Contribution of Non-Canonical DNA G-quadruplex Structures to Premature Ageing

Denise Liano

A thesis presented for the degree of

Doctor of Philosophy

Department of Chemistry

Imperial College London

December 2021

Declaration of Originality

I, Denise Liano, hereby declare that this Thesis is my own work and reports original results conducted by myself. Any work performed in collaboration with others, both experimentally and intellectually, has been recognised within the text. Sources obtained from others have been referenced and acknowledged.

Denise Liano,

December 2021

Declaration of Copyright

Contribution of Non-Canonical DNA G-quadruplex Structures to Premature Ageing

Denise Liano

Abstract

Previous studies have identified Cockayne Syndrome B (CSB) as a helicase that can resolve non-canonical DNA structures, called G-quadruplexes (G4s). The aim of this study is to investigate the properties of CSB as a G4-binder and -resolvase, and examine the correlation between the G4-helicase activity of CSB and premature ageing phenotype observed in CSB-deficient cells. Accordingly, the recombinant CSB full-length protein (FL) and its helicase-“like" domain (HD) were respectively expressed from insect and bacterial cells, and their resolvase and binding activities were tested over a large panel of DNA substrates. Native gel analysis and biophysical characterisations revealed that ribosomal DNA (rDNA) sequences, that typically act as CSB substrate, can form intermolecular G4s. We discovered that intermolecular G4s were strongly bound by CSB with picomolar affinity, whilst negligible binding to intramolecular G4s was observed. *In vitro* and cellular data demonstrated that G4-ligands can compete with CSB for binding to intermolecular rDNA G4, which results in CSB being displaced off the nucleoli of cells upon treatment with G4-ligands. Immunostaining with the selective G4-antibody BG4 revealed a lack of BG4-staining in the nucleoli of CSB-deficient cells after exogenous expression of recombinant CSB, further corroborating the hypothesis that CSB can bind intermolecular rDNA G4s in the nucleoli and compete with BG4 for the binding of such DNA-substrate. The work presented in this thesis allowed us to observe that (I) intermolecular G4s are likely to form from long-range distant rDNA sequences within the nucleoli of cells, and (II) CSB specifically binds and resolves these structures. Our results provide the first evidence of an endogenous protein that specifically interacts with intermolecular G4s, suggesting potential biological significance of these structures. The biological relevance of intermolecular rDNA G4s could be key in rare genetic disorders like Cockayne Syndrome, where senescence and premature ageing is observed when CSB is functionally mutated.

Acknowledgments

I thank my supervisor Dr. Marco Di Antonio for choosing me as first member of the group. You gave me trust and inspiration. Thank to your endless optimism and enthusiasm. You always picked me up from my insecurities and taught to see the best part of every difficulty. You have been the best supervisor I could have asked for, both personally and scientifically. I will always be proud to have been part of your group from the beginning and saw how warm and welcoming is the group now.

Thanks to all past and present members of the Di Antonio's group for your support during these three years. Thank you to Dr. Aisling Minard for your patience with my English. Thanks to Souro for supporting me throughout the ups and downs of the biochemical experiments and help me with the FRET analysis. Thank you, Jenna, for helping with the ELISAs optimisation. Thanks to Federica for the artistic inspirations.

I would like to thank Anna Di Porzio for our priceless friendship. We shared the same worries during the pandemic and supported each other during the joy and bad moments of our PhDs.

Thank you, Prof. Ramon Vilar, for being my co-supervisor and thanks to all the Vilar's group for sharing the lab during these three years.

Thank you to the proofreaders of my thesis: Marco, Ramon, Dr. Ludovica Monti, and Dr. D. Mager. Thanks for your time and the precious advice.

Thank you to my parents, my brother and my grandparents for your support and endless love. Your presence is constant and reassuring even from far away.

Thank you to my lifelong friends: Martina, Aurora, Maria Chiara and thanks to my university friends, Elena and Rosaria, to be as excited as me when things worked well and for being on my side every day of my life.

Finally, I would like to thank my boyfriend, and now husband Roberto for being my greatest support and for your patience. I know is not easy staying on my side but nevertheless you always encourage me at most.

Presentations

Oral Presentations:

- Nucleic Acids Forum (09/07/2021). Title: “Cockayne Syndrome B protein selectively interacts and resolves intermolecular DNA G-quadruplex structures”.
- Department of Chemistry PG Symposium, Imperial College London (Chemical Biology & Healthcare Theme) (07/07/2021). Title: “Cockayne Syndrome B protein selectively interacts and resolves intermolecular DNA G-quadruplex structures”.
- Chemical Biology and Drug Discovery Seminar (16/03/2021). Title: “CSB as an endogenous interactor of intermolecular G4s formed within ribosomal DNA”.
- G4-webinar (21/01/2021). Title: “Cockayne Syndrome B protein (CSB) interacts selectively with ribosomal DNA G-quadruplex secondary structures”.
- ICB CDT Nucleic Acid Chemical Biology Colloquium (30/08/2019). Title: Combining small-molecule and CRISPR-Cas9 for gene selective G-quadruplex stabilisation.

Poster Presentations:

- Poster presentation at the Department of Chemistry PG Symposium, Imperial College London (Chemical Biology & Healthcare Theme) (July 2020). Title: “Biochemical, structural and genomic implications of DNA G-quadruplex formation in premature ageing”. (**Awarded best poster prize**)

Publications

Liano, D., Chowdhury and S., Di Antonio, M. Cockayne Syndrome B protein selectively resolves and interacts with intermolecular DNA G-quadruplex structures. *J. Am. Chem. Soc.* 143, 20988-21002 (2021).

Robinson, J., Raguseo, F., Nuccio, S. P., **Liano, D**. and Di Antonio, M. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.* 49, 8419-8431 (2021).

Minard, A., Morgan, D., Raguseo, F., Di Porzio, A., **Liano, D**., Jamieson, A. J. and Di Antonio, M. A short peptide that preferentially binds c-MYC G-quadruplex DNA. *Chem. Commun. 56,* 8940-8943 (2020).

Minard, A., **Liano, D.**, Wang, X. and Di Antonio, M. The unexplored potential of quinone methides in chemical biology. *Bioorganic & Medicinal Chemistry.* 27, 2298-2305 (2019).

Abbreviations

5'– UTR - 5'-untranslated region

A – Adenine

aa – Amino acids

AD – Acidic amino acid stretch

ACF – ATP-dependent chromatin assembly factor

AcMNPV – *Autographa californica* multiple nuclear polyhedrosis virus

AP – Apurinic site

APE1 – Apurinic/apyrimidinic endonuclease

ATP – Adenosine triphosphate

BER – Base excision repair

bp – Base pair

BLM – Bloom syndrome helicase

BRCA1/2 – Breast cancer gene 1/2

BRACO-19 – *N*-[9-[4-(dimethylamino) anilino] -6-(3-pyrrolidin-1-ylpropanoylamino) acridin-3-yl] -3-pyrrolidin-1-ylpropanamide

BSA – Bovine serum albumin

C – Cytosine

CD – Circular dichroism

ChIP-Seq – Chromatin immunoprecipitation followed by high-throughput DNA sequencing

CIM – CSA-interaction motif

CNBP – Cellular nucleic acid-binding protein

cryo-EM – Cryo-electron microscopy

CS – Cockayne Syndrome

CS1AN – Cockayne Syndrome 1 ANn Arbor cells

CSA – Cockayne Syndrome A

CSB – Cockayne Syndrome B

CSB-FL – CSB full-length protein

CSB-HD – CSB helicase- “like” domain alone

CX-5461– 2-(hexahydro-4-methyl-1H-1,4-diazepin-1-yl)-N-[(5-methyl-2-pyrazinyl) methyl]-5-oxo-5H-benzothiazolo[3,2-a] [1,8] naphthyridine-6-carboxamide

Cy5 – Cyanine-5

DAOTA – Diazaoxatriangulenium

DDB1– DNA damage-binding protein 1

DDR – DNA damage response

DEAH/RHA – Asp-Glu-Ala-His and RNA helicase A (*RHA)*

DHX36 – Helicase protein DEAH-box helicase 36

DMEM – Dulbecco's Modified Eagle Medium

DNA – Deoxyribonucleic acid

DNMT1 – (cytosine-5)-DNA methyltransferase 1 enzyme

DOG1 – Gene discovered on gastrointestinal stromal tumors: GIST

ds – Double strand

DSBs – Double strand breaks

DTT – Dithiothreitol

EDTA – Ethylenediaminetetraacetic acid

EGFP – Enhanced green fluorescent protein

El – Elution

ELISA – Enzyme-linked immunosorbent assay

EMSA – Electrophoretic mobility shift assay

FAM – 6-FAM (Fluorescein)

FANCJ – Fanconi anaemia complementation group J

FBS – Fetal bovine serum

FLIM – Fluorescence lifetime imaging microscopy

FP – Fluorescence polarisation

FRET – Fluorescence resonance energy transfer

FT – Flow through

G4 – G-quadruplex

GFP – Green fluorescence protein

G4LDB – G4-ligand database

G4-Seq – Sequencing of G-quadruplexes

G – Guanine

γH2AX – Phosphorylation of histone H2AX on Ser-139

GMP – Guanylic acid

H3K4me3 – Trimethylation of histone H3 at Lysine 4

H3K9me3 – Trimethylation of histone H3 at Lysine 9

HA – Hemagglutinin antigen

HeLa – Henrietta Lacks' cells (human cervical cancer cells)

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

His_6 – 6X Histidine tag

HT1080 – Human fibrosarcoma cell line

hTERT – Human telomerase reverse transcriptase

Ig – Immunoglobulin

IMAC – Immobilised metal affinity chromatography

IPTG – Isopropyl ß-D-1-thiogalactopyranoside

iPyPDS – PyPDS isomer

Kb – Kilobase

Kd – Dissociation constant

kDa – Kilo Dalton

KU – Kilo Units

MAZ – Myc-associated zinc-finger protein

mRNA – Messenger ribonucleic acid

MW – Molecular weight

NAD^+ – Nicotinamide adenine dinucleotide

NHEJ – Non-homologous DNA end-joining repair mechanism

NLS – Nuclear localisation sequence

NM23-H2 – Non-metastatic cell 2 protein

NMM – N-methylmesoporphyrin IX

NMR – Nuclear magnetic resonance

NoLS – Nucleolar localisation sequence

NTHL1 – Nth like DNA glycosylase 1

OG – 8-oxo-7,8-dihydroguanine

PARP1 – Poly-ADP ribose-polymerase 1 enzyme

PBS – Phosphate-buffered saline

PCR – Polymerase chain reaction

PDS – Pyridostatin

Phen-DC3 – 3,3′-[1,10-Phenanthroline-2,9-diylbis(carbonylimino)] bis[1-methylquinolinium] 1,1,1-trifluoromethanesulfonate

PMSF – Phenylmethylsulphonyl fluoride

POT1 – Protection of telomeres 1

PQS – G4-forming sequence

rDNA – ribosomal DNA

RF-cloning – Restriction-free cloning

RNAPI – RNA polymerase I

RNAPII – RNA Polymerase II

ROS – Reactive oxygen species

rRNA – ribosomal RNA

SBS – Sequencing-by-synthesis

SD – Standard deviation

SDS – Sodium dodecyl sulphate

SDS-PAGE – Sodium dodecyl sulphate-polyacrylamide gel electrophoresis

SEC – Size exclusion chromatography

SF – Soluble fraction

SFM – Scanning force microscopy

SiR-PyPDS – Silicon-rhodamine (SiR), pyridostatin (PyPDS)

ss – Single strand

SSBs – Single-strand DNA breaks

SUMO-tag – Small ubiquitin-like modifier-tag

SWI/SNF – Switch/sucrose non-fermentable

T7-RNAP – T7 RNA polymerase

T – Thymine

TAMRA – Tetramethylrhodamine

TBE – Tris-borate-EDTA

TCEP – Tris(2-carboxyethyl) phosphine

TC-NER – Transcription coupled nucleotide excision repair

TCR – Transcription-coupled repair pathway

TDS – Thermal difference spectrum

TF – Total fraction

TFs – Transcription factors

TFIIH – Transcription factor IIH

TMEJ – DNA Polymerase theta (Pol θ) mediated end-joining repair mechanism

TMPyP2 – 5,10,15,20-tetra-(N-methyl-2-pyridyl) porphine

TMPyP4 – 5,10,15,20-tetra-(N-methyl-4-pyridyl) porphine

TRF2 – Telomeric repeat-binding factor 2

Tris-HCl – Tris(hydroxymethyl)aminomethane hydrochloride

TSS – Transcription start site

U2OS – Human bone osteosarcoma epithelial cells

UBD – Ubiquitin binding domain

UV – Ultraviolet

UVSSA – UV stimulated scaffold A

VEGF – Vascular endothelial growth factor

W – Wash

WB – Western blot

WDR61 – WD Repeat Domain 61

WHD – Winged-helix domain

WRN – Werner syndrome helicase

WT – Wild type

x-Gal – 5-Bromo-4-Chloro-3-Indolyl-beta-D-Galactoside

Table of Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

1.1 DNA Structure and Heterogeneity

Deoxyribonucleic acid (DNA) is a macromolecule that carries the genetic information that plays essential roles in the life of all living organisms, including humans, bacteria, and viruses. A single strand (ss) of DNA contains one of four possible nitrogen bases: the purines Adenine (A) and Guanine (G), and the pyrimidines Thymine (T) and Cytosine (C) (Figure 1.1A). The four bases establish unique interactions where A pairs with T, and C with G [(1)] (Figure 1.1B). Each base is also attached to a sugar moiety functionalised with a phosphate group to form a nucleotide (Figure 1.1A). Nucleotides can polymerise one with the other to form long chains of ssDNA. When two complementary strands of DNA interact through hydrogen bonds between the bases, they coil around each other to form a right-handed spiral called double helix (Figure 1.1C). The first model of the DNA's α-helix conformation was proposed in the early 1950's by R. Franklin, J. Watson and F. Crick. The suggested model is known as 'Watson-Crick DNA topology' (Figure 1.1B and C) and it represents most of the DNA found in the nucleus of all eukaryotic cells [(2,3)]. The double helix described by Watson and Crick is also known as the B-form of DNA. The characteristic features of the B-form are a helical diameter of approximately 2 nm, a rise of 0.34 nm per base pair, and each helical twist corresponds to 10 base pairs (bp) per turn (Figure 1.1C). The cellular DNA is generally very long, which can extend up to 2 meters. Therefore, DNA is packaged in a compact form that allows its containment into the cell. The fundamental units of eukaryotic DNA packaging are known as nucleosomes [(4-6)]. Nucleosomes consist of a histone core, formed by two pairs of each of the histone proteins H2A, H2B, H3, and H4, that is surrounded by double stranded DNA. Approximately 150 bps of dsDNA are wrapped around each histone core [(7,6)]. Nucleosomes are then organised into higher-level compaction by H1 histone-associated proteins [(8,9)]. The high degree of compaction reduces the DNA accessibility and ensures maintenance of an inactive state of DNA, also called heterochromatin. The organisation of the nucleosomes across the genome is highly dynamic and regulated by specific factors and post-translational modifications that allow the continuous transition of the DNA between heterochromatin and euchromatin, the latter being a more accessible and relaxed state of chromatin. Open and more accessible single stranded (ss) DNA regions are associated to active processes where cellular machineries, for instance, replication or transcription factors, can interact and modulate the activity of a particular gene [(10,11)].

In addition, depending on the base composition of specific genomic sequences, such as the presence of repetitive regions or high CG content, the more accessible ssDNA topology can be rearranged into non-canonical DNA secondary structures [12]. For example, G-quadruplexes (G4s) [13-15], triplex DNA [16], and i-motifs [17] (Figure 1.1E-G) have been shown to have crucial roles in gene expression regulation and other essential cellular processes (e.g. replication, transcription and translation) [18].

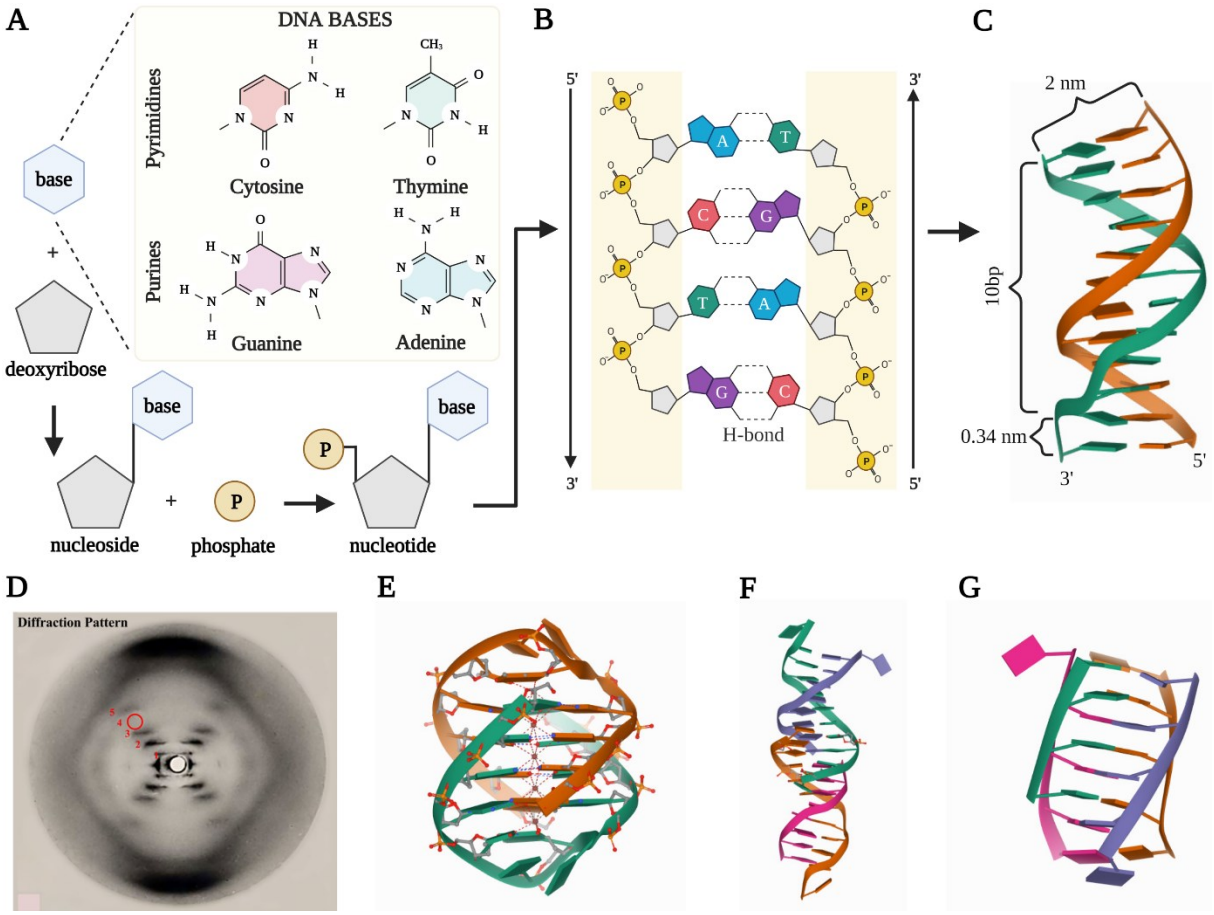

Figure 1.1. DNA structures. (**A**) Each single nucleotide that constitutes the DNA filament is made by three components: a nitrogen base (i.e. Adenine, Thymine, Guanine, or Cytosine), a deoxyribose sugar, and a phosphate group. This representation was redrawn and adapted from Pray et al. [19]. (**B**) Representation of the nucleotides interacting with each other through hydrogen bonds (H-bond). Adenine (A, blue) interacts with Thymine (T, green), while Cytosine (C, red) interacts with Guanine (G, purple). (**C**) Crystal structure of a dsDNA fragment (PDB ID: 1D28, 2.70 Å-resolution) [20]. (**D**) X-shape or distorted rhombus of the X-ray diffraction partner of DNA obtained by R. Franklin (adjusted from Thompson et al.) [21]. (**E**) Crystal structure of a G4 structure (PDB ID: 2HBN, 1.55 Å-resolution) [22]. (**F**) Crystal structure of a triplex DNA (PDB ID: 1D3R, 1.80 Å-resolution) [23]. (**G**) Crystal structure of an i-motif (PDB ID: 1CN0, 2.20 Å-resolution) [24].

1.2 G-quadruplex DNA Secondary Structures

G-quadruplexes (G4s) are non-canonical DNA (or RNA) secondary structures that can arise from single stranded (ss) guanine-rich DNA sequences [25,26]. In 1962, M. Gellert and colleagues reported the first evidence of aggregation of guanylic acid (GMP). Examination of the optical properties and the X-ray diffraction pattern of these aggregates (Figure 1.2A) suggested helical formation by the GMP where four guanines (Gs) interact with each other through hydrogen bonding. Based on this data, the authors proposed a unique association of the four Gs that results in the formation of a planar tetrameric structure, which can stack one on top of another to form a stable helical aggregate [27]. Later, D. Sen and W. Gilbert used synthetic ssDNA sequences extracted from G-rich immunoglobulin (Ig) switch regions to observe the formation of aggregates able to move towards higher molecular weight on a non-denaturing gel under physiological salt conditions. These results further suggested that the G-rich sequences were associated in tetrameric inter-strand aggregates, referred to as G4s [25]. Moreover, fibre diffraction analyses confirmed that Hoogsteen hydrogen-bonding and cations coordination between the Gs stabilise the four G-rich strands, thus forming a characteristic planar tetrameric structure, also known as G-tetrad [28], which is the core element of a G4 structure.

The first crystal structure of a telomeric G4 (Figure 1.2B and C) [29] confirmed that ss G-rich sequences containing at least one G-stretch can form G-tetrads, and that two or more G-tetrads can stack one on top of the other generating the G4 scaffold (Figure 1.2D and E) [13,15]. Oxygen atoms (O) of Gs are oriented to the center of each G4-tetrad. Consequently, the resulting G4 scaffold presents a tubular empty space in the center, named ion channel. When an alkali metal cation is placed in this ion channel, it interacts with the O atoms promoting the stabilisation of the G4 structure [30]. Different monovalent cations have different efficiencies in stabilising G4 structures. Specifically, K^+ is the most potent stabiliser compared to Na^+, Rb^+, or Li^+, which is the least effective cation promoting G4s stabilisation [14,31].

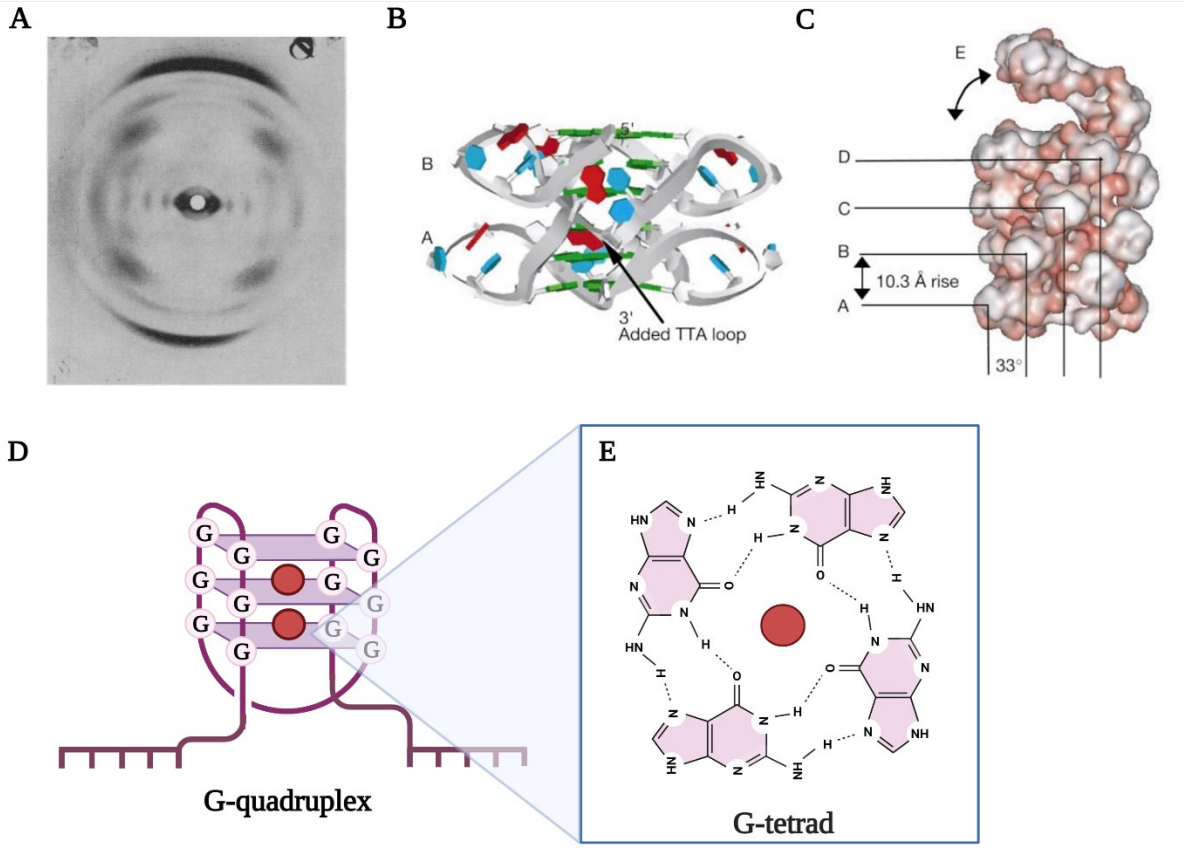

Figure 1.2. G4s formed from ss G-rich sequences. (**A**) X-ray diffraction pattern of a 5'-GMP. Image taken and adapted from M. Gellert et al. [(27)]. (**B**) Model of two telomeric G-tetrads (A and B) stacked 3' to 5' with an unmodified TTA loop modelled to link them (modified from Parkinson et al. [(29)]). (**C**) Model of the human telomeric G4 where four G-tetrads stack one on the top of the other (A to D) and a fifth G-tetrad (**E**) is folding onto the stack (adapted from Parkinson et al. [(29)]). (**D**) Schematic representation of a G4 structure arising from a ss G-rich sequence with a zoomed G-tetrad. (**E**). In each G-tetrad, four guanines (showed in pink) are associated through Hoogsteen hydrogen bonds (lanes between Gs). Stabiliser monovalent cations are represented as red spheres [(30)].

1.2.1 G-quadruplex DNA Structural Features

Numerous *in vitro* studies revealed that G4s can adopt different conformations, depending on the length and composition of the G-rich single stranded sequence of DNA used. Moreover, environmental conditions, such as local molecular crowding or the presence of different metal cations, can influence the topology of G4s making the structures very heterogeneous [(18,32,33)]. G4s could be potentially formed from a single G-rich sequence, classified as unimolecular or intramolecular G4s (Figure 1.3A). Typically, a unimolecular G4-forming sequence is described as:

$G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ [34,35]

Where the number of Gs in each G-tract should be ≥3 [35,36], while N is any combination of bases (including G) between 1 and 7 that are not involved in the formation of the G-tetrad but form a loop (Figure 1.3A and B) [32,35,37]. This description has been revised since first proposed as stable G4 structures that contain runs of two Gs (rather than a minimum of 3) have been reported. Thus, a minimum of two stacked quartets are required, but G4 stability generally increases with additional quartets [38].

When two or more distinct G-rich sequences interact with each other forming a G4 structure, the G4 is classified as multimeric (dimeric or tetrameric) or intermolecular (bimolecular or tetramolecular) (Figure 1.3B). *In vitro*, bimolecular or tetramolecular G4s can originate from the association of two or four identical or different G-rich sequences [32], respectively. However, the cellular environment is enriched of condensed heterochromatin, therefore, the probability of multiple long-range distant G-rich sequences to assemble has often been considered highly unlikely. For this reason, most of the G4 research has focused on intramolecular G4s, with only a few intermolecular G4s biologically characterised to date [32,39].

Depending on the orientation of the DNA strands, G4 structures can adopt a variety of topologies. Parallel G4s present all the G-rich strands oriented in the same direction; antiparallel G4s are characterised by either two strands with opposite directionality or two strands oriented in the same direction and the other two in the opposite direction; the hybrid-type (or mixed) topology has only one strand or one G-stretch oriented in an opposite direction compared to the other three strands or G-stretches (Figure 1.3C-E) [15,32,40].

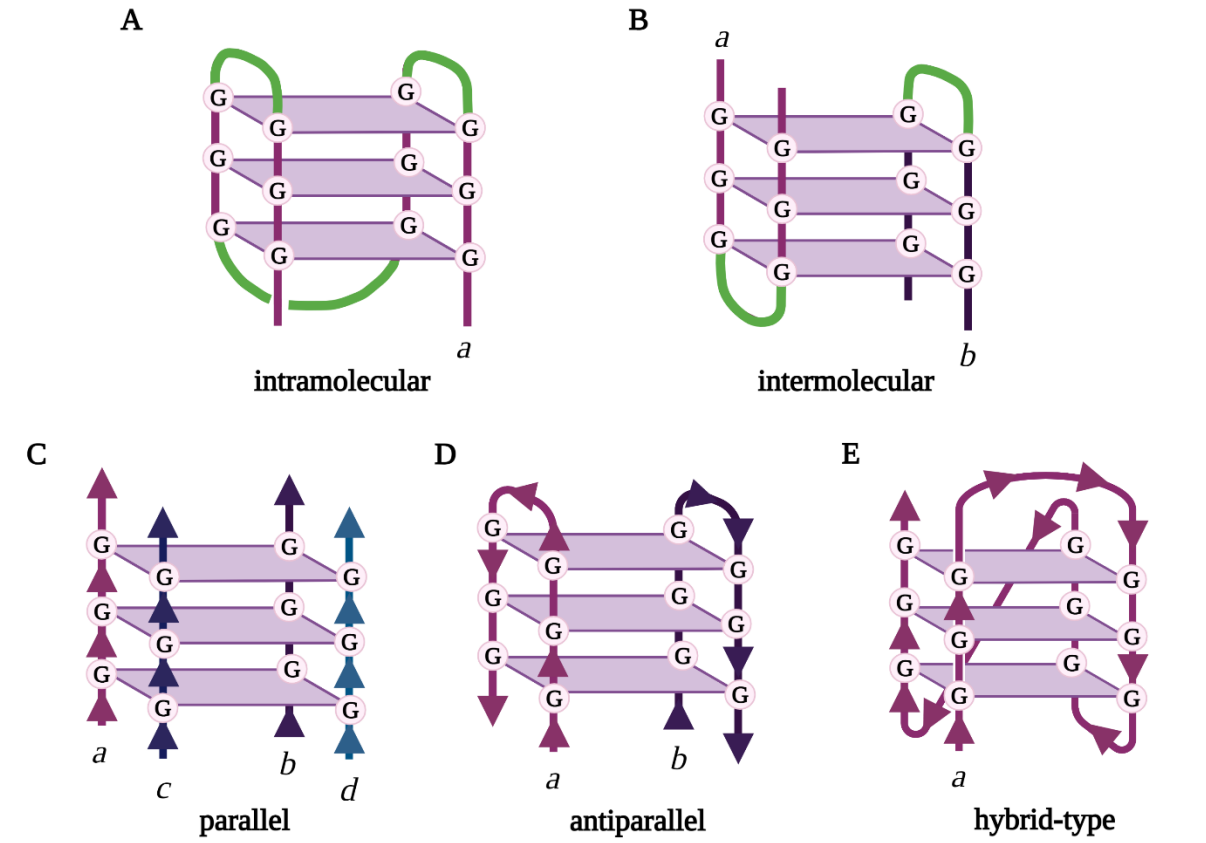

Figure 1.3. G4s different topologies. (**A**) Schematic representation of an intramolecular (unimolecular) G4, which arises from the same G-rich ss sequence (a, in light purple). The loops are indicated in green. (**B**) Schematic representation of an intermolecular (dimeric) G4, which arises from two different G-rich ss sequences (a and b, in light purple and dark purple, respectively). The loops are indicated in green. (**C**) Schematic representation of a parallel intermolecular (tetrameric) G4, where all the strands (a, b, c and d, in light purple, dark purple, blue, and light blue, respectively) are oriented in the same direction. (**D**) Schematic representation of an antiparallel intermolecular (dimeric) G4, where the ss G-rich sequences (a and b, in light purple and dark purple, respectively) are oriented in opposite directions. (**E**) Schematic representation of a hybrid-type (unimolecular) G4, where three G-stretches are oriented in parallel and the other one in antiparallel (in a, light purple).

1.2.2 Methods to Assess G-quadruplex Topology and Structural Conformation

Over the years, numerous methods have been developed to study the formation of G4s and their structural features. For example, ultraviolet (UV) spectroscopy is used to confirm the formation of a G4 by analysing the typical UV curves obtained at 295 nm [41,42]. Moreover,

circular dichroism spectroscopy (CD) can be used to understand the nature of the G4-topology[40]. Despite the UV and CD techniques allowing for a rapid assessment of the G4 formation and its main conformational characteristics, only NMR spectroscopy and X-ray crystallography allow to fully characterise the G4 structures with atomic resolution [42,43].

ULTRAVIOLET (UV) SPECTROSCOPY

UV-spectroscopy is a technique used to confirm the formation of G4 structures and evaluate their stability. Commonly, the maximum absorption wavelength of nucleic acids is observed at 260 nm. A difference of approximately 25% of absorption is observed between folded or unfolded (melted) dsDNA. Contrarily, the absorption difference between folded and unfolded G4s is not so evident and the precise determination of melting temperature of G4s is not accurate when recorded at 260 nm. The quality of the G4 melting profiles was optimised by Mergny and colleagues who observed a 50% change in absorbance amplitude using a wavelength of 295 nm [44]. Although absorbance of the G4 structure at 295 nm is lower compared to its absorbance at 260 nm, G4s undergo a hypochromic shift (i.e. decrease in the UV absorbance) at 295 nm upon melting (Figure 1.4A). Measurements of melting profiles under various conditions allow to characterise the G4 structure and determine the thermodynamic parameters that influence its formation and stability [42]. Subtraction of the UV spectra above and below the G4 melting temperature such as, 90 °C and 20 °C, respectively, allows the generation of a thermal difference spectrum (TDS), which is unique for each type of nucleic acid structure. A typical TDS of G4s is characterised by a negative peak at 295 nm and positive peaks approximately at 240 nm and 270 nm (Figure 1.4B). The negative peak is usually less intense for parallel G4s compared to that of the antiparallel one. The main drawback of this technique is that other DNA secondary structures can form the negative peak at 295 nm for instance, i-motifs, Hoogsteen duplexes, and pyrimidine triplex structures. Consequently, UV spectroscopy alone is not sufficient to prove G4-formation, which should be confirmed by additional techniques [42,45].

CIRCULAR DICHROISM (CD)

Circular dichroism (CD) is an absorption spectroscopy originated when linearly polarised light passes through optical active chiral molecules. The chiral molecule can preferentially

absorb a right- or left-hand direction of the circularly polarised components and this difference in absorption can be measured [46]. CD spectroscopy has been widely used to determine the secondary structure of proteins and it is one of the primary methods used to validate the formation of G4s in potential G4-forming sequences and to determine the topology of G4 structures. Indeed, unique CD spectral signatures are associated to the different orientation of the strands that constitute the G4 structure [40,46]. Further, observation of multiple G4s of known topologies allowed to establish key CD spectral signatures, which have been widely accepted for the characterisation of G4s (Figure 1.4C):

≈*264 nm max, ≈245 nm min* is associated to parallel topology;

≈*295 nm max, ≈260 nm min* identifies the antiparallel topology;

≈*295 nm max, ≈260 nm max, ≈245 nm min* defines the hybrid G4[40].

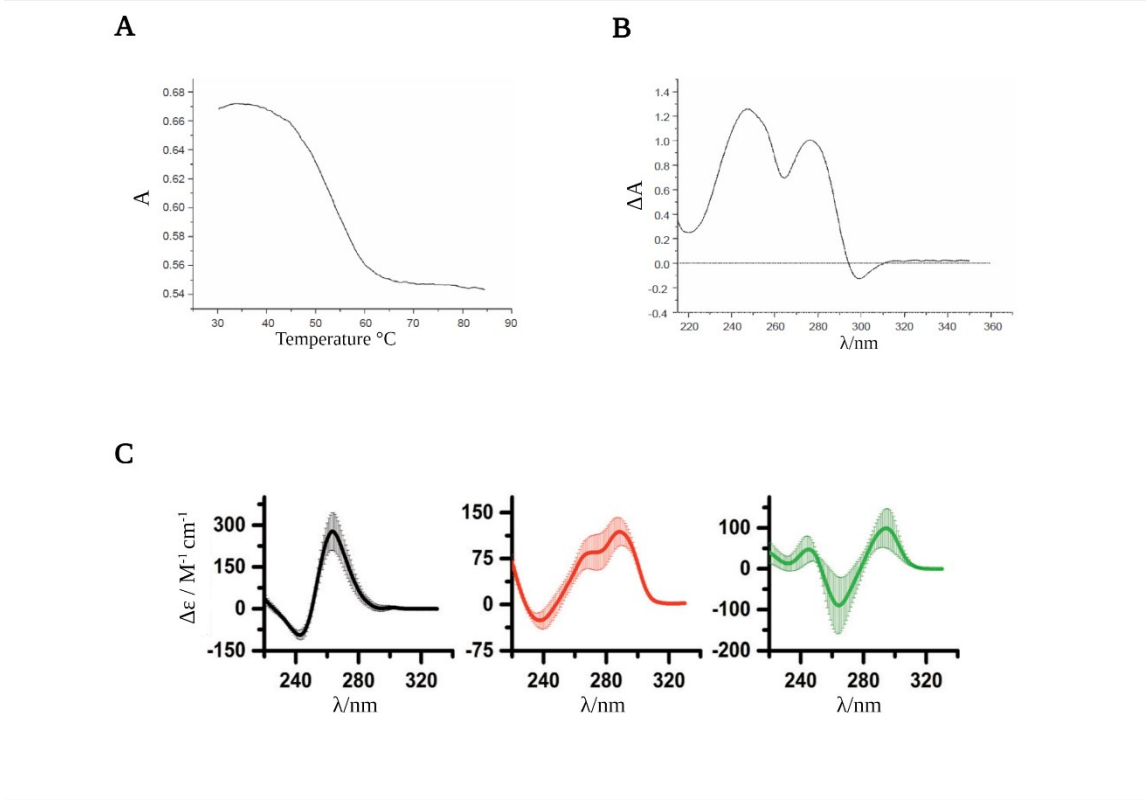

Figure 1.4. Methods to assess G4 topology and conformation. (**A**) UV denaturation profile (hypochromic profile) of a G4 structure. Figure adapted from Małgowska et al. [42]. (**B**) Thermal difference spectrum (TDS) of a G4 structure. Figure adapted from Małgowska et al. [42]. (**C**) Reference CD spectra referring to parallel (black), hybrid-type (red), and antiparallel (green) G4s. Figure adapted from R. del Villar-Guerra et al. [40].

NUCLEAR MAGNETIC RESONANCE (NMR) and X-RAY CRYSTALLOGRAPHY

Nuclear Magnetic Resonance (NMR) spectroscopy and X-ray crystallography are distinct methods used to investigate the structural conformation of molecules. NMR spectroscopy analyses the alignment of the nuclei in an atom by using strong local magnetic fields, while X-ray crystallography determines the three-dimensional structure of a crystallised molecule using X-ray diffraction [43]. To obtain the desired structural information, the two techniques require kinetically stable homogeneous species in solution. Importantly, the presence of multiple species in the same solution decreases the probability to obtain accurate information. To overcome this limitation, the G4-forming sequence can be modified to promote the formation of a unique and stable G4 structure in solution, however, using a non-endogenous sequence may bias the structural analysis. Indeed, various structures can be resolved after mutations of a G4-forming sequence that may differ from endogenous ones. Therefore, the results obtained from NMR and X-ray analyses need to be carefully evaluated based on such commonly introduced artificial mutations [32].

Nevertheless, thanks to these techniques, many G4s structures have been characterised over the years. For example, in 2004 Phan and colleagues reported the NMR structure under K^+ stabilisation of a very stable intramolecular parallel G4 that originates from the G-rich promoter region of *c-myc* oncogene [47,48], named c-MYC G4. More recently, Stump and co-workers reported the crystal structure of the same c-MYC G4, which is, overall, in agreement with the previously reported NMR structure (Figure 1.5A and B) [49]. As indicated by Stump and colleagues, the main difference between the NMR and the crystal structures resided in the 5'-head of the G4, where the crystal structure showed a more extended 5'-head compared to that of the NMR structure (Figure 1.5 A and B). The 3'-tail is also more extended in the crystal structure compared to that of the NMR structure, confirming the flexibility of the 5'-head and the 3'-tail [49].

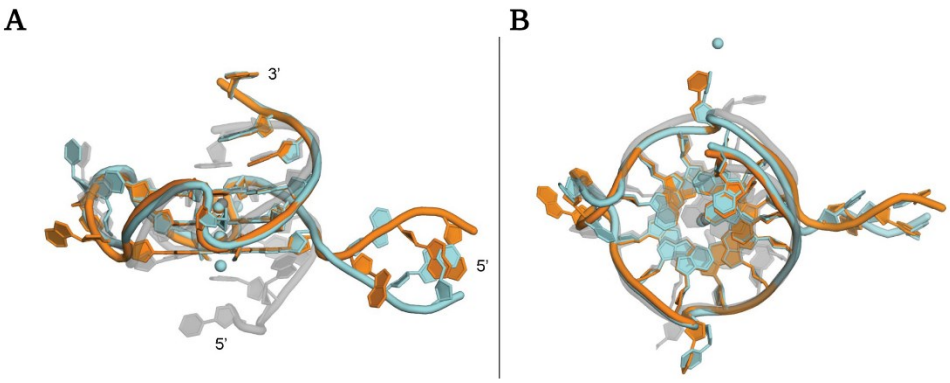

Figure 1.5. Comparison between the NMR and crystal structures of c-MYC G4. This figure has been modified from Stump et al. [49] and shows the overlay of the c-MYC G4 NMR structure (PDB ID:1XAV, in grey [48]) with the crystal structure (2.35 Å) of two independent c-MYC G4s-forming strands (represented in orange and cyan). The side view is shown in (**A**) while the top view is shown in (**B**).

1.3 G-quadruplex DNA Mapping in Chromatin

The ability of G-rich telomeric sequences to form G4s *in vitro* was first demonstrated in *Tetrahymena* and *Oxytricha* in the late 1980s [50,51]. The increasing number of evidence demonstrating the formation of G4 structures *in vitro* under physiological salt conditions has prompted the research community to develop new methods to visualise and map endogenous G4s also in human cells. Therefore, different techniques have been recently developed to predict, visualise, and assess G4s across the genome of multiple species. These studies suggested that the G4 structures are not randomly distributed across the genome of different species, but rather enriched in specific functional regions such as telomeres and repeat expansions associated with diseases and gene-promoters. In addition, the high evolutionary enrichment of these structures might reflect their biological roles in regulating essential cellular pathways, and potential as novel targets for therapeutic intervention.

1.3.1 Computational Prediction

The first algorithm able to predict putative G4-forming sequences (PQSs) in a genomic context was Quadparser, developed by J. L. Huppert and S. Balasubramanian in 2005 [35]. Quadparser is based on the folding rule $G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$ [34,35] described in Section 1.2.1. One year later, J. Eddy and N. Maizels reported a similar approach to predict the ability of a PQS to form G4 structures, which is also based on the recurrence of repeating G_n units ($n \geq 3$) [52]. These computational approaches predicted over 370,000 PQSs in the human genome [35]. Further, these methods not only detected enrichment of G4s at telomeres, but also at immunoglobulin gene class-switch recombination sites and other regulatory regions, such as promoters of genes and oncogenes where the chromatin is open and G4s may be involved in transcriptional regulation [35, 52-54]. These algorithms can directly predict the potential G4 formation from the primary DNA sequence, giving a simple ‘yes’ or ‘no’ answer based on a fixed number of possible consecutive G-tract numbers (4) and length (≥3) or on a limited maximum loop size (N_{1-7}). However, the algorithms do not consider other genomic or functional parameters (like molecular crowding or base modifications), therefore, these approaches can only provide a global picture of the G4-landscape. In fact, many sequences

not detected have been demonstrated to form G4s *in vitro* (false negative), while few sequences predicted to form G4s do not (false positive) [54].

More recently, J-L. Mergny and colleagues developed a different algorithm, called G4-Hunter. G4-Hunter predicts the potential of a genomic sequence to form G4 structures by providing a score of the G4-propensity as output. Such G4-score is calculated based on *G-richness* and *G-skewness* of a given sequence, where *richness* is the number of Gs in the sequence, while *skewness* is the G/C asymmetry between the complementary strands. Moreover, G4-Hunter allows the optimisation of the search by choosing different thresholds and window values. Despite the significantly higher number of stable G4s predicted with this algorithm (2-10 times) compared to that predicted by previous computational approaches, G4-Hunter is also a context-independent approach that predicts stable G4-formation without considering the grade of stability given by loops and genomic localisation of the G4-structure [54].

1.3.2 High-throughput Sequencing of G-quadruplexes (G4-Seq)

Ten years after the development of Quadparser, S. Balasubramanian's group introduced a high-resolution sequencing-based method to experimentally detect G4s in the human genome, which allowed to overcome limitations associated with the use of computational approaches to detect G4s on a genomic scale [55]. High-throughput sequencing of G4s (G4-Seq) is used for mapping the formation of G4 structures within single stranded purified human genomic DNA sequences, and relies on polymerase stalling [56] in the context of next generation sequencing [57].

Polymerase stop assays have been widely exploited to show the ability of G-rich sequences to form G4 structures under stabilising K^+ conditions [58]. This assay leverages the presence of a G4 in a ssDNA template to act as a knot blocking the activity of a DNA polymerase in processing the template, stalling DNA synthesis in a G4-dependent fashion. By gel electrophoresis, the prematurely terminated DNA products run faster than fully elongated ones do. Therefore, this assay allows a rapid assessment of G4-formation on a gel and analyses the conditions that affect their formation and stability (Figure 1.6A) [56,58].

In addition to polymerase stalling, G4-Seq relies on a modified version of the classical Illumina next generation sequencing. In a standard Illumina sequencing experiment, the DNA or RNA sequences are prepared through fragmentation of genomic DNA to a desired length (usually 100–5,000 bp). Oligonucleotide adapters are then attached to the end of the target sequences, while the ligated fragments are PCR amplified and gel purified to generate a sequencing library. This library can be attached to a surface through complementarity with the adaptors and amplified through bridge amplification cycles. Finally, the amplified library is sequenced through a sequencing-by-synthesis (SBS) method where, after the addition of primers, the templates are sequenced by repeated cycles of polymerase-directed single base extension using nucleotides with base-specific fluorescent markers (i.e. Illumina Sequencing) [59]. In a G4-seq experiment, the templates are sequenced twice. The first sequencing (read 1) is the accurate sequencing of the template that would normally be performed during the standard Illumina sequencing using a mid G4-stabiliser Na^+ buffer, while the second run (read 2) is performed under conditions that promote G4 formation such as the use of K^+ buffers or after addition of molecules that stabilise the G4 structures. The formation of a G4 structure arrests the polymerase extension and results in mismatches between reads 1 and 2, with reduced sequencing quality from the G4-forming point to the end of the sequence, which can be used to detect genomic fragments capable to fold into stable G4s (Figure 1.6B) [55,57].

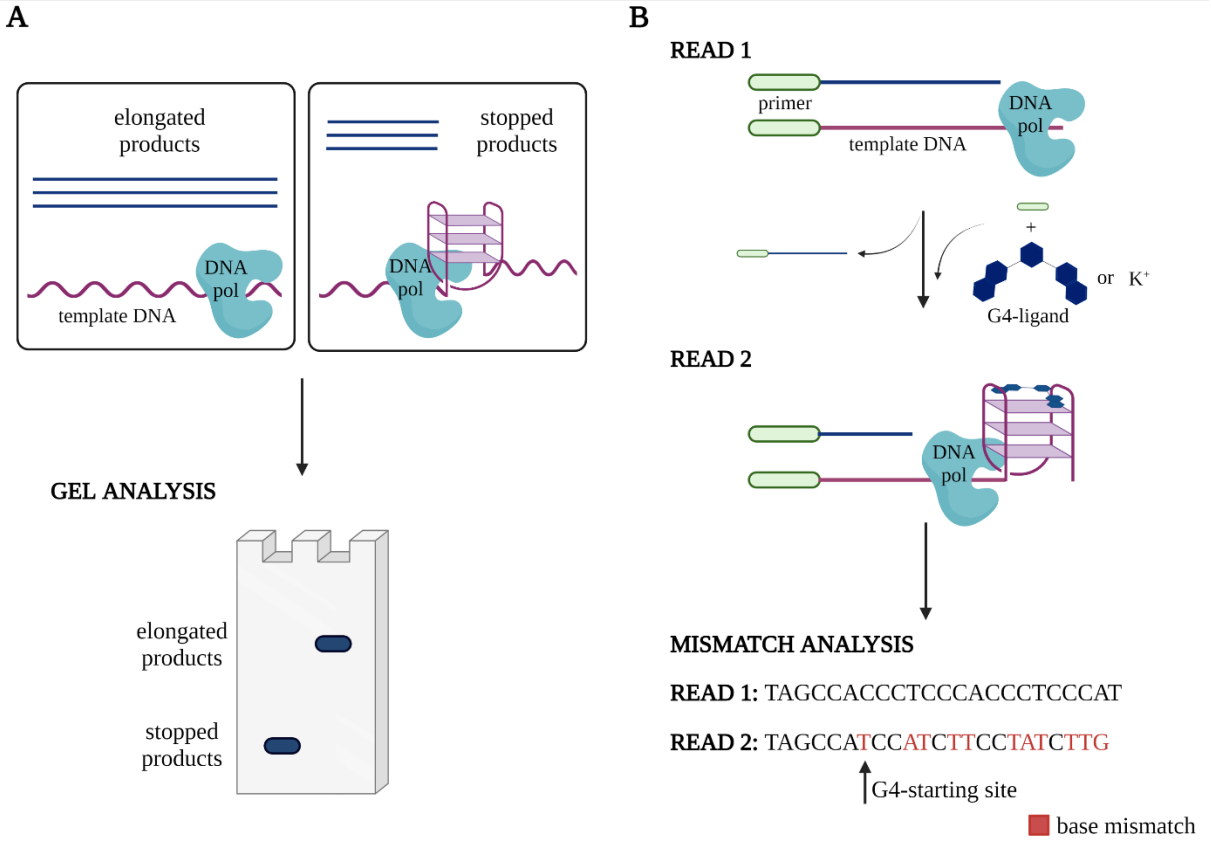

Figure 1.6. G4 structures block the polymerase processivity. (**A**) Representation of a polymerase stop assay. In absence of a G4 structure (left), the template DNA is fully elongated by the DNA polymerase, generating fully elongated products. Contrarily, the template DNA (right) forms a G4 structure (in purple), which impedes the DNA polymerase elongation, where only stopped products are obtained. All the products could be separated by gel electrophoresis, where the early stopped products generate bottom bands that run faster than the fully elongated ones do, which are observed at the top of the gel. (**B**) Schematic representation of the G4-Seq method. The first sequencing run (read 1) uses primers (in green) to allow the elongation of the DNA polymerase and provides a first reading of the template DNA. Then, the product is removed from the reaction and followed by the second sequencing run (read 2) after the addition of the primer alongside with either the G4-stabiliser ligand or K^+ stabiliser cation. These conditions promote the folding of the G4 within the template DNA with consequent stalling of the polymerase that starts to insert mismatches during the sequencing. Finally, during the mismatch analysis, the two reads are aligned and compared to identify the G4-forming region by looking at the base mismatches between reads 1 and 2. Figure B is a modification of Spiegel et al.[15] and Marsico et al. [60] figures.

The number of G4s detected by G4-Seq is over 700,000 and considerably exceeds the computational PQS predicted by Quadparser. This number includes G4s formed within functional regions such as 5′-untranslated regions (5'-UTR) and splicing sites, as well as G4s detected within many genes that have few or no PQS such as *brca1*, *brca2*, and *map3k8*, which are important cancer-related genes. High G4 density has been also found in oncogenes such as *myc*, *tert*, *akt1*, *fgfr3*, and *bcl2l1* underlying a link between G4s and cancer [55]. Interestingly, G4 formation has been detected between sequences that highly deviate from the canonical G4-forming sequence described in Quadparser, such as G4s with long loops or with discontinuous G-tetrads that form bulges, which cannot be detected by computational predictions [15,55,61]. Evidence for the existence of non-canonical G4s further underlines the complexity of these structures and the uncertainty about their specific biological role(s) *in cells* and *in vivo,* which is yet to be fully unravelled.

An improved version of G4-Seq, which includes the use of Li^+ instead of Na^+ for the first sequencing (read 1) run, allowed G. Marsico and colleagues to generate whole genome G4 maps for 12 species, including *human*, *mouse*, *E. coli*, *zebrafish*, and other model organisms and pathogens, with the aim to elucidate global G4-formation across species for the generation of improved predicting tools [60]. Interestingly, this analysis revealed a lack of G4s in bacterial and yeast genomes, suggesting that their formation might have been lost during the evolution. However, the same G4-Seq upon stabilisation with G4-ligands revealed the formation of G4s also in yeast and bacteria genomes, suggesting the potential to form G4s and their possible impact in cellular processes. Contrarily, higher species such as *human* and *mouse* present strong G4 enrichment at promoters and transcription start site (TSS) regions. A similar enrichment was also observed in the evolutionary distant *Trypanosoma*, suggesting similarities between *human*, *mouse*, and *Trypanosoma* species [60].

1.4 G-quadruplex Visualisation in Cells

Despite G4-Seq helped to detect G4s formation within purified DNAs and generate G4-maps for multiple species, over the years the lack of direct evidence proving G4-formation within cells always raised criticism about the ability of these structures to actually form under physiological cellular conditions. Therefore, it has been necessary to develop techniques to detect and visualise G4s in cells to fully understand the processes associated to these structures, their distribution, and dynamics in physiological contexts. To date, many tools rely on the widely accepted use of BG4 antibody that specifically recognises and binds G4 structures [62]. However, the main issue associated with BG4 is the need to fix the cells before applying the antibody, which could artificially perturb the G4 landscape. Therefore, there has been significant interest in generating new methods to visualise G4s in living cells. Recently, the generation of new probes for fluorescence lifetime imaging microscopy [63,64] and the development of small molecules for living-cell single-molecule fluorescence imaging of G4s under non-perturbative conditions [65], allowed to visualise G4s in real-time in living cells. The ability to visualise G4s in living cells not only confirms their dynamic formation but will also help to characterise the precise activity of individual G4s within the human genome [65].

1.4.1 BG4 Antibody

In 2013, the screening of an antibody phage display library [66] composed of 2.3×10^{10} single-chain antibody clones over a panel of intramolecular G4s, allowed G. Biffi and co-workers to isolate BG4, the first engineered structure-specific antibody that specifically binds G4 structures and can be used to visualise their formation in human cells [62]. The ability of BG4 to interact with intramolecular and intermolecular DNA G4s has been tested through enzyme-linked immunosorbent assay (ELISA) (see section 2.2). The calculated dissociated constants (K_d) were in the 0.5–2.0 nM range, indicating a high affinity of BG4 for all the G4s tested, including parallel (c-MYC, c-KIT1 and c-KIT2), anti-parallel (SPB1 and TBA), and hybrid-type (hTELO). Negligible binding was detected to RNA hairpin, ssDNA or dsDNA [62]. Subsequential binding and competitional analysis of BG4 confirmed its specificity for G4 DNAs and G4 RNAs [67], and its ability to bind both intramolecular and intermolecular G4s with a slight preference for parallel over antiparallel G4 conformations [68]. Given the

specificity of BG4 for binding G4 structures, this antibody has been heavily used for G4 visualisation in cells using fluorescence microscopy, and to map endogenous G4 structures in chromatin of fixed human cells through chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-Seq) experiments.

1.4.2 G-quadruplex Visualisation by Fluorescence Microscopy

As mentioned, the BG4 antibody has been heavily used for immunofluorescence staining of G4 structures within fixed human cells. BG4 has been designed with a Flag epitope tag, which could be recognised by a secondary fluorescently labelled antibody, generating a fluorescent signal that can be observed by confocal microscopy indicative of the presence of a G4 structure [62,67]. The immunofluorescence results showed punctate nuclear staining in presence of BG4, which is not observed in the absence of the primary antibody (Figure 1.7A), indicating that G4s mainly localise in the nucleus of cells [62]. To confirm the ability of BG4 to target G4s in human cells, cells were either treated with DNase or transfected with pre-folded G4s or single stranded oligonucleotides. The number of punctate nuclear staining strongly decreased after treatment with DNase, whilst an increased number of BG4 foci was observed in cells transfected with pre-folded G4s. Moreover, no changes were observed in cells transfected with ssDNAs, corroborating the cellular targeting of G4s by BG4 antibody. Immunofluorescence experiments showed that G4 formation is highly modulated during the cell cycle with the maximum number of BG4 foci when cells are replicating (during the cellular S phase). These results confirmed G4 formation mainly during DNA replication, when the DNA is single stranded and more easily folded into secondary structures [62].

BG4 antibody has also allowed to visualise RNA G4 structures within the cytoplasm of human cell lines [67]. The intense nuclear BG4 foci observed by fluorescence microscopy were obtained with short exposure times. Interestingly, longer exposure times identified BG4 staining throughout the cytoplasm of cells (Figure 1.7B). Such cytoplasmatic staining disappeared when cells were treated with RNase A before fixation, while negligible reduction was observed with Dnase treatment, confirming that G4s can be detected within the human transcriptome [67]. Different fixation protocols (such as the use of formaldehyde or ethanol) have been tested to prove that fixation had no impact on the results and cytoplasmatic foci

observed. However, the possible influence of the fixative on the DNA and RNA G4-landscape is something that cannot be excluded [67].

Although BG4 allowed to obtain direct evidence of G4 formation within cells, real-time detection of G4 formation in living cells is not possible using antibodies. As mentioned in Section 1.4, alternative approaches could be used to overcome these limitations. One valid approach was developed by M. K. Kuimova and R. Vilar. This approach relies on chemical probes that can change their fluorescence lifetime upon binding to different nucleic acid structures, which can be measured by fluorescence lifetime imaging microscopy (FLIM) [63,64]. Specifically, DAOTA-M2 was developed as a low cytotoxic molecule with good live cell permeability and a significant different lifetime when bound to G4s (9–12 ns) compared to that of dsDNA (5–7 ns) or RNA (7–11 ns) (Figure 1.7C) [63,64]. Longer lifetimes [11.1 ± 0.7 ns] were observed after depletion of DNA helicases such as FancJ, which resolves G4 structures, compared to those of control cells [10.5 ± 0.7 ns], confirming the ability of DAOTA-M2 to monitor G4-distribution in live cells [63]. One of the main limitations of this technique is the high (μM) concentrations of probe required for the analysis. The high concentration of probe could indeed influence the dynamics of the G4-landscape or promote their formation [65]. More recently, M. Di Antonio and co-workers developed a non-toxic G4-fluorogenic probe, called SiR-PyPDS, which can be used for single-molecule fluorescence imaging of G4s in live cells (Figure 1.7D). As further discussed in Section 1.6.1, low concentrations (nM) of the molecule are sufficient for detecting G4s, preventing any possible effects on the global G4-landscape. With this approach, approximately 3,000 G4s have been identified within a single cell. The highest number of SiR-PyPDS binding events was observed during the S phase, confirming the previous findings made in fixed cells, hence demonstrating the powerful nature of this method to understand the biological role of G4s in real time in living cells [65].

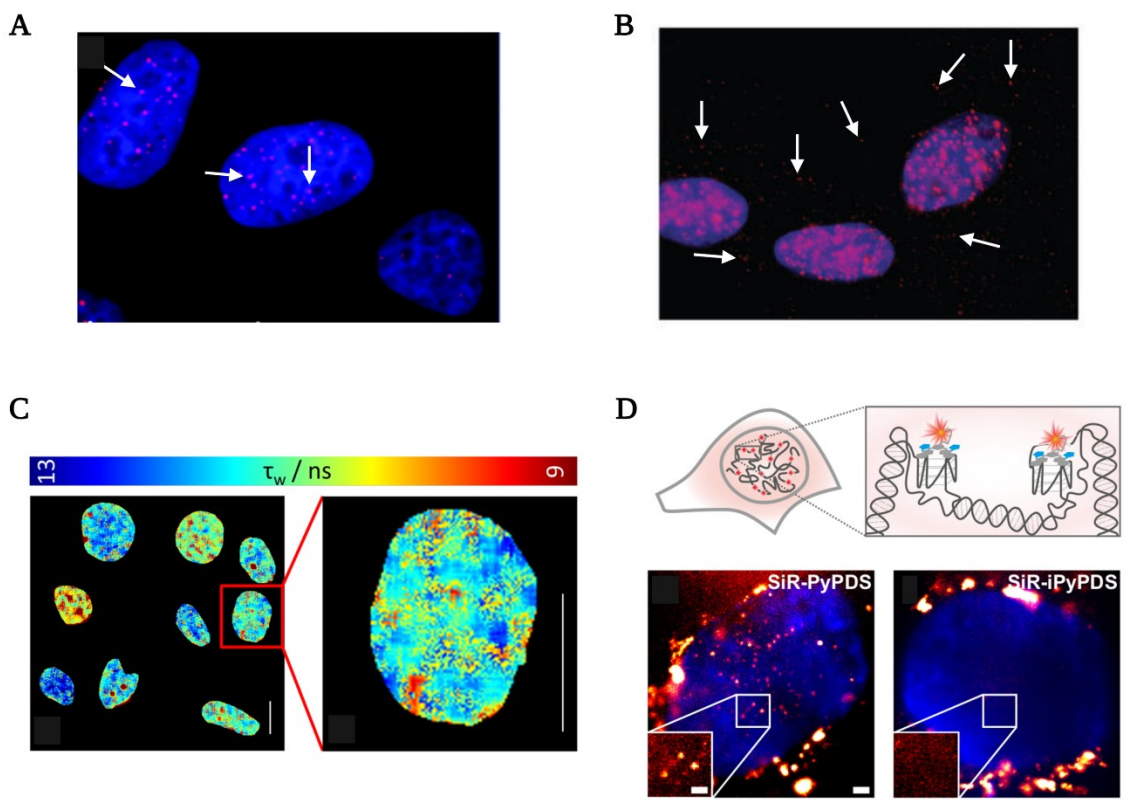

Figure 1.7. G4 visualisation in human cells. (**A**) Fluorescent microscopy image adapted from Biffi et al. [62] showing DNA G4 localisation as red BG4 spots mainly localised in the nuclei of fixed cells (some BG4 spots are indicated with white arrows). (**B**) Fluorescent microscopy image taken after a long exposure showing RNA G4 staining in the cytoplasm of fixed cells using BG4 antibody. Some BG4 cytoplasmic foci are indicated with white arrows. This picture has been modified from Biffi et al. [67]. (**C**) FLIM map of human U2OS living cells after staining with DAOTA-M2 (20µM, 24h), adapted from Summer et al. [63]. The FLIM map of a single nucleus is zoomed-in on the right-side. The colour gradient bar between 9 nm (red) and 13 ns (blue) represents the different lifetimes. (**D**) Single molecule fluorescence imaging of G4s in living cells after staining with the fluorescent probe SiR-PyPDS. The image has been adapted from Di Antonio et al. [65]. Schematic representation of a cell (top) with a zoomed-in nucleus where only G4s are bound and stained by the fluorescent probe SiR-PyPDS. G4 staining in a living U2OS cell (bottom) treated with 20 nM of SiR-PyPDS for 30 min (red spots in the nucleus on the left-side), and the absence of G4 foci in living U2OS cells treated with a control probe SiR-iPyPDS uncapable of binding (right).

1.4.3 Chromatin Immunoprecipitation Followed by High-throughput DNA Sequencing (ChIP-Seq)

Evidence that G4s can form in human cells raised many questions about how these structures can affect chromatin architecture. Therefore, the generation of genome-wide maps of G4s in cells became increasingly important [62].

Targeting proteins that are known G4 binders allowed chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-Seq) to investigate genomic locations of G4 structures. In 2011, K. Paeschke and co-workers demonstrated that the *Saccharomyces cerevisiae* Pif1 DNA helicase binds G4 motifs, whereas lack of Pif1 causes stalling of the replication fork with frequent DNA breakages near the G4 regions that are resolved by the helicase. These results suggested that G4s can form in living cells and, as further discussed in Section 1.5.2, these structures could block replication if they are not resolved by Pif1 [69]. ChIP-Seq studies by L. T. Gray and colleagues, showed that in human cells, 40% of the binding sites of the transcription-associated helicases XPB and XPD overlap with G4 motifs, and the regulation of these structures is associated to cancer signalling and regulatory pathways [70].

Previous ChIP-Seq on known G4-binding proteins helped to elucidate the role of G4s in the chromatin context. However, a direct characterisation of the relationship between G4s and chromatin has been possible only after the development of a G4 ChIP-Seq protocol with the synthetic BG4 antibody (Figure 1.8A) [62,71,72]. Interestingly, this technique revealed approximately 10,000 G4 peaks, which is a remarkably lower number compared to the number of G4s computationally predicted [35] or detected by G4-Seq [15,55,71]. These results strongly reflect the influence of chromatin on the G4 landscape, especially considering that most G4s were detected at promoters of highly transcribed genes and in nucleosome-depleted regions (Figure 1.8B).

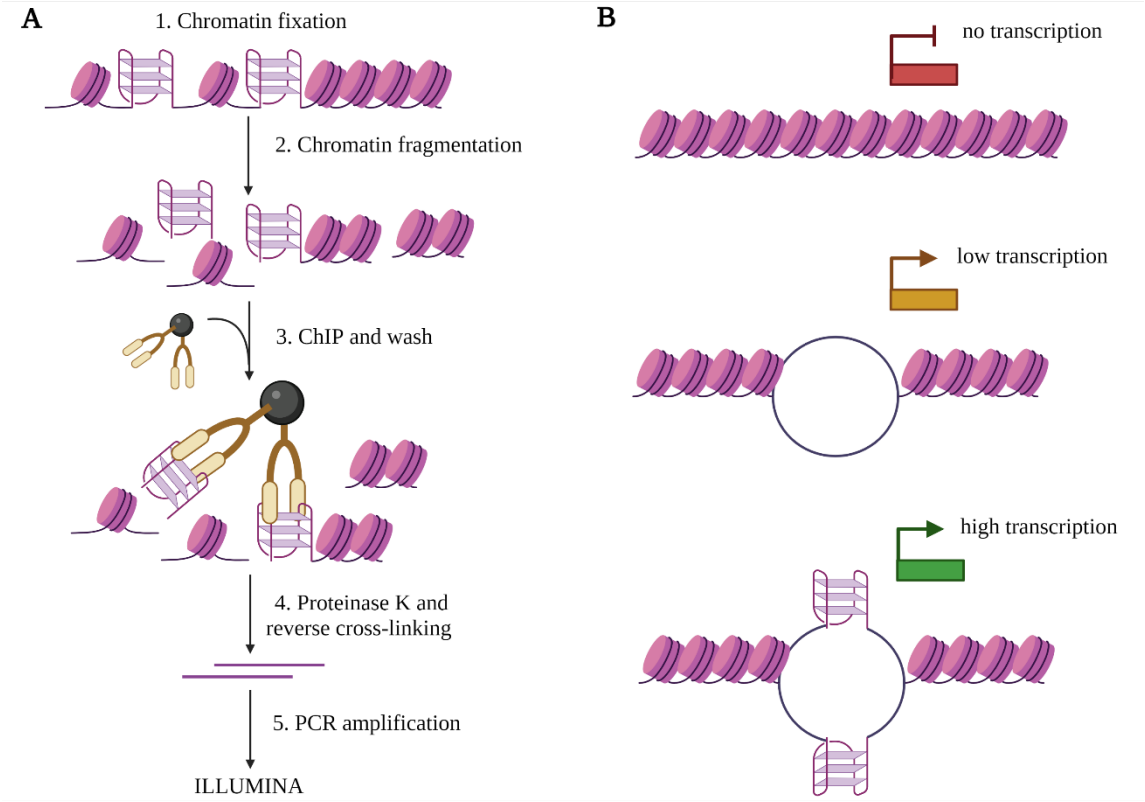

Figure 1.8. ChIP-Seq using BG4 antibody identifies G4s in high transcribed genes. (**A**) Schematic representation of ChIP-Seq protocol described in Hänsel-Hertsch et al. [72]. The first step requires chromatin fixation to stabilise DNA-nucleosome interactions. Nucleosomes are represented as purple rounded molecules. Then, the chromatin is sonicated to generate short fragments of 100-500 bp. The sample is then treated with RNase A and blocked with BSA prior to adding BG4 antibody (represented in brown), which has been previously attached to magnetic beads (represented with a black sphere) to allow immunoprecipitation. Beads are washed to remove unspecific interaction and the immunoprecipitated sample is treated with protease K and reverse cross-linked. Finally, the library for Illumina sequence is prepared by fragmentation, addiction of adaptors and PCR amplification. (**B**) Schematic representation of the effects that the presence of G4s could have during gene transcription. Heterochromatic gene promoter in a "closed" status does not allow transcription (transcription off; top). When the gene promoter is opened, the transcription is generally low (low transcription; middle). However, when the nucleosome-depleted promoter contains G4 structures the transcription is activated (transcription on; bottom). This figure has been modified from the figure present in Hänsel-Hertsch et al. reference [71].

Nucleosome-depleted regions are associated with high transcriptional levels, raising the question on whether G4s arise as a consequence of active transcription, or their formation is independent from gene expression levels [71,73]. Very recently, J. Shen and co-workers found

that G4-formation at gene promoters and their stabilisation with G4-ligands leads to the retention of RNA Polymerase II (RNAPII). This suggested that G4s could act as a site for the recruitment of key components of the transcriptional machinery by either direct binding between RNAPII and G4s or through recruitment of transcription factors (TFs) [73]. Conversely, G4 formation is unfavoured by compacted chromatin, and lack of G4s promoters is accompanied by loss of RNAPII at the same site [73]. In support, ChIP-Seq experiments and biochemical assays showed that a zinc-finger transcription factor Sp1 binds the G4 formed with high affinity within the *c-kit* promoter [74]. Similarly, a Myc-associated zinc-finger protein (MAZ) promotes transcriptional activation through its binding on the G4-forming sequences present in the *kras* promoter, while disruption of the G4 conformation causes down-regulation of *kras* oncogene expression [75]. Altogether these results are consistent with previous observations obtained in cells deficient for G4-resolving helicases, such as Werner and Bloom syndrome helicase (WRN and BLM, respectively), which displayed transcriptional upregulation of predicted G4-forming genes [71,76].

Interestingly, only 26% of G4s predicted within nucleosome-depleted regions overlapped with G4 ChIP-Seq, suggesting that stable G4 formation is influenced not only by the suppressive role of nucleosomes that inhibits G4s formation, but also additional features are likely to contribute. For instance, helicases, which resolve these structures, make the G4-landscape very dynamic and possibly tightly regulated [71], as described in Section 1.7. Such strict regulation is needed in light of G4s' enrichment in cancer-related genes and their association to DNA damage, genome instability, and cancer progression.

1.5 Biological Role of G-quadruplex DNA

G4s might have different roles in cells, given the different cellular localisation and contexts where they have been observed. G4s have often been considered as knots in ssDNA sequences acting like physical impediments to polymerases during replication, transcription, and translation [77]. However, as discussed in Section 1.4.3, ChIP-Seq data revealed that G4s are enriched at promoters and TSS of highly transcribed genes and in nucleosome-depleted regions [71]. Moreover, co-localisation of these structures with histone modifications that are associated to active genes such as trimethylated histone H3 Lysine 4 (H3K4me3), and lack of G4s in the presence of trimethylation of histone H3 Lysine 9 (H3K9me3), which is a modification associated to heterochromatin, suggest that these structures may be involved in signalling active gene expression [71,78]. G4-ChIP-Seq in the human genome showed that G4s can form in unmethylated DNA CpG Islands and co-localise with (cytosine-5)-DNA methyltransferase 1 (DNMT1) enzyme, which is responsible for the maintenance of the CpG repressive-methylation marks in human cells [79,80]. SQ. Mao and colleagues, demonstrated that the methylation activity of DNMT1, which establishes a repressive state of the chromatin, is inhibited by G4 formation [80]. G4s might sequester DNMT1 due to the high binding affinity of the enzyme to these structures (Figure 1.9) [80], suggesting new roles of G4s in the regulation of epigenetic modifiers such as DNMT1. G4s might be involved in many other different mechanisms for epigenetic and chromatin regulations, which are dependent on their wide and different genome distribution and cellular activities [81]. An overview of the different roles that G4s could have within the human genome is further presented in this section.

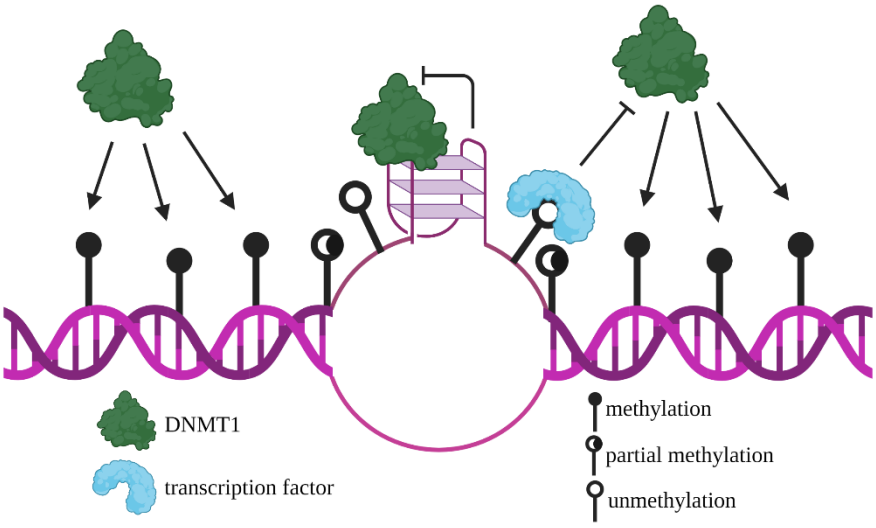

Figure 1.9. A possible model for the role of G4s in methylation control. This model has been proposed by Mao and co-workers [80]. Due to its high binding affinity for G4 structures, DNMT1 is sequestered by G4, and this DNMT1-G4 interaction inhibits the methylation activity of the enzyme in a similar way as the methylation inhibition is promoted by transcription factors. Contrarily, when DNMT1 is not bound to a G4, its methylation activity is active and promotes the heterochromatic state of the genome. This image has been modified from Mao et al.[80].

1.5.1 G-quadruplex in Telomeres

The end of eukaryotic chromosomes is characterised by single strand overhangs, called telomeres. The length of human telomeres is around 5-25 kilobases (kb) and consists of tandem repeats of the hexanucleotide $(TTAGGG)_n$. The 35–600 bases that constitute the 3'-overhang of telomeres are single stranded and prone to form G4 structures [51,82,83]. NMR analysis of the 22-nucleotides (nt) human telomeric DNA 5′-$AGGG(TTAGGG)_3$ sequence revealed the formation of an intramolecular antiparallel G4 in Na^+ solution [84]. However, the crystal structure of the same 22-nt sequence in K^+ solution showed a three-tetrad parallel G4 [29]. Interestingly, further structural characterisations identified that human telomeric sequences typically form hybrid-type intramolecular structures in K^+ solution, which are different from the Na^+ NMR structure or K^+ crystal structure [85]. These structural differences suggested that G4s are very polymorphic at telomeres and such plasticity might reflect sophisticated regulations mediated by protein interactions [86]. Fluorescence microscopy using BG4 antibody in human cells further demonstrated G4 formation in telomeres (Figure 1.10A) [62].

In somatic cells, about 50 to 200 bases are lost at telomeres during each replication cycle, causing shortening of telomeres with ageing and consequent apoptosis of the cells [87]. One of the characteristics of the cancerogenic transformation is the re-activation of the enzyme telomerase. This reverse transcriptase is normally active only in stem cells and inactivated in somatic cells. After transformation, telomerase is re-activated in 80%–85% of cancer cells and elongates the telomeric sequences, avoiding telomeric shortening upon replication [85,88,89]. Pivotal studies of *Oxytricha nova* demonstrated that G4 stabilisation with K^+ or Na^+ inhibits the primer elongation activity of telomerase [90], suggesting that G4 structures formed at telomeric repeats might be targeted by small molecules as new approaches for cancer therapeutics [85]. Telomerase is composed of two main subunits, the catalytic protein, and a telomerase RNA. The promoter region of the catalytic subunit of the telomerase, also called human telomerase reverse transcriptase (hTERT), is enriched of G4 motifs that can form G4 structures [91]. hTERT promoter also contains binding sites for the transcription factor Sp1, which is essential for hTERT expression in several cancer cell lines. Mutational experiments demonstrated that G4 formation prevents all the promoter regions to be bound by Sp1 with consequent inhibition of the telomerase activity [91]. Therefore, stabilisation of telomeric G4s in the hTERT promoter could serve as an additional mechanism to block the cancerogenic cell growth. Some examples of G4-ligands that induce senescence and apoptosis in human cancer cells by inhibiting the telomerase activity are presented in Section 1.6.

1.5.2 G-quadruplex in Replication and Genome Instability

As discussed in Section 1.1, alternative DNA secondary structures, including G4s, are prone to form when the dsDNA is transiently opened during replication and transcription.

Evidence has suggested that G4s can form during replication and impede DNA polymerases progression *in vitro* [92,93]. As a consequence of the replication fork stalling, cells undergo increased level of chromosomal rearrangements and recombination, which could be potentially deleterious [94]. The first evidence of polymerase impediment by G4s *in vivo* came from E. Kruisselbrink and co-workers in their loss of function study of DOG1, a *Caenorhabditis elegans* homolog of the human FANCJ helicase known to resolve G4 structures [95,96]. A *LacZ* reporter system containing a C-rich tract in the non-template strand

(G-stretch in the template strand) followed by multiple stop codons upstream *LacZ* start codon, allowed to identify G4-induced genomic deletions through expression of β-galactosidase. Indeed, only when the deletions remove the C-rich tract and stop codons, the *LacZ* start codon results in-frame with the downstream ORF and is actively expressed. PCR-based analysis of individual worms determined increased accumulation of small deletions (50–300 nt) at the G4-forming regions only in *dog1* deficient animals, but not in the WT. This suggested that DOG1 was necessary to resolve G4s formed during replication. Therefore, in absence of the helicase, the polymerase is unable to bypass the G4 structure causing the formation of a ssDNA gap on one parental strand. The other strand maintains the G4 structure through multiple cell cycles and this generates deletions during each replication round and double strand breaks (DSBs) (Figure 1.10B), which are normally resolved by two error-prone DBS repair mechanisms: either a DNA Polymerase theta (Pol θ) mediated end-joining (TMEJ) [97] or a non-homologous DNA end-joining (NHEJ) [98] with the risk of inserting mutations or chromosomal aberrations [95,98-100]. A very well-known example of a G4-induced chromosomal translocation occurs between chromosomes 14 and 18, named [t(14;18)], and it is associated with the development of follicular lymphoma. In particular, the *bcl2* locus on chromosome 18 and enhancer element of the IgH locus on chromosome 14 are fused together, resulting in overexpression of the antiapoptotic protein, BCL2 [101]. The majority of DNA breaks within the *bcl2* locus occur within a very G-rich region that forms G4s, and cause polymerase stalling *in vitro* [102].

Recently, the progression rates of single replication forks through different G4-containing sequences have been measured by live-cell imaging in *Saccharomyces cerevisiae* [103]. The replication rate was calculated using a lacI-GFP and tetR-tdTomato replication assay containing a G4-forming sequence between the two fluorescent reporters. The ratio between GFP and tdTomato signals showed a significant reduction of the replication rate in absence of the helicase Pif1. Contradictorily, the data revealed that the genetic instability caused by the replication block in absence of Pif1 occurs either when the G4s arise on the discontinuous lagging strand [103] or on the leading strand [104], suggesting that G4s could form on both the DNA strands during replication and these two possibilities are not mutually exclusive [105].

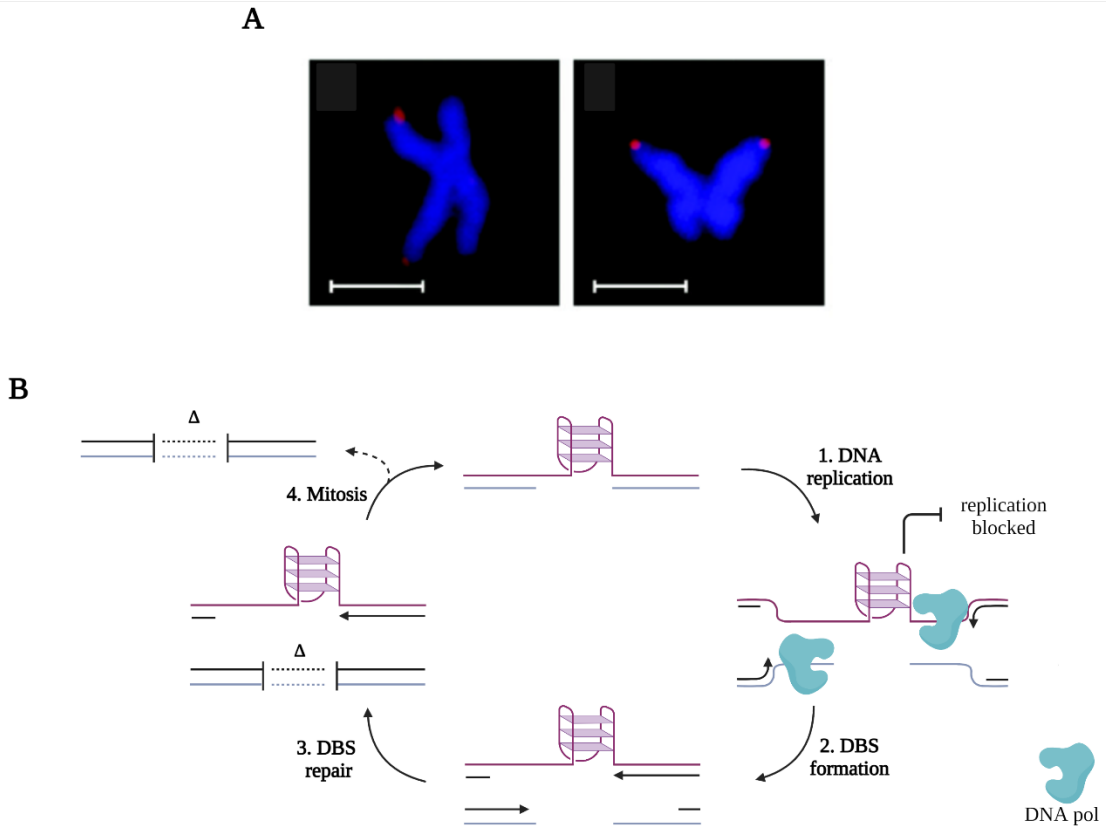

Figure 1.10. G4s are involved in different biological processes. (**A**) Fluorescence microscopy picture showing G4 formation in telomeres (red signals) of fixed human cells (from Biffi et al. [62]). (**B**) Model for mitotic inheritance of G4s, which causes accumulation of deletions among proliferating cells reported by B. Lemmens et al [100]. The presence of a G4 structure blocks the processivity of the DNA polymerase, causing a local potent block of the synthesis of the nascent strand during replication (1). Failed DNA replication of the G4-region causes a ssDNA gap in the nascent strand. However, if the28pposete parental strand also presents a gap due to replication of the stalled template, the presence of two gaps at the same region generates DSBs (2) which are repaired by TMEJ (3) generating small deletions (Δ). Therefore, with the next mitotic cycle, one cell will inherit the deletion, while the other will receive stable G4 and the ssDNA gap. Figure modified from B. Lemmens et al. reference [100].

1.5.3 G4-quadruplex and Transcription and Their Role in Epigenetic Regulation

The negative effects of G4 formation on the replicative machinery have been well documented by polymerase stop assays *in vitro* and reporter assays in living cells or *in vivo*. Conversely, the effects of G4s on the transcriptional machinery suggest either a disruptive or

a constructive influence meaning that G4s are associated to either increased or decreased transcriptional activity [106]. Although initial studies identified G4s as inhibitors of gene expression, the lack of direct correlations between G4s in promoters and transcriptional repression cannot certainly prove their role as antagonists of transcription [81]. Moreover, the demonstration that G4s enriched at gene promoters of transcriptionally active genes acted as transcriptional enhancers rather than repressors [81], highlighted the necessity of further investigations to unravel the role of G4s in transcriptional regulation.

RNA-Seq data revealed that genes containing G4s in promoters, TSS and 5'-UTR regions are more expressed compared to genes containing G4s at more than 1 kilobase (kb) from the TSS [106], suggesting that the transcriptional effects of G4s might be based on genomic location (Figure 1.11A). *In vitro* studies using T7 RNA polymerase (T7-RNAP) on G-rich repeats (5'-GGA)$_4$ templates located within the human *c-myb* proto-oncogene suggested transcriptional arrest under physiological conditions. Transcriptional blockage was not observed in the presence of LiCl or after mutation of the G-tract that prevents G4 formation, indicating that similar to replication, intragenic G4s can act as transcriptional repressors [107]. Conversely, ongoing studies have shown the role of G4s as transcriptional activator elements when formed within promoters of genes, upon interaction with some nucleic acid-binding proteins, for instance, DNMT1 [80]. Other different proteins such as the cellular nucleic acid-binding protein (CNBP) and non-metastatic cell 2 (NM23-H2) protein, present strong binding affinity for G4 structures. *In vitro* and *in cells* data demonstrated that these two proteins are responsible for an increased *c-myc* transcription after interaction with the G4s formed within the *c-myc* promoter [108-110].

Depending on which strand contains the G-rich tract, during transcription G4 structures could form either on the template strand (intramolecular G4 DNA) or within the non-template DNA strand and nascent mRNA (intermolecular G4 DNA:RNA), physically interfering with the polymerase processivity (Figure 1.11B) [78,102]. Using a reconstituted T7 transcription model K-W. Zheng and co-workers reported the first intermolecular DNA:RNA hybrid G4 in a plasmid system *in vitro* [111]. To better understand the implication of DNA:RNA G4s in cells, human HEK293 cells were transfected with a luciferase reporter vector containing a G-rich NRAS sequence downstream of a SV40 promoter. The luciferase expression was compared to a control plasmid in the absence of the G-tract. Decreased luciferase expression was observed only in absence of a recombinant RNase H [112]. Expression of the luciferase was further reduced in the presence of a G4-ligand, confirming that DNA:RNA G4s are able to stall the

T7-RNAP processivity [111]. In contrast, recent *in vitro* studies demonstrated that R-loop/G4 structures significantly enhance RNAP initiation and elongation only when located downstream of the TSS in the non-template strand (Figure 1.11C). R-loops are secondary structures formed when the nascent RNA invades the dsDNA and interacts with the template strand forming a DNA:RNA hybrid [113]. R-loop formation seems to be favoured by the formation of G4s within the non-template strand. Experimental evidence showed increased mRNA level upon formation of these R-loop/G4 structures, opposite to what was observed when the G4 was located within the DNA template (Figure 1.11C) [114] . Overall, these results confirmed that the different positioning of G4 structures can result in different transcriptional outputs.

The findings discussed are examples of how G4s formation might regulate the gene expression with distinct mechanisms. Since G4 formation in promoters and regulatory regions is reversible and dependent on chromatin compaction, increasing studies have suggested the potential role of G4s as epigenetic regulators (Figure 1.11D). For example, it can be speculated that G4s might represent binding sites for histone-modifying proteins [71] or act as repressors of the epigenetic silencer DNMT1 to sustain the gene expression [80]. Increasing evidence suggests that G4s may recruit chromatin remodelling proteins and histone chaperones such as SWI/SNF and locally promote nucleosome disassembly or reassembly [115,116].

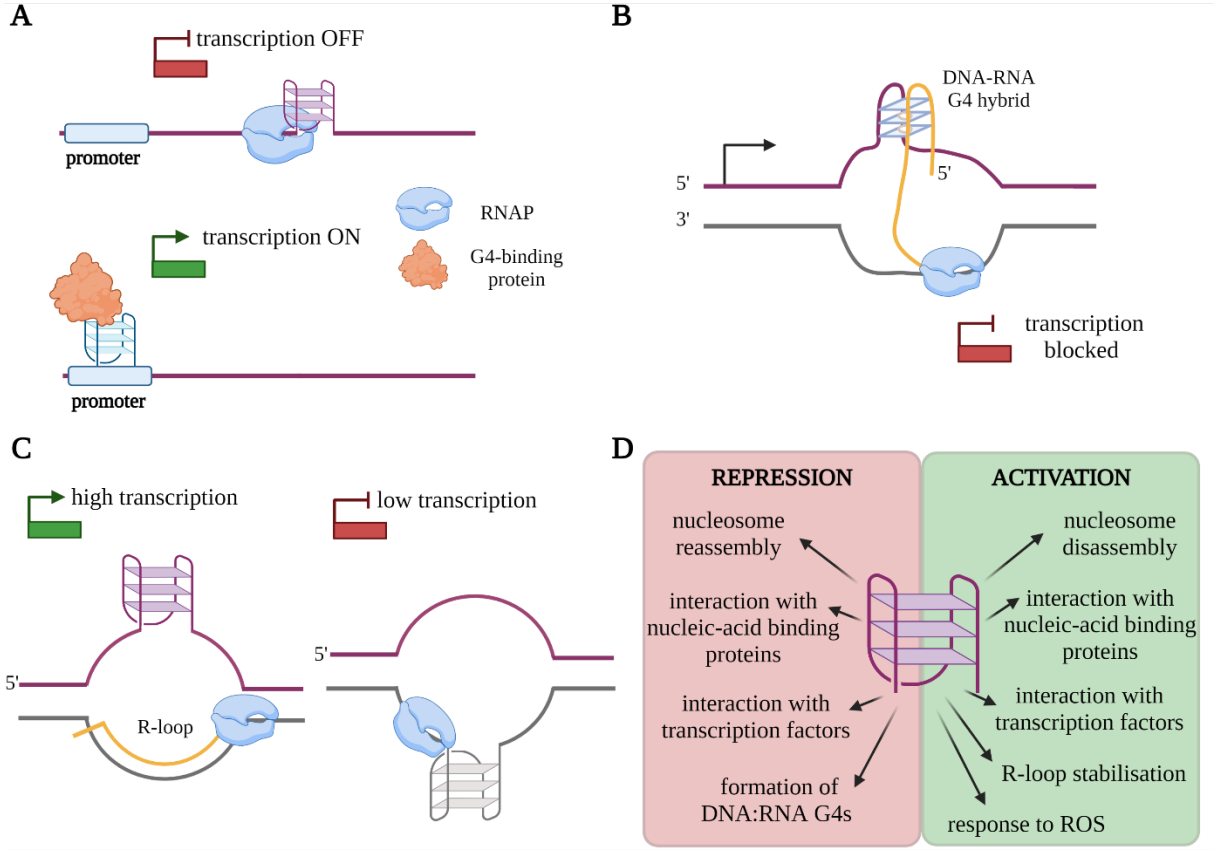

Figure 1.11. G4s are associated to either increased or decreased transcriptional activity. (**A**) Schematic representation about the different impacts of G4s in transcriptional regulation depending on their distance from the gene promoter or TSS. When the G4 is found more than 1 kb from the promoter, it blocks the RNAP processivity, causing a reduction of transcription (upper part). If the G4 is localised within the promoter region of the corresponding gene, it is bound by G4-binding proteins with consequent transcriptional activation (bottom part). (**B**) DNA-RNA G4 hybrids block the transcription process. Figure adapted from D. Varshney et al. [78]. (**C**) G4 structures within the non-template DNA (in purple) promote the R-loops stabilisation, which enhances transcription (right-side). G4 structures within the template DNA (in grey) inhibit the RNAP progression resulting in low transcription. Figure adapted from C-Y. Lee et al. [114]. (**D**) Summary of some transcriptional repressor and activator effects of G4s within gene promoters and other regulatory regions.

An additional example of the epigenetic regulation given by G4s is within the response to environmental stress such as guanine oxidation induced by reactive oxygen species (ROS) [81]. G4 structures are highly sensitive to guanine oxidation, which could potentially lead to deleterious events for the cells, such as cancer and ageing if not correctly repaired [117,118]. However, recent studies demonstrated that 8-oxo-7,8-dihydroguanine (OG) lesions generated in G-rich gene promoters upon ROS exposure could promote gene activation through base

excision repair (BER) [119,120]. Luciferase reporter assays in the presence of G-rich *VEGF* or *NTHL1* promoters containing OG bases within their five G-tracts showed formation of an apurinic site (AP), which is extruded into a loop. This looped G4 structure is recognised and bound by an apurinic/apyrimidinic endonuclease (APE1), which leads to gene activation by recruiting transcription factors.

Altogether these findings place G4 structures as an interconnected network between biomolecules, rather than suggesting G4s as isolated entities limited to act as transcriptional repressors as previously suggested.

1.6 G-quadruplex Ligands and Their Role as Therapeutic Treatment for Cancer

The potential for G4s to be formed at telomeres, promoters and other highly regulated regions of the genome strongly suggest their implication in cellular processes and consequently, in human diseases such as cancer and ageing [15,35,77,121]. In particular, G4s are more enriched in promoters of proto-oncogenes compared to promoters of tumour suppressor or regulatory genes, suggesting that they might be used as targets for anticancer therapies [122]. Among the numerous promoters where G4s have been detected, the most studied G4-forming proto-oncogenes include *VEGF* [123], *bcl2* [124], *kras* [125], *kit* [126] and *myc* [127]. MYC is a transcription factor involved in the alteration of cell proliferation, metabolism, and immune evasion during tumour transformation, which is upregulated in around 70% of cancers [128,129]. Polymerase stop assays and gene-expression data using luciferase reporters decorated with an upstream G-rich *c-myc* promoter, demonstrated that treatment with G4-ligands causes high polymerase arrest and relevant reduction of luciferase expression [127]. Similarly, treatment with G4-ligands caused reduction of the mRNA level of the oncogene *kras* [125], suggesting that stabilisation of G4s formed at *c-myc* and *kras* promoters impede the transcriptional machinery to proceed, resulting in a significant down-regulation of these oncogenes.

Therefore, there has been a growing interest in generating G4-ligands that can selectively interact with *myc* and other G4s that could potentially act as novel therapeutic and diagnostic agents to reduce tumour growth [122,130]. Gomez and co-workers demonstrated that a potent G4-ligand telomestatin (Figure 1.12A) induces senescence and apoptosis in cancer cells [131,132]. Treatment of HT1080 tumour cell line with telomestatin caused strong stabilisation of telomeric G4s with consequent degradation of telomeres and delocalisation of POT1 and TRF2 telomeric proteins from telomeres. This activated a damage response that triggers the cancer cells to senescence and apoptosis [131]. Similar to telomestatin, treatment of human glioblastoma cells with the G4 stabiliser BRACO-19 (Figure 1.12B) [133] caused displacement of TRF2 and POT1 and inhibited telomerase activity, leading to senescence and apoptosis of the cells [134]. Pyridostatin (PDS) (Figure 1.12C) is a very well-known small molecule able to target not only telomeric G4s but also G4s disperse within the cellular genome, inducing DNA damage responses (DDRs), which decreases the proliferation of cancer cells [135,136]. Similarly, CX-5461 (Figure 1.12D) [137] is a small molecule, which interacts with G4s with high affinity, and the stabilisation of these structures activates DDRs [138]. Other two well-

known G4-ligands are 5,10,15,20-tetra-(N-methyl-4-pyridyl)porphine (TMPyP4) and porphyrin 5,10,15,20-tetra-(N-methyl-2-pyridyl)porphine (TMPyP2) (Figure 1.12E and F). TMPyP4 binds intramolecular telomeric 5'-GGGTTA G4s through π–π stacking with the terminal G-tetrad, while TMPyP2 binds the TTA loop of the same telomeric sequence. These different interaction modes on the same telomeric G4 allow to establish external stacking to the G-tetrad (Figure 1.13A), which is more efficient in producing telomerase inhibition compared to binding on the loop from outside [139].

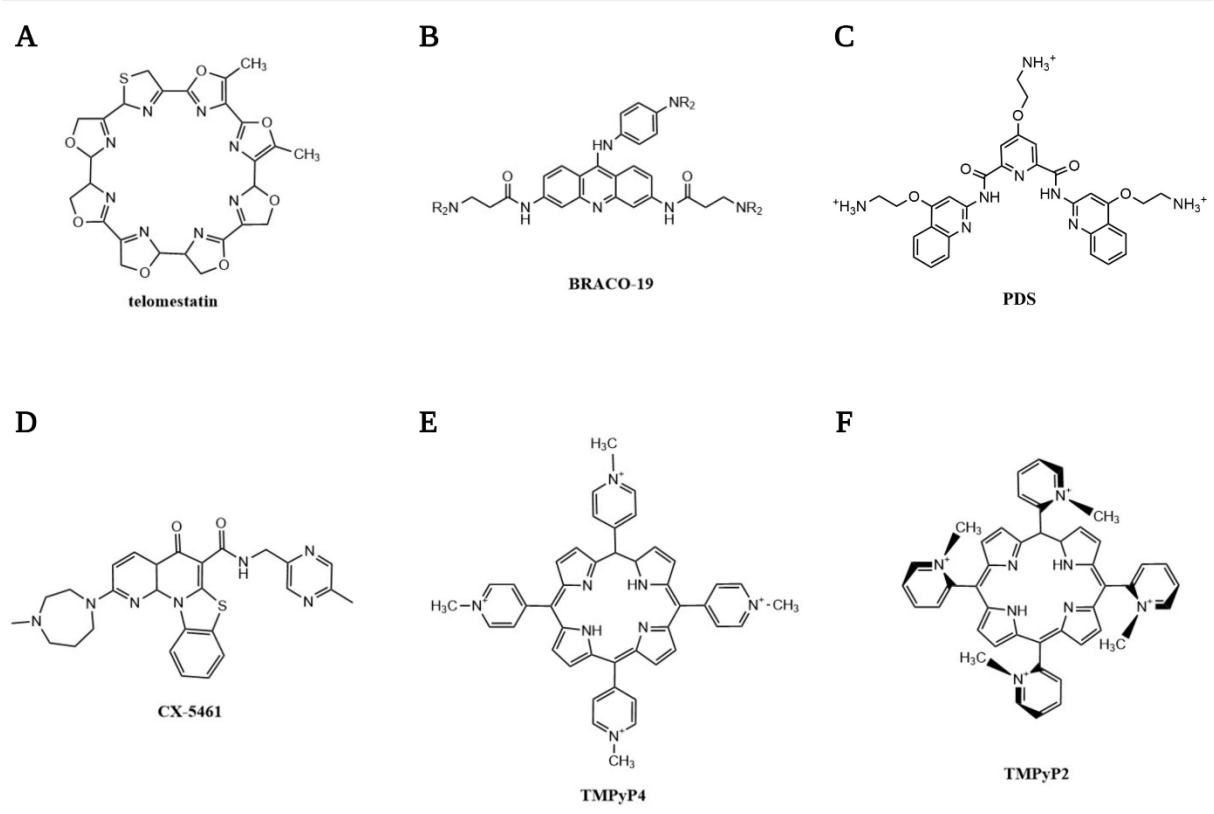

Figure 1.12. Molecular structures of different G4-ligands. (**A**) Telomestatin [132], (**B**) BRACO-19 [133] (**C**) PDS [136], (**D**) CX-5461[137], (**E**) TMPyP4 [139], and (**F**) TMPyP2 [139].

From these studies, other investigations have been performed to assess the best properties that make a synthetic small molecule a good G4-interactor that could be safely delivered in human cells. Most of the G4-ligands contain a large flat planar aromatic surface that can lead to π–π interactions with the terminal G-tetrad, while others present dimeric compounds that bind to two opposite grooves of the G4 [140,141]. The combined knowledge on about 1,000 reported G4

ligands has been collected by L. Qian and colleagues on a database, called G4LDB (http://www.g4ldb.org) with the aim to improve the efficiency of the G4-ligand research [140]. However, one of the main problems with G4-ligands is that they are not specific for a particular G4 and interact with the whole population of G4s within a cell. Therefore, it is not clear which G4s they actually bind to or which pathways are involved in the effect caused by G4-stabilisation with G4-ligands [128]. Another issue related to G4-ligands is their selectivity for G4s over dsDNA in the cellular context. In fact, even if these compounds are more selective for G4 *in vitro*, there is so much dsDNA in cells that the probability for these ligands to interact with dsDNA is highly probable. For instance, TmPyP4 is known to have very poor G4-selectivity [142]. Hence, the biological studies with this ligand should be treated with caution.

1.6.1 Pyridostatin (PDS)

PDS (Figure 1.12C) is a well characterised planar polyaromatic G4-ligand that binds and stabilises G4 structures through π–π interactions and electrostatic interactions with the terminal G-tetrad (Figure 1.13A) [135]. Fluorescence resonance energy transfer (FRET)-melting experiments showed that PDS stabilises human telomeric G4s with a maximum ΔTm of 35 K in 60 mM K^+ at <0.5 μM compound with negligible stabilisation of dsDNA (ΔTm of 0.5 K in 60 mM K^+ at 1 μM compound) [135]. Electrophoretic mobility shift assays (EMSA) and cellular data using HT1080 tumour cells demonstrated that, similar to telomestatin [131], 1 μM PDS is able to displace POT1 from the telomeric G4-repeats.

As mentioned in Section 1.6, treatment of cells with PDS revealed the ability of this G4-ligand to elicit a DDR. Using ChIP-Seq, Rodriguez and colleagues observed that PDS induced approximately 60 domains of phosphorylation of histone H2AX on Ser-139 (called γH2AX), which is a marker of DDR activation [135,136,143]. The γH2AX domains were enriched in non-telomeric regions of chromosomes that contained high numbers of predicted PQS. In particular, they observed that PDS could repress the expression of the oncogene *src*, which reduced the SRC-dependant cellular motility in human breast cancer cells [136]. This work proposed the druggability of certain cancer genes with small molecules, which decreases cancer progression by activating DDRs upon G4 stabilisation. Similarly, an increased number of γH2AX *foci* was observed in neuronal cells after treatment with PDS compared to the number in untreated controls, causing neurite retraction, decreased neuronal survival, synaptic

loss and formation of DSBs [144]. Further analysis revealed that PDS downregulates *brca1* expression in neurons by stabilising the G4 structures formed within this gene. BRCA1 is a protein involved in neuronal genome repair mechanisms, at the transcriptional level and it is generally lost in the brains of patients with Alzheimer's disease, which suffer from neural senescence [144]. Therefore, transcriptional gene downregulation through stabilisation of G4 structures with G4-ligands not only has the potential to revert the tumorigenic process but also pointed out a correlation between G4s and brain ageing, which may be important to ageing biology in general.

Surprisingly, PDS did not cause any effect on *hras* expression, which is the oncogene with the highest number of PQS [71,72], clearly indicating that not all genes with high PQS levels are targeted by PDS. Thus, additional mechanisms could impact the binding of the G4-ligand to certain G4s, and the G4-folding might be influenced by complex regulations.

Finally, the high selectivity of PDS for G4s inspired the generation of the SiR-PyPDS (Figure 1.13B) and SiR-iPyPDS (Figure 1.13C) probes used for the live-cell single-molecule fluorescence imaging of G4s presented in Section 1.4.2 [65]. In particular, SiR-PyPDS was prepared using an analogue of PDS with improved lipophilicity [73], called PyPDS that has been bound to a red fluorophore Silicon-Rhodamine (SiR) (Figure 1.13B, D and E). The non G4-binding control, SiR-iPyPDS presents an amino side-chain on the quinoline ring (Figure 1.13C), which prevents the formation of the flat molecular conformation required for optimal G4-binding (Figure 1.13F). Beside the direct visualisation of individual G4s in real time in living cells, this work confirmed that G4s are important for replication and transcription as the chemical inhibition of these processes reduced the presence of G4s in cells, therefore confirming the ChIP-Seq evidence obtained by Hänsel-Hertsch et al. [71].

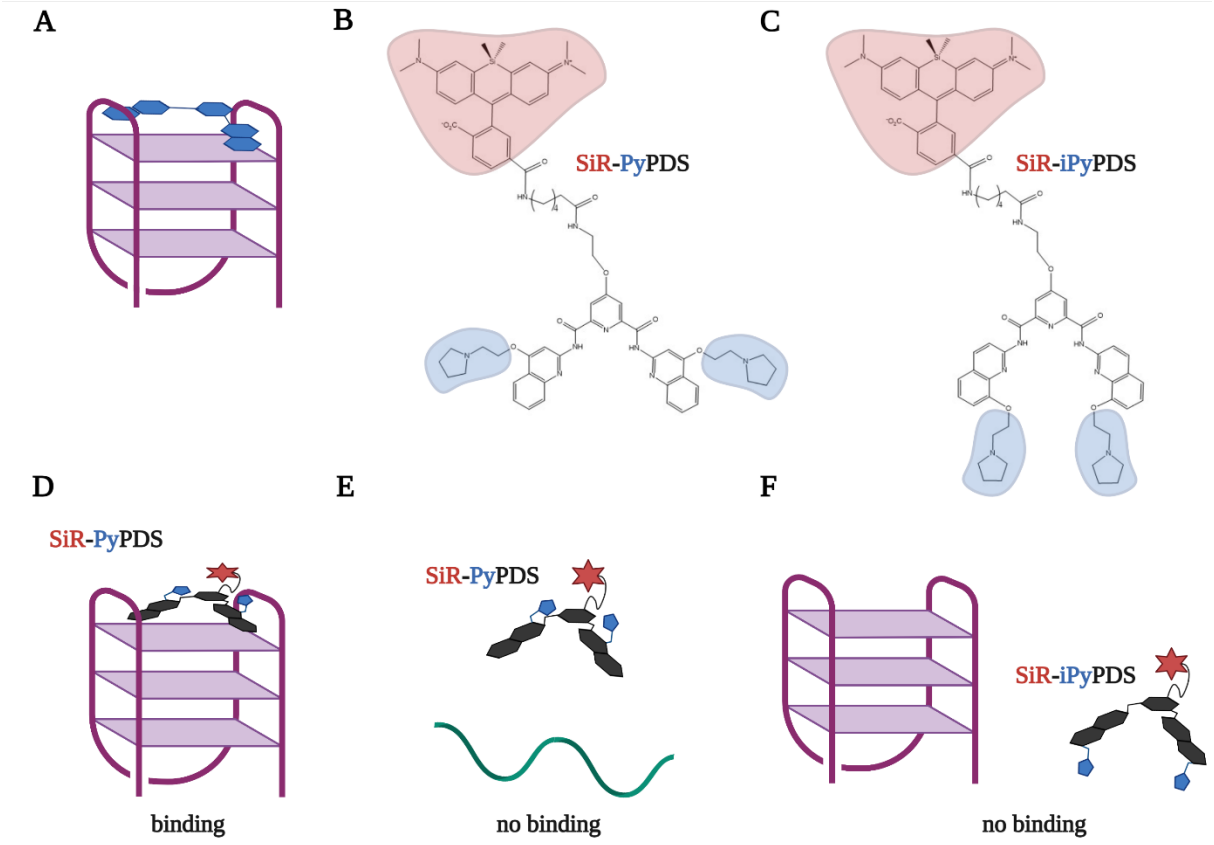

Figure 1.13. PDS is selective for G4 structures. (**A**) PDS (blue molecule) binds G4s (in purple) through interactions with the terminal G-tetrad. (**B**) Representation of SiR-PyPDS small molecule where in blue are highlighted the more lipophilic regions (Py) attached to PDS while the SiR fluorophore is highlighted in red. (**C**) Representation of SiR-iPyPDS control where the amino side-chains are attached on the quinoline ring (iPy, highlighted in blue) on the backbone of PDS. The SiR fluorophore is highlighted in red. (**D**) Representation of the interaction between SiR-PyPDS and a G4 structure. (**E**) SiR-PyPDS is not able to interact with ssDNA (in green). (**F**) SiR-iPyPDS control, is not able to bind G4s. Figures B-F have been adapted from Di Antonio et al. [65].

1.6.2 CX-5461

Quarfloxin (also known as CX-3543) inhibits the RNA Polymerase I (RNAPI) transcription and induces apoptosis in cancer cells through binding to ribosomal DNA (rDNA) G4s, displacing the interaction between Nucleolin protein and G4s [145]. Despite clinical trial studies showing the ability to reduce tumour growth in xenograft models of breast and pancreatic cancers, issues with the bioavailability of the molecule required the development of a next generation molecule, which needed to be well tolerated and with an antitumor response against solid tumours *in vivo* [137,146]. This led to the generation of CX-5461 (Figure 1.12D).

Lymphoma caused by the hyperexpression of *myc* oncogene, were extremely sensitive to treatment with CX-5461, without affecting the normal B-lymphocyte population [146,147]. Further analysis revealed that CX-5461 reduces the interaction of SL1 pre-initiation complex and RNAPI to rDNA promoters with consequent induction of a nucleolar stress pathway, which activates a p53-mediated apoptosis of the malignant cell [147]. Although CX-5461 is the first in-human molecule tested to treat patients with advanced haematological cancers and is used mainly as RNAPI inhibitor [138,146,148], this molecule originates from an established G4-ligand, suggesting that its similarity to quarfloxin could make it a good G4-ligand. This hypothesis was confirmed by H. Xu and co-workers who demonstrated that, in addition to the inhibition of rRNA transcription, CX-5461 is a potent G4s binder and stabiliser, which causes replication dependent DNA damages upon binding to G4 structures [138]. Interestingly, treatment of different cell lines with CX-5461 revealed more apoptotic cell death in *brca1/2* deficient ($^{-/-}$) cells (Figure 1.14A). Treatment of these cells with inhibitors for RNAPI, RNAPII, and protein translation elongation showed that *brca2*$^{-/-}$ cells present high specific sensitivity only to CX-5461 with lower sensitivity to transcription and translation inhibition. The fact that CX-5461 was known to inhibit RNAPI transcription but inhibition of this pathway in absence of the molecule did not affect the vitality of *brca2*$^{-/-}$ cells, suggested that inhibition of rDNA transcription is not an important mechanism of CX-5461 toxicity in *brca2*$^{-/-}$ cells. Because *brca1/2*$^{-/-}$ cells present unfunctional homologous recombination (HR) DNA damage repair mechanisms, Xu and colleagues investigated DNA damage response upon exposure to CX-5461. This analysis revealed stronger γH2AX and 53BP1 DNA damage *foci* signal in *brca2*$^{-/-}$ cells compared to the wide-type (WT) (Figure 1.14B). The increased number of DNA damage *foci* in *brca2*$^{-/-}$ cell suggested that this deficiency in repairing CX-5461-induced DNA damages could lead to chromosome aberration and lethality. As expected, increased chromosome abnormalities were observed in *brca2*$^{(-/-)}$ cells in the presence of the ligand (Figure 1.14C). Interestingly, ChIP-Seq data showed that the presence of DNA damage *loci* after the treatment with CX-5461 are enriched at G4 sequences in human genome.

Biophysical and immunofluorescence analysis using BG4, finally proved the ability of the molecule to interact with G4s and showed an increased number of BG4 foci in *brca2*$^{-/-}$ cells after treatment with either CX-546 or quarfloxin (Figure 1.14D), suggesting that the sensitivity of *brca1/2*$^{-/-}$ cells to this molecule is likely due to their inability to repair the DNA damage caused by G4 stabilisation and G4 accumulation during DNA replication upon treatment with CX-5461. In addition, the tests of the molecule in xenograft models with *brca2*

knockout tumour cells showed a specific reduction of the tumour compared to the WT tumorigenic cells. These encouraging results proposed the employment of G4-ligands as a possible therapeutic treatment for cancers with deficiencies in DNA repair and replication pathways [138]. Therefore, the G4-ligand CX-5461 has been tested as therapeutic strategy in patients with *brca1/2*$^{-/-}$ breast cancer and it is now undergoing interventional clinical trials (NCT04890613).

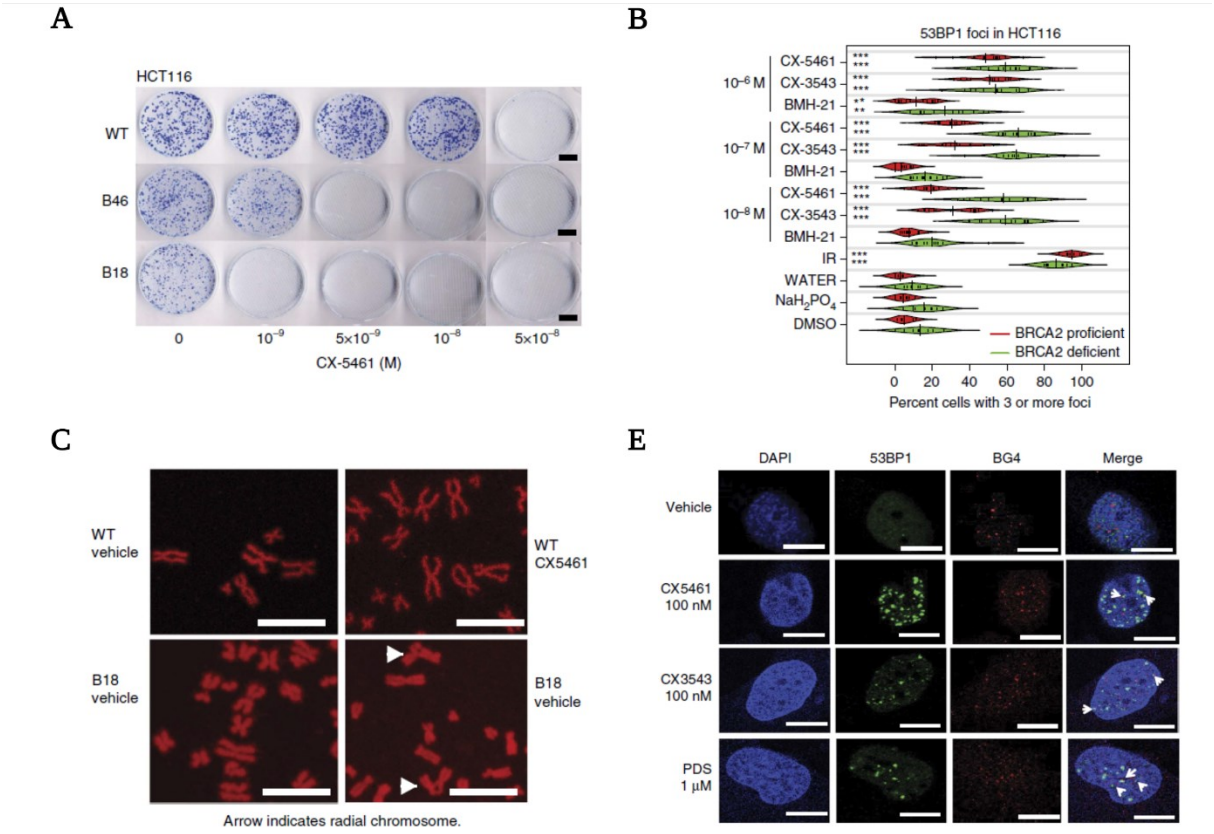

Figure 1.14. CX-5461 is a G4s binder that causes replication dependent DNA damages upon G4-binding in *brca1/2* deficient cells. (**A**) Assay showing the reduction in the colony formation capacity of HCT116 *brca2* deficient cells (B46 and B18) compared to the formation capacity of WT after treatment with CX-5461. (**B**) Bean plot showing the percentage of HCT116 cells presenting three or more 53BP1 *foci* after 24 hours treatment with the indicated drug. (**C**) Mitotic chromosome spread showing increased chromosome abnormalities in *brca2* deficient cells (B18) in presence of CX5461. Chromosome abnormalities are indicated with white arrows. (**D**) Increased number of 53BP1 and BG4 *foci* after treatment with G4-ligands CX-5461, quarfloxin (CX-3543) or PDS. White arrows indicate co-localisation between 53BP1 and BG4. The images have been adapted from H. Xu et al. [(138)].

1.7 Regulation of G-quadruplex Formation by Helicases

G4-formation is highly dynamic in cells and is carefully regulated by cellular proteins. Affinity proteomic experiments and pull-downs from nuclear lysates allowed the identification of a wide range of proteins that interact with G4s [15], including chaperones and helicases. Chaperones can promote G4-formation, whilst helicases can resolve them [77]. In Section 1.5.2, the helicase DOG1 was introduced as a *C. elegans* homolog of the *human* FANCJ helicase. DOG1 is a member of Rad3/XPD family presenting a 5' to 3' processivity and it has been the first helicase observed to regulate G4s structures *in vivo* [96,149]. Mutations of FANCJ helicase give rise to a human chromosomal instability disorder, called Fanconi anaemia. Works in *Xenopus egg* extracts showed that the helicase FANCJ maintains DNA replication through unwinding of G4 structures, and in absence of this helicase, the genomic aberrations are concentrated in G4-rich regions [150,151].

Most of the proteins associated with G4-resolution belong to families of canonical and well-conserved helicases, such as RecQ-like and DEAD box or DEAH box helicase families [78]. *In vitro* analysis demonstrated that RecQ helicases can resolve G4s with a 3' to 5' directionality, and their dysfunction is associated with increased predisposition to cancer and genetic disorders [149]. For instance, Werner syndrome helicase (WRN) and Bloom syndrome helicase (BLM) belong to the human RecQ helicase family and are involved in telomere replication through resolution of telomeric G4s. Loss of these proteins or treatment with G4-ligands decreases the replication level at telomere causing the two syndromes. Beside the ability to resolve telomeric G4s, these two helicases are also associated to G4-resolution during transcription [149,152,153].

M. C. Chen and co-workers reported the first X-ray crystallographic structure of a G4-helicase bound to MYC G4 and proposed a mechanism for G4-unwinding (Figure 1.15A) [154]. The helicase protein DEAH-box helicase 36 (DHX36, also known as G4R1/RHAU) belongs to a DEAH/RHA family of helicases and binds with a very high affinity to DNA G4s (~77 pM) and RNA G4s (~39 pM) displaying a 3' to 5' adenosine triphosphate (ATP)-dependent G4-specific resolvase activity *in vitro* [155,156]. Experimental evidence in cell lysates [157,158] showed the ability of DHX36 to specifically resolve 3'-tailed (9 nt-tail), both tetramolecular G4s, and unimolecular DNA and RNA G4s with a preference for parallel

topology compared to antiparallel or mixed topology [159]. As a member of the DEAH/RHA family, DHX36 presents two RecA-like domains (RecA1 and RecA2) in its core, preceded by the N-terminal domain, which contains a glycine-rich region and a conserved region (DHX36-specific motif, DSM) necessary for the G4 binding [155,160]. The co-crystal structure of bovine DHX36 bound to MYC G4 showed that the MYC G4 structure is reorganised upon DHX36 binding where a single G is pooled out of the folded G4. This rearrangement decreases the stability of the G4 instead of completely resolve the secondary structure of the MYC G4 [154]. ATP hydrolysis might be required to release the single stranded residue from DHX36, as suggested by the crystal structure and single-molecule fluorescence resonance energy transfer (FRET) analysis [154]. DHX36 is a multifunctional helicase expressed in two isoforms, which is mainly localised in the nucleus and cytoplasm of cells [155]. For example, this protein plays an important role in replication, as treatment of *Xenopus* egg lysate with G4-ligands showed increased enrichment of G4-resolving helicases, with DHX36 being the most strongly enriched. Interestingly, studies in *Xenopus* egg lysate using dsDNA plasmids that contain a G4 motif or a control sequence (Figure 1.15B) also revealed that DHX36 could promote resolution of G4 structures formed either within the leading or lagging strand during replication. A plasmid-based ChIP technique (Figure 1.15C) revealed a joint activity between DHX36 and FANCJ in resolving G4 structures and promoting efficient DNA replication [161]. DHX36 is also involved in telomere maintenance. As briefly mentioned in Section 1.5.1, one subunit of telomerase is composed of telomeric RNA (or called hTERC). The 5'-end of hTERC can adopt RNA G4s and is key in the regulation of the enzyme [162]. Interestingly, DHX36 binds and resolves these G4s promoting the formation of a helical structure within hTERC, called P1, which is essential for the accurate reverse transcriptase activity of the enzyme [163]. DHX36 might also play a regulatory role in oncogenesis as its G4-resolvase activity has been also associated with transcriptional activation, post-transcriptional regulation, and RNA degradation of several oncogenes and genes involved in cell growth and differentiation [155].

Another example of G4-resolvase is ATRX, which belongs to the family of SWI2/SNF2 DNA helicase/ATPase. Patients presenting alteration in ATRX activity suffer a X-linked mental retardation syndrome (ATR-X syndrome). In these patients, the expression of G-rich genes involved in the regulation of alpha-globin is compromised, supporting the ATRX role in G4-handling [164,165].

Cockayne syndrome B (CSB) protein is another example of a helicase belonging to the SWI2/SNF2 family [166,167]. Recently, biochemical studies have revealed a modest activity of CSB in resolving a rDNA G4 and suggested that this is essential to prevent the premature ageing observed in patients lacking functional CSB [168].

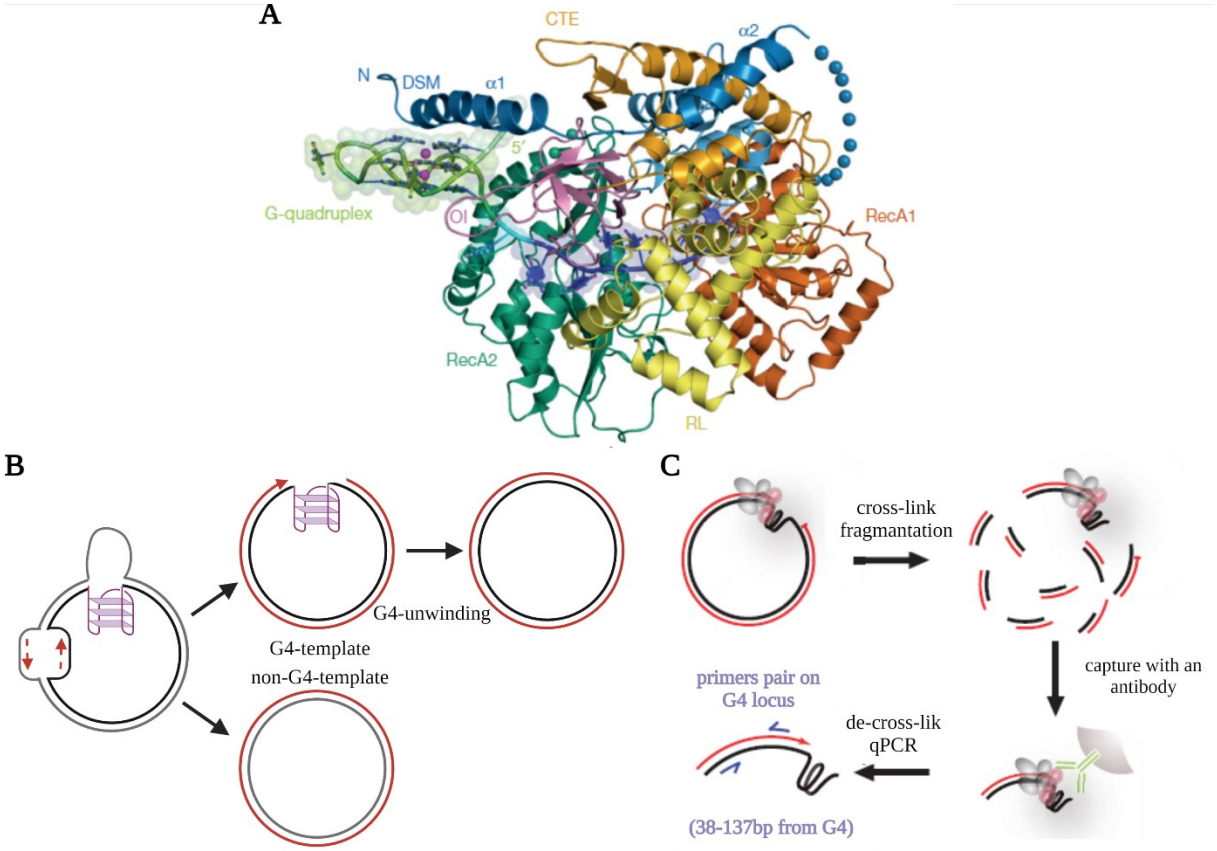

Figure 1.15. G4-helicases can bind and resolve G4 structures. (**A**) Cartoon representation of the crystal structure (3.79 Å-resolution, PDB ID: 5VHE) of DHX36 bound to MYC G4 (in light green) where different domains are shown with different colours (adapted from M. C. Chen et al. [154]). (**B**) dsDNA plasmid containing a G4 motif used for replication studies. (**C**) Scheme of the plasmid-ChIP technique. B and C were redrawn (B) or adapted (C) from Sato et al. [161].

1.8 Loss of G-quadruplex Regulation and Premature Ageing

The main knowledge of G4s and ageing arose from the treatment with G4-ligands or studies of diseases and syndromes caused by loss of activity of cellular G4-helicases, such as WRN and CSB. These syndromes are characterised by premature ageing, suggesting that G4 stabilisation and accumulation causes transcriptional stalling, and transcriptional alteration, which might be the key to understanding the mechanisms of cellular senescence.

Within eukaryotic genomes, ribosomal genes, or rDNA, are among the most enriched for PQS and have the potential to form G4 structures. rRNA together with ribosomal proteins, generates ribosomes. Therefore, tight control of G4 formation is necessary for cells to maintain proper protein synthesis and growth rates [169]. Although direct evidence of the role of G4 helicases such as FANCJ, RTEL1, WRN or BLM in resolving rDNA G4 is missing, Scheibye-Knudsen et al. found defective rDNA transcription due to loss of either Cockayne Syndrome (CS) complementation group gene A (*CSA*) or CS complementation group gene B (*CSB)* [168,169]. Treatment of a *Caenorhabditis elegans* model with G4-ligands revealed accelerated ageing in this model. Moreover, they observed a modest activity of CSB protein in resolving one particular rDNA sequence, suggesting that lack of CSB may be involved in defective rDNA transcription caused by G4 accumulation, and consequently, accelerated ageing [168].

1.9 Cockayne syndrome B (CSB)

CSB, also known as excision repair cross-complementing protein group 6 (ERCC6), is a 168 kDa protein made by 1,493 amino acids and codified by *ercc6* on chromosome 10q11 [166]. As mentioned in Section 1.7, CSB belongs to a family of chromatin remodelling proteins, namely SWI2/SNF2-family, which are DNA-dependent ATPases containing a well-conserved helicase domain that includes seven canonical ATPase motifs typical of DNA and RNA helicases (Figure 1.16A) [166,167]. Alignment of CSB proteins from different chordates identified conserved basic residues and a small glycine-rich region within the N-terminal region of the protein. Moreover, the N-terminal region of CSB presents a stretch of acidic amino acids (AD), which may be important for the activity of CSB. In fact, this domain is commonly observed in numerous nuclear proteins, including different SWI2/SNF2 chromatin remodelling proteins and transcriptional activators where the acidic domain interacts directly with components of the RNAP machinery or gene-specific activators [170]. Deletion experiments showed that CSB self-regulates its association with chromatin and that the CSB-chromatin interaction depends on the ability of the protein to hydrolyse ATP. A proposed mechanism suggested that the stabilisation of the interaction between CSB and chromatin requires the contact between DNA, the ATPase domain, and a DNA-binding surface present within a helicase- "like" C-terminal region of CSB (Figure 1.16A). This autoregulative model proposed that the N-terminal region of CSB could normally prevent the stable chromatin association by occluding the DNA-binding surface within the helicase- "like" C-terminal region. Subsequently, ATP hydrolysis is necessary to promote conformational changes within the protein that relieve the inhibitory effect imposed by the N-terminal region [171]. The C-terminal domain of CSB contains an ubiquitin binding domain (UBD) (Figure 1.16A), which is part of a larger winged-helix domain (WHD) [166,172]. UBD is important for the recruitment of CSB to double strand breaks (DSBs), and during transcription coupled nucleotide excision repair (TC-NER) mechanism [166,173]. Recently, pull-down and mutational experiments identified a CSA-interaction motif (CIM) upstream the UBD domain (Figure 1.16A), which is necessary for the recruitment of the CSA protein to the RNAPII stalling-site [166,174].

Biochemical evidence revealed that CSB can bind both dsDNA and ssDNA as a homodimer [175]. However, CSB lacks dsDNA resolvase activity, proposing that, despite the presence of a helicase- "like" domain, CSB is not a canonical helicase [176]. Contrarily, CSB showed ATPase activity with a turnover number (k_{cat}) of ~3 min^{-1}, which increased to 45–53 min^{-1} in

the presence of ss- or dsDNA [176]. Scanning force microscopy (SFM) of CSB bound to DNA revealed that CSB can modify the double helix conformation of DNA by promoting wrapping and unwrapping of dsDNA around itself (Figure 1.16B) [177]. This observation is further supported by the shortened length of the DNA strand measured upon CSB binding, which occurs in an ATP-dependent fashion. For example, SFM showed that wrapping of the DNA happens when the ATP binds to the protein, whereas ATP hydrolysis results in the unwrapping of the DNA. These results suggest that CSB can alter nucleosome positioning and affect the interaction between DNA-binding proteins and DNA, thus resulting in gene expression changes [177].

Additional experimental evidence revealed that CSB is involved in the regulation of numerous cellular pathways, including transcription, TC-NER, DNA base excision repair (BER), chromatin remodelling, RNAPII processing, Nucleolin regulation, redox homeostasis, mitochondrial functions, and rDNA transcription [166]. Consistent with the wide range of cellular pathways in which CSB is involved, cellular localisation studies revealed that CSB is mainly localised in the nucleus, with some enrichment at the nucleolar level (Figure 1.16C) [178]. A small fraction of CSB has also been found in the mitochondria, while only traces (below 10%) have been observed in the cytoplasm of cells [178]. Computational predictions and confocal imaging of cells transfected with different GFP-tagged portions of CSB, revealed a precise regulation of the intranuclear distribution of the protein, which relies on the presence of different nuclear and nucleolar localisation sequences (NLSs and NoLSs, respectively) (Figure 1.16B) [178].

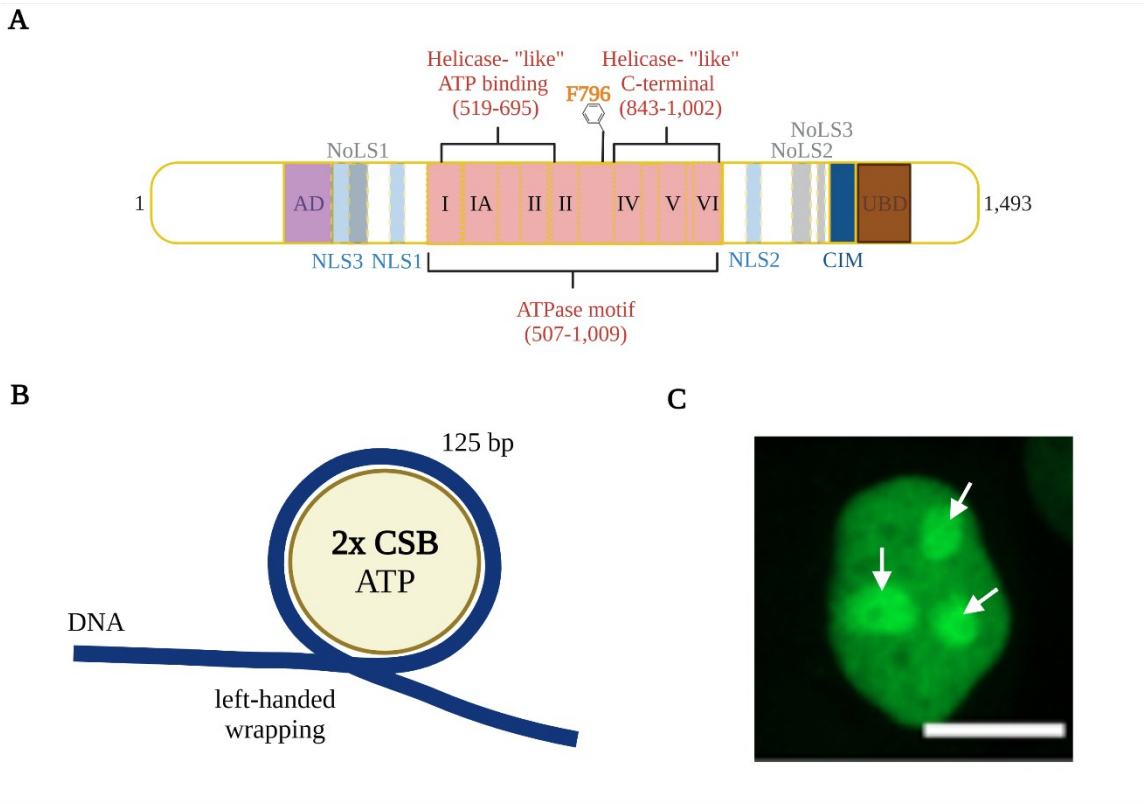

Figure 1.16. CSB is not a canonical helicase, and is localised within the nucleoli of cells. (**A**) Schematic domain organisation of CSB protein (1–1,493). The acidic domain (AD) is indicated in purple, the three nuclear localisation sequences (NLS1-3) and three nucleolar localisation sequences (NoLS1-3) are indicated in light blue and grey, respectively. The conserved ATPase domain (507–1,009) containing the canonical seven ATPase motifs (I, IA-VI) is represented in red. Within the ATPase motif, two regions have been identified. A helicase- "like" ATP binding site (519–695) and a helicase- "like" C-terminal region (843–1,002). The CIM and UBD at the C-terminal of the protein are indicated in blue and brown, respectively. A "pulling hook" residue [179] (phenylalanine 796) is reported as F796 in orange. This figure has been redrawn from Tiwari et al. and Iyama et al. [166,178]. (**B**) Model showing a CSB dimer (yellow) wrapping 125bp of DNA around its surface in the presence of ATP. This figure was redrawn from Beerens et al. [177]. (**C**) Picture of a single cell taken from Iyama et al. [178] showing nuclear localisation of CSB (green staining) with enhanced localisation of the protein within the nucleoli of the cell (white arrows). Bar: 10 µm.

1.9.1 Role of CSB in TC-NER Pathway

Among the various cellular activities associated to CSB, TC-NER is the main mechanism regulated by this protein. TC-NER is a repair pathway that removes DNA lesions induced by ultraviolet (UV) radiations [180], which cause the stalling of RNAPII during transcription [171,181]. CSB initiates TC-NER by binding to the stalled RNAPII and promoting the assembly of the repair machinery through recruitment of CSA, a UV stimulated scaffold A (UVSSA), and other NER factors [182]. Upon binding to the blocked RNAPII, CSB recruits CSA to the lesion through its CIM domain. CSA is a subunit of the E3-ubiquitin ligase complex, which triggers CSB ubiquitylation releasing transcription-coupled repair (TCR) factors anchored via CSB. Moreover, CSB assists UVSSA activity to recruit the NER core complex transcription factor IIH (TFIIH) [174,182]. Then, the dsDNA is unwound by TFIIH with consequent repair by excision, followed by DNA synthesis and ligation [182] (Figure 1.17A). Recently, the resolution of the structures of both *S. cerevisiae* RNAPII-Rad26 (the yeast human homologues of RNAPII and CSB) and human RNAPII-TCR (Figure 1.17B) complexes by cryo-electron microscopy (cryo-EM) proposed a novel translocation model where CSB promotes the movement of the stalled RNAPII through a "pulling hook" residue that alters the DNA trajectory (Figures 1.16A and 1.17B). In detail, the binding of ATP induces the remodelling activity of CSB, which pulls the template strand of the upstream DNA, whereas the "pulling hook" within the ATPase domain pulls the non-template strand in the same direction. As CSB is anchored to the RNAPII, the pulling of the upstream DNA promotes forward translocation of RNAPII, thus confirming the role of CSB in the transcription elongation upon TC-NER resolution [179,182,183].

Transcription elongation in the absence of DNA lesions has also been investigated for both CSB and Rad26. The intrinsic chromatin remodelling nature of these proteins allows for an ATP-dependent directional DNA translocase activity, which assists RNAPII to bypass nucleosomes or non-bulky lesions (Figure 1.17C) [182,184,185]. Upon interaction with histone chaperones, the nucleosome remodelling activity of CSB reaches levels comparable to a human ATP-dependent chromatin assembly factor (ACF). Notably, under physiological conditions, the nucleosome remodelling activity of CSB is typically 10-fold lower than that of ACF [166], suggesting that the main cellular role of CSB might not be related to physiological chromatin remodelling. Conversely, the localisation of CSB within the nucleoli [178] suggests a

specific activity of this protein within these regions, which may be associated to nucleolar TC-NER or to a role in the nucleolar organisation.

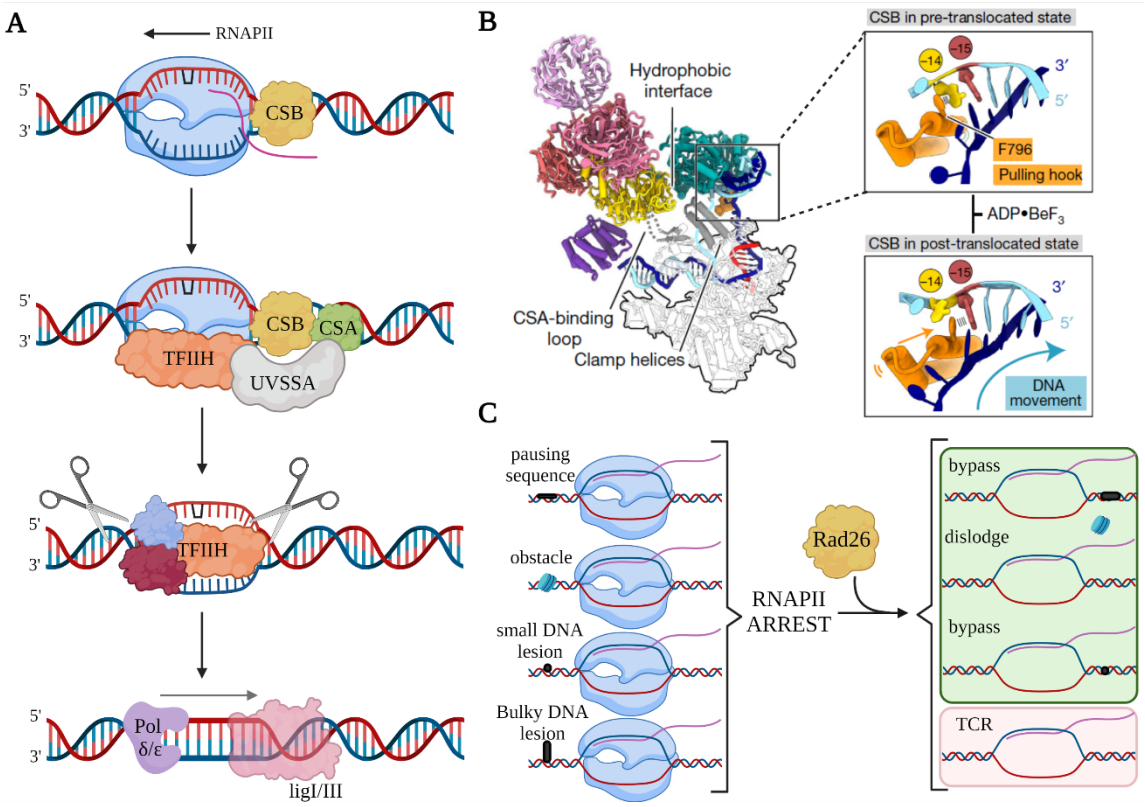

Figure 1.17. CSB and repair of DNA lesions. (**A**) Representation of the mammal TC-NER. RNAPII (light blue) blocked at a DNA lesion site (black) is bound by CSB, which recruits CSA and UVSSA. UVSAA recruits TFIIH, which promotes the unwinding and the cut of the side regions of the lesion (represented with two scissors). Finally, the DNA is synthesised by DNA polymerase δ and ε and the gap is ligated by ligases. This figure has been redrawn from Duan et al. [182]. (**B**) Ribbon model (right-side) of the RNAPII-CSB-CSA-DDB1-UVSSA complex obtained from cryo-EM resolution (PDB ID: 7OO3, 2.80 Å-resolution), including RNAPII (white-transparent element), CSB (green), CSA (yellow), DDB1 (DNA damage-binding protein 1; multiple pink subunits), and UVSSA (purple). The two zoomed-in boxes report the upstream DNA fork interacting with the "pulling hook" (orange residue) either in the pre-translocated (top) or post-translocated (bottom) states. This figure has been adapted from Kokic et al. [179]. (**C**) Schematic representation of the ability of Rad26 to bypass and dislodge pausing sequences or small lesions and obstacles such as nucleosomes (top green box), which arrest RNAPII (represented in light blue). Bulky DNA lesions are recognised by Rad26 but necessitate the activation of TCR for transcriptional recovery (bottom red box). Figure redrawn from Xu et al. [183].

1.9.2 Role of CSB in Ageing

Cockayne Syndrome (CS) is a rare, autosomal-recessive disorder, characterised by UV-sensitivity, growth retardation, progressive neurological degeneration, and premature ageing, which ultimately cause premature death typically in the second decade of the life of patients [186]. The observed phenotype is mediated by different biological processes, including genetic and epigenetic alterations that cause loss of functionality of cells and tissues with subsequent development of premature ageing-related diseases [187]. Approximately 70% of CS cases have been associated with at least 78 different mutations of the *CSB* gene, including, for example, deletions, missense mutations, and frameshifts [188,189]. One key characteristic of *CSB*-deficient cells is a persistent activation of the enzyme poly-ADP ribose-polymerase 1 (PARP1) [168]. PARP1 is recruited to the sites of DNA damage where it recognises and binds single-strand DNA breaks (SSBs). PARP1 uses nicotinamide adenine dinucleotide (NAD^+) as a cofactor to covalently modify target proteins by adding linear or branched chains of ADP-ribose (PARylation). These modifications result in chromatin relaxation and recruitment of DNA damage repair factors [166,187,190,191]. Furthermore, the activity of PARP1 increases the production of lactate, a phenomenon observed in both normal ageing conditions and models lacking CSB, and it is also associated with mitochondrial dysfunction [168,192].

Increasing evidence shows loss of heterochromatin during ageing, with reduction of H3K9me3. Consistently, immunofluorescence imaging and biochemical analysis of fibroblast derived from CS patients (CS1AN) showed enlarged nuclei, reduced level of H3K9me3, and decreased expression of histone H3 [187,192]. *In vitro* studies demonstrated that the high level of compaction of heterochromatin protects DNA from ROS-induced damages and genome instability [187]. Thus, the increased UV-sensitivity observed in CS patients might be due to enhanced vulnerability of the cells to DNA damage caused by the loss of heterochromatin.

Bioinformatic and ChIP-Seq analyses in CSB-deficient cells identified increased PARP and low H3K9me3 levels in TSS, TSS flanking regions, and rDNA coding regions (18S, 5.8S and 28S) [192]. Furthermore, RNA sequencing (RNA-Seq) analyses performed by M. Scheibye-Knudsen and co-workers revealed increased transcriptional stalling at rDNA level in CSB-deficient cells [168], strongly suggesting that the lack of heterochromatin and nucleolar TC-NER may explain the high predisposition of these cells to DNA damage. Interestingly, studies performed with normal human lung fibroblasts showed loss of CSB and heterochromatin during normal replicative senescence. This evidence further supports the correlation between

loss of CSB and senescence, however, the exact mechanism describing the decreased expression of CSB during healthy aging still remains unclear [187,193].

Recent studies revealed that the interaction between CSB and Nucleolin is required to stimulate rDNA synthesis in the nucleoli [194]. Notably, rDNA sequences are highly rich in Gs and prone to form G4 structures, which are stabilised upon binding to Nucleolin resulting in decreased rDNA transcription [145,195]. Immunocytochemistry data showed increased levels of both Nucleolin and PARP in the nucleoli of CSB-deficient cells. This is particularly relevant especially considering that PARP1 also binds G4s with high affinity and stimulates its enzymatic activity [168,196]. Moreover, a direct correlation between PARP1 activation and premature ageing was also observed in *C. elegans* models treated with G4-ligands, suggesting that stabilisation of G4s causes stalling of RNAPI and high and persistent activation of PARP1 that triggers premature ageing [168]. M. Scheibye-Knudsen and colleagues questioned whether CSB could resolve G4s and promote rDNA transcription. They demonstrated that CSB can resolve a single rDNA G4 in an ATP-independent fashion [168].

Taken together, these data propose that lack of CSB within the nucleolus triggers two main effects (Figure 1.18):

- Lack of heterochromatic compaction at TSS and rDNA levels, which makes the DNA more prone to DNA damage [187]; and
- Formation of rDNA G4s that are not resolved by CSB and cause persistent activation of PARP1, lactate accumulation, and mitochondrial damage [168].

M. Scheibye-Knudsen and colleagues observed rDNA G4 resolution mediated by CSB, revealing a G4-helicase activity for this protein, albeit this protein has never been classified as a canonical helicase. Therefore, the G4-resolution mechanism of CSB might be different from what has been observed with the other canonical helicases and requires additional characterisation. Furthermore, in the report published by Scheibye-Knudsen, CSB displayed only a modest rDNA G4 resolution activity that was limited to a single rDNA substrate tested. The modest G4-resolvase activity of CSB is in stark contrast with the strong ageing phenotype caused by lack of the protein in CS patients. Moreover, G4 helicases are highly abundant in cells (see Section 1.7), hence CSB G4-resolvase activity could be easily compensated by the presence of other G4s helicases and prevent CS to be developed.

Altogether these observations indicate a more complex role of CSB in regulating rDNA homeostasis, which has been the primary focus of this post-graduate research, recently published in the *Journal of the American Chemical Society* [197].

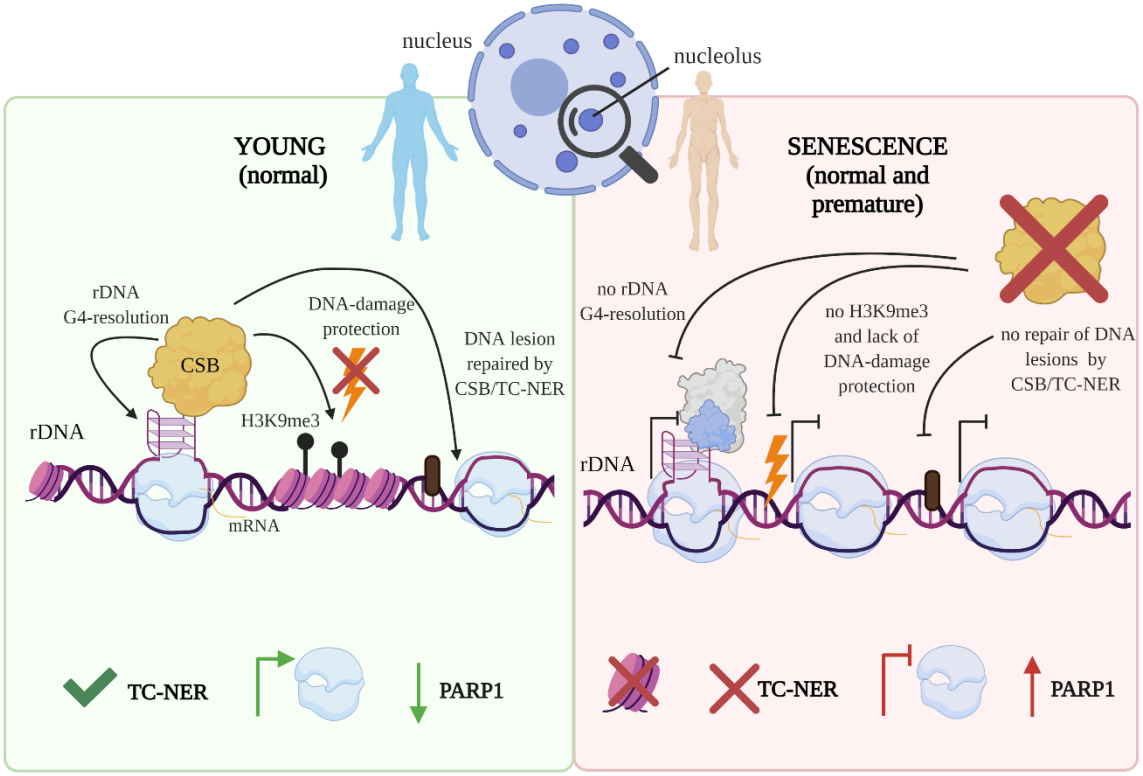

Figure 1.18. Overview of chromatin changing and rDNA G4 regulation during ageing. On the left, the nucleoli of young (normal) individuals contain active CSB. Here, CSB protects rDNA from DNA damage (UV, ROS or radiation) by maintaining the heterochromatic state of rDNA (H3K9me3). TC-NER machinery is allowed by CSB, therefore, in presence of a bulky lesion, TC-NER could repair the damage. G-rich rDNA sequences are prone to form G4s that are resolved by CSB. This rDNA regulation given by CSB prevents RNAPI (represented as a light blue protein) stalling and assures low PARP1 levels. Contrarily, on the right, during normal or premature senescence, lack of CSB reduced the heterochromatic state of rDNA, which is more susceptible to DNA damages that cannot be repaired by TC-NER due to the lack of CSB. In absence of CSB, rDNA G4s are not resolved and accumulate within rDNA causing RNAPI pausing. Unresolved G4s might be stabilised by Nucleolin and bound by PAPR1 (represented as blue and grey globular proteins bound to G4), which activate the PARP1-cascade that causes loss of NAD^+ and increased of lactate, and are typical features of ageing.

1.10 Aims of the Project

Uncontrolled G4 accumulation has a strong impact on DNA stability and cellular homeostasis, with an increase in G4-prevalence being associated with cancer and premature ageing. The correlation between G4s and cancer is more broadly acknowledged and has consequently been investigated in depth. In contrast, evidence correlating G4s and ageing is much more limited.

The aim of this project is to investigate the fundamental role of G4s in promoting cellular senescence and ageing phenotypes, using Cockayne Syndrome as a model. To achieve this, we chose CSB as an endogenous model to further our understanding of the mechanisms by which proteins recognise, bind to, and resolve G4s and the consequences of the absence of this protein in Cockayne Syndrome patients.

To address this, this work has focused on the following objectives:

1 **Production** of the full length CSB protein (CSB-FL) and a smaller helicase- "like" domain of the protein (CSB-HD).

2 **Full biochemical characterisation** of the G4-binding and resolvase activity of CSB-FL and CSB-HD towards a panel of different nucleic acids structures and sequences.

3 **Cellular investigation** of CSB binding to G4s in normal and CSB-derived cell lines by immunofluorescence experiments.

Chapter 2

2 General Techniques

2.1 Approaches for Protein Purification

AFFINITY PURIFICATION USING NICKEL/COBALT RESIN

Nickel and Cobalt resins are widely used for immobilised metal affinity chromatography (IMAC) and allow purification of recombinant proteins containing a poly-histidine tag on either N-terminus or C-terminus [198] (Figure 2.1A-D). Nickel-NTA and Cobalt-NTA are the most common type of resins. The resin consists of beaded agarose derivatised with nitrilotriacetic acid (NTA) chelation moieties loaded with either nickel (Ni^{2+}) or cobalt (Co^{2+}) ions. Generally, Nickel resins allow higher protein yields as his-tagged proteins have a higher affinity for this metal compared to cobalt. However, in a cellular lysate, multiple interactors that contain histidine might be present and weakly interact with the resin (Figure 2.1B) with consequent presence of contaminants in the eluted sample. Compared to Nickel ones, Cobalt resins give purer eluates but lower protein yields. Washing steps are usually recommended to remove interactors that are weakly bound to the resin. Because of its structural similarity with histidine, low concentrations of imidazole could be used to remove unspecific binders, as it also interacted with nickel or cobalt, competing with the unspecific interactors for the binding to the resin (Figure 2.1C). Recombinant target proteins are usually eluted using a high concentration of imidazole, which assures complete competition with the target his-tagged protein (Figure 2.1D). Alternatively, strong chelating agents, such as EDTA could be used for elution. However, the use of strong chelating agents could strip the ions from the resin, with consequent elution of protein-nickel/cobalt complexes.

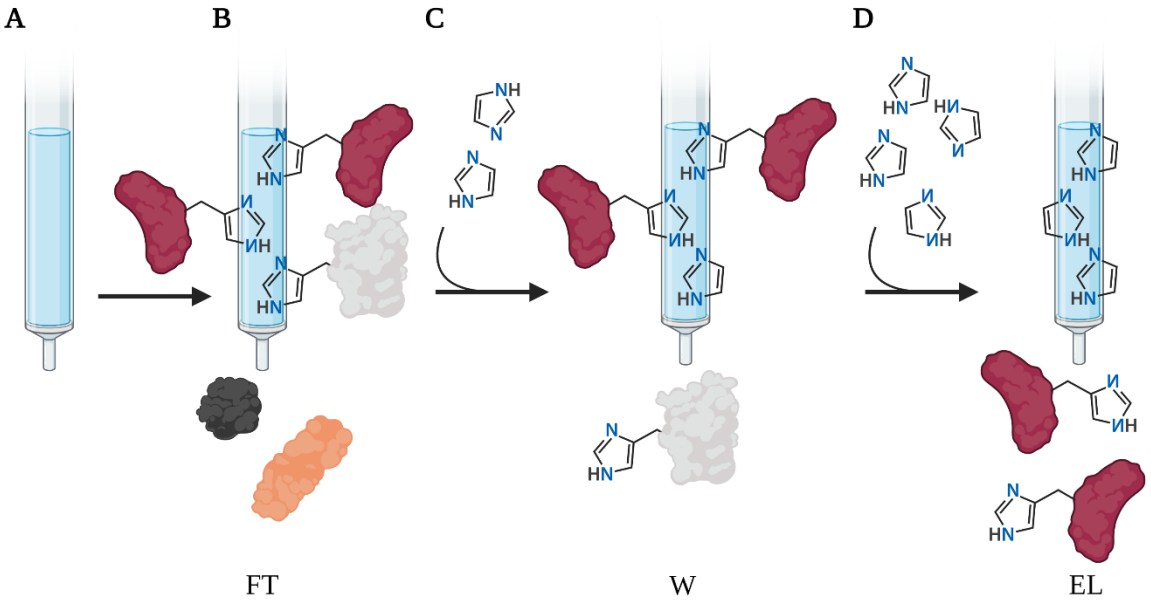

Figure 2.1 Affinity purification of histidine-tagged molecules using Nickel-NTA (or Cobalt) resins. (**A**) The free nickel resin is added into an empty column and equilibrated with the desired buffer. The clarified cellular lysate is then added to the resin. (**B**) Only the histidine-tagged molecules (in red and grey) can bind the resin while all the untagged molecules (in black and orange) pass through the resin (FT). (**C**) To avoid unspecific contaminations, the resin is washed with low concentrations of imidazole, which competes with weak interacting molecules removed from the resin (W). (**D**) Finally, higher concentrations of imidazole are generally used to detach the target histidine tagged protein from the resin (EL).

AFFINITY PURIFICATION USING HEPARIN RESIN

Heparin chromatography is an affinity chromatography method, which allows the purification of proteins through their interaction with heparin that is immobilised on a porous bead matrix (Figure 2.2) [(199)]. Heparin belongs to a family of glycosaminoglycans, which present very high negatively charged sulphate polysaccharide groups. Heparin could either mimic the polyanionic structure of the nucleic acid and allow specific binding of nucleic acid-binding protein, and other specific affinity ligands, such as enzymes, growth factors, and lipoproteins, or it could act as a cation exchanger through its high content of anionic sulphate groups. In both cases, the interaction can be weakened by increased ionic strength, such as using a linear gradient from 0 to 1-2 M NaCl or KCl. The stronger the interaction with heparin, the higher the concentrations of NaCl/KCl where the protein will be eluted. One of the main advantages of this technique is that it is a simple and very effective chromatography, which allows the

purification of a wide range of proteins without requiring the presence of affinity tags [200]. Because of its high purification potential, it is usually used as a second purification step to further clean the target recombinant protein.

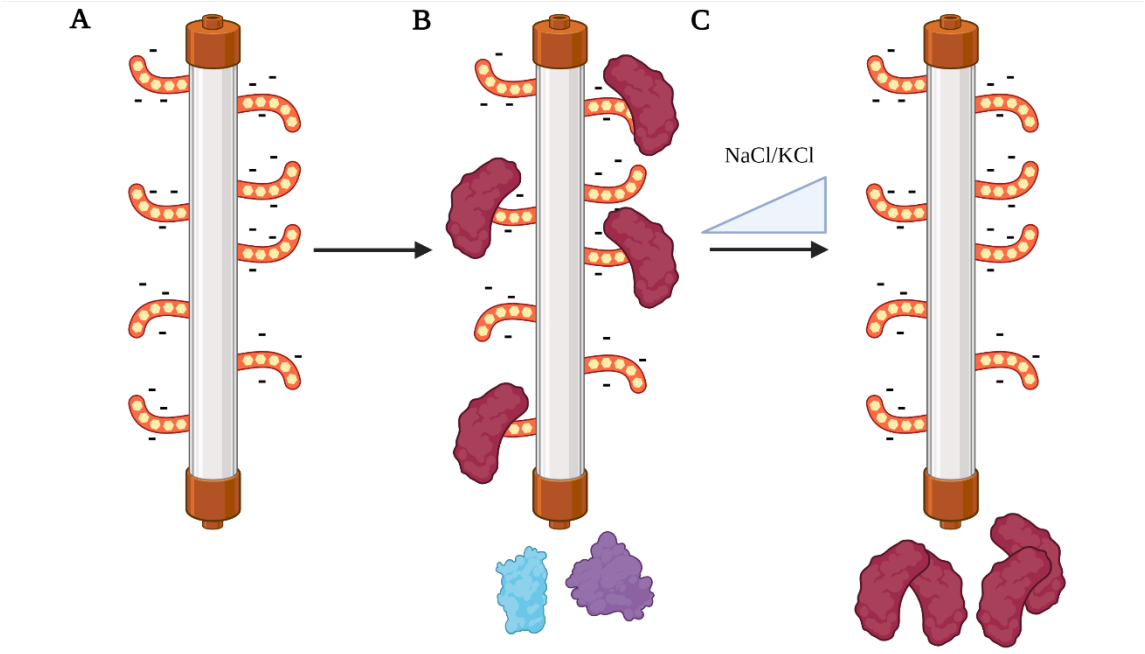

Figure 2.2. Affinity chromatography using heparin column. (**A**) The heparin negatively charged sulphate polysaccharide groups are immobilised on the surface of the column. (**B**) DNA-binding proteins (red interactors) bind the negatively charged polysaccharide groups. (**C**) The bound proteins are eluted with increased concentrations of either NaCl or KCl.

SIZE EXCLUSION CHROMATOGRAPHY

Size exclusion chromatography (SEC) allows the separation of molecules depending on their molecular weight (Figure 2.3) [201]. SEC resins consist of a porous matrix of spherical particles packed into a column. Unlike affinity or ion exchange chromatography, SEC does not require changes in buffer composition and the samples could be directly diluted into a suitable buffer for assay or storage. Moreover, during SEC, the molecules do not chemically interact with the resin, but they remain trapped into the pores of the matrix. Big molecules larger than the biggest pores in the matrix, simply pass through the column as they are too large and cannot enter the pores, and are eluted together in the void volume. Conversely, smaller molecules that have access to the pores are separated and eluted in order of decreasing

size. Depending on the size of the proteins to be separated, different SEC columns with optimal size ranges are available. For example, Superdex-peptide allows the separation of peptides within 100–7,000 Da, Superdex 75 separates proteins within 3,000–70,000 Da, Superdex 200 for proteins between 10,000–600,000 Da and Superose column separates proteins of 5,000–5,000,000 Da. SEC could be either preparative or analytical. Resolution of preparative SEC varies from high to moderate. Analytic SEC is usually performed without fractionation and allows a high-resolution separation used for quality screening of a molecule or its molecular characterisation. SEC could be also used to measure the molecular size of molecules or proteins by comparative inspection between the chromatogram (elution profile) obtained from a target recombinant protein or molecule, and calibration chromatograms created with known standards. Generally, the SEC elution profiles indicate the variation of sample components as they elute in order of their molecular weight, where higher molecular weights can elute earlier than molecules with lower molecular weights do (Figure 2.3).

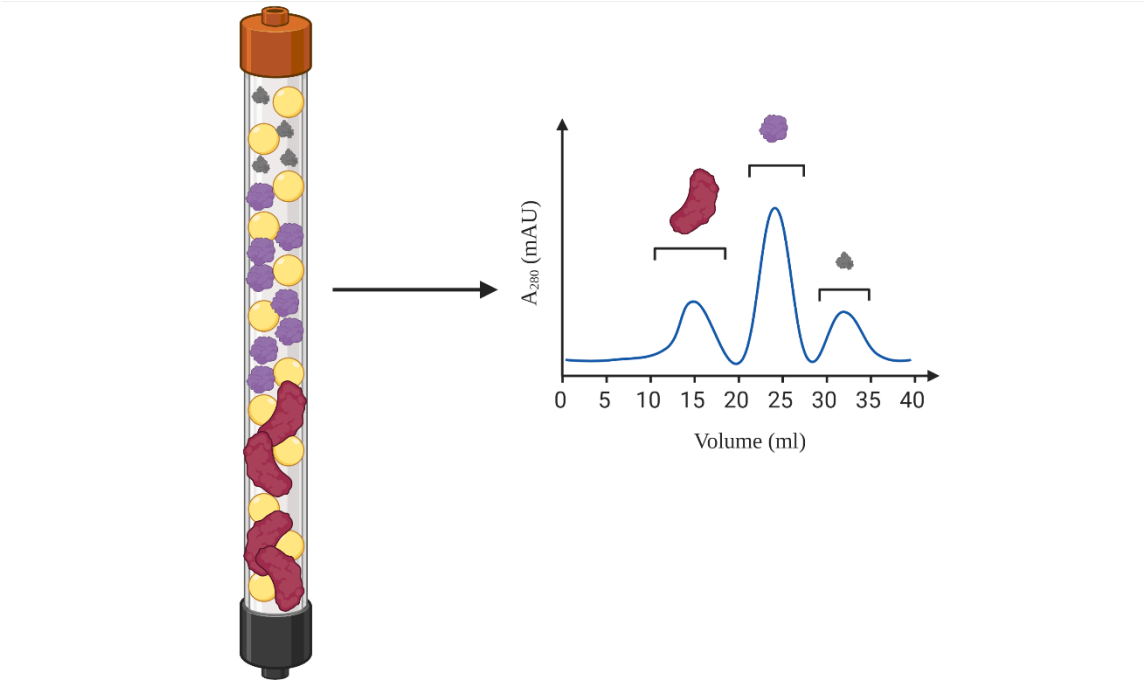

Figure 2.3. Size exclusion chromatography separates molecules depending on their molecular weight. Bigger molecules elute earlier than smaller ones do.

2.2 Approaches to Investigate G-quadruplex Interactors and G-quadruplex Resolution

Over the years, many methods have been developed to identify and characterise the interaction between proteins and between proteins and nucleic acids. As discussed in Section 1.2.2, NMR and X-ray crystallography allow to characterise the three-dimensional structure of G4s and G4-associated molecular complexes. However, both these techniques are laborious, necessitate isolated macromolecules and present some disadvantages related to the experimental equipment, the extreme purity and the high amount of sample required for the analysis [43]. Therefore, experimentally less-demanding techniques that allow to study G4-protein interactions within heterogeneous samples are ideal to investigate their biochemistry. Some of these techniques relevant to this thesis will thus be described in this section.

ELECTROPHORETIC MOBILITY-SHIFT ASSAY (EMSA), FLUORESCENCE POLARISATION (FP), and ENZYME-LINKED IMMUNOSORBENT ASSAY (ELISA)

EMSA, FP, and ELISA are three rapid and sensitive techniques used to detect protein-protein complexes or protein-nucleic acid interactions. These techniques could be used qualitative but typically are leveraged to quantify the binding affinity between different interactors, allowing to extract dissociation constants (K_ds) between a substrate (S) and a ligand (L). The equilibrium between S and L is described as follows:

$$[S]+[L] \underset{Koff}{\overset{Kon}{\rightleftharpoons}} [SL] \quad ,$$

$$Kd = \frac{Koff}{Kon}$$

where [S] is the molar concentration of S, [L] is the molar concentration of L, [SL] is the molar concentration of the complex, K_{off} is the dissociation rate constant in min^{-1}, and K_{on} is the association rate constant in $min^{-1} * M^{-1}$. *Kd* is expressed in molar (M) and it is inversely related to the affinity of the substrate for the ligand [202,203].

As the name suggests, electrophoretic mobility-shift assay (EMSA) allows the separation of mixtures of proteins, molecules or protein and nucleic acids by electrophoresis under non-denaturing native conditions through polyacrylamide or agarose gels. In this assay, an oligonucleotide, typically fluorescently labelled at either its 3'- or 5'-end, is incubated with increasing concentrations of a protein of interest for a fixed period of time, before running the sample on a non-denaturing gel. Generally, oligonucleotide-protein complexes migrate more slowly, due to the mass increase, than the corresponding free fluorophore-labelled oligonucleotide do, resulting in a shift in gel bands when the protein binds to the oligonucleotide substrate (Figure 2.4A) [(204)]. EMSAs are experimentally easy to perform and quickly allow to obtain information about the relative binding affinity between the protein of interest and tested oligonucleotide sequence. However, the main drawback of EMSAs is the reliability on a stable molecular complex being formed during the electrophoretic run. In fact, some protein-oligonucleotide complexes could dissociate as a consequence of the gel run, causing underestimation of the ability of a ligand to interact with its substrate. Therefore, optimisation of the protocol is essential to correctly estimate a K_d with EMSAs [(204)].

Fluorescence polarisation (FP) can be alternatively used to measure K_d values, without relying on the formation of a stable protein-DNA complex as needed for EMSAs. FP is based on the basic principle that when a small and fluorescently labelled molecule (such as an oligonucleotide) in solution is excited by polarised light, it emits low depolarised light because of its rapid tumbling promoted by its relatively small size, which results in a low FP signal. However, when the fluorophore-labelled oligonucleotides interact with a larger molecule (such as protein) a larger complex form, which rotates more slowly in solution and emits higher polarised light, resulting in a detectable FP signal (Figure 2.4B) [(205)]. This technique allows high throughput screening of substrates and ligands. However, FP might be problematic when measuring low-affinity interactions as it requires very high concentrations of the unlabelled ligand, which could lead to artificial crowding effects [(206)].

As mentioned in Section 1.4.1, ELISA allowed to confirm the binding specificity of BG4 antibody for G4 structures, and is an alternative strategy to assess K_d for nucleic acids/protein interactions [(62)]. This technique is an immunoassay, which relies on the use of a primary antibody (Ab1) that recognises its specific antigen (nucleic acids, proteins or molecules). The antigen is usually biotinylated and attached to a streptavidin-coated surface through strong biotin-streptavidin interaction. After incubation with Ab1, different washing steps are required to remove unbound Ab1. The Ab1 bound to the antigen is then recognised by a

secondary antibody (Ab2), which carries an enzyme-conjugate (e.g. Horseradish Peroxidase (HRP)). After re-washing to remove any unbound Ab2, a substrate (e.g. TMB (3,3',5,5'-Tetramethylbenzidine)) is added for the enzyme and a measurable signal (usually colour change) is produced. An absorbance signal of the chromogen is then measured and used to calculate K_d between the Ab1 and the antigen (Figure 2.4C). Although this technique is generally highly sensitive, one of the main problems is the possibility to detect false positives caused by unspecific binding of Ab1 to the surface or unspecific recognition of Ab2 [207,208]. Each of the assays described in this section has its advantages and limitations, which is why this work has used a combination to thoroughly assess binding affinity between CSB and the panel of synthetic oligonucleotides tested.

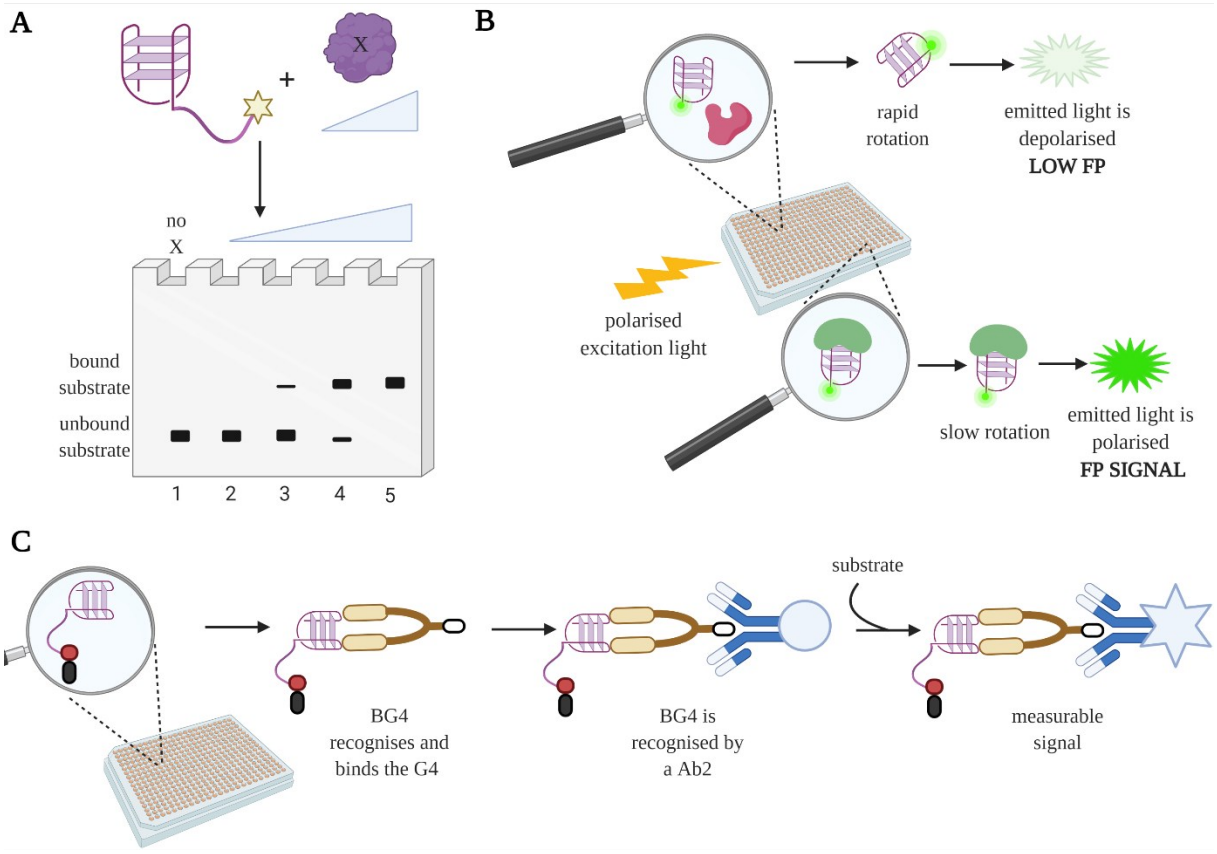

Figure 2.4. Approaches to investigate G4-interactors. (**A**) Schematic representation of EMSA. A fluorophore labelled G4-forming sequence is incubated with increasing concentrations of an examined protein (represented as a purple "X" globular protein). After incubation, the samples are separated on a native gel. In absence of the protein (no X, lane 1) only a bottom band corresponding to the unbound substrate is formed. Molecular interactions between the G4 and X protein are identified as upper bands (molecular shifts) that run slower (lanes 3-5) compared to the unbound substrate (lanes 1 and 2). (**B**) Schematic representation of the FP assay. Fluorophore-labelled G4s are incubated with different proteins (represented in magenta and green). The interaction between the protein and the G4 decreases the rotation of the G4, resulting in increased polarised light (FP signal). Contrarily, the unbound G4 retains its fast rotation in solution, resulting in a depolarised signal (low FP). (**C**) Scheme of a typical BG4-ELISA. Biotinylated (red circle) G4-forming oligonucleotides trapped on the surface of a plate through streptavidin (black bar) interaction. Then, the G4 is bound by the BG4 (represented in brown). BG4 presents a flag-tag (white bar) recognised by a secondary antibody (Ab2; blue). Ab2 is conjugated with an enzyme (light blue sphere), which after the addition of the substrate, emits a measurable signal (light blue star).

FÖRSTER RESONANCE ENERGY TRANSFER (FRET)

Förster resonance energy transfer (FRET) is a physical phenomenon that can be leveraged to assess distance-dependent energy-transfer interaction between a pair of fluorescent molecules, where one act as an energy donor and the other one as an acceptor. When close (few nanometres), the excitation of the donor can result in energy transfer (FRET) to the acceptor, which will, in turn, emit despite not being directly excited. This can be used to study G4-folding, using synthetic sequences containing a donor and an acceptor fluorophore at its 3'- and 5'- ends. When the G4 is fully folded, the two fluorophores are relatively close to each other and they will generate a high FRET signal, resulting from the energy transfer from the excited donor to the acceptor (Figure 2.5A). When the reverse complement of the G4-folding sequence is added, the resulting duplex will form in a time-dependent manner, bringing the donor and acceptor distant in the space (Figure 2.5A), causing the FRET-signal to gradually decrease in a time-dependent fashion [(209)]. Therefore, G4-unfolding can be followed and quantified by measuring the increasing emission of the donor over time, as the FRET is abrogated with the duplex being formed. A plateau of the donor fluorescence emission is indicative of the full resolution of the G4 structure. As an excess of unlabelled reverse complementary strand is also added into the mixture, the unfolded G4 sequence is stabilised by its reverse complementary strand with consequent formation of dsDNA, which contributes to maintaining the plateau signal emitted from the donor constant. Alternatively, a quencher instead of an acceptor could be attached at either the 3'- or 5'- end of the G4 sequence. In this case, when the G4 is fully folded, the fluorescent signal will be quenched, and an increased fluorescence emission will be indicative of G4 resolution as the quencher will be far from the fluorophore in the duplex [(210)]. This same principle could also be exploited to calculate the melting temperatures of G4-structures in FRET-melting assays, which has been extensively used to identify and validate G4-ligands and small-molecule interactors. In summary, the stabilisation induced by these G4-ligands causes an increase in melting temperature (T_m) between the unbound G4 and after interaction of the same G4 with the ligand [(211,212)].

GEL-BASED RESOLVASE ASSAY TO DETERMINE G4 RESOLUTION

Alongside FRET-based techniques, the ability of a protein to resolve G4 structures can also be evaluated using a more conventional gel-based assay. This technique relies on the different gel-mobility between G4s and dsDNA. Both the G4-forming sequence and dsDNA should be fluorescently labelled to follow their migration on a gel and quantify resolvase activity. Typically, a dsDNA control is used as a reference for the G4-resolution and is prepared by mixing equal concentration of the fluorophore-labelled G4-forming sequence and its unlabelled reverse complementary strand. Both the G4-forming sequence alone and the G4-forming sequence mixed with its unlabelled reverse complementary strand are boiled and slowly cooled to room temperature to allow the formation of the G4 and dsDNA, respectively (annealing of the oligonucleotides). The annealed G4 is subsequently incubated with an excess of its reverse complementary DNA in the presence or absence of a fixed concentration of the protein of interest. The reactions are incubated for different timing prior denaturation of the protein-nucleic acid interactions with EDTA and SDS. Then, the samples are separated on a polyacrylamide gel where the dsDNA product migrates slower compared to the folded G4. The relative ratio between these two species (G4s vs dsDNA) can be measured under the different conditions tested to assess the ability of the protein to promote G4-resolution (i.e. increased dsDNA formation) (Figure 2.5B) [168]. Although this is a good method that allows the visualisation of the G4-resolution by looking at dsDNA formation on a gel, the main drawback is the manual time required by the operator to add all the reagents in all the samples and stop the reactions. In fact, very unstable G4s could form dsDNA in seconds after addition of only an excess of the complementary strand, making the assay uninformative for intrinsically fast unfolding G4s. Moreover, the resolution of the G4 substrates is estimated after fixed time points, with a consequent lack of precise real-time information about the G4 resolution. Therefore, a more quantitative real-time resolution using the FRET system is recommended. Indeed, in most of the cases a complete G4 analysis has been carried out using both the FRET-based and gel-based approaches in parallel to overcome the above-mentioned limitations.

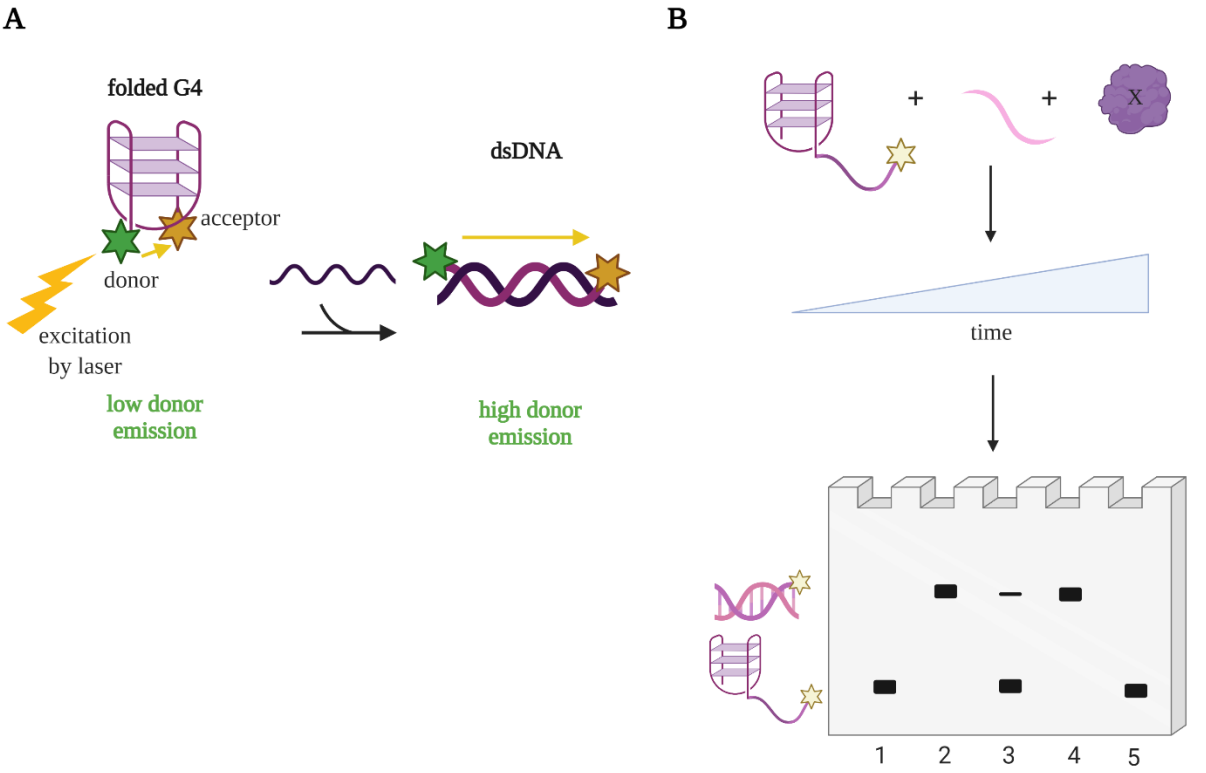

Figure 2.5. Approaches to determine G4-resolution. (**A**) Schematic representation of the FRET principle. When a G4 is folded, the two fluorophores (represented as donor and acceptor (green and yellow stars, respectively)) are close in the space. Excitation of the donor with a laser can result in energy transfer to the acceptor with consequent low emission signal from the donor. Conversely, destabilisation of the G4 structure after incubation with a reverse complementary sequence (in dark purple) causes increased emission of the donor as the two fluorophores are too distant for energy transfer to occur. (**B**) Schematic representation of a gel-based resolvase assay. The fluorophore labelled G4-forming sequence is incubated for increasing timing with its unlabelled reverse complementary sequence and fixed concentrations of an examined protein (represented as a "X" globular protein in purple). The reactions are stopped prior to the separation of the samples on a polyacrylamide gel. Lane 1 represents the pre-annealed G4 control, which generates fast running bands at the bottom of the gel. Lane 2 indicates the pre-annealed dsDNA reference, which runs slower compared to the folded G4 and generates upper bands. Lane 3-4 represents two increasing time-points incubations. In lane 3, only a small portion of the G4 is resolved to dsDNA, while longer incubation allows the protein to fully resolve the G4 which forms dsDNA with its complementary, as shown in lane 4. Lane 5 shows a control sample without incubation with protein X.

Chapter 3

3 Results and Discussion

3.1 CSB-FL and CSB-HD Proteins Production

The main aim of this study was to biochemically investigate the activity of CSB in recognising and resolving G4 structures. To achieve this, the first step was to produce the recombinant protein. CSB full length (CSB-FL) was expressed and purified from insect cells, while its smaller helicase- "like" domain (CSB-HD, see 3.1.4 and 3.1.5 for details) was expressed and purified from bacteria. The production of both CSB-FL and CSB-HD proteins required different several expression and purification steps, which are explained in detail in the following 68aragraphhs.

3.1.1 Production of CSB-FL from Baculovirus/ *Spodoptera frugiperda 9* (*Sf9*)

CSB-FL is a 1,493 amino acids (aa) protein with a molecular weight (MW) of 168.4 kDa, and it is known to exist in a dimeric form in solution [175]. Several approaches were used by P. Selby and A. Sancar to generate the recombinant CSB-FL protein. However, the protein was not appreciably expressed by *Escherichia coli*, and its production from *human* HeLa cells resulted in protein degradation. In contrast, CSB was abundantly produced using baculovirus/*Spodoptera frugiperda 21 (Sf21)* insect cells, which showed particularly high efficiency in producing multi-subunit proteins [176]. In addition, subsequent work demonstrated that CSB can also be efficiently expressed in *Sf9* and *Trichopulsia ni* (*Hi5)* insect cells after infection with the recombinant baculovirus [179]. The main advantages of the baculovirus/insect system include high yields of secreted proteins, improved solubility, and eukaryotic post-translational modifications, which ensures the correct folding of the recombinant proteins.

TRANSPOSITION

The first step to express recombinant proteins from insect cells system was to generate a recombinant baculovirus containing the desired *CSB-FL* gene, which can be used to infect insect cells. CSB-FL coding sequence (NCBI Reference Sequence: NC_000010.11 [213]) was cloned into a pFASTBac donor vector (Invitrogen), where the expression of *CSB-FL* was

controlled by a baculovirus-specific promoter, called P_{PH} (Figure 3.1A). *CSB-FL* was inserted between an N-terminal hemagglutinin antigen (HA) epitope, which allows the detection and the purification of the protein using HA antibody, and a C-terminal six-histidine tag (His_6), which also allows the purification of the protein by affinity chromatography using nickel or cobalt resins that bind histidine [214] (Figure 3.1A). In this work, the construct used for CSB expression (pFASTBac_HA-CSB-His6) was kindly donated by the Scheibye-Knudsen group (University of Copenhagen).

Baculovirus production required a site-direct transposition of the P_{PH}-HA-CSB-His_6 region of the donor vector into a baculovirus shuttle vector, called bacmid. The bacmid contained a recombinant viral DNA from *Autographa californica* multiple nuclear polyhedrosis virus (AcMNPV) and was autonomously replicating in DH10Bac cells (Figure 3.1B) [215]. DH10Bac is an *E. coli* bacterial strain which contains a helper plasmid encoding a transposase. Transposase allowed the site-direct transposition between the P_{PH}-HA-CSB-His_6 region of the pFASTBac donor vector, which was flanked by Tn7 sequences at its 5'- and 3'-ends, and a mini-Att Tn7 site included in the bacmid (Figure 3.1B) [216]. Positive transposition (Figure 3.1C) was confirmed by blue-white screening of the cells. This screening relies on the activity of β-galactosidase, which is codified by a *LacZ* operon expressed within the mini-Att Tn7 site of the bacmid in DH10Bac cells. For the visualisation screening, a chromogenic X-gal substrate was supplemented in the media. If the β-galactosidase is expressed, it will hydrolyse the chromogenic substrate (X-gal) generating a blue insoluble pigment. Thus, non-recombinant cells appeared as blue colonies. Conversely, when the plasmid was correctly inserted within the mini-Att Tn7 site containing the *LacZ* operon, the coding region of *LacZ* was compromised and the β-galactosidase was not expressed, resulting in the formation of white colonies (Figure 3.1D). Two positive white colonies were selected and expanded to extract the required bacmids containing the *CSB-FL* gene (Figure 3.1D). The two bacmids were named CSB (1) and CSB (2).

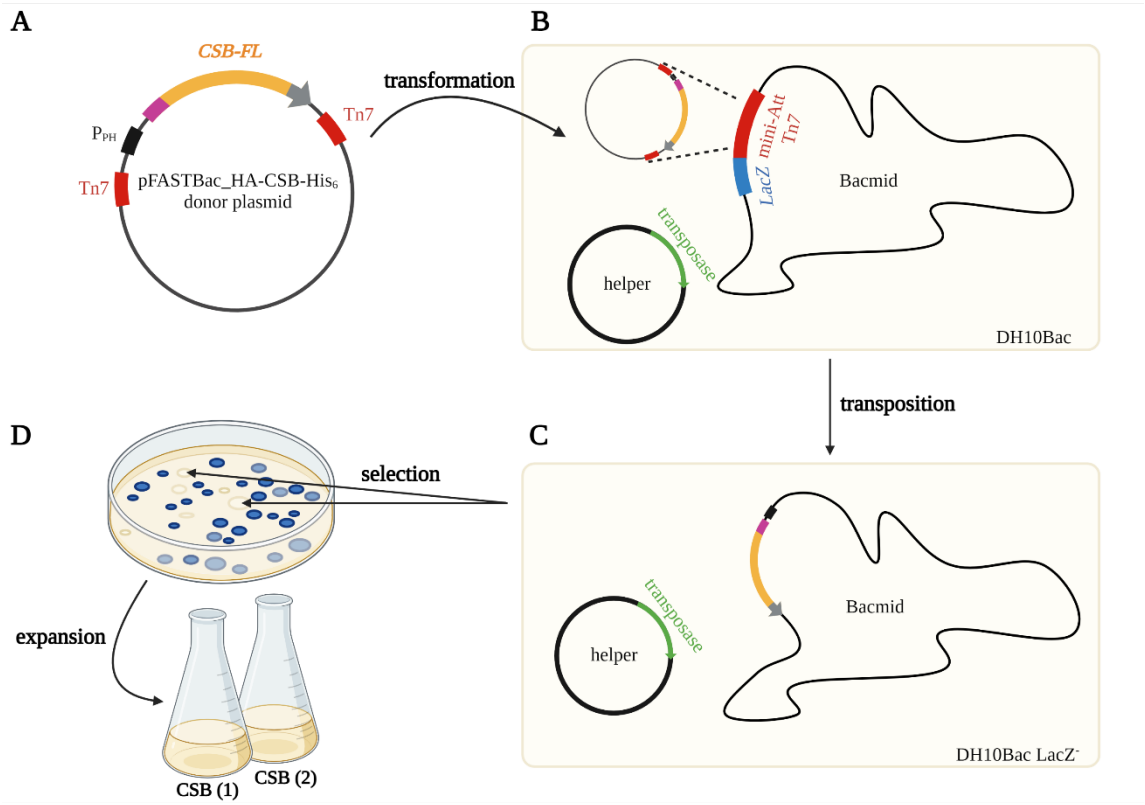

Figure 3.1. Generation of the recombinant bacmid. (**A**) The donor pFASTBac plasmid containing the *CSB-FL* gene was provided by the Scheibye-Knudsen group. This plasmid presented a baculovirus-specific promoter (P_{PH}) and two Tn7 sites at the 5'- and 3'- ends of the HA-CSB-His_6 insertion. The HA-tag, *CSB-FL,* and His_6-tag are reported in magenta, ochre, and grey, respectively. (**B**) The pFASTBac_HA-CSB-His_6 donor plasmid was transformed in DH10Bac cells that contained a bacmid shuttle and helper vectors. The bacmid presented a mini-Att Tn7 region, while the helper plasmid codified for a transposase, which allowed a site-direct transposition of the P_{PH}-HA-CSB-His_6 region within the bacmid. The *LacZ* operon within the mini-Att Tn7 region of the bacmid assured production of a β-galactosidase. (**C**) Correct transposition of the P_{PH}-HA-CSB-His_6 within the bacmid compromised the production of the *LacZ* operon (DH10Bac *LacZ* $^-$). (**D**) In absence of the *LacZ* operon, the β-galactosidase cannot be produced, and the cells are not able to hydrolyse the chromogenic x-gal substrate. Therefore, the positive transposed appeared as white colonies. Two white colonies were expanded to amplify the HA-CSB-His_6 bacmid. The final concentration of the bacmids after extraction was: CSB (1): 4072 ng/µl and CSB (2): 106.65 ng/µl.

BACULOVIRUS COLLECTION AND DOT BLOTS

CSB (1) and CSB (2) bacmids containing *CSB-FL* were separately transfected in *Sf9* insect cells as described in Section 5.1.3. As the transfection advanced, insect cells were subjected to several morphological changes, which reflected the initial viral synthesis (from 0 to 6 hours (h) post-transfection), and release of extracellular viruses (from 18 to 96 h post-transfection). Then, 72 h post transfection, the release of the virus from the cells was confirmed by looking at signs of viral infection, which included increased cellular diameter, enlarged nuclei, vesicular appearance within the cells, and decreased cellular growth compared to the non-transfected controls (Figure 3.2A) [(217)]. The secreted baculoviruses, called P1, were collected by centrifugation and their presence in the cellular supernatant was confirmed by dot-blot on a nitrocellulose membrane using an anti-gp64 antibody that recognises a baculoviral gp64 protein necessary for viral entry by endocytosis [(218,219)]. The gp64 antibody can be in turn recognised by an HRP-enzyme conjugated secondary antibody. Thus, development of the membrane using a high sensitivity enhanced chemiluminescence (ECL) substrate will generate a luminescent signal that can be detected and used to estimate the concentration of the virus in the collected cellular supernatant. The viral concentration could be indeed qualitatively determined by comparison of the signal from the unknown sample to that of a positive control (WDR61).

The dot blot results (Figure 3.2B) obtained from transfection with CSB (1) bacmid showed a very low concentration of secreted virus, named P1-1, suggesting a poor virulence of this virus. Conversely, P1-2 that resulted from transfection with CSB (2) bacmid had a higher signal compared to both P1-1 and the positive control WDR61 P1.

A subsequent P1 amplification was necessary to achieve a sufficient amount of virus to infect high volumes of insect cells. An approximate amount of virus required for P1 amplification could be determined based on the dot-blot results. Usually, 100-300 µl P1 is required to efficiently amplify viruses that are highly concentrated. Therefore, because the luminescence intensity of P1-2 suggested a high concentration of the virus, 100 µl of P1-2 was used to infect 50 ml of *Sf9* cells. After 48 h post-infection, the amplified virus, called P2, was secreted from the cells, collected by centrifugation, and its presence and concentration were determined by dot blot. Unfortunately, the dot blot results obtained for the P2 originated from transfection with 100 µl P1-2 (P2-1) (Figure 3.2B) reported a very low luminescent signal compared to that of the WDR61 P2 control, indicating inefficient virulence.

To improve the concentration of the amplified P2 virus, both longer infection time of 72 h and exposure of *Sf9* to higher amount (200 µl) of either P1-1 or P1-2 were attempted. An additional viral expansion (P3) was also tested by infecting *Sf9* cells with 500 µl of P2-1. The dot-blot test (Figure 3.2C) clearly showed improved P2-2 and P2-3 concentrations after 72 h infection, with higher volumes of either P1-1 or P1-2, respectively, while almost absence of P3 was observed. To understand whether the higher volume or the longer infection time were the limiting factors, *Sf9* cells were treated with either 200 µl or 400 µl of P1-2 for 48 h. Again, the dot-blot (Figure 3.2D) showed lower signals for both P2-4 and P2-5 compared to those of the positive control. However, the infection with 400 µl of P1-2 provided better results than the infection with 200 µl of the same P1-2, indicating that increased P1 volumes helped to obtain a more concentrated P2-5. An additional dot-blot (Figure 3.2E) confirmed a higher concentration of P1-2 compared to P1-1 and displayed a more concentrated P2-6 after infection of *Sf9* cells with 800 µl of P1-2 for 48 h, compared to the previous P2-5 collected after infection with 400 µl P1-2 for the same incubation time. Based on these results, the most concentrated P2-6 was used for the subsequent gene expression.

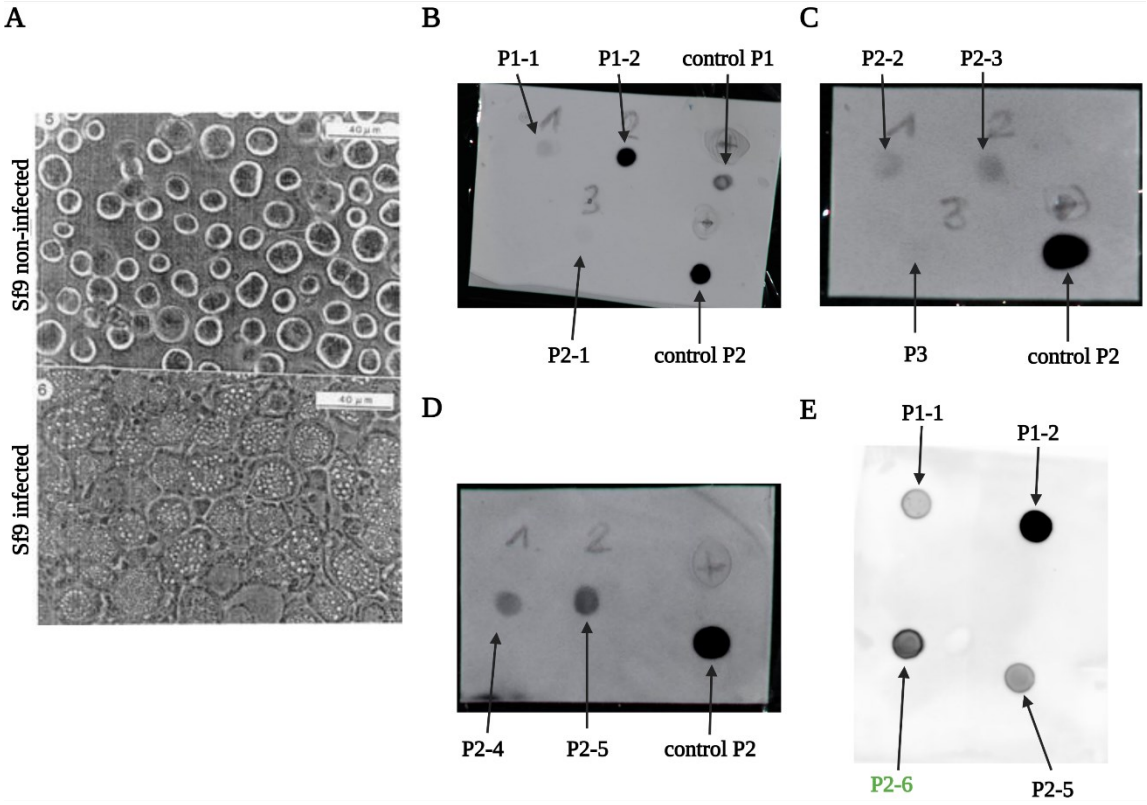

Figure 3.2. Recombinant viruses released from *Sf9* cells, and virulence assessed by dot-blot. (**A**) Photomicrograph showing morphological changes between non-infected *Sf9* cells and *Sf9* cells infected with recombinant viruses (photo from Unger and Peleg [217]). (**B**) Dot-blot results reporting higher concentration of P1-2 compared to P1-1 and the control P1. P2-1 amplified from P1-2 showed low concentration of the virus. (**C**) Dot-blot showing improved concentration of P2-3 virus compared to P2-2. Negligible P3 was observed. (**D**) Dot-blot reporting improved P2-5 concentration compared to P2-4. © Final dot-blot confirming higher concentration of P1-2 compared to that of P1-1, and higher P2-6 concentration compared to that of the previous most concentrated P2-5 tested. P2-6 was used for the subsequent protein expression.

3.1.2 CSB-FL Protein Expression and Purification

CSB-FL SMALL-SCALE PROTEIN PRODUCTION

With the aim to find the best conditions for expression and purification of CSB-FL, a first small-scale expression was conducted. Recombinant baculovirus gene products have been generally observed between 72-96 h post-infection of *Sf9* cells with the amplified P2[219]. Accordingly, 50 ml of *Sf9* cells were infected with 800 µl P2-6 for 72 h.

As described in section 5.2, after 72 h of infection the *Sf9*, cells were resuspended and lysed by sonication. After lysis, the total fraction lysate (TF) was subjected to high-speed centrifugation to sediment all the cellular debris. The supernatant (soluble fraction: SF) was then purified by affinity chromatography on 1 ml Nickel-NTA resin, which allows the purification of recombinant proteins containing a poly-histidine tag (see Section 2.1 for a more detailed overview about his-tag affinity chromatography). Once all the unbound material passed through the column (flow through: FT), the resin was washed with low concentrations of imidazole. Imidazole competes with the his-tagged proteins for binding to the Nickel-NTA resin considering its structural similarity to histidine. Therefore, several washes with low concentration of imidazole allowed to remove the unspecific interactors (wash: W). Finally, the protein was eluted (elution: EL) with high concentration of imidazole. All the fractions were collected and analysed by 8% SDS-PAGE (Figure 3.3A). A band around the expected CSB-FL molecular weight (MW) at 168.4 kDa was visible in the eluted sample and was concentrated using dedicated spin columns. The presence of CSB protein was confirmed by Western Blot (WB) using an anti-CSB antibody that specifically recognises a fragment within human CSB-FL (aa: 300-750). The WB (Figure 3.3B) indicated the presence of the expected CSB-FL band, suggesting a successful expression of the protein. However, the presence of multiple bands at lower MW indicated either unspecific interactions of the antibody to the contaminants present within the EL, or protein degradation.

Overall, the preliminary tests allowed to optimise conditions for CSB-FL expression and achieve purification of the protein. Therefore, large-scale production of the protein was performed by expression and multiple-step purification, and yielded a high amount of protein suitable for biochemical analysis.

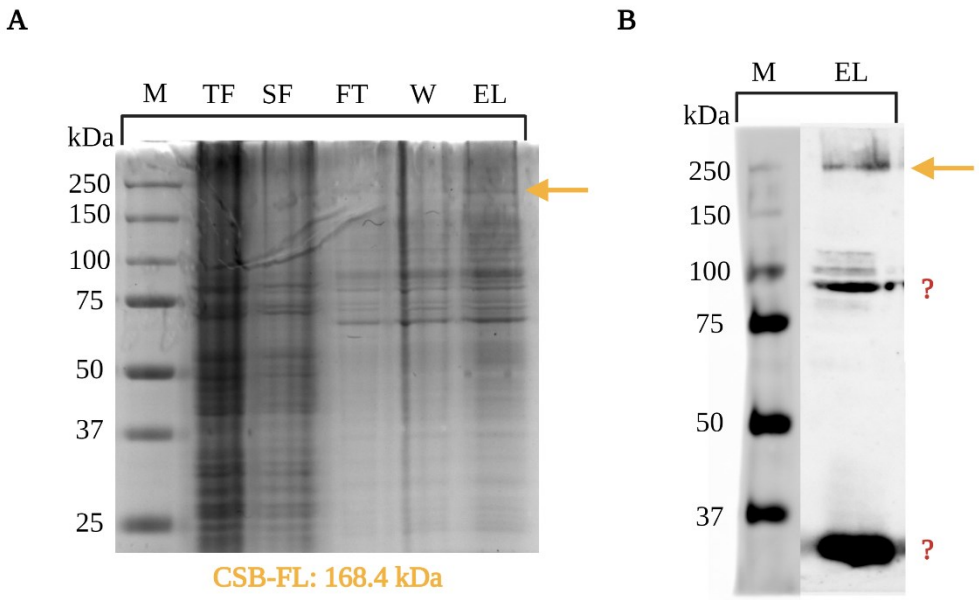

Figure 3.3. CSB-FL small-scale expression and purification. (**A**) 8% SDS-PAGE showing the bands related to the proteins contained in the samples collected after cell lysis and Nickel-NTA purification (TF: total fraction, SF: soluble fraction, FT: flow through, W: wash, EL: elution; M is the protein MW marker which is used to estimate the MW of the proteins run on the gel). The arrow indicates the presence of the band related to at the expected MW of CSB-FL protein (168.4 kDa), with 7 µl of TF, SF and FT and 10 µl of W and EL samples used to load the gel. (**B**) WB image obtained after 2 min development of the membrane. The arrow indicates the CSB-FL band while the question marks (?) indicate the bands related to unknown products (either degradation of the protein or contaminants). M is the protein MW marker.

CSB-FL LARGE-SCALE PROTEIN PRODUCTION

Based on the results obtained from the small scale expression test, four different 500 ml *Sf9* cellular batches were infected with 7.5 ml P2-6 virus each, for a total volume of 2 l. After 72 h of infection, the cells were lysed and centrifuged. The supernatant was then purified by his-tag affinity chromatography using 10 ml Nickel-NTA resin. All the fractions including TF, SF, FT, W (1-5), EL were collected and analysed by 8% SDS-PAGE (Figure 3.4A). Although numerous contaminants were detected, the EL fraction clearly showed the expected CSB-FL band at 168.4 kDa. An additional affinity purification using a heparin column allowed to separate nucleic acid-binding proteins (including CSB) from the contaminants (see Section 2.1 for a more detailed overview about heparin affinity purification). The heparin elution profile (Figure 3.4B) showed a main peak that eluted around 700 mM KCl. The fractions

corresponding to this peak were then analysed by SDS-PAGE, revealing the presence of the band at 168.4 kDa, which identifies CSB-FL (Figure 3.4C). Fractions from E4 to G2 were pooled together and concentrated to ~300 µl.

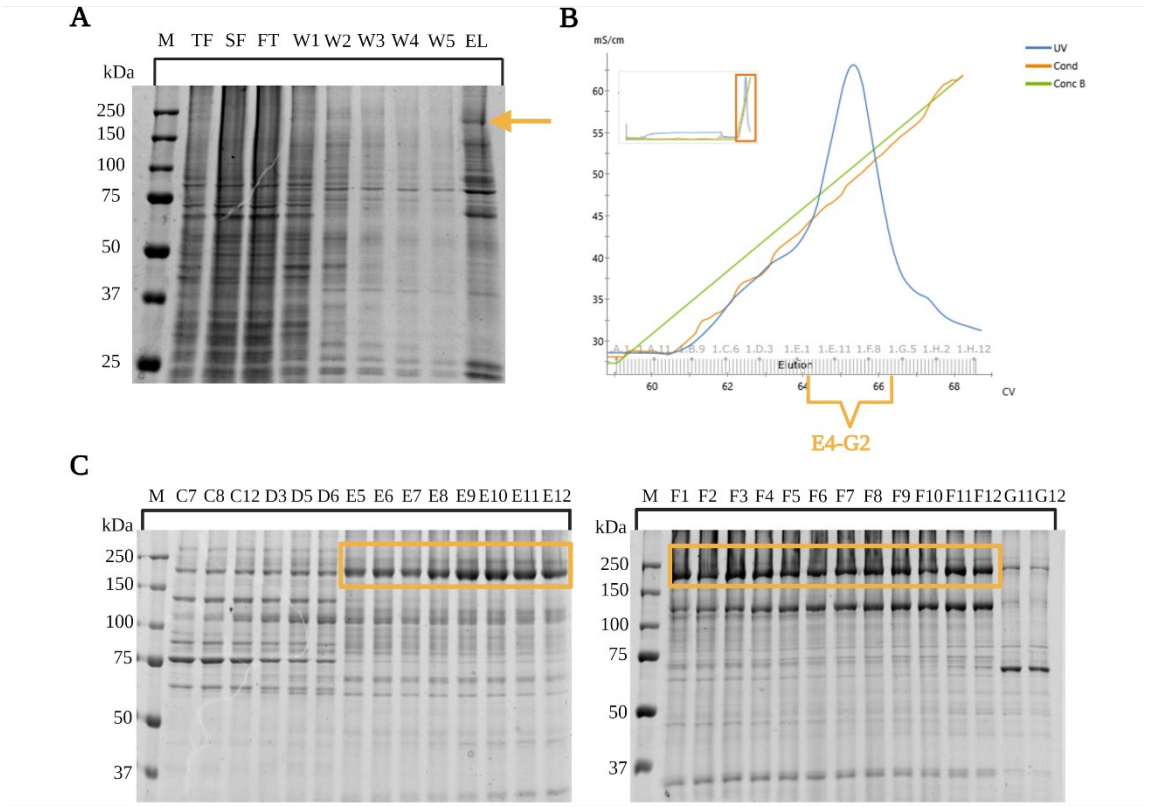

Figure 3.4. CSB-FL large-scale expression and purification. (**A**) SDS-PAGE showing the presence of CSB-FL (ochre arrow) within the EL sample obtained after Nickel-NTA purification of the 2 l cell lysate. The TF, SF, FT wells of the gel were loaded with 7 µl of each sample, while the other wells were loaded with 10 µl of the related sample. (**B**) Elution profile obtained from heparin purification. Fractions from E4 to G2 were pooled together, concentrated and further purified. (**C**) SDS-gel of the peak eluted from the heparin column confirmed the presence of CSB-FL (ochre box). The protein ladder used to estimate the MW of the proteins separated on the gel is indicated as M.

Finally, a size exclusion chromatography (SEC, also called gel filtration) was performed to further improve the purity of CSB-FL. Distinct molecules can be separated depending on their different size when injected into a gel filtration column, where molecules with higher MW elute earlier compared to molecules with lower MW (see Section 2.1 for a more detailed overview about size exclusion chromatography). Because CSB is a big protein of approximately 170 kDa, and it is known to dimerise [175], the protein was expected to elute within the initial volume of elution. To evaluate which SEC column was the most efficient in

separating CSB from the contaminants, the sample that was concentrated after heparin affinity chromatography was divided into two aliquots and purified using two different SEC columns.

The first aliquot (200 µl) was directly purified by SEC using a Superdex 200 (S200) Increase 10/300 column. Comparing the S200 elution profile of CSB-FL (Figure 3.5A) with the S200 10/300 elution profiles of proteins with known MW (calibration curves, Figure 3.5B) allowed the identification of two main peaks eluted around 10.3–11.5 ml, which corresponded to the dimeric CSB-FL (336.8 kDa) and monomeric CSB-FL, respectively. The presence of CSB-FL within these fractions was further confirmed by the denaturing SDS-PAGE that reported strong bands at 168.4 kDa for both peaks (Figure 3.5C). Fractions from B3 to B5 were pooled together, concentrated to 1 mg/ml, and stored at -80 °C.

The rest of the protein obtained from the heparin purification (100 µl) was subsequently purified in a different experiment by a different gel filtration column, called Superose 6 Increase 10/300, which better separates proteins with high MW from smaller ones. The elution profile obtained after Superose 6 purification (Figure 3.5D) was less clear than the S200 one, suggesting some aggregation or degradation of the protein after thawing. Nevertheless, a comparison between the Superose 6 elution profile and Superose 6 Increase 10/300 calibration curves (Figures 3.5E) suggested the presence of CSB-FL within the fractions eluting at approximately 15 ml of elution. The presence of CS-FL within these fractions was further confirmed by SDS-PAGE (Figure 3.5F), therefore, the clearest fractions containing CSB-FL (from C5 to C7) were collected, concentrated to 1 mg/ml and stored at -80 °C. Finally, WB analysis confirmed the presence of CSB-FL in both the samples collected from S200 (EL1) and Superose 6 columns (EL2) (Figure 3.5G).

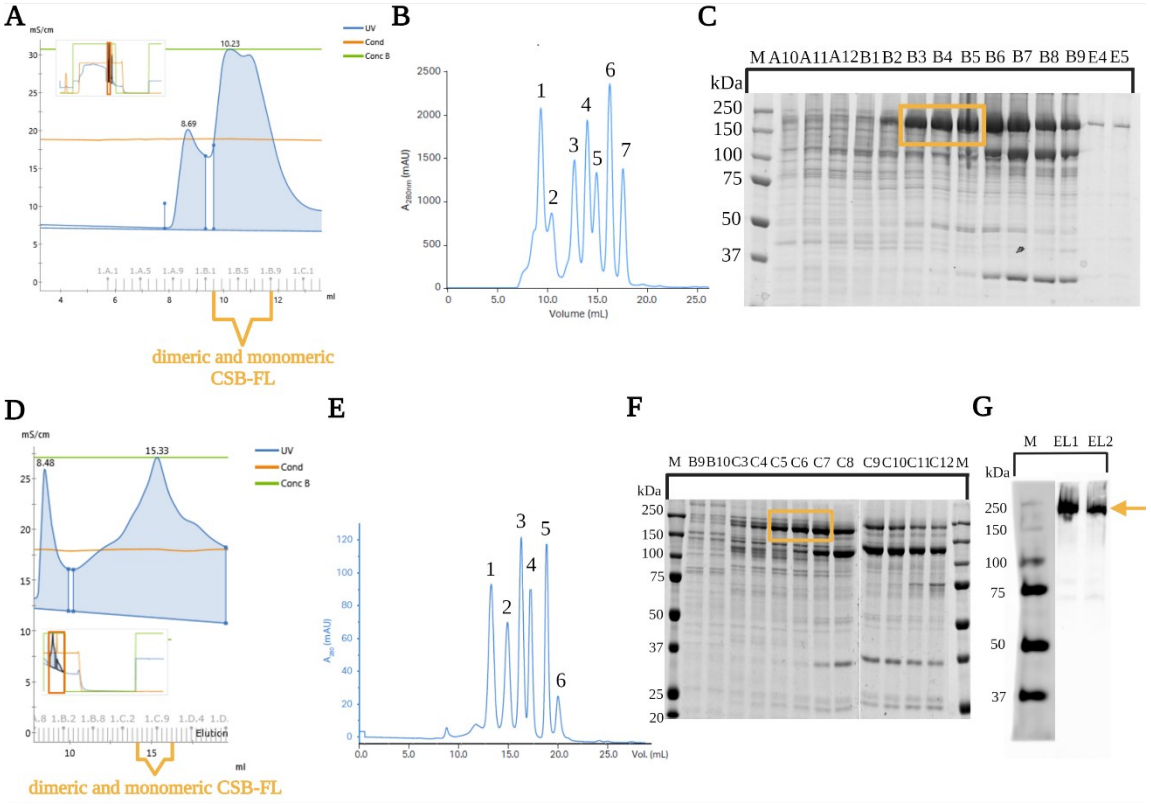

Figure 3.5. CSB-FL size exclusion chromatography. (**A**) Elution profile obtained from CSB-FL purification using Superdex 200 Increase 10/300 column. The two main peaks correspond to the dimeric and monomeric CSB-FL. (**B**) S200 Increase 10/300 calibration curve. 1: Thyroglobulin (MW:669 kDa), 2: Ferritin (MW: 440 kDa), 3: Aldolase (MW: 158 kDa), 4: Conalbumin (MW: 75 kDa), 5: Ovalbumin (MW: 44 kDa), 6: Carbonic anhydrase (MW: 29 kDa), and 7: Ribonuclease A (MW: 13.7 kDa) [220]. (**C**) SDS-PAGE of selected fractions obtained from S200 purification confirmed the presence of CSB-FL within the elution profile. Fractions B3-B5 were pooled together and concentrated. (**D**) Elution profile obtained from CSB-FL purification using Superose 6 Increase 10/300 column. The main peak at approximately 15 ml of elution contains both the dimeric and monomeric CSB-FL. (**E**) Superose 6 Increase 10/300 calibration curve. 1: Thyroglobulin (MW:669 kDa), 2: Ferritin (MW: 440 kDa), 3: Aldolase (MW: 158 kDa), 4: Ovalbumin (MW: 44 kDa), 5: Ribonuclease A (MW: 13.7 kDa), and 6: Aprotinin (MW: 6.5 kDa) [221]. (**F**) SDS-gel of selected fractions obtained from the Superose 6 purification confirmed the presence of CSB-FL within the elution profile. Fractions C5-C7 were pooled together and concentrated. (**G**) WB analysis of the eluted samples from the S200 column (EL1) and Superose 6 column (EL2). The arrow indicates the CSB-FL band. M is the protein MW marker.

3.1.3 Bioinformatic Analysis

Although the baculovirus/insect cell system allowed the production of a good amount of CSB-FL suitable for biochemical studies, the expression of the FL protein is challenging and not compatible with the production of the large amounts of protein required for our biochemical investigations. Hence, we decided to explore whether a smaller CSB domain could be expressed more easily using *E. coli* culture, and still be used to study CSB-G4 interactions. To this end, we performed a multiple sequence alignment [222] between different homologues of CSB-FL (see Section 5.5) to identify the most conserved DNA-binding region containing the ATP binding domain and a signature DEGH box (aa: 646–649), which we assumed being important for CSB translocase activity and G4 resolution [168,223]. The region between 519 and 1,002 aa presented almost 50% identity and very high similarity between all the organisms tested (Figure 3.6). Thus, the *human* nucleotide sequence corresponding to this highly conserved region was synthesised by Genscript service using the *E. coli* codon usage to increase translational efficiency and protein expression [224]. The CSB-HD (519–1,002) sequence codon optimised for bacterial expression was then inserted within a pCS46_His_6-SUMO vector by RF-cloning, as described in Section 5.6. The quality of the cloning was confirmed by Sanger sequencing (Genewiz). The His_6-tag at the N-terminal of SUMO-CSB-HD (519–1,002) allowed its purification by affinity chromatography while the SUMO-tag is reported to increase the solubility of recombinant proteins [225].

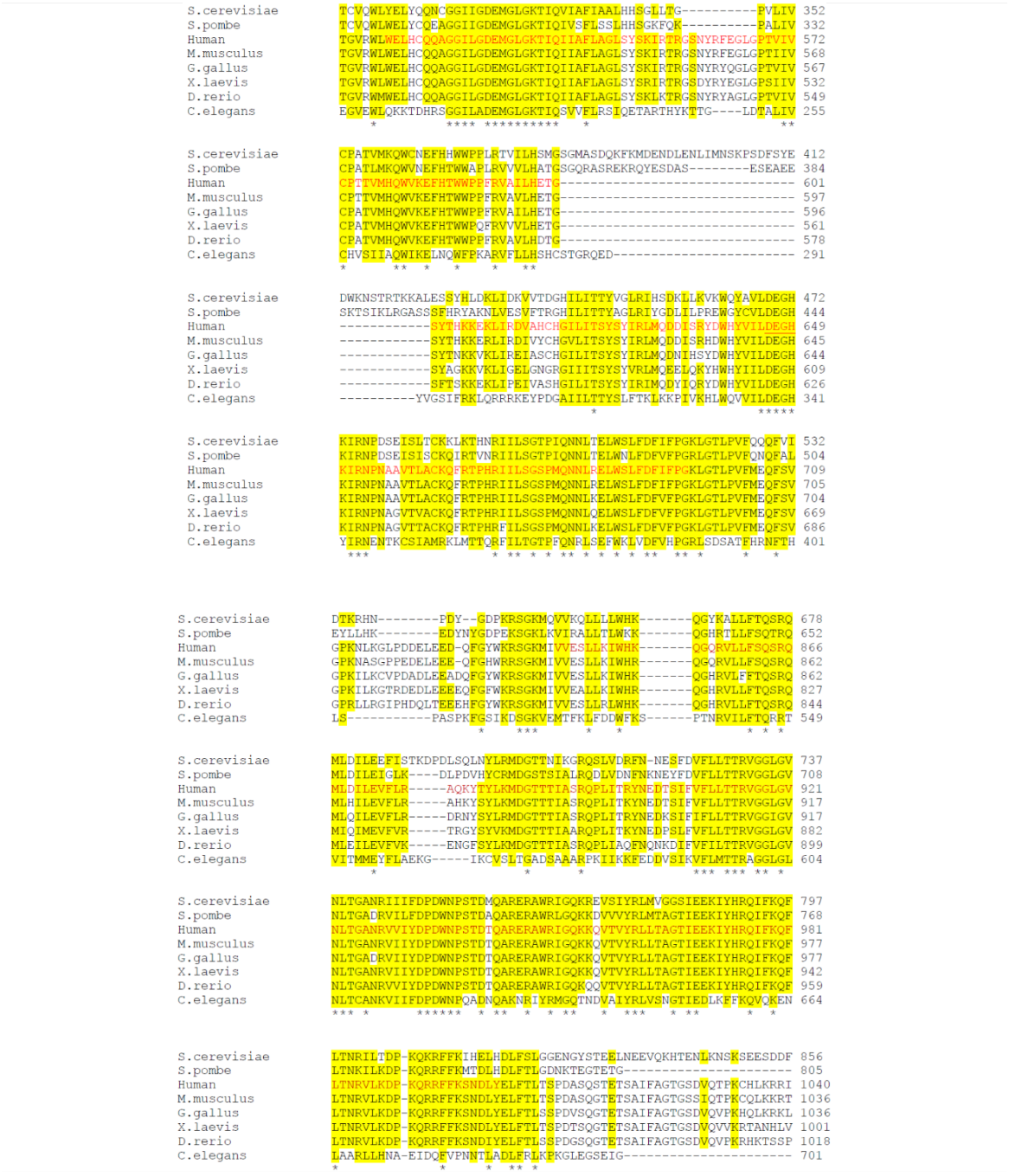

Figure 3.6. Multiple sequence alignment between different homologues of the CSB helicase- "like" domain. The ATP-binding region containing the DEGH box and the C-terminal region of the *human* helicase- "like" domain are reported in red. The residues with either high similarity or 100% identity are highlighted in yellow. Residues with 100% identity are indicated with asterisks (*).

3.1.4 CSB-HD (519–1,002) Expression and Purification

CSB-HD (519–1,002) SMALL-SCALE PROTEIN PRODUCTION

Once the pCS46_His_6-SUMO-CSB-HD (519–1,002) was generated (Figure 3.7A) and sequenced, preliminary small-scale expression and purification tests were performed to confirm the potential to express and purify the protein in bacteria using this construct. The expression of CSB-HD (519–1,002) was tested in 250 ml BL21(DE3) *E. coli* cells grown in a nutrient-enriched Terrific Broth (TB) media. The induction of the protein expression was carried out overnight (O/N) with 0.1 mM IPTG at 18 °C. A sample of cells before the induction was collected and lysed for the subsequent SDS-PAGE electrophoresis. After incubation, the inducted cells were lysed by sonication (TF) and centrifugated to clarify the lysate from the cellular debris (SF). The supernatant obtained after centrifugation was then purified by his-tag affinity purification using Nickel-NTA resin. All the fractions, including the W and EL, were collected and run on a 10% SDS-PAGE (Figure 3.7B). The presence of a strong band around 67 kDa (His_6-SUMO-CSB-HD (519–1,002) MW: 66.6 kDa) in the TF suggested that CSB-HD (519–1,002) was expressed by the cells. In fact, this band was absent in the pre-induced sample, further confirming that it corresponded to CSB-HD expression upon IPTG induction.

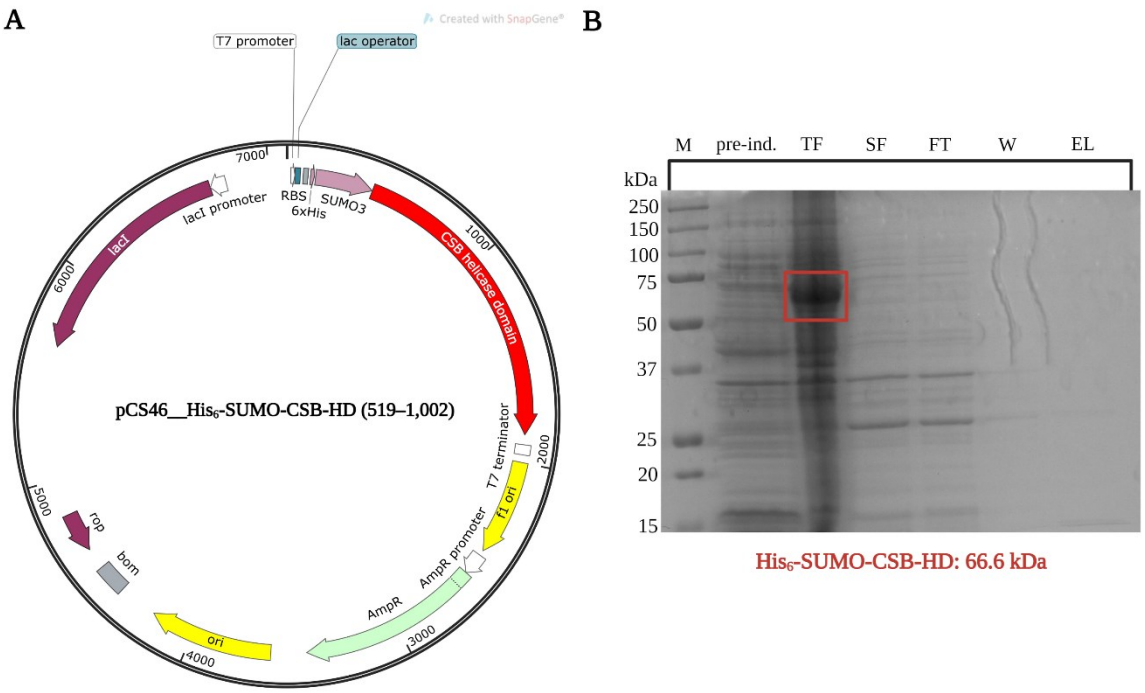

Figure 3.7. CSB-HD (519–1,002) expression. (**A**) Vector map showing the CSB-HD (519–1,002) sequence (in red) inserted at the C-terminal of His_6-SUMO tag (in purple). Vector map created with SnapGene. (**B**) SDS-gel of the samples obtained from small scale expression of CSB-HD (519–1,002). M, protein ladder; pre-ind, sample collected before IPTG induction. The His_6-SUMO tagged CSB-HD band around 67 kDa is marked with the red box. To load the gel, 5 µl of pre-ind., 7 µl of TF, SF and FT and 10 µl of W and EL samples were used.

However, lack of a CSB-HD band in the SF and EL samples indicated that the protein was not soluble in the tested conditions and further optimisations were required, including optimised IPTG concentration, expression temperatures, and buffers composition. A summary of all the tested conditions is reported in Table 3.1A and 3.1B. Among all, the best conditions found to obtain soluble protein were: 0.05 mM IPTG at 36 °C, 4 h in nutrient-rich media and cell lysis buffer containing modest salt concentration (300 mM KCl) supplemented with 10% glycerol.

A

	50mM Hepes pH 7.5, 300mM NaCl, 10mM imidazole, 5% glycerol, 5mM TCEP	50mM Hepes pH 8.2, 300mM KCl, 10mM imidazole, 10% glycerol, 1mM TCEP								
	0.1mM IPTG	0.01mM IPTG			0.05mM IPTG			0.1mM IPTG		
	TB	LB	TB/2xYT		LB	TB/2xYT		LB	TB/2xYT	
	18 °C, 16h	36 °C, 4h	15 °C, O/N	36 °C, 4h	36 °C, 4h	15 °C, O/N	36 °C, 4h	36 °C, 4h	15 °C, O/N	36 °C, 4h
BL21 (DE3)	X	X	0	X	X	X	XX	X	X	XX

B

	25mM Tris pH 9.0, 300mM KCl, 10mM imidazole, 10% glycerol, 1mM TCEP								
	0.01mM IPTG			0.05mM IPTG			0.1mM IPTG		
	LB	TB/2xYT		LB	TB/2xYT		LB	TB/2xYT	
	36 °C, 4h	15 °C, O/N	36 °C, 4h	36 °C, 4h	15 °C, O/N	36 °C, 4h	36 °C, 4h	15 °C, O/N	36 °C, 4h
BL21 (DE3)	X	0	X	X	0	XX	X	X	XX

Table 3.1 (A and B). List of the small-scale expression conditions used for his_6_SUMO_CSB-HD (519–1,002) expression. The best conditions chosen for the expression and lysis of the protein are highlighted in green. 0: no expression of the protein; X: the protein is expressed but it is not soluble; XX: the protein is expressed, and it is soluble.

The solubility of the protein was further improved with the addition of a mild non-ionic detergent, Triton-X (0.1%) (Figure 3.8A). CSB-HD (519–1,002) was then purified by his-tag affinity chromatography using Nickel-NTA resin (Figure 3.8B). However, the presence of multiple bands in the elution around 67 kDa made difficult to unequivocally identify CSB-HD. Therefore, a subsequent cut of the His_6-SUMO tag (MW: 11.3 kDa) using a SUMO protease was necessary to further characterise this protein. In fact, a removal of the His_6-SUMO tag will cause a decreased MW in the untagged CSB-HD (519–1,002) (MW: 56.1 kDa), which could be easily identified by SDS-PAGE.

CLEAVAGE OF SUMO-TAG

The CSB-HD (519–1,002) eluate obtained after affinity chromatography was diluted to decrease the imidazole concentration prior to cleavage of the his_6-SUMO tag with a SUMO protease. The SUMO protease hydrolyses the peptide bond at a Gly-Gly bond in the C-terminal of the SUMO tag, releasing the CSB-HD protein by SUMO domain and the His_6 tag. The cut CSB-HD (519–1,002) protein (MW: 56.1 kDa) could be isolated by his-tag affinity purification and was expected in the FT (cut-FT, Figure 3.8C). Surprisingly, the SUMO protease was not able to completely cut the SUMO domain, as the 67 kDa band corresponding to the SUMO-CSB-HD (519–1,002) was still observed in the elution obtained from purification of the cut sample (cut-EL) (red arrow, Figure 3.8C). However, both the pre-cut sample (EL conc., Figure 3.8C) and cut-FT showed a band around the expected untagged CSB-HD (519–1,002) MW (white arrow, Figure 3.8C) which was absent in the cut-EL, suggesting either cleavage of the tag by SUMO protease or spontaneous instability of the protein. In support of this, both the pre-cut (EL conc., Figure 3.8C) and cut-EL samples presented multiple bands below 25-20 kDa that were not present in the cut-FT, which could be ascribed to protein residues that still carried the His_6-SUMO tag and were originated from spontaneous instability in the proximity of the tags (Figure 3.8C).

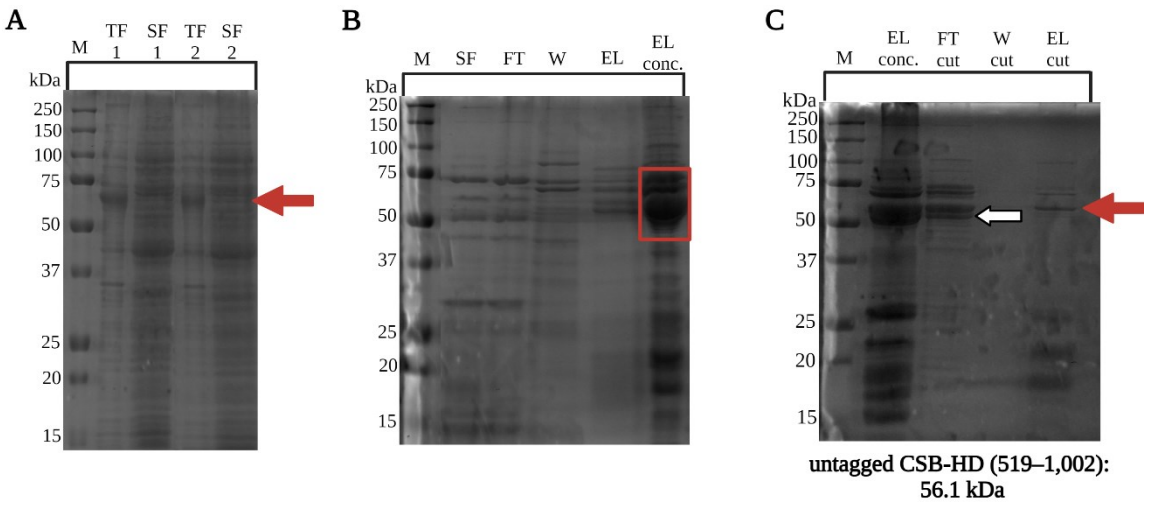

Figure 3.8. Cut of the His_6-SUMO tag revealing instability of the CSB-HD (519–1,002) construct. (**A**) SDS-gel showing improved expression and solubility of CSB-HD using lysis buffer containing 0.1% Triton-X. M indicates the protein MW marker. TF1: total fraction using lysis buffer without 0.1% Triton-X; SF1: soluble fraction using lysis buffer without 0.1% Triton-X; TF2: total fraction using lysis buffer with 0.1% Triton-X; SF2: soluble fraction using lysis buffer with 0.1% Triton-X. The red arrow indicates the expected tagged CSB-HD protein (**B**) Affinity purification using Nickel-NTA resin showed the presence of multiple bands around the expected CSB-HD (519–1,002) MW in the concentrated eluted sample (EL conc.; red box). SF, soluble fraction; FT, flow through; W, wash; and EL indicates the elution sample. (**C**) SDS-PAGE obtained after cleavage of the His_6-SUMO tag. The white arrow in the FT cut sample indicates the untagged CSB-HD (519–1,002) band, while the red arrow indicates either the tagged CSB-HD protein or a tagged degradation product present in the cut EL.

3.1.5 Secondary Structure Prediction and Generation of CSB-HD (498–1,002) Construct

To understand whether the problematic expression and low stability of CSB-HD (519–1,002) were caused by truncation of essential structural features, such as α-helices, the secondary structure of CSB-FL was predicted using PSIPRED [226]. The structural prediction showed truncation of an α-helices at the N-terminal of the original CSB-HD (519–1,002) construct designed (Figure 3.9A), which could potentially justify the low stability of the protein. Therefore, an additional 21-aa sequence was inserted by RF-cloning at the N-terminal of the His_6-SUMO-CSB-HD (519–1,002) construct (Section 5.6) with the aim to reconstitute the truncated α-helices and increase the stability of the final protein (Figure 3.9A). This new

CSB-HD (498–1,002) construct was cloned as previously described (Figure 3.9B) and its identity was confirmed by Sanger sequencing (Genewiz).

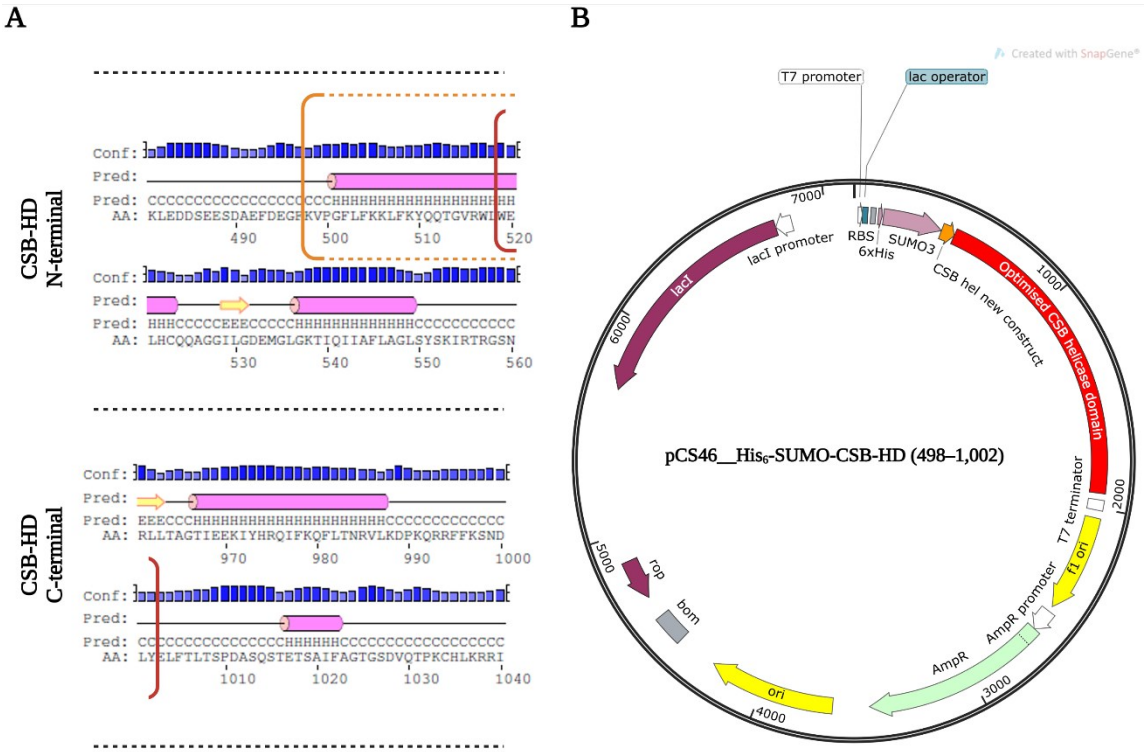

Figure 3.9. Secondary structure prediction showing α-helix truncation at the N-terminal of CSB-HD (519–1,002) construct. (**A**) Schematic representation of the N- and C-terminal regions of CSB-HD. Red brackets indicate the start aa and the end aa of CSB-HD (519–1,002), while the orange bracket indicates the start aa of the new CSB-HD (498–1,002) construct. In the predicted structure, the α-helices are represented as pink cylinders while β-sheets are represented as yellow arrows. (**B**) Vector map showing the new CSB-HD (498–1,002) construct, where a 21 aa sequence (orange) was inserted between the at the His_6-SUMO tag (purple) and the N-terminal of CSB-HD (519–1,002) (red). Vector map created with SnapGene.

3.1.6 CSB-HD (498–1,002) Expression and Purification

The novel His_6-SUMO-CSB-HD (498–1,002) was expressed from BL21(DE3) cells using 0.05 mM IPTG and purified following the same protocol described for the previous CSB-HD (519–1,002) construct. The first his-tag affinity purification using the Nickel-NTA column showed the presence of few bands close to the expected MW of the tagged-CSB-HD (498–

1,002) protein (MW:70 kDa) (Figure 3.10A). However, multiple contaminants were still present in the EL and a further affinity purification using heparin column was necessary to isolate the expected CSB-HD band. The elution profile obtained from heparin purification presented two main peaks eluting at 700–750 mM KCl (Figure 3.10B). Gel visualisation of these peaks clearly revealed the presence of a ~70 kDa band within the first peak, which is likely indicating the presence of tagged-CSB-HD (498–1,002) (Figure 3.10C). Thus, the heparin fractions between B7-B12 were combined and concentrated to the final concentration of 1 mg/ml.

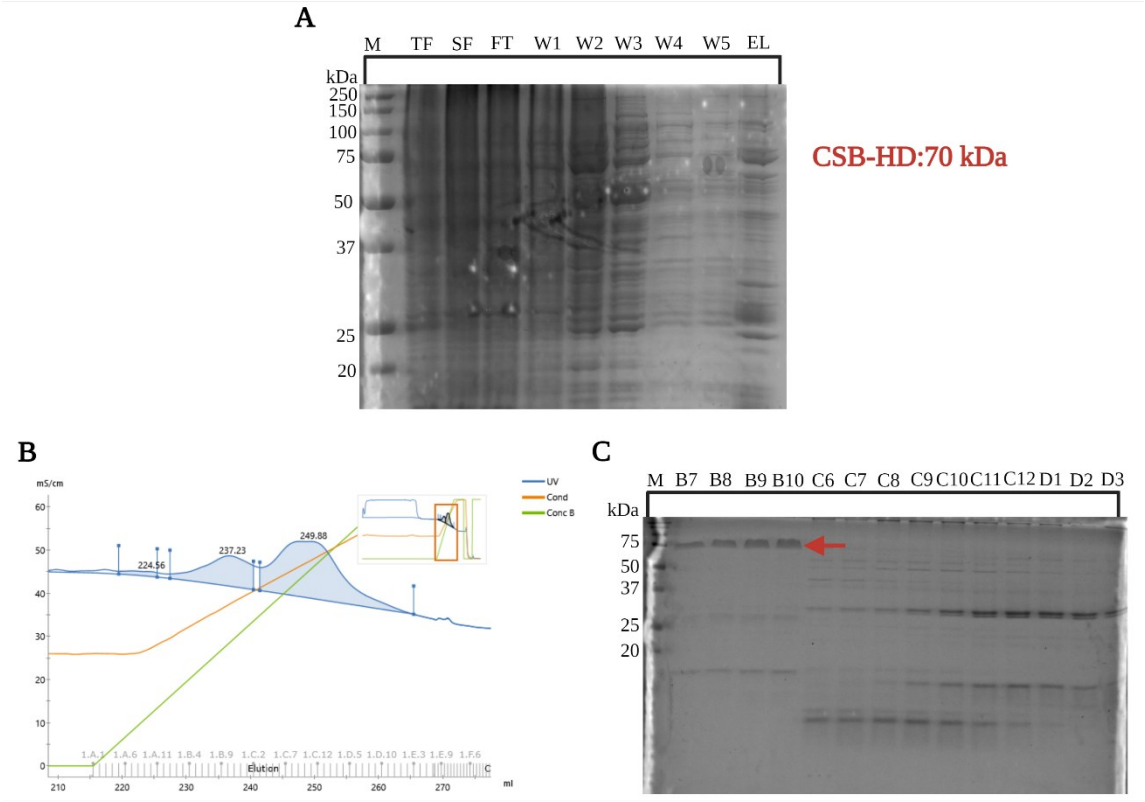

Figure 3.10. Expression and purification of CSB-HD (498–1,002) construct. (**A**) 12%-SDS gel showing the bands related to the proteins contained in the samples collected after cell lysis and Nickel-NTA affinity purification (TF: total fraction, SF: soluble fraction, FT: flow through, W1-5: wash, EL: elution; M is the protein marker used to estimate the MW of the proteins separated on the gel). To load the gel, 10 µl of each sample was used. (**B**) Elution profile obtained from heparin affinity purification of CSB-HD (498–1,002). (**C**) 15% SDS-PAGE of selected fractions obtained from the heparin purification. The arrow indicates the presence of the related band at the expected MW of CSB-HD (498–1,002) protein (70 kDa).

3.2 Cockayne Syndrome B (CSB) as a Selective Intermolecular rDNA G4 Binder and Resolvase

Following their expression and purification, the ability of both CSB-FL and CSB-HD to interact and resolve G4 structures was investigated. In particular, the proteins were biochemically and biophysically tested using a vast panel of nucleic acid substrates, including different rDNA G4-forming sequences, non-rDNA G4s, and non-G4 forming sequences. The different analyses performed, and the results obtained are reported in the following sections.

3.2.1 Gel-based Resolvase Assays Confirmed that CSB-FL and CSB-HD Can Partially Resolve a rDNA G4

Considering the recent results showing the ability of CSB to partially resolve G4-structures formed by a rDNA sequence[168], both CSB-FL and CSB-HD G4-resolvase activities were initially tested using the same rDNA-1 G4-forming sequence used by M. Scheibye-Knudsen and colleagues in their seminal paper [(168)] (more details on the gel-based resolvase assay technique are reported in Section 2.2).

The rDNA-1 G4-forming sequence used for this assay presented both a 5'-Cyanine-5 (Cy5) fluorophore, which allows the visualisation of the oligonucleotide on a gel, and a 20 base pairs (bp) tail at its 5'-end, for consistency with the published substrate (see Table 5.7 for the oligonucleotide sequence). Briefly, the 5'-tailed rDNA-1 substrate was pre-annealed in KCl buffer and incubated with a 5-fold excess of its reverse complementary sequence. Subsequently, a fixed concentration (10 nM) of either the CSB-FL or the CSB-HD were added into the mixture and incubated for increasing times. After incubation, the products were separated on a polyacrylamide gel. On the gel, the formation of bands corresponding to dsDNA that run slower compared to the G4 structure alone (G4 lane, Figure 3.11A and B), was indicative of the resolution of the substrate.

The gels revealed increased formation of the dsDNA bands after incubation with either CSB-FL or CSB-HD (from 0.5 min to 40 min incubation, Figure 3.11A and B) compared to that of the control without the protein (0.5 min and 40 min in absence of CSB-FL/-HD, Figure 3.11A and B). The increased dsDNA formation upon incubation with either one of the two proteins

indicated resolution of the 5'-tailed rDNA-1 G4 structure which, once linearised, could pair with its reverse complementary strand, confirming that both CSB-FL and CSB-HD were able to partially promote rDNA-1 G4 resolution in absence of ATP, as previously reported and observed in the gels in Figure 3.11A and B [197].

Consistent with previous reports [168], most of the G4-related band was still strongly visible in the gel after 40 min incubation with the protein, indicating that most of the substrate was still present in a folded G4 conformation (Figure 3.11A and B). Quantification of the ratio between the intensity of the newly formed dsDNA bands and the fast-running G4 bands was used to assess the percentage of dsDNA formation in the presence or absence of the proteins. The results showed that CSB-FL can significantly increase the dsDNA formation of about 16% when incubated with 5'-tailed rDNA-1 in the presence of the reverse complementary strand compared to the control in absence of CSB-FL (Figure 3.11C and Table 3.2A). A statistically significant difference between the percentage of dsDNA formation in the presence or absence of the protein was calculated using a two-tailed Student's *t*-test, providing a *p*-value of p=0.02. Similarly, incubation with CSB-HD (498–1,002) promoted resolution of the 5'-tailed rDNA-1 G4 (Figure 3.11B) with increased dsDNA formation of approximately 15% compared to the control without CSB-HD, which is comparable to the 16% increase of dsDNA formation observed for the full-length protein. For CSB-HD the *p*-value obtained after the two-tailed Student's *t*-test was p=0.0378 (Figure 3.11D and Table 3.2B) [197].

Overall, the gel-based resolvase results recapitulated the reported ability of CSB to partially resolve G4 structures formed within the 5'-tailed rDNA-1 substrate, confirming that the produced proteins were functional and could be used for the next investigations. Moreover, the G4-resolvase activity of CSB-HD indicated that the helicase- "like" domain of CSB could resolve the 5'-tailed rDNA-1 G4 in a comparable fashion to that of the full-length protein under the tested experimental conditions.

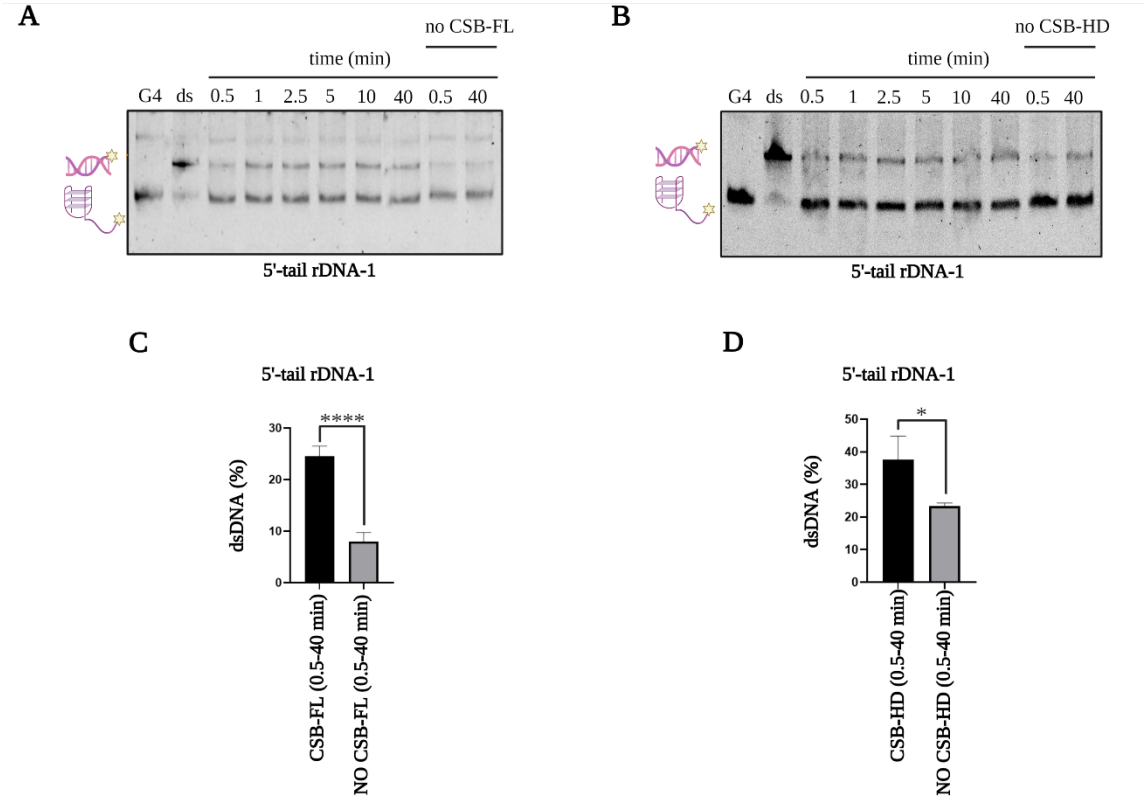

Figure 3.11. CSB-FL and CSB-HD display modest rDNA-1 resolvase activity. (**A**) Gel based-resolvase assay in presence or absence of CSB-FL. G4 indicates the pre-annealed G4-control, ds is referred to the pre-annealed dsDNA control using 1:1 ratio between the G4-forming substrate and its complementary sequence. CSB-FL was incubated with 5'-tailed rDNA-1 and 5-fold excess of its reverse complementary sequence for increasing time (from 0.5 min to 40 min). The last two lanes are the controls without CSB-FL. (**B**) Gel based-resolvase assay using 5'-tailed rDNA-1 in presence or absence of CSB-HD (498–1,002). The fluorophore-labelled G4-forming sequence is represented as a purple G4 that runs faster compared to the dsDNA which runs slower on the gel, and it is represented as a pink/purple double helix. (**C**) Column graph of quantified gel-based resolvase assays using 5'-tailed rDNA-1 with or without CSB-FL. (**D**) Column graph of quantified gel-based resolvase assays using 5'-tailed rDNA-1 with or without CSB-HD. All quantified gel-based resolvase assays were based on the average of three independent experiments. Significance was calculated based on a two-tailed Student's *t*-test. Asterisks indicate statistical significance at 95% CI between the data with ****p <0.0001 and *p < 0.05. The results presented in this figure have been taken from Liano et al. [197].

3.2.2 CSB Can Partially Resolve Different rDNA G4s But Requires a 5'- end Tail

Once the activity of the produced proteins was assessed, a deeper characterisation of the resolvase activity of CSB towards G4s was necessary to better understand the mechanism by which this protein processes G4 structures.

To further investigate the ability of CSB to resolve different rDNA G4s, a novel rDNA-2 G4-forming sequence [195] was tested [197]. For consistency with the previous rDNA-1 used, the fluorescently-labelled rDNA-2 presented the same 20-bp tail at its 5'-end (see Table 5.7 for the oligonucleotide sequence used) [197]. As expected, gel-based helicase assays revealed partial resolution of the 5'-tailed rDNA-2 G4 upon incubation with CSB-FL and increased dsDNA formation of approximately 15% in the presence of the protein (Figure 3.12A and B and Table 3.2A) [197]. The similar results obtained with two different rDNA G4s confirmed that CSB can only partially resolve rDNA G4s.

Next, the need for the 5'- tail for G4-resolution was evaluated by testing an untailed rDNA-1 substrate (see Table 5.7 for the oligonucleotide sequence used). Interestingly, CSB-FL was not able to resolve the untailed rDNA-1 G4 as no difference in dsDNA production was observed by gel after incubation of the untailed rDNA-1 substrate with its reverse complementary strand and CSB-FL (Figure 3.12C). In agreement with this result, quantification of the dsDNA formation did not show any significant increase in dsDNA formation in presence of CSB-FL compared to that of the control in absence of the protein (Figure 3.12D and Table 3.3). The inability of CSB-FL to resolve the untailed rDNA-1 substrate suggested that, like other helicases reported in literature, CSB requires a single-stranded overhang tail to initiate the structural modification of the G4 substrate [197,227].

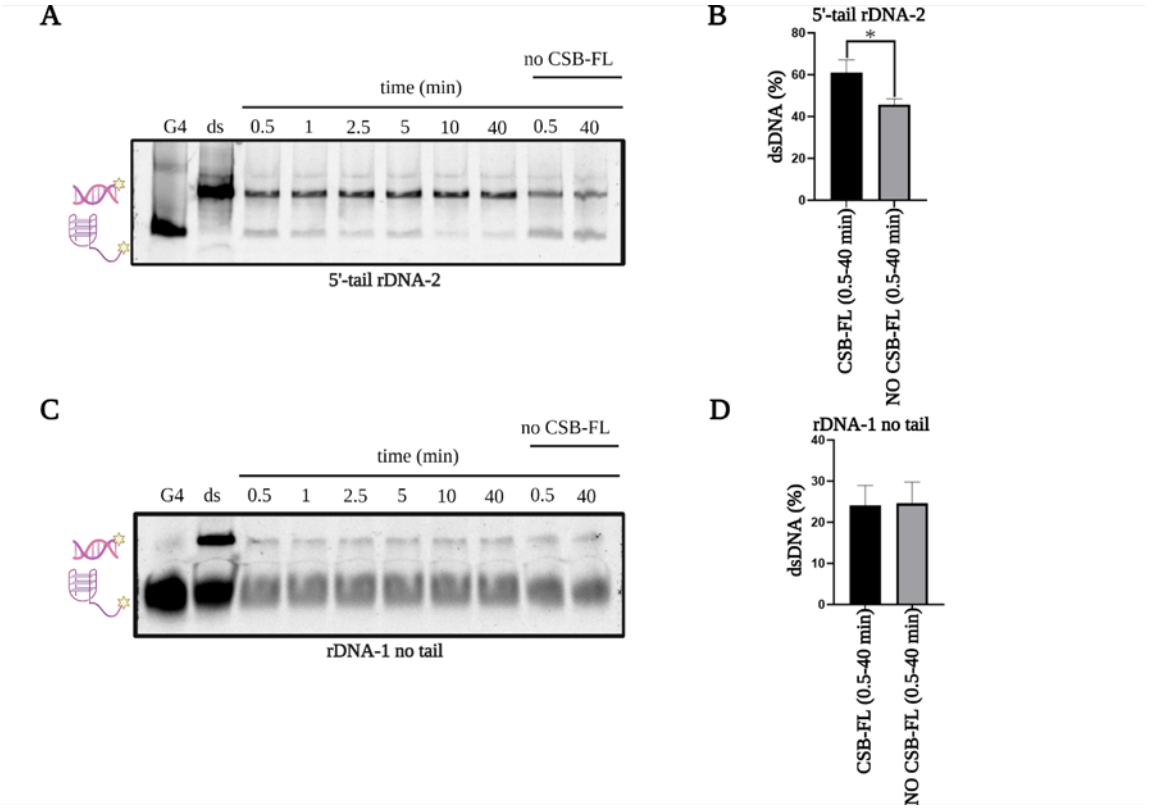

Figure 3.12. An overhang tail is essential for the rDNA G4 resolvase activity of CSB. (**A**) Gel based-resolvase assay using 5'-tailed DNA-2 substrate in presence or absence of CSB-FL. G4 indicates the pre-annealed G4-control, ds is referred to the pre-annealed dsDNA control. CSB-FL was incubated with 5'-tailed rDNA-2 and its reverse complementary for increasing time (from 0.5 min to 40 min). The last two lanes are the controls without CSB-FL. The fluorophore-labelled G4-forming sequence is represented as a purple G4. dsDNA is represented as pink/purple double helix. (**B**) Column graph of quantified gel-based resolvase assays using 5'-tailed rDNA-2 with or without CSB-FL. The results are expressed as percentage of dsDNA formation. All quantified gel-based resolvase assays were based on the average of three independent experiments. Significance was calculated based on two-tailed Student's *t*-test. Asterisks indicate statistical significance at 95% CI between the data with $*p < 0.05$. (**C**) Gel based-resolvase assay using untailed DNA-1 substrate in the presence or absence of CSB-FL. (**D**) Column graph of quantified gel-based resolvase assays using untailed rDNA-1 with or without CSB-FL. This figure has been redrawn from Liano et al.[197].

A

G4 sequence	CSB-FL (0.5-40 min)	NO CSB-FL (0.5-40 min)	double strand (%)
rDNA -1 5'-tail	24.6****	8****	
rDNA-2 5'-tail	61.1*	45.7*	

B

G4 sequence	CSB-HD (0.5-40 min)	NO CSB-FL (0.5-40 min)	double strand (%)
rDNA-1 5'-tail	37.6*	23.4*	

Table 3.2. Quantification of the resolvase assay gels testing different rDNA-1 and rDNA-2 G4-forming sequences. The gel quantification is represented as the percentage of the ratio between the intensities of dsDNA formation and the G4 bands at 0.5 min and 40 min in the presence or absence of either CSB-FL (**A**) or CSB-HD (**B**). Significance was calculated based on two-tailed Student's *t*-test. Asterisks indicate statistical significance at 95% CI between the data with ****p <0.0001 and *p < 0.05. Tables adapted from Liano et al.[197].

3.2.3 CSB Does Not Resolve Other G4s

Gel-based resolvase assay confirmed the ability of CSB to partially unwind different rDNA G4s in presence of a 5'-tail. To further investigate the resolution of different G4 structures, we studied whether the G4-resolvase activity of CSB was limited to rDNA substrates or if partial resolution could also be observed for non-rDNA G4s.

Therefore, a panel of different fluorophore-labelled G4s was tested with the same gel-based resolvase assay employed for rDNA. The panel of G4s used included c-KIT1, hTELO, HRAS, and c-MYC, all functionalised with the same 20 bp 5'-tail used before (see Table 5.7 for the oligonucleotide sequences used). Interestingly, negligible G4 resolution was observed for all the tested 5'-tailed G4s after incubation with CSB-FL compared to that of the controls without the protein (Figure 3.13 A-H) suggesting that the resolvase activity of CSB was limited to tailed rDNA G4s under the tested experimental conditions[197]. Unlike the others, the G-runs are longer in rDNA sequences which might be the reason for their selective resolution promoted by CSB.

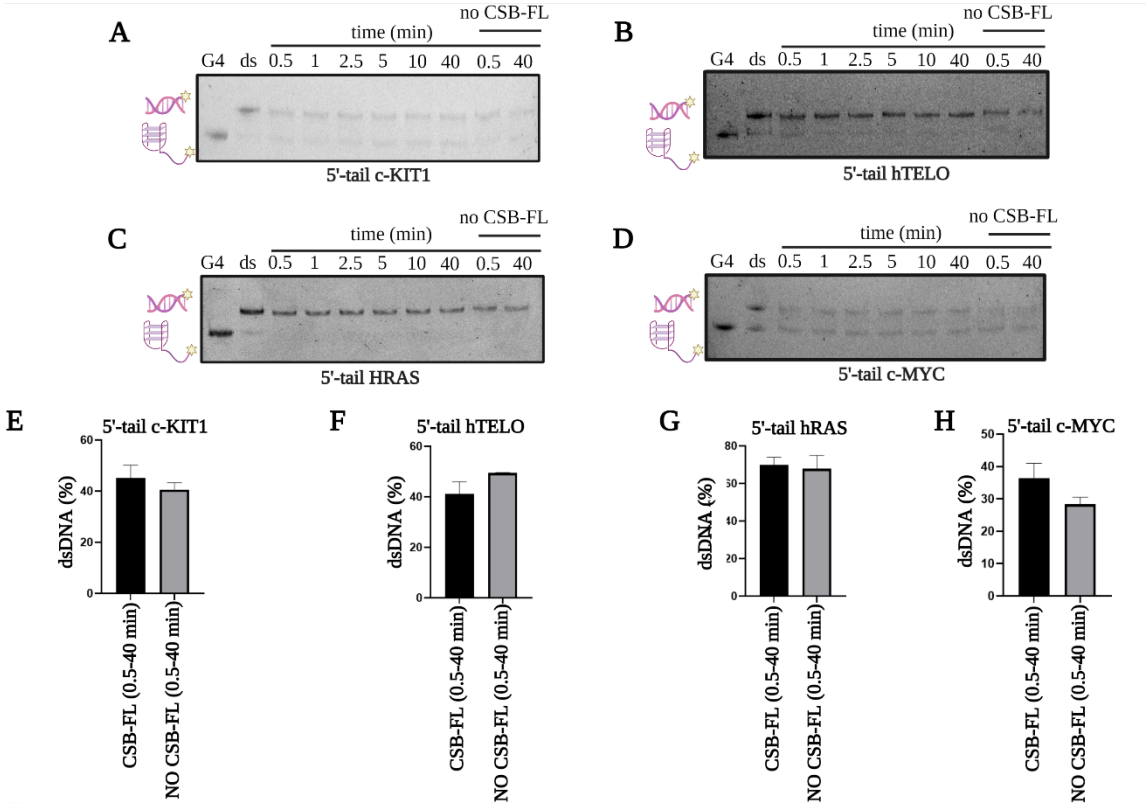

Figure 3.13. Gel-based resolvase assay of a panel of 5'-tailed non-rDNA substrates and their quantification. Gel-based resolvase assays of: (**A**) 5'-tail c-KIT1; (**B**) 5'-tail hTELO; (**C**) 5'-tail HRAS; (**D**) 5'-tail c-MYC. The fluorophore-labelled G4-forming sequence is represented as a purple G4 that runs faster compared to the slower run of dsDNA, and it is represented as pink/purple double helix. Quantification of the gel-based helicase assay using: (**E**) 5'-tail c-KIT1; (**F**) 5'-tail hTELO; (**G**) 5'-tail HRAS; (**H**) 5'-tail c-MYC. The column graphs of quantified gel-based resolvase assays presented in this figure are expressed as percentage of dsDNA formation in the presence or absence of CSB-FL. All quantified gel-based resolvase assays were based on the average of three independent experiments. Significance was calculated based on two-tailed Student's *t*-test. This figure has been adapted from Liano et al.[197].

Unsurprisingly, a similar lack of G4 resolution was also observed for untailed c-MYC, untailed c-KIT, and a 3'-tailed cMYC[197] (Figure 3.14 A-F). Indeed, quantification of the gel-based resolvase assays did not report a statistically significant difference between the percentage of dsDNA formed in the presence or absence of CSB-FL (Table 3.3).

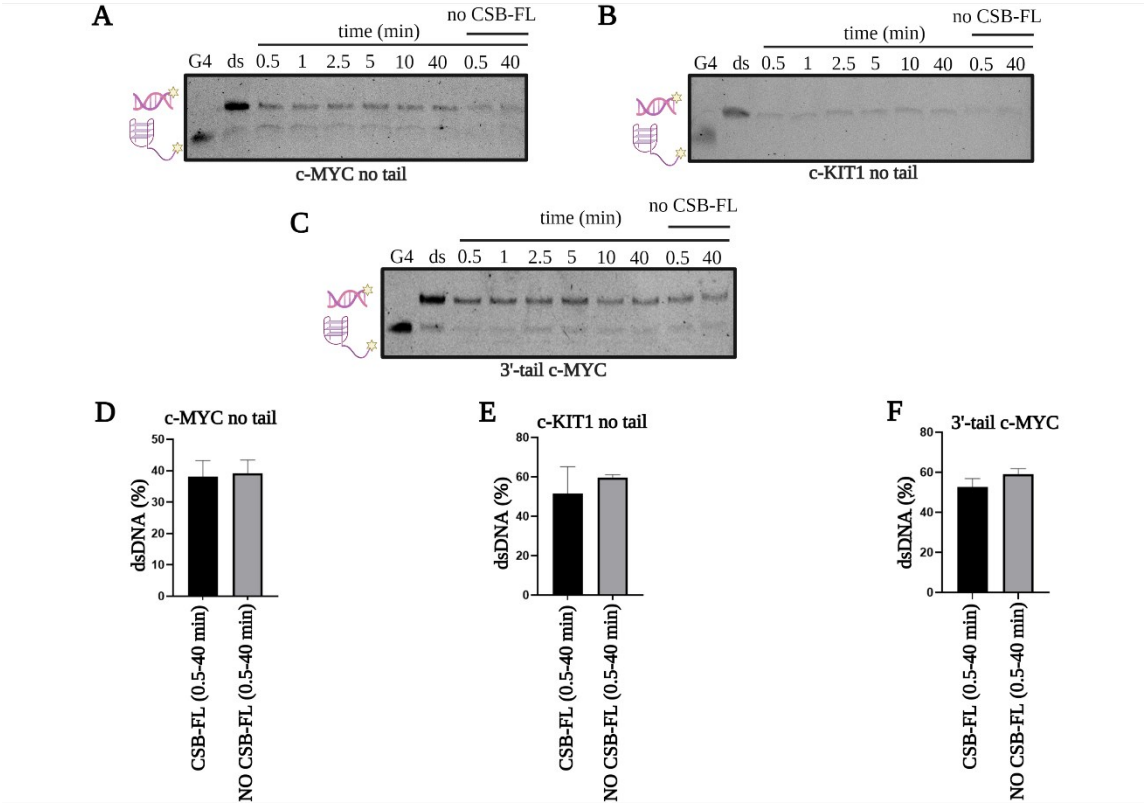

Figure 3.14 Gel-based resolvase assay of a panel of untailed or 3'- tailed G4s and their quantification. Gel-based resolvase assays of: (**A**) untailed c-MYC; (**B**) untailed c-KIT; (**C**) 3'-tail c-MYC; Quantification of the gel-based helicase assay using: (**D**) untailed c-MYC; (**E**) untailed c-KIT; and (**F**) 3-tail c-MYC. The column graphs of quantified gel-based resolvase assays presented in this figure are expressed as percentage of dsDNA formation in the presence or absence of CSB-FL. This figure has been adapted from Liano et al.[197].

G4 sequence	CSB-FL (0-40 min)	NO CSB-FL (0-40 min)
cKIT1 5'-tail	45.1	40.6
hTELO 5'-tail	41.2	49.4
HRAS 5'-tail	69.8	67.8
cMYC 5'-tail	36.3	28.3
cMYC 3'-tail	52.7	59.2
rDNA -1 No tail	24.1	24.6
cMYC No tail	38.1	39.2
cKIT1 No tail	34.8	42.9

double strand (%)

Table 3.3. Quantification of the gel-based resolvase assay gels testing different non-rDNA G4s and untailed-rDNA-1 G4-forming sequences. The gel quantification is represented as percentage of the ratio between the intensities of dsDNA formation and the G4 bands at 0.5 min and 40 min in the presence or absence of CSB-FL.

3.2.4 FRET-based Unfolding Assays Confirmed the Inability of CSB to Resolve Non-rDNA G4s

Although the quantification of the gel-based helicase assays revealed negligible G4-resolvase activity of CSB-FL on non-rDNA G4 substrates, a careful observation of the resolvase gels indicated that some of the non-rDNA G4s tested, for instance c-KIT, hTELO and HRAS (Figure 3.13), could fully form dsDNA in presence of their reverse complementary strand after 0.5 min or 40 min incubation in absence of CSB. The formation of dsDNA in absence of CSB suggested instability of these G4s under the experimental conditions used. Therefore, an additional more quantitative FRET-based unfolding assay that followed the G4-resolution in real time was needed to overcome the time points limitation associated to the gel-based resolvase assays and further assure the accuracy of the results obtained from interpretation of the resolvase gels. In more detail, FRET requires a dually fluorophore-labelled G4-forming sequence containing different fluorophores at its 5'- and 3'-ends, called donor and acceptor, respectively. When the G4 structure is folded, the two fluorophores are relatively close to each other. Consequently, excitation of the donor will result in emission of the closer acceptor, causing an overall low fluorescence of the donor. Addition of a G4-resolvase in the

sample will instead promote a gradual increase in the donor signal as the two fluorophores will be more separated in the space upon resolution of the G4 structure. The gradual increase of the donor emission can be followed in real time and it is indicative of the G4-unfolding in presence of the G4-resolvase. The linearised substrate is then stabilised by its reverse complementary strand, which is also added into the sample, with consequent formation of a dsDNA substrate. The comparison between the G4-unfolding rate in the presence and in absence of the protein provides a quantitative indication of the ability of the protein to resolve the G4 structure, which we could have missed in the single points gel-based resolvase assays (a more detailed overview on FRET-based unfolding assay is described in Section 2.2).

In this work, the FRET-based resolvase assays were performed with the help of Souroprobho Chowdhury (a third-year PhD student in M. Di Antonio's group). The G4-resolvase activity of CSB towards non-rDNA G4s was tested using a parallel c-KIT1 G4-forming sequence and mixed-type hTELO G4 (see Table 5.7 for the oligonucleotide sequences used), both annealed under KCl conditions to promote stable G4-formation. These G4s were both dually labelled with 5'-Fluorescein (FAM, donor) and 3'-Carboxytetramethylrhodamine (TAMRA, acceptor) fluorophores, as FRET-pair. The unfolding rate of c-KIT1 and hTELO G4s was measured in real time either in the presence or absence of CSB-FL or CSB-HD (498–1,002). Notably, both proteins were not able to accelerate the unfolding rate of the two G4s tested, as no difference in rate was detected in the presence of CSB-FL or CS-HD (Figure 3.15A-D). The assay was also run in the presence of ATP to further confirm the lack of CSB-HD resolvase activity towards c-KIT1.

Altogether, these results indicate that the G4-resolvase activity of both CSB-FL and CSB-HD is only partial and limited to tailed rDNA G4s.

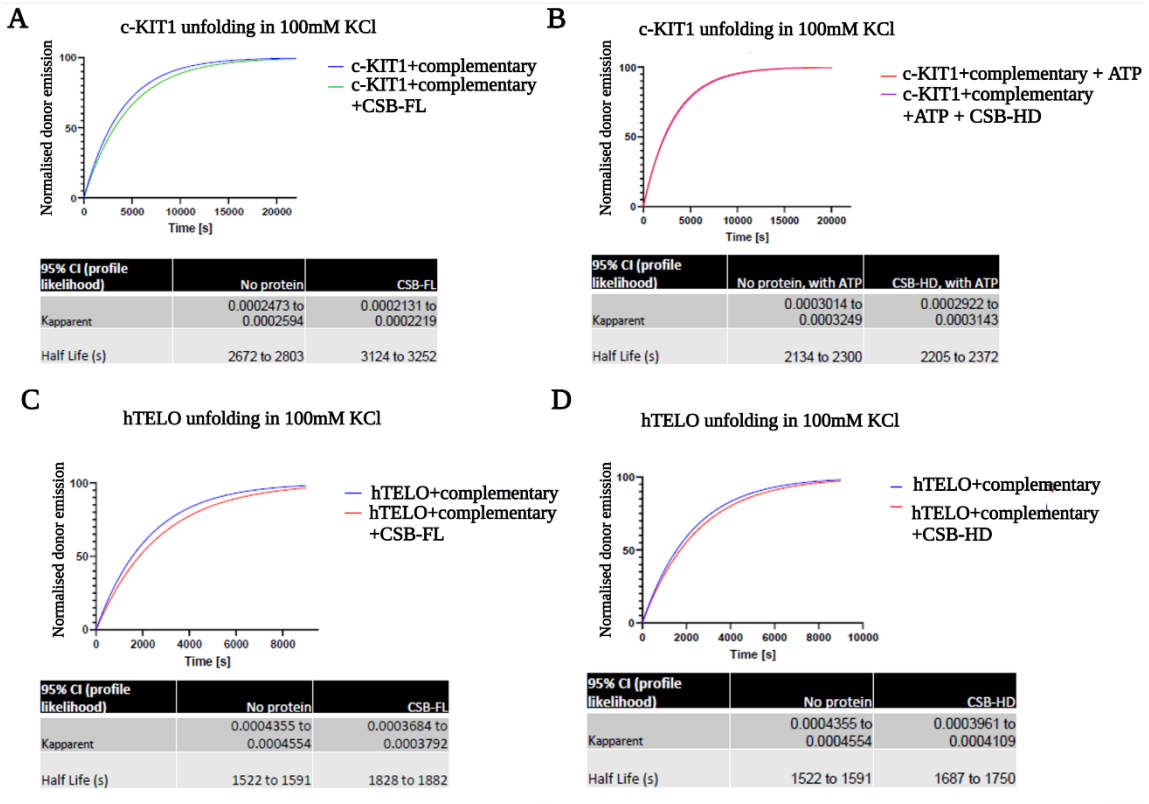

Figure 3.15. FRET-based helicase assays confirmed the absence of non-rDNA G4 resolution after incubation with either CSB-FL or CSB-HD. (**A**) FRET-based helicase assay performed using c-KIT1 substrate in presence of its complementary sequence. The experiment was performed with 100 mM KCl either in the presence or absence of CSB-FL. (**B**) FRET-based helicase assay performed using c-KIT1 in the presence or absence of CSB-HD. In this specific case, the sample was supplemented with 5 mM ATP to further confirm the lack of G4 resolution. (**C**) FRET-based helicase assay performed using hTELO substrate in the presence or absence of CSB-FL. (**D**) FRET-based helicase assay performed using hTELO substrate in the presence or absence of CSB-HD. All experiments were performed in triplicates. Curves represent the best-fit from single-exponential curve-fitting to triplicate data. Half-lives expressed in seconds as 95% credible interval profile likelihood (computed using Graphpad Prism), with constraints and plateau set as y0=0 and 100, respectively [(197)].

3.2.5 rDNA G4-Resolvase Activity of CSB Is Not Dependent on CSB Concentrations or Incubations Times

Once the selectivity of CSB for rDNA G4s was confirmed, multiple screenings were performed to investigate the conditions under which CSB was able to fully resolve rDNA G4s. Either higher concentrations of CSB-FL were used for the gel-based resolvase assays, or the protein was incubated for longer times with the 5'-tailed rDNA-1 substrate to investigate time-dependency.

Interestingly, exposure to either 20 nM or 40 nM CSB-FL did not improve the rDNA G4 resolution. An intense signal from the fast-running G4 band was observed in all the samples (Figure 3.16A), indicating that most of the G4 was still folded after 40 min incubation with higher concentrations of the protein. Similarly, longer incubation times of the 5'-tailed rDNA-1 with CSB-FL extended up to 160 min or overnight (O/N) were not sufficient to fully resolve the G4 structure (Figure 3.16B) [197]. Based on these observations we speculated that the presence of a specific structural G4 topology exclusively formed within rDNA sequences under the tested experimental conditions was responsible for the partial resolvase activity displayed by the protein [197].

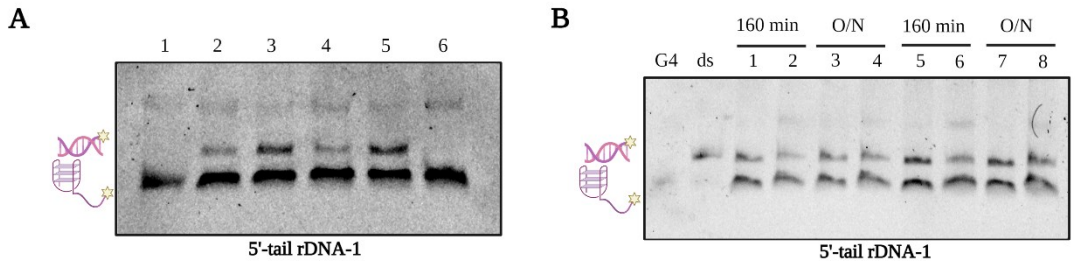

Figure 3.16 The resolvase activity of CSB towards rDNA G4 is not dependent on CSB concentrations or incubation times. (**A**) Resolvase gel using higher CSB-FL concentrations. Lane 1: control 5'-tailed rDNA-1 incubated with its reverse complementary for 0.5 min in absence of CSB-FL. Lane 2: incubation of 5'-tailed rDNA-1 and its reverse complementary with 20 nM CSB-FL for 0.5 min. Lane 3: incubation with 20 nM CSB-FL for 40 min. Lane 4: incubation with 40 nM CSB-FL for 0.5 min. Lane 5: incubation with 40 nM CSB-FL for 40 min. Lane 6: control 5'-tailed rDNA-1 incubated with its reverse complementary for 40 min in absence of CSB-FL. (**B**) Gel-based resolvase assay testing longer incubation time. G4 and the ds lane indicate the controls. Lane 1 and 3: incubation with 5 nM CSB-FL for 160 min or O/N, respectively. Lane 2 and 4: incubation without CSB-FL for 160 min or O/N, respectively. Lane 5 and 7: incubation with 10 nM CSB-FL for 160 min or O/N, respectively. Lane 6 and 8: incubation without CSB-FL for 160 min or O/N, respectively. The gels shown in this figure have been adapted from Liano et al. [197].

3.2.6 CD Analysis Failed to Identify Structural Features that are Specific of rDNA G4s

To understand whether the rDNA sequences could form a specific G4-topology selectively resolved by CSB, the CD analysis of the rDNA-1 and rDNA-2 sequences was performed (CD) (see section 1.2.2 for an overview of the CD technique). The CD spectra of all the sequences tested identified a clear parallel G4-conformation, as assessed by the maxima ~263 nm and minimum ~240 nm [228] (Figure 3.17A-B and D-E), except for the untailed rDNA-1 which revealed a mixed topology [229] (Figure 3.17C). The CD spectra of an additional rDNA-3 (see Table 5.7 for the oligonucleotide sequence) confirmed that rDNA G4s are mostly folded in a parallel conformation both in the presence and absence of the tail (Figure 3.17F-G), suggesting that the selectivity of CSB towards rDNA G4s is unlikely to be caused by a topology-based selectivity of the protein [197].

Interestingly, the CD spectra of the tested rDNAs revealed that these sequences were still able to form G4 structure when annealed in buffer containing LiCl. The Li^+ cation is a weak G4-stabiliser and lower G4-formation is normally expected under LiCl conditions. However, the ability of rDNA sequences to form G4s even when annealed in LiCl buffer, suggested that rDNA G4s are generally very stable and can promptly form parallel G4s under non G4-stabilising conditions either in the presence or absence of the 5'- tail, with the exception of untailed rDNA-1, which forms a mixed-type G4 under K^+ stabilisation.

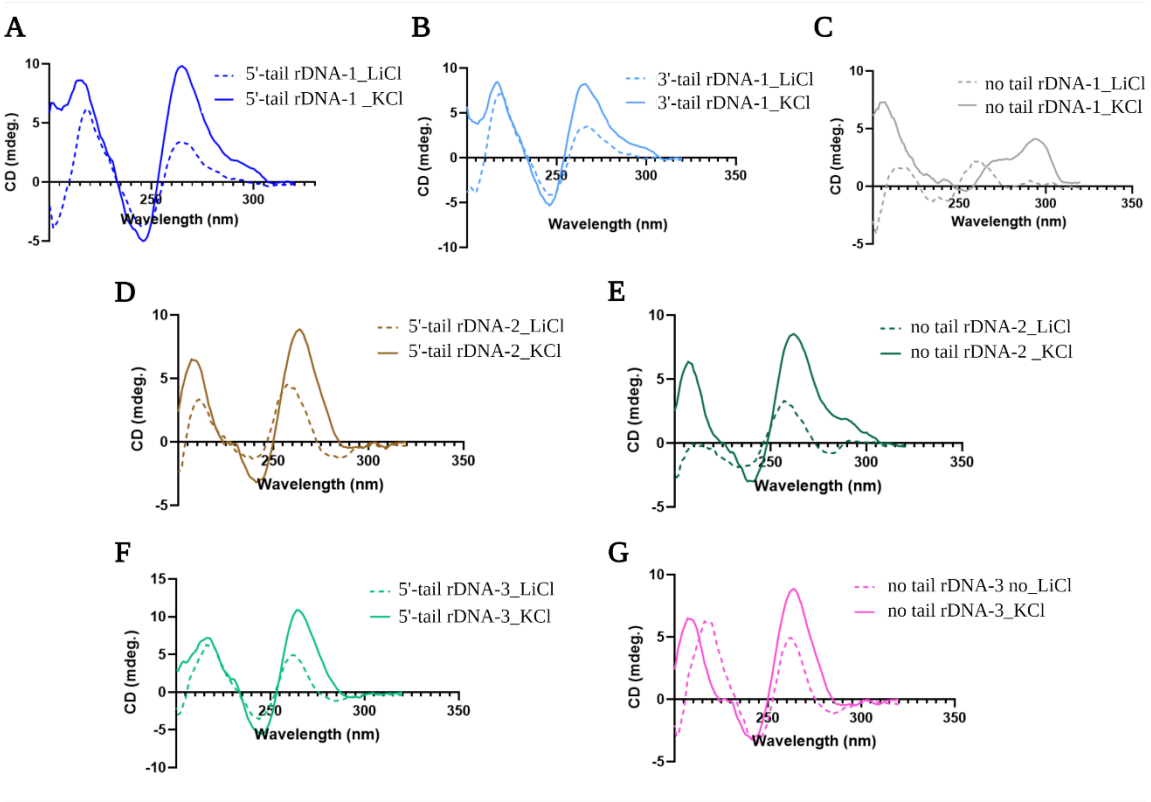

Figure 3.17. CD analysis revealing mostly parallel rDNA G4-topology in LiCl and KCl buffers. (**A**) CD spectra of 5'-tailed rDNA-1. (**B**) CD spectra of 3'-tailed rDNA-1. (**C**) CD spectra of untailed rDNA-1. (**D**) CD spectra of 5'-tailed rDNA-2. (**E**) CD spectra of untailed rDNA-2. (**F**) CD spectra of 5'-tailed rDNA-3. (**G**) CD spectra of untailed rDNA-3. The recorded spectra represent the average of three different reads corrected by using the CD-spectra of buffer alone as a blank. This figure has been adapted from Liano et al. [197].

3.2.7 Polymerase Stop Assays Confirmed the High Stability of rDNA G4s

To validate the high propensity of rDNA sequences to form G4 even in absence of the stabiliser K^+ cation, polymerase stop assays were performed to assess the relative stability of these G4-structures. The polymerase stop assay is based on the principle that a G4 within a DNA template acts as a knot and blocks a DNA polymerase processing and replicating the DNA template. In fact, the DNA polymerase is incapable to resolve the stabilised G4 and pauses replication at the G4-forming site. Hence, the stalling of the DNA polymerase is indicative of G4-stability, with more stable G4s capable of stalling DNA polymerase to a greater extent. The stalling of the polymerase could be estimated by gel electrophoresis. Indeed, the stalling of the DNA polymerase at the G4-forming site will generate prematurely terminated DNA products that run faster than the fully elongated ones [138] (Figure 3.18A and Section 1.3.2 for a further overview about the polymerase stop assay technique).

To confirm the stability of rDNA G4s, two different rDNA G4-forming sequences were tested by the polymerase stop assay. In particular, either the rDNA-1 or a different rDNA-4 G4-forming sequence [195,197] was inserted within two separate DNA templates, which present a region complementary to a fluorescently labelled primer (see Table 5.8 for the oligonucleotide sequences used). Each G4-forming template was mixed with the primer and annealed by boiling and slow cooling the mixture to RT either in KCl or LiCl buffers. The annealing of the mixture allowed both the primer to bind the template and the G4 to form (Figure 3.18A). Then, the annealed samples were incubated with either the G4-stabiliser PDS or increasing concentrations of CSB-FL (or -HD) (0.5 nM, 5 nM, and 50 nM). Starting from the free 3'-OH of the primer, the polymerase will extend the template. This template will be fully elongated if the G4 structure is not stabilised whilst premature terminated products are expected if the G4 blocks the processivity of the polymerase (Figure 3.18A). After incubation, the samples were run on an electrophoretic gel. Fast-running premature terminated products, which are indicative of polymerase arrest in the presence of the G4 structure, were observed equally in KCl and LiCl indicating that rDNA G4s are highly stable also under LiCl conditions. As expected, maximum stalling of the polymerase was observed in the presence of PDS (Figure 3.18B-E) [197]. The polymerase stop assay also revealed that the highest concentration (50 nM) of CSB-FL (Figure 3.18B-C) or CSB-HD (Figure 3.18D-E) was generally required to observe full elongation of the DNA polymerase, while prematurely terminated products were still

observed in the presence of lower concentrations (0.5 nM and 5 nM) of either CSB-FL or CSB-HD.

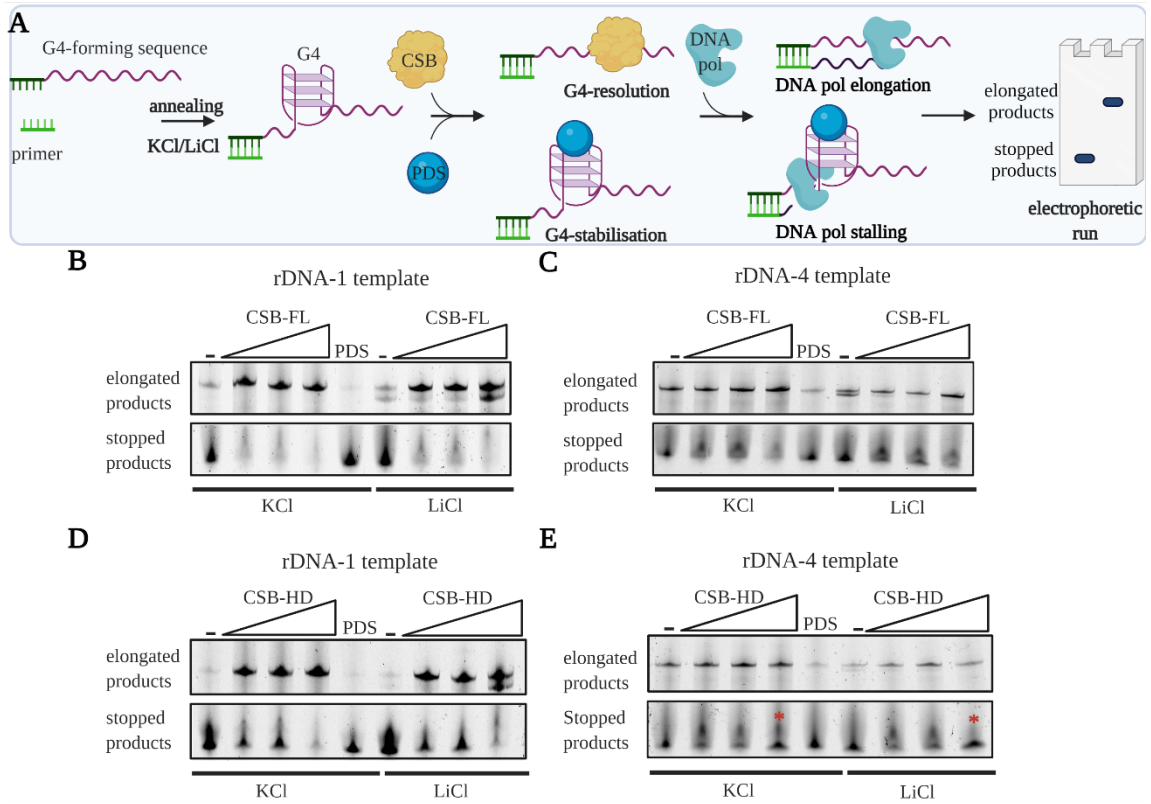

Figure 3.18. Polymerase stop assay indicating polymerase arrest at the level of the rDNA G4-forming sequence both in KCl and LiCl. (**A**) Schematic representation of the polymerase stop assay. (**B**) Electrophoretic gel showing the formation of elongated and stopped rDNA-1 products in presence or absence (-) of increasing concentrations of CSB-FL or pyridostatin (PDS) either under KCl or LiCl conditions. (**C**) Experiment performed using rDNA-4 sequence. (**D**) Electrophoretic gel showing the formation of elongated and stopped rDNA-1 products in the presence or absence (-) of increasing concentrations of CSB-HD or pyridostatin (PDS) under KCl or LiCl conditions. (**E**) Experiment performed using rDNA-4 sequence. The red asterisks indicate a possible loss of activity of the protein or experimental variability. This figure has been adapted from Liano et al. [197]

3.2.8 CSB Selectively Binds Intermolecular rDNA G4s

rDNA CAN FORM INTERMOLECULAR G4s

To further investigate the nature of the resolution of rDNA G4s elicited by CSB, EMSAs were performed to assess the binding affinity of CSB towards the same panel of G4s tested in

the resolvase assays. The interaction between CSB and different G4-forming sequences could be identified on a native gel using fluorescently labelled oligonucleotides. The fluorophore-labelled oligonucleotide was incubated with increasing concentrations of the protein for a fixed time prior to the electrophoretic run. The oligonucleotide-CSB complex presented less mobility through a native gel than the free oligonucleotide, therefore, a shift in the oligonucleotide band is indicative of the interaction between the protein and the substrate (further details about the EMSA technique are provided in Section 2.2).

In more details, a Cy5- labelled 5'-tailed rDNA-1 annealed in KCl buffer was firstly incubated with increasing concentrations (0–150 nM) of CSB-FL for 30 min. The samples were then run on a native polyacrylamide gel. The gel showed a shift in the G4 band starting at approximately 40 nM of the protein (Figure 3.19A) indicating interaction between CSB-FL and the tailed rDNA-1 G4. Furthermore, on the same gel, two slow-moving low-intensity bands were observed in the free rDNA sample (green arrows, Figure 3.19A), suggesting that alternative higher MW structures could form after annealing of rDNA-1 G4 that were not observed in the gels used in the resolvase assays.

To better elucidate the nature of these different bands within the free rDNA-1, the samples were also run on an agarose gel, as an efficient and effective way of separating nucleic acids and large protein complexes [230]. Similar to the polyacrylamide gel, the agarose gel revealed clear formation of two higher MW bands in the free rDNA-1 G4 sample annealed in KCl buffer (Figure 3.19B) [197]. Interestingly, these higher MW bands disappeared when rDNA-1 was annealed using LiCl buffer (Figure 3.19C). The formation of slow-running bands only in the presence of the G4-stabiliser cation K^+ further suggested the hypothesis of formation of alternative G4 structures within the tailed rDNA sample that requires KCl to be observed. Specifically, the higher MW bands formed could be indicative of the formation of intermolecular G4s between multiple rDNA sequences, while the fast-moving bands present in both KCl and LiCl conditions could be indicative of more stable intramolecular G4-formation.

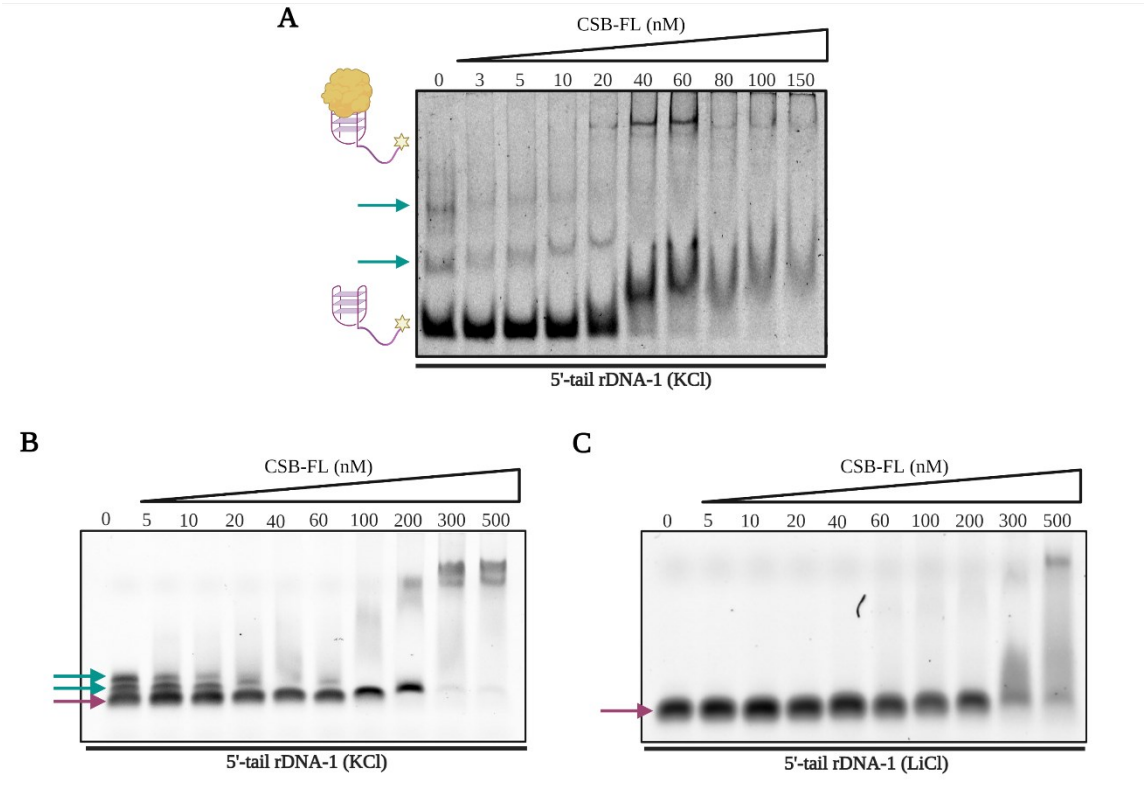

Figure 3.19. 5'-tail rDNA-1 can form higher MW bands under KCl condition. (**A**) Polyacrylamide EMSA gel showing molecular shift of the free G4-bands from ~40 nM CSB-FL indicating the interaction between the protein and 5'-tail rDNA substrate. The G4-related band is indicated with a fluorophore-labelled purple G4 at the bottom of the gel. The protein-G4 complex is represented as a globular yellow protein bound to the fluorophore-labelled G4 structure. The two green arrows indicate formation of multiple higher MW structures within the free rDNA sample. (**B**) Agarose EMSA gel confirmed the formation of higher MW bands within the 5'-tail rDNA substrate under KCl conditions. (**C**) Agarose EMSA gel of the 5'-tail rDNA substrate under LiCl conditions. The fast-running band corresponds to intramolecular G4 (purple arrow), while the slow-running bands correspond to intermolecular G4s (green arrows). The gels shown in (B) and (C) have been adapted from Liano et al.[197].

To understand whether the two slow-moving bands were intermolecular G4s, the unlabelled 5'-tailed rDNA-1 was annealed either in KCl or LiCl buffer and run in a native agarose gel in the presence of a dsDNA ladder. Again, the gel revealed only a single fast-running band when the rDNA substrate was annealed in LiCl buffer, while the formation of higher MW bands was observed using KCl. The apparent MW of these slow-running bands was either twice or four times the one expected for 5'-tailed rDNA-1 (48 bases) (Figure 3.20A) [197], confirming that the 5'-tailed rDNA-1 substrate forms intermolecular G4s (bimolecular or tetramolecular,

respectively) under stabilising KCl conditions. The fast-moving bands observed both in KCl and LiCl could instead indicate the formation of more stable intramolecular G4s, as those formed under LiCl conditions were observed by CD analysis and polymerase stop assays [197].

To validate that all the bands detected in the gel were indeed representative of folded G4s (intra and inter molecular), the same agarose gel was stained with a G4-specific probe N-methylmesoporphyrin IX (NMM) [231] (Figure 3.20B). This experiment showed NMM staining of both the higher and lower MW bands formed from all the three different 5'-tailed rDNAs that were previously tested, ascribing both the fast- and slow-moving bands to G4 conformations [197]. Negligible staining was observed for a ssDNA control that was not able to form a G4 structure (see Tables 5.7 and 5.8 for all the oligonucleotide sequences used in this experiment). Moreover, a truncated version of the 5'-tailed rDNA-1 sequence was designed by removing one of its four G-tracts. Due to the lack of four G-tracts, this 5'-tailed truncated rDNA-5 bimolecular sequence was no longer able to form an intramolecular G4 but could form intermolecular G4s. Indeed, this oligonucleotide was still generating higher MW bands when annealed under KCl conditions (Figure 3.20C). The CD spectra of the truncated rDNA-5 bimolecular sequence annealed in KCl buffer confirmed parallel G4-formation (Figure 3.20D), further validating that rDNA could form intermolecular G4 structures [197].

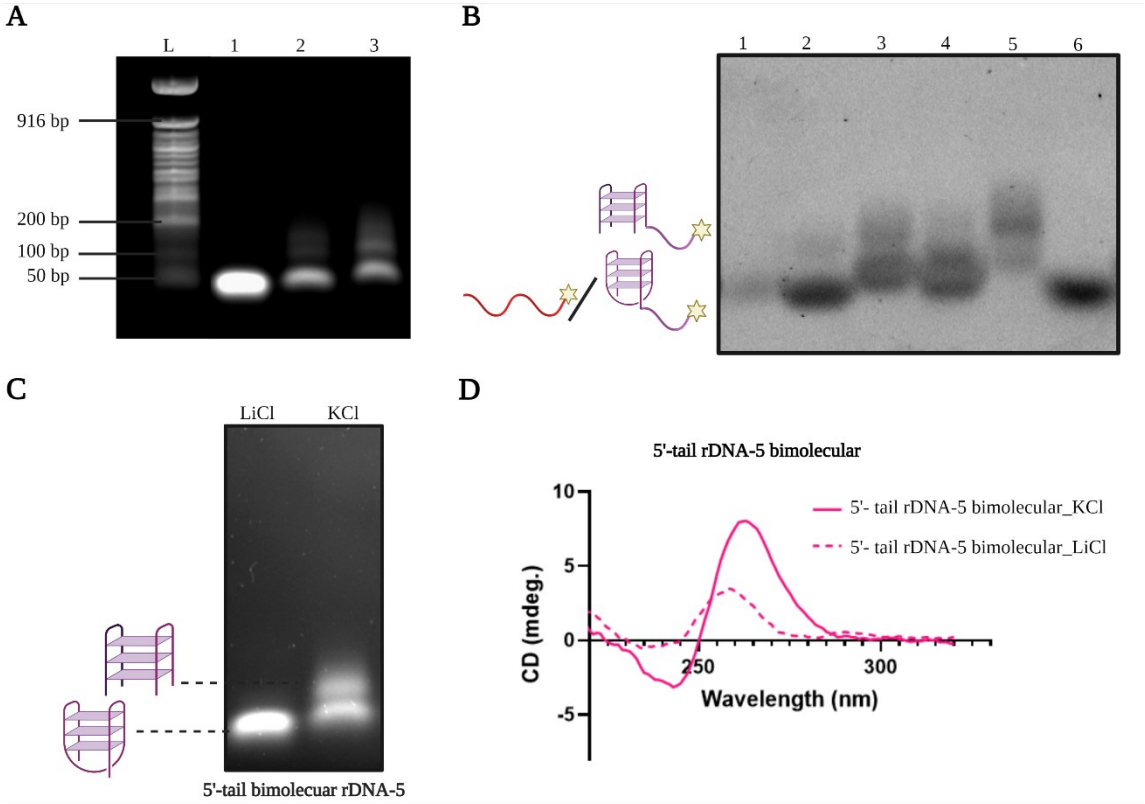

Figure 3.20. rDNA can form intermolecular G4-structures under KCl condition. (**A**) Agarose gel of unlabelled 5'-tail rDNA-1 substrate annealed in buffer containing either LiCl or KCl. L is the dsDNA ladder. Lane 1: 5'-tail rDNA-1 annealed in buffer containing 100 mM LiCl. Lane 2: 5'-tail rDNA-1 annealed in buffer containing 100 mM KCl. Lane 3: 5'-tail rDNA-1 annealed in buffer containing 100 mM KCl+30% PEG200. (**B**) NMM staining of agarose gel. Lane 1: ssDNA, Lane 2: untailed rDNA-1 in KCl buffer, Lane 3: 5'-tail rDNA-1 in KCl buffer, Lane 4: 5'-tail rDNA-2 in KCl buffer, Lane 5: 5'-tail rDNA-3 in KCl buffer, Lane 6: untailed rDNA-1 in LiCl buffer. ssDNA band is indicated with a red linear strand, intramolecular G4s are indicated with a light purple G4 while intermolecular G4s are indicated with a G4 formed by light and dark purple strands. (**C**) Agarose gel of 5- tail truncated bimolecular rDNA-5 sequence annealed either in LiCl or KCl. The fast-running band corresponding to the intramolecular G4 is indicated with a light purple G4, while the slow-running band corresponding to the intermolecular G4 is indicated with a G4 formed by a light and a dark purple strand. (**D**) CD spectra indicating G4-formation within the truncated 5'-tail rDNA-5 substrate. The spectra were recorded both using KCl buffer and LiCl buffer. The absorbance of the buffers was subtracted from the recorded spectra. The final spectra represent the average of three different reads. The results presented in this figure have been adapted from Liano et al.[(197)].

CSB SPECIFICALLY INTERACTS WITH INTERMOLECULAR rDNA G4s

The agarose EMSAs obtained after incubation of CSB-FL with the 5'-tailed rDNA-1 not only revealed formation of intramolecular and intermolecular G4s under KCl conditions, but the ability of CSB to interact with these rDNA G4 substrates, as suggested by the polyacrylamide EMSA. Moreover, the agarose EMSA also indicated a specific interaction between CSB-FL and the intermolecular G4s, whilst binding to the intramolecular G4 required a higher concentration of the protein, both under KCl or LiCl conditions (Figure 3.19B and C).

Similarly, low concentrations of CSB-HD (498–1,002) were able to interact only with the intermolecular rDNA-1 G4 subtype that was formed under KCl conditions (Figure 3.21A), while higher concentrations of CSB-HD were required for the interaction with the intramolecular G4 substrate, both in KCl and LiCl buffers (Figure 3.21A and B). This specific interaction of CSB-FL and CSB-HD with the intermolecular G4s formed within the 5'-tailed rDNA-1 suggested higher specificity of the proteins for these particular G4 subtypes than that of the intramolecular ones.

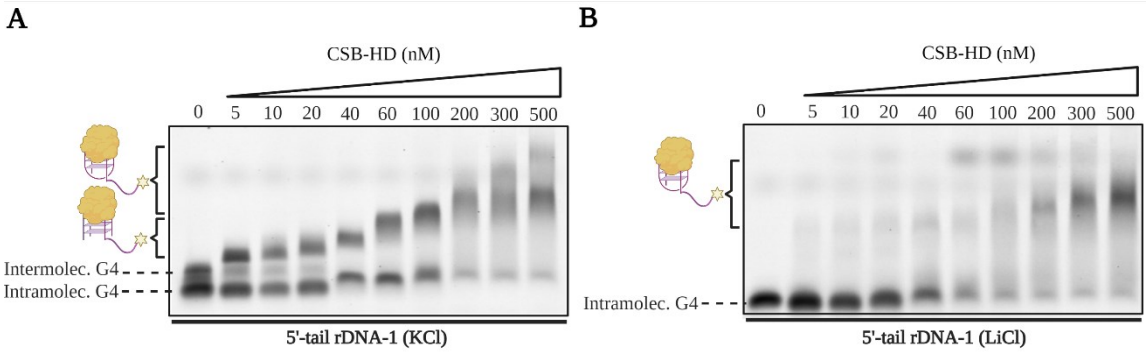

Figure 3.21. CSB interacts with intermolecular rDNA G4s with higher affinity than that of intramolecular G4s. (**A**) Agarose EMSA performed using 5'-tailed rDNA-1 substrate annealed in KCl buffer and increasing concentrations of CSB-HD. The interaction between the slow-moving intermolecular G4 and low concentrations of CSB-HD is indicated with a yellow globular protein that interacts with the G4 formed from light and dark purple strands. The interaction between the fast-moving intramolecular G4 and high concentrations of CSB-HD is indicated with a yellow globular protein that interacts with the G4 formed from a light purple substrate. (**B**) Agarose EMSA performed using 5'-tailed rDNA-1 substrate annealed in LiCl buffer and increasing concentrations of CSB-HD. This figure has been redrawn by adapting the results presented in Liano et al. [(197)].

Notably, the interaction between CSB and intermolecular G4 was saturated at 5 nM CSB-HD, suggesting sub-nanomolar affinity. More EMSA experiments were performed using a lower concentration range of CSB to assess the K_d of CSB binding to intermolecular rDNA G4s [(197)]. Since the helicase- "like" domain of CSB displayed the same activity as the full-length protein in all the assays performed, CSB-HD (498–1,002) was used thereafter. Initially, the 5'-tailed rDNA-1 sequence annealed either in KCl or LiCl was incubated with increasing concentration of CSB-HD (0–5 nM). As expected, CSB-HD selectively bound the intermolecular G4 bands formed in KCl buffer (Figure 3.22A), whilst negligible interaction was observed between CSB-HD and the intramolecular G4 bands formed both in KCl and LiCl buffers (Figure 3.22A and B). The K_d calculation for this interaction revealed an astonishing picomolar affinity of CSB-HD for the intermolecular rDNA-1 G4s with a K_d of 557.5 pM [442.2–698.0 pM – 95% CI] (Figure 3.22C) [(197)].

The high selectivity and affinity of CSB towards intermolecular rDNA G4s was further investigated by testing different rDNA substrates. The 5'-tailed rDNA-2 and the 5'-tailed rDNA-3 G4-forming sequences previously tested by gel-based resolvase assays and CD were annealed both in KCl and LiCl buffers and incubated with increasing concentrations of CSB-

HD (0–5 nM). Agarose gel showed the formation of intermolecular G4s within both the 5'-tailed rDNA-2 and 5'-tailed rDNA-3 substrates under KCl conditions (Figure 3.22D and E). Although less intense compared to that with KCl, the controls also annealed in LiCl showed the formation of intermolecular G4s to a lesser extent, suggesting high stability of intermolecular G4s formed within rDNA-2 and rDNA-3. Consistent with the results obtained for rDNA-1, CSB-HD was able to bind the intermolecular G4 subtype with high affinity, yielding K_ds of 359.9 pM [205.2–594.1 pM – 95% CI] and 977.1 pM [476.3–1924 pM – 95% CI] towards intermolecular rDNA-2 and rDNA-3 G4s, respectively (Figure 3.22C-E) [197].

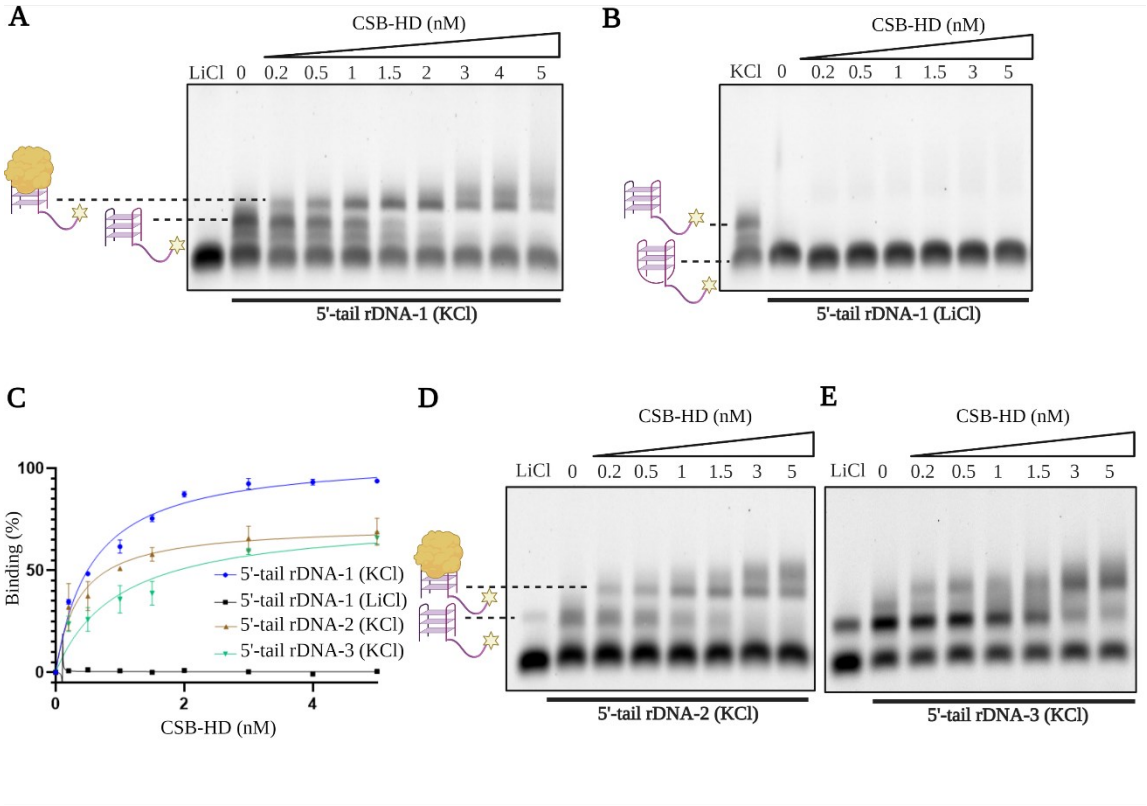

Figure 3.22. CSB interacts with intermolecular rDNA G4s with picomolar affinity. (**A**) Agarose EMSA gel on 5'-tailed rDNA-1 G4 using 0 to 5 nM CSB-HD in KCl buffer. The first lane is a control with the oligonucleotide annealed in LiCl. The intermolecular G4s formed within rDNA-1 sequence in KCl are represented with a fluorophore-labelled G4 formed between a light and a dark purple sequence. The interaction between the intermolecular G4 and CSB-HD is indicated with a yellow globular protein that interacts with the G4 formed from light and dark purple strands. (**B**) Agarose EMSA gel on 5'-tailed rDNA-1 G4 using 0 to 5 nM CSB-HD in LiCl buffer. The first lane is a control with the oligonucleotide annealed in KCl. Intramolecular G4s are represented with a G4 formed within the same light purple sequence. Intermolecular G4s are represented with a light and dark purple G4. (**C**) Binding curves expressing the percentage of different rDNA sequences bound to increasing concentrations of CSB-HD (0 to 5 nM) in KCl or LiCl buffers. (**D**) Agarose EMSA gel on 5'-tailed rDNA-2 G4 using 0 to 5 nM CSB-HD in KCl buffer. The first lane is a control with the oligonucleotide annealed in LiCl. (**E**) Agarose EMSA gel on 5'-tailed rDNA-3 G4 using 0 to 5 nM CSB-HD in KCl buffer. The first lane is a control with the oligonucleotide annealed in LiCl. All the EMSAs were performed in duplicates and the K_ds were calculated using a one site specific binding equation using GraphPad Prism 9.0.1 software and 95% CI. This figure has been adapted from Liano et al. [197].

3.2.9 CSB Does Not Bind Intramolecular G4s

To understand whether CSB could also bind different G4s or non-G4 structures, the same EMSAs using native agarose gels were performed incubating CSB-HD with a panel of different non-rDNA substrates.

To achieve this, the binding activity of CSB-HD was tested for the same panel of G4-forming sequences used for the gel-based resolvase assays. Interestingly, CSB-HD was not able to bind 5'-tailed c-KIT1, c-MYC, HRAS, and hTELO at concentrations ranging between 0 and 5 nM (Figure 3.23A-D). Unsurprisingly, the agarose gels revealed the absence of intermolecular G4 formation within all the sequences tested with the only formation of a single fast-moving intramolecular G4-band, both in KCl and LiCl buffer, as expected for these very well characterised G4-forming sequences [197]. In agreement with what was observed for rDNA-1, CSB-HD was able to interact with the intramolecular G4 structure only at high protein concentrations (>50 nM CSB-HD), confirming the lower binding affinity of CSB towards intramolecular G4s.

Similarly, high concentrations of CSB-HD were required to observe the binding between the protein and either a non-G4-forming single stranded (ss) DNA sequence (Figure 3.23E) or a mutated rDNA-1 sequence, which was no longer able to form G4s (Figure 3.23F) (see Table 5.8 for the oligonucleotide sequences used), suggesting that under these conditions, CSB was unable to discriminate between different types of DNA structures, leading to unspecific binding [197].

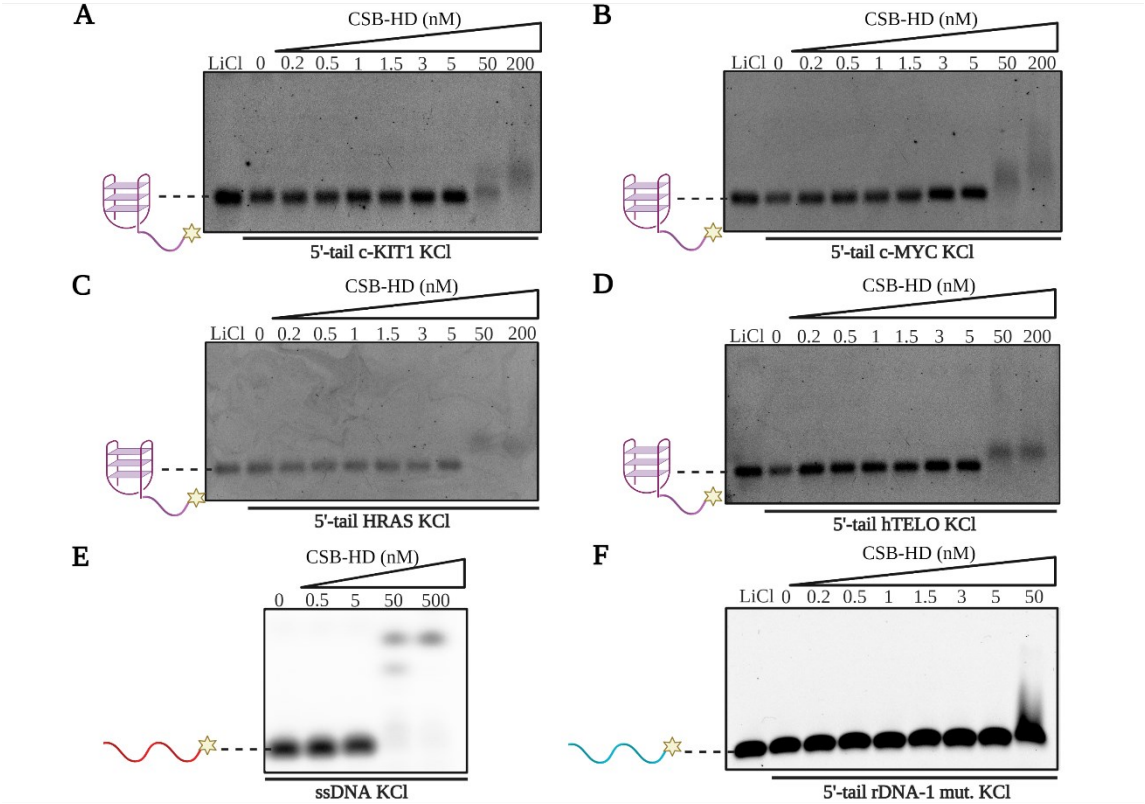

Figure 3.23. CSB binds intramolecular G4s with low affinity and this interaction is not specific over ssDNA. (**A**) Agarose EMSA incubating increasing concentrations of CSB-HD with 5'-tailed c-KIT1 substrate annealed in KCl buffer. The first lane is the control annealed in LiCl. The intramolecular G4 band is represented with a G4 structure formed within the same fluorophore-labelled sequence. (**B**) Agarose EMSA using 5'-tailed c-MYC substrate. (**C**) Agarose EMSA using 5'-tailed HRAS substrate. (**D**) Agarose EMSA using 5'-tailed hTELO substrate. (**E**) Agarose EMSA using a non-G4-forming ssDNA sequence. The bands corresponding to the fluorophore-labelled ssDNA are represented with a red ss sequence. (**F**) Agarose EMSA using a mutated 5'-tailed rDNA-1 sequence. The band on the gel corresponding to the 5'-tailed rDNA-1 free substrate is represented as a light blue ss sequence. Figure adapted from Liano et al. [(197)].

3.2.10 Necessary 5'- or 3' Tail for Intermolecular rDNA G4s to Form

We next wanted to understand the influence of the tail on the CSB-G4 interaction and if different positioning of the tail, or its absence, could perturb the binding affinity of CSB-HD towards intermolecular rDNA G4s.

To achieve this, we tested CSB binding against a 3'-tailed rDNA-1 and a 3'-tailed c-MYC sequence using the same EMSAs conditions described in Section 3.2.8. The 3'-tailed rDNA-1 showed the formation of intermolecular G4s bands when annealed in KCl buffer, similar to the observed for the 5'-tailed counterpart (Figure 3.24A). As expected, CSB-HD was able to only bind with high affinity the intermolecular G4 structure formed by the 3'-tailed oligonucleotide. Calculation of the K_d using the EMSA gel, revealed picomolar affinity of CSB-HD on the intermolecular G4 substrate with K_d of 175.5 pM [82.7–315.9 pM – 95% CI] (Figure 3.24A and B). The K_d value indicated ~3-fold higher affinity of CSB-HD for the 3'-tailed rDNA-1 than that of its 5'-tailed counterpart, indicating that the G4-binding of CSB might have a slight preference for a 3'- to 5'- orientation[197]. As expected, only a single fast-running band indicative of an intramolecular G4 was observed with the 3'-tailed c-MYC substrate, with negligible interaction detected between CSB-HD and the 3'-tail c-MYC G4 (Figure 3.24C) [(197)].

Subsequently, the relevance of the tail for binding was investigated by testing untailed templates, including untailed rDNA-1, rDNA-2, and c-MYC. Interestingly, the agarose EMSAs displayed negligible formation of intermolecular G4s for both rDNA-1 and rDNA-2 in absence of the tail annealed in K^+, with consequent lack of binding to these untailed substrates. These results confirmed the lack of intermolecular G4-formation for untailed rDNA-1 as observed by NNM staining (Figure 3.20B), and suggested that either a 5'- or a 3'-tail is necessary for the intermolecular rDNA G4s to form under our experimental conditions[197]. Unsurprisingly, the untailed c-MYC template showed only the formation of a single intramolecular G4 band also not bound by CSB-HD.

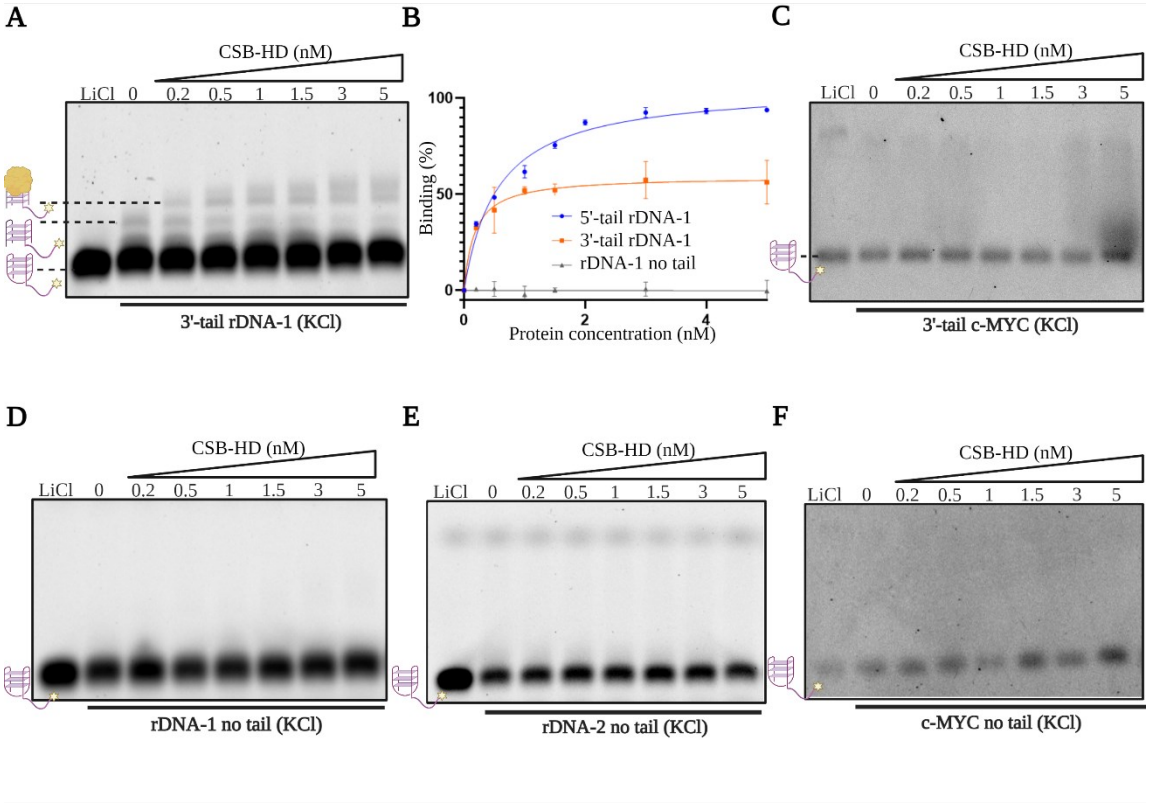

Figure 3.24. Essential 5'- or -3' tail for intermolecular rDNA G4 formation. (**A**) Agarose EMSA showing interaction between low concentrations of CSB-HD and the 3'-tail rDNA-1. The first lane is the control annealed in LiCl. The intramolecular G4 band is represented with a G4 structure formed within the same fluorophore-labelled sequence, while the intermolecular G4 band is represented with a G4 structure formed from two different purple sequences. CSB is represented as a yellow globular protein. (**B**) Binding curves expressing the percentage of 5'-, 3'- or untailed rDNA sequences bound to increasing concentrations of CSB-HD (0 to 5 nM) in KCl or LiCl buffers. EMSAs were performed in duplicates and the K_ds were calculated using a one site specific binding equation using GraphPad Prism 9.0.1 software and 95% CI. (**C**) EMSA gel using 3'-tail c-MYC G4. (**D**) EMSA gel using untailed rDNA-1. (**E**) EMSA gel using untailed rDNA-2. (**F**) EMSA gel using untailed c-MYC. The results presented in this figure have been adapted from Liano et al. [197].

TAIL OF AT LEAST 3-5bp IS NECESSARY FOR INTERMOLECULAR rDNA G4s TO FORM

To understand the minimal length of the tail necessary for the intermolecular G4-formation, different rDNA-1 sequences with either a 5'- or a 3'- tail of increasing lengths (from 0 to 20 bp, see Table 5.7 for all the oligonucleotide sequences) were annealed in KCl or LiCl buffers and then analysed by agarose gel electrophoresis.

As expected, either a 5'- or a 3'-tail 20 bp-long increased the formation of intermolecular G4s when annealed in KCl buffers, while negligible formation of intermolecular G4s was observed in absence of the tail. Weaker intermolecular G4 formation was observed for rDNA-1 with 5'- or 3'-tail either 10 bp or 5 bp long. The 3 bp tail further weakened the intermolecular G4s when placed at the 5'-end and no intermolecular G4 formation was observed for the 3 bp tail placed at the 3'-end of rDNA-1 (Figure 3.25A and B).

Altogether these results suggested that either a 5'- or a 3'- tail of at least 3 or 5 bp is essential for intermolecular rDNA G4s to form under these experimental conditions. Moreover, the lack of CSB unwinding activity towards untailed rDNA-1 detected by gel-based resolvase assay (Figure 3.12C and D) is probably explained by the lack of an intermolecular G4-substrate for CSB to bind to rather than pointing to an essential role of the tail per se.

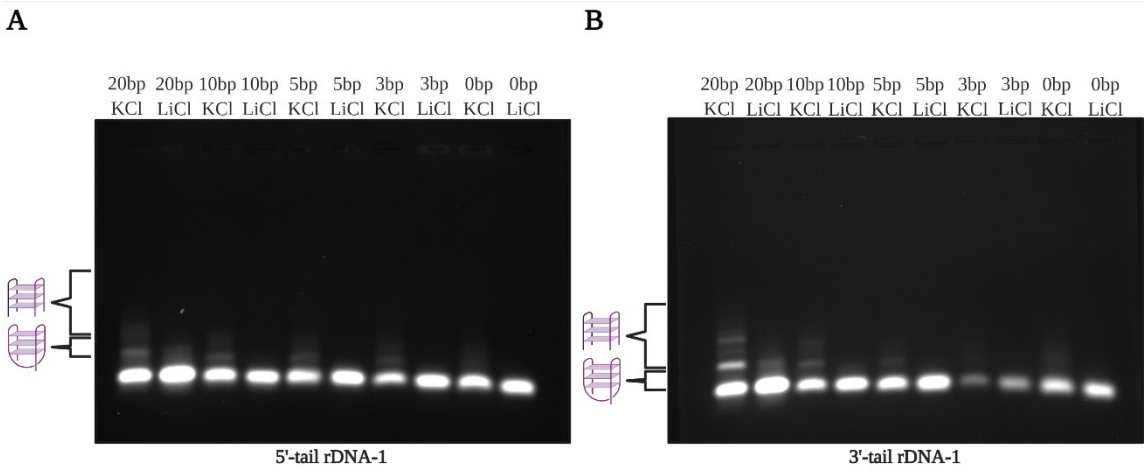

Figure 3.25. Necessary tail of at least 3–5 bp for intermolecular G4s to form. (**A**) Gel agarose showing intermolecular G4 formation from the rDNA-1 sequence annealed in either KCl or LiCl buffer that presents 5'-tail of different lengths. (**B**) Gel agarose showing intermolecular G4 formation from the rDNA-1 sequence annealed in either KCl or LiCl buffer that presents 3'-tail of different lengths. The intramolecular G4 band is represented with a G4 structure formed within the same light purple sequence, while the intermolecular G4 band is represented with a G4 structure formed from light and dark purple sequences.

3.2.11 CSB Resolves Any Intermolecular G4s

The binding analysis revealed that CSB interacts selectively and with strong picomolar affinity with the intermolecular G4s formed within rDNA sequences under KCl conditions without binding intramolecular G4s formed from either the same rDNA substrates or the panel of intramolecular G4-forming sequences tested. The selectivity of CSB for intermolecular G4s was further investigated by assessing its interaction towards a non-rDNA intermolecular G4. Specifically, a 5'-tailed *O. nova* d($G_4T_4G_4$) [232,233] telomeric sequence (see Table 5.8 for the oligonucleotide sequence used), known to form bimolecular G4s, was tested by agarose EMSAs [197]. Since d($G_4T_4G_4$) contains only two guanine-rich tracts in its sequence, this substrate could only generate bimolecular G4s or single strand DNA, and lacks the ability to form intramolecular G4s. As expected, the 5'-tail d($G_4T_4G_4$) sequence generated both slow- and fast-running bands when annealed in KCl buffer, which were assigned to the bimolecular G4 and ssDNA, respectively. Conversely, only a single fast-running band was observed in LiCl buffer, which confirmed that in the absence of K^+, the oligonucleotide was exclusively found in its ssDNA form. CSB-HD was able to interact exclusively with the slow-moving band corresponding to the bimolecular G4s formed upon K^+ stabilisation (Figure 3.26A), with a calculated K_d of 9.878 pM [2.872–23.91 pM – 95% CI] (Figure 3.26B). Consistently to what was observed for the rDNA substrate, untailed d($G_4T_4G_4$) could not form intermolecular G4s and only the ssDNA was observed for the untailed *O. nova* substrate (Figure 3.26C), which was not bound by CSB-HD (Figure 3.26B) [197].

Finally, the ability of CSB to resolve d($G_4T_4G_4$) G4s was also investigated by gel-based resolvase assay. Unsurprisingly, this assay revealed a modest increase of d($G_4T_4G_4$) G4 resolution after incubation with CSB-HD (Figure 3.26D and E). The ~11% increased resolution of d($G_4T_4G_4$) G4 agreed with the small percentage of bimolecular G4 formed within this tailed substrate, suggesting specific intermolecular G4 resolution.

Overall, these results demonstrated the ability of CSB to interact selectively with any type of intermolecular G4s besides the ones formed by rDNA substrates, suggesting that CSB's biological role could span beyond rDNA, and raising interesting possible avenues that could be investigated in the future, such as the role of CSB in resolving R-loops type intermolecular G4 that can be formed during active transcription. Moreover, the gel-based resolvase assays confirmed that the picomolar affinity interaction between CSB and intermolecular G4s is

necessary to promote the selective resolution of such intermolecular structures in an ATP-independent fashion [197].

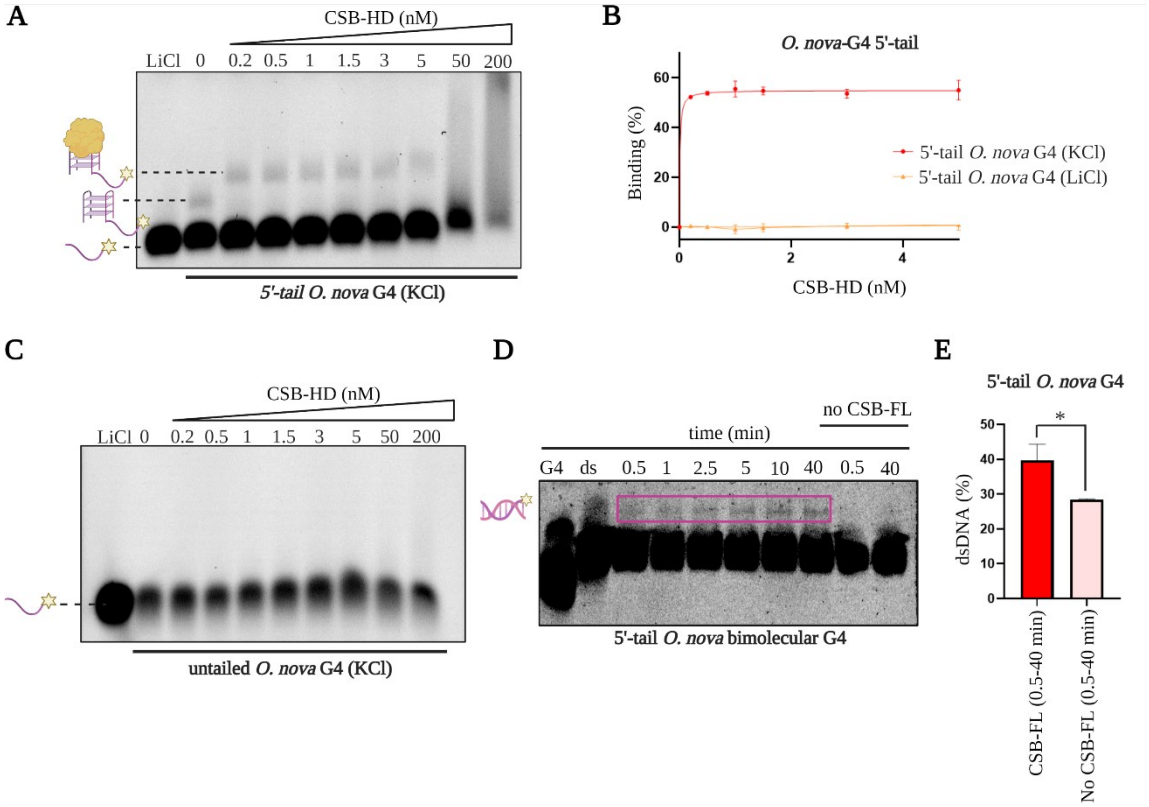

Figure 3.26. CSB binds and resolves any type of intermolecular G4s. (**A**) Agarose EMSA showing interaction between CSB-HD and bimolecular G4s formed within 5'-tailed *O. nova* d($G_4T_4G_4$) sequence in KCl conditions. The first lane is the control annealed in LiCl. The intermolecular G4 band is represented as a fluorophore-labelled G4 formed by light and dark purple sequences. CSB is represented as a yellow globular protein. The faster running ssDNA band is indicated with a light purple ss strand. (**B**) Binding curves expressing the percentage of 5'- or untailed *O. nova* d($G_4T_4G_4$) bound to increasing concentrations of CSB-HD (0 to 5 nM) in KCl buffer. EMSAs were performed in duplicates, while the K_ds were calculated using a one site specific binding equation using GraphPad Prism 9.0.1 software and 95% CI (**C**) Agarose EMSA of untailed *O. nova* d($G_4T_4G_4$) sequence in the presence of increasing concentrations of CSB-HD, in KCl conditions. (**D**) Gel-based helicase assay using 5'-tailed *O. nova* d($G_4T_4G_4$) sequence in the presence or absence of CSB-FL for increasing times (0.5 to 40 min). The purple box indicates the formation of dsDNA in the presence of CSB. (**E**) Column graph of quantified gel-based resolvase assays using 5'-tailed *O. nova* d($G_4T_4G_4$). All quantified gel-based resolvase assays were based on the average of three independent experiments. The significance was calculated based on two-tailed Student's *t*-test. Asterisks indicate statistical significance at 95% CI between the data with $*p < 0.05$. This figure was adapted from Liano et al.[197].

3.3 G4-Ligands Displace the Interaction Between CSB and Intermolecular G4s *in vitro*

Cellular and *in vivo* studies performed by Scheibye-Knudsen and colleagues proposed that treatment with PDS and CX-5461 G4-ligands can accelerate premature ageing both in cells and *C. elegans* models [168]. Interestingly, confocal imaging performed by Iyama and co-workers revealed displacement of CSB from the nucleolus of cells after treatment with CX-5461 and microirradiation[234], suggesting that transcriptional inhibition caused by CX-5461 could cause transcriptional pausing and displacement of the protein from the nucleoli. However, as discussed in Section 1.6.2, CX-5461 has been originally designed and further validated as an effective G4-ligand [235], suggesting that this molecule could also compete with CSB for the binding to intermolecular rDNA G4s in the nucleoli of cells.

To investigate whether G4-ligands could displace CSB bound to intermolecular rDNA G4s, competitive EMSAs were performed using native agarose gels. In this experiment, CSB-HD was pre-bound to the intermolecular rDNA-1 G4. The intermolecular G4-CSB-HD complex was then incubated with increasing concentrations (from 0 to 5,000 nM) of CX-5461. Subsequently, the incubated samples were analysed by agarose gel. Formation of a band running like the free intermolecular G4 was expected if the G4-ligand was displacing CSB from its intermolecular oligonucleotide substrate (Figure 3.27A).

Interestingly, replenishment of the band corresponding to the free intermolecular rDNA G4 was observed upon increasing concentration of CX-5461 (Figure 3.27B) [197]. The displacement of CSB from its intermolecular G4 substrate indicated that this G4-ligand can compete with CSB for binding to intermolecular rDNA G4s. Since previous evidence revealed that treatment with G4-ligands can accelerate premature ageing [168], the ability of CX-5461 to displace the interaction between CSB and rDNA G4s suggested that the binding of CSB to intermolecular rDNA G4s may thus be important to avoid premature ageing. Moreover, we postulate that the absence of nucleolar CSB upon treatment with CX-5461 observed by Iyama and colleagues [178] could actually be caused by the displacement of the protein from its intermolecular G4 substrate rather than a transcriptional inhibition effect. Supporting this, the same competitive EMSA was performed using PDS, which is a very well-known G4-ligand and not a transcriptional inhibitor. PDS was also able to displace CSB-HD from the interaction with the intermolecular rDNA-1 G4 tested (Figure 3.27C) [197], indicating

that both the premature ageing phenotype and nucleolus displacement of CSB upon treatment with G4-ligands could be due to the loss of intermolecular rDNA G4 interaction.

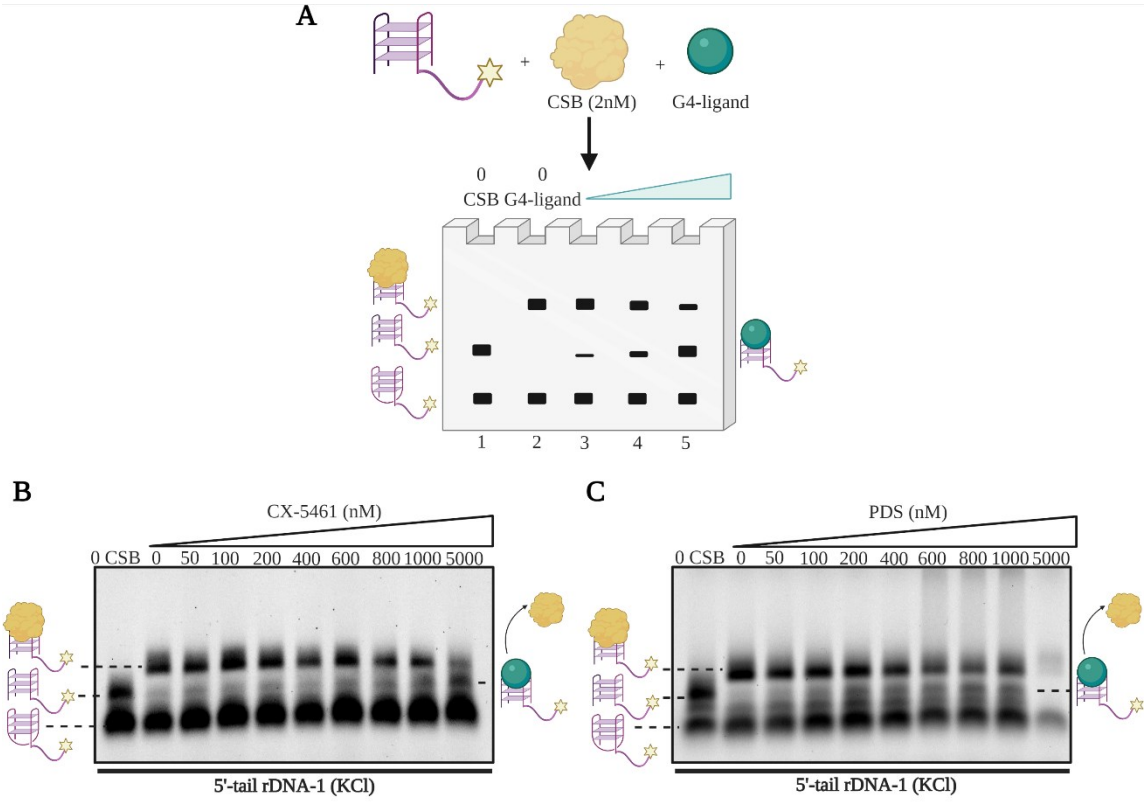

Figure 3.27. G4-ligands can displace the strong interaction between CSB-HD and intermolecular rDNA G4s. (**A**) Schematic representation of the competitive EMSA. The 5'-tailed rDNA-1 G4 was incubated with a fixed concentration of CSB-HD (2 nM). This complex was then incubated with increasing concentration of the G4-ligand. Then, the samples were separated on the agarose gel. The first lane is a control of free 5'-tail rDNA-1 showing the formation of both intramolecular and intermolecular G4. CSB-HD protein is indicated as a globular yellow protein, while the CSB displacement at increasing concentrations of the small molecule (green sphere) is represented on the right-side of the gel. Intermolecular G4 bound to the small molecule runs at the same high as the free intermolecular G4. (**B**) Competitive EMSA using CX-5461. (**C**) Competitive EMSA using PDS. The last lane (5000 nM PDS) is showing PDS precipitation on the well of the gel. The gels in this figure have been taken from Liano et al.[197].

3.4 G4-Ligands Displace the Interaction Between CSB and Intermolecular G4s *in cells*

Given that G4-ligands can displace CSB bound to intermolecular rDNA G4s *in vitro*, this phenomenon was also investigated in cells by fluorescence microscopy, using an EGFP labelled CSB construct [197].

3.4.1 CSB-HD (285–1,009) Insertion Into an EGFP-Reporter Vector

To follow the localisation of CSB in human cells, prior or after treatment with G4-ligands, we needed to generate a construct to express full-length CSB fused to an EGFP-reporter to fluorescently label the protein and track it during microscopy experiments. Generation of such a construct embedding the entire CSB-FL sequence (4,476 bp) by RF cloning was problematic, therefore the smaller helicase- "like" domain (2,175 bp) was cloned into the EGFP-reporter vector. As previously described, CSB-HD necessitates both a nuclear and nucleolar localisation sequence 1 (NLS1 and NoLS1, respectively) to display nucleolar localisation[178]. Therefore, both the CSB-FL (aa: 1–1,493, Figure 1.16A) and the CSB-HD including the NLS1 and NoLS1 (aa: 285–1,009, Figure 1.16A), were extracted from the original pFASTBac_HA-CSB-His_6 plasmid using PCR, and subsequently fused to the C-terminal of the EGFP reporter gene by RF-cloning (Figure 3.28A and B). Size comparison between the linearised original recipient pEGFP-C1 plasmid (4,731 bp) and linearised constructs obtained after RF-cloning revealed that only the CSB-HD (285–1,009) was successfully inserted into the reporter vector as the size of the original plasmid increased from 4,731 bp to 6,906 bp (Figure 3.28C). The linearisation of the plasmids obtained after RF-cloning was performed by digestion of the products using NdeI restriction enzyme, as described in Section 5.18.

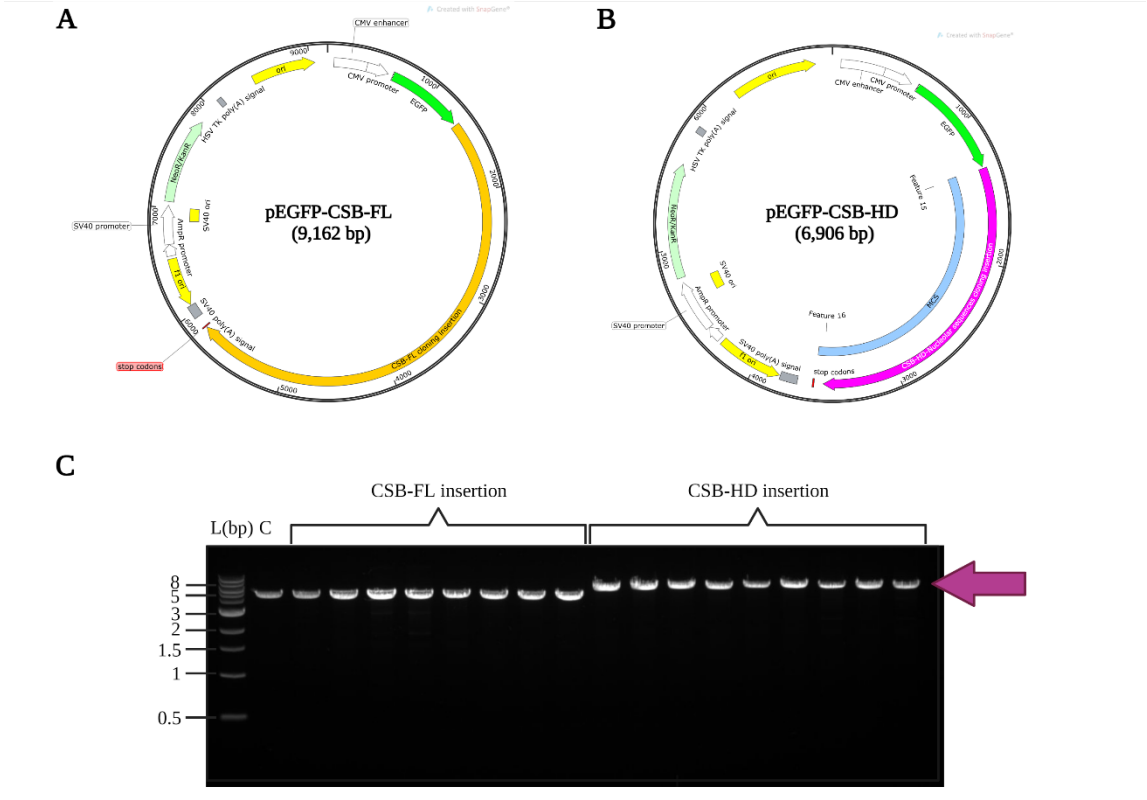

Figure 3.28. CSB-HD insertion into pEGFP-C1. (**A**) Vector map showing the pEGFP-CSB-FL construct with CSB-FL insertion (in ochre) at the C-terminal of the EGFP coding sequence (green). Vector map created with SnapGene. (**B**) Vector map showing the pEGFP-CSB-HD construct with CSB-HD (285–1,009) insertion (magenta) at the C-terminal of the EGFP coding sequence (green). Vector map created with SnapGene. (**C**) Agarose gel of the linearised vectors after RF-cloning of CSB-FL or CSB-HD (285–1,009) into the pEGFP-C1 vector. L indicates the ladder. C is the linearised pEGFP-C1. All the other different lanes indicate different clones. The CSB-FL insertion did not show any increase in the size of the original pEGFP-C1vector, indicating inefficient insertion. Conversely, the CSB-HD (285-1009) insertion showed higher MW compared to that of the original pEGFP-C1vector (magenta arrow), indicating efficient insertion.

3.4.2 CSB-HD (285–1,009) Localisation After Treatment with G4-Ligands

Once the fidelity of the newly generated pEGFP-C1-CSB-HD construct was confirmed by Sanger sequencing, this plasmid was applied to HeLa cells by transient transfection using FuGENE® HD transfection reagent (Promega) and a standard protocol. On a standard transfection protocol, cells are plated on day one and transfected after 24 h by addition of the

desired plasmid mixed with the transfection reagent directly to the cells. The transfection reagent allows the delivery of DNA to the cells with high efficiency and low toxicity by formation of a complex with the DNA that can be internalised within the cell. After 48 h of transfection, CSB-HD localisation was mainly observed in the nucleoli of HeLa cells, as expected and previously described[178] (Figure 3.29A and B). The nucleolar localisation of CSB-HD could potentially indicate the formation of intermolecular G4s within rDNA sequences in living cells that are bound tightly by the protein.

To further explore this hypothesis, HeLa cells transfected with the pEGFP-C1-CSB-HD plasmid were then treated with G4-ligands to investigate whether the application of these molecules could affect CSB localisation. Specifically, cells were treated for 24 h with either 1 μM CX-5461 or 10 μM PDS. Gratifyingly, EGFP-CSB-HD nucleolar localisation was significantly reduced after treatment with either the G4-ligands compared to that of the untreated controls (Figure 3.29A and C) [(197)]. After performing this experiment in two biological and two technical replicates, we calculated the percentage of cells showing nucleolar localisation of CSB under the different conditions tested. Statistical significance was determined using a paired two-tailed Student's *t*-test, yielding *p*-values of $p<0.0005$ for all conditions. Conversely, negligible reduction of nucleolar CSB was observed after 4 h treatment with either CX-5461 or PDS G4-ligands (Figure 3.29B and D), suggesting that the displacement of CSB required longer incubation times with these small-molecules [(197)].

The displacement of CSB from the nucleoli upon treatment with G4-ligands strongly supported that the nucleolar localisation of the protein could be, at least partially, promoted/regulated by the selective binding of CSB to intermolecular G4s formed in rDNA sequences, and this binding could be abrogated by G4-ligands as observed *in vitro*.

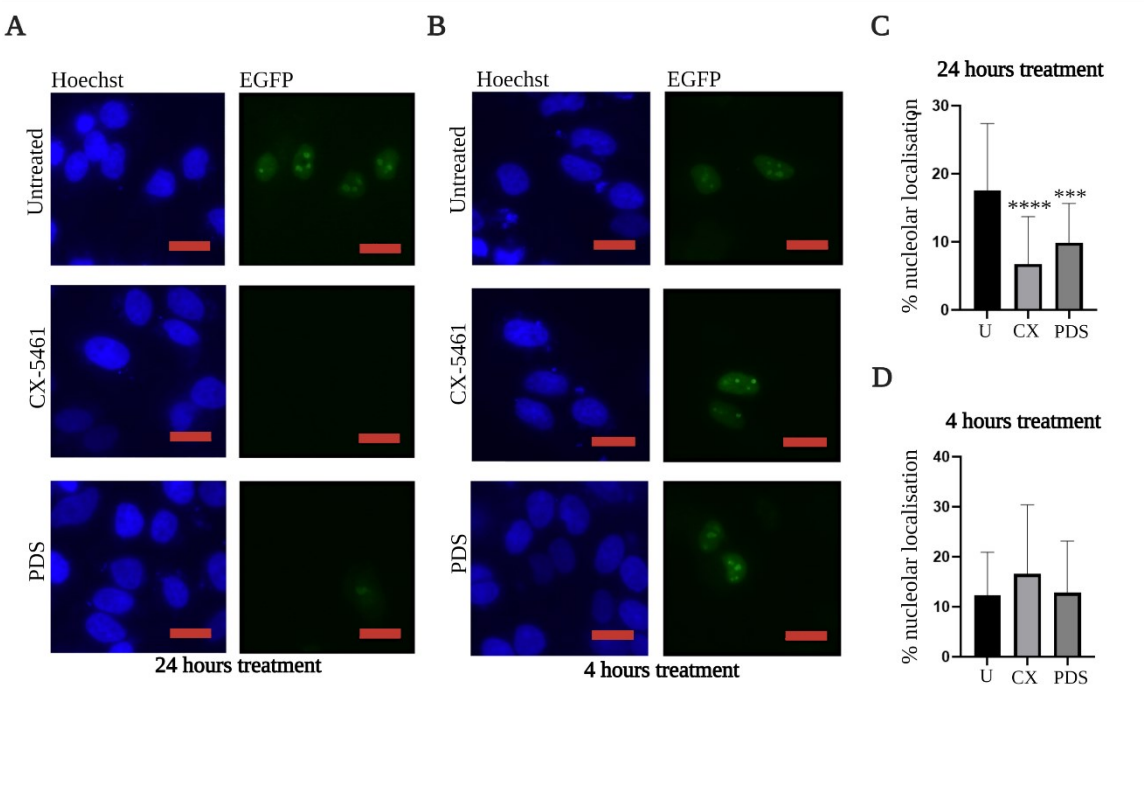

Figure 3.29. G4-ligands displace CSB from its nucleolar localisation in HeLa cells. (**A**) Localisation pattern of EGFP-CSB-HD (285–1,009) in HeLa cells without treatment with G4-ligands (Untreated) or after 24 h treatment with either CX-5461 or PDS. (**B**) Localisation pattern of EGFP-CSB-HD (285–1,009) in HeLa cells without treatment with G4-ligands (Untreated) or after 4 h treatment with either CX-5461 or PDS. Hoechst is the nuclear marker recorded at 400 nm, EGFP indicates the EGFP signal recorded at 470 nm. Red bars indicate 20 µm. (**C**) Quantification of the nucleolar localisation EGFP-CSB-HD (285–1,009) without treatment with G4-ligands (U) or after 24 h treatment with either CX-5461 (CX) or PDS. The percentage of nucleolar localisation was calculated from eight cellular images for each condition, performed in two biological and two technical replicates. Significance was calculated using a paired two-way Student's *t*-test. ***p=0.0003; ****p<0.0001. (**D**) Quantification of the nucleolar localisation of EGFP-CSB-HD (285–1,009) without treatment with G4-ligands (U) or after 4 h treatment with either CX-5461 (CX) or PDS. This figure has been adapted from Liano et al. [197].

Finally, a control transfection with the empty pEGFP-1C vector was performed to validate that treatment with G4-ligands did not affect the level of expression of the EGFP gene. After 24 h transfection with the empty pEGFP-1C vector, HeLa cells were treated with 1 µM CX-5461 for either 24- or 4- h. Negligible reduction of the general EGFP cellular signal was expected if treatments with CX-5461 did not affect the expression of the reporter gene.

Unsurprisingly, similar levels of EGFP expression were observed after either 24- or 4- h incubation with the G4-ligand (Figure 3.30A-D). This result further confirmed that the nucleolar displacement of EGFP-CSB-HD is due to the G4-binding activity of the G4-ligands that compete with CSB for the interaction with intermolecular rDNA G4s rather than their transcriptional inhibition properties [197].

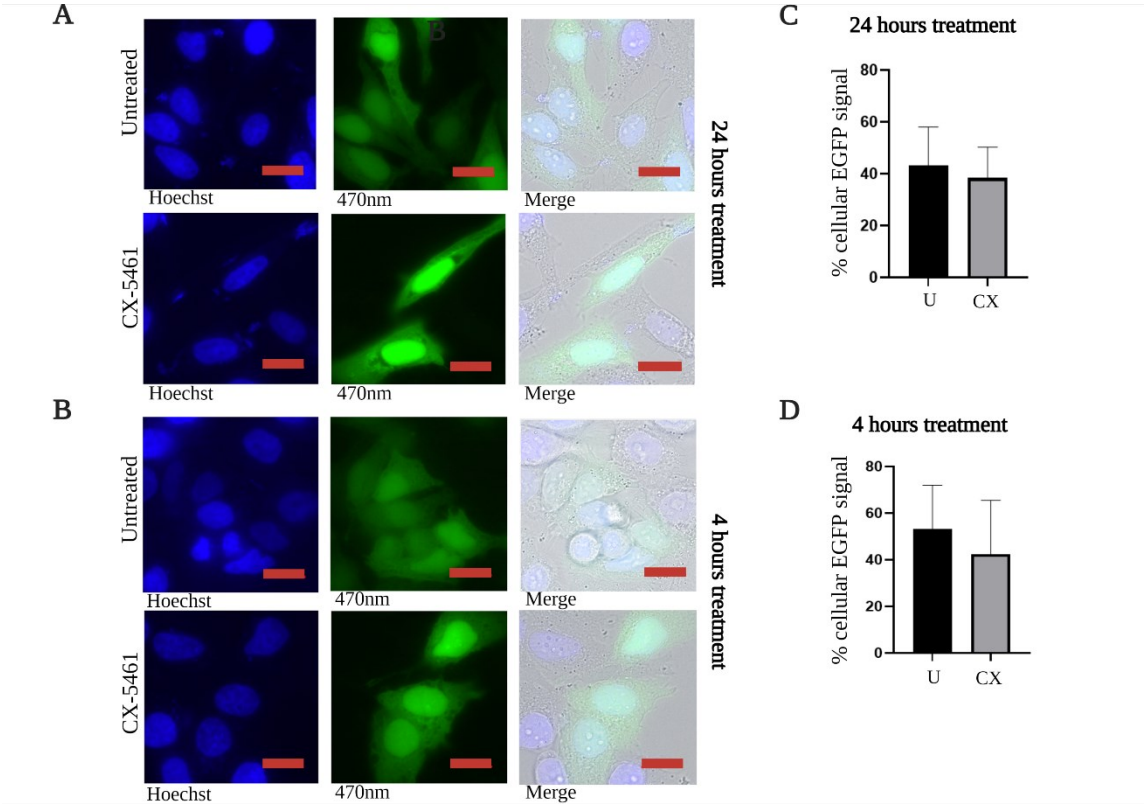

Figure 3.30. CSB displacement is not caused by transcriptional inhibition of EGFP-CSB-HD (285–1,009). (**A**) Localisation pattern of the control EGFP-1C in HeLa cells without treatment with G4-ligands (Untreated) or after 24 h treatment with CX-546. (**B**) Localisation pattern of the control EGFP-1C in HeLa cells without treatment with G4-ligands (Untreated) or after 4 h treatment with CX-5461. (**C**) Quantification of the percentage of cells presenting cellular EGFP-1C signal without treatment with CX-5461 (U) or after 24 h treatment with CX-5461 (CX). (**D**) Quantification of the cellular EGFP-1C signal without treatment with G4-ligands (U) or after 4 h treatment with CX-5461 (CX). The percentage of cells presenting EGFP-1C signal in the presence or absence of the ligand was calculated from eight different cellular images for each condition performed in two biological and two technical replicates. This figure has been taken and adapted from Liano et al.[197].

3.5 Nucleolar Localisation of CSB Prevents Nucleolar BG4 Staining

The observation that G4-ligands could displace CSB from the nucleoli of HeLa cells prompted to validate more directly the interaction between CSB and the intermolecular rDNA G4s in the nucleoli of human cells. To achieve this, it was necessary to visualise G4s' distribution either in CSB positive human cells (HeLa) or in CSB deficient cells derived from a Cockayne Syndrome patient (CS1AN) [236] by means of the selective G4 antibody BG4 [62].

3.5.1 Production and Validation of the Synthetic BG4 Antibody

Production of the G4 specific BG4 antibody was necessary to perform immunofluorescence experiments. Therefore, the plasmid containing the BG4 coding sequence (Figure 3.31A) was transformed into *E. coli* cells for the expression of the protein from bacteria (the pSANG10-3F-BG4 plasmid was kindly donated by S. Balasubramanian's group, University of Cambridge). The cells expressing BG4 were then lysed and spun down by centrifugation to separate the cellular debris. The supernatant was purified through multiple chromatography steps (see Section 5.23). As the BG4 construct contains a His_6-tag, a first his-tag affinity purification was performed using 10 ml of Cobalt resin. Samples from all the lysis and purification steps were collected and separated by 12% SDS-gel, which showed clear formation of a ~30 kDa band corresponding to the MW of BG4 (Figure 3.31B). The elution obtained after the first purification step was further subjected to chromatography using a 5 ml heparin column. The elution profile obtained after the heparin purification revealed two main peaks eluting approximately 500–600 mM KCl (Figure 3.31C). The separation of these peaks by SDS-PAGE revealed the presence of BG4 within the two peaks (Figure 3.31D). Fractions from C2 to D6 were combined and concentrated prior to further SEC purification using a Superdex 75 column. The elution profile of SEC revealed the presence of a main peak eluting at 10.78 ml, and the SDS-gel confirmed the presence of BG4 antibody in this peak (Figure 3.32A and B). Fractions from B5 to B8 were combined and concentrated to 30 mg/ml and stored at -20 °C.

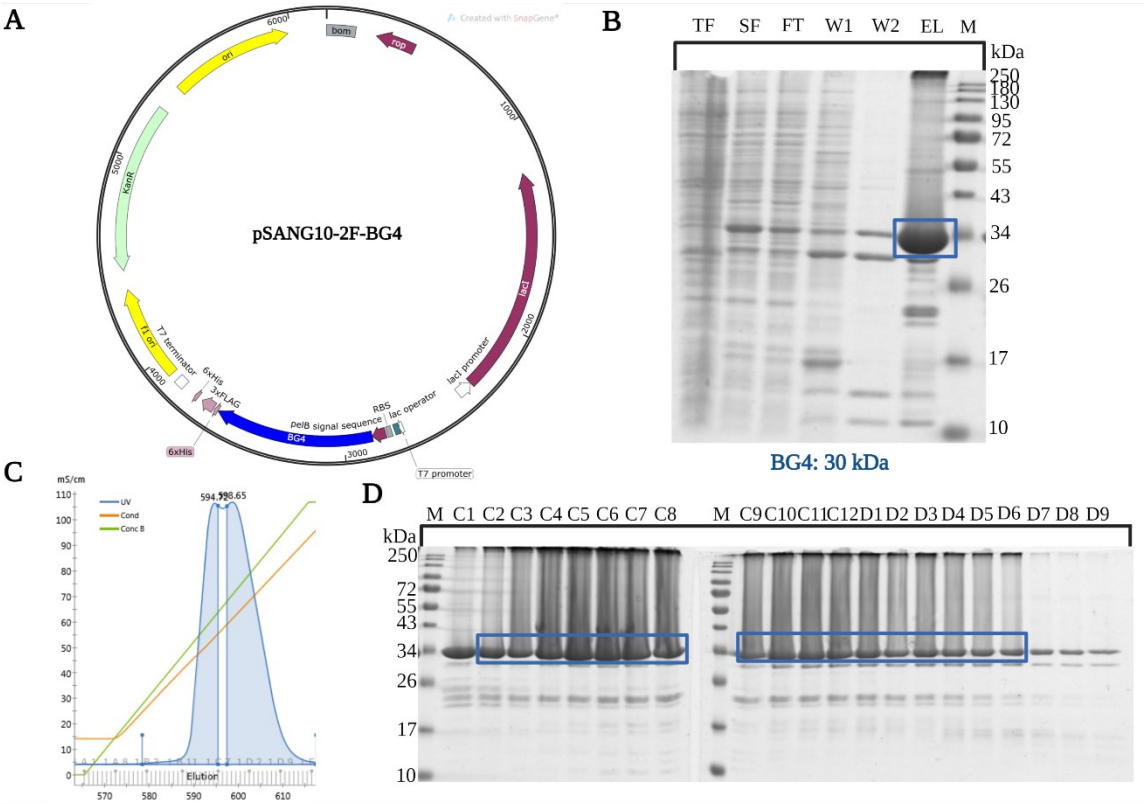

Figure 3.31. BG4 expression and purification. (**A**) Vector map of the pSANG10-2F-BG4 containing the BG4 coding sequence (blue), a pelB sequence (dark purple at the N- terminal of BG4) that is necessary for its cytoplasmatic localisation and a 3X FLAG-His_6-tag at its C-terminal (light purple). Vector map created with SnapGene. (**B**) 12% SDS-PAGE of the samples obtained from cell lysis and BG4 his- affinity purification using Cobalt resin. M is the marker, TF: total fraction, SF: soluble fraction, FT: flow through, W1-W2: washes and EL: elution. To load the gel 5 µl of TF, SF, and FT and 10µl of W1, W2, and EL samples were used. The BG4 band around 30 kDa is indicated with a blue box. (**C**) Elution profile obtained from heparin purification of BG4 (**D**) 12% SDS-PAGE of selected fractions obtained from the heparin purification. The fractions C2-D6 containing BG4 are indicated with a blue box.

To confirm the ability of the purified BG4 antibody to specifically interact with G4 structures, ELISAs experiments were performed using the same oligonucleotide sequences used in the BG4 seminal paper [62]. Briefly, either a biotinylated c-MYC G4-forming sequence or a biotinylated ssDNA sequence unable to form G4 was immobilised onto a streptavidin-coated surface (see Table 5.9 for the oligonucleotide sequences). The surface was subsequently incubated with serial dilutions of BG4, which was expected to bind only to the c-MYC G4. After removal of the unbound BG4, the surface was incubated with a secondary anti-FLAG antibody conjugated with a TMB substrate, which produces a water-soluble blue reaction

product. Stopping the reaction with the addition of sulfuric acid changed the colour to yellow with an absorbance peak at 450 nm. Its intensity was measured using a plate reader (a more detailed overview about the ELISA technique is reported in Section 2.2) and was indicative of binding between BG4 and either c-MYC or ssDNA. As previously described, ELISAs indicated very high affinity of BG4 for c-MYC G4 with a calculated K_d of 0.5 nM and negligible affinity for ssDNA (Figure 3.32C), which aligns well with the affinity values reported by G. Biffi and co-workers [62]. The ELISA quantification shown in Figure 3.32C was generated by Jenna Robinson (a second-year PhD student in M. Di Antonio's group) who supported the optimisation of the ELISA's protocol for BG4 validation.

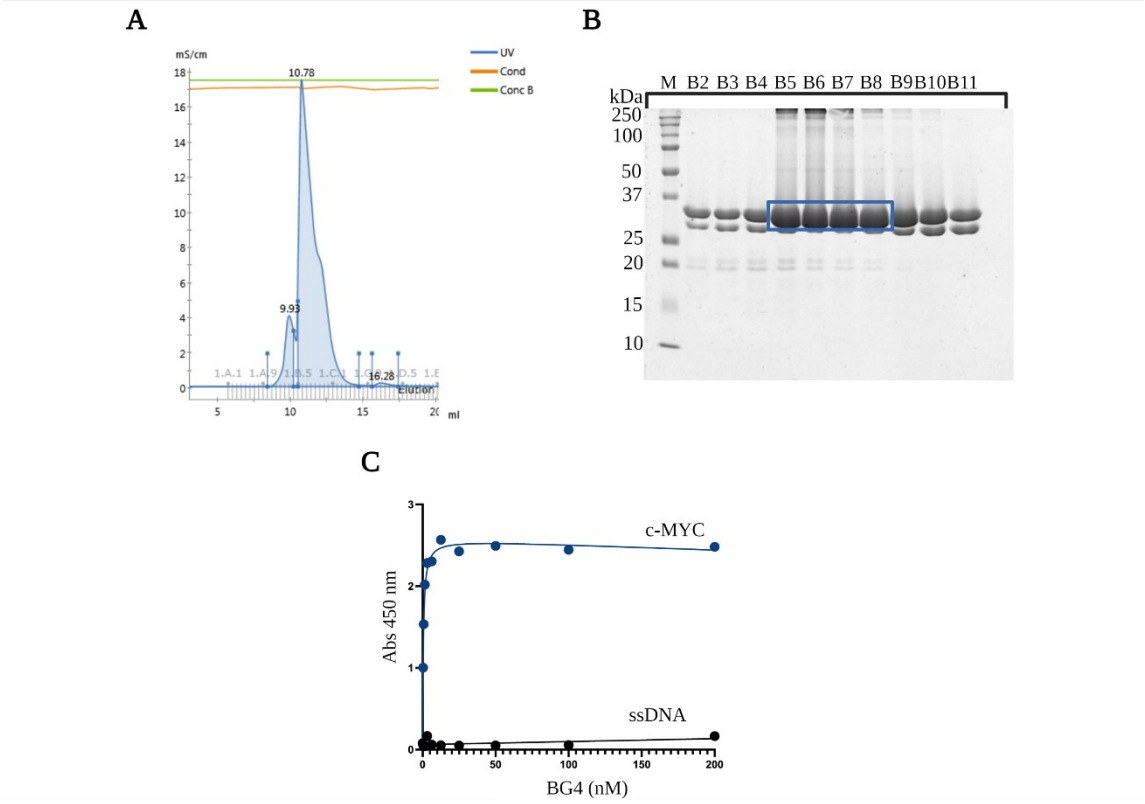

Figure 3.32. Production and validation of BG4 antibody. (**A**) Elution profile obtained after BG4 purification using S75 column. (**B**) SDS-gel of the eluted fractions presenting the peak. Fractions B5-B8 containing BG4 (blue box) were collected. (**C**) Quantification of ELISA performed with either biotinylated c-MYC G4 or biotinylated ssDNA sequence in the presence of increasing concentrations of BG4. Quantification performed using a one site specific binding equation using GraphPad Prism 9.0.1 software.

3.5.2 CSB Localisation in the Nucleolus Prevents BG4 Staining

To validate the direct binding of CSB to the G4-formed in nucleoli of cells, BG4 immunostaining [62,71] was performed in both non-transfected CS1AN cells that are CSB-deficient and in CS1AN cells that were previously transfected with functional EGFP-CSB-HD for 48 h. If CSB interacts with the intermolecular G4s present in the nucleoli, a reduction of the BG4 nucleolar staining was expected between cells that re-express CSB after transfection and wild-type CS1AN cells. Indeed, a clear reduction of BG4 staining in CS1AN cells that re-express functional CSB would be indicative of direct competition between CSB and BG4 for binding to rDNA G4s.

CS1AN cells transfected with EGFP-CSB-HD construct re-express CSB and fluorescence microscopy revealed the accumulation of CSB in the nucleoli (Figure 3.33A), similarly to what has been observed for HeLa cells transfected with the same construct (Figure 3.29A and B) [197]. Subsequently, the transfected CS1AN cells were fixed and incubated with BG4 antibody, which was expected to selectively bind the G4 structures present within the cells. The nucleolar accumulation of CSB was associated with a clear lack of BG4 staining in the nucleoli of the transfected CS1AN cells (Figure 3.33A). To understand whether the lack of nucleolar BG4 was caused by the presence of CSB in these organelles, the same immunofluorescence analysis was performed both in non-transfected CS1AN cells that do not express functional CSB, and in HeLa cells that express endogenous CSB, as control. Gratifyingly, non-transfected CS1AN cells displayed clear nucleolar BG4 staining. A statistically significant increase in the fraction of cells that lack BG4 staining was observed in the nucleoli of CS1AN cells transfected with EGFP-CSB-HD plasmid (~50% nucleoli that lack BG4 staining) compared to those non-transfected (~14% nucleoli that lack BG4 staining) (Figure 3.33 B and C) [197]. The ability of the CSB protein to compete with BG4 for the binding to nucleolar G4s was further suggested by a similar lack of nucleolar BG4 staining observed between HeLa cells that express endogenous CSB and the CS1AN cells that re-express CSB (Figure 3.33B and C) [197]. Indeed, a statistically significant increase in the fraction of cells that lack BG4 staining was calculated between the nucleoli of HeLa cells (~30% nucleoli that lack BG4) compared to non-transfected CS1AN (~14% nucleoli that lack BG4) [197]. These experiments were performed in two biological and two technical replicates for each condition and eight images were taken from each replicate. The percentage of

nucleoli that lack BG4 staining was determined from the ratio between the number of nucleoli without BG4 staining and the total number of nucleoli counted within a cell. Quantification of the difference between the percentage of cells lacking nucleolar BG4 was calculated using a two-tailed Student's *t*-test.

Altogether, this suggested that when CSB is either re-expressed in CS1AN CSB-deficient cells or is already active in HeLa cells, the protein binds the G4s formed within nucleoli with high affinity and prevents the BG4 staining of these organelles. Nucleoli are dense organelles with very high concentration of rDNA [237]. Consequently, the G-rich rDNA must be highly compacted within the nucleoli, making the long-range interaction rDNA sequences possible. The picomolar affinity of CSB for intermolecular rDNA G4s observed *in vitro*, strongly supports the potential for CSB to bind intermolecular G4s formed from long-range interactions within the rDNA sequences in the nucleoli of living cells, with endogenous CSB potentially involved in regulating the homeostasis of these structures.

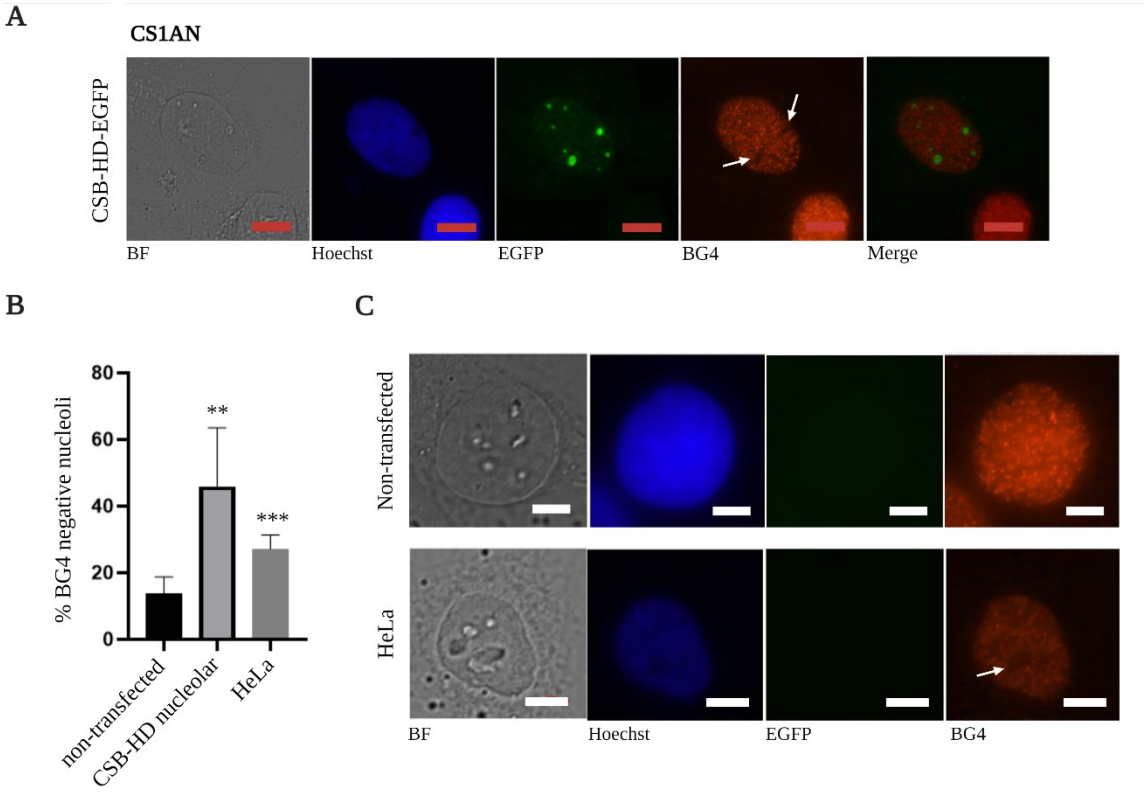

Figure 3.33. Presence of active CSB reduces BG4 staining within the nucleoli of human cells. (**A**) Localisation pattern of EGFP-CSB-HD (285–1,009) after 48 h transiently transfection in CS1AN cells. Nucleolar localisation of CSB is confirmed by the green EGFP staining of the nucleoli (EGFP signal recorded at 470 nm). BG4 localisation in CS1AN re-expressing CSB is visualised by immunostaining and revealed the presence of black nucleoli only in cells positively transfected with CSB-HD-EGFP, suggesting inefficient BG4 staining in nucleoli occupied by CSB (white arrows). (**B**) Quantification of the absence of nucleolar BG4 staining expressed by the different percentage of BG4 negative nucleoli between CS1AN transfected with EGFP-CSB-HD (285–1,009) plasmid or HeLa cells and non-transfected CS1AN cells. Significant difference calculated using two-way Student's *t*-test. ***p*=0.003; ****p*=0.0002. (**C**) Immunostaining using BG4 reveals abundant BG4 staining in the nucleoli of non-transfected CS1AN cells (CSB negative, first raw), while lack of nucleolar BG4 staining is observed in HeLa cells (CSB positive, second raw), as previously reported[62]. BF indicates the brightfield, while Hoechst is the nuclear marker recorded at 400 nm. BG4 signal is recorded at 590 nm. Merge is the overlap between 470 nm and 590 nm signals. Red and white bars indicate 10 µm and 5 µm, respectively. This figure has been taken and adapted from Liano et al.[(197)].

Chapter 4

4 Conclusions and Future Work

G4s are non-canonical nucleic acid secondary structures that arise from single stranded G-rich motifs [15]. G4-folding sequences can be computationally predicted or mapped within different genomes using G4-Seq and ChIP-Seq [72]. Visualisation of G4 structures in fixed cells was achieved using the synthetic G4-specific antibody BG4 [62]. Probes for both fluorescence lifetime imaging microscopy and live-cell single-molecule fluorescence imaging of G4s allowed to directly visualise these structures in real time, further proving the formation of G4s within living cells [63,65]. G4s are widely distributed across genomes, with enrichment of these structures at telomeres, promoters, TSS, and nucleosome depleting regions [71]. Experimental evidence demonstrated the ability of these structures to regulate essential cellular pathways such as replication, transcription, translation, and genome maintenance. Because their wide roles in cells, G4s require precise regulation, which is mainly maintained by chaperones and G4-helicases [77]. Lack of activity of G4-helicases or treatment with G4-stabilising ligands can alter the regulation of G4-formation, which could lead to cancer and accelerated ageing. Although the potential roles of G4s in cancer progression have been investigated, the impact of G4s on ageing biology has been understudied and thus requires further research.

Recently, RNA-Seq experiments in cells that lack functional CSB revealed increased transcriptional pausing at rDNA G4-forming sequences, which has been linked to a series of biological responses typical of ageing [168]. It was also found that CSB can resolve rDNA G4s, suggesting a close correlation between functional mutations of CSB, rDNA G4s, and premature ageing [168]. On the other hand, CSB belongs to a family of chromatin remodelling proteins, and it has not been classified as a canonical helicase. This aspect is particularly relevant by close inspection of CSB G4-resolvase activity that is very limited and never leads to full G4-unwinding, thus indicating a more complex mechanism behind CSB functional mutations and G4-homeostasis behind the premature ageing observed in Cockayne Syndrome patients.

In this PhD Thesis, CSB and its interaction with a panel of different G4-folding sequences were fully investigated biochemically. Firstly, both the full-length CSB protein and its helicase- “like” domain were expressed and purified. Once the proteins were produced, their

activity as G4-resolvase was confirmed by gel-based assays. Interestingly, both the CSB-FL and CSB-HD displayed an incomplete G4-resolvase activity, which was limited to either 5'-tailed or 3'-tailed rDNA G4s and not observed for canonical DNA G4s such as c-MYC, c-KIT1, hTELO, and HRAS [197]. The limited ability of CSB to resolve G4s was incongruous with the strong G4-mediated transcriptional stalling observed in CSB impaired cells [168] and with the premature ageing phenotype typical of Cockayne Syndrome patients [186]. Therefore, further characterisations of the ability of CSB to bind to G4 structures were performed by EMSAs.

The agarose gel used for this investigation revealed that rDNA G4-forming sequences were able to fold into multiple conformations when annealed in KCl buffer. These were identified as slow- and fast-running bands on the agarose gels. Further characterisation revealed that rDNA sequences could form mainly intramolecular G4s (fast-running bands) and a small sub-population of intermolecular (bi- or tetrameric) G4 structures (slow-running bands) when annealed using the G4-stabilising K^+ [197]. Interestingly, intermolecular rDNA G4s were formed only when the G4-folding sequence was decorated with a ssDNA tail of at least 3-5 bp either at the 5'- or 3'-end. CSB-HD showed an astonishing picomolar affinity for intermolecular rDNA G4s, whilst negligible binding was observed against the intramolecular G4s formed by the same rDNA sequence or against a panel of different intramolecular G4s. Furthermore, CSB-HD was also able to bind and resolve the non-rDNA bimolecular G4s originated from the telomeric *O. nova* sequence. Moreover, CSB failed to interact with ssDNA, corroborating a selective and high binding affinity of the protein to intermolecular G4s.

Altogether, our results suggest that the partial rDNA G4 resolution observed by gel-based resolvase assays is associated to the selective resolution of the intermolecular G4s bound by CSB [197]. Higher concentrations of CSB (>50 nM) were required to observe binding against both intramolecular G4s and ssDNA, suggesting that the interaction between CSB and intramolecular G4 was not specific over ssDNA [197]. The strong picomolar affinity of CSB for intermolecular G4s, hinted, for the first time, at the possibility that rDNA sequences could adopt multimeric G4-conformations in cells as well as they do *in vitro*, and that CSB could be the dedicated protein involved in regulating the formation of these structures. Supporting this, previous reports identified nucleolar localisation of CSB [178]. Therefore, we speculated that CSB might be directed to the nucleoli driven by strong affinity against intermolecular G4s. Nucleoli are very dense and rDNA enriched organelles within the nucleus of cells [237], further

indicating a good probability for distal rDNA sequences to interact with each other in the space to form intermolecular G4s.

Previous experiments on *C. elegans* models revealed that treatment with G4-stabilising ligands such as PDS [238] and CX-5461 [235] accelerates ageing of the worms. Both *in vitro* and cellular studies were performed to understand whether premature ageing could be triggered by ligand-displacement of CSB from its intermolecular rDNA G4s. Interestingly, this research combination demonstrated that both CX-5461 and PDS can compete with CSB for binding to intermolecular G4s. Treatment of HeLa cells with these small molecules showed significant displacement of CSB from the nucleoli, supporting a model where the interaction between CSB and intermolecular rDNA G4s might be relevant in cells and of importance to the premature ageing phenotypes observed in CSB impaired cells [197].

Finally, the interaction between CSB and rDNA G4s within nucleoli was further validated by immunofluorescence with BG4 antibody in Cockayne Syndrome cells (CS1AN). Re-expression of CSB in CS1AN cells triggered nucleolar localisation of CSB and was associated with the suppression of BG4-staining in the nucleoli. These results provided the first evidence for the potential biological function of intermolecular G4s, which could be of relevance for the maintenance of cellular homeostasis and to prevent premature ageing (Figure 4.1) [197].

Future work in this space will be aimed at understanding the molecular mechanism behind the selectivity of CSB for intermolecular G4s from a structural perspective. Therefore, obtaining a high-resolution structure of CSB bound to intermolecular rDNA G4s from either a crystal (X-ray) or using cryo-EM, would be pivotal to understanding how CSB interacts and resolves G4-structures. This can be further leveraged to design small-molecule ligands that can mimic such interactions and restore intermolecular G4s homeostasis lost upon functional mutation of CSB. Further optimisation of G4-destabilisers, such as a recently reported phenylpyrrolocytosine (PhpC)-based G-clamp analog, opens the possibility to design G4-disrupting small molecules that could be specifically driven at the G4 sites that are misregulated due to the absence of CSB [239]. Moreover, a deeper investigation of the dual role between Nucleolin and CSB will be useful to unravel the mechanisms that regulate G4 structures within nucleoli. For instance, a deficiency of CSB strongly suppresses the binding of Nucleolin to the H8 coding region of rDNA, suggesting that CSB, together with Nucleolin participates in the process of transcription elongation. Nucleolin is known as a G4-stabiliser,

proposing an opposite and highly regulated control of the CSB-Nucleolin activity in modulating rDNA G4 formation [194]. ChIP-seq analysis of CSB will also reveal the interactors of the protein in the nucleoli and shed light on its activity in maintaining the intermolecular rDNA G4 homeostasis. Finally, the long-range rDNA G4 interactions could be further detected by a high-throughput HiC-sequencing technology that allows the exploration of the biophysical proprieties of the chromatin [240]. Hence, HiC-sequencing could provide a picture of the spatially interacting rDNA sequences both in normal conditions and in prematurely aged cells, revealing the effects caused by the lack of CSB in the rDNA landscape of Cockayne Syndrome patients.

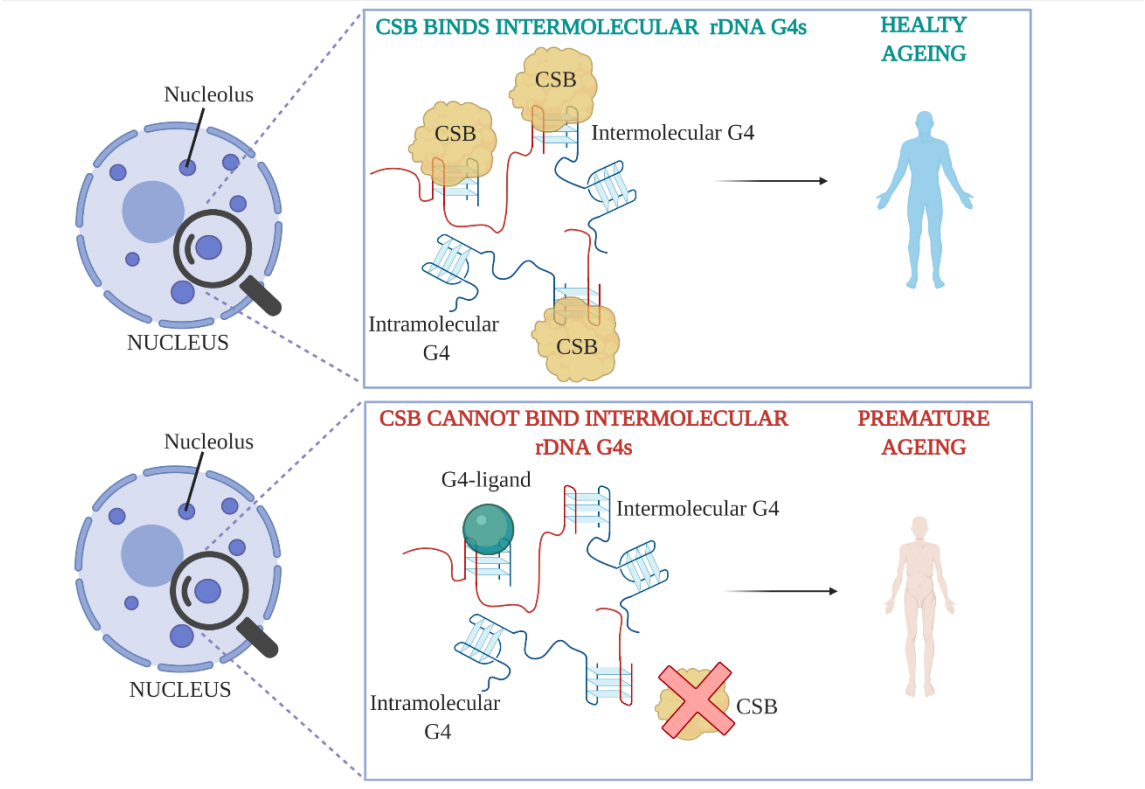

Figure 4.1. Functional CSB binds rDNA intermolecular G4s in the nucleoli promoting healthy ageing. When CSB is not functional or after the displacement by G4-ligands, rDNA intermolecular G4s accumulate triggering premature ageing. This scheme has been redrawn from Liano et al. [197].

Chapter 5

5 Material and Methods

5.1 CSB-FL Baculovirus Expression

The pFASTBac-HA-his$_6$ vector system was used to overexpress the CSB-FL (UniProtKB: Q03468) protein in a baculovirus system. This construct was kindly donated by the Scheibye-Knudsen group (University of Copenhagen) and was expressed in *Spodoptera frugiperda (Sf9)* insect cells [241] following previously reported protocols [242,243]. The steps required for CSB-FL expression are described below.

5.1.1 Transposition

To generate recombinant CSB-FL protein, the pFASTBac-HA-CSB-His$_6$ vector was transposed into a baculovirus shuttle vector, called bacmid [215].

Briefly, 50 µl of DH10Bac™ cells were transformed via a heat shock using a 5 µl aliquot of pFASTBac-HA-CSB-His$_6$ plasmid (between 350-600 ng/ul). Specifically, after the addition of the plasmid, cells were incubated on ice for 30 min following heat shock for 45 seconds (sec) at 42 °C, without shaking. Then, the cells were incubated on ice for 2 min prior to the addition of 900 µl of recovery SOC medium. Transformed cells were grown for 5 h at 37 °C before plating a 30 µl aliquot on a GTK IPTG XGAL LB-agar plate and incubated at 37 °C for 3 days. Positive white cells were picked and grown overnight (O/N) at 37 °C in LB-GKT medium (7 µg/ml gentamycin, 10 µg/ml tetracycline, and 50 µg/ml kanamycin). Finally, the bacmids were purified using a Midi-prep kit (QIAGEN®) following the supplier's protocol. The concentration of each bacmid was detected by Nanodrop (Thermo Scientific) set to read DNA at 260 nm. Precisely, CSB (1): 4072 ng/µl and CSB (2): 106.65 ng/µl.

Recipe for GTK IPTG XGAL LB-agar plate:

- Gentamycin (10 mg/ml) 266 ul to 380 ml (7 µg/ml)
- Tetracyclin (100 mg/ml) 30.4 µl to 380 ml (10 µg/ml)
- Kanamycin (25 mg/ml) 760 µl to 380 ml (50 µg/ml)
- IPTG (0.5 M) 400 µl to 380 ml (0.5 mM)
- XGAL (40 mg/ml) 400 µl to 380 ml (42 µg/ml)

5.1.2 *Sf9* Cell Culture

Suspension growths of *(Sf9)* insect cells were grown at a cell density of 1.5×10^6–2.5×10^6 cells/ml in a complete medium (Insect-XPRESS^{TM} Protein-free Insect Cell Medium with L-Glutamine (Lonza) supplemented with 5 ml penicillin-streptomycin solution (5,000 U/ml, Gibco)) at 27 °C, 130 rpm, humidified incubator, no CO_2. Cells were divided after three days and passaged every four days to a cell density of 0.5×10^6 cells/ml.

5.1.3 Transfection *(TransIT-Insect Transfection Reagent-Mirus)*

To achieve optimal transfection, 95%–98% cell viability is required.

Cells were diluted in complete medium to a cell density of 8×10^5 cells/ml. A volume of 176 µl of cells was added to each well of a 6-well plate containing 2.5 ml of medium/well. After a gentle agitation of the plate, cells were incubated for 26 min at 27 °C, without shaking to allow them to settle. Meanwhile, the transfection reaction was prepared by adding either 0.6 µl CSB (1) or 23.4 µl CSB (2) bacmid and 5 µl of *Trans*IT-Insect Transfection reagent into 250 µl of media. The mixtures were incubated at room temperature (RT) for 25 min following drop-wise addition to the cells in the 6-well plate. The plate was incubated in a humidified box at 27 °C, without shaking for 72 h.

5.1.4 Isolation of Primary Virus (P1)

The primary virus (P1) is expected to be released after 72 h incubation from the initial transfection. Therefore, cells were observed under the microscope to identify signs of viral infection (e.g. increased nucleus and cellular diameter, vesicular appearance within the cells, detachment of the cells from the surface, and lower number of cells compared to the control). Once signs of infection were observed, P1-1 and P1-2 viruses (2.5 ml) were collected in four microcentrifuge tubes and supplemented with cold 5% fetal bovine serum (FBS).

5.1.5 Dot Blot

Prior to the expansion of the P1 viruses, it was necessary to confirm their virulence by dot blot by means of their concentration within the sample.

A 2 µl of an undiluted known high titer (positive control) baculovirus and 2 µl of undiluted test samples were drop-wise added to a nitrocellulose membrane and left 10 min to completely dry. The membrane was then treated with blocking solution (5% milk in PBST (phosphate-buffered saline (PBS) + 1% Tween-20)) for 15 min at RT under gentle shaking. The membrane was incubated with a mouse anti-gp64 antibody that recognises a gp64 baculoviral protein (eBioscience #14-6995) diluted 1:1,000 in blocking solution for 1 h at RT, with gentle shaking. Then, the membrane was washed twice with PBST (5 min each wash) with gentle shaking at RT. The membrane was then incubated with anti-mouse:HRP (Pierce™, Thermo Scientific #1858413) secondary antibody diluted 1:5,000 in blocking solution for 1 h, at RT, under gentle shaking. The membrane was washed again three times in PBST (5 min each wash) with gentle shaking at RT. Finally, the membrane was developed with horseradish peroxidase (HRP) substrate for enhanced chemiluminescence (ECL) (Pierce™, Thermo Scientific) and visualised using Image Quant LAS 4000 (Cytiva). The concentration and required volume of virus for amplification/infection of 50 ml of cells were roughly determined by comparing the concentration of the tested sample to the positive control. Typically, a volume of 100–300 µl of P1 is used for highly concentrated viruses.

5.1.6 Virus Amplification

Sf9 cells (50 ml) were seeded to a cell density of 7.5×10^5 and incubated at 27 °C for 24 h and 130 rpm. Either 100 µl, 200 µl, 400 µl or 800 µl of P1-2 virus was added in 50 ml of baculovirus cells and incubated at 27 °C, 130 rpm, for 48–72 h. The cells were then harvested at 1,000 rpm, RT, for 10 min to collect the supernatant (50 ml) containing the amplified secondary virus (P2). Finally, 5% of FBS was added to P2 and stored at 4 °C.

As previously described for P1, a dot blot test was performed with P2, to confirm the presence of the virus in the cellular supernatant and evaluate its concentration.

5.1.7 Small Scale Protein Expression

Sf9 cells at a cell density of 0.75×10^6 were seeded into one flask containing 50 ml of complete media and incubated at 27 °C and 130 rpm for 24 h. After incubation, 800 µl P2-6 was added to the growth and incubated at 27 °C, 130 rpm for three days.

5.1.8 Large Scale Protein Expression

Sf9 cells at a cell density of 0.75×10^6 were seeded into four flasks containing 500 ml complete media each (2 l total volume), and incubated at 27 °C, 130 rpm for 24 h. A volume of 7.5 ml CSB (2) P2 was added into each of the 500 ml aliquots of the growth and incubated at 27 °C, 130 rpm for three days.

5.2 CSB-FL Protein Extraction and Purification

5.2.1 Cell Lysis

Three days post infection, either 50 ml or 2l of *Sf9* cells were collected and harvested by centrifugation at 1,000 xg, at 4 °C for 10 min. The supernatant was removed, and the cellular pellets were washed twice with ice-cold PBS. The pellets were then resuspended into eight (packed cells) volumes of ice-cold Lysis buffer and protease inhibitors (25 mM Tris-HCl pH 9.0, 1 mM EDTA, 10% glycerol, 1% Triton-X, 300 mM KCl, 1 mM TCEP, 0.1 mM PMSF, and 0.2 µM chemostatin, leupeptin, antipapain, and pepstatin A) supplemented with 5 mM $MgCl_2$ and 10 µl 5KU Benzonase nuclease (Merck Millipore). The suspension was gently stirred for 30 min at 4 °C prior to completing lysis with a gentle sonication on ice for 15 min, 20% amplitude, 2 sec on, 58 sec off (sonicator from Thermo Fisher Scientific). A 5 µl aliquot of the total lysate (total fraction: TF) was collected for SDS-PAGE gel, diluted in 15 µl milliQ water and stored at 4 °C. Once the cells were broken, the lysate was clarified by centrifugation at 4 °C, 12,000 xg, for 30–40 min, to remove all the cellular debris. Then, 10 µl of the clarified supernatant (soluble fraction: SF) was collected for SDS-PAGE gel and stored at 4 °C.

5.2.2 His-tag Affinity Purification Using Nickel Resin

The clarified supernatant obtained from cellular lysis was incubated with either 1 ml (small-scale expression) or 10 ml (large-scale expression) of free HisPur Nickel-NTA Resin (Thermo Fisher Scientific) at 4 °C, for 30 min under gentle shacking. Then, all the unbound material was let to pass through the resin by using a gravitational column (flow through: FT). A 10 µl volume of FT was collected for SDS-PAGE gel and stored at 4 °C.

To remove unspecific interactors, the resin was washed five times with two column volumes (CV) of Wash buffer (25 mM HEPES pH 7.9, 10% glycerol, 0.01% Triton-X, 300 mM KCl, 20 mM imidazole, 1 mM TCEP, 0.1 mM PMSF) supplemented with a low percentage of imidazole. After the washing steps (W), 10 µl of the sample from each W was collected for SDS-PAGE gel and stored at 4 °C.

The column was closed to avoid loss of material, and the resin was incubated with 10CV of Elution buffer containing a high concentration of imidazole (25 mM HEPES pH 7.9, 10%

glycerol, 0.01% Triton-X, 300 mM KCl, 250 mM imidazole, 1 mM TCEP and 0.1mM PMSF) at 4 °C for 15 min. The column was then re-opened. The high concentration of imidazole allowed to remove the CSB-FL from the resin, which was found in the eluted sample (EL) that passed through the column, of which 10 µl was collected for SDS-PAGE gel and stored at 4 °C.

The presence of CSB-FL protein in all the collected samples was then analysed by 8% SDS-gel (Table 5.1).

5.2.3 Affinity Purification Using Heparin Column (Only for Large Scale Expression)

The EL obtained after nickel chromatography was 10-times diluted in Heparin buffer A (25 mM HEPES pH 7.9, 10% glycerol, 0.01% Triton-X, 300 mM KCl, 1 mM TCEP, 0.1 mM PMSF) to decrease the concentration of imidazole. Heparin buffer A was also used to equilibrate a HiTrap Heparin High Performance column (Cytiva life sciences) attached to an ÄKTA pure protein purification system (Cytiva life sciences).

Using a sample pump, the diluted EL was loaded on the pre-equilibrated 5 ml HiTrap Heparin column. After the loading, the column was washed with 10CV Heparin Buffer A. Finally, the protein was eluted using a linear gradient (0–100%) of Heparin buffer B (25 mM HEPES pH 7.9, 10% glycerol, 0.01% Triton-X, 800 mM KCl, 1 mM TCEP, and 0.1 mM PMSF). The El was collected using a fractionator system and 10 µl of each fraction was collected and analysed by 8% SDS page.

The purest fractions containing CSB bands were combined and concentrated to <500 µl volume using a 15 ml 100K Amicon device (Merk Millipore) at 5,000 xg (fixed rotor)/4,000 xg (swinging rotor), for several 10 min rounds at 4 °C.

5.2.4 Size Exclusion Purification Using Superdex200 / Superose 6 Column (Only for Large Scale Expression)

The concentrated CSB sample was injected either on a pre-equilibrated Superdex 200 Increase 10/300 GL column (Cytiva life sciences) or on a pre-equilibrated Superose 6 Increase 10/300 GL column (Cytiva life sciences) using a 500 µl sample loop injector. The columns were pre-equilibrated in Gel Filtration buffer (25mM HEPES pH 7.9, 10% glycerol, 0.01% Triton-X, 200 mM KCl, 1 mM TCEP, and 0.1 mM PMSF). Finally, CSB-FL was eluted in Gel Filtration buffer and the purity of the protein was assessed by 8% SDS page.

The purest fractions containing CSB-FL were concentrated using 100 KDa Amicon device to a final concentration of 1 mg/ml and stored at -80 °C.

5.3 Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis (SDS-PAGE)

Protein samples were analysed by electrophoresis under denaturing, reducing conditions using discontinuous SDS-polyacrylamide gels. Vertical gels were prepared following the recipe indicated in Table 5.1, and left to dry for at least 30 min at RT prior to use. Then, a small volume (mostly10 µl) of each sample was mixed with 2 µl of SDS-dye (6X DNA Loading Dye & SDS Solution, Thermo Scientific) and heated at 75 ºC for 10 min. Then, samples and 3 µl of ready-to-use protein marker (either Precision Plus Protein™ All Blue Prestained Protein Standards, Bio-Rad or Color Prestained Protein Standard-Broad Range 11-245 kDa, New England Biolab) were loaded on the gel and run at constant voltage (120 V) for 60–90 min in 1X Novex™ Tris-glycine-SDS buffer (Thermo Scientific).

Finally, gels were stained using QC Colloidal Coomassie Stain (Bio-Rad), at RT, gentle shaking, for either 1 h or O/N. Then, gels were de-stained by incubation in milliQ water at RT for 1 h, with gentle shacking and visualised using Image Quant LAS 4000 (Cytiva).

	8%	10%	12%	15%	Stacking buffer
MilliQ water	4.75 ml	4 ml	3.4 ml	2.4 ml	3.05 ml
Resolving Buffer (VWR)/ Stacking buffer (National diagnostics)	2.5 ml	2.5 ml	2.5 ml	2.5 ml	1.25 ml
Protogel 30% (National diagnostics)	2.65 ml	3.4 ml	4 ml	5 ml	650 µl
Ammonium persulfate (APS 10%)	100 µl	100 µl	100 µl	100 µl	25 µl
N,N,N′,N′-tetramethylethylenediamine (TEMED)	10 µl	10 µl	10 µl	10 µl	5 µl
	Total: 10 ml	Total: 10 ml	Total: 10 ml	Total: 10 ml	Total: 5 ml

Table 5.1. Recipe for resolving and stacking components of two SDS gels of different percentage of polyacrylamide. The resolving solution is the first reagent to be prepared. Once the resolving gel is solidified, the stacking gel solution is prepared and poured on the top of the resolving gel. Finally, the desired combs are added, and the gel is left to solidify for additional 30 min.

5.4 Western Blotting (WB)

The presence of the recombinant protein was confirmed by WB. First, proteins were separated from each other based on their size by SDS-PAGE.

Meanwhile, a nitrocellulose membrane and four filter papers were cut with the following dimensions: 10 cm × 7.5 cm and soaked in Transfer buffer (14.2 g glycine, 3.03 g trizma base, 200 ml methanol, and 800 ml milliQ water).

Then, a transfer "sandwich" was prepared by subsequentially assembling the layers of a sandwich-type structure. From the bottom, the layers were composed by placing a support grid, a black sponge, two filter papers, the nitrocellulose membrane, the SDS-gel, two filter papers and an additional black sponge on the top. Finally, the "sandwich" was closed with a second grid. The air bubbles between the layers were gently removed with a roller. The "sandwich" was placed into a transfer cassette and then into a tank filled with Transfer buffer. The transferring was performed at 100 V for 1 h, on ice.

After the transfer, the membrane was stained for 5 min with Ponceau S (Sigma-Aldrich) to confirm that protein transfer was successful. Later, the membrane was washed twice with milliQ water. The membrane was then treated with Blocking buffer (5% milk in TBS-T) for 1 h at RT with gentle shaking. A rabbit anti-CSB primary antibody (ABCAM: ab96089) was diluted to 1:500 in Blocking buffer and incubated with the membrane under gentle shaking at either 4 °C, O/N, or at RT, for 1 h. After incubation with the primary antibody, the membrane was washed four times with TBS-T for 5 min with gentle shaking and incubated with a goat anti-rabbit HRP secondary antibody (ABCAM: ab205718) diluted 1:10,000 in Blocking buffer for 1 h, at RT, with gentle shaking. After this second incubation, the membrane was washed four times in TBS-T for 5 min each wash, at RT and gentle shaking. Finally, 1 ml of the ECL substrate for HRP (Pierce™, Thermo Scientific) was added on the top of the membrane, which was developed using Image Quant LAS 4000 (Cytiva) set to chemiluminescence for the WB signal, and to Cy5 for visualising the protein marker.

5.5 Bioinformatic Analysis

To investigate the conserved sequence of CSB-HD domain within different organisms, a Multiple Protein Sequence Alignment by Clustal Omega [222] online software was performed. FASTA sequences of the CSB homologues entered as input are reported below (accession numbers provided in parentheses):

1. *Saccharomyces cerevisiae*, Rad26 (GenBank: CAA57290.1)
2. *Schizosaccharomyces pombe*, Rhp26 (GenBank: CAB62827.1)
3. *Homo sapiens*, Excision repair cross-complementing rodent repair deficiency, complementation group (ERCC6) (GenBank: AAO13487.1)
4. *Mus musculus*, ERCC6 (NCBI Reference Sequence: NP_001074690.1)
5. *Gallus gallus*, ERCC6 (NCBI Reference Sequence: XP_040530141.1)
6. *Xenopus laevis*, ERCC6 (NCBI Reference Sequence: XP_018081111.1)
7. *Danio rerio*, ERCC6 (NCBI Reference Sequence: XP_017214320.1)
8. *Caenorhabditis elegans*, ERCC6 (GenBank: CAB03135.2)

The secondary structure prediction was carried out using the PSIPRED website [226]. The DNA sequence of CSB (UniProtKB: Q03468 [244]) was submitted to PSIPRED in a FASTA format. The results of the prediction software were received both textually via e-mail and graphically via the web.

5.6 Restriction-Free Cloning to Insert CSB-HD Into pET_SUMO-tag Vector

The CSB-HD sequence optimised for bacterial expression was cloned into a pCS46_His_6-SUMO vector for *E. coli* expression by restriction-free (RF) cloning [245].

Two different RF-cloning reactions were performed. The first RF-cloning reaction inserted the optimised CSB-HD (519–1,002) sequence into the pCS46_His_6-SUMO vector at the C-terminal of the SUMO sequence (Figure 3.7) using Forward primer 1 and Reverse primer 1 (see Table 5.2). The second RF-cloning reaction inserted an optimised short (63 bp) sequence at the N-terminal of the CSB-HD (519–1,002) that was previously cloned in pCS46_His_6-SUMO (Figure 3.9), to generate a more stable CSB-HD (498–1,002) construct. The Forward primer 2 and the Reverse primer 2 used in the second cloning are reported in Table 5.2. The optimised CSB-HD sequence was synthetised by Genescript service, while the short (63 bp) sequence and primers were designed using rf-cloning.org service and purchased from Eurogentec.

Each RF-cloning reaction required two different PCR steps. The first PCR generated the "megaprimer", while the second one inserted the "megaprimer" into the desired position of the vector. The PCR mixtures and thermal programs are shown below.

	Sequence (5'-3')
Forward 1	GTTCCAGCAGCAGACGGGAGGTTGGGAGCTGCATTGCCAA
Reverse 1	CAGCGGTGGCAGCAGCCAACTCTTAATACAGGTCATTGCTCTTAAAG
Forward 2	GTTCCAGCAGCAGACGGGAGGTAAAGTGCCGGGTTTCCTG
Reverse 2	CTTGTTGGCAATGCAGCTCCCACAGCCAACGAACACCGGT

Table 5.2. Nucleotide sequence of the primers used for the CSB-HD (519–1,002) and CSB-HD (498–1,002) RF insertion in pCS46_His_6-SUMO vector.

CSB-HD (519–1,002)

RF I PCR	1x		Program			
5x Phusion HF Buffer	10	µl	Initial denaturation	98 °C	30 sec	x1
dNTPs (10mM)	1	µl				
FW primer (10µM)	2.5	µl	Denaturation	98 °C	8 sec	
RV primer (10µM)	2.5	µl				x35
DNA template (100µM)	0.5	µl	Extension	72 °C	20 sec (30s/kb)	
Phusion polymerase	0.5	µl				
diH2O	33	µl	Final extension	72 °C	5 min	x1
			Hold	4 °C	∞	
total	50	µl (x2)				

Table 5.3. Sample composition and thermal protocol for the first PCR steps (2X reactions) required to generate the CSB-HD (519–1,002) "megaprimer".

RF II PCR	1x		Program		
5x Phusion HF Buffer	4	µl	Initial denaturation	98°C	2min
dNTPs (10mM)	0.4	µl			
Plasmid (101.7ng/µl)	0.4	µl	Denaturation	98°C	8sec
MegaPrimer (13.4ng/µl)	1.2	µl			
Phusion polymerase	0.2	µl	Extension	72°C	12min
diH2O	13.8	µl			
			Final extension	72°C	5min
total	20	µl	Hold	4°C	∞

Table 5.4. Sample composition and thermal protocol for the second PCR required to insert the CSB-HD (519–1,002) "megaprimer" (insert) into pCS46_His_6-SUMO vector (plasmid).

CSB-HD (498–1,002)

RF I PCR	1x		Program			
5x Phusion HF Buffer	10	µl	Initial denaturation	98 °C	30 sec	x1
dNTPs (10mM)	1	µl				
FW primer (10µM)	2.5	µl	Denaturation	98 °C	8 sec	
RV primer (10µM)	2.5	µl				x35
DNA template (322 ng/µl)	1	µl	Extension	72 °C	60 sec (30s/kb)	
Phusion polymerase	0.5	µl				
diH2O	32.5	µl	Final extension	72 °C	5 min	x1
			Hold	4 °C	∞	
total	50	µl (x2)				

Table 5.5. Sample composition and thermal protocol for the first PCR steps (2X reactions) required to generate the short "megaprimer" sequence.

RF II PCR	1x		Program		
5x Phusion HF Buffer	4	µl	Initial denaturation	98°C	2min
dNTPs (10mM)	0.4	µl			
Plasmid (113.5ng/µl)	0.4	µl	Denaturation	98°C	8sec
MegaPrimer (66ng/µl)	6.2	µl			
Phusion polymerase	0.2	µl	Extension	72°C	12min
diH2O	8.8	µl			
			Final extension	72°C	5min
total	20	µl	Hold	4°C	∞

Table 5.6. Sample composition and thermal protocol for the second PCR required to insert the short "megaprimer" sequence (insert) into the pCS46_His_6-SUMO-CSB-HD (519–1,002) vector (plasmid).

Two identical reactions of 50 µl volume each were necessary to obtain enough material from the first PCR. After the first PCR (Table 5.3 or 5.5), 8.3 µl of 6X purple gel loading dye (New England BioLab) was added into each product. Then, 50 µl of the two products were loaded

and separated on 3% agarose gel (see Section 5.10). After the run, the bands corresponding to the expected size of the amplified CSB-HD were cut and purified by gel extraction (Monarch® DNA Gel Extraction Kit, New England BioLab). The concentration of the product was calculated using Nanodrop (Thermo Scientific) set at 260 nm. Then, the purified product (insert) was inserted into the empty pCS46_His_6-SUMO vector by a second PCR (Table 5.4 or 5.6) using an insert to vector ratio of 1:40.

The product obtained after the second PCR was digested by adding 1 µl of DpnI enzyme (New England BioLab) directly into the PCR mix followed by incubation at 37 °C for 2 h. The enzyme was then inactivated at 80 °C for 20 min.

A 2 µl aliquot of the sample was transformed into 50 µl of H5α competent *E. coli* cells (New England Bio-lab) following the supplier's heat shock protocol. After transformation, 100 µl of cells were plated on LB-agar (Fisher BioReagents) plates supplemented with 100 µg/ml ampicillin and incubated at 37 °C, O/N. The resulting colonies were then expanded in LB-broth (Fisher BioReagents) supplemented with 100 µg/ml ampicillin at 37 °C at 200 rpm, O/N. Finally, the pCS46_His_6-CSB-HD-SUMO was extracted from the cells by miniprep (PureYield™, Promega) following the supplier's protocol. The quality of the product was determined by Sanger sequencing.

5.7 CSB-HD Bacterial Expression

In brief, a 1 µl volume of either pCS46_His_6-CSB-HD (519–1,002) -SUMO or pCS46_His_6-CSB-HD (498–1,002)-SUMO plasmid was transformed into BL21(DE3) competent *E. coli* cells (New England Bio-lab) following the supplier's heat shock protocol. After transformation, 900 µl of cells were directly added to 120 ml of LB-broth (Fisher BioReagents) supplemented with ampicillin (100 µg/ml) and incubated at 37 °C, 250 rpm, O/N (small inoculum). From the small inoculum, 12.5 ml (dilution 1:40) was expanded in 500 ml (8 flasks, 4 l in total) of 2XYT media supplemented with 100 µg/ml ampicillin at 37 °C, 250 rpm, until reaching an optical density (OD) of 0.6 at 600 nm, after which the protein expression was induced with 0.05 mM IPTG either at 36 °C for 4 h, or at 18 °C, O/N, 200 rpm.

5.8 CSB-HD Cell Lysis and Purification

5.8.1 Cell Lysis

Cells were harvested at 4 °C, 4,000 xg for 30 min. The pellet was then resuspended in 300 ml of Lysis buffer (25 mM Tris-HCl pH 9.0, 1 mM EDTA, 10% glycerol, 1% Triton-X, 300 mM KCl, 1 mM TCEP, and 0.1 mM PMSF) supplemented with 90 mg of Lysozyme (300 µg/ml; Thermo Scientific) and 6X protease inhibitor tablets (Sigmafast-EDTA-free, Sigma-Aldrich). The mixture was stirred at 4 °C for 20 min prior to add 5 nM $MgCl_2$ and 20 µl Benzonase (5 KU). The mixture was gently stirred at 4 °C for an additional 30 min. The suspension was gently sonicated on ice for 5 sec, with short pulses of 5 sec each followed by 55 sec pause, at 20% amplitude (sonicator from Thermo Fisher Scientific). Finally, the cell lysate (TF) was centrifugated at 4 °C, 16,000 xg, for 40 min.

5.8.2 Protein Purification

The clarified supernatant obtained from cellular lysis (soluble fraction: SF) was purified by his-affinity purification using 5 ml of Nickel resin (HisPur™ Ni-NTA Resin, Thermo Scientific), and left gently stirring either for an hour or O/N at 4 °C prior to adding the mixture in an empty column.

Once all the unbound material passed through the column (flow through: FT), the resin was washed five times with 2CV of Wash buffer (25 mM HEPES pH 7.9, 10% glycerol, 0.01% Triton-X, 300 mM KCl, 20 mM imidazole, 1 mM TCEP, and 0.1 mM PMSF) supplemented with a low percentage of imidazole to remove the non-specific interactions. An additional washing step with 2CV High salt concentration buffer (1 M NaCl) was performed before the last wash, to remove any presence of DNA.

After the washing steps (W), the resin was incubated with 5CV of Elution buffer containing a high concentration of imidazole (25 mM HEPES pH 7.9, 10% glycerol, 0.01% Triton-X, 300 mM KCl, 250 mM imidazole, 1 mM TCEP, and 0.1 mM PMSF) for 15 min at 4 °C. CSB-HD was then gravitationally eluted (El) from the column.

The presence of the protein (CSB-HD (519–1,002) MW: 66.6 kDa / (498–1,002) MW: 70 kDa) in all the collected samples was analysed by 10% SDS-gel (Table 5.1). After performing

nickel purification, the protein was further purified by heparin affinity column, using the same protocols and buffers described in Section 5.2.3. Finally, the purest fractions containing CSB-HD were pooled and concentrated using Amicon device 30–50K (Merck Millipore) and stored at -80 °C.

5.8.3 SUMO-Tag Removal

Proteins expressed from SUMO-tagged expression vectors were subjected to removal of SUMO-tag by SUMO Protease. SUMO protease (Sigma-Aldrich) was diluted in 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 10% glycerol, and 1 mM DTT. After the Nickel affinity purification, samples were diluted 10 times with Lysis buffer to reduce the concentration of imidazole, and treated O/N with 20 µl of SUMO protease at 4 °C. Following digestion with SUMO Protease, cleaved proteins carrying the his-tag and SUMO-tag were separated by his-tag affinity chromatography. The quality of cleavage was checked by SDS-PAGE gel. Because the His_6-tag was still attached to the N-terminal of the cleaved SUMO-tag, the removed tags were expected to be in the fraction eluted from the resin at high concentration of imidazole, whilst the untagged protein that was not able to interact with the resin was expected in the FT.

5.9 Oligonucleotide Preparation

The annealing of the oligonucleotides used for the biochemical and biophysical experiments (see Tables 5.7, 5.8 and 5.9) was performed in either 25 mM HEPES pH 7.6, 100 mM KCl or 25 mM HEPES pH 7.6, 100 mM LiCl buffer at 95 °C for 10 min, following slow-cooling to RT, O/N. c-MYC, ssDNA (c-MYC mutant) and dsDNA used for BG4 validation were annealed in 10 mM Tris-HCl pH 7.4, and 100 mM KCl buffer at 95 °C for 10 min, following slow-cooling to RT, O/N.

NAME	SEQUENCE (5'→3')
(5'-Cy5) rDNA-1 5'-tail	ATAATTATAAATAAATAATGGGGCCGGGGGTGGGGTCGGCGGGGAAA
Complementary rDNA-1 5'-tail	TTTCCCCGCCGACCCCACCCCCGGCCCCTATTATTTATTTATAATTAT
(5'-Cy5) rDNA-2 5'-tail	ATAATTATAAATAAATAATAGGGTCGGGGGGTGGGGCCCGGGCCGGGG
Complementary rDNA-2 5'-tail	CCCCGGCCCGGGCCCCACCCCCCGACCCTATTATTTATTTATAATTAT
(5'-Cy5) rDNA No tail	GGGGCCGGGGGTGGGGTCGGCGGGGAAA
Complementary rDNA No tail	TTTCCCCGCCGACCCCACCCCCGGCCCC
(5'-FAM) c-KIT1 5'-tail	ATAATTATAAATAAATAATAGGGAGGGCGCTGGGAGGAGGGAAA
Complementary c-KIT1 5'-tail	TTTCCCTCCTCCCAGCGCCCTCCCTTATTATTTATTTATAATTAT
(5'-FAM) hTELO 5'-tail	ATAATTATAAATAAATAATGGGTTAGGGTTAGGGTTAGGGAAA
Complementary hTELO 5'-tail	TTTCCCTAACCCTAACCCTAACCCTATTATTTATTTATAATTAT
(5'-FAM) HRAS 5'-tail	ATAATTATAAATAAATAATATCGGGTTGCGGGCGCAGGGCACGGGCGAAA
Complementary HRAS 5'-tail	TTTCGCCCGTGCCCTGCCCTGCGCCCGCAACCCGATATTATTTATTTATAATTAT
(5'-FAM) c-MYC 5'-tail	ATAATTATAAATAAATAATTGGGTGGGTAGGGTGGGTAAA
Complementary c-MYC 5'-tail	TTTACCCACCCTACCCACCCATATTATTTATTTATAATTAT
(5'-FAM) c-MYC 3'-tail	AAATGGGTGGGTAGGGTGGGTATAATTATAAATAAATAATA
Complementary c-MYC 3'-tail	TATTATTTATTTATAATTATACCCACCCTACCCACCCATTT
(5'-FAM) c-MYC No tail	TGAGGGTGGGTAGGGTGGGTAA
Complementary c-MYC No tail	TTACCCACCCTACCCACCCTCA
(5'-FAM) c-KIT1 No tail	TGGGAGGGCGCTGGGAGGAGGG
Complementary c-KIT1 No tail	CCCTCCTCCCAGCGCCCTCCCA
FRET: (5'-FAM/3'-TAMARA) c-KIT1	AGGGAGGGCGCTGGGAGGAGGGGC
FRET: Complementary c-KIT1	GCCCCTCCTCCCAGCGCCCTCCCT
FRET: (5'-FAM/3'-TAMARA) hTELO	GGGTTAGGGTTAGGGTTAGGG
FRET: Complementary hTELO	CCCTAACCCTAACCCTAACCC
(5'-Cy5) rDNA-1 3'-tail	AAAGGGGCCGGGGGTGGGGTCGGCGGGGATAATTATAAATAAATAATA
(5'-Cy5) rDNA-3 5'-tail	ATAATTATAAATAAATAATAGGGAGGGAGACGGGGGGG
rDNA-2 No tail	AGGGTCGGGGGGTGGGGCCCGGGCCGGGG
rDNA-3 No tail	AGGGAGGGAGACGGGGGGG
rDNA-1 3bp tail at 5'	ATAGGGGCCGGGGGTGGGGTCGGCGGGGAAA
rDNA-1 5bp tail at 5'	TAATAGGGGCCGGGGGTGGGGTCGGCGGGGAAA
rDNA-1 10bp tail at 5'	ATAAATAATAGGGGCCGGGGGTGGGGTCGGCGGGGAAA
rDNA-1 3bp tail at 3'	AAAGGGGCCGGGGGTGGGGTCGGCGGGGATA
rDNA-1 5bp tail at 3'	AAAGGGGCCGGGGGTGGGGTCGGCGGGGATAAT
rDNA-1 10bp tail at 3'	AAAGGGGCCGGGGGTGGGGTCGGCGGGGATAATAAATA

Table 5.7. Oligonucleotide sequences. The tail end of the oligonucleotide is indicated in red.

NAME	SEQUENCE (5'→3')
rDNA-1 G4 template	TCTGCTTTGGGAACCCGAGAGGAGCGCTTATGGGGCCGGGGGTGGGG TCGGCGGGGGACTCAGCCGAGCAGCCGAGCACTCTAGCTCTAG
rDNA-4 G4 template	TCTGCTTTGGGAACCCGAGAGGAGCGCTTATGGGTGGCGGGGGGGAG AGGGGGGGACTCAGCCGAGCAGCCGAGCACTCTAGCTCTAG
Primer (Cy5)	CTAGAGCTAGAGTGTCGGC
rDNA-5 bimolecular 5'-tail	TAATAGGGGCCGGGGGTGGGGT
(5'-Cy5) ssDNA	GGCATAGTGCGTGGGCG
(5'-Cy5) *O. nova* G4 5'-tail	ATAATTATAAATAAATAATGGGGTTTTGGGG
Complementary *O. nova* G4 5'-tail	CCCCAAAACCCCATTATTTATTTATAATTAT
(5'-Cy5) rDNA-1 mut. 5'-tail	ATAATTATAAATAAATAATAGGTGCCGTGGTTGGTGTCGTCGGTGAAA

Table 5.8. Oligonucleotide sequences. The complementary nucleotides between the primer and the sequences used for the polymerase stop assay are indicated in green. The G4-forming sequence within the oligonucleotides used for polymerase stop assays is reported in blue, and the tails are reported in red.

NAME	SEQUENCE (5'→3')
c-MYC	TGAGGGTGGGTAGGGTGGGTAA
ssDNA	GGCATAGTGCGTGGGCG
Complemetary ssDNA	CGCCCACGCACTATGCC

Table 5.9. Oligonucleotide sequences used to validate the BG4 antibody.

5.10 Agarose Gel

Gels were prepared in 50 ml final volume following the recipe reported in Table 5. The solution was boiled for three min using a microwave and poured into a 10 cm × 10 cm horizonal tray with a 10-well comb, and left to dry for at least 60 min at RT prior to use.

Then, 10 µl of each sample was mixed with 2 µl 6X purple gel loading dye either with or without SDS (New England BioLab). When used, 1 µl of DNA ladder (either 50 bp or 1 kb DNA Ladder, New England BioLab) was mixed with 9 µl of milliQ water and 2 µl of the selected 6X purple gel loading dye.

Gels were run at a constant voltage (either 80 V or 100 V) for 50–90 min in 1X Tris-borate-EDTA buffer (Thermo Scientific™) and visualised by G:BOX F3 set to detect Gel-red. For fluorophore-labelled samples, gels were scanned on a GE Typhoon imager (Cytiva) set to detect the desired fluorophore.

	0.6% agarose	**0.8% agarose**	**1.5% agarose**	**3% agarose**
Agarose	0.3 g	0.4 g	0.75 g	1.5 g
Gel-Red (3X) (Biotium)	16.7 ml	16.7 ml	16.7 ml	16.7 ml
TBE (10X)	5 ml	5 ml	5 ml	5 ml
Up to 50 ml with milliQ water				

Table 5.10. Agarose gel recipe with different percentages of agarose.

5.10.1 NMM Staining

For the N-methylmesoporphyrin IX (NMM) staining, a 0.6% agarose gel (see Table 5.10) was loaded with 10 µl of pre-annealed oligonucleotides (see Tables 5.7 and 5.8 for the oligonucleotide sequences used) at final concentrations of 10 µM and run for 50 min in 1X TBE at 100 V. The gel was then incubated in 50 ml 1X TBE containing 2 µl NMM (from original stock of 10 mM) for 30 min, at RT with gentle shaking. Then, the gel was washed for 20 min in 1X TBE at RT and gentle shaking, and visualised by GE Typhoon imager (Cytiva) set to detect EtBr.

5.11 Polyacrylamide EMSA

EMSA reactions (10 µl) were prepared by mixing 25 nM of pre-annealed fluorophore-labelled oligonucleotide in the presence of increasing concentrations of protein in 1X Reaction buffer. The Reaction buffer used for the BG4 EMSAs contained: 10 mM Tris-HCl pH 7.4, 1 mM EDTA, 0.1 mg/ml BSA, 100 mM KCl, and 0.02% Tween-20 (BG4). The Reaction buffer used for CSB EMSAs was composed by 20 mM HEPES pH 8.0, 40 ng/µl BSA, 1 mM DTT, 1 mM $MgCl_2$, and 100 mM KCl (or LiCl). The samples were incubated for 30 min either at RT (BG4) or at 30°C (CSB). After 30 min incubation, 2 µl of 50% glycerol was added into each reaction before loading the gel with 10 µl of the sample.

A 6% polyacrylamide gel was prepared (8.8 ml 2X TBE, 7 ml 30% acrylamide 29:1, 280 µl 10% APS, and 35 µl TEMED (4X gels)) and pre-run at 100 V for 15 min in 0.5X TBE. The gel was run in 0.5X TBE, 100 V, for approximately 30-35 min at RT and visualised by GE Typhoon imager (Cytiva) set to detect the correct fluorophore. Quantification of gel bands was processed using ImageJ software.

5.12 CSB Agarose EMSA

The final volume required for each EMSA reaction was 10 µl. Each sample was prepared by adding 0.5 µl (25 nM final concentration) of pre-annealed fluorophore labelled nucleotide (see section 5.9), 2.5 µl EMSA reaction buffer 4X (20 mM HEPES pH 8.0, 40 ng/µl BSA, 1 mM DTT, 1 mM $MgCl_2$, and 100 mM KCl or LiCl), and increasing concentrations of either CSB-FL or CSB-HD. The reactions were incubated for 30 min at 30 °C and then supplemented with 2 µl of either 50% glycerol or 6X no SDS-purple gel loading dye (New England BioLabs). Samples were then separated on a pre-run 0.6% agarose gel in 1X TBE buffer with the electrophoretic run set at 100 V for 50 min at RT. Gels were visualised by GE Typhoon imager (Cytiva) set to detect the desired fluorophore. Quantitation of the gel bands was achieved using ImageJ software and, data analysis was performed with GraphPad Prism 9.0.1 using one-site specific binding equation and 95% CI (see section 2.2).

5.13 Competitive EMSA

The competitive EMSAs were performed following the same protocol as the standard described in Section 5.12. However, in this case, the concentration of CSB-HD was kept constant (2 nM). Samples were incubated 30 min prior to adding increasing concentrations (from 0 to 5 μM) of the selected G4-ligand (either PDS or CX-5461) diluted in milliQ water. Samples were incubated at 30 °C for an additional 2 h before the addition of 2 μl of 6X no SDS-purple gel loading dye to each reaction (New England BioLabs). Samples were then separated on a pre-run 0.6% agarose gel in 1X TBE at 100 V for 50 min, at RT, and gels were visualised using a GE Typhoon imager.

5.14 Gel-based Resolvase Assay

Each sample for the resolvase assay was prepared in a 20 µl volume by mixing 5 µl of 4X Resolvase reaction buffer (20 mM HEPES pH 8.0, 40 ng/µl BSA, 1 mM DTT, 1 mM $MgCl_2$, and 25 mM KCl), 1 nM of G4 and 5X fold excess of the unlabelled complementary strand in milliQ water. Following, 10 nM (or 20/40 nM) of either CSB-FL or CSB-HD were added into each sample and incubated at 30 °C for different times (0.5–40 min or 160 min/O/N). Reactions were then stopped with 5 µl of Stop solution (0.5% SDS, 50 mM EDTA in milliQ water) and separated using pre-run 10% polyacrylamide (19:1 acrylamide-bisacrylamide) gels at 100 V for 1 h in 0.5X TBE. The gel recipe is reported in Table 5.11. Gels were visualised using an GE Typhoon imager (Cytiva) set to detect the desired fluorophore. Quantitation of the gel bands was performed using ImageJ software and analysed with GraphPad Prism 9.0.1.

	10% polyacrylamide gel
Acrylamide/Bis 19:1, 40% (w/v) solution	3.75 ml
TBE (5X)	2.4 ml
APS 10%	200 µl
milliQ water	5.65 ml

Table 5.11. Recipe for 2X 10% resolvase polyacrylamide gels.

5.15 FRET-based Resolvase Assay

The FRET-based Resolvase assay was performed at 25 °C in 100 mM KCl, 25 mM Tris-HCl pH 7.4 buffer containing either 50 nM of pre-annealed 5'-FAM, 3'-TAMRA -labelled hTELO or 50 nM of pre-annealed 5'-FAM, 3'-TAMRA -labelled c-KIT1 oligonucleotides (see Section 5.9). The reaction mixtures were supplemented with 5 mM ATP (where applicable) and 10 nM of either CSB-FL or CSB-HD. The unfolding reactions were triggered by the addition of 250 nM complementary DNA sequence at time (t)=0. The reactions were performed using a BMG Labtech Clariostar Plus instrument, where samples were excited at 488 nm wavelength with an 8 nm excitation slit and emission measured at 518 nm with an 8 nm emission slit at regular time points. The increasing donor fluorescence signal (FAM) at 518 nm was measured to determine the unfolding progression. Data were normalised by subtracting the t=0 signal. All the experiments were performed in triplicates and data were analysed by GraphPad Prism 9.0.1 using single-exponential decay functions. The 95% confidence intervals for apparent rate constants ($K_{apparent}$) and half-lives (reported in seconds, s) were obtained using constraints as: y0=0 and plateau=100.

5.16 Polymerase Stop Assay

Polymerase stop assays were adapted from a reported protocol [138]. Each reaction was prepared in a PCR tube by adding 8 µl of 1X NEB buffer 1 (New England BioLabs) supplemented with either 100 mM KCl or 100 mM LiCl. Then, 500 nM of the template oligonucleotide and 50 nM of 5'-Cy5 fluorophore-labelled primer were added to the tube and annealed at 95 °C for 5 min following slow cooling at RT, O/N. The sequences used for these assays are reported in Table 5.8. After the annealing, 1 µl of either CSB-FL or CSB-HD at the final concentration of 0.5 nM, 5 nM or 50 nM was added to the annealed material and incubated at 37 °C for 30 min prior to adding 1 µl of Polymerase mix containing 89 µl 1X NEB buffer 1, 10 µl 10 mM dNTP mix (New England BioLabs) and 1 µl Klenow fragment^{exo-} (New England BioLabs). The reactions were incubated at 37 °C for 2 h and then stopped with 20 µl of 2X orange gel loading dye (New England BioLabs). The samples were boiled at 95 °C for 5 min and 7 µl of each sample was run on a pre-run 15% TBE-Urea gel at 300 V for 25 min in 1X TBE, RT. The recipe for the 15% TBE-Urea gel is reported in Table 5.12. Gels were visualised using GE Typhoon imager (Cytiva) set to detect Cy5, processed and analysed using ImageJ software and GraphPad Prism 9.0.1, respectively.

	15% TBE-Urea gels
Urea	10 g
TBE (10X)	2 ml
Acrylamide/Bis 19:1, 40% (w/v) solution	7.48 ml
APS 10%	160 µl
TEMED	20 µl
milliQ water	10.34 ml

Table 5.12. Recipe for preparing 2X 15% TBE-Urea gels

5.17 CD Analysis

CD spectra of 10 μM final concentration of pre-annealed oligonucleotides (see Section 5.9) were recorded at 25 °C on a JASCO J-810 circular dichroism spectrophotometer using a 1 mm path length quartz cuvette. The final recorded spectra represented the average of three different reads where the absorbance of the buffers (25 mM HEPES pH 7.6, 100 mM KCl, or LiCl) was subtracted. The CD measurements were performed over a range of 200–320 nm using a response time of 2 sec, 1nm pitch and 0.5 nm bandwidth, and analysed using GraphPad Prism 9.0.1.

5.18 Restriction-Free Cloning to Insert CSB-HD Nucleolar Sequence Into pEGFP-1C Vector

The CSB-HD containing both NoLS1 and NLS1 from the full-length CSB sequence[178] (residues: 285–1,009; 2,175 bp) was inserted into a pEGFP-1C empty vector for mammal expression by RF-cloning using Forward primer 3 and Reverse primer 3. The sequence of the primers used is reported in Table 5.13. The sequences were designed using the online rf-cloning.org service and purchased from Merck Life Science UK.

The RF-cloning protocol used followed the same steps of the protocol reported in Section 5.6. The components of the first and second PCR reactions and their thermal profiles are shown in Table 5.14 and Table 5.15, respectively.

	Sequence (5'-3')
Forward 3	AGCTGTACAAGTCCGGACTCAGAAAGAAGCAAGGTTGTAATAAAAGAG
Reverse 3	GCAGAATTCGAAGCTTGAGCTCGAGAAGGACTAGTCAGAGTAAATAGCTCA

Table 5.13. Nucleotide sequence of the primers using to insert CSB-HD (285–1,009) into pEGFP-1C empty vector by RF-cloning.

RF I PCR	**1x**		**Program**			
5x Phusion HF Buffer	10	µl	Initial denaturation	98 °C	30 sec	x1
dNTPs (10mM)	1	µl				
FW primer (10µM)	2.5	µl	Denaturation	98 °C	8 sec	
RV primer (10µM)	2.5	µl	Annealing	55 °C	22 sec	x35
DNA template (300ng/µl)	1	µl	Extension	72 °C	90 sec (30s/kb)	
Phusion polymerase	0.5	µl				
diH2O	32.5	µl	Final extension	72 °C	5min	x1
			Hold	4 °C	∞	
total	50	µl				

Table 5.14. Sample composition and thermal profile of the first PCR step (4X different reactions) to generate the CSB-HD (285–1,009) megaprimer.

RF II PCR	1x		Program			
5x Phusion HF Buffer	4	µl	Initial denaturation	95 °C	2 min	x1
dNTPs (10mM)	0.4	µl				
Plasmid (180ng/µl)	0.4	µl	Denaturation	98 °C	30 sec	
MegaPrimer (139.8 ng/µl)	3	µl	Annealing	61 °C	1 min	x18
Phusion polymerase	0.2	µl	Extension	72°C	12 min	
diH2O	12	µl				
			Final extension	72 °C	5 min	x1
total	20	µl	Hold	4 °C	∞	

Table 5.15. Sample composition and thermal profile of the second PCR step to insert the CSB-HD (285–1,009) megaprimer into pEGFP-1C empty vector.

After the second PCR, 3 µl of DpnI enzyme was added into the product and incubated at 37 °C for 1 h. Then, 2 µl of the treated PCR product was transformed into a 50 µl NEB-H5α cell following the supplier's heat shock protocol. The cells were then plated on LB-agar plates supplemented with 50 µg/ml kanamycin. The colonies were expanded in LB-broth (Fisher BioReagents) supplemented with 50 µg/ml kanamycin at 37 °C, 200 rpm, O/N. Finally, the pEGFP-1C_CSB-HD (285–1,009) plasmid was extracted from the cells using miniprep (PureYield™, Promega) following the supplier's protocol. The final plasmids were then linearised by enzymatic digestion using NdeI restriction enzyme (New England BioLabs). For this, 1 µl of the enzyme was added into 50 µl reaction composed by 1-5 µg of plasmid, 5 µl 10X NEB buffer, and milliQ water. The reactions were incubated at 37 °C for 1 h. The reactions were then run on 0.8% Gel-red-agarose gel and the size of the linearised plasmids was compared to the size of the linearised recipient vector. The products that corresponded to the insertion size by gel were sent for Sanger sequencing for quality check.

5.19 Cellular Maintenance

HeLa cells were grown in high glucose Dulbecco's Modified Eagle Medium (DMEM) media (Sigma-Aldrich) supplemented with 10% FBS at a cellular density of $1.5–1.7\times10^6$ using T75 flasks for adherent cells. Cells were maintained in a cell culture incubator at 37 °C, 5% CO_2 and passaged every three or four days. Media was replaced with fresh media every two days. Cells were detached from the flask using 2 ml 0.25% trypsin solution for 5 min.

CS1AN cells were grown in DMEM media (Sigma-Aldrich) supplemented with 10% FBS and 2 mM glycine at a cellular density of $1.2–1.4\times10^6$ cells/ml using T75 flasks for adherent cells. Cells were maintained in a cell culture incubator at 37 °C, 5% CO_2, and passaged every three or four days. Media was replaced with fresh media every two days. Cells were detached from the flask using 2 ml 0.05% trypsin solution for 5 min.

5.20 Cellular Transient Transfection

Both HeLa and CS1AN cells were seeded for 24 h in 8-well chamber slides (0.8cm²/well; size: 25 x 75mm, Nunc™ Lab-Tek® Chamber Slide™) at cellular densities of 0.4×10^5 cells/ml or 0.5×10^5 cells/ml, respectively. After incubation, cells were transfected with the desired plasmid (pEGFP-1C-CSB-HD or empty pEGFP-1C vector) using FuGENE HD Transfection Reagent (Promega). The transfection mixture was prepared by adding: 4 µl plasmid (500-600 ng/µl) and 6 µl FuGENE reagent in 90 µl serum-free media, following incubation for5 min at RT. Then, 15 µl of the transfection mixture was added in the selected wells of the plate containing the seeded cells, and incubated at 37 °C with 5% CO_2 for 48 h.

5.21 Cell Treatment with G4-Ligands

After 24 h from transfection, the selected cells were treated with either 10 μM PDS or 1 μM CX-5461 small molecules. The small molecules were directly diluted into each well to reach the desired final concentrations. Cells were incubated for an additional 24 h prior to fixation.

5.22 Cellular Fixation and Imaging

The growing media was removed from the wells, and cells were carefully washed twice with 250 µl of 1X PBS. Then, the cells were fixed using 150 µl 4% Paraformaldehyde diluted in 1X PBS for 10 min at RT. The cells were then rinsed with 250 µl 1X PBS prior to permeabilization using 150 µl of 1X PBS supplemented with 0.1% Triton-X for 10 min at RT. The cells were washed twice with 250 µl PBS and the cell nuclei were stained using 150 µl Hoechst 33342 (Invitrogen) diluted to 1 µg/ml in PBS for 15 min. Finally, cells were rinsed three times with 1X PBS prior to confocal imaging. The images were acquired using a Nikon Ti2 Microscope (objective magnification: 60×, oil). EGFP signal (EmW:515.0 nm, ExW:470 nm) was selected to identify cellular localisation of CSB-HD (285-1009), while the cell nuclei were detected using Hoechst 33342 staining (EmW:432.0 nm). Cellular images were analysed by NIS element viewer (Nikon) and ImageJ programs and quantified using GraphPad Prism 9.0.1. Each cellular experiment was performed in two independent biological replicates, each consisting of two technical replicates. The quantification was performed on eight different images for each replicate.

5.23 BG4 Antibody Production

The expression and purification steps required to produce BG4 antibody were optimised from previously reported protocols [62,71]. pSANG10-3F-BG4 expression plasmid was transformed in BL21(DE3) pLysS Competent Cells (Promega) following the supplier's heat shock protocol. Bacteria cells were expanded in 2XTY media supplemented with 2% glucose and 50 µg/ml kanamycin at 37 °C, O/N. Subsequently, 23 ml of the bacteria growth was inoculated in 1 l of autoinduction media (8X l) supplemented with kanamycin (MagicMedia™ *E. coli* Expression Medium, Invitrogen) and incubated for 6 h at 37 °C. Then, the temperature was decreased to 18 °C and the mixture was agitated at 200 rpm for 24 h. The cell culture was spun down at 4,000 xg for 30 min at 4 °C and the pellet was resuspended in TES buffer (50 mM Tris-HCl pH 8.0, 20% sucrose, and 1 mM EDTA) supplemented with EDTA-free protease inhibitor tablets (Sigma-Aldrich) for 15 min at 4 °C. Then, a diluted TES (1:5) buffer supplemented with 2 mM $MgCl_2$ and 10 µl of Benzonase was added to the lysate. The lysate was gently stirred for 30 min at 4 °C prior to 20 min centrifugation at 8,000 xg (4 °C). The soluble fraction was then incubated with 10 ml pre-equilibrated HisPur™ Cobalt Resin for his-tag affinity purification (Thermo Fisher Scientific) and stirred for 30 min at 4 °C. The resin was washed twice with 3CV Wash buffer (PBS pH 8, 100 mM NaCl, and 10 mM Imidazole). Finally, BG4 was eluted using 3CV Elution buffer (PBS pH 8.0, and 250 mM Imidazole). The eluate was diluted to 500 ml to decrease the imidazole concentration and directly loaded on 5 ml HiTrap Heparin High Performance column (Cytiva life sciences) equilibrated with Heparin buffer A (25 mM HEPES pH 7.6, 100 mM KCl) for the second affinity purification. The protein was diluted using a linear gradient (0–100%) of Heparin buffer B (25 mM HEPES pH 7.6, and 1 M KCl). The heparin fractions containing BG4 were further purified using a pre-equilibrated Superdex 75 10/300 GL SEC column (Cytiva life sciences) and eluted in Inner Cell Salt buffer (25 mM HEPES pH 7.6, 110 mM KCl, 10.5 mM NaCl, and 1 mM $MgCl_2$). The quality of the protein was evaluated by 12% SDS-PAGE. Finally, the antibody was concentrated using Amicon device 10K (Merck Millipore) and stored at -20 °C.

5.24 BG4 ELISA

The day before, biotinylated oligonucleotides (Table 5.9) were annealed to 500 nM final concentrations, as indicated in Section 5.9.

A streptavidin-coated plate (Life Technologies) was washed three times with 200 µl of PBS. The wells of the plate were then hydrated for 30 min with 200 µl PBS at RT.

The biotinylated oligonucleotides were diluted to 50 nM in annealing buffer (see Section 5.9). Then, 200 µl of each diluted oligonucleotide was added into each well and incubated for 1 h at RT, shaking at 450 rpm (no oligonucleotide was added in the 11th column of the plate but buffer). The plate was washed three times with 200 µl of ELISA buffer (100 mM KCl, 50 mM H_2KPO_4), shaking the plate for 1 min at 450 rpm during each wash.

The wells were then treated with 100 µl Block buffer (50 ml of Blocking buffer prepared by dissolving 1.5 g of BSA in ELISA buffer) for 1 h at RT. Meanwhile, a stock dilution of 200 nM BG4 was prepared. The blocking buffer was discarded from the wells, and replaced with 100 µl of fresh Blocking buffer, except for the first well. Following, 200 µl of the diluted BG4 was added into the first well. Serial dilutions of the antibody were added into each well by taking 100 µl of BG4 out from the first well and transferring it into the next well, pipetting up and down and repeating. To maintain the volumes consistent across the plate, 100 µl volume was removed from the last well. No BG4 was placed in the 12th column. The plate was then incubated for 1 h at RT.

After incubation, the samples were removed from the wells and the plate was washed three times with 200 µl ELISA buffer supplemented with 0.1% Tween-20, 1 min per wash, 450 rpm. The plate was then incubated with 100 µl of HRP-antibody (ab1238, ABCAM, UK) diluted at 1:20,000 in Blocking buffer for 1 h at RT.

The samples were discarded, and the plate was washed with 200 µl ELISA buffer supplemented with 0.1% Tween-20 three times for one min per wash at 450 rpm. The plate was dried carefully prior to the addition of 100 µl TMB (HRP substrate, Thermo Scientific) for at least 2 min. Finally, 50 µl of H_2SO_4 was added into each well.

The absorbance was measured at 450 nm and the data were plotted using one site-specific binding equation on GraphPad Prism 9.0.1 (see Section 2.2).

5.25 BG4 Immunofluorescence

The BG4 immunofluorescence protocol used was adapted from R. Hänsel-Hertsch and co-workers[71]. After fixation (described in Section 5.22), cells were blocked using 150 µl of Immunofluorescence block buffer (5% milk in PBS-T) for 1 h, in a humidity chamber at 37 °C. After incubation, the block buffer was aspirated and the cells were incubated with 150 µl of purified BG4 antibody diluted 1:100 in Immunofluorescence block buffer for 1 h at 37 °C, in the humidity chamber. Cells were then rinsed three times (5 min per wash) with 250 µl PBS-T at RT and incubated with 150 µl of rabbit anti-FLAG antibody (Cell Signalling Technology, #2368) diluted to 1:800 in Immunofluorescence block buffer for 1 h at 37 °C, in the humidity chamber. Cells were washed again three times (5 min per wash) with 250 µl PBS-T at RT. Cells were then incubated with 150 µl of Rhodamine anti-rabbit IgG (H+L) (Life Technologies, #R6394) antibody diluted 1:800 in Immunofluorescence block buffer for 1 h at 37 °C, in the humidity chamber. Then, cells were rinsed three times (5 min per wash) with 250 µl PBS-T at RT. The second wash contained Hoechst 33342 (Invitrogen) diluted 1:10000 in PBS-T. Finally, cells were rinsed three times with milliQ water and cellular imaging, performed using Nikon Ti2 Microscope (objective magnification: 60×, oil). Cellular localisation of CSB-HD (285–1,009) was detected by EGFP signal (EmW:515.0 nm, ExW:470 nm), BG4 staining was identified by Alexa Fluor 594 (EmW:624 nm; ExW:555 nm) signal, whilst the cell nuclei were detected using Hoechst 33342 staining (EmW:432.0 nm). Each immunofluorescence experiment was performed in two independent biological replicates, each consisting of two technical replicates. The quantification was performed on eight different images for each replicate.

5.26 Statistical Analysis

5.26.1 Resolvase Assays

Each experiment was performed in triplicates and analysed using GraphPad Prism 9.0.1. Data were presented as the mean ± SD between triplicates and statistical significance between two groups (presence/absence of CSB) was calculated based on a two-way Student's *t*-test [246], where the statistically significant difference was defined as $p<0.05$. The percentage of dsDNA formation was calculated from the ratio between the intensity of the newly formed dsDNA bands and fast-running G4 bands either in the presence or absence of CSB (-FL or -HD).

5.26.2 Cellular Localisation

Each experiment was performed in two biological and two technical replicates. The percentage of cells showing nucleolar localisation was calculated from the ratio between the number of cells with CSB-positive nucleoli and the total number of nucleoli present in the image. Statistically significant differences between the percentage of cells showing nucleolar localisation of CSB in the presence or absence of G4-ligands were calculated from eight different cellular images for each replicate using a paired two-tailed Student's *t*-test. Statistically significant difference was defined at $p<0.05$.

5.26.3 BG4 Immunofluorescence

The percentage of cells that lack nucleolar BG4 staining was calculated from the ratio between the number of nucleoli that lack BG4 staining and the total number of nucleoli (brightfield signal) observed in the microscopy image. Eight different images were collected for each tested condition. The percentage of cells lacking nucleolar BG4 was calculated for either CS1AN, CS1AN transfected with CSB-HD (285–1,009), and HeLa cells. Each experiment was performed in two technical and two biological replicates. The statistical significance ($p<0.05$) between the data was calculated using a two-way Student's *t*-test on GraphPad Prism 9.0.1 software.

6 Bibliography

1 Chargaff, E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **6**, 201-240 (1950).

2 Franklin, R. E. & Gosling, R. G. Molecular configuration in sodium thymonucleate. *Nature* **171**, 740-741 (1953).

3 Watson, J. D. & Crick, F. H. C. A structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).

4 Olins, D. E. & Olins, A. L. Chromatin history: our view from the bridge. *Nat. Rev. Mol. Cell Biol.* **4**, 809-814 (2003).

5 Luger, K., Dechassa, M. L. & Tremethick, D. J. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat. Rev. Mol. Cell Biol.* **13**, 436-447 (2012).

6 Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structureof thenucleosome core particle at 2.8A˚ resolution. *Nature* **389**, 251-260 (1997).

7 Kornberg, R. D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974).

8 Hergeth, S. P. & Schneider, R. The H1 linker histones: multifunctional proteins beyond the nucleosomal core particle. *EMBO Rep.* **16**, 1439-1453 (2015).

9 Fyodorov, D. V., Zhou, B. R., Skoultchi, A. I. & Bai, Y. Emerging roles of linker histones in regulating chromatin structure and function. *Nat. Rev. Mol. Cell Biol.* **19**, 192-206 (2018).

10 Tsompana, M. & Buck, M. J. Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin* **7**, 1-16 (2014).

11 Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207-220 (2019).

12 Tateishi-Karimata, H. & Sugimoto, N. Chemical biology of non-canonical structures of nucleic acids for therapeutic applications. *Chem. Commun. (Camb.)* **56**, 2379-2390 (2020).

13 Raguseo, F., Chowdhury, S., Minard, A. & Di Antonio, M. Chemical-biology approaches to probe DNA and RNA G-quadruplex structures in the genome. *Chem. Commun. (Camb.)* **56**, 1317-1324 (2020).

14 Sen, D. & Gilbert, W. A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature* **344**, 410-414 (1990).

15 Spiegel, J., Adhikari, S. & Balasubramanian, S. The structure and function of DNA G-quadruplexes. *Trends. Chem.* **2**, 123-136 (2019).

16 Asensio, J. L., Brown, T. & Lane, A. N. Solution conformation of a parallel DNA triple helix with 5′ and 3′ triplex–duplex junctions. *Structure* **7**, 1-11 (1999).

17 Phan, A. T. & Mergny, J. Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* **30**, 4618-4625 (2002).

18 Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13** (2012).

19 Pray, L. A. Discovery of DNA structure and function: Watson and Crick. *Nature Education* **1**, 100 (2008).

20 Narayana, N., Ginell, S. L., Russu, I. M. & Berman, H. M. Crystal and molecular structure of a DNA fragment: dCCGTGAATTCACG). *Biochemistry* **30**, 4449-4455, (1991).

21 Thompson, J. *et al.* Rosalind Franklin's X-ray photo of DNA as an undergraduate optical diffraction experiment. *American Journal of Physics* **86**, 95-104 (2018).

22 Gill, M. L., Strobel, S. A. & Loria, J. P. Crystallization and characterization of the thallium form of the Oxytricha nova G-quadruplex. *Nucleic Acids Res.* **34**, 4506-4514 (2006).

23 Rhee, S., Han, Z., Liu, K., Miles, H. T. & Davies, D. R. Structure of a triple helical DNA with a triplex-duplex junction. *Biochemistry* **38**, 16810-16815 (1999).

24 Weil, J. *et al.* Stabilization of the i-motif by intramolecular adenine±adenine±thymine base triple in the structure of d(ACCCT). *Acta Cryst.* **D55**, 422-429 (1998).

25 Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364-366 (1988).

26 Fay, M. M., Lyons, S. M. & Ivanov, P. RNA G-quadruplexes in biology: principles and molecular mechanisms. *J. Mol. Biol.* **429**, 2127-2147 (2017).

27 Gellert, M., Lipsett, M. N. & Davies, D. R. Helix formation by guanylic acid *Proc Natl Acad Sci USA* **48**, 2013-2018 (1962).

28 Zimmerman, S. B., Cohen, G. H. & Davies, D. R. X-ray fiber diffraction and model-building study of polyguanylic acid and aolyinosinic acid. *J. Mol. Biol.* **92**, 181-192 (1975).

29 Parkinson, G. N., Lee, M. P. H. & Neidle, S. Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **417**, 876-880 (2002).

30 Ortiz de Luzuriaga, I., Lopez, X. & Gil, A. Learning to Model G-Quadruplexes: Current Methods and Perspectives. *Annu. Rev. Biophys* **50**, 209-243 (2021).

31 Bhattacharyya, D., Mirihana Arachchilage, G. & Basu, S. Metal cations in G-quadruplex folding and stability. *Front. Chem.* **4**, 38 (31-14), doi:10.3389/fchem.2016.00038 (2016).

32 Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402-5415 (2006).

33 Pipier, A. *et al.* Constrained G4 structures unveil topology specificity of known and new G4 binding proteins. *Sci. Rep.* **11**, 13469 (2021).

34 Wong, H. M., Stegle, O., Rodgers, S. & Huppert, J. L. A toolbox for predicting g-quadruplex formation and stability. *J. Nucleic Acids* **2010** (2010).

35 Huppert, J. L. & Balasubramanian, S. Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **33**, 2908-2916 (2005).

36 Maizels, N. & Gray, L. T. The G4 genome. *PLoS Genet.* **9** (2013).

37 Hazel, P., Huppert, J., Balasubramanian, S. & Neidl, S. Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **126**, 16405-16415 (2004).

38 Johnson, F. B. Fundamentals of G-quadruplex biology. *Annu. Rep. Med. Chem.* **54**, 3-44 (2020).

39 Kolesnikova, S. & Curtis, E. A. Structure and function of multimeric G-quadruplexes. *Molecules* **24** (2019).

40 del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-quadruplex secondary structure obtained from circular dichroism spectroscopy. *Angewandte Chemie International Edition* **57**, 7171-7175 (2018).

41 Mergny, J.-L., Phanb, A.-T. & Lacroix, L. Following G-quartet formation by UV-spectroscopy. *FEBS Letters* **435**, 74-78 (1998).

42 Małgowska, M., Gudanis, D., Teubert, A., Dominiak, G. & Gdaniec, Z. How to study G-quadruplex structures. *BioTechnologia* **4**, 381-390 (2012).

43 Brunger, A. T. X-ray crystallography and NMR reveal complementary views of structure and dynamics. *Nat. Struct. Biol.* **4**, 862-865 (1997).

44 Mergny, J., Phan, A. & Lacroix, L. Following G-quartet formation by UV-spectroscopy. *FEBS Letters* **435** 74-78 (1998).

45 Mergny, J. L., Li, J., Lacroix, L., Amrane, S. & Chaires, J. B. Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* **33**, e138 (2005).

46 Kypr, J., Kejnovska, I., Renciuk, D. & Vorlickova, M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res.* **37**, 1713-1725 (2009).

47 Phan, A. T., Modi, Y. S. & Dinshaw, J. P. Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *JACS* **126**, 8710-8716 (2004).

48 Ambrus, A., Chen, D., Dai, J., Jones, R. A. & Yang, D. Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry* **44**, 2048-2058 (2005).

49 Stump, S., Mou, T. C., Sprang, S. R., Natale, N. R. & Beall, H. D. Crystal structure of the major quadruplex formed in the promoter region of the human c-MYC oncogene. *PLoS One* **13**, 1-15 (2018).

50 Blackburn, E. H. Structure and function of telomeres. *Nature* **350**, 569-573 (1991).

51 Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I. J. & Blackburn, E. H. Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine Base pairs. *Cell* **51**, 899-908 (1987).

52 Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887-3896 (2006).

53 Huppert, J. L. & Balasubramanian, S. G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.* **35**, 406-413 (2007).

54 Bedrat, A., Lacroix, L. & Mergny, J. L. Re-evaluation of G-quadruplex propensity with G4Hunter. *Nucleic Acids Res.* **44**, 1746-1759 (2016).

55 Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877-881 (2015).

56 Weitzmann, M. N., Woodford, K. J. & Usdin, K. The development and use of a DNA polymerase arrest assay for the evaluation of parameters affecting intrastrand tetraplex formation. *J. Biol. Chem.* **271**, 20958-20964 (1996).

57 Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456**, 53-59 (2008).

58 Han, H., Hurley, L. H. & Salazar, M. A DNA polymerase stop assay for G-quadruplex-interactive compounds. *Nucleic Acid Res.* **27**, 537–542 (1999).

59 Head, S. R. *et al.* Library construction for next-generation sequencing: overviews and challenges. *Biotechniques* **56**, 61-77 (2014).

60 Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* **47**, 3862-3874 (2019).

61 Mukundan, V. T. & Phan, A. T. Bulges in G-quadruplexes: broadening the definition of G-quadruplex-forming sequences. *J. Am. Chem. Soc.* **135**, 5017-5028 (2013).

62 Biffi, G., Tannahill, D., McCafferty, J. & Balasubramanian, S. Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.* **5**, 182-186 (2013).

63 Summers, P. A. *et al.* Visualising G-quadruplex DNA dynamics in live cells by fluorescence lifetime imaging microscopy. *Nat. Commun.* **12**, 162 (161-111) (2021).

64 Shivalingam, A. *et al.* The interactions between a small molecule and G-quadruplexes are visualized by fluorescence lifetime imaging microscopy. *Nat. Commun.* **6**, 8178 (8171-8110) (2015).

65 Di Antonio, M. *et al.* Single-molecule visualization of DNA G-quadruplex formation in live cells. *Nat. Chem.* **12**, 832-837 (2020).

66 Schofield, D. J. *et al.* Application of phage display to high throughput antibody generation and characterization. *Genome. Biol.* **8**, R254 (2007).

67 Biffi, G., Di Antonio, M., Tannahill, D. & Balasubramanian, S. Visualization and selective chemical targeting of RNA G-quadruplex structures in the cytoplasm of human cells. *Nat. Chem.* **6**, 75-80 (2014).

68 Javadekar, S. M., Nilavar, N. M., Paranjape, A., Das, K. & Raghavan, S. C. Characterization of G-quadruplex antibody reveals differential specificity for G4 DNA forms. *DNA Res.* **27** (2020).

69 Paeschke, K., Capra, J. A. & Zakian, V. A. DNA replication through G-quadruplex motifs is promoted by the Saccharomyces cerevisiae Pif1 DNA helicase. *Cell* **145**, 678-691 (2011).

70 Gray, L. T., Vallur, A. C., Eddy, J. & Maizels, N. G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **10**, 313-318 (2014).

71 Hansel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267-1272 (2016).

72 Hansel-Hertsch, R., Spiegel, J., Marsico, G., Tannahill, D. & Balasubramanian, S. Genome-wide mapping of endogenous G-quadruplex DNA structures by chromatin immunoprecipitation and high-throughput sequencing. *Nat. Protoc.* **13**, 551-564 (2018).

73 Shen, J. *et al.* Promoter G-quadruplex folding precedes transcription and is controlled by chromatin. *Genome Biol* **22**, 143 (141-114) (2021).

74 Raiber, E. A., Kranaster, R., Lam, E., Nikan, M. & Balasubramanian, S. A non-canonical DNA structure is a binding motif for the transcription factor SP1 in vitro. *Nucleic Acids Res.* **40**, 1499-1508 (2012).

75 Cogoi, S., Paramasivam, M., Membrino, A., Yokoyama, K. K. & Xodo, L. E. The KRAS promoter responds to Myc-associated zinc finger and poly(ADP-ribose) polymerase 1 proteins, which recognize a critical quadruplex-forming GA-element. *J. Biol. Chem.* **285**, 22003-220016 (2010).

76 Johnson, J. E., Cao, K., Ryvkin, P., Wang, L. S. & Johnson, F. B. Altered gene expression in the Werner and Bloom syndromes is associated with sequences having G-quadruplex forming potential. *Nucleic Acids Res.* **38**, 1114-1122 (2010).

77 Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627-8637 (2015).

78 Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell. Biol.* **21**, 459-474 (2020).

79 Du, Q., Wang, Z. & Schramm, V. L. Human DNMT1 transition state structure. *Proc. Natl. Acad. Sci. USA* **113**, 2916-2921 (2016).

80 Mao, S. *et al.* DNA G-quadruplex structures mold the DNA methylome. *Nat. Struct. Mol. Biol.* **25**, 951-957 (2018).

81 Robinson, J., Raguseo, F., Nuccio, S. P., Liano, D. & Di Antonio, M. DNA G-quadruplex structures: more than simple roadblocks to transcription? *Nucleic Acids Res.* **49**, 8419-8431 (2021).

82 Moyzis, R. *et al.* A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc. Natl. Acad. Sci. USA* **85**, 6622-6626 (1988).

83 Wright, E. W., Tesmer, M. T., Huffman, K. E., Levene, S. D. & Shay, J. W. Normal human chromosomes have long G-rich telomeric overhangs at one end. *GENES & DEVELOPMENT* **11**, 2801–2809 (1997).

84 Wang, Y. & Patell, J. P. Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex. *Structure* **1**, 263-282 (1993).

85 Lin, C. & Yang, D. Human telomeric G-quadruplex structures and G-quadruplex-interactive compounds. *Methods Mol. Biol.* **1587**, 171-196 (2017).

86 Dai, J., Carver, M. & Yang, D. Polymorphism of human telomeric quadruplex structures. *Biochimie* **90**, 1172-1183 (2008).

87 Harley, C. B., Futchert, A. B. & Greidert, C. W. Telomeres shorten during ageing of human fibroblasts. *Nature* **345**, 458-460 (1990).

88 Greider, C. W. & Blackburn, E. H. Identification of a specific telomere terminal transferase activity in Tetrahymena extracts. *Cell* **43**, 405-413 (1985).

89 Kim, N. W. *et al.* Specific association of human telomerase activity with immortal cells and cancer. *Science* **266**, 2011-2015 (1994).

90 Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DNA structures. *Nature* **350**, 718-720 (1991).

91 Palumbo, S. L., Ebbinghaus, S. W. & Hurley, L. H. Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J. Am. Chem. Soc.* **131**, 10878-10891 (2009).

92 Kaguni, L. S. & Clayton, D. Template-directed pausing in *in vitro* DNA synthesis by DNA polymerase a from *Drosophila melanogaster* embryos. *Proc. NatL Acad. Sci. USA* **79**, 983-987 (1982).

93 Kamath-Loeb, A. S., Loeb, L. A., Johansson, E., Burgers, P. M. J. & Fry, M. Interactions between the Werner Syndrome helicase and DNA polymerase δ specifically facilitate copying of tetraplex and hairpin structures of the d(CGG) trinucleotide repeat sequence. *Journal of Biological Chemistry* **276**, 16439-16446 (2001).

94 Lambert, S., Watson, A., Sheedy, D. M., Martin, B. & Carr, A. M. Gross chromosomal rearrangements and elevated recombination at an inducible site-specific replication fork barrier. *Cell* **121**, 689-702 (2005).

95 Valton, A. L. & Prioleau, M. N. G-quadruplexes in DNA replication: a problem or a necessity? *Trends Genet.* **32**, 697-706 (2016).

96 Kruisselbrink, E. *et al.* Mutagenic capacity of endogenous G4 DNA underlies genome instability in FANCJ-defective *C. elegans*. *Current Biology* **18**, 900-905 (2008).

97 Wyatt, D. W. *et al.* Essential roles for polymerase theta-mediated end joining in the repair of chromosome breaks. *Mol. Cell.* **63**, 662-673 (2016).

98 Chang, H. H. Y., Pannunzio, N. R., Adachi, N. & Lieber, M. R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat. Rev. Mol. Cell Biol.* **18**, 495-506 (2017).

99 van Gent, D. C., Hoeijmakers, J. H. J. & Kanaar, R. Chromosomal stability and the DNA double-stranded break connection. *Nat. Rev. Genet.* **2**, 196–206 (2001).

100 Lemmens, B., van Schendel, R. & Tijsterman, M. Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat. Commun.* **6**, 8909 (8901-8908) (2015).

101 Nambiar, M. *et al.* Formation of a G-quadruplex at the BCL2 major breakpoint region of the t(14;18) translocation in follicular lymphoma. *Nucleic Acids Res.* **39**, 936-948 (2011).

102 Kim, N. The Interplay between G-quadruplex and transcription. *Curr. Med. Chem.* **26**, 2898-2917 (2019).

103 Dahan, D. *et al.* Pif1 is essential for efficient replisome progression through lagging strand G-quadruplex DNA secondary structures. *Nucleic Acids Res.* **46**, 11847-11857 (2018).

104 Lopes, J. *et al.* G-quadruplex-induced instability during leading-strand replication. *EMBO J.* **30**, 4033-4046 (2011).

105 Lerner, L. K. & Sale, J. E. Replication of G Quadruplex DNA. *Genes (Basel)* **10** (2019).

106 Lago, S. *et al.* Promoter G-quadruplexes and transcription factors cooperate to shape the cell type-specific transcriptome. *Nat. Commun.* **12**, 3885 (3881-3813) (2021).

107 Broxson, C., Beckett, J. & Tornaletti, S. Transcription arrest by a G quadruplex forming-trinucleotide repeat sequence from the human c-myb gene. *Biochemistry* **50**, 4162-4172 (2011).

108 Tian, T., Cheng, Y.-Q., Wang, S.-R. & Zhou, X. G-Quadruplex: A Regulator of Gene Expression and Its Chemical Targeting. *Chem* **4**, 1314-1344 (2018).

109 Chen, S. *et al.* Mechanistic studies for the role of cellular nucleic-acid-binding protein (CNBP) in regulation of c-myc transcription. *Biochim. Biophys. Acta* **1830**, 4769-4777 (2013).

110 Thakur, R. K. *et al.* Metastases suppressor NM23-H2 interaction with G-quadruplex DNA within c-MYC promoter nuclease hypersensitive element induces c-MYC expression. *Nucleic Acids Res.* **37**, 172-183 (2009).

111 Zheng, K. W. *et al.* Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.* **41**, 5533-5541 (2013).

112 Belotserkovskii, B. P., Soo Shin, J. H. & Hanawalt, P. C. Strong transcription blockage mediated by R-loop formation within a G-rich homopurine-homopyrimidine sequence localized in the vicinity of the promoter. *Nucleic Acids Res.* **45**, 6589-6599 (2017).

113 De Magis, A. *et al.* DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl. Acad. Sci. USA* **116**, 816-825 (2019).

114 Lee, C. Y. *et al.* R-loop induced G-quadruplex in non-template promotes transcription by successive R-loop formation. *Nat. Commun.* **11**, 3392 (3391-3315) (2020).

115 Vlasenok, M. *et al.* Data set on G4 DNA interactions with human proteins. *Data Brief.* **18**, 348-359 (2018).

116 Reina, C. & Cavalieri, V. Epigenetic modulation of chromatin states and gene expression by G-quadruplex structures. *Int. J. Mol. Sci.* **21** (2020).

117 Clark, D. W., Phang, T., Edwards, M. G., Geraci, M. W. & Gillespie, M. N. Promoter G-quadruplex sequences are targets for base oxidation and strand cleavage during hypoxia-induced transcription. *Free Radic. Biol. Med.* **53**, 51-59 (2012).

118 Finkel, T., Serrano, M. & Blasco, M. A. The common biology of cancer and ageing. *Nat. Rev.* **448**, 767-774 (2007).

119 Fleming, A. M., Zhu, J., Ding, Y. & Burrows, C. J. Location dependence of the transcriptional response of a potential G-quadruplex in gene promoters under oxidative stress. *Nucleic Acids Res.* **47**, 5049-5060 (2019).

120 Fleming, A. M., Ding, Y. & Burrows, C. J. Oxidative DNA damage is epigenetic by regulating gene transcription via base excision repair. *Proc. Natl. Acad. Sci. USA* **114**, 2604-2609 (2017).

121 Malousi, A. *et al.* Age dependent methylation in epigenetic clock CpGs is associated with G quadruplex co transcriptionally formed RNA structures and tentative splice. *Epigenetics* **13**, 808-821 (2018).

122 Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug. Discov.* **10**, 261-275 (2011).

123 Sun, D., Guo, K., Rusche, J. J. & Hurley, L. H. Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **33**, 6070-6080 (2005).

124 Dexheimer, T. S., Sun, D. & Hurley, L. H. Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* **128**, 5404-5015 (2006).

125 Cogoi, S. & Xodo, L. E. G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **34**, 2536-2549 (2006).

126 Rankin, S. *et al.* Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **127**, 10584–10589 (2005).

127 Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. USA* **99**, 11593-11598 (2002).

128 Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. G-quadruplexes: a promising target for cancer therapy. *Mol. Cancer.* **20**, 40 (41-18) (2021).

129 Dang, C. V. MYC on the path to cancer. *Cell* **149**, 22-35 (2012).

130 Collie, G. W. & Parkinson, G. N. The application of DNA and RNA G-quadruplexes to therapeutic medicines. *Chem. Soc. Rev.* **40**, 5867-5892 (2011).

131 Gomez, D. *et al.* Telomestatin-induced telomere uncapping is modulated by POT1 through G-overhang extension in HT1080 human tumor cells. *J. Biol. Chem.* **281**, 38721-38729 (2006).

132 Kim, M.-Y., Vankayalapati, H., Shin-ya, K., Wierzba, K. & Hurley, L. H. Telomestatin, a potent telomerase inhibitor that interacts quite specifically with the human telomeric intramolecular G-quadruplex. *J. Am. Chem. Soc.* **124**, 2098–2099 (2001).

133 Read, M. *et al.* Structure-based design of selective and potent G-quadruplex-mediated telomerase inhibitors. *Proc Natl Acad Sci USA* **98**, 4844 – 4849 (2001).

134 Zhou, G. *et al.* Telomere targeting with a novel G-quadruplex-interactive ligand BRACO-19 induces T-loop disassembly and telomerase displacement in human glioblastoma cells. *Oncotarget* **7**, 14925-14939 (2016).

135 Rodriguez, R. *et al.* A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **130**, 15758-15789 (2008).

136 Rodriguez, R. *et al.* Small-molecule-induced DNA damage identifies alternative DNA structures in human genes. *Nat. Chem. Biol.* **8**, 301-310 (2012).

137 Drygin, D. *et al.* Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res.* **71**, 1418-1430 (2011).

138 Xu, H. *et al.* CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.* **8**, 14432 (14431-14418) (2017).

139 Han, F. X., Wheelhouse, R. T. & Hurley, L. H. Interactions of TMPyP4 and TMPyP2 with quadruplex DNA. Structural basis for the differential effects on telomerase inhibition. *J. Am. Chem. Soc.* **121** (1999).

140 Li, Q. *et al.* G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* **41**, D1115-1123 (2013).

141 Monchaud, D. & Teulade-Fichou, M. P. A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem.* **6**, 627-636 (2008).

142 Martino L, Pagano B, Fotticchia I, Neidle S & C., G. Shedding light on the interaction between TMPyP4 and human telomeric quadruplexes. *J. Phys. Chem. B.* **113**, 14779-14786 (2009).

143 Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071-1078 (2009).

144 Moruno-Manchon, J. F. *et al.* The G-quadruplex DNA stabilizing drug pyridostatin promotes DNA damage and downregulates transcription of Brca1 in neurons. *Aging* **9**, 1957–1970 (2017).

145 Drygin, D. *et al.* Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res.* **69**, 7653-7661 (2009).

146 Quin, J. E. *et al.* Targeting the nucleolus for cancer intervention. *Biochim. Biophys. Acta* **1842**, 802-816 (2014).

147 Bywater, M. J. *et al.* Inhibition of RNA polymerase I as a therapeutic strategy to promote cancer-specific activation of p53. *Cancer Cell* **22**, 51-65 (2012).

148 Sanij, E. *et al.* CX-5461 activates the DNA damage response and demonstrates therapeutic efficacy in high-grade serous ovarian cancer. *Nat. Commun.* **11**, 2641 (2641-2618) (2020).

149 Sauer, M. & Paeschke, K. G-quadruplex unwinding helicases and their function *in vivo*. *Biochem. Soc. Trans.* **45**, 1173-1182 (2017).

150 Castillo Bosch, P. *et al.* FANCJ promotes DNA synthesis through G-quadruplex structures. *EMBO J.* **33**, 2521-2533 (2014).

151 London, T. B. *et al.* FANCJ is a structure-specific DNA helicase associated with the maintenance of genomic G/C tracts. *J. Biol. Chem.* **283**, 36132-36139 (2008).

152 Crabbe, L., Verdun, R. E., Haggblom, C. I. & Karlseder, J. Defective telomere lagging strands synthesis in cells lacking WRN helicase activity. *Science* **306**, 1951-1953 (2004).

153 Drosopoulos, W. C., Kosiyatrakul, S. T. & Schildkraut, C. L. BLM helicase facilitates telomere replication during leading strand synthesis of telomeres. *J. Cell. Biol.* **210**, 191-208 (2015).

154 Chen, M. C. *et al.* Structural basis of G-quadruplex unfolding by the DEAH/RHA helicase DHX36. *Nature* **558**, 465-469 (2018).

155 Schult, P. & Paeschke, K. The DEAH helicase DHX36 and its role in G-quadruplex-dependent processes. *Biol. Chem.* **402**, 581-591 (2021).

156 Chen, M. C., Murat, P., Abecassis, K., Ferre-D'Amare, A. R. & Balasubramanian, S. Insights into the mechanism of a G-quadruplex-unwinding DEAH-box helicase. *Nucleic Acids Res.* **43**, 2223-2231 (2015).

157 Vaughn, J. P. *et al.* The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* **280**, 38117-38120 (2005).

158 Creacy, S. D. *et al.* G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* **283**, 34626-34634 (2008).

159 Yangyuoru, P. M., Bradburn, D. A., Liu, Z., Xiao, T. S. & Russell, R. The G-quadruplex (G4) resolvase DHX36 efficiently and specifically disrupts DNA G4s via a translocation-based helicase mechanism. *J. Biol. Chem.* **293**, 1924-1932 (2018).

160 Lattmann, S., Giri, B., Vaughn, J. P., Akman, S. A. & Nagamine, Y. Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.* **38**, 6219-6233 (2010).

161 Sato, K., Martin-Pintado, N., Post, H., Altelaar, M. & Knipscheer, P. Multistep mechanism of G-quadruplex resolution during DNA replication. *Sci. Adv.* **7**, 1-16 (2021).

162 Li, X. *et al.* Structure, interactions and effects on activity of the 5'-terminal region of human telomerase RNA. *J. Biochem.* **141**, 755-765 (2007).

163 Park, S., Park, H.-E. H., Son, H. G. & Lee, S.-J. V. The role of RNA helicases in aging and lifespan regulation. *Translational Medicine of Aging* **1**, 24-31 (2017).

164 Law, M. J. *et al.* ATR-X syndrome protein targets tandem repeats and influences allele-specific expression in a size-dependent manner. *Cell* **143**, 367-378 (2010).

165 Mitson, M., Kelley, L. A., Sternberg, M. J., Higgs, D. R. & Gibbons, R. J. Functional significance of mutations in the Snf2 domain of ATRX. *Hum. Mol. Genet.* **20**, 2603-2610 (2011).

166 Tiwari, V., Baptiste, B. A., Okur, M. N. & Bohr, V. A. Current and emerging roles of Cockayne syndrome group B (CSB) protein. *Nucleic Acids Res.* **49**, 2418-2434 (2021).

167 Troelstra, C. *et al.* Molecular cloning of the human DNA excision repair gene ERCC-6. *Molecular and Cellular Biology* **10**, 5806-5813 (1990).

168 Scheibye-Knudsen, M. *et al.* Cockayne syndrome group A and B proteins converge on transcription-linked resolution of non-B DNA. *Proc. Natl. Acad. Sci. USA* **113**, 12502-12507 (2016).

169 Datta, A., Pollock, K. J., Kormuth, K. A. & Brosh, R. M., Jr. G-quadruplex assembly by ribosomal DNA: emerging roles in disease pathogenesis and cancer biology. *Cytogenet Genome Res.* **161**, 285-296 (2021).

170 Sunesen, M. *et al.* Molecular characterization of an acidic region deletion mutant of Cockayne syndrome group B protein. *Nucleic Acids Res.* **28**, 3151–3159 (2000).

171 Lake, R. J., Geyko, A., Hemashettar, G., Zhao, Y. & Fan, H. Y. UV-induced association of the CSB remodeling protein with chromatin requires ATP-dependent relief of N-terminal autorepression. *Mol. Cell.* **37**, 235-246 (2010).

172 Takahashi, T. S. *et al.* Structural basis of ubiquitin recognition by the winged-helix domain of Cockayne syndrome group B protein. *Nucleic Acids Res.* **47**, 3784-3794 (2019).

173 Anindya, R. *et al.* A ubiquitin-binding domain in Cockayne syndrome B required for transcription-coupled nucleotide excision repair. *Mol. Cell.* **38**, 637-648 (2010).

174 van der Weegen, Y. *et al.* The cooperative action of CSB, CSA, and UVSSA target TFIIH to DNA damage-stalled RNA polymerase II. *Nat. Commun.* **11**, 2104 (2101-2116) (2020).

175 Christiansen, M., Thorslund, T., Jochimsen, B., Bohr, V. A. & Stevnsner, T. The Cockayne syndrome group B protein is a functional dimer. *FEBS J.* **272**, 4306-4314 (2005).

176 Selby, C. P. & Sancar, A. Human transcription-repair coupling factor CSB/ERCC6 is a DNA-stimulated ATPase but is not a helicase and does not disrupt the ternary transcription complex of stalled RNA polymerase II. *J. Biol. Chem.* **272**, 1885-1890 (1997).

177 Beerens, N., Hoeijmakers, J. H., Kanaar, R., Vermeulen, W. & Wyman, C. The CSB protein actively wraps DNA. *J. Biol. Chem.* **280**, 4722-4729 (2005).

178 Iyama, T. *et al.* Regulation of the Intranuclear Distribution of the Cockayne Syndrome Proteins. *Sci. Rep.* **8**, 17490 (17491-17412) (2018).

179 Kokic, G., Wagner, F. R., Chernev, A., Urlaub, H. & Cramer, P. Structural basis of human transcription-DNA repair coupling. *Nature* **598** (2021).

180 Mao, P., Wyrick, J. J., Roberts, S. A. & Smerdon, M. J. UV-Induced DNA Damage and Mutagenesis in Chromatin. *Photochem. Photobiol.* **93**, 216-228 (2017).

181 Citterio, E. *et al.* ATP-dependent chromatin remodeling by the Cockayne Syndrome B DNA repair-transcription-coupling factor. *Molecular and Cellular Biology* **20**, 7643–7653 (2000).

182 Duan, M., Speer, R. M., Ulibarri, J., Liu, J. K. & Mao, P. Transcription-coupled nucleotide excision repair: New insights revealed by genomic approaches. *DNA Repair* **103**, 1-8 (2021).

183 Xu, J. *et al.* Structural basis for the initiation of eukaryotic transcription-coupled DNA repair. *Nature* **551**, 653-657 (2017).

184 Selby, C. P. & Sancar, A. Cockayne syndrome group B protein enhances elongation by RNA polymerase II. *Proc. Natl. Acad. Sci. USA* **94**, 11205–11209 (1997).

185 Lee, S. K., Yu, S. L., Prakash, L. & Prakash, S. Requirement for yeast RAD26, a homolog of the human CSB gene, in elongation by RNA polymerase II. *Mol. Cell Biol.* **21**, 8651-8656 (2001).

186 Wilson, B. T. *et al.* The Cockayne Syndrome Natural History (CoSyNH) study: clinical findings in 102 individuals and recommendations for care. *Genet. Med.* **18**, 483-493 (2016).

187 Lee, J. H., Kim, E. W., Croteau, D. L. & Bohr, V. A. Heterochromatin: an epigenetic point of view in aging. *Exp. Mol. Med.* **52**, 1466-1474 (2020).

188 Laugel, V. *et al.* Mutation update for the CSB/ERCC6 and CSA/ERCC8 genes involved in Cockayne syndrome. *Hum. Mutat.* **31**, 113-126 (2010).

189 Wang, S. *et al.* Rescue of premature aging defects in Cockayne syndrome stem cells by CRISPR/Cas9-mediated gene correction. *Protein Cell* **11**, 1-22 (2020).

190 Thorslund, T. *et al.* Cooperation of the Cockayne syndrome group B protein and poly(ADP-ribose) polymerase 1 in the response to oxidative stress. *Mol. Cell. Biol.* **25**, 7625-7636 (2005).

191 Boetefuer, E. L., Lake, R. J., Dreval, K. & Fan, H. Y. Poly(ADP-ribose) polymerase 1 (PARP1) promotes oxidative stress-induced association of Cockayne syndrome group B protein with chromatin. *J. Biol. Chem.* **293**, 17863-17874 (2018).

192 Lee, J. H. *et al.* Cockayne syndrome group B deficiency reduces H3K9me3 chromatin remodeler SETDB1 and exacerbates cellular aging. *Nucleic Acids Res.* **47**, 8548-8562 (2019).

193 Crochemore, C., Fernandez-Molina, C., Montagne, B., Salles, A. & Ricchetti, M. CSB promoter downregulation via histone H3 hypoacetylation is an early determinant of replicative senescence. *Nat. Commun.* **10**, 5576 (5571-5517) (2019).

194 Okur, M. N. *et al.* Cockayne syndrome group A and B proteins function in rRNA transcription through nucleolin regulation. *Nucleic Acids Res.* **48**, 2473-2485 (2020).

195 Chiarella, S. *et al.* Nucleophosmin mutations alter its nucleolar localization by impairing G-quadruplex binding at ribosomal DNA. *Nucleic Acids Res.* **41**, 3228-3239 (2013).

196 Edwards, A. D., Marecki, J. C., Byrd, A. K., Gao, J. & Raney, K. D. G-Quadruplex loops regulate PARP-1 enzymatic activation. *Nucleic Acids Res.* **49** (2021).

197 Liano, D., Chowdhury, S. & Di Antonio, M. Cockayne Syndrome B protein selectively resolves and interact with intermolecular DNA G-quadruplex structures. *J. Am. Chem. Soc.*, 1-15 (2021).

198 Bornhorst, J. A. & Falke, J. J. Purification of proteins using polyhistidine affinity tags. *Methods Enzymol.* **326**, 245–254 (2000).

199 Farooqui, A. A. Purification of enzymes by heparin-sepharose affinity chromatography. *Journal of cromatography A* **184**, 335-345 (1980).

200 Bolten, S. N., Rinas, U. & Scheper, T. Heparin: role in protein purification and substitution with animal-component free material. *Appl. Microbiol. Biotechnol.* **102**, 8647-8660 (2018).

201 Hong, P., Koza, S. & Bouvier, E. S. P. A review size-exclusion chromatography for the analysis of protein biotherapeutics and their aggregates. *Journal of Liquid Chromatography & Related Technologies* **35**, 2923-2950 (2012).

202 Corzo, J. Time, the forgotten dimension of ligand binding teaching. *Biochemistry and Molecular Biology Education* **34**, 413–416 (2006).

203 Bjorkelund, H., Gedda, L. & Andersson, K. Comparing the epidermal growth factor interaction with four different cell lines: intriguing effects imply strong dependency of cellular context. *PLoS One* **6**, e16536 (2011).

204 Hellman, L. M. & Fried, M. G. Electrophoretic mobility shift assay (EMSA) for detecting protein-nucleic acid interactions. *Nat. Protoc.* **2**, 1849-1861 (2007).

205 Du, Y. Fluorescence polarization assay to quantify protein-protein interactions in an HTS format. *Methods in Molecular Biology* **1278**, 529-544 (2015).

206 Dharadhar, S., Kim, R. Q., Uckelmann, M. & Sixma, T. K. Quantitative analysis of USP activity in vitro. *Methods Enzymol.* **618**, 281-319 (2019).

207 Aydin, S. A short history, principles, and types of ELISA, and our laboratory experience with peptide/protein analyses using ELISA. *Peptides* **72**, 4-15 (2015).

208 Engvall, E. The ELISA, enzyme-linked immunosorbent assay. *Clin. Chem.* **56**, 319-320 (2010).

209 Lee, C. Y., McNerney, C. & Myong, S. G-quadruplex and protein binding by single-molecule FRET microscopy. *Methods Mol. Biol.* **2035**, 309-322 (2019).

210 Wang, K., Flaherty, D. P., Chen, L. & Yang, D. High-throughput screening of G-quadruplex ligands by FRET assay. *G-Quadruplex Nucleic Acids* **2035**, 323-331 (2019).

211 Renciuk, D. *et al.* A FRET-based screening assay for nucleic acid ligands. *Methods* **57**, 122-128 (2012).

212 De Rache, A. & Mergny, J. L. Assessment of selectivity of G-quadruplex ligands via an optimised FRET melting assay. *Biochimie* **115**, 194-202 (2015).

213 Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 10. *Nature* **429**, 375-381 (2004).

214 Citterio, E. *et al.* Biochemical and biological characterization of wild-type and ATPase-deficient Cockayne syndrome B repair protein. *J. Biol. Chem.* **273**, 11844-11851 (1998).

215 Kost, T. A., Condreay, J. P. & Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* **23**, 567-575 (2005).

216 Luckow, E. A., Lee, S. C., Barry, G. F. & Olins, P. O. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in Escherichia coli. *Journal of Virology* **67**, 4566-4579 (1993).

217 Unger, T. & Peleg, Y. Recombinant protein expression in the baculovirus-infected insect cell system. *Chemical Genomics and Proteomics* 187-199.

218 Zhou, J. & Blissard, G. W. Mapping the conformational epitope of a neutralizing antibody (AcV1) directed against the AcMNPV GP64 protein. *Virology* **352**, 427-437 (2006).

219 Invitrogen, L. t. Guide to baculovirus expression vector systems (BEVS) and insect cell culture techniques. *Instruction Manual*.

220 Cytiva. Superdex 200 Increase columns. *Manual*.

221 Cytiva. Superose 6 Increase columns. *Manual*.

222 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (531-536) (2011).

223 Ghosh-Roy, S., Das, D., Chowdhury, D., Smerdon, M. J. & Chaudhuri, R. N. Rad26, the transcription-coupled repair factor in yeast, is required for removal of stalled RNA polymerase-II following UV irradiation. *PLoS One* **8**, e72090 (2013).

224 Alexaki, A. *et al.* Effects of codon optimization on coagulation factor IX translation and structure: Implications for protein and gene therapies. *Sci. Rep.* **9**, 15449 (15441-15415) (2019).

225 Kuo, D., Nie, M. & Courey, A. J. SUMO as a solubility tag and in vivo cleavage of SUMO fusion proteins with Ulp1. *Protein Affinity Tags*, 71-80 (2014).

226 Jones, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195±202 (1999).

227 Mendoza, O., Bourdoncle, A., Boule, J. B., Brosh, R. M., Jr. & Mergny, J. L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989-2006 (2016).

228 Huppert, J. L. Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* **37**, 1375-1384, doi:10.1039/b702491f (2008).

229 Del Villar-Guerra, R., Trent, J. O. & Chaires, J. B. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew. Chem. Int. Ed. Engl.* **57**, 7171-7175, doi:10.1002/anie.201709184 (2018).

230 Lee, P. Y., Costumbrado, J., Hsu, C. Y. & Kim, Y. H. Agarose gel electrophoresis for the separation of DNA fragments. *J. Vis. Exp.* (2012).

231 Sabharwal, N. C. *et al.* N-methylmesoporphyrin IX fluorescence as a reporter of strand orientation in guanine quadruplexes. *FEBS J.* **281**, 1726-1737 (2014).

232 Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402-5415, doi:10.1093/nar/gkl655 (2006).

233 Guo, X., Liu, S., . & Yu, Z. Bimolecular quadruplexes and their transitions to higher-order molecular structures detected by ESI-FTICR-MS. *J. Am. Soc. Mass Spectrom.* **18**, 467–1476, doi:doi:10.1016/j.jasms.2007.05.003 (2007).

234 Iyama, T. *et al.* Regulation of the Intranuclear Distribution of the Cockayne Syndrome Proteins. *Sci. Rep.* **8**, 17490, doi:10.1038/s41598-018-36027-6 (2018).

235 Xu, H. *et al.* CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun.* **8**, 14432, doi:10.1038/ncomms14432 (2017).

236 Troelstra, C. *et al.* ERCC6, a member of a subfamily of putative helicases, is involved in Cockaynes Syndrome and preferential repair of active genes. *Cell* **71**, 939-953 (1992).

237 Maiser, A. *et al.* Super-resolution in situ analysis of active ribosomal DNA chromatin organization in the nucleolus. *Scientific Reports* **10** (2020).

238 Rodriguez, R. *et al.* A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J. Am. Chem. Soc.* **130**, 15758-15759, doi:10.1021/ja805615w (2008).

239 Mitteaux, J. *et al.* Identifying G-Quadruplex-DNA-Disrupting Small Molecules. *J. Am. Chem. Soc.* **143**, 12567-12577 (2021).

240 Belton, J. M. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268-276 (2012).

241 Shu, B. *et al.* Transcriptome analysis of Spodoptera frugiperda Sf9 cells reveals putative apoptosis-related genes and a preliminary apoptosis mechanism induced by azadirachtin. *Sci. Rep.* **7**, 13231 (13231-13213) (2017).

242 van Gool, A. J. *et al.* The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. *The EMBO Journal* **16**, 5955–5965 (1997).

243 Cho, I., Tsai, P. F., Lake, R. J., Basheer, A. & Fan, H. Y. ATP-dependent chromatin remodeling by Cockayne syndrome protein B and NAP1-like histone chaperones is

required for efficient transcription-coupled DNA repair. *PLoS Genet* **9**, e1003407 (2013).

244 UniProt, C. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480-D489 (2021).

245 van den Ent, F. & Lowe, J. RF cloning: a restriction-free method for inserting target genes into plasmids. *J. Biochem. Biophys. Methods* **67**, 67-74 (2006).

246 Huang, J. W. *et al.* MCM8IP activates the MCM8-9 helicase to promote DNA synthesis and homologous recombination upon DNA damage. *Nat. Commun.* **11**, 2948 (2941-2918) (2020).