

Biochemical Characterization of C15ORF41: A Novel Nuclease

A thesis submitted to the University of Oxford for the degree of Masters of Research – Michaelmas 2016



Rami M. Ezzibdeh

New College

Weatherall Institute of Molecular Medicine

University of Oxford

Abstract

Congenital dyserythropoietic anaemia type I (CDA-I) is a rare autosomal recessive form of anaemia that is characterized by ineffective erythropoiesis and abnormal morphologies of bone marrow erythroblasts. While previous studies have indicated mutations in the *CDANI* gene for a large group of patients with CDA-I, new research has revealed that mutations in the *C15ORF41* gene are also linked to patients with the condition.

With the recent purification of the C15ORF41 protein by the Structural Genomics Consortium at the University of Oxford, a series of biochemical assays were conducted in order to characterize the catalytic and mechanistic behaviour of the protein. Nuclease assays indicated that the protein, when co-purified with the *CDANI* protein Codanin1, acts as a single-stranded endonuclease that acts on both DNA and RNA substrates. These assays also revealed that the complex incises in 5-6 nucleotide increments in a 3'-to-5' directionality and exhibits a preference for 3' flaps. Electrophoretic Mobility Shift Assays indicated that there is likely a 2:1 protein to substrate stoichiometry. Fluorescence anisotropy data showed that the C15ORF41-Codanin complex appeared to have much higher affinity to DNA substrates than the individual constituent proteins.

The results of this study provided initial evidence into the predicted endonucleolytic behaviour of the protein and established a solid foundation for further investigations of the mechanistic basis of its activity. This is absolutely critical for the molecular understanding of CDA-I for the formulation of future treatments, and is the first step in examining a novel endonuclease that could further our grasp on DNA damage repair proteins.

Acknowledgements

First, I want to thank my advisors Prof. Peter McHugh and Prof. Tom Brown for giving me the opportunity to work on this project. Their continued support and guidance throughout my degree is extremely appreciated.

I am also very grateful to all members of the McHugh lab, current and past: Anna, Lianne, Ghadeer, Sal, Susie, and Erkin for their support and help. Sheriff Lonnie for bringing law and order to the lab. Sanja for being a great bench-mate. Hannah for having a laugh that defies the laws of acoustics and somehow matching the decibel count of a Concorde sonic boom. I would also like to thank Ronan, the honorary McHugh lab member, for his advice and expertise. Special thanks to Sook for being an incredible teacher and supervisor, and for not complaining every time I asked her to remind me how to do an essay.

I also have to thank my brothers Emad and Tariq for always being there for me. Finally, this thesis is dedicated to my parents, whose unwavering support is beyond description.

Table of Contents

I. Cover Page	1
II. Abstract	2
III. Acknowledgments	3
IV. Table of Contents	4
V. Table of Abbreviations	7
1. Introduction	10
1.1. Definition of Anaemia	10
1.2. Clinical Presentation and Treatment of Anaemia	11
1.3. Structure and Functional Role of Haemoglobin	12
1.4. Erythropoiesis	12
1.5. Types of Anaemia	14
1.5.1. Microcytic Anaemia	15
1.5.2. Normocytic Anaemia	18
1.5.3. Macrocytic Anaemia	19
1.6. Congenital Dyserythropoietic Anaemia – General Overview	20
1.6.1. Congenital Dyserythropoietic Anaemia Type II	22
1.6.2. Congenital Dyserythropoietic Anaemia Type III	24
1.6.3. Congenital Dyserythropoietic Anaemia Type I	24
1.7. DNA Damage	34
1.7.1. Sources of DNA Damage	34
1.8. DNA Repair Mechanisms	35
1.8.1. Nucleotide Excision Repair	35

1.8.2. Double Strand Break Repair	38
1.8.2.1. Non-Homologous End Joining	38
1.8.2.2. Homologous Recombination	38
1.8.3. Interstrand Crosslink Repair	41
1.9. Aims of Project	44
2. Materials and Methods	45
2.1. Protein Purification	45
2.2. Substrate Labelling	46
2.2.1. α (3') Radiolabelling of Oligos	46
2.2.2. γ (5') Radiolabelling of Oligos	46
2.2.3. Substrate Annealing	47
2.3. Radioactive Nuclease Assay	47
2.4. Electrophoretic Mobility Shift Assay (EMSA)	48
2.5. Plasmid Assay	48
2.6. Fluorescence Assay	49
2.7. Fluorescence Anisotropy	49
3. Biochemical Behaviour of Protein Complex	52
3.1. Protein Purification	52
3.2. Plasmid Assay	53
3.3. Fluorescent DNA Substrates Assay	55
3.4. Radioactive Substrate Assays	59
3.5. Comparative Gels	73
3.6. Electrophoretic Mobility Shift Assay (EMSA)	74

3.7. Fluorescence Anisotropy	76
3.8. Discussion	78
4. Protein Mutants and Biochemical Behaviour of Individual Proteins in Non-Complex Form	80
4.1. Nuclease Inactive Mutant	80
4.2. Activity of Individual Proteins Within Complex	85
4.3. Discussion and Conclusion	89
5. Future Work	92
6. References	94
7. Appendix	108

Table of Abbreviations

6-4PP	6-4 Pyrimidine Photoproducts
ACD	Anaemia of Chronic Disease
AML	Acute Myeloid Leukemia
CDA	Congenital Dyserythropoietic Anaemia
CNS	Central Nervous System
CPD	Cyclobutane Pyrimidine Dimer
CS	Cockayne Syndrome
CtIP	C-terminal Binding Protein (CtBP) Interacting Protein
DBA	Diamond-Blackfan Anaemia
DDT	Dithiothreitol
DHF	Dihydrofolate
DHJ	Double Holliday Junction
DSB	Double Strand Break
dsDNA	Double Stranded DNA
EM	Electron Microscopy
EMSA	Electrophoretic Mobility Shift Assay
EPO	Erythropoietin
FA	Fanconi Anaemia
GG-NER	Global Genome Nucleotide Excision Repair
Hb	Haemoglobin

hHRad23B	Human Homolog of Rad23B
HIF-2 α	Hypoxia-Inducible Factor 2 α
HJ	Holliday Junction
HR	Homologous Recombination
HtH	Helix-turn-helix
ICL	Interstrand Crosslink
IDA	Iron Deficiency Anaemia
IR	Infrared
MCV	Mean Corpuscular Volume
MHF	MPH1-Associated Histone-Fold Protein
NER	Nucleotide Excision Repair
NHEJ	Non-Homologous End-Joining
PCNA	Proliferating Cell Nuclear Antigen
RBC	Red Blood Cell
ROI	Region of Interest
RPA	Replication Protein A
SSB	Single Strand Break
SSDA	Synthesis Strand-Dependent annealing
ssDNA	Single Stranded DNA
T4 PNK	T4 Polynucleotide Kinase
TC-NER	Transcription-Coupled Nucleotide Excision Repair
THF	Tetrahydrofolate
TLS	Translesion Synthesis

TTD	Trichothiodystrophy
UV	Ultraviolet
UV-DDB	UV-Damaged DNA Binding Protein
XP	Xeroderma pigmentosum
XPC	Xeroderma pigmentosum Group C Protein
XPF-ERCC1	XPF-Excision Repair Cross Complementing 1

1. Introduction

Part I: Anaemia – A General Overview

1.1 Definition of Anaemia

Anaemia is a medical condition characterized by decreased levels of haemoglobin (Hb) when compared to reference values for the patient's sex and age (Hugh Jones et al. 2008). In and of itself, anaemia is not a final clinical diagnosis but rather a manifestation of a disease that has in turn led to a decline in the red blood cell (RBC) count. The normal ranges for haemoglobin levels are indicated in Table 1.1.

Table 1.1 (Adapted: Pagana et al. 2015): Normal Ranges for Haemoglobin Levels

Hb Concentration (g/dL)	
Adult Male	14.0 – 18.0
Adult Female	12.0 – 16.0
Pregnant Female	> 11.0
Newborn	14.0 – 24.0
0 – 2 Weeks	12.0 – 20.0
2 – 6 Months	10.0 – 17.0
6 Months – 6 Years	9.5 – 14.0
6 – 18 Years	10.0 – 15.5

According to the World Health Organization, anaemia affects approximately 1.62 billion people from around the world, or 24.8% of the global population, as of 2007 (de Benoist et al. 2008). As will be discussed in detail later on, anaemia can arise from a large number of intrinsic disorders and conditions as well as extrinsic and societal factors such as malnutrition.

1.2 Clinical Presentation and Treatment of Anaemia

The presence and severity of symptoms in anaemia is heavily dependent on the degree of decline in haemoglobin levels. The National Cancer Institute has created a 4-grade classification system for anaemia: mild is defined as Hb of 10.0 g/dL till normal levels indicated in Table 1.1, moderate from (8.0 – 10.0) g/dL, severe from (6.5 – 7.9) g/dL, and life-threatening for Hb < 6.5 g/dL (Bohlius et al. 2006).

Since haemoglobin is necessary for circulating oxygen from the lungs throughout the body, the symptoms of anaemia are diverse and affect a wide range of organ systems. Those generally include fatigue, dizziness, irregular bowel movement, brittle nails, decline in cognitive function, and others depending on the root cause of the anaemia (Ludwig and Strasser 2001). In terms of specific organ symptomology, tachycardia may arise due to the decline in blood viscosity caused by the decreased haematocrit, as well as to compensate for the fall in oxygen levels in the peripheral organs. Increase in respiratory rates is also promoted by the lowered oxygen in the blood, which is accompanied by a rise in 2,3-bisphosphoglycerate as a counterbalance that eases oxygen dissociation from the erythrocytes to the tissues (Ludwig and Strasser 2001).

The treatment of anaemia depends on both its severity and cause. In general, the management could range from dietary changes to blood transfusion in minor cases to surgical intervention in more serious, life-threatening situations. The specific treatments for the types of anaemia discussed later will be outlined accordingly.

1.3 Structure and Functional Role of Haemoglobin

The functional role of red blood cells is possible due to haemoglobin's ability to bind to oxygen atoms. Human haemoglobin is a tetramer that consists of 4 globin polypeptides that each associate with a haem group. In adults, the majority of haemoglobin is composed of 2 α and 2 β globin units ($\alpha_2\beta_2$) with a small percentage of $\alpha_2\delta_2$ and $\alpha_2\gamma_2$ (Mehta et al. 2009), with each haem containing an iron ion coordinated in its core (Figure 1.1). The haem unit is synthesized from a protoporphyrin precursor. As erythrocytes make their way through the blood stream to the capillaries in the peripheral organs, the gradient in oxygen concentration promotes its release into the tissues for use in oxidative phosphorylation and metabolic processes.

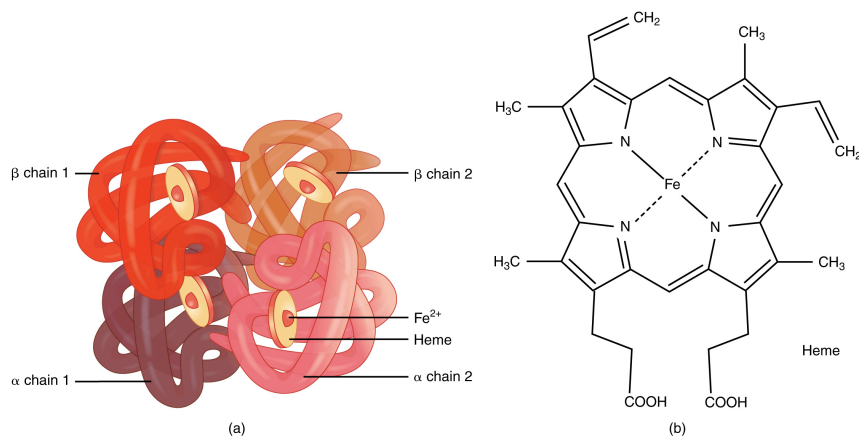


Figure 1.1: a) Quaternary structure of Haemoglobin showing the 2 alpha and 2 beta globin units and their associated haem groups. b) Molecular structure of haem with the coordinated Fe ions (OpenStax 2013).

1.4 Erythropoiesis

The body is constantly generating erythrocytes, with approximately 1 trillion cells produced on a daily basis in a normal adult bone marrow (Doulatov et al. 2012). The developmental process of erythrocyte production begins with the haematopoietic stem cell that goes through

a myeloid progenitor stage before the formation of the erythrocyte precursor pronormoblast in the red bone marrow. The pronormoblast then progressively divides into smaller normoblasts (erythroblasts) that contain higher concentrations of haemoglobin in the cytoplasm. The nucleus is then ejected from the normoblast, which leads to the formation of a reticulocyte that then circulates the blood stream for 1 – 2 days before finally maturing into an erythrocyte in the spleen (Hoffbrand et al. 2001). This process is regulated by a number of factors, most importantly erythropoietin (EPO), a hormone released by the peritubular interstitial cells in the kidneys in response to hypoxic conditions (Hattangadi et al. 2011). There is evidence for thyroxine (T4) hormone stimulating the release of EPO, which explains the elevated levels of Hb in males, as testosterone is known to promote T4 secretion (Kawa and Machalinski 2014).

Figure 1.2 presents a schematic of erythropoiesis.

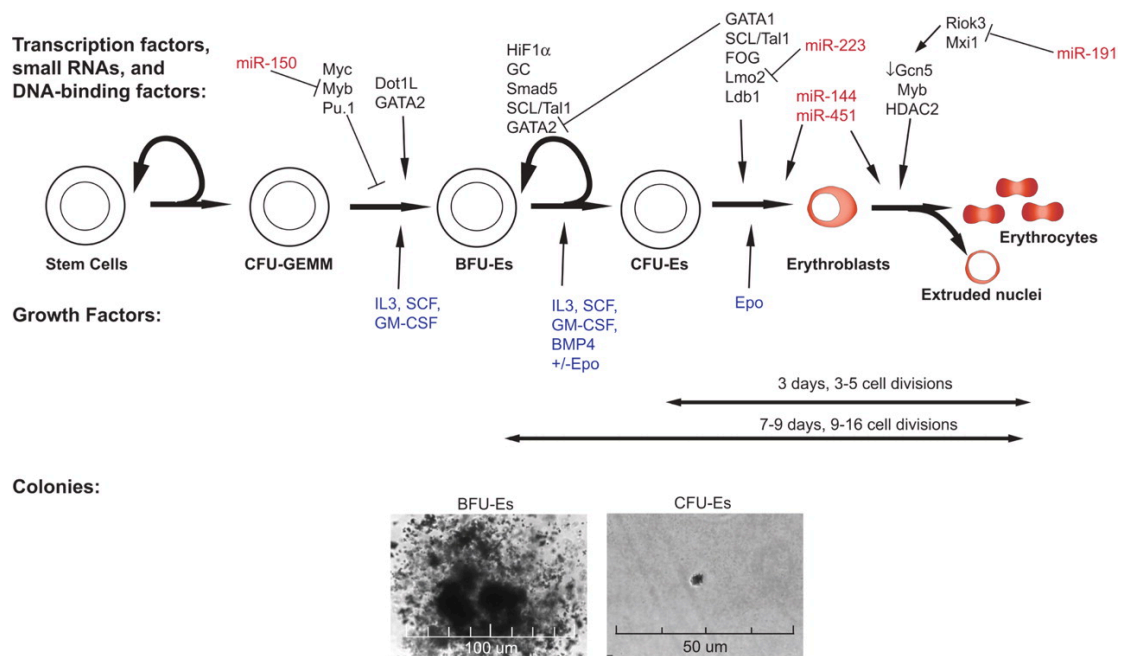


Figure 1.2: Diagram illustrating the process of erythropoiesis starting with stem cells in the red bone marrow to the erythrocytes (Hattangadi et al. 2011).

1.5 Types of Anaemia

There are many different types of anaemia that have distinct molecular and cellular features, and only the relevant ones will be discussed in this work. In general, anaemia could arise from one of two reasons: destruction of erythrocytes or the underproduction/malformation of erythrocytes. Haemolytic anaemia is characterized by excessive degradation of erythrocytes and is usually caused by autoimmune responses (Gehrs and Friedberg 2002). In this case, patients have a high reticulocyte count as the bone marrow attempts to compensate for the decline in blood haemoglobin. The underproduction of erythrocytes, on the other hand, is independent of erythrocyte degradation and is rooted in the erythropoiesis developmental process outlined earlier. Accordingly, patients have a lower reticulocyte count as fewer progenitor cells proceed through erythropoiesis intact. There are three broad categories of such anaemia: microcytic, normocytic, and macrocytic anaemia. These are defined by the average size of the erythrocytes, measured as the mean corpuscular volume (MCV) in femtoliter. Microcytic anaemia occurs when the MCV is less than 80 fL, normocytic for an MCV between 80 – 100 fL, and macrocytic anaemia for MCV values greater than 100 fL, and can be calculated using equation 1.1 (Aslinia et al. 2006).

$$\textbf{Equation 1.1} \quad MCV (fL) = \frac{Hematocrit \% \times 10}{RBC \text{ count } \left(\frac{10^6}{\mu L} \right)}$$

The following will contain a discussion of each type of anaemia.

1.5.1 Microcytic Anaemia

Microcytosis occurs when there is a disruption in the production of any of the subunits of haemoglobin. When the defect occurs in protoporphyrin, an important precursor in haem production, Sideroblastic anaemia occurs, as the iron cannot associate with its binding partner, resulting in iron accumulation in mitochondria (Gattermann 2000). If either the α or β globin units are defective, Thalassemia occurs. Accordingly, a fall in Hb levels occurs followed with further cell divisions to maintain a constant Hb concentration producing smaller sized erythrocytes. The inheritance pattern for Thalassemia is autosomal recessive, and depending on the severity it can be classified as Thalassemia major, intermedia, or minor (Taher et al. 2006).

Iron deficiency anaemia (IDA) is a hypochromic form of the condition that is associated with a decrease in iron levels. It is by far the most common form of anaemia, affecting millions of people worldwide, with premenopausal women and children being the most affected groups (Killip et al. 2007). The physiological regulation of iron is unusual since it is conducted via changes in absorption levels, with the body having no controlled and systematic mechanism for getting rid of excess iron (Clark 2008).

Regulation of iron absorption takes place through controlling the intake of dietary iron in the proximal gastrointestinal (GI) tract, in which iron passes from the intestinal lumen into the enterocyte via a divalent metal transporter and finally through the iron exporter ferroportin where it is bound to transferrin and released into the blood stream (Ganz 2011). When blood oxygen levels fall, the kidneys increase the production of EPO as a response to elevated levels

of hypoxia-inducible factor 2 α (HIF-2 α), which then promotes erythropoiesis in the bone marrow (Camaschella 2015). As a consequence of increased erythropoiesis, erythroblasts suppress hepcidin, a regulatory hormone that leads to the degradation of ferroportin, in order to increase the uptake of dietary iron via the intestinal lumen (Ganz 2011). Figure 1.3 shows a schematic of iron regulation and the role hepcidin plays in controlling the intake of dietary iron.

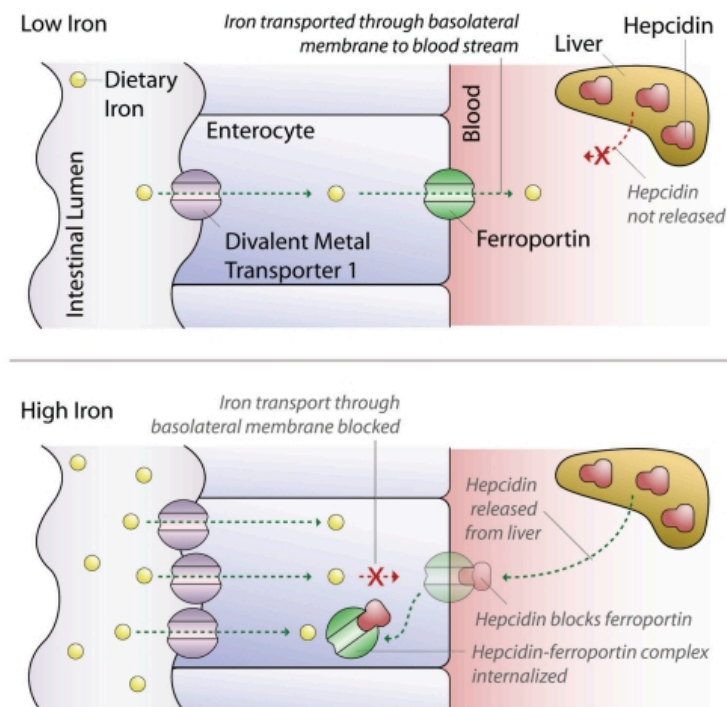


Figure 1.3: Illustrates the pathway in which dietary iron is absorbed in the duodenum of the proximal GI. It indicates the suppression of hepcidin in response to a decline in iron, which in turn allows the transport of iron through ferroportin into the blood (Naigamwalla 2012).

In patients with Anaemia of Chronic Disorder (ACD), there is evidence to suggest that the inflammation, which could be caused by cancer, infections, or autoimmune disorders, triggers the secretion of hepcidin by an increase in cytokine levels (Nemeth et al. 2004). This then has the effect outlined earlier of blocking ferroportin by the internalization of the hepcidin-ferroportin complex, thus lowering iron intake into the blood. Figure 1.4 provides a simple illustration of the main types of microcytic anaemia discussed and the underlying molecular causes. One clinical difference between IDA and ACD is the level of ferritin, as iron deficiency leads to a decrease in iron binding ferritin in liver macrophages. On the other hand, ACD patients have increased levels of ferritin as the body attempts to elevate the amount of stored iron through storage in the liver macrophages (Recalcati et al. 1998).

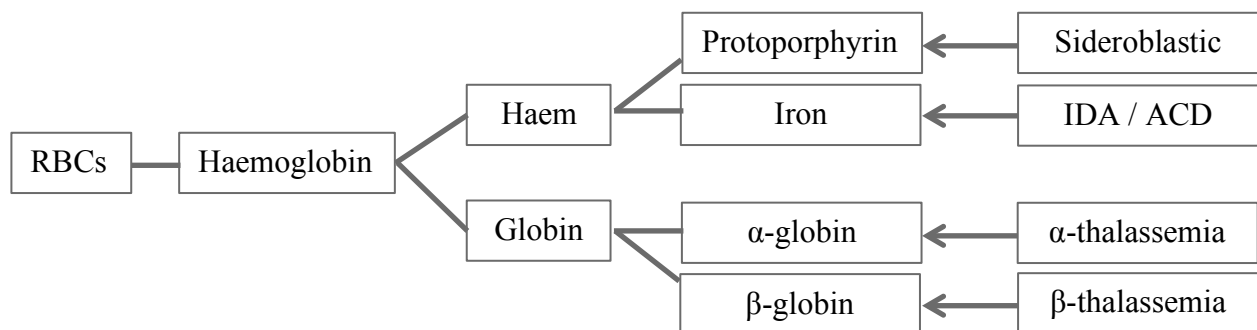


Figure 1.4: A summary of the types of microcytic anaemia discussed in the earlier section. Malformations or absence of the indicated subunits can cause the identified types of anaemia.

1.5.2 Normocytic Anaemia

Normocytic anaemia occurs when there is a decline in the haemoglobin levels with the erythrocytes maintaining a normal MCV of 80 – 100 fL. There are several ways this form of anaemia could arise, including bleeding, haemolysis, and bone marrow failure (Armitage 2004). Diamond-Blackfan (DBA) anaemia is an autosomal dominant congenital disorder characterized by erythrocyte aplasia (Ball 2011). The genetic basis for the arrest in erythrocyte progenitor cells has been linked to mutations in ribosomal proteins. In the case of Fanconi anaemia (FA), the bone marrow failure leads to aplastic anaemia, in which the haematopoietic stem cells are damaged and accordingly a drop in all blood cell types results, also known as pancytopenia (Greer et al 2009, Glanz and Fraser 1982). FA is a rare congenital disease that has an autosomal recessive inheritance pattern (Meetei et al. 2003). It is an interesting form of anaemia since it has been used extensively as a model in oncology and mechanistic DNA repair research.

Fanconi anaemia patients have mild to severe congenital abnormalities, short stature, and marrow failure (Kutler et al. 2003, and Garcia-Higuera et al. 2001). Acute myeloid leukaemia (AML) and squamous cell carcinomas, usually in the head and neck regions, develop in a significant proportion of FA patients (Howlett 2002). A detailed description of the molecular aetiology of FA will be provided when discussing DNA repair mechanisms in later chapters.

1.5.3 Macrocytic Anaemia

When the mean corpuscular volume of erythrocytes is greater than 100 fL, the patient suffers from one of the forms of macrocytic anaemia. Most cases of macrocytosis occur from vitamin B₁₂ or folate deficiencies, with excessive alcohol consumption and liver damage constituting a proportion of macrocytic anaemia cases (Armitage 2004). Folate is the conjugate base of folic acid, also known as vitamin B9. Folic acid is converted to dihydrofolate (DHF) and then tetrahydrofolate (THF) by dihydrofolate reductase, which acts as a 1-carbon donor used in the production of DNA, RNA, and amino acids (Shane and Stokstad 1985). Vitamin B₁₂ regenerates THF from Methyl-THF, forming methyl-B₁₂ that is then utilized to transfer the methyl group to homocysteine to form methionine (Baik and Russell 1999, and Tiemeier et al. 2002). The lack of DNA precursors and nucleotides significantly affects cellular division, which is very pronounced in the rapidly dividing cells of haematopoiesis. Therefore, the G₂-M transition is delayed and excess growth before cellular division occurs, leading to large erythrocytes (Kozyraki et al. 1998). Congenital dyserythropoietic anaemia type I (CDA-I) is a form of macrocytic anaemia that will be discussed in great detail in the following chapter.

Part II: Congenital Dyserythropoietic Anaemia

1.6 Congenital Dyserythropoietic Anaemia – General Overview

Congenital dyserythropoietic anaemia (CDA) is a group of extremely heterogeneous and rare hereditary anaemias (Iolascon et al. 2011 and Carreno-Tarragona et al. 2015). They are characterized mainly by ineffective erythropoiesis, anomalies in the morphological features of erythroblasts in the bone marrow, and secondary haemochromatosis (Marks and Mitus 1996). Clinically, the patients typically present with symptoms of anaemia, enlarged spleens, jaundice, and suboptimal reticulocyte responses. In the 1960s, Wendt and Heimpel created a classification system for the different forms of CDA with three major categories: I, II and III (Kamiya and Manabe 2010). That system is used clinically to this day, with some minor modifications and the addition of other variant types as some patients did not match any of the original three classifications. Table 1.2 shows a modern version of Heimpel's types of CDA.

Table 1.2 (Iolascon et al. 2013): Summary of the clinical, molecular, and cellular features of CDA

CDA type	I	II	III Familial	III Sporadic	Variants
Inheritance	Autosomal recessive	Autosomal recessive	Dominant	Variable	Autosomal dominant or X linked or recessive
Cases reported	>300	>450	2 families	<20	~70
BM morphology (light microscopy)	Abnormal chromatin structure, chromatin bridges	Binuclearity, multinuclearity of mature erythroblasts	Giant multinucleated erythroblasts	Giant multinucleated erythroblasts	CDA I–like, CDA II–like, others
BM EM findings	“Spongy” heterochromatin, invagination of cytoplasm into the nucleus	Peripheral cisternae beneath the plasma membrane	Clefts in heterochromatin, autophagic vacuoles, intranuclear cisternae	Various	Various
Mutated gene	<i>CDAN1</i> , <i>C15ORF41</i>	<i>SEC23B</i>	<i>KIF23</i>	Unknown	<i>KLF1</i> , <i>GATA-1</i> , unknown
Associated dysmorphology/organ involvement	Skeleton	Variable, rare	Monoclonal gammopathy, myeloma, angiod streaks	Variable	CNS, others

In general, a CDA diagnosis is conducted by matching the patient's clinical presentation with 4 criteria. First, there has to be evidence of congenital or hereditary anaemia. Ineffective erythropoiesis and abnormal morphology of BM erythroblasts also have to be established. Finally, the physician has to eliminate other forms of hereditary anaemias that could be responsible for the dyserythropoiesis, such as hereditary sideroblastic anaemia and thalassemia (Heimpel 2004).

The next sections will outline the clinical and molecular features of the three major types of CDA, starting with CDAs II and III, in order to build a holistic picture of the CDA family. That will be followed by a more detailed description of CDA type I for its clinical and translational relevance to the project. Images of bone marrow show large and multinucleated erythroblasts with no signs of enlarged spleens.

1.6.1 Congenital Dyserythropoietic Anaemia Type II

As indicated by Table 2.1.1, CDA-II is the most common form of CDA, with more than 450 cases identified as of 2013, and it has an autosomal recessive inheritance pattern. Patients with CDA-II have moderate to severe anaemia, with Hb levels between 8 – 11 g/dL and normal MCV values, and often present with gallstones and splenomegaly (Kamiya and Manabe 2010). Patients also show iron overload, with secondary cirrhosis due to the iron accumulation occurring in approximately 20% of cases (Heimpel et al. 2003).

From a molecular perspective, CDA-II is associated with atypical erythroblast morphology with binuclearity and multinuclearity showing in up to 50% of those cells in the bone marrow

as seen under light microscopy (Queisser et al. 1971). Electron microscopy of BM tissue reveals double membranes in the mature erythroid cells from residual endoplasmic reticulum (Heimpel et al. 2003). A recent investigation identified mutations in the component of the *SEC23B* gene coding COPII that lead to the formation of CDA-II (Schwarz et al. 2009). COPII is a coat protein involved in the trafficking of proteins from the endoplasmic reticulum to the Golgi apparatus (Boyadjiev et al. 2006). It plays a key role in membrane homeostasis and protein localization, which could explain the multinuclearity observed in CDA-II (Kamiya and Manabe 2010). Some patients require blood transfusions, depending on the severity of the anaemia, and splenectomy surgeries are sometimes conducted.

Figure 1.5 shows an EM image of an erythroblast from a CDA-II patient and demonstrates the peripheral cistern characteristic of this condition.

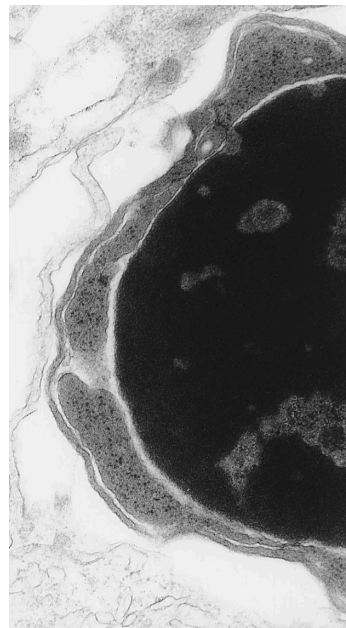


Figure 1.5: Electron microscope image of a CDA-II patient's erythroblast, indicating the peripheral cistern (Wickramasinghe 2000).

1.6.2 Congenital Dyserythropoietic Anaemia Type III

There have been only a few cases of CDA-III identified so far, making it the rarest form of CDA. CDA-III's inheritance is dominant, but there are a few cases of sporadic CDA-III that have reported in the literature recently (Wickramasinghe and Wood 2005). The gene identified in the case of CDA-III is *KIF23*, which encodes mitotic kinesin-like protein 1 (MKLP1), which is crucial for cytokinesis (Iolascon et al. 2013).

1.6.3 Congenital Dyserythropoietic Anaemia Type I

CDA-I is inherited in an autosomal recessive pattern, and as of 2013 there have been approximately 300 cases reported. It is typically a macrocytic anaemia, with MCV values ranging between 100 – 120 fL, although there have been several cases with normocytic MCV values. Hb levels, similar to CDA-II, are between 8 – 11 g/dL (Kamiya and Manabe 2010). Symptoms can begin showing neonatally, with signs of hepatosplenomegaly, jaundice, and growth retardation (Shaley et al. 2004). Some patients also present with skeletal abnormalities, especially of the distal limbs, with cases of missing nails or additional fingers. A significant percentage of patients show signs of iron overloading and Pappenheimer inclusions, which are most likely caused by increased iron absorption due to hepcidin inhibition as a result of the ineffective erythropoiesis (Nemeth and Ganz 2006). Studies have indicated that elevated levels of growth factor 15 (GF15) could be responsible for the stimulation of hepcidin, with a 38-fold increase noted in one paper (Tamary et al. 2008).

When investigated using a light microscope, CDA-I patients have approximately 10% binuclear erythroblasts with few that are more than binuclear, and up to 8% of intermediate

erythroblasts show inter-nuclear bridges (Heimpel et al. 2010). A defining feature for this form of anaemia is the spongy (or “Swiss cheese”) appearance of the erythroblast heterochromatin when imaged under electron microscopy, with up to 60% of cells showing this characteristic (Babbs et al. 2013), as indicated in Figure 1.6.

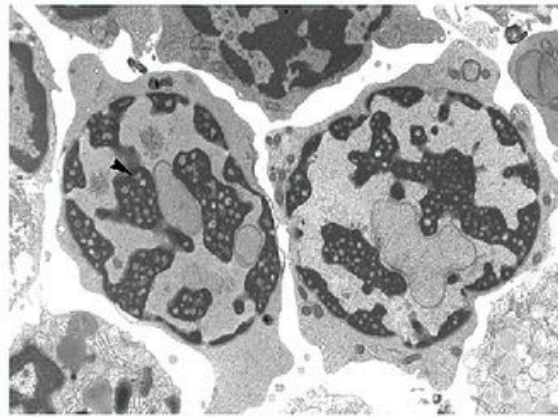


Figure 1.6: EM image of CDA-I erythroblast shows the spongy heterochromatin structure. These cells were differentiated in culture from CD34+ progenitors and separated from peripheral blood (Babbs et al. 2013).

A study in 2002 found homozygous mutations in the *CDANI* gene that were responsible for CDA-I in a group of patients from the Middle East (Gandy et al. 2002). *CDANI*, on chromosome 15q15, has 28 highly-conserved exons, 15kb of genomic DNA, and encodes a 4738 nucleotide mRNA that yields Codanin-1, a 1227 amino acid protein (Noy-Lotan et al. 2008). This ubiquitous protein is evolutionarily highly-conserved and has no orthologous relatives, proposing a non-redundant cellular role (Kamiya and Manabe 2010). There are no homozygous null-type mutations in *CDANI*, which suggests it is essential for survival. This is supported by the fact that most of the 30 mutations identified so far are missense mutations

leading to single nucleotide alterations in Codanin-1. Homozygous gene traps for *CDANI* in mice are lethal, with early termination in embryogenesis (Renella et al. 2011).

The functional role of Codanin-1 is still yet to be fully understood. Recent studies have shown this protein to localises to heterochromatin in interphase, and tracking of endogenous expression levels indicated significant increases in S-phase followed by a drop at mitosis. Five putative E2F binding sites in *CDANI* gene promoter region were found to be a target E2F1, a key G₁ to S regulatory transcription factor (Noy-Lotan et al. 2008). Previous work on cell cycle progression dynamics indicated that E2F knockout in erythroid cells cause S-phase arrest and dyserythropoiesis, with cell cycle abnormalities being linked to macrocytosis (Humbert et al. 2000 and Von Lindern 2006). These studies provide a plausible explanation as to why mutations in *CDANI* and its associated cell cycling irregularities could lead to the dyserythropoiesis and macrocytosis observed in patients suffering from CDA-I. Table 1.3 shows common benign and pathogenic allelic variants in *CDANI*.

Table 1.3 (Sparks et al. 2005): Allelic variations in the *CDAN1* gene, identifying the amino acid changes.

Variant Classification	DNA Nucleotide Change	Protein Amino Acid Change
Benign	c. 320A>T	p. Gln107Leu
	c. 386G>A	p. Arg129His
	c. 1787A>G	p. Gln596Arg
	c. 2671C>T	p. Arg891Cys
Pathogenic	c. 156C>G	p. Phe52Leu
	c. 2012C>T	p. Pro671Leu
	c. 2150C>T	p. Arg714Trp
	c. 3124C>T	p. Arg1042Trp
	c. 3128A>T	p. Asp1043Val
	c. 3389C>T	p. Pro1130Leu

While mutations in *CDAN1* were observed in most CDA-I cases, approximately 20% of patients did not show any mutations in this gene, suggesting the involvement of another locus. Through whole genome sequencing, a paper published in 2013 by Christian Babbs from the Buckle group in Oxford's MRC Molecular Haematology Unit discovered mutations in *C15ORF41* in a Middle Eastern pedigree (Babbs et al. 2013). This gene is located in chromosome locus 15q14 with 11 exons, and gene expression data indicate that *C15ORF41* is widely transcribed with increased levels in CD34+ cells, cardiomyocytes, and B lymphoblasts, which indicate a requirement in haematopoiesis (Su et al. 2004). Gene expression studies have shown that the expression level of *C15ORF41* is uniform throughout the process of erythropoiesis, which could point to its continual requirement in the maturation of erythrocytes (Merryweather-Clarke et al. 2011).

Babbs *et al* also commented that *C15ORF41* is evolutionarily widely conserved with several orthologs in eukaryotes, archaea, and viruses. Through profile-to-profile comparisons, the group provided evidence that the C15ORF41 protein, which is 281 amino acids long, has two N-terminal AraC/XylS-like HtH domains that are followed by a PD-(D/E)XK nuclease domain. The mutations identified were 281A>G, leading to Y94C missense alteration in the DNA binding domain, and 533T>A causing a L178Q missense mutation in the nuclease domain. L178 is more conserved than Y94, but the latter causes a significant change in the hydrophobicity by switching a tyrosine to a cysteine, which could also form disulfide bonds that disrupt the structural integrity of the protein. The predicted structure of the DNA binding domain of C15ORF41 is shown in Figure 1.7-A, which indicates the putative DNA contacting helices H3 and H6. Part B of the figure shows multiple sequence alignments for the two N-terminal helix-turn-helix domains, with the Y94C mutation indicated in green. Based on the structure, the authors of the paper make the prediction that the C15ORF41 protein is a novel endonuclease, a claim that has been tested in this project and will be elaborated on in the results section.

C150RF43
(161-259)
Model

**PD-(D/E)XK
Nuclease
Domain**

**Active Site
Residues**

C-term

E214

K216

D196

E166

L178

N-term

(A)

E166 **L178Q** **D196** **E214** **K216**

Q9Y2V0 HUMAN 161 **L**CH**E**HE**V**LL**R**DL**L**LE**K**N**L**S**F**LD**L**D**L**R**A**K**G**Y**D**K**T**P**D**F**I**LQ**V**P**V**A**V**E**C**H**I**H**W**I**S**K**A**S**G**D**E**C**S**T**H**A**L**H**D****C**F**W**S**W**N**R**F**C**---P**C**L**V**I**V**Y**W**G-F**I**Q**E**LD**C**N**R** 259
E9HMN8 DAPPU 162 **V**E**Y**E**I**R**L**R**M**E**L**E**K**C**R**V**S**F**L**R**E**H**L**R**V**K**G**Y**D**K**T**P**D**I**K**L**E**I**P**V**A**V**D****C**O**V**I**N****I**S**K**A**L**F**G**D**E**E**V**E**R**G**L**K**C**D**L**Y**S**W**N**R**F**C---P**C**L**V**I**V**Y**W**G-Y**V**E**D**L**E**K**I**H 260
B7F753 IXOSU 157 **L**CH**E**HE**V**Y**L**L**K**Q**L**R**A**L**G**V**A**F**D**E**V**L**R**E**R**G**Y**D**K**T**P**D**V**K**L**E**V**P**V**V**D****T**V**N****I**S**K**A**L**F**G**D**E**S**E**H**G**Y**M**R**D**Q**S**W**N**R**F**C---R**C**L**V**I**V**Y**W**G-Y**V**E**D**V**A**T**G**D 255
D6WU13 TRICA 155 **L**Q**E**Y**G**I**K**L**H**N**E**V**L**K**L**I**A**F**R**D**E**H**L**R**K**F**G**Y**D**K**T**P**D**V**K**L**D**V**P**V**A**V**D****C**V**I**N**I**S**K**A**L**F**G**D**D**V**E**I**E**Y**K**W**N**R**F**C---P**C**L**V**I**V**Y**W**G-F**V**K**S**I**V**Q**A**D 253
Q70613 ANOGA 156 **L**S**E**Y**G**V**R**L**K**M**A**R**D**A**G**M**V**F**Y**D**G**D**L**R**T**G**D**K**T**P**D**K**M**A**V**F**M**Y**R****C**O**V**I**N****I**S**K**A**S**G**D**M**E**S**K**R**V**K**D****C**L**S**S**G**N**R**F**C**---P**C**L**V**I**V**Y**W**G-Y**L**E**T**I**A**D**C**P 254
Q55184 DICDI 184 **T**S**E**Y**I**L**L**Q**E**K**F**K**L**M**T**P**L****S****D**L**R**M**D**Y**G****K**H**P**I**K**L**E**I**P**S**T****C**V**I**N**I**S**K**A**S**G**D****D**N**N**L**K**-I**T**R**T**I**C**I**C****N**R**F**C---P**C**L**V**I**V**Y**W**G-Y**I**D**D**L**N**L**Q** 281
Q9VU06 DROME 172 **V**E**Y**E**G**L**K**L**N**L**K**A**E**A**G**I**H**F**D****K**D**L**R**M**D**Y****K**T**P**D**I**K**M**L**D**F**L**Y**K****S**V**I**N**I**S**K**A**S**G**D****C**K**E**S**N**L**Q****L**Q**S****C**N**R**F**C**---P**C**H**I**Y**W**G-Y**H**E**E**T**P**A**L**P 270
D5E891 METMS 135 **K**I**F**G**M**I**V**A**R**W**L**E**Y**R**N**I**E**Y**L**D**E**L**R**K**Q**A**E**K**T**P**D**F**L**L**P****P**Y**V**E**I**R**C**O**V**N**V**S**K**A**V**G**N**S**T**P**T**O**T**Y**L**E**K****F**S**R**E**E**L**Y**C---S**C**M**V**Y**W**G-Y**V**D**G**I**S**L**E**G 233
F7XK67 METZD 135 **R**K**A**G**S**L**L**E**K**W**L**G**I**R**K**V**O**F**L**D**K**D**L**R**K**R**G**T**S**K**T**P**D**F**L**L**E**N**A**L**E**I**D****C**N**L**I**S****I**W**I**S**K**A**L**F**G**S**D**R**E**F**H**G**H**R**L**R**H**Q**S**R**Y**C---K**M**V**V**Y**W**G-F**L**E**S**I**D**N**R** 233
Q8PUY2 METMA 140 **K**E**M**G**S**E**I**I**R**W**L**D**D**R**E**I**S**Y**C**H**E**A**I**R**A**Q**G**D**K**T**P**D**V**L**S**D**I**P**I**E**D**H**M**V**I**S**K**A**L**F**G**D**F**E**K**H**Y**A**K**K**O**F**R**E**A**D**I**G**C**---E**C**M**V**Y**W**G-Y**L**E**D**L**S**S**E**D 238
F2KPU2 ARCVS 133 **R**K**L**G**S**E**I**I**R**W**L**E**E**R**G**I**E**Y**R**D**K**D**L**R**K**E-F**D**K**T**P**D**F**Y**F**D**E**L**V**D**I**A****C**E**K**V**K**W**I**S**K**A**L**F**G**D**P**R**T****H**W**L**D**V**K**I****K**S**R**E**E**M**F**C---E**C**Y**V**Y**W**G-F**W**G**L**N**R**S**I** 230
O29402 ARCFC 135 **R**K**L**G**S**R**G**L**A**Y**E**L**K**A**G**I**F**L**T****K**D**L**R**K**-F**S**K**T**P**D**F**Y**F**E**E**P**V**E**F**M****C**E**L**K**W**I**S**K**A**L**F**G**D**P**R**S**D**L**V**W**K**K**O**Y**S**K**Y**E**M**F**C**---N**C**L**V**I**V**Y**W**G-C**V**E**S**I**E**A**S**D 232
Q12UW9 METBU 140 **K**Q**M**G**S**D**I**S**K**W**I**D**H**Q**L**G**Y**S**T****E**D**L**R**A**M**G**M**K**T**P**D**F**L**D**K**E**L**Y**D**D**T**E**V**G**W**I**S**K**A**M**G**D**R**E**S**S**H**Y**L**K**K**O**F**O**R**E**R**E**Y**C**---T**O**M**V**I**V**Y**W**G-H**I**D**L**E**F**E**G** 238
D38125 FERPA 133 **R**K**L**G**S**K**I**V**E**R**Y**L**I**A**R**G**V**F**I**S**K**E**I**R**G**---D**H**K**P**D**L**Q**I**Q**E**L**K**I**C**R**I**V**N****I**S**K**A**S**G**D****F**A**E**D**L**K**F**E**K**S**S**E**S**E---E**A**V**L**E**G**F**L**D**V**L**D**K**E** 227
D2RFH1 ARCPA 125 **R**K**L**G**S**K**I**L**E**D**I**M**G**L**D**V**D**F**K****S**E**L**K**T**---R**K**K**T**P**D**F**L**L**K**D**C**I**E**F**L****C**K**V**R**W**I**S**K**A**L**F**G**D**P**R**T**S**F**Y**W**R**R**Y**H**F**K**D**A**G**---K**C**V**V**Y**W**G-H**I**D**T**P**F**S**L**S 220
2wcw_A 4 **K**T**R**P**E**R**D**L**V**E**L**W**K**A**G**F**A**I**R**V**A**G**A**G**S**P**F**P**C**-P**D**I**V**A**G**N-----**C**R**T**Y**L**A**D**V**P**W**K**R**K**E**L**P**L**V**L**A**D**E**V**E**L**V**T****G**A

Figure 1.9 shows the 11 exons in the *C15ORF41* gene as well as the translated protein with its DNA binding and nuclease domains. The red arrows at the bottom indicate the locations and nature of the two missense mutations that have been identified in patients.

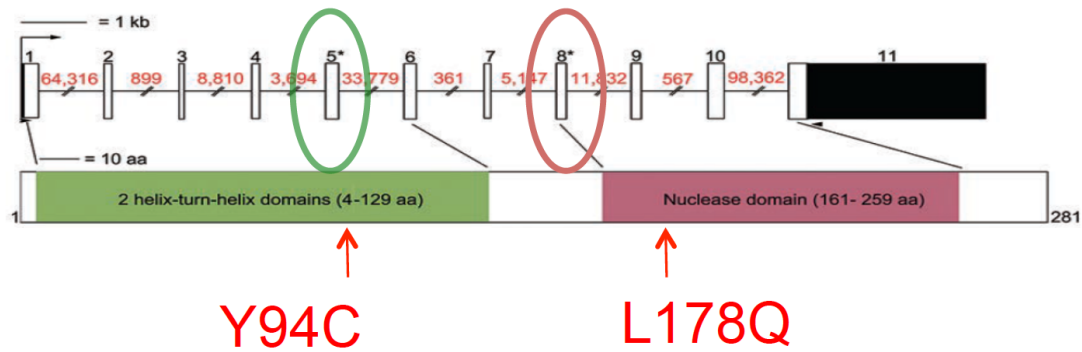


Figure 1.9: Exons of *C15ORF41*, the two HtH domains, and the nuclease domain along with the respective mutations (adapted from Babbs et al. 2013).

In the same paper Babbs provided haematological data for patients with *C15ORF41* mutations, which are shown in Table 1.4. The families and individuals Babbs refers to are indicated in pedigrees in Figure 1.9, taken from the publication (Babbs et al. 2013).

Table 1.4: Haematological data for patients with *C15ORF41* mutations.

Individual	Age at Testing	Hb (g/L)	MCV (fL)	Reticulocytes (10 ⁹ /L or %)	Bilirubin (mM/L)	Ferritin (ng/mL)
Family 1 V-1	13 yrs	90	98	1.5%	68	170
Family 1 V-6	4 yrs	79	84	3%	12	198
Family 1 V-8	1 yr 4 Months	59	79	2.5%	139	-
Family 2 II-2	30 yrs	77	95	61	63	260
Family 2 II-3	23 yrs	103	97	106	-	232
Family 3 II-2	14 yrs	82	96	-	38	377
Family 3 II-3	< 1 yr	58	79	-	-	-

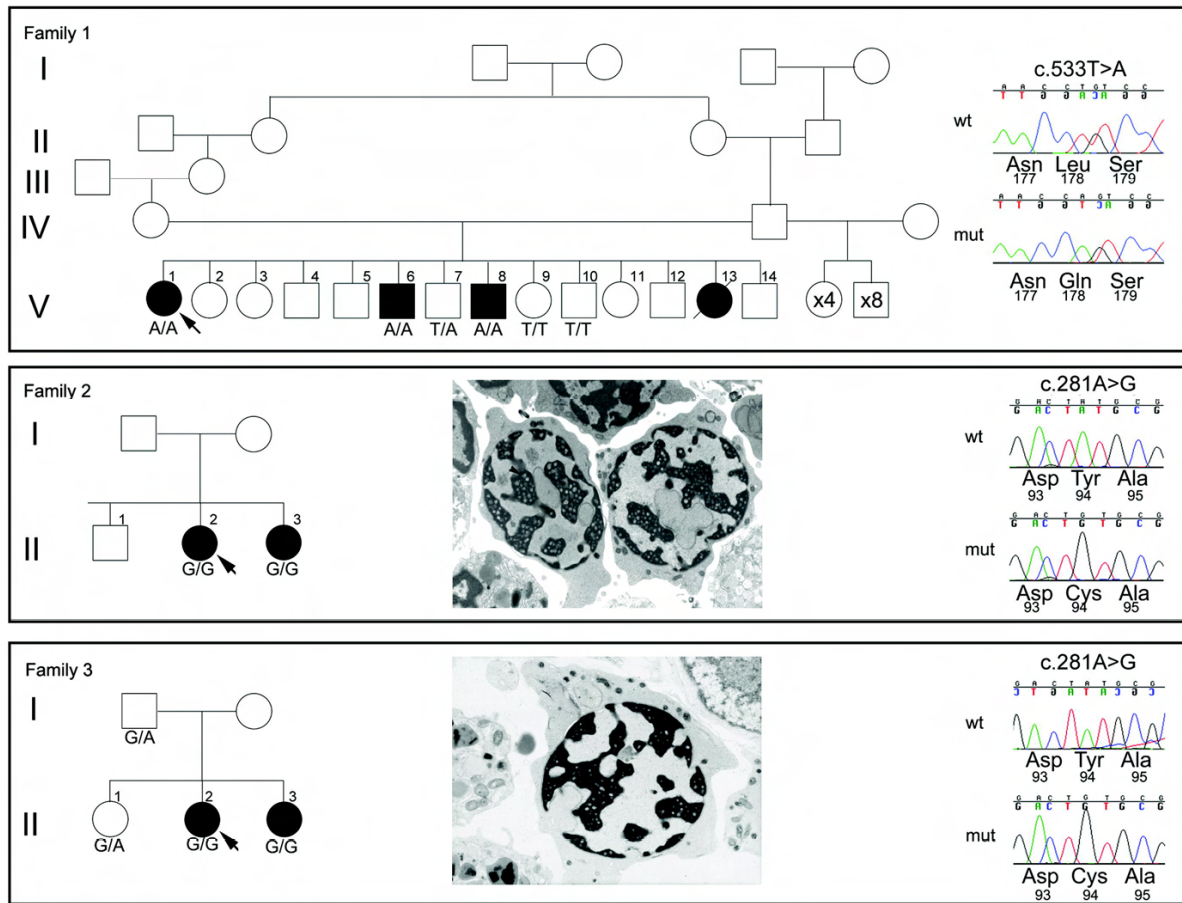


Figure 1.9: Pedigrees for three families identified with the two different C15ORF41 mutations indicated on the right side. EM images show the spongy heterochromatin that is characteristic for CDA-I patients (Babbs et al. 2013).

PD-(D/E)XK motif proteins represent a superfamily of phosphodiesterases that is diverse both structurally and functionally. Proteins in this family are involved in DNA damage repair, Holliday junction resolution, and RNA post-translational modification (Steczkiewicz et al. 2012). The conserved core is made of 4 beta sheets with 2 alpha helices on both ends, forming

a $\alpha\beta\beta\alpha\beta$ topology (Knizewski et al. 2007). Figure 1.10 shows a representation of the catalytic core of the superfamily along with the highly conserved active site residues.

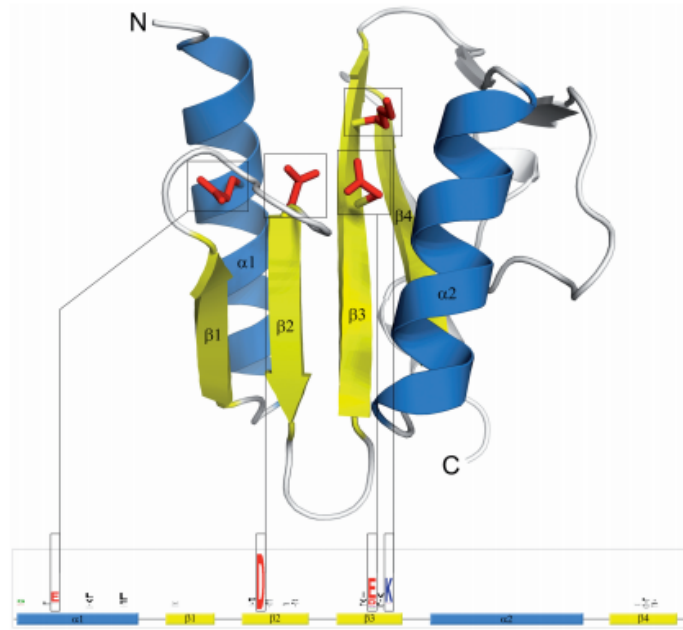


Figure 1.10: Conserved core of the PD-(D/E)XK superfamily with the important active site residues shown in red (Steczkiewicz et al. 2012).

Members of the family include important DNA damage and repair enzymes such as MUS81, XPF^{ERCC4/FANCDQ}, and ERCC1 (Kosinski et al. 2005). A detailed discussion of some of those proteins is included in the next chapter as the topic of DNA damage and repair is tackled.

Part III: DNA Damage and Repair Pathways

1.7 DNA Damage

With approximately 3 billion base pairs in the human genome, there is a high probability that damaging agents can compromise the integrity of our DNA. Both exogenous and endogenous sources of damage affect the genome up to thousands of times a day (Lindahl 1993). Immediate repair of DNA damage is absolutely critical, as impairment of repair has been shown to contribute to premature aging syndromes, congenital defects, neurological disorders, and cancer (Martin 2008). In the case of cancer, the mutations lead to abnormal and rapidly dividing cells that have the potential, through the acquisition of further genetic changes, to metastasise. As a measure to counteract DNA damage, our cells have evolved a number of repair pathways that are deployed to tackle specific types of damage and maintain genomic integrity. The following will contain a discussion of types and sources of DNA damage, as well as various molecular mechanisms cells have to repair them.

1.7.1 Sources of DNA Damage

There are several sources of endogenous damage that could affect our genome. A common one is DNA depurination caused by the hydrolysis of glycosidic bonds, producing apurinic (AP) sites (Loeb and Preston 1986). Cytosines can also be spontaneously hydrolysed in the aqueous cellular environment, which causes a deamination of cytosine into a uracil. Another common source are reactive species that are a by-product of metabolic processes, such as oxygen or hydroxyl radical containing molecules, that can convert guanines into the

mutagenic 8-oxoguanine, induce single strand breaks (SSB), double strand breaks (DSB), or react with cellular metabolites to induce interstrand cross links (ICLs) (Abdullah 2016).

Exogenous sources of DNA damage can be divided into two subgroups: natural and induced. Natural sources include ultraviolet (UV) and ionizing (IR) radiations. Solar UV radiation can lead to the formation of covalent bonds between pyrimidine bases to form bulky products, such as 6-4 pyrimidine photoproducts (6-4PP) and cyclobutane pyrimidine dimers (CPD), that disrupt replication and translation (Sinha and Hader 2002). IR is known to introduce double-strand breaks, which are detrimental to cell function (Coin et al. 2016). Induced damage can be achieved using chemotherapeutic agents such as cisplatin that utilize the rapid division of cancerous cells in order to selectively damage them, as well as therapeutically applied radiation.

1.8 DNA Repair Mechanisms

As discussed earlier, it is vital that cells quickly identify and repair damaged DNA, as that is important for the avoidance of cancer development or abnormalities in organ development and function. The following sections will discuss a select group of repair pathways that are pertinent to the selected members of the PD-(D/E)XK superfamily of nucleases mentioned in section 1.6.3.

1.8.1 Nucleotide Excision Repair

Nucleotide Excision Repair (NER) is involved in resolving damage that leads to bulky, helix distorting lesions, such as CPDs from UV damage mentioned earlier. NER deficiency leads to

the inherited disorder Xeroderma Pigmentosum (XP), as well as involvement in Cockayne Syndrome (CS) and Trichothiodystrophy (TTD) (Abdullah 2016). This mechanism is active during all phases of the cell cycle and it has two pathways through which it can function: global genome NER (GG-NER) and transcription-coupled NER (TC-NER). The difference is that TC-NER is only initiated when lesions on actively transcribed DNA cause a block to the progression of RNA polymerase II (RNAPII). The only difference between these two NER pathways is in the recognition procedure, with everything that is downstream being identical. Depending on the degree of distortion, GG-NER either recognizes the distortion through XPC/hHRad23B (Xeroderma Pigmentosum/human homolog of Rad23B) for major distorting lesions or in conjunction with UV-DDB1/2 (UV-Damaged DNA Binding Proteins 1 and 2) for more minor distortions like the ones caused by CPDs (Petruseva et al. 2014). As alluded to earlier, in TC-NER the recognition is done by RNAPII stalling, which then recruits Cockayne Syndrome group A and group B proteins (CSA/CSB) (Moser et al. 2016), mutated in CS.

After recognition occurs, both paths converge, where a non-canonical form of the transcription factor IIIH (TFIIH) complex is recruited through XPC-hHRad23B. The XPD subunit of the complex unwinds the DNA in a 3' to 5' directionality, creating a 'bubble', with XPD unwinding DNA around the lesion with the opposite polarity. This facilitates the stable association of the pre-incision complex composed of XPA, XPG, and replication protein A (RPA). XPF-ERCC1 and XPG are guided by RPA, bound to the single stranded DNA (ssDNA) region within the bubble. XPF-ERCC1 and XPG incise the lesion, a damage-containing oligonucleotide is released, and the resulting gap is then filled by a DNA polymerase (δ , ϵ or κ) with proliferating cell nuclear antigen (PCNA). LIG1 (ligase 1) or

LIG3 (ligase 3) then completes the processes by sealing the nick (Edifizi and Schumacher 2015). The entire process is summarized schematically in Figure 1.11.

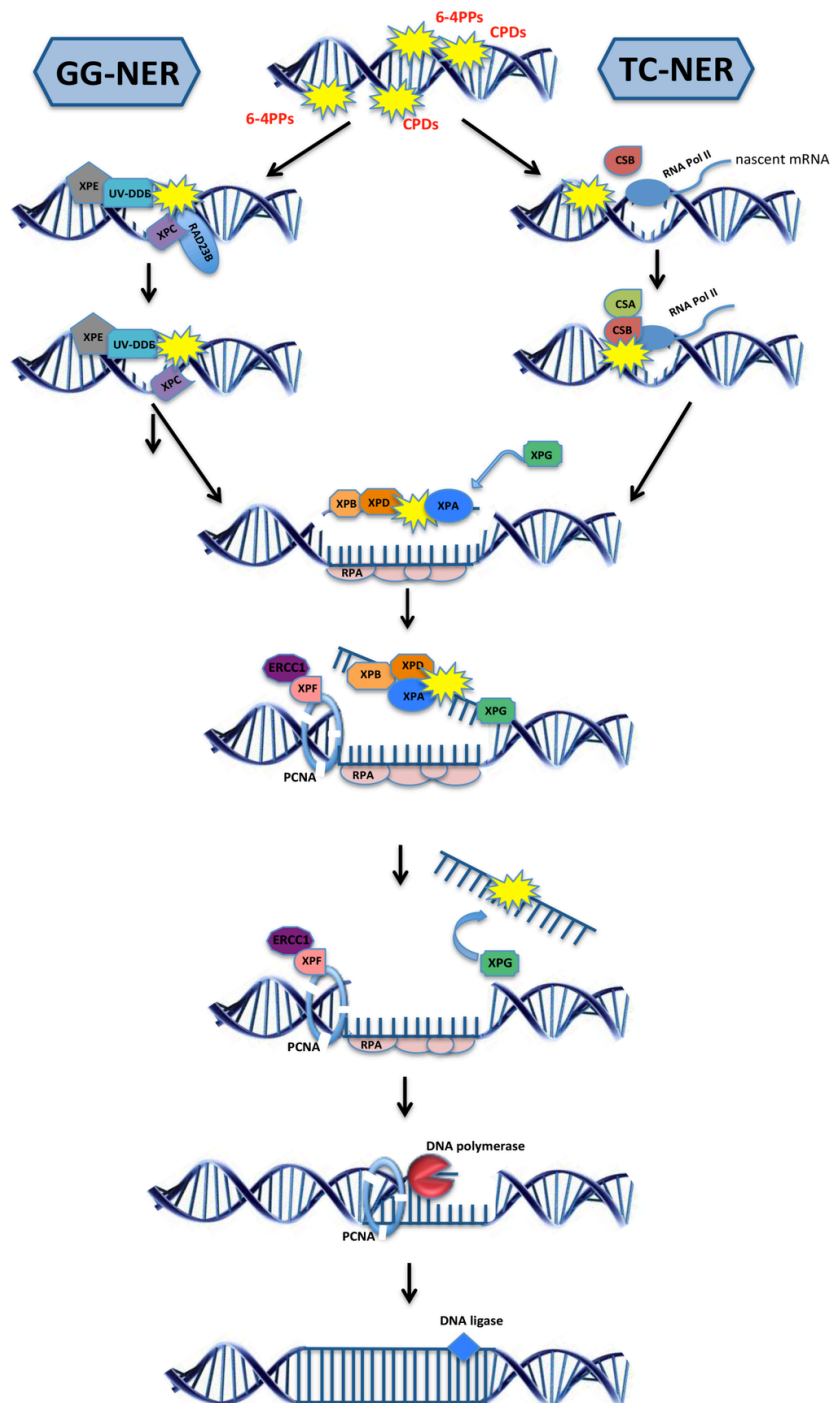


Figure 1.11: Mechanistic details of GG-NER and TC-NER as described in section 1.8.1 (Rothkamm et al. 2003).

1.8.2 Double Strand Break Repair

DSBs are arguably the most toxic damage that can occur to DNA, as they can lead to permanent cell-cycle arrest, apoptosis, and impairment of chromosome function and architecture (Iyama and Wilson 2013). There are two separate major pathways that cells utilize to repair DSBs: non-homologous end-joining (NHEJ) and homologous recombination (HR). NHEJ operates in all cellular phases, while HR is specific to late S and G₂ phases as it requires the presence of a sister chromatid for repair (Lee 2016).

1.8.2.1 Non-Homologous End Joining

NHEJ operates by direct ligation of the broken strands, and is initiated via the recruitment of the Ku70/Ku80 complex to both broken ends. This then allows for the action of DNA-Protein kinase catalytic subunit (DNA-PKcs), which associates with the Ku proteins in an event that initiates its kinase activity, forming DNA-PK. DNA-PK can activate Artemis, an endonuclease that can process damaged break termini to yield the 3'-hydroxyl and 5'-phosphate in order for DNA polymerases to fill the gaps and ligation to occur through the X-ray cross complementation group 4-ligase 4 (XRCC4-LIG4) complex (Abdulla 2016). NHEJ is error prone as the DNA ligation can occur in the incorrect orientation or the sequence might have been changed as a result of processing required for end joining.

1.8.2.2 Homologous Recombination

HR is a much more accurate method for processing DSBs. It requires a sister chromatid that will be used as a template for rebuilding the lost genetic information. The Mre11/Rad50/Nbs1 (MRN) complex recognizes the broken ends, which then allows the recruitment of CtIP (C-

terminal binding protein (CtBP) interacting protein). CtIP with Mre11 initiates an exonucleolytic 5' to 3' resection of the termini, producing 3' ssDNA overhangs that are coated with RPA. Rad51 then replaces RPA on these overhangs, facilitating the homologous sister chromatid invasion that creates a D-loop. BRCA2 promotes the loading and strand-invasion capabilities of Rad51 'recombinase'. Following the capture of the second DSB end, a Holliday junction (HJ) may form. If Holliday junction (HJ) architecture is not formed, the synthesized strand dissociates and re-anneals to its original partner strand, in a process known as synthesis strand-dependent annealing (SSDA), and non-cross over products are formed. If a HJ is formed, it can be processed in ways that lead to crossover and non-cross over products. Dissolution of the junction via the BLM-TOPIII-RMI1-RMI2 complex results in non-cross over products. Resolution, on the other hand, of the junction can be conducted via endonucleases such as GEN1, MUS81-EME1, and SLX4-SLX1. Incision of the DHJ in the same strand orientation results in non-crossover products, while incision in different orientations produces crossover products, which also occurs in meiotic division (Abdullah 2016). Both NHEJ and HR are summarized in Figure 1.12.

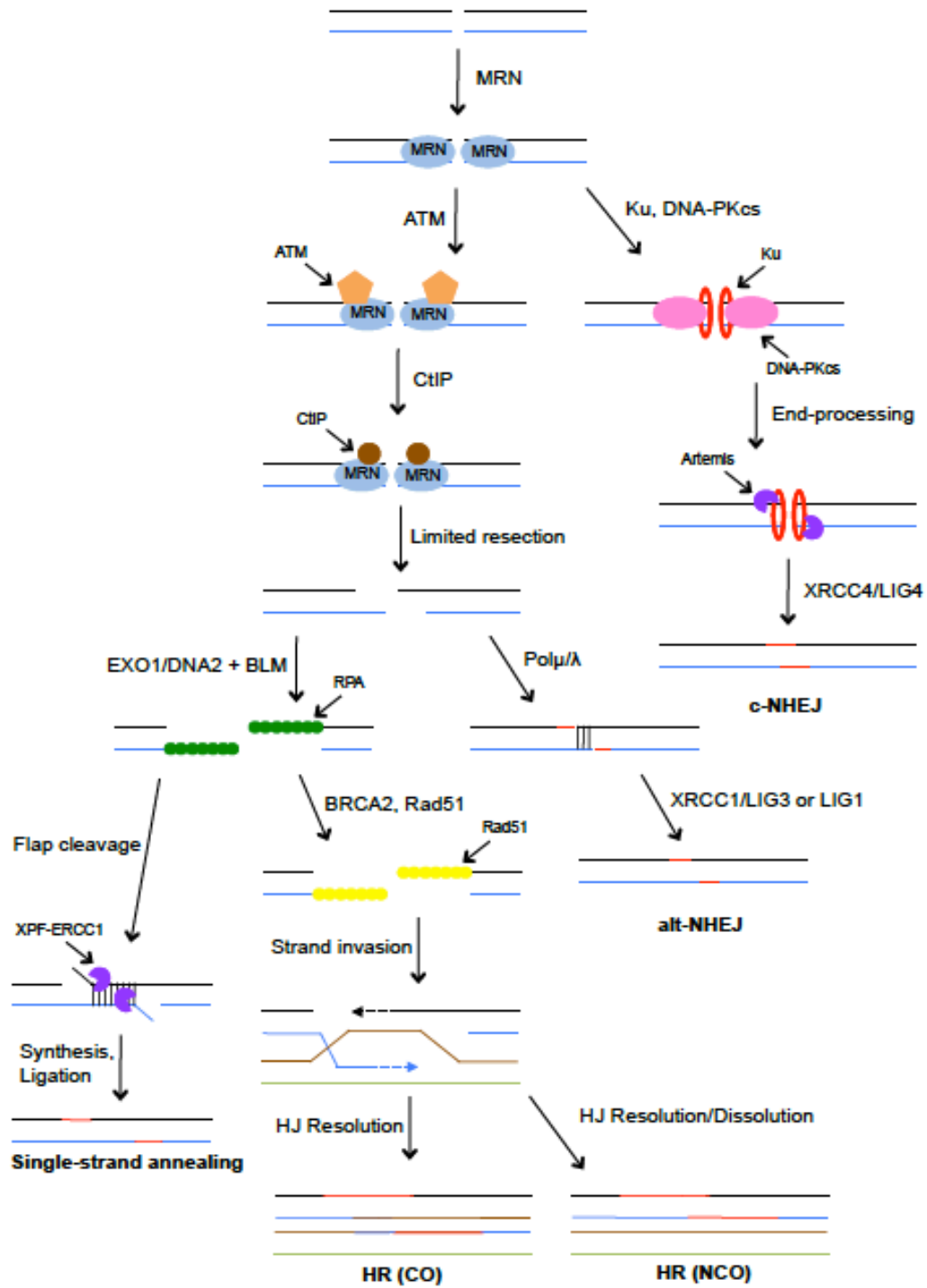


Figure 1.12: Schematic summarizing the key steps in NHEJ and HR (Lee 2016).

1.8.3 Interstrand Crosslink Repair

Interstrand crosslinks (ICLs) are covalent bonds between the two strands of DNA, and accordingly are extremely toxic as replication and transcription are interrupted (Raschle et al. 2008). There are both endogenous and exogenous sources of ICLs. Endogenous sources include reactive aldehydes from the metabolism of alcohol and lipids, while exogenous sources include chemotherapeutic agents like cisplatin, carboplatin, and mitomycin C. Fanconi anaemia patients are extremely sensitive to ICL-inducing agents, as are cells derived from affected individuals, but are also prone to cancer development, and this appears to be due to defects in a key ICL repair pathway described later.

ICL repair in mammalian cells is quite complex and can occur in a replication-dependent or independent fashion, where the replication-dependent pathway is by far best understood. In a simplified sense, during replication-coupled repair the ICL is detected via the FA core complex consisting of the 8 units: FANCA, B, C, E, F, G, L, and M by the association of FANCM with MHF (MPH1-associated histone-fold protein). FANCM-MHF complex exposes ssDNA by promoting branch migration of the fork, allowing RPA to bind the single-stranded regions. This is then followed by FANCL monoubiquitinating the FANCI-FANCD2 complex after it has been phosphorylated by ataxia telangiectasia and Rad3-related protein (ATR) kinase on chromatin. This allows the recruitment of FANCD1 (also known as XPF) and FANCD2, and possibly other nucleases such as SLX4-SLX1, MUS81-EME1, FANCD1, and SNM1A, which create the incisions for ICL unhooking. SLX4 appears to be involved by providing a scaffold for XPF-ERCC1 and MUS81-EME1. Evidence for the involvement of

these nucleases is supported by the increased sensitivity of depleted cells to ICL agents. The unhooking process produces a DSB in the S/G₂ pathway.

After unhooking, translesion synthesis (TLS) polymerases are required to extend past the ICL. REV3/REV7 and REV1 are two important TLS polymerases involved in mammalian cells. Since the unhooking process produces a DSB, the repair process then proceeds through homologous recombination (HR) with the aid of units such as Rad51, BRCA1, BRCA2, PCNA, and RPA. After the resolution of the ICL damage, the FA core complex dissociates, allowing deubiquitinating enzyme ubiquitin specific peptidase 1 (USP1) to deubiquitinate FANCI/FANCD2, which completes the repair (Abdullah 2016). This model is summarized in Figure 1.13.

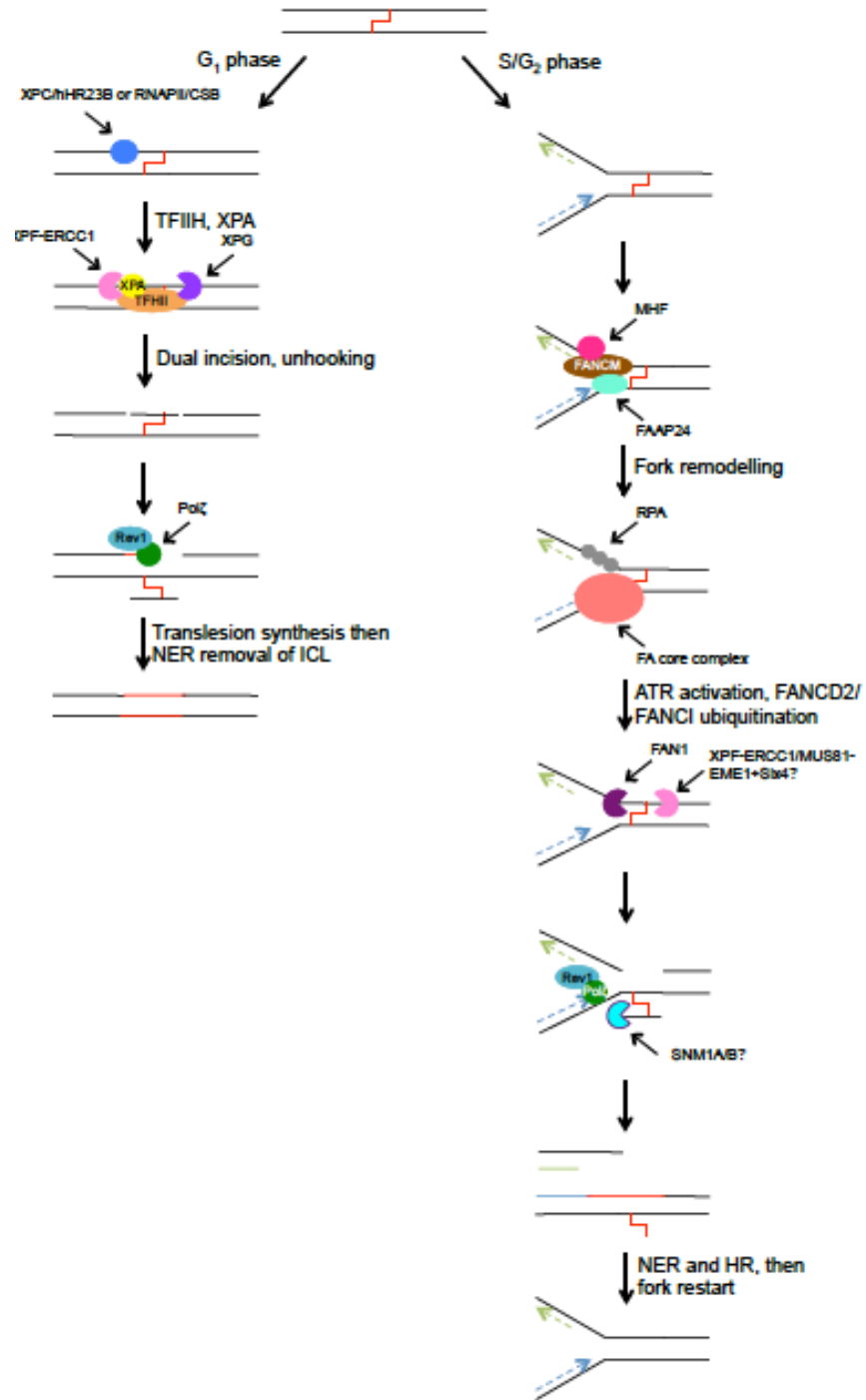


Figure 1.13: Diagram summarizing ICL repair pathway in mammalian cells

(Lee 2016).

1.9 Aims of the Project

The publication by Babbs *et al* has provided convincing arguments for the involvement of *C15ORF41* mutations in CDA type I. It is likely that these would account for a proportion of the cases in which patients with CDA-I had no mutations in the *CDANI* gene. Based on modelling data and by conducting spatial comparisons, the authors predicted that the protein is part of the PD-(D/E)XK superfamily of nucleases, and accordingly would be involved in processes like DNA damage and repair, RNA processing, or Holliday Junction resolution. The Structural Genomics Consortium (SGC) at Oxford managed to purify a complex of the full-length C15ORF41 protein along with the C-terminus of Codanin-1, the reasons of which will be elaborated on in the results and discussion section, which made it available for experiments. Therefore, the aims of this project are:

- 1) Providing a biochemical characterization of the C15ORF41/Codanin-1 complex by conducting a variety of nuclease assays with a number of different DNA and RNA substrate architectures.
- 2) Investigate some of the kinetic properties of the complex as well as its substrate binding behaviour.
- 3) Provide a starting point for future cell biology experiments to be conducted such that the role of this complex in cellular processes can be evaluated. This could then act as a starting point for studying whether C15ORF41 is involved in DNA damage and repair pathways.

2. Materials and Methods

2.1 Protein Purification

All the protein samples used in this project were purified by Hazel Aitkenhead of the Structural Genomics Consortium (SGC) at the University of Oxford. It should be noted that 3 mutations were made to the C15ORF41 structure in order to increase stabilization for enhanced purification: W237R, C223S, and V18M. The protein was expressed in BL21 (DE3) Rosetta-R3 *E. coli* cells with an N-terminal His₆ tag – TEV cleavage site (indicated by an asterisk): MHHHHHHSSGVDLG TENLYFQ*SM-[continuation]. A brief summary of the protocol followed is included below.

An initial culture was made by inoculating cell from the Rosetta stock plates into Terrific Broth (TB), containing kanamycin and chloramphenicol, followed by overnight growth in an orbital shaker at 37°C. This starter culture was placed in 2 litres of 1X kanamycin for C15ORF41 constructs and 1 litre of 1X kanamycin for CDAN1. Those were then shaken at 37°C until the OD₆₀₀ reached a value between 2 and 3. After centrifugation at 4000g for 25 minutes, the pellets were re-suspended in lysis buffer (50mM HEPES pH7.5, 500mM NaCl, 10mM Imidazole, 5% glycerol, 1mM TECP), and then pulse sonication was performed (5s pulses separated by 10s pauses) for 25 minutes, with stirring every 5 minutes.

The proteins were purified by immobilising with metal affinity chromatography followed by gel filtration. For single protein purifications, only the CDAN1 tag was cleaved. For the complex, metal affinity chromatography was conducted for CDAN1 (with cleavage of the

tag), and anion exchange was conducted for C15ORF41. The proteins were then mixed with a 1:10 molar ratio for CDAN1 to C15ORF41 and passed through a gel filtration column.

2.2 Substrate Labelling

A range of DNA and RNA substrate types have been used in this project, such as radiolabeled and fluorescent oligos. Table 2.1 provides detailed descriptions and sequences of all the substrates used.

2.2.1 α (3') Radiolabeling of Oligos

5 μ L of 10 μ M (50pmol) single-stranded DNA (ssDNA) substrate (Eurofins MWG Operon, Ebersberg, Germany) was labelled with 3.3pmol of [α -³²P]-dATP (Perkin Elmer) using terminal deoxynucleotidyl transferase (TdT) (20U, Thermo Scientific). The mixture was incubated at 37°C for 1 hour. To get rid of unlabelled radiation, the mixture was filtered through a P-6 spin column (Bio-Rad) at 1000-rcf for 4 minutes. Finally, the labelled substrate was diluted to 500nM using nuclease free water.

2.2.2 γ (5') Radiolabeling of Oligos

10 μ L of 1 μ M (10pmol) ssDNA substrate was labelled with 6.8pmol [γ -³²P]-ATP with T4 Polynucleotide Kinase (T4 PNK). The mixture was incubated at 37°C for 1 hour, run through a P-6 spin column at 1000-rcf for 4 minutes to get rid of unbound label, and then eluted in concentrations of 500nM. RNA samples were prepared in the same way.

2.2.3 Substrate Annealing

10 μ L of the 500nM radiolabeled oligo was incubated with 1.5 μ L of 10 μ M cold oligo, along with 2.5 μ L of 10X annealing buffer (10mM Tris pH 7.5, 50mM NaCl, 1mM EDTA) and 33.5 μ L of nuclease free water, and heated up to 95°C for 6 minutes before slowly cooling down to room temperature. Final concentration of annealed product was 100nM.

2.3 Radioactive Nuclease Assay

The gel-based nuclease assays were carried out in a 10 μ L mixture made by adding 1 μ L of 100nM radioactive substrate (10nM final concentration), 2 μ L of 10X nuclease buffer (50mM HEPES – pH 7.5, 500mM NaCl, 5% glycerol, 0.5mM dithiothreitol (DTT), 200mM MgCl₂) 6 μ L of nuclease free water, and 1 μ L of 5 μ M protein substrate (final concentration of 500nM), unless stated otherwise by the specific reaction. Controls had the same composition except the protein is not added and hence 7 μ L of water were added instead.

The reactions were conducted at 37°C for the indicated amount of time, and then quenched by adding 2 μ L of stop buffer (80% formamide, 1mg/mL bromophenol blue, 1 mg/mL xylene cyanol, 10mM EDTA pH 8.0) and transferring the reaction tube to a heat block at 95°C for 4 minutes. After allowing the samples to cool to room temperature, they were run on a 20% denaturing polyacrylamide gel for 1.5 hours at 525 volts in a 1X TBE solution. The gels were composed of 40% 19:1 acrylamide:bis-acrylamide with 7M urea and 1X TBE. Afterwards, the gels were fixed for 1 hour (50% methanol, 10% glacial acetic acid) on a rocker and dried at 80°C for 2 hours or at 50°C for 2 hours while under constant vacuum. Finally, the gels were placed on a phosphorimager screen and scanned (General Electric Typhoon 9400).

Whenever substrates were first annealed, a 10% native (non-denaturing) gel was run to check whether the annealing process had occurred properly. This was done by making sure only one band appeared at the correct point in the gel when compared to a reference marker. The gel was made by mixing 6.25mL of 40% acrylamide (19:1 acrylamide:bis) with 1.25mL 10X TBE, and 17.5mL water. Then 13.5μL TEMED was added to the solution along with 220μL of 10% APS. After the gel was cast and the samples loaded, it was run at 400V for 1 hour, placed on a phosphorimager screen and scanned using the Typhoon 9400.

2.4 Electrophoretic Mobility Shift Assay (EMSA)

A non-denaturing gel was cast (5% 19:1 acrylamide:bis, 0.5X TBE), and pre-cooled in the cold room (4°C) overnight. 10nM of substrate were added to 2μL of nuclease buffer (50mM HEPES – pH 7.5, 500mM NaCl, 5% glycerol, 0.5mM dithiothreitol (DTT), and 200mM MgCl₂ when indicated – some reactions did not include Mg⁺² ions), and protein at a concentration of 500nM. The reaction was incubated at 4°C for 15 minutes before being quenched with 2μL ice-cold stop buffer (50% glycerol, 1mg/mL bromophenol blue). Samples were added to the gel and run in the cold room at 100V for 2 hours, and then fixed (50% methanol, 10% glacial acetic acid) for 10 minutes. Afterwards, the gel was dried for 2 hours at 80°C under vacuum, and then placed on a phosphorimager screen and finally scanned using the Typhoon 9400.

2.5 Plasmid Assay

Three different forms of the pUC18-shortGAP46 plasmid substrate were used in this assay: supercoiled (S), linearized (L), and nicked (N). Plasmid was nicked with NbBbvC1 and

NEBuffer2 (BioLabs) and linearized with HindIII and NEBuffer2. Then 20nM of the desired pG64 plasmid was incubated with 500nM of protein for 1 hour at 37°C and nuclease buffer (50mM HEPES – pH 7.5, 500mM NaCl, 5% glycerol, 0.5mM dithiothreitol (DTT), 200mM MgCl₂). A 1% agarose gel was made by adding 1.5g agarose to 150mL TAE and 12.5μL ethidium bromide. The samples were loaded in the gel and run at 130V for 2 hours.

2.6 Fluorescence Assay

Fluorescence reactions were run in the same conditions as radioactive gels. The difference is that these assays had a fluorescently labelled substrate instead of a radiolabeled one. Samples were also run in a 20% acrylamide gel (see section 2.3 for details). Gel was scanned using the Typhoon 9400.

2.7 Fluorescence Anisotropy

Fluorescence anisotropy experiments were conducted with the fluorescein moiety containing BS20 substrate (refer to Table 2.1). A final concentration of 10nM substrate was added to buffer (50mM HEPES – pH 7.5, 500mM NaCl, 5% glycerol, 0.5mM dithiothreitol (DTT), 200mM MgCl₂) in black 384 well plates, with a final volume of 15μL. Unless otherwise stated, a serial dilution was conducted for the protein starting at 1μM. 15μL of the protein samples were then added to the substrate/buffer mixture and the plates were centrifuged at 3000rpm for a few second to get rid of bubbles. The samples were then immediately read on a PHERAstar machine (BMG LabTech).

Table 2.1: Lists the names and sequences of all the substrates used in this project.

Substrate Name	Sequence
BS20	5' P-ATA ATT TGA TCA [FluoT]CT ATT ATA-OH 3'
BS49	5' ATA ATT TGA TCA TCT ATT ATA – [FLOU] 3'
BS76	5' P – [FLOU-T]TA ATT TGA TCA TCT ATT AT 3'
BS40	5' ATA AAT ATT TTT TAT TAA TAA TAG ATC ACC TTT CTT TCT CTT CTC CCC TT 3'
BS40-B	5' ATA AAT ATT TTT TAT TAA TAA TAG ATC ACC TTT CTT TCT CTT CTC CCC TT [B] 3'
BS41	5' TTC CCC TCC TCT CCT TCC TTC CTG ATC TAT TAT TAA TAA AAA ATA TTT AT 3'
UA1	5' AAG GGG AGA AGA GAA AGA AAG GTG ATC TAT TAT TAA TAA AAA ATA TTT AT 3'
UA2	5' AAG GGG AGA AGA GAA AGA AAG G 3'
UA3	5' ATT ATT AAT AAA AAA TAT TTA T 3'
UA6	5' AAG GGG AGG AGA GGA AGG AAG G 3'
HJ1	5' GAC GCT GCC GAA TTC TTC TGG ATT GCT AGG ACA TCT ATT AAG CTT ACG TTA ACC CA 3'
HJ2	5' AAG AAA TTG CAT CTA GAG GGT CAT GTC CTA GCA ATC CAG AAG AAT TCG GCA GCG TC 3'
HJ3	5' AGG CTC ATA ATG ATA TCT CGT CTT GCT AGG ACA TGA CCC TCT AGA TGC AAT TTC TT 3'
HJ4	5' TGG GTT AAC GTA AGC TTA ATA GAT GTC CTA GCA AGA CGA GAT ATC ATT ATG AGC CT 3'
Bubble Top	5' TCC AGC GCT CGG TTT TTT TTT TTT TTT TTT TTT GGC TCG CGA CCT 3'

Bubble Bottom	5' AGG TCG CGA GCC TTT TTT TTT TTT TTT TTT TTT CCG AGC GCT GGA 3'
ICL 1	5'[B]TTCCCCTCCTCTCCTTCCTTCCTGATCTATTATTAATAAAAAATATTTAT- P 3' 3' TTCCCCTCTTCTCTTTCTTTCC <u>A</u> CTAGATAATAATTATTTTTTATAAATA 5'
BS51	5' AUA AUU UGA UCA UCU AUU AU 3'

[B] = biotin, P = Phosphate

3. Results: Biochemical Behaviour of C15ORF41 Complex

3.1 Protein Purification

As indicated before, all the purifications were conducted by Hazel Aitkenhead from the SGC. The mass spectrometry data from those purifications are presented in this section. The predicted mass of the C15ORF41 protein including the tag is 34802.7 Da, while that for tag-cleaved C-terminus of Codanin1 is 39910.1 Da. Figure 3.1 shows multiple purifications for the complex, with the first row in red being selected for the biochemical assays presented later in this chapter.

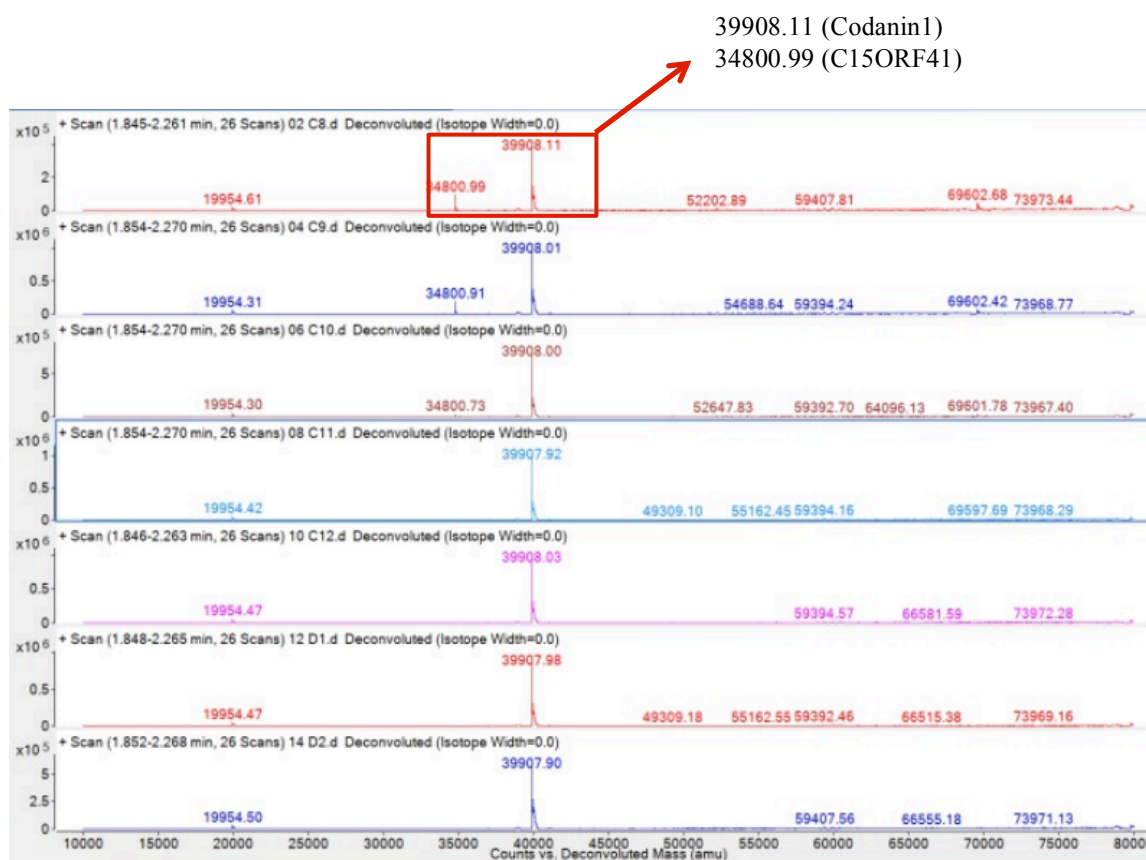


Figure 3.1: Mass spectrometry data for various purifications of the C15ORF41-Codanin1 complex. The first sample was chosen for later

experiments as it contained the expected masses for both proteins shown in the red box.

3.2 Agarose Plasmid Substrate Assay

As mentioned in the introduction, Babbs *et al* proposed the hypothesis that C15ORF41 is probably an endonuclease, as apparent from in silico analysis of the protein sequence and predicted structure. In order to examine this hypothesis, the protein had to be tested with a range of nucleic acid substrates. First, the protein complex (C15ORF41/Codanin) from the purification indicated in section 3.1 was incubated for 1 hour at 37°C with a DNA plasmid (pUC18-shortGAP46, 2852 base pairs long) and 10X nuclease buffer (50mM HEPES – pH 7.5, 500mM NaCl, 5% glycerol, 0.5mM dithiothreitol (DTT), 200mM MgCl₂). Three different forms of the pUC18-shortGAP46 plasmid were used: supercoiled (S), nicked (N, via NbBbvC1), and linearized (L, via HindIII). The concentration of the DNA substrates was 20nM and that of the protein was 500nM. A 1% agarose gel stained with ethidium bromide was used to visualize the results.

The results of a time course experiment are shown in Figure 3.2. There does not appear to be any activity on the nicked or linearized plasmids, as the control lanes (the ones with the negative (-) sign) are not substantially different than the ones with the protein (positive lanes). However, looking at the supercoiled lanes in all time points, the supercoiled band is replaced by a second form of the molecule, a transition from ~ 1800 up to ~ 2700 bp. This indicates that the protein has linearized the supercoiled DNA, which would require two nicks on both strands of the plasmid. As a result, the mobility of the plasmid is affected, with supercoiled

plasmids having greater mobility through the gel than linearized forms. The strength of the supercoiled signal increases with time, from 30 minutes to 180 minutes. This is expected, as the protein would have interacted with more substrates.

No specific information on the behaviour of the protein is obtained from this assay, however, it gave an indication that the complex could both interact with and metabolise DNA substrates. This acted as the stepping-stone for further experiments with substrates with defined features.

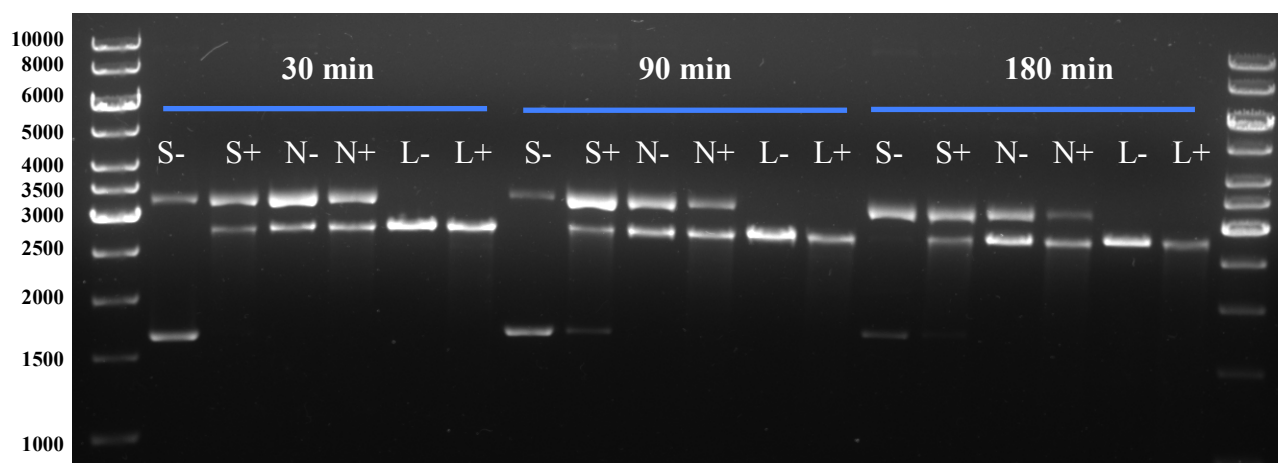


Figure 3.2: Plasmid assay of the protein complex incubated with the DNA plasmid (S = supercoiled, N = nicked, L = linearized). The negative (-) lanes are controls that do not contain the protein, and the positive (+) lanes are with the protein in the mixture. The bands in the positive lanes for the supercoiled match those of the nicked plasmid, indicating the protein introduced nicks to the plasmid.

3.3 Fluorescent DNA Substrates

After the indication that C15ORF41-Codanin complex has some activity on DNA plasmid substrates, assays with fluorescently labelled substrates were conducted to delve deeper into the function of the protein. A range of divalent metal ions was tested (Mn^{+2} , Mg^{+2} , Ca^{+2} , and Zn^{+2} in Figure 3.3 and Ni^{+2} in Figure 3.4) in the nuclease buffer. As discussed in section 1.6.3, the protein is part of the PD-(D/E)XK superfamily and as such is predicted to share the common feature of being active with divalent ions. The substrate was a simple 20-mer ssDNA oligonucleotide, with a fluorescein moiety on the 3' end. The assay was carried out at 37°C for 1 hour with the same nuclease buffer used in the previous plasmid experiment.

Figure 3.3 shows bands consistent with substrate digestion in the presence of Mn^{+2} and Mg^{+2} ions as indicated with the arrows, with increasing intensity as the concentration of the ions goes up. The strength of the signal is only slightly stronger for manganese than magnesium. A different pattern appears with Zn, in which a weak signal shows only for the lower concentrations (0.01 and 0.1 mM). This could indicate some sort of inhibitory action for the higher zinc concentrations. Figure 3.4 shows the same assay but with nickel ions in the nuclease buffer. No products are generated for any of the concentrations tested with Ni^{+2} .

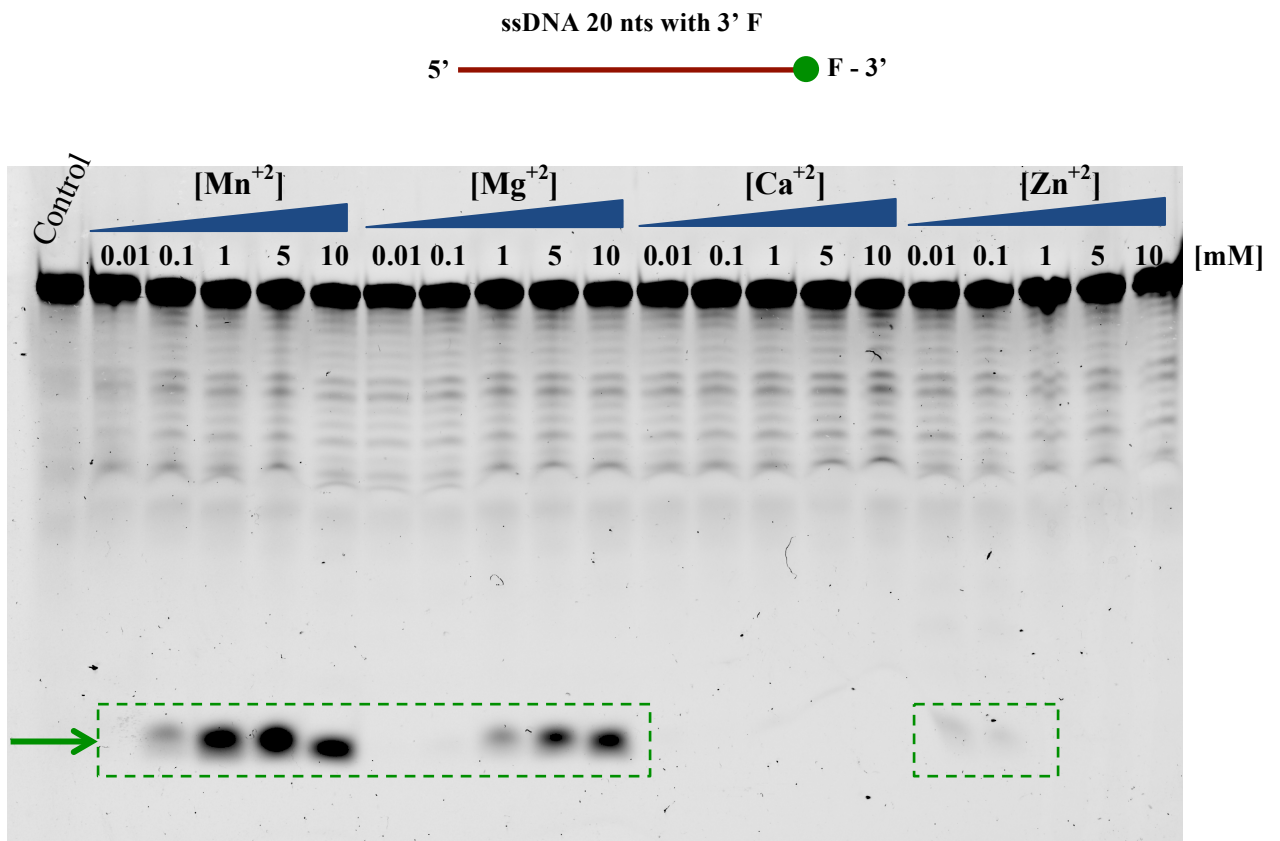


Figure 3.3: 20% acrylamide nuclease gel showing the behaviour of the protein with a range of divalent metal ions (concentration in mM) on a 20-mer ssDNA substrate (BS49 in Table 2.1). [Protein] = 500nM, [substrate] = 200nM.

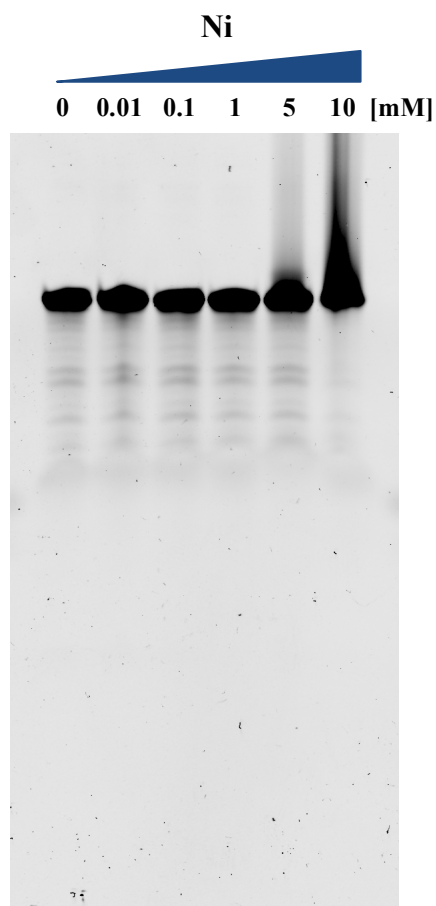


Figure 3.4: 20% acrylamide gel showing of the result of protein incubated for an hour at 37°C with 20-mer fluorescently labelled ssDNA substrate (BS49). Ion concentrations in mM, [protein] = 500nM, and [substrate] = 200nM.

The results from the two acrylamide gels shown before suggest that the protein is cleaving the 20-mer fluorescent substrate in the presence of Mn^{+2} or Mg^{+2} ions, which results in the products shown. Molecular weight markers were not included in these pilot experiments. The small size of the cleaved product suggests the incision occurs close to the 3' end of the substrate. To gain further insight, another assay was performed, this time using 5' and 3' fluorescently labelled substrates with only magnesium and manganese ions. EDTA was also

introduced in certain reactions to chelate the metal ions in order to confirm the ionic dependence.

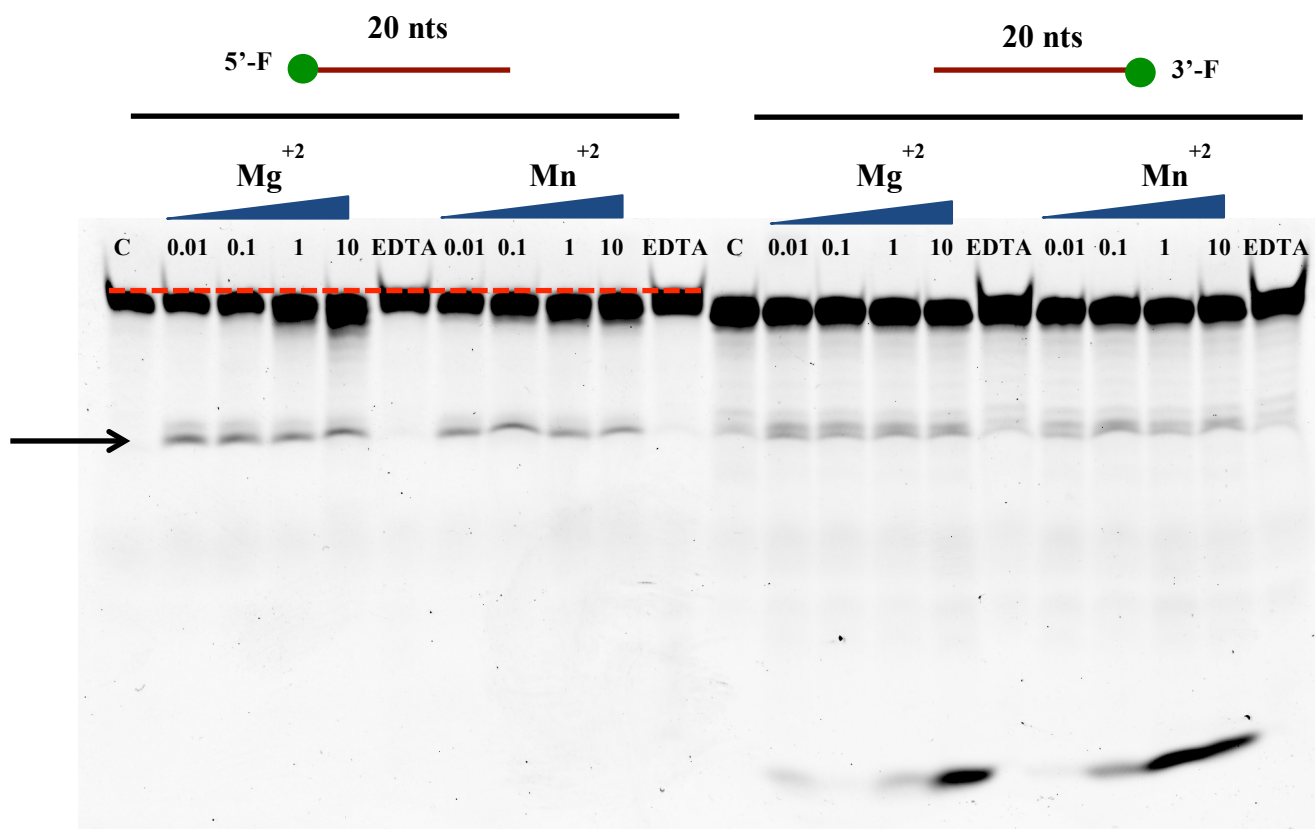


Figure 3.5: Assay of a 20% acrylamide gel with 5' and 3' fluorescently labelled 20-mer ssDNA substrates (BS76 and BS49). [Protein] = 500nM, [substrate] = 100nM, [EDTA] = 10mM, and ion concentrations are in mM.

The same result is obtained for the 3' labelled substrates as in the previous gel (Figure 3.3). As the concentration of the divalent ions increases, a greater quantity of product is released as evident by the strengthening in the band signals. The cleavage appears to be happening a few nucleotides from the 3' end of the substrate. The curvature observed in the signals is due to edge effects as the reaction mixtures run through the gel. EDTA reactions do not show any product, which confirms the necessity of the divalent ions in the protein carrying out its

catalytic function. The 5' labelled substrate reactions, shown on the left hand side of the gel, do not have any signal at the bottom of the gel, in contrast to the 3' labelled substrate. It does appear, however, that there is cleavage occurring on the 3' end, as the bands in the non-control lanes are lower suggesting a lower molecular weight product. The dotted red line on the gel show how the control and EDTA lanes run at the same location, suggesting that the ions are an important cofactor in the function of the protein. The same pattern is not observed in the 3' labelled substrates on the right hand side of the gel. The bands indicated by the black arrow in the figure are extremely weak, and are approximately (1-2)% the strength of the original control bands as measured using ImageJ. They could be due to some form of minor cleavage occurring in the middle of the substrate.

An important thing to point out at, however, is that Figure 3.3 does not exclude the possibility of the protein cleaving the fluorescent tag (glycosidic activity) without acting on the DNA substrate. Therefore, to be able to confirm the exact behaviour and exclude the possibility of the Fluor being cleaved, radiolabeled substrates were used.

3.4 Radioactive Substrate Assays

The design of the different substrate architectures used is detailed in section 2.2.4. Before using the substrates in any experiments, it is crucial to run a native (non-denaturing) gel to check whether strand annealing took place and was quantitative. Depending on the size and shape of the DNA substrate, the migration through the gel matrix would differ. Not surprisingly, larger and less compact substrates would have reduced migration in the native gel.

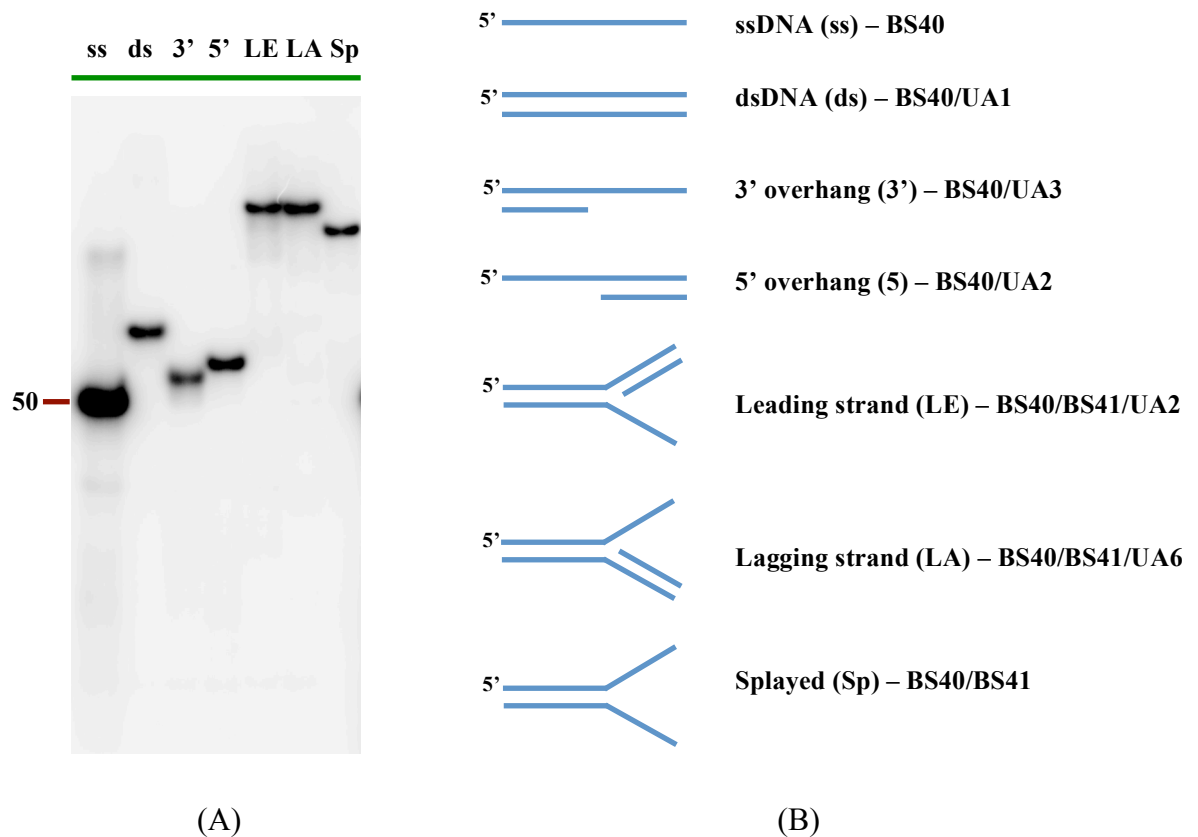


Figure 3.6: Gives an example of annealed substrates made in this project. Part A shows the annealed substrates in a native 10% acrylamide gel. Part B shows the different structures annealed in part A.

After confirming that the structures have been annealed properly, the substrates can be used in nuclease assays. We know that members of the nuclease superfamily, like XPF and MUS81, have structures with splayed arm, replication forks, and overhang structures as preferred substrates. Therefore, a set of such substrates were labelled on the 3' end ($\alpha - ^{32}\text{P}$) and incubated with the protein for 1 hour at 37°C. Figure 3.7 contains 8 substrates (splayed, fork, leading, and lagging) labelled on both strands.

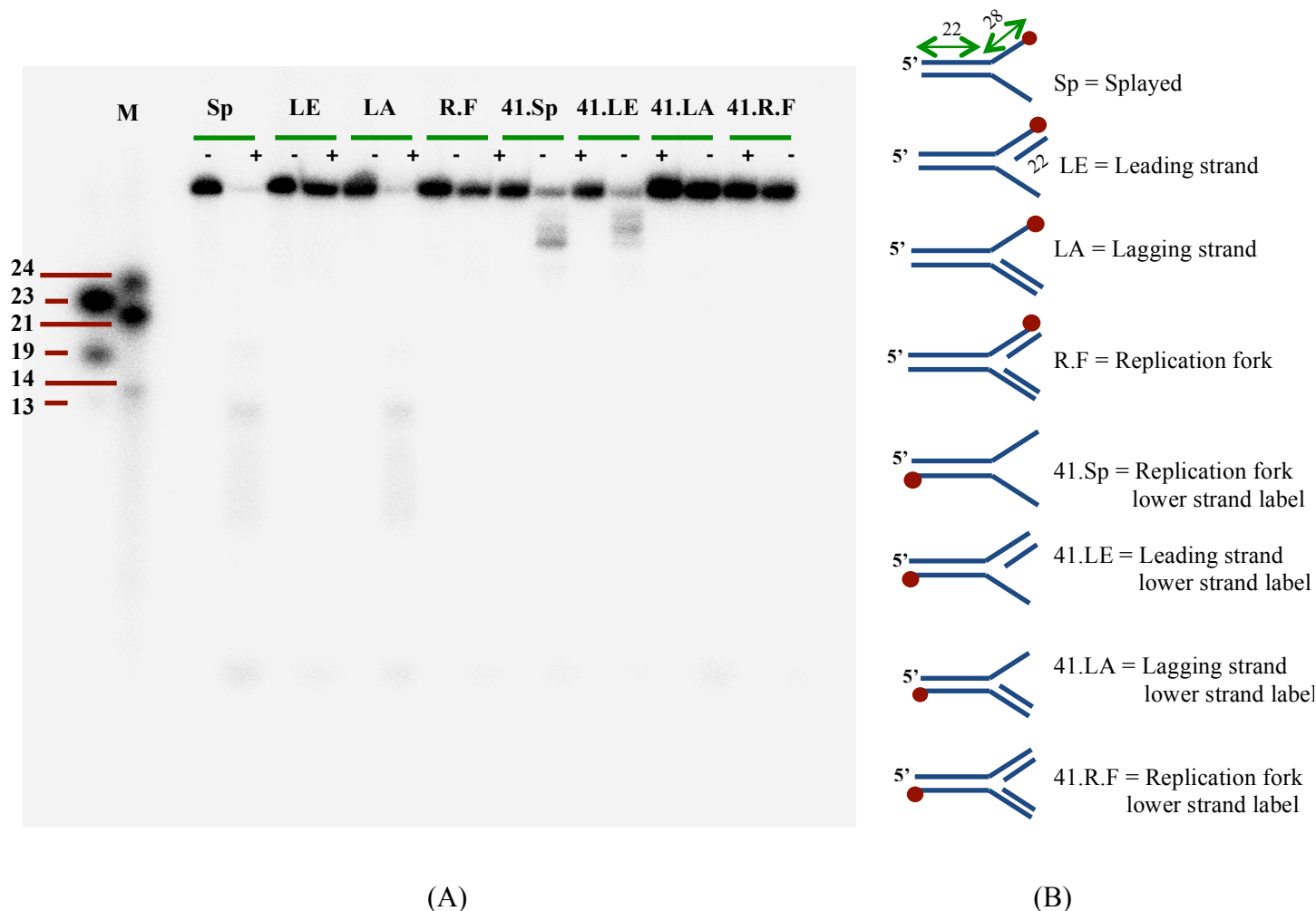


Figure 3.7: Part A shows a 20% acrylamide denaturing gel for the 3' labelled substrates shown in part B, with a marker for reference on the left side. Reactions occurred for an hour at 37°C with the nuclease buffer described in section 2.3. [Protein] = 500nM and [substrate] = 100nM.

All eight structures shown have a double-stranded region (annealed segment) that is 22 nucleotides long and a single stranded flap that is 28 nucleotides long. What is immediately clear from Figure 3.7 is that there is activity on both splayed structures (Sp and 41.Sp), lagging, as well as lower strand leading (LA and 41.LE). It is also apparent that no activity

occurs for leading, lower strand lagging, and both fork structures. This suggests that the protein does not have any observable activity with structures that do not contain single stranded regions of DNA. The annealed strands in the fork and leading strands create a double stranded region that seems to prevent the catalytic activity of the C15ORF41/Codanin1 complex. This gel also confirms that the cleavage seen in Figure 3.3 is not glycosidic in nature, with the fluorescent signal being resultant from an endonucleolytic cleavage.

In terms of the incision pattern, looking at 41.Sp and 41.LE lanes, it can be deduced that the protein has cleaved in the 5' ssDNA region. There are several bands that, even though not entirely resolved, suggest the protein might be cleaving in a sequential manner, with the cleaved product then acting as the new substrate. The bands stop near the 24-nucleotide mark, which is 2 nucleotides away from the double-stranded region. This again suggests that the protein can only function in the single stranded region, with its activity prohibited in the double stranded (annealed) region. A similar pattern emerges when the labelling occurs on the top strand such as in the Sp and LA lanes, with a faint but visible band approximately near the 4 nucleotide region. This would correspond to the protein cleaving near the 3' end. Other faint bands appear in higher regions (near 12 nucleotides and one near 19 that is only visible with increased contrast). Again, all of these cleavage points occur in the ssDNA region as opposed to dsDNA. Therefore, it can be concluded that the protein can cleave both 3' and 5' flaps. Based on the spacing of the bands, the cleaved products are approximately 5-6 nucleotides apart. Figure 3.8 shows the areas where cleavage is occurring in the different substrates described in the gel in Figure 3.7.

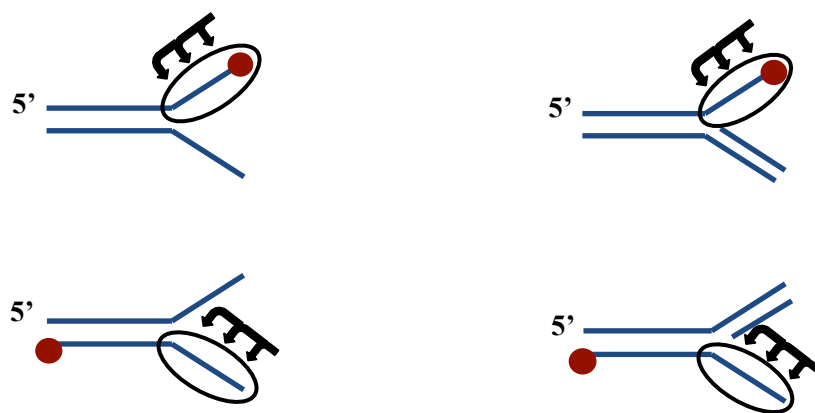


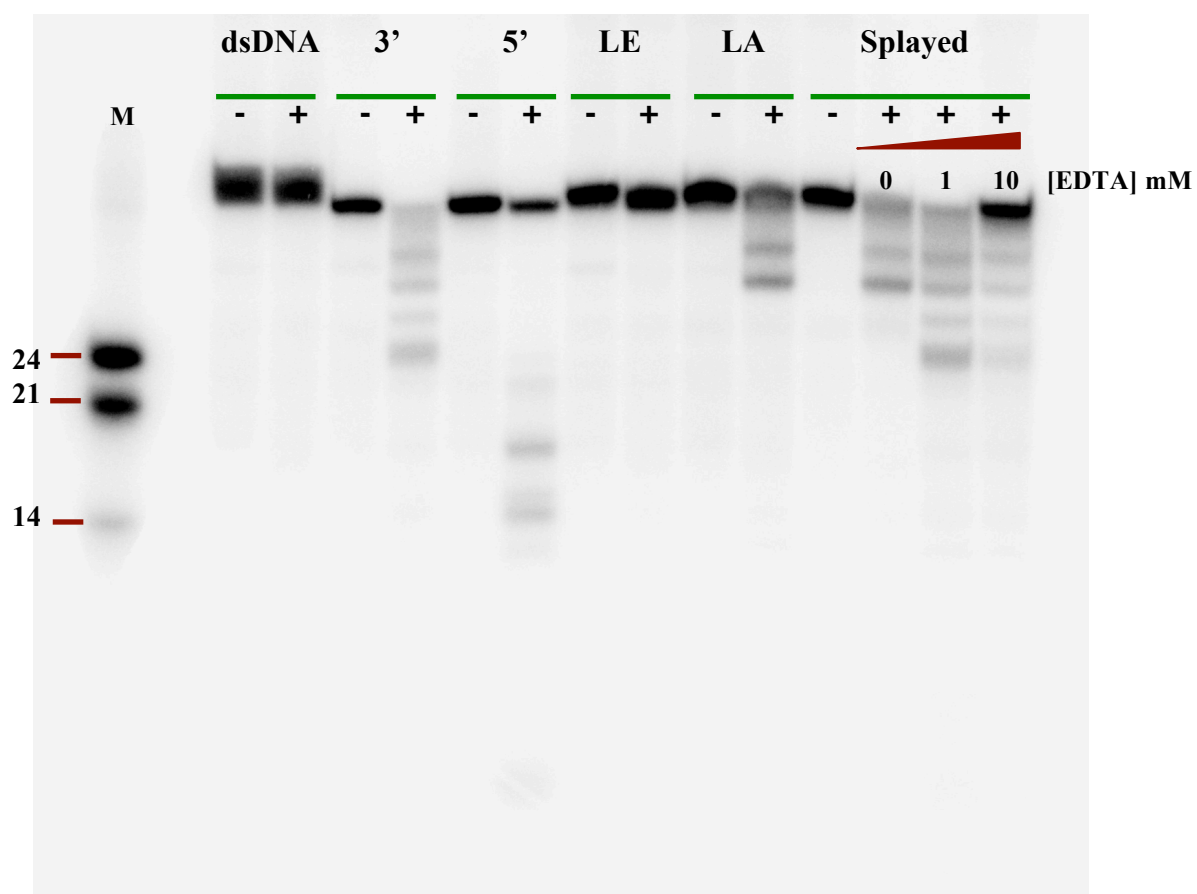
Figure 3.8: Indicates the substrates that showed cleavage products and the regions in which the cleavage occurred.

The reduced signal strength in the case of top strand labelled structures (Sp and LA) is plausibly due to the small size cleaved product being lost in either the fixing or drying process, a problem sometimes encountered in acrylamide gel analysis. This could also explain why the signal strength of all product bands combined is lower than that of the control sample, indicating a net loss of radioactive signal. This is not the case for the lower strand labels, in which the signal strength of the products matches that of the control, explained by the larger cleaved products being retained in the gel.

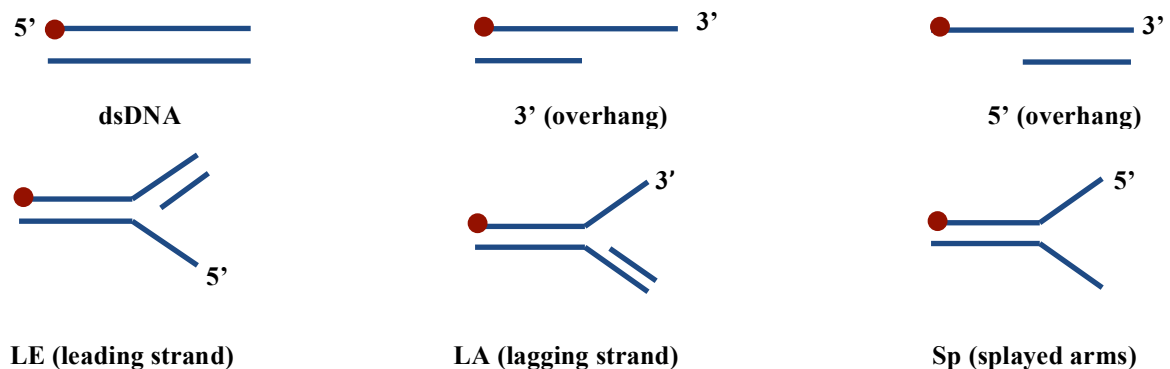
Another notable observation is that while the protein appears to act on both 3' and 5' flaps, 3' flaps appear to be a preferred substrate. Looking at the + lanes for Sp and LA, almost no signal appears for the original substrate with 8.15% and 25.8% of substrate left, respectively, a difference that could probably be explained by the steric hindrance introduced from the annealed strand to the 5' flap in the LA structure. Contrast this with 41.Sp and 41.LE, with 50.1% and 47.1% of substrate remaining respectively. The percentage of substrate left at the

end of the reaction was approximated by drawing a region of interest (ROI) in ImageJ software and then dividing the signal strength of the positive lane over that of the control.

To confirm the observations made with the 3' α -radiolabeled gels, and to delve deeper into substrate preferences, a range of architectures were made with 5' γ - ^{32}P labels. The reaction was carried out at 37°C for an hour with the aforementioned nuclease buffer. The results are summarized in Figure 3.9.



(A)



(B)

Figure 3.9: Part A shows a 20% acrylamide gel with the substrates diagrammed in part B after incubation with the protein at 37°C for an hour. All substrates in part B are labelled on the 5' end.

The initial observation from Figure 3.9 is that it confirms the observation made with the 3' labelled substrates, such that the protein does not act on regions with double stranded DNA, as apparent in lanes with dsDNA and LE. Furthermore, the protein acts on both 3' and 5' flaps (single stranded regions). Again, the level of affinity to 3' flaps is higher, as can be observed from the percentages of substrates left at the end of the reaction when conducting ImageJ ROI analysis. Only 18.7% of the 3' overhang substrate remains and 31.9% for the splayed structure, as opposed to 53.1% and 52.2% for the 5' overhang and lagging ones respectively. For the splayed structure, the ionic chelation from EDTA reduces the activity, which reveals the dependence of the functionality on the divalent cation as occurs with all other members of the PD-(E/D)XK superfamily.

In Figure 3.7, it was pointed out that there was a loss of signal occurring, where the strength of the signal in the control lane was significantly greater than that of the sum of all bands in the reaction lane. As discussed, this could be explained by an initial incision on the 3' end that detaches the radiolabel in a low molecular weight product, followed by other incisions on a now unlabelled substrate. In Figure 3.9, an ROI measurement reveals that the entire signal is retained in the products. It could be concluded then that the protein acts primarily in a 3'-to-5' fashion. Identical patterns appear in the 3' overhang and the splayed structure, both of which have a single-stranded region of 28 nucleotides and a double-stranded region of 22 nucleotides. Therefore, these two substrates are effectively identical to the protein. The protein appears to be cleaving in 5-6 nucleotide increments starting from the 3' flap and working towards the 5' end until the dsDNA region is encountered, at which point the protein's activity ceases.

The pattern can be observed from the 5' overhang lane, which can be thought of as a reversal of the 3' overhang structure. Several bands appear near 23, 19, and 14 nucleotides. Based on the same model proposed earlier, this could be explained by the protein docking on the single stranded region, incising approximately 5 nucleotides from the double-stranded region, as it does in the other structures, and then continuing the activity in 5 nucleotide increments. The protein would probably not be able to function with any substrate smaller than 12-13 nucleotides as it requires nearly 5 nucleotides to dock on. This would also correspond to what appears in Figure 3.7 with the 41.Sp lane, in which the clearest band is near 28 nucleotides. That would in turn correspond to the protein incising 5-6 nucleotides from the double

stranded region, working from a 3'-to-5' manner. A simple schematic summarizing the activity on 3' and 5' flaps is shown in Figure 3.10.

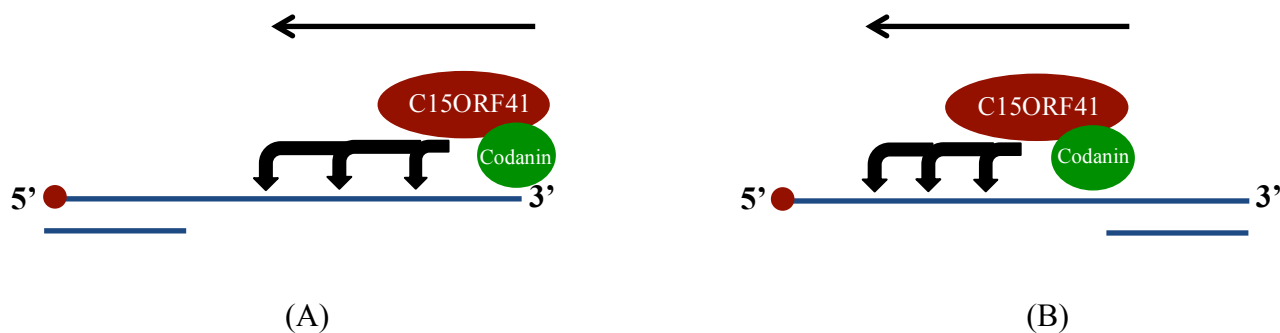


Figure 3.10: Shows the activity of the C15ORF41/Codanin complex on substrates with 3' and 5' single stranded flaps, with incision points are represented as arrows. The arrows on top show the 3'-to-5' directionality predicted for this complex.

It is apparent from the experiments presented thus far that the protein complex acts on single-stranded regions of DNA and has $3' \rightarrow 5'$ directionality in its endonucleolytic activity. However, one of the questions that remain is the possibility of exonucleolytic activity for the protein. One way to exclude this would be by conducting a reaction in which the substrate has a biotin attached to the 3' end, such that the substrate does not terminate with a nucleotide that the protein can recognize. Another question is whether or not the protein activity is contingent on having a single-stranded flap, and therefore if activity can be observed in ssDNA regions situated within a 'bubble' structure. Figure 3.11 shows reactions with a range of substrates that aim to address these questions in addition to an ICL substrate to test the effect, if any, of a cross-link.

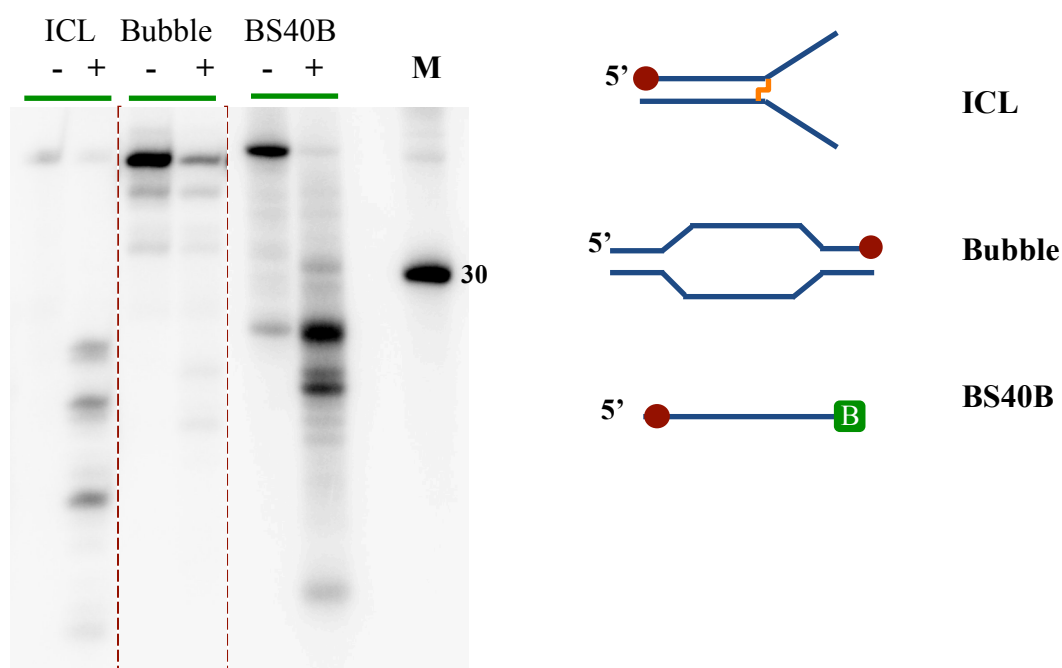


Figure 3.11: Shows a 20% acrylamide gel biochemically characterizing the behaviour of the protein on 3' end biotin-blocked substrates, a bubble substrate excluding 3' flaps, and an ICL.

Examination of Figure 3.11 shows that the protein still has activity on substrate BS40B, which has a biotin attached to the 3' end. This indicates that exonucleolytic activity is unlikely, as the biotin moiety introduces steric hindrance and replaces the terminal nucleotide recognized by exonucleases. However, a more robust test would include another ssDNA without the biotin (BS40) as a control in which the digested products can be directly compared. The bubble structure reveals another interesting element of the complex's behaviour, as it clearly does not require a single-stranded flap to recognize the substrate and initiate its catalytic activity. Finally, the ICL does not seem to influence the incision patterns of the protein, and it can maintain its activity distal to the ICL in the substrate.

In section 1.6.3, the general cellular roles taken by the PD-(D/E)XK superfamily of proteins were described. As mentioned, some of these proteins are involved in Holliday junction resolution. From the previous several experiments presented, it might be predicted that the C15ORF41/Codanin complex is probably not involved in resolving HJs as the protein only functions in regions of single stranded DNA. To confirm this, the complex was incubated with HJs in the same reaction environments as before (37°C, 1 hour, and same Mg^{+2} nuclease buffer), and the product was run in a 20% acrylamide gel to visualize the products. In this case, all four strands in the HJ were labelled on the 5' end. Figure 3.12 shows that, as predicted, the protein had no activity on Holliday junctions.

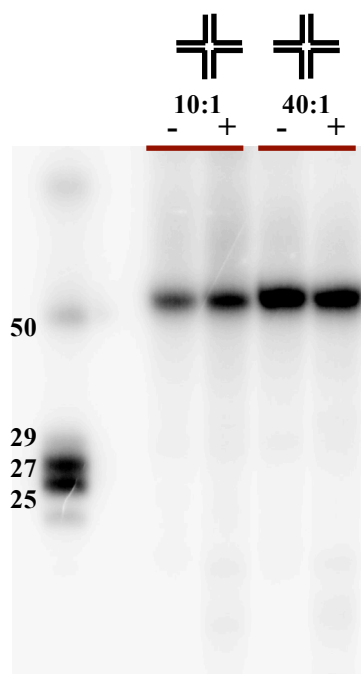


Figure 3.12: 20% acrylamide gel showing the protein complex incubated with the diagrammed Holliday junctions at two different protein:substrate ratios.

One of the common forms of endogenous DNA damage described previously is the formation of 8-oxoguanine due to reactive oxygen species. To test whether this common form of damage affects the behaviour of the protein, several substrates with 8-oxoguanine were designed, with the damage occurring at different locations. Figure 3.13 shows how there is no significant difference between the substrates that contained the damage and those that did not (confirmed via ROI analysis on ImageJ). All substrates were labelled on the 5' end and the number 8 in the figure symbolizes the damage. Any distortion that might be caused by 8-oxoguanine in the double stranded DNA case did not appear to activate the protein's catalytic activity.

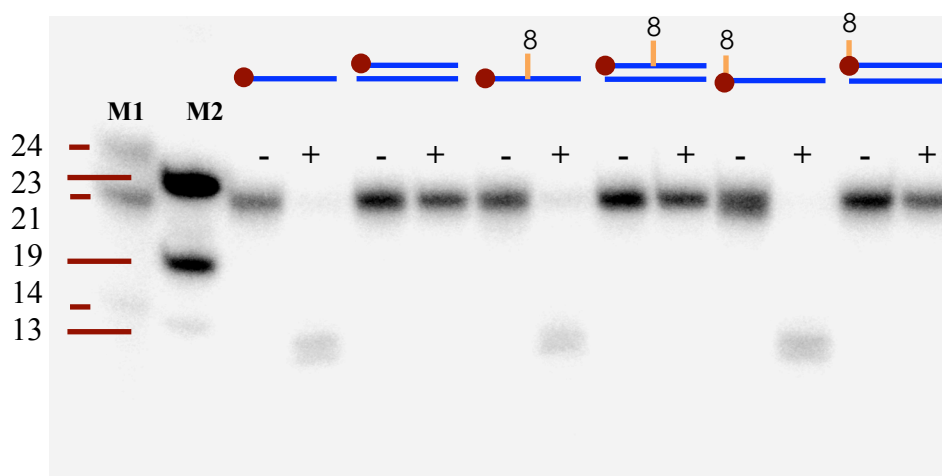


Figure 3.13: A range of substrates with 8-oxoguanine damage in different locations and the corresponding damage-free substrates, all labelled on the 5' end. Shows a 20% acrylamide gel with the reaction occurring at 37°C for 1 hour and the same magnesium (Mg^{+2}) nuclease buffer.

In section 1.6.3, RNA processing was also mentioned as a substrate for this protein superfamily. To test the possibility that the C15ORF41/Codanin complex can process RNA, the protein was incubated with RNA substrates under the same conditions as the previous substrates. The labelling process for RNA was conducted in the same way as that described in the methods section for DNA substrates.

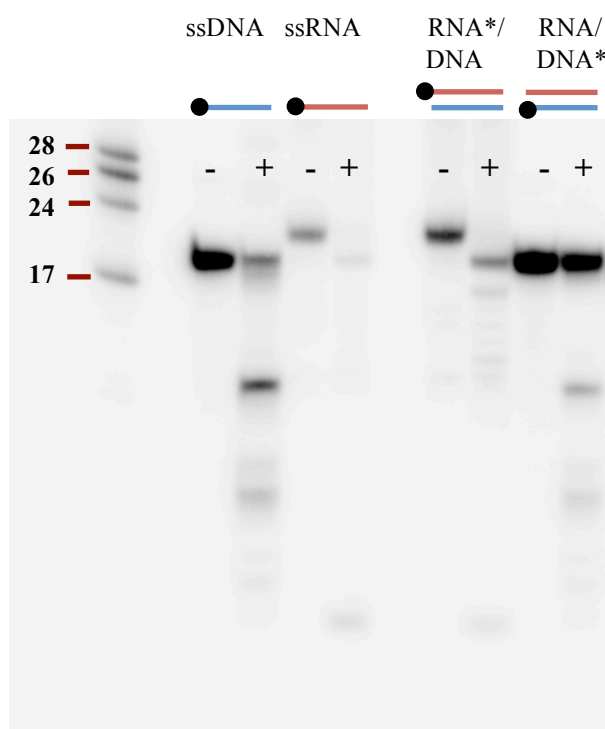


Figure 3.14: A 20% acrylamide gel with a range of DNA and RNA substrates labelled at the 5' end. Reaction carried out at 37°C for an hour with substrates that are 20 nucleotides long.

First, it should be pointed out that the difference in the band positions for the substrates is due to the variation in the molecular weight between the RNA and DNA substrates. It should be

noted, furthermore, that DNA labelling tends to be stronger and produces less diffuse signals when viewed in a gel when compared to RNA substrates, likely due to the higher inherent instability of RNA molecules. This adds a layer of complexity to comparing differential protein activity on DNA and RNA substrates.

As can be seen from Figure 3.14, the protein has activity on all single stranded substrates, both ssDNA and ssRNA. The ability of the protein to cleave RNA is very interesting as members of the nucleolytic superfamily are involved to varying extents in RNA processing functions in cells. In the annealed double-stranded hybrids, minimal activity appears when the DNA is labelled, while a relatively stronger cleavage seems to be occurring when the RNA is labelled. The cleavage pattern appears to be different in the case of RNA and DNA, with no major products that match those in the case of ssDNA. There is also a significant loss of signal for the RNA substrate suggesting a cleavage near the 5' end, however this cannot be confirmed. There certainly are examples of nucleases that act on both DNA and RNA substrates, with one example being E. Coli Exo III, which has 3' $\rightarrow$ 5' exonuclease function on DNA. Exo III also has RNase H behaviour (Rogers and Weiss, 1980), showing preference to RNA in DNA/RNA hybrids, similar to what is observed in figure 3.14. Further experiments are necessary to obtain a clearer picture of the protein's interaction with RNA, such as testing with various RNA substrate architectures. This could lead to a greater understanding of the protein's potential involvement in RNA processing mechanisms.

3.5 Comparative Gels

In order to gain a greater understanding of the function of this protein within the context of the wider superfamily, the protein was run side-by-side with MUS81-EME1 complex using substrates established to be efficiently processed by both C15ORF41-Codanin and MUS81-EME1. Figure 3.15 (A) shows the complex and MUS81-EME1, with the 3' (α) labelled substrates indicated in part (B). The concentration of both proteins was the same (500nM) with a substrate concentration of 100nM.

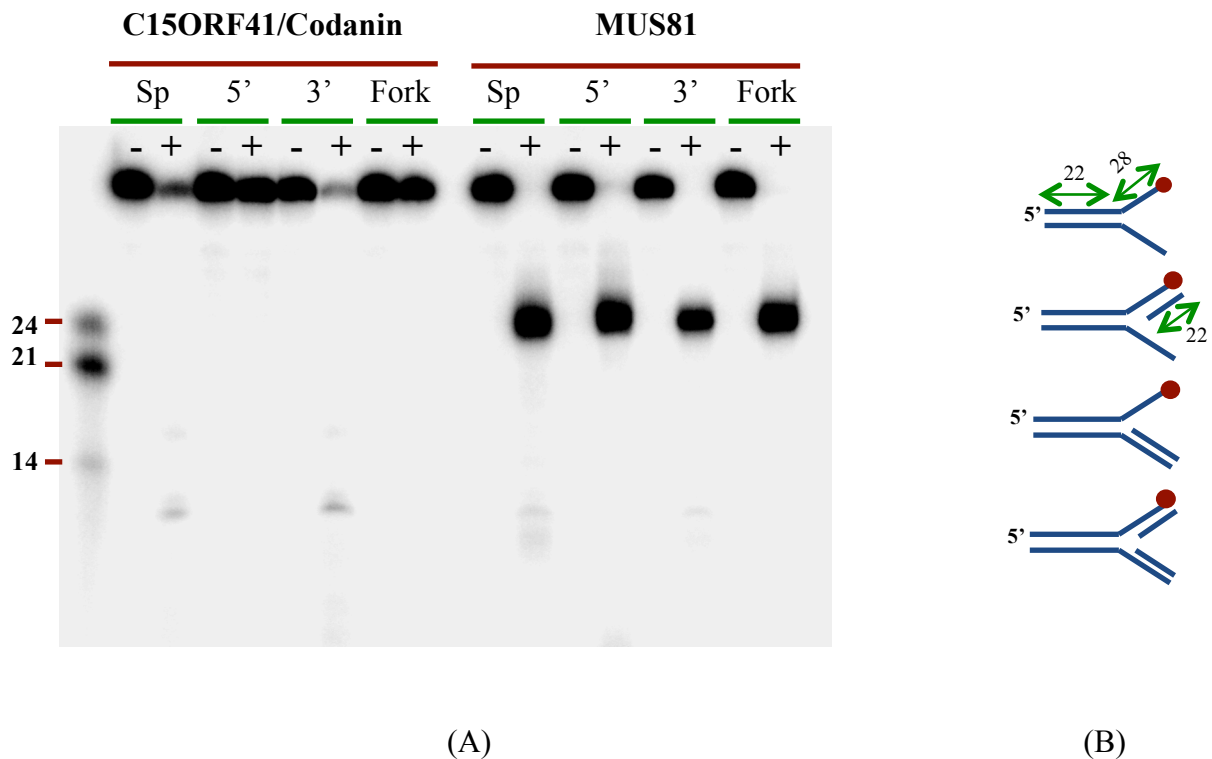


Figure 3.15: Part A: shows a comparative 20% acrylamide gel with the C15ORF41/Codanin complex and MUS81-EME1 reacting with various substrates. Part B: the architectures of the substrates in part A, all labelled at the 3' end.

Figure 3.15 shows that MUS81 and C15ORF41 complexes behave quite differently. MUS81-EME1 is efficient at incising in the double stranded region of the substrates, in contrast to C15ORF41. MUS81-EME1 also makes one major incision near the junction, which accounts for the entire product. ImageJ ROI analysis shows that 100% of the signal is present in the reaction lane as compared to the control lane. As mentioned earlier, this is in contrast to C15ORF41, where a significant loss of signal occurs, probably due to the small, cleaved product being lost in the fixing and drying processes (Figure 3.7). Moreover, for the same concentration and under the same reaction conditions, MUS81 appears to be a much stronger nuclease than C15ORF41. This could be due to our lack of understanding of the characteristics of the protein, possible absence of an unknown binding partner in the reaction mixture, or that further refinement of reaction conditions is required. In order to gain a fuller understanding of the differences in protein behaviours, however, other substrates would have to be tested, with one example including Holliday Junctions nicked at various locations.

3.6 Electrophoretic Mobility Shift Assay (EMSA)

EMSAs provide visual, semi-quantitative data regarding binding between proteins and DNA or RNA substrates. The attachment of protein to the substrate creates a product with a higher mass and size, which then experiences greater impeding forces in a gel. Therefore, the affinity causes what is often referred to as a ‘band shift’ that can be easily visualized in the gel. As detailed in the methods section of the thesis, the EMSAs were run in a 10% native (non-denaturing) gel at 4°C. The colder temperature should decrease the functionality of the protein, accordingly hindering its nuclease activity and stabilising the bound form. Figure 3.16 is an EMSA conducted for both MUS81 and C15ORF41/Codanin, with the mixtures

incubated for 15 minutes before loading. All substrates used in this assay were labelled on the 3' end.

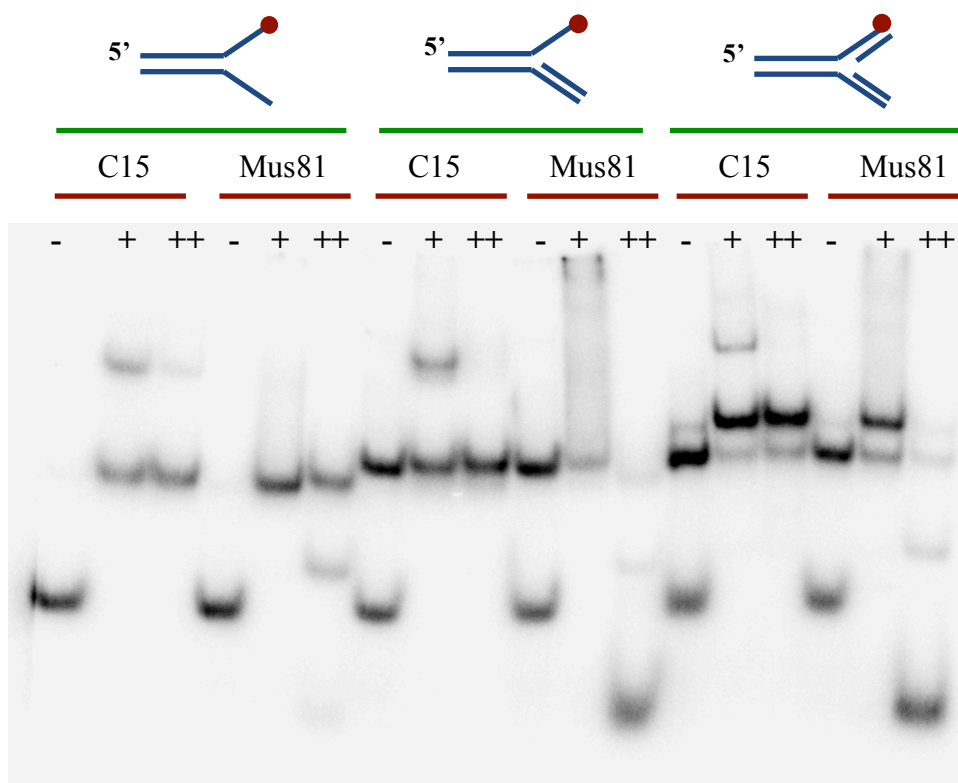


Figure 3.16: 10% native EMSA gel for MUS81 and C15ORF41/Codanin with an incubation time of 15 minutes at 4°C. The negative (-) lanes are for substrates without the protein, positive (+) for the substrate with the protein but without Mg^{+2} in the buffer, and double positive (++) for the substrate with the protein and Mg^{+2} containing buffer.

Looking at Figure 3.16, the first thing that is clear is both C15ORF41 and MUS81 complexes are capable of attaching to the substrate in the absence of the divalent metal ion. The nuclease assays shown previously, however, indicated that the activity of the C15 protein was

abolished when chelating the ions with EDTA. This reveals that, as is consistent in the nuclease superfamily, the divalent ion is necessary for catalytic activity but not for binding to the substrate.

An interesting observation in the case of the C15 complex is that it appears two proteins can bind to one substrate. The lanes containing the protein but not the metal ion have two upper bands, which indicates that some substrates are bound to one protein and others to two. The distance separating those bands is the same, providing further evidence for the two-protein attachment model. The strength of the highest band (corresponding to two proteins) weakens going from the splayed structure to leading and fork. This could be due to the steric hindrance introduced by the annealed strand, such that the majority of the substrates are bound to only one protein. The same pattern does not appear for MUS81, with all lanes containing one band in the single positive lanes. For the double positive lanes, containing the magnesium ions, digestion is likely occurring and interfering with the results of the EMSA. This is more pronounced for MUS81, with clear digestion bands in the ++ lanes, expected from its superior nuclease activity.

3.7 Fluorescence Anisotropy

EMSAs only provide a semi-quantitative indication of protein/DNA interaction. Moreover, the analysis can be compromised by digestion of the substrate, as occurred in the gel in Figure 3.16. A more quantitative and accurate method for understanding binding patterns and kinetic behaviour is fluorescence anisotropy. This is a technique that utilizes the polarization of light in order to investigate the biomolecular interactions. A strong light beam with random

polarization is passed through a filter in order to obtain one specific polarization. The light is then absorbed by the fluorescein moiety in the substrate, which then emits a light of the same wavelength. As molecules tumble in solution, the emitted light again has a random polarization orientation. Passing that light through another filter selects the light that has the same orientation, and this allows for a measurement of the rate of tumbling in solution of the molecules. If another molecule is added that interacts with the substrate, a decrease in the average tumbling velocity occurs, leading to a longer temporal conservation of the polarization signal (Jameson and Sawyer 1995).

This setup was used in this project to measure the interaction of proteins to a DNA substrate (BS20: ssDNA with fluorescein tag – refer to Table 2.1). Figure 3.17 shows the fluorescence polarization curve for the C15ORF41/Codanin1 complex. The data was obtained using PHERAstar (BMG LabTech) and plotted in Prism (GraphPad Software Inc.). The dissociation constant (K_d) can be measured using this graph by finding the IC_{50} value. Using Prism, the dissociation constant of the protein was found to be approximately (12.74 ± 0.7) nM. When compared to other members of the superfamily, this is a relatively low K_d value. XPF, for example, has a reported dissociation constant of $(2-5)$ μ M, which is 2 orders of magnitude higher. This suggests a tighter binding to the substrate for the C15ORF41 complex.

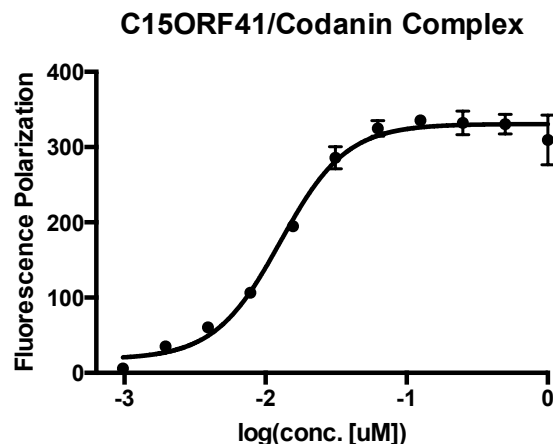


Figure 3.17: Polarization vs. log concentration curve obtained from fluorescence anisotropy for the protein with 20-mer ssDNA substrate. Dissociation constant obtained from this graph is 12.74 nM.

3.8 Discussion

The results from this section reveal several interesting elements regarding the biochemical behaviour of the C15ORF41/CDAN1 complex. First, the fluorescence gels (Figure 3.3 and 3.4) clearly show the importance of divalent cations in the activity of the protein. The strength of the cleaved signals was highest for Mn^{+2} and Mg^{+2} , consistent with other members of the superfamily. The acrylamide gels in Figures 3.7 and 3.9 show that the protein is an endonuclease that acts on single-stranded regions of DNA, as overhangs and flaps revealed a much stronger level of activity than annealed or double-stranded regions. The sequential products indicated in Figure 3.9 point to the fact that the protein is likely successively incising the single-stranded region of DNA. Since 5' labelled oligos were used in the gel in Figure 3.9, this provides evidence that the protein acts in a 3' to 5' manner. This is further supported by the presence of only single cleaved products (bands) in 3' labelled oligos, since those are cleaved off first (Figure 3.7). The activity also appears to be enhanced in the case of 3' flaps

as compared to 5' flaps, which is evident in Figure 3.7 and explained further in that segment of the thesis. Figure 3.10 provides a diagram that summarizes the proposed activity of the complex and its mechanism in cleaving products as deduced from the gels presented in section 3. Figure 3.14 also reveals the ability of the complex to incise RNA substrates, though more work is needed to fully understand the protein's cleavage pattern on RNA oligos.

The EMSA in Figure 3.16 showed the importance of the divalent metal cation in the biochemical activity of the protein but not in its ability to attach to the DNA substrates. The appearance of two bands in the DNA-bound lanes for the complex suggests that it is likely there is a 2:1 ratio of protein to substrate. The fluorescence anisotropy data from Figure 3.17 suggest tighter binding with the substrates as compared to other members of the superfamily.

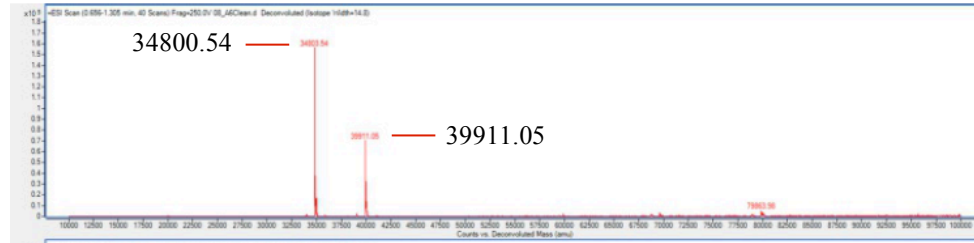
The data in this section provides evidence for the predicted catalytic activity put forward by Babbs *et.al* and provides an outline for the general characteristics of its behaviour with various oligos and nucleotide architectures. Further experiments are necessary to provide a fuller image for the protein's activity, which will be discussed in the future work segment of this thesis.

4. Protein Mutants and Biochemical Behaviour of Individual Proteins in Non-Complex Form

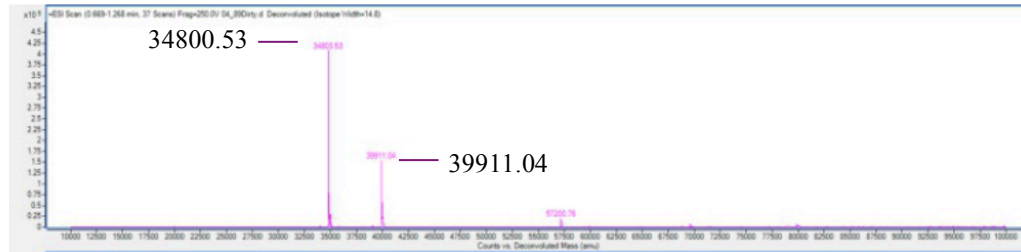
4.1 Nuclease Inactive Mutant Form of C15ORF41 Complex

With the help of Hazel Aitkenhead and Joseph Newman from Professor Opher Gileadi's lab at the Structural Genomics Consortium (SGC), mutated forms of the protein were purified and tested for catalytic activity. Two mutations were introduced that, based on the bioinformatics data on the protein and the knowledge of the catalytic core of the superfamily (refer to section 1.6.3), should theoretically produce an inactive or 'nuclease dead' version of the protein complex. The mutations were both in the catalytic core of C15ORF41: D196A and E214A. The proteins were purified in the same way as described in the methods section.

Two batches of proteins were purified, a mutant and a control wild type. As mentioned before, the "wild type" protein contains mutations introduced to increase stability and enhance purification efforts. Based on bioinformatics analysis, it is predicted that those mutations are unlikely to be detrimental to the biochemical behaviour of the protein. However, it should be acknowledged that this is a limitation of the study and more tests on the protein structure and function have to be conducted to confirm this prediction. For each purified protein, two samples were made; one that was purified with ion exchange chromatography (MonoQ) and one that was not. The mass spectrometry data is shown in Figures 4.1 and 4.2, both showing the presence and expected masses for C15ORF41 and Codanin1.

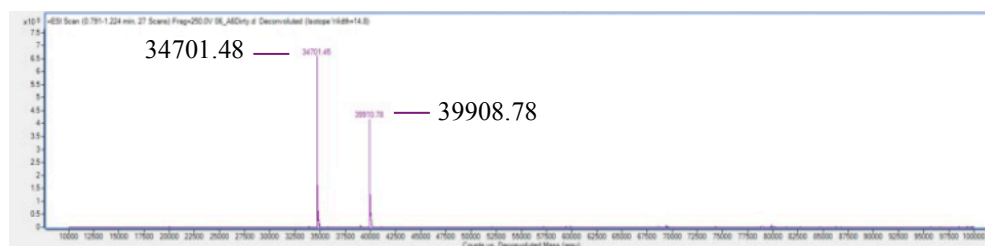


(A)

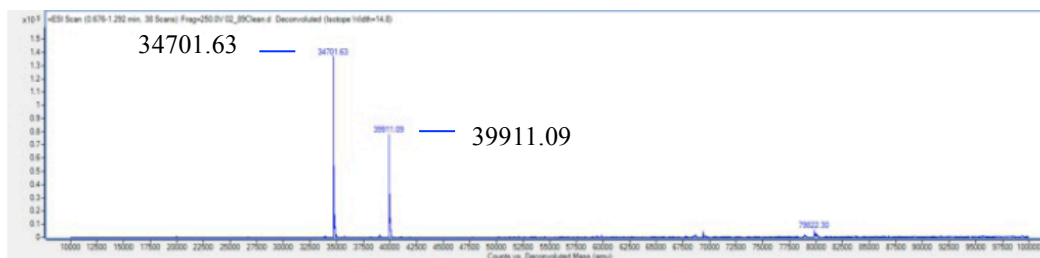


(B)

Figure 4.1: Part A shows the mass spectrometry data for the wild type C15ORF41/Codanin protein complex after ion exchange filtration (MonoQ). Part B shows the mass spectrometry data for the complex before ion exchange filtration. Both show expected mass data for the C15ORF41 and Codanin1.



(A)



(B)

Figure 4.2: Part A shows the mass spectrometry data for the mutant C15ORF41/Codanin (D196A, E214A) complex after ion exchange filtration. Part B shows the mass spectrometry data for the complex before ion exchange filtration. Both show expected mass data for the C15ORF41 and Codanin1.

The full sequence of the mutant protein is listed below:

```

      10          20          30          40          50          60
MHHHHHHSSG VDLGTENLYF QSMILTKAQY DEIAQCLVSM PPTRQSLRKL KQRFPSQSQA

      70          80          90         100         110         120
TLLSIFSQEY QKHIKRTHAK HHTSEAIESY YQRYLNGVVK NGAAPVLLDL ANEVDYAPSL

     130         140         150         160         170         180
MARLILERFL QEHEETPPSK SIINSMRLDP SQIPDGVLAN QVYQCIVNDC CYGPLVDCIK

     190         200         210         220         230         240
HAIGHEHEVL LRDLLLEKNL SFLDEDQLRA KGYDKTPAFI LQVPVAVEGH IHWIASKAS

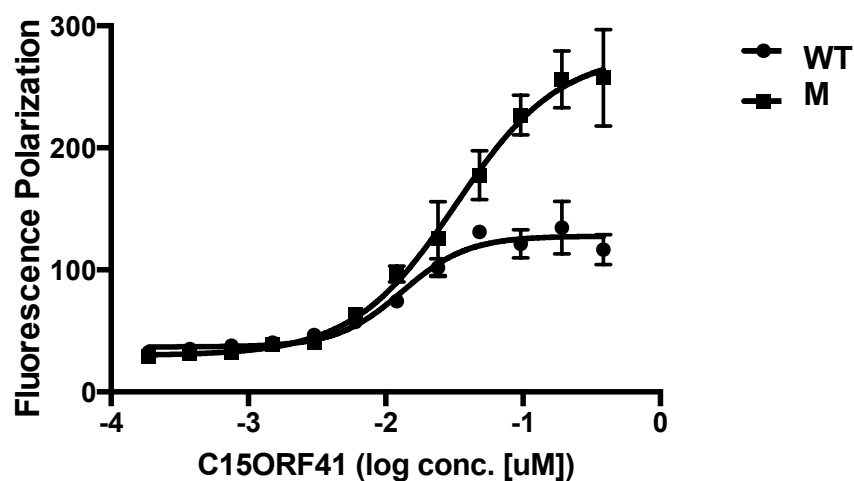
     250         260         270         280         290         300
FGDESSHAY LHDQFWSYRN RFGPGLVIYW YGFIQELDCN RERGILLKAC FPTNIVTLCH

```

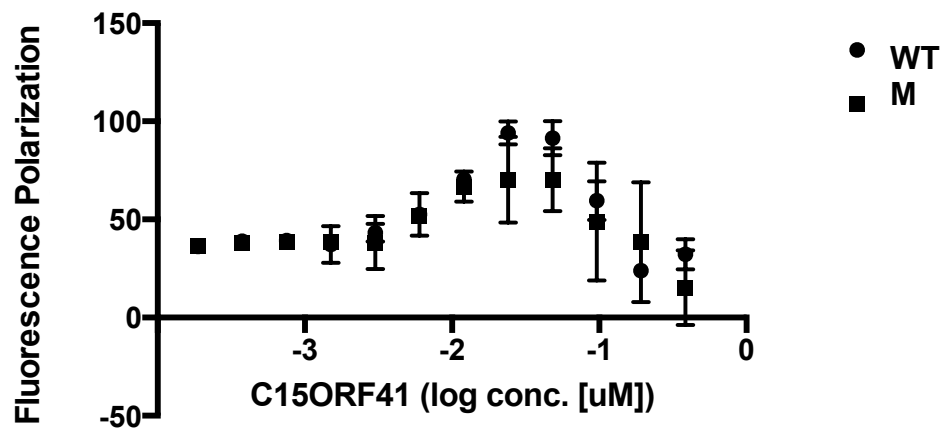
SIA

The importance of making a nuclease dead version of a protein is that it provides evidence for the importance of the catalytic residues in its activity. It also gives further confidence that the activity observed in the assays is actually intrinsic to the protein and not contamination carried over from the purification process. Unfortunately, not enough protein was obtained in order to allow for thorough biochemical investigation of the effect of the mutations on the catalytic activity of the complex. A preliminary assay did show slight activity with the presumably nuclease-inactive mutant, but with the poor quality of the preparation more purifications and experiments would have to be conducted to make any conclusions. This will be addressed further in the future experiments segment of this thesis.

Fluorescence anisotropy assays, however, were conducted on both the wild type and mutant proteins in order to test any changes to substrate binding. The results shown in Figure 4.3 A and B reveal the comparative behaviour of both proteins before and after the ion exchange purification step.



(A)



(B)

Figure 4.3: Shows the fluorescence anisotropy data obtained for the C15ORF41 (WT and mutants) with a 20-mer ssDNA substrate before ion exchange in part A and after ion exchange purification in part B. K_d in part A for mutant is 32.5nM while for wild type it is 13.2nM.

The results from Figure 4.3 indicate that in the case of the proteins prior to ion exchange purification, the mutant version has lower binding affinity to the DNA substrate, as indicated by the higher K_d value (32.5nM for mutant vs 13.2nM for wild type). This decrease in binding affinity could in turn lead to a disruption in the protein functionality for the mutant. Even though the mutations are in the catalytic core, resultant conformational changes could disrupt the DNA binding. As for the proteins after ion exchange purification (part B), no binding behaviour is apparent, which could be due to multiple reasons. The fluorescence anisotropy graph in section 3.7 was for the protein complex purified from a different batch, and was passed through an ion exchange column. The discrepancy in behaviour could be caused by an

error in the purification of the second batch shown in Figure 4.3 (B) or an excess chelation of ions necessary for substrate binding. The experiment would have to be repeated to investigate the root cause of this discrepancy in results.

4.2 Activity of Individual Proteins Within Complex

Since in previous experiments C15ORF41 protein on its own did not show any nucleolytic activity, it was co-purified with the C-terminus of Codanin1 to study the behaviour of the complex. Furthermore, Codanin1 does not have any predicted nuclease domains, which supplemented the notion of a complex being necessary for proper binding and catalytic activity. To confirm this hypothesis, fluorescence anisotropy was conducted on the individual proteins, both wild type and mutated forms. Two mutants and one wild type C15ORF41 protein and one mutant and wild type of Codanin1 were tested, with the results being shown in Figure 4.4 (A) and (B). The C15ORF41 mutants are 913 with Y94C (binding domain) and 914 with L178Q (nuclease domain), which are the mutations found in patients in the Babbs paper (Babbs et al. 2013). The Codanin-1 mutation, R1042W, has been found in previous studies to disrupt Codanin's interaction with Asf1 assembly protein. The data was then combined with that previously obtained for the complex and shown collectively in Figure 4.4 (C) to give a comparative view of the binding kinetics. The K_d values are summarized in the table 4.1.

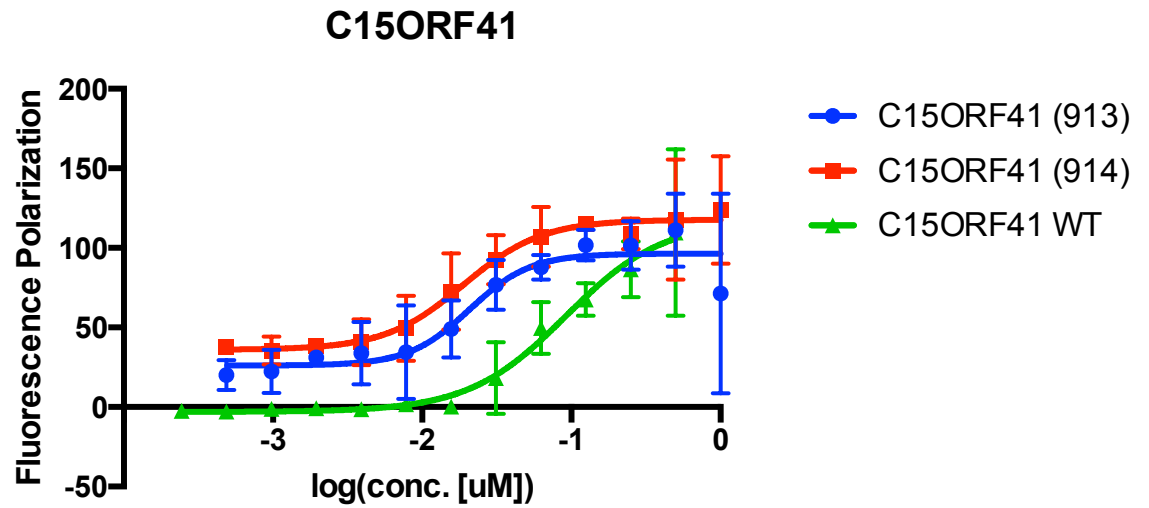
Table 4.1: K_d values for the individual proteins shown in Figure 4.4 A and B.

Protein	Dissociation Constant (nM)
C15ORF41 WT	91.71
C15ORF41 Mutant 913	20.73
C15ORF41 Mutant 914	18.93
Codanin WT	96.5
Codanin Mutant	96.6

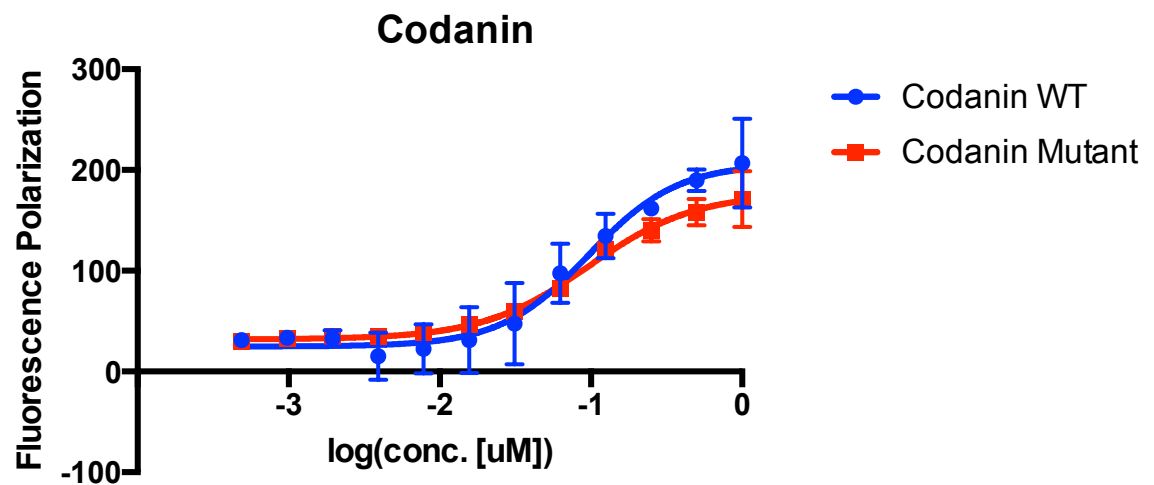
The results of Table 4.1 show that K_d for the complex (approximately 12.7nM) is lower than any of the values of the individual proteins. This indicates that the complex has a much higher affinity for the substrates than C15ORF41 or Codanin1 or any of their mutants individually, reinforcing the view that the binding is significantly improved for the complex. It can be assumed then that Codanin1 enhances C15ORF41's ability to bind to the substrate via its binding domain, allowing it to carry out its catalytic function through its nuclease domain (refer to Figure 1.9 for a detailed view of C15ORF41's sequence and domains). The table also shows that the dissociation constant for Codanin is unchanged for the mutant, remaining equally high. This is consistent with the fact that Codanin does not have a DNA binding domain, and hence an extremely low affinity to DNA.

Interestingly, the patient mutations in C15ORF41 lower the dissociation constant, indicating an enhancement in the DNA affinity. Both values are still higher than that of the complex, however. The physiological significance of this change in affinity is not fully understood, and requires several experiments to build a complete understanding between CDA-I and the

binding affinity of C15ORF41. Due to time and protein availability restrictions, the anisotropy assay was not repeated sufficient times to obtain errors in K_d values.



(A)



(B)

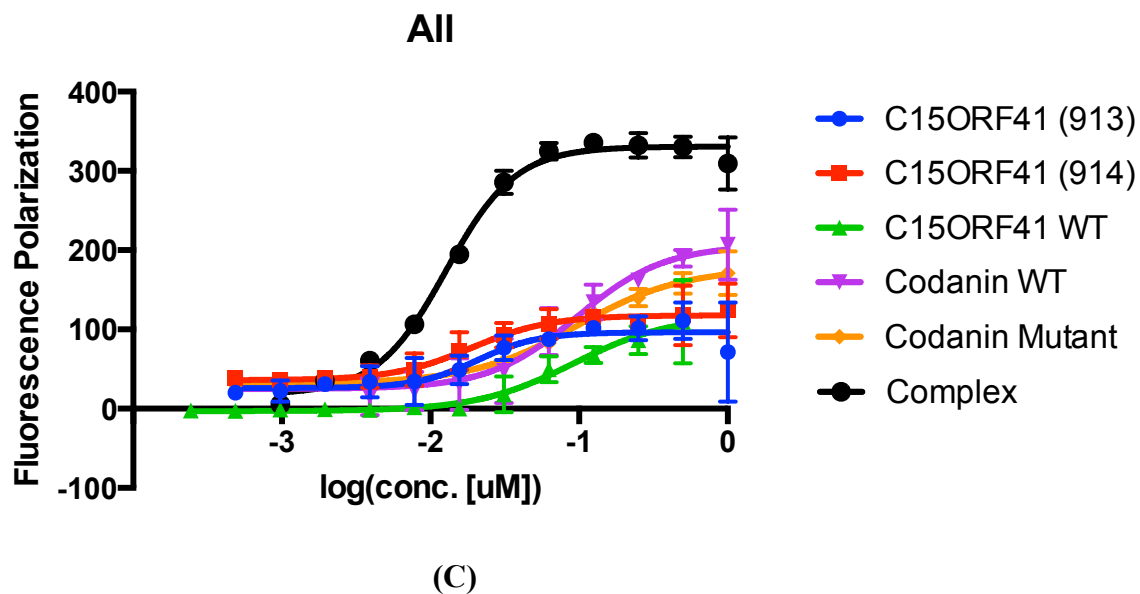


Figure 4.4: Shows the fluorescence anisotropy data obtained for the C15ORF41 (WT and mutants) in part A, Codanin1 (WT and mutant) in part B, and a combined graph with all the individual proteins as well as the complex in part C.

4.3. Discussion and Conclusion

The results from sections 3 and 4 of this thesis offer a significant first step in the characterisation of the C15ORF41 complex from both biochemical and kinetic perspectives, showing how distinctive it is from other members of the superfamily. The radioactive and fluorescent gels indicate that the complex is an endonuclease that acts in a 3'-to-5' manner on both DNA and RNA substrates. The step size for the incisions is shown to be around 5 nucleic acids and substrates with a ssDNA 3' flap appear to be the preferred substrate for the complex. As consistent with the superfamily of proteins, divalent metal cations are crucial to the catalytic function, with the strongest observed for manganese and magnesium. Figure 3.10 provides a summary of the expected behaviour of the protein and its incision patterns.

The EMSA in Figure 3.16 suggests that the cations are not required for substrate binding but are necessary for nucleolytic activity. It also indicates that there is probably a 2:1 stoichiometric ratio between the protein and substrate, with 2 proteins binding to each DNA substrate. The fluorescence anisotropy experiments conducted provide a more quantitative look into the binding between the complex and the substrate. For the wild type protein, a K_d of 12.7nM was obtained in the first purification (Figure 3.14) and 13.2nM in the second purification (Figure 4.3 – A), with a 3.86% deviation in the values. When mutating the catalytic core (D196A and E214A), an increase in K_d to 32.5nM is observed, which could result in a disruption in catalytic function. Figure 4.4 shows that there is a clear enhancement in substrate binding for the complex as opposed to each individual protein on its own, which would explain why C15ORF41 does not show the catalytic activity on its own.

The results of this work also provide an initial understanding as to how this protein could be linked to CDA-I. It is known that nucleosome assembly is key in the restoration of chromatin in new DNA, and that the histone chaperone anti-silencing function 1 (Asf1) is pivotal in supplying H3.1-4 to Chromatin Assembly Factor 1 (CAF-1) during replication-coupled nucleosome assembly (Ask *et al.*, 2012). As mentioned earlier in the introduction, Codanin interacts with Asf1 in S-phase, and its depletion increases the levels of histone-bound Asf1. Ask *et al.* indicates that Codanin-1 acts as a negative regulator of histone-binding Asf1 by showing, amongst other things, that Codanin-1 is found in cytosolic complexes with histones H3.1-4, Importin-4, and Asf1, and by showing that RNAi depletion of Codanin-1 increases amounts of chromatin-bound Asf1.

Gene expression in developing erythroid cells is likely controlled by tightly regulated changes in chromatin structure, which might be sensitive to perturbations in chromatin/histone deposition. The data in section 4 suggest that Codanin-1 enhances C15ORF41's ability to bind DNA substrates. This, coupled with the possible repair of damaged forks introduced by C15ORF41 nuclease activity, could explain the involvement of C15ORF41 in the development of chromatin abnormalities in erythroid cells and the subsequent development of CDA-I via damaged erythropoiesis.

More work is obviously required in order to reach a fuller understanding of the biochemical nature of C15ORF41 in both its native form and when associated with Codanin. Moreover, a transition into biological assays is extremely important to probe the role of C15ORF41 on a physiological level and to fully establish its connection with DNA damage repair. The next

section outlines a few of the experiments that could be conducted in the future to gain such an understanding of this newly discovered protein, which would prove pivotal in creating therapeutics for CDA-I.

5. Future Work

In order to build a complete picture of the kinetic and biochemical characteristics of this protein, several assays need to be conducted. First, in order to more precisely prove the cleavage patterns of the complex, thin layer chromatography (TLC) and/or mass spectrometry (MS) can be used to confirm the low molecular weight products resulting from the 3' incisions observed on the DNA substrates. TLC or MS could clearly identify the 1-3 nucleotide long products, which are not reliably observed on acrylamide gels, helping to confirm the 3'-to-5' directionality deduced. Furthermore, it is important to build on the existing work to produce a nuclease-inactive version of the protein. Achieving an inactive form through the mutation of key residues in the catalytic core would indicate that the activity observed in the gels is in fact due to the protein and not contaminations introduced in the purification procedure. Also, it would provide insight into the accuracy of the modelling predictions to the nature of the catalytic amino acids that lead to its characterisation as a PD-(D/E)XK superfamily.

Nuclease assays on the individual proteins would also be valuable in confirming the catalytic activity as a function of the complex rather than C15ORF41 or Codanin1 on their own. Codanin1 does not have a nuclease domain, and previous experiments with C15ORF41 by members of the McHugh lab did not reveal any catalytic activity. The heterodimer would also follow the same feature in other members of the superfamily, such as the MUS81-EME complex and XPF-ERCC1. An interesting study also would be to repeat the fluorescence

anisotropy assays without the metal cation and compare the results with those from the EMSA in Figure 3.16.

Another interesting direction for the research would be to conduct cell biology assays in order to gain further insights into the cellular role of the protein. As of yet, there are no C15ORF41 antibodies with optimal function, making siRNA knockdowns difficult to validate. An interesting experiment would be to knockdown other DNA repair enzymes such as MUS81 and XPF so as to discover any redundancies or dependencies in the pathways of both proteins. 6 lymphoblast cell lines were provided by the Buckle group – MRC Molecular Haematology Unit for these experiments; 2 normal and 4 CDA-I patient derived lines. Since lymphoblasts are suspension cells, clonogenic survival assays cannot be conducted to investigate knockdown effects, and thus other methods that could be used include MTS cell proliferation assays. Due to time restrictions, these experiments were started but not completed. Table 4.1 lists the mutations in each lymphoblast cell line obtained.

Table 5.1: Lists all lymphoblast cell lines with the corresponding mutations.

Cell Line	Mutation
JT	(Normal): No mutations
BH	(Normal): No mutations
SB	(Patient): C15ORF41 Y914C homozygous
NJ	(Patient): C15ORF41 L178Q homozygous
ES	(Patient): CDAN1 - Compound heterozygote Pro672L and Ala944Ser
EA	(Patient): CDAN1 - Pro51Leu homozygous

6. References

Abdullah, U. *Biochemical Studies on the Role of the XPF-ERCC1 Endonuclease in Repairing Damaged DNA Replication Forks*. Thesis. University of Oxford, 2016. Print.

Adapted: Pagana, KD, Pagana, TJ, & Pagana, TN. "Hemoglobin". *Mosby's Diagnostic and Laboratory Test Rerefence*. 12th ed. St. Louis, MO: Mosby Elsevier, 2015. N. pag. Print.

Armitage, J. *Atlas of Clinical Hematology*. Philadelphia: Lippincot Williams & Wilkins, 2004. Print.

Aslinia, F., Mazza, J. J., & Yale, S. H. (2006). Megaloblastic Anemia and Other Causes of Macrocytosis. *Clinical Medicine and Research*, 4(3), 236–241.

Babbs, C., N. A. Roberts, L. Sanchez-Pulido, S. J. McGowan, M. R. Ahmed, J. M. Brown, M. A. Sabry, D. R. Bentley, G. A. Mcvean, P. Donnelly, O. Gileadi, C. P. Ponting, D. R. Higgs, and V. J. Buckle. Homozygous Mutations in a Predicted Endonuclease Are a Novel Cause of Congenital Dyserythropoietic Anemia Type I. *Haematologica* 98.9 (2013): 1383-387. Web.

Baik, H. W., & Russell, R. M. (1999). Vitamin B12 Deficiency in the Elderly. *Annual Review of Nutrition*, 19(1), 357–377. <http://doi.org/10.1146/annurev.nutr.19.1.357>

Ball, S. (2011). Diamond Blackfan Anemia. *ASH Education Program Book*, 2011 (1), 487–491. <http://doi.org/10.1182/asheducation-2011.1.487>

Bohlius, J., Weingart, O., Trelle, S., & Engert, A. (2006). Cancer-related anemia and recombinant human erythropoietin an updated overview. *Nat Clin Prac Oncol*, 3(3), 152–164.

Retrieved from <http://dx.doi.org/10.1038/ncponc0451>

Boyadjiev, S. A., Fromme, J. C., Ben, J., Chong, S. S., Nauta, C., Hur, D. J., ... Eyaid, W. (2006). Cranio-lenticulo-sutural dysplasia is caused by a SEC23A mutation leading to abnormal endoplasmic-reticulum-to-Golgi trafficking. *Nat Genet*, 38(10), 1192–1197.

Camaschella, C. (2015). Iron-Deficiency Anemia. *New England Journal of Medicine*, 372(19), 1832–1843. <http://doi.org/10.1056/NEJMra1401038>

Carreno-Tarragona, G., Trapiello-Valbuena, F., Aranda-Salom, J., Mart'inez-López, J., Lumbreras, C., Oliveira-Ram'irez, E., & D'iaz-Pedroche, C. (2015). Antiphospholipid syndrome in a patient suffering from congenital dyserythropoietic anemia type III. *Annals of Hematology*, 94(8), 1411–1412. <http://doi.org/10.1007/s00277-015-2379-y>

Clark, S. F. (2008). Iron Deficiency Anemia. *Nutrition in Clinical Practice*, 23 (2), 128–141. <http://doi.org/10.1177/0884533608314536>

Coin, F., Oksenych, V., Mocquet, V., Groh, S., Blattner, C., & Egly, J. M. (2016). Nucleotide Excision Repair Driven by the Dissociation of CAK from TFIIH. *Molecular Cell*, 31(1), 9–20. <http://doi.org/10.1016/j.molcel.2008.04.024>

Rogers, S.G., Weiss, B. *Exonuclease III of Escherichia coli K-12, an AP Endonuclease 1*.
Methods in Enzymology, vol 65, ISBN 0-12-1965-5 (1980).

de Benoist B et al. *Worldwide prevalence of anaemia 1993-2005*. WHO Global Database on
Anaemia Geneva, World Health Organization, 2008.

Doulatov, S., Notta, F., Laurenti, E., & Dick, J. E. (2012). Hematopoiesis: A Human
Perspective. *Cell Stem Cell*, 10(2), 120–136.

<http://doi.org/http://dx.doi.org/10.1016/j.stem.2012.01.006>.

Edifizi, D., & Schumacher, B. (2015). Genome Instability in Development and Aging:
Insights from Nucleotide Excision Repair in Humans, Mice, and Worms. *Biomolecules* .
<http://doi.org/10.3390/biom5031855>

Gadny O, et al. Congenital Dyserythropoietic Anemia type I is caused by mutations in
condanin-1. *Am J Hum Genet*. 2002; 71(6):1467-74

Ganz, T. (2011). Heparidin and iron regulation, 10 years later. *Blood*, 117(17), 4425–4433.
<http://doi.org/10.1182/blood-2011-01-258467>

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M. S., Timmers, C., Hejna, J., D'Andrea, A. D. (2001). Interaction of the Fanconi Anemia Proteins and BRCA1 in a Common Pathway. *Molecular Cell*, 7(2), 249–262.

[http://doi.org/http://dx.doi.org/10.1016/S1097-2765\(01\)00173-3](http://doi.org/http://dx.doi.org/10.1016/S1097-2765(01)00173-3)

Gattermann, N. (2000). From sideroblastic anemia to the role of mitochondrial DNA mutations in myelodysplastic syndromes. *Leukemia Research*, 24(2), 141–151.

[http://doi.org/http://dx.doi.org/10.1016/S0145-2126\(99\)00160-5](http://doi.org/http://dx.doi.org/10.1016/S0145-2126(99)00160-5)

Gehrs, B. C., & Friedberg, R. C. (2002). Autoimmune hemolytic anemia. *American Journal of Hematology*, 69(4), 258–271. <http://doi.org/10.1002/ajh.10062>

Glanz, A., & Fraser, F. C. (1982). Spectrum of anomalies in Fanconi anaemia. *Journal of Medical Genetics*, 19(6), 412–416. Retrieved from

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1048952/>

Greer, J., J. Foerster, G. Rodgers, and R. Means, Jr. *Wintrrobe's Clinical Hematology*. 12th ed. Vol. 1. N.p.: Lippincot Williams & Wilkins, 2009. Print.

Ask, K., *et al.* (2012). Codanin-1, mutated in the anaemic disease CDAI, regulates Asf1 function in S-phase histone supply. *The EMBO Journal*. Vol 31. No 8.

Hattangadi, S. M., Wong, P., Zhang, L., Flygare, J., & Lodish, H. F. (2011). From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood*, 118(24), 6258 LP – 6268. Retrieved from <http://www.bloodjournal.org/content/118/24/6258.abstract>

Heimpel, et al. The morphological diagnosis of CDA: results of a quantitative analysis of peripheral blood and bone marrow cells. *Haematologica* 2010; 95(6):1034-6.

Heimpel, H. (2004). Congenital dyserythropoietic anemias: epidemiology, clinical significance, and progress in understanding their pathogenesis. *Annals of Hematology*, 83(10), 613–621. <http://doi.org/10.1007/s00277-004-0892-5>

Heimpel, H., Anselstetter, V., Chrobak, L., Denecke, J., Einsiedler, B., Gallmeier, K., Kohne, E. (2003). Congenital dyserythropoietic anemia type II: epidemiology, clinical appearance, and prognosis based on long-term observation. *Blood*, 102(13), 4576 LP – 4581.

Hoffbrand, A. V., P. A. H. Moss, and J. E. Pettit. *Essential Haematology*. 4th ed. Malden, MA: Blackwell Pub., 2001. Print.

Howlett, N. G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., de Die-Smulders, C., D;Andrea, A. D. (2002). *Science*, 297(5581), 606 LP – 609. Retrieved from <http://science.sciencemag.org/content/297/5581/606.abstract>

Hughes-Jones, NC, Wickramasinghe, SN, & Hatton, CSR. *Lecture Notes Haematology*. 8th ed. N.p: John Wiley & Sons, 2008. Print

Humbert PO, Rogers C, Ganiatsas S, Landsberg RL, Trimarchi JM, Dandapani S, et al. (2000). E2F4 is essential for normal erythrocyte maturation and neonatal viability. *Mol Cell* 6:281-91.

Iolascon, A., Heimpel, H., Wahlin, A., & Tamary, H. (2013). Congenital dyserythropoietic anemias: molecular insights and diagnostic approach. *Blood*, 122(13), 2162–2166.
<http://doi.org/10.1182/blood-2013-05-468223>

Iolascon, A., Russo, R., & Delaunay, J. (2011). Congenital dyserythropoietic anemias. *Current Opinion in Hematology*, 18(3).

Iyama, T., & Wilson III, D. M. (2013). DNA repair mechanisms in dividing and non-dividing cells. *DNA Repair*, 12(8), 620–636.
<http://doi.org/http://dx.doi.org/10.1016/j.dnarep.2013.04.015>

Jameson, D. M., & Sawyer, W. H. (1995). [12] Fluorescence anisotropy applied to biomolecular interactions. In B. T.-M. in Enzymology (Ed.), *Biochemical Spectroscopy* (Vol. Volume 246, pp. 283–300). Academic Press. [http://doi.org/http://dx.doi.org/10.1016/0076-6879\(95\)46014-4](http://doi.org/http://dx.doi.org/10.1016/0076-6879(95)46014-4)

Kamiya, T., & Manabe, A. (2010). Congenital dyserythropoietic anemia. *International Journal of Hematology*, 92(3), 432–438. <http://doi.org/10.1007/s12185-010-0667-9>

Kawa, M. and Machaliński, B. (2014). Hematopoiesis Dysfunction Associated with Abnormal Thyroid Hormones Production, Thyroid Disorders - Focus on Hyperthyroidism, Dr. Gonzalo Diaz Soto (Ed.), InTech, DOI: 10.5772/58531.

Killip, S., Bennett, J., and Chambers, M. Iron Deficiency Anemia. *American Family Physician* 75.5 (2007): n. pag. Web.

Knizewski, L., Kinch, L. N., Grishin, N. V, Rychlewski, L., & Ginalski, K. (2007). Realm of PD-(D/E)XK nuclease superfamily revisited: detection of novel families with modified transitive meta profile searches. *BMC Structural Biology*, 7(1), 40. <http://doi.org/10.1186/1472-6807-7-40>

Kosinski, J., Feder, M., & Bujnicki, J. M. (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(1), 172. <http://doi.org/10.1186/1471-2105-6-172>

Kozyraki, R., Kristiansen, M., Silahtaroglu, A., Hansen, C., Jacobsen, C., Tommerup, N., Moestrup, S. K. (1998). The Human Intrinsic Factor-Vitamin B₁₂ Receptor, *Cubilin*: Molecular Characterization and Chromosomal Mapping of the Gene to 10p Within the Autosomal Recessive Megaloblastic Anemia (*MGAI*) Region *Blood*, 91(10), 3593 LP – 3600.

Kutler, D. I., Singh, B., Satagopan, J., Batish, S. D., Berwick, M., Giampietro, P. F., Auerbach, A. D. (2003). A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood*, 101(4), 1249 LP – 1256. Retrieved from <http://www.bloodjournal.org/content/101/4/1249.abstract>

Lee, SY. *Structure and Inhibition of the Human SNM1A and SNM1B/Apollo DNA Repair Exonucleases*. Thesis. University of Oxford, 2016. Print.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, 362(6422), 709–715.

Loeb, L. A., & Preston, B. D. (1986). Mutagenesis by Apurinic/Apyrimidinic Sites. *Annual Review of Genetics*, 20(1), 201–230. <http://doi.org/10.1146/annurev.ge.20.120186.001221>

Ludwig, H., & Strasser, K. (2001). Symptomatology of anemia. *Seminars in Oncology*, 28, 7–14. [http://doi.org/http://dx.doi.org/10.1016/S0093-7754\(01\)90206-4](http://doi.org/http://dx.doi.org/10.1016/S0093-7754(01)90206-4)

Marks, P. W. and Mitus, A. J. (1996), Congenital dyserythropoietic anemias. *Am. J. Hematol.*, 51: 55–63. doi:10.1002/(SICI)1096-8652(199601)51:1<55::AID-AJH9>3.0.CO;2-6

Martin, L. J. (2008). DNA Damage and Repair: Relevance to Mechanisms of Neurodegeneration. *Journal of Neuropathology and Experimental Neurology*, 67(5), 377–387. <http://doi.org/10.1097/NEN.0b013e31816ff780>

Meetei, A. R., de Winter, J. P., Medhurst, A. L., Wallisch, M., Waisfisz, Q., van de Vrugt, H. J., Wang, W. (2003). A novel ubiquitin ligase is deficient in Fanconi anemia. *Nat Genet*, 35(2), 165–170. Retrieved from <http://dx.doi.org/10.1038/ng1241>

Mehta, AB & Hoffbrand AV. *Haematology at a Glance*. 2nd ed. Chicester, UK: Wiley-Blackwell, 2009. Print.

Merryweather-Clarke, A. T., Atzberger, A., Soneji, S., Gray, N., Clark, K., Waugh, C., Robson, K. J. H. (2011). Global gene expression analysis of human erythroid progenitors. *Blood*, 117(13), e96 LP – e108.

Moser, J., Kool, H., Giakzidis, I., Caldecott, K., Mullenders, L. H. F., & Foulster, M. I. (2016). Sealing of Chromosomal DNA Nicks during Nucleotide Excision Repair Requires XRCC1 and DNA Ligase IIIα in a Cell-Cycle-Specific Manner. *Molecular Cell*, 27(2), 311–323. <http://doi.org/10.1016/j.molcel.2007.06.014>

Naigamwalla, D. Z., Webb, J. A., & Giger, U. (2012). Iron deficiency anemia. *The Canadian Veterinary Journal*, 53(3), 250–256.

Nemeth E, Ganz T. (2006) Heparin and iron-loading anemias. *Haematologica*. 91:727-732

Nemeth, E., Rivera, S., Gabayan, V., Keller, C., Taudorf, S., Pedersen, B. K., & Ganz, T.

(2004). IL-6 mediates hypoferrremia of inflammation by inducing the synthesis of the iron regulatory hormone hepcidin. *Journal of Clinical Investigation*, 113(9), 1271–1276.

<http://doi.org/10.1172/JCI200420945>

Noy-Lotan S, Dgany O, Lahmi R, Marcoux N, Krasnov T, Yissachar N, Ginsberg D, Motro B, Resnitzky P, Yaniv I, Kupfer GM, and Tamary H. (2008). Codanin-1, the protein encoded by the gene mutated in congenital dyserythropoietic anemia type I (CDAN1), is cell cycle regulated. *Haematologica* 2009; 94:629-637. doi:10.3324/haematol.2008.003327

OpenStax, ed. "Erythrocytes." *OpenStax College*. N.p., 28 June 2013. Web.

Petruseva, I. O., Evdokimov, A. N., & Lavrik, O. I. (2014). Molecular Mechanism of Global Genome Nucleotide Excision Repair. *Acta Naturae*, 6(1), 23–34.

Queisser W, Spiertz E, Jost E, Heimpel H. (1971) Proliferation disturbances of erythroblasts in congenital dyserythropoietic anemia type I and II. *Acta Haematol*. 45:65-76.

Räschle, M., Knipscheer, P., Enoiu, M., Angelov, T., Sun, J., Griffith, J. D., Walter, J. C. (2008). Mechanism of Replication-Coupled DNA Interstrand Crosslink Repair. *Cell*, 134(6), 969–980. <http://doi.org/http://dx.doi.org/10.1016/j.cell.2008.08.030>

Recalcati, S., Taramelli, D., Conte, D., & Cairo, G. (1998). Nitric Oxide–Mediated Induction of Ferritin Synthesis in J774 Macrophages by Inflammatory Cytokines: Role of Selective Iron Regulatory Protein-2 Downregulation. *Blood*, 91(3), 1059 LP – 1066. Retrieved from <http://www.bloodjournal.org/content/91/3/1059.abstract>

Renella R, Roberts NA, Brown JM, De Gobbi M, Bird LE, Hassanali T, Sharpe JA, Sloane-Stanley J, Ferguson DJ, Cordell J, Buckle VJ, Higgs DR, Wood WG (2011). Codanin-1 mutations in congenital dyserythropoietic anemia type 1 affect HP1 {alpha} localization in erythroblasts. *Blood* 117: 6928–6938

Rothkamm, K., Krüger, I., Thompson, L. H., & Löbrich, M. (2003). Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Molecular and Cellular Biology*, 23 (16), 5706–5715. <http://doi.org/10.1128/MCB.23.16.5706-5715.2003>

Schwarz, K., Iolascon, A., Verissimo, F., Trede, N. S., Horsley, W., Chen, W., Heimpel, H. (2009). Mutations affecting the secretory COPII coat component SEC23B cause congenital dyserythropoietic anemia type II. *Nat Genet*, 41(8), 936–940.

Shalev H, Kapelushnik J, Moser A, et al. A comprehensive study of the neonatal manifestations of congenital dyserythropoietic anemia type I. *J Pediatr Hematol Oncol*. 2004;26:746–8

Shane, B., & Stokstad, E. L. R. (1985). Vitamin B12-Folate Interrelationships. *Annual Review of Nutrition*, 5(1), 115–141. <http://doi.org/10.1146/annurev.nu.05.070185.000555>

Sinha, R. P., and D. P. Hader. 2002. UV-induced DNA damage and repair: a review. *Photochem Photobiol Sci*. 1: 225-36.

Sparks SE, Krasnewich DM. Congenital Disorders of N-linked Glycosylation Pathway Overview. 2005 Aug 15 [Updated 2014 Jan 30]. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2016.

Steczkiewicz, K., A. Muszewska, L. Knizewski, L. Rychlewski, and K. Ginalski. Sequence, Structure and Functional Diversity of PD-(D/E)XK Phosphodiesterase Superfamily. *Nucleic Acids Research* 40.15 (2012): 7016-045. Web.

Su, A. I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K. A., Block, D., ... Hogenesch, J. B. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America*, 101 (16), 6062–6067. <http://doi.org/10.1073/pnas.0400782101>

Taher, A., Isma'eel, H., & Cappellini, M. D. (2006). Thalassemia intermedia: Revisited. *Blood Cells, Molecules, and Diseases*, 37(1), 12–20. <http://doi.org/http://dx.doi.org/10.1016/j.bcmed.2006.04.005>

Tamary, H., H. Shalev, G. Perez-Avraham, M. Zoldan, I. Levi, D. W. Swinkels, T. Tanno, and J. L. Miller. Elevated Growth Differentiation Factor 15 Expression in Patients with Congenital Dyserythropoietic Anemia Type I. *Blood* 112.13 (2008): 5241-244. Web.

Tiemeier, H., van Tuijl, H. R., Hofman, A., Meijer, J., Kiliaan, A. J., & Breteler, M. M. B. (2002). Vitamin B12, Folate, and Homocysteine in Depression: The Rotterdam Study. *American Journal of Psychiatry*, 159(12), 2099–2101. <http://doi.org/10.1176/appi.ajp.159.12.2099>

Vignard, J., Mirey, G., & Salles, B. (2016). Ionizing-radiation induced DNA double-strand breaks: A direct and indirect lighting up. *Radiotherapy and Oncology*, 108(3), 362–369. <http://doi.org/10.1016/j.radonc.2013.06.013>

Von Lindern, M. (2006). Cell-cycle control in erythropoiesis. *Blood*, 108(3), 781 LP – 782.

Wickramasinghe, S. N. (2000). Congenital dyserythropoietic anemias. *Current Opinion in Hematology*, 7(2).

Wickramasinghe, S. N. and Wood, W. G. (2005), Advances in the understanding of the congenital dyserythropoietic anaemias. *British Journal of Haematology*, 131: 431–446.

doi:10.1111/j.1365-2141.2005.05757.x

7. Appendix

Repeated experiments:

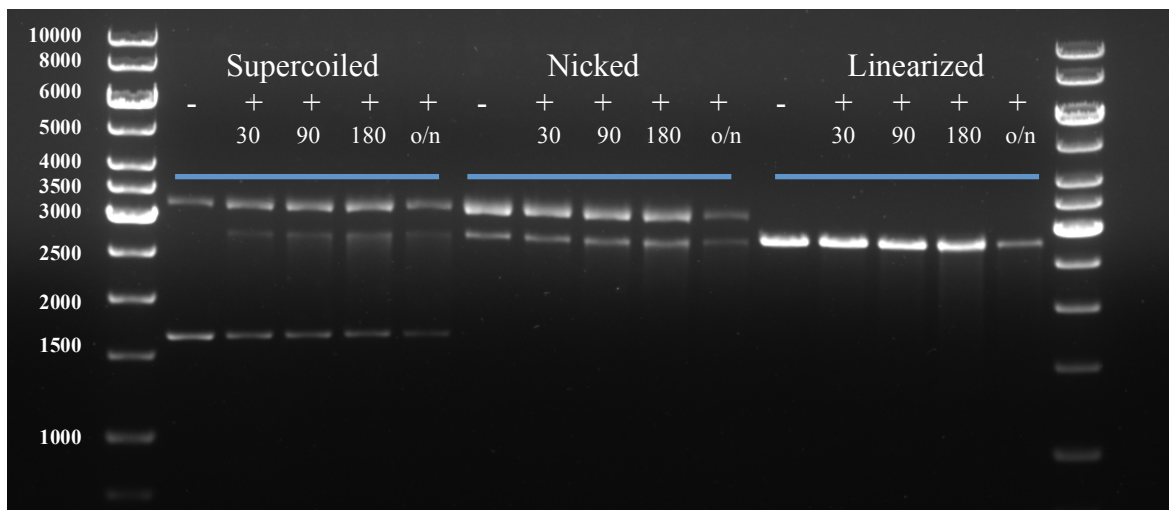
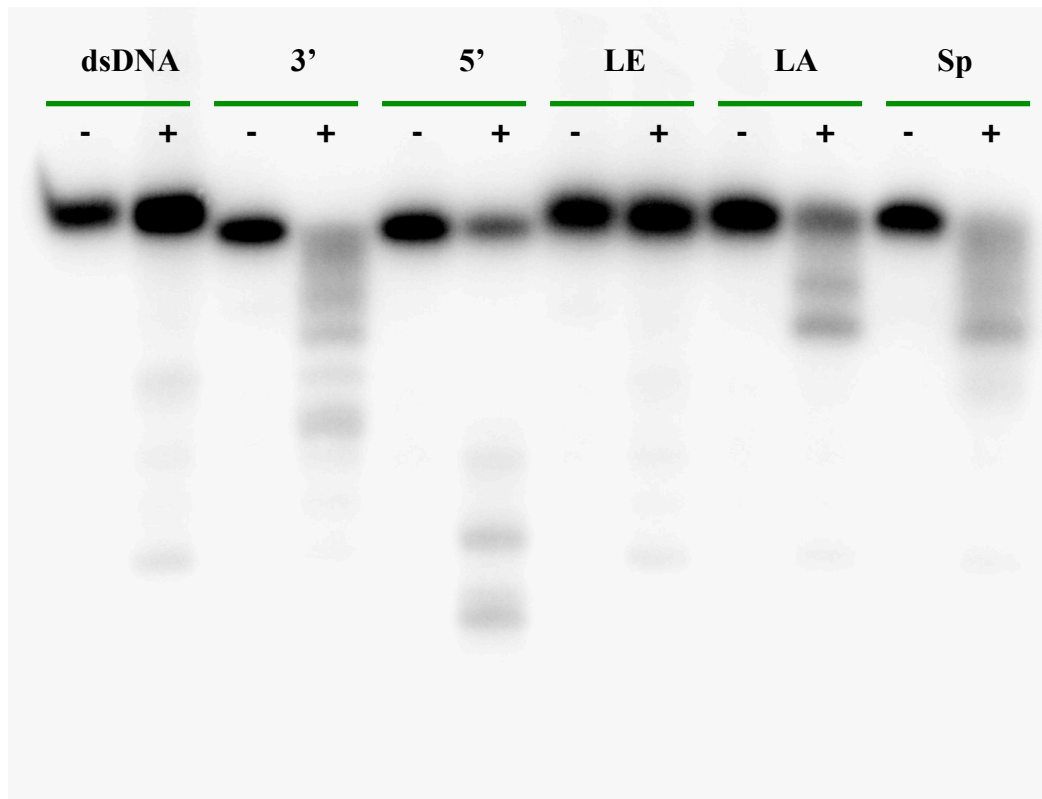
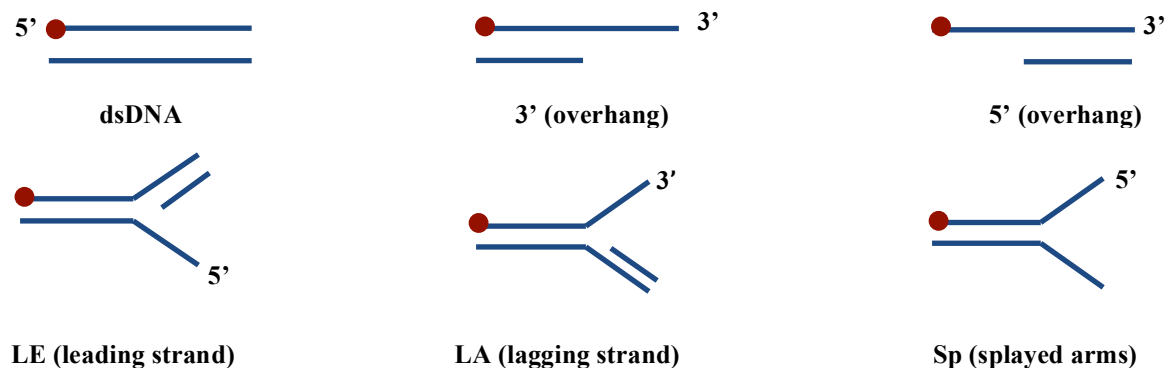


Figure A.1: Plasmid assay of the protein complex incubated with the DNA plasmid (o/n = overnight). The negative (-) lanes are controls that do not contain the protein, and the positive (+) lanes are with the protein in the mixture. The bands in the positive lanes for the supercoiled match those of the nicked plasmid, indicating the protein introduced nicks to the plasmid.



(A)



(B)

Figure A.2: Part A shows a 20% acrylamide gel with the substrates diagrammed in part B after incubation with the protein at 37°C for an hour. All substrates in part B are labelled on the 5' end.