

University of Kent

MSc by Research Thesis

Biochemical and Biophysical Characterisation of Transcription Regulators Associated with Myosin VI

Author – Hannah Caroline Wendy Reed

Degree – MSc-R Biochemistry

Word Count (Excluding References) – 23,501

Thesis Advisor – Dr Christopher Toseland

2019

1 Declaration

No part of this thesis has been submitted in support of an application for any degree or qualification of the University of Kent or any other University or institute of learning.

Hannah Reed



August 2019

2 Acknowledgements

I would like to take this opportunity to thank my MSc by Research supervisor – Dr Christopher Toseland – for his academic support and patience. I am particularly grateful for opportunities to attend conferences, present posters and be involved with both national and international collaboration work. I believe these experiences have been instrumental in helping me to secure a PhD position and in developing my confidence both in and outside the laboratory. I would also like to thank you for your encouragement to continue my education onto PhD level and for all of your help with applications and references.

My sincerest gratitude also extends to members of the Toseland laboratory, namely Ália dos Santos, Alexander Cook, Yukti Hari-Gupta, Natalia Fili and Rosemarie Gough, who have made me feel extremely welcome and have provided endless academic and practical advice. I am proud to call you all good friends and am very grateful for your support. I would like to thank Ália dos Santos in particular for her guidance and support whilst working with nuclear dot protein 52 (NDP52). Again, I would also like to thank her for her personal encouragement, support during PhD applications and for being a good friend.

I would also like to thank members of the Moreno-Herrero laboratory at the Spanish National Centre for Biotechnology for making us feel welcome during our visit to Madrid and teaching us specialist techniques, namely magnetic tweezers. I would especially like to thank Carolina Carrasco and Clara Aicart-Ramos. I learnt a lot during my time here and am extremely grateful for the opportunity.

I would also like to express my thanks to Lin Wang at the Research Complex at Harwell for his patience with my Health and Safety inductions, being extremely friendly, chasing up our requests and helping us to resolve technical difficulties with stochastic optical reconstruction microscopy (STORM). I would also like to thank Laura Zanetti Domingues and Benji Bateman for their practical advice and Michael Hirsch for his data analysis.

3 Contents

1 Declaration 2

2 Acknowledgements 3

3 Contents 4

4 List of Figures 7

5 List of Tables 10

6 List of Abbreviations 11

7 Abstract 15

8 Introduction 16

8.1 Myosins are actin-based molecular motors 16

8.2 Myosin VI demonstrates unique directionality and a plethora of biological functions 18

8.3 Overview of transcriptional regulation 21

8.4 Myosin VI transcriptional regulation extends beyond androgen receptors 23

8.5 Myosin VI and NDP52 – a putative transcription co-activator 26

8.6 Project aims 35

9 Materials and Methods 37

9.1 Materials and reagents 37

9.2 Standard expression protocol for recombinant proteins 38

9.2.1 Transformation of recombinant DNA 38

9.2.2 Overnight/starter culture 39

9.2.3 Expression and resuspension 39

9.3 Immobilised metal-ion affinity chromatography and size-exclusion chromatography 40

9.4 Denaturing SDS-PAGE 41
9.5 Beer-Lambert law 42
9.6 DNA-binding assays 42
9.6.1 Sample preparation 42
9.6.2 Data fitting 43
9.7 Circular dichroism 44
9.7.1 Sample preparation 44
9.7.2 Analysis 45
9.8 Size-exclusion chromatography-multi-angle light scattering 46
9.9 Total internal reflection fluorescence microscopy 49
9.9.1 Etch-cleaning of coverslips 49
9.9.2 Sample preparation 49
9.9.3 Imaging 50
9.9.4 Photobleaching analysis 50
9.10 Cryogenic stochastic optical reconstruction microscopy 51
9.10.1 Sample preparation 51
9.10.2 Analysis 52
9.11 SEC-SAXS 53
9.11.1 Sample preparation 53
9.11.2 Analysis 54
10 Results 55
10.1 Optimising the purification of oestrogen receptor α 55
10.1.1 Standard expression protocol 55
10.1.2 Modified expression protocols 56
10.1.3 Incubation of cell lysate with estradiol 58
10.1.4 Culturing in the presence of estradiol 61
10.1.5 Regular additions of estradiol to the culture medium 65
10.2 Biochemical and biophysical characterisation of nuclear dot protein 52 . 67
10.2.1 Expression and purification of NDP52 structural domains 67
10.2.2 Secondary structure content of NDP52 1-190 and 120-end 71

10.2.3 NDP52 structural domain DNA-binding affinities 72
10.2.4 Investigating NDP52 oligomeric states using SEC-MALS 74
10.2.5 eGFP-labelled NDP52 expression and purification 76
10.2.6 TIRF microscopy 82
10.2.7 Preliminary results for a novel GFP-based reporter assay using cryogenic super-resolution microscopy 85
10.2.8 SAXS analysis on NDP52 89

11 Discussion 94

11.1 Culturing in the presence of estradiol significantly improves purification results for ERα 94

11.2 Full-length ERα purification: limitations and future direction 96

11.3 Residues 120-end of NDP52 likely confer binding to dsDNA *in vivo* 99

11.4 NDP52 likely functions as a homodimer *in vivo* 101

11.5 NDP52 is largely linear with an N-terminal globular domain 103

11.6 The coiled-coil domain of NDP52 likely confers dimerisation and provides stability 103

11.7 The presence of dsDNA likely promotes the formation of higher oligomeric states of NDP52 104

11.8 GFP-NDP52-GFP and GFP-NDP52 show the same clustering behaviour. 104

11.9 SAXS suggests NDP52 adopts a parallel homodimeric state *in vitro* 107

12 Conclusions 112

13 References 113

14 Supplementary Data 121

4 List of Figures

Figure 1: Myosin superfamily structure 17

Figure 2: Myosin VI structure 18

Figure 3: Myosin VI has a range of biological functions thanks to multiple binding partners 20

Figure 4: Oestrogen receptor mechanisms of signalling and structure 25

Figure 5: NDP52 structure 28

Figure 6: NDP52 is well-characterised in macro-autophagy 31

Figure 7: NDP52 activates myosin VI to enhance RNAPII-mediated transcription 35

Figure 8: Principle of Ni Sepharose® purification 41

Figure 9: PageRuler™ Plus Pre-Stained Protein Ladder molecular weight markers 42

Figure 10: A typical circular dichroism reference spectrum for protein secondary structure 44

Figure 11: Setup and analysis using SEC-MALS 48

Figure 12: Oxygen scavenger system used in TIRF buffer 49

Figure 13: Setup of *superSIL* microscope 51

Figure 14: Standard expression of ERα 55

Figure 15: Modified expression conditions for ERα 58

Figure 16: Expression and purification results when incubating cell lysate with estradiol 61

Figure 17: Expression results when culturing in the presence of estradiol 62

Figure 18: Affinity chromatography and gel filtration results following purging of the ÄKTA™ pump 64

Figure 19: Expression and affinity chromatography results following regular additions of 500nM estradiol to the culture medium 66
Figure 20: Schematic of NDP52 domains – NDP52 1-190 (A) and NDP52 120-end (B) – not drawn to scale 67
Figure 21: Expression of NDP52 120-end and 1-190 following the standard expression protocol modified to culture at 27°C overnight..... 68
Figure 22: Modified expression protocol results for NDP52 120-end 70
Figure 23: Circular dichroism analysis for NDP52 1-190..... 71
Figure 24: Circular dichroism analysis for NDP52 120-end..... 72
Figure 25: Binding assay results for NDP52 1-190 and 120-end..... 73
Figure 26: SEC-MALS elution profiles for NDP52 1-190 and 120-end..... 75
Figure 27: Expression and affinity chromatography results for GFP-NDP52 and GFP-NDP52-GFP culturing at 27°C post-induction 78
Figure 28: Affinity chromatography and gel filtration results for GFP-NDP52 and GFP-NDP52-GFP following the standard expression protocol 80
Figure 29: Affinity chromatography (gradient elution) and gel filtration (0.2mL/min flowrate) purification results for GFP-NDP52 and GFP-NDP52-GFP..... 81
Figure 30: TIRF microscopy analysis used to assess oligomeric state of NDP52..... 83
Figure 31: Results of photobleaching analysis with TIRF microscopy..... 85
Figure 32: Schematic of N-terminally eGFP-tagged NDP52 in parallel and anti-parallel dimeric states 86
Figure 33: Raw cryo-STORM data for GFP-NDP52 (test sample) 88
Figure 34: Kinhom(H') plot for cryogenic super-resolution microscopy conditions . 89

Figure 35: Gel filtration of full-length NDP52, expressing following the standard expression protocol 90

Figure 36: Guinier fitting for full-length NDP52 used to estimate I(0) and R_g values 91

Figure 37: Intensity plots for peak 1 and 2 of NDP52 sample 92

Figure 38: SAXS envelope for full-length NDP52 93

Figure 39: Summary of optimal expression and purification conditions used for ERα 95

Figure 40: Interpretation of SAXS data 111

Figure 41: Positive and negative controls for ds40 NDP52 binding assays 122

Figure 42: Positive and negative controls for ds15 NDP52 binding assays 123

Figure 43: Analysis of SEC-MALS protein samples at different angles 124

5 List of Tables

Table 1: Source list for materials and reagents 37

Table 2: List of expression vectors 39

Table 3: List of oligonucleotide sequences employed for DNA-binding assays 43

Table 4: List of oligonucleotide sequences employed for TIRF analysis 50

Table 5: Problems encountered during the optimisation of full-length ERα in BL21 *E.coli* 97

Table 6: Experimental issues with cryogenic super-resolution microscopy 106

Table 7: Extinction coefficients and molecular weights of recombinant proteins.. 121

6 List of Abbreviations

ADU	Analogue-to-digital-units
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid
AF	Activation function
AIB1	Amplified in breast cancer 1
Amp	Ampicillin
APS	Ammonium persulfate
AP-1	Activator protein 1
AR	Androgen receptor
Atg8	Autophagy-related protein 8
BSA	Bovine serum albumin
CBD	Cargo-binding domain
CD	Circular dichroism
cDNA	Complementary DNA
ChIP	Chromatin immunoprecipitation
CHO	Chinese hamster ovary
CLIR	Noncanonical LC3-interacting region
COPD	Chronic obstructive pulmonary disease
CSR	Complete spatial randomness
CTD	C-terminal domain
DBD	DNA-binding domain
DLS	Dynamic light scattering
dsDNA	Double-stranded DNA
DTT	Dithiothreitol
E1	Estrone
E2	Estradiol
E3	Estriol
EGF	Epidermal growth factor
eGFP	Enhanced GFP
EM	Electron microscopy

EMCCD	Electron-multiplying charge-coupled device
ERE	Oestrogen response element
ERα	Oestrogen receptor α
ERβ	Oestrogen receptor β
FERM	4.1, ezrin, radixin and moesin
FF	Fast flow
FRET	Förster resonance energy transfer
FT	Fourier transform
GFP	Green fluorescent protein
GPCR	G-protein-coupled receptor
GTF	General transcription factor
H12	Helix 12
hnRNPU	Heterogeneous nuclear ribonucleoprotein U
HPLC	High performance liquid chromatography
IDT	Integrated DNA Technologies
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IQ	Isoleucine-glutamate
K_A	Equilibrium association constant
K_D	Equilibrium dissociation constant
LB	Luria Bertani
LBD	Ligand-binding domain
LC3	Microtubule-associated protein light chain 3
LI	Large insert
LIM	Lin-1, isl-1 and mec-3
LIM-L	LIM-like
LIR	LC3-interacting region
LZ	Leucine zipper
mAb	Monoclonal antibody
MISS	Membrane-initiated steroid signalling
MVI	Myosin VI
MW	Molecular weight

NA	Numerical aperture
ND	Nuclear dot
ND10	Nuclear domain 10
NDP52	Nuclear dot protein 52
NF-κβ	Nuclear factor-κβ
NI	Non-insert
NLS	Nuclear localisation signal
NMR	Nuclear magnetic resonance
NM1	Nuclear myosin 1
OPTN	Optineurin
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PDB	Protein data bank
P_i	Inorganic phosphate
PIPP	Proline-rich inositol-polyphosphate 5-phosphatase
PML	Promyelocytic leukaemia
PMSF	Phenylmethylsulfonyl fluoride
RBP	RNA-binding protein
Rf	Relative mobility
RFP	Red fluorescent protein
R_g	Radius of gyration
RNAi	RNA interference
RNAP	RNA polymerase
RT	Room temperature
SAP97	Synapse-associated protein 97
SAXS	Small-angle X-ray scattering
sCMOs	Scientific complementary metal-oxide semiconductor
SDS	Sodium dodecyl sulphate
SEC-MALS	Size-exclusion chromatography-multi-angle light scattering
SFDA	Single fluorophore detection algorithm
SH3	Src homology domain 3

SI	Small insert
SIL	Solid immersion lens
SKICH	SKIP carboxyl homology
SKIP	Skeletal muscle and kidney enriched inositol phosphatase
SM	Second messenger
SMRT	Silencing mediator for retinoid and thyroid hormone receptor
SNARE	Soluble N-ethylmaleimide sensitive factor attachment protein
SP-1	Stimulating protein 1
ssDNA	Single-stranded DNA
STORM	Stochastic optical reconstruction microscopy
TCI	Tokyo Chemical Industry
TEMED	Tetramethylethylenediamine
TF	Transcription factor
TIRF	Total internal reflection fluorescence
T_m	Midpoint temperature
TRAF6	Tumour necrosis factor receptor-associated factor 6
TSS	Transcription start site
T6BP	TRAF6-binding protein
ZF	Zinc finger

7 Abstract

Myosin VI (MVI) is implicated in many biological processes including, cell migration, endocytosis and transcription. MVI achieves its plethora of cellular functions owing to its association with multiple binding partners. Here, two proteins associated with MVI – nuclear dot protein 52 (NDP52), a putative transcription regulator, and the established transcription factor, oestrogen receptor (ERα) – are biochemically and biophysically characterised to further elucidate the role of this motor protein in transcription. Optimisation of expression and purification conditions has enabled the successful purification of full-length ERα using BL21 *E.coli*. According to current literature, this is the first time this has been reported. This will set the foundation for future purification optimisation required for binding assays between MVI and ERα.

NDP52 has only recently been characterised as a putative transcription co-activator, where it was shown to promote RNA polymerase II (RNAPII) transcription by relieving the auto-inhibition of MVI. Here, novel insights into the biochemical and biophysical properties of NDP52 are presented, including DNA-binding characteristics of different structural domains using fluorescence-based binding assays. Expectedly, C-terminal DNA-binding motifs likely confer binding to double-stranded DNA (dsDNA) *in vivo*. Small-angle X-ray scattering (SAXS) data for full-length NDP52 is also presented, providing important structural information, given that no full-length crystal structure for NDP52 is currently available. Methods such as total internal reflection fluorescence (TIRF) microscopy, size-exclusion chromatography-multi-angle light scattering (SEC-MALS) and cryogenic super-resolution microscopy have also provided an understanding of the oligomeric state of NDP52 and its tertiary structure. Together, this evidence suggests that NDP52 functions as a parallel homodimer *in vivo*, which provides important structure-function insights into NDP52's role in transcriptional regulation. For instance, the parallel arrangement of the NDP52 dimer may facilitate tight binding to dsDNA, where each monomer binds either side of DNA, effectively clamping it.

8 Introduction

8.1 Myosins are actin-based molecular motors

Kinesin, dynein and myosin constitute the 3 classes of cytoskeletal molecular motors, converting chemical energy from the hydrolysis of ATP into mechanical energy.[1] The myosin superfamily is diverse and categorised into 35 distinct subclasses[2], including conventional and non-conventional myosins. Together, these perform a variety of functions such as muscle contraction[3], cell division[4] and endocytosis[5] to name but a few.

Structurally, myosins are classified according to their C-terminal tail domain, which, unlike the N-terminal motor domain, is highly divergent (**Figure 1**) and permits different biological functions by binding to different cargo[6] and associated regulatory proteins[7]. The N-terminal ATPase domain of myosin heavy chains, also known as the 'motor domain' or 'head', is highly conserved in evolution owing to its ability to bind both actin and ATP (**Figure 1A**).[8] Myosin heavy chains are also characterised by a neck region (also known as the 'lever arm'), which consists of varying numbers of isoleucine-glutamate (IQ) motifs (from 1-6) depending on the class (**Figure 1A**).[6] IQ motifs conventionally facilitate binding to calmodulin/myosin light chains.[6,9] Myosin VI (MVI) has 2 calmodulin binding sites – an IQ motif and insert-2.[10]

The lever arm of myosins is so called since it is believed to transduce force and movement from the N-terminal head, during the ATPase cycle, to the C-terminal tail; hence, acting as a lever or converter.[11] Binding of calmodulin to IQ motifs in the lever arm provides rigidity and structural support to this otherwise long and flexible protein domain, which enables walking along actin filaments.[12] Many myosins conform to the lever arm hypothesis, which predicts that myosin step size is directly proportional to the length of the lever arm.[11] Although, some challenge it. For instance, experiments using optical tweezers have demonstrated that MVI takes a larger step size (on average between 30-36nm) than expected, considering its relatively short lever arm with only 2 IQ motifs.[13]

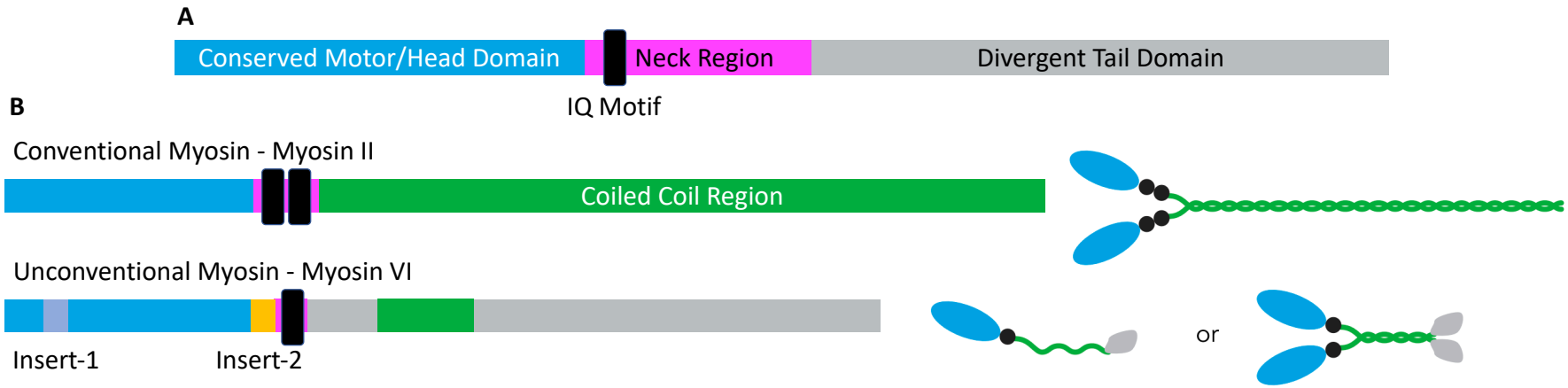


Figure 1: Myosin superfamily structure

(A) Schematic of myosin structural domains. All myosins have 3 key structural elements: a N-terminal motor domain, a neck region and a C-terminal tail domain. Depending on the myosin class, there will be varying numbers of IQ motifs in the neck region.[6] **(B)** Comparison of a conventional and unconventional myosin structure. Myosin II is well-characterised in muscle contraction, where it provides the power to facilitate the sliding filament mode of muscle contraction.[14,15] Myosin II is the only conventional class, while all other myosins are termed 'non-conventional'.[6] Coiled-coil motifs in the tails of some myosin classes are believed to facilitate dimerisation.[6] Myosin VI, for example, can exist in both a monomeric or dimeric form, which is mediated by binding partners[16,17], whilst myosin II is only ever monomeric[6]. Myosins can contain other structural domains, such as 4.1, ezrin, radixin and moesin (FERM) domains, which enable binding to transmembrane proteins and are found in myosins such as myosin VIIa and X.[6] Myosin VI has unique amino acid sequences, labelled insert-1 and -2, which are discussed in text. Other myosin classes are not shown. Figure re-drawn from Krendel & Mooseker, 2005.

8.2 Myosin VI demonstrates unique directionality and a plethora of biological functions

Actin filaments are orientated so that their barbed (positive) end points towards the plasma membrane, while the pointed (minus) end faces away from the plasma membrane towards the nucleus.[18] MVI, unlike other myosins, displays unique directionality, walking towards the minus end of actin filaments.[19,20] Repositioning/reversal of the lever arm by 120° is achieved in MVI by a unique 53-amino acid insert (insert-2) between the motor domain and IQ motif in the neck of the protein (**Figure 2A**).[21] It is this repositioning of the lever arm which is responsible for the unique directionality of MVI.[21]

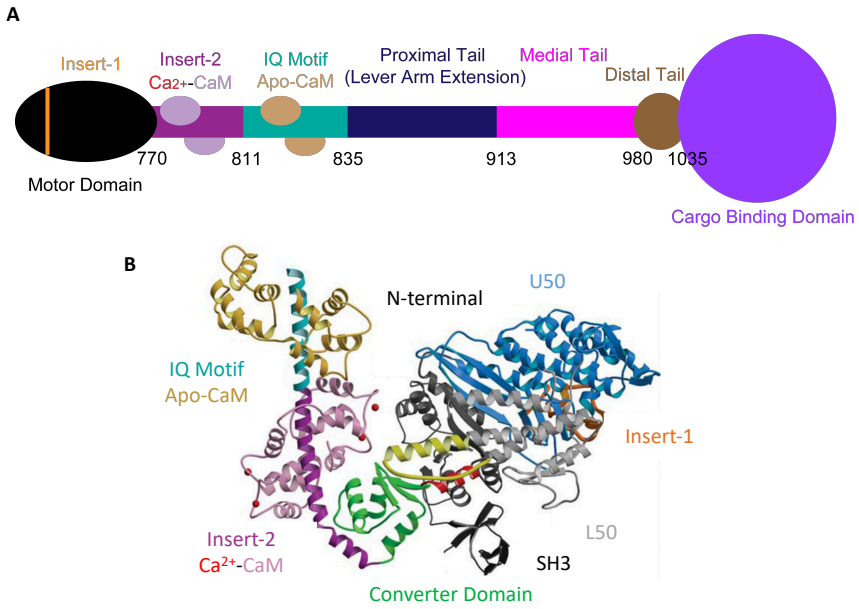


Figure 2: Myosin VI structure

(A) Schematic of myosin VI structural domains. Key structural domains of myosin VI include: the N-terminal motor domain, insert-2 (discussed in text) which can bind Ca^{2+}-calmodulin[21], another calmodulin-binding IQ motif and the tail domain. The tail domain can be split into the proximal tail/lever arm extension, medial tail, distal tail and C-terminal cargo-binding domain.[16] The lever arm extension contains a 3-helix bundle thought to help facilitate the large step-size of myosin VI dimers upon unfolding (although remains a topic of controversy)[22], while the medial tail contains a short coiled-coil and then single α-helix[16]. Figure largely inspired by Phichith *et al*, 2009. **(B)** Crystal structure (ribbon diagram) of nucleotide-free myosin VI. The N-terminal motor domain of myosin VI comprises the N-terminal, U50, L50 and converter domains.[21] The Src homology domain 3 (SH3) of myosin VI is also shown in black. Figure taken directly from Ménétrey *et al*, 2005.

The unique directionality of MVI suggests that it may have important biological functions, which has been the driving force behind research in the Toseland laboratory, as well as many others. This is in fact true, as MVI is implicated in transcription (along with nuclear myosin I and myosin Vb).[23] The range of functions of MVI is commonly attributed to its splice isoforms and range of binding partners, which will be discussed below. To give an overview of the range of biological processes MVI is implicated in, MVI has functional roles in both clathrin-coated and non-clathrin coated vesicle endocytosis, hearing through maintenance of stereocilia, autophagy, cell migration, cytokinesis, synaptic transmission, maintenance of Golgi morphology and the secretory pathway (**Figure 3**).[24]

MVI is alternatively spliced resulting in the so-called 'large insert' (LI) and 'small insert' (SI) isoforms, where there are 31-residue and 8-residue insertions before and in the cargo-binding domain (CBD) respectively.[7,25] In combination, these inserts result in the SI+LI isoform. These isoforms have distinct distributions, where the non-insert (NI) isoform can enter the nucleus whilst the LI isoform cannot.[7] MVI isoforms are also believed to be selective for different binding partners, where NDP52 has been shown to bind and regulate the backfolding of the NI isoform[7] (**Figure 7A**), whilst the LI isoform can only bind partners at the WWY motif (excludes NDP52)[25]. Hence, splice isoforms of MVI contribute to its range of biological functions.

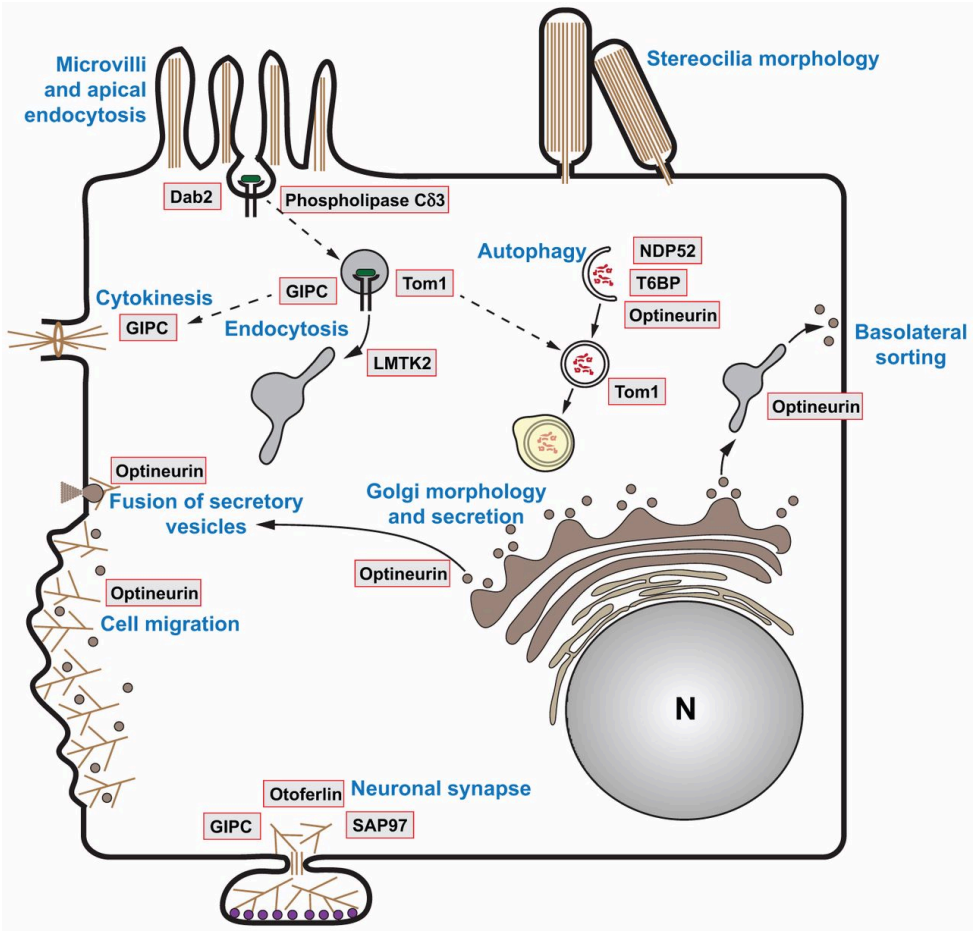


Figure 3: Myosin VI has a range of biological functions thanks to multiple binding partners

Myosin VI has many biological functions owing to its range of binding partners, as shown, including maintaining stereocilia morphology and autophagy (discussed in text). Myosin VI also functions in synaptic transmission, such as glutamergic transmission[26], through its interaction with α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid (AMPA)–type glutamate receptors via synapse-associated protein 97 (SAP97)[27]. Figure taken directly from Tumbarello *et al*, 2013.

As discussed, MVI, along with nuclear myosin I (NM1) and myosin Vb, is implicated in transcription.[23] Using chromatin immunoprecipitation (ChIP), MVI has been shown to be recruited to the promoters of actively transcribed genes.[28] MVI has also been shown to complex with RNA polymerase (RNAP) II upon transcription activation through immunoprecipitation assays and immunofluorescence.[28] Expression levels of MVI also correlate with RNAPII-mediated transcription.[28] Recently, MVI has been further characterised in transcription, where it was shown to co-localise with transcriptionally active regions upon PC12 cell stimulation.[29] Many MVI binding partners are also thought to be implicated in processes associated with gene

expression, such as heterogeneous nuclear ribonucleoprotein U (hnRNPU) implicated in pre-mRNA metabolism and transport[29], as well as nuclear dot protein 52 (NDP52) and oestrogen receptor α (ERα) – which are transcriptional regulators[7].

ERα has only recently been characterised as a MVI binding partner[7] but has long been recognised as a nuclear receptor transcription factor[30]. NDP52, however, has only recently been characterised as a putative transcription factor[7] but is well-characterised as a MVI binding partner[31]. These MVI binding partners are of great interest with regard to MVI's role in transcriptional regulation, which will be discussed in the following chapters.

8.3 Overview of transcriptional regulation

There are 3 RNA polymerases (RNAPs) in eukaryotes, where RNAPI, II and II transcribe class I, II and III genes respectively. RNAPII is also responsible for the transcription of mRNA, in addition to class II genes, and hence is a key regulator of eukaryotic gene expression. RNAPII is a formed of 12 distinct subunits, where RNA-binding protein (RBP) 5, 6, 8, 10 and 12 are common to the 3 classes of eukaryotic RNA polymerases.[32] These 12 subunits must come together in a coordinated fashion to enable transcription of class II genes.[33] General transcription factors (GTFs) that mediate this assembly are TFIID, TFIIB, TFIIF, TFIIE and TFIIH, which assemble in this order.[33] GTFs have roles in promoter recognition, RNAPII recruitment and transcription start site (TSS) recognition, where TFIIH specifically has helicase and kinase roles.[33]

Transcription is divided into 3 stages: initiation, elongation and termination. The C-terminal domain (CTD) of RBP1 (largest subunit of RNAPII) acts as a binding scaffold for nuclear factors, regulated by phosphorylation events at Ser2 and 5.[33] Phosphorylation of Ser2 and 5, of the tandemly repeated heptad sequence (YSPTSPS), is particularly important during the transition between initiation and elongation, as well as during transcription termination.[33] This CTD is therefore an important regulatory domain in RNAPII.

The term 'transcription factor' refers to any protein that can bind DNA in a sequence-specific manner and regulate transcription.[34] The diversity of transcription factors is vast, which is expected considering that they account for approximately 8% of all human genes.[34] Transcription factors achieve transcriptional regulation through chromatin remodelling, via histone acetylation/deacetylation, interactions with co-factors and directly by recruiting RNAPII to the promoters of target genes.[34]

Transcription factors and transcription factor binding proteins are implicated in many human diseases, including cancer and asthma. For instance, pro-inflammatory gene expression, regulated by transcription co-activators such as nuclear factor-κβ (NF-κβ) and activator protein-1 (AP-1), is enhanced in chronic inflammation.[35] Actions of these transcription co-activators results in the acetylation of core histones, hence promoting transcription of pro-inflammatory genes.[35] Corticosteroids are an anti-inflammatory therapy used to effectively treat asthma but are relatively ineffective for chronic obstructive pulmonary disease (COPD), where they act by reversing this histone acetylation.[35] Similarly, mutations in transcription factors are often implicated in cancer, such as those in N-MYC[36] and ERα[30].

Transcription factors are usually divided into 2 classes – co-activators and repressors – which respectively promote or repress transcription. Transcription factors are usually modular in structure, containing a DNA-binding domain, as well as either an activator or repressor domain connected by a flexible linker.[37] These domains are believed to act independently of each other, as demonstrated initially with GAL4.[38] Repressor domains usually have high proportions of hydrophobic amino acid residues, while activator domains exhibit large structural diversity.[37]

Well-characterised DNA binding motifs in transcription factors include zinc fingers (ZFs) and basic leucine zippers (LZs).[37] Zinc fingers, as the name suggests, coordinate 1 or more Zn^{2+} ions via cysteine and histidine residues, where the structure formed enables insertion into the major groove of DNA.[37] There are different classes of ZFs, including the C_2H_2, C_4 and C_6 classes, where the C_2H_2 class is one of the most common found in eukaryotic transcription factors.[37] Leucine zippers, however, are characterised by leucine at every 7^{th} position; since most transcription factors with

these domains are dimeric, this motif forms a zipper-like structure.[37] Other DNA-binding motifs include homeodomains and helix-loop-helix motifs.[37]

8.4 Myosin VI transcriptional regulation extends beyond androgen receptors

ERα and ERβ both belong to the class of steroid hormone nuclear receptor transcription factors, along with androgen receptors (ARs), where signalling between ERα and ERβ is said to be in balance[30,39]. Transcription factors within this class translocate to the nucleus, where they bind directly to hormone response elements in the promoters of target genes.[30] ERα is no exception to this, where, in the classical signalling pathway, ERα homodimers bind directly to oestrogen response elements (ERE) in the promoters of target genes.[30] ERα homodimer binding is believed to be facilitated by the palindromic nature of EREs, which have the consensus sequence GGTCAnnnTGACC[40,41] ('nnn' denotes a 3-nucleotide spacer); although, non-consensus EREs are still highly responsive to oestrogen[41]. Circulating oestrogens include estrone (E1), estradiol (E2) and estriol (E3), where E2 is the primary circulating oestrogen.[30] ERα acts as a RNAPII-mediated transcription co-activator as it modifies chromatin through recruitment of co-regulatory proteins, stabilises the pre-initiation complex and promotes the recruitment of RNAPII.[30]

Structurally, ER classes have conserved DNA-binding and ligand-binding domains (LBDs) but divergent N-terminal domains (**Figure 4B**).[30] The N-terminal domain and the LBD, which are known as activation function(AF)-1 and -2 respectively (**Figure 4B**), recruit co-regulatory proteins to DNA-bound ER to promote transcription.[30]

ER signalling is implicated in growth, migration, differentiation and apoptosis of cells, as well as angiogenesis, and is hence often implicated in cancer, such as that of breast tissue.[30] In addition to the classical pathway, ER signalling can also occur through 3 distinct pathways, which are classified as being either genomic or non-genomic, as well as either ligand-dependent or ligand-independent.[30] Although, the classical signalling pathway (genomic and ligand-dependent) remains the best characterised.[30]

The first of these pathways is termed the 'cross-talk' pathway (genomic and ligand-dependent). In the absence of EREs, ER homodimers can cross-talk with other transcription factors such as Fos/Jun[42] or stimulating protein-1 (SP-1)[43]. In the genomic but ligand-independent pathway, growth factor signalling is believed to activate kinases which in turn phosphorylate and activate the ER instead of oestrogens, which would ordinarily bind and activate ERs in the classical signalling pathway.[44] This ligand-independent mechanism is believed to be implicated in hormone-independent growth of some cancers.[30] Finally, in the non-genomic but ligand-dependent pathway, which is considered the least understood[30], G-protein-coupled receptors (GPCRs), like GPCR-30[45], are activated through oestrogen-bound ERs. This results in a signalling cascade via secondary messengers (SMs) and a rapid physiological response in the cell.[30] It is this response, such as an increased intracellular NO concentration, that is thought to enable non-genomic transcriptional regulation.[30] This is also called the membrane-initiated steroid signalling (MISS) pathway and is believed to account for the rapid action of oestrogens observed in some cell types.[46] A summary of these signalling mechanisms is shown in **Figure 4A**.

In the classical pathway, ligand binding induces a conformational change in ERs which enables translocation to the nucleus. This conformational change occurs in the AF-2 domain of ERs, specifically helix 12 (H12).[30] ER ligands can either be agonistic or antagonistic, depending on what conformational change is brought about in H12.[30] For instance, binding of agonistic ligands causes H12 to adopt an open conformation with a shallow hydrophobic binding site for the leucine-rich LxxLL motif of co-activators.[30] Whereas, when antagonists bind, some believe H12 mimics the LxxLL motif, hence blocking the co-activator docking site.[30] Others believe binding of co-repressors, such as silencing mediator for retinoid and thyroid hormone receptor (SMRT), is prevented by an extended co-repressor box.[30] The recruitment of co-regulatory proteins to AF-2 is important for correct ER signalling. For instance, over-expression of 'amplified in breast cancer 1' (AIB1) – an ER transcriptional co-activator – is observed in breast cancer.[30]

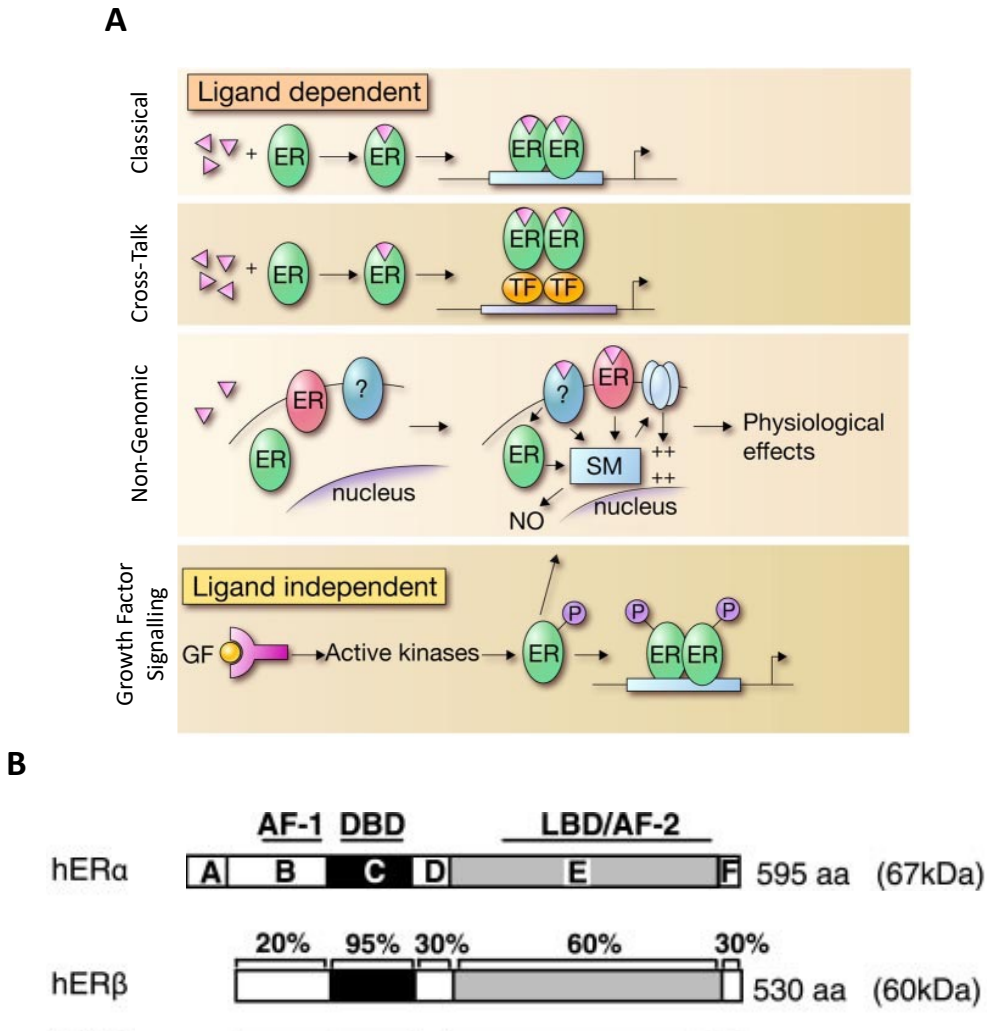


Figure 4: Oestrogen receptor mechanisms of signalling and structure

(A) Oestrogen receptor (ER) signalling is classified into ligand-dependent and independent pathways. In the classical pathway, ligands bind to the ER causing homodimerisation and subsequent translocation to the nucleus, where ERs bind to oestrogen response elements in the promoters of target genes.[30] ERs can also signal through 'cross-talk' with other transcription factors such as Fos/Jun.[30] In the non-genomic pathway, either a classical ER, an ER isoform or a distinct receptor is activated upon ligand binding causing a rapid physiological response via second messengers inside the cell, such as increased NO levels or an influx of ions.[30] In the ligand-independent pathway, signalling is achieved through growth factor signalling, which in turn activates kinases which phosphorylate/activate ERs.[44] **(B)** Schematic of the structural organisation of human (h) ERα and ERβ, where the structure of ERα has been sub-divided into domains A to F. ERα and ERβ have a highly conserved DNA-binding domain (DBD), with 2 ZFs, a less-well conserved ligand-binding binding domain or LBD (also known as activation function (AF-2) and a highly divergent N-terminal AF-1 domain.[30] The percentage sequence homology between the receptors is shown. Figures taken from Heldring *et al*, 2007.

Fili *et al* (2017) demonstrated, through pull-down assays, that ERα binds MVI at the nuclear receptor LxxLL binding motif in the CBD, where site-directed mutagenesis of this motif prevents binding.[7] MVI was shown to have a functional role in the regulation of ERα-driven transcription, since knockdown of the motor protein reduced expression of ERα target genes – *PS-2* and *GREB-1* – by 85% and 40% respectively.[7] Using a luciferase reporter assay, Fili *et al* (2017) also demonstrated that MVI has the potential to regulate expression from ERE promoters, supporting a role for MVI in ER-driven gene expression.[7]

Nuclear MVI has been shown previously to have a functional role in androgen-dependent gene expression, where MVI and ARs have been shown to bind through co-immunoprecipitation assays in LNCaP cells; MVI is also thought to recruit ARs to the promoters of target genes.[47] Therefore, the finding that MVI also binds to ERα demonstrates a functional role for MVI, in relation to steroid hormone nuclear receptor transcription factors, beyond just ARs.[7] This is particularly significant given the over-expression of MVI in prostate, breast[48] and ovarian[49] cancer, where signalling through hormone receptors is implicated in driving cancer progression.

8.5 Myosin VI and NDP52 – a putative transcription co-activator

The name NDP52 comes from initial immunohistochemical studies that showed dot-like nuclear staining and co-localisation with nuclear domain 10 (ND10) proteins, such as promyelocytic leukaemia (PML) and Sp100.[50] The term '52' comes from a molecular weight (MW) of 52kDa (446 amino acids), which was predicted through isolation and sequencing of cDNA.[50] NDP52, along with tumour necrosis factor receptor-associated factor 6 (TRAF6)-binding protein (T6BP), also known as TAX1BP1, were identified as binding partners of MVI using two-hybrid screens in yeast and Chinese hamster ovary (CHO) cells, as well as co-immunoprecipitation assays.[31] Site-directed mutagenesis revealed that both NDP52 and T6BP bind to MVI at the RRL motif in the C-terminal CBD, where the ZF motifs of both NDP52 and T6BP are required for binding.[31]

Initial Kyte and Doolittle[51] plots predicted a central hydrophilic region of approximately 200 amino acids in NDP52, between amino acids 134-350, with high probability of forming an α-helical coiled-coil (supported by the absence of prolines).[50] Within the coiled coil, between amino acids 324-345, a LZ is predicted.[50] NDP52 is also predicted to have small clusters of hydrophobic amino acids at the N- and C-termini, with high probability of forming β-sheets and -turns.[50] This is consistent with the presence of a skeletal muscle and kidney enriched inositol phosphatase (SKIP) carboxyl homology (SKICH) domain and a lin-1, isl-1 and mec-3 (LIM)-like (LIM-L) domain at the N- and C-termini respectively (**Figure 5A**).[50] NDP52 is hence predicted to have globular structures at its ends (**Figure 5B**).[50]

Other proteins, such as CALCOCO1 (or CoCoA), SKIP, and T6BP, are also characterised by a SKICH domain[52], where NDP52 and T6BP are both autophagy receptors[53]. This 128-amino acid domain is responsible for translocation of SKIP from the endoplasmic reticulum to sub-membranous actin at plasma membrane ruffles upon epidermal growth factor (EGF) stimulation (in COS7 cells).[54] The same phenomenon is observed for proline-rich inositol-polyphosphate 5-phosphatase (PIPP)[54] but unexpectedly was not observed with T6BP and NDP52 upon EGF stimualtion[31]. Although, RNAi-mediated knockdown of both T6BP and NDP52 shows that these proteins are required for actin filament organisation because there is reduced membrane ruffling at the leading edge and an increase in stable actin stress fibres.[31] The SKICH domain of NDP52 is hence believed to facilitate plasma membrane localisation.[31]

Aforementioned, NDP52 is also characterised by the presence of a C-terminal LIM-L domain, as well as a SKICH domain. The LIM domain is a unique double-ZF motif, with the consensus sequence $C\text{-}X_2\text{-}C\text{-}X_{16\text{-}23}\text{-}H/C\text{-}X_2\text{-}C/H\text{-}X_2\text{-}C\text{-}X_2\text{-}C\text{-}X_{15\text{-}30}\text{-}C\text{-}X_{1\text{-}3}\text{-}C/H/D$ responsible for coordinating 2 zinc (Zn^{2+}) ions.[55] The C-terminus of NDP52 demonstrates high sequence homology with the consensus sequence of a LIM domain, sharing 6 cysteine residues.[50] Hence, NDP52 has a C-terminal double-ZF motif. LIM domains have been characterised in kinases, adaptors and importantly, with regard to NDP52, transcription factors.[55]

Despite the structural information presented above, there are currently no crystallography structures available for full-length NDP52. However, the structure of the NDP52 ubiquitin-binding ZF has been determined by nuclear magnetic resonance (NMR; PDB entry – 2MXP) and the SKICH domain of NDP52 similarly determined by X-ray crystallography (PDB entry – 3VVV). These structures have been incorporated into the schematic shown in **Figure 5**.

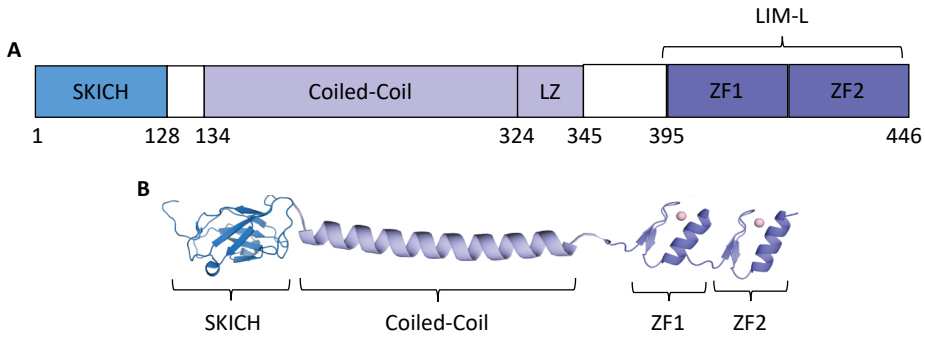


Figure 5: NDP52 structure

(A) Schematic of NDP52 structure. NDP52 is characterised by the presence of a skeletal muscle and kidney enriched inositol phosphatase (SKIP) carboxyl homology (SKICH) domain, a central coiled-coil region containing a leucine zipper (LZ) at the C-terminal and a double zinc finger (ZF) lin1, isl-1 and mec-3 (LIM)-like (LIM-L) domain.[50] Not drawn to scale. Figure adapted from Korioth *et al*, 1995. **(B)** Cartoon representation of expected secondary structure content of NDP52. SKICH domain – PDB entry 3VVV; ZFs – PDB entry 2MXP.

LIM domains are believed to facilitate protein-protein interactions, such as those observed between paxillin and binding partners including α- and γ-tubulin.[55] When NDP52 was initially shown to co-localise with ND10 proteins, the identification of a LIM-L domain (between amino acids 395-446) supported the conclusion that NDP52 could mediate interactions between ND10 proteins such as Sp100 and PML.[50] Subsequently, however, it was found that this observed co-localisation, through use of mAb C8A2 to stain for NDP52, was an artefact of C8A2 cross-reacting with Sp100 and NDP52 does not associate with nuclear dots (NDs).[56] NDP52 instead exhibits mainly cytoplasmic and diffuse nuclear staining, suggesting that NDP52 can be localised to the nucleus but not to NDs, despite its original name.[56] Two-hybrid assays also demonstrated that NDP52 forms a homodimer; NDP52 is hence likely to exist as a non-ND-associated dimeric or multimeric protein.[56]

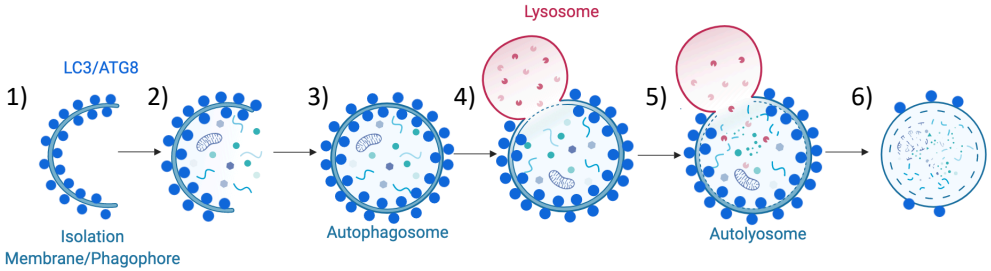
NDP52 interestingly has no nuclear localisation signal (NLS)[50], depsite being characterised as a putative transcription factor[7]. Nuclear localisation can therefore be explained by unregulated diffusion through nuclear pores or binding to a cytoplasmic protein targeted to the nucleus.[50] Interestingly, green fluorescent protein (GFP)-tagged NDP52 has not been observed entering the nucleus[31], suggesting that GFP may impair nuclear import of NDP52.

NDP52 is well-characterised in biological processes including cell adhesion[31] and macro-autophagy[57]. For instance, RNAi-mediated knockdown of NDP52 reduces focal adhesion formation.[31] For the remainder of this section, the focus will be NDP52's role in autophagy.

Autophagy, or the self-degradative process in cells, is responsible for clearing cells of aggregated proteins, damaged organelles and intracellular pathogens.[58] The 3 distinct types of autophagy – macro-, micro- and chaperone-mediated autophagy – all converge at the lysosome, where proteolysis takes place by lysosomal proteases such as cathepsins.[58] Macro-autophagy contrasts micro-autophagy, where cytoplasmic components are endocytosed directly by the lysosome.[58] Macro-autophagy is instead characterised by the formation of the autolysosome by the fusion of the double-membraned autophagosome with the lysosome (**Figure 6A**).[58] By-products of degradation, including amino acids, are recycled during autophagy where they are exported back into the cytoplasm by the action of lysosomal permeases and transporters, where they can be used for metabolism and protein synthesis.[58]

NDP52 is a well-characterised autophagic receptor.[59,60] NDP52 binds to ubiquitinated cargoes, such as pathogens or mitochondria, through its double-ZF motif (LIM-L domain) and selectively recruits them to microtubule-associated protein light chain 3 (LC3)-positive autophagosomal membranes.[59] The way that NDP52 achieves this is through binding to autophagy-related protein 8 (Atg8) or LC3 family proteins, such as LC3C, though its noncanonical LC3-interacting region (CLIR).[60] NDP52 is, therefore, along with optineurin (OPTN) and T6BP, a cargo-specific autophagy receptor.[60] Interestingly, OPTN, NDP52 and T6BP are all binding partners of MVI.[31,60]

Figure 6A


LC3/ATG8
1)
2)
3)
4)
5)
6)
Lysosome
Isolation Membrane/Phagophore
Autophagosome
Autolysosome

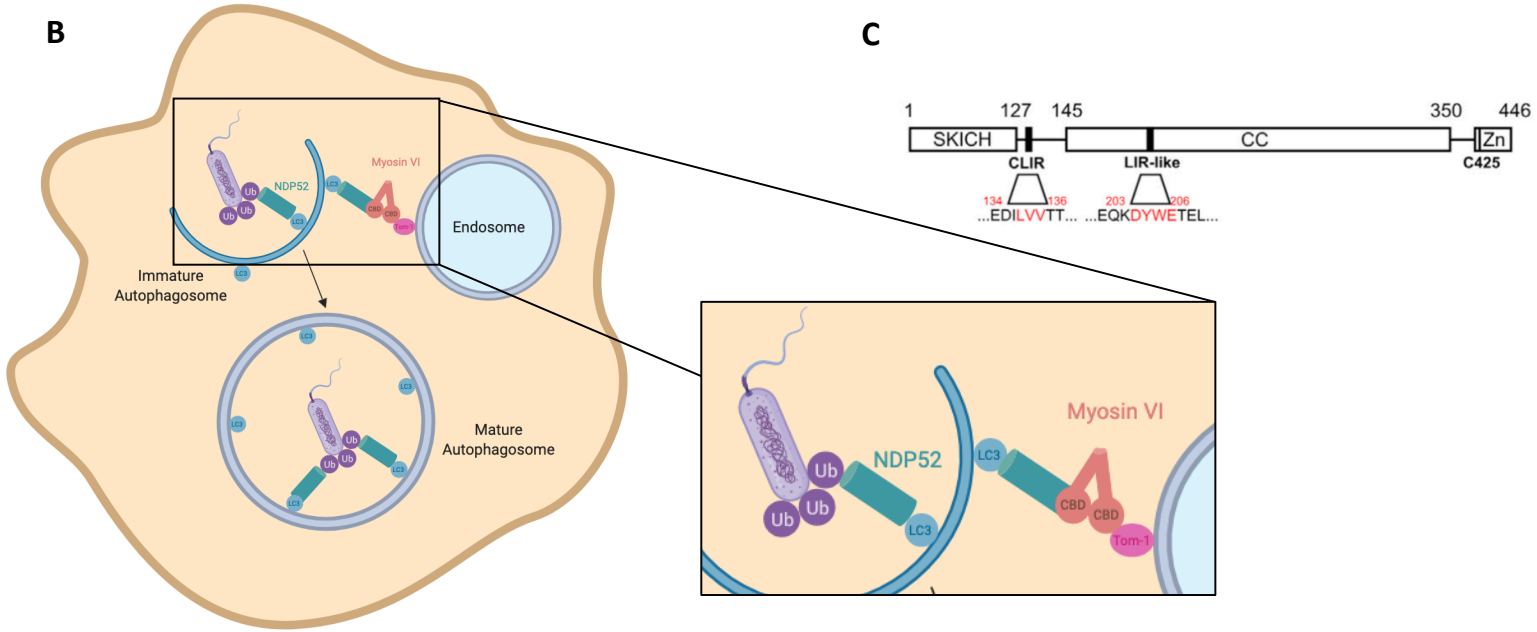

B
C
Myosin VI
NDP52
Endosome
Immature Autophagosome
Mature Autophagosome
Ub
LC3
CBD
Tom-1
1
127
145
350
446
SKICH
CC
Zn
CLIR
LIR-like
C425
134
136
203
206
...EDILVVTT...
...EQKDYWETEL...

Figure 6: NDP52 is well-characterised in macro-autophagy

(A) Schematic of macro-autophagy. 1) The phagophore (also known as the isolation membrane) expands (2) to engulf its target, such as damaged mitochondria, ubiquitinated bacteria or protein aggregates. The subsequent autophagosome (3) fuses with the lysosome (4) resulting in an autolysosome (5), which promotes cargo degradation via lysosomal enzymes (6).[58] LC3 is a characteristic marker of autophagosome membranes.[60] **(B)** NDP52 promotes autophagosome maturation in macro-autophagy. NDP52, a well-characterised autophagy receptor, binds to ubiquitinated cargoes, such as intracellular pathogens during xenophagy, as well as LC3 family proteins on autophagosomal membranes, hence targeting specific cargoes for autophagy.[60] NDP52, bound to LC3-positive autophagosomal membranes, also binds myosin VI, linked to endosomal membranes via Tom-1[57,61], to promote autophagosome maturation. **(C)** NDP52 binds different LC3 family members via different domains. NDP52 has a noncanonical LC3-ineracting region (CLIR), which binds LC3C, as well as a LIR-like domain that binds LC3A, LC3B and or GABARAPL2 to promote autophagosome maturation.[57] **(A)** and **(B)** made with BioRender; **(C)** taken directly from Verlhac *et al*, 2015.

In xenophagy (autophagic removal of pathogens), NDP52 has been shown to be important for the maturation of bacteria-containing autophagosomes.[57] NDP52 promotes autophagosome maturation by linking MVI, which is bound to endosomal membranes via Tom-1, with Atg8 orthologs such as LC3A, LC3B and/or GABARAPL2, which are characteristic of autophagosomal membranes (**Figure 6B**).[57,61] Hence, NDP52, along with other MVI binding partners, such as OPTN and T6BP, are linkers that promote fusion of the endosome and autophagosome.[57,61]

NDP52 selectively binds LC3C via its CLIR and displays only weak or no affinity for other LC3 family proteins at this region; however, NDP52 binds LC3A, LC3B and/or GABARAPL2 (but not LC3C) through a LIR-like motif that lacks hydrophobic residues of the conventional LIR motif.[57] It was shown that this LIR-like motif is required for autophagosome maturation, along with MVI binding, while the CLIR motif is required to selectively bind substrates (**Figure 6C**).[57] It is hypothesised that binding to LC3C uncovers the LIR-like motif, which is otherwise masked by the coiled-coil domain, hence enabling autophagosome maturation through interactions with LC3A, LC3B and/or GABARAPL2.[57]

Tom-1 binds MVI at the WWF motif[53], via a conserved MVI-binding site[62], while OPTN, NDP52 and T6BP bind MVI at the RRL motif in the CBD[53,62]. In the case of

NDP52, binding to MVI is achieved through the double-ZF (LIM-L) motif[62]. Alone, Tom-1 cannot associate with autophagy receptors (OPTN, NDP52 and T6BP) but can in the presence of MVI.[62] MVI therefore brings its 2 cargo-binding domains, RRL and WWF, in close proximity to enable soluble N-ethylmaleimide sensitive factor attachment protein (SNARE)-mediated fusion of the endosomal membrane with the autophagosomal membrane, as discussed.[24] Interestingly, MVI does not dimerise upon binding to Tom-1 which supports an anchoring/tethering role (functioning as a non-processive motor) during autophagosome maturation.[62] This contributes to the growing body of evidence that monomer to dimer transition, regulated by binding partners of MVI, is important for the range of functions of this motor protein.[62]

Interestingly, NDP52 (CALCOCO2) shares 70% sequence homology with a known transcription co-activator – CoCoA (CALCOCO1) – hence suggesting NDP52 may have functional roles in transcription.[7] This has recently been shown to be the case, where NDP52 has been identified as a putative transcription co-activator demonstrating roles beyond autophagy and cell adhesion.[7] MVI has been shown to exist in an auto-inhibited conformation, where there is backfolding between the CBD and tail domain.[7] This was demonstrated *in vitro* using Förster resonance energy transfer (FRET)-based assays, as well as *in vivo* using fluorescence lifetime imaging.[7] In this auto-inhibited state, the DNA-binding sites are occluded.[7] However, upon unfolding, MVI is capable of binding DNA.[7] Calcium, actin and ATP are not responsible for this unfolding event, and instead unfolding is attributed to the MVI binding partner – NDP52.[7]

It is suggested that there are 2 NDP52-dependent routes for MVI dimerisation.[7] In one pathway, NDP52 unfolds MVI first before recruiting a second MVI monomer (I; **Figure 7A**), whereas in the other pathway, 2 MVI monomers are unfolded before dimerisation (II; **Figure 7A**).[7] MVI dimerisation involves either a single NDP52 dimer (stoichiometry of 2:2) or 2 NDP52 dimers (stoichiometry of 2:4).[7] These routes are dependent on the concentration of NDP52, which implies that different pathways may be employed for different functions and at different times.[7]

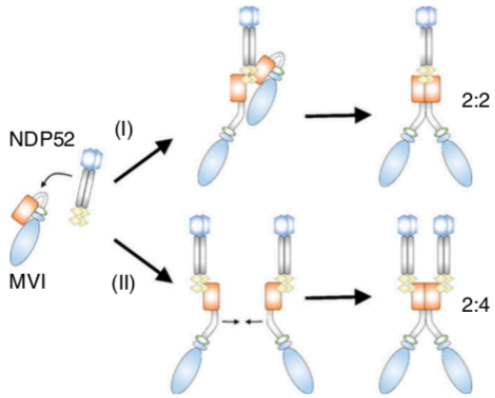
Binding of MVI to RNAPII is dependent on DNA binding; hence, through unfolding MVI, which promotes DNA binding, NDP52 enhances RNAPII-mediated transcription and can be described as a transcription co-activator.[7] The functional role of NDP52 in transcription is further supported by:

- co-immunoprecipitation of RNAPII with full-length NDP52 from HeLa extracts;
- partial co-localisation of RNAPII and NDP52, shown by immunofluorescence (**Figure 7B**);
- strong binding affinity of full-length NDP52 for double-stranded DNA or dsDNA (**Figure 7C**);
- and the observation that transcription decreases by ~50% *in vitro* upon depletion (sequestering) of NDP52, using HeLaScribe extracts (**Figure 7D**).[7]

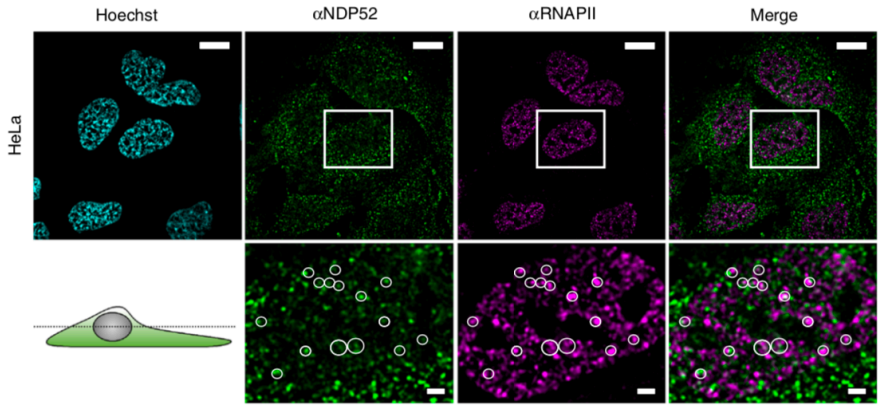
Similar to NDP52, depletion of MVI reduced transcription by ~75% (**Figure 7D**), supporting a functional role for MVI in transcription.[7]

MVI is believed to bind RNAPII through both its CBD and motor domain; it is believed these domains bind DNA and/or binding partners, as well as nuclear actin, respectively (**Figure 7E**).[7] Therefore, MVI is thought to form a bipartite association with RNAPII, where the unfolded state of MVI could enable the motor to act processively during transcription or form a molecular anchor.[7] For instance, it is hypothesised that nuclear MVI may have functional roles in anchoring RNAPII in so-called 'transcription factories', where it is thought RNAPII is localised in discrete foci to enhance transcription rates.[63]

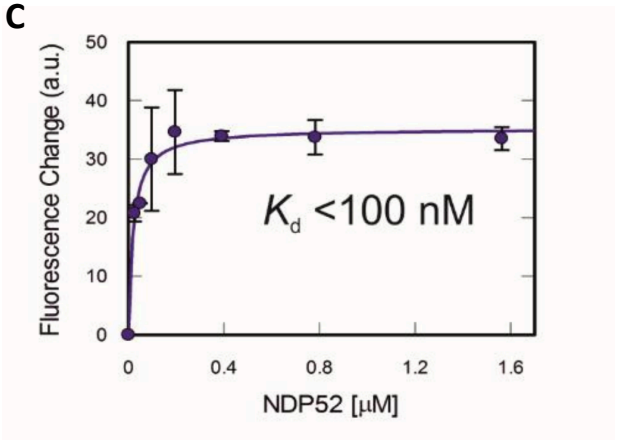
Figure 7A



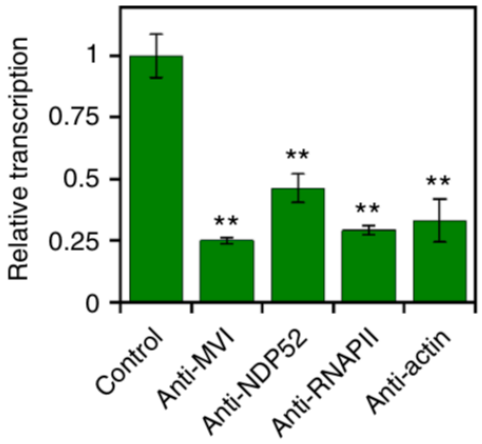
B



C



D



E

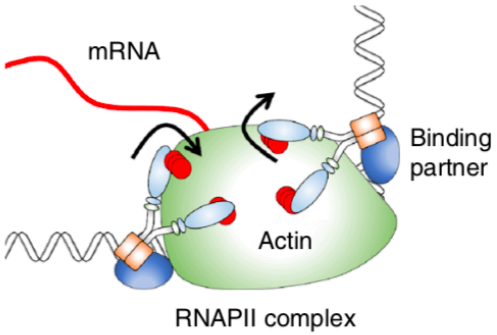


Figure 7: NDP52 activates myosin VI to enhance RNAPII-mediated transcription

(A) Proposed mechanism of NDP52-facilitated dimerisation of myosin VI. Myosin VI could first dimerise before the second myosin VI is unfolded by NDP52. Alternatively, respective myosin VI monomers could unfold before NDP52-facilitated dimerisation. **(B)** NDP52 (green) partially co-localises with RNAPII (magenta) in punctate structures. DNA shown in cyan; scale bar of 10µm for whole images and 1µm for inserts. **(C)** NDP52 binds double-stranded DNA (dsDNA) with high affinity. NDP52 was titrated against 50nM ds40. Data fitted as described in Fili *et al*, 2017 (average of 3 independent experiments). **(D)** Depletion of NDP52 and myosin VI reduces transcription in HeLaScribe transcription assays. Antibodies against myosin VI and NDP52 were used to sequester the proteins respectively. Depletion of myosin VI reduces transcription to 25%, while depletion of NDP52 reduces transcription to 50%. Error bars represent the standard error of the mean, where ** denotes a p-value of <0.001 by a 2-tailed t-test from 5 independent experiments. **(E)** Myosin VI forms a bipartite association with RNAPII. Myosin VI is believed to bind a binding partner and/or DNA at its C-terminal cargo-binding domain, while the N-terminal motor domain is believed to bind nuclear actin. Figures taken from Fili *et al*, 2017.

8.6 Project aims

As discussed, NDP52 and ERα are binding partners of MVI, binding at the RRL and LxxLL motifs in the CBD respectively, hence supporting a role for MVI in transcription.[7] The main aim of this work is to take a multi-disciplinary approach to biochemically and biophysically characterise these transcription regulators associated with MVI, in the hope that this will elucidate further the role of MVI in transcription and ultimately cancer.

As part of this work, purification methods of ERα need to be optimised to achieve a reproducible, high-yield, reliable and relatively inexpensive method. Previous attempts at purifying full-length hERα have proven difficult in a range of expression systems, including bacteria[64,65], yeast[66] and baculovirus[65] cell lines. One of the main aims of this project is hence to successfully optimise the purification of full-length ERα using the BL21 *Escherichia coli* (*E.coli*) strain as the expression vector, owing to its short generation times and relative low cost. This will enable further binding assays and other biochemical analysis to better understand the interaction between ERα and MVI, particularly in relation to transcriptional regulation.

NDP52 is well-characterised in cell adhesion[31] and autophagy[57] but little is known about its role in transcription, where it has only recently been identified as a putative transcription co-activator.[7] Fili *et al* (2017) demonstrated that full-length NDP52 can bind dsDNA with high affinity (**Figure 7C**).[7] Hence, this work aims to employ biochemical methods to characterise the DNA-binding characteristics of different structural domains.

NDP52 is also believed to function as a homodimer *in vivo*[56]. This project, therefore, aims to use size-exclusion chromatography-multi-angle light scattering (SEC-MALS), as well as photobleaching analysis using total internal reflection fluorescence (TIRF) microscopy, to probe the structure and oligomeric state of NDP52. This project also aims to develop and trial a novel GFP-based assay, with cryogenic super-resolution microscopy, to investigate whether NDP52 homodimers adopt a parallel or anti-parallel state, as well as successfully use small-angle X-ray scattering (SAXS) to gain structural insights into full-length NDP52. This is particularly important to help elucidate NDP52's role as a regulator of transcription. For instance, a parallel conformation may suggest NDP52 can bind tightly to DNA, potentially clamping down on it, whereas an anti-parallel configuration may suggest that NDP52 acts as a linker.

9 Materials and Methods

9.1 Materials and reagents

Table 1: Source list for materials and reagents

Material	Source	Composition
A0106	Integrated DNA Technologies (IDT)	100µM
A0107		
A0108		
A0110		
Acrylamide 37.5:1, 30% solution	Alfa Aesar Chemicals	-
Amicon Ultra centrifugal filter	Merck-Millipore	
Ammonium persulfate (APS)	Fisher Scientific™ UK	10%w/v stock
Ampicillin (Amp)	Melford Laboratories Ltd	100mg/mL stock
Bovine serum albumin (BSA)	Sigma-Aldrich UK	-
Catalase		17mg/mL stock
Coomassie Brilliant Blue G-250	BIO-RAD	0.1%w/v Coomassie stain, 40%v/v methanol, 10%v/v acetic acid
Dithiothreitol (DTT)	Melford Laboratories Ltd	1M stock
Glucose	Fisher Scientific™ UK	-
Glucose oxidase	Sigma-Aldrich UK	-
HiLoad™ 16/600 Superdex™ 200	GE Healthcare	-
HisTrap™ Fast Flow (FF) column		-
Imidazole	Acros Organics™	-
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Melford Laboratories Ltd	1M stock
Luria Bertani (LB) broth	Thermo Fisher Scientific	25g diluted in 1L Milli-Q™ H_2O
$MgCl_2$		-
NaCl		-

Material	Source	Composition
Nu-PAGE	Novex®	-
Phenylmethylsulfonyl fluoride (PMSF)	Melford Laboratories Ltd	100mM stock
Sodium dodecyl sulphate (SDS)	Thermo Fisher Scientific	10%w/v stock
Spin-X centrifuge tube filter	Costar®	-
Sucrose	Thermo Fisher Scientific	-
Superdex™ 200 Increase 10/300 GL column	GE healthcare	-
Superose™ 6 Increase 3.2/300 column		-
Tetramethylethylenediamine (TEMED)	Thermo Fisher Scientific	-
TetraSpeck™ beads	Invitrogen – Thermo Fisher Scientific	
Tris (hyrdoxymethyl methylamine)	Thermo Fisher Scientific	-
Vivaspin® 20mL 10K device	Sartorius	-

9.2 Standard expression protocol for recombinant proteins

9.2.1 Transformation of recombinant DNA

Using a 1:25 dilution of DNA vector to competent BL21 DE3 (Invitrogen™) *E.coli* cells, recombinant DNA was transformed using heat shock at 42°C for 45s. Both before and after heat shock, cells were incubated on ice for 30min and 2min respectively. Post-heat shock and recovery on ice, cells were incubated for 1hr at 37°C in LB broth (1:5 dilution) and plated aseptically onto LB agar before overnight incubation (37°C). **Table 2** lists all expression vectors employed.

Table 2: List of expression vectors

Expression Vector (Residue Numbers)	Affinity Tag	Antibiotic Resistance
Human pET151 His-tag ERα (1-end)	N-terminal 6 histidine residues (His-tag)	Amp^R
Human pET151 His-tag NDP52 (1-190)		
Human pET151 His-tag NDP52 (120-end)		
Human pET151 GFP-NDP52 (1-end)		
Human pET151 GFP-NDP52-GFP (1-end)		

Residue numbers are shown in brackets. pET151 refers to the T7 expression system in *E.coli.*

9.2.2 Overnight/starter culture

Transformed single colonies were cultured overnight at 37°C in LB broth supplemented with 100µg/mL Amp (**Table 2**).

9.2.3 Expression and resuspension

Starter culture was added (1:100 dilution) to LB broth and incubated at 37°C, on rotation at 150rpm, until an OD_{600} of between 0.5 and 0.7AU was reached. Expression was induced upon the addition of 1mM IPTG – a molecular mimic of allolactose – which displaces the repressor from the T7 promoter. Following induction, the temperature was reduced to 18°C and cells incubated overnight.

Cells were harvested by centrifugation at 4,000rpm for 20min at 4°C and resuspended in resuspension buffer (50mM Tris-HCl pH 7.5, 40mM imidazole, 200mM NaCl, 1mM DTT, 20% w/v sucrose, 100µM PMSF). Cell pellets were defrosted and lysed on ice by sonication at 30s intervals (at 14 Amplitude Microns) for a total of 5min 30s (5 repeats), where thawing aids lysis. The supernatant (soluble cell lysate) from subsequent centrifugation (18,000rpm for 30min at 4°C) was used for purification.

9.3 Immobilised metal-ion affinity chromatography and size-exclusion chromatography

The supernatant was purified using a 5mL HisTrap™ FF column (**Table 1**), which was pre-packed with pre-charged Ni Sepharose® hence enabling rapid purification of His-tagged recombinant proteins. The Ni^{2+} ions (on Ni Sepharose®) coordinate histidine residues, which are out-competed upon increasing concentrations of imidazole (**Figure 8**). The HisTrap™ column was equilibrated in Buffer A (50mM Tris-HCl pH 7.5, 40mM imidazole, 500mM NaCl, 1mM DTT) before stepwise increases (20, 40, 75 and 100%) in Buffer B (50mM Tris-HCl pH 7.5, 400mM imidazole, 500mM NaCl, 1mM DTT) hence generating an elution profile. A gradient elution was also employed, where stated. Although, stepwise elution should be assumed as the default method. Protein elution was monitored on the ÄKTA™ chromatography system by measuring the intrinsic fluorescence of tryptophan residues at 280nm, where imidazole also absorbs at 280nm. The purity of fractions yielding a large UV peak were assessed by SDS-PAGE and purified further by gel filtration/size-exclusion chromatography if necessary. This protocol was modified, where stated, by equilibrating the column in Buffer A without any imidazole.

Gel filtration (size-exclusion chromatography) separates proteins according to molecular size (Stokes radius) and was employed to obtain a higher degree of purity compared with affinity chromatography alone. Affinity chromatography fractions were passed through a HiLoad™ 16/600 Superdex™ 200 (**Table 1**) column, equilibrated with gel filtration buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 1mM DTT). The Superdex™ column is a matrix of dextran and highly cross-linked agarose, where, depending on the interaction time with the matrix, proteins elute at different time points, with larger molecules eluting first from the column. As with affinity chromatography, protein elution was tracked measuring absorbance at 280nm.

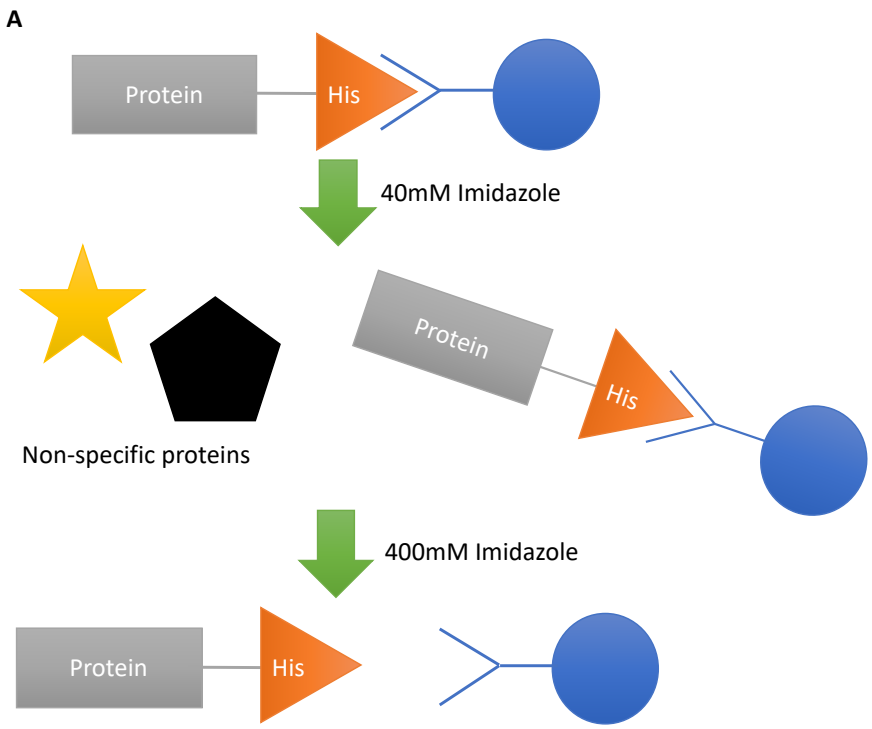


Figure 8: Principle of Ni Sepharose® purification

Nickel ions coordinate His-tagged proteins. Upon treatment with Buffer A (low imidazole), unwanted/non-specifically bound proteins are eluted, while the His-tagged protein remains bound to the column matrix. Upon treatment with Buffer B (high imidazole), imidazole out-competes histidine for the nickel ion meaning that the protein of interest (His-tagged) is eluted.

9.4 Denaturing SDS-PAGE

SDS-PAGE gels were cast by hand, consisting always of a 4% stacking gel and either an 8 or 12% resolving component. Protein samples were denatured prior to loading by heating at 95°C for 10 minutes. Gels were ran in SDS-PAGE running buffer (25mM Tris, 1.4% w/v glycine, 1% SDS), stained with Coomassie Brilliant Blue solution (**Table 1**) and imaged using Image Lab™ software. PageRuler™ Plus Pre-Stained Protein Ladder (**Table 1**) was used throughout (**Figure 9**). MWs of unknown proteins were determined manually (where stated) using calibration curves, where the logMW (in kDa) of known protein standards was plotted against relative mobility (Rf). The intensity of colouration of SDS-PAGE bands was assessed by eye.



Figure 9: PageRuler™ Plus Pre-Stained Protein Ladder molecular weight markers

9.5 Beer-Lambert law

Absorbance was measured at 280nm using the LVis Plate and the CLARIOstar® High Performance Monochromator Multimode Microplate Reader. Protein extinction coefficients (ε) in $M^{-1}cm^{-1}$ were determined using ProtParam (Supplementary **Table 7**). The extinction coefficient, absorbance and pathlength (1cm) were used in the Beer-Lambert Law to determine protein concentration.

9.6 DNA-binding assays

9.6.1 Sample preparation

Changes in fluorescence intensity were used to assess the affinity of protein-DNA interactions. Protein samples (at concentrations stated in text, always in excess of DNA) were serial diluted (1:2 dilution) against 100µL of 100nM DNA, either dsDNA or single-stranded DNA (ssDNA), where ssDNA acted as a negative control.

dsDNA samples were annealed by mixing equimolar and equal volumes of complementary strands (**Table 3**), heating at 95°C for 2min and allowing to cool for 2-3hrs at room temperature (RT), whereas ssDNA samples were not. 3' fluorescein-labelled DNA was employed for binding assays with non-fluorescent protein.

Protein and DNA samples were diluted in binding assay buffer (50mM Tris-HCl pH 7.5, 100mM NaCl, 3mM $MgCl_2$) to achieve desired concentrations and experiments

were performed at RT. Fluorescence intensity readings were taken using the following experimental setup: excitation wavelength – 489nm, excitation bandwidth – 16nm, emission wavelength – 517-656nm and emission bandwidth – 10nm. Data presented is at 517nm and an average of ≥3 independent experiments. Raw data was divided by the dilution factor and normalised before percentage change calculations. For all graphs, error bars are plotted as the standard error of the mean.

Table 3: List of oligonucleotide sequences employed for DNA-binding assays

Referred to in Text	Description	Trade Name	Oligonucleotide Sequence
ds40	Random dsDNA of 40bp	A0107	TTAGTTGTTCGTAGTGCTCGT CTGGCTCTGGATTACCCGC*
		A0106	GCGGGTAATCCAGAGCCAGA CGAGCACTACGAACAACTAA
ss40	Random ssDNA of 40bp	A0107	TTAGTTGTTCGTAGTGCTCGT CTGGCTCTGGATTACCCGC*
ds15	Random dsDNA of 15bp	A0108	TTAGTTGTTCTCTGG*
		A0110	CCAGAGAACAACTAA
ss15	Random ssDNA of 15bp	A0108	TTAGTTGTTCTCTGG*

Oligonucleotide sequences are written in the conventional 5' to 3' direction.

9.6.2 Data fitting

Average readings were fitted to the quadratic equation (assuming 1 binding site) using GraFit[67], which accounts for all titration conditions:

$$[P.DNA] = \frac{([DNA] + [P_T] + [K_D]) - \sqrt{([DNA] + [P_T] + [K_D])^2 - 4\,([DNA] \times [P_T])}}{2}$$

Where:

[P.DNA] = protein-DNA complex concentration (M)

[DNA] = concentration of DNA (M)

K_D = equilibrium dissociation constant (M)

$[P_T]$ = total protein concentration (M)

9.7 Circular dichroism

9.7.1 Sample preparation

Circular dichroism (CD) is defined as the differential absorbance of left-handed and right-handed circularly polarised light, where the differential absorbance (ΔA) can be employed to deduce information about the secondary structure content of proteins and hence assess protein folding. A typical reference spectrum for CD is shown in **Figure 10**.

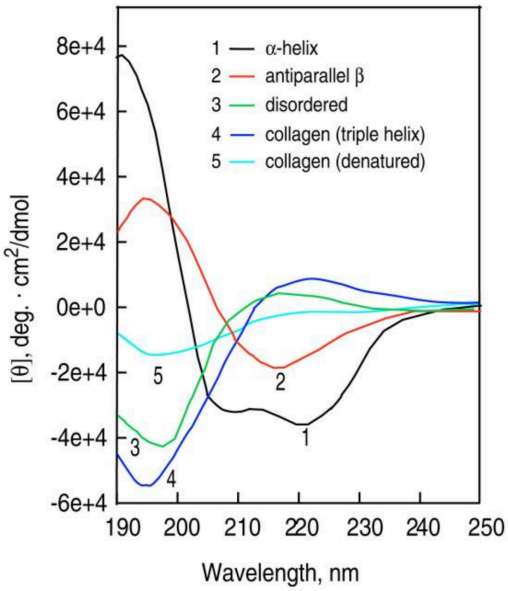


Figure 10: A typical circular dichroism reference spectrum for protein secondary structure
Taken directly from Greenfield, 2007.

ΔA and ellipticity (θ) values, which are the conventional output for CD spectra, can be interconverted using the equation:

$$\Delta A = \frac{\theta}{32982}$$

Where:

θ = ellipticity (mdeg) A = absorbance (AU)

Experiments were performed in the far-UV (190-250nm), which provides information relating to the peptide bond; whereas, near-UV (250-330nm) provides information relating to aromatic side chains. The following parameters were used for full spectrum measurements: temperature – 20°C, wavelength range – 190-270nm, data pitch – 1nm, scanning mode – continuous, scanning speed – 100nm/min, response – 0.5sec, bandwidth – 20nm, 1mm pathlength quartz cuvette and an accumulation of 4 (4 readings taken before being averaged). Whereas, for thermal stability assays, the same parameters were employed but in the range of 20-80°C, using a pitch of 2°C. Melting curves were fitted using the Boltzmann Equation; for proteins containing largely α-helices, the equation was fitted at 222nm while for proteins containing largely β-sheet, data was fitted at 214-217nm. All measurements were performed using Jasco J715 software and data ≥600V was ignored (represents detector saturation). Protein concentrations used are stated in text and dilutions were achieved with gel filtration buffer unless otherwise stated.

9.7.2 Analysis

CD outputs or ellipticity values (θ, measured in mdeg) were converted to mean residue ellipticity ([θ], measured in deg.cm^2.dm^{-1}) using the equation below. This accounts for the number of amino acid residues and enables proteins of different MW to be compared:

$$[\theta] = \frac{\left(\frac{MW}{N-1}\right)\theta}{10cl}$$

Where:

[θ] = mean residue ellipticity (deg.cm^2.dm^{-1})

θ = ellipticity (mdeg)

MW = molecular weight (g/mol)

N = number of amino acid residues

c = concentration (g/L)

l = pathlength (cm)

Protein secondary structure estimation was quantified using KD23. K2D2 (older version of K2D3[68]) estimates protein secondary structure using a reference set of CD spectra from proteins with solved crystallography structures[69]. K2D3, however, employs a larger reference set of PDB entries for which the CD spectra have been predicted.[68] For input into K2D3, ϑ values were converted into mean residue circular dichroism (ΔεMR) using the following equation. Again, ΔεMR allows CD spectra from proteins of different MW to be easily compared:

$$\Delta\varepsilon MR = \frac{\Delta A}{clN}$$

Where:

ΔεMR = mean residue circular dichroism ($M^{-1}cm^{-1}$)

A = absorbance (AU)

c = concentration (M)

l = pathlength (cm)

N = number of amino acid residues

9.8 Size-exclusion chromatography-multi-angle light scattering

SEC-MALS couples SEC, which separates proteins according to their molecular size, with MALS (**Figure 11A**); hence, enabling accurate determination of MW and investigations into the oligomeric state of proteins. Photons from the incident beam are absorbed by the macromolecule and re-emitted in all directions (**Figure 11B**). Using the Rayleigh Equation (**Figure 11C**), which defines the relationship between light scattering and MW, the MW of the macromolecule can be determined by measuring the intensity of light scattering. The intensity of the scattered light is proportional to the MW of macromolecules, hence scattered light from a dimer is twice as intense as that from a monomer.

NDP52 protein samples were subjected to SEC-MALS, using a Superdex™ 200 Increase 10/300 GL column (**Table 1**). All buffers were de-gassed before use and the column washed in Milli-Q™ H_2O at 0.1mL/min overnight before use. The SEC column was then equilibrated in gel filtration buffer (50mM Tris-HCl pH 7.5, 150mM NaCl, 1mM DTT) by running at 0.75mL/min for at least 2hrs, which protein samples were also diluted in. For each run, equipment was 'autozeroed' and protein samples were injected at a volume of at least 200µL. Once all runs were completed, the column was placed in water and then ethanol, running at 0.1mL/min overnight and 0.3mL/min for 2-3hrs respectively. Importantly, all protein samples were filtered Spin-X centrifuge tube filters (**Table 1**) before use to prevent aggregates entering and blocking the column. Measurements and analysis were performed using OmniSecBio software, where calibration was achieved with bovine serum albumin (BSA) standards (1mg/mL, 2.5mg/mL and 5mg/mL).

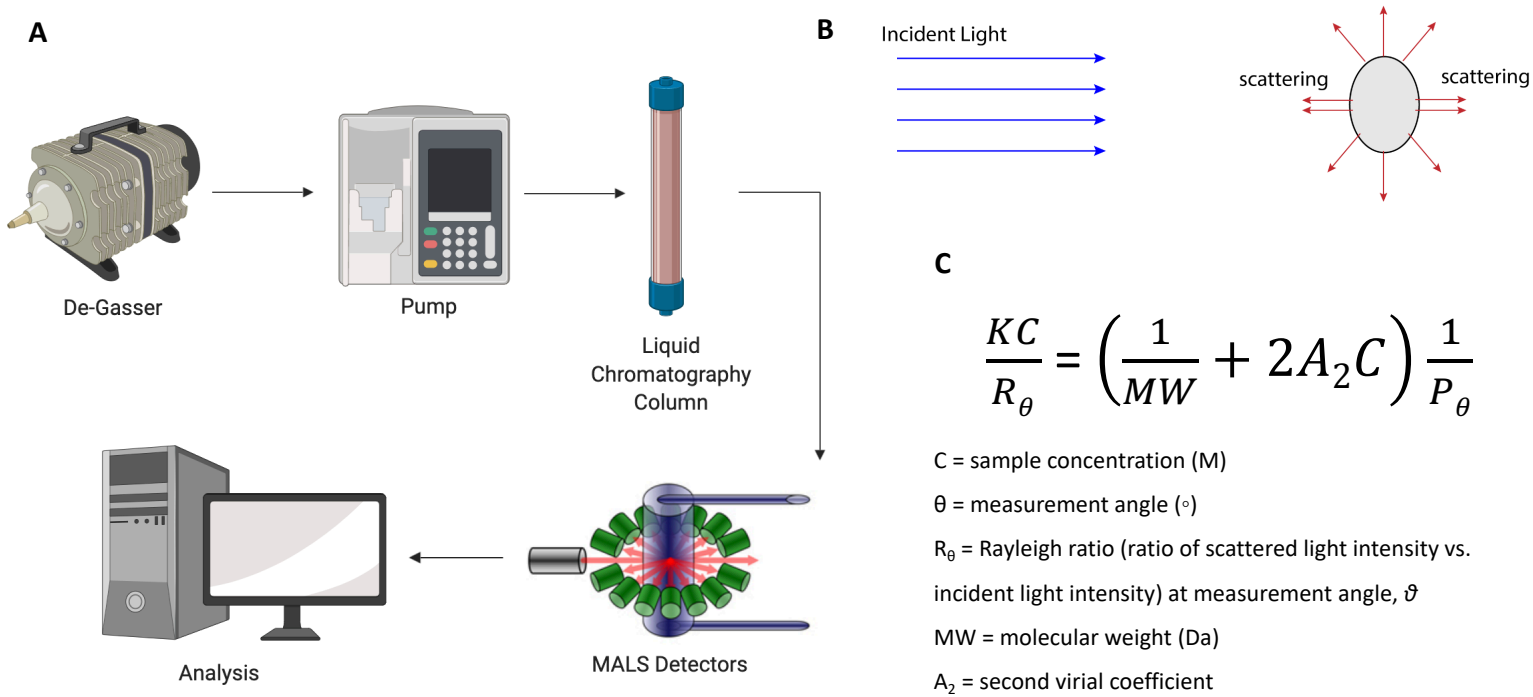


Figure 11: Setup and analysis using SEC-MALS

(A) Setup of SEC-MALS equipment. Buffers are de-gassed before being pumped/loaded onto a SEC column where MALS detectors measure the intensity of scattered light at different directions. Analysis is then performed using OmniSecBio software. Figure generated in BioRender. **(B)** Macromolecules emit photons from the incident beam in different directions, referred to as light scattering. **(C)** The Rayleigh Equation defines the relationship between light scattering and molecular weight.

9.9 Total internal reflection fluorescence microscopy

9.9.1 Etch-cleaning of coverslips

Glass coverslips were cleaned using stabilised Etch solution at 70°C, composed of a 5:1:1 ratio of H_2O:H_2O_2:NH_4OH. Coverslips were incubated for 2-3hrs at 70°C and then washed thrice in Milli-Q™ H_2O and once in absolute ethanol before being allowed to dry overnight. Before use, coverslips were placed under UV light for 10-15 minutes.

9.9.2 Sample preparation

100µL of sample (at the concentration stated in text), diluted in TIRF microscopy buffer (25mM NaCl, 20mM imidazole pH 7.4, 20mM Tris-HCl. pH 7.5, 0.1mg/mL glucose oxidase, 0.02mg/mL catalase, 3mg/mL glucose), was placed onto the coverslip before covering with an opaque object for 5 minutes (as glucose oxidase is light-sensitive). The sample was then removed, and the coverslip completely covered with TIRF buffer for imaging.

TIRF microscopy analysis makes use of photobleaching properties of fluorophores, hence the presence of an oxygen scavenger system (**Figure 12**) is important. Glucose oxidase sequesters oxygen, preventing bleaching, while catalase promotes protein activity by preventing hydrogen peroxide (H_2O_2) accumulation (**Figure 12**).

$$\text{Glucose} + O_2 \xrightarrow{\text{Glucose Oxidase}} \text{Gluconic Acid} + H_2O_2$$

$$H_2O_2 \xrightarrow{\text{Catalase}} H_2O + ½\ O_2$$

Figure 12: Oxygen scavenger system used in TIRF buffer

dsDNA for TIRF microscopy (**Table 4**) was annealed in the same way as for DNA-binding assays but was non-fluorescent as proteins were labelled with enhanced GFP (eGFP). Non-fluorescent dsDNA was used at 100nM (diluted in TIRF buffer) and incubated with protein samples for 10 minutes (in the dark) before being loaded onto coverslips for a further 5 minutes (15-minute total incubation time).

Table 4: List of oligonucleotide sequences employed for TIRF analysis

Referred to in Text	Description	Trade Name	Oligonucleotide Sequence
Non-fluorescent ds40	Random dsDNA of 40bp	A0106	GCGGGTAATCCAGAGCCAGACGA GCACTACGAACAACTAA
		A0105	TTAGTTGTTCGTAGTGCTCGTCTGG CTCTGGATTACCCGC

9.9.3 Imaging

Imaging was performed in TIRF buffer, in the presence of an oxygen scavenger (as described). Images were acquired using a ZEISS Elyra P1 system with a Plan Apochromat 100x 1.45 numerical aperture (NA) oil immersion lens. Samples were illuminated with a 488nm laser and images were acquired at 20hz with an electron-multiplying charge-coupled device (EMCCD) camera (Andor). A total of 1000 frames were recorded for each time series/cycle.

9.9.4 Photobleaching analysis

Individual GFP photobleaching events were tracked over the 1000-frame time series using a single-fluorophore detection algorithm (SFDA) and extracted using GMimPro[70]. A frame size of 5x5 and a 10-frame window were employed. The threshold for photobleaching events was set at 7 counts/pix based on the change in intensity between fluorescent and background states. This was calculated based on the number of spots detected within a record; for instance, to limit the number of false positives, while still enabling spot detection. Traces were assessed manually to identify likely (real) GFP molecules, considering GFP is expected to exist for ~5s *in vitro*. Fluorescence traces were classified as either a single or multiple photobleaching event, where a drop of intensity to 0 counts/pix was indicative of a discrete photobleaching event. The fluorescence lifetime was also measured manually by taking point-to-point measurements. A total of 10 (1000-frame) cycles were analysed for each condition.

9.10 Cryogenic stochastic optical reconstruction microscopy

9.10.1 Sample preparation

Under cryogenic (cryo-) conditions, the native ultrastructure of proteins is preserved and fluorophore properties change, namely increased brightness and fluorescence lifetime.[71] Using a super-hemispherical solid immersion lens (*superSIL*), made of both cubic zirconia and platinum (**Figure 13A**), super-resolution of 12nm of can be achieved when plunge-freezing the lens together with the sample.[71]

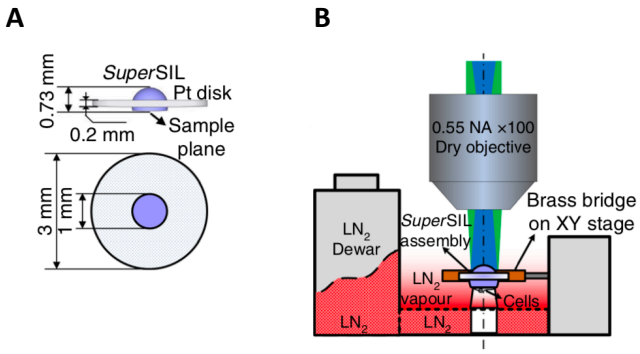


Figure 13: Setup of *superSIL* microscope

(A) Side- and top-view of solid immersion lens. **(B)** Schematic of *superSIL* microscope. Figure taken directly from Wang *et al*, 2019.

GFP-NDP52-GFP (positive control) and GFP-NDP52 (test condition) at 0.3mg/mL were diluted 1:150 in gel filtration buffer. GFP-LifeAct (negative control) at 0.085mg/mL was diluted 1:2 in gel filtration buffer and filtered by centrifugation at 4000g for 10min (Eppendorf table-top centrifuge) using a 50kDa (cut-off) Amicon Ultra centrifugal filter (**Table 1**). Following this, GFP was diluted using a 1:150,000 dilution in gel filtration buffer to obtain single-particle labelling on the *superSIL* substrate (0.042mg/mL starting concentration).

SuperSIL lenses were glow discharged for 120s at 40mA on a Quorum GlowCube machine. 2.5µl of the following was added in the order stated: sample, 100nm TetraSpeck™ fluorescent beads (**Table 1**; 1:100 dilution) and sample (again). Blotting was performed manually at every step using Vitrobot-compatible blotting paper. *SuperSIL* lenses were then plunge-frozen, together with the sample, into liquid

ethane using a Vitrobot plunge-freezer (Field Electron and Ion (FEI), ThermoFisher Scientific) before loading onto a magnetic cassette. Cassettes were then secured to a translatable stage inside a Linkam CMS196M cryogenic module for imaging. The setup of the microscope is shown in **Figure 13B**.

A 50x Mitutoyo objective, coupled to a custom microscope[71], was focussed through the *superSIL* on to the sample plane. Samples were then illuminated with a 488nm continuous wavelength laser at an approximate power density of 1 kW/cm^2 with an isotropic top-hat profile. The fluorescence signal was then collected by the same objective before being sent through an appropriate dichroic beam splitter. The signal was then projected onto a Photometrics Prime 95B scientific complementary metal-oxide semiconductor (sCMOS) camera though a narrowband Semrock 525-25nm fluorescence filter. A time series of 10,000 frames was collected per sample using an exposure time of 50ms per frame. The datasets were then saved as TIFF stacks for data analysis.

9.10.2 Analysis

The single molecule localisation software – ThunderSTORM (an ImageJ plug-in) – was chosen to analyse the datasets. The effective pixel size was defined as 46.8nm and the sCMOS camera was calibrated to 0.98 photoelectrons per analogue-to-digital-unit (ADU) count. The resultant single molecule localisations were then saved in a comma separated variable file for analysis.

The Ripley's K-function is referred to as the 'gold standard' for spatial point analysis, where the original K-function is given by the equation below[72]:

$$K(r,n) = \frac{1}{\lambda} \cdot E$$

Where:

E = expectation (average) number of events within radius, *r*

N = total number of events within the study area

λ = intensity (density) of events

r = radius

Cluster-analysis was performed on all frames using an inhomogeneous Ripley's *K*-function.[73] The inhomogeneous *K*-function considers both stationary and non-stationary points within a set area, to investigate the inhomogeneity of the distribution.[73] K(r) is calculated by looking at the average number of neighbour points within a given threshold distance (r) of each other.[74] Clustering is identified by comparing this value to that expected for a random distribution; with $K(r)= \pi r^2$ for a homogeneous Poisson distribution and $K(r)> \pi r^2$ for a clustered distribution.[74]

The L- and H-functions are based on K(r), where the L-function is the *K*-function normalised by the circle area so that the expected value is *r*, i.e. linear[74]:

$$L(r) = \sqrt{(K(r)/\pi}$$

Similarly, the H-function is the *K*-function normalised further using the equation below to get an approximate value of 0, independent of *r*, for complete spatial randomness (CSR). A positive value for H(r) is indicative of clustering while a negative H(r) is indicative of dispersion[74]:

$$H(r) = L(r) - r$$

9.11 SEC-SAXS

9.11.1 Sample preparation

SEC-SAXS, which will henceforth be referred to as SAXS, employs X-ray scattering at small angles (typically 0.1-0.5°) to deduce structural information on the nanometre scale (in the range of 1-300nm). 5mg/mL NDP52 (diluted in gel filtration buffer) was sent to Diamond Light Source (UK) for SAXS, where high intensity X-rays from the synchrotron were employed. For high performance liquid chromatography (HPLC), coupled to SAXS, a Superose™ 6 Increase 3.2/300 column (**Table 1**), which resolves proteins in the range of 5kDa-5MDa, was employed using a 0.16mL/min flowrate. As proteins elute off this sizing column, they directly enter the SAXS beamline enabling analysis by X-ray scattering.

9.11.2 Analysis

ScÅtter, developed by Robert Rambo at Diamond Light Source[75], was employed for analysis of solution X-ray scattering data. SAXS scattering images were initially plotted as the intensity of scattering (normalised to the concentration of protein) against sample number (signal plot). A threshold was then set and subtracted from readings to account for the buffer blank using the 'trace' function. A Guinier fitting was then employed to assess the quality of the sample (discussed in Results).

Raw SAXS data was then plotted as an intensity plot (intensity vs. scattering vector, q) before plotting the average and median of these results (to enable easy identification of outliers in the sample). Intensity plots are plotted in reciprocal space, where *q* is measured in $Å^{-1}$. Hence, small values of *q* correspond to large distance vectors and hence provide information relating to overall size and shape of the protein. However, larger values of *q* provide information at the atomic level.

Using an indirect Fourier transform (FT), a P(r) model was generated from the intensity plot. The P(r) model was refined by varying values of D_{max} (maximum particle dimension), where a smooth curved was achieved that approaches the x-axis gradually. Values for the radius of gyration (R_g), determined by the Guinier fitting and the P(r) function, were compared to determine if the P(r) model is well-fitted (compare R_g in real and reciprocal space). The intensity plot was also truncated to remove noise at higher *q* ($Å^{-1}$) values to improve the P(r) model.

Once the P(r) model was refined, 23 bead models were generated and then averaged. A FT was employed to back-transform onto the original intensity plot, hence improving the fit (χ^2 values used to assess fitting). From this (improved fit), another 23 bead models were generated and averaged to create the final SAXS envelope.

10 Results

10.1 Optimising the purification of oestrogen receptor α

10.1.1 Standard expression protocol

Following the standard expression protocol (Materials and Methods), ERα was successfully expressed without degradation, where expression bands are observed (**Figure 14A**) consistent with the expected MW of ERα at ~67kDa (Supplementary **Table 7**). Although, **Figure 14B** shows that there is considerable localisation to the pellet, which prevented purification by affinity chromatography. ERα, however, was shown to be soluble through localisation to the supernatant, as shown by the presence of a band in line with the 70kDa marker (**Figure 14B**). Notably, a ~35kDa band was observed to increase in band intensity during expression and was also present in both the pellet and supernatant (**Figure 14**), suggesting potential non-specific over-expression or degradation/cleavage.

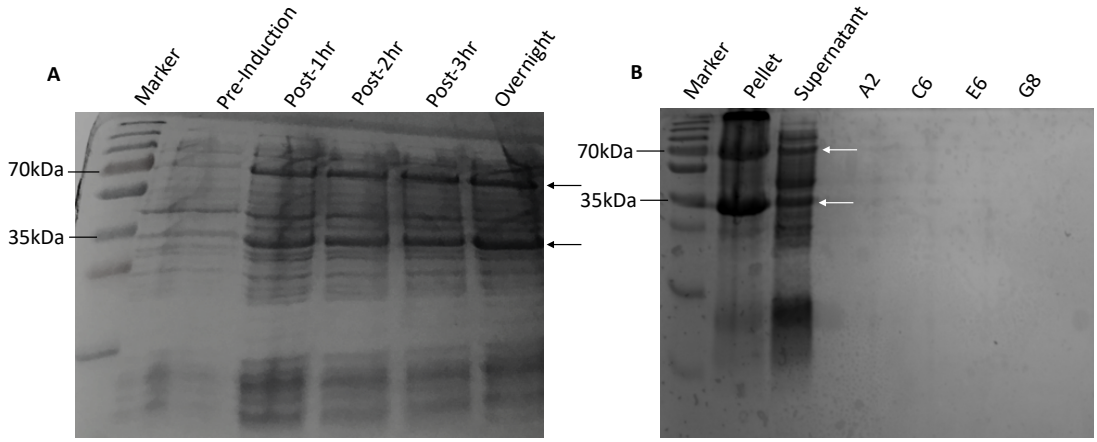


Figure 14: Standard expression of ERα

(A) Expression results. Lane headings denote timings before and after induction with IPTG. Arrows mark positions of proteins of interest and relevant molecular weight markers are labelled (as with all subsequent SDS-PAGE gels). **(B)** Results of affinity chromatography. The pellet and supernatant following cell lysis are shown, as well as elution fractions (numbered) from affinity chromatography.

10.1.2 Modified expression protocols

In an attempt to reduce localisation to the pellet, the standard expression protocol was modified so that: 1) induction was achieved with only 0.5mM IPTG and 2) culturing post-induction was only performed for 3hrs at 37°C. Other than these two distinct modifications, the standard expression protocol remained the same. These modifications will henceforth be referred to as modified protocol 1 and 2. The hypothesis was that these modifications would decrease the rate of protein synthesis. Hence, decreasing folding intermediates prone to aggregation and preventing localisation to the pellet.

To ensure that ERα was not eluting in 40mM imidazole buffer (Buffer A), imidazole was removed from the affinity chromatography equilibration buffer (Materials and Methods). Accordingly, cell resuspensions were achieved in 0M imidazole resuspension buffer (Materials and Methods). This method was employed for all ERα purifications discussed subsequent in this report.

Both modified protocols resulted in successful expression of ERα (**Figure 15A** and **B**). However, there was still a large degree of localisation to the pellet, most notably in line with both the 70 and 35kDa markers (**Figure 15C** and **D**). This was unexpected since the modifications were hypothesised to reduce localisation to the insoluble fraction (discussed above). Accordingly, ERα was not purified using modified protocol 1, as shown by the absence of bands in affinity chromatography elution fractions (**Figure 15C**).

For modified protocol 2, a faint band (red arrow) at the expected MW (~70kDa; Supplementary **Table 7**) is seen in **Figure 15D**, suggesting successful purification of ERα. This was unexpected considering the degree of localisation to the pellet (**Figure 15D**) and the results for modified protocol 1. However, this re-iterates that ERα can be solubilised and suggests shorter culturing times may be preferential. Interestingly, a band at ~35kDa (blue arrow; **Figure 15D**) is observed, following affinity chromatography, at higher intensity compared to the purified ERα band (red arrow; **Figure 15D**), suggesting degradation/cleavage of ERα.

Figure 15A

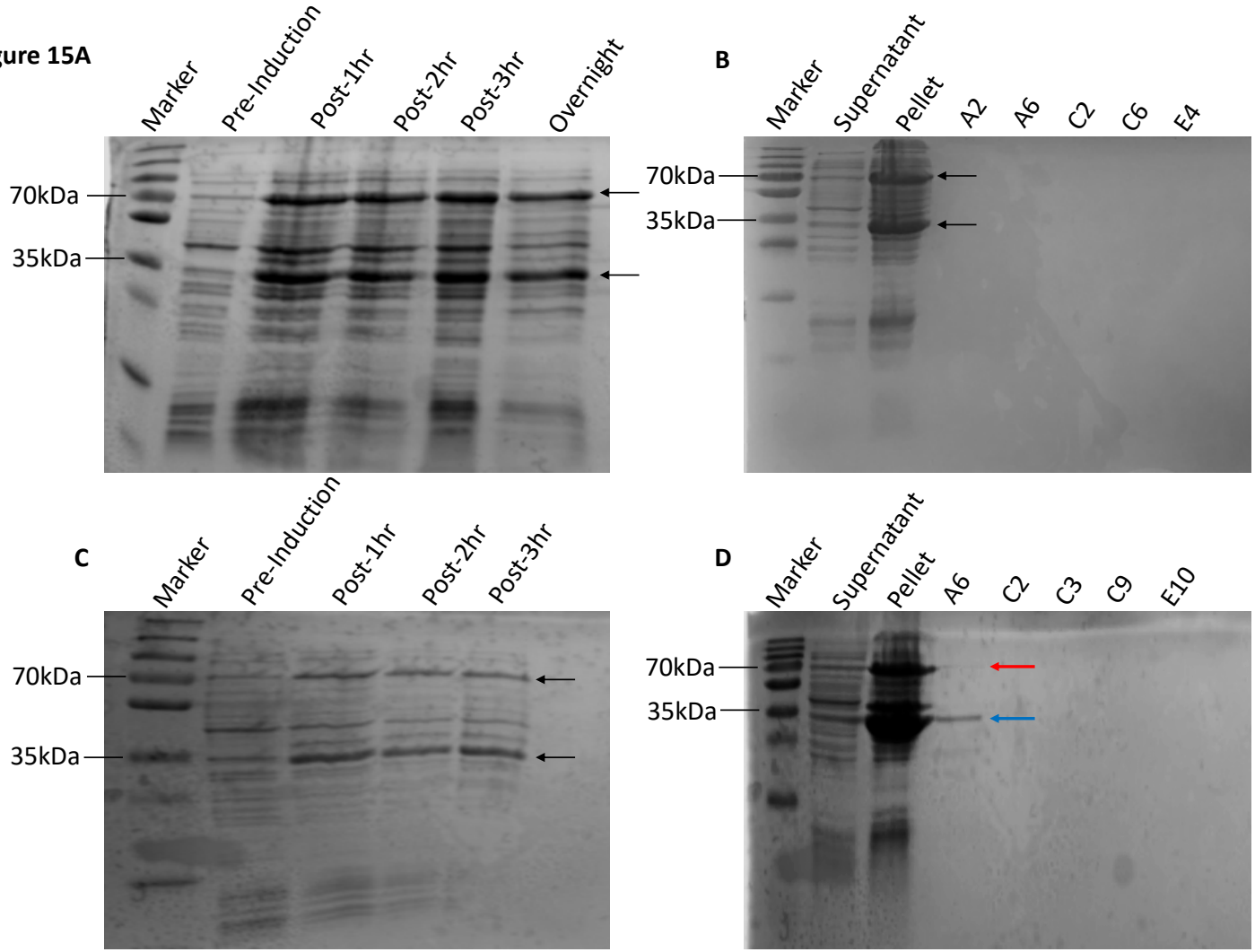

Marker
Pre-Induction
Post-1hr
Post-2hr
Post-3hr
Overnight
70kDa
35kDa
B
Marker
Supernatant
Pellet
A2
A6
C2
C6
E4
70kDa
35kDa
C
Marker
Pre-Induction
Post-1hr
Post-2hr
Post-3hr
70kDa
35kDa
D
Marker
Supernatant
Pellet
A6
C2
C3
C9
E10
70kDa
35kDa

Figure 15: Modified expression conditions for ERα

Expression **(A)** and affinity chromatography **(B)** results for ERα following modified protocol 1 (discussed in text). **(C)** and **(D)**, as for **(A)** and **(B)**, except following modified protocol 2. Timings relative to induction are given and affinity chromatography elution fractions are labelled. 'Pellet' and 'supernatant' refer to insoluble and soluble fractions following cell lysis.

10.1.3 Incubation of cell lysate with estradiol

To further optimise the expression and purification protocol, and ultimately increase purification yields, 10μM E2 was incubated with the cell lysate (following sonication) for 20 minutes prior to centrifugation and affinity chromatography. Incubation of E2 with the lysate was hypothesised to stabilise ERα, hence preventing localisation to the pellet due to degradation, something which was observed previously for both expression rounds (Section 10.1.1 and 10.1.2). Modifications to the standard expression protocol, outlined in Section 10.1.2, otherwise remained the same.

With the additional incubation of E2 with the lysate, protein bands were observed in line with the 70kDa marker (expected for ERα; Supplementary **Table 7**) following affinity chromatography. This was true for both modified expression protocols (**Figure 16A** and **B**). Although, there is again considerable localisation to the insoluble cell fraction, in line with the 70 and 35kDa markers (**Figure 16A** and **B**). This was unexpected since incubation of E2 with the lysate was hypothesised to stabilise ERα to prevent localisation to the pellet. However, solubility of ERα was confirmed, for both modified expression protocols, by the presence of 70kDa bands in the supernatant (**Figure 16A, B** and **C**), suggesting E2 promotes solubilisation of ERα.

Following modified protocol 1, a faint protein band (red box; **Figure 16A**) is observed. However, due to the intensity of the band, protein concentration was not determined as it was apparent that further optimisation of the expression protocol was required. This was expected considering the degree of localisation to the pellet (**Figure 16A**), although incubation with E2 was hypothesised to decrease this. Comparatively, when following modified protocol 2, protein bands are observed in line with both the 70 and 35kDa markers following affinity chromatography (**Figure 16B**).

Gel leakage for **Figure 16B** was questionable; hence, relevant pooled affinity chromatography fractions were re-analysed by SDS-PAGE (**Figure 16C**). However, potential gel leakage was again observed (**Figure 16C**), especially in line with the 70 and 35kDa markers. Upon measuring absorbance, an Abs_{280} corresponding to 0.6mg/mL was recorded, hence confirming the presence of protein in pooled affinity chromatography fractions. Interestingly, when re-examining the concentration of purified ERα (Materials and Methods), approximately a week following purification, a negative Abs_{280} was recorded indicative of precipitation. This was unsurprising considering consistent localisation of ERα to the insoluble cell fraction.

Due to likely precipitation (discussed above), modified protocol 2 was repeated. However, degradation was observed from 2hrs post-induction (**Figure 16D**) during the expression, which was not observed previously (**Figure 15C**). Accordingly, the soluble fraction (supernatant; **Figure 16E**) is largely devoid of ERα. In an attempt to mitigate degradation, the protocol was repeated but the post-induction culturing time was reduced to 2hrs (again, the lysate was incubated with E2). Expectedly, there was no degradation during expression (data not shown). However, following affinity chromatography, ERα was seen localised to the pellet (**Figure 16F**) consistent with previous purifications (**Figure 16B**). Unexpectedly, affinity chromatography failed to yield purified protein (**Figure 16F**). This was despite the same expression conditions (**Figure 16B** and **C**), except reducing the post-induction culturing time to mitigate degradation during expression, yielding purified protein.

Figure 16A

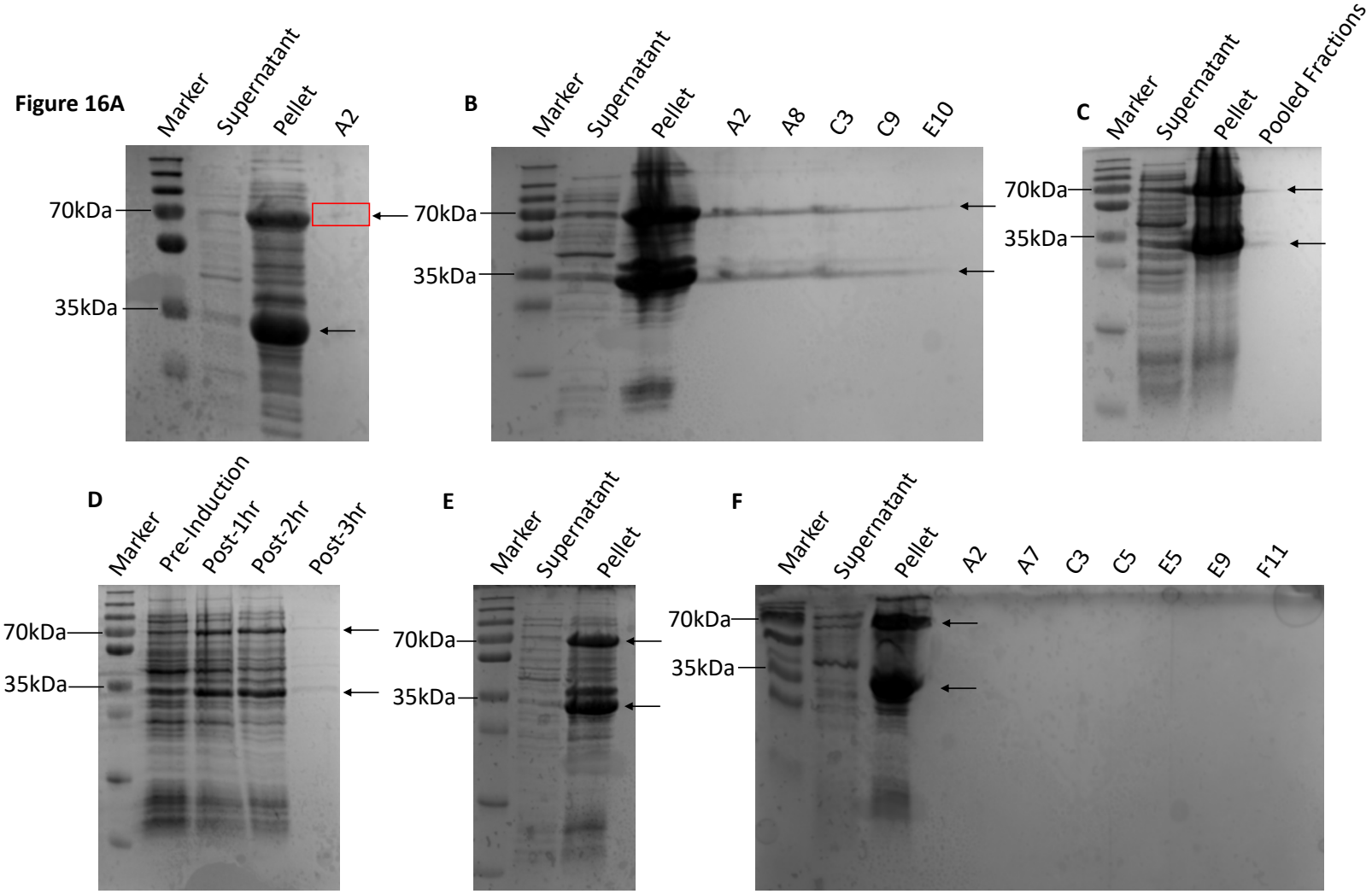

Marker
Supernatant
Pellet
A2
70kDa
35kDa
B
Marker
Supernatant
Pellet
A2
A8
C3
C9
E10
70kDa
35kDa
C
Marker
Supernatant
Pellet
Pooled Fractions
70kDa
35kDa
D
Marker
Pre-Induction
Post-1hr
Post-2hr
Post-3hr
70kDa
35kDa
E
Marker
Supernatant
Pellet
70kDa
35kDa
F
Marker
Supernatant
Pellet
A2
A7
C3
C5
E5
E9
F11
70kDa
35kDa

Figure 16: Expression and purification results when incubating cell lysate with estradiol

(A) Affinity chromatography results following modified expression protocol 1 (and incubating the cell lysate with 10µM estradiol for 20 minutes, as discussed in text). The elution fraction - A2 - in addition to the supernatant and pellet, are shown. **(B)** Affinity chromatography results for modified expression protocol 2 (again, cell lysate incubated with estradiol). Elution fractions are numbered. **(C)**, as for **(B)**, except pooled elution fractions are shown. **(D)** Re-expression of ERα, following modified protocol 2. **(E)** Corresponding pellet and supernatant following centrifugation. **(F)** Affinity chromatography results upon reducing the post-induction culturing time of modified protocol 2 to 2hrs.

10.1.4 Culturing in the presence of estradiol

Since a reproducible purification method had not been achieved, 500nM E2 was added to the culture medium (LB broth; Materials and Methods) during the expression of ERα. This was in addition to incubating the lysate with 10µM E2 (20 minutes) after sonication. It was hypothesised, like with incubation of the lysate, that the presence of E2 during culturing would stabilise the ERα hence preventing degradation and improving purification results. The sonication time was also increased to a total time of 15 minutes and 30s, employing 30s intervals with 1-minute breaks, to encourage DNA and membrane fragmentation with the hypothesis that this would also improve purification yields. All other parameters, including resuspension and column equilibration, in buffer without any imidazole, were kept the same.

For both modified expression protocols (Section 10.1.2), culturing in the presence of E2, degradation was observed from 2hrs post-induction (**Figure 17A** and **B**). On the SDS-PAGE gels, this was observed as a band intensity decrease despite the same loading volumes (**Figure 17A** and **B**). This was unexpected for modified protocol 1, since during the equivalent expression, without E2 in the culture medium, degradation was not observed (**Figure 15A**). However, degradation from 2hrs post-induction during modified protocol 2 had previously been observed, in the absence of E2 (**Figure 16D**), and this result was therefore unsurprising.

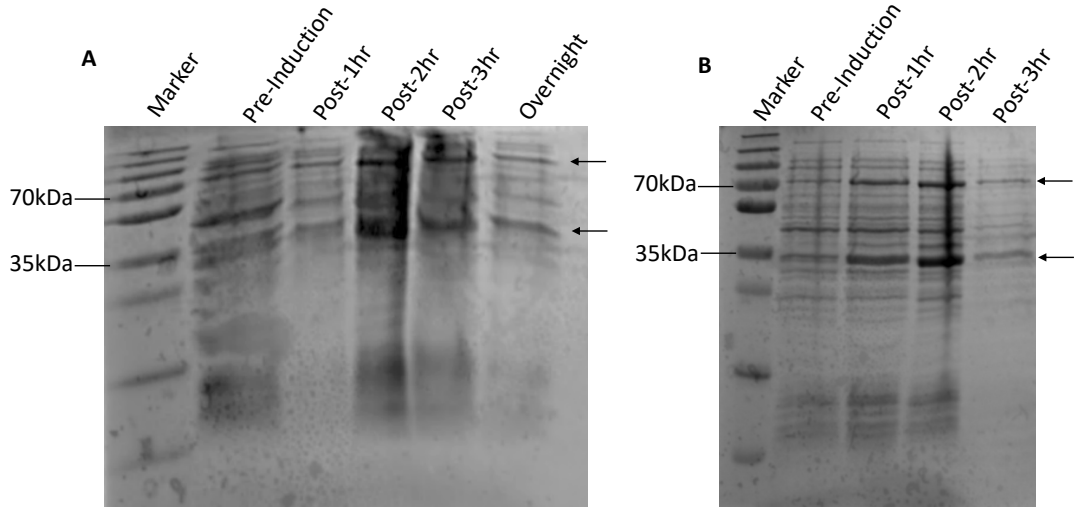


Figure 17: Expression results when culturing in the presence of estradiol

(A) Expression results when following modified protocol 1, culturing in the presence of 500nM E2 (discussed in text). **(B)**, as for **(A)**, except following modified protocol 2.

Both modified expression protocols were hence refined to prevent degradation during expression. Cells were only cultured for: 1) 2hrs at 18°C post-induction, inducing with 0.5mM IPTG and 2) 2hrs at 37°C post-induction, inducing with 1mM IPTG. Increased sonication time and incubating the cell lysate with 10µM E2 remained the same. These protocols will be referred to modified protocol 3 and 4 respectively. Expression following these protocols expectedly prevented degradation during culturing (data not shown).

Nickel-ion affinity chromatography was performed on the supernatant from both modified protocol 4 and 3, in this order. However, affinity chromatography did not yield purified protein due to an air block in ÄKTA^{TM} Pure chromatography system, which meant no high (400mM) imidazole buffer was being injected into the chromatography system (data not shown). After purging, this issue was resolved, and the resulting elution profile/chromatogram is shown in **Figure 18C**. The peaks and troughs of the chromatogram have been aligned with the SDS-PAGE gel (**Figure 18A** and **B**) through the use of coloured boxes (**Figure 18**).

Figure 18A

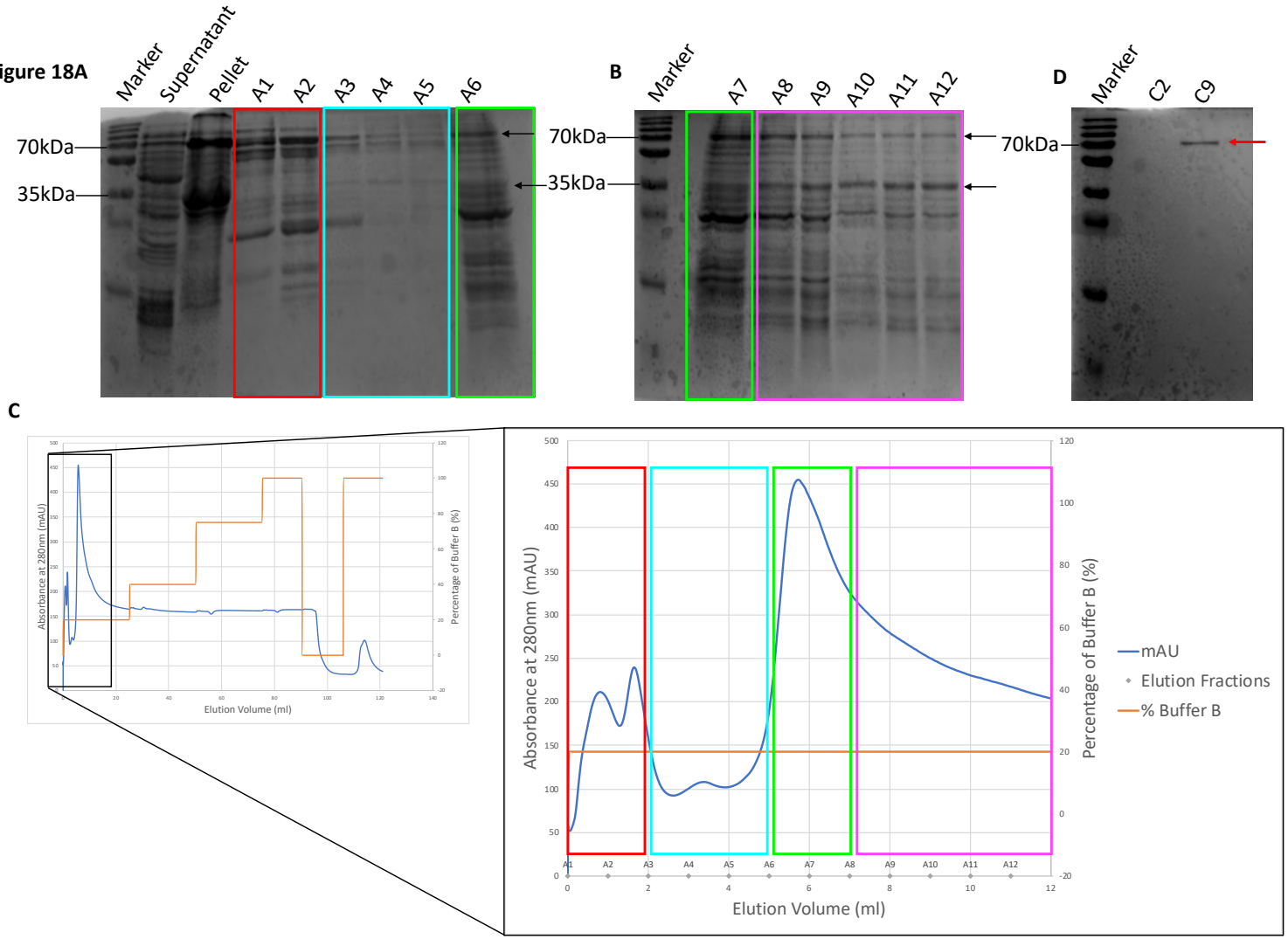

Marker
Supernatant
Pellet
A1
A2
A3
A4
A5
A6
70kDa
35kDa
B
Marker
A7
A8
A9
A10
A11
A12
70kDa
35kDa
D
Marker
C2
C9
70kDa
C
Absorbance at 280nm (mAU)
Percentage of Buffer B (%)
Elution Volume (ml)
mAU
Elution Fractions
% Buffer B

Figure 18: Affinity chromatography and gel filtration results following purging of the ÄKTA™ pump

Affinity chromatography results for the purification of ERα after purging pump B, corresponding to modified expression protocols 4 **(A)** and 3 **(B)**, as discussed in text. **(C)** Chromatogram corresponding to affinity chromatography, showing the elution profile for purification. Corresponding regions of the chromatogram and SDS-PAGE analysis are shown using colour-coordinated boxes. **(D)** Gel filtration results corresponding to pooled affinity chromatography fractions A1-12.

When monitoring the Abs_{280} using the ÄKTA™ Pure chromatography system, absorbance values were high in fractions A1-2 before decreasing in fractions A3-5 (**Figure 18C**). Similarly, in fractions A6-7, absorbance values are high before a gradual decline in Abs_{280} in elution fractions A8-12 (**Figure 18C**). This was supported by SDS-PAGE band intensities following the same pattern (**Figure 18A** and **B**).

In the SDS-PAGE gel, the positions of protein bands in line with the 70kDa marker (as well as the 35kDa marker) have been marked by arrows (**Figure 18A** and **B**). Protein bands in line with the 70kDa marker likely correspond to ERα at ~67kDa (Supplementary **Table 7**). Elution fractions A1-2 likely correspond to elution of ERα from the first affinity chromatography run (corresponding to modified protocol 4), as shown in **Figure 18A**. Elution fractions A3-5, however, likely correspond to degradation between affinity chromatography runs as there is a decrease in Abs_{280} (**Figure 18C**) and band intensity (**Figure 18A**). Elution fractions A6-12 likely represent elution of ERα from modified protocol 3, where the highest protein concentration is in fractions A6-7, supported by both SDS-PAGE band intensities (**Figure 18A** and **B**) and Abs_{280} readings (**Figure 18C**).

Gel filtration (Materials and Methods) on pooled affinity fractions A1-12 demonstrates that purified ERα elutes in gel filtration fraction C9 (red arrow; **Figure 18D**), where the corresponding protein band is estimated at ~87kDa using the SDS-PAGE calibration curve. Following concentration of appropriate gel filtration fractions, a final protein concentration was achieved of 15.8mg/mL. However, precipitation was observed *in vitro*, and, despite centrifugation, large variability in absorbance values was observed hence questioning the reliability of the protein concentration estimate. Expectedly, when measuring the protein concentration

approximately a week after purification (after centrifugation), a negative absorbance was recorded, demonstrating precipitation. This was despite a clear band on the SDS-PAGE gel (**Figure 18D**). This was unsurprising since degradation had previously been observed following successful purification (Section 1.1.1).

Successful purification of ERα, when culturing in the presence of 500nM E2 (**Figure 18D**), supported the hypothesis that E2 stabilises native ERα during folding. Although, there is still localisation to the pellet in line with the 70 and 35kDa markers (**Figure 18A**) suggesting a degree of degradation and/or precipitation even with E2 present. Both successful expressions (modified protocols 3 and 4) involved culturing for 2hrs post-induction, re-iterating that shorter culturing times are likely to be preferential (**Figure 15D**), as well as the presence of E2 in the culture medium. However, differences in temperature do not appear to significantly influence purification results (**Figure 18A** and **B**).

10.1.5 Regular additions of estradiol to the culture medium

During modified protocol 1 (culturing at 18°C overnight post-induction and inducing with 0.5mM IPTG) with E2 present in the culture medium, there is degradation of ERα 2hrs after induction (**Figure 17A**) – something that is not observed with E2 absent from the culture medium (**Figure 15A**). This is surprising since E2 is hypothesised to stabilise native ERα during folding and hence it is unexpected to see that the presence of E2 in the culture medium promotes degradation. However, it can be hypothesised that this result is due to the degradation of E2 itself during overnight culturing. To investigate this, 500nM E2 was added continually to the culture medium during expression (modified protocol 1). E2 was added at the following time points: 0hr, 3hr, 3hr post-induction and 4.5hr post-induction.

Figure 19A demonstrates that regular additions of E2 prevents degradation of ERα during culturing, as expected. Subsequently, there is purification of ERα by affinity chromatography, where the lysate was incubated with E2 prior to purification (as discussed in Section 10.1.4). This was observed on 2 distinct occasions, but protein concentrations are low demonstrated by weak band intensities (**Figure 19B** and **C**).

This re-iterates that the presence of E2 in the culture medium improves purification results and that shorter culturing times are preferential.

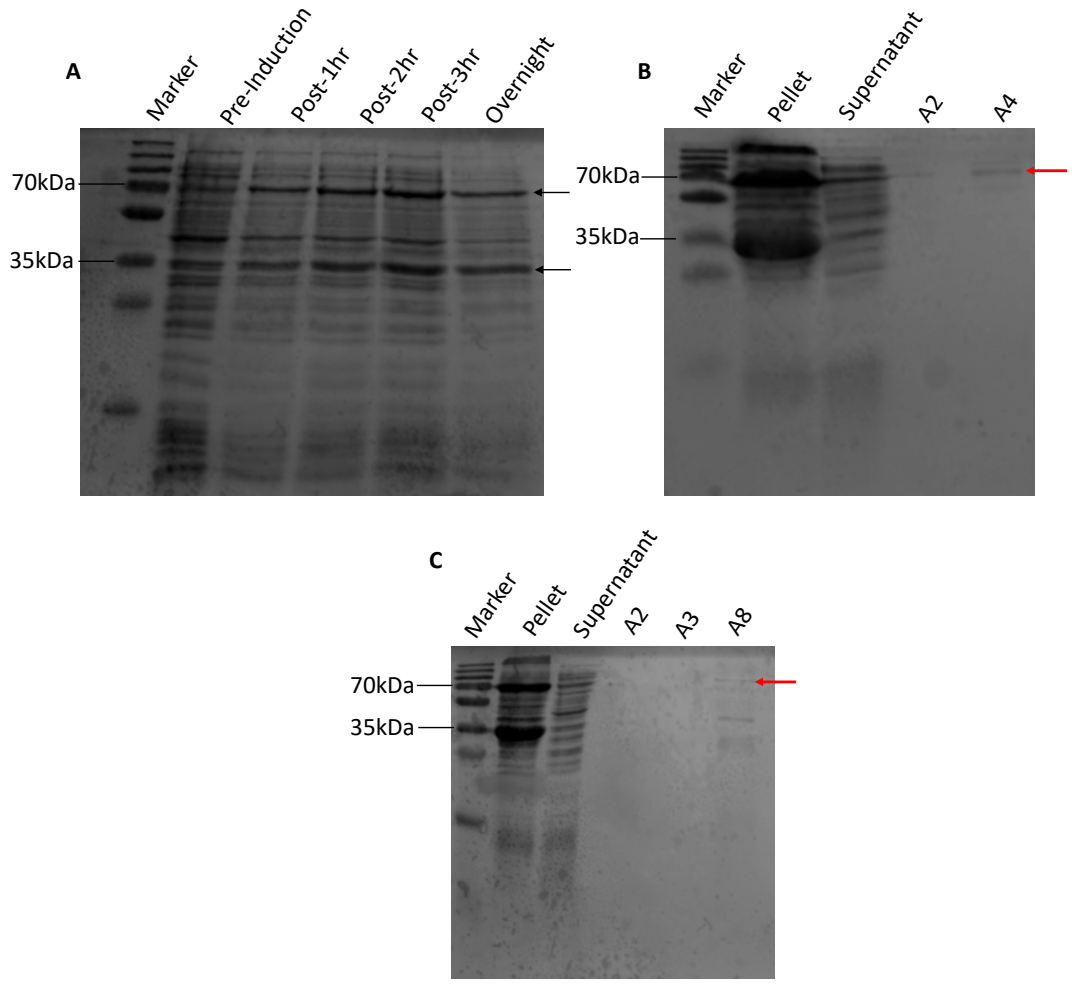


Figure 19: Expression and affinity chromatography results following regular additions of 500nM estradiol to the culture medium

(A) Expression results following modified protocol 1, where the cell lysate was incubated with 10µM E2 and 500nM E2 added to the culture medium at 0hr, 3hr, 3hr post-induction and 4.5hr post-induction. **(B)** and **(C)**. Affinity chromatography results, from 2 distinct expressions, following this protocol. Purified ERα is marked with a red arrow.

Despite the range of expression and purification conditions trialled, it was not possible to gain a high yield of stable purified protein for biochemical and biophysical characterisation. Efforts were therefore turned to another transcription regulator associated with MVI – NDP52 – which will be discussed in the following sections.

10.2 Biochemical and biophysical characterisation of nuclear dot protein 52

10.2.1 Expression and purification of NDP52 structural domains

To biochemically and biophysically characterise structural domains of NDP52, residues 1-190 and 120-end were expressed. These domains largely represent the N-terminal SKICH domain (1-190) and the remainder of the protein (120-end), with overlap (**Figure 20**). These protein domains will be referred to as NDP52 1-190 and 120-end throughout the remainder of the document.

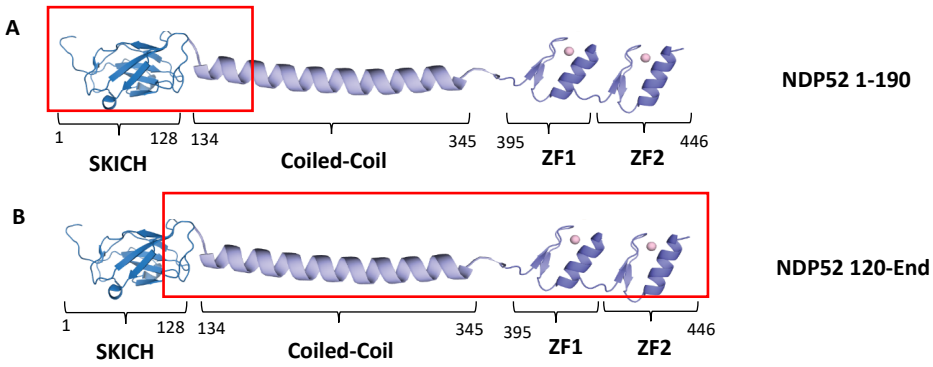


Figure 20: Schematic of NDP52 domains – NDP52 1-190 (A) and NDP52 120-end (B) – not drawn to scale

10.2.1.1 Modified standard expression protocol

NDP52 1-190 and NDP52 120-end were expressed following the standard expression protocol (Materials and Methods), modified so that cells were incubated at 27°C overnight. This decision was based on previous NDP52 expressions performed in the Toseland laboratory. Expression was successful in both cases, where protein bands are observed between the 35 and 55kDa markers (blue arrow; **Figure 21A**) and in line with the 25kDa marker (red arrow; **Figure 21B**), consistent with the predicted MWs of NDP52 120-end and 1-190 at ~40kDa and ~23kDa respectively (Supplementary **Table 7**). Degradation was observed for both NDP52 120-end and 1-190 (**Figure 21A** and **B**) as a decrease in band intensity from 3hrs post-induction, which was more apparent for NDP52 120-end (**Figure 21A**), expectedly considering its length.

Correspondingly, no protein was obtained by affinity chromatography for NDP52 120-end, which was confirmed by SDS-PAGE analysis and negative Abs_{280} readings (data not shown).

However, following affinity chromatography and gel filtration (**Figure 21C** and **D**), a high protein concentration was obtained for NDP52 1-190 suggesting degradation during expression was insignificant. On repeat purifications, similar results to **Figure 21C** and **D** were observed (data not shown). Notably, gel filtration did not result in completely pure fractions for NDP52 1-190 due to the presence of protein bands other than those in line with the 25kDa marker, such as a ~70kDa band marked with a green arrow (**Figure 21D**). This may suggest a potential protein complex or oligomerisation.

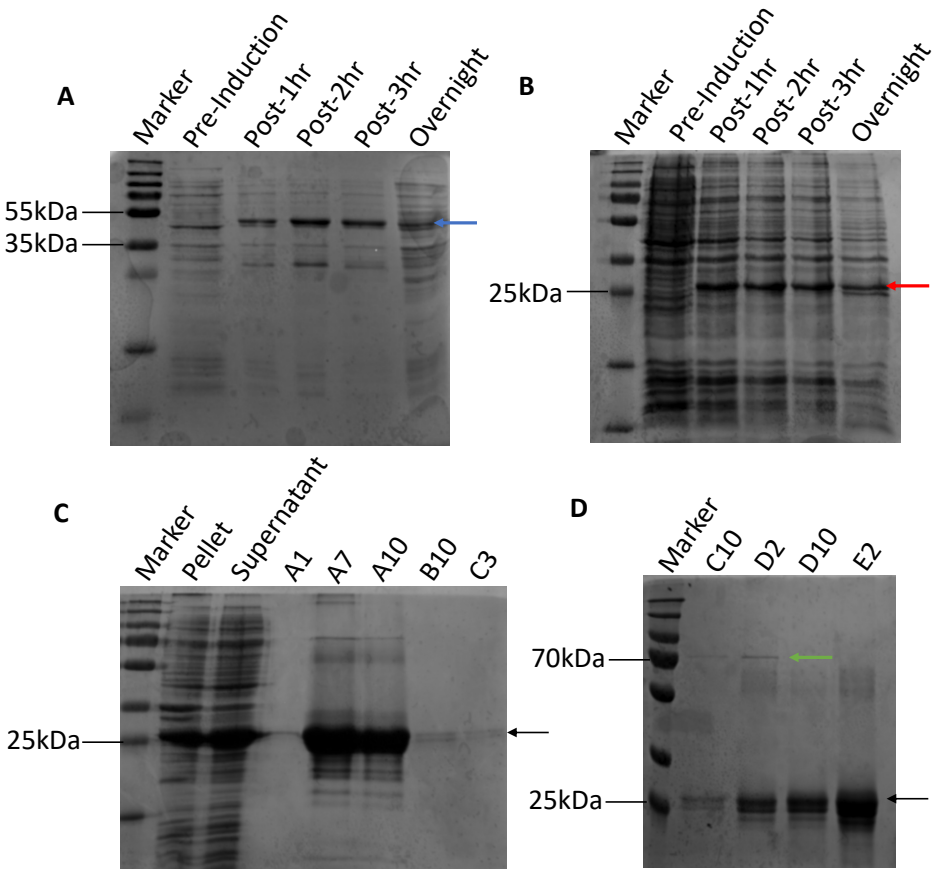


Figure 21: Expression of NDP52 120-end and 1-190 following the standard expression protocol modified to culture at 27°C overnight

Expression results for NDP52 120-end **(A)** and NDP52 1-190 **(B)**, where a representative expression SDS-PAGE gel for NDP52 1-190 is shown. Affinity chromatography **(C)** and gel filtration **(D)** results for NDP52 1-190. Elution fractions are numbered.

10.2.1.2 Modified expression protocol for NDP52 120-end

In an attempt to mitigate degradation, in order to purify NDP52 120-end, the standard expression protocol, as described in Materials and Methods, was modified so cells were only cultured for 3hrs post-induction at both 18°C and 37°C. NDP52 120-end was expectedly successfully expressed with both modified protocols (**Figure 22A** and **E**). Following cell lysis, NDP52 120-end localises to both the soluble and insoluble cell fractions (for both modified protocols) demonstrating solubilisation of NDP52 120-end and some potential degradation (**Figure 22B** and **F**). Again, for both modified protocols, affinity chromatography (**Figure 22B**, **C** and **F**) and gel filtration (**Figure 22D** and **G**), yielded purified NDP52 120-end. Although, complete purification was not obtained for neither expression protocol (**Figure 22D** and **G**) as multiple protein bands are present on SDS-PAGE gels, much like NDP52 1-190, such as those in line with the 130 and 100kDa markers (again marked with green arrows). Again, potentially reflecting oligomerisation or binding partners of NDP52 120-end.

When repeating NDP52 120-end expressions, cells were cultured for 3hrs at 18°C post-induction. Correspondingly, all NDP52 120-end results presented from this point onwards refer to these expression conditions. This decision was based on the observation that higher protein concentrations were obtained at this temperature, suggesting it is optimal.

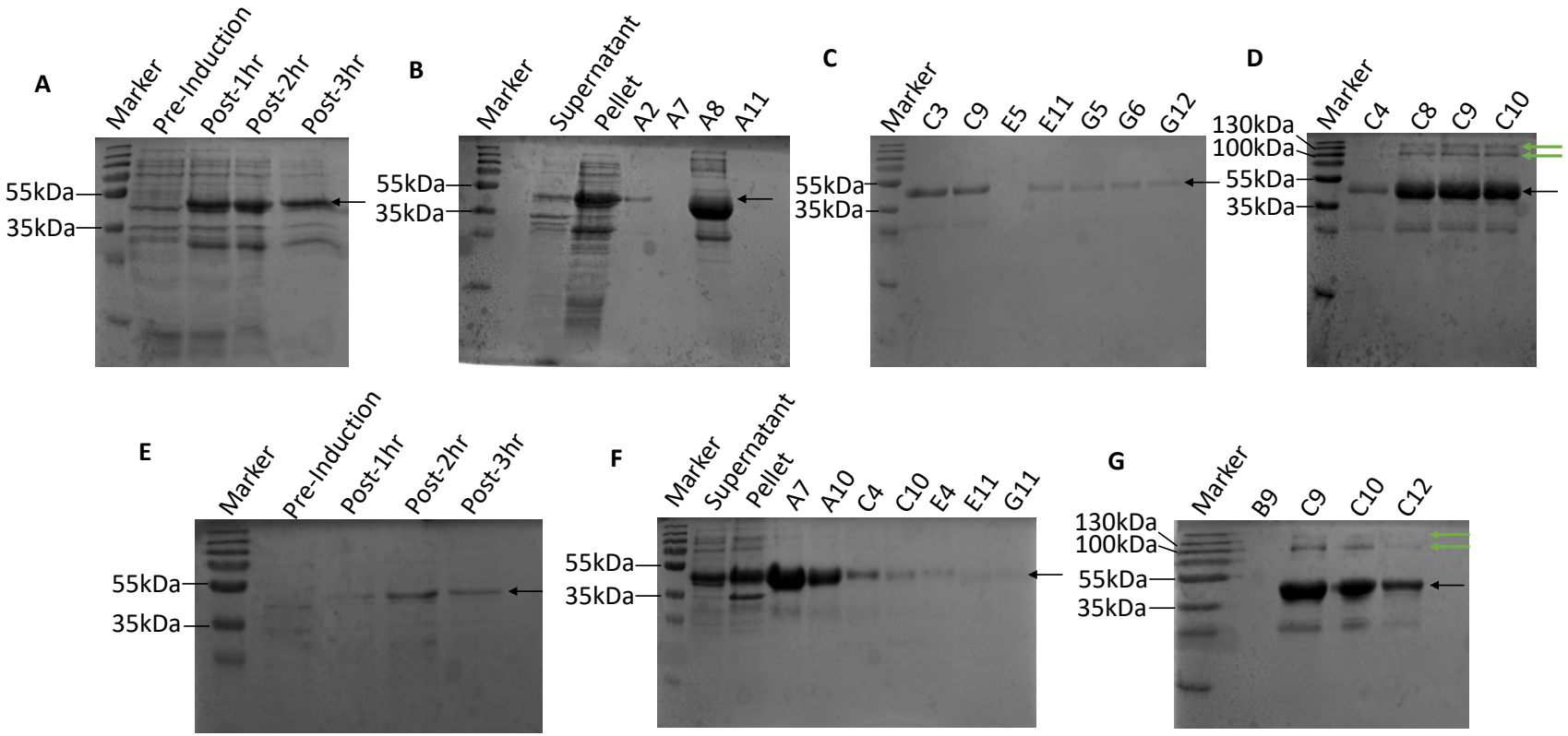


Figure 22: Modified expression protocol results for NDP52 120-end

Expression **(A)**, affinity chromatography (**B** and **C**) and gel filtration **(D)** results corresponding to the expression protocol where cells were cultured for 3hrs at 18°C post-induction but otherwise following the standard expression protocol. Representative expression **(E)**, affinity chromatography **(F)** and gel filtration **(G)** results when culturing cells at 3hrs at 37°C post-induction. Elution fractions are numbered, and expression gel headings correlate to timings relative to induction.

10.2.2 Secondary structure content of NDP52 1-190 and 120-end

CD was performed on NDP52 1-190 and 120-end to assess protein folding after purification, to determine if further characterisation was possible. NDP52 1-190 (0.1mg/mL diluted in 50mM Tris-HCl pH 7.5, 50mM NaCl) was shown to be 0.7% α-helical and 44% β-strand by analysis with K2D3[68] (**Figure 23B**). These estimates are consistent with the expected secondary structure of the SKICH domain (**Figure 23A**), suggesting purified NDP52 1-190 is suitable for biochemical and biophysical analysis (folded correctly).

Fitting the Boltzmann Equation at 215nm (Materials and Methods), gave a midpoint temperature (T_m) of 57±0.2°C (**Figure 23D**). This is clear on **Figure 23C**, where there is an obvious shift in thermal stability between 50 and 60°C. The midpoint temperature is the temperature where 50% of protein molecules are said to be in an unfolded state. This demonstrates that NDP52 1-190 is stable at RT, which fluorescence-based binding assays are performed at.

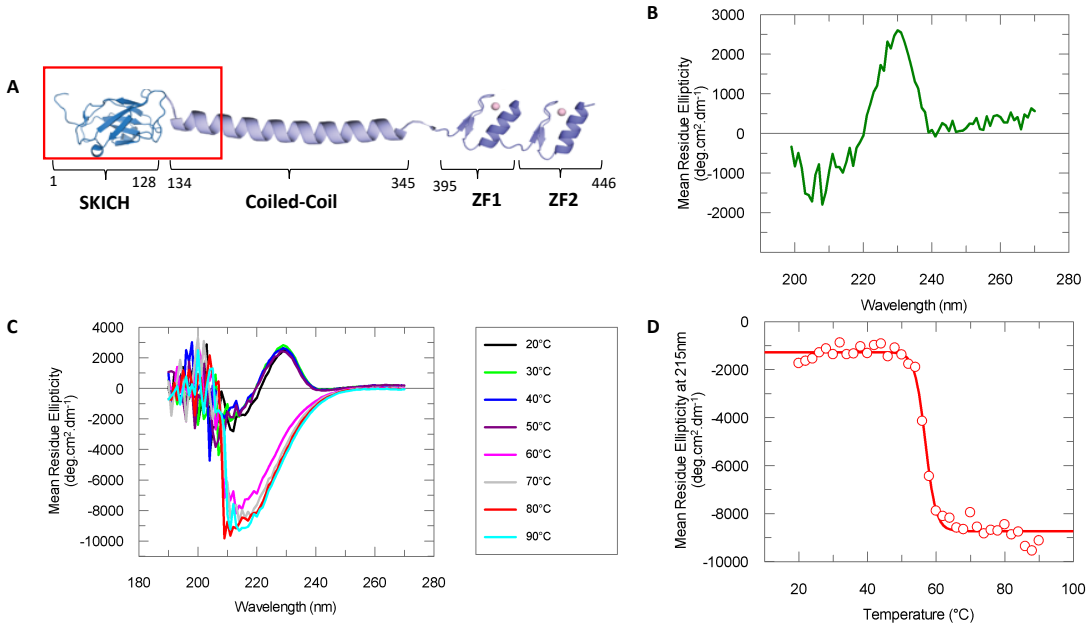


Figure 23: Circular dichroism analysis for NDP52 1-190

(A) Cartoon schematic of secondary structure content of full-length NDP52, with residues 1-190 shown in a red box (not drawn to scale). **(B)** Full spectrum measurements for 0.1mg/mL NDP52 1-190 (diluted in 50mM Tris-HCl pH 7.5, 50mM NaCl) at 20°C. **(C)** Circular dichroism spectra for 1mg/mL NPD52 1-190 (in gel filtration buffer) at varying temperatures (20-80°C) used to determine the melting temperature by fitting to the Boltzmann Equation **(D)**.

NDP52 120-end (0.1mg/mL diluted in 50mM Tris-HCl pH 7.5, 50mM NaCl), however, was shown to be 61% α-helical and 4% β-strand according to K2D3[68] (**Figure 24B**). Again, this is consistent with the expected secondary structure content of NDP52 120-end, where the coiled-coil region is α-helical, whilst the ZFs contribute anti-parallel β-sheet (**Figure 24A**). The Boltzmann Equation was fitted at 222nm giving a T_m of 50±0.2°C (**Figure 24C** and **D**). Again, this demonstrates NDP52 120-end is folded correctly and stable enough for further characterisation.

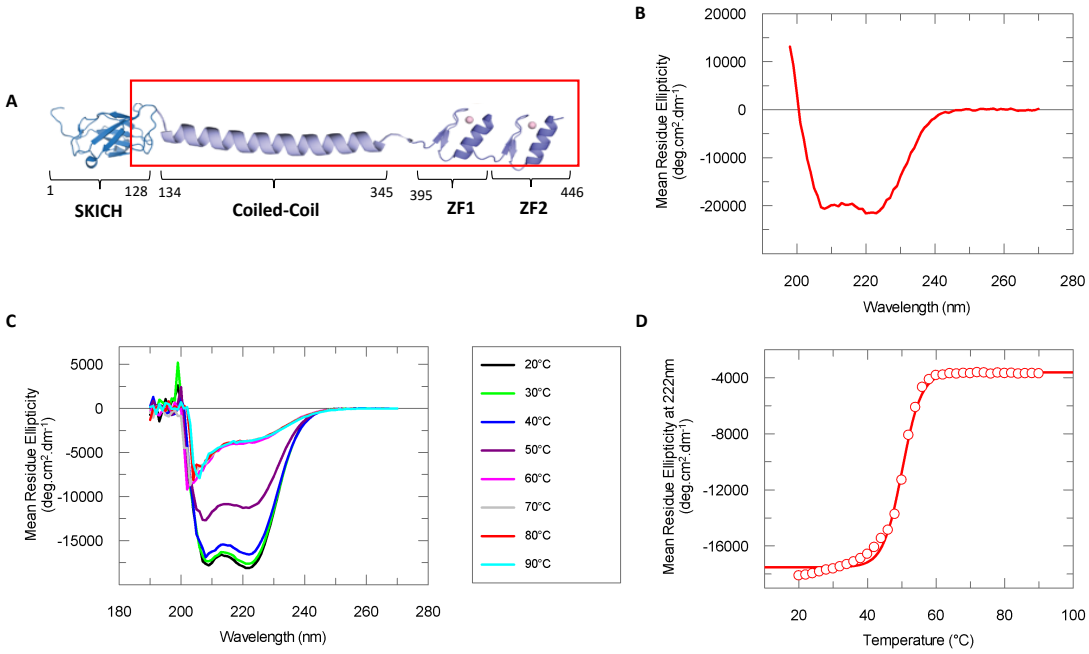


Figure 24: Circular dichroism analysis for NDP52 120-end

(A) Red box highlighting predicted secondary structure content of NDP52 120-end (not drawn to scale). 0.1mg/mL NDP52 120-end (diluted in 50mM Tris-HCl pH 7.5, 50mM NaCl) was subjected to circular dichroism analysis at 20°C **(B)**. 1mg/mL NDP52 120-end, diluted in gel filtration buffer, was also subjected to circular dichroism analysis in the range of 20-80°C **(C)**. **(D)** The Boltzmann Equation was fitted to circular dichroism data at 222nm to estimate the melting temperature for NDP52 120-end.

10.2.3 NDP52 structural domain DNA-binding affinities

Having successfully purified both NDP52 1-190 and 120-end, fluorescence intensity binding assays, as described in the Materials and Methods, were employed to investigate binding affinities for different structural domains of NDP52. Previous work in the Toseland laboratory has shown that full-length NDP52 binds dsDNA with

high affinity but, here, efforts were made to further refine DNA-binding characteristics of the protein domains. Binding curves for positive and negative controls are shown in Supplementary **Figure 41** and **Figure 42**.

Figure 25A and **B** show that NDP52 1-190 and 120-end bind to ds40 with high affinity, where both have a dissociation constant (K_D) of <100nM when fitted to the quadratic equation. K_D estimates in this order of magnitude are expected given NDP52's role in transcriptional regulation, where tight binding to DNA would be expected. Like with ds40, NDP52 120-end binds to ds15 with high affinity with a K_D of 1.1±0.3µM (**Figure 25D**). Although GraFit estimates the K_D for 1-190 binding to be 188±157nM, the error in the prediction is large owing to poor fitting of the binding curve meaning that an accurate K_D cannot be determined (**Figure 25C**).

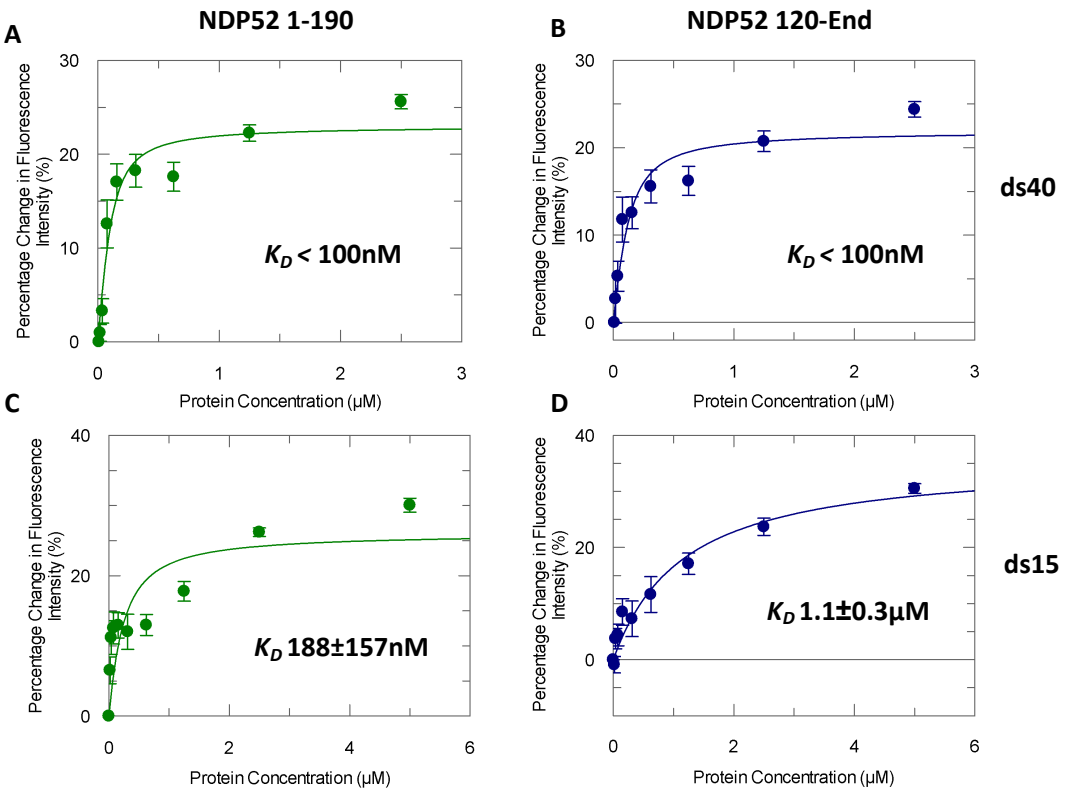


Figure 25: Binding assay results for NDP52 1-190 and 120-end

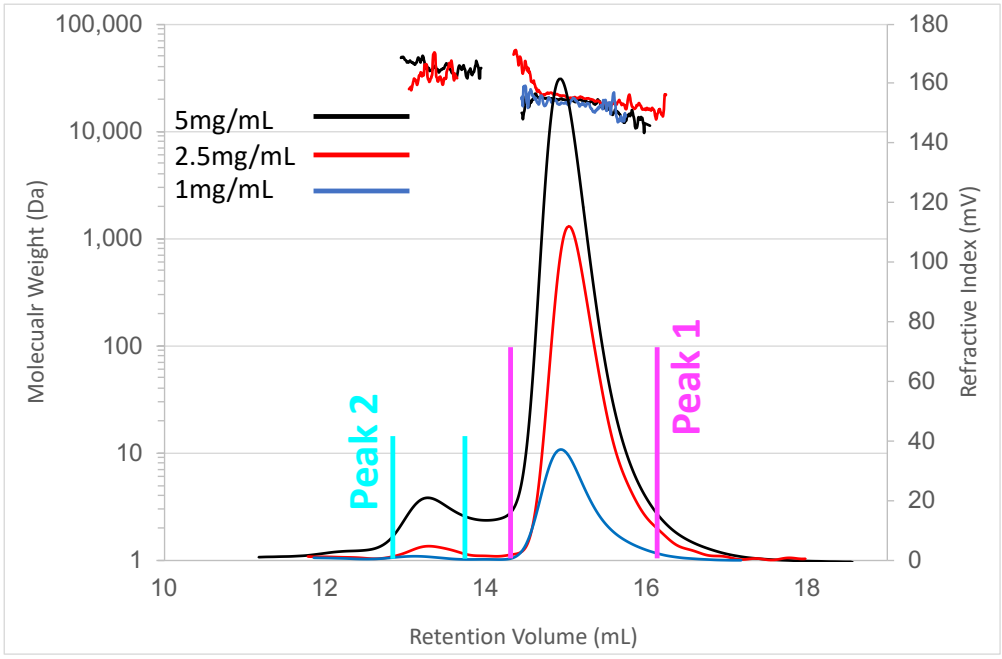
5µM NDP52 1-190 **(A)** and NDP52 120-end **(B)** were titrated against 100nM fluorescein-labelled ds40. 10µM NDP52 1-190 **(C)** and NDP52 120-end **(D)** were titrated against 100nM fluorescein-labelled ds15. Data presented is an average of at least 3 separate experiments. Dissociation constant (K_D) estimates are shown.

10.2.4 Investigating NDP52 oligomeric states using SEC-MALS

To investigate oligomerisation of NDP52 and make predictions about its tertiary structure, SEC-MALS was performed on NDP52 1-190 and 120-end domains. **Figure 26A** and **B** respectively show the SEC-MALS elution profiles for NDP52 1-190 and NDP52 120-end at 1, 2.5 and 5mg/mL. For both NDP52 1-190 and 120-end, the average MW of peak 2 was taken at 2.5mg/mL and 5mg/mL, whereas the average MW for peak 1 was taken using all protein concentrations. Anisotropic readings were excluded from the data analysis (Supplementary **Figure 43**).

For NDP52 1-190, peak 1 has an average MW of 19kDa, while peak 2 an average MW of 42kDa (**Figure 26A**). Therefore, peak 2 for NDP52 1-190 is approximately double that of peak 1. Similarly, for NDP52 120-end, peak 1 has an average MW of 33kDa whereas peak 2 has an average MW of 74kDa (**Figure 26B**); again, peak 2 has approximately double the MW of peak 1. Notably, peak 2 for NDP52 120-end does not follow a characteristic bell-shape curve, with a shoulder marked by a grey box in **Figure 26B**. This shoulder is at the limit of the resolution of the column, suggesting aggregation. This shoulder was hence excluded from calculations determining the MW of peak 2.

A



B

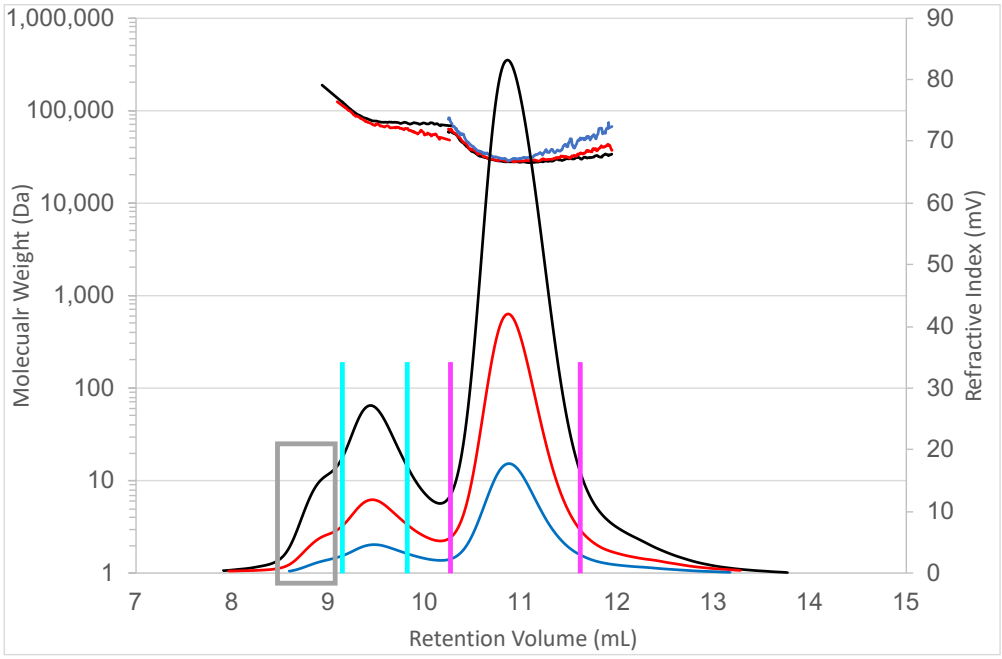


Figure 26: SEC-MALS elution profiles for NDP52 1-190 and 120-end

(A) Elution profile for NDP52 1-190 at 1, 2.5 and 5mg/mL coloured blue, red and black respectively. **(B)** Same for **(A)** except for NDP52 120-end. The shoulder of peak 2 for NDP52 120-end is marked by a grey box. Peak limits used for analysis are defined in magenta for peak 1 and cyan for peak 2.

10.2.5 eGFP-labelled NDP52 expression and purification

10.2.5.1 Modified standard expression protocol

Results from SEC-MALS elucidated that NDP52 is likely dimeric *in vitro* (see Discussion). Efforts were therefore made to further investigate the arrangement of the NDP52 dimer, namely whether it adopts a parallel or anti-parallel form, to further elucidate the role of the protein in transcriptional regulation. For instance, as an anti-parallel dimer, NDP52 may act as a linker between DNA and another protein. Whereas a parallel conformation may confer tight binding to DNA.

For these investigations, a novel GFP-based reporter assay was developed, using cryogenic super-resolution microscopy. The concept behind this design was that distances between GFP fluorophores could be directly measured. Both N-terminally eGFP-tagged NDP52 (GFP-NDP52; Supplementary **Table 7**) and doubly-tagged NDP52 (GFP-NDP52-GFP; Supplementary **Table 7**) were therefore required to act as the test sample and positive control respectively (see Section 10.2.7). GFP-NDP52 was also required for TIRF microscopy to support investigations into the oligomeric state of NDP52.

eGFP-tagged variants of NDP52 were expressed in *E.coli* following the standard expression protocol, modified so cells were cultured at 27°C post-induction, like initial expression conditions for NDP52 1-190 and 120-end. Affinity chromatography was performed using stepwise increases of imidazole (Materials and Methods). The affinity chromatography column was equilibrated in Buffer A, without any imidazole, to avoid eGFP-tagged NDP52 variants eluting in low imidazole concentrations (as observed for ERα previously). This decision was taken in the interest of time for collaboration work. All purifications for eGFP-tagged NDP52 variants, presented from here onwards, therefore employ 0M imidazole Buffer A and 0M imidazole resuspension buffer (Materials and Methods).

GFP-NDP52 was successfully expressed (**Figure 27A**) and localised to the supernatant (**Figure 27B**), demonstrating solubility. Bands consistent with the expected MW of GFP-NDP52 at ~80kDa (Supplementary **Table 7**) are marked with arrows (**Figure 27A** and **B**). Expectedly, GFP-NDP52 was purified by affinity chromatography (**Figure 27C**). However, the presence of multiple protein bands at lower MW suggests degradation (**Figure 27C**). GFP-NDP52 was, therefore, expectedly lost during gel filtration (gel not shown) likely due to degradation.

The successful expression of GFP-NDP52-GFP (~100kDa; Supplementary **Table 7**), however, is questioned by the absence of a clear expression band around the 100kDa marker (**Figure 27D**). However, following affinity chromatography, protein bands consistent with the expected MW (red arrow; **Figure 27E**) are observed in both the pellet and supernatant (demonstrating solubility), as well as in elution fractions, hence confirming successful expression. However, band intensities (**Figure 27E**) for affinity chromatography elution fractions were weak and accordingly, upon determining protein concentration, negative Abs_{280} readings were recorded.

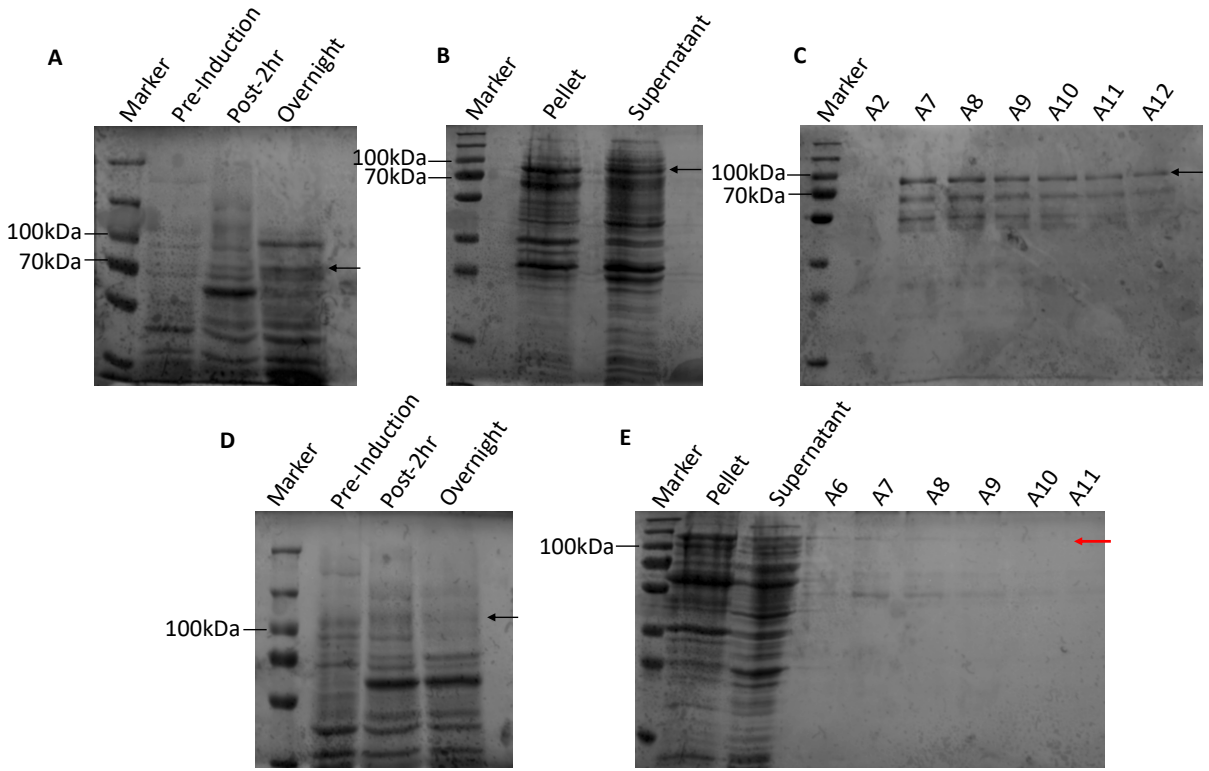


Figure 27: Expression and affinity chromatography results for GFP-NDP52 and GFP-NDP52-GFP culturing at 27°C post-induction

Expression of GFP-NDP52 **(A)** following a modified version of the standard expression protocol (culturing at 27°C post-induction) and analysis of the supernatant and pellet **(B)** following cell lysis. **(C)** Affinity chromatography results for GFP-NDP52, namely equilibrating the Ni Sepharose® column with 0M imidazole Buffer A and using stepwise elutions to elute protein fractions. Expression **(D)** and affinity chromatography results **(E)** for GFP-NDP52-GFP using the same conditions as for GFP-NDP52. The expected position of the GFP-NDP52-GFP expression band is marked by an arrow **(D)**.

10.2.5.2 Standard expression protocol

In an attempt to improve purification yields for both GFP-NDP52 and GFP-NDP52-GFP, the post-induction culturing temperature was reduced from 27°C to 18°C (standard expression protocol; Materials and Methods). The hypothesis being this would slow expression rate and hence improve purification results. Both GFP-NDP52 and GFP-NDP52-GFP were expectedly successfully expressed following this expression protocol (data not shown) and purified by affinity chromatography (using 0M imidazole equilibration buffer). Although, unexpectedly, the degree of purification by affinity chromatography (**Figure 28A** and **C**) was less compared to culturing at 27°C post-induction observed previously (**Figure 27C** and **E**).

GFP-NDP52 and GFP-NDP52-GFP were then subjected to gel filtration, and unlike culturing at 27°C post-induction, yielded purified protein (**Figure 28B** and **D**). Notably, a higher degree of purification was observed for GFP-NDP52-GFP (elution fraction C5; **Figure 28D**) compared to GFP-NDP52 (elution fraction C2; **Figure 28B**). Again, complete purification was not observed for GFP-NDP52 nor GFP-NDP52-GFP (**Figure 28B** and **D**).

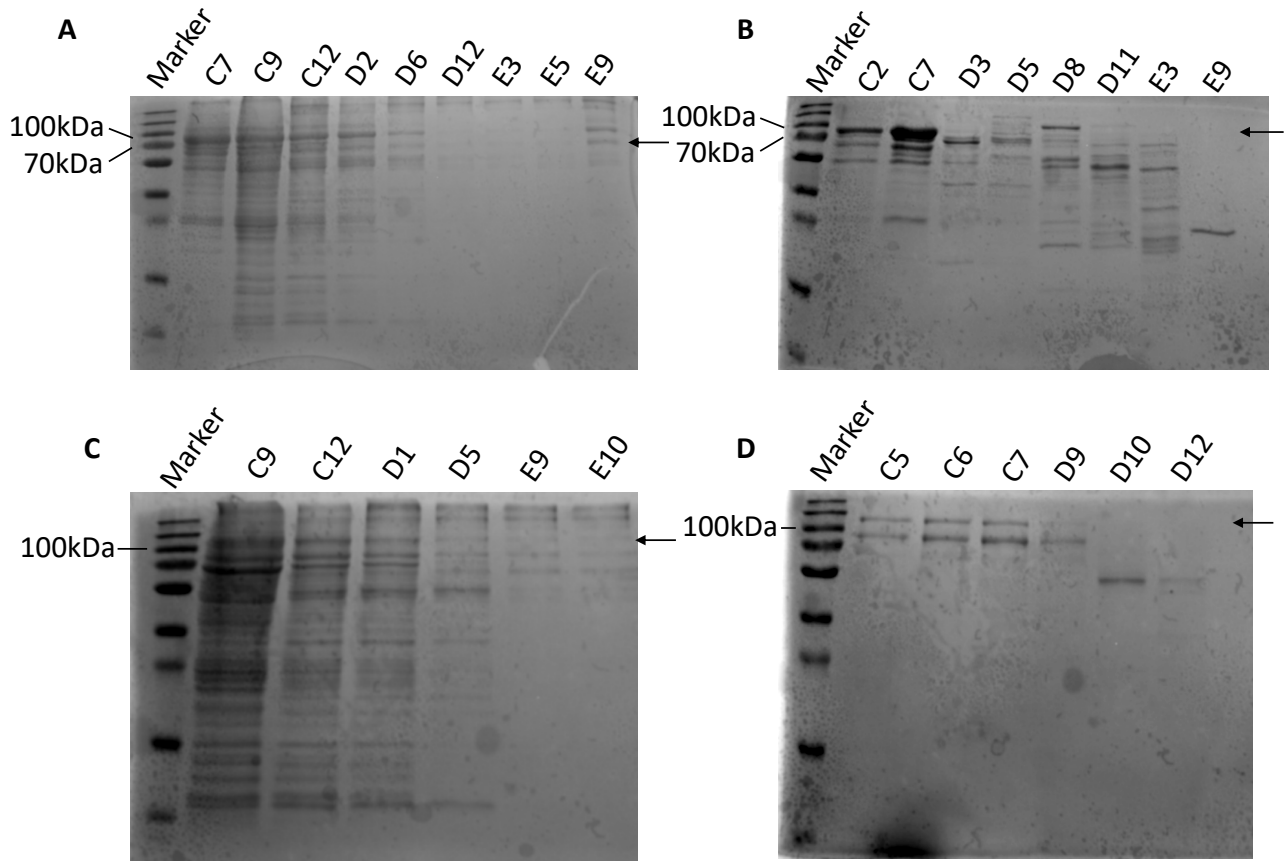


Figure 28: Affinity chromatography and gel filtration results for GFP-NDP52 and GFP-NDP52-GFP following the standard expression protocol

Affinity chromatography and gel filtration results for GFP-NDP52 (**A** and **B**) and GFP-NDP52-GFP (**C** and **D**) respectively.

10.2.5.3 Gradient elution affinity chromatography

To achieve greater resolution and yield pure fractions (not achieved prior), affinity chromatography was performed using a gradient elution (from 0-100% Buffer B) compared to stepwise elutions used previously (Materials and Methods). Similarly, gel filtration was performed at a lower flowrate of 0.2mL/min compared with 1mL/min used previously. The protocol for expressions, culturing at 18°C post-induction (standard expression protocol), as well as equilibrating the affinity chromatography column in 0M imidazole Buffer A, remained the same. Obtaining pure fractions was important for further characterisation using cryogenic super-resolution microscopy and TIRF. Results are shown in **Figure 29**.

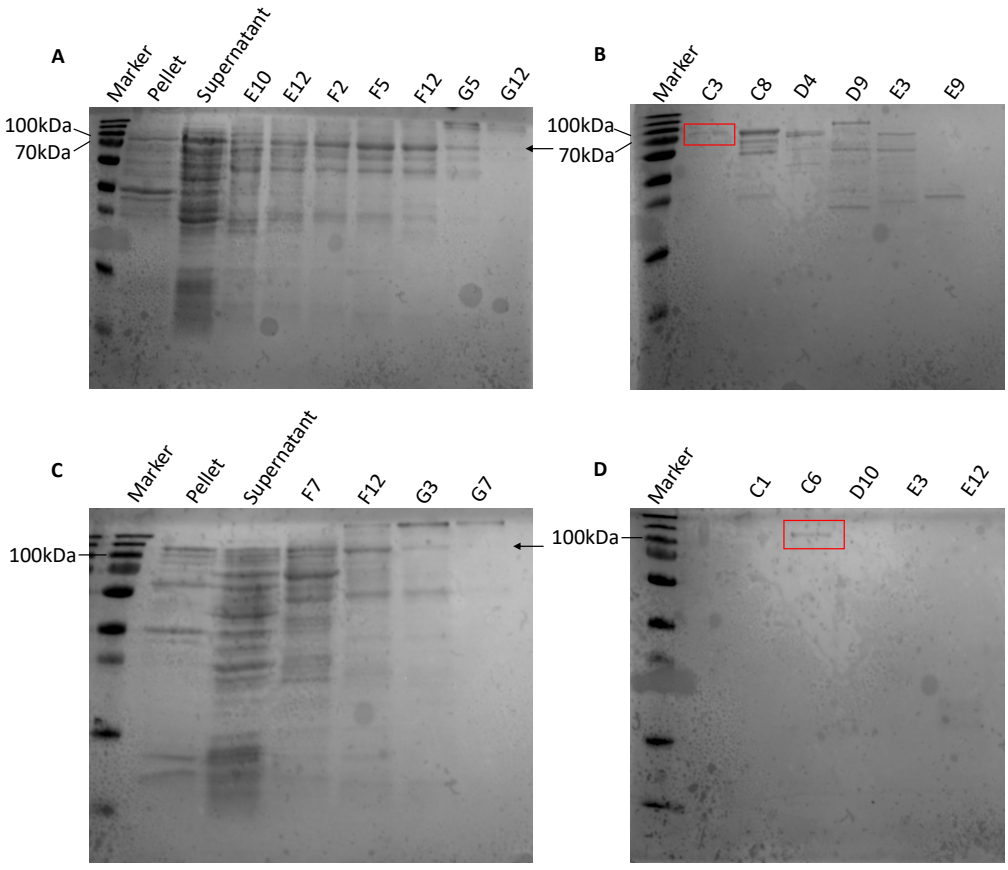


Figure 29: Affinity chromatography (gradient elution) and gel filtration (0.2mL/min flowrate) purification results for GFP-NDP52 and GFP-NDP52-GFP

Affinity chromatography and gel filtration results for GFP-NDP52 (**A** and **B**) and GFP-NDP52-GFP (**C** and **D**) respectively. Elution fractions C3 and C6, referred to in text, are marked with a red box. Other affinity chromatography and gel filtration elution fractions are labelled, as are molecular weight markers of interest.

Following the modified protocols for affinity chromatography and gel filtration, pure fractions C3 and C6 were obtained for GFP-NDP52 (**Figure 29B**) and GFP-NDP52-GFP (**Figure 29D**) respectively. This was expected since the modifications were designed to improve resolution of fractions. Unfortunately, due to the low protein concentration, accurate secondary structure content predictions from CD analysis could not be made for GFP-NDP52 nor GFP-NDP52-GFP.

10.2.6 TIRF microscopy

To further investigate if NDP52 forms a dimer *in vitro*, 5nM GFP-NDP52 was used for TIRF microscopy (Materials and Methods). **Figure 30A** shows an even distribution of GFP-NDP52 at 5nM. Photobleaching events were monitored using photobleaching analysis described in Materials and Method. It is hypothesised that a single photobleaching event likely represents a monomeric GFP-NDP52 (**Figure 30B**), while double photobleaching events likely represent dimeric NDP52 (**Figure 30C**). Example single and multiple photobleaching events are shown (**Figure 30D** and **E**).

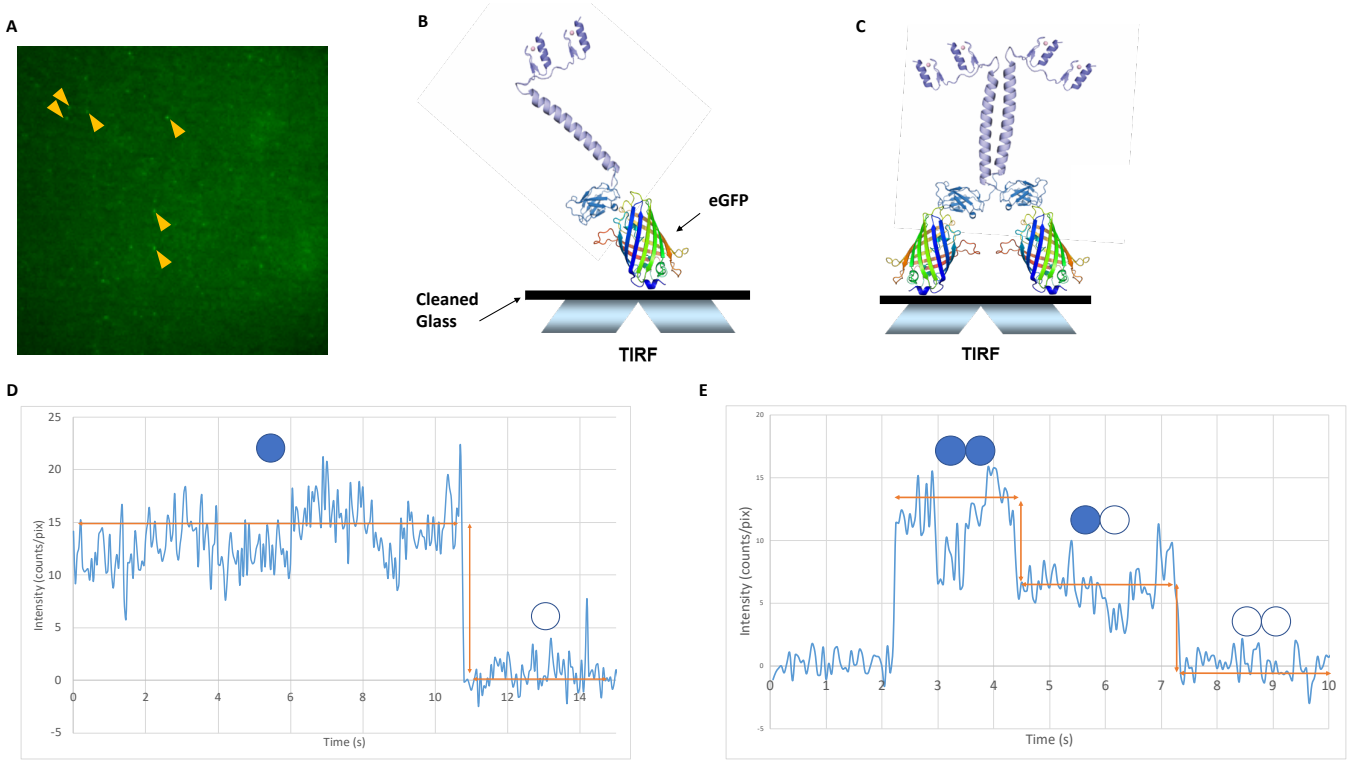


Figure 30: TIRF microscopy analysis used to assess oligomeric state of NDP52

(A) 5nM GFP-NDP52 plated onto Etch-cleaned coverslips and imaged using a TIRF microscope. Yellow arrows highlight positions of single GFP-NDP52 molecules. Schematic of GFP-NDP52 in monomeric **(B)** and dimeric **(C)** form on the coverslip. Results from photobleaching analysis showing representative single **(D)** and double photobleaching events **(E)**. Arrows approximately denote where end-to-end time measurements were taken, as well as the drop in intensity. Blue circles represent fluorescent GFP-NDP52 monomers, while white circles represent photobleaching. **(D)** and **(E)** inspired by Cabukusta *et al*, 2016.[76]

Figure 31A shows that upon incubation with 100nM DNA (non-fluorescent ds40; **Table 4**) more single photobleaching events are observed, with the total number of fluorescent events (both single and double photobleaching events) more than doubling from 104 to 211 events. However, the percentage of dimers does not change considerably both in the presence and absence of DNA at 1% and 0.5% respectively, where only 1 dimer was observed in both datasets (**Figure 31A**). **Figure 31B** also demonstrates that both in the presence and absence of DNA, there is a right (positive) skew, showing that more GFP-NDP52 molecules exhibit a short lifetime. However, in the presence of DNA, the number of GFP-NDP52 molecules exhibiting a longer lifetime increases (**Figure 31B**). In percentages, this equates to 3% of single photobleaching events with a lifetime >5s (expected for GFP) in the absence of DNA but 7% in the presence of DNA (more than double). There is also an increase in the mean lifetime of fluorescent events (single photobleaching events) from 1.4±0.1s to 2±0.2s in the presence of DNA (error denotes the standard error of the mean). The mode lifetime also increases from 0.7s to 1s, as does the maximum lifetime from 6s to 20s.

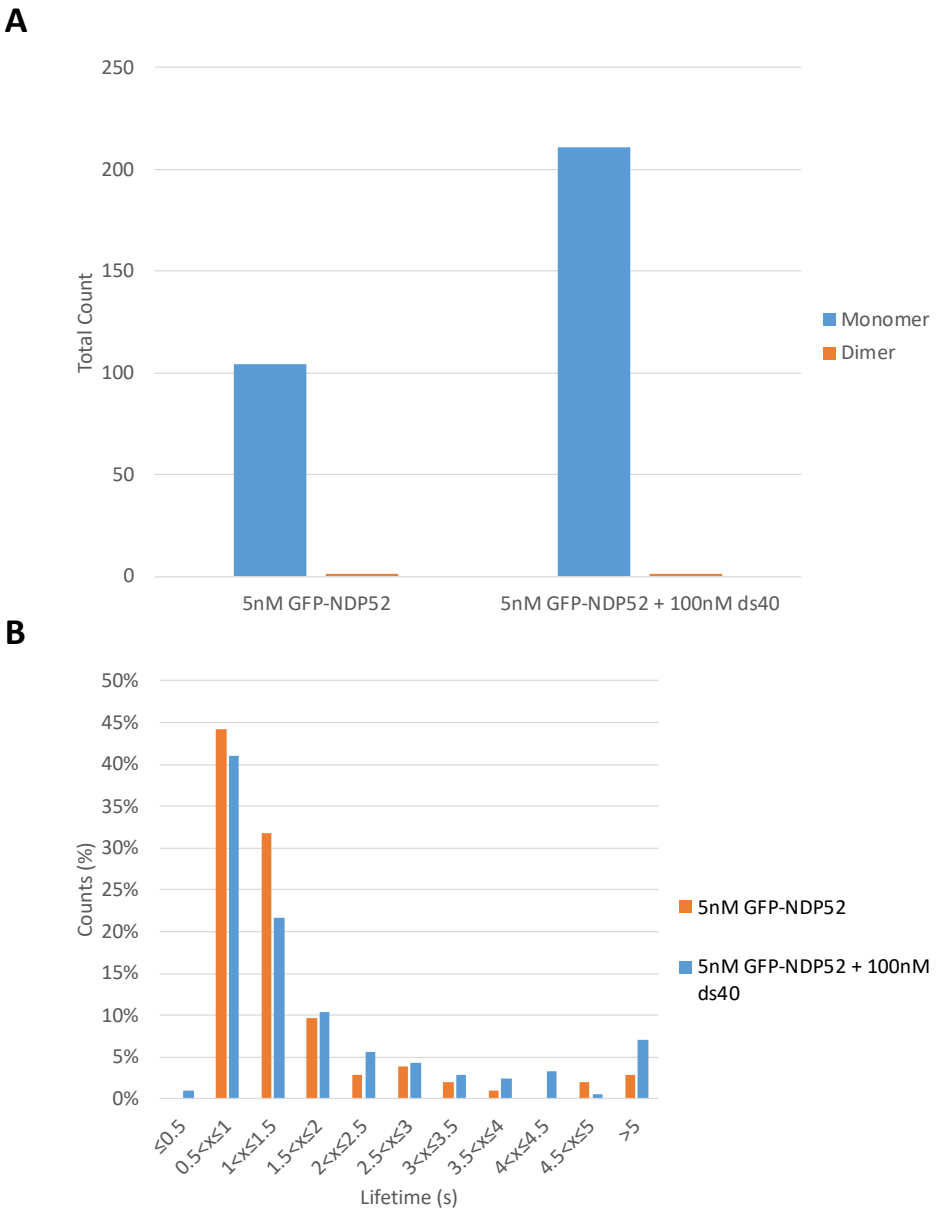


Figure 31: Results of photobleaching analysis with TIRF microscopy

(A) Bar chart showing how the total count of single and double photobleaching events (corresponding to GFP-NDP52 monomers and dimers) changes in the presence and absence of DNA (100nM non-fluorescent ds40). **(B)** Bar chart showing the fluorescence lifetime of single photobleaching events, again in the presence and absence of DNA.

10.2.7 Preliminary results for a novel GFP-based reporter assay using cryogenic super-resolution microscopy

As discussed, to investigate if NDP52 adopts a parallel or anti-parallel dimeric conformation, a novel GFP-based methodology using *superSIL* microscopy (super-

resolution) was employed. This technique will be referred to as cryogenic stochastic optical reconstruction microscopy (cryo-STORM), which makes use of photoblinking properties of fluorophores. It was hypothesised that in the parallel conformation, the distance between eGFP tags would be approximately 5nm, while in the anti-parallel conformation approximately 30-40nm (**Figure 32**). Therefore, clusters a distance of ~40nm apart would be expected for anti-parallel GFP-NDP52 homodimers. These estimates were based on previous studies with dynamic light scattering (DLS) in the Toseland laboratory. To investigate if NDP52 forms a parallel or anti-parallel dimer *in vitro*, pure gel filtration fractions (Section 10.2.5.3) were used for GFP-NDP52 and GFP-NDP52-GFP.

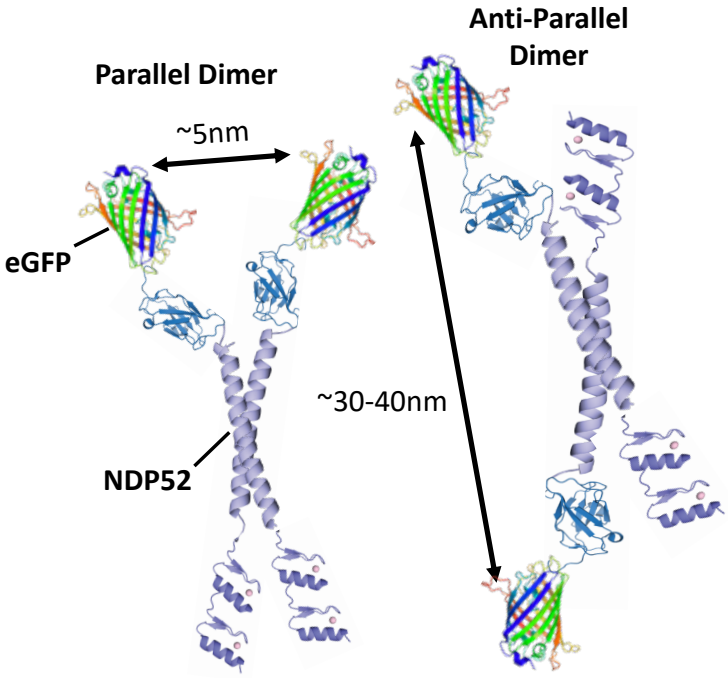
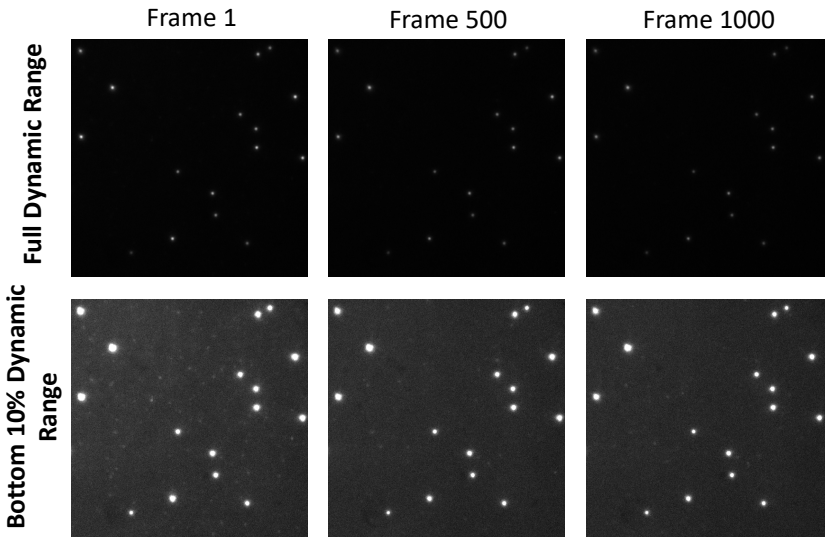


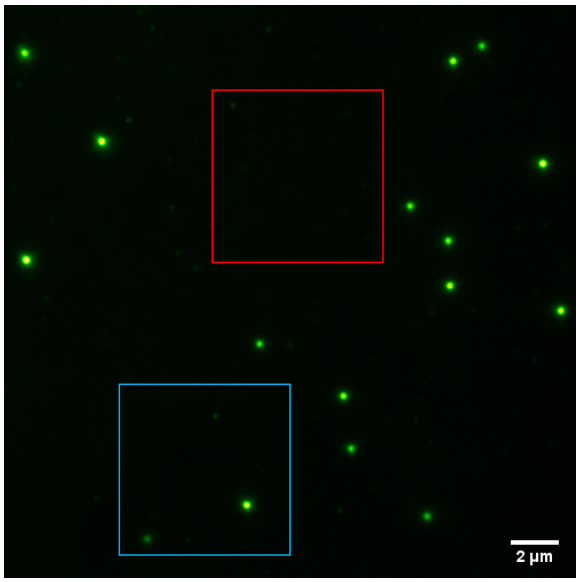
Figure 32: Schematic of N-terminally eGFP-tagged NDP52 in parallel and anti-parallel dimeric states

3 datasets were obtained for GFP-NDP52-GFP (positive control) and GFP-LifeAct (negative control), while only 1 dataset was obtained for GFP-NDP52 (test data). Importantly, this is preliminary work hence why few datasets were obtained. **Figure 33** shows raw data for cryo-STORM corresponding to GFP-NDP52 (test sample).

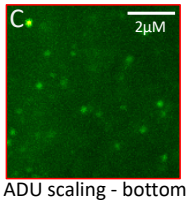
Figure 33A



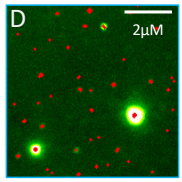
B



ADU scaling - full dynamic range



ADU scaling - bottom 10%



ADU scaling - bottom 10%

E

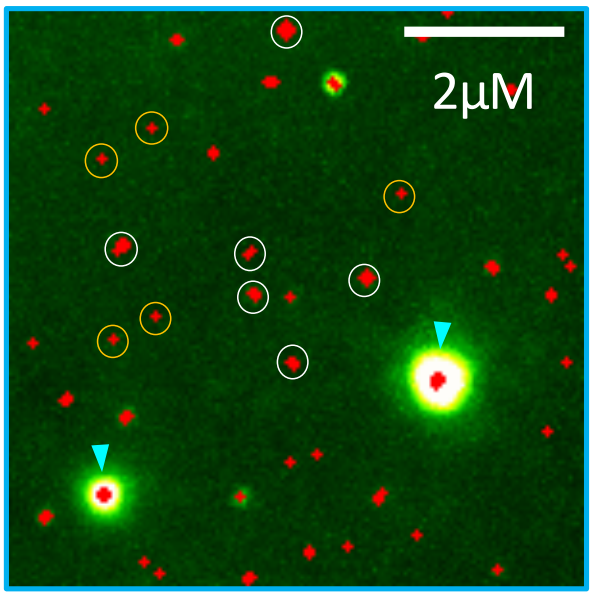


Figure 33: Raw cryo-STORM data for GFP-NDP52 (test sample)

(A) Raw cryo-STORM data for GFP-NDP52, taken at 1, 500 and 1000 frames. The first pane shows the full dynamic range while the second pane shows the bottom 10% of the dynamic range. **(B)** A maximum intensity projection of 1000 frames of cryo-STORM data showing the full dynamic range of the dataset with the TetraSpeck™ fiducial. The red inset extracted in **(C)** has been re-scaled to show the bottom 10% of the dynamic range, showing NDP52 single molecules. The valid single molecule detections from ThunderSTORM are overlaid in red on extracted inset **(D)**. **(E)** Enlargement of **(D)**, where single GFP-NDP52 molecules are circled in yellow, multiple (clusters) of GFP-NDP52 circled in white and TetraSpeck™ beads highlighted with cyan arrows.

Figure 33E shows that there is a mixed population of single GFP-NDP52 molecules and GFP-NDP52 clusters. However, under non-cryogenic conditions, at the same concentration, only single GFP-NDP52 molecules are observed (**Figure 30A**) suggesting that GFP-NDP52 aggregates under cryogenic conditions. GFP-tagged NDP52 molecules are also currently in 3D orientation, where immobilisation uniformly to the surface would improve distance measurements (Discussion). Nevertheless, some conclusions can be made regarding similarities in clustering behaviour. **Figure 34** shows a Kinhom(H') plot comparing the H(r) values (Materials and Methods) for different conditions tested. The black curve (pos_ctrl-neg_ctrl; **Figure 34**) shows that there is a difference in clustering behaviour comparing GFP-NDP52-GFP (positive control) and GFP alone (negative control). The blue curve (test_cond-neg_ctrl; **Figure 34**) also shows that there is a difference in clustering behaviour comparing GFP-NDP52 (test condition) and GFP-LifeAct (negative control). However, interestingly, when comparing GFP-NDP52 (test condition) to GFP-NDP52-GFP (positive control), the H'(r) value is approximately zero (test_cond-pos_ctrl; red curve; **Figure 34**), suggesting that GFP-NDP52 and GFP-NDP52-GFP cluster in the same way.

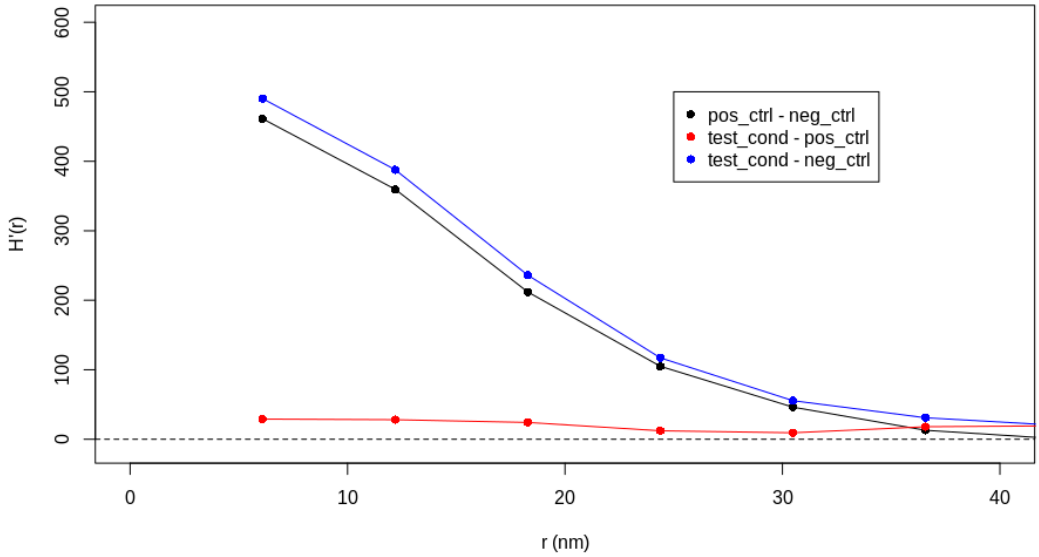


Figure 34: Kinhom(H') plot for cryogenic super-resolution microscopy conditions

Positive, negative and test conditions are discussed in text. Conditions being compared are shown in the figure legend. Data analysis and graph by Michael Hirsch.

10.2.8 SAXS analysis on NDP52

Currently, there are no X-ray crystallography structures for full-length NDP52. Therefore, SAXS was employed for structural investigations. Although, super-resolution microscopy was initially employed to investigate the dimeric conformation of NDP52, experimental and data analysis limitations (Discussion), meant that SAXS served as a good alternative to also probe the dimeric orientation of NDP52.

Full-length NDP52 was successfully expressed following the standard expression protocol and purified using nickel-ion affinity chromatography and gel filtration (Materials and Methods), as shown in **Figure 35**. Elution fractions C3-C12 were pooled and concentrated to a final concentration of 5mg/mL for SAXS.

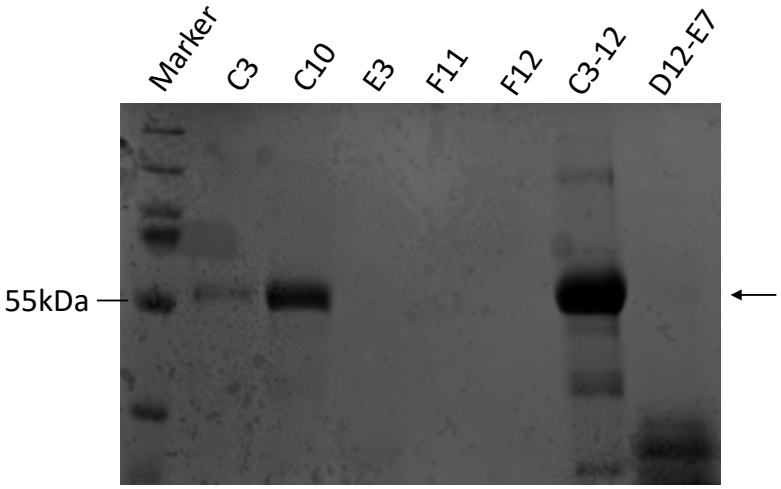


Figure 35: Gel filtration of full-length NDP52, expressing following the standard expression protocol

Figure 36 shows the signal plot overlay with R_g estimates from the Guinier equation for peak 1 and 2 of full-length NDP52 SAXS data. Using the Guinier equation, values for R_g and intensity at 0° scattering angle (I(0)) can be directly estimated (from SAXS images above the buffer blank threshold). This can provide information relating to the quality of the sample (Materials and Methods). R_g *'describes the mass distribution of a macromolecule around its centre of gravity'*[75] and can provide important information relating to the shape of the macromolecule. From the Guinier fitting, a straight line for R_g (for the sample peak) is indicative of a homogeneous sample and hence enables assessment of the quality of the data to be made.

Figure 36B shows that, for peak 1, the R_g values are extremely varied, suggesting aggregation and heterogeneity. However, for peak 2, the R_g values are largely constant enabling fitting of a straight line (red dashed line; **Figure 36C**), suggesting homogeneity of the sample.

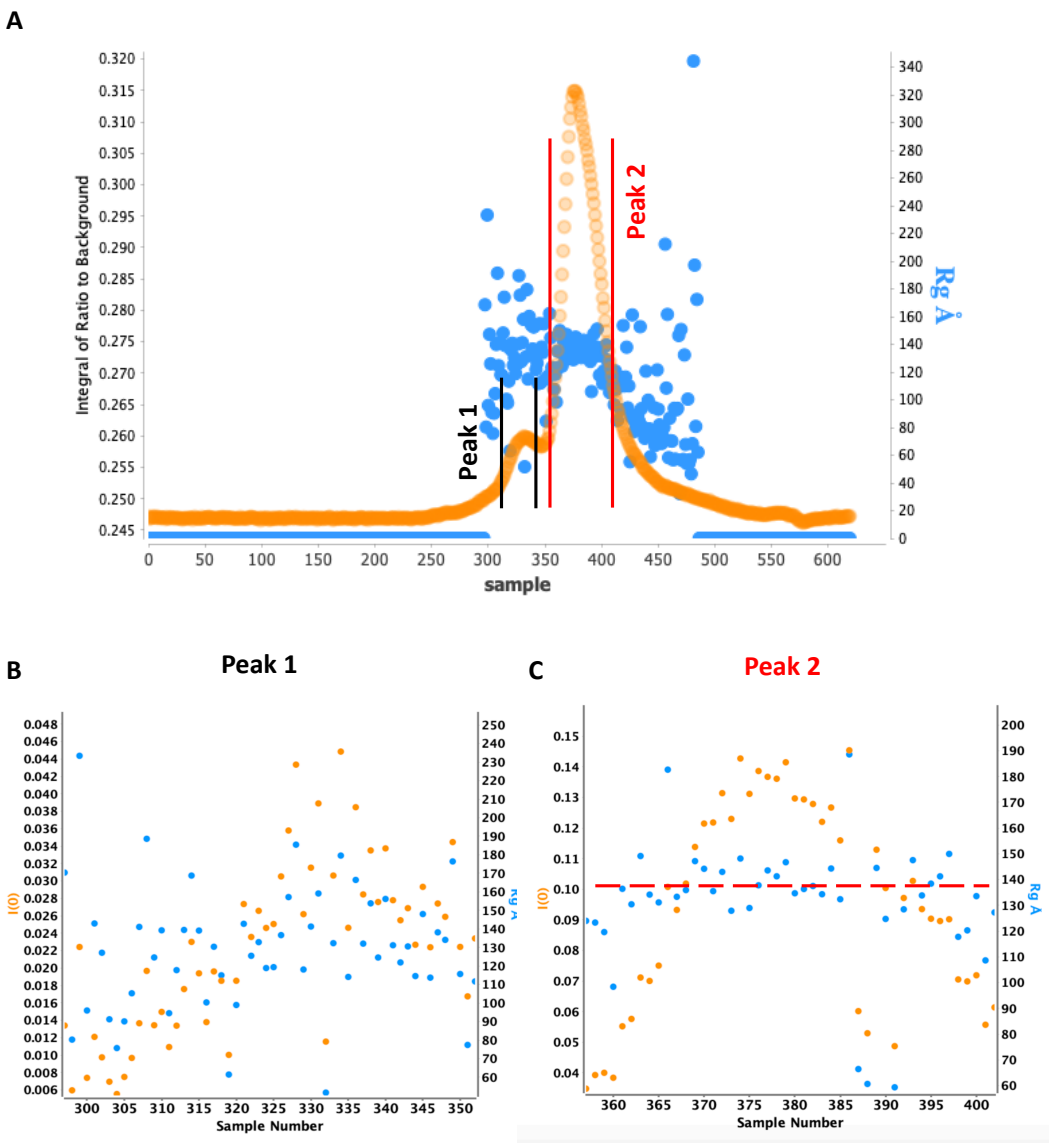


Figure 36: Guinier fitting for full-length NDP52 used to estimate I(0) and R_g values

(A) SEC-SAXS signal plot for full-length NDP52 (orange) overlay with R_g estimates from Guinier fitting (blue). In the signal plot, each point represents the integrated area of the ration of the sample SAXS curve to the estimated background. **(B)** and **(C)** represent zoomed-in versions of plot **(A)** for peak 1 and 2, where limits are defined as shown in **(A)**. The red dashed line in **(C)** shows an approximate fit for R_g values.

Figure 37 shows the intensity plots for peak 1 and 2 corresponding to raw and averaged SAXS data. Noticeably, for peak 1, both curves are noisier, especially at higher *q* (scattering vector, measured in $Å^{-1}$) values, compared to peak 2. This would be expected for a less homogenous sample, consistent with the Guinier fitting.

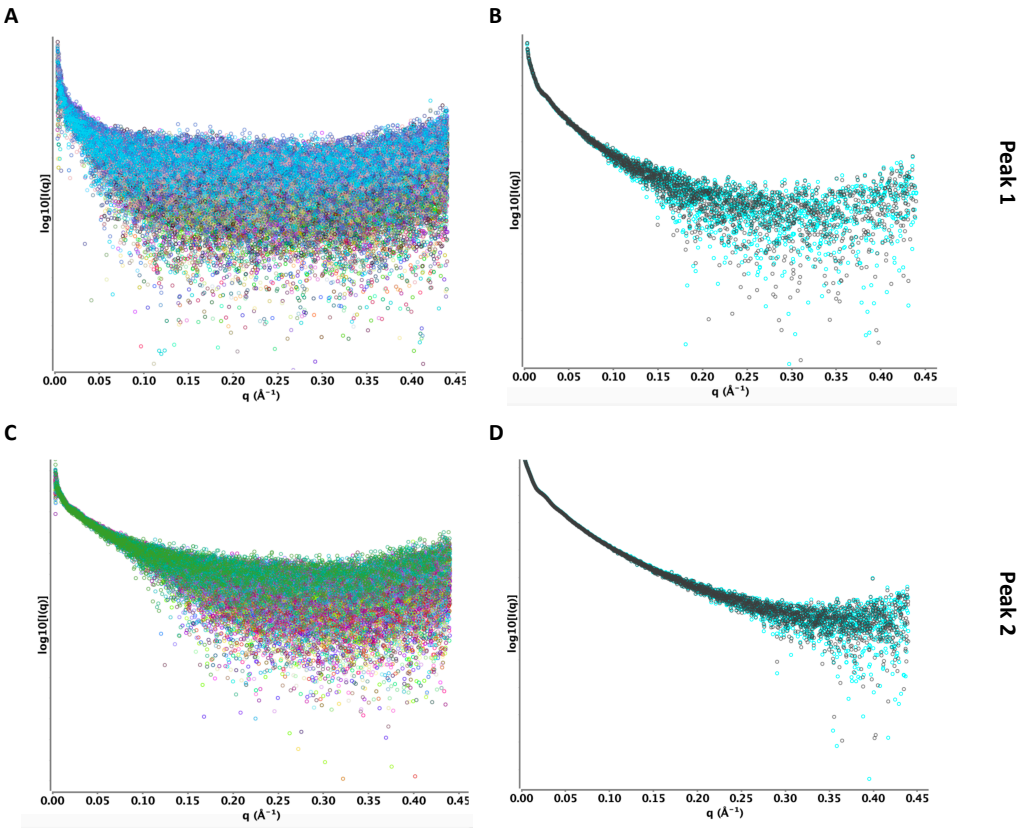


Figure 37: Intensity plots for peak 1 and 2 of NDP52 sample

(A) Raw SAXS data from the signal plot (peak 1), accounted for buffer, plotted in reciprocal space as intensity values vs. *q* (scattering vector). **(B)** Intensity plot of averaged SAX frames for peak 1. Averaged (black) and median (cyan) buffer frames are subtracted from the averaged sample frames. **(C)** and **(D)**, as for **(A)** and **(B)**, but for peak 2.

Figure 38 shows the intensity plot (**Figure 38B**) and P(r) function (**Figure 38C**) used to generate a SAXS envelope for NDP52 (**Figure 38D**). The intensity plot, as described in Materials and Methods, is plotted in reciprocal space ($Å^{-1}$) and is hence converted to real space (Å) by means of an indirect FT to create a P(r) model. The shape of the P(r) model is indicative of the shape of the protein; for instance, a bell-shaped curve is indicative of a globular structure. The shape of the P(r) model for full-length NDP52 reveals an elongated structure (**Figure 38C**), which is supported by the SAXS envelope, where full-length NDP52 is seen to adopt an asymmetric linear/elongated shape (**Figure 38D**). This is expected considering SEC-MALS data (**Figure 26**). From SAXS, NDP52 is also predicted to have an end-to-end length of 50nm (**Figure 38D**), consistent with expectation from previous DLS experiments with NDP52, where the expected end-to-end length is ~40nm.

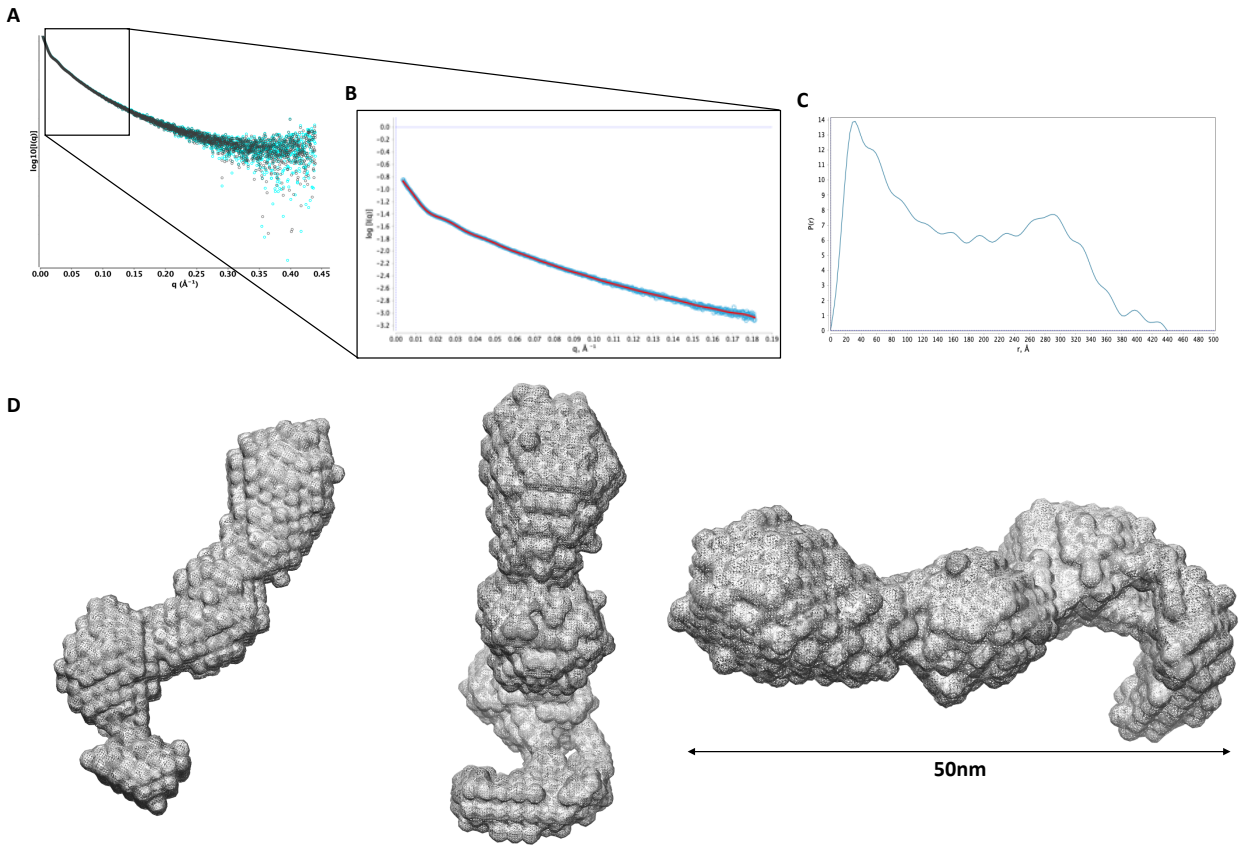


Figure 38: SAXS envelope for full-length NDP52

(A) Intensity plot of averaged SAX frames for peak 2 (as for Figure 37D). **(B)** Truncated intensity plot for P(r) function analysis. Fitting is shown by the red line. **(C)** Refined P(r) function corresponding to **(B)** and resultant SAXS envelope averaged from bead models **(D)**. The SAXS envelope is shown in different orientations.

11 Discussion

11.1 Culturing in the presence of estradiol significantly improves purification results for ERα

Many proteins of interest are present in low abundance *in vivo*, insufficient for biochemical studies, hence prompting many researchers to turn to recombinant methods and expression in a foreign host.[77] This is particularly true for mammalian cells, where it is well-characterised that steroid hormones are expressed at low levels (although over-expression can occur in cancer).[64] *E.coli* as an expression system has long been employed for such work as it demonstrates many advantages such as short generation times, efficient nutrient conversion and vast practical and theoretical knowledge regarding the organism's biochemistry, genetics and physiology.[77]

In this study, *E.coli* was employed for expression of recombinant hERα for reasons outlined above. Previously, the expression and purification of the LBD of ERα has been optimised in *E.coli* BL21 DE3 cells using a pET expression vector[78], as well as in *E.coli* TOPP-3 cells[77]. Similarly, full-length ERα has been purified from *E.coli* strains AR58 and AR68[64], as well as yeast[66] (*Saccharomyces cerevisiae*), as a ubiquitin-fused protein. While Eiler *et al* (2001)[78] employ the same expression system as that used in this study (*E.coli* BL21 DE3 cells), only the LBD is purified; similarly, when full-length ERα is purified, it is so only as a ubiquitin-fusion protein. Here, however, for the first time in the literature, full-length ERα (without ubiquitin-fusion) is purified using the *E.coli* BL21 strain.

Different expression and purification conditions were trialled to optimise the purification yield for ERα. A summary of these steps is shown in **Figure 39**. Culturing in the presence of 500nM E2 and increased sonication time markedly increased purification results compared to other methods, including incubating the lysate with 10µM E2 prior to sonication (only) as seen in **Figure 16**. This was true culturing both at 37°C and 18°C for 2hrs post-induction (**Figure 18**). This result is consistent with other studies that have demonstrated the presence of this ligand during culturing (both at 500nM and 10µM) improves purification results.[77,78] It is believed that E2

increases the stability of native ERα during folding[77,78], which, in addition to increased solubility facilitated by longer sonication times, improves the purification result. For instance, by increasing the stability of native ERα, there is a reduced population of partially denatured or unfolded ERα prone to aggregation.[77]

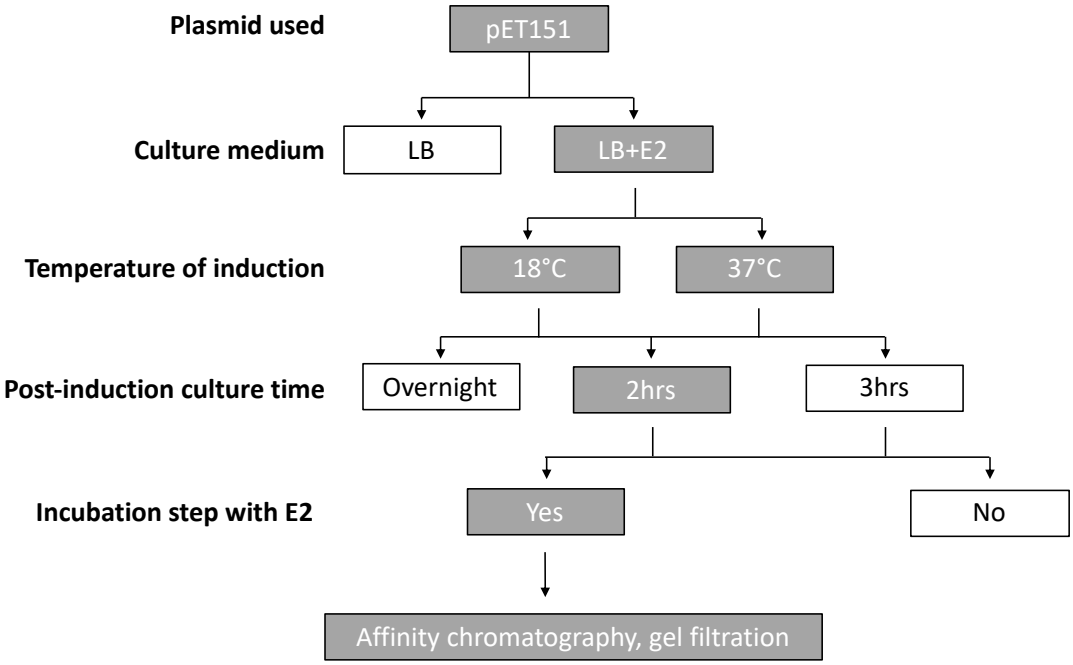


Figure 39: Summary of optimal expression and purification conditions used for ERα

Optimal conditions used are shaded in grey. Figure inspired by Eiler *et al*, 2001.

Interestingly, degradation of ERα is observed in the presence of E2 in the culture medium (**Figure 17A**) but not in its absence (**Figure 15A**) when following modified protocol 1 (culturing at 18°C overnight post-induction and inducing with 0.5mM IPTG). It was hypothesised that this was due to the degradation of E2 itself during culturing. Results presented in this study show that regular additions of E2 during culturing prevents such degradation of ERα (**Figure 19**), supporting the hypothesis that E2 is important for stabilising ERα during folding. However, ERα degrades both in the presence (**Figure 17B**) and absence (**Figure 16D**) of E2 in the culture medium when following modified protocol 2 (culturing at 37°C for 3hrs post-induction and inducing with 1mM IPTG). Presumably, this is because the expression time is too

short for degradation of E2 during culturing and there must be other contributing factors towards the degradation of ERα. It is hypothesised, therefore, that even in the presence of E2 during culturing, ERα is highly unstable and/or may be being degraded due to partial folding (due to the expression of a human gene in a bacterial expression system). This is strengthened by the observation that even when culturing in the presence of E2, there is consistently localisation to the insoluble cell fraction (**Figure 18**). Other factors that likely contribute towards difficult purification of full-length ERα are discussed in the following section.

11.2 Full-length ERα purification: limitations and future direction

Insolubility, aggregation, instability and degradation were major problems faced during optimisation and prevented further work on full-length ERα due to time. Examples of these problems are given in **Table 5**, as well as potential resolutions which should be strongly considered for future optimisations. Although, the optimal methodology presented (**Figure 39**) provides a strong foundation for further work. In the future, upon the obtainment of higher purification yields, the affinity of the interaction between MVI and ERα should be investigated using fluorescence-based binding assays, to build upon initial results presented by Fili *et al* (2017).[7] A better understanding of this interaction is particularly important if it is to be exploited in cancer therapeutics in the future, where MVI is over-expressed in breast and ovarian cancers and activation of ER target genes is associated with increased tumorigenic potential.[7]

Table 5: Problems encountered during the optimisation of full-length ERα in BL21 *E.coli*

Problem	Example(s)	Resolution
Degradation/Instability	Low protein concentrations are achieved even when culturing in the presence of E2 (**Figure 19**), which is hypothesised to stabilise ERα	Presence of E2 in lysis and purification buffers, in addition to culture medium, which has been previously shown to promote the production of soluble ERα-LBD[78]
	ERα is shown to degrade during expression both in the absence (**Figure 16D**) and presence of E2 (**Figure 17B**)	
	Band at ~35kDa observed on all expression and affinity chromatography gels, sometimes more intense than the 70kDa band for ERα (e.g. **Figure 15D**). The 35kDa band may represent non-specific over-expression or degradation/cleavage of ERα, where degradation is more likely since the 35kDa band (half of 70kDa expected for ERα) likely represents a monomer of the ERα homodimer.	Addition of protease inhibitors before, during and after sonication, which has previously been shown to be optimal for the purification of ERα-LBD[77]
	Shorter culturing times post-induction have been shown to be optimal (**Figure 15** and **Figure 18**)	
Aggregation	On 2 distinct occasions, ERα was shown to precipitate (by measuring negative Abs_{280} values) following purification	Use of 2% ethanol, in addition to E2 during culturing, where E2 is believed to reduce protein aggregation by reducing hydrophobic interactions[77]

Problem	Example(s)	Resolution
Aggregation	*E.coli* cannot perform post-translational modifications, potentially making ERα prone to aggregation due to incorrect or partial folding	Purify in insect or mammalian cell lines to ensure that necessary post-translational modifications are performed
Insolubility	Solubility of ERα has been demonstrated by localisation to the supernatant but the majority of the ERα population consistently localises to the insoluble fraction	Use of 2% ethanol, in addition to E2 during culturing, where ethanol is believed to acts as a solubilising agent for E2[77]
		Use of detergents such as Sarkosyl or Triton X
	Increased sonication times have been shown to promote purification (**Figure 18**) and reduce localisation to the insoluble fraction (**Figure 18A** vs. **Figure 15D**)	Ubiquitin-fusion to improve solubilisation, where full-length ubiquitin-fused ERα has been successfully purified using *E.coli*[64] and yeast[66]. Ubiquitin fusion is believed to improve the solubilisation of fused proteins due to its hydrophilic outer surface.[64]
		Longer sonication times/harsher sonication
Inaccessibility of His-tag	ERα is expressed well and can be solubilised (**Figure 18A**), suggesting that the N-terminal His-tag is inaccessible hence limiting the purification of ERα	Trial different affinity tag to reduce steric hinderance which potentially limits binding to the affinity chromatography column
		Expression with no affinity tag and purification using a Heparin column[65] or estradiol-Sepharose® column[66]

11.3 Residues 120-end of NDP52 likely confer binding to dsDNA *in vivo*

Previously, full-length NDP52 has been shown to bind to dsDNA (ds40) with high affinity (K_D < 100nM) in the Toseland laboratory.[7] A strong binding interaction has a small equilibrium dissociation constant (K_D) and a large association constant (K_A). In this study, the DNA-binding characteristics of structural domains of NDP52 were investigated – namely, NDP52 1-190 and NDP52 120-end. NDP52 1-190 largely represented the SKICH domain, while NDP52 120-end largely represented the coiled-coil domain, inclusive of a LZ and double-ZF motif.

NDP52 1-190 and NDP52 120-end are both shown to bind ds40 with high affinity, with K_D estimates <100nM (**Figure 25A** and **B**). Estimates cannot be made with greater precision since the DNA (ligand) concentration was 100nM. Notably, however, data points for NDP52 120-end lie closer to the binding curve, suggesting there is tighter binding achieved in this structural domain compared to NDP52 1-190. For NDP52 1-190, there is a poor fit to the binding curve (**Figure 25A**) hence it is likely that the binding observed *in vitro* is an artefact of electrostatic interactions that would usually facilitate lipid-binding by the SKICH domain[54] and binding to dsDNA is non-specific. Therefore, it is highly likely that residues 120-end confer binding to dsDNA *in vivo*, which is expected due to the presence of the well-characterised DNA-binding motifs – LZs and ZFs.

For ds15, a similar observation is made, where NDP52 1-190 and 120-end bind DNA tightly with K_D estimates in the nano and micro molar range. Considering K_D estimates, NDP52 1-190 binds ds15 tighter compared with NDP52 120-end. However, like with ds40, the poor fit (**Figure 25C**) for NDP52 1-190 likely contributes to inaccurate K_D estimation (large error in the estimation). **Figure 25C** shows that full-saturation is not reached for NDP52 1-190 (but could not be fitted due to limitations with GraFit software) as a plateau is not observed. Although, this is again likely to be an artefact of electrostatic interactions which facilitate lipid-binding *in vivo*. Similarly, for NDP52 120-end, a plateau is not achieved (K_D 1.1±0.3µM). The datapoints for NDP52 120-end (**Figure 25D**) lie close to the binding curve and the error for the K_D estimate is significantly lower compared to NDP52 1-190, demonstrating that this

interaction is likely specific (unlike NDP52 1-190). This re-iterates that NDP52 120-end is likely to confer DNA-binding *in vivo*.

For both NDP52 1-190 and 120-end, tighter binding is observed with ds40 compared to ds15 (**Figure 25**) suggesting NDP52 preferentially binds longer oligonucleotide sequences. This agrees with the numerous C-terminal DNA-binding domains of NDP52, where ds15 may not be able to facilitate all DNA-binding interactions hence resulting in lower binding affinities. The binding observed with NDP52 1-190, for both ds40 and ds15, is unlikely to be an artefact of NDP52 not being folded correctly. This is since NDP52 1-190 was purified multiple times and was predicted to have the same secondary structure content by CD analysis (data not shown) and displayed the same DNA-binding characteristics *in vitro* regardless of the purification round.

All results were compared to positive and negative controls, where NDP52 1-190 and 120-end were shown not to bind to ss40 (Supplementary **Figure 41A** and **B**) and ss15 (Supplementary **Figure 42A** and **B**) *in vitro*. Positive controls were chosen to be NDP52 295-end and ZF1 based on previous work conducted in the Toseland laboratory that demonstrated these constructs bind dsDNA reproducibly with high affinity. In this study, both 295-end and ZF1 were shown to bind ds40 (Supplementary **Figure 41C** and **D**) and ds15 (**Figure 42C** and **D**), confirming DNA was annealed correctly. However, it was observed that the binding curves did not plateau and saturation was not reached for neither ds40 (Supplementary **Figure 41C** and **D**) nor ds15 (**Figure 42C** and **D**), where fitting and K_D estimates are again limited by GraFit software.

Unlike NDP52 295-end (Supplementary **Figure 41C** and **Figure 42C**), binding curves did saturate for NDP52 120-end (**Figure 25**). Both constructs encompass the LZ domain, as well as the double-ZF motif at the C-terminus; however, NDP52 120-end has an additional 175 residues of the coiled-coil domain compared to NDP52 1-190. This suggests that the additional residues in NDP52 120-end provide stability and favour DNA binding.

In the future, DNA-binding affinities of structural domains of NDP52 should be evaluated in the presence of MVI, as well as the lipid-binding affinity of the SKICH domain *in vitro* to help further elucidate if NDP52 1-190 binding to dsDNA is an artefact of electrostatic interactions. Similarly, binding assays with NDP52 and MVI should form the basis of future work given the identification of NDP52 as a putative transcription co-activator and MVI over-expression in some cancers.[7] Having now refined the DNA-binding characteristics of NDP52 structural domains, mutagenesis to residues in DNA-binding motifs in NDP52 120-end, should also be assessed using binding assays to map key residues important for DNA-binding and hence NDP52 function as a transcription co-activator. Going forward, it will also be important to elucidate the genes that NDP52 transcriptionally regulates with MVI, particularly in relation to cancer.

11.4 NDP52 likely functions as a homodimer *in vivo*

The SEC-MALS profile for NDP52 1-190 (**Figure 26A**) shows that these residues exhibit concentration-dependent oligomerisation, where the average MW of peak 2 is 2.2 times greater than the average MW of peak 1 at 2.5mg/mL and 5mg/mL suggesting that a dimer forms at these concentrations. Peak 1 therefore likely represents the monomeric state of NDP52 1-190, while peak 2 likely represents the dimeric state (**Figure 26A**). Similarly, NDP52 120-end exhibits concentration-dependent oligomerisation, where again peak 2 is 2.2 times greater than the average MW of peak 1 (**Figure 26B**) showing that a dimer forms at 2.5mg/mL and 5mg/mL. This supports that NDP52 functions as a homodimer *in vivo.*[50]

SEC-MALS is a method for accurate MW predictions, however, for both NDP52 1-190 and 120-end the MW estimates show a degree of error compared to the expected MWs. For instance, NDP52 1-190 is predicted to be ~23kDa (Supplementary **Table 7**) while SEC-MALS predicts the monomeric state to be 19kDa (Section 10.2.4), meaning there is a ~4kDa difference. Similarly, for NDP52 120-end, SEC-MALS predicts a MW of 33kDa (Section 10.2.4) for the monomer, while NDP52 120-end is expected to be ~39kDa (Supplementary **Table 7**), meaning there is a difference of

~6kDa. Errors with BSA calibration are likely to account for these differences, where BSA was not predicted at the correct MW of 66,700Da.

SDS-PAGE analysis (data not shown) confirmed that these issues with BSA calibration were not due to degradation of BSA. Instead, reducing the flowrate of buffer to 0.5mL/min compared to 0.75mL/min improved BSA calibrations by improving the resolution. Unfortunately, SEC-MALS was not repeated due to time limitations and low protein concentrations obtained, meaning that all 1mg/mL, 2.5mg/mL and 5mg/mL concentrations could not be achieved. When repeating SEC-MALS in the future, a reduced flowrate should be employed to improve BSA calibrations and a Superose™ 6 increase 3.2/300 column, which has a 5kDa-5MDa range, should be used to improve the resolution between SEC-MALS peaks. For instance, currently the shoulder of the dimer peak (peak 2) for NDP52 120-end is observed at the limit of the column volume (**Figure 26B**).

Dimerisation of NDP52, as shown by SEC-MALS (**Figure 26**), is further supported by photobleaching analysis with TIRF microscopy, where likely dimeric (double) photobleaching events are observed. The observation of dimers at this low concentration (5nM) supports the hypothesis that NDP52 is dimeric *in vivo*.[50] Although, as discussed in Section 10.2.6, dimers only account for 1% (at most) of the observed GFP-NDP52 population. Unfortunately, higher concentrations of GFP-NDP52 could not be employed for TIRF microscopy due to over-crowding of the coverslip slide. In the future, immobilisation of GFP-NDP52 onto coverslips may improve the percentage of dimers observed at low concentrations.

The evidence presented here, which supports a homodimeric conformation for NDP52, is particularly important in relation to MVI. For instance, NDP52 has been shown to unfold/relieve the auto-inhibited state of MVI with a stoichiometry of either 2:2 or 2:4.[7] Therefore, NDP52 must be dimeric to unfold MVI, which is important for transcriptional regulation.[7]

11.5 NDP52 is largely linear with an N-terminal globular domain

For SEC-MALS, the bed volume for the Superdex™ 200 Increase 10/300 GL column is 24mL (according to manufacturer specifications), while the void volume is 8-9mL (approximately a third of the column volume). NDP52 1-190 elutes between ~12-17mL (**Figure 26A**; within the range of the bed and void volumes), suggesting a globular state, consistent with the crystal structure for the SKICH domain (PDB entry – 3VVV).[50] NDP52 120-end, however, elutes between ~9-12mL (**Figure 26B**) which is comparatively earlier compared to NDP52 1-190. This suggests that NDP52 120-end adopts a linear conformation, eluting earlier due to its Stokes radius. This observation is also consistent with the prediction of a α-helical coiled-coil domain between residues 134-350 of NDP52.[50] These results from SEC-MALS are also concurrent with SAXS, where the both the P(r) function and SAXS envelope (**Figure 38**) suggest NDP52 is elongated/linear.

11.6 The coiled-coil domain of NDP52 likely confers dimerisation and provides stability

Originally, the constructs for investigation were NDP52 1-120, 1-190 and 120-end. This design aimed to probe the SKICH domain in isolation (NDP52 1-120), where the SKICH domain is from residues 1-128, as well as the remainder of NDP52 (NDP52 120-end). NDP52 1-190 was therefore designed to represent the overlap between the 2 constructs. However, NDP52 1-120 could not be purified alone (work conducted by colleagues in the Toseland laboratory; data not shown) whereas 1-190 could. This suggests that the presence of the coiled-coil region between amino acids 134-190 (where the entire coiled-coil is between 134-350) increases stability, hence enabling purification in *E.coli*. Similarly, NDP52 120-end contains part of the SKICH domain (residues 120-128) to aid protein folding.

Figure 26 demonstrates that the dimerisation efficiency for NDP52 120-end is greater than that for NDP52 1-190, since peak 2 for NDP52 120-end has a greater refractive index compared to NDP52 1-190, especially at 1mg/mL. Although, NDP52 1-190 is also still shown to dimerise at 2.5mg/mL and 5mg/mL (**Figure 26**). Since NDP52 1-190

is characterised by the presence of the N-terminal of the coiled-coil domain (residues 134-190) of NDP52, it is likely that the coiled-coil domain facilitates dimerisation, concurrent with present belief[50].

11.7 The presence of dsDNA likely promotes the formation of higher oligomeric states of NDP52

Upon the addition of 100nM dsDNA (non-fluorescent ds40), the lifetimes of single photobleaching events for GFP-NDP52 markedly increase (**Figure 31B**). For instance, the percentage of events with a lifetime >5s more than doubles, the mean and mode lifetimes increase, and events are observed up to 20s in the presence of DNA. This suggests that the presence of 100nM non-fluorescent ds40 promotes higher oligomeric states of NDP52, although this is not observed as a percentage increase in the number of dimers, only as an increase in fluorescence lifetime. For instance, the longer lifetime suggests there may be more GFP-NDP52 molecules per spot when DNA is present, which is consistent with the presence of a dimer.

Also, the observation of more monomeric photobleaching events (**Figure 31A**) in the presence of DNA suggests NDP52 is recruited to DNA, consistent with the function of NDP52 as a putative transcription co-activator.[7] In the future, SEC-MALS elution profiles could be employed to investigate if there is a shift in oligomeric states of NDP52 in the presence of dsDNA, where you would expect a peak with a higher refractive index if more dimers were present. Similarly, in the future (as discussed) GFP-NDP52 should be immobilised on the glass surface, potentially through use of biotin-conjugated antibodies, to improve the number of fluorophores detected in the TIRF setup, which may show a percentage increase in the dimer population.

11.8 GFP-NDP52-GFP and GFP-NDP52 show the same clustering behaviour

As discussed, NDP52 has been shown to be dimeric through both TIRF and SEC-MALS analysis. A GFP-based reporter assay was hence developed to further investigate the dimeric state of NDP52, namely whether it adopts either a parallel or anti-parallel

state, which is important for structure-function insights into NDP52. For instance, an anti-parallel configuration may suggest a linker role for NDP52, with binding sides at different ends of the molecule, while a parallel configuration may suggest tight binding at one end. The hypothesis was to employ *superSIL* microscopy, which achieves 12nm precision, to measure distances between GFP fluorophores, where the use of cryogenic conditions would preserve the native structure of NDP52 and increase fluorescence lifetime and brightness of GFP.[71]

If employing a FRET-based assay, there would be high FRET between the donor and acceptor pair if NDP52 were in a parallel conformation, while in the anti-parallel conformation, there would be no FRET as the distance between fluorophores would be too great. However, a lack of FRET signal may also imply that that there are no dimers and would be indistinguishable from anti-parallel NDP52 homodimers. Also, since NDP52 is known to dimerise in the nano molar range, from previous experiments in the Toseland laboratory using microscale thermophoresis (where the dimerisation K_D is between 200-500nM), there would be a low concentration of mixed FRET populations. For instance, there would be a low concentration of dimers tagged with both GFP and red fluorescent protein (RFP), making FRET changes difficult to measure. Therefore, development of a novel GFP-based methodology, utilising cryogenic super-resolution microscopy (as discussed), was preferred over a FRET-based assay. This was despite the limitations discussed below, which are mainly due to recent development, and hence limited user knowledge, of the *superSIL* microscope.

A key error, which was overlooked when developing the GFP-based methodology was the localisation error of the microscope. For this analysis, a localisation error of ~10nm was achieved for GFP. However, the localisation error is still greater than the predicted distance between GFP fluorophores in the parallel conformation (~5nm). It was therefore hypothesised that if the distance between GFP fluorophores in the test condition (GFP-NDP52) was similar to the positive control (GFP-NDP52-GFP), this would infer an anti-parallel NDP52 homodimer.

Table 6 outlines further experimental issues relating to *superSIL* microscopy, as well as resolutions which have either been already employed or could be in the future.

Table 6: Experimental issues with cryogenic super-resolution microscopy

Issue	**Resolution**
Drift of the stage	Use of TetraSpeck™ beads (employed)
Aggregation of GFP-LifeAct (negative control)	Dilution and filtering by centrifugation (employed)
3D orientation of NDP52 through blot/freezing process	Use of surface chemistry, such as biotin-conjugated antibodies
Uncontrollable protein concentrations due to manual blotting	No current resolution as the entire *superSIL* is plunge-frozen and no alternative plunge-freezing protocol has been developed
Aggregation of 5nM GFP-NDP52 under cryogenic conditions, not apparent with TIRF microscopy (**Figure 30A**)	No current resolution

Another major issue was data analysis. Currently, cluster-analysis can be performed but direct distances between GFP fluorophores cannot be measured. Therefore, while you can make comparisons regarding the clustering behaviour, conclusions relating to the dimeric conformation of NDP52 cannot be drawn, especially considering the clustering behaviour of NDP52 appears to change under cryogenic conditions. Neither can conclusions be made about the clustering structure of samples. Nevertheless, results show that the clustering behaviour in the positive and test condition is the same. Since the positive control is tagged at both the N- and C-termini with GFP, while the test condition is only tagged at the N-termini, from the clustering behaviour, it can be suggested that NDP52 adopts an anti-parallel dimeric state at 5nM. However, as discussed, currently only conclusions regarding the clustering behaviour can be made.

There are also issues determining the statistical significance of the results. For instance, there is no confidence level associated with K(r) itself. Also, variance in the results cannot be estimated because the number of datasets is limited, where only 1 dataset was acquired for the test condition due to time, meaning that an error

calculation would not be possible. To overcome this, the frame ranges could be broken down, but the data is time-dependent meaning this is not feasible.

It is important to stress here that the data presented in this study is preliminary and clearly more work needs to be done to optimise this technique. Therefore, in the future, if this technique is to be employed to investigate the dimeric conformation of NDP52, the issues, discussed above, would need to be addressed. This is particularly important given that there are major limitations with alternative methods, such as FRET-based assays. However, this methodology does show promise and is something that should be continually developed going forward. Although, this is not to say that other methods, such as X-ray crystallography, should not also be pursued.

11.9 SAXS suggests NDP52 adopts a parallel homodimeric state *in vitro*

As discussed, cryogenic super-resolution microscopy was limited by providing only comparisons regarding the clustering behaviour of different GFP-tagged NDP52 variants. Therefore, SAXS was employed to provide novel insights not only into the structure of full-length NDP52 (currently unavailable) but also the dimeric arrangement of NDP52 homodimers.

SEC-MALS has shown previously that NDP52 is homodimeric in solution (**Figure 26**). Therefore, peak 2 of SAXS data (**Figure 36A**) likely represents the dimeric form of NDP52. However, peak 1 (**Figure 36A**) is at the limit of the SEC column (void volume) suggesting that this peak is aggregation of NDP52. This is supported by Guinier fitting, where, for peak 1, R_g values are very varied (**Figure 36B**) suggesting heterogeneity. However, R_g values for peak 2 are constant (**Figure 36C**) suggesting homogeneity and that sample preparation is of good quality (can be used for interpretation). Intensity plots further support that peak 1 likely corresponds to aggregates and peak 2 the NDP52 dimer. For instance, peak 1 is considerably nosier at larger values of *q*, compared to peak 2, where average (black) and median (cyan) values largely agree (**Figure 37**); again, suggesting that peak 2 is homogeneous.

From the SAXS envelope, the structure of NDP52 is shown to be asymmetrical (**Figure 38D**), suggesting that NDP52 likely adopts a parallel homodimeric conformation *in*

vitro (**Figure 40C**). For instance, if both N-terminal SKICH domains were aligned in the anti-parallel configuration, NDP52 would be symmetrical (**Figure 40D**). An end-to-end orientation of the dimer (**Figure 40E**) is also dismissed since you would again expect to see symmetry.

Interestingly, one end of NDP52 is shown to be curved (**Figure 38D**) but the other globular. It is likely that these correspond to the C-terminal double-ZF motif (or LIM-L domain) and the N-terminal SKICH domain respectively (**Figure 40A**). This is since the SKICH domain is well-characterised as globular (PDB entry – 3VVV). The potential curvature of the ZF domain may confer tight binding to DNA, by 'wrapping around', to enable NDP52 function as a transcriptional regulator.

SAXS was performed on 3 different gel filtration columns (Shodex KW405-4F, Superdex™ 200 increase 3.2/300 and Superose™ 6 increase 3.2/300). However, only results from the Superose™ 6 column are shown in this study. Data from the Shodex column showed higher heterogeneity of the sample (assessed by R_g values) and poorer resolution of peaks 1 and 2, compared to the Superose™ 6 column. Despite the lower quality of data from the Shodex column, the SAXS envelope generated using this data also (**Figure 40B**) supports a linear/elongated shape for NDP52. Although, one end of the structure (red box; **Figure 40B**) shows some variation, compared to the model generated using the Superose™ 6 column, suggesting that this could be another state of this domain (since SAXS envelopes are an average of multiple bead models) and hence inferring flexibility. This provides important functional insights. For instance, if this flexible region corresponds to the ZF-motif (as discussed), the flexibility may confer the ability of the curved structure to wrap around DNA and enable tight binding for NDP52's role in transcription. Similarly, this flexibility may be important for NDP52-facilitated unfolding of myosin VI[7], as well as autophagosome maturation where NDP52 binds both MVI and ubiquitinated targets via its tandem ZF motif[59,61].

Interestingly, the central domain of the SAXS envelope shows some globular arrangement (**Figure 38D**). This topology may arise from the LZ or the arrangement of the α-helices in the coiled-coil domain of NDP52. Again, this may provide

important structure-function insights into the action of NDP52 as a putative transcription regulator.[7] For instance, this conformation may be essential for docking of other proteins or binding to DNA.

Future work should therefore focus on obtaining higher resolution structures for full-length NDP52 using techniques such as X-ray crystallography (previously tried but unsuccessful[79]) and NMR. Such efforts to improve our structural knowledge of NDP52 are currently ongoing in the Toseland laboratory. For instance, cryo-electron microscopy (EM) data has been collected and is currently being analysed, which will provide a much higher resolution structure of NDP52 compared to SAXS. Higher resolution insights are important as SAXS envelopes only give an outline structure and are hence limited in the information they provide. For instance, SAXS envelopes provide no information regarding the secondary structure and require models, pre-determined by alternative methods such as X-ray crystallography, to be fitted to them. It is therefore important that going forward current crystal and NMR structures (for the SKICH and ZF domains respectively) are fitted to the SAXS envelope to identify which end represents the N-terminus. However, it is acknowledged that it may be difficult to obtain full-length crystal structures for NDP52 (as previously observed[79]) due to the flexibility inferred by SAXS.

The finding that NDP52 is likely a parallel homodimer is particularly important for structure-function observations in relation to transcriptional regulation. For instance, the parallel nature of NDP52 may suggest that NDP52 can achieve tighter binding with DNA, where NDP52 could clamp down on DNA from both sides. This also makes sense due to potential flexibility observed in the NDP52. Whereas, if an anti-parallel dimeric conformation was inferred, this may suggest that NDP52 acts as a linker between DNA and another protein, since residues 120-end have been shown (in this study) to bind DNA with high affinity.

The parallel configuration of NDP52 may also be of functional significance with regard to MVI unfolding[7]. For instance, MVI is a parallel homodimer, with motor domains orientated in the same direction to enable hand-over-hand walking along actin[12]. Therefore, the parallel arrangement of NDP52 presumably enables binding

to MVI to facilitate unfolding, dimerisation and subsequent binding to DNA for transcriptional regulation[7].

Notably, the results from cryogenic super-resolution microscopy (Section 11.8) and SAXS starkly contradict each other – one result suggests an anti-parallel arrangement while the other parallel. However, the results from SAXS are considered more trustworthy, given that *superSIL* microscopy is currently limited as direct measurements cannot yet be made between GFP fluorophores (meaning only statements regarding similarities or differences in clustering behaviour can be made). The anti-parallel arrangement, suggested by *superSIL* microscopy, is therefore likely to come from NDP52 aggregating under cryogenic conditions, which is likely to be an artefact of the sample preparation rather than a true reflection of NDP52 clustering since this behaviour is not observed under non-cryogenic conditions.

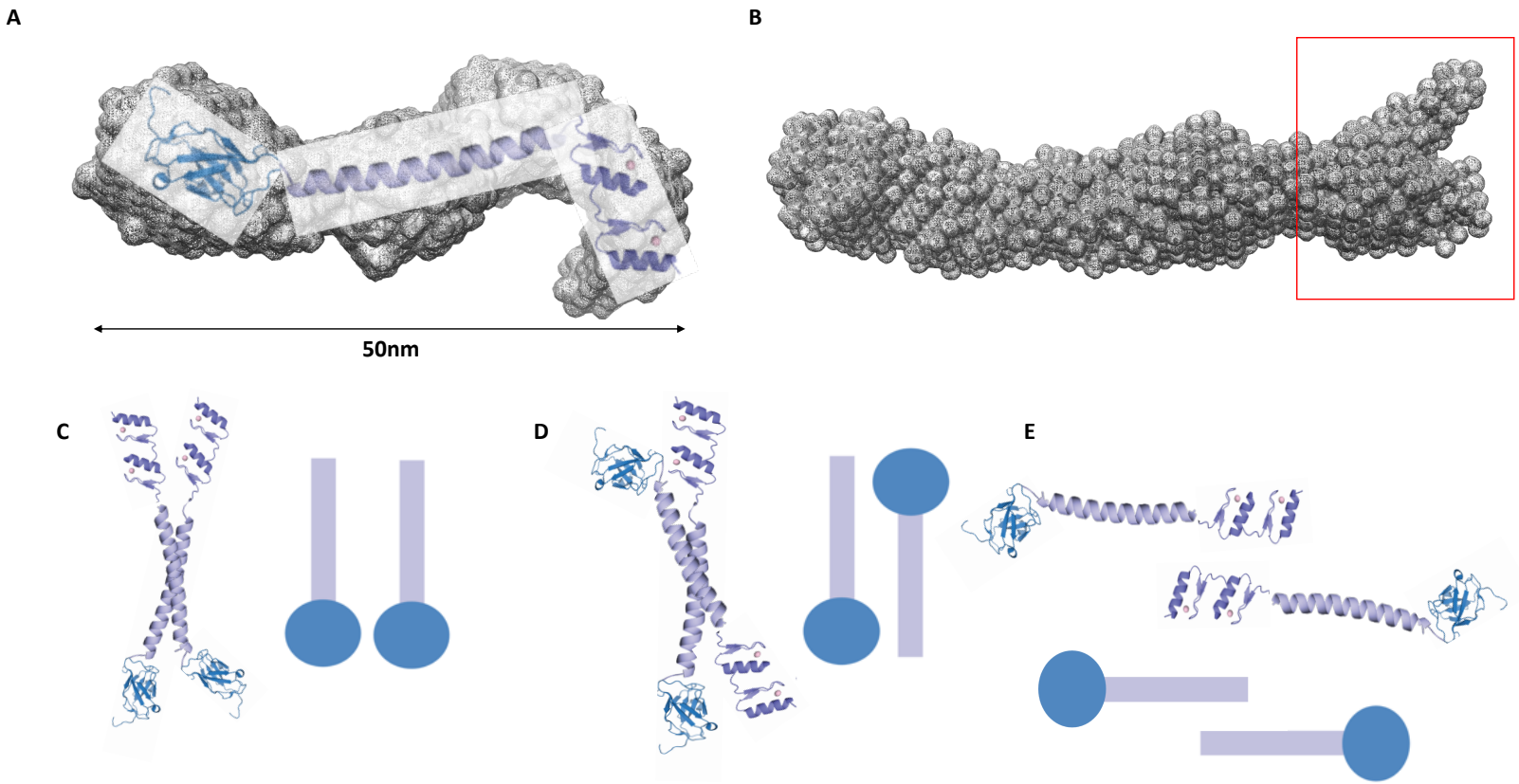


Figure 40: Interpretation of SAXS data

(A) Cartoon representation of NDP52 overlaid onto the SAXS envelope (using the Superose™ 6 increase 3.2/300 column) by eye. **(B)** Resultant SAXS envelope using the Shodex KW405-4F column. Potential flexible domain marked with a red box. NDP52 schematically shown as a parallel **(C)**, anti-parallel **(D)** or end-to-end **(E)** dimer. The N-terminal SKICH domain of NDP52 is shown by a blue circle while the remainder of the structure is shown in stick format.

12 Conclusions

For the first time, novel methodologies are reported for the purification of full-length ERα using the *E.coli* BL21 stain, where the presence of estradiol in the culture medium significantly improves purification results consistent with other studies[77,78]. Although, major problems were identified, including aggregation and degradation. However, this methodology provides a robust foundation for further optimisation of ERα purification, using the *E.coli* BL21 strain as an inexpensive and time-efficient expression system, required for further biochemical and biophysical characterisation of ERα focussing on its interaction with MVI.

Biochemical assays have elucidated the DNA-binding characteristics of NDP52 structural domains. These results importantly contribute to our understanding of NDP52 as a transcriptional co-activator, as the DNA-binding site has been mapped to the coiled-coil domain (inclusive of the LZ), together with the double-ZF motif. This domain is therefore likely to be crucial for NDP52's function as a transcription co-activator. In the future, mutagenesis of key residues within this structural domain can be investigated with regard to dsDNA binding.

Evidence presented here also provides important insights not only into the tertiary structure of NDP52 but also the oligomeric state of the protein. For instance, NDP52 has been shown to be largely linear in structure, with globular N-terminal domain, and to be a homodimer *in vitro.* For the first time, the conformation of this dimer was probed using a recently-developed super-resolution microscope and SAXS. Although *superSIL* requires much optimisation, particularly with regard to data analysis, this is a promising technique which could be used in the future to support findings from SAXS, which suggest that NDP52 functions as a parallel homodimer *in vivo*. This finding provides important structure-function insights into NDP52's role in transcriptional regulation, where a parallel conformation may enable NDP52 to clamp down either side of DNA, enabling tight binding. Future work should be centralised on achieving higher resolution structures for full-length NDP52, through methods such as X-ray crystallography, cryo-EM and NMR.

13 References

1. Berg, J. S., Powell, B. C. & Cheney, R. E. A Millennial Myosin Census. *Mol. Biol. Cell* **12**, 780–794 (2001).

2. Odronitz, F. & Kollmar, M. Drawing the tree of eukaryotic life based on the analysis of 2,269 manually annotated myosins from 328 species. *Genome Biol.* **8**, (2007).

3. Geeves, M. A. & Holmes, K. C. The molecular mechanism of muscle contraction. *Adv. Protein Chem.* **71**, 161–193 (2005).

4. Scholey, J. M., Brust-Mascher, I. & Mogilner, A. Cell division. *Nature* **422**, 746–752 (2003).

5. Buss, F. Myosin VI isoform localized to clathrin-coated vesicles with a role in clathrin-mediated endocytosis. *EMBO J.* **20**, 3676–3684 (2001).

6. Krendel, M. & Mooseker, M. S. Myosins: Tails (and Heads) of Functional Diversity. *Physiology* **20**, 239–251 (2005).

7. Fili, N. *et al.* NDP52 activates nuclear myosin VI to enhance RNA polymerase II transcription. *Nat. Commun.* **8**, 1871 (2017).

8. Rayment, I. *et al.* Three-dimensional structure of myosin subfragment-1: a molecular motor. *Science* **261**, 50–58 (1993).

9. Roberts, R. *et al.* Myosin VI: Cellular functions and motor properties. *Philos. Trans. R. Soc. B Biol. Sci.* **359**, 1931–1944 (2004).

10. Bahloul, A. *et al.* The unique insert in myosin VI is a structural calcium–calmodulin binding site. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 4787–4792 (2004).

11. Uyeda, T. Q., Abramson, P. D. & Spudich, J. A. The neck region of the myosin motor domain acts as a lever arm to generate movement. *Proc. Natl. Acad. Sci.* **93**, 4459–4464 (1996).

12. Yu, C. *et al.* Membrane-induced lever arm expansion allows myosin VI to walk with large and variable step sizes. *J. Biol. Chem.* **287**, 35021–35035 (2012).

13. Rock, R. S. *et al.* Myosin VI is a processive motor with a large step size. *Proc. Natl. Acad. Sci.* **98**, 13655–13659 (2001).

14. Huxley, H. & Hanson, J. Changes in the Cross-Striations of Muscle during Contraction and Stretch and their Structural Interpretation. *Nature* **173**, 973–976 (1954).

15. Huxley, A. F. & Niedergerke, R. Structural Changes in Muscle During Contraction: Interference Microscopy of Living Muscle Fibres. *Nature* **173**, 971–973 (1954).

16. Phichith, D. *et al.* Cargo binding induces dimerization of myosin VI. *Proc. Natl. Acad. Sci.* **106**, 17320–17324 (2009).

17. Yu, C. *et al.* Myosin VI Undergoes Cargo-Mediated Dimerization. *Cell* **138**, 537–548 (2009).

18. Buss, F., Spudich, G. & Kendrick-Jones, J. Myosin VI: cellular functions and motor properties. *Philos. Trans. R. Soc. B Biol. Sci.* **359**, 1931–1944 (2004).

19. Sweeney, H. L. & Houdusse, A. Myosin VI Rewrites the Rules for Myosin Motors. *Cell* **141**, 573–582 (2010).

20. Wells, A. L. *et al.* Myosin VI is an actin-based motor that moves backwards. *Nature* **401**, 505–508 (1999).

21. Ménétrey, J. *et al.* The structure of the myosin VI motor reveals the mechanism of directionality reversal. *Nature* **435**, 779–785 (2005).

22. Sweeney, H. L. & Houdusse, A. What can myosin VI do in cells? *Curr. Opin. Cell Biol.* **19**, 57–66 (2007).

23. Primal de Lanerolle. Nuclear actin and myosins at a glance. *J. Cell Sci.* **125**, 4945–4949 (2012).

24. Tumbarello, D. A., Kendrick-Jones, J. & Buss, F. Myosin VI and its cargo adaptors - linking endocytosis and autophagy. *J. Cell Sci.* **126**, 2561–2570 (2013).

25. Fili, N. *et al.* Binding partner regulation of Myosin VI: Loss of tumour-suppressor Dab2 leads to enhanced activity of nuclear myosin. *bioRxiv* 1–9 (2019). doi:10.1101/639963

26. Osterweil, E., Wells, D. G. & Mooseker, M. S. A role for myosin VI in postsynaptic structure and glutamate receptor endocytosis. *J. Cell Biol.* **168**, 329–338 (2005).

27. Buss, F., Luzio, J. P. & Kendrick-Jones, J. Myosin VI, an Actin Motor for Membrane Traffic and Cell Migration. *Traffic* **3**, 851–858 (2002).

28. Vreugde, S. *et al.* Nuclear Myosin VI Enhances RNA Polymerase II-Dependent Transcription. *Mol. Cell* **23**, 749–755 (2006).

29. Majewski, L. *et al.* Myosin VI in the nucleus of neurosecretory PC12 cells: Stimulation-dependent nuclear translocation and interaction with nuclear proteins. *Nucleus* **9**, 125–141 (2018).

30. Heldring, N. *et al.* Estrogen Receptors: How Do They Signal and What Are Their Targets. *Physiol. Rev.* **87**, 905–931 (2007).

31. Morriswood, B. *et al.* T6BP and NDP52 are myosin VI binding partners with potential roles in cytokine signalling and cell adhesion. *J. Cell Sci.* **120**, 2574–2585 (2007).

32. Hasegawa, J. A Rapid Purification Method for Human RNA Polymerase II by Two-Step Affinity Chromatography. *J. Biochem.* **133**, 133–138 (2003).

33. Sims, R. J., Mandal, S. S. & Reinberg, D. Recent highlights of RNA-polymerase-II-mediated transcription. *Curr. Opin. Cell Biol.* **16**, 263–271 (2004).

34. Lambert, S. A. *et al.* The Human Transcription Factors. *Cell* **172**, 650–665

(2018).

35. Barnes, P. J. How corticosteroids control inflammation: Quintiles Prize Lecture 2005. *Br. J. Pharmacol.* **148**, 245–254 (2006).

36. Rickman, D. S., Schulte, J. H. & Eilers, M. The Expanding World of N-MYC–Driven Tumors. *Cancer Discov.* **8**, 150 LP – 163 (2018).

37. Lodish, H., Berk, A. & Zipursky, S. Section 10.5, Eukaryotic Transcription Activators and Repressors. in *Molecular Cell Biology* (New York: W. H. Freeman, 2000).

38. Brent, R. & Ptashne, M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* **43**, 729–736 (1985).

39. Acconcia, F., Fiocchetti, M. & Marino, M. Xenoestrogen regulation of ERα/ERβ balance in hormone-associated cancers. *Mol. Cell. Endocrinol.* **457**, 3–12 (2017).

40. Kumar, V. & Chambon, P. The estrogen receptor binds tightly to its responsive element as a ligand-induced homodimer. *Cell* **55**, 145–156 (1988).

41. O'Lone, R., Frith, M. C., Karlsson, E. K. & Hansen, U. Genomic Targets of Nuclear Estrogen Receptors. *Mol. Endocrinol.* **18**, 1859–1875 (2004).

42. Kushner, P. J. *et al.* Estrogen receptor pathways to AP-1. *J. Steroid Biochem. Mol. Biol.* **74**, 311–317 (2000).

43. Saville, B. *et al.* Ligand-, Cell- and Estrogen Receptor Subtype (α/β)-dependent Activation at GC-rich (Sp1) Promoter Elements. *J. Biol. Chem.* **275**, 5379–5387 (2000).

44. Kato, S. *et al.* Activation of the estrogen receptor through phosphorylation by mitogen-activated protein kinase. *Science* **270**, 1491–4 (1995).

45. Revankar, C. M., Cimino, D. F., Sklar, L. A., Arterburn, J. B. & Prossnitz, E. R. A transmembrane intracellular estrogen receptor mediates rapid cell signaling.

Science **307**, 1625–30 (2005).

46. Gourdy, P. *et al.* Estrogen receptor subcellular localization and cardiometabolism. *Mol. Metab.* **15**, 56–69 (2018).

47. Vaarala. Myosin VI is a modulator of androgen-dependent gene expression. *Oncol. Rep.* **22**, 991–995 (2009).

48. Dunn, T. A. *et al.* A Novel Role of Myosin VI in Human Prostate Cancer. *Am. J. Pathol.* **169**, 1843–1854 (2006).

49. Yoshida, H. *et al.* Lessons from border cell migration in the Drosophila ovary: A role for myosin VI in dissemination of human ovarian cancer. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 8144–8149 (2004).

50. Korioth, F., Gieffers, C., Maul, G. G. & Frey, J. Molecular characterization of NDP52, a novel protein of the nuclear domain 10, which is redistributed upon virus infection and interferon treatment. *J. Cell Biol.* **130**, 1–13 (1995).

51. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

52. Yang, Y., Wang, G., Huang, X. & Du, Z. Crystallographic and modelling studies suggest that the SKICH domains from different protein families share a common Ig-like fold but harbour substantial structural variations. *J. Biomol. Struct. Dyn.* **33**, 1385–1398 (2015).

53. Fu, T. *et al.* Mechanistic insights into the interactions of NAP1 with the SKICH domains of NDP52 and TAX1BP1. *Proc. Natl. Acad. Sci.* **115**, E11651--E11660 (2018).

54. Gurung, R. *et al.* Identification of a Novel Domain in Two Mammalian Inositol-polyphosphate 5-Phosphatases that Mediates Membrane Ruffle Localization. *J. Biol. Chem.* **278**, 11376–11385 (2003).

55. Velyvis, A. & Qin, J. LIM Domain and Its Binding to Target Proteins. in *Zinc*

Finger Proteins 99–105 (Springer US). doi:10.1007/0-387-27421-9_15

56. Sternsdorf, T., Jensen, K., Züchner, D. & Will, H. Cellular localization, expression, and structure of the nuclear dot protein 52. *J. Cell Biol.* **138**, 435–448 (1997).

57. Verlhac, P. *et al.* Autophagy Receptor NDP52 Regulates Pathogen-Containing Autophagosome Maturation. *Cell Host Microbe* **17**, 515–525 (2015).

58. Glick, D., Barth, S. & Macleod, K. F. Autophagy: cellular and molecular mechanisms. *J. Pathol.* **221**, 3–12 (2010).

59. Kruppa, A. J. *et al.* Myosin VI-Dependent Actin Cages Encapsulate Parkin-Positive Damaged Mitochondria. *Dev. Cell* **44**, 484-499.e6 (2018).

60. Furuya, N. *et al.* NDP52 interacts with mitochondrial RNA poly(A) polymerase to promote mitophagy. *EMBO Rep.* **19**, e46363 (2018).

61. Tumbarello, D. A. *et al.* Autophagy receptors link myosin VI to autophagosomes to mediate Tom1-dependent autophagosome maturation and fusion with the lysosome. *Nat. Cell Biol.* **14**, 1024–35 (2012).

62. Hu, S. *et al.* Structure of Myosin VI/Tom1 complex reveals a cargo recognition mode of Myosin VI for tethering. *Nat. Commun.* **10**, 3459 (2019).

63. Hari Gupta, Y. *et al.* Nuclear Myosin VI Stabilizes RNA Polymerase II in Transcription Factories. *Biophys. J.* **116**, 208a-209a (2019).

64. Wittliff, J. L., Wenz, L. L., Dong, J., Nawaz, Z. & Butt, T. R. Expression and characterization of an active human estrogen receptor as a ubiquitin fusion protein from *Escherichia coli*. *J. Biol. Chem.* **265**, 22016–22 (1990).

65. Verrier, C. S. *et al.* High-Mobility Group (HMG) Protein Associated Factor TAF II 30 Affect Transcriptional Activation. *Mol. Endocrinol.* **11**, 1009–1019 (1997).

66. McDonnell, D. P. *et al.* High level expression of biologically active estrogen receptor in *Saccharomyces cerevisiae*. *J. Steroid Biochem. Mol. Biol.* **39**, 291–

297 (1991).

67. Leatherbarrow, R. J. GraFit Version 5, Erithacus Software Ltd. (2001).

68. Louis-Jeune, C., Andrade-Navarro, M. A. & Perez-Iratxeta, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins Struct. Funct. Bioinforma.* **80**, 374–381 (2012).

69. Perez-Iratxeta, C. & Andrade-Navarro, M. A. K2D2: Estimation of protein secondary structure from circular dichroism spectra. *BMC Struct. Biol.* **8**, 1–5 (2008).

70. Mashanov, G. I. & Molloy, J. E. Automatic Detection of Single Fluorophores in Live Cells. *Biophys. J.* **92**, 2199–2211 (2007).

71. Wang, L. *et al.* Solid immersion microscopy images cells under cryogenic conditions with 12 nm resolution. *Commun. Biol.* **2**, 74 (2019).

72. Amgad, M., Itoh, A. & Tsui, M. M. K. Extending Ripley's K-Function to Quantify Aggregation in 2-D Grayscale Images. *PLoS One* **10**, e0144404 (2015).

73. Baddeley, A. J., Møller, J. & Waagepetersen, R. Non- and semi-parametric estimation of interaction in inhomogeneous point patterns. *Stat. Neerl.* **54**, 329–350 (2000).

74. Kiskowski, M. A., Hancock, J. F. & Kenworthy, A. K. On the Use of Ripley's K-Function and Its Derivatives to Analyze Domain Size. *Biophys. J.* **97**, 1095–1103 (2009).

75. Rambo, R. P. BIOISIS. Available at: http://bioisis.net. (Accessed: 8th August 2019)

76. Cabukusta, B., Kohlen, J. A., Richter, C. P., You, C. & Holthuis, J. C. M. Monitoring Changes in the Oligomeric State of a Candidate Endoplasmic Reticulum (ER) Ceramide Sensor by Single-molecule Photobleaching. *J. Biol. Chem.* **291**, 24735–24746 (2016).

77. Nygaard, F. B. & Harlow, K. W. Heterologous Expression of Soluble, Active Proteins in *Escherichia coli*: The Human Estrogen Receptor Hormone-Binding Domain as Paradigm. *Protein Expr. Purif.* **21**, 500–509 (2001).

78. Eiler, S., Gangloff, M., Duclaud, S., Moras, D. & Ruff, M. Overexpression, Purification, and Crystal Structure of Native ERα LBD. *Protein Expr. Purif.* **22**, 165–173 (2001).

79. Kim, B.-W., Beom Hong, S., Hoe Kim, J., Hoon Kwon, D. & Kyu Song, H. Structural basis for recognition of autophagic receptor NDP52 by the sugar receptor galectin-8. *Nat. Commun.* **4**, 1613 (2013).

14 Supplementary Data

Table 7: Extinction coefficients and molecular weights of recombinant proteins

Expression Vector (Residue Number)	UniProt Entry Name	UniProt Identifier	Protein Name	Molecular Weight (Da)	Extinction Coefficient Assuming all Disulphide Bonds Present ($M^{-1}cm^{-1}$)	Extinction Coefficient Assuming all Disulphide Bonds Absent ($M^{-1}cm^{-1}$)
Human pET151 His-tag* ERα (1-end)	ESR1_HUMAN	P03372	ERα-His	67,039	62,520	61,770
Human pET151 His-tag* NDP52 (1-190)	CACO2_HUMAN	Q13137	NDP52 1-190	23,288	30,160	29,910
Human pET151 His-tag* NDP52 (120-end)			NDP52 120-end	38,931	13,575	12,950
Human pET151 His-tag GFP-NDP52** (1-end)			GFP-NDP52	79,945	65,625	64,750
Human pET151 His-tag* GFP-NDP52-GFP (1-end)			GFP-NDP52-GFP	106,813	87,640	86,640
Human pET151 His-tag* NDP52			NDP52	53,076	43,610	42,860

Extinction coefficient and molecular weight estimates taken from ProtParam using the first isoform listed for each entry on UniProt.

* 'His-tag' refers to 6 N-terminal histidine residues **GFP-tags linked to NDP52 via a 4-glycine linker

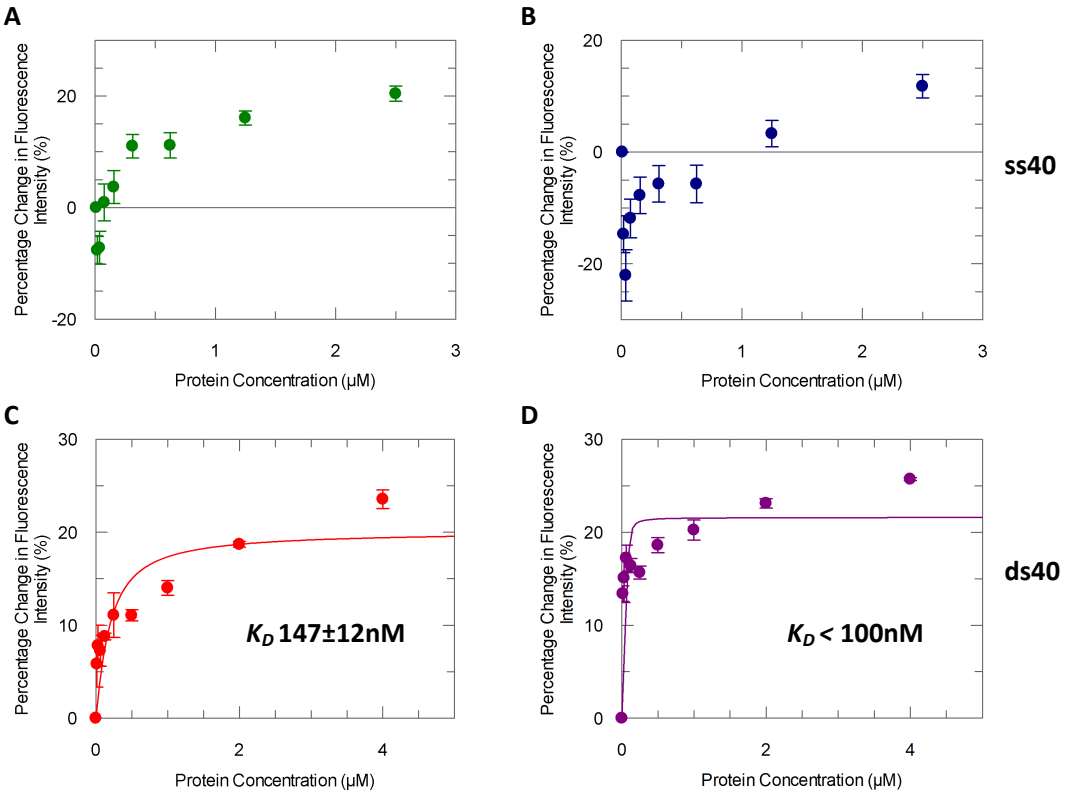


Figure 41: Positive and negative controls for ds40 NDP52 binding assays

5μM NDP52 1-190 **(A)** and NDP52 120-end **(B)** were titrated against 100nM ss40 to act as negative controls. 8μM NDP52 295-end **(C)** and ZF1 **(D)** were also titrated against 100nM fluorescein-labelled ds40 to act as positive controls.

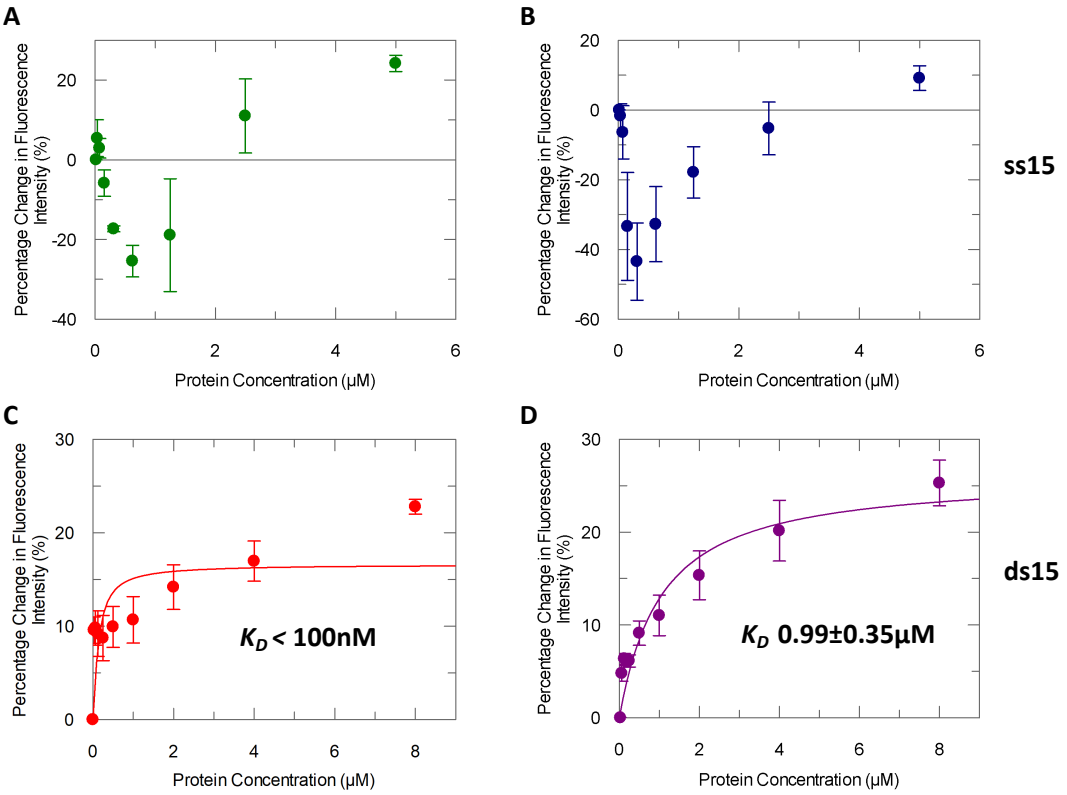


Figure 42: Positive and negative controls for ds15 NDP52 binding assays

10μM NDP52 1-190 **(A)** and NDP52 120-end **(B)** were titrated against 100nM ss15 to act as negative controls. 16μM NDP52 295-end **(C)** and ZF1 **(D)** were also titrated against 100nM fluorescein-labelled ds15 to act as positive controls.

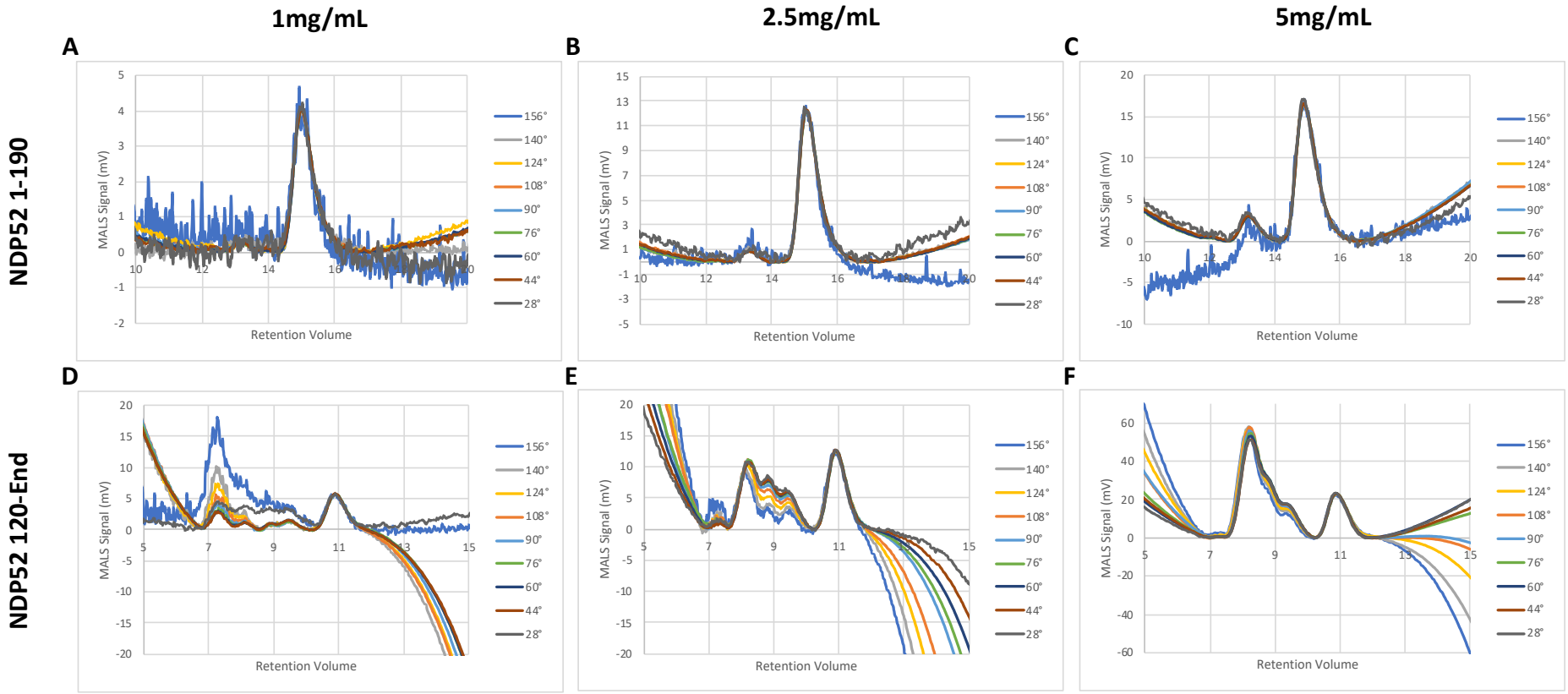


Figure 43: Analysis of SEC-MALS protein samples at different angles

Analysis of NDP52 1-190 **(A-C)** and NDP52 120-end **(D-F)** at respective concentrations. Anisotropic results were excluded from data analysis.