Understanding the Role of RNA Nucleases in Human Disease

Sanja Brolih

Green Templeton College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Trinity 2021

Abstract

Congenital dyserythropoietic anaemia type-I (CDA-I) is a rare inherited disorder of erythropoiesis, in which patient erythroblasts uniquely display a pathognomonic pattern of heterochromatin, referred to as 'spongy' heterochromatin. The molecular basis of CDA-I remains unknown. Most cases of CDA-I are caused by mutations in a protein of unknown function, termed Codanin-1. However, the recent identification of a second disease gene in CDA-I, namely *C15orf41*, has marked a breakthrough in providing new insights into CDA-I. This study describes the previously-uncharacterised protein C15orf41 as a novel RNA exonuclease that associates *in vivo* and *in vitro* with Codanin-1. Using the newly generated CDA-I disease model in an erythroid-specific HUDEP-2 cell background, C15orf41 and Codanin-1 complex formation is determined at physiologically relevant levels. This interaction is then also confirmed *in vitro*. The C15orf41-Codanin-1 complex is a versatile RNA nuclease complex, able to digest a variety of RNA structures. We also show the association and nucleolytic activity of C15orf41-Codanin-1 are disrupted by the introduction of CDA-I patient mutations p. Y94C and p. L178Q in C15orf41. This suggests a direct link between the disease phenotype observed and C15orf41's novel RNAse activity, a potential new target for alternative treatment development.

In light of the 2020 SARS-CoV-2 global pandemic, the focus of this study was broadened to perform essential research on the SARS-CoV-2 virus, in particular the nsp14-nsp10 RNA nuclease complex. This work provides the first complete biochemical activity report for the SARS-CoV-2 nsp14-nsp10 complex, characterising its exonuclease activity as well as a identifying a novel, endonuclease activity. These results have served as a basis for inhibitor development and screening, essential considering the global health impact of this virus.

Understanding the Role of RNA Nucleases in Human Disease



Sanja Brolih

Green Templeton College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2021

Za mamu i tatu.

Abstract

Congenital dyserythropoietic anaemia type-I (CDA-I) is a rare inherited disorder of erythropoiesis, in which patient erythroblasts uniquely display a pathognomonic pattern of heterochromatin, referred to as 'spongy' heterochromatin. The molecular basis of CDA-I remains unknown. Most cases of CDA-I are caused by mutations in a protein of unknown function, termed Codanin-1. However, the recent identification of a second disease gene in CDA-I, namely *C15orf41*, has marked a breakthrough in providing new insights into CDA-I. This study describes the previously-uncharacterised protein C15orf41 as a novel RNA exonuclease that associates *in vivo* and *in vitro* with Codanin-1. Using the newly generated CDA-I disease model in an erythroid-specific HUDEP-2 cell background, C15orf41 and Codanin-1 complex formation is determined at physiologically relevant levels. This interaction is then also confirmed *in vitro*. The C15orf41-Codanin-1 complex is a versatile RNA nuclease complex, able to digest a variety of RNA structures. We also show the association and nucleolytic activity of C15orf41-Codanin-1 are disrupted by the introduction of CDA-I patient mutations p. Y94C and p. L178Q in C15orf41. This suggests a direct link between the disease phenotype observed and C15orf41's novel RNAse activity, a potential new target for alternative treatment development.

In light of the 2020 SARS-CoV-2 global pandemic, the focus of this study was broadened to perform essential research on the SARS-CoV-2 virus, in particular the nsp14-nsp10 RNA nuclease complex. This work provides the first complete biochemical activity report for the SARS-CoV-2 nsp14-nsp10 complex, characterising its exonuclease activity as well as a identifying a novel, endonuclease activity. These results have served as a basis for inhibitor development and screening, essential considering the global health impact of this virus.

Acknowledgements

First, thank you to my supervisors, Prof. Peter McHugh and Dr Christian Babbs, your knowledge, support, and guidance has been invaluable. Thank you both for your scientific rigour, enthusiasm, and encouragement. A special thank you to Peter for agreeing to keep me in his lab and for all the support through my scientific journey from MRes student to post-doc.

I am extremely grateful to all members in the McHugh lab. Thanks must go to the following: Hannah B, for being the world's best sci-wife and providing me with endless care and kindness, market trips and tea breaks; Lonnie, for your support and teaching me to be better; Will, for being my fellow DPhil; Malitha, for being the best desk neighbour and lab pal; Abi, for all the purification fun, Jenni and Ceren for being my fellow post-doc social secs as well as all other members, present and past.

Thank you also to the Babbs lab, particularly Aude and Caroline, for all your scientific input, support and the opportunity to collaborate on this new and interesting project as well as the Gileadi lab for all your help with protein purification, particularly to Lia, for being the best science-guru and third member of the sci-wife club. Thank you also to the Blackford, in particular Matthew, for being my lunch-buddy for the last 4 years, Clynes and Fairfax groups for being excellent labbies and making the Oncology floor a fun place to work in; being your social-sec has been exciting. I am also very grateful to the MRC for funding my studies over the last four years and for providing me with this opportunity.

Thanks must also go to my OUWBC, rowing, GTC and University friends for their support outside the lab: to Vira, for making me a better person to this day; Camille and Ghalia, for being my favourite GTC members; Phil, for making sure I am alive just enough times and being the most patient person I know; Sander, for all the fun summer adventures; Ruth, for all the wholesome fun; Zoulikha and Baby Karl, for being my adoptive family throughout lockdown and my favourite travel buddies and band members; Carl and Sean, for the endless hospitality and kindness and to Lise, for you have been here from the start to the end, and all the turbulence in between.

Final thanks go to my family back home– your endless support throughout all this time means the world to me.

Declaration

This thesis is entirely my own work except where otherwise indicated.

Peer-reviewed publications

1. *CDIN-1 is a novel RNA nuclease involved in CDA-I.* **Brolih S***, Olijnik A-A*, Yosaatmadja Y, Scott C, Gavard A, Johnson E, Newman J, Downes D J, Brown J, Gileadi O, Buckle V J, McHugh P J and Babbs C. 2021. [Manuscript in preparation.]
   * co-first authorship

2. *New Perspectives on the Clinical Use of Interferons in CDA-I Treatment.* Frey K*, **Brolih S***, Da Costa L, Roy N & Babbs C. 2021. [Manuscript in preparation]
   * co-first authorship

3. *Characterisation of the SARD-CoV-2 ExoN (nsp14ExoN-nsp10) complex: implications for its role in viral genome stability and inhibitor identification.* Baddock T*, **Brolih S***, Yosaatmadja Y*, Ratnaweera M, Bielinski M, Swift L P, Cruz-Migoni A, Morris G M, Schofield C J, Gileadi O, McHugh P J. 2021. [Manuscript in review at NAR]; doi.org/10.1101/2020.08.13.248211
   * co-first authorship

4. *Genetic and Functional Insights into CDA-I Prevalence and Pathogenesis.* Olijnik A-A, Roy N, Scott C, Marsh J A, Lauschke K, Ask K, Roberts N, Downes D J, Brown J, **Brolih S**, Johnson E, Xella B, Proven M, Hipkiss R, Ryan K, Frisk P, Mäkk J, Stattin E, Sadasivam N, McIlwain L, Hill Q A, Renella R, Hughes J R, Gibbons R, Groth A, McHugh P J, Higgs D R, Buckle V J & Babbs C. 2020. Journal of Medical Genetics, Available at: doi.org/10.1136/jmedgenet-2020-106880

5. *Optimised oligonucleotide substrates to assay XPF-ERCC1 nuclease activity for the discovery of DNA repair inhibitors.* Thomas AM*, **Brolih S***, McGouran JF, El-Sagheer AH, Ptchelkine D, Jones M, McDonald NQ, McHugh PJ & Brown T. 2019. TL - 55. Chem. Commun. (Camb). 55 VN-readcube.com: 11671–11674, Available at: doi.org/10.1039/C9CC05476F
   * co-first authorship

6. *RPA activates the XPF-ERCC1 endonuclease to initiate processing of DNA interstrand crosslinks.* Abdullah UB, McGouran JF, **Brolih S**, Ptchelkine D, El-Sagheer AH, Brown T & McHugh PJ. 2017. EMBO J. Available at: doi.org/10.15252/embj.201796664

Contents

List of Abbreviations **x**

List of Figures **xiv**

List of Tables **xviii**

1 Introduction I **1**

1.1 General Introduction to Anaemia 1
1.1.1 Anaemia: General overview 1
1.1.2 Haemoglobin (Hb) 1
1.1.2.1 Structure and Function 1
1.1.2.2 Haemoglobin and Disease 3
1.1.3 Erythropoiesis 3
1.1.3.1 Erythropoietin (Epo) 6
1.2 Congenital Dyserythropoietic Anaemia 7
1.2.1 General Overview 7
1.2.2 Sub-types of CDA 7
1.2.3 CDA-I 8
1.2.3.1 Disease Phenotype 8
1.2.3.2 Founder mutation 10
1.2.3.3 Clinical Diagnosis 12
1.2.3.4 Current treatments 13
1.2.4 CDA-II 14
1.2.4.1 Disease Phenotype 14
1.2.4.2 Diagnosis 15
1.2.5 CDA-III 16
1.3 Disease Genes in CDA-I 18
1.3.1 Codanin-1 (CDAN1) 18
1.3.1.1 Biological Characterisation 18
1.3.1.2 3D Structure and function prediction 19
1.3.1.2.1 Drosophila Homologue 21
1.3.1.2.2 Codanin-1 3D Structure predictions 21
1.3.1.3 Asf1 Interaction 23
1.3.2 C15orf41 25
1.3.2.1 Discovery 25
1.3.2.2 Structure predictions 27
1.3.2.3 The PD-(D/E) XK Superfamily of Nucleases . . . 29
1.3.3 C15orf41-Codanin-1 Complex 31
1.4 Summary & Aims 34

2 Introduction II **36**
2.1 General Introduction . . . 36
2.1.1 Coronavirus disease 19 . . . 36
2.1.2 The biology of Coronaviruses . . . 38
2.1.3 SARS-CoV-1 and SARS-CoV-2 lifecycle . . . 39
2.1.4 Non-structural proteins . . . 40
2.1.4.1 The replication-transcription complex or replisome 40
2.1.4.2 nsp14(ExoN)-nsp10 . . . 43
2.2 Aims . . . 46

3 Materials and Methods **47**
3.1 Protein Purification methods . . . 47
3.1.1 C15orf41-Codanin-1 purifications . . . 47
3.1.2 nsp14-nsp10 purification . . . 50
3.2 Biochemistry methods . . . 52
3.2.1 Generation of 5′ radiolabelled substrates . . . 52
3.2.2 Substrate Annealing . . . 52
3.2.3 Generation of a single nucleotide RNA ladder . . . 52
3.2.4 Nuclease assays . . . 54
3.2.4.1 C15orf41-Codanin-1 nuclease assays . . . 54
3.2.4.2 Nsp14-nsp10 nuclease assays . . . 56
3.2.5 Electrophoretic Mobility Shift Assay (EMSA) . . . 57
3.2.6 SEC-MALS . . . 57
3.2.7 SAXS . . . 58
3.2.8 Negative-staining electron microscopy (neg-staining EM) . . 58
3.3 Cell Biology methods . . . 59
3.3.1 Tissue Culture . . . 59
3.3.2 Transfections . . . 60
3.3.3 Immunoprecipitations (IP) . . . 61
3.3.4 Western Blotting, protein transfers and ECL detection . . . 61

4 Cellular characterisation of C15orf41-Codanin-1 **64**
4.1 C15orf41 and Codanin-1, the two disease proteins in CDA-I, form a stable complex . . . 64
4.1.1 Complex formation in HEK293-FT cells with over-expressed proteins . . . 65
4.1.2 Complex formation in HUDEP-2 cells at endogenous levels . 70
4.1.3 Patient mutations in C15orf41 disrupt C15orf41-Codanin-1 complex formation in HUDEP-2 cells . . . 77

4.2 Patient mutations in C15orf41 and Codanin-1 affect protein levels of C15orf41 and Codanin-1, albeit differentially 79
4.2.1 Antibody Validation 79
4.2.2 Analysis of C15orf41 and Codanin-1 protein levels in HUDEP-2 cells 80
4.3 Conclusion 82

5 Purification and Biophysical Characterisation of the C15orf41-Codanin-1 Complex 85
5.1 Protein Purification: purification of wild-type C15orf41-Codanin-1 complex and mutant versions 85
5.2 Biophysical characterisation of C15orf41-Codanin-1 91
5.2.1 SEC-MALS 91
5.2.2 SAXS 93
5.3 Attempts at structural studies of C15orf41-Codanin-1 97
5.3.1 X-Ray crystallography 97
5.3.1.1 Crystallography screen set-up 97
5.3.1.2 Expression in LOBSTR E. coli strain 98
5.3.2 Negative-staining EM 99
5.4 Conclusion 102

6 Biochemical characterisation of C15orf41-Codanin-1 103
6.1 C15orf41 is a novel RNA nuclease 103
6.1.1 C15orf41 and Codanin-1 form a functional RNA nuclease complex with enhanced activity 104
6.2 C15orf41-Codanin-1 is a versatile RNA nuclease 109
6.3 C15orf41-Codanin-1 is able to process a variety of complex RNA-containing structures 116
6.4 CDA-I patient mutations in C15orf41 reduce C15orf41 nuclease activity 120
6.5 Conclusion 123

7 Biochemical characterisation of nsp14-nsp10 126
7.1 SARS-CoV-2 nsp14-nsp10 is an RNA processing nuclease 126
7.2 SARS-CoV-2 nsp14-nsp10 is able to process dsRNA and dsRNA substrates containing mismatches 132
7.2.1 Nsp14-nsp10 exhibits a ribonuclease activity that reaches beyond its classical role as a proof-reading nuclease 135
7.3 SARS-CoV-2 nsp14-nsp10 displays both 3′-exonuclease and a newly-described endonucleolytic activity 137

7.4 The SARS-CoV-2 polymerase nsp12-7-8 complex enhances nsp14-nsp10 nuclease activity on a variety of RNA substrates 141
7.5 Conclusion 145

8 Discussion and Future Perspectives for the C15orf41-Codanin-1 study 148
8.1 Discussion 148
8.2 Conclusion 159

9 Discussion and Future Perspectives for the nsp14-nsp10 study 161
9.1 Discussion 161
9.2 Conclusion 164

Appendices

A Vector Maps 166

List of Abbreviations

2,3-DPG . . . 2,3-diphosphoglycerate

2019-nCoV . . 2019 novel coronavirus

aa . . . . . amino acid

ACE2 . . . . . angiotensin-converting enzyme 2

Asf1 . . . . . anti-silencing function 1

ATP . . . . . adenosine triphosphate

BFU-E . . . . burst-forming unit erythroid

BSA . . . . . bovine serum albumin

$CaCl_2$ . . . . . calcium chloride

CAF1 . . . . . chromatin assembly factor 1

Cas9 . . . . . CRISPR-associated protein 9

CDA . . . . . Congenital Dyserythropoietic Anaemia

CDA-I . . . . . Congenital Dyserythropoietic Anaemia Type I

CDA-II . . . . Congenital Dyserythropoietic Anaemia Type II

CDA-III . . . Congenital Dyserythropoietic Anaemia Type III

CFU-E . . . . colony-forming unit erythroid

CoV . . . . . coronavirus

COVID-19 . . coronavirus disease 19

CRISPR . . . clustered regularly interspaced short palindromic repeats

D-loop . . . . . displacement loop

dATP . . . . . deoxyadenosine triphosphate

DBA . . . . . Diamond Blackfan Anaemia

dL . . . . . decilitre

dlt . . . . . discs lost

DNA . . . . . deoxyribonucleic acid

dNTP . . . . . deoxynucleotide triphosphate

dsDNA . . . . double-stranded DNA

dsRNA . . . . double-stranded RNA

DTT . . . . . dithriothreitol

E.coli	Escherichia coli
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
EM	electron microscopy
EMSA	electrophoretic mobility shift assay
Epo	erythropoietin
Epo-R	erythropoietin receptor
ESI-Q-TOF	electrospray-ionisation quadruple time-of-flight
FCS	fetal calf serum
fL/cell	femtolitre per cell
g	gramme
GFP	green fluorescent protein
h	hour
Hb	haemoglobin
HEK	human embryonic kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His_6	hexa-histidine
HSC	haematopoietic stem cell
HUDEP-2	human umbilical cord blood-derived erythroid progenitor
IFN-alpha	interferon alpha
IMAC	immobilised metal affinity chromatography column
IP	immuno-precipitation
IPTG	beta-D-1-thiogalactoyranoside
JAK2	Janus kinase 2
L	litre
LC-MS/MS	liquid chromatography tandem mass spectrometry
M	molar
MAPK	mitogen-activated protein kinase
MCV	mean corpuscular volume
MERS	Middle East respiratory syndrome
$MgCl_2$	magnesium chloride

Min . . . . . minute

$MnCl_2$ . . . . . manganese chloride

mRNA . . . . messenger RNA

MS . . . . . mass spectrometry

NaCl . . . . . sodium chloride

NiRAN . . . . nidovirus RdRp-associated nucleotidyltransferase

Nsp . . . . . non-structural proteins

ORF . . . . . open reading frame

PBS . . . . . phosphate-buffered saline

Pi3K . . . . . phosphoinositide 3-kinase

PNK . . . . . polynucleotide kinase

RdRp . . . . . RNA-dependent RNA polymerase

RNA . . . . . ribonucleic acid

RT-qPCR . . . reverse transcription quantitative PCR

rTEV . . . . . recombinant tobacco etch virus

SAM . . . . . s-adenosyl methionine

SARS . . . . . severe acute respiratory syndrome

SARS-CoV-2 . severe acute respiratory syndrome coronavirus 2

SAXS . . . . . small angle X-ray scattering

SDS-PAGE . . sodium dodecyl sulphate polyacrylamide gel electrophoresis

Sec . . . . . second

SEC-MALS . size-exclusion chromatography multiple angle light scatter

ssDNA . . . . single-stranded DNA

ssRNA . . . . single-stranded RNA

STAT5 . . . . signal transducer and activator of transcription 5

STREP . . . . streptavidin

TBE . . . . . tris-borate EDTA

TCEP . . . . . tris(2-carboxyethyl)phosphine

TEM . . . . . transmission electron microscopy

UV . . . . . ultraviolet

v/v . . . . . volume per volume

w/v . . . . . weight per volume

wHtH . . . . . winged helix-turn-helix

WT . . . . . wild-type

ZB . . . . . Z-basic

$ZnCl_2$ . . . . . zinc chloride

List of Figures

1.1 Schematic of Haemoglobin 2
1.2 Overview of Erythropoiesis 5
1.3 Clinical features of Congenital Dyserythropoiesis Type-I 9
1.4 Cellular features of Congenital Dyserythropoiesis Type-I 10
1.5 Clinical profile of a CDA-I patient with the founder mutation R1042W in *CDAN1* 11
1.6 Clinical features of Congenital Dyserythropoiesis Type-II 15
1.7 Clinical features of Congenital Dyserythropoiesis Type-III 16
1.8 Summative diagram of CDA 17
1.9 Schematic of the predicted structural domains of Codanin-1 19
1.10 Representation of all known pathogenic variants associated with *CDAN1* 20
1.11 Schematic of predicted Codanin-1 domains based on CNOT 3D structure similarity 22
1.12 Proposed model for Codanin-1 and Asf1 interaction 24
1.13 Clinical profile of CDA-I patients harbouring Y94C and L178Q substitutions in C15orf41 26
1.14 Predicted model of the tertiary structure of the Helix-turn-Helix domain in C15orf41 (I) and the nuclease domain in C15orf41 (II) . 27
1.15 Representation of all known pathogenic variants in *C15orf41* 29
1.16 The PD-(D/E) XK nuclease fold 30
1.17 Schematic of the newly-identified C15orf41-Codanin-1 complex . . . 33

2.1 The SARS-CoV-2 pandemic mapped 37
2.2 The SARS-CoV virion 39
2.3 The Coronavirus lifecycle 41
2.4 Schematic of the SARS-CoV replisome 42
2.5 Schematic of the SARS-CoV-1 nsp14-nsp10 complex 45

4.1 Transfected FLAG-GFP-C15orf41 plasmid in HEK293-FT cells . . . 66
4.2 Transfected FLAG-GFP-C15orf41 is not detectable by western blotting for GFP 68
4.3 FLAG-GFP-C15orf41 only is detectable by western blotting for FLAG 69
4.4 Co-immunoprecipitations of FLAG-GFP-C15orf41 with Codanin-1 in HEK293-FT cells 71
4.5 FLAG-GFP-C15orf41 and Codanin-1 form a tight complex in HEK293-FT cells 72

4.6 CDA-I cellular defects are recapitulated in culture HUDEP-2 cell model 74
4.7 Co-immunoprecipitations of 3xFLAG-C15orf41 with Codanin-1 in HUDEP-2 cells 76
4.8 The L178Q patient mutation in C15orf41 disrupts complex formation with Codanin-1 in HUDEP-2 cells 78
4.9 Testing anti-Codanin-1 and anti-C15orf41 primary antibodies 80
4.10 The L178Q patient mutation in C15orf41 affects both levels of C15orf41 and Codanin-1 whereas patient mutation R1042W in Codanin-1 only affects C15orf41 levels in HUDEP-2 cells 82

5.1 Purification of full-length C15orf41 and the C-terminal fragment of Codanin-1 87
5.2 C15orf41-Codanin-1 complex formation in the nickel-rebind step . . 88
5.3 Wild-Type C15orf41-Codanin-1 complex purification 89
5.4 Predicted 'nuclease-inactive' C15orf41$^{D196A\text{-}E214A}$-Codanin-1 complex purification 90
5.5 Patient Mutant C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 complexes purification 92
5.6 SEC-MALS studies of C15orf41-Codanin-1 94
5.7 SAXS model of wild-type C15orf41 and Codanin-1 95
5.8 SAXS model of C15orf41-Codanin-1 complex 96
5.9 Optimisation of C15orf41-Codanin-1 complex purification 100
5.10 Negative staining EM studies of C15orf41-Codanin-1 101

6.1 C15orf41 is a novel RNA nuclease 104
6.2 C15orf41-Codanin-1 form a RNA nuclease complex with enhanced activity 106
6.3 Mutant C15orf41^{D196A,E214A}- Codanin-1 does not have any nuclease activity 107
6.4 Observed nuclease activity is intrinsic to C15orf41 and not Codanin-1 108
6.5 C15orf41-Codanin-1 activity coincides with the gel filtration elution peak 109
6.6 C15orf41-Codanin-1 nuclease activity is reproducible 110
6.7 C15orf41-Codanin-1 has an ion-dependent RNA nuclease activity . 111
6.8 C15orf41-Codanin-1 shows no nuclease activity on various DNA substrates 112
6.9 C15orf41-Codanin-1 is able to bind both RNA and DNA substrates 113
6.10 C15orf41-Codanin-1 shows preferential activity on 20-mer RNA substrates 114

6.11 C15orf41-Codanin-1 displays a distinct digestion pattern 115
6.12 Figure 12: C15orf41-Codanin-1 is an RNA exonuclease 116
6.13 C15orf41-Codanin-1 is not able to process substrates with embedded ribonucleotides 117
6.14 C15orf41-Codanin-1 is not able to process substrates mimicking 'Okazaki-like' structures 118
6.15 C15orf41-Codanin-1 is able to process R-loop but not D-loop structures 119
6.16 C15orf41-Codanin-1 is able to process some common chemical modification of RNA 121
6.17 Patient mutations Y94C and L178Q in C15orf41 disrupt its nuclease activity 122
6.18 Patient mutant complexes C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 show partially restored activity in the presence of Codanin-1 124

7.1 Purity of Sars-CoV-2 nsp14-nsp10 proteins used in this study 128
7.2 SARS-CoV-2 nsp10 confers additional activity to nsp14 129
7.3 Nsp14-nsp10 activity coincides with the gel filtration elution peak . 130
7.4 Nsp14-nsp10 is an ion-dependent RNA nuclease 131
7.5 Nsp14-nsp10 is an RNA nuclease with a preference for 20-mer substrates 132
7.6 Nsp14-nsp10 is an RNA nuclease with a complex digestion pattern . 133
7.7 Sars-CoV-2 nsp14-nsp10 is able to digest substrates containing mismatches 134
7.8 Nsp14-nsp10 activity is not sequence specific on a mixed-sequence RNA substrate 136
7.9 Nsp14-nsp10 shows altered activity profiles on repetitive sequences . 137
7.10 Nsp14-nsp10 is an RNA nuclease able to process a variety of RNA-containing substrates 138
7.11 Nsp14-nsp10 activity is specific to RNA only 139
7.12 Nsp14-nsp10 activity is able to process common chemical modifications of RNA 140
7.13 Nsp14-nsp10 exhibits both 3′-exonuclease activity and a newly-described endonucleolytic activity 140
7.14 Purity of the SARS-CoV-2 nsp12-7-8 polymerase complex 141
7.15 The proteins of the SARS-CoV-2 nsp12-7-8 polymerase complex have no intrinsic nuclease activity 143
7.16 The SARS-CoV-2 polymerase nsp12-nsp7-nsp8 complex enhances nsp14-nsp10 nuclease activity on a variety of RNA substrates 144
7.17 The SARS-CoV-2 polymerase nsp12-nsp7-nsp8 complex enhances nsp14-nsp10 nuclease activity particularly on uracil-containing substrates 145

7.18 The SARS-CoV-2 primase nsp8 is the key factor in nsp14-nsp10 activity stimulation 146
7.19 The SARS-CoV-2 primase nsp8 enhances the activity of nsp14-nsp10 147

A.1 Maps of purification vectors 167
A.2 Map of purification vector used for nsp14-nsp10 purification 168

List of Tables

3.1 List of RNA and DNA substrates used for C15orf41-Codanin-1 nuclease assays 53
3.2 Codes to generate complex RNA and DNA structures used for C15orf41-Codanin-1 nuclease assays 54
3.3 List of RNA and DNA substrates used for nsp14-nsp10 nuclease assays 55
3.4 Codes to generate complex RNA and DNA structures used for nsp14-nsp10 nuclease assays 55
3.5 List of antibodies used during western blotting experiments 63

1

Introduction I: Characterisation of the novel RNA nuclease complex C15orf41-CDAN1

1.1 General Introduction to Anaemia

1.1.1 Anaemia: General overview

Anaemia is a widespread medical condition, affecting approximately 24.8% of the global population as defined by the World Health Organisation (de Benoist et al. 2008). Anaemia is a disease that affects almost every organ and tissue of the human body causing a variety of different symptoms such as fatigue, loss of energy, dizziness and general weakness (Ludwig and Strasser 2001). Clinically, a low level of haemoglobin (Hb) when compared to reference values for the patient's sex and age is indicative of this condition (Jones et al. 2008). The degree of anaemia may be scaled as mild, moderate and severe by Hb level (Ludwig and Strasser 2001). Following levels of Hb apply to each category: Hb 10.0 g/dL for mild anaemia, Hb 8.0 to 10.0 g/dL for moderate anaemia and Hb 6.5 to 7.9 g/dL for severe anaemia (Ludwig and Strasser 2001). A more detailed description of haemoglobin will be provided in the next section.

1.1.2 Haemoglobin (Hb)

1.1.2.1 Structure and Function

In humans, each erythrocyte contains around 300 million molecules of haemoglobin (Hb) (Thomas and Lumb 2012). The primary function of Hb in erythrocytes is

transport of oxygen from the lungs to other tissues. This is achieved through Hb's unique structure, consisting of a tetramer of two α- and two β- chains with a haem molecule embedded in each globin chain (Figure 1.1) (Thomas and Lumb 2012). Each α chain has 141 amino acids and each β chain has 146 amino acids. It is the iron molecule of the haem that carries the oxygen and confers the molecule its characteristic red colour (Marengo-Rowe 2006; Ahmed, Ghatge, and Safo 2020).

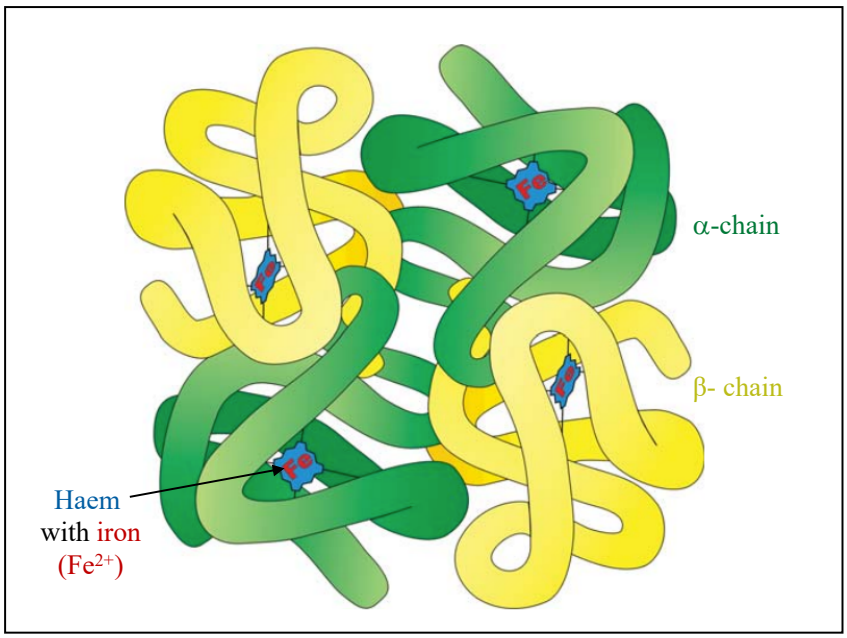


Figure 1.1: Schematic of Haemoglobin (Hb)
Diagram showing the basic quaternary structure of Haemoglobin. Haemoglobin is comprised of two α- and two β- chains (in green and yellow respectively) with a haem molecule embedded in each globin chain. The iron molecule of the haem (in blue) that carries the oxygen (as shown by Fe in red). *Adapted from Thomas and Lumb, 2012.*

Each globin chain of Hb contains a haem molecule that reversibly binds to oxygen in a cooperative fashion: the binding of oxygen to one haem group increases the oxygen affinity within the remaining haem groups through quaternary conformational changes. This molecular cooperativity within Hb is responsible for the sigmoidal-shaped oxygen dissociation curve and is influenced by pH, carbon dioxide and 2,3-diphosphoglycerate (2,3-DPG) (Thomas and Lumb 2012; Marengo-Rowe 2006; Ahmed, Ghatge, and Safo 2020). 2,3-DPG is a highly anionic organic phosphate that promotes the release of oxygen from Hb. Its deregulation promotes anaemia. In fact, 2,3-DPG is increased in anaemia, causing 2,3-DPG to bind

between the β-globin chains of Hb thereby reducing oxygen affinity of the molecule through protein structure alterations (Thomas and Lumb 2012).

1.1.2.2 Haemoglobin and Disease

Haemoglobin structure and function are highly regulated processes. Any genetic defects in Hb therefore have impacting health consequences, leading to a variety of inherited erythroid anaemias. Genetic defects in Hb are the most common of all genetic disorders, resulting either in impaired production of globin chains or abnormal globin chain structures which cause Thalassaemia(s) or Haemoglobinopathies respectively (Thomas and Lumb 2012). α- and β- Thalassaemia are disorders characterized by transcriptional defects of either α- and β- globin. Whereas a healthy individual has equal quantities of α- and β-globin chains produced, α- and β-Thalassaemia patients have an excess of either one of the chains causing haemolysis and anaemia (Muncie and Campbell 2009).

The best-known example of a Haemoglobinopathy is sickle-cell disease. In sickle-cell disease, a point mutation in the amino acid sequence of the β-globin gene leads to a substitution of a glutamic acid residue to a valine at the sixth position of the β-globin chain resulting in the formation of abnormal Hb such as Hb S. This impairs normal functioning of erythrocytes and distorts the cells into a sickle or crescent shape, characteristic of the disease (Thomas and Lumb 2012; Mandal et al. 2020). Interestingly, individuals heterozygous for this mutation display a much milder condition termed sickle cell trait that has been shown to confer protection against certain forms of malaria, a possible explanation as to why this allele is present at a high frequency in human populations, particularly in Central Africa (Elguero et al. 2015).

1.1.3 Erythropoiesis

As described above, the most important function of erythrocytes is transport and delivery of oxygen to peripheral tissues. This requires a delicate balance between constant replenishment of erythrocytes and degradation of old cells achieved through

a tightly-regulated and complex process termed erythropoiesis (Ludwig and Strasser 2001; Zivot et al. 2018). Erythropoiesis is initiated in the bone marrow to produce 2 million mature, enucleated erythrocytes per second in healthy adults in a multi-step process originating from multipotent haematopoietic stem cells (HSCs) (Hattangadi et al. 2011; Moras et al. 2017). HSCs also have the ability to produce a variety of other functionally distinct lineages of differentiated blood cells such as lymphocytes, granulocytes, megakaryocytes or macrophages (Dzierzak and Philipsen 2013).

To produce erythrocytes, HSCs differentiate into more committed erythroid progenitors termed common myeloid progenitors and finally into the burst-forming unit erythroid (BFU-E) that commits the progenitor cell to the erythroid lineage. The BFU-E then further differentiates into the rapidly-dividing colony-forming unit erythroid (CFU-E) following terminal differentiation (Hattangadi et al. 2011; Zivot et al. 2018). CFU-Es undergo a series of drastic changes such as decrease in cell size, chromatin condensation and gradual accumulation of haemoglobin as part of their maturation process where they evolve from proerythroblasts to more mature cells termed basophilic and then from polychromatophilic and orthochromatic erythroblasts (Doulatov et al. 2012; Mei, Liu, and Ji 2021). This process occurs in the bone marrow within erythroblastic islands, composed of a central macrophage surrounded by erythroblasts. Macrophages play an essential role in erythropoiesis as they facilitate proliferation and erythroid differentiation and provide necessary co-factors such as iron to the erythroblasts (Moras et al. 2017).

Orthochromatic erythroblasts then undergo enucleation and expulsion of organelles such as the endoplasmic reticulum, the mitochondria, Golgi apparatus and ribosomes to form reticulocytes, also referred to as immature red blood cells. The extruded nucleus is referred to as the pyrenocyte and is rapidly eliminated by the macrophages present in the erythroblastic island (Moras et al. 2017; Zivot et al. 2018).

The final phase of erythroid development involves the maturation of reticulocytes into erythrocytes or mature red blood cells. The final maturation of reticulocytes takes place in the bone marrow for 3 days where the cells lose the remainder of

their organelles and undergo extensive membrane remodelling, losing 30% of the cell surface before eventually circulating into the blood stream for 1 day until they fully mature into the characteristically-shaped biconcave erythrocytes (Moras et al. 2017; Zivot et al. 2018). A summary of erythropoiesis is shown below in Figure 1.2.

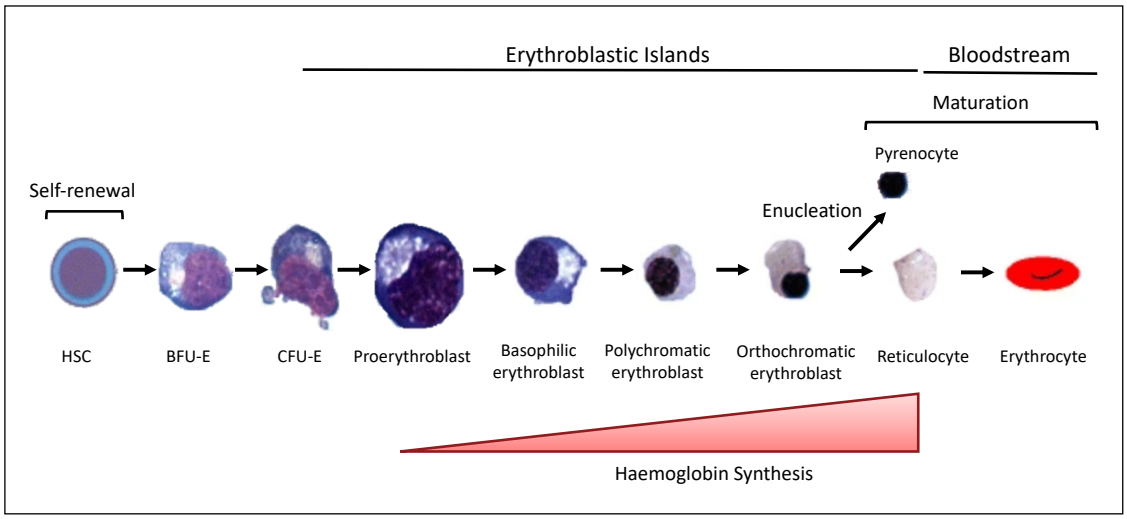


Figure 1.2: Overview of Erythropoiesis
Haematopoietic stem cells (HSCs) differentiate from committed erythroid progenitors such as burst-forming unit erythroids (BFU-Es) to various erythroblast stages to finally form orthochromatic erythroblasts. Orthochromatic erythroblasts then undergo terminal maturation characterized by enucleation and cellular remodeling to finally yield a mature erythrocyte circulating in the blood stream. *Adapted from Zivot et al. 2018.*

During erythropoiesis, haematopoietic progenitors execute a cell division program in conjunction with a differentiation program suggesting a key interplay between cell cycle and differentiation (Liang Zhu and Skoultchi, 2001). As work by Pop et al. demonstrates, the rapidly-dividing CFU-E progenitors will undergo a total of three cell divisions in 48 h. This progression through S-phase is a key commitment step in differentiation, in which progenitor cells become dependent on the cytokine Erythropoietin and undergo important chromatin remodelling at erythroid-specific gene loci (Weintraub, 1979; Pop et al., 2010). Arrest in S-phase progression is therefore pathogenic and disruptive to correct erythropoiesis. As the progenitors reach terminal differentiation, a rapid decline of S-phase cells is observed, in accordance with cell cycle exit, highlighting the tight coordination between cell cycle and key steps in erythroid differentiation (Pop et al., 2010).

Control of erythropoiesis is primarily regulated by the kidney-derived cytokine Erythropoietin (Epo). Epo is induced under hypoxic conditions and stimulates terminal differentiation (Hattangadi et al. 2011). The mechanism of regulation of erythropoiesis by Epo is described in the Section 1.1.3.1.

1.1.3.1 Erythropoietin (Epo)

Erythropoietin (Epo) has been established as the major growth factor governing erythropoiesis. Epo is a blood-secreted cytokine targeting erythroid progenitor cells in the bone marrow where it binds to the erythropoietin receptor (Epo-R) causing receptor homodimerization and downstream activation of erythroid-specific genes (Hattangadi et al. 2011; Zivot et al. 2018). The Epo/Epo-R interaction targets numerous intracellular pathways including the JAK2 (Janus kinase 2)/STAT5 (Signal transducer and activator of transcription 5) pathway. The JAK2/STAT5 pathway has been shown to activate genes fundamental for erythroid progenitor survival, proliferation and differentiation. STAT5 disruption leads to anaemia, low erythroid progenitor counts and diminished Epo-response proving STAT5 plays an essential role in erythroid differentiation (Hattangadi et al. 2011; Dzierzak and Philipsen 2013; Zivot et al. 2018). Other downstream pathways regulated by Epo include the mitogen-activated protein kinase (MAPK) and phosphoinositide 3-kinase (PI3K) pathway that are similarly involved in erythroid proliferation and differentiation (Zivot et al. 2018). Erythropoiesis regulation by Epo is temporally regulated as Epo binds to Epo-R from the CFU-E stage to the polychromatophilic stage whereas the later stages of erythropoiesis are not Epo-dependent. Epo expression is believed to be negatively regulated by the transcription factors GATA1 and GATA2 that simultaneously drive the expression of other target genes essential for terminal differentiation (Zhao et al. 2006; Zivot et al. 2018).

Regulation of erythropoiesis is therefore a very complex and fine-tuned process involving the interplay of several factors. Any deregulation of erythropoiesis can cause a variety of disorders such as thalassemia syndromes, congenital anaemias, leukaemia and myeloproliferative bone marrow disorders. An example of such

an erythropoietic disorder is Congenital Dyserythropoietic Anaemia, which is described in detail in Section 1.2.

1.2 Congenital Dyserythropoietic Anaemia

1.2.1 General Overview

In healthy individuals, 150 billion red blood cells are produced daily in the erythroid compartment of bone marrow in a highly-regulated process called erythropoiesis. Dyserythropoiesis is the derangement of this process, leading to maturation arrest and reduction of red blood cell counts and as a consequence, anaemia (Iolascon, Esposito, and Russo 2012). Dyserythropoiesis underlies several, rare, congenital erythroid disorders, some of which are the Congenital Dyserythropoietic Anaemias (CDAs) which will be described in more details in the following sections.

The CDAs are a heterogeneous group of rare disorders, primarily characterised by ineffective erythropoiesis, erythroblast abnormalities and secondary haemochromatosis (Tamary et al. 2008; Russo et al. 2019). The hallmark of the CDAs is failure of terminal erythropoiesis, however the exact disease mechanism remains unknown.

1.2.2 Sub-types of CDA

In clinical practice, CDA has been divided into several subgroups as classified by Heimpel and Wendt, based on the identified disease genes. There are currently four well-defined types of CDA: CDA-I, II, III and IV caused by mutations in Codanin-1 (CDAN1)/ C15orf41, SEC23B, KIF23 and KLF1 respectively (Arnaud et al. 2010; Iolascon et al. 2013). It has to be noted that some cases display clinical features that do not fit into the classical four subtypes described above, a consequence of the heterogeneity of the disease. To this end, four additional CDA groups have been created, with only very few individual cases for each of these (Wickramasinghe 1998). As such, only the main sub-groups of CDA will be described in more detail, below.

1.2.3 CDA-I

1.2.3.1 Disease Phenotype

Congenital Dyserythropoietic anaemia type I (CDA-I) is an inherited recessive macrocytic anaemia displaying bone marrow abnormalities, usually diagnosed in childhood or late adolescence. In some cases, patients also display vertebral abnormalities including severe scoliosis and short stature (Tamary et al. 2005). On average, CDA patients have low haemoglobin concentrations between 8 and 11 g/dL although numbers can vary significantly on a case by case basis. In 80% of affected individuals, the anaemia is accompanied by jaundice and splenomegaly as well as gallstones present in $50-60\%$ of affected adults (Heimpel 2004; Tamary and Dgany, 2009; Shalev et al. 2017).

CDA-I is a rare and very heterogenous disease, often misdiagnosed. The accurate prevalence is therefore hard to conclude, but all combined data to date suggests the incidence of CDA-I is ~ 5 cases per million live births, indicating there may be ~ 35000 cases world-wide, however, the current diagnostic rate suggests CDA-I may be under-diagnosed (Roy and Babbs 2019).

CDA-I is often associated with a variety of limb and physical abnormalities: syndactyly in hands or feet, absence of distal phalanges and nails, duplication of distal phalanx (present in 30% of patients), short stature, patches of skin depigmentation, flattened bodies of vertebrae and sensory deafness and impairment of speech (Figure 1.3) (Iolascon and Delaunay 2009).

Along with CDA-II, CDA-I is one of the best documented CDA-subtype to date. Light microscopy studies of patient bone marrow reveal megaloblastic erythropoiesis, binuclear or multi-nuclei erythroblasts ($< 10\%$) as well as some features of inter-nuclear bridges between erythroblasts ($1.4-7.9\%$) in up to 79% of patients (Figure 1.4.a) (Tamary et al. 2008; Babbs et al. 2013; Resnitzky et al. 2017). Uniquely, electron microscopy (EM) studies of the bone marrow show a particular pattern of spongy heterochromatin present in up to 60% of the intermediate erythroblasts in all patients. This pathognomonic feature is termed 'spongy' or 'swiss-cheese like' heterochromatin (Iolascon et al. 2013; Babbs et al. 2013; Resnitzky et al.

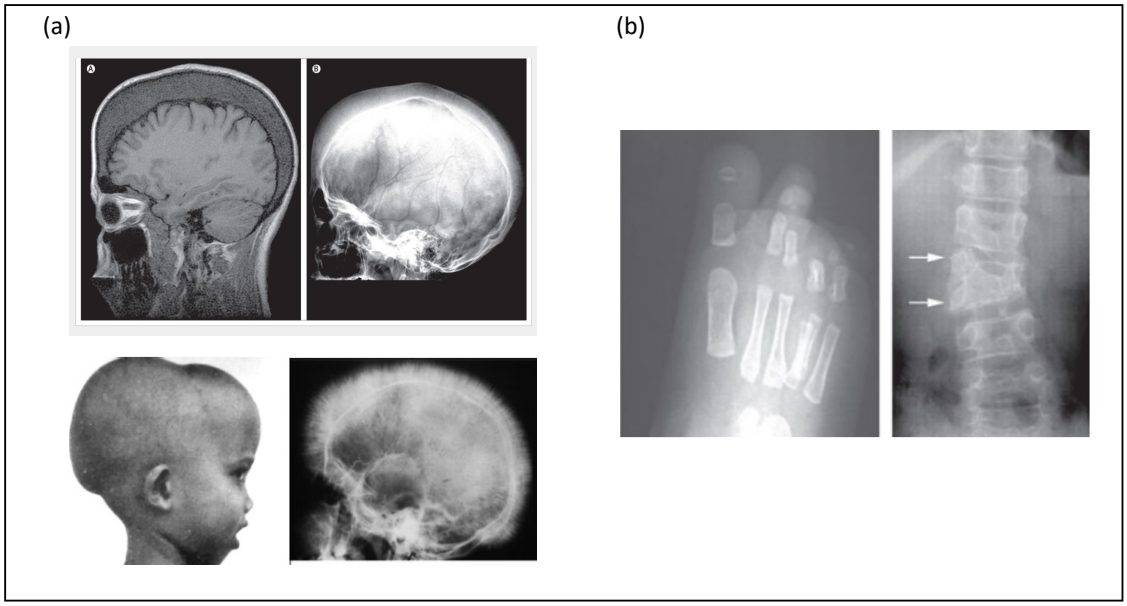


Figure 1.3: Clinical features of Congenital Dyserythropoiesis Type-I
(a) Cranial MRIs of CDA-I patients showing strong thickening of the cranial bones, bone abnormalities and widening of the diploic space.
(b) X-rays illustrating hypoplastic nails, third finger and syndactyly of the fourth and fifth toes and lumbar scoliosis of a CDA-I patient. *Adapted from Asin et al., 2015; Agrigento et al., 2017; Rathe et al., 2018.*

2017). The chromatin appears with light and dark patches in early and late erythroblasts with the nuclear envelope lined with cytoplasmic intrusions and cytoplasmic organelles (Figure 1.4.b) (Iolascon and Delaunay 2009). To date, this still remains the most useful diagnostic feature of CDA-I.

In 2002, *Codanin-1 (CDAN1)* was discovered as a causative gene for CDA-I (Dgany et al. 2002). To date, 80% of patients with CDA-I harbour bi-allelic mutations in *CDAN1*, with 51 causative *CDAN1* mutations identified to date (Babbs et al. 2013; Roy and Babbs 2019). Whole-genome sequencing studies undertaken by Babbs et al. revealed a second disease locus for CDA-I, accounting for 10% of the remaining 20% of the cases (Babbs et al. 2013). In 2013, two homozygous mutations in a previously uncharacterised gene, *C15orf41*, were identified. The remaining 10% of the cases do not harbour mutations in either *CDAN1* or *C15orf41*, suggesting the existence of a yet unidentified third disease locus.

Codanin-1 and C15orf41, the proteins encoded by *CDAN1* and *C15orf41* respectively, are both expressed in all cell types as well as being cell-essential,

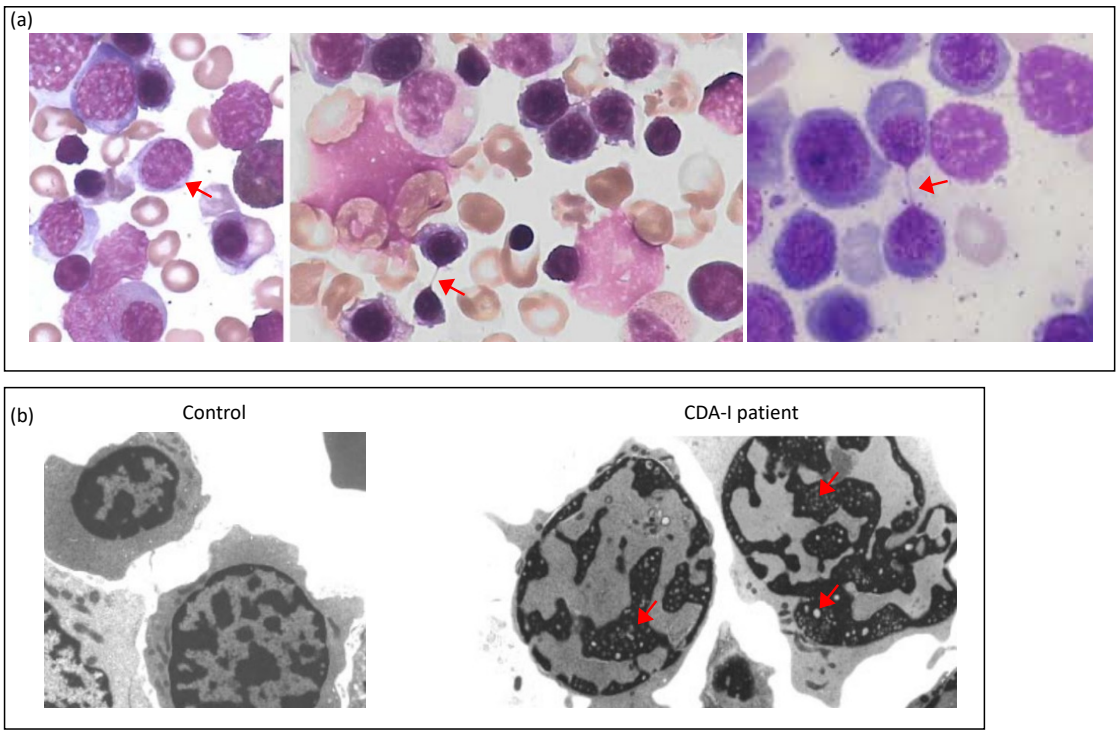


Figure 1.4: Cellular features of Congenital Dyserythropoiesis Type-I
(a) Light microscopy image of patient bone marrow erythroblasts showing the characteristic dyserythropoietic features of CDA-I such as megaloblastic erythroblasts and inter-nuclear bridges.
(b) Electron microscopy of bone marrow smear of a CDA-I patient with Codanin-1 mutations p. V690A and p. R714W showing erythroblasts with abnormal chromatin structure termed 'Spongy' or 'Swiss-Cheese' -like heterochromatin. This clinical feature is unique to CDA-I and not present in the control image. *Adapted from Wickramasinghe and Wood, 2005; Heimpel et al., 2006.*

suggesting a crucial role for both proteins in a fundamental cellular pathway that remains yet to be elucidated. Very little is known about the function of either protein and a lack of molecular reagents such as antibodies and appropriate cell lines have proven a challenge in better understanding the molecular mechanism of CDA-I. A more in-depth analysis of both disease genes can be found in Section 1.3.

1.2.3.2 Founder mutation

In this section, we present an example of a CDA-I patient profile. We take the example of a patient with the founder mutation in Codanin-1, the R1042W substitution, as an illustration of a characteristic CDA-I case (Figure 1.5).

This mutation in Codanin-1 was identified in the Bedouin population in 2002

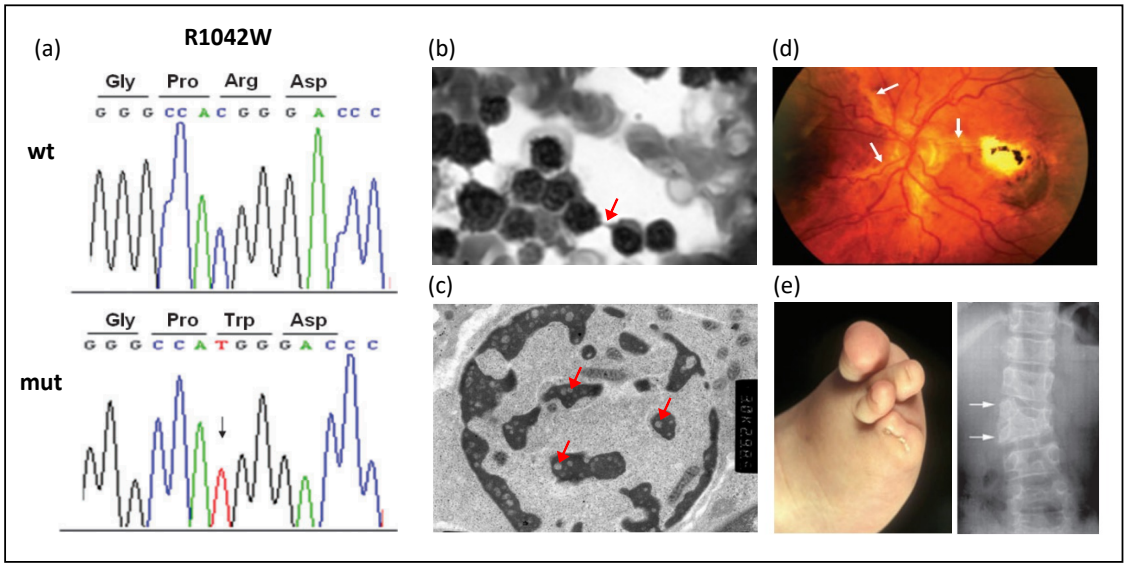


Figure 1.5: Clinical profile of a CDA-I patient with the founder mutation R1042W in *CDAN1*
(a) Sequence analysis in a normal control (wt) and the patient (mut) showing a C>T substitution in exon 24 of the patient, resulting in an arginine (Arg) to tryptophan (Trp) substitution at amino acid 1042 of Codanin-1.
(b) Light microscopy image of a blood smear of patient erythroblasts showing the characteristic inter-nuclear bridges found in CDA-I patient samples.
(c) Bone marrow electron microscopy showing heterogenous chromatin with heavily stained areas and small translucent patches termed 'spongy' heterochromatin. The nuclear envelope had folds and the cytoplasm penetrated these finger-like protrusions.
(d) Retinography of patient outlining angioid streaks (white arrows).
(e) Photograph of CDA-I patient revealing characteristic bone abnormalities such as hypoplastic nails, third finger and syndactyly of the fourth and fifth toes and lumbar scoliosis (shown by X-ray).
Adapted from Tamary et al. 2008; Agrigento et al. 2017

(Dgany et al. 2002). This CDA-I patient was diagnosed at 10 years old with chronic haemolytic anaemia. The patient also suffered from splenomegaly and jaundice. Patient blood smears displayed structural abnormalities such as deformation of cellular shape and integrity. Haemoglobin fluctuated between 8 and 9 g/dL (Tamary et al. 2008). Light microscopy of patient bone marrow showed major signs of dyserythropoiesis including bi-and trinucleate erythroblasts and some internuclear bridges. Electron-microscopy of the bone marrow confirmed CDA-I diagnosis: a high proportion of erythroblasts displayed the diagnostic spongy heterochromatin phenotype. It was also noted that the nuclear envelope had folds and nuclear pores were enlarged. Cytoplasmic organelles were also identified in the nucleus (Tamary

et al. 2008). Interestingly, the patient also displayed retinal angioid streaks which eventually led to deterioration of vision. This has previously been described in various different inherited anaemias such as sickle cell anaemia and β-thalassemia (Paton 1959; Aessopos, Farmakis, and Loukopoulos 2002).

1.2.3.3 Clinical Diagnosis

With CDA-I being a rare disorder, there is a constant need to re-evaluate and update diagnosis and patient management. This next section details the current clinical diagnosis for CDA-I.

Currently, most cases of CDA-I present in childhood or young adulthood, with patients showing low haemoglobin levels (Iolascon, Russo, and Delaunay 2011). Albeit very heterogeneous in phenotypic expression, examinations by physicians is the first step and usually reveals pallor, mild jaundice, limb deformations and splenomegaly (Iolascon, Russo, and Delaunay 2011; Shalev et al. 2017). A blood smear test is usually also performed with patient erythroblasts revealing abnormal and dismorphic red blood cells (Roy and Babbs, 2019). CDA-I often has secondary iron-overload, as such iron kinetics can be a good marker for ineffective erythropoiesis. Biochemical tests evaluating at unconjugated bilirubin and lactate dehydrogenase levels can be employed as levels will be elevated as a result of severe iron overload and ineffective erythropoiesis (Heimpel 2004).

CDA-I patient bone marrow uniquely display a spongy heterochromatin or 'Swiss cheese appearance' when analysed by EM. A bone marrow aspirate and EM analysis are therefore essential to not only confirm a CDA-I diagnosis but also allow a rapid differentiation between other types of rare anaemia such as Diamond Blackfan Anaemia (DBA) (Roy and Babbs, 2019).

Considering all clinical and biochemical tests available, the ultimate approach still remains genetic testing, which is now routinely performed to confirm CDA-I diagnosis and can replace the need for bone marrow aspiration and EM analysis which can often be invasive to the patient. 90% of CDA-I cases are accounted for by mutations

in *Codanin-1* or *C15orf41* with 51 described causative mutations in *Codanin-1* and 6 in *C15orf41* (Hamada et al. 2018; Roy and Babbs, 2019; Olijnik et al., 2021).

1.2.3.4 Current treatments

Considering CDA-I is an iron-overloading anaemia, directly related to ineffective erythropoiesis, current treatment of CDA-I focuses on normalising haemoglobin levels and iron chelation therapy. This is managed by regular blood transfusions and titrations of interferon-α (IFNα), that varies on a patient-to-patient basis. True transfusion dependence only occurs in $3-4\%$ of CDA-I cases (Heimpel 2004).

Apart from blood transfusions and iron chelation, the current most targeted and effective treatment for CDA-I is administration of interferon-α (IFNα). The positive effect of interferon-α was first noticed when interferon-α was used to treat post-transfusion chronic hepatitis C in a repeatedly transfused patient that also had CDA-I (Lavabre-Bertrand et al. 1995). The mechanism of action by which IFNα helps alleviate CDA-I symptoms is currently not known. Possible modes of action may be correcting a deficit in IFNα production in CDA-I patients or differential expression of IFN-responsive genes or via modulation of erythroblast cell cycle or basic cellular maintenance pathways. Further experiments will need to be conducted to better understand IFNα mechanism to provide better suited and tolerated therapies for CDA-I patients. Nevertheless, resistance to IFN therapy is possible and, in those cases, a bone marrow transplant or stem cell therapy might be necessary (Rathe et al. 2018). IFN therapy is not always effective and often not well tolerated, with short term side effects such as flu-like symptoms and, on a long-term, an increased incidence of autoimmune disorders such as systemic lupus erythematosus (SLE) (Niewold 2011; Pinto and Andrade 2016; Roy and Babbs 2019). Interestingly, CDA-I patients have undetectable levels of hepcidin which leads to enhanced iron absorption and deposition in organs. This inhibition of hepcidin levels reveals a potential new avenue of therapy for CDA-I patients and will require future insight into its exact molecular function and link to CDA-I (Kawabata et al., 2012).

Sustained and possibly life-long therapy is required for these patients, leading to various side effects and a poor quality of life. There is, therefore, a pressing need to further characterise the disease mechanism to provide alternative, better-tolerated therapies for patients.

1.2.4 CDA-II

1.2.4.1 Disease Phenotype

CDA-II is the most common type of CDA, characterised by mild to severe anaemia, jaundice and splenomegaly in 50% of affected individuals (Tamary and Dgany, 2009). Patients' haemoglobin is between 8 and 11 g/dl with a normal MCV count (Kamiya and Manabe, 2010). Patient erythroblast blood smears reveal characteristic CDA bone marrow deformities such as $10 - 35\%$ binucleated erythroblasts and erythroblasts with a double plasma membranes (Heimpel, 2004). EM studies of the bone marrow show the presence of vesicles with endoplasmic reticulum proteins found beneath the double plasma membrane (Figure 1.6) (Iolascon et al., 2013).

CDA-II is an autosomal recessive anaemia caused by mutation in the SEC23B gene, encoding for the cytoplasmic COPII component SEC23B (coat protein) which is a multi-subunit complex involved in accumulation of secretory cargo, deformation of the membrane and anterograde transport of correctly folded cargo from the endoplasmic reticulum towards the Golgi apparatus (De Matteis and Luini, 2011; Iolascon et al., 2013). Defects in vesicle transport therefore offer a possible explanation for the presence of the malignant vesicles found within the plasma membrane (Bianchi et al., 2009; Wang et al., 2018). How this links to the disease phenotype still remains to be understood.

To date, over 400 cases of CDA-II were identified: 52% are caused by missense mutations, 20% are due to nonsense, 13% to intronic and 13% to small indel mutations in SEC23B (Iolascon et al., 2010; Russo et al., 2011). Homozygosity for null mutations were never found, presumably lethal. In fact, SEC23B-deficient mice have no apparent anaemia phenotype but they die shortly after birth with

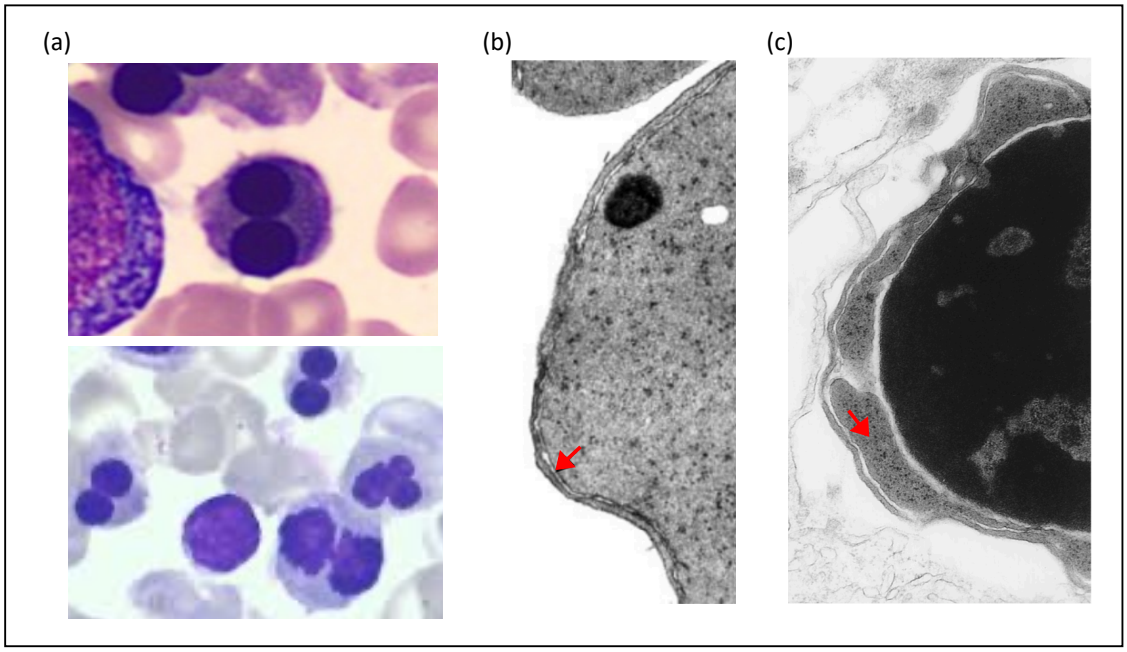


Figure 1.6: Clinical features of Congenital Dyserythropoiesis Type-II
(a) Light microscopy image of a blood smear of patient bone marrow erythroblasts showing the characteristic bi-nuclearity seen in CDA-II.
(b) & (c) Electron microscopy images of erythroblasts from a CDA-II patient demonstrating the peripheral cistern characteristic of this condition. The images also show the presence of vesicles with endoplasmic reticulum proteins found beneath the double plasma membrane. *Adapted from Wickramasinghe and Wood, 2005.*

degeneration of secretory tissues such as pancreas and salivary glands, confirming that SEC23B is likely to be cell essential (Iolascon et al., 2010; Tao et al., 2012).

1.2.4.2 Diagnosis

Aside from blood smears and EM studies of bone marrow to identify erythroblast abnormalities, the most specific and sensitive test to diagnose CDA-II remains the analysis of glycosylation abnormalities of band 3 and band 4.5, an anion exchanger 1 and glucose transporter 1 respectively, using SDS-PAGE (Iolascon et al., 2013). A band shift is observed due to clustering of incompletely-processed N-linked oligosaccharides as seen through structural studies of the carbohydrates in band 3 of CDA type II red cell membranes only (Lowenthal et al., 1980). CDA-II is therefore also viewed as a congenital disorder of glycosylation (Iolascon and Delaunay, 2009).

Most CDA-II patients require blood transfusions on a regular basis although the majority is not transfusion-dependent for life. Patients also show iron overload

and iron accumulation, sometimes requiring iron chelation therapy to mediate those symptoms (Heimpel, 2004).

1.2.5 CDA-III

CDA-III is the rarest type of CDA. Inherited dominantly, it is characterised by large multinucleated erythroblasts (with up to 12 nuclei per cell), erythroid hyperplasia and usually a lack of splenomegaly (Figure 1.7) (Lowenthal et al., 1980). In CDA-III, anaemia is mild and iron overload does not seem to be clinically significant so no transfusions are usually required (Tamary and Dgany, 2009 ; Kamiya and Manabe, 2010).

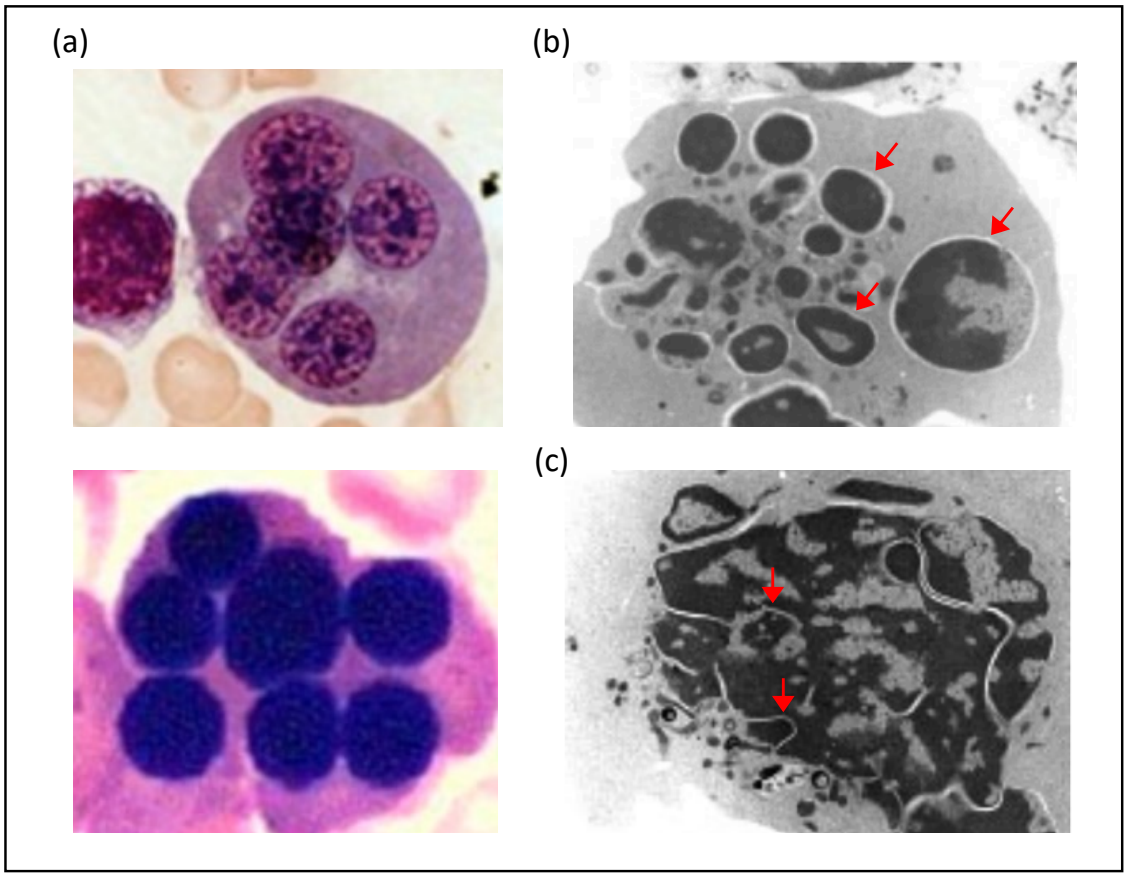


Figure 1.7: Clinical features of Congenital Dyserythropoiesis Type-III
(a) Light microscopy image of patient bone marrow erythroblasts showing the characteristic multi-nuclearity seen in CDA-III.
(b) & (c) Low magnification electron microscopy image of a giant erythroblast from a patient with CDA-III showing karyorrhexis (fragmentation of the nucleus) and intranuclear clefts. Adapted from Wickramasinghe and Wood, 2005.

On a genetic level, mutations in *KIF23* were shown causitive with CDA-III (Liljeholm et al., 2013). *KIF23* encodes a kinesin-superfamily protein, mitotic kinesin-like protein 1 (MKLP1), essential for cytokinesis (Iolascon et al., 2013). MKLP1 forms a complex with Arf6 (adenosine 5′-diphosphate-ribosylation factor 6). Coincidentally, Arf6 knockdown results in bi- and multi-nucleated cells which is a clinical hallmark of CDA-III erythroblasts (Liljeholm et al., 2013). However, the exact molecular mechanism of how these proteins lead to CDA-III still remains unknown.

In summary, Congenital Dyserythropoietic Anaemia is a rare, heterogenous group of bone marrow failure syndromes characterised by very specific dyserythropoietic features and haematological abnormalities. These main CDA features are outlined in the diagram below (Figure 1.8).

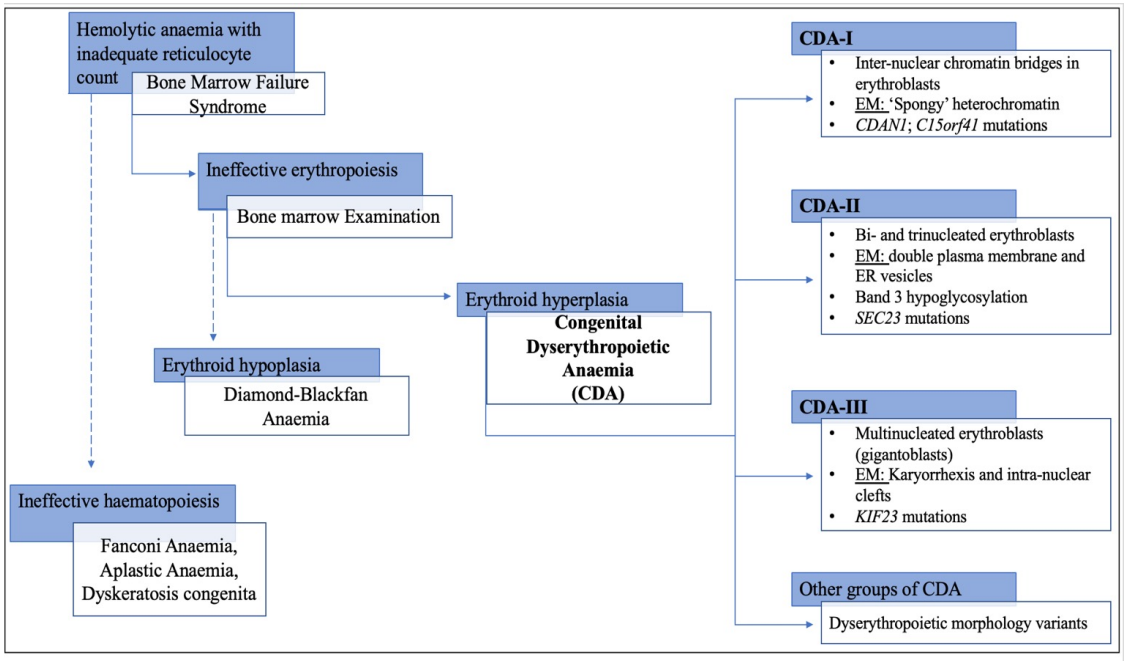


Figure 1.8: Summative diagram of CDA
CDA is a group of bone marrow failure syndromes characterized by ineffective erythropoiesis and erythroblast abnormalities as outlined for each type of CDA. *Adapted from Kaushansky et al., 2006.*

1.3 Disease Genes in CDA-I

1.3.1 Codanin-1 (CDAN1)

1.3.1.1 Biological Characterisation

Codanin-1 or *CDAN1*, the main disease gene of CDA-I was the first of all CDA-genes to be described (Dgany et al., 2002). No patient homozygous for null-type mutations has yet been identified, suggesting that a complete absence of Codanin-1 is lethal. In addition, Codanin-1 knockout mice embryos die in utero at 6.5 days before erythropoiesis onset revealing a critical role of Codanin-1 in an essential, yet unknown cellular pathway in addition to its role in erythropoiesis (Renella et al., 2011; Iolascon et al., 2013).

Codanin-1 is a 28 exon, highly evolutionarily-conserved gene that encodes for a relatively large, poorly characterised protein, Codanin-1. Codanin-1 has identifiable homologues in fish, frogs and flies, but not in worms and yeast (Dgany et al., 2002). Very little is known about the structure of the 134 kDa protein. Structure predictions reveal a mainly disordered protein with only one putative ordered domain identified, namely the B domain. Codanin-1 is thought to interact with the well-described histone chaperone Asf1 through this B domain in the N-terminus of the protein (Figure 1.9) (Tamary et al., 2010; Ask et al., 2012). The interaction between Codanin-1 and Asf1 is detailed in Section 1.3.1.3. It is yet to be determined whether this known interaction occurs in erythroblasts and is causative of CDA-I or whether Codanin-1 has other functions independent from its role in CDA-I.

To date, 51 causative *Codanin-1* mutations have been recorded in The Human Gene Mutation Database, with the majority being missense mutations (Figure 1.10.a) (Lin et al., 2020). Interestingly, these mutations cluster into six distinct regions of Codanin-1, mainly within the C-terminus of the protein: 340-398aa, 594-626aa, 648-754aa, 852-910aa, 1014-1071aa and 1101-1142aa (Figure 1.10.b). Considering the structure of Codanin-1 is unknown and Codanin-1 is thought to be a mainly disordered protein, it can be hypothesised that these six mutational clusters represent functional domains (Tamary et al., 2005; Olijnik et al., 2021).

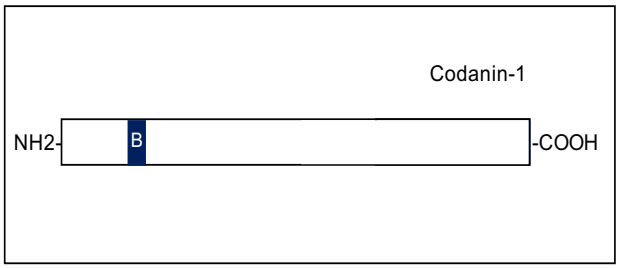


Figure 1.9: Schematic of the predicted structural domains of Codanin-1
Codanin-1 is a mainly disordered protein with the only structural domain identified so far being the B domain (in blue) in the N-terminus. The B domain is predicted to mediate the binding of Codanin-1 and the histone chaperone Asf1. *Adapted from Ask et al. 2012.*

Interestingly, the predicted B-domain and binding site to Asf1 holds no patient mutation reported thus far (Figure 1.10.c) (Olijnik et al., 2021).

There is currently conflicting data on localisation of Codanin-1 within the cell. Some reports show Codanin-1 is enriched in the nucleus, specifically the nucleolus although over-expression experiments place Codanin-1 in the cytoplasm (Noy-Lotan et al., 2009; Ask et al., 2012; Olijnik et al., 2021). In HeLa cells, Codanin-1 has been shown to associate with heterochromatin during interphase and be excluded from mitotic condensing chromosomes (Noy-Lotan et al., 2009). Cellular expression and levels of Codanin-1 are also unclear. In a patient with a IVS-12+5G>A substitution which leads to a splice-site mutation and two Codanin-1 transcripts, a decrease of 65% of Codanin-1 synthesis was identified by Western blotting (Tamary et al., 2005). This specific mutation was associated with a severe manifestation of CDA-I. There is currently no evidence that reduced levels of Codanin-1 lead to more or less severe cases of CDA-I; further investigations will need to be conducted to better understand phenotype-genotype correlations.

1.3.1.2 3D Structure and function prediction

Very little is known about the structure and function of Codanin-1, with only little information available through homologue studies or structure predictions. To date, the best characterised homologue of Codanin-1 is discs lost, its *Drosophila melanogaster* homologue (Pielage et al., 2003). Furthermore, a recent study used

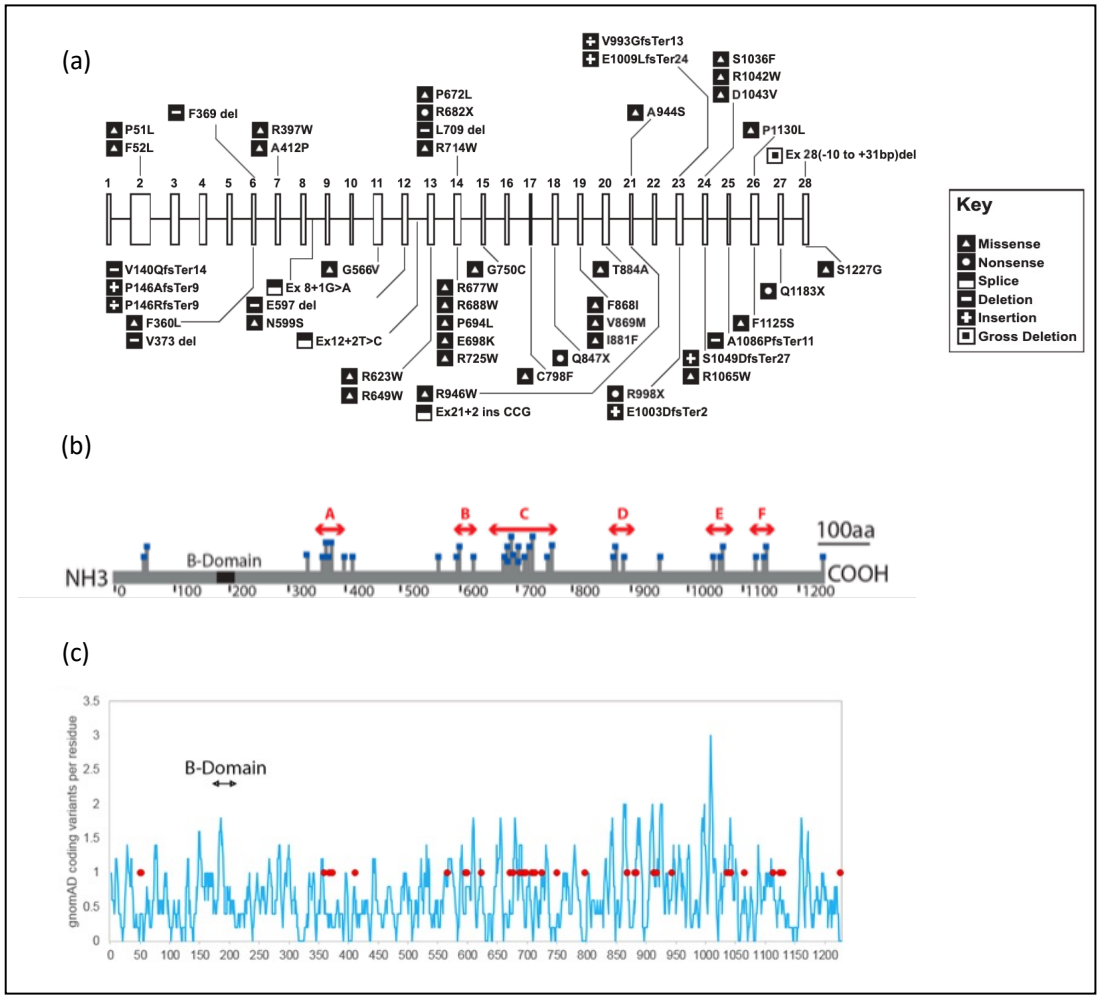


Figure 1.10: Representation of all known pathogenic variants associated with *CDAN1*

(a) Representation of all known patient mutations in *CDAN1* associated with CDA-I. The exons are numbered to show the location of each mutation.

(b) Patient mutations in *CDAN1* cluster into 6 distinct domain labelled A to F (red arrows) highlighting potential functional domains of Codanin-1. Interestingly, no patient mutations cluster in the B-domain of the protein.

(c) Graph representing the distribution of missense variants in the general population showing missense and in-frame indels per residue. Variability is equivalent across *CDAN1*, including the B domain. *Adapted from Olijnik et al., 2021.*

3D predictions to identify potential new domains in Codanin-1, homologous to the protein CNOT, a conserved subunit of the CCR4-NOT complex involved in RNA-related processing (Swickley et al., 2020). Both of these homology studies are described in detail in the following sections (1.3.1.2.1 & 1.3.1.2.2).

1.3.1.2.1 Drosophila Homologue The *Drosophila* gene *discs lost (dlt)* has been reported to encode a homolog of Codanin-1, a cytoplasmic protein predicted to be involved in cell survival and cell cycle progression, albeit to varying levels. (Pielage et al. 2003). The discs lost protein comprises 1240 amino acids and homologous, similar-sized proteins are present in vertebrates but not *C-elegans*, yeast or plants. Codanin-1 was identified as a vertebrate homolog of a similar size (1226 amino acids), with the first 300 amino acids being the least conserved (Pielage et al., 2003).

Discs lost is expressed ubiquitously during development and has been identified as a cytoplasmic protein. Cells mutated for *discs lost* appear to have a growth disadvantage, affecting the survival of differentiated cells. RNAi experiments suggest that *discs lost* is required for progression through both G1 and S phase and that lack of discs lost halters S phase completion thereby blocking mitosis (Pielage et al. 2003). *Discs lost* is therefore predicted to function in control of cell cycle progression. Interestingly, in mutant *discs lost* larvae, a 30% decrease in the number of haemocytes were noted, suggesting a possible role of discs lost in erythropoiesis (Pielage et al. 2003).

Therefore, *discs lost* might hold a role in coordinating cell cycle dynamics and erythropoiesis, a possible function for its homologue, Codanin-1.

1.3.1.2.2 Codanin-1 3D Structure predictions Although Codanin-1 is a large and mostly disordered protein, a recent study identified a potential new structural homologue for Codanin-1. Using 3D structure simulations and homology models, a high level of similarity was identified between Codanin-1 and the scaffold CNOT1 subunit of the CCR4-NOT complex, offering new hypothesis for the function of Codanin-1 (Swickley et al., 2020). CCR4-NOT is a complex involved in RNA metabolism and function including mRNA deadenylation, translational repression and transcriptional control.

CNOT1 and Codanin-1 appear to be structurally similar, with three domains showing the highest homology scores and conservation during evolution. The three domains correspond to amino acids 299-418, 478-758 and 853-994 of Codanin-1.

Interestingly, four of the six CDA-I patient mutation clusters identified in Codanin-1 fall within those three domains, suggesting a possible clinical relevance for these structural predictions (Figure 1.11) (Olijnik et al., 2021).

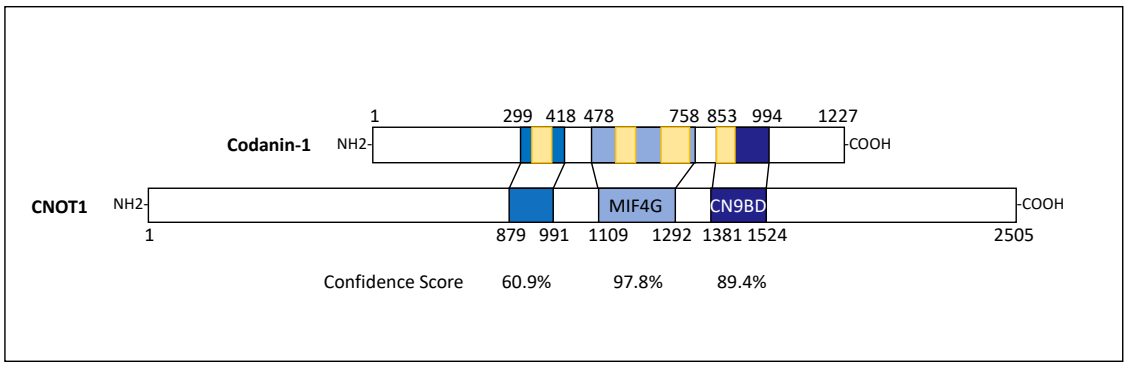


Figure 1.11: Schematic of predicted Codanin-1 domains based on CNOT 3D structure similarity
Three domains with high structural similarity were identified. Two of those domains correspond to the MIF4G and CN9BD domains of CNOT1.
In Codanin-1, four out of the six CDA-I patient mutation clusters (in yellow: amino acids 340-398, 594-626, 648-754 and 852-910) fall within the three predicted functional domains.
Adapted from Swickley et al., 2020 and Olijnik et al., 2021.

Within CNOT, two of those homology domains are well characterised: the MIF4G and the CN9BD domains, the second and third domain identified, respectively. The MIF4G is the region that models Codanin-1 with the highest confidence. This domain is involved in chromatin assembly factor 1 (CAF1) binding. CAF1 associates with CCR4-NOT complex to form a deadenylase complex, involved in mRNA poly(A) tail shortening (Petit et al., 2012; Maryati, Airhihen, and Winkler, 2015; Swickley et al., 2020). The CN9BD domain is designated as the CAF40/NOT9-binding domain. CAF40 binds to CNOT1 and acts as a scaffold protein for several factors involved in mRNA metabolism including miRNA degradation, degradation of mRNAs containing AU-rich elements and mRNA decay (Raisch et al., 2019; Swickley et al., 2020).

CNOT1 has also been shown to bind and repress transcriptional activation by oestrogen receptor α (Swickley et al., 2020). This function of CNOT1 is dependent on nine nuclear receptor boxes which are L-XX-LL sequence motifs where L is leucine and X is any amino acid. Further strengthening the similarity between the two

proteins, eight L-XX-LL motifs were identified in Codanin-1. This finding suggests that Codanin-1 and CNOT1 could be involved in similar pathways, suggesting involvement of Codanin-1 in nuclear receptor signalling and transcription (Plevin, Mills, and Ikura, 2005; Swickley et al., 2020).

Altogether, this data suggests that Codanin-1 could act as a scaffold protein, bringing together potential binding partners and modulating their activity enzymatically or through cellular localisation modulation. Further experiments will have to be conducted to identify Codanin-1's basic cellular function and role in CDA-1 and confirm any predictions made to date.

1.3.1.3 Asf1 Interaction

Characterisation of Asf1 complexes identified Codanin-1 as a new cytosolic binding partner to Asf1 in HeLa cells by interactome mass-spectrometry, a high-throughput proteomic screen and reciprocal immunoprecipitations of Asf1 and Codanin-1 (Ask et al., 2012). The only structural domain identified on Codanin-1 is a binding-domain known as the B domain on the N-terminus of the protein, where Codanin-1 binds to Asf1. It is unclear whether this interaction with Asf1 occurs in erythroblasts.

Nucleosome assembly in a key step towards restoration of chromatin on newly-replicated DNA. A key player in histone deposition is the histone H3-H4 chaperone anti-silencing function 1 (Asf1). Asf1 binds to histones H3-H4 in the cytoplasm and in complex with Importin-4 coordinates nuclear import of the histone dimers to further downstream processing factors such as the histone chaperone chromatin assembly factor 1 (CAF1) (Mello et al., 2002; Campos et al., 2010). The exact molecular mechanism of this pathway is not well understood. Codanin-1 has been hypothesised to act as a negative regulator of Asf1, sequestering the protein in the cytoplasm by binding directly to it thereby limiting the supply of histones bound to Asf1 to the nucleus (Figure 1.12) (Ask et al., 2012). The B domain has been shown to have homology to CAF-1's binding domain to Asf1 further strengthening the hypothesis that Codanin-1 and Asf1 interact directly. Mutations in Asf1 key binding residues (D36A and D37A) abolish interaction of Asf1 with Codanin-1 and

CAF-1 all together suggesting Asf1 binds to either interaction partner in a mutually exclusive manner (Ask et al., 2012). In agreement with this, Codanin-1 depleted cells show a striking increase of chromatin-bound Asf1 suggesting Codanin-1 can indeed act as a negative regulator of Asf1 (Ask et al., 2012). It should be noted that Codanin-1's interaction with Asf1 has also been shown to be histone independent as Codanin-1 still binds to a mutant version of Asf1 with impaired histone binding (V94R mutant) (Jasencakova et al., 2010; Ask et al., 2012). The exact contribution of Codanin-1 to histone deposition therefore still remains to be elucidated.

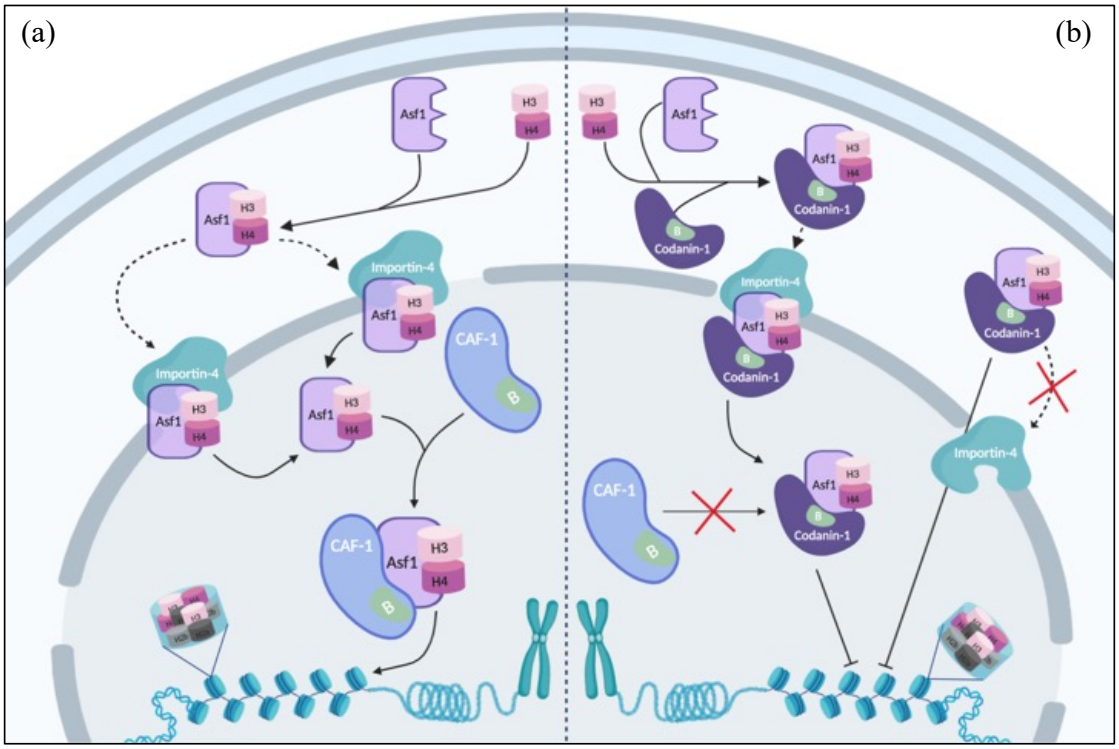


Figure 1.12: Proposed model for Codanin-1 and Asf1 interaction
(a) In absence of Codanin-1, more Asf1 is present on chromatin, perhaps because Asf1-H3-H4 shuttling to the nucleus is accelerated. (Ask et al., 2012)
(b) Codanin-1 acts as a negative regulator of Asf1 function in provision of histones to replicating chromatin. Codanin-1 interacts directly with Asf1 via the B domain and can sequester the chaperone in the cytoplasm away from replicating DNA. Codanin-1 binding to Asf1 is mutually exclusive with CAF-1, making it a potential inhibitor of Asf1 histone donor function. *Adapted from Ask et al., 2012.*

There is also evidence that patient mutations in Codanin-1 impair Asf1 complex formation thereby providing a potential disease mechanism. Immunoprecipitations with patient mutant Codanin-1 proteins with C-terminal R714W and R1042W

substitutions failed to show interaction between Codanin-1 and Asf1 (Ask et al., 2012). With reports that Codanin-1 is a cell cycle-regulated protein, highly expressed in S-phase, it may be that Codanin-1 is indeed involved in histone deposition and remodelling during DNA replication (Noy-Lotan et al., 2009). Abnormal histone delivery could therefore lead to the specific heterochromatin phenotype seen in CDA-I. It should also be noted that although Codanin-1 and Asf1 form a complex, there is no evidence to date for C15orf41, the other CDA-I disease gene, to interact with Asf1.

Very little is known about the cellular functions of Codanin-1. While it could be that the interaction between Codanin-1 and Asf1 is involved in the disease mechanism of CDA-I, this still remains unconfirmed and it cannot be excluded that this interaction is independent of the role of Codanin-1 in CDA-I. Further experiments will need to be conducted in order to shed more details on the cellular function of Codanin-1 and the role of Asf1 and histone deposition in CDA-I.

1.3.2 C15orf41

1.3.2.1 Discovery

The discovery of *C15orf41* as a disease gene for CDA-I was prompted by the report of a family with CDA-I but no identifiable mutations in *Codanin-1*. Affected siblings showed the characteristic haematological features of CDA-I: megaloblastic erythroblasts, bi- and multi-nuclear erythroblasts and inter-nuclear chromatin. EM studies confirmed the classic spongy heterochromatin phenotype of CDA-I. All three patients required blood transfusion as they were severely affected (Figure 1.13.b) (Babbs et al., 2013).

Whole-genome sequencing revealed two homozygous mutations in a previously uncharacterised gene, *C15orf41*. One is a T to A transversion in exon 8 (c.533T>A) leading to an L178Q substitution and the other is a A>G transition in exon 5 (c.281A>G) leading to a Y94C missense change in C15orf41, the protein encoded by *C15orf41* (Figure 1.13.a) (Babbs et al., 2013). Since then, four other cases of CDA-I revealed novel mutations in C15orf41: P20R, C156Y, Y238C and H230P (Palmblad et al., 2018; Russo et al., 2019). No null mutations were identified, suggesting

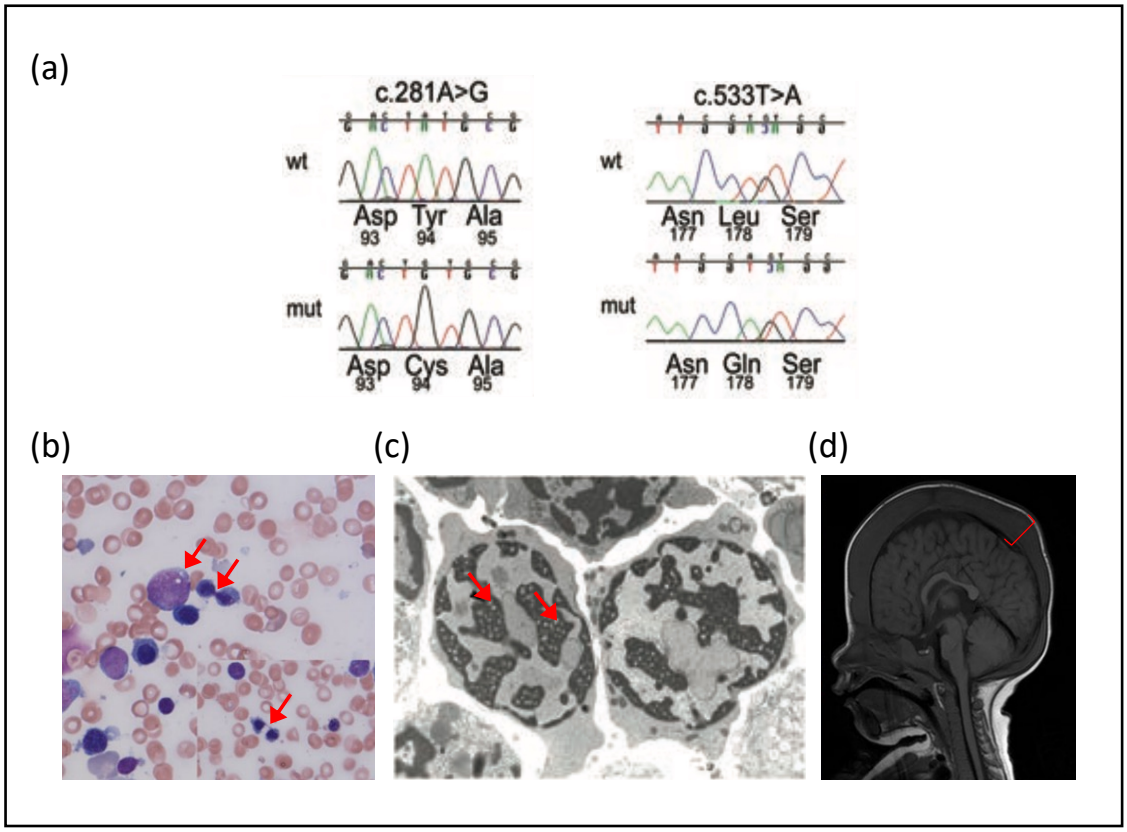


Figure 1.13: Clinical profile of CDA-I patients harbouring Y94C and L178Q substitutions in C15orf41
(a) Chromatograms from control (wt) and patient (mut) showing sequence changes identified, together with the amino acids encoded by the change and by the adjacent codons. Positions of nucleotide changes are shown above the chromatogram and are given according to numbering of the base within the open reading frame in the C15orf41 cDNA (ENST00000566621).
(b) Light microscopy image of a blood smear of patient erythroblasts with prominent dyserythropoietic features (red arrows) including megaloblastic changes, bi-nuclearity and interchromatin bridges, characteristic of CDA-I.
(c) Electron microscopy images of patient erythroblasts showing the spongy heterochromatin (red arrows) , uniquely indicative of CDA-I.
(d) Cranial MRI showing strong thickening of the cranial bones and widening of the diploic space (red bar). *Adapted from Babbs et al., 2013; Rathe et al., 2018.*

C15orf41 is cell essential. The effect of these six *C15orf41* variants on the disease mechanism is yet to be determined but all mutations lead to impaired erythroid maturation and the characteristic spongy heterochromatin phenotype of CDA-I (Babbs et al., 2013; Palmblad et al., 2018; Russo et al., 2019).

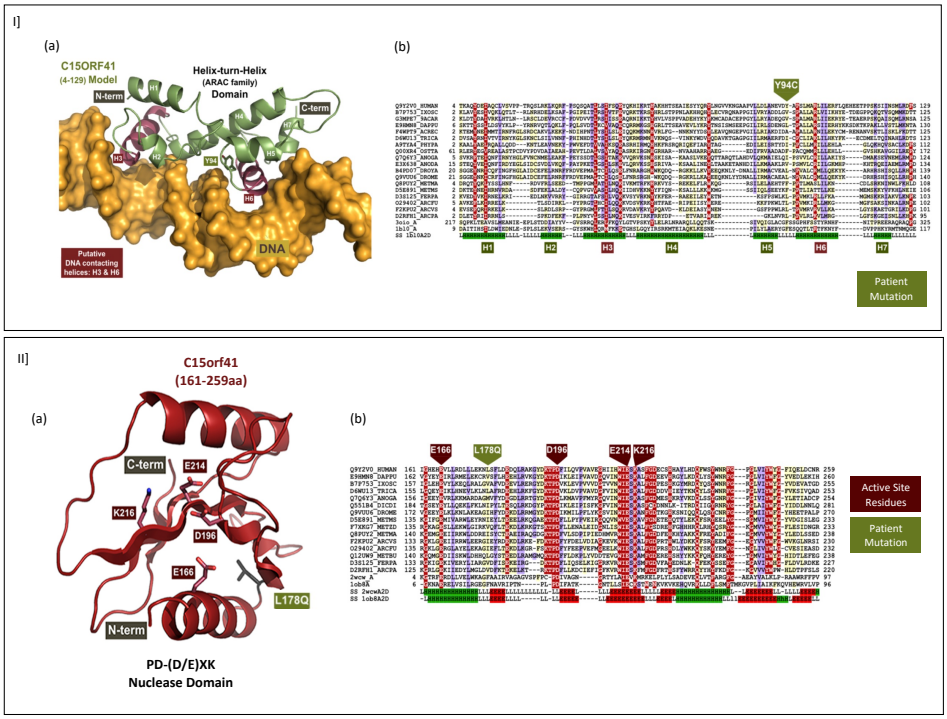


Figure 1.14: Predicted model of the tertiary structure of the Helix-turn-Helix domain in C15orf41 (I) and the nuclease domain in C15orf41 (II)
(I.a) The two predicted helix-turn-helix domains (HTH in green) docked onto a DNA molecule (in yellow). Helices are numbered and putative DNA interaction helices shown in red (H3 and H6). The founder CDA-I patient mutation, missense change Y94C, is indicated in green. (DNA was taken from Rhee et al., model adapted from Babbs et al., 2013).
(I.b) Multiple sequence alignments showing conserved residues in the HTH from the following species: Homo sapiens; B7P753_IXOSC, Ixodes scapularis; E9HMN8_DAPPU, Daphnia pulex; D6WU13_TRICA, Tribolium castaneum; Q7Q6Y3_ANOGA, Anopheles gambiae; Q9VUU6_DROME, Drosophila melanogaster; Q8PUY2_METMA, Methanosarcina mazei; D5E891_METMS, Methanohalophilus mahii; D3S125_FERPA, Ferroglobus placidus; O29402_ARCFU, Archaeoglobus fulgidus; and, D2RFH1_ARCPA, Archaeoglobus profundus. The patient mutation, Y94C, is shown in green (Babbs et al., 2013).
(II.a) The PD-(D/E)XK nuclease domain predicted for the C-terminal region of C15orf41 (161-259aa). Highly conserved residues in the PD-(D/E)XK nuclease superfamily that form part of the active site are labeled in red : E166, D196, E214 and K216. The founder mutation of CDA-I, L178Q, is shown in green.
(II.b) Multiple sequence alignments showing conserved residues of the nuclease domain from the same species as outlined above (Babbs et al., 2013).

1.3.2.2 Structure predictions

C15orf41 encodes a completely uncharacterised protein comprised of 11 exons, termed C15orf41, located on chromosome region 15q14 (Babbs et al., 2013). C15orf41

is ubiquitously expressed in all cell types as well as all throughout erythroid differentiation, suggesting a constant requirement for this protein. Secondary structure predictions reveal the 34 kDa protein contains 2 N-terminal AraC/XylS-like winged helix-turn-helix (wHtH) domains followed by a PD-(D/E) XK-like nuclease domain (Babbs et al., 2013). Homologous to other members of the PD-(D/E) XK superfamily of nucleases, C15orf41's wHTH domain is likely to be involved in substrate binding whereas the PD-(D/E) XK core constitutes predicted catalytic residues characteristic of other nucleases within this superfamily (Figures 1.14). In C15orf41, those catalytic residues are E166, D196, E214 and K216 at the C-terminus of the protein.

The significance of the newly-discovered patient mutations come into play here. The founder L178Q mutation is mapped within the predicted nuclease domain of C15orf41, altering a highly conserved hydrophobic leucine to a polar glutamine. The other founder Y94C mutation is located in the predicted helix-turn-helix motif of the protein, likely to disrupt substrate interaction. The P20R mutation is also located in the substrate binding domain and C156Y, Y238C and H230P are all located adjacent or within the predicted nuclease domain (Figure 1.15) (Palmblad et al., 2018; Russo et al., 2019). As patient mutations have been shown to not destabilise the entire protein, it is highly probable that the mutations would disrupt the function of C15orf41 and should be considered when characterising C15orf41 (Olijnik et al., 2021).

C15orf41 is predicted to hold a highly conserved function, with homologues broadly distributed in eukaryotes and in members of the archaea and viruses and interestingly, in all the same species as Codanin-1 (Roy and Babbs, 2019). So far, three potential post-translational modification sites have been described: a K50 acetylation, a T114 phosphorylation and a K176 ubiquitination sites although the biological relevance of these still remains unknown (M. S. Ahmed et al., 2018).

There is currently conflicting data on C15orf41 expression of mutant forms of the protein. Protein expression seems to vary depending on the mutation found within the gene. Whereas the Y94C variant does not lead to any variation in gene

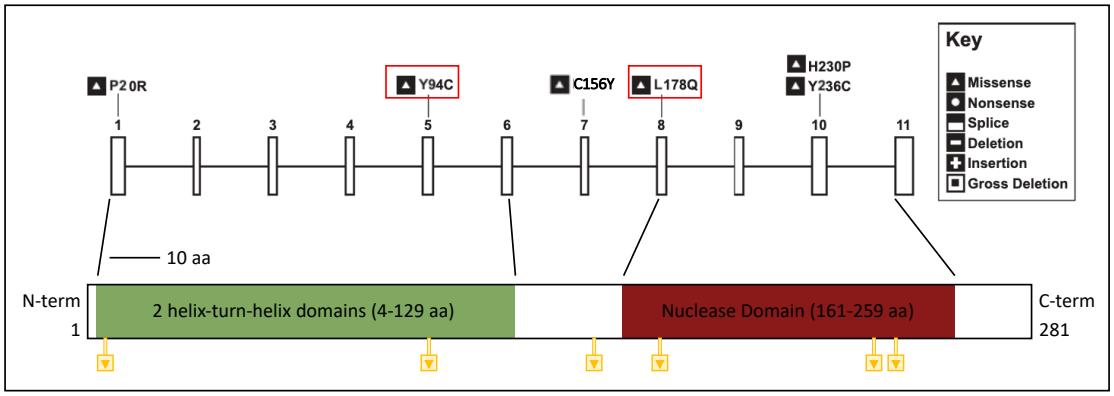


Figure 1.15: Representation of all known pathogenic variants in *C15orf41* Schematic of *C15orf41* predicted domains with exons numbered to show the location of each patient mutations. 6 patient missense mutations have been identified in *C15orf41*: P20R, Y94C, C156Y, L178Q, H230P and Y236C. Founder mutations Y94C and L178Q are indicated in red. It can be hypothesized that patient mutations impair the predicted structure or function of C15orf41. *Adapted from Roy et al., 2019 and Olijnik et al., 2021.*

expression, mutations within the predicted nuclease domain such as H230P show a decrease in *C15orf41* expression (Russo et al., 2019). There is also conflicting data on the localisation of the protein: while some data suggests C15orf41 is a nuclear protein enriched in the nucleolus, others have reported cytoplasmic localisation of the protein (Ask et al., 2012; Russo et al., 2019; Olijnik et al., 2021). Further research is therefore necessary to confirm the exact expression pattern and localisation of the protein.

Very little is known about C15orf41 and its role in CDA-I as well as a yet unknown essential cellular pathways required for normal cellular maintenance. Further experiments will need to be conducted to better understand the newly-discovered, yet uncharacterised disease-related protein.

1.3.2.3 The PD-(D/E) XK Superfamily of Nucleases

The PD-(D/E) XK superfamily of nucleases constitutes a diverse group of phosphodiesterases involved in a variety of genome housekeeping processes, ranging from DNA repair to tRNA-intron splicing. Although very varied in sequence similarity, all members of this superfamily contain a common structural core consisting of a central, four-stranded β-sheets flanked by two α-helices on both sides (Knizewski et al., 2007). This motif constitutes the active site, exposing the conserved catalytic

residues, namely an aspartic acid, glutamic acids and lysine, hence the name PD-(D/E) XK (Figure 1.16) (Kosinski, Feder, and Bujnicki, 2005). These residues are predicted to be involved in various catalytic processes as well as coordination of several essential divalent metal ion cofactors (Steczkiewicz et al., 2012).

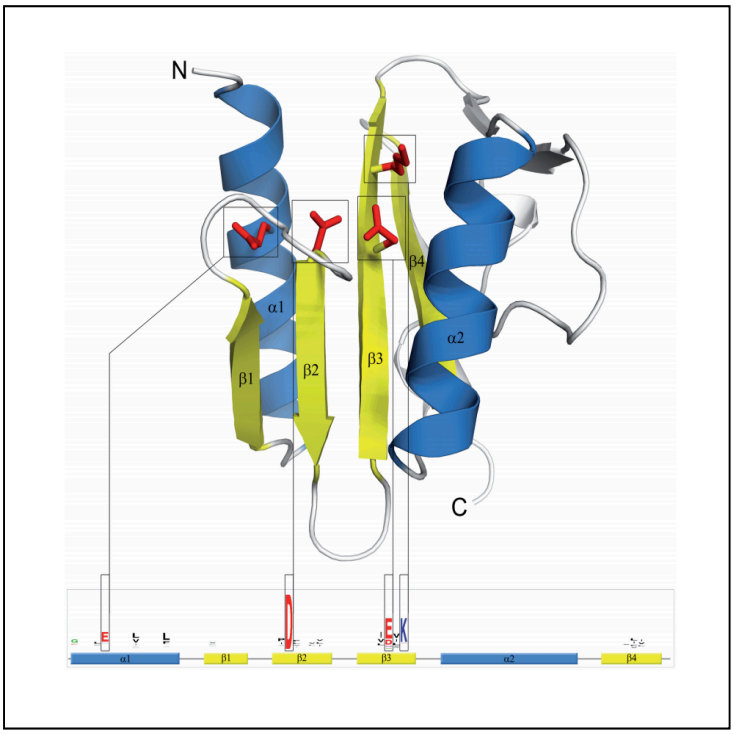


Figure 1.16: The PD-(D/E) XK nuclease fold
The PD-(D/E) XK superfamily has a common structural core consisting of a central, four-stranded β-sheets (in yellow) flanked by two α-helices (in blue) on both sides. Critical active sides residues are shown in red: an aspartic acid, glutamic acids and lysine. *Adapted from Steczkiewicz et al., 2012.*

Currently, there are 60 diverse families grouped under this large umbrella of enzymes, all of which play a role in nucleic acid maintenance and processing. This includes restriction enzymes, mismatch repair enzymes, Holliday Junction resolvases and DNA repair nucleases. Notable members include Type II restriction endonucleases BamHI or EcoRI, tRNA intron endonuclease Sen15, mismatch repair enzyme MutH and DNA repair factor XPF-ERCC1 (Knizewski et al., 2007; Steczkiewicz et al., 2012).

Malfunction of other members of the PD-(D/E) XK superfamily is known to cause several severe, hereditary diseases. For instance, mutations in RNA processing

and tRNA splicing lead to Pontocerebellar Hypoplasia (PCH) which is motor neuron disease severely affecting brain development (Steczkiewicz et al., 2012). Mutations in the DNA repair nuclease from the PD-(D/E) XK superfamily XPF-ERCC1 cause a variety of genetic disorders such as Xeroderma pigmentosum and Fanconi Anaemia. Patients with those syndromes display enhanced sensitivity to DNA damaging agents, developmental abnormalities as well as predisposition to cancer (Kosinski, Feder, and Bujnicki, 2005; Steczkiewicz et al., 2012).

1.3.3 C15orf41-Codanin-1 Complex

Mutations in *C15orf41* and *CDAN1* both cause CDA-I and result in very similar and specific erythroid abnormalities. Due to the specific disease phenotype of CDA-I, it is highly likely both factors function in the same pathway.

Unpublished data from our collaborators at the Gileadi group in the Structural Genomics Consortium at the University of Oxford hinted at a possible interaction between C15orf41 and Codanin-1 using immunoprecipitation experiments in which HEK293 whole cell extracts were passed through immobilised synthetic Codanin-1 peptides and the outputs purified and analysed by mass spectrometry. C15orf41 was identified as a binding partner to Codanin-1 and C15orf41's binding region was narrowed down to the C-terminus of Codanin-1 through partial trypsin digestion (Newman et al. unpublished). This hypothesis has now been confirmed in three separate studies recently published. Studies by Shroff et al. 2020 and Swickley et al. 2020 use immunoprecipitation experiments coupled to mass spectrometry to identify the C15orf41-Codanin-1 complex whereas Olijnik et al. 2021 showed a direct interaction between Codanin-1 and C15orf41 using immunoprecipitation experiments coupled with western blotting in HeLa cells. In Shroff et al., stable HEK293 cell lines expressing GFP-tagged C15orf41 were generated to perform an immunoprecipitation experiment using GFP-Trap resin. The GFP-C15orf41 protein was immuno-purified with a major additional band that was identified as Codanin-1 by liquid chromatography tandem mass spectrometry (LC-MS/MS) (Shroff et al., 2020). The interaction between C15orf41 and Codanin-1 was shown

to be robust, withstanding salt concentrations of up to 1M (Shroff et al., 2020). Reciprocal co-immunoprecipitation using over-expressed Flag-tagged Codanin-1 from HeLa cells confirmed Codanin-1 and C15orf41 physically interact (Swickley et al., 2020; Olijnik et al., 2021). Interestingly, immunoprecipitations with segmental deletion fragments of Codanin-1 re-iterate that the C-terminal 222 amino acid residues of Codanin-1 are essential for its interaction with C15orf41. No domains for C15orf41 responsible for Codanin-1 interaction could be identified in these studies, probably due to the small size of the protein as truncations may be destabilising (Shroff et al., 2020; Swickley et al., 2020).

The effect of CDA-I patient mutations on complex formation were also assessed in U2OS cells using over-expression plasmids encoding for epitope-tagged C15orf41 and Codanin-1 harbouring founder mutations L178Q or Y94C for C15orf41 and R1042W for Codanin-1. In all cases, complex formation was not disrupted suggesting a different pathogenic mechanism for these mutations or that different effects occur at physiologically relevant protein levels (Shroff et al., 2020; Swickley et al., 2020).

Very little is known about the properties of the C15orf41-Codanin-1 complex. Native size exclusion experiments suggest the complex elutes at a size close to 670 kDa suggesting oligomerisation of the complex or the presence of yet unknown interaction partners (Swickley et al., 2020). Localisation studies of the C15orf41-Codanin-1 complex using over-expressed fluorescently tagged proteins suggest Codanin-1 serves as a scaffold, holding C15orf41 in the cytoplasm thereby shifting its expression from the nucleus to the cytoplasm. This is reminiscent of the observation that Codanin-1 acts as a negative regulator of Asf1, also sequestering the protein to the cytoplasm (Shroff et al., 2020). The interaction between C15orf41 and Codanin-1 was also confirmed using *in vitro* studies. To this end, hexa-histidine tagged-C15orf41 bound to a C-terminal fragment of Codanin-1 (corresponding to amino acids P800 to S1227) was purified in bacteria using nickel affinity chromatography and a stable complex was identified (Swickley et al., 2020).

Altogether, affinity-purification of epitope-tagged C15orf41 and Codanin-1 in human cells and *in vitro* confirm C15orf41 and Codanin-1 form a tight complex, with

the C-terminal 227 amino acids of Codanin-1, namely amino acids 1000 to 1227, being essential for this interaction (Figure 1.17). It can be hypothesised that both proteins share a yet unknown common biological role, likely to be pathogenic in CDA-I.

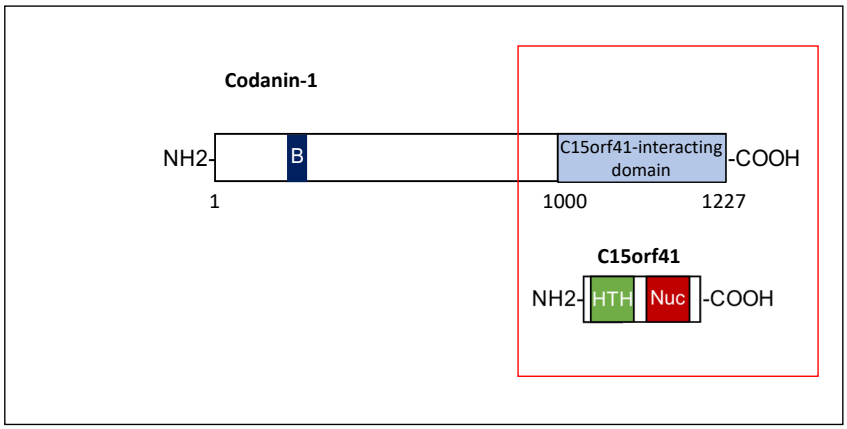


Figure 1.17: Schematic of the newly-identified C15orf41-Codanin-1 complex
C15orf41 binds to the C-terminus of Codanin-1 (amino acids 1000-1227) forming a tight complex. The red box indicated the essential part of Codanin-1 complexed to C15orf41.
B: B domain of Codanin-1, predicted to be the Asf1 binding site located in the N-terminus of the protein.
HtH: Predicted helix-turn-helix domain of C15orf41.
Nuc: Predicted nuclease domain of C15orf41.
Adapted from Shroff et al., 2020 and Swickley et al., 2020.

Although there is now convincing data showing that Codanin-1 and C15orf41 form a complex, how this relates to CDA-I pathogenesis and the function of the complex still remain unknown. It should be added that Codanin-1 is a large protein, likely to act as a scaffold and it is conceivable that Codanin-1 may have several cellular functions, some of which may be independent from C15orf41. As with localisation studies of each protein individually, there is also contradicting localisation data of the complex. Whilst the over-expression studies described above suggest that Codanin-1 may possibly sequester C15orf41 to the cytoplasm thereby acting as an activity regulator, others report nuclear and more specifically nucleolar signals for C15orf41 and Codanin-1 (Noy-Lotan et al., 2009; Ask et al., 2012; Russo et al., 2019; Shroff et al., 2020; Swickley et al., 2020; Scott et al., 2020; Olijnik et al., 2021). Current reports also suggest no change in localisation of patient mutant C15orf41 or Codanin-1 (Russo et al., 2019). It should be noted that most of these

localisation experiments have been executed by over-expression of tagged versions of the proteins in transformed human cancer cells. It is possible that the over-expression method as well as cell line choice influence the results obtained, causing protein mis-localisation. It is therefore necessary to perform these experiments in a relevant cell line, ideally an erythroid progenitor cell line as well as using an endogenous tag in order to better preserve *in vivo* conditions and fully establish the true localisation pattern of C15orf41, Codanin-1 and the C15orf41-Codanin-1 complex. In fact, only Olijnik et al. performed immunofluorescence localisation experiments of both C15orf41 and Codanin-1 in erythroblasts at endogenous levels and found both proteins to be enriched in the nucleolus (Olijnik et al., 2021). Still, how this finding relates to CDA-I pathogenesis remains to be investigated. The same caveat applies to the study of mutant C15orf41 or mutant Codanin-1 complex formation. Current data suggests patient mutations in C15orf41 or Codanin-1 have no effect on C15orf41-Codanin-1 complex formation, yet these experiments have only been performed using over-expressed forms of the respective mutants in genetically abnormal cancer cell lines. Further experiments using endogenous protein levels in relevant cell lines therefore need to be performed.

There still remain some key questions about the structure and function of C15orf41 and Codanin-1 as well as their role in CDA-I and in a broader, yet unknown biological process, some of which will be explored in this thesis.

1.4 Summary & Aims

CDA-I is a hereditary disorder characterised by unique morphological abnormalities in bone marrow erythroblasts, primarily caused by mutations in *Codanin-1*. Very little is known about the molecular pathways affected in CDA-I or the structure and function of Codanin-1, slowing down progress in understanding the biology of CDA-I.

The recent identification of a second disease gene, *C15orf41*, has marked a breakthrough in unveiling new insights into CDA-I and its molecular mechanism. Secondary structure predictions identify C15orf41 as a potential new member of the

PD-(D/E) XK superfamily of nucleases, offering new avenues for further research. Still, C15orf41 remains completely uncharacterised.

With recent reports showing that Codanin-1 and C15orf41 are both cell essential and form a tight complex, it can be hypothesised that both proteins operate together in a crucial cellular pathway, whose disruption leads to CDA-I. As both proteins were largely uncharacterised at the beginning of this thesis, basic biological tools including physiologically-relevant cellular tools and antibodies needed to be established and validated. Most current studies are performed in transformed cancer cell lines and have yielded conflicting data about cellular characterisation of both proteins. CDA-I is a rare disease and patient samples are scarce and use of primary erythroid cell lines difficult, limiting the scope of experiments feasible.

Considering the caveats of the current research on CDA-I, this thesis will address the following research questions with its ultimate aim to understand the molecular mechanism of CDA-I which will lead to an improved understanding of this disease:

1. Identify and characterise C15orf41-Codanin-1 complex formation in a relevant erythroid cell line model
2. Purify C15orf41 and Codanin-1 as well as the C15orf41-Codanin-1 complex
3. Perform an *in vitro* biochemical characterisation of the predicted nuclease activity of C15orf41

All results obtained will be described in Chapters 4, 5 and 6. This will be followed by a discussion and a future experiments and discussion section where results will be placed into context and perspectives offered. All experiments laid out in this thesis have been conducted in a collaboration between three different research groups at the University of Oxford: the Buckle group (MRC Molecular Haematology Unit at the MRC Weatherall Institute of Molecular Medicine), the McHugh group (Oncology Laboratories at the MRC Weatherall Institute of Molecular Medicine) and the Gileadi group (Centre for Medicines Discovery). This interdisciplinary collaboration offers a unique approach to characterising C15orf41, combining the expertise of each individual group to obtain a broader understanding of CDA-I.

2

Introduction II: Biochemical study of the SARS-CoV-2 nsp14-nsp10 RNA nuclease complex

In response to the 2020 SARS-CoV-2 global pandemic, scientists all around the world have had to shift their research focus towards developing a better understanding of coronaviruses and providing potential therapeutic strategies to combat this virus. This study focuses on characterising the SARS-CoV-2 nsp14-nsp10 nuclease complex. Very little was previously known about this complex rendering a complete biochemical characterisation of this nuclease essential to not only better understand the mechanism of this virus but also suggest new, alternative therapeutic avenues.

2.1 General Introduction

2.1.1 Coronavirus disease 19

The years 2019, 2020 and 2021 have been marked by the emergence of a global pandemic, Coronavirus disease 19 or COVID-19, which was declared a public health emergency of international concern by the World Health Organisation. COVID-19 is a disease caused by the SARS-CoV-2 virus, a member of the coronavirus family of viruses. This virus has affected up to ninety million people and caused two million deaths to date (John Hopkins national public health agency; Figure 2.1), shifting medical science research all around the world to focus on understanding the exact mechanism of this virus and identifying drugs and therapies to provide a cure for the disease.

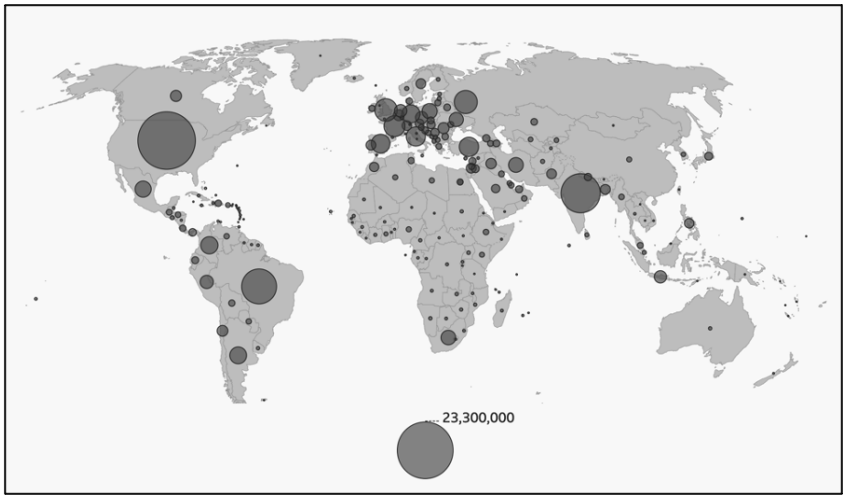


Figure 2.1: The SARS-CoV-2 pandemic mapped
COVID-19 is continuing its spread across the world with more than 90 million confirmed cases in 190 countries and two million deaths. Circles show number of confirmed coronavirus cases per country. Source: Johns Hopkins University, national public health agencies.

COVID-19 emerged in late December 2019 in the province of Wuhan in China when patients with severe respiratory conditions of unknown cause were admitted to hospital. Genome sequencing quickly identified a new coronavirus variant, 2019 novel coronavirus (2019-nCoV) now named SARS-CoV-2, as the underlying cause of this disease (Li et al., 2020). Although the origin of the virus is not fully understood, it has been suggested that the virus was passed onto humans from wild animals, including bats or pangolins, but it still remains to be investigated whether this transmission is direct or via an intermediate host (Li et al., 2020; Cevik et al., 2020).

SARS-CoV-2 is transmittable between humans via respiratory droplets between individuals in close contact, affecting all populations but in particular, elderly and people with a variety of underlying health conditions or compromised immune systems. It is characterised by a relatively long incubation time that can last up to 14 days (compared to SARS-CoV-1 which is 5 days) and a high rate of asymptomatic yet infectious cases, causing this disease to be particularly contagious (Fung and Liu, 2019; Li et al., 2020). Clinical representations of COVID-19 include fatigue, fever, dry persistent cough as well characteristic loss of taste and smell (Cevik et al., 2020). Reverse transcription quantitative PCR (RT-qPCR) tests on nasopharyngeal

and oropharyngeal swabs are used to detect nucleic acids of SARS-CoV-2 as the primary detection mode for COVID-19 (Li et al., 2020).

Currently, COVID-19 patients are treated with broad anti-viral therapies, such as interferon-α, the protease inhibitor Lopinavir/ritonavir, used in HIV therapy, the anti-malaria drug Chloroquine or nucleoside-analogues Ribavirin and Remdesivir that have shown some inhibition towards SARS-CoV-1 and MERS-CoV *in vivo*. Although the safety and efficacy of all these treatments is yet to be fully determined, it should be noted that the combination of several anti-viral drugs can cause severe side effects, therefore there is clearly the need for improved targeted and better-tolerated therapies for COVID-19 (Fung and Liu, 2019; Li et al., 2020; Cevik et al., 2020).

2.1.2 The biology of Coronaviruses

Coronaviruses are viruses from the *coronavirinae* family in the *Nidovirales* order. They are characterised by their large single-stranded RNA genomes (ranging between $\sim$ 26 kb to $\sim$ 32 kb), the largest genome of known RNA viruses (Fung and Liu, 2019). Cryo-electron microscopy studies of coronaviruses reveal the virions are spherical (of a diameter of 80-90 nm) and are characterised by protruding spike proteins on their membranes that confer this virus its name (Figure 2.2) (Klein et al., 2020).

Aside from the current COVID-19 pandemic, two other recent epidemics of viral pneumonia were caused by coronaviruses: severe acute respiratory syndrome (SARS) in the early 2000s and Middle East respiratory syndrome (MERS) in 2012 (Li et al., 2020).

The first SARS-CoV epidemic of 2003 was caused by the SARS-CoV-1 virus and the current COVID-19 pandemic by the SARS-CoV-2 virus, that has an 80% genetic sequence similarity with the SARS-CoV-1 virus. Both viruses have almost identical lifecycles and mechanisms, which will be described in the next sections.

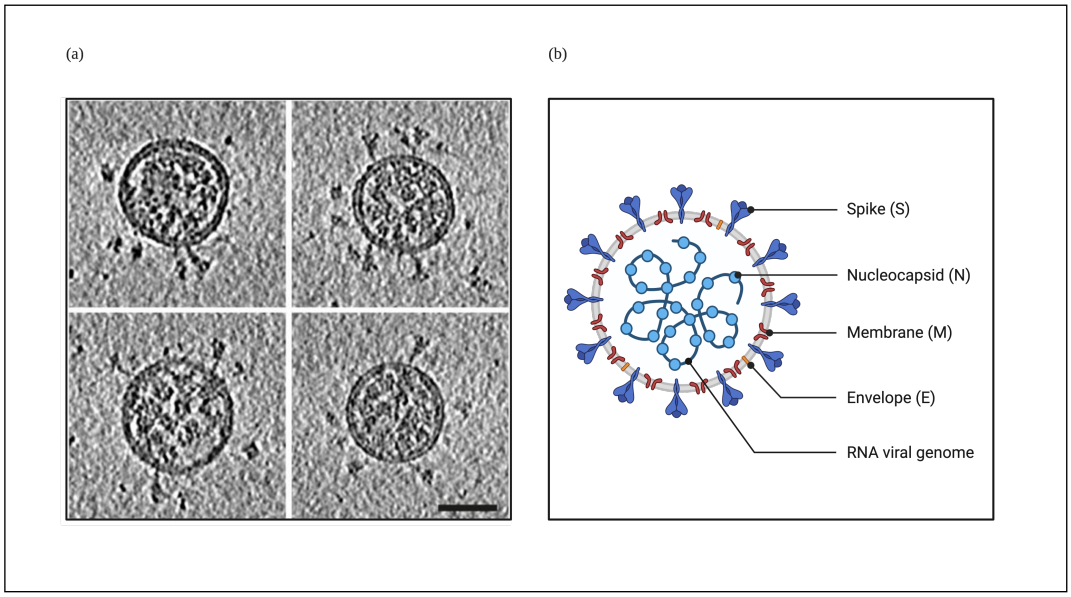


Figure 2.2: The SARS-CoV virion
(a) Four representative tomographic slices of SARS-CoV-2 virions. Virions are approximately spherical, contain granular density corresponding to the RNA genome, and have trimeric spike protein protruding at variable angles from their surfaces creating a "crown" effect, hence the name coronavirus. Scale bar corresponds to 50 nm. (Ke et al. 2020)
(b)Schematic of the SARS-CoV virion. Designated here are the four structural proteins of the SARS-CoV viruses: the spike (S), the membrane (M), the envelope (E) and the nucleocapsid (N) proteins, all of which are essential for virus assembly and infection. *Adapted from Ke et al., 2020.*

2.1.3 SARS-CoV-1 and SARS-CoV-2 lifecycle

The SARS-CoV viruses are enveloped β-coronaviruses. The viral envelope is coated by spike glycoprotein, envelope and membrane proteins. Viral entry is mediated by the spike proteins, that bind to the peptidase domain of angiotensin-converting enzyme 2 (ACE2) which is principally expressed in the airway epithelial cells and vascular endothelial cells. These spike proteins in SARS-CoV-2 have an increased capacity for the ACE2 receptor binding compared to the SARS-CoV-1, conferring the SARS-CoV-2 virus a more infectious nature. Viral entry leads to membrane fusion and releases the viral genome into the host cytoplasm (Cevik et al., 2020).

Once internalised into a host organism, the capped and polyadenylated incoming SARS-CoV genome gets immediately translated. The positive-sense genomic

RNA encodes for two large, open reading frames, ORF1a and ORF1b that result in polyproteins pp1a and pp1b, the latter due to a ribosomal frameshift that occurs in about 30% of translation events. Those non-structural polyproteins are further proteolytically cleaved into 16 non-structural proteins, 4 structural proteins and 5 accessory proteins (ORF3a, ORF6, ORF7, ORF8 and ORF9) (Li et al., 2020; V'kovski et al., 2020). The four structural proteins encode for the spike, the membrane, the envelope and the nucleocapsid proteins, all of which are essential for virus assembly and infection (Cevik et al., 2020; Li et al., 2020). The non-structural proteins form the viral replication and transcriptional complex or replisome, necessary for virus propagation. The replisome uses the positive-sense genome as a template to generate negative-sense subgenome and genome-length RNAs, which are in turn used as templates for synthesis of positive-sense subgenomic mRNAs. These subgenomic mRNAs are translated into structural and accessory proteins. The viral components are subsequently assembled, the virus matured and finally released (Klein et al., 2020) (Figure 2.3).

2.1.4 Non-structural proteins

The SARS-CoV viruses are characterised by their high-fidelity during replication of viral genome. In order to maintain this aspect, the viruses rely upon a set of carefully-coordinated protein complexes, referred to as non-structural proteins or nsps. The 16 non-structural proteins of the SARS-CoV viruses that result from proteolytic processing of the ORFs encode a variety of RNA replicating, RNA-processing and RNA-modifying enzymes (V'kovski et al., 2020). nsp3 and nsp5 (also referred to as M^{pro}) are two proteases involved in proteolytic cleavage on the polyprotein resulting from ORFs translation. nsp2 to nsp16 compose the viral replication-transcriptional complex or replisome, which is described in more detail in Section 2.1.4.1.

2.1.4.1 The replication-transcription complex or replisome

Coronavirinae viruses have a tightly regulated replication machinery, conferring a high replication fidelity rate (errors in the order of 10^{-6} to 10^{-7}) as opposed

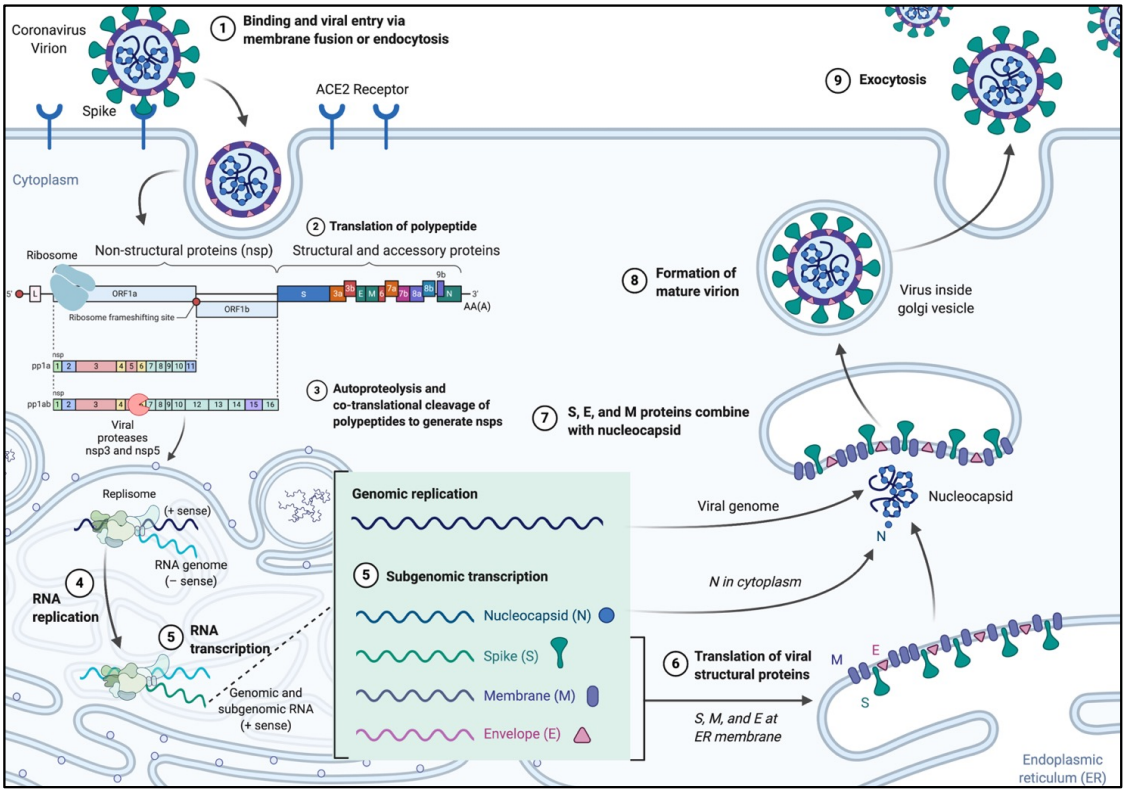


Figure 2.3: The Coronavirus lifecycle
Coronavirus virions bind to cellular attachment factors such as angiotensin-converting enzyme 2 (ACE2) receptors through their characteristic spike proteins leading to viral uptake and fusion at the cellular membrane (1). Following entry, the incoming genomic RNA is released and its two large open reading frames, ORF1a and ORF1b, immediately translated (2). The resulting polyproteins pp1a and pp1ab are co-translationally and post-translationally processed by proteases nsp3 and nsp5 to form sixteen non-structural proteins (nsps 1-16), fours structural proteins and five accessory proteins. nsp 2-16 form the viral replication and transcription complex or replisome (3). The replisome uses the positive- sense genome as a template to generate negative-sense subgenome and genome-length RNAs (4), which are in turn used as templates for synthesis of positive-sense subgenomic mRNAs (5). These subgenomic mRNAs are then translated into structural and accessory proteins (6). The viral components are subsequently assembled, the virus matured and finally released (7-9).
Adapted from Klein et al., 2020 and Hartenian et al., 2020.

to most other RNA viruses ($\sim 10^{-3}$ to 10^{-5}) (Klein et al., 2020). This requires a carefully regulated and tightly controlled replication-transcription complex or replisome (Figure 2.4).

The replisome is composed of proteins nsp12 to nsp16 that contain the core enzymatic functions involved in RNA synthesis, RNA proofreading and RNA modification (V'kovski et al., 2020). RNA synthesis is performed by the RNA-

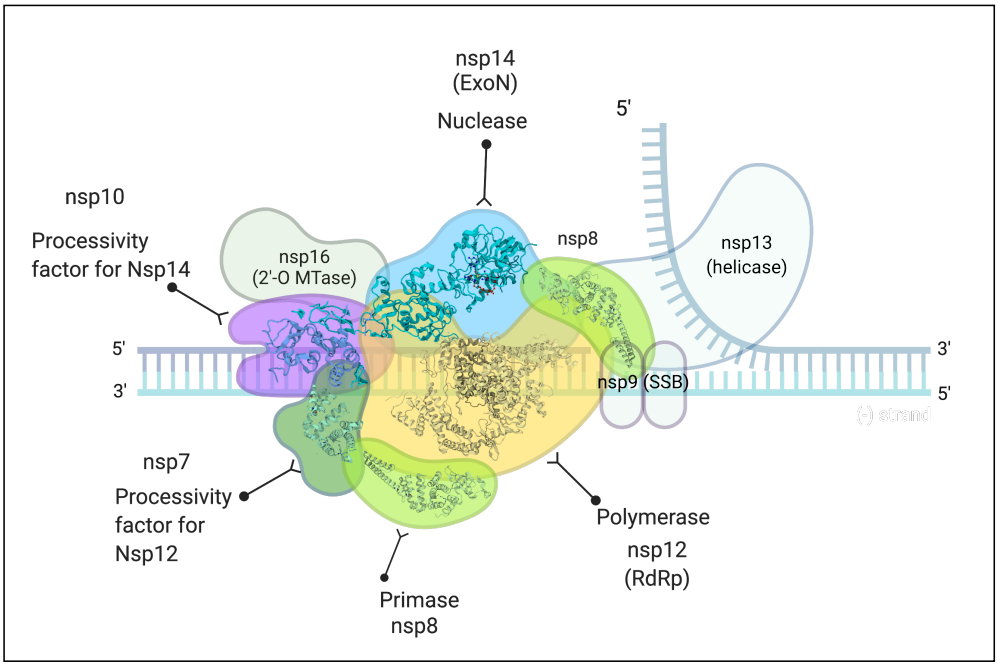


Figure 2.4: Schematic of the SARS-CoV replisome
The replisome is a large, well-organised unit consisting of nsp12 as the main replicative polymerase (RdRp), which requires its cofactors nsp8 (also acting as a primase), and nsp7. nsp13 is a dsRNA helicase. nsp16 is a 2′-O methyltransferase (contributing an essential step in 5′ RNA capping to avoid immune detection). nsp9 coats single-stranded RNA (SSB). nsp14(ExoN) and its binding partner nsp10 act as a proofreading nuclease complex and are essential for maintaining high fidelity in viral replication.
PDB files used are: SARS-CoV-1 nsp14 (5C8U), SARS-CoV-1 nsp10 (5C8U), SARS-CoV-2 nsp12 (6M71), SARS-CoV-2 nsp7 (6M71) and SARS-CoV-2 nsp8 (6M71).
Adapted from Hillen et al., 2020 and Ma et al., 2015.

dependent RNA polymerase (RdRp) nsp12 and its two, required cofactors nsp7 and nsp8, the latter acting as a putative primase. Recently, the structure of SARS-CoV-2 nsp12 bound to the nsp7 and nsp8 cofactors has been determined using cryo-EM, showing the tight regulation of RNA replication and conferring a better understanding the SARS-CoV viruses (Kirchdoerfer and Ward, 2019). The N-terminal domain of nsp12, named NiRAN (after nidovirus RdRp-associated nucleotidyltransferase) is conserved in all nidoviruses and is essential for viral growth. The C-terminal moiety is termed the RNA polymerase domain or RdRp and holds the catalytic residues involved in RNA replication (Kirchdoerfer and

Ward, 2019). nsp9 is a single-stranded RNA binding unit and an accessory protein to the polymerase nsp12 (Egloff et al., 2004). nsp13 is a zinc-binding RNA helicase that assists the polymerase complex in replication (Romano et al., 2020). Genomic capping is also performed by nsp13 in its capacity as an RNA 5′-triphosphatase and nsp14 and nsp16, which are *N*7-methyltransferases and 2′-O-methyltransferases respectively (V'kovski et al., 2020).

In order to assure accurate replication of such a large genome, the SARS-CoV non-structural proteins nsp14 and nsp10 have been implied to act as proofreading nuclease complex. nsp14-nsp10 is a RNA nuclease complex that holds a 3′-5′ exonuclease activity likely to be essential in assuring fidelity in replication, acting on mismatched bases introduced during replication (Mickaël Bouvet et al., 2010; Kim et al., 2020). The endoribonuclease nsp15 (NendoU) is an RNA endoribonuclease that is sequence-specific to uridines and holds a crucial role in evasion of host defence mechanisms (Deng and Baker, 2018; Liu et al., 2019). A possible cooperativity between those two nucleases still remains to be explored.

How exactly all the nsp proteins cooperate together and their exact interaction profiles still remain to be fully understood but it has become apparent that the polymerase complex nsp12-nsp7-nsp8 lies in the heart of the viral replication machinery, mediating most protein-protein interactions described thus far (Romano et al., 2020).

2.1.4.2 nsp14(ExoN)-nsp10

Nsp14-nsp10 is a RNA nuclease complex that is highly conserved between the SARS-CoV-1 and SARS-CoV-2 viruses with a sequence identity of 95.1% and 97.1% for nsp14 and nsp10 respectively (Yoshimoto, 2020). The structure of the SARS-CoV-1 nsp14-nsp10 has been described and will be explored below whereas the structure of the SARS-CoV-2 still remains unknown (Ma et al., 2015).

In SARS-CoV-1, nsp14-nsp10 has been shown to possess a 3′-5′ exonuclease activity that is selective for dsRNA and able to excise terminally mismatched ribonucleotide structures (M. Bouvet et al., 2012; Ma et al., 2015). It has therefore been implied the ribonuclease acts as a proof-reading nuclease. Nsp14-nsp10 plays a

key role in virus replication, as studies disrupting or inactivating its activity in mouse and cell models reveal reduced virulence and pathogenesis (Eckerle et al., 2007; Denison et al., 2011; Smith et al., 2013). In line with these observations, some reports infer an interaction between the SARS-CoV-1 polymerase complex nsp12-nsp7-nsp8 and the nuclease complex nsp14-nsp10 although no exact molecular details of this interaction have been elucidated (Ferron et al., 2018; Ogando et al., 2019). nsp14-nsp10 is a member of the DEDD superfamily of nucleases. The DEDD superfamily of exonucleases comprises a variety of RNA processing enzymes, notably proofreading domains of many DNA polymerases and DNA repair exonucleases (Zuo, 2001).

Structural studies of SARS-CoV-1 nsp14-nsp10 reveal a number of key conformational aspects of this complex. nsp14 is a bifunctional protein containing two distinct domains: the N-terminal domain ExoN and the C-terminal methyl transferase domain and requires its binding partner nsp10 for full activity and stability (Figure 2.5).

The ExoN domain confers nsp14 its nuclease activity, characterised by three conserved catalytic motifs: DE (Asp90 and Glu92), E (Glu191) and D (Asp272), placing it within the DEDD superfamily of nucleases. Those catalytic residues coordinate a metal ion, predicted to be magnesium, essential for nsp14 nuclease activity. Accordingly, alanine substitution mutations in these residues abolish nsp14 nuclease activity, resulting in a significant reduction of viral RNA synthesis. In comparison, other DEDD superfamily exoribonucleases share a catalytic mechanism characterised by the presence of two metal ions (Zuo, 2001). The nsp14-ExoN domain presents a classic α-β fold, with a β sheet flanked by α-helices, characteristic of this superfamily (Ma et al., 2015; Kirchdoerfer and Ward, 2019). Interestingly, this domain also holds three zinc-finger binding motifs coordinating three zinc atoms, setting it apart from other DEDD nucleases. Those three zinc-fingers are essential for nsp14 nuclease activity and stability, with mutations in either causing abolishment of nsp14 enzymatic activity and insoluble protein (Ma et al., 2015).

Separated by a flexible loop region, the N7 guanine-methyl transferase domain of nsp14 is functionally uncoupled from its nuclease activity and is involved in viral

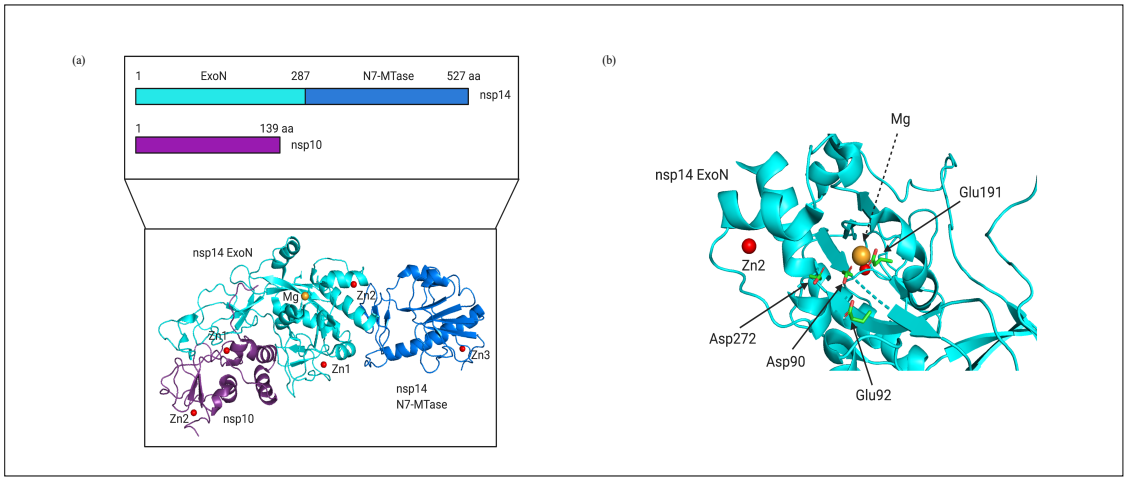


Figure 2.5: Schematic of the SARS-CoV-1 nsp14-nsp10 complex
(a) nsp14 forms a functional RNA nuclease complex with nsp10. nsp14 is a bifunctional protein containing two distinct domains: the N-terminal nuclease domain ExoN (in cyan) and the C-terminal methyl transferase domain N7-MTase (in blue). The ExoN domain confers nsp14 its nuclease activity. This domain presents a classic α-β fold. Interestingly, only one magnesium ion in the active site and three zinc finger binding motifs are present, setting nsp14 apart from other DEDD nucleases. nsp10 (in purple) has a helical domain and a β-sheet region with a distinct loop region in the C-terminus, crucial for nsp14 interaction. It also has two zinc-finger binding motifs.
(b) The ExoN domain confers nsp14 its nuclease activity, characterised by three conserved catalytic motifs: DE (Asp90 and Glu92), E (Glu191) and D (Asp272), placing it within the DEDD superfamily of nucleases (residues are highlighted in stick form). Magnesium ion is shown in orange; zinc atoms are shown in red; aa: amino acids.
PDB files used are: SARS-CoV-1 nsp14 (5C8U), SARS-CoV-1 nsp10 (5C8U).
Adapted from Ma et al., 2015 and Ferron et al., 2018.

mRNA cap synthesis through its binding partner S-adenosyl methionine (SAM) (Chen et al., 2009; Jin et al., 2013; Mickaël Bouvet et al., 2010). nsp14 does not fall within any classes of SAM-dependent methyl-transferases, due to its atypical fold of five β-sheets compared to the usual seven β-sheets. Together with nsp13, nsp16 and nsp10, this domain of nsp14 forms part of the coronavirus RNA cap machinery (Ma et al., 2015; Kirchdoerfer and Ward, 2019).

nsp10 is composed of a helical domain and a β-sheet region with a distinct loop region in the C-terminus. This loop forms critical contacts with the ExoN domain of nsp14, not only essential for stability of both components but also for full activity of nsp14, likely by causing conformational changes within nsp14. nsp10 also has two zinc-finger binding motifs, providing structural integrity to the protein (Minskaia et

al., 2006; M. Bouvet et al., 2012; Ferron et al., 2018; Kirchdoerfer and Ward, 2019).

2.2 Aims

The aim of this chapter is to complete a biochemical characterisation of the SARS-CoV-2 nsp14-nsp10 complex, detailing its activity profiles and substrate specificities. Results will be outlined in Chapter 7. Very little research has been done on this complex thus far, rendering these observations essential for better understanding of the virus' functioning.

3
Materials and Methods

3.1 Protein Purification methods

3.1.1 C15orf41-Codanin-1 purifications

For the C15orf41-Codanin-1 project, proteins were purified with the help of A. Gavard at the Centre for Medicines Discovery at the University of Oxford using following method. Codon-optimised full-length C15orf41 was cloned into a pNIC28-Bsa4 expression vector containing a Histidine (His_6) tag in a 22-amino acid N-terminal fusion peptide with a recombinant tobacco etch virus (rTEV) protease cleavage site (see map in Appendix A). A C-terminal fragment of Codanin-1 (amino acids 866-1227) was cloned into a pNIC-ZB expression vector with N-terminal His_6 and Z-basic (ZB) tags followed by a rTEV protease cleavage site (see map in Appendix A). All proteins were generated by transformation into *E. coli* BL21 Rosetta2 cells and expression in Terrific Broth media supplemented with 30 mg/ mL kanamycin and 25 mg/ mL chloramphenicol. The cultures were incubated with shaking (200 rpm) at 37°C until an OD_{600} of 2.0 was reached. Cultures were then transferred to an incubator at 18°C for 30 min before induction with 0.1 mM Isopropyl β-D-1-thiogalactoyranoside (IPTG) and incubated for a further 18 hours with shaking (200 rpm).

Cells were harvested by centrifugation at 4000 x g for 30 min at 4°C. The cell pellets were resuspended in lysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 10 mM imidazole, 5% v/v glycerol and 1 mM tris(2-carboxyethyl)phosphine (TCEP) and lysed by sonication for 20 min (5 sec on, 10 sec off) with intermediate stirring. The lysates were clarified by centrifugation at >20,000 rpm for 30 min. The

supernatants were applied to pre-washed nickel resin (Ni-NTA resin, GE healthcare) and incubated on a rotator at 4°C for minimum 2 hours.

The resin was washed twice with lysis buffer then poured onto a gravity column and further washed with wash buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 30 mM imidazole, 5% Glycerol, 1 mM TCEP). Proteins were eluted in elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 300 mM imidazole, 5% glycerol and 1 mM TCEP). All protein-containing fractions were then dialysed overnight at 4°C in dialysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, 5% glycerol). Samples containing Codanin-1 were supplemented with rTEV protease for cleavage of the N-terminal His_6 tag.

In order to form the C15orf41-Codanin-1 complex, a nickel rebind step was performed where His_6-C15orf41 was immobilised on the nickel affinity column and an excess of untagged Codanin-1 added to the same column to allow for complex formation on the column. To this end, His6-C15orf41 was loaded onto nickel beads on a gravity column, washed twice with low salt buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5% Glycerol, 1 mM TCEP) before adding untagged Codanin-1 to the same column to allow complex formation. The column was then washed with low salt buffer (50 mM HEPES pH 7.5 150 mM NaCl, 5% Glycerol, 1 mM TCEP) followed by low salt wash buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5% Glycerol, 1 mM TCEP, 30 mM imidazole).

The complex was co-eluted in low salt elution buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 1 mM TCEP and 300 mM imidazole). The purified complex was then concentrated to 1 ml using a 50 kDa MWCO centrifugal concentrator and further purified on to a Superdex 200 increase 10/300 GL column equilibrated with gel filtration buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP and 5% Glycerol) at 0.5 mL/min. Fractions containing the purified complex were collected, concentrated and snap frozen in liquid nitrogen for later use.

For single protein purification, proteins were immobilised on a nickel affinity chromatography, eluted in elution buffer, concentrated and run on a gel filtration

Superdex 200 increase 10/300 GL column for further purification using the same buffers as indicated above.

The presence of all proteins was confirmed via SDS-PAGE, and ESI-Q-TOF (electrospray-ionisation quadrupole time-of-flight) mass spectroscopy (MS).

The following proteins and protein complexes were purified:

1. *Wild-type full-length C15orf41 with a N-terminal His_6 tag (C15orf41 or C15orf41^{WT})*
   It should be noted that three solubility mutations were added to C15orf41 (W237R, C223S and V18M) in order to enhance solubility and purification yield. These mutations were incidentally discovered by Dr. Opher Gileadi, Centre for Medicines Discovery, in test expression experiments. This construct is referred to as wild-type C15orf41.
2. *Truncated, C-terminal fragment of Codanin-1 (Codanin-1)*
   A C-terminal fragment of Codanin-1 (amino acids 866-1227) was purified with N-terminal His6 and Z-basic (ZB) tags. This construct is referred to as Codanin-1 throughout this study.
3. *Wild-type C15orf41-Codanin-1 complex (C15orf41-Codanin-1)*
   Full length His_6-C15orf41 was co-purified with a C-terminal fragment of Codanin-1 (amino acids 866-1227) to obtain a pure, stable complex.
4. *'Nuclease-dead' C15orf41-Codanin-1 complex (C15orf41^{D196A,E214A}-Codanin-1)*
   Full length His_6-C15orf41 with two point mutations in the predicted nuclease domain, namely D196A and E214A, was co-purified with a C-terminal fragment of Codanin-1 to obtain a predicted catalytically inactive or 'nuclease-dead' version of the complex.
5. *Patient mutant Y94C C15orf41 (C15orf41^{Y94C})*
   Full length His_6-C15orf41 with the patient point mutation Y94C was purified.
6. *Patient mutant L178Q C15orf41 (C15orf41^{L178Q})*
   Full length His_6-C15orf41 with the patient point mutation L178Q was purified.
7. *Patient mutant C15orf41^{Y94C}-Codanin-1 complex (C15orf41^{Y94C}-Codanin-1)*
   Full length His_6-C15orf41 with the patient point mutation Y94C was co-purified with the C-terminal fragment of Codanin-1.

8. *Patient mutant C15orf41^{L178Q}-Codanin-1 complex (C15orf41^{L178Q}-Codanin-1)*
   Full length His_6-C15orf41 with the patient point mutation L178Q was co-purified with the C-terminal fragment of Codanin-1.

3.1.2 nsp14-nsp10 purification

For the nsp14-nsp10 project, proteins were purified by Dr. Y. Yosaatmadja at the Centre for Medicines Discovery and Dr. A. Cruz-Migoni at the MRC Weatherall Institute of Molecular Medicine at the University of Oxford as part of our collaboration, using the method outlined next. N-terminally His-tagged nsp14-nsp10 protein complex was expressed as a bi-cistronic mRNA in which nsp14 bearing a N-terminal Histidine tag is followed by a ribosome binding site ORF (see map in Appendix A). The protein was generated by transformation into *E. coli* BL21 Rosetta2 cells and expression in Terrific Broth media supplemented with 30 mg/ mL kanamycin and 10 mM ZnCl2. The culture was incubated with shaking (200 rpm) at 37°C until an OD_{600} of 2.0 was reached. Cultures were then transferred to an incubator at 18°C for 30 min before induction with 1 mM Isopropyl β-D-1-thiogalactoyranoside (IPTG) and incubated for 18 hours with shaking (200 rpm).

Cells were harvested by centrifugation at 6000 x g for 30 min at 4°C. The cell pellet was resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, 5% v/v glycerol and 1 mM tris(2-carboxyethyl)phosphine (TCEP) and lysed by sonication. The lysate was clarified by centrifugation at 40,000 x g for 30 min, and the supernatant was loaded onto an equilibrated (lysis buffer) immobilised metal affinity chromatography column (IMAC) (Ni-IDA Sepharose, GE healthcare).

The immobilised protein was washed with lysis buffer and eluted with elution buffer (50 mM Tris-HCl pH 8.0, 500 mM NaCl, 300 mM imidazole, 5% v/v glycerol and 1 mM TCEP). The protein-containing fractions were pooled and dialysed overnight at 4°C in dialysis buffer (25 mM Tris-HCl pH 8.0, 150 mM NaCl, 5% v/v glycerol, and 1 mM TCEP) supplemented with recombinant tobacco etch virus (rTEV) protease for cleavage of the N-terminal His_6 tag. The protein was then

subjected to a second IMAC step to remove His-tagged rTEV protease, cleaved His tag, and uncleaved His-tagged nsp14-nsp10 complex.

The cleaved nsp14-nsp10 complex was concentrated to 1 mL using a 50 kDa MWCO centrifugal concentrator. The protein was then further purified on to a Superdex 200 increase 10/300 GL column equilibrated with SEC buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 5% v/v glycerol, 2 mM TCEP) at 0.8 mL/min. The purified protein was concentrated to 0.5 mg/ mL and was snap frozen in liquid nitrogen for later use. Monomeric nsp14 was purified from the same vector under different induction conditions (overnight growth without adding IPTG).

The presence of the nsp14-nsp10 complex was confirmed via SDS-PAGE, and ESI-Q-TOF (electrospray-ionisation quadrupole time-of-flight) mass spectroscopy (MS).

The following proteins and protein complexes were generated:

1. *Wild-type full-length nsp14 referred to as nsp14*
2. *Wild-type nsp14-nsp10 (nsp14-nsp10)*

   Full-length nsp14 complex with full-length nsp10.
3. *'Nuclease-dead' nsp14^{D113A,E115A}-nsp10*

   Full-length nsp14 with two point mutations in the predicted nuclease domain, D113A and E115A, complexed with full-length nsp10 to form a catalytically inactive nsp14-nsp10 complex.

Wild-type and mutant nsp12, nsp7 and nsp8 used in this study were a kind gift from Dr. H. Fan and Dr. J. Keown at the Division of Structural Biology (STRUBI) at the University of Oxford. Following proteins were obtained: Wild-type nsp12, nsp7 and nsp8 as well as two mutant forms of nsp12: nsp12^{D126A} and nsp12^{D670A,D761A}. nsp12^{D126A} is a NiRan domain active site mutant and nsp12^{D670A,D761A} a polymerase active site double mutant.

3.2 Biochemistry methods

3.2.1 Generation of 5′ radiolabelled substrates

10 pmol of single-stranded RNA or DNA (Eurofins MWG Operon, Germany) were incubated with 6.8 pmol γ-^{32}P-ATP (Perkin Elmer), and 10 U T4 PNK (ThermoFisher Scientific) at 37°C for 1 hour. This solution was passed through a P6 Micro Bio- Spin chromatography column (BioRad) for 4 min at 1000 x rcf to remove unbound label and flow-through diluted accordingly in nuclease-free ultrapure H_2O. For all assays, a final substrate concentration of 10 nM was used unless indicated otherwise.

3.2.2 Substrate Annealing

For double-stranded structures, radiolabelled RNA or DNA was annealed to unlabelled complementary molecules at a 1 : 1.5 ratio to give a final concentration of 100 nM in annealing buffer (10 mM Tris-HCl (pH 7.5), 50 mM NaCl and 1 mM EDTA) by heating the sample to 95°C for 3 min before cooling to room temperature.

A detailed list of all oligonucleotide sequences and complex substrates used is provided in Tables 3.1, 3.2, 3.3 and 3.4. For each oligonucleotide, the minimum free energy was computed using RNAFold 2.4.7 as part of the ViennaRNA suite to confirm no unwanted secondary structure formation (Lorenz et al., 2011).

3.2.3 Generation of a single nucleotide RNA ladder

10 pmol of 5′ radiolabelled single-stranded RNA was incubated in 50 mM potassium acetate (pH 7.9), 20 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate (pH 7.9) and 1 mM DTT for 2, 5, 10, and 15 min at 95°C respectively. 5 µL of stop solution (95% formamide, 10 mM EDTA, 0.25% xylene cyanol, 0.25% bromophenol blue) was added to each sample, samples were then combined and run on a 20% denaturing polyacrylamide gel (40% solution of 19:1 acrylamide:bis-acrylamide, BioRad) and 7 M urea (Sigma Aldrich)) in 1 x TBE (Tris-borate EDTA) buffer. Electrophoresis was carried out at 525 V for 1.5 hours; gels were subsequently fixed for 60 min in fixing solution (50% methanol, 10% acetic acid), and dried at 80°C for two hours

Table 3.1: List of RNA and DNA substrates used for C15orf41-Codanin-1 nuclease assays

Each nucleotide sequence is numbered and a general description included. Minimum free energy (MFE, in kcal/mol) was calculated using RNAFold 2.4.7 as part of the ViennaRNA suite to determine the ability of oligonucleotides to form secondary structures.

Name	Structure	Sequence (5′ to 3′)	MFE
1	10mer RNA	AUA AUU AGA U	0
2	10mer RNA	AUC AAA UUC U	0
3	10mer RNA with 3′-Phosphate	AUA AUU UGA U [PHO]	0
4	10mer RNA with 3′-Biotin	AUA AUU UGA U [BIO]	0
5	20mer RNA	AUA AUU UGA UCA UCU AUU AU	-0.17
6	20mer RNA	AUA AUA GAU GAU CAA AUU AU	-0.27
7	20mer RNA with 3′-Phosphate	AUA AUU UGA UCA UCU AUU AU [PHO]	-0.17
8	20mer RNA with 3′-Biotin	AUA AUU UGA UCA UCU AUU AU [BIOTEG]	-0.17
9	30mer RNA	GUU GAG AGA GAG AGA GUU UGG AGA GUU UGU	0
10	30mer RNA	ACA AAC UCU CCA AAC UCU CUC UCU CUC AAC	0
11	30mer RNA with 3′-Phosphate	GUU GAG AGA GAG AGA GUU UGG AGA GUU UGU [PHO]	0
12	30mer RNA with 3′-Biotin	GUU GAG AGA GAG AGA GUU UGG AGA GUU UGU [BIOTEG]	0
13	10mer DNA	ATC AAA TTA T	0
14	10mer DNA with 3′-Phosphate	ATC AAA TTA T [PHO]	0
15	10mer DNA with 3′-Biotin	ATC AAA TTA T [BIO]	0
16	20mer DNA	ATA ATA GAT GAT CAA ATT AT	-0.27
17	20mer DNA with 3′-Phosphate	ATA ATA GAT GAT CAA ATT AT [PHO]	-0.27
18	20mer DNA with 3′-Biotin	ATA ATA GAT GAT CAA ATT AT [BIO]	-0.27
19	30mer DNA	ACA AAC TCT CCA AAC TCT CTC TCT CTC AAC	0
20	30mer DNA with 3′-Phosphate	ACA AAC TCT CCA AAC TCT CTC TCT CTC AAC [PHO]	0
21	30mer DNA with 3′-Biotin	ACA AAC TCT CCA AAC TCT CTC TCT CTC AAC [BIO]	0
22	20mer poly(A) RNA	AAA AAA AAA AAA AAA AAA AA	0
23	20mer poly(A) DNA	AAA AAA AAA AAA AAA AAA AA	0
24	18mer DNA with a single ribonucleotide	GAT CTG AGC CTG GG[A] GCT	-1.3
25	18mer DNA	AGC TCC CAG GCT CAG ATC	-2.1
26	10mer RNA-13mer DNA	[UGC CUG CAG G]- TCG ACT CTA GAG G	-3.09
27	40mer DNA	CC TCT AGA GTC GAC CTG CAG GCA TGC AAG CTT TTG TTC CC	-5
28	50mer DNA	TCT TCA CAC TCT TCC GTA GGA CCG ACG GAT TTG GTG ACC CTC AGC TAA TT	-3.74
29	50mer DNA	TCT TCA CAC TCT TCC GTA GGA CCG ACG GAT TTG GTG ACC CTC AGC TAA TT	-3.74
30	10mer DNA	CCG TCG GTC C	-0.05
31	10mer RNA	CCG UCG GUC C	-0.05
32	50mer RNA	UCU UCA CAC UCU UCC GUA GGA CCG ACG GAU UUG GUG ACC CUC AGC UAA UU	-5.74
33	49mer DNA	ATA AAT ATT TTT TAT TAA TAA TAG ATC ACC TTT CTT TCT CTT CTC CCC T	-1.19
34	40mer DNA	TTC CCC TCC TCT CCT TCC TTC CTG ATC TAT TAT TAA TAA AAA ATA TTT A	-0.31
35	22mer DNA	AAG GGG AGA AGA GAA AGA AAG G	0
36	23mer DNA	GGA AGG AAA GGA GAG GAG GGG AA	0
37	20mer RNA with a 2-methyladenosine	AUA AUU UGA UCA UCU AUU A*U	-0.17
38	20mer RNA with a N6-methyladenosine	AUA AUU UGA UCA UCU AUU A*U	-0.17
39	20mer RNA with an inosine	AUA AUU UGA UCA UCU AUU IU	-0.17
40	22mer RNA- 28mer DNA	[UUC CCC UCC UCU CCU UCC UUC C]- TGA TCT ATT ATT AAT AAA AAA TAT TTA T	-0.39

Table 3.2: Codes to generate complex RNA and DNA structures used for C15orf41-Codanin-1 nuclease assays
To generate complex RNA and DNA structures, single-stranded oligos were annealed as designated by the numbers using the protocol described in the Materials and Methods.

Structure	Code
10mer RNA-DNA hybrid	1 + 13
10mer double-stranded RNA	1 + 2
20mer RNA-DNA hybrid	5 + 16
20mer double-stranded RNA	6 + 7
30mer RNA-DNA hybrid	9 + 19
30mer double-stranded RNA	9 + 10
R-loop structure	28 + 29 + 31
D-loop structure	28 + 29 + 30
RNA loop structure	31 + 32
Ribonucleotide structure	24 + 25
Okazaki fragment flap structure	26 + 27
DNA fork	33 + 34
DNA fork with leading strand	33 + 34 + 35
DNA fork with lagging strand	33 + 34 + 36
Replication fork	33 + 34 + 35 + 36
RNA-DNA hybrid with m2A	37 + 16
RNA-DNA hybrid with m6A	38 + 16
RNA-DNA hybrid with Inosine	39 + 16

under a vacuum. The dried gels were exposed to a Kodak phosphorimager screen and scanned using a Typhoon 9400 instrument (GE).

3.2.4 Nuclease assays

3.2.4.1 C15orf41-Codanin-1 nuclease assays

Nuclease assays were carried out in 10 µL reactions containing 50 mM HEPES pH 7.5, 50 mM NaCl, 5 mM MgCl2, 5% glycerol, 0.5 mM DTT, 10 nM RNA or DNA substrate and serially diluted two-fold C15orf41-Codanin-1 complex or individual proteins from 500 nM unless indicated otherwise. Reactions were incubated at 37°C for 60 min, and quenched by the addition of 5 µL stop solution (95% formamide

Table 3.3: List of RNA and DNA substrates used for nsp14-nsp10 nuclease assays
Each nucleotide sequence is numbered and a general description included. Minimum free energy (MFE in kcal/mol) was calculated using RNAFold 2.4.7 as part of the ViennaRNA suite to determine the ability of oligonucleotides to form secondary structures.

Name	Description	Sequence (5′ to 3′)	MFE
1	10mer RNA	AUA AUU AGA U	0
2	20mer RNA	AUA AUU UGA UCA UCU AUU AU	-0.17
3	20mer RNA	AUA AUA GAU GAU CAA AUU AU	-0.27
4	20mer RNA	AUC UAA ACG AAC AAA CUA AA	-0.05
5	20mer RNA	CAA CAG UUC AAG AAA UCA AA	-0.01
6	20mer RNA	AUA AUA GAU GAU CAA AUU AG	-0.19
7	20mer RNA	AUA AUA GAU GAU CAA AUU CG	-0.53
8	20mer RNA	AUA AUA GAU GAU CAA AUG CG	-0.07
9	20mer RNA	AUA AUA GAU GAU CAA ACG CG	-0.06
10	20mer RNA	AUA AUA GAU GAC CAA AUU AU	-0.18
11	20mer RNA	UUU UUU UUU UUU UUU UUU UU	0
12	20mer RNA	AAA AAA AAA AAA AAA AAA AA	0
13	20mer RNA	CCC CCC CCC CCC CCC CCC CC	0
14	20mer RNA	GGG GGG GGG GGG GGG GGG GG	0
15	20mer RNA with 3'-Phosphate	AUA AUU UGA UCA UCU AUU AU [PHO]	-0.17
16	20mer RNA with 3'- Biotin	AUA AUU UGA UCA UCU AUU AU [BIOTEG]	-0.17
17	30mer RNA	GUU GAG AGA GAG AGA GUU UGG AGA GUU UGU	-0.41
18	20mer DNA	ATA ATA GAT GAT CAA ATT AT	-0.27
19	18mer DNA with a single ribonucleotide at position 15	GAT CTG AGC CTG GG[A] GCT	-1.3
20	20mer RNA with a 2-methyladenosine at position 19	AUA AUU UGA UCA UCU AUU A*U	-0.17
21	20mer RNA with a N6-methyladenosine at position 19	AUA AUU UGA UCA UCU AUU A*U	-0.17
22	20mer RNA with an inosine at position 19	AUA AUU UGA UCA UCU AUU I*U	-0.17
23	20mer RNA	GUC AUU CUC CUA AGA AGC UA	-1

Table 3.4: Codes to generate complex RNA and DNA structures used for nsp14-nsp10 nuclease assays
To generate complex RNA and DNA structures, single-stranded oligos were annealed as designated by the numbers using the protocol described in the Materials and Methods.

Structure	Code
RNA-DNA hybrid	2 + 18
double-stranded RNA	2 + 3
double-stranded RNA with a 2-methyladenosine on the top strand	20 + 3
double-stranded RNA with a N6-methyladenosine on the top strand	21 + 3
double-stranded RNA with a inosine on the top strand	22 + 3
double-stranded RNA with one terminal mismatch	6 + 2
double-stranded RNA with two terminal mismatch	7 + 2
double-stranded RNA with three terminal mismatch	8 + 2
RNA Flap	9 + 2
double-stranded RNA with one internal mismatch at position 12	10 + 2

(v/v), 10 mM EDTA, 0.25% xylene cyanol, 0.25% bromophenol blue) and boiling at 95°C for 5 min.

Reactions were then analysed by 20% denaturing polyacrylamide gel electrophoresis (40% solution of 19:1 acrylamide:bis-acrylamide, BioRad) and 7 M urea (Sigma Aldrich)) in 1 x TBE (Tris-borate EDTA) buffer. Electrophoresis was carried out at 525 V for 1.5 hours; gels were subsequently fixed for 60 min in fixing solution (50% methanol, 10% acetic acid), and dried at 80°C for two hours under a vacuum. The dried gels were exposed to a Kodak phosphorimager screen and scanned using a Typhoon 9400 instrument (GE).

3.2.4.2 Nsp14-nsp10 nuclease assays

Nuclease assays were carried out in 10 µL reactions containing 25 mM HEPES-KOH, pH 7.5, 50 mM NaCl, 5 mM MgCl2, 5% glycerol, 1.0 mM DTT, 10 nM RNA or DNA substrate and serially diluted two-fold nsp14-nsp10 complex or nsp14 from 500 nM unless indicated otherwise. Reactions were incubated at 37°C for 45 min, and quenched by the addition of 5 µL stop solution (95% formamide (v/v), 10 mM EDTA, 0.25% xylene cyanol, 0.25% bromophenol blue) and boiling at 95°C for 5 min.

Reactions were then analysed by 20% denaturing polyacrylamide gel electrophoresis (40% solution of 19:1 acrylamide:bis-acrylamide, BioRad) and 7 M urea (Sigma Aldrich)) in 1 x TBE (Tris-borate EDTA) buffer. Electrophoresis was carried out at 525 V for 1.5 hours; gels were subsequently fixed for 60 min in fixing solution (50% methanol, 10% acetic acid), and dried at 80°C for two hours under a vacuum. The dried gels were exposed to a Kodak phosphorimager screen and scanned using a Typhoon 9400 instrument (GE).

Nuclease assays with the addition of the nsp12-7-8 complex (as indicated) were carried out as above, with the following modifications. To facilitate complex formation, 200 nM of nsp12, 600 nM nsp7, and 600 nM of nsp8 (a molar ratio of 1:3:3) were incubated together with the relevant RNA substrate (10 nM) for 5 min at 37°C before the nuclease reaction was started with the addition of nsp14-nsp10. Nsp14-nsp10 concentrations (unless stated otherwise) were maintained at 100 nM.

Where applicable, gel-based nuclease assays were quantified by analysing the product formation (digestion) of the radiolabelled RNA substrate(s). Gel images collected on a Typhoon scanner were analysed in Image J to determine the amount of substrate remaining in the well (undigested) in comparison to the amount of substrate that entered the gel (product) reported as a percent digested. This was plotted versus protein concentration to indicate digestion rate. Error is standard error. Three or more gels were analysed for each substrate.

3.2.5 Electrophoretic Mobility Shift Assay (EMSA)

EMSA reactions were carried out in 10 µL reactions containing 50 mM HEPES pH 7.5, 50 mM NaCl, 5% Glycerol, 0.5 mM DTT, 50 mM L- arginine, 0.01% Tween20 and 1 nM RNA or DNA substrate with either 20 mM Mg^{2+} or 5 mM Ca^{2+} and serially diluted two-fold C15orf41-Codanin-1 complex or individual proteins from 500 nM unless indicated otherwise. The reaction was incubated on ice for 15 min before the addition of 5 µL ice-cold loading buffer (50% glycerol, 1 mg/mL bromophenol blue).

Reactions were then analysed by 6% non-denaturing polyacrylamide gel electrophoresis in 0.5 x TBE (Tris-borate EDTA) buffer. Electrophoresis was carried out at 4°C at 100 V for 3 hours in 0.5 x TBE; gels were subsequently fixed for 10 min in fixing solution (50% methanol, 10% acetic acid), and dried at 80 oC for two hours under a vacuum. The dried gels were exposed to a Kodak phosphorimager screen and scanned using a Typhoon 9400 instrument (GE).

3.2.6 SEC-MALS

Size-exclusion chromatography multiple angle light scatter (SEC-MALS) was performed by Dr. D. Staunton at the Department of Biochemistry at the University of Oxford. Briefly, C15orf41-Codanin-1 complex as well as individual proteins were purified as described above. Purified proteins were then serially diluted from 2 mg/mL to 0.125 mg/mL in 25 mM HEPES pH 7.5, 150 mM NaCl, 5% Glycerol and 2 mM TCEP and run on a Superose 6 HR10/30 analytic column coupled to a Wyatt

Dawn HELEOS-II 8-angle light scattering detector and Wyatt Optilab rEX refractive index monitor. Data was collected and then analysed using the ASTRA v6 software.

3.2.7 SAXS

Small angle X-ray scattering (SAXS) was performed by Dr. Y. Yosaatmadja at the Centre for Medicines Discovery at the University of Oxford as part of our collaboration and data collection undertaken at the Diamond light source at the Harwell Science and Innovation Campus in Oxford. In brief, C15orf41-Codanin-1 complex as well as individual proteins were purified as described above. Purified proteins were then serially diluted from 4 mg/mL to 0.125 mg/mL in 25 mM HEPES pH 7.5, 100 mM NaCl, 1% Glycerol and 1 mM TCEP. Protein-buffer solutions were aliquoted into a 96-well plate and mounted on the beamline and held at 283 K for auto sampling and capillary flow data acquisition. The SAXS camera length was set to 3340 mm and images taken at 1 s intervals of X-ray exposure with a sample flow rate of 4 µL/s using the ScatterBrain software package. Matched buffer samples were run as a control to account for background noise. The raw data then analysed using the ATSAS (2.7.1) suite of software packages.

3.2.8 Negative-staining electron microscopy (neg-staining EM)

Negative-staining EM was performed by Dr. J. Gurnett at the MRC Weatherall Institute of Molecular Medicine as part of our collaboration. In brief, C15orf41-Codanin-1 complex as well as individual proteins were purified as described above. Purified C15orf41-Codanin-1 complex was serially diluted in 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP and 5% Glycerol to reach final concentrations between 0.03 mg/mL and 0.06 mg/mL. Diluted protein samples were then applied to glow discharged carbon film 300 mesh grids (TAAB laboratories). Grids were then washed with the dilution buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP and 5% Glycerol) and stained using 2% (w/v) uranyl acetate. Grids were then visualised and screened on a FEI Technai 12 TEM with a 16-megapixel Gatan OneView

camera operated at 120 kV in the Sir William Dunn School of Pathology Bioimaging Facility, University of Oxford. In general, 200 micrographs were collected per grid at a magnification of 98,000x, with a pixel size of 1.2 Å/pixel and used as the input for data processing performed by Dr. J. Gurnett. In summary, particles were manually picked using the built in RELION3.1 particle picker and several rounds of 2D and 3D class averages were generated using RELION3.1. In total, >10000 particles were classified into 2D and 3D averages and further refined into a final 3D model.

3.3 Cell Biology methods

3.3.1 Tissue Culture

Human HEK 293-FT cells (transformed human embryonic kidney cells expressing SV-40 large T-antigen) used in this thesis were a kind gift from Dr. L. Swift, MRC Weatherall Institute of Molecular Medicine. Cells were grown in DMEM (Gibco, Thermo Scientific) media supplemented with 10% (v/v) fetal calf serum and L-glutamine and no antibiotics and incubated at 37°C and 5% CO2. When appropriate, cells were harvested using Trypsin solution (Sigma Aldrich). Cells were regularly tested for mycoplasma and general tissue culture good practice respected.

HUDEP-2 (Human umbilical cord derived progenitors-2) cells used in this thesis were grown by A-A. Olijnik and Dr. C. Scott at the MRC Weatherall Institute of Molecular Medicine as part of our collaboration. HUDEP-2 cells have been engineered by extracting CD34+ cells from umbilical cord blood and immortalising them by transduction of the Tetracycline-inducible (Tet) Human PapillomaVirus 16-E6/E7 oncoprotein cassette (HPV16-E6/E7). HUDEP-2 cells were maintained as proerythroblasts at a density of 0.5x106 cells/mL to 2x106 cells/mL at 37°C and 5% CO2. Every 48 hours, cells were counted, washed with PBS and resuspended in “expansion media” (StemSpan SFEM) supplemented with L-Glutamine (2mM), Penicillin/Streptomycin (100 U Penicillin and 100 µg/mL Streptomycin), human Stem Cell Factor (SCF) (50 ng/mL), erythropoietin (3 IU/mL) and Dexamethasone (330 µg/L). Doxycycline was added fresh to the culture at the time of passaging, at a final concentration of 2 µg/mL, due to the low stability of the compound.

Following HUDEP-2 cell lines were available during this study:

1. Wild-type HUDEP-2 cells *(WT HUDEP-2)*
2. HUDEP-2 cells with a N-terminal endogenous 3xFLAG tag in the C15orf41 locus *(3xFLAG-C15orf41)*
3. HUDEP-2 cells with a C-terminal endogenous STREP tag in the Codanin-1 locus *(STREP-Codanin-1)*
4. HUDEP-2 cells with a N-terminal endogenous 3xFLAG tag in the C15orf41 locus and a L178Q patient mutant missense mutation in C15orf41 *(3xFLAG-C15orf41^{L178Q})*
5. HUDEP-2 cells with a C-terminal endogenous STREP tag in the Codanin-1 locus and a R1042W patient mutant missense mutation in Codanin-1 *(STREP-Codanin-1^{R1042W})*

All above endogenously-tagged cell lines were engineering by A-A. Olijnik and Dr. C. Babbs at the MRC Weatherall Institute of Molecular Medicine using the CRISPR-Cas9 gene editing technology. After CRISPR-Cas9 editing, positively targeted clones have been differentiated through the erythroid pathway through a well-defined protocol, as per Scott et al., 2019. All HUDEP-2 cell lines were differentiated to produce mature late erythroblasts when indicated.

3.3.2 Transfections

The plasmids below were transfected into HEK 293-FT cells using Lipofectamine 2000 transfection reagent (Thermo Fisher) according to the manufacturer's manual. In brief, 5x105 HEK203-FT cells/well were plated in a 6-well plate and incubated overnight at 37°C and 5% CO2. 4 µg of plasmid DNA was incubated with Lipofectamine 2000 in OptiMEM media at room temperature for 30 min then added to the cells. Cells were incubated for minimum 2 hours at 37°C and 5% CO2 before media was changed and fresh media was added. Whole cell extracts were harvested 48 hours after and analysed by western blotting. Following plasmids were a kind gift from Dr. C. Babbs:

1. Wild-type FLAG-HA C15orf41-GFP in VPN5
2. FLAG-HA in VPN5

3.3.3 Immunoprecipitations (IP)

HUDEP-2 cells ($5x10^6$ containing 3xFLAG-C15orf41) or transfected HEK293-FT cells with the plasmids described above were harvested and cell pellets washed twice with ice-cold PBS. Cells were then lysed on ice for 45 min using lysis buffer (50 mM Tris (pH 7.5), 150 mM NaCl, protease inhibitor (cOmplete mini, EDTA-free, Sigma), 2 mM Mg^{2+}, 5 U Benzonase (Sigma), 0.5% NP-40 and 5% Glycerol). After 45 min on ice, 1 mM EDTA was added to each sample and samples incubated on ice for a further 15 min.

Once lysis is complete, samples were centrifuged at >20000 x g for 20 min at 4°C. The supernatant was applied to 100 µL of pre-washed FLAG-M2 magnetic beads (Sigma, M8823) and incubated on a rotator at 4°C for minimum 2 hours. This was followed by washing steps using wash buffer (50 mM Tris (pH 7.5), 1 mM EDTA, protease inhibitor (cOmplete mini, EDTA-free, Sigma), 0.5% NP-40 and either 150 mM, 250 mM or 500 mM NaCl as indicated).

Immuno-precipitated protein was eluted using 3xFLAG peptide (F4799, Sigma) in TBS buffer (50 mM Tris (pH 7.5), 150 mM NaCl). To this end, 6 µL of 5 µg/µL 3xFLAG peptide was added to 200 µL of TBS buffer and incubated with the beads on a rotator at 4°C for minimum 30 min and the supernatant collected. Alternatively, beads were mixed with loading buffer and reducing agent according to the manufacturer's guidelines (Thermo Scientific NuPAGE Sample buffer Kit) and boiled for 5 min at 95°C. Samples were then analysed by western blotting.

Where indicated, 3xFLAG-C15orf41^{L178Q} HUDEP-2 cells were used to perform IPs using the protocol above. WT HUDEP-2 cells were used as a control.

3.3.4 Western Blotting, protein transfers and ECL detection

Whole cell extracts of either HEK 293-FT or HUDEP-2 cells were used to perform Western blot experiments. Cells were centrifuged for 5 min at 2400 rpm at 4°C,

pellets washed with phosphate buffered saline (PBS) solution and lysed on ice for 45 min using lysis buffer (50 mM Tris (pH 7.5), 150 mM NaCl, protease inhibitor (cOmplete mini, EDTA-free, Sigma), 2 mM Mg^{2+}, 5 U Benzonase, 0.5% NP-40 and 5% Glycerol). After 45 min on ice, 1 mM EDTA was added to each sample and samples incubated on ice for a further 15 min. Once lysis is complete, samples were centrifuged at >20000 x g for 20 min at 4°C. The supernatant was collected and mixed with loading buffer and reducing agent according to the manufacturer's guidelines (Thermo Scientific NuPAGE Sample buffer Kit) and boiled for 5 min at 95°C. Samples were then loaded onto a 4-12% Bis-Tris SDS-PAGE gel (Thermo Scientific) and run at 120 V for 90 min in 1x MOPS buffer.

Protein gels were then transferred onto a PVDF membrane at 4°C at 30 V for 18 hours in 1x transfer buffer (25 mM Tris-HCl (pH 7.5), 192 mM glycine and 20% methanol). The membrane was then blocked for 1 hour in PBS with 5% milk on a shaker at room temperature. The membrane was blotted with the appropriate antibody (as indicated in Table 3.5) overnight at 4°C followed by washing for 3x10 min in PBS with 0.2% Tween. The membrane was then incubated with the appropriate HRP-conjugated secondary antibody (see Table 3.5) for 90 min on a shaker at room temperature and then washed for 3x10 min in PBS with 0.2% Tween before being visualised with ECL western blotting substrate kit (Thermo Fisher) as per manufacturer's instructions.

Table 3.5: List of antibodies used during western blotting experiments
A general description of all antibodies used in this study is provided in this table.

Antibody	Code	Make	Species	Dilution
Actin-HRP	A3854	Sigma	Mouse	1:10000
C15orf41	ab107352	Abcam	Rabbit	1:500
C15orf41	PA897474LA01HU	Cusabio	Rabbit	1:500
Codanin-1	A304-951A	Bethyl	Rabbit	1:100
FLAG	F1804	Sigma	Mouse	1:2000
GFP	11814460001	Roche	Mouse	1:1000
STREP	ab76949	Abcam	Rabbit	1:1000
STREP	ab184224	Abcam	Mouse	1:1000
STREP	34850	Qiagen	Mouse	1:1000
α-mouse HRP	P0260	Agilent Tech.	-	1:2000
α-rabbit HRP	P0448	Agilent Tech.	-	1:2000

4

Cellular characterisation of C15orf41-Codanin-1

CDA-I is a rare hereditary erythroid disorder characterised by unique morphological abnormalities in bone marrow erythroblasts, most often caused by mutations in *CDAN1*. The recent identification of *C15orf41*, the second disease gene in CDA-I, offers new avenues of research into the molecular mechanism underlying CDA-I.

Very little is known about the cellular function of Codanin-1 or C15orf41, the two CDA-I-associated proteins, both of which are cell-essential. It is therefore crucial to understand the role of Codanin-1 and C15orf41, not only to elucidate their involvement in the disease mechanism of CDA-I but also in a yet unidentified, essential and basic cellular maintenance pathway. The aim of this chapter is therefore to provide some new insights into the cellular roles of C15orf41 and Codanin-1.

4.1 C15orf41 and Codanin-1, the two disease proteins in CDA-I, form a stable complex

Mutations in *C15orf41* and *CDAN1* both cause CDA-I and result in identical erythroid abnormalities. Due to the very specific disease phenotype of CDA-I, which includes the characteristic "spongy" heterochromatin defect, it is extremely likely both factors participate in the same pathway and this possibility explored in this section.

Both published and unpublished work from our laboratory as well as our collaborators have identified an interaction between C15orf41 and Codanin-1 in non-erythroid cell lines. Olijnik et al., 2021 showed a direct interaction between Codanin-1

and C15orf41 using immunoprecipitation experiments coupled with western blotting in HeLa cells. Unpublished data from our collaborators in the Gileadi group in the Centre for Medicines Discovery, University of Oxford identified a possible interaction between C15orf41 and Codanin-1 using immunoprecipitation experiments in which HEK293 whole cell extracts were passed through immobilised synthetic Codanin-1 peptides and the outputs purified and analysed by mass spectrometry. C15orf41 was identified as a binding partner to Codanin-1 and C15orf41's binding region was narrowed down to the C-terminus of Codanin-1 through partial trypsin digestion (Dr. J. Newman, Centre for Medicines Discovery, unpublished).

These findings have since been confirmed in the literature, where recent reports have shown that Codanin-1 and C15orf41 form a tight complex in transformed human cancer cell lines, using over-expressed tagged proteins (Shroff et al., 2020; Swickley et al., 2020).

We therefore sought to determine the interaction between both proteins in an erythroid cell-type at physiologically relevant levels. To confirm published results in our hands, C15orf41-Codanin-1 complex formation was assessed using over-expressed C15orf41 protein in transformed human cancer cell lines in agreement with current reports. Importantly, these results were then validated using physiologically appropriate, erythroid-specific cell lines with endogenous protein levels to eliminate confounding factors such as artifacts of over-expression or lineage-specific caveats and further investigate this interaction.

4.1.1 Complex formation in HEK293-FT cells with over-expressed proteins

C15orf41 and Codanin-1 have been shown to interact in transformed cell lines using over-expressed proteins (Shroff et al., 2020; Swickley et al., 2020; Olijnik et al., 2021). In order to confirm this result in our hands, HEK293-FT cells were grown in DMEM media supplemented with 10% (v/v) fetal calf serum and L-glutamine and no antibiotics and maintained at 37°C and 5% CO2. Cells were then transfected with a VPN5 plasmid containing wild-type C15orf41 with N-terminal FLAG and GFP

tags using Lipofectamine 2000 as per manufacturer's instructions (Figure 4.1.a). Cells were harvested using Trypsin solution and transfection efficiency determined by microscopic observation. Different transfection lengths were employed to assess for optimal expression levels. As shown in Figure 4.1.b, both 2 hours and 4 hours transfection length were sufficient, as seen by presence of a large proportion of green cells, confirming expression of the desired FLAG-GFP-C15orf41 protein.

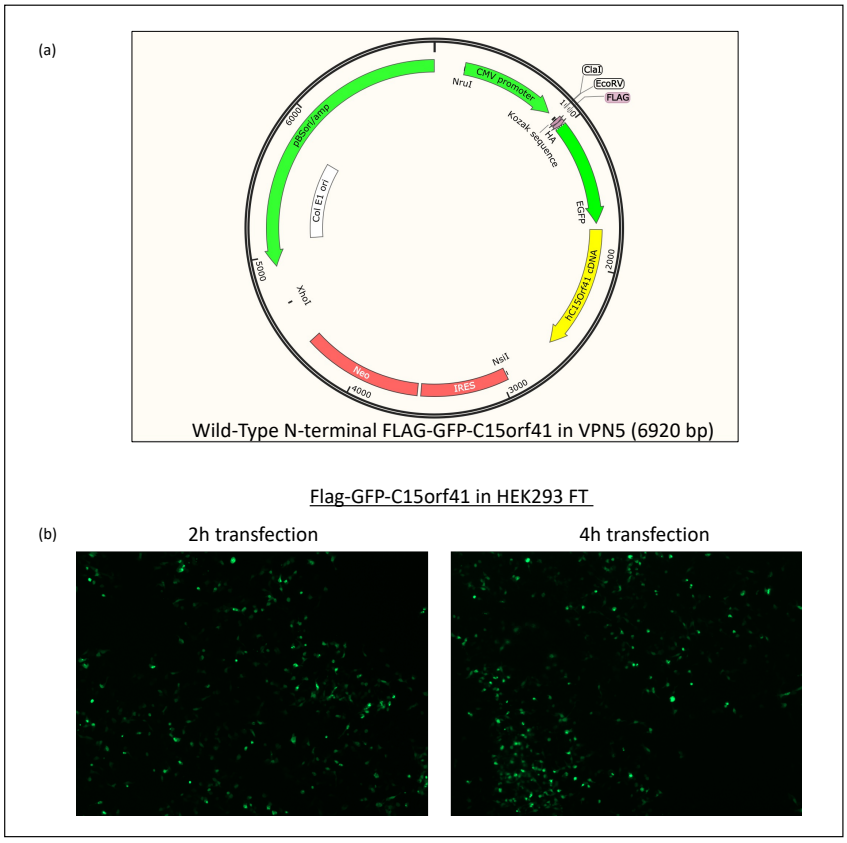


Figure 4.1: Transfected FLAG-GFP-C15orf41 plasmid in HEK293-FT cells
(a) N-terminal FLAG-GFP-C15orf41 (wild-type) in VPN5 plasmid map.
(b) FLAG-GFP-C15orf41 was transfected in HEK293-FT cells and transfection efficiency visualised for both 2 hours and 4 hours of transfection length. In both cases, transfection was successful as shown by the high proportion of visible green cells, confirming efficient update of the plasmid.

Before attempting to perform pull-down experiments with the over-expressed and tagged C15orf41 protein, western-blot experiments were undertaken to determine whether protein was detectable. Because specific C15orf41 antibodies were not commercially available at that time, a primary antibody specific to GFP was used.

Intriguingly, no GFP was detectable in extracts from cells over-expressing FLAG-GFP-C15orf41 (Figure 4.2). Instead, GFP was only detectable in the positive control (consisting of cell extracts transfected with a VPN5 vector expressing only GFP) confirming the result obtained is not due to ineffective Western blotting technique or defective antibody. Untransfected cells were used as a negative control to account for antibody background. It is unclear why GFP could not be identified in cells transfected with the VPN5 vector expressing FLAG-GFP-C15orf41. A possible explanation is that C15orf41 adopts a fold that causes masking of the GFP tag in a way that renders the epitope unavailable for antibody detection. It should also be noted that C15orf41 is a relatively small protein (around 34 kDa) and GFP (around 26 kDa) quite a large tag such that C15orf41 protein misfolding (with associated protein instability and degradation) is also possible.

Since the plasmid used also harbours a FLAG tag, untransfected and transfected HEK293 cells with FLAG-GFP-C15orf41 were blotted using a primary FLAG antibody and samples analysed by Western blotting. In this case, FLAG-GFP-C15orf41 was detected in the transfected samples, both after 2 hours and 4 hours of transfection. Considering the GFP tag remained undetectable, an alternative approach of transfecting HEK 293-FT cells with a plasmid containing a FLAG-C15orf41 construct, was employed. Curiously, when extracts from these cells were analysed alongside those expressing FLAG-GFP-C15orf41 by Western blotting, FLAG signal was not detected in the FLAG-C15orf41 cells, only in the FLAG-GFP-C15orf41 cells (Figure 4.3). Some degradation bands of smaller molecular weight were also detected in the 4h transfection sample, possibly corresponding to cleaved FLAG-GFP tag (expected to be around 29 kDa). The 2h transfection time was therefore chosen as optimal as it yielded near to no visible degradation bands. The reason why FLAG signal was not detected in the FLAG-C15orf41 cells is unclear. The FLAG tag is very small, consisting of an octapeptide with the sequence DYKDDDDK. When directly attached to the N-terminus of C15orf41, it might be masked by the adjacent C15orf41 protein thereby masking the epitope. The GFP tag situated between the terminal FLAG tag and the C15orf41 protein

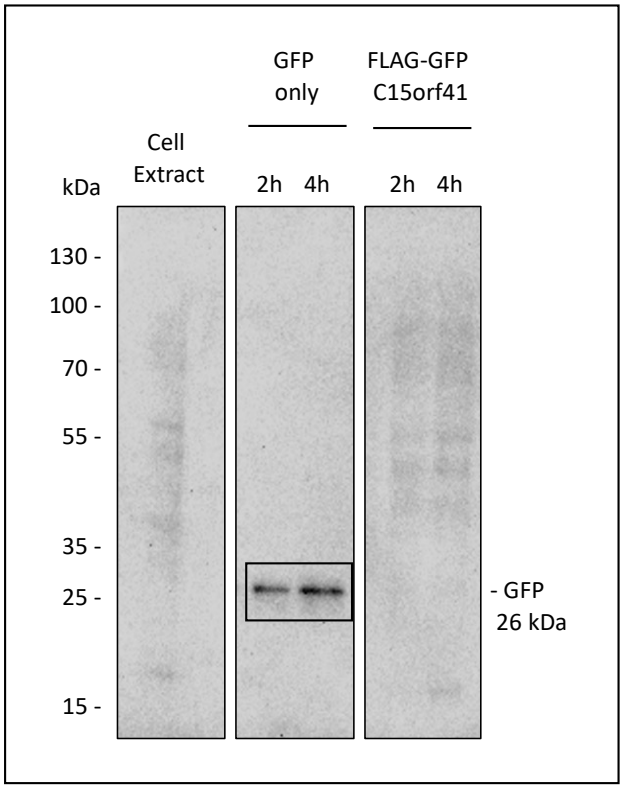


Figure 4.2: Transfected FLAG-GFP-C15orf41 is not detectable by western blotting for GFP
GFP alone as well as FLAG-GFP-C15orf41 were transfected in HEK293-FT cells and cell extracts run on a 4-12% Bis-Tris protein gel and blotted with GFP (Roche,11814460001) antibody. GFP was only visible in the GFP only control and not in samples containing FLAG-GFP-C15orf41. Un-transfected HEK293-FT cell extract was used as a negative control. In all cases, both 2 hours and 4 hours transfection lengths were performed (indicated as 2h and 4h).

may, therefore, acts as a linker, thereby allowing for epitope detection and signal on the western blots obtained.

Next, C15orf41-Codanin-1 complex formation was investigated using immuno-precipitation experiments. To this end, HEK cells transfected with FLAG-GFP-C15orf41, lysed and applied to pre-washed FLAG-M2 beads to allow for specific binding. This was then followed by consecutive washes with wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, protease inhibitor and 0.5% NP-40). Two different methods of elution were tested to determine the optimum conditions: a non-denaturing 3xFLAG peptide elution method and a denaturing, boiling method. Fractions were collected before and after the immuno-precipitation step, termed

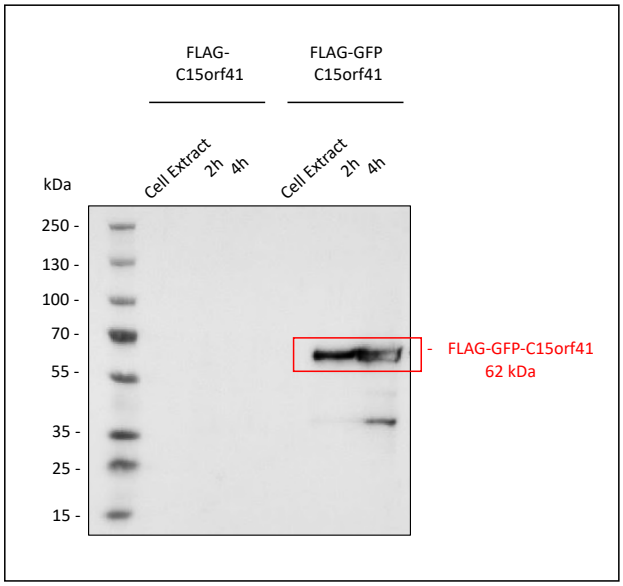


Figure 4.3: FLAG-GFP-C15orf41 only is detectable by western blotting for FLAG
FLAG-C15orf41 as well as FLAG-GFP-C15orf41 were transfected in HEK293-FT cells and cell extracts run on a 4-12% Bis-Tris protein gel and blotted with FLAG (Sigma,F1804) antibody. FLAG-GFP-C15orf41 was visible, at the predicted size of 62 kDa. No signal was observed for the FLAG-C15orf41 samples, suggesting masking of the tag. Un-transfected HEK293-FT cell extract was used as a negative control. In all cases, both 2 hours and 4 hours transfection lengths were performed (indicated as 2h and 4h).

Input and IP: FLAG, respectively, and run on a denaturing SDS-PAGE gel and subsequently blotted with primary anti-FLAG antibody to visualise pulled-down FLAG-GFP-C15orf41 (Figure 4.4.a). With both elution methods, FLAG-GFP-C15orf41 was successfully pulled-down, where bands were detected at approximately 62 kDa. With the boiling elution method, the heavy chain (around 55 kDa) and light chain (around 25 kDa) of the FLAG antibody linked to the beads are also visible on the gel. In order to confirm the specificity of the immuno-precipitation experiment and FLAG antibody, the same protocol was followed for FLAG-GFP-C15orf41 as well as a short FLAG peptide-containing VPN5 plasmid (Figure 4.4.b). As Figure 4.4.c shows, C15orf41 is only visible in the samples transfected with the FLAG-GFP-C15orf41 plasmid and the control plasmid pulled-down the FLAG peptide only, confirming the specificity of the antibody and beads used in this experiment. In order to investigate C15orf41-Codanin-1 complex formation, the output samples were blotted for Codanin-1 using a newly-generated commercial Codanin-1 antibody

(Bethyl). Empty FLAG plasmid was used as a negative control and HEK293-FT cell extract as a positive control. As expected, Codanin-1 is co-eluted with C15orf41 only and not the FLAG control plasmid confirming C15orf41-Codanin-1 complex formation in HEK293-FT cell (Figure 4.4.d). It should also be noted that two main bands (of similar molecular weight) are visible when Codanin-1 is blotted with the Bethyl antibody in the Input sample. The origin of this double-banding is currently unknown. It remains to be investigated whether this represents two different isoforms of Codanin-1, a possible post-translational modification of Codanin-1 or whether this is an artifact of the newly-developed antibody. Future experiments, whereby Codanin-1 is silenced using siRNA for instance, will need to be conducted in order to fully understand this observation.

Subsequently, the interaction between C15orf41 and Codanin-1 was examined in more detail. To this end, immuno-precipitation experiments were performed with increasing salt concentrations and samples checked for complex formation. The following salt concentrations were used in the washing steps: 150 mM NaCl, 250 mM NaCl and 500 mM NaCl. As seen in Figure 4.5, both C15orf41 and Codanin-1 are co-eluted at all salt concentrations, albeit faintly at 500 mM, implying a robust (salt-resistant) interaction between the two binding partners.

4.1.2 Complex formation in HUDEP-2 cells at endogenous levels

Current published data shows C15orf41-Codanin-1 complex formation in transformed human cancer cell lines using over-expressed forms of the proteins. Previously, over-expression experiments for both Codanin-1 and C15orf41 have led to confounding results for both localisation and complex formation when patient mutations were analysed, with conflicting data presented in the published literature (Shroff et al., 2020.; Swickley et al., 2020). In particular, the cytoplasmic localisation of Codanin-1 when over-expressed strongly suggests Codanin-1 behaves aberrantly when over-expressed. CDA-I is an erythroid-specific disease, it is therefore necessary to explore the relationship between C15orf41 and Codanin-1 in an erythroid-specific

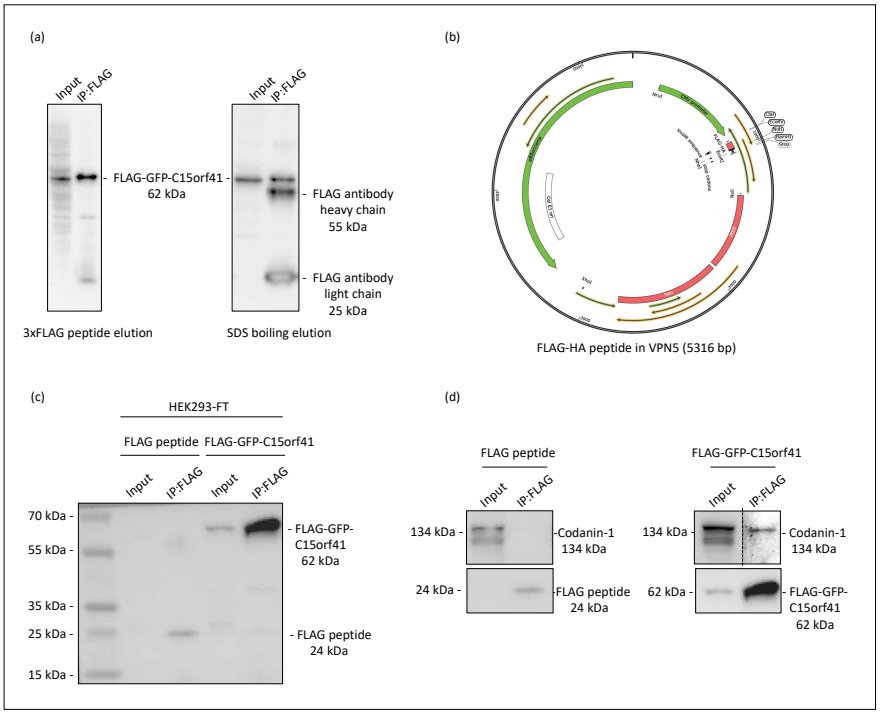


Figure 4.4: Co-immunoprecipitations of FLAG-GFP-C15orf41 with Codanin-1 in HEK293-FT cells
(a)FLAG-GFP-C15orf41 from transfected HEK293-FT cells can be successfully immunoprecipitated using FLAG-M2 beads. Two different elution methods were tested: a non-denaturing 3xFLAG-peptide and a denaturing boiling method. In both cases, FLAG-GFP-C15orf41 was obtained in the output sample.
(b)Map of FLAG-HA peptide in VPN5.
(c)Using FLAG-M2 beads, FLAG-peptide as well as FLAG-GFP-C15orf41 was successfully immunoprecipitated in transfected HEK293-FT cells. The tagged C15orf41 protein is only present in the samples transfected with the FLAG-GFP-C15orf41, confirming the band observed is our protein of interest and not an artifact of the beads/antibody.
(d)FLAG-GFP-C15orf41 and Codanin-1 form a complex in HEK293-FT cells. Codanin-1 was co-eluted with FLAG-GFP-C15orf41 only. No bands for Codanin-1 were observed when HEK293-FT cells were transfected with FLAG peptide plasmid showing the specificity of the FLAG IP.
HEK293-FT cells were transfected with either FLAG-peptide or FLAG-GFP-C15orf41 plasmid before lysis and immunoprecipitation using FLAG-M2 beads. Immunoprecipitates were then run on a 4-12% Bis-Tris protein gel and blotted with appropriate antibodies.
FLAG-GFP-C15orf41: 62 kDa, blotted with FLAG antibody (Sigma, F1804).
FLAG peptide: 24 kDa, blotted with FLAG antibody (Sigma, F1804).
Codanin-1: 134 kDa, blotted with anti-Codanin-1 antibody (Bethyl, A304-951A).

context at endogenous levels and under their natural regulatory control, to provide a clearer picture of the physiological role of these proteins and help elucidate the potential disease mechanism of CDA-I.

To this end, HUDEP-2 cells (Human Umbilical cord Derived Erythroid Progenitor-

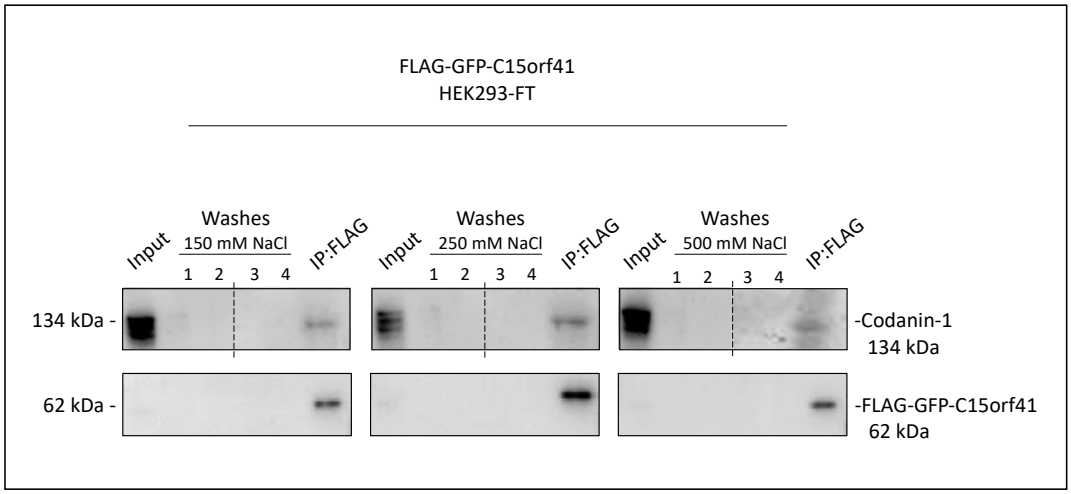


Figure 4.5: FLAG-GFP-C15orf41 and Codanin-1 form a tight complex in HEK293-FT cells
FLAG-GFP-C15orf41 and Codanin-1 form a complex in HEK293-FT cells. Codanin-1 was co-eluted with FLAG-GFP-C15orf41 withstanding NaCl concentrations up to 500 mM. Lanes representing the different washes at 150 mM, 250 mM and 500 mM NaCl were included to show there is no complex dissociation.
HEK293-FT cells were transfected with FLAG-GFP-C15orf41 plasmid before lysis and immunoprecipitation using FLAG-M2 beads. Immunoprecipitates were then run on a 4-12% Bis-Tris protein gel and blotted with appropriate antibodies.
FLAG-GFP-C15orf41: 62 kDa, blotted with FLAG antibody (Sigma, F1804).
Codanin-1: 134 kDa, blotted with anti-Codanin-1 antibody (Bethyl, A304-951A).

2) (Kurita, 2014) were genome engineered by Dr. C. Babbs and A-A. Olijnik at the MRC Weatherall Institute of Molecular Medicine to develop a clinically-relevant erythroid model of CDA-I. Although working with HUDEP-2 cells can prove to be challenging due to low transfection rates and specific growth requirements, HUDEP-2 cells offer major advantages for studying erythroid-specific diseases as they can be maintained as proerythroblasts in culture. Proerythroblasts are fully committed cells that have entered terminal erythroid differentiation. HUDEP-2 cells can be expanded to large numbers and present an invaluable resource to study the early stages of terminal erythroid differentiation, which becomes crucial when investigating erythroid-specific diseases such as CDA-I.

HUDEP-2 cells were genome engineered using CRISPR/Cas9-based methodology to add tags at the endogenous locus of both proteins as well as the introduction of patient mutations in both C15orf41 and Codanin-1. For both proteins, the founder mutations that were first discovered in each protein in CDA-I patients were

chosen as a representation for CDA-I patient models. The following HUDEP-2 cell lines were available during this study:

1. Wild-type HUDEP-2 cells (WT HUDEP-2)
2. HUDEP-2 cells with a N-terminal 3xFLAG tag homozygously targeted to the endogenous C15orf41 loci (3xFLAG-C15orf41)
3. HUDEP-2 cells with a C-terminal STREP tag homozygously targeted to the Codanin-1 loci (STREP-Codanin-1)
4. (4) HUDEP-2 cells with a N-terminal 3xFLAG tag homozygously targeted to the C15orf41 loci and the p.L178Q patient mutant missense mutation in C15orf41 (3xFLAG-C15orf41^{L178Q})
5. (5) HUDEP-2 cells with a C-terminal STREP tag homozygously targeted to the endogenous Codanin-1 loci and the p.R1042W patient mutant missense mutation in Codanin-1 (STREP-Codanin-1^{R1042W})

To confirm these cell lines represent a suitable disease model for CDA-I, the generated HUDEP-2 CDA-I cell lines were differentiated through the erythroid pathway to check for the appearance of the characteristic CDA-I heterochromatin defects of CDA-I. Differentiation was performed by A-A. Olijnik at the MRC Weatherall Institute of Molecular Medicine using a modified version of the protocol described in Scott et al., 2019 and monitored by immunophenotyping and observation of cell morphology. At day 10 of differentiation (which represents the late erythroblast stage), cells were analysed by transmission electron microscopy (TEM) by Dr. E. Johnson at the Sir William Dunn School of Pathology, University of Oxford (Figure 4.6). Whilst the WT and 3xFLAG-C15orf41 HUDEP-2 cells showed no chromatin abnormalities, the HUDEP-2 cell lines homozygous for patient mutations in either *C15orf41* or *CDAN1* (both of 3xFLAG-C15orf41^{L178Q} and STREP-Codanin-1^{R1042W}) exhibited the characteristic pathognomonic heterochromatin abnormalities in CDA-I. To further confirm this phenotype, CDA-I patient erythroblasts (CD34^{+} cells) were also differentiated using the same protocol and revealed the characteristic heterochromatin abnormalities seen in the engineered CDA-I patient HUDEP-2 cell

lines. CD34$^+$ cells from a healthy individual were used as a control. Interestingly, it was noted that while pathogenic missense mutations in both known causative CDA-I genes produce the same nuclear abnormalities in HUDEP-2 cells, the C15orf41 L178Q patient mutation has a more marked effect then the Codanin-1 R1042W mutation. Whilst >50% of nuclei showed 'spongy' heterochromatin in 3xFLAG-C15orf41^{L178Q} cells, STREP-Codanin-1^{R1042W} cells only exhibit about 20% of abnormal nuclei (A-A. Olijnik, unpublished). Further studies will be required to determine whether this is a result of clonal variation or whether mutations of C15orf41 cause a higher frequency of abnormal erythroblast.

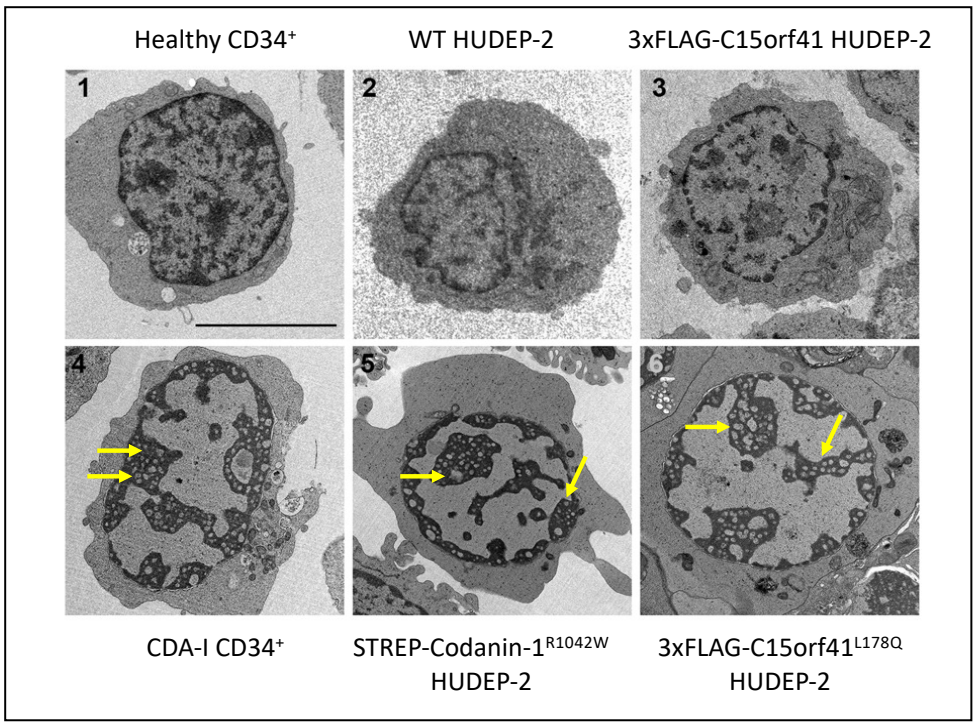


Figure 4.6: CDA-I cellular defects are recapitulated in culture HUDEP-2 cell model
Transmission electron micrographs were taken from *in vitro* cultured erythroblasts (CD34$^+$) from a healthy individual (1), WT HUDEP-2 cell line (2) as well as 3xFLAG-C15orf41 HUDEP-2 cells (3). In 1-3, no chromatin abnormalities are visible. In contrast, 4-6 show the characteristic 'spongy' heterochromatin phenotype observed in CDA-I: dense, 'Swiss-cheese like' heterochromatin with electron lucent patches is visible (yellow arrows). (4) CDA-I patient CD34$^+$ cells, (5) STREP-Codanin-1^{R1042W} HUDEP-2 cells, (6) 3xFLAG-C15orf41^{L178Q} HUDEP-2 cells.

After establishing that the mutant HUDEP-2 cell lines recapitulate the key aspects of CDA-I, complex formation between C15orf41 and Codanin-1 was in-

vestigated next, using proteins at native levels. Firstly, 3xFLAG-C15orf41 was immunoprecipitated using the same protocol described in the previous section and the output analysed by Western blotting. In short, FLAG-C15orf41 cells were collected and lysed and incubated with FLAG-M2 magnetic beads. This was followed by consecutive washes with wash buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, protease inhibitor and 0.5% NP-40) and non-denaturing elution using the 3xFLAG peptide elution method. Fractions were collected before and after the immuno-precipitation step, termed Input and Output respectively and run on a denaturing SDS-PAGE gel and blotted with primary FLAG antibody to visualise pulled-down 3xFLAG-C15orf41 (Figure 4.7.a). Results reveal 3xFLAG-C15orf41 can be immunoprecipitated quantitatively from HUDEP-2 cells, migrating at approximately 38 kDa on SDS-PAGE gels. WT HUDEP-2 cells with no endogenous tags were used as a negative control to confirm specificity of the FLAG beads and tag. In those samples, no 3xFLAG-C15orf41 is visible (Figure 4.7.a). Furthermore, blotting the output sample for Codanin-1 confirmed that C15orf41 and Codanin-1 form a complex, since Codanin-1 co-elutes with C15orf41 only in 3xFLAG-C15orf41 HUDEP-2 cells and not in WT HUDEP-2 (Figure 4.7.a). To investigate the salt-stability of the native complex, co-immunoprecipitation experiments were performed by washing the bound complexes at increasing salt concentrations, namely 150 mM NaCl, 250 mM NaCl and 500 mM NaCl (Figure 4.7.b). Again, C15orf41-Codanin-1 complex is visible in all output samples, withstanding high concentration of salt (0.5M NaCl), implying a strong direct interaction between the two proteins.

This work represents the first report of C15orf41-Codanin-1 complex formation at endogenous levels in an erythroid cell line. The generation of the HUDEP-2 CDA-I disease model represents a useful novel resource in the CDA-I field as it allows researchers to explore cellular functions of C15orf41 and Codanin-1 at endogenous levels in a context relevant to the disease. This section shows complex formation between C15orf41 and Codanin-1 using C15orf41 as bait, the reverse co-immunoprecipitation was attempted but not successful. STREP-Codanin-1 cells were incubated with beads coupled to a STREP antibody but STREP-Codanin-1

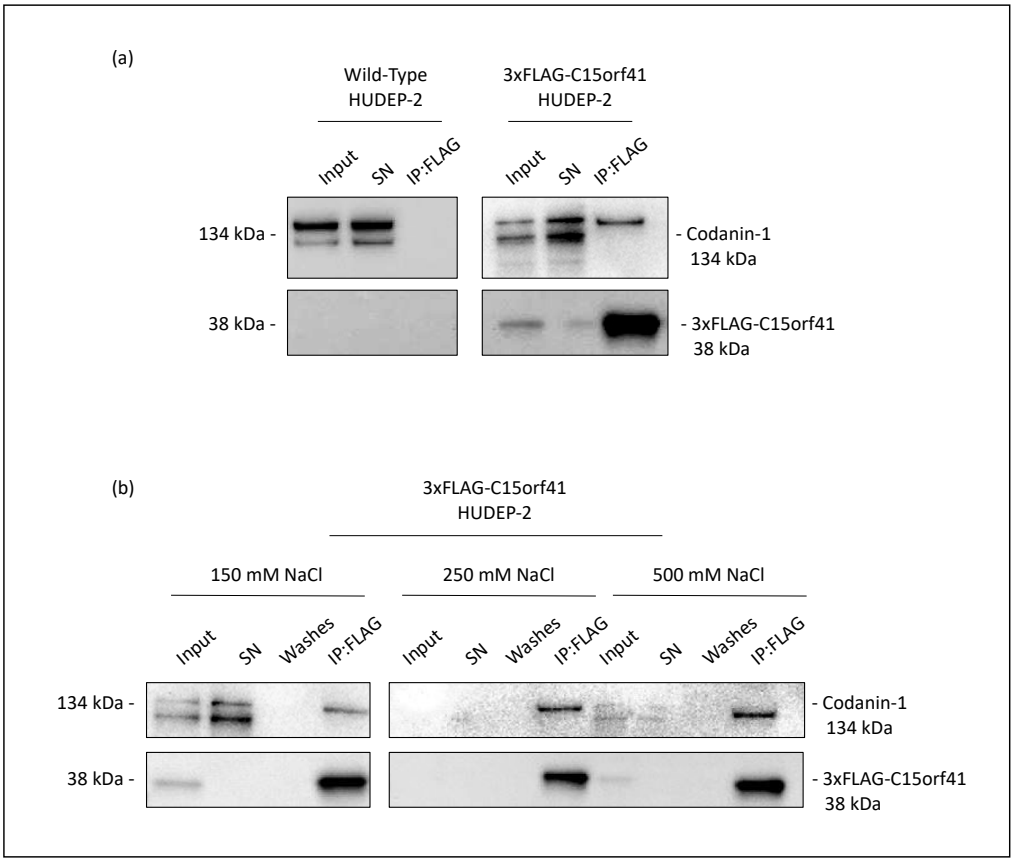


Figure 4.7: Co-immunoprecipitations of 3xFLAG-C15orf41 with Codanin-1 in HUDEP-2 cells
(a) 3xFLAG-C15orf41 and Codanin-1 form a complex in HUDEP-2 cells. Codanin-1 is co-eluted with 3xFLAG-C15orf41 at endogenous levels in HUDEP-2 cells. No bands for C15orf41 or Codanin-1 were observed in the output in WT HUDEP-2 cells confirming specificity of the FLAG-IP step. SN: supernatant collected after lysate was incubated with FLAG-M2 beads showing efficiency of the IP step.
(b) Codanin-1 was co-eluted with 3xFLAG-C15orf41 withstanding NaCl concentrations up to 500 mM. Lanes representing the different washes at 150 mM, 250 mM and 500 mM NaCl were included to show there is no complex dissociation. SN: supernatant collected after lysate was incubated with FLAG-M2 beads showing efficiency of the IP step.
Wild-type HUDEP-2 and 3xFLAG-C15orf41 HUDEP-2 cells were lysed and immunoprecipitated using FLAG-M2 beads. Immunoprecipitates were then run on a 4-12% Bis-Tris protein gel and blotted with appropriate antibodies.
3xFLAG-C15orf41: 38 kDa, blotted with FLAG antibody (Sigma, F1804).
Codanin-1: 134 kDa, blotted with anti-Codanin-1 antibody (Bethyl, A304-951A).

could not be immunoprecipitated and no band was visible by Western blotting (see Section 4.2 below). This may be sue to the small size of the STREP tag (around 3 kDa) that is likely being masked within the large, 134 kDa Codanin-1 protein. The remaining of the immuno-precipitation experiments were, therefore, performed

using the wild-type and patient mutant 3xFLAG-C15orf41 HUDEP-2 cell lines.

4.1.3 Patient mutations in C15orf41 disrupt C15orf41-Codanin-1 complex formation in HUDEP-2 cells

The section above shows wild-type C15orf41 and Codanin-1 interact and form a stable complex. Next, the effect of the L178Q patient mutation in C15orf41 on C15orf41-Codanin-1 complex formation was assessed.

The effect of CDA-I patient mutations on complex formation has previously been assessed in U2OS (osteosarcoma) cells using over-expression plasmids encoding epitope-tagged C15orf41 and Codanin-1 harbouring founder mutations L178Q or Y94C for C15orf41 and R1042W for Codanin-1. In all cases, complex formation was not disrupted, although results should be interpreted carefully considering the excess of protein present (Shroff et al., 2020; Swickley et al., 2020).

In order to define at the effect of CDA-I patient mutations on C15orf41-Codanin-1 complex formation, immuno-precipitation experiments were performed on 3xFLAG-C15orf41^{L178Q} HUDEP-2 cells. HUDEP-2 cell expressing wild-type 3xFLAG-C15orf41 were used as a positive control to control for immuno-precipitation efficiency.

Whilst we are able to efficiently pull-down wild-type C15orf41 and confirm complex formation with Codanin-1, using the 3xFLAG-C15orf41^{L178Q} cell line revealed a different outcome. Although the C15orf41^{L178Q} protein was detected in the output lane, albeit at a lower level than the WT protein (reflecting the reduced C15orf41 levels in HUDEP-2 cells harbouring this patient mutation discussed in Section 4.2, below), no Codanin-1 was co-eluted in the mutant cell line (Figure 4.8). This suggested that mutation L178Q in C15orf41 disrupts complex formation with Codanin-1. Considering the role of both proteins in CDA-I pathogenesis, it may be that loss of C15orf41-Codanin-1 interaction contributes to CDA-I, although the exact molecular details of this remain unknown and additional defects, beyond interaction, might also be impacted by mutation L178Q. The result also further highlights the importance of studying these disease proteins in a relevant erythroid

context, at native levels. Whilst previous studies using over-expressed proteins showed no effects of patient mutations on C15orf41-Codanin-1 complex formation, this work clearly demonstrates their detrimental effect, suggesting part of the mechanism underlying CDA-I pathology.

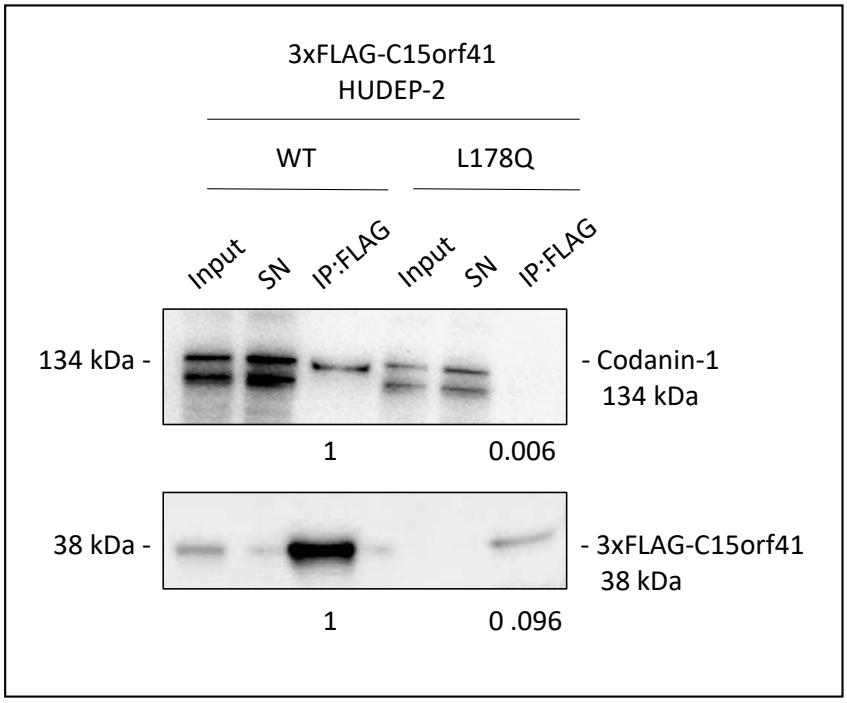


Figure 4.8: The L178Q patient mutation in C15orf41 disrupts complex formation with Codanin-1 in HUDEP-2 cells
The L178Q patient mutation in C15orf41 abolishes C15orf41-Codanin-1 complex formation at endogenous levels in HUDEP-2 cells. Codanin-1 only co-eluted with wild-type 3xFLAG C15orf41. In 3xFLAG-C15orf41^{L178Q} cells, only mutant 3xFLAG-C15orf41 was immunoprecipitated but no band for Codanin-1 was co-eluted. The lower immunoprecipitation efficiency in 3xFLAG-C15orf41^{L178Q} cells reflects the lower C15orf41 protein levels described in this cell line.
SN: supernatant collected after lysate was incubated with FLAG-M2 beads showing efficiency of the IP step.
3xFLAG-C15orf41 and 3xFLAG-C15orf41^{L178Q} HUDEP-2 cells were lysed and immunoprecipitated using FLAG-M2 beads. Immunoprecipitates were then run on a 4-12% Bis-Tris protein gel and blotted with appropriate antibodies.
3xFLAG-C15orf41: 38 kDa, blotted with FLAG antibody (Sigma, F1804).
Codanin-1: 134 kDa, blotted with anti-Codanin-1 antibody (Bethyl, A304-951A).
Band intensities were analysed using ImageJ and ratios of protein levels compared to wild-type 3xFLAG-C15orf41 HUDEP-2 cells indicated below each band where applicable.

4.2 Patient mutations in C15orf41 and Codanin-1 affect protein levels of C15orf41 and Codanin-1, albeit differentially

4.2.1 Antibody Validation

The generation of the CDA-I disease model using the HUDEP-2 cell lines was also used as an opportunity to validate newly-available antibodies for both Codanin-1 and C15orf41. Considering the erythroid-specificity of CDA-I, these cell-lines were chosen to allow for visualisation of endogenous levels of C15orf41 and Codanin-1 in an erythroid-specific context. Using the STREP-Codanin-1 cell line, whole-cell extracts were lysed and blotted with both the anti-Codanin-1 antibody (Bethyl) which had previously been validated and several anti-STREP antibodies to check for tag detection. As shown in Figure 4.9.a, Codanin-1 detection was observed with the anti-Codanin-1 antibody from Bethyl whereas only very weak signal was achieved with any of the STREP antibodies. This result offers a possible explanation as to why immuno-precipitation experiments with Codanin-1 immobilisation were unsuccessful, likely due to masking of the STREP tag in the Codanin-1 protein at the endogenous locus. Masking may persist following denaturation for loading and analysis by SDS-PAGE, implying the STREP epitope is highly inaccessible when added to the C-terminus of Codanin-1 and tagging on the N-terminus would represent an alternative approach to circumvent this.

Thus far, C15orf41 detection has been via the 3xFLAG-tag using the FLAG antibody. In order to be able to visualise C15orf41 in non-tagged cell lines, several newly-produced C15orf41 antibodies from Abcam and Cusabio were tested and compared to detection with the FLAG antibody. This experiment was performed on both cell extracts from 3xFLAG-C15orf41 HUDEP-2 cells (endogenous levels) as well as immuno-precipitates from HEK293-FT cells transfected with FLAG-GFP-C15orf41 (over-expressed levels). In all cases, C15orf41 was detected at the correct molecular weight validating those antibodies (Figure 4.9.b). This

represents an advance in CDA-I biology as basic tools were previously missing, now published in Olijnik et al., 2021.

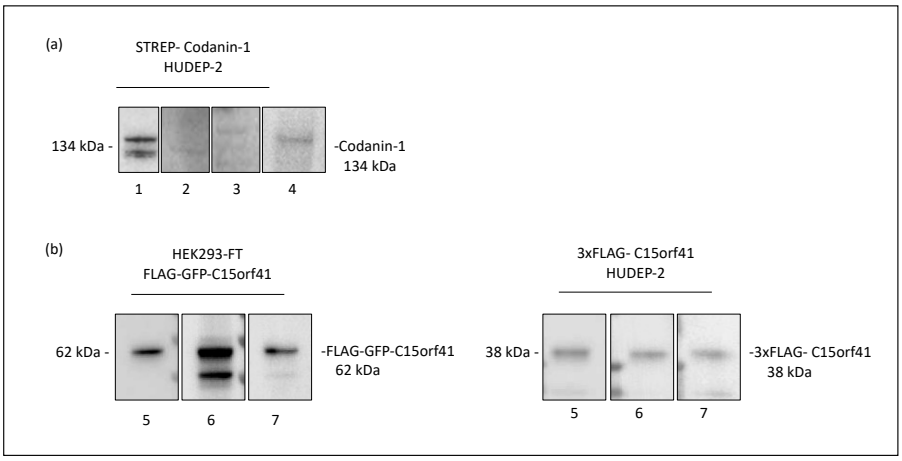


Figure 4.9: Testing anti-Codanin-1 and anti-C15orf41 primary antibodies
(a) Whole-cell extracts of wild-type STREP-Codanin-1 HUDEP-2 cells were blotted for Codanin-1 (134 kDa) with anti-Codanin-1 antibody (Bethyl, A304-951A) (1), anti-STREP (abcam, Ab76949) (2), anti-STREP (Abcam, ab184224) (3) and anti-STREP (Qiagen, 34850) (4). Strong signal is only observed with anti-Codanin-1 antibody suggesting masking of the STREP-tag.
(b) Whole cell extracts of 3xFLAG-C15orf41 HUDEP-2 cells as well as immunoprecipitates from HEK293-FT cells transfected with FLAG-GFP-C15orf41 were blotted for 3xFLAG-C15orf41 (38 kDa) or FLAG-GFP-C15orf41 (62 kDa) with FLAG antibody (Sigma, F1804) (5), anti-C15orf41 (Abcam, ab107352) (6) and anti-C15orf41 (Cusabio, CSBPA897474LA01HU) (7). In all cases, C15orf41 was detectable, validating the newly-generated antibodies.
All sample were run on a 4-12% Bis-Tris protein gel and blotted with the antibodies indicated above.

4.2.2 Analysis of C15orf41 and Codanin-1 protein levels in HUDEP-2 cells

Using the HUDEP-2 CDA-I disease model, endogenous proteins levels of C15orf41 and Codanin-1 were investigated in both wild-type cell lines and in cell lines harbouring patient mutations p.L178Q in C15orf41 and p.R1042W in Codanin-1. Firstly, immuno-blotting experiments were performed using 3xFLAG-C15orf41 and 3xFLAG-C15orf41^{L178Q} cells. This revealed the introduction of the L178Q mutation in C15orf41 leads to a significant decrease in protein level compared to WT 3xFLAG-C15orf41 (Figure 4.10.a). Quantification showed the levels of C15orf41

in the patient-mutant cell line were only approximately 10% of those seen in the wild-type. These numbers agree with the observed decreased immuno-precipitation efficiency in 3xFLAG-C15orf41^{L178Q} cells. Considering the strong disease phenotype of 3xFLAG-C15orf41^{L178Q} cells, with over 50% exhibiting 'spongy' heterochromatin nuclei, a hypothesis can be inferred between low levels of C15orf41 and CDA-I pathogenesis. The L178Q mutation in C15orf41 also led to decreased levels of Codanin-1 (only around 30% relative to the wild-type control), suggesting a mutual role of C15orf41 and Codanin-1 in promoting their stability. Published reports find that knockdown of Codanin-1 also destabilised C15orf41, further strengthening the hypothesis that integrity of Codanin-1 and C15orf41 are linked (Olijnik et al., 2021). Interestingly, when CDA-I patient erythroblasts with C15orf41 mutations p.Y94C and p.C157Y were assessed for C15orf41 and Codanin-1 levels, no reduction in either protein levels were observed (Olijnik et al., 2021). This indicates that whilst some patient mutations, such as p.L178Q in C15orf41, destabilise both CDA-I proteins, others likely have a different effect, potentially by disrupting a yet unknown interaction. Both mechanisms point to a lack of protein functionality and suggest CDA-I pathogenesis can arise from more than one cause.

Western blotting analysis of WT STREP-Codanin-1 and STREP-Codanin-1^{R1042W} HUDEP-2 cells revealed that the R1042W mutation does not affect Codanin-1 levels as Codanin-1 was present at equal levels in both cell lines. The Codanin-1 patient mutation did however lead to a significant decrease in C15orf41 levels, with roughly 50% loss (Figure 4.10.b). These results suggest that Codanin-1 integrity might be required to ensure stability of C15orf41 and that some patient mutations in Codanin-1 can destabilise C15orf41. The R1042W mutation in Codanin-1 is located at the C-terminus of the protein. This region also constitutes the C15orf41 binding region, leading to the hypothesis that disruption of C15orf41-Codanin-1 association leads to destabilisation of C15orf41 alone as well (Swickley et al., 2020; Olijnik et al., 2021).

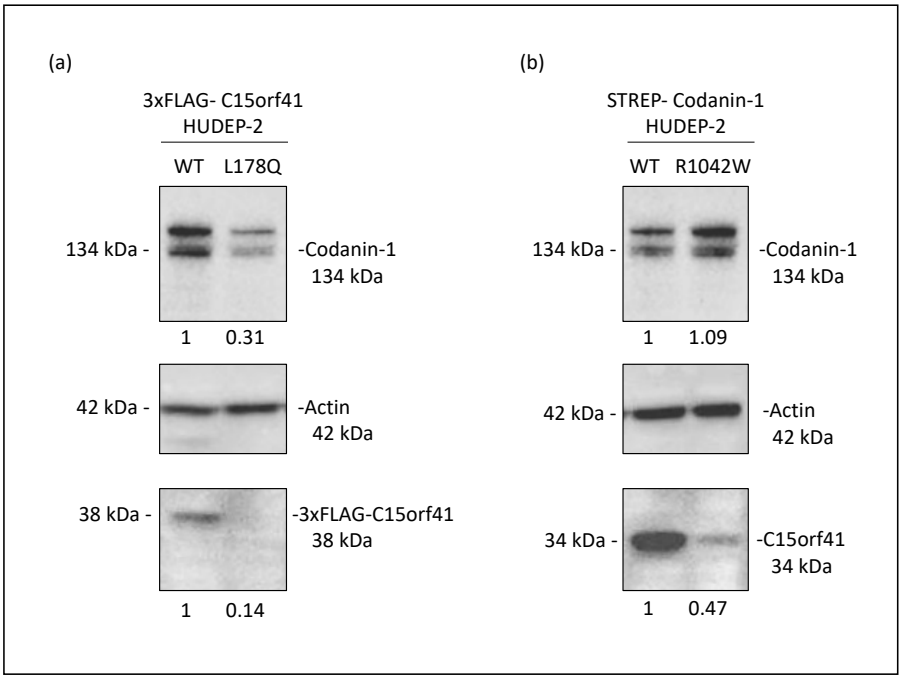


Figure 4.10: The L178Q patient mutation in C15orf41 affects both levels of C15orf41 and Codanin-1 whereas patient mutation R1042W in Codanin-1 only affects C15orf41 levels in HUDEP-2 cells
(a) The L178Q patient mutation in C15orf41 causes a reduction in both C15orf41 and Codanin-1 protein levels. Only around 10% of C15orf41 was detectable in the patient mutant cell line compared to WT C15orf41. The L178Q mutation in C15orf41 also led to decreased levels of Codanin-1 (only around 30% detectable), suggesting a mutual role of C15orf41 and Codanin-1 in promoting their stability.
(b) The R1042W patient mutation in Codanin-1 doesn't affect Codanin-1 levels but causes a reduction in C15orf41 protein levels, with roughly 50% loss of expression.
Whole cell extracts of all HUDEP-2 cell lines indicated were lysed and samples then run on a 4-12% Bis-Tris protein gel and blotted with appropriate antibodies. Actin was used as a loading control.
3xFLAG-C15orf41: 38 kDa, blotted with FLAG antibody (Sigma, F1804).
C15orf41: 34 kDa, blotted with anti-C15orf41 antibody (Abcam, ab107352).
Codanin-1: 134 kDa, blotted with anti-Codanin-1 antibody (Bethyl, A304-951A).
Actin: 42 kDa, blotted with anti-Actin HRP antibody (Sigma, A3954).
Band intensities were analysed using ImageJ and ratios of protein levels compared to the appropriate wild-type HUDEP-2 cell line indicated below each band where applicable.

4.3 Conclusion

In summary, this chapter provides new insights into C15orf41 and Codanin-1, the two main disease proteins in CDA-I. Using both over-expressed proteins in an epithelial cell line and endogenously-tagged proteins in a relevant, erythroid HUDEP-2 background, we show C15orf41 and Codanin-1 form a stable complex, that is able

to withstand increasing salt concentrations up to 500 mM. Interestingly, C15orf41 and Codanin-1 complex formation is disrupted by the L178Q patient mutation in C15orf41 in HUDEP-2 cells. The interaction between Codanin-1 and C15orf41 is mediated by a C-terminal fragment of Codanin-1 yet no clearly defined interacting domains were identified in C15orf41. The L178Q patient mutation in C15orf41 is located in the predicted nuclease domain of the protein, a key functional domain, yet it still disrupts complex formation with Codanin-1. Considering the severe phenotype of the L178Q C15orf41 patient cells, it may be that complex formation between Codanin-1 and C15orf41 is crucial for maintenance of correct erythropoiesis and disruption of complex formation is highly pathogenic. Furthermore, it is also likely that the interaction between Codanin-1 and C15orf41 will have an effect of a, yet undescribed, function of C15orf41 as the patient mutation is located in the predicted nuclease domain of the protein.

Although previous reports using over-expression experiments did not find that patient mutations in C15orf41 disrupt C15orf41-Codanin-1 complex formation, our data supports this idea. This discrepancy highlights the relevance of studying CDA-I in an erythroid background such as HUDEP-2 cells using native protein levels through the introduction of epitope tags at the endogenous loci. The development of our HUDEP-2 CDA-I disease model therefore provides an important tool for current and future cellular studies of C15orf41 and Codanin-1, without relying on extremely limited patient cellular material, that cannot be easily expanded to the scale required to probe mechanism, or through over-expression experiments that are prone to artefact.

We also characterised the effects of patient mutations L178Q and R1042W in C15orf41 and Codanin-1, respectively, on protein levels in our HUDEP-2 cell line. Importantly, we found that C15orf41 levels are affected by both the L178Q patient mutation in C15orf41 as well as the R1042W patient mutation in Codanin-1, whereas Codanin-1 levels are only affected by the L178Q mutation in C15orf41 suggesting a role for both proteins in ensuring mutual stability and integrity. This finding also suggests C15orf41 may be the key determinant of CDA-I, highlighting

the importance of identifying its cellular role. Still, the exact cellular function of C15orf41 and Codanin-1, both individually and as a complex, remains to be investigated. As such the next part of this thesis will focus on *in vitro* studies of the two CDA-I disease proteins. In light of secondary structure predictions revealing C15orf41 is potentially a nuclease, Chapters 5 and 6 will investigate this hypothesis.

5

Purification and Biophysical Characterisation of the C15orf41-Codanin-1 Complex

The structure and function of both C15orf41 and Codanin-1, the two disease proteins of CDA-I, was unknown before work on this thesis began. In the previous chapter, experiments revealed the two proteins form a stable complex *in vivo* that is disrupted if certain disease-associated mutations are introduced within C15orf41. Unpublished data from our collaborators (Gileadi group in the Centre for Medicines Discovery, University of Oxford) narrowed down C15orf41's binding region to the C-terminus of Codanin-1 through partial trypsin digestion (Experiments performed by Dr. J. Newman, Centre for Medicines Discovery, University of Oxford, unpublished). The aim of this chapter is therefore to purify full-length C15orf41 and the interacting C-terminal Codanin-1 fragment, both individually and as a complex to create pure protein that will then be used for downstream *in vitro* studies of the proteins, including a biophysical and biochemical characterisation as well as structural analysis.

5.1 Protein Purification: purification of wild-type C15orf41-Codanin-1 complex and mutant versions

Our *in vivo* studies as well as recent literature reports and unpublished data from our collaborators show C15orf41 forms a stable complex with Codanin-1,

interacting with the C-terminal region of Codanin-1, ranging from amino acids 866 to 1227. Considering the large size of full-length Codanin-1 (134 kDa) and the challenges this would present for purification, it was chosen to only focus on this C-terminal fragment of Codanin-1 for this study. From here on within this thesis, in all biochemical and biophysical studies the C-terminal Codanin-1 fragment (866-1227aa) interacting with C15orf41 will be referred to as 'Codanin-1' for clarity.

To purify the C15orf41-Codanin-1 complex, full-length C15orf41 and the C-terminal fragment of Codanin-1 were initially purified separately followed by a 'complex formation' step and a final purification step. Protein purification was done with the help of A. Gavard at the Centre for Medicines Discovery, University of Oxford and results described below. More details can be found in Chapter 3 (Materials & Methods).

Firstly, codon-optimised full length C15orf41 in a pNIC28-Bsa4 expression vector with a Histidine (His_6) tag in a 22-amino acid N-terminal fusion peptide with a recombinant tobacco etch virus (rTEV) protease cleavage site was expressed and purified from *E. coli* BL21 Rosetta2 cells. Briefly, the purification protocol involved an initial nickel resin binding step allowing for selective binding of C15orf41 to the resin through its N-terminal His-tag, exploiting the intrinsic metal-binding capacity of the tag. Bound C15orf41 was then eluted in elution buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 300 mM imidazole, 5% glycerol and 1 mM TCEP). Samples for each step outlined above were collected and analysed by SDS-PAGE to check for His_6-C15orf41 presence and purity (Figure 5.1 a&b). Next, the C-terminal fragment of Codanin-1 (amino acids 866-1227) in an expression vector with N-terminal His6 and Z-basic (ZB) tags followed by a rTEV protease cleavage site was also expressed and purified from *E. coli* using the same method as for C15orf41. Again, samples for each step purification step were collected and analysed by SDS-PAGE to check for His_6-ZB-Codanin-1 presence and purity (Figure 5.1 a&b).

All fractions containing C15orf41 and Codanin-1 were dialysed overnight at 4°C in dialysis buffer (50 mM HEPES pH 7.5, 500 mM NaCl, 1 mM TCEP, 5% glycerol). The Codanin-1 sample was supplemented with rTEV protease for cleavage

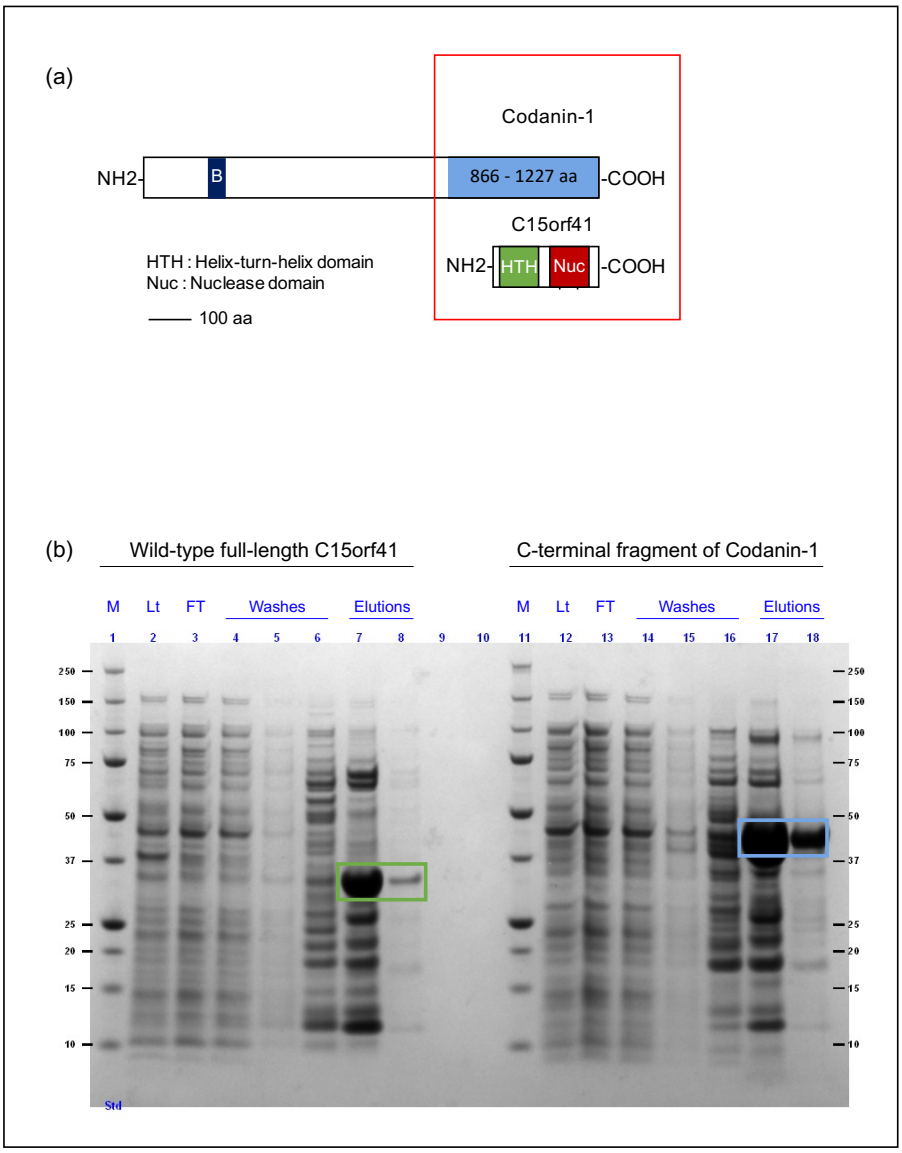


Figure 5.1: Purification of full-length C15orf41 and the C-terminal fragment of Codanin-1
(a) Schematic of C15orf41 and Codanin-1. The red box indicates the regions purified in this study, namely full length C15orf41 and the C-terminus of Codanin-1 (866-1227 amino acids (aa)).
(b) SDS-PAGE of purified C15orf41 and Codanin-1 after the initial nickel resin batch purification. Purified His_6-C15orf41 is indicated in the green box ($\sim$ 35 kDa) and purified His_6-ZB-Codanin-1 in the blue box ($\sim$ 48 kDa).
Lt: lysate; FT: flow-through; M: molecular weight marker in kDa.

of the N-terminal His_6 tag. The His-tag on C15orf41 was retained to allow for subsequent complex formation.

To form the C15orf41-Codanin-1 complex, a nickel rebind step was performed where the purified His6-C15orf41 was immobilised on the nickel affinity column

and an excess of untagged Codanin-1 (Codanin-1 fractions after TEV cleavage) added to the same column to allow for complex formation on the column. The complex is then co-eluted in low salt elution buffer (50 mM HEPES pH 7.5, 150 mM NaCl, 5% glycerol, 1 mM TCEP and 300 mM imidazole) and samples for each step outlined above were collected and run on an SDS-PAGE to check for C15orf41-Codanin-1 presence and purity (Figure 5.2 a&b).

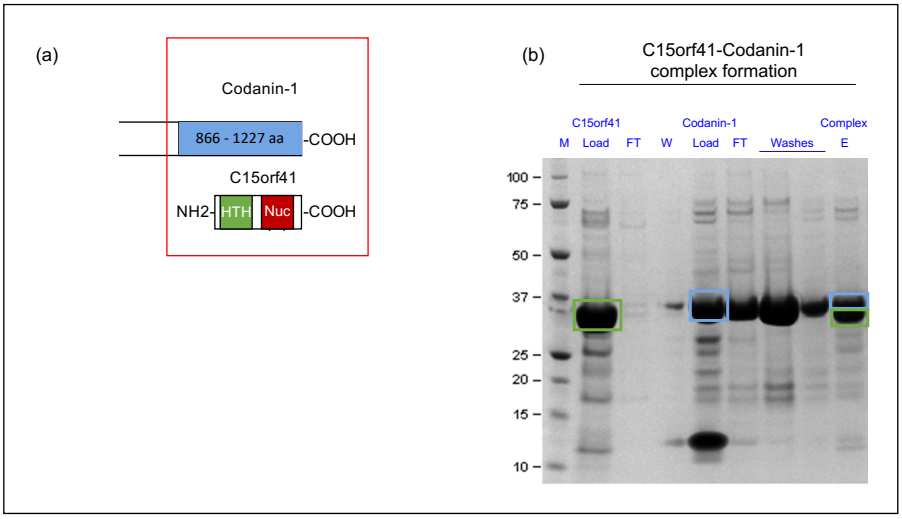


Figure 5.2: C15orf41-Codanin-1 complex formation in the nickel-rebind step
(a) Schematic of the proteins included in the C15orf41-Codanin-1 complex formed in the nickel-rebind step.
(b) SDS-PAGE of purified C15orf41-Codanin-1 complex. Purified His_6-C15orf41 (C15orf41 Load) was immobilised on a nickel affinity column and excess of untagged Codanin-1 (Codanin-1 Load) added to the same column to allow for complex formation on the column. The purified complex was elution as shown in Complex E lane.
Purified His_6-C15orf41 is indicated in the green box ($\sim$ 35 kDa) and purified untagged Codanin-1 in the blue box ($\sim$ 39 kDa).
FT: flow-through; W: Washes; E: Elution; M: molecular weight marker in kDa.

The purified complex was concentrated and further purified using a Superdex 200 size exclusion chromatography. Fractions containing purified C15orf41-Codanin-1 were identified by absorbance at A280 nm and collected, concentrated and snap frozen in liquid nitrogen for later use. The presence of C15orf41-Codanin-1 was confirmed via SDS-PAGE and ESI-Q-TOF MS (Figure 5.3).

C15orf41 is predicted to be a new member of the PD-(D/E) XK superfamily of nucleases, characterised by a highly-conserved active site motif, namely an aspartic acid, glutamic acids and lysine (Kosinski, Feder, and Bujnicki, 2005).

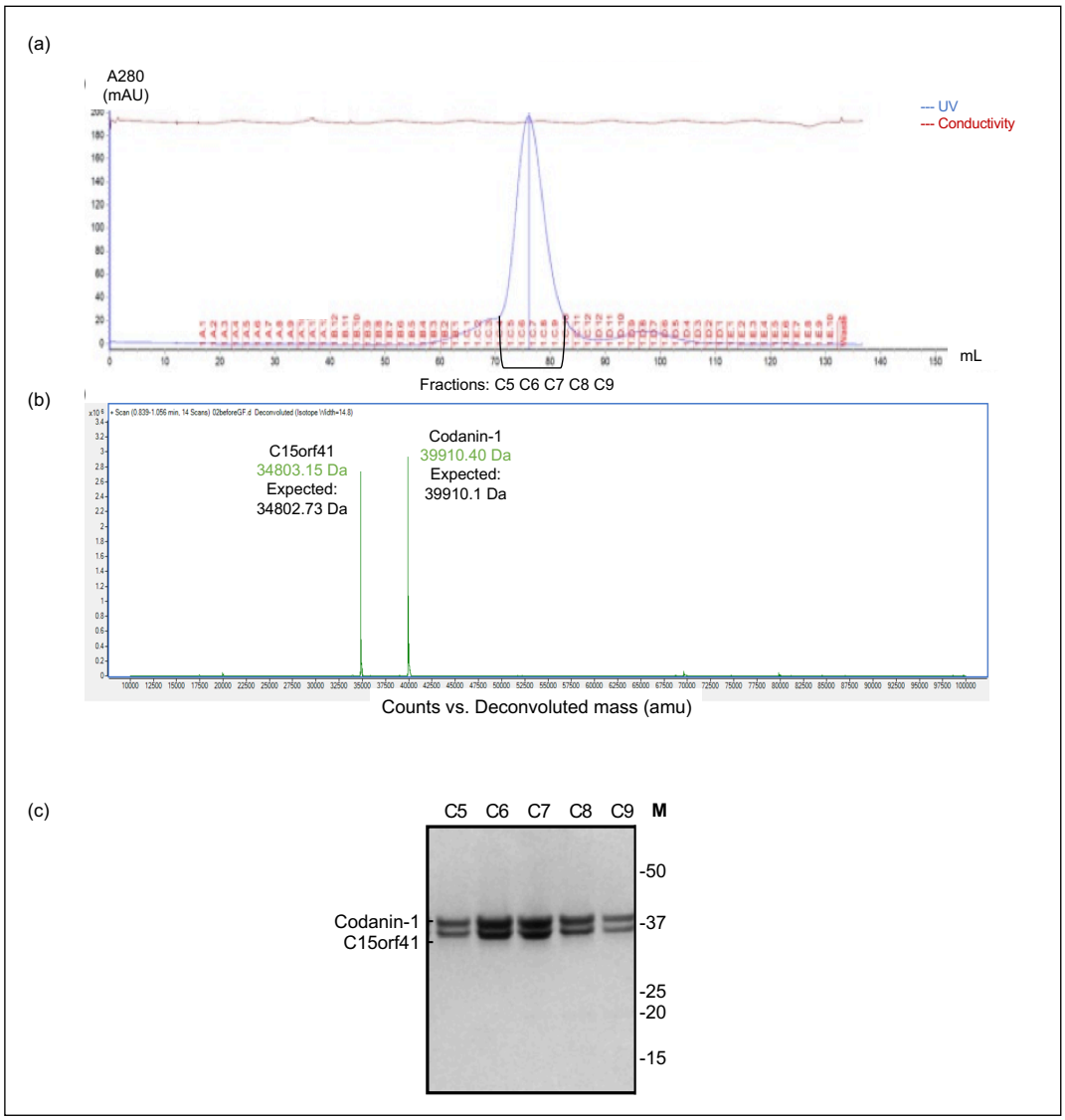


Figure 5.3: Wild-Type C15orf41-Codanin-1 complex purification
(a) Superdex 200 gel filtration trace of wild-type C15orf41-Codanin-1 complex. C15orf41-Codanin-1 elutes in a single peak confirming complex formation.
(b) Intact mass spectrometry analysis of C15orf41-Codanin-1 complex. The chromatogram shows the observed mass of all proteins is within the range of the calculated mass.
(c) SDS-PAGE of fractions across the gel filtration peak confirms the presence of pure C15orf41 and Codanin-1 in fractions across the gel filtration peak (C5-C9).
Purified His_6-C15orf41: $\sim$ 35 kDa; purified untagged Codanin-1: $\sim$ 39 kDa.
M: molecular weight marker in kDa.

In C15orf41, those catalytic residues are E166, D196, E214 and K216 at the C-terminus of the protein, forming a conserved, acidic residue core, likely involved in metal ion coordination and substrate catalysis as seen in other members of the PD-(D/E) XK superfamily (Steczkiewicz et al., 2012). Therefore, in order

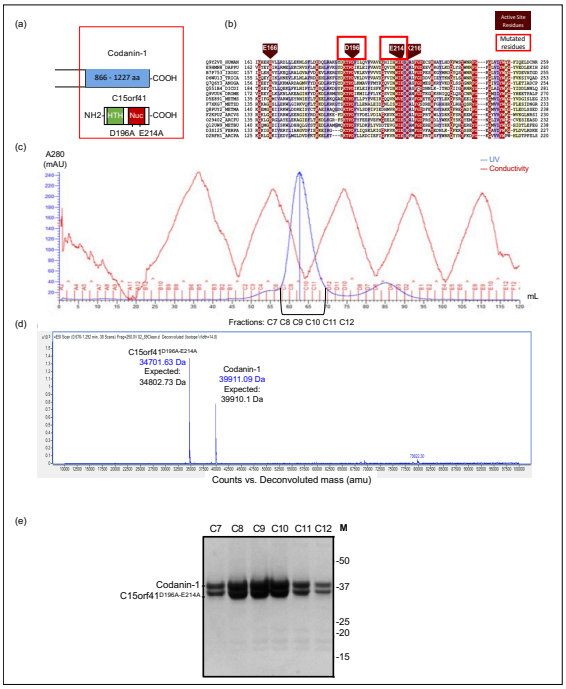


Figure 5.4: Predicted 'nuclease-inactive' C15orf41$^{D196A\text{-}E214A}$-Codanin-1 complex purification
(a) Schematic of the proteins present in C15orf41$^{D196A\text{-}E214A}$-Codanin-1 complex.
(b) Multiple sequence alignments showing conserved residues of the PD-(D/E) XK domain of C15orf41 from the following species: Homo sapiens; B7P753_IXOSC, Ixodes scapularis; E9HMN8_DAPPU, Daphnia pulex; D6WU13_TRICA, Tribolium castaneum; Q7Q6Y3_ANOGA, Anopheles gambiae; Q9VUU6_DROME, Drosophila melanogaster; Q8PUY2_METMA, Methanosarcina mazei; D5E891_METMS, Methanohalophilus mahii; D3S125_FERPA, Ferroglobus placidus; O29402_ARCFU, Archaeoglobus fulgidus; and, D2RFH1_ARCPA, Archaeoglobus profundus. (Babbs et al., 2013). The active site residues are highly conserved (in red).
(c) Superdex 200 gel filtration trace of C15orf41$^{D196A\text{-}E214A}$-Codanin-1 complex. C15orf41$^{D196A\text{-}E214A}$-Codanin-1 elutes in a single peak confirming complex formation.
(d) Intact mass spectrometry analysis of C15orf41$^{D196A\text{-}E214A}$-Codanin-1 complex. The chromatogram shows the observed mass of all proteins is within the range of the calculated mass.
(e) SDS-PAGE of fractions across the gel filtration peak confirms the presence of pure C15orf41$^{D196A\text{-}E214A}$ and Codanin-1 in fractions across the gel filtration peak (C7-C12). Purified His$_6$-C15orf41$_{D196A\text{-}E214A}$: $\sim$ 35 kDa; purified untagged Codanin-1: $\sim$ 39 kDa. HTH: helix-turn-helix domains; Nuc: predicted nuclease domain; M: molecular weight marker in kDa.

to generate the predicted 'nuclease-inactive' C15orf41 protein, full-length His$_6$-C15orf41 with two-point mutations in the predicted catalytic core, D196A and E214A, was co-purified with a C-terminal fragment of Codanin-1. By altering two key, charged residues to non-reactive alanine residues, we aim to disrupt metal ion

coordination within the PD-(D/E) XK catalytic core of C15orf41, thereby yielding a 'nuclease-inactive' version of the complex, an important control for biochemical activity assays. The presence of C15orf41^{D196A,E214A}-Codanin-1 was confirmed via SDS-PAGE and ESI-Q-TOF MS (Figure 5.4).

Additionally, complexes harbouring patient mutations in C15orf41-Codanin-1 were also generated. The CDA-I founder mutations Y94C and L178Q were introduced into C15orf41 and C15orf41^{Y94C} and C15orf41^{L178Q} patient mutant proteins co-purified with a C-terminal fragment of Codanin-1 to obtain stable complexes purified to near homogeneity (Figure 5.5). Finally, samples of each individual proteins generated, specifically wild-type C15orf41, wild-type Codanin-1 as well as C15orf41^{Y94C} and C15orf41^{L178Q}, were also collected prior complex formation in order to test the activity of each individual subunit (see Chapter 6).

5.2 Biophysical characterisation of C15orf41-Codanin-1

C15orf41-Codanin-1 is an uncharacterised protein complex with no current ascribed function. To understand some properties of this complex, a range of biophysical assays were undertaken to determine the complex's native conformation and molecular properties. To this end, multi-angle laser light scattering (SEC-MALS) and small angle X-ray scattering (SAXS) were performed on purified wild-type C15orf41-Codanin-1 complex and results described in the next section.

5.2.1 SEC-MALS

Very little is known about the properties of the C15orf41-Codanin-1 complex. Native size exclusion chromatography experiments using whole-cell HEK293 cell extracts with proteins at endogenous levels suggest the complex elutes at a size close to 670 kDa suggesting oligomerisation of the complex or the presence of yet unknown interaction partners (Shroff et al., 2020). In order to investigate this hypothesis, size exclusion chromatography combined with multi-angle laser light scattering (SEC-MALS) was performed in collaboration with Dr. D. Staunton at the Department of

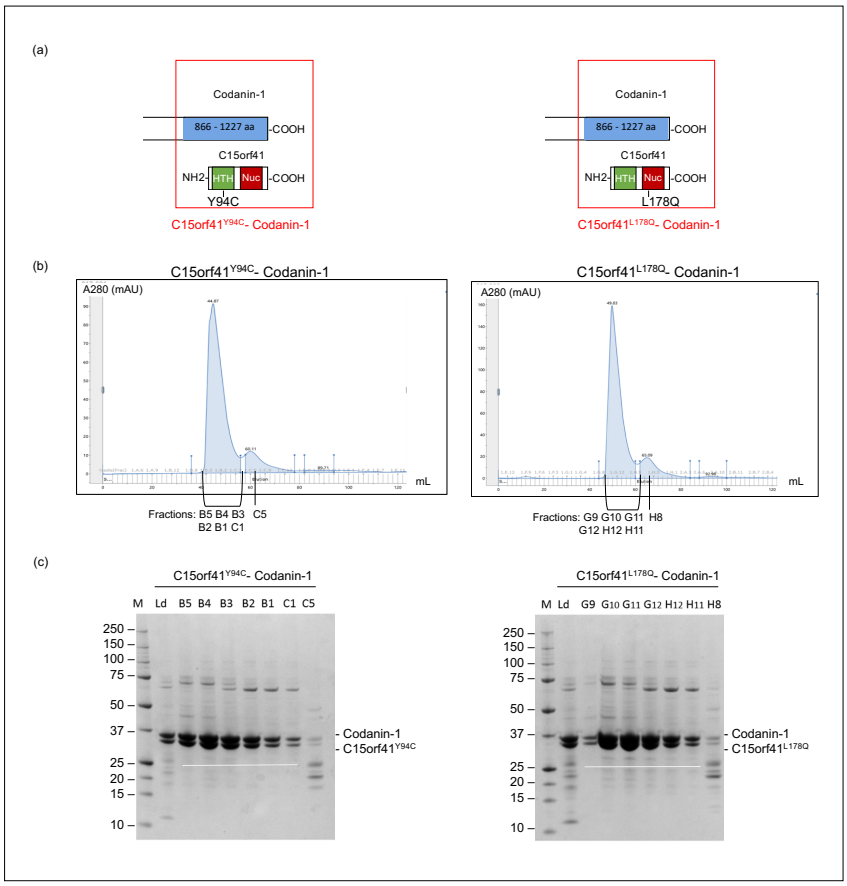


Figure 5.5: Patient Mutant C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 complexes purification
(a) Schematic of the C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 complexes. Point mutations Y94C and L178Q were introduced in the predicted helix-turn-helix and nuclease domains respectively to create patient mutant complexes.
(b) Superdex 200 gel filtration trace of patient mutant C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 complexes. C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 elute in a single peak confirming complex formation.
(c) SDS-PAGE of fractions across the gel filtration peak confirm the presence of pure C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 in fractions across the gel filtration peak (B5-C1 for C15orf41^{Y94C}-Codanin-1 and G9-H11 for C15orf41^{L178Q}-Codanin-1). Purified His$_6$-C15orf41^{Y94C} and His$_6$-C15orf41^{L178Q}: $\sim$ 35 kDa; purified untagged Codanin-1: $\sim$ 39 kDa ; Ld: Load into gel filtration column; M: molecular weight marker in kDa.
HTH: helix-turn-helix domains.
Nuc: predicted nuclease domain.

Biochemistry, University of Oxford. By combining size-exclusion chromatography with multi-angle laser light scattering, we aim to obtain information the molecular weight and stoichiometry of the C15orf41-Codanin-1 complex.

Wild-type C15orf41-Codanin-1 complex was analysed in 25 mM HEPES pH

7.5, 150 mM NaCl, 5% Glycerol and 2 mM TCEP on a Superose 6 HR10/30 analytic column coupled to a light scattering detector and refractive index monitor at a final concentration of 1 mg/mL. Results suggest the C15orf41-Codanin-1 complex is present in different oligomerisation states corresponding to the two peaks observed: one peak corresponds to a complex of one Codanin-1 subunits for two C15orf41 subunits (around 120 kDa, elution volume 32 mL) whilst the second peak suggest this complex dimerises to form a supercomplex of two Codanin-1 and four C15orf41 molecules (around 280 kDa, elution volume 27 mL) (Figure 5.6.a). Increasing concentrations of complex at 0.5 mg/mL, 1 mg/mL and 2 mg/mL were also tested. In all cases, two peaks were observed again, suggesting the presence of two oligomerisation species (Figure 5.6.b). Whether the oligomerisation states are concentration dependent remains ambiguous and further analyses would be required to determine this. It also should be noted that high levels of aggregation were observed during the analyses, most likely due to prior freeze/thawing of the proteins. In fact, when fresh and frozen C15orf41-Codanin-1 complex were run on a size-exclusion column alongside Biorad size standards, a shift in oligomerisation state was also observed. While the fresh protein complex eluted at around 120 kDa, suggesting the presence of two C15orf41 molecules for each Codanin-1 molecule, the previously-frozen complex eluted at a much higher molecular weight, corresponding to either dimerization or further oligomerisation of the 2:1 C15orf41-Codanin-1 complex (Figure 5.6.c). The results should, therefore, be interpreted with some caution and where possible, repeated with fresh protein to avoid freeze/thaw cycles.

5.2.2 SAXS

Although preliminary, the SEC-MALS results suggest C15orf41-Codanin-1 form a complex at a 2:1 ratio that has the ability to oligomerise to a 4:2 ratio (C15orf41: Codanin-1), and potentially even to a higher oligomeric state. In order to further investigate this, small angle X-ray scattering (SAXS) was performed with Dr. Y. Yosaatmadja at the Centre for Medicines Discovery, University of Oxford as part of

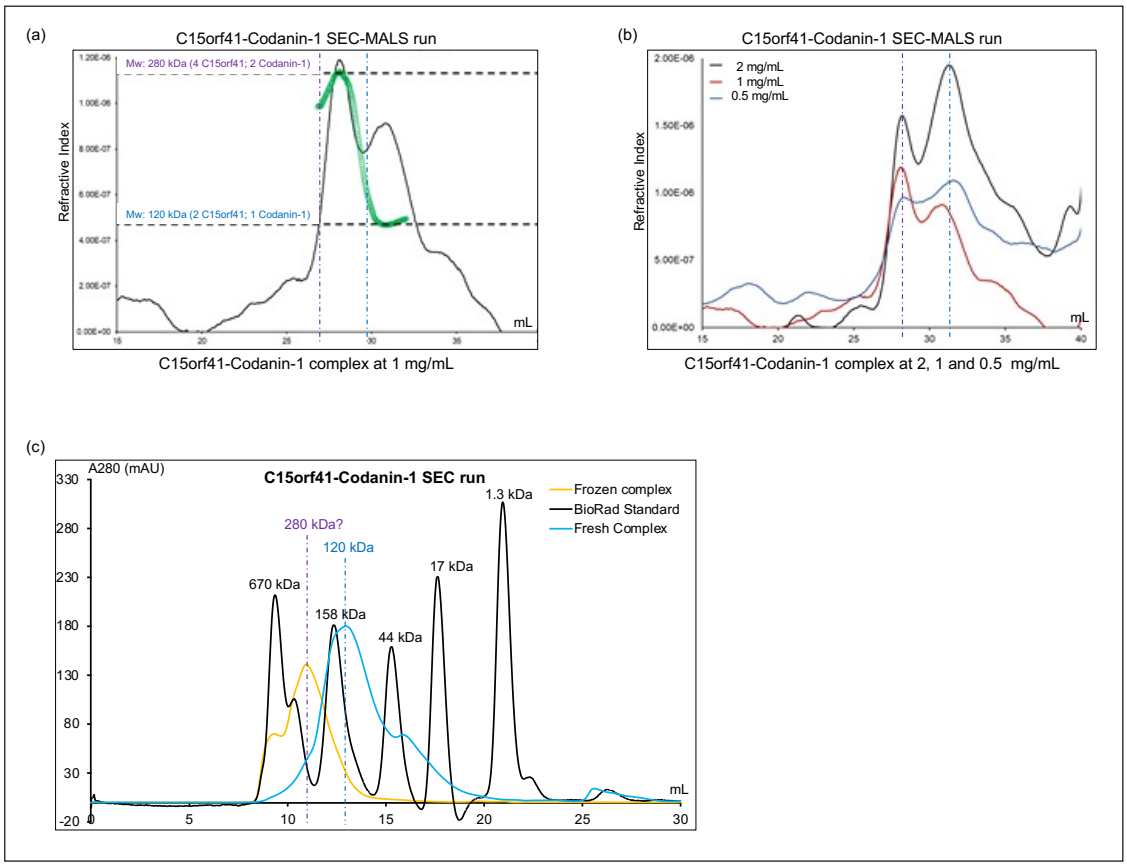
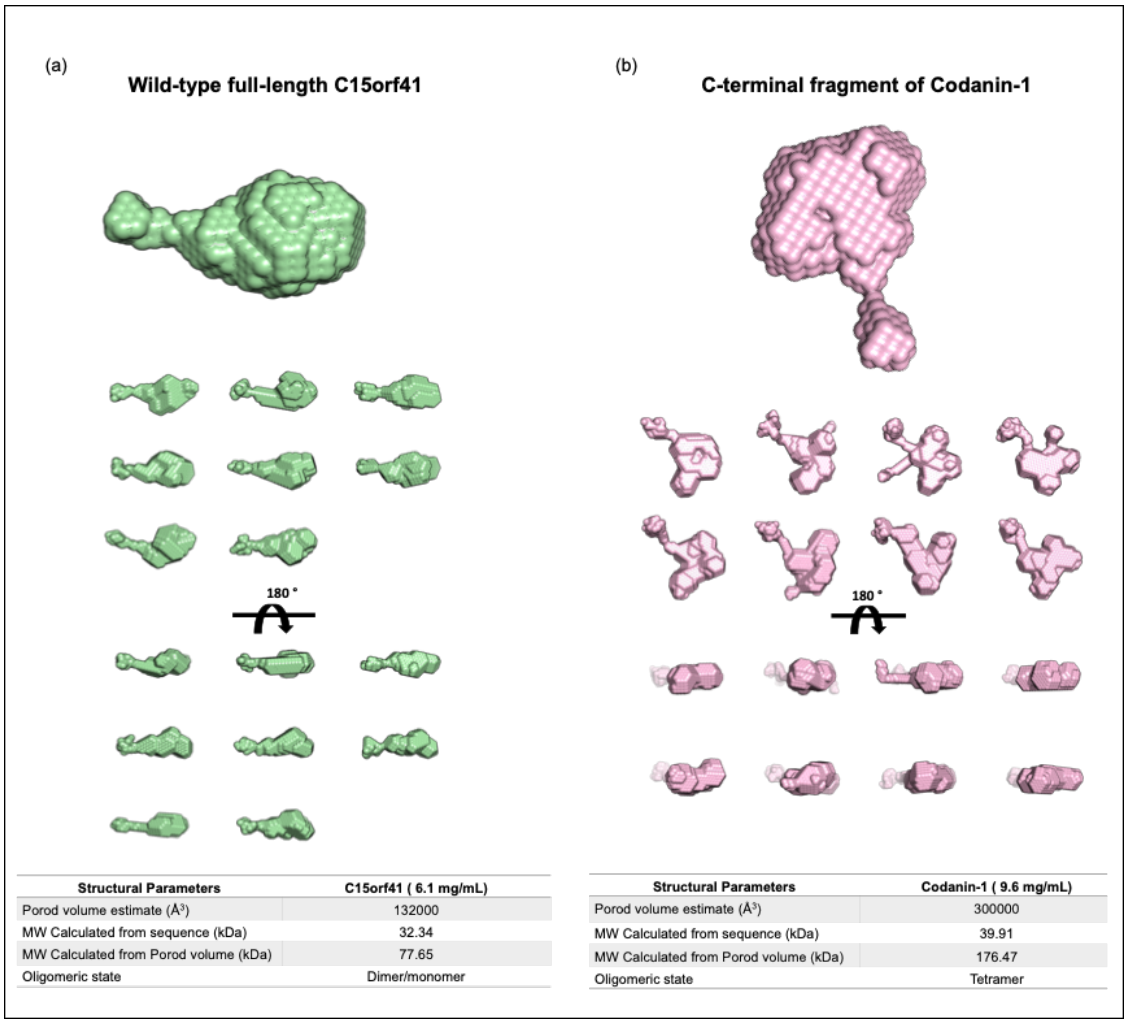


Figure 5.6: SEC-MALS studies of C15orf41-Codanin-1
(a) Trace of wild-type C15orf41-Codanin-1 complex at 1 mg/mL, run on a Superose 6 HR10/30 analytic column coupled to a light scattering detector and refractive index monitor. Two peaks are observed, corresponding to two oligomerisation states of C15orf41-Codanin-1: the complex at a 2:1 ratio of C15orf41-Codanin-1 (120 kDa, blue line) and at a 4:2 ratio (280 kDa, purple line).
(b) The same two peaks are observed when a serial dilution of C15orf41-Codanin-1 (2, 1 and 0.5 mg/mL) is run on SEC-MALS suggesting the complex has two different oligomeric states.
(c) The presence of various oligomeric states is likely an artifact of the freeze/thaw process. When run on SEC with molecular weight standards, fresh C15orf41-Codanin-1 complex elutes at around 120kDa but frozen complex appears to form higher molecular weight species. Results should therefore be interpreted with care and repeated with fresh protein for more conclusive data.

our collaboration. SAXS is a method to determine overall structure or shape of a molecule, providing low-resolution structural information and global size.

In brief, C15orf41-Codanin-1 complex as well as individual proteins were purified as described above. Purified proteins were concentrated between 6 mg/mL and 10 mg/mL in 25 mM HEPES pH 7.5, 100 mM NaCl, 1% Glycerol and 1 mM TCEP,

mounted on the beamline and images collected. The raw data then analysed and refined by Dr. Y. Yosaatmadja using the ATSAS (2.7.1) suite of software packages.



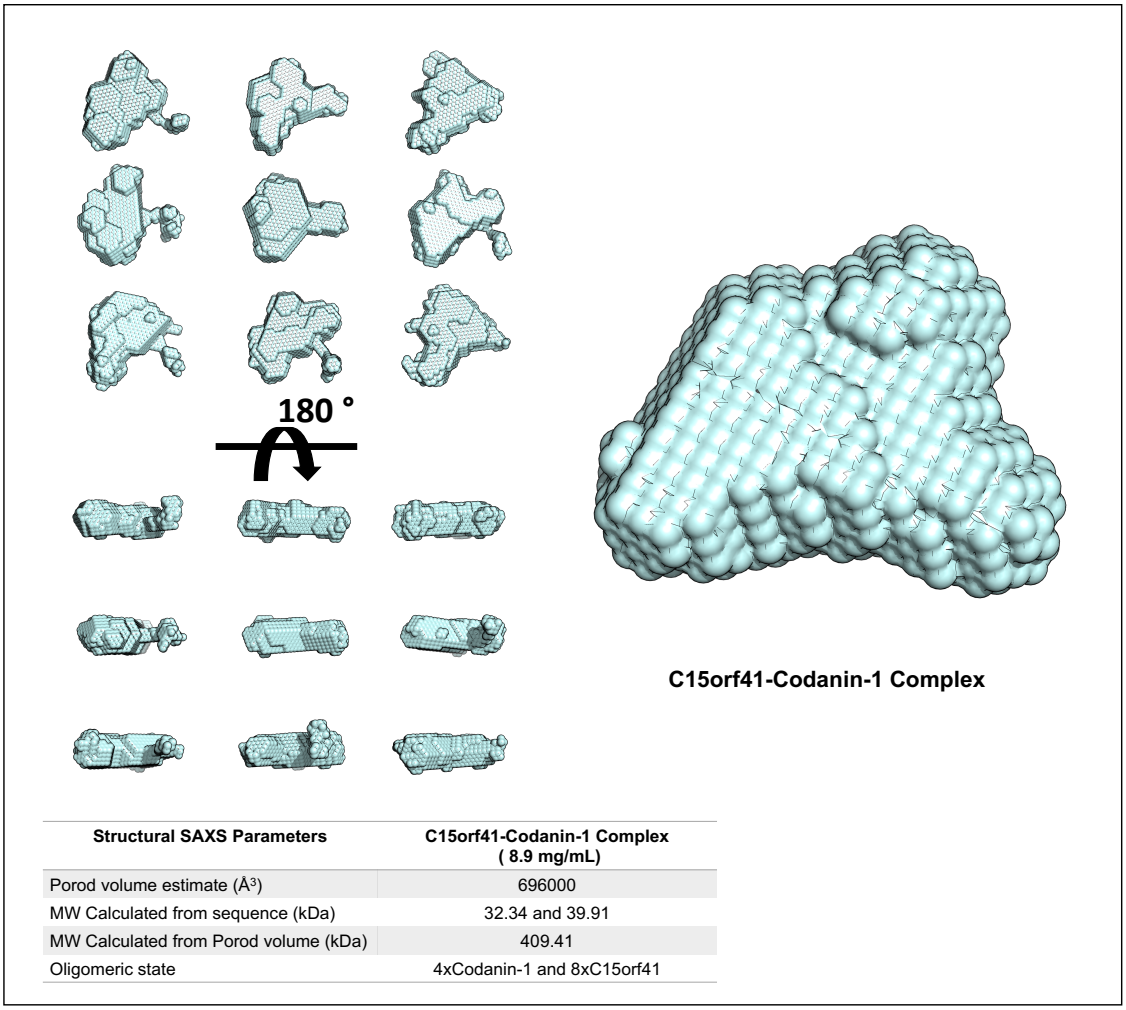
Structural Parameters	C15orf41 (6.1 mg/mL)
Porod volume estimate (Å^3)	132000
MW Calculated from sequence (kDa)	32.34
MW Calculated from Porod volume (kDa)	77.65
Oligomeric state	Dimer/monomer

Structural Parameters	Codanin-1 (9.6 mg/mL)
Porod volume estimate (Å^3)	300000
MW Calculated from sequence (kDa)	39.91
MW Calculated from Porod volume (kDa)	176.47
Oligomeric state	Tetramer

Figure 5.7: SAXS model of wild-type C15orf41 and Codanin-1
(a) SAXS model of full-length wild-type His_6-C15orf41. His_6-C15orf41 is predicted to form a dimer, folding into a compact, globular shape.
(b) SAXS model of the C-terminal region of Codanin-1. Codanin-1 is present in tetrameric form, with protrusions likely reflecting the more flexible and dynamic state of this protein.

Initially, SAXS on C15orf41 alone was performed (Figure 5.7.a). Results revealed C15orf41 forms a dimer (approximate molecular weight calculated was 77 kDa which is close to a dimer predicted at 68 kDa) that folds into relatively flat, globular and compact shape. The extension observed on the SAXS structure could possibly represent the His_6 tag still present on C15orf41 although still remains to be investigated. Next, the structure of Codanin-1 was investigated. Codanin-1 SAXS

analysis revealed Codanin-1 is present in a tetrameric form, with each subunit stacking up into a flat shape (Figure 5.7.b). There is also a protrusion present, likely due to some protein unfolding, flexibility or reflecting a dynamic oligomerisation state leading to a degree of protein heterogeneity.



Structural SAXS Parameters	C15orf41-Codanin-1 Complex (8.9 mg/mL)
Porod volume estimate (Å^3)	696000
MW Calculated from sequence (kDa)	32.34 and 39.91
MW Calculated from Porod volume (kDa)	409.41
Oligomeric state	4xCodanin-1 and 8xC15orf41

Figure 5.8: SAXS model of C15orf41-Codanin-1 complex
SAXS model of full-length wild-type His_6-C15orf41 complexed with the C-terminal region of Codanin-1. The proteins form a super-complex with 8xC15orf41 and 4xCodanin-1, interlocked into a flat, stacked shape.

Finally, C15orf41-Codanin-1 complex SAXS results were analysed (Figure 5.8). As predicted from the calculations of the porod volume obtained (approximate molecular weight is 400 kDa), C15orf41-Codanin-1 may form a super-complex which oligomerises to an 8:4 ratio, likely a dimer of the 4:2 C15orf41-Codanin-1 complex identified by SEC-MALS. The super-complex forms a very flat structure, with each

subunit interlinking together resembling a stack. During data collection, it was noted that some heterogeneity in the sample was present, likely reflecting protein unfolding and dynamic oligomerisation states.

Altogether, the SAXS data revealed C15orf41-Codanin-1 forms a large, flat, mostly tight complex with some degree of flexibility and unfolding present. To confirm these results, SAXS should be repeated using fresh protein to minimise the protein unfolding and heterogeneity in oligomerisation states. Again, the caveat of using a previously frozen protein batch needs to be considered. Only large, oligomerised species were seen by SAXS (around 400 kDa, suggesting a 8:4 ratio of C15orf41 to Codanin-1), which could be an artefact from using frozen protein. Yet, the results still allow to visualise the C15orf41-Codanin-1 complex for the very first time, providing a foundation for further structural studies.

5.3 Attempts at structural studies of C15orf41-Codanin-1

Before this thesis, the C15orf41-Codanin-1 structure remained completely unexplored. Whilst secondary structure predictions indicate C15orf41 may contain winged helix-turn-helix (wHtH) domains followed by a PD-(D/E) XK -like nuclease domain, Codanin-1 is predicted to mainly be a disordered protein (Hannah Tamary et al., 2010; Babbs et al., 2013; Swickley et al., 2020). Nucleases notoriously present a challenge to structural studies due to their intrinsically flexible active site structure that allows processing on a variety of nucleic acid structures. This section will focus on the different strategies employed for C15orf41-Codanin-1 structure determination.

5.3.1 X-Ray crystallography

5.3.1.1 Crystallography screen set-up

For crystallisation trials, various concentrations of purified C15orf41-Codanin-1 complex were tested in a variety of crystallisation screens using the sitting drop vapour diffusion method. To this end, C15orf41-Codanin-1 was concentrated between 7.5 mg/mL and 15 mg/mL and drops set up in a variety of broad

crystallisation screens at both 4°C and 20°C to cover a maximum number of different conditions with varying polyethylene glycol (PEG), salt, neutralised organic acids or organic precipitants concentrations across a pH range from 4.0 to 10.5. Five core screens from the Structural Genomics Consortium at the University of Oxford were used to screen for potential hits: the Hampton Research Index Screen, Hampton Research Crystal Screen, JCSG+ Screen, Morpheus Screen and the Basic ChemSpace Screen. For crystallisation, 50-100 nL of C15orf41-Codanin-1 complex was mixed with 100-150 nL reservoir solution and plates incubated at 4°C and 20°C. Unfortunately, no crystals were obtained in any of the screens, possibly due to the flexible nature of nucleases and heterogeneity of oligomerisation states of the complex. Further optimisation will be required, in particular co-crystallisation with a suitable substrate or the generation of truncated C15orf41 protein should be considered to maximise crystal formation.

5.3.1.2 Expression in LOBSTR E. coli strain

One possible explanation for the lack of crystals obtained is the presence of impurities in the C15orf41-Codanin-1 preparation. In support of this hypothesis, SDS-PAGE analysis of purified C15orf41-Codanin-1 complex shows the presence of additional bands (Figure 5.9.a). MS analysis of those contaminating bands reveals the presence of a common *E. coli* contaminant (likely to be heat-shock protein 70 (Hsp70) at 70 kDa) and the bifunctional polymyxin resistance protein ArnA at 74 kDa, both of which have been previously reported to co-purify with proteins derived from transformation into *E. coli* BL21 Rosetta2 cells. Therefore, C15orf41 and Codanin-1 were re-expressed in the optimised *E. coli* expression strain LOBSTR. The low background strain has been optimised to minimise common contaminants derived from His-tag purifications (Andersen, Leksa, and Schwartz, 2013). Test expressions of both C15orf41 and Codanin-1 using the LOBSTR strain were performed and compared to expression in BL21 cells, confirming successful expression of both proteins at higher purity. For both C15orf41 and Codanin-1, the initial nickel resin purification step was performed and samples from each step (including lysate,

flow-through, different washed and elution) collected and analysed by SDS-PAGE to visualise product formation (Figure 5.9.b). In both cases, using the LOBSTR strain increased the expression of the proteins with less contaminating protein bands visible, particularly noticeable for C15orf41. Unfortunately, a full-scale purification and subsequent complex formation was never performed due to time constraints but this represents an avenue for future experiments with a more optimised purification method.

5.3.2 Negative-staining EM

Considering C15orf41 and Codanin-1 form a relatively large protein complex depending on its oligomerisation state (>200 kDa if a 4:2 ratio of C15orf41:Codanin-1), cryo-electron microscopy (cryo-EM) studies were undertaken to obtain further structural data on the complex. To start, negative-staining EM was completed. This was performed with the help of Dr. J. Gurnett at the MRC Weatherall Institute of Molecular Medicine.

Purified C15orf41-Codanin-1 complex was serially diluted in 50 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP and 5% Glycerol to concentrations ranging from 0.03 mg/mL to 0.06 mg/mL and applied to glow discharged carbon film grids. Grids were then washed and stained using uranyl acetate before visualisation on a screening microscope. Grids were screened on a FEI Technai 12 TEM in the Sir William Dunn School of Pathology Bioimaging Facility, University of Oxford. Images were taken throughout the screening process to assess the staining level and particle distribution in chosen areas on a grid (Figure 5.10.a). While some grids were too concentrated and showed too high heterogeneity within the particles, grids with optimal concentration and homogenous particles were chosen for data collection. In total, 200 micrographs were collected per grid and used as the input for data processing, performed by Dr. J. Gurnett. In summary, particles were manually picked using the built in RELION3.1 particle picker and several rounds of 2D and 3D class averages were generated using RELION3.1 to discard bad particles. In total, 10,225 particles were classified into 2D and 3D averages and further refined

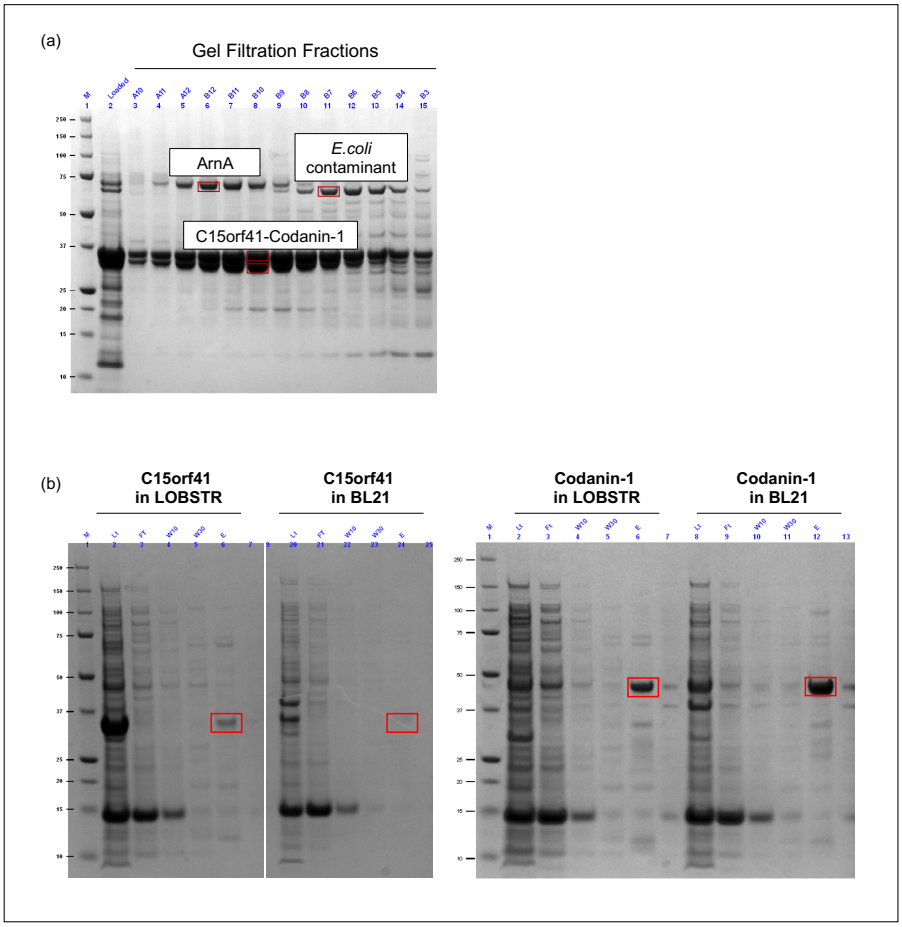


Figure 5.9: Optimisation of C15orf41-Codanin-1 complex purification
(a) SDS-PAGE of purified C15orf41-Codanin-1 complex showing the presence of additional contaminating bands from the purification process. Fractions of the purified complex were collected during gel filtration and analysed alongside the gel filtration input (termed 'Loaded'). The nature of the products present was determined by MS.
Following bands are visible: ArnA at 74 kDa, a common *E. coli* contaminant at 70 kDa and our complex of interest, C15orf41 (34.8 kDa) and Codanin-1 (39.9 kDa).
(b) Test expression of C15orf41 and Codanin-1 comparing expression in the LOBSTR versus BL21 *E. coli* strains. A small-scale nickel resin batch purification was performed on both C15orf41 and Codanin-1 and expression levels and purity assessed by SDS-PAGE. Expression in the LOBSTR strain improved the yield and purity of both C15orf41 and Codanin-1.
Lt: lysate; FT: flow-through; W10 and W30: washes with increasing imidazole concentrations; E: elution; M: molecular weight marker in kDa.

into a final 3D model (Figure 5.10 b, c&d). This provisional negative staining-EM model of C15orf41-Codanin-1 reveals our proteins of interest form a very large tight complex, in a similar flat shape as suggested by the SAXS model. The complex is predicted to be around 420 kDa, suggesting the species seen by SEC-MALS

are able to further oligomerise and form a super-complex with eight C15orf41 and four Codanin-1 subunits. These data are in agreement with the SAXS model of C15orf41-Codanin-1, where this same large oligomeric species was observed.

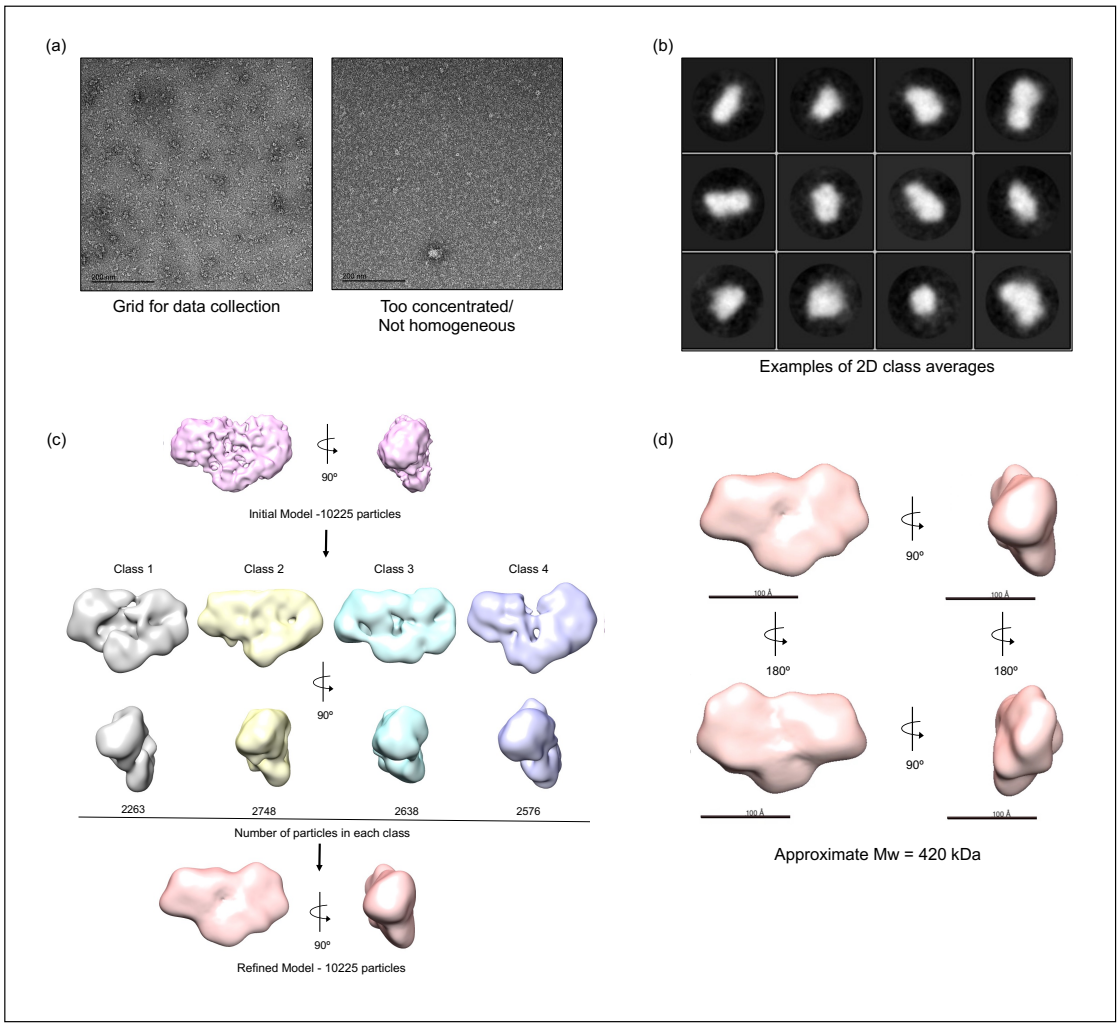


Figure 5.10: Negative staining EM studies of C15orf41-Codanin-1
(a) Examples of micrographs taken of C15orf41-Codanin-1 negative-staining grids.
(b) 2D class averages of models of C15orf41-Codanin-1 generated using RELION3.1.
(c) 3D class averages of models generated of C15orf41-Codanin-1 generated using RELION3.1.
(d) Final refined 3D model of C15orf41-Codanin-1 that was generated from the initial 10 225 particles picked. The complex forms a flat, compact shape. The predicted volumed calculated was 713 900 A3 which relates to a predicted molecular weight of 420 kDa. This suggests C15orf41-Codanin-1 forms a super-complex with 8 C15orf41 and 4 Codanin-1 subunits.

Whilst not conclusive, the negative staining data obtained complements the results obtained by SEC-MALS and SAXS and reveals C15orf41-Codanin-1 is likely

forming a complex with different oligomerisation states. Some degree of disparity was also observed during grid screening, likely a caveat of using frozen protein leading to heterogenous mix of protein species. Importantly, it therefore should be noted that experiments using fresh proteins will need to be conducted to eliminate potential confounding results due to freeze/thawing before cryo-EM studies are performed.

5.4 Conclusion

Altogether, successful purification of the C15orf41-Codanin-1 complex, as well as of catalytically-inactive and patient-mutant versions was achieved to near homogeneity, allowing further biochemical assays to be discussed in Chapter 6. Currently, the data suggest C15orf41-Codanin-1 is able to form a complex with two C15orf41 molecules and one Codanin-1 molecule. This complex has the capacity to further oligomerise, forming larger species with up to eight C15orf41 molecules and four Codanin-1 molecules that fold into a flat, stacked shape as seen by both SAXS and negative-staining EM. It should be noted that the experiments outlined above were undertaken during the COVID-19 pandemic, thereby limiting the extent of the assays able to be performed. We are also aware of the limitations of using frozen protein and low/no repeats but believe these results offer a first basis into C15orf41-Codanin-1 complex properties and structure and should be used as a complement to the cellular and biochemical characterisation of the proteins. Nonetheless, these results provide potentially useful insights into the biophysical and structural characteristic of C15orf41-Codanin-1, which will be further expanded on in Chapter 8.

6

Biochemical characterisation of C15orf41-Codanin-1

When this thesis began, C15orf41 was a completely uncharacterised protein, with no previous reports of assays for biochemical activity. Some clues of potential functions can be found in secondary structure prediction studies. The 34 kDa protein contains 2 N-terminal AraC/XylS-like winged helix-turn-helix (wHtH) domains followed by a PD-(D/E) XK -like nuclease domain and was previously predicted to be a novel nuclease (Babbs et al., 2013).

Thanks to the successful purification of C15orf41 and its binding partner Codanin-1 to near homogeneity (as described in Chapter 5), this chapter will focus on the in-vitro biochemical characterisation of C15orf41 and the C15orf41-Codanin-1 complex. This aims to shed some light on a role for C15orf41 and the complex, not only in the context of CDA-I but also in a broader, more fundamental, biological role.

6.1 C15orf41 is a novel RNA nuclease

In order to determine whether C15orf41 holds any nuclease activity, reminiscent of the PD-(D/E) XK superfamily of nucleases, purified C15orf41 was incubated with a variety of 5′-end terminally radiolabelled DNA and RNA substrates. Increasing concentrations of protein (as indicated) were incubated with 10 nM substrate at 37°C for 60 min in the presence of 50 mM HEPES pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 5% glycerol and 0.5 mM DTT and reactions were subsequently analysed by 20% denaturing PAGE to visualize product formation. Somewhat surprisingly, nucleolytic digestion products were only observed when 20-mer double-stranded

RNA and 20-mer RNA-DNA hybrid were used as substrates (with the RNA strand of the latter being processed). Results show a single cut at concentrations of protein higher than 250 nM, leading to an 18-mer product. This is the first report of any nuclease activity of C15orf41 and suggest this protein might act as an RNA nuclease, with a preference for double-stranded RNA-containing substrates (Figure 6.1).

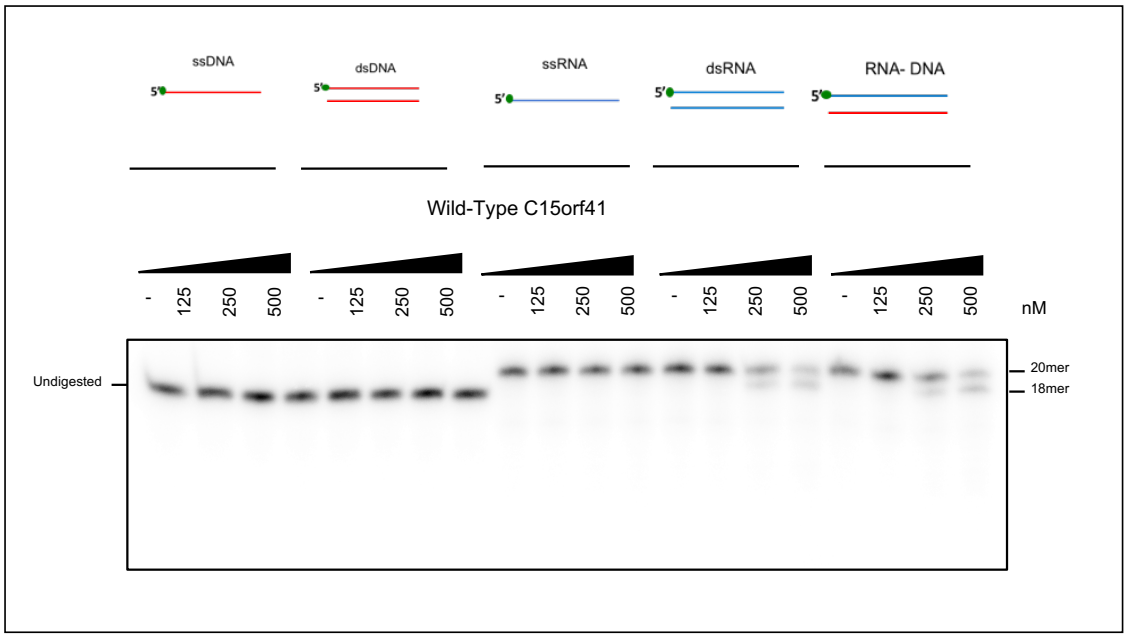


Figure 6.1: C15orf41 is a novel RNA nuclease
Wild-type C15orf41 is a novel RNase able to digest 20-mer dsRNA and the RNA strand of an RNA-DNA hybrid. A single cut 2 nucleotides (nt) from the 3′-end (leading to a 18-mer product) is observed at protein concentrations above 250 nM. No activity is observed on any DNA substrates.
Increasing concentrations of protein (as indicated) were incubated with substrate at 37°C for 60 min; reactions were subsequently analysed by 20% denaturing PAGE to visualize product formation. The size of products was determined as shown in Figure 6.10. Oligonucleotide sequences used are indicated in Tables 3.1 and 3.2 in the Materials and Methods section. All oligonucleotides were 5′-end terminally radiolabelled (in green).

6.1.1 C15orf41 and Codanin-1 form a functional RNA nuclease complex with enhanced activity

Considering the accumulating evidence for formation of a C15orf41-Codanin-1 complex, both *in vivo* and *in vitro*, the nuclease activity of this complex was investigated next. To this end, purified C15orf41 alone as well as C15orf41 complexed with a C-terminal fragment of Codanin-1 (also referred to as the C15orf41-Codanin-1

complex) was incubated with a 20-mer ssRNA and an RNA-DNA hybrid substrate. Interestingly, the addition of the C-terminal fragment of Codanin-1 greatly enhanced and modulated the nuclease activity of C15orf41 on both substrates. The complex is not only able to digest the substrates at lower concentrations than C15orf41 alone (62 nM versus 250 nM), and it also shows activity on ssRNA, a substrate that was not processed by C15orf41 alone (Figure 6.2). On the RNA-DNA hybrid substrate, C15orf41-Codanin-1 shows enhanced activity, processing the RNA-strand beyond a single 2-nt cut to generate a more complex, laddering digestion pattern as the protein concentration increases. Considering that association between C15orf41 and Codanin-1 is disrupted in CDA-I patients bearing mutation L178Q in C15orf41, this is an interesting observation when thinking about a possible link between C15orf41's nuclease activity and the contribution of Codanin-1 to the mechanistic deficit in CDA-I.

The rest of this chapter will focus on a comprehensive biochemical characterisation of the C15orf41-Codanin-1 complex, designed to provide insight into its activity requirements and preferences.

To confirm the observed nuclease activity is intrinsic to the complex, several complementary methods were employed. Firstly, the activity of the predicted 'nuclease-inactive' complex, where key residues D196 and E214 of the PD-(D/E) XK catalytic motif of C15orf41 were mutated (named C15orf41^{D196A,E214A}-Codanin-1), was assessed using the 20-mer RNA-DNA hybrid substrate. Consistent with the PD-(D/E) XK domain being required for RNA hydrolysis, no activity was observed with C15orf41^{D196A,E214A}-Codanin-1, even at concentrations 10-fold higher than the highest concentration employed for wild-type C15orf41-Codanin-1 complex (Figure 6.3). Second, the nuclease activity of each individual component of the complex (C15orf41 alone and Codanin-1 alone) was assessed on the same substrate and C15orf41-Codanin-1 complex was used as a positive control. Nucleolytic digestion was only observed for C15orf41 and the complex, confirming that the enhanced nuclease activity of the complex is not due to any intrinsic nuclease activity of Codanin-1 or any contaminant within that sample (Figure 6.4). Next,

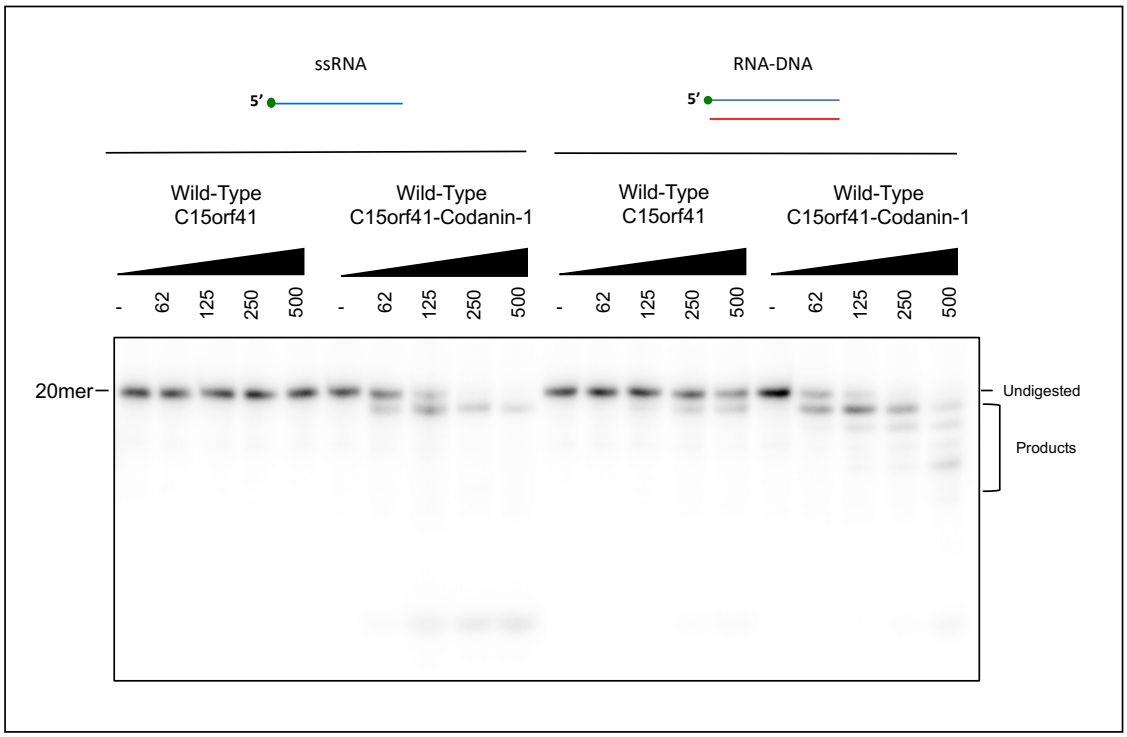


Figure 6.2: C15orf41-Codanin-1 form a RNA nuclease complex with enhanced activity
The addition of a C-terminal fragment of Codanin-1 greatly enhances the activity of C15orf41, both quantitatively and qualitatively. The C15orf41-Codanin-1 complex is now able to digest ssRNA, creating a single cut at protein concentration from 62 nM. The C15orf41-Codanin-1 complex is also able to further nucleolytically process the RNA strand of an RNA-DNA hybrid, creating a laddering pattern of digestion that increases with protein concentration.
Increasing concentrations of protein (as indicated) were incubated with substrate at 37°C for 60 min; reactions were subsequently analysed by 20% denaturing PAGE to visualize product formation. Oligonucleotide sequences used are indicated in Tables 3.1 and 3.2 in the Materials and Methods section. All oligonucleotides were 5′-end terminally radiolabelled (in green).

the elution fractions of wild-type complex obtained from the final gel filtration step of purification (see Chapter 5) were assayed for nuclease activity. As shown in Figure 6.5, the RNAse activity of this complex coincides precisely with the peak of the elution profile, further strengthening the hypothesis that the observed RNA nuclease activity results from activities intrinsic to the C15orf41-Codanin-1 complex, and does not result from co-purification of a contaminant. Finally, the nuclease activity of three, independently purified wild-type complexes was examined on ssRNA and an RNA-DNA hybrid and revealed the nuclease activity observed is

reproducible and equivalent across different purifications (Figure 6.6).

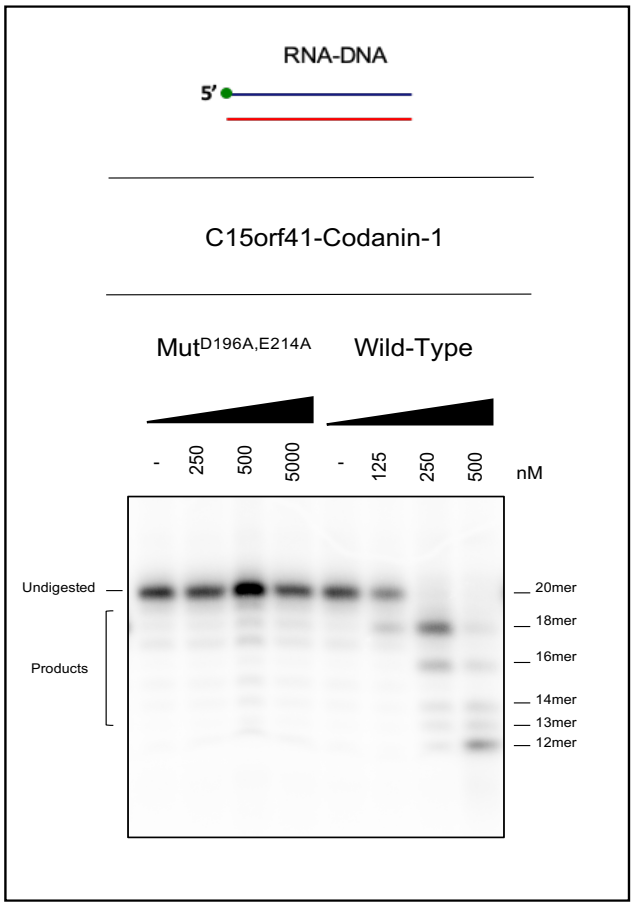


Figure 6.3: Mutant C15orf41^{D196A,E214A}- Codanin-1 does not have any nuclease activity
The activity of C15orf41-Codanin-1 where key residues D196 and E214 of C15orf41 within the PD-(D/E) XK catalytic motif were mutated (named MutD^{196A,E214A}) was assessed on an RNA-DNA hybrid. No nuclease activity was observed, even at concentrations 10-fold higher than for the wild-type complex, confirming the activity observed is intrinsic to C15orf41 and not due to a contaminant. Wild-type C15orf41-Codanin-1 complex was used as a positive control.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

PD-(D/E) XK domain-containing nucleases typically employ divalent cations as an important co-factor. Therefore, the metal ion-dependency of C15orf41-Codanin-1 activity was investigated. To this end, 125 nM of wild-type complex was incubated with increasing concentrations (as indicated) of various divalent metal cations, namely Mg^{2+}, Mn^{2+} and Zn^{2+} on ice for 15 min before adding 10 nM RNA-DNA substrate and incubating the reaction at 37°C for 60 min.

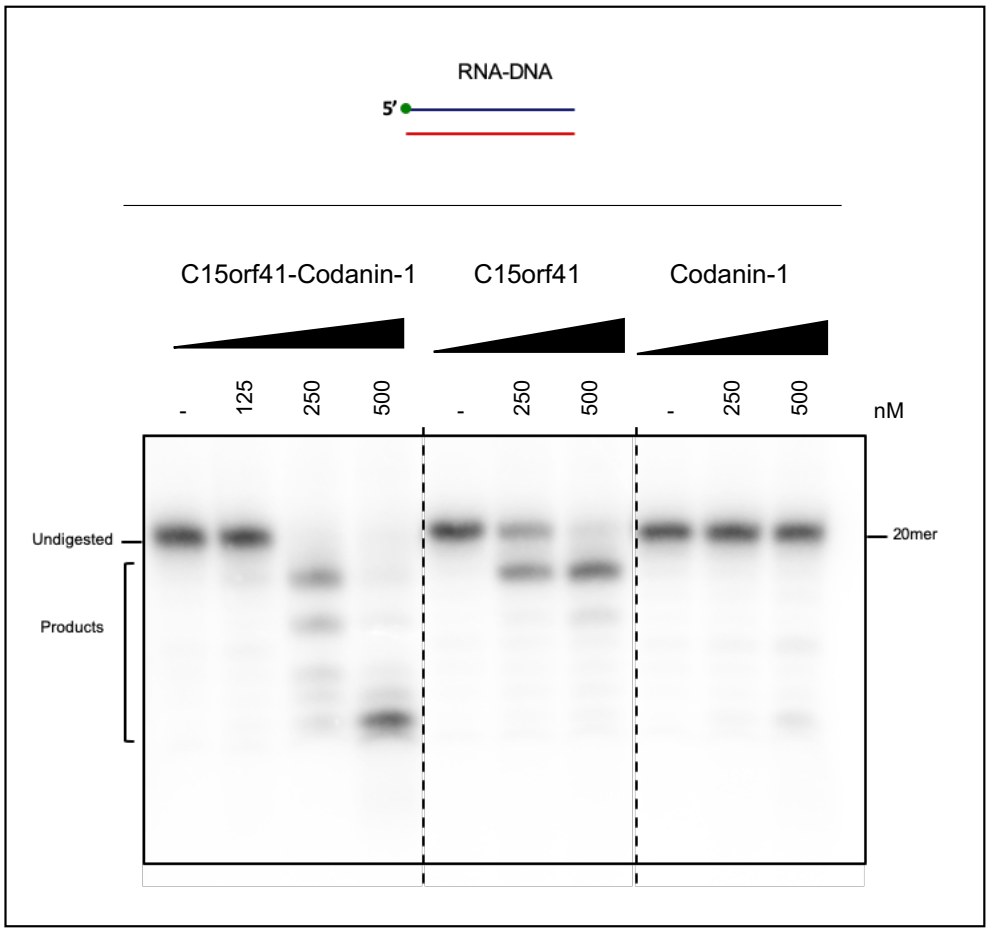


Figure 6.4: Observed nuclease activity is intrinsic to C15orf41 and not Codanin-1
The nuclease activity of C15orf41 and Codanin-1 individually was assessed using an RNA-DNA hybrid substrate. Nuclease activity was only observed for C15orf41 alone and not Codanin-1, confirming that the enhanced nuclease activity of the complex is not due to any intrinsic nuclease activity of Codanin-1 or any contaminant. Wild-type C15orf41-Codanin-1 complex was used as a positive control.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

Reactions were subsequently analysed by 20% denaturing PAGE to visualise product formation. Mutant C15orf41^{D196A,E214A}-Codanin-1 was used as a negative control. All three metal ions promoted the RNase activity of C15orf41-Codanin-1, although qualitatively slightly different (Figure 6.7.a). The requirement for metal ions is further highlighted as EDTA, EGTA and *o*-phenanthroline, three metal-chelating agents, were also inhibitory (Figure 6.7.b). For physiological relevance, 5 mM magnesium (that is more abundant in cells than manganese) was chosen as the

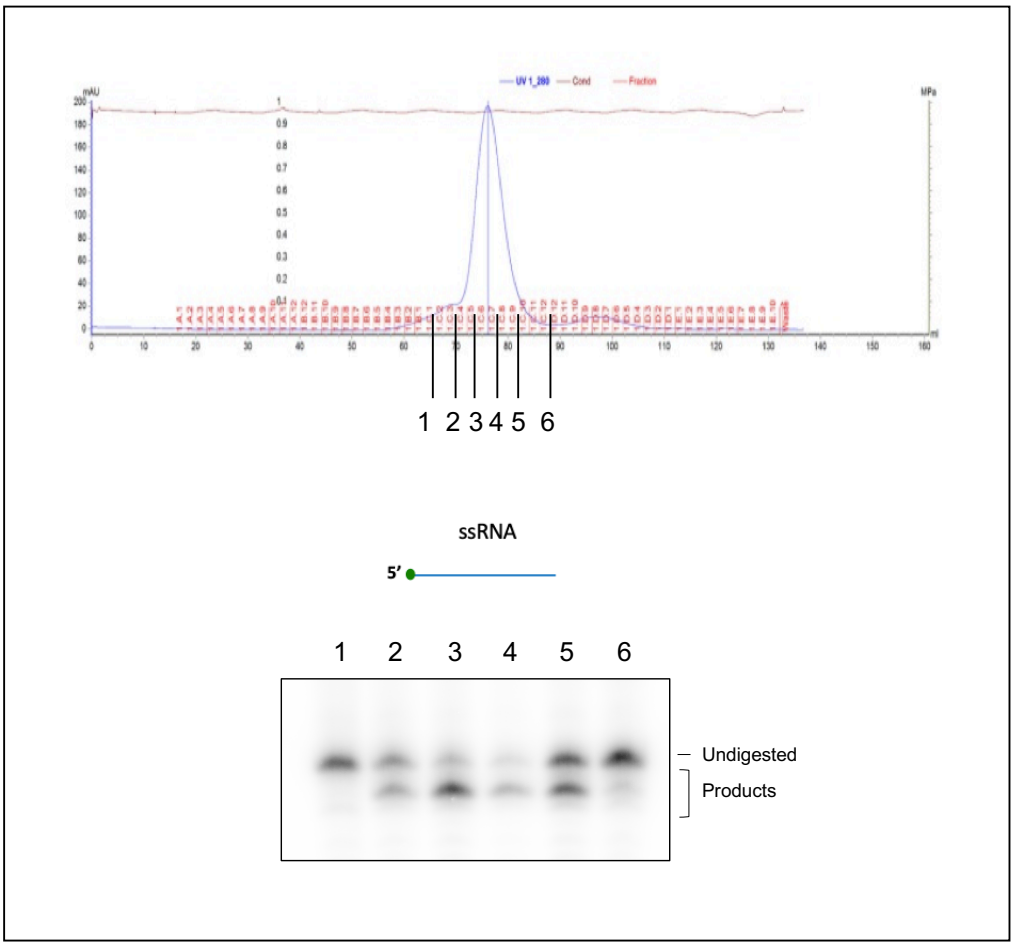


Figure 6.5: C15orf41-Codanin-1 activity coincides with the gel filtration elution peak
Fractions before (1), during (2-5) and after (6) the gel filtration peak were collected and tested for nuclease activity. C15orf41-Codanin-1 activity observed coincided precisely with the RNAse activity observed.
Fractions (as indicated) were incubated with substrate at 37oC for 60 min; reactions were subsequently analysed by 20% denaturing PAGE to visualize product formation.

optimal concentration of cation to perform all further biochemical experiments with.

6.2 C15orf41-Codanin-1 is a versatile RNA nuclease

In order to determine whether C15orf41-Codanin-1 activity is limited to RNA only, a variety of DNA substrates of various lengths and structures were tested on a broad range of protein concentrations (Figure 6.8 a&b). DNA substrates reminiscent of DNA replication intermediates were prioritised as notable members of the PD-(D/E)

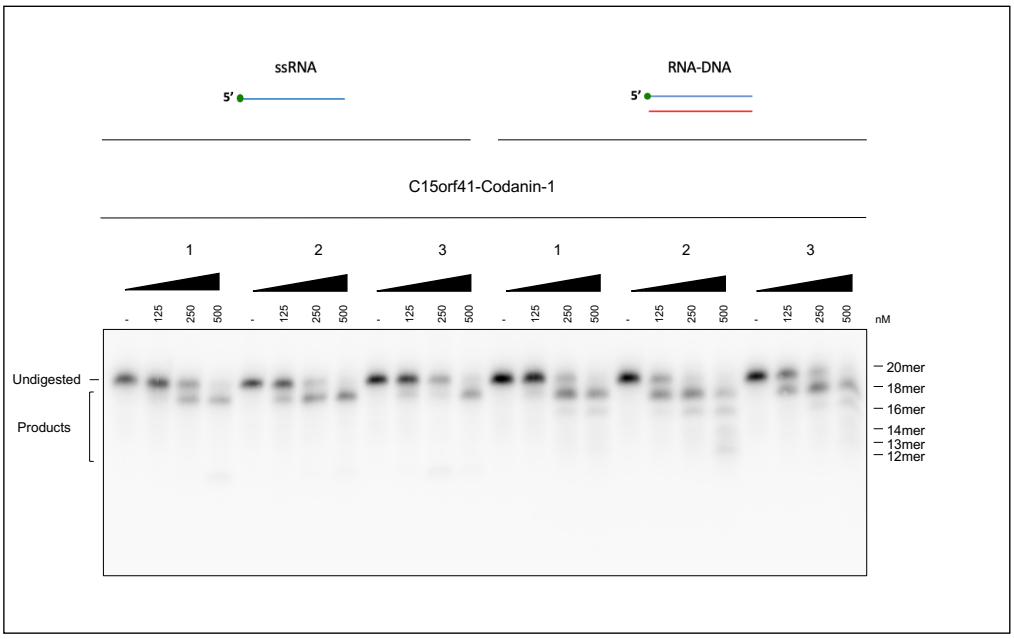


Figure 6.6: C15orf41-Codanin-1 nuclease activity is reproducible
The nuclease activity of three, independently purified C15orf41-Codanin-1 complexes (1, 2, 3) was assessed on a ssRNA and an RNA-DNA hybrid. Nuclease activity was reproducible amongst all three purifications.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

XK family of nucleases, such as DNA repair nucleases Mus81-EME1 and XPF-ERCC1, act on these substrates. These include structures mimicking replication forks, both with and without leading and lagging strands. In no case was DNAse activity observed, even at concentrations 10-fold higher than previously tested for RNA substrates. Importantly, a DNA-RNA hybrid was also tested where the DNA strand was radiolabelled. Again, no DNase activity could be observed, further confirming C15orf41 activity is specific to RNA strands, even when in complex with DNA (Figure 6.8 c). Interestingly, however, when electrophoretic mobility shift assays (EMSAs) were performed to assess the ability of the C15orf41-Codanin-1 complex to bind to RNA or DNA, the complex showed capability of binding to both RNA and DNA substrates, even though no DNAse activity has been found (Figure 6.9). EMSAs for wild-type and mutant C15orf41-Codanin-1 (as indicated) were executed in the presence of both magnesium and calcium ions. Magnesium was chosen as C15orf41 activity was shown to be ion-dependent (see Figure 6.7),

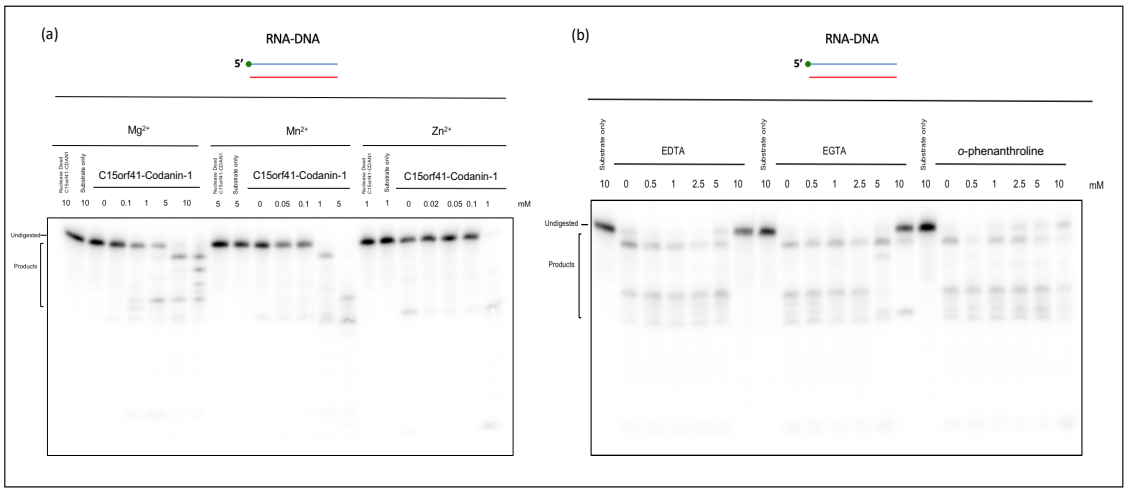


Figure 6.7: C15orf41-Codanin-1 has an ion-dependent RNA nuclease activity
(a) C15orf41-Codanin-1 activity is enhanced by the addition of Mg^{2+}, Mn^{2+} and Zn^{2+}. We determined 5 mM Mg^{2+} as optimal for activity.
(b) Consistent with its requirement for metal ions, C15orf41-Codanin-1 activity is inhibited by the addition of increasing concentrations (0 to 10 mM) of metal ion chelators EDTA, EGTA, and *o*-phenanthroline, with a particular sensitivity towards EDTA and EGTA.

whereas calcium was chosen for its inhibitory effect (an effect previously described for members of the PD-(D/E) XK superfamily), preventing substrate digestion and promoting nuclease-substrate complex formation under native conditions. Since nucleolytic digestion was only observed for the wild-type complex on RNA substrates and not DNA substrates but C15orf41 is able to bind to both, this suggest C15orf41 holds an intrinsic capability to differentiate between deoxyribose and ribose sugars, with the ability to bind to either but process only ribonucleic acids.

Subsequent studies focussed on the activity of C15orf41-Codanin-1 on structural and sequence variation of RNA substrates. Firstly, the activity of C15orf41-Codanin-1 was compared on RNA substrates of different lengths to investigate a potential length specificity. 10-mer, 20-mer and 30-mer ssRNA were tested and revealed the complex is only active on a 20-mer ssRNA substrate (Figure 6.10.a). The same pattern is observed when analysing C15orf41-Codanin-1 activity on dsRNA and RNA-DNA hybrids of the same lengths (Figure 6.10 b&c). Again, a strong preference for the 20-mer substrates is visible with no distinguishable digestion observed on any of the 10-mer substrates, possibly because the complex requires a minimal footprint to allow for stable binding. In contrast, weak activity (only

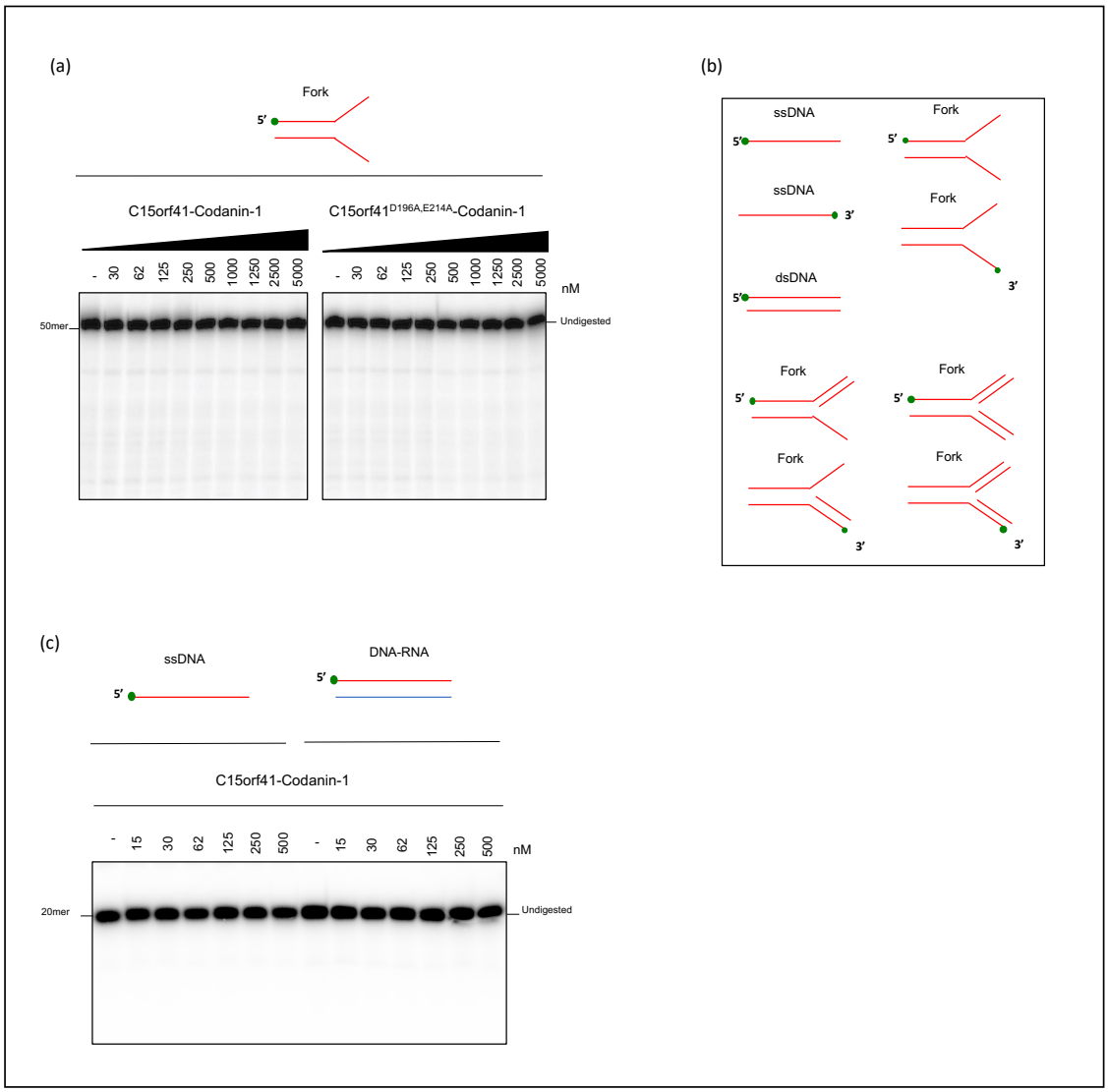


Figure 6.8: C15orf41-Codanin-1 shows no nuclease activity on various DNA substrates
(a) Nuclease activity of wild-type C15orf41-Codanin-1 as well as catalytically-inactive C15orf41^{D196A,E214A}-Codanin-1 complex were assessed on various ssDNA and complex DNA structures (as indicated in (b)). In all cases, no digestion is observed, confirming C15orf41-Codanin-1 is an RNase.
(b) C15orf41-Codanin-1 is also inactive on ssDNA and the DNA strand of an RNA-DNA hybrid, again confirming its specificity to RNA substrates.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

observed with protein concentrations above 250 nM) was observed on the 30-mer RNA-DNA hybrid only (Figure 6.10 b&c). This might be attributed to the increased conformation flexibility (or "breathing") of an RNA-DNA hybrid as opposed to duplex RNA, allowing for some interaction with C15orf41-Codanin-1, although

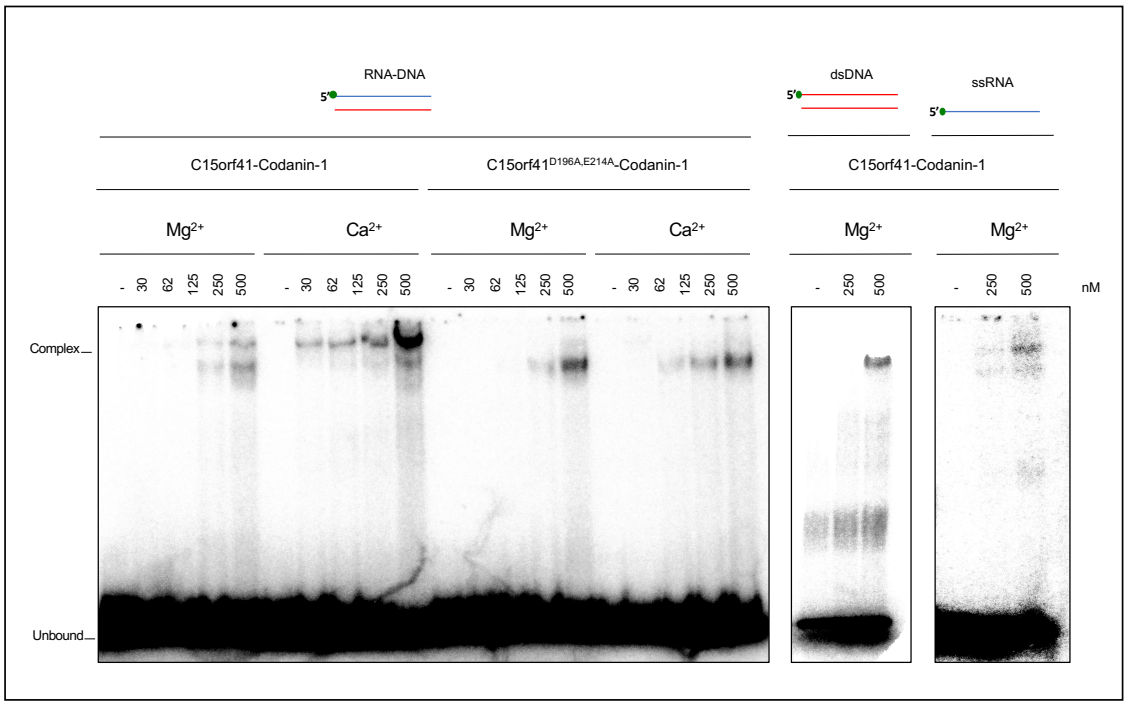


Figure 6.9: C15orf41-Codanin-1 is able to bind both RNA and DNA substrates
Electrophoretic mobility shift assays show both wild-type C15orf41-Codanin-1 and mutant C15orf41^{D196A,E214A}-Codanin-1 are able to bind to an RNA-DNA hybrid, both with the presence of Mg^{2+} and Ca^{2+}. C15orf41-Codanin-1 is also able to bind to dsDNA and ssRNA, even though the complex is inactive on DNA substrates, suggesting C15orf41 holds an intrinsic capacity to differentiate between ribose and deoxyribose nucleic acids. Increasing concentrations of protein (as indicated) were incubated with substrates for 15min on ice; reactions were subsequently analysed by 20% denaturing PAGE to visualize product formation. Oligonucleotide sequences used are indicated in Tables 3.1 and 3.2 in the Materials and Methods section.

further experiments will need to be conducted to fully understand this observation.

C15orf41-Codanin-1 show preferential nuclease activity on 20-mer double-stranded RNA-containing substrates, displaying a distinct digestion pattern: at lower protein concentrations (<125 nM), an initial 2-nucleotide cut is visible (forming at 18-mer product), reminiscent of the protein's activity on ssRNA. As the protein concentration is increased, the RNA is further cleaved in 2-nucleotide steps from the 3′-terminus until it reaches a 14-mer length which initiates single-nucleotide RNA laddering to a final 12-mer product (Figure 6.11). No products smaller than a 12-mer were observed, possibly relating to the minimum footprint necessary for the enzyme to bind. For 20-mer ssRNA, C15orf41-Codanin-1 only apparently incises

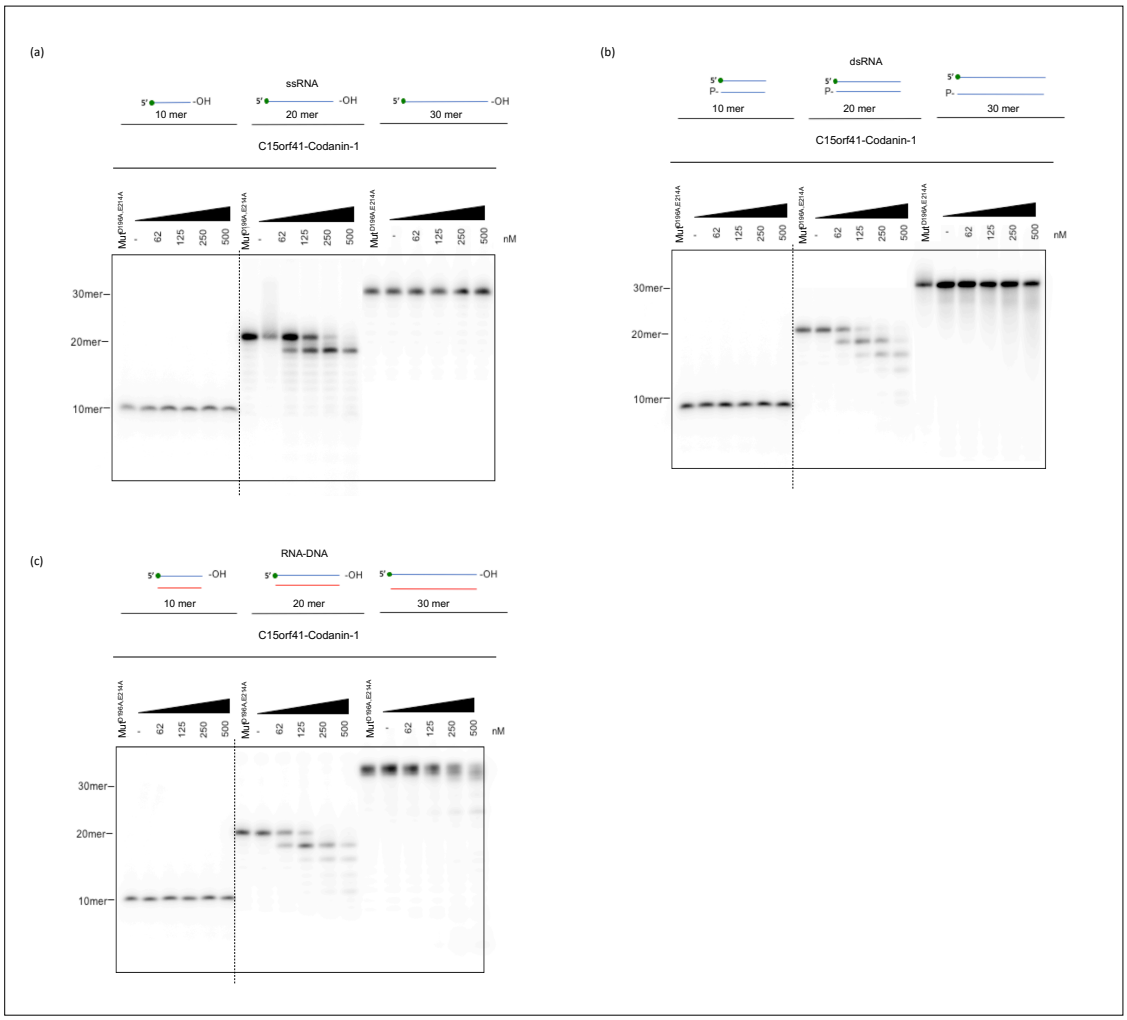


Figure 6.10: C15orf41-Codanin-1 shows preferential activity on 20-mer RNA substrates
The nuclease activity of C15orf41-Codanin-1 was tested on 10-mer, 20-mer and 30-mer ssRNA (a), dsRNA (b) and RNA-DNA hybrids (c). In all cases, the complex displays the most activity on 20-mer RNA substrates with no distinguishable digestion observed on any of the 10-mer substrates and only weak activity on the 30-mer RNA-DNA substrate. Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

on the substrate at one site, at 2-nucleotides away from the 3′-terminus, creating a 18-mer product. The size of the digestion products was precisely determined using a single-nucleotide RNA ladder. The ladder was generated through hydrolysis of the 20-mer substrate in 50 mM potassium acetate (pH 7.9), 20 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate (pH 7.9) and 1 mM DTT.

To better understand the digestion pattern of the C15orf41-Codanin-1 complex,

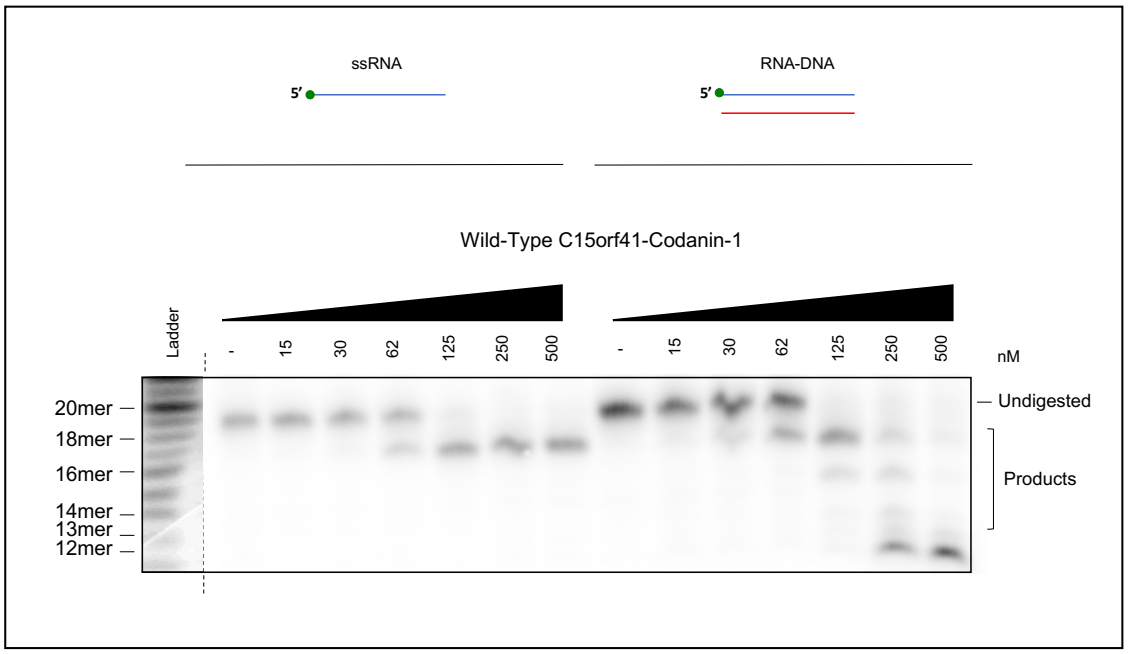


Figure 6.11: C15orf41-Codanin-1 displays a distinct digestion pattern
On ssRNA, C15orf41-Codanin-1 performs a single cut, 2 nucleotides away from the 3′-terminus, releasing a 18-mer product. In contrast, C15orf41-Codanin-1 displays a more complex digestion pattern on an RNA-DNA hybrid: at lower protein concentrations (<125 nM), an initial 2-nucleotide cut is visible (forming at 18-mer product). As the protein concentration is increased, the RNA is further cleaved in 2-nucleotide steps from the 3′-terminus until reaching a 14-mer length which initiates single-nucleotide RNA laddering to a final 12-mer product.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

its activity was compared on sequence-identical 20-mer ssRNA substrates containing either a 3′-hydroxyl, 3′-phosphate, or 3′-biotin moiety (Figure 6.12.a). Interestingly, activity was observed only on the substrates containing the 3′-hydroxyl and 3′-phosphate but not the 3′-biotin moiety. The characteristic 2-nucleotide cut was only present on the RNA substrates with a free 3′-end, with C15orf41-Codanin-1 showing the ability to process both the 3′-hydroxyl and the 3′-phosphate substrates equally. In contrast, no activity was observed for the 3′-biotin substrate, suggesting C15orf41-Codanin-1 is a novel RNA exonuclease whose activity is dependent upon the presence of an unblocked 3′- terminus. To further confirm this, the same substrates as detailed above where incubated with a known RNA exonuclease, Exonuclease T. As expected, Exonuclease T was able to efficiently digest both the 3′-hydroxyl and the

3′-phosphate substrates but its activity was clearly blocked by the 3′-biotin group, similar to the results obtained for C15orf41-Codanin-1 (Figure 6.12.b).

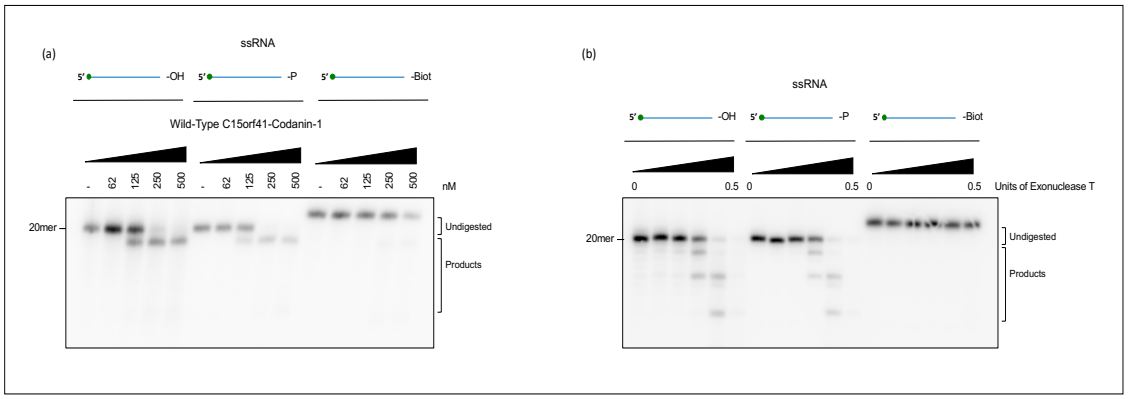


Figure 6.12: Figure 12: C15orf41-Codanin-1 is an RNA exonuclease
(a) C15orf41-Codanin-1 is an RNA exonuclease. With a substrate containing a 3′-biotin group, no nuclease activity can be observed while the substrates with a 3′- hydroxyl or phosphate exhibit the single incision pattern expected on ssRNA.
(b) Exonuclease T was used to demonstrate a classic RNA exonuclease activity. With a substrate containing a 3′-biotin group, no activity is observed confirming ExoT has no endonuclease activity while the substrates with a 3′- hydroxyl or phosphate exhibit nearly identical product formation.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

6.3 C15orf41-Codanin-1 is able to process a variety of complex RNA-containing structures

This work has shown C15orf41-Codanin-1 is an exonuclease with a nucleolytic activity specific to RNA and double-stranded RNA-containing substrates. To shed light on a possible biological function linking the biochemical activity of C15orf41-Codanin-1 and its cellular role, complex RNA-containing substrates mimicking common RNA-DNA structures found in fundamental cellular processes were tested next. To this end, RNA structures representative of a broad variety of cellular secondary structures were tested: structures mimicking embedded ribonucleotides, ‘Okazaki-like’ structures, R-loops and substrates containing common RNA modifications.

Initially, the activity of C15orf41-Codanin-1 was tested on a ssDNA and dsDNA substrate containing an embedded ribonucleotide (analogous to the types of substrate

processed by the ribonucleotide excision repair (RER) enzymes). Interestingly, no activity was observed on this substrate, suggesting C15orf41-Codanin-1 is not able to incise a single ribonucleotide within a DNA substrate. It is likely C15orf41-Codanin-1 activity is dependent on a free RNA 3′- terminus (as opposed to a DNA 3′- terminus), to allow the enzyme to load onto the substrate (Figure 6.13).

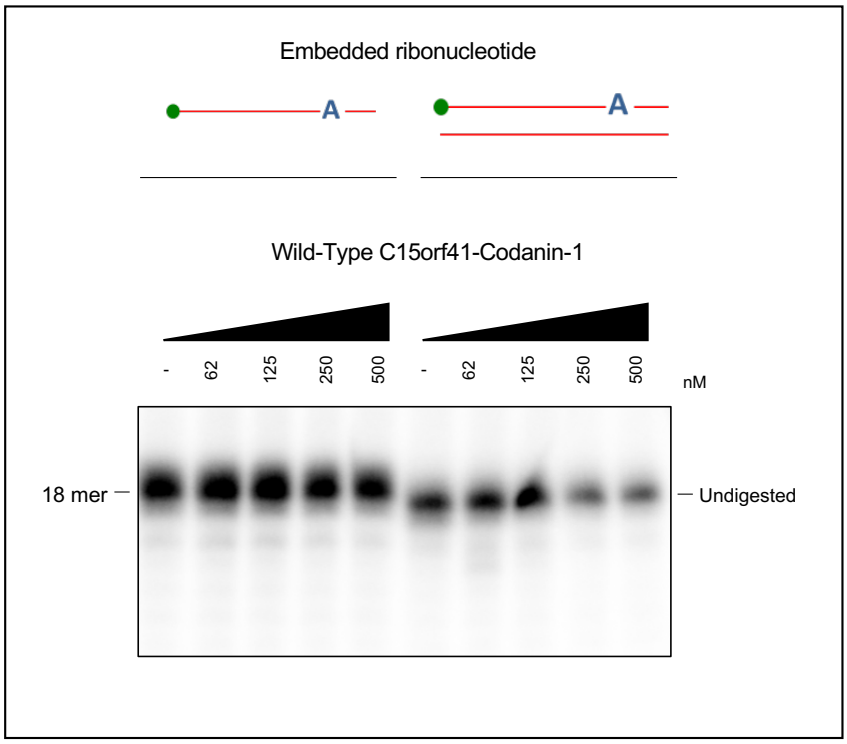


Figure 6.13: C15orf41-Codanin-1 is not able to process substrates with embedded ribonucleotides
C15orf41-Codanin-1 is not able to incise a single ribonucleotide within a DNA substrate (both for ss- and ds-DNA), likely due to its requirement for a free RNA 3′- terminus (as opposed to a DNA 3′- terminus) to load onto the substrate.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

Next, 'Okazaki-like' DNA replication intermediates were examined (Zheng and Shen, 2011). Substrates with a 5′- RNA section linked to a 3′- DNA section mimicking an Okazaki fragment were designed, both single-stranded and annealed to a DNA lagging-like structure, and were incubated with C15orf41-Codanin-1. Similar to the embedded nucleotide structure, no digestion was observed; confirming that, even at the highest concentrations tested, C15orf41-Codanin-1 is not able to process substrates containing a 3′- DNA end (Figure 6.14).

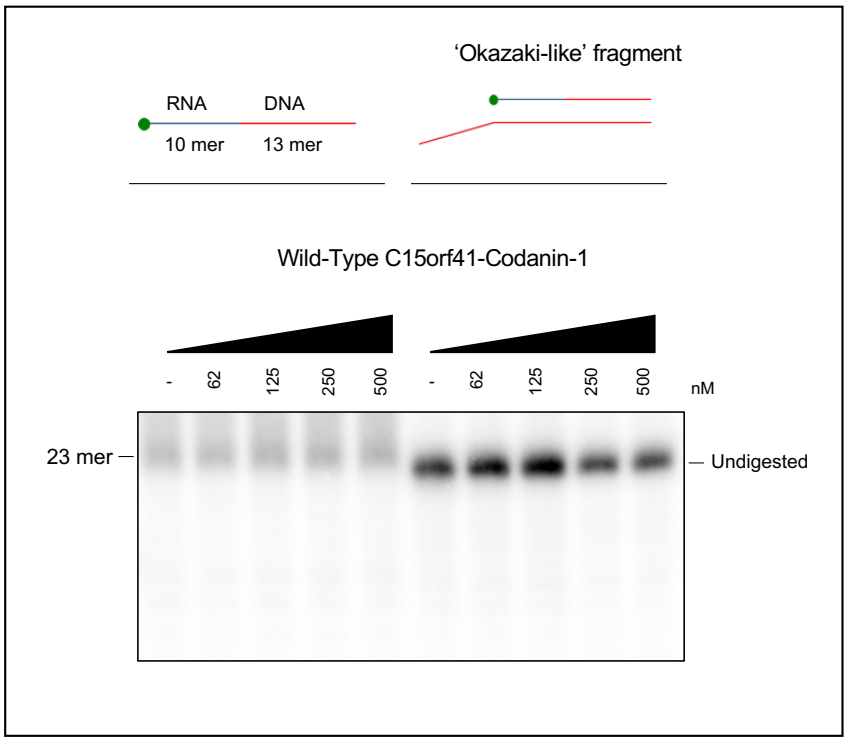


Figure 6.14: C15orf41-Codanin-1 is not able to process substrates mimicking 'Okazaki-like' structures
C15orf41-Codanin-1 was not able to process structures containing an RNA primer (10 mer) fused to a 5′ DNA fragment (13 mer), again highlighting its requirement for a free RNA 3′- terminus (as opposed to a DNA 3′- terminus) to load onto the substrate. Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

R-loops are common RNA-containing structural intermediates, primarily formed during transcription. An R-loop is a three-stranded structure formed of an RNA-DNA hybrid that results in the displacement of the second DNA and thereby forming a loop. Therefore, C15orf41-Codanin-1 activity was assessed on substrates mimicking R-loops, consisting of a DNA loop and a small 10-mer radiolabelled RNA fragment annealed within the loop region of the DNA. Results reveal C15orf41-Codanin-1 is able to make a single incision within the RNA fragment. Curiously, C15orf41-Codanin-1 is not able to nucleolytically incise the ssRNA fragment alone, in agreement with earlier results showing the enzyme requires a minimum length of substrate to allow for digestion. Furthermore, the ability of C15orf41-Codanin-1 to incise the RNA fragment is not dependent on the actual DNA loop structure, but rather on the longer bottom DNA strand that likely provides a scaffold for the

enzyme to bind and overcome its minimum length requirement (Figure 6.15.a). As a control, a D-loop structure was also tested where the short RNA fragment was replaced with a short ssDNA fragment. As expected, no activity is observed on any of the DNA components of this structure, further confirming C15orf41-Codanin-1 activity is specific to RNA only (Figure 6.15.b).

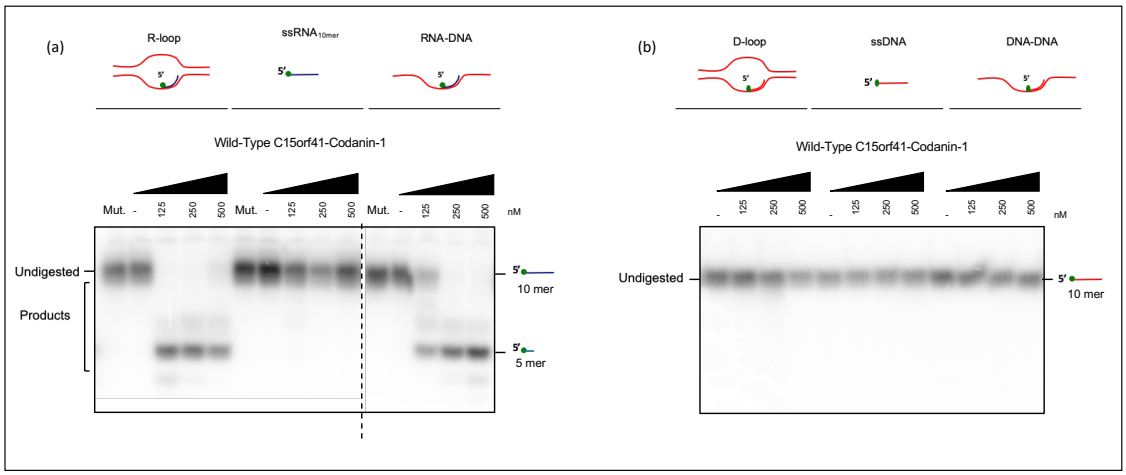


Figure 6.15: C15orf41-Codanin-1 is able to process R-loop but not D-loop structures
(a) C15orf41-Codanin-1 is able to incise a 10-mer RNA fragment annealed to a DNA-loop structure mimicking an R-loop, creating a 5-mer product. The same activity is observed when the RNA fragment is annealed to a long DNA strand but not when the RNA fragment is present alone.
(b) C15orf41-Codanin-1 shows no activity on any DNA structure, including D-loop confirming the nuclease activity of the complex is RNA specific.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.
Mut.: Mutant C15orf41^{D196A,E214A}-Codanin-1 catalytically inactive mutant was used as a negative control.

Finally, RNA substrates containing base modifications were tested. The impact of following modifications on C15orf41-Codanin-1 were investigated: 6-methyladenine, the artificial base modification 2-methyladenine and the variant RNA base inosine. 6-methyladenine is the most common RNA modification, mainly present in messenger RNA, where it serves as a regulatory marker for translation and stability (Frye et al., 2018). It is also of particular interest to understand the activity of C15orf41-Codanin-1 on this modification as the 6-methyladenine marker has been implicated in regulation of R-loop stability, a potential role for C15orf41-Codanin-1 that

will be discussed further in Chapter 8 (Abakir et al., 2020; Marnef and Legube, 2020). Inosine, generated by A-to-I editing has been reported to be involved in innate immune sensing (Nakahama and Kawahara, 2020). C15orf41-Codanin-1 activity was differentially affected by the presence of certain RNA modifications two ribonucleotides from the 3′-terminus of the 20-mer substrates when compared with an unmodified RNA substrate (Figure 6.16). Whilst the 6-methyladenine and the unmodified substrate were processed almost identically, the presence of the 2-methyladenine was slightly inhibitory, with digestion observed only at the highest protein concentration (500 nM). In contrast, the addition of the inosine base to the substrate was strongly inhibitory to C15orf41-Codanin-1, both quantitatively and qualitatively. Partial digestion can only be observed at the highest protein concentration (500 nM), with only a single incision present and no characteristic RNA laddering.

Overall, these results confirm C15orf41-Codanin-1 is a versatile RNA nuclease that is able to digest a variety of RNA secondary structures and common modifications. Its activity is specific to RNA only and digestion only possible if the substrate contained a free, 3′- RNA terminus. These newly-described biochemical activities provide some useful insight into a possible role for C15orf41-Codanin-1 in CDA-I biology, which will be elaborated on in the Discussion section of this thesis.

6.4 CDA-I patient mutations in C15orf41 reduce C15orf41 nuclease activity

The sections above have provided a basic biochemical characterisation of the C15orf41-Codanin-1 complex, involved in CDA-I, showing C15orf41-Codanin-1 is a novel RNA exonuclease with a broad activity spectrum. In order to further investigate how the newly-identified biochemical activity of C15orf41 relates to the cellular abnormalities in CDA-I, the nuclease activity of patient mutant C15orf41 proteins was tested.

The following patient mutations were introduced to C15orf41: the founder p.L178Q mutation, mapped within the predicted nuclease domain of the protein;

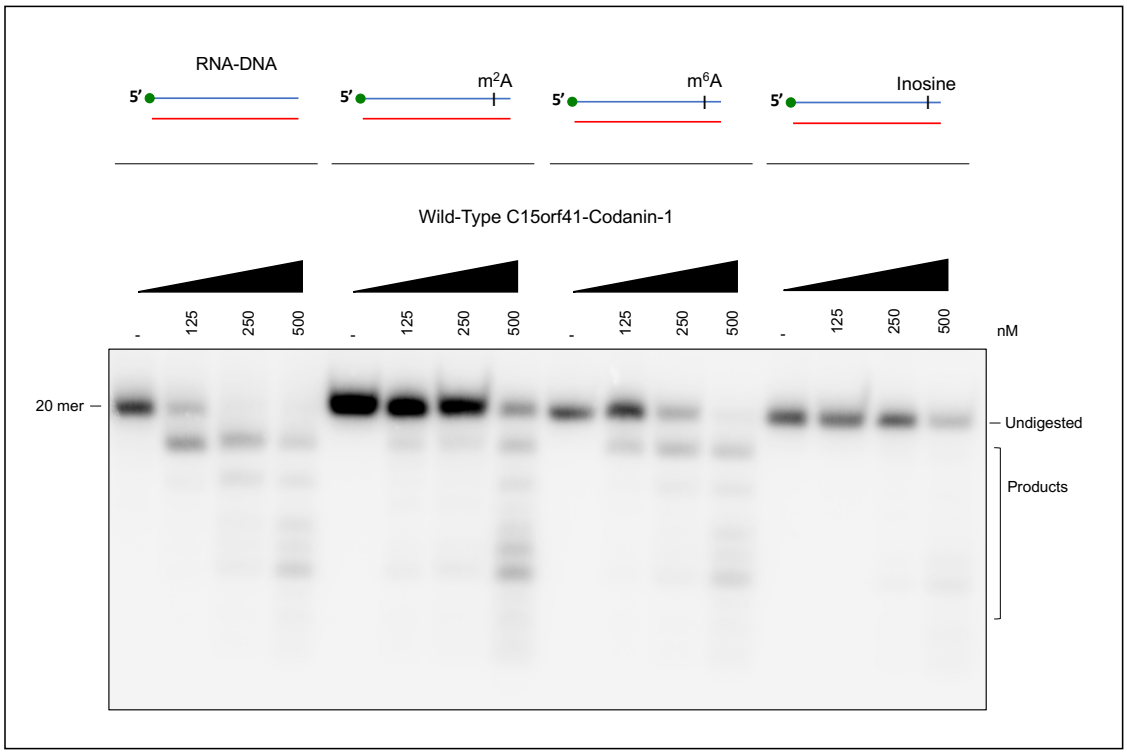


Figure 6.16: C15orf41-Codanin-1 is able to process some common chemical modification of RNA
C15orf41-Codanin-1 is able to process common chemical modifications of RNA if present on the RNA strand of an RNA-DNA hybrid, albeit differentially. Whilst 2-methyladenine (m^2A) is mildly inhibitory (with digestion observed only at 500 nM protein), inosine (I) was strongly inhibitory with almost no activity observed. In contrast, C15orf41-Codanin-1 was able to process 6-methyladenine (m^6A) almost as efficiently as a simple RNA-DNA hybrid.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

and second founder p.Y94C mutation, located in the predicted substrate-binding domain, referred to as C15orf41^{L178Q} and C15orf41^{Y94C} respectively. Considering that these patient mutations in C15orf41 abolished C15orf41-Codanin-1 complex formation *in vivo*, wild-type and patient mutant C15orf41 alone were purified to near-homogeneity and tested on an RNA-DNA hybrid substrate to compare for nuclease activity (Figure 6.17). The catalytically-inactive complex (C15orf41^{D196A,E214A}-Codanin-1) was used as a negative control and the wild-type complex as a positive control, to show a complete digestion profile. From the results above, we know wild-type C15orf41 alone (C15orf41^{WT}) has reduced digestion products compared to the complex. Therefore, as expected, C15orf4^{1WT} protein alone shows activity on this

substrate, albeit less than the complex, with a single incision 2 nucleotides away from the 3′-terminus. In comparison, both patient mutations in C15orf41 severely affect its RNA nuclease activity. While the Y94C mutation almost completely abolished the nuclease activity of C15orf41^{Y94C}, likely by distorting the predicted helix-turn-helix motif of the protein thereby disrupting substrate binding, C15orf41^{L178Q} only shows very weak digestion at the highest concentration of protein (500 nM). It is likely that altering a highly conserved hydrophobic leucine to a polar glutamine affects the conformational of the active site and therefore the catalytic mechanism of C15orf41, causing the reduced activity observed.

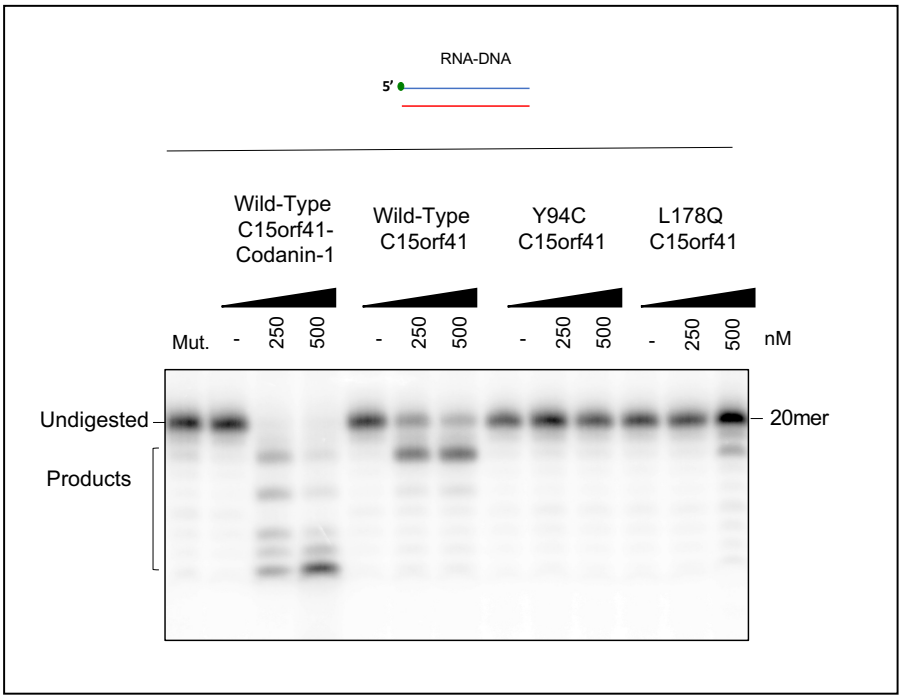


Figure 6.17: Patient mutations Y94C and L178Q in C15orf41 disrupt its nuclease activity
The patient mutations Y94C and L178Q, located in the predicted substrate-binding domain and nuclease domain respectively in C15orf41, both severely disrupt its nuclease activity. Compared to wild-type C15orf41, which is able to incise the substrate at all concentrations tested, almost no digestion is observed for the patient mutant C15orf41s proteins.
Mut.: Mutant C15orf41^{D196A,E214A}-Codanin-1 catalytically inactive mutant was used as a negative control.
Wild-type C15orf41-Codanin-1 was used as a positive control.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

Although patient mutations in C15orf41 impair complex formation with Codanin-1 *in vivo*, complex formation between patient mutant C15orf41^{Y94C} and C15orf41^{L178Q} and Codanin-1 was achieved *in vitro* (see purification data in Chapter 5). Patient mutant complexes C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 were therefore purified to near homogeneity and assessed for nuclease activity. Interestingly, the addition of the C-terminal fragment of Codanin-1 modulated the nuclease activity for both C15orf41 mutants, reminiscent of the effect of this fragment of Codanin-1 on C15orf41^{WT} activity (Figure 6.18). Both patient mutant complexes show increased digestion, albeit still at lower levels than wild-type C15orf41-Codanin-1 complex. In particular, the Y94C patient mutation appears to affect the nuclease activity of the mutant complex to a greater extent than the L178Q mutation, suggesting the Y94C residue could be involved in both substrate mediation as well as Codanin-1 binding. It should also be noted that in all cases, the patient mutations within C15orf41 did not fully abolish its nuclease activity. This is in agreement with reports suggesting C15orf41 activity is essential and that CDA-I may result from impaired biochemical activity. Altogether, these results suggest impaired complex formation between C15orf41 and Codanin-1 is likely to be pathogenic in CDA-I, with correct complex formation necessary to assure full C15orf41 nuclease activity. Disruption of complex formation by patient mutations, as shown by *in vivo* experiments, could therefore lead to decreased nuclease activity of C15orf41, likely causative of the phenotypes observed in CDA-I. Still, the cellular pathway affected by reduced nuclease activity of C15orf41 that leads to CDA-I remains to be discovered. In addition, identifying the native cellular substrates of the C15orf41-Codanin-1 complex would likely also shed light on this.

6.5 Conclusion

Taken together, the data presented in this chapter provides the first activity report for C15orf41, a previously undescribed protein involved in CDA-I. C15orf41 is a novel RNA nuclease, with no discernible activity on DNA. Without its binding partner Codanin-1, C15orf41 is only able to digest double-stranded RNA and

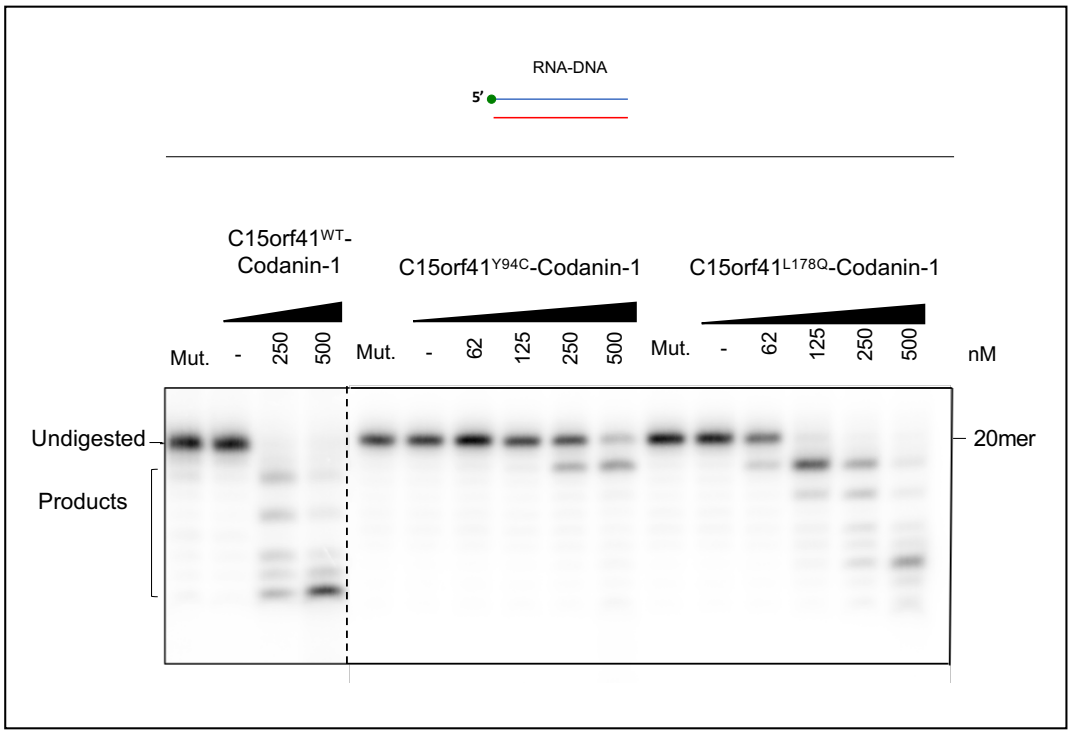


Figure 6.18: Patient mutant complexes C15orf41^{Y94C}-Codanin-1 and C15orf41^{L178Q}-Codanin-1 show partially restored activity in the presence of Codanin-1
The addition of the C-terminal fragment of Codanin-1 to patient mutant C15orf41^{Y94C} and C15orf41^{L178Q} restores some nuclease activity lost due to the presence of the patient mutations in C15orf41, albeit not fully to the level of C15orf41^{WT}-Codanin-1 complex. This suggests C15orf41-Codanin-1 complex formation is pathogenic in CDA-I.
Mut.: Mutant C15orf41^{D196A,E214A}-Codanin-1 catalytically inactive mutant was used as a negative control.
Wild-type C15orf41^{WT}-Codanin-1 was used as a positive control.
Protein concentrations as indicated; reactions analysed by 20% denaturing PAGE; see Materials and Methods section for more details.

the RNA strand of an RNA-DNA hybrid, incising once 2-nucleotides from the 3′-terminus of the 20-mer substrates. We find the activity of C15orf41 is greatly enhanced and modulated by the addition of a C-terminal fragment of Codanin-1, the second disease protein in CDA-I. The C15orf41-Codanin-1 complex is a versatile RNA nuclease complex, able to digest a variety of RNA substrates including single-stranded RNA, double-stranded RNA as well as more complex replication and transcription intermediates that contain an unblocked 3′-RNA terminus. Both C15orf41 alone and the C15orf41-Codanin-1 complex show no discernible activity on any DNA substrates tested.

The effect of the founder p.L178Q mutation, mapped within the predicted nuclease domain of the protein; and second founder p.Y94C mutation, located in the predicted substrate-binding domain was also assessed on the RNAse activity of C15orf41. In both cases, the presence of patient mutations Y94C and L178Q in C15orf41 disrupts its nuclease activity, likely interfering with substrate binding or processing. This finding suggesting an important role between disease pathogenesis and the nuclease activity of C15orf41. Further perspectives on links between the biochemical activity of this complex and its cellular role will be discussed in Chapter 8.

7

Biochemical characterisation of nsp14-nsp10

In response to the 2020 SARS-CoV-2 global pandemic, the focus of this thesis was shifted towards characterisation of the SARS-CoV-2 nsp14-nsp10 complex, an RNA nuclease complex essential for viral replication. The SARS-CoV-1 nsp14-nsp10 has been shown to possess a $3'$-$5'$ exonuclease activity that is selective for dsRNA and able to excise terminally mismatched ribonucleotide structures (M. Bouvet et al., 2012; Ma et al., 2015). It has therefore been implied the ribonuclease acts as a proof-reading nuclease.

nsp14-nsp10 is highly conserved between the SARS-CoV-1 and SARS-CoV-2, implying SARS-CoV-2 nsp14-nsp10 holds a similar activity and function. The aim of this chapter is therefore to complete a biochemical characterisation of the SARS-CoV-2 nsp14-nsp10 complex, detailing its activity profiles and substrate specificities. Very little research has been done on this complex thus far, rendering these observations essential for better understanding of the virus' functioning but also suggest new, alternative therapeutic avenues.

7.1 SARS-CoV-2 nsp14-nsp10 is an RNA processing nuclease

To biochemically characterise the SARS-CoV-2 nsp14-nsp10 nuclease complex, wild-type full-length nsp14 and wild-type full-length nsp10, referred to as nsp14-nsp10 from here on, was purified by Dr. Y. Yosaatmadja as part of our collaboration with the Structural Genomics Consortium and Centre for Medicines Development,

University of Oxford. In brief, codon optimised constructs were expressed in *E. coli* and purified on an immobilised nickel affinity chromatography column (Ni-IMAC) exploiting the nickel-affinity of the N-terminal His-tag on Nsp14. The His-tag was cleaved using TEV-protease and the protein complex passed through the Ni-IMAC column for a second time to allow for further purification of the un-cleaved nsp14-nsp10 complex. The protein was then subject to gel filtration using a Superdex 200 increase 10/300 GL column to obtain pure protein. Using the method outlined above, a predicted enzyme-inactive ('nuclease-dead') control complex was also purified, bearing substitution mutations D113A and E115A active-site residues. This complex is referred to as nsp14$^{\mathrm{D113A,E115A}}$-nsp10 and by comparison with a form of the SARS-CoV-1 nsp14 protein mutated at the cognate residues, mutation of these residues in SARS-CoV-2 is predicted to be catalytically inactive (Ma et al., 2015). Finally, wild-type full-length nsp14 (ExoN) subunit of the complex was also purified alone to allow for comparative nuclease activity studies between the individual subunit and the nsp14-nsp10 complex.

To assess the purity of the proteins obtained, an SDS-PAGE gel was performed with the nsp14-nsp10 proteins and protein complexes (Figure 7.1). As pictured, all proteins were purified to near homogeneity, with few faint protein bands present that were determined to be nsp14 degradation products by tryptic digest MS/MS (data not shown).

The next step was to determine whether the protein and protein complexes that were generated showed any RNA nuclease activity, analogous to SARS-CoV-1 complex activity (Ma et al., 2015). To achieve this, Nsp14 alone as well as the wild-type nsp14-nsp10 complex and the predicted nuclease-dead complex were incubated with a simple 20-mer ssRNA substrate terminally radiolabelled at the 5′-end to perform a nuclease assay. Within this assay, increasing concentrations of protein (as indicated in Figure 7.2) were incubated with 10 nM substrate at 37°C for 45 min in the presence of 25 mM HEPES-KOH, pH 7.5, 50 mM NaCl, 5 mM $MgCl_2$, 5% glycerol and 1.0 mM DTT and reactions were subsequently analysed by 20% denaturing PAGE to visualize product formation. As predicted, nucleolytic

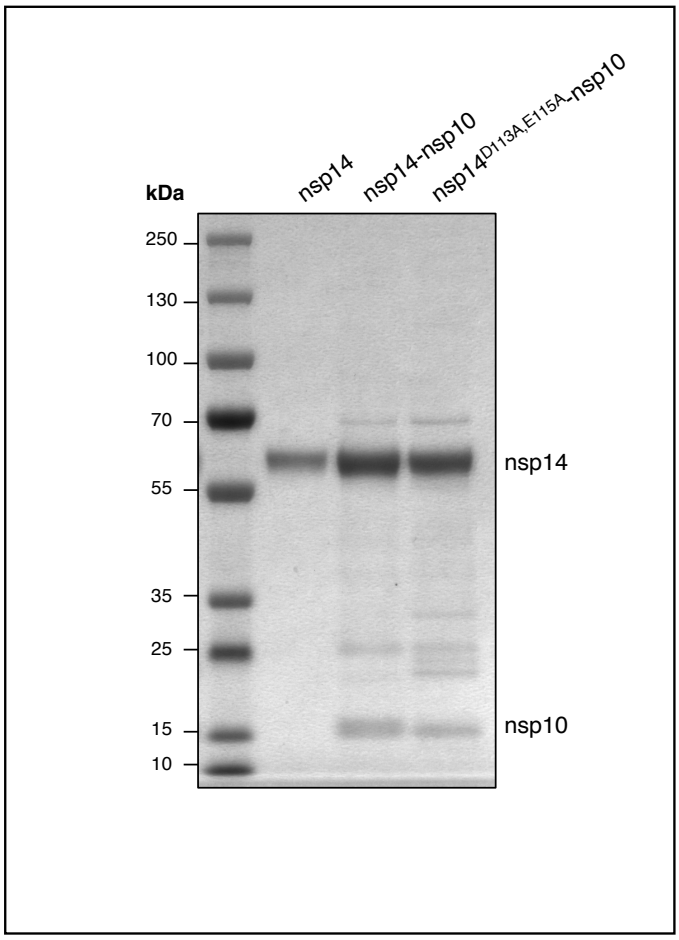


Figure 7.1: Purity of Sars-CoV-2 nsp14-nsp10 proteins used in this study
SDS-PAGE of purified nsp14 alone as well as WT nsp14-nsp10 (designated nsp14-nsp10) and a control 'nuclease-dead' complex bearing alanine substitutions at residues D113 and E115 (nsp14^{D113A,E115A}-nsp10) showing the purity of the purified proteins. Predicted molecular weights are 59 463 Da for nsp14 and 15 281 Da for nsp10.

digestion products were observed for nsp14 alone, albeit only at concentrations higher than 250 nM. Interestingly, the addition of its binding partner nsp10 greatly enhanced the nuclease activity of nsp14, with nsp14-nsp10 digestion visible even at the lowest concentration employed (15 nM) (Figure 7.2). This is in accordance with reports of SARS-CoV-1 nsp10 association being essential for enhanced nsp14 activity (M. Bouvet et al., 2012).

In order to confirm that the observed nuclease activity is intrinsic to the nsp14-nsp10 complex, two complementary methods were employed. Firstly, the activity of the predicted nuclease-dead complex, namely nsp14^{D113A,E115A}-nsp10, was assessed on the 20-mer ssRNA substrate. As hypothesised, no activity was observed, even at

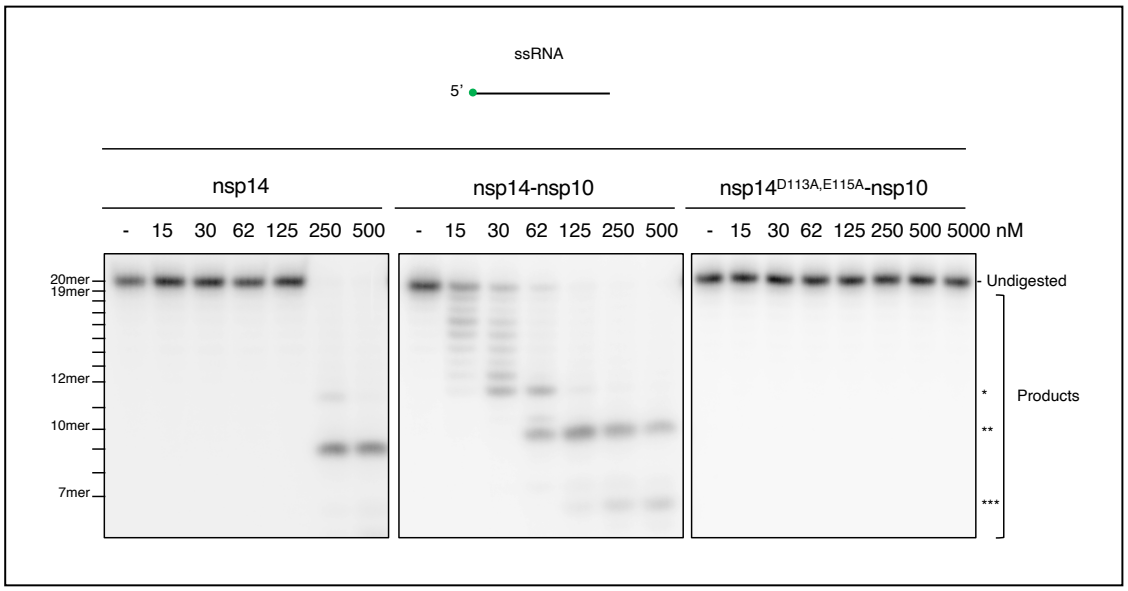


Figure 7.2: SARS-CoV-2 nsp10 confers additional activity to nsp14
Nsp14-nsp10 is an RNase able to digest a 20-mer ssRNA oligo (oligo 2 as indicated in materials and methods) in a single-nucleotide fashion from the 3′-end terminating at the 8^{th} ribonucleotide from the 3′-end (labelled *) and further incising closer to the 5′-end to generate products of 10-mer and 7-mer (labelled ** and *** respectively). Nsp14 alone is able to generate the 12-mer and 10-mer products at significantly higher protein concentrations. The predicted nuclease-dead nsp14^{D113A,E115A}-nsp10 mutant complex exhibits no discernible nuclease activity, even at ten-fold higher concentrations compared with the wild-type complex.
Size of products was determined as shown in Figure 7.6. Main products are labelled *, ** and *** corresponding to 12-mer, 10-mer and 7-mer respectively.

concentrations 10-fold higher than the highest concentration employed for wild-type nsp14-nsp10 (Figure 7.2). Second, the elution fractions of wild-type nsp14-nsp10 obtained from the gel filtration were assayed for nuclease activity. As shown in Figure 7.3, the characteristic RNAse activity of this complex coincides precisely with the peak of the elution profile, further indicating the observed activity results from nsp14-nsp10 activity and not due to a contaminant.

Next, the metal ion dependency of nsp14-nsp10 was investigated. To this end, 125 nM of nsp14-nsp10 complex was incubated with increasing concentrations (as indicated in Figure 7.4) of the divalent metal cations $MgCl_2$, $MnCl_2$ and $ZnCl_2$, on ice for 15 min before adding 10 nM ssRNA substrate and incubating the reaction at 37°C for 45 min. Reactions were subsequently analysed by 20% denaturing PAGE to visualise product formation. Both magnesium and manganese promote the RNase

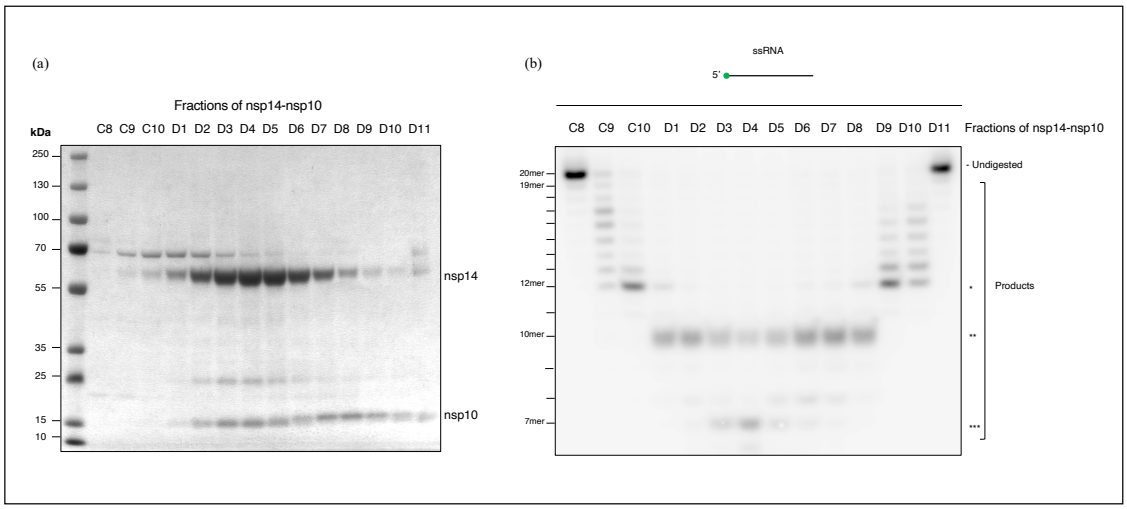


Figure 7.3: Nsp14-nsp10 activity coincides with the gel filtration elution peak
(a) SDS-PAGE of the fractions before, during and after the peak used from gel filtration for wild-type nsp14-nsp10. The predicted molecular weight is 59 463 Da for nsp14 and 15 281 Da for nsp10.
(b) The elution peak (D1–D8) coincides precisely with the characteristic RNAse activity that we observe in our activity gel. Oligo 2 was used in this gel.
Size of products was determined as shown in Figure 6. Main products are labelled *, ** and *** corresponding to 12-mer, 10-mer and 7-mer respectively.

activity of nsp14-nsp10 whereas zinc was inhibitory (Figure 7.4). These results are in line with previous biochemical studies that imply the nuclease activity of the ExoN family of nucleases is dependent upon divalent metal cations (M. Bouvet et al., 2012; Ma et al., 2015; Minskaia et al., 2006). This requirement is further highlighted as EDTA, EGTA and *o*-phenanthroline, three metal-chelating agents, were strongly inhibitory (Figure 7.4). For physiological relevance, 5 mM magnesium was chosen as the optimal concentration of cation to perform all further biochemical experiments.

To gain a more detailed understanding of the activity of nsp14-nsp10 on RNA, the complex was incubated with ssRNA of varying lengths: 10-mer, 20-mer and 30-mer to investigate a potential length specificity. A differential response was observed where nsp14-nsp10 was not able to digest a 10-mer substrate, possibly because it requires a minimal footprint to allow for stable binding, but digested a 20-mer and a 30-mer substrate (Figure 7.5). Interestingly, the 20-mer substrate shows preferential activity, whereby digestion is observed even at the lowest concentration of protein (15 nM), displaying a distinctive pattern: initially, RNA laddering is observed close

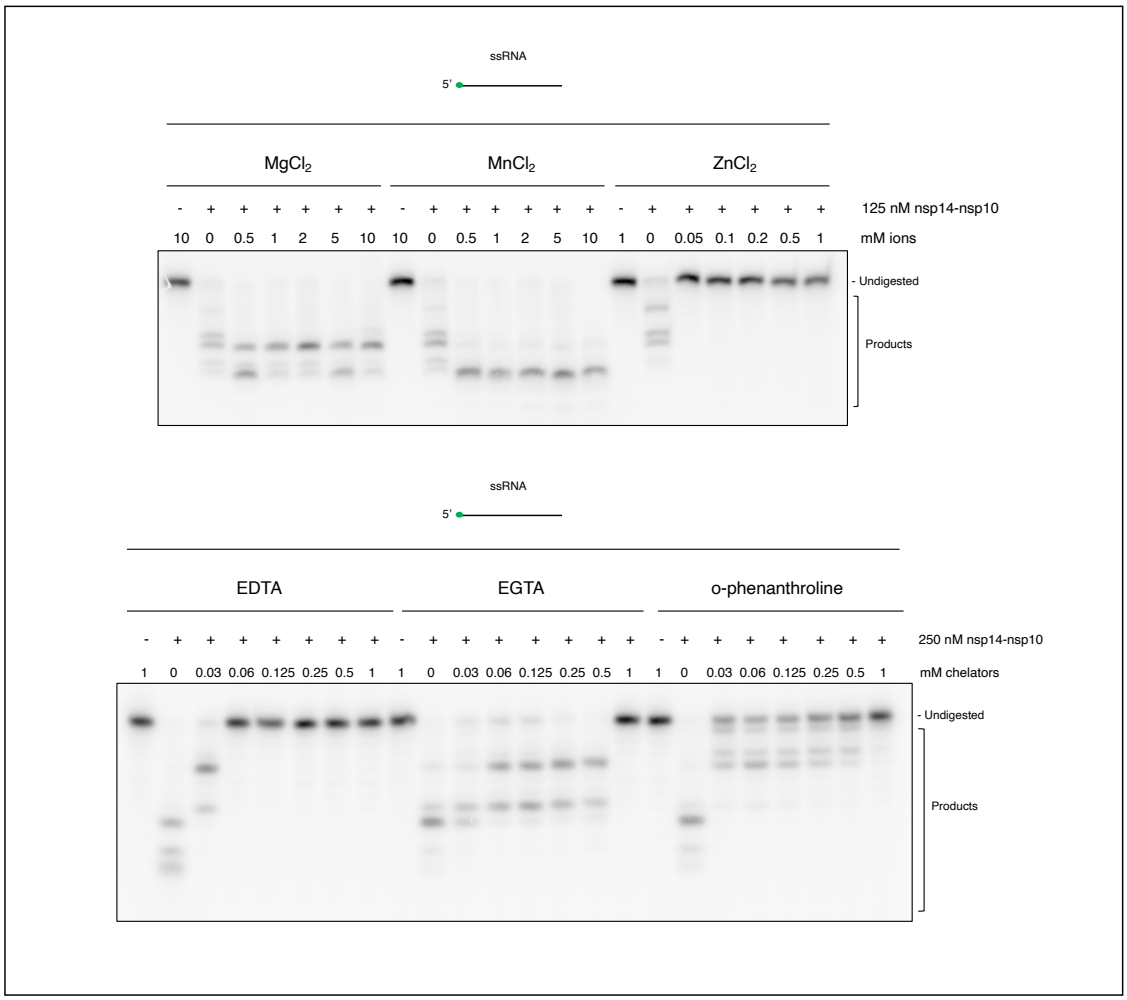


Figure 7.4: Nsp14-nsp10 is an ion-dependent RNA nuclease
Nsp14-nsp10 activity is enhanced by the addition of both $MgCl_2$ and $MnCl_2$ but is inhibited by $ZnCl_2$. We determined 5 mM $MgCl_2$ as optimal for activity. Consistent with its requirement for metal ions, nsp14-nsp10 activity is inhibited by the addition of metal ion chelators EDTA, EGTA, and *o*-phenanthroline, with a particular sensitivity towards EDTA. Oligo 2 was used in these gels.

to the 3′-terminus followed by the appearance of several additional prominent bands further from the 3′-end as the concentration of nsp14-nsp10 increases. The size of the digestion products was precisely determined using a single-nucleotide RNA ladder. The ladder was generated through hydrolysis of the 20-mer substrate in 50 mM potassium acetate (pH 7.9), 20 mM Tris-acetate (pH 7.9), 10 mM magnesium acetate (pH 7.9) and 1 mM DTT. This enabled the determination of following digestion pattern for nsp14-nsp10: nsp14-nsp10 digests 20-mer ssRNA in a single-nucleotide fashion from the 3′-end until the 8^{th} ribonucleotide (labelled with *)

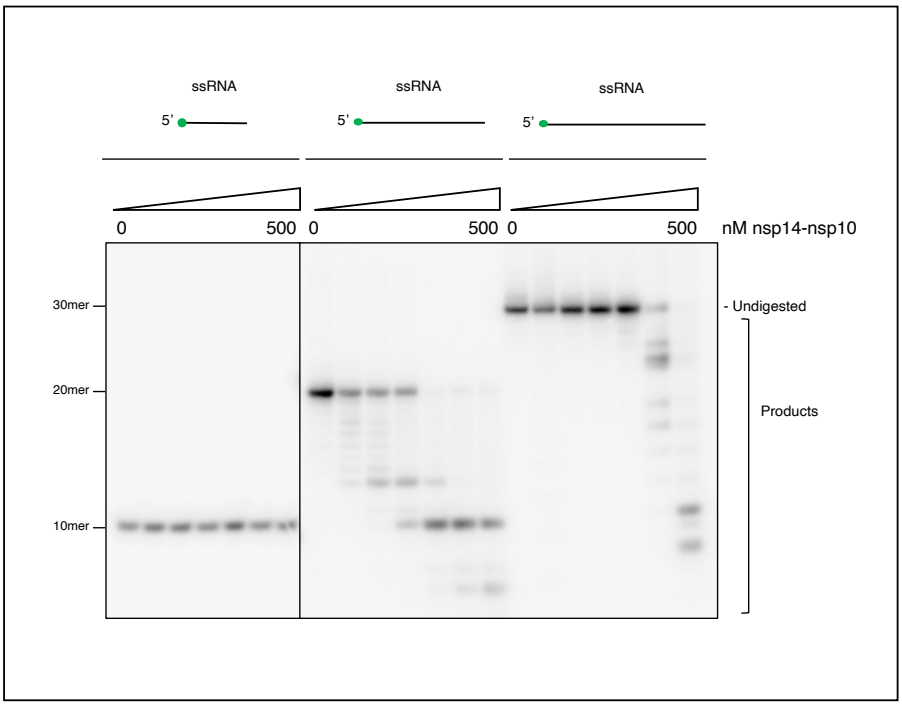


Figure 7.5: Nsp14-nsp10 is an RNA nuclease with a preference for 20-mer substrates
Nsp14-nsp10 is able to digest a 20-mer and a 30-mer ssRNA substrate, but not a 10-mer ssRNA. Oligos 1, 2 and 17 were used respectively.

followed by two additional major cuts, corresponding to cleaved 10^{th} and 13^{th} ribonucleotides from the 3′-end, releasing a 10-mer and 7-mer product (labelled with ** and *** respectively; Figure 7.6).

7.2 SARS-CoV-2 nsp14-nsp10 is able to process dsRNA and dsRNA substrates containing mismatches

Current reports identify the SARS-CoV-1 nsp14-nsp10 complex as a mismatch exonuclease, with a preferential activity on double-stranded RNA and substrates containing mismatches of varying lengths at their 3′-termini (M. Bouvet et al., 2012; Minskaia et al., 2006). In order to establish whether this is the case for the SARS-CoV-2 nsp14-nsp10 complex, RNA substrates containing 1-, 2-, 3- or 4-ribonucleotide mismatches at their 3′-termini as well as a substrate containing one internal mismatch at position 8 from the 3′-end were generated and used to test the

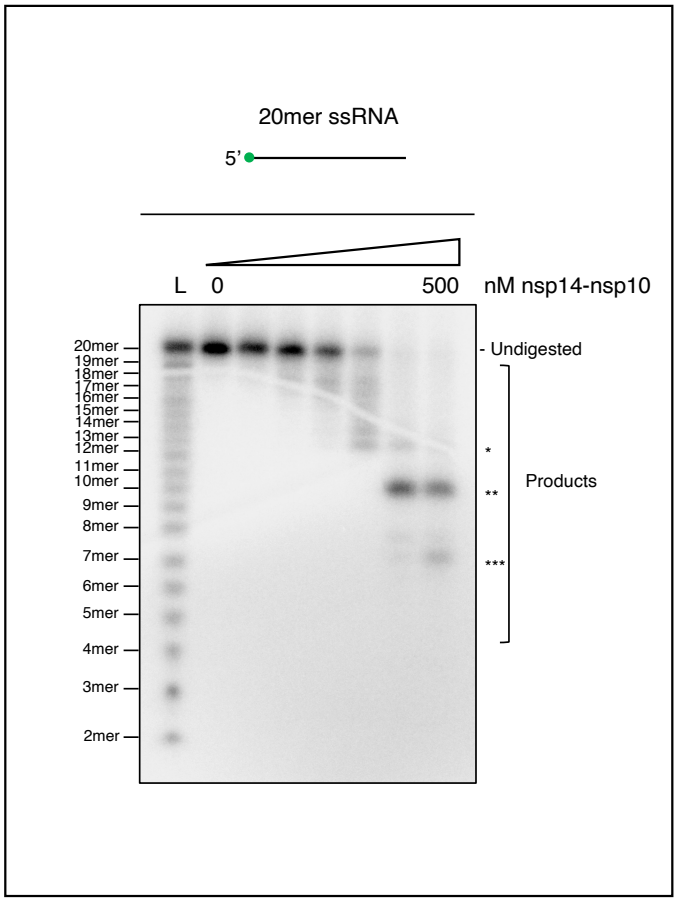


Figure 7.6: Nsp14-nsp10 is an RNA nuclease with a complex digestion pattern Nsp14-nsp10 is an RNA nuclease with a complex digestion pattern, digesting from the 3′ end in a single-nucleotide fashion until the 8^{th} ribonucleotide, then cleaving at the 10^{th} and 13^{th} ribonucleotide. A single nucleotide ladder was used to determine the size of the released products.

activity of the complex. Increasing amounts of nsp14-nsp10 were incubated with 10 nM of varying duplex RNA substrates with and without mismatches in order to observe nucleolytic digestion. As expected, SARS-CoV-2 nsp14-nsp10 is able to digest all structures containing mismatches, although nucleolytic activity was slightly reduced compared to standard dsRNA (Figure 7.7). Slight qualitative differences in the digestion products can be observed, likely due to the structural variations in the substrates introduced by the mismatches. Taken together, data supports the capacity of SARS-CoV-2 nsp14-nsp10 to process 3′ mismatched substrates although no preference for substrates containing mismatches was noted.

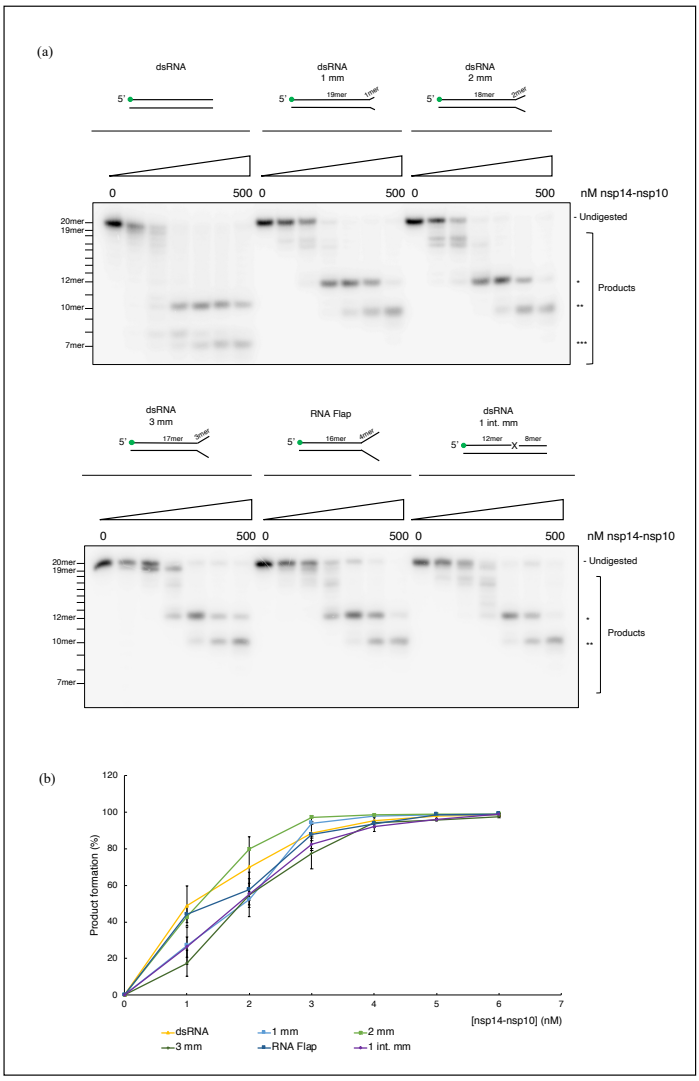


Figure 7.7: Sars-CoV-2 nsp14-nsp10 is able to digest substrates containing mismatches

(a) Nsp14-nsp10 is an RNA nuclease able to incise a variety of RNA substrates, including RNA substrates with mismatched termini and flaps with no preference for mismatched ribonucleotides. mm: mismatch; int. mm: internal mismatch.

(b) Product formation (%) was quantified for the gel above comparing nsp14-nsp10 nuclease activity on dsRNA and RNA substrates containing termini and internal mismatches as outlined in methods and materials. All data are shown as mean ± s.e.m, and at least three biological replicates were used for each substrate. Nsp14-nsp10 shows no preference for mismatched oligonucleotides compared to dsRNA. mm: mismatch; int. mm: internal mismatch.

7.2.1 Nsp14-nsp10 exhibits a ribonuclease activity that reaches beyond its classical role as a proof-reading nuclease

Although SARS-CoV-2 is able to process RNA substrates containing mismatches, these do not appear to be preferred to simplex duplex substrates. Considering how little the biochemical activity of the SARS-CoV-2 nsp14-nsp10 has been investigated, further experiments were performed to obtain an extensive image of the requirement for nsp14-nsp10 activity on a variety of RNA structures, sequences and modifications.

Initially, the activity of nsp14-nsp10 was compared on a variety of single-stranded, 20-mer RNA substrates to test the RNA sequence dependency of the complex. In addition to the 20-mer ssRNA that has been used thus far, three additional RNA substrates were selected on the basis of mixed sequences and having a negligible capacity to form stable secondary structures under the reaction conditions used (as determined by the calculated minimum free energy values, see materials and methods). Results show no distinguishable difference in pattern of digestion between all oligos, suggesting that the activity of nsp14-nsp10 is not sequence specific on RNA substrates of mixed sequences (Figure 7.8).

Interestingly, the activity of nsp14-nsp10 is dramatically reduced and qualitatively altered when presented with 20-mer poly(A), poly(U), poly(C) or poly(G) ssRNA substrates. For poly(U), poly(A) and poly(C), step-wise digestion was observed at higher concentrations of nsp14-nsp10 (above 125 nM) curtailing at nucleotide 11 (for poly(U)) or nucleotide 8/9 (for poly(A) and poly(C)). Intriguingly, the poly(G) substrate was completely inhibitory to nsp14-nsp10 (Figure 7.9).

Next, nsp14-nsp10 activity was tested on sequence-identical ssRNA, dsRNA and hybrid RNA:DNA substrates (where the RNA strand is 5′-radiolabelled) to examine whether the SARS-CoV-2 nsp14-nsp10 complex has a preference for a certain structures of RNA-containing substrate (Figure 7.10). Contrary to the SARS-CoV-1 complex, which exhibits a clear preference for dsRNA, the SARS-CoV-2 complex shows no preference for dsRNA over ssRNA (M. Bouvet et al., 2012, Figure 7.10). The complex appears to be agnostic to the nature of the strand

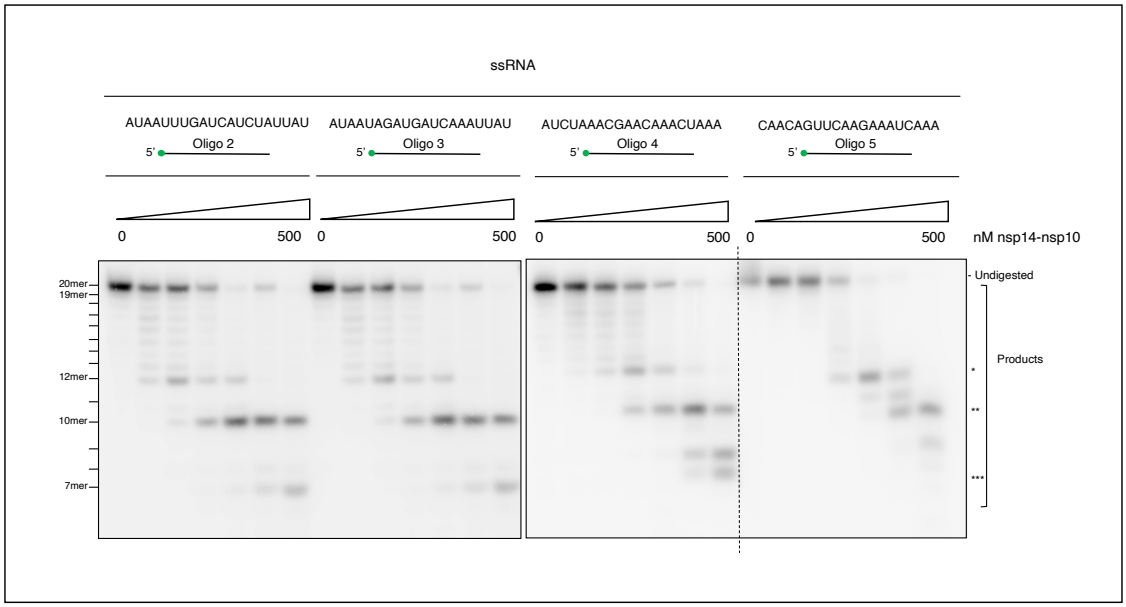


Figure 7.8: Nsp14-nsp10 activity is not sequence specific on a mixed-sequence RNA substrate
The nuclease activity of nsp14-nsp10 is not sequence specific on a mixed-sequence substrate. The complex shows indistinguishable digestion patterns on four 20-mer ssRNA substrates of different sequences and containing all four ribonucleotides. Oligos 2, 3, 4 and 5 were used respectively (see materials and methods).

annealed to the RNA substrate strand, since the RNA component of the RNA:DNA hybrid was digested as efficiently as the ssRNA substrate (Figure 7.10).

Consistent with its role as an RNA nuclease, the SARS-CoV-2 nsp14-nsp10 was unable to digest a ssDNA substrate, or the DNA strand of an RNA:DNA hybrid. Strikingly, on a ssDNA substrate containing an embedded ribonucleotide (analogous to the types of substrate processed by the ribonucleotide excision repair (RER) enzymes (Williams, Lujan, and Kunkel, 2016), nsp14-nsp10 was able to incise the substrate adjacent to the ribonucleotide at higher concentrations (above 125 nM; Figure 7.11). In mammalian cells, RER, the most common repair pathway for the removal of genomic embedded rNMPs, is initiated by RNase H2, essential for genomic integrity (Sparks et al., 2012; Nava et al., 2020). By analogy, this suggests an important role for nsp14-nsp10 in viral genome maintenance.

Finally, RNA substrates containing common modifications were tested. The impact of following modifications on nsp14-nsp10 were assessed: 6-methyladenine, the artificial base modification 2-methyladenine and the variant RNA base inosine.

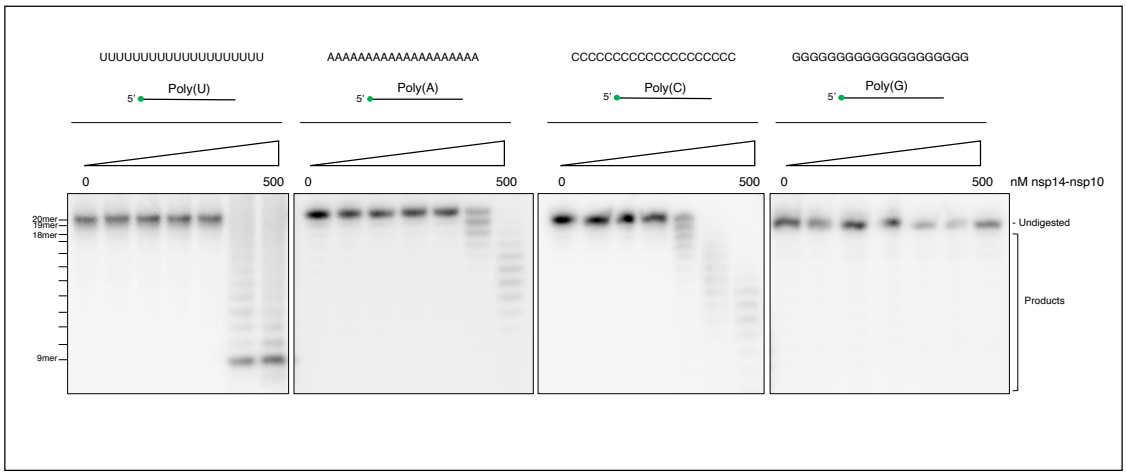


Figure 7.9: Nsp14-nsp10 shows altered activity profiles on repetitive sequences
When presented with 20-mer Poly(U), Poly(A), Poly(C) and Poly(G) ssRNA (oligos 11,12, 13 and 14 respectively, see materials and methods), nsp14-nsp10 shows reduced and qualitatively altered activity, with a single nucleotide step-wise digestion from the 3′-end curtailing at the 9^{th}-11^{th} nucleotide from the 3′-end.

6-methyladenine is the most common RNA modification found in cytoplasmic RNA and has been associated with viral evasion by the host innate immune system (Manners, Baquero-Perez, and Whitehouse, 2019). Inosine, generated by A-to-I editing has been reported to enhance viral recognition by the innate immune sensors (Netzband and Pager, 2020). When any of these modified bases was placed two ribonucleotides from the 3′-terminus of the 20-mer substrate, none affected, either qualitatively or quantitatively, activity of nsp14-nsp10 when compared with an unmodified RNA substrate (Figure 7.12). The SARS-CoV-2 nsp14-nsp10 complex is therefore able to digest past some of the common chemical modifications of RNA.

7.3 SARS-CoV-2 nsp14-nsp10 displays both 3′-exonuclease and a newly-described endonucleolytic activity

To better understand the complex digestion pattern of the SARS-CoV-2 nsp14-nsp10 complex, its activity was compared on sequence-identical 20-mer ssRNA substrates containing either a 3′-hydroxyl, 3′-phosphate, or 3′-biotin moiety (Figure 7.13). Unexpectedly, activity was observed on all three substrates, albeit qualitatively and

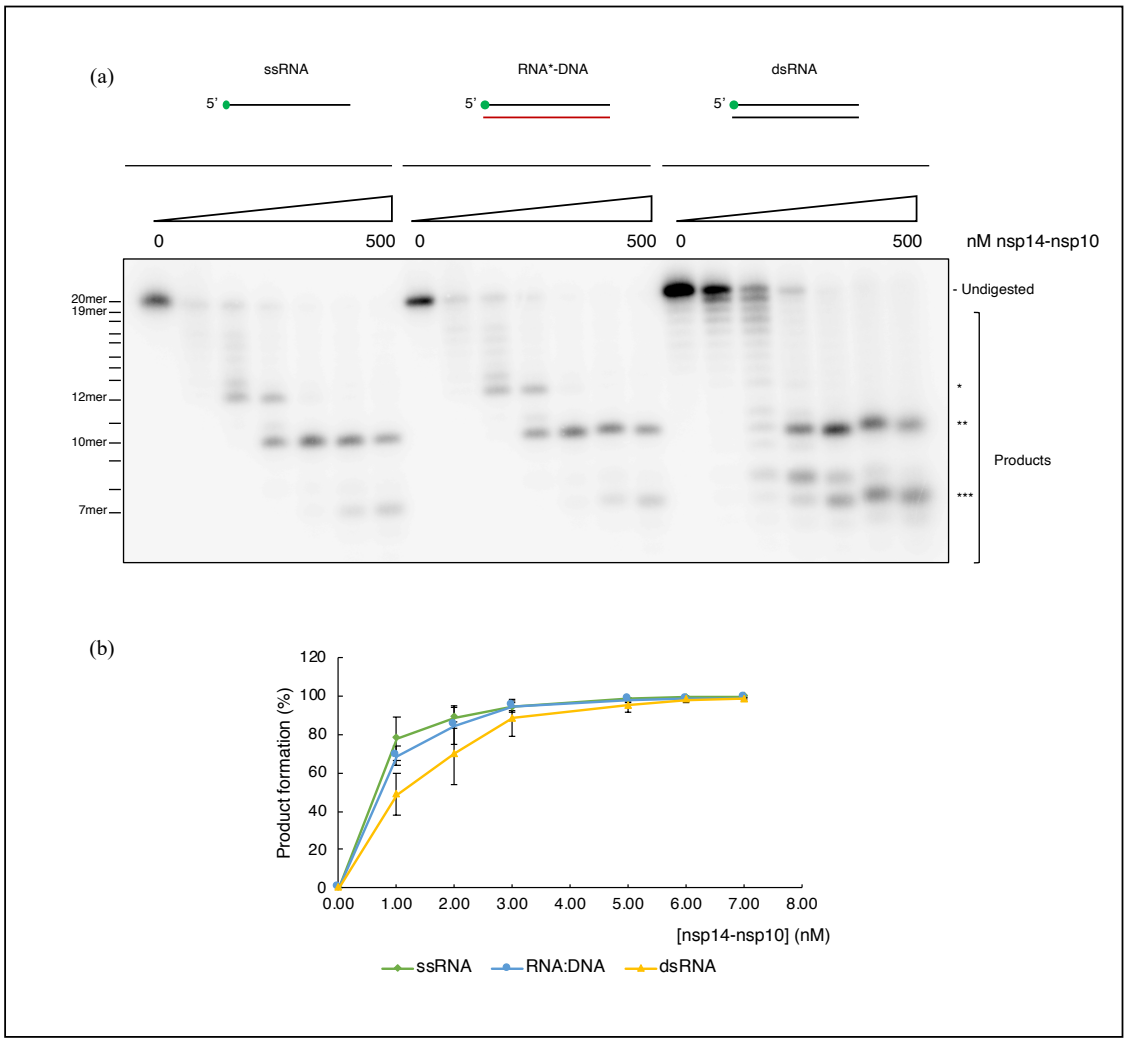


Figure 7.10: Nsp14-nsp10 is an RNA nuclease able to process a variety of RNA-containing substrates
(a) Nsp14-nsp10 is able to process ssRNA, the RNA strand of an RNA:DNA hybrid, and dsRNA with no preference for double-stranded substrates. For all structures, labelled oligo is 2 (see Materials and Methods).
(b) Product formation (%) was quantified for the gel above comparing nsp14-nsp10 nuclease activity on ssRNA, the RNA strand of an RNA:DNA hybrid and dsRNA as outlined in materials and methods. All data are shown as mean ± s.e.m, and at least three biological replicates were used for each substrate. Nsp14-nsp10 shows equivalent activity on ssRNA and the RNA strand of an RNA:DNA hybrid with a slight decrease in activity for dsRNA.

quantitatively variable. The characteristic laddering of products was present on the substrates bearing a 3′-hydroxyl and a 3′-phosphate group but were abrogated when the 3′-end was replaced with a biotin group. While the activity of nsp14-nsp10 is comparable on the 3′-hydroxyl and 3′-phosphate substrates, only the 10-mer product

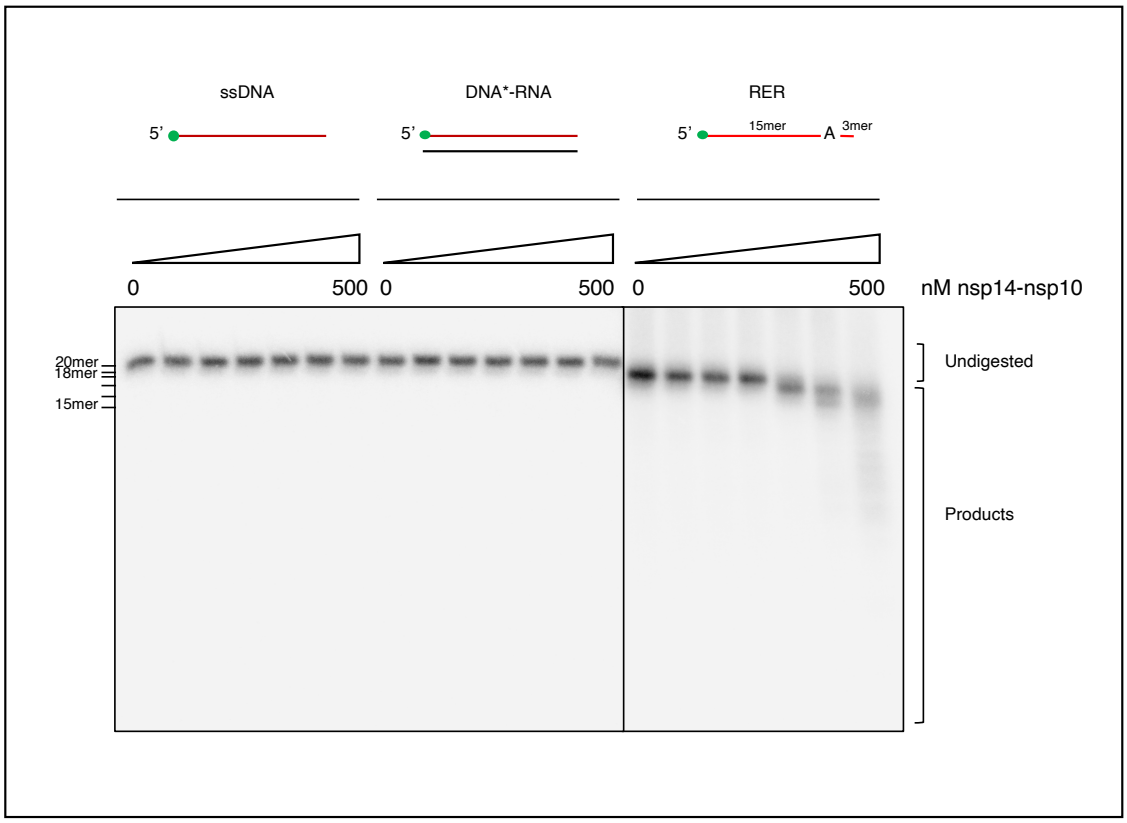


Figure 7.11: Nsp14-nsp10 activity is specific to RNA only
Nsp14-nsp10 has no discernible nuclease activity on ssDNA or the DNA strand of an DNA:RNA hybrid. Interestingly, the complex is able to incise around an embedded ribonucleotide in a ssDNA oligo at concentrations above 125 nM.

was predominantly observed at higher concentrations (above 250 nM) of nsp14-nsp10 on the 3′-biotin substrate. On this substrate, the 7-mer product was also observed at the highest concentration of nsp14-nsp10, albeit very faintly. Together, these results strongly suggest that the SARS-CoV-2 nsp14-nsp10 not only has a 3′-exonuclease activity, which is dependent upon the presence of an unblocked 3′-RNA terminus (leading to the laddering products), but also an endonucleolytic activity, cleaving at positions further from the 3′-end. These incisions occur independently of the presence of a 3′-hydroxyl or phosphate group, and therefore do not require a free 3′-terminus. This newly-identified endonucleolytic activity has not been previously reported, neither for the SARS-CoV-1 nor SARS-CoV-2 nsp14-nsp10 complex. In fact, previous studies of the SARS-CoV-1 nsp14-nsp10 complex concluded it primarily exhibits solely a 3′-exonuclease activity, that was reduced even in the

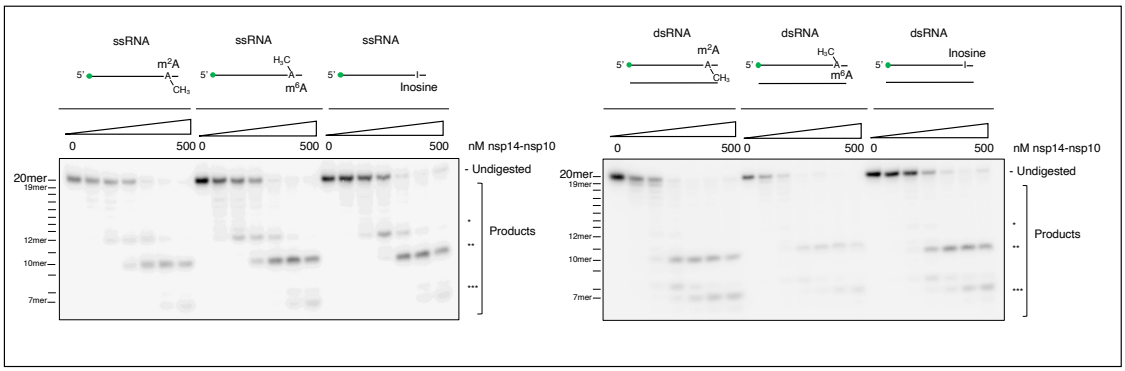


Figure 7.12: Nsp14-nsp10 activity is able to process common chemical modifications of RNA
Nsp14-nsp10 is able process common chemical modifications of RNA, namely 2-methyladenine (m^2A), 6-methyladenine (m^6A) and inosine (I), both ssRNA and dsRNA with no apparent change compared to ssRNA or dsRNA.

presence of a phosphate group at the 3′-termini (M. Bouvet et al., 2012).

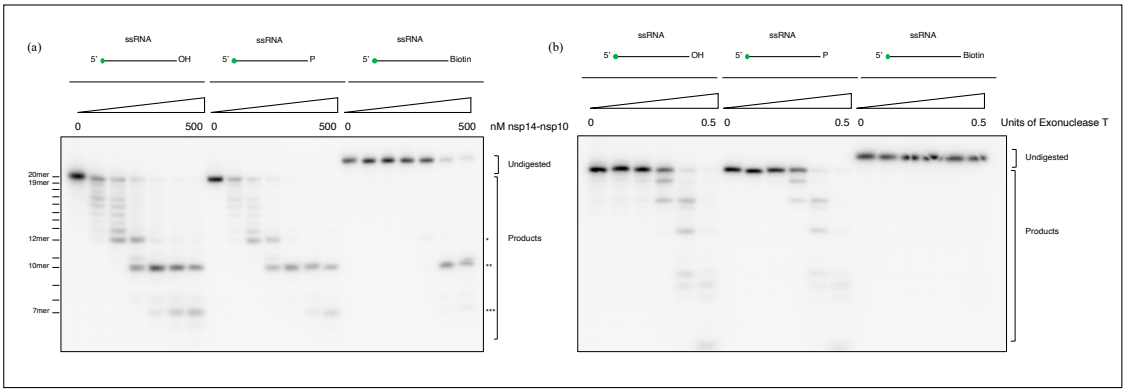


Figure 7.13: Nsp14-nsp10 exhibits both 3′-exonuclease activity and a newly-described endonucleolytic activity
(a) Nsp14-nsp10 is an RNA exo- and endo- nuclease. With a substrate containing a 3′-biotin group, the characteristic laddering of the substrate is lost and only endonucleolytic cleavage at the positions furthest from the 3′-end can be observed while the substrates with a 3′- hydroxyl or phosphate exhibit nearly identical product formation. Oligos 2, 15 and 16 were used respectively (see materials and methods).
(b) Exonuclease T was used to demonstrate a classic RNA exonuclease activity. Increasing concentrations of protein (0 to 0.5 U) were incubated with substrate at 30oC for 45 min and reactions were subsequently analysed by 20% denaturing PAGE. With a substrate containing a 3′-biotin group, no activity is observed confirming ExoT has no endonuclease activity while the substrates with a 3′- hydroxyl or phosphate exhibit nearly identical product formation. Oligos 2, 15 and 16 were used respectively (see Materials and Methods).

To further validate this finding, the same substrates as detailed above where incubated with a known RNA exonuclease, Exonuclease T. As expected, Exonu-

clease T was able to efficiently digest both the 3′-hydroxyl and the 3′-phosphate substrates but its activity was clearly blocked by the 3′-biotin group (Figure 7.13). These observations confirm that nsp14-nsp10 does indeed possess *bona fide* endoribonuclease activity.

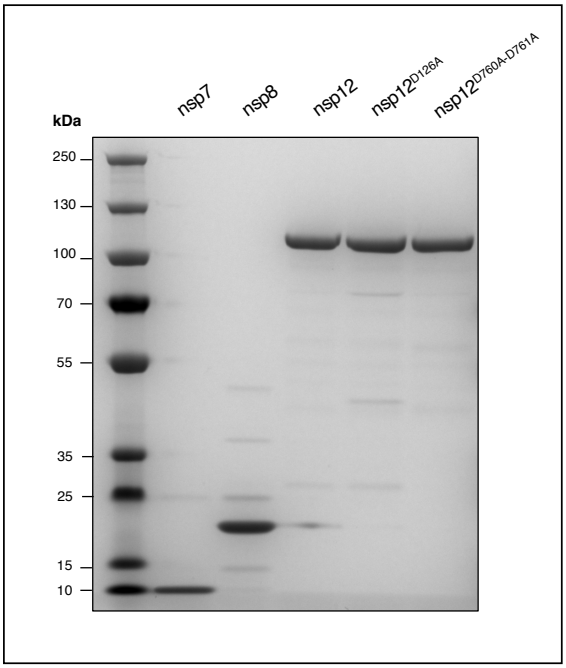


Figure 7.14: Purity of the SARS-CoV-2 nsp12-7-8 polymerase complex
SDS-PAGE of purified nsp7, nsp8, np12, nsp12^{D126A} and nsp12$^{D760A\text{-}D761A}$ showing the purity of the purified proteins. Predicted molecular weights are 9 000 Da for nsp7, 22 000 Da for nsp8, 106 000 Da for nsp12, nsp12^{D126A} and nsp12$^{D760A\text{-}D761A}$. Purified nsp7, nsp8, np12, nsp12^{D126A} and nsp12$^{D760A\text{-}D761A}$ were supplied by Dr. H. Fan and Dr. J. Keown at the Division of Structural Biology (STRUBI) at the University of Oxford.

7.4 The SARS-CoV-2 polymerase nsp12-7-8 complex enhances nsp14-nsp10 nuclease activity on a variety of RNA substrates

It has previously been proposed that the SARS-CoV-1 nsp14-nsp10 proof-reading nuclease plays an important role in maintaining viral genome stability, possibly through an interaction with the SARS-CoV-1 polymerase complex nsp12-7-8 (Ferron et al., 2018; Subissi et al., 2014; Imbert et al., 2008). Nsp12 designates the polymerase subunit, nsp8 is the primase and nsp7 is an essential accessory unit to

both nsp8 and nsp12. Given the high homology between the SARS-CoV-1 and SARS-CoV-2 replication machinery, the effect of the SARS-CoV-2 polymerase complex nsp12-7-8 on the nuclease activity of SARS-CoV-2 nsp14-nsp10 was investigated.

Purified protein from the SARS-CoV-2 polymerase complex, referred to as nsp12-7-8, were obtained from Dr. H. Fan and Dr. J. Keown at the Division of Structural Biology (STRUBI) at the University of Oxford as part of our collaboration. The following proteins were obtained: Wild-type nsp12, nsp7 and nsp8 as well as two mutant forms of nsp12: nsp12^{D126A} and nsp12^{D670A,D761A}. nsp12^{D126A} is a NiRan domain active site mutant and nsp12^{D670A,D761A} a polymerase active site double mutant. The purity of all components obtained was assessed by SDS-PAGE, showing all proteins to be purified to near homogeneity (Figure 7.14). To further confirm no contaminants that were able to digest RNA were present, a nuclease assay on a simple RNA duplex was performed on each subunit obtained (Figure 7.15). Nsp14-nsp10 was used a positive control. No other degradation of the RNA substrate was observed, ruling out the possibility of any intrinsic nuclease activity of the nsp12-7-8 polymerase subunits or purification contaminants.

Initially, the effect of adding the nsp12-7-8 polymerase complex to the SARS-CoV-2 nsp14-nsp10 was investigated on both simple duplex RNA and duplex RNA containing a variety of mismatches. To this end, SARS-CoV-2 nsp12-7-8 polymerase complex (at a 1:3:3 molar ratio) was allowed to form in the presence of 10 nM substrate at 37°C for 5 min before the addition of 100 nM of nsp14-nsp10 at a 1:2 molar ratio of nsp14-nsp10 complex to nsp12-7-8 complex. Reactions were then incubated at 37°C for 45 min and subsequently analysed by 20% denaturing PAGE to visualize product formation. Interestingly, a marked enhancement of the nuclease activity of nsp14-nsp10 can be observed for all substrates (compare lanes + and ++ in Figure 7.16.

This novel stimulatory effect was then further tested on a variety of sequence diverse RNA substrates. In particular, it was of interest to investigate whether the presence of the nsp12-7-8 polymerase complex could overcome the largely inhibitory effect of repetitive RNA sequences on the activity of nsp14-nsp10. To this effect,

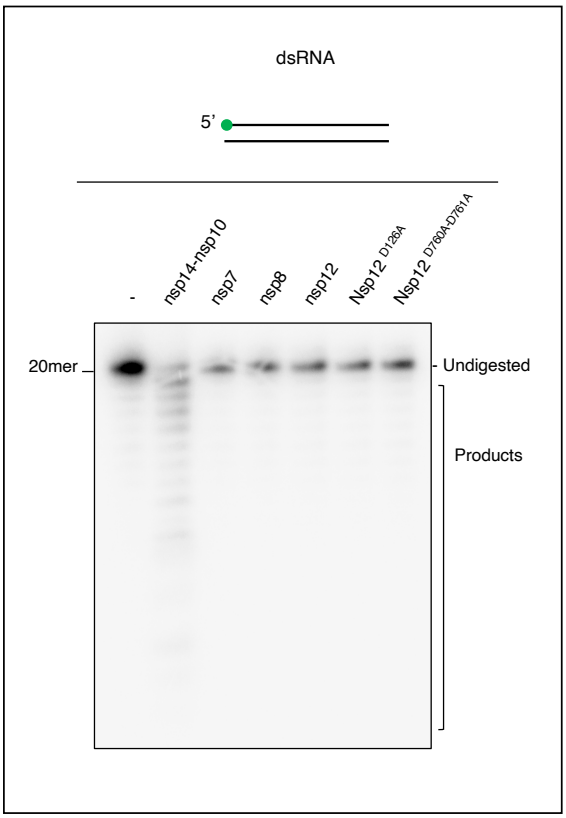


Figure 7.15: The proteins of the SARS-CoV-2 nsp12-7-8 polymerase complex have no intrinsic nuclease activity
All components of the SARS-CoV-2 polymerase complex used in this study (nsp7, nsp8, nsp12 as well as catalytic mutants nsp12^{D126A} and nsp12$^{D760A\text{-}D761A}$; 500nM) were incubated with substrate at 37°C for 45 min, reactions were subsequently analysed by 20% denaturing PAGE to visualize product formation. No digestion of the RNA substrate was observed. Nsp14-nsp10 (100nM) was used as a positive control. All oligos used are indicated in materials and methods.

nsp12-7-8 was added to nsp14-nsp10 (as described above) in the presence of poly(A), poly(G), poly(C) and poly(U) ssRNA (Figure 7.17). In these cases, nsp12-7-8 was still able to differentially stimulate the activity of nsp14-nsp10 but not in a uniform manner: while the poly(A), poly(C) and poly(G) RNA substrates showed very little nucleolytic digestion, both with and without nsp12-7-8, the effect was much more pronounced on uracil-containing oligonucleotides: Poly (U), a uracil-rich ssRNA sequence as well as duplex poly(A) and poly(U) substrates (with either the poly(U) or poly(A) strand labelled) exhibited the greatest stimulatory effect of nsp14-nsp10 activity in the presence of nsp12-7-8.

To gain insight into this marked stimulatory effect of nsp12-7-8 on nsp14-nsp10,

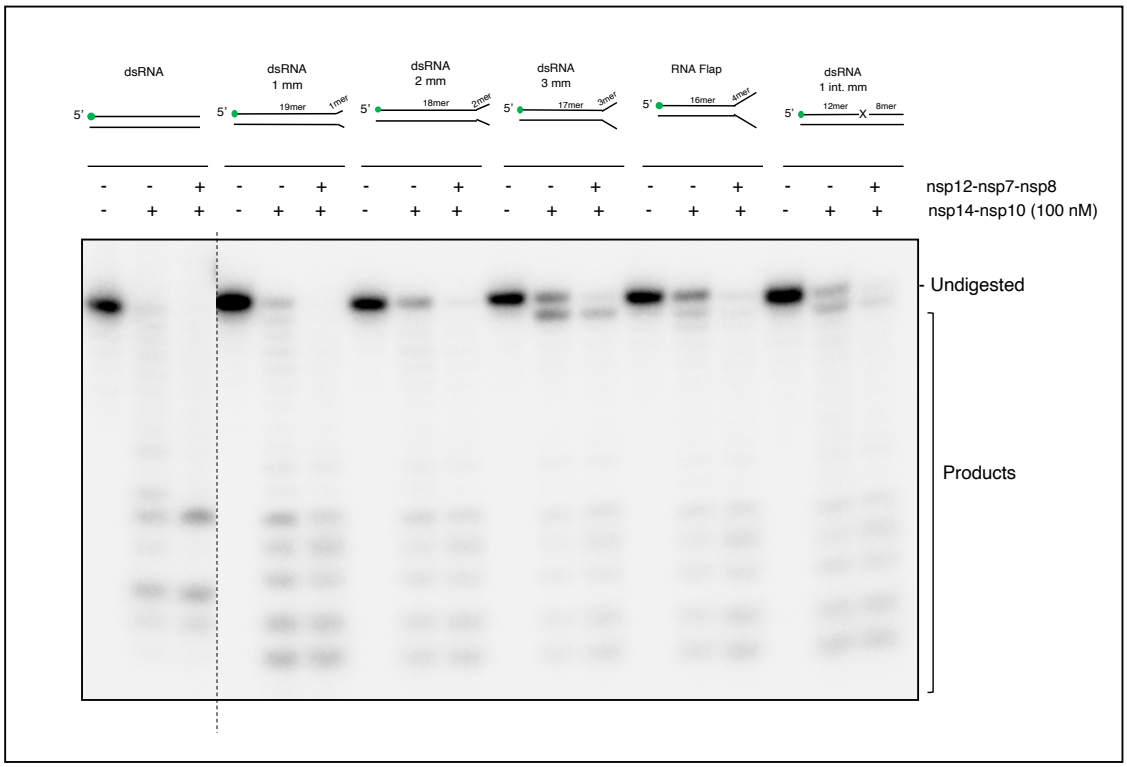


Figure 7.16: The SARS-CoV-2 polymerase nsp12-nsp7-nsp8 complex enhances nsp14-nsp10 nuclease activity on a variety of RNA substrates
In the presence of the nsp12-7-8 polymerase complex, nsp14-nsp10 shows enhanced nuclease activity on a variety of RNA substrates, including RNA substrates with mismatched termini and flaps with no observed differential effects. mm: mismatch; int. mm: internal mismatch. All oligos used are indicated in the Materials and Methods.

the effect of each individual subunit of the polymerase complex, nsp12, nsp7 and nsp8 as well as different combinations of thereof was tested. Strikingly, the nucleolytic stimulation of nsp14-nsp10 is only observed in the presence of the primase nsp8 (and any combination of nsp8 and other polymerase components). To further confirm the stimulatory effect observed is uncoupled from nsp12 polymerase activity, two mutant forms of nsp12 were also tested: nsp12^{D126A}, and nsp12$^{D760A\text{-}D761A}$; the former being a NiRan domain active site mutant, and the latter a polymerase active site double mutant. Neither of these mutant forms of nsp12 affected the stimulation observed in the presence of nsp8 (Figure 7.18). This result is further strengthened by a titration experiment in which the activity of nsp14-nsp10 alone is compared to the activity of the nuclease complex in the presence of nsp8. Increasing concentrations of nsp14-nsp10 were incubated with ss- or dsRNA and compared to the same range

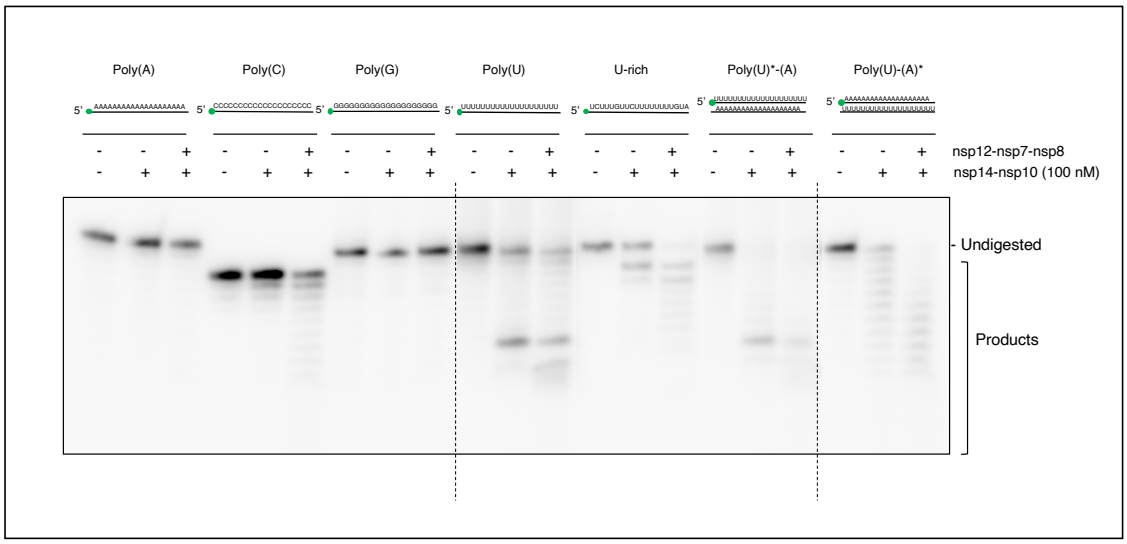


Figure 7.17: The SARS-CoV-2 polymerase nsp12-nsp7-nsp8 complex enhances nsp14-nsp10 nuclease activity particularly on uracil-containing substrates
When presented with 20-mer Poly(A), Poly(C), Poly(G), Poly(U) and U-rich ssRNA (oligos 12, 13, 14, 11 and 25 respectively, see materials and methods) as well as double-stranded Poly(U)*-(A) and Poly(U)-(A)*, nsp14-nsp10 shows a more profound stimulation on repetitive sequences when nsp12-7-8 is present. In particular, enhancement of nsp14-nsp10's nuclease activity by nsp12-7-8 is the most pronounced on uracil-containing substrates.
* indicates labelled strand.

of concentrations of nsp14-nsp10 but with the addition of a fixed concentration of 300 nM nsp8 (Figure 7.19). In the lanes present with nsp8, a clear left shift in the pattern of nucleolytic digestion as the activity of nsp14-nsp10 becomes significantly more marked at lower concentrations (15 nM) on both ss- and dsRNA further confirming that the primase nsp8 is essential in the stimulation of nsp14-nsp10 nuclease activity. Whether this stimulation is mediated via protein-protein interactions or is substrate-mediated remains to be investigated.

7.5 Conclusion

Together, this section presents the first detailed biochemical analysis of the SARS-CoV-2 nsp14-nsp10 protein complex, a key factor in genome duplication of the SARS-CoV-2 virus. Analogous to the SARS-CoV-1 nsp14-nsp10 complex, the SARS-CoV-2 nsp14-nsp10 is an RNA exonuclease, where the presence of nsp10 enhances

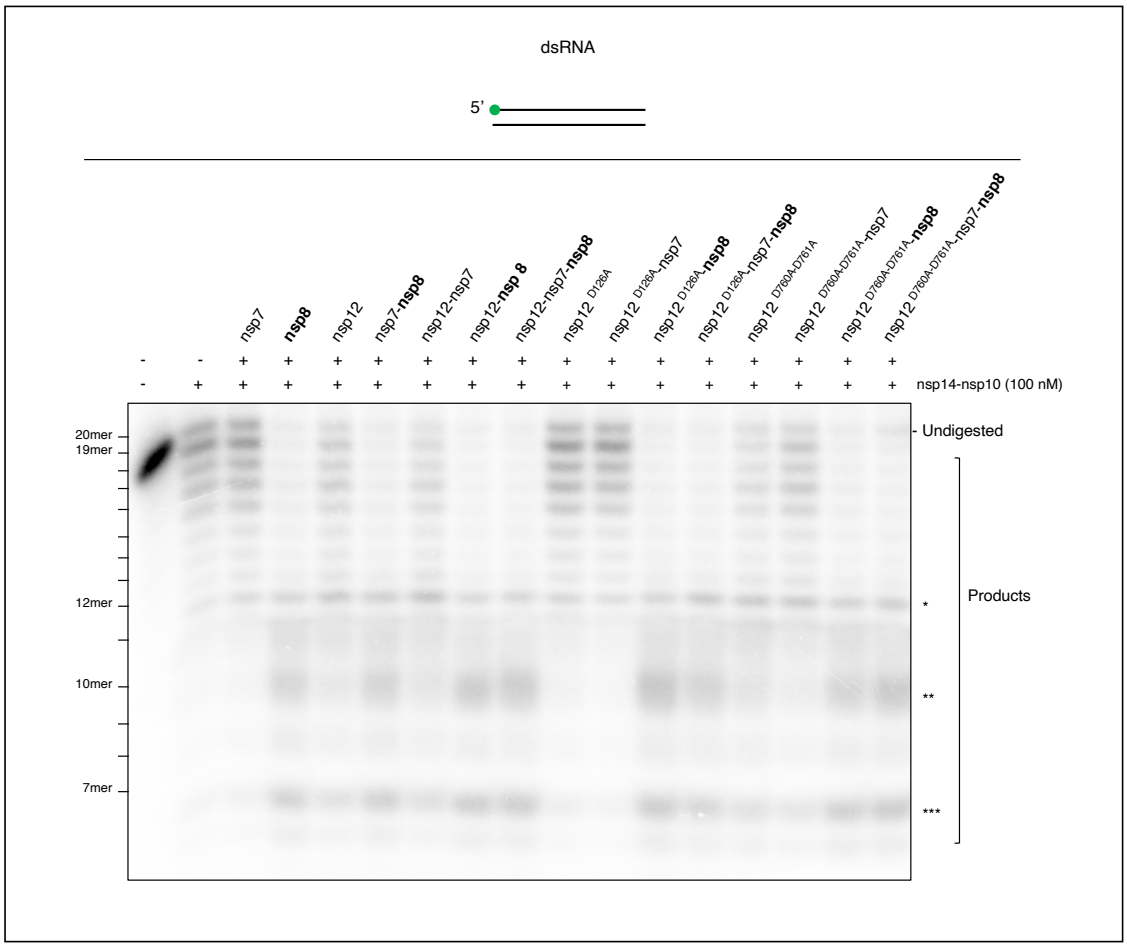


Figure 7.18: The SARS-CoV-2 primase nsp8 is the key factor in nsp14-nsp10 activity stimulation
The SARS-CoV-2 primase, nsp8, is the key factor in nsp14-nsp10 activity stimulation. The addition of each individual polymerase subunit, namely nsp12, nsp7, and/or nsp8, as well as different combinations reveal nsp8 as the major enhancer of nsp14-nsp10 activity. Mutant nsp12^{D126A} and nsp12$^{D760A\text{-}D761A}$ show no major stimulation of activity of nsp14-nsp10 compared to nsp8 suggesting stimulation of nuclease activity is uncoupled from polymerase activity.

the activity of nsp14. nsp14-nsp10 preferentially digests a 20-mer substrate in a 3′ to 5′ direction. Interestingly, we also find a previously undescribed endonuclease activity: nsp14-nsp10 is still able to incise an RNA substrate with a 3′-biotin end, contributing to the complex digestion pattern observed.

Nsp14-nsp10 is able to digest a variety of RNA substrates, including RNA substrates containing terminal mismatches and common RNA modifications. Consistent with a role in maintaining viral genome integrity during replication, we show the activity of nsp14-nsp10 is enhanced by the addition of the SARS-CoV-2

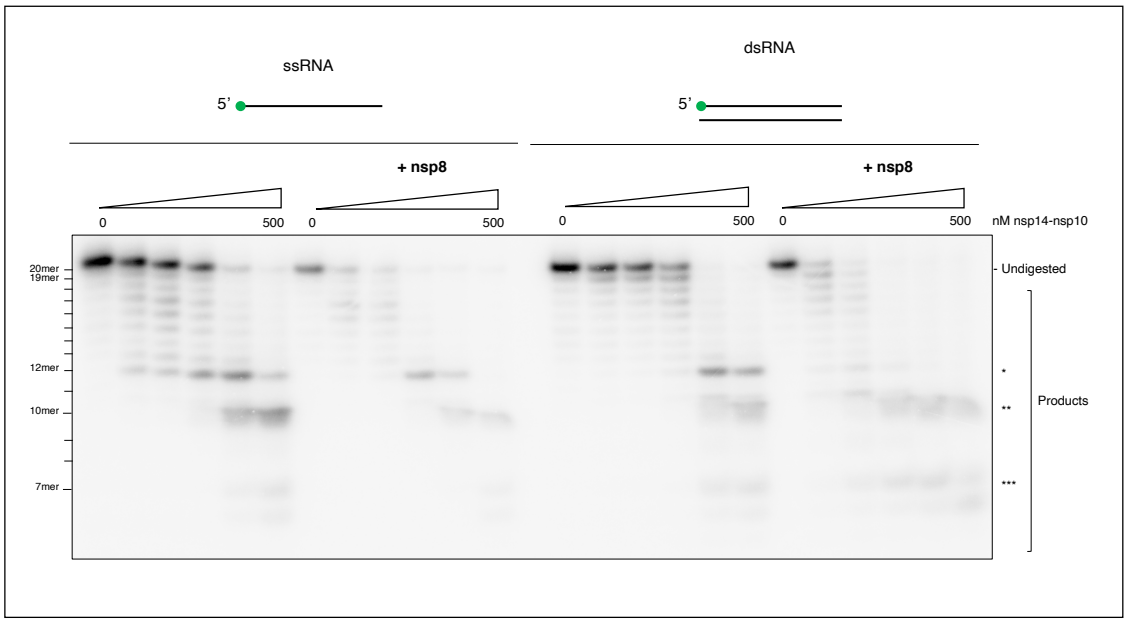


Figure 7.19: The SARS-CoV-2 primase nsp8 enhances the activity of nsp14-nsp10
The presence of nsp8 stimulates the nuclease activity of SARS-CoV-2 nsp14-nsp10. Increasing concentrations of nsp14-nsp10 were incubated with ss- or dsRNA (Oligo 2 and 3, see Materials and Methods) and a fixed concentration of 300nM nsp8 at 37°C for 45 min and reactions subsequently analysed by 20% denaturing PAGE to visualize product formation.

polymerase complex nsp12-nsp7-nsp8. In particular, we identify nsp8 as the key factor in modulating this enhancement in activity.

This suggests a role of nsp14-nsp10 in maintaining replication fidelity that goes beyond classical proofreading, which will be further discussed in Chapter 9.

8

Discussion and Future Perspectives for the C15orf41-Codanin-1 study

8.1 Discussion

C15orf41 is a previously uncharacterised protein causative of CDA-I, a rare inherited anaemia uniquely characterised by spongy heterochromatin abnormalities in patient erythroblasts (Babbs et al., 2013). This thesis identifies the function of C15orf41 as an RNA exonuclease and characterises its activity using a variety of complementary cellular, biophysical and biochemical methods.

The proteins encoded by *C15orf41* and *CDAN1*, the two known disease genes causing CDA-I, were previously very poorly characterised. Although C15orf41 and Codanin-1 have recently been reported to form a tight complex, these results were performed using over-expression studies and/or in transformed cell lines, somewhat clouding the interpretation of the results (Shroff et al., 2020; Swickley et al., 2020; Olijnik et al., 2021). We therefore establish complex formation of C15orf41-Codanin-1 at endogenous levels in our newly-engineered HUDEP-2 cell lines, that serve as erythroid-specific cell models for CDA-I. C15orf41 and Codanin-1 form a stable complex, able to withstand high concentrations of salt. This interaction is disrupted by the p.L178Q patient mutation in C15orf41, located in the predicted nuclease domain of the protein, suggesting disruption of C15orf41-Codanin-1 complex likely plays a pathogenic role in CDA-I. C15orf41 and Codanin-1 complex formation has been shown to be mediated by the C-terminus of Codanin-1 (Shroff et al., 2020; Swickley et al., 2020). The results presented in this thesis show how the

addition of the C-terminus of Codanin-1 enhances and modulates the nuclease activity of C15orf41 dramatically. Loss of this complex formation would therefore directly impair C15orf41 function.

Using the HUDEP-2 CDA-I model, we also find that C15orf41 levels are affected by both the p.L178Q patient mutation in C15orf41 as well as the p.R1042W patient mutation in Codanin-1, whereas Codanin-1 levels are only affected by the p.L178Q mutation in C15orf41. C15orf41 and Codanin-1 likely contribute to each other's mutual stability and integrity, this is supported by Codanin-1 knockdown experiments destabilising C15orf41 levels (Olijnik et al., 2021). Considering the 3xFLAG-C15orf41^{L178Q} HUDEP-2 cells display more severe erythroblast abnormalities STREP-Codanin-1^{R1042W} HUDEP-2, this strengthens the idea that adequate levels of C15orf41 are essential for basic erythroblast maintenance and that disruption of this balance is pathogenic. Interestingly, Olijnik et al. report no change in C15orf41 or Codanin-1 protein levels in CDA-I patient erythroblasts compared to healthy control samples when CDA-I patient mutations p.Y94C and p.C156Y are present in C15orf41. This observation suggests an additional, C15orf41-level independent, mechanism that leads to CDA-I. It would be of interest to assess C15orf41-Codanin-1 complex formation in cells harbouring the above-mentioned mutations in C15orf41 in order to establish whether the disease phenotype observed in these cases results (at least partially) to loss of complex formation.

Secondary structures of C15orf41 predict the protein possesses a nuclease domain analogous to nucleases of the PD-(D/E) XK superfamily of nucleases. We therefore purified full-length C15orf41 and performed a full biochemical characterisation which identified C15orf41 as a novel RNA exonuclease. We show that C15orf41 alone is able to process ssRNA only with no discernible activity on DNA substrates. We also confirm *in vitro* complex formation between C15orf41 and the C-terminus of Codanin-1 and provide a full biochemical characterisation of the complex. C15orf41-Codanin-1 is able to process RNA-containing structures, both of single-stranded and double-stranded nature with the main requirement being a free 3′ RNA terminus for the complex to initiate its characteristic RNA laddering in a 2- then 1-nucleotide step

fashion until it reaches a critical minimal length where it terminates. The complex exhibits a preference for 20-mer RNA substrates, although it is able to process larger RNA fragments (albeit at a lower efficiency) but not shorter, presumably requiring a minimal length to interact with the substrate.

In order to link the newly-identified biochemical activity of C15orf41 to a possible biological role for the protein, two complementary methods were employed. Firstly, a variety of RNA-containing structures representing relevant biological intermediates were tested for C15orf41 processing. In particular, common RNA-containing structures such as simple RNA-DNA hybrids, R-loops, embedded ribonucleotides or Okazaki-like structures were designed and assessed for C15orf41-Codanin-1 nuclease activity. To start, a simple RNA-DNA hybrid was tested and digestion comparable to double-stranded RNA observed. Both double-stranded RNA and the RNA-DNA hybrid constitute C15orf41's preferred substrates, revealing a possible biological role for C15orf41 in RNA-DNA hybrid resolution. This hypothesis is supported by the ability of C15orf41-Codanin-1 to incise the RNA fragment of an R-loop, as well as 6-methyladenine-modified RNA, a marker also present on the RNA moiety of R-loops (Abakir et al., 2020). Both RNA-DNA hybrids and R-loops represent essential transcription intermediates that, when unresolved, can lead to genomic instability and trigger chromatin state changes (Groh and Gromak, 2014; Richard and Manley, 2017). Considering the molecular basis of CDA-I is unknown, it is possible that accumulation of either RNA-DNA hybrids or RNA-containing structural intermediates due to ineffective C15orf41 activity could be a molecular trigger for the chromatin defects observed. However, intriguingly, the presence of the top DNA strand constituting the loop was not a requirement for C15orf41 activity, suggesting that the activity of this nuclease is probably not structure selective but rather just specific to the presence of ribonucleotides. We also investigate the ability of C15orf41 to discern individual and fragments of ribonucleotides embedded or adjacent to DNA. To this end, substrates mimicking an embedded ribonucleotide substrate as well as an Okazaki-like flap consisting of an RNA primer followed by a 3′-DNA end were tested. In both cases, incorporation

of ribonucleotides or defects in RNA primer removal during DNA replication can lead to the accumulation of DNA damage and aberrant chromosome alterations, causing devastating health conditions (Zheng and Shen, 2011; Reijns et al., 2012; Sassa, Yasui, and Honma, 2019). For both embedded ribonucleotides and Okazaki fragment-like structures, C15orf41 does not have the capacity to initiate digestion using a 3′-DNA end nor to specifically detect individual ribonucleotides within a DNA substrate. It is therefore unlikely C15orf41 plays a key role in either embedded-ribonucleotide resolution or DNA substrate resolution. Finally, we assessed the effect of common RNA modification of C15orf41-Codanin-1 nuclease activity, specifically 6-methyladenine, the most abundant internal modification present on mRNA and the modified base Inosine, critical in host immune system evasion (Roundtree et al., 2017; Nakahama and Kawahara, 2020). While C15orf41-Codanin-1 was able to process 6-methyladenine (and 2-methyladenine) modified bases, the presence of Inosine was found to be inhibitory. Considering the importance of the 6-methyladenine modification in mRNA stability and metabolism, it is possible that C15orf41 is an mRNA processing nuclease although how this relates to the CDA-I phenotype observed requires further investigation. Interestingly, the 6-methyladenine modification has been suggested to be present in most RNA:DNA hybrids, including the RNA strand of R-loops (Abakir et al., 2020; Marnef and Legube, 2020). In R-loops, this mark is able to modulate their stability, thereby regulating the accumulation and degradation of R-loops (Abakir et al., 2020). It could therefore also be hypothesised that C15orf41-Codanin-1 is a nuclease complex that specifically acts on 6-methyladenine modified RNA moieties, such as in R-loops (or RNA:DNA hybrids), thereby maintaining genomic stability. The inhibitory effect of Inosine is striking, likely due to conformational changes when an Inosine base is introduced, blocking C15orf41 binding and processing. Inosine is mainly present in transfer RNAs (tRNAs) or has been implicated to play a role in immune response, two roles that could possibly be excluded for C15orf41 (Alseth, Dalhus, and Bjørås, 2014; Nakahama and Kawahara, 2020).

Secondly, we also assessed the effects of the two patient founder mutations in C15or41, the p.Y94C and the p.L178Q mutants, on RNA nuclease activity. In both cases, patient mutant C15orf41 show dramatically reduced nuclease activity, albeit not completely abolished. This residual activity is expected and further strengthens the idea that complete loss of C15orf41 activity is not viable. It is likely the patient mutations in C15orf41 affect substrate interactions or cause misfolding of the nuclease domain. In fact, the L178Q mutation is within the predicted nuclease domain of C15orf41, altering a highly conserved hydrophobic leucine to a polar glutamine whilst the Y94C mutation is located in the predicted helix-turn-helix motif of the protein, likely to disrupt substrate interaction. Interestingly, both patient mutant forms of C15orf41 retained their ability to complex to Codanin-1 *in vitro*, even though *in vivo* studies show loss of complex formation when patient mutations are present within C15orf41. The defects in nuclease activity in the patient mutant C15orf41 complexed with Codanin-1 were less severe than when considering mutant C15orf41 alone, further highlighting the fact that C15orf41 and Codanin-1 complex formation is an essential step in conferring C15orf41 its full nuclease activity and loss of complex binding highly pathogenic. It is likely that the impairment of nuclease activity of C15orf41 is causative of the CDA-I phenotype.

It should also be noted that RNA substrates are likely to be labile and require careful handling during experiments, somewhat limiting the possible structures we could test. Labelling RNA substrates presents its own challenge as well, in particular when it comes to 3′-end labelling. 3′-end RNA labelling was attempted using ^{32}PcP and T4 RNA ligase but almost no signal was obtained, hindering nuclease assay performance. In this study, all substrates were therefore labelled with 5′-end ^{32}P-ATP to allow for product visualisation during these assays.

Together, we find that C15orf41-Codanin-1 is a very versatile RNA nuclease complex that is able to digest a variety of RNA structures and specifically requires a free 3′-RNA end to initiate nucleolytic processing. The association between C15orf41 and Codanin-1 is crucial, not only for maintaining C15orf41 levels but also for conferring C15orf41 its full nuclease activity. CDA-I patient mutations

p.Y94C and p.L178Q have a direct detrimental effect on this nuclease activity. Considering the p.L178Q mutation in C15orf41 also abolishes C15orf41-Codanin-1 complex formation in an erythroid specific CDA-I cell model, it can be inferred that part of the disease mechanism of CDA-I could be due to maintaining a balance between adequate C15orf41 levels and activity, both of which are likely modulated by Codanin-1. Considering the promiscuous nuclease activity of C15orf41, C15orf41 could be acting on replication/transcription intermediates containing RNA specifically that would arise from defect in cell-cycle progression in erythroblasts, a hypothesis that will be discussed next.

CDA-I is a dyserythropoietic disorder. By definition, CDA-I patients present defects in erythropoiesis leading to less abundant and abnormal erythroblast production. Erythropoiesis has previously been closely linked with cell cycle progression, where erythroblast progenitor progression through S-phase has been identified as the committing step to initiate terminal erythroid differentiation (Pop et al., 2010). Stalled replication and S-phase arrest, due to the accumulation of aberrant DNA or RNA intermediates, that are not being processed efficiently by C15orf41 (likely in complex with Codanin-1), could explain the defects in erythropoiesis progression resulting in anaemia as well as the observed heterochromatin phenotype. In fact, appearance of dsRNA or RNA:DNA hybrids in the form of R-loops at certain gene loci, both preferred substrates of the nuclease C15orf41, can result in transient formation of heterochromatin (Gullerova and Proudfoot, 2008). In agreement with the idea that CDA-I may arise from defects in S-phase progression during erythropoiesis, CDA-I patient erythroblasts display internuclear bridges, characteristic of regions associated with genomic instability (Y. Liu et al., 2014). The importance of cell cycle in erythropoiesis is further emphasised with reports of interferon alpha (IFNα) and its ability to induce cell cycle arrest (Szeps et al., 2003). IFNα is currently the only effective therapy against CDA-I but its mechanism is unknown. A possible hypothesis is that IFNα induces cell cycle arrest or slows down the rapidly-dividing erythroid progenitors. This would allow for repair of possibly accumulated aberrant replication-transcription intermediates

due to deficient C15orf41 activity and allow for more erythroid progenitors to complete terminal differentiation. This hypothesis remains to be investigated, possibly by understanding the nuclease's native substrate and analysis of cell-cycle and replication fork dynamics in CDA-I patient erythroblasts. It should be noted that previous experiments performed by Dr. L. Swift as part of our collaboration between the McHugh and Buckle lab at the MRC Weatherall Institute of Molecular Medicine revealed CDA-I patient erythroblasts showed no sensitivity to various DNA damaging agents, suggesting no inherit DNA damage and repair defects in this disease. Still, a possible role for C15orf41 in basic genome maintenance, possibly related to RNA processing in erythroblasts is conceivable and should be further investigated.

The unique heterochromatin defect in CDA-I patients further points towards a replication or heterochromatinisation defect in these cells. In fact, DNA replication and histone deposition are a tightly-coupled process. Heterochromatin factors have been reported to associate with histone chaperones such as chromatin assembly factor 1 (CAF1) during S-phase (Cam et al., 2009). CAF1 mediates nucleosome formation by depositing histones H3/H4 onto newly-replicated DNA (W. H. Liu et al., 2016). CAF1 has also been shown to be a downstream partner of the histone H3-H4 chaperone anti-silencing function 1 (Asf1) (Mello et al., 2002; Campos et al., 2010). Asf1 is a known binding partner to Codanin-1 so a possible interplay between C15orf41, Codanin-1 and Asf1 could possibly explain the observed heterochromatin defect (Ask et al., 2012; Olijnik et al., 2021). Considering the rapid S-phase present in erythroid progenitors, it is possible the high rate of DNA synthesis leads to structural changes in the chromatin, particularly if S-phase progression is affected in any way (Pop et al., 2010). C15orf41 could be acting on any aberrant structures that form during DNA replication and histone deposition. Low levels of C15orf41 or impaired C15orf41 function would therefore have a direct detrimental effect on those pathways, potentially leading to the phenotype observed. Considering Codanin-1 has been reported to have homology to CNOT1, which acts as a scaffold subunit localising its binding partners, a similar role for Codanin-1 can be envisioned where

Codanin-1 assures correct localisation of C15orf41 and/or Asf1 to assure for correct nucleosome deposition, progress through S-phase and ultimately, erythropoiesis (Swickley et al., 2020; Pop et al., 2010).

Another possible role for C15orf41 and Codanin-1 can be found when looking at localisation experiments. In erythroblasts, C15orf41 and Codanin-1 are enriched in the nucleolus (Olijnik et al., 2021). The nucleolus is a key organelle involved in transcription and processing of ribosomal RNA (rRNA) and ribosome assembly (Lian Zhu et al., 2019). This is an interesting observation as defects in ribosome biogenesis, localised in the nucleolus, have previously been linked to impaired erythropoiesis as outlined by diseases such as Diamond Blackfan Anaemia (DBA) or Shwachman-Diamond Syndrome (SDS) (Lian Zhu et al., 2019; Kampen et al., 2020). DBA and SDS belong to the class of congenital ribosomopathies characterised by mutations in ribosomal proteins or defects in rRNA processing (Kampen et al., 2020). Considering the high rate of transcription of rRNA which would require RNA processing nucleases such as C15orf41, it could be possible C15orf41 and Codanin-1 act as an rRNA processing complex thereby classifying CDA-I a possible new ribosomopathy. Defects in rRNA transcription could lead to accumulation of transcription intermediates such as R-loops which promotes genomic instability leading to disease (Allison and Wang, 2019). Identifying the native substrate that C15orf41 binds and processes would be very informative in this case, a future experiment discussed below. Although DBA and CDA share some common phenotypes such as anaemia and some congenital malformations, one of the hallmarks of ribosomopathies is an increased risk in incidence in cancer for these patients, a feature that is not currently thought to be shared by CDA-I patients (Kampen et al., 2020).

Altogether, a role for C15orf41 in maintaining genome stability is possible. This role is likely modulated by Codanin-1, that acts as a scaffold able to modulate protein levels, activity and likely the localisation of the nuclease. By disrupting this association, such as with the presence of CDA-I patient mutations, C15orf41 is not able to fulfil its cellular role leading to the observed disease phenotype. Complete loss of C15orf41 is believed to be lethal; this stands in agreement with

the observation that patient mutations in C15orf41 severely disrupt but not abolish nuclease activity. The same relationship is observed when looking at protein levels, whereby some CDA-I patient mutations in C15orf41 lower C15orf41 and Codanin-1 levels to still detectable levels. Furthermore, the fact that only 40-60% of patient erythroblasts display the characteristic 'spongy' heterochromatin defect suggests the remainder have functional levels of C15orf41 and Codanin-1, thereby ensuring correct erythropoiesis, albeit at a lower efficiency. This suggests a role for C15orf41 and Codanin-1 in erythropoiesis regulation and could potentially explain the erythroid specific nature of CDA-I.

From a structural biology stand-point, RNA nucleases notoriously present a challenge in the nuclease field. The proteins themselves are usually quite unstable in the absence of their binding partners or specific cofactors and are particularly affected by freeze/thaw cycles, limiting the feasibility of some experiments. Nucleases generally present with a degree of flexibility to allow for various substrate processing slowing down structural studies in particular. With those limitations in mind, we have attempted to perform some biophysical and structural studies of C15orf41-Codanin-1 and whilst not fully conclusive, find that the complex forms a multi-oligomeric state (2 C15orf41 molecules for 1 Codanin-1 molecule) that is able to further oligomerise into a super-complex (with up to 8 C15orf41 molecules for 4 Codanin-1 molecules). Whilst we acknowledge the caveat of using a frozen protein sample which could account for the large, oligomeric species present, it should also be noted that members of the PD-(D/E)-XK family of nucleases have been described to contain peripheral domains and extensions promoting oligomerisation of these nucleases (Kosinski, Feder, and Bujnicki, 2005; Laganeckas, Margelevičius, and Venclovas, 2011).

Using both SAXS and negative-staining EM, we find the complex folds in a relatively flat, stacked shape although the resolution obtained was not high enough to decern specific domains. Further studies will therefore have to be performed to obtain a full structure of the complex and understand its exact mechanism. In particular, it would be interesting to obtain a co-crystal with C15orf41 and RNA

substrate to visualise active-site folding. To this end, a future experiment would be to re-clone a truncated version of C15orf41 that comprises solely of the predicted nuclease domain of the protein. The nuclease domain of C15orf41 has been identified to have high homology to nuclease domains from the PD-(D/E) XK superfamily of nucleases (Babbs et al., 2013). The common structural core is highly structured, consisting of a central, four-stranded β-sheets flanked by two α-helices on both sides, representing a good target for crystallography studies (Knizewski et al., 2007). Generating a truncated C15orf41 protein in a new expression vector with a different tag not only would minimise the disordered regions present; it would also account for possible cleavage of the tag. Although the previous His_6-tag construct was designed to encode a 22-amino acid N-terminal fusion peptide with a TEV protease cleavage site to allow for cleavage of the His_6-tag, this has proven unsuccessful for the full-length C15orf41. It is likely due to obstruction of the TEV protease site or misfolding of the N-terminus of C15orf41, causing burial of the TEV linker and cleavage site. Whilst the presence of the His_6-tag and linker does not impair any downstream biochemical assays, it contributes to structural flexibility which could be clouding some of the biophysical and structural experiments performed. Using a smaller STREP tag adjacent to a TEV protease cleavage site might therefore be a better strategy, allowing not only for cleavage of the tag but also for a smaller, less interfering tag in case cleavage is still not successful. Furthermore, performing comparative studies between the wild-type, nuclease-dead and the patient mutants of C15orf41 would also be informative, shedding light into a possible relationship between the disease phenotype of CDA-I and biochemical activity of C15orf41.

In order to fully elucidate a potential relationship between the newly-identified RNAase activity of C15orf41 and its *in vivo* function and confirm some of the hypothesis outlined above, we believe future experiments should aim to identify the endogenous RNA substrate(s) to C15orf41. Considering the newly-available relevant erythroid HUDEP-2 cell line with endogenously tagged C15orf41, we intend to perform a modified RNA Immuno-Precipitation Sequencing experiment or (RIP-Seq) experiment to extract and sequence this set of RNA molecules unique to

C15orf41. We believe this could offer clues on C15orf41's cellular role and provide insights into a possible disease mechanism of CDA-I. We propose to crosslink endogenously-tagged C15orf41 to its native substrate(s) using UV crosslinking and extract the protein-substrate complexes using immuno-precipitation. This would be followed by extraction of this specific set of RNA molecule(s) and high-throughput sequencing in order to identify a potential class or sequence of RNA molecule(s) that C15orf41 is likely to bind to.

CDA-I is an erythroid specific disease so we believe using an appropriate cell line where C15orf41 is present under its endogenous promoter, such as the 3xFLAG-C15orf41 HUDEP-2 cell line, that serves as a good disease model is essential in this context. We are aware of the limitations of using the HUDEP-2 suspension cell line to perform crosslinking experiments but considering the different cell lines available (tagged C15orf41 with and without patient mutations) and the efficacy of the immuno-precipitation step as demonstrated in Chapter 4, we believe this set-up would yield the most accurate and results. Over-expression experiments have previously yielded very conflicting data, particularly when it comes to localisation of C15orf41 so minimising any confounding factors and running appropriate controls, coupled with biological and technical replicates will be crucial.

The limitations of performing such a complex multi-step experiment are not negligeable. C15orf41 expression levels are relatively low and large cell numbers would be required to obtain enough material to sequence successfully. As already mentioned above, UV-C crosslinking in suspension cells presents a particular challenge and will have to be carefully controlled. Still, we believe that with the appropriate controls, namely the WT HUDEP-2 cells as well as non-crosslinked samples, this experiment could be pivotal in identifying the cellular function of C15orf41. Considering its role as an RNAse, it would be of interest to identify whether C15orf41 binds to a certain class of RNAs or a specific sequence enriched within a specific genomic locus, important findings that could help understand the role of C15orf41.

Altogether, this section provides some valuable new insights into C15orf41 and Codanin-1 biology, biochemistry and structural biology leading to important new research avenues in the CDA-I field. Understanding how C15orf41 is linked to the disease phenotype will represent an important milestone in CDA-I biology and erythropoiesis more generally, and specifically when attempting to develop alternative therapies for patients. Currently, the only treatment is interferon alpha (IFNα) which is not always effective or well-tolerated and presents with significant side effects. The role of IFNα in CDA-I is unknown. Future research should therefore investigate the effect of IFNα on C15orf41-Codanin-1 complex formation as well as the newly-identified RNAse activity of C15orf41. It is possible IFNα has a direct effect on either C15orf41 or Codanin-1, for instance by modulating expression, activity or interactions although a more systemic effect of IFNα on erythroblast biology is also possible. The development of previously-lacking cell and molecular biology tools as well as the insight that C15orf41 is a RNAse will help in providing some valuable insight in future.

Considering C15orf41 was previously completely uncharacterised, we believe this study presents a valuable addition towards better understanding CDA-I and developing more effective and better-tolerated therapies for patients.

8.2 Conclusion

To date, this study presents the first functional characterisation of the recently identified protein encoded by C15orf41, mutated in CDA-I. Using a combination of cellular, biophysical, structural and biochemical methods, C15orf41 is identified as a novel RNA exonuclease that is able to digest a variety of RNA-containing structures. We also confirm native complex formation between C15orf41 and Codanin-1, the two disease proteins of CDA-I, in an erythroid-specific cell line and characterise the nuclease activity of the complex, providing valuable insights into a possible disease mechanism for CDA-I. This has led to the renaming of C15orf41, which is now been designated as CDAN-1 Interacting Nuclease-1 or CDIN-1.

This work also shows the detrimental effects of CDA-I patient mutations in C15orf41 on protein levels, C15orf41-Codanin-1 complex formation and nuclease activity, likely relating to the disease phenotype observed. We hope these findings will lead to interesting novel therapeutic avenues that can be considered if conventional treatments are ineffective. Still, the exact role of C15orf41 will need further investigating to completely elucidate its role in CDA-I and in fundamental cell biology processes.

9

Discussion and Future Perspectives for the nsp14-nsp10 study

9.1 Discussion

This study presents, to the best of our knowledge, the first biochemical analysis of the SARS-CoV-2 nsp14-nsp10 nuclease complex, essential in the viral genome duplication. Whilst previous studies reported the 3′-exonuclease activity of the SARS-CoV-1 nsp14-nsp10 virus, no extensive biochemical analysis was undertaken (Ma et al., 2015; Ferron et al., 2018).

While the SARS-CoV-2 nsp14-nsp10 nuclease complex presents itself as a proof-reading nuclease that is able to digest dsRNA as well RNA substrates containing mismatches, further testing of its nuclease activity paints a much more complex picture. Nsp14-nsp10 emerges as a versatile RNA nuclease, that not only displays a characteristic exonuclease digestion profile (as observed by the RNA laddering displayed on ssRNA substrates) but also holds a previously uncharacterised endonuclease activity. The ability of nsp14-nsp10 to process not only various RNA substrates of mixed sequences but also mixed structures and various common RNA modifications that are associated with viral evasion (6-methyladenine) or induction (inosine) of innate immune pathways, only accentuate the fact that nsp14-nsp10's role extend much beyond that of a classical proofreading nuclease. Nsp14-nsp10 emerges as vital for viral replication and likely represents a simple and efficient replication-repair nuclease complex, analogous to various DNA damage and repair nuclease complexes in the context of mammalian DNA replication. It could be

conceived that nsp14-nsp10 acts on the elongating strand when replication stalls due to mismatch incorporation or the presence of altered ribonucleotides, thereby removing the aberration and allowing resumption of viral replication.

This work also investigates the effect of the SARS-CoV-2 nsp12-7-8 polymerase complex on nsp14-nsp10 nuclease activity. A stimulatory effect of nsp14-nsp10 activity by nsp12-7-8 is reported, further strengthening the hypothesis that replication and repair in the viral genome are closely linked. This enhancement in nsp14-nsp10 nuclease activity was observed across a variety of RNA substrates but was particularly striking on uracil-containing RNA substrates. This could become relevant when considering the physiological life-cycle of the virus. In fact, the plus strand of the coronavirus genomes usually contain poly(A) tails, of up to 130 residues (Hofmann and Brain, 1991; Hackbart, Deng, and Baker, 2020). It is therefore conceivable that poly(U)-poly(A) duplexes are formed through complementary poly(U) deposition on the minus strand and could accumulate during viral genome by the replisome. Poly(U)-pol(A) duplexes have been shown to act as pathogenic-associated molecular patterns or PAMPs, triggering an immune response (Hackbart, Deng, and Baker, 2020). It can therefore be hypothesised that the nuclease activity of nsp14-nsp10 is targeted towards PAMPs in the context of viral replication, thereby limiting poly(U)-poly(A) accumulation and consequent innate immune signalling (Hackbart, Deng, and Baker, 2020; Tay et al., 2020; Deng and Baker, 2018). The need to eliminate uracil-containing structures during viral replication becomes even more apparent when focusing on SARS-CoV-2 nsp15, an endoribonuclease (endoU) which preferentially cleaves uracil residues (Hofmann and Brain, 1991; Deng and Baker, 2018; Hackbart, Deng, and Baker, 2020; Tay et al., 2020). A possible interplay between the two nucleases can be envisioned but still remains to be investigated.

Finally, this study identifies the primase nsp8 as the key factor in nsp14-nsp10 activity stimulation. The exact mechanism of this stimulation still remains unclear. It can be hypothesised that this is due to either direct protein-protein interaction between nsp8 and nsp14 or via a substrate-mediated pathway. Preliminary work

from our lab has identified a weak interaction between nsp8 and nsp14 using nickel pull-down assays, although further experiments will need to be conducted to confirm this (Baddock et al., 2020). Interestingly, several structural studies of nsp8 could potentially shed a light on the stimulation mechanism (Gao et al., 2020; Hillen et al., 2020; Kirchdoerfer and Ward, 2019). Nsp8 structure within the replisome is characterised by large, protruding positively-charged 'poles' that extend out from the main nsp12-7-8 core complex. These poles are essential for efficient RNA substrate engagement with the replisome (Hillen et al., 2020). It can therefore be posited that the nsp8 poles serve as a possible platform for nsp protein interactions. Furthermore, it can be hypothesised that the engagement of nsp8 with the RNA substrates leads to conformational changes within this substrate, rendering it more accessible to nsp14-nsp10 digestion. Further studies, in particular structural studies with nsp8, nsp14-nsp10 and a present RNA substrate, will be necessary to shed light on an exact mechanism of function. Additional biochemical assays with different, truncated versions of nsp8 containing mutations on the pole-like structures would be of interest to determine the role of those structures in nsp14-nsp10 enhancement.

Due to the urgency of identifying new drugs and candidate agents that can treat COVID-19, this biochemical work has laid the ground for inhibitor screening with nsp14-nsp10 as a target. Using *in silico* docking simulations, two potential targets of nsp14-nsp10 were identified: the active site magnesium ion and the several zinc-finger binding motifs of nsp14-nsp10. This led to the discovery of three major inhibitor classes against nsp14-nsp10: hydroxamic acids, hydroxypyrimidinones and zinc-ejecting compounds (Baddock et al., 2020). Hydroxamic acids and hydroxypyrimidinones, such as the HIV integrase inhibitor raltegravir are predicted to chelate the active site ion while selenium-based zinc-ejecting compounds, such as the FDA-approved drug ebselen, might interact with the zinc-finger binding motifs and induce conformational changes detrimental to nsp14-nsp10 activity (Baddock et al., 2020). Ebselen emerged as an interesting target, particularly as it has been identified as the most potent inhibitor of the major SARS-CoV-1 protease (M^{pro}) (Z. Jin et al., 2020). Nsp14, nsp10 and M^{pro} all present zinc-binding motifs, further

strengthening the idea ebselen disrupts these motifs, likely inducing conformational changes affecting enzymatic activities (Ma et al., 2015; Z. Jin et al., 2020).

All together, these findings form the basis for further pharmacological studies, suggesting a poly-pharmacological therapy model for SARS-CoV-2 where both key enzymes nsp14-nsp10 and M^{pro} could be targeted at the same time using ebselen or other multi-target agents (Baddock et al., 2020).

9.2 Conclusion

In all, this study provides a thorough biochemical characterisation of the SARS-CoV-2 nsp14-nsp10 ribonuclease complex. We not only confirm its predicted exonuclease activity on a variety of RNA substrates including substrates containing mismatches but also identify a previously undescribed endonuclease activity of the complex. We also report a previously un-identified stimulatory effect of the SARS-CoV-2 polymerase complex, nsp12-7-8, on the nuclease activity of nsp14-nsp10 that is mediated by the primase component, nsp8.

This work has provided a basis for the design and testing of a variety of inhibitors and drug-like molecules through our collaborators which has allowed for the identification of three reactive chemical compound classes: zinc-finger binding compounds, hydroxamic acids and hydroxypyrimidinones. These results have now been written up as part of a publication, for which co-first authorship has been obtained, that can be found here: `doi.org/10.1101/2020.08.13.248211`.

Appendices

A
Vector Maps

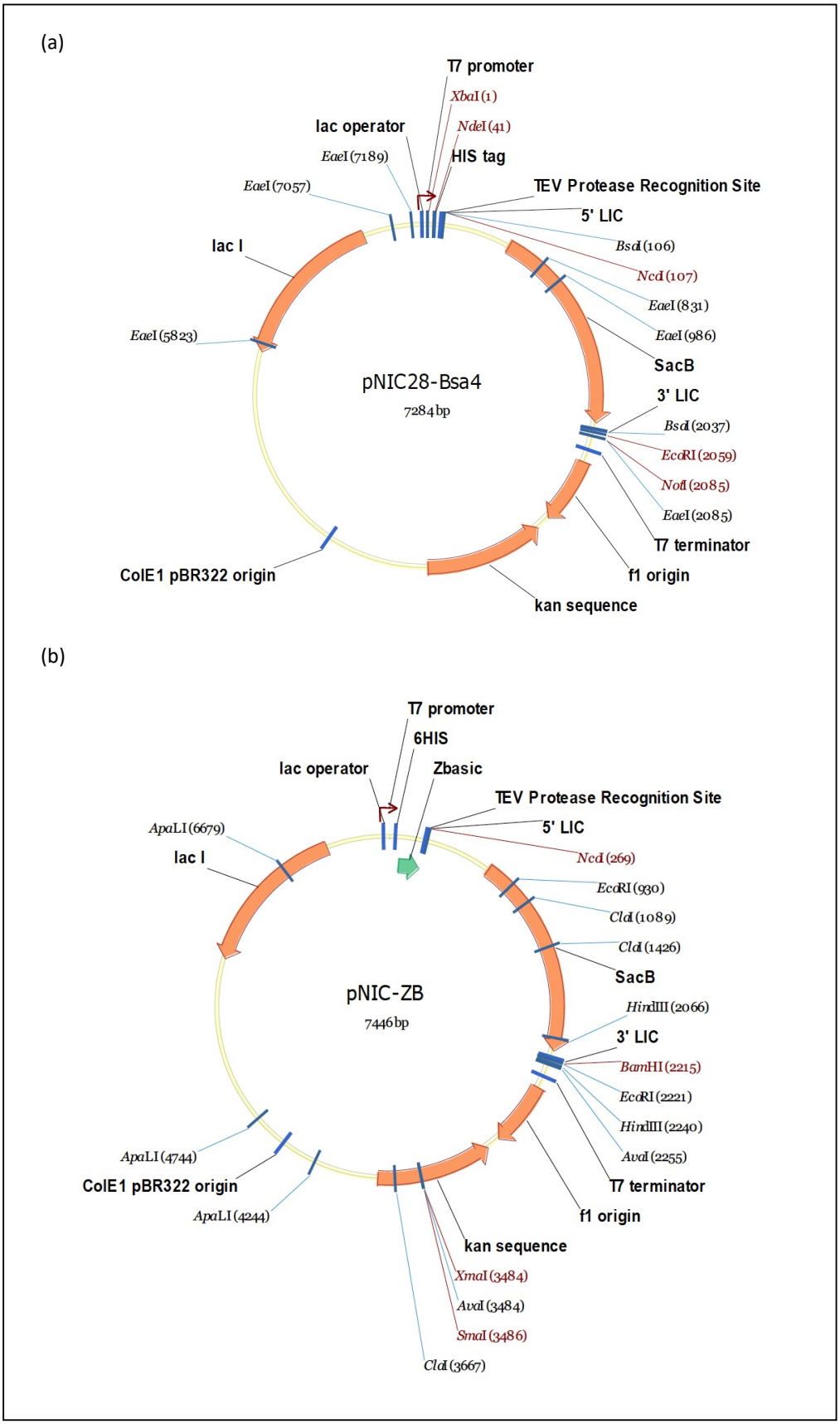


Figure A.1: Maps of purification vectors used for C15orf41 purification (a) and Codanin-1 purification (b).

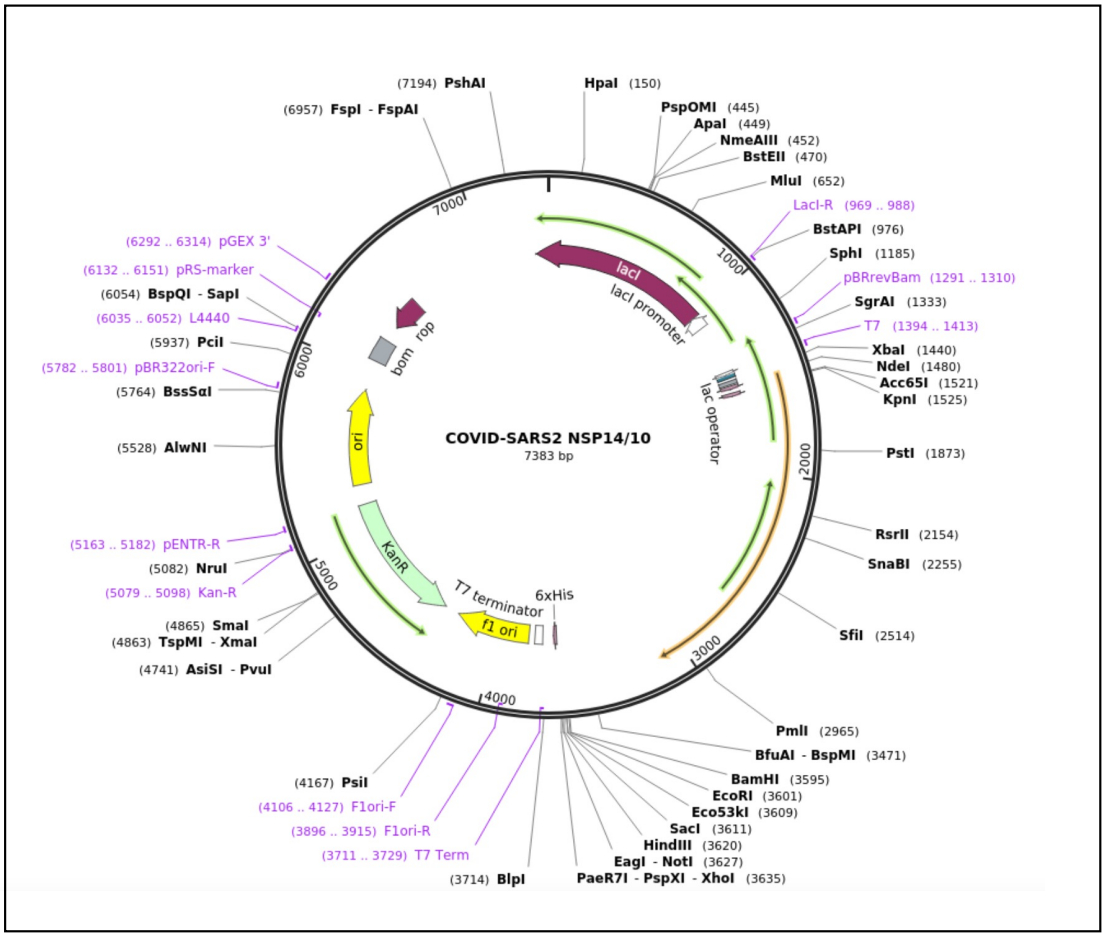


Figure A.2: Map of purification vector used for nsp14-nsp10 purification.

Bibliography

Aessopos, Athanasios, Dimitrios Farmakis, and Dimitris Loukopoulos. 2002. 'Elastic Tissue Abnormalities Resembling Pseudoxanthoma Elasticum in Beta Thalassemia and the Sickling Syndromes'. Blood 99 (1): 30–35. https://doi.org/10.1182/blood.v99.1.30.

Agrigento, V., R. Barone, S. Sclafani, R. Di Maggio, M. Sacco, A. Maggio, and E. D'Alcamo. 2017. 'Response to Alpha-Interferon Treatment of the Congenital Dyserythropoietic Anemia Type I in Two Sicilian Beta Thalassemia Carriers'. Indian Journal of Hematology & Blood Transfusion 33 (4): 621–23. https://doi.org/10.1007/s12288-016-0765-9.

Ahmed, Md Shakil, Md Shahjaman, Enamul Kabir, and Md Kamruzzaman. 2018. 'Structure Modeling to Function Prediction of Uncharacterized Human Protein C15orf41'. Bioinformation 14 (5): 206–12. https://doi.org/10.6026/97320630014206.

Ahmed, Momin R., Aref Chehal, Laila Zahed, Ali Taher, Joud Haidar, Ali Shamseddine, Anne-Marie O'Hea, et al. 2006. 'Linkage and Mutational Analysis of the CDAN1 Gene Reveals Genetic Heterogeneity in Congenital Dyserythropoietic Anemia Type I'. Blood 107 (12): 4968–69. https://doi.org/10.1182/blood-2006-01-0081.

Ahmed, Momin R., Mohamed Zaki, Mohamed A. Sabry, Douglas Higgs, Paresh Vyas, William Wood, and Sunitha N. Wickramasinghe. 2006. 'Evidence of Genetic Heterogeneity in Congenital Dyserythropoietic Anaemia Type I'. British Journal of Haematology 133 (4): 444–45. https://doi.org/10.1111/j.1365-2141.2006.06089.x.

Ahmed, Mostafa H., Mohini S. Ghatge, and Martin K. Safo. 2020. 'Hemoglobin: Structure, Function and Allostery'. In Vertebrate and Invertebrate Respiratory Proteins, Lipoproteins and Other Body Fluid Proteins, edited by Ulrich Hoeger and J. Robin Harris, 94:345–82. Subcellular Biochemistry. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-41769-7_14.

Alliegro, Mary Anne, Jonathan J. Henry, and Mark C. Alliegro. 2010. 'Rediscovery of the Nucleolinus, a Dynamic RNA-Rich Organelle Associated with the

Nucleolus, Spindle, and Centrosomes'. Proceedings of the National Academy of Sciences 107 (31): 13718–23. https://doi.org/10.1073/pnas.1008469107.

Allison, David F, and Gang Greg Wang. 2019. 'R-Loops: Formation, Function, and Relevance to Cell Stress'. Cell Stress 3 (2): 38–46. https://doi.org/10.15698/cst2019.02.175. Alseth, Ingrun, Bjørn Dalhus, and Magnar Bjørås. 2014. 'Inosine in DNA and RNA'. Current Opinion in Genetics & Development 26 (June): 116–23. https://doi.org/10.1016/j.gde.2014.07.008.

Andersen, Kasper R., Nina C. Leksa, and Thomas U. Schwartz. 2013. 'Optimized E. Coli Expression Strain LOBSTR Eliminates Common Contaminants from His-tag Purification'. Proteins: Structure, Function, and Bioinformatics 81 (11): 1857–61. https://doi.org/10.1002/prot.24364.

Antony, Sonia, and Craig A. Bayse. 2013. 'Density Functional Theory Study of the Attack of Ebselen on a Zinc-Finger Model'. Inorganic Chemistry 52 (24): 13803–5. https://doi.org/10.1021/ic401429z.

Arnaud, Lionel, Carole Saison, Virginie Helias, Nicole Lucien, Dominique Steschenko, Marie-Catherine Giarratana, Claude Prehu, et al. 2010. 'A Dominant Mutation in the Gene Encoding the Erythroid Transcription Factor KLF1 Causes a Congenital Dyserythropoietic Anemia'. The American Journal of Human Genetics 87 (5): 721–27. https://doi.org/10.1016/j.ajhg.2010.10.010.

Ask, Katrine, Zuzana Jasencakova, Patrice Menard, Yunpeng Feng, Geneviève Almouzni, and Anja Groth. 2012. 'Codanin-1, Mutated in the Anaemic Disease CDAI, Regulates Asf1 Function in S-Phase Histone Supply'. The EMBO Journal 31 (8): 2013–23. https://doi.org/10.1038/emboj.2012.55.

Babbs, Christian, Nigel A. Roberts, Luis Sanchez-Pulido, Simon J. McGowan, Momin R. Ahmed, Jill M. Brown, Mohamed A. Sabry, et al. 2013. 'Homozygous Mutations in a Predicted Endonuclease Are a Novel Cause of Congenital Dyserythropoietic Anemia Type I'. Haematologica 98 (9): 1383–87. https://doi.org/10.3324/haematol.2013.089490.

Baddock, Hannah T., Sanja Brolih, Yuliana Yosaatmadja, Malitha Ratnaweera, Marcin Bielinski, Lonnie P. Swift, Abimael Cruz-Migoni, et al. 2020. 'Characterisation of the SARS-CoV-2 ExoN (Nsp14 ExoN -Nsp10) Complex: Implications for Its Role in Viral Genome Stability and Inhibitor Identification'. Preprint. Biochemistry. https://doi.org/10.1101/2020.08.13.248211.

Bianchi, Paola, Elisa Fermo, Cristina Vercellati, Carla Boschetti, Wilma Barcellini, Alessandra Iurlo, Anna Paola Marcello, Pier Giorgio Righetti, and Alberto Zanella. 2009. 'Congenital Dyserythropoietic Anemia Type II (CDAII) Is Caused by Mutations in the SEC23B Gene'. Human Mutation 30 (9): 1292–98. https://doi.org/10.1002/humu.21077.

Boulon, Séverine, Belinda J. Westman, Saskia Hutten, François-Michel Boisvert, and Angus I. Lamond. 2010. 'The Nucleolus under Stress'. Molecular Cell 40 (2): 216–27. https://doi.org/10.1016/j.molcel.2010.09.024.

Bouvet, M., I. Imbert, L. Subissi, L. Gluais, B. Canard, and E. Decroly. 2012. 'RNA 3'-End Mismatch Excision by the Severe Acute Respiratory Syndrome Coronavirus Nonstructural Protein Nsp10/Nsp14 Exoribonuclease Complex'. Proceedings of the National Academy of Sciences 109 (24): 9372–77. https://doi.org/10.1073/pnas.1201130109.

Bouvet, Mickaël, Claire Debarnot, Isabelle Imbert, Barbara Selisko, Eric J. Snijder, Bruno Canard, and Etienne Decroly. 2010. 'In Vitro Reconstitution of SARS-Coronavirus MRNA Cap Methylation'. Edited by Michael J. Buchmeier. PLoS Pathogens 6 (4): e1000863. https://doi.org/10.1371/journal.ppat.1000863.

Busch, Anke, Mirko Brüggemann, Stefanie Ebersberger, and Kathi Zarnack. 2019. 'ICLIP Data Analysis: A Complete Pipeline from Sequencing Reads to RBP Binding Sites'. Methods (San Diego, Calif.), November. https://doi.org/10.1016/j.ymeth.2019.11.008.

Cam, Hugh P., Ee Sin Chen, and Shiv I.S. Grewal. 2009. 'Transcriptional Scaffolds for Heterochromatin Assembly'. Cell 136 (4): 610–14. https://doi.org/10.1016/j.cell.2009.02.004.

Campos, Eric I., Jeffrey Fillingham, Guohong Li, Haiyan Zheng, Philipp Voigt, Wei-Hung W. Kuo, Harshika Seepany, et al. 2010. 'The Program for Processing Newly-Synthesized Histones H3.1 and H4'. Nature Structural & Molecular Biology 17 (11): 1343–51. https://doi.org/10.1038/nsmb.1911.

Cevik, Muge, Krutika Kuppalli, Jason Kindrachuk, and Malik Peiris. 2020. 'Virology, Transmission, and Pathogenesis of SARS-CoV-2'. BMJ, October, m3862. https://doi.org/10.1136/bmj.m3862.

Chapman, Timothy M., Claire Wallace, Kevin J. Gillen, Preeti Bakrania, Puneet Khurana, Peter J. Coombs, Simon Fox, et al. 2015. 'N-Hydroxyimides and Hydroxypyrimidinones as Inhibitors of the DNA Repair Complex ERCC1–XPF'. Bioorganic & Medicinal Chemistry Letters 25 (19): 4104–8. https://doi.org/10.1016/j.bmcl.2015.08.024.

Chen, Y., H. Cai, J. Pan, N. Xiang, P. Tien, T. Ahola, and D. Guo. 2009. 'Functional Screen Reveals SARS Coronavirus Nonstructural Protein Nsp14 as a Novel Cap N7 Methyltransferase'. Proceedings of the National Academy of Sciences 106 (9): 3484–89. https://doi.org/10.1073/pnas.0808790106.

Chichinadze, Konstantin, Ann Lazarashvili, and Jaba Tkemaladze. 2013. 'RNA in Centrosomes: Structure and Possible Functions'. Protoplasma 250 (1): 397–405. https://doi.org/10.1007/s00709-012-0422-6.

Chonat, Satheesh, Morgan L. McLemore, Silvia T. Bunting, Shannon Nortman, Kejian Zhang, and Theodosia A. Kalfa. 2018. 'Congenital Dyserythropoietic Anaemia Type I Diagnosed in a Young Adult with a History of Splenectomy in Childhood for Presumed Haemolytic Anaemia'. British Journal of Haematology 182 (1): 10–10. https://doi.org/10.1111/bjh.15217.

D'Alcamo, Elena, V. Agrigento, L. Pitrolo, S. Sclafani, R. Barone, G. Calvaruso, V. Buffa, and A. Maggio. 2016. 'New Codanin-1 Gene Mutations in a Italian Patient with Congenital Dyserythropoietic Anemia Type I and Heterozygous Beta-Thalassemia'. Indian Journal of Hematology & Blood Transfusion: An Official Journal of Indian Society of Hematology and Blood Transfusion 32 (Suppl 1): 278–81. https://doi.org/10.1007/s12288-015-0633-z.

De Benoist, Bruno, World Health Organization, and Centers for Disease Control and Prevention (U.S.). 2008. Worldwide Prevalence of Anaemia 1993-2005 of: WHO Global Database of Anaemia. Geneva: World Health Organization. http://whqlibdoc.who.int/publications/2008/9789241596657_eng.pdf.

De Koning, Leanne, Armelle Corpet, James E. Haber, and Geneviève Almouzni. 2007. 'Histone Chaperones: An Escort Network Regulating Histone Traffic'. Nature Structural & Molecular Biology 14 (11): 997–1007. https://doi.org/10.1038/nsmb1318.

De Matteis, Maria Antonietta, and Alberto Luini. 2011. 'Mendelian Disorders of Membrane Trafficking'. The New England Journal of Medicine 365 (10): 927–38. https://doi.org/10.1056/NEJMra0910494.

Deng, Xufang, and Susan C. Baker. 2018. 'An "Old" Protein with a New Story: Coronavirus Endoribonuclease Is Important for Evading Host Antiviral Defenses'. Virology 517 (April): 157–63. https://doi.org/10.1016/j.virol.2017.12.024.

Denison, Mark R., Rachel L. Graham, Eric F. Donaldson, Lance D. Eckerle, and Ralph S. Baric. 2011. 'Coronaviruses: An RNA Proofreading Machine Regulates Replication Fidelity and Diversity'. RNA Biology 8 (2): 270–79. https://doi.org/10.4161/rna.8.2.15013.

Dgany, Orly, Nili Avidan, Jean Delaunay, Tatyana Krasnov, Lea Shalmon, Hanna Shalev, Tal Eidelitz-Markus, et al. 2002. 'Congenital Dyserythropoietic Anemia Type I Is Caused by Mutations in Codanin-1'. American Journal of Human Genetics 71 (6): 1467–74. https://doi.org/10.1086/344781.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. 'STAR: Ultrafast Universal RNA-Seq Aligner'. Bioinformatics 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Doulatov, Sergei, Faiyaz Notta, Elisa Laurenti, and John E. Dick. 2012. 'Hematopoiesis: A Human Perspective'. Cell Stem Cell 10 (2): 120–36. https://doi.org/10.1016/j.stem.2012.01.006.

Drozdetskiy, Alexey, Christian Cole, James Procter, and Geoffrey J. Barton. 2015. 'JPred4: A Protein Secondary Structure Prediction Server'. Nucleic Acids Research 43 (W1): W389–94. https://doi.org/10.1093/nar/gkv332.

Dzierzak, Elaine, and Sjaak Philipsen. 2013. 'Erythropoiesis: Development and Differentiation'. Cold Spring Harbor Perspectives in Medicine 3 (4). https://doi.org/10.1101/cshperspect.a011601.

Eckerle, Lance D., Xiaotao Lu, Steven M. Sperry, Leena Choi, and Mark R. Denison. 2007. 'High Fidelity of Murine Hepatitis Virus Replication Is Decreased in Nsp14 Exoribonuclease Mutants'. Journal of Virology 81 (22): 12135–44. https://doi.org/10.1128/JVI.01296-07.

Egloff, Marie-Pierre, François Ferron, Valérie Campanacci, Sonia Longhi, Corinne Rancurel, Hélène Dutartre, Eric J. Snijder, Alexander E. Gorbalenya, Christian Cambillau, and Bruno Canard. 2004. 'The Severe Acute Respiratory Syndrome-Coronavirus Replicative Protein Nsp9 Is a Single-Stranded RNA-Binding Subunit Unique in the RNA Virus World'. Proceedings of the National Academy of Sciences of the United States of America 101 (11): 3792. https://doi.org/10.1073/pnas.0307877101.

Elguero, Eric, Lucrèce M. Délicat-Loembet, Virginie Rougeron, Céline Arnathau, Benjamin Roche, Pierre Becquart, Jean-Paul Gonzalez, et al. 2015. 'Malaria Continues to Select for Sickle Cell Trait in Central Africa'. Proceedings of the National Academy of Sciences of the United States of America 112 (22): 7051–54. https://doi.org/10.1073/pnas.1505665112.

El-Sheikh, Ayman A., Hasan Hashem, Carol Holman, and Yatin M. Vyas. 2014. 'Congenital Dyserythropoietic Anemia Type I Presenting as Persistent Pulmonary Hypertension with Pigeon Chest Deformity'. Pediatric Blood & Cancer 61 (8): 1460–62. https://doi.org/10.1002/pbc.24945.

Exell, Jack C, Mark J Thompson, L David Finger, Steven J Shaw, Judit Debreczeni, Thomas A Ward, Claire McWhirter, et al. 2016. 'Cellularly Active N-Hydroxyurea FEN1 Inhibitors Block Substrate Entry to the Active Site'. Nature Chemical Biology 12 (10): 815–21. https://doi.org/10.1038/nchembio.2148.

Ferron, François, Lorenzo Subissi, Ana Theresa Silveira De Morais, Nhung Thi Tuyet Le, Marion Sevajol, Laure Gluais, Etienne Decroly, et al. 2018. 'Structural and Molecular Basis of Mismatch Correction and Ribavirin Excision from Coronavirus RNA'. Proceedings of the National Academy of Sciences 115 (2): E162–71. https://doi.org/10.1073/pnas.1718806115.

Franscina Pinto, Ekta, and Chittaranjan Andrade. 2016. 'Interferon-Related Depression: A Primer on Mechanisms, Treatment, and Prevention of a Common Clinical Problem'. Current Neuropharmacology 14 (7): 743–48. https://doi.org/10.2174/1570159X14666160106155129.

Frye, Michaela, Bryan T. Harada, Mikaela Behm, and Chuan He. 2018. 'RNA Modifications Modulate Gene Expression during Development'. Science 361 (6409): 1346–49. https://doi.org/10.1126/science.aau1646.

Fujino, Hisanori, Sayoko Doisaki, Young-Dong Park, Asahito Hama, Hideki Muramatsu, Seiji Kojima, and Shinichi Sumimoto. 2013. 'Congenital Dyserythropoietic Anemia Type 1 with a Novel Mutation in the CDAN1 Gene Previously Diagnosed as Congenital Hemolytic Anemia'. International Journal of Hematology 97 (5): 650–53. https://doi.org/10.1007/s12185-013-1338-4.

Fung, To Sing, and Ding Xiang Liu. 2019. 'Human Coronavirus: Host-Pathogen Interaction'. Annual Review of Microbiology 73 (1): 529–57. https://doi.org/10.1146/annurev-micro-020518-115759.

Gao, Yan, Liming Yan, Yucen Huang, Fengjiang Liu, Yao Zhao, Lin Cao, Tao Wang, et al. 2020. 'Structure of the RNA-Dependent RNA Polymerase from COVID-19 Virus'. Science 368 (6492): 779–82. https://doi.org/10.1126/science.abb7498.

Gatsiou, Aikaterini, Nikolaos Vlachogiannis, Federica Francesca Lunella, Marco Sachse, and Konstantinos Stellos. 2018. 'Adenosine-to-Inosine RNA Editing in Health and Disease'. Antioxidants & Redox Signaling 29 (9): 846–63. https://doi.org/10.1089/ars.2017.7295.

Goede, Jeroen S., Rudolf Benz, Joerg Fehr, Klaus Schwarz, and Hermann Heimpel. 2006. 'Congenital Dyserythropoietic Anemia Type I with Bone Abnormalities, Mutations of the CDAN I Gene, and Significant Responsiveness to Alpha-Interferon

Therapy'. Annals of Hematology 85 (9): 591–95. https://doi.org/10.1007/s00277-006-0143-z.

Groh, Matthias, and Natalia Gromak. 2014. 'Out of Balance: R-Loops in Human Disease'. Edited by Nancy Maizels. PLoS Genetics 10 (9): e1004630. https://doi.org/10.1371/journal.pgen.1004630.

Gullerova, Monika, and Nick J. Proudfoot. 2008. 'Cohesin Complex Promotes Transcriptional Termination between Convergent Genes in S. Pombe'. Cell 132 (6): 983–95. https://doi.org/10.1016/j.cell.2008.02.040.

Haberman, Nejc, Ina Huppertz, Jan Attig, Julian König, Zhen Wang, Christian Hauer, Matthias W. Hentze, et al. 2017. 'Insights into the Design and Interpretation of ICLIP Experiments'. Genome Biology 18 (1): 7. https://doi.org/10.1186/s13059-016-1130-x.

Hackbart, Matthew, Xufang Deng, and Susan C. Baker. 2020. 'Coronavirus Endoribonuclease Targets Viral Polyuridine Sequences to Evade Activating Host Sensors'. Proceedings of the National Academy of Sciences 117 (14): 8094–8103. https://doi.org/10.1073/pnas.1921485117.

Hallick, Richard B., Barry K. Chelm, Patrick W. Gray, and Emil M. Orozco. 1977. 'Use of Aurintricarboxylic Acid as an Inhibitor of Nucleases during Nucleic Acid Isolation'. Nucleic Acids Research 4 (9): 3055–64. https://doi.org/10.1093/nar/4.9.3055.

Hamada, Motoharu, Sayoko Doisaki, Yusuke Okuno, Hideki Muramatsu, Asahito Hama, Nozomu Kawashima, Atsushi Narita, et al. 2018. 'Whole-Exome Analysis to Detect Congenital Hemolytic Anemia Mimicking Congenital Dyserythropoietic Anemia'. International Journal of Hematology 108 (3): 306–11. https://doi.org/10.1007/s12185-018-2482-7.

Hartenian, Ella, Divya Nandakumar, Azra Lari, Michael Ly, Jessica M. Tucker, and Britt A. Glaunsinger. 2020. 'The Molecular Virology of Coronaviruses'. Journal of Biological Chemistry 295 (37): 12910–34. https://doi.org/10.1074/jbc.REV120.013930.

Hattangadi, Shilpa M., Piu Wong, Lingbo Zhang, Johan Flygare, and Harvey F. Lodish. 2011. 'From Stem Cell to Red Cell: Regulation of Erythropoiesis at

Multiple Levels by Multiple Proteins, RNAs, and Chromatin Modifications'. Blood 118 (24): 6258–68. https://doi.org/10.1182/blood-2011-07-356006.

Heimpel, Hermann. 2004. 'Congenital Dyserythropoietic Anemias: Epidemiology, Clinical Significance, and Progress in Understanding Their Pathogenesis'. Annals of Hematology 83 (10): 613–21. https://doi.org/10.1007/s00277-004-0892-5.

Heimpel, Hermann, Klaus Schwarz, Monika Ebnöther, Jeroen S. Goede, Detlev Heydrich, Torsten Kamp, Lothar Plaumann, et al. 2006. 'Congenital Dyserythropoietic Anemia Type I (CDA I): Molecular Genetics, Clinical Appearance, and Prognosis Based on Long-Term Observation'. Blood 107 (1): 334–40. https://doi.org/10.1182/blood-2005-01-0421.

Hentges, Lance D., Martin J. Sergeant, Damien J. Downes, Jim R. Hughes, and Stephen Taylor. 2021. 'LanceOtron: A Deep Learning Peak Caller for ATAC-Seq, ChIP-Seq, and DNase-Seq'. Preprint. Bioinformatics. https://doi.org/10.1101/2021.01.25.428108.

Hildebrandt, Andrea, Mirko Brüggemann, Cornelia Rücklé, Susan Boerner, Jan B. Heidelberger, Anke Busch, Heike Hänel, et al. 2019. 'The RNA-Binding Ubiquitin Ligase MKRN1 Functions in Ribosome-Associated Quality Control of Poly(A) Translation'. Genome Biology 20 (1): 216. https://doi.org/10.1186/s13059-019-1814-0.

Hillen, Hauke S., Goran Kokic, Lucas Farnung, Christian Dienemann, Dimitry Tegunov, and Patrick Cramer. 2020. 'Structure of Replicating SARS-CoV-2 Polymerase'. Nature 584 (7819): 154–56. https://doi.org/10.1038/s41586-020-2368-8.

Hiller, Bjoern, Martin Achleitner, Silke Glage, Ronald Naumann, Rayk Behrendt, and Axel Roers. 2012. 'Mammalian RNase H2 Removes Ribonucleotides from DNA to Maintain Genome Integrity'. Journal of Experimental Medicine 209 (8): 1419–26. https://doi.org/10.1084/jem.20120876.

Hofmann, M A, and D A Brain. 1991. 'The 5' End of Coronavirus Minus-Strand RNAs Contains a Short Poly(U) Tract.' Journal of Virology 65 (11): 6331–33. https://doi.org/10.1128/JVI.65.11.6331-6333.1991.

Ilik, Ibrahim Avsar, Tugce Aktas, Daniel Maticzka, Rolf Backofen, and Asifa Akhtar. 2019. 'FLASH: Ultra-Fast Protocol to Identify RNA–Protein Interactions

in Cells'. Nucleic Acids Research, December, gkz1141. https://doi.org/10.1093/nar/gkz1141.

Imbert, Isabelle, Eric J. Snijder, Maria Dimitrova, Jean-Claude Guillemot, Patrick Lécine, and Bruno Canard. 2008. 'The SARS-Coronavirus PLnc Domain of Nsp3 as a Replication/Transcription Scaffolding Protein'. Virus Research 133 (2): 136–48. https://doi.org/10.1016/j.virusres.2007.11.017.

Iolascon, Achille, and Jean Delaunay. 2009. 'Close to Unraveling the Secrets of Congenital Dyserythropoietic Anemia Types I and II'. Haematologica 94 (5): 599–602. https://doi.org/10.3324/haematol.2009.005785.

Iolascon, Achille, Maria Rosaria Esposito, and Roberta Russo. 2012. 'Clinical Aspects and Pathogenesis of Congenital Dyserythropoietic Anemias: From Morphology to Molecular Approach'. Haematologica 97 (12): 1786–94. https://doi.org/10.3324/haematol.2012.072207.

Iolascon, Achille, Hermann Heimpel, Anders Wahlin, and Hannah Tamary. 2013. 'Congenital Dyserythropoietic Anemias: Molecular Insights and Diagnostic Approach'. Blood 122 (13): 2162–66. https://doi.org/10.1182/blood-2013-05-468223.

Iolascon, Achille, Roberta Russo, and Jean Delaunay. 2011. 'Congenital Dyserythropoietic Anemias'. Current Opinion in Hematology 18 (3): 146–51. https://doi.org/10.1097/MOH.0b013e32834521b0.

Iolascon, Achille, Roberta Russo, Maria Rosaria Esposito, Roberta Asci, Carmelo Piscopo, Silverio Perrotta, Madeleine Fénéant-Thibault, Loïc Garçon, and Jean Delaunay. 2010. 'Molecular Analysis of 42 Patients with Congenital Dyserythropoietic Anemia Type II: New Mutations in the SEC23B Gene and a Search for a Genotype-Phenotype Relationship'. Haematologica 95 (5): 708–15. https://doi.org/10.3324/haematol.2009.014985.

Jaramillo, Catalina, Anna K Ermarth, Angelica R Putnam, and Mark Deneau. 2019. 'Neonatal Cholestasis and Hepatosplenomegaly Caused by Congenital Dyserythropoietic Anemia Type 1: A Case Report'. World Journal of Hepatology 11 (5): 477–82. https://doi.org/10.4254/wjh.v11.i5.477.

Jasencakova, Zuzana, Annette N. D. Scharf, Katrine Ask, Armelle Corpet, Axel Imhof, Geneviève Almouzni, and Anja Groth. 2010. 'Replication Stress Interferes with Histone Recycling and Predeposition Marking of New Histones'. Molecular Cell 37 (5): 736–43. https://doi.org/10.1016/j.molcel.2010.01.033.

Jin, Xu, Yu Chen, Ying Sun, Cong Zeng, Yi Wang, Jiali Tao, Andong Wu, et al. 2013. 'Characterization of the Guanine-N7 Methyltransferase Activity of Coronavirus Nsp14 on Nucleotide GTP'. Virus Research 176 (1–2): 45–52. https://doi.org/10.1016/j.virusres.2013.05.001.

Jin, Zhenming, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, et al. 2020. 'Structure of Mpro from SARS-CoV-2 and Discovery of Its Inhibitors'. Nature 582 (7811): 289–93.
https://doi.org/10.1038/s41586-020-2223-y.

Kamiya, Takahiro, and Atsushi Manabe. 2010. 'Congenital Dyserythropoietic Anemia'. International Journal of Hematology 92 (3): 432–38.
https://doi.org/10.1007/s12185-010-0667-9.

Kawa, Miłosz P., and Bogusław Machaliński. 2014. 'Hematopoiesis Dysfunction Associated with Abnormal Thyroid Hormones Production'. Thyroid Disorders - Focus on Hyperthyroidism, May. https://doi.org/10.5772/58531.

Kawabata, Hiroshi, Sayoko Doisaki, Akio Okamoto, Tatsuki Uchiyama, Soichiro Sakamoto, Asahito Hama, Kiminori Hosoda, et al. 2012. 'A Case of Congenital Dyserythropoietic Anemia Type 1 in a Japanese Adult with a CDAN1 Gene Mutation and an Inappropriately Low Serum Hepcidin-25 Level'. Internal Medicine (Tokyo, Japan) 51 (8): 917–20. https://doi.org/10.2169/internalmedicine.51.6978.

Ke, Zunlong, Joaquin Oton, Kun Qu, Mirko Cortese, Vojtech Zila, Lesley McKeane, Takanori Nakane, et al. 2020. 'Structures, Conformations and Distributions of SARS-CoV-2 Spike Protein Trimers on Intact Virions'. Preprint. Microbiology. https://doi.org/10.1101/2020.06.27.174979.

Kelly, Timothy J., Hiroshi I. Suzuki, Jesse R. Zamudio, Megumu Suzuki, and Phillip A. Sharp. 2019. 'Sequestration of MicroRNA-Mediated Target Repression

by the Ago2-Associated RNA-Binding Protein FAM120A'. RNA (New York, N.Y.) 25 (10): 1291–97. https://doi.org/10.1261/rna.071621.119.

Kim, Youngchang, Robert Jedrzejczak, Natalia I. Maltseva, Mateusz Wilamowski, Michael Endres, Adam Godzik, Karolina Michalska, and Andrzej Joachimiak. 2020. 'Crystal Structure of Nsp15 Endoribonuclease NENDOU from SARS-COV -2'. Protein Science 29 (7): 1596–1605. https://doi.org/10.1002/pro.3873.

Kirchdoerfer, Robert N., and Andrew B. Ward. 2019. 'Structure of the SARS-CoV Nsp12 Polymerase Bound to Nsp7 and Nsp8 Co-Factors'. Nature Communications 10 (1): 2342. https://doi.org/10.1038/s41467-019-10280-3.

Klein, Steffen, Mirko Cortese, Sophie L. Winter, Moritz Wachsmuth-Melm, Christopher J. Neufeldt, Berati Cerikan, Megan L. Stanifer, Steeve Boulant, Ralf Bartenschlager, and Petr Chlanda. 2020. 'SARS-CoV-2 Structure and Replication Characterized by in Situ Cryo-Electron Tomography'. Nature Communications 11 (1): 5885. https://doi.org/10.1038/s41467-020-19619-7.

Klockenbusch, Cordula, and Juergen Kast. 2010. 'Optimization of Formaldehyde Cross-Linking for Protein Interaction Analysis of Non-Tagged Integrin 1'. Edited by Chun Tang. Journal of Biomedicine and Biotechnology 2010 (June): 927585. https://doi.org/10.1155/2010/927585.

Knizewski, Lukasz, Lisa N Kinch, Nick V Grishin, Leszek Rychlewski, and Krzysztof Ginalski. 2007. 'Realm of PD-(D/E)XK Nuclease Superfamily Revisited: Detection of Novel Families with Modified Transitive Meta Profile Searches'. BMC Structural Biology 7 (June): 40. https://doi.org/10.1186/1472-6807-7-40.

Koch, Linda. 2015. 'New CLIP Pipeline Improves Interactome Discovery'. Nature Reviews Genetics 16 (1): 2–2. https://doi.org/10.1038/nrg3877.

König, Julian, Kathi Zarnack, Gregor Rot, Tomaž Curk, Melis Kayikci, Blaž Zupan, Daniel J. Turner, Nicholas M. Luscombe, and Jernej Ule. 2010. 'ICLIP Reveals the Function of HnRNP Particles in Splicing at Individual Nucleotide Resolution'. Nature Structural & Molecular Biology 17 (7): 909–15. https://doi.org/10.1038/nsmb.1838.

Kosinski, Jan, Marcin Feder, and Janusz M. Bujnicki. 2005. 'The PD-(D/E)XK Superfamily Revisited: Identification of New Members among Proteins Involved in DNA Metabolism and Functional Predictions for Domains of (Hitherto) Unknown Function'. BMC Bioinformatics 6 (July): 172. https://doi.org/10.1186/1471-2105-6-172.

Laganeckas, Mindaugas, Mindaugas Margelevičius, and Česlovas Venclovas. 2011. 'Identification of New Homologs of PD-(D/E)XK Nucleases by Support Vector Machines Trained on Data Derived from Profile–Profile Alignments'. Nucleic Acids Research 39 (4): 1187–96. https://doi.org/10.1093/nar/gkq958.

Lavabre-Bertrand, Thierry, Carole Exbrayat, Jeannie Liautard, Jean-Philippe Gaillard, Pierre-Paul Baskevitch, Nicolas Poujol, Christophe Duperray, Pascal Bourqoard, and Jean Brochier. 1995. 'Detection of Membrane and Soluble Interleukin-6 Receptor in Lymphoid Malignancies'. British Journal of Haematology 91 (4): 871–77. https://doi.org/10.1111/j.1365-2141.1995.tb05403.x.

Li, Heng, Shang-Ming Liu, Xiao-Hua Yu, Shi-Lin Tang, and Chao-Ke Tang. 2020. 'Coronavirus Disease 2019 (COVID-19): Current Status and Future Perspectives'. International Journal of Antimicrobial Agents 55 (5): 105951. https://doi.org/10.1016/j.ijantimicag.2020.105951.

Liljeholm, Maria, Andrew F. Irvine, Ann-Louise Vikberg, Anna Norberg, Stacy Month, Herbert Sandström, Anders Wahlin, Masanori Mishima, and Irina Golovleva. 2013. 'Congenital Dyserythropoietic Anemia Type III (CDA III) Is Caused by a Mutation in Kinesin Family Member, KIF23'. Blood 121 (23): 4791–99. https://doi.org/10.1182/blood-2012-10-461392.

Lin, Pei-Chin, Chao-Neng Cheng, Hsi-Yuan Huang, Yu-Hsin Tseng, Ya-Sian Chang, Chien-Yu Lin, and Jan-Gowth Chang. n.d. 'Congenital Dyserythropoiesis Anemia Type Ia with a Novel CDAN1 Mutation Diagnosed by Whole Exome Sequencing'. Molecular Genetics & Genomic Medicine n/a (n/a): e1220. https://doi.org/10.1002/mgg3.1220.

Liu, Wallace H., Sarah C. Roemer, Yeyun Zhou, Zih-Jie Shen, Briana K. Dennehey, Jeremy L. Balsbaugh, Jennifer C. Liddle, et al. 2016. 'The Cac1

Subunit of Histone Chaperone CAF-1 Organizes CAF-1-H3/H4 Architecture and Tetramerizes Histones'. ELife 5 (September). https://doi.org/10.7554/eLife.18023.

Liu, Xiaorong, Puxian Fang, Liurong Fang, Yingying Hong, Xinyu Zhu, Dang Wang, Guiqing Peng, and Shaobo Xiao. 2019. 'Porcine Deltacoronavirus Nsp15 Antagonizes Interferon-β Production Independently of Its Endoribonuclease Activity'. Molecular Immunology 114 (October): 100–107. https://doi.org/10.1016/j.molimm.2019.07.003.

Liu, Ying, Christian F Nielsen, Qi Yao, and Ian D Hickson. 2014. 'The Origins and Processing of Ultra Fine Anaphase DNA Bridges'. Current Opinion in Genetics & Development 26 (June): 1–5. https://doi.org/10.1016/j.gde.2014.03.003.

Lorenz, Ronny, Stephan H Bernhart, Christian Höner zu Siederdissen, Hakim Tafer, Christoph Flamm, Peter F Stadler, and Ivo L Hofacker. 2011. 'ViennaRNA Package 2.0'. Algorithms for Molecular Biology 6 (1): 26. https://doi.org/10.1186/1748-7188-6-26.

Lowenthal, R. M., K. A. Marsden, C. L. Dewar, and G. R. Thompson. 1980. 'Congenital Dyserythropoietic Anaemia (CDA) with Severe Gout, Rare Kell Phenotype and Erythrocyte, Granulocyte and Platelet Membrane Reduplication: A New Variant of CDA Type II'. British Journal of Haematology 44 (2): 211–20. https://doi.org/10.1111/j.1365-2141.1980.tb01203.x.

Ludwig, Heinz, and Kathrin Strasser. 2001. 'Symptomatology of Anemia'. Seminars in Oncology, Evolving Issues in Oncology: What is the @'Optimal@' Hemoglobin Level? Part one, 28 (April): 7–14. https://doi.org/10.1016/S0093-7754(01)90206-4.

Lv, Xueyan, Shanwu Dong, Fengli Lan, Bingyu Zhang, Hongbo Chen, and Runming Jin. 2020. 'Congenital Dyserythropoietic Anemia Type I Mimicking Myelodysplasia Syndrome with a Novel CDAN1 Mutation'. Annals of Hematology 99 (1): 197–99. https://doi.org/10.1007/s00277-019-03848-y.

Ma, Yuanyuan, Lijie Wu, Neil Shaw, Yan Gao, Jin Wang, Yuna Sun, Zhiyong Lou, Liming Yan, Rongguang Zhang, and Zihe Rao. 2015. 'Structural Basis and Functional Analysis of the SARS Coronavirus Nsp14–Nsp10 Complex'. Proceedings

of the National Academy of Sciences 112 (30): 9436–41.
https://doi.org/10.1073/pnas.1508686112.

Mandal, Amit Kumar, Amrita Mitra, and Rajdeep Das. 2020. 'Sickle Cell Hemoglobin'. Sub-Cellular Biochemistry 94: 297–322. https://doi.org/10.1007/978-3-030-41769-7_12.

Manners, Oliver, Belinda Baquero-Perez, and Adrian Whitehouse. 2019. 'M6A: Widespread Regulatory Control in Virus Replication'. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1862 (3): 370–81.
https://doi.org/10.1016/j.bbagrm.2018.10.015.

Marengo-Rowe, Alain J. 2006. 'Structure-Function Relations of Human Hemoglobins'. Proceedings (Baylor University. Medical Center) 19 (3): 239–45. https://doi.org/10.1080/08998280.2006.11928171.

Maryati, Maryati, Blessing Airhihen, and G. Sebastiaan Winkler. 2015. 'The Enzyme Activities of Caf1 and Ccr4 Are Both Required for Deadenylation by the Human Ccr4-Not Nuclease Module'. The Biochemical Journal 469 (1): 169–76. https://doi.org/10.1042/BJ20150304.

Maticzka, Daniel, Ibrahim Avsar Ilik, Tugce Aktas, Rolf Backofen, and Asifa Akhtar. 2018. 'UvCLAP Is a Fast and Non-Radioactive Method to Identify in Vivo Targets of RNA-Binding Proteins'. Nature Communications 9 (1): 1142. https://doi.org/10.1038/s41467-018-03575-4.

Mei, Yang, Yijie Liu, and Peng Ji. 2021. 'Understanding Terminal Erythropoiesis: An Update on Chromatin Condensation, Enucleation, and Reticulocyte Maturation'. Blood Reviews 46 (March): 100740.
https://doi.org/10.1016/j.blre.2020.100740.

Mello, Jill A., Herman H. W. Silljé, Daniele M. J. Roche, Doris B. Kirschner, Erich A. Nigg, and Geneviève Almouzni. 2002. 'Human Asf1 and CAF-1 Interact and Synergize in a Repair-Coupled Nucleosome Assembly Pathway'. EMBO Reports 3 (4): 329–34. https://doi.org/10.1093/embo-reports/kvf068.

Meznarich, Jessica A., Lauren Draper, Robert D. Christensen, Hassan M. Yaish, Nick D. Luem, Theodore J. Pysher, Grace Jung, Elizabeta Nemeth, Tomas Ganz,

and Diane M. Ward. 2018. 'Fetal Presentation of Congenital Dyserythropoietic Anemia Type 1 with Novel Compound Heterozygous CDAN1 Mutations'. Blood Cells, Molecules & Diseases 71: 63–66. https://doi.org/10.1016/j.bcmd.2018.03.002.

Miller, Jeffery L. 2013. 'Iron Deficiency Anemia: A Common and Curable Disease'. Cold Spring Harbor Perspectives in Medicine 3 (7). https://doi.org/10.1101/cshperspect.a011866.

Minskaia, E., T. Hertzig, A. E. Gorbalenya, V. Campanacci, C. Cambillau, B. Canard, and J. Ziebuhr. 2006. 'Discovery of an RNA Virus 3'->5' Exoribonuclease That Is Critically Involved in Coronavirus RNA Synthesis'. Proceedings of the National Academy of Sciences 103 (13): 5108–13. https://doi.org/10.1073/pnas.0508200103.

Moras, Martina, Sophie D. Lefevre, and Mariano A. Ostuni. 2017. 'From Erythroblasts to Mature Red Blood Cells: Organelle Clearance in Mammals'. Frontiers in Physiology 8. https://doi.org/10.3389/fphys.2017.01076.

Moreno-Carralero, María-Isabel, Saul Horta-Herrera, Marta Morado-Arias, María-Pilar Ricard-Andrés, Angelina Lemes-Castellano, Mariola Abio-Calvete, María-Teresa Cedena-Romero, et al. 2018. 'Clinical and Genetic Features of Congenital Dyserythropoietic Anemia (CDA)'. European Journal of Haematology 101 (3): 368–78. https://doi.org/10.1111/ejh.13112.

Muncie, Herbert L., and James Campbell. 2009. 'Alpha and Beta Thalassemia'. American Family Physician 80 (4): 339–44. Nadel, Julie, Rodoniki Athanasiadou, Christophe Lemetre, N. Ari Wijetunga, Pilib Ó Broin, Hanae Sato, Zhengdong Zhang, et al. 2015. 'RNA:DNA Hybrids in the Human Genome Have Distinctive Nucleotide Characteristics, Chromatin Composition, and Transcriptional Relationships'. Epigenetics & Chromatin 8 (1): 46. https://doi.org/10.1186/s13072-015-0040-6.

Nakahama, Taisuke, and Yukio Kawahara. 2020. 'Adenosine-to-Inosine RNA Editing in the Immune System: Friend or Foe?' Cellular and Molecular Life Sciences 77 (15): 2931–48. https://doi.org/10.1007/s00018-020-03466-2.

Netzband, Rachel, and Cara T. Pager. 2020. 'Epitranscriptomic Marks: Emerging Modulators of RNA Virus Gene Expression'. WIREs RNA 11 (3). https://doi.org/10.1002/wrna.1576.

Niewold, Timothy B. 2011. 'Interferon Alpha as a Primary Pathogenic Factor in Human Lupus'. Journal of Interferon & Cytokine Research 31 (12): 887–92. https://doi.org/10.1089/jir.2011.0071.

Niranjanakumari, Somashe, Erika Lasda, Robert Brazas, and Mariano A. Garcia-Blanco. 2002. 'Reversible Cross-Linking Combined with Immunoprecipitation to Study RNA–Protein Interactions in Vivo'. Methods 26 (2): 182–90. https://doi.org/10.1016/S1046-2023(02)00021-X.

Noguchi, Noriko. 2016. 'Ebselen, a Useful Tool for Understanding Cellular Redox Biology and a Promising Drug Candidate for Use in Human Diseases'. Archives of Biochemistry and Biophysics 595 (April): 109–12.
https://doi.org/10.1016/j.abb.2015.10.024.

Noy-Lotan, Sharon, Orly Dgany, Roxane Lahmi, Nathaly Marcoux, Tanya Krasnov, Nissan Yissachar, Doron Ginsberg, et al. 2009. 'Codanin-1, the Protein Encoded by the Gene Mutated in Congenital Dyserythropoietic Anemia Type I (CDAN1), Is Cell Cycle-Regulated'. Haematologica 94 (5): 629–37.
https://doi.org/10.3324/haematol.2008.003327.

Ogando, Natacha S., Francois Ferron, Etienne Decroly, Bruno Canard, Clara C. Posthuma, and Eric J. Snijder. 2019. 'The Curious Case of the Nidovirus Exoribonuclease: Its Role in RNA Synthesis and Replication Fidelity'. Frontiers in Microbiology 10 (August): 1813. https://doi.org/10.3389/fmicb.2019.01813.

Olijnik, Aude-Anais, Noémi B. A. Roy, Caroline Scott, Joseph A. Marsh, Jill Brown, Karin Lauschke, Katrine Ask, et al. 2021. 'Genetic and Functional Insights into CDA-I Prevalence and Pathogenesis'. Journal of Medical Genetics 58 (3): 185–95. https://doi.org/10.1136/jmedgenet-2020-106880.

Palmblad, Jan, Birgitta Sander, Barbara Bain, Monika Klimkowska, and Erik Björck. 2018. 'Congenital Dyserythropoietic Anemia Type 1: A Case with Novel Compound Heterozygous Mutations in the C15orf41 Gene'. American Journal of Hematology 93 (8): E213–15. https://doi.org/10.1002/ajh.25157.

Paton, David. 1959. ‘Angioid Streaks and Sickle Cell Anemia: A Report of Two Cases’. A.M.A. Archives of Ophthalmology 62 (5): 852–58. https://doi.org/10.1001/archopht.1959.04220050112018.

Petit, Alain-Pierre, Lara Wohlbold, Praveen Bawankar, Eric Huntzinger, Steffen Schmidt, Elisa Izaurralde, and Oliver Weichenrieder. 2012. ‘The Structural Basis for the Interaction between the CAF1 Nuclease and the NOT1 Scaffold of the Human CCR4–NOT Deadenylase Complex’. Nucleic Acids Research 40 (21): 11058–72. https://doi.org/10.1093/nar/gks883.

Pielage, Jan, Tobias Stork, Ingrid Bunse, and Christian Klämbt. 2003. ‘The Drosophila Cell Survival Gene Discs Lost Encodes a Cytoplasmic Codanin-1-like Protein, Not a Homolog of Tight Junction PDZ Protein Patj’. Developmental Cell 5 (6): 841–51. https://doi.org/10.1016/s1534-5807(03)00358-7.

Piperno, Alberto, Sara Pelucchi, and Raffaella Mariani. 2020. ‘Inherited Iron Overload Disorders’. Translational Gastroenterology and Hepatology 5: 25. https://doi.org/10.21037/tgh.2019.11.15.

Plevin, Michael J., Morgon M. Mills, and Mitsuhiko Ikura. 2005. ‘The LxxLL Motif: A Multifunctional Binding Sequence in Transcriptional Regulation’. Trends in Biochemical Sciences 30 (2): 66–69. https://doi.org/10.1016/j.tibs.2004.12.001.

Pop, Ramona, Jeffrey R. Shearstone, Qichang Shen, Ying Liu, Kelly Hallstrom, Miroslav Koulnis, Joost Gribnau, and Merav Socolovsky. 2010. ‘A Key Commitment Step in Erythropoiesis Is Synchronized with the Cell Cycle Clock through Mutual Inhibition between PU.1 and S-Phase Progression’. PLoS Biology 8 (9). https://doi.org/10.1371/journal.pbio.1000484.

Potužník, Jiří František, and Hana Cahová. 2020. ‘It’s the Little Things (in Viral RNA)’. Edited by Vinayaka R. Prasad and Danielle A. Garsin. MBio 11 (5): e02131-20, /mbio/11/5/mBio.02131-20.atom. https://doi.org/10.1128/mBio.02131-20.

Raunser, and Eugene Valkov. 2019. ‘Reconstitution of Recombinant Human CCR4-NOT Reveals Molecular Insights into Regulated Deadenylation’. Nature Communications 10 (1): 1–14. https://doi.org/10.1038/s41467-019-11094-z.

Rathe, Mathias, Michael Boe Møller, Pernille Wied Greisen, and Niels Fisker. 2018. 'Successful Management of Transfusion-Dependent Congenital Dyserythropoietic Anemia Type 1b with Interferon Alfa-2a'. Pediatric Blood & Cancer 65 (3): e26866. https://doi.org/10.1002/pbc.26866.

Reijns, Martin A.M., Björn Rabe, Rachel E. Rigby, Pleasantine Mill, Katy R. Astell, Laura A. Lettice, Shelagh Boyle, et al. 2012. 'Enzymatic Removal of Ribonucleotides from DNA Is Essential for Mammalian Genome Integrity and Development'. Cell 149 (5): 1008–22. https://doi.org/10.1016/j.cell.2012.04.011.

Renella, Raffaele, Nigel A. Roberts, Jill M. Brown, Marco De Gobbi, Louise E. Bird, Tasneem Hassanali, Jacqueline A. Sharpe, et al. 2011. 'Codanin-1 Mutations in Congenital Dyserythropoietic Anemia Type 1 Affect HP1alpha Localization in Erythroblasts'. Blood 117 (25): 6928–38. https://doi.org/10.1182/blood-2010-09-308478.

Resnitzky, Peretz, Dina Shaft, Hanna Shalev, Joseph Kapelushnik, Orly Dgany, Tanya Krasnov, and Hannah Tamary. 2017. 'Morphological Features of Congenital Dyserythropoietic Anemia Type I: The Role of Electron Microscopy in Diagnosis'. European Journal of Haematology 99 (4): 366–71.
https://doi.org/10.1111/ejh.12931.

Richard, Patricia, and James L. Manley. 2017. 'R Loops and Links to Human Disease'. Journal of Molecular Biology 429 (21): 3168–80.
https://doi.org/10.1016/j.jmb.2016.08.031.

Romano, Maria, Alessia Ruggiero, Flavia Squeglia, Giovanni Maga, and Rita Berisio. 2020. 'A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping'. Cells 9 (5): 1267.
https://doi.org/10.3390/cells9051267.

Roundtree, Ian A., Molly E. Evans, Tao Pan, and Chuan He. 2017. 'Dynamic RNA Modifications in Gene Expression Regulation'. Cell 169 (7): 1187–1200. https://doi.org/10.1016/j.cell.2017.05.045.

Roy, Noémi B. A., and Christian Babbs. 2019. 'The Pathogenesis, Diagnosis and Management of Congenital Dyserythropoietic Anaemia Type I'. British Journal of Haematology 185 (3): 436–49. https://doi.org/10.1111/bjh.15817.

Russo, Roberta, Antonella Gambale, Maria Rosaria Esposito, Maria Luisa Serra, Annaelena Troiano, Ilaria De Maggio, Mario Capasso, et al. 2011. 'Two Founder Mutations in the SEC23B Gene Account for the Relatively High Frequency of CDA II in the Italian Population'. American Journal of Hematology 86 (9): 727–32. https://doi.org/10.1002/ajh.22096.

Russo, Roberta, Roberta Marra, Immacolata Andolfo, Gianluca De Rosa, Barbara Eleni Rosato, Francesco Manna, Antonella Gambale, et al. 2019. 'Characterization of Two Cases of Congenital Dyserythropoietic Anemia Type I Shed Light on the Uncharacterized C15orf41 Protein'. Frontiers in Physiology 10: 621. https://doi.org/10.3389/fphys.2019.00621.

Sassa, Akira, Manabu Yasui, and Masamitsu Honma. 2019. 'Current Perspectives on Mechanisms of Ribonucleotide Incorporation and Processing in Mammalian DNA'. Genes and Environment 41 (1): 3.
https://doi.org/10.1186/s41021-019-0118-7.

Sauna, Zuben E., Suneet Shukla, and Suresh V. Ambudkar. 2005. 'Disulfiram, an Old Drug with New Potential Therapeutic Uses for Human Cancers and Fungal Infections'. Molecular BioSystems 1 (2): 127. https://doi.org/10.1039/b504392a.

Savarino, Andrea. 2007. 'In-Silico Docking of HIV-1 Integrase Inhibitors Reveals a Novel Drug Type Acting on an Enzyme/DNA Reaction Intermediate'. Retrovirology 4 (1): 21. https://doi.org/10.1186/1742-4690-4-21.

Sekirnik, Rok, Nathan R. Rose, Armin Thalhammer, Peter T. Seden, Jasmin Mecinović, and Christopher J. Schofield. 2009. 'Inhibition of the Histone Lysine Demethylase JMJD2A by Ejection of Structural Zn(Ii)'. Chemical Communications, no. 42: 6376. https://doi.org/10.1039/b916357c.

Sengerová, Blanka, Charles K. Allerston, Mika Abu, Sook Y. Lee, Janet Hartley, Konstantinos Kiakos, Christopher J. Schofield, John A. Hartley, Opher Gileadi, and Peter J. McHugh. 2012. 'Characterization of the Human SNM1A and

SNM1B/Apollo DNA Repair Exonucleases'. Journal of Biological Chemistry 287 (31): 26254–67. https://doi.org/10.1074/jbc.M112.367243.

Sevajol, Marion, Lorenzo Subissi, Etienne Decroly, Bruno Canard, and Isabelle Imbert. 2014. 'Insights into RNA Synthesis, Capping, and Proofreading Mechanisms of SARS-Coronavirus'. Virus Research 194 (December): 90–99. https://doi.org/10.1016/j.virusres.2014.10.008.

Shalev, Hanna, Kaid Al-Athamen, Itai Levi, Aviva Levitas, and Hannah Tamary. 2017. 'Morbidity and Mortality of Adult Patients with Congenital Dyserythropoietic Anemia Type I'. European Journal of Haematology 98 (1): 13–18. https://doi.org/10.1111/ejh.12778.

Shalev, Hanna, Joseph Kapelushnik, Asher Moser, Orly Dgany, Tatyana Krasnov, and Hannah Tamary. 2004. 'A Comprehensive Study of the Neonatal Manifestations of Congenital Dyserythropoietic Anemia Type I'. Journal of Pediatric Hematology/Oncology 26 (11): 746–48. Shroff, Maithili, Axel Knebel, Rachel Toth, and John Rouse. n.d. 'A Complex Comprising C15ORF41 and Codanin-1- the Products of Two Genes Mutated in Congenital Dyserythropoietic Anemia Type I (CDA-I)'. Biochemical Journal. Accessed 28 April 2020. https://doi.org/10.1042/BCJ20190944.

Skourti-Stathaki, K., and N. J. Proudfoot. 2014. 'A Double-Edged Sword: R Loops as Threats to Genome Integrity and Powerful Regulators of Gene Expression'. Genes & Development 28 (13): 1384–96. https://doi.org/10.1101/gad.242990.114.

Smith, Everett Clinton, Hervé Blanc, Marco Vignuzzi, and Mark R. Denison. 2013. 'Coronaviruses Lacking Exoribonuclease Activity Are Susceptible to Lethal Mutagenesis: Evidence for Proofreading and Potential Therapeutics'. Edited by Michael S. Diamond. PLoS Pathogens 9 (8): e1003565. https://doi.org/10.1371/journal.ppat.1003565.

Sparks, Justin L., Hyongi Chon, Susana M. Cerritelli, Thomas A. Kunkel, Erik Johansson, Robert J. Crouch, and Peter M. Burgers. 2012. 'RNase H2-Initiated Ribonucleotide Excision Repair'. Molecular Cell 47 (6): 980–86. https://doi.org/10.1016/j.molcel.2012.06.035.

Steczkiewicz, Kamil, Anna Muszewska, Lukasz Knizewski, Leszek Rychlewski, and Krzysztof Ginalski. 2012. 'Sequence, Structure and Functional Diversity of PD-(D/E)XK Phosphodiesterase Superfamily'. Nucleic Acids Research 40 (15): 7016–45. https://doi.org/10.1093/nar/gks382.

Subissi, Lorenzo, Clara C. Posthuma, Axelle Collet, Jessika C. Zevenhoven-Dobbe, Alexander E. Gorbalenya, Etienne Decroly, Eric J. Snijder, Bruno Canard, and Isabelle Imbert. 2014. 'One Severe Acute Respiratory Syndrome Coronavirus Protein Complex Integrates Processive RNA Polymerase and Exonuclease Activities'. Proceedings of the National Academy of Sciences 111 (37): E3900–3909. https://doi.org/10.1073/pnas.1323705111.

Sugimoto, Yoichiro, Julian König, Shobbir Hussain, Blaž Zupan, Tomaž Curk, Michaela Frye, and Jernej Ule. 2012. 'Analysis of CLIP and ICLIP Methods for Nucleotide-Resolution Studies of Protein-RNA Interactions'. Genome Biology 13 (8): R67. https://doi.org/10.1186/gb-2012-13-8-r67.

Swickley, Grace, Yehoshua Bloch, Lidor Malka, Adi Meiri, Sharon Noy-Lotan, Amiel Yanai, Hannah Tamary, and Benny Motro. 2020. 'Characterization of the Interactions between Codanin-1 and C15Orf41, Two Proteins Implicated in Congenital Dyserythropoietic Anemia Type I Disease'. BMC Molecular and Cell Biology 21 (March). https://doi.org/10.1186/s12860-020-00258-1.

Tamary, H., H. Offret, O. Dgany, B. Foliguet, S. N. Wickramasinghe, T. Krasnov, F. Rumilly, et al. 2008. 'Congenital Dyserythropoietic Anaemia, Type I, in a Caucasian Patient with Retinal Angioid Streaks (Homozygous Arg1042Trp Mutation in Codanin-1)'. European Journal of Haematology 80 (3): 271–74. https://doi.org/10.1111/j.1600-0609.2007.01004.x.

Tamary, Hannah, and Orly Dgany. n.d. 'Congenital Dyserythropoietic Anemia Type I', 12. Tamary, Hannah, Orly Dgany, Alexis Proust, Tatyana Krasnov, Nili Avidan, Tal Eidelitz-Markus, Gil Tchernia, et al. 2005. 'Clinical and Molecular Variability in Congenital Dyserythropoietic Anaemia Type I'. British Journal of Haematology 130 (4): 628–34. https://doi.org/10.1111/j.1365-2141.2005.05642.x.

Tamary, Hannah, Nathaly Marcoux, Sharon Noy-Lotan, Isaac Yaniv, and Orly Dgany. 2010. 'Codanin-1, the Product of the Gene Mutated In Congenital Dyserythropoietic Anemia Type I (CDA I), Binds to Histone Chaperone Asf1a and Inhibits Its Nucleosome Assembly Activity'. Blood 116 (21): 1004–1004. https://doi.org/10.1182/blood.V116.21.1004.1004.

Tan, Siun Chee, and Beow Chin Yiap. 2009. 'DNA, RNA, and Protein Extraction: The Past and The Present'. Journal of Biomedicine and Biotechnology 2009: 1–10. https://doi.org/10.1155/2009/574398.

Tao, Jiayi, Min Zhu, He Wang, Solomon Afelik, Matthew P. Vasievich, Xiao-Wei Chen, Guojing Zhu, Jan Jensen, David Ginsburg, and Bin Zhang. 2012. 'SEC23B Is Required for the Maintenance of Murine Professional Secretory Tissues'. Proceedings of the National Academy of Sciences of the United States of America 109 (29): E2001–9. https://doi.org/10.1073/pnas.1209207109.

Tay, Matthew Zirui, Chek Meng Poh, Laurent Rénia, Paul A. MacAry, and Lisa F. P. Ng. 2020. 'The Trinity of COVID-19: Immunity, Inflammation and Intervention'. Nature Reviews Immunology 20 (6): 363–74. https://doi.org/10.1038/s41577-020-0311-8.

Thomas, Caroline, and Andrew B. Lumb. 2012. 'Physiology of Haemoglobin'. Continuing Education in Anaesthesia Critical Care & Pain 12 (5): 251–56. https://doi.org/10.1093/bjaceaccp/mks025.

Trott, Oleg, and Arthur J. Olson. 2009. 'AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading'. Journal of Computational Chemistry, NA-NA. https://doi.org/10.1002/jcc.21334.

Tumey, L. Nathan, David Bom, Bayard Huck, Elizabeth Gleason, Jianmin Wang, Daniel Silver, Kurt Brunden, et al. 2005. 'The Identification and Optimization of a N-Hydroxy Urea Series of Flap Endonuclease 1 Inhibitors'. Bioorganic & Medicinal Chemistry Letters 15 (2): 277–81. https://doi.org/10.1016/j.bmcl.2004.10.086.

Van Nostrand, Eric L, Gabriel A Pratt, Alexander A Shishkin, Chelsea Gelboin-Burkhart, Mark Y Fang, Balaji Sundararaman, Steven M Blue, et al. 2016.

‘Robust Transcriptome-Wide Discovery of RNA-Binding Protein Binding Sites with Enhanced CLIP (ECLIP)’. Nature Methods 13 (6): 508–14. https://doi.org/10.1038/nmeth.3810.

V’kovski, Philip, Annika Kratzel, Silvio Steiner, Hanspeter Stalder, and Volker Thiel. 2020. ‘Coronavirus Biology and Replication: Implications for SARS-CoV-2’. Nature Reviews Microbiology, October. https://doi.org/10.1038/s41579-020-00468-6.

Wang, Yongwei, Yongxin Ru, Gang Liu, Shuxu Dong, Yuan Li, Xiaofan Zhu, Fengkui Zhang, Yan-Zhong Chang, and Guangjun Nie. 2018. ‘Identification of CDAN1, C15ORF41 and SEC23B Mutations in Chinese Patients Affected by Congenital Dyserythropoietic Anemia’. Gene 640 (January): 73–78. https://doi.org/10.1016/j.gene.2017.10.027.

Weintraub, Harold. 1979. ‘Assembly of an Active Chromatin Structure during Replication’. Nucleic Acids Research 7 (3): 781–92. https://doi.org/10.1093/nar/7.3.781.

Wickramasinghe, S. N. 1998. ‘Congenital Dyserythropoietic Anaemias: Clinical Features, Haematological Morphology and New Biochemical Data’. Blood Reviews 12 (3): 178–200. https://doi.org/10.1016/s0268-960x(98)90016-9.

Wickramasinghe, Sunitha N., and William G. Wood. 2005. ‘Advances in the Understanding of the Congenital Dyserythropoietic Anaemias’. British Journal of Haematology 131 (4): 431–46. https://doi.org/10.1111/j.1365-2141.2005.05757.x.

Williams, Jessica S., Scott A. Lujan, and Thomas A. Kunkel. 2016. ‘Processing Ribonucleotides Incorporated during Eukaryotic DNA Replication’. Nature Reviews Molecular Cell Biology 17 (6): 350–63. https://doi.org/10.1038/nrm.2016.37.

Wrapp, Daniel, Nianshuang Wang, Kizzmekia S. Corbett, Jory A. Goldsmith, Ching-Lin Hsieh, Olubukola Abiona, Barney S. Graham, and Jason S. McLellan. 2020. ‘Cryo-EM Structure of the 2019-NCoV Spike in the Prefusion Conformation’. Science 367 (6483): 1260–63. https://doi.org/10.1126/science.abb2507.

Yoshimoto, Francis K. 2020. 'The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19'. The Protein Journal 39 (3): 198–216. https://doi.org/10.1007/s10930-020-09901-4.

Zhao, Wei, Claire Kitidis, Mark D. Fleming, Harvey F. Lodish, and Saghi Ghaffari. 2006. 'Erythropoietin Stimulates Phosphorylation and Activation of GATA-1 via the PI3-Kinase/AKT Signaling Pathway'. Blood 107 (3): 907–15. https://doi.org/10.1182/blood-2005-06-2516.

Zheng, L., and B. Shen. 2011. 'Okazaki Fragment Maturation: Nucleases Take Centre Stage'. Journal of Molecular Cell Biology 3 (1): 23–30. https://doi.org/10.1093/jmcb/mjq048.

Zhu, Lian, Tiffany M. Richardson, Ludivine Wacheul, Ming-Tzo Wei, Marina Feric, Gena Whitney, Denis L. J. Lafontaine, and Clifford P. Brangwynne. 2019. 'Controlling the Material Properties and RRNA Processing Function of the Nucleolus Using Light'. Proceedings of the National Academy of Sciences 116 (35): 17330–35. https://doi.org/10.1073/pnas.1903870116.

Zhu, Liang, and Arthur I Skoultchi. 2001. 'Coordinating Cell Proliferation and Differentiation'. Current Opinion in Genetics & Development 11 (1): 91–97. https://doi.org/10.1016/S0959-437X(00)00162-3.

Zivot, Andrea, Jeffrey M. Lipton, Anupama Narla, and Lionel Blanc. 2018. 'Erythropoiesis: Insights into Pathophysiology and Treatments in 2017'. Molecular Medicine 24 (1): 11. https://doi.org/10.1186/s10020-018-0011-z.

Zuo, Y. 2001. 'Exoribonuclease Superfamilies: Structural Analysis and Phylogenetic Distribution'. Nucleic Acids Research 29 (5): 1017–26. https://doi.org/10.1093/nar/29.5.1017.