis* (green, PDB id: 1VE1), *Y. pestis* (white, PDB id: 4DYU), *E.coli* (black, PDB id: 1F30), *S. enterica* (yellow, PDB id: 3AK8), *B. melitensis* (blue, PDB id: 3GE4) and *A. temefaciens* (purple, PDB id: 109R). **B,** The ferroxidase centres of *E.coli* showing the high affinity (red) and low affinity (orange) iron binding sites generated from two monomers shown in (yellow) and (light green).

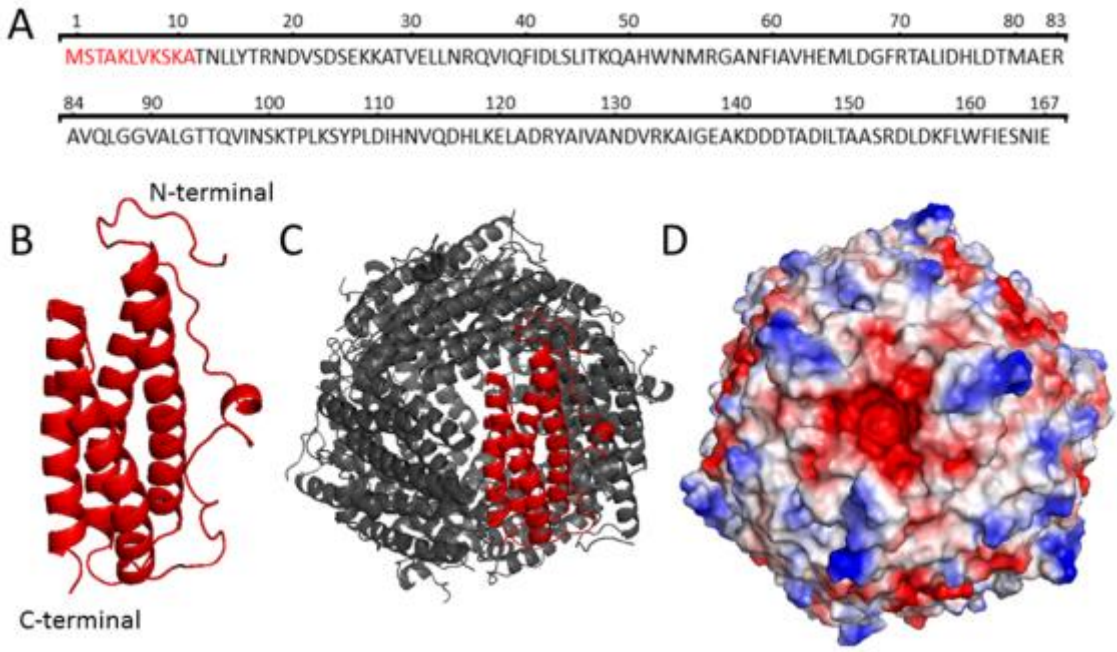


Figure 8.3 Sequence, structure and vacuum electrostatics of *E.coli* DPS. A, The full *E.coli* DPS sequence and residues un-resolved in the structure (red). **B,** The DPS monomer (PDB id: 1F30) . **C,** The DPS dodecamer complex with the flexible N-terminal exposed to the surface. **D,** The vacuum electrostatic model representing, negative (red),neutral (white) and positive (blue) from the remaining N-terminal.

8.8 Elucidating the DPS:DNA crystal complex and neighbouring dodecamer contacts

The structure of DPS has been determined through X-ray crystallography methods for *E.coli* and a range of prokaryotic organism figures as illustrated in Figure 8.2. At present there is no structure which shows the DNA interaction region of DPS, since the N-terminal residues are non-visible in all DNA binding structures.

A high resolution DNA protein complex is a worthy pursuit to prove that DNA interactions are mediated by lysine and Mg^{2+} bridging as suggested in the literature [169,186]. The study showing that the lysines confer the DNA interaction and self-aggregation did so by either removing the first 8 or 18 amino acids which also exclude other potentially important residues under the assumption that they had no affect to binding.

In conjunction with obtaining a structure of the DPS:DNA complex this study will also assess the interactions to neighbouring *E.coli* DPS dodecamers to assess whether these residues can make direct polar interactions with neighbouring dodecamers within the crystal structure. Interestingly other non-DNA binding species of DPS also adopt the hexagonal crystal packing to imply that DPS self-aggregation may not be entirely mediated by DNA binding. Therefore, through comparing the structures of the most conserved DPS species to *E.coli* (Figure 8.1) it is possible to search for conserved inter-protein interactions and to determine whether polar contacts could have any significance during self-aggregation.

These goals may be of interest to develop targets suitable to disrupt the effectiveness of bacterial stress responses to enhance current bacterial infections that tend to be growing in resistance on a seemingly yearly basis. DPS has already been shown to confer resistance to bacterial treatments and has been considered as a drug target [187]. Given the similarities between DPS sequences and structure in prokaryotes, these drugs could have the potential to treat a wide variety of bacterial infection. Disrupting just one function of DPS should make organisms more susceptible to new or old treatments e.g. penicillin.

Chapter IX

Expression, purification and preliminary analysis of DPS structure

9.1 Over-expression of DPS

To over-express DPS two different plasmids were constructed one containing an N-terminal 6xHis tag and one without any additional residues. The 6xHIS DPS construct contained an extra 2.2 kDa incorporating the sequence MAHHHHHHSSGLEVLFQ*GP with a 3C protease cleavage site that could cleave after glutamine illustrated by (*). The rationale to include this tag was to reduce the number of chromatography steps during purification and since the tag was relatively small it wasn't anticipated to disrupt the DPS dodecamer assembly. Alongside this construct a plasmid containing untagged DPS was also made to produce protein identical to that described by Almiron in 1992 [161].

The main difference between the Almiron procedure and this study was the type of vector used to over-express DPS in *E.coli*. Almiron used a plasmid adapted to control DPS expression by stress related Sigma factors which activate during nutrient deprivation during early stationary-phase. As DPS is only normally expressed under stress in prokaryotes the advantage is DPS is less likely to be removed by bacterial degradation machinery. During Almirons' protocol *E. coli* was shown to over-express DPS in abundance when the bacteria entered the exponential stage of its growth cycle [161].

During this study protein expression for both DPS plasmids were under the control of the lac operon active in the presence of IPTG as used previously for TDP-43S, chapter III. This methodology aimed to reduce the total time required to over-express DPS, since starvation was not required, however complications could arise from expressing DPS during an irregular period in the *E.coli* life-cycle. Therefore an important preliminary experiment was to first establish whether IPTG could induce detectable amounts of DPS and whether this was soluble.

9.2 IPTG concentration and incubation time

IPTG was utilised to induce expression of both DPS constructs during the study using different concentrations (0.1-2.0 mM) and IPTG incubation times (3h,5h & overnight) initially on the N-terminally 6xHIS tagged DPS construct to probe the minimal levels required to detect DPS over-expression and accumulation of inclusion-bodies [188](Figure 9.1). DPS could over-express in all IPTG concentrations between the 15-20 kDa markers on an SDS-PAGE verified by an in-gel trypsin digest and fingerprint using MALDI-TOF mass spectrometry (Figure 9.2). The untagged construct also over-expressed using a low IPTG concentration (0.4 mM) over a relatively short time (5h) to provide two options for DPS overexpression and purification (Figure 9.1, B).

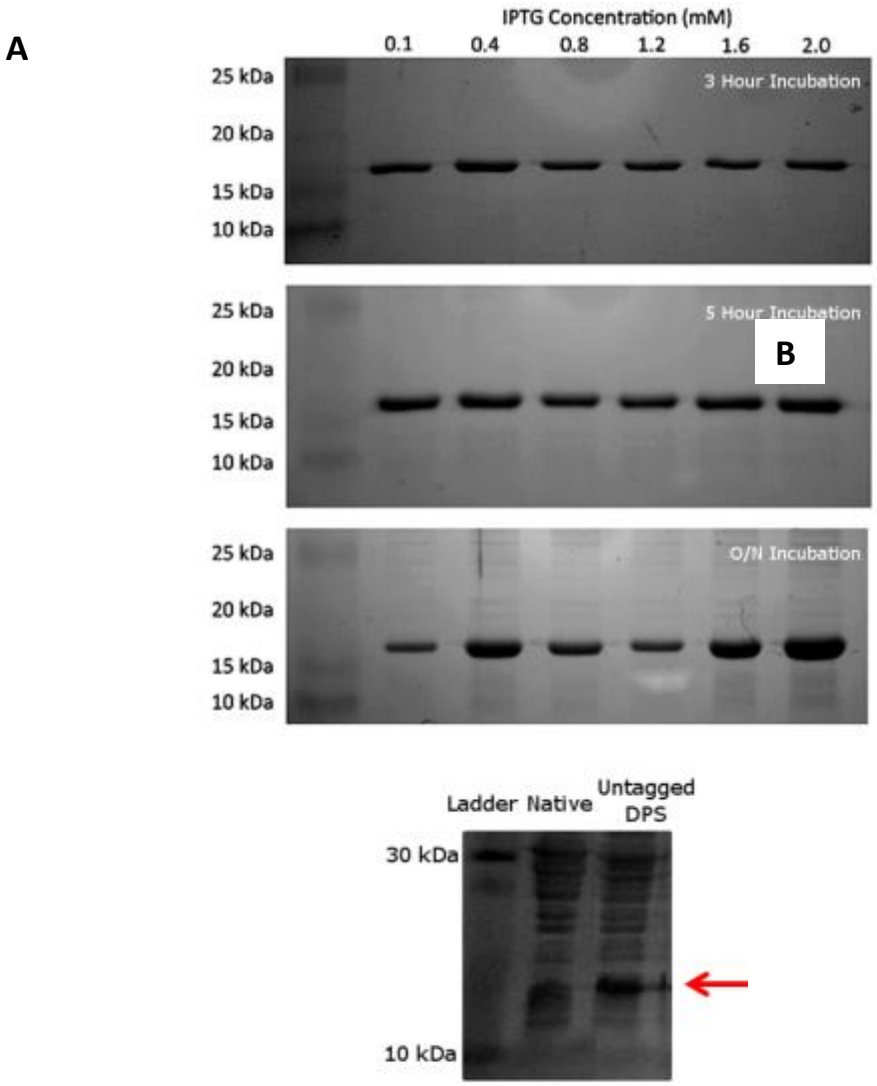
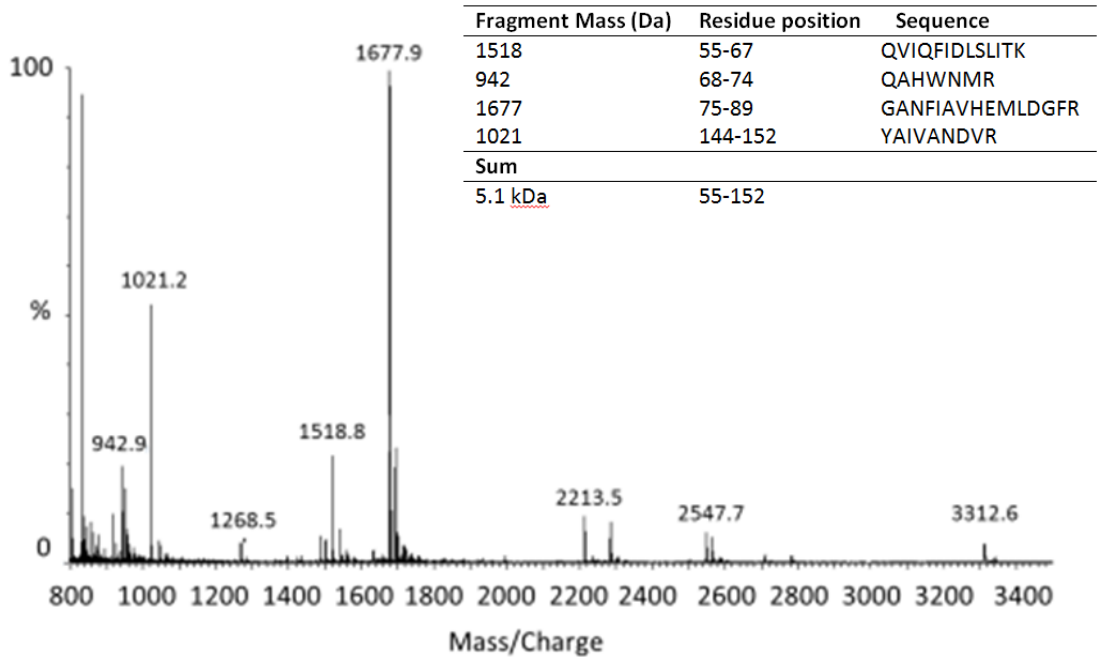


Figure 9.1 DPS over-expression under different IPTG concentrations and incubation times. A, N-terminally HIS tagged DPS was induced using a variable IPTG concentrations and time. All conditions lead to over-production of DPS. **B,** The untagged DPS construct also over-expressed (indicated by red arrow) when induced with 0.4 mM IPTG and a 5 hour incubation period.



Fragment Mass (Da)	Residue position	Sequence
1518	55-67	QVIQFIDLSLITK
942	68-74	QAHWNMR
1677	75-89	GANFIAVHEMLDGFR
1021	144-152	YAIVANDVR
Sum		
5.1 kDa	55-152	

Figure 9.2 MALDI-TOF mass spectrometry of DPS trypsin fingerprint taken from 12% SDS PAGE. N-terminal 6xHis tagged DPS (18.7 kDa) was identified by comparing against the peptide fragments expected from a theoretical trypsin digest of its sequence (inset).

9.3 Solubility of N-terminal HIS tagged DPS

Despite the confirmation that DPS could over-express it was uncertain whether these constructs were soluble in solution. To this end the lower IPTG concentrations were repeated but separated into soluble and insoluble fractions by centrifugal force (90 minutes, 47.7xg).

Unfortunately the 6xHIS tagged construct expressed in the insoluble fraction despite inducing with low levels of IPTG and varying incubation times (Figure 9.3). This suggested that 6xHIS tagged DPS was prone to form inclusion bodies during expression which wasn't expected compared to other studies [189].

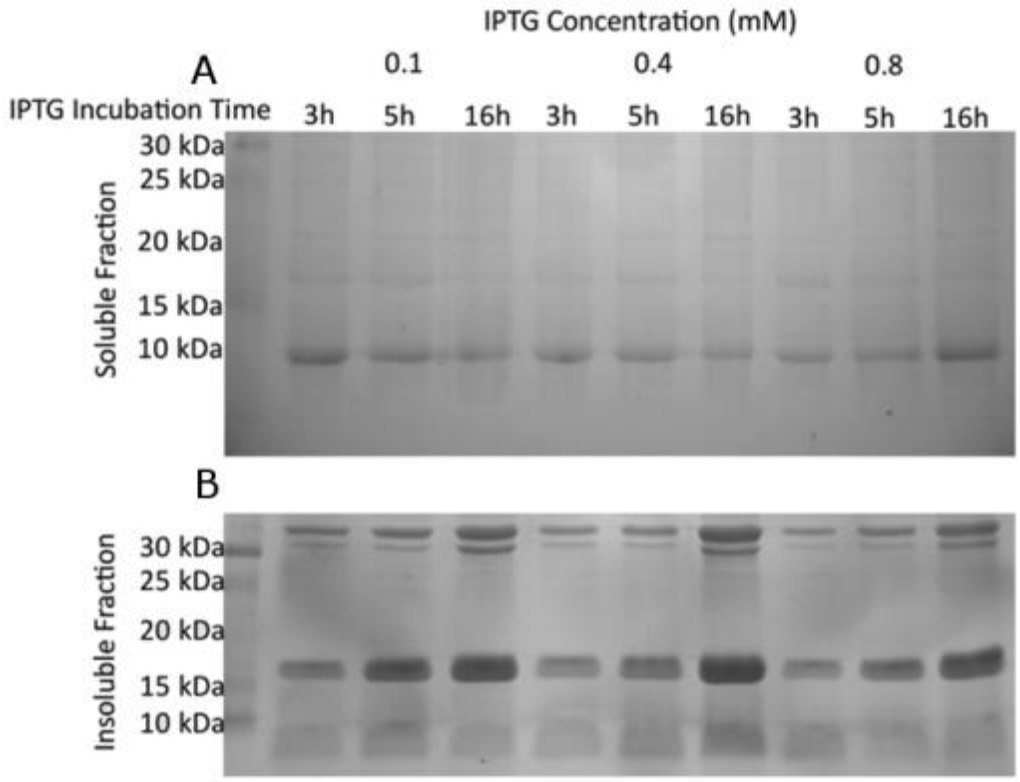


Figure 9.3 Solubility of 6xHIS tagged DPS. Tagged DPS cultures were split into **A,** soluble fraction and **B,** insoluble (lower gel) fractions by sonication and centrifugation at 19,000 rpm. The pellet containing insoluble proteins show that a large proportion of tagged DPS is being expressed between 15 kDa and 20 kDa to suggest DPS may be expressing as inclusion bodies. The soluble sample was dilute but DPS should still resolve considering that it was the most abundant protein and other bands can be seen.

Given the quantity of over-expression a refolding protocol was employed adapted from the Terzic *et. al.* 2010 methodology [190] in an attempt to rescue the protein. This protocol had previously been used successfully to purify large octomeric complexes of the nucleotide binding domains of sulfonylurea receptors (SURA). Though the protocol re-solubilised 6xHIS tagged DPS it failed to bind to a nickel affinity column and ssDNA cellose column described in Almirons 1992 protocol

[161] despite the tag being intact as shown through a western blot (Figure 9.4). This suggested that DPS was not refolding correctly and could not be used.

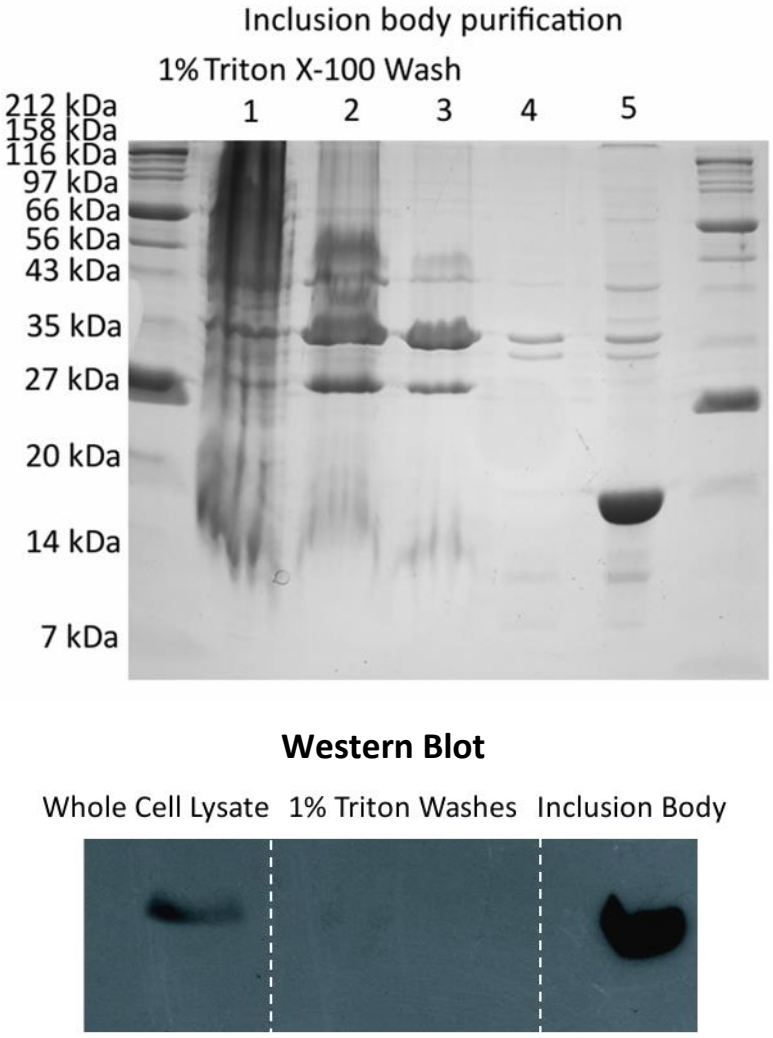


Figure 9.4 Insoluble pellet wash and 6xHIS tag specific western blot for tagged DPS. construct. The pellet was washed in 1%Triton X-100 3 times (1-3) and then again in the same buffer without Triton X-100 which contained 50 Mm Tris, 50 mM NaCl, 1 mM TCEP, 0.5 mM EDTA and 5% glycerol pH 8.0 (4). The remaining insoluble pellet contained the remaining 6xHIS DPS and an intact N-terminal 6xHIS tag as shown in the Western blot.

9.4 Production of untagged DPS

Fortunately untagged DPS expressed in the soluble fraction under similar conditions to the 6xHIS tagged version (IPTG 4mM, 5h). This suggests that tagging of the DPS

N-terminal is disruptive to the structure during expression. The purification then followed the original protocol developed by Almirion *et. al* 1992.

During the 30%, 60% and 90% ammonium sulphate (AS) cuts DPS was found in the 60% and 90% fractions as described by the original authors with four fold more DPS precipitating in the 60% AS sample. The gel filtration was performed using pre-packed gel filtration columns with high resolution ranges since DPS is prone to self-aggregation, shown to elute near the void volume using a Superdex 200 16/60 (Appendix 10). A Sephracryl 300 16/60 column was sufficient to separate DPS to elute the protein at ~55 ml in buffer containing 2M NaCl to dissociate the DNA (Figure 9.5). The fractions were pooled and returned into its original buffer by dialysis to produce >70% pure DPS (Figure 9.6). Unfortunately the 2M NaCl wash wasn't sufficient in removing all the DNA contaminant as judged through comparing the 260:280 nm wavelength ratios as previously described for TDP-43S, Chapter III. Through scanning the DPS spectra it was clear that there were differences between the DNA content of 60% and 90% AS cuts (Appendix 7) with 260:280 nm of 1.57 and 1.01 respectively, indicating more DNA in the 60% fraction. DNA was present in both cuts indicative of a high absorption at 260nm (Appendix 7) so were independently screened by crystal trials in an attempt to obtain DPS:DNA complexes to determine the specific binding contacts that nteract with DNA and surrounding dodecamers. The DPS peaks eluted between the 56-62ml elution volume and pooled as shown (Figure 9.5).

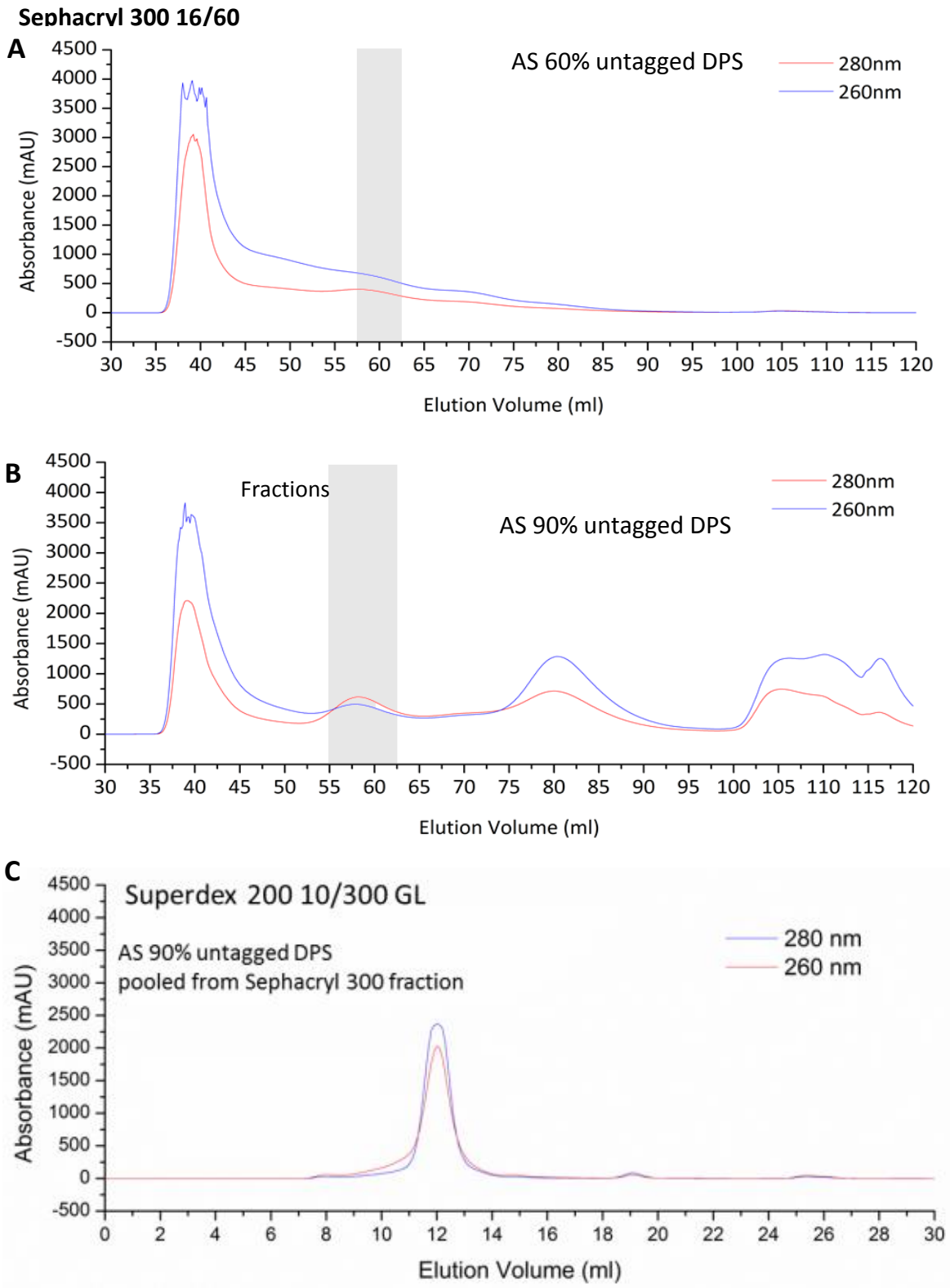


Figure 9.5 Gel filtration profiles of 60 and 90% ammonium sulphate cuts. The AS fractions were purified individually down a Sephracryl 300 16/60 gel filtration column at 0.5ml/min^{-1} where DPS eluted at ~55ml. **A,** The elution profile of the 60% AS fraction and **B, t**he 90% AS elution profile. The latter shows a better resolution of DPS to allow more volume to be pooled (shown in grey) for subsequent dialysis, concentrating and crystal trials.

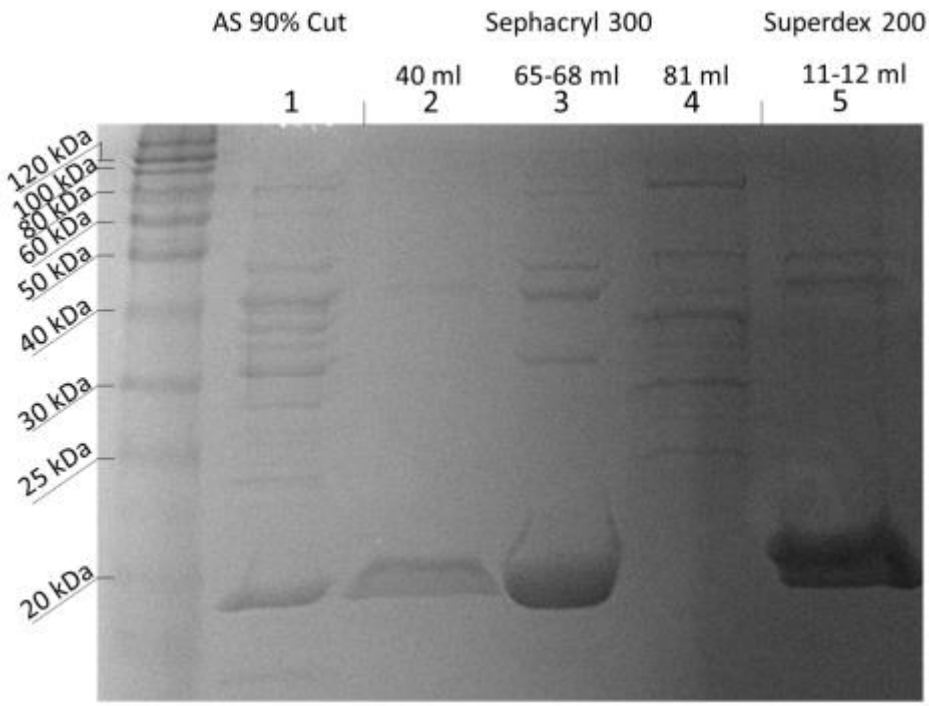


Figure 9.6 DPS purity after Size exclusion chromatography. The purity of the DPS used in determining the crystal structure after 2 gel filtration runs. This shows (Lane 1) purity after 90% ammonium sulphate precipitation. Lanes 2-4 show the peak fractions from the AS 90% DPS fraction using the Sephacryl 300 10 30 column (Figure 9.5). Lane 5 Shows the final purity of DPS used in the successful protein crystallography of DPS after Superdex 200 10 60 purification.

Chapter X

X-ray Crystallography study of *E.coli* DPS

Purified DPS crystallised and analysed by X-ray crystallography. This is a standard method in determining protein structure used to extract atomic co-ordinates from protein crystals. There is a large list of X-ray crystallography literature available to describe the technique in detail [191–193], however the following sections will provide the brief methodology used in context of the DPS project.

10.1 X-ray crystallography

X-ray crystallography (PX) is well suited for the study of DPS since the complex readily forms ordered crystals. PX works by determining the electron density contained within a single asymmetric units (unit cell) that is self-repeated throughout the structure of the crystal. This self-propagating unit cell acts to amplify X-ray diffraction signal that would otherwise be too weak to determine the parameters necessary to solve the macromolecular structure. The detected diffraction is measured in reciprocal space with co-ordinates *h, k, l* and this relates inversely to the electron density of the unit cell with co-ordinates *x, y, z* (Figure 10.1) by the following relationship;

$$\rho(x,y,z) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|e^{-2\pi i(hx+ky+lz-\propto'_{hkl})} \qquad \text{Equation (10.1)}$$

Where F_{hkl} is the structure factor describing the contribution of scatterers to reciprocal reflections, i is the imaginary number $(-1)^{1/2}$, 'h, k and l' are the reciprocal coordinates of the real space and ∝ describes the phase. The phase is the only parameter that cannot be measured or determined from the reflection pattern or X-ray source, which as noted in equation 10.1 is a prerequisite for solving the electron density of the unit cell (Equation 10.1) [194]. During this study the phase was obtained by molecular replacement using a previously solved structure though

other methods including isomorphous replacement and anomalous diffraction can also be performed [194].

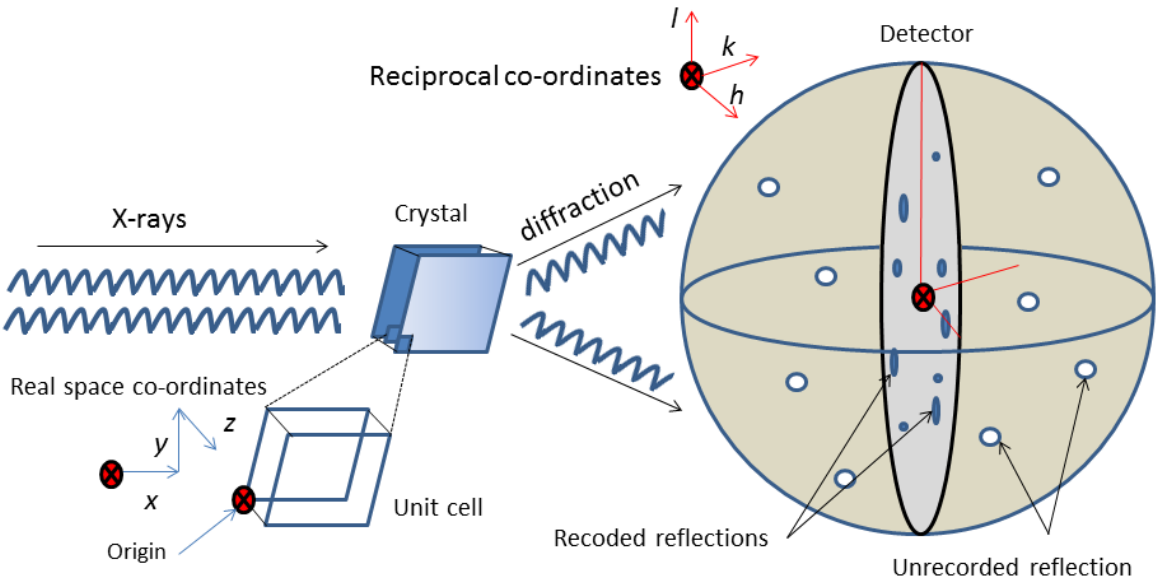


Figure 10.1 Cystallographic data collection. The real space co-ordinates (x,y,z) relate to the reflections measured in the reciprocal space by an inverse relationship (equation 10.1). The furthest reflections measured describe the closest distances in real-space related to the maximum resolution that can be described by the data. The crystal needs to be rotated in order to record all reflections corresponding to the electron density in order to build the 3D structure.

10.2 Data manipulation

The first step is concerned with understanding the unit cell and quality of the data. Because a crystal and unit cell conformation is static in one arrangement an angular range probing different crystal orientations is required during data collection to provide diffraction of the 3D co-ordinates. Therefore a crystal dataset can consist of many diffraction images. Through time and rotating the crystal the intensities even for the same reflections can differ between images due to different crystal lengths or beam-fluctuations. Firstly the data is reduced performed by indexing the

reflections to their respective *h, k, l* co-ordinates. From these Miller indices the cell type, dimensions, mosaicity, Intensity (I_{hkl}) and uncertainty (σI) can be derived which are performed by the programs iMOSFILM [195] or HKL2000 [196]. To normalise the intensities of each reflection the *h, k, l* co-ordinates are compared against multiple frames to assign relative I_{hkl} values performed by SCALA [197] or HKL2000. Data collection quality is commonly assessed by the R_{symm} value as described in Equation 10.2.

R_{Symm}

R_{symm} is a quality measure for the collected data which compares the extent each unique reflection deviates from being equal to each other commonly shown as a percentage.

$$R = \frac{\sum_{hkl}\sum_i |I_i(hkl) - < I(hkl) >|}{\sum_{hkl}\sum_i (hkl)} \qquad \text{(Equation 10.2)}$$

Where I_i is the intensity of a single reflection.

10.3 Phasing by 'Molecular Replacement'

As mentioned previously the phases of the reflections are a prerequisite to solving the structure which cannot be measured directly from the diffraction images. This is because F_{hkl} is a wave function and as such must include information regarding amplitude, frequency and phase. The former two functions are known from the experiment; the frequency is the same as the X-ray source and amplitude is related to the I_{hkl} which is related to the intensities of all reflections. The most efficient method of obtaining phases is to substitute them from a similar crystallographic structure. Through combining the experimental indexed co-ordinates and phases from a similar structure the diffraction data containing all wave functions can be

solved using algorithms in the MOLREP [198] software to generate a rough guide of electron distribution.

10.4 Refinement.

The structure from molecular replacement using the experimental I_{hkl} is still a draft map of the electron density derived indicative of the discrepancies between the actual and the phases used. The draft map can be improved upon by adding residues or explaining unknown density of the model by adding/ rotating amino acids (known from the sequence) or adding structure of known buffer components. This refinement aims to lower the R values that measure the accuracy of the model to the data (Equations 10.3). Gradually filling the unknown density by this method gradually improves the model to resemble the real space co-ordinates of the protein judged through improved statistics.

During this study non-crystallographic symmetry was applied and the data was refined using rigid-body refinement to a single monomer of *E.coli* DPS structure (PDB id: 1F30).

Models were built using COOT [199] and refined by REFMAC [200]. See below on how the R-values and important statistics are used in interpreting the model quality;

R-factor

$$R = \frac{\sum\big||F_{obs}| - |F_{calc}|\big|}{\sum|F_{obs}|} \qquad \text{(Equation 10.3)}$$

F_{obs} is the structure factor amplitude from the native dataset and F_{calc} is the structure factor calculated from the model.

R-free

R-free measures how well the current atomic model fits with a random subset of measured intensities not used during refinement which is less biased then R-factor.

Sigma I (σ(I))

Mean intensity divided by the error, a measure of 1 implies that only background has been measured. The Sigma cut-off was set at 0.00, meaning that intensities which have a negative value on the detector were ignored.

B-factor

Otherwise known as the temperature factor. This is a measure of the oscillation around a position specified in the model that naturally occurs in protein structure. Since diffraction can be affected it is appropriate to assign a value to the positional variation of atoms in the model.

Chapter XI

DPS structure and polar contacts involved in crystal formation

11.1 DPS crystallisation

The DPS crystallisation trails uncovered a variety of suitable conditions for DPS crystallisation however only a few of these conditions produced crystals large enough (typically > 0.1mm at its longest dimension [201]) for PX analysis. The largest crystals were taken to the Diamond synchrotron source in Oxford to maximise the diffraction and intensity profile. The DPS crystals that diffracted were grown from the 90% ammonium sulphate cut obtained from a 1:1 ratio containing 4mg/ml DPS and 0.2M $NaNO_2$, 20% w/v PEG 3350.

The resolution was limited to 2.74 Å which is comparable to previously solved DPS structures, including *E.coli* (PDB id: 1F30). The data collection statistics and parameters for the *E.coli* DPS structure are shown in table 11.1. Waters were absent during rigid-body refinement, using isotropic B-factors and non-crystallography symmetry to a single monomeric unit (residues 1-167) of E.coli DPS (PDB id:1F30). Unfortunately DNA was absent in this crystal structure as observed in other DPS complexes solved so far. The N-terminal regions of DPS monomers were also missing. The last N-terminal residue visible in the electron density was Thr12. The fact that the remaining N-terminal residues are missing and known to be bound by DNA supports the notion as in other DPS structures that DNA and N-terminal residues are flexible within the hexagonal crystal packing.

11.2 DPS unit cell

The P1 asymmetric unit cell contained three full DPS dodecamers, with maximum unit cell dimensions of 222.95 Å (Figure 11.1). The proteins positioning within this

cell shows that the orientation between individual DPS dodecamers differs. This is to presumably enhance the positions of functionally important residues important to bind DNA and neighbouring DPS, to form the large crystalline DNA containing structures observed *in vivo* during bacterial stress responses [170]. Interestingly the rotation of each neighbouring DPS complex differed by a ~120˚ rotation at a ~45˚ tangent axis to each neighbouring complex mimicking the three-fold symmetry of the ferritin-like pores that coincidently align close to the gaps generated between dodecamers. This packing is tight between neighbouring complexes showing that the DPS assemblies interact. The positive charge is caused by the remaining N-terminal which protrudes next to dodecamer binding sites. The majority of the complex is neutral apart from the ferritin-like channels that are negative.

Table 11.1 DPS crystal structure statistics

Data collection and parameters	***E.coli* DPS (During this project)**
Wavelength (Å)	0.98
Space Group	P1
Dimensions (Å, ˚)	a= 86.95, b=90.18, c=222.95 α=*80.40, β=86.69, γ=61.44*
Resolution (Å)	2.74
Observed Reflections	134843
Number of atoms used during refinement (non-hydrogen)	44496
Redundancy	3.0 (2.6)
Completeness (%)	92.20 (52.44)
R_{symm} (%)	28.77 (40.70)
I/σ(I)	17.78 (4.59)
Resolution range	30.00-276
R-factor/R-free (%)	28.77/32.87
Bond length (Å)	0.011
Bond Angle (˚)	1.292
Average B-factor ($Å^2$)	60.97

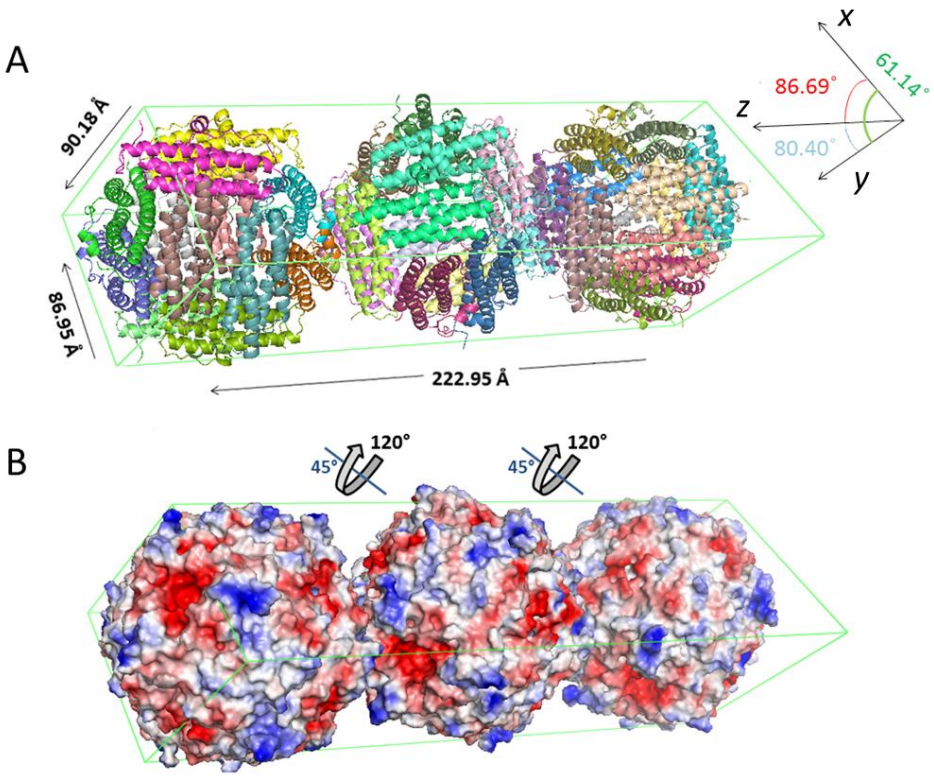


Figure 11.1 ***E.coli*** **DPS unit cell packing, complex charge and orientation.** The unit cell of DPS. **A**, Showing the dimensions of DPS cell and monomer orientation within. Three complexes form within the P1 space group. **B**, The vacuum electrostatics (PyMol http://pymol.com/) show the distribution of charge and orientation to neighbouring complexes. Note that the N-terminal is positive (blue), the surface of the complex is neutral (white) except for ferritin-like iron channels (red).

11.3 DPS-DPS interactions

From the crystal structure (analysed in PyMol) the central DPS dodecamer forms three polar interactions to each neighbour in the same plane and two to a single DPS dodecamer in the plane behind (Figure 11.2). In the longest dimension (z axis) polar interactions can occur between residues N99, S106, K134 to D20, T12 and N99 respectively. The contacts involve two monomeric subunits from each complex and interestingly involve Thr12 to suggest that this residue may be important in organising the self-aggregation when DNA is bound in *E.coli* DPS. This residues involvement in forming this contact likely explains why it can be resolved by crystallography despite its position on an unstructured region of the protein.

The hexagonal packing arrangement means that dodecamers form a cleft involving three other dodecamers (Figure 11.2). However interactions only form to two of these using two polar interactions; R18, D20 to S100 and T12 respectively meaning these interactions are likely weaker than the neighbours separated about a 120˚ angle (Figure 11.2, C & D). This is representative of the DPS layers described by Wolf in *E.coli* during crystallisation [170], suggested to be important during DNA packing.

The weaker interaction to dodecamers on this plane means that DPS form a gap (Figure 11.3). The N-terminals of surrounding dodecamers are positioned to face this space. Because only two polar contacts form to just one of these dodecamers more space is generated which may be important to localise the remaining 33 missing residues from each dodecamer and the condensed DNA (Figure 11.3).

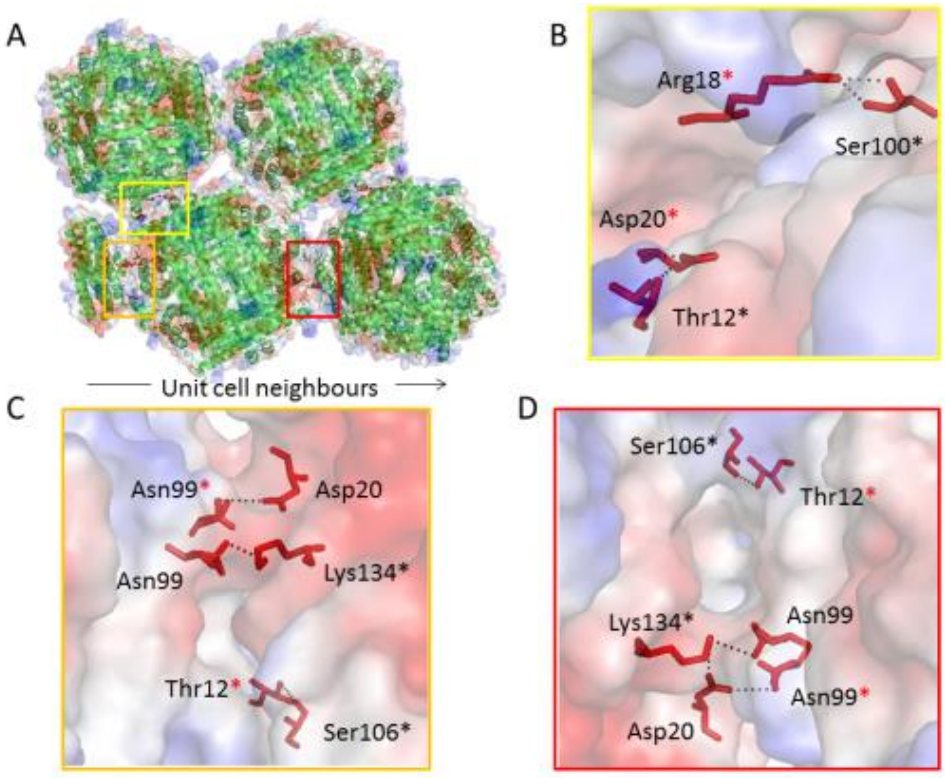


Figure 11.2 Polar contacts between neighbouring *E.coli* DPS complexes'. The polar contacts formed between the central DPS dodecamer in the unit cell. **A,** The DPS cartoon representation of the monomer organisation, neighbour contacts and electrostatic potentials. The polar contacts formed between **B,** the left dodecamer and **C,** the right dodecamer. Stars (*) indicate contacts are mediated on the same monomer.

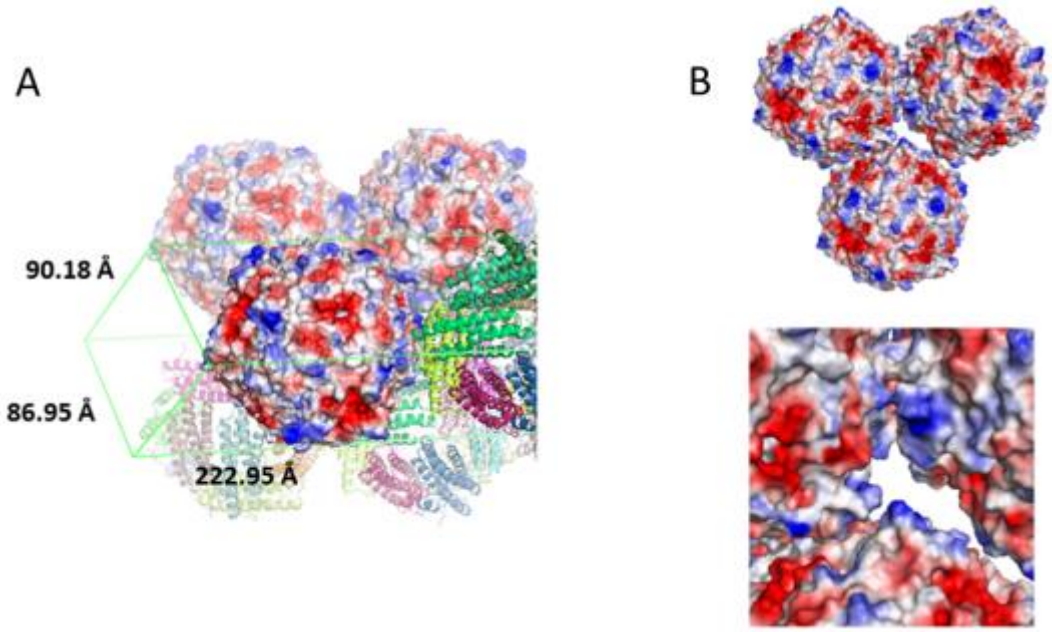


Figure 11.3 *E.coli* DPS complex packing and suspected DNA containing cleft. The DPS hexgonal arrangement. **A**, The hexagonal layers arrangement. **B**, The top view of this packing assembly, in the blown up image below you can observe the elongated cleft generated between interacting dodecamers.

11.4 DPS structure comparisons

Through comparing the DPS monomeric subunits from different species there was little observed differences between the structural alignment and RMSD values (Figure 11.4 & Table 11.2). The largest differences were from the remaining N-terminals which were anticipated given that this region lacks secondary structure and is likely to be flexible. The lowest RMSD is unsurprisingly from *E.coli* at 0.14 Å and the largest is from *M. smegmatis* at 0.73 Å which is due to its elongated C-terminal which extends nine residues further than *E.coli* in the structure and twenty seven in sequence.

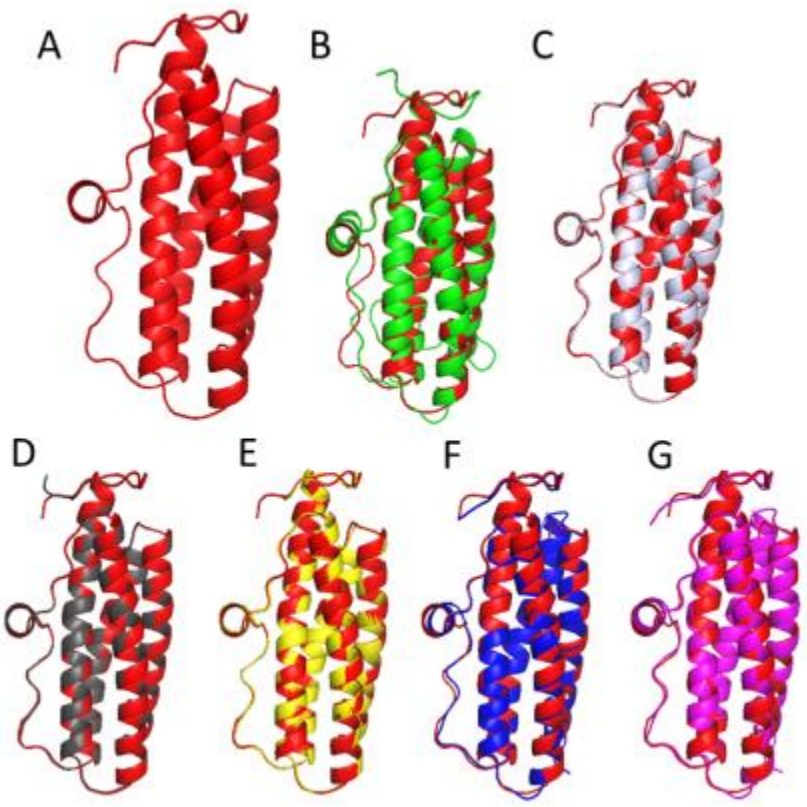


Figure 11.4 Alignment of E.coli DPS with previously solved DPS monomers. The superimposition of the monomeric DPS structure between different species. **A,** DPS in this study was compared against species with known structures (shown in red). To include **B,** *Myobacterium smegmatis* PDB Id: 1VEI. **C**, *Yersinia pestis* PDB Id: 4DYU . **D,** *Escherichia coli* PDB Id: 1F30. **E**, *Salmonella enterica* PDB Id: 3AK8. **F,** *Brucella melitensis* PDB Id: 3GE4 *and* **G**, *Agrobacterium tumefaciens* PDB Id: 1O9R.

Table 11.2 The structural conservation between DPS monomers.

DPS Structure (Residues)	No. atoms used IN RMSD	RMSD to 2.8Å DPS structure *	Model No.
M. smegmatis (1-175)	125-125	0.73	B
Y. pestis (13-166)	149-149	0.34	C
E.coli (12-166)	142-142	0.14	D
S. enterica (14-166)	146-146	0.24	E
B. melitasis (6-166)	128-128	0.47	F
A. tumefaciens (1-162)	133-133	0.52	G

* Note that RMSD values may differ due to crystal structure resolution and are shown here as a rough guide to DPS similarity

11.5 DPS packing similarities between other species

The hexagonal packing from this study is consistent with that of other species (Figure 11.5). The packing of *M. smegmatis* DPS which binds DNA through its C-terminal domain can also adopt both cubic and a hexagonal cells [202](not shown) to suggest that DNA binding may not be the only determinant in crystal packing and self-aggregation.

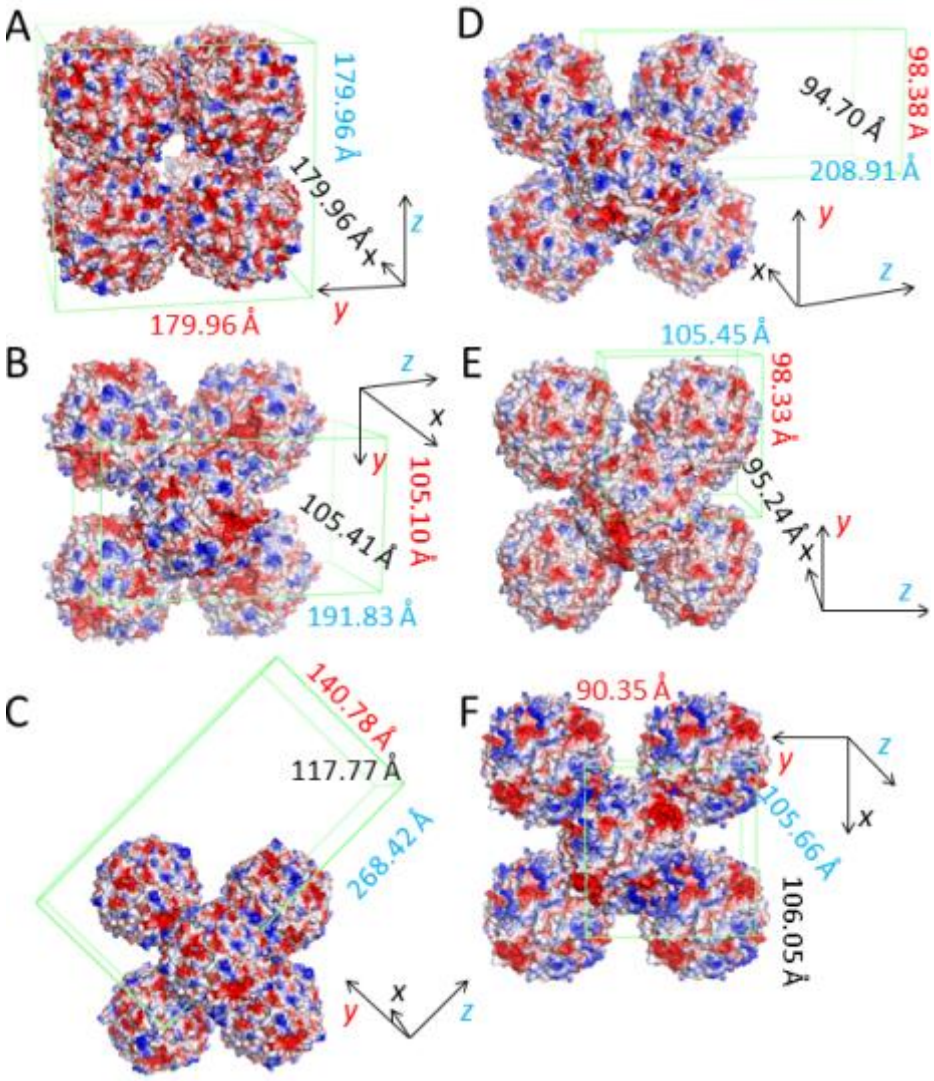


Figure 11.5 The organisation of complex assembles of DPS crystals between species. The unit cells of DPS crystal complexes were compared to show packing differences, unit cell parameters and charge. The crystals are from **A,** *Myobacterium smegmatis* PDB Id: 1VEI. **B***, Yersinia pestis* PDB Id: 4DYU . **C,** *Escherichia coli* PDB Id: 1F30*.* **D***, Salmonella enterica* PDB Id: 3AK8. **E,** *Brucella melitensis* PDB Id: 3GE4 *and* **F***, Agrobacterium tumefaciens* PDB Id: 1O9R.

Except for *M. smegmatis* the DNA binding species with conserved N-terminal lysines in *E.coli* [161], *Y. pestis*, and *S. enterica* all crystallised with the largest unit cells. This suggests that being bound to DNA does have some effect on the hexagonal crystal packing arrangement. Using the non-DNA binding species of *A. tumefaciens* as a comparison, the hexagonal arrangement is more uniform to reveal just one conformation of the DPS dodecamer mediated by 3 polar contacts to dodecamers on different planes (Figure 11.6). In the packing of *E.coli* DPS these neighbouring contacts are variable which may be an adaptation to optimise the amount of DNA which can be stored, facilitated by the ability of DPS to form multiple polar interactions when rotated (Figure 8.1). Though DNA binding has been shown as critical for self-assembly it is likely that when DPS dodecamers come into contact with DNA the polar interactions may further stabilise its positioning.

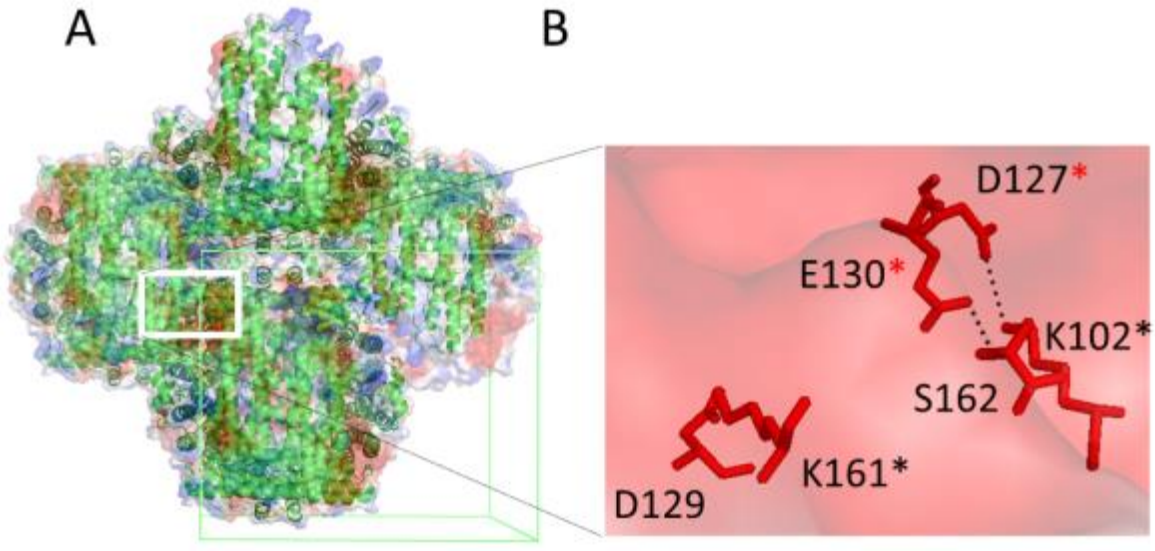


Figure 11.6 Non-DNA binding model of *A.tumefaciens* hexagonal crystal packing. The small unit cell is indicative of a simple DPS organisation which is mediated by 3 polar contacts at each corner of the dodecamer and none with the DPS above and below on the same plane. Stars (*) indicate contacts are mediated on the same monomer. Using PDB model; 1O9R.

11.6 Polar anchors.

Polar interactions were observed in every DPS crystal. Since the sequence identities between *E.coli* and the DNA binding species; *Y. pestis* and *S.enterica* were well

conserved (80-95% identical, expasy BLAST) it was thought that perhaps these polar contacts may also be maintained to describe a common self-assembly mechanism involving polar interactions. Using the polar contacts of *E.coli* observed during this study, 3/7 residues involved also formed interactions in either of these two species. Using the numbering of *E.coli* conserved residues involved in these interactions were Asp99, Ser106 in *Y.pestis* and Arg18 in *S.enterica.* To suggest that polar contacts can vary during self-aggregation.

There is a high level of conserved charged surface residues accessible in DPS throughout all DPS species (Figure 8.1) capable of forming polar contacts (Table 11.3). Given this and the fact that the surface of DPS does not mediate any identified interaction with DNA suggests that perhaps the function of these residues is to form stable interactions which can form using a combination of residues depending on the amount of DNA.

Table 11.3 Residues involved in forming polar contacts between DPS dodecamers

DPS species	**Residues involved in forming contacts**
DNA-Binding Species	**(Numbering as ClustalOmega alignment)**
E.coli (*E.coli* from 1F30)	T15, R21, D23, N102, S103, S109, K137, (T12, D23, D26, S25, L17, D26, K146)
Y.pestis	H27, N102, S109, E146, L107
S.enterica	R21, S27, D28, D119, K122, E173
M. smegmatis	K103, D108
Non-DNA binding species	
A.tumefaciens	N27, K116, D122, D140, E144, S176
B. melitensis	S26, K108, D119, E126, D130, N176

11.7 Significance of polar contacts and DNA in DPS packing

The crystal structure of the *E.coli* DPS:DNA complex could not be resolved despite a strong indication that it is DNA bound. This confirms as with other DPS structures that at least the first 11 residues are flexible in this structure. It is likely that DNA is contained within the cleft generated between layers of weaker contact as this was shown to generate a larger pore and is the most likely place for DNA to be stored as previously suggested by Grant *et. al* 1998 [185].

The hexagonal packing is observed in all DPS species however the larger unit cells are associated with the DNA-binding species. Since many of the polar residues responsible in forming interactions with neighbouring dodecamers are largely conserved there is no other reason why this packing should change other than the addition of DNA. Through comparing the hexagonal packing arrangements between DNA and non-DNA binding species we observe that dodecamers rotate 120˚ which allows them to come into closer contact when compared against a non-binding species. This is likely to mimic the three-fold symmetry of the ferritin-like pores and to orient N-terminal residues into the clefts between dodecamers. This would allow the N-terminals to interact to possibly the same condensed DNA and convey the very stable aggregates described by Almirion *et. al.* 1992 [161]. This would also explain why the longer unit cell dimensions are also observed within DNA-binding species.

The actual polar contacts observed in the structure are poorly conserved, even in the same species as shown between the *E.coli* in this study and previously solved structures (PDB id; 1F30, 1DPS). This implies that the positioning of DPS within the crystal and possibly the *in vivo* aggregates are not mediated by specific polar interactions. However given the large amount of polar residues accessible on the dodecamer surface which is made equally from negative and positive residues (Figure 8.1) not involved in DNA binding or ferritin-like activity [169,176], it is possible that they provide points of contact that aid to stabilise the dodecamer assembly. This may also explain why DPS is readily crystallisable in a variety of species.

Chapter XII

Materials and methods

12.1 Generating the wild-type TDP-43S plasmid

The 'mTDP-43SpGEX3X' plasmid components were generated by amplification of residues 101-265 of Mouse TDP-43 DNA (GI: 156616305) and additional base pairs incorporating the 3' BAMHI and 5' AatII restriction enzyme sites (New England Biolabs (NEB)). This was performed by the polymerase chain reaction (PCR) utilising the 'pGEX3XmTDP-43SFwd' and 'pGEX3XmTDP-43SRvs' primers (Appendix 8) and KOD Hotstart DNA polymerase (Novagen) under the following cycling conditions;

Table 12.1 PCR cycling conditions used to amplify TDP-43S genes

	Stage	Temperature (°C)	Time (s)
	DNA Separation	95°C	60
x30 Cycles	Denaturation	95°C	30
	Annealing	60°C	30
	Elongation	65°C	300

The amplified fragment was isolated by 1% TAE agorose gel electrophoresis set at a constant 100V and excised. The 513bp fragment was collected using the Wizard® SV gel and PCR clean-up system (Promega) and digested with restriction enzymes to generate the 5' and 3' sticky ends as described for the enzymes (NEB). The insert was ligated into a similarly treated linearized pGEX3X expression vector (GE Healthcare) using a 3:1 v/v insert/vector ratio as components in the reaction mix containing; 1x T4 ligase buffer, and 1µl T4 DNA ligase (NEB) in a 10µl total volume. This reaction was incubated at 16°C and left over-night.

The reaction was transformed into competent *Omnimax2* strain of *E. coli* cells (Appendix 3) and incubated on selective Luria Broth (LB) agar overnight at 37˚C. All

TDP-43S media selectivity was achieved through addition of 100µg/ml Ampicillin. Single colonies were grown in 5-10ml LB for >16 hours and isolated using Wizard® SV Miniprep DNA Purification system (Promega). The plasmid was sequenced verified using DNA terminator reaction techniques at GATC biotechnology using custom-made sequencing primers, 'pGEX3X Sequence Rvs' and 'pGEX3X Sequence Fwd' (Appendix 9). Purified plasmids were stored at -20˚C.

12.2 Generation of D169G and K263E mutant TDP-43S plasmids

Primers were designed to infer D169G and K263E protein mutations to create two separate mutant constructs (Appendix 1) from the 'mTDP-43SpGEX3X' plasmid. Site directed mutagenesis was performed using the QuikChange Lightning Site-Directed Mutagenesis Kit (Stratagene). The Primers; 'MGD169GmTDP-43SFwd', 'MGD169GmTDP-43SRvs', 'MGK263EmTDP-43SFwd' and 'MGK263EmTDP-43SRvs' were designed respectively to substitute a bp to generate 'D169GmTDP-43SpGEX3x' and 'K263EmTDP-43SpGEX3X' plasmids (Appendix 3). The mutated codon corresponding to the residue change is highlighted in Appendix 8.

12.3 Expression of wild-type, K263E and D169G TDP-43S proteins

All mTDP-43S and variant plasmids were transformed into *E. coli* BL21 DE3 (Appendix 3) cells and plated as described previously for *Omnimax2*. Colonies were grown over-night using 1% of the desired growth volume in selective LB media. This overnight was then used to inoculate a larger growth volume (0.5-1L) and left at 37°C in a shaking incubator set at 220 rpm, until an optical density measured at a 600nm wavelength (OD_{600}) of 0.5-0.8 was achieved. Protein production was initialised by addition of 1mM Isopropyl β-D-1-thiogalactopyranoside (IPTG). Protein production was left for 5 hours at 30°C, shaking at 220 rpm, the cells were then pelleted at 3000g, at 4°C for 20 minutes. Pellets were flash frozen and stored at -20˚C.

12.4 Mouse TDP-43S purification

TDP-43S was purified from the frozen -20˚C pellets which were thawed and resuspended in 1:3 w/v ratio of TDP-43S buffer (50 mM Tris-HCl, 150 mM NaCl and 2 mM $CaCl_2$ and pH 7.4), 0.2mg/ml lysozyme, 10µg/ml DNase and 10x concentrate protease inhibitor cocktail set III (Calbiochem). Cells were lysed by sonication and cell debris removed by centrifugation at 30,000g for 90 minutes, TDP-43S proteins were contained in the soluble fraction. This fraction was filtered through a 0.2µm syringe filter (Milipore) and applied to an equilibrated GSTrapFF affinity chromatography column in the same buffer (GE Healthcare). The TDP-43S sample flow was set at 1ml/min^{-1} and ran through twice. The GST-TDP-43S bound GSTrapFF column was washed with 10 column volumes of high salt (0.5M NaCl) TDP-43S buffer, to remove electrostatic interactions. The GST column was then returned into standard TDP-43S buffer.

The GST tag was removed on column using Factor Xa as described in the GSTrapFF manual (GE Healthcare). To elute and remove proteases the sample was passed through an attached BenzamidineFF column (GE Healthcare) and 95% pure TDP-43S was collected. Concentrated TDP-43S proteins were further purified to 99% by passing down a pre-equilibrated Superdex 200 16/60 (GE Healthcare) gel filtration column in TDP-43S buffer at 0.8ml/min^{-1}. TDP-43S eluted between 92-96ml.

12.5 Calibration of Superdex 200 16/60 gel filtration column

Molecular weight standards were calibrated to the elution volume of a Superdex 200 16/60 gel filtration chromatography column using the ÄKTA purifier system (GE Healthcare). The standards include; BSA (67 kDa), myoglobin (17 kDa) (Sigma) and SOD1 dimer (32 kDa) (UoL) which were individually ran at 2mg/ml concentration and 2.5ml sample volume.

12.6 Mouse TDP-43S crystallisation trials

Concentrated Mouse TDP-43S protein (5-10mg/ml) were screened using hanging or sitting drop vapour diffusion methods. The hanging drop used pre-greased hanging-drop plates and cover slips (Hampton Research) whilst the sitting drops were set in 96 well sitting plates, and covered using clear sealing film (Hampton). Drops were mixed in a with reservoir solution in either a 1:1 or 2:1 ratio as described in Chapter III. The plates were sealed and incubated at 4°C or 18˚C and observed weekly using a light microscope. The list of trials used are included in 'Chapter III, Table 3.2'.

12.7 Differential scanning fluorimetry melting temperature assay

DSF was performed using specific 96 well MicroAmp® plates (Life technologies) in a 60µl volume with 10x Sypro-orange (Life technologies) and TDP-43S variants at a concentration of 0.6 mg/ml, repeated seven times for each. Samples were measured using a real-time PCR thermal cycler (StepOnePlus *Applied Biosystems*) covering the temperature range 14-95°C, increasing at a rate of 1.2 °C/min^{-1}. Fluorescence was measured at ~470 and ~610 nm for excitation and emission respectively. The 7 melt curves for each; wild-type, K263E and D169G constructs were obtained and averaged and their first derivative plotted and analysed to obtain their respective melting temperatures.

12.8 Unfolding tracking by aromatic fluorescence

Aromatic fluorescence of TDP-43S samples (0.45mg/ml) were monitored in a fluorimeter using excitation and emission wavelengths of 280 and 340 nm respectively. A total of 6 repeats for each construct and buffer were performed.

Samples and buffer were preheated for 5 minutes over 2 °C increments between 37 to 65°C. The samples were buffer subtracted and averaged to produce the data points and standard deviations. Boltzmann curve fitting was calculated using Origin 8.6 (OriginLab).

12.9 Small angle X-ray solution scattering data collection

The small angle X-ray solution scattering (SAXS) data for TDP-43S and variants were collected at St Aubin, France using the Soleil, synchrotron facility and their dedicated SAXS station, SWING [106]. X-ray wavelengths were set at 1.03 Å and the CCD sample detector (AVIEX 170x170mm) was positioned 180cm from the sample, covering the momentum transfer range (q) =0.01 - 0.57 $Å^{-1}$, where $q=4\pi\sin\theta\lambda^{-1}$ (λ is the wavelength of the incident radiation and θ is half the angle between the incident and scattered radiation). The SWING beam-line was coupled to a Shodex KW402.5-4F HPLC system set at 25°C and equilibrated with TDP-43S buffer. A 40µl sample volume of concentrated TDP-43S (10mg/ml) was injected and its elution set at 300µl/min^{-1} flow was collected over 255 frames. Each frame was measured for one second and separated by 0.5 second dead time. The buffer profile for subtraction was achieved through averaging 20 frames prior to the proteins elution. The data were integrated, averaged, buffer subtracted and reduced using the in-built Foxtrot software developed in house at Soleil to generate 1D profiles of individual frames.

12.10 SAXS data interpretation and modelling

The 1D data selection was performed and averaged by PRIMUS [203] and radii of gyration (Rg) estimated to within qRg limits of <1.3. The p(r) profile and Dmax were calculated with Gnom, which generated the input for *ab inito* modelling. The *ab initio* modelling software, GASBOR was used to generate TDP-43S bead models of the Cα trace. The default software parameters were ran to fit the entire curve, with no symmetry (P1) constraints. This was repeated 10 times for each construct and averaged to produce the average electron distribution, calculated by DAMAVER [133].

The rigid-body modelling using the experimental 1D curves was performed by BUNCH [115]. Rigid-bodies for mouse TDP-43S RRM1 and RRM2 were derived from the PDB ids; 2CQG and 3D2W respectively. DNA was removed and structures truncated to contain residues 106-177 and 193-259. The Cα trace for residues 101-105, 178-192 and 260-265 were reconstituted using the online software SABBAC and copied into the PDB file. All 1D SAXS curves derived from *ab initio* and rigid-body models were fitted against the original 1D SAXS curves by CRYSOL [117], goodness of fit assessed by χ^2 minimisation. All model visualisations and RMSD alignments were performed using PyMol (http://pymol.com/).

12.11 TDP-43S heat mediated aggregation

The radiusi of gyration was estimated by the Guinier approximation using data collected in-house at the Barkla X-ray Laboratory of Biophysics at the University of Liverpool, at 20 and 40°C for each construct. TDP-43S samples (3mg/ml) static measurements performed in a mica cell in TDP-43S buffer with the addition of 20mM DTT. Scattering were collected on a MAR300 image plate at a distance of 1.25 m from the sample. Protein and buffer scattering were recorded over 3 separate 20 minute exposures which were then averaged. Data integration and analysis were performed using Fit2D [204] and PRIMUS [203]. The bounds of the Guinier region (qRg) all lay below 1.3 for each construct except the wild-type at 40°C which due to the large size was <2.0.

12.12 Circular dichroism

The optical activity of TDP-43S variants was determined at the Soleil synchrotron using the DISCO station [205]. Wild-type and mutant TDP-43S proteins were dialysed against 50mM phosphate buffer (pH 7.2) and 50mM NaF to a 0.8mg/ml final concentration. Measurements were taken using a 0.02cm quartz cell and 9 second acquisition time in range 180-260 nm. Data acquisition was repeated 3

times for each construct before buffer subtraction, averaging and smoothing by averaging two consecutive points. The data points were calibrated to CSA at a concentration of 6.3 mg/ml to correspond to peaks at -37.54 at 192.3nm and 18.13 at 291.4nm. CDSSTR secondary structure prediction was performed using the DichroWeb On-line server, using the SP175 reference set [136].

12.13 *E. coli* DPS cloning

The DPS gene was amplified directly from the genome of *E. coli* DE3 strain using primers designed for use with the in-fusion protocol (Clontech). Two DPS constructs were generated containing an N-terminal His tag amplified using '*pOpinFE coli*DPSFwd' and '*pOpinFE.coli*DPSRvs' or without using '*pOpineGFP*DPSFwd' and '*pOpineGFP*DPSRvs' primers (Appendix 1). Amplification was performed as described for 'mTDP-43SpGEX3X' except for a reduction to 60˚C in the elongation temperature.

PCR reactions were ran on 100V TAE gel corresponding to ~500bp which were, excised and extra contaminants removed (PCR clean-up system, Promega). The 6xHis tagged insert was ligated into a double digested pOpinF plasmid performed with HINDIII and KpnI restriction enzymes (NEB). The untagged insert was ligated into a similarly treated pOpineGFP plasmid and both adhered to the In-fusion® HD cloning kit procedure (Clontech). The reaction was transformed into Omnimax2 cells and incubated in selective agar plates overnight. The selectivity for DPS was achieved with 50µg/ml ampicillin. Colonies were grown overnight in 5-10ml selective LB broth, isolated, sequenced and stored as described for 'mTDP-43SpGEX3X', using the 'T7Minus1' sequencing primer.

12.14 DPS expression

Both DPS plasmids were transformed into DE3 *E.coli* cells and plated overnight on selective media (50µg/ml ampicillin). Colonies were grown over-night using 1% of the desired growth volume in selective LB media. This overnight was then used to inoculate a larger growth volume (0.5-1L) and left at 37°C in a shaking incubator set at 220 rpm, until an OD_{600} of 0.5-0.8 was achieved. Protein production was initialised by addition of 1mM IPTG. Protein production was left for 5 hours at 18°C, shaking at 220 rpm, the cells were then pelleted at 3000g, at 4°C for 20 minutes. Pellets were flash frozen and stored at -20˚C.

12.15 Purification of untagged DPS protein

DPS was purified from the frozen -20˚C pellets thawed and resuspended in 1:3 w/v ratio of DPS buffer (50 mM Tris-HCl (Ph 8.0), 500 mM NaCl and 0.1 mM EDTA), 0.2mg/ml lysozyme, 10µg/ml DNase and 10x concentrate protease inhibitor cocktail set III. Cells were lysed by sonication and cell debris removed by centrifugation at 30,000g for 90 minutes, untagged DPS was found in the soluble fraction and filtered through 0.2 µm pores. The least soluble proteins in this fraction were removed by 30%, 60% and 90% ammonium sulphate (AS) precipitation. DPS was found within the 60 and 90% AS cuts. Cuts containing DPS were purified further using a sephacryl 300 16/60 gel filtration column (GE healthcare) set at 0.5ml/min^{-1}, equilibrated in DPS buffer containing 2M NaCl.

12.16 Tagged DPS

The purification of tagged DPS was performed as described for untagged DPS except proteins were found in the insoluble fraction. The pellet was resuspended into 50 mM Tris–HCl (pH 8.0), 50 mM NaCl, 5 mM TCEP and incubated with 4.5% N-lauroylsarcosine which resolubolised the pellet after 30 minutes. Remaining

aggregates were removed by filtering through a 0.45 µm filter and refolded by dialysis into 2 litres of 10 mM Tris–HCl (pH 8.0), 0.05 mM EDTA, 0.1 Mm TCEP and 0.03% N-lauroylsarcosine at least 3 times. The sample was found in the void volume after gel filtration as described for both TDP-43S proteins and untagged DPS. The western blot analysis was performed as described [206] only using anti-6xHis-peroxidase antibodies (Roche) to avoid the use of secondary antibody.

12.17 Isothermal titration calorimetry

ITC was carried out at 25 °C with an iTC^{200} Microcal calorimeter (GE Healthcare). The TDP-43S proteins were tested in TDP-43S buffer in cell concentrations of 4.6µM, 2.0µM and 2.6µM respectfully for wtTDP-43S, k263eTDP-43S and d169gTDP-43S. Titrations consisted of 1 preinjection of 0.3µl and 18 consecutive 1.2µl syringe injections of thymine guanine repeated ssDNA ($(TG)_8$ ssDNA) at 60µM (Invitrogen). The stirring speed was set at 1000 rpm, and injection intervals set 180s apart. Anomalous points were removed and single-site boltzmann curve fitting performed by the integrated Microcal ITC200 software and Origin8 (OriginLabs).

12.18 DPS crystal screening and model refinement

Sitting drop vapour diffusion of gel filtrated and concentrated untagged DPS (3-6 mg/ml) from 60 and 90% AS fractions were screened. DPS protein and reservoir were incubated in a 1:1 v/v ratio with a total of 288 conditions from the Morpheus (Molecular Dimensions), PACT (Qiagen) and NATRIX (Hampton Research) commercial screens at 20˚C. Crystals were observed in each screen for both the 60 and 90% AS cuts. After two weeks crystals were large enough in PACT, with reservoir condition 53 (0.2M $NaNO_{2,}$ 20% w/v PEG 3350) for 90% AS cuts. Crystals were washed in 30% glycerol for cryoprotectant prior to crystal mounting and X-ray data collection.

Data collection was performed on the IO4 station at the Diamond light source (http://www.diamond.ac.uk/Beamlines/Mx/I04.html), Oxford, at 100 K using a Pilatus 6M-F detector at a wavelength set at 0.98 Å. The data were integrated using iMOSFLM [195] and scaled by SCALA [197]. Phasing was performed by molecular replacement using the PDB id: 1F30 crystallographic DPS model of *E.coli* by PHASER [207]. Model building was performed by COOT [199] and refined with REFMAC [200] which built a reasonable model at 2.74 Å with parameters described in 'Chapter IX'.

Appendices

Appendix 1 – Clustal X colour scheme

Colour	Residue at position	Conservation Rules
Blue	A,I,L,M,F,W,V,C	{+60%, WLVIMAFCHP}
Red	R,K	{+60%,KR},{+80%, K,R,Q}
Green	N	{+50%, N}, {+85%, N,Y}
Green	Q	{+60%,KR},{+50%,QE},{+85%,Q,E,K,R}
Green	S,T	{+60%, WLVIMAFCHP}, {+50%, TS}, {+85%,S,T}
Pink	C	{100%, C}
Magenta	E	{+60%,KR},{+50%,QE},{+85%,E,Q,D}
Magenta	D	{+60%,KR}, {+85%, K,R,Q}, {+50%,ED}
Orange	G	{+0%, G}
Cyan	H,Y	{+60%, WLVIMAFCHP}, {+85%,W,Y,A,C,P,Q,F,H,I,L,M,V}
Yellow	P	{+0%, P}

Residues are conserved if subsequent alignments succeed the minimum % quoted in brackets for the described amino acid. *Adapted from http://www.jalview.org/help/html/ colourSchemes/clustal.html.*

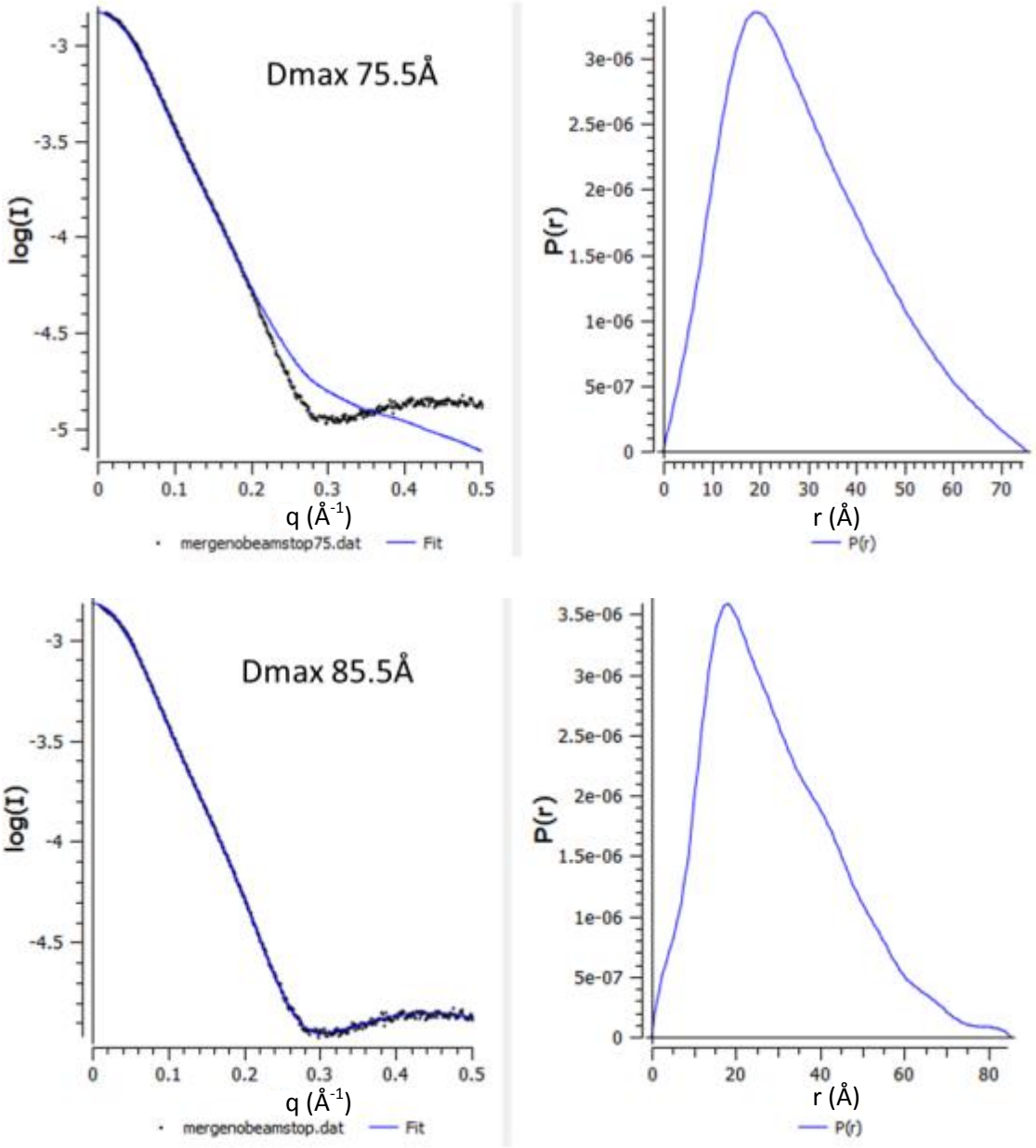
Appendix 2 – Bacterial strains

Strain	Source	Details
Omnimax2 (*E. Coli)*	Invitrogen	F´ {*pro*AB+ *lac*Iq *lac*ZΔM15 Tn10(TetR) Δ(*ccd*AB)} *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *end*A1 *rec*A1 *sup*E44 *thi*-1 *gyr*A96 *rel*A1 *ton*A *pan*D
BL21 DE3 (*E. coli*)	Novagen	F^- ompT gal dcm lon $hsdS_B(r_B^- m_B^-)$ λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])
M10616 (*E. coli*)	University of Liverpool - *Gifted*	F^- Δ(ara-leu)7697 $[araD139]_{B/r}$ Δ(codB-lacI)3 galK16 galE15 λ^- $e14^-$ mcrA0 relA1 rpsL150(strR) spoT1 mcrB1 hsdR2(r^-m^+)

Appendix 3 – Cloning and expression plasmids

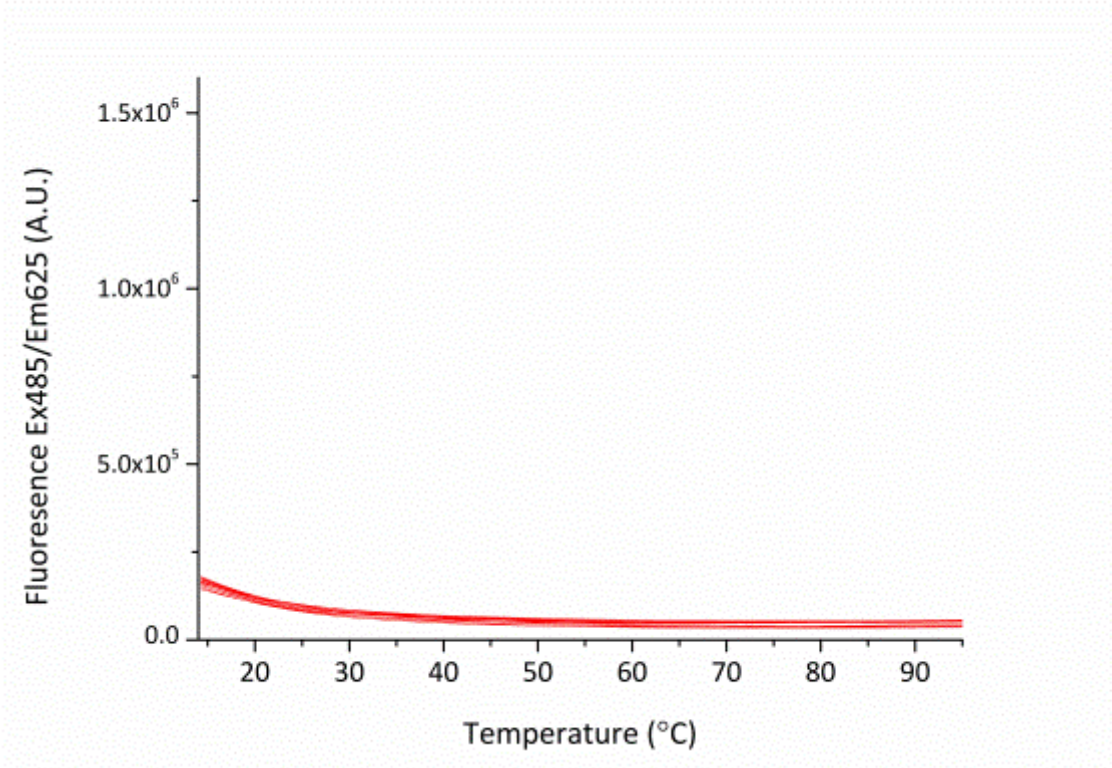
Name	Source	Details	Size (bp)	Selective Resistance
pGEX3X	GE Healthcare	N-terminal GFP tag Factor Xa cleavable	4,952	Ampicillin
pOPINF	OPPF	N-terminal HIS tag 3C protease cleavable	5,457	Ampicillin
pOPINEGFP	OPPF	C-terminal GFP tag 3C Protease cleavable	6,265	Ampicillin
mTDP-43SpGEX3X	N/A	Contains WT TDP-43S DNA sequence	5,158	Ampicillin
K263EmTDP-43SpGEX3X	N/A	Contains K263E TDP-43S mutation in DNA sequence	5,158	Ampicillin
D169GmTDP-43SpGEX3X	N/A	Contains D169E TDP-43S mutation in DNA sequence	5,158	Ampicillin
Mtdp-43f popinEGFP	N/A	Contains the TDP-43F DNA sequence	5,605	Ampicillin
e.coli DPS popin F	N/A	Contains the DPS gene for *E. coli*	5,758	Ampicillin
R. pickettii popin F	N/A	Contains the DPS gene for *R. pickettii*	5,685	Ampicillin

Appendix 4 – TDP-43S Dmax assignment



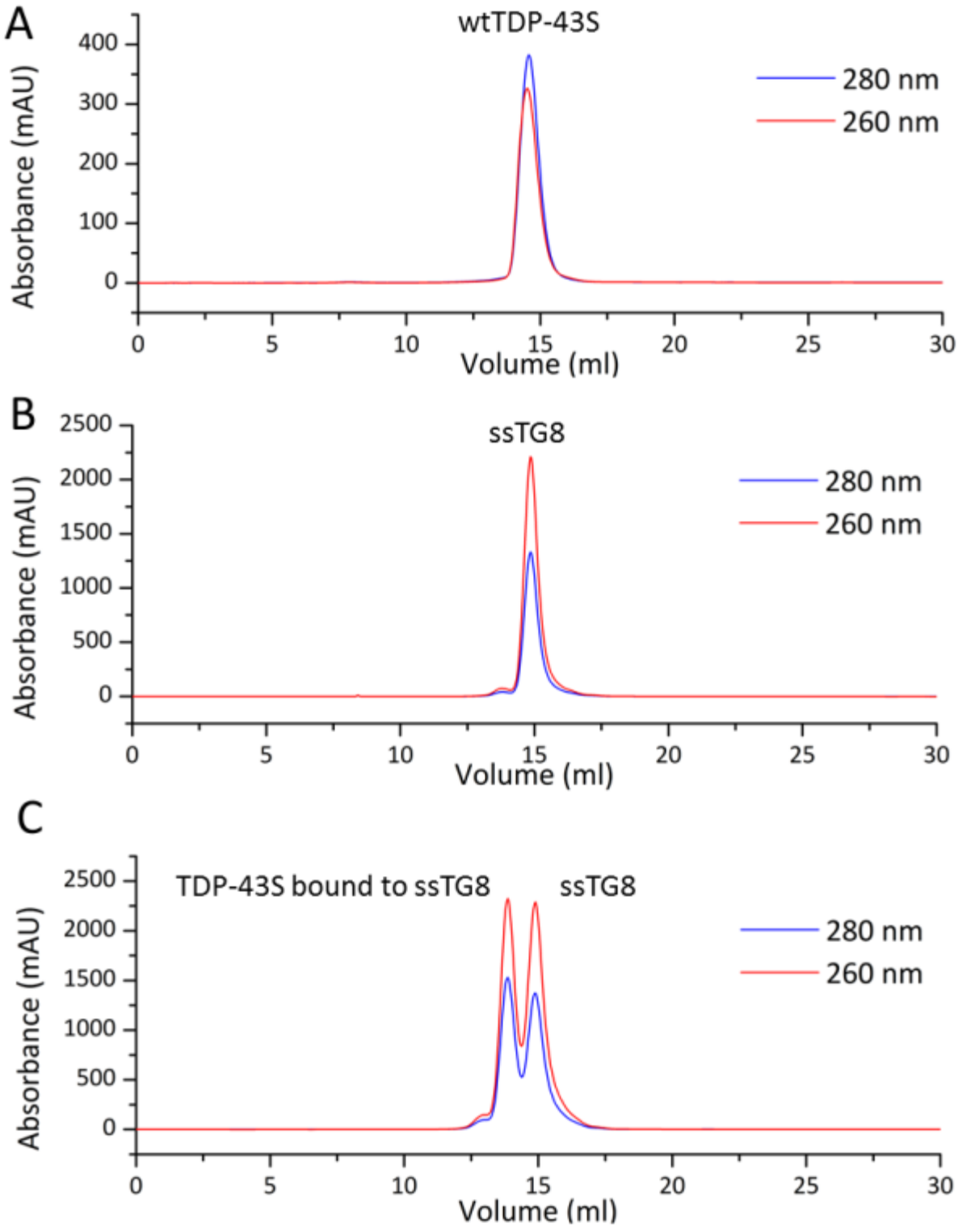
Wild-type TDP-43S P(r) Profiles and respective fits. Dmax comparisons at 75.5 and 85.5Å and their respective fits to the wild-type TDP-43S 1D profiles. Using points 1-925 using a lower 75.5 Å Dmax does not converse with the 1D data.

Appendix 5 – DSF blank



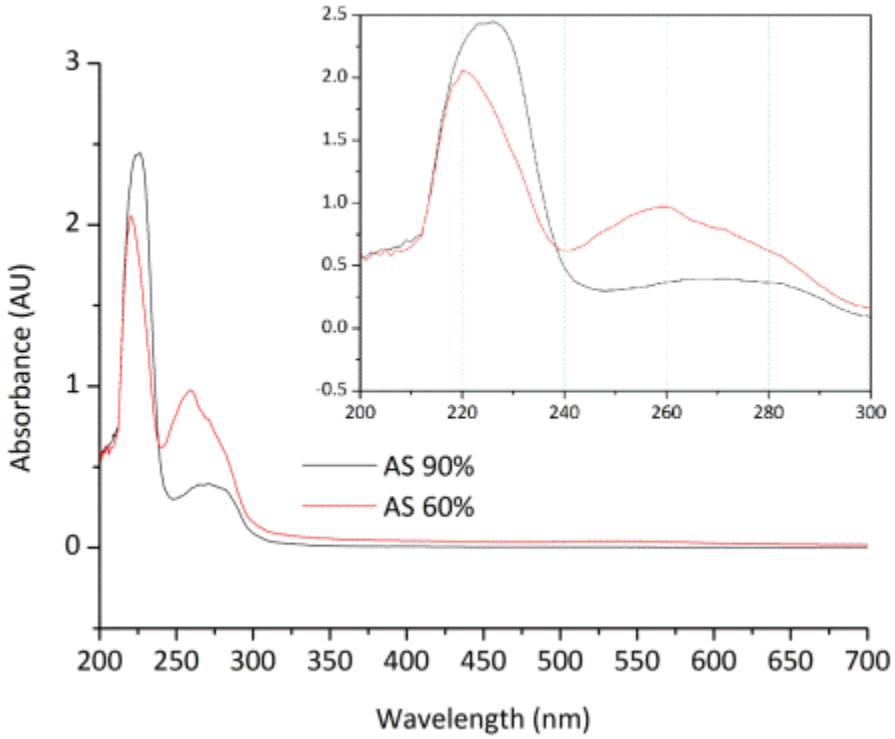
DSF Blank. DSF fluoresence of 10xSypro Orange in buffer without protein with 7 repeats showing that the dye had no affect to signal when elevating the temperature.

Appendix 6 – TDP-43S and ss$(TG)_8$ stoichiometry by gel filtration



TDP-43S and ss$(TG)_8$ stoichiometry by gel filtration. A, wtTDP-43S elution alone eluted at ~14 ml volume. **B,** Single stranded $(TG)_8$ alone eluted at a similar position ~14.5 ml of volume. **C,** The TDP-43S bound sample with excess $(TG)_8$ eluted with two peaks. This shows that only one wtTDP-43S molecule can bind $(TG)_8$.

Appendix 7 – Absorbance of DPS 60 and 90% AS cuts at different wavelengths



Wavelength spectra of DPS 60 and 90% AS cuts. The DPS samples were pooled, concentrated for crystal trials and diluted 1/20 for a spectrum to be tested. The AS60% shows a higher absorbance at 260nm then its 280nm which suggests nucleic acid contamination

Appendix 8 – Cloning plasmids

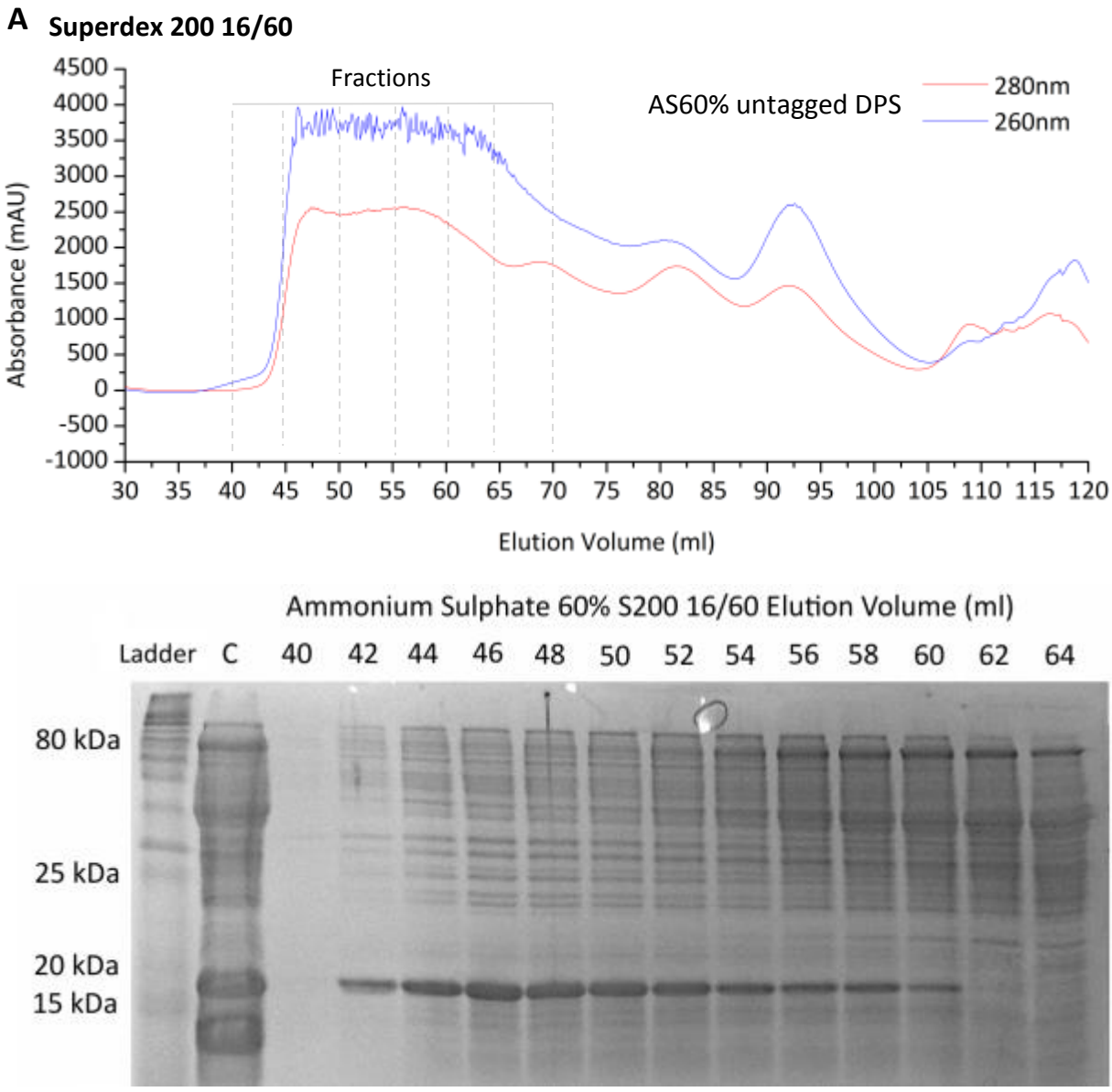
Cloning Primers	
pGEX3XmTDP-43SFwd	5’ - GCG GAT CCT TCA GAA AAC ATC TGA CCT CAT AGT GTT – 3’
pGEX3XmTDP-43SRvs	5’ - GAC GTC TTA ATT ATG CTT AGG TTC AGC ATT GGA TAT ATG CAC – 3’
MGD169GmTDP-43SFwd	5’ – TGTCACAACGACATATGATAGGTGGGCGATGGTGTGAC TGTAA – 3’
MGD169GmTDP-43SRvs	5’-TTACAGTCACACCATCGCCCACCTATCATATGTCGTTGTGACA – 3’
MGK263EmTDP-43SFwd	5’ - ATCCAATGCTGAACCTGAGCATAATTAAGACGTCA – 3’
MGK263EmTDP-43SRvs	5’ - TGACGTCTTAATTATGCTCAGGTTCAGCATTGGAT - 3’
*pOpinFE coli*DPSFwd	5’– AGTTCTGTTTCAGGGCCCGATGAGTACCGCTAAATTAGT TAAATCAAAAGCGAC – 3’
*pOpinFE.coli*DPSRvs	5’ – ATGGTCTAGAAAGCTTTATTCGATGTTAGACTCGATAAAC CACAG – 3’
*pOpineGFP*DPSFwd	5’-AGGAGATATACCATGAGTACCGCTAAATTAGTTAAATCAAAA-3’
*pOpineGFP*DPSRvs	5’- CAGAACTTCCAGTTTTTATTCGATGTTAGACTCGATAAA CCACAG-3’

Amino acids shown in red are the codons responsible for the mutant change.

Appendix 9 – DNA sequencing primers

Sequencing Primers	
pGEX3X Sequence Rvs	5’ – GACACGGAAATGTTGAATACTC – 3’
pGEX3X Sequence Fwd	5’ – GGAGCTGCATGTGTCAGAG – 3’
T7Minus1	5’ – AATACGACTCACTATAGGG – 3’

Appendix 10 - Aggregation of AS60% DPS as shown through gel filtration



DPS Aggregation of AS60% DPS as shown through gel filtration. The untagged DPS fractions ran on the superdex 200 (S200) 16/60 eluted near the maximum resolution limits of the column and with a substantial amount of protein contamination still remaining from the sample loaded (**C**). The same preparation was repeated on a sephacryl 300 column and was found to benefit from higher resolution and with the majority of protein contamination removed.

References

1. Neumann M, Sampathu DM, Kwong LK, Truax AC, Micsenyi MC, et al. (2006) Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science 314: 130–133. doi:10.1126/science.1134108.

2. Arai T, Hasegawa M, Akiyama H, Ikeda K, Nonaka T, et al. (2006) TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem Biophys Res Commun 351: 602–611. doi:10.1016/j.bbrc.2006.10.093.

3. Wang H-Y, Wang I-F, Bose J, Shen C-KJ (2004) Structural diversity and functional implications of the eukaryotic TDP gene family. Genomics 83: 130–139. doi:10.1016/S0888-7543(03)00214-3.

4. Neumann M, Kwong LK, Sampathu DM, Trojanowski JQ, Lee VM-Y (2007) TDP-43 proteinopathy in frontotemporal lobar degeneration and amyotrophic lateral sclerosis: protein misfolding diseases without amyloidosis. Arch Neurol 64: 1388–1394. doi:10.1001/archneur.64.10.1388.

5. Talbot K, Ansorge O (2006) Recent advances in the genetics of amyotrophic lateral sclerosis and frontotemporal dementia: common pathways in neurodegenerative disease. Hum Mol Genet 15 Spec No 2: R182–187. doi:10.1093/hmg/ddl202.

6. Rezania K, Roos RP (2013) Spinal cord: motor neuron diseases. Neurol Clin 31: 219–239. doi:10.1016/j.ncl.2012.09.014.

7. Silani V, Messina S, Poletti B, Morelli C, Doretti A, et al. (2011) The diagnosis of Amyotrophic lateral sclerosis in 2010. Arch Ital Biol 149: 5–27.

8. Simmons Z (2005) Management strategies for patients with amyotrophic lateral sclerosis from diagnosis through death. The neurologist 11: 257–270.

9. Neudert C, Oliver D, Wasner M, Borasio GD (2001) The course of the terminal phase in patients with amyotrophic lateral sclerosis. J Neurol 248: 612–616.

10. Schymick JC, Talbot K, Traynor BJ (2007) Genetics of sporadic amyotrophic lateral sclerosis. Hum Mol Genet 16 Spec No. 2: R233–242. doi:10.1093/hmg/ddm215.

11. Rabinovici GD, Miller BL (2010) Frontotemporal lobar degeneration: epidemiology, pathophysiology, diagnosis and management. CNS Drugs 24: 375–398. doi:10.2165/11533100-000000000-00000.

12. Snowden JS, Thompson JC, Stopford CL, Richardson AMT, Gerhard A, et al. (2011) The clinical diagnosis of early-onset dementias: diagnostic accuracy and clinicopathological relationships. Brain J Neurol 134: 2478–2492. doi:10.1093/brain/awr189.

13. Snowden JS, Neary D, Mann DMA (2002) Frontotemporal dementia. Br J Psychiatry J Ment Sci 180: 140–143.

14. Xu Z-S (2012) Does a loss of TDP-43 function cause neurodegeneration? Mol Neurodegener 7: 27. doi:10.1186/1750-1326-7-27.

15. Hirtz D, Thurman DJ, Gwinn-Hardy K, Mohamed M, Chaudhuri AR, et al. (2007) How common are the "common" neurologic disorders? Neurology 68: 326–337. doi:10.1212/01.wnl.0000252807.38124.a3.

16. Lattante S, Rouleau GA, Kabashi E (2013) TARDBP and FUS Mutations Associated with Amyotrophic Lateral Sclerosis: Summary and Update. Hum Mutat 34: 812–826. doi:10.1002/humu.22319.

17. Sreedharan J, Blair IP, Tripathi VB, Hu X, Vance C, et al. (2008) TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science 319: 1668–1672. doi:10.1126/science.1154584.

18. Rutherford NJ, Zhang Y-J, Baker M, Gass JM, Finch NA, et al. (2008) Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. PLoS Genet 4: e1000193. doi:10.1371/journal.pgen.1000193.

19. Borroni B, Archetti S, Del Bo R, Papetti A, Buratti E, et al. (2010) TARDBP mutations in frontotemporal lobar degeneration: frequency, clinical features, and disease course. Rejuvenation Res 13: 509–517. doi:10.1089/rej.2010.1017.

20. Prince M, Bryce R, Albanese E, Wimo A, Ribeiro W, et al. (2013) The global prevalence of dementia: A systematic review and metaanalysis. Alzheimers Dement 9: 63–75.e2. doi:10.1016/j.jalz.2012.11.007.

21. Foulds P, McAuley E, Gibbons L, Davidson Y, Pickering-Brown SM, et al. (2008) TDP-43 protein in plasma may index TDP-43 brain pathology in Alzheimer's disease and frontotemporal lobar degeneration. Acta Neuropathol (Berl) 116: 141–146. doi:10.1007/s00401-008-0389-8.

22. Nakashima-Yasuda H, Uryu K, Robinson J, Xie SX, Hurtig H, et al. (2007) Co-morbidity of TDP-43 proteinopathy in Lewy body related diseases. Acta Neuropathol (Berl) 114: 221–229. doi:10.1007/s00401-007-0261-2.

23. Buratti E, Brindisi A, Pagani F, Baralle FE (2004) Nuclear Factor TDP-43 Binds to the Polymorphic TG Repeats in CFTR Intron 8 and Causes Skipping of Exon 9: A Functional Link with Disease Penetrance. Am J Hum Genet 74: 1322–1325.

24. Fallini C, Bassell GJ, Rossoll W (2012) The ALS disease protein TDP-43 is actively transported in motor neuron axons and regulates axon outgrowth. Hum Mol Genet 21: 3703–3718. doi:10.1093/hmg/dds205.

25. Abhyankar MM, Urekar C, Reddi PP (2007) A novel CpG-free vertebrate insulator silences the testis-specific SP-10 gene in somatic tissues: role for TDP-43 in insulator function. J Biol Chem 282: 36143–36154. doi:10.1074/jbc.M705811200.

26. Ou SH, Wu F, Harrich D, García-Martínez LF, Gaynor RB (1995) Cloning and characterization of a novel cellular protein, TDP-43, that binds to human immunodeficiency virus type 1 TAR DNA sequence motifs. J Virol 69: 3584–3596.

27. Wang I-F, Wu L-S, Chang H-Y, Shen C-KJ (2008) TDP-43, the signature protein of FTLD-U, is a neuronal activity-responsive factor. J Neurochem 105: 797–806. doi:10.1111/j.1471-4159.2007.05190.x.

28. Colombrita C, Zennaro E, Fallini C, Weber M, Sommacal A, et al. (2009) TDP-43 is recruited to stress granules in conditions of oxidative insult. J Neurochem 111: 1051–1061. doi:10.1111/j.1471-4159.2009.06383.x.

29. Ayala YM, Misteli T, Baralle FE (2008) TDP-43 regulates retinoblastoma protein phosphorylation through the repression of cyclin-dependent kinase 6 expression. Proc Natl Acad Sci U S A 105: 3785–3789. doi:10.1073/pnas.0800546105.

30. Buratti E, Baralle FE (2001) Characterization and functional implications of the RNA binding properties of nuclear factor TDP-43, a novel splicing regulator of CFTR exon 9. J Biol Chem 276: 36337–36343. doi:10.1074/jbc.M104236200.

31. Fiesel FC, Voigt A, Weber SS, Van den Haute C, Waldenmaier A, et al. (2010) Knockdown of transactive response DNA-binding protein (TDP-43) downregulates histone deacetylase 6. EMBO J 29: 209–221. doi:10.1038/emboj.2009.324.

32. Polymenidou M, Lagier-Tourenne C, Hutt KR, Huelga SC, Moran J, et al. (2011) Long pre-mRNA depletion and RNA missplicing contribute to neuronal vulnerability from loss of TDP-43. Nat Neurosci 14: 459–468. doi:10.1038/nn.2779.

33. Tollervey JR, Curk T, Rogelj B, Briese M, Cereda M, et al. (2011) Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. Nat Neurosci 14: 452–458. doi:10.1038/nn.2778.

34. Sephton CF, Good SK, Atkin S, Dewey CM, Mayer P 3rd, et al. (2010) TDP-43 is a developmentally regulated protein essential for early embryonic development. J Biol Chem 285: 6826–6834. doi:10.1074/jbc.M109.061846.

35. Wu L-S, Cheng W-C, Hou S-C, Yan Y-T, Jiang S-T, et al. (2010) TDP-43, a neuro-pathosignature factor, is essential for early mouse embryogenesis. Genes N Y N 2000 48: 56–62. doi:10.1002/dvg.20584.

36. Kraemer BC, Schuck T, Wheeler JM, Robinson LC, Trojanowski JQ, et al. (2010) Loss of murine TDP-43 disrupts motor function and plays an essential role in embryogenesis. Acta Neuropathol (Berl) 119: 409–419. doi:10.1007/s00401-010-0659-0.

37. Li Y, Ray P, Rao EJ, Shi C, Guo W, et al. (2010) A Drosophila model for TDP-43 proteinopathy. Proc Natl Acad Sci U S A 107: 3169–3174. doi:10.1073/pnas.0913602107.

38. Feiguin F, Godena VK, Romano G, D'Ambrogio A, Klima R, et al. (2009) Depletion of TDP-43 affects Drosophila motoneurons terminal synapsis and locomotive behavior. FEBS Lett 583: 1586–1592. doi:10.1016/j.febslet.2009.04.019.

39. Kabashi E, Lin L, Tradewell ML, Dion PA, Bercier V, et al. (2010) Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo. Hum Mol Genet 19: 671–683. doi:10.1093/hmg/ddp534.

40. Elvira G, Wasiak S, Blandford V, Tong X-K, Serrano A, et al. (2006) Characterization of an RNA granule from developing brain. Mol Cell Proteomics MCP 5: 635–651. doi:10.1074/mcp.M500255-MCP200.

41. Ayala YM, De Conti L, Avendaño-Vázquez SE, Dhir A, Romano M, et al. (2011) TDP-43 regulates its mRNA levels through a negative feedback loop. EMBO J 30: 277–288. doi:10.1038/emboj.2010.310.

42. Lee EB, Lee VM-Y, Trojanowski JQ (2012) Gains or losses: molecular mechanisms of TDP43-mediated neurodegeneration. Nat Rev Neurosci 13: 38–50. doi:10.1038/nrn3121.

43. Volkening K, Leystra-Lantz C, Yang W, Jaffee H, Strong MJ (2009) Tar DNA binding protein of 43 kDa (TDP-43), 14-3-3 proteins and copper/zinc superoxide dismutase (SOD1) interact to modulate NFL mRNA stability. Implications for altered RNA processing in amyotrophic lateral sclerosis (ALS). Brain Res 1305: 168–182. doi:10.1016/j.brainres.2009.09.105.

44. Moisse K, Volkening K, Leystra-Lantz C, Welch I, Hill T, et al. (2009) Divergent patterns of cytosolic TDP-43 and neuronal progranulin expression following axotomy: implications for TDP-43 in the physiological response to neuronal injury. Brain Res 1249: 202–211. doi:10.1016/j.brainres.2008.10.021.

45. Sato T, Takeuchi S, Saito A, Ding W, Bamba H, et al. (2009) Axonal ligation induces transient redistribution of TDP-43 in brainstem motor neurons. Neuroscience 164: 1565–1578. doi:10.1016/j.neuroscience.2009.09.050.

46. Barmada SJ, Skibinski G, Korb E, Rao EJ, Wu JY, et al. (2010) Cytoplasmic Mislocalization of TDP-43 Is Toxic to Neurons and Enhanced by a Mutation Associated with Familial Amyotrophic Lateral Sclerosis. J Neurosci 30: 639–649. doi:10.1523/JNEUROSCI.4988-09.2010.

47. Igaz LM, Kwong LK, Lee EB, Chen-Plotkin A, Swanson E, et al. (2011) Dysregulation of the ALS-associated gene TDP-43 leads to neuronal death and degeneration in mice. J Clin Invest 121: 726–738. doi:10.1172/JCI44867.

48. Johnson BS, McCaffery JM, Lindquist S, Gitler AD (2008) A yeast TDP-43 proteinopathy model: Exploring the molecular determinants of TDP-43 aggregation and cellular toxicity. Proc Natl Acad Sci 105: 6439–6444. doi:10.1073/pnas.0802082105.

49. Johnson BS, Snead D, Lee JJ, McCaffery JM, Shorter J, et al. (2009) TDP-43 is intrinsically aggregation-prone, and amyotrophic lateral sclerosis-linked mutations accelerate aggregation and increase toxicity. J Biol Chem 284: 20329–20339. doi:10.1074/jbc.M109.010264.

50. Tatom JB, Wang DB, Dayton RD, Skalli O, Hutton ML, et al. (2009) Mimicking aspects of frontotemporal lobar degeneration and Lou Gehrig's disease in rats via TDP-43 overexpression. Mol Ther J Am Soc Gene Ther 17: 607–613. doi:10.1038/mt.2009.3.

51. Wegorzewska I, Bell S, Cairns NJ, Miller TM, Baloh RH (2009) TDP-43 mutant transgenic mice develop features of ALS and frontotemporal lobar degeneration. Proc Natl Acad Sci U S A 106: 18809–18814. doi:10.1073/pnas.0908767106.

52. McDonald KK, Aulas A, Destroismaisons L, Pickles S, Beleac E, et al. (2011) TAR DNA-binding protein 43 (TDP-43) regulates stress granule dynamics via differential regulation of G3BP and TIA-1. Hum Mol Genet 20: 1400–1410. doi:10.1093/hmg/ddr021.

53. Jiang L-L, Che M-X, Zhao J, Zhou C-J, Xie M-Y, et al. (2013) Structural transformation of the amyloidogenic core region of TDP-43 protein initiates its aggregation and cytoplasmic inclusion. J Biol Chem 288: 19614–19624. doi:10.1074/jbc.M113.463828.

54. Pesiridis GS, Lee VM-Y, Trojanowski JQ (2009) Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. Hum Mol Genet 18: R156–162. doi:10.1093/hmg/ddp303.

55. D'Ambrogio A, Buratti E, Stuani C, Guarnaccia C, Romano M, et al. (2009) Functional mapping of the interaction between TDP-43 and hnRNP A2 in vivo. Nucleic Acids Res 37: 4116–4126. doi:10.1093/nar/gkp342.

56. Watanabe S, Kaneko K, Yamanaka K (2013) Accelerated Disease Onset with Stabilized Familial Amyotrophic Lateral Sclerosis (ALS)-linked Mutant TDP-43 Proteins. J Biol Chem 288: 3641–3654. doi:10.1074/jbc.M112.433615.

57. Ling S-C, Albuquerque CP, Han JS, Lagier-Tourenne C, Tokunaga S, et al. (2010) ALS-associated mutations in TDP-43 increase its stability and promote TDP-43 complexes with FUS/TLS. Proc Natl Acad Sci 107: 13318–13323. doi:10.1073/pnas.1008227107.

58. Suzuki H, Lee K, Matsuoka M (2011) TDP-43-induced death is associated with altered regulation of BIM and Bcl-xL and attenuated by caspase-mediated TDP-43 cleavage. J Biol Chem 286: 13171–13183. doi:10.1074/jbc.M110.197483.

59. Che M-X, Jiang Y-J, Xie Y-Y, Jiang L-L, Hu H-Y (2011) Aggregation of the 35-kDa fragment of TDP-43 causes formation of cytoplasmic inclusions and alteration of RNA processing. FASEB J Off Publ Fed Am Soc Exp Biol 25: 2344–2353. doi:10.1096/fj.10-174482.

60. Wang Y-T, Kuo P-H, Chiang C-H, Liang J-R, Chen Y-R, et al. (2013) The truncated C-terminal RNA recognition motif of TDP-43 protein plays a key role in forming proteinaceous aggregates. J Biol Chem 288: 9049–9057. doi:10.1074/jbc.M112.438564.

61. Yang C, Tan W, Whittle C, Qiu L, Cao L, et al. (2010) The C-Terminal TDP-43 Fragments Have a High Aggregation Propensity and Harm Neurons by a Dominant-Negative Mechanism. PLoS ONE 5: e15878. doi:10.1371/journal.pone.0015878.

62. Chen AK-H, Lin RY-Y, Hsieh EZ-J, Tu P-H, Chen RP-Y, et al. (2010) Induction of Amyloid Fibrils by the C-Terminal Fragments of TDP-43 in Amyotrophic Lateral Sclerosis. J Am Chem Soc 132: 1186–1187. doi:10.1021/ja9066207.

63. Nonaka T, Masuda-Suzukake M, Arai T, Hasegawa Y, Akatsu H, et al. (2013) Prion-like Properties of Pathological TDP-43 Aggregates from Diseased Brains. Cell Rep 4: 124–134. doi:10.1016/j.celrep.2013.06.007.

64. Pesiridis GS, Tripathy K, Tanik S, Trojanowski JQ, Lee VM-Y (2011) A "two-hit" hypothesis for inclusion formation by carboxyl-terminal fragments of TDP-43 protein

linked to RNA depletion and impaired microtubule-dependent transport. J Biol Chem 286: 18845–18855. doi:10.1074/jbc.M111.231118.

65. Zhang Y-J, Caulfield T, Xu Y-F, Gendron TF, Hubbard J, et al. (2013) The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. Hum Mol Genet 22: 3112–3122. doi:10.1093/hmg/ddt166.

66. Winton MJ, Van Deerlin VM, Kwong LK, Yuan W, Wood EM, et al. (2008) A90V TDP-43 variant results in the aberrant localization of TDP-43 in vitro. FEBS Lett 582: 2252–2256. doi:10.1016/j.febslet.2008.05.024.

67. Neumann M, Kwong LK, Lee EB, Kremmer E, Flatley A, et al. (2009) Phosphorylation of S409/410 of TDP-43 is a consistent feature in all sporadic and familial forms of TDP-43 proteinopathies. Acta Neuropathol (Berl) 117: 137–149. doi:10.1007/s00401-008-0477-9.

68. Hasegawa M, Arai T, Nonaka T, Kametani F, Yoshida M, et al. (2008) Phosphorylated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Ann Neurol 64: 60–70. doi:10.1002/ana.21425.

69. Dormann D, Capell A, Carlson AM, Shankaran SS, Rodde R, et al. (2009) Proteolytic processing of TAR DNA binding protein-43 by caspases produces C-terminal fragments with disease defining properties independent of progranulin. J Neurochem 110: 1082–1094. doi:10.1111/j.1471-4159.2009.06211.x.

70. Zhang Y-J, Xu Y-F, Cook C, Gendron TF, Roettges P, et al. (2009) Aberrant cleavage of TDP-43 enhances aggregation and cellular toxicity. Proc Natl Acad Sci 106: 7607–7612. doi:10.1073/pnas.0900688106.

71. Zhang Y-J, Gendron TF, Xu Y-F, Ko L-W, Yen S-H, et al. (2010) Phosphorylation regulates proteasomal-mediated degradation and solubility of TAR DNA binding protein-43 C-terminal fragments. Mol Neurodegener 5: 33. doi:10.1186/1750-1326-5-33.

72. Deng H-X, Chen W, Hong S-T, Boycott KM, Gorrie GH, et al. (2011) Mutations in UBQLN2 cause dominant X-linked juvenile and adult-onset ALS and ALS/dementia. Nature 477: 211–215. doi:10.1038/nature10353.

73. Lee J-A, Gao F-B (2008) Roles of ESCRT in autophagy-associated neurodegeneration. Autophagy 4: 230–232.

74. Winton MJ, Igaz LM, Wong MM, Kwong LK, Trojanowski JQ, et al. (2008) Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. J Biol Chem 283: 13302–13309. doi:10.1074/jbc.M800342200.

75. Cassel JA, Reitz AB (2013) Ubiquilin-2 (UBQLN2) binds with high affinity to the C-terminal region of TDP-43 and modulates TDP-43 levels in H4 cells: characterization of inhibition by nucleic acids and 4-aminoquinolines. Biochim Biophys Acta 1834: 964–971. doi:10.1016/j.bbapap.2013.03.020.

76. Maris C, Dominguez C, Allain FH-T (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. FEBS J 272: 2118–2131. doi:10.1111/j.1742-4658.2005.04653.x.

77. Kuo P-H, Doudeva LG, Wang Y-T, Shen C-KJ, Yuan HS (2009) Structural insights into TDP-43 in nucleic-acid binding and domain interactions. Nucleic Acids Res 37: 1799–1808. doi:10.1093/nar/gkp013.

78. Swanson MS, Nakagawa TY, LeVan K, Dreyfuss G (1987) Primary structure of human nuclear ribonucleoprotein particle C proteins: conservation of sequence and domain structures in heterogeneous nuclear RNA, mRNA, and pre-rRNA-binding proteins. Mol Cell Biol 7: 1731–1739.

79. Adam SA, Nakagawa T, Swanson MS, Woodruff TK, Dreyfuss G (1986) mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. Mol Cell Biol 6: 2932–2943.

80. Dreyfuss G, Swanson MS, Piñol-Roma S (1988) Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. Trends Biochem Sci 13: 86–91. doi:10.1016/0968-0004(88)90046-1.

81. Ayala YM, Pantano S, D'Ambrogio A, Buratti E, Brindisi A, et al. (2005) Human, Drosophila, and C.elegans TDP43: nucleic acid binding properties and splicing regulatory function. J Mol Biol 348: 575–588. doi:10.1016/j.jmb.2005.02.038.

82. Kawahara Y, Mieda-Sato A (2012) TDP-43 promotes microRNA biogenesis as a component of the Drosha and Dicer complexes. Proc Natl Acad Sci U S A 109: 3347–3352. doi:10.1073/pnas.1112427109.

83. Ikai A (1980) Thermostability and aliphatic index of globular proteins. J Biochem (Tokyo) 88: 1895–1898.

84. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7. Available: http://msb.embopress.org/content/7/1/539. Accessed 5 February 2014.

85. Crystal structure of human UP1, the domain of hnRNP A1 that contains two RNA-recognition motifs (1997). Structure 5: 559–570. doi:10.1016/S0969-2126(97)00211-6.

86. Allain FH, Gilbert DE, Bouvet P, Feigon J (2000) Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target. J Mol Biol 303: 227–241. doi:10.1006/jmbi.2000.4118.

87. Crowder SM, Kanaar R, Rio DC, Alber T (1999) Absence of interdomain contacts in the crystal structure of the RNA recognition motifs of Sex-lethal. Proc Natl Acad Sci U S A 96: 4892–4897.

88. Cassel JA, Blass BE, Reitz AB, Pawlyk AC (2010) Development of a novel nonradiometric assay for nucleic acid binding to TDP-43 suitable for high-throughput screening using AlphaScreen technology. J Biomol Screen 15: 1099–1106. doi:10.1177/1087057110382778.

89. Kabashi E, Valdmanis PN, Dion P, Spiegelman D, McConkey BJ, et al. (2008) TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet 40: 572–574. doi:10.1038/ng.132.

90. Kovacs GG, Murrell JR, Horvath S, Haraszti L, Majtenyi K, et al. (2009) TARDBP variation associated with frontotemporal dementia, supranuclear gaze palsy, and chorea. Mov Disord Off J Mov Disord Soc 24: 1843–1847. doi:10.1002/mds.22697.

91. McLendon G, Radany E (1978) Is protein turnover thermodynamically controlled? J Biol Chem 253: 6335–6337.

92. Inoue I, Rechsteiner M (1994) On the relationship between the metabolic and thermodynamic stabilities of T4 lysozymes. Measurements in eukaryotic cells. J Biol Chem 269: 29247–29251.

93. Kwon WS, Da Silva NA, Kellis JT Jr (1996) Relationship between thermal stability, degradation rate and expression yield of barnase variants in the periplasm of Escherichia coli. Protein Eng 9: 1197–1202.

94. Hammarstrom M, Hellgren N, van den Berg S, Berglund H, Hard T (2002) Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli. Protein Sci Publ Protein Soc 11: 313–321.

95. Luo ZH, Hua ZC (1998) Increased solubility of glutathione S-transferase-P16 (GST-p16) fusion protein by co-expression of chaperones groes and groel in Escherichia coli. Biochem Mol Biol Int 46: 471–477.

96. Smyth DR, Mrozkiewicz MK, McGrath WJ, Listwan P, Kobe B (2003) Crystal structures of fusion proteins with large-affinity tags. Protein Sci Publ Protein Soc 12: 1313–1322. doi:10.1110/ps.0243403.

97. Smith DB, Johnson KS (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. Gene 67: 31–40.

98. Kuge M, Fujii Y, Shimizu T, Hirose F, Matsukage A, et al. (1997) Use of a fusion protein to obtain crystals suitable for X-ray analysis: crystallization of a GST-fused protein containing the DNA-binding domain of DNA replication-related element-binding factor, DREF. Protein Sci Publ Protein Soc 6: 1783–1786. doi:10.1002/pro.5560060822.

99. De Boer HA, Comstock LJ, Vasser M (1983) The tac promoter: a functional hybrid derived from the trp and lac promoters. Proc Natl Acad Sci U S A 80: 21–25.

100. Warburg O, Christian W (1942) Insulation and crystalisation of the fermenting process of Enolase. Biochem Z 310: 384–421.

101. Sambrook J (2001) Molecular Cloning: A Laboratory Manual, Third Edition. 3rd ed. Cold Spring Harbor Laboratory Press. 2344 p.

102. Chang C, Wu T-H, Wu C-Y, Chiang M, Toh EK-W, et al. (2012) The N-terminus of TDP-43 promotes its oligomerization and enhances DNA binding affinity. Biochem Biophys Res Commun 425: 219–224. doi:10.1016/j.bbrc.2012.07.071.

103. De Yoreo JJ (2003) Principles of Crystal Nucleation and Growth. Rev Mineral Geochem 54: 57–93. doi:10.2113/0540057.

104. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006) *ATSAS* 2.1, a program package for small-angle scattering data analysis. J Appl Crystallogr 39: 277–286. doi:10.1107/S0021889806004699.

105. Svergun DI, Koch MHJ (2003) Small-angle scattering studies of biological macromolecules in solution. Rep Prog Phys 66: 1735. doi:10.1088/0034-4885/66/10/R05.

106. David G, Pérez J (2009) Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. J Appl Crystallogr 42: 892–900. doi:10.1107/S0021889809029288.

107. Debye P. (1915) Scattering from non-crystalline substances. Ann Phys 46: 809–823.

108. Jacques DA, Trewhella J (2010) Small-angle scattering for structural biology--Expanding the frontier while avoiding the pitfalls. Protein Sci Publ Protein Soc 19: 642–657. doi:10.1002/pro.351.

109. Moncoq K, Broutin I, Craescu CT, Vachette P, Ducruix A, et al. (2004) SAXS study of the PIR domain from the Grb14 molecular adaptor: a natively unfolded protein with a transient structure primer? Biophys J 87: 4056–4064. doi:10.1529/biophysj.104.048645.

110. Receveur-Brechot V, Durand D (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. Curr Protein Pept Sci 13: 55–75.

111. Svergun DI (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J Appl Crystallogr 25: 495–503. doi:10.1107/S0021889892001663.

112. Ciccariello S, Schneider J-M, Schönfeld B, Kostorz G (2000) Generalization of Porod's law of small-angle scattering to anisotropic samples. EPL Europhys Lett 50: 601. doi:10.1209/epl/i2000-00312-y.

113. Durand D, Vivès C, Cannella D, Pérez J, Pebay-Peyroula E, et al. (2010) NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. J Struct Biol 169: 45–53. doi:10.1016/j.jsb.2009.08.009.

114. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. Biophys J 80: 2946–2953.

115. Petoukhov MV, Svergun DI (2005) Global rigid body modeling of macromolecular complexes against small-angle scattering data. Biophys J 89: 1237–1250. doi:10.1529/biophysj.105.064154.

116. Maupetit J, Gautier R, Tufféry P (2006) SABBAC: online Structural Alphabet-based protein BackBone reconstruction from Alpha-Carbon trace. Nucleic Acids Res 34: W147–151. doi:10.1093/nar/gkl289.

117. Svergun D, Barberato C, Koch MHJ (1995) *CRYSOL* – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. J Appl Crystallogr 28: 768–773. doi:10.1107/S0021889895007047.

118. Bernadó P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. J Am Chem Soc 129: 5656–5664. doi:10.1021/ja069124n.

119. Kim HS, Wilce MCJ, Yoga YMK, Pendini NR, Gunzburg MJ, et al. (2011) Different modes of interaction by TIAR and HuR with target RNA and DNA. Nucleic Acids Res 39: 1117–1130. doi:10.1093/nar/gkq837.

120. Kelly SM, Jess TJ, Price NC (2005) How to study proteins by circular dichroism. Biochim Biophys Acta 1751: 119–139. doi:10.1016/j.bbapap.2005.06.005.

121. Whitmore L, Wallace BA (2008) Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. Biopolymers 89: 392–400. doi:10.1002/bip.20853.

122. Manavalan P, Johnson WC Jr (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. Anal Biochem 167: 76–85.

123. Compton LA, Johnson WC Jr (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Anal Biochem 155: 155–167.

124. Sreerama N, Woody RW (2000) Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. Anal Biochem 287: 252–260. doi:10.1006/abio.2000.4880.

125. Van Stokkum IH, Spoelder HJ, Bloemendal M, van Grondelle R, Groen FC (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. Anal Biochem 191: 110–118.

126. Sreerama N, Woody RW (1993) A self-consistent method for the analysis of protein secondary structure from circular dichroism. Anal Biochem 209: 32–44. doi:10.1006/abio.1993.1079.

127. Brahms S, Brahms J (1980) Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. J Mol Biol 138: 149–178. doi:10.1016/0022-2836(80)90282-X.

128. Dhatwalia R, Singh H, Oppenheimer M, Karr DB, Nix JC, et al. (2012) Crystal structures and small-angle x-ray scattering analysis of UDP-galactopyranose mutase from the pathogenic fungus Aspergillus fumigatus. J Biol Chem 287: 9041–9051. doi:10.1074/jbc.M111.327536.

129. Vestergaard B, Sanyal S, Roessle M, Mora L, Buckingham RH, et al. (2005) The SAXS solution structure of RF1 differs from its crystal structure and is similar to its ribosome bound cryo-EM structure. Mol Cell 20: 929–938. doi:10.1016/j.molcel.2005.11.022.

130. Petoukhov MV, Monie TP, Allain FH-T, Matthews S, Curry S, et al. (2006) Conformation of polypyrimidine tract binding protein in solution. Struct Lond Engl 1993 14: 1021–1027. doi:10.1016/j.str.2006.04.005.

131. Receveur V, Czjzek M, Schülein M, Panine P, Henrissat B (2002) Dimension, shape, and conformational flexibility of a two domain fungal cellulase in solution probed by small angle X-ray scattering. J Biol Chem 277: 40887–40892. doi:10.1074/jbc.M205404200.

132. Ghosh A, Hartung S, van der Does C, Tainer JA, Albers S-V (2011) Archaeal flagellar ATPase motor shows ATP-dependent hexameric assembly and activity stimulation by specific lipid binding. Biochem J 437: 43–52. doi:10.1042/BJ20110410.

133. Volkov VV, Svergun DI (2003) Uniqueness of *ab initio* shape determination in small-angle scattering. J Appl Crystallogr 36: 860–864. doi:10.1107/S0021889803000268.

134. Fischer H, de Oliveira Neto M, Napolitano HB, Polikarpov I, Craievich AF (2009) Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J Appl Crystallogr 43: 101–109. doi:10.1107/S0021889809043076.

135. Wang Y-F, Ding J, Wang F, Bu D-F (2004) [Proteins structure changes of COL4A5 gene point mutations and structure-phenotype relations in Alport syndrome]. Zhonghua Er Ke Za Zhi Chin J Pediatr 42: 589–594.

136. Lees JG, Miles AJ, Wien F, Wallace BA (2006) A reference database for circular dichroism spectroscopy covering fold and secondary structure space. Bioinforma Oxf Engl 22: 1955–1962. doi:10.1093/bioinformatics/btl327.

137. Greenfield NJ, Fasman GD (1969) Computed circular dichroism spectra for the evaluation of protein conformation. Biochemistry (Mosc) 8: 4108–4116. doi:10.1021/bi00838a031.

138. Greenfield NJ (2006) Using circular dichroism spectra to estimate protein secondary structure. Nat Protoc 1: 2876–2890. doi:10.1038/nprot.2006.202.

139. Anderluh G, Gökçe I, Lakey JH (2004) A natively unfolded toxin domain uses its receptor as a folding template. J Biol Chem 279: 22002–22009. doi:10.1074/jbc.M313603200.

140. Tam S, Spiess C, Auyeung W, Joachimiak L, Chen B, et al. (2009) The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. Nat Struct Mol Biol 16: 1279–1285. doi:10.1038/nsmb.1700.

141. Andersson FI, Werrell EF, McMorran L, Crone WJK, Das C, et al. (2011) The effect of Parkinson's-disease-associated mutations on the deubiquitinating enzyme UCH-L1. J Mol Biol 407: 261–272. doi:10.1016/j.jmb.2010.12.029.

142. Ni C-L, Shi H-P, Yu H-M, Chang Y-C, Chen Y-R (2011) Folding stability of amyloid-beta 40 monomer is an important determinant of the nucleation kinetics in fibrillization. FASEB J Off Publ Fed Am Soc Exp Biol 25: 1390–1401. doi:10.1096/fj.10-175539.

143. Vassall KA, Stubbs HR, Primmer HA, Tong MS, Sullivan SM, et al. (2011) Decreased stability and increased formation of soluble aggregates by immature superoxide dismutase do not account for disease severity in ALS. Proc Natl Acad Sci. Available: http://www.pnas.org/content/early/2011/01/20/0913021108. Accessed 14 July 2013.

144. Münch C, Bertolotti A (2010) Exposure of hydrophobic surfaces initiates aggregation of diverse ALS-causing superoxide dismutase-1 mutants. J Mol Biol 399: 512–525. doi:10.1016/j.jmb.2010.04.019.

145. Soto C (2003) Unfolding the role of protein misfolding in neurodegenerative diseases. Nat Rev Neurosci 4: 49–60. doi:10.1038/nrn1007.

146. Uniewicz KA, Ori A, Xu R, Ahmed Y, Wilkinson MC, et al. (2010) Differential scanning fluorimetry measurement of protein stability changes upon binding to glycosaminoglycans: a screening test for binding specificity. Anal Chem 82: 3796–3802. doi:10.1021/ac100188x.

147. Teale FWJ, Weber G (1957) Ultraviolet fluorescence of the aromatic amino acids. Biochem J 65: 476–482.

148. Eftink MR (1994) The use of fluorescence methods to monitor unfolding transitions in proteins. Biophys J 66: 482–501.

149. Lakowicz JR (2006) Principles of fluorescence spectroscopy. New York, N.Y.: Springer.

150. Vivian JT, Callis PR (2001) Mechanisms of Tryptophan Fluorescence Shifts in Proteins. Biophys J 80: 2093–2109. doi:10.1016/S0006-3495(01)76183-8.

151. Wiseman T, Williston S, Brandts JF, Lin LN (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. Anal Biochem 179: 131–137.

152. Todd MJ, Gomez J (2001) Enzyme Kinetics Determined Using Calorimetry: A General Assay for Enzyme Activity? Anal Biochem 296: 179–187. doi:10.1006/abio.2001.5218.

153. C. J, Murciano-Calles J, S. E, Iglesias-Bexiga M, Luque I, et al. (2013) Isothermal Titration Calorimetry: Thermodynamic Analysis of the Binding Thermograms of Molecular Recognition Events by Using Equilibrium Models. In: Elkordy AA, editor. Applications of Calorimetry in a Wide Context - Differential Scanning Calorimetry, Isothermal Titration Calorimetry and Microcalorimetry. InTech. Available: http://www.intechopen.com/books/howtoreference/applications-of-calorimetry-in-a-wide-context-differential-scanning-calorimetry-isothermal-titration-calorimetry-and-microcalorimetry/isothermal-titration-calorimetry-thermodynamic-analysis-of-the-binding-thermograms-of-molecular-reco. Accessed 30 May 2013.

154. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. Nat Protoc 2: 2212–2221. doi:10.1038/nprot.2007.321.

155. Oh YK, Shin KS, Yuan J, Kang SJ (2008) Superoxide dismutase 1 mutants related to amyotrophic lateral sclerosis induce endoplasmic stress in neuro2a cells. J Neurochem 104: 993–1005. doi:10.1111/j.1471-4159.2007.05053.x.

156. Raychaudhuri S, Sinha M, Mukhopadhyay D, Bhattacharyya NP (2008) HYPK, a Huntingtin interacting protein, reduces aggregates and apoptosis induced by N-terminal Huntingtin with 40 glutamines in Neuro2a cells and exhibits chaperone-like activity. Hum Mol Genet 17: 240–255. doi:10.1093/hmg/ddm301.

157. Uryu M, Karino A, Kamihara Y, Horiuchi M (2007) Characterization of prion susceptibility in Neuro2a mouse neuroblastoma cell subclones. Microbiol Immunol 51: 661–669.

158. Shinjyo N, Kita K (2007) Relationship between reactive oxygen species and heme metabolism during the differentiation of Neuro2a cells. Biochem Biophys Res Commun 358: 130–135. doi:10.1016/j.bbrc.2007.04.071.

159. Jäger M, Dendle M, Kelly JW (2009) Sequence determinants of thermodynamic stability in a WW domain--an all-beta-sheet protein. Protein Sci Publ Protein Soc 18: 1806–1813. doi:10.1002/pro.172.

160. Lukavsky PJ, Daujotyte D, Tollervey JR, Ule J, Stuani C, et al. (2013) Molecular basis of UG-rich RNA recognition by the human splicing factor TDP-43. Nat Struct Mol Biol 20: 1443–1449. doi:10.1038/nsmb.2698.

161. Almirón M, Link AJ, Furlong D, Kolter R (1992) A novel DNA-binding protein with regulatory and protective roles in starved Escherichia coli. Genes Dev 6: 2646–2654.

162. Battesti A, Majdalani N, Gottesman S (2011) The RpoS-mediated general stress response in Escherichia coli. Annu Rev Microbiol 65: 189–213. doi:10.1146/annurev-micro-090110-102946.

163. Altuvia S, Almiron M, Huisman G, Kolter R, Storz G (1994) The dps promoter is activated by OxyR during growth and by IHF and σs in stationary phase. Mol Microbiol 13: 265–272. doi:10.1111/j.1365-2958.1994.tb00421.x.

164. Zhang A, Altuvia S, Tiwari A, Argaman L, Hengge-Aronis R, et al. (1998) The OxyS regulatory RNA represses rpoS translation and binds the Hfq (HF-I) protein. EMBO J 17: 6061–6068. doi:10.1093/emboj/17.20.6061.

165. Weber H, Polen T, Heuveling J, Wendisch VF, Hengge R (2005) Genome-Wide Analysis of the General Stress Response Network in Escherichia coli: ?S-Dependent Genes, Promoters, and Sigma Factor Selectivity. J Bacteriol 187: 1591–1603. doi:10.1128/JB.187.5.1591-1603.2005.

166. Loewen PC, Hu B, Strutinsky J, Sparling R (1998) Regulation in the rpoS regulon of Escherichia coli. Can J Microbiol 44: 707–717.

167. Ali Azam T, Iwata A, Nishimura A, Ueda S, Ishihama A (1999) Growth phase-dependent variation in protein composition of the Escherichia coli nucleoid. J Bacteriol 181: 6361–6370.

168. Haikarainen T, Papageorgiou AC (2010) Dps-like proteins: structural and functional insights into a versatile protein family. Cell Mol Life Sci CMLS 67: 341–351. doi:10.1007/s00018-009-0168-2.

169. Ceci P, Cellai S, Falvo E, Rivetti C, Rossi GL, et al. (2004) DNA condensation and self-aggregation of Escherichia coli Dps are coupled phenomena related to the properties of the N-terminus. Nucleic Acids Res 32: 5935–5944. doi:10.1093/nar/gkh915.

170. Wolf SG, Frenkiel D, Arad T, Finkel SE, Kolter R, et al. (1999) DNA protection by stress-induced biocrystallization. Nature 400: 83–85. doi:10.1038/21918.

171. Papinutto E, Dundon WG, Pitulis N, Battistutta R, Montecucco C, et al. (2002) Structure of two iron-binding proteins from Bacillus anthracis. J Biol Chem 277: 15093–15098. doi:10.1074/jbc.M112378200.

172. Bozzi M, Mignogna G, Stefanini S, Barra D, Longhi C, et al. (1997) A novel non-heme iron-binding ferritin related to the DNA-binding proteins of the Dps family in Listeria innocua. J Biol Chem 272: 3259–3265.

173. Pierpaolo Ceci AI (2003) The Dps Protein of Agrobacterium tumefaciens Does Not Bind to DNA but Protects It toward Oxidative Cleavage. J Biol Chem 278: 20319–20326.

174. Ceci P, Ilari A, Falvo E, Giangiacomo L, Chiancone E (2005) Reassessment of protein stability, DNA binding, and protection of Mycobacterium smegmatis Dps. J Biol Chem 280: 34776–34785. doi:10.1074/jbc.M502343200.

175. Ceci P, Mangiarotti L, Rivetti C, Chiancone E (2007) The neutrophil-activating Dps protein of Helicobacter pylori, HP-NAP, adopts a mechanism different from Escherichia coli Dps to bind and condense DNA. Nucleic Acids Res 35: 2247–2256. doi:10.1093/nar/gkm077.

176. Chiancone E, Ceci P (2010) The multifaceted capacity of Dps proteins to combat bacterial stress conditions: Detoxification of iron and hydrogen peroxide and DNA binding. Biochim Biophys Acta 1800: 798–805. doi:10.1016/j.bbagen.2010.01.013.

177. Oexle H, Gnaiger E, Weiss G (1999) Iron-dependent changes in cellular energy metabolism: influence on citric acid cycle and oxidative phosphorylation. Biochim Biophys Acta 1413: 99–107.

178. Doherty CP (2007) Host-pathogen interactions: the role of iron. J Nutr 137: 1341–1344.

179. Andrews SC (1998) Iron storage in bacteria. Adv Microb Physiol 40: 281–351.

180. Zhao G, Ceci P, Ilari A, Giangiacomo L, Laue TM, et al. (2002) Iron and hydrogen peroxide detoxification properties of DNA-binding protein from starved cells. A ferritin-like DNA-binding protein of Escherichia coli. J Biol Chem 277: 27689–27696. doi:10.1074/jbc.M202094200.

181. Nair S, Finkel SE (2004) Dps protects cells against multiple stresses during stationary phase. J Bacteriol 186: 4192–4198. doi:10.1128/JB.186.13.4192-4198.2004.

182. Ilari A, Ceci P, Ferrari D, Rossi GL, Chiancone E (2002) Iron incorporation into Escherichia coli Dps gives rise to a ferritin-like microcrystalline core. J Biol Chem 277: 37619–37623. doi:10.1074/jbc.M206186200.

183. Abdul-Tehrani H, Hudson AJ, Chang YS, Timms AR, Hawkins C, et al. (1999) Ferritin mutants of Escherichia coli are iron deficient and growth impaired, and fur mutants are iron deficient. J Bacteriol 181: 1415–1428.

184. Calhoun LN, Kwon YM (2011) Structure, function and regulation of the DNA-binding protein Dps and its role in acid and oxidative stress resistance in Escherichia coli: a review. J Appl Microbiol 110: 375–386. doi:10.1111/j.1365-2672.2010.04890.x.

185. Grant RA, Filman DJ, Finkel SE, Kolter R, Hogle JM (1998) The crystal structure of Dps, a ferritin homolog that binds and protects DNA. Nat Struct Mol Biol 5: 294–303. doi:10.1038/nsb0498-294.

186. Frenkiel-Krispin D, Levin-Zaidman S, Shimoni E, Wolf SG, Wachtel EJ, et al. (2001) Regulated phase transitions of bacterial chromatin: a non-enzymatic pathway for generic DNA protection. EMBO J 20: 1184–1191. doi:10.1093/emboj/20.5.1184.

187. Gautam A, Vinson HM, Gibbs PS, Olet S, Barigye R (2011) Proteomic analysis of multidrug resistant Escherichia coli strains from scouring calves. Vet Microbiol 151: 363–371. doi:10.1016/j.vetmic.2011.03.032.

188. Sriubolmas N, Panbangred W, Sriurairatana S, Meevootisom V (1997) Localization and characterization of inclusion bodies in recombinant Escherichia coli cells overproducing penicillin G acylase. Appl Microbiol Biotechnol 47: 373–378.

189. Ishikawa T, Mizunoe Y, Kawabata S, Takade A, Harada M, et al. (2003) The iron-binding protein Dps confers hydrogen peroxide stress resistance to Campylobacter jejuni. J Bacteriol 185: 1010–1017.

190. Park S, Terzic A (2010) QUATERNARY STRUCTURE OF KATP CHANNEL SUR2A NUCLEOTIDE BINDING DOMAINS RESOLVED BY SYNCHROTRON RADIATION X-RAY SCATTERING. J Struct Biol 169: 243. doi:10.1016/j.jsb.2009.11.005.

191. Rhodes G (2006) Chapter 2 - An Overview of Protein Crystallography. Crystallography Made Crystal Clear (Third Edition). Burlington: Academic Press. pp. 7–30. Available: http://www.sciencedirect.com/science/article/pii/B9780125870733500040. Accessed 23 July 2013.

192. Massa PDW (2004) Introduction. Crystal Structure Determination. Springer Berlin Heidelberg. pp. 1–1. Available: http://link.springer.com/chapter/10.1007/978-3-662-06431-3_1. Accessed 20 February 2014.

193. Crystal Structure Analysis: Principles and Practice (International Union of Crystallography Texts on Crystallogra) by Main, Peter; William, Clegg; Blake, Alexander J.; Gould, Robert O.: Oxford University Press, USA 9780198506188 Hardcover - AbeBooks.com (n.d.). Available: http://www.abebooks.com/Crystal-Structure-Analysis-Principles-Practice-International/11041148733/bd. Accessed 20 February 2014.

194. Cowtan K (2001) Phase Problem in X-ray Crystallography, and Its Solution. eLS. John Wiley & Sons, Ltd. Available: http://onlinelibrary.wiley.com/doi/10.1038/npg.els.0002722/abstract. Accessed 10 October 2013.

195. Evolving Methods for Macromolecular Crystallography - The Structural Path to the Understanding (n.d.). Available: http://www.springer.com/materials/book/978-1-4020-6314-5. Accessed 4 November 2013.

196. Otwinowski Z, Minor W (1997) [20] Processing of X-ray diffraction data collected in oscillation mode. Methods in Enzymology. Elsevier, Vol. 276. pp. 307–326. Available: https://www.google.co.uk/search?q=Processing+of+X-ray+Diffraction+Data+Collected+in+Oscillation+Mode&ie=utf-8&oe=utf-8&aq=t&rls=org.mozilla:en-GB:official&client=firefox-a&channel=fflb. Accessed 5 July 2013.

197. Evans P (2006) Scaling and assessment of data quality. Acta Crystallogr D Biol Crystallogr 62: 72–82. doi:10.1107/S0907444905036693.

198. Vagin A, Teplyakov A (1997) *MOLREP* : an Automated Program for Molecular Replacement. J Appl Crystallogr 30: 1022–1025. doi:10.1107/S0021889897006766.

199. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of *Coot*. Acta Crystallogr D Biol Crystallogr 66: 486–501. doi:10.1107/S0907444910007493.

200. Murshudov GN, Skubák P, Lebedev AA, Pannu NS, Steiner RA, et al. (2011) *REFMAC* 5 for the refinement of macromolecular crystal structures. Acta Crystallogr D Biol Crystallogr 67: 355–367. doi:10.1107/S0907444911001314.

201. Smyth MS, Martin JH (2000) x ray crystallography. Mol Pathol MP 53: 8–14.

202. Roy S, Gupta S, Das S, Sekar K, Chatterji D, et al. (2004) X-ray analysis of Mycobacterium smegmatis Dps and a comparative study involving other Dps and Dps-like molecules. J Mol Biol 339: 1103–1113. doi:10.1016/j.jmb.2004.04.042.

203. Konarev PV, Volkov VV, Sokolova AV, Koch MHJ, Svergun DI (2003) *PRIMUS* : a Windows PC-based system for small-angle scattering data analysis. J Appl Crystallogr 36: 1277–1282. doi:10.1107/S0021889803012779.

204. Hammersley AP (1997) FIT2D: An introduction and overview.

205. Giuliani A, Jamme F, Rouam V, Wien F, Giorgetta J-L, et al. (2009) DISCO: a low-energy multipurpose beamline at synchrotron SOLEIL. J Synchrotron Radiat 16: 835–841. doi:10.1107/S0909049509034049.

206. Mahmood T, Yang P-C (2012) Western Blot: Technique, Theory, and Trouble Shooting. North Am J Med Sci 4: 429–434. doi:10.4103/1947-2714.100998.

207. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, et al. (2007) *Phaser* crystallographic software. J Appl Crystallogr 40: 658–674. doi:10.1107/S0021889807021206.