Investigating a human DNA repair helicase that supports CRISPR editing

Andrew P. Cubbon, MSci.(Hons)

Thesis submitted to the University of Nottingham for the degree of

Doctor of Philosophy

School of Life Sciences

University of Nottingham

United Kingdom

August 2021

Acknowledgements

Firstly, I would like to thank my funders, the BBSRC and the University of Nottingham, for providing the means to carry out this project. I would also like to thank Dr. Ed Bolt for his supervision and his support through my extension period. His guidance during this time has been essential to submitting this thesis. Further thanks go to Prof. Thorsten Allers, Rachel Van Krimpen and Prof. James McInerney for their support in securing my extension.

Next, I would like to thank all members of the Bolt and Chalmers labs, past and present, for their help and support. Particular thanks go to Tabi for providing HelQ protein despite its scarcity, and to Jack, Ryan B and Ryan R for their moral support in and out of the lab. Re-establishing human cell culture in the Bolt lab, with no prior experience, was incredibly daunting and would not have been possible without the help I received from George Blundell-Hunter, Deb Briggs and Colin Nicholson.

I knew that completing a PhD would be hard, but did not expect the rollercoaster of success, failure and personal circumstances I would experience. I often wondered if it was worth it, but the support I received from friends and colleagues got me to the end.

Without the support of my mum and sister throughout the PhD, I may well have quit. Their unending support got me through some very tough days.

Most of all, I would like to thank my partner Paulina who has supported me through it all and motivated me to keep going no matter how hard things got. This project wouldn't have gotten where it did without you.

Abstract

In the last decade, CRISPR-associated proteins such as Cas9 and Cas12a have become powerful gene-editing tools for research, with great potential for therapeutic applications. However, barriers remain to their successful implementation, one of which is understanding the interaction of these systems with DNA repair processes. Previous research has identified HelQ, an ATP-dependent single-stranded DNA helicase implicated in replication-coupled repair, as a major factor in the efficient integration of template DNA through homology-directed pathways. HelQ has also been identified as a prognostic biomarker for several different cancers and a promising target in platinum-resistant ovarian tumours.

We hypothesised that HelQ removes Cas-proteins from the sites of double-strand breaks, but *in vitro* assays utilising both synthetic DNA substrates and supercoiled plasmid DNA were unable to confirm this. Alongside this, a system of human-derived cell-free extracts was reconstituted to study the impact of individual DNA repair proteins on integration efficiency. Further exploration into the role of HelQ in the resolution of DNA:RNA hybrids was also undertaken, ultimately finding conflicting evidence for the participation of the protein.

Previous research has demonstrated sensitisation to DNA-crosslinking agents in HelQ-deficient cells. To further characterise the importance of this, research was conducted using HelQ-depleted cell lines. This generated preliminary data laying the groundwork for a future functional genomic screen against HelQ.

Finally, using fragment-based drug discovery techniques, a small molecule

screen identified putative inhibitors against HelQ *in vitro*. Subsequent testing to validate and characterise hits identifying several candidates with micromolar IC50 values. Experiments to determine the mode-of-action suggest that many compounds were competitive with ATP, although several candidates were identified which may disrupt HelQ DNA binding or helicase activities.

Contents

List of Figures

List of Tables

Abbreviations

Abbreviation	Definition
5mC	5-methyl cytosine
8-oxo-G	7,8 dihydro-8-oxoguanine
A	Adenine
AA	amino-acid
ADP	Adenosine diphosphate
APS	Ammonium per sulphate
ATP	Adenosine triphosphate
ATM	Ataxia-telangiectasia mutated
ATR	Ataxia telangiectasia and Rad3-related
BER	Base excision repair
BIR	Break induced replication
BLM	Bloom syndrome protein
Bp	Base pair
BRCA	Breast cancer associated gene
BSA	Bovine serum albumin
C	Cytosine
C-HelQ	C-terminal domain of HelQ fragment
C2H2	Cys_2-His_2
Cas	CRISPR associated protein
CDK	Cyclin dependent kinases
CMG	Cdc45-MCM-GINS
CPD	Cyclobutane pyrimidine dimers
CRD	Central repeat domain
CRISPR	Clustered Regularly interspaced short palindromic repeats

Continued on next page

Abbreviation	Definition
CRISPRi	CRISPR-interference
crRNA	CRISPR RNA
D-loop	Deoxyribonucleic acid loop
DDR	DNA damage response
DNA	Deoxyribonucleic acid
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
dNTP	Deoxyribonucleoside triphosphate
dsDNA	Double stranded DNA
dsRNA	Double-stranded RNA
DSB	Double strand break
DSBR	Double strand break repair
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
EMSA	Electrophoretic mobility shift assay
ETC	Electron-transport chain
FA	Fanconi anaemia
FBDD	Fragment-based drug-discovery
FBS	Foetal bovine serum
FPC	Fork protection complex
G	Guanine
GINS	Go-ichi-ni-san protein
GOI	Gene-of-interest
HR	Homologous recombination
HTS	High-throughput screening
ICL	Inter-strand crosslink
IMAC	Immobilised metal affinity chromatography
InDel	Insertion/deletion

Abbreviation	Definition
IPTG	Isopropyl b-D-1thiogalactopyranoside
IR	Ionising radiation
KO	Knock out
LB	Luria-Bertani
MBP	Maltose-binding protein
MCM	Mini-chromosome maintenance
MegN	Meganuclease
MGE	mobile genetic element
MMC	Mitomycin C
MMR	Mismatch repair
MMEJ	Microhomology mediated end joining
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
N-HelQ	N-terminal domain of HelQ fragment
NUC	Nuclease
OD	Optical density
ORC	Origin recognition complex
PAGE	Polyacrylamide gel electrophoresis
PARP	poly(ADP-ribose) polymerase
PBS	phosphate buffered saline
PDB	Protein data bank
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PTMs	Post translational modifications
R-loop	Ribonucleic acid loop
REC	Recognition
RNP	Ribonucleoprotein complex
ROS	Reactive oxygen species
RPA	Replication Protein A

Abbreviation	Definition
RPMI	Roswell Park Memorial Institute (culture medium)
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RT	Room temperature
SDS	Sodium dodecyl sulphate
SDSA	Synthesis dependent strand annealing
SDW	Sterile distilled water
SEC	Size exclusion chromatography
SF1	Superfamily 1
SF2	Superfamily 2
sgRNA	Single guide RNA
siRNA	short interfering DNA
ssDNA	Single-stranded DNA
ssODN	single-stranded donor oligonucleotide
ssRNA	Single-stranded RNA
SSA	Single strand annealing
SSTR	Single-strand template repair
S-phase	Synthesis phase
T	Thymine
TALE	transcription activator-like effector
TALENs	TALE-nuclease
TBE	Tris-aminomethane borate ethylenedi-aminetetraacetic acid
TEMED	Tetramethylethylenediamine
TLS	Translesion synthesis
tracrRNA	trans-activating crRNA
TRC	Transcription-replication collision
TRIS	Trisaminomethane
UV	Ultraviolet

Abbreviation	Definition
WHD	Winged helix domain
WRN	Werner syndrome ATP-dependent helicase
WT	Wild-type
ZF	Zinc-finger
ZFN	Zinc-finger nuclease

1
Introduction

1.1 Genome editing

1.1.1 The development of genome editing

The manipulation of genes and genomes has been carried out by humans for millennia, beginning with the earliest attempts at selective breeding of crops and livestock. The turn of the 20th century saw an explosion in the study of gene function, including pioneering work on heredity in plants by Bateson as an extension of the prior work of Mendel[1,2]. Early discovery was limited to the observation of spontaneous mutations, later aided by the application of chemical or radioactive treatments and eventually recombinases and transposases to expedite this process, albeit randomly[3–6]. One of the key aims of gene-function research for several decades has been the precise, programmable, and targeted mutation of genomes which would open doors to both biological discovery and clinical advancement in the treatment of diseases, such as haemophilia and cystic fibrosis, through gene-therapy[7,8].

1.1.2 Early homologous gene targeting

The first experiments to achieve direct gene editing by insertion were carried out in yeast in the 1970s and 80s, using plasmids containing a gene-of-interest (GOI) flanked by regions of homology corresponding to the desired insertion site[9–11]. This research demonstrated the rescue of a *Saccharomyces cerevisiae LEU2* mutant by complementation with a wild-type (WT) copy of

the gene contained within a plasmid[11]. The results from this work found that exogenous *LEU2* was incorporated into the endogenous *LEU2* locus, demonstrating the replacement of DNA sequences by homologous recombination.

The technique was successfully transferred to mammalian cell and mouse models in the 1980s. Experiments demonstrated the repair of an integrated, mutated neomycin-resistance cassette via homologous repair pathways using an exogenous, correct cassette[12]. Despite this success, the technique was limited by a dependence on precise, but low-efficiency homologous repair pathways, need for powerful selection and a requirement for in depth characterisation[12,13]. A success rate of ≤1% limited applications to gene-therapy, leaving it predominantly as a tool for research, although one that has still seen use in recent years[14,15].

1.1.3 Transposases

First described in 1948 by McClintock, transposons are a family of mobile genetic elements (MGEs) found almost ubiquitously in both prokaryotes and eukaryotes[16]. In nature, transposons exist as 'selfish' elements which exist to replicate in the host genome by a variety of mechanisms. The class II DNA transposons are of particular interest in gene-editing, moving via an enzyme-mediated cut-and-paste mechanism which facilitates scarless excision and integration of DNA throughout the genome[17].

Transposon vectors have been widely used for gene delivery, beneficially being less sensitive to reduced integration efficiency with increasing cargo size than host-mediated homologous recombination[18–20]. However, due to their random integration and its associated mutagenesis, their use for precise insertions are severely limited. To overcome this, multiple attempts have been made to fuse transposases such as Sleeping Beauty, Casposase, and HsMar with powerful targetable nuclease technologies, such as CRISPR,

MegNs, ZFNs and TALENs, that are described in more detail below.[21–24].

1.1.4 Recombinases

Site-specific recombinase technology was introduced in the 1980s for the insertion of genetic material into genomes, being successfully applied in yeasts, mammalian cell models, and mice[25–27]. Recombinases are more specific than transposons, requiring the recognition of predetermined sequences prior to carrying out strand cleavage, exchange, and ligation of DNA[28]. The most commonly used site-specific recombinases are *Cre* from P1 bacteriophage and *FLP* from *S. cerevisiae*[29,30]. These enzymes share structurally common recognition sites which include two 13 bp palindromic sequences separated by an 8 bp asymmetric spacer sequence[28].

Recombinases have been successfully used as editing tools for decades, particularly in mouse models[29]. The approach has been considered a promising candidate for targeted gene therapy, but is limited by the rigidity of recognition sites. While pseudo-sites for some recombinases such as ΦC31, exist in mammalian cells, allowing limited targeting, the approach lacks programmability[31]. Attempts have been made to redesign specificity for recombinases, but currently *loxP* or *FRT* sites must first be introduced into the genome[32]. Doing so relies on technologies such as homologous gene targeting to insert recognition sites at known 'safe-harbour' locations.

1.1.5 Targeting nucleases

Since the early 1990s, there has been a sustained interest in exploiting natural DNA-repair processes within cells using targetable nuclease proteins, described in more detail in section1.1.5.1. These enzymes introduce double-strand breaks (DSBs) at specific points in a genome, triggering host DNA-repair mechanisms and thus promoting genetic change (Fig.1.1). The flexibility and programmability of these systems has seen multiple nuclease

technologies explored over the past three decades, gradually developing a gene-editing toolkit.
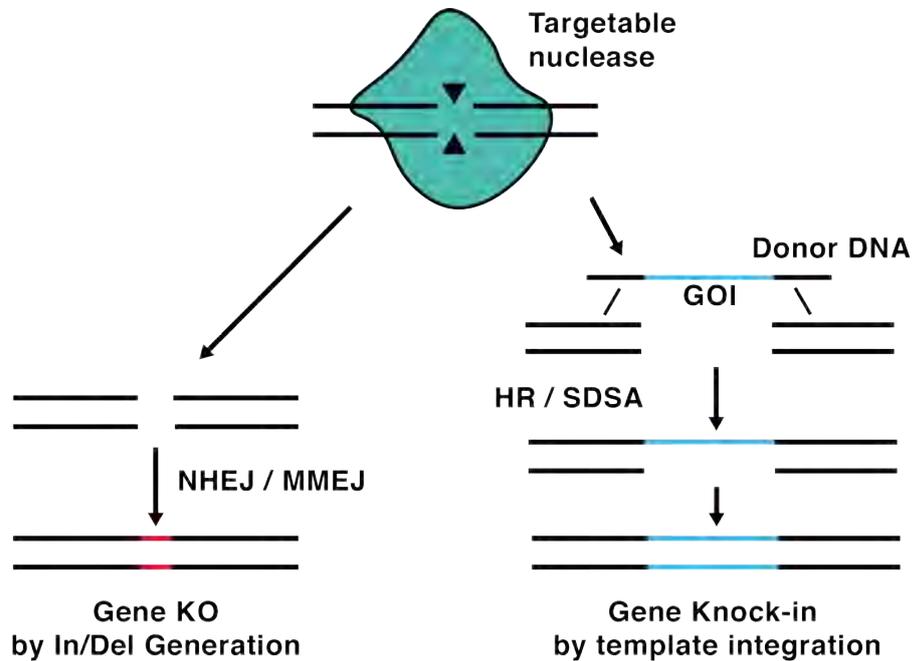
Figure 1.1: Schematic representation of pathways exploited by targetable nucleases. Targetable nucleases generate DSBs by cleaving DNA at predetermined sequences (black triangles). Cleavage at these sites stimulates host DNA repair mechanisms which can be exploited to achieve gene knockouts (red) through error-prone pathways or the insertion of genes of interest (GOI, blue). through homology-directed pathways. Figure produced by the author.

1.1.5.1 Meganucleases

Meganucleases (MegNs) are highly specific DNA cleaving enzymes found within all forms of microbial life, mitochondria, chloroplasts, and various eukaryotes[33]. First described in 1978, MegNs were eventually adapted as a tool for genome editing in the mid 1990's[34,35].

Naturally occurring MegNs exist within self-splicing elements such as the group I introns or inteins[36]. The mobility of these regions labels them as MGEs, similar to but biologically distinct from, transposons. MegNs can be characterised into six families based upon their sequences and structural motifs: LAGLIDADG, HNH, His-Cys box, GIY-YIG, PD-(D/E) XK, and

EDxHD, so named for conserved residues distinct to each[37,38]. A hallmark of all MegNs is a contrast between their small size, at typically fewer than 200 residues, and their long DNA target sites, which operate over a range of 12–45 bp[33,39].

LAGLIDADG is the largest and best characterised MegN family, and is most commonly employed in gene editing[40]. Within this family, the most well-known example is I-SceI from the mitochondria of *Saccharomyces cerevisiae*[41] which cleaves DNA containing an18 bp sequence, generating a 4-bp overhang (Fig. 1.2). The protein achieved early success in the editing of embryonic stem cells, as well as in cell lines such as HEK293[42,43].

The small size of MegNs makes them attractive candidates for gene editing in research and clinical applications, but their long recognition sequences limit flexibility. While hundreds of MegNs have been characterised, presenting a diverse array of target sequences, they lack the programmability of other targeting nucleases and so have seen diminished use over time[44]. Several attempts have been made at protein engineering to re-define the recognition site of meganucleases to overcome this inflexibility[45].

1.1.5.2 Zinc-finger nucleases

For many years, the best option for carrying out targeted genome editing was zinc-finger nucleases (ZFNs). First published in 1996, ZFNs are modular, programmable nucleases composed of a series of zinc-finger (ZF) motifs attached to a non-specific nuclease domain[46].

The ZF motif was discovered in 1985 as part of transcription factor IIIa in *Xenopus laevis* oocytes[47]. Subsequent research revealed that it is wide-spread among eukaryotic transcription factors, with the classical Cys_2-His_2 (C2H2) family being the largest class of DNA-binding proteins in metazoans and the most commonly used in the design of ZFNs[48,49].

Each ZF comprises a 30 amino-acid (aa) repeating region containing
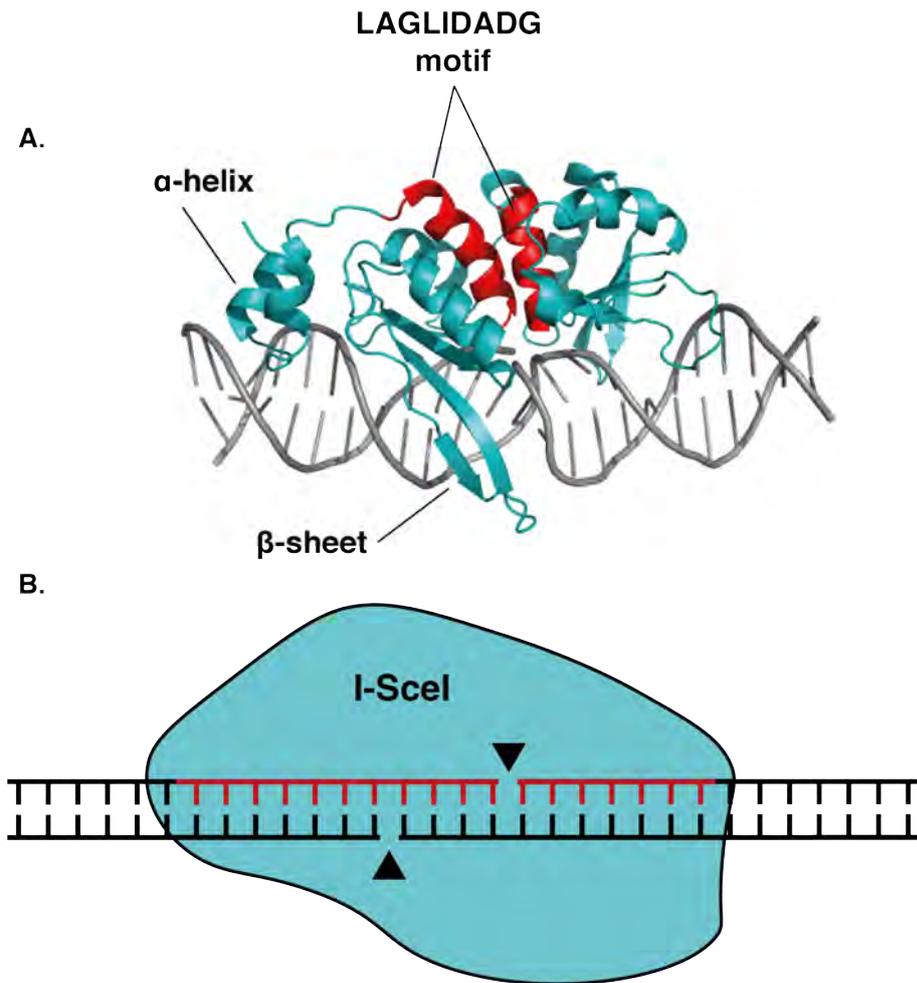
Figure 1.2: Interaction of meganuclease with DNA. (A) Crystal structure for the meganuclease I-SceI cleaving dsDNA, PDB accession number 5A0M. Highlighted in red are the two α-helices which contain the residues of the LAGLIDADG motif responsible for DNA cleavage. (B) Schematic demonstrating cleavage of dsDNA by a monomeric meganuclease. The DNA target for the enzyme is highlighted in red. The sites of DNA cleavage are indicated by black triangles. Figure produced by the author.

two anti-parallel β sheets and opposing α-helix (Fig. 1.3A)[50]. Within this structure, two cysteine and two histidine residues coordinate zinc for stability (Fig. 1.3B)[51]. Other residues, located on the α-helix at the tip of the ZF, determine specificity and participate in binding through interactions with the DNA major groove (Fig. 1.3C)[51,52]. Through these interactions ZFs are able to interact with 3–4 bp of DNA[52,53].

ZFNs are created by the fusion of 3–6 ZFs to a non-specific nuclease domain. The original, and most popular, choice for this was from the Type IIS

restriction enzyme FokI (Fig. 1.3D)[46,54]. Once bound to DNA, this domain creates a single-stranded nick. To create DSBs at the target site FokI nuclease must dimerise, with each monomer binding to opposite DNA strands[54]. The requirement for dimerization doubles the targeted sequence, bringing with it increased specificity (Fig. 1.3). Successful generation of DSBs requires ZFN monomers be designed to target regions of DNA 5–7 bp apart to allow the nuclease access to DNA[55,56].

ZFNs have progressed to clinical trials, including the treatment of patients with HIV by editing the CCR5 receptor gene associated with cell-to-cell spread of the virus[57]. Despite this success, the approach has several issues. Assembly of ZFNs is difficult, requiring experience and specialist knowledge to achieve, limiting wide-ranging application in research[58]. The system also suffers from incomplete programmability, as not all 64 possible triplet combinations have been developed, restricting the total targetable number of sites[59]. Alongside this, off-target effects may occur due to context-depended interactions between certain types of zinc finger[60,61].

1.1.5.3   TALENS

In 2009, it was discovered that transcription activator-like effectors (TALEs) secreted by the plant-pathogen *Xanthamonas* could be programmed to target individual DNA nucleotides[63,64]. TALEs were fused to a nuclease to form a modular, programmable system relying upon protein-DNA interactions for targeting (TALENs)[65,66].

TALEs comprise N- and C-terminal domains, and a TALE central repeat domain (CRD)[67]. The N-terminal domain contains a translocation signal and four repeats which are required for the initiation of DNA binding and confer a strong preference for thymine at position N_0; the 5′-most base of the bound sequence. The C-terminal region contains three conserved NLS sequences and a transcriptional activator domain.
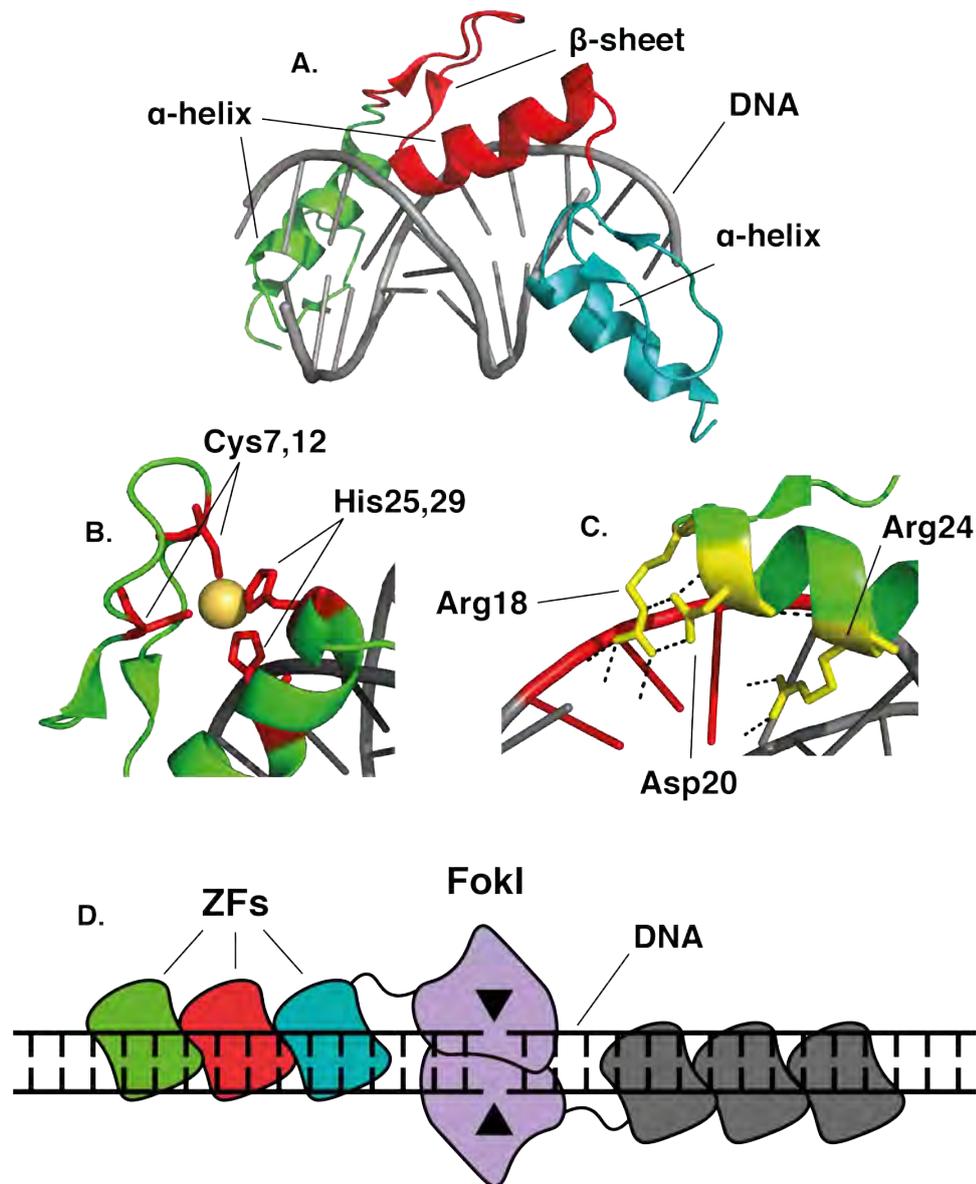
Figure 1.3: Interaction of a ZFN with DNA. (A) Strucutre of a tandem repeat of three ZF domains (green, red, blue) interacting with the major groove of DNA (grey). PDB accession number 1AAY; ZIF268 zinc finger-DNA complex[62]. (B) Struture of the Cys_2-His_2 motif in which a Zinc ion (yellow) is coordinated for stability. (C) Structure of a single ZF (green) targeting DNA. The nucleotide triplet targeted is shown in red while active residues are shown as sticks in yellow. Interactions with DNA shown as dashed lines. (D) A schematic demonstrating ZFN-mediated cleavage of dsDNA (black) by paired FokI domains (purple). Each colour (green, red, blue, grey) is representative of an individual zinc-finger domain composed of two β-sheets and a single α-helix. The sites of DNA cleavage are indicated by black triangles. Figure produced by the author.

The CRD is typically composed of 13—29 repeat units which determine TALE sequence-specificity[63,68,69]. Each repeat is composed of 34 aa, 32 of which are highly conserved (Fig. 1.4). The remaining two residues, at positions 12 and 13, are variable and responsible for determining the target nucleotide[70]. This leads to a simple set of recognition sequences with, for example, ND binding to cytosine, HN variably to alanine or guanine, NH to guanine, and NP non-specifically to all nucleotides[64].

TALENs are engineered by generating a CRD containing the desired repeats and fusing this at the C-terminal with the FokI nuclease domain (Fig. 1.4)[65]. Similar to ZFNs, this requires two different TALENs to target a specific site. The benefit of TALENs over ZFNs lies in the simplicity of targeting, only having to select modules corresponding to single nucleotides rather than triplets. TALENs have also been shown to have increased specificity, generating less off-target effects than ZFNs[71].

TALENs have progressed to clinical trials, enabling targeted insertion of CAR into the TRAC locus to enhance CAR-engineered T-cell therapy[72]. This indirect use of TALENs highlights a key disadvantage, their size. TALENs are typically three-times larger than ZFNs, increasing the difficulty of packaging and delivering them as a pair, reducing their attractiveness for clinical applications[71]. A further limitation is the highly-repetitive sequence of TALEs which are prone to spontaneous rearrangements, causing difficulty for some viral vectors in packaging them for delivery[73].

Despite success in applications to research and reaching clinical trials, the spotlight on TALENs was short-lived. Following several high-impact papers in 2012–13 a new technique would rapidly be adopted as the new standard for molecular biology applications and the pursuit of gene editing.
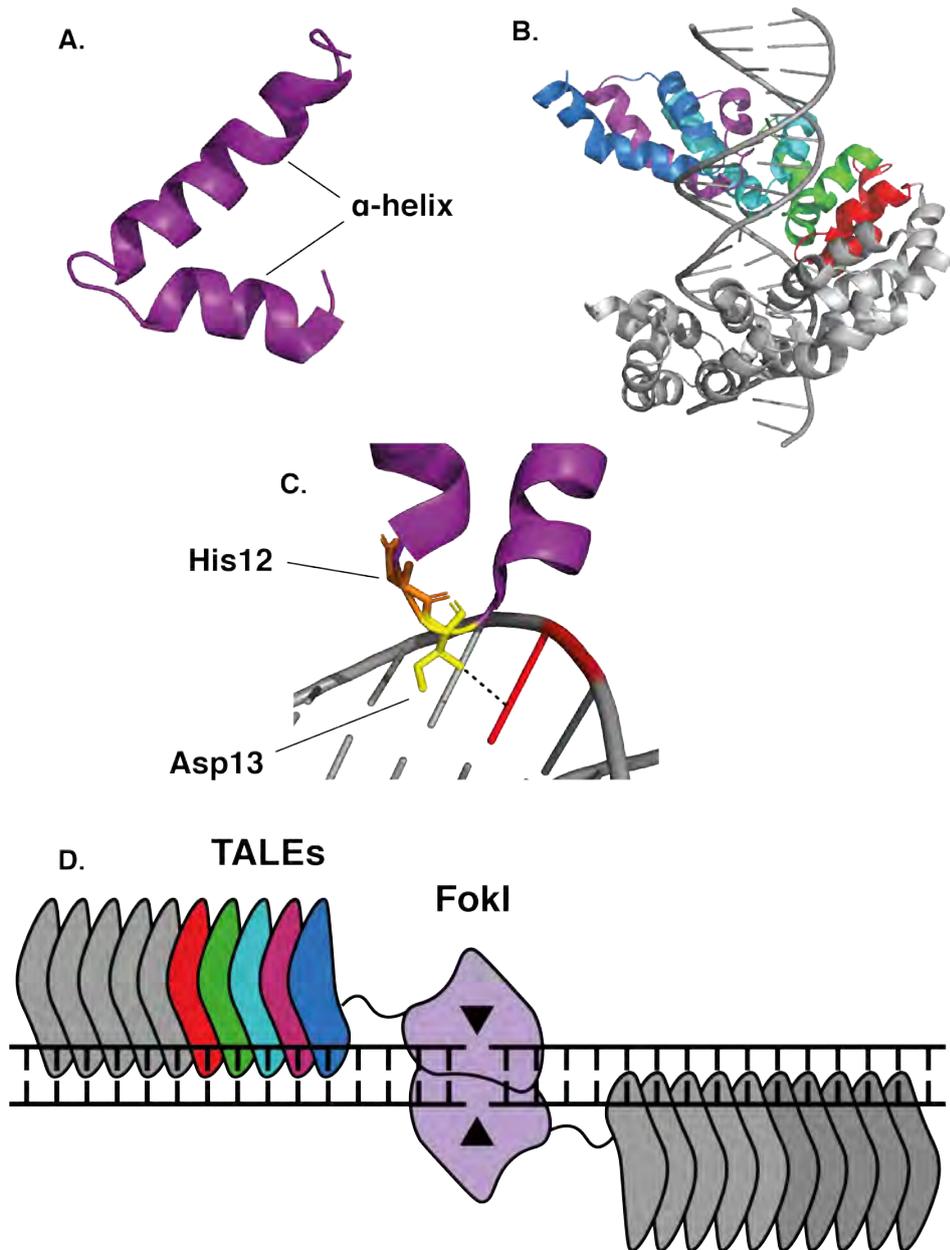
Figure 1.4: Interaction of a TALEN with DNA. (A) Strucutre of a single TALE repeat, displaying its dual-helix format. PDB accession number 3UGM; TAL Effector PthXo1 Bound to its DNA target[69]. (B) TALE repeat-array binding to individual bases in the major groove of DNA (grey). Each repeat (colours) targets a different base via a two-residue recognition sequence. (C) Structure of a single TALE repeat (purple) targeting DNA. Targeted nucleotide shown in red, active residues shown as sticks in yellow and orange. DNA interactions shown as dashed lines. (D) Schematic representation of TALEN-mediated dsDNA cleavage by paired FokI domains (purple). Several TALE repeat units are highlighted (colours) to correspond to panel B, with the remainder of the array also shown (grey). Each TALE domain (colours) is composed of the dual α-helices as shown in panel A. Sites of DNA cleavage are indicated by black triangles. Figure produced by the author.

1.2 CRISPR-Cas systems

1.2.1 The development of CRISPR research

Modules of clustered regularly interspaced short palindromic repeat DNA sequences (CRISPR) form the basis of an adaptive immune system in prokaryotes. Alongside CRISPR-associated (Cas) proteins, they provide sequence-specific protection from predatory elements such as phage, as well as limiting the movement of MGEs[74,75].

CRISPR sequences were first observed in prokaryotes in 1987 with the discovery of a repeat array downstream of the E. coli *iap* gene[76]. Similar arrays were subsequently discovered in diverse prokaryotes, including both bacteria and archaea, although their significance was not immediately realised[77–79]. Study of the arrays throughout the 2000s revealed key information about CRISPR including its function as an adaptive immune system, its targeting of DNA, the diversity of system types, and the functions of adaptation and effector proteins[80–84].

2012 saw the reconstitution of Cas9 cleavage reactions *in vitro* and the demonstration of programmable DSB generation using custom spacer sequences[85,86]. This was quickly developed as a system to carry out targeted mutation and insertion in mammalian cells and animal models, revolutionising approaches to gene-editing[87,88].

1.2.2 Prokaryotic immunity and the CRISPR locus

Upon invasion by MGEs, a prokaryotic cell can defend itself via the CRISPR immune response. This is coordinated by the CRISPR locus, a series of direct repeats connected by short, variable DNA sequences of foreign origin termed 'spacers', which define response specificity[82,83]. The repeat array is flanked by a set of type-specific *cas* genes 1.6 essential to the three stages of CRISPR response: adaptation, expression, and interference (Fig.1.5).
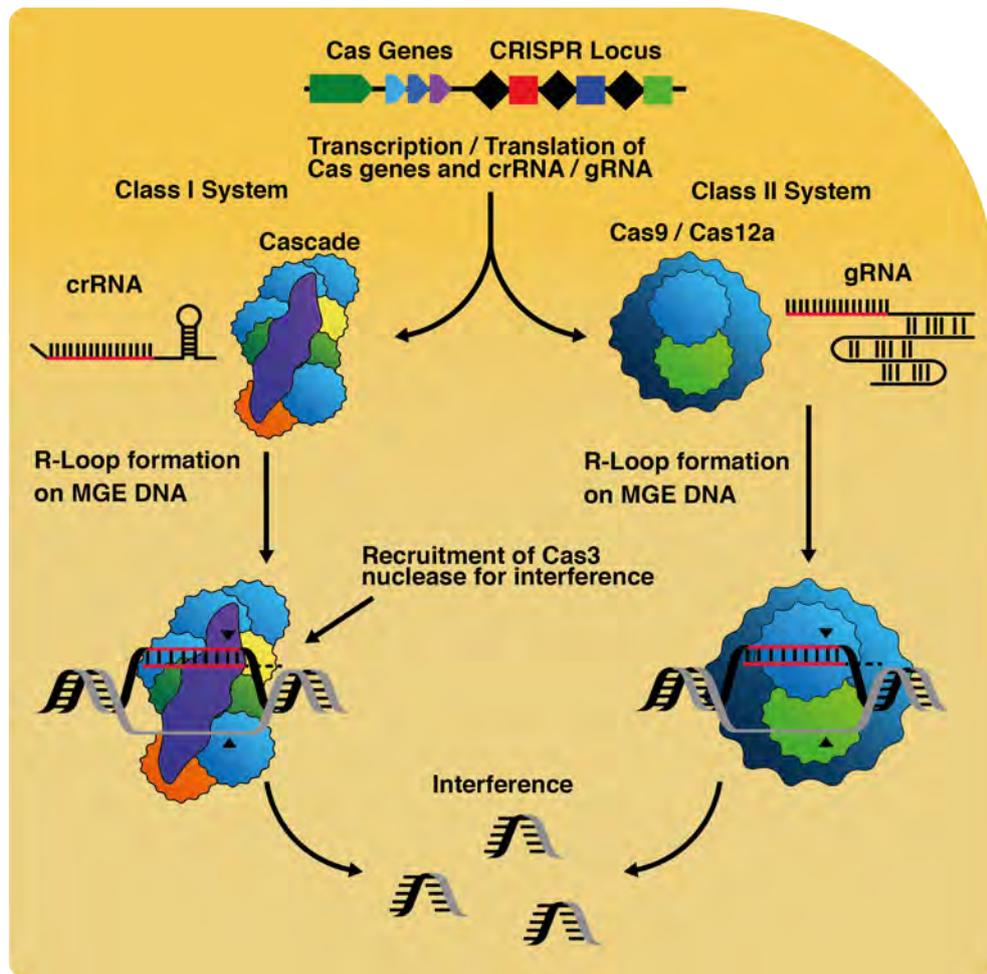
Figure 1.5: Naturally occurring CRISPR-mediated interference reactions. DNA from MGEs provides small protospacer fragments for acquisition into a CRISPR-locus during adaptation, generating immunity. Transcription and processing create crRNAs, which associate with and guide Interference complexes. Cas9 uses a two component system of crRNA and tracrRNA, to produce a mature sequence. Cas-protein:crRNA complexes catalyse the formation of R-loops on target DNA leading to nuclease activity. Cleavage is catalysed by Cas3 recruited to Cascade in Class I systems, or by Cas9 in Class II systems. Figure taken from Cubbon et al.[89] with permission.

1.2.2.1 Adaptation

Adaptation is the process by which new spacer sequence is obtained from an invading element. This is catalysed by a type-specific set of Cas-proteins, typically led by the Cas1-Cas2 complex[90]. In some systems spacer acquisition begins with the degradation of invading DNA through a combination of RecBCD helicase activity, assisted by one or more species-dependent nucleases and followed by Cas1-mediated strand-nicking[91,92]. In other systems,

the Cas1-2 complex is able to degrade target DNA independently, or in conjunction with a Cas-effector protein and so do not rely on host DNA-repair factors[93]. Degraded DNA is captured by the Cas1-2 complex and subsequently integrated into the host genome at the 5′ end of the CRISPR array. This process also duplicates the repeat DNA to maintain spacer separation. Spacer generation from host-DNA by RecBCD is limited by the presence of *Chi* sites which slow protein activity[94].

1.2.2.2   Expression

The expression stage sees the CRISPR repeat array transcribed as a single transcript known as pre-CRISPR RNA (crRNA)[74]. This is processed in a type-dependent manner by either Cas-ribonucleases or by host-factors such as RNase III, resulting in mature crRNA comprising a single spacer typically flanked by partial repeat sequence[95]. Once processed, the crRNA associates with a Cas-effector protein forming a ribonucleoprotein complex (RNP) which is able to target and initiate cleavage of DNA through interference reactions[96].

1.2.2.3   Interference

Interference utilises mature RNPs to interrogate and degrade foreign DNA. This is achieved by the RNA-guided formation of R-loops in which crRNA base-pairs with complementary MGE DNA[85,86]. Before an effector complex can cleave it must detect a short, typically 2–5 bp, sequence known as the proto-spacer-adjacent motif (PAM)[97,98]. Once the PAM has been detected, DNA degradation by the effector complex proceeds in a type-specific manner (Fig. 1.5). In some systems, R-loop formation is stimulated by a multi-protein complex which subsequently recruits the Cas3 nuclease-translocase to catalyse degradation. Other systems catalyse both R-loop formation and cleavage activity via a single effector protein[86].

1.2.3 The classification of CRISPR-Cas systems

CRISPR systems are widespread among prokaryotes and diverse in composition. Functionally, it is possible to divide the Cas genes into distinct modules (Fig. 1.6)[99]. Adaptation encompasses spacer acquisition enzymes, Cas1 and Cas2, the Cas4 nuclease present in several subtypes, and reverse transcriptase found in type III systems[100]. Expression comprises pre-crRNA processing. In most Class I systems, this is controlled by Cas6. Class II systems vary, with Type II utilising an endogenous, non-Cas RNase III, while many type V, and all type VI systems, possess effector proteins capable of independent processing[101,102]. The effector module encompasses target recognition and interference. Class I systems utilise multi-protein targeting complexes coupled with an effector nuclease, while Class II systems use single, combined targeting and interference proteins[103]. The signal transduction/ancillary module comprises CRISPR-associated genes, most of which are currently only hypothesised to play a role in immunity[99,104].

1.2.3.1 Class 1 CRISPR systems

Class I CRISPR–Cas systems are comprised of Types I, III and IV, and related sub-types. Class I is defined by its reliance on a multi-subunit effector complex, although the composition of these differs between types[99,103].

Type I loci achieve interference by combining a multi-protein targeting complex, Cascade, with a recruited secondary effector, Cas3. This enzyme is a superfamily-2 (SF-2) helicase able to unwind double-stranded DNA (dsDNA) and RNA–DNA hybrid duplexes, fused to a HD family endonuclease domain for cleavage of target DNA[105].

Type III systems possess Cas10 in place of Cas3[105,106]. Cas10 possesses a fused HD nuclease domain that is distinct from those of the Type I systems. Types III also differ from Type I in that Cas10, unlike Cas3, is associated with the multi-protein targeting complex making it similar in principle to Class II

systems[103].

Type IV systems are the most recent and evolutionarily divergent addition to Class I, typified by a lack of adaptation and interference-mediating proteins (Fig. 1.6)[107]. Currently, only Cas5 and Cas7 have been identified by sequence comparison to other Class I systems[108]. Structural comparisons among Class I systems suggests that Type IV may be mutated or derived from Type I and/or III[99,108].

1.2.3.2 Class 2 CRISPR systems

Class II CRISPR-Cas systems are comprised of Types II, V, VI and their related sub-types. Unlike Class I, this class uses a single effector protein to catalyse interference[99,103]. This function is often coupled with participation in adaptation or expression, including pre-crRNA processing[102].

Type II CRISPR systems catalyse interference using the multi-domain effector Cas9. These systems universally utilise Cas1 and Cas2 for adaptation[101]. A key feature of the Type II systems is a two-component RNA guide-sequence, which combines spacer sequences in the form of crRNA alongside a trans-activating crRNA (tracrRNA), required for maturation[84].

Type V systems utilise the effector Cas12, which is both structurally and mechanistically different to Cas9[101,109]. Most Type V systems also do not rely on endogenous RNase III for pre-crRNA processing, possessing the ability to do this independently. Some Type V systems have also been noted to lack certain adaptation proteins such as Cas1 and Cas2. The observed impact of this was an increase in non-CRISPR locus integrations[110].

Type VI systems are functionally unrelated to Types II and V, uniquely targeting ssRNA via the signature effector protein Cas13[111,112]. The enzyme possesses two distinct ribonuclease activities: one responsible for the processing of pre-crRNA to assist in mature effector complex formation, and a second for the degradation of target RNA[112].
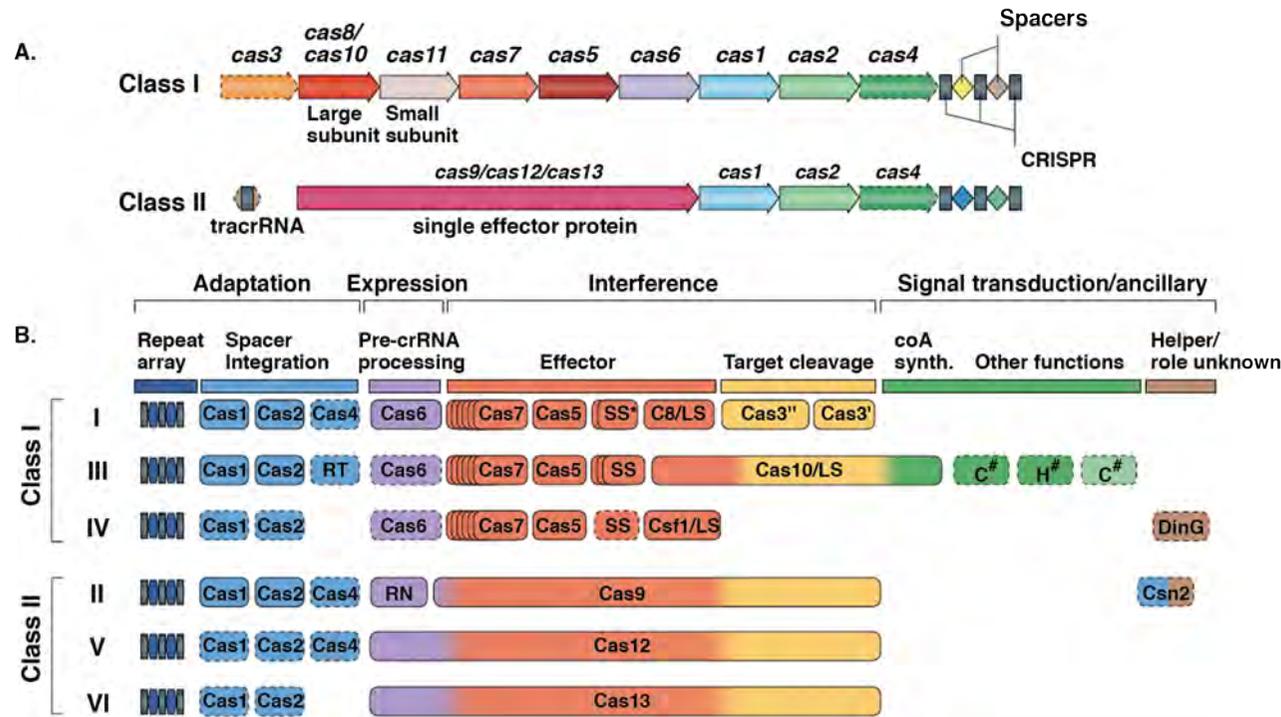
Figure 1.6: Classification and gene organisation of different CRISPR-Cas system. (A) Gene organisation of Class I and Class II CRISPR loci. Class I CRISPR–Cas systems are typified by multiple-component effector modules that form complexes capable of binding and processing a target. Class II systems rely on a single, multidomain effector protein for target binding and processing. (B) The functional modules of different types of CRISPR–Cas system. An asterisk (*) indicates a putative smaller subunit that may fuse to large subunits in several type I subtypes. Pound symbols (#) indicate unknown sensor, effector and ring nuclease protein families which may participate in signalling. Cas9, Cas10, Cas12 and Cas13 contribute to different stages of the CRISPR–Cas response and as such are multicoloured. The CRISPR-associated Rossmann fold (CARF) and higher eukaryotes and prokaryotes nucleotide-binding (HEPN) domain proteins are common sensors and effectors, respectively, in type III ancillary modules. C8, Cas8; RT, reverse transcriptase; RN, RNaseIII; LS, large subunit; SS, small subunit; tracrRNA, trans-activating crRNA.Figure adapted from Makarova et al.[99].

1.2.3.3 Cas9

Cas9 is a Type II CRISPR system and as such exists as a multidomain, multi-functional DNA endonuclease. Since the seminal work of Siksnys, Doudna, and Zhang in 2012/13 *Streptococcus pyogenes* Cas9 has been the flagship for CRISPR-Cas genome editing[85–87].

The enzyme is comprised of two different functional lobes: the recognition (REC) nuclease (NUC) lobes, linked by the bridge-helix a disordered linker in the Rec1 domain (Fig. 1.7A,B)[113]. The REC lobe is an α-helix-rich region unique to Cas9, comprising the split Rec1 and Rec2 domain, also referred to as Hel-I, -II, and -III[113,114]. The lobe participates in RNP complex formation with crRNA through contacts within the Rec1 domain, bridge helix, and the PI domain of the NUC lobe (Fig. 1.7B)[115]. Unless associated with crRNA as a RNP complex, the REC lobe is largely disordered, rendering Cas9 inactive. Upon RNP formation, the lobe undergoes conformational changes to allow its function[114,115].

The NUC lobe contains two nuclease domains responsible for DNA-target cleavage. The HNH nuclease domain is cleaves the guide-complementary strand of DNA. Conversely, the RuvC-like domain cleaves the non-complementary strand[113,114]. The PI domain of the NUC lobe is responsible for PAM recognition (NGG) and thus the initiation of target cleavage[113,114]. When not associated with crRNA, the NUC lobe is inactive. Upon RNP formation and target recognition, conformational changes occur which activate the RuvC-like and HNH nuclease domains[115]. These changes, mediated by linker regions between the two and assisted by the Rec2 domain of the REC lobe, shift the protein to an active state enabling dsDNA cleavage 3 bp upstream of the PAM (Fig. 1.7C,D)[115,116]. The dual-nuclease structure of Cas9 has led to the use of point mutations such H840A or D10A, to inactivate HNH and RuvC-like, respectively, creating either strand-specific nickases or catalytically dead protein (dCas9)[86].
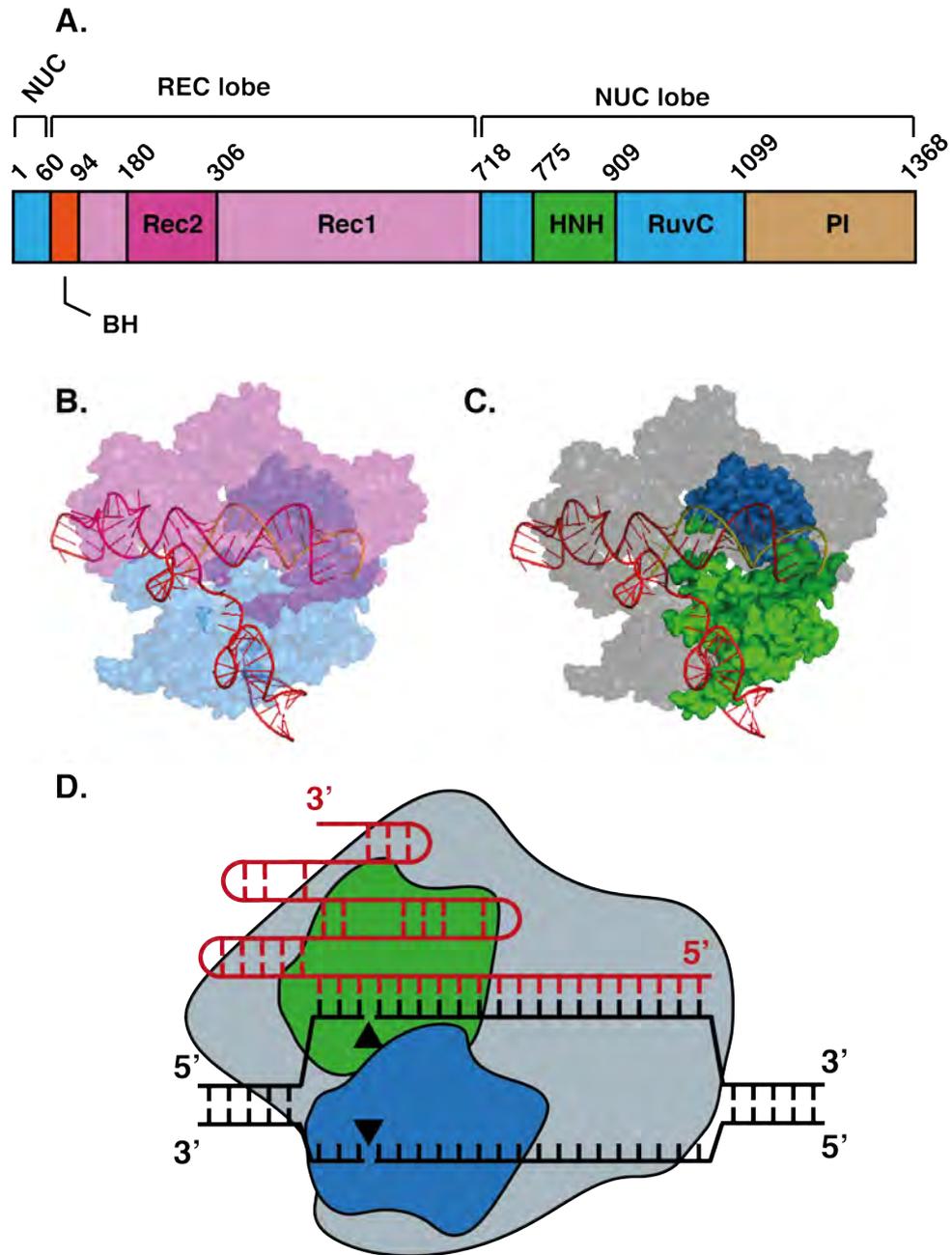
Figure 1.7: Schematic and crystal structure images of *S. pyogenes* Cas9. (A) Domain organisation of *S. pyogenes* Cas9. Split domains are colour coded: RuvC (blue), Rec1 (pink). BH, bridge helix. (B) Crystal structure of *S. pyogenes* Cas9 RNP bound to DNA (yellow), accession number 4OO8. Both the REC (pink) and NUC (blue) domains are highlighted. The sgRNA complexed with DNA is highlighted in red. (C) Crystal structure of *S. pyogenes* Cas9 RNP bound to DNA (yellow). The two nuclease domains of Cas9 are highlighted: RuvC (blue), HNH (green). (D) Schematic of Cas9 RNP forming an R-loop on dsDNA. Both nuclease domains are again highlighted: RuvC (blue), HNH (green). Cleavage sites are represented by black triangles. Figure produced by the author.

1.2.3.4 Cas12a

Cas12a is a Type V CRISPR system and so like the Type II Cas9 exists as a multidomain, multifunctional DNA endonuclease. Published in 2015 under the name CpfI, the characterisation of Cas12a expanded the gene-editing toolkit even further, offering a different mechanism of cleavage to Cas9[109].

Like Cas9, Cas12a adopts a bi-lobed structure comprising the REC and Nuc lobes[117]. The REC lobe contains Rec1 and Rec2 domains. Rec1 works in conjunction with the WED split domain and the PI domain of the NUC lobe to achieve two functions: the formation of RNPs, and the recognition of PAM sequences (TTTV)[117,118]. Upon crRNA association, the interaction is stabilised through contacts within the Rec2 domain and the WED and RuvC domains of the NUC lobe[119].

The NUC lobe of Cas12a differs from Cas9, possessing a single, RuvC-like nuclease domain[120]. It also contains PI, WED, and bridge-helix domains. The WED domain contains an RNase processing site which enables Cas12a to mature its own crRNA, unlike Cas9[102]. The split RuvC-like domain catalyses DNA cleavage via an interface between itself and the Nuc domain. Despite possessing only one nuclease domain, Cas12a still cleaves both strands of DNA. The mechanism of cleavage remains unknown, but the prevailing hypothesis is that conformational changes direct the target and non-target strands to the catalytic site through different pathways[121,122]. For instance, the Nuc domain has been suggested to guide the non-target strand and conformational changes to the RuvC-like domain that facilitate cleavage[121].

Cas12a also possesses indiscriminate ssDNA degradation activity upon activation at a target[123]. Structural studies have revealed this to be the result of a 'lid' region at the interface of the RuvC-like and Nuc domains, which controls access to the catalytic core[124]. Following target detection and cleavage the lid becomes disordered allowing access to ssDNA, which is then degraded[117,122].
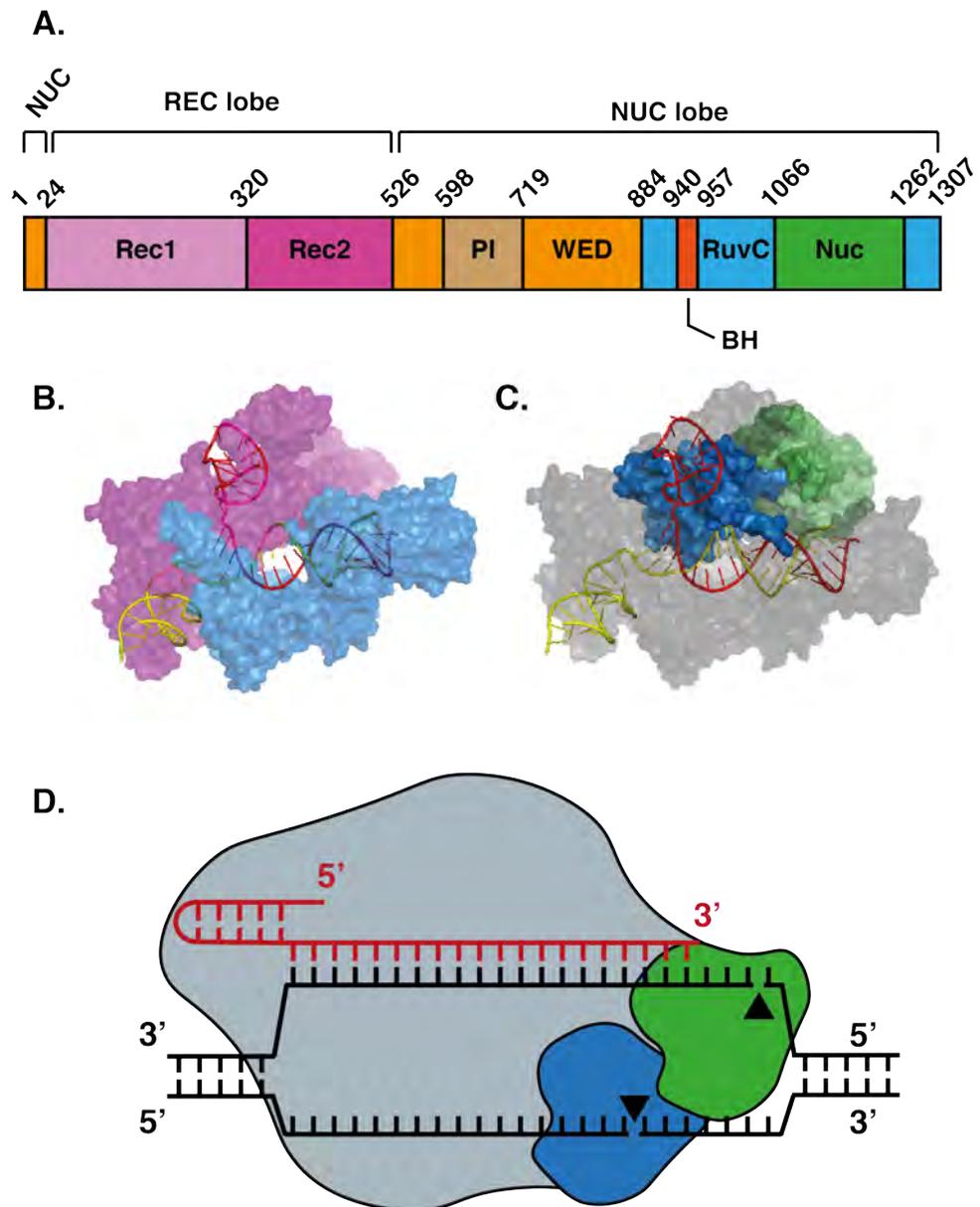
Figure 1.8: Schematic and crystal structure images of *Acidaminococcus* Cas12a. (A) Domain organisation of AsCas12a. Split domains are colour coded: RuvC (blue), WED (orange). BH, bridge helix. (B) Crystal structure of AsCas12a RNP bound to DNA (yellow), accession number 5B43. Both the REC (pink) and NUC (blue) domains are highlighted. The sgRNA complexed with DNA is highlighted in red. (C) Crystal structure of AsCas12a RNP bound to DNA (yellow). The RuvC nuclease domain is highlighted (blue) as is the Nuc domain (green). (D) Schematic of AsCas12a RNP forming an R-loop on dsDNA. The RuvC (blue) and Nuc (green) domain are again highlighted. Cleavage sites are represented by black triangles. Figure produced by the author.

1.2.4 The CRISPR revolution: Cas-enzymes as tools for gene-editing

The *in vitro* reconstitution of CRISPR interference reactions and subsequent application to targeted cleavage in bacteria and eukaryotes led to an explosion in the development of editing technologies. Cas-proteins such as Cas9 and Cas12a have been adapted and applied in a multitude of ways in the last decade to produce a vast tool-kit for the manipulation and exploration of genomes. As targeted nucleases, these methods rely on host DNA repair systems to achieve editing. The role of CRISPR-Cas enzymes as gene editing tools will be further explored in Section 1.7, but first it is important to understand the pathways that these tools can exploit.

1.3 DNA replication

The complete and accurate replication of DNA is essential for proliferation in all domains of life. DNA replication is a tightly regulated process carried out by a diverse collection of functionally conserved proteins known as the replisome. The process can be divided into three distinct stages: initiation, elongation and termination, all of which take place around a branched DNA structure known as a replication fork (Fig. 1.9).

1.3.1 Initiation

In eukaryotes, replication is initiated from multiple sites across the genome. These are detected by the origin recognition complex (ORC), a heterohexamer composed of the subunits ORC1–6[125]. In G1 phase of the cell cycle ORC, alongside Cdt1 and Cdc6, recruits and assists in the loading of the double hexameric minichromosome maintenance (MCM) complex[126]. This forms the pre-replicative complex, designating origins for the initiation of replication in the subsequent S phase[125].

During S phase, phosphorylation of MCM by the kinases Cdc7 and Cdk2
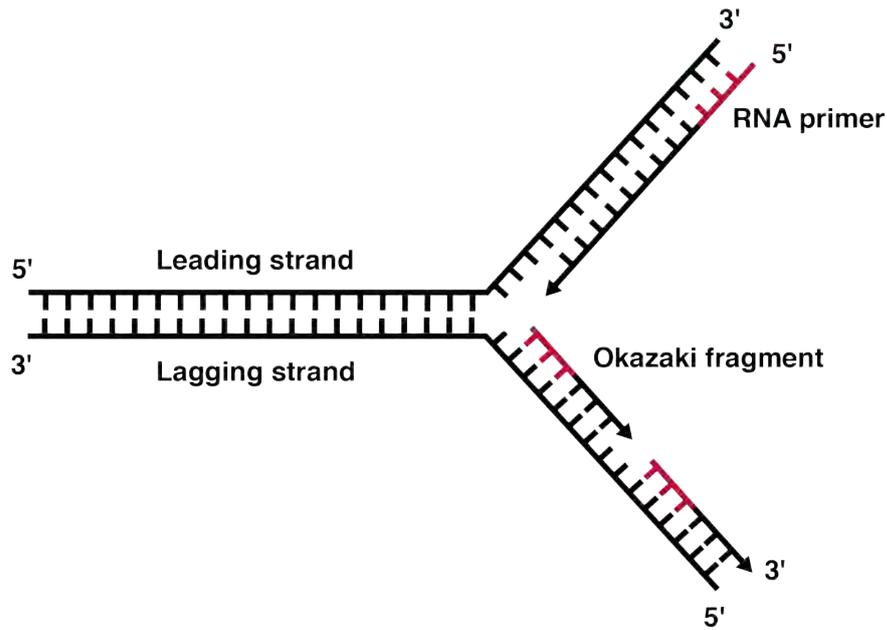
Figure 1.9: Schematic of a replication fork. The replicative complex translocates along DNA creating a forked structure, composed of a leading and lagging strand, which facilitates the bi-directional synthesis of DNA, utilising RNA primers (Red). The leading strand is synthesised continuously in a 5′–3′ direction. The lagging strand is synthetised discontinuously creating a series of Okazaki fragments. Figure produced by the author.

results in the recruitment of Cdc45 and the heterotetrameric *go-ichi-ni-san* (GINS), respectively[127,128]. The three proteins are assembled in a process mediated by the Sld3-–Sld7 complex to form the Cdc45-MCM-GINS (CMG) replicative helicase[129,130]. This assembly results in separation of the MCM double hexamer, forming bi-directional pre-initiation complexes[131]. CMG encloses the DNA leading strand and translocates along it with 3′ to 5′ polarity resulting in the melting of surrounding dsDNA, and steric exclusion of the lagging strand[131]. The resulting regions of ssDNA are then bound by replication protein A (RPA) and a suite of replicative enzymes are recruited to begin the process of elongation[132].

1.3.2 Elongation

Elongation begins with the CMG replicative helicase translocating along DNA. Downstream topological stress generated by unwinding is relieved by topoisomerases[133]. The replication fork is stabilised by the presence of the fork protection complex (FPC), composed of Claspin and other associated proteins[134]. This is crucial for preventing fork collapse during stalling or if a DNA lesion is encountered[135]. Within the replication fork, 7–12 nt RNA primers are synthesised by the DNA α-primase complex (Fig. 1.10, red) which then extends an additional 30 nt before being replaced by the more highly processive, strand-specific replicative polymerases[136,137].

DNA synthesis along the leading strand is a continuous process, requiring a single RNA primer. It is mediated by the proteins proliferating cell nuclear antigen (PCNA) and Polϵ. PCNA functions as a sliding clamp, facilitating the association of Polϵ with DNA (Fig. 1.10)[138]. A growing body of evidence suggests that this is not always the case and that there is a distinction between stable and recovering DNA replication. Prior research has established that in response to UV radiation, subsequent leading strand synthesis contained ssDNA gaps indicative of discontinuous replication[139,140]. More recent work has also shown that in response to stalling at sites of DNA damage, the core replisome is capable of recovering synthesis by re-priming downstream of the lesion[141]. In eukaryotes this can be initiated by the primase-polymerase, PrimPol, which re-primes the DNA downstream of replication-stalling lesions, including abasic sites and crosslinks[142,143]. Notably, the protein is capable of synthesising and elongating from a DNA, rather than RNA, primer which is thought to facillitate a more efficient restart of replication[142].

Synthesis along the lagging strand is discontinuous, requiring multiple RNA primers and leading to the formation of approximately 200 bp structures called Okazaki fragments (Fig. 1.10)[144,145]. DNA is synthesised beyond the initial 30 nt by a complex of PCNA and Polδ until the next RNA primer

is reached[146]. Here the replicative complex is removed and reassociates with the DNA on the next Okazaki fragment. The RNA primer sequence is processed by a combination of RNase H, FEN1 and DNA ligase I to fill in gaps between fragments[147].
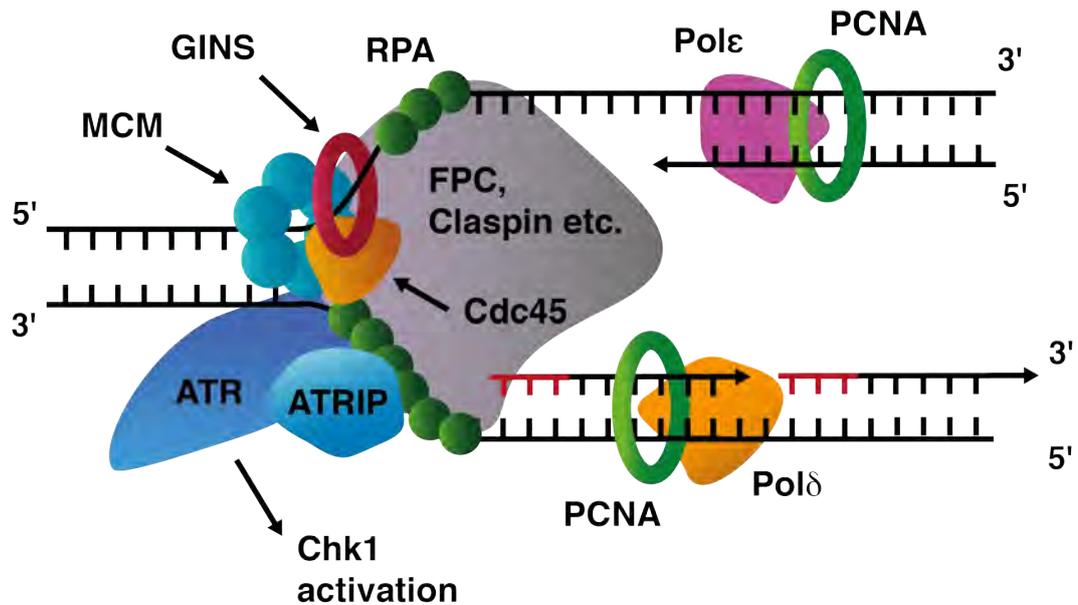
Figure 1.10: Schematic of the eukaryotic replisome. The replicative complex is composed of a hexameric helicase (MCM) and the accessory factors Cdc45 and GINS. Translocation of this complex forms a replication fork to which other replicative proteins are recruited. This includes the fork protection complex (FPC), clamp-loader protein PCNA and associated replicative polymerases for the leading and lagging strand, Polϵ and Polδ, respectively. ssDNA at the replication fork is bound by RPA, which also facillitates the ATP checkpoint kinase response via a docking interaction with the ATR-ATRIP complex. Figure produced by the author.

1.3.3 Termination

The termination of DNA replication remains poorly understood, lacking a true mechanism of action. Current research suggests that termination typically occurs with the convergence of two forks. During this, opposing leading-strand replication complexes pass one another, proceeding until they encounter a downstream Okazaki fragment and are disassembled from the

DNA[148,149]. Okazaki fragments are then removed through the recruitment of DNA Polδ and the flap-endonuclease FEN1 before the synthesised strands are ligated together by DNA ligase I[150]. During fork convergence, the parental DNA between forks becomes too short to supercoil and topological stress is instead relieved by the formation of pre-catenanes in which the replicated sisters cross over each other[151,152]. For the full termination of replication, the pre-catenates are resolved by topoisomerase II to separate the replicated DNA[153,154].

Site-specific mechanisms for termination have also been described for several classes of DNA. At highly transcribed regions, such as ribosomal gene-clusters, structural barriers stall the replication machinery long enough for fork convergence to occur outside of the region[155]. In prokaryotes this occurs when Tus protein binds at *Ter* sequences, blocking the progress of replicative helicase DNaB[156–158]. In the eukaryote budding yeast, this occurs at replication fork barrier (RFB) sites, which are bound by Fob1 to form a polar barrier wherein replisome advance is stalled from one side, allowing the region to be faithfully replicated without fork collision[155,159]. This also appears to prevent genomic instability which can be associated with fork collision between the replication and transcription machinery[155]. At the telomeres, termination occurs when the replication machinery reaches the end of the DNA and falls off of the chromosome leaving overhangs which lead to the recruitment of the specialised telomerase enzymes for end processing[160].

1.3.4 Replication stress

Faithful replication of the human genome during S-phase is essential to cell health, being carried out through the tightly timed coordination of approximately 1×10^5 replication forks at 3–5×10^4 origins, across an 8 hour period[161]. DNA however, is highly prone to damage with estimates placing

the total events per cell per day at 10^4–10^5 [162,163]. This damage is unavoidable, caused by endogenous and exogenous factors which lead to lesions such as strand breaks and crosslinking. During replication this can also be further exacerbated by collision with other DNA-bound proteins, such as the transcription machinery[164].

Left unchecked, this damage leads to replication stress due to fork stalling. If this remains unresolved then replisome stability may be compromised, leading to complex disassembly and replication fork collapse[165]. The occurrence of fork collapse risks incomplete replication, which is a great source of genome instability, resulting in deletions and large chromosomal rearrangements that can contribute to the development of cancers[164]. To combat fork stalling and collapse, the replisome is stabilised by the recruitment of the FPC (Fig. 1.10)[166]. This assists in coordinating the activation of lesion-specific, checkpoint-activated pathways such as double-strand break repair, which have evolved to specifically address different forms of DNA lesion[167]. This is coupled with programmed death responses in the event that repair mechanisms fail to successfully address damage[164,167].

1.4   DNA repair

1.4.1   Repair of DNA double-strand breaks

DSBs caused accidentally, by endogenous/exogenous sources, or intentionally, due to cellular events, or the use of targetable nucleases such as Cas9, can result in chromosomal breakage and rearrangements, leading to cell death[168]. To combat this, multiple repair pathways have evolved to process these lesions based upon the structure of DNA at the site of breakages (Fig. 1.11), as well as the current stage of the cell cycle.
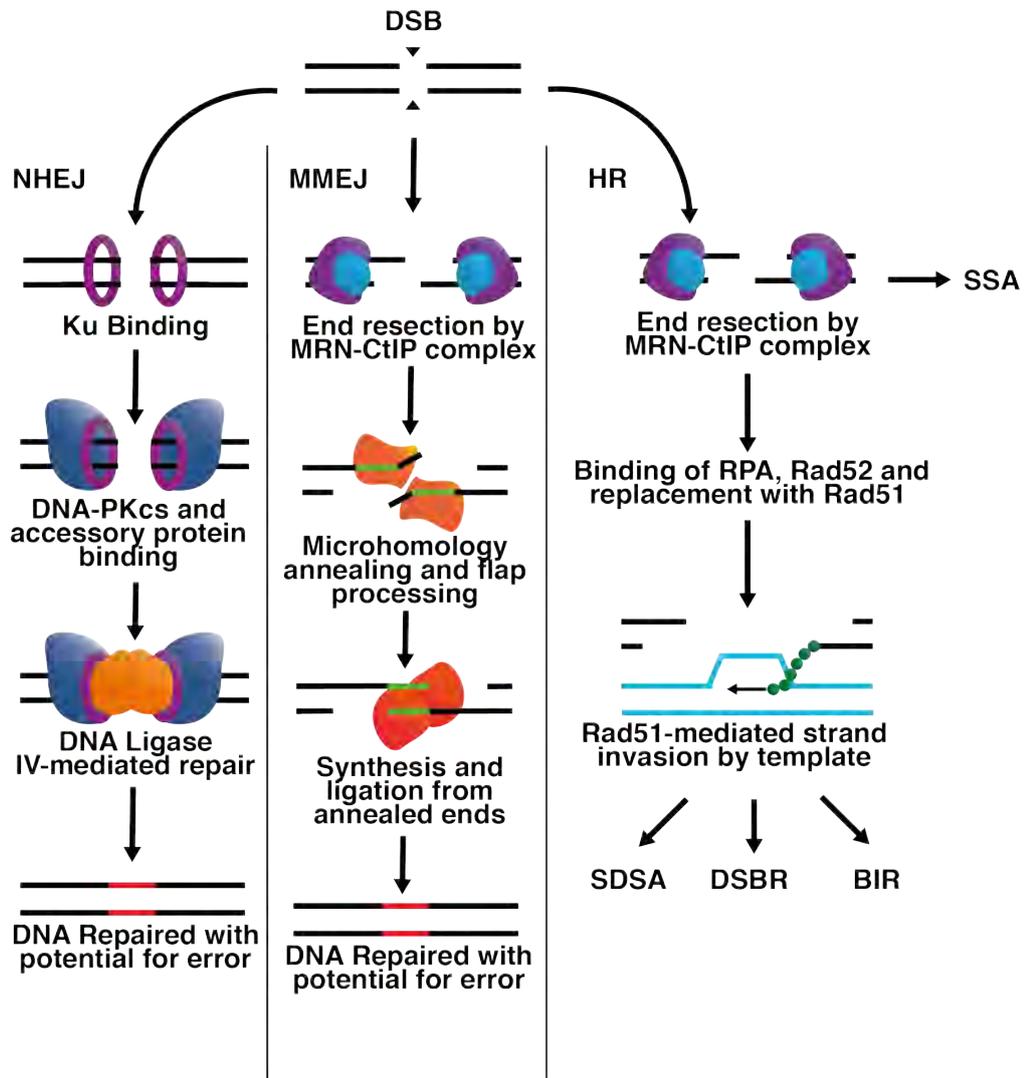
Figure 1.11: Summary of pathways for the repair of DSBs. The pathways to DSB repair can be broadly defined by three mechanisms. During non-homologous end joining (NHEJ), Ku heterodimers stabilise free DNA ends and recruit DNA-PKcs among other accessory factors to bridge, process and ligate the break site. During microhomology-mediated end joining (MMEJ), DNA ends anneal at short regions of homology. This is followed by flap processing and gap filling by several exonucleases and polymerase before the nicked ends are ligated together. Homologous recombination (HR) comprises several pathways during which end resection reveals homologous template DNA which is sequestered by recombinases to form filaments. These invade homologous DNA to be used as templates for repair. Figure produced by the author.

1.4.1.1 Repair pathway choice throughout the cell cycle

The most direct determinant of DSB repair pathway is the type of end-stabilisation complex that forms at the site of DNA damage, which is expanded upon in sections 1.4.1.2–1.4.3. Equally important however, is the phase of the cell cycle at which damage occurs, as this can limit the available pathways for repair by controlling which end-resection proteins are available at the site of a break[169]. This is a complex process coordinated by several arrays of proteins, carrying out multiple functions, many of which are yet to be understood. As such, only a few key examples will be highlighted in this review.

During the first growth phase of the cell cycle (G1), the predominant repair pathway for DSBs is non-homologous end joining (NHEJ, see section 1.4.1.2), initiated by the Ku complex[170]. At this time, repair through homologous recombination (HR) is carefully repressed due to a lack of sister chromatids to be used as repair templates for the DSB[171]. This has been shown to be, in part, moderated through the activity of KEAP1, a E3 ubiquitin-ligase, which prevents interaction between BRCA1 and PALB2 and thus the recruitment of BRCA2 and Rad51 as elaborated on in section 1.4.1.4[172].

Outside of G1, NHEJ is down-regulated relative to HR. This has been partly associated with the E3 ubiquitin ligase, RNF138, which simultaneously inhibits the Ku complex, while promoting the HR-associated resection factor, CtIP[173,174]. CtIP is further regulated through phosphorylation by CDK1[174]. The helicase RecQL4 has been observed to impact pathway fate interchangeably; interacting with the NHEJ-associated Ku complex during G1, but switching to promote end resection by the HR-associated MRN complex during S and G2 phases[175]. This switch is the result of phospohrylation by CDK1/2 which changes the interacting partner of RecQL4[175].

During the mitosis (M) phase of the cell cycle, all DNA repair pathways

are down-regulated to allow cell division to take place, leaving only truncated repair mechanisms in place to process DSBs[176]. This is regulated by the checkpoint proteins CDK1 and PikI, which repress the activity of key NHEJ and HR proteins[176].

Regulation of repair pathways throughout the cell cycle is essential to fitness, preventing deleterious events which may lead to cancer. Utilisation of HR in response to damage during G1, could lead to recombination between parental chromatids resuting in a loss of allele heterozygosity and, by extension, genetic redundancy in cells[177]. As a result, the cell would become reliant on a single allele, which could similarly be lost during future damage events. Wild-type allele loss has been strongly associated with tumour development in individuals with inherited predispositions to cancer and those carrying germline mutations in genes such as BRCA1 in breast and ovarian cancers[178,179].

1.4.1.2 Non-homologous end-joining (NHEJ)

Non-homologous end joining (NHEJ) is the repair of DNA in the absence of template to inform the reattachment of free ends. The pathway is most active in G1 phase of the cell cycle and is inhibited during mitosis to prevent telomere fusions which would disrupt the chromosome[180,181]. NHEJ begins with the binding of the hetero-dimeric Ku complex, composed of Ku70 and Ku80, to exposed DNA ends (Fig. 1.11). This serves as a scaffold for the recruitment of multiple lesion-specific accessory proteins[170]. The most common of these is the DNA-PKcs complex which aids in the stabilisation of broken ends and assists in bridging the synaptic complex[182]. Not all lesions will produce blunt, NHEJ-compatible ends. The Artemis exonuclease complex is recruited for the end processing for the removal of overhanging nucleotides to form true blunt-ends, or the removal of non-ligatable groups such as 5′ hydroxyls or 3′ phosphates. Incompatible ends can also be processed through

gap-filling by DNA Pol μ and λ[180]. Repair is completed by DNA Ligase IV which seals the remaining nicks (Fig. 1.11). NHEJ is associated with a common incidence of small InDel mutations, usually several base-pairs long, which can induce frameshift in the coding regions of proteins, leading to their truncation and inactivity following translation[170].

1.4.1.3 Microhomology-mediated end-joining (MMEJ)

MMEJ is an error prone DNA repair process that repairs DSBs by annealing stretches of 2-–20-bp of homologous DNA flanking a lesion. It has been reported as a destabiliser of the genome as it can result in small InDels, telomeric fusions, and hypermutations of up to 10 kb[183].

MMEJ is tethered by the association of poly(ADP-ribose) polymerase 1 (PARP1) to the site of DSBs[184]. This recruits the Mre11-Rad50-Nbs1 (MRN)– C-terminal Binding Protein Interacting Protein (CtIP) complex, which catalyses end resection through 3′–5′ exonuclease activity (Fig. 1.11)[185]. More extensive end resection can be catalysed by a complex of Bloom syndrome helicase (BLM) and exonuclease 1 (EXO1) through 5′–3′ exonuclease activity, but this has been shown to be dispensable and can instead suppress MMEJ if there is microhomology within 2 kb of break-site[185].

Following end-resection, the exposed regions of microhomology anneal while remaining ssDNA is sequestered by RPA[186]. The annealed intermediate is further processed by the XPF/ERCC1 structure-specific nuclease complex which removes non-homologous tails to leave blunt, 3′-hydroxyls compatible with polymerase-mediated extension (Fig. 1.11)[187]. XPF/ERCC1 is also implicated in the recruitment of DNA helicase-polymerase θ (PolQ), which fills gaps in the duplex before they are then ligated redundantly by DNA ligases LIG1 and LIG3[184,188,189].

1.4.1.4 Homologous recombination (HR)

HR comprises multiple pathways conserved across bacteria, archaea and eukaryotes for the repair of DSBs, single-strand lesions and interstrand crosslinks (Fig. 1.11)[176]. HR pathways rely on the availability of homologous DNA template that can base-pair with damage sites. For this reason, it is considered an error-free method of DNA repair. Alongside this, HR also supports other processes such as the reactivation of stalled replication forks at DNA[139].

Unrestrained HR has been shown to cause rearrangements in DNA[190]. In response, regulatory mechanisms have evolved to constrain it. This means that HR in eukaryotes occurs only during the S and G2 phases of the cell cycle[191]. Failure to resolve DNA damage and/or replication stress through HR mechanisms can lead to genomic instability, a driving factor in the development of cancers[176]. Mechanistically the process can be broken down into three phases: pre-synapsis, synapsis, and post-synapsis.

During pre-synapsis in humans, free DNA ends are resected by the MRN-CtIP complex with a $5'$–$3'$ polarity to produce overhanging $3'$ ends (fig. 1.12)[192]. Where required, the nuclease Exo1 can carry out more extensive end-resectioning. At this stage, interactions with other effector proteins such as BRCA1 commit repair to the HR pathways. ssDNA exposed during resection is bound by RPA stabilising and protecting it from exonuclease activity, and recruiting downstream effectors[192]. This includes the eukaryotic recombinase Rad51, which displaces RPA with assistance from intermediaries such as BRCA2, Rad52 and Rad54. Rad51 loads onto the exposed ssDNA forming a right-handed helix, composed of 6.4 Rad51 monomers and eighteen nucleotides of DNA per turn, known as the nucleoprotein filament (NPF) (Fig. 1.12)[193]. This composition causes the associated DNA to stretch to 150%, which assists in downstream homology-searching by the filament[193,194].

The second phase of HR, synapsis, begins with strand exchange between

homologous DNA by a Rad51-NPF. This facilitates homology-searching in which a filament migrates along template dsDNA to identify complementary sequence[195]. The ssDNA of an invading NPF then base-pairs with complimentary sequence through Rad54-mediated displacement of Rad51, resulting in the formation of a triple-stranded DNA intermediate known as a displacement-loop (D-loop). This strand is able to prime extension, known as first-end synthesis, via an exposed 3′ hydroxyl group for a suite of polymerases, with DNA Pol ζ, λ, ν, κ, and η all being implicated in various studies[196].

The final stage of HR is post-synapsis in which several context-dependent pathway choices are confirmed. Three major pathways can begin resolving repair from this point: single-strand template repair (section 1.4.1.5), break-induced replication (section 1.4.1.6) and double-strand break repair (DSBR). Post-synapsis via the DSBR pathway begins with second-end capture in which the newly extended strand and the second free-end of the DSB are annealed through Rad52-dependent interactions. This is coupled with second-end synthesis in which the gap remaining on the second strand of the DSB is filled by DNA pol δ[196]. The result is the formation of a double-Holliday junction(dHJ), a branched structure comprising two linked DNA duplexes. To resolve the dHJ and complete repair, a set of structure-specific endonucleases are employed. Multiple endonuclease complexes have been implicated as HJ-resolvases in eukaryotes, including: Mus81-Mms4, Slx1-Slx4, XPF-ERCC1, GEN1, and Sgs1/MutLγ/Exo1[197].

The manner of cleavage by HJ-resolving endonucleases determines whether or not repair products will be crossover or non-crossover. Symmetrical cleavage of the dHJ (Fig. 1.12, black, clear) results in non-crossover products with no exchange of genetic material[198]. Asymmetric cleavage meanwhile (Fig. 1.12, black, red) results in crossover products and the exchange of genetic material[198].

A final, alternative approach to the resolution of dHJs is through dissolution. This is facilitated in humans by the helicase BLM, which migrates the two junctions towards each other to form a hemi-catenane, which is then processed by the topoisomerase TOPIIIα[139]. Resolution of the hemi-catenane in this way releases the two duplexes resulting in completed DNA repair with non-crossover products.

1.4.1.5 Synthesis-dependent strand annealing (SDSA)

SDSA is a sub-pathway of HR-mediated repair which branches from the main pathway during post-synapsis (Fig. 1.11). The result of SDSA repair is always a non-crossover product, greatly reducing the likelihood of genomic rearrangements[176]. As a result of this the pathway in eukaryotes is predominant in mitotic cells[199].

Mechanistically SDSA follows the same primary route as DSBR. It diverges following extension of the invading strand, but prior to second-end capture. At this point helicases, including BLM and RTEL1, reverse D-loop annealing by disrupting Rad51 on the NPF, causing dissociation of the invading strand (Fig. 1.12)[200]. The now extended strand then anneals to the remaining resected free DNA end where repair is completed through gap filling by DNA Pol δ (Fig. 1.12)[200].

1.4.1.6 Break-induced replication (BIR)

BIR is a specialised sub-pathway of HR-mediated repair, most often taking place at the site of one-ended DSBs such as telomere ends and stalled replication forks[201]. As in DSBR and SDSA, BIR begins with end resection and the formation of Rad51-mediated NPFs (Fig. 1.11). Following strand invasion, synthesis capable of extending ≥ 100 kb in yeast, or ≤ 4 kb in mammalian cells, is initiated by Pif1-mediated recruitment of DNA Pol δ to D-loops (Fig. 1.12)[202]. When synthesis is completed, strand exchange and the resolution
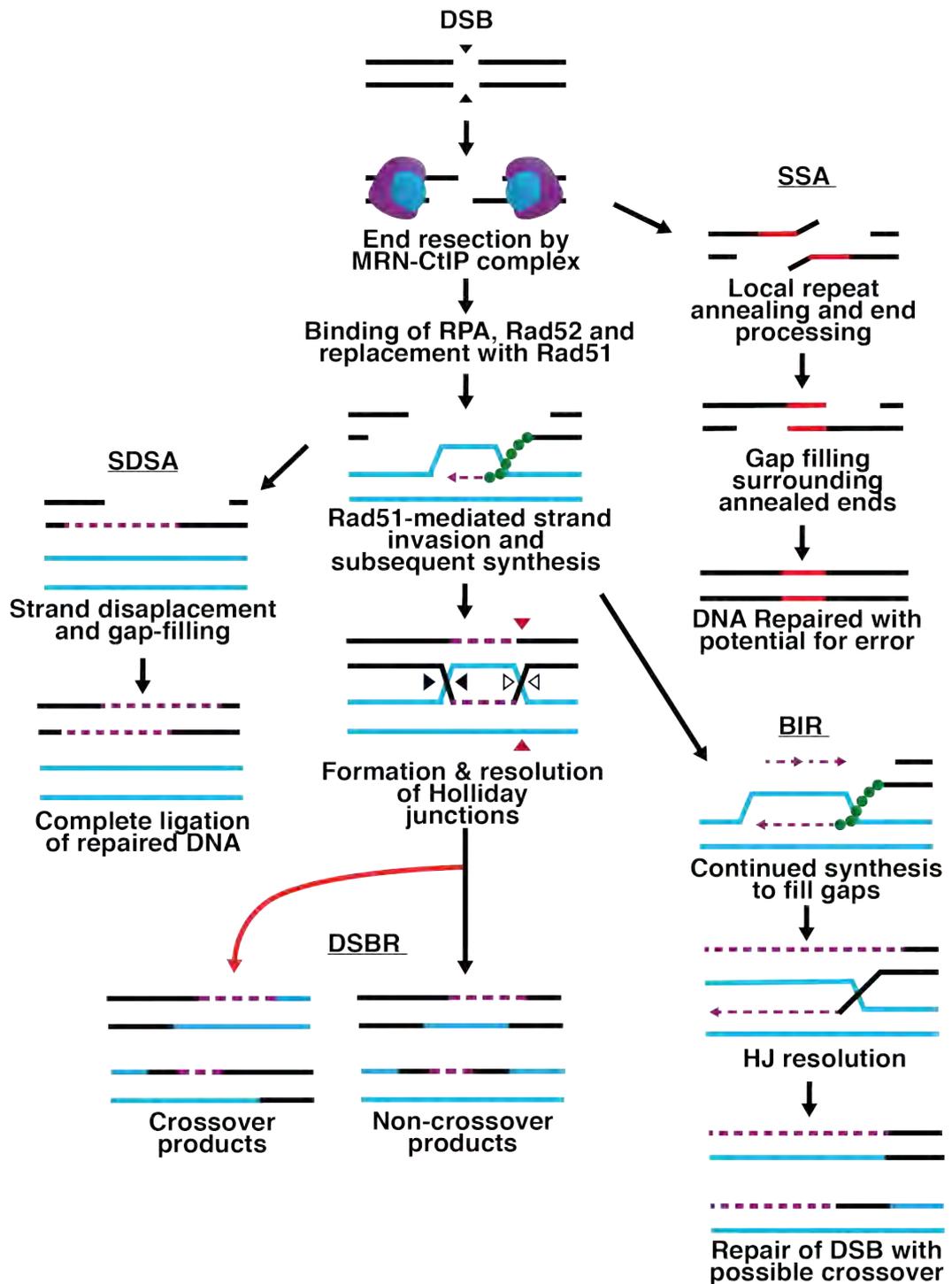
Figure 1.12: Outcomes of repair by homologous recombination. Repair of double-strand breaks (DSB) by homologous recombination can take place through a series of context-specific pathways. In all cases, the site of the break is processed by having the free ends resected to facilitate the binding of HR-associated protein. Single-strand annealing (SSA), Non-homologous end joining (NHEJ) Break-induced replication (BIR), synthesis-dependent strand annealing (SDSA), Double-strand break repair (DSBR). Figure produced by the author.

of Holliday junction intermediates is coordinated by HR-associated enzymes as discussed in section 1.4.1.4.

Beyond this basic mechanism, BIR is poorly understood in eukaryotes. It is however mechanistically distinct from typical replication, utilising a migrating bubble in place of a replication fork and synthesising the leading and lagging strands asynchronously[203,204]. This leads to a large accumulation of ssDNA which is is stabilised by RPA[205]. BIR is a highly error-prone pathway, demonstrated to have a mutation rate one thousand times that of normal DNA replication[201].

1.4.1.7 Single-strand annealing (SSA)

SSA is another specialised, HR-adjacent repair pathway, typically taking place when there are long regions of homology in proximity to the site of the DSB. This pathway diverges after the completion of end resection forgoing strand invasion in favour of Rad51-independent repair (Fig. 1.11)[206]. This involves the annealing of long repeat stretches of ssDNA, mediated by Rad52. Following this, any remaining 3′ tails are processed by the ERCC1/XPF complex[207]. Gaps in the annealed sequences are then filled by a currently unclear suite of polymerases and ligases to complete repair (Fig. 1.12). SSA often results in deletions at repeat regions, making it a relatively mutagenic repair pathway[206].

SSA can be distinguished from the seemingly similar MMEJ pathway by several factors. First is the extent of end-resection carried out and the length of the associated annealing intermediates as MMEJ relies on ≤ 10 bp of homology, while SSA complementarity can be much longer[208]. Second is the proteins required to mediate the annealing. In MMEJ this is done by PARP1/PolQ, whilst in SSA the length of the annealed sequence requires Rad52.

1.4.2 Single-strand template repair (SSTR)

Single-stranded template repair (SSTR) is a homology-directed pathway, similar to SDSA and BIR. SSTR is not yet well understood in humans, but emerging research is increasing our understanding in budding yeast.

SSTR in yeast is initiated through end resection by the MRN-CtIP complex, facilitating the Rad52-mediated annealing of the ssDNA 3′ end to local regions of homology (Fig. 1.13).This annealing is Rad51-independent, with one study demonstrating Rad51-mediated repression of the pathway when expressed[209,210]. Following this, the ssDNA is used as a template for DNA synthesis by DNA Pol δ and Rad52 preferentially anneals the new strand to the other end of the DSB, displacing the ssDNA template. Finally, Pol δ completes second-strand synthesis and gaps on the opposing strand are filled by DNA Pol ζ[211].

It is unknown how well this translates into humans, but increasingly differences are emerging between organisms. A 2018 study in human cells demonstrated that efficient integration of template ssDNA into the genome was reliant on the FA repair enzymes and related proteins such as CtIP and HelQ[212]. To date, only one yeast homolog of the FA repair pathway, Mph1 (FANCM) has been identified and was found non-essential for SSTR in yeast[210]. The role of FA repair proteins in human SSTR remains unknown, but it has been suggested that the FA core complex determines pathway fate by localising to DSBs to repress NHEJ and other HR pathways.

SSTR has gained interest in the last decade as an important factor in successful gene-editing. Multiple studies in eukaryotes have now demonstrated that single-stranded templates are integrated through this pathway[210,212]. This will be explored further in section 1.7.2.
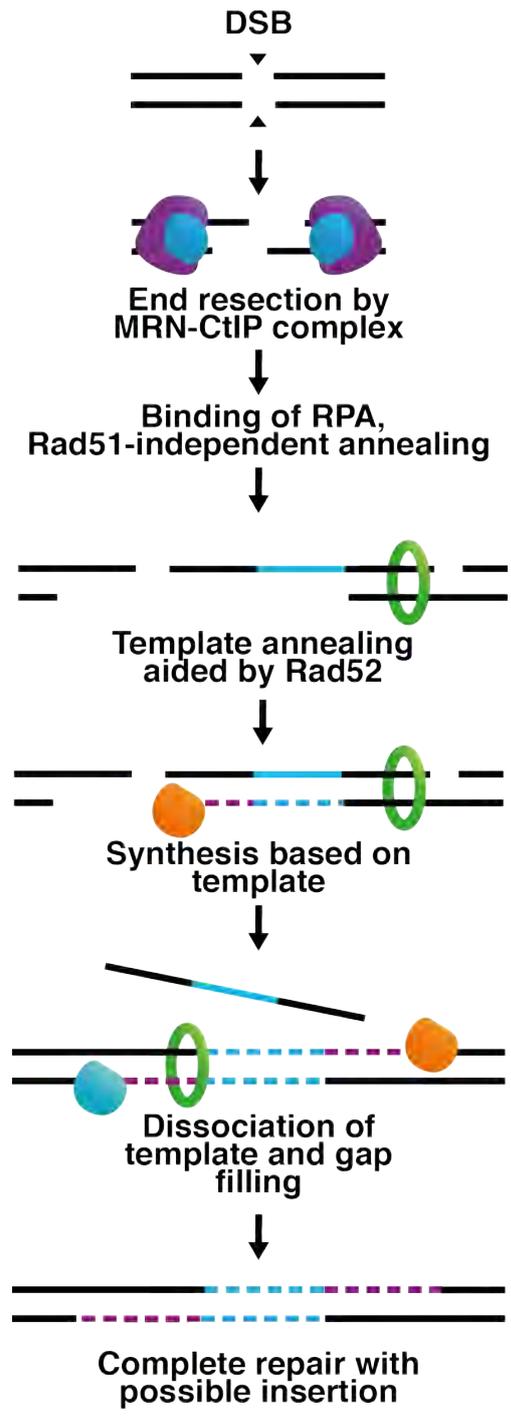
Figure 1.13: A model for eukaryotic single-strand template repair. Following end-resection by the MRN-CtIP complex (Purple-blue), ssDNA anneals to local regions of exposed homology, assisted by Rad52 (green). The template is speculated to be copied by low-fidelity polymerase DNA Pol δ (orange). The ssDNA template then dissociates and copied DNA anneals to the resected end of the break site. Gaps are then predicted to be filled by DNA Pol ζ (blue). Figure produced by the author.

1.4.3 Interstrand cross-link (ICL) repair

ICL repair is a complex process centred on a network of proteins drawn from multiple repair mechanisms commonly referred to as the Fanconi Anaemia (FA)/BRCA pathway. Aberrant expression of these proteins results in FA, an autosomal recessive disorder typified by bone marrow failure and a range of physical and/or neurological abnormalities[213]. FA also presents as a predisposition to multiple cancers, including those of the head and neck, breast, ovaries, and acute myeloid leukaemia[213,214].

The canonical route for ICL resolution is replication-coupled repair through the FA/BRCA pathway, which is active in S-phase cells[215]. This begins when a replication fork encounters an ICL, activating ATR/CHK1 and the FANCM/MHF1/MHF2/FAAP24 complex[216]. These mediate formation of the FANC core complex at ICLs alongside the FANCI-FANCD2 complex. Assembly of the core complex leads to strand nicking, or "unhooking", by a nuclease, suggested to be SLX4/SLX1, MUS81/EME1, XPF/ERCC1 or FAN1[216,217]. XPF/ERCC1 appears particularly important in ICL repair, with multiple studies of *XPF -/-* models demonstrating severe sensitivity to ICL-inducing agents such as cisplatin[217]. Depending on the structural context, unhooking results in the generation of single, or double-strand breaks which are repaired through the BIR or DSBR repair pathways, respectively.

Other, non-canonical pathways have also been observed for the resolution of ICLs during G1/G2 phases of the cell cycle[218]. This has been demonstrated in eukaryotes to be completed through the NER pathway, which carries out the ICL-incision steps of repair before remaining gaps are filled by translesion polymerases such as DNA Pol ζ, κ, and η[219,220].

The FANC proteins have also been observed to facilitate replisome traversal of ICLs, with the complex extensively remodelled and reassembled on the other side of the lesion without immediate repair[221]. This is largely dependent on the translocase activity of FANCM and its partners, notably

PCNA and BLM[222].

ICLs are powerfully cytotoxic lesions, often exploited in the treatment of cancers. Many clinically relevant compounds including nitrogen mustards, mitomycin C and cisplatin achieve their antiproliferative effects through the generation of ICLs[223]. The FA repair pathway is increasingly important in cancer research as multiple sub-species of ICL-resistant tumours have been discovered, requiring deeper understanding of the proteins responsible.

1.5 HelQ

Helicases comprise a suite of multi-functional enzymes ubiquitous in biological processes[224]. They are essential for maintaining genome stability and defects in their expression or function are defining factors in multiple developmental and neurodegenerative diseases, as well as cancers. Functionally, helicases combine ATP and ssDNA/RNA binding-dependent DNA/RNA translocation to disrupt hydrogen bonds between nucleotides[224].

HelQ is an SF2, Ski-2-like helicase originally described in a study searching for homologs of the *Drospohila* DNA repair helicase *Mus308*[225]. Research indicates that the protein participates in human DNA repair and replication recovery, both directly and through interactions with other proteins, as will be discussed in more detail below[225–229]. Research has also implicated HelQ in successful CRISPR-mediated gene-editing, alongside multiple components of the FA repair pathway, via SSTR. Using CRISPRi, one study identified that knockdown of HelQ resulted in a substantial decrease in the integration efficiency of single-stranded donor oligonucleotides (ssODNs)[212]. This poses the question, how does HelQ function, and what is the significance of that role in successful gene-editing?

1.5.1 Overview of the DNA repair helicase, HelQ

1.5.1.1 Helicase classification

Helicases can be divided into six superfamilies (SF1–6) based upon sequence similarity, oligomeric state, and the composition and organisation of nine signature motifs (see section 1.5.1.2)[230]. Superfamilies 1 and 2 comprise the majority of helicase enzymes, covering a range of families interacting with DNA and/or RNA. SF1 and SF2 are typically monomeric, translocating along oligonucleotides with 3′–5′ polarity[230]. The SF3–6 helicases meanwhile, form hexameric, ring-like complexes typically involved in DNA replication. The families also differ in polarity, with the SF4 helicases translocating in a 5′–3′ direction, whilst the SF3 and SF6 translocate 3′–5′[231]. SF5 helicases are an exception to functions in DNA replication, instead participating in other processes such as transcription termination, translocating in a 5′–3′[231].

1.5.1.2 Domain and motif organisation of SF1–2 helicases

The SF1 and SF2 helicase families can be defined by the presence or absence of nine conserved motifs: Q, I, Ia, Ib, II, III, IV, V, and VI, which form a 200–700 aa helicase 'core' responsible for translocation[232,233]. This consists of two similar domains that resemble recombination protein RecA[230]. Motifs Q, I, II, III, V and VI are most commonly associated with processes required for ATP binding and hydrolysis, while motifs Ia, Ib and IV are associated with DNA/RNA binding[230,233]. Q has also been identified as a regulator of NTP-binding specificity, while III and V have been associated with intramolecular interactions such as coordination of the ATP and DNA/RNA binding pockets[233,234].

The most highly conserved regions of the SF1 and SF2 helicases are motifs I and II, commonly annotated as Walker A and B ATPase boxes, respectively[235]. These motifs bind to the α and β phosphates of NTPs as

well as coordinating the NTP-associated Mg^{2+} required for hydrolysis[235]. This assists in coupling energy generation from NTP-hydrolysis to helicase activity via other motifs.

1.5.1.3 HelQ

As an SF2 helicase, HelQ possesses a highly-conserved core composed of two RecA-like domains containing both the Walker A and Walker B motifs crucial to ATP binding and hydrolysis (Fig. 1.14)[225,229]. This region also contains the ratchet and winged-helix domains (WHD) (Fig. 1.14) which participate in unwinding coupled to the activity of ATPase domains[236,237]. Mutagenesis studies of the WHD have shown that altering key residues such as Y818 and K819 removes the ability of the protein to bind to dsDNA, supporting the domain's predicted contribution to substrate selection and unwinding[236].

The N-terminal region of HelQ has been shown to be intrinsically disordered relative to the rest of the protein[229]. Despite this, the N-terminus also possesses a conserved PWI-like domain (Fig. 1.14) which appears to be responsible for the displacement of RPA from ssDNA[229].

The current oligomeric status of HelQ is uncertain, with early work suggesting that the enzyme exists as a hexamer based on the results of gel-filtration experiments[225]. This would be unusual however as SF2 helicases typically function as monomers. More recent work using size-exclusion chromatography coupled multi-angle light scattering (SEC-MALS), alongside gel-filtration has suggested a model where the enzyme exists in an apo-state as tetramers ready to be deployed as an activated dimer upon being recruited to ssDNA[229].
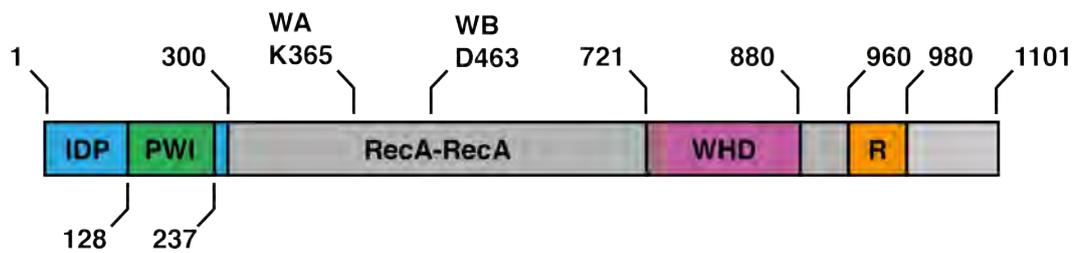
Figure 1.14: Domain map for HelQ. The first 300aa region is predicted to be intrinsically disordered protein (IDP). The remainder forms a 'core' helicase domain of 826 aa and comprises dual RecA-like domains, a winged helix domain (WHD), and a helicase 'ratchet' (R). Also indicated are the key residues of the ATP binding Walker A (WA) and ATP hydrolysis Walker B (WB) active sites. Figure produced by the author.

1.5.2 Overview of HelQ activity

1.5.2.1 Helicase mechanisms

Helicases utilise chemical energy obtained from the hydrolysis of nucleotide triphosphates (NTPs) to power translocation along DNA/RNA. The strand separation associated with the enzymes can result from either passive or active mechanisms, so called for the specific impact that the protein has on the stability of a bound duplex[238,239].

1.5.2.2 Helicase mechanisms: passive unwinding

Passive helicases have no role in melting the DNA duplex instead relying on transient, upstream fluctuations to expose sections of ssDNA which are then bound and trapped by the enzyme[239]. As the junction of dsDNA and ssDNA continues to move, the helicase utilises NTPs to translocate along the exposed strand. This is seen in the SF-4 bacteriophage T4 helicase gp41[240]

1.5.2.3 Helicase mechanisms: active models for unwinding

The inchworm model for unwinding describes coordinated, alternating binding of DNA by two distinct sites within the helicase. Here, one site binds to

exposed ssDNA and the second binds both ss- and dsDNA. Translocation along the bound strand is coupled to ATP binding and hydrolysis, which causes conformational shifts between 'open' and 'closed' states in the helicase (Fig. 1.15), as is seen in the *Bacillus subtillis* helicase, PcrA[241].

A variant model, the cooperative inchworm, described the progression of monomeric helicases through obstacles on the DNA[242]. This is seen with the T4 phage helicase Dda, which can more efficiently overcome streptavidin roadblocks when utilising several monomers[243]. This increase thought to result from a backlog of multiple monomers translocating in the same direction, preventing the leader from rebounding upon stalling[244].

Another model, steric exclusion is employed by many ring-like helicases of SF-3–6 and relies on ATP hydrolysis and step-wise conformational change to unwind DNA. The exact mechanism of translocation and unwinding remains unclear, but it is thought that the helicase complex encloses the bound strand whilst preventing the unbound from doing the same, causing separation as the helicase translocates[245].

The active rolling model of helicase translocation has been proposed for enzymes which function as oligomeric complexes. Here, the constituent proteins of the complex translocate in a 'hand-over-hand' movement, alternating the bound subunit at exposed ssDNA or ss-dsDNA junctions via an ATP-dependent isomerisation step, as observed with the *E. coli* Rep helicase[239].

1.5.2.4   HelQ

Research into the function of HelQ has identified it as a DNA-associated helicase translocating with 3′–5′ polarity[225]. This work also established the processivity of HelQ, demonstrating that the protein could translocate through up to 70 nt, although with drastically reduced efficiency once the length exceeded 40 nt. Subsequent work established that HelQ unwinding has a strong preference for substrates resembling stalled replication forks,

and in particular those with a nascent lagging strand[226]. The recruitment of HelQ to these regions is through interaction with RPA[225,226]. While the precise mechanism and context of HelQ remains ambiguous, the recruitment of the protein to replication forks and its promotion of, or participation in, several DNA repair pathways, including HR and FA, suggests that the protein functions as a mediator of replication-coupled repair[226–229].
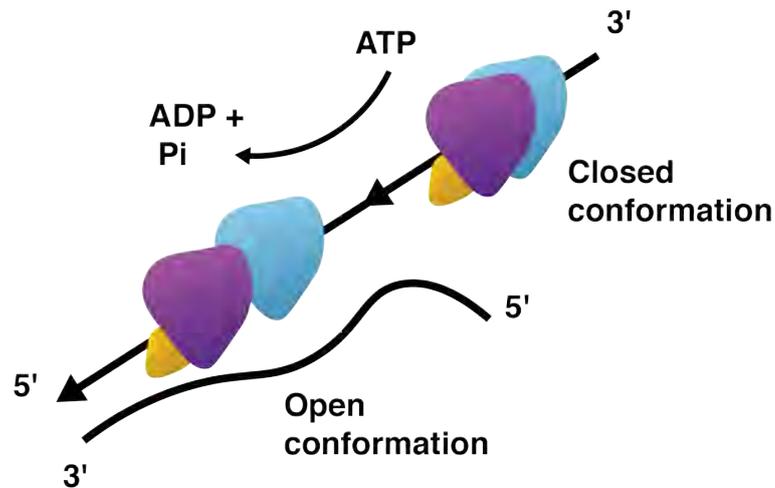
Figure 1.15: Schematic representation of SF1-2 helicases translocating on DNA SF-1/2 helicases utilising the inchworm model translocate by the alternating movement of separate domains (blue, purple). As the protein translocates along the strand, base-pairing is disrupted to produce more ssDNA. The Ski-like and DEAH/RHA helicases of SF2 achieve this through a conserved β-hairpin (yellow). Figure produced by the author.

1.5.3   Overview of HelQ interactions with proteins and pathways

In elucidating the function of HelQ it is important to consider its interactions with other proteins and the potential for these to be significant in contributing to the success of CRISPR-mediated gene-editing. Multiple studies have shown that HelQ co-localises *in vivo* at ssDNA with a number of key HR and FA repair associated proteins including RPA, ATR, FANCD2, and subunits of the Rad51 paralogue complexes BCDX2 and CX3 (Fig. 1.16).

1.5.3.1 RPA

Early studies of HelQ identified that the presence of RPA stimulated the recruitment and unwinding of DNA by HelQ[225]. HelQ has also been shown to displace RPA from ssDNA through and interaction with RPA70, the ssDNA binding subunit of the heterotrimeric RPA complex[227,228]. This interaction is thought to occur via a predicted PWI domain in HelQ, which may trigger remodelling of the RPA complex and may provide the mechanism for HelQ recruitment to stressed/stalled replisomes[229].

1.5.3.2 Rad51 paralogues

HelQ has been shown to interact with subunits of the Rad51 paralogue complexes BCDX2 and CX3, participating in the disruption of Rad51 binding to dsDNA[227,228]. Also observed is an absence of change in the frequency of crossover recombination when HelQ is compromised[228]. Recent work has suggested that these interactions may be indicative of a function for HelQ in MMEJ and/or SDSA[229]. This interaction is supported by research in *C. elegans* which reported that the combined loss of HelQ and Rad51 paralogues blocks the progression of HR, likely due to a lack of Rad51 filament disassembly from stalled forks which may act as a checkpoint for HR progression[246].

Interaction between HelQ and the subunits of the Rad51 paralogue complexes, particularly Rad51C, may hold further significance in the context of successful gene-editing. Research has established that the CX3 subunits Rad51C and XRCC3 are required for SSTR, suggesting that these complexes could promote recombination between the ssODN and genomic DNA[212]. As such, a deeper exploration of the interactions between these proteins may yield insight into the role of HelQ and SSTR in successful gene-editing.

1.5.3.3 Other interactors

HelQ participation in HR has also been observed through interactions with several other proteins. This includes HR regulation through epistatic expression with the histone demethylase JMJD5, as observed in *C. elegans*, which is thought to facilitate Rad51 removal from stalled recombination intermediates, thus progressing HR[247]. HelQ is also redundant in the completion of HR with the HROB-Mcm8-Mcm9 helicase complex[248]. The depletion of both proteins was observed in human cells to result in a cumulative increase in sensitivity to cisplatin, suggesting that both helicases act through independent pathways to resolve postsynaptic HR intermediates[248]. Finally, HelQ has been shown through proteomic analysis to associate with the replication checkpoint kinase ATR, as HelQ enrichment on chromatin in response to ICL-inducing agent-mediated replication fork stalling was compromised by ATR inhibition[228].

HelQ has also been associated with the FA pathway ICL repair[226–228]. Multiple studies have observed that HelQ deficient cells display increased sensitivity to crosslinking agents such as MMC, with one also identifying colocalisation between HelQ and FancD2[226–228]. This is supported by studies in model organisms involving analogous proteins. A study in *C. elegans* reported decreased survival following cross-linking treatments in HelQ deficient strains[249]. Likewise, a murine study found no epistasis between HelQ and FANCC, despite a deficiency of both resulting in FA-like symptoms[250].

The association of HelQ with FA repair was also observed in research implicating FA proteins in CRISPR-mediated ssDNA template integration[212]. The work found that alongside silencing FA proteins, the knockdown of HelQ had a profound impact on integration efficiency[212]. Despite these associations, no direct interaction has been detected between the proteins leading to the prediction that HelQ acts independently or in parallel to the FA repair proteins[250].
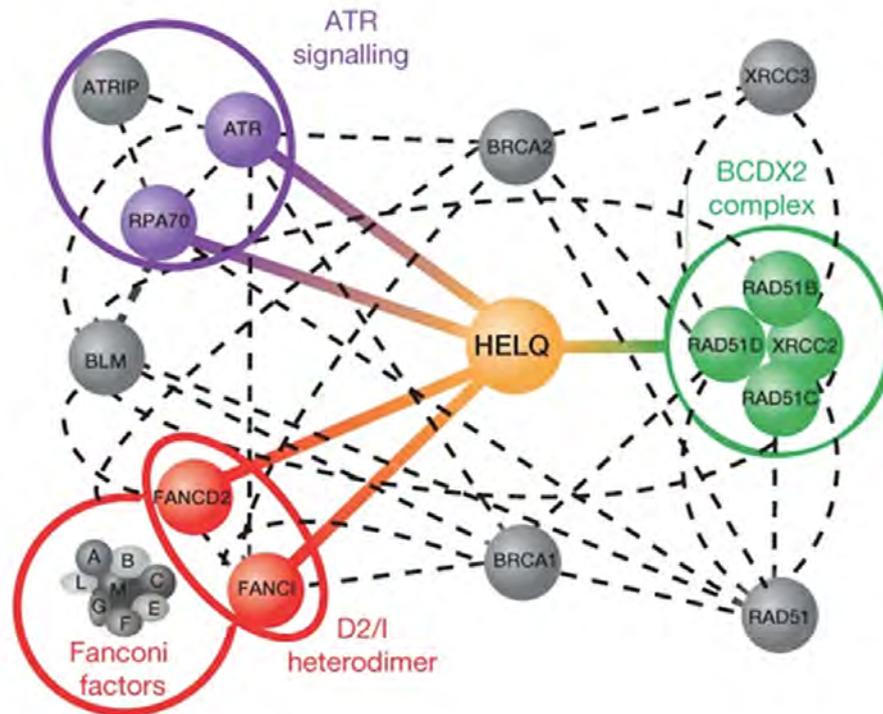
Figure 1.16: Map of PPIs for HelQ. An interaction network for HelQ was modelled based upon mass spectroscopy results (solid lines) as well as reported interactions from BIOGRID, STRING and MINT databases (dashed lines). Image taken from Adelman et al.[228] with permission.

1.6 Drug Discovery

1.6.1 A role for HelQ and other DNA-repair helicases in disease

The role of helicases in DNA processing and remodelling is essential to cell health and perturbations to expression or function have been associated with a range of diseases. Examples of this include the replication/recombination-associated helicases WNR and BLM, for which aberrant expression results in Werner's syndrome and Bloom's syndrome, respectively[224,251]. These conditions are both typified by an increase in illegitimate recombination events and chromosomal rearrangements that result in an increased chance of, and in some cases predisposition to, many types of cancers[224,252–254].

The role of HelQ in DNA repair and the maintenance of genome stability

has been shown to influence the development of cancers in the event of deregulated expression[255–257]. Several papers have identified the protein as a biomarker for the prognosis and likelihood of reoccurrences of multiple cancers included those of the ovary, testes, head, and neck[258,259]. HelQ has also been associated directly with cancer progression and treatment, being observed in a sub-set of ovarian tumours to generate resistance to chemotherapuetic treatment with Cisplatin[260].

1.6.2 Modern drug discovery

Cytotoxic agents, such as nitrogen mustard and mitomycin C, have been used for decades as a primary cancer treatment[261,262]. Their mechanisms of action however are incredibly broad, causing indiscriminate damage to cells throughout the body. Alongside this, research is increasingly identifying chemotherapy-resistant subclasses of tumours which are able to overcome cytotoxic effects[263,264]. As such, focus is gradually shifting toward the development of tumour-specific agents with reduced toxicity to non-cancerous cells and increased efficacy[264,265].

Additionally, the success of PARP inhibitors (described in more detail in section 1.6.3.1), has directed the development of new second-generation anti-cancer drugs towards proteins participating in the DNA-damage response, particularly those that modulate the progress of pathways such as HR. This includes helicases such as BLM, for which recent work has identified a new class of isaindigotone derivatives which are able to disrupt unwinding activity and regulate HR[266].

This search is supported by powerful sequencing technologies, which have made available wide ranging data on areas of the genome yet to be explored, and new screening platforms such as RNAi and CRISPRi[267]. Together, these approaches have enabled the launch of accessible databases which can define targets by metrics such as synthetic lethal interactions (see

section 1.6.3), providing both academic and pharmaceutical research with an abundance of new targets[268,269]. This has opened up a large amount of chemical space to explore, requiring powerful methods to identify promising candidates for development.

Among the many DNA repair proteins being targeted both academically and commercially for the development of anti-cancer therapies is the HelQ homolog, PolQ[270]. Recent research has found that this protein has a synthetic lethal association with *BRCA*-mutant cancers and also that inhibitors against PolQ act synergistically with PARP inhibitors to overcome PARP-resistant tumours[271]. The similarities between HelQ and PolQ, as well as the entry of PolQ inhibitors to clinical trials in 2021, positions HelQ well as a potential target for the treatment of cancers, which will be further explored in section 5.1.

1.6.2.1 High-throughput put screening

High-throughput screening (HTS) technologies saw great development throughout the 1990s and are currently the dominant drug discovery approach for the pharmaceutical industry. As of 2019 multiple companies reported libraries containing up to 4×10^6 compounds[272].

HTS can be applied to biochemical functional assays using purified proteins, or cell-based assays using immortalised or primary cell lines[273]. They are most powerful when little is known about the protein target(s), preventing structure-based drug-design. The approach has seen success in identifying leads that would go on to become licensed therapies. This includes the chemokine receptor agonist maraviroc, an antiretroviral drug used in the treatment of HIV[274].

HTS is limited by high start-up costs, as with increasing library size comes an increase in the resources and equipment required for screening[275]. They have also historically been limited by low compound diversity as

early libraries predominantly comprised chemical collections synthesised by pharmaceutical companies for other commercial activities[276]. Efforts have since been made to increase library diversity to improve efficiency[277,278].

Efforts have also been made to increase HTS library size, but the enormity of chemical space makes this futile, with estimates placing the number of small drug-like molecules at 10^{63}[279]. This has led to approaches which seek to increase the efficiency with which chemical space is explored rather than simply increasing compound number.

1.6.2.2 Fragment-based drug discovery

The principles that would become fragment-based drug discovery (FBDD) were proposed in the 1980s, but proof-of-principle was not published until 1996[280,281]. In the 25 years since, FBDD has developed rapidly to emerge as a powerful approach to modern drug discovery.

Compared to HTS, FBDD libraries are incredibly small, typically containing around 2 000 compounds[282]. Despite this, the strength of FBDD lies in its diversity, utilising a wide array of functional groups and scaffolds to efficiently assay chemical space[282,283]. This provides a statistically higher probability of identifying hits from which to build drug-like molecules. Careful design is important however and FBDD libraries often contain some similar compounds to enhance hit clarification and certainty[284].

Compounds within libraries typically fit the "rule-of-three" which describes ideal compounds as having a molecular weight $\leq$300 Da, a ClogP value $\leq$3, and $\leq$3 hydrogen donors/acceptors[285]. The resulting compounds typically have a higher hydrophilicity, and therefore greater solubility, than those found in HTS libraries. They also have the disadvantage of reduced specificity due to their small size, making them less effective at class-specific targeting.

Fragments often represent the ideal binding motif for inhibition, avoiding

unfavourable interactions such as steric clashes due to their small size[286,287]. This facilitates interactions with hard-to-target allosteric sites or small binding pockets that form the basis of many protein-protein interactions[287]. Fragments have lower affinities for target molecules, typically in the high-μM to low mM range, versus the low nM of the larger compounds in HTS libraries[284]. This is overcome during hit development, which fuses fragments using combinatorial chemistry to achieve higher-affinity binding.

The power of FBDD in generating successful lead compounds was demonstrated with the approval of vemurafenib in 2011, six years after precursors were identified through screening[288]. Vemurafenib inhibits the B-RAF^{V600} mutant which is implicated in tumour suppressor inactivation and the inhibition of apoptosis through destabilisation of the MAPK signalling pathway[289,290]. This is a potent chemotherapy as V600 accounts for 90% of all BRAF mutations and has been identified in 50% of melanomas[291]. The success of FBDD has since been repeated; as of 2021, four drugs derived from FBDD-screens have been approved, with 40 additional candidates undergoing clinical trials[292–295].

1.6.3 Synthetic lethality

First reported in 1922, synthetic lethality describes a relationship between two genes in which the aberrant expression of one remains viable, but the mutation of both is lethal (Fig.1.17). This was developed from genetic studies in *Drosophila* which observed that certain combinations of mutations did not produce viable offspring upon cross-breeding[296,297].

In 1997 synthetic lethality was proposed as an avenue for the development of anti-cancer drugs by identifying pairings between cancer-specific mutations and partners required for viability[298,299]. This is an attractive route to therapy as only cancer cells would be sensitised, potentially facilitating treatment of otherwise undruggable targets[299]. For example, the

Myc family of transcription factors are overexpressed in most cancers but are essential to proliferation and so cannot be directly targeted. Synthetic lethality screens have identified several relationships, such as transcriptional regulator BRD4, which when inhibited lead to the down-regulation of MYC expression and the initiation of apoptosis[300].

Synthetic lethality is a promising approach to drug-discovery, but currently only one new class, PARP inhibitors, has been successfully licensed for clinical applications. Several more clinical trials are ongoing to determine the efficacy of candidates identified through synthetic lethal screens such as the relationship between *KRAS*-mutant cancers and MEK inhibitors[301,302].

1.6.3.1 Case study: PARP inhibitors

Using synthetic lethality methods, a relationship was discovered between PARP and the breast cancer susceptibility genes BRCA1 and BRCA2[304]. Subsequent development resulted in the use of PARP inhibitors in clinical trials against BRCA1/2 germline-mutated tumours, with the first successful drug, olaparib, approved in 2014[305].

The PARP family is involved in several critical cellular processes, including the stress response, chromatin remodelling, DNA repair and apoptosis[306–308]. PARP1 has been implicated in multiple repair pathways, predominantly SSBR and BER[306,309]. Upon detection of a DNA strand break, PARPs catalyse the addition of ADP-ribose to several proteins, including XRCC1 and PolQ, making it an important recruitment factor for repair initiation[310]. PARP inhibitors disrupt this by competing with NAD for the catalytic pocket of the enzyme, although the exact biological mechanism of disruption and resulting lethality remains controversial. Multiple mechanisms have been suggested including the accumulation of SSBs, replication fork stalling and the upregulation of NHEJ[311–313].

Since 2014, four PARP inhibitors: olaparib, rucaparib, niraparib, and
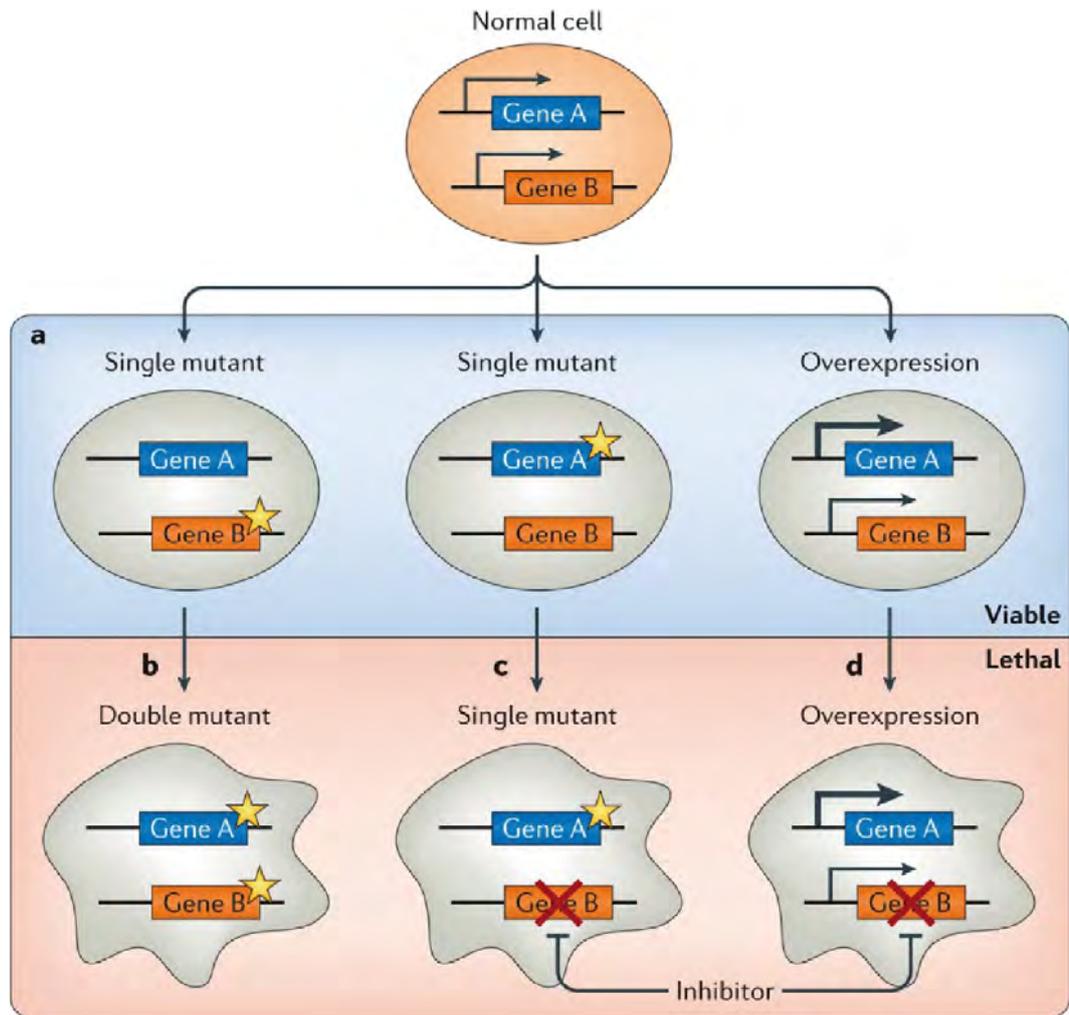
Figure 1.17: The basic pathway of synthetic lethality. (A) Mutations leading to the inactivation, or overexpression, of a single gene in the pair remains viable. (B-D) An additional mutation or pharmacological inhibition of the second gene in the pairing is lethal to the cell, resulting in death. Figure taken from O'Neil et al.[303] with permission.

talazoparib have been approved for the treatment of BRCA1/2 mutated cancers and also recurrent, epithelial ovarian, fallopian tube, or primary peritoneal cancers irrespective of BRCA status[314]. Their success has spurred interest in synthetic lethality and small-molecule screens, leading to the development of new candidates from the DDR pathways. This includes inhibitors of PolQ, set to enter clinical trials in 2021, and pre-clinical screens against WRN which has been identified in several studies as a synthetic-lethal target in microsatellite-unstable cancers[270,315–319].

1.7 The nexus of CRISPR technology, DNA repair and drug discovery

The fields of drug discovery and DNA repair research were for some time limited to known phenotypes for gene target identification, later relying on targetable nucleases that are laborious to program and validate. The versatility of CRISPR-Cas systems has facilitated simple and rapid phenotypic model generation for the study of DNA repair proteins in cancer and has become a powerful tool for research and drug discovery.

The simplicity and programmability of CRISPR give it an edge over other modular nucleases such as ZFNs or TALENS. This has given rise to an ever-increasing suite of functional and discovery tools that elevate Cas-proteins above their nuclease activity to assist in a range of disciplines.
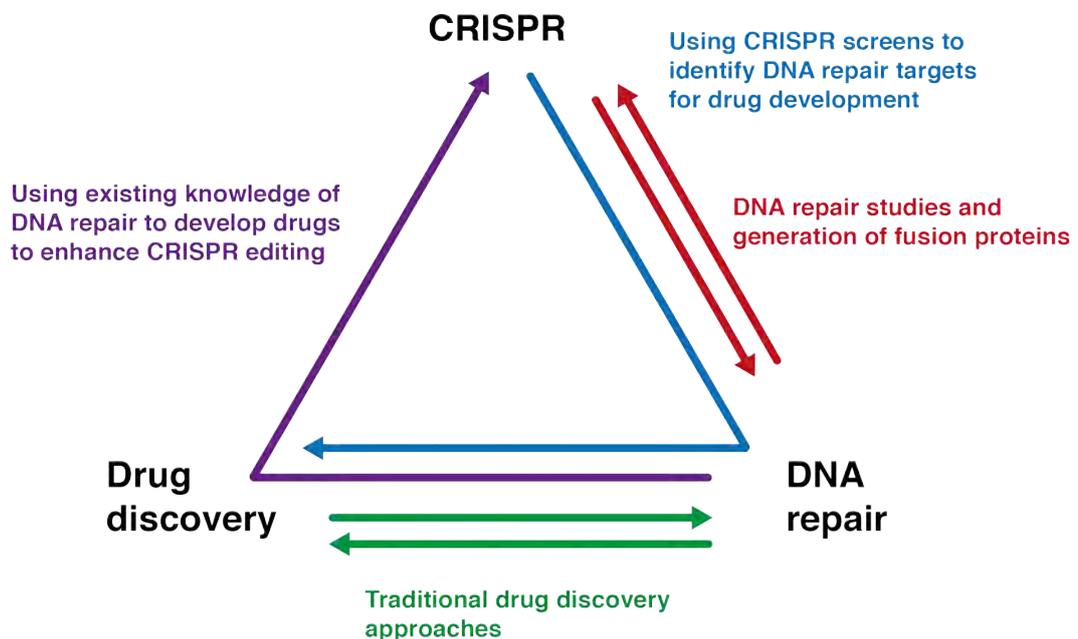
Figure 1.18: Cooperativity between DNA repair, CRISPR and drug discovery approaches. Interdisciplinary research can have powerful outcomes, most commonly seen in traditional drug discovery. The adoption of gene editing has lent even greater potential to molecular and therapeutic research, but it is currently limited by several milestones. Figure produced by the author.

1.7.1 NHEJ and MMEJ-based editing tools

The error-prone nature of NHEJ has seen it exploited for gene knockout generation since the adoption of CRISPR as an editing tool. Alongside this, alternative tools have emerged, exploiting the high incidence of the pathway. CRISPaint (CRISPR-assisted insertion tagging), facilitates tagging of target proteins by editing near a target gene to integrate sequence encoding an in-frame 'tag' such as a fluorescent protein[320]. Another tool, VIKING (Versatile non-homologous end joining-based knock-in module for genome editing) uses Cas9 RNPs to linearise transfected DNA at a VKG1 sequence widely found in plasmid vectors[321]. Linearisation optimises Ku binding and increases the probability of donor incorporation into genomic DSB sites.

NHEJ-derived techniques are limited by a risk of off-target integrations and InDel generations at the site of Cas-mediated DSBs. For integration-based technologies such as CRISPaint and VIKING, the use of non-homologous repair carries a risk of integration out of frame or in the reverse conformation, resulting in a failure to correctly express inserted DNA.

Due to increased activity relative to HR throughout the cell cycle, and a lower rate of error that NHEJ, MMEJ is also being developed as a possible alternative editing route. A notable development for this pathway is PITCh (Precise Insertion into Target Chromosome), which is able to insert microhomology-flanked DNA cassettes into a Cas9-induced DSBs at higher efficiencies than HR[322,323]. A further technology, mHAX (microhomology assisted excision) is capable of scar-less selectable marker removal from insertion sites[324].

1.7.2 HDR-based editing tools

Pathways which require template DNA, such as DSBR, are often grouped under the broad term "homology-directed repair" (HDR) in the context of gene-editing. Due to the lower incidence of HDR pathways relative to NHEJ,

most tools for CRISPR-mediated editing aim to assist higher efficiencies of integration by optimising template DNA.

Research has demonstrated donor-dependent differences in editing efficacy based upon the chosen template format, finding ssODNs to be the most efficient[325]. These can be further optimised by end-protection from exonucleases by phosphorothioate-modifications at the 5′ and 3′ ends[326]. Further optimisation is possible by asymmetric design of homology arms, with multiple studies finding that ssODNs comprising 36 bp homology on the PAM-distal side, and 91-bp homology on the PAM-proximal side of the break exhibit increased integration[327,328].

1.7.3 Cas-protein fusions

The advent of Cas9 as a targetable nuclease also saw the development of nuclease-dead and nicking alternatives. These proteins, particularly dCas9 have been used to develop tools for as varied tasks as transcriptional regulation and fluorescent imaging[329].

Cas-protein fusions have also been created to develop powerful screening tools such as CRISPR-interference (CRISPRi) and CRISPR-activation (CRISPRa) in which dCas9 is fused to transcriptional repressors such as KRAB, or activators such as SAM, respectively. These have seen wideranging use in synthetic lethality screens and mechanistic studies[212,330,331].

Further tools have been developed to enhance editing by HR. One study generated Cas9-CtIP fusion proteins, the first utilising full-length CtIP to stimulate increased editing compared with standard HR, and the second enhancing editing using an N-terminal fragment of CtIP crucial to its initiation of HR[332]. Another study developed a RAD52 fusion with Cas9 demonstrating enhanced reporter insertion at the desired site[333].

Further developments includex base-modifying tools that exploit pathways such as BER and MMR to achieve small, precise edits by fusion with

enzymes such as cytidine deaminase[334,335]; RNA editing enzymes that utilise catalytically inactive Cas13 in place of dCas9[336]; and finally, some systems that have sought to harness site-specific recombinases for large-scale alterations[22,23].

1.7.4 Drug discovery approaches

CRISPR gene editing tools are currently limited by repair pathway determination. Therapeutic approaches relying upon homology-directed repair are powerful, but their incidence is much lower than NHEJ. To counter this, several studies have attempted to suppress NHEJ using small molecule inhibitors or silencing the expression of associated proteins[337–340]. One study utilised small molecules NU7441 and KU-0060648 to inhibit DNA-PKcs, achieving a reduction in NHEJ repair by 40% and a two-fold increase in successful HR[337]. Another study used Scr7, a small molecule inhibitor of DNA ligase IV, to achieve a several-fold increase in HR-mediated template integration at multiple genomic loci in human cells[340]. The drawback to these approaches is a lack of specificity. By globally inhibiting NHEJ, a treatment may cause unintended instability elsewhere in the genome.

1.7.5 Barriers to progress

CRISPR-Cas has revolutionised many approaches to drug discovery and molecular research, and holds strong promise for therapeutic applications. Despite this, several crucial roadblocks remain before it can reach its full potential, namely: PAM stringency, off-target effects, repair pathway choice, efficiency of integration, and immunogenicity in therapeutics[341]. Several of these issues are not with CRISPR-Cas systems themselves, but with our understanding of DNA repair. To date, multiple enzymes of unknown, or poorly understood, function have been associated with strong impacts on the efficiency of CRISPR-mediated gene editing[212].

Work on ssODNs has raised interesting questions about how repair pathways recombine the template into a chromosome. The prevailing theory is that the host relies on HR-mediated pathways such as SDSA or SSTR, with the latter in particular gaining new interest due to its increased activity relative to standard HR. A significant study identified the FA pathway and its associated proteins as crucial to ssODN-based gene-editing[212]. Knockdowns of HelQ were shown to have a strong impact on the incidence of SSTR-mediated template integration, possibly due to interaction with FANCD2 and the Rad51-paralogues. Without further study of such proteins, it may not be possible to overcome the barriers remaining to the full potential of CRISPR systems.

1.8 Research aims

This project aimed to better understand the mechanism of HelQ, for its perceived contributions to the success of CRISPR-based gene editing, and as a new target in the treatment of cancers. To achieve this, two major areas were focussed on: (a) an hypothesis that HelQ may aid Cas9-based editing by modulation of the Cas R-loop structure, (b) development of inhibitors against HelQ activities that could be refined into potential therapeutics. To this end, the specific research aims were:

- To study the interaction of HelQ with Cas9 R-loops using synthetic DNA substrates and supercoiled plasmids

- To study the interaction of HelQ with non-CRISPR R-loops using synthetic DNA and RNA substrates.

- To adapt the methods of Sansbury et al.[342,343] to create an *in vitro* cell-free model to study the impact of DDR proteins on editing efficiency, with HelQ as a model.

- The screening of a small fragment library to identify and characterise putative inhibitors of FL-HelQ for further development

2

Materials and Methods

2.1   Materials

2.1.1   Chemicals and Reagents

Unless otherwise stated, all chemicals used in this work were purchased from the following companies: Merck/Sigma-Aldrich (Darmstadt, Germany), Thermo Fisher Scientific (Thermo Fisher, Waltham, Massachusetts, USA) or VWR International (Radnor, Pensylvania, USA). Restriction enzymes, other commercial enzymes, and molecular weight markers (DNA and protein) were from obtained from New England BioLabs (NEB, Ipswich, Massechusetts, USA). DNA preparation and purification kits were purchased from QIAGEN (Hilden, Germany). Cell culture media were from Lonza (Basel, Switzerland) and selective reagents from Stratech (Ely, UK).

2.1.2   Consumables

Unless otherwise stated, all plasticware including microplates and dishes used for tissue culture and assays was purchased from Corning (Corning, New York, USA). Tissue culture flasks were purchased from Sarstedt, (Nümbrecht, Germany).

2.1.3   Small-molecule inhibitor library

The small molecule screen described in Chapter 5 was carried out using a 320-compound Essential Fragment Library from Enamine (Monmouth

Junction, NJ, USA). The library was a kind gift from Nanna Therapeutics (Cambridge, UK). The 19 candidate molecules also characterised in Chapter 5 were synthesised by Nanna Therapeutics and assessed for purity using mass spectrometry. The 27 compounds further characterised in Chapter 5 were synthesised by Sygnature Discovery (Nottingham, UK). Upon receipt, all small-molecules were resuspended to a concentration of 100 mM in 100% dimethyl sulphoxide (DMSO) and stored at -20 °C. Subsequent dilutions for assays, were made from the 100 mM stock.

2.1.4 Plasmids and oligonucleotides

Unless otherwise stated, all oligonucleotides used in this study were purchased from Sigma-Aldrich custom oligonucleotide synthesis.

2.1.4.1 Plasmids

Table 2.1: Plasmids used in this study

Name	Other name(s)	Information / use	Selection marker
pAC01	pSDC74	Addgene vector (pMJ806) containing His-MBP-tagged Cas9.	Kanamycin
pAC02	pSDC75	Addgene vector (pMJ841) containing His-MBP-tagged dCas9.	Kanamycin

Continued on next page

Table 2.1 – *Continued from previous page*

Name	Other names(s)	Information / use	Selection marker
pAC21	6His-MBP-TEV-huAsCpf1	Addgene vector (90095) pET-28 backbone containing humanised His-MBP-tagged Cas12a derived from Acidaminococcus sp.	Ampicillin
pAC23	pNLS	pACYC-Duet backbone altered to contain N-terminal SV40 and C-terminal nucleoplasmin NLS tags	Chloramphenicol
pAC29	pNLS_Cas9	Cas9 cloned into pAC23	Chloramphenicol
pAC36	pcDNA3.1_HELQ	GeneArt plasmid, human HelQ sequence cloned into pcDNA3.1 (+)	Ampicillin
pAC37		pAC36 with HelQ sequence mutated to encode D142F143A	Ampicillin
pAC38		pAC36 with HelQ sequence mutated to encode D463A	Ampicillin
pAC39		pAC36 with HelQ sequence mutated to encode Y642A	Ampicillin

Table 2.1 – *Continued from previous page*

Name	Other names(s)	Information / use	Selection marker
pAC40		pAC36 with HelQ sequence mutated to encode Q965A	Ampicillin
pAC41		pAC36 with HelQ sequence mutated to encode F965A	Ampicillin
pAC42		pAC36 with HelQ sequence mutated to encode V306I	Ampicillin
pAC43		pAC36 with HelQ sequence mutated to encode Y991XX	Ampicillin
pAC44		pAC36 with HelQ sequence mutated to encode K54G	Ampicillin
pAC45		pAC36 with HelQ sequence mutated to encode K365M	Ampicillin

2.1.4.2 Oligonucleotides

Table 2.2: Oligonucleotides used in this study

Name	Sequence (5′–3′)	Details / use
oAC49	[Cyanine5]AGGATCCGAC TTTCTCATAGACGATTA CATTGCTACATGGAGCT GTCTAG	plus strand for substrate cleavable by CAS9, Cas12a, CasX. Anneal to oAC50.
oAC50	CTAGACAGCTCCATGTA GCAATGTAATCGTCTAT GAGAAAGTCGGATCCT	minus strand for substrate cleavable by CAS9, Cas12a, CasX. Anneal to oAC49
oAC51	TAATACGACTCACTATA GGTAATTTCTACTCTTGT AGATTCATAGACGATTA CATTGCTA	plus strand for T7 HiScribe template expressing AsCas12a gRNA targeting dsAC49
oAC52	TAGCAATGTAATCGTCT ATGAATCTACAAGAGTA GAAATTACCTATAGTGA GTCGTATTA	minus strand for T7 HiScribe template expressing AsCas12a gRNA targeting dsAC49
oAC57	TAATACGACTCACTATA GGTAATTTCTACTCTTGT AGATCCCAGTCACGACG TTGTAAAA	plus strand for T7 HiScribe template expressing AsCas12a gRNA targeting pUC19

Continued on next page

Table 2.2 – *Continued from previous page*

Name	Sequence (5′–3′)	Details / use
oAC58	TTTTACAACGTCGTGAC TGGGATCTACAAGAGTA GAAATTACCTATAGTGA GTCGTATTA	minus strand for T7 HiScribe template expressing AsCas12a gRNA targeting pUC19
oAC59	CTATGCGGCATCAGAGC AG	Primer F to amplify fragment LacZa to confirm integration
oAC60	CGTATGTTGTGTGGAAT TGTGAGC	Primer R to amplify fragment LacZa to confirm integration
oAC61	T*GACTGGGAAAACCCT GGCGTTACCCAACTGCG GCCGCAATAATCGCCTT GCAGCACATCCCCCTTT CG*C	ssODN for Cas9 integration into pUC19 LacZa
oAC64	G*GGTTTTCCCAGTCACG ACGTTGTAAAACGTTGC GGCCGCACGGCCAGTG AATTCGAGCTCGGTACC CG*G	ssODN for AsCas12a integration into pUC19 LacZa
oAC67	GCTGAGCAATAACTAGC ATAAC	Primer F to amplify pACYCDuet-1 for gibson assembly

Table 2.2 – *Continued from previous page*

Name	Sequence (5′–3′)	Details / use
oAC68	GCTGTGGTGATGATGGTG	Primer R to amplify pACYCDuet-1 for gibson assembly
oAC69	ACCTGCCTTCTTTGTTGCAGCAGGACGTTTTGCGGCCGCAAGCTTGTCGACCTGCAGCGAATTCGGGACTTTGCGTTTCTTTTTTGGCTGGCTGTGGTGATGATGGTGATGGCTGCTGCC	Fragment for Gibson assembly containing half of NLS-cassette. Anneals with the + strand of pACYC fragment
oAC70	AAACGTCCTGCTGCAACAAAGAAGGCAGGTCAAGCCAAAAAGAAAAAGTGCTTATGGAGCCACCCGCAGTTCGAAAAAAGCGCGTAAAGGCTGAGCAATAACTAGCATAACCCCTTGGG	Fragment for Gibson assembly containing half of NLS-cassette. Anneals with the - strand of the pACYC fragment
oAC71	GATTCTGCAGGATAAGAAATACTCAATAGGCTTAGATATCG	Primer F to amplify Cas9 with PstI site for cloning into NLS plasmid

Table 2.2 – *Continued from previous page*

Name	Sequence (5′–3′)	Details / use
oAC72	GATTGCGGCCGCGTCAC CTCCTAGCTGACTC	Primer R to amplify Cas9 with NotI site for cloning into NLS plasmid
oAC94	TCACCTAGATCCTTTTA AACTTCACCTAGATCCT TTTAAACATTTCCCCGA AAAGTGCTAGTGGTGCT AGCCCCGCGAAATTAAT ACGACTCACTATAGGTA ATTTCTACTCTTGTAGAT	template DNA for Cas12a sgRNA cassette production
oAC95	TCACCTAGATCCTTTTA AACTTCAC	primer F to amplify Cas12a sgRNA cassette by PCR
oAC96	GTTTTACAACGTCGTGA CTGGATCTACAAGAGTA GAAATTACC	primer R to amplify Cas12a sgRNA cassette by PCR and add guide sequence targeting pUC19
oAC97	TAGCAATGTAATCGTCT ATGAATCTACAAGAGTA GAAATTACC	primer R to amplify Cas12a sgRNA cassette by PCR and add guide sequence targeting pUC19primer R to amplify Cas12a sgRNA cassette by PCR and add guide sequence targeting dsAC49

Table 2.2 – *Continued from previous page*

Name	Sequence (5′–3′)	Details / use
oAC98	AACGATTGCGGCCGC	test sequence for integration reactions using Cas12a (+ strand)
oAC99	TCGTTGCGGCCGCAA	test sequence for integration reactions using Cas12a (- strand)
oAC100	CCATTCGCCATTCAGGCTGC	primer to confirm integration of ssODN sequences using Cas12a
oAC103	CCATAATTGCATAGAGCAACC	sequencing primer for SDM V306I
oAC110	TCGGATCCTCTAGACAGCTCCATGTAGCAATGTAATCGTCTATGACGTTG	DNA complementary to MW14 used a trap DNA in helicase assays
oAC111	AGGATCCGACTTTCTCATAGACGATTACATTGCTACATGGAGCTGTCTAG	unlabelled version of oAC49
oAC112	[Cynanine5]UUGCUAAGAGCAAGAUGUUCUAUAAAAGAUGUCCUAGCAAGGCAC	labelled RNA for formation of R-loop with 5′ flap

Continued on next page

Table 2.2 – *Continued from previous page*

Name	Sequence (5′–3′)	Details / use
oAC113	[Cyanine5]AAAGAUGUCCUAGCAAGGCACGAUCGAGCGGAUAUCUAUGACCAU	labelled RNA for formation of R-loop with 3′ flap
oAC114	[Cyanine5|AAAGATGTCCTAGCAAGGCAC	labelled DNA for formation of D-loop with no flap
oAC115	[Cyanine5]TTGCTAAGAGCAAGATGTTCTATAAAAGATGTCCTAGCAAGGCAC	labelled DNA for formation of R-loop with 5′ flap
oAC116	[Cyanine5]AAAGATGTCCTAGCAAGGCACGATCGACCGGATATCTATGACCAT	labelled DNA for formation of R-loop with 3′ flap
oAC117	GGGTGAACCTGCAGGTGGGCGGCTGCTCATCGTAGGTTAGTTGGTAGAATTCGGCAGCGTC	oligo for formation of D/R-loops, + strand
oAC118	GACGCTGCCGAATTCTACCAGTGCCTTGCTAGGACATCTTTGCCCACCTGCAGGTTCACCC	oligo for formation of D/R-loops, - strand

Table 2.2 – *Continued from previous page*

Name	Sequence (5′–3′)	Details / use
oAC129	[Cyanine5]TCGGATCCTCT AGACAGCTCCATGATCA CTGGCACTGGTAGAATT CGGC	Cy-5 labelled MW12 ssDNA
oAC130	CAACGTCATAGACGATT ACATTGCTACATGGAGC TGTCTAGAGGATCCGA	partial complement to MW12, used to form Fork 2B
oAC131	[Cyanine5]CAACGTCATA GACGATTACATTGCTAC ATGGAGCTGTCTAGAGG ATCCGA	Cy-5 labelled MW14 ssDNA
oAC132	TGCCGAATTCTACCAGT GCCAGTGATCATGGAGC TGTCTAGAGGATCCGA	Complementary sequence for MW12 use to make dsDNA competitor for use in Cas-protein assays
oAC138	TAATACGACTCACTATA GG	primer F to add T7 promoter to Cas9 sgRNA cassettes
oAC139	AACCACCGACTCCCT	primer R to amplify Cas9 sgRNA cassettes

Table 2.2 – *Continued from previous page*

Name	Sequence (5′–3′)	Details / use
oAC140	TAATACGACTCACTATA GGTAGACGATTACATTG CTACAGTTTTAGAGCTA GAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTA TCAACTTGAAAAAGTGG CACCGAGTCGGTGCTT	template DNA for Cas9 sgRNA cassette containing guide sequence for dsAC49
oAC141	TAATACGACTCACTATA GGGTGCTGCAAGGCGAT TAAGTGTTTTAGAGCTA GAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTA TCAACTTGAAAAAGTGG CACCGAGTCGGTGCTT	template DNA for Cas9 sgRNA cassette containing guide sequence for pUC19
oAC145	ACTAACCTACGATGAGC AGCC	DNA complement to PM4, mimicking oAC114
oAC146	[Cyanine]ACTAACCTACG ATGAGCAGCC	Cy-5 labelled DNA complement to PM4, mimicking oAC114
oAC147	AAAGATGTCCTAGCAA GGCAC	unlabelled equivalent of oAC114

Table 2.2 – *Continued from previous page*

Name	Sequence (5′–3′)	Details / use
oAC148	CCAGCTGGCGAAAGGG GGATGTGCTGCAAGGCG ATTAAGTTGGGTAACGC CAGGGTTTTC	DNA containing sgpUC19 sequence to act as trap DNA for roadblock removal assays (+) strand
oAC149	GAAAACCCTGGCGTTAC CCAACTTAATCGCCTTG CAGCACATCCCCCTTTC GCCAGCTGG	DNA containing sgpUC19 sequence to act as trap DNA for roadblock removal assays (-) strand

2.1.5 Bacterial Strains

Table 2.3: *E. coli* strains used in this study.

Strain	Genotype
NEB5α	fhuA2 Δ(argF-lacZ)U169 phoA glnV44 ϕ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17
DH5α	F– ϕ80lacZΔM15 Δ (lacZYA-argF)U169 recA1 endA1 hsdR17(rK–, mK+) phoA supE44 λ– thi-1 gyrA96 relA1
BL21 AI	F– ompT hsdSB (rB–, mB–) gal dcm araB::T7RNAP-tetA

Plasmids were transformed into either NEB5α (NEB) or DH5α (Thermo Fisher). For each plasmid high quality preparations were stored after propagation. Unless otherwise stated, for protein overexpression relevant plasmids were transformed into BL21 AI (Thermo Fisher).

2.1.6 Media and supplements for culturing *E.coli*

All *E. coli* strains were cultured in Luria Bertoni broth (LB, 1% (w/v) tryptone, 0.5% (w/v) yeast extract, 340 mM NaCl, 2 mM NaOH, pH 8.0). Where required, cultures were plated on LB agar (LB, 0.5% (w/v) Agar). All *E. coli* media were sterilised by autoclaving. Both LB and LB agar were stored away from direct sunlight at room temperature until needed.

When appropriate, media was supplemented for antibiotic selection or induction of protein expression. Unless otherwise stated, supplements were dissolved in sterile distilled water (SDW). Chloramphenicol was dissolved in ethanol. Supplements were used at the concentrations indicated in Table 2.4.

Table 2.4: Media supplements used in *E. coli* culture.

Supplement	Final concentration
Ampicillin	100 μg/ml
Kanamycin	50 μg/ml
Chloramphenicol	25 μg/ml
Tetracycline	10 μg/ml
Spectinomycin	50 μg/ml
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	0.5 mM
L-arabinose	0.2 % (w/v)
5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal)	100 μg/ml

2.1.7 Human Cell-lines

All U2OS-derived 5G6 cell-lines were a kind gift from the laboratory of Richard Wood (MD Anderson Cancer Centre, University of Texas, Texas, USA). RKO and derived cell lines were a kind gift from Nanna Therapeutics (Cambridge, UK).

Table 2.5: Human cell-lines

Cell line	Information
U2OS	Human osteosarcoma cell line derived in 1964 from a moderately differentiated sarcoma of the tibia of a 15 year old girl.
5G6	U2OS clone containing homogeneous (-/-) knockout of HelQ
5G6-E	5G6 cell line expressing pEGFP-C1
5G6-Q	5G6 cell line expressing pEGFP-HelQ
5G6-W	5G6 cell line expressing pEGFP-K365M HelQ ATPase-activity deficient mutant
RKO	Poorly differentiated human colon carcinoma cell line developed by Michael Brattain
R-101	RKO clone with homogeneous (-/-) knockout of HelQ
R-172	RKO clone with homogeneous (-/-) knockout of HelQ
R-DA339	RKO clone carrying D463A substitution in HELQ (D463A/D463A)
R-DA93	RKO clone carrying D463A substitution in HELQ (D463A/D463A)

2.1.8 Media and supplements used for human cell culture

U2OS and derived cell lines were maintained in Dulbecco's modified Eagle's medium (DMEM,4.5 g/L glucose, without L-glutamine, Lonza) supplemented with 10% heat inactivated Foetal Bovine Serum (FBS, Sigma-Aldrich), 2 mM L-glutamine, 100 Units (U)/ml of streptomycin and 100 μg/ml of penicillin. The media for cell lines expressing pEGFP-C1-derived plasmids was supplemented with 100 μg/ml G418 for maintenance.

RKO and derived cell lines were maintained in complete Roswell Park Memorial Institute 1640 medium (RPMI 1640, 2.0 g/L glucose, 2 mM L-glutamine, Lonza) supplemented with 10% FBS, 100 U/ml of streptomycin and 100 μg/ml of penicillin.

Table 2.6: Supplements used in human cell culture.

Supplement	Final concentration
Foetal bovine serum (FBS)	10%
L-glutamine	2 mM
Penicillin/streptomycin	100 μg/ml / 100 U/ml
G418	100 μg/ml
Mitomycin C	0–300 μM
Cisplatin	0–100 μM
Aphidicolin*	0–50 μM
Hydroxyurea	0–100mM

2.2 Methods

2.2.1 General DNA and RNA manipulation

2.2.1.1 Polymerase chain reaction (PCR)

DNA was amplified using either Vent (M0254S, NEB) or Q5 polymerase (M0491, NEB). Annealing temperature (Tm) was calculated using an online tool (Tm calculator, NEB, https://tmcalculator.neb.com/). Primer design was carried out using SnapGene software (GSL Biotech LLC, San Diego, California, USA).

Q5 DNA polymerase amplification was carried out in 1× Q5 reaction buffer with 500 nM of each primer (Table 2.2), 200 μM dNTPs, 1-10 ng of template DNA, and 1 unit (U) of enzyme per 50 μl reaction. Here one unit is defined as the amount of enzyme that will incorporate 10 nmol of dNTP into acid insoluble material in 30 minutes at 74 °C. For standard PCR, reactions were denatured at 98 °C for 30 s followed by 30 cycles of: denaturation for 10 s at 98 °C, annealing for 20 s at Tm °C and extension for 30 s/kb at 72 °C, before a final extension step of 72 °C for 3 min.

Vent DNA polymerase amplification was carried out in $1\times$ ThermoPol Reaction buffer with 200 nM of each primer (Table 2.2), 200 μM dNTPs, 1-10 ng of template DNA, and 1 U of enzyme per 50 μl reaction. One unit is defined as the amount of enzyme that will incorporate 10 nmol of dNTP into acid-insoluble material in 30 minutes at 75 °C. For standard PCR, reactions were denatured at 95 °C for 5 minutes followed by 30 cycles of: denaturation for 30 s at 95 °C, annealing for 30 s at Tm °C and extension for 1 min/kb at 72 °C, before a final extension step of 72 °C for 5 min.

Colony PCR was carried out as described using Vent DNA polymerase. In place of template DNA, single bacterial colonies were inoculated into the PCR reaction mix. DNA was released by boiling during the initial heat denaturation step at 95 °C.

PCR products were resolved on Tris-borate-EDTA ($1\times$ TBE, 89 mM Tris, 89 mM boric acid, 2 mM EDTA))-agarose (1% w/v) gels, containing 0.5 μg/ml ethidium bromide, at a voltage of 10 V/cm for 60-90 mins to achieve desired band separation. Product size was determined using either Quick Load® 1 kb (N0552S, NEB) or 100 bp markers (N0467S, NEB). DNA was purified using the using QIAquick® PCR purification kit (28104, Qiagen) or by resolving on a 1% agarose gel followed by QIAquick® gel extraction kit (Qiagen) according to the manufacturer's guidelines.

2.2.1.2 Plasmid construction

To construct recombinant plasmids, purified PCR products and plasmid DNA were digested with appropriate restriction enzymes (NEB) according to the manufacturers' guidance and incubated at 37 °C for 1 – 16 hrs. Digested fragments were purified by resolving on a 1% (w/v) agarose gel and then using the QIAquick® gel extraction kit. The 5 ′ ends of digested vectors were dephosphorylated by the addition of calf intestinal phosphatase (CIP, M0290, NEB) to reactions and subsequent incubation for 1 hr at 37 °C. Vector and

insert DNA were covalently joined together using T4 DNA ligase (M0202, NEB) in 1× ligase reaction buffer and were incubated for either 2 hrs at room temperature or at 16 °C overnight. Ligation reactions were transformed into *E. coli* and DNA was extracted using the QIAprep Spin miniprep kit (27104, Qiagen) following the manufacturer guidelines. Successful plasmid construction was confirmed by restriction digest and subsequent Sanger sequencing.

For site-directed mutagenesis (SDM), PCR was carried out as described in section 2.2.1.1 using Q5 DNA polymerase and mutagenic primers annealed against an appropriate plasmid template. Successful amplification of PCR products was confirmed by visibility of DNA bands in ethidium bromide stained TBE-agarose gels. To remove template DNA and circularise the plasmid, 3μl of PCR product was treated with 1 U each of DpnI (R0176S, NEB), T4 polynucleotide kinase (PNK, M0201S, NEB), and T4 DNA ligase in 1× ligase reaction buffer at room temperature for 1 hr.

One unit of DpnI is defined as the amount of enzyme required to digest 1 μg of DNA (dam methylated) in 1 hour at 37 °CC. One unit of T4 PNK is defined as the amount of enzyme catalysing the incorporation of 1 nmol of acid-insoluble [32P] in 30 minutes at 37 °CC. One unit of T4 DNA ligase is defined as the amount of enzyme required to give 50% ligation of HindIII fragments of λ DNA in 30 minutes at 16 °CC.

Reactions were transformed into *E. coli* and DNA was extracted using the QIAprep Spin miniprep kit. Successful plasmid construction was confirmed by sanger sequencing.

The plasmid pAC22 was constructed using NEBuilder® HiFi DNA assembly master mix (E2621S, NEB) according to the manufacturer's guidelines. The vector used was pACYCDuet-1 (Sigma-Aldrich) and NLS sequences were inserted using two ssDNA oligos (oAC69 and oAC70).

2.2.1.3   Annealing of DNA strands into substrates

Unless otherwise stated, all DNA:DNA, and DNA:RNA substrates used in assays were produced by incubating 5 μM of complementary ssDNA oligonucleotides at 95 °C for 10 minutes before cooling to 37 °C at a rate of 0.5 °C per minute.

DNA:DNA substrates were purified by polyacrylamide gel electrophoresis (PAGE). Annealed substrates were migrated in PAGE gels (1× TBE, 10% (v/v) polyacrylamide (37.5:1)) at 8 V/cm for 120 mins and visualised by UV shadowing if unlabelled or by scanning using an Amersham Typhoon 5 biomolecular imager (GE Healthcare, laser LD635, filter-set Cy5 Fltr 670BP30) if labelled. Substrates were purified, following band excision, by diffusion over 48 hours in 250 μl annealing buffer (4 mM Tris-HCl pH8.0 and 10 mM NaCl) then concentrated to 50 μl using an Eppendorf Concentrator 5301 vacuum concentrator (Eppendorf, Hamburg, Germany). DNA:RNA substrates were stored on ice following annealing and were used in assays immediately due to their instability during PAGE gel purification.

Substrate concentration was determined by spectroscopic measurement in a Nanodrop 2000 (Thermo FIsher). The sample A_{260} was measured and the value obtained applied to the Beer-Lambert law in which A is absorbance, I is light intensity, ϵ is extinction coefficient, c is concentration and l is the length of the light path. The extinction coefficient of substrates was calculated using the online tool Oligoanalyzer 3.1 (Integrated DNA Technologies, https://eu.idtdna.com/calc/analyzer, IDT, Coralville, Iowa, USA).

$$A = \log_{10}(\frac{I_0}{I}) = \epsilon c l$$

2.2.1.4   Ethanol precipitation of DNA and RNA

Ethanol precipitation was used to purify or concentrate DNA. To each sample, 0.1 volume of sodium acetate (3M, pH5.2) and 2.5 volumes of ice-cold 100%

ethanol were added. Samples were incubated at -20 °C for 16 hours and then centrifuged at 20 000 ×g, 4 °C for 30 minutes before decanting the supernatant. Pellets were washed with 70% ethanol and centrifuged for a further 5 minutes before decanting the supernatant and allowing to air dry. Finally, pellets were resuspended in an appropriate volume of Tris-EDTA (TE, 10 mM Tris-HCl, pH 7.5, 1 mM EDTA) solution.

2.2.2   Production of sgRNAs by *in vitro* transcription

All sgRNAs used in this study were produced using the HiScribe™ T7 High Yield RNA Synthesis Kit (E2040S, NEB). The sgRNAs were synthesised using the standard protocol for the production of short oligonucleotides (<300 nt), according to the manufacturer's guidelines. Following synthesis, HiScribe reactions were treated with 1U DNaseI (M0303S, NEB) in 1× reaction buffer at 37 °C for 30 min to remove template DNA. Here one unit of DNaseI is defined as the amount of enzyme which will completely degrade 1 μg of DNA in 10 minutes at 37 °CC.

The synthesised sgRNAs were purified by denaturing-PAGE. HiScribe reaction products were mixed 1:1 with 2x RNA loading dye (B0363S, NEB) and incubated at 95 °C for 5 mins before loading onto a denaturing-PAGE gel (1× TBE, 15% polyacrylamide (19:1, v/v), 5% (v/v) formamide, 7M Urea). Denaturing gels were migrated at 10 W for 180 mins and sgRNA identified by UV shadowing. Bands were excised and soaked in nuclease-free water (AM9930, Thermo Fisher) for 24-48 hours to allow diffusion before concentrating the supernatant containing sgRNA by ethanol precipitation. The final concentration of sgRNA was determined by measuring the sample A_{260} using a Nanodrop 2000 (Thermo Fisher) and applying the obtained value(s) to the Beer-Lambert law.

2.2.3 General Microbiology

2.2.3.1 Growth and storage of *Escherichia coli*

Liquid cultures of *E. coli* were incubated at 37 °C overnight with shaking at 180rpm, unless otherwise stated. For small-scale work such as plasmid propagation, 5 ml culture volumes were used. DNA was extracted using the QIAprep Spin miniprep kit (27104, Qiagen) following the manufacturer guidelines. For large-scale plasmid purification, or where highly pure DNA was required, 500 ml culture volumes were used. DNA was purified using the QIAGEN Plasmid Maxi kit (12162, Qiagen) was used according to the manufacturers' guidance.

For long-term storage, glycerol was added to *E. coli* cultures at final concentration of 20% (v/v) and flash frozen on dry ice. These stocks were then stored at -80 °C. LB agar plates were cultured overnight at 37 °C and stored at 4 °C to minimise colony outgrowth.

2.2.3.2 Preparation of chemically competent *E. coli*

E. coli strains were streaked out from glycerol stocks on LB agar supplemented with antibiotics. Single colonies were selected, inoculated into LB broth and cultured overnight. *E. coli* culture was inoculated 1:100 into fresh LB broth supplemented with antibiotics to maintain selective pressure and cultured to an OD_{600} of 0.6. The culture was centrifuged at 4000 rpm, 4 °C for 10 minutes and the supernatant discarded. Pellets were resuspended in ice-cold calcium chloride ($CaCl_2$, 0.1 mM) and incubated on ice for 2 hours. The culture was centrifuged at 4000 rpm, 4 °C for 10 minutes and resuspended in fresh $CaCl_2$. Sterile glycerol was added to a final concentration of 20% (v/v) and aliquots were flash frozen on dry ice before storing at -80 °C.

2.2.3.3 Transformation of chemically competent *E. coli*

Chemically competent *E. coli* were transformed by heat shock. Briefly, 1-10 ng of plasmid DNA was added to 100 μl of chemically competent *E. coli* and incubated on ice for 30 minutes. The mix was then heat-shocked at 42 °C for 45 seconds before incubation on ice for a further 3 minutes. 900 μl of LB was added and cells were incubated at 37 °C with shaking for 60 minutes to recover. The cells were pelleted by centrifugation at 13 000 rpm for 1 minute and resuspended in 100 μl of fresh LB. The cells were then plated on LB agar, supplemented with selection agents where appropriate, and incubated overnight.

2.2.4 Protein overexpression and purification in *E. coli*

2.2.4.1 General protocol for protein overexpression

Prior to overexpression, *E. coli* strain BL21 AI was transformed with a plasmid containing the protein of interest and plated on LB agar plus selection antibiotic. Single colonies were selected, inoculated into 50 ml LB supplemented with antibiotics and incubated at 37 °C overnight. The following day, liquid culture was inoculated 1:100 into 2-4L fresh LB supplemented with antibiotics and grown to an OD_{600} of 0.6. The media was then further supplemented with IPTG (0.5 mM) and L-arabinose (0.2 %, w/v) and cultured overnight at 18 °C. The following day, cultures were pelleted at 4 000 $\times$g, 4 °C for 10 minutes and resuspended in column equilibration buffer A (composition dependent on the column to be used) supplemented cOmplete™ protease inhibitor cocktail tablet (11836153001, Merck) and/or 1mM phenylmethylsulfonyl fluoride (PMSF). If biomass was not to be used immediately, then instead of resuspending in equilibration buffer, the pelleted biomass was flash frozen and stored at -80 °C for future use.

2.2.4.2 Purification of *Streptococcus pyogenes* Cas9, dCas9, and NLS-Cas9

Cas9 proteins were purified using their tandem affinity tags and based upon the method published by Anders et al.[344]. Briefly, culture resuspended in Ni^{2+} equilibration buffer (20mM Tris-HCl pH8.0, 250mM NaCl, 20mM Imidazole pH8.0, plus protease inhibitor) was lysed using a Vibra Cell VC 50T sonicator (Sonics & Materials, Newtown, Connecticut, USA) and pelleted by centrifugation at 30 000 $\times$g, 4 °C for 30 minutes before incubating on ice with DNaseI (10 μg/ml) for a further 30 minutes.

All protein purification was carried out using an ÄKTA Start system (GE Healthcare, Chicago, Illinois, USA). The clarified lysate was loaded onto an equilibrated HiTrap Chelating HP column (GE Healthcare) charged with Ni^{2+} before eluting in Ni^{2+} buffer B (20mM Tris-HCl pH8.0, 250mM NaCl, 500mM Imidazole pH8.0) across a gradient. Fractions corresponding to a UV absorbance peak were analysed using SDS-PAGE, with size being confirmed using Blue Prestained Protein Standard, Broad Range (P7706, NEB). Fractions containing the desired protein were pooled and dialysed overnight at 4 °C in Heparin Buffer A (20mM HEPES-KOH pH7.5, 150 mM KCl, 10% (v/v) glycerol and 1mM Dithiothreitol (DTT)) using dialysis tubing with an 8 kDa molecular weight cut-off (MWCO). The dialysed sample was loaded onto a HiTrap Heparin HP column (GE Healthcare) and eluted in Heparin buffer B (20mM HEPES-KOH pH7.5, 1 mM KCl, 10% (v/v) glycerol and 1mM Dithiothreitol (DTT)) across a gradient. Samples were again analysed by SDS-PAGE before loading protein containing fractions onto a HiPrep 16/60 Sephacryl S-300 HR (GE Healthcare) column equilibrated with SEC Buffer (20 mM HEPES-KOH pH7.5, 500mM KCl, 10% glycerol (v/v) and 1mM DTT). Purified His-MBP-Cas9 or His-MBP-dCas9 were concentrated using a centrifugal concentrator (Pierce, 100kDa MWCO). Protein concentration was quantified using a combination of the Bradford assay (see section 2.2.4.4) and

measurement of A_{280} using a Nanodrop 2000 (Thermo Fisher). The purified protein was aliquoted, flash frozen and stored at -80 °C.

2.2.4.3 Purification of *Acidaminococcus sp.* Cas12a

AsCas12a was purified using a method adapted from Anders et al.[344]. Briefly, culture resuspended in Ni^{2+} lysis buffer A was sonicated using a Vibra Cell VC 50T sonicator (Sonics & Materials) and pelleted by centrifugation at 30 000 $\times$g, 4 °C for 30 minutes before incubating on ice with DNaseI (10 μg/ml) for a further 30 minutes. The clarified lysate was loaded onto an equilibrated HiTrap Chelating HP column (GE Healthcare) charged with Ni^{2+} before eluting in Ni^{2+}buffer B across a gradient. Fractions corresponding to a UV absorbance peak were analysed by SDS-PAGE to confirm the presence of the desired protein before being pooled and dialysed overnight at 4 °C in Heparin Buffer A. The sample was then loaded onto a HiTrap Heparin HP column(GE Healthcare) and eluted in Heparin buffer B. Samples were again analysed using SDS-PAGE before loading onto a HiPrep 16/60 Sephacryl S-200 HR (GE Healthcare) equilibrated with SEC Buffer (20 mM HEPES-KOH pH7.5, 500mM KCl, 30% glycerol (v/v) and 1mM DTT). Purified His-NLS-Cas9 / His-NLS-Cas12a was concentrated using a centrifugal concentrator (Pierce, 100kDa MWCO). The concentration of protein was measured using the Bradford assay (see section 2.2.4.4) from the A_{280} using a Nanodrop 2000 (Thermo Fisher) in conjunction with the Beer-Lambert law. Purified protein was then aliquoted, flash frozen and stored at -80 °C.

2.2.4.4 Bradford Assay

Protein concentrations were measured by the Bradford assay using Bradford Reagent (B6916, Sigma-Aldrich) according to the manufacturer's guidance. Briefly, protein standards at concentrations from 0.1 – 1.4 mg/ml were prepared using bovine serum albumin (BSA). Where a high total protein

concentration was suspected, samples were measured at dilutions of 1:5 and 1:10. Both the standard curve and sample solutions were mixed with Bradford reagent as instructed and incubated at room temperature for 20 mins. Sample absorbance was measured at 595 nm using a DeNovix DS-11 FX spectrophotometer (Cambridge Biosciences, Cambridge, UK). A standard curve was generated using Prism software (Graphpad, San Diego, California, USA), from which sample concentrations were interpolated. All samples were measure in duplicate.

2.2.4.5 Pierce bicinchoninic acid (BCA) Protein Assay

When the Bradford assay was incompatible with buffer components, the Pierce^{TM} bicinchoninic acid (BCA, Pierce Biotechnology, Rockford, Illinois, USA) assay was used to determine protein concentration, according to the manufacturer's guidelines. In brief, protein standards were generated using BSA at concentrations from 0.25 – 1.4 mg/ml. Samples and standards were combined with BCA working reagent and then incubated at 37 °C for 30 minutes. Where a high total protein concentration was suspected, samples were measured at dilutions of 1:5 and 1:10. Assays were carried out in a 96-well microplate. Absorbance was measured at 562 nm using a SPECTROstar Nano microplate reader (BMG labtech, Ortenberg, Germany). A standard curve was generated, using Prism software, and sample concentrations were interpolated from it. All samples were measure in duplicate.

2.2.5 Human Cell Culture

2.2.5.1 Routine cell culture

All cell lines listed in Table 2.5 were cultured at 37°C with 5% CO_2. Media for the culture of U2OS-derived cell lines stably expressing EGFP, EGFP-HelQ, or EGFP-HelQ K365M from a non-integrated pEGFP-C1 plasmid were supplemented with 100μg/ml of G418 to maintain selection.

When seeding a new flask of cells, cryovials were retrieved from storage in vapour phase liquid nitrogen and rapidly thawed in a 37°C water bath. Cells were transferred to a fresh 50 ml tube (Sartsedt) and 20 ml pre-warmed media was slowly added to dilute DMSO in the freezing medium. Cells were pelleted by centrifugation at 300 ×g for 5 minutes. The supernatant was aspirated, and cells were resuspended in 5 ml media. Cells were transferred to a fresh 25 cm^2 (Sarstedt) flask and incubated at 37°C with 5% CO_2. Growth was monitored until cells reached 80% confluency, wherein they were transferred to a new 75 cm^2 flask (Sarstedt).

Cell lines were passaged at 60-80% confluency by washing twice with phosphate buffered saline (PBS, no calcium or magnesium, Sigma-Aldrich) and incubation with 1× trypsin/EDTA solution (Sigma-Aldrich) at 37°C until detachment from the flask was visible. Trypsin was neutralised by the addition of media. Cell viability and number was assessed by mixing a small sample 1:1 with 0.4% Trypan blue (w/v, 15250061, Thermo FIsher)and counting using a haemocytometer (Marienfeld, Lauda-Königshofen, Germany). The desired number of cells were seeded to a new 75 cm^2 flask (Sarstedt) and incubated at 37°C with 5% CO_2. Cell lines were kept for 10-15 passages before a new culture was thawed out from a frozen aliquot.

When preserving cell lines, cells were seeded into 100 mm dishes and grown to 80% confluency. Cells were then processed as described above. Once counted, cells were resuspended in freezing medium (80% complete medium, 10% FBS, 10% DMSO) to a density of 2 ×10^6 cells/ml and aliquoted into cryotubes. Tubes were frozen slowly over 24 hours at -80°C before placing in the vapour phase of liquid nitrogen for long-term storage.

2.2.5.2 Cell proliferation and cytotoxicity assays

Assays to determine cell proliferation or cytotoxicity were carried out using WST-1 cell proliferation reagent (Sigma). Assays were carried out according

to the manufacturer's guidelines. In brief, cells were seeded at the required number in 96-well plates. Additives for the assay, such as those in Table 2.6, were supplemented into complete media at the indicated concentrations. Cells were seeded in 100 μl total volume and incubated as described above. Following the incubation period, 10 μl of WST-1 reagent was added to each well and microplates were incubated at 37 °C, 5 % CO_2 for 3 hours. Absorbance was read at 440 nm using a SPECTROstar Nano microplate reader (BMG Labtech). Background absorbance was corrected against a media-only blank for each assay condition.

2.2.5.3 Preparation of cell-free extracts

Cell-free extracts were prepared according to the protocol stated in Sansbury et al.[342]. Briefly, 8 $\times 10^6$ cells were harvested from 75 cm^2 flasks or 100 mm dishes, following two washes with ice-cold PBS, by scraping. Cells were pelleted at 300 $\times$g for 10 minutes and then washed with ice-cold hypotonic buffer (20 mM HEPES pH7.5, 5 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 250 mM sucrose). The cells were resuspended in the same buffer minus sucrose and left to resuspend on ice for 15 minutes with occasional agitation. Cells were lysed by 25 strokes of a Dounce homogeniser and then incubated for a further 60 minutes, all on ice. The crude lysate was pelleted at 12 000 $\times$g, 4 °C for 15 minutes before aliquoting and freezing at -80 °C for later use.

2.2.6 Immunocytochemistry for the detection of R-loops in human cell lines

Immunocytochemistry to detect DNA:RNA hybrids was performed as described in Abakir et al.[345] and Hamperl et al.[346]. U2OS and derived cell-lines were cultured as described in 2.2.5.1 and then seeded into separate wells in 8-well chamber slides (Sarstedt) at a density of 1.3 $\times 10^4$ and cultured for a further 48 hours. Cells were fixed to the slide surface by incubation with ice-cold 4% (v/v) paraformaldehyde (PFA) for 15 minutes at room tempera-

ture. Slides were washed twice with PBS for 5 minutes at room temperature to remove excess PFA. Fixed cells were permeabilised by incubating slides in PBT buffer (1× PBS, 0.5% (v/v) Triton X-100) for 15 min. Slides were then blocked in a solution of PBS supplemented with 10% (w/v) BSA for one hour in a humidity chamber at room temperature. The slides were washed three times with 1× PBS before being incubated with S9.6 antibody (mouse, monoclonal, 1:200 dilution, Sigma-Aldrich, MABE1095) at 4 °C overnight. The following day, slides were washed three times with PBS for 5 minutes each before incubating with anti-mouse IgG (goat, polyclonal, Alexafluor-633 conjugated, 1:400 dilution, Invitrogen, A-21052) for 1 hr at room temperature in a humidity chamber. Slides were washed three times for 5 minutes each with 1 × PBS before incubation with DAPI stain (Boster, Pleasanton, California, USA) for 10 minutes at room temperature. The slides were washed again three times for 5 minutes each with 1 × PBS. Slides were mounted in 1 × PBS and sealed for imaging. Cell images were acquired with a Zeiss LSM 710 AxioObserver confocal microscope (Zeiss,Oberkochen, Germany) using a PlanApochromat ×63/1.40 numerical aperture Oil DIC M27 objective. S9.6 intensity per nucleus was calculated using ImageJ/Fiji (National Institutes of Health, NIH, Bethesda, Maryland, USA) and Adobe Photoshop (Adobe, San Jose, California, USA), where DAPI is used as a mask for the nucleus.

2.2.7 Quantification of S9.6 signal

The signal intensity for S9.6 was quantified using the method of Rajani et al. [347]. Briefly, Fiji software[348] was set to record measurements for area, standard deviation, min & max gray value, mean gray value, and median. Images obtained by confocal microscopy (See Section 2.2.6) were then imported and the channels split to produce separate DAPI and S9.6-stained images. Masks were created for the nuclei using the DAPI channel for each image. This was achieved by applying Gaussian blur to each image before

using the thresholding tool to create a mask of each nucleus. Nuclei not completely in the field of view and those overlapping or irregular in shape were excluded. Masks were then transferred to the corresponding S9.6 image and used as the boundaries within which to measure the aforementioned values. The data obtained were then transferred to Prism software for further analysis.

2.2.8 Biochemical Assays

2.2.8.1 Cas-protein nuclease activity assays

To determine Cas-protein activity, purified protein was incubated with sgRNA for 10 minutes at room temperature in $1\times$ cleavage buffer (20 mM HEPES-KOH pH 7.5, 100 mM KCl, 5% glycerol (v/v), 1 mM DTT, 0.5 mM EDTA and 2 mM MgCl2) to allow RNP complexes to form. Complexes were then incubated with 25 nM dsAC49 and 75 nM competitor DNA (MW14 dsDNA). Reactions were incubated at 37 °C for 15 minutes before quenching with proteinase K stop buffer (2% (w/v) SDS, 200 mM EDTA, 2mg/ml Proteinase K). Assays requiring an absence of nuclease activity of Cas9 omitted $MgCl_2$ from reaction buffer. Reactions were electrophoresed through an 8% (v/v) polyacrylamide, native PAGE gel at 120 V for 2 hours. The labelled DNA substrate was imaged using an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30). Band intensity was quantified, using ImageJ software and the percentage of substrate cleaved by the nuclease calculated against a control reaction that contained no Cas-protein.

2.2.8.2 R-loop formation assays

R-loop formation was assessed by targeting Cas9 or Cas12a RNPs to (insert the oligo name here). 150 nM Cas-protein was incubated with 150 nM sgRNA in $1\times$ binding buffer (25 mM Tris pH 7.5, 106 mM KCl, 50 μg/ml BSA, 5 mM EDTA, 2 mM DTT, 9% (v/v) glycerol) for 10 minutes at room

temperature to allow RNP formation. Complexes were then incubated with 25 nM dsAC49 and 75 nM competitor DNA (MW14 dsDNA) for 20 minutes at 37°C. The reaction was stopped by the addition of proteinase K stop buffer and migrated on an 8% (v/v) polyacrylamide, native PAGE gel at 120 V for 2 hours. The gel was imaged using an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30). The formation of R-loops was validated by the addition of two controls: one containing no protein and another supplemented with RNaseH (NEB, M0297S).

2.2.8.3   Helicase unwinding assays

Helicase unwinding assays were carried out using 12.5 nM 5′ Cy5 labelled forked DNA substrate in $1\times$ unwinding buffer (20 mM Tris-HCl pH 7.5, 5% glycerol (v/v), 100 μg/ml BSA) supplemented with 5 mM $MgCl_2$, 5 mM ATP and 25 mM DTT. Helicases were added to the desired final concentration(s). Reactions proceeded at 37°C for 30 minutes before adding stop buffer. Substrate re-annealing was prevented by the addition of unlabelled competitor DNA, identical to the fluorescently labelled strand, at 10-fold excess. The reactions were visualised by migration on a 10% polyacrylamide TBE gel and migrated at 140 V for 40 minutes. The labelled DNA substrate was then imaged using an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30). Unwinding was quantified in ImageJ by analysing band-intensity relative to either a no-protein dsDNA control lane, or a boiled sample simulating fully unwound substrate.

For helicase-inhibition assays, small-molecule inhibitors were added to desired final concentrations and control assays were supplemented to contain the equivalent final concentration of DMSO.

The initial small-molecule screen for inhibitors of HelQ was visualised by running assays through a 2.5% (w/v) agarose gel using sodium-borate buffer (10 mM NaOH, 39 mM boric acid). Gels were migrated at 200 V for 10

minutes. Assay products were visualised using SYBR-Gold stain (Thermo Fisher).

2.2.8.4 Electrophoretic Mobility Shift Assays (EMSAs)

The association of proteins with DNA to form nucleoprotein complexes was observed using EMSAs. Reactions were carried out in a $1\times$ buffer suitable for the protein of interest. Where required, protein activity was prevented by removing $MgCl_2$ from the reaction buffer. Reactions were carried out at 37°C for 10 minutes to allow complexes to form. Reactions were then migrated through a 5% (v/v) polyacrylamide native PAGE gel at 120 V for 180 minutes to allow complex separation from unbound DNA substrate. The labelled DNA was imaged using an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30). Band intensity was measured using ImageJ and the percentage of substrate in complex with the protein of interest was calculated against a control lane containing no protein.

To determine the impact of putative HelQ / C-HelQ inhibitors on DNA binding, candidate compounds were added to reactions at the desired final concentration (Maximum 5 mM) and control assays were supplemented to contain the equivalent final percentage of DMSO.

2.2.8.5 ATPase assays

ATPase activity for proteins was measured using the Malachite Green Phosphate Assay Kit (Sigma), according to the manufacturer's guidelines. In brief, standards were prepared at concentrations ranging from $0 - 40$ μM phosphate. Samples were prepared by diluting helicase unwinding reactions, described in section 2.2.8.3, to an ATP concentration that was <0.25 mM. Samples and standards were then mixed with Malachite Green Assay working reagent in a 96-well microplate and incubated at room temperature for 30 minutes. Absorbance was measured at 620 nm using a SPECTROstar

Nano microplate reader. A phosphate standard curve was then generated, using Prism software, from which sample concentrations were interpolated. All assays were carried out in triplicate. To detect possible phosphate contamination, control wells were included for each reagent used in the assay.

2.2.8.6 Roadblock removal assays

To assess if HelQ could remove a Cas9 roadblock from dsDNA, a hybrid unwinding assay-EMSA was used. Reactions were carried out in $1\times$ unwinding buffer supplemented with 5 mM $MgCl_2$,5 mM ATP, 25 mM DTT and 250 nM MW12 trap ssDNA, or in $1\times$ NEBuffer 3.1 (NEB, B7203S) supplemented with 5 mM ATP, 25 mM DTT, and 250 nM MW12 trap ssDNA. dCas9 was incubated with an sgRNA capable of targeting MW14 fork DNA at room temperature for 10 minutes to form RNP complexes. These were then incubated with 25 nM MW14 fork to allow the formation of dCas9-mediated R-loops. To this reaction, HelQ was added to a final concentration of 20-240 nM. The assays proceeded at $37^{\circ}C$ for 30 minutes. Reactions were then electrophoresed through an 8% polyacrylamide native PAGE gel at 120 V for 180 minutes. The labelled DNA was imaged using an Amersham Typhoon 5 biomolecular imager (laser LD635, filter-set Cy5 Fltr 670BP30). The assays were compared to controls lanes containing DNA only, Cas9 only, Cas9 only minus $MgCl_2$ and HelQ only.

Plasmid-based assays were also used to determine the ability of HelQ to remove Cas-protein roadblocks. Reactions were carried out using a method outlined in Killelea et al.[75]. Briefly, reactions were carried out in $1\times$ unwinding buffer supplemented with 5 mM $MgCl_2$,5 mM ATP, and 25 mM DTT. dCas9 was incubated with an sgRNA capable of targeting pUC19 plasmid DNA at room temperature for 10 minutes to form RNP complexes. These were then incubated with 100 ng of pUC19 to allow the formation of dCas9-mediated R-loops. To this reaction, HelQ was added to a final concentration

of 80-320 nM. The assays proceeded at 37°C for 30 minutes. After 15 minutes, the assays were supplemented with 250 nM of appropriate trap ssDNA to prevent dCas9 re-annealing to pUC19. Reactions were then electrophoresed through an 0.8% agarose gel in 1× TAE buffer at 15 V for 16 hours.

2.2.8.7   ssODN integration assays

Assays to integrate donor DNA into pUC19 were carried out following the protocol published by Sansbury et al.[342], Cole-Strauss[349]. Cell-free extracts were produced as described in Section 2.2.5.3. Cas-protein RNPs were formed by incubating either Cas9 or Cas12a with an appropriate sgRNA in 1× NEBuffer 3.1 at room temperature for 10 minutes. RNPs were then added to 250 ng pUC19 plasmid DNA, also in 1× NEBuffer 3.1, and incubated at 37 °C for 15 minutes to allow cleavage. DNA was subsequently purified using QIAquick® gel extraction columns. Recovered DNA was concentrated by ethanol precipitation.

In vitro integration assays contained DNA recovered from cleavage reactions, 20 μg of CFE, and 400 cohesive end units of T4 DNA ligase in a 1× reaction buffer (20mM Tris, 15mM MgCl2, 0.4mM DTT, and 1.0mM ATP). Where appropriate, reactions also contained 4.464 μg of single- or double-stranded donor DNA. Reactions were incubated at 37 °C for 15 minutes before being recovered using QIAquick® gel extraction columns.

Recovered DNA was transformed into 50 μl chemically competent DH5α via heat shock as described in Section 2.2.3.3. Cells were plated on selective media supplemented with IPTG and X-gal at the concentrations described in Table 2.4 and incubated at 37 °C overnight. The next day, single white colonies of interest were selected for colony PCR. Successful amplicons were then digested by supplementing reactions with 2U NotI-HF (NEB, R3189S). Digestion reactions were incubated at 37 °C for two hours before products were resolved on a 2% agarose gel at 120 V for 90 minutes.
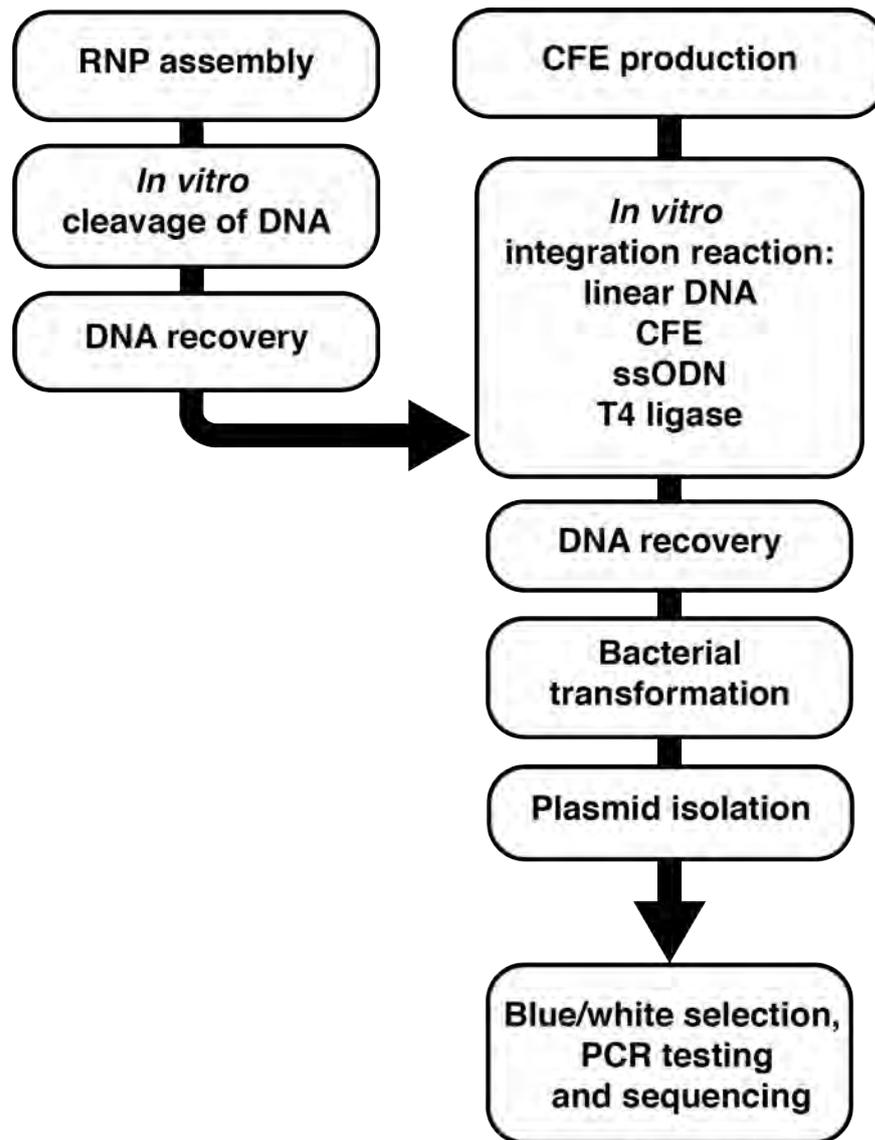
Figure 2.1: Workflow diagram for ssODN integration assays. RNPs were formed by incubating recombinant Cas9 or Cas12a with appropriate sgRNA. RNPS were then used to cleave plasmid DNA, which was recovered and used in integration assays. Separately produced CFEs were supplemented into integration assays along with template DNA and T4 DNA ligase. Recovered DNA was then transformed in *E. coli* and white colonies screened by colony PCR.

2.2.9   Calculation of IC50 values

IC50 values were calculated using GraphPad Prism. The model of analysis chosen was the four-parameter logistic curve, a regression model often used to analyse bioassays. This is because such assays typically follow a sigmoidal curve, only being linear across a specific range of concentrations before

plateauing at either end towards minimum and maximum values. When using a four-parameter logistic curve, a and d are equal to the maximum and minimum asymptotes, c is the IC50 and b is the Hill's slope of the curve.

$$y = d + \frac{a - d}{1 + 10^{(c-x) \times b}}$$

The parameters can be further defined as such: the minimum value is the lowest possible response usually corresponding to a control assay which is not supplemented with inhibitor, the maximum value is the largest observed response usually corresponding to the highest dosage, the IC50 is the point half-way between the minimum and maximum values, and the Hill's slope is the gradient of the curve at the point of the IC50. The four-parameter logistic curve assumes that a fit will be symmetrical meaning that the IC50 will typically fall within the linear range of the cure, as such the value for the Hill's slope is a fixed value. For inhibition assays this value is -1.0.

2.2.10 Dynamic light scattering

Dynamic light scattering (DLS) was used to analyse the aggregation properties of candidate small-molecule inhibitors. Hydrodynamic radii were measure using a Zetasizer Ultra (Malvern Panalytical Ltd., Malvern, UK). Particles were dispersed in 100% DMSO. Sample measurements were taken at a fixed scattering angle of 173 rad at 25°C after a 3 min incubation in the DLS instrument. Intensity-weighted size distributions were converted into number-weighted size distributions to eliminate the influence of small particle population, assuming particle sphericity and the accuracy of optical constants. The number-weighted size distributions for each sample were averaged over 5 consecutive measurements.

2.2.11 Statistical tests

2.2.11.1 Student's two-tailed t-test

For quantification of S9.6 signal intensities (Section 3.2.7) and of ATPase, EMSA, and helicase unwinding assays (Chapter 5.3), the statistical significance between signal intensities under different conditions was determined using unpaired two-tailed Student's t-tests. All data were plotted and analyzed in GraphPad Prism v7.0 (Graphpad, San Diego, California, USA). Outliers from measurements of cell-based fluorescence imaging were removed using the GraphPad ROUT method.

3

HelQ in CRISPR-mediated gene-editing and resolution of DNA:RNA hybrids

3.1   Introduction

3.1.1   HelQ in DNA repair

HelQ is a ssDNA-dependent 3′–5′ helicase exhibiting a strong preference for unwinding substrates resembling stalled replication forks that possess a nascent lagging strand[225,226]. HelQ recruitment at these regions involves interaction with the RPA complex, which the protein displaces from ssDNA, via the subunit RPA70[225–228]. This is thought to occur via a predicted PWI domain in HelQ, which may trigger remodelling of the RPA complex and subsequently release it from DNA[229].

HelQ has been implicated in HR through interaction with RPA, and the Rad51 paralogue complexes BCDX2 and CX3, participating in the disruption of Rad51 binding to dsDNA[227,228]. The combined loss of these proteins was shown to block the progression of HR, likely due to a lack of Rad51 filament disassembly from stalled forks[246]. Recent work has suggested that this may be indicative of participation in MMEJ and/or SDSA[229]. HelQ has also been associated with HR through association with key proteins such as the replication checkpoint kinase ATR, the histone demethylase JMJD5, and the HROB-Mcm8-Mcm9 helicase complex[228,247,248].

Multiple studies have associated HelQ with the FA pathway and the repair of ICLs, observing that HelQ-deficient cells display increased sensitivity to crosslinking agents such as MMC[226–228]. The same studies also demonstrated co-localisation between HelQ and known marker of stalled replication forks, FANCD2[226–228].

While the precise purpose of HelQ remains ambiguous, the recruitment of the protein to replication forks and its promotion of, or participation in, several DNA repair pathways, including HR and FA, suggests that the protein functions as a mediator of replication-coupled repair[226–229].

3.1.2 A role for HelQ in successful gene-editing

The use of CRISPR-associated (Cas) nuclease proteins for knock-in/knock-out gene editing results in the generation of DSBs. This can be used to harness NHEJ to create InDel mutations at the cut site, disrupting expression of a protein of interest. It can also be used for gene replacement, in which exogenous donor DNA sequences are integrated into the host genome. This can facilitate the correction of aberrant gene expression, or the insertion of new genetic material[350].

The efficient replacement of genes has been long sought-after for the treatment of genetic diseases. Despite a decade of rapid advancement in genome-editing technologies, our knowledge of the repair processes involved in Cas-mediated integration remains poor. One promising method for gene replacement is repair using single-stranded oligodeoxyribonucleotides (ssODNs) which are integrated into the host genome via the single-stranded template repair (SSTR) pathway[351,352].

Previous research has demonstrated that Cas9-induced editing using ssODNs utilises the FA repair pathway[212]. This paper identified a possible role for HelQ in gene-editing, demonstrating using CRISPR-interference (CRISPRi) that knockdown of the protein, as well as multiple FA complex

proteins, substantially decreased SSTR efficiency. This included FANCD2, with which HelQ is known to interact[227]. The study also identified that the Rad51C subunit, but not the BCDX2 complex with which HelQ is known to interact, is required for SSTR[212].

3.1.3 R-loops as a source of genome instability

R-loops are triple-stranded DNA:RNA hybrid structures formed when RNA invades and displaces a strand from dsDNA (Fig. 3.1A)[353]. These are widespread throughout cells, participating in many natural processes including transcription, where they form at the site of RNA polymerase activity, and replication as primers for DNA synthesis. While these activities have been identified, information regarding the precise role of R-loops is often conflicting and their significance in many contexts remains unknown.

Aberrant R-loops formation can generate a plethora of DNA lesions, which act as a source of genome instability. During R-loop formation, a strand of ssDNA is exposed (Fig. 3.1B) and can be targeted by DNA deaminases, resulting in mutagenesis by C>U conversion[354]. It can also lead to the generation of SSBs and DSBs, by endonucleases such as XPG, XPF and FEN1[355]. R-loop associated ssDNA has been observed to form complex secondary structures including G-quadruplexes, which can act as obstacles to replication[356].

R-loops can also cause transcription stress due to stalling of the RNA polymerase machinery (Fig. 3.1B)[357]. This typically results from encountering an existing roadblock or lesion and has the potential to cause strand breaks, which require repair by the DNA-damage response, or results in polymerase back-tracking[358]. The latter is especially dangerous when coupled with an approaching replication fork as it can lead to transcription-replication collisions (TRC).

As both the transcription and replication machinery require access to

the same template DNA during S-phase, TRCs which require repair are inevitable. This is thought to occur through two different mechanisms: Co-directional collision, in which the replication and transcription machinery are progressing in the same direction, and head-on collision in which they are travelling in the opposite direction (Fig. 3.1B)[346,359] It remains unknown whether the formation of R-loops are a cause or consequence of TRCs, but failure to resolve these structures can result in fork collapse, resulting in genomic rearrangements and/or chromosomal breakage[358].

3.1.4   Helicases in R-loop resolution

Multiple helicases have been shown to associate with R-loops, likely to participate in processes which maintain genome stability. The RecQ-like helicase BLM has been shown to efficiently unwind R-loops *in vitro* and to co-localise with the structures, likely functioning to suppress hybrid-associated genome stability[360,361]. Research has also found that the helicase DHX9 is required along with the splicing factors SF3B3 and SFPQ for R-loop formation as the absence of all three protein together suppressed the structure entirely[362]. Alongside this, Pif1 has been suggested to act as an accessory protein alongside Rrm3 in dissociating the pre-initiation transcription complex and removing R-loops at tRNA genes to prevent replication fork stalling[363].

Of particular interest is the reported role of PolQ in interactions with R-loops. The polymerase-helicase has been shown to efficiently unwind DNA:RNA hybrids, exhibiting a preference for unwinding the lagging strand at replication forks, much like HelQ[364]. Subsequent research has also show that PolQ takes part in RNA-templated DNA repair[365]. Given the reported similarity between the helicase domain of PolQ and HelQ it remains to be seen whether the protein is capable of the same functions.

A.

Aberrant R-loop

B.

Replication stress

Transcription stress

C   **H**

ssDNA damage

DSB

Genome Instability
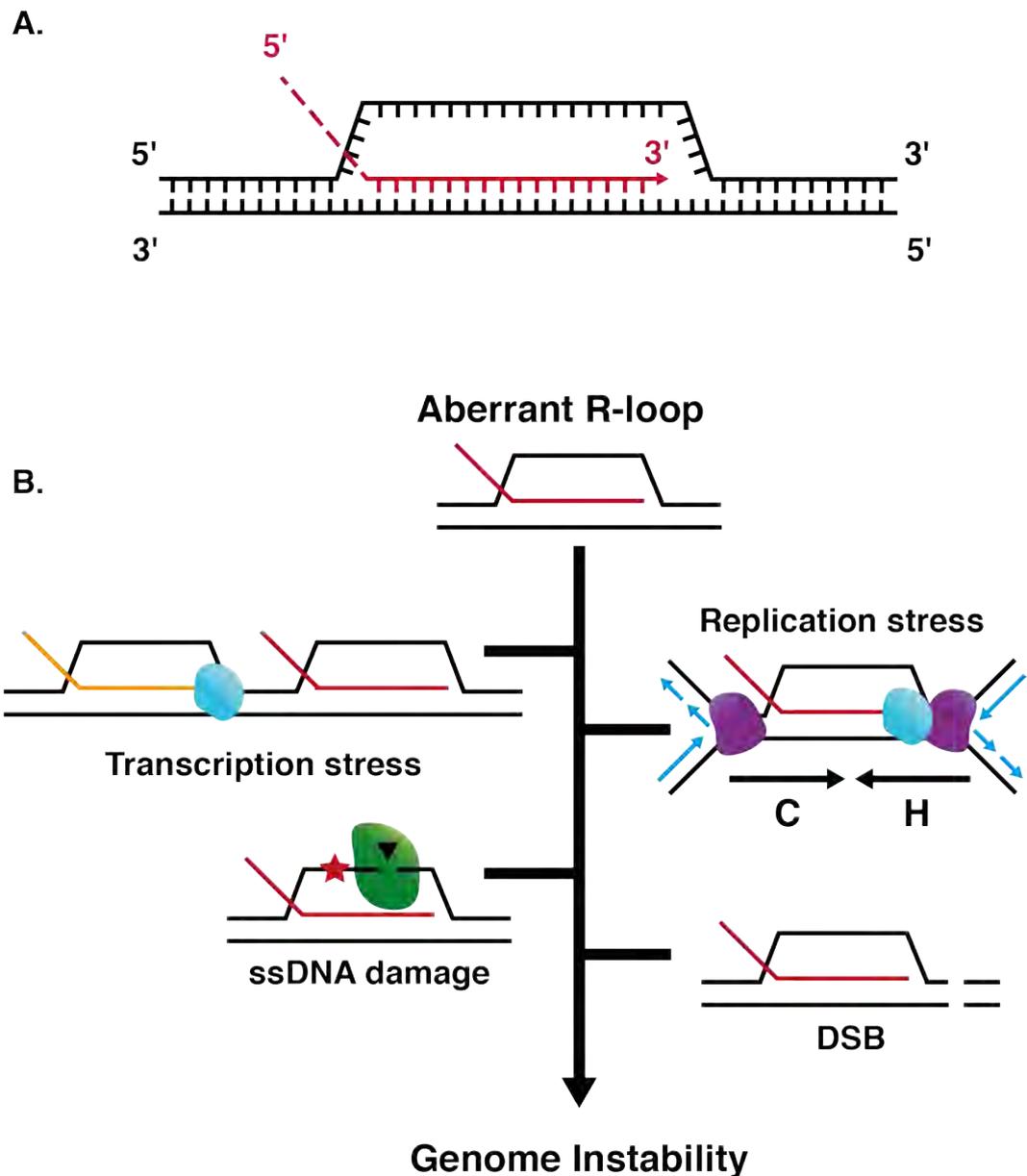
Figure 3.1: The role of R-loops in generating DNA damage and genome instability. (A) Schematic of an R-loop. The structure comprises a strand of RNA which is synthesised on or has invaded dsDNA. These can form as part of transcription initiation and termination, mitochondrial DNA replication, and epigenetic modifications. (B) The aberrant formation of R-loops can lead to genome instability. Exposed ssDNA can be cleaved by nucleases (green) or can undergo spontaneous mutations (red star). They can also cause replication stress, wherein the transcription machinery (blue) causes the replication complex (purple) to stall by Co-directional (C) or Head-on (H) collision, or transcription stress when the synthesis of new RNA (orange) is stalled by an existing R-loop. Finally, R-loops can result in the generation of double-strand breaks (DSBs).

3.1.5   Aims and Objectives

The role of HelQ in efficient integration of template DNA during gene-editing reported by Richardson et al.[212] gives us early quantitative insight into its importance, but does not indicate the mechanism by which the protein functions to stimulate repair. This chapter focusses on identifying a role for HelQ during gene-editing using *in vitro* biochemical approaches, including reconstituting a cell-free system as a model for repair, before further exploring a more general role for the protein in resolving R-loops. The objectives of this work were:

- To assess the ability of HelQ to remove Cas-protein roadblocks from DNA as a possible role in gene-editing.

- To reconstitute a cell-free system for modelling template during gene-editing and use it to model the impact of HelQ-deficiency.

- To assess whether HelQ has a role in resolving R-loops using both biochemical and cell-based models.

3.2   Results

3.2.1   Purification of CRISPR-Cas proteins

3.2.1.1   Purification of *Streptococcus pyogenes* Cas9 and dCas9

To purify Cas9 and dCas9, the relevant plasmids (see Table 2.1) were transformed and overexpressed in *E. coli*. All proteins were purified using a method adapted from Anders et al.[344] in which an ion-exchange purification step was substituted for heparin, which acts as both an affinity ligand and a cation-exchanger[366].

Both Cas9 and dCas9 possessed an N-terminal His-tag and following lysis were purified using a Ni^{2+} chelating column. This was insufficient to fully

purify the Cas-proteins, with a number of contaminant bands remaining present (Fig. 3.2B, 3.3A). Following a dialysis step to remove imidazole and reduce the salt concentration in the buffer, proteins were further separated through a heparin column. This also concentrated the sample, highlighting remaining impurities (Fig. 3.2C). The final step, SEC, appears to have removed many of the previous contaminants while keeping the sample concentrated across approximately 10 fractions in the case of Cas9 (Fig. 3.2D, A3) and a single peak constituting approximately 10 fractions in the case of dCas9 (Fig. 3.3B, A5).The pooled protein was further concentrated using a centrifugal filter with a MWCO of 100 kDa, which also removed smaller remaining contaminant bands. Each purification was successful in producing pure recombinant protein which ran at the expected sizes of 201.7 kDa for both proteins (Fig. 3.2E, 3.3C).

3.2.1.2 Purification of *acidaminococcus spp.* Cas12a

The purification of Cas12a utilised the same adapted method as the Cas9 proteins. This proceeded similarly to the Cas9 variants, with a single nickel affinity step proving insufficient (Fig. 3.4A). During the heparin-affinity step, the purity of eluted Cas12a appeared to be much higher than that of His-MBP-Cas9, but this could be due to a lower sample concentration, resulting in contaminant bands being less visible when observed by SDS-PAGE (Fig. 3.4B). The final stages of the purification, a SEC column (Fig. 3.4C) followed by centrifugal concentration with a 100 kDA MWCO filter, resulted in the production of pure recombinant protein which ran at the expected size of 199.6 kDa, as observed by SDS-PAGE (Fig. 3.4D).

3.2.2 Testing the activity of recombinant Cas-proteins

Purified Cas proteins were tested for activity using three different approaches: nuclease cleavage assays, DNA binding assays, and R-loop formation assays.
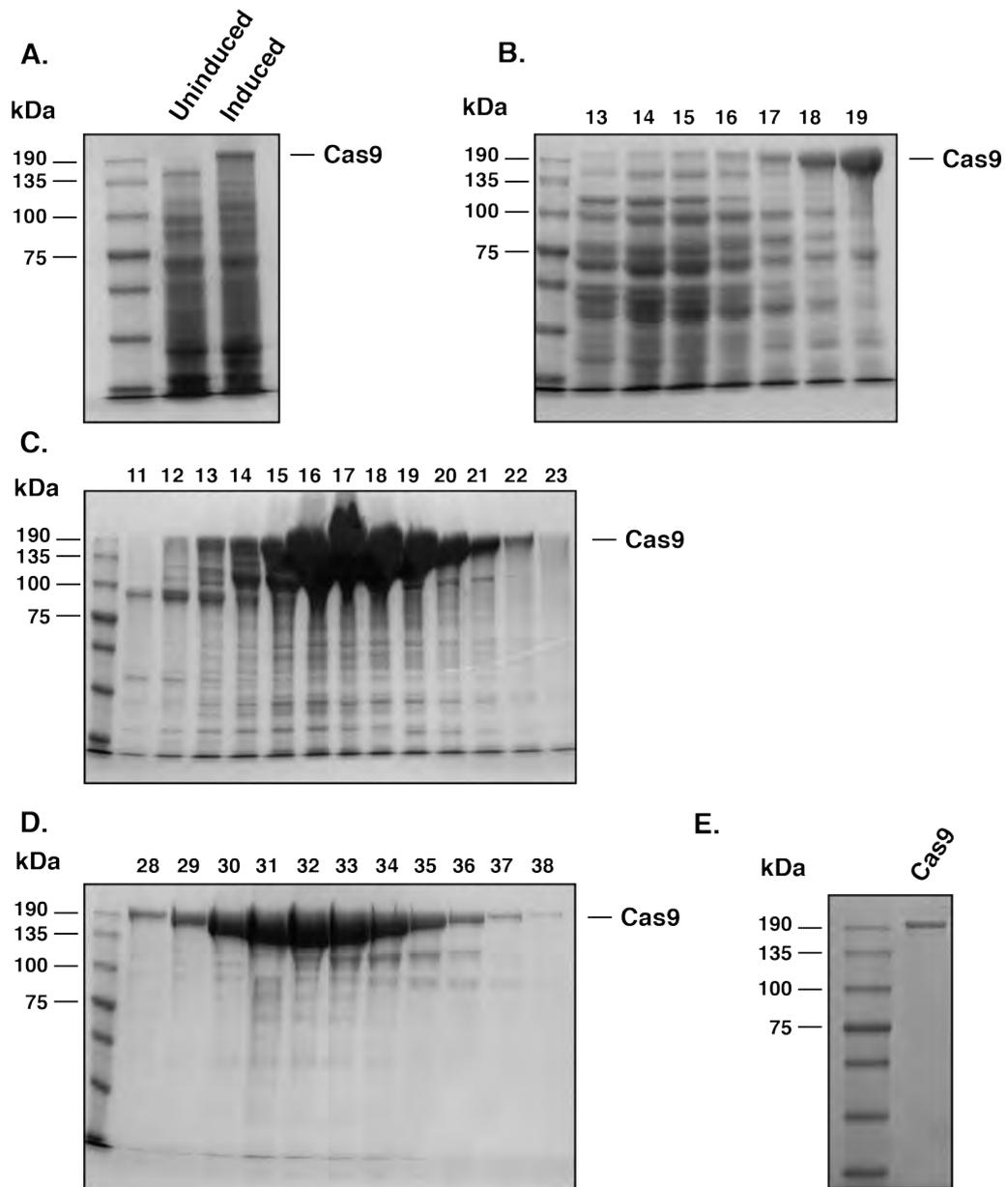
Figure 3.2: Purification of His6-MBP-Cas9. Cas9-proteins were overexpressed in *E. coli*. (A) Protein overexpression was first confirmed in a small-scale assay, comparing lysate from transformed cells that were induced or uninduced. Successful protein expression was determined by the appearance of a band corresponding to 201.7 kDa in the induced fraction. (B) Lysate containing overexpressed protein was purified using a Ni^{2+} charged chelating column. Fractions containing the overexpressed protein were pooled and dialysed overnight. (C) The sample was then passed through a heparin column and protein containing fractions pooled. (D) Finally, the protein was passed through a SEC column. Protein containing fractions were pooled and concentrated using a centrifugal concentrator (100 kDa MWCO). Chromatograms corresponding to panels B-D supplied in Figs. A1–A3. (E) Purified recombinant Cas9.

A.

B.

C.
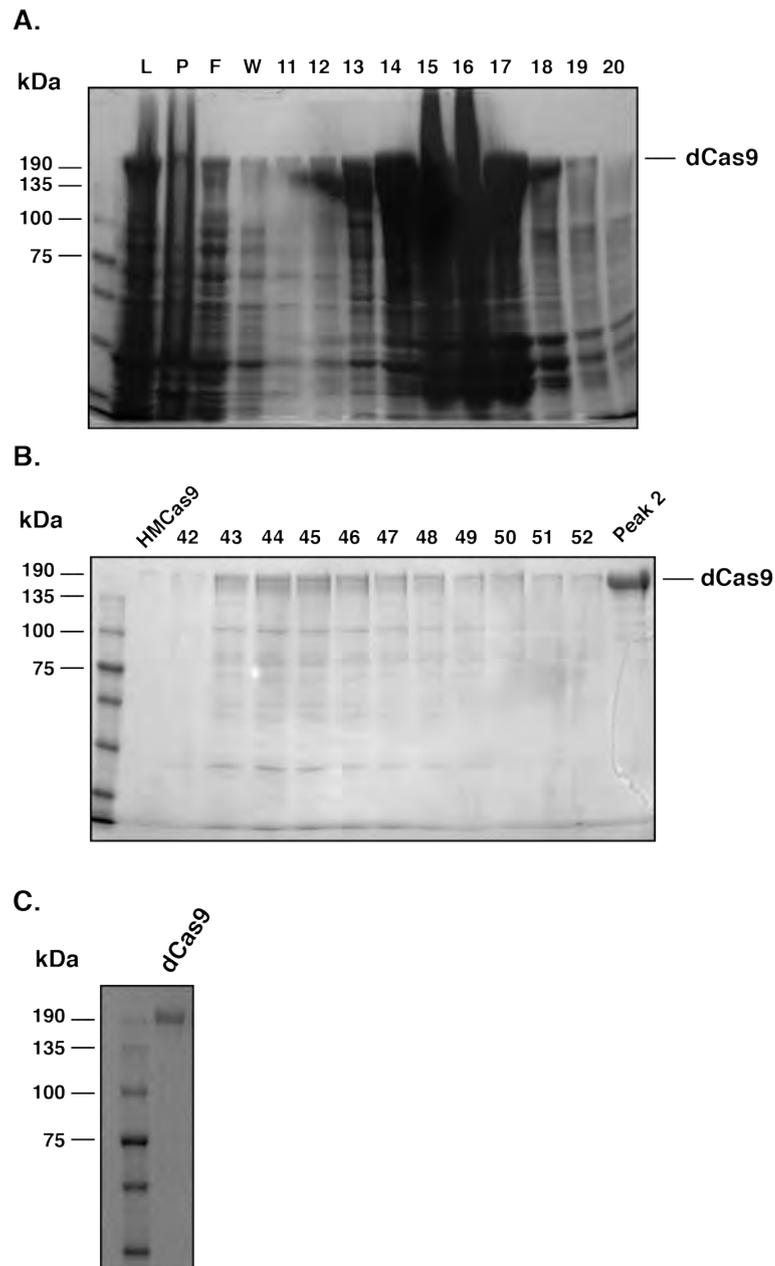
Figure 3.3: Purification of His6-MBP-dCas9. Cas9-proteins were overexpressed in *E. coli*. Successful protein expression was determined by the appearance of a band corresponding to 201.7 kDa in induced lysate. (A) Lysate containing overexpressed protein was purified using a Ni^{2+} charged chelating column. Fractions containing the overexpressed protein were pooled and dialysed overnight. The sample was then passed through a heparin column (data not shown) and protein containing fractions pooled. L, lysate sample; P, pellet sample, F, flow-through sample; W, wash-through sample. (B) Finally, the protein was passed through a SEC column. Protein containing fractions were pooled and concentrated using a centrifugal concentrator (100 kDa MWCO). Chromatograms corresponding to panels A and B supplied in Figs. A4–A5. (C) Purified recombinant dCas9 at the expected size of 201.7 kDa.
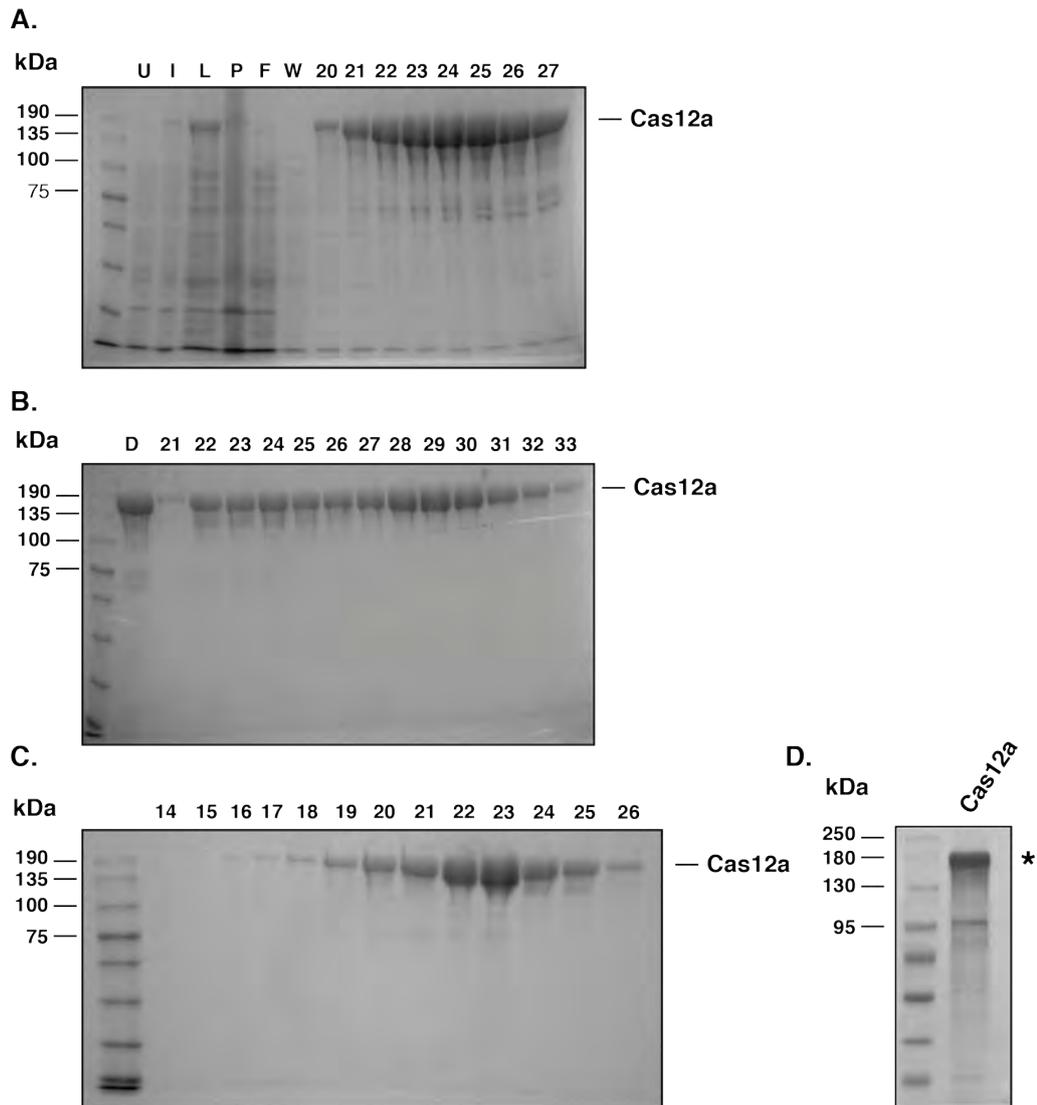
Figure 3.4: Purification of His6-MBP-Cas12a. Cas12a was overexpressed in *E. coli* comparing lysate from transformed cells that were induced or uninduced. Successful protein expression was determined by the appearance of a band corresponding to 199.6 kDa in the induced fraction. (A) Lysate containing overexpressed protein was purified using a Ni^{2+} charged chelating column . U, uninduced sample; I, induced sample; P, pellet sample, F, flow-through sample; W, wash-through sample;. Fractions containing the overexpressed protein were pooled and dialysed overnight. (B) The sample was then passed through a heparin column and protein containing fractions pooled. D, dialysed sample. (C) Finally, the protein was passed through a SEC column. Protein containing fractions were pooled and concentrated using a centrifugal concentrator (100 kDa MWCO). Chromatograms corresponding to panels A-C supplied in Figs. A6–A8. (D) Purified recombinant Cas12a. * denotes the position of Cas12a at 199.6 kDa.

In the case of dCas9, due to point mutations which deactivate the nuclease domains of the protein, cleavage assays were not necessary.

First, a baseline for nuclease activity by Cas9 and Cas12a was established using a dsDNA substrate containing PAM sites recognisable by both proteins. Controls were also included containing protein and sgRNA only to ensure that nuclease activity was complex specific (Fig. 3.5A, Fig. 3.6A). The results confirm, for both Cas9 and Cas12a, that RNP assembly is required to cleave the substrate. Cas9 was shown to cut, a mean of 67.8% of the substrate at a concentration of 40 nM, increasing to 87.1% at 80 nM. Cas12a demonstrated 4.8-fold lower activity than Cas9, cleaving an average of 18.2% of the substrate DNA at a concentration of 80 nM.

All three Cas proteins were confirmed to be capable of forming DNA:protein complexes (Fig. 3.5B, 3.6B, 3.7A), and R-loops (Fig. 3.5C, 3.6C, 3.7B) using EMSAs. This was shown to be RNP specific, with sgRNA only lanes displaying no complex formation and Cas protein-only lanes showing some smearing. This was unsurprising as it has been established that Cas apoenzymes are capable of making non-specific, transient contacts with DNA in the absence of an RNA guide[367]. Due to complex stability and its purpose in later assays, dCas9 EMSAs were quantified showing that 80 nM protein was able to complex with a mean of 64.5% of the substrate DNA, with the remainder lost to dissociation.

3.2.3 HelQ is unable to remove a Cas9 roadblock from short, forked substrate DNA

Building upon its association with high gene-editing efficiency, we hypothesised that HelQ removes Cas-protein roadblocks to facilitate access for other DNA repair proteins. To determine this, we conducted unwinding assays using forked DNA substrates complexed with dCas9-RNPs to create an obstacle.
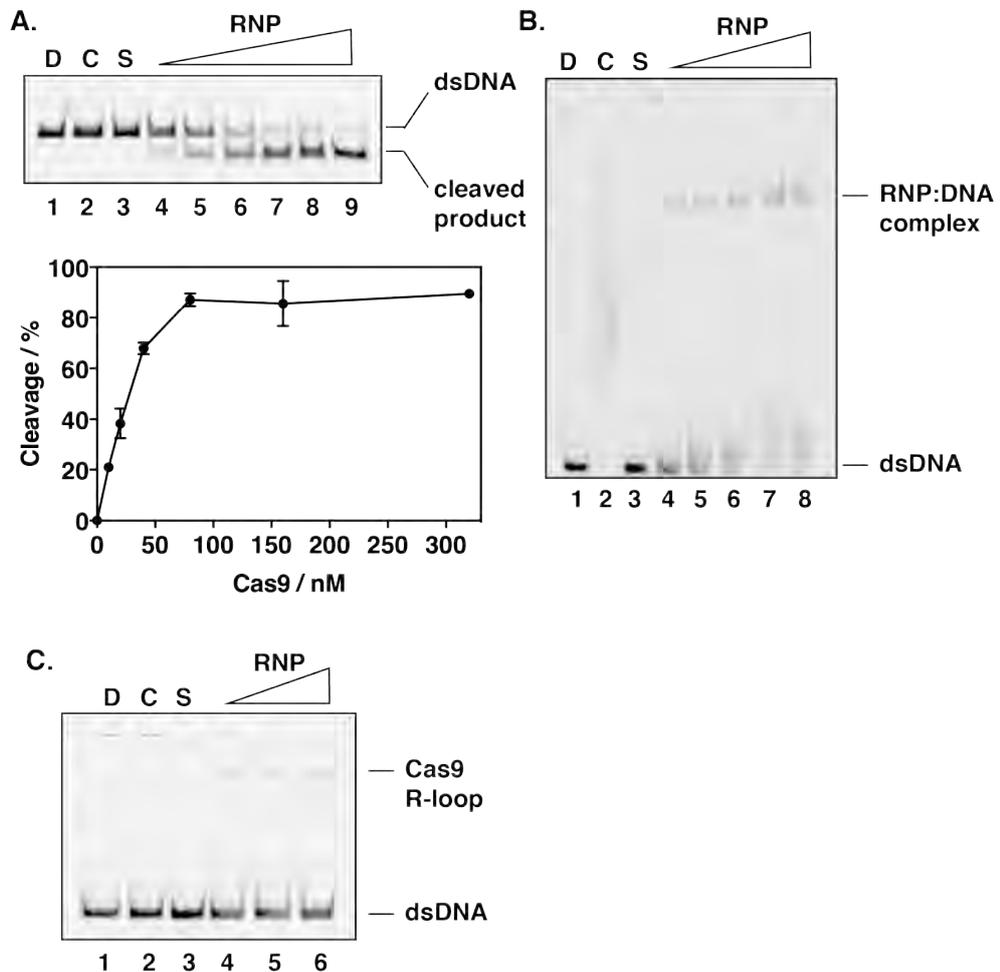
Figure 3.5: Assessing the activity of recombinant Cas9. (A) Cleavage activity of Cas9. RNPs were formed by combining purified Cas9 with *in vitro* transcribed sgRNA. RNPs were combined with 25 nM dsDNA substrate and incubated for 30 minutes at 37. RNPs were used at concentrations of 10, 20, 40, 80, 160, and 320 nM. Mean cleavage was plotted and error bars represent standard deviation of assays performed in triplicate. Controls to demonstrate proper complex formation for cleavage included: DNA only, D; DNA incubated with Cas9 only, C; and DNA incubated with sgRNA only, S. (B) DNA binding of a dsDNA substrate by Cas9. RNPs were incubated together with 25 nM dsDNA substrate in the absence of Mg^{2+} to prevent cleavage. RNPs were used at concentrations of 10, 20, 40, 80, and 160 nM. (C) R-loop formation by Cas9 on a dsDNA substrate. RNPs were formed as in (B) but were then incubated with proteinase K to degrade Cas-protein. RNPs were used at concentrations of 250 nM, 500 nM, and 1000 nM. All images are representative of assays done in triplicate.
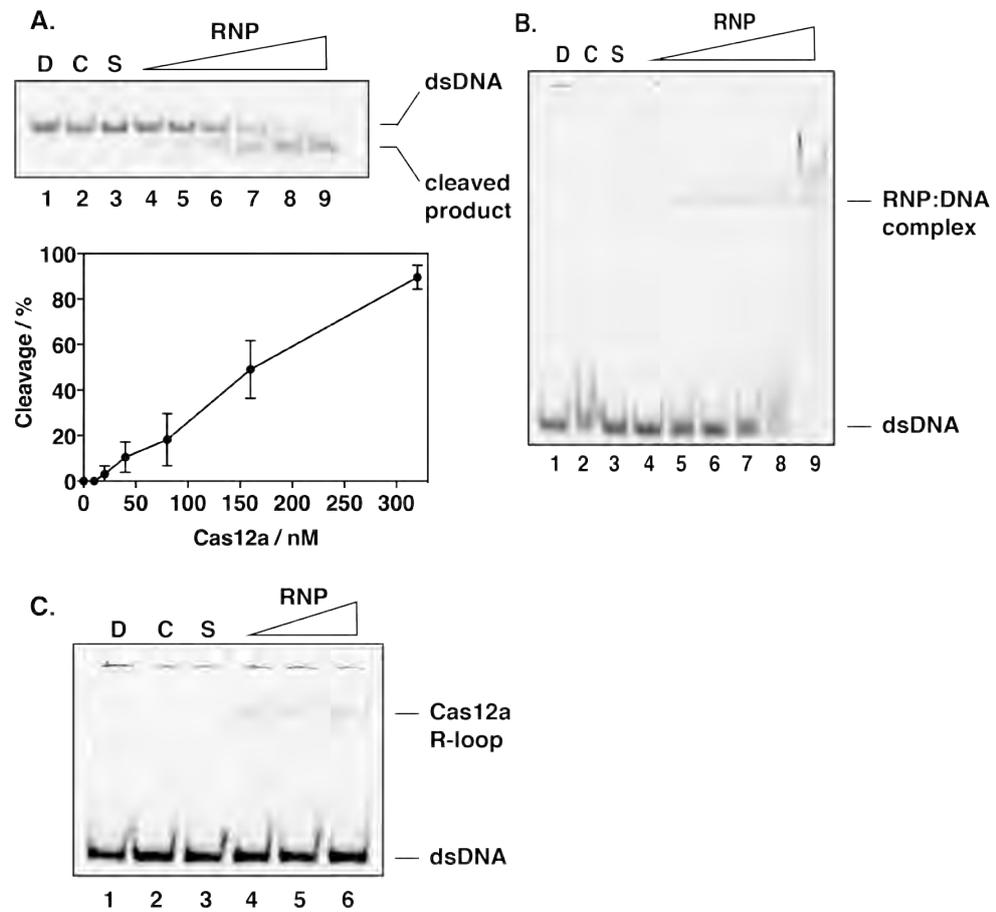
Figure 3.6: Assessing the activity of recombinant Cas12a. (A) Cleavage activity of Cas12a. RNPs were formed by combining purified Cas12a with *in vitro* transcribed sgRNA. RNPs were combined with 25 nM dsDNA substrate and incubated for 30 minutes at 37 °C. RNPs were used at 10, 20, 40, 80, 160, and 320 nM. Mean cleavage was plotted and error bars represent standard deviation of assays performed in triplicate. Controls to demonstrate proper complex formation for cleavage included DNA only (D), DNA incubated with Cas9 only (C), and DNA incubated with sgRNA only (S). (B) DNA binding of a dsDNA substrate by Cas12a. RNPs were incubated with 25 nM dsDNA substrate in the absence of Mg^{2+} to prevent cleavage. RNPs were used at concentrations of 10, 20, 40, 80, and 160 nM. (C) R-loop formation by Cas12a on a dsDNA substrate. RNPs were formed as in (B) but were then incubated with proteinase K to degrade Cas-protein. RNPs were used at concentrations of 250 nM, 500 nM, and 1000 nM. All images are representative of assays done in triplicate.
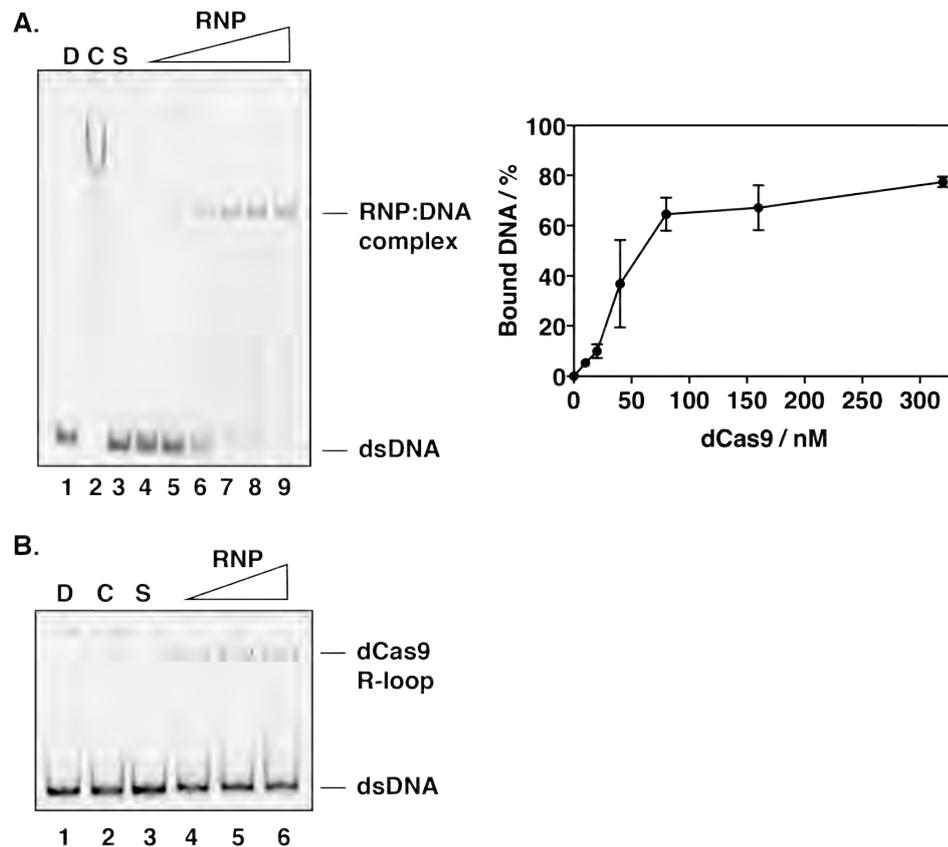
Figure 3.7: Assessing the activity of recombinant dCas9. (A) DNA binding of a dsDNA substrate by dCas9. RNPs were incubated together with 25 nM dsDNA substrate. RNPs were used at concentrations of 10, 20, 40, 80, and 160 nM. Mean binding was plotted and error bars represent standard deviation of assays performed in triplicate. Controls to demonstrate proper complex formation for cleavage included DNA only (D), DNA incubated with Cas9 only (C), and DNA incubated with sgRNA only (S). (B) R-loop formation by dCas9 on a dsDNA substrate. RNPs were formed as in (A) but were then incubated with proteinase K to degrade Cas-protein. RNPs were used at concentrations of 250 nM, 500 nM, and 1000 nM. All images are representative of assays done in triplicate.

First, HelQ unwinding of the forked substrate in the absence of dCas9 a roadblock was quantified using acrylamide gel-based assays (Fig. 3.8A). HelQ was confirmed as capable of unwinding the forked substrate, achieving a mean maximum of 77.3 % strand dissociation at a concentration of 80 nM.

Having confirmed this, we next optimised the system in which the assays would take place. Both HelQ and dCas9 have been established to be optimally active in different reaction buffers. As such, we tested the activity

of both proteins to in HB and NEBuffer 3.1 to determine the most suitable to conduct the roadblock assay. The performance of HelQ was assessed using time-course assays taking place over 30 minutes (Fig. 3.8B). The activity of dCas9 was observed using EMSAs to confirm DNA:protein complex formation (Fig. 3.8C).

HelQ, at 80 nM, was able to unwind the forked substrate in both buffers over the course of the 30 minute reaction, with increased unwinding at earlier timepoints in HB relative to NEBuffer 3.1 (Fig. 3.8B). After 5 minutes, HelQ in HB was able to unwind a mean of 42 % of the substrate DNA, while the HelQ in NEBuffer 3.1 unwound 18.1 %. After 30 minutes, the difference in protein activity in both buffers narrowed, with both unwinding at means over 75 %. Likewise, dCas9 was capable of forming stable DNA:protein complexes in both HB and NEBuffer 3.1 (Fig. 3.8C). The subsequent roadblock assays were carried out in HB.

Having confirmed the activity of both proteins in the optimised system, HelQ removal of a dCas9 roadblock was determined using a hybrid EMSA-unwinding assay (Fig. 3.9). In this assay, removal of the dCas9 roadblock would be observable by a loss of DNA super-shift and a return to free, ds-DNA substrate, or complete unwinding resulting in dissociated ssDNA. The reassociation of dCas9-RNPs with substrate DNA was prevented by the addition of a 10-fold excess of unlabelled 'trap' DNA containing sequence complementary to the sgRNA. Controls were included to confirm substrate unwinding by HelQ (Fig. 3.9, lanes 1–3) and the correct formation of RNP:DNA complexes (Fig. 3.9, lanes 4–6). The results suggest that HelQ is unable to unwind past, or fully dissociate a dCas9 roadblock on DNA as there was no visible reduction in RNP:DNA complexes and no increase in free or unwound substrate.
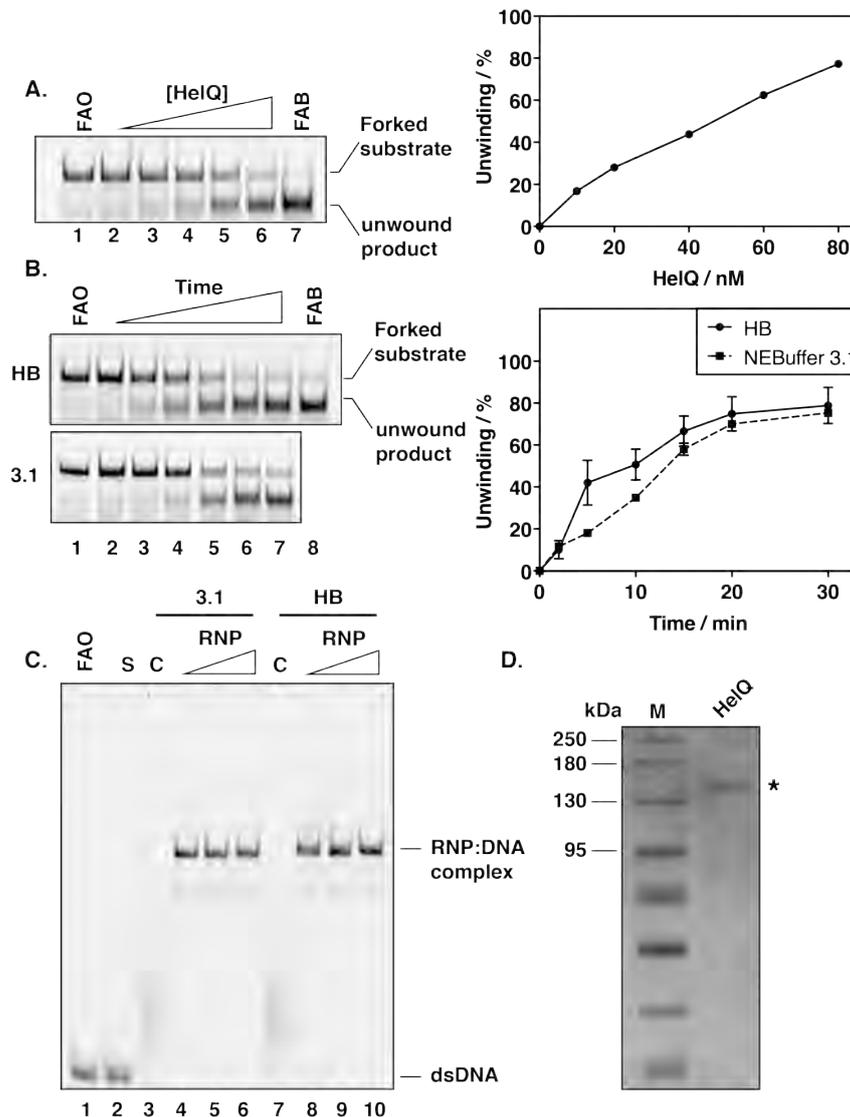
Figure 3.8: Assay optimisation for roadblock removal assays. (A) Titrating the concentration of HelQ in helicase buffer (HB) to optimise observable unwinding. HelQ was added to reactions at 10, 20, 40, 80, and 160 nM. Unwinding was quantified against a DNA-only control assays (FAO). Also included as a control was a boiled, substrate-only reaction (FAB) to represent full fork dissociation. (B) Unwinding of forked dsDNA substrate by 80 nM HelQ as a function of time. The compatibility of unwinding in NEBuffer 3.1 was assessed against HB. Assays were stopped at timepoints of 0, 2, 5, 10, 15, and 20 minutes. (A,B) Error bars represent standard deviation for experiments performed in triplicate. (C) Comparison of dCas9 RNP binding in NEBuffer 3.1 and HB. RNPs were incubated with the same substrate in (A-B) at concentrations of 40, 80, and 160 nM. All images are representative of assays done in triplicate. (D) SDS-PAGE gel image demostrating the purity of HelQ protein (*, 137 kDa) obtained from Tabitha Jenkins.
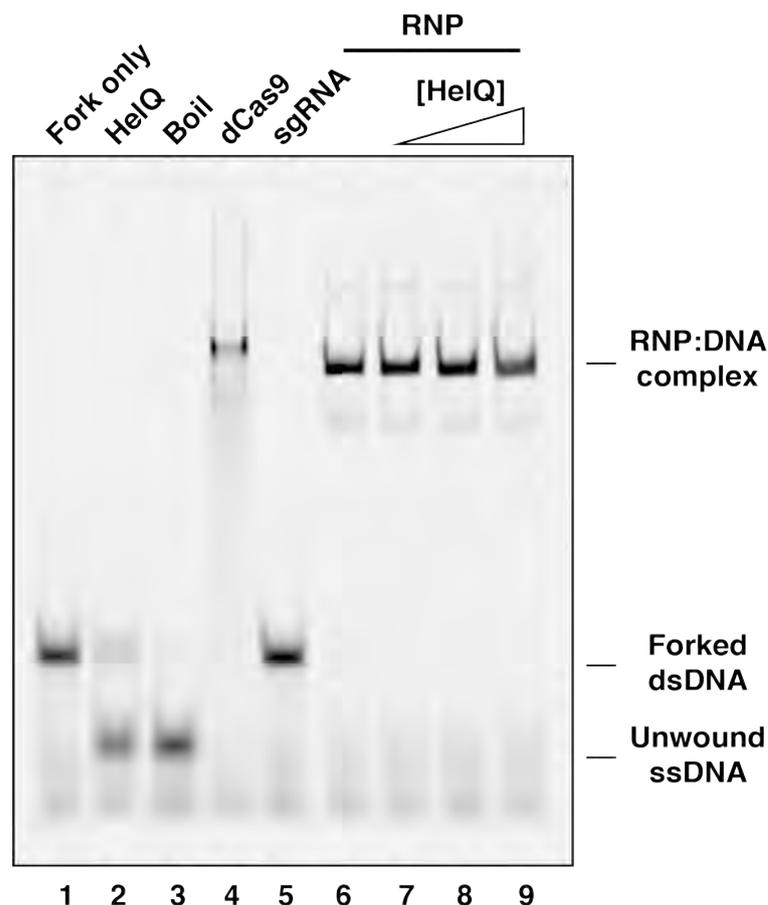
Figure 3.9: HelQ is unable to remove dCas9 roadblocks from a short, synthetic DNA substrate. Unwinding was confirmed using control assays (lanes 1–3) comprising fork only and boiled fork samples, as well as an assay containing 160 nM HelQ. RNP formation and DNA binding was confirmed by controls (lanes 4–6) containing dCas9, sgRNA, and RNP only. RNP:DNA complexes were formed prior to HelQ addition. HelQ was added to reactions at concentrations of 40, 80, and 160 nM (lanes 7–9). Gel image is representative of assays performed in triplicate.

3.2.4 HelQ is unable to remove a dCas9 roadblocks from supercoiled plasmid DNA

Having failed to observe HelQ-mediated removal of dCas9 RNPs from a short, synthetic substrate, we next tested the ability of the helicase to remove a roadblock using supercoiled plasmid DNA. This was informed by the work of Killelea et al. [75] which reported the removal of Cascade roadblocks by the bacterial helicase RecG.

To accommodate the increased size of plasmid DNA, agarose gels were

used to conduct the same unwinding assay employed in Section 3.2.3. RNPs were titrated into reactions containing plasmid DNA and demonstrated an observable shift, with complete substrate binding achieved at a concentration of 20 nM (Fig. 3.10A). Control lanes containing sgRNA and dCas9-only behaved as expected, with the the former having no effect on plasmid DNA mobility, while the impact of the latter was due to non-specific binding, as reported by Sternberg et al.[367].

HelQ was then titrated into reactions containing DNA:RNP complexes to assess its ability to remove the roadblock. Assays were also supplemented with a 10-fold excess of unlabelled ssDNA 'trap' to prevent RNPs reannealing to plasmid DNA once removed. Controls included assays to establish proper RNP formation and DNA-binding, as well as an assay supplemented with proteinase K to demonstrate the loss of super-shift by a plasmid from which dCas9 had been removed. These behaved as expected, with the sgRNA and dCas9-only lanes having no impact on the expected shift, despite dCas9 not causing aberrant shifting due to non-specific binding on this occasion. Assays containing HelQ demonstrated a failure to remove the RNP roadblock during the 30 minute reaction (Fig. 3.10B), suggesting that its function is not in roadblock removal during gene editing.

3.2.5 Reconstituting a cell-free system for the study of template integration following Cas-protein cleavage

In a 2018 paper, Sansbury et al.[342] described a system for the integration a donor template into plasmid DNA using Cas-proteins and human cell derived extracts (Fig. 3.11). We aimed to reconstitute this system for our own purposes of studying the mechanistic basis for the DSB repair that achieved the integration. This began with the production of CFEs and carrying out *in vitro* integration assays using template DNA analogous to those in the published work as proof-of-principle.
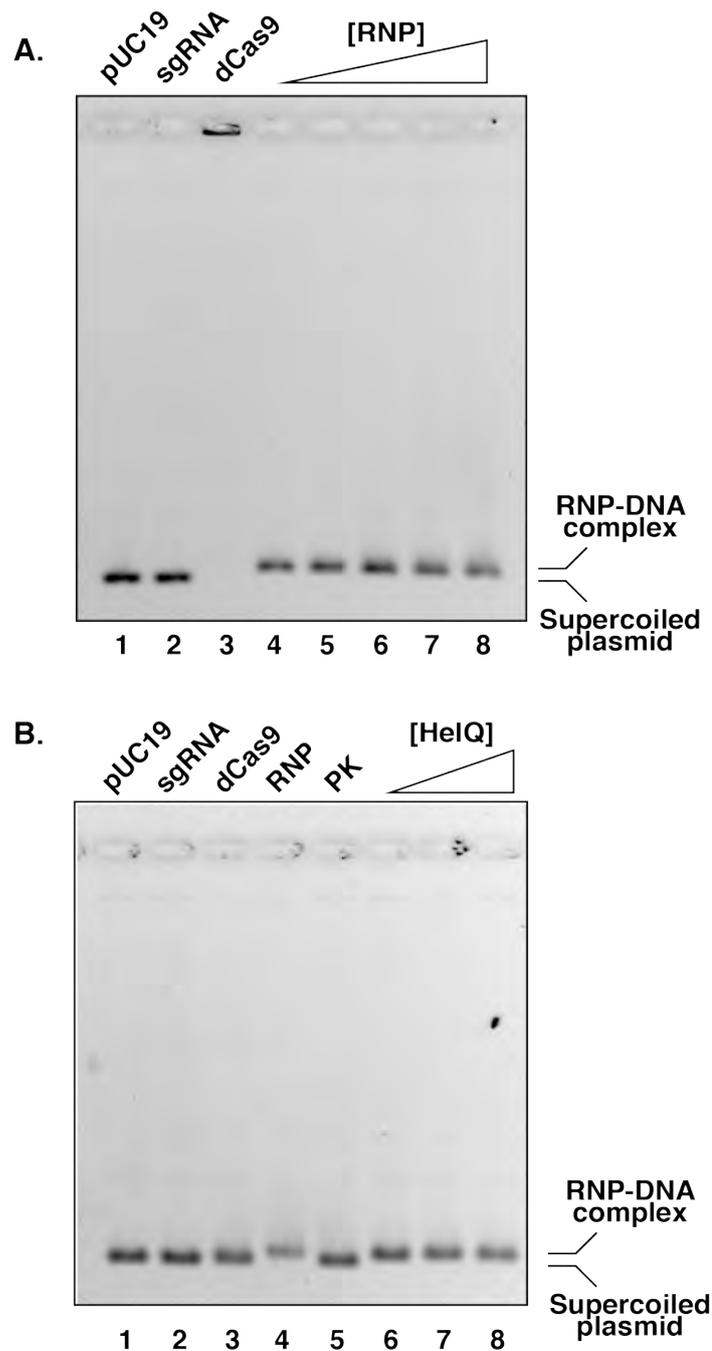
Figure 3.10: HelQ is unable to remove dCas9 roadblocks from a super-coiled plasmid DNA substrate. (A) Optimisation of RNP required for complete RNP:DNA complex formation. dCas9 RNPs were added to plasmid DNA at concentrations of 20, 40, 80, 160, and 320 nM. (B) HelQ unwinding assay against a dCas9 RNP roadblock. Complexes between dCas9 RNPs and plasmid DNA were formed prior to the addition of HelQ protein. As a control for dissociation, a lane containing RNP was supplemented with proteinase K (lane 5, PK). dCas9 RNPs were added to plasmid DNA at 20 nM. HelQ was added to reactions at 80, 160, and 320 nM. Gels are representative of assays carried out in duplicate.
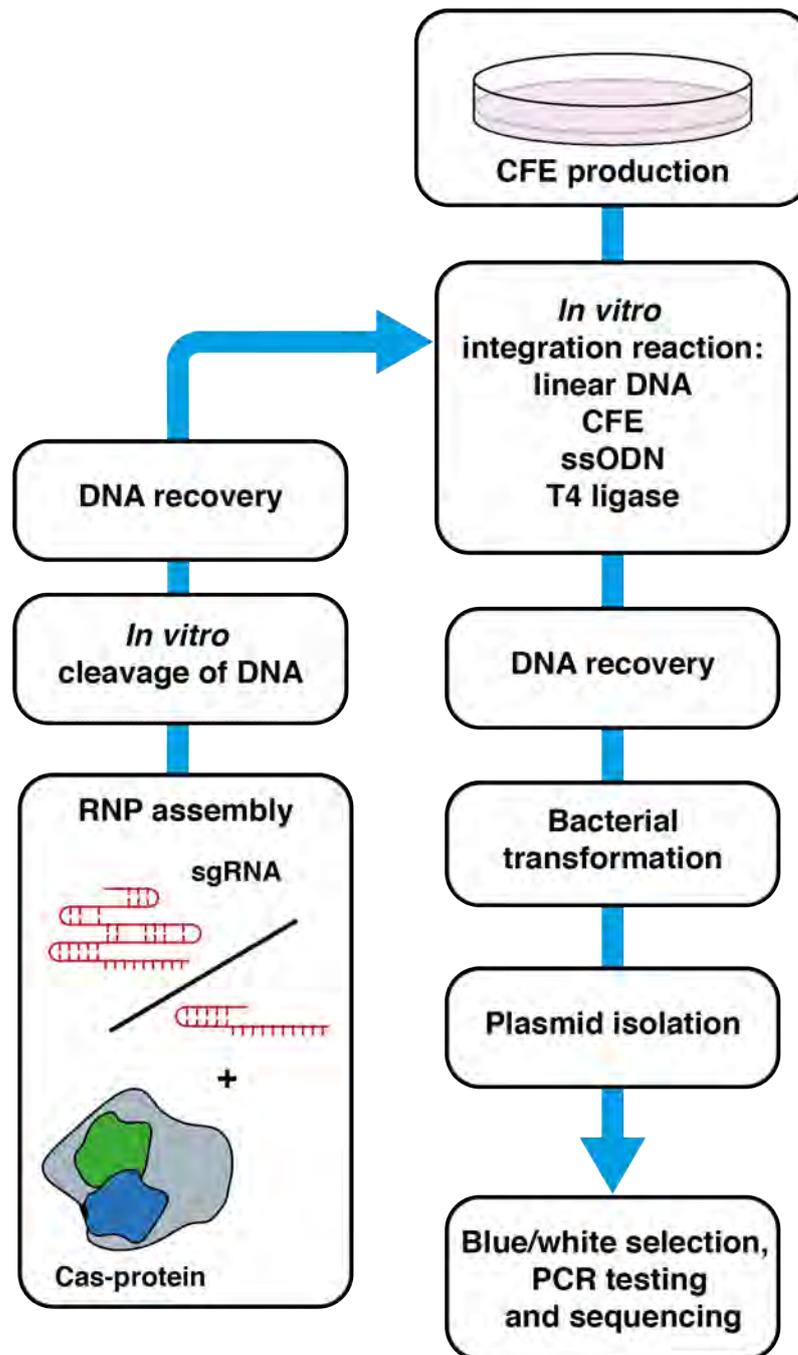
Figure 3.11: A workflow diagram for ssODN integration using cell-free extracts. To form RNPs, purified Cas9 or Cas12a are combined with the appropriate sgRNA. These are used to cleave the target DNA, which is then recovered via spin column. Separately, cell-free extracts (CFEs) are produced using the desired cell lines. The RNP and CFE are combined in an *in vitro* reaction with template DNA and ligase to complete integration. The resulting DNA is recovered via spin column and transformed into *E. coli*. Successful disruption of *LacZα* is easily visible by blue/white screening. White colonies can then be selected for subsequent testing by PCR and sequencing.

3.2.5.1 Generation of Cas9 and Cas12a DSBs

To reconstitute the system used by Sansbury et al.[342], plasmid DNA was first linearised using recombinant cas-proteins, creating a DSB into which template could be integrated. RNPs were formed by combining recombinant Cas9/12a with appropriate purified sgRNA targeting *LacZα* in the pUC19 plasmid. Alongside controls to monitor correct RNP formation were a pUC19-only assay to establish the mobility of uncut DNA as well as an assay containing pUC19 supplemented with BamHI to indicate the position of correctly linearised plasmid DNA.

Agarose gel electrophoresis of the samples demonstrated successful cleavage of pUC19 plasmid DNA by both Cas9 and Cas12a (Fig. 3.12A and B, respectively). Both proteins were able to completely cleave the target DNA in a 15 minute reaction at a concentration of 40 nM. Samples containing Cas12a-RNPs were enriched for nicked plasmid DNA compared to Cas9 (Fig. 3.12B). This is unsurprising given the single-nuclease domain mechanism employed by Cas12a and suggested that a longer incubation time would be required.

3.2.5.2 Production of human cell-free extracts (CFE)

Alongside the production of linearised plasmid DNA, CFEs with which to supplement integration reactions were produced using U2OS cells. This was achieved by harvesting cultured U2OS cells and gently lysing them in a cold hypotonic buffer. To meet the required protein concentration to supplement reactions, the resuspension volume of CFEs was optimised (Table 3.1). The results determined that a resuspension volume of 250 μl gave a consistent protein concentration around 1 mg/ml, which was sufficiently high to supplement integration assays.
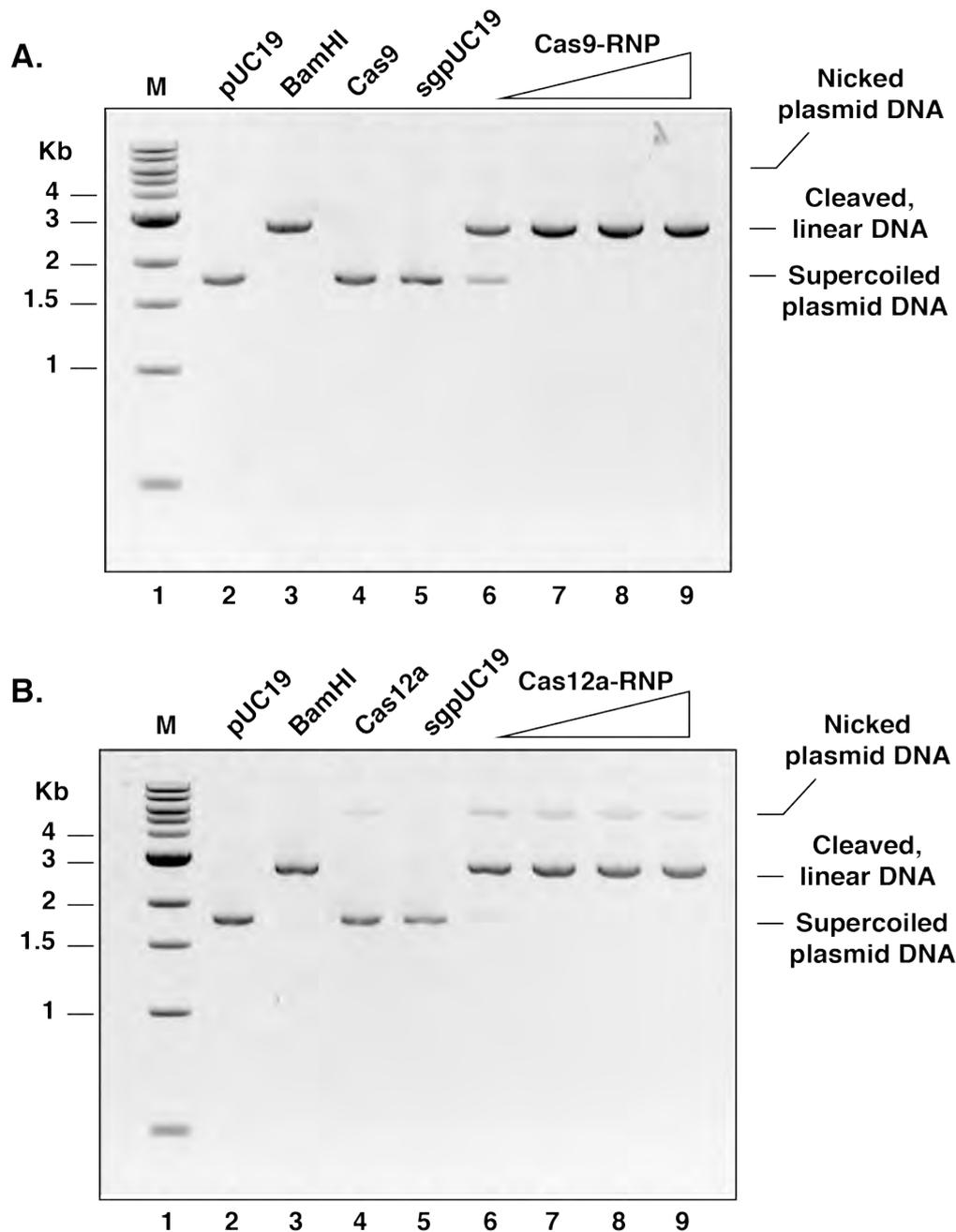
Figure 3.12: Cleavage of pUC19 by Cas9 and Cas12a. (A) Cas9 cleavage of pUC19. (B) Cas12a cleavage of pUC19. In both assays, each lane contains 250 ng of pUC19. A plasmid-only sample (lane 1) was included as a negative control to indicate the location of supercoiled, circular DNA. Plasmid DNA incubated with BamHI (lane 3) was also included, as a positive control for cleavage. Cas-protein or sgRNA (lanes 4, 5) were used as additional controls to show that cleavage is RNP specific. RNPs were used at concentrations of Cas-protein used were 20, 40, 80, and 160 nM. All images are representative of assays done in duplicate.

Table 3.1: Optimising the production of cell-free extracts. Yields are presented as mean values calculated from three independent batches of U2OS-derived extract.

Resuspension volume	Mean concentration (mg/ml)	SD (±)
100 μl	2.283	0.351
250 μl	1.166	0.128
500 μl	0.783	0.183

3.2.5.3 CFE-mediated integration of ssODNs at the site of Cas9 and Cas12a DSBs

We next attempted to reconstitute the method used by Sansbury et al.[342] with the intention of using the system to study differences between editing in extracts made from HelQ-depleted cells. Integration assays were first carried out using short ssODNs which possessed short regions of homology to Cas9 or Cas12a cut-sites and a NotI recognition sequence which would later be used to confirm template integration. Two different ssODNs were tested for each protein, corresponding to the sense and anti-sense strands of the cut DNA. Alongside this a dsDNA oligo comprised of the two ssODNs annealed together was tested.

Multiple controls assays were also included to: confirm the success of transformations (pUC19 only, (Fig 3.13A), the activity of T4 ligase (HindIII-only and HindIII plus T4 DNA ligase, (Fig 3.13B–C), and to observe sources of background mutation from the components of the assay ((Fig 3.13D–F).

DNA recovered from *in vitro* integration reactions was transformed into LacZα-deficient *E. coli* strain DH5α, which were subsequently plated on X-gal. The target site for both Cas9 and Cas12a was inside the LacZα complementation fragment contained by pUC19. Disruption of this gene due to

the integration of template DNA should therefore result in white colonies as the bacteria would be unable to metabolise the X-gal.

The results show successful disruption of LacZα in pUC19 (Fig. 3.14) with white colonies visible on all three test plates. There was a visible bias towards successful disruption in the reactions containing the anti-sense ssODN, which resulted in several-fold more colonies than assays containing the sense-ssODN and dsDNA .

White colonies were also visible on several of the control plates, suggesting that mutations disrupting the LacZα fragment in pUC19 had taken place. There was particularly high disruption of LacZα in the plates containing Cas-protein, CFE, and T4 ligase (Fig. 3.14).

Following successful transformation, plasmids were tested for successful integration of template DNA. Colonies were selected from each plate and lysed by boiling to provide template DNA for colony PCRs. The primers used in this reaction were designed to amplify a 376 bp fragment flanking the cut site for both Cas9 and Cas12a in the LacZα fragment of pUC19, with successful integration resulting in a 10 bp insertion containing a NotI site. As such, PCR products were digested with NotI to produce fragments of size 236 bp and 150 bp for DNA cut with Cas12a (Fig. 3.15), or 196 and 190 bp for DNA cut with Cas9 (Fig. 3.16).

Two sets of controls were included in this test. First, two blue colonies from each plate were examined to confirm the presence of unaltered pUC19 plasmid DNA. Second, white colonies from plates not containing any template DNA were included to determine whether they were spontaneous mutants and that cleaved products were NotI specific.

In the case of assays containing pUC19 cleaved by Cas12a, all control PCRs from blue colonies were successful, producing bands migrating at approximately 400 bp, consistent with the expected 376 bp PCR product. PCRs from white colonies were diverse in the efficacy of their amplification.
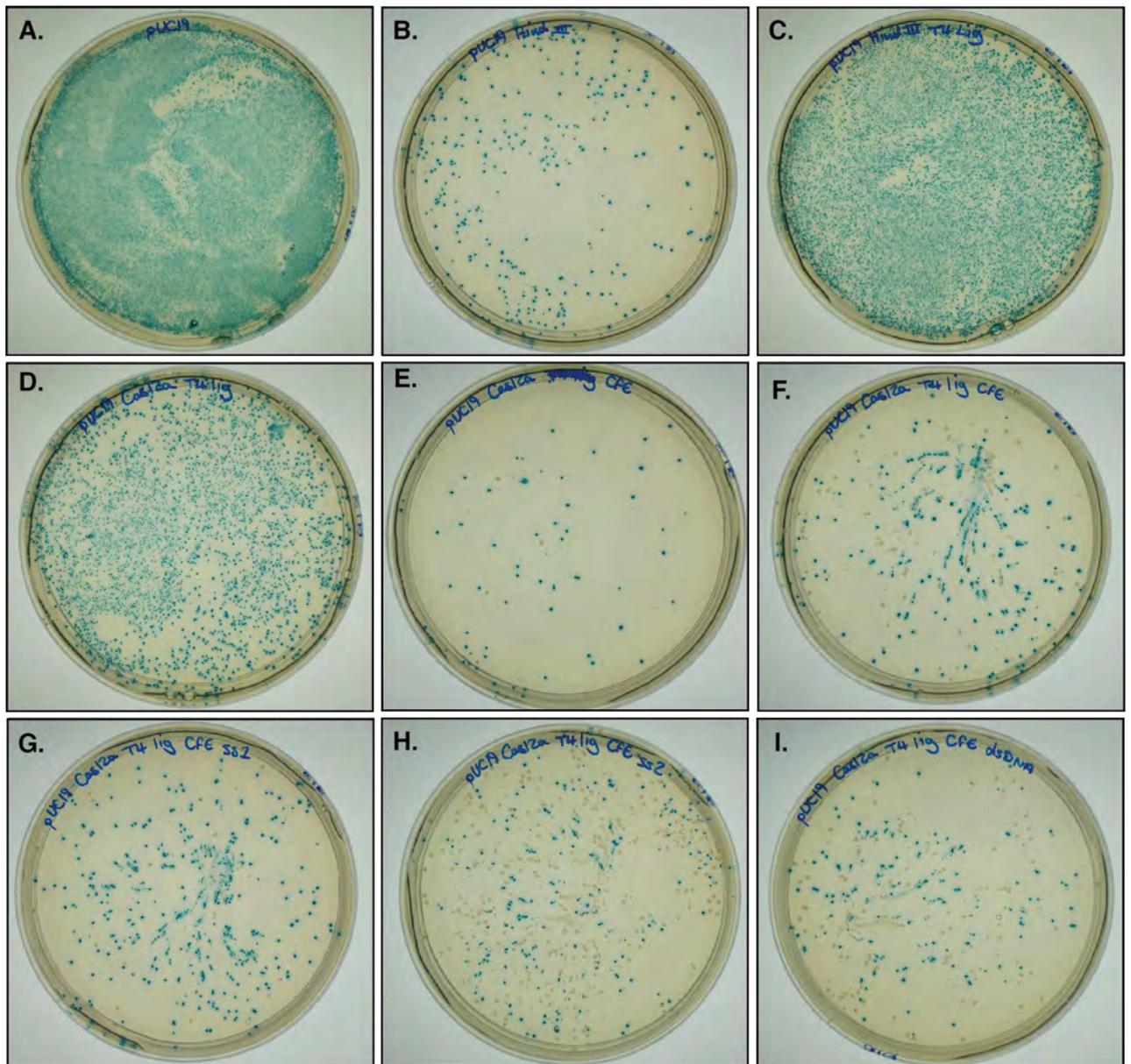
Figure 3.13: Blue-white screening to identify putative integration invents from *in vitro* assays. (A) pUC19-only. (B) pUC19 digested with HindIII. (C) pUC19-HindIII digest supplemented with T4 DNA ligase. (D-I) pUC19 digested with Cas-protein and supplemented with: T4 DNA ligase (D); cell-free extract (CFE, E); T4 DNA ligase and CFE (F); T4 DNA ligase, CFE, and an ssODN equivalent to the sense strand (G); T4 DNA ligase, CFE, and an ssODN equivalent to the anti-sense strand (H), T4 DNA ligase, CFE, and a dsDNA donor formed by annealing the two ssODNs (I). Images are representative of a full set of test plates.

pUC19	HindIII	HindIII + T4 ligase

Cas9/12a + T4 ligase	Cas9/12a + CFE	Cas9/12a + CFE + T4 ligase

	Cas9	Cas12a
Sense ssODN		
Anti-sense ssODN		
dsDNA donor		
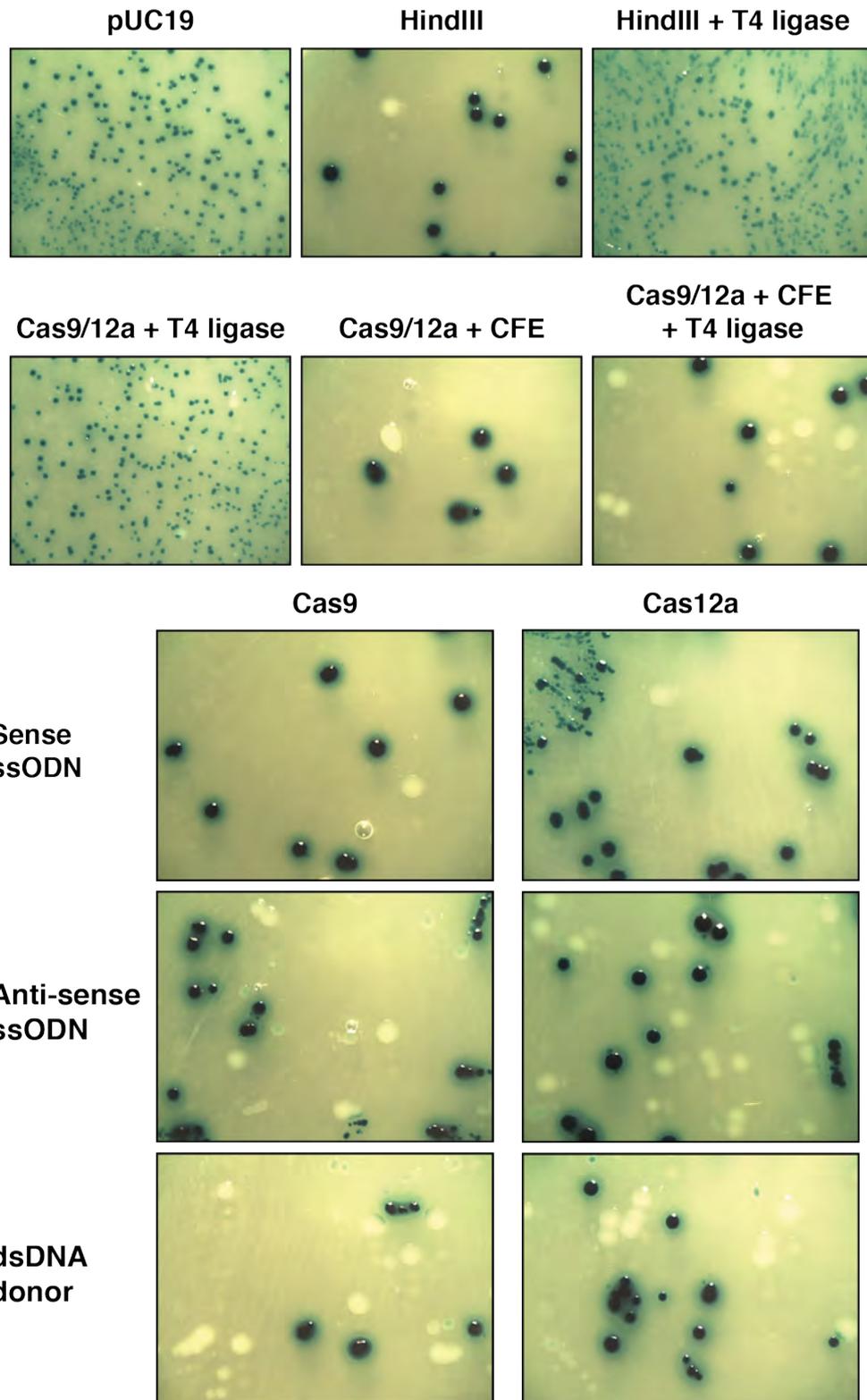
Figure 3.14: Blue-white colonies containing putative template integrations. Images are representative of blue-white colony compositions for each control and assay plate.

The control white colonies included from plate E failed to amplify and are therefore inconclusive. The white colony controls from plate F successfully amplified, and with the exception of one colony produced bands at a size of approximately 400 bp, which is suggestive of mutation during the *in vitro* assays or transformation. The exception to this is colony 3 from plate F which produced a single, faint band similar in size to the products of successful NotI digestion.

The results for Cas12a assays containing template DNA appear to show successful integration of the anti-sense and dsDNA templates DNA (Fig. 3.15). Of the six colonies tested, there was no evidence of template integration into pUC19 in reactions containing the ssODN comprising the sense strand. The anti-sense ssODN on the other hand appeared to have been successfully integrated in all 15 of the colonies tested. Results for the dsDNA template are harder to confirm due to poor amplification from several of the selected colonies. Despite this, it appeared that at least 12 of the 15 test colonies contained pUC10 into which the template DNA was successfully integrated.
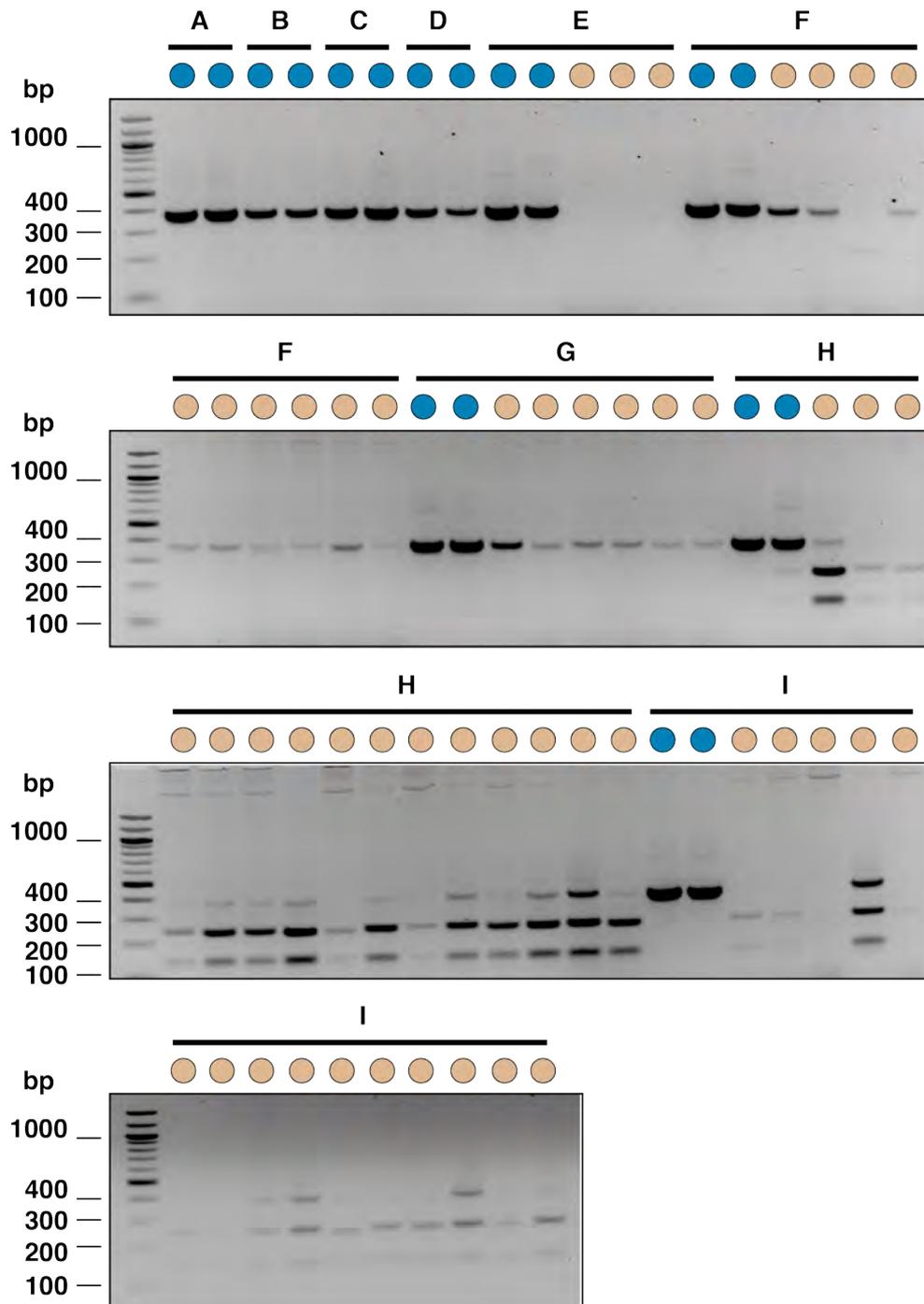
Figure 3.15: Digests to confirm integration events resulting from Cas12a *in vitro* assays. Colonies were picked from plates to confirm integration by NotI digest. Two control colonies were included to confirm the presence of plasmid DNA and successful PCR (blue). (A) pUC19-only. (B) pUC19 digested with HindIII. (C) pUC19-HindIII digest supplemented with T4 DNA ligase. (D-I) pUC19 digested with Cas-protein and supplemented with: T4 DNA ligase (D); cell-free extract (CFE, E); T4 DNA ligase and CFE (F); T4 DNA ligase, CFE, and an ssODN equivalent to the sense strand (G); T4 DNA ligase, CFE, and an ssODN equivalent to the anti-sense strand (H), T4 DNA ligase, CFE, and a dsDNA donor formed by annealing the two ssODNs (I).

The results for Cas9 integration assays also appear to show some successful insertion of the template DNA, albeit to a lesser extent that those using Cas12a (Fig. 3.16). All control PCRs from blue colonies were successful, producing bands migrating at approximately 400 bp, which is consistent with the expected 376 bp PCR product. The majority of PCRs from white colonies displayed poor amplification. DNA obtained from control white colonies on plates E and F amplified poorly, but did produce bands at approximately 400 bp, again suggesting mutation during the *in vitro* assays or transformation. The visible amplicons from control white colonies also do not appear to have been digested by NotI.

The results for Cas9 integrations assays appear to show successful integration of the anti-sense and dsDNA templates (Fig. 3.15). Of the ten colonies transformed with reactions using sense ssODN, only three amplified of these, there was no evidence of template integration. Likewise, the DNA obtained from reactions containing the anti-sense ssODN amplified poorly, with 4 of 15 producing a visible product. Of these, all four appeared to have successfully integrated the DNA template. DNA obtained from reactions using dsDNA template also amplified poorly, with only 5 of 15 colonies producing visible PCR products. From these , four produced digestion products, suggesting successful template DNA integration.

Following successful reconstitution of the CFE system, experiments were planned wherein integration would be compared in extracts produced from parental U2OS and HelQ-depleted cell lines. This aimed to support the findings of Richardson et al.[212], demonstrating reduced integration efficiency. We hypothesised that complementing HelQ-depleted extracts with recombinant HelQ would increase or restore integration efficiency and provide us a starting point for further biochemical screens to unpick the role of the protein in gene editing. Unfortunately, these assays were not able to be carried out and will need to be followed up on in future work.
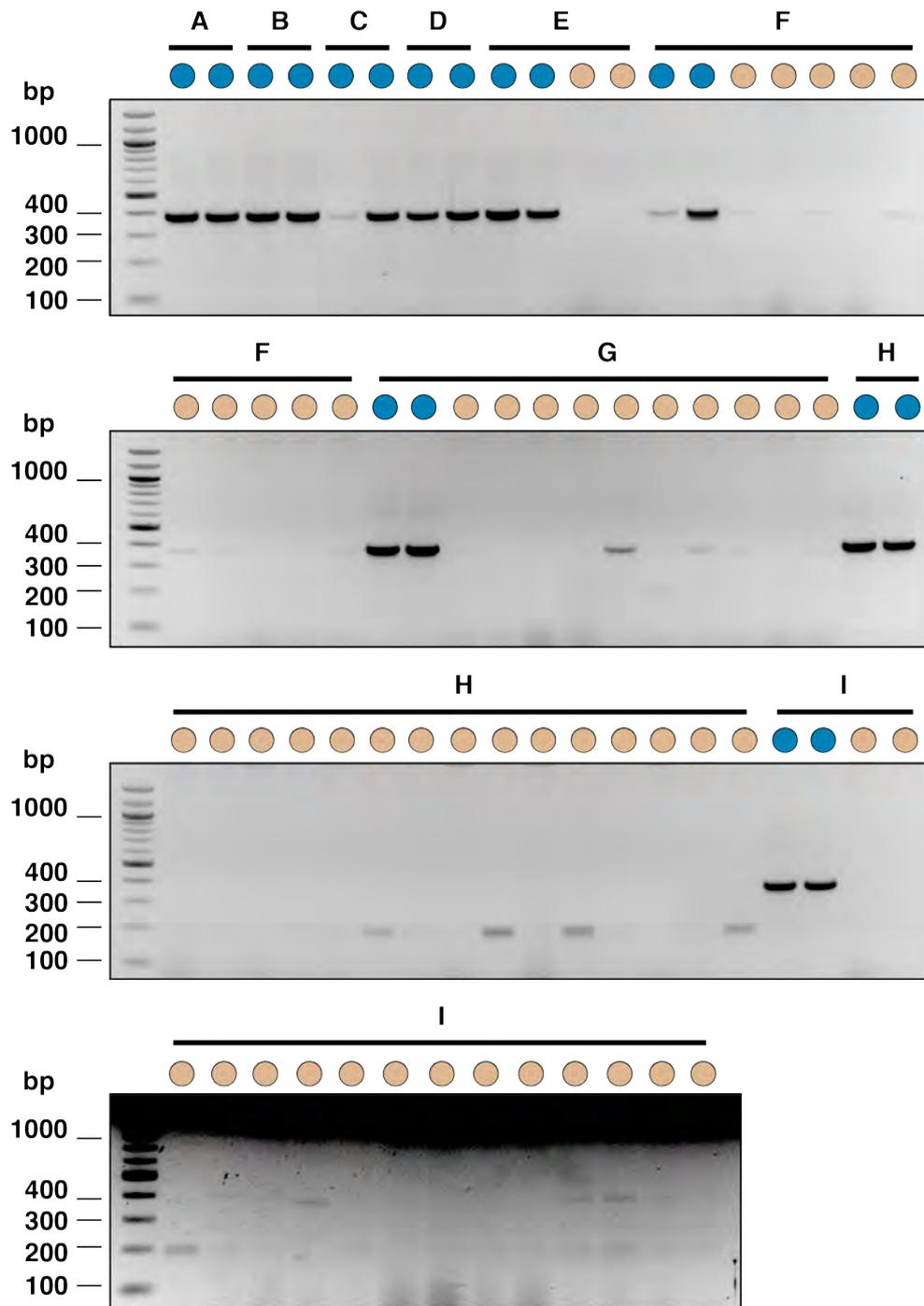
Figure 3.16: Digests to confirm integration events resulting from Cas9 *in vitro* assays. Colonies were picked from plates to confirm integration by NotI digest. Two control colonies were included to confirm the presence of plasmid DNA and successful PCR (blue). (A) pUC19-only. (B) pUC19 digested with HindIII. (C) pUC19-HindIII digest supplemented with T4 DNA ligase. (D-I) pUC19 digested with Cas-protein and supplemented with: T4 DNA ligase (D); cell-free extract (CFE, E); T4 DNA ligase and CFE (F); T4 DNA ligase, CFE, and an ssODN equivalent to the sense strand (G); T4 DNA ligase, CFE, and an ssODN equivalent to the anti-sense strand (H), T4 DNA ligase, CFE, and a dsDNA donor formed by annealing the two ssODNs (I).

3.2.6 *in vitro* assays to determine whether HelQ is capable of unwinding DNA:RNA hybrids

Alongside exploring the role of HelQ in successful gene-editing, we also became interesting in a more general role for the protein in resolving DNA:RNA hybrids such as R-loops. This was studied using unwinding assays on a variety of DNA:RNA hybrids including both simple forked substrates (Fig. 3.17) and more complex, looped substrates (Fig. 3.18). All unwinding assays were carried out alongside a control dsDNA fork known to be unwound by HelQ. Also included in each set of assays were substrate-only and boiled controls to indicate complete and dissociated substrate, respectively.

We first tested a variety of forked DNA:DNA and DNA:RNA substrates to determine the ability of HelQ to unwind hybrids (Fig. 3.17). The results show that HelQ could unwind both dsDNA and hybrid substrates to a similar degree. This was surprising in the case of substrates B and C, representing no fork and a 5, fork, which HelQ should unwind poorly, if at all. This is likely the result of a free 3′ end present in substrates B–D which allowed HelQ to unwind from an alternate, unintended site.
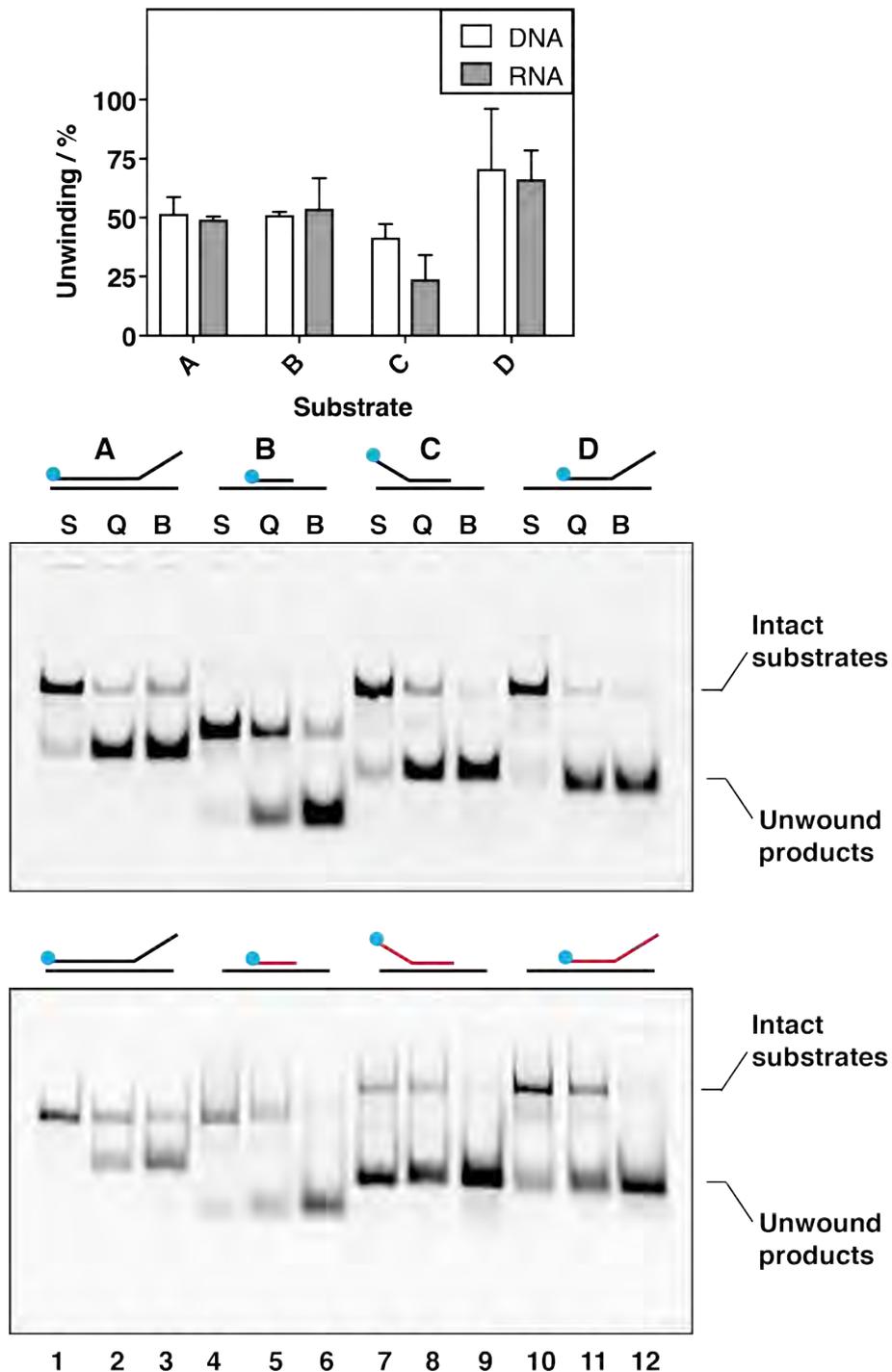
Figure 3.17: Assessing the ability of HelQ to unwind short, synthetic DNA:RNA hybrid structures. (A) HelQ unwinding of DNA substrates. (B) HelQ unwinding of DNA:RNA hybrid substrates. Unwinding was analysed using three different assays: One containing DNA only (S), one incubated with 80 nM HelQ (Q), and one a boiled DNA only sample to indicate unwinding (B)

Following the analysis of forked substrate unwinding, we then examined

the ability of HelQ to unwind a range of substrates simulating DNA-loops

(D-loops) and RNA-loops (R-loops). As before, these substrates were tested alongside a control forked dsDNA substrate known to be unwound by HelQ, as well as substrate-only, and boiled controls (Fig. 3.18).

In this series there were some notable differences between substrates, but no visible preference towards DNA-only, or DNA:RNA hybrid compositions. HelQ was able to unwind the structures similarly well, with mean values of 27.8 % (DNA) and 34.8 % (DNA:RNA), respectively, for substrate E. This is also seen in the case of substrate G, where HelQ unwound 35.9 % (G, DNA) and 22.0 % (G, DNA:RNA) of substrate. Notably, in the case of both D- and R-loops, HelQ largely failed to unwind substrate F which contained a 5′ tail, achieving means of 14.1 % and 1.1 %, respectively.

While a preference for substrates lacking a 5′ tail would make sense in the context of previous HelQ literature, we considered that HelQ may be able to unwind along the exposed ssDNA of the loop structure, creating an additional 3′ end that could subsequently be unwound as was potentially seen in Fig. 3.17.
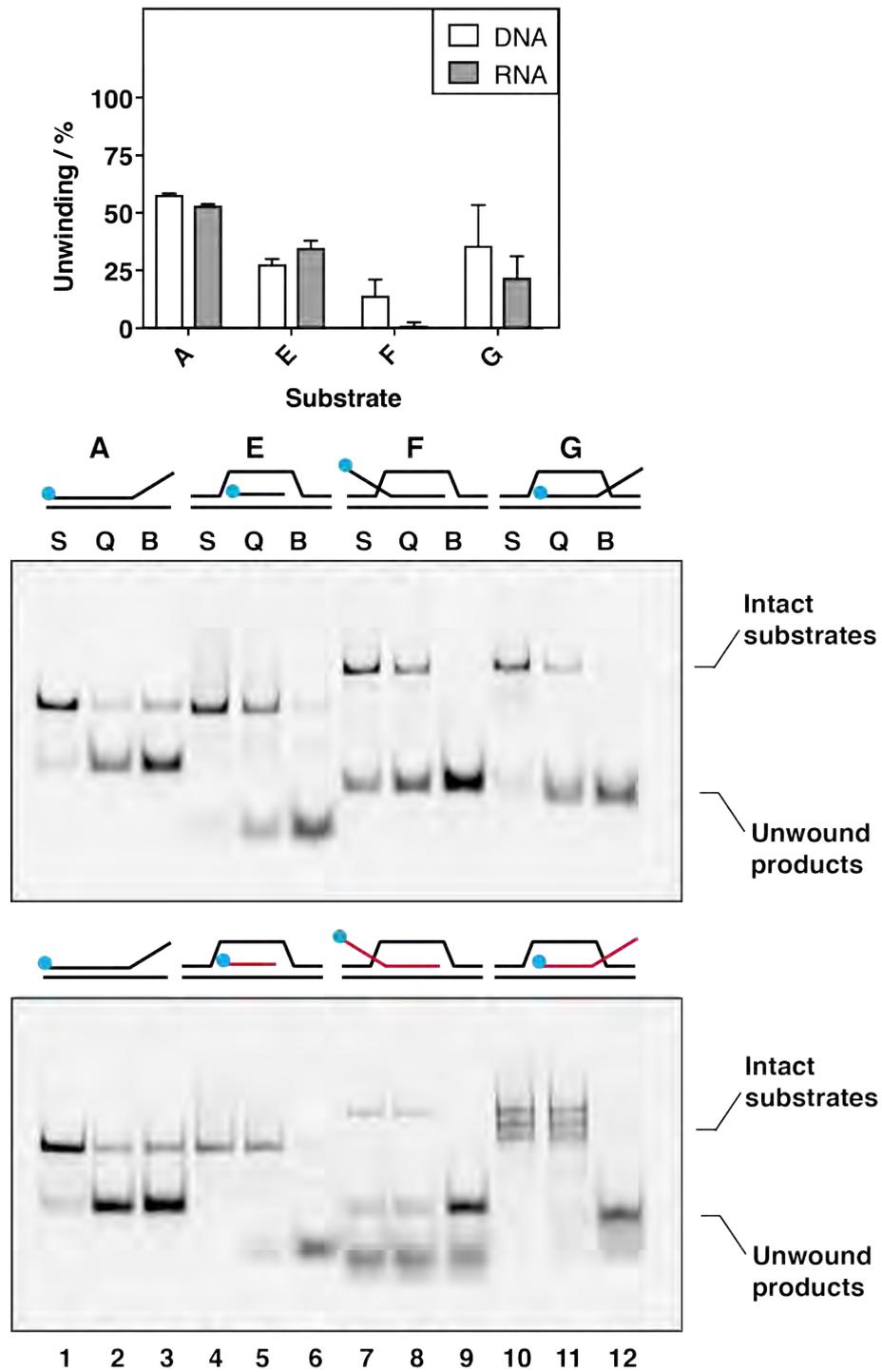
Figure 3.18: Assessing the ability of HelQ to unwind looped, synthetic DNA:RNA hybrid structures. (A) HelQ unwinding of looped DNA substrates. (B) HelQ unwinding of looped DNA:RNA hybrid substrates. Unwinding was analysed using three different assays: One containing DNA only (S), one incubated with 80 nM HelQ (Q), and one a boiled DNA only sample to indicate unwinding (B).

3.2.7 DNA:RNA hybrids accumulate in cells with compromised HelQ expression

Alongside *in vitro* assays, we also explored the impact of HelQ depletion and mutation on the accumulation of DNA:RNA hybrids in human cell lines. To do this, U2OS cells, alongside HelQ$^{-/-}$ and HelQ^{K365M} mutants, were fixed and probed with S9.6, a DNA:RNA hybrid-specific antibody. As a marker for nuclei and to indicate the regions to be quantified, cells were also stained with DAPI. Cultured cells were fixed to multi-well chamber slides and imaged using confocal microscopy.

The images show the successful staining of nuclei (Fig. 3.19, blue). Stained wild-type U2OS cells were observed to produce a diffuse, low-intensity signal from the S9.6-antibody. Both the 5G6 (HelQ$^{-/-}$) and 5G6W (HelQ^{K365M}) cell lines display an overall higher signal-intensity for S9.6 and also the appearance of foci in the nuclei. Quantification of the nuclear S9.6 signal for each cell line demonstrated a significant difference between the wild-type U2OS cells and both the HelQ$^{-/-}$ (P= 0.0124)and HelQ^{K365M} (P= <0.0001) mutants (Fig. 3.19B). Alongside the increase in nuclear signal, there was also an accumulation of high-intensity S9.6 spots outside of the nucleus which could be indicative of cellular instability.
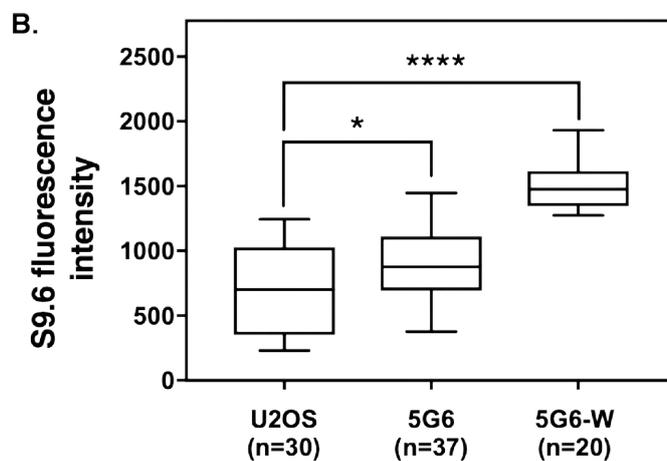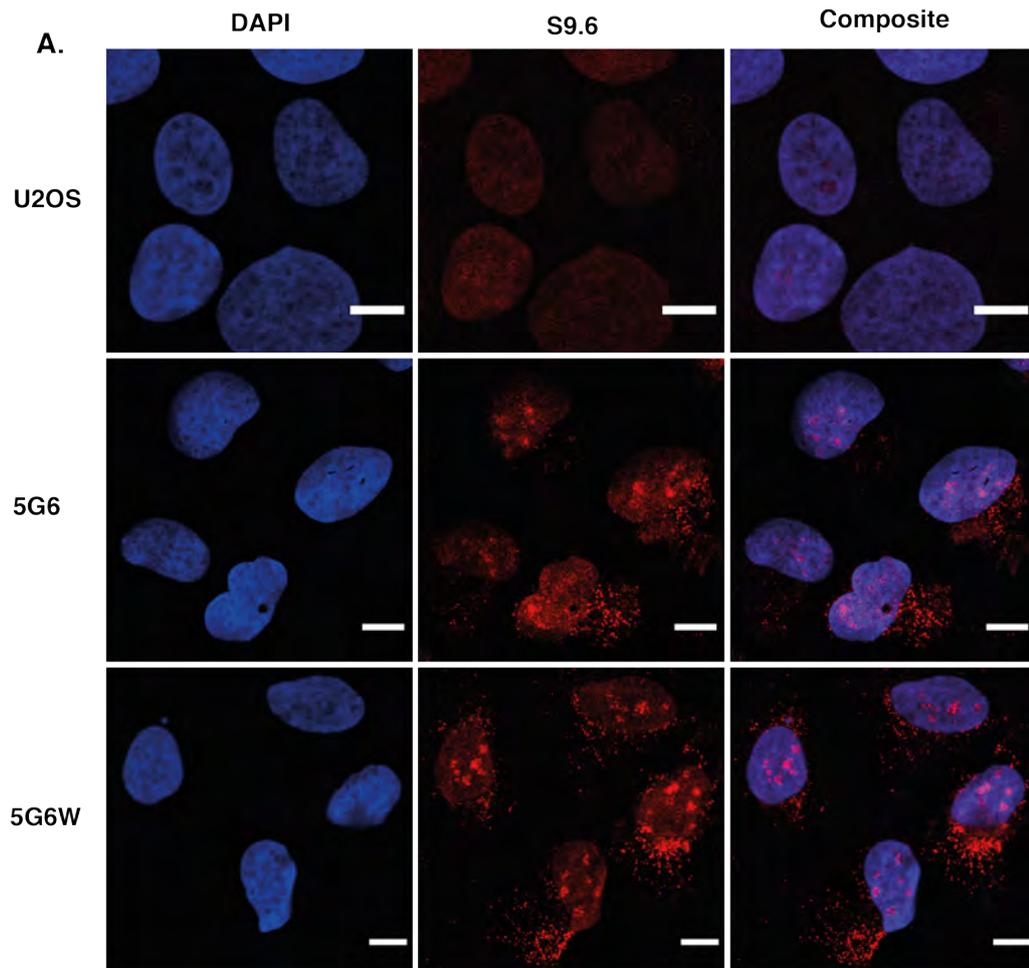
Figure 3.19: Immunostaining of U2OS and HelQ mutants by S9.6 antibody. (A) Nuclei were stained with DAPI (blue); DNA:RNA hybrids were stained with S9.6 (red). All scale bars, 10 μm. (B) Box plot showing quantification of S9.6 signal intensity. Elements displayed are: centre line, median; box limits, upper and lower quartiles; whiskers, minimum and maximum of all data. Significance was determined by unpaired two-tailed Student's t-test (*, P= 0.0124; ****, P= <0.0001). S9.6 signal in nuclei was quantified using the method of Rajani et al.[347].

3.3 Discussion

3.3.1 HelQ is unable to remove a dCas9 roadblock from dsDNA

A key gap in our knowledge of gene-editing is the interaction, if any, between exogenous Cas-nucleases and host DNA repair processes. Cas9 has been shown to remain bound to DNA at target sites for multiple hours following cleavage[367]. While dissociation characteristics such as strand release order have been determined *in vitro*, it remains unclear whether DNA repair processes have any involvement in the removal of the protein *in vivo*.

We hypothesised that a helicase such as HelQ would be a logical choice to displace a Cas-nuclease following gene-editing. This was supported by previous work which demonstrated helicase-mediated Cas-protein roadblock removal, as a study by Killelea et al.[75] observed that the *E. coli* helicase, RecG, was capable of displacing a Cascade roadblock from plasmid DNA. Alongside this, HelQ has already been established to have a role in the efficient integration of template DNA during editing via the FA pathway[212]. However, we were unsuccessful in observing such an interaction taking place with HelQ using both short, synthetic substrates and supercoiled plasmid DNA. This is in line with previously published research in which HelQ was shown to be unable to remove other protein blockages on DNA such as catalytically dead BamHI[229].

Work has however begun to emerge which may provide insight into interactions between gene-editing tools and endogenous repair pathways. Single-molecule analysis on a core fragment of BLM helicase observed displacement of a dCas9 roadblock from DNA through disruption of a weak interaction downstream of the PAM site, on the same strand as the recognition sequence[368]. This study, alongside similar work, raises the possibility that displacement may be protein or polarity dependent, establishing that DNA polymerases were unable to displace or bypass dCas9 roadblocks,

while RNA polymerases were only able to remove the blockage if translocating from downstream[368,369].

This work has been expanded on further by research demonstrating that the 5′–3′ helicase Pif1, which interacts with the human CMG complex, assists in bypassing a dCas9 roadblock to restart arrested DNA replication[370]. This not only deepens our understanding of the interactions between gene-editing tools and DNA repair proteins, but also suggests that an alternate approach could be taken to explore the role of HelQ further. It is possible that by attempting to study HelQ in isolation, we removed key context from any possible function it may have in assisting repair during editing. Based on the established role of HelQ in efficient integration during editing, existing knowledge of its interactions with proteins in the FA pathway such as FANCD2, and a functionally unclear association with the polymerase Pol ν, it may be interesting to explore a similar role for the protein as that of Pif1 in replication restart[371].

3.3.2 *In vitro* gene editing of plasmid DNA using Cas-nucleases and a mammalian cell-free extract

This work set out to reconstitute a cell-free system published by Sansbury et al.[342] to use as a model for the Cas-protein mediated integration of template DNA in mammalian cells. In this, the work appears to have partly succeeded, being able to demonstrate integration but ultimately not utilising the system to explore the impact of HelQ-depletion efficiency.

3.3.2.1 The production and yield of cell-free extracts

A key limitation of the approach used in this work is the overall yield of cell-free extract given the input required to produce it. Following optimisation of the buffer resuspension volume, it was possible to produce extract at concentrations consistent with the assays published in the initial, 2018

Sansbury et al.[342] paper. However, the volume required per assay versus the total volume produced meant that each batch could only be used for a single set of experiments. While this was easily solved by increasing the amount cells cultured for extract preparation, the system in its current form is not easily compatible with scaling up to include multiple cell lines or much larger batch production. Similar systems such as those used for cell-free protein synthesis are often produced using suspension cultures, which are more easily scaled-up[372,373].

It should also be noted that CFEs were not able to be produced at concentrations consistent with assays published in the subsequent 2019 Sansbury et al.[343] paper. Here the protocol was altered, increasing the CFE supplement to 175 μg, up from the original 20 μg, whilst simultaneously reducing the reaction volume from 35 μl to 20 μl. This is despite not addressing a modification to the protocol for producing CFEs or the benefits that such a change would bring.

3.3.2.2 The successful reconstitution of a cell-free system to study CRISPR-directed editing *in vitro*.

PCR screening of colonies, following transformation with DNA recovered from integration assays, appears to show successful insertion of the template into pUC19 with a heavy bias towards Cas12a. This is unsurprising given that Cas12a generates overhangs upon cutting DNA, which would have been complementary to ssODNs and dsDNA templates. This likely provides an advantage in this system relative to the blunt cuts generated by Cas9.

Interestingly, our observation of putative integration at the site of DSBs generated by Cas9 contrasts with the findings of Sansbury et al.[342]. This does however highlight a limitation to the findings of this work. While control assays worked as expected, leading to high confidence in successful integration, without sequencing there is no definitive answer. Sansbury

et al. [342] sequenced 18 colonies from their own Cas9 screen and were unable to identify a single instance of successful integration. This may partially explain the observed differences in amplification during colony PCR, as colonies which underwent mutation rather than integration may still appear white during blue-white screening.

One concern for this system was that the original publication includes no data demonstrating that integration is CFE-dependent. As such control assays were included to assess background mutation and ensure that phenomena such as ligation-independent cloning, which can be used to insert template sequence into plasmids in *E. coli* were not responsible for integration events. While controls to capture background mutation worked as expected, insufficient controls were included to assess CFE-dependent integration. As such, questions remain about the success of the system. In future, additional controls would be prudent, including a complete integration assay minus the CFE supplement, as well as assays containing HindIII digested DNA, plus ssODN, plus and minus T4 ligase. These should simulate a typical ligation-independent cloning reaction and be sufficient to provide answers about the CFE-dependence of integration assays.

3.3.3 Exploring a role for HelQ in the resolution of DNA:RNA hybrids

While conducting research into the HelQ-mediated removal of dCas9 roadblocks, we also became interested in a role for the protein in the resolution of R-loops. These structures are a source of genome instability due to their ability to cause replication fork stalling; a structure with which HelQ is known to interact [229]. Published research has not yet established a preference for the protein between DNA and RNA, but PolQ, a polymerase possessing a helicase domain sharing high sequence homology with HelQ, has been demonstrated to unwind DNA-RNA hybrids [364]. Several papers have also discovered roles for HelQ-associated proteins. A 2017 study iden-

tified RPA as a sensor of R-loops which recruits and stimulates the activity of RNaseH1[374]. Other work has also identified activation of the FA repair pathway and monoubiquitylation of the FANCI-FANCD2 complex in the presence of R-loop substrates[375,376].

3.3.3.1 HelQ does not convincingly unwind synthetic DNA:RNA hybrid structures

Using unwinding assays, we performed a biochemical analysis of the activity of HelQ on DNA:RNA hybrid substrates, including those that resembled structures such as R-loops. The results for forked substrates (Fig. 3.17) appear to show no strong preference for unwinding between DNA and RNA strands. Only a slight reduction in unwinding is visible for DNA:RNA substrates with a 5′ fork on the labelled strand. However, this result is limited by substrate design as all forks used in Fig. 3.17 possess an exposed 3′ end on the unlabelled strand, meaning that unwinding of the DNA:RNA substrates is likely contributed to, if not wholly accomplished in the case of substrate C, through an interaction with the DNA strand.

These assays also lack a DNA:RNA fork akin to substrate A (Fig 3.17), in which the RNA strand possesses an exposed ssDNA 3′ end. This would have been an important additional sample, providing insight not only into whether HelQ could unwind a substrate using RNA, but also whether this could be done with equal efficacy to the DNA control fork. This has however been demonstrated in recent research where HelQ was shown to unwind a DNA:RNA fork with equal efficacy to the DNA-only equivalent[229]. This suggests that further experiments using appropriately designed substrates may be able to provide clearer answers.

There is a more noticeable difference in overall unwinding between loop substrates with substrate F, which possesses a ssDNA 5′ end, is dissociated far less efficiently than substrates E (no free ssDNA end) and G (3′ ssDNA

end). Among the DNA substrates, this is in line with expectations given the polarity of HelQ. However, these substrates may also be limited by design as the presence of exposed ssDNA on the displaced strand, may facilitate a secondary reaction in which the two parental strands of the loop are separated. This would then generate an exposed 3′ end to be subsequently unwound by HelQ, releasing the labelled strand and giving a false-positive result.

In both sets of assays (Fig. 3.17–3.18), there was a large amount of degradation observed for both forked and looped substrates which possessed a strand with a labelled ssDNA 5′ end. This was observed across assays carried out in duplicate for both DNA and DNA:RNA substrates. It is unknown why this was the case, but must be accounted for when considering the relative signal from which measurements were taken.

3.3.3.2 HelQ depletion leads to an accumulation of S9.6 signal in nuclei

In parallel to biochemical assays using synthetic substrates, we also attempted to determine a role for HelQ in DNA:RNA hybrid resolution in cell-based models. Using U2OS-derived HelQ$^{-/-}$ cells (5G6) we observed in nuclei the accumulation of DNA:RNA hybrids, as signalled by the antibody S9.6, relative to the parental line. This effect was greatly enhanced when comparing the parental U2OS cells to 5G6 expressing an ATPase mutant, EGFP-HelQ^{K365M}. This could begin to suggest a role for HelQ in the resolution of DNA:RNA hybrids in cells, although it is not sufficient evidence on its own. The data would be more convincing were an additional experiment included in which wild-type HelQ was expressed in the knockout 5G6 cell line. Depletion of the S9.6 signal due to such a rescue of HelQ expression would strengthen the case for the observed effect being caused by absence of the protein.

When conducting this work, DNA:RNA hybrids were identified by prob-

ing with the S9.6 antibody, which has seen widespread use in similar research[377]. However, accumulating evidence suggests that the antibody is not as reliable as previously thought. Multiple studies have shown that the antibody binds not only DNA:RNA hybrids but also to ss RNA and dsRNA[378,379]. Recently published research has identified that this lack of specificity results in the majority of S9.6 signal to come from the ribosomal RNA (rRNA) rather than DNA:RNA hybrids[380]. Additional work has also revealed that the affinity of the antibody for hybrids is not solely structure specific, but also relies on local sequence composition[378].

It is too early to say whether or not this work provides insight into a possible role for HelQ in DNA:RNA hybrid resolution. While depletion of the protein and complementation with an ATPase defective mutant correlate with an accumulation of S9.6 signal, it is unclear what this means. Given the reported issues with specificity for the antibody, it is not necessarily DNA:RNA hybrids which have been upregulated, but could instead be ssRNA or dsRNA of unknown function. Alongside this, it is still poorly understood how hybrid structures such as R-loops contribute to DNA-damage and repair, including whether they are a cause or consequence of genome instability. Multiple studies have shown that R-loops can be causative of replication stress in the form of TRCs[346,358], however research has also identified that hybrids can form at the sites of broken DNA ends when long non-coding RNAs are transcribed at the site of the break and appear to act in recruiting damage-response proteins[381,382].

3.4   Future Perspectives

3.4.1   Further exploration of a role for HelQ in efficient integration of ssODNs

This work failed to demonstrate the HelQ-mediated removal of a dCas9 R-loop from short synthetic substrates, or supercoiled plasmid DNA. How-

ever, roadblock displacement is only a single stage of the pathway to full DNA repair following cleavage by a Cas-nuclease. HelQ has already been associated with repair during editing through the work of[212], and as such it is likely that we haven't yet explored the right conditions. Recent work demonstrating Pif1-mediated replisome bypass of a dCas9 R-loop may be able to inform future experimental development, utilising an assay containing components of the replisome, or FA complex to better discern the role of HelQ in editing-based repair[370].

3.4.2 Utilising cell-free extracts to study the role of HelQ in gene-editing and DNA repair

This work set out to reconstitute a cell-free system for modelling CRISPR-mediated template DNA integration and use it to study the impact of HelQ on editing *in vitro*. While this goal was ultimately not achieved, there is much to be gained from developing this system further.

The most basic development would be the use of sequencing to confirm whether integration events truly occurred and if they were perfect, or resulted in mutations. The model could also be developed further to achieve the integration of larger template DNA. The system, as published, utilises short homology regions of 5 nt, but for an experimentally relevant ssODN this would typically be up to 90 nt. This work had been planned for, but was also unable to be completed (Table.3.2).

Outside of the CFE itself, the reporter used in the assay may prove a point for development. Multiple CRISPR-focussed studies in human cell lines have utilised an assay in which a single codon change converts BFP to GFP[212]. Translating this reporter into the CFE system, which in this work used blue-white screening, could prove powerful as it would rely on perfect integrations to produce a result and could be coupled to high-throughput analytical techniques such as flow-cytometry.

Relevant to this work, the rescue of gene editing efficiency could be explored by supplementing *in vitro* integration assays containing HelQ-deficient CFE with recombinant HelQ. This also raises the possibility of supplementation with mutant variants of the protein, which could lead to a deeper understanding of the mechanisms to which HelQ may be a crucial contributor.

Finally, while the system can model DSB-repair and to some extent assess its efficiency, the separation of cleavage and integration reactions means that information about any direct interactions between Cas-proteins and DNA repair enzymes is lost. The current protocol could be developed into a parallel assay in which cleavage and integration are combined, resulting in a complete system. This might then provide a more physiologically relevant environment which could, for instance, be applied to roadblock assays such as those carried out in Sections 3.2.3–3.2.4.

This CFE-based system for integration is versatile and could be applied to any protein-of-interest, provided that a knockout cell-line was available. This could facilitate the study of an array of DNA-repair proteins similar, in a way, to other cell-free systems which have reconstituted *in vitro*, among other things, the eukaryotic replisome[383].

3.4.3 Complementary approaches to assist in the study of DNA repair proteins during gene-editing

Alongside work utilising the CFE system, complementary research could be carried out using human cell models. As stated in Section 3.4.2, a BFP-GFP reporter assay has been used in CRISPR-focused research to determine the efficiency of donor DNA integration. Similar reporter assays have been applied in this way for a variety of applications, including the homologous recombination assays for ICL repair[384]. As with the CFE system, it could be prudent to explore a range of HelQ mutants in HelQ-deficient cell lines.

Table 3.2: Substrates to model ssODN integration in cell-free extracts.

Name	Sequence (5′–3′)	Details / use
oAC61	T*GACTGGGAAAACCCT GGCGTTACCCAACTGCG GCCGCAATAATCGCCTT GCAGCACATCCCCCTTT CG*C	ssODN for Cas9 integration into pUC19 LacZa
oAC64	G*GGTTTTCCCAGTCACG ACGTTGTAAAACGTTGC GGCCGCACGGCCAGTG AATTCGAGCTCGGTACC CG*G	ssODN for AsCas12a integration into pUC19 LacZa

* denotes a phosphorothioate linkage.

3.4.4 Exploring a role for HelQ in DNA:RNA hybrid resolution

This work demonstrated some preliminary evidence of DNA:RNA hybrid accumulation in cells either depleted for, or containing inactive, HelQ. A notable omission from this work was the staining of 5G6Q cells (HelQ$^{-/-}$ complemented with EGFP-HelQ) to see if the phenotype was rescued. To develop this work further, repeat assay would need to be done including this potential rescue.

The S9.6 antibody has been shown in multiple studies to be less specific to DNA:RNA hybrids than previously thought. The continued use of it in this work would require a series of tight controls alongside the experiments to ensure that hybrid signal is being detected. Work by Smolka et al.[380] has suggested that pre-treatment of the samples with RNase T1 and RNase III, which degrade specifically ssRNA and dsRNA, respectively, can remove the majority of false, background signal from imaging. Utilising these, with RNase H1, which is specific to hybrids, as a control would be crucial to ensuring that this work makes the correct observations.

Further experiments using different antibody probes could also support

this work. One example would be to probe against m6a, a modification seen to accumulate on R-loops during G_2/M; or $\gamma H2aX$, a known marker for DNA damage[345]. Further experimentation in this area was continued by Hannah Betts in the Soultanas group.

3.5   Chapter Summary

This chapter attempted to address the role of HelQ in successful gene editing and the resolution of DNA:RNA hybrids. To achieve this, *in vitro* biochemical assays were combined with cell-free systems and cell-biology approaches. The progress made during this project only partially addressed the original aims. The work successfully established the inability of HelQ to remove Cas-protein roadblocks from DNA, but was less successful in establishing a more general role for the protein in the resolution of DNA:RNA hybrids. Moreover, while the reconstitution of a cell-free system for modelling donor DNA integration during appears to have been successful, circumstances meant that it was unfortunately not applied to the study of HelQ. Overall, this work has demonstrated:

- That HelQ is unable to remove Cas-protein roadblocks from DNA.

- The successful reconstituting of a cell-free system for modelling Cas-protein mediated integration of donor DNA.

- A correlation between HelQ-deficiency and the accumulation of DNA:RNA hybrids in the nuclei of human cells.

4

A functional genomics approach to study the role of HelQ in human cell lines

4.1 Introduction

HelQ has been implicated as a participant in multiple DNA-repair pathways, including HR and FA, with suspected roles in MMEJ and SDSA[225,229]. Supported by biochemical assays, research in human cell lines has made several key observations into the interactions of HelQ during DNA-repair. This includes the colocalisation of HelQ at replication site with RPA foci, as well as at the sites of DNA damage with Rad51 and FANCD2, suggestive of fork restart and induced HR[226]. Subsequent studies have also identified that HelQ-depletion in cells results in sensitivity to ICL-inducing agents such as cisplatin, as well as an increase in chromatid breaks and radial chromosomes Takata et al.[227], Adelman et al.[228].

The aberrant expression of HelQ has been implicated in the development of cancers due to increased genome instability[255–257]. Several papers have identified the protein as a biomarker for the prognosis and likelihood of reoccurrence of multiple cancers included those of the ovary, testes, head, and neck[258,259]. HelQ has also been associated directly with cancer progression and treatment, being observed in a sub-set of ovarian tumours to generate resistance to chemotherapuetic treatment with cisplatin[260].

4.2 Aims and Objectives

Multiple associations with cancer development and progression, as well as a role in resistance to common chemotherapies, make HelQ a clear target for further research. Functional genomics represents a powerful toolset to gain insight into important cellular interactions. By studying the impact of toxic agents on HelQ-deficient or mutated cells, the pathways and interactions that rely on the protein may be elucidated.

This chapter begins to lay the groundwork for a knockout-rescue study of HelQ in two different cell lines and several derived knockout and mutant clones. The objectives of this work were:

- To assess the impact of HelQ depletion on cell lines relative to parental cell lines with and without damage generated by toxic agents.

- To observe whether HelQ complementation in KO cells could rescue the phenotype and restore typical function

- To develop a suite of HelQ mutant plasmids for further complementation studies and use these to determine their impact on cell health.

4.3 Results

4.3.1 Characterisation of HelQ-edited cell lines

4.3.1.1 HelQ KO and mutant genotypes

The impact of HelQ in DNA damage and repair was explored using two different sets of cell lines as described in Table 2.5. The first set of cells was derived from U2OS cells and was obtained from the Wood lab as published in Takata et al.[227]. This set was comprised of the parental U2OS cell line, a homozygous knockout (KO, 5G6) and two cell lines derived from 5G6 stably expressing GFP-tagged versions of WT-HelQ (5G6-Q) and an ATPase

deficient mutant (*HELQ^{K365M}*, 5G6-W), respectively. The 5G6 KO cell line was produced using ZFNs (Fig. 4.1), generating a homozygous 40nt deletion which resulted in an early frameshift in the protein[227].
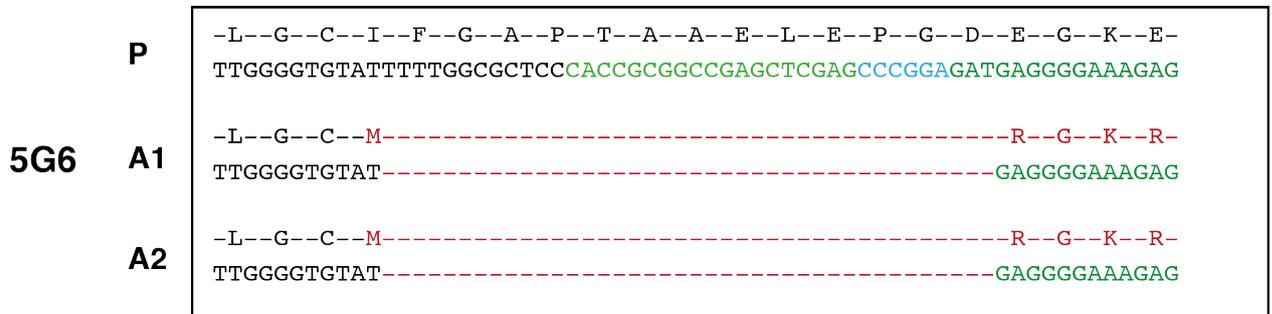
Figure 4.1: Genotype information for a *HELQ^{-/-}* U2OS cell line. Knockouts were generated from U2OS cells using a targeted ZFN. The target region is highlighted in green and the cut site in blue. The outcome was a 40 nt deletion in both alleles, shown in red, resulting in a frame shift. Parental sequence (P); Allele 1 (A1); Allele 2 (A2). Genotype information was obtained from the original publication by Takata et al.[227].

The second cell line used in this work was originally purchased from Horizon Discovery and was obtained from Nanna Therapeutics for this work. This set of cell lines comprised the parental RKO cell line, two clones containing a homozygous *HELQ^{-/-}* deletion (R-101, R-172), and two cell lines containing a homozygous *HELQ^{D463A/D463A}* mutation.

Sequence data shows that the two *HELQ^{-/-}* clones differ in their editing outcome (Fig. 4.2, Table 4.1). Clone R-101 possesses a net 7 bp deletion, causing a frameshift on one allele, while the other contains an 8 bp deletion which produces a stop codon, leading to early termination of protein expression. Clone R-172 on the other hand has a 1 bp In/Del on each allele, both of which encode stop codons, leading to early termination of protein expression. Sequence data for the clones R-DA339 and R-DA93, which possess the *HELQ^{D463A/D463A}* mutation can be seen to both contain the same two homozygous edits on both alleles.

Figure 4.2: Genotype information for *HELQ⁻/⁻* and *HELQ^D463A/D463A* RKO cell lines. Knockouts were generated from RKO cells using Cas9. Point mutations were induced using an ssODN template. The Cas9 binding region is highlighted in green and the PAM site in blue. The outcomes of editing are shown in red and silent mutations in orange. Parental sequence (P); Allele 1 (A1); Allele 2 (A2).

Table 4.1: Genotype information for RKO-derived cell lines

Cell line	Genotype	Allele	Modification
R-101	HELQ (-/-)	1	10 bp deletion + 3 bp insertion
		2	8 bp deletion
R-172	HELQ (-/-)	1	1 bp deletion
		2	1 bp insertion
R-DA339	HELQ (D463A/D463A)	1	D463A (GAC>GCC point mutation, V462V (GTA>GTC) mutation
		2	8 bp deletion
R-DA93	HELQ (D463A/D463A)	1	D463A (GAC>GCC point mutation, V462V (GTA>GTC) mutation
		2	8 bp deletion

4.3.1.2 Morphology and growth of HelQ KO cell lines

The depletion or deletion of genes can have an impact on overall cell health, manifesting as changes to the rate of proliferation or morphology. We began our characterisation of HelQ by monitoring the growth and morphology of cells which either had HelQ knocked out, or possessed mutant copies of the gene.

Cell growth for U2OS cell lines was monitored for 7 days (Fig. 4.3A). Compared to the wild-type U2OS cells, 5G6 ($HELQ^{-/-}$) and 5G6Q ($HELQ^{-/-}$, complemented with EGFP-HelQ) showed little difference in growth with only a slight reduction in signal as observed by WST-1 assay. Interestingly, the 5G6W cell line ($HELQ^{-/-}$, complemented with EGFP-HelQ^{K365M}) displayed significantly reduced growth, producing a fold-change in growth of -0.29 compared to the other three cell lines.

Separately to growth assays, cells were seeded in 6-well plates to observe their morphology (Fig. 4.3B). Images taken at 10× magnification did not

display any significant differences in cell morphology, with all three 5G6 cell lines appearing the same as the wild-type. The detection of EGFP signal from both of the complemented cell lines, 5G6Q and 5G6W, was surprisingly low however and may be indicative of the removal of the stably transfected EGFP-HelQ-expressing cassette from the genome.

Cell growth for RKO cell lines was monitored for 9 days (Fig. 4.4A). Compared to the parental RKO cell line, the clones R-101 (HelQ $^{-/-}$) and R-DA93 (HelQ $^{D463A/D463A}$) saw a slight, but not significant decrease in growth rate. A more severe reduction in proliferation was seen in the clones R-172(HelQ $^{-/-}$) and R-DA339 (HelQ $^{D463A/D463A}$), which displayed a fold-reduction in growth of -0.12 and -0.18, respectively, after 9 days.

Separately to growth assays, cells were seeded in 6-well plates to observe their morphology (Fig. 4.4B). Images taken at $10\times$ magnification did not display any significant differences in morphology for cell lines R-101, R-DA339, or R-DA93 (both HelQ $^{D463A/D463A}$) compared to the parental cell line. The clone R-172 (HelQ $^{-/-}$) seems to present a mildly altered morphology compared to the parental cell line, with multiple cells appearing irregular in shape.

4.3.2   Tolerance of HelQ KO cell lines to DMSO

The next step in establishing a phenotype for HelQ depletion was to monitor cell response to DNA damaging agents. In some cases this required solubilisation in DMSO. As this is toxic to cells even at low concentrations, the impact of the solvent on cell viability was assessed against all cell lines to avoid false-positive signal. Data were normalised to 100 % against control wells containing no DMSO.

Among the U2OS cell lines, the results show a uniform decline in cell viability between 0.5–1 % (v/v) DMSO, increasing in severity with solvent concentration. At the maximum of 5% DMSO, viability fell below 60% for all
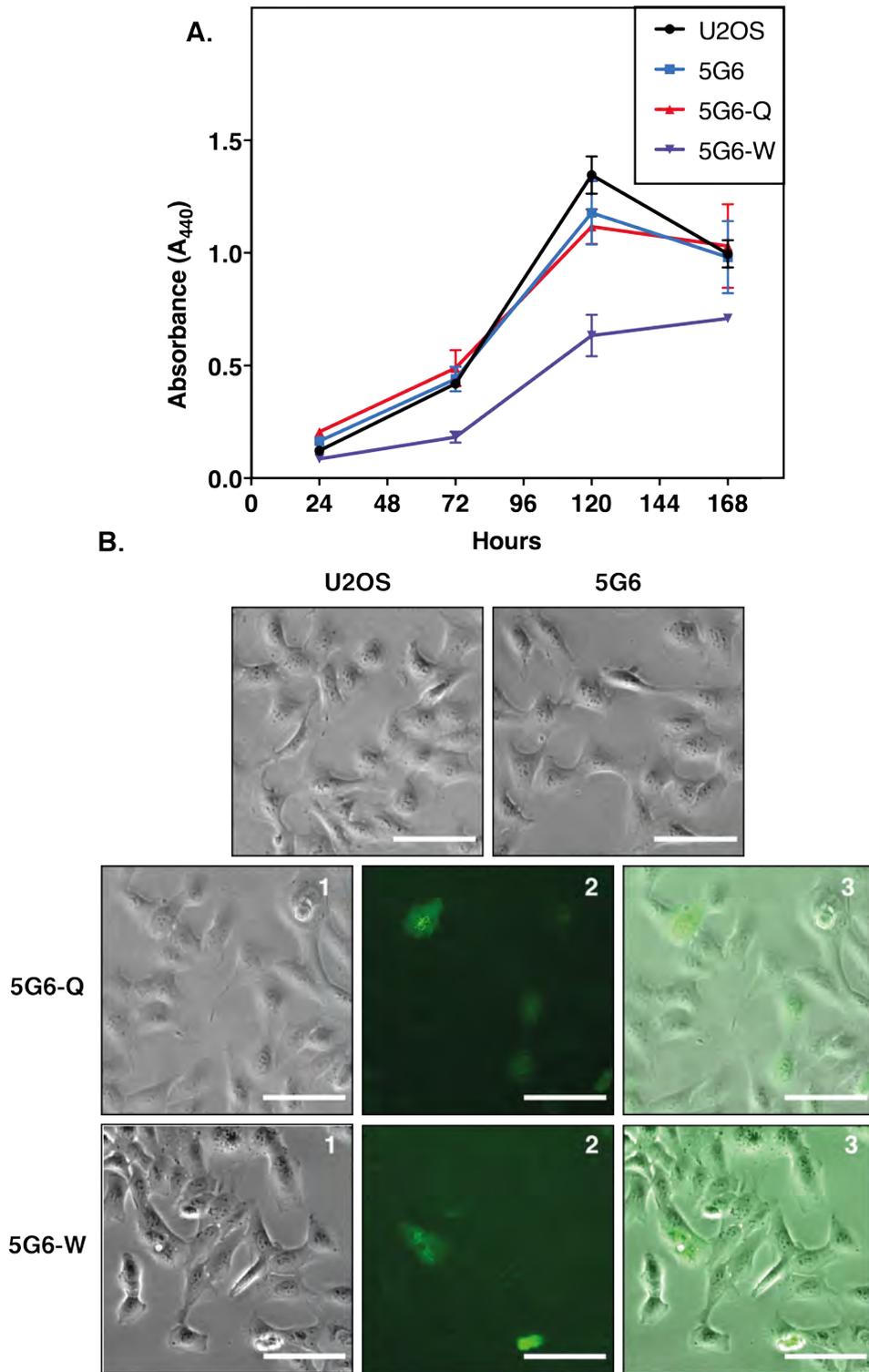
Figure 4.3: Growth and cell morphology comparison of U2OS and HelQ [-/-] **cell lines.** (A) Growth curves for U2OS cell lines. Cells were cultured for 7 days with proliferation being measured every 48 hours using the WST-1 assay as a measure of growth. Timepoints were plotted as mean values, with error bars representing standard deviation for assays performed in duplicate. (B). Cell size and morphology of U2OS and derived cell lines. Box 1, phase contrast; Box 2, fluorescence; Box 3, composite image. All scale bars, 100 μm.
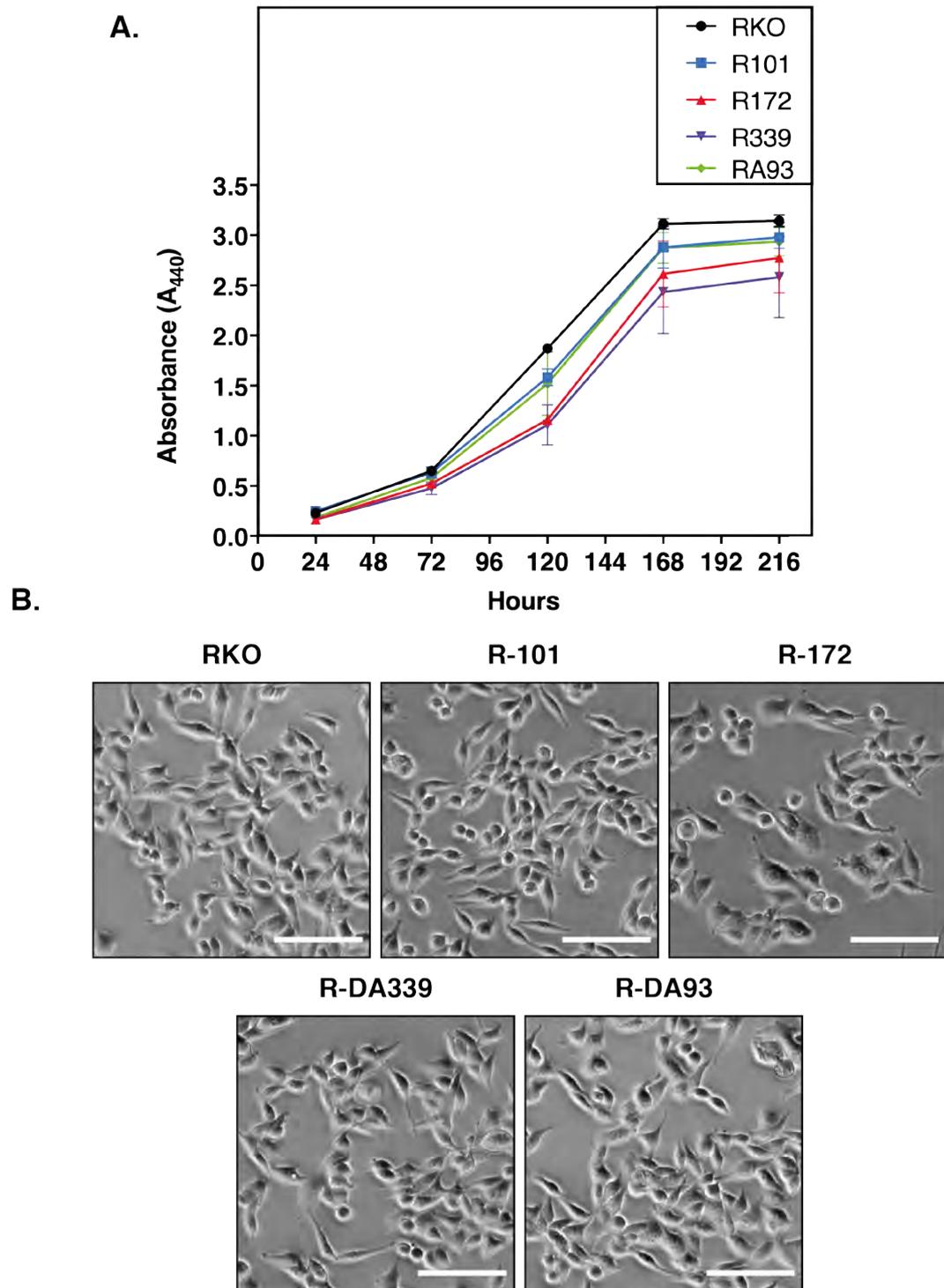
Figure 4.4: Growth and cell morphology comparison of RKO, HelQ [-/-] and HelQ [D463A/D463A] cell lines. (A) Growth curves for RKO cell lines. Cells were cultured for 9 days with proliferation being measured every 48 hours using the WST-1 assay as a measure of growth. Timepoints were plotted as mean values, with error bars representing standard deviation for assays performed in duplicate. (B). Cell size and morphology of RKO and derived cell lines. All scale bars, 100 μm.

four cell lines. Among the RKO cell lines, the results show a sharp decrease in viability beginning from DMSO concentrations of 1%. At the maximum concentration of 5% DMSO all five of the cell lines displayed a viability below 35%. The HelQ-deficient clone R-172 appears to sensitive to lower concentrations of DMSO, experiencing a decline in viability between 0.5–1%. Following these experiments, the total concentration of DMSO per well was set at a maximum of 0.5%.
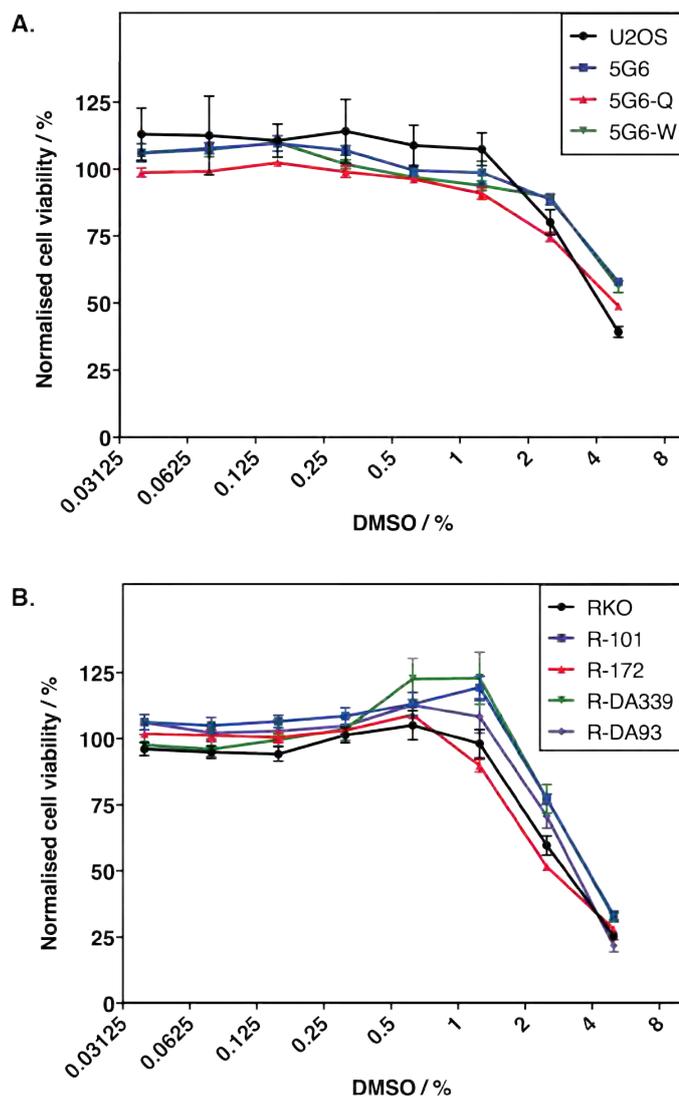
Figure 4.5: Tolerance of cell lines to increasing concentrations of DMSO.
(A) Cell viability of U2OS and derived cell lines in the presence of DMSO.
5G6, HelQ $^{-/-}$; 5G6-Q, 5G6 stably transfected with EGFP-HelQ; 5G6-W, 5G6
stably transfected with EGFP-HELQ$_{K365M}$. (B) Cell viability of RKO and
derived cell lines. R-101 and R-172, HelQ$^{-/-}$ clones; R-DA339 and R-DA93,
HelQ$^{D463A/D463A}$ mutants. Cells were incubated with DMSO for 24 hours
before measuring viability. All assays were carried out in duplicate with
both technical and biological replicates. Error bars show the standard error
of the mean.

4.3.3 Response of HelQ KO cell lines to toxic agents

Having established a maximum tolerable assay concentration of DMSO, we
next assessed the impact of DNA-damaging agents on the proliferation of
U2OS and RKO cell lines. Cells were seeded in 96-well plates with media
containing DNA-damaging agents as indicated in Fig. 4.6 (U2OS) and Fig.

4.7 (RKO). Data for all cell lines were normalised against a no-drug control.

HelQ-deficient U2OS cells were observed to show increasing differences in viability at high concentrations of mitomycin C (MMC) and cisplatin (Fig. 4.6). Interestingly, in the case of MMC, the homozygous knockout 5G6 and EGFP-HelQ complemented 5G6Q displayed a greater tolerance for high concentrations of the drug than the parental cell line or the ATPase mutant-complemented 5G6W. There was no apparent difference in sensitivity to either aphidicolin or hydroxyurea.

In contrast, HelQ-deficient RKO cells did not show any apparent sensitisation to cisplatin, aphidicolin or hydroxyurea when compared to the parental cell line (Fig. 4.7). Cells treated with MMC appeared to be displaying differences in their tolerance to the drug between concentrations of 37.5–75 μM, however this effect plateaued up to concentrations of 300 μM.
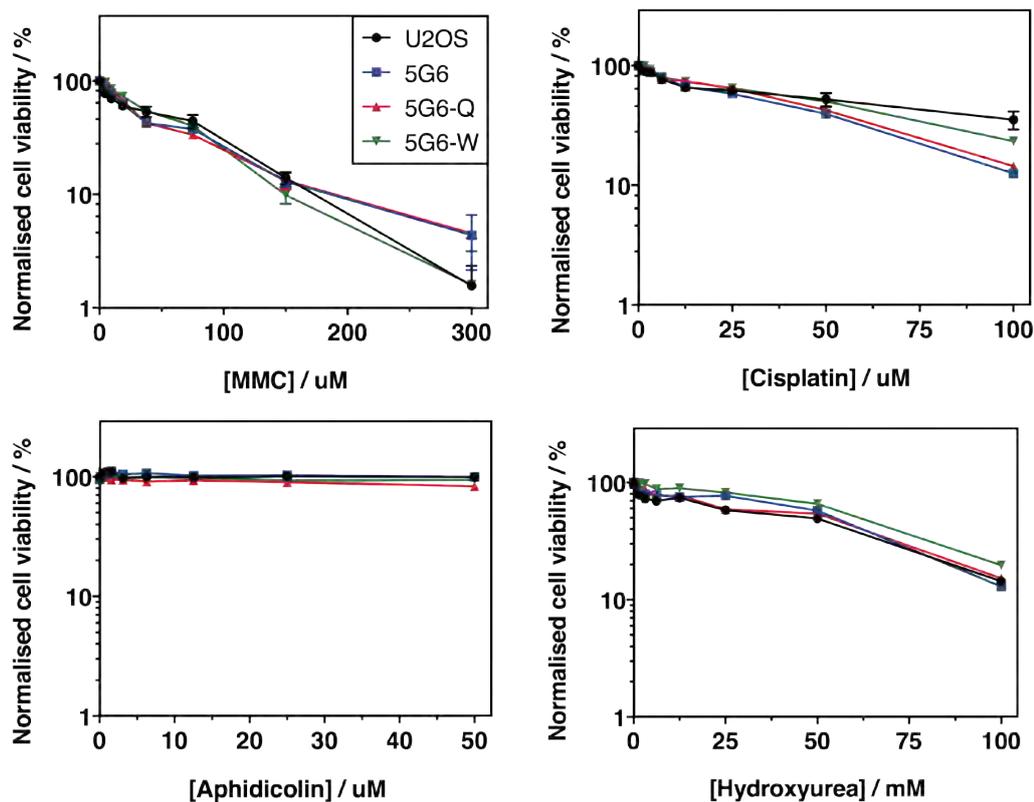
Figure 4.6: Impact of DNA-damaging agents on the proliferation of U2OS and derived HelQ KO cell lines. Cells were incubated with DNA-damaging agents at the indicated dosages for 24 hours before measuring absorbance from the WST-1 assay. Data were normalised against controls not treated with the indicated agents. All assays were carried out in duplicate. Error bars show the standard deviation. MMC, Mitomycin C.
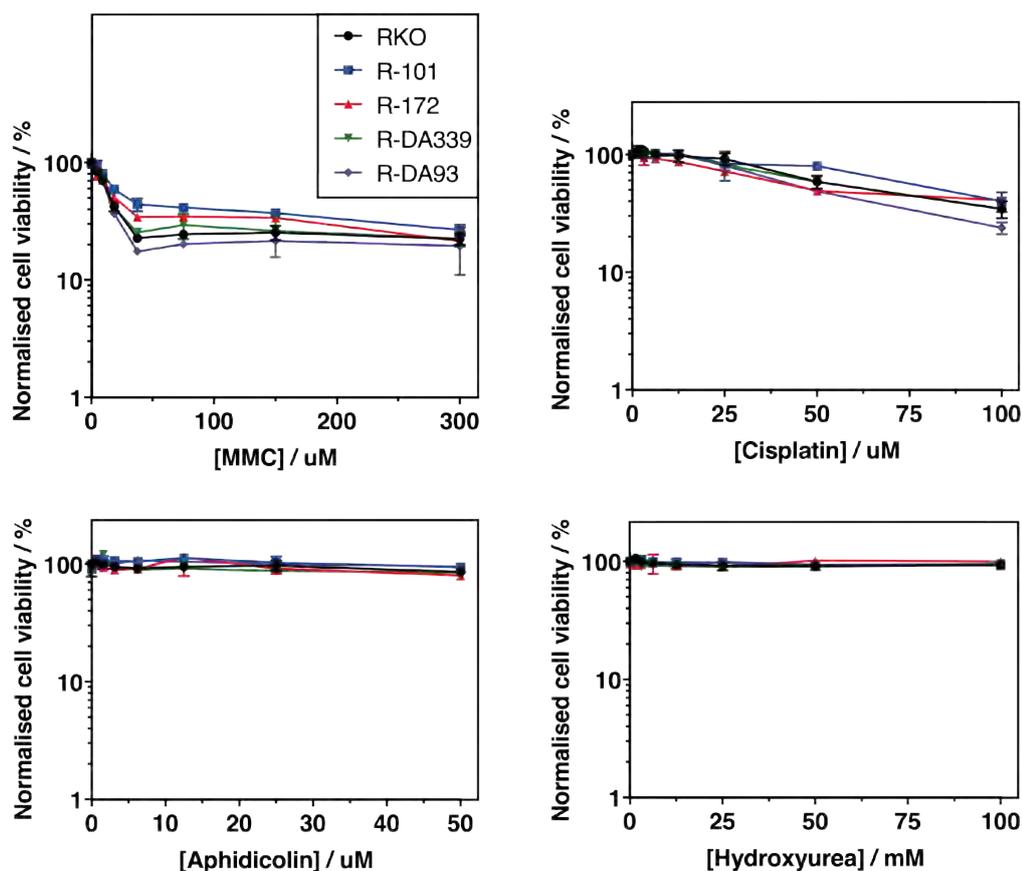
Figure 4.7: Impact of DNA-damaging agents on the proliferation of RKO and derived HelQ KO and mutant cell lines. Cells were incubated with DNA damaging agents at the indicated dosages for 24 hours before measuring absorbance from the WST-1 assay. Data were normalised against controls not treated with the indicated agents. All assays were carried out in duplicate. Error bars show the standard deviation. MMC, Mitomycin C.

4.4 Discussion

4.4.1 The impact of HelQ KO on cell morphology and growth

This study began by characterising the proliferation of each cell line as previous studies have not indicated a growth phenotype for HelQ-depleted or mutated cells. For the majority of mutant clones, while overall slower growth was measured, no significant difference in proliferation was observed when compared to the parental cell line. This could be indicative either of a mild

impact to growth, or of no true proliferation phenotype for HelQ-depletion. While a growth phenotype in untreated cells is not directly reported in the literature, it could be feasible given the information available. Adelman et al.[228] demonstrated that replication fork extension rates were significantly lowered in HelQ$^{-/-}$ cells, which could lead to modest increases in proliferation time.

The exception to the observed growth patterns was cell line 5G6-W, which expressed the Walker A-compromised mutant HelQ^{K365M}. It is possible that this produces a more severe reduction in proliferation due to HelQ itself effectively becoming a roadblock, as it binds to DNA but is unable to translocate along it. Previous work may support this as it has been shown that a mutation to the Walker B domain (D463A), did not alter the DNA binding activity of HelQ *in vitro*[229]. It should be noted however, that the same phenotype was not observed for Walker B mutants R-DA339 and R-DA93 in the RKO-derived cell lines, which might be expected to perform similarly.

Each of these assays was only performed in duplicate and as such there may be insufficient data to draw conclusions. Alongside this, more assay development may be required to be certain of current findings. For instance, while both R-172 and R-DA339 are displayed as the slowest-growing cell lines, the severity of the measured phenotype is inconsistent with observations made during routine cell culture. Coupled with the change in morphology observed in R-172, this could be indicative of limitations in the assay readout used to measure cell viability, as will be further discussed in Section 4.4.3.

4.4.2 The response of cell lines to toxic agents

In order to further characterise the impact of HelQ-depletion in cells, we first aimed to reconstitute the published phenotype of sensitivity to ICL-inducing

agents seen by Takata et al. [227] and Adelman et al. [228]. Alongside this we also screened against known inhibitors of DNA synthesis/replication and repair, hydroxyurea and aphidicolin. Neither of these two DNA-damaging agents produced an effect in either cell line, which in the case of the former, is consistent with previous reports [227].

While at higher concentrations of 300 μM MMC and 100 μM cisplatin, a phenotype did appear to be emerging in the U2OS set of cells, there are several caveats to the data. First, the concentrations at which these effects became most apparent was twice the maximum dose used by Takata et al. [227]. Alongside this, the phenotype observed appears much milder than the previous study despite this work using the same cell lines. The same cannot be said for the RKO set of cells in which no sensitivity phenotype was visible. The overall lack of an ICL-sensitivity phenotype in this work could be explained by several factors, which will be further explored in Section 4.4.3.

A limitation of this work was the lack of positive control cell lines, which could confirm that both the drugs used and the dosages chosen were effective. Repeats of this work would therefore require the use of cell lines derived from the parental U2OS and RKO which were deleted for genes known to induce drug sensitivity or lethality. The possible controls to be used are summarised in table 4.2.

In its current state, the data cannot begin to suggest a lack of phenotype, which would run contrary to published work; more likely is that the drugs failed to function as intended in the first place. This could be the result of several issues, including a failure to appropriately dissolve the drug compounds, or their precipitation when stored at 4°C, resulting in sub-lethal concentrations. In future close attention to the solvent used and avoiding storage of drug compounds should rectify this. A further possibility is the possible hetergeneous status of test cell lines, which is discussed further in section 4.4.4.

Table 4.2: Positive controls for drug phenotype testing

Drug used	Example Gene KO for control
Mitomycin C	*XPF*[385]
Cisplatin	*XPF*[386]
Hydroxyurea	*CDK6*[387]
Aphidicolin	*CHK1*[388]

4.4.3 A lack of clear phenotype for HelQ-deficient or mutated cell lines

Overall, the screening of HelQ-depleted cells against DNA-damaging agents was not successful in identifying a phenotype. This comes despite multiple studies from different research groups demonstrating sensitivity to ICL-inducing agents in HelQ$^{-/-}$ cell lines[226–228]. There could be multiple reasons for this including unexpected expression of HelQ in the cell lines, or the assay conditions used.

One of the key principles of knockout-generation through targeted gene-editing is to target early exons in order to introduce either nonsense mutations, resulting in a frameshift, or stop codons to halt translation. A 2016 paper reported on the ability of cells to circumvent the effects of nonsense mutations through so-called 'illegitimate translation' in which the mutated early exons are skipped over, and translation resumed from a downstream start codon, which can result in the expression of a truncated version of a given protein[389]. This has been expanded on by subsequent studies which

established that in some cases the truncated proteins can retain catalytic activity[390].

HelQ has been shown to be active when truncated at the N-terminal, with a C-terminal fragment (C-HelQ) exhibiting similar activity to that of the full length protein[229]. The RKO cell lines used in this work included two different clones, R-101 and R-172, with the genotype HelQ[-/-] between which the precise edits differed (Fig. 4.2). Given the lack of a significant viability phenotype between these cell lines and the parental RKO cells, and with the knowledge of C-HelQ activity, it would be tempting to suggest that illegitimate translation could be the cause. While this may be an interesting phenomenon to explore in future, to determine whether HelQ-depletion may be more impactful than previously thought, the inability to reproduce the ICL-sensitivity phenotype observed by Takata et al.[227], using the same cell lines as published, suggests the issue lies elsewhere.

More likely is that the assay conditions used to determine cell viability when treating HelQ with DNA-damaging agents were sub-optimal for observing the desired phenotype. This work utilised only a single approach in determining cell viability, the colorimetric WST-1 assay, in which the tetrazolium salt WST-1 is cleaved to create formazan by cellular mitochondrial dehydrogenases. This measurement of cell viability relies on the fact that a dead cell will have no mitochondrial dehydrogenase activity and as such will produce less formazan. However, this is limited for cytotoxic assays as it is unable to distinguish between cells which are viable, and those in the process of dying. This leads to underestimation of cytotoxicity, as has been noted in multiple studies across a range of cell lines and DNA-damaging agents[391–393].

This underestimation could explain our inability to reproduce the published ICL-sensitivity phenotype for HelQ-depleted cells. The approaches used to determine this previously have ranged from luminometric assays

that measure absolute ATP levels at a fixed endpoint, and clonogenic assays[227,228]. Another method frequently used to assess DNA damage, which could be appropriated as a measure for the impact of HelQ-depletion following treatment, is the comet assay which can be modified to analyse specific lesions such as ICLs[394,395]. Future work should use these methods to replicate the phenotype, exploring other assays which may be more amenable to scaling up.

4.4.4 Issues arising from possible gene-silencing in HelQ KO cell lines

In addition to the preceeding discussion, a further systematic issue likely played a significant role regarding the lack of observable phenotypes from HelQ KO cell lines. Whilst conducting microscopy to capture the images displayed in Figure 4.3, it was observed that EGFP expression in the 5G6-Q and 5G6-W cell lines was much lower than would be anticipated from a constitutively expressing cassette that had been stably integrated. This was overlooked at the time and in hindsight may explain the lack of observable phenotypes, including those already published using these cell lines, such as[227].

While stably transfected cell-lines were cultured in a constant low concentration of G418, as advised by the authors of Takata et al.[227], it is possible that the relative age of the cell lines led to gene-silencing of the expression cassette and thus a mixed population of cells from which no phenotype could be observed. This is a common phenomenon shown to be primarily the result of epigenetic modulation by methylation and histone modifications[396,397]. Future work could circumvent this by the expansion of clonal populations of cells, as will be further discussed in section 4.5.1.

4.5   Future perspectives

This work began to explore the impact of HelQ depletion and mutation in human cell lines, but only scratches the surface of the requirements for in-depth characterisation. Continued assay development, a wider array of DNA-damaging compounds, complementation of HelQ-deficient cells with both the wild-type and mutant copies of the gene, and the expansion of the work into wider-scale screening, would provide a solid platform upon which to further develop our knowledge.

4.5.1   Clonal selection and expansion of U2OS and 5G6-derived cell lines to ensure experimental consistency

One of the major issues of this work was low expression of HelQ, as inferred from levels of fluorescent EGFP tag. The resulting polyclonal population most likely interfered with the results of growth and drug-response assays, confounding the data. This work should therefore be repeated following clonal selection. For EGFP-HelQ-expressing cells this would be achieved by cell sorting based on fluorescence into a 96-well plate, from which new stocks could be expanded and stored in liquid nitrogen. For the non-EGFP expressing cells , U2OS could be sorted and expanded easily as the parental cell line, while 5G6 would be sorted and then the absence of HelQ expression examined by western blot.

4.5.2   Assay development to observe true HelQ-deficiency phenotypes

As discussed in Section 4.4.2, reliance on the WST-1 assay as the sole measurement for sensitivity to DNA-damaging agents comes with several limitations. Chief among these is the possible loss of information regarding the true phenotype of cell lines due to a lack of specificity in distinguishing live and dying cells. In future, the use of different assays that are more specific

to detecting DNA damage, might be better suited to gathering good data. Alongside this, a deeper exploration of dosage and treatment time may be able to give us better insight into the role played by HelQ in DNA damage repair.

4.5.3 Testing the impact of a wide array of toxic agents on HelQ deficient cells

This work briefly explored the impact of four different DNA damaging agents on HelQ-deficient or mutated cell lines, largely informed by previous work which had identified a sensitivity to ICL-inducing agents. Going forward it may be prudent to also test a wide array of damaging agent classes. This could include nucleoside reverse transcriptase inhibitors such as azidothymidine, PARP inhibitors such as Olaparib, or topoisomerase inhibitors such as SN-38. Exposing cells to a wide variety of DNA damage in this way may be able to better develop our understanding of the extent of HelQ involvement in DNA repair. It may also be interesting to explore combination treatments, for example using hydroxyurea with cisplatin, as the two have been shown to function synergistically[398]. This could also contribute considerably to the work discussed later in Chapter 5.

4.5.4 Complementation studies using HelQ deficient cell lines

One of the aims of this work was to attempt to rescue any observed sensitivity phenotypes in HelQ-deficient cells by complementation with the wild-type gene. Alongside this we aimed to produce and study a suite of mutant HelQ plasmids to better inform our understanding of HelQ function in cells and how different mutations may contribute to pathogenic phenotypes (Table 4.3). Unfortunately this was not achieved during the project, but remains an essential set of experiments.

Table 4.3: Proposed mutagenesis of HelQ for further complementation study.

Residue	Feature	Function
D142F143A	PWI-like fold in the N-terminal domain of HelQ	Required for interaction with RPA
V306I	RecA domain	Result of SNP associated with ESCC*[399]
K365M	Walker A motif	Abolishes ATPase and DNA helicase activity
D463A	Walker B motif	Abolishes DNA helicase activity
Y642A	Helicase motif IV	Motif IV thought to be important for translocation[229,400]
F974A	Ratchet domain	Mutation of residue associated with endometrioid carcinoma[401]
Y991XX	HLH domain*	Mutation of residue associated with ampullary carcinoma[402]

* ESCC, oesophageal squamous cell carcinoma; HLH, helix-loop-helix

4.5.5   Synthetic lethal screening against HelQ-deficient cell lines

Studies of HelQ as a singular entity within cells, such as the planned complementation assays, are crucial to our understanding of DNA repair. However this can be enhanced with powerful approaches such as synthetic lethality screens which may inform us on the wider interactions of HelQ in human cells. This would not only develop our knowledge of HelQ, but may also provide new targets for drug development against cancers, as will be further elaborated on in Chapter 5.

4.6   Chapter Summary

The focus of this chapter was on characterising the phenotype of HelQ-deficient cells as well as their response to DNA-damaging agents, with a particular focus on those responsible for the generation of ICLs. While the principle aims were not met, the initial experiments did lay the groundwork for a larger study of HelQ in human cell lines, including complementation assays to study specific mutations of interest.

5

The development of small-molecule inhibitors against HelQ

5.1 Introduction

5.1.1 A role for HelQ in ICL-repair and cancer

Since the discovery of Mus308 from *Drosophila melanogaster* in 1976, the Hel308 family SF2 helicases, including Hel308 and HelQ, have had a significant role in the study of DNA repair[403–405]. Despite significant research, the precise role of HelQ in maintaining genome stability remains unclear[225–229]. Presently, the protein is thought to promote repair at stalled replication forks through HR-related pathways via an as-of-yet unknown mechanism. Several studies have implicated HelQ as a participant in the repair of ICLs, working synergistically with the Fanconi anaemia (FA) pathway[212,227,228,250].

This function indicates a caretaker role for HelQ which could prove significant to the development of cancers. Disruption of protein expression may interfere with correct lesion repair, leading to increased genome instability[255–257]. This has been shown in studies implicating HelQ in the development of cancers of the upper aerodigestive tract, head and neck[406–408]. There is also further potential for the involvement of HelQ in the development of cancers in areas of the body in which expression is highest: the ovaries, heart and skeletal tissue[406].

5.1.2 Validation of proteins as targets for cancer treatments

Many proteins may be predicted as cancer targets based upon the outcomes that changes to their expression would have on genome stability and cell proliferation. However care must be taken to ensure that proteins of interest are valid targets. A valid target can be defined as a gene upon which a tumour relies for growth, proliferation and survival[409]. Determining this requires robust characterisation of the impact disruption of the gene has on tumourigenesis[410].

Small-molecule screens to identify inhibitors against targets are a staple of the pharmaceutical industry and are increasingly common in academic research projects. Typically utilising compounds between 200-300 Da, HTS approaches have led to the successful development of several anti-cancer drugs including Olaparib, a PARP-1 inhibitor identified in 2003 and licensed in 2015 to treat ovarian cancer[411,412]. HTS of small-molecule libraries continues to be relevant; a 2019 study focussed on WRN was able to identify several promising small-molecule candidates for further development[319].

Functional genomics can validate cancer targets by identifying synthetic-lethal genes. First described in *Drosophila* in 1922, this is a cumulative process in which disruption of one gene is compatible with cell survival, but subsequent perturbation of a second gene is not[296,413]. Screens to identify targets that are lethal when combined with known cancer genes provide a powerful tool for validation. Previously limited to loss-of-function screens in model organisms, the rise of CRISPR-Cas as a molecular tool has facilitated large-scale, screening for cancer targets[414,415]. Several studies have used this approach to identify the helicase WRN as a synthetic-lethal drug target in microsatellite-unstable cancers[315–318].

Data generated from high-throughput small-molecule screens and functional genomics approaches also has a secondary use. Repositories of data such as the Cancer Genome Atlas and Cancer Dependency Map can facilitate

predictive models through the use of machine learning, leading to future streamlining of target identification[416,417].

5.1.3 HelQ as a prognostic biomarker in cancers

HelQ has been associated in several recent papers as a prognostic biomarker for the outcomes of several types of cancer[258,259]. A study by Long *et al.* found that alongside POLM, NUDT15, and AEN, overexpression of HelQ could be associated with biochemical recurrence of prostate cancers that have a high risk of progression and metastasis[259]. A study by Guo *et al.* found that HelQ and EGR3 expression were strongly associated with overall survival in cases of chronic lymphocytic leukaemia. In this case however, down-regulated expression of HelQ was associated with poor prognosis[258].

5.1.4 HelQ in chemotherapy-resistant tumours

The participation of HelQ in ICL repair as a synergistic partner to the FA pathway may prove crucial to the treatment of cancers. Several common classes of chemotherapeutic drugs such as cisplatin, carboplatin and mitomycin C function by catalysing the formation of ICLs to kill cancer cells[418,419]. Any up-regulation in ICL-associated repair pathways could therefore lead to reduced efficacy of these treatments.

Previously, HelQ has been associated with ovarian cancer through both genome-wide association studies and bioinformatics[399,408,420,421]. This was directly observed in a 2018 study which implicated HelQ in epithelial ovarian carcinoma. The study observed that overexpression of HelQ in a sub-set of ovarian tumours resulted in resistance to chemotherapuetic treatment with the crosslinking agent Cisplatin[260]. Furthermore, a strong association was found between up-regulated HelQ expression a lower rate of survival in patients presenting this subset of tumours.

5.2   Aims and Objectives

The role of up-regulated HelQ expression in tumour populations resistant to crosslinking agents could potentially be overcome by a dual-drugging approach. This would allow the continued use of existing chemotherapies, such as cisplatin, alongside an inhibitor for HelQ.

This chapter focusses on the screening and development of such inhibitors for HelQ. This was achieved by the screening of a small-molecule library and subsequent testing to characterise early candidates. The objectives of this work were:

- To screen a small-fragment library for potential inhibitors of full-length HelQ (FL-HelQ) .

- To assess the performance of potent inhibitory compounds in additional assays, including impact on DNA binding and ATPase activity.

- To generate and assay the performance of lead compounds using promising candidates as a scaffold.

5.3   Results

5.3.1   A small molecule library for the development of inhibitors against FL-HelQ by screening *in vitro*

To begin development for inhibitors of FL-HelQ, a small-fragment screen was carried out using a commercially available library. This had been pre-screened by the manufacturer for desirable properties such as high aqueous solubility and lack of aggregating compounds (Fig. 5.1). These metrics form the basis of libraries considered to have a high chance of successful hit development. A promising library can be seen to conform to the 'Rule of 3' (Ro3), which is defined as a molecular weight below 300 Da and a number of hydrogen-bond acceptors (HBA) and donors (HBD) less than 3.

Library diversity was assessed by generating frequency distributions (Fig.5.1) and global averages (Table 5.1) for a range of desirable small-fragment properties. These were then used to determine the fit of the library against ideal values for each metric. Property distributions and whole-library averages for both Ro3 and other desirable properties were found to fall within the range for efficient hit detection. Several fragments were also selected which fell outside of the strict definition of these metrics as this has been found to be beneficial in finding suitable candidates[422].
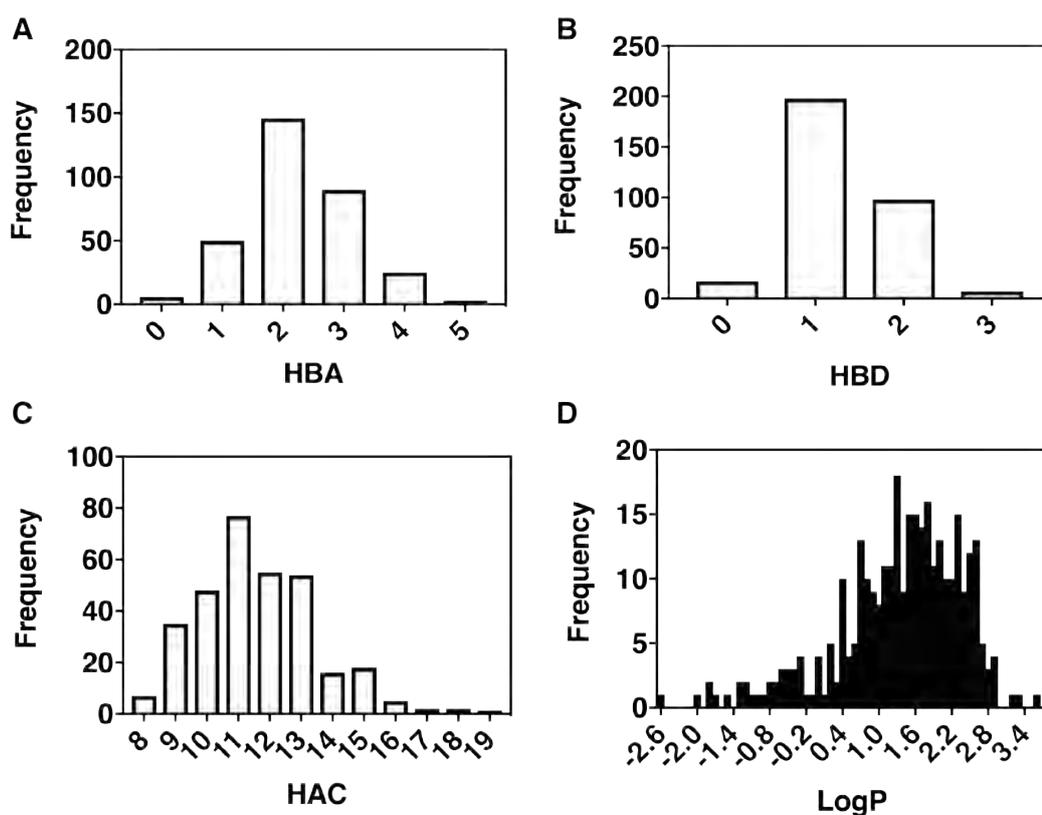
Figure 5.1: Frequency distributions of compound properties in a small-fragment library to be screened for inhibitor candidates The data displayed represent distributions of hydrogen-bond acceptor (**A**) and donor (**B**) atoms, as well as heavy atoms per compound (**C**) and lipophilicity (Log P, (**D**)). Log P values were binned in intervals of 0.1.

Table 5.1: Whole-library mean values for essential small-fragment properties.

Metric	Average value
Molecular weight (MW)	166.466
Log P	1.337
H-bond donors (HBD)	1.297
H-bond acceptors (HBA)	2.272
Log D	-0.038
Topological polar surface area (TPSA)	47.603
Fsp^3	0.169
Heavy atom count (HAC)	11.634
Rotatable Bonds (RTB)	1.269

5.3.2 HelQ is active as a helicase in low-percentage concentrations of DMSO

Library compounds used for the small-fragment screen were dissolved in DMSO. Despite being the standard solvent for the preparation of drug solution stocks, it has previously been observed that it can alter protein properties and lead to anomalous assay results if not accounted for[423].

First, a baseline for unwinding by HelQ was established using a forked DNA substrate. Helicase assays were carried out as described in (Section 2.2.8.3). Alongside this were control assays containing either forked DNA substrate only, or a boiled, fully dissociated substrate. The results of a concentration titration experiment (Fig. 5.2A) show that HelQ is able to unwind the substrate to a mean of 49% completion at a concentration of 80 nM and to 60.5% at 160 nM.

HelQ was subsequently used at 80 nM in a time-course assay to determine the best assay length in which to observe good unwinding. Consistent with the result of the concentration titration assay, HelQ unwound the forked DNA substrate to a mean of 45.5% after 10 minutes, reaching a maximum of 72.2% unwinding after 20 minutes. To ensure sufficient unwinding was

observed, the time for the reaction was set at 10 minutes.
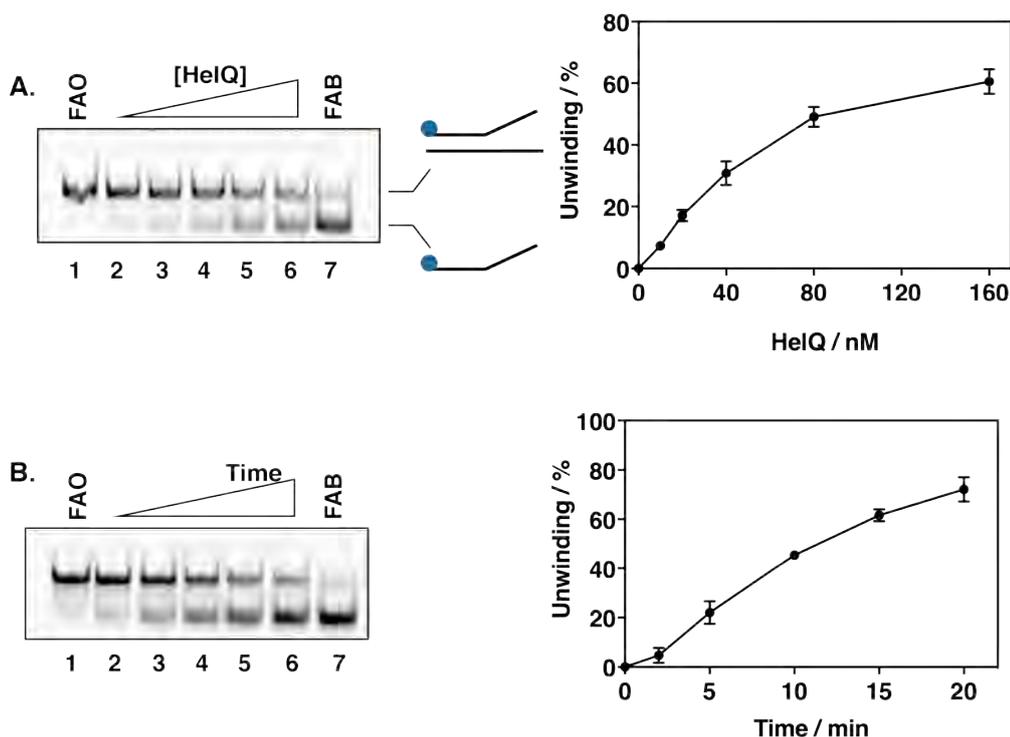
Figure 5.2: Unwinding of a forked DNA substrate by full-length HelQ helicase. Unwinding assays were carried out using a Cy-5 labelled, forked DNA substrate. (A) Unwinding of forked substrate at different concentrations of HelQ. Protein was added to reactions at concentrations of 10, 20, 40, 80, and 160 nM. (B) Unwinding of forked substrate by 80 nM HelQ as a function of time. Assays were stopped at timepoints of 2, 5, 10, 15, and 20 minutes. Unwinding was quantified against a DNA-only control assays (FAO). Also included as a control was a boiled, substrate-only reaction (FAB) to represent full fork dissociation. Error bars represent standard deviation for experiments performed in triplicate.

To eliminate the possibility that inhibition of FL-HelQ helicase activity was caused by DMSO alone, DNA fork unwinding assays were carried out determine the enzyme's tolerance to it. Helicase assays were supplemented with concentrations of DMSO from 0.5–16 % (v/v) (Fig.5.3). The data show an inverse relationship between FL-HelQ activity and increasing concentrations of solvent. Total substrate unwinding was reduced from a mean of 57.7% in reactions lacking DMSO, to 22.0% at the maximum concentration of 16% (v/v).

DNA unwinding by FL-HelQ was minimally impacted up to 4 % DMSO,

as at this concentration the protein retained 85% of relative activity compared to the control reaction. This informed our approach towards future assays. To ensure sufficient unwinding for assay quantification the maximum concentration of DMSO was set at 5%. As stock solutions for compounds were at 100 mM, this percentage of DMSO equates to a maximum concentration of 5 mM inhibitor.
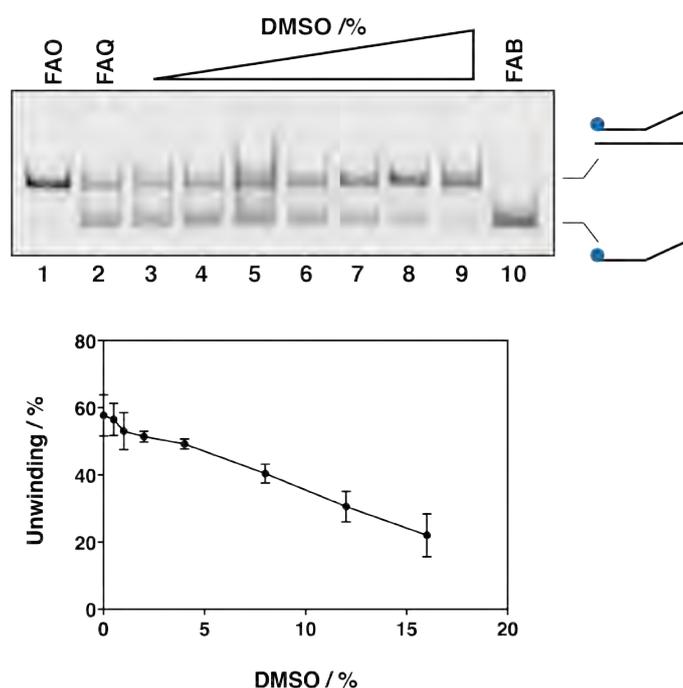
Figure 5.3: Determining the tolerance of FL-HelQ to increasing concentrations of DMSO. Unwinding assays were carried out using a Cy-5 labelled, forked DNA substrate. Reactions were supplemented with DMSO to concentrations of 0.5, 1, 2, 4, 8, 12 and 16% (v/v). Bands were quantified against a control assay lacking DMSO (FAQ). Also included as controls were a substrate-only (FAO) reaction lacking FL-HelQ and a boiled, substrate-only reaction (FAB) to represent full fork dissociation. Error bars represent standard deviation for experiments performed in triplicate.

5.3.3 A small molecule screen for inhibitors of FL-HelQ helicase

To identify initial candidates, FL-HelQ was screened against a small molecule library of 300 compounds, with results obtained for 282 candidates (Fig.5.4). Compounds were assessed for inhibitory capability qualitatively by compar-

ing helicase assays supplemented with 5 mM of each small-fragment against a control assay containing 5 % DMSO. Alongside these assays were controls containing either forked DNA substrate only, or a boiled, fully dissociated substrate (Fig.5.5).

The screen identified a total of 54 putative inhibitor candidates (Fig.5.5, red), equating to 19% of the total compounds tested. Candidates were then analysed based upon structural characteristics and chemical properties (see Fig. 5.1 and Table. 5.1) by Nanna Therapeutics (Cambridge, UK). Inhibitors containing known promiscuous structures such as those prone to aggregation, and/or functional groups, including those similar to ATP, were removed. This approach allowed us to focus on inhibitors that are potentially novel and specific for targeting of FL-HelQ.

Plate-01

Plate-02

Plate-03

Plate-04
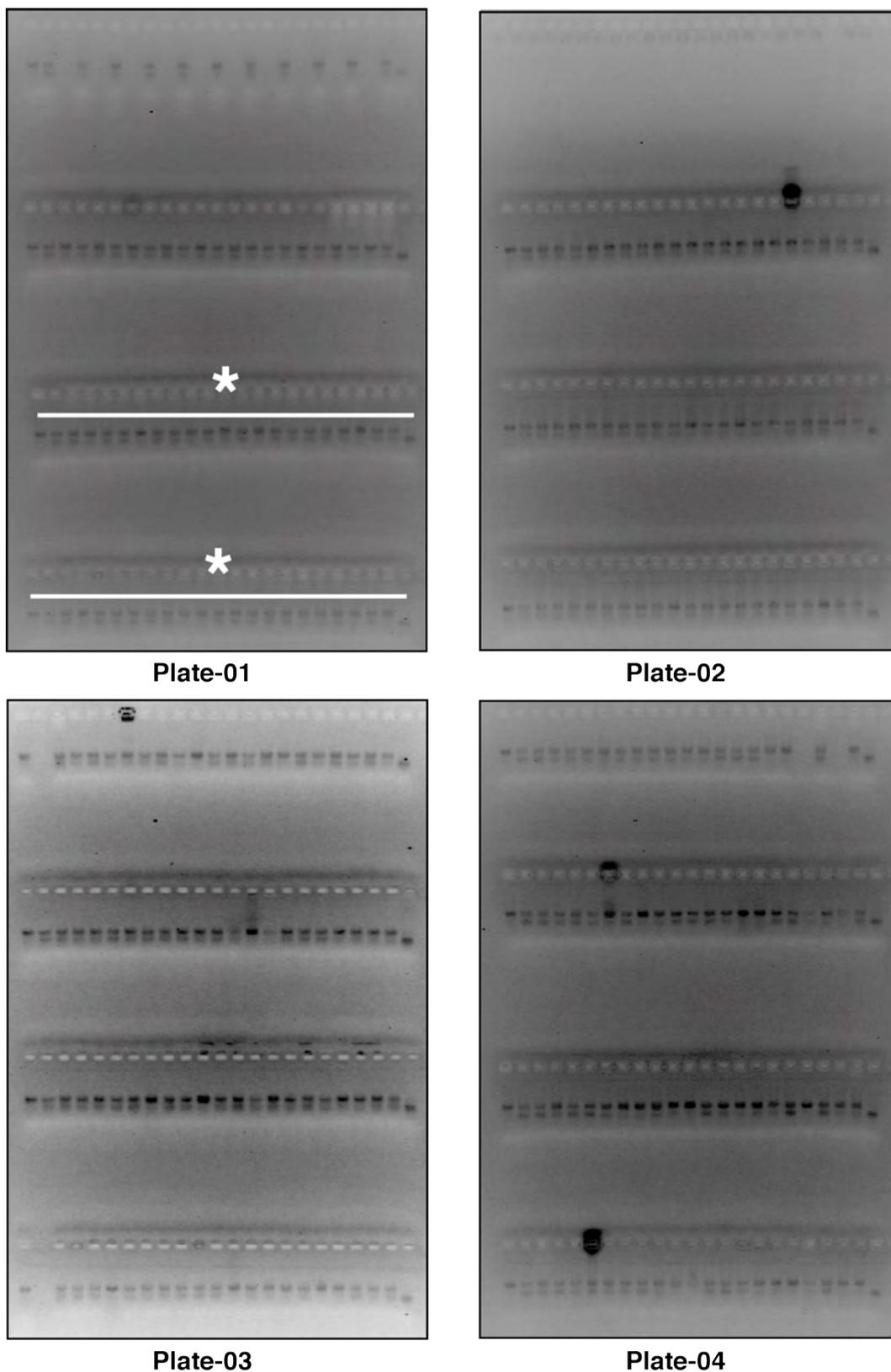
Figure 5.4: Full gel images of a small molecule screen for inhibitors of the human helicase FL-HelQ. Helicase unwinding assays were carried out in the presence of small-fragments to identify whether they were capable of inhibiting FL-HelQ activity. Assays were resolved using 2.5% agarose gels White * indicate candidates displayed in Fig. 5.4.

Plate-01 E-H
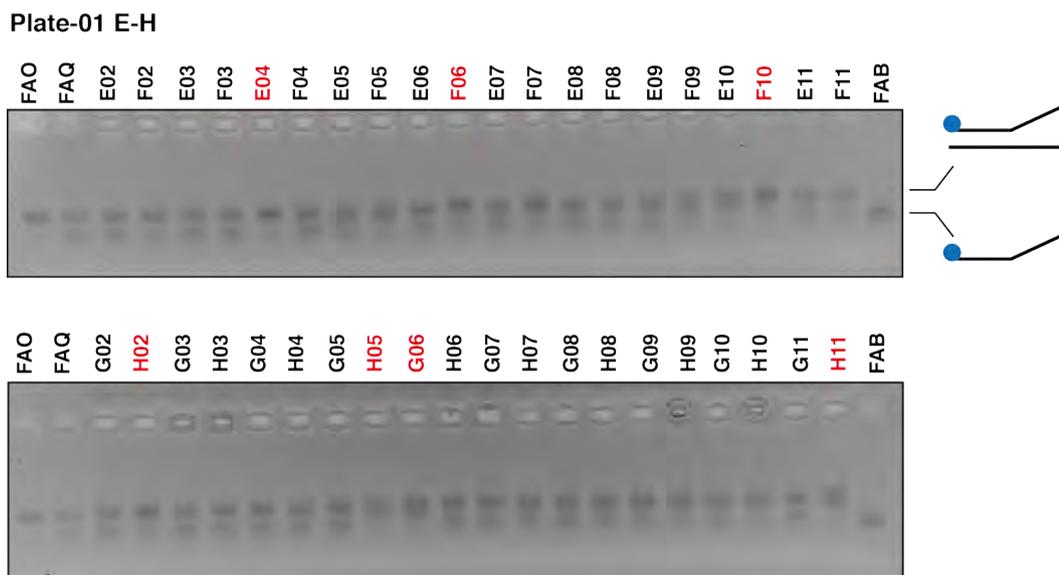
Figure 5.5: A small-molecule screen for inhibitors of human FL-HelQ helicase. Reactions contained 80 nM FL-HelQ and 5 mM of each putative inhibitor. Control lanes were supplemented with 5% DMSO. Candidate molecules were qualitatively assessed for inhibition by comparison with control lanes containing forked DNA substrate only (FAO), a standard helicase unwinding assay (FAQ), and a boiled, completely dissociated substrate (FAB). A number of candidates (red) were identified as impacting FL-HelQ unwinding of the substrate ranging from moderate (F06) to seemingly total inhibition (E04).

5.3.4 Confirmation of 19 candidate molecules as inhibitors of FL-HelQ unwinding activity

Of the 54 putative inhibitor candidates observed in section 5.3.3, 19 were identified as promising for detailed characterisation. As a first step to more effectively quantify helicase inhibition, polyacrylamide gels were substituted in place of lower-resolution agarose gels.

Inhibition of FL-HelQ activity was confirmed using DNA helicase assays supplemented with 5 mM of each inhibitor and quantifying changes against a DMSO-containing control (Fig.5.6). Several candidates were confirmed to have a strong impact on FL-HelQ activity. The greatest effect was seen in assays containing compounds N254, N107 and N144, which reduced FL-HelQ unwinding from the mean total of 56.7% observed in the control, to means of 3.02%, 12.86% and 14.76%, respectively. Less potent were compounds N215

and N326, which reduced the amount of substrate unwound to 28.61 and 27.74%.

The effect of the remaining candidates on FL-HelQ activity was less significant in these assays. Compound N182 was observed to reduce unwinding to 35.85%, while the others allowed FL-HelQ catalysed DNA unwinding of 42 – 47% compared to the 57% uninhibited control. Several compounds, such as N175, appear to have been false positives in the original screen as they exhibited no appreciable inhibition. A clear outlier is N193 which appears to stimulate, rather than inhibit, FL-HelQ activity, achieving an average total unwinding of 76.51%.

Candidates exhibiting poor inhibition were eliminated from characterisation and development. Several further compounds were also discounted due to issues with precipitation during the reaction which could complicate the development process. This included the most strongly inhibitory compound in these assays, N254.

From the remaining candidates, seven fragments were selected for further characterisation, summarised in Table 5.2. This included two strong inhibitors, N107 and N144, as well as weaker inhibitors N212 and N078. One compound which had shown no appreciable inhibition, N214, was also selected to observe its performance in subsequent in the development assays.
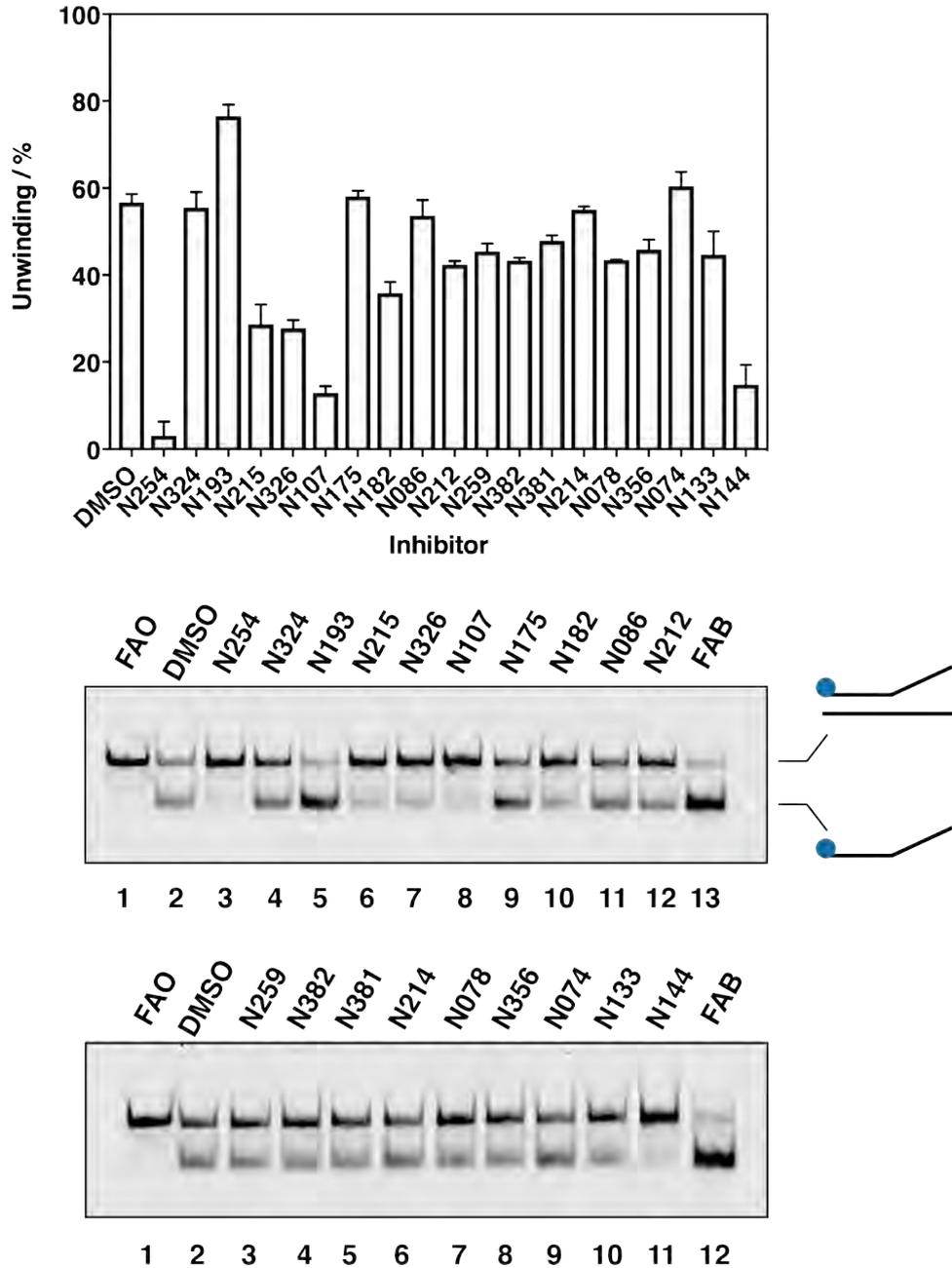
Figure 5.6: Confirming 19 candidate compounds as inhibitors of FL-HelQ
Unwinding assays were carried out in the presence of 5 mM of each inhibitor candidate and incubated at 37 °C for 10 minutes. Inhibition was quantified against control reactions containing substrate-only (FAO) and a standard assay supplemented with DMSO (FAQ). A boiled, fully dissociated controls (FAB) was included to demonstrate complete unwinding. Error bars represent standard deviation of assays carried out in duplicate.

5.3.5 The chemical properties of selected small-molecule inhibitor candidates

Having identified several compounds to study further, chemical property data introduced in section 5.3.1 was gathered for each candidate to better ascertain druglikeness (Table 5.2).

As expected given the pre-screening of the library, all seven of the candidate compounds fall within the acceptable bounds for the majority of metrics. This includes those used in the Ro3 and Ro5: MW, LogP, HBD, and HBA. Analysis of the metrics for individual compounds raised some interesting features, such as the negative Log D values of N078 and N214 which are suggestive of a lower overall lipophilicity and therefore the potential for poor membrane permeability. All candidates were observed to fall below the conventional cut-off value of <140 Å for TPSA, another metric for uptake and transport, which is suggestive of reduced passive membrane permeability. Finally, all of the candidate molecules have Fsp^3 values below the typical cut-off of $\geq$0.42. This may indicate a small 3D footprint, meaning much more planar structures. Higher Fsp^3 values have previously been correlated with candidate success[424].

Table 5.2: Molecular properties of small-molecule inhibitors selected for further study

	N078	N107	N144	N212	N214	N215	N326
MW	151.2	181.6	169.2	133.2	175.2	216.2	164.2
Log P	1.97	1.83	2.85	0.82	2.16	1.26	2.48
HBD	2	2	0	2	2	2	1
HBA	3	1	1	2	2	2	2
log D	-0.63	1.83	2.84	0.82	-1.18	1.26	2.48
TPSA	63.32	36.02	12.89	54.7	53.09	41.13	38.91
Fsp³	0.12	0.11	0.08	0	0.1	0.22	0.12
HAC	11	12	13	10	13	15	11
RTB	1	1	2	0	1	1	0

5.3.6 The impact of small-molecule inhibitors on FL-HelQ ATPase activity

The inhibition of ATPase activity is often used for high-throughput screening of chemical libraries to identify candidates[425–427]. As a second form of hit confirmation, ATPase assays were used to observe inhibition by the selected candidates.

The data show modest inhibition profiles at a concentration of 5 mM for six of the seven candidates, which achieved a relative reduction in ATPase activity across a range of 72.1 – 85.2 %. For several inhibitors this is a severely reduced response to that seen in section 5.3.4. This was particularly surprising for compounds N326 and N144, which had previously been observed to have a strong impact on FL-HelQ activity in unwinding assays. Compound N107 was the only inhibitor to exhibit a strong reduction in the relative ATPase activity of FL-HelQ, displaying a decrease to 27.4 %.
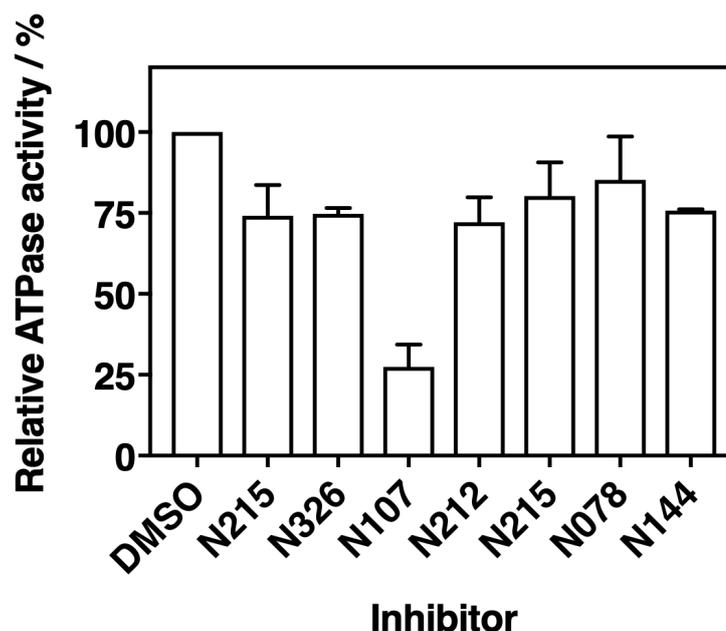
Figure 5.7: Determining the impact of inhibitor candidates on FL-HelQ ATPase activity. Unwinding assays were carried out in the presence of 5 mM of each inhibitor. Relative ATPase activity is represented as a percentage against control assays supplemented with DMSO. Background absorbance was corrected for using a blank reaction lacking FL-HelQ. Error bars represent the standard deviation from the mean of assays carried out in triplicate.

5.3.7 Some small-molecule inhibitors affect the DNA binding ability of FL-HelQ

To further assess inhibitor mode of action against FL-HelQ, electrophoretic mobility shift assays (EMSAs) were used to determine if the compounds altered enzyme binding to the forked DNA substrate (Fig.5.8). Previous work in the Bolt lab has established the conditions for these assays and as such a concentration of 100 nM FL-HelQ was used to observe DNA binding [229].

The impact of increasing DMSO concentrations on DNA binding by FL-HelQ was assessed using EMSAs (Fig.5.8A). The data display an inverse relationship between DMSO concentration and FL-HelQ:DNA binding with a reduction from an average of 89.8% to 51.0% across the concentration gradient. It should be noted that the error observed at 20% DMSO is high (SD $\pm$ 21.2 %). At the maximum concentration of 5% DMSO used in unwinding assays, DNA binding decreased to an average of 82.9%, which is sufficient to

characterise the impact of inhibitor candidate molecules on FL-HelQ:DNA binding.

Finally, the impact of inhibitors at 5 mM on FL-HelQ:DNA binding was examined using EMSAs (Fig.5.8B). The data show that four inhibitors have an impact on the ability of FL-HelQ to bind DNA. Compared to the mean of 99.1% binding by the DMSO control, N214 and N078 reduced to 71.4% and 75.1% respectively. Notably, the promising inhibitors N107 and N144 saw a large reduction in FL-HelQ:DNA binding to 68.9% and 69.7%. An unpaired t-test was used to compare the difference between binding in the control assays and each inhibitor. The result showed no significant difference in binding for six of the seven inhibitors. The binding difference between the control and N107 was however shown to be significant (p= 0.005).
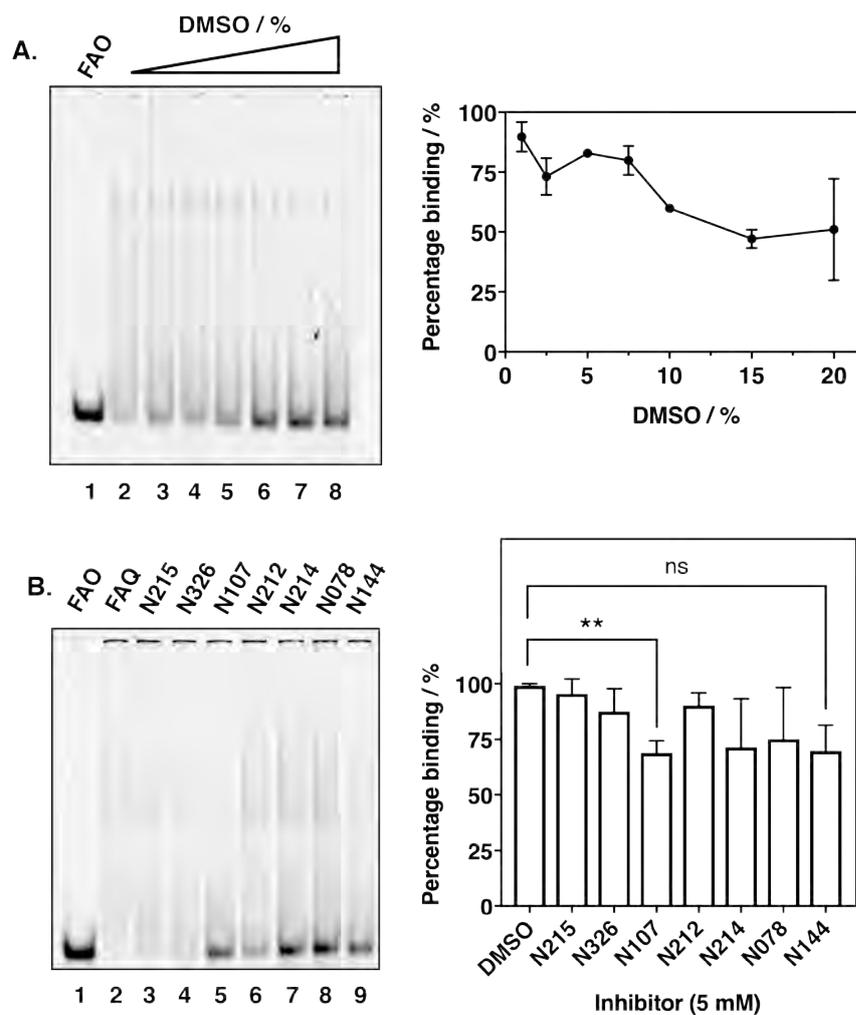
Figure 5.8: Determining the impact of inhibitor candidates on FL-HelQ binding to a forked DNA substrate. (A) Titration of DMSO into DNA binding assays. FL-HelQ was added to reactions at 100 nM. DMSO concentrations used were 0, 1, 2.5, 5, 7.5, 10, 15 and 20%. Error bars show standard deviation for assays carried out in duplicate. **(B)** DNA binding assays supplemented with 5 mM small-molecule inhibitors. Error bars represent standard deviation of assay carried out in triplicate. Significance between assays containing inhibitors and a DMSO-only control was determined by an unpaired t-test. ** represents a p value ≤ 0.01. DNA binding was quantified against controls containing DNA only (FAO) and an assay lacking either DMSO (**A**, FAQ) or inhibitor (**B**,FAQ).

5.3.8 Determining IC50 values for small-molecule inhibitors effect on FL-HelQ

Characterisation of inhibitors in previous sections was conducted entirely at the maximum dosage of 5 mM. To better assess candidate performance, unwinding assays were supplemented with lower concentrations of inhibitor

across a range of 0.1 – 5 mM and the data obtained used to calculate IC50 values (Fig.5.9). To address variance in the total unwinding between assays, and more easily observe differences in the activity profiles of each candidate, data was normalised to 100% against control unwinding assays for each dataset (Fig.5.9, FAQ).

Inhibition profiles obtained for candidate compounds were varied. The most potent inhibitor was candidate N107, which reduced total unwinding by 94.25%. It also achieved inhibition at much lower concentrations than the other compounds, reducing total unwinding to 13.6% at a concentration of 2 mM. Candidates N144 and N326 showed less significant inhibition, achieving a total reduction in unwinding of 78.8% and 72.1%, respectively, when compared to the control. Inhibition by the remaining candidates was shown to be lower even at higher concentrations.
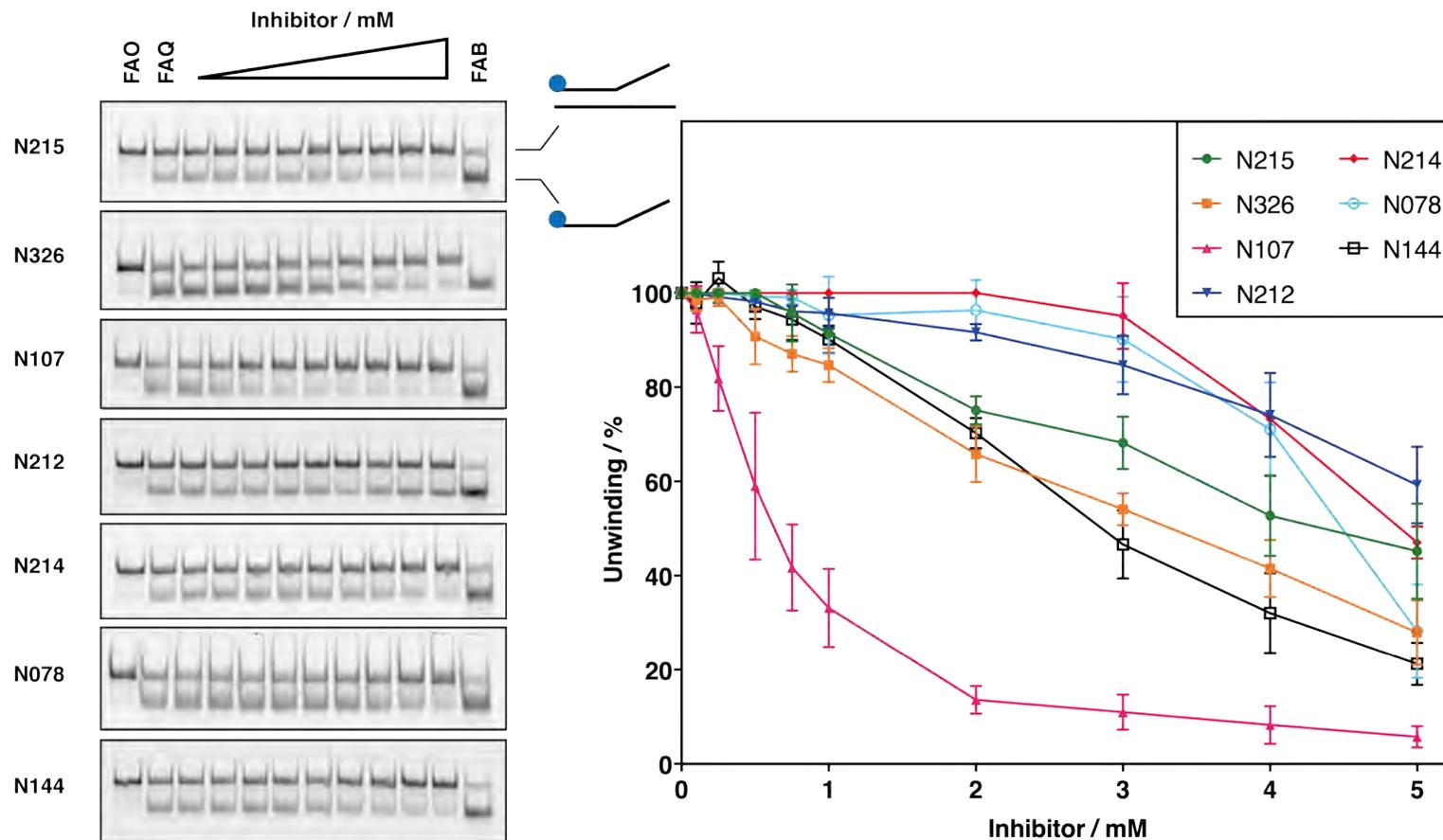
Figure 5.9: Titration of inhibitor candidates into DNA helicase assays in order to determine IC50 values. Inhibitors were titrated into reactions at concentrations of 0.1 0.25, 0.5, 0.75, 1, 2, 3, 4 and 5 mM. Bands were quantified by comparison against substrate only (FAO), standard unwinding assay (FAQ) and boiled, dissociated (FAB) controls. Error bars represent standard deviation for assays carried out in triplicate. Data was normalised to 100% against the control assays seen in lane 2 of all gels.

Unwinding data obtained from concentration titration assays were then used to determine IC50 values for each inhibitor candidate. Raw unwinding data were plotted against logarithms ($\log_{10}$) of inhibitor concentration (Fig.5.10). The data were then analysed in Graphpad Prism using a four-parameter logistic curve, where a and d are equal to the maximum and minimum asymptotes, c is the IC50 and b is the Hill's slope of the curve. This allows an IC50 value to be interpolated from a curve constrained by the assay data. For further comment on the four-parameter logistic curve see section 2.2.9

$$y = d + \frac{a - d}{1 + 10^{(c-x) \times b}}$$

The curves generated display a broad range of putative IC50 values and highlight some interesting differences between the inhibitory capability of each compound. N107 was the only candidate to achieve a sub-millimolar IC50 value at 0.557 mM; a 10.6-fold difference in efficacy to the least effectve inhibitor, N212, which was calculated to have an IC50 of 5.897 mM. This value however may not be reliable as it exceeds the maximum assay concentration and was extrapolated from the trajectory of the IC50 curve. N107 is also superior to other promising candidates, with a 5.15-fold greater efficacy than N326 and 5.05-fold difference to N144, the nearest competing inhibitors. The IC50 values obtained for each candidate are summarised in Table 5.3.
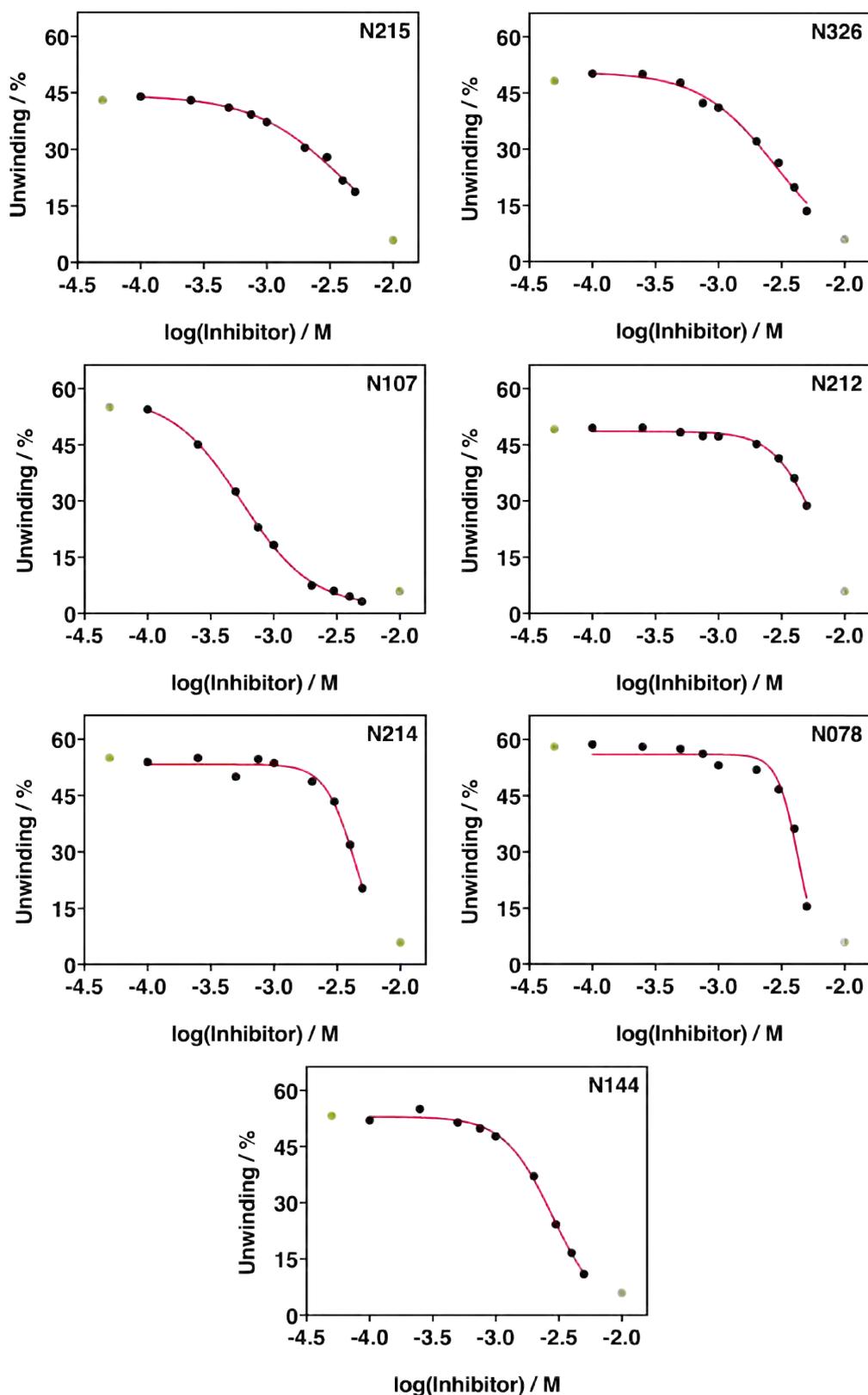
Figure 5.10: Determining putative IC50 values for seven small-molecule inhibitor candidates.. Inhibitors were added to reactions at 0.1 0.25, 0.5, 0.75, 1, 2, 3, 4 and 5 mM. Unwinding data were converted to logarithms (Log(10)) and analysed using a 4-parameter logistic curve (red). IC50 values were calculated from assays done in triplicate, mean values are displayed. Controls values (green) are displayed for unwinding assays lacking inhibitor and boiled, fully dissociated substrate.

Table 5.3: IC50 values calculated for small-molecule inhibitors

Inhibitor candidate	IC50 (mM)
N215	4.026
N326	2.869
N107	0.557
N212	5.897*
N214	4.423
N078	4.301
N144	2.812

* denotes an extrapolated value.

5.3.9 Small-molecule inhibitor candidates differentially impact the unwinding activity of C-HelQ and bacterial RecQ helicases

Having calculated IC50 values to assess inhibitor potency, the candidates were further analysed to evaluate their specificity to FL-HelQ. To achieve this, each compound was assayed for ability to inhibit two different proteins: a C-terminal fragment (C-HelQ) of FL-HelQ reported to display the same unwinding activity, and the *E. coli* RecQ protein. *E. coli* RecQ, was used for this test because like HelQ it is a superfamily-2 helicase with 3′ to 5′ translocating polarity, it has several homologues in humans, but unlike HelQ it is not a Ski-2 helicase and is therefore distinct.

A comparison of unwinding activity was made between FL-HelQ, C-HelQ, and RecQ in the absence of small-molecule inhibitors, in endpoint and time-course assays (Fig. 5.11). End-point data show that FL- and C-HelQ have similar profiles for unwinding, with the latter capable of achieving a higher percentage of substrate dissociation over the course of the assay (Fig. 5.11A). RecQ was far more active than either HelQ, reaching 100 % unwinding at a concentration of 40 nM. To account for the difference in unwinding, RecQ was used at 10 nM in subsequent assays as at this concentration it

reached 70.5 % unwinding which is comparable with that of C-HelQ.

Time-course data shows that C-HelQ and RecQ unwind forked substrate at a similar rate (Fig. 5.11B). The data also show that C-HelQ unwinds the forked DNA substrate at a faster rate than FL-HelQ when tested at 80 nM. C-HelQ was able to achieve over 50% unwinding by the 5 minute timepoint of the assay whilst FL-HelQ only unwound an average of 22.0% of the substrate in the same time. There is also a notable disparity in unwinding between the two proteins at 80 nM in Fig. 5.11A, where an average of 49.24% of substrate was unwound, compared with the 10 minute timepoint in Fig. 5.11B, where 77.29% of the substrate was unwound.
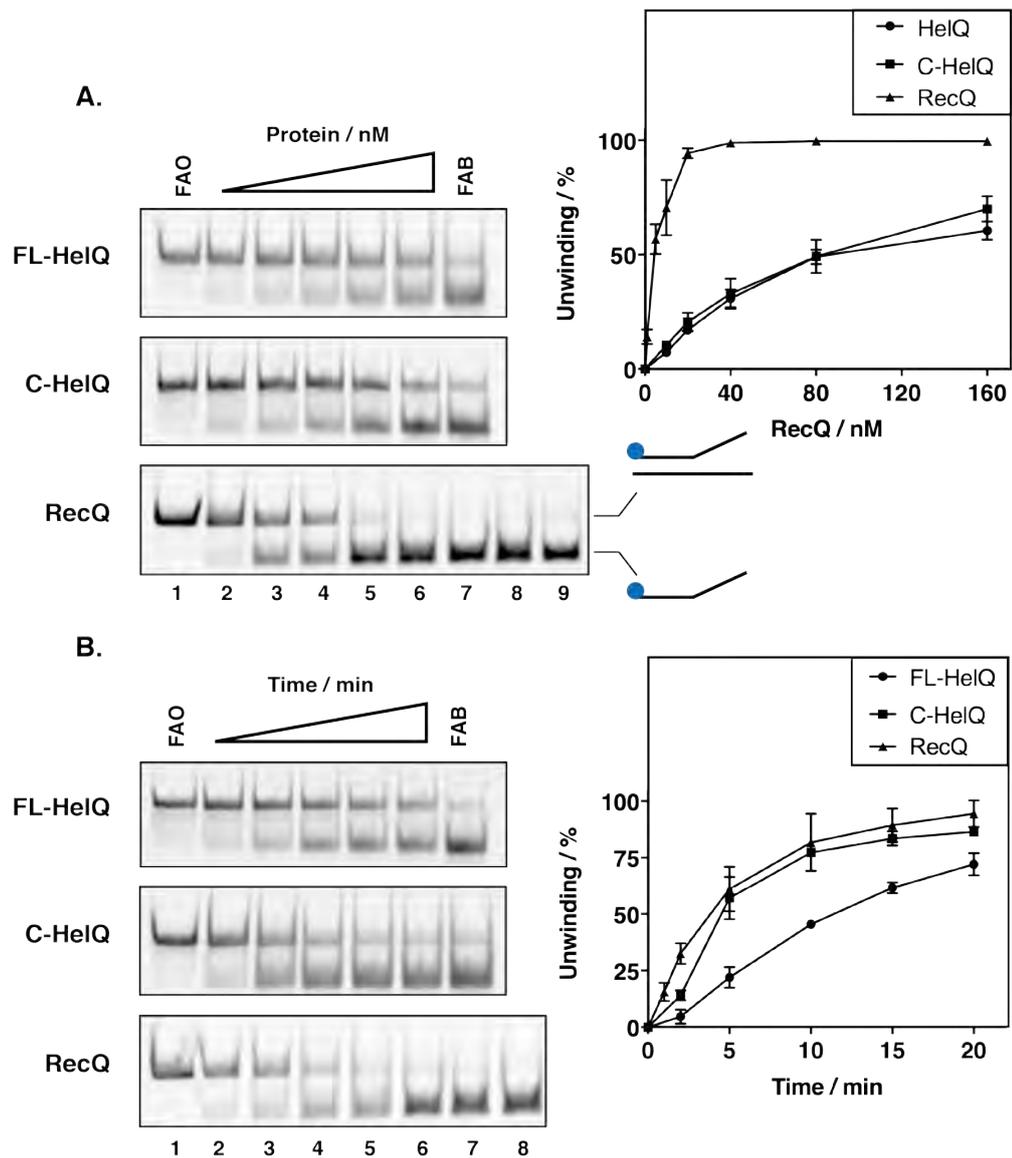
Figure 5.11: Comparing the unwinding activity of full-length FL-HelQ and a C-terminal fragment. (**A**) Unwinding activity measured at varying concentrations of FL-HelQ, C-HelQ and RecQ. Protein was added to reactions at concentrations of 10, 20, 40, 80, and 160 nM for HelQ. RecQ was tested at additional concentrations of 1 and 5 nM. (**B**) Unwinding activity of FL- and C-HelQ measured over time. Protein was added to reactions at 80 nM. Assays were stopped at timepoints of 0, 2, 5, 10, 15, and 20 minutes. Timepoints of 1 and 5 minutes were additionally taken for RecQ. Error bars represent standard deviation of assays carried out in triplicate.

Having established unwinding profiles for C-HelQ and RecQ relative to FL-HelQ, drug interaction and specificity comparisons were then made between the proteins. Unwinding assays were carried out as previously described, using 80 nM FL- or C-HelQ and 10 nM RecQ, in the presence of 5

mM inhibitor. The values obtained were then normalised to 100% against control unwinding assays to allow for direct comparison (Fig 5.12).

C-HelQ presents a similar inhibition profile to HelQ, with multiple candidates appearing to have more of an impact on the C-terminal fragment than they do on the full-length protein. To determine whether these differences were statistically different, an unpaired t-test was used. The results show that the differences observed between FL- and C-HelQ for both N078 (p= 0.035) and N144 (p= 0.047) were significant.

The data obtained for RecQ show that five of the seven inhibitors have little-to-no impact on helicase activity, reporting relative unwinding across a range of 92.8–94.5 %. The remaining two candidates achieved modest inhibition of RecQ activity with N078 and N212 reducing relative unwinding to 45.3 % and 72.1 %, respectively. In the case of N212, this reduction was the largest decrease in activity observed while analysing the compound.
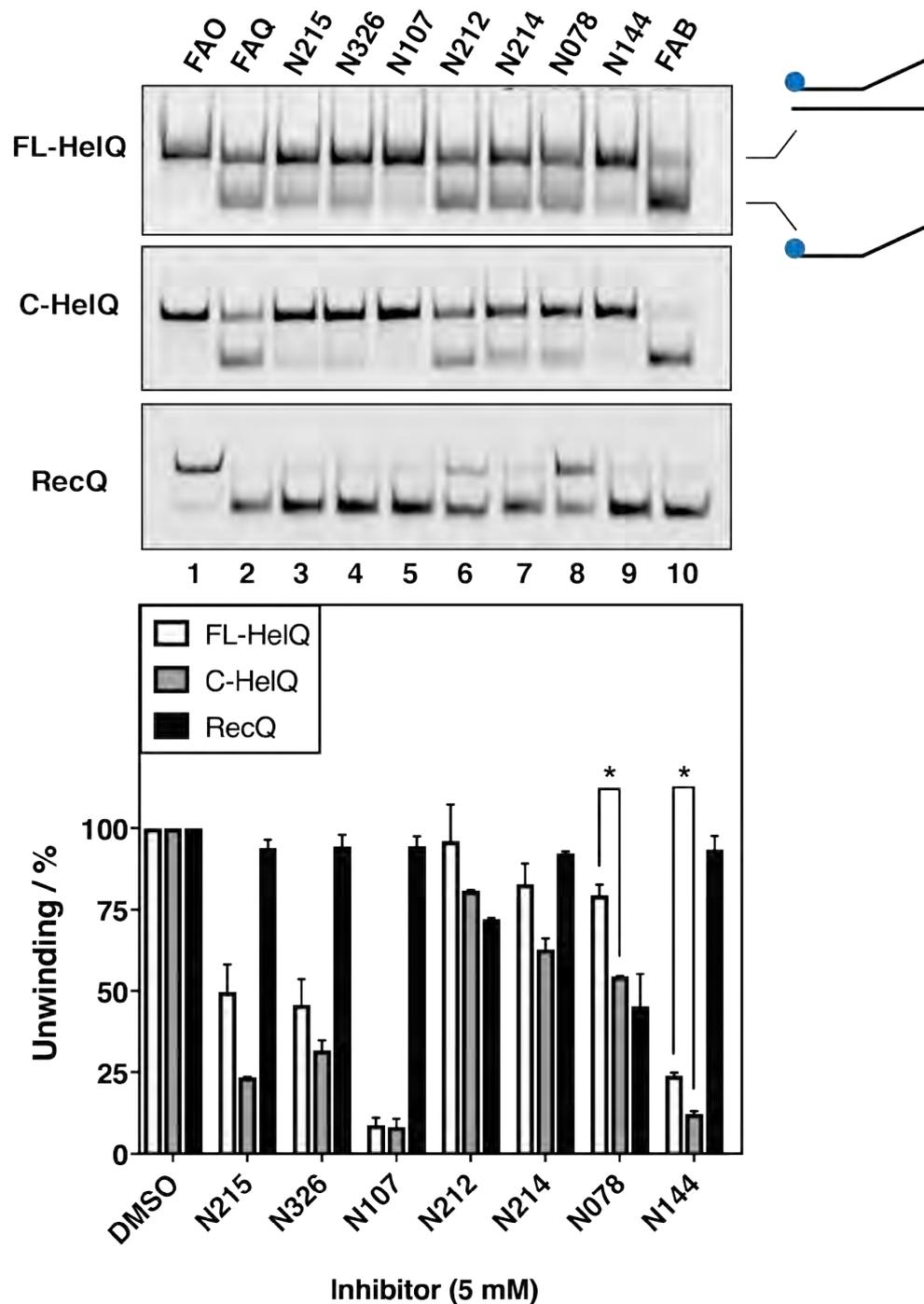
Figure 5.12: A comparison of inhibitor efficacy between full-length FL-HelQ and a hyperactive C-terminal fragment. Assays were incubated at 37 °C for 10 minutes. FL-HelQ and C-HelQ were used at a concentration of 80 nM. RecQ was added to reactions at 10 nM. Unwinding was quantified against controls containing DNA only (FAO), a standard assay supplemented with DMSO (FAQ) and a boiled sample (FAB). All data were normalised to 100 % against control unwinding assays supplemented with DMSO. Significance between the observed means of FL- and C-HelQ assays was determined by an unpaired t-test

5.3.10 Assessing small-molecule inhibitors as drug-development candidates using dynamic light scattering

A common source of false-positive results during drug development is colloidal aggregation of candidate compounds, which often results in promiscuous interaction with proteins. Under assay conditions this gives the appearance of inhibition, when the effect is actually denaturing to the protein, or non-specific.

To exclude aggregation as a possible source of inhibition, the size of candidate compounds was analysed using dynamic light scattering (DLS). Compounds were tested at a concentration of 100 mM, dispersed in DMSO. Each sample was measured five times to monitor reproducibility. Data quality was assessed against measurements of a nanopshere size standard (200 nm) as a positive control. Biotin, a known non-aggregator with a similar size to candidate compounds (MW 242.3), was included as a negative control[428]. The results for inhibitors of interest are displayed in Fig 5.13 and Table 5.4. Complete data for all inhibitor candidates can is provided as an appendix (Appendix A.1.2).

The data obtained for the majority of inhibitor candidates suggested that they were beneath the instrument detection range of 0.3 nm. A representative example of this can be seen in measurements for candidate N107 which displays a low y-intercept value of 0.05-0.1 in the correlogram of Fig 5.13, as well as a noisy and dispersed signal similar to that of the biotin negative control. Similarly, it displays a major peak in number intensity at or below 1 nm. This could demonstrate a lack of aggregation, or a weak signal. This data uncertainty is compounded in the recorded values for average diameter and polydispersity collected in Table 5.4. The values appear much larger for both N107 and the biotin control sample, including large values for standard deviation when compared to the nanosphere positive control.

Measurements for candidate N212 displayed y-intercept values between

0.7 and 0.8, indicative of a strong signal, as well as a low signal-to-noise ratio similar that of the nanopshere size standard (Fig 5.13). The data suggest aggregation of the inhibitor which recorded a major number intensity peak at 1×10^3 nm and a lower occurrence, secondary population at 1×10^2 nm (Fig 5.13). This is supported by the recorded polydispersity index value of 0.371, which is suggestive of multiple populations, falling between the typical cut-off values for mono-dispersed samples (≤ 0.05) and samples that are too broadly dispersed for reliable measurement by DLS (≥ 0.7)[429]. The values obtained for these inhibitors and controls are summarised in Table 5.4 or in appendix A1 for all nineteen compounds.
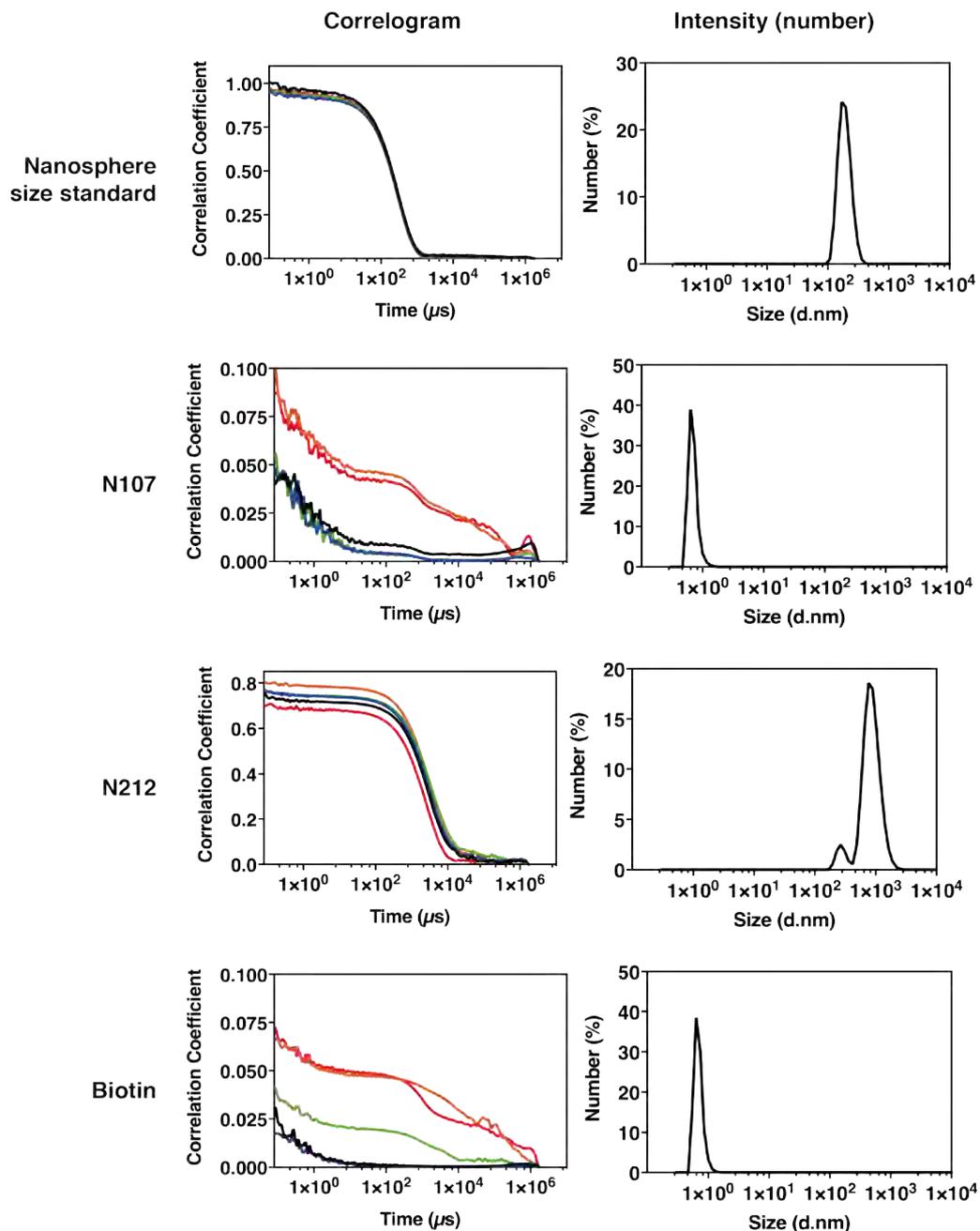
Figure 5.13: Dynamic light scattering to determine the aggregation properties of inhibitor candidates. Correlograms indicate the quality of data in terms of both signal strength and reproducibility. In each correlogram, a different coloured trace represents an independent measurement of the same sample. Intensity plots display diameter based upon the number of particles in the solution. Values shown on these plots are the mean of five separate measurements. A nanosphere size standard (200 nm) was used as a control to produce positive test data, while biotin was used as a small, non-aggregating negative control.

Table 5.4: Particle diameter and dispersity data obtained from DLS

Sample name	Intensity averaged diameter (nm)	SD	Polydispersity Index	SD
Nanosphere size standard	209.3	1.799	0.045	0.059
N107	255	344.5	0.677	0.219
N212	1211	171.2	0.371	0.112
Biotin	557.1	746.1	0.871	0.533

5.3.11 The development of lead compounds from the first phase of inhibitor testing

Having identified and characterised several promising candidates from the initial screen, a second phase of inhibitors was developed by Sygnature Discovery in conjunction with Nanna Therapeutics. New compounds were synthesised using the previously described N107, N214 and N078 as a scaffold. From this new phase, 27 candidates were identified for further testing. Personal correspondence with Sygnature Discovery indicated that all compounds displayed inhibition at 1 mM concentrations as observed by Transcreener® APD^2 assay.

Prior to characterisation of the new candidate compounds as described previously, inhibition of FL-HelQ was confirmed using helicase unwinding assays (Fig.5.14). As in section 5.3.4, the assays were supplemented with each small molecule inhibitor, although in this case the final concentration of each was 1 mM.

The data show that of the twenty-seven candidates, only two were capable of strong inhibition of FL-HelQ. Compound N274 reduced unwinding to an average of 15.8 % compared to the control reaction which achieved a mean of 60.3 %. N186 meanwhile, was able to reduce total unwinding to an average of 16.2 %.

Several other candidates were able to achieve modest inhibition of FL-HelQ activity. Compounds N381, N147, N145 and N701 all reduced the total observed unwinding below 50%, reporting values across a range of 44.8–49.7 %. Full unwinding data for all twenty-seven compounds are summarised in Table 5.5.
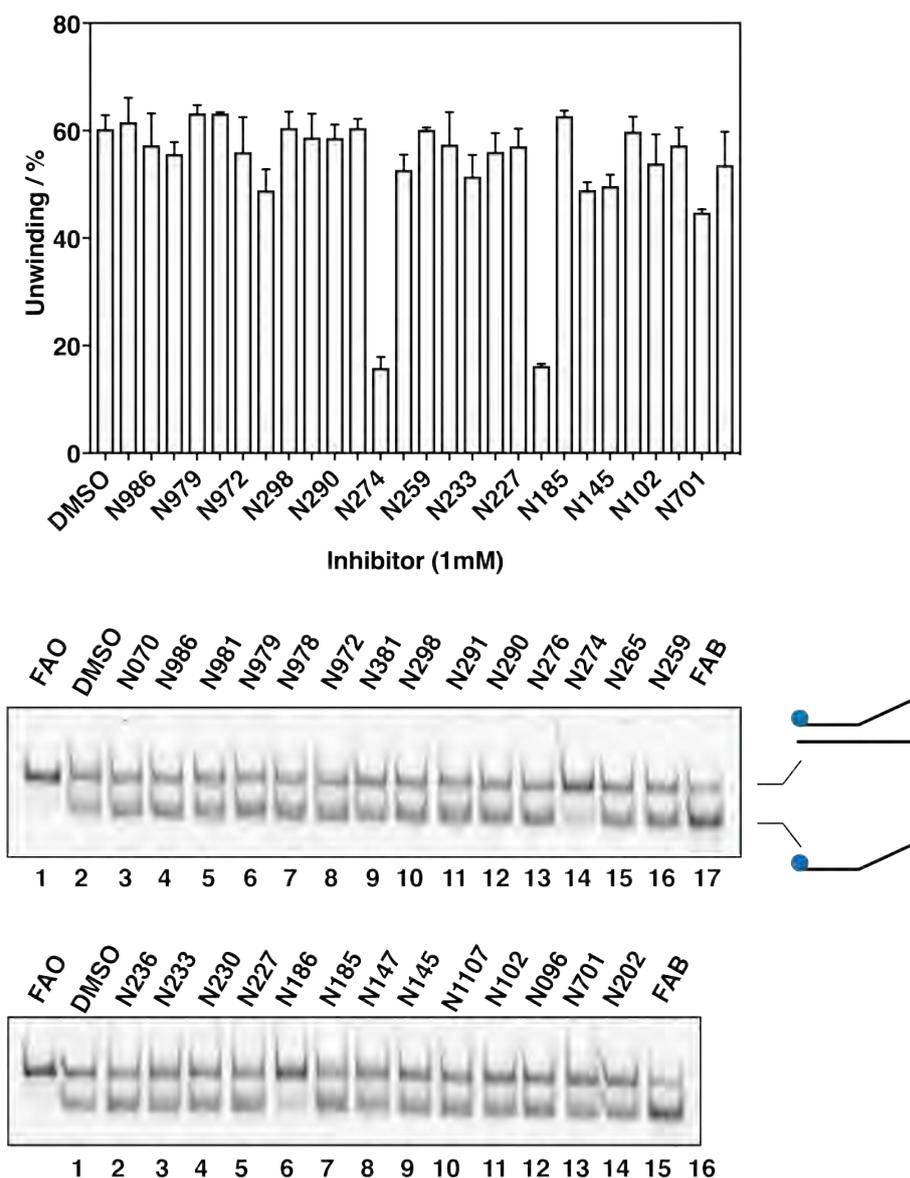
Figure 5.14: Confirming FL-HelQ unwinding inhibition by twenty-seven small-molecule inhibitor candidates. Unwinding assays were carried out in the presence of 1 mM of each inhibitor candidate and incubated at 37 °C for 10 minutes. Inhibition was quantified against control reactions containing substrate-only (FAO) and a standard assay supplemented with DMSO (FAQ). A boiled, fully dissociated control (FAB) was also included. Error bars represent standard deviation of assays carried out in duplicate.

Table 5.5: Mean percentage unwinding values for phase two inhibitors at 1 mM.

Inhibitor ID	Average unwinding (%)	SD
N070	61.598	4.510
N986	57.306	5.870
N981	55.677	2.168
N979	63.202	1.549
N978	63.179	0.229
N972	56.013	6.485
N381	48.886	3.899
N298	60.506	3.002
N291	58.734	4.393
N290	58.658	2.450
N276	60.504	1.659
N274	15.841	2.007
N265	52.693	2.797
N259	60.161	0.417
N236	57.413	5.999
N233	51.447	4.023
N230	56.068	3.474
N227	57.112	3.211
N186	16.204	0.379
N185	62.751	0.966
N147	48.960	1.438
N145	49.680	2.113
N1107	59.808	2.773
N102	53.901	5.418
N096	57.251	3.348
N701	44.780	0.553
N202	53.627	6.143

5.3.12 The impact of inhibitor lead compounds on FL-HelQ ATPase activity

As in section 5.3.6, ATPase assays were used as a secondary method to attempt the confirmation of candidates and to gain some early insight into whether they might have any impact on FL-HelQ ATPase activity. Seven candidates were chosen for testing including N274 and N186 which were confirmed for inhibition in the previous section, modest candidates N381 and N202, as well as three additional candidates which had achieved strong inhibition when tested by Sygnature Discovery (personal correspondence).

Several of the inhibitors were observed to impact on FL-HelQ ATPase activity at 1 mM (Fig. 5.15). The strongest response came from candidates N274 and N186 which achieved mean reductions in relative activity of 41.8 % and 46.8 %. This reflects the performance of the two candidates seen in section 5.3.11, albeit with a less powerful effect.

Candidates N381 and N202 were also observed to effect the ATPase activity of FL-HelQ, albeit to a lesser extent. N381 reduced relative activity to 59.8 %, while N202 achieved a less effective 50.9 %. This effect appears to be more potent than the observed decrease to unwinding observed in section 5.3.11. The remaining three candidates N290, N276 and N236 were shown to have little impact on overall ATPase activity, supporting the lack of inhibition observed in section 5.3.11.
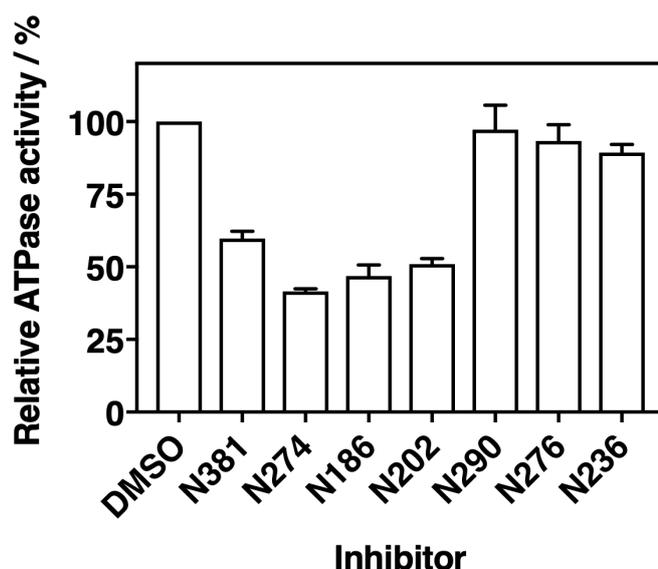
Figure 5.15: Determining the impact of inhibitor lead compounds on FL-HelQ ATPase activity. Unwinding assays were supplemented with 1 mM of each inhibitor. Relative ATPase activity is represented as a percentage against control assays supplemented with DMSO. Background absorbance was corrected for using a blank reaction lacking FL-HelQ. Error bars represent the standard deviation from the mean of assays carried out in triplicate.

5.3.13   Determining IC50 values for inhibitor lead compounds

Despite the apparent poor performance of the second phase of inhibitors, two lead compounds still achieved significant inhibition. These two candidates, N274 and N186, were further analysed alongside N381 which reported some weak inhibition(see section 5.3.11). The inhibitors were assayed at concentrations from $50 - 1000 \ \mu$M to assess their impact on unwinding (Fig. 5.16)i. The data was normalised to 100 % against control reactions to facilitate comparisons between inhibitors.

The strongest overall inhibition was achieved by inhibitor N274 which was able to reduce relative unwinding activity to 22.9 % at the maximum concentration of 1 mM. Candidate N186 also recorded strong inhibition, achieving an overall decrease in relative unwinding to 38.9 %, following a similar trajectory to N274 across the gradient. N381 remained the weakest of the three candidates, achieving a mean relative unwinding of 50.1 % at the maximum concentration of 1 mM.
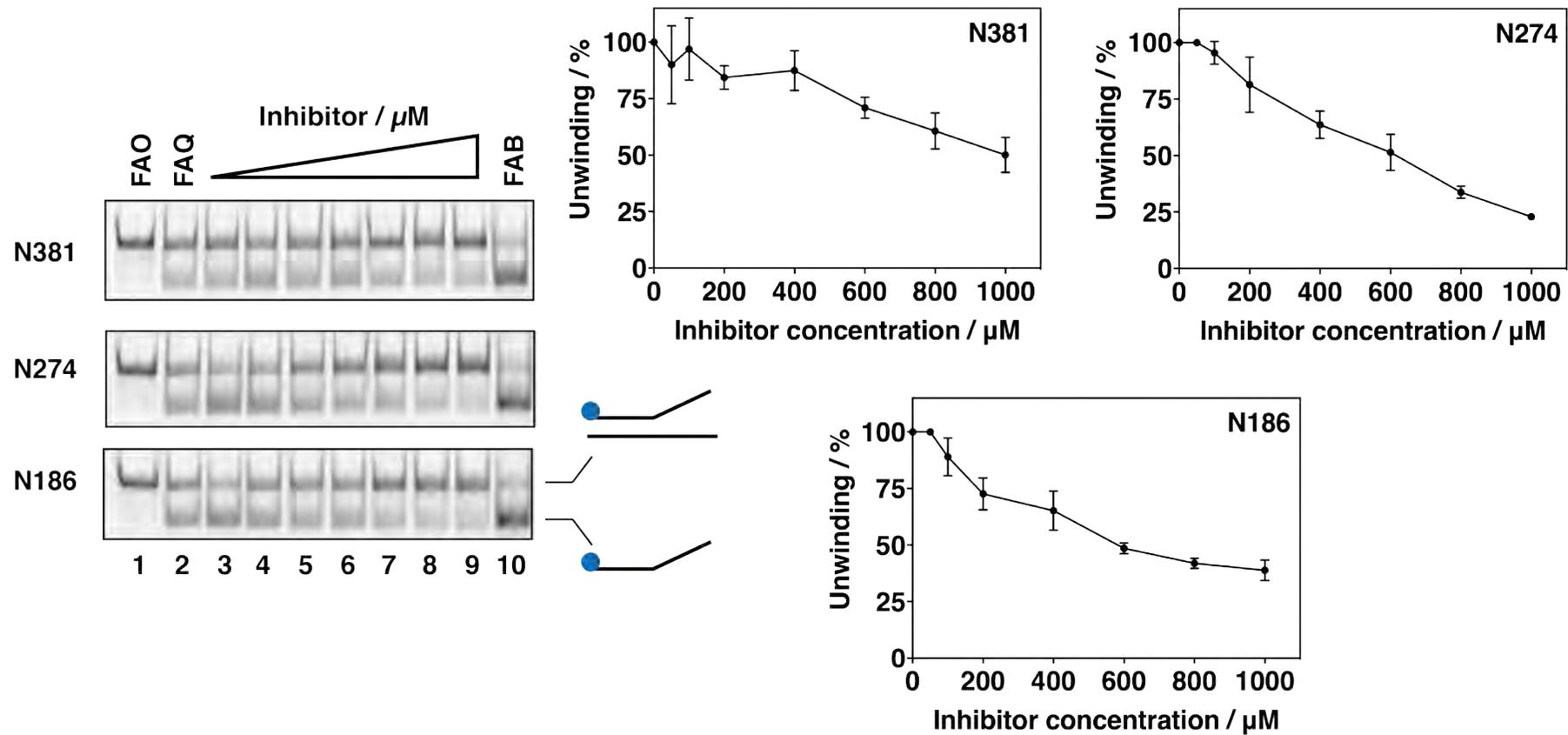
Figure 5.16: Titration of second phase inhibitor candidates into DNA helicase assays in order to determine IC50 values All assays were incubated at 37 °C for 10 minutes. Concentrations of inhibitors used were 50, 100, 200, 400, 600, 800 and 1000 μM. Bands were quantified by comparison against substrate only (FAO), standard unwinding assay (FAQ) and boiled, dissociated (FAB) controls. Data was normalised to 100% against control assays seen in lane 2 of all gels. Error bars represent standard deviation from assays carried out in triplicate.

Using raw unwinding data from these assays, IC50 values were calculated for the three inhibitors using a four-parameter logistic curve (Fig. 5.17), as described in section 5.3.8.

The values obtained for N278 and N186 were very similar, with calculated IC50s of 547.8 μM and 564.9 μM, respectively. A reliable IC50 value was unable to be determined for candidate N381, which failed to reduce relative unwinding below 50 % at the highest concentration of 1 mM. Instead, an IC50 of 1054 μM was extrapolated from the curve. Values obtained for each candidate are summarised in Table 5.6.
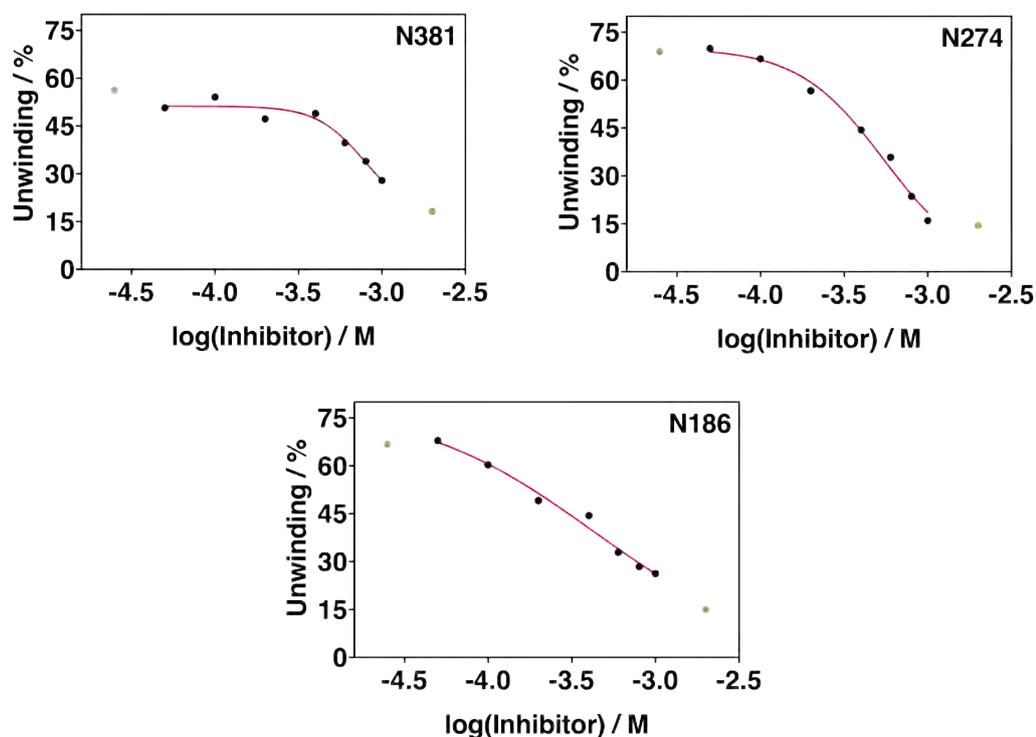
Figure 5.17: Determining putative IC50 values for a second phase of small-molecule inhibitor candidates. Inhibitors were added to reactions at 50, 100, 200, 400, 600, 800 and 1000 μM . Unwinding data were converted to logarithms (Log(10)) and analysed using a 4-parameter logistic curve (red). IC50 values were calculated from assays done in triplicate, mean values are displayed. Controls values (green) are displayed for unwinding assays lacking inhibitor and boiled, fully dissociated substrate.

Table 5.6: IC50 values calculated for the second phase of small-molecule inhibitors

Inhibitor candidate	IC50 (μM)
N381	1054.0*
N274	547.8
N186	564.9

* denotes an extrapolated value.

5.3.14   Some inhibitors appear to be competitors for ATP

The inconsistency in results for successful inhibition between Sygnature Discovery and the assays conducted in this work remained an issue throughout the second phase of testing. Aside from the assay method for observing inhibition, a key difference in the experimental conditions used was the concentration of ATP added to each assay. This raised the possibility that inhibitors were competing with ATP and that this was overcome by the higher concentrations used in this work.

To address this potential issue, unwinding assays were carried out for the seven inhibitor candidates tested in sections 5.3.12 and 5.3.13. To maximise the chance of observing successful inhibition, compounds were added to assays at a final concentration of 1 mM. Each set of assays was also supplemented with different concentrations of ATP and incubated for an extended assay time of 1 hour. To allow for comparison between the different ATP concentrations, data were normalised to 100 % against control reactions (Fig. 5.18.

Inhibitors N274 and 186 can be seen to consistently inhibit FL-HelQ activity across the concentration gradient, consistent with their previous performance. Inhibitors N202, N290, and N236 were all able to achieve inhibition of HelQ at 0.25–0.5 mM ATP, but this inhibition was not observed when ATP concentration was increased to 1 mM.
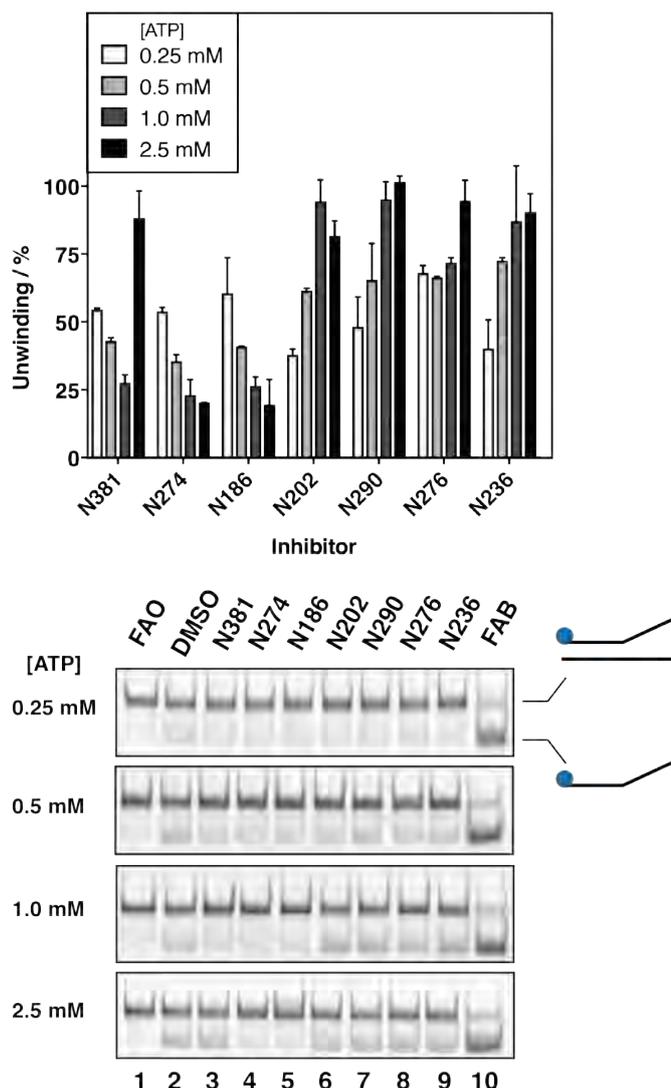
Figure 5.18: Determining inhibitor impact on forked DNA unwinding at differing concentrations of ATP. Rections were carried out using a constant 1 mM of each inhibitor. Assays were carried out using 0.25, 0.5, 1.0 or 2.5 mM ATP. Reactions were incubated at 37 °C for 60 minutes. Percentage unwinding was calculated against a control unwinding assay supplemented with DMSO (FAQ). Also included were a DNA only control (FAO) and a boiled, fully dissociated control (FAB). Data were normalised to 100% against control assays to account for disparity in the unwinding at different concentrations of ATP. Error bars represent standard deviation for assays carried out in duplicate.

Interestingly, both compounds N381 and N276 were able to maintain inhibition of FL-HelQ up to 1 mM ATP. At this concentration N381 reduced relative unwinding to an average of 27.6 %, while N276 achieved a more modest decrease to 71.9 %. In both cases, this inhibition was overcome at the maximum assay concentration of 2.5 mM ATP, with relative decreases in

unwinding falling to an average of 88.2 % and 94.7 %, respectively.

Having observed that inhibition by several of the phase two candidates can be overcome by sufficiently high concentrations of ATP, ATPase assays were revisited to determine whether candidate compounds could be detected at lower concentrations (Fig. 5.19).

The data obtained do not show the same differences in inhibitor efficacy seen in (Fig. 5.18. Inhibitors N290 and N276 appeared to have no change in effect on activity. Candidates N274 and N186 remain promising inhibitors, achieving a stronger effect on relative ATPase activity at 1 mM ATP. N381 was shown to be more potent at 1 mM ATP versus 5 mM, although the difference is not as drastic as observed in (Fig. 5.18.

To determine whether this change was significant, unpaired t-tests were used to compare the mean relative ATPase activity obtained for each inhibitor at 1 mM and 5 mM ATP. The result shows that while not as large as expected, the difference in relative activity for candidate N381 was significant (p= 0.013). The t-test also show significant differences between the means for candidates N274 and N186 (P= 0.006, P= 0.044, respectively).
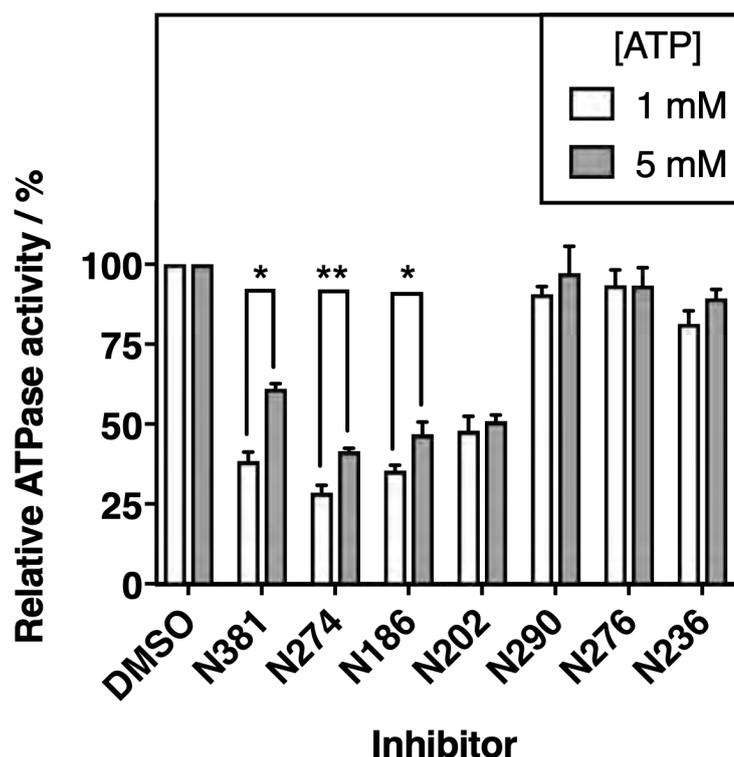
Figure 5.19: Determining the impact of inhibitor lead compounds on FL-HelQ ATPase activity at different concentrations of ATP. Unwinding assays were carried out in the presence of 1 mM of each inhibitor and either 1 or 5 mM ATP. Relative ATPase activity is represented as a percentage against control assays supplemented with DMSO. Background absorbance was corrected for using a blank reaction lacking FL-HelQ. Error bars represent the standard deviation from the mean of assays carried out in triplicate. Significance between means was assessed using unpaired t-tests. * represents a p value ≤ 0.05. ** represents a p value ≤ 0.01.

5.4 Discussion

5.4.1 Drug discovery in context

Novel drug development from putative candidate to finished product is complex and expensive, commonly taking 12-15 years to fully realise[430]. This is highlighted in a 2020 study analysing data from 2009–2018 which found that bringing a new drug to market had a mean investment value of £962 million[431].

A key issue in this process is correct protein and pathway target identification for diseases lacking extensive prior research or understanding[432].

Many approaches typically used to identify targets are limited by high false discovery rates in pre-clinical study, leading to less than 10 % of industry-developed compounds progressing to Phase I clinical trials and an overall failure rate of 96 %[432].

Academic research is increasingly vital to innovation in the development of new drugs[433]. Specialist insight into proteins, pathways and mechanisms removes much uncertainty from target identification and may inform early assay development[434]. The success of academia-led projects is reflected in an increased rate of progression to Phase I clinical trials of 29.5 % between 1991–2015[435]. However, these projects often take much longer to advance, limited by the scale of screening, lack of pharmaceutical expertise and less aggressive deadlines than in industry[435].

Collaborative approaches, as utilised in this work, are becoming much more common and evidence is emerging in support of their success. Combined industry-academia projects operating between 1991–2010 were found to have an elevated pre-clinical to Phase I success rate of 36.7 % compared to single-sector led studies[434].

5.4.2 Library size and the efficient exploration of chemical space

The efficacy of small-fragment libraries can be evaluated by their success in producing candidates against targets and by their efficient coverage of chemical space. The 300 fragment library screened in this work is small compared to those used in the pharmaceutical industry. Typically these contain around 2 000 compounds and can be used to rapidly produce new lead compounds as required[436–438]. Despite this, several other factors determinant of library success should be considered and ultimately an argument could be made in favour of the small size of the screen.

It has been suggested that one of the greatest pitfalls in fragment screening is the urge to prematurely increase compound complexity aiming to drive

down IC50 values rather than properly characterising candidates[439]. This leads to less efficient screening of the chemical space and increases the chance for false positives results[440]. The small size of this study and lack of access to compound synthesis has prevented such unnecessary increases in complexity.

Fragment properties are also crucial to the exploration of chemical space. Those used in this work were assessed for druglikeness using multiple metrics, predominantly the Ro3 and Ro5. Due to the harsh cuttoffs applied to fragments using these metrics, they are increasingly being challenged as to their validity[441]. Despite this, it has been argued that when applied sensibly, the real strength of these metrics is in constraining library complexity, whilst also keeping the fit within optimal drug-like parameters[439].

5.4.3   Small-fragment screening to identify inhibitors of FL-HelQ

A 2016 report found that hit-rates from biochemical screens carried out by leading pharmaceutical companies, including AstraZeneca and GSK, fell in the range of 7–26.2 %[442]. As such, the observed hit rate of 19% from the initial small-fragment library is consistent with previous outcomes.

Despite this success, the data output of the screen was limited. Using agarose gels to assess fragment performance was both low-throughput and low-resolution, presenting a problem for quantitative analysis. Determining the band intensity of each lane used for the screen would have been incredibly time consuming and attempting this with a several-fold larger library would be impractical. This meant that compounds were assessed qualitatively which, can be seen in Fig. 5.6 to lead to several false-positive results.

A better approach would have been methods with a quantitative rather than visual output. For example, several studies have utilised high-sensitivity techniques such as liquid-chromatographic mass-spectrometry (LC-MS) to carry out high-throughput screening of fragment libraries[443,444]. Another

approach with increased throughput is the use of Förster resonance energy transfer (FRET) which allows assays to be conducted in microplates ranging from 96–1536 wells[445]. This approach also raises the possibility of kinetic and interaction-based measurements using time-resolved FRET (TR-FRET)[446,447].

5.4.4 Unsuccessful hit and lead inhibitor compounds

Assays following the initial small-fragment screen aimed to confirm fragment inhibition of FL-HelQ and succeeded in detecting multiple false-positive results. As discussed in section 5.4.3, this is likely due to the qualitative nature of the screen. Several of the confirmed inhibitors were observed to precipitate during the assay, resulting in their removal from development.

Aggregation is one of the most common sources of false-positive results for small-fragment screening[448,449]. This was surprising as the commercially available library used had been pre-screened to limit the presence of aggregating compounds. This may be explained by experimental factors as most biochemical assays require high concentrations of fragments due to their weak protein:ligand interactions[450]. It is also unknown if assay buffer components could play a role in promoting aggregation in some inhibitors.

The second phase of screening assayed 27 lead compounds aiming to improve on the potency of the original fragments. Surprisingly, the majority of compounds failed to produce any inhibition at all. Further exploration determined that several fragments achieved inhibition at reduced concentrations of ATP. This could explain discrepancies between the results reported in this work and those of the supplier. It may also allude to such compounds functioning as competitive inhibitors of ATP.

Competition with ATP does not necessarily preclude a compound from development, but does require additional caution. Due to the ubiquity of ATP usage within cells, competitive inhibitors may promiscuously interact

with other proteins. This has been observed in the development of kinase inhibitors as the ATP binding pocket is often highly conserved[451]. Therefore, strict specificity testing must be utilised for inhibitors which function by this mechanism.

5.4.5 Identification of promising candidates for the inhibition of FL-HelQ

Of the 327 total compounds studied across two phases of screening, five promising inhibitors were identified, each assessed in turn in sections below. The efficacy of these compounds can be considered using several metrics including: chemical properties, selectivity and potency. The most valuable metric established for select compounds in this work was the IC50. For small-fragment based screens, IC50 values close to 1 mM are considered promising[452]. As development progresses, this value typically decreases, reaching the nM to low μM range. Cisplatin serves as a good example of a fully developed drug, having a maximum plasma concentration of 14.4 μM[453].

5.4.5.1 Compound N107

Compound N107 was the most promising inhibitor identified from the initial small-fragment screen. Results obtained from titration experiments show that the compound greatly outperformed all other candidates, achieving inhibition at several-fold lower concentrations. This is reflected in its low IC50 value of 557 μM ranking it as the second most potent inhibitor across all screening. The strong performance of this compound may be tempered by results from DNA-binding and ATPase assays which demonstrated significant impact on FL-HelQ:DNA interaction and ATP turnover, respectively, possibly suggesting less desirable modes of action. Conversely, it could be considered that as HelQ is an ssDNA binding dependent ATPase , an inhibitor that prevents DNA binding will also inhibit ATPase activity. This

could mean that rather than a non-specific inhibitor that is competitive with ATP, the compound may be very useful as a way to prevent DNA binding. Further testing is required to ascertain mode-of-action and specificity, despite good performance in assays comparing FL-HelQ and RecQ.

5.4.5.2 Compounds N326 and N144

These two compounds consistently demonstrated similar activity across all assays. Results obtained from titration experiments displayed closely related inhibition profiles across the concentration gradient. Also similar were the IC50 values for the two inhibitors with N326 calculated at 2.87 mM and N144 at 2.81 mM. These values are several-fold lower in potency when compared to N107. Little difference was observed between the two candidates in ATPase assays. The only separation between the two is results obtained for DNA-binding in which N326 was observed to have less impact on FL-HelQ than N144. This could hint at differing modes of action, with N326 a competitive inhibitor of ATP and N144 potentially having a secondary effect on ATPase activity by disrupting DNA-binding. While it is too early to tell, were further profiling to be carried out, such a difference could be a deciding factor in whether or not to continue development of these compounds.

5.4.5.3 Compound N274

From the second phase of testing, this compound was by a small margin the highest performing. Results from titration assays show a similar inhibition profile to N186, although N274 appears to have a greater potency at higher concentrations. The calculated IC50 value of 547.8 μM places this compound as the most potent inhibitor observed in this study across both phases of screening, by a narrow margin. As with compound N107, there is some doubt as to the mode-of-action of the compound as in ATPase assays, the candidate significantly reduced FL-HelQ ATP turnover.

5.4.5.4 Compound N186

This compound was the second promising lead identified during the second phase of testing, performing only marginally less effectively than candidate N274. The candidate was calculated to have an IC50 of 564.9 μM, making it the third most potent inhibitor observed across all screening. As seen with compounds N107 and N274, this candidate also significantly impacted the ATPase activity of FL-HelQ during assays meaning that further development should be carried out with caution, taking care to undertake thorough specificity screening.

5.4.6 Study limitations for determining inhibitor specificity and mode-of-action

This chapter partially profiled candidate inhibitors of FL-HelQ, but little was determined regarding probable mode-of-action. Supershift assays were utilised to determine the impact of inhibitors on FL-HelQ:DNA binding. While these assays are a good way of visualising interactions, they may be insensitive to transient or low-affinity interactions[454]. An alternative would be to use high-sensitivity biophysical techniques with increased throughput such as microscale thermophoresis (MST) and fluorescence anisotropy or polarisation[455,456].

ATPase assays were used as a secondary method of confirming inhibition. Their inclusion was in part informed by their common use in fragment-based drug discovery, however the results produced were often inconsistent with those obtained from unwinding assays. This could either indicate fragment mode-of-inhibition, or be the result of experimental variance. Also of note is the potential for ATPase assays to produce false-positive or false-negative results. Major decreases in ATP turnover may signal strong inhibition only for further profiling to reveal compounds as ATP competitors with a high likelihood of promiscuity. Conversely, compounds bound allosterically to

FL-HelQ in such a way as to prevent translocation along the DNA substrate, but not interfere with ATP hydrolysis, could be disregarded as candidates.

To truly determine inhibition mechanisms, kinetic assays would be required to determine baseline values for metrics such as the K_m and V_{max} of FL-HelQ. This could then be used to assess the same assays supplemented with inhibitor candidates. Common approaches to these kinds of assays are the aforementioned MST and fluorescence anisotropy or polarisation.[455,456]. Profiling in this manner would further streamline the development of compounds by selecting those with a higher probability of specific inhibition.

5.5   Future Perspectives

This work successfully progressed from the initial stages of small-molecule screening to the development and characterisation of lead compounds. However, much work remains to develop inhibitor candidates to a stage suitable for progression along the drug-discovery pipeline.

5.5.1   Increasing the throughput and power of library screening

This work was predominantly conducted using assays resolved by native-PAGE. While this is a powerful technique for visualising assay outcomes, the throughput is incredibly low with a maximum of seventeen samples per gel plus controls. Microplate-based assays utilising techniques such as FRET would increase throughput, and in doing so the efficiency of exploring chemical-space. This should in turn increase the likelihood of identifying successful candidates for further development.

5.5.2   Using kinetic assays to fully assess inhibitor mode-of-action

This work generated a range of useful data for characterising inhibitor candidates but is limited in that it can only allude to, rather than definitively determine, candidate mode-of-action. This is evident in comparing unwind-

ing and ATPase assays across sections 5.3.6 and 5.3.8 where candidate N107 appears to impact ATPase activity while N144 does not. Kinetics-based methods would be able to determine mode-of-action, whether they were biochemical assays, akin to time-courses seen in this work, or biophysical approaches such as microscale thermophoresis.

5.5.3 Parallel screening of inhibitors using human DNA repair proteins to assess specificity

Candidate specificity to FL-HelQ inhibition is briefly touched upon in this work, but was not adequately explored further. Due to its availability in the lab, *E. coli* RecQ was used to gauge compound promiscuity, but the insight we can gain from this is limited. While sufficient as a preliminary check, a more comprehensive approach would be to test human DNA repair proteins. This could include pathologically related RecQ-like helicases such as BLM and WRN, as well as other classes of protein from a variety of pathways, to ensure the specific interaction of candidates with HelQ.

5.5.4 Exploring a range of assay conditions for more physiologically relevant data

The disparity in observed inhibition between our own assays and those of the second phase inhibitor manufacturer, Sygnature Discovery, highlighted the importance of assay design. Troubleshooting assays demonstrated that ATP concentration could influence the confirmation of inhibition. Therefore, it may be prudent in future to screen libraries at several physiologically relevant concentrations of ATP to capture all possible candidates. This could be applied across a range of different assays such as those used in this work i.e. unwinding, supershift and ATPase assays.

5.5.5 In-depth *In vitro* characterisation of inhibitor action using human cell lines

This work was conducted using biochemical approaches and lacks assays to assess drug response in human cell lines. Expanding on this work would require determining the cytotoxicity of inhibitors and calculation of cellular IC50s for each compound in cells. Further development would explore the cellular response to inhibitor compounds in the presence and absence of DNA damage. The obvious choice of damaging agent would be ICL-inducing chemotherapeutic drugs such as mitomycin C and cisplatin as was briefly explored in Chapter 4. It would also be useful to include compounds that do not induce ICLs such as aphidicolin and hydroxurea for comparison.

Such work would initially be carried out using 2D models and immortalised cell lines. Eventually, a move to more physiologically representative models would be the most sensible approach to achieve relevant data. This could include the development and use of 3D models such as spheroids or cells encapsulated in matrices[457].

Another approach to gathering powerful physiological data would be the use of phenotypically relevant cell lines. A good example of this is the OVCAR-8 cell line derived from a patient with high grade ovarian serous adenocarcinoma, which has been use to study ovarian cancer for decades as seen in Johnson *et al.* and Alley *et al.*[458,459].

5.6   Chapter Summary

The focus of this chapter was on the screening of small-fragments to identify inhibitors of HelQ, combining the principles of early-stage drug-discovery with biochemical approaches to characterise candidates and assess their potency. The progress made during this satisfied the initial aims described in section 5.2 and to a limited degree explores beyond them by troubleshooting issues related to hit detection and thus informing approaches to assay design. Overall, this work has shown:

- That basic drug-discovery principles can be used in conjunction with small-molecule libraries to identify inhibitors for HelQ.

- The characterisation of promising candidates and their development towards lead compounds.

- Basic specificity and mode-of-action testing to identify promising inhibitors in both the initial hit and lead compound phase of pre-clinical testing.

6
Conclusion and Perspectives

6.1 Summary of the research

HelQ is an SF2, Ski-2-like helicase with ATP and ssDNA-dependent activity that translocates along DNA with 3′ to 5′ polarity. Research has clearly established a role for HelQ in maintaining genome stability, implicating it in the promotion of DNA repair at stalled replication forks through HR-related pathways[225–229]. Several studies have also implicated HelQ as a participant in the repair of ICLs, working synergistically with the FA pathway[212,227,228,250]. The precise role and mechanism of HelQ in maintaining genome stability remains unclear, but research indicates a caretaker role for the protein which could prove significant to the development of cancers[255–257,406–408].

Research has also associated HelQ, through its synergy with the FA repair pathway, with the efficiency of CRISPR-mediated gene-editing using ssODNs[212]. In a separate vein, the protein has also been implicated as a prognostic biomarker for the outcomes of several different types of cancer, and a promising target in the treatment of platinum-resistant ovarian tumours[258–260]. This work sought to explore the role of HelQ in DNA repair and the maintenance of genome-stability, using biochemical assays and model cell-lines, with specific focus on its participation in CRISPR-mediated gene-editing and cancers.

216

Overall, the key aims of this research project were::

- To study the interaction of HelQ with Cas9 R-loops using synthetic DNA substrates and supercoiled plasmids

- To study the interaction of HelQ with non-CRISPR R-loops using synthetic DNA and RNA substrates.

- To adapt the methods of Sansbury et al.[342,343] to create an *in vitro* cell-free model to study the impact of DDR proteins on editing efficiency, with HelQ as a model.

- The screening of a small fragment library to identify and characterise putative inhibitors of FL-HelQ for further development

6.1.1   HelQ does not appear to remove roadblocks during gene-editing

Research using CRISPRi screens successfully established that Cas9-induced editing using ssODNs integrates template sequence into the genome via the FA pathway[212]. The research identified a possible role for HelQ in gene-editing demonstrating that knockdown of the protein, as well as the FANCI-FANCD2 complex with which it interacts, substantially decreased SSTR efficiency[212].

Previous work has established a role for several helicases in the removal of protein roadblocks on DNA, including one study which specifically showed the RecG-mediated removal of Cascade from supercoiled plasmid DNA[460]. Using biochemical approaches, including functional unwinding assays and EMSAs, we explored the activity of HelQ in removing a dCas9 roadblock from both synthetic model substrates and plasmid DNA. Ultimately this proved unsuccessful, with the protein unable to remove dCas9 from either of the substrates used. This is however supported by recently published work which demonstrated that HelQ was unable to remove other protein roadblocks from DNA[229].

217

Emerging data on the function of other helicases, such as the 5′–3′ translocating Pif1, suggests that some proteins facilitate roadblock bypass rather than displacement[370]. This raises the possibility that while HelQ does cannot remove a roadblock in isolation, it may instead have another function that the assays used were not able to observe. In future, it may be more informative to explore the role of HelQ using more complex assays taking into account previously established interactions with proteins such as RPA, FANCD2, and the polymerase Pol ν[371].

6.1.2 Cell-free extracts have potential as a model for DNA-repair during gene-editing

To further develop assessing the function and significance of HelQ in gene-editing, template integration assays in a cell-free extract were reconstituted from the work of Sansbury et al.[342,343]. These aimed to model gene-editing by preparing cell-free extracts from both wild-type and HelQ-depleted cell lines. While extracts were successfully produced and integration demonstrated, circumstances meant that the system was not utilised to explore the impact of HelQ-depletion, and subsequent rescue with recombinant protein, on integration efficiency.

Colony PCR screening of plasmid DNA recovered from integration assays demonstrated successful template insertion into pUC19 with a heavy bias towards Cas12a. This is unsurprising given that Cas12a generates overhangs upon cutting, providing one complementary end to ssODNs and two to dsDNA templates, likely providing an advantage relative to the blunt cuts generated by Cas9. The fact that apparent integrations into Cas9-cleaved DNA were observed at all contrasts with the observations made by Sansbury et al.[342]. This highlights the importance of sequencing the PCR products amplified from potential integrations to determine whether or not they are true insertions.

Several issues were identified with the system including extract yield when following the base protocol, and the practicalities of scaling up production when testing multiple different cell lines. This will need future development to overcome, possibly by transitioning to suspension culture where possible, as is often utilised by other cell-free systems including those used for *in vitro* protein synthesis[372,373]. Beyond this, there remains some ambiguity as to whether integrations are truly extract-dependent. The control assays used in this work fall short of being able to definitively confirm this and as such future experiments should include these to increase confidence in the results.

The CFE system shows great potential for the biochemical screening of protein interactions with DNA, provided a knockout of the GOI is available. The degree of control over assay conditions facilitates experiments such as rescue assays in which recombinant HelQ would be supplemented in KO extracts. The assay could be developed further by transitioning from blue-white screens to fluorescent reporter assays, which may facilitate the gathering of higher-throughput, quantitative data using techniques such as flow-cytometry. Finally, the system could merge the existing, separate cleavage and integration steps into one to provide a complete system in which to model gene-editing.

6.1.3   A potential role for HelQ in the resolution of DNA:RNA hybrids?

Alongside researching the involvement of HelQ in DNA repair during CRISPR-mediated gene-editing, a role for the protein in resolving DNA:RNA hybrid structures such as R-loops was also explored. Multiple helicases that participate in repair have previously been shown to interact with R-loops, although it remains ambiguous whether their formation is a cause or consequence of DNA damage[360–363]. Of particular interest was the observation that the helicase domain of polymerase PolQ was capable of unwinding

DNA:RNA hybrids and that the protein takes part in RNA-templated DNA-repair[364,365]. HelQ has high sequence homology to the helicase domain of PolQ and as such we hypothesised that it may also be capable of this activity.

To ascertain whether HelQ was able to unwind hybrid structures, biochemical screening for unwinding activity was conducted against several DNA and DNA:RNA model substrates. The results for forked substrates were marred by a lack of hybrid fork comparable to the control assay and unplanned secondary reactions caused by alternate, free 3′ ends which the helicase could unwind. Analysis of D- and R-loop substrates demonstrated a reduced unwinding preference for those with a free 5′ end, as expected, and some ability to unwind R-loops with a free 3′ end, although this requires further testing to confirm. Loop substrate assays were also somewhat limited as unwinding of a control substrate with no free ends suggested that HelQ was translocating along the ssDNA of the displaced DNA strand, again facilitating a secondary unwinding reaction.

In parallel, immunofluorescence staining was used to observe whether DNA:RNA hybrids accumulated in cell-models with HelQ depletion, and when complemented with an ATPase-inactive mutant. In both cases a significant increase in signal intensity was observed relative to the parental U2OS cell line. The effect observed in HelQ-depleted cells suggests that the protein may play a role in resolving hybrids structures. The increase in signal was most severe in the mutant-complemented cells, which may suggest that inactive HelQ is interfering with repair processes, causing a more severe accumulation of hybrids than in the complete absence of the protein. This needs to be followed up with more extensive experiments such as a rescue in which HelQ-depleted cells are transfected with a plasmid expressing wild-type protein to observe whether this reverses the phenotype. Additional experiments utilising a range of antibodies for different markers would also be beneficial to overcome the limitations of the S9.6 antibody.

6.1.4 Developing a cell-based model phenotype for HelQ-depletion and ICL-inducing agent sensitivity

Previous research has implicated HelQ in multiple DNA repair pathways including HR and FA[225–228]. Crucial insight into the function of the protein has been made using cell-based models, including colocalisation of the protein at stalled replication forks and sites of DNA damage, promotion of HR, synergy with FA, and a phenotype of sensitivity to ICL-inducing agents upon HelQ depletion[226]. Functional genomics in cell models represents a powerful approach to gain insight into important cellular interactions. In this work, attempts were made to study the impact of toxic agents on HelQ-deficient or mutated cells to further elucidate the pathways and interactions that rely on the protein.

Characterisation of cell models identified a mild growth phenotype upon HelQ depletion as well as a mild morphology phenotype for one HelQ$^{-/-}$ clone (R-172). The most severe effect was observed in cells complemented with the ATPase mutant HelQ^{K365M} (U2OS), while a milder phenotype was seen from HelQ^{D463A} (RKO). Work observing the effect of toxic agents such as cisplatin and MMC on cell viability were unable to replicate the previous observations of Takata et al.[227] and Adelman et al.[228]. This is likely due to the method used to measure viability which may underestimate the impact of toxic agents on cells. No effect on cell viability was observed when screening cells against aphidicolin and hydroxyurea.

Continuing this work would require significant assay development, beginning with a change in the method used to measure cell viability. This should also be paired with other methods to measure DNA damage more directly. Beyond this, a wider exploration of DNA damaging agents and complementation assays with wild-type protein to rescue phenotypes, if possible, may deepen our understanding of the function of HelQ in DNA-damage repair.

6.1.5 Small-fragment screening to identify inhibitors of HelQ

The involvement of HelQ with the HR and FA pathways indicates a caretaker role for the protein, which could prove significant to the development of cancers. This has been demonstrated by studies implicating HelQ in the development of cancers of the ovaries, testes, head, and neck[406–408]. HelQ has also been identified as a prognostic biomarker for the outcomes of several types of cancer, including overall survival in cases of chronic lymphocytic leukaemia.[258,259]. Furthermore, the protein has been associated with ovarian cancer through both genome-wide association studies and bioinformatics[399,408,420,421]. This was directly observed in a sub-set of ovarian tumours in which HelQ overexpression resulted in resistance to chemotherapuetic treatment with Cisplatin[260].

Through applying drug discovery principles, this work aimed to develop inhibitors against HelQ using small-molecule screening techniques supported by biochemical assays to characterise candidates. An initial screen of 300 small-molecules identified 55 candidate compounds, equating to a 19% 'hit' rate, which is in line with previous commercial screens[442]. Of these, 19 were further screened to check for false-positives, resulting in 7 promising compounds being biochemically characterised. A combination of unwinding, EMSA, and ATPase assays identified two compounds: N107 and N144 as the most potent candidates, with IC50's of 0.557 μM and 2.81 μM, respectively. However, both of these compounds may indicate a mode of action reliant to some extent on competition with ATP. While this does not exclude them from being effective inhibitors, extensive specificity testing would be required to ensure that they are not promiscuous. Some insight into the specificity of compounds was gained by activity screening against a C-terminal fragment of HelQ and *E. coli* RecQ. The data demonstrated that the most promising compounds produced similar inhibitory effects against both HelQ and C-HelQ, whilst not impacting on the activity of RecQ.

Despite successfully identifying multiple promising hits, the data output of the initial screen was limited by the use of low resolution agarose gels, which led to qualitative assessment of results. Future screens to identify inhibitors against HelQ could instead utilise quantitative techniques such as LC-MS to carry out high-throughput screening or microplate-based FRET assays[443–445]. The latter approach also raises the possibility of kinetic and interaction-based measurements using TR-FRET[446,447].

6.1.6 The development of promising candidate molecules

Following the first phase of testing a second phase of inhibitors, comprising 27 new compounds, was produced using candidates N078, N107 and N214 as scaffolds. Surprisingly, when assayed to check for false-positive results, it was found that the majority of this series displayed no inhibition of HelQ. Deeper characterisation determined that several fragments achieved inhibition at reduced concentrations of ATP, indicating that these compounds function as competitive inhibitors of ATP. Among the second phase candidates, three compounds, N186, N274, and N326, were identified which performed similarly to the first phase with respective IC50s of 0.565 μM, 548 μM, and 2.87 μM.

This work succeeded in identifying inhibitors of HelQ but only basic data was obtained regarding probable mode-of-action. To further characterise this, kinetic assays would be required to determine metrics such as the K_m and V_{max} of HelQ plus and minus putative inhibitors. The data obtained are also limited in their physiological evidence, comprised entirely of biochemical screens. Future work should also incorporate work in human cell models to determine inhibitor cytotoxicity and cellular IC50 values. This could be developed by assessing inhibitors in the presence of DNA damaging agents and eventually alongside cisplatin, to which HelQ overexpression has been observed to cause resistance[258,260].

6.2 Outlook

6.2.0.1 Understanding the function of HelQ and its significance in CRISPR-mediated gene-editing

Despite extensive efforts, the precise role of HelQ in DNA-damage repair remains elusive. This study has taken a focussed approach to study the role of HelQ in DNA-damage repair and the significance of this during CRISPR-mediated gene editing. While this did not produce clear answers, the progress made has raised fresh questions and opportunities to further explore the function of HelQ.

HelQ in isolation has been demonstrated as incapable of directly removing dCas9 roadblocks. Research also indicates that the protein does not interact with other protein:DNA roadblocks[229]. Further work however could aim to determine the possible existence of any indirect interactions between HelQ and such structures. This could include participation in roadblock bypass as is the case with other proteins such as Pif1. Cell-free extracts present a promising biochemical model for this kind of study, but significant development is needed to truly explore the mechanism of DNA-repair during gene-editing. It is possible that DNA-repair systems do not directly interact at all with the gene-editing machinery and probable that the significance of HelQ during this process is not unique, but rather indicative of its function in SSTR. Repair by SSTR is still not fully understood, but further research into its mechanism may reveal a role for HelQ as well as the context in which it participates. Through this, it may be possible to infer more about the overall function of the protein, as well as its significance to gene-editing.

Ultimately, biochemical approaches alone are insufficient to unpick the function and context of HelQ participation in DNA-repair and should be combined with biophysical techniques and cell-based models. This work made attempts to reproduce a sensitivity phenotype to ICL-inducing agents.

With further development, this could be extended to model the rescue of HelQ expression both transiently, from plasmid DNA, and stably, from a knock-in of the gene in depleted cells. Cell-models have also proven useful in identifying a possible corelation of HelQ with R-loops. This requires careful exploration to both confirm that an association between the two. Further research into DNA:RNA hybrids in general will also be useful in determining whether their accumulation is a cause or consequence of elevated DNA damage.

6.2.0.2 Developing inhibitors against HelQ for combination therapies in platinum-resistant tumours

While the successful identification and early characterisation of small-molecule inhibitors against HelQ is an exciting development, it is just the first step on the road to a complete drug compound. Substantial work remains before progress along the drug-discovery pipeline becomes a reality. Larger library screens using higher throughput technologies, assay development to eliminate false-positive results, and the incorporation of cell-based models to obtain more physiologically relevant data are all approaches which will need to be explored to maximise the potency and specificity of candidate compounds. In this, the work presented in establishing a model for HelQ response to DNA damaging agents could contribute, providing a platform to test drug compounds for both general cytotoxicity in a parental cell-line and also for a phenotype in disease-like models.

Aside from the work presented here, several powerful tools remain by which the role of HelQ could be explored. Driven by next-generation sequencing technologies as well as screens using RNAi and CRISPRi, an ever-expanding library of publicly available tumour meta-data and protein interactions is providing increasing power to bioinformatic approaches. If combined with a crystal structure for HelQ, this would enhance modelling

for structural, mechanistic, and proteomic research. Moreover, coupling this approach to large-scale synthetic lethality screens would facilitate the construction of a more complete functional network for the protein, as well as driving future drug discovery through the identification of promising new targets and enhanced modelling for the docking of putative drug candidates. Several of these approaches would not only contribute to the development of inhibitors against HelQ, but also to our understanding of the structure and function of the protein. This would better place us to determine the context of the protein in DNA damage, recombination-coupled repair, and through this, its significance in gene-editing.

6.3   Thesis summary

This thesis has explored the role of HelQ, a helicase implicated in multiple DNA repair pathways, platinum-resistant cancers, and high-efficiency, CRISPR-mediated gene editing. Through a multifaceted approach we have:

- Determined that HelQ is unable to remove Cas-protein roadblocks from DNA.

- Successfully reconstituted a cell-free system for modelling Cas-protein mediated integration of donor DNA.

- Identified a possible correlation between HelQ-deficiency in human cells and the accumulation of DNA:RNA hybrids in nuclei.

- Begun to characterise the impact of HelQ-deficiency in cells, laying the groundwork for complementation/rescue studies.

- Applied basic drug-discovery principles in conjunction with small-molecule screening to identify putative inhibitors for HelQ.

- Characterised candidates using basic specificity and mode-of-action testing to identify promising inhibitors in both the initial hit and lead compound phase of pre-clinical testing.

References

1. Bateson, W. Hybridisation and cross-breeding as a method of scientific investigation. *J. R. Hortic. Soc.* **1900**, 59–66.

2. Bateson, B. *William Bateson, Nat.*; Cambridge University Press: Cambridge, 2010; pp 161–170.

3. Muller, H. J. Artificial transmutation of the gene. *Science (80-.).* **1927**, *66*, 84–87.

4. Auerbach, C.; Robson, J. M. Chemical Production of Mutations. *Nature* **1946**, *157*, 302–302.

5. Rudin, N.; Haber, J. E. Efficient repair of HO-induced chromosomal breaks in Saccharomyces cerevisiae by recombination between flanking homologous sequences. *Mol. Cell. Biol.* **1988**, *8*, 3918–3928.

6. Rausch, H.; Lehmann, M. Structural analysis of the actinophage phi C31 attachment site. *Nucleic Acids Res.* **1991**, *19*, 5187–9.

7. Nathwani, A. C. et al. Long-Term Safety and Efficacy of Factor IX Gene Therapy in Hemophilia B. *N. Engl. J. Med.* **2014**, *371*, 1994–2004.

8. Griesenbach, U.; Pytel, K. M.; Alton, E. W. Cystic Fibrosis Gene Therapy in the UK and Elsewhere. *Hum. Gene Ther.* **2015**, *26*, 266–275.

9. Orr-Weaver, T. L.; Szostak, J. W.; Rothstein, R. J. Yeast transformation: a model system for the study of recombination. *Proc. Natl. Acad. Sci.* **1981**, *78*, 6354–6358.

10. Rothstein, R. J. *Methods Enzymol.*; 1983; pp 202–211.

11. Hinnen, A.; Hicks, J. B.; Fink, G. R. Transformation of yeast. *Proc. Natl. Acad. Sci.* **1978**, *75*, 1929–1933.

12. THOMAS, K. High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **1986**, *44*, 419–428.

13. Smithies, O.; Gregg, R. G.; Boggs, S. S.; Koralewski, M. A.; Kucherlapati, R. S. Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination. *Nature* **1985**, *317*, 230–234.

14. Capecchi, M. R. Generating mice with targeted mutations. *Nat. Med.* **2001**, *7*, 1086–1090.

15. Gerlai, R. Gene Targeting Using Homologous Recombination in Embryonic Stem Cells: The Future for Behavior Genetics? *Front. Genet.* **2016**, *7*, 1–10.

16. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci.* **1950**, *36*, 344–355.

17. Munoz-Lopez, M.; Garcia-Perez, J. DNA Transposons: Nature and Applications in Genomics. *Curr. Genomics* **2010**, *11*, 115–128.

18. Way, J.; Davis, M.; Morisato, D.; Roberts, D.; Kleckner, N. New Tn10 derivatives

for transposon mutagenesis and for construction of lacZ operon fusions by transposition. *Gene* **1984**, *32*, 369–379.

19. Lamberg, A.; Nieminen, S.; Qiao, M.; Savilahti, H. Efficient Insertion Mutagenesis Strategy for Bacterial Genomes Involving Electroporation of In Vitro-Assembled DNA Transposition Complexes of Bacteriophage Mu. *Appl. Environ. Microbiol.* **2002**, *68*, 705–712.

20. Grabundzija, I.; Irgang, M.; Mátés, L.; Belay, E.; Matrai, J.; Gogol-Döring, A.; Kawakami, K.; Chen, W.; Ruiz, P.; Chuah, M. K.; VandenDriessche, T.; Izsvák, Z.; Ivics, Z. Comparative Analysis of Transposable Element Vector Systems in Human Cells. *Mol. Ther.* **2010**, *18*, 1200–1209.

21. Strecker, J.; Ladha, A.; Gardner, Z.; Schmid-Burgk, J. L.; Makarova, K. S.; Koonin, E. V.; Zhang, F. RNA-guided DNA insertion with CRISPR-associated transposases. *Science (80-.).* **2019**, *365*, 48–53.

22. Bhatt, S.; Chalmers, R. Targeted DNA transposition in vitro using a dCas9-transposase fusion protein. *Nucleic Acids Res.* **2019**, *47*, 8126–8135.

23. Lau, C. H.; Bolt, E. L. Integration of diverse DNA substrates by a casposase can be targeted to R-loops in vitro by its fusion to Cas9. *Biosci. Rep.* **2021**, *41*, 1–14.

24. Owens, J. B.; Mauro, D.; Stoytchev, I.; Bhakta, M. S.; Kim, M.-S.; Segal, D. J.; Moisyadi, S. Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. *Nucleic Acids Res.* **2013**, *41*, 9197–9207.

25. Hoess, R.; Wierzbicki, A.; Abremski, K. Formation of small circular DNA molecules via an in vitro site-specific recombination system. *Gene* **1985**, *40*, 325–329.

26. Sauer, B.; Henderson, N. Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc. Natl. Acad. Sci.* **1988**, *85*, 5166–5170.

27. Gronostajski, R. M.; Sadowski, P. D. The FLP recombinase of the Saccharomyces cerevisiae 2 microns plasmid attaches covalently to DNA via a phosphotyrosyl linkage. *Mol. Cell. Biol.* **1985**, *5*, 3274–3279.

28. Amin, A.; Roca, H.; Luetke, K.; Sadowski, P. D. Synapsis, strand scission, and strand exchange induced by the FLP recombinase: analysis with half-FRT sites. *Mol. Cell. Biol.* **1991**, *11*, 4497–4508.

29. Song, A. J.; Palmiter, R. D. Detecting and Avoiding Problems When Using the Cre–lox System. *Trends Genet.* **2018**, *34*, 333–340.

30. Nagy, A. Cre recombinase: The universal reagent for genome tailoring. *genesis* **2000**, *26*, 99–109.

31. Branda, C. S.; Dymecki, S. M. Talking about a Revolution. *Dev. Cell* **2004**, *6*, 7–28.

32. Buchholz, F.; Angrand, P.-O.; Stewart, A. F. Improved properties of FLP recombinase evolved by cycling mutagenesis. *Nat. Biotechnol.* **1998**, *16*, 657–662.

33. Stoddard, B. L. Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mob. DNA* **2014**, *5*, 7.

34. Bos, J. L.; Heyting, C.; Borst, P.; Arnberg, A. C.; Bruggen, E. F. J. V. An insert in the single gene for the large ribosomal RNA in yeast mitochondrial DNA.

Nature **1978**, *275*, 336–338.

35. Choulika, A.; Perrin, A.; Dujon, B.; Nicolas, J. F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol. Cell. Biol.* **1995**, *15*, 1968–1973.

36. Lazowska, J.; Jacq, C.; Slonimski, P. P. Sequence of introns and flanking exons in wild-type and box3 mutants of cytochrome b reveals an interlaced splicing protein coded by an intron. *Cell* **1980**, *22*, 333–348.

37. Orlowski, J.; Boniecki, M.; Bujnicki, J. M. I-Ssp6803I: the first homing endonuclease from the PD-(D/E)XK superfamily exhibits an unusual mode of DNA recognition. *Bioinformatics* **2007**, *23*, 527–530.

38. Stoddard, B. L. Homing Endonucleases: From Microbial Genetic Invaders to Reagents for Targeted DNA Modification. *Structure* **2011**, *19*, 7–15.

39. Thierry, A.; Dujon, B. Nested chromosomal fragmentation in yeast using the meganuclease I- Sce I: a new method for physical mapping of eukaryotic genomes. *Nucleic Acids Res.* **1992**, *20*, 5625–5631.

40. Haugen, P. The spread of LAGLIDADG homing endonuclease genes in rDNA. *Nucleic Acids Res.* **2004**, *32*, 2049–2057.

41. Bellaiche, Y.; Mogila, V.; Perrimon, N. I-SceI Endonuclease, a New Tool for Studying DNA Double-Strand Break Repair Mechanisms in Drosophila. *Genetics* **1999**, *152*, 1037–1044.

42. Grizot, S.; Smith, J.; Daboussi, F.; Prieto, J.; Redondo, P.; Merino, N.; Villate, M.; Thomas, S.; Lemaire, L.; Montoya, G.; Blanco, F. J.; Pâques, F.; Duchateau, P. Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.* **2009**, *37*, 5405–5419.

43. Rouet, P.; Smih, F.; Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol. Cell. Biol.* **1994**, *14*, 8096–8106.

44. Silva, G.; Poirot, L.; Galetto, R.; Smith, J.; Montoya, G.; Duchateau, P.; Paques, F. Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy. *Curr. Gene Ther.* **2011**, *11*, 11–27.

45. Chevalier, B. S.; Kortemme, T.; Chadsey, M. S.; Baker, D.; Monnat, R. J.; Stoddard, B. L. Design, Activity, and Structure of a Highly Specific Artificial Endonuclease. *Mol. Cell* **2002**, *10*, 895–905.

46. Kim, Y. G.; Cha, J.; Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci.* **1996**, *93*, 1156–1160.

47. Miller, J.; McLachlan, A.; Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J.* **1985**, *4*, 1609–1614.

48. Persikov, A. V.; Singh, M. De novo prediction of DNA-binding specificities for Cys2His2 zinc finger proteins. *Nucleic Acids Res.* **2014**, *42*, 97–108.

49. Rubin, G. M. Comparative Genomics of the Eukaryotes. *Science (80-.).* **2000**, *287*, 2204–2215.

50. Gibson, T. J.; Postma, J. P.; Brown, R. S.; Argos, P. A model for the tertiary structure of the 28 residue DNA-binding motif ('zinc finger') common to many

eukaryotic transcriptional regulatory proteins. "*Protein Eng. Des. Sel.* **1988**, *2*, 209–218.

51. Zhang, W.; Xu, C.; Bian, C.; Tempel, W.; Crombet, L.; MacKenzie, F.; Min, J.; Liu, Z.; Qi, C. Crystal structure of the Cys2His2-type zinc finger domain of human DPF2. *Biochem. Biophys. Res. Commun.* **2011**, *413*, 58–61.

52. Buck-Koehntop, B. A.; Stanfield, R. L.; Ekiert, D. C.; Martinez-Yamout, M. A.; Dyson, H. J.; Wilson, I. A.; Wright, P. E. Molecular basis for recognition of methylated and specific DNA sequences by the zinc finger protein Kaiso. *Proc. Natl. Acad. Sci.* **2012**, *109*, 15229–15234.

53. Fairall, L.; Schwabe, J. W. R.; Chapman, L.; Finch, J. T.; Rhodes, D. The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* **1993**, *366*, 483–487.

54. Guo, J.; Gaj, T.; Barbas, C. F. Directed Evolution of an Enhanced and Highly Efficient FokI Cleavage Domain for Zinc Finger Nucleases. *J. Mol. Biol.* **2010**, *400*, 96–107.

55. Bibikova, M.; Carroll, D.; Segal, D. J.; Trautman, J. K.; Smith, J.; Kim, Y.-G.; Chandrasegaran, S. Stimulation of Homologous Recombination through Targeted Cleavage by Chimeric Nucleases. *Mol. Cell. Biol.* **2001**, *21*, 289–297.

56. Händel, E.-M.; Alwin, S.; Cathomen, T. Expanding or Restricting the Target Site Repertoire of Zinc-finger Nucleases: The Inter-domain Linker as a Major Determinant of Target Site Selectivity. *Mol. Ther.* **2009**, *17*, 104–111.

57. Tebas, P. et al. Gene Editing of CCR5 in Autologous CD4 T Cells of Persons Infected with HIV. *N. Engl. J. Med.* **2014**, *370*, 901–910.

58. Ramirez, C. L.; Foley, J. E.; Wright, D. A.; Müller-Lerch, F.; Rahman, S. H.; Cornu, T. I.; Winfrey, R. J.; Sander, J. D.; Fu, F.; Townsend, J. A.; Cathomen, T.; Voytas, D. F.; Joung, J. K. Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods* **2008**, *5*, 374–375.

59. Wu, J.; Kandavelou, K.; Chandrasegaran, S. Custom-designed zinc finger nucleases: What is next? *Cell. Mol. Life Sci.* **2007**, *64*, 2933–2944.

60. Gabriel, R. et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.* **2011**, *29*, 816–823.

61. Pattanayak, V.; Ramirez, C. L.; Joung, J. K.; Liu, D. R. Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat. Methods* **2011**, *8*, 765–770.

62. Elrod-Erickson, M.; Rould, M. A.; Nekludova, L.; Pabo, C. O. Zif268 protein-DIVA complex refined at 1.6Å: a model system for understanding zinc finger-DNA interactions. *Chem. Biol.* **1996**, *3*, 958–959.

63. Boch, J.; Scholze, H.; Schornack, S.; Landgraf, A.; Hahn, S.; Kay, S.; Lahaye, T.; Nickstadt, A.; Bonas, U. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science (80-.).* **2009**, *326*, 1509–1512.

64. Moscou, M. J.; Bogdanove, A. J. A simple cipher governs DNA recognition by TAL effectors. *Science (80-.).* **2009**, *326*, 1501.

65. Li, T.; Huang, S.; Jiang, W. Z.; Wright, D.; Spalding, M. H.; Weeks, D. P.; Yang, B. TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.* **2011**, *39*, 359–372.

66. Miller, J. C. et al. A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.* **2011**, *29*, 143–148.

67. Joung, J. K.; Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. *Nat. Rev. Mol. Cell Biol.* **2013**, *14*, 49–55.

68. Boch, J.; Bonas, U. Xanthomonas AvrBs3 Family-Type III Effectors: Discovery and Function. *Annu. Rev. Phytopathol.* **2010**, *48*, 419–436.

69. Mak, A. N.-S.; Bradley, P.; Bogdanove, A. J.; Stoddard, B. L. TAL effectors: function, structure, engineering and applications. *Curr. Opin. Struct. Biol.* **2013**, *23*, 93–99.

70. Bogdanove, A. J.; Voytas, D. F. TAL Effectors: Customizable Proteins for DNA Targeting. *Science (80-.).* **2011**, *333*, 1843–1846.

71. Gupta, R. M.; Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J. Clin. Invest.* **2014**, *124*, 4154–4161.

72. Osborn, M. J. et al. Evaluation of TCR Gene Editing Achieved by TALENs, CRISPR/Cas9, and megaTAL Nucleases. *Mol. Ther.* **2016**, *24*, 570–581.

73. Holkers, M.; Maggio, I.; Liu, J.; Janssen, J. M.; Miselli, F.; Mussolino, C.; Recchia, A.; Cathomen, T.; Gonçalves, M. A. F. V. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res.* **2013**, *41*, e63–e63.

74. Hille, F.; Richter, H.; Wong, S. P.; Bratovič, M.; Ressel, S.; Charpentier, E. The Biology of CRISPR-Cas: Backward and Forward. *Cell* **2018**, *172*, 1239–1259.

75. Killelea, T.; Hawkins, M.; Howard, J. L.; McGlynn, P.; Bolt, E. L. DNA replication roadblocks caused by Cascade interference complexes are alleviated by RecG DNA repair helicase. *RNA Biol.* **2019**, *16*, 543–548.

76. Ishino, Y.; Shinagawa, H.; Makino, K.; Amemura, M.; Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J. Bacteriol.* **1987**, *169*, 5429–33.

77. Groenen, P. M.; Bunschoten, A. E.; van Soolingen, D.; van Embden, J. D. Nature of DNA polymorphism in the direct repeat cluster of Mycobacterium tuberculosis; application for strain differentiation by a novel typing method. *Mol. Microbiol.* **1993**, *10*, 1057–65.

78. Mojica, F. J.; Ferrer, C.; Juez, G.; Rodríguez-Valera, F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Mol. Microbiol.* **1995**, *17*, 85–93.

79. Kamerbeek, J.; Schouls, L.; Kolk, A.; van Agterveld, M.; van Soolingen, D.; Kuijper, S.; Bunschoten, A.; Molhuizen, H.; Shaw, R.; Goyal, M.; van Embden, J. Simultaneous detection and strain differentiation of Mycobacterium tuberculosis for diagnosis and epidemiology. *J. Clin. Microbiol.* **1997**, *35*, 907–914.

80. Mojica, F. J. M.; Díez-Villaseñor, C.; García-Martínez, J.; Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.* **2005**, *60*, 174–82.

81. Makarova, K. S.; Grishin, N. V.; Shabalina, S. A.; Wolf, Y. I.; Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computa-

tional analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol. Direct* **2006**, *1*, 7.

82. Barrangou, R.; Fremaux, C.; Deveau, H.; Richards, M.; Boyaval, P.; Moineau, S.; Romero, D. A.; Horvath, P. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **2007**, *315*, 1709–12.

83. Marraffini, L. A.; Sontheimer, E. J. CRISPR Interference Limits Horizontal Gene Transfer in Staphylococci by Targeting DNA. *Science (80-.).* **2008**, *322*, 1843–1845.

84. Deltcheva, E.; Chylinski, K.; Sharma, C. M.; Gonzales, K.; Chao, Y.; Pirzada, Z. A.; Eckert, M. R.; Vogel, J.; Charpentier, E. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **2011**, *471*, 602–7.

85. Gasiunas, G.; Barrangou, R.; Horvath, P.; Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E2579–86.

86. Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **2012**, *337*, 816–21.

87. Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, R.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science (80-.).* **2013**, *339*, 819–823.

88. Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-guided human genome engineering via Cas9. *Science* **2013**, *339*, 823–6.

89. Cubbon, A.; Ivancic-Bace, I.; Bolt, E. L. CRISPR-Cas immunity, DNA repair and genome stability. *Biosci. Rep.* **2018**, *38*, 1–10.

90. Nuñez, J. K.; Kranzusch, P. J.; Noeske, J.; Wright, A. V.; Davies, C. W.; Doudna, J. A. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. *Nat. Struct. Mol. Biol.* **2014**, *21*, 528–534.

91. Ivančić-Baće, I.; Cass, S. D.; Wearne, S. J.; Bolt, E. L. Different genome stability proteins underpin primed and naïve adaptation in E. coli CRISPR-Cas immunity. *Nucleic Acids Res.* **2015**, *43*, 10821–10830.

92. Radovcic, M.; Killelea, T.; Savitskaya, E.; Wettstein, L.; Bolt, E. L.; Ivancic-Bace, I. CRISPR-Cas adaptation in Escherichia coli requires RecBCD helicase but not nuclease activity, is independent of homologous recombination, and is antagonized by 5' ssDNA exonucleases. *Nucleic Acids Res.* **2018**, *46*, 10173–10183.

93. Xiao, Y.; Ng, S.; Nam, K. H.; Ke, A. How type II CRISPR–Cas establish immunity through Cas1–Cas2-mediated spacer integration. *Nature* **2017**, *550*, 137–141.

94. Levy, A.; Goren, M. G.; Yosef, I.; Auster, O.; Manor, M.; Amitai, G.; Edgar, R.; Qimron, U.; Sorek, R. CRISPR adaptation biases explain preference for acquisition of foreign DNA. *Nature* **2015**, *520*, 505–510.

95. Carte, J.; Wang, R.; Li, H.; Terns, R. M.; Terns, M. P. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev.* **2008**,

22, 3489–3496.

96. Charpentier, E.; Richter, H.; van der Oost, J.; White, M. F. Biogenesis pathways of RNA guides in archaeal and bacterial CRISPR-Cas adaptive immunity. *FEMS Microbiol. Rev.* **2015**, *39*, 428–441.

97. Bolotin, A.; Quinquis, B.; Sorokin, A.; Ehrlich, S. D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **2005**, *151*, 2551–2561.

98. Deveau, H.; Barrangou, R.; Garneau, J. E.; Labont{\'e}, J.; Fremaux, C.; Boyaval, P.; Romero, D. A.; Horvath, P.; Moineau, S. Phage Response to CRISPR-Encoded Resistance in Streptococcus thermophilus. *J. Bacteriol.* **2008**, *190*, 1390–1400.

99. Makarova, K. S. et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **2020**, *18*, 67–83.

100. Silas, S. et al. On the Origin of Reverse Transcriptase-Using CRISPR-Cas Systems and Their Hyperdiverse, Enigmatic Spacer Repertoires. *MBio* **2017**, *8*, e00897–17.

101. Chylinski, K.; Makarova, K. S.; Charpentier, E.; Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **2014**, *42*, 6091–105.

102. Fonfara, I.; Richter, H.; Bratovič, M.; Le Rhun, A.; Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **2016**, *532*, 517–521.

103. Makarova, K. S. et al. An updated evolutionary classification of CRISPR–Cas systems. *Nat. Rev. Microbiol.* **2015**, *13*, 722–736.

104. Shah, S. A.; Alkhnbashi, O. S.; Behler, J.; Han, W.; She, Q.; Hess, W. R.; Garrett, R. A.; Backofen, R. Comprehensive search for accessory proteins encoded with archaeal and bacterial type III CRISPR- cas gene cassettes reveals 39 new cas gene families. *RNA Biol.* **2019**, *16*, 530–542.

105. He, L.; St. John James, M.; Radovcic, M.; Ivancic-Bace, I.; Bolt, E. L. Cas3 Protein—A Review of a Multi-Tasking Machine. *Genes (Basel).* **2020**, *11*, 208.

106. Makarova, K. S. A DNA repair system specific for thermophilic Archaea and bacteria predicted by genomic context analysis. *Nucleic Acids Res.* **2002**, *30*, 482–496.

107. Pinilla-Redondo, R.; Mayo-Muñoz, D.; Russel, J.; Garrett, R. A.; Randau, L.; Sørensen, S. J.; Shah, S. A. Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. *Nucleic Acids Res.* **2020**, *48*, 2000–2012.

108. Özcan, A.; Pausch, P.; Linden, A.; Wulf, A.; Schühle, K.; Heider, J.; Urlaub, H.; Heimerl, T.; Bange, G.; Randau, L. Type IV CRISPR RNA processing and effector complex formation in Aromatoleum aromaticum. *Nat. Microbiol.* **2019**, *4*, 89–96.

109. Zetsche, B.; Gootenberg, J. S.; Abudayyeh, O. O.; Slaymaker, I. M.; Makarova, K. S.; Essletzbichler, P.; Volz, S. E.; Joung, J.; van der Oost, J.; Regev, A.; Koonin, E. V.; Zhang, F. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **2015**, *163*, 759–771.

110. Wright, A. V.; Wang, J. Y.; Burstein, D.; Harrington, L. B.; Paez-Espino, D.;

Kyrpides, N. C.; Iavarone, A. T.; Banfield, J. F.; Doudna, J. A. A Functional Mini-Integrase in a Two-Protein Type V-C CRISPR System. *Mol. Cell* **2019**, *73*, 727–737.e3.

111. Abudayyeh, O. O.; Gootenberg, J. S.; Konermann, S.; Joung, J.; Slaymaker, I. M.; Cox, D. B. T.; Shmakov, S.; Makarova, K. S.; Semenova, E.; Minakhin, L.; Severinov, K.; Regev, A.; Lander, E. S.; Koonin, E. V.; Zhang, F. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **2016**, *353*, aaf5573.

112. East-Seletsky, A.; O'Connell, M. R.; Knight, S. C.; Burstein, D.; Cate, J. H. D.; Tjian, R.; Doudna, J. A. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **2016**, *538*, 270–273.

113. Nishimasu, H.; Ran, F. A.; Hsu, P. D.; Konermann, S.; Shehata, S. I.; Dohmae, N.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell* **2014**, *156*, 935–949.

114. Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **2014**, *343*, 1247997.

115. Palermo, G.; Miao, Y.; Walker, R. C.; Jinek, M.; McCammon, J. A. CRISPR-Cas9 conformational activation as elucidated from enhanced molecular simulations. *Proc. Natl. Acad. Sci.* **2017**, *114*, 7260–7265.

116. Palermo, G.; Chen, J. S.; Ricci, C. G.; Rivalta, I.; Jinek, M.; Batista, V. S.; Doudna, J. A.; McCammon, J. A. Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain. *Q. Rev. Biophys.* **2018**, *51*, e9.

117. Stella, S.; Alcón, P.; Montoya, G. Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature* **2017**, *546*, 559–563.

118. Yamano, T.; Nishimasu, H.; Zetsche, B.; Hirano, H.; Slaymaker, I. M.; Li, Y.; Fedorova, I.; Nakane, T.; Makarova, K. S.; Koonin, E. V.; Ishitani, R.; Zhang, F.; Nureki, O. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. *Cell* **2016**, *165*, 949–962.

119. Dong, D. et al. The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature* **2016**, *532*, 522–526.

120. Shmakov, S.; Smargon, A.; Scott, D.; Cox, D.; Pyzocha, N.; Yan, W.; Abudayyeh, O. O.; Gootenberg, J. S.; Makarova, K. S.; Wolf, Y. I.; Severinov, K.; Zhang, F.; Koonin, E. V. Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **2017**, *15*, 169–182.

121. Saha, A.; Arantes, P. R.; Hsu, R. V.; Narkhede, Y. B.; Jinek, M.; Palermo, G. Molecular Dynamics Reveals a DNA-Induced Dynamic Switch Triggering Activation of CRISPR-Cas12a. *J. Chem. Inf. Model.* **2020**, *60*, 6427–6437.

122. Stella, S.; Mesa, P.; Thomsen, J.; Paul, B.; Alcón, P.; Jensen, S. B.; Saligram, B.; Moses, M. E.; Hatzakis, N. S.; Montoya, G. Conformational Activation Promotes CRISPR-Cas12a Catalysis and Resetting of the Endonuclease Activity. *Cell* **2018**, *175*, 1856–1871.e21.

123. Chen, J. S.; Ma, E.; Harrington, L. B.; Da Costa, M.; Tian, X.; Palefsky, J. M.; Doudna, J. A. CRISPR-Cas12a target binding unleashes indiscriminate single-

stranded DNase activity. *Science (80-.).* **2018**, *360*, 436–439.

124. Swarts, D. C.; van der Oost, J.; Jinek, M. Structural Basis for Guide RNA Processing and Seed-Dependent DNA Targeting by CRISPR-Cas12a. *Mol. Cell* **2017**, *66*, 221–233.e4.

125. Seki, T.; Diffley, J. F. X. Stepwise assembly of initiation proteins at budding yeast replication origins in vitro. *Proc. Natl. Acad. Sci.* **2000**, *97*, 14115–14120.

126. Donovan, S.; Harwood, J.; Drury, L. S.; Diffley, J. F. X. Cdc6p-dependent loading of Mcm proteins onto pre-replicative chromatin in budding yeast. *Proc. Natl. Acad. Sci.* **1997**, *94*, 5611–5616.

127. Moyer, S. E.; Lewis, P. W.; Botchan, M. R. Isolation of the Cdc45/Mcm2-7/GINS (CMG) complex, a candidate for the eukaryotic DNA replication fork helicase. *Proc. Natl. Acad. Sci.* **2006**, *103*, 10236–10241.

128. Aparicio, T.; Guillou, E.; Coloma, J.; Montoya, G.; Méndez, J. The human GINS complex associates with Cdc45 and MCM and is essential for DNA replication. *Nucleic Acids Res.* **2009**, *37*, 2087–2095.

129. Boos, D.; Sanchez-Pulido, L.; Rappas, M.; Pearl, L. H.; Oliver, A. W.; Ponting, C. P.; Diffley, J. F. Regulation of DNA Replication through Sld3-Dpb11 Interaction Is Conserved from Yeast to Humans. *Curr. Biol.* **2011**, *21*, 1152–1157.

130. Tanaka, T.; Umemori, T.; Endo, S.; Muramatsu, S.; Kanemaki, M.; Kamimura, Y.; Obuse, C.; Araki, H. Sld7, an Sld3-associated protein required for efficient chromosomal DNA replication in budding yeast. *EMBO J.* **2011**, *30*, 2019–2030.

131. Aria, V.; Yeeles, J. T. Mechanism of Bidirectional Leading-Strand Synthesis Establishment at Eukaryotic DNA Replication Origins. *Mol. Cell* **2019**, *73*, 199–211.e10.

132. Szambowska, A.; Tessmer, I.; Prus, P.; Schlott, B.; Pospiech, H.; Grosse, F. Cdc45-induced loading of human RPA onto single-stranded DNA. *Nucleic Acids Res.* **2017**, *45*, 3217–3230.

133. Lucas, I. Topoisomerase II can unlink replicating DNA by precatenane removal. *EMBO J.* **2001**, *20*, 6509–6519.

134. Leman, A. R.; Noguchi, E. Local and global functions of Timeless and Tipin in replication fork protection. *Cell Cycle* **2012**, *11*, 3945–3955.

135. Noguchi, E.; Noguchi, C.; Du, L.-L.; Russell, P. Swi1 Prevents Replication Fork Collapse and Controls Checkpoint Kinase Cds1. *Mol. Cell. Biol.* **2003**, *23*, 7861–7874.

136. Chilkova, O.; Stenlund, P.; Isoz, I.; Stith, C. M.; Grabowski, P.; Lundstrom, E.-B.; Burgers, P. M.; Johansson, E. The eukaryotic leading and lagging strand DNA polymerases are loaded onto primer-ends via separate mechanisms but have comparable processivity in the presence of PCNA. *Nucleic Acids Res.* **2007**, *35*, 6588–6597.

137. Muzi-Falconi, M.; Giannattasio, M.; Foiani, M.; Plevani, P. The DNA Polymerase alpha-Primase Complex: Multiple Functions and Interactions. *Sci. World J.* **2003**, *3*, 21–33.

138. Pursell, Z. F.; Isoz, I.; Lundstrom, E.-B.; Johansson, E.; Kunkel, T. A. Yeast DNA Polymerase Participates in Leading-Strand DNA Replication. *Science (80-.).* **2007**, *317*, 127–130.

139. Lopes, M.; Foiani, M.; Sogo, J. M. Multiple Mechanisms Control Chromosome Integrity after Replication Fork Uncoupling and Restart at Irreparable UV Lesions. *Mol. Cell* **2006**, *21*, 15–27.

140. Lehmann, A. R. Postreplication repair of DNA in ultraviolet-irradiated mammalian cells. *J. Mol. Biol.* **1972**, *66*, 319–37.

141. Taylor, M. R.; Yeeles, J. T. The Initial Response of a Eukaryotic Replisome to DNA Damage. *Mol. Cell* **2018**, *70*, 1067–1080.e12.

142. Bainbridge, L. J.; Teague, R.; Doherty, A. J. Repriming DNA synthesis: an intrinsic restart pathway that maintains efficient genome replication. *Nucleic Acids Res.* **2021**, *49*, 4831–4847.

143. Boldinova, E. O.; Yudkina, A. V.; Shilkin, E. S.; Gagarinskaya, D. I.; Baranovskiy, A. G.; Tahirov, T. H.; Zharkov, D. O.; Makarova, A. V. Translesion activity of PrimPol on DNA with cisplatin and DNA–protein cross-links. *Sci. Rep.* **2021**, *11*, 17588.

144. Ogawa, T.; Okazaki, T. Discontinuous DNA Replication. *Annu. Rev. Biochem.* **1980**, *49*, 421–457.

145. Okazaki, R.; Okazaki, T.; Sakabe, K.; Sugimoto, K.; Kainuma, R.; Sugino, A.; Iwatsuki, N. In Vivo Mechanism of DNA Chain Growth. *Cold Spring Harb. Symp. Quant. Biol.* **1968**, *33*, 129–143.

146. Zhou, Z.-X.; Lujan, S. A.; Burkholder, A. B.; Garbacz, M. A.; Kunkel, T. A. Roles for DNA polymerase δ in initiating and terminating leading strand DNA replication. *Nat. Commun.* **2019**, *10*, 3992.

147. Garg, P.; Stith, C. M.; Sabouri, N.; Johansson, E.; Burgers, P. M. Idling by DNA polymerase δ maintains a ligatable nick during lagging-strand DNA replication. *Genes Dev.* **2004**, *18*, 2764–2773.

148. Fu, Y. V.; Yardimci, H.; Long, D. T.; Guainazzi, A.; Bermudez, V. P.; Hurwitz, J.; van Oijen, A.; Schärer, O. D.; Walter, J. C. Selective Bypass of a Lagging Strand Roadblock by the Eukaryotic Replicative DNA Helicase. *Cell* **2011**, *146*, 931–941.

149. Maric, M.; Maculins, T.; De Piccoli, G.; Labib, K. Cdc48 and a ubiquitin ligase drive disassembly of the CMG helicase at the end of DNA replication. *Science* **2014**, *346*, 1253596.

150. Goulian, M.; Richards, S. H.; Heard, C. J.; Bigsby, B. M. Discontinuous DNA synthesis by purified mammalian proteins. *J. Biol. Chem.* **1990**, *265*, 18461–18471.

151. Postow, L.; Crisona, N. J.; Peter, B. J.; Hardy, C. D.; Cozzarelli, N. R. Topological challenges to DNA replication: Conformations at the fork. *Proc. Natl. Acad. Sci.* **2001**, *98*, 8219–8226.

152. Vafabakhsh, R.; Ha, T. Extreme Bendability of DNA Less than 100 Base Pairs Long Revealed by Single-Molecule Cyclization. *Science (80-.).* **2012**, *337*, 1097–1101.

153. Dewar, J. M.; Budzowska, M.; Walter, J. C. The mechanism of DNA replication termination in vertebrates. *Nature* **2015**, *525*, 345–350.

154. Baxter, J.; Diffley, J. F. Topoisomerase II Inactivation Prevents the Completion of DNA Replication in Budding Yeast. *Mol. Cell* **2008**, *30*, 790–802.

155. Kobayashi, T. Ribosomal RNA gene repeats, their stability and cellular senes-

cence. *Proc. Japan Acad. Ser. B* **2014**, *90*, 119–129.

156. Neylon, C.; Kralicek, A. V.; Hill, T. M.; Dixon, N. E. Replication Termination in Escherichia coli : Structure and Antihelicase Activity of the Tus- Ter Complex. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 501–526.

157. Elshenawy, M. M.; Jergic, S.; Xu, Z.-Q.; Sobhy, M. A.; Takahashi, M.; Oakley, A. J.; Dixon, N. E.; Hamdan, S. M. Replisome speed determines the efficiency of the Tus-Ter replication termination barrier. *Nature* **2015**, *525*, 394–398.

158. Dewar, J. M.; Walter, J. C. Mechanisms of DNA replication termination. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 507–516.

159. Shyian, M.; Shore, D. Approaching Protein Barriers: Emerging Mechanisms of Replication Pausing in Eukaryotes. *Front. Cell Dev. Biol.* **2021**, *9*, 1–17.

160. Pfeiffer, V.; Lingner, J. Replication of Telomeres and the Regulation of Telomerase. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a010405–a010405.

161. Cortez, D. Preventing replication fork collapse to maintain genome integrity. *DNA Repair (Amst).* **2015**, *32*, 149–157.

162. Yousefzadeh, M.; Henpita, C.; Vyas, R.; Soto-Palma, C.; Robbins, P.; Niedernhofer, L. DNA damage-how and why we age? *Elife* **2021**, *10*, 1–17.

163. De Bont, R. Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis* **2004**, *19*, 169–185.

164. Cortez, D. Replication-Coupled DNA Repair. *Mol. Cell* **2019**, *74*, 866–876.

165. Alexander, J. L.; Orr-Weaver, T. L. Replication fork instability and the consequences of fork collisions from rereplication. *Genes Dev.* **2016**, *30*, 2241–2252.

166. Escorcia, W.; Forsburg, S. L. Destabilization of the replication fork protection complex disrupts meiotic chromosome segregation. *Mol. Biol. Cell* **2017**, *28*, 2978–2997.

167. Gaillard, H.; García-Muse, T.; Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **2015**, *15*, 276–289.

168. Nguyen, T. V.; Pawlikowska, P.; Firlej, V.; Rosselli, F.; Aoufouchi, S. V(D)J recombination process and the Pre-B to immature B-cells transition are altered in Fanca-/- mice. *Sci. Rep.* **2016**, *6*, 36906.

169. Brandsma, I.; Gent, D. C. Pathway choice in DNA double strand break repair: observations of a balancing act. *Genome Integr.* **2012**, *3*, 9.

170. Marini, F.; Rawal, C. C.; Liberi, G.; Pellicioli, A. Regulation of DNA Double Strand Breaks Processing: Focus on Barriers. *Front. Mol. Biosci.* **2019**, *6*, 1–8.

171. Fugger, K.; West, S. C. Keeping homologous recombination in check. *Cell Res.* **2016**, *26*, 397–8.

172. Orthwein, A.; Noordermeer, S. M.; Wilson, M. D.; Landry, S.; Enchev, R. I.; Sherker, A.; Munro, M.; Pinder, J.; Salsman, J.; Dellaire, G.; Xia, B.; Peter, M.; Durocher, D. A mechanism for the suppression of homologous recombination in G1 cells. *Nature* **2015**, *528*, 422–6.

173. Langerak, P.; Russell, P. Regulatory networks integrating cell cycle control with DNA damage checkpoints and double-strand break repair. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **2011**, *366*, 3562–71.

174. Her, J.; Bunting, S. F. How cells ensure correct repair of DNA double-strand breaks. *J. Biol. Chem.* **2018**, *293*, 10502–10511.

175. Lu, H.; Shamanna, R. A.; de Freitas, J. K.; Okur, M.; Khadka, P.; Kulikow-icz, T.; Holland, P. P.; Tian, J.; Croteau, D. L.; Davis, A. J.; Bohr, V. A. Cell cycle-dependent phosphorylation regulates RECQL4 pathway choice and ubiq-uitination in DNA double-strand break repair. *Nat. Commun.* **2017**, *8*, 2039.

176. Scully, R.; Panday, A.; Elango, R.; Willis, N. A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 698–714.

177. Nichols, C. A.; Gibson, W. J.; Brown, M. S.; Kosmicki, J. A.; Busanovich, J. P.; Wei, H.; Urbanski, L. M.; Curimjee, N.; Berger, A. C.; Gao, G. F.; Cherni-ack, A. D.; Dhe-Paganon, S.; Paolella, B. R.; Beroukhim, R. Loss of heterozy-gosity of essential genes represents a widespread class of potential cancer vulnerabilities. *Nat. Commun.* **2020**, *11*, 2517.

178. Ryland, G. L.; Doyle, M. A.; Goode, D.; Boyle, S. E.; Choong, D. Y. H.; Row-ley, S. M.; Li, J.; Australian Ovarian Cancer Study Group,; Bowtell, D. D. L.; Tothill, R. W.; Campbell, I. G.; Gorringe, K. L. Loss of heterozygosity: what is it good for? *BMC Med. Genomics* **2015**, *8*, 45.

179. Merajver, S. D.; Frank, T. S.; Xu, J.; Pham, T. M.; Calzone, K. A.; Bennett-Baker, P.; Chamberlain, J.; Boyd, J.; Garber, J. E.; Collins, F. S. Germline BRCA1 mutations and loss of the wild-type allele in tumors from families with early onset breast and ovarian cancer. *Clin. Cancer Res.* **1995**, *1*, 539–44.

180. 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.* **2017**, *18*, 495–506.

181. Orthwein, A.; Fradet-Turcotte, A.; Noordermeer, S. M.; Canny, M. D.; Brun, C. M.; Strecker, J.; Escribano-Diaz, C.; Durocher, D. Mitosis Inhibits DNA Double-Strand Break Repair to Guard Against Telomere Fusions. *Science (80-.).* **2014**, *344*, 189–193.

182. Uematsu, N.; Weterings, E.; Yano, K.-i.; Morotomi-Yano, K.; Jakob, B.; Taucher-Scholz, G.; Mari, P.-O.; van Gent, D. C.; Chen, B. P.; Chen, D. J. Autophosphory-lation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J. Cell Biol.* **2007**, *177*, 219–229.

183. Sinha, S.; Li, F.; Villarreal, D.; Shim, J. H.; Yoon, S.; Myung, K.; Shim, E. Y.; Lee, S. E. Microhomology-mediated end joining induces hypermutagenesis at breakpoint junctions. *PLOS Genet.* **2017**, *13*, e1006714.

184. Sharma, S.; Javadekar, S. M.; Pandey, M.; Srivastava, M.; Kumari, R.; Raghavan, S. C. Homology and enzymatic requirements of microhomology-dependent alternative end joining. *Cell Death Dis.* **2015**, *6*, e1697–e1697.

185. Truong, L. N.; Li, Y.; Shi, L. Z.; Hwang, P. Y.-H.; He, J.; Wang, H.; Razavian, N.; Berns, M. W.; Wu, X. Microhomology-mediated End Joining and Homologous Recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc. Natl. Acad. Sci.* **2013**, *110*, 7720–7725.

186. Deng, S. K.; Gibb, B.; de Almeida, M. J.; Greene, E. C.; Symington, L. S. RPA antagonizes microhomology-mediated repair of DNA double-strand breaks. *Nat. Struct. Mol. Biol.* **2014**, *21*, 405–412.

187. Seol, J.-H.; Shim, E. Y.; Lee, S. E. Microhomology-mediated end joining: Good, bad and ugly. *Mutat. Res. Mol. Mech. Mutagen.* **2018**, *809*, 81–87.

188. Mateos-Gomez, P. A.; Gong, F.; Nair, N.; Miller, K. M.; Lazzerini-Denchi, E.; Sfeir, A. Mammalian polymerase θ promotes alternative NHEJ and suppresses recombination. *Nature* **2015**, *518*, 254–257.

189. Yousefzadeh, M. J.; Wyatt, D. W.; Takata, K.-i.; Mu, Y.; Hensley, S. C.; Tomida, J.; Bylund, G. O.; Doublié, S.; Johansson, E.; Ramsden, D. A.; McBride, K. M.; Wood, R. D. Mechanism of Suppression of Chromosomal Instability by DNA Polymerase POLQ. *PLoS Genet.* **2014**, *10*, e1004654.

190. Elango, R.; Sheng, Z.; Jackson, J.; DeCata, J.; Ibrahim, Y.; Pham, N. T.; Liang, D. H.; Sakofsky, C. J.; Vindigni, A.; Lobachev, K. S.; Ira, G.; Malkova, A. Break-induced replication promotes formation of lethal joint molecules dissolved by Srs2. *Nat. Commun.* **2017**, *8*, 1790.

191. Takata, M.; Sasaki, M. S.; Sonoda, E.; Morrison, C.; Hashimoto, M.; Utsumi, H.; Yamaguchi-Iwai, Y.; Shinohara, A.; Takeda, S. Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO J.* **1998**, *17*, 5497–5508.

192. Li, X.; Heyer, W.-D. Homologous recombination in DNA repair and DNA damage tolerance. *Cell Res.* **2008**, *18*, 99–113.

193. Short, J. M.; Liu, Y.; Chen, S.; Soni, N.; Madhusudhan, M. S.; Shivji, M. K. K.; Venkitaraman, A. R. High-resolution structure of the presynaptic RAD51 filament on single-stranded DNA by electron cryo-microscopy. *Nucleic Acids Res.* **2016**, *44*, 9017–9030.

194. Chen, Z.; Yang, H.; Pavletich, N. P. Mechanism of homologous recombination from the RecA–ssDNA/dsDNA structures. *Nature* **2008**, *453*, 489–494.

195. Lee, J. Y.; Terakawa, T.; Qi, Z.; Steinfeld, J. B.; Redding, S.; Kwon, Y.; Gaines, W. A.; Zhao, W.; Sung, P.; Greene, E. C. DNA RECOMBINATION. Base triplet stepping by the Rad51/RecA family of recombinases. *Science* **2015**, *349*, 977–81.

196. McVey, M.; Khodaverdian, V. Y.; Meyer, D.; Cerqueira, P. G.; Heyer, W.-D. Eukaryotic DNA Polymerases in Homologous Recombination. *Annu. Rev. Genet.* **2016**, *50*, 393–421.

197. Wyatt, H. D. M.; West, S. C. Holliday junction resolvases. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a023192.

198. Daley, J. M.; Gaines, W. A.; Kwon, Y.; Sung, P. Regulation of DNA pairing in homologous recombination. *Cold Spring Harb. Perspect. Biol.* **2014**, *6*, a017954.

199. Hastings, P. Mechanisms of Ectopic Gene Conversion. *Genes (Basel).* **2010**, *1*, 427–439.

200. Zapotoczny, G.; Sekelsky, J. Human Cell Assays for Synthesis-Dependent Strand Annealing and Crossing over During Double-Strand Break Repair. *G3 (Bethesda).* **2017**, *7*, 1191–1199.

201. Kramara, J.; Osia, B.; Malkova, A. Break-Induced Replication: The Where, The Why, and The How. *Trends Genet.* **2018**, *34*, 518–531.

202. Li, S.; Wang, H.; Jehi, S.; Li, J.; Liu, S.; Wang, Z.; Truong, L.; Chiba, T.; Wang, Z.;

Wu, X. PIF1 helicase promotes break-induced replication in mammalian cells. *EMBO J.* **2021**, *40*, e104509.

203. Saini, N.; Ramakrishnan, S.; Elango, R.; Ayyar, S.; Zhang, Y.; Deem, A.; Ira, G.; Haber, J. E.; Lobachev, K. S.; Malkova, A. Migrating bubble during break-induced replication drives conservative DNA synthesis. *Nature* **2013**, *502*, 389–392.

204. Wilson, M. A.; Kwon, Y.; Xu, Y.; Chung, W.-H.; Chi, P.; Niu, H.; Mayle, R.; Chen, X.; Malkova, A.; Sung, P.; Ira, G. Pif1 helicase and Polδ promote recombination-coupled DNA synthesis via bubble migration. *Nature* **2013**, *502*, 393–396.

205. Lydeard, J. R.; Jain, S.; Yamaguchi, M.; Haber, J. E. Break-induced replication and telomerase-independent telomere maintenance require Pol32. *Nature* **2007**, *448*, 820–823.

206. Blasiak, J. Single-Strand Annealing in Cancer. *Int. J. Mol. Sci.* **2021**, *22*, 2167.

207. Al-Minawi, A. Z.; Saleh-Gohari, N.; Helleday, T. The ERCC1/XPF endonuclease is required for efficient single-strand annealing and gene conversion in mammalian cells. *Nucleic Acids Res.* **2007**, *36*, 1–9.

208. Bhargava, R.; Onyango, D. O.; Stark, J. M. Regulation of Single-Strand Annealing and its Role in Genome Maintenance. *Trends Genet.* **2016**, *32*, 566–575.

209. Davis, L.; Maizels, N. Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proc. Natl. Acad. Sci.* **2014**, *111*, E924–E932.

210. Gallagher, D. N.; Pham, N.; Tsai, A. M.; Janto, A. N.; Choi, J.; Ira, G.; Haber, J. E. A Rad51-independent pathway promotes single-strand template repair in gene editing. *PLOS Genet.* **2020**, *16*, e1008689.

211. Gallagher, D. N.; Haber, J. E. Single-strand template repair: key insights to increase the efficiency of gene editing. *Curr. Genet.* **2021**, *67*, 747–753.

212. Richardson, C. D.; Kazane, K. R.; Feng, S. J.; Zelin, E.; Bray, N. L.; Schäfer, A. J.; Floor, S. N.; Corn, J. E. CRISPR–Cas9 genome editing in human cells occurs via the Fanconi anemia pathway. *Nat. Genet.* **2018**, *50*, 1132–1139.

213. Kee, Y.; D'Andrea, A. D. Expanded roles of the Fanconi anemia pathway in preserving genomic stability. *Genes Dev.* **2010**, *24*, 1680–1694.

214. Kreutmair, S. et al. Loss of the Fanconi anemia-associated protein NIPA causes bone marrow failure. *J. Clin. Invest.* **2020**, *130*, 2827–2844.

215. De Silva, I. U.; McHugh, P. J.; Clingen, P. H.; Hartley, J. A. Defining the Roles of Nucleotide Excision Repair and Recombination in the Repair of DNA Interstrand Cross-Links in Mammalian Cells. *Mol. Cell. Biol.* **2000**, *20*, 7980–7990.

216. Renaudin, X.; Rosselli, F. The FANC/BRCA Pathway Releases Replication Blockades by Eliminating DNA Interstrand Cross-Links. *Genes (Basel).* **2020**, *11*, 585.

217. McHugh, P. J. XPF–ERCC1: Linchpin of DNA crosslink repair. *PLOS Genet.* **2020**, *16*, e1008616.

218. Enoiu, M.; Jiricny, J.; Schärer, O. D. Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-

coupled repair and translesion synthesis. *Nucleic Acids Res.* **2012**, *40*, 8953–8964.

219. Sarkar, S.; Davies, A. A.; Ulrich, H. D.; McHugh, P. J. DNA interstrand crosslink repair during G1 involves nucleotide excision repair and DNA polymerase ζ. *EMBO J.* **2006**, *25*, 1285–1294.

220. Williams, H. L.; Gottesman, M. E.; Gautier, J. Replication-Independent Repair of DNA Interstrand Crosslinks. *Mol. Cell* **2012**, *47*, 140–147.

221. Huang, J.; Liu, S.; Bellani, M. A.; Thazhathveetil, A. K.; Ling, C.; de Winter, J. P.; Wang, Y.; Wang, W.; Seidman, M. M. The DNA Translocase FANCM/MHF Promotes Replication Traverse of DNA Interstrand Crosslinks. *Mol. Cell* **2013**, *52*, 434–446.

222. Ling, C.; Huang, J.; Yan, Z.; Li, Y.; Ohzeki, M.; Ishiai, M.; Xu, D.; Takata, M.; Seidman, M.; Wang, W. Bloom syndrome complex promotes FANCM recruitment to stalled replication forks and facilitates both repair and traverse of DNA interstrand crosslinks. *Cell Discov.* **2016**, *2*, 16047.

223. McHugh, P. J.; Spanswick, V. J.; Hartley, J. A. Repair of DNA interstrand crosslinks: molecular mechanisms and clinical relevance. *Lancet Oncol.* **2001**, *2*, 483–490.

224. Brosh, R. M.; Matson, S. W. History of DNA Helicases. *Genes (Basel).* **2020**, *11*, 255.

225. Marini, F.; Wood, R. D. A Human DNA Helicase Homologous to the DNA Cross-link Sensitivity Protein Mus308. *J. Biol. Chem.* **2002**, *277*, 8716–8723.

226. Tafel, A. A.; Wu, L.; McHugh, P. J. Human HEL308 localizes to damaged replication forks and unwinds lagging strand structures. *J. Biol. Chem.* **2011**, *286*, 15832–15840.

227. Takata, K. I.; Reh, S.; Tomida, J.; Person, M. D.; Wood, R. D. Human DNA helicase HELQ participates in DNA interstrand crosslink tolerance with ATR and RAD51 paralogs. *Nat. Commun.* **2013**, *4*, 1–11.

228. Adelman, C. A.; Lolo, R. L.; Birkbak, N. J.; Murina, O.; Matsuzaki, K.; Horejsi, Z.; Parmar, K.; Borel, V.; Skehel, J. M.; Stamp, G.; D'Andrea, A.; Sartori, A. A.; Swanton, C.; Boulton, S. J. HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and tumorigenesis. *Nature* **2013**, *502*, 381–384.

229. Jenkins, T.; Northall, S. J.; Ptchelkine, D.; Lever, R.; Cubbon, A.; Betts, H.; Taresco, V.; Cooper, C. D. O.; McHugh, P. J.; Soultanas, P.; Bolt, E. L. The HelQ human DNA repair helicase utilizes a PWI-like domain for DNA loading through interaction with RPA, triggering DNA unwinding by the HelQ helicase core. *NAR Cancer* **2021**, *3*, 1–15.

230. Singleton, M. R.; Dillingham, M. S.; Wigley, D. B. Structure and Mechanism of Helicases and Nucleic Acid Translocases. *Annu. Rev. Biochem.* **2007**, *76*, 23–50.

231. O'Donnell, M. E.; Li, H. The ring-shaped hexameric helicases that function at DNA replication forks. *Nat. Struct. Mol. Biol.* **2018**, *25*, 122–130.

232. Banroques, J.; Doère, M.; Dreyfus, M.; Linder, P.; Tanner, N. K. Motif III in Superfamily 2 "Helicases" Helps Convert the Binding Energy of ATP into a High-Affinity RNA Binding Site in the Yeast DEAD-Box Protein Ded1. *J. Mol. Biol.* **2010**, *396*, 949–966.

233. Ding, H.; Guo, M.; Vidhyasagar, V.; Talwar, T.; Wu, Y. The Q Motif Is Involved

in DNA Binding but Not ATP Binding in ChlR1 Helicase. *PLoS One* **2015**, *10*, e0140755.

234. Wu, Y.; Sommers, J. A.; Loiland, J. A.; Kitao, H.; Kuper, J.; Kisker, C.; Brosh, R. M. The Q motif of Fanconi anemia group J protein (FANCJ) DNA helicase regulates its dimerization, DNA binding, and DNA repair function. *J. Biol. Chem.* **2012**, *287*, 21699–716.

235. Hall, M. C.; Matson, S. W. Helicase motifs: the engine that powers DNA unwinding. *Mol. Microbiol.* **1999**, *34*, 867–877.

236. Northall, S. J.; Buckley, R.; Jones, N.; Penedo, J. C.; Soultanas, P.; Bolt, E. L. DNA binding and unwinding by Hel308 helicase requires dual functions of a winged helix domain. *DNA Repair (Amst).* **2017**, *57*, 125–132.

237. Woodman, I. L.; Briggs, G. S.; Bolt, E. L. Archaeal Hel308 Domain V Couples DNA Binding to ATP Hydrolysis and Positions DNA for Unwinding Over the Helicase Ratchet. *J. Mol. Biol.* **2007**, *374*, 1139–1144.

238. Lohman, T. M.; Bjornson, K. P. Mechanisms of Helicase-Catalyzed DNA Unwinding. *Annu. Rev. Biochem.* **1996**, *65*, 169–214.

239. Manosas, M.; Xi, X. G.; Bensimon, D.; Croquette, V. Active and passive mechanisms of helicases. *Nucleic Acids Res.* **2010**, *38*, 5518–5526.

240. Ribeck, N.; Saleh, O. A. DNA Unwinding by Ring-Shaped T4 Helicase gp41 Is Hindered by Tension on the Occluded Strand. *PLoS One* **2013**, *8*, e79237.

241. Velankar, S. S.; Soultanas, P.; Dillingham, M. S.; Subramanya, H. S.; Wigley, D. B. Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism. *Cell* **1999**, *97*, 75–84.

242. Mackintosh, S. G.; Raney, K. D. DNA unwinding and protein displacement by superfamily 1 and superfamily 2 helicases. *Nucleic Acids Res.* **2006**, *34*, 4154–9.

243. Byrd, A. K.; Raney, K. D. Increasing the Length of the Single-Stranded Overhang Enhances Unwinding of Duplex DNA by Bacteriophage T4 Dda Helicase. *Biochemistry* **2005**, *44*, 12990–12997.

244. Byrd, A. K.; Raney, K. D. Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA. *Nat. Struct. Mol. Biol.* **2004**, *11*, 531–538.

245. Chaban, Y.; Stead, J. A.; Ryzhenkova, K.; Whelan, F.; Lamber, E. P.; Antson, A.; Sanders, C. M.; Orlova, E. V. Structural basis for DNA strand separation by a hexameric replicative helicase. *Nucleic Acids Res.* **2015**, *43*, 8551–8563.

246. Ward, J. D.; Muzzini, D. M.; Petalcorin, M. I.; Martinez-Perez, E.; Martin, J. S.; Plevani, P.; Cassata, G.; Marini, F.; Boulton, S. J. Overlapping Mechanisms Promote Postsynaptic RAD-51 Filament Disassembly during Meiotic Double-Strand Break Repair. *Mol. Cell* **2010**, *37*, 259–272.

247. Amendola, P. G.; Zaghet, N.; Ramalho, J. J.; Vilstrup Johansen, J.; Boxem, M.; Salcini, A. E. JMJD-5/KDM8 regulates H3K36me2 and is required for late steps of homologous recombination and genome integrity. *PLOS Genet.* **2017**, *13*, e1006632.

248. Hustedt, N.; Saito, Y.; Zimmermann, M.; Álvarez-Quilón, A.; Setiaputra, D.; Adam, S.; McEwan, A.; Yuan, J. Y.; Olivieri, M.; Zhao, Y.; Kanemaki, M. T.; Jurisicova, A.; Durocher, D. Control of homologous recombination by the

HROB–MCM8–MCM9 pathway. *Genes Dev.* **2019**, *33*, 1397–1415.

249. Muzzini, D. M.; Plevani, P.; Boulton, S. J.; Cassata, G.; Marini, F. Caenorhabditis elegans POLQ-1 and HEL-308 function in two distinct DNA interstrand cross-link repair pathways. *DNA Repair (Amst).* **2008**, *7*, 941–950.

250. Luebben, S. W.; Kawabata, T.; Akre, M. K.; Lee, W. L.; Johnson, C. S.; O'Sullivan, M. G.; Shima, N. Helq acts in parallel to Fancc to suppress replication-associated genome instability. *Nucleic Acids Res.* **2013**, *41*, 10283–97.

251. Shamanna, R. A.; Lu, H.; de Freitas, J. K.; Tian, J.; Croteau, D. L.; Bohr, V. A. WRN regulates pathway choice between classical and alternative non-homologous end joining. *Nat. Commun.* **2016**, *7*, 13785.

252. Lebel, M.; Lavoie, J.; Gaudreault, I.; Bronsard, M.; Drouin, R. Genetic Coopera-tion between the Werner Syndrome Protein and Poly(ADP-Ribose) Polymerase-1 in Preventing Chromatid Breaks, Complex Chromosomal Rearrangements, and Cancer in Mice. *Am. J. Pathol.* **2003**, *162*, 1559–1569.

253. Opresko, P. L.; Otterlei, M.; Graakjær, J.; Bruheim, P.; Dawut, L.; Kølvraa, S.; May, A.; Seidman, M. M.; Bohr, V. A. The Werner Syndrome Helicase and Exonuclease Cooperate to Resolve Telomeric D Loops in a Manner Regulated by TRF1 and TRF2. *Mol. Cell* **2004**, *14*, 763–774.

254. Bachrati, C. Z. Mobile D-loops are a preferred substrate for the Bloom's syn-drome helicase. *Nucleic Acids Res.* **2006**, *34*, 2269–2279.

255. Hanahan, D.; Weinberg, R. A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144*, 646–674.

256. Kinzler, K. W.; Vogelstein, B. Gatekeepers and caretakers. *Nature* **1997**, *386*, 761–763.

257. van Heemst, D.; den Reijer, P.; Westendorp, R. Ageing or cancer: A review. *Eur. J. Cancer* **2007**, *43*, 2144–2152.

258. Guo, C.; Gao, Y.-y.; Ju, Q.-q.; Zhang, C.-x.; Gong, M.; Li, Z.-l. HELQ and EGR3 expression correlate with IGHV mutation status and prognosis in chronic lymphocytic leukemia. *J. Transl. Med.* **2021**, *19*, 42.

259. Long, G.; Ouyang, W.; Zhang, Y.; Sun, G.; Gan, J.; Hu, Z.; Li, H. Identification of a DNA Repair Gene Signature and Establishment of a Prognostic Nomogram Predicting Biochemical-Recurrence-Free Survival of Prostate Cancer. *Front. Mol. Biosci.* **2021**, *8*, 1–13.

260. Long, J.; Zhu, J. Y.; Liu, Y. B.; Fu, K.; Tian, Y.; Li, P. Y.; Yang, W. Q.; Yang, S. Y.; Yin, J. Y.; Yin, G.; Zhang, Y. Helicase POLQ-like (HELQ) as a novel indicator of platinum-based chemoresistance for epithelial ovarian cancer. *Gynecol. Oncol.* **2018**, *149*, 341–349.

261. Goodman, L. S. Nitrogen mustard therapy. *J. Am. Med. Assoc.* **1946**, *132*, 126.

262. Den Hartigh, J.; Van Oort, W.; Bocken, M.; Pinedo, H. High-performance liquid chromatographic determination of the antitumor agent mitomycin c in human blood plasma. *Anal. Chim. Acta* **1981**, *127*, 47–53.

263. Wang, X.; Zhang, H.; Chen, X. Drug resistance and combating drug resistance in cancer. *Cancer Drug Resist.* **2019**, *2*, 141–160.

264. Bukowski, K.; Kciuk, M.; Kontek, R. Mechanisms of Multidrug Resistance in

Cancer Chemotherapy. *Int. J. Mol. Sci.* **2020**, *21*, 3233.

265. Rocha, C. R. R.; Silva, M. M.; Quinet, A.; Cabral-Neto, J. B.; Menck, C. F. M. DNA repair pathways and cisplatin resistance: an intimate relationship. *Clinics (Sao Paulo).* **2018**, *73*, e478s.

266. Yin, Q.-K.; Wang, C.-X.; Wang, Y.-Q.; Guo, Q.-L.; Zhang, Z.-L.; Ou, T.-M.; Huang, S.-L.; Li, D.; Wang, H.-G.; Tan, J.-H.; Chen, S.-B.; Huang, Z.-S. Discovery of Isaindigotone Derivatives as Novel Bloom's Syndrome Protein (BLM) Helicase Inhibitors That Disrupt the BLM/DNA Interactions and Regulate the Homologous Recombination Repair. *J. Med. Chem.* **2019**, *62*, 3147–3162.

267. Premsrirut, P.; Martin, G.; Dow, L.; Kim, S.; Zuber, J.; Hannon, G.; Lowe, S. RNAi and CRISPR/Cas9 based in vivo models for drug discovery. *Eur. J. Cancer* **2016**, *61*, S208.

268. Innocenti, F.; Cox, N. J.; Dolan, M. E. The Use of Genomic Information to Optimize Cancer Chemotherapy. *Semin. Oncol.* **2011**, *38*, 186–195.

269. Berger, M. F.; Mardis, E. R. The emerging clinical relevance of genomics in cancer medicine. *Nat. Rev. Clin. Oncol.* **2018**, *15*, 353–365.

270. Schrempf, A.; Slyskova, J.; Loizou, J. I. Targeting the DNA Repair Enzyme Polymerase θ in Cancer Therapy. *Trends in Cancer* **2021**, *7*, 98–111.

271. Zatreanu, D. et al. Polθ inhibitors elicit BRCA-gene synthetic lethality and target PARP inhibitor resistance. *Nat. Commun.* **2021**, *12*, 3636.

272. Volochnyuk, D. M.; Ryabukhin, S. V.; Moroz, Y. S.; Savych, O.; Chuprina, A.; Horvath, D.; Zabolotna, Y.; Varnek, A.; Judd, D. B. Evolution of commercially available compounds for HTS. *Drug Discov. Today* **2019**, *24*, 390–402.

273. Blay, V.; Tolani, B.; Ho, S. P.; Arkin, M. R. High-Throughput Screening: today's biochemical and cell-based approaches. *Drug Discov. Today* **2020**, *25*, 1807–1821.

274. Dorr, P. et al. Maraviroc (UK-427,857), a Potent, Orally Bioavailable, and Selective Small-Molecule Inhibitor of Chemokine Receptor CCR5 with Broad-Spectrum Anti-Human Immunodeficiency Virus Type 1 Activity. *Antimicrob. Agents Chemother.* **2005**, *49*, 4721–4732.

275. Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty years on: the impact of fragments on drug discovery. *Nat. Rev. Drug Discov.* **2016**, *15*, 605–619.

276. Barker, A.; Kettle, J. G.; Nowak, T.; Pease, J. E. Expanding medicinal chemistry space. *Drug Discov. Today* **2013**, *18*, 298–304.

277. Lane, S. J.; Eggleston, D. S.; Brinded, K. A.; Hollerton, J. C.; Taylor, N. L.; Readshaw, S. A. Defining and maintaining a high quality screening collection: the GSK experience. *Drug Discov. Today* **2006**, *11*, 267–272.

278. Macarron, R.; Banks, M. N.; Bojanic, D.; Burns, D. J.; Cirovic, D. A.; Garyantes, T.; Green, D. V. S.; Hertzberg, R. P.; Janzen, W. P.; Paslay, J. W.; Schopfer, U.; Sittampalam, G. S. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **2011**, *10*, 188–195.

279. Bohacek, R. S.; McMartin, C.; Guida, W. C. The art and practice of structure-based drug design: A molecular modeling perspective. *Med. Res. Rev.* **1996**, *16*, 3–50.

280. Jencks, W. P. On the attribution and additivity of binding energies. *Proc. Natl. Acad. Sci.* **1981**, *78*, 4046–4050.

281. Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science (80-.).* **1996**, *274*, 1531–1534.

282. Shi, Y.; von Itzstein, M. How Size Matters: Diversity for Fragment Library Design. *Molecules* **2019**, *24*, 2838.

283. Hall, R. J.; Mortenson, P. N.; Murray, C. W. Efficient exploration of chemical space by fragment-based screening. *Prog. Biophys. Mol. Biol.* **2014**, *116*, 82–91.

284. Kirsch, P.; Hartman, A. M.; Hirsch, A. K. H.; Empting, M. Concepts and Core Principles of Fragment-Based Drug Design. *Molecules* **2019**, *24*, 4309.

285. Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'Rule of Three' for fragment-based lead discovery? *Drug Discov. Today* **2003**, *8*, 876–877.

286. Turnbull, A.; Boyd, S.; Walse, B. Fragment-based drug discovery and protein–protein interactions. *Res. Reports Biochem.* **2014**, 13.

287. Si, Y.; Xu, D.; Bum-Erdene, K.; Ghozayel, M. K.; Yang, B.; Clemons, P. A.; Meroueh, S. O. Chemical Space Overlap with Critical Protein–Protein Interface Residues in Commercial and Specialized Small-Molecule Libraries. *ChemMedChem* **2019**, *14*, 119–131.

288. Kim, G. et al. FDA Approval Summary: Vemurafenib for Treatment of Unresectable or Metastatic Melanoma with the BRAF V600E Mutation. *Clin. Cancer Res.* **2014**, *20*, 4994–5000.

289. Sumimoto, H.; Imabayashi, F.; Iwata, T.; Kawakami, Y. The BRAF–MAPK signaling pathway is essential for cancer-immune evasion in human melanoma cells. *J. Exp. Med.* **2006**, *203*, 1651–1656.

290. Borst, A.; Haferkamp, S.; Grimm, J.; Rösch, M.; Zhu, G.; Guo, S.; Li, C.; Gao, T.; Meierjohann, S.; Schrama, D.; Houben, R. BIK is involved in BRAF/MEK inhibitor induced apoptosis in melanoma cell lines. *Cancer Lett.* **2017**, *404*, 70–78.

291. Davies, H. et al. Mutations of the BRAF gene in human cancer. *Nature* **2002**, *417*, 949–954.

292. Tsai, J. et al. Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. *Proc. Natl. Acad. Sci.* **2008**, *105*, 3041–3046.

293. Souers, A. J. et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nat. Med.* **2013**, *19*, 202–8.

294. Zhang, C. et al. Design and pharmacology of a highly specific dual FMS and KIT kinase inhibitor. *Proc. Natl. Acad. Sci.* **2013**, *110*, 5689–5694.

295. Murray, C. W.; Newell, D. R.; Angibaud, P. A successful collaboration between academia, biotech and pharma led to discovery of erdafitinib, a selective FGFR inhibitor recently approved by the FDA. *Medchemcomm* **2019**, *10*, 1509–1511.

296. Bridges, C. B. The Origin of Variations in Sexual and Sex-Limited Characters. *Am. Nat.* **1922**, *56*, 51–63.

297. Dobzhansky, T. Genetics of natural populations. XIII. Recombination and variability in populations of drosophila pseudoobscura. *Genetics* **1946**, *31*, 269–290.

298. Hartwell, L. H. Integrating Genetic Approaches into the Discovery of Anticancer Drugs. *Science (80-.).* **1997**, *278*, 1064–1068.

299. Nagel, R.; Semenova, E. A.; Berns, A. Drugging the addict: non-oncogene addiction as a target for cancer therapy. *EMBO Rep.* **2016**, *17*, 1516–1531.

300. Delmore, J. E. et al. BET Bromodomain Inhibition as a Therapeutic Strategy to Target c-Myc. *Cell* **2011**, *146*, 904–917.

301. Adjei, A. A. et al. Phase I Pharmacokinetic and Pharmacodynamic Study of the Oral, Small-Molecule Mitogen-Activated Protein Kinase Kinase 1/2 Inhibitor AZD6244 (ARRY-142886) in Patients With Advanced Cancers. *J. Clin. Oncol.* **2008**, *26*, 2139–2146.

302. Jänne, P. A.; Shaw, A. T.; Pereira, J. R.; Jeannin, G.; Vansteenkiste, J.; Barrios, C.; Franke, F. A.; Grinsted, L.; Zazulina, V.; Smith, P.; Smith, I.; Crinò, L. Selumetinib plus docetaxel for KRAS-mutant advanced non-small-cell lung cancer: a randomised, multicentre, placebo-controlled, phase 2 study. *Lancet Oncol.* **2013**, *14*, 38–47.

303. O'Neil, N. J.; Bailey, M. L.; Hieter, P. Synthetic lethality and cancer. *Nat. Rev. Genet.* **2017**, *18*, 613–623.

304. Bryant, H. E.; Schultz, N.; Thomas, H. D.; Parker, K. M.; Flower, D.; Lopez, E.; Kyle, S.; Meuth, M.; Curtin, N. J.; Helleday, T. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **2005**, *434*, 913–917.

305. Kaufman, B. et al. Olaparib Monotherapy in Patients With Advanced Cancer and a Germline BRCA1/2 Mutation. *J. Clin. Oncol.* **2015**, *33*, 244–250.

306. Hanzlikova, H.; Gittens, W.; Krejcikova, K.; Zeng, Z.; Caldecott, K. W. Overlapping roles for PARP1 and PARP2 in the recruitment of endogenous XRCC1 and PNKP into oxidized chromatin. *Nucleic Acids Res.* **2017**, *45*, 2546–2557.

307. Strom, C. E.; Johansson, F.; Uhlen, M.; Szigyarto, C. A.-K.; Erixon, K.; Helleday, T. Poly (ADP-ribose) polymerase (PARP) is not involved in base excision repair but PARP inhibition traps a single-strand intermediate. *Nucleic Acids Res.* **2011**, *39*, 3166–3175.

308. Ray Chaudhuri, A.; Nussenzweig, A. The multifaceted roles of PARP1 in DNA repair and chromatin remodelling. *Nat. Rev. Mol. Cell Biol.* **2017**, *18*, 610–621.

309. Dantzer, F.; de La Rubia, G.; Ménissier-De Murcia, J.; Hostomsky, Z.; de Murcia, G.; Schreiber, V. Base excision repair is impaired in mammalian cells lacking Poly(ADP-ribose) polymerase-1. *Biochemistry* **2000**, *39*, 7559–69.

310. Liu, C.; Vyas, A.; Kassab, M. A.; Singh, A. K.; Yu, X. The role of poly ADP-ribosylation in the first wave of DNA damage response. *Nucleic Acids Res.* **2017**, *45*, 8129–8141.

311. Horton, J. K.; Stefanick, D. F.; Prasad, R.; Gassman, N. R.; Kedar, P. S.; Wilson, S. H. Base excision repair defects invoke hypersensitivity to PARP inhibition. *Mol. Cancer Res.* **2014**, *12*, 1128–39.

312. Hopkins, T. A.; Shi, Y.; Rodriguez, L. E.; Solomon, L. R.; Donawho, C. K.; DiGiammarino, E. L.; Panchal, S. C.; Wilsbacher, J. L.; Gao, W.; Olson, A. M.; Stolarik, D. F.; Osterling, D. J.; Johnson, E. F.; Maag, D. Mechanistic Dissection of PARP1 Trapping and the Impact on In Vivo Tolerability and Efficacy of PARP

Inhibitors. *Mol. Cancer Res.* **2015**, *13*, 1465–1477.

313. Choi, Y. E.; Meghani, K.; Brault, M.-E.; Leclerc, L.; He, Y. J.; Day, T. A.; Elias, K. M.; Drapkin, R.; Weinstock, D. M.; Dao, F.; Shih, K. K.; Matulonis, U.; Levine, D. A.; Konstantinopoulos, P. A.; Chowdhury, D. Platinum and PARP Inhibitor Resistance Due to Overexpression of MicroRNA-622 in BRCA1-Mutant Ovarian Cancer. *Cell Rep.* **2016**, *14*, 429–439.

314. Friedlander, M. et al. Long-term efficacy, tolerability and overall survival in patients with platinum-sensitive, recurrent high-grade serous ovarian cancer treated with maintenance olaparib capsules following response to chemotherapy. *Br. J. Cancer* **2018**, *119*, 1075–1085.

315. Chan, E. M. et al. WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature* **2019**, *568*, 551–556.

316. Lieb, S. et al. Werner syndrome helicase is a selective vulnerability of microsatellite instability-high tumor cells. *Elife* **2019**, *8*, 1–22.

317. Kategaya, L.; Perumal, S. K.; Hager, J. H.; Belmont, L. D. Werner Syndrome Helicase Is Required for the Survival of Cancer Cells with Microsatellite Instability. *iScience* **2019**, *13*, 488–497.

318. van Wietmarschen, N. et al. Repeat expansions confer WRN dependence in microsatellite-unstable cancers. *Nature* **2020**, *586*, 292–298.

319. Sommers, J. A.; Kulikowicz, T.; Croteau, D. L.; Dexheimer, T.; Dorjsuren, D.; Jadhav, A.; Maloney, D. J.; Simeonov, A.; Bohr, V. A.; Brosh, R. M. A high-throughput screen to identify novel small molecule inhibitors of the Werner Syndrome Helicase-Nuclease (WRN). *PLoS One* **2019**, *14*, e0210525.

320. Schmid-Burgk, J. L.; Höning, K.; Ebert, T. S.; Hornung, V. CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat. Commun.* **2016**, *7*, 12338.

321. Sawatsubashi, S.; Joko, Y.; Fukumoto, S.; Matsumoto, T.; Sugano, S. S. Development of versatile non-homologous end joining-based knock-in module for genome editing. *Sci. Rep.* **2018**, *8*, 593.

322. Nakamae, K.; Nishimura, Y.; Takenaga, M.; Nakade, S.; Sakamoto, N.; Ide, H.; Sakuma, T.; Yamamoto, T. Establishment of expanded and streamlined pipeline of PITCh knock-in – a web-based design tool for MMEJ-mediated gene knock-in, PITCh designer, and the variations of PITCh, PITCh-TG and PITCh-KIKO. *Bioengineered* **2017**, *8*, 302–308.

323. Sakuma, T.; Nakade, S.; Sakane, Y.; Suzuki, K.-I. T.; Yamamoto, T. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat. Protoc.* **2016**, *11*, 118–33.

324. Kim, S.-I.; Matsumoto, T.; Kagawa, H.; Nakamura, M.; Hirohata, R.; Ueno, A.; Ohishi, M.; Sakuma, T.; Soga, T.; Yamamoto, T.; Woltjen, K. Microhomology-assisted scarless genome editing in human iPSCs. *Nat. Commun.* **2018**, *9*, 939.

325. Song, F.; Stieger, K. Optimizing the DNA Donor Template for Homology-Directed Repair of Double-Strand Breaks. *Mol. Ther. - Nucleic Acids* **2017**, *7*, 53–60.

326. Renaud, J.-B. et al. Improved Genome Editing Efficiency and Flexibility Using Modified Oligonucleotides with TALEN and CRISPR-Cas9 Nucleases. *Cell Rep.*

2016, *14*, 2263–2272.

327. Richardson, C. D.; Ray, G. J.; DeWitt, M. A.; Curie, G. L.; Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.* **2016**, *34*, 339–344.

328. Di Stazio, M.; Foschi, N.; Athanasakis, E.; Gasparini, P.; D'Adamo, A. P. Systematic analysis of factors that improve homologous direct repair (HDR) efficiency in CRISPR/Cas9 technique. *PLoS One* **2021**, *16*, e0247603.

329. Knott, G. J.; Doudna, J. A. CRISPR-Cas guides the future of genetic engineering. *Science (80-.).* **2018**, *361*, 866–869.

330. Mengwasser, K. E.; Adeyemi, R. O.; Leng, Y.; Choi, M. Y.; Clairmont, C.; D'Andrea, A. D.; Elledge, S. J. Genetic Screens Reveal FEN1 and APEX2 as BRCA2 Synthetic Lethal Targets. *Mol. Cell* **2019**, *73*, 885–899.e6.

331. Kampmann, M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chem. Biol.* **2018**, *13*, 406–416.

332. Charpentier, M. et al. CtIP fusion to Cas9 enhances transgene integration by homology-dependent repair. *Nat. Commun.* **2018**, *9*, 1133.

333. Shao, S.; Ren, C.; Liu, Z.; Bai, Y.; Chen, Z.; Wei, Z.; Wang, X.; Zhang, Z.; Xu, K. Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing Saccharomyces cerevisiae Rad52. *Int. J. Biochem. Cell Biol.* **2017**, *92*, 43–52.

334. Komor, A. C.; Kim, Y. B.; Packer, M. S.; Zuris, J. A.; Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **2016**, *533*, 420–424.

335. Gaudelli, N. M.; Komor, A. C.; Rees, H. A.; Packer, M. S.; Badran, A. H.; Bryson, D. I.; Liu, D. R. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* **2017**, *551*, 464–471.

336. Cox, D. B. T.; Gootenberg, J. S.; Abudayyeh, O. O.; Franklin, B.; Kellner, M. J.; Joung, J.; Zhang, F. RNA editing with CRISPR-Cas13. *Science (80-.).* **2017**, *358*, 1019–1027.

337. Robert, F.; Barbeau, M.; Éthier, S.; Dostie, J.; Pelletier, J. Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med.* **2015**, *7*, 93.

338. Kostyushev, D.; Kostyusheva, A.; Brezgin, S.; Zarifyan, D.; Utkina, A.; Goptar, I.; Chulanov, V. Suppressing the NHEJ pathway by DNA-PKcs inhibitor NU7026 prevents degradation of HBV cccDNA cleaved by CRISPR/Cas9. *Sci. Rep.* **2019**, *9*, 1847.

339. Chu, V. T.; Weber, T.; Wefers, B.; Wurst, W.; Sander, S.; Rajewsky, K.; Kühn, R. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.* **2015**, *33*, 543–548.

340. Li, H.; Beckman, K. A.; Pessino, V.; Huang, B.; Weissman, J. S.; Leonetti, M. D. Design and specificity of long ssDNA donors for CRISPR-based knock-in. *bioRxiv* **2017**, 178905.

341. You, L.; Tong, R.; Li, M.; Liu, Y.; Xue, J.; Lu, Y. Advancements and Obstacles of CRISPR-Cas9 Technology in Translational Research. *Mol. Ther. - Methods Clin. Dev.* **2019**, *13*, 359–370.

342. Sansbury, B. M.; Wagner, A. M.; Nitzan, E.; Tarcic, G.; Kmiec, E. B. CRISPR-Directed <i>In Vitro</i> Gene Editing of Plasmid DNA Catalyzed by Cpf1 (Cas12a) Nuclease and a Mammalian Cell-Free Extract. *Cris. J.* **2018**, *1*, 191–202.

343. Sansbury, B. M.; Wagner, A. M.; Tarcic, G.; Barth, S.; Nitzan, E.; Goldfus, R.; Vidne, M.; Kmiec, E. B. CRISPR-Directed Gene Editing Catalyzes Precise Gene Segment Replacement <i>In Vitro</i> Enabling a Novel Method for Multiplex Site-Directed Mutagenesis. *Cris. J.* **2019**, *2*, 121–132.

344. Anders, C.; Niewoehner, O.; Duerst, A.; Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **2014**, *513*, 569–573.

345. Abakir, A. et al. N6-methyladenosine regulates the stability of RNA:DNA hybrids in human cells. *Nat. Genet.* **2020**, *52*, 48–55.

346. Hamperl, S.; Bocek, M. J.; Saldivar, J. C.; Swigut, T.; Cimprich, K. A. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **2017**, *170*, 774–786.e19.

347. Rajani, S.; Gell, C.; Abakir, A.; Markus, R. Computational Analysis of DNA Modifications in Confocal Images. *Methods Mol. Biol.* **2021**, *2198*, 227–254.

348. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676–682.

349. Cole-Strauss, A. Targeted gene repair directed by the chimeric RNA/DNA oligonucleotide in a mammalian cell-free extract. *Nucleic Acids Res.* **1999**, *27*, 1323–1330.

350. Sternberg, S. H.; Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol. Cell* **2015**, *58*, 568–574.

351. Chen, F.; Pruett-Miller, S. M.; Davis, G. D. Gene editing using ssODNs with engineered endonucleases. *Methods Mol. Biol.* **2015**, *1239*, 251–65.

352. Okamoto, S.; Amaishi, Y.; Maki, I.; Enoki, T.; Mineno, J. Highly efficient genome editing for single-base substitutions using optimized ssODNs with Cas9-RNPs. *Sci. Rep.* **2019**, *9*, 4811.

353. Thomas, M.; White, R. L.; Davis, R. W. Hybridization of RNA to double-stranded DNA: formation of R-loops. *Proc. Natl. Acad. Sci. U. S. A.* **1976**, *73*, 2294–8.

354. Gomez-Gonzalez, B.; Aguilera, A. Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc. Natl. Acad. Sci.* **2007**, *104*, 8409–8414.

355. Cristini, A.; Ricci, G.; Britton, S.; Salimbeni, S.; Huang, S.-y. N.; Marinello, J.; Calsou, P.; Pommier, Y.; Favre, G.; Capranico, G.; Gromak, N.; Sordet, O. Dual Processing of R-Loops and Topoisomerase I Induces Transcription-Dependent DNA Double-Strand Breaks. *Cell Rep.* **2019**, *28*, 3167–3181.e6.

356. Freudenreich, C. H. R-loops: targets for nuclease cleavage and repeat instability. *Curr. Genet.* **2018**, *64*, 789–794.

357. Sollier, J.; Stork, C. T.; García-Rubio, M. L.; Paulsen, R. D.; Aguilera, A.; Cimprich, K. A. Transcription-Coupled Nucleotide Excision Repair Factors Promote R-Loop-Induced Genome Instability. *Mol. Cell* **2014**, *56*, 777–785.

358. Dutta, D.; Shatalin, K.; Epshtein, V.; Gottesman, M. E.; Nudler, E. Linking RNA Polymerase Backtracking to Genome Instability in E. coli. *Cell* **2011**, *146*, 533–543.

359. Lang, K. S.; Hall, A. N.; Merrikh, C. N.; Ragheb, M.; Tabakh, H.; Pollock, A. J.; Woodward, J. J.; Dreifus, J. E.; Merrikh, H. Replication-Transcription Conflicts Generate R-Loops that Orchestrate Bacterial Stress Survival and Pathogenesis. *Cell* **2017**, *170*, 787–799.e18.

360. Popuri, V.; Bachrati, C. Z.; Muzzolini, L.; Mosedale, G.; Costantini, S.; Giacomini, E.; Hickson, I. D.; Vindigni, A. The Human RecQ Helicases, BLM and RECQ1, Display Distinct DNA Substrate Specificities. *J. Biol. Chem.* **2008**, *283*, 17766–17776.

361. Chang, E. Y.-C.; Novoa, C. A.; Aristizabal, M. J.; Coulombe, Y.; Segovia, R.; Chaturvedi, R.; Shen, Y.; Keong, C.; Tam, A. S.; Jones, S. J. M.; Masson, J.-Y.; Kobor, M. S.; Stirling, P. C. RECQ-like helicases Sgs1 and BLM regulate R-loop-associated genome instability. *J. Cell Biol.* **2017**, *216*, 3991–4005.

362. Chakraborty, P.; Huang, J. T. J.; Hiom, K. DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat. Commun.* **2018**, *9*, 4346.

363. Tran, P. L. T.; Pohl, T. J.; Chen, C.-F.; Chan, A.; Pott, S.; Zakian, V. A. PIF1 family DNA helicases suppress R-loop mediated genome instability at tRNA genes. *Nat. Commun.* **2017**, *8*, 15025.

364. Ozdemir, A. Y.; Rusanov, T.; Kent, T.; Siddique, L. A.; Pomerantz, R. T. Polymerase θ-helicase efficiently unwinds DNA and RNA-DNA hybrids. *J. Biol. Chem.* **2018**, *293*, 5259–5269.

365. Chandramouly, G. et al. Polθ reverse transcribes RNA and promotes RNA-templated DNA repair. *Sci. Adv.* **2021**, *7*, 1–11.

366. Guan, D.; Chen, Z. Challenges and recent advances in affinity purification of tag-free proteins. *Biotechnol. Lett.* **2014**, *36*, 1391–1406.

367. Sternberg, S. H.; Redding, S.; Jinek, M.; Greene, E. C.; Doudna, J. A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **2014**, *507*, 62–67.

368. Zhang, Q.; Wen, F.; Zhang, S.; Jin, J.; Bi, L.; Lu, Y.; Li, M.; Xi, X.-G.; Huang, X.; Shen, B.; Sun, B. The post-PAM interaction of RNA-guided spCas9 with DNA dictates its target binding and dissociation. *Sci. Adv.* **2019**, *5*, eaaw9807.

369. Clarke, R.; Heler, R.; MacDougall, M. S.; Yeo, N. C.; Chavez, A.; Regan, M.; Hanakahi, L.; Church, G. M.; Marraffini, L. A.; Merrill, B. J. Enhanced Bacterial Immunity and Mammalian Genome Editing via RNA-Polymerase-Mediated Dislodging of Cas9 from Double-Strand DNA Breaks. *Mol. Cell* **2018**, *71*, 42–55.e8.

370. Schauer, G. D.; Spenkelink, L. M.; Lewis, J. S.; Yurieva, O.; Mueller, S. H.; van Oijen, A. M.; O'Donnell, M. E. Replisome bypass of a protein-based R-loop block by Pif1. *Proc. Natl. Acad. Sci.* **2020**, *117*, 30354–30361.

371. Takata, K.-i. et al. Analysis of DNA polymerase ν function in meiotic recombination, immunoglobulin class-switching, and DNA damage tolerance. *PLOS Genet.* **2017**, *13*, e1006818.

372. Gregorio, N. E.; Levine, M. Z.; Oza, J. P. A User's Guide to Cell-Free Protein

Synthesis. *Methods Protoc.* **2019**, *2*, 24.

373. Heide, C.; Buldum, G.; Moya-Ramirez, I.; Ces, O.; Kontoravdi, C.; Polizzi, K. M. Design, Development and Optimization of a Functional Mammalian Cell-Free Protein Synthesis Platform. *Front. Bioeng. Biotechnol.* **2021**, *8*, 1–10.

374. Nguyen, H. D.; Yadav, T.; Giri, S.; Saez, B.; Graubert, T. A.; Zou, L. Functions of Replication Protein A as a Sensor of R Loops and a Regulator of RNaseH1. *Mol. Cell* **2017**, *65*, 832–847.e4.

375. García-Rubio, M. L.; Pérez-Calero, C.; Barroso, S. I.; Tumini, E.; Herrera-Moyano, E.; Rosado, I. V.; Aguilera, A. The Fanconi Anemia Pathway Protects Genome Integrity from R-loops. *PLOS Genet.* **2015**, *11*, e1005674.

376. Liang, Z.; Liang, F.; Teng, Y.; Chen, X.; Liu, J.; Longerich, S.; Rao, T.; Green, A. M.; Collins, N. B.; Xiong, Y.; Lan, L.; Sung, P.; Kupfer, G. M. Binding of FANCI-FANCD2 Complex to RNA and R-Loops Stimulates Robust FANCD2 Monoubiquitination. *Cell Rep.* **2019**, *26*, 564–572.e5.

377. Boguslawski, S. J.; Smith, D. E.; Michalak, M. A.; Mickelson, K. E.; Yehle, C. O.; Patterson, W. L.; Carrico, R. J. Characterization of monoclonal antibody to DNA · RNA and its application to immunodetection of hybrids. *J. Immunol. Methods* **1986**, *89*, 123–130.

378. König, F.; Schubert, T.; Längst, G. The monoclonal S9.6 antibody exhibits highly variable binding affinities towards different R-loop sequences. *PLoS One* **2017**, *12*, e0178875.

379. Phillips, D. D.; Garboczi, D. N.; Singh, K.; Hu, Z.; Leppla, S. H.; Leysath, C. E. The sub-nanomolar binding of DNA-RNA hybrids by the single-chain Fv fragment of antibody S9.6. *J. Mol. Recognit.* **2013**, *26*, 376–381.

380. Smolka, J. A.; Sanz, L. A.; Hartono, S. R.; Chédin, F. Recognition of RNA by the S9.6 antibody creates pervasive artifacts when imaging RNA:DNA hybrids. *J. Cell Biol.* **2021**, *220*.

381. Francia, S.; Cabrini, M.; Matti, V.; Oldani, A.; d'Adda di Fagagna, F. DICER, DROSHA and DNA damage response RNAs are necessary for the secondary recruitment of DNA damage response factors. *J. Cell Sci.* **2016**, *129*, 1468–76.

382. D'Alessandro, G. et al. BRCA2 controls DNA:RNA hybrid level at DSBs by mediating RNase H2 recruitment. *Nat. Commun.* **2018**, *9*, 5376.

383. Yeeles, J. T.; Janska, A.; Early, A.; Diffley, J. F. How the Eukaryotic Replisome Achieves Rapid and Efficient DNA Replication. *Mol. Cell* **2017**, *65*, 105–116.

384. Nakanishi, K.; Cavallo, F.; Brunet, E.; Jasin, M. Homologous recombination assay for interstrand cross-link repair. *Methods Mol. Biol.* **2011**, *745*, 283–91.

385. Sabatella, M.; Theil, A. F.; Ribeiro-Silva, C.; Slyskova, J.; Thijssen, K.; Voskamp, C.; Lans, H.; Vermeulen, W. Repair protein persistence at DNA lesions characterizes XPF defect with Cockayne syndrome features. *Nucleic Acids Res.* **2018**, *46*, 9563–9577.

386. Arora, S.; Kothandapani, A.; Tillison, K.; Kalman-Maltese, V.; Patrick, S. M. Downregulation of XPF-ERCC1 enhances cisplatin efficacy in cancer cells. *DNA Repair (Amst).* **2010**, *9*, 745–53.

387. Schneeweiss-Gleixner, M. et al. CDK4/CDK6 inhibition as a novel strategy to suppress the growth and survival of BCR-ABL1T315I+ clones in TKI-resistant

CML. *EBioMedicine* **2019**, *50*, 111–121.

388. Rogers, R. F.; Walton, M. I.; Cherry, D. L.; Collins, I.; Clarke, P. A.; Garrett, M. D.; Workman, P. CHK1 Inhibition Is Synthetically Lethal with Loss of B-Family DNA Polymerase Function in Human Lung and Colorectal Cancer Cells. *Cancer Res.* **2020**, *80*, 1735–1747.

389. Makino, S.; Fukumura, R.; Gondo, Y. Illegitimate translation causes unexpected gene expression from on-target out-of-frame alleles created by CRISPR-Cas9. *Sci. Rep.* **2016**, *6*, 39608.

390. Tuladhar, R.; Yeu, Y.; Tyler Piazza, J.; Tan, Z.; Rene Clemenceau, J.; Wu, X.; Barrett, Q.; Herbert, J.; Mathews, D. H.; Kim, J.; Hyun Hwang, T.; Lum, L. CRISPR-Cas9-based mutagenesis frequently provokes on-target mRNA misregulation. *Nat. Commun.* **2019**, *10*, 4056.

391. Stepanenko, A.; Dmitrenko, V. Pitfalls of the MTT assay: Direct and off-target effects of inhibitors can result in over/underestimation of cell viability. *Gene* **2015**, *574*, 193–203.

392. O'Donovan, M. A critique of methods to measure cytotoxicity in mammalian cell genotoxicity assays. *Mutagenesis* **2012**, *27*, 615–621.

393. Jo, H. Y.; Kim, Y.; Park, H. W.; Moon, H. E.; Bae, S.; Kim, J.; Kim, D. G.; Paek, S. H. The Unreliability of MTT Assay in the Cytotoxic Test of Primary Cultured Glioblastoma Cells. *Exp. Neurobiol.* **2015**, *24*, 235–245.

394. Ostling, O.; Johanson, K. Microelectrophoretic study of radiation-induced DNA damages in individual mammalian cells. *Biochem. Biophys. Res. Commun.* **1984**, *123*, 291–298.

395. Wu, J. H.; Wilson, J. B.; Wolfreys, A. M.; Scott, A.; Jones, N. J. Optimization of the comet assay for the sensitive detection of PUVA-induced DNA interstrand cross-links. *Mutagenesis* **2008**, *24*, 173–181.

396. Zimak, J.; Wagoner, Z. W.; Nelson, N.; Waechtler, B.; Schlosser, H.; Kopecky, M.; Wu, J.; Zhao, W. Epigenetic silencing directs expression heterogeneity of stably integrated multi-transcript unit genetic circuits. *Sci. Rep.* **2021**, *11*, 2424.

397. Alhaji, S. Y.; Ngai, S. C.; Abdullah, S. Silencing of transgene expression in mammalian cells by DNA methylation and histone modifications in gene therapy perspective. *Biotechnol. Genet. Eng. Rev.* **2019**, *35*, 1–25.

398. Albain, K. S.; Swinnen, L. J.; Erickson, L. C.; Stiff, P. J.; Fisher, S. G.; Fisher, R. I. Cytotoxic synergy of cisplatin with concurrent hydroxyurea and cytarabine: summary of an in vitro model and initial clinical pilot experience. *Semin. Oncol.* **1992**, *19*, 102–9.

399. Gao, Y.; He, Y.; Xu, J.; Xu, L.; Du, J.; Zhu, C.; Gu, H.; Ma, H.; Hu, Z.; Jin, G.; Chen, X.; Shen, H. Genetic variants at 4q21, 4q23 and 12q24 are associated with esophageal squamous cell carcinoma risk in a Chinese population. *Hum. Genet.* **2013**, *132*, 649–656.

400. Banroques, J.; Cordin, O.; Doere, M.; Linder, P.; Tanner, N. K. A Conserved Phenylalanine of Motif IV in Superfamily 2 Helicases Is Required for Cooperative, ATP-Dependent Binding of RNA Substrates in DEAD-Box Proteins. *Mol. Cell. Biol.* **2008**, *28*, 3359–3371.

401. Sjoblom, T. et al. The Consensus Coding Sequences of Human Breast and

Colorectal Cancers. *Science (80-.).* **2006**, *314*, 268–274.

402. Yachida, S. et al. Genomic Sequencing Identifies ELF3 as a Driver of Ampullary Carcinoma. *Cancer Cell* **2016**, *29*, 229–240.

403. Boyd, J. B.; Golino, M. D.; Nguyen, T. D.; Green, M. M. Isolation and characterization of X-linked mutants of Drosophila melanogaster which are sensitive to mutagens. *Genetics* **1976**, *84*, 485–506.

404. Fairman-Williams, M. E.; Guenther, U.-P.; Jankowsky, E. SF1 and SF2 helicases: family matters. *Curr. Opin. Struct. Biol.* **2010**, *20*, 313–324.

405. Richards, J. D.; Johnson, K. A.; Liu, H.; McRobbie, A.-M.; McMahon, S.; Oke, M.; Carter, L.; Naismith, J. H.; White, M. F. Structure of the DNA Repair Helicase Hel308 Reveals DNA Binding and Autoinhibitory Domains. *J. Biol. Chem.* **2008**, *283*, 5118–5126.

406. Marini, F.; Kim, N.; Schuffert, A.; Wood, R. D. POLN, a nuclear PolA family DNA polymerase homologous to the DNA cross-link sensitivity protein Mus308. *J. Biol. Chem.* **2003**, *278*, 32014–32019.

407. Liang, C.; Marsit, C. J.; Houseman, E. A.; Butler, R.; Nelson, H. H.; McClean, M. D.; Kelsey, K. T. Gene-environment interactions of novel variants associated with head and neck cancer. *Head Neck* **2012**, *34*, 1111–1118.

408. Li, X.; Burnight, E. R.; Cooney, A. L.; Malani, N.; Brady, T.; Sander, J. D.; Staber, J.; Wheelan, S. J.; Joung, J. K.; McCray, P. B.; Bushman, F. D.; Sinn, P. L.; Craig, N. L. piggyBac transposase tools for genome engineering. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2279–87.

409. Weinstein, I. B.; Joe, A. Oncogene Addiction. *Cancer Res.* **2008**, *68*, 3077–3080.

410. Kaelin, W. G. Common pitfalls in preclinical cancer target validation. *Nat. Rev. Cancer* **2017**, *17*, 441–450.

411. Dillon, K. J.; Smith, G. C. M.; Martin, N. M. B. A FlashPlate Assay for the Identification of PARP-1 Inhibitors. *J. Biomol. Screen.* **2003**, *8*, 347–352.

412. Lamoree, B.; Hubbard, R. E. Current perspectives in fragment-based lead discovery (FBLD). *Essays Biochem.* **2017**, *61*, 453–464.

413. Sturtevant, A. H. A Highly Specific Complementary Lethal System in Drosophila Melanogaster. *Genetics* **1956**, *41*, 118–123.

414. Huang, A.; Garraway, L. A.; Ashworth, A.; Weber, B. Synthetic lethality as an engine for cancer drug target discovery. *Nat. Rev. Drug Discov.* **2020**, *19*, 23–38.

415. Topatana, W.; Juengpanich, S.; Li, S.; Cao, J.; Hu, J.; Lee, J.; Suliyanto, K.; Ma, D.; Zhang, B.; Chen, M.; Cai, X. Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. *J. Hematol. Oncol.* **2020**, *13*, 118.

416. Dezső, Z.; Ceccarelli, M. Machine learning prediction of oncology drug targets based on protein and network properties. *BMC Bioinformatics* **2020**, *21*, 104.

417. Bazaga, A.; Leggate, D.; Weisser, H. Genome-wide investigation of gene-cancer associations for the prediction of novel therapeutic targets in oncology. *Sci. Rep.* **2020**, *10*, 10787.

418. Huang, Y.; Li, L. DNA crosslinking damage and cancer - a tale of friend and foe. *Transl. Cancer Res.* **2013**, *2*, 144–154.

419. Rycenga, H. B.; Long, D. T. The evolving role of DNA inter-strand crosslinks in

chemotherapy. *Curr. Opin. Pharmacol.* **2018**, *41*, 20–26.

420. McKay, J. D. et al. A Genome-Wide Association Study of Upper Aerodigestive Tract Cancers Conducted within the INHANCE Consortium. *PLoS Genet.* **2011**, *7*, e1001333.

421. Li, W.-Q. et al. Genetic variants in DNA repair pathway genes and risk of esophageal squamous cell carcinoma and gastric adenocarcinoma in a Chinese population. *Carcinogenesis* **2013**, *34*, 1536–1542.

422. Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *J. Med. Chem.* **2011**, *54*, 7784–7796.

423. Tjernberg, A.; Markova, N.; Griffiths, W. J.; Hallén, D. DMSO-related effects in protein characterization. *J. Biomol. Screen.* **2006**, *11*, 131–137.

424. Kombo, D. C.; Tallapragada, K.; Jain, R.; Chewning, J.; Mazurov, A. A.; Speake, J. D.; Hauser, T. A.; Toler, S. 3D molecular descriptors important for clinical success. *J. Chem. Inf. Model.* **2013**, *53*, 327–342.

425. Kumar, M.; Lowery, R. G. A High-Throughput Method for Measuring Drug Residence Time Using the Transcreener ADP Assay. *SLAS Discov.* **2017**, *22*, 915–922.

426. Kumar, M.; Lowery, R. G. Development of a High-Throughput Assay to Identify Inhibitors of ENPP1. *SLAS Discov. Adv. life Sci. R D* **2021**, *26*, 740–746.

427. Ahmad, S.; Hughes, M. A.; Johnson, G. L.; Scott, J. E. Development and validation of a high-throughput intrinsic ATPase activity assay for the discovery of MEKK2 inhibitors. *J. Biomol. Screen.* **2013**, *18*, 388–399.

428. Chan, L. L.; Lidstone, E. A.; Finch, K. E.; Heeres, J. T.; Hergenrother, P. J.; Cunningham, B. T. A Method for Identifying Small-Molecule Aggregators Using Photonic Crystal Biosensor Microplates. *J. Lab. Autom.* **2009**, *14*, 348–359.

429. Danaei, M.; Dehghankhold, M.; Ataei, S.; Hasanzadeh Davarani, F.; Javanmard, R.; Dokhani, A.; Khorasani, S.; Mozafari, M. Impact of Particle Size and Polydispersity Index on the Clinical Applications of Lipidic Nanocarrier Systems. *Pharmaceutics* **2018**, *10*, 57.

430. Hughes, J.; Rees, S.; Kalindjian, S.; Philpott, K. Principles of early drug discovery. *Br. J. Pharmacol.* **2011**, *162*, 1239–1249.

431. Wouters, O. J.; McKee, M.; Luyten, J. Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018. *JAMA* **2020**, *323*, 844.

432. Hingorani, A. D.; Kuan, V.; Finan, C.; Kruger, F. A.; Gaulton, A.; Chopade, S.; Sofat, R.; MacAllister, R. J.; Overington, J. P.; Hemingway, H.; Denaxas, S.; Prieto, D.; Casas, J. P. Improving the odds of drug development success through human genomics: modelling study. *Sci. Rep.* **2019**, *9*, 18911.

433. Galkina Cleary, E.; Beierlein, J. M.; Khanuja, N. S.; McNamee, L. M.; Ledley, F. D. Contribution of NIH funding to new drug approvals 2010–2016. *Proc. Natl. Acad. Sci.* **2018**, *115*, 2329–2334.

434. Takebe, T.; Imai, R.; Ono, S. The Current Status of Drug Discovery and Development as Originated in United States Academia: The Influence of Industrial and

Academic Collaboration on Drug Discovery and Development. *Clin. Transl. Sci.* **2018**, *11*, 597–606.

435. Frearson, J.; Wyatt, P. Drug discovery in academia: the third way? *Expert Opin. Drug Discov.* **2010**, *5*, 909–919.

436. Messick, T. E. et al. Structure-based design of small-molecule inhibitors of EBNA1 DNA binding blocks Epstein-Barr virus latent infection and tumor growth. *Sci. Transl. Med.* **2019**, *11*.

437. Böttcher, J. et al. Fragment-based discovery of a chemical probe for the PWWP1 domain of NSD3. *Nat. Chem. Biol.* **2019**, *15*, 822–829.

438. Shi, Y.; von Itzstein, M. How Size Matters: Diversity for Fragment Library Design. *Molecules* **2019**, *24*, 2838.

439. Jhoti, H.; Williams, G.; Rees, D. C.; Murray, C. W. The 'rule of three' for fragment-based drug discovery: where are we now? *Nat. Rev. Drug Discov.* **2013**, *12*, 644–644.

440. Hann, M. M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 856–864.

441. Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A small nonrule of 3 compatible fragment library provides high hit rate of endothiapepsin crystal structures with various fragment chemotypes. *J. Med. Chem.* **2011**, *54*, 7784–7796.

442. Keserű, G. M.; Erlanson, D. A.; Ferenczy, G. G.; Hann, M. M.; Murray, C. W.; Pickett, S. D. Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J. Med. Chem.* **2016**, *59*, 8189–8206.

443. Duong-Thi, M.-D.; Bergström, M.; Fex, T.; Isaksson, R.; Ohlson, S. High-Throughput Fragment Screening by Affinity LC-MS. *J. Biomol. Screen.* **2013**, *18*, 160–171.

444. Chan, D. S.-H.; Whitehouse, A. J.; Coyne, A. G.; Abell, C. Mass spectrometry for fragment screening. *Essays Biochem.* **2017**, *61*, 465–473.

445. Rogers, M. S.; Cryan, L. M.; Habeshian, K. A.; Bazinet, L.; Caldwell, T. P.; Ackroyd, P. C.; Christensen, K. A. A FRET-Based High Throughput Screening Assay to Identify Inhibitors of Anthrax Protective Antigen Binding to Capillary Morphogenesis Gene 2 Protein. *PLoS One* **2012**, *7*, e39911.

446. Chung, C.-w.; Dean, A. W.; Woolven, J. M.; Bamborough, P. Fragment-Based Discovery of Bromodomain Inhibitors Part 1: Inhibitor Binding Modes and Implications for Lead Discovery. *J. Med. Chem.* **2012**, *55*, 576–586.

447. Miller, T. W.; Amason, J. D.; Garcin, E. D.; Lamy, L.; Dranchak, P. K.; Macarthur, R.; Braisted, J.; Rubin, J. S.; Burgess, T. L.; Farrell, C. L.; Roberts, D. D.; Inglese, J. Quantitative high-throughput screening assays for the discovery and development of SIRPα-CD47 interaction inhibitors. *PLoS One* **2019**, *14*, e0218897.

448. Vom, A.; Headey, S.; Wang, G.; Capuano, B.; Yuriev, E.; Scanlon, M. J.; Simpson, J. S. Detection and Prevention of Aggregation-based False Positives in STD-NMR-based Fragment Screening. *Aust. J. Chem.* **2013**, *66*, 1518.

256

449. Blevitt, J. M.; Hack, M. D.; Herman, K. L.; Jackson, P. F.; Krawczuk, P. J.; Lebsack, A. D.; Liu, A. X.; Mirzadegan, T.; Nelen, M. I.; Patrick, A. N.; Steinbacher, S.; Milla, M. E.; Lumb, K. J. Structural Basis of Small-Molecule Aggregate Induced Inhibition of a Protein–Protein Interaction. *J. Med. Chem.* **2017**, *60*, 3511–3517.

450. Feng, B. Y.; Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors. *Nat. Protoc.* **2006**, *1*, 550–553.

451. Hanson, S. M.; Georghiou, G.; Thakur, M. K.; Miller, W. T.; Rest, J. S.; Chodera, J. D.; Seeliger, M. A. What Makes a Kinase Promiscuous for Inhibitors? *Cell Chem. Biol.* **2019**, *26*, 390–399.e5.

452. Murray, C. W.; Rees, D. C. The rise of fragment-based drug discovery. *Nat. Chem.* **2009**, *1*, 187–192.

453. Liston, D. R.; Davis, M. Clinically Relevant Concentrations of Anticancer Drugs: A Guide for Nonclinical Studies. *Clin. Cancer Res.* **2017**, *23*, 3489–3498.

454. Valuchova, S.; Fulnecek, J.; Petrov, A. P.; Tripsianes, K.; Riha, K. A rapid method for detecting protein-nucleic acid interactions by protein induced fluorescence enhancement. *Sci. Rep.* **2016**, *6*, 39653.

455. Rachman, M.; Bajusz, D.; Hetényi, A.; Scarpino, A.; Merő, B.; Egyed, A.; Buday, L.; Barril, X.; Keserű, G. M. Discovery of a novel kinase hinge binder fragment by dynamic undocking. *RSC Med. Chem.* **2020**, *11*, 552–558.

456. Hendrickson, O. D.; Taranova, N. A.; Zherdev, A. V.; Dzantiev, B. B.; Eremin, S. A. Fluorescence Polarization-Based Bioassays: New Horizons. *Sensors* **2020**, *20*, 7132.

457. Langhans, S. A. Three-Dimensional in Vitro Cell Culture Models in Drug Discovery and Drug Repositioning. *Front. Pharmacol.* **2018**, *9*, 1–14.

458. Johnson, S. W.; Laub, P. B.; Beesley, J. S.; Ozols, R. F.; Hamilton, T. C. Increased platinum-DNA damage tolerance is associated with cisplatin resistance and cross-resistance to various chemotherapeutic agents in unrelated human ovarian cancer cell lines. *Cancer Res.* **1997**, *57*, 850–6.

459. Alley, M. C.; Scudiero, D. A.; Monks, A.; Hursey, M. L.; Czerwinski, M. J.; Fine, D. L.; Abbott, B. J.; Mayo, J. G.; Shoemaker, R. H.; Boyd, M. R. Feasibility of drug screening with panels of human tumor cell lines using a microculture tetrazolium assay. *Cancer Res.* **1988**, *48*, 589–601.

460. Killelea, T.; Bolt, E. CRISPR-Cas adaptive immunity and the three Rs. *Biosci. Rep.* **2017**, *37*, BSR20160297.

A
Appendices

A.1 Supplementary figures

See overleaf.

A.1.1 Chromatograms for protein purification

A.1.1.1 Purification of *Streptococcus pyogenes* Cas9
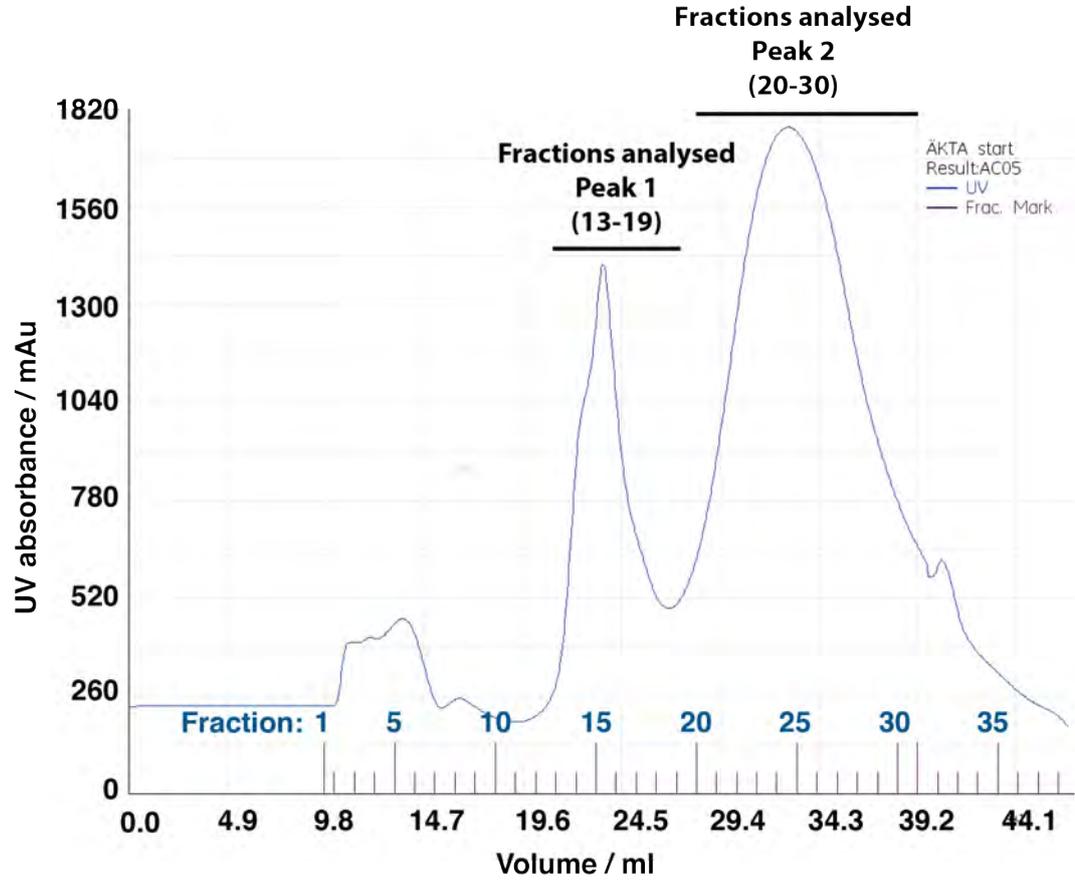
Figure A1: Chromatogram for Ni²⁺-affinity purification of Cas9 from clarified cell lysate.
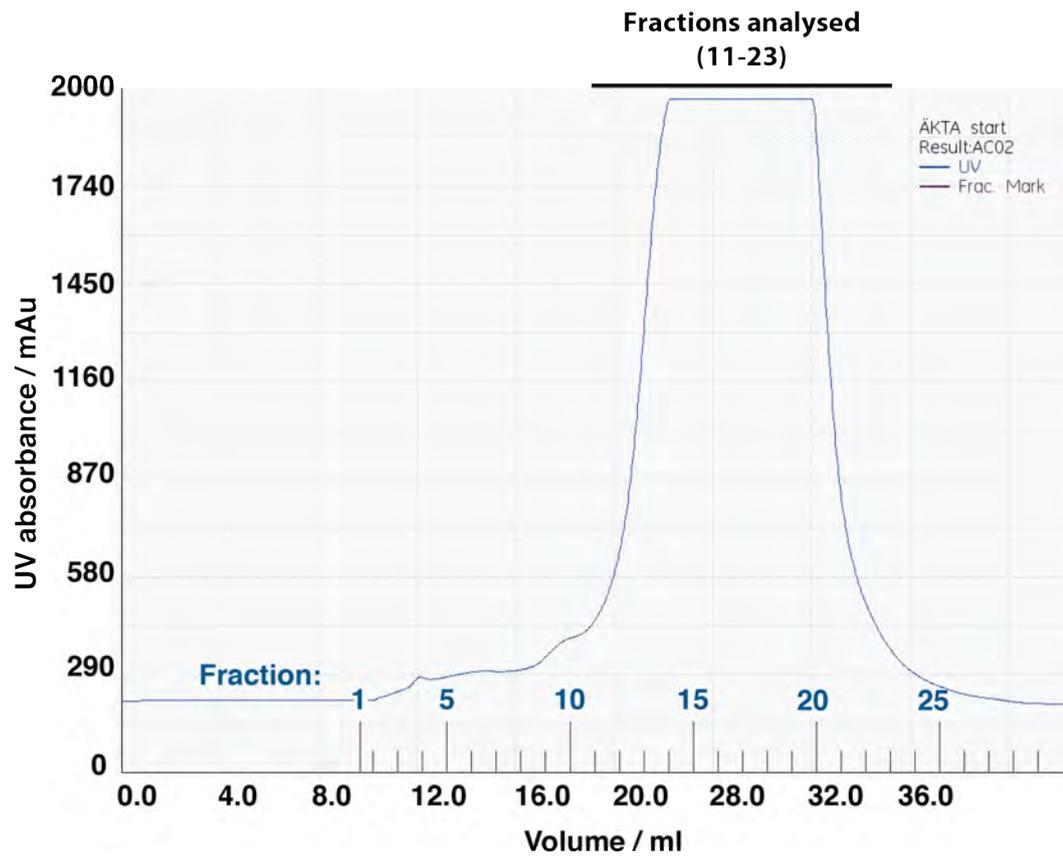
Figure A2: Chromatogram for Heparin purification of Cas9 following Ni^{2+}-affinity purification.
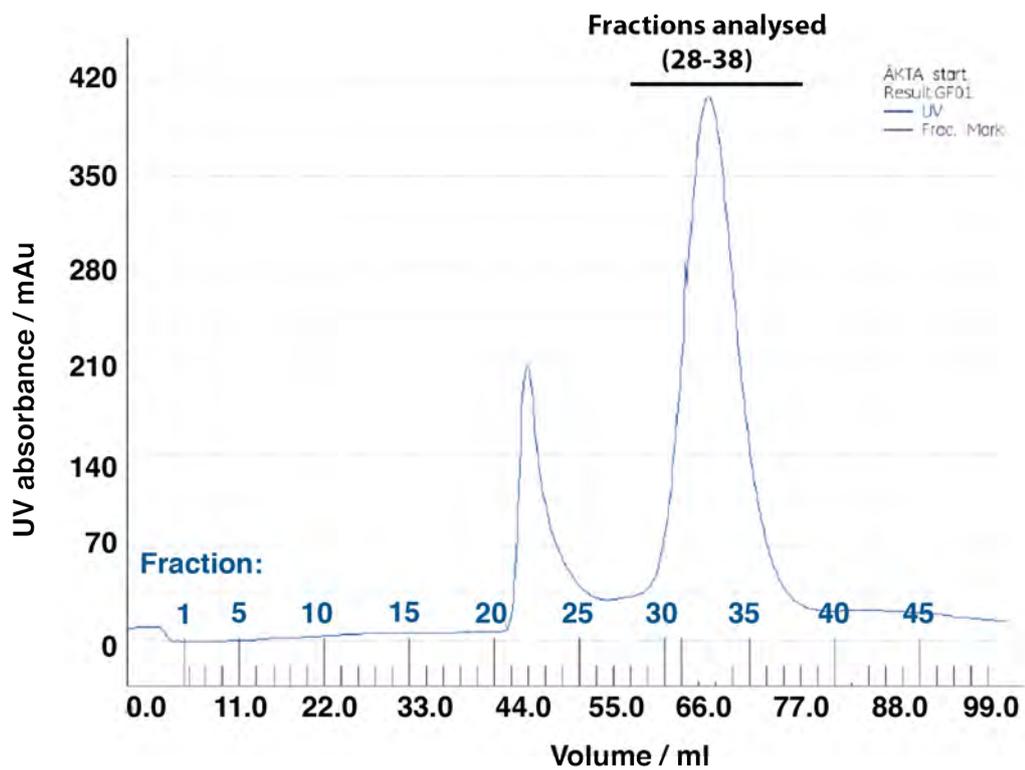
Figure A3: Chromatogram for size-exclusion purification of Cas9 following heparin-affinity purification.
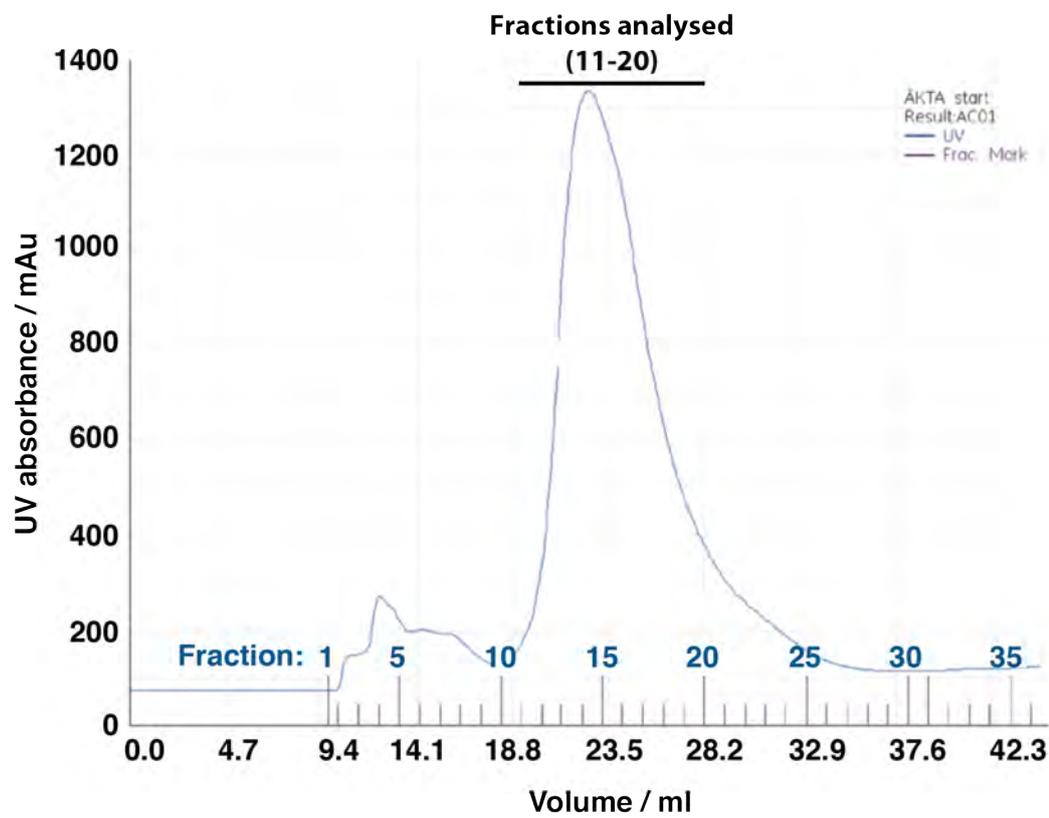
Figure A4: Chromatogram for Ni^{2+}-affinity purification of dCas9 from clarified cell lysate.
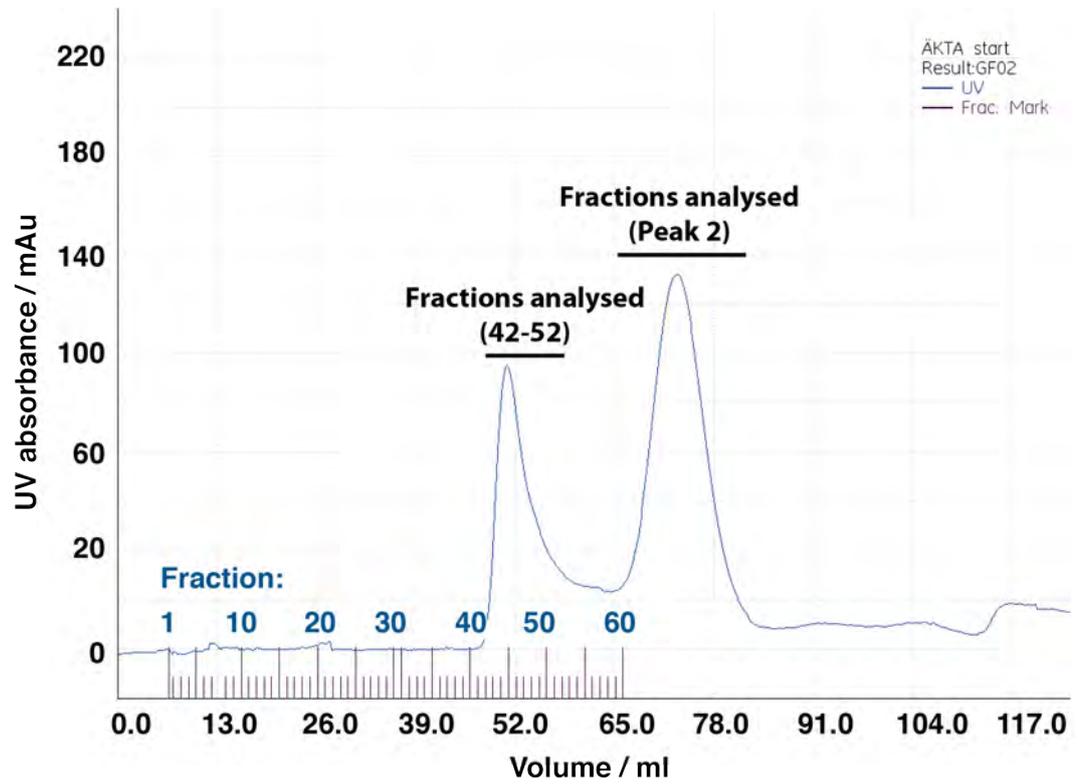
Figure A5: Chromatogram for size-exclusion purification of dCas9 following Ni^{2+} and heparin-affinity purification.
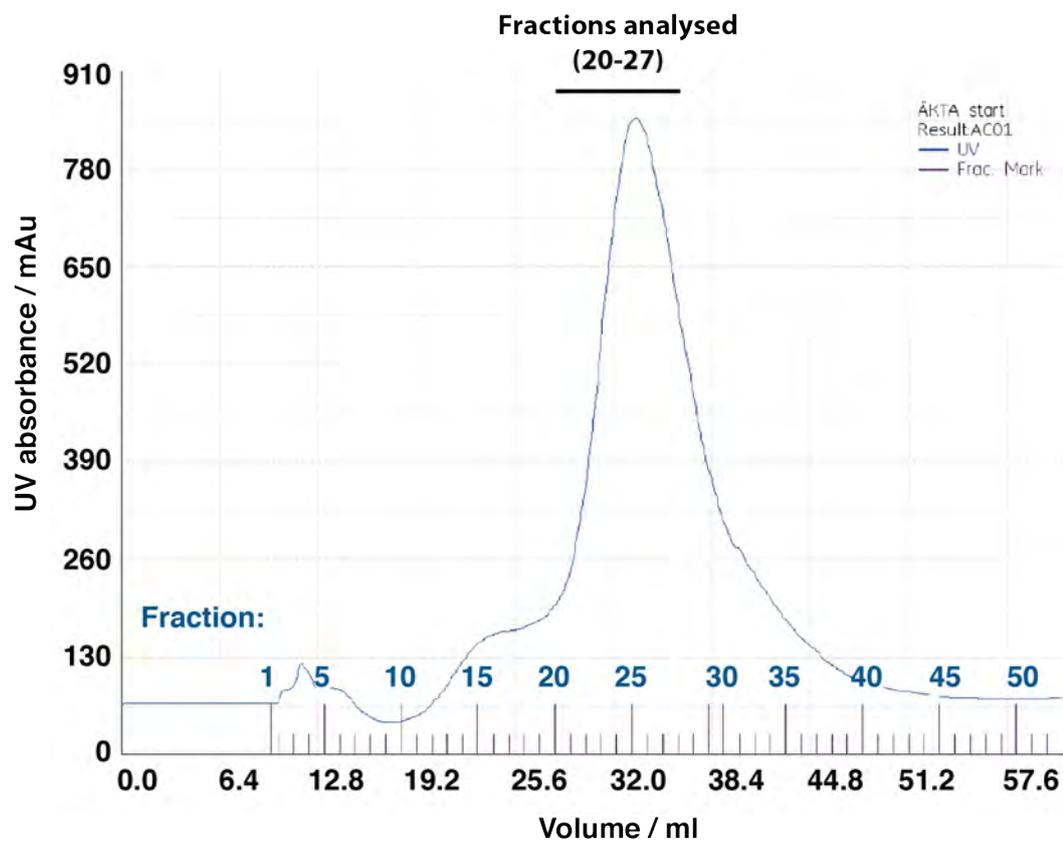
Figure A6: Chromatogram for Ni^{2+}-affinity purification of Cas12a from clarified cell lysate.
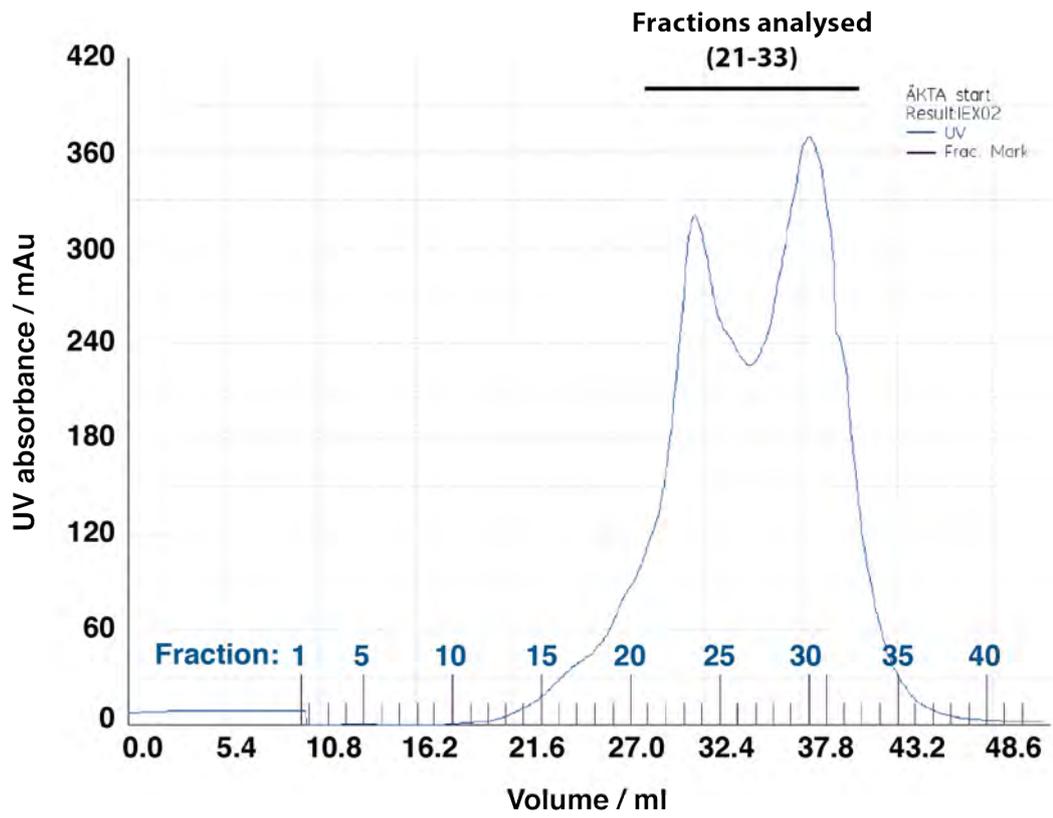
Figure A7: Chromatogram for Heparin purification of Cas12a following Ni²⁺-affinity purification.
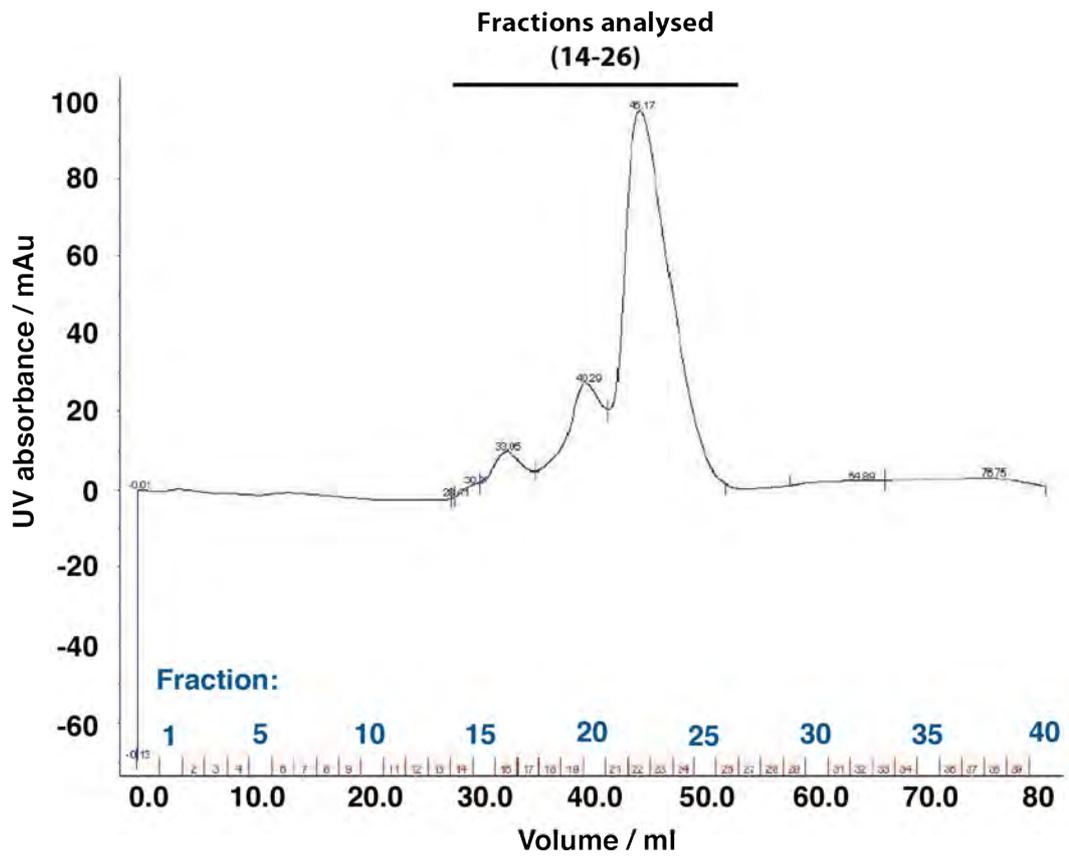
Figure A8: Chromatogram for size-exclusion purification of Cas12a following heparin-affinity purification.
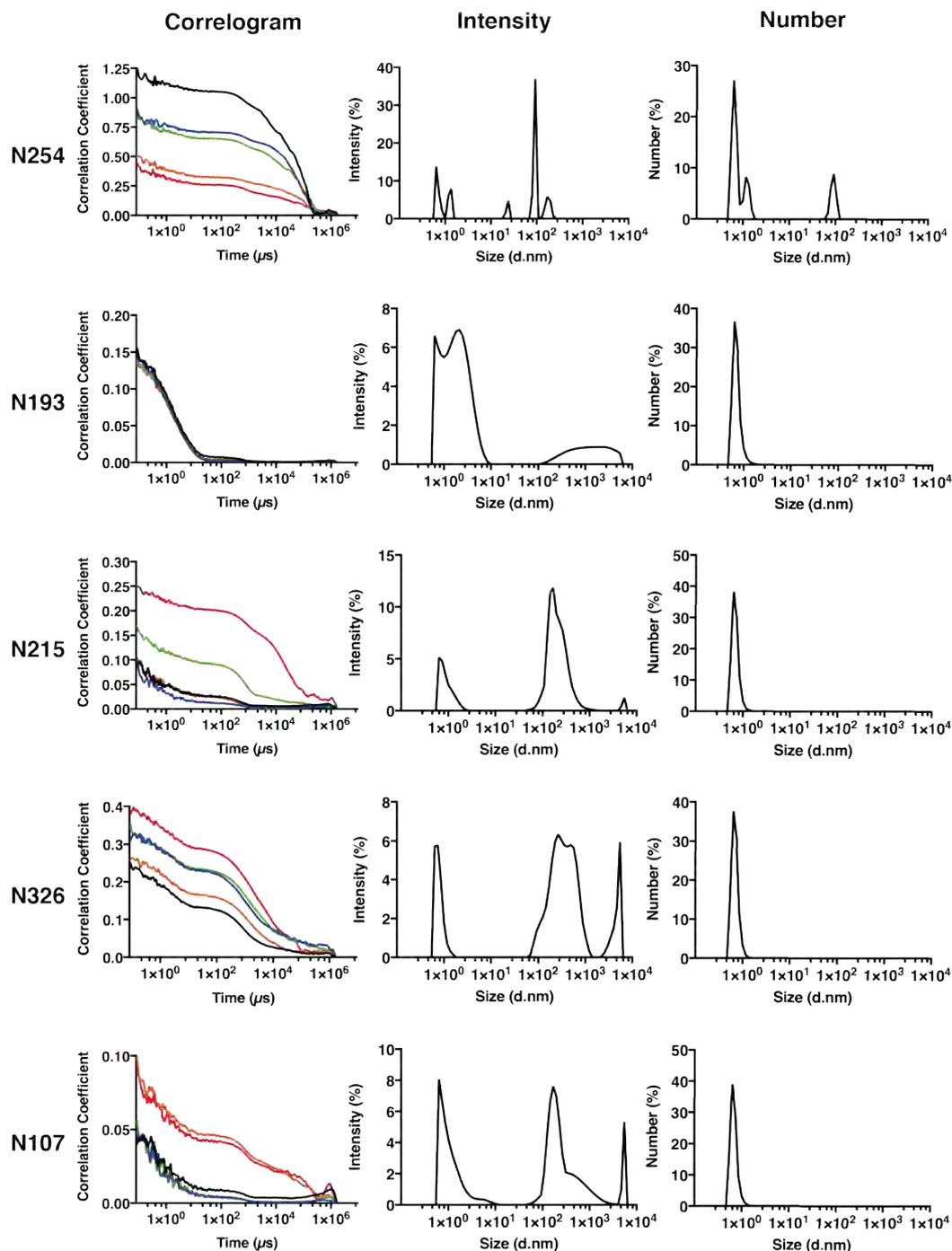
Figure A9: Complete DLS data for inhibitors 1–6 Shown are correlograms, intensity profiles and number intensity plots. Number plots are derived from the initial intensity signal.
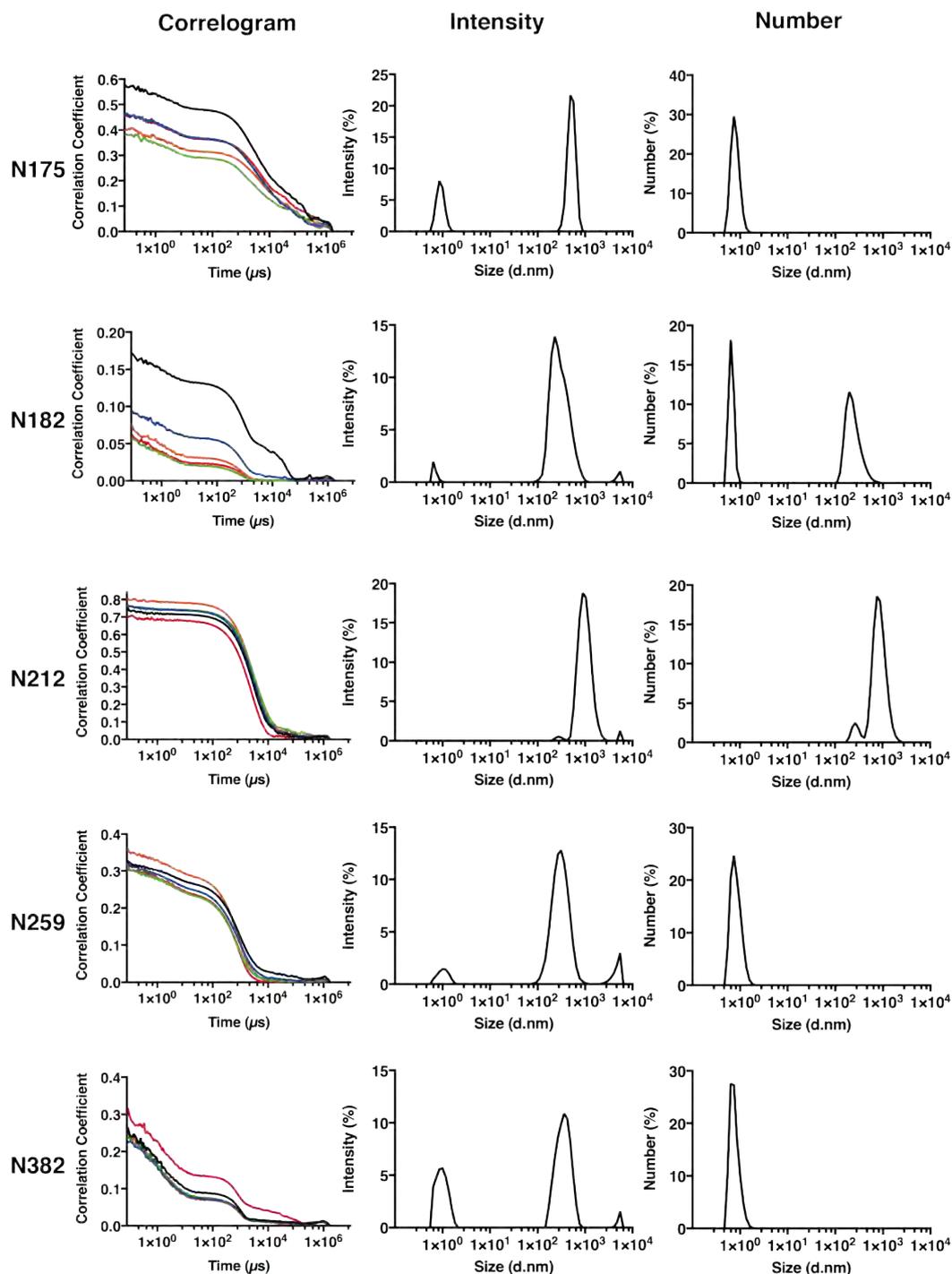
Figure A10: Complete DLS data for inhibitors 7–12 Shown are correlograms, intensity profiles and number intensity plots. Number plots are derived from the initial intensity signal.
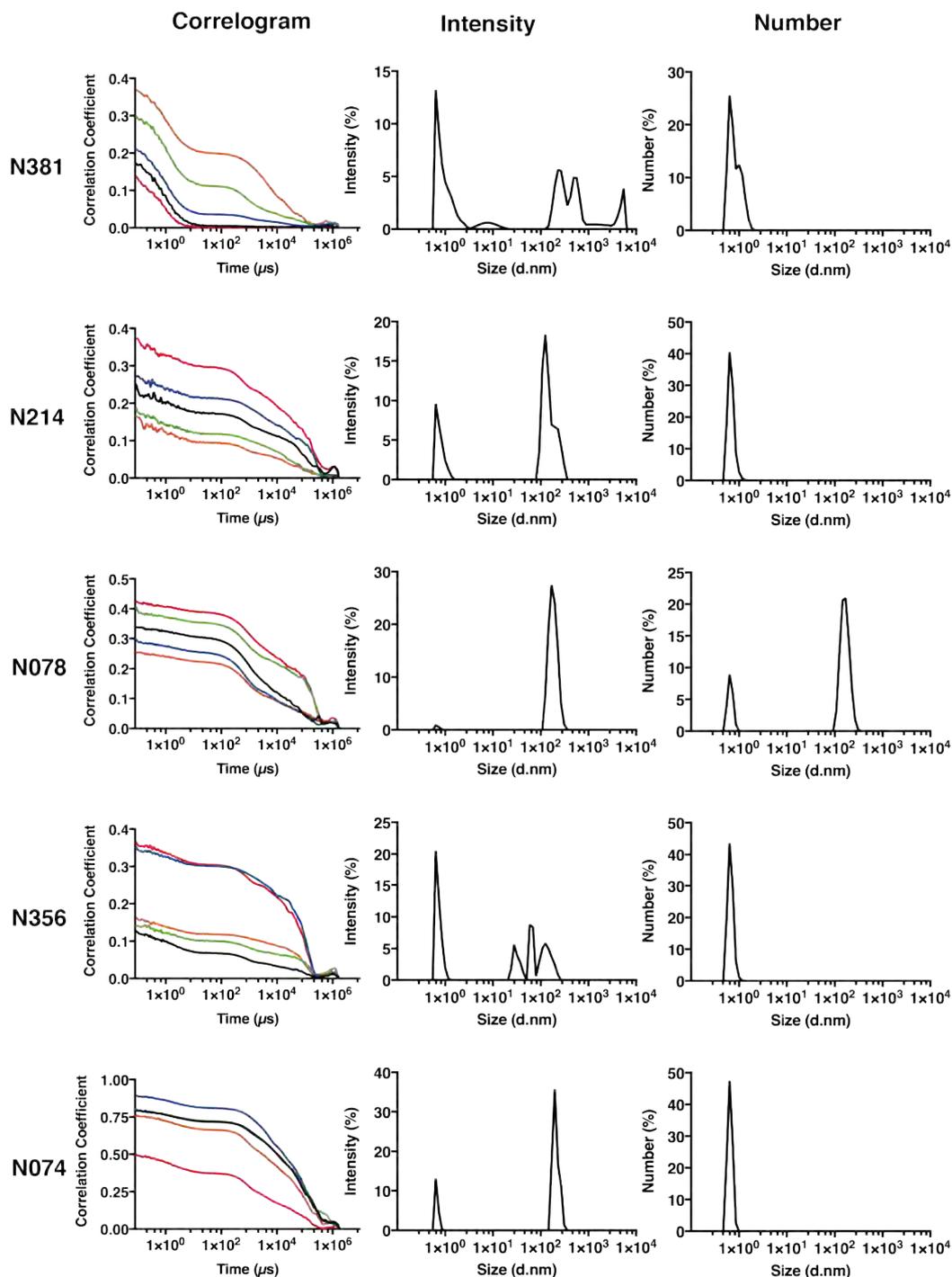
Figure A11: Complete DLS data for inhibitors 13–17 Shown are correlograms, intensity profiles and number intensity plots. Number plots are derived from the initial intensity signal.
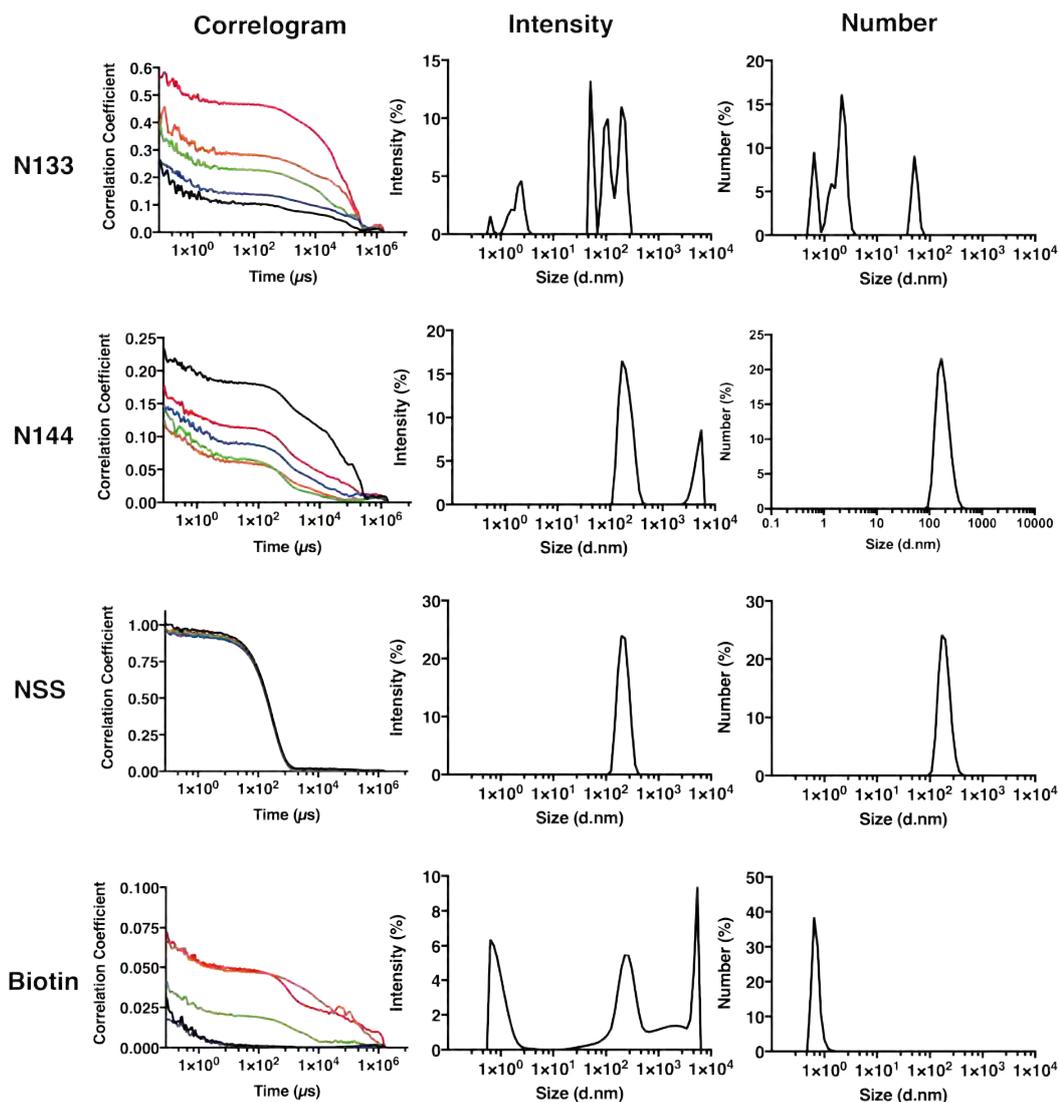
Figure A12: Complete DLS data for inhibitors 18–19 plus controls Shown are correlograms, intensity profiles and number intensity plots. Number plots are derived from the initial intensity signal. Nanosphere size standard (200 nm) is represented by acronym NSS.

Table A1: Particle diameter and dispersity data obtained by DLS

Sample name	Intensity averaged diameter (nm)	SD	Polydispersity index	SD
N254	3422	3939	1.459	0.4591
N193	1.912	0.3734	0.2344	0.04422
N215	580.3	1083	1.154	0.478
N326	622.1	239.2	0.8258	0.2066
N107	255	344.5	0.677	0.2197
N175	1501	230.3	1.082	0.1131
N182	266.4	171.3	0.7875	0.5431
N212	1211	171.2	0.3705	0.1122
N259	296	47.66	0.4586	0.1092
N382	247.6	100.9	0.5097	0.07844
N381	429.9	616.6	0.5758	0.426
N214	1570	244	1.367	0.1319
N078	884.2	252.4	0.941	0.1046
N356	2728	1678	1.761	0.5784
N074	2718	1060	1.493	0.2898
N133	3806	3597	1.388	0.3421

Sample name	Intensity averaged diameter (nm)	SD	Polydispersity index	SD
N144	770.6	653.9	0.9728	0.2125
Nanosphere size standard	209.3	1.799	0.04482	0.05907
Biotin	557.1	746.1	0.871	0.5334

Sample name	Intensity averaged diameter (nm)	SD	Polydispersity index	SD

A.2 Permissions

Several figures were adapted or reproduced for use in this thesis, with the permission of the original authors.

Figure 1.5 taken from Cubbon et al.[89] was published under open access with a a CC BY licence and as such no formal permissions are required. However, the relevant article is cited appropriately and authors credited for the re-use.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Dec 01, 2021

This Agreement between Mr. Andrew Cubbon ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5098870038272
License date	Jun 30, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	HELQ promotes RAD51 paralogue-dependent repair to avert germ cell loss and tumorigenesis
Licensed Content Author	Carrie A. Adelman et al
Licensed Content Date	Sep 4, 2013
Type of Use	Thesis/Dissertation
Requestor type	non-commercial (non-profit)
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no

Will you be translating?	no

Circulation/distribution	1 - 29

Author of this Springer Nature content	no

Title	Doctoral Thesis: Investigating a human DNA repair helicase that supports CRISPR editing

Institution name	University of Nottingham

Expected presentation date	Aug 2021

Portions	Use of graphic from Fig.3b

Requestor Location	Mr. Andrew Cubbon 32 Hassocks Close Beeston Nottingham, Nottinghamshire NG9 2GH United Kingdom Attn: Mr. Andrew Cubbon

Total	0.00 GBP

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of

another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. *BOOKS ONLY:*Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Dec 01, 2021

This Agreement between Mr. Andrew Cubbon ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5067240196757
License date	May 13, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Microbiology
Licensed Content Title	Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants
Licensed Content Author	Kira S. Makarova et al
Licensed Content Date	Dec 19, 2019
Type of Use	Thesis/Dissertation
Requestor type	non-commercial (non-profit)
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no

Will you be translating?	no

Circulation/distribution	1 - 29

Author of this Springer Nature content	no

Title	Doctoral Thesis: Investigating a human DNA repair helicase that supports CRISPR editing

Institution name	University of Nottingham

Expected presentation date	Jun 2021

Portions	Figure from Box 1

Requestor Location	Mr. Andrew Cubbon 32 Hassocks Close Beeston Nottingham, Nottinghamshire NG9 2GH United Kingdom Attn: Mr. Andrew Cubbon

Total	0.00 GBP

Terms and Conditions

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any

other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. *BOOKS ONLY:*Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication papers:**
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

Questions? <u>customercare@copyright.com</u> or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Dec 01, 2021

This Agreement between Mr. Andrew Cubbon ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5200281061449
License date	Dec 01, 2021
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Genetics
Licensed Content Title	Synthetic lethality and cancer
Licensed Content Author	Nigel J. O'Neil et al
Licensed Content Date	Jun 26, 2017
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no

Will you be translating?	no
Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	Doctoral Thesis: Investigating a human DNA repair helicase that supports CRISPR editing
Institution name	University of Nottingham
Expected presentation date	Jan 2022
Portions	Use of figure 1 illustrating synthetic lethality in schematic form.
Requestor Location	Mr. Andrew Cubbon 32 Hassocks Close Beeston Nottingham, Nottinghamshire NG9 2GH United Kingdom Attn: Mr. Andrew Cubbon
Total	0.00 GBP

Terms and Conditions

Springer Nature Customer Service Centre GmbH
Terms and Conditions

This agreement sets out the terms and conditions of the licence (the **Licence**) between you and **Springer Nature Customer Service Centre GmbH** (the **Licensor**). By clicking 'accept' and completing the transaction for the material (**Licensed Material**), you also confirm your acceptance of these terms and conditions.

1. Grant of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, world-wide licence to reproduce the Licensed Material for the purpose specified in your order only. Licences are granted for the specific use requested in the order and for no other use, subject to the conditions below.

1. 2. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. However, you should ensure that the material you are requesting is original to the Licensor and does not carry the copyright of another entity (as credited in the published version).

1. 3. If the credit line on any part of the material you have requested indicates that it was reprinted or adapted with permission from another source, then you should also seek permission from that source to reuse the material.

2. Scope of Licence

2. 1. You may only use the Licensed Content in the manner and to the extent permitted by these Ts&Cs and any applicable laws.

2. 2. A separate licence may be required for any additional use of the Licensed Material, e.g. where a licence has been purchased for print only use, separate permission must be obtained for electronic re-use. Similarly, a licence is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the licence. Any content owned by third parties are expressly excluded from the licence.

2. 3. Similarly, rights for additional components such as custom editions and derivatives require additional permission and may be subject to an additional fee. Please apply to Journalpermissions@springernature.com/bookpermissions@springernature.com for these rights.

2. 4. Where permission has been granted **free of charge** for material in print, permission may also be granted for any electronic version of that work, provided that the material is incidental to your work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version.

2. 5. An alternative scope of licence may apply to signatories of the STM Permissions Guidelines, as amended from time to time.

3. Duration of Licence

3. 1. A licence for is valid from the date of purchase ('Licence Date') at the end of the relevant period in the below table:

Scope of Licence	Duration of Licence
Post on a website	12 months
Presentations	12 months
Books and journals	Lifetime of the edition in the language purchased

4. Acknowledgement

4. 1. The Licensor's permission must be acknowledged next to the Licenced Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract, and must be hyperlinked to the journal/book's homepage. Our required acknowledgement format is in the Appendix below.

5. Restrictions on use

5. 1. Use of the Licensed Material may be permitted for incidental promotional use and minor editing privileges e.g. minor adaptations of single figures, changes of format, colour and/or style where the adaptation is credited as set out in Appendix 1 below. Any

other changes including but not limited to, cropping, adapting, omitting material that affect the meaning, intention or moral rights of the author are strictly prohibited.

5. 2. You must not use any Licensed Material as part of any design or trademark.

5. 3. Licensed Material may be used in Open Access Publications (OAP) before publication by Springer Nature, but any Licensed Material must be removed from OAP sites prior to final publication.

6. Ownership of Rights

6. 1. Licensed Material remains the property of either Licensor or the relevant third party and any rights not explicitly granted herein are expressly reserved.

7. Warranty

IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL OR INDIRECT DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE MATERIALS REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION, DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND
WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION SHALL APPLY NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

8. Limitations

8. 1. *BOOKS ONLY:*Where **'reuse in a dissertation/thesis'** has been selected the following terms apply: Print rights of the final author's accepted manuscript (for clarity, NOT the published version) for up to 100 copies, electronic rights for use only on a personal website or institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/).

8. 2. For content reuse requests that qualify for permission under the STM Permissions Guidelines, which may be updated from time to time, the STM Permissions Guidelines supersede the terms and conditions contained in this licence.

9. Termination and Cancellation

9. 1. Licences will expire after the period shown in Clause 3 (above).

9. 2. Licensee reserves the right to terminate the Licence in the event that payment is not received in full or if there has been a breach of this agreement by you.

Appendix 1 — Acknowledgements:

For Journal Content:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For Advance Online Publication papers:
Reprinted by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].)

For Adaptations/Translations:
Adapted/Translated by permission from [**the Licensor**]: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

Note: For any republication from the British Journal of Cancer, the following credit line style applies:

Reprinted/adapted/translated by permission from [**the Licensor**]: on behalf of Cancer Research UK: : [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication)

For **Advance Online Publication** papers:
Reprinted by permission from The [**the Licensor**]: on behalf of Cancer Research UK: [**Journal Publisher** (e.g. Nature/Springer/Palgrave)] [**JOURNAL NAME**] [**REFERENCE CITATION** (Article name, Author(s) Name), [**COPYRIGHT**] (year of publication), advance online publication, day month year (doi: 10.1038/sj. [JOURNAL ACRONYM])

For Book content:
Reprinted/adapted by permission from [**the Licensor**]: [**Book Publisher** (e.g. Palgrave Macmillan, Springer etc) [**Book Title**] by [**Book author**(s)] [**COPYRIGHT**] (year of publication)

Other Conditions:

Version 1.3

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

A.3 PIP reflective statement

PIP Host Organisation: Nanna Therapeutics, Cambridge, UK

Throughout my undergraduate degree I found that insight into industry-led research was lacking despite the large part of the life sciences sector it occupies. In four years I think I only saw one or two lectures summarising experimental scaling or the basic pipeline of drug development. While applying for PhD positions, I considered this gap in my knowledge and experience and looked for ways to fill it. When I started on the BBSRC DTP, I was excited by the opportunity to gain industry experience through the PIP scheme. I reached out to Nanna Therapeutics, a drug-discovery company, focussing on mitochondrial diseases, oncology and the development of high-throughput screening platforms.

Upon arrival, the goals of my project were:

- Developing functional assays to be probe targets of interest using the company's high-throughput screening platform.

- To design Cas9 sgRNAs for desirable gene targets and to then use these in the generation of CRISPR-ko human cell lines.

- To shadow key employees in their development of a high-throughput screening system using techniques such as microfluidics and FACS.

By the end of my placement I had successfully developed an assay for target screening, including tolerance testing in a variety of conditions, using low-throughput methods. I had also designed sgRNAs and used them in combination with Cas9 for the generation of knockout cell lines, although utlimately this proved unsuccessful. The positive outcomes from this project led to the company exploring further collaboration with the University, including a future studentship and grant application for a new project.

This placement provided me with exactly the insight that I was looking for, as well as allowing me to develop my skills as a researcher. Throughout the PIP I was responsible for planning and troubleshooting my own project, allowing me to develop critical thinking and management skills in a different environment to my PhD project. I also further developed skills related to my PhD, such as human tissue culture and biochemical assay development. Perhaps one of the most important skills I developed was multi-disciplinary collaboration and communication as I often had to convey my progress and ideas to scientists in other fields, such as synthetic chemists. I also gained careers insight by observing elements of their hiring processes and learning from staff about what they considered to be a good or bad CV, interview, or candidate in general.

Overall, the PIP gave me a different experience of science as I found industry much more tightly focussed in both the scope of the research and also the timeframe for achieving objectives. The environment of a small biotechnology company seemed the ideal fit for the way that I enjoy working, with the tight focus meaning that research objectives were clearly defined, whilst still leaving some room for exploration. The trade-off however was that this exploration was extremely limited in scope.

While I currently intend to apply for post-doctoral research positions, the experience I gained during the PIP including discussions with staff about working in industry, has definitely informed my future career plans, opening up my search for jobs in the sector.

A.4   COVID-19 impact statement

Included at the suggestion of the University of Nottingham.

This is a statement regarding the COVID-19 pandemic and the impact that university closure had on the progress of this research. The form is provided with the intention to provide examiners with some context to issues incurred both throughout the university closure and also during the study period prior.

See overleaf.

COVID19 Impact Statement 2020
For use only by UKRI funded PGRs with funding end date between 1st March 2020 and 31st March 2021

The University of Nottingham aims to support all our PGRs to complete their degrees within their period of study, by meeting our Doctoral Outcomes. We recognise, and aim to take into account, personal circumstances that may affect a PGR's ability to achieve this.

This Impact Statement should be used to provide details and evidence of impact for:

- applications for an additional funded period of registered study;
- applications for a funded extension to Thesis Pending;
- the thesis examination.

It will also be used to determine both the case and length of a COVID extension to funding **up to a maximum of six months** (twelve for part time PGRs). **Please note that it is expected that most approved extensions will reflect the duration of enforced change in activity during lockdown, and that extensions of longer duration will be the exception rather than the rule.**

Please keep a completed copy of this form as you may want to refer to it as evidence of impact in your thesis examination. You will also need to use it copy/paste your responses for submission in the online version of the form which will be the format used to make your extension request. The online form will be made available during the week of 11 May.

Please carefully consider your case for any extension with reference to the University's online Policy on Circumstances Affecting Students' Ability to Study and Complete Assessments (under Exceptional Guidance to Extenuating Circumstances Panels) and section 16 of the PhD Regulations (see Appendix 2, section 1), relating to existing regulations on circumstances that may or may not be usual grounds for an extension, and the Exceptional Regulations for UKRI funded PGR extensions (which can be found on the same site as this form).

In accordance with UKRI guidance, **we strongly encourage you to discuss the completion of this form with your supervisors.** If you prefer, you can alternatively discuss the form with an appropriate member of PGR support staff such as your DTP/CDT Director or Manager, DTP/CDT Welfare Officer, School Postgraduate Student Advisor, School PGR Director or other member of the Welfare team, or the Researcher Academy Faculty Lead (formerly Associate Dean for the Graduate School).

To ensure that you cover the full impact of the COVID-19 pandemic on you and your research **since March 15th 2020**, please complete all relevant sections of the form. You can be very brief but please include all relevant information even in note or bullet form.

When applying for an extension to either your period of registered study (i.e. when active data collection is to be done) or to Thesis Pending, or both you should show, briefly how/whether your work to date already meets some of the University and QAA Doctoral Outcomes, and clarify which doctoral outcomes are not currently met and how your plan will enable you to meet these (Appendix 1).

Under the exceptional conditions of the COVID-19 pandemic, in addition to the usual circumstances that may be grounds for an extension, you can and should also consider, and evidence if asked to do so, the additional circumstances listed in Section 1. These include but are not limited to:

- your ability to work effectively now that you are not in your usual working environment;
- any change in access to research settings or facilities, such as archives, field-sites, laboratories, software, or databases;
- any changes in your personal circumstances or environment resulting from remote working, or national restrictions , including those related to:
  - caring responsibilities,
  - disability and/or being at higher risk from coronavirus

292

o impacts on your supervisory team that have affected your research progress
o your mental health, and whether you have access to mental health support if needed,

o any financial impacts, either personal or on the research in progress or planned.
o any other considerations that should be taken into account, whether these do or do not relate to any protected characteristics.

This form should capture the impact of the pandemic on you and your research progress, not solely any impact of the University closure itself.

For further information including addressing future impacts; privacy and confidentiality of information submitted, and additional notes and guidance please see Appendix 2.

The information collected in this form will be used for the purposes of assessing your case for a funded extension to your doctoral studies, to provide information to your funder; to inform the University of the range of impacts that our PGRs have experienced, and to inform policy decisions on how to support our PGRs in future. The document will also aid discussion and decision making, to ensure consistency in evaluation of the impacts for different people.

All information used for other than the stated purposes will be anonymised, and all personal information through which anyone could be identified removed. The information on this form will not be shared with anyone, including supervisory teams, for other than the stated purposes, without your permission.

Background Information – your details

Family Name:	Cubbon	First Name(s)	Andrew
ID:	4274584	School:	Life Sciences

If this form is competed for an extension request:
I am applying for:
☐ an additional period of registered study; ☐ extension to Thesis Pending; ☒ both

Please identify your relevant funder(s)	AHRC/ BBSRC/ ESRC/EPSRC/MRC/NERC/ STFC	Dates of impact: (the date from which the impact has had an effect).	06/03/20
Start date	28/09/16	Funding end date	30/09/20
Length of extension requested: (up to a maximum of twenty four (26) weeks)	24 weeks	Programme length (3, 3.5, 4 years) and full time or part time	4 years, full time

The primary areas of impact:

Please tick all that are relevant for the ways in which you have been affected by the COVID pandemic and the resulting effect(s) on you and/or your research progression. You can give more details on these impacts, if you wish, on the next page.

Note: We will ask you to explain whether and how you have been able to manage or reduce any of these impacts in Section 2, on p.5.

The ways in which you have been affected (choose all that apply)

☒ additional/new caring responsibilities
☐ specific impact resulting from remote working as a result of a disability*
☐ being at higher risk of coronavirus;
☒ personal financial impact;
☒ new illness, accident or hospitalisation, including any mental health problems
☒ lack of access to mental health support (if needed);
☐ death or illness of a partner/close relative*
☐ illness of a relative for whom you are a carer
☐ impacts related to any protected characteristics*
☐ an impact on your supervisory team that has affected your supervision or progress
☐ military or other service (e.g. NHS) that has not already been accommodated
☐ parental leave that has not already been accommodated
☐ redeployment to work in another area (e.g. COVID) where this has not already been   accommodated.
☐ other events not on this list that are specifically related to the COVID pandemic (please describe below)

The ways in which your research activity has been affected
(for each that applies, please also indicate whether you have tried to mitigate the effect in this area).

	Was any mitigation possible?
☒ Disruption of planned activities	Yes/No
☒ Access to facilities/archives/lab/equipment/field sites etc	Yes/No
☒ Postponement of critical activities where alternatives are not available	Yes/No
☐ Access to other research resources including financial impact	Yes/No
☒ Ability to achieve a planned outcome/ milestone/deliverable	Yes/No
☒ Access a research partner, including research-related placements	Yes/No
☒ Inability to devote your usual time to research activity	Yes/No
☐ *Lack of usual supervisory support for thesis progression/writing	Yes/No
☐ *Lack of usual supervisory support to help manage risk and mitigate plans	Yes/No
☐ Other (please describe below)	Yes/No

*We are collecting this information in order to fully understand how you have been affected. Any information that you give here will only be used as information to inform us and will not be shared with anyone other than the teams considering the cases for extension and collating information for submission to UKRI.

1. DESCRIBING THE IMPACT
(Please complete this section to provide us with more detailed information)

For example you could write a short clear description of the nature of the impacts or problems that you face/have faced, make making this description as brief, and specific as possible. You could also give more detail on the nature of the impacts on your research progress.

We understand that personal and research impacts will be related, so if it helps you could structure the content in line with the impacts you identified in the tick boxes above.

Error! Reference source not found. additional guidance

The impact **on you:**

My partner suffers from diagnosed depression (2015 onwards) and Chronic Fatigue Syndrome (Mar 2018 onwards), which require ongoing treatment. For a large portion of my studies I have had some caring duties, often amounting to several hours a day, as required during flare-ups where they are incapable of doing many things for themselves. My partner's CFS means they are also considered to be at higher risk from COVID-19, requiring more caution when going outside and requiring outdoors activities e.g. food shopping, to be done by me.

On 06/03/20 my partner sustained a workplace injury requiring hospital treatment including X-rays and an MRI scan. This has further increased my caring duties.

Throughout the PhD I have struggled with Generalised Anxiety Disorder which has been exacerbated by the stress of long-term caring, particularly in light of their recent accident. Juggling my research activities with this has meant that often work takes me longer to accomplish than expected, making the completion of goals / milestones more difficult.

I have previously sought the help of the university counselling service (from July 2017), which helped for a time and since have attempted to use the free self-help materials provided by the University. The uncertainty generated by the closure impacted my anxiety enough that I sought mental health advice outside of the university but was unable to receive any due to the effects of the lockdown. Overall this has impacted the amount of work I have been able to achieve during the closure period.

Further stress and anxiety have arisen from the loss of income from undergraduate teaching and also from my partner being furloughed (Apr 14th 2020), not only reducing pay but also preventing the receipt of sales bonuses which supplemented their pay.

The impact **on your research:**

The closure period has meant no access to lab space in which I can complete the experiments required to finish the original work for my PhD. This includes the disruption of planned activities, which had been carefully laid out during meetings with my supervisor. This included the inability to image cells fixed to microscope slides; an experiment crucial to the completion of one of my results chapters and which takes several weeks to set up and complete. The closure has also meant the postponement of critical activities, specifically the generation of gene knockout cell lines; a process which takes 2-3 months to complete and which is crucial to another of my results chapters. Taken together, this means I have been unable to complete a large portion of work required to bring my thesis up to a standard to satisfy peer review and merit publication.

The closure period and subsequent lockdown by the government have prevented access to a research partner who provides materials which are the basis of a third results chapter. These experiments have had to be delayed due to the closure and upon return to work the receipt of these materials may be delayed.

Throughout the closure period it has been difficult to devote the planned amount of time to study, writing, and research due to caring responsibilities, anxiety and the associated mental health costs that come with them. This has meant that delivering on milestones, such as thesis chapter completion and online bioinformatics-based research is harder and often takes longer to achieve.

2. ACTIONS TAKEN TO MINIMISE THE IMPACT

a) How have you tried to mitigate the risk to your project?

Please **briefly** explain how you are trying to minimise the impact of the situation on your research activities and progress. **If you have not tried to alter your plans ,it's particularly important to explain here why you have taken this decision.**

For example,

- have you discussed how to do this with your supervisors?
- have you considered different ways to get the research done, such as changing your research plans to alter the order in which you do different elements?
- have you altered your research design, for example to conduct research online, or using other digital resources?
  what constraints or barriers did you have to try to remove, modify or overcome?
- **If you have not tried to alter your plans at all, why not?**

Try to show how/whether your work to date already meets some of the University and QAA Doctoral Outcomes, clarify which doctoral outcomes are not currently met and how your plan will enable you to meet these.
up to 200 words

Section 2 additional guidance

Prior to the lockdown, my supervisor and I met to lay out plans in case of a university closure. We have also met regarding several data sets and their interpretation throughout the closure period.

As a final year student, the plan in case of closure was to compile and analyse existing data, create figures for my thesis and to write up this data alongside the methods and introduction chapters, aiming to minimise writing-up time at the end of the research period.

We have also reconsidered parts of the project and narrowed its scope, making the research less expansive but more realistic for completion. For example, testing only the two most promising candidates in a drug-development trial rather than the seven that were initially planned and testing them only on the two most interesting proteins rather than the original four. We have also devised a bioinformatic component to one chapter to maximise data gathering during this time, as this information can be acquired online.

Currently, the work fails to meet the doctoral outcomes points (i) and (ii) in that the results generated are neither substantial enough nor of a quality high enough to satisfy peer review and merit publication.

b) List the aspects of your research plan that you have managed to achieve or progress during the period of impact.

Completed during the period of impact:

- Materials and methods chapter
- One data chapter, minus a small number of figures which require additional, essential experimentation to create.

Achieved / progressed during the period of impact:

- Analysis of previously acquired data
- Creation of figures for the remaining two results chapters.
- Began bioinformatic analysis required for a chapter, which is ongoing.
- Extensively planned and began writing the introductory chapter.
- Ongoing planning of experiments, design of custom materials (e.g. oligonucleotides) and noting of other materials required for completion of research upon the end of closure period.

3. NEXT STEPS

Please list what you have planned to do, in order to continue to lessen the impact on your research once you are able to resume the specific activities listed in Section 1

For example, what plans do you have to make sure that elements of your research that you have been unable to undertake due to the University closure restart quickly, or to efficiently complete the work you started during the closure?

up to 200 words

Section 3 additional guidance

Once I am able to resume research activities there are several things, I plan to do to streamline the completion of work including:

- Immediately commencing human cell culture again as this takes time to get started.
- If given enough notice by the university, ordering required materials prior to returning to work.
- Experiments which are controls, replicates or similar to work previously carried out are already / are being planned so that the work can take place swiftly following re-opening.
- Several of the experiments can be run in tandem, thereby achieving multiple objectives in the research plan simultaneously, maximising efficiency.

4. EVIDENCE

List any evidence that you have to demonstrate the impact you have detailed in section 1.

Please do not provide the evidence with this form, we will request it from you if we need to see it.

Please also provide here:
- a brief bullet list of the doctoral work completed prior to COVID-19 impact
- a revised research plan that shows how the requested length of extension is justified by the work that remains to be done to enable you to meet the Doctoral Outcomes;
- only if available, a previous work plan for comparison
up to 200 words

Section 4 additional guidance

Evidence:
- Partners medical history,
- Seeking (NCS, 06/04/20) / attending counselling.
- Welfare meetings with Alice Haslam
- Research plans

Doctoral work completed:
- Preparative work for testing a cell-free system for studying genome editing and DNA repair.
- Anti-cancer small-molecule drug screening and development.
- Biochemical and cell-line analysis of HelQ (protein of interest) interaction with R-loops (DNA:RNA hybrid).

Revised plan:

Chapter: CRISPR-DNA Repair
- Activity and R-loop formation Cas12a (set up: 1wk, completion: 1wk)
- Cell-free system study (20wks):
    - Proof-of-concept integration assays
    - Test of Cas9,12a with own ssODN designs
    - Test of integration with nuclear extract using Cas9,12a with ssODNs
    - Test of Integration with HelQ KO cell-lines
    - Test KO cell extract complemented with HelQ

Chapter: R-loops
- pUC19 roadblock removal assays (set up: 2wks, completion: 4wks)
- HelQ-unwinding RNA:DNA hybrids (set up: 2wks, completion: 4wks)
- Confocal microscopy (set up: 2wks, completion: 4wks):
    - DNA:RNA hybrid
    - m6a presence
    - anti-HelQ and yH2aX

Chapter: Inhibitors
- C-HelQ:Inhibitor analysis (set up:2wks, completion: 4wks):
    - DNA binding
    - ATPase activity
    - IC50s (I6, I19)
    - RecQ IC50's (I10, I15)
- In Vivo tolerance of DMSO and inhibitors (set up: 2wks, completion 6 wks)

5. CONFIDENTIAL INFORMATION

Please use this section to provide any confidential information that you would like to be considered. Information given here will only be shared with the team assessing your case and returning the information to UKRI for their consideration of your application.

Since the beginning of the closure period, I have taken part in two welfare check up sessions with Alice Haslam, to address my anxiety. Due to the nature and severity of the anxiety it was noted that self-help materials may be insufficient for coping. It was suggested as a further

course of action that I should consider undertaking cognitive behavioural therapy to try and cope better with these issues. This is something I am now actively pursuing.

☒ I confirm that I have completed this form after/in discussion with:
(indicate all those that apply, discussion with only one person is required)

☒ Primary supervisor/other supervisor   ☐ SPSA   ☐ School PGR Director   ☐ DTP/CDT Director
☐ DTP/CDT Manager   ☐ DTP/CDT Welfare Officer   ☐ other member of the Welfare Team
☐ Researcher Academy Faculty Lead (RAFL, aka Associate Dean of the Graduate School)

RAFLs are: Prof A Grabowska (MHS), Dr L Bradnock (Arts), Prof R Graham (Science) and Dr N Porter (Eng)

☒ I confirm that the information provided in this form is true and request an extension for the reasons and purposes outlined above.

Additional impact notes from 11/08/20 onwards:

Frequent changes to working hours have made it very difficult to plan work or to know whether work requiring consecutive days / multiple weeks can even be carried out. From 10/08/20 to 23/08/20 the system was week on / off with the lab split in half. From 24/08/20 to 30/08/20 the system was operated in half days but was changed after one week to accommodate increased occupancy. This meant that from 31/08/20 to 11/10/20 the system was 4-4.5 days per week which was good for work. The system changed from 12/10/20 until 01/12/20 to operate largely unchanged at 4 days per week to accommodate two new arrivals. From 01/12/20 the system had to accommodate a new start in the lab group sharing our space, which reduced time to 3 days per week. The stress associated with constant changes to work time and the decreasing hours has made it impossible to actually achieve the work approved for the extension, largely due to failures on the part of UoN to address required changes in day-to-day working practices. They were too busy putting up signs and patting themselves on the back about it to actually make sensible change that would work for everyone.

Ordering delays also reduced work capacity. While this isn't to be unexpected given the situation, it did make activities such as assays and receiving sequencing results slower and much more difficult to achieve.

A.5 Statement regarding extension to registered period of study

Included at the suggestion of the University of Nottingham.

This is a statement regarding the extension to my registered period of study awarded from 31/03/21 to 31/08/21. This statement is provided with the intention of providing examiners with some context in two key areas: the continued impacts that the COVID-19 pandemic had on this work from August 2020 onwards, and perhaps more importantly factors outside of the lab that had a severe impact on my ability to conduct research both before and after the pandemic.

See overleaf.

**Request to Register a Research Student
for an Additional Period of Study**

Where it is agreed that a student's course should be extended into a fourth year of study (for students registered as PhD candidates) or a third year of study (for students registered as MPhil candidates and intending to submit for that degree). Once the form is complete e-mail it to studentservices@nottingham.ac.uk (alternatively you can take this to in person or post to one of the Student Service Centres).

PLEASE READ: The University's regulations stipulate the time students should spend studying as a registered student: **PhD** – full time course *minimum* 2 years and part-time course *minimum* 4 years. However, in general most full-time PhD students are registered for 3 years. The maximum period of study allowed is 4 years full-time and 8 years part-time. **MPhil** – full-time course *minimum* 1 year and part-time 2 years. Most full-time MPhil students are registered for 2 years. The maximum period of study allowed is 3 years full-time, 6 years part-time.
Section 1 – Student Information: Extra period of study requested (see above): 3 Months Surname: Cubbon   Student ID: 4274584 First Name: Andrew   School/Dept: School of Life Sciences Candidate for (please tick ✓):   PhD ☑   MPhil ☐
Section 2 - School/Student Declaration: I request permission to register for an additional period of study and this has the support of my supervisor(s). Student Signature: *A. Cubbon*   Date: Jan 11th 2020 Signatures of   _____   Date: Jan. 11th 2020 Supervisor(s): *E. c Bolt*   Date: _____ Name of principal supervisor (in capitals): ED BOLT Head of School signature: _____*James McClory*_____   Date: _____Jan 14th 2021_____

Section 3 – International Office Approval (overseas students only)

On behalf of the International Office, I can confirm that the student has been fully informed of the implications this additional period of registration will have on their immigration status in the UK. The student understands that the University is obliged to report this additional period of registration to the Home Office.

Are you currently sponsored by an officially recognised sponsor?   Yes ☐   No ☐

If yes, please give the name of your sponsor and file number: _____

If you are sponsored, this change may have implications to your funding. If you have sought permission from your sponsor, you must attach written evidence; otherwise the International Office will be obliged to inform your sponsor.

Does this course require ATAS clearance: Yes ☐   No ☐

In some cases new ATAS clearance may be required, even if you have applied before. Your advisor will inform you if this applies to you.

Signed: _____ (On behalf of the International Office) Date: _____

Office use only

Form complete	☐	Signed by HoS	☐
Saturn updated	☐	IO Advice (overseas only)	☐
Notified student	☐	SAM decision	☐
Notified other departments	☐	Approved by: _____	
Reported to Home Office	☐	Date: _____	
Signed by Supervisor	☐		

Andrew Cubbon - Supporting statement for an additional period of study:

Throughout the PhD I have suffered from generalised anxiety disorder and depression, twice seeking counselling in 2017 and 2020 to help me move forward. I have also had caring responsibilities, looking after my disabled partner who suffers from severe CFS. This care has often amounted to several hours per day during flareups, requiring me to cut my days short. I have often tried to make up for lost time by working late evenings and weekends, but this has not been sustainable and has exacerbated my own mental health issues. Since January 2021 my partner has also been undergoing oncology testing for potential cancer, contributing significantly more stress. Juggling these problems with work has taken an immense toll on my mental health and reduced the amount of time I could spend working on, researching, writing up or thinking about my project, resulting in the slow completion of project milestones.

Despite my supervisor and I streamlining the project to meet the March 2021 deadline afforded by my COVID extension, delays in resuming work were much more severe than we anticipated. This impacted both the rate of progress and affected my mental health and ability to work due to stress. As such the work remains insufficient to meet the doctoral outcomes set out by the University. The requested extension to 30/06/21 would enable completion of essential data gathering from the two most disrupted areas of the project, described below, and also time to write up as reflected in the provided thesis plan.

The first remaining topic of my project is drug development for a cancer-related protein. The March and November 2020 lockdowns, and resulting issues with resuming business and deliveries, severely delayed the production of compounds required to finish the chapter. While the compounds have been delivered (as of 19/02/21) and the remaining work begun immediately, there is not sufficient time to complete planned, essential experiments, process and write up the data.

The second essential element is a screen regarding gene-editing associated proteins. These experiments require the culture of human cells to generate biomass. Upon resuming work in August 2020, delays on the delivery of reagents and consumables and the return of liquid nitrogen dewars containing cell stocks from storage, as well as limited access to the safety cabinets required to carry out the work, severely impacted my ability to complete experiments. We aimed for this work to be completed by December 2020, but this quickly became impossible. Having remained in Nottingham to work over the Christmas break and with greater access to the required safety cabinets the work is

on track to be completed in April, but this means that there is insufficient time to generate, process and write up the essential data for thesis completion without the requested extension.

Since returning to the lab I have worked incredibly hard to be flexible in my approach to carrying out experiments and to minimise the impact of delays in other areas of the project, despite it being incredibly taxing for my mental health. Through this effort, and extensive work on dealing with my mental health issues, I have been able to keep the project moving forward. Now that I have all of the missing pieces required to finish the work available to me, the remaining limitation is the time to complete these experiments and to process and write up the data. The provided thesis plan reflects this as it lists only essential remaining experiments that can be completed within this timeframe, with time also accounted for in the analysis and writing up of the final results for the completion of the thesis.

Edward L. Bolt
Associate Professor of Biochemistry
School of Life Sciences
University of Nottingham
Queens Medical Centre
Nottingham NG7 2UH

Telephone: (0115) 8230194
e-mail: ed.bolt@nottingham.ac.uk

March 4th 2021

To whom it concerns – supporting statement for Andrew Cubbon, extension until June 30th for his BBSRC DTP deadline.

I wholeheartedly support Andy's request to be granted a deadline extension until end of June 2021. I can confirm all of the comments in Andy's statement, which I believe provide rock-solid justification for further support. I would like to add that Andy is an outstanding early career/PhD researcher, resourceful and diligent, who has had by a combination of circumstances an extremely difficult last 12 months. I think Andy has perhaps had a more gruelling time than many of his peers.

After discussions with Andy frequently over the last months I am in agreement with the revised plan going forward, as detailed in the GANNT chart. I know that Andy is enthusiastic to complete this, now that he has a clear run at it, and if it is at all possible he should now be given that clear opportunity to do so.

E. L. Bolt

(Edward L. Bolt)

+44 (0)115 74 86500
SS-PGR-QMC@nottingham.ac.uk

Mr Andrew Phillip Cubbon
32 Hassocks Close
Beeston
Nottingham
NG9 2GH
United Kingdom

25 March 2021

Dear Mr Cubbon

PhD Life Science

We are pleased to confirm that your request for an additional period of study has been approved. Your period of registered study now ends on 30th June 2021 with a latest thesis submission date of 30th June 2021.

At the start of the next session, you will need to reregister with the University and you should do this online through the Portal at http://my.nottingham.ac.uk (logging in with your existing IS username and password). If you have difficulty registering online, please contact the IT Helpline (email: itservicedesk@nottingham.ac.uk / tel: 0115 9516677).

Following completion of your period of directed study, you will then commence the thesis pending period. Please visit the following website for the submission pack which contains all of the forms and information you will need to complete the submission of your thesis:
http://www.nottingham.ac.uk/academicservices/currentstudents/examinations/researchstudent/index-research.aspx

For all international students and students from the EU/EEA/Switzerland who started a course on or after 1 January 2021, studying in the UK on student visa, this may have an impact on your immigration status. For more information on your specific situation please see: www.nottingham.ac.uk/go/statuschange-immigrationadvice. It is your responsibility to ensure that you understand what this means for your immigration status in the UK and that you act in accordance with immigration rules and regulations. If you have any questions please contact an Immigration Advisor by email: immigration-support@nottingham.ac.uk, by telephone at +44 (0)115 84 66125 or in person by attending the drop-in service or by making an appointment. Information on how to contact an Immigration Advisor is available at: https://www.nottingham.ac.uk/studywithus/international-applicants/visa-help/adviser.aspx.

If you are an international student (on a visa of any kind) and your course requires Academic Technology Approval Scheme (ATAS) clearance and the end date of your studies has been extended by more than 3 months, you are required to obtain a fresh ATAS certificate now to cover this additional time: https://www.academic-technology-approval.service.gov.uk/. The Visa and Immigration team will contact you to check you have applied and can answer any questions about this process.

If you think any of the information contained in this letter is incorrect please contact Student Services, using the contact details at the top of the letter, immediately.

Yours sincerely

Victoria Pooley
Student Services Senior Administrator

Student Services
QMC PGR Student Services
University of Nottingham
University Park
Nottingham
NG7 2RD

+44 (0)115 74 86500 **|** nottingham.ac.uk

Copy to: Life Sciences; Student-Fees; Visa and Immigration Team (International students only); File (Invoice)

Our ref. 14274584

+44 (0)115 74 86500
SS-PGR-QMC@nottingham.ac.uk

Mr Andrew Phillip Cubbon
32 Hassocks Close
Beeston
Nottingham
NG9 2GH
United Kingdom

30 June 2021

Dear Mr Cubbon,

PhD Life Science

We are pleased to inform you that we now have formal agreement for an extension of time until 31/08/2021 in which to submit your thesis for the degree of PhD Life Science.

In accordance with University regulations, students granted an extension of time are required to pay a fee of £160 per session. This will be raised within ten days and is payable through the "My Finance" tab on the Student Portal (http://my.nottingham.ac.uk/cp/home/loginf). If you do not know your username or password or cannot access the Portal please contact the Student IT Helpline on 0115 9516677. It would also be helpful if you could let us know when you have paid this.

For international students in the UK on student visas issued under Tier 4 this may have an impact on your immigration status. For more information on your specific situation please see: www.nottingham.ac.uk/go/statuschange-immigrationadvice. It is your responsibility to ensure that you understand what this means for your immigration status in the UK and that you act in accordance with Immigration Rules and regulations. If you have any questions please contact one of the Immigration Advisors at the University Park Central Student Service Centre (Cherry Tree Lodge) (+44 (0)115 846 6125/email: immigration-support@nottingham.ac.uk).

Please note that you will no longer be a registered student during your extension period as registration is only available during the one year thesis pending period.

A full submission pack which gives all the forms and information you will need to complete the submission of your thesis is available at: http://www.nottingham.ac.uk/academicservices/qualitymanual/researchdegreeprogrammes/procedures-for-assessment.aspx

If you think any of the information contained in this letter is incorrect please contact Student Services, using the contact details at the top of the letter, immediately.

Yours sincerely,

Paulina Hutchinson
Student Services

cc: Life Sciences; Visa and Immigration Team (International students only); Research Examination Administrator; Graduate School; File (Invoice)

A.6   Publications

See overleaf.

PORTLAND PRESS

Review Article

CRISPR-Cas immunity, DNA repair and genome stability

Andrew Cubbon[1], Ivana Ivancic-Bace[2] and Edward L. Bolt[1]

[1]School of Life Sciences, University of Nottingham, QMC Medical School, Nottingham, U.K.; [2]Department of Molecular Biology, Faculty of Science, University of Zagreb, Croatia

Correspondence: Edward L. Bolt (ed.bolt@nottingham.ac.uk)

OPEN ACCESS

Co-opting of CRISPR-Cas 'Interference' reactions for editing the genomes of eukaryotic and prokaryotic cells has highlighted crucial support roles for DNA repair systems that strive to maintain genome stability. As front-runners in genome editing that targets DNA, the class 2 CRISPR-Cas enzymes Cas9 and Cas12a rely on repair of DNA double-strand breaks (DDSBs) by host DNA repair enzymes, using mechanisms that vary in how well they are understood. Data are emerging about the identities of DNA repair enzymes that support genome editing in human cells. At the same time, it is becoming apparent that CRISPR-Cas systems functioning in their native environment, bacteria or archaea, also need DNA repair enzymes. In this short review, we survey how DNA repair and CRISPR-Cas systems are intertwined. We consider how understanding DNA repair and CRISPR-Cas interference reactions in nature might help improve the efficacy of genome editing procedures that utilise homologous or analogous systems in human and other cells.

Interplay of DNA repair and CRISPR-Cas immunity: the fundamentals
Overview

CRISPR-Cas is a naturally occurring adaptive immunity system in prokaryotes [1,2]. Operational efficiency of CRISPR-Cas enzymes is closely associated with active DNA repair and replication, in natural CRISPR-Cas systems to promote building of adaptive immunity, processes called 'Adaptation', and in biotechnology where genome-editing reactions that utilise 'Interference' reactions also trigger DNA repair and their associated reactions. Identities of DNA repair enzymes involved in supporting native CRISPR-Cas systems in bacteria are becoming clearer, but the molecular mechanisms are not known or are inferred from known DNA repair and genome stability functions. Understanding these mechanisms might aid development of strategies for interpolating CRISPR-Cas enzymes (e.g. Cas9 and Cas12a, the latter formerly known as Cpf1) into eukaryotic cells, including in humans. This rationale is based on conservation of fundamental principles, and some specific properties, of DNA repair in bacterial and eukaryotic cells. New information is also emerging on how DNA repair processes in human cells support genome editing, which deepens understanding of how DNA repair systems are triggered and function in human cells, which can help to protect against cancers and other aging syndromes.

CRISPR-Cas adaptive immunity and Cas9-based editing

Deliverance of CRISPR-Cas immunity in native systems is through 'Interference' reactions that feature nucleotide base pairing of CRISPR-encoded RNA (crRNA) with an 'invader' mobile genetic element (MGE, e.g. a phage, plasmid). This is catalysed by a ribonucleoprotein Interference complex, also called an effector complex (Figure 1), reviewed in [3]. The molecular events within interference complexes vary according to the class and subtype of CRISPR-Cas system [4], but they incapacitate the MGE by binding to it
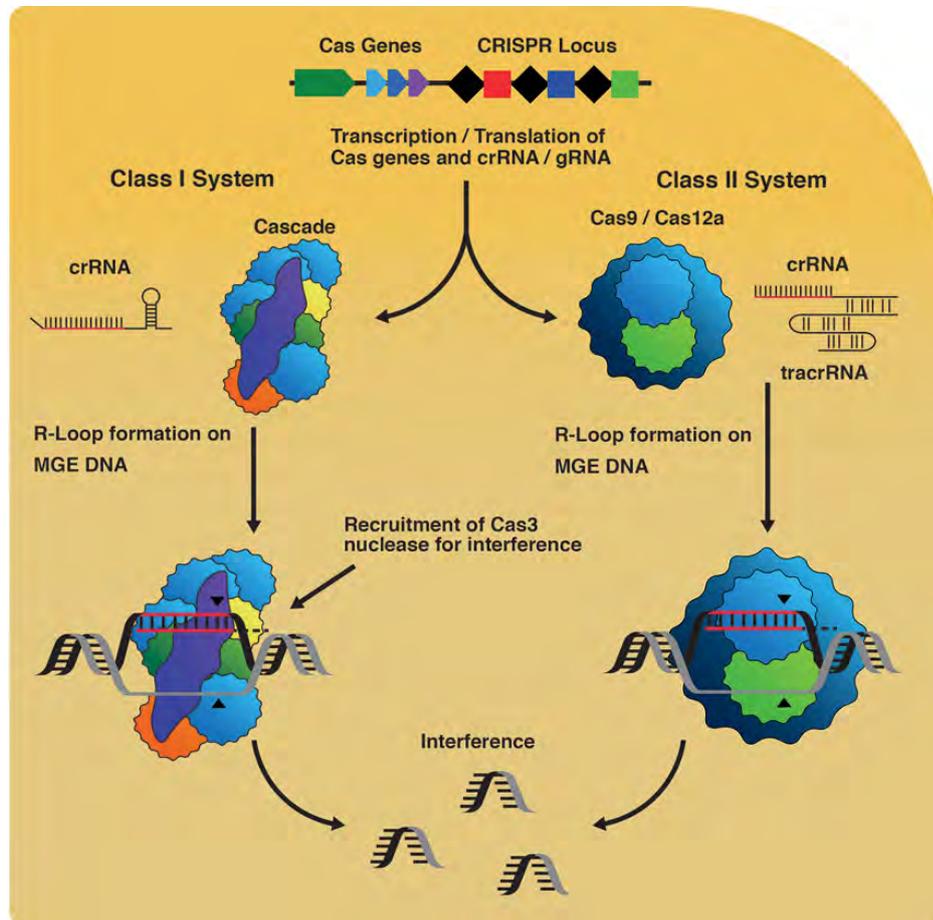
1

Figure 1. CRISPR-mediated interference reactions

DNA from MGEs provide small 'protospacer' fragments for acquisition into a CRISPR-locus during 'Adaptation' that generates immunity. Transcription and processing of CRISPRs create crRNAs, which are loaded into Interference complexes. Cas9 requires a second, long trans-activating crRNA (tracrRNA), which base pairs with crRNA to produce a mature sequence. These complexes catalyse R-loop formation on target DNA leading to nuclease activity targeted to the R-loop site, catalysed by Cas3 recruited to Cascade, or by Cas9.

stably, and triggering nucleolytic degradation of the MGE. Multi-subunit Cascade interference complexes of class I CRISPR systems [5], lack intrinsic nuclease activity but recruit the Cas3 nuclease-translocase enzyme to complete interference reactions [6,7]. Unlike Cascade, class 2 interference complexes have intrinsic DNA cutting activities. Co-incident DNA nicks generated by two nuclease active sites in Cas9 interference complex (RuvC-like and HNH-like) generates a DNA double-strand break (DDSB) [8,9]. Cas12a also generates DDSBs via two nuclease active sites and possesses potent ssDNA endo/exonuclease activity that has spawned further useful applications [10]. Structure, function and detailed mechanism of Cas9 and Cas12a are presented in a recent review [11].

Understanding molecular details of RNA-DNA base pairing in Cas9 interference reactions allowed for engineering of programmable single-guide RNAs to target DNA sequences of choice (sgRNA, Figure 1) [8], opening up the simplified DNA editing process that is now widely used for targeting individual genes. The effectiveness of using Cas9 for gene editing in cells is highlighted in landmark papers describing the first methods for editing genes in bacteria [12] and in human and mouse cells [13-15].

PORTLAND
PRESS

In native CRISPR-Cas systems, including the class 2 systems utilising Cas9, the crRNA payload that base pairs to DNA during interference is derived from transcription and processing of a CRISPR locus, in which each crRNA sequence is stored as a DNA 'spacer' (Figure 1). By engineering a CRISPR locus with multiple desired spacer sequences and transplanting into cells the engineered CRISPR with Cas9, and associated Cas proteins from the native system, it was possible to 'multiplex' Cas9 for targeting multiple genes as part of the same process [14].

RNA-DNA pairing by Cas9 forms the basis for genome editing, exploiting the molecular biology of native Interference reactions that target MGE DNA in R-loop nucleoprotein complexes [16,17] (Figure 1). The details of interference R-loop formation have been assessed in detail elsewhere [18,19]. There is currently a great deal of interest in how specificity for targeting of precise DNA sequences is achieved by Cas9 et al., and in off-site or genome instability effects of editing processes [20]. Both DDSBs and R-loops generated by Cas9, and other editing enzymes, have the potential to provoke genome instability by disrupting polymerases and helicases of DNA replication and transcription. Therefore CRISPR-Cas interference may trigger genome instability and cell death analogously to naturally occurring endogenous and exogenous genotoxins [20,21]. These lesions and blocks are detected globally or when linked to replication and transcription and dealt with by DNA repair systems that also impact on CRISPR-Cas interference reactions.

DNA repair at DNA breaks

DNA repair systems most relevant to this summary are illustrated in Figure 2. Repair of DDSBs by **Non-Homologous End Joining** (NHEJ) proteins does not require DNA sequence homology but instead ligates broken DNA ends together. NHEJ is not present in many bacterial clades, but is characterised in *Bacillus subtilis*, *Pseudomonas aeruginosa* and *Mycobacteria* [22,23]. The genetic requirements and biochemical mechanisms for NHEJ in bacteria differ from eukaryotes. In eukaryotes NHEJ predominantly occurs in G_1 phase of the cell cycle, reviewed in [24], and is inhibited during mitosis to prevent undesirable chromosome fusions at telomeres [25]. NHEJ is promoted by the Shieldin complex [26-29] and initiates from the Ku protein complex accessing exposed DNA ends. This serves as a scaffold for the recruitment of an assortment of lesion-specific accessory proteins including DNA-PKcs that stabilises broken DNA ends [30]. Artemis nuclease complex is recruited for DNA end processing, and DNA Ligase IV seals processed DNA ends to fix the break [31]. NHEJ is associated with insertion/deletion (In/Del) mutations several base-pairs long, which can induce frameshift in the coding regions of proteins, leading to their truncation and inactivity following translation [32].

DNA repair by **microhomology-mediated end joining** (MMEJ), also called Alternative End-Joining (A-EJ) [33], is also error-prone, and like NHEJ results in In/Del mutations, but MMEJ relies on 5–25 bp of sequence complementarity ('microhomology') between DNA strands by mechanisms that are still being worked out. A complex of Mre11, Rad50 and Nbs proteins (MRN complex) with C-terminal Binding Protein Interacting Protein (CtIP) is required for MMEJ to remove blockages at the DSB and to initiate end-resection, forming a 3′ overhanging flap [34]. DNA microhomology within the flap allows DNA annealing, with the rest of the flap either being resected or filled in by DNA synthesis activity. More detailed information about MMEJ can be gained from recent reviews [35,36], and both NHEJ and MMEJ repair Cas9 DDSBs in bacteria [37].

Homologous Recombination (HR), or homology-directed DNA repair, is a collection of processes in bacteria, archaea and eukaryotes for DDSB repair and to support DNA replication by reactivating replication forks at DNA nicks and barriers [38-41]. HR depends on availability of homologous DNA molecules that can base pair, and in some instances involves strand invasion catalysed by recombinases, RecA, RadA or Rad51 and their accessory and regulatory proteins [42,43]. Two major modes of recombinase-mediated HR operate in many, but not all, bacteria exemplified in *Escherichia coli* by RecBCD and RecFOR pathways [44,45]. RecBCD is a helicase-nuclease that targets dsDNA that is blunt-ended, or recessed by a few nucleotides, and converts them into 3′ ssDNA tailed molecules that are coated by RecA recombinase, through specific interaction between RecBCD and RecA in response to DNA sequences called Chi, reviewed in [45]. The RecFOR complex targets ssDNA gaps that may arise if RecQ helicase unwinds DNA at stalled replication forks or in other contexts such as G4 DNA [46]. RecFOR replaces single-strand DNA binding protein (SSB) with RecA provoking strand invasion into a homologous duplex that can lead to later stages of HR, including Holliday Junction formation by the RuvABC resolvasome.

In human cells, HR is predominant during DNA synthesis (S) and the second growth (G_2) phases of the cell cycle [47,48]. In HR, DDSBs are recognised by the MRN complex, which recruits CtIP initiating end resection of the DDSB in a 5′–3′ direction, leaving 3′ overhangs. RPA is recruited to 3′ tailed ssDNA that protects DNA and recruits ATR protein. RPA is exchanged with Rad51 recombinase through interaction with BRCA2 and/or RAD52, preparing the 3′ ssDNA tail DNA for strand invasion into homologous DNA template [49,50]. Strand invasion forms a D-loop ('Displacement-loop') intermediate, which can be resolved through multiple pathways of double-strand break repair
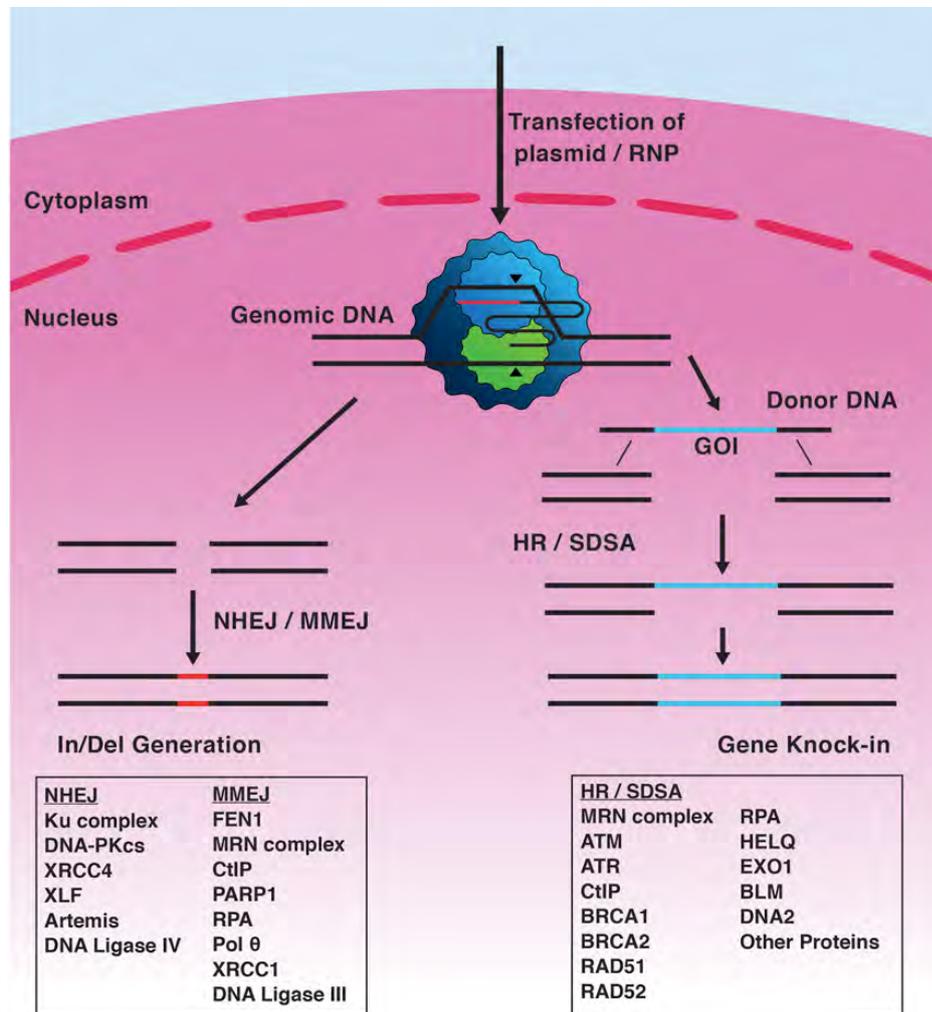
Figure 2. Overview of CRISPR-Cas9 triggered DNA repair

Cas9 targets and cuts DNA at R-loops. Editing procedures arising from the resulting DDSB depend on host DNA repair systems; NHEJ or microhomology-mediated end joining (MMEJ) pathways can lead to the generation of insertions or deletions in the DNA, generating frameshifts, preventing gene expression. Homologous recombination (HR) or synthesis-dependent strand annealing (SDSA) can be utilised for the precise knockin of genetic material, for example a 'gene of interest' (GOI). All of these processes require the co-ordination of a large suite of proteins, the interactions of which with Cas proteins is poorly documented.

(DSBR), synthesis-dependent strand annealing (SDSA) or break-induced replication (BIR). In these instances re-combinase catalysed strand invasion to form a D-loop is a pre-requisite. HR in the guise of single-stranded template repair (SSTR) is similar to SDSA and BIR but does not require a recombinase and therefore does not generate a D-loop [51]. Knowledge of SSTR in cells other than yeasts is very limited but it is a process that is potentially significant for HR-based genome editing that utilises ssDNA as a donor for insertion into a target site.

PORTLAND
PRESS

Functional interplay of natural CRISPR-Cas and DNA repair in bacteria

Little is known about interactions of DNA repair and Cas9 in *Streptococcus* species, the original source of Cas9, or with Cas12a in its host species of *Acidaminococcus, Francisella* and *Lachnospiraceae*. Most knowledge about how DNA repair and CRISPR-Cas systems interact physically and functionally is currently from the CRISPR-Cas system of *E. coli*, where they have been studied for their effects on CRISPR-Cas 'Adaptation', processes that lead to new DNA being integrated into a CRISPR locus by Cas1–Cas2 proteins thus generating or updating immunity. The helicase/translocase activities of RecBCD seem to be important for adaptation but the mechanism is not yet known [52-54]. Adaptation is stimulated in *E. coli* by R-loop interference complexes in 'primed' or 'targeted' adaptation [55,56], which seems to be a more general effect of interference on adaptation in bacteria [57]. Binding of Cas ribonucleoprotein complexes to target DNA forms R-loops that are sites positively identifying an MGE or other sequence to which the cell has acquired immunity. In this way cross-talk between interference complexes and adaptation reactions can stimulate immunity in response to incursion by an MGE. Genetic analyses in *E. coli* identified that loss of RecG or PriA helicase activities resulted in a loss of primed adaptation, but had no effect on naïve adaptation that occurs without interference R-loops [54]. This, and other genetic data, indicated that RecG and possibly PriA support adaptation through having an effect on R-loop interference complexes. *In vitro*, RecG protein dissociates Cascade interference R-loops that had blocked reconstituted DNA replication [58]. A model was proposed that RecG processes interference R-loops, as part of its intrinsic response to maintaining genome stability, and in doing so generates DNA substrates suitable for capture as new spacers for Adaptation [58]. R-loops trigger genome instability, therefore, mechanisms to dissolve R-loops are widespread across species [59,60]. The principle that RecG helicase can remove Cascade interference R-loops in *E. coli* may be of interest for potential effects of analogous helicases in eukaryotic cells because such enzymes might antagonise genome editing by targeting Cas9 R-loops for removal.

Interplay of CRISPR-Cas and DNA repair: genome editing

The CRISPR-Cas interference enzyme Cas9 has been used for a variety of gene editing applications in many species, including human cells as reviewed most recently in [61]. The CRISPR-Cas adaptation protein complex Cas1–Cas2 has been used in novel ways to create a CRISPR locus with DNA-based digital witness and recorder properties [62]. We herein focus on Cas9-based editing procedures, for which there is rapidly growing body of information about interplay with DNA repair, and which is likely to be relevant and extended to other genome editing enzymes, most notably Cas12a, as more is understood about their use. Genome editing using Cas9 relies on its natural enzymatic activities, forming an interference R-loop (modified as sgRNA) and generating DDSB, but also activities from Cas9 variants that nick only one DNA strand, lack any nuclease activity or are fused to other enzyme functionalities, the latter described more below.

NHEJ-based editing: DNA cut, disrupt or re-write

The error-prone nature of NHEJ has been widely used with Cas9-sgRNA for generating gene knockouts, since its inception in 2013 [12]. Two more recent applications of NHEJ, CRISPaint (CRISPR-assisted insertion tagging) and VIKING have been used to re-write DNA sequence information at Cas9-sgRNA targets and are readily available in kit-form. CRISPaint has been used to facilitate tagging of target proteins by editing the target gene with sequence encoding an in-frame 'tag' (e.g. luciferase or a coloured fluorescent protein) [63]. VIKING technology was developed from principles of NHEJ that were used in ZFN and Talen-based genome editing [64], for example the ObLiGaRe method [65]. This is modified in VIKING by use of Cas9-sgRNA to direct linearisation of DNA to a VKG1 sequence that is shared widely among plasmid vectors. This optimises binding of the Ku complex and the likelihood that donor DNA will be incorporated into the cut site on the human genome. However, DNA integrations off-target are a concern as is the production of In/Del mutations at DNA junctions surrounding the large inserts and also notes the potential for inserts to be inserted in the reverse conformation, resulting in a failure to express the large cassette which has been inserted [64]. MMEJ is also being exploited for gene editing, benefiting from its activity throughout the cell cycle. MMEJ-based PITCh (Precise Insertion into Target Chromosome) [66,67] has been used to insert custom DNA cassettes flanked by arms of microhomology into a Cas9-induced DDSB with higher efficiency than HR, and mHAX (microhomology assisted excision) [68] is used for scar-less removal of selectable markers from DNA insertions, for example removal of a puromycin cassette from *HPRT1* in human stem cells. Procedures based on MMEJ offer an alternative-editing route to NHEJ and HR, with potentially reduced error rate relative to NHEJ and higher rate of incidence compared with HR.

HR and Fanconi anaemia pathway based editing: DNA cut and replace

HR can be exploited for 'genetic replacement' by insertion of new DNA sequence at the Cas9 R-loop target site. A great deal of effort has been made to optimise HR-based genome editing because it has the ability to accurately swap undesired DNA sequence for a desired sequence, for example to achieve therapeutic editing within mutated cells. The low prevalence of HR throughout the eukaryotic cell cycle and difficulty preparing suitable DNA donor for successful insertion and resistance to cellular assault have stimulated research to optimise genome editing that is underpinned by HR. Strategies for optimisation include: (i) promoting HR over NHEJ in cells, (ii) determining the most suitable combination of genome editing tools for use by insertion into recipient cells, and (iii) determining host cell DNA repair enzymes that promote HR at editing sites, and which antagonise it.

(i) NHEJ can be suppressed using small molecule inhibitors or gene silencing of genes encoding NHEJ proteins, concomitantly promoting HR-based editing [69-71]. Use of NU7441 and KU-0060648 to inhibit DNA-PKcs in human HEK293 T/17 cells achieved this, with HR measured in a 'Traffic Light Reporter' assay as green fluorescence, and NHEJ as Red fluorescence [69]. The study observed a reduction in NHEJ events by 40% and a two-fold increase in successful HR. Use of Scr7, a small molecule inhibitor of the DNA ligase IV DNA binding domain, achieved several-fold increased HR-mediated insertions into various genetic loci in eukaryotic cells [72]. A potential drawback to NHEJ inhibitors however may be that although achieving the intended effect on genome editing it may also lead to problems for DNA repair elsewhere in the genome that requires NHEJ, leading to unintended genome instability away from the Cas9 target site.

(ii) The tools necessary for HR-based genome editing are Cas9 and one or more sgRNAs alongside a donor DNA molecule that contains desired sequence for insertion (Figure 2). One area for optimisation of editing is the composition of donor DNA, which can be linear or circular dsDNA, or a single-stranded donor oligonucleotide (ssODN). Various studies have shown profound differences in editing efficacy depending on donor DNA used, for example [73]. The DNA used comprises DNA arms of sequence homologous to the genome target site that flank the desired DNA sequence. This allows homologous DNA pairing to be initiated after Cas9 has generated a DDSB. Use of phosphothioate-modified oligonucleotides in donor DNA can improve the efficiency of gene modification by stabilising donor DNA against host cell degradation [74]. This is thought to stimulate HR because a higher concentration of template persists for prolonged availability for successful insertion. Similarly, in ssODN higher efficiencies of gene insertion can be achieved compared with dsDNA donor, and can be further improved by chemical modification of ssODN donor [74,75].

(iii) The enhanced effect of ssODN donor DNA on genome editing has placed it at the forefront of establishing how cell DNA repair systems recombine this donor into the chromosome. Genetic replacement using ssODNs is thought to rely on HR repair pathways SDSA and/or SSTR, and the latter is gaining significant new interest because it is many times more active in human cells than recombinase-dependent HR that relies on synapsis forming D-loops [76-79]. Fanconi anaemia (FA) pathway proteins are identified as crucial for genome editing via ssODNs independently of Rad51-mediated HR, and this may rely on localisation of FANCD2 to sites of Cas9-catalysed DDSBs [75]. In this model, FA proteins marshal DNA break repair away from NHEJ and towards SSTR. DNA repair enzymes downstream of FA proteins were also identified as being important for SSTR including Rad51 paralogues, CtIP and HelQ helicase [75]. Knockdowns of HelQ had a strong negative effect on the incidence of SSTR that may be related to its physical interactions with FANCD2 and Rad51 paralogues [75]. HelQ in human cells helps to maintain genome stability by repair of broken down DNA replication and it may act to limit HR from progressing into Holliday junctions [80,81]. The importance of HelQ for this type of genome editing may help to identify more precisely its cellular role.

Designer Cas9 proteins

Cas9-sgRNA 'off-the-peg' catalyses R-loop interference reactions triggering a DDSB at the site of the R-loop. This has also facilitated development of gene editing technologies based on modifying the protein architecture of wild-type Cas9, nuclease inactivated Cas9 (dCas9) and single-strand cutting 'nickase' Cas9 (nCas9) including transcriptional regulation and imaging, reviewed recently in [61]. Cas9 protein fusions have also been generated to enhance HR in human cells, by biasing DNA repair pathway choice at the site of the DDSB. Two examples have fused CtIP and RAD52 to Cas9 [82,83].

The CtIP fusion protein was explored using two different methods. An active Cas9-CtIP fusion was able to stimulate an increase in editing compared with standard HR, and a second Cas9-fusion enhanced HR further by fusing an N-terminal fragment of CtIP, deemed the HR-enhancer domain (HE), that is crucial to its initiation of HR [82]. The
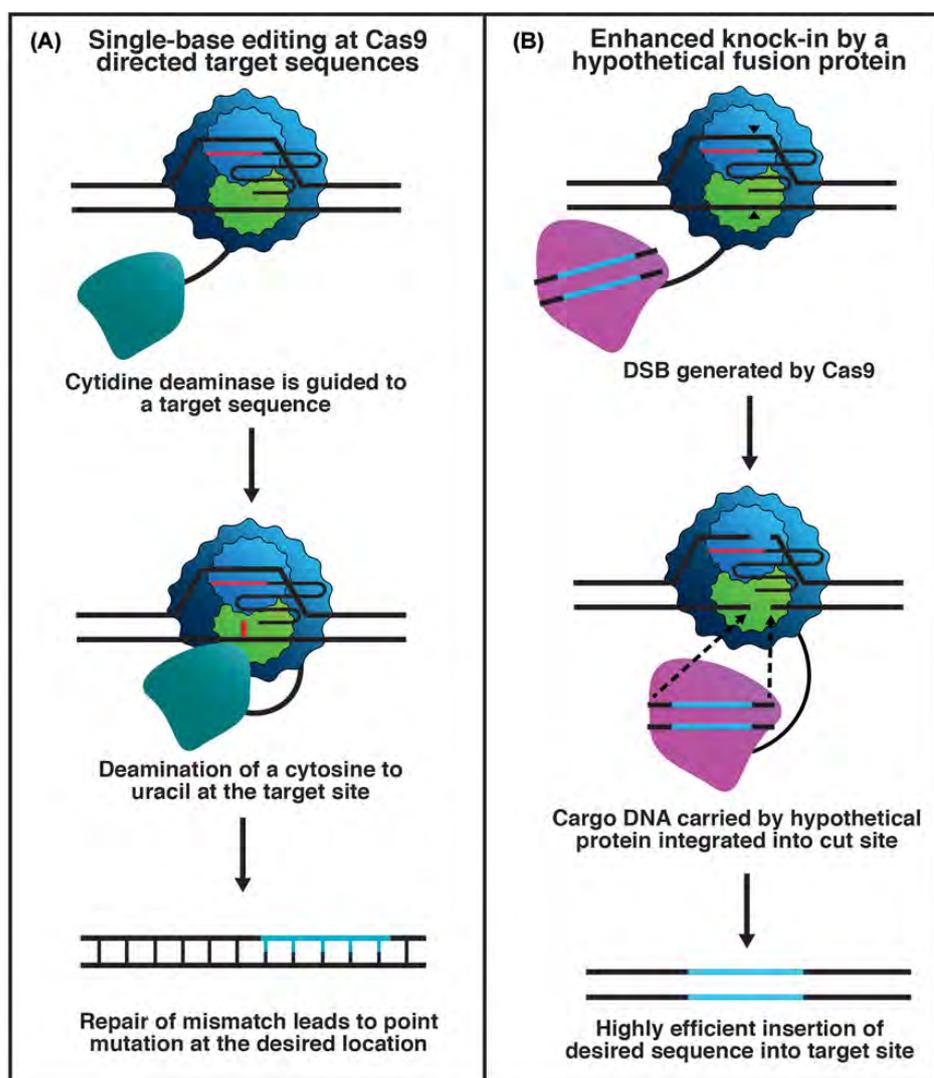
Figure 3. Schematic of genome editing triggered by Cas9-fusions

(**A**) A Cas9—CDA fusion protein precisely targets a cytosine base for conversion into uracil. Then, in cahoots with other factors, this can promote C to T transitions in Cas9—targeted DNA sequences. (**B**) Speculative model showing Cas9 fused to a protein for delivering duplex DNA into a site of Cas9 DDSB that may not need to rely on host cell HR processes.

RAD52 fusion protein was designed with the same rationale of forcing a protein crucial to the completion of HR close to the site of a Cas9 DDSB and was found to enhance the efficiency of reporter cassette insertion [83].

Fusions of dCas9 or nCas9 to DNA base modifying enzymes have facilitated editing of single bases. A cytidine deaminase (CDA)—dCas9 fusion has generated a C to T transition at R-loop targeted cytosine residues by generation of uracil, which is replaced with thymine during subsequent DNA repair (Figure 3) [84]. The R-loop generated window of ssDNA allows the deaminase to convert C into U. By fusing uracil glycosylase inhibitor (UGI) to the C-terminus of CDA—dCas9 base-excision repair was prevented, allowing mismatch repair (MMR) to complete the C to T change [84]. The system was enhanced by fusion of a second UGI, to further favour MMR, and the Gam

protein derived from bacteriophage, which binds the free ends of DDSBs, minimising In/Del generation. An adenine deaminase fused to dCas9 has been effective at targeted conversion of adenine into inosine, which is in turn converted into guanine [85]. A similar system, RNA Editing for Programmable A to I Replacement (REPAIR), has also been reported for the single-base editing of adenosine to guanine through an inosine intermediate in RNA transcripts in mammalian cells utilising catalytically inactive Cas13, a class 2 CRISPR-Cas RNA editing enzyme [86]. Finally, the reliance of HR-based CRISPR-Cas9 genome editing on host cell HR enzymes might make it attractive to develop a Cas9 fusion to proteins that are active as site-specific recombinases or have similar DNA integration activity (Figure 3). Cas9-mediated R-loop formation would in this scenario target DNA for integration of a duplex DNA payload carried by the fusion enzyme.

Concluding remarks

DNA repair was first implicated in cell survival in the 1930s [87], recombination as a form of DNA repair was modelled first in the 1970s [88], but CRISPR-Cas immunity was discovered recently, in 2005–2007. Their combined study is mutually beneficial for improving genome editing towards therapeutic advances in many organisms, but also for understanding human DNA repair processes, which when faulty lead to diseases associated with genome instability, and when activated are obstacles to cancer treatments through helping cancerous cells overcome chemotherapeutic agents.

Competing interests

The authors declare that there are no competing interests associated with the manuscript.

Abbreviations

BIR, break-induced replication; Cas, CRISPR-associated; CDA, cytidine deaminase; CRISPaint, CRISPR-assisted insertion tagging; CRISPR, clustered regularly interspaced short palindromic repeat; crRNA, CRISPR-encoded RNA; CtIP, C-terminal binding protein interacting protein; DDSB, DNA double-strand break; D-loop, displacement-loop; FA, Fanconi anaemia; HR, homologous recombination; In/Del, insertion/deletion; MGE, mobile genetic element; MMEJ, microhomology-mediated end joining; MMR, mismatch repair; MRN, complex of Mre11, Rad50 and Nbs protein; NHEJ, non-homologous end joining; RPA, replication protein A; SDSA, synthesis-dependent strand annealing; sgRNA, single-guide RNA; SSTR, single-stranded template repair; ssODN, single-stranded donor oligonucleotide; UGI, uracil glycosylase inhibitor.

References

1 Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P. et al. (2007) *Science* **315**, 1709–1712, https://doi.org/10.1126/science.1138140
2 Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S. and Charpentier, E. (2018) *Cell* **172**, 1239–1259, https://doi.org/10.1016/j.cell.2017.11.032
3 Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V. and van der Oost, J. (2016) *Science* **353**, aad5147, https://doi.org/10.1126/science.aad5147
4 Makarova, K.S., Wolf, Y.I., Alkhnbashi, O.S., Costa, F., Shah, S.A. et al. (2015) *Nat. Rev. Microbiol.* **13**, 722–736, https://doi.org/10.1038/nrmicro3569
5 Makarova, K.S., Zhang, F. and Koonin, E.V. (2017) *Cell* **168**, 946–946.e1, https://doi.org/10.1016/j.cell.2017.02.018
6 Sinkunas, T., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. (2011) *EMBO J.* **30**, 1335–1342, https://doi.org/10.1038/emboj.2011.41
7 Westra, E.R., van Erp, P.B., Kunne, T., Wong, S.P., Staals, R.H. et al. (2012) *Mol. Cell* **46**, 595–605, https://doi.org/10.1016/j.molcel.2012.03.018
8 Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) *Science* **337**, 816–821, https://doi.org/10.1126/science.1225829
9 Gasiunas, G., Barrangou, R., Horvath, P. and Siksnys, V. (2012) *Proc. Natl. Acad. Sci. U.S.A.* **109**, E2579–E2586, https://doi.org/10.1073/pnas.1208507109
10 Chen, J.S., Ma, E., Harrington, L.B., Da Costa, M., Tian, X. et al. (2018) *Science* **360**, 436–439, https://doi.org/10.1126/science.aar6245
11 Swarts, D.C. and Jinek, M. (2018) *Wiley Interdiscip. Rev. RNA* **9**, e1481, https://doi.org/10.1002/wrna.1481
12 Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L.A. (2013) *Nat. Biotechnol.* **31**, 233–239, https://doi.org/10.1038/nbt.2508
13 Jinek, M., East, A., Cheng, A., Lin, S., Ma, E. and Doudna, J. (2013) *Elife* **2**, e00471, https://doi.org/10.7554/eLife.00471

PORTLAND
PRESS

14 Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R. et al. (2013) *Science* **339**, 819–823, https://doi.org/10.1126/science.1231143

15 Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M. et al. (2013) *Science* **339**, 823–826, https://doi.org/10.1126/science.1232033

16 Jore, M.M., Lundgren, M., van Duijn, E., Bultema, J.B., Westra, E.R. et al. (2012) *Nat. Struct. Mol. Biol.* **18**, 529–536, https://doi.org/10.1038/nsmb.2019

17 Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E. et al. (2014) *Science* **343**, 1247997, https://doi.org/10.1126/science.1247997

18 Rutkauskas, M., Sinkunas, T., Songailiene, I., Tikhomirova, M.S., Siksnys, V. and Seidel, R. (2015) *Cell. Rep.* **10**, 1534–1543, https://doi.org/10.1016/j.celrep.2015.01.067

19 Xiao, Y., Luo, M., Hayes, R.P., Kim, J., Ng, S. et al. (2017) *Cell* **170**, 48–60.e11, https://doi.org/10.1016/j.cell.2017.06.012

20 Kosicki, M., Tomberg, K. and Bradley, A. (2018) *Nat. Biotechnol.* **36**, 765–771, https://doi.org/10.1038/nbt0918-899c

21 Cui, L. and Bikard, D. (2016) *Nucleic Acids Res.* **44**, 4243–4251, https://doi.org/10.1093/nar/gkw223

22 Weller, G.R., Kysela, B., Roy, R., Tonkin, L.M., Scanlan, E. et al. (2002) *Science* **297**, 1686–1689, https://doi.org/10.1126/science.1074584

23 Bowater, R. and Doherty, A.J. (2006) *PLoS Genet.* **2**, e8, https://doi.org/10.1371/journal.pgen.0020008

24 Chang, H.H.Y., Pannunzio, N.R., Adachi, N. and Lieber, M.R. (2017) *Nat. Rev. Mol. Cell Biol.* **18**, 495–506, https://doi.org/10.1038/nrm.2017.48

25 Orthwein, A., Fradet-Turcotte, A., Noordermeer, S.M., Canny, M.D., Brun, C.M. et al. (2014) *Science* **344**, 189–193, https://doi.org/10.1126/science.1248024

26 Gupta, R., Somyajit, K., Narita, T., Maskey, E., Stanlie, A. et al. (2018) *Cell* **173**, 972–988.e23, https://doi.org/10.1016/j.cell.2018.03.050

27 Dev, H., Chiang, T.W., Lescale, C., de Krijger, I., Martin, A.G. et al. (2018) *Nat. Cell Biol.* **20**, 954–965, https://doi.org/10.1038/s41556-018-0140-1

28 Ghezraoui, H., Oliveira, C., Becker, J.R., Bilham, K., Moralli, D. et al. (2018) *Nature* **560**, 122–127, https://doi.org/10.1038/s41586-018-0362-1

29 Noordermeer, S.M., Adam, S., Setiaputra, D., Barazas, M., Pettitt, S.J. et al. (2018) *Nature* **560**, 117–121, https://doi.org/10.1038/s41586-018-0340-7

30 Uematsu, N., Weterings, E., Yano, K., Morotomi-Yano, K., Jakob, B. et al. (2007) *J. Cell Biol.* **177**, 219–229, https://doi.org/10.1083/jcb.200608077

31 Wilson, T.E., Grawunder, U. and Lieber, M.R. (1997) *Nature* **388**, 495–498, https://doi.org/10.1038/41365

32 Heidenreich, E., Novotny, R., Kneidinger, B., Holzmann, V. and Wintersberger, U. (2003) *EMBO J.* **22**, 2274–2283, https://doi.org/10.1093/emboj/cdg203

33 Sallmyr, A. and Tomkinson, A.E. (2018) *J. Biol. Chem.* **293**, 10536–10546, https://doi.org/10.1074/jbc.TM117.000375

34 Truong, L.N., Li, Y., Shi, L.Z., Hwang, P.Y., He, J. et al. (2013) *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7720–7725, https://doi.org/10.1073/pnas.1213431110

35 Sfeir, A. and Symington, L.S. (2015) *Trends Biochem. Sci.* **40**, 701–714, https://doi.org/10.1016/j.tibs.2015.08.006

36 Seol, J.H., Shim, E.Y. and Lee, S.E. (2018) *Mutat. Res.* **809**, 81–87, https://doi.org/10.1016/j.mrfmmm.2017.07.002

37 Jasin, M. and Haber, J.E. (2016) *DNA Repair (Amst.)* **44**, 6–16, https://doi.org/10.1016/j.dnarep.2016.05.001

38 Wright, W.D., Shah, S.S. and Heyer, W.D. (2018) *J. Biol. Chem.* **293**, 10524–10535, https://doi.org/10.1074/jbc.TM118.000372

39 Klein, H.L. and Kreuzer, K.N. (2002) *Mol. Cell* **9**, 471–480, https://doi.org/10.1016/S1097-2765(02)00493-8

40 Verma, P. and Greenberg, R.A. (2016) *Genes Dev.* **30**, 1138–1154, https://doi.org/10.1101/gad.280545.116

41 Ceccaldi, R., Rondinelli, B. and D'Andrea, A.D. (2016) *Trends Cell Biol.* **26**, 52–64, https://doi.org/10.1016/j.tcb.2015.07.009

42 Lin, Z., Kong, H., Nei, M. and Ma, H. (2006) *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10328–10333, https://doi.org/10.1073/pnas.0604232103

43 Cox, M.M. (2007) *Crit. Rev. Biochem. Mol. Biol.* **42**, 41–63, https://doi.org/10.1080/10409230701260258

44 Morimatsu, K. and Kowalczykowski, S.C. (2003) *Mol. Cell* **11**, 1337–1347, https://doi.org/10.1016/S1097-2765(03)00188-6

45 Dillingham, M.S. and Kowalczykowski, S.C. (2008) *Microbiol. Mol. Biol. Rev.* **72**, 642–671, https://doi.org/10.1128/MMBR.00020-08

46 Morimatsu, K., Wu, Y. and Kowalczykowski, S.C. (2012) *J. Biol. Chem.* **287**, 35621–35630, https://doi.org/10.1074/jbc.M112.397034

47 Branzei, D. and Foiani, M. (2005) *Curr. Opin. Cell Biol.* **17**, 568–575, https://doi.org/10.1016/j.ceb.2005.09.003

48 Branzei, D. and Foiani, M. (2008) *Nat. Rev. Mol. Cell Biol.* **9**, 297–308, https://doi.org/10.1038/nrm2351

49 Jasin, M. and Rothstein, R. (2013) *Cold Spring Harb. Perspect. Biol.* **5**, a012740, https://doi.org/10.1101/cshperspect.a012740

50 Iyama, T. and Wilson, III, D.M. (2013) *DNA Repair (Amst.)* **12**, 620–636, https://doi.org/10.1016/j.dnarep.2013.04.015

51 Gallagher, D.N. and Haber, J.E. (2018) *ACS Chem. Biol.* **13**, 397–405, https://doi.org/10.1021/acschembio.7b00760

52 Radovcic, M., Killelea, T., Savitskaya, E., Wettstein, L., Bolt, E.L. and Ivancic-Bace, I. (2018) *Nucleic Acids Res.*, https://doi.org/10.1093/nar/gky799

53 Levy, A., Goren, M.G., Yosef, I., Auster, O., Manor, M. et al. (2015) *Nature* **520**, 505–510, https://doi.org/10.1038/nature14302

54 Ivancic-Bace, I., Cass, S.D., Wearne, S.J. and Bolt, E.L. (2015) *Nucleic Acids Res.* **43**, 10821–10830, https://doi.org/10.1093/nar/gkv1213

55 Sternberg, S.H., Richter, H., Charpentier, E. and Qimron, U. (2016) *Mol. Cell* **61**, 797–808, https://doi.org/10.1016/j.molcel.2016.01.030

56 Datsenko, K.A., Pougach, K., Tikhonov, A., Wanner, B.L., Severinov, K. and Semenova, E. (2012) *Nat. Commun.* **3**, 945, https://doi.org/10.1038/ncomms1937

57 Jackson, S.A., McKenzie, R.E., Fagerlund, R.D., Kieper, S.N., Fineran, P.C. and Brouns, S.J. (2017) *Science* **356**, eaal5056, https://doi.org/10.1126/science.aal5056

58 Killelea, T., Hawkins, M., Howard, J.L., McGlynn, P. and Bolt, E.L. (2018) *RNA Biol.*, https://doi.org/10.1080/15476286.2018.1496773

59 Li, X. and Manley, J.L. (2006) *Genes Dev.* **20**, 1838–1847, https://doi.org/10.1101/gad.1438306

60 Aguilera, A. and Garcia-Muse, T. (2012) *Mol. Cell* **46**, 115–124, https://doi.org/10.1016/j.molcel.2012.04.009

61 Knott, G.J. and Doudna, J.A. (2018) *Science* **361**, 866–869, https://doi.org/10.1126/science.aat5011

62 Shipman, S.L., Nivala, J., Macklis, J.D. and Church, G.M. (2016) *Science* **353**, aaf1175, https://doi.org/10.1126/science.aaf1175

63 Schmid-Burgk, J.L., Honing, K., Ebert, T.S. and Hornung, V. (2016) *Nat. Commun.* **7**, 12338, https://doi.org/10.1038/ncomms12338

64 Sawatsubashi, S., Joko, Y., Fukumoto, S., Matsumoto, T. and Sugano, S.S. (2018) *Sci. Rep.* **8**, 593, https://doi.org/10.1038/s41598-017-18911-9

65 Maresca, M., Lin, V.G., Guo, N. and Yang, Y. (2013) *Genome Res.* **23**, 539–546, https://doi.org/10.1101/gr.145441.112

66 Nakamae, K., Nishimura, Y., Takenaga, M., Nakade, S., Sakamoto, N. et al. (2017) *Bioengineered* **8**, 302–308,
    https://doi.org/10.1080/21655979.2017.1313645
67 Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.T. and Yamamoto, T. (2016) *Nat. Protoc.* **11**, 118–133, https://doi.org/10.1038/nprot.2015.140
68 Kim, S.I., Matsumoto, T., Kagawa, H., Nakamura, M., Hirohata, R. et al. (2018) *Nat. Commun.* **9**, 939, https://doi.org/10.1038/s41467-018-03044-y
69 Robert, F., Barbeau, M., Ethier, S., Dostie, J. and Pelletier, J. (2015) *Genome Med.* **7**, 93, https://doi.org/10.1186/s13073-015-0215-6
70 Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S. et al. (2015) *Nat. Biotechnol.* **33**, 543–548, https://doi.org/10.1038/nbt.3198
71 Li, G., Zhang, X., Zhong, C., Mo, J., Quan, R. et al. (2017) *Sci. Rep.* **7**, 8943, https://doi.org/10.1038/s41598-017-09306-x
72 Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R. and Ploegh, H.L. (2015) *Nat. Biotechnol.* **33**, 538–542,
    https://doi.org/10.1038/nbt.3190
73 Song, F. and Stieger, K. (2017) *Mol Ther Nucleic Acids* **7**, 53–60, https://doi.org/10.1016/j.omtn.2017.02.006
74 Renaud, J.B., Boix, C., Charpentier, M., De Cian, A., Cochennec, J. et al. (2016) *Cell Rep.* **14**, 2263–2272,
    https://doi.org/10.1016/j.celrep.2016.02.018
75 Richardson, C.D., Kazane, K.R., Feng, S.J., Zelin, E., Bray, N.L. et al. (2018) *Nat. Genet.* **50**, 1132–1139, https://doi.org/10.1038/s41588-018-0174-0
76 Chen, F., Pruett-Miller, S.M., Huang, Y., Gjoka, M., Duda, K. et al. (2011) *Nat. Methods* **8**, 753–755, https://doi.org/10.1038/nmeth.1653
77 Richardson, C.D., Ray, G.J., Bray, N.L. and Corn, J.E. (2016) *Nat. Commun.* **7**, 12463, https://doi.org/10.1038/ncomms12463
78 Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L. and Corn, J.E. (2016) *Nat. Biotechnol.* **34**, 339–344, https://doi.org/10.1038/nbt.3481
79 DeWitt, M.A., Magis, W., Bray, N.L., Wang, T., Berman, J.R. et al. (2016) *Sci. Transl. Med.* **8**, 360ra134, https://doi.org/10.1126/scitranslmed.aaf9336
80 Adelman, C.A., Lolo, R.L., Birkbak, N.J., Murina, O., Matsuzaki, K. et al. (2013) *Nature* **502**, 381–384, https://doi.org/10.1038/nature12565
81 Takata, K., Reh, S., Tomida, J., Person, M.D. and Wood, R.D. (2013) *Nat. Commun.* **4**, 2338, https://doi.org/10.1038/ncomms3338
82 Charpentier, M., Khedher, A.H.Y., Menoret, S., Brion, A., Lamribet, K. et al. (2018) *Nat. Commun.* **9**, 1133,
    https://doi.org/10.1038/s41467-018-03475-7
83 Shao, S., Ren, C., Liu, Z., Bai, Y., Chen, Z. et al. (2017) *Int. J. Biochem. Cell Biol.* **92**, 43–52, https://doi.org/10.1016/j.biocel.2017.09.012
84 Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. (2016) *Nature* **533**, 420–424, https://doi.org/10.1038/nature17946
85 Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H. et al. (2017) *Nature* **551**, 464–471, https://doi.org/10.1038/nature24644
86 Cox, D.B.T., Gootenberg, J.S., Abudayyeh, O.O., Franklin, B., Kellner, M.J. et al. (2017) *Science* **358**, 1019–1027,
    https://doi.org/10.1126/science.aaq0180
87 Friedberg, E.C. (2008) *Cell Res.* **18**, 3–7, https://doi.org/10.1038/cr.2007.113
88 Resnick, M.A. (1976) *J. Theor. Biol.* **59**, 97–106, https://doi.org/10.1016/S0022-5193(76)80025-2