# Insights into the Ubiquitylation System and DNA Damage Repair

by

Rabih Abou Farraj

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry

University of Alberta

# Abstract

DNA serves as a detailed blueprint for cellular functions, but continuous exposure to damaging agents poses a threat to genomic stability, especially through DNA double-strand breaks. Cells utilize two main repair pathways, non-homologous end joining (NHEJ) and homologous recombination (HR), to mitigate this risk. Ubiquitylation, a widespread cellular process, is vital for DNA damage repair, involving the attachment of ubiquitin to proteins. This mechanism relies on E2 conjugating enzymes and E3 ligases, particularly RING E3 ligases, which target substrates and activate E2 enzymes for ubiquitin transfer. The E2 ubiquitin conjugating enzyme Ubc13, along with its binding partner Mms2, construct ubiquitin chains with distinctive lysine 63 (K63) linkage. These chains are responsible for recruiting downstream DNA repair proteins. Another notable RING E3 ligase, RNF138, plays a crucial role in ubiquitylating target proteins to promote HR. Here we investigate Ubc13 and RNF138. Ubc13 has been implicated in several different cancer pathways. A new structure of Ubc13 revealed a novel conformation of its active site pointing towards a dynamic active site loop. We suggest that distinctive structural characteristics of the active site loop of Ubc13 may serve as a foundation for the rational development and design of targeted Ubc13 inhibitors. We characterized two compounds, one which showed covalent reactivity with Ubc13's active site, but also inhibited a mutant of Ubc13 which resembles other E2 conjugating enzymes, suggesting it is a non-specific inhibitor. The second compound, a small fragment, is believed to non-covalently bind to Ubc13's active site, selectively inhibiting the wildtype form of the enzyme. We suggest that this small molecule fragment represents a promising starting point for developing next-generation inhibitors. Our studies continued characterizing RNF138, a RING E3 ligase, also implicated in disease pathways. We first characterized the DNA binding of RNF138, which showed strong preference for long single stranded 3' and 5' DNA overhangs. AlphaFold2 structural modeling unveiled an overall flexible architecture, with N-terminal domains, including the RING domain and the first zinc finger, crucially involved in E2-ubiquitin coordination. Finally, we characterized the regulation of RNF138. We showed RNF138 is ubiquitylated at K158 and phosphorylated by CDK2, and ATM kinase at residues T27, and S124 respectively. Interfering with the three post-translational modifications significantly diminishes RNF138's capacity to facilitate HR. Overall, our study provides significant insights into the mechanism underlying the essential roles of the E2 enzyme Ubc13 and the RING E3 ligase RNF138.

# Preface

The studies in this thesis are my original works, with collaborations and guidance primarily by Dr. Mark Glover (Department of Biochemistry, University of Alberta), Dr. Leo Spyracopoulos (Department of Biochemistry, University of Alberta), and Dr. Ismail Ismail (Department of Oncology, University of Alberta). The appended studies were collaborations with Dr. Basil Hubbard (Department of Pharmacology and Toxicology, University of Toronto), Dr. John Ussher (Faculty of Pharmacy & Pharmaceutical Sciences, University of Alberta), and Dr. Marek Michalak (Department of Biochemistry, University of Alberta). Chapters 2 and 3 of our work are currently in preparation for submission, and chapter 4 has been published.

Chapter 2 of this thesis is currently in preparation for submission as "Abou Farraj, R., Fan, Y., Cholak, S., Edwards, R., Hodge, C., Spyracopoulos, L, Glover J.N.M. Characterizing the Active Site of Ubc13". The manuscript was written and edited by Dr. Mark Glover, and myself. All the experiments were performed by myself, with the help of Yue Fan, and Spencer Cholak in protein purification and assay setup. Ross Edwards assisted in x-ray diffraction collection, and refinement. Dr. Curtis Hodge provided all the DNA constructs for protein purification and provided insights into the experiments. Dr. Leo Spyracopoulos provided molecular dynamic resources, conceptual insights, and critical feedback. Dr. Glover conceived and acquired funding for and supervised the project.

Chapter 3 of this thesis is currently in preparation for submission as "Abou Farraj, R., Cholak, S., Locke. A., Spyracopoulos, L, Ismail I.H., Glover J.N.M. Mechanistic Insights into DNA Damage Sensing RING E3 Ligase RNF138". The manuscript was written and edited by Dr. Mark Glover and myself. I performed all the experiments, with the help of Spencer Cholak in protein purification, and EMSA experiments. I assisted Dr. Andrew Locke in establishing the mammalian expression system and mammalian protein purification. Dr. Leo Spyrocopolous provided the DNA constructs of RNF138, and oligonucleotides used in the initial experiments. Dr. Ismail provided crucial insights and supervised Dr. Andrew Locke. Dr. Glover conceived and acquired funding for and supervised the project.

Chapter 4 of this thesis has been published as "Locke, A. J., Abou Farraj, R., Tran, C., Zeinali, E., Mashayekhi, F., Ali, J. Y. H., Glover, J. N. M., and Ismail, I. H. (2024) The role of RNF138 in DNA end resection is regulated by ubiquitylation and CDK phosphorylation. *J. Biol. Chem. 300. 105709.*" The experiments were designed by Dr. Ismail and Dr. Andrew Locke. Dr. Locke and myself preformed the formal analysis and visualization. Under the supervision of Dr. Locke I conducted immunofluorescence, clonogenic survival, and immunoblotting assays. Caroline Tran and I also assisted with maintaining cell cultures. I performed the AlphaFold modeling and sequence alignments and under the supervision Dr. Glover, offered extensive insights into RNF138 expression and purification. Caroline and Dr. Locke conducted the immunoprecipitation, cycloheximide chase, and His pull-down assays. Jana Ali carried out the laser microirradiation experiments, while Dr. Ismail and Dr. Locke performed the homologous recombination reporter assay. Dr. Locke and Dr. Ismail wrote the original draft, and Dr. Glover and I contributed to writing and editing the final draft. Dr. Glover and Dr. Ismail provided funding, and Dr. Ismail conceptualized the project.

Appendix A of this thesis has been published as "Chen, J., Abou Farraj, R., Limonta, D., Amir, S., Dakhili, T., Kerek, E. M., Bhattacharya, A., Reformat, F. M., Mabrouk, O. M., Brigant, B., Pfeifer, T. A., Mcdermott, M. T., Ussher, J. R., Hobman, T. C., Glover, J. N. M., and Hubbard, B. P. (2023) Reversible and irreversible inhibitors of coronavirus Nsp15 endoribonuclease. *J. Biol.*

*Chem. 299*, 105341." Jerry Chen and Dr. Basil Hubbard conceptualized the project. I crystallized, collected, and solved the structure of mutant Nsp15. I also preformed mass spectrometry, florescence polarization, DLS, and purified and provided WT and mutant Nsp15 for further inhibitor assays. I contributed to writing and editing the original draft and finalized the manuscript under Dr. Glover's supervision. Dr. Hubbard provided funding.

Appendix B of this thesis has been published as "Amirhossein, S., Dakhili, T., Greenwell, A. A., Yang, K., Abou Farraj, R., Saed, C. T., Gopal, K., Chan, J. S. F., Chahade, J. J., Eaton, F., Lee, C., Vel, C. A., Crawford, P. A., Glover, J. N. M., Batran, R. Al, and Ussher, J. R. (2023) The Antipsychotic Dopamine 2 Receptor Antagonist Diphenylbutylpiperidines Improve Glycemia in Experimental Obesity by Inhibiting Succinyl-CoA : 3-Ketoacid CoA Transferase. *Diabetes 72*, 126–134." I purified, crystallized, and solved the structure of SCOT. I also provided pure SCOT protein for further experimental characterization, and conducted crystallization trials with promising drug molecules, under the supervision of Dr. Glover. Dr. Amirhossein and Dr. Ussher conceived and provided funding for the project.

Appendix C of this thesis is a review and has been published as "Paskevicius, T., Abou Farraj, R., Michalak, M., and Agellon, L. B. (2023) Calnexin, More Than Just a Molecular Chaperone. *Cells 12*, 403." I preformed AlphaFold modelling and made figures for the text. Dr. Paskevicius, Dr. Michalak, and Dr. Agellon conceptualized the review. I edited and contributed to writing the final draft of the paper.

# Acknowledgements

I would like to extend my heartfelt gratitude to my supervisor, Dr. Mark Glover, for providing the mentorship and freedom crucial to my development as a scientist. Mark created an environment that I always looked forward to being a part of. There is no doubt in my mind that the mentorship and stimulating scientific atmosphere he cultivated were invaluable gifts that laid the foundation for my career. I will always remember and appreciate the support and opportunities he provided, which have been instrumental in shaping my professional journey. I'll miss the strong coffees made every morning that kept me energized throughout the week (one on Monday was all it took).

I would also like to thank my supervisory committee members, Dr. Leo Spyracopoulos and Dr. Ismail Ismail. I am eternally grateful to both for taking me under their wings and directly mentoring me. Leo was instrumental in guiding me on molecular dynamic simulations, often spending afternoons debugging scripts with me. He also provided invaluable advice on all the projects in this thesis, frequently offering fresh perspectives. Ismail accepted me into his lab during my final year and directly mentored me on cellular biological techniques, allowing me to broaden my skill set and contribute to exciting projects. Our conversations often went beyond science, and I can confidently say Ismail played a crucial role in my professional career development.

To all my lab mates and friends, thank you for celebrating the good days with me and for helping make the bad days better. A special thanks to Dr. Ross Edwards for his great patience in hands-on training, from protein purification to X-ray crystallography and SAXS techniques.

I would also like to thank my family: Mom, Dad, Rami, Rayan, and David. Since we immigrated to Canada, we have collectively sacrificed and tirelessly worked to ensure we all met our goals. This achievement is a collective success that would not have been possible without each of you playing a crucial role in my personal development.

To my wife, Madison: I met you a month into my PhD, and you have been there from the start, through late nights and weekends of work in the lab. You were always so optimistic and encouraged me through stressful days. Your continuous support has directly contributed to my success, and I am forever grateful for you. I love you.

I would also like to thank the funding agencies Canadian Institutes of Health Research (CIHR) and Natural Sciences and Engineering Research (NSERC). Additionally, I received funding opportunities through the Faculty of Graduate Studies and Research at the University of Alberta: the Novartis Pharmaceuticals Canada Inc. Graduate Scholarship, the Queen Elizabeth II Graduate Scholarship, the Alberta Graduate Excellence Scholarship, and the David Lawson Award.

# Table of Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| DNA | Deoxyribonucleic acid |
| ATP | Adenosine triphosphate |
| DSB | Double strand break |
| DDR | DNA Damage response |
| IR | Ionizing Radiation |
| UV | Ultraviolet |
| ROS | Reactive oxygen species |
| NHEJ | Non-homologous end joining |
| HR | Homologous recombination |
| NER | Nuclear excision repair |
| BER | Base excision repair |
| MMR | Mismatch Repair |
| FA | Fanconi Anemia |
| TLS | Translesion DNA synthesis |
| A-NHEJ | Alternative Non-homologous end joining |
| CPT | Camptothecin |
| Ku | KU70/KU80 |
| MRN | Mre11-RAD50-Nbs1 |
| CtIP | CtBP interacting protein |
| PTM | Post translational modification |
| ATM | Ataxia telangiectasia mutated |
| ATR | Ataxia telangiectasia and Rad3 related |

| | |
|---|---|
| DNA-PKcs | DNA-dependent protein kinase, catalytic subunit |
| BRCA1 | Breast cancer type 1 susceptibility protein |
| BRCA2 | Breast cancer type 2 susceptibility protein |
| CDK1 | Cyclin dependent kinase 1 |
| RPA | Replication protein A |
| ssDNA | Single stranded DNA |
| dsDNA | Double stranded DNA |
| E1 | E1 activating enzyme |
| E2 | E2 conjugating enzyme |
| E3 | E3 ubiquitin ligase |
| HECT | Homology to E6AP C terminus |
| Ub | Ubiquitin |
| polyUb | Polyubiquitin |
| UBD | Ubiquitin binding domain |
| UEV | Ubiquitin enzyme variant |
| UIM | Ubiquitin-interacting motif |
| RING | Really interesting new gene |
| RNF | Ring finger protein |
| Ubc | Ubiquitin conjugating enzyme |
| NF-κB | Nuclear factor kappa-light-chain enhancer of activated B cells |
| siRNA | Small interfering ribonucleic acid |
| ZNF | Zinc finger |
| FHA | Forkhead-associated |
| FAM | Fluorescein amidites |

| | |
|---|---|
| *E.coli* | *Escherichia* coli |
| GST | Glutathione S-transferase |
| PDB | Protein data bank |
| $T_m$ | Melting temperature |
| GFP | Green fluorescent protein |
| LB | Luria broth |
| SDS-PAGE | Sodium dodecyl sulfate–polyacrylamide gel electrophoresis |
| EMSA | Electrophoretic mobility shift assay |
| SAXS | Small angle x-ray scatter |
| WT | Wildtype |
| DLS | Dynamic light scattering |
| RMSD | Root-mean-square deviation |
| pTM | Predicted template modelling |
| ipTM | Interface predicted template modelling |
| pLDDT | Predicted local distance difference test |
| PAE | Predicted aligned error |

# Chapter 1 – Introduction

## 1.1 Types of DNA Damage and Links to Human Health

Deoxyribonucleotide (DNA) serves as the genetic blueprint, encoding essential information for cellular function. Cells depend on an intact genome for nominal function, but genetic integrity is constantly challenged by endogenous and exogenous damaging agents, both of which are estimated to contribute to $10^5$ DNA damage events per cell per day[1,2]. The barrage of endogenous DNA damaging sources can arise from metabolic processes which may produce reactive metabolites or oxygen species (ROS), DNA replication error and stress, and the presence of water as a solvent may lead to hydrolysis[3]. Exogenous sources of DNA damage can range from ultraviolet light (UV), chemotherapies, and ionizing radiation (IR)[3]. DNA damage can manifest in different forms: misrepaired, or misincorporated DNA bases from replication errors or stress, base, and sugar modification from reactive metabolites, inter and interstrand crosslinks generated from chemotherapeutics, and single-strand or double-strand breaks from ionizing radiation. If DNA lesions persist or are repaired erroneously sites of DNA damage can lead to genomic instability highlighted by chromosomal abnormalities, and mutations leading to cell cycle arrest and cell death, or mutations which overtime give rise to cancer[4]. Cells have evolved an intricate network of signalling pathways to sense DNA damage and repair it in a timely matter preventing cell death or cancer.

## 1.2 DNA Damage Response

As mentioned above DNA damage can manifest in a variety of ways. The DNA damage response or DDR, is the network of proteins responsible for detecting DNA damage, recruiting factors, and repairing the lesions[4]. Reactive oxygen species ($O_2^{-\cdot}$) from metabolic processes can react with nitric oxide (NO), which is involved in neurotransmission, immune system function, and vasodilation to form peroxynitrite ($ONO_2$) which can lead to DNA base oxidation and strand breaks[5,6]. The modified bases can be repaired via base excision repair pathways (BER). Reactive species can also form bulky base adducts disrupting the DNA helical structure by blocking Watson-crick base pairing. Tobacco carcinogens, more specifically polycyclic aromatic hydrocarbons can react and form bulky adducts, and the persistence of bulky adducts have been found to be highly mutagenic[7]. Furthermore, the formation of bulky base adducts can be caused by UV rays forming pyrimidine dimers, which if unrepaired can lead to DNA strand breaks. These DNA modifications are typically repaired via nucleotide excision repair (NER)[8]. In addition to the mentioned DNA damage response pathways (BER and NER), other mechanisms include mismatch repair (MMR), the Fanconi anemia (FA) pathway, translesion DNA synthesis (TLS), non-homologous end joining (NHEJ), and homologous recombination (HR)[4]. The MMR mechanism operates on DNA mismatches and DNA replication-dependent insertion/deletion loops[9]. For the repair of DNA inter-strand crosslinks, the FA pathway is activated[10]. TLS functions as a bypass mechanism, facilitating the continuation of DNA replication despite base damage by interchanging polymerases, followed by subsequent repair[11]. HR and NHEJ are two distinct DDR pathways utilized for the repair of DNA double-strand breaks (DSBs), each functioning in different stages of the cell cycle[4]. To align with the primary focus of this thesis, only HR and NHEJ will be explored in greater detail.

## 1.3 DNA Double Strand Breaks

DNA DSBs occur when both strands of the DNA duplex are severed on the phosphodiester backbone broken creating DNA lesions close in space. Regardless of the source, DNA DSBs are among the most dangerous DNA lesions, driving chromosomal translocation, genomic instability and may lead to cancer or cell death[12,13]. Additionally, faulty DSB repair pathways contribute to various disorders involving developmental, immunological, and neurological pathways[12–14]. DSBs can arise from pathological exogenous sources, such as IR which can directly ionize DNA, damaging the bases, or cleaving the backbone or indirectly by ionizing water molecules producing hydroxyl radicals[15]. Furthermore, they can arise from the exposure to topoisomerase poisons[16]. Topoisomerase enzymes play a role in relieving the topological strain caused by processes like the winding and unwinding of DNA duplexes during DNA replication. These enzymes create breaks in the DNA strands to alleviate torsional strain, and then reconnect the ends to restore the continuity of the DNA molecule. Inhibitors of topoisomerases, such as camptothecin (CPT), interfere with the re-ligation step, preventing the reconnection of DNA ends and leaving topoisomerases bound to DNA containing a single-strand break. If a DNA replication fork encounters a single-strand break stabilized by CPT, it halts and the replication fork collapses, resulting in the generation of a one- or single-ended DSB[16,17]. Endogenous sources such as single strand breaks related to ROS DNA damage, can escalate to DSBs if the lesions are within proximity and on opposite strands[18]. Additionally, unrepaired DNA lesions when encountered by DNA replication forks trigger replication fork collapse[13,16–18]. Indeed, spontaneous DSBs may arise from programmed cellular processes, which is the case during meiosis. At the first meiotic division, DSB repair is necessary for correct chromosomal segregation, which in turn generates gametes with different allele combinations from the parent germ line[13]. The two main pathways responsible for the detection and repair of DSBs are non-homologous end joining (NHEJ) and, homologous recombination (HR)[4]. The choice between the two main pathways is cell cycle dependent[13]. NHEJ can typically function in all stages of the cell cycle but is most active in Gap 1 (G1) phase in the cell cycle where eukaryotic cells are growing and is the phase in between cell division and DNA replication[19]. NHEJ is an efficient DNA repair pathway that can act in as little as half an hour and does not require a template strand to ligate the breaks. Instead the two ends flanking the break are directly ligated together[20]. While it is a practical process, NHEJ is considered error-prone due to the lack of homology in ligating the DNA ends which may lead to chromosomal translocation, due to the DNA ends being fused to different locations within the genome[4,12,13,21]. Furthermore, chemically mismatched DNA ends (3' phosphate group and a 5' hydroxyl) which are not suitable for ligation may require additional enzymatic processes that may lead to insertions, deletions, or substitutions at the break[22]. HR repair pathways are homology driven, and rely on a donor sequence, which is usually the sister chromatid from newly replicated cellular genome[4]. This reliance on homology restricts HR mechanisms to be active within the synthesis phase (S) and post synthesis and in between cell replication and cell division, Gap 2 phase (G2)[19,21]. Although HR is not as efficient as NHEJ (~ 7 hours), the need for HR machinery to sift through the genome and find a homologous region to repair the DSB makes it a less error prone pathway[1,4,7,13,21].

## 1.4 DSB Repair via Homologous Recombination

HR begins with the rapid recruitment of Ku (Ku70/80) as the Ku heterodimeric protein is abundant throughout all cell cycles, and binds DNA tightly[23,24] (

**Figure *1.1***). In addition to Ku, the MRN complex (Mre11, RAD50 and Nbs1) and ataxia-telangiectasia mutated (ATM) are also recruited. Mre11 in the MRN complex has endo- and exonuclease activity and is active in the presence CtBP interacting protein (CtIP) which is phosphorylated by ATM when recruited[21]. The CtIP-MRN complex then carries out short range 3'- 5' end resection through its Mre11 subunit generating a single stranded region which favours HR and displaces Ku and prevents further binding of NHEJ proteins, making DNA end resection a key step in determining pathway choice[25–27]. Longer range DNA processing enzymes DNA2, and EXO1 are further recruited which initiate 5' – 3' DNA end resection to generate single stranded DNA (ssDNA) 3' overhang[21,27]. Replication protein A (RPA) filaments are then recruited to protect the DNA ends, followed by the recruitment of RAD51, which eventually replaces RPA with the help of PALB2 (partner and localizer of BRCA2) and BRCA2 (breast cancer gene 2) proteins. This involves disrupting the bond between RPA and single-stranded DNA (ssDNA) while promoting the association of RAD51 with ssDNA[13]. RAD51 oligomers initiate strand invasion on a sister chromatid scans a homologous region allowing for DNA synthesis by DNA polymerases, followed by annealing and ligation completing the repair process[4,13,21].

## 1.5 DSB Repair via Canonical Non-Homologous End Joining

The initial critical stage of non-homologous end joining (NHEJ) involves the association of the damaged DNA ends with the scaffold protein Ku heterodimer[23,24] (

**Figure *1.1***). Structurally, the Ku70/80 heterodimer forms a ring-shaped DNA binding domain capable of accommodating two DNA double-strand turns, and its binding activity protects DNA ends from endonuclease activity[21]. Subsequently, a DNA-dependent serine/threonine protein kinase catalytic subunit (DNA-PKcs) is recruited to the damaged sites upon interaction with DNA-bound Ku[28]. This complex facilitates the close proximity and stabilization of the broken DNA ends to facilitate further processing, and its autophosphorylation signals recruitment of down stream factors involved in DNA end processing. Given the unplanned and undesired nature of DNA breaks, these resulting ends often cannot be directly rejoined. Several repair factors, including polynucleotide kinase/phosphatase (PNKP) and Artemis, are believed to participate in the repair of these irregular DNA ends[29]. The remaining gaps are then filled in by μ and λ polymerases. The ligation of the ends together is carried out by the DNA ligase IV, XRCC4, and XLF complex[28,29]. Additionally, an alternative pathway known as alternative non-homologous end joining (A-NHEJ) exists as a backup or alternative, particularly in cases involving mutated or absent classical NHEJ factors. A-NHEJ utilizes the MRN complex, XRCC1, and PARP1 to create microhomology regions on the broken DNA ends through nucleolytic degradation, followed by the formation of short complementary regions[30]. Polymerase β can then fill in the gaps, and ligation can occur via ligase I or III[31].

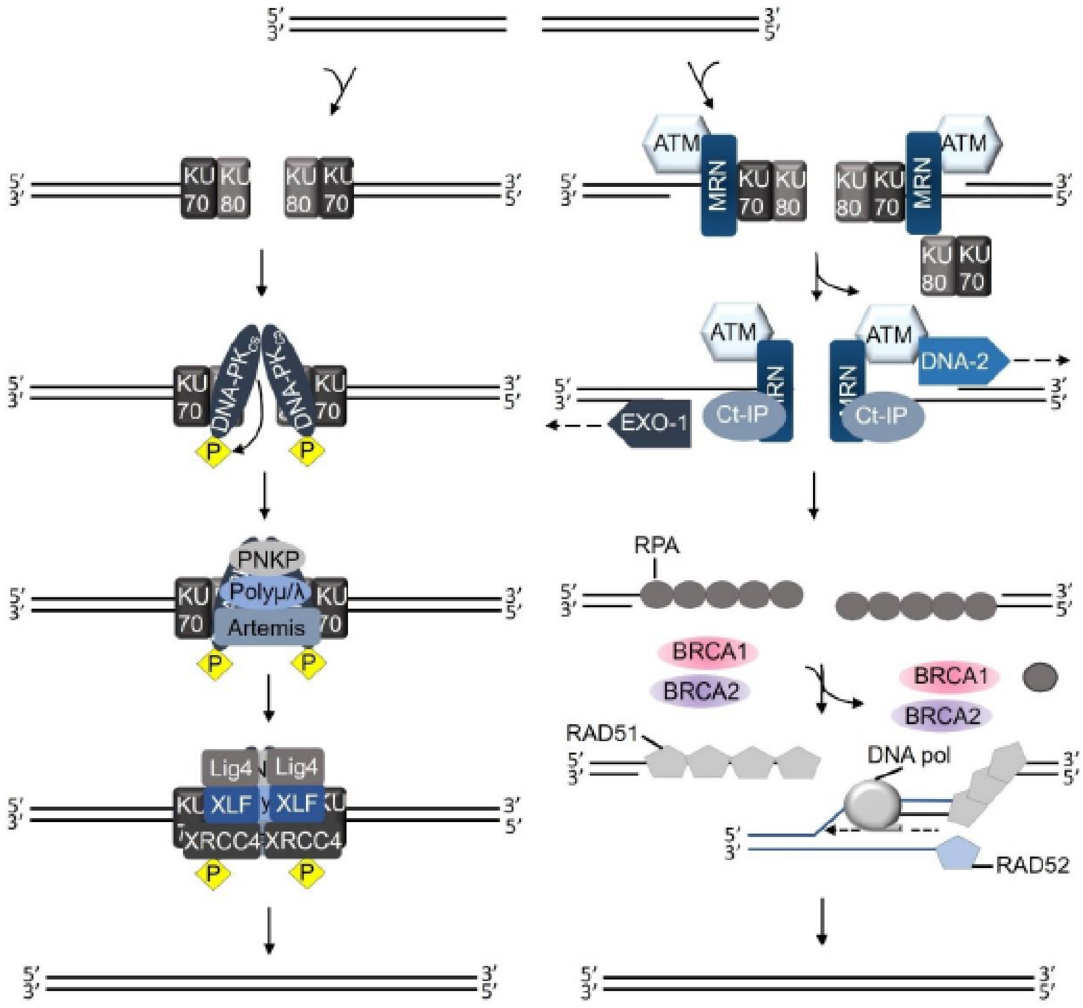


**Figure 1.1. Major Pathways of DSB repair.**

Left panel: In NHEJ, Ku protein is rapidly recruited to DSB, followed by DNA-PKcs recruitment and autophosphorylation. Polynucleotide kinase 3'phosphatase (PNKP), Artemis, and polymerases are then recruited. PNKP is able to repair non-ligatable ends containing 5'OH or 3'P and Artemis has nuclease activity. X-ray repair cross-complementing protein 4 (XRCC4) and XRCC4-like factor (XLF) are recruited and act as scaffolds, as DNA ligase 4 ligates the DNA ends. Right panel: HR Ku and the MRN complex are both recruited independently to DSB, followed by ATM recruitment and activation. MRN then carries out 3'-5' end resection displacing Ku. ATM phosphorylates and recruits CtIP followed by the recruitment of exonuclease-1 (EXO-1) and

DNA2, and extensive 5'-3' end resection. RPA filaments bind ssDNA overhangs for protection, followed by recruitment of BRCA1 and BRCA2 which replace the filaments with RAD51. RAD51 scans and initiates strand invasion on a sister chromatid with RAD52, and polymerases synthesize the nucleotides in an error free manner.

## 1.6 Posttranslational Modifications and DSB Signaling

Extensive evidence suggests that DNA repair effectiveness and timeliness are regulated through reversible post-translational modifications (PTMs)[32–34]. These modifications involve the addition or removal of small chemical groups (such as methyl, acetyl, phosphoryl, adenosine diphosphate or proteins like ubiquitin) to biomolecules by specific enzymatic processes. These mechanisms, such as conjugation (for example, methylation, phosphorylation, ubiquitylation) and deconjugation, act as molecular toggles that synchronize the functions of various proteins, both near and far to DNA damage sites and happen rapidly in a response to DNA lesions[35,36]. PTMs are known to facilitate interactions between proteins, regulate enzymatic functions, influence protein localization within cells, and target proteins for degradation[32–34]. A cascade of PTMs is triggered upon DSB detection, the two main PTMs we discuss in our work are phosphorylation and ubiquitylation.

Phosphorylation involves the conjugation of a phosphoryl group onto serine, threonine, or tyrosine residues imparting a negative charge which can mediate or prevent protein-protein interactions by affecting binding affinities or introducing conformational changes in target proteins[37]. Kinases transfer phosphoryl groups on to proteins, and phosphatases catalyze their removal. Phosphatidylinositol-3 kinase-related kinases (PIKKs), a major family of kinases, play key roles in DSB response mainly: DNA-PKcs, ATM, and ATR (ataxia telangiectasia-mutated and Rad3-related)[38–40]. Remarkably, a similar feature of PIKK enzymes is their size, all of which are within 280-480 kDa[40]. Additionally, they are serine/threonine kinases, and the kinase domains are much more like phosphatidylinositol 3-kinase family than other serine/threonine kinase domains, but interestingly, all evidence points towards specificity of PIKK to protein targets, over lipid[40]. DNA-PKcs is involved in the coordination of NHEJ **(section 1.4)**[38,39]. ATM is an essential kinase which phosphorylates histone H2A variant H2AX on residue Ser139 (also known as γH2AX) upon its recruitment after MRN complex detects the DSB, triggering a cascade of protein recruitment and amplifying the signalling far from the DSB[38]. Additionally, ATM phosphorylation of checkpoint kinase 2 (CHK2) and p53, activates certain cell programs in response to DSB, mainly leading to cell cycle arrest in G1 phase, and preventing programmed cell death (apoptosis)[38,41]. ATR is a kinase activated mainly by UV-induced damage, or ssDNA coated by RPA complexes, such structures arising from helicases and polymerases during DNA replication or during HR respectively[38,42]. The phosphorylation of checkpoint 1 (CHK1) by ATR slows cell cycle progression and assists in resolving replication stress[42]. Cyclin-dependent kinases (CDKs) are key kinases that control transitions in the cell cycle[43]. The accumulation of CDKs at specific cell cycle phases, and their activation by binding to their regulatory subunits, the cyclins, trigger progression through the cell cycle stages[43,44]. The progression through S, G2, and mitotic phases are triggered by the phosphorylation activity of CDK2 and CDK1[43]. Furthermore, CDK- dependent phosphorylation events ensure HR can proceed during S/G2 by phosphorylating DNA end resection proteins such as Mre11, Nbs1, CtIP, activating HR and DNA end resection[45]. Ubiquitin (Ub) is a signaling protein utilized by cells. One of the uses of Ub is protein recruitment and amplification of DSB signaling when Ub is transferred to proteins on chromatin. Ubiquitylation is the focus of our work and therefore is further reviewed in detail in **section 1.6.**

The cascade of PTMs extend beyond the actual repair of DNA DSB and contribute to the amplification of signalling many DNA bases downstream and upstream of DSB[46]. Furthermore, several types of PTMs simultaneously participate in regulation. A key pathway exemplifying the crescendo of PTMs starts when the MRN complex detects a DSB, and ATM is recruited, phosphorylating the histone H2A, becoming γ-H2AX[33,46–50]**(Figure 1.2)**. This facilitates the assembly of a large adaptor protein, MDC1, which binds γH2AX via its BRCT[50–52]. Subsequently, MDC1 undergoes two crucial phosphorylation events. The first phosphorylation, mediated by casein kinase 2 (CK2), involves the modification of Ser-Asp-Thr motifs that facilitate binding interactions with Nbs1, a component of the MRN nuclease complex[50]. The second phosphorylation, carried out by ATM kinase, enables the recruitment of RNF8 through its phospho-protein binding FHA domain[50,53,54]. RNF8, an E3 ubiquitin ligase, forms Lysine-63 (K63) linked polyubiquitin (polyUb) chains with the Ubc13/Mms2 E2 heterodimer[54–58]. These ubiquitin chains likely target an unidentified substrate, leading to the recruitment of RNF168, which amplifies the K63 chains through its ubiquitin-interacting motifs (UIMs)[48]. Once recruited, RNF168 collaborates with the E2 enzyme UbcH5c to monoubiquitinate histones H2A/H2AX on K13-15[59,60]. The resulting ubiquitylation of H2A at K15 (H2AK15ub) serves as a docking site for 53BP1, potentially favoring non-homologous end joining (NHEJ) over BRCA1-mediated repair pathways[61]. Additionally, the extended K63 chains generated by the cooperative activity of RNF8 and RNF168 recruit RAP80 via its UIMs which in turn facilitates the recruitment of ABRAXAS1 and BRCA1, promoting homologous recombination (HR)[62–64].

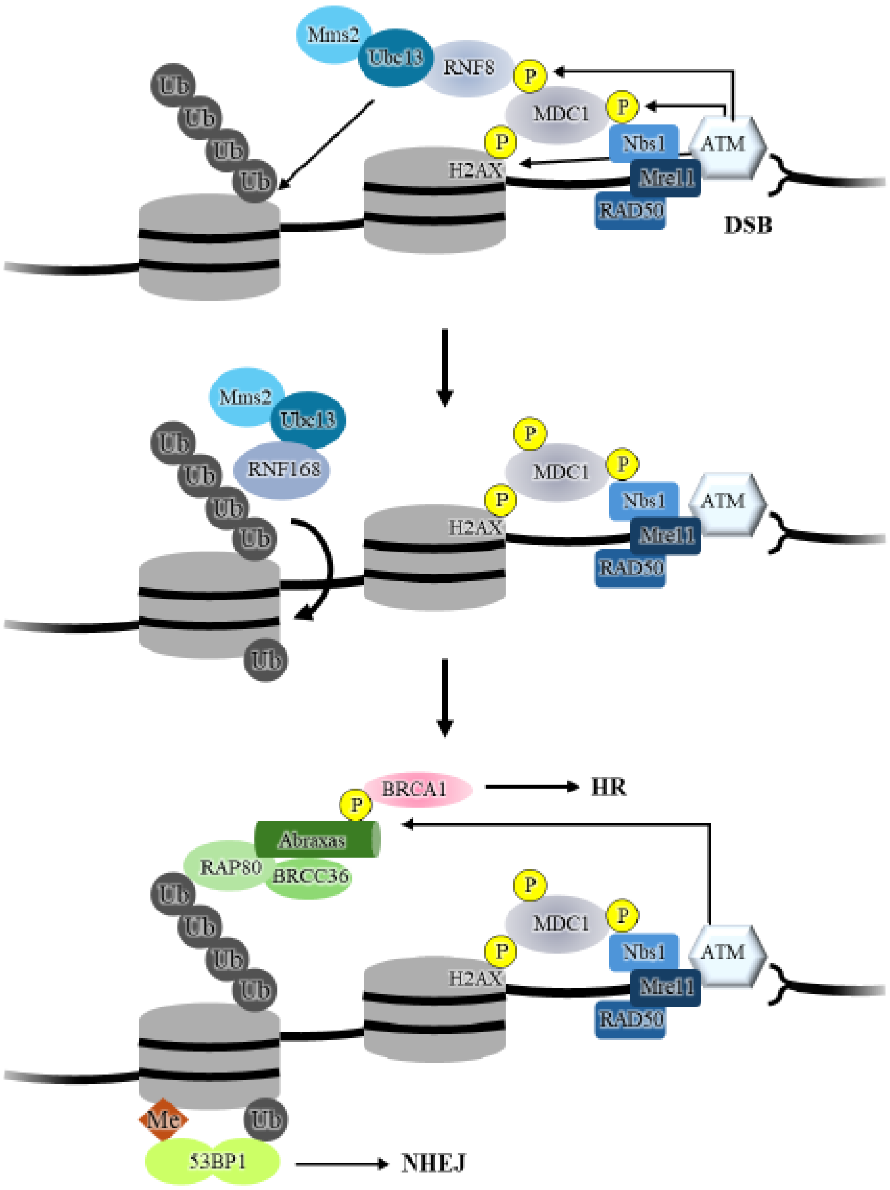


**Figure 1.2. DNA DSB signal amplification via PTM cascade.**

When DNA damage occurs, a signaling cascade primarily regulated by ATM is set in motion. This process entails the ubiquitylation of nearby histones through the actions of RNF8/RNF168, in addition to various other post-translational modifications. Depending on the cell cycle phase, the initial pathway can result in either homologous recombination (HR) repair, which occurs exclusively during the S/G2 phases, or non-homologous end joining (NHEJ), involving BRCA1 or 53BP1, respectively.

## 1.7 Cellular Ubiquitylation:

Ubiquitin, a 76 amino acid signaling protein, is highly conserved and found across all life forms and compared to other PTMs (methylation, acetylation, phosphorylation) its size makes it unique allowing for a wider range of signaling activity[65,66]. Ubiquitin is highly stable, adopting β -grasp fold (four stranded β -sheet curls around a central alpha helix, and a flexible 6 residue C-terminal tail)[65]**(Figure 1.4)**. Ubiquitin can be transferred onto a substrate lysine and conjugated via its C-terminal carboxylate tail forming an isopeptide bond resulting in monoubiquitinated substrate[65–67]. A substrate can be monoubiquitylated on one site, or monoubiquitylated on multiple sites. In addition to monoubiquitylation of substrates, polyubiquitin chains can form through the successive conjugation of ubiquitin molecules on a substrate forming isopeptide bonds via its seven lysines (K6, K11, K27, K29, K33, K48, K63) and its N-terminal methionine (M1)[66] **(Figure 1.4A)**. The polyubiquitin chains can form homogenous (ubiquitin connected via K48 only), linear (K63), mixed and branched polyubiquitin chains (K11/K48)[66]**(Figure 1.4D)**. A remarkable number of genes in the human genome encode the ubiquitin system: two E1 ubiquitin activating proteins (E1), ~40 Ubiquitin conjugating enzymes (E2), and greater than 700 ubiquitin ligases (E3) genes, accounting for around 5% of human protein-coding genes[66,68]. Initially, in the first stage, E1 adenylates the carboxy-terminus of Ub (Gly76), activating it for the formation of a thioester linkage to the active-site Cysteine (Cys) of E1[66,69]**(Figure 1.3)**. Subsequently, Ub is transferred from E1 to the active-site Cys of an E2[66,69]. Interestingly, the linkage between subsequent ubiquitin molecules within a chain is determined by the specific E2 involved in the reaction. Finally, the E2, often working in conjunction with an E3, coordinates the creation of an isopeptide bond between the carboxy-terminus (C-terminus) of ubiquitin and a lysine ε-amino group on a target protein[66,68,69]. Through a step-by-step approach, polyubiquitin chains are formed through successive rounds of the E1-E2 steps facilitated by E3. In this process, the C-terminus of the donor ubiquitin molecule is linked to a lysine residue on an acceptor ubiquitin instead of the lysine residue on the substrate (target) protein[66]. A more in-depth look at the various chain types of ubiquitin is demanded to gain insights into just how effectively nature has made use of ubiquitylation, to regulate and transmit information in the cell.

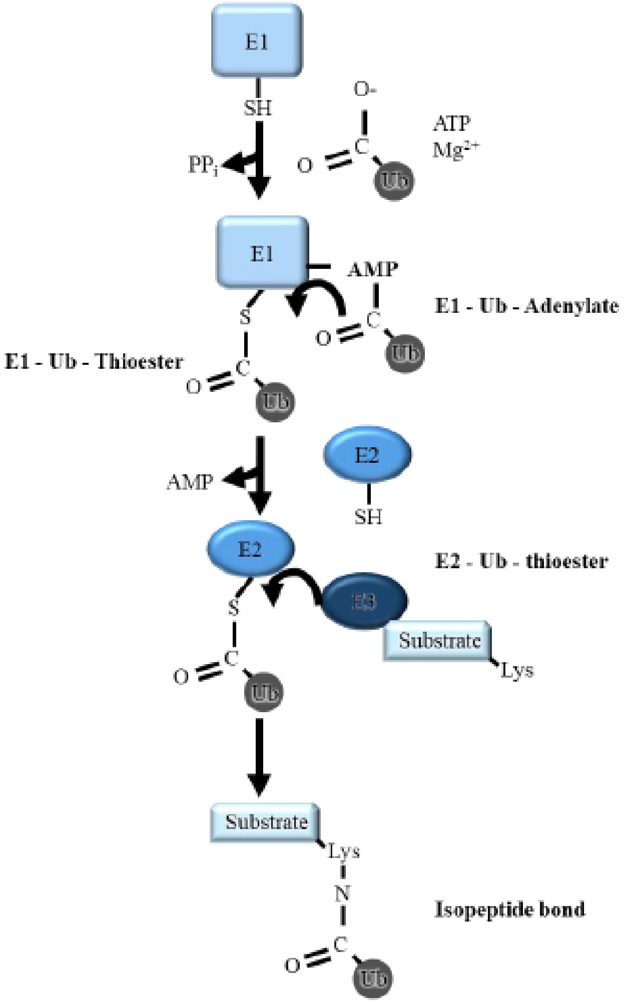


**Figure 1.3 Ubiquitylation Cascade**

Ubiquitylation of a substrate occurs in a step-by-step process involving E1-E2-E3 ubiquitlyation cascade targetting substrate lysines to form isopeptide bonds. Synthesis of polybubiqutylation ivolves reiterative process targetting lysine residues on the ubiquitin tethered to the substrate.

PolyUb chains are specifically recognized by ubiquitin binding domains (UBDs) in proteins that recognize the different topologies and conformations of polyUb[66,70]. Recently, ubiquitin has been found to be modified by phosphorylation and acetylation, which adds another layer of complexity to ubiquitin signaling[70]. Furthermore in the last several years, ubiquitin linkages to residues other than lysine on substrates have been identified[71].

The two extensively studied types of ubiquitin chains involve ubiquitin lysine 48 (K48), which serves as the canonical signal for proteasomal degradation, and K63 which has non-degradative roles, implicated in DNA damage response (DDR) and NF-κB immunological signalling[66,72,73]. However, various types of ubiquitin chains have been detected in studies of eukaryote organisms **(Figure 1.4D)**. Linear M1 chains are essential for activating the NF-κB pathway through NF-κB essential modulator (NEMO) recognition and binding via the ubiquitin-binding in ABIN and NEMO (UBAN) motif, which was found to bind linear ubiquitin chains[74,75]. K29-linked chains participate in proteasomal degradation and innate immunity responses[76,77]. K6-linked chains participate in DDR and mitochondrial homeostasis[70]. Additionally, K11-linked chains were originally thought to target substrates for proteasomal degradation, but it was found that K11-linked chains are poor substrates of the proteosome[70]. More recently K11-linked chains have been implicated in forming the anaphase-promoting complex cyclosome (APC/C), contributing to cell cycle regulation, additionally participating in DDR, and protein stabilization[78,79]. K33-linked chains have been implicated in regulating protein-protein interactions, immune response, and cellular trafficking processes[73]. The activity of K27-linked chains has remained ambiguous but more recently has been implicated in DDR pathways. It has been proposed that two E3 ligases are involved in the formation of K27-linked chains within cells. RNF168, a RING E3 ligase, has been documented to catalyze the assembly of K27-linked ubiquitin chains on histones H2A and H2A.X within cellular contexts, aligns with the established functions of RNF168, particularly in response to DNA damage[70,80]. These chains have been observed to facilitate the recruitment of DNA damage repair factors, such as p53 binding protein 1 (53BP1), to sites of DNA damage[80]. Emerging insights into the functions of mixed and branched polyubiquitin chains have further revealed complex signalling functions. K11/K48 branched polyUb chains for example, strengthen ubiquitin signalling emanating from substrates and function as proteasomal priority signals in cell cycle and protein quality control pathways[70]. Recent proteomic studies have revealed the covalent modification of ubiquitin by phosphorylation and acetylation, and the first studies into the physiological relevance of the modifications only came about in 2014[70]. Eleven phosphorylation sites were discovered on ubiquitin in proteomic screens **(Figure 1.4B)**. The most well studied site is the phosphorylation at S65 on ubiquitin which has been implicated in mitophagy signalling[81,82]. Ubiquitin phosphorylation by protein kinase 1 (PINK1), a well studied protein linked to Parkinson's disease, was shown to act as a recruitment platform for Parkin (an E3 ligase which accumulates in response to mitochondrial damage, and PINK1 signalling), and autophagy adaptors which begin the process of mitophagy[83]. Phosphorylation on S65 changes the structure of ubiquitin, and more importantly the surface potential due to the negative nature of the phosphate group and these alterations are thought to be recognized by Parkin and other unknown proteins driving mitophagy[70,81,82]. Ubiquitin acetylation occurs at the same lysine residues which typically would form polyubiquitin chains, therefore acetylation competes with ubiquitin chain formation, impacting the overall shape and architecture of polyubiquitin chains[84] **(Figure 1.4C)**. K6, and K48 acetylated ubiquitin have been observed in cells, but acetylation occurs on other lysine residues as well. Acetylation of K6 and K48 appears to affect the ability of the ubiquitylation machinery to transfer ubiquitin onto substrates. Notably, acetylated ubiquitin has been found on histone proteins, specifically K6, K48 acetylated ubiquitin. An acetyl memetic mutation of K6Q stabilized monoubiquitinated histone H2B[70,84]. Acetylases of ubiquitin are yet to be characterized and further characterization is required. In the most recent years, evidence of ubiquitin chains forming on substrate residues other than lysine residues have emerged[71]. For instance, the ubiquitylation of threonine by the human E3 ligase MYCBP2 has been observed, where the ubiquitin is linked to the threonine via an ester linkage rather than the canonical isopeptide linkage[85,86]. Additionally, evidence of ubiquitylation of non-protein molecules, such as the ubiquitylation of ADP-ribose (ADPR) has been observed. RING E3 ligase DELTEX3L interacts with the ADP-ribosyltransferase PARP9, indicating a potential link between these two post-translational modifications. Furthermore, under conditions where E1, E2, and the DELTEX3L–PARP9 complex were present, it was observed that the C-terminus of ubiquitin underwent ADP-ribosylation[87]. Further investigation into the physiological roles of these modifications is necessary, setting the stage for exciting discoveries. There is still much to learn about how the cell uses ubiquitin and its diverse chains. It should be noted that in addition to ubiquitin, there are a number of ubiquitin-like modifiers, such as SUMO and NEDD that share similarities to ubiquitin but possess distinct differences[67]. Much still remains to be characterized. A more in-depth look at the E2-E3 machinery will provide further insight into just how effectively nature has made use of ubiquitylation to regulate and transmit information in the cell.

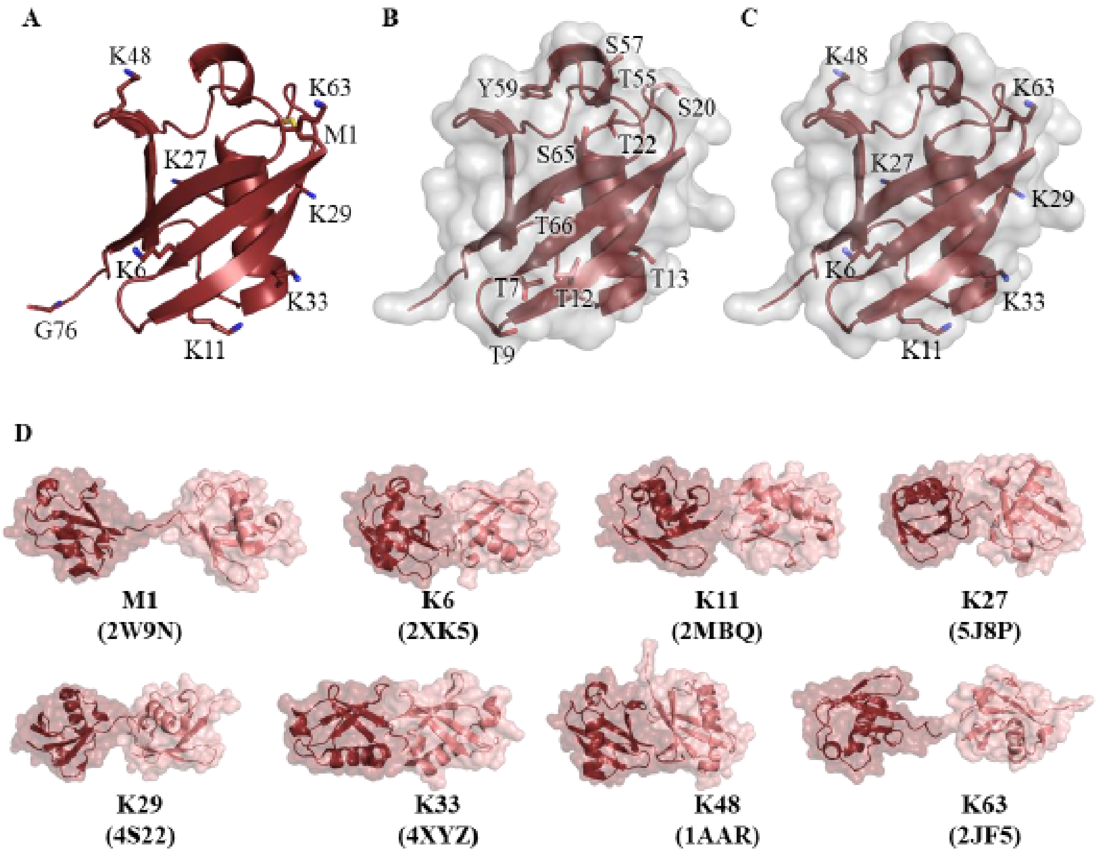


**Figure 1.4. Ubiquitin chains and modification sites.**

**A**. Crystal structure of ubiquitin with the highlighted lysine residues that are involved in forming polyubiquitin chains, and the carboxyl terminal glycine. Protein Data Bank accession code (PDB): 1UBQ. **B.** 11 Phosphorylation sites identified on ubiquitin, with surface of ubiquitin shown in white. **C.** 7 detected acetylation sites on ubiquitin, with surface of ubiquitin shown in white **D.** Different diubiquitin chain linkages. The acceptor ubiquitin is colored dark red, and donor ubiquitin is salmon colored. PDBs of each crystal structure is reported below each chain. Starting from top left to bottom right: M1 chains are open, flexible and linear. K6 Chains form compact chains. K11, K27, and K29, K33 chains can be compact and open with different orientations of the ubiquitins chainging the topology. K48 is highly compact compared to other chains. K63 chains are linear, and flexible.

Overall our work has focused around understanding the mechanisms of E2 conjugating enzymes, specifically Ubc13 **(Chapter 2)** and E3 ubiquitin ligases, focusing on RNF138 in **(Chapters 3 and 4)**. Therefore, a more detailed examination centred around the around the roles of the two proteins, their signalling activities, and the therapeutic relevance is necessary.

## 1.8 Ubc13 and K63-Linked Ubiquitin Chains

The network of signaling interactions emanating from DSB is a result of the ubiquitylation of chromatin associated proteins around the break[54,57,61]. K63-linked ubiquitin chains play a crucial role recruiting downstream factors. They are formed by a specific ubiquitin conjugating enzyme E2 (named UBE2N or Ubc13), and either of its non-catalytic, E2-like binding partners, Mms2 or UEV1A[55,88]. In the nucleus, Ubc13 bound to Mms2 is targeted to the DSBs through its interactions with RNF8 a RING E3 ligase. This complex plays a crucial role in both HR and NHEJ, through recruitment of key downstream factors BRCA-1, or 53BP1 respectively in a cell cycle dependent manner **(section 1.5)**. Ubc13/Mms2 are also involved in nucleotide excision repair, in response to UV-induced DNA damage. Ubc13/Mms2 interacts with RNF111 which in response to DNA damage binds UV-DNA damage sensing protein XPC, regulating its accumulation on damaged DNA[89,90]. Ubc13/Mms2 is also involved in DNA damage tolerance (DDT) pathways which are triggered when the DNA replication machinery come across DNA damage on the template strands at replication forks. Upon DNA damage of the template strand, DNA polymerases halt and detach from the helicases, and RPA rapidly builds up on the resulting ssDNA[91]. RPA recruitment drives further recruitment of proteins which modify proliferating cell nuclear antigen (PCNA), promoting the recruitment and interaction with polymerases involved in the error prone translesion DNA synthesis (TLS) repair pathway[91,92]. The error-free lesion bypass or template switching (TS) repair pathway is engaged when Ubc13/Mms2, with E3s SHPRH and helicase-like transcription factor (HTLH), ubiquitylates and elongates K63-linked ubiquitin chains on PCNA, driving the switch from TLS to TS repair[89,92]. Ubc13/Mms2 is further implicated in Fanconi anemia repair pathways, which are responsible for repairing interstrand crosslinks[89]. RNF8 is thought to be the E3 ligase involved in this pathway, K63-linked ubiquitin chains are recognized by FA core protein complex

FA-associated protein 20 kDa (FAAP20), which upon recruitment leads to further signalling and downstream recruitment of DNA damage repair proteins[93]. Additionally, the activity of Ubc13/Mms2 and RNF8 have been shown to play an integral role in telomere end protection[89]. The ubiquitylation of Tpp1, a protein involved in binding and protecting telomere ends, by the Ubc13 complex stabilizes and retains Tpp1 on telomeres[94]. Ubc13's activity extends beyond the nucleus and is implicated in inflammation pathways due to its involvement in NF-κB signalling[95–97]. In the cytosol, Ubc13 interacts with UEV1A which is nearly identical to Mms2, and the heterodimer interacts with the E3 TRAF6, to generate cytoplasmic K63-linked polyubiquitin chains that recruit and activate I-kB kinases[96]. Ubc13/Uev1a are also known to interact with RNF8, and TRAF2, leading to NF-κB activation and translocation into the nucleus[89,95–97]. Ubc13 signalling may also lead to activation of apoptotic pathways[89]. The many cellular signalling pathways Ubc13 is implicated in warrant a closer look at its mechanism of action.

### 1.8.1 The Structure and Mechanism of Ubc13/Mms2

Structural studies of Ubc13 have revealed significant insights into the mechanism of K63-linked ubiquitin chain synthesis, and why Ubc13 is a unique E2[55,88,98–100] **(Figure 1.5)**. Ubc13 forms a tight heterodimer with Mms2 ($K_D$ = 49 ± 7 nM), and mutations to residues within the binding interface disrupting the heterodimer have detrimental effects on K63-linked ubiquitin chain formation[101,102] **(Figure 1.5A)**. Another unique characteristic of Ubc13 compared to other E2 proteins, lies within the active site **(Figure 1.5B)**. Structural comparison of E2 configurations reveals that the active site of Ubc13 takes on a distinct conformation within the E2 family[103]. A structural alignment with one of the other well characterized E2 enzymes, UbcH5c, reveals an overall high similarity between UbcH5c and Ubc13. A closer look at the active site reveals variations in the conformation of the active site loop residues ranging from 114-124 adjacent to the active site cysteine 87 (C87) in the two structures **(Figure 1.5C)**. This is consistent across evolutionary lines as both human and yeast Ubc13 exhibit the same active site loop conformations[104]. Interestingly, Ubc13 has an exposed L121 in the active site which allows for a more open conformation. In comparison, the canonical loop conformation in other E2 family members (UbcH5c in this case), the leucine is packed against the core fold adopting a closed catalytically active conformation **(Figure 1.5C)**. The active site loop in Ubc13 transitions from the open form, to closed conformation to in the presence of ubiquitin to facilitate catalysis[89,103] **(Figure 1.5D,E)**.

The previously mentioned mechanism of ubiquitylation involving an E1 activates ubiquitin through a trans-thioesterfication reaction. The donor ubiquitin is then transferred to the active site cysteine of the E2 enzyme, specifically C87 in Ubc13[55,88]. A second acceptor ubiquitin binds to Mms2 non-covalently, mediated by the hydrophobic patch on ubiquitin (L8, I44, V70) and surface residues of Mms2 M54, I56, and I67[55,88] **(Figure 1.5D)**. Given the interaction, the acceptor ubiquitin is positioned in way where K63 is positioned for attack on the thioester of the donor ubiquitin linked to C87 **(Figure 1.5E)**. In this interaction, K63 of the acceptor ubiquitin is hydrogen-bonded by N123, likely driving the conformational change to accommodate the incoming ubiquitin[55,89]. Markin et al. characterized the chemical environment of K63, and suggested that the suppression of the pKa and deprotonation of K63 on the acceptor ubiquitin promotes the nucleophilic attack on the thioester linkage in the active site of Ubc13, and the resulting reaction intermediate is likely stabilized by N79, which is also implicated in maintaining the active site loop structure[105,106]. Mutations in residues affecting the gating activity of Ubc13 have detrimental effects on catalysis[89]. Mutations in N79 reduced diubiquitin formation[106]. Furthermore, series of mutations in the active site loop D118 and A122 to glycine increased loop flexibility and had a direct effect on Ubc13's catalytic activity, increasing the rates of thioester hydrolysis, and impaired aminolysis of the ubiquitin thioester bond[103]. The mechanism of ubiquitylation by Ubc13/Mms2 points towards a mechanism of extending polyubiquitin chains on already monoubiquitylated substrates. *In vitro*, Ubc13/Mms2 can build free ubiquitin chains however the physiological relevance of this activity is unclear[59,107,108].

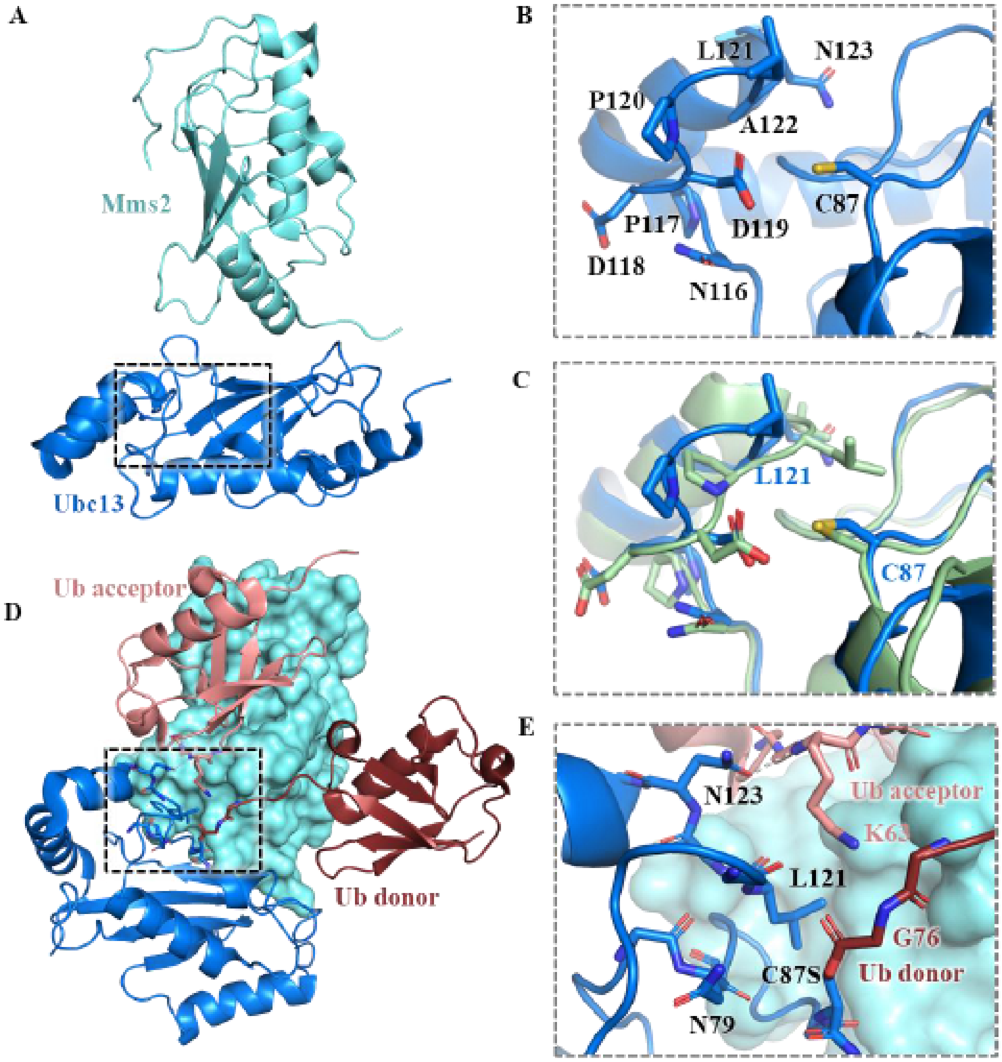


**Figure 1.5. Ubc13 active site and K63- linked diubiquitin formation with Mms2.**

**A.** Crystal structure of the heterodimeric E2 conjugating enzyme Ubc13 (blue) with its binding partner Mms2 (cyan), with the active site highlighted with a box corresponding to the active site. PDB accession code: 1J7D. **B.** Active site of Ubc13, with the active site loop residues and the catalytic cysteine labelled. **C.** Structural alignment of Ubc13's active site (blue), with UbcH5c (green). The active site leucine is highlighted, showing Ubc13 is in a open inactive state blocking access to the active site cysteine, whereas in UbcH5c, the leucine is in the closed conformation and cysteine is available for reaction with the C-terminal tail of ubiquitin. PDB accession code: 1X23. D. Crystal Structure of yeast Ubc13~Ub/Mms2~Ub. The acceptor ubiquitin interacts with Mms2 noncovalently, positioned to attack the donor ubiquitin covalently linked to C87 in Ubc13 PDB accession code: 2GMI. **E.** zoomed in snapshot of the interaction showing N123 directly engaging the acceptor ubiquitin, and N79 available to stabilize the reaction intermediates. Importantly, L121 is in the active closed form.

### 1.8.2 Ubc13/Mms2 Activity with RNF8

The role of E3 ligases for Ubc13/Mms2 is crucial, as they catalyze the last step in the E3 ligase cascade by targeting the E2~Ub complex to the protein target[66]. To aid in ubiquitin transfer, RING E3 ligases interact with both the substrate and a ubiquitin-conjugating enzyme (E2) linked to ubiquitin via a thioester bond[66]. Additionally, RING E3 ligases, play a crucial role in stimulating the catalytic activity of the E2[99,109,110]. Without an E3 ligase, the ubiquitin thioester attached to the E2 enzyme can assume various conformations, including a 'folded-back' configuration. The hydrophobic interaction of RING E3 ligases aligns the C-terminus of ubiquitin within the groove of an E2 enzyme, potentially facilitating the activation of the thioester by the lysine residue of the acceptor. This was observed in RNF4, a dimeric RING E3 ligase which was found to stimulate ubiquitylation activity of UbcH5a. Initially, one of the RING domains engages UbcH5a, while the thioester linked Ub would be engaged by the other RING. This secures the C-terminal tail of Ub in the active site groove of the E2, positioning the carbonyl of G76 thioester optimally for nucleophilic attack by the incoming substrate lysine. Ubc13 interacts with many E3 ligases some mentioned above, and our lab has previously looked at Ubc13/Mms2 interaction with RNF8. Hodge et al. showed RNF8 achieves conformational activation of E2-Ub complex, by the non-covalent interaction between RNF8 and Ubc13/Mms2~Ub. In a similar mechanism to RNF4, RNF8 orients the E2~Ub thioester linkage in a conformation favoring a catalytic attack from a nucleophilic substrate lysine residue. Mutations in RNF8 that specifically affect the E2 stimulating activity impair the DNA damage response.

### 1.8.3 Therapeutic Implications and Inhibitors of Ubc13/Mms2

Understanding the mechanism of activity of Ubc13 is crucial due to recent implications of Ubc13 in disease pathways. Briefly highlighted above are some of the functions Ubc13 plays in cellular signalling, specifically DNA damage repair pathways, and NF-κB signalling[67,108]. Additionally, Ubc13 is implicated in mitogen-activated protein kinase (MAPK) signalling[111]. Ubc13 is up-regulated in metastatic breast cancer cells, and Ubc13/Uev1a TRAF6 K63-linked dependent tumor metastasis was observed in mouse model xenograft studies of breast cancer[112]. K63-linked ubiquitin chains are required for activating mitogen-activated protein kinase kinase 1 (MEKK1), transforming growth factor β (TFGβ-activating kinase 1 (TAK1) and down stream MAPK cascades[111–113]. TFGβ pathways have been found to have the most potent effects on breast cancer metastases, by promoting migration, intravasation, and epithelial mesenchymal transition (EMT) in cancer cells[113]. In another breast cancer cell line (MDA-MB-231), it was shown that Uev1A promoted metastasis attributed to the activation of the NF-κB transcription factor in a manner dependent on Ubc13[112]. Further evidence has emerged of Ubc13 upregulation in prostate, colon, pancreatic, and lymphoma tumor tissues[113]. Ubc13 has also been implicated in chemotherapeutic resistance in certain cancer cells. Nasopharyngeal carcinomas treated with the DNA crosslinker cisplatin, which leads to collapse of replication forks and generation of DNA DSB, results in upregulation of DNA repair pathways, leading to therapeutic resistance[114]. Cells resistant to cisplatin show an increase in Ubc13 expression, and depletion of Ubc13 leads to desensitization to cisplatin[114]. There have been significant efforts aimed at exploring and creating Ubc13 inhibitors, primarily due to Ubc13's involvement in DDR and cytoplasmic signal transduction pathways associated with cancer. Certain approaches aimed at targeting the binding interface between Ubc13/Uev1A have been employed, of which two compounds were identified with varying binding affinities. A natural compound isolated from marine sponge species specifically *Luffariella variabilis* was reported to inhibit Ubc13 interaction with Uev1A , and also had effects on the proteosome[115]. More recently high throughput screens for compounds that inhibit NF-κB-mediated signalling, revealed a series of compounds that inhibit Ubc13[116]. Our group characterized the Ubc13 specific inhibitor NSC697923, along with a less specific inhibitor BAY 11-7082[100,117]. Both compounds were found to covalently modify Ubc13's active site cysteine and inhibit ubiquitin chain synthesis[100]. Interestingly, four mutations in Ubc13 D81N/R85S/A122V/N123P (quadruple mutant) resulted in a flipped Ubc13 active site loop. A crystal structure of the mutant revealed the mutant mimics the conformation present in UbcH5c[100]. NSC697923 was found to only inhibit Ubc13 and was not reactive against the Ubc13 quadruple mutant, suggesting NSC697923 is not reactive against other members of the E2 family of proteins[100]. The BAY compound inhibited both Ubc13, and the quadruple mutant. This is likely due to its smaller size which allows it to evade active site restrictions and interact with C87. Both compounds inhibited NF-κB activation and the DDR, albeit with off target affects in cells to varying degrees[100]. Cellular investigations utilizing the recently identified compound NSC697923 are still in their early stages, but initial findings suggest potential anti-cancer effects against neuroblastoma and diffuse large B-cell lymphoma cells, and more recently the use of NSC696923 has been investigated in the context of sensitizing colorectal cancer cells to radiation[116,118,119]. A wealth of knowledge about the active site dynamics of Ubc13 has been gained using the mentioned covalent compounds. Specifically, NSC697923 binds covalently to C87, with the nitrofuran moiety forming hydrogen bonds with N123. The pocket created by the open conformation of

Ubc13 facilitates specific interactions unique to Ubc13, thereby locking the protein in its inactive form[100]. In the canonical E2 enzyme conformations this pocket is blocked by the active site leucine, preventing the binding of NSC697923, providing an opportunity for future rational drug design strategies and inhibitor development[103].

To round off our discussion on ubiquitylation, a closer look at E3 ligases is warranted. Particularly focusing on RNF138, a significant participant in DNA damage response featured prominently in two chapters of our study.

## 1.9 RING Finger Protein 138 E3 ligase

E3 ligases are crucial for the final step of the ubiquitin cascade and catalyze the covalent transfer of ubiquitin from an E2 enzyme to a substrate lysine[69]. E3 ligases fall into two categories based on their mechanisms, those that act as intermediates and covalently attach ubiquitin before transferring it to a substrate, as seen in Homologous to E6AP C-terminus (HECT), and Really Interesting New Gene (RING)-in-between-RING (RBR) and those that do not form a covalently bound ubiquitin intermediate as is the case for most RING class of E3 ligases, and U-box E3 ligases[69]. RING E3 ligases typically contain a RING domain which interacts with an E2 enzyme stimulating ubiquitin transfer. The RING family of E3 ligases does not have a catalytically active site to receive the ubiquitin from the E2 enzyme, but instead play a role in positioning ubiquitin E2-ubiquitin complex in the optimal position for the nucleophilic attack by a lysine of a substrate[120]. RING domains can form dimers, or function as monomers and considered are a type of zinc finger. A canonical RING domain is about 40 to 60 amino acids in length and contains three cysteines and histidine bound to one zinc ion, and another 4 cysteines bound to a zinc ion (C3HC4)[120]. RING E3 ligases are prevalent in DNA repair signalling, and one of their many functions involves recruiting DNA repair machinery to sites of DSB, such as RNF8 and

RNF168[21,48,56,59,63,99]. RNF8 forms a homodimer through its coiled-coil domain upon activation and interacts through its RING domain with the heteromeric E2 Ubc13-Mms2 to synthesize lysine63-linked polyubiquitin chains on histone proteins surrounding DNA breaks[56,89]. These chains are recognized by monomeric RNF168 which contains two MIU **(section 1.5)**. The RING domain in RNF168 also interacts with the Ubc13-Mms2 and amplifies the ubiquitin signal to recruit down stream factors involved in DSB repair[48,59]. RNF138 is a conserved RING type E3 ligase initially discovered as a negative regulator of the Wnt-β pathway where it cooperates with E2-25K E2 conjugating enzyme and synthesizes K48-linked polyubiquitin chains on T cell factor/lymphoid enhancer factor targeting the complex to proteasomal degradation[121]. RNF138 ubiquitylates G2/M-phase specific E3 ubiquitin ligase (G2E3). This was initially hypothesized to lead to G2E3 proteasomal degradation but may perhaps alter its enzymatic activity or localization[122]. Additionally, RNF138 ubiquitylates voltage-gated calcium channels, specifically $Ca_v2.1$ channels present in neurons[123]. The over expression of RNF138 promotes polyubiquitylation of $Ca_v2.1$ channels in presynaptic and postsynaptic regions within the neuron, leading to higher protein turnover. Furthermore, RNF138 is implicated in synthesizing K63-linked ubiquitin chains facilitating NFκB and Type I interferon signaling[124,125]. Interestingly, RNF138 has also been implicated in spermatogenesis, as its downregulation led to apoptosis of spermatogonia[126]. Recent work points to the involvement of RNF138 in DSB repair, as the depletion of RNF138 sensitizes cells to DNA damaging agents such as IR[127–130]. Indeed, RNF138 was found to promote HR by directly binding DNA overhangs post Mre11 nuclease activity, but prior to CtIP recruitment[129]. Cooperating with UbcH5a (also UBE2Da) E2, RNF138 ubiquitylates CtIP, which facilitates its recruitment to DSB sites[130]. Ct-IP is an activator of DNA end resection[4,19,27]. In parallel, Ku80, a subunit of the Ku (Ku70/Ku80) heterodimeric complex is ubiquitylated and displaced from the DSB sites[129]. Ku is highly abundant, and has remarkably high binding affinity to DNA DSB, acting in a cell cycle independent manner[23,129]. Additionally, Ku acts as a protector of DNA ends from DNA end resection, and as a scaffold for NHEJ machinery[22]. Therefore, the combined actions facilitating the recruitment of Ct-IP, and the displacement of Ku, ensure that DNA end resection can proceed with the recruitment of HR downstream factors **(Figure 1.6).** A third target more downstream in the HR pathway has been identified. The ubiquitylation of Rad51D, which is a paralogue of Rad51 recombinase, is thought to contribute to Rad51 filament assembly in response to ionizing radiation through an unknown mechanism[35,131].

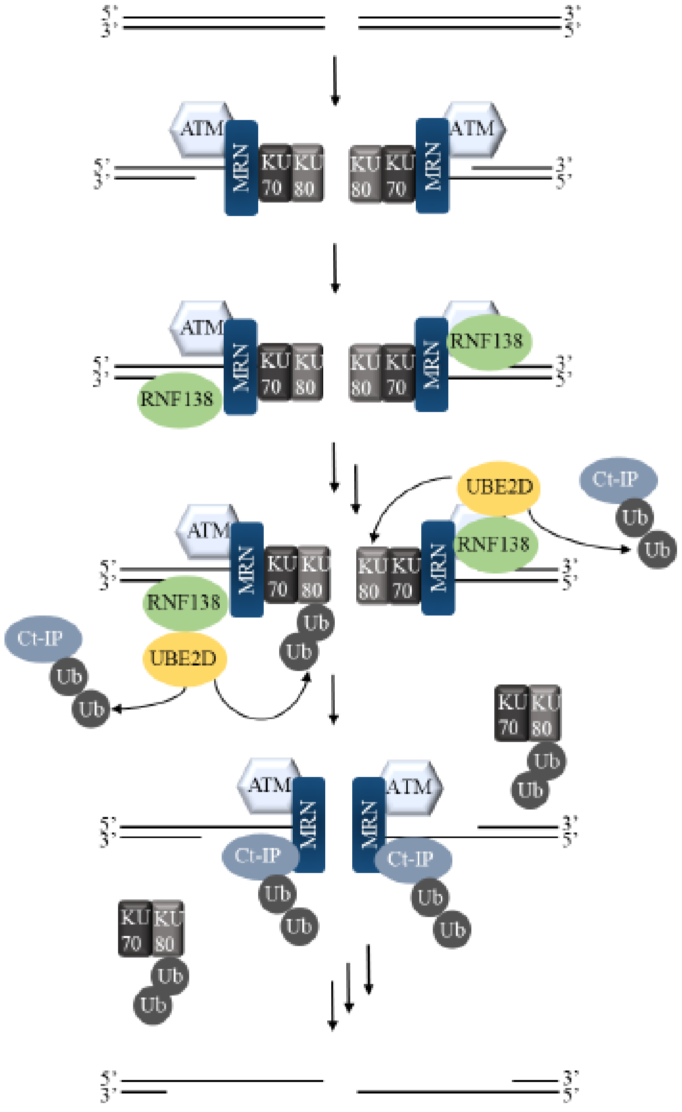


**Figure 1.6. RNF138 promotes HR by promoting DNA end resection.**

After a double-strand break (DSB) forms, Ku and the MRN complex quickly and independently bind to the DNA ends. The MRN initially performs limited 3′–5′ directed resection, creating binding sites for the RNF138–UBE2D complex. Subsequently, RNF138 ubiquitylates Ku80, displacing Ku from the DNA ends. Additionally, RNF138-UBE2D ubiquitylates CtIP promoting its recruitment to DSB breaks. This process creates conditions favorable for more extensive end resection by additional nucleases, promoting DSB repair HR.

RNF138 is related to a poorly characterized subfamily of RING E3 ligases which includes RNF125, RNF114, and RNF166[132]. All four proteins are similar in size and have a RING domain and three zinc fingers (ZNF), and a UIM[132]. RNF138's N-terminal RING domain plays a key role in protein-protein interaction and its three zinc fingers are thought to play a crucial role in recruitment to DNA DSB[121,129,130]. RNF138 is the only member of the subfamily to have been implicated in DNA binding, and DSB break repair so far. The first zinc finger (ZNF1) is atypical in nature (C2HC) and differs from the canonical DNA binding zinc fingers with C2H2 coordinating one zinc ion[133]. The C-terminal region contains two C2H2 zinc fingers but diverges from classical DNA binding zinc fingers by the spacing of conserved hydrophobic residues, and the spacing between the second cysteine and the first histidine[133]. RNF138 was found to directly bind DNA, with a preference to single stranded DNA overhangs, over double stranded blunt ends, or single stranded DNA, suggesting the atypical ZNF domains in RNF138 may interact with the DNA in a unique manner[129].

The structure of RNF138 remains to be characterized, therefore many questions persist about its mechanism, and its atypical functional domains. A recent study published a partial structure of RNF125 and revealed the importance of the RING domain and the first zinc finger (ZNF1) in the interaction with its E2 conjugating enzyme, and its autoubiquitylation activity[134]. RING and ZNF1 have extensive intramolecular interactions, and mutations in important residues were deleterious to its E3 activity[134]. Indeed, an alpha fold model of RNF138 shows similar overall intramolecular interactions, with a much longer linker region separating the N-terminal RING and ZNF from the C-terminal ZNFs, and UIM discussed extensively in **(Chapter 3)**.

RNF138 is a therapeutically relevant E3 ligase as it is overexpressed in glioma cells, a very aggressive type of brain cancer, promoting proliferation, metastasis, and radioresistance[135]. This is thought to be due to the activation of extracellular-signal regulated kinase (ERK) pathways, but the direct mechanism remains unclear[135]. What is clear though, is the down regulation of RNF138 in glioma cells leads to cancer cell apoptosis. More recently, a group mapped a possible pathway implicating RNF138 in the observed phenotype in glioma cells[127]. RNF138 ubiquitylates and degrades ribosomal proteins S3 (rpS3) which regulates the maturation and initiation of translation of elongation factors, and many more other cellular activities. One of the main functions of rpS3 is the interaction with TNF receptor type1-associated DEATH domain protein in response to radiation, inducing apoptosis and activation of kinases, and caspase activity[127]. In glioma cells, rpS3 was found to be down regulated, which is thought to be the driver of radioresistance. RNF138 is implicated in the degradation of the nuclear translocating rpS3 in response to radiation, which ultimately inhibits rpS3 mediated apoptosis[127]. RNF138 is further implicated in cisplatin resistance in gastric cancer cells, as its activity activates Chk1 signaling pathways[135]. Furthermore, the aberrant ubiquitylation of Cav2.1 channels present in neurons leads to aberrant protein degradation, which is associated with the cerebellar disease episodic ataxia type 2 (EA2)[123].

It is evident that there is an abundance of data highlighting the significant roles Ubc13 and RNF138 fulfill in the DNA damage response and cellular ubiquitin signaling. The findings presented in this thesis encompass numerous inquiries into the structure and functions of these pivotal proteins, leading to exciting insights and further questions to investigate.

## 1.10 Chapter 2 Study Overview and Aims

In chapter 2 we set out to study Ubc13, we aimed to answer the questions surrounding the gating activity of the protein, is it more dynamic than previously thought? Can we probe the active site loop with small molecule compounds to gain insights into the dynamics, helping our understanding Ubc13's unique E2 activity, and ultimately in the future open the way for rational drug design strategies. We uncovered a new C2 monoclinic crystal form of the unbound Ubc13-Mms2 heterodimer through batch crystallization at room temperature. Through a comparison of this new structure with previous structures, we propose that the active site loop of Ubc13 exhibits greater flexibility than previously anticipated, existing in multiple conformational states. To delve deeper into the adaptability of this active site, we investigated two newly identified compounds, NSC291068 and a 2-nitrofuran fragment. Each demonstrated a degree of specificity, with the NSC291068 showing evidence of covalent reactivity to the active site cysteine, while 2-nitrofuran interacts non-covalently in the binding pocket thought to be unique to Ubc13, H-bonding to N123. The dynamic behavior of Ubc13's active site is critical and should be considered in future studies aimed at designing next-generation inhibitors.

## 1.11 Chapter 3 Study Overview and Aims

In chapter 3, we aimed to unveil the molecular details underlying RNF138's DNA binding and E3 ligase activity in promoting HR. We aimed to address the following questions: 1) How does RNF138 specifically recognize DNA overhangs? 2) What roles do the individual functional domains play in direct DNA binding and E3 ligase activity? 3) What type of ubiquitin chains are synthesized on RNF138 with regards to its autoubiquitylation activity?

We initially set out to purify RNF138 from *Escherichia coli (E.coli)* and it was clear RNF138 was unstable *in vitro* proving to be challenging to express and purify. Previous studies purified RNF138 using insect cell approaches however this protein proved to be poorly behaved for biochemical studies. Instead, we set out to establish a mammalian cellular protein expression for RNF138 which potentially might supply key mammalian cell chaperones and appropriate PTMs for proper folding. However, we found that overexpressed RNF138 was toxic in mammalian cells. We therefore reverted to *E.coli* expression, to characterize the binding activity of RNF138 to DNA, by utilizing various protein truncations of RNF138. What became clear is RNF138 prefers DNA overhangs as substrates, and requires all the protein domains for specific DNA binding. Furthermore, RNF138 bound to DNA in a size dependent manner, preferring larger overhangs. An extensive look at RNF138's AlphaFold model reveals a "beads on a string" structure, with a canonical N-terminal RING domain, followed by ZNF1 with extensive intramolecular interactions between the domains, followed by a large linker, connecting two more ZNFs, and the UIM. The flexibility of RNF138 offers an explanation for its vast interaction network, and involvement in different cellular signalling pathways, and reveals possible PTM sites, required for its regulation and activity.

## 1.12 Chapter 4 Study Overview and Aims

In chapter 4 we continued looking at RNF138 activity, more specifically how RNF138 is regulated. Previously RNF138 has been shown to function in a cell cycle dependent manner post IR exposure, being active in S/G2, but not G1 phase. We predicted RNF138 activity is under cell cycle control coinciding with the preferential use of HR. In this study, we demonstrate that RNF138 undergoes phosphorylation at threonine residue 27 by cyclin-dependent kinase activity during the S and G2 phases of the cell cycle. Additionally, we observe constitutive ubiquitylation of RNF138, with some ubiquitylation occurring on lysine residue 158. Notably, the ubiquitylation of RNF138 decreases in response to genotoxic stress. Through mutation of RNF138 at residues

T27, K158, and the previously identified S124 ATM phosphorylation site, we reveal that post-translational modifications at all three positions play a role in mediating DNA double-strand break repair. Our findings provide insights into the functional mechanism and regulation of RNF138 by the cell during HR.

# Chapter 2 – Characterizing the Active Site of Ubc13

## 2.1 Abstract

Ubc13 is an E2 conjugating enzyme involved in DNA damage and proinflammatory signaling. Together with its partner protein Mms2/Uev1a, Ubc13 builds K63-linked polyubiquitin chains which play roles in cellular signaling. Ubc13 has shown promise to be a therapeutic target. Understanding its active site structure and dynamics is crucial to both understanding the mechanism of K63-linked polyubiquitin chain formation and for the development of specific inhibitors. Comparing the structures of different E2 enzymes highlights a unique conformation of Ubc13's active site. When aligned with UbcH5c, another well-studied E2 enzyme, Ubc13 shows high overall similarity, but there are differences in the conformation of the active site loop residues (114-124) near the active site cysteine. This difference is unique to Ubc13 making it an attractive therapeutic target. Here we determine a structure of Ubc13 crystallized under near native solution conditions that reveals a new active site conformation compared to previous structures which were crystallized at lower temperatures. The new structure suggests that the active site loop is more flexible than previously thought. To probe these dynamics, we characterized a novel covalent compound NSC291068 and a noncovalent fragment 2-nitrofuran that inhibit Ubc13 *in vitro*. NSC291068 inhibited wild type Ubc13, and a mutant with the active site loop in a conformation similar to other members of the E2 family suggesting it is a promiscuous compound. The 2-nitrofuran fragment showed specificity for Ubc13 at high concentrations, and its inability to inhibit mutant Ubc13 suggests specificity. The fragment offers a new avenue of noncovalent drug development. Understanding active dynamics of Ubc13 is crucial in developing tight binding specific inhibitors of Ubc13.

## 2.2 Introduction

Ubiquitylation is a common post-translational modification which regulates signal transduction pathways. Initially, ubiquitin (Ub) is activated by forming a thioester linkage between C-terminal carboxylate of Ub and the active site cysteine of an E1 in the presence of ATP[65,66,136]. The activated Ub is transferred to a ubiquitin conjugating enzyme (E2) at the active site cysteine, and with the help of an E3 ligase, the Ub-E2 complex recognizes and ubiquitylates specific protein targets[66,120].

Protein ubiquitylation can lead to diverse downstream effects which are driven in part by different forms of Ub chains that are linked to the substrate[66]. The K63-linked polyubiquitin (polyUb) chains are linked by an isopeptide bond between the amino group of K63 of one Ub and the C-terminal carboxylate of the next Ub. Among the ~35 E2 enzymes in the human genome[137,138], Ubc13 (UBE2N) is the only known E2 dedicated to the synthesis of K63-linked chains, and it does so with the help of its binding partner: an E2-like Ub enzyme variant protein Mms2[55,98,139,140]. The C-terminal tail of an activated Ub is transferred from an E1 enzyme to the active site cysteine of Ubc13 to form a thioester bond (the acceptor ubiquitin), whilst the Mms2 binds a second Ub (the donor Ub) and positions its K63 to attack the thioester bond to form K63-linked diubiquitin[55,88]. K63-linked polyubiquitin (polyUb) chains, synthesized by the Ubc13/Mms3 complex in the nucleus in collaboration the E3 ligase RNF8, play a crucial role in recruiting downstream effectors essential for the DNA damage response (DDR)[4]. In the cytoplasm, Ubc13 cooperates with Uev1A, a protein almost identical to Mms2. Cytoplasmic K63-linked polyUb chains generated by the Ubc13/Uev1A complex, in collaboration with TRAF6, promote NF-kB signaling pathways[74,97].

E2 enzymes contain a loop adjacent to the active site cysteine (the gating loop: residues 114-124 in Ubc13) that protects the active site from solvent and regulates access of Ub to the catalytic cysteine[105]. While this loop is found in all E2s, the Ubc13 gating loop adopts a different conformation compared to other E2s. However, once it is charged with Ub at C87, the gating loop shifts into the conformation observed in the other E2s, allowing the attack of the incoming donor K63 on the thioester-linked Ub[103,105]. More specifically, L121, which is well conserved within the family, occupies a position that would sterically block the attack of the acceptor ubiquitin K63 on the active site C87. In other E2 enzymes, and in the structure of Ubc13 when C87 is tethered to the C-terminus of a donor Ub, the gating loop adopts a conformation in which L121 reorients so that the acceptor K63 can access the active site C87. The gating mechanism is based on the hypothesis that two conformations exist and the dynamics between the two drive the specific synthesis of K63-linked ubiquitin chains[55,88,103].

The unique structural properties of the Ubc13 gating loop present the possibility for the selective inhibition of Ubc13 by small molecule inhibitors[89,100]. The small molecule NSC697923 was uncovered in a high-throughput cell-based assay to identify inhibitors of NF-kB signaling and was shown to inhibit Ubc13 but not other E2s in biochemical ubiquitylation assays[100,116]. The mechanism of this inhibition was shown to involve the covalent modification of the active C87 by the compound, dependent upon the presence of a pocket formed by the gating loop[100]. A mutant of Ubc13 harbouring 4 amino acid substitutions ($\mathrm{Ubc13^{QD}}$) which stabilized the gating loop in the standard conformation and thus lacked the NSC697923-binding pocket was shown to be resistant against NSC697923 inhibition in biochemical assays[100]. In addition, cells in which wild type Ubc13 ($\mathrm{Ubc13^{WT}}$) was replaced with $\mathrm{Ubc13^{QD}}$ were significantly resistant to inhibition of both NF-kB and DNA damage signaling by NSC697923. Taken together, this information indicates that the uniquely structured gating loop of Ubc13 could provide a route for the selective inhibition of this critical enzyme.

Ubc13's active site loop has unique gating functions. We hypothesize there is either a transitional state conformation between the unbound and bound state, or the loop is flexible and occupies different conformations at physiological temperature. In this study we determined a new C2 monoclinic crystal form of the unbound Ubc13-Mms2 heterodimer through batch crystallization at room temperature. The active site loop shows a conformational deviation from the existing structures of Ubc13 in either the ubiquitin bound or unbound forms. We suggest that Ubc13's active site loop is more flexible than previously thought and exists in multiple conformational states. To further probe the plasticity of the active site, we characterized two novel compounds, NSC291068 and 2-nitrofuran fragment. NSC291068 reacts with the active site cysteine specifically in a pH dependent reaction. The compound inhibited ubiquitin chain formation and inhibited both the Ubc13$^{WT}$ and Ubc13$^{QD}$. Surprisingly, the 2-nitrofuran fragment inhibited polyUb chain formation of Ubc13$^{WT}$, but the mutant was resistant to the fragment suggesting specificity to Ubc13's unique active site.

NSC291068 shows promise as an inhibitor of E2s, but not specifically Ubc13. The 2-nitrofuran fragment shows promise as a specific inhibitor of Ubc13, taking advantage of unique active site characteristics of Ubc13. The active site dynamics of Ubc13 are crucial and should be considered in the future when designing next generation inhibitors.

## 2.3 Materials and Methods

### 2.3.1 Protein Constructs and Expression

All the proteins utilized in our study were cloned in pGEX-6p1 vector, with a GST tag to the N-terminus of the protein of interest. *E. coli* BL21 Gold cells were transformed with all the constructs, and colonies were selected with ampicillin and kanamycin.

### 2.3.2 Bacterial Cell Growth and Induction

Transformed colonies were selected and inoculated in 5 mL of Luria Broth (LB) supplemented with 5 µL of 50 mg/mL Kanamycin (50 µg/mL) and 5 µL of 100 mg/mL ampicillin (100 µg/mL) and incubated overnight at 37 °C. The starter culture was subsequently transferred to 500 mL of LB broth containing 500 µL each of ampicillin and kanamycin and incubated for at least 2-4 hours at 37 °C. Next, 25 mL of the 500 mL culture was transferred to 8 separate bottles, each containing 1 L of LB broth supplemented with 1 mL of kanamycin and 1 mL of ampicillin (same final concentrations as above). These 1 L preparations were then incubated at 37 °C for an additional 2-4 hours. The optical density at 600 nm was measured to determine whether induction was appropriate, with a target optical density of 0.6-0.7. Cells were induced by adding 200 µL of 1 M IPTG (200 µL of 0.5 M $ZnCl_2$ supplement for RNF8). The preparations were then incubated overnight at 17 °C. The cells were pelleted by centrifugation at 4000 RPM (JLA-8.1 rotor) for 15 minutes at 4 °C. After discarding the supernatants, the pellet was frozen using liquid nitrogen and stored at -80 °C.

### 2.3.3 Protein Production and Purification

The pelleted cells, whether freshly grown or retrieved from -80 °C storage, were resuspended in lysis buffer (20 mM HEPES pH 6.8, 400 mM NaCl, 1 µL/mL BME, and 10 µM $ZnCl_2$ only for RNF8), along with 200 µL of 1X protease inhibitor solution (Sigma Aldrich). Resuspension was done at 4 °C with a magnetic stir bar for approximately 35 minutes. After resuspension, cells were lysed by sonication at 70% amplitude, with 15-second pulses followed by 45-second pause intervals, totaling 3 minutes of sonication. The lysate was centrifuged at 13000 RPM (JA-17 Rotor) for 40 minutes at 4 °C, and the pellet was discarded. The supernatant was collected and incubated with Glutathione Sepharose Beads (Sigma Aldrich) for 1 hour, then transferred to a falcon tube and centrifuged at 2000 RPM (JA-7.5) for 2 minutes at 4 °C. After discarding the supernatant, the beads were rinsed with lysis buffer, and the centrifugation step was repeated. The beads were then transferred back to a column and washed with lysis buffer. GST-tagged constructs were eluted with 10 mL elution buffer (20 mM HEPES pH 6.8, 400 mM NaCl, 1 µL/mL BME, 10 mM glutathione, and 10 µM $ZnCl_2$ only for RNF8). Four 10 mL fractions were collected and pooled. The pooled fractions were digested with Human Rhinovirus 3C Protease (HRV-3C) overnight, and the volume of the 50 mL overnight digests was concentrated to 1-2 mL using 10 kDa centricon concentrator tubes (Sigma Aldrich). The proteins were then further purified by size exclusion chromatography using a Superdex 75 16/60 column, with buffers adjusted for each protein (20 mM TRIS HCl pH 7.5, 150 mM NaCl, 1 mM TCEP for Ubc13 and Mms2; 20 mM TRIS HCl pH 7.5, 150 mM NaCl, 10 µM $ZnSO_4$, and 1 mM DTT for RNF8). Protein elution was monitored at 280 nm, and fractions corresponding to peaks were collected and pooled. The success of each purification step was confirmed by running samples on a 5-15% polyacrylamide gel, which was subsequently stained with Coomassie blue dye. All proteins were quantified using an absorbance assay at 280 nm utilizing the extinction coefficient of each protein, with guanidinium hydrochloride. Ubc13 and Mms2, were purified separately, and combined at 1:1 concentration. The complex was subjected to another round of size exclusion chromatography.

### 2.3.4 Crystallization:

Ubc13-Mms2 heterodimer was concentrated to 12 mg/mL and stored in 50 mM TRIS pH 7.5, 150 mM NaCl, 1 mM TCEP at 4 °C, and then slowly equilibrated to room temperature. Crystals were observed after 16 hours. Data was collected on a home source and refined using Python-based Hierarchical Environment for Integrated crystallography (PHENIX)[141], and the active site loop was constructed using Crystallographic Object-Oriented Toolkit (COOT)[142].

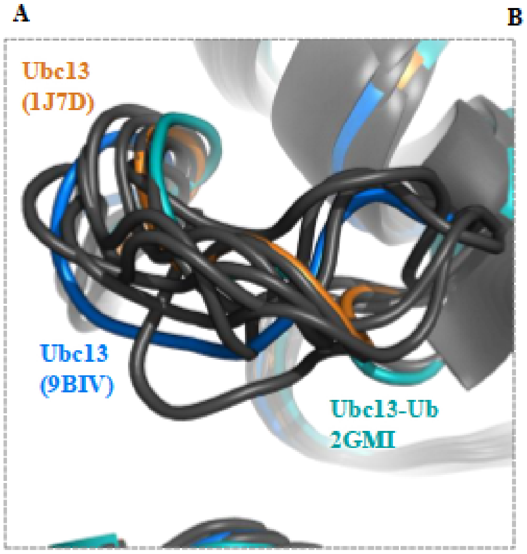
Protein modeling and analysis was done using PyMOL. Structural coordinates were submitted to the Protein Data Bank (PDB) accession code: 9BIV **(Table 2.1)**.

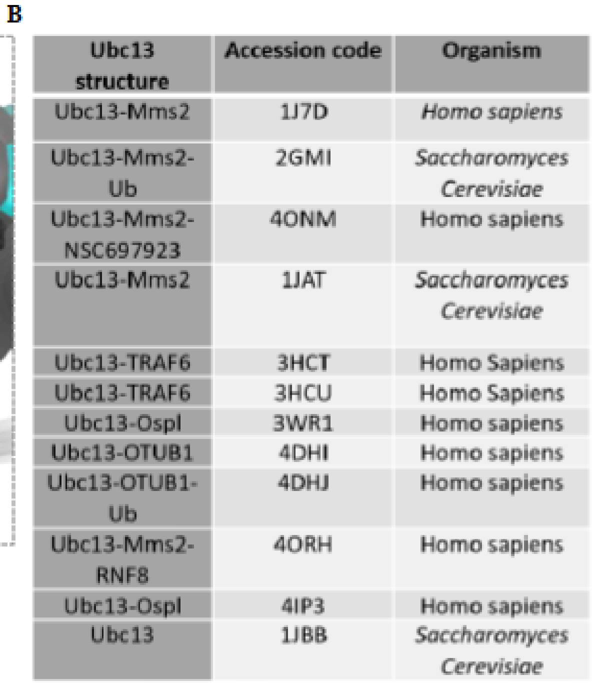
**Table 2.1. Data Collection and Refinement Statistics of the New Structure of Ubc13.**

| **Data Collection** | |
|---|---|
| Space Group | C 1 2 1 |
| Cell Dimensions | |
| a, b, c (Å) | 86.361, 42.942, 93.79 |
| α, β, γ (°) | 90, 108.37, 90 |
| Resolution (Å) | 50.0 - 1.68 (1.72-1.68) |
| $R_{meas}$ | 0.286 |
| I/σ (I) | 4.43 (1.68) |
| Completeness (%) | 99.9 (99.4) |
| Redundancy | 7.6 (3.9) |
| $CC_{1/2}$ | 0.94 |
| **Refinement** | |
| Resolution (Å) | 42.68-1.68 |
| Reflections used in refinemnet | 37360 (2564) |
| Rwork | 0.1458 (0.1868) |
| Rfree | 0.1781 (0.2042) |
| Number of non-hydrogen atoms | 2735 |
| Macromolecules | 2293 |
| Ligands | 0 |
| Solvent | 442 |
| Protein residues | 288 |
| Nucleic acid bases | 0 |
| B-factor | 21.83 |
| RMS bond length | 0.004 |
| RMS bond angles | 0.72 |
| Ramachandran favored (%) | 98.59 |
| Ramachandran allowed (%) | 1.41 |
| Ramachandran outliers (%) | 0 |
| Rotamer outliers (%) | 0.4 |
| Clashscore | 1.96 |

### 2.3.5 Structural analysis of Ubc13 gating loop diversity in PDB entries.

All Ubc13 structures with resolutions below 3 Å were downloaded from the PDB and analyzed using PyMOL. **(Figure 2.1A,B)**. A conformational similarity matrix for the gating loop for all the structures was determined by aligning all structures against all using all C-alpha atoms except those in the gating loop (A114-D124). The average RMSD for all C-alpha within the gating loop was then determined using the measurement tool in PyMOL for all of the structure pairs. The all-in-all correlation matrix between the different average RMSD values were plotted in Biovinci.





| Ubc13 structure | Accession code | Organism |
|---|---|---|
| Ubc13-Mms2 | 1J7D | *Homo sapiens* |
| Ubc13-Mms2-Ub | 2GMI | *Saccharomyces Cerevisiae* |
| Ubc13-Mms2-NSC697923 | 4ONM | Homo sapiens |
| Ubc13-Mms2 | 1JAT | *Saccharomyces Cerevisiae* |
| Ubc13-TRAF6 | 3HCT | Homo Sapiens |
| Ubc13-TRAF6 | 3HCU | Homo Sapiens |
| Ubc13-Ospl | 3WR1 | Homo sapiens |
| Ubc13-OTUB1 | 4DHI | Homo sapiens |
| Ubc13-OTUB1-Ub | 4DHJ | Homo sapiens |
| Ubc13-Mms2-RNF8 | 4ORH | Homo sapiens |
| Ubc13-Ospl | 4IP3 | Homo sapiens |
| Ubc13 | 1JBB | *Saccharomyces Cerevisiae* |

**Figure 2.1. Alignment of Novel Loop with All Existing Ubc13 Yeast and Human Structures.**

**A.** The colored active site loops correspond to the ones compared in this study. The remaining are to show the conservation of the conformation of the loop with or without substrate binding, Mms2, or E3 ligase binding. **B.** All structures of Ubc13 in yeast and in humans used in the alignment in panel A, with their PDB accession codes.

### 2.3.6 Thermal Shift Assay

360 µM of Ubc13$^{WT}$, Ubc13$^{C87S}$ or Ubc13$^{QD}$ were incubated with 5x and 10x molar excess of inhibitors (NSC291068, NSC291057, 2-nitrofuran) and then diluted to 130 µM with 500x of SYPRO orange dye, to a final concertation of 250x. 5 µL of the reacted protein and dye solution was added to 15 µL of 20 mM Tris pH 7.5, 150 mM NaCl, 1 mM TCEP. The solutions were

prepared in triplicate in 96 well plates. The temperature was increased from 20 °C to 95 °C degrees in 0.5 °C increments over 45 minutes, and all sample read outs were done in triplicate using MasterCycler RealPlex (Eppendorf) filter set at 550 nm. Readouts are fluorescence as a function of temperature, and the melting temperature is inferred from the inflection point.

### 2.3.7 Absorbance assay to monitor covalent addition of inhibitors to Ubc13

100 µM of Ubc13 or Ubc13 C87S mutant were incubated with 100 µM of the inhibitor in 50 mM HEPES, 75 mM NaCl. The absorbance at 420 nm was monitored over 15 minutes as a measure of the degree of reaction, and each measurement was done in triplicate. A pH series was determined over a range of pH 6.0 – 9.0.

### 2.3.8 Mass Spectroscopy

Matrix- assisted laser desorption/ionization (MALDI) analysis was conducted to measure the molecular weight of protein with or without compounds. The samples underwent a tenfold dilution in a solution comprising 50% acetonitrile and water with 0.1% trifluoroacetic acid. Subsequently, 1 µL of each diluted sample was mixed with 1 µL of sinapinic acid solution (10 mg/mL in 50% acetonitrile/water + 0.1% trifluoroacetic acid). The resulting sample/matrix solutions were then spotted onto a stainless-steel target plate and allowed to air dry. Mass spectra were acquired using an Autoflex Speed MALDI-ToF mass spectrometer (Bruker Daltonik, Bremen, Germany) equipped with a Smartbeam-II laser operating at a frequency of 2 kHz. Ion analysis was conducted in positive mode, and external calibration was performed using a standard protein mixture.

### 2.3.9 Ubiquitylation Inhibition Assay

200 nM of Ubc13 was pre-incubated with 0, 1, 2, 4, 6, 8, 10, 15 µM inhibitor or higher concentration in the case of the 2-nitrofuran (0, 40, 80, 100, 200, 400, 600, 800 µM) for 10 minutes

on ice. The remaining components were added in the following final concentrations to a final volume of 50 µL in 50 mM TRIS-HCl pH 7.5, 100 mM NaCl, 10 mM $MgSO_4$ and 1 mM $ZnCl_2$: 500 nM E1, 1 µM RNF8, 200 nM Mms2, 50 µM ubiquitin and 2 mM ATP. The reactions were incubated at 37 °C for 90 min. Reactions were quenched with 2X SDS-PAGE loading buffer and loaded on a 16% polyacrylamide gel. The proteins were transferred to an Immobilon PVDF membrane, and blocked overnight at 4 °C. The membrane was probed using mouse anti-ubiquitin IgG (Santa Cruz SC-8017) for an hour at room temperature. Membranes were washed and incubated with a 1:5000 dilution of anti-mouse linked horse radish peroxidise (Abcam), and the chemiluminescence reaction was activated using the Clarity ECL substrate (BIO-RAD ab205719). Western blots were imaged using ImageQuant LAS 4000, and membranes were exposed for 1 minute.

## 2.4 Results

### 2.4.1 New Active Site Loop has Altered Hydrogen Bonding Interactions

To probe the conformational states of Ubc13-Mms2, we sought out new crystallization conditions for the complex. Our previous structural work was performed with protein that was stored at 4 °C immediately prior to crystallization with PEG precipitant[98]. We discovered that complex incubated for ~24 hr at room temperature could be batch crystallized without added precipitant in a new C2 crystal form. Determination of the structure of this form by molecular replacement at 1.7 Å resolution revealed a new conformation for the gating loop **(Figure 2.2A,B)** that is distinct from either the previous structure crystallized in the absence of ubiquitin **(Figure 2.2C)**, or the structure in the presence of Ub **(Figure 2.2D).** A more detailed comparison of the structures of the previous unbound structure of Ubc13 (1J7D) with the Ub-conjugated form (2GMI) and the new structure (9BIV), suggests that the new structure may represent an intermediate between 1J7D and the Ub-bound state **(Figure 2.3)** In 1J7D, the gating loop adopts a conformation stabilized by two adjacent type 1b turns, stabilized by internal hydrogen bonding between N79 and turn 1, and N123 and the interior of the protein, resulting in L121 lying above the active site C87 **(Figure 2.3B)**. In the new structure, turn 1 rotates away from N79, and turn 2 is broken, adopting a more extended conformation **(Figure 2.3C)**. In spite of these changes, L121 remains in the "up" conformation and N123 is internally hydrogen bonded, similar to 1J7D. With the binding of Ub, the conformation changes again, likely triggered by the interaction of N123 with the incoming K63 of the donor Ub **(Figure 2.3A,D)**. This rotation of N123 is coupled to the rotation of L121 so that it no longer occludes the active site cysteine and instead packs against the body of the protein. Like the new crystal form, the region 119-122 adopts an extended conformation and does not adopt the turn observed in 1J7D.

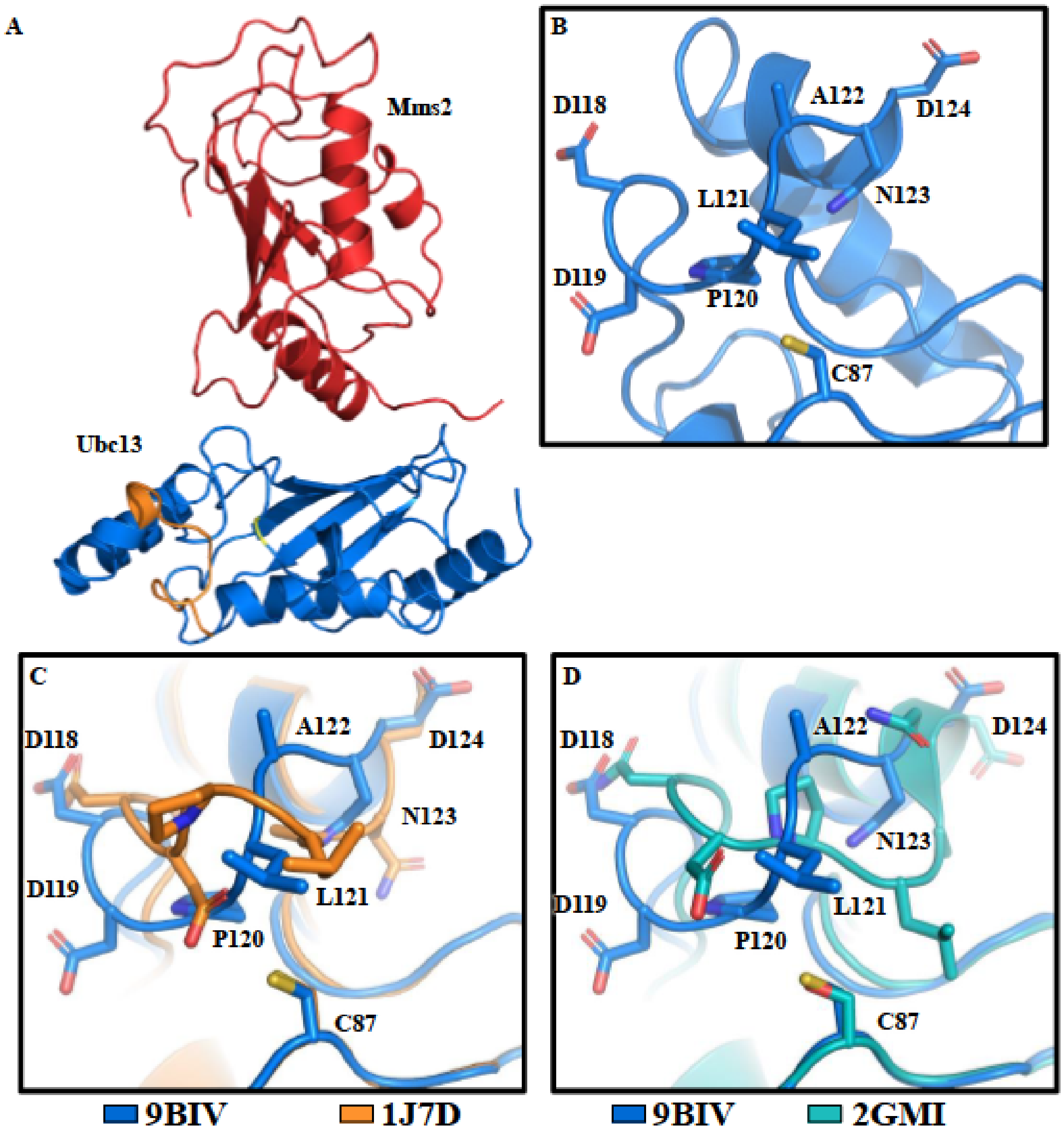


**Figure 2.2. New Ubc13 Structure Reveals New Active Site Loop Conformation.**

**A.** Structure of the heterodimer Ubc13 (blue) - Mms2 (red) in the new crystallographic form C2 monoclinic. Orange denotes the active site loop (A114-D124). Yellow denotes active site cysteine. **B.** Zoomed in active site of new crystallographic form. **C.** Alignment of the novel structure with orange: Ubc13$^{WT}$(1J7D). **D.** Alignment of novel structure with cyan: Ubc13 (2GMI) bound to Ub.

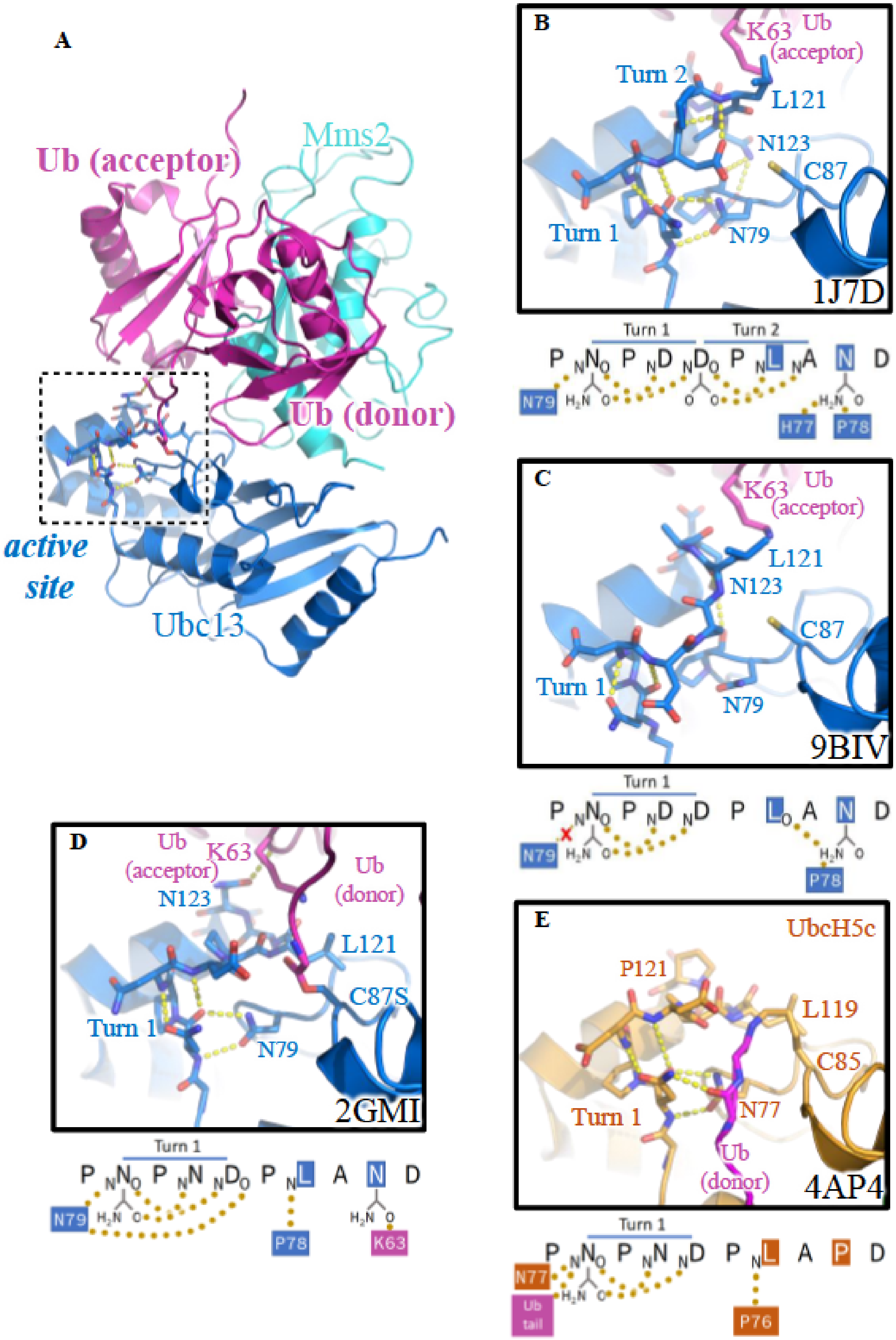

A
Ub (acceptor)
Mms2
Ub (donor)
active site
Ubc13
B
K63
Ub (acceptor)
Turn 2
L121
N123
C87
Turn 1
N79
1J7D
Turn 1
Turn 2
P N P D D P L A N D
N79
H77
P78
C
K63
Ub (acceptor)
L121
N123
C87
Turn 1
N79
9BIV
Turn 1
P N P D D P L A N D
N79
P78
D
Ub (acceptor)
K63
Ub (donor)
N123
L121
C87S
Turn 1
N79
2GMI
Turn 1
P N P N D P L A N D
N79
P78
K63
E
UbcH5c
P121
L119
C85
Turn 1
N77
Ub (donor)
4AP4
Turn 1
P N P N D P L A P D
N77
Ub tail
P76

**Figure 2.3. Analysis of Active Site Loop Hydrogen Bonding patterns Between Different States of Ubc13 and UbcH5c.**

**A.** Ub (purple) conjugated structure of Ubc13/Mms2 (blue, teal respectively) with the active site highlighted (PDB:2GMI). **B.** Active site loop of the inactive form of Ubc13, with the L121 pointing upwards blocking the donor Ub from accessing C87. The hydrogen bonding pattern in the turns near the active site loop is shown below (PDB: 1J7D). Notably, the active site loop consisting of two turns, are held together by an extensive network of hydrogen bonds between active site loop residues, and nearby residues, N79, H77, and P78. **C.** Active site loop of the new form of Ubc13, crystallized in the C2 monoclinic space group (PDB:9BIV). L121 is still in the "up" conformation blocking access to the active site C87, but the hydrogen bonding pattern is altered in the residues near the active site loop. Notably, N79 is no longer hydrogen bonded to N119. Additionally, L121 is shifted and is hydrogen bonded to N123. **D.** The structure of Ubc13 active site conjugated to Ub. L121 moves into the canonical conformation found in other E2s allowing for the donor Ub to access C87. The acceptor Ub is directly engaged by N123 which is thought to be the driver of the conformational change. **E.** The structure of the active of UbcH5c conjugated to Ub. The hydrogen bonding with the Ub tail occurs with N119, as UbcH5c does not bind Mms2, and does not form K63-linked chains, therefore N123 is not engaged. Overall, the hydrogen bonding activity resembles the activated form Ubc13 (2GMI).

### 2.4.2 New Active Site Loop Conformation Points to More Flexibility in the Ubc13 Active Site

To better document the range of conformations of the Ubc13 gating loop, we performed an all against all comparison using all published PDB structures of Ubc13 meeting the resolution cutoff of 3 Å **(Figure 2.4)**. All the structures were aligned against each other and C-alpha-C-alpha distances within the Ala114-Asp124 gating loop were measured and averaged for each structure pair **(Figure 2.4A,B)**.The structural differences are displayed in a correlative heat map in which a positive correlation indicates the structures are more similar, and a negative value indicates the two structures diverge from one another **(Figure 2.4C)**. This analysis reveals that most of the previously determined structures of Ubc13 determined without a covalently tethered Ub or other modifying binding partner form a family in which L121 guards the active site cysteine and this conformation is conserved between human and *S. cerevisiae* Ubc13. It has been reported that the apo conformation of the active site loop observed in the crystal structure represents a ground state and once Ub is introduced the conformation of the loop changes to the active state as observed in 2GMI[55,103,109,110]. As expected, the Ub bound form of Ubc13, as seen in the matrix, highly correlates across other ubiquitin bound structures where the ubiquitin C-terminal tail stabilizes the loop, holding it in place (Figure 2c). E3 binding appears not to greatly impact the gating loop; structures of the E3 TRAF6 bound to Ubc13 resemble the Ubc13 free conformation, while structures of the E3s TRIM21 and TRIM26 bound to Ubc13~Ub resemble the Ubc13~Ub conformation determined in the absence of an E3. However, conformations of Ubc13 determined with the inhibitory factors OTUB1 or OspI show intermediate conformations. This analysis furthermore reveals that the new C2 form of Ubc13 is distinct from the Ubc13~Ub forms and it is more similar to the non-ubiquitylated forms of Ubc13, including those structures bound to OTUB1, OspI, or TRAF6. Taken together, this evidence suggests that the active site loop is more flexible than previously thought to be. As seen in the matrix, static shots of every previously deposited Ubc13 structure vary to some degree **(Figure 2.4C)**. Structures with more than one Ubc13 in the asymmetric unit differ from one another. Different factors binding to Ubc13 induce diverse conformations within the loop as seen with various binding partners such as TRAF6 and the TRIM proteins. We hypothesize the lack of precipitant and crystal equilibration at room temperature yielded an unbiased crystallographic structure in a more dynamic system, allowing the loop to sample physiological conformations, yielding a perhaps more physiologically relevant structure.

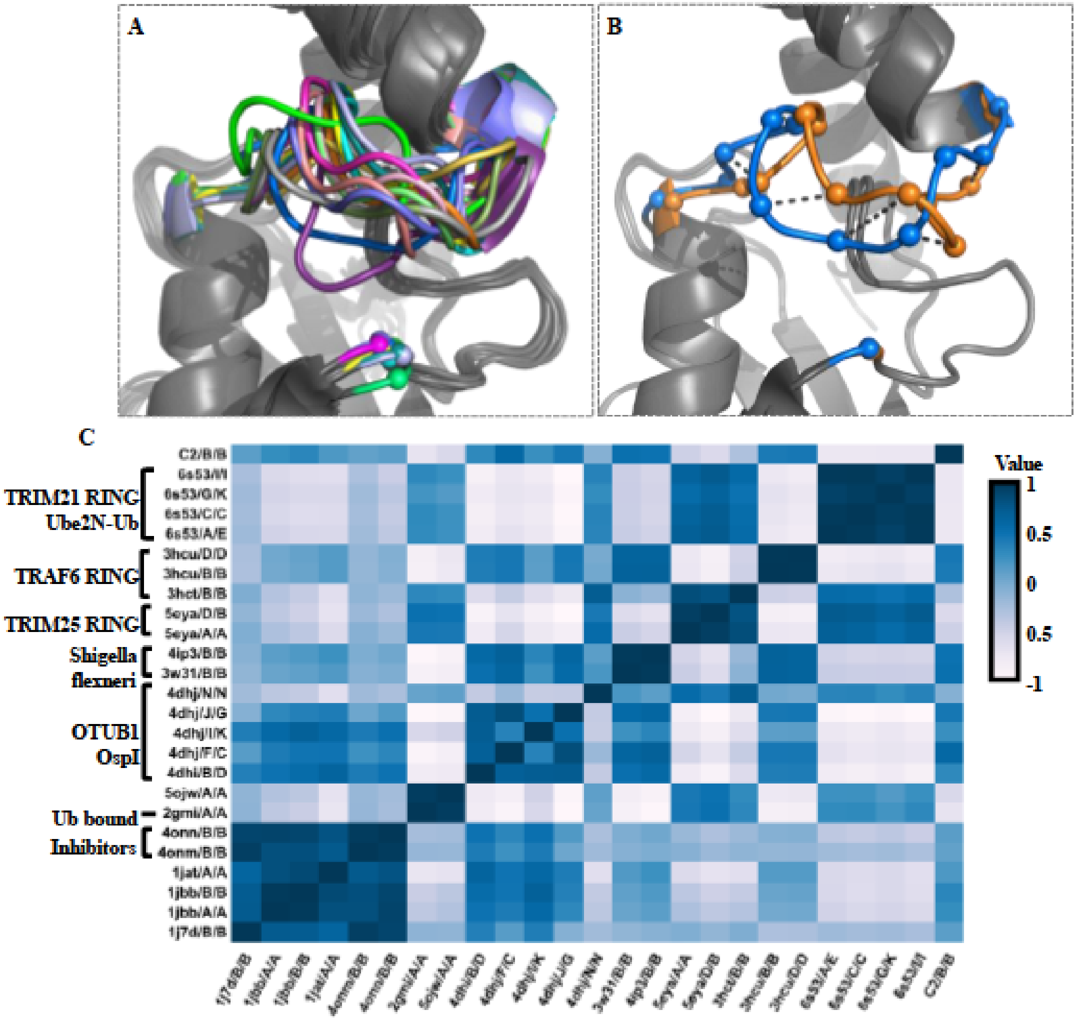


**Figure 2.4. All Against All Comparison of Ubc13 Structures Reveals a Highly Flexible System.**

**A.** All available high-resolution structures aligned in PyMOL. **B.** Representative image of how the measurements were carried out in PyMOL. **C.** Correlative matrix generated in Biovinci of the all against all comparison of the active site residues with darker blue denoting stronger correlation, and white denoting negative correlation. The various bound forms are denoted - the remainder of PDB accessions codes are apo forms of Ubc13.

### 2.4.3 New Loop Conformation is Not a Crystal Packing Artifact

To determine if the conformations of the gating loops were influenced by crystal packing, we compared the crystal environments of the different crystal forms (**Figure 2.5**). This analysis revealed that the gating loop in human Ubc13 (1J7D) approaches an Mms2 symmetry mate, which may bias its conformation (**Figure 2.5**), however the loop is not in a tight packing.

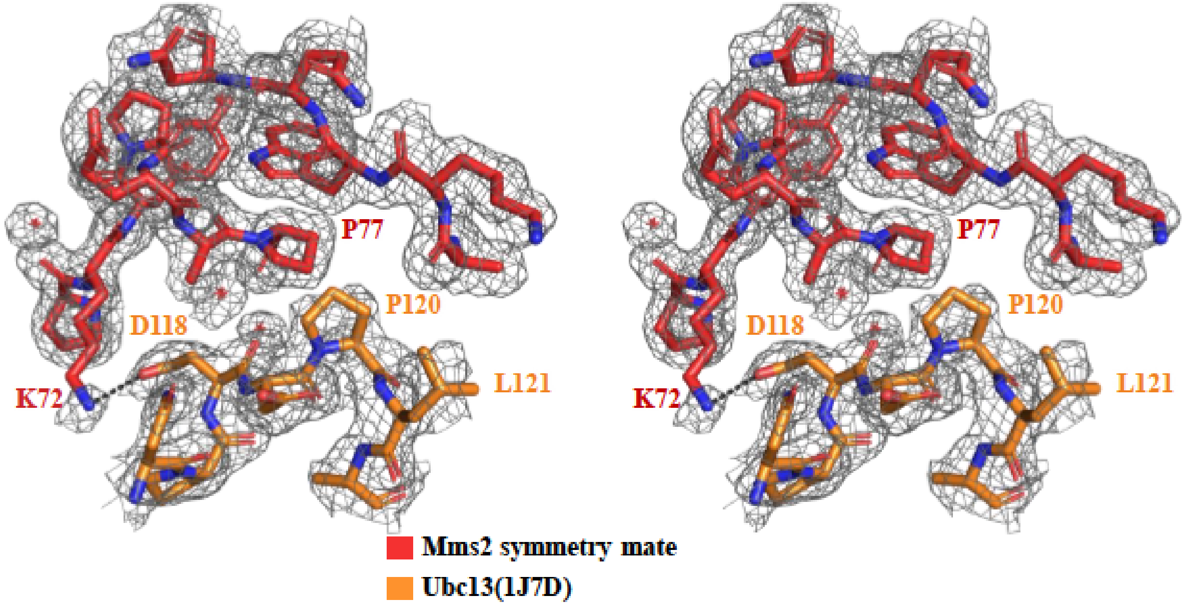


**Figure 2.5. Electron Density Maps of the Crystal Packing Environment of Ubc13 (1J7D).**

Cross-eyed display of the active site loop (orange) of Ubc13 in 1J7D, with its electron density map (grey mesh). The symmetry mate, Mms2 from a different heterodimer within the crystal is shown in red. Waters are represented as red asterisk.

The active site of the new crystal form is well resolved (**Figure 2.6A**), and in contrast to 1J7D, there was no evidence of crystal packing on the loop conformation. The loop is farther away from its symmetry mate, and water interrupts any possible interactions (**Figure 2.6B**).

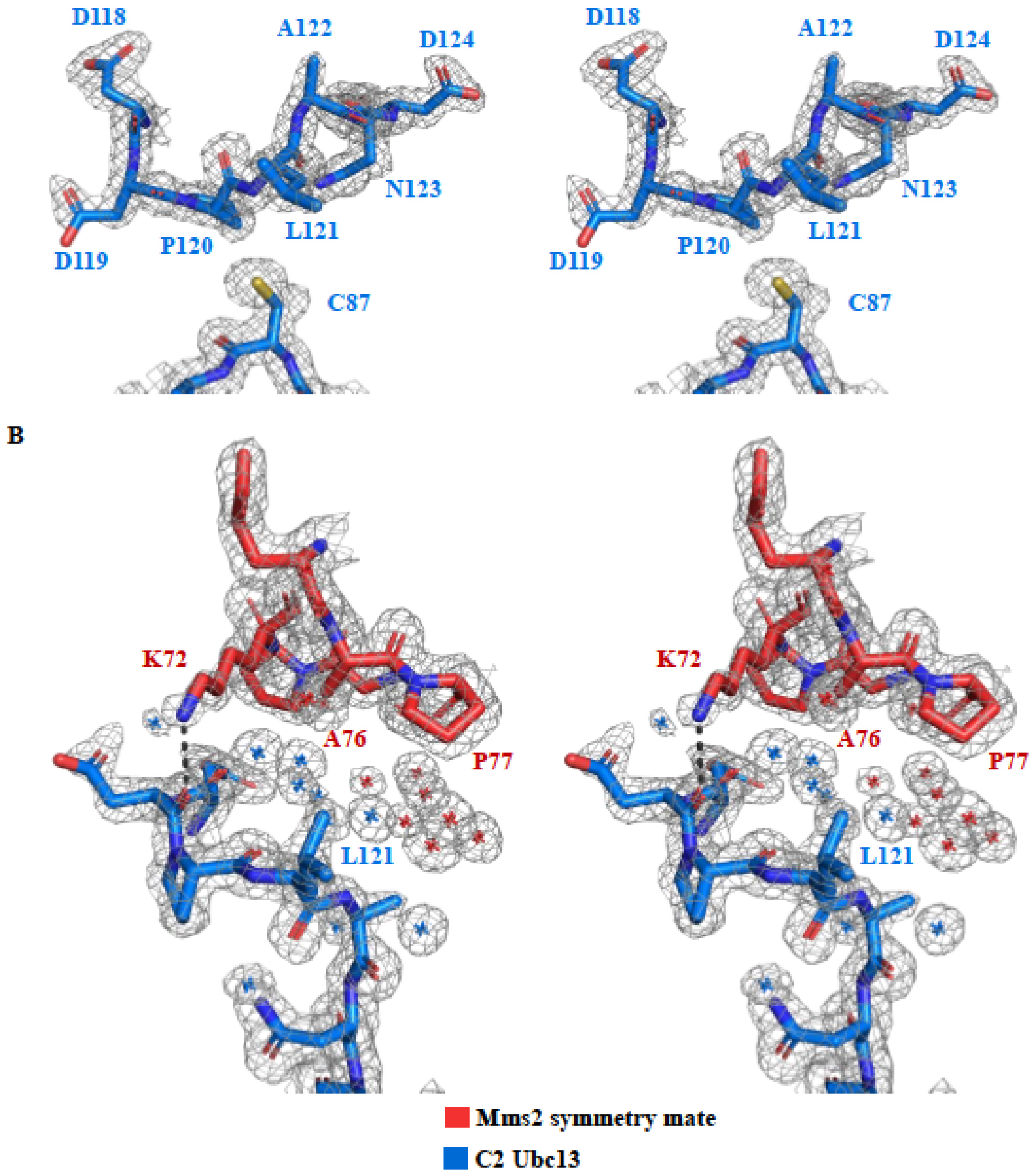


**Figure 2.6. Electron Density Maps of the Newly Fitted Loop and the Crystal Packing Environment.**

**A.** The newly built model with the 2Fo-Fc map in grey, and constructed cartoon model in blue at 1.7 Å displayed in cross eyed stereo. **B.** Cross-eyed display of the active site loop (blue) in proximity with a symmetry mate from a different heterodimer, Mms2 (red), and the respective electron density maps. Waters from each respective model are also displayed. A possible electrostatic interaction is highlighted with a dark grey dashed line. Waters are represented in asterks Ubc13 (blue), and Mms2 symmetry mate (red)

It is important to note that crystal packing involving the loop was not observed in the yeast isoforms (1JBB and 1JAT), however the structure of the active site loop in these yeast structures are very similar to the 1J7D human isoforms. Additionally, the loop was not closely packed with symmetry mates in the Ub conjugated structure of Ubc13 (2GMI) **(Figure 2.7A)**. We suggest that the different loop conformations observed in these structures likely arises from inherent conformational preferences rather than crystal packing effects. Furthermore, the lack of precipitant and crystal equilibration at room temperature used to determine our new structure suggests that this structure might represent more physiologically relevant structure of Ubc13 in its ground state.

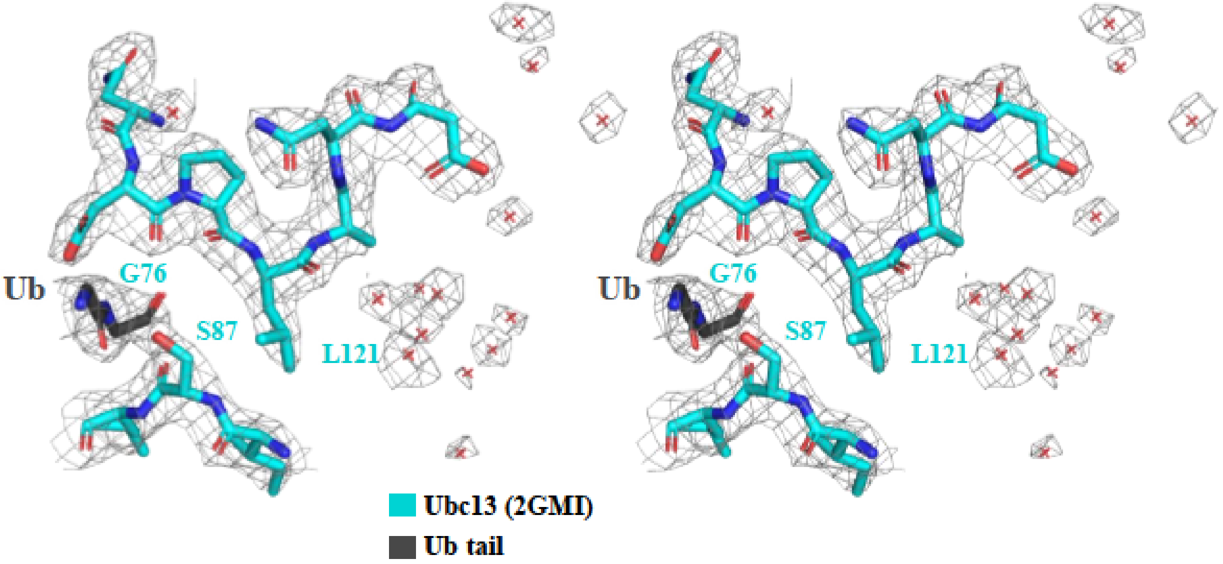
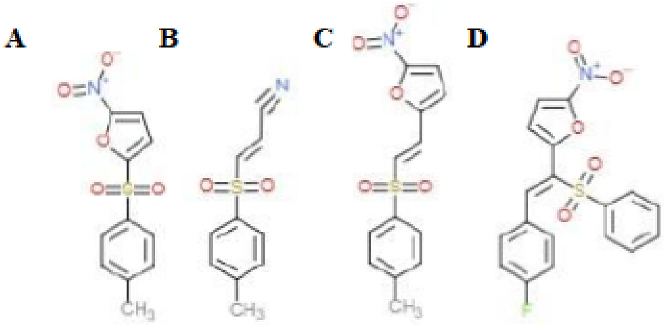


**Figure 2.7. Electron Density Maps and Crystal Packing Environment of Ub bound Ubc13 (2GMI).**

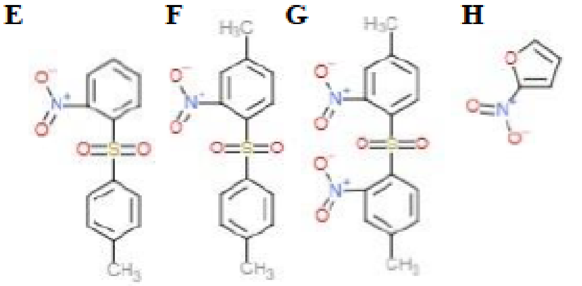
Cross-eyed display of the crystal packing condition. No symmetry mate within proximity was observed. The tail of Ub tail is seen in dark grey. Waters are represented with red asterisks.

### 2.4.4 Compound Binding Stabilizes Ubc13

The critical roles of Ubc13 in NF-kB and DNA damage signaling has made this E2 a particularly interesting target for therapeutic development[89] and a nitrofuran-based small molecule Ubc13 inhibitor has been discovered that covalently modifies the active site cysteine[100,116]. Interestingly, the active site loop plays a key role in the specificity of compounds targeting Ubc13. Our finding that the Ubc13 gating loop can occupy an additional conformation in the new crystal structure demonstrates that this loop is even more flexible than previously expected. This additional flexibility suggests that Ubc13 could be reactive to a larger range of Michael addition compounds than only NSC697923 (**Figure 2.8**). Therefore, we set out to probe the active site with novel inhibitors.



**NSC697923 BAY 11 -7082 NSC291068 NSC291057**



**NSC627708 NSC646124 NSC656835 2-Nitrofuran**

**Figure 2.8. Previously Characterized Compounds that Inhibit Ubc13.**

**A.**NSC697923 **B.** BAY 11-7082. **Compounds utilized in our study C.** NSC291068, **D.** NSC291057, **E.** NSC627708, **F.** NSC646124, **G.** NSC656835 and **H.** 2-nitrofuran.

NSC291068 (**Figure 2.8C**) and NSC291057 (**Figure 2.8D**), were reported to inhibit Ubc13 in a cellular based assay, similar to the previously published compounds NSC697923 and BAY 11-7082[100,116] **(Figure 2.8A,B)**. A common feature seen in all compounds, except for BAY11-7082, is a 2-nitrofuran moiety. We utilized a thermal shift assay to probe the ability of these compounds to affect the folding stability of wildtype Ubc13 ($\text{Ubc13}^{\text{WT}}$). This assay relies on the binding of the hydrophobic dye, SYPRO orange to hydrophobic residues within the protein that become exposed as temperature is increased and the protein unfolds allowing for the measurement of the melting temperature of the protein ($T_m$)[143]. An increase in the $T_m$ of a protein upon interaction with ligand suggests a favourable interaction and stabilization of the protein fold[144]. In the case of Ubc13, its intrinsic melting temperature in the presence of 2% DMSO is 55.4°C **(Figure 2.9A)**. Pre-reacting $\text{Ubc13}^{\text{WT}}$ with an excess of NSC291068 led to a 4.5 °C increase in $T_m$ to 59.9 °C (**Figure 2.9A**), suggesting that the compound stabilizes Ubc13. This shift is not observed in $\text{Ubc13}^{\text{C87S}}$, consistent with a covalent modification of the active site cysteine **(Figure 2.9B)**. In the case of NSC291057, no shift was observed in either $\text{Ubc13}^{\text{WT}}$ or the $\text{Ubc13}^{\text{C87S}}$ **(Figure 2.9C,D)**. Interestingly, when a 10-fold excess of 2-nitrofuran was incubated with Ubc13, an increase in $T_m$ of 1.4 °C was observed **(Figure 2.9E)**. This shift was not observed when 2-nitrofuran was incubated with the $\text{Ubc13}^{\text{QD}}$ mutant **(Figure 2.9F)**. The resistance of the $\text{Ubc13}^{\text{QD}}$ mutant to the compound suggests that the nitrofuran is specific to the Ubc13 active site groove structure, and not to other E2s.

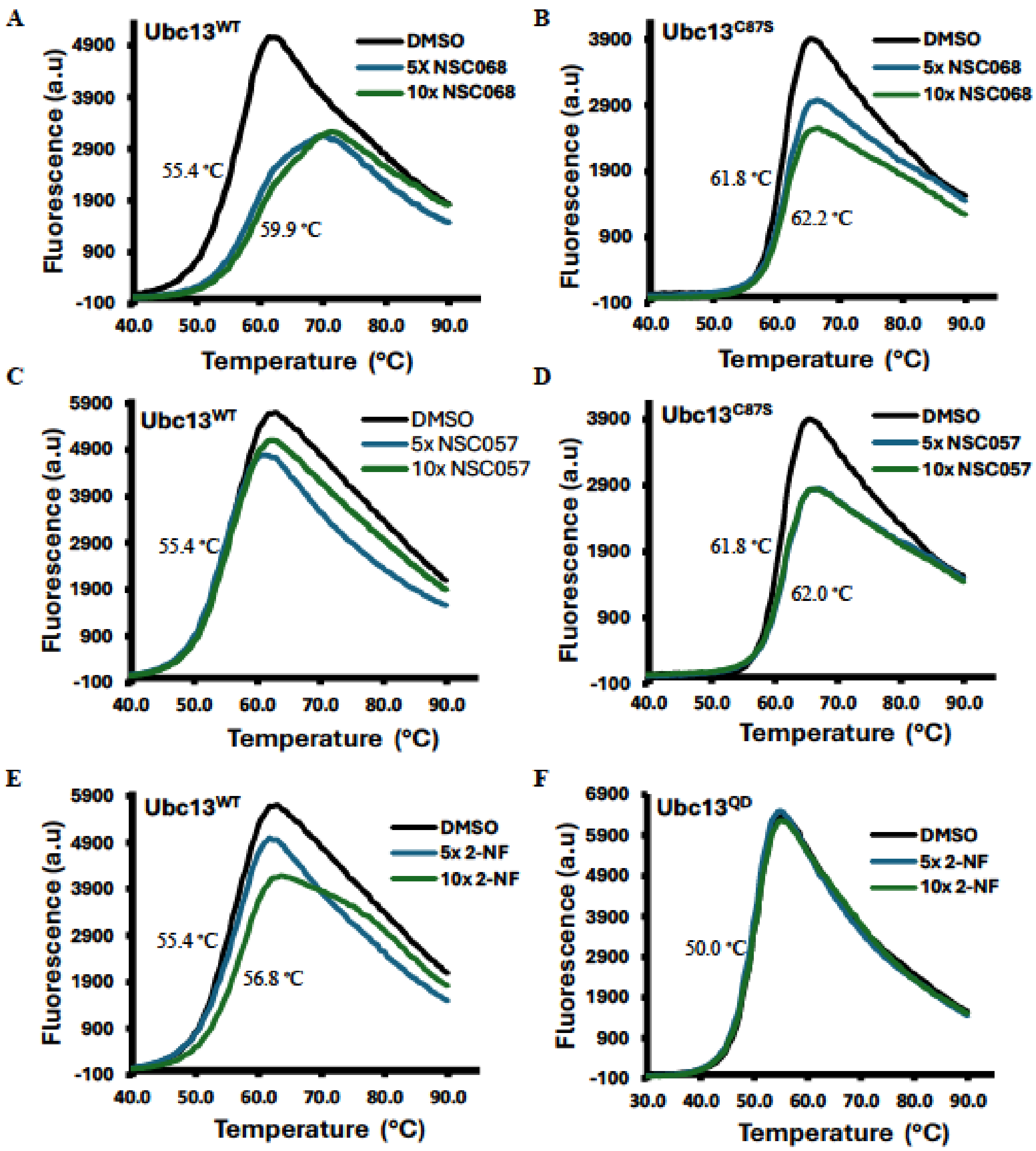


**Figure 2.9. Compound Binding Induces Stabilization of the Melting Temperature of Ubc13.**

**A.** Unfolding transition of 130 μM Ubc13$^{WT}$ with DMSO, 5x NSC291068, and 10x NSC291068 with T$_m$ reported on the graph. **B.** Unfolding transition of 130 μM Ubc13$^{C87S}$ with DMSO, 5x, and 10x NSC068. **C.** Unfolding transition of 130 μM Ubc13$^{WT}$ with DMSO, 5x, and 10x NSC291057. **D.** unfolding transition of 130 μM Ubc13$^{C87S}$ with DMSO, 5x, and 10x of NSC291057. **E.** unfolding transition of 130 μM Ubc13$^{WT}$ with DMSO, 5x and 10x of 2-nitrofuran. **F.** unfolding transition of 130 μM Ubc13$^{QD}$ with DMSO, 5x and 10x of 2-nitrofuran.

### 2.4.5 NSC291068 Reacts with the Active Site Cysteine of Ubc13, and Inhibits Ubiquitin Chain Formation

We previously used an absorbance assay to monitor the covalent modification of Ubc13 with the nitrofuran-containing inhibitor NSC697923[100]. In this assay, Ubc13 was incubated with inhibitor and the reaction was monitored over 15 minutes. Like NSC697923, treatment of Ubc13 with NSC291068 led to an absorbance peak at 420 nm and a reaction was observed **(Figure 2.10A)**. The absorbance peaks were different between the two compounds, with NSC697923 having a maximal absorption peak at 380 nm. We predict this difference is due to the new bond formed with C87, despite both compounds being predicted to have the same leaving group. This reaction required the active site cysteine, as no reaction was observed with the Ubc13C87S mutant (**Figure 2.10A**). An increased rate of reactivity was observed in more alkaline conditions, consistent with a mechanism where the active site C87 nucleophilically attacks the compound (**Figure 2.10B**). NSC291068 was also able to inhibit Ubc13 Lys63-linked polyubiquitin chain building in an *in vitro* assay (**Figure 2.10C**). To test the specificity for Ubc13 we utilized Ubc13$^{QD}$. NSC21068 inhibited the mutant with similar affinity to Ubc13$^{WT}$, suggesting that, unlike NSC697923, the compound is not selective for the native Ubc13 active site configuration (**Figure 2.10D**). To further confirm the reactivity of NSC291068 we utilized mass spectrometry. Reaction of Ubc13 with NSC291068 led to an increase in mass of 137 Da **(Figure 2.11A, B)**, consistent with the covalent addition of a nitrofuran-containing adduct to the active site cysteine (**Figure 2.10G**). In contrast, the related NSC291057 compound showed no effect on Ubc13 $T_M$ and no inhibition of Ub chain building, suggesting that this compound is not a Ubc13 inhibitor (**Figure 2.10E, F**). We predict the overall size and geometry of the compound does not allow it access into the active site of Ubc13.

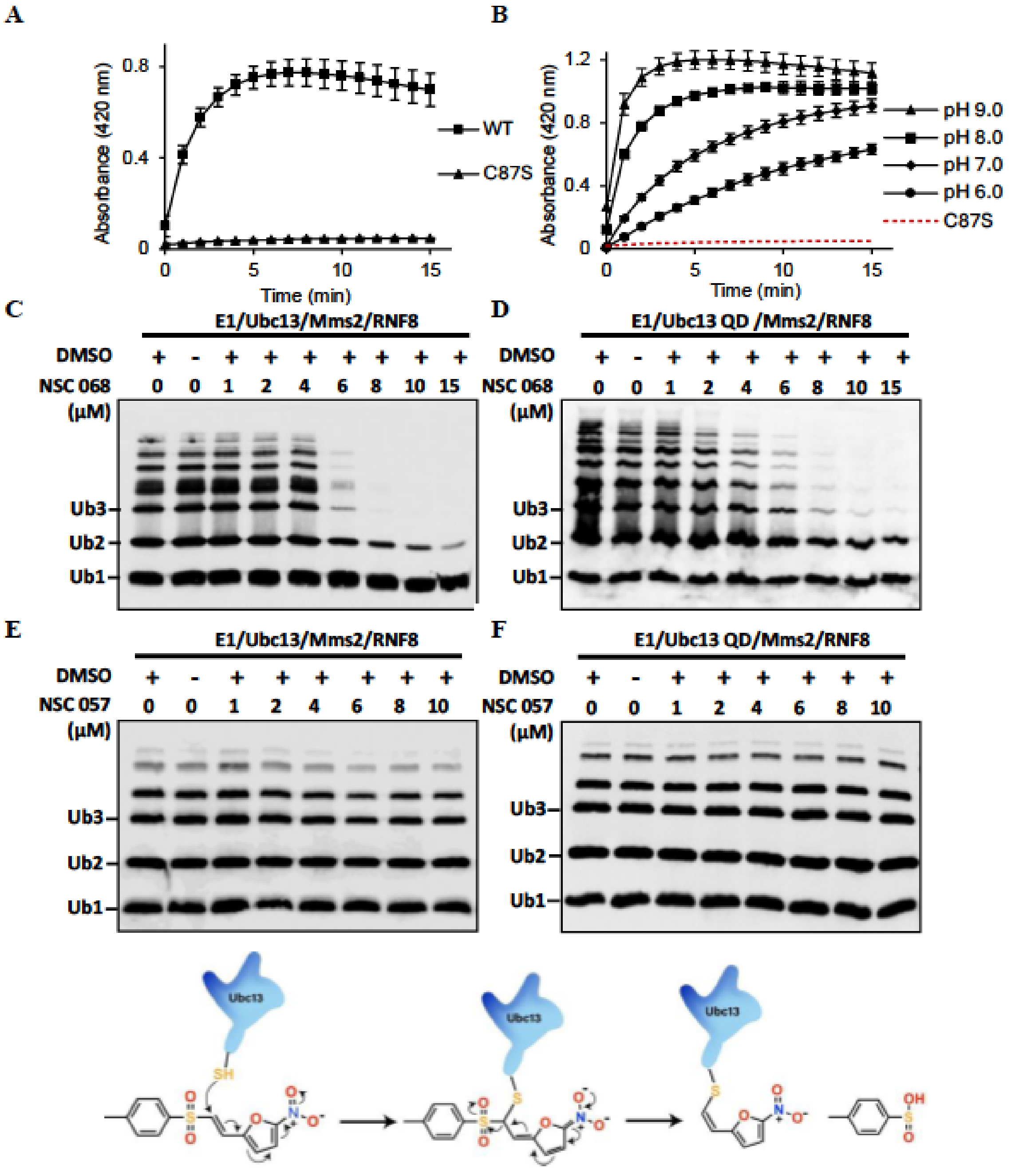


**Figure 2.10. NSC291068 reacts in a pH dependent manner and inhibits ubiquitin chain formation.**

**(A)** Absorbance assay with a wavelength fo 420 nm using Ubc13$^{WT}$ and Ubc13$^{C87S}$ over 15 minutes. **(B)** NSC291068 reactivity pH 6.0 – 9.0. **(C)** Western blot of ubiquitin inhibition assay with Ubc13$^{WT}$, first lane is a 2% DMSO control, 2nd lane nothing but reaction components, 3rd lane to 7th lane are reactions pre-reacted with NSC291068 concentrations ranging from 1-15 µM. Bands correspond to ubiquitin. **(D)** Ubc13$^{QD}$ utilized in the assay with NSC291068. **(E)**Ubc13$^{WT}$ pre-reacted with NSC291057 concentrations ranging from 1 µM-10 µM. **(F)** Ubc13$^{QD}$ with NSC29105. **(G)** Proposed mechanism of inhibition of NSC291068.

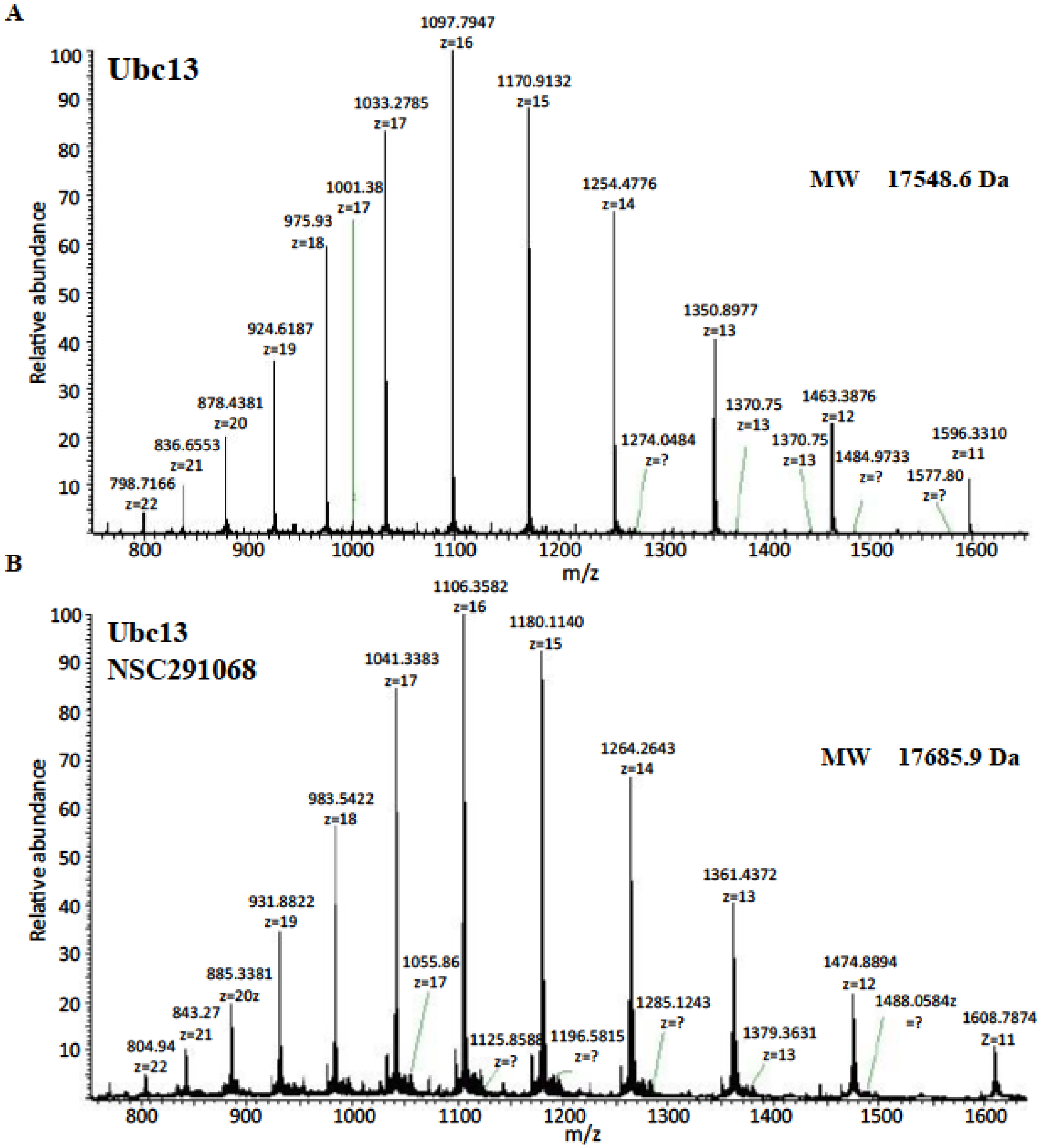


**Figure 2.11. Mass analysis shows the addition of the adduct to Ubc13.**

**A.** Mass spectrometry data collected using orbitrap analyzer, and the molecular weight of Ubc13$^{WT}$ was measured to be 17548.6 Da, in agreement with the theoretical molecular weight. **B.** Ubc13$^{WT}$ reacted with NSC291068, and the molecular weight was measured to be 17685.9 Da, in agreement with the theoretical molecular weight post adduct reaction with active site cysteine.

### 2.4.5 2-Nitrofuran specifically inhibits Ubc13.

The 2-nitrofuran moiety is a common feature of the inhibitors NSC291068 and NSC697923. The thermal shift assay suggested 2-nitrofuran could interact with Ubc13 WT but not the active site mutant Ubc13$^{QD}$ **(Figure 2.9E,F)**, suggesting 2-nitrofuran might provide a lead towards a non-covalent inhibitor targeting the unique Ubc13 active site structure. It was previously observed in cellular based assays that compounds lacking the nitrofuran moiety showed no reduction in Ubc13's activity[116] **(Figure 2.8E, F, G)**, and we found that these compounds also were not able to inhibit Ubc13 in the *in vitro* ubiquitylation assay **(Figure 2.12A, B, C)**. Taken together, these results pointed to the importance of the 2-nitrofuran group for inhibition and we therefore decided to test its ability to inhibit Ub chain formation. Consistent with the results of the thermal shift assay, we found that 2-nitrofuran inhibited the ubiquitylation activity of Ubc13$^{WT}$ but not Ubc13$^{QD}$ **(Figure 2.12D, E)**.

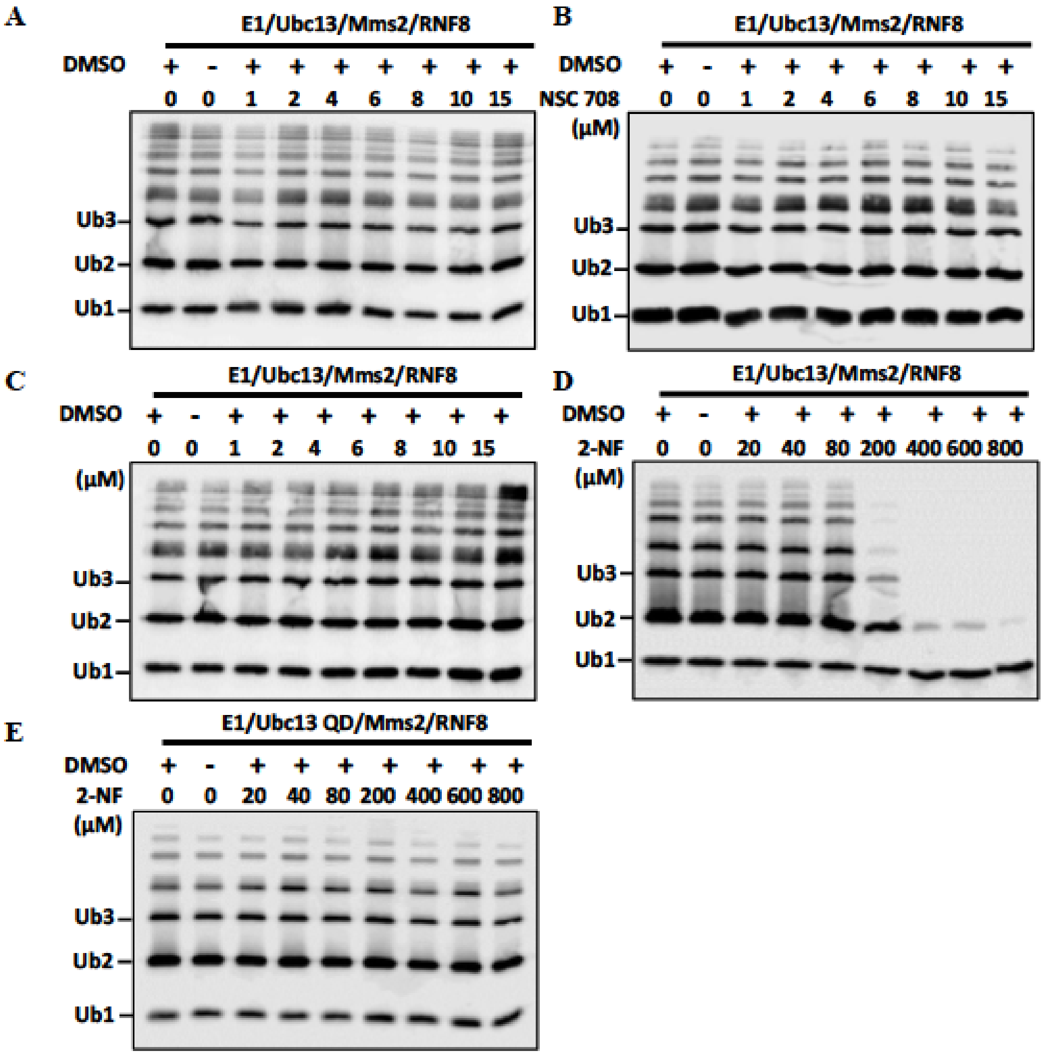


**Figure 2.12. The Nitrofuran Moiety Plays a Key Role in Inhibiting Ubc13.**

Western blots of Ub inhibition assay with Ubc13$^{WT}$ incubated with compounds lacking the nitrofuran moiety **A.** NSC646124, **B.** NSC627708, **C.** NSC656835. **D.** Western blot of Ub inhibition assay using elevated concentrations of the 2-nitrofuran moiety. **E.** Ub inhibition assay using Ubc13$^{QD}$ incubated with 2-nitrofuran.

## 2.5 Discussion

Ubc13 is an E2 conjugating enzyme responsible for synthesizing K63-linked polyUb chains which play a key role in DNA damage response, and NF kB signalling[89,95,96]. The active site loop of Ubc13 is central in how it carries out its function[55,101]. The current model suggests residue L121, which exhibits significant conservation within the enzyme family, occupies a position that would obstruct the attack of acceptor ubiquitin K63 on the active site C87 due to steric hindrance[103,105]. In other E2 enzymes, and in the Ubc13 structure when C87 is linked to the C-terminus of a donor ubiquitin, the gating loop undergoes a conformational change wherein L121 repositions to allow access of acceptor K63 to the active site C87[55,101]. This gating mechanism is postulated to involve two distinct conformations, with dynamics between them driving the selective synthesis of K63-linked ubiquitin chains[103,105]. Here we present a new conformation, crystallized in a more native environment lacking any precipitants. Notably, L121 maintains the ability to block access to C87 in the presented structure, suggesting that the loop is flexible, but it remains in the inactive state. Interestingly, a different hydrogen binding pattern is observed compared to the Ub bound and unbound forms of Ubc13. In the 1J7D structure, the gating loop assumes a conformation held by two adjacent type 1b turns, stabilized by internal hydrogen bonding between N79 and turn 1, as well as N123 and the protein's interior. Consequently, L121 is positioned above the active site C87. In the updated structure, turn 1 shifts away from N79, while turn 2 becomes disrupted, adopting a more elongated conformation. Despite these alterations, L121 remains in the "up" position, and N123 maintains internal hydrogen bonding, mirroring the arrangement in 1J7D. Upon Ub binding, the conformation undergoes further modifications, likely initiated by the interaction between N123 and the incoming K63 of the acceptor ubiquitin. This reorientation of N123 is linked to the repositioning of L121, which no longer obstructs the active site cysteine but instead aligns with the protein's body. Like the revised crystal structure, the region spanning 119-122 adopts an elongated conformation and does not exhibit the turn observed in 1J7D. A comparative analysis of all the available structures of Ubc13 in the PDB revealed the various conformations the loop can occupy with different binding partners. Notably, the newly characterized structure remains unique in comparison to all the structures but is ultimately closer in conformation to the unbound, than the Ub-bound forms. We hypothesize there is either a transitional state conformation between the unbound and bound state, or the loop is flexible and occupies different conformations at physiological temperature. In past studies, two compounds were crystallized covalently linked to Ubc13, providing insights into the dynamics of the active site. BAY 11-7082 left a smaller adduct attached to the cysteine and it was able to evade Ubc13$^{QD}$ loop, and still inhibit its activity[100]. In contrast NSC697923 only inhibited Ubc13$^{WT}$ *in vitro* and not Ubc13$^{QD}$ and specificity was validated *in vivo*[100]. Structural analysis reveals NSC697923 reacted with C87, and the nitrofuran moiety formed a hydrogen bond with N123 **(Figure 2.13A)**. N123 was found to play a key role in interacting with covalent compounds found to inhibit Ubc13, further stabilizing the adducts post reaction as previously shown with NSC697923 and BAY 11-7082. Notably, there were no significant conformational changes induced in the active site loop when compared to the unbound structure of Ubc13[100].

To gain more insights into Ubc13 active site dynamics, we pursued two similar compounds that were discovered in the same screen as NSC697923. NSC291057 showed no direct binding, or reactivity with Ubc13. It was also unable to inhibit polyUb chain formation. NSC291068 showed direct interaction with Ubc13, stabilizing Ubc13's melting temperature by 4 °C. Moreover, the reactivity with the active cysteine, rather than a serine mutant, provides evidence of interaction at the active site. The mass spectrometry data confirmed the adduct left over in the active site is 137 g/mol. When compared to the adduct left over when NSC697923 reacts at the active site (111 g/mol)[100], the active site cysteine must attack at a different carbon than at the nitrofuran, leaving two more carbons extending the 2-nitrofuran in the active site. This points to a new mechanism of covalent inhibition. To further investigate NSC291068 specificity we studied its effects on *in vitro* Ub chain synthesis. Its ability to inhibit both Ubc13$^{WT}$ and Ubc13$^{QD}$ suggests NSC291068 is a nonspecific E2 inhibitor. We hypothesize it evades the obstructed active site groove in Ubc13$^{QD}$, while tethered to the active site cysteine. This could be due to the two additional carbons on the adduct left on C87. Excitingly, we saw evidence of the interaction of 2-nitrofuran with Ubc13$^{WT}$ due to the observed shift in $T_m$, and, at higher concentrations, the fragment was able to inhibit polyubiquitin chain formation. Ubc13$^{QD}$ was resistant to 2-nitrofuran, suggesting it specifically inhibits the active site groove of Ubc13$^{WT}$. Structural comparison between the newly crystallized form (9BIV) and the previously crystallized Ubc13 bound to the 2-nitrofuran adduct post-reaction with NSC697923 (4ONM) shows that N123 likely remains available for hydrogen bonding with the 2-nitrofuran **(Figure 2.13A)**. L121 does not sterically hinder the covalent adduct and is closer to the "up" position like 1J7D. The overlay of Ubc13$^{QD}$ and 4ONM reveals the L121 occupies the groove in the mutant which sterically hinders the 2-nitrofuran **(Figure 2.13B)**. We predict the 2-nitrofuran fragment likely is sterically hindered by L121 in Ubc13$^{QD}$, making it a non-covalent specific inhibitor of Ubc13.

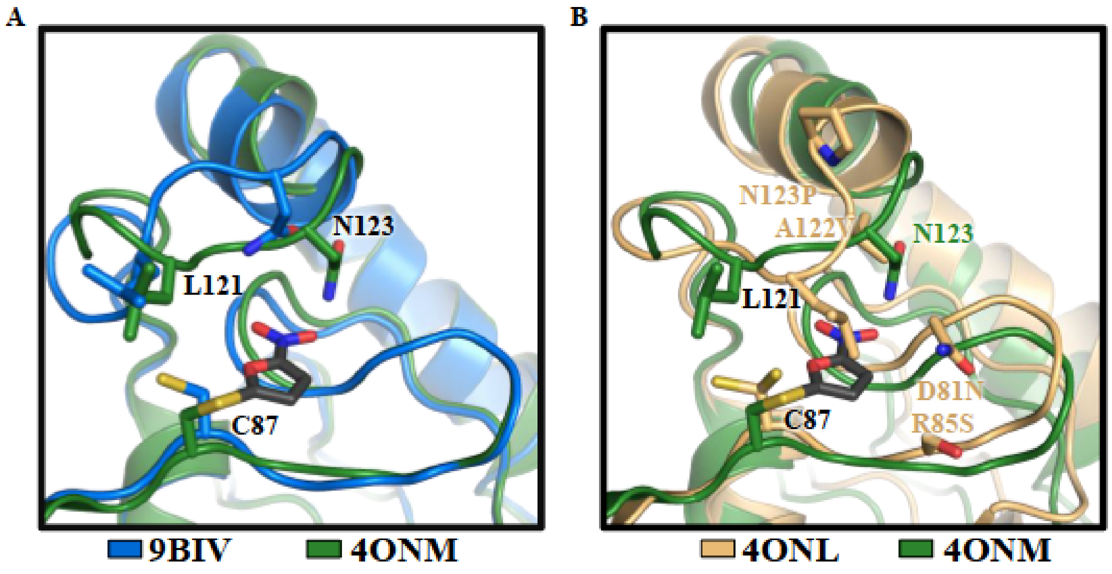


**Figure 2.13. Active Site Analysis with NSC697923.**

The crystal structure of Ubc13 reacted with NSC697923, with the 2-nitrofuran adduct covalently linked in the active site (grey) overlayed on the new form of the active site loop (blue) **A.** and the structure of Ubc13$^{QD}$ (orange) **B.** using PyMOL. Key residues are shown in sticks. Notably, the residues mutated in Ubc13$^{QD}$ D81N, R85S, A122V and N123P.

Certain limitations of our study stem from our inability to determine the structures of the new compounds bound to Ubc13. Future studies should focus on characterizing the structures of compounds for future development of next generation inhibitors. Furthermore, *in vitro* characterization showed promising results in the compounds ability to inhibit Ubc13, but *in vivo* studies are necessary to study the effects of inhibiting Ubc13 on DDR.

Knowing the dynamics of the active site of Ubc13 can help in further developing the inhibitors to more specific compounds. For example, NSC697923 is specific to Ubc13 but off target effects were observed in cells. Modifying certain chemical groups while maintaining hydrogen bonds within the active site could be key in increasing specificity and reducing reactivity. The nitro group on the nitrofuran can be substituted with primary amides, nitrile, or a methyl sulfone. The 2-nitrofuran fragment opens a non-covalent avenue in drug design. The unique conformation of the active site loop, and the presence of N123 within the groove, offers up a selectivity determinant. There have been several strategies employed in assembling fragments into larger specific and tight binding compounds[145]. The 2-nitrofuran could be modified, and several components can be added that take advantage of Ubc13's relatively deep groove compared to other E2s to improve specificity and affinity, while eliminating the prerequisite reactivity of a covalent inhibitor. Understanding drug targets such as Ubc13 with non-redundant characteristics is key due to the opportunity of having a selectivity determinant. Our work suggests that exploring the intricate details of protein structure and dynamics could enhance our ability to develop highly specific and potent inhibitors.

## 2.6 Acknowledgements

We are grateful for the NIH and the Developmental Therapeutics Program for the compounds. We are also grateful for the Alberta Proteomics and Mass Spectrometry facility for their technical assistance.

## 2.7. Funding

The work in the J.N.M.G. laboratory is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant RGPIN-2016-05163) and the Canadian Institutes of Health Research (CIHR168972). I was also supported by Novartis Pharmaceuticals Canada Inc Graduate Scholarship, Queen Elizabeth II Graduate Scholarship, The Department of Biochemistry Doctoral Recruitment Scholarship, Alberta Graduate Excellence Scholarship, and David Lawson Award.

## 2.8 Declaration of Interests

The authors declare no conflicts.

## 2.9 Author Contributions

R.A.F performed most of the experiments and wrote the manuscript. S.C. and Y.F. helped with protein purification. Y.F. preformed the absorbance assays and helped with *in vitro* ubiquitylation assay. J.N.M.G. conceived and supervised the project.

# Chapter 3 – Mechanistic Insights into DNA Damage Sensing RING E3 Ligase RNF138

## 3.1 Abstract

RNF138 is a ubiquitin E3 ligase that binds DNA double-strand breaks, in a cell cycle-dependent manner, promoting homologous recombination during the S2 and G phases. These actions are carried out by ubiquitylating and displacing the heterodimer Ku from DNA ends. RNF138 also ubiquitylates Ct-IP, facilitating DNA end resection. Here we characterized the DNA binding mechanism of RNF138 by testing its binding with various DNA constructs. Our binding studies revealed a strong preference for 3' and 5' overhang DNA, which RNF138 binds with a strong preference for larger overhangs. Furthermore, truncating RNF138 leads to a notable weakening of DNA binding, suggesting that many regions of the protein are involved in DNA binding. AlphaFold modelling of RNF138 suggests a flexible N-terminal region followed by localized ordered domains consisting of a RING domain followed by a zinc finger, connected to two C-terminal zinc fingers and a ubiquitin interacting motif, in a "beads on a string" like configuration.

## 3.2 Introduction

The maintenance of a cell's DNA is important to protect against a variety of diseases and disorders[146]. DNA constantly subjected to insults from both endogenous and exogenous forces; therefore, organisms have implemented a variety of mechanisms to deal with the different kinds of damage that can occur[147]. A particularly problematic form of damage is a double-strand break (DSB)[28]. DSBs are highly cytotoxic lesions, and the occurrence of such a break in the DNA can have multiple consequences, such as chromosomal instability, missense, nonsense mutations, and even cell death[4,28,46]. The two most prevalent methods for dealing with DSBs are homologous recombination (HR) and non-homologous end joining (NHEJ)[1,4,19,21]. Homologous recombination uses a sister chromatid as a template to repair the double-strand break[148]. Non-homologous end joining will process the two raw ends of the strand, removing nucleotides to produce blunt ends and sealing the new gap[30]. A reduction or absence of the employment of HR can lead to the accumulation of mutations over time and often results in an increasingly unstable genome[4]. The use of HR is carefully managed within the cell, through posttranslational modifications (PTMs), and mismanagement of the levels of HR has been associated with tumorigenesis and, in rare instances, resistance to chemotherapy and radiation treatments[149].

Ubiquitylation has been identified as an important PTM at several points in the HR pathway and is used in both degradative and non-degradative signalling[139]. The generally accepted mechanism starts with the activation of ubiquitin by an E1 enzyme, followed by the transfer of the activated ubiquitin to a ubiquitin-conjugating enzyme, E2, which, with the help of an E3 ligase, transfers the activated ubiquitin to a target substrate[65,66]. E3 ligases are crucial in the final step of the ubiquitin cascade, facilitating the transfer of ubiquitin from an E2 enzyme to a substrate lysine[68,120,150]. These ligases fall into two main categories based on their mechanisms: those that transiently bind ubiquitin before transferring it to the substrate, such as HECT and RBR, and those that do not form such an intermediate, typical of most Really Interesting New Gene (RING)-class E3 ligases[120]. RING E3 ligases usually feature a RING domain that interacts with an E2 enzyme, promoting ubiquitin transfer[151]. Unlike HECT ligases, the RING family lacks a catalytically active site to directly accept ubiquitin from the E2 enzyme[151,152]. Instead, they position the E2-ubiquitin complex optimally for the substrate's lysine to undergo nucleophilic attack[151].

Our focus is a DNA-binding ubiquitin E3 ligase, known as RING finger protein 138 or RNF138[129,130]. RNF138, identified as a conserved RING-type E3 ligase, was first recognized for its role as a suppressor of the Wnt pathway[121]. More recently, RNF138 has been associated with DNA damage responses and has been shown experimentally to correlate with increased sensitivity of a cell to DNA damaging agents upon its depletion[35,129,130]. RNF138 was found to promote HR by directly binding DNA overhangs post MRN end resection and ubiquitylates Ku, displacing the heterodimeric complex which protects DSB from end resection and promotes recruitment of proteins involved in NHEJ[129]. In parallel, RNF138 cooperates with the UBE2D (UbcH5) family of E2 enzymes to ubiquitylate and recruit CtIP to DSB which promotes HR and extensive DNA end resection[130].

RNF138 is related to a poorly characterized family of RING E3 ligases, RNG125, RNF114, and RNF166 all of which share similar size and domain distribution[132]. All four proteins consist of a RING domain on the amino (N)-terminus, connected by a short linker (L1) to a zinc finger (ZNF1), followed by a linker (L2) which separates the N-terminal domains from 2 more zinc fingers (ZNF2, ZNF3), a 3rd linker (L3) and a ubiquitin interacting motif (UIM) on the carboxy (C)-terminus[129,132,134]. Recent characterization of RNF125's N-terminal regions has revealed insights into the catalytic functions of this family of E3 ligases, but the structure and molecular function of RNF138 is yet to be characterized[134,153]. More recently, several studies have revealed an important role for RNF138 in cancer. It is overexpressed in glioma cells, a very aggressive type of brain cancer, promoting proliferation, metastasis, and radioresistance[127,135]. It is further implicated in cisplatin resistance in gastric cancer cells[128]. Therefore, it is imperative to characterize RNF138's activity and understanding the functions of the domains of RNF138 is key to understanding RNF138's DNA and protein binding activities.

In this study, we set out to biochemically characterize RNF138. We initially truncated RNF138 to varying degrees to investigate the impact they may have on DNA binding. followed by investigating the length of DNA required for RNF138 binding. Our results reveal that RNF138 is a relatively unstable protein, which prefers larger DNA substrates, specifically DNA substrates containing large overhangs. Furthermore, truncation of any domains within RNF138 reduced or eliminated DNA binding. AlphaFold2 analysis of RNF138 revealed a highly flexible protein, with organized domains separated by linker regions. Utilizing AlphaFold2 with the multimer implementation, reveals the RING and ZNF2 with elements from L2 are involved in E2 and ubiquitin interaction.

# 3.3 Methods

### 3.3.1 Protein Constructs and Expression

Four constructs of RNF138 were gifted from Dr. Leo Spyracopoulos's lab from Dr. Brian Lee, full-length RNF138, RING-ZNF1, ZNF1, ZNF2, ZNF3, and ZNF2, ZNF3. The fifth construct truncated at the UIM was generated in our lab. All constructs were generated with Glutathione S-Transferase (GST) tags, connected by a 3C protease cut site for tag removal. Both genes were introduced into the plasmid pGEX6p1, which also contained the genes for ampicillin resistance. These genes were expressed in a bacterial expression system. The plasmids were transformed into BL21 Gold *E. coli* cells, which contained genes that provided kanamycin resistance.

### 3.3.2 DNA Constructs

All oligonucleotides were purchased from Integrated DNA Technologies (IDT) **(Table 3.1)**. Four 3' labelled 6-carboxyfluorescein (FAM), with a 5'phosphate group, at 1 mM concentrations. The remaining oligonucleotides were ordered at 100 µM. Upon arrival, DNA substrates were resuspended in water. For annealing reactions 1:1 of FAM labelled DNA with a complementary unlabelled DNA strand were combined in an Eppendorf tube. The tubes were incubated at 95 °C for 5 min. The tubes were then allowed to cool down slowly overnight to room temperature.

**Table 3.1. Oligonucleotide Sequences Utilized in Our Study**

| Number of bases | Oligonucleotide Sequence |
|---|---|
| 40 | P 5 TAACGGAGAGGCTAAGGATGTACTGCTTGTTAGGCTCCAT 3 FAM |
| 40 | 5 ATGGAGCCTAACAAGCAGTACATCCTTAGCCTCTCCGTTA 3 |
| 60 | 5 ATGGAGCCTAACAAGCAGTACATCCTTAGCCTCTCCGTTAAGTGCCAATTTCACTGTAAG 3 |
| 20 | 5 ATGGAGCCTAACAAGCAGTA 3 |
| 30 | P 5 TAACGGAGAGGCTAAGGATGTACTGCTTG 3 FAM |
| 30 | 5 ACAAGCAGTACATCCTTAGCCTCTCCGTTA 3 |
| 40 | 5 ACAAGCAGTACATCCTTAGCCTCTCCGTTAATGGAGCCTA 3 |
| 20 | 5 ACAAGCAGTACATCCTTAGC 3 |
| 20 | P 5 AACGGAGAGGCTAAGGATGT3 FAM |
| 20 | 5 ACATCCTTAGCCTCTCCGTT 3 |
| 30 | 5 ACATCCTTAGCCTCTCCGTTATGGAGCCTA 3 |
| 10 | 5 ACATCCTTAG 3 |
| 15 | P 5 TAACGGAGAGGCTAA 3 FAM |
| 15 | 5 TTAGCCTCTCCGTTA 3 |
| 20 | 5 TTAGCCTCTCCGTTAATGGA 3 |
| 10 | 5 TTAGCCTCTC 3 |

### 3.3.3 Bacterial Cell Growth and Induction

In the case of all RNF138 constructs, we started with an *E. coli* BL21 Gold cell line. This cell line was transformed with a plasmid pGEX6P1, which contains ampicillin-resistant genes, as well as the gene for RNF138-GST constructs. Our initial culture was inoculated from bacterial colonies on ampicillin and kanamycin plates, in 5 mL of LB Broth treated with 5 µL of 50 mg/mL kanamycin (50 µg/mL) and 5 µL of 100 mg/mL ampicillin (100 µg/mL) and incubated overnight at 37 °C. We transferred the entire 5 mL preparation to 500 mL of LB Broth pre-treated with 500 µL of kanamycin and ampicillin. This culture was incubated for at least 2-4 hours at 37 °C. The optical density for our sample was measured at 600 nm to determine whether or not to move on to the next step. Once the optical density fell between 0.6 and 0.7, we moved on to the next culture.

8 bottles, each containing 1 L of LB broth, were prepared, and treated with 1 mL of kanamycin and ampicillin. 50 mL from the 500 mL preparation was added to each bottle. The 1 L preparations were incubated at 37 °C for a further 2-4 hours. Again, optical density at 600 nm was used to determine whether to move on to induction. 0.6-0.7 was once again the target optical density. The cells were induced by the addition of 200 μL of 0.5 M $ZnCl_2$ and 200 μL of 1 M IPTG to each 1L bottle. The preparations were then incubated overnight at 17 °C. The cells were pelleted out of solution via centrifugation at 4000 RPMs (JLA-8.1 rotor) for 15 minutes at 4 °C. Supernatants were discarded, and the pellet was either used right away or frozen via liquid nitrogen and stored at -80 °C.

### 3.3.4 Bacterial Expressed Protein Purification

The following protocol was used for all protein constructs. The pelleted cells, used directly following growth or retrieved from -80 °C, were resuspended in RNF138 lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 10 μM $ZnCl_2$, 1 μL/mL BME). To the sample, 0.04 g of lysozyme and 200 μL of a protease inhibitor cocktail (Sigma Aldrich) were added. Resuspension was carried out at 4 °C, using a magnetic stir bar set up for ~35 minutes. Following resuspension, the cells were lysed by sonication at an amplitude of 70%, using 15 seconds on 45 seconds off pulses, for a total of 3 minutes of sonication. The lysate was subsequently centrifuged at 12000-13000 RPMs (JA-17 Rotor) for 40 minutes at 4 °C. The pellet was discarded. The supernatant was then treated with 4.5 mL of polyethyleneimine (PEI) (dropwise to a final concentration of 0.05% w/v over approximately 15 minutes). The sample was left to mix for an additional 20 minutes, then centrifuged at 13000 RPMs (JA-17 rotor) for 30-35 minutes at 4 °C to pellet out any precipitated DNA. The supernatant is collected to proceed to GST affinity chromatography. The supernatant was then transferred to a bottle and incubated with glutathione sepharose beads (Sigma Aldrich) for 1 hour. This mixture was transferred to a falcon tube following incubation and was centrifuged at 4500 RPMs (JA-7.5) for 2 minutes at 4 °C. The resulting supernatant was discarded, the bottle was rinsed with lysis buffer and the previous centrifugation was repeated. The beads were transferred back to a column and washed with lysis buffer (50 mM Tris pH 7.5, 500 mM NaCl, 10 μM $ZnCl_2$, 1 μL/mL BME). We eluted the GST-tagged constructs with 10 mL volumes of elution buffer (50 mM Tris pH 7.5, 500 nM NaCl, 10 μM $ZnCl_2$, 10 mM glutathione, 1 μL/mL BME). Four of these 10 mL fractions were collected in total. Each of these fractions was collected in test tubes. One additional 15 mL fraction was also collected and was always collected last. These fractions were pooled and digested with 3C protease overnight. At this point in the experiment, we needed to separate free GST from RNF138 as they are far too close in size to be effectively separated in size exclusion chromatography. We concentrated the 50 mL overnight 3C digests volume down to 1-2 mL using 10 kDa concentrator tubes, added 20 mL of lysis buffer lacking glutathione and repeated cycles of centrifugation at 4000 RPMs (JA-7.5 rotor) three times, effectively buffer exchanging glutathione out. After the 3rd run, the samples were further diluted with lysis buffer and incubated on fresh glutathione beads for an hour and a half at 4 °C. The flow through after the incubation was collected, the column was washed twice with lysis buffer bringing the total volume to 40 mL, the proteins were then concentrated to 2 mL using 10 kDa centricon concentrator tubes (Sigma Aldrich). Finally, size exclusion chromatography was applied to the sample. We employed a Superdex 75 16/60 column, and protein elution was monitored using a 280 nm UV lamp, fractions corresponding to a peak in 280 nm were collected. These fractions were pooled for further use. The success of the steps of purification was established by running samples taken from each step in a 5-15% polyacrylamide gel. The resulting gel(s) were run at 240 volts for 30 minutes and then stained using Coomassie blue dye.

### 3.3.5 Mammalian Cell Cloning and Transfection

Expi293F cells were gifted to us from Dr. Marek Michalak's lab. Mammalian cell construct of RNF138 was cloned into pCDNA3.1 vector, with a CMV promoter followed by N-terminal GST tag, followed by a Tobacco Etch Virus (TEV) protease cleavage site, 3C protease, and 10x His-tag. Fresh Expi293 cells were thawed and cultured in Nalgene Single-Use PETG 125 mL Erlenmeyer Flasks (Thermofisher Scientific) and passaged 3 times in CDM4HEK293 media (Cytiva), supplemented with 1X GlutaMAX before transfection. 125 mL flasks containing 35 mL of cells were grown on a shaker at 37 °C, 80% humidity environment, and 8% $CO_2$ atmosphere with a shaker speed of 125 rpm. For the first passage, cells were counted using trypan blue, and a hemocytometer, to determine cell numbers and viability. At the first passage, 1-3 x $10^6$ viable cells/mL were acceptable. Following the first passage, routine general maintenance of cells was carried out when they reached a density of 3-5x$10^6$ viable cells/mL. For transfection, cells were grown to 3-5 mvc (million viable cells)/mL >95% viable. They were pelleted, at 600 xg for 2 min at room temperature and the media was carefully removed. Fresh warm media was added to give a final of 1 mcv/mL and incubated overnight. On the second day cells were counted and spun down at 600 x g for 2 min, at room temperature and media was gently removed. Growth media was replaced with warm transfection media HyCell TransFx-H (Cytiva), supplemented with +0.1% Pluronic F-68, to give a final volume of 20 mvc/mL. 1.25 µg DNA/mvc was added while gently swirling, 3.75 µL of 1mg/ml 40 kDa PEI/mvc was added slowly while swirling and cells were incubated for 3 hours. In a separate flask, transfection media + 0.1% Pluronic F-68 were warmed in the expression flask to give 1mvc/ml final, cells containing the DNA PEI mixture, were then transferred into the expression flask, 3.5 mM of valproic acid (VPA) was added. Cells were incubated for 4 days. After 4 days, cells were counted, pelleted, and flash-frozen in liquid nitrogen.

### 3.3.6 Expi293 GST protein purification

Around a 160 million cells were utilized. Cells were lysed by resuspension in 5 mL of phosphate-buffered saline (PBS) supplemented with 1 mM EDTA, 1% Triton X-100; 1 mM DTT and 1 tablet of cOmplete™ Protease Inhibitor Cocktail (Millipore Sigma). Cells were pipetted up and down on ice and allowed to tumble for 1 hour at 4 °C. Cells were then sonicated, with short pulses at 70% amplitude, 1 second on, 5 seconds off. Cells were then spun down at 13000 RPMs (JA-17 rotor) for 30-35 minutes at 4 °C, the supernatant was collected, the pellet was washed again with 3 mL of lysis buffer, spun down again, and the supernatant was collected. GST beads were equilibrated with lysis buffer at 4 °C, and the supernatant was incubated with 1 mL of GST beads and allowed to incubate for 1 hour at 4 °C. The protein was then eluted using the lysis buffer, except 0.5% Triton, and supplemented with 10 mM reduced glutathione. Histidine purifications followed the same protocol, but instead of GST beads, we utilized HisPur™ Ni-NTA Resin (Thermofisher Scientific), and 100 mM imidazole was used for elution.

### 3.3.7 SDS-PAGE and Immunoblotting

The samples were loaded into small, custom-cast gel trays containing Tris solutions with specific concentrations and pH levels (37.5 mM at pH 8.8 for the main layer and 12.5 mM at pH 6.8 for the supporting layer), along with 0.1% sodium dodecyl sulphate (SDS), and 5-12% polyacrylamide. Precision Plus protein dual colour standards from Bio-Rad were used as reference markers. Electrophoresis was conducted at 150 volts in a running buffer solution containing 25 mM Tris pH 8.3, 192 mM glycine, and 0.1% SDS. The separated proteins were then transferred onto a nitrocellulose membrane via wet electro-transfer for one hour at 110 volts using a transfer buffer solution containing 25 mM Tris pH 8.3, 192 mM glycine, and 20% methanol. For immunoblotting, the nitrocellulose membrane was initially blocked with a solution of 4% fish skin gelatin (FSG) dissolved in TBS (Tris-buffered saline) at room temperature. The primary antibodies were diluted in freshly prepared TBS containing 0.1% Tween-20 (TBST) or 2% FSG for RNF138 antibodies and incubated overnight at 4 °C on a rocker. The membranes were subsequently washed three times, each for 10 minutes, in TBST. They were then incubated for 1 hour at room temperature with secondary antibodies conjugated with horseradish peroxidase (HRP), IRDye 680RD, or IRDye 800CW (all from LI-COR Biosciences) in TBST. Following incubation, the membranes were washed again three times, each for 10 minutes, in TBST, followed by a single wash in TBS for 10 minutes. HRP activity was detected by immersing the membranes in Amersham ECL Prime Western Blotting Detection Reagent (from Cytiva) for 2 minutes. Chemiluminescence signals were acquired using the Odyssey Fc Imaging System and quantified by densitometry using Image Studio software, both from LI-COR Biosciences. Fluorescence signals from IRDye were also detected using the same imaging system and quantified similarly.

### 3.3.8 Electrophoretic Mobility Shift Assay

Serially diluted concentrations of RNF138 constructs (0/0.02/0.04/0.08/0.2/0.4/0.8/2/4/8 µM) were tested against a consistent concentration (20 nM) of DNA substrate. Each reaction mixture comprised 1 µL of 3'FAM DNA substrate, 1 µL of protein in storage buffer, 5 µL of EMSA binding buffer (containing, 25 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM DTT, and 10% glycerol), and 2 µL of $H_2O$, which were mixed and then incubated for 10 minutes. Following this, 2 µL of 4X EMSA loading dye (consisting of 10% glycerol, 240 mM Tris-HCl (pH 7.5), and 4 mg/mL bromophenol blue) was added to the reaction mixture before loading it onto a pre-run 6% native polyacrylamide gel (prepared with a ratio of 19:1 acrylamide/bisacrylamide). Electrophoresis was conducted at 100 V for 55 minutes at 4 °C in 1X TBE buffer. Free and bound

DNA substrates were then visualized using a Typhoon™ phosphorimager and qualitatively analyzed.

### 3.3.9 AlphaFold2 Modelling and Analysis

Structural modeling and predictions for RNF138 (Uniprot accession: Q8WVD3), UbH5b (P62837) and Ubiquitin (P0CG48) were conducted using the ColabFold implementation of AlphaFold2 (https://github.com/sokrypton/ColabFold)[154–157]. The protein sequences were input in AlphaFold_MMseq2 Google Colab notebook (version 1.5.5). Output models were assessed and ranked based on predicted template modelling (pTM) and interface predicted template modelling (ipTM) scores, with the highest scoring model selected for further analysis. Model visualization, analysis, and image creation were performed using PyMOL (Version 2.3.3, Schrödinger, LLC). Confidence metrics were generated and plotted using Microsoft Excel (version 2403) and Morpheus (available at: https://software.broadinstitute.org/morpheus). All structural alignments were done with experimental structures deposited in the protein data bank (PDB) **(Table 3.2).**

**Table 3.2. Proteins Used in the structural Alignments.**

| PDB AC | Proteins | RMSD ( ) |
|---|---|---|
| 8GBQ | RNF125-UbcH5b | 0.387 |
| 4QP1 | RNF146-UbcH5a | 0.692 |
| 4AP4 | RNF4-UbcH5a | 0.706 |
| 2YHO | MYLIP-UbcH5 | 0.905 |
| 4A4C | Cbl-UbcH5b | 0.667 |
| 3RPG | BMI-RING1b-UbcH5b | 0.501 |
| 4V3K | RNF38-UbcH5b~Ub | 0.460 |
| 4AUQ | BIRC8-UbcH5b~Ub | 0.409 |
| 8GCB | RNF125-UbcH5b~Ub | 0.486 |
| 3ZNI | Cbl-UbcH5b~Ub | 0.556 |
| 6W9D | RNF12-UbcH5b~Ub | 0.477 |
| 5ULK | RNF165-UbcH5b~Ub | 0.512 |

### 3.3.10 Crystallization

We carried out several crystallization screens with ZNF2,3, using screens from Anatrace's MCSG crystallization suite. We made use of MCSG-1, MCSG-2, and MCSG-3. The trays were periodically checked for potential development of crystals via examination under a microscope. Just as with ZNF2,3, we carried out microlytic crystallization screens with the full-length RNF138 complexed with DNA. The DNA strand used in this complex had a 40 nucleotide double-stranded region and a 20 nucleotide 3' overhang. The same screens from Anatrace's MCSG crystallization suite were used. We made use of MCSG-1, MCSG-2, and MCSG-3.

### 3.3.11 Thermal Shift Stability Assays

This procedure was the same for the two constructs. First, 360 µM of protein was diluted to 130 µM with 500x of SYPRO orange dye. The final concentration of the dye was 250x. A SYPRO assay consists of a library of low and high salt buffers, at varying pH values ranging from pH of 3 to 10. 5 µL of the reacted protein and dye solution was added to 15 µL of each buffer. The solutions were prepared in triplicates in 96 well plates. The temperature was increased from 20 °C to 95 °C degrees in 0.5 °C increments over 45 minutes, and all sample readouts were done in triplicates using MasterCycler RealPlex (Eppendorf) filter set at 550 nm. The results produced are fluorescence values as a function of temperature, wherein the melting temperature is inferred from the inflection point[144].

## 3.4 Results

### 3.4.1 Purification of RNF138 Constructs

To characterize RNF138 biochemically, the protein had to be isolated. We utilized different constructs as this approach allowed us to study the various domains of RNF138 in isolation to try to understand their functions and gain insights into their stability **(Figure 3.1A)**. Initially, we decided to utilize *E. coli* as the expression host as it is the most cost-effective way to express large amounts of protein and purify it. We received four constructs as a gift from Dr. Leo Spyropoulos's lab: RNF138$^{WT}$, RING-ZNF1, ZNF1,2,3, and ZNF2,3, and we created 1 more truncation, a construct lacking the UIM, RNF138$^{1\text{-}225}$ **(Figure 3.1A)**.

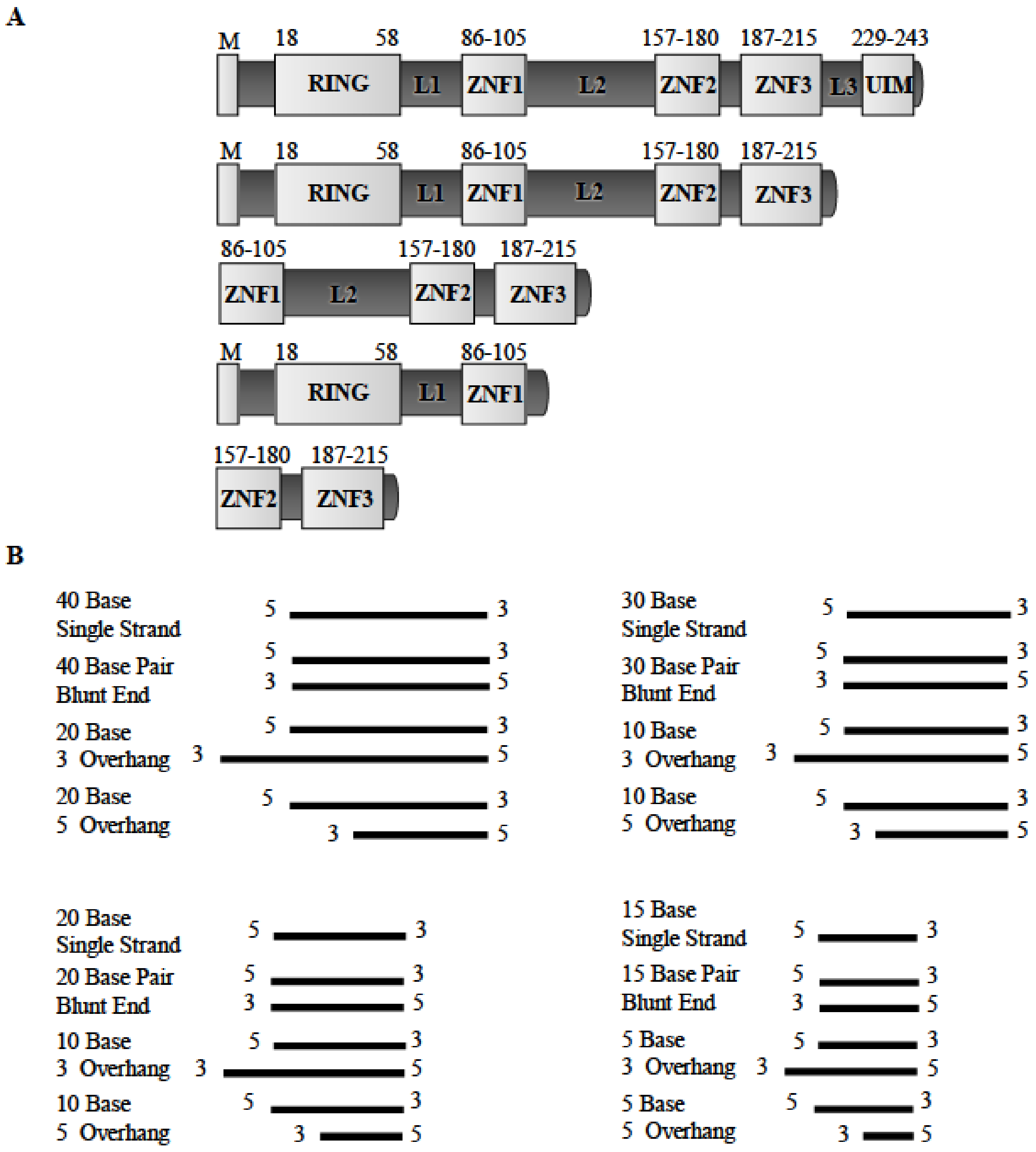


**Figure 3.1. Protein Constructs, and Schematic Representation of DNAs Used in Our Study.**

All samples were subjected to GST-affinity tag purification, followed by polyethyleneimine (PEI) precipitation, to precipitate contaminating DNA and extraneous DNA binding proteins out of our samples, followed by sizing exclusion chromatography. All the truncations were successfully purified with greater than 80% purity **(Figure 3.2A-E)**. Except for ZNF2,3, all constructs were relatively unstable over time and expressed in lower yields. To mitigate instability, proteins were kept at high salt, lower concentrations, and flash frozen, and were only buffer exchanged and concentrated when they were to be used **(Figure 3.3A)**. ZNF2,3 was used in crystallization screens due to its promising yields and stability **(Figure 3.3B)**, but we have not been successful in our efforts to crystallize this construct. We decided to try a mammalian cell expression system concurrent with *E.coli* purification to investigate whether a more native environment will yield more stable forms of RNF138. *E. coli* lacks mammalian protein chaperones, and our protein may require specific post-translational modifications (PTMs) for proper folding, stability, and overall function. We utilized Expi293F cells, which are suspension-adapted human embryonic kidney (HEK293) cells. Two constructs containing RNF138$^{WT}$, with differing tags, were cloned under the control of a CMV promoter. The GST- RNF138$^{WT}$-His construct, yielded ambiguous results, as antibodies for both GST, and His tags were non-specific and the current antibodies for RNF138, were also very cross-reactive when probing cell extracts. We were unable to confirm if there was any protein expression using that construct. Another construct was generated, by cloning an N-germinal GFP tag instead of GST and we kept the C-terminal tag. Anti-GFP bodies are well characterized and are relatively non-cross-reactive. We were able to detect protein expression using this construct and carry out purification, but the yields were far lower than we anticipated **(Figure 3.2F)**. Further optimization is required which may yield enough for biochemical, and structural characterization of RNF138.

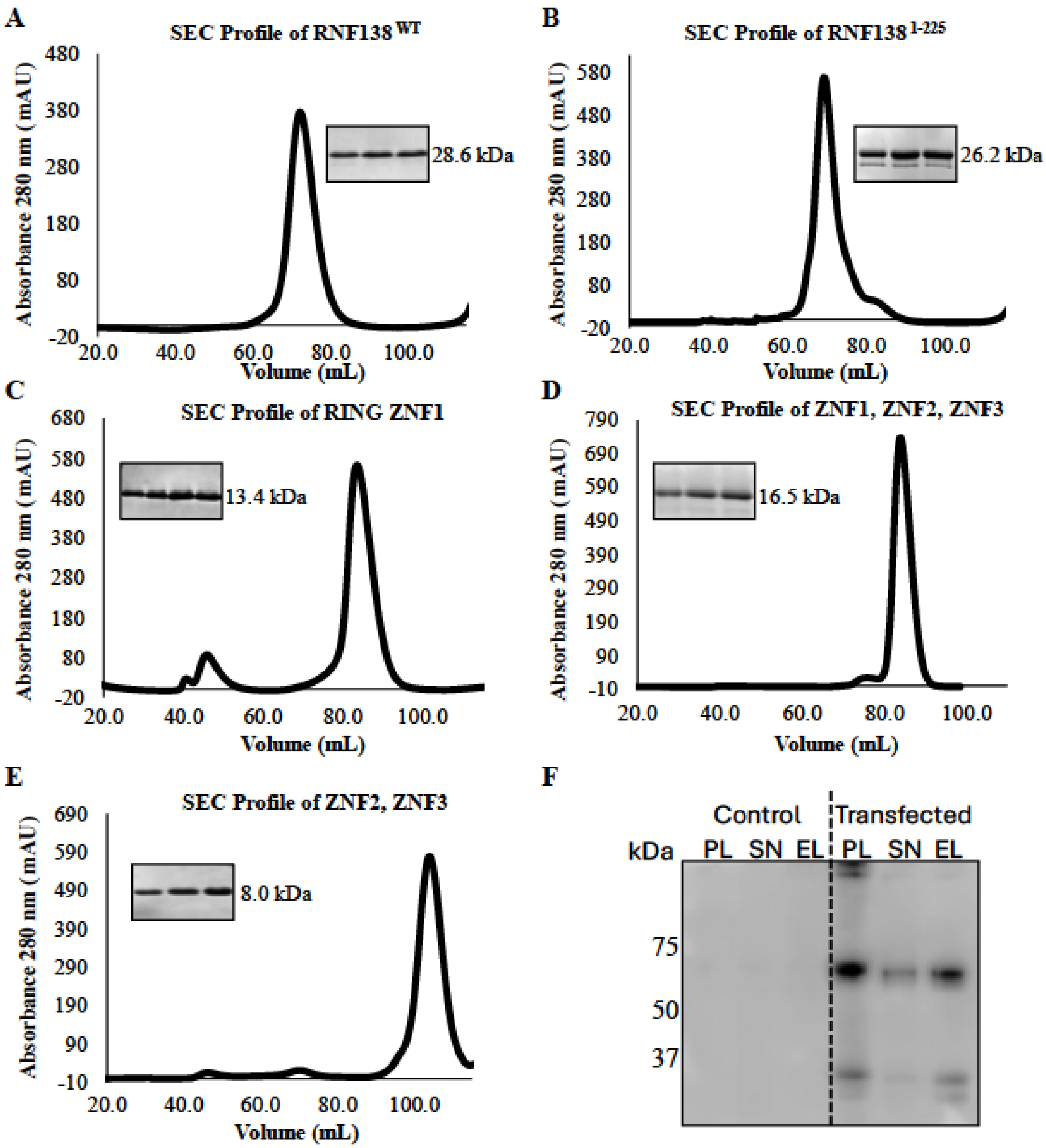


**Figure 3.2. RNF138 Purification Results.**

**A.** *E.Coli* expressed protein size exclusion chromatogram and a representative image of an SDS-PAGE gel of the final purification step of RNF138$^{WT}$. **B** RNF138$^{1\text{-}225}$ **C.** RING ZNF1 **D.** ZNF1, ZNF2, ZNF3 **E.** ZNF2,ZNF3. **F.** Immunoblot of non-transfected Expi cells (left, control) and GFP-His-RNF138 construct (right panel) with the following samples: pellet post lysis and spin down (PL), supernatant post lysis and spin down (SN), post nickel column elution (EL). Primary mouse anti-GFP (1/2000) was used, and secondary anti-mouse HRP (1/5000).

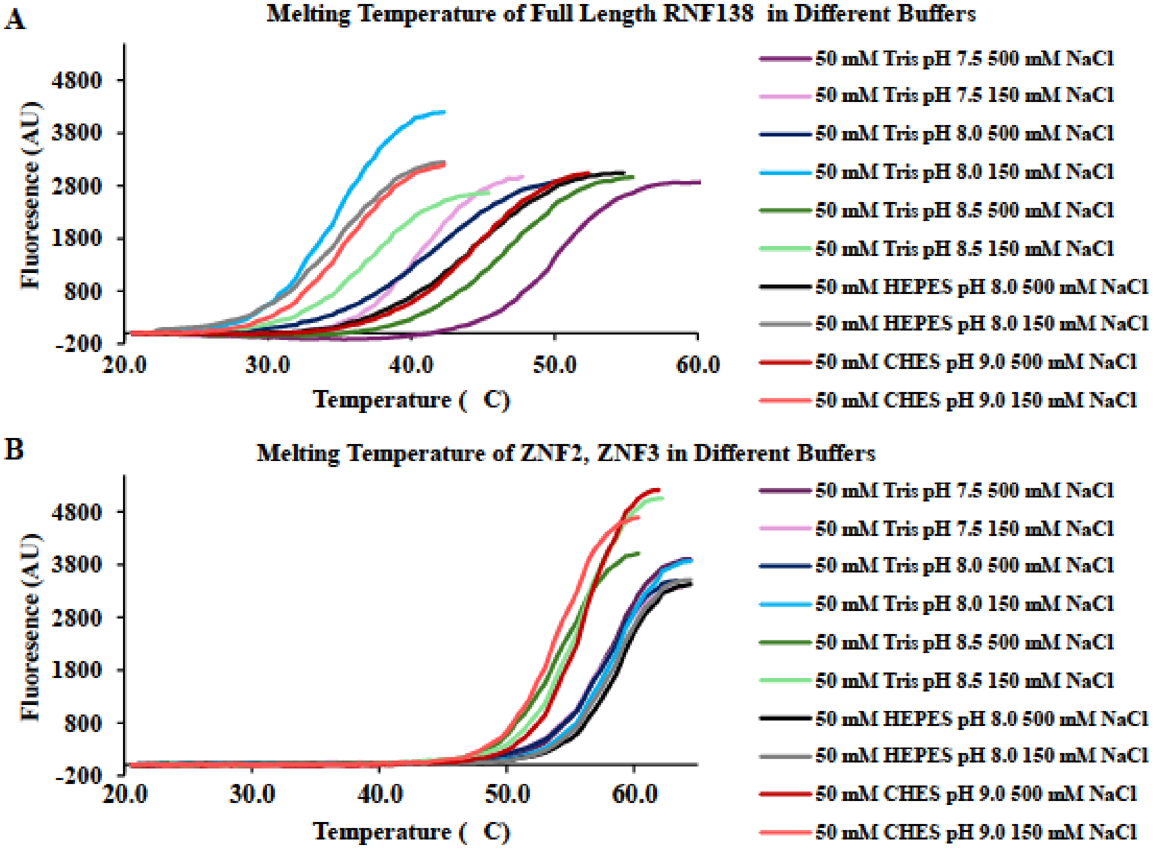


**Figure 3.3. Thermostability Shift Assay Using SYPRO Orange Dye of RNF138$^{WT}$, ZNF2, ZNF3, and RING ZNF1 in Different Buffer Conditions.**

Proteins were diluted with SYPRO orange dye and incubated with a library of buffers ranging from pHs of 3 to 10, with each pH having two different NaCl 150 mM and 500 mM NaCl. Temperatures were gradually increased, and fluorescence was measured as a function of temperature. Melting temperatures are inferred from the inflection point. A lower melting temperature suggests an overall less stable protein in the respective solution. Both constructs were consistently less stable, the most stable buffers are shown. **A.** Overall RNF138$^{WT}$ was most stable in high salt buffer conditions, with the most stable buffer being the selected as our lysis buffer in our experiments: 50 mM Tris pH 7.5 500 mM NaCl, with a melting temperature of: 50.8±0.3°C. The same buffer with 150 mM NaCl, was the most stable (41.5±0.6°C) in the low salt buffers therefore this was the selected buffer for our biochemical analysis of RNF138. **B.** ZNF2, ZNF3 was more stable, with the highest melting temperature (58.9±0.3°C) in 50 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) and 500 mM NaCl. A lower melting temperature was observed with lower salt concentrations, but the melting temperature remained higher than RNF138$^{WT}$, suggesting ZNF2, ZNF3 construct is more stable. In comparison to RNF138$^{WT}$ in 50 mM Tris pH 7.5 500 mM and 150 mM NaCl ZNF2, ZNF3 had melting temperatures of 58.5±0.6°C, and 57.9±0.3°C respectively.

### 3.4.2 RNF138 Preferentially Binds DNA Overhangs

RNF138 is recruited to DSB by directly binding DNA[129]. To characterize the DNA binding mechanism of RNF138, we utilized an electrophoretic mobility shift assay (EMSA) using purified RNF138 full length, and different protein truncations which were incubated with different fluorescently labelled DNA constructs and were visualized on native gels starting with a 40 base 3' FAM labelled DNA **(Figure 3.4)**. The two constructs which showed significant DNA binding activity were RNF138$^{WT}$ and RNF138$^{1\text{-}225}$, and both showed a preference for DNA with 20 nucleotides 3' and 5' overhang. A shift is observed at 200 nM and 2 µM for RNF138$^{WT}$ and RNF138$^{1\text{-}225}$ respectively, suggesting a 10-fold stronger interaction with the wildtype protein compared to the RNF138 lacking a UIM **(Figure 3.4)**. This pattern is also observed in the case of the 5' overhang **(Figure 3.4)**. A 2-fold difference between 3' and 5' overhang DNA was observed in the case of RNF138$^{WT}$ with binding happening at 400 nM of protein. RNF138$^{1\text{-}225}$ again had a 10-fold weaker interaction as a shift was observed at 4 µM **(Figure 3.4)**. Interestingly, although the overhang region remains 20 nucleotides long in the 5' overhang DNA construct, the double stranded region of the DNA gets smaller. This is because all the DNA constructs were designed to complement the 40 nucleotide FAM labelled DNA strand, which may explain the discrepancy observed between 3' and 5' overhang DNA binding. This suggests that the duplexed region of the overhang constructs may affect protein binding. In the case of 40 base single-stranded (ssDNA) and 40 base-paired double-stranded (dsDNA), RNF138$^{WT}$ bound both with similar affinities, shifting the DNA at 800 nM of protein, whereas RNF138$^{1\text{-}225}$ showed binding at 8 µM to the ssDNA and 4 µM to dsDNA **(Figure 3.4)**. Overall, the results reveal that RNF138 has a strong preference for DNA overhangs, and removing the UIM leads to a marked drop in DNA binding affinity. Additionally, the shortening of the duplexed region in the 5'overhang construct was observed to reduce the binding affinity by 2-fold, signalling an involvement of the double-stranded region of DNA. Interestingly, the DNA binding activity of RNF138 does not plateau at higher concentrations, suggesting a non-specific binding, wherein increasing concentrations of RNF138 lead to the loading of more protein molecules onto the DNA.

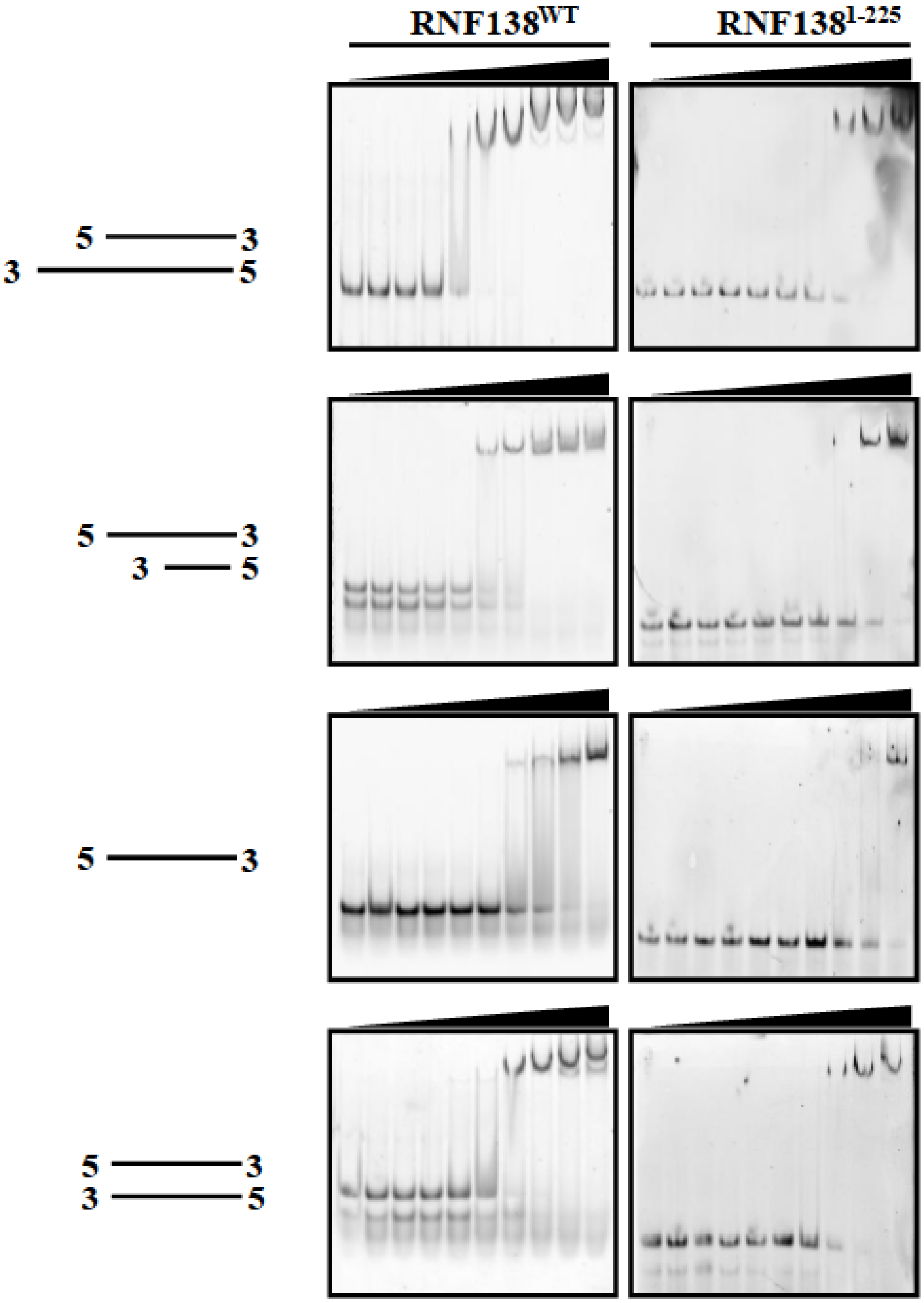


**Figure 3.4. Analyzing RNF138$^{WT}$ and RNF138$^{1-225}$ Binding to DNA Substrates by EMSA.**

The left panel corresponds to RNF138$^{WT}$ and the right to RNF138$^{1-225}$. Top to bottom in both panels the DNA constructs used correspond to 40 base strands of FAM labelled DNA annealed or single stranded to 60 bases (20 bp 3'overhang), 20 bases (5'overhang), 40 base ssDNA, and 40 bp blunt end indicated to the left of each gel row. Increasing concentrations of protein (0/0.02/0.04/0.08/0.2/0.4/0.8/ 2/4/8 μM) indicated by a black wedge above each gel, were titrated against 20 nM of 3' FAM labelled DNA indicated by the asterisk.

### 3.4.3 RNF138 binds DNA in a size-dependent manner

Mre11 endonuclease activity creates large overhangs before the recruitment of subsequent DNA end resection machinery[27]. It has been previously shown that RNF138 binds to DNA on the scale of the minimally resected DNA initiated by Mre11 (~100 base overhang)[129]. We generated a library of DNA in which the lengths of both the overhang and the duplex regions were varied to examine the dependence of the size of the DNA on RNF138 binding. Annealing a 40-nucleotide strand of DNA to a 30 nucleotide FAM labelled DNA (**Figure 3.5**, left column), which halved the size of the 3'overhang from 20 to 10 nucleotides, led to a 4-fold reduction in binding affinity to DNA, with a shift occurring at 800 nM. The same pattern was also observed when a 20-nucleotide strand was annealed to the 30-nucleotide creating a 10 base 5'overhang, with a shift occurring at 2 μM of RNF138. There was no binding observed for the 30-base ssDNA. A 10-fold reduction in binding was noted for the 30-base pair blunt end construct, compared to the 40 base pair duplexes, with a shift observed at 4 μM of RNF138.

The length of the duplexed region was reduced by 10 nucleotides by shortening the single-stranded FAM-labeled DNA from 30 to 20 nucleotides (**Figure 3.5**, center column). This shortened strand was then annealed to a 30-base strand to create a 3' overhang, or to a 10-base strand to create 5' overhangs. When maintaining a 10-nucleotide overhang, the reduction in the duplexed region of DNA resulted in decreased binding affinity for the 3' overhang to 2 μM. Similarly, another 10-base pair reduction in the duplexed region decreased binding to the 5' overhang to 4 μM. No binding was observed with ssDNA, and when the length of the blunt-end DNA was reduced to 20 bases, the binding affinity shifted to 8 μM of RNF138.

Finally, RNF138 showed a markedly lower affinity for both 3' and 5' overhang DNA, with binding affinities at 4 μM, when the overhang regions were halved to 5 nucleotides **(Figure 3.5**, right column). No binding was observed for the 15 base pair double-stranded blunt end DNA. Collectively, these results point to a size dependence for RNF138 DNA binding activity with a consistent preference for larger single-stranded overhang DNA.

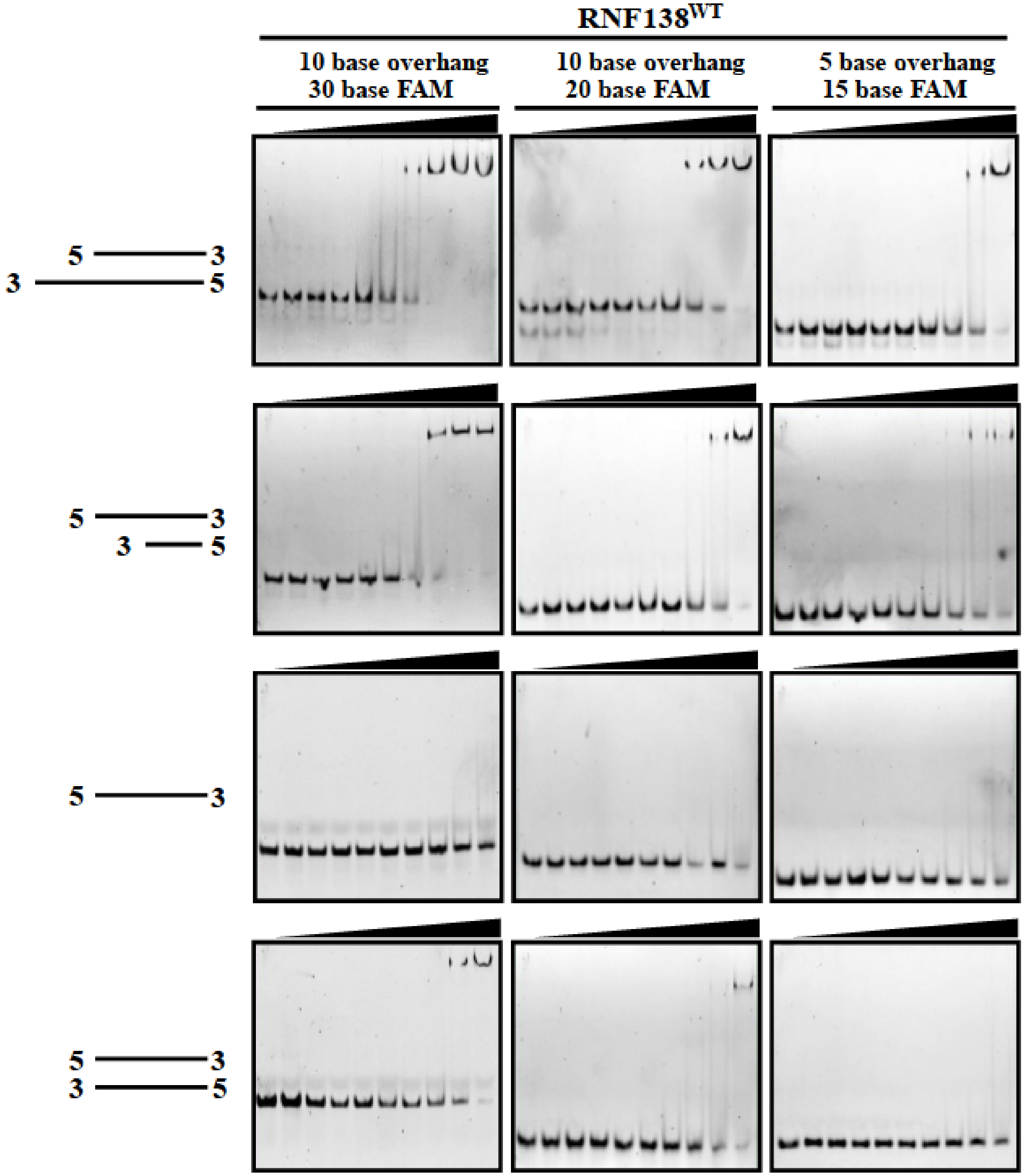


**Figure 3.5. Analyzing RNF138$^{WT}$ DNA binding affinity to smaller DNA constructs.**

Systematic reduction in sizes of DNA were titrated with RNF138$^{WT}$ with increasing concentrations (0/0.02/0.04/0.08/0.2/0.4/0.8/ 2/4/8 μM) indicated by a black wedge above each gel, against 20 nM of 3' FAM labelled DNA indicated by the asterisk on the schematics on the left. The numbers on top of the gels refer to the number of nucleotides of each overhang (top), and the number of nucleotides in each FAM labelled DNA constructs (bottom).The left panel corresponds to

RNF138$^{WT}$ top to bottom in both panels the DNA constructs used correspond top to bottom: 30 base strands of FAM labelled DNA annealed to or single stranded 40 bases (10 bp 3'overhang), 20 bases (5'overhang), 30 base ssDNA, and 30 bp blunt end indicated to the left of each gel row. Middle panel top to bottom: 20 base strands of FAM labelled DNA annealed to or single stranded 30 bases (10 bp 3'overhang), 10 bases (5'overhang), 20 base ssDNA, and 20 bp blunt end. Right panel top to bottom: 15 base strands of FAM labelled DNA annealed to or single stranded 20 bases (10 bp 3'overhang), 10 bases (5'overhang), 15 base ssDNA, and 15 bp blunt end.

### 3.4.4 Significant truncations of RNF138 ablate DNA binding

Zinc fingers are motifs implicated in DNA binding. More specifically, RNF138's zinc fingers have been implicated in the recruitment of RNF138 and retention at DNA DSB[129]. To test for the extent of involvement of the zinc fingers in DNA binding, three more truncations were tested for DNA binding using EMSA: RING ZNF1, ZNF1-3 (ZNF1, 2, 3) and ZNF2,3 **(Figure 3.6)**. Interestingly, the N-terminal region RING ZNF1 construct showed some binding to DNA **(Figure 3.6**, left column), although the pattern observed with each of the DNA constructs signals non-specific binding as it occurs at high protein concentrations and bands were observed to be near the wells of the gel with smearing in other lanes, signalling aggregation. ZNF1-3 contains the ZNF1, L1, and ZNF2, ZNF3 **(Figure 3.6**, center column). Interestingly, the protein appeared to also aggregate, sequestering DNA in the well. This construct was the least stable of the truncations, suggesting a poorly folded protein, consistent with the observed aggregation in EMSA. The last construct we tested, ZNF2,3, had the two C-terminal zinc fingers, ZNF2 and ZNF3, and showed no DNA binding **(Figure 3.6**, right column). All in all, our findings suggest all the domains contribute to RNF138's DNA binding activity, whether directly interacting with DNA or indirectly, via structural stabilization.

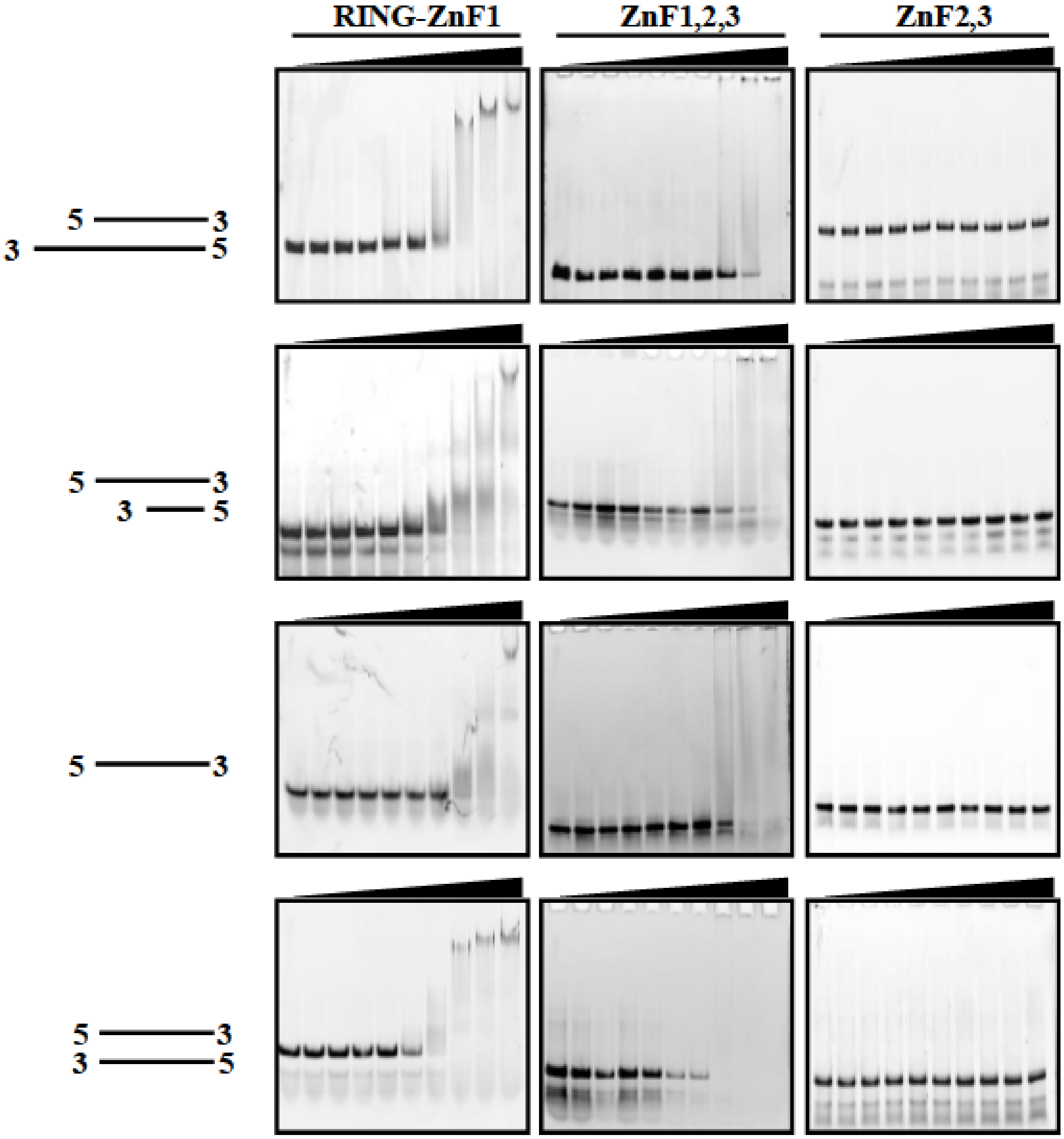


**Figure 3.6. Analyzing RNF138 Truncations DNA Binding Activity.**

RNF138 constructs from left to right: RING ZnF1, ZNF1, ZNF2, ZNF3 and ZNF2,ZNF3. Top to bottom in both panels the DNA constructs used correspond to 40 base strands of FAM labelled DNA annealed to or single stranded 60 bases (20 bp 3'overhang), 20 bases (5'overhang), 40 base ssDNA, and 40 bp blunt end indicated to the left of each gel row. Increasing concentrations of protein (0/0.02/0.04/0.08/0.2/0.4/0.8/ 2/4/8 µM) indicated by a black wedge above each gel, were titrated against 20 nM of 3' FAM labelled DNA indicated by the asterisk.

### 3.4.5 AlphaFold2 Modelling Reveals a Flexible Protein with Localized Structural Domains

Characterizing RNF138's DNA binding activity revealed an involvement of all of the domains of RNF138 which demanded structural characterization to aid us in understanding more about its function. RNF138 remains to be structurally characterized, and ongoing trials have yet to yield an experimental structure. Predicting protein structures solely based on amino acid sequence has been an ongoing problem for many years. Recently, a computational method employing machine learning called AlphaFold2 scored highly in the Critical Assessment of Protein Structure Prediction – an international competition aimed at benchmarking protein folding algorithms – and since has become the gold standard for protein structural prediction[155,158]. We employed AlphaFold2 to glean insights into RNF138's structure alongside crystallographic trials **(Figure 3.7)**. The modelled structure revealed a flexible N-terminal region, followed by a RING domain, L1, and ZNF1 **(Figure 3.7A)**. The RING domain coordinates two zinc ions, with a C2CH configuration for one binding site and a C4 configuration for the other. The C2HC ZNF exhibits a ββα fold, characteristic of ZNF domains, with α-helices from flanking linkers capping both ends **(Figure 3.7A)**. The N-terminal domains are connected to two more zinc fingers, ZNF2 and ZNF3 predicted to be packed against each other, followed by another linker connecting the UIM **Figure 3.7A)**. A closer look at N-terminal RING and ZNF1 reveals an intricate network of hydrophobic and electrostatic and hydrogen bonding interactions, stemming from the RING domain, ZNF1 and some contributions from the linker following ZNF1 **(Figure 3.7B)**. More recently, a partial structure of RNF125, a protein related to RNF138 in size, amino acid sequence, and domain organization, has been characterized[132,134,153]. The N-terminal region of RNF125 was crystallized, revealing a nearly identical network of interactions to RNF138 **(Figure 3.7C)**. The residues involved, specifically methionine 98 (M98 in RNF138, M112 in RNF125) in the center of the hydrophobic interaction, are highly conserved between both proteins, and these residues are also conserved in two other family members related to RN125 and RNF138 – RNF166, and RNF114 **(Figure 3.8)**. It is noteworthy that the construct containing only the three zinc fingers without the RING domain exhibited the lowest stability, suggesting a potential misfolding of the N-terminal domain when it lacks the RING domain. Analysis of the two C-terminal zinc fingers reveals a network of hydrophobic residues emanating from ZNF2, and ZNF3, which were modelled to pack against one another. The residues involved with the intricate network are highly conserved across other family members related to RNF138 **(Figure 3.8)**. The UIM was predicted to be alpha-helical, which is consistent with previously characterized structures.

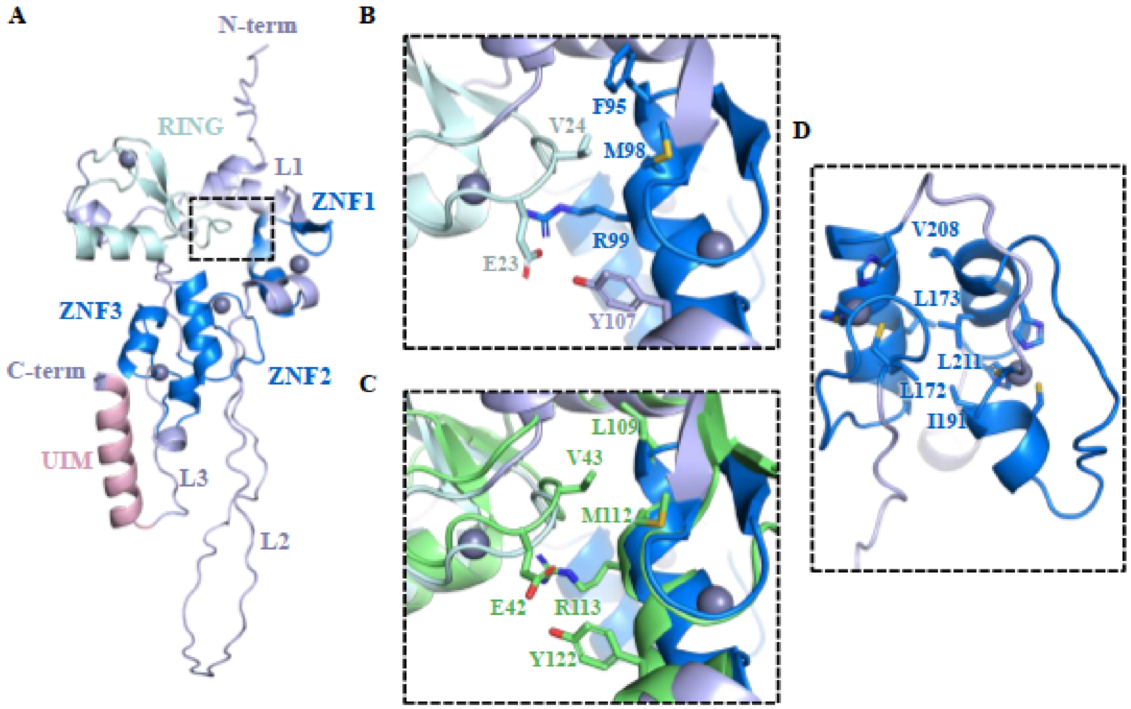


**Figure 3.7. AlphaFold2 Modelling of RNF138$^{WT}$.**

**A.** Highest ranked AlphaFold2 predicted model of RNF138 with an overall predicted template modelling (pTM) value of 0.699. The individual domains are coloured respectively: linkers in white, RING in cyan, zinc fingers (ZNF) in blue, and the ubiquitin interacting motif (UIM) in dark yellow. Zinc ions are modelled as grey spheres. The RING domain folds onto ZNF1 (highlighted by black box), with a linker (L1) in between acting like a hinge. A second longer linker (L2), which is 54 amino acids long joins the N-terminal RING and ZNF1 to ZNF2 and ZNF3, which appear to pack and form a folded structure and are followed by a short linker (L3) and the UIM. **B.** Interdomain interactions involved between the RING domain and ZNF1. M98, and F98 in ZNF1, form hydrophobic interactions with V24. R99 forms electrostatic interactions with the RING E23. Additionally, Y107 from L2 is extended to form h-bonding with the RING domain. **C.** Superimposed crystal structure of RNF125's RING ZNF1 domains in green. Protein data bank (PDB) accession code: 5DKA. Met, and Val are conserved, and Leu replaces Phe in RNF125. The polar charged residues Glu, Arg, and Tyr (from L2) are also conserved in RNF125. **D.** Close-up on the predicted packing of ZNF2, and ZNF3. A network of hydrophobic interactions is observed with L172, and L173 from ZNF2, are packed closely to I191, V208, and L211 of ZNF3.

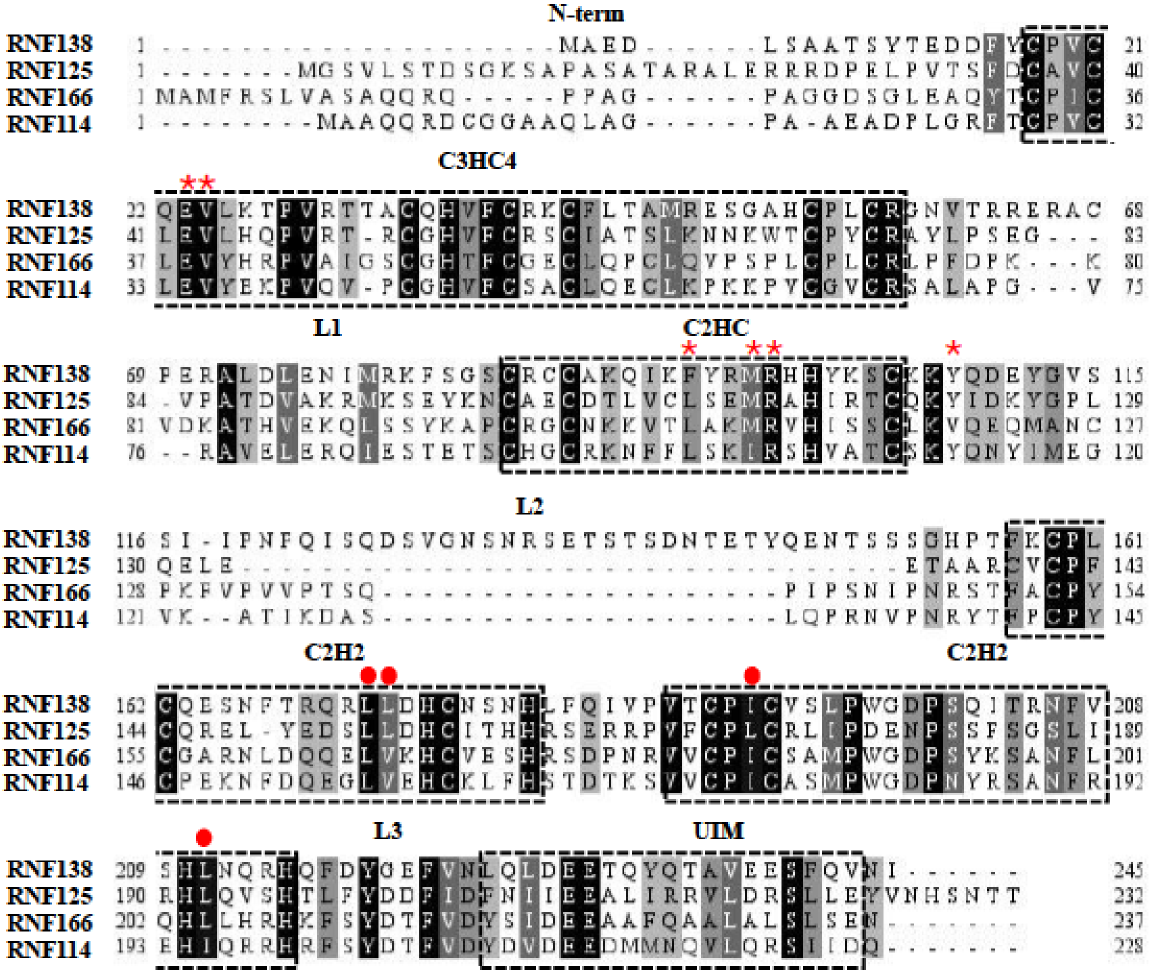


**Figure 3.8. Sequence Alignment of RNF138 with Related Proteins.**

Clustal Omega protein sequence alignment of the RING E3 ligase family which RNF138 is part of. All 4 E3 ligases are similar in size and have similar domains. Different shadings correspond to different thresholds of conservation, black being conserved in all 4 sequences. Dark Grey corresponds conserved mutations with an overall better similarity between the 4 sequences, compared to lighter grey which may include missense mutations. the black boxes highlight the different domains starting with RING domain indicated by C3HC4 (the residues involved in coordinating 2 zinc ions), ZNF1 (C2HC), ZNF2, ZNF3 both coordinating zinc ions with a C2H2 configuration, and the UIM. Red asterisks point to the residues observed in the inter-domain interactions between the RING and ZNF1, in the crystal structures of RNF125 (PDB: 5DKA) and our AlphaFold2 model of RNF138, evidently relatively conserved across all four members. The red dots indicate the hydrophobic core residues between ZNF2, and ZNF3 which were observed to fold closely, also conserved across all four members of the family.

### 3.4.6 AlphaFold Fold Modelling Reveals RING ZNF1 E2 Interaction

Recent innovations in AlphaFold2 have allowed for the prediction of protein complexes, allowing for the structural prediction of protein-protein interactions using the multimer implementation[154,157]. RING domains in RING E3 ligases have a primary function of protein-protein interactions[151]. A well-characterized interaction in RING E3 ligases is the non-covalent interaction with E2 conjugating enzymes, with covalently linked ubiquitin, for substrate transfer[150,151]. RNF138 has been found to interact with several E2 conjugating enzymes, synthesizing different ubiquitin chains[121,130,132]. We utilized AlphaFold-multimer to model RNF138 interaction with UbcH5b, an E2 conjugating enzyme shown to work with RNF138 to ubiquitylate substrates[130] **(Figure 3.9)**. The model reveals the involvement of N-terminal elements in binding to the E2, specifically the RING domain, ZNF1, and some binding contributions from the linker region **(Figure 3.9A)**. A closer look reveals canonical interactions typically involved between the E2 and RING domains, mainly between highly conserved hydrophobic residues on the loop coordinating zinc ions within the RING domain: V20, A46, and P57 with the hydrophobic patch on the E2 enzyme **(Figure 3.9B)**. Additional electrostatic, and hydrogen bonding interactions were observed between Q22 and the highly conserved R58 residue in RNF138 **(Figure 3.9B)**. Non-canonical E2-E3 interactions were also observed, mainly interactions stemming from ZNF1 and the following linker **(Figure 3.9C)**. Specifically, R87 from ZNF1 forms a crucial hydrogen bond with the alpha helix, below the hydrophobic patch of the E2. Three residues from the linker, Y108, Y112, and N111, further contribute to bonding with polar residues on the same helix **(Figure**

**3.9C)**. A structural alignment with RNF138 in comparison to other crystallized E2-E3 ligase complexes reveals a similar overall pattern **(Figure 3.9D)**. Crucial residues involved in the E2-E3 complexes are conserved **(Figure 3.9E)**. More recently the structure of the RING-ZNF1 of RNF125 was crystallized with UbcH5b, showing canonical RING interactions between RNF125, and UbcH5b, and the non-canonical residues as well[153], consistent with AlphaFold modelling of RNF138. In conclusion, modeling of RNF138's N-terminal regions with the E2 UbcH5b, suggest this may form a complex to activate the E2. The role of ZNF1 may extend beyond structurally stabilizing the RING domain and could play a role in direct interaction with the E2, with some residues from the linker coming into contact as well.

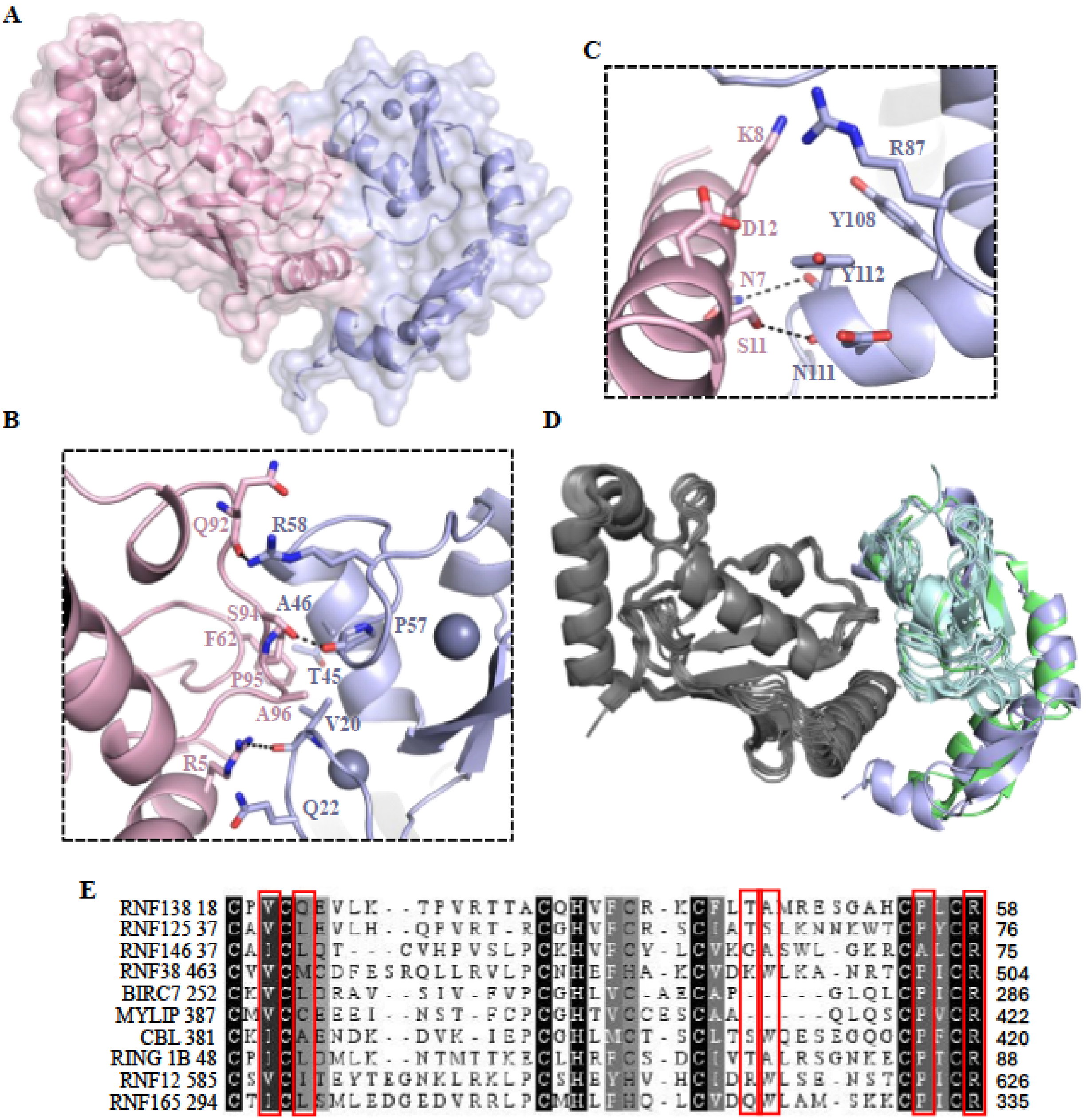


**Figure 3.9. AlphaFold2-Multimer Modelling of RNF138 Interaction with UbcH5b.**

**A.** Highest ranked model shown in surface representation of UbcH5b (pink) bound to RNF138 (purple) with an interface (i) pTM of 0.873. Only the N-terminal region with some residues of L2 are shown for clarity. The overall structure shows a direct interaction of the RING domain, as L1 acts like a hinge, bringing ZNF1 and elements of L2 closer to E2 creating a clamp like structure. **B.** Close up on the canonical RING interactions modelled by AlphaFold. Interactions between V20, P57, and A46 on the RING domain with A96 P85, and F62 on Ubch5b create a hydrophobic core flanked by hydrogen bonding and electrostatic interactions. Mainly: R5 and S94 of Ubch5b hydrogen bond the main chain atoms of P57, and V20 of the RING domain. R58, and Q22 extending from the RING interact with the main chain of Q92, and hydrogen bond with R5 on

UbcH5b, respectively. **C.** Extended non-canonical E2-E3 interactions of charged residues from ZNF1 (R87) and L2 (Y108, Y112, and N111) with polar residues and charged residues (N7, K8, S11, D12) on the helix packed against RNF138. **D.** Structural alignment of monomeric RING E3 ligases (cyan) crystallized with their respective E2s (grey). RNF138 is in purple, and RNF125 (RING ZNF1) crystallized with UbcH5b is in green (PDB accession code: 8GBQ). **E.** Clustal Omega sequence alignment of RING domains structurally aligned with in panel D. Residues observed in the experimental structures and our model are highlighted in the red boxes.

### 3.4.7 Residues Located at the N-terminal of the RING Domain Contribute to Stabilizing the Closed Conformation

To facilitate the transfer of ubiquitin, it's widely acknowledged that RING E3 ligases need to stabilize the closed conformation of the bound E2~Ub conjugate[120]. This involves establishing contacts with both the E2 and the conjugated Ub, positioning ubiquitin in a way which is essential for making the thioester bond between the E2 and ubiquitin open to nucleophilic attack. In a recent study, the RING-ZNF1 domains of RNF125 bound to UbcH5b-Ub, but not the E2 alone, in pulldown experiments suggesting RNF125 interacts with ubiquitin[153].

We utilized AlphaFold-multimer to model interactions of RNF138 E2~Ub **(Figure 3.10)**. The modeling showed interactions between ubiquitin and the RING domain of RNF138, consistent with interactions observed in other RING E3 proteins. Notably, R58, a highly conserved residue found in other monomeric RING E3 ligases, is positioned in close proximity to both the E2 enzyme and ubiquitin **(Figure 3.10A)**. R58 plays a crucial role in stabilizing the closed conformation by interacting with both proteins and positioning the ubiquitin molecule in a conformation that facilitates nucleophilic attack[153]. Furthermore, L56, another highly conserved position in RING domains, packs against I36 in ubiquitin **(Figure 3.10A)**. Additionally, residues N-terminal to the RING domain, such as D15, contact and form electrostatic interactions with K11 on ubiquitin, consistent with similar interactions found between RNF125 and ubiquitin[134,153] **(Figure 3.10A)**. An intriguing observation arises from a structural overlay of previously crystallized RING domains bound to E2~Ub, which revealed a similar pattern of protein-protein interaction with RNF138 (**Figure 3.10C**). Upon closer examination of ZNF1 in our model, we observed its interactions with the N-terminal region of the E2 enzyme. It appeared to wrap around and open the RING-ZNF1 interface to facilitate this interaction (**Figure 3.10B**). This interaction was also observed in the RNF125 experimental structure and was found to be crucial for the activation of the E2 enzyme[153] (**Figure 3.10C,D**). Collectively, our model suggests the N-terminal extension of RNF138, in conjunction with characteristics shared by other RING E3 ligases, works to stabilize the closed conformation of the E2~Ub conjugate, and likely facilitate catalytic activity.

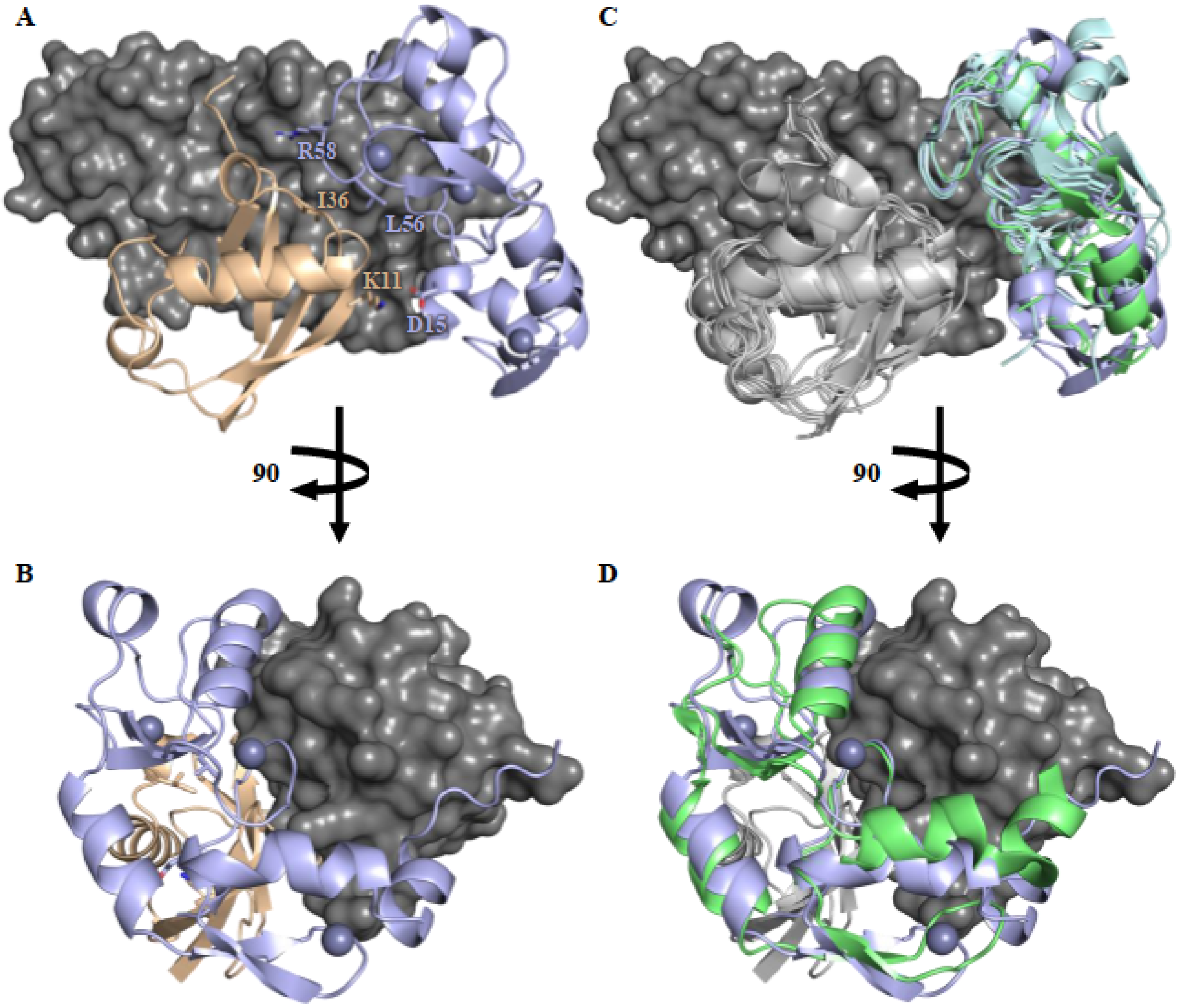


**Figure 3.10. AlphaFold-Multimer Model of RNF138 Bound to E2~Ub.**

**A.** RING ZNF1 of RNF138 (purple) modelled with E2 (grey) and ubiquitin (brown). Conserved R56 interacts with both the E2, and ubiquitin packing against both ipTM 0.885. L56 forms a hydrophobic interaction with I36 hydrophobic patch on ubiquitin. Electrostatic interactions between D15 in the flexible N-terminal region of RNF138, and K11 from ubiquitin are also observed. **B.** 90 rotation of panel **A.** ZNF1 and L2 extend and wrap around to the backside (N-terminal region) of the E2 enzyme suggesting allosteric activation in the presence of the E3 ligase. **C.** Alignment of experimentally determined structures of RING E3 domains (cyan), bound to an E2 protein (surface grey), and ubiquitin (white). RNF138 is in blue, and RNF125 is green. Overall, all structures had an RMSD less than 0.800 Å when aligned on RNF138, suggesting an overall conserved binding interface between E3-E2~ubiquitin. **D.** 90 rotation of panel **C** excluding all structures but RNF125 RING ZNF1-UbchH5b~Ub. The extended interactions with the backside of the E2 enzyme is evident in the experimental structure of RNF125, conserved in our model of RNF138.

### 3.4.8 AlphaFold2 Statistical Metrics and Analysis

AlphFold2 statistical metrics includes pTM and iPTM, which stem from a measure called template modelling (TM) score, which evaluates the precision of the overall protein structure and tends to be less affected by localized inaccuracies[159]. pTM specifically is a comprehensive metric indicating the accuracy of AlphaFold model[156,157]. Overall, a pTM value greater than 0.5 indicates an accurate overall fold of the predicted structure, meaning it might be similar to the “true” structure[156,159]. A score below 0.5 is indicative of a poorly predicted or the prediction is likely wrong[156,159]. Conversely, ipTM evaluates the precision of the predicted positions of the constituent subunits within a protein-protein complex[154]. Scores exceeding 0.8 indicate confident, high-quality predictions, whereas scores below 0.6 indicate potential prediction failure[154]. The predicted alignment error (PAE) and the predicted local distance difference test (pLDDT) are two additional metrics to take into consideration. PAE represents a comprehensive assessment of domain position confidence spanning long distances, computed for each residue in comparison to all others[154–157]. It quantifies the predicted error in Angstroms (Å) for residue x when the model is aligned with the "true" structure at residue y. Importantly, PAE's applicability extends beyond residues within the same monomer, enabling convenient analysis of both intra- and inter-position confidence[154–157]. A smaller PAE value between residues of distinct domains suggests a lower expected error, indicating AlphaFold2's confidence in the positioning of these residues, and vice versa[155]. pLDDT plot is a measure of local confidence of each individual residue[160]. The scale ranges from 0 to 100, where higher scores signify greater confidence and typically a more precise prediction[154–157,160]. Specifically, scores exceeding 70 indicate a high degree of confidence in the positioning of the Cα atom, while scores surpassing 90 indicate a high level of confidence in the arrangement of side chain atoms[154–157,160].

Overall the pTM value for the top ranked model of RNF138 is 0.699 suggesting a correct solution. The median pLDDT score for the whole structure was 88.9, suggesting an overall high local confidence in our model **(Figure 3.11B)**. Specifically, the N-terminal region, had a median pLDDT score of 88.9 **(Figure 3.11B)**. Furthermore, the C-terminal domains had a median pLDDT score of 87.94 **(Figure 3.11B)**, indicative of high local confidence. Interestingly, the median pLDDT score for L2 was 26.1 **(Figure 3.11B)**, consistent with a disordered region, represented as a ribbon in our model. More specifically M98, the residue in the center of the hydrophobic interactions between the RING and ZNF1 had a pLDDT score of 90.9 **(Figure 3.11B)**. Additionally, L212 involved in key interactions between ZNF2, and ZNF3 had a pLDDT score of 88.1 **(Figure 3.11B)**. The high pLDDT scores indicate a high level of confidence in the predicted structure and interactions at these specific sites. The reliability of these predictions suggests that the modeled structure accurately represents the spatial arrangement of these residues. To validate the interaction and positioning of the overall domains, we conducted an analysis of the PAE plot generated from our model **(Figure 3.11A)**. Overall, the median value of PAE for the whole structure was 22.36 Å. However, in the RING and ZNF1 regions of the plot, a median PAE value of 2.82 Å is observed **(Figure 3.11A)**, similarly in the ZNF2, and ZNF3 regions a median PAE error value of 2.12 Å is observed **(Figure 3.11A)**. The low PAE values indicate that there is accurate placement of domains relative to one another in these regions, but overall RNF138 is flexible due to the highly dynamic linker separating the N-terminal and C-terminal domains.

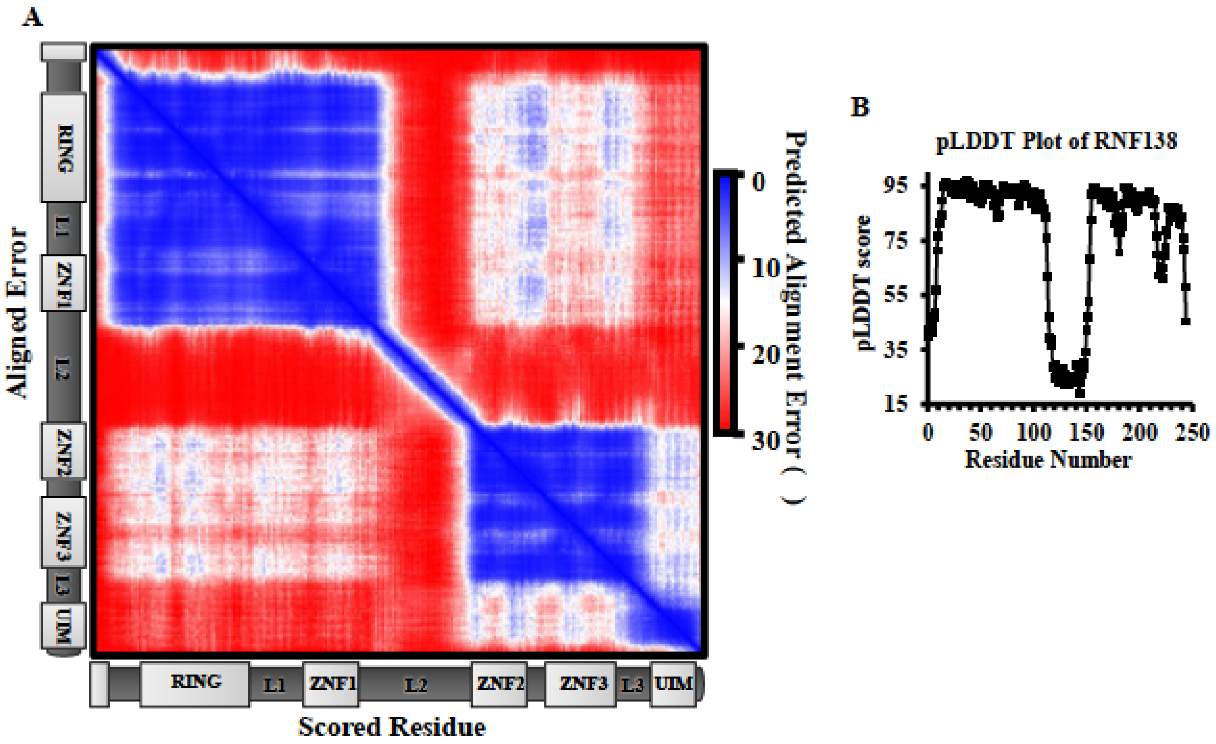


**Figure 3.11. AlphFold2 Confidence Metrics of RNF138**

**A.** A plot illustrating the PAE was generated for the structure of RNF138. PAE serves as a long-range metric for domain position confidence, evaluating the confidence of each residue relative to all others The N-terminal domains of RNF138 exhibited high-confidence PAE values, with a median predicted error of 2.82 Å. Conversely, the flexible linker linking the N-terminal and C-terminal domains showed a high predicted error, with a median pAE value of 22.36 Å, indicating a flexible region within the protein. For the C-terminal ZNF2, ZNF3, the median pAE was 2.12 Å, indicating a high-confidence prediction. PAE is useful for assessing both intra- and inter-position confidence since it does not require residues to belong to the same monomer. Although the N- and C-terminal domains reside within the same monomer and exhibit folding and interaction, they are separated by a large flexible linker. The pAE error between the N-terminal and C-terminal domains is relatively high (16.77 Å) due to the dynamic linker, making it challenging to confidently predict residues involved in their interaction. Overall, this modeling suggests that RNF138 is a flexible and dynamic protein susceptible to post-translational modification. **B.** pLDDT plot of RNF138 the model. The model was confidently predicted overall, as indicated by a median pLDDT score of

88.9. Residues within the N-terminal RING domain, as well as ZNF1, ZNF2, ZNF3, and UIM consistently showed high pLDDT values (>80). However, L2 exhibited low scores (<50), suggesting disorder.

When bound to UbcH5b and ubiquitin, RNF138 had a median pLDDT value of 63.8 **(Figure 3.12B)**. The local confidence plot shows reduction in confidence of the C-terminal domains. Overall, the N-terminal domain maintained high pLDDT value, with a median score of 92.9, suggesting high local confidence **(Figure 3.12B)**. Interestingly crucial residues in the RING domain, ZNF1, and L2 (D15, V20, Q22, T45, A46, P57, and R58, R87, Y108, Y112, and N111) identified to be part of the interface with UbcH5b and ubiquitin had median pLDDT values of 92.3. The high pLDDT values associated with these residues signify a high level of confidence in their predicted positions within the protein structure. The PAE plot generated for RNF138 bound to UbcH5b, and ubiquitin revealed the structured regions RING ZNF1, and ZNF2, ZNF3 had median PAE values of 2.92 Å and 9.23 Å **(Figure 3.12A)**, suggesting a strong prediction of the N-terminal domains of RNF138. The binding of UbcH5b distorted the positions of ZNF2, ZNF3 as AlphaFold2 is unable to model the disordered second linker reliably, therefore increasing error in the model. UbcH5b and ubiquitin were both predicted accurately with median PAE values of 2.35 Å and 1.89 Å respectively **(Figure 3.12C,D,B)**. Importantly, the median PAE of 4.44 Å observed for the RING, ZNF1, L2, and UbcH5b highlights the precision of their spatial arrangement within the predicted model **(Figure 3.12B)**. Additionally, the median PAE between RING ZNF1 L2, and Ubiquitin is 4.39 Å **(Figure 3.12B)**. Finally, the median PAE value between Ubc5Hb and ubiquitin is 3.91 Å **(Figure 3.12B)**. Collectively the low median PAE error values combined with an iPTM score of 0.885 suggest a highly accurate model of the interactions between N-terminal regions of RNF138 with UbcH5b and ubiquitin.

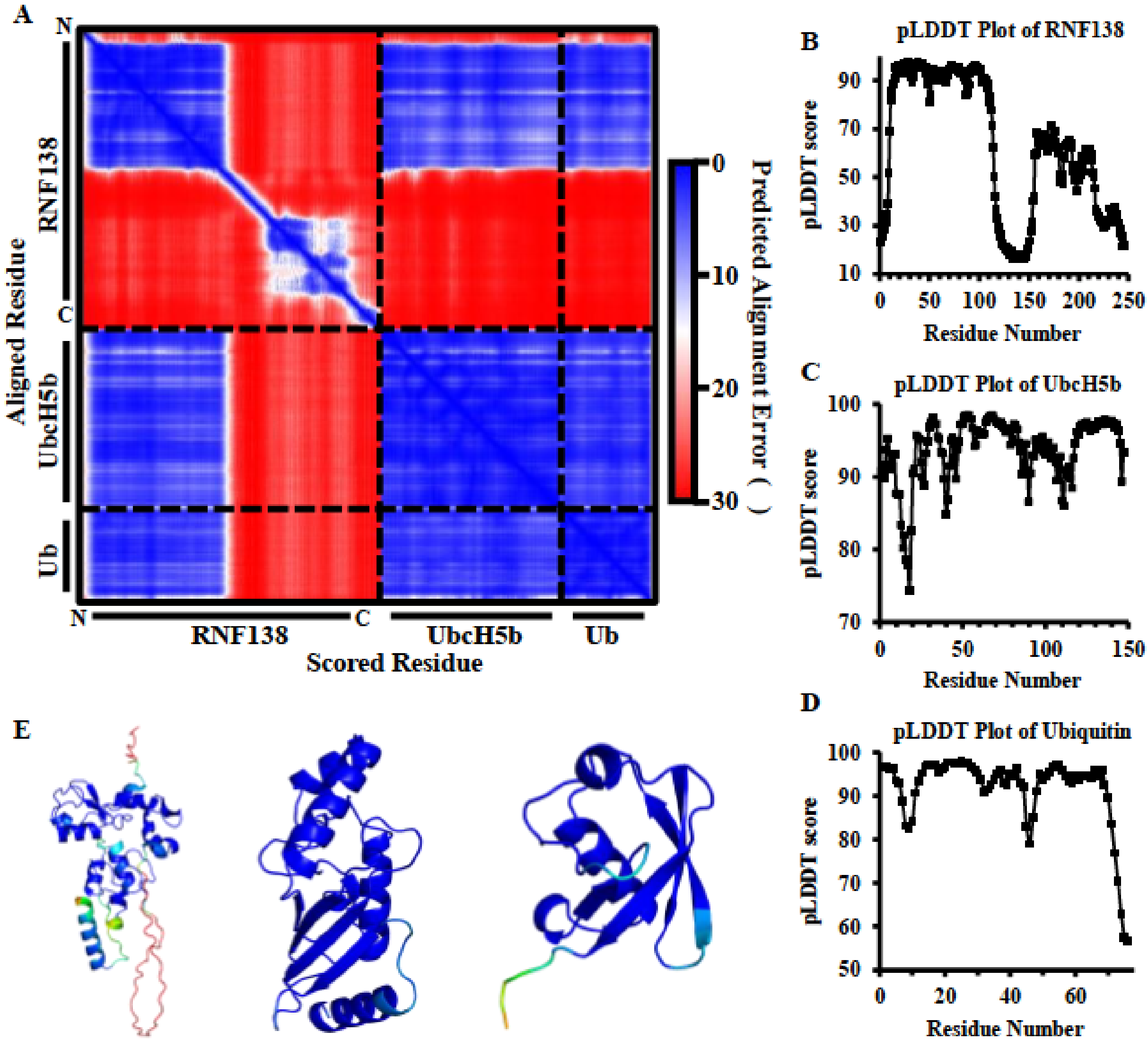


**Figure 3.12. Confidence Metrics of the Modelled Proteins: RNF138, UbcH5b and Ubiquitin**

**A.** PAE plot for the predicted structure of RNF138, UbcH5b, and ubiquitin. The quadrans corresponding to each aligned and scored structure are labelled besides, or below respectively. pLDDT plots of RNF138$^{WT}$, UbcH5b, and ubiquitin **(A., B., and C.)** respectively. **E.** From left to right: RNF138, UbcH5b, and ubiquitin coloured according pLDDT scores. The colors dark blue and red represent pLDDT scores indicating high confidence (>90) and low confidence (<50) respectively.

## 3.5 Discussion

In this study, we assessed the ability of RNF138 to bind different DNAs relevant to DNA damage repair, and to understand the contributions of different protein domains to these interactions. Purifying RNF138 and the various truncations proved to be challenging, but these challenges were circumvented by varying buffer conditions, and keeping concentrations lower until more is required, which then fresh protein was used promptly. Both RNF138WT and RNF138$^{1-225}$ showed DNA binding, with RNF138$^{WT}$ binding observed at 200 nM and RNF138$^{1-225}$ binding at 4 μM when incubated with 3'overhang DNA, suggesting strong involvement of the entire protein in DNA binding activity. Supershifting was observed at higher concentrations of RNF138, suggesting multiple RNF138 molecules might bind these DNA substrates. Interestingly our results showed lower binding affinity relative to previously studies of the DNA binding of RNF138. This could be attributed to the larger DNA substrates used in the previous study. Our DNAs in general contained ~20 base single-stranded regions and ~40 bp duplexed regions, while the previous study used 50 base single-stranded DNA, with a 50 bp duplexed region. Additionally, the difference could be due to the different expression systems. The previous study used insect-cell expressed RNF138, which might have additional post-translational modifications, or may be better folded than our *E. coli*-expressed material[129]. Our Expi293F expression system was our attempt at circumventing this discrepancy, but we were unable to express useable amounts of RNF138 in this way. Further truncation of RNF138 strongly diminished DNA binding, suggesting the involvement of all RNF138's domains in DNA binding activity. The systematic reduction in the size of the DNA substrate further diminished the DNA binding affinity. We initially thought the zinc fingers would likely bind DNA regardless of the presence or absence of other domains, but this is not the case. Likely, there is direct involvement of several different elements within RNF138 responsible for DNA binding, and in addition, there are likely structural contributions from the other domains required for overall protein stability which are not directly involved in DNA binding.

Insights from AlphaFold2 models about the overall structure of RNF138 reveal a highly flexible protein with localized ordered domains. RING E3 ligases, such as TRAF6, are known in certain cases to work as dimers. No evidence of dimerization was observed during protein purification, and AlphaFold failed to model a dimeric structure therefore we proceeded to treat RNF138 as a monomeric RING E3 ligase. The N-terminal region of RNF138 has an intricate network of interactions between the RING domain and ZNF1. The residues involved in the interactions were conserved across related proteins RNF125, RNF166, and RNF114. Interestingly, the experimental structure of RNF125 RINGZNF1 reveals a nearly identical pattern of binding between the RING domain, and ZNF1, suggesting an important structural role of ZNF1 and overall stabilization of the RING domain. Due to the flexibility in L2, the exact overall structure of the C-terminal regions is not confidently modelled, but we can comment on ZNF2,3 which were modelled to be packed against each other with conserved hydrophobic residues facilitating the interaction.

The binding interaction modelled by AlphaFold2 between the RING domain and the E2 enzyme mirrors similar interactions observed in experimentally characterized structures of RING-E2 complexes. Interestingly, the recently published structure of RNF125 RING-ZNF1, and E2 revealed the important canonical interactions of the RING domain with the E2. Those interacting residues are conserved in RNF138 and were observed in the AlphaFold2 model of the RNF138-E2 complex. In addition to the RING domain, non-canonical interactions were also observed with the ZNF1 and L2. Our model is consistent with previous experimental findings in the structure RNF125-UbcH5b, suggesting an important role of the N-terminal region in catalytic activity. Interestingly it was found that the RING alone is not enough for the catalytic activity of RNF125, further implicating ZNF1 and residues in L2 in the catalytic activity of RNF138[134,153]. Like

RNF138, two extensively studied RING-E3s, LNX1 and TRAF6, each feature at least one ZNF positioned C-terminally to the RING domain[161–163]. In both LNX1 and TRAF6, the RING domains form dimers, while the C-terminal ZNF does not tightly interact with the RING domain. Instead, the ZNF located C-terminal to the RING domain interacts with the ubiquitin molecule to maintain the activated state of the E2~Ub conjugate bound to the RING domain of its dimeric counterpart. LNX1 is particularly noteworthy as it not only contains a C-terminal ZNF domain but also harbours another ZNF domain N-terminal to the RING domain[162,163]. In this structural arrangement, these two ZNFs pack together, extending the RING dimer interface and stabilizing the activated state of the conjugate. Interestingly, RNF138 as a monomeric RING E3 ligase, interacts with the ubiquitin and E2 through a conserved R58 residue, predicted to be the linchpin residue stabilizing and positioning ubiquitin for nucleophilic attack[153]. Another conserved residue is the L56, which interacts with the hydrophobic patch on ubiquitin. Both are canonical RING ubiquitin interactions observed in other monomeric RING E3 ligases[120]. An interesting observation is the importance of region N-terminal of the RING domain, where important electrostatic interactions were observed in our model contributing to the interaction with ubiquitin. In addition to the activation of UbcH5b by interaction with the E3, a number of studies have shown that the interaction of ubiquitin with the N-terminal region of the E2, often referred to as the backside-binding site, significantly enhances activity[153,164–166]. The ubiquitin molecule that binds to the backside of UbcH5b enhances activity, at least in part, because it restricts the flexibility of UbcH5b[153]. This form of allosteric regulation warranted a closer look at non-RING elements of RNF138 as from the AlphaFold model an extended interaction between RNF138 and UbcH5b is observed. Upon rotating the structural model 90 degrees, it became evident that ZNF1 is extended and wraps around the N-terminal region of the E2. Interestingly, this was also observed in the experimental structure of the RNF125 N-terminal region with E2. Biochemical characterization further revealed this extended interaction facilitate E2 activity. Mutations in residues contacting the E2 from ZNF1, lead to a reduction in catalytic activity. Indeed, the back-side interaction of non-RING elements with E2 has been observed in other E3 ligases, which is the case for the RING E3s A07 and gp78[165,166]. In addition to the core RING interaction with the E2, they have an additional helix which interacts with the backside N-terminal region of the E2 enzyme. Collectively, our model underscores the importance of the N-terminal region of RNF138 in catalytic activity, and likely overall structure of RNF138. Through the overall structural stabilization of RNF138, which may mediate DNA binding and retention on DSBs.

## 3.6 Acknowledgements

We are grateful for Dr. Brian Lee, and Dr. Leo Spyracopoulos for the DNA constructs for both the protein constructs and initial set of oligonucleotides used in our EMSA studies.

## 3.7 Funding

The work in the J.N.M.G. laboratory is funded by the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery Grant RGPIN-2016-05163) and the Canadian Institutes of Health Research (CIHR168972). I was also supported by Novartis Pharmaceuticals Canada Inc Graduate Scholarship, Queen Elizabeth II Graduate Scholarship, The Department of Biochemistry Doctoral Recruitment Scholarship, Alberta Graduate Excellence Scholarship, and David Lawson Award.

## 3.8 Conflict of Interest

The authors declare no conflicts.

## 3.9 Author Contributions

J.N.M.G and R.A.F wrote and edited the manuscript. R.A.F conducted all the experiments, with S.C. assisting in protein purification and EMSA experiments. R.A.F and A.J.L established the mammalian expression system and purified mammalian expressed proteins. L.S. supplied the DNA constructs of RNF138 and the oligonucleotides used in the initial experiments. I.H.I. provided essential insights and supervised A.J.L. J.N.M.G conceived the project, secured funding, and provided overall supervision.

# Chapter 4 – The Role of RNF138 in DNA End Resection is Regulated by Ubiquitylation and CDK Phosphorylation

## 4.1 Abstract

Double-strand breaks (DSBs) are DNA lesions that pose a significant threat to genomic stability. The repair of DSBs by the homologous recombination (HR) pathway is preceded by DNA end resection, the 5' to 3' nucleolytic degradation of DNA away from the DSB. We and others previously identified a role for RNF138, a RING finger E3 ubiquitin ligase, in stimulating DNA end resection and HR. Yet, little is known about how RNF138's function is regulated in the context of DSB repair. Here, we show that RNF138 is phosphorylated at residue T27 by cyclin-dependent kinase (CDK) activity during the S and G2 phases of the cell cycle. We also observe that RNF138 is ubiquitylated constitutively, with ubiquitylation occurring in part on residue K158 and rising during the S/G2 phases. Interestingly, RNF138 ubiquitylation decreases upon genotoxic stress. By mutating RNF138 at residues T27, K158, and the previously identified S124 ATM phosphorylation site (Han *et al.*, 2016, ref. 22), we find that post-translational modifications at all three positions mediate DSB repair. Cells expressing the T27A, K158R, and S124A variants of RNF138 are impaired in DNA end resection, HR activity, and are more sensitive to ionizing radiation compared to those expressing wildtype RNF138. Our findings shed more light on how RNF138 activity is controlled by the cell during HR

## 4.2 Introduction

DNA double-strand breaks (DSBs) occur when both DNA strands are severed in close proximity and are considered the most dangerous DNA lesion[167]. They are mainly repaired by two major mechanisms, namely homologous recombination (HR) and non-homologous end joining (NHEJ). HR typically uses the sister chromatid as a template for repair[168]. As the sister chromatid is generated upon DNA replication, this restricts HR activity to the S and G2 phases of the cell cycle[169,170]. HR commences with DNA end resection, the 5' to 3' nucleolytic degradation of DNA away from the DSB[171]. The process generates 3' single-stranded DNA (ssDNA) overhangs which are rapidly coated by RPA (replication protein A) complexes[171]. End resection is initiated by the nuclease activities of the MRN (Mre11-Rad50-Nbs1) complex[171], the Mre11 endonuclease activity being activated upon binding to CtIP[172–174]. The overhangs are further extended by the exonucleases Exo1 and Dna2[171], after which RPA is exchanged for the Rad51 recombinase. Rad51 activity then drives the search for the homologous locus in the sister chromatid, initiating strand invasion and the eventual restoration of the site of damage by DNA synthesis[168].

If DSBs are instead repaired by NHEJ, the two DNA ends are directly ligated together once they are made to be chemically compatible by end processing[22]. NHEJ requires binding of the DNA ends by the Ku70-Ku80 (Ku) heterodimer[175], which serves as a platform to assemble the NHEJ machinery[22,176,177]. NHEJ is active during all phases of the cell cycle[170], but is especially important in G1 phase as that is when HR is not active. The decision of whether to perform NHEJ or HR is a dynamic process governed by multiple decision points[178]. End resection biases cells to performing HR, as the resulting ssDNA overhangs are not amenable to ligation by NHEJ, and Ku itself has low affinity to ssDNA[175]. However, Ku recruits to DSBs regardless of the cell cycle phase[129], and its presence on chromatin is a block to end resection[24,179]. Thus, the removal of Ku is required for both end resection and HR to proceed[24,179].

A recurring theme in the regulation of DSB repair is post-translational modification (PTM), the reversible covalent conjugation of protein or chemical groups onto biomolecules. PTMs include phosphorylation, catalyzed by kinases such as the cyclin-dependent kinases (CDKs)[180], and by members of the PI-3-kinase-related kinase (PIKK) family, such as ATM (ataxia telangiectasia mutated), which is activated by DNA damage[38]. Ubiquitylation also plays major roles in the DSB response[181]. Here, the ubiquitin protein is conjugated to its target substrates through sequential activity of three classes of enzyme, E1, E2 and E3. We and others have shown that the E3 ubiquitin ligase RNF138 promotes HR[35,129–131]. Originally found to inhibit Wnt-β-catenin signaling[121], RNF138 belongs to a family of E3s with similar domain structure[132]. This includes an N-terminal RING (Really Interesting New Gene) finger domain, which interacts with the E2 ubiquitin conjugating enzyme[121,130,132], three zinc finger (ZNF) domains, and a C-terminal ubiquitin interacting motif (UIM). The zinc finger domains mediate its recruitment to DNA damage [129–131] and bind DNA[129,130], showing preference for ssDNA overhangs[129]. Mechanistically, RNF138 promotes HR by stimulating DNA end resection[129,130]. It promotes the ubiquitylation of Ku80, which evicts Ku from chromatin[129]. It also mediates the ubiquitylation of CtIP, facilitating CtIP's accumulation at DSB sites[130]. These parallel actions – promoting the recruitment of CtIP, a stimulator of resection, and facilitating the dissociation of Ku, which blocks resection – help ensure end resection can proceed[182]. Downstream of end resection, a third target of RNF138-dependent ubiquitylation was found to be Rad51D[35,131], a Rad51 paralogue that may contribute to Rad51 filament assembly[183]. Although it is not clear how Rad51D ubiquitylation contributes to HR[35], the recruitment of Rad51D to DNA damage is dependent on RNF138[131].

While it is established that RNF138 participates in HR, how its activity is controlled in HR has not been extensively investigated. Intriguingly, in response to IR, RNF138-dependent ubiquitylation of Ku80 occurs in S/G2 phase, but not G1 phase.[129] This hints RNF138 activity might be under cell cycle regulation to coincide with the onset of HR. We also wondered if RNF138 could be regulated by ubiquitin conjugation. In this study, we find that RNF138 is phosphorylated in S and G2 phase by CDK activity on residue T27, and is also ubiquitylated on residue K158. Both sites mediate RNF138 function in DSB repair, as cells expressing the T27A and K158R mutants exhibit defective DNA end resection, HR activity, and heightened sensitivity to ionizing radiation relative to those expressing wildtype RNF138. We also investigate whether the same processes are impacted by mutations at S124, a previously identified ATM phosphorylation site on RNF138[131]. Our findings uncover how RNF138 activity is governed by the cell, providing more insight into its contribution to DSB repair.

## 4.3 Methods

### 4.3.1 DNA Constructs and siRNAs

The FLAG-RNF138 vector (containing the full-length RNF138 ORF and a single FLAG tag (DYKDDDDK) directly C-terminal to it, within the AbVec2.0 expression vector) was a gift from Michael Hendzel (University of Alberta). pEGFP-RNF138-WT and -ΔUIM (with the RNF138 ORF C-terminal of the GFP tag) were generated previously[129]; the ΔUIM mutant contains residues 1-228 of full-length RNF138. pCDNA3-HA-ubiquitin-WT and -L73P (both containing residues 2-76 of ubiquitin) plasmids were gifts from Tony T. Huang (New York University School of Medicine)[184]. pcDNA4-TO-hygromycin-sfGFP-MAP was a gift from Dannel McCollum (Addgene plasmid #44100; http://n2t.net/addgene:44100; RRID:Addgene_44100)[185]. The siRNA-resistant RNF138 ORF was inserted into the pcDNA4-TO-hygromycin-sfGFP-MAP vector between the superfolder GFP (sfGFP)-N175 and His8 modules by GenScript Biotech (Piscataway, New Jersey, USA), producing the pcDNA4-TO-hygromycin-sfGFP-RNF138-MAP (abbreviated sfGFP-RNF138) construct. The siRNA-resistant RNF138 ORF was inserted into the pmCherry-C1 vector by Biomatik Corporation (Kitchener, Ontario, Canada), producing the mCherry-RNF138 construct. The sfGFP-RNF138-T27A, -S124A, and -K158R mutants were generated by GenScript Biotech (Piscataway, New Jersey, USA). All other mutants were generated using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs) according to the manufacturer's instructions. Sequences were verified by Sanger sequencing performed by the Molecular Biology Service Unit (Dept. of Biological Sciences, University of Alberta). DNA primers used for Sanger sequencing **(Table 4.1)** and site-directed mutagenesis **(Table 4.2)** and siRNA **(Table 4.3)** were custom synthesized by Sigma-Aldrich (St. Louis, Missouri, USA).

**Table 4.1. Primers for DNA Sequencing.**

| Name | Sequence (5' – 3') | Construct | Direction |
|---|---|---|---|
| RNF138 21 For | GGCCACGTCCTACACCGA | any containing RNF138 | forward |
| RNF138 210 For | ACGGGCCTTAGACCTTGAAA | any containing RNF138 | forward |
| RNF138 169 Rev | ATAGGGGACAATGTGCTCCG | any containing RNF138 | reverse |
| FLAG-RNF138 845 Rev | TGAAATTTGTGATGCTATTGCTTT | FLAG-RNF138 | reverse |
| pEGFP-C1 For | CATGGTCCTGCTGGAGTTCGTG | GFP-RNF138 | forward |
| pEGFP-C1 Rev | CAGGTTCAGGGGGAGGTGTGG | GFP-RNF138 | reverse |

**Table 4.2 Primers for Q5 Site-Directed Mutagenesis (New England Biolabs).**

| Mutation | Templates | Sequences of Primer Pairs (5' – 3') | Annealing Temp. |
|---|---|---|---|
| **RNF138-T27A** | FLAG-RNF138-WT,<br>GFP-RNF138-WT,<br>mCherry-RNF138-WT | Forward: GGTGCTCAAAgcgCCCGTGCGGA<br>Reverse: TCCTGACAGACGGGGCAGTAG | 72°C |
| **RNF138-T27E** | GFP-RNF138-WT,<br>sfGFP-RNF138-WT,<br>mCherry-RNF138-WT | Forward: GGTGCTCAAAgagCCCGTGCGGAC<br>Reverse: TCCTGACAGACGGGGCAG | 67°C |
| **RNF138-K158R** | GFP-RNF138-WT,<br>mCherry-RNF138-WT/-T27A/-S124A/-T27A-S124A | Forward: CCTACTTTTaggTGTCCCCTG<br>Reverse: ATGACCAGAAGAACTTGTATTC | 60°C |
| **RNF138-S124A** | FLAG-RNF138-WT,<br>GFP-RNF138-WT,<br>mCherry-RNF138-WT/-T27A | Forward: CTTTCAGATCgctCAAGATTCAG<br>Reverse: TTTGGAATGATAGAAGAAACAC | 58°C |
| **RNF138-S124E** | sfGFP-RNF138-WT,<br>mCherry-RNF138-WT | Forward: CTTTCAGATCgagCAAGATTCAGTAGGG<br>Reverse: TTTGGAATGATAGAAGAAACAC | 57°C |

**Table 4.3. siRNAs.**

| Name | Target | Sense Sequence (5' – 3') |
|---|---|---|
| **siCTRL** | Luciferase<br>(negative control) | CGUACGCGGAAUACUUCGA |
| **siRNF138** | RNF138<br>coding region (20) | CCAAACUGCUGUUGAAGAA |
| **siRNF138 UTR** (used only in Fig. S1A) | RNF138 3' UTR (sequences were selected by Dharmacon Inc.) | The following siRNAs pooled together:<br>GGAGGGAAUUGUAUUGAUA<br>AAAGAGUGGUGUUUACUAU<br>GGGAAUAGGGAUAGACUUU<br>AGCCAUACAUCUUAAUGAA |
| **siCDK1** | CDK1 | CCUAGUACUGCAAUUCGGGAAAUUU |
| **siCDK2** | CDK2 | CCUAUUCCCUGGAGAUUCUGAGAUU |

### 4.3.2 Cell Lines, Tissue Culture, and Transfection of Nucleic Acids

All cells were maintained at 37°C in a humidified atmosphere containing 5% $CO_2$. Unless indicated, all cells were cultured in low glucose Dulbecco's Modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Gibco), 50 units/mL penicillin and 50 µg/mL streptomycin (both Gibco). Cells approaching confluency were detached using trypsin-EDTA solution, 0.25% (Sigma-Aldrich) prior to subculture. All cell lines were tested for *Mycoplasma* using DAPI staining. HeLa cells were a gift from Alfred C.O. Vertegaal (Leiden University). The HeLa HB-ubiquitin cell line[186] was a gift from Peter Kaiser (University of California, Irvine). HEK293 cells were a gift from Michael Hendzel (University of Alberta). U2OS cells stably expressing doxycycline-inducible I-*Sce*I and the DR-GFP reporter (TRI-DR-U2OS) were a gift from Philipp Oberdoerffer (Johns Hopkins University)[187]. U2OS cells stably integrated with FRT (flippase recognition target) sites and the TetR tetracycline repressor-expressing pcDNA6/TR vector (U2OS-TREx cells) were a gift from Armin Gamper (University of Alberta). The Flp recombinase (Flp-In) system was not exploited to generate stable cell lines in U2OS-TREx. Instead, pcDNA4-TO-hygromycin-sfGFP-RNF138-MAP constructs were stably integrated into U2OS-TREx cells upon transient transfection and selection in DMEM supplemented with 10%

charcoal-stripped FBS (Sigma-Aldrich) and 200 μg/mL hygromycin B (Invitrogen). To enrich for cells expressing the sfGFP-RNF138 constructs, 5 μg/mL doxycycline (Sigma-Aldrich) was added to the culture medium for 20-24 hours, then GFP$^+$ cells were isolated by fluorescence-activated cell sorting (FACS). U2OS-TREx cells stably expressing RNF138 constructs were maintained in DMEM with 10% FBS, 150 to 200 μg/mL hygromycin B, and 10 μg/mL blasticidin S (Gibco). Unless indicated otherwise, plasmid DNA was transfected into cells using Effectene Transfection Reagent (Qiagen) according to the manufacturer's instructions 18-24 hours before assays were performed. siRNA was transfected into freshly seeded cells using Lipofectamine RNAiMax Transfection Reagent (Invitrogen) once ~48 hours before assays were to be performed. Unless indicated otherwise, siRNA **(Table 4.3)** to RNF138 was transfected at a final concentration of 60 nM, while siRNAs to CDK1 and CDK2 were transfected at a final concentration of 50 nM. Control siRNA targeting luciferase was transfected at the same final concentration as the targeted siRNA.

### 4.3.3 Cell Cycle Synchronization

HeLa cells were synchronized by the double thymidine block method. Thymidine (4 mM final concentration) was added to the culture medium of asynchronous cells at ~40% confluency for 16-18 hours (block #1). The cells were then washed twice with room temperature sterile PBS, replaced with warmed DMEM + 10% FBS, and incubated at 37°C (release). 7-8 hours post-release, cells were transfected with DNA constructs, if necessary. 2-3 hours post-transfection (or 9-11 hours post-release), thymidine (4 mM final concentration) was again added to the culture medium and kept for 12-14 hours (block #2). At this point the cells were considered synchronized to the G1/S transition. To allow synchronous progression through the cell cycle, the cells were released by 2 washes with ice-cold PBS followed by 37°C incubation in warm DMEM + 10% FBS. Cells were then harvested at various timepoints post-release to enrich for specific cell cycle phases (e.g. 3 hours for S phase, 7 hours for G2 phase, 11 hours for G1 phase).

### 4.3.4 Harvesting Cells

For HEK293, the cells were dislodged from the culture vessel by flushing the surface with the culture medium. Cells were then pelleted by centrifugation at 525*g* for 5 minutes at 4°C. The other cell lines were dislodged via trypsinization: they were washed twice in ice-cold PBS (137 mM NaCl, 2.7 mM KCl, 10 mM $Na_2HPO_4$, 1.8 mM $KH_2PO_4$, pH 7.4), detached with trypsin-EDTA solution, 0.25% (Sigma-Aldrich) at 37°C for 5 minutes, resuspended into 4 volumes of ice-cold DMEM + 10% FBS, and pelleted by centrifugation at 525*g* for 5 minutes at 4°C. All cells were then resuspended into ice-cold PBS and centrifuged again (525*g*, 5 minutes, 4°C). After removing the supernatant, the pellet was flash frozen in liquid nitrogen before storage at -80°C.

### 4.3.5 Cell Cycle Profiling by Flow Cytometry

If needed, this procedure was performed during the above cell harvesting method. When cells were resuspended in PBS after the first centrifugation, 10-20% of the cell pellet was saved. To this fraction, much of the supernatant was removed, after which the cells were vortexed into an ice-cold mixture of PBS prepared with 70% ethanol as the solvent. The cells were fixed by -20°C incubation for at least 30 minutes. The cells were washed once in PBS, then tumbled end-over-end for 30 minutes at room temperature in PBS containing 100 µg/mL RNase A (Invitrogen) and 3.8 mM sodium citrate. Propidium iodide was added to a final concentration of 50 µg/mL, and the cells were again tumbled end-over-end at room temperature for at least 30 minutes. The propidium iodide intensity was then measured for single cells by flow cytometry using a FACSCanto II (BD Biosciences).

### 4.3.6 Cell Treatments

All inhibitors were purchased from Millipore-Sigma or Selleck Chemicals and dissolved in DMSO (or, for hydroxyurea, water). Inhibitors were diluted in warmed (37°C) culture medium immediately prior to cell treatment. Vehicle controls contained only the solvent of the inhibitors diluted to the same extent. Unless indicated otherwise, cells were treated with the following concentrations of inhibitors: 25 µM roscovitine, 2.5 µM AZD5438, 10 µM RO-3306, 10 µM SB203580. For treatment with ultraviolet light (UV), cell monolayers were washed in PBS, which was then removed, and exposed to 20 seconds of UV (equivalent to ~60 $J/m^2$). The culture medium was quickly re-added and cells were incubated at 37°C for 1 hour after which they were harvested. For treatment with ionizing radiation, cells were exposed to 10 Gy from a $^{60}Co$ source (Gammacell 220 Irradiation Unit, purchased 1978, Atomic Energy of Canada Limited), allowed to recover for 1 hour at 37°C, then harvested. For the remaining DNA damaging agents, cells were replaced with culture medium containing the agents and incubated at 37°C (1 µM camptothecin for 1 hour; 25 µM phleomycin for 1 hour; 2 mM hydroxyurea for 4 hours), then harvested.

### 4.3.7 Preparation of Whole Cell Extracts

Frozen pellets (obtained from the above cell harvesting protocol) were resuspended into ice-cold High SDS Lysis Buffer (25 mM HEPES pH 7.4, 500 mM NaCl, 2% sodium dodecyl sulfate, 1% Triton X-100, 0.5% sodium deoxycholate, 1 mM EDTA) supplemented with 2X cOmplete protease inhibitor cocktail, EDTA-free (cOmplete, Roche) and 1X phosSTOP phosphatase inhibitor cocktail (phosSTOP, Roche). The mixture was then sonicated with a Fisher Scientific Model 705 Sonic Dismembrator with microtip probe (at amplitude 1 to 5, for 1 minute). 4X SDS Sample Buffer (250 mM Tris pH 6.8, 8% sodium dodecyl sulfate, 40% glycerol, 0.2% bromophenol blue) was then added to attain a final concentration of 1X, while 2-mercaptoethanol

(BME) was added to a final concentration of 5%. The samples were treated at 95°C for 5 minutes at 900 rpm on a ThermoMixer F1.5 (Eppendorf) prior to resolution by SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis).

### 4.3.8 SDS-PAGE and Immunoblotting

Samples were loaded into hand-cast mini-gels comprising Tris (37.5 mM, pH 8.8 for the resolving layer; 12.5 mM, pH 6.8 for the stacking layer), 0.1% sodium dodecyl sulfate (SDS), and 5-12% polyacrylamide. Precision Plus protein dual color standards (Bio-Rad) were loaded as the molecular weight ladder. Electrophoresis was performed at 150 V in Running Buffer (25 mM Tris pH 8.3, 192 mM glycine, 0.1% SDS). The resolved proteins were then wet electro-transferred onto 0.2 µm nitrocellulose membrane for 1 hour at 110 V in Transfer Buffer (25 mM Tris pH 8.3, 192 mM glycine, 20% methanol). To perform immunoblot, the nitrocellulose was first blocked in 4% fish skin gelatin (FSG) dissolved in TBS (50 mM Tris pH 7.5, 150 mM NaCl) at room temperature. TBS with 5% bovine serum albumin (BSA) was used as the blocking solution when phospho-specific antibodies were used as the primary antibody. The primary antibodies were diluted in freshly prepared TBS + 0.1% Tween-20 (TBST). For phospho-specific antibodies, 5% BSA was included in this solution, while for the anti-RNF138 antibodies, 2% FSG was included. The diluted antibodies were incubated with the membranes for either 1 hour at room temperature or overnight at 4°C under gentle rocking. The membranes were then shaken in TBST (3 times, 10 minutes each), incubated for 1 hour at room temperature with horseradish peroxidase (HRP)-, IRDye 680RD- or IRDye 800CW-conjugated secondaries (all LI-COR Biosciences) in TBST under gentle rocking, and shaken in TBST (3 times, 10 minutes each) and TBS (once, 10 minutes). HRP activity was detected by incubating the membranes in Amersham ECL Prime Western Blotting Detection Reagent (Cytiva) for 2 minutes. Enhanced chemiluminescence (HRP) or fluorescence (IRDye) signals were acquired on the Odyssey Fc Imaging System and quantified by densitometry with Image Studio software (both LI-COR Biosciences). If re-probing was required, membranes were treated with Stripping Buffer (100 mM glycine pH 2.2, 1% SDS) for 30 minutes with vigorous shaking, rinsed with distilled water, and air-dried overnight. They were then re-blocked and probed with the necessary primary and secondary antibodies. To ensure immunoblots for loading controls of whole cell extracts (actin, tubulin) and immunoprecipitations (anti-GFP on GFP-RNF138, anti-FLAG on FLAG-RNF138) could be quantified without saturation, in such situations primary antibodies were used at low concentrations **(Table 4.4)** and blots were detected via fluorescence (IRDye) instead of enhanced chemiluminescence.

**Table 4.4. Antibodies.**

| **Antibody** | **Source** | **Catalogue Number** | **Identifier** | **Dilution (Application)** |
|---|---|---|---|---|
| rabbit anti-β-actin | Sigma | A5060 | RRID:AB_476738 | 1:10000 – 1:20000 for 1 hour (IB) |
| mouse anti-BrdU | Cytiva Life Sciences | RPN202 | RRID:AB_2314032 | 1:1000 overnight (IF) |
| mouse anti-CDK1 | Santa Cruz Biotechnology | sc-54 | RRID:AB_627224 | 1:1000 overnight (IB) |
| rabbit anti-CDK2 | Santa Cruz Biotechnology | sc-163 | RRID:AB_631215 | 1:1000 overnight (IB) |
| rabbit anti-phospho-Chk1 (Ser345) | Cell Signaling Technology | 2348 | RRID:AB_331212 | 1:5000 for 1 hour (IB) |
| rabbit anti-phospho-Chk2 (Thr68) | Cell Signaling Technology | 2197 | RRID:AB_2080501 | 1:1000 overnight (IB) |
| rabbit anti-Cyclin A | Santa Cruz Biotechnology | sc-751 | RRID:AB_631329 | 1:1000 overnight (IB) |
| mouse anti-FLAG tag | Millipore Sigma | F1804 | RRID:AB_262044 | 1:4000 – 1:8000 for 1 hour (IB) |
| rabbit anti-FLAG tag | Cell Signaling Technology | 14793 | RRID:AB_2572291 | 1:4000 – 1:8000 for 1 hour (IB) |
| rabbit anti-Geminin | Cell Signaling Technology | 5165 | RRID:AB_10623289 | 1:1000 overnight (IB) |
| mouse anti-GFP | Santa Cruz Biotechnology | sc-9996 | RRID:AB_627695 | 1:2000 – 1:6000 for 1 hour (IB) |
| rabbit anti-GFP | Proteintech | 50430-2-AP | RRID:AB_11042881 | 1:5000 for 1 hour (IB) |
| rabbit anti-phospho-H2AX (Ser139) | Cell Signaling Technology | 2577 | RRID:AB_2118010 | 1:3000 for 1 hour (IF) |
| rabbit anti-HA tag | Abcam | ab9110 | RRID:AB_307019 | 1:4000 – 1:8000 overnight (IB) |
| rabbit anti-RNF138 | St John's Laboratory | STJ112342 | RRID:AB_2938982 | 1:1000 – 1:4000 overnight (IB) |
| rabbit anti-RNF138 | Abcam | ab92730 | RRID:AB_2238719 | 1:1000 overnight (IB) |

| | | | | |
|---|---|---|---|---|
| (used only in Fig. S3A) | | | | |
| mouse anti-RPA2 | Abcam | ab2175 | RRID:AB_302873 | 1:8000 for 1 hour (IF)<br>1:2000 overnight (IB) |
| rabbit anti-phospho-RPA2 (Ser4/Ser8) | Bethyl Laboratories | A300-245A | RRID:AB_210547 | 1:4000 – 1:12000 overnight (IB) |
| rabbit anti-phospho-Ser/Thr (P-S/TQ; ATM/ATR substrate) | Cell Signaling Technology | 2851 | RRID:AB_330318 | 1:1000 overnight (IB) |
| mouse anti-phospho-Thr-Pro (P-TP) | Cell Signaling Technology | 9391 | RRID:AB_331801 | 1:500 – 1:1000 overnight (IB) |
| mouse anti-α-Tubulin | Genscript | A01410 | RRID:AB_1968943 | 1:20000 – 1:25000 for 1 hour (IB) |
| mouse anti-Ubiquitin | Santa Cruz Biotechnology | sc-8017 | RRID:AB_2762364 | 1:1000 overnight (IB) |
| donkey anti-mouse IgG—IRDye 680RD | LI-COR Biotechnology | 926-68072 | RRID:AB_10953628 | 1:20000 for 1 hour (IB) |
| donkey anti-rabbit IgG—IRDye 680RD | LI-COR Biotechnology | 926-68073 | RRID:AB_10954442 | 1:20000 for 1 hour (IB) |
| donkey anti-mouse IgG—IRDye 800CW | LI-COR Biotechnology | 926-32212 | RRID:AB_621847 | 1:20000 for 1 hour (IB) |
| donkey anti-rabbit IgG—IRDye 800CW | LI-COR Biotechnology | 926-32213 | RRID:AB_621848 | 1:20000 for 1 hour (IB) |
| goat anti-mouse IgG—HRP | LI-COR Biotechnology | 926-80010 | RRID:AB_2721263 | 1:5000 for 1 hour (IB) |
| goat anti-rabbit IgG—HRP | LI-COR Biotechnology | 926-80011 | RRID:AB_2721264 | 1:5000 for 1 hour (IB) |
| goat anti-mouse IgG—Cy3 | Jackson ImmunoResearch Laboratories | 115-165-146 | RRID:AB_2491007 | 1:500 for 1 hour (IF) |
| goat anti-rabbit IgG—Alexa Fluor 647 | Invitrogen | A-21244 | RRID:AB_2535812 | 1:250 for 1 hour (IF) |

### 4.3.9 FLAG Immunoprecipitation

Pellets of FLAG-RNF138-expressing HeLa cells from a 100 mm dish were resuspended into ice-cold NETN-500 (50 mM Tris pH 8.0 at 4°C, 500 mM NaCl, 0.5% IGEPAL CA-630, 1 mM EDTA) supplemented with fresh 2X cOmplete, 1.25X phosSTOP, and 50 mM N-ethylmaleimide (NEM) and shaken on ice (250 rpm, 20 minutes). The lysate was clarified by centrifugation at 20000*g* for 15 minutes at 4°C, and the resulting pellet was discarded. 10% of the supernatant was saved as an input control and mixed with an equal volume of 2X SDS Sample Buffer (125 mM Tris pH 6.8, 4% sodium dodecyl sulfate, 20% glycerol, 0.1% bromophenol blue) along with BME to a final concentration of 5%. The input control was then heated to 95°C for 5 minutes at 900 rpm on a ThermoMixer F1.5. The remaining supernatant (90%) was diluted to reduce the NaCl concentration to 150 mM, then mixed with 20 µL of anti-FLAG M2 magnetic beads (Sigma) that were pre-washed twice in ice-cold TBS + 0.5% IGEPAL CA-630 (TBSN). The mixture was tumbled end-over-end for 2 hours at 4°C. Non-specific binding was removed by 4 washes in ice-cold TBSN. Each wash involved vortexing for 20 seconds, centrifuging at 2700*g* for 2 minutes, placing the sample on a magnetic rack, and aspirating the supernatant. Bound proteins were eluted off the beads by adding 2X SDS Sample Buffer and heating for 10 minutes at 95°C at 1200 rpm (ThermoMixer F1.5). The eluate and input control fractions were then processed for SDS-PAGE and immunoblotted. For the eluate fraction, the modification of interest (e.g. phosphorylation) was blotted for first, then the membrane was stripped and re-probed to detect immunoprecipitated FLAG-RNF138.

### 4.3.10 GFP Immunoprecipitation

Pellets of GFP construct-expressing HEK293 cells from a 100 mm dish were resuspended into ice-cold RIPA Buffer (50 mM Tris pH 7.4 at 4°C, 150 mM NaCl, 1% IGEPAL CA-630, 1% sodium deoxycholate, 0.1% SDS, 1 mM EDTA) supplemented with fresh 2X cOmplete, 1.25X phosSTOP, and 50 mM NEM and shaken on ice (250 rpm, 20 minutes). The lysate was centrifuged (20000*g*, 15 minutes, 4°C), and the pellet discarded. 10% of the supernatant was taken out as an input control and processed as in the FLAG Immunoprecipitation procedure (above). The remaining supernatant (90%) was mixed with 15 µL of GFP Selector agarose beads (NanoTag Biotechnologies, Göttingen, Germany) that were pre-washed twice in ice-cold RIPA Buffer, then tumbled end-over-end at 4°C for 1 hour. To remove non-specific binding, the beads were washed twice in ice-cold RIPA Buffer, then 4 times in ice-cold Stringent Wash Buffer (50 mM Tris pH 8.0 at 4°C, 2 M NaCl, 1% IGEPAL CA-630, 0.25% sodium deoxycholate, 0.1% SDS, 1 mM EDTA).

Each wash involved vortexing for 20 seconds, centrifuging at 3000*g* for 2 minutes, and aspirating the supernatant. Bound proteins were eluted by adding to the beads 2X SDS Sample Buffer with 5% BME and heating on the Thermomixer (95°C, 30 minutes, 1200 rpm). Both the input and eluate fractions were then subjected to SDS-PAGE and immunoblot analysis. For the eluate fraction, the modification of interest (e.g. ubiquitylation) was blotted for first, then the membrane was stripped and re-probed to detect the immunoprecipitated GFP-tagged protein.

### 4.3.11 Co-Immunoprecipitation

Pellets of GFP construct-expressing HEK293 cells from a 100 mm dish were processed as in the procedure for GFP Immunoprecipitation (above), with the following changes: NETN-150 (50 mM Tris pH 8.0 at 4°C, 150 mM NaCl, 0.5% IGEPAL CA-630, 1 mM EDTA) was used in place of RIPA Buffer, 20 µL of GFP Selector agarose beads (NanoTag Biotechnologies) were used, and the beads were washed 4 times with only NETN-150 to remove non-specific interactions.

### 4.3.12 CDK2 *in vitro* Kinase Assay

HEK293 cells were transfected with DNA constructs encoding GFP, GFP-RNF138-WT, or GFP-RNF138-T27A (2 150 mm dishes per construct) and harvested the next day. The cell pellets were then resuspended into ice-cold RIPA Buffer (as above) supplemented with fresh 2X cOmplete and 1X phosSTOP and shaken on ice (250 rpm, 20 minutes), after which the lysates were clarified by centrifugation (20000*g*, 15 minutes, 4°C) and tumbled end-over-end at 4°C for ~3-4 hours with 20 µL of GFP Selector agarose beads that had been pre-washed twice in RIPA Buffer. The beads were then washed three times in ice-cold RIPA Buffer and three times in ice-cold TBSN buffer (as above). Washes were performed as per the GFP Immunoprecipitation procedure (above). Any residual TBSN above the beads was then removed, and the bead suspension for each construct was evenly split into two portions of ~13 µL each. Each portion was promptly mixed on ice with components from a CDK2 Assay Kit (catalogue number 79599, BPS Bioscience, San Diego, CA, USA) to a final volume of 45 µL: 6 µL of 5X kinase assay buffer 1, 1 µL of 500 µM ATP, 5 µL of sterile double-distilled water, and 20 µL of either CDK2/Cyclin A2 mixture (concentration of stock: 2.5 ng/µL in 1X kinase assay buffer 1) or 1X kinase assay buffer 1. The reaction mixtures were then incubated at 30°C (300 rpm, ThermoMixer F1.5). The reaction was quenched 45 minutes later by adding 45 µL of 2X SDS Sample Buffer and heating the mixture to 95°C for 5 minutes (900 rpm, ThermoMixer F1.5). Upon SDS-PAGE, the relative kinase activity of CDK2 was detected by immunoblotting for phospho-Thr-Pro (P-TP) and GFP.

### 4.3.13 Isolation of Ubiquitin Conjugates by Nickel Affinity Purification

HeLa HB-ubiquitin cells from 150 mm dishes were harvested as described above, except prior to pelleting for flash freezing, the cells were resuspended into ice-cold PBS and 10% of each sample was taken out to serve as the input control. All samples were then pelleted by centrifugation (525*g*, 5 minutes, 4°C) and flash frozen in liquid nitrogen before storage at -80°C. The 10% input control was processed separately for lysis to prepare whole cell extract (as above) while the remainder was processed for nickel affinity purification, described here. All buffers were prepared at most 4 hours before use. All washes entailed vortexing in the indicated buffer for at least 20 seconds, centrifugation at 750*g* for 2 minutes, and removal of the supernatant via vacuum aspiration. Cell pellets were dissociated into ice-cold Guanidine Lysis Buffer (6 M guanidine-HCl, 100 mM sodium phosphate buffer pH 8.0, 10 mM Tris, 5 mM imidazole, 5 mM BME) by vortex and sonicated for 1 minute (amplitude 25, Fisher Scientific Model 705 Sonic Dismembrator with microtip probe). The lysate was then mixed with 150 µL of Ni-NTA agarose beads (Qiagen) that were pre-washed three times in Guanidine Lysis Buffer. The mixture was agitated on a rocker for 4 hours at room temperature. Non-specific interactions were removed with sequential washes at room temperature: once in Guanidine Wash Buffer (6 M guanidine-HCl, 100 mM sodium phosphate buffer pH 8.0, 10 mM Tris, 10 mM imidazole, 0.1% Triton X-100, 5 mM BME), once in pH 8 Urea Wash Buffer (8 M urea, 100 mM sodium phosphate buffer pH 8.0, 10 mM Tris, 10 mM imidazole, 0.1% Triton X-100, 5 mM BME), and three times in pH 6.3 Urea Wash Buffer (8 M urea, 100 mM sodium phosphate buffer pH 6.3, 10 mM Tris, 0.1% Triton X-100, 5 mM BME). Bound proteins were eluted off the beads with Ni-NTA Elution Buffer (150 mM Tris pH 6.7, 200 mM imidazole, 5% sodium dodecyl sulfate, 30% glycerol, 0.05% bromophenol blue, 5% BME) at 60°C for 30 min on a ThermoMixer F1.5 (Eppendorf) at 1200 rpm and resolved by SDS-PAGE for subsequent immunoblotting.

### 4.3.14 Cycloheximide Chase Assay

HEK293 cells were transfected with the indicated GFP-RNF138 DNA constructs. 20 hours later, cycloheximide was added to the culture medium to 100 µg/mL with or without MG132 (5 µM final concentration). Cells were incubated for various timepoints up to 16 hours and then harvested. Whole cell extracts were prepared, and protein expression levels were analyzed by fluorescence-based immunoblotting, with GFP intensity normalized to the intensity of actin.

### 4.3.15 Laser Microirradiation of Live Cells

U2OS cells were seeded on 35 mm glass-bottom dishes (MatTek Corporation). 24 hours later, they were transfected with 200 ng of the indicated DNA construct and incubated for ~16 hours. Prior to imaging, cells were pre-sensitized with 0.5 µg/mL Hoechst 33258 for 30 min at 37°C, washed with PBS, and replaced with warmed phenol red-free DMEM containing 25 mM HEPES + 10% FBS (both Gibco). The dishes were maintained at 37°C in a humidified 5% $CO_2$ atmosphere while being imaged on a Nipkow spinning disk confocal system (UltraVIEW ERS, PerkinElmer) mounted on an Axiovert 200M inverted microscope (Zeiss) and equipped with an sCMOS camera (Prime BSI, Photometrics). Localized DNA damage was induced in a single 1 µm thick line spanning the width of the cell nucleus using a 5 mW 405 nm diode laser coupled to a FRAP (fluorescence recovery after photobleaching) module (UltraVIEW Photokinesis, PerkinElmer) with the following settings: 20% power output, 20 iterations. The GFP fusion proteins were then excited with a 488 nm argon laser and seen through a 63X, 1.4 numerical aperture oil immersion DIC Plan-Apochromat objective lens (Zeiss). Time-lapse images in the GFP emission channel were recorded using Volocity 6.3 software (PerkinElmer). The fluorescence intensity at the laser stripe over time was determined via Image J software. Measurements from 30-75 cells pooled from three independent experiments were averaged.

### 4.3.16 Immunofluorescence Staining

Stable cell lines generated in U2OS-TREx were seeded onto sterilized glass coverslips (#1.5 thickness, Electron Microscopy Sciences) and transfected with siRNA 1 hour later. ~28 hours post-transfection, 5 µg/mL doxycycline was added to the culture medium for ~16-20 hours to induce sfGFP-RNF138 expression. After a 1-hour treatment with 1 µM camptothecin, the cells were washed twice with ice-cold PBS, then incubated twice, 3 minutes each, with ice-cold RPA Extraction Buffer (25 mM HEPES pH 7.9 at 4°C, 300 mM sucrose, 50 mM NaCl, 0.5% Triton X-100, 1 mM EDTA, 3 mM $MgCl_2$). They were then washed again twice with ice-cold PBS prior to fixation at room temperature for 20 minutes in 2% paraformaldehyde in PBS. The reaction was quenched for 10 minutes in 100 mM $NH_4Cl$ in PBS, and the cells were permeabilized for 5 minutes in PBS + 0.5% Tween-20. The cells were then incubated with primary antibodies for 1 hour at room temperature or overnight at 4°C, incubated in PBS + 0.1% Tween-20 for 5 minutes, washed 6 times in PBS, and incubated with secondary antibodies for 1 hour at room temperature. All antibodies were diluted in PBS, and antibody incubations were performed with the coverslips being inverted into 75 µL droplets of the antibody solution. The cells were then counterstained in 10 ng/µL DAPI in PBS for 20 minutes, washed 6 times in PBS, and mounted onto microscopy slides in Mounting Medium (2% propyl gallate in PBS prepared with 10% DMSO and 80% glycerol as the solvent). Images were acquired on a fluorescence microscope (Zeiss AxioImager.Z2) using version 7.10.4 of MetaMorph software (Molecular Devices). The microscope utilized a 1.4 numerical aperture 40X oil immersion DIC M27 Plan-Apochromat objective lens (Zeiss) and Prime BSI sCMOS camera (Teledyne Photometrics). Extraction-resistant nuclear foci were quantified using the Cell Intensity Mean of Vesicles feature of the Statistics module of Imaris x64 software version 9.9.1 (Oxford Instruments).

### 4.3.17 In vivo Homologous Recombination (DR-GFP) Reporter Assay

Per condition, ~4 X $10^6$ TRI-DR-U2OS cells were electroporated with 60 nM siRNA and if necessary, 2 µg of mCherry-RNF138 DNA via a 4D-Nucleofector X Unit (program CM-104) with SE Cell Line 4D-Nucleofector X Kit L (both Lonza Bioscience). 8 hours post-transfection, 1 µg/mL doxycycline (Dox) was added to the culture medium for 24 hours to induce expression of I-*Sce*I. The culture medium was then replaced and cells were cultured without doxycycline for another 24 hours. Cells were collected according to the above cell harvesting procedure except that instead of being flash frozen, the cells were vortexed into 2% paraformaldehyde in PBS and incubated for 20 minutes for fixation. The cells were then washed 3 times in PBS. The frequency of $GFP^+$ cells was measured by flow cytometry (FACSCanto II, BD Biosciences) on at least 100,000 cells.

### 4.3.18 Clonogenic Survival Assay

Stable cell lines of U2OS-TREx were seeded in 100 mm dishes and transfected with siRNA 1 hour later. ~40 hours post-transfection, the cells were detached by trypsinization and kept on ice.

The cell density was measured on an automated cell counter (Corning) equipped with CytoSMART Cloud (CytoSMART Technologies). Cells were then seeded in duplicate onto 60 mm dishes and incubated at 37°C with or without 5 µg/mL doxycycline for ~16 hours. The number of cells seeded per dish were as follows: parental + siCTRL – 900 cells, parental or sfGFP-RNF138-WT + siRNF138 – 4000 cells, all other cell lines + siRNF138 – 12000 cells. The dishes were exposed to up to 5 Gy of ionizing radiation from a $^{60}$Co source (Gammacell 220 Irradiation Unit, purchased 1978, Atomic Energy of Canada Limited), after which the culture medium was replaced. The cells were incubated at 37°C for 7-10 days; 5 µg/mL doxycycline was included in the medium if sfGFP-RNF138 expression was required. The medium was then removed and cells were fixed and stained in 0.5% Crystal Violet/25% methanol. Colonies of ≥50 cells were counted. The surviving fraction was calculated by dividing the number of colonies formed at a given dose by the number that formed at 0 Gy.

### 4.3.19 Sequence Alignment

Amino acid sequences of RNF138 orthologues were obtained from UniProt, aligned with Clustal Omega (European Bioinformatics Institute, European Molecular Biology Laboratory), and annotated in ESPript 3.0[188].

### 4.3.20 AlphaFold Modeling

RNF138 modeling and predictions were performed using the ColabFold implementation of AlphaFold (https://github.com/sokrypton/ColabFold)[154–157]. The AlphaFold_MMseqs2 Google Colab notebook (version 1.5.2) was used as previously described[189]. Confidence metrics were plotted with Microsoft Excel (version 2204) and Morpheus (https://software.broadinstitute.org/morpheus).

### 4.3.21 Image and Data Processing

Raw micrographs and immunoblot scans were adjusted with the Levels tool and cropped in Adobe Photoshop 2023 and 2021, respectively. For immunofluorescence micrographs, images from the same biological replicate were scaled to identical settings. For GFP immunoprecipitation assays, GFP-RNF138 ubiquitylation was quantified from densitometry readings of immunoblots with Image Studio software (LI-COR Biosciences). HA signal from immunoprecipitates was normalized to the GFP signal immunoprecipitated. All graphs were generated in Prism 9 (GraphPad) and display the mean with error bars showing the standard deviation. Unless indicated otherwise, ordinary one-way ANOVA with Šídák's multiple comparisons test was performed to determine statistical significance. Asterisks depict statistically significant differences: ns (not significant), ($p \leq 0.05$), ($p \leq 0.01$), ($p < 0.001$), ($p < 0.0001$). Figures were arranged and labeled using Adobe Illustrator 2023.

## 4.4 Results

### 4.4.1 RNF138 Protein Expression is Maintained Over the Course of the Cell Cycle

To ascertain how RNF138 is regulated, we first asked if its expression was controlled in a cell cycle-dependent manner. As the expression of the HR factor BRCA1 peaks during the S and G2 phases[190–192], we surmised RNF138 protein levels could behave similarly, coinciding with its role in mediating Ku80 ubiquitylation and eviction from chromatin in S/G2[129]. We chose to examine RNF138 expression in HeLa cells as they can be efficiently synchronized to the G1/S transition by double thymidine block. HeLa can then be released for different timepoints to approach specific cell cycle phases (30). Flow cytometric analysis confirmed that the chosen timepoints sufficiently enriched for cells in either S, G2 or G1 phase **(Figure 4.1A)**. When whole cell extracts from these samples were immunoblotted, we detected a prominent immunoreactive band above 25 kDa in all cell cycle phases **(Figure 4.1B)**. As RNF138's molecular weight is predicted to be 28 kDa, and transfecting cells with short interfering RNA (siRNA) targeting both coding and non-coding regions of the *RNF138* gene reduced detection of the band **(Figure 4.2A)**, the immunoblot signal just above 25 kDa represents endogenous RNF138. Interestingly, while a minor increase in RNF138 expression was seen at G2 phase, RNF138 was still adequately expressed in G1 phase, and overall, substantial changes in expression were not seen at any particular phase **(Figure 4.1C)**. We thus conclude that in HeLa cells, RNF138 protein expression is relatively constant over the course of the cell cycle.

### 4.4.2 RNF138 is Phosphorylated at Residue T27 by CDK-Dependent Activity

We reasoned RNF138 activity might be regulated during the cell cycle by a different mechanism. Transitions in the cell cycle are controlled by CDKs, whose activities accumulate at specific phases, and are stimulated when bound to their regulatory subunits, the cyclins[44,193]. CDK activity promotes HR and DNA end resection[45,194,195], and players in DNA end resection such as Mre11, Nbs1, CtIP, Dna2, and Exo1 are indeed targets of CDK phosphorylation[196–203]. Analysis of RNF138's primary structure revealed a single putative CDK consensus phosphorylation motif (S/T-P-X-K/R)[43], with the potential phosphorylation site at threonine 27 (T27). This motif is conserved in RNF138 orthologues spanning *Xenopus laevis*, chicken, and mammals **(Figure 4.1D)**, and was predicted to be the only CDK site on RNF138 by the algorithm GPS 6.0 [204] **(Figure 4.2B)**. Further, the AlphaFold prediction of RNF138's structure shows T27, despite being located within the RING domain, is solvent-accessible and potentially available for phosphorylation **(Figure 4.8A-E)**. T27 is also at the centre of a positively charged surface of the RING (K26, R40, K41, R48, R80) **(Figure 4.8B)**, the addition of a phosphoryl group potentially altering electrostatics in the region. We thus hypothesized RNF138 is phosphorylated at residue T27 by CDK activity.

To uncover if RNF138 was phosphorylated, we first transfected HeLa cells with FLAG-tagged RNF138, and performed anti-FLAG immunoprecipitation to enrich for exogenous RNF138. Expressing FLAG-RNF138 yielded a protein that migrated at 37 kDa upon sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) **(Figure 4.1E)**, despite RNF138 being predicted to be 28 kDa. While endogenous RNF138 migrated at the expected position upon SDS-PAGE **(Figure 4.1E,Figure *4.2*A)**, immunoprecipitated FLAG-RNF138 was detected at 37 kDa by an anti-RNF138 antibody **(Figure 4.4A)**. We thus attribute the reduced electrophoretic mobility of FLAG-RNF138 to the FLAG tag, speculating the tag's five negatively charged aspartate residues hinder SDS binding to the protein, impeding its migration during electrophoresis. We next immunoblotted the FLAG immunoprecipitates for the presence of phospho-threonine immediately N-terminal of a proline residue (phospho-Thr-Pro, or P-TP), a motif shared by the substrates of both CDKs and mitogen-activated protein kinases (MAPKs). We observed P-TP signal on FLAG-RNF138 **(Figure 4.1E)**, suggesting RNF138 was a substrate for phosphorylation at TP sites. To detect if RNF138 was phosphorylated in a cell cycle-dependent manner, we transfected HeLa cells with FLAG-RNF138 and synchronized them to S, G2, or G1 phase. The cells were suitably enriched for the phases of interest, although the synchronization efficiency was less than in untransfected HeLa cells **(Figure 4.1F**, compare to **Figure 4.1A)**, likely from minor cytotoxicity resulting from FLAG-RNF138 overexpression. FLAG immunoprecipitation revealed P-TP signal that peaked at S phase and progressively weakened as cells approached the G2 and then G1 phases **(Figure 4.1G)**. In support of this, P-TP signal was partially reduced in cells that were not synchronized by double thymidine block **(Figure 4.1G)**, which primarily contain cells in G1 phase **(Figure 4.1F)**. Importantly, the P-TP signal occurred at 37 kDa, and did not appear in immunoprecipitates from untransfected cells **(Figure 4.1E,G)**, indicating the P-TP signal was associated with FLAG-RNF138. We thus conclude RNF138 is phosphorylated in a cell cycle-dependent manner, with the modification occurring primarily in S and G2 phase.

CDK2 and CDK1 activity trigger progression through the S and G2 phases (31). To confirm that the P-TP signal observed on RNF138 was dependent on CDK activity, we treated HeLa cells expressing FLAG-RNF138 with roscovitine and AZD5438, inhibitors that target both CDK1 and CDK2 activity[205,206], and RO-3306, an inhibitor of CDK1 activity[207]. The P-TP signal in FLAG immunoprecipitates was reduced upon treatment with each inhibitor **(Figure 4.1H)**. As a control, treating the cells with SB203580, an inhibitor of MAPK activity, did not affect the P-TP signal, indicating the phosphorylation on RNF138 arose solely from CDK-dependent activity **(Figure 4.4B)**. We also assessed the P-TP signal on FLAG-RNF138 when cells were transfected with siRNA targeting CDK1 or CDK2. Knocking down either kinase capably reduced RNF138 TP phosphorylation **(Figure 4.1I)**.While CDK1 or CDK2 depletion did decrease the proportion of cells in S or G2 phase, in both cases from 49.5% to ~45% **(Figure 4.4C)**, accounting for this difference still resulted in the P-TP signal dropping ~50% when either CDK was depleted **(Figure 4.1I)**. Thus, the P-TP signal on RNF138 is dependent on CDK1 and CDK2. In support of a role for CDK2 in RNF138 phosphorylation, CDK2 and its binding partner Cyclin A could co-immunoprecipitate with RNF138 **(Figure 4.1J)**, indicating RNF138 may form a complex with CDK2-Cyclin A. Consistent with this, GFP-RNF138 was capably phosphorylated by CDK2/Cyclin A2 *in vitro*. Moreover, compared to GFP-RNF138, GFP itself was only minimally phosphorylated by CDK2/Cyclin A2, demonstrating the specificity of CDK2 activity toward

RNF138 **(Figure 4.1K)**. Overall, our findings suggest RNF138 is phosphorylated by the action of CDK1 and CDK2.

To demonstrate that T27 is the site of TP phosphorylation on RNF138, we ablated the site by mutating T27 to a non-phosphorylatable alanine residue (T27A) in the FLAG-RNF138 construct. P-TP signal was severely impaired in immunoprecipitates of T27A relative to wildtype (WT) FLAG-RNF138 **(Figure 4.1L)**. At first, we could not completely eliminate P-TP signal from the T27A mutant **(Figure 4.1L**, right panel), speculating residual signal arises from other proteins co-precipitating with FLAG-RNF138 at the same molecular weight. In line with this, loading a smaller amount of the immunoprecipitates for SDS-PAGE completely abrogated P-TP signal in the T27A mutant **(Figure 4.4D)**. As well, unlike GFP-RNF138-WT, the T27A variant was not phosphorylatable by CDK2/Cyclin A2 *in vitro* **(Figure 4.1K).** Altogether, our results suggest RNF138 is phosphorylated at position T27 in a CDK1- and CDK2-dependent manner.

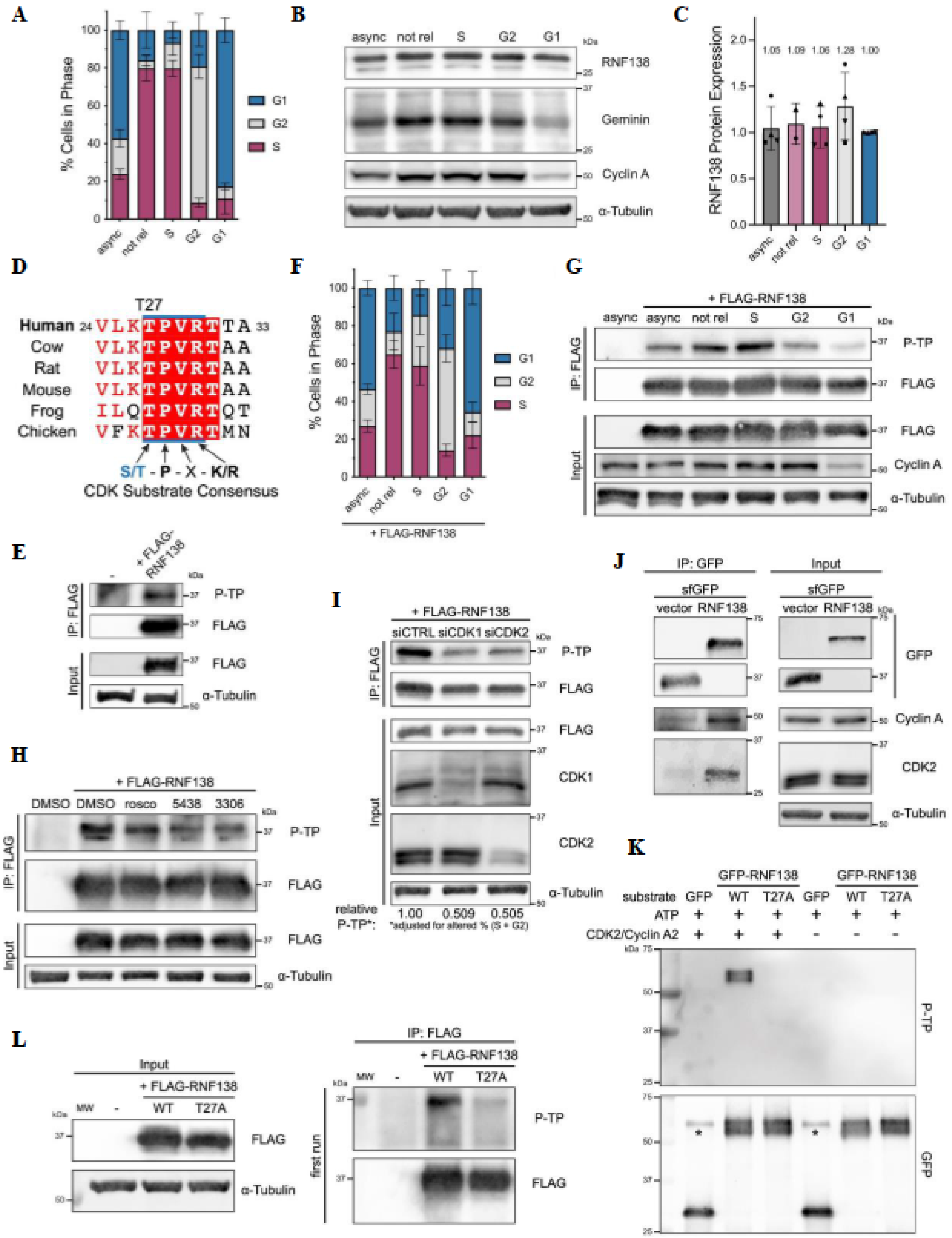
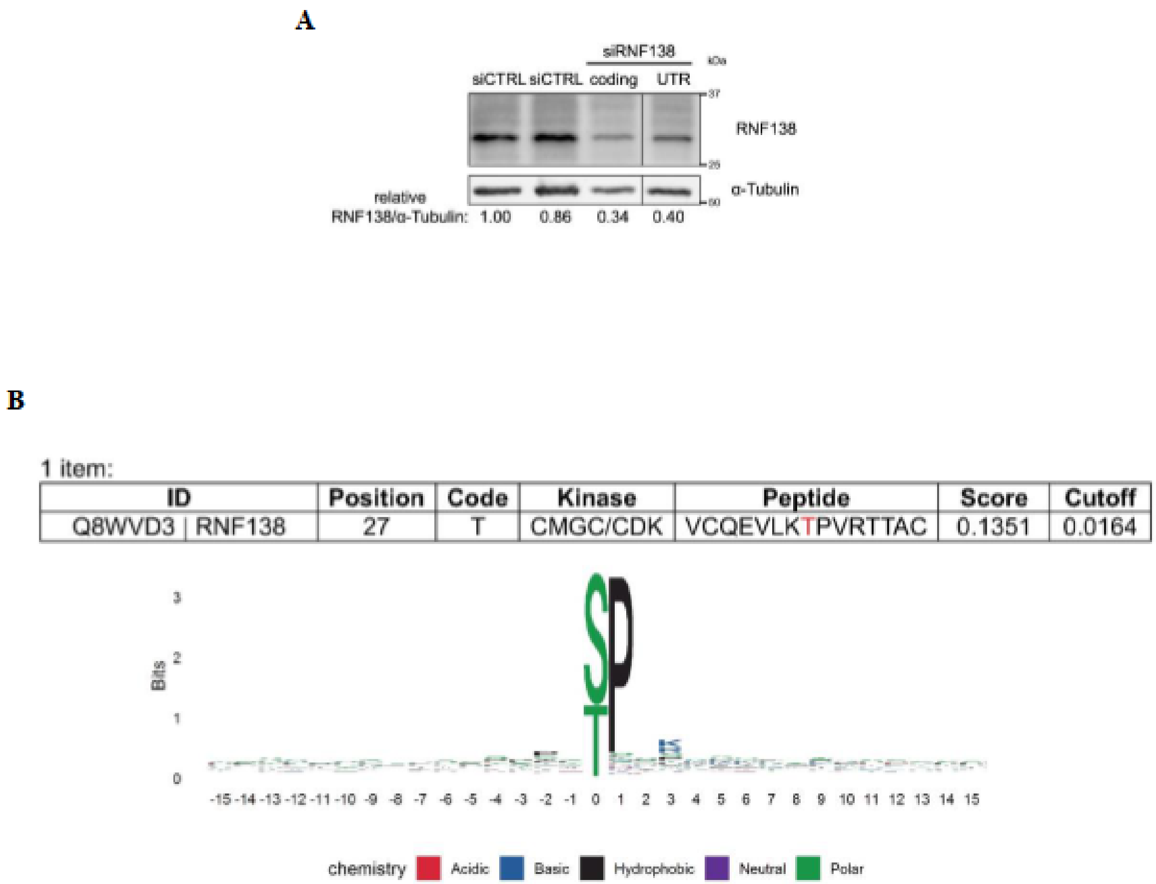

A
% Cells in Phase
G1
G2
S
async
not rel
S
G2
G1
B
async
not rel
S
G2
G1
kDa
RNF138
Geminin
Cyclin A
α-Tubulin
C
RNF138 Protein Expression
1.05
1.09
1.06
1.28
1.00
D
T27
Human 24 VLKTPVRTTA 33
Cow VLKTPVRTAA
Rat VLKTPVRTAA
Mouse VLKTPVRTAA
Frog ILQTPVRTQT
Chicken VFKTPVRTMN
S/T - P - X - K/R
CDK Substrate Consensus
F
% Cells in Phase
+ FLAG-RNF138
G
+ FLAG-RNF138
IP: FLAG
Input
P-TP
FLAG
Cyclin A
α-Tubulin
E
-
+ FLAG-RNF138
H
DMSO
rosco
5438
3306
I
siCTRL
siCDK1
siCDK2
CDK1
CDK2
relative P-TP*:
1.00
0.509
0.505
*adjusted for altered % (S + G2)
J
IP: GFP
Input
sfGFP
vector
RNF138
GFP
K
substrate
GFP
GFP-RNF138
WT
T27A
ATP
CDK2/Cyclin A2
L
MW
first run

**Figure 4.1. RNF138 is Phosphorylated by CDK-Dependent Activity.**

**A.** Flow cytometric analysis of propidium iodide signal in HeLa cells that were left asynchronous (async) or were synchronized by double-thymidine block without release (not rel) or with release to the S, G2, and G1 phases. **B.** Immunoblot of whole cell extracts from cells treated as in A). Geminin expression was used to confirm enrichment for cells in S phase, while Cyclin A expression was used to confirm enrichment in the S/G2 phases. **C)** Quantification of endogenous RNF138 expression from B). RNF138 expression was normalized to α-tubulin levels. **D.** Clustal Omega amino acid sequence alignment of the CDK consensus phosphorylation motif from human RNF138 and its orthologues. UniProt accession numbers are indicated in **Figure 4.6A**. **E.** FLAG immunoprecipitation (IP) from asynchronous HeLa cells expressing FLAG-RNF138, immunoblotted for phosphorylated TP sites (P-TP). **F.** As in A), but with HeLa transfected with FLAG-RNF138 during the first release step of double thymidine block. **G.** FLAG IP and IB of cells processed as in F). **H.** FLAG-RNF138-expressing asynchronous HeLa cells treated with the CDK inhibitors roscovitine (rosco), AZD5438, or RO-3306 or vehicle control (DMSO) for 4 hours before harvest, FLAG IP, and IB. **I.** HeLa cells were transfected with siRNA (si) to luciferase (CTRL), CDK1 or CDK2, along with FLAG-RNF138 DNA, and subjected to FLAG IP and IB. Quantification of relative P-TP signal is adjusted for the altered proportion of cells in S or G2 phase resulting from knockdown of CDK1 or CDK2 from the same biological replicate **(Figure 4.4C)**. **J.** GFP co-immunoprecipitation for endogenous CDK2 and Cyclin A in HEK293 cells transfected with sfGFP-MAP-tagged (58) RNF138 or the empty vector (vector). **K.** CDK2 *in vitro* kinase assay, with GFP, GFP-RNF138-WT or -T27A as potential substrates. The GFP constructs were expressed in HEK293 cells and immunoprecipitated. Kinase activity was detected by immunoblotting for P-TP. : a non-specific band detected by the GFP antibody. **L.** As in E., but with FLAG-RNF138-WT and -T27A. The IP eluates were loaded at different volumes (first and second runs; the second run, with 1/4 the volume loaded, is shown in **(Figure 4.4D)**. MW: molecular weight standards. Shown are representative results from at least 2 (I), 3 (G, H, K, L), at least 3 (E, J), and 4 (B) biological replicates. Averages were calculated from at least 2 (A, C) or at least 5 (F) biological replicates pooled together.



1 item:

| ID | Position | Code | Kinase | Peptide | Score | Cutoff |
|---|---|---|---|---|---|---|
| Q8WVD3 \| RNF138 | 27 | T | CMGC/CDK | VCQEVLKTPVRTTAC | 0.1351 | 0.0164 |

**Figure 4.2. Related to Figure 1**

**A)** Immunoblots (IBs) of U2OS cells were transfected with siRNA targeting luciferase (siCTRL) or the *RNF138* gene (siRNF138) within its coding or 3' untranslated region (UTR). siRNA targeting the *RNF138* UTR was comprised of a pool of four siRNAs, each at 40 nM. **B)** Results from the scan of the primary sequence of full-length human RNF138 by the GPS (Group-based Prediction System) 6.0 web server (44) for consensus phosphorylation motifs of the cyclin-dependent kinase (CDK) subset of the CMGC family of kinases (http://gps.biocuckoo.cn/online.php).

### 4.4.3 RNF138 is a Target for Polyubiquitylation

The repair of DSBs is coordinated by a cascade of ubiquitylation events, contributing to protein recruitment to sites of damage, the assembly and disassembly of complexes involved in repair, and protein turnover[181,208]. We consequently were curious if RNF138 was also a target of ubiquitylation. Previously, we studied SUMOylation of CtIP in HeLa cells stably expressing His-tagged SUMO-2; SUMO-2 is a ubiquitin-like modifier also conjugated to proteins[209]. From these cells, we isolated the pool of His-tagged proteins by nickel affinity purification ("His pull-down"), which represented those that had incorporated SUMO-2. We attempted to use the same strategy to study the ubiquitylation of endogenous RNF138, this time using HeLa cells stably expressing 6xHis-biotin-tagged ubiquitin (HeLa HB-ubiquitin)[186], with the goal of exploiting the 6xHis component for nickel affinity purification. Nickel beads could successfully enrich ubiquitylated proteins from extracts of these cells relative to plain HeLa cells not expressing HB-ubiquitin **(Figure 4.4E)**. However, despite our best efforts, we were unable to detect ubiquitylated RNF138 in the His pull-down fraction, even though RNF138 expression was observed in whole cell extracts **(Figure 4.4E)**, rightmost panel). Monoubiquitylated RNF138 should increase in molecular weight by ~10 kDa, and such a species (~38 kDa) was not clearly observed in HeLa HB-ubiquitin relative to plain HeLa cells **(Figure 4.4E)**. Perhaps ubiquitylated RNF138 is of low abundance, such that these species are below the detection limit of our immunoblotting approach. The issue is exacerbated by RNF138 antibodies detecting non-specific signals during immunoblotting **(Figure 4.4E**, rightmost panel). It is difficult to discern which higher order species in the His pull-down arise from ubiquitylated RNF138 or are simply non-specific in nature, complicating our analysis. Thus, in our hands, we do not consider His pull-down a viable approach to study RNF138 ubiquitylation.

As an alternate strategy to detect RNF138 ubiquitylation, we turned to exogenous co-expression of green fluorescent protein (GFP)-tagged RNF138 and HA-tagged ubiquitin (HA-Ub) in cells followed by immunoprecipitation for GFP. This approach would boost expression of both ubiquitin and RNF138, improving the detection of ubiquitylated RNF138. In addition to WT-ubiquitin, we also used an HA-tagged mutant, with leucine 73 substituted with proline (HA-Ub-

L73P), that is conjugable to substrates but resistant to deubiquitinating enzymes (51). We hoped this would stabilize ubiquitylated RNF138, increasing the likelihood of its detection. We also performed these experiments in HEK293 cells, which capably tolerated overexpression of GFP-RNF138. GFP-RNF138 was isolated by immunoprecipitation with anti-GFP beads under stringent conditions. Eluates from the beads were then processed for SDS-PAGE and subjected to immunoblotting for the HA tag. When HA-Ub-WT and GFP-RNF138 were co-expressed, we detected a smear of immunoreactivity to HA appearing under 75 kDa and extending beyond 250 kDa, where the signal greatly intensified **(Figure 4.3A)**, strongly suggesting GFP-RNF138 (theoretically ~55 kDa in size but migrating at ~63 kDa to start) was polyubiquitylated. This signal was absent in cells when HA-Ub-WT was co-expressed with empty vector GFP **(Figure 4.3A)**. Notably, higher order HA signal was also present when the same experiment was performed with HA-Ub-L73P **(Figure 4.3A)**. Similarly, this signal was much reduced when GFP was used in place of GFP-RNF138. However, using HA-Ub-L73P shifted the bulk of the HA smear between <75 to ~150 kDa, and reduced the intensity of the signal appearing >250 kDa **(Figure 4.3A)**. At lower molecular weights, the smear resolved to a distinct laddering pattern perhaps revealing different species of ubiquitylated RNF138, each band suggestive of a different number of ubiquitin moieties attached to the protein. The discrepancy in how RNF138 conjugates appear using HA-Ub-WT or -L73P is consistent with the observation that L73P-Ub is conjugated to targets less efficiently than WT-Ub *in vitro* (51). On ubiquitin, L73 sits in a hydrophobic patch used by some E3 ligases to assemble polyubiquitin chains, and the proline substitution disrupts hydrophobic packing in this region[184]. In our case, it is clear expressing L73P- instead of WT-Ub restricts the extent by which RNF138 can be polyubiquitylated, but it is still polyubiquitylated nonetheless. While smearing above unmodified GFP-RNF138 did not appear when we first re-probed the immunoprecipitates for GFP **(Figure 4.3A)**, these immunoblots were performed using infrared-fluorescent secondary antibodies, which we find are less sensitive to signals of low abundance. When chemiluminescence was instead used to develop the GFP immunoblots, we observed smearing above the unconjugated GFP-RNF138 band, although the laddering was not as distinct as that detected by HA immunoblotting **(Figure 4.3B)**.

Continuing on, to confirm RNF138 is ubiquitylated, we observed that the ladder of HA signal in immunoprecipitates of GFP-RNF138 was absent if HA-Ub-L73P was not transfected into cells **(Figure 4.3C)**, indicating the signal resulted from HA-Ub-L73P co-expression. We also co-expressed a truncated GFP-RNF138 deleted of its UIM (ΔUIM) **(Figure 4.4E)** with HA-Ub-L73P. Here, the bands in the HA-Ub ladder exhibited a downward shift in size relative to those from GFP-RNF138-WT, consistent with the decrease in molecular weight resulting from deleting the UIM **(Figure 4.3D)**. This definitively showed that the HA signal on GFP-RNF138 arises from ubiquitylation events on RNF138 itself, and not from proteins that are inadvertently co-immunoprecipitated. We thus conclude that RNF138 is constitutively polyubiquitylated in cells.

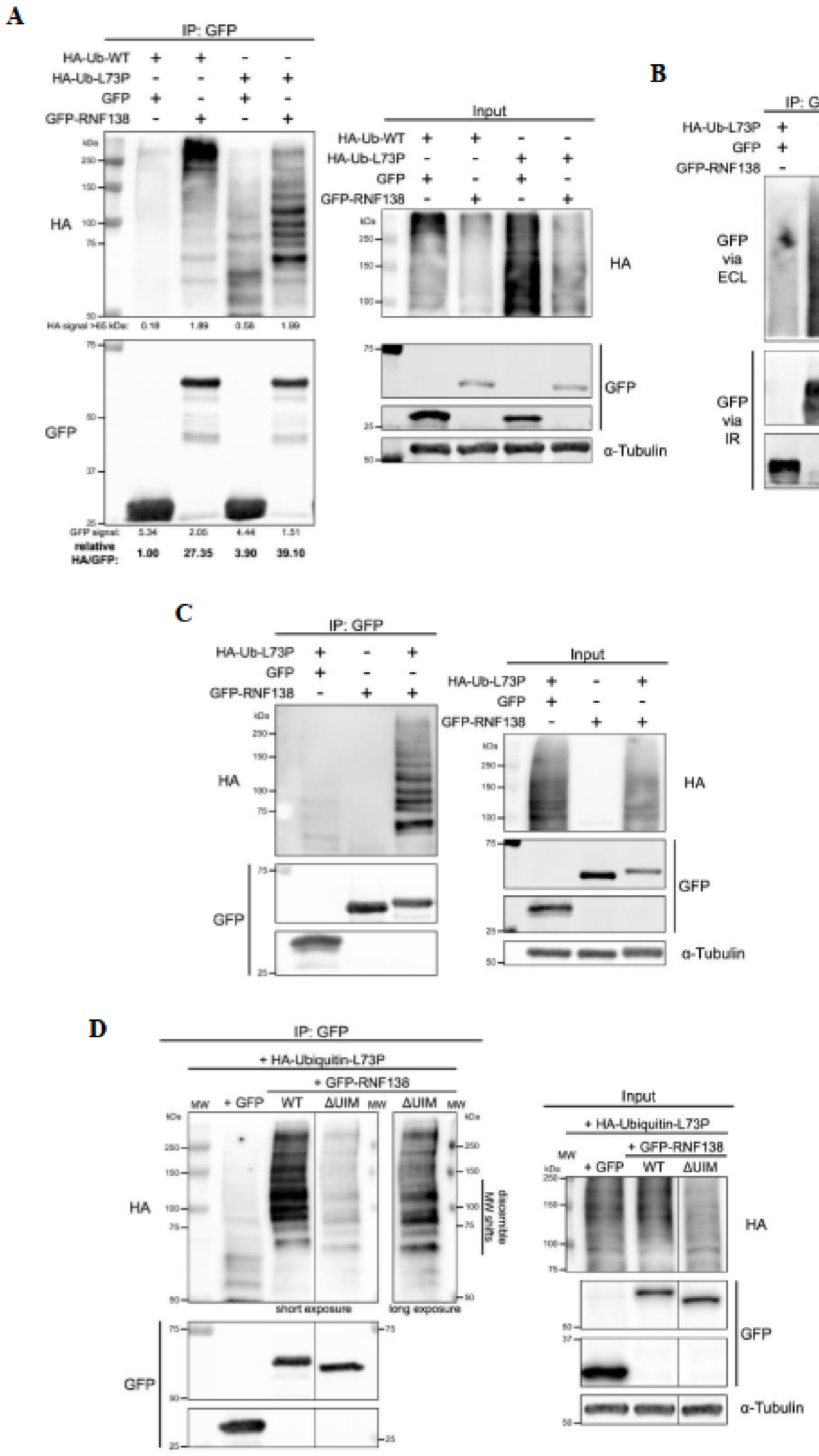

A
IP: GFP
HA-Ub-WT + + - -
HA-Ub-L73P - - + +
GFP + - + -
GFP-RNF138 - + - +
HA
HA signal >65 kDa: 0.18 1.89 0.58 1.99
GFP
GFP signal: 5.34 2.05 4.44 1.51
relative HA/GFP: 1.00 27.35 3.90 39.10
Input
HA-Ub-WT + + - -
HA-Ub-L73P - - + +
GFP + - + -
GFP-RNF138 - + - +
HA
GFP
α-Tubulin
B
IP: GFP
HA-Ub-L73P + +
GFP + -
GFP-RNF138 - +
MW
GFP via ECL
GFP via IR
C
IP: GFP
HA-Ub-L73P + - +
GFP + - -
GFP-RNF138 - + +
HA
GFP
Input
HA-Ub-L73P + - +
GFP + - -
GFP-RNF138 - + +
HA
GFP
α-Tubulin
D
IP: GFP
+ HA-Ubiquitin-L73P
+ GFP-RNF138
MW
+ GFP
WT
ΔUIM
MW
ΔUIM
MW
HA
discernible MW shifts
short exposure
long exposure
GFP
Input
+ HA-Ubiquitin-L73P
+ GFP-RNF138
MW
+ GFP
WT
ΔUIM
HA
GFP
α-Tubulin

**Figure 4.3. RNF138 is Constitutively Polyubiquitylated.**

Stringent GFP immunoprecipitations (IP) were performed for HEK293 cells expressing **A. and B.** HA-ubiquitin-WT or -L73P and GFP-RNF138 or free GFP, **C.** GFP or GFP-RNF138 with or without HA-Ub-L73P, or **D.** HA-Ub-L73P with GFP or GFP-RNF138-WT or -ΔUIM, the products of which were then immunoblotted. **B.** contains select immunoprecipitate samples from A), with GFP immunoblot signals below 75 kDa developed by infrared-fluorescent (IR) secondary antibodies, and GFP signals above 75 kDa developed by horseradish peroxidase-coupled secondary antibodies and enhanced chemiluminescence (ECL). Shown are representative results from at least 4 biological replicates (each of A, C, D). MW: molecular weight standards.

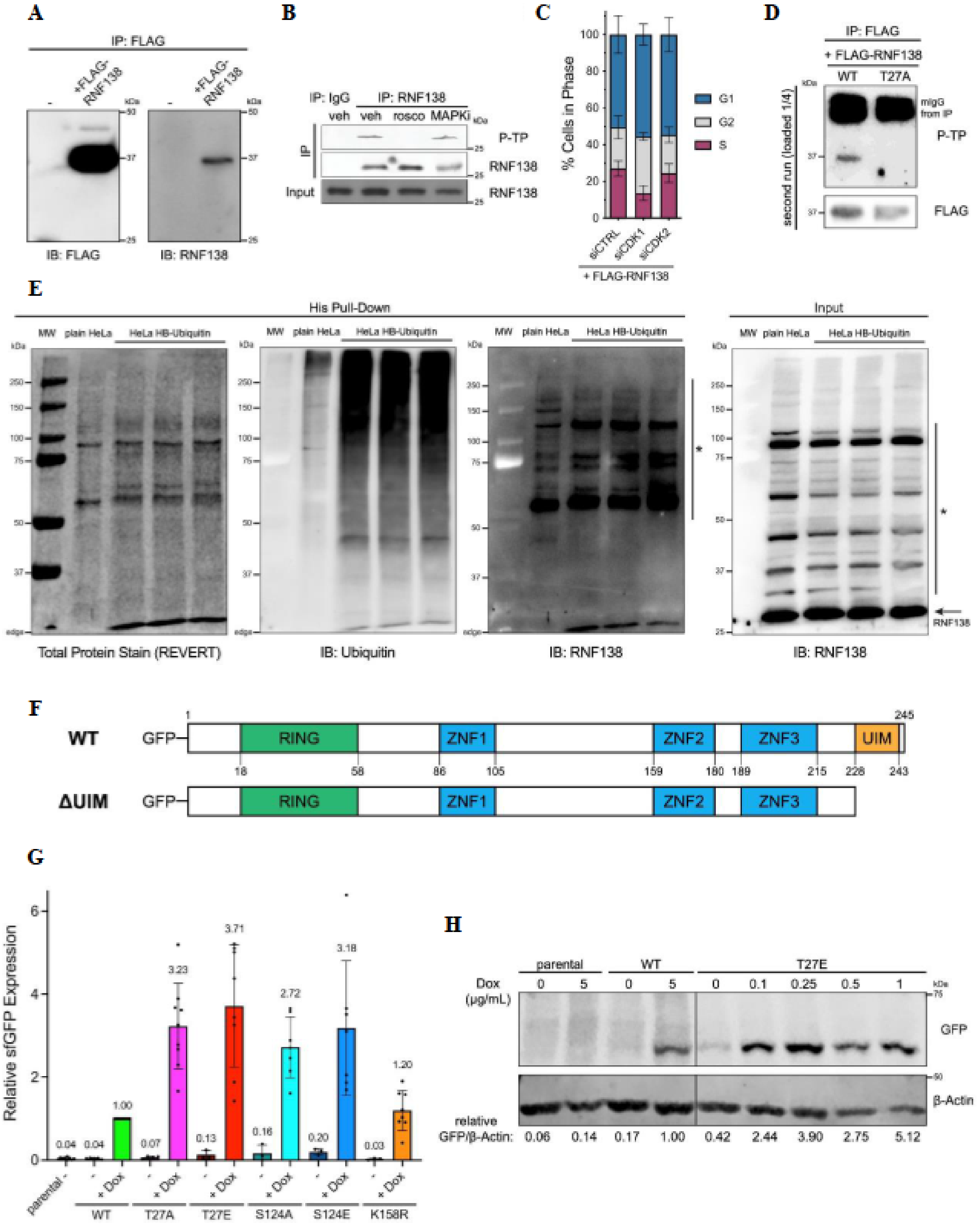


**Figure 4.4. Related to Figures 4.1, 4.2 and 4.5.**

**A.** FLAG immunoprecipitates (IPs) of HEK293 cells expressing FLAG-RNF138 or not were immunoblotted for the FLAG epitope or RNF138 (Abcam antibody). **B.** G2-synchronized HeLa cells were treated with 0.2% DMSO vehicle control (veh), 50 µM roscovitine (rosco), or the MAP kinase inhibitor SB203580 (MAPKi) at 10 µM for 4 hours. Cell extracts were then IP'd for RNF138 or with control immunoglobulin G (IgG) and immunoblotted for phosphorylated TP sites (P-TP) or RNF138. **C)** Flow cytometric analysis of propidium iodide signal in HeLa cells treated as in **Figure 4.1I**. Averages in C) were calculated from 3 biological replicates pooled together. **D.** FLAG IP from asynchronous HeLa cells expressing FLAG-RNF138-WT or -T27A and immunoblotted for P-TP and FLAG. The first run of the IP and the input control is shown in **Figure 4.1L**. The second run of the IP is shown here; 1/4 the volume of the IP eluate from the first run is loaded. mIgG: mouse anti-FLAG immunoglobulin G fragments used for IP, detected by the anti-mouse—HRP secondary antibody. **E.** 90% of a single pellet of plain HeLa cells and 3 cell pellets of HeLa stably expressing 6XHis-biotin-ubiquitin (HeLa HB-ubiquitin) was subjected to nickel affinity purification ("His pull-down"), while the remaining 10% was processed to generate whole cell extract (input). The samples were resolved by SDS-PAGE and transferred to nitrocellulose membrane. The His pull-down fraction was first stained for total protein with REVERT total protein stain (LI-COR Biosciences), then immunoblotted (IB'd) for RNF138, and finally stripped and IB'd for ubiquitin. For the remaining 10% of cells, whole cell extracts were prepared for the input control (rightmost panel) and IB'd for RNF138. Note the RNF138 antibody detects other species beyond RNF138 ( ) . RNF138 itself (predicted molecular weight: 28 kDa) is detected just above 25 kDa (arrow). MW: molecular weight standards. **F.** Schematic diagrams of the structural domains in wildtype (WT) and the ΔUIM mutant of GFP-tagged RNF138. **G.** Quantification of IBs (such as **Figure 4.7G**) for doxycycline (Dox)-inducible expression of sfGFP-RNF138 variants in U2OS-TREx cells stably integrated with their DNA. **H.** Example titration of Dox concentrations to induce sfGFP-RNF138-WT and -T27E expression, detected by IB. D. is a representative result from at least 6 biological replicates. Averages in G. were calculated from at least 7 biological replicates pooled together.

### 4.4.4 Insights into the Dynamics of RNF138 Ubiquitylation

Having determined that RNF138 is a target for ubiquitylation, we sought to understand how different cellular conditions would affect RNF138 ubiquitylation. We first asked if the cell cycle factored into the degree of RNF138 ubiquitylation. We synchronized HeLa cells to the S, G2, and G1 phases, and during the process co-expressed in them low levels of HA-Ub-WT and GFP-RNF138. Despite transfecting in low levels of GFP-RNF138, its expression was slightly cytotoxic to HeLa, resulting in a reduction in synchronization efficiency **(Figure 4.5**, compare with **Figure 4.1A)**. Still, GFP immunoprecipitation consistently revealed more RNF138 ubiquitylation in samples enriched for S phase cells compared to ones enriched for G1 phase cells **(Figure 4.5B-C)**. As well, samples enriched for cells in G2 had similar or greater levels of ubiquitylation than those enriched for S phase cells (on average, 1.986 times more). We infer then that RNF138 ubiquitylation is cell cycle-dependent, with ubiquitylation increasing in S and G2 phase and decreasing in G1 phase.

We also examined if genotoxic stress would alter RNF138 ubiquitylation. We expressed HA-Ub-WT and GFP-RNF138 in HEK293 cells and treated the cells with various DNA damaging agents. We used ionizing radiation from a gamma source (IR) and phleomycin (phleo) to induce DSBs. We also treated cells with ultraviolet light (UV), along with the replication stress-inducing agents camptothecin (CPT, which also induces DSBs) and hydroxyurea (HU). Treating cells with any of the agents reduced the signal of higher order HA-Ub on GFP-RNF138 **(Figure 4.5D-E)**. This was not due to variable expression of HA-Ub-WT between the samples, as signals from HA-Ub-WT conjugates were comparable in whole cell extracts **(Figure 4.5D)**. Therefore, DNA damage induces a reduction in RNF138 ubiquitylation. Overall, our data reveal differential ubiquitylation of RNF138 depending on cell status, with ubiquitylation rising when cells are in S/G2 phase and lessening upon genotoxic stress.

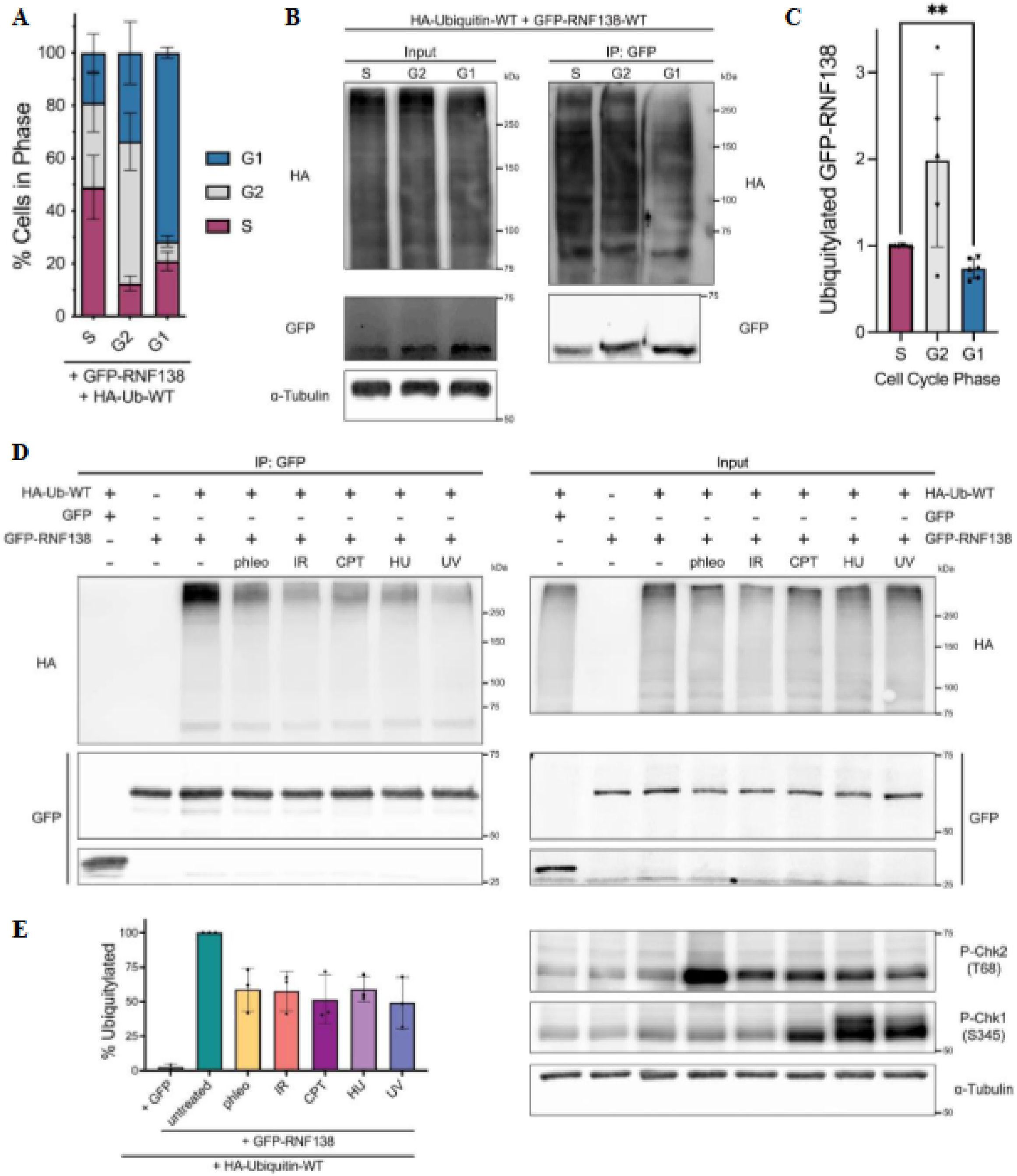


**Figure 4.5. The Dynamics of RNF138 Ubiquitylation Upon Cell Cycle Progression and DNA Damage.**

**A.** Flow cytometric analysis of propidium iodide signal in HeLa cells synchronized by double-thymidine block and release. The cells were transfected with GFP-RNF138 and HA-ubiquitin-WT during the first release. **B.** GFP immunoprecipitation (IP) assay from cells treated as in A). **C.** Quantification of ubiquitylated GFP-RNF138 from B). HA signal in the IP fraction was normalized to GFP signal in the IP fraction. : a paired, two-tailed, parametric *t*-test comparing the S and G1 phase conditions for ubiquitylated GFP-RNF138 gives $p = 0.0043$ ($p \leq 0.01$). **D.** HEK293 cells were transfected with HA-ubiquitin-WT and GFP or GFP-RNF138 and exposed to phleomycin (phleo), ionizing radiation (IR), camptothecin (CPT), hydroxyurea (HU), or ultraviolet light (UV). The cells were then subjected to GFP immunoprecipitation and immunoblot analysis. Chk1 and Chk2 phosphorylation (P) at S345 and T68, respectively, are markers for activation of the DNA damage response. **E)** Quantification of D. as per C. Shown are representative results from 3 (D) and at least 6 (B) biological replicates. Averages were calculated from 3 (E), at least 4 (A), or at least 5 (C) biological replicates pooled together.

### 4.4.5 K158 is a Site of RNF138 Ubiquitylation

Four proteomic screens have identified residue K158 as a ubiquitylation site on RNF138 expressed in HEK293, HeLa cells, and Jurkat cells[210–213]. K158 was also a site of ubiquitylation in a study that found RNF138 promotes oncogenic signaling in lymphomas expressing L265P-mutated MyD88[124]. Interestingly, the K158 residue is conserved across RNF138 orthologues **(Figure 4.6A)**. It is also predicted to be solvent-exposed **(Figure 4.8A, C-E)**, in line with it being accessible for modification. To determine if K158 was indeed a ubiquitylation site in our system, we generated a K158R mutant of GFP-RNF138, the arginine substitution maintaining the positive charge but ablating the site of ubiquitylation. GFP immunoprecipitates of the construct exhibited a notable reduction in HA-Ub signal relative to WT-RNF138 when co-expressed with HA-Ub-L73P **(Figure 4.6B)**, substantiating K158 as a site for RNF138 ubiquitylation.

As polyubiquitin chains are known to target proteins for degradation by the proteasome[214], we determined the impact of disrupting K158 ubiquitylation on RNF138 stability by a cycloheximide chase assay. We expressed GFP-RNF138-WT or -K158R in HEK293 cells, inhibited protein synthesis with cycloheximide, and monitored exogenous RNF138 levels over time. We found that the K158R mutant was turned over at a reduced rate compared to WT-RNF138 **(Figure 4.6C-D)**. In the same experiment, we also inhibited proteasome activity with the compound MG132. MG132 treatment partially and fully restored WT- and K158R-RNF138 stability, respectively (**Figure 4.6C,E**), indicating RNF138 turnover is at least partly dependent on proteasomal activity. Taken together, these findings indicate K158 is a site of constitutive ubiquitylation on RNF138, and modification of this site by ubiquitin can promote proteasome-mediated turnover of the protein.

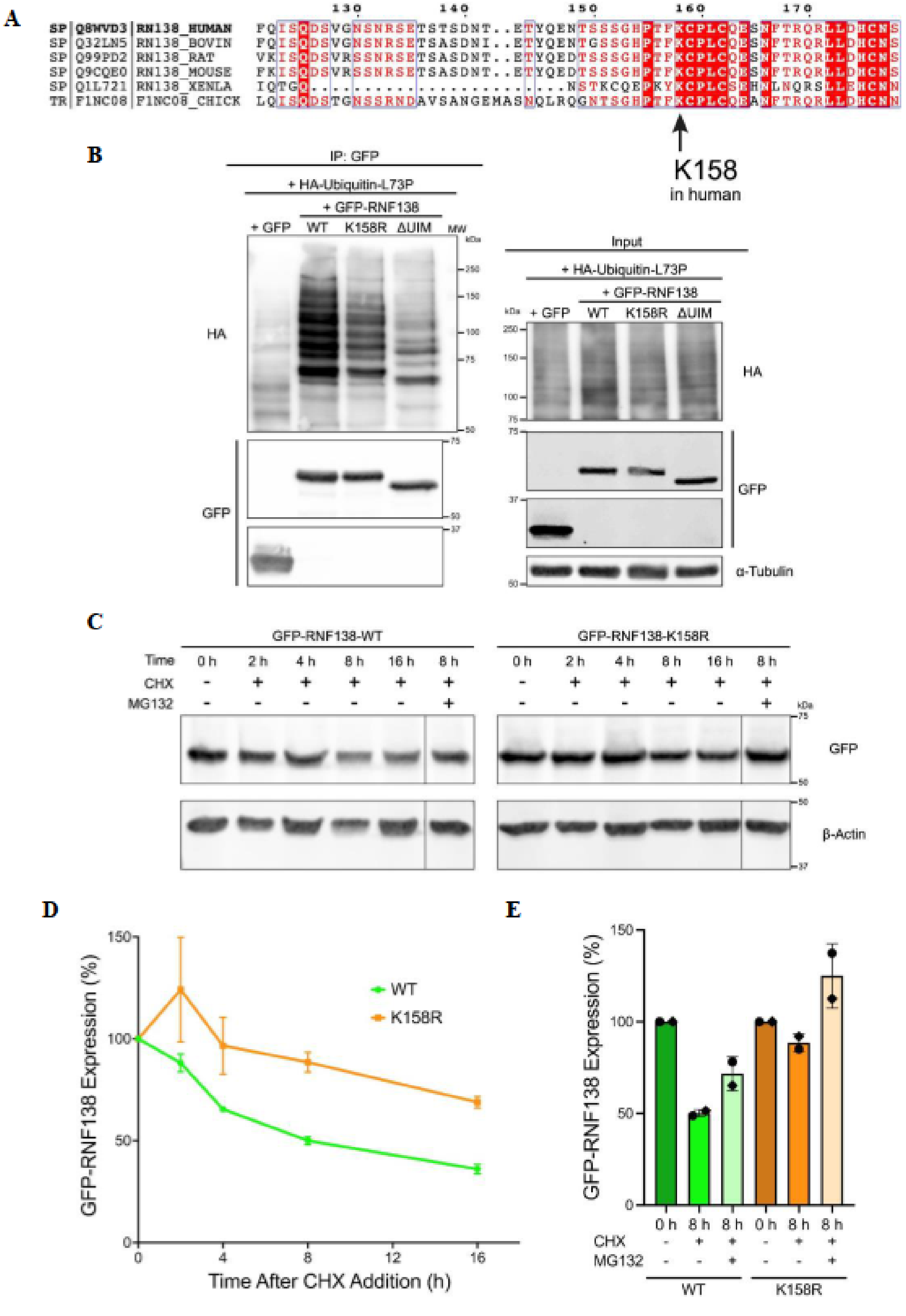

A
130
140
150
160
170
SP|Q8WVD3|RN138_HUMAN FQISQDSVGNSNRSETSTSDNT..ETYQENTSSSGHPTFKCPLCQESNFTRQRLLDHCNS
SP|Q32LN5|RN138_BOVIN FQISQDSVGNSNRSETSASDNI..ETYQENTGSSGHPTFKCPLCQESNFTRQRLLDHCNS
SP|Q99PD2|RN138_RAT VKISQDSVRSSNRSETSASDNT..ETYQEDTSSSGHPTFKCPLCQESNFTRQRLLDHCNS
SP|Q9CQE0|RN138_MOUSE FKISQDSVRSSNRSETSASDNT..ETYQEDTSSSGHPTFKCPLCQESNFTRQRLLDHCNS
SP|Q1L721|RN138_XENLA IQTGQ.....................NSTKCQEPKYKCPLCSEHNLNQRSLLEHCNN
TR|F1NC08|F1NC08_CHICK LQISQDSTGNSSRNDAVSANGEMASNQLRQGNTSGHPTFKCPLCQEANFTRQRLLDHCNN
K158
in human
B
IP: GFP
+ HA-Ubiquitin-L73P
+ GFP-RNF138
+ GFP
WT
K158R
ΔUIM
MW
kDa
250
150
100
75
50
37
HA
GFP
Input
α-Tubulin
C
GFP-RNF138-WT
GFP-RNF138-K158R
Time
0 h
2 h
4 h
8 h
16 h
8 h
CHX
MG132
kDa
75
50
37
GFP
β-Actin
D
GFP-RNF138 Expression (%)
150
100
50
0
0
4
8
12
16
Time After CHX Addition (h)
WT
K158R
E
GFP-RNF138 Expression (%)
150
100
50
0
0 h
8 h
8 h
CHX
MG132
WT
K158R

**Figure 4.6. K158 is a Site of RNF138 Ubiquitylation.**

**A.** Clustal Omega amino acid sequence alignment of the region containing the K158 residue from human RNF138 with its orthologues. **B.** Stringent GFP immunoprecipitation (IP) was performed for HEK293 cells transfected with HA-Ub-L73P and GFP or GFP-RNF138-WT, -K158R, or -ΔUIM and then immunoblotted. MW: molecular weight standards. **C)** Immunoblot of whole cell extracts of HEK293 cells transfected with GFP-RNF138-WT or -K158R and treated with cycloheximide (CHX) with or without MG132 for various timepoints. **D. and E.** Quantifications of C). Shown are representative results from 2 (C) and at least 5 (B) biological replicates. Averages in D) and E) are of 2 biological replicates pooled together.

### 4.4.6 T27 and K158 are Not Required for the Recruitment of RNF138 to Sites of Damage

We so far observed CDK-dependent phosphorylation of RNF138 on residue T27 and, in support of previous findings[124,210–213], ubiquitylation on RNF138 at K158. To study how these post-translational modifications impact RNF138's function in the DSB response, we continued our investigations in U2OS osteosarcoma cells, a standard cell line used to study the DSB response. The three zinc finger domains (ZNF1, ZNF2, and ZNF3) are together essential for targeting RNF138 to DNA damage sites[129]. ZNF2 and ZNF3 are predicted to pack together in a single domain; this ZNF2/3 domain is flexibly tethered to the rest of the protein by a long linker **(Figure 4.8A, C-E)**. As K158 is situated immediately N-terminal of ZNF2 in RNF138 **(Figure 4.7A)**, we asked if ubiquitylation at K158 might contribute to RNF138 recruitment to DNA damage. We transiently transfected U2OS cells with minimal amounts of DNA encoding GFP-RNF138-WT or -K158R. To induce DNA damage, we pre-sensitized cells with Hoescht 33358, then irradiated nuclei with a stripe of 405 nm light. Timelapse fluorescence imaging was used to monitor the recruitment of the fusion proteins to the regions of damage over 5 minutes. Both GFP-RNF138-WT and -K158R effectively accumulated at laser stripes **(Figure 4.7B)**, although the K158R mutant exhibited a minor impairment when the recruitment kinetics were examined (~25%

decrease in signal relative to WT at 200 s) **(Figure 4.7C)**. We conclude then that the K158 site promotes but is not essential for RNF138's accumulation on damaged DNA.

We next asked if T27 played a role in RNF138 recruitment to damage. This time, we generated phospho-ablating alanine (T27A) and phospho-mimicking glutamate (T27→E27, T27E) substitutions for T27 in the context of GFP-RNF138. While the WT and T27A GFP-RNF138 constructs were capably enriched at laser microirradiated stripes **(Figure 4.7D)**, the T27A mutant displayed a subtle defect in retention over time (~10% decrease in signal relative to WT at 240 s) **(Figure 4.7E)**. Meanwhile, the T27E variant accumulated at laser stripes, but exhibited a slight defect in accrual (14.75% decrease in signal relative to WT at 100 s) and a discernible defect in retention (36.22% decrease in signal relative to WT at 300 s) **(Figure 4.7D-E)**. It appears then that a permanent negative charge at T27 may impact sustained localization of RNF138 to chromatin, perhaps owing to electrostatic repulsion impacting DNA binding potentially contributed by the positively charged surface of the RING domain **(Figure 4.8A, B)**. Regardless, it is clear the function of residue T27 is not required for RNF138's localization to sites of DNA damage. Collectively, these data suggest modifications at K158 and T27 are not required for RNF138's accrual at DNA damage, but may have minor roles in promoting recruitment to (K158) and regulating retention on (T27) chromatin.

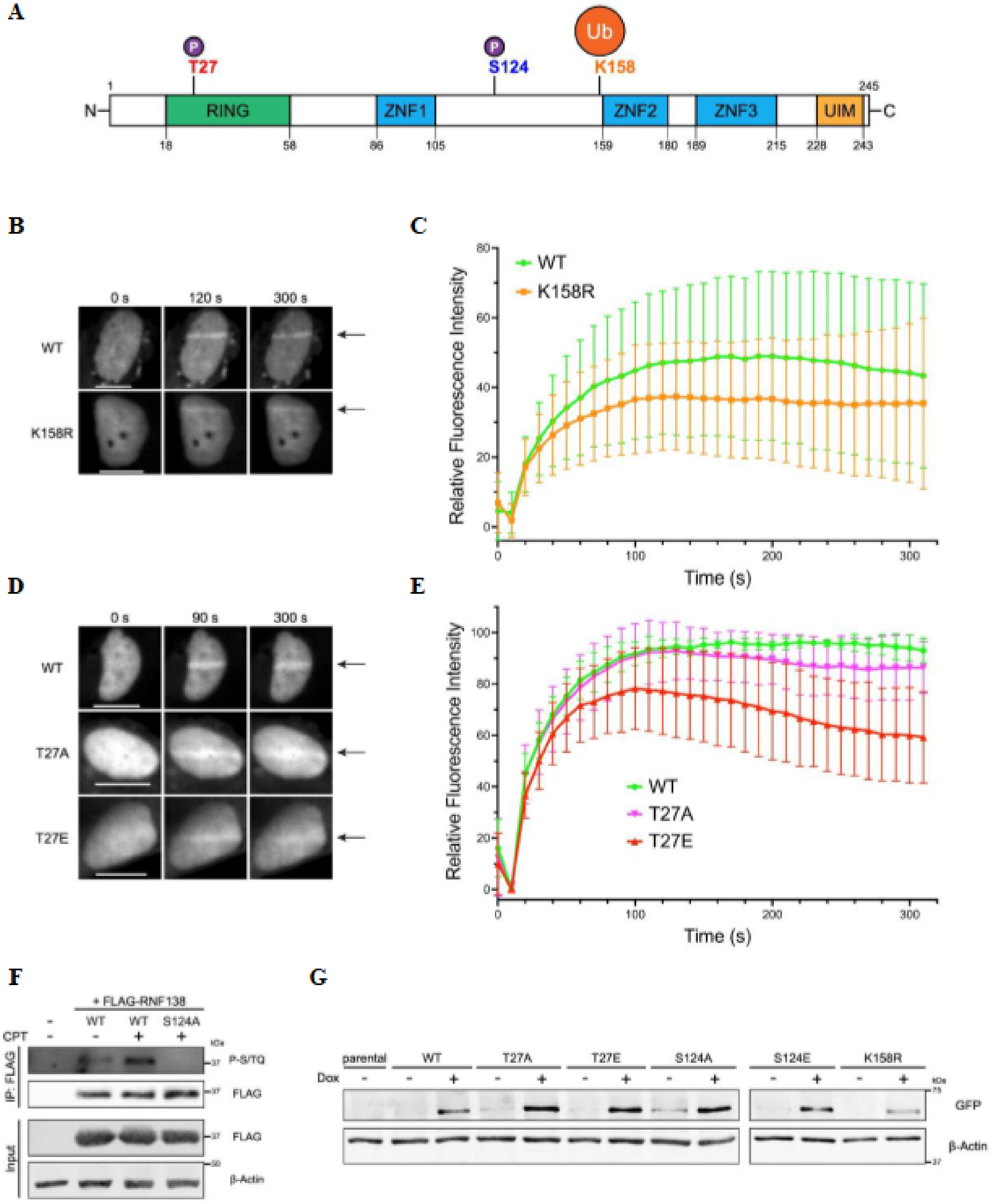


**Figure 4.7. Recruitment Kinetics of RNF138 Mutants at T27 and K158 to Sites of DNA Damage; Stable Expression of T27, S124, and K158 Variants in U2OS-TREx Cells.**

**A.** Schematic diagram of the structural domains in wildtype (WT) RNF138 and the positions of the post-translational modification sites investigated. **B.** Representative micrographs of live U2OS cells transfected with GFP-RNF138-WT or -K158, microirradiated in a line across the nucleus with a 405 nm laser, and monitored by time-lapse microscopy (arrow: region irradiated). **C.** Quantification of relative fluorescence intensity over time at the microirradiated region in B). The relative recruitment kinetics of 30 cells per construct, pooled from 3 biological replicates, were averaged. **D.** As in B) but for U2OS expressing GFP-RNF138-WT, -T27A, and -T27E. **E.** Quantification of relative fluorescence intensity over time at the microirradiated region in D). For a single replicate, the mean recruitment kinetics of ~25 cells were calculated for each construct. The normalized means from 3 biological replicates were then averaged and plotted (~75 cells total per construct). **F.** FLAG immunoprecipitation (IP) and subsequent immunoblot of HeLa cells expressing FLAG-RNF138-WT or -S124A and treated with camptothecin (CPT) or not for 1 hour. **G.** IB of parental or stable sfGFP-RNF138 variant-expressing U2OS-TREx cells induced with doxycycline (Dox) or not. Shown are representative results from 3 (B, D), 4 (F), and at least 7 (G) biological replicates. Scale bars denote 10 µm.

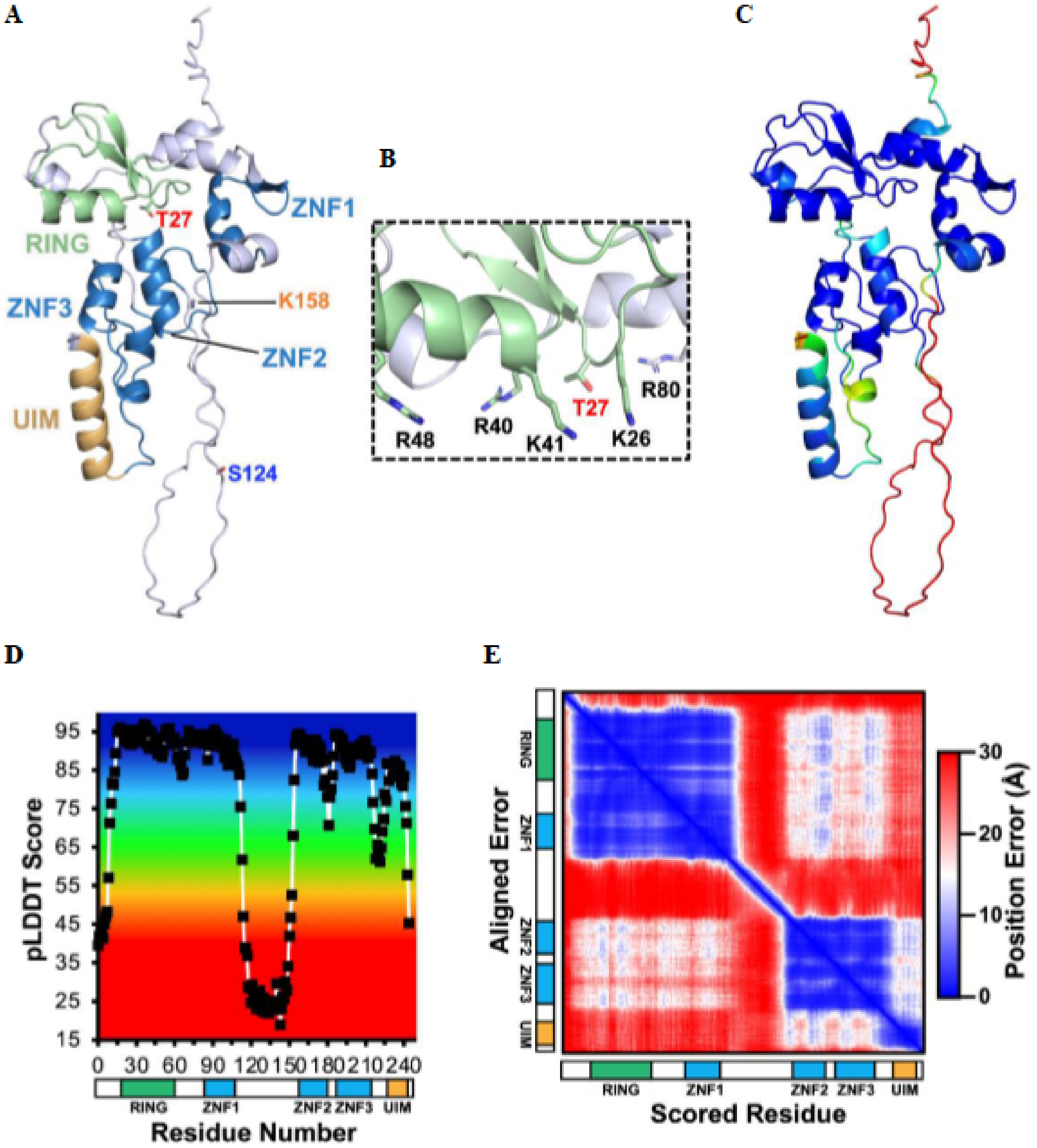


**Figure 4.8. AlphaFold-Predicted Model of RNF138's Structure.**

**A**. Predicted model[154–156] of human RNF138 (UniProt accession Q8WVD3) with individual domains coloured respectively: linkers in white, RING in green, zinc fingers (ZNF) in blue, ubiquitin interacting motif (UIM) in dark yellow. The RING domain folds onto ZNF1, linked together by a hinge-like linker. A longer linker (54 residues) connects the N-terminal RING and ZNF1 to ZNF2 and ZNF3, which appear to pack and form a folded structure and are followed by the UIM. The post-translational modification sites described in this study are highlighted: T27 (red), S124 (dark blue), and K158 (orange). **B.** Closeup view of residue T27 in A), highlighting positive residues in close proximity and forming a positively charged surface in the RING domain. **C., D.** Predicted Local Distance Difference Test (pLDDT)[160] plot of the predicted model of

RNF138. pLDDT scores range from 0 to 100, where scores above 70 indicate a high level of confidence in the position of the Cα atom, and scores above 90 indicate a high level of confidence in the placement of side chain atoms. Scores below 50 are generally considered as indications of low confidence and suggest disorder[160]. Dark blue and red correspond to high confidence (>90) and low confidence (<50) pLDDT scores, respectively. The colour gradient on the graph (D) corresponds to the same colour scheme in panel C). Overall, the model was predicted with high confidence with a median pLDDT of 88.9. The residues in the N-terminal RING domain and ZNF1 had consistently high pLDDT high values. Notably, T27 had a score of 91.6. The linker connecting the N-terminal domains to the C-terminal domains was predicted with low confidence, suggesting flexibility within those residues. S124 is positioned in the flexible linker with a pLDDT score of 26.05, suggesting a dynamic or flexible region[155–157,160]. The C-terminal zinc fingers and UIM were predicted with high confidence. Interestingly, residue K158, preceding ZNF2, had a pLDDT score of 94.12. **E.** Plot representing the predicted alignment error (pAE) for the predicted structure of RNF138. In contrast to pLDDT, which measures local confidence, pAE is a long-range domain position confidence metric that calculates the confidence of each residue in relation to all other residues[154–157]. It quantifies the predicted error in Angstroms (Å) for residue *x* if the model were aligned onto the true structure at residue *y*. The pAEs within the N-terminal domains of RNF138 were of high confidence, with a median predicted error of 2.82 Å. The flexible linker connecting the N-terminal and C-terminal domains had high predicted error, with a median pAE value of 22.36 Å, further suggesting a flexible region in the protein. The C-terminal domains had a median pAE of 5.1 Å, suggesting a high confidence prediction. pAE is practical for measuring and analyzing both intra- and interposition confidence because it does not impose the requirement for the residues to belong to the same monomer[155,157]. Although the N- and C-terminal domains are within the same monomer and appear to fold and interact, they are separated by the large flexible linker. The pAE error between the N-terminal domains and C-terminal domains is relatively high (16.77 Å) due to the dynamic linker making it difficult to confidently predict residues involved in the interaction between the N- and C-terminal domains. Overall, this modeling suggests RNF138 is a flexible and dynamic protein accessible to post-translational modification.

### 4.4.7 The RNF138 PTM Sites T27, S124 and K158 are Important for DNA End Resection

An additional residue on RNF138, S124, has been reported to be phosphorylated in an ATM kinase- and DNA damage-dependent manner[131]. S124 resides in the long, flexible linker between ZNF1 and ZNF2 (**Figure 4.7A,Figure *4.8*A,C-E**), so it would also be available for phosphorylation. We confirmed that S124 was the sole DNA damage-induced phosphorylation site on RNF138 by immunoblotting immunoprecipitates of FLAG-RNF138-WT and -S124A for the phosphorylated PIKK substrate consensus motif (phospho-Ser/Thr-Gln, or P-S/TQ) (**Figure 4.7F**).

Treating HeLa cells with CPT induced S/TQ phosphorylation on FLAG-RNF138-WT, while this signal was completely abolished in the S124A mutant. While it has been reported that the S124A mutant is proficient in recruitment to laser-microirradiated stripes[131], the effects of S124 phosphorylation on other aspects of HR have not been investigated.

As RNF138 promotes DNA end resection in HR[129,130], we next examined what impact mutations at T27, S124, and K158 would have on the efficiency of DNA end resection. We opted for a simultaneous knockdown and complementation approach in U2OS cells, depleting endogenous RNF138 and re-expressing exogenous, siRNA-resistant WT-RNF138 or versions where the PTM sites were mutated. We employed the tetracycline-regulated expression (TREx) system to enable inducible expression of exogenous RNF138 in the bulk of cells. This system was comprised of U2OS cells stably expressing the TetR tetracycline repressor (U2OS-TREx) being transfected with the pCDNA4-TO-hygromycin-sfGFP-MAP vector[185]. The vector contains Tet operator sequences which are bound by TetR, enabling transcriptional repression of the encoded transgene in the absence of tetracycline or doxycycline, but inducible expression in their presence. Using this vector, we generated U2OS-TREx cells stably expressing siRNA-resistant RNF138-WT or its mutants tagged with all four of the FLAG, 8XHis, streptavidin binding peptide (SBP), and superfolder GFP (sfGFP) tags[185]. When induced with doxycycline, these cell lines produced fusion proteins of ~65 kDa **(Figure 4.7G)**, from here on referred to as sfGFP-RNF138 constructs. The WT and K158R variants of RNF138 were expressed at similar levels as measured by quantitative immunoblotting **(Figure 4.7G,Figure *4.7*G)**. On the other hand, the T27A, T27E, S124A, and S124E mutants were expressed, on average, 2.7-3.7 times higher than WT. We attempted to adjust the doxycycline concentrations to induce similar expression levels of the variants, but these differences could not be eliminated at the concentrations tested (0.1 µg/mL – 5 μg/mL) **(Figure 4.4H)**. Perhaps the differences in expression reflect other factors among the cell lines, such as the frequency of vector integration or integration in heterochromatic versus euchromatic regions. Nonetheless, we achieved inducible expression of exogenous RNF138 in U2OS-TREx cells, and importantly the expression levels of WT/K158R, T27A/T27E, and S124A/S124E were comparable within each pair, respectively.

With the TREx system ready, we examined the ability for U2OS-TREx cells expressing these mutants to undergo end resection in response to treatment with CPT. CPT was chosen as it inhibits DNA Topoisomerase I (TOP1), stabilizing TOP1 cleavage complexes that, upon collision with a DNA replication fork, are converted into single-ended DSBs[215]. In this way, CPT-induced DSBs are generated in an S phase-dependent manner[215], biasing repair pathway choice to DNA end resection and HR. To monitor end resection, we detected the occurrence of ssDNA by immunofluorescence staining (IF). U2OS-TREx cells were labeled with the brominated nucleoside 5-bromo-2'-deoxyuridine (BrdU), which is incorporated into DNA. The process of resection generates ssDNA, exposing an epitope on BrdU that can be detected by IF under non-denaturing conditions. The appearance of BrdU foci in response to CPT is thus indicative of active DNA end resection[216,217]. As expected, treating parental U2OS-TREx cells with CPT significantly increased the intensity of BrdU foci **(Figure 4.11A-B,Figure 4.9A-C)**. In accordance with RNF138 being shown to promote DNA end resection[129,130], this intensity was significantly reduced when the cells were transfected with siRNA targeting RNF138 **(Figure 4.11A-B, Figure 4.9A-C)**. We next performed flow cytometric analysis to check if the difference was due to changes in the ratio of S phase cells. Depleting RNF138 in U2OS-TREx did not substantially alter the proportion of cells in S phase **(Figure 4.11E)**. Therefore, the change in BrdU focal intensity did not arise from a shift in the frequency of S phase-cycling cells, rather reflecting a perturbation in efficient DNA end resection.

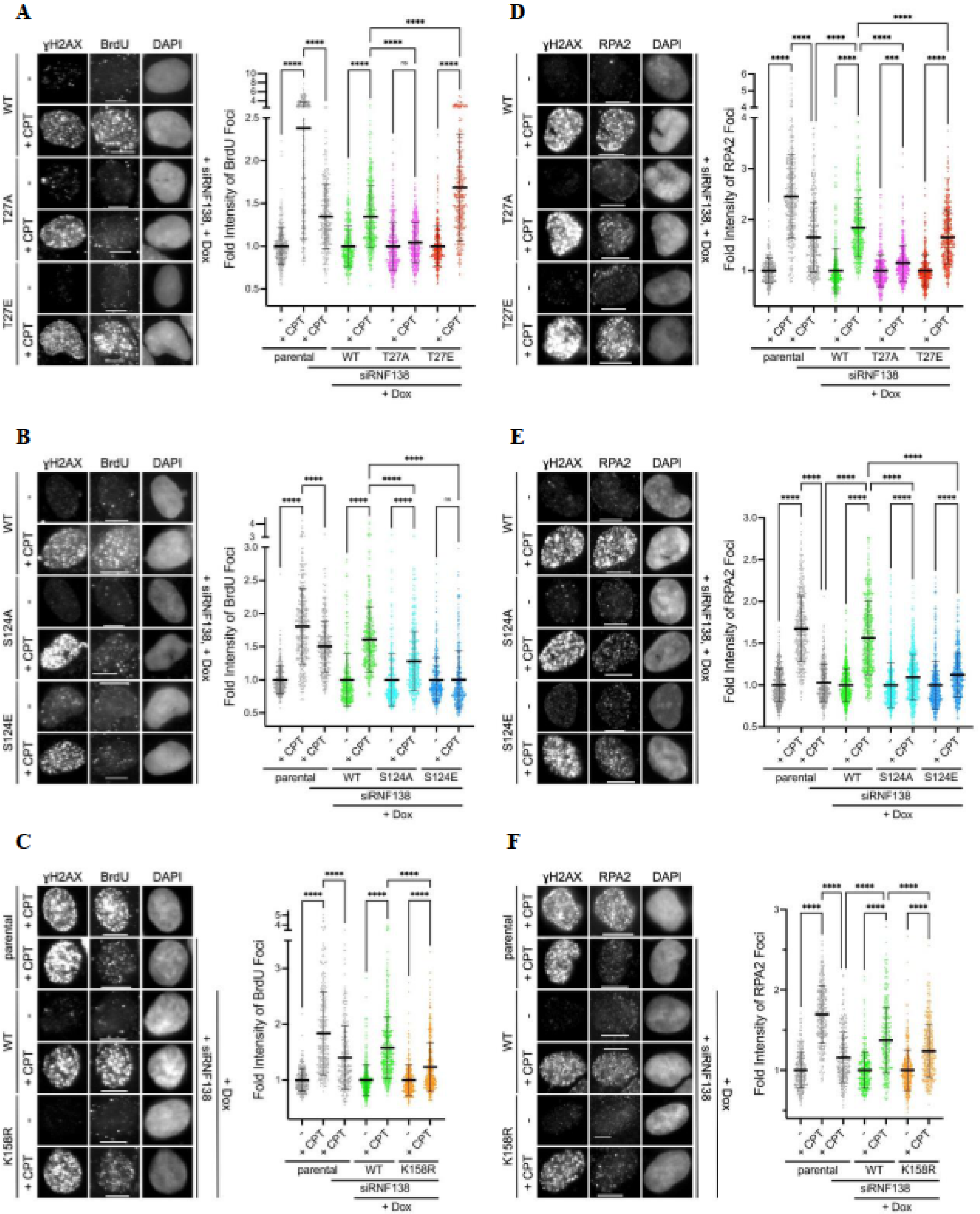

A
γH2AX
BrdU
DAPI
WT
T27A
T27E
-
+ CPT
+ siRNF138, + Dox
Fold Intensity of BrdU Foci
****
ns
parental
WT
T27A
T27E
siRNF138
+ Dox
B
γH2AX
BrdU
DAPI
WT
S124A
S124E
+ CPT
+ siRNF138, + Dox
Fold Intensity of BrdU Foci
parental
WT
S124A
S124E
siRNF138
+ Dox
C
γH2AX
BrdU
DAPI
parental
WT
K158R
+ CPT
+ siRNF138
+ Dox
Fold Intensity of BrdU Foci
parental
WT
K158R
siRNF138
+ Dox
D
γH2AX
RPA2
DAPI
WT
T27A
T27E
+ CPT
+ siRNF138, + Dox
Fold Intensity of RPA2 Foci
***
parental
WT
T27A
T27E
siRNF138
+ Dox
E
γH2AX
RPA2
DAPI
WT
S124A
S124E
+ CPT
+ siRNF138, + Dox
Fold Intensity of RPA2 Foci
parental
WT
S124A
S124E
siRNF138
+ Dox
F
γH2AX
RPA2
DAPI
parental
WT
K158R
+ CPT
+ siRNF138
+ Dox
Fold Intensity of RPA2 Foci
parental
WT
K158R
siRNF138
+ Dox

**Figure 4.9. The RNF138 PTM Sites T27, S124 and K158 are Important for DNA End Resection.**

**A., B., C.** Left panel: Representative BrdU immunofluorescence (IF) micrographs of sfGFP-RNF138 variant-expressing U2OS-TREx cells transfected with siRNA to RNF138 (siRNF138) and treated with doxycycline (Dox) to induce expression of sfGFP-RNF138 from a single biological replicate. C. includes parental U2OS-TREx cells transfected with or without siRNF138. In all, cells were treated with camptothecin (CPT) or not for 1 hour. ϒH2AX was used to indicate DNA damage, while DAPI stain labeled the nucleus. Right panel: quantifications of relative intensity of nuclear BrdU foci. For CPT-treated cells, only ϒH2AX$^+$ cells were quantified. The fluorescence intensities of foci for conditions within a given cell line were normalized to each other, with the average fluorescence intensity set at 1 for the untreated sample (no CPT). Averages were derived from the focal intensity of at least 150 cells each from 2 biological replicates (at least 300 cells total per condition). **D., E., F.** As in A.-C., but performing IF for RPA2 foci. Scale bars denote 10 µm. See **Figure 4.11A-D** for additional representative micrographs from experiments depicted in **Figure 4.9A,B,DE**. In all, ordinary one-way ANOVA with Šídák's multiple comparisons test was performed to determine whether differences between conditions were statistically significant. : $p$ values were <0.0001. For A), ns: $p = 0.8921$. For B), ns: $p > 0.9999$. For D), : $p = 0.0002$.

We next examined the impact of RNF138 mutations on DNA end resection. U2OS-TREx cells were depleted of endogenous RNF138 and complemented with WT or mutated sfGFP-RNF138 by treatment with doxycycline. In agreement with T27 being an important phosphorylation site, average BrdU focal intensity in the cell line expressing the T27A mutant was substantially reduced compared to cells expressing WT-RNF138, and at levels barely above that of the same cells not treated with CPT **(Figure 4.9A)**, indicating dramatically reduced end resection. The contrary was observed for the T27E mutant, with mean BrdU focal intensity extensively higher than WT **(Figure 4.9A)**. Perhaps the negatively charged glutamate substitution results in constitutively active RNF138, constantly promoting resection. Hence, the phosphorylation of RNF138 at T27 plays a crucial role in RNF138's ability to promote DNA end resection. Performing the same assay on the S124A mutant showed reduced BrdU focal intensity relative to WT **(Figure 4.9B)**, demonstrating a role for S124 phosphorylation in RNF138 function.

Interestingly, unlike T27E, the S124E mutation was unable to restore BrdU focal intensity **(Figure 4.9B)**. Perhaps glutamate substitution at S124 is insufficient to cause the same conformational changes elicited by serine phosphorylation. Nevertheless, the fact that both S124A and S124E mutations reduce BrdU intensity underscores the importance of the S124 site in RNF138 function. We also performed the same assay for the K158R substitution mutant. U2OS-TREx cells expressing sfGFP-RNF138-K158R exhibited reduced intensity in BrdU foci relative to cells expressing WT **(Figure 4.9C)**, indicating K158 plays a role in RNF138's ability to promote DNA end resection. As a control, we confirmed that concurrently depleting endogenous RNF138 while inducing expression of sfGFP-RNF138-WT with doxycycline did not severely alter the proportion of cells in S phase **(Figure 4.11E)**. We also checked if this proportion was affected by expressing any of the RNF138 mutants. The S phase fraction was similar (~31-37%) among all the U2OS-TREx cell lines, whether expression of the sfGFP-RNF138 mutants was induced or not **(Figure 4.11F)**. Thus, the observed differences in BrdU focal intensity were not from cells being more or less responsive to CPT from changes in the fraction in S phase.

As active end resection generates ssDNA overhangs that are rapidly coated and protected by the binding of RPA complexes[171,216,218,219], we asked if the observed changes in BrdU focal intensity would correlate with RPA binding. To do so, we performed a similar IF assay in U2OS-TREx, this time detecting extraction-resistant foci of the RPA complex subunit RPA2 in response to CPT. As seen with BrdU foci **(Figure 4.9A-C,Figure *4.11*A,B)**, knocking down RNF138 reduced RPA focal intensity **(Figure 4.9D-F,Figure *4.11*C,D)**, indicating the inhibition of end resection. Notably, CPT-induced RPA focal intensity was partially rescued upon concurrent expression of sfGFP-RNF138-WT **(Figure 4.9D-F)**. As well, the T27A, S124A, and K158R variants exhibited significantly reduced RPA2 focal intensity relative to WT-RNF138 **(Figure**

**4.9D-F)**. Likewise, RPA2 focal intensity was rescued by the T27E mutant, but not by the S124E mutant **(Figure 4.9D)**. Our results with RPA2 foci thereby recapitulate our findings seen with native BrdU foci.

The N-terminal region of RPA2 is phosphorylated by the PIKKs in response to DNA damage[219]. As Ser4 and Ser8 phosphorylation on RPA2 (P-S4/S8-RPA2) is dependent on CtIP [130,174] and occurs after the binding of RPA complexes to ssDNA[219–221], it has been used as a readout for end resection[130,200]. To further support that end resection is regulated by PTMs on RNF138, we detected CPT-induced P-S4/S8-RPA2 in the same cell lines by immunoblot **(Figure 4.10A,B)**. Consistent with our previous findings[129], cells transfected with siRNA to RNF138 exhibited reduced RPA2 phosphorylation upon stimulation with CPT. Cells knocked down of RNF138 and expressing sfGFP-RNF138-T27A, -S124A, -S124E, or -K158R could not restore P-S4/S8-RPA2 to the levels seen with WT-RNF138, agreeing with our IF data with BrdU and RPA2 foci. In fact, the cells exhibited P-S4/S8-RPA2 levels below that of parental U2OS-TREx cells treated with RNF138-targeting siRNA. Unlike the IF data however, the T27E mutant did not recover P-S4/S8-RPA2 to levels near or above WT-RNF138 **(Figure 4.10A,B,Figure *4.11*A,D)**. It may be that the resection induced by T27E, seen by IF, somehow cannot trigger RPA2 phosphorylation as effectively as having a phosphoryl group on T27, and that RPA2 phosphorylation requires events additional to RPA binding ssDNA. Still, this data reveal the importance of the T27 site in early RNF138-dependent DSB signaling. Overall, the analysis of BrdU foci, RPA2 foci, and phospho-RPA2 levels suggest phosphorylation at T27 and S124 and ubiquitylation at K158 are important for RNF138's role in promoting DNA end resection.

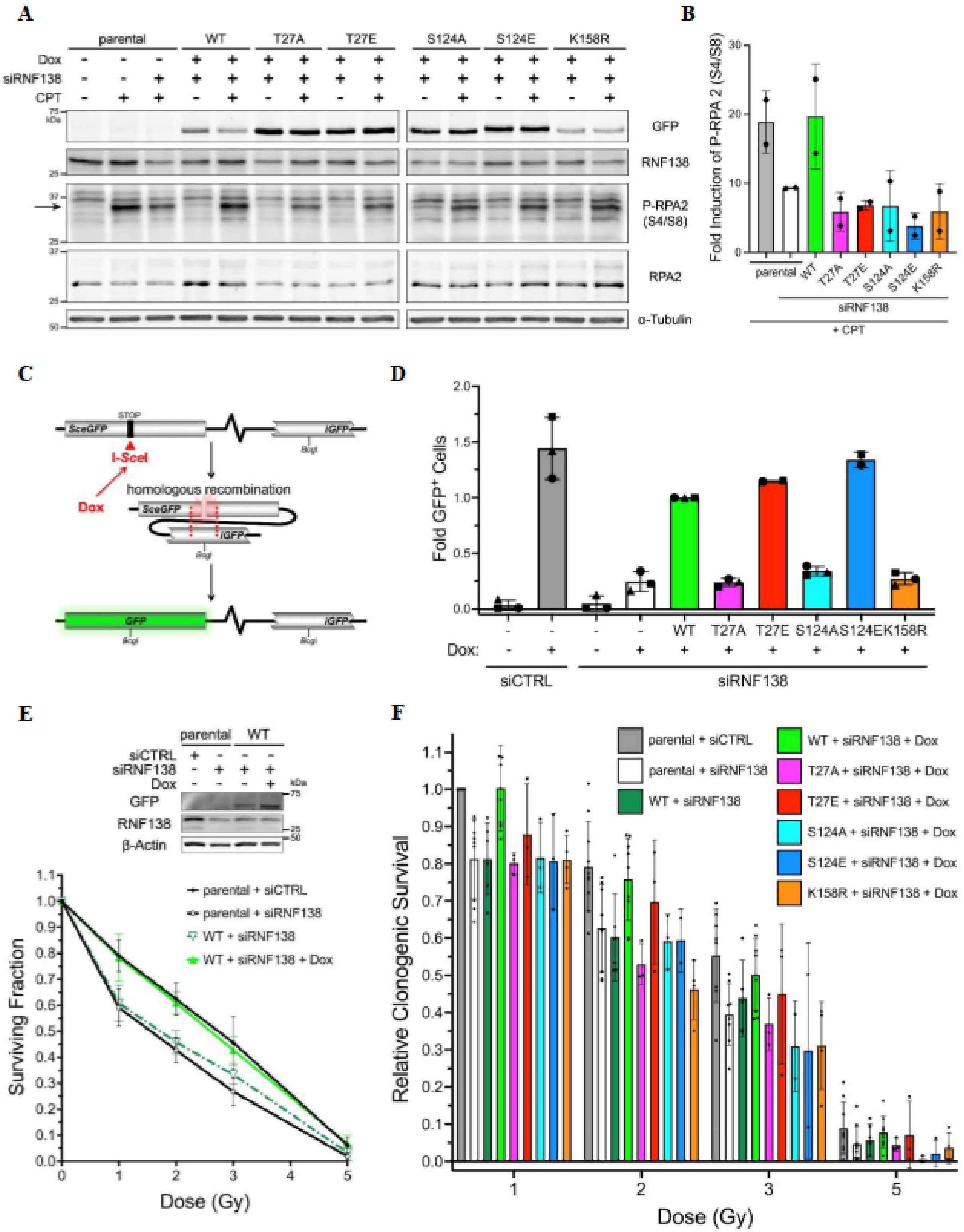

A
parental
WT
T27A
T27E
S124A
S124E
K158R
Dox
siRNF138
CPT
GFP
RNF138
P-RPA2 (S4/S8)
RPA2
α-Tubulin
75 kDa
25
37
25
37
25
50
B
Fold Induction of P-RPA 2 (S4/S8)
30
20
10
0
parental
WT
T27A
T27E
S124A
S124E
K158R
siRNF138
+ CPT
C
SceGFP
STOP
I-SceI
iGFP
Bcgl
Dox
homologous recombination
SceGFP
iGFP
Bcgl
GFP
Bcgl
iGFP
Bcgl
D
Fold GFP+ Cells
2.0
1.5
1.0
0.5
0.0
WT
T27A
T27E
S124A
S124E
K158R
Dox:
siCTRL
siRNF138
E
parental
WT
siCTRL
siRNF138
Dox
GFP
RNF138
β-Actin
kDa
75
25
50
Surviving Fraction
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
parental + siCTRL
parental + siRNF138
WT + siRNF138
WT + siRNF138 + Dox
Dose (Gy)
0
1
2
3
4
5
F
Relative Clonogenic Survival
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
parental + siCTRL
parental + siRNF138
WT + siRNF138
WT + siRNF138 + Dox
T27A + siRNF138 + Dox
T27E + siRNF138 + Dox
S124A + siRNF138 + Dox
S124E + siRNF138 + Dox
K158R + siRNF138 + Dox
1
2
3
5
Dose (Gy)

**Figure 4.10. T27, S124, and K158 on RNF138 are important for RPA2 Phosphorylation, HR activity, and cell survival.**

**A.** Immunoblot of whole cell extracts of parental or sfGFP-RNF138 variant-expressing U2OS-TREx cells transfected with or without siRNA to RNF138 and treated with camptothecin (CPT) or not for 1 hour. Arrow: main band for P-RPA (S4/S8); note its reduced electrophoretic mobility compared to unmodified RPA2. **B.** Quantification of fold induction of P-RPA2 (S4/S8) signal from A). P-RPA2 (S4/S8) signal was normalized to that of RPA2 expression. The resulting values from CPT-treated samples were then ratioed to those left untreated. **C.** Schematic diagram of the DR-GFP reporter assay in TRI-DR-U2OS cells expressing doxycycline (Dox)-inducible I-*Sce*I. **D.** DR-GFP reporter assay in TRI-DR-U2OS cells transfected with siRNA to luciferase (siCTRL) or RNF138 (siRNF138) and complemented with siRNA-resistant mCherry-RNF138 variants or not. $GFP^+$ cells were quantified by flow cytometry. **E.** Bottom: Clonogenic survival assay for parental or sfGFP-RNF138-WT-expressing U2OS-TREx cells transfected with siCTRL or siRNF138, induced with Dox if necessary, and treated with increasing doses of ionizing radiation (IR). Top: immunoblot of endogenous and exogenous RNF138 expression of cells used for the clonogenic survival assay at time of irradiation. **F.** Clonogenic survival assay for parental or sfGFP-RNF138 variant-expressing U2OS-TREx cells transfected with siCTRL or siRNF138, induced with Dox or not, and treated with increasing doses of IR. Within each biological replicate, the surviving fractions obtained were normalized to the surviving fraction of the parental cells + siCTRL + 1 Gy condition. A) is a representative result from at least 2 biological replicates. Averages were calculated from 2 (B), at least 2 (D, F), and 6-10 (E) biological replicates pooled together.

### 4.4.8 T27, S124 and K158 on RNF138 are Important for HR and Cell Survival

In HR, the processes of homology search, strand invasion, and DNA synthesis occur downstream of DNA end resection (2). As T27, S124, and K158 affected the efficiency of end resection, we next determined if the frequency of HR was impacted by the same mutations. To do so, we used the direct repeat GFP (DR-GFP) reporter assay (67), utilizing U2OS cells integrated with two repeats of the GFP open reading frame (ORF) **(Figure 4.10C)**. These copies are mutated such that neither produces a fluorescent protein product; the upstream copy contains stop codons and a recognition site for the restriction endonuclease I-*Sce*I, while the downstream copy is truncated, with the N- and C-terminal regions removed. Expression of I-*Sce*I induces a site-specific DSB in the upstream copy of the GFP ORF. If the DSB is repaired by HR using the downstream copy as a template, the sequence encoding intact, fluorescent GFP is generated. Consequently, the proportion of GFP$^+$ cells after I-*Sce*I expression reflects the efficiency of HR. We generated siRNA-resistant, mCherry-tagged constructs of WT-RNF138 and its PTM site mutants (T27A, T27E, S124A, S124E, K158R), then transfected them into TRI-DR-U2OS cells stably expressing doxycycline-inducible I-*Sce*I and depleted of endogenous RNF138. As expected, I-*Sce*I induction alone markedly increased the frequency of GFP$^+$ cells, and this frequency was sharply inhibited when the cells were depleted of RNF138 (14) **(Figure 4.10D)**. Consistent with our IF data for end resection detected by RPA2 foci **(Figure 4.9D-F,Figure *4.11*C,D)**, adding back mCherry-RNF138-WT and -T27E in the presence of RNF138 siRNA partially rescued the frequency of HR to levels seen in cells expressing endogenous RNF138 **(Figure 4.10D)**. On the other hand, cells expressing the -T27A, -S124A, and -K158R mutants exhibited severely reduced HR frequencies **(Figure 4.10D)**. Interestingly, the S124E mutant was fully capable of restoring HR frequency, in fact to levels surpassing that of WT-RNF138, similar to the T27E mutant **(Figure 4.10D)**. It appears that in this specific context (DR-GFP reporter assay, mCherry-tagged RNF138), S124E resembles the actions of constitutively phosphorylated RNF138. Thus, the T27, S124, and K158 sites are important for RNF138's ability to promote HR.

With end resection and the occurrence of HR dependent on the aforementioned PTM sites, we asked if cell survival upon DNA damage would also be impacted by mutations at the sites. To do so, we performed clonogenic survival assays on U2OS-TREx cells treated with 1 to 5 Gy of IR. Knocking down RNF138 impeded the colony forming ability of parental U2OS-TREx **(Figure 4.10E)**, in line with previous observations (14, 20). This was not due to indirect effects on the cell cycle from RNF138 depletion, as the proportion of cells in S or G2 phase (and therefore conducive to HR) increased slightly when RNF138 was depleted **(Figure 4.11B)**. Conversely, clonogenic survival was rescued in sfGFP-RNF138-WT-expressing U2OS-TREx cells depleted of endogenous RNF138 **(Figure 4.10E)**. As a control, when doxycycline was not added to the culture medium, preventing expression of sfGFP-RNF138-WT, survival was reduced **(Figure 4.10E)**. Again, these effects were not from indirect changes on the cell cycle, as the S/G2 fraction of cells was similar whether doxycycline was added or not **(Figure 4.11E)**. Having confirmed that survival to IR is dependent on RNF138, we tested how the PTM site mutants would fare in colony forming ability using U2OS-TREx cells stably expressing the sfGFP-RNF138 variants. Consistent with our IF results for resection **(Figure 4.9A-F)**, cells expressing the T27A, S124A, S124E, and K158R mutants exhibited relative surviving fractions similar to or below that of parental U2OS-TREx cells depleted of RNF138 **(Figure 4.10F)**. At 1, 2, and 3 Gy, cells expressing the T27E mutant showed intermediate survival, greater than the former mutants but less than WT-RNF138. None of these effects were from fluctuations in the cell cycle distribution, as the stable cell lines did not show appreciable differences in the S/G2 phase fraction (~58% to 63.6%) upon doxycycline induction **(Figure 4.11F)**.

All in all, these data exemplify the role of the T27, S124, and K158R residues in promoting HR and cellular survival in the face of DSBs. They are also mostly concordant with the phenotypes observed for the mutants in DNA end resection, assayed by RPA2 and BrdU foci formation **(Figure 4.9A-F,Figure 4.11A-D)**.

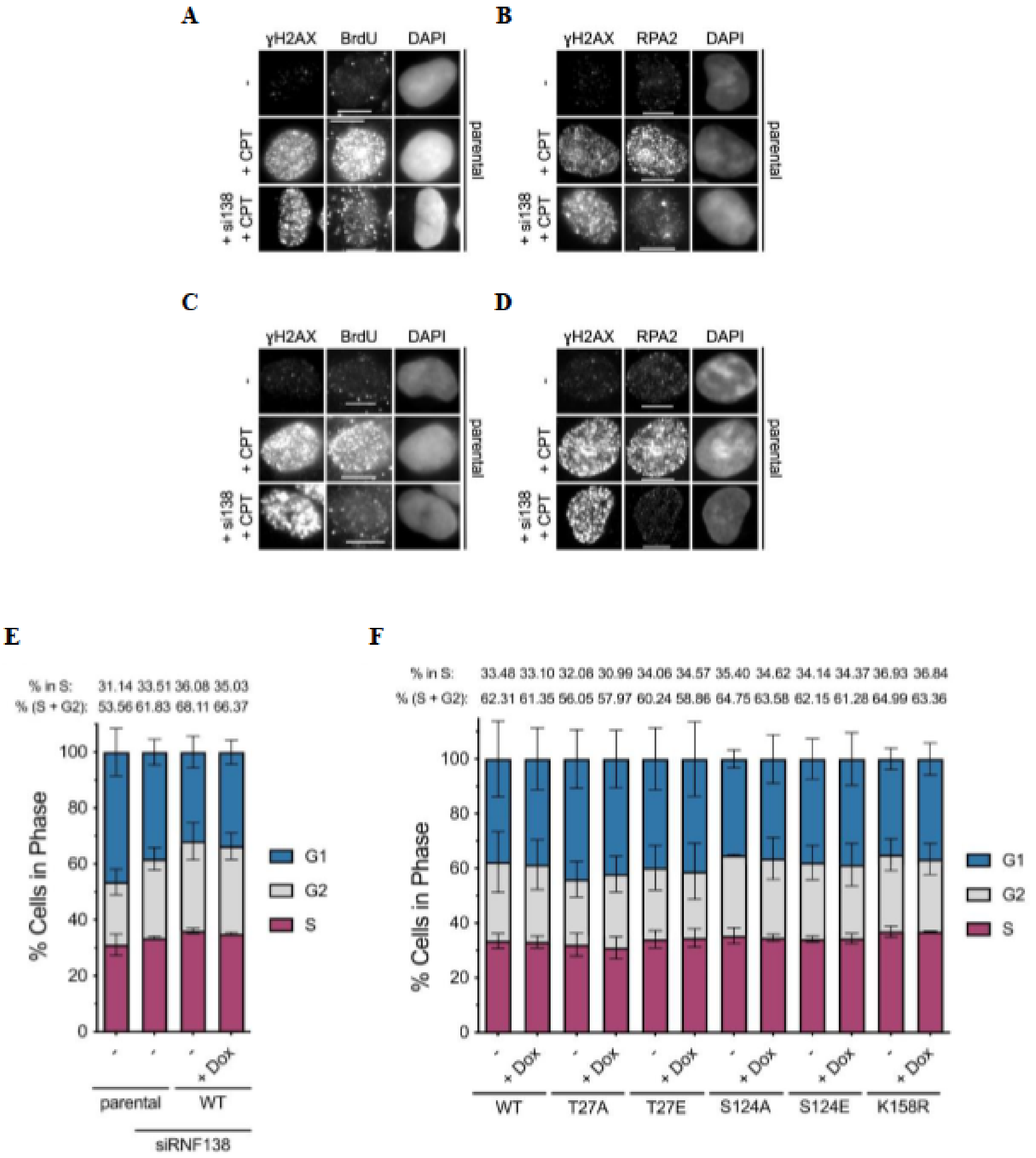


**Figure 4.11. Related to Figures Figure 4.9 and Figure 4.10.**

**A. and B.** Additional representative micrographs from **Figure 4.9**. Representative BrdU immunofluorescence (IF) micrographs of parental U2OS-TREx cells transfected with or without siRNA to RNF138 (siRNF138) and treated with camptothecin (CPT) or not for 1 hour. ƔH2AX was used to indicate DNA damage, while DAPI stain labeled the nucleus. Micrographs are derived from the same biological replicates presented in the left panels of **Figure 4.9** (A) and **Figure 4.9B** (B), respectively. **C. and D.** As per A) and B), but for RPA2 foci instead. Micrographs are derived from the same biological replicates presented in the left panels of **Figure 4.9D** (C) and **Figure**

**4.9E** (D), respectively. **E.** Flow cytometric analysis of propidium iodide signal in parental or sfGFP-RNF138-WT-expressing U2OS-TREx cells transfected with siRNA to RNF138 or not, and with or without doxycycline (Dox) induction. **F.** Flow cytometric analysis of propidium iodide signal in sfGFP-RNF138 variant-expressing U2OS-TREx cells induced with Dox or not. Averages were calculated from 2 (F) or at least 2 (E) biological replicates pooled together. Scale bars denote 10 µm.

### 4.4.9 Investigating the Interplay Between Post-Translational Modifications on RNF138

Lastly, we investigated whether interplay occurs between the PTMs on RNF138. We first performed immunoprecipitation experiments on the single PTM site substitution variants of GFP-RNF138 expressed in HEK293 cells. We assessed the impact of S124A and K158R on the P-TP signal **(Figure 4.12A)**, the impact of T27A and K158R on the CPT-induced P-S/TQ signal **(Figure 4.12B)**, and the impact of T27A and S124A on RNF138 ubiquitylation by HA-Ub-WT **(Figure 4.12C)**. We observed that the K158R substitution had no effect on the P-TP and P-S/TQ signals on RNF138 **(Figure 4.12A,B)**, concluding that K158 is dispensable for phosphorylation events at T27 and S124. Next, the S124A substitution partially reduced the P-TP signal **(Figure 4.12A)**, suggesting S124 phosphorylation may support but is not required for phosphorylation on T27. However, the T27A mutation strongly impaired the P-S/TQ signal **(Figure 4.12B)**. As well, basal higher-order ubiquitylation on GFP-RNF138 was notably reduced by the T27A and S124A mutations, preventing the decrease in RNF138 ubiquitylation induced by CPT from being seen **(Figure 4.12C,Figure *4.5*D,E)**. Therefore, T27 phosphorylation may serve as a prerequisite for S124 phosphorylation, and functional T27 and S124 sites seem to be important for RNF138 ubiquitylation.

With this interdependency in mind, we asked what effect expressing the T27A-S124A (TS), T27A-K158R (TK), and S124A-K158R (SK) double mutants, as well as the T27A-S124A-K158R (TSK) triple mutant, would have on the ability for cells to perform HR. The aforementioned mutations were incorporated in siRNA-resistant mCherry-RNF138 by sequential site-directed mutagenesis, the products of which were used in the DR-GFP HR reporter assay upon concurrent depletion of endogenous RNF138. Expressing the double and triple mutants resulted in similar low HR frequencies upon I-*SceI* expression **(Figure 4.12D)**. These values were also similar to those for the T27A, S124A, and K158R single mutants, which were already low to start and similar to HR frequencies for cells depleted of endogenous RNF138. As ablation of any of the PTM sites, or any combination of them, was sufficient to impair HR to baseline levels, this finding emphasizes the importance of all three sites in the function of RNF138 and suggests they could all act within the same pathway on RNF138. This is in agreement with the dependence of RNF138 ubiquitylation on the T27 and S124 residues **(Figure 4.12C)**, and the dependence of S124 phosphorylation on the T27 site **(Figure 4.12B)**.

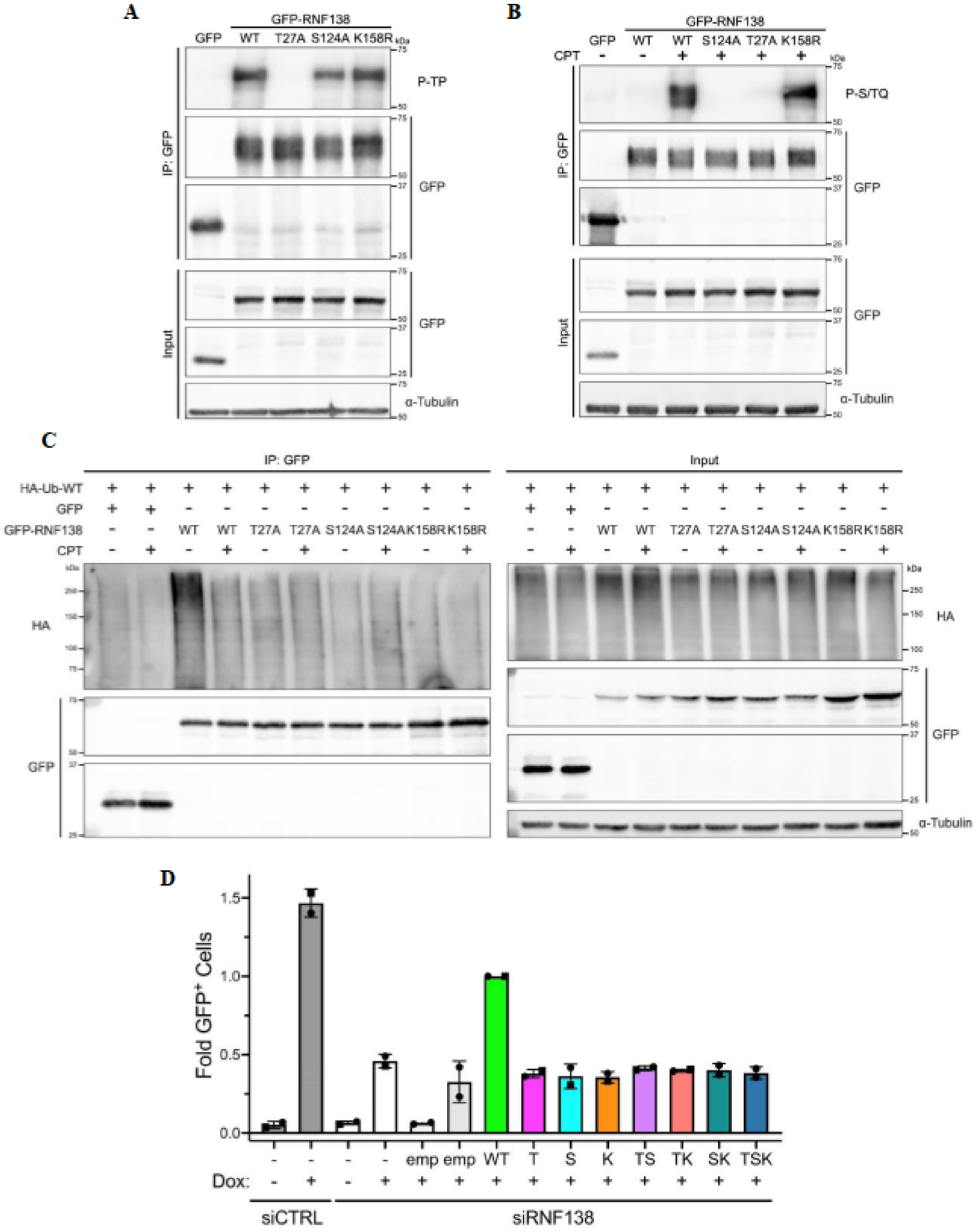

A
GFP-RNF138
GFP
WT
T27A
S124A
K158R
kDa
IP: GFP
P-TP
GFP
Input
GFP
α-Tubulin
B
GFP-RNF138
GFP
WT
WT
S124A
T27A
K158R
CPT
IP: GFP
P-S/TQ
GFP
Input
GFP
α-Tubulin
C
IP: GFP
Input
HA-Ub-WT
GFP
GFP-RNF138
CPT
HA
GFP
α-Tubulin
D
Fold GFP⁺ Cells
1.5
1.0
0.5
0.0
emp
WT
T
S
K
TS
TK
SK
TSK
Dox:
siCTRL
siRNF138

**Figure 4.12. Investigating the Interplay Between Post-Translational Modifications on T27, S124, and K158R on RNF138.**

**A., B.** Stringent GFP immunoprecipitations (IP) were performed for HEK293 cells expressing GFP or GFP-RNF138 variants and immunoblotted for phosphorylated TP sites (P-TP) (A) or treated with camptothecin (CPT) or not for 1 hour and immunoblotted for phosphorylated S/TQ sites (P-S/TQ) (B). **C):** Stringent GFP IP and subsequent immunoblotting was performed for HEK293 cells co-expressing HA-Ub-WT and GFP or GFP-RNF138 variants and treated with CPT or not for 1 hour. In all samples, 10 µM MG132 was added to the culture medium starting 1 hour prior to CPT treatment to improve preservation of HA-Ub signals. **D.** DR-GFP reporter assay in TRI-DR-U2OS cells transfected with siRNA to luciferase (siCTRL) or RNF138 (siRNF138) and complemented with mCherry empty vector (emp) or siRNA-resistant mCherry-RNF138 variants or not. I-*Sce*I expression was induced by doxycycline (Dox). $GFP^+$ cells were quantified by flow cytometry. WT: wildtype, T: T27A, S: S124A, K: K158R, TS: T27A-S124A, TK: T27A-K158R, SK: S124A-K158R, TSK: T27A-S124A-K158R. Shown are representative results from 2 (A, B) and 3 (C) biological replicates. Averages in D) were calculated from 2 biological replicates pooled together.

## 4.5 Discussion

Multiple reports have implicated a role for RNF138 in the DNA damage response[35,127–131]. As an E3 ubiquitin ligase, RNF138 facilitates the ubiquitylation of Ku80, CtIP, and Rad51D, actions that promote the occurrence of HR[35,129–131]. Little however is known about how its activity is regulated in the context of DSB repair. While RNF138 was reported to be phosphorylated by ATM in response to IR at position S124[131], the functional significance of this modification remained unclear, as it did not affect RNF138's recruitment to sites of DNA damage[131]. Furthermore, as RNF138 stimulates Ku80 ubiquitylation in S/G2, but not G1 phase[129], it appeared RNF138 activity could be regulated by the cell cycle, although how this occurred was not previously explored. Thus, we sought to elucidate additional mechanisms by which RNF138 could be regulated. In HeLa cells, we detected a slight increase in RNF138 protein levels during G2 phase (**Figure 4.1C**), consistent with the reported increase in RNF138 messenger RNA transcripts in G2[122]. However, RNF138 protein was expressed throughout the cell cycle, and at relatively constant levels (**Figure 4.1A-C**). As HR is specific to the S/G2 phases[178], we speculate RNF138 expression in G1 may serve its roles outside of HR, such as binding interactors and ubiquitylating substrates in NFκB or Wnt-β-catenin signaling[121,124,125,222]. Regardless, it became clear that factors beyond protein level regulate RNF138 activity in HR. We located a putative CDK substrate consensus motif within RNF138's amino acid sequence **(Figure 4.1D)**. In addition, RNF138 was a hit in proteomic screens for ubiquitylated proteins[210–213], and ubiquitylation events are well known to govern the response to DSBs[181]. We therefore addressed if ubiquitylation and CDK-dependent phosphorylation occurred on RNF138 and were involved in its function in HR.

We report here that RNF138 is phosphorylated in a CDK-dependent manner. This phosphorylation peaked during S phase, was dependent on CDK1 and CDK2, occurred on residue T27 **(Figure 4.1G-I,K,L,Figure *4.4*D)**, and, importantly, stimulated RNF138-mediated DNA end resection **(Figure 4.9A,D,Figure *4.10*A,B)**. While the precise mechanism by which CDK phosphorylation activates RNF138 requires further study, it aligns with a pro-resection function for CDK activity in S/G2. RNF138 has been shown to mediate ubiquitylation of the pro-resection factor CtIP, enabling CtIP's recruitment to DSBs[130]. RNF138-dependent ubiquitylation of Ku80 also promotes the eviction of the DNA end-protecting and NHEJ-promoting[176,177] Ku heterodimer from chromatin[129]. As Ku displacement is required for DNA end resection to begin[24,179], this contributes a second mode to trigger end resection. Notably, CtIP itself is phosphorylated by CDK activity, the phosphorylated version activating the endonuclease activity of Mre11 to initiate DNA end resection[172,173,199–201,223]. The Ku heterodimer is also a substrate for CDK activity; Ku70 is phosphorylated by CDK1 and CDK2 during S, G2, and M phase[224], and in budding yeast, CDK-phosphorylated Yku80 promotes HR and impairs NHEJ[225]. Phosphorylated Ku, similar to ubiquitylated Ku[226], has been found to dissociate from DNA[24,224]. The evidence points to a model where concerted CDK activity in S and G2, acting on multiple fronts, like RNF138, CtIP, Ku, and other resection factors[196–198,202,203], culminates to drive the process of DNA end resection.

Interestingly, our laser microirradiation data also suggests that beyond promoting DNA end resection, T27 phosphorylation may serve a minor role in regulating RNF138's retention on chromatin **(Figure 4.7D,E)**.

We also find that RNF138 is constitutively polyubiquitylated, with K158 serving as a site of modification **(Figure 4.3A-D,Figure *4.6*B,Figure *4.12*C)**. This is in line with a recent report identifying K158 as a ubiquitylation site on RNF138, with the K158R substitution able to suppress negative regulation of oncogenic MyD88 signaling[124]. As the K158R mutant could not completely eliminate HA-Ub-L73P signal in GFP-RNF138 immunoprecipitates **(Figure 4.6B)**, there are likely other residues by which RNF138 is ubiquitylated. Indeed, four additional lysine residues on RNF138 were detected to be ubiquitylated in one proteomic screen[212]. Yet, K158 is probably an abundant site for ubiquitylation, as it was the only ubiquitylation site predicted by UbPred software[124], and was consistently ubiquitylated in four proteomic screens, detected in three of these as the only site of modification on RNF138[210–213]. Phenotypically, despite exhibiting reduced turnover and prolonged stability, the K158R mutant was still clearly defective in promoting DNA end resection **(Figure 4.6C,D,Figure 4.9C,F,Figure *4.10*A,B)**. Although K158 was not essential for RNF138 recruitment to laser stripes, the K158R mutant also showed a minor impairment in recruitment **(Figure 4.7C)**, and this change could contribute at least partly to the defect in end resection. Overall, the data suggests ubiquitylation on K158 activates RNF138 for end resection. This is supported by the increase in ubiquitylated RNF138 ubiquitylation in the S and G2 phases **(Figure 4.5B,C)**, overlapping with the times end resection and HR are active. As RNF138 is appreciably expressed during all cell cycle phases **(Figure 4.1C)**, we suspect the enhanced ubiquitylation in S/G2 serves more to activate the protein for end resection and HR rather than actively target it for proteasomal degradation. Further study, particularly of the linkages in the ubiquitin chains conjugated to RNF138, is required to validate this idea. We also observed that RNF138 ubiquitylation was reduced upon DSBs, UV, and replication stress **(Figure 4.5D,E, Figure 4.12C)**. In support of this, a proteomic screen found ubiquitylation at K158 was reduced 3 hours after HEK293 cells were exposed to UV irradiation[210]. Since K158 ubiquitylation promotes DNA end resection **(Figure 4.9C,F,Figure *4.10*A,B)**, the reduction in ubiquitylation, seen after 1 hour of genotoxic stress **(Figure *4.5*D,E)**, may reflect a negative regulatory mechanism, constraining RNF138 activity after actions to resolve the stress have commenced. For HR, this might prevent overstimulation of DNA end resection, which could lead to a loss of genetic information. Uncontrolled resection might also deplete local RPA pools, causing aberrant annealing, secondary structures, and degradation in unprotected ssDNA[219]. Together, our data suggest RNF138 is regulated by ubiquitylation, with ubiquitin conjugation to K158 serving to activate the protein in DSB repair.

Our work thus identifies two additional PTMs that contribute to RNF138's function in DSB repair, the aforementioned K158 ubiquitylation and CDK-dependent T27 phosphorylation. Both promote RNF138's role in stimulating DNA end resection, as the T27A and K158R variants inhibit RNF138-dependent resection in response to CPT **(Figure 4.9A,C-D,F,Figure *4.10*A,B)**. Using the S124A mutant, we also demonstrate S124 phosphorylation positively regulates RNF138's role in end resection **(Figure 4.9B,E,Figure *4.10*A,B)**, providing functional significance to the known ATM-dependent modification[131]. Aligning with these observations, the T27A, S124A, and K158R mutations dramatically reduced HR frequency *in vivo* **(Figure 4.10D)** and sensitized cells to DNA damage by IR **(Figure 4.10F)**, suggesting the defects in end resection translated to negative consequences downstream. Supporting the importance of T27 phosphorylation, the phospho-mimicking T27E variant could restore DNA end resection, HR, and clonogenic survival **(Figure**

**4.9A,D,Figure *4.10*D,F)**, despite having a partial defect in retention on chromatin **(Figure 4.7D,E)**. Intriguingly, like T27E, the S124E mutant was also capable of restoring HR **(Figure 4.10D)**, although it was unable to rescue end resection and clonogenic survival **(Figure 4.9B,E,Figure *4.10*A,B,F)**. We note that the end resection and survival assays were conducted with sfGFP-RNF138, while the HR reporter assay utilized the mCherry-RNF138 fusion. Beyond the different fluorescent proteins, the sfGFP construct also contains three other epitope tags[185]. We speculate construct-specific conformational differences could explain the discrepancy in results for S124E. Alternatively, it could be that the S124E variant does not sufficiently promote end resection yet is capable of activating HR through a separate mechanism. Indeed, we did not assess if the stability or recruitment of Rad51D, another target of RNF138 ubiquitylation in HR[35,131], were impacted by the PTM site mutations in RNF138. We thus cannot conclude if the effects on HR and survival arise solely from changes in end resection or if alterations in Rad51D function also contribute. Nevertheless, our functional readouts reiterate that PTMs at T27, S124, and K158 are important to the role of RNF138 in protecting cells from DSBs.

Our data ultimately point to a scenario where phosphorylations at T27 and S124 and ubiquitylation at K158 all positively regulate RNF138 activity in end resection and HR. This may enable RNF138 to integrate signals of both DNA damage and cell cycle stage, ensuring it is fully active when DSB breaks occur during S/G2 phase. An attractive idea is that cell cycle-dependent ubiquitylation and CDK-dependent T27 phosphorylation prime RNF138 to function in S/G2. ATM-dependent phosphorylation at S124, triggered by DSBs, could perhaps give the final go-ahead signal to license RNF138 activity. Support for this notion may come from the observation that a functional T27 site is important for the DNA damage-dependent phosphorylation of RNF138 on S124 **(Figure 4.12B)**, suggesting T27 phosphorylation could precede phosphorylation on S124.

This of course needs to be tested further, and more study is needed to decipher the molecular mechanisms by which each of the PTMs activate RNF138. In the end, our data provide additional intricacies to the tightly orchestrated molecular events triggered upon DSB damage.

## 4.6 Funding

Funding for this work was provided by a CIHR (Canadian Institutes of Health Research) Project Grant (grant number 154485) to I.H.I., a CIHR (Canadian Institutes of Health Research) Project Grant (grant number 168972) to J.N.M.G., an NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant (RGPIN-2017- 05752) to I.H.I, and CRINA (Cancer Research Institute of Northern Alberta) Bridge Funding to I.H.I. provided by Terry and Betty Davis. A.J.L. was supported by the Rachel Mandel Scholarship in Lymphoma and Other Blood Cancers, the Dr. Herbert Meltzer Memorial Fellowship, and the Yau Family Foundation Award. R.A.F was supported by Alberta Graduate Excellence Scholarship, and David Lawson Award.

## 4.7 Acknowledgements

We are grateful to Joseph Drake, Allison Tran, Adham Shousha, Kristi Baker, and Anne Galloway for technical assistance.

## 4.8 Conflict of Interest

The authors declare that they have no conflicts of interest with the contents of this article.

## 4.9 Author Contributions

A.J.L. and R.A.F. formal analysis and visualization; A.J.L., R.A.F., C.T., E.Z., F.M., J.Y.H.A. and I.H.I. investigation; A.J.L. and I.H.I. methodology and writing – original draft; A.J.L., R.A.F., J.N.M.G. and I.H.I. writing – review & editing; I.H.I. and J.N.M.G. funding acquisition and supervision; I.H.I. conceptualization.

# Chapter 5 – Conclusions and Future Directions

## 5.1 Brief Study Overview

This thesis presents comprehensive investigations into the E2 conjugating enzyme Ubc13 and the RING E3 ligase RNF138. Structural characterization of Ubc13 was conducted along with a study of the modulation of its activity using small molecule inhibitors. Additionally, we explored the DNA binding activity of the RING E3 ligase RNF138, gained insights into its structure using AlphaFold2, and examined its regulation in mammalian cells, particularly in response to DNA damage.

## 5.2 Conclusions and Future Directions: Characterization of Ubc13

### 5.2.1 New Ubc13 Structure Reveals New Active Site Loop Conformation

Our work began when we crystallized a new form of Ubc13-Mms2 under more native conditions than had been used previously. E2 enzymes feature a gating loop near the active site cysteine, shielding it from solvent effects and controlling ubiquitin access[88,100,103,105,106]. Although all E2s possess this loop, Ubc13's gating loop (residues 114-124) displays a unique conformation within the E2 family[89,100,103]. L121, a conserved residue, typically obstructs the attack of the acceptor ubiquitin K63 on the active site C87[89,103]. When C87 is bound to the C-terminus of a donor ubiquitin, the gating loop shifts to enable access to C87 by reorienting Leu121, rearranging Ubc13 to a conformation similar to that seen in other E2s for example, UbcH5c[55,88,89]. This gating mechanism is believed to be exclusive to Ubc13 and plays a pivotal role in its ability to synthesize Lys63 ubiquitin chains[55,88,89]. Previous work by our group has explored the active site of Ubc13 using small molecule compounds[100]. This approach is based on the premise that the unique structure of the Ubc13 gating loop enables selective inhibition by small molecules[89,100]. NSC697923 – discovered through a high-throughput cell-based assay targeting NF-kB signaling inhibitors – specifically inhibits Ubc13 without affecting other E2s in ubiquitination assays[116]. This inhibition mechanism involves the covalent modification of active C87 by NSC697923, dependent on a pocket formed by the gating loop[100]. A Ubc13 mutant (Ubc13$^{QD}$) with four amino acid substitutions stabilized the gating loop in the conformation found in other E2s lacks the NSC697923-binding pocket, rendering it resistant to NSC697923 inhibition in biochemical assays[100]. In our newly characterized structure (PDB: 9BIV), the active site loop exhibits a structural difference compared to known structures of Ubc13, whether ubiquitin is bound (2GMI)[55] or not (1J7D)[98]. The newly obtained structure may represent an intermediate state between the unbound and ubiquitin-bound forms. In the unbound form, a specific conformation of the gating loop is stabilized by internal hydrogen bonding while, in the ubiquitin-bound form, interaction with ubiquitin triggers conformational changes in the loop, allowing it to no longer obstruct the active site. The newly obtained structure exhibits characteristics of both states, indicating a transitional state. An analysis of structures of Ubc13 when bound to various E3 ligases or other regulatory factors such as OTUB1 and OspI reveals diverse conformations of the active site loop. This implies that the binding of these factors may impact Ubc13's catalytic activity by affecting the conformation of the active site loop. This suggests the active site loop of Ubc13 may be more flexible than previously believed, potentially adopting multiple conformational states.

### 5.2.2 Probing Ubc13 Active Site with Small Molecule Compounds

To explore the flexibility of the active site, we investigated two new compounds: NSC291068 and a 2-nitrofuran fragment. Both compounds exhibited a thermal shift upon incubation with Ubc13, indicating favorable interactions with the protein, and displayed a degree of specificity. NSC291068 specifically targeted the active site cysteine, and it hindered ubiquitin chain formation *in vitro*, but it also inhibited Ubc13$^{QD}$, suggesting an inhibitory mechanism not dependent on the binding groove in Ubc13. The 2-nitrofuran fragment showed specificity, only inhibiting ubiquitin chain synthesis for Ubc13$^{WT}$ but not Ubc13$^{QD}$. A structural overlay between 9BIV and 1J7D reveals the availability of N123 in the active site. This suggests 2-nitrofuran is likely capable of inhibiting both forms of Ubc13 as it can likely form the hydrogen bond in the active site. The compounds NSC627708, NSC646124, and NSC656835, lacking the 2-nitrofuran moiety, all demonstrated an inability to inhibit Ubc13, underscoring the crucial role of the 2-nitrofuran moiety in Ubc13 inhibition. An attractive mechanism of action is that 2-nitrofuran inhibits Ubc13 similarly to NSC697923, thereby inhibiting Ubc13 specifically, but through a non-covalent interaction. This discovery opens promising avenues for the development of non-covalent inhibitors targeting Ubc13 in future research. Interestingly, NSC291057, observed to react with reducing agents, was not observed to directly bind or inhibit Ubc13. This is perhaps due to the large size of the compound which is unable to be accommodated in the active site.

### 5.2.3 Next Generation Inhibitors of Ubc13

The motivation for developing small-molecule inhibitors targeting Ubc13 has been discussed previously in Chapter 2. Recent research has highlighted the involvement of this E2 enzyme in the pathogenesis of various cancers, as well as its role in the development of resistance to conventional chemotherapies[113,114-118,227,228]. Ubc13 possesses several characteristics making it an attractive target for inhibition[89]. Specifically, it is a unique enzyme in humans, meaning there are no alternative proteins capable of performing of its function of synthesizing K63-linked ubiquitin chains[89]. Ubc13 plays a significant role in two critical pathways, NF-κB signaling and the DNA damage response, both crucial for cancer cell proliferation and survival[58,67,91,95,96,139]. Another compelling aspect of Ubc13 as a target for chemical inhibition in cancer treatment is the discovery of its distinct active site[89,103,105,106]. As discussed in Chapter 2, traditional E2 enzymes are not typically amenable to drug targeting due to their lack of deep, complex active site pockets commonly targeted by small molecule inhibitors[89]. Moreover, a subset of human E2 enzymes (~17 out of approximately 34) share a high degree of structural similarity and amino acid conservation[137]. Therefore, our identification of a unique active site loop conformation and small binding pocket is essential for specifically inhibiting Ubc13, distinguishing it from other E2 enzymes[89,98,100]. Our work has characterized a covalent compound NSC291068 showing increased promiscuity compared to the previously characterized compound NSC697923[100]. While these results are promising, the high levels of reactivity of both compounds to sulfhydryl groups limit the therapeutic potential of these compounds. For next generation inhibitors, considering a more electron withdrawing or electron donating group would render the aromatic sulfone (aryl sulfone) group a poorer leaving group, making the compound less reactive[229]. We have also explored the possibility of finding a noncovalent inhibitor of Ubc13. We found that 2-nitrofuran could bind and inhibit Ubc13 catalytic activity. While this compound only bound with low affinity, it was shown to bind and specifically inhibit Ubc13 and not other E2s, providing some promise that it might offer a potential inhibitor lead. We hypothesize that 2-nitrofuran inhibits by binding the Ubc13 pocket in much the same way that the nitrofuran docks in the case of covalent inhibition, forming a hydrogen bond with N123. This would immobilize the loop in its inactive conformation to block Ubc13 catalytic activity. From a pharmaceutical standpoint, nitro groups are frequently less favored because of their metabolic instability and known toxicities[230]. Substitutions for the nitro group in the furan ring with alternative electron-withdrawing groups that are similar in size and structure might still maintain hydrogen bonding interactions with the Ubc13. It is expected that substituting the nitro group with less electron-withdrawing groups will reduce the compound's overall reactivity, potentially enhancing specificity.

### 5.2.4 Synthetic Lethality

An additional characteristic of Ubc13 which makes it a desirable target lies in the principal of synthetic lethality. Synthetic lethality refers to the phenomenon where the combination of two genetic alterations, neither of which alone is lethal to the organism, results in cell death or impaired viability[231]. Essentially, it describes a situation where the simultaneous presence of two specific mutations or perturbations leads to a phenotype that is lethal or unable to survive, while each individual alteration alone is viable[231]. This concept is particularly relevant in cancer research and drug development[231–234]. Recently, Ubc13 knockdown has been identified as synthetically lethal with the G-quadruplex (G4) ligand currently in clinical trials, CX-5461[235]. G-quadruplex (G4) structures, which form transiently at guanine-rich sequences in DNA and RNA across various organisms, play crucial roles in biological processes like gene regulation, telomere maintenance, and DNA replication[236–238]. During DNA replication and transcription, the presence of stable G-quadruplex (G4) structures can lead to replication stalling and transcriptional dysregulation[231]. Various factors are involved in resolving these stable structures to facilitate duplex refolding, replication, or transcription[239,240]. When factors responsible for G4 resolution are lacking, genome integrity can be compromised, leading to chromosomal deletions, rearrangements, or mutations[235,236,239,240]. In mammalian cells which lack homologous recombination (HR) factors BRCA1 and BRCA2, instability at G4 templates is observed in the presence of G4 ligands like CX-5461[235,241,242]. CX-5461 exhibits two significant activities: stabilizing G-quadruplex (G4) structures and inhibiting ribosomal RNA (rRNA) synthesis through RNA polymerase I[243–245]. Although these activities may not be mutually exclusive, CX-5461 has shown an initial human safety profile in advanced myelomas, non-Hodgkin lymphomas, and acute leukemias, as demonstrated in a phase I clinical trial[245]. Furthermore, preclinical studies have indicated promising antitumor effects against various cancer types, including MYCN-amplified neuroblastomas, ovarian, and prostate cancers[246–252]. Recent findings suggest that CX-5461's G4 stabilizing activity induces DNA damage, partly through replication stalling, which is synthetically lethal with loss of homologous recombination repair and independent of RNA polymerase I activity[235]. Specifically, knocking down Ubc13, and RNF168 sensitized cells to G4 stabilizers[235]. Additionally, the combination of CX-5461 with NSC697923 was shown to enhance toxicity in cancer cells[235]. Synthetic lethality provides a strategy for developing targeted cancer therapies[231–234]. By exploiting the genetic vulnerabilities of cancer cells, synthetic lethal interactions can be targeted to specifically kill cancer cells while sparing normal cells[231–234]. Furthermore, synthetic lethality offers a way to overcome drug resistance mechanisms observed in cancer treatment[253]. Exploring synthetic lethal pathways involving Ubc13 holds great promise for future research, especially as next-generation inhibitors continue to be developed.

## 5.3 Conclusions and Future Directions: Characterization of RNF138 DNA Binding Activity and Structure

### 5.3.1 RNF138 Preferentially Binds DNA Overhangs

This study focused on understanding the functional domains of RNF138 and their interactions with various DNA substrates. With a remarkable improvement in cost, and increased availability of the Expi293F expression system, we opted to express RNF138 in mammalian cells. Expressing RNF138 in Expi293F cells was an attempt to address the challenges encountered with RNF138 expressed in *E. coli*. Through Expi293F cells, we aimed to provide a more native environment conducive to proper protein folding and stability. This cellular system offers the presence of mammalian chaperones and facilitates key post-translational modifications, potentially resulting in enhanced stability and purity of RNF138[254,255]. Unfortunately, we initially encountered challenges in expressing sufficient levels of RNF138 after introduction into Expi293F cell lines. When RNF138 was expressed with a GFP and His tag, we were able to detect the protein at every step of purification with anti-GFP antibodies, but further optimization of the system is required as protein yields were too low to be useful. Despite facing challenges in purifying RNF138 in human cells, we were able to obtain sufficient quantities from *E. coli* expression for our studies by fine-tuning buffer conditions and meticulously controlling protein concentrations. Both RNF138$^{WT}$ and RNF138$^{1\text{-}225}$ exhibited DNA binding activity, with varying affinities, with RNF138$^{WT}$ exhibiting stronger binding interactions at lower concentrations compared to RNF138$^{1\text{-}225}$. The remaining truncations did not show DNA binding activity. This is likely due to the direct and indirect involvement of all domains within RNF138 to DNA binding affinity. A strong preference for overhang DNA was observed, with the size of the overhangs on both ends the single stranded and duplexed regions affecting binding affinities. This finding aligns with a previous study, which demonstrated that RNF138 exhibited stronger binding affinity to larger DNA molecules, particularly those with extended overhangs[129]. Perhaps this preference points to the recruitment of RNF138 post Mre-11 nuclease activity[27,129]. RNF138 contains a UIM, which are known to recognize and bind ubiquitin[256]. Previous cellular studies have indicated that the UIM plays a crucial role in DNA strand retention, underscoring the significance of substrate recognition in this process[129]. Interestingly, as higher concentrations of protein were titrated, a noticeable super-shifting of the bands on the gels was observed, indicating multiple protomers binding the DNA at higher concentrations[257].

### 5.3.2 Insights into the Structure of RNF138 Using AlphaFold2

An experimental structure of RNF138 has yet to be determined, and ongoing efforts to achieve this remain unsuccessful. The structural insights obtained from AlphaFold2[156–158] models provided valuable information about the flexibility of RNF138 and the organization of its ordered domains. Likely the overall flexibility of the protein leads to solubility issues. Flexibility becomes apparent upon closer examination of PAE statistical reports, which show highly localized areas of confidence interspersed with predicted regions of disorder. The interdomain interactions in RNF138 further emphasizes issues observed in our different constructs. The construct ZNF1, ZNF2, ZNF3 was unstable, perhaps because it lacks the RING domain that is required to interact with ZNF1 for proper folding. ZNF2 and ZNF3, like the N-terminal domain, were confidently predicted, and our study's construct exhibited stability, implying potential interplay in the folding of the ZNFs. Surprisingly, this construct did not show DNA binding, likely due to the requirement of other elements within RNF138. It is possible that elements from the second linker (L2), which connects the N-terminal domains to the C-terminal domains (ZNF2 and ZNF3), directly participate in DNA binding. These elements may rely on the presence of other domains for structural support and proper folding. AlphaFold2 modeling of the N-terminal RING and ZNF1 domains strongly suggests their involvement in E3 catalytic activity, as evidenced by the robust interdomain interactions observed, akin to those seen in RNF125 and TRAF6[132,134,153]. Additionally, AlphaFold2 multimer[154,157] confidently predicted that the RING, ZNF1, and elements from L2 interact with the E2 and ubiquitin. The crystal structure of RNF125 with E2, combined with experimental data from techniques like nuclear magnetic resonance, indicate that RNF125 functions as a monomer[134,153]. Additionally, it is observed that elements beyond the RING domain play a role in stabilizing and activating the ubiquitin-E2 complex[134,153]. This suggests that these elements outside of the RING domain likely function independently of dimerization. AlphaFold2 modeling of RNF138 strongly predicted a monomeric structure, and the results obtained from size exclusion chromatography during protein purification indicates that RNF138 does not form dimers in solution. An interesting observation is the placement of ubiquitin bound to the UIM in RNF138 in one of the models **(Figure 5.1)**. UIMs are short helical structures, involved in ubiquitin recognition[256]. Typical UIM interactions occur between hydrophobic residues in the UIM helix and hydrophobic residues in the ubiquitin hydrophobic patch[256]. Further interactions are supplemented with a C-terminal positively charged patch on ubiquitin and N-terminal acidic patch in UIMs[256].

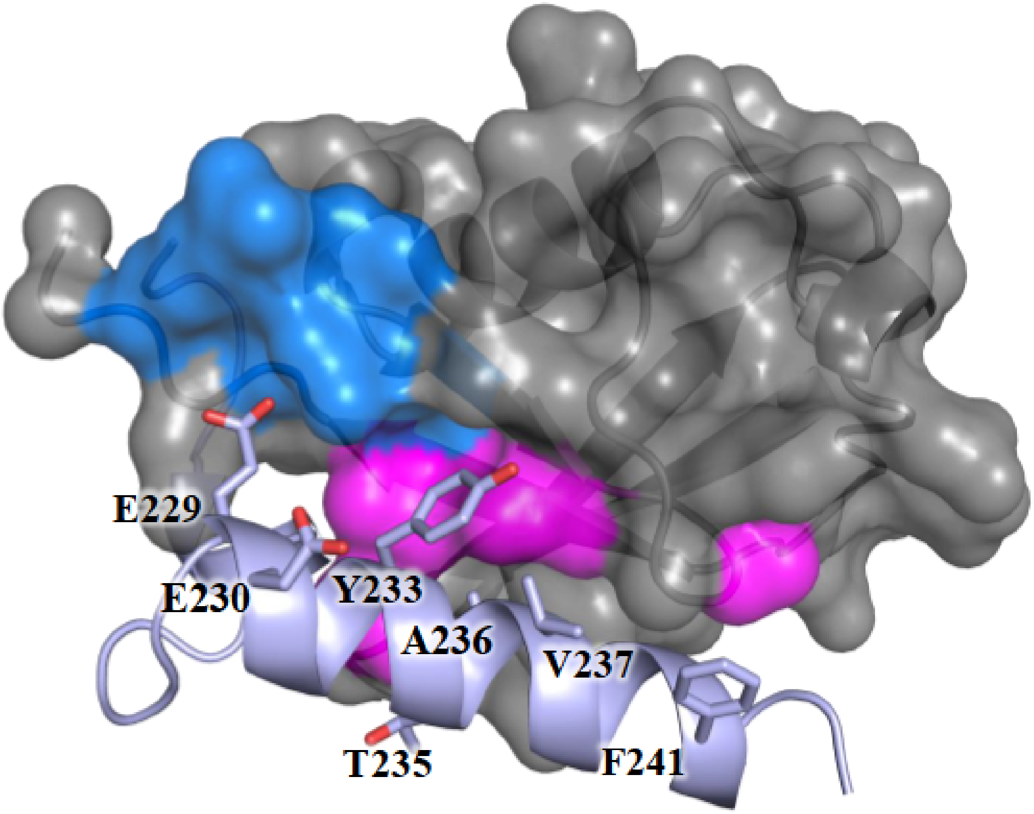


**Figure 5.1. RNF138 UIM in Complex with Ubiquitin.**

The UIM helix (blue) complex with ubiquitin (surface representation grey). Purple surface represents the hydrophobic patch on ubiquitin. The blue surface represents a positively charged C-terminal patch on ubiquitin. A canonical binding interface of Y233 T235, A236, V237, and F241 of RNF138's UIM are predicted to complex with ubiquitin. Additional interactions are typically observed in all UIMs N-terminal to the hydrophobic residues, with a conserved motif (DEEE) which form electrostatic interactions with the C-terminal region of ubiquitin. E229, and E230 are shown complexed[256].

### 5.3.3 A model of RNF138 Activity on DNA Overhangs.

Based on our study, we can speculate and propose a model for the mechanism of how RNF138 activates ubiquitylation in response to DNA strand break binding **(Figure 5.2)**. The N-terminal region is involved in catalysis. Specifically, the RING domain, ZNF1, and several residues from L2, bind and promote ubiquitylation of substrate with the E2~ubiquitin complex. The remainder of L2, ZNF2 and ZNF3 are involved in DNA binding and recruitment to DNA overhangs. The DNA was modelled using zinc finger 1 of Poly [ADP-ribose] polymerase 1 (PARP1) bound to DNA, which was overlayed over zinc finger 2 of RNF138. All the domains are involved in overall structural stabilization of the linker, therefore are necessary for stable DNA binding. Finally, the UIM is involved in recognizing ubiquitylated substrates, and is involved in substrate recognition.

Ultimately, the models generated by AlphaFold2 require further experimental validation. Challenges such as the inability to model post-translational modifications and mutations, which can significantly impact protein function, pose difficulties in elucidating the mechanism[155,156]. Future studies should aim to solve the experimental structure of RNF138 aided by AlphaFold2 for construct design. This may reveal insights into the function of the independent domains. Furthermore, co-crystallization trials with the preferred DNA substrates could be significant in revealing DNA binding activity of RNF138. Strategies to address cooperative binding, may require solution-based analysis of RNF138. A powerful approach to investigate the mechanism of RNF138 DNA binding is Small Angle X-Ray Scattering (SAXS)[258,259]. Despite its low resolution, SAXS is a technique used to visualize proteins in solution, providing dynamic insights into protein behavior rather than static snapshots[258,259]. AlphaFold2 models serve as valuable tools for molecular replacement in phasing novel structures or can be employed for SAXS envelope fitting in the absence of experimental structures[260–263]. All models can then be tested by structural guided mutagenesis, via *in vitro* studies such as EMSAs, mutating key residues identified in the models to see the outcome. Additionally, *in vivo* studies can be carried out for model verification. This strategy is not limited for DNA binding activity but can also be extended to characterizing RNF138's E3 ligase activity.

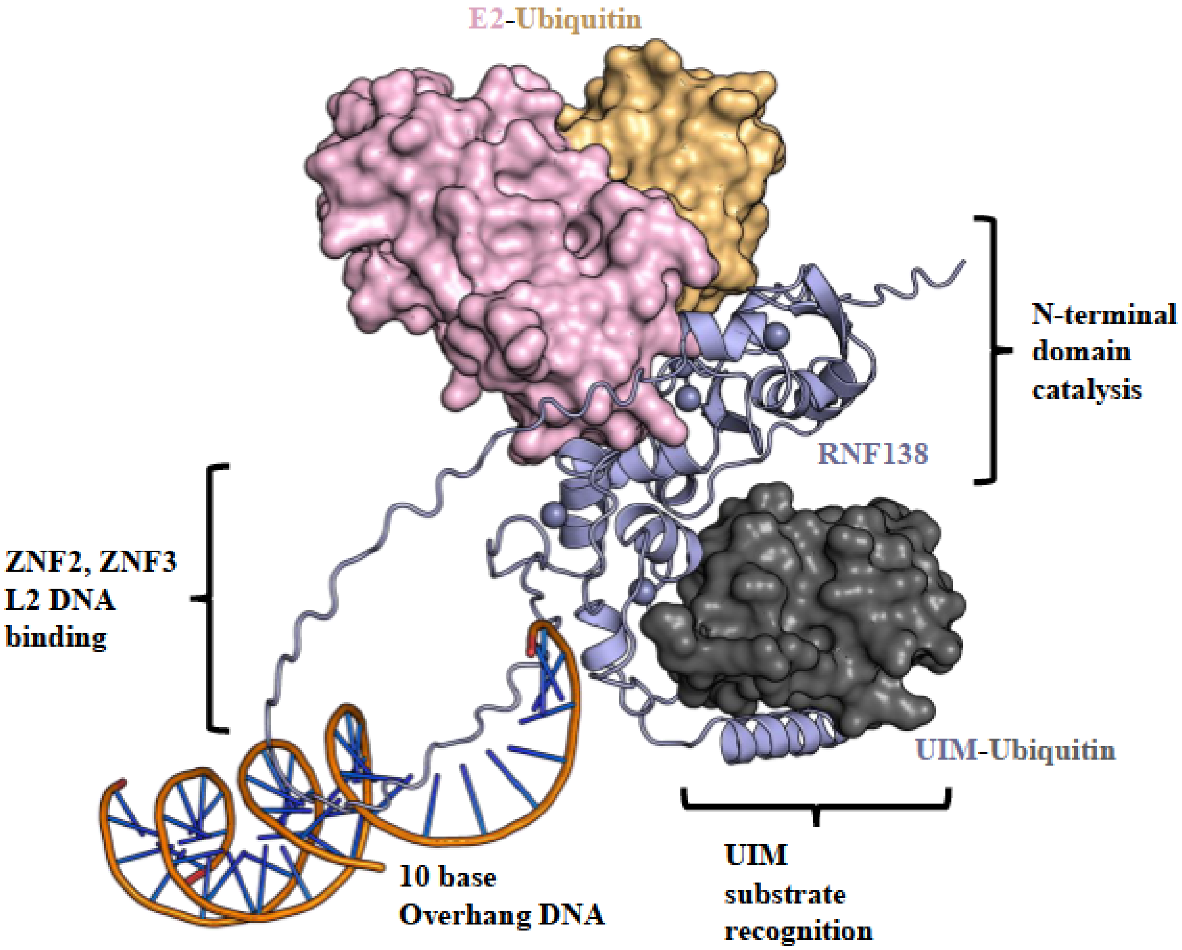


**Figure 5.2. A model of RNF138 Activity on DNA Overhangs.**

The RNF138 model (blue) was constructed based on the modeled structure of RNF138 with E2-Ubiquitin (pink and orange) and aligned to show ubiquitin bound to UIM in the second-ranked model. DNA was modeled using DNA-bound PARP1 zinc fingers (PDB: 4DQY), with the zinc finger portion aligned to ZNF2 of RNF138. However, caution is advised in interpreting the position of the linker, as it is disordered and likely flexible. AlphaFold2 cannot model disordered regions, so this model remains speculative and requires validation.

## 5.4 Investigating the Regulation of RNF138

### 5.4.1 RNF138 is Regulated by Ubiquitylation and Phosphorylation

We proceeded with our exploration of RNF138, focusing on specific post-translational modification (PTM) sites that appeared to be critical for its regulation. We observe a regulatory role for the E3 ubiquitin ligase RNF138 in end resection, controlled by both ubiquitylation and CDK-dependent phosphorylation events. Interestingly, RNF138 ubiquitylation levels fluctuate during different cell cycle phases, being elevated during S and G2 phases but decreased under prolonged genotoxic stress. This suggests a potential role as a switch, to prime and restrict RNF138 activity. Phosphorylation also plays a crucial role in regulating RNF138 activity in end resection and HR progression. Phospho-ablation mutations at T27 and S124 sites impair HR-dependent gene conversion, while phospho-mimicking substitutions at these positions rescue HR activity beyond wildtype levels. These findings provide deeper insights into the molecular signals and events governing DNA end resection. In future investigations of the RNF138 study, the focus should initially be on understanding the ubiquitylation process of RNF138. We hypothesize that increased ubiquitylation during the S/G2 phases activates RNF138 rather than marking it for proteasomal degradation. To support this hypothesis, we plan to identify the types of ubiquitin chain linkages formed on RNF138. One approach involves performing GFP immunoprecipitations on HEK293 cells co-expressing GFP-RNF138 and HA-Ub, followed by probing with antibodies specific to different ubiquitin chain linkages. Alternatively, deubiquitinases (DUBs) can be used to determine ubiquitin chain linkages[264]. DUBs are enzymes that cleave ubiquitin molecules from proteins or disassemble polyubiquitin chains[184,264,265]. By selectively targeting specific ubiquitin linkages, DUBs can reveal the types of ubiquitin chains present on a protein or within a cellular environment[184,264,265]. Ubiquitylated RNF138 can be immunoprecipitated and subject to a treatment with a library of DUBs to determine specific ubiquitin linkages[264].

Another area of exploration involves identifying the E3 ubiquitin ligase responsible for adding ubiquitin molecules to RNF138. Since RNF138 itself is an E3 ligase, it is worth investigating whether the ubiquitylation of the protein occurs through autocatalytic activity. To address this question, we could disrupt the function of the RING domain by mutating specific residues that were identified to play a role in catalysis, for example Met98, while minimizing structural instability, and then assess whether ubiquitin signals are still present. If autocatalysis is not responsible for RNF138 ubiquitylation, we could explore other potential E3 ligases for RNF138 by depleting them using siRNA and examining RNF138 ubiquitylation once again. One potential candidate E3 ligase is A20, known for its role in inhibiting NFκB signaling[36,124]. In certain B cell malignancies, A20's E3 activity is involved in ubiquitylating RNF138, thereby suppressing the oncogenic activity of L265P-mutated MyD88[124]. Interestingly, A20 has been found to negatively regulate DSB signaling independently of its catalytic activity by inhibiting H2A ubiquitylation and the accumulation of RNF168 and 53BP1 through direct binding to RNF168. We have previously attempted to analyze RNF138 ubiquitylation activity *in vitro* by using Ubc13 as the E2 enzyme in the reaction. RNF138 seems to only build ubiquitin chains at elevated concentrations, compared to RNF8, a well characterized E3 ligase involved in synthesizing K63-linked ubiquitin chains. Moving forward, our research can extend this *in vitro* assay by including both known substrates of RNF138 and by varying the E2 conjugating enzymes used. Such approaches can contribute to a comprehensive understanding of its catalytic activity, potentially shedding light on its autocatalytic properties. Such approaches can contribute to a comprehensive understanding of its catalytic activity, potentially shedding light on its autocatalytic properties as well.

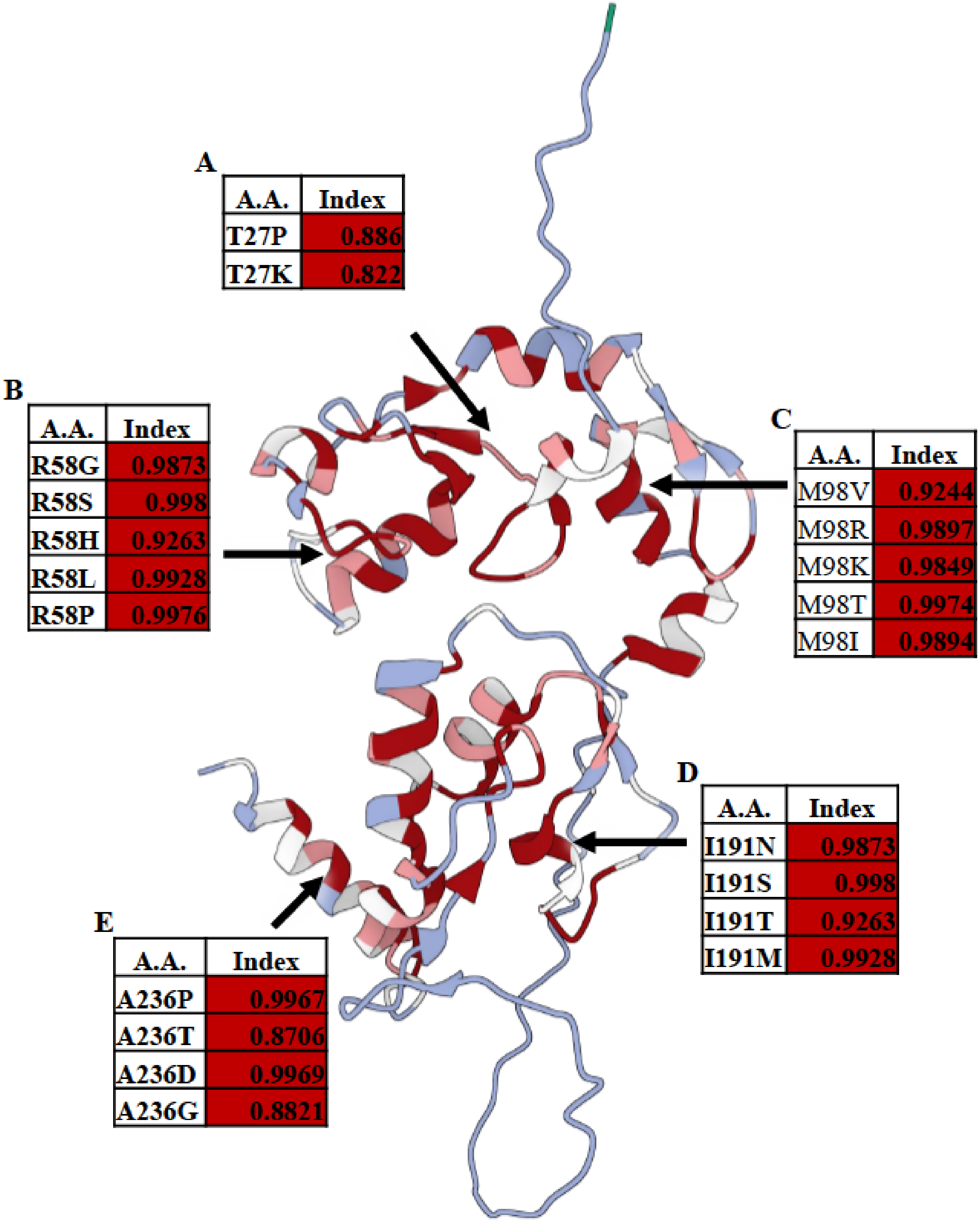
### 5.4.2 Cell Cycle Dependent Regulation of RNF138

Our research demonstrates that the ubiquitylation and CDK-mediated phosphorylation of RNF138 vary with the cell cycle. We observed increases in both ubiquitylation and TP motif phosphorylation during the S/G2 phases. We explored potential interactions between these PTMs in RNF138. Our observations revealed that the K158R substitution did not alter the p-TP and p-S/TQ signals on RNF138. This indicates that K158 is not essential for phosphorylation events at T27 and S124. Conversely, the S124A substitution led to a partial reduction in the p-TP signal, suggesting that S124 phosphorylation may contribute to, but is not indispensable for, phosphorylation at T27. However, the T27A mutation significantly diminished the p-S/TQ signal. Additionally, the basal levels of higher-order ubiquitylation on GFP-RNF138 were notably decreased by the T27A and S124A mutations, masking the expected reduction in RNF138 ubiquitylation induced by CPT. Consequently, T27 phosphorylation may be a prerequisite for S124 phosphorylation, and both functional T27 and S124 sites appear crucial for RNF138 ubiquitylation. Considering this interconnectedness, we sought to investigate the impact of expressing the T27A-S124A (TS), T27A-K158R (TK), and S124A-K158R (SK) double mutants, as well as the T27A-S124A-K158R (TSK) triple mutant, on the cell's ability to engage in HR. Expression of both double and triple mutants led to similarly low HR frequencies. These frequencies were comparable to those observed for the T27A, S124A, and K158R single mutants, which were initially low and akin to HR frequencies in cells lacking endogenous RNF138. The depletion of any PTM site, or any combination thereof, was adequate to reduce HR to baseline levels, underscoring the significance of all three sites in RNF138 function and indicating their potential involvement in a shared pathway. Interestingly, K158 is not the only ubiquitylation site detected on RNF138 as K26 and K41 were also detected. AlphaFold2 modelling reveals key insights into the posttranslational modification sites[155–157]. S124 and K158 are both exposed and fall in L2. T27 is within the RING domain. The AlphaFold model indicates that while T27 is exposed to solvent, the proline and arginine residues of the CDK phosphorylation motif are buried, potentially obstructing recognition by CDKs. For CDKs to phosphorylate T27, the RING domain would likely need to be unfolded. It is possible that ubiquitylation at K26 and/or K41 facilitates this process by inducing conformational changes that unfold the RING domain and allow full accessibility of RNF138's phosphorylation motif to CDK activity. Therefore, further investigation into the roles of K26 and/or K41 in RNF138 ubiquitylation, and whether T27 phosphorylation relies on the integrity of one or both lysine sites is warranted.

Ultimately, our findings suggest that phosphorylation events at T27 and S124, along with ubiquitylation at K158, collectively enhance RNF138 activity in end resection and HR processes. The question remains how, and as previously mentioned, *in vitro* characterization of RNF138 is imperative to answer the questions. For example, to test the effects of the PTMs *in vitro,* phosphorylated RNF138 can be used in *in vitro* assays to determine if E3 ligase activity is increased. Furthermore, DNA binding affinity studies could be conducted with phosphorylated or ubiquitylated RNF138 to determine the affects on DNA binding affinity. AlphaFold2 is limited in that is unable to model post translational modifications, therefore an experimental structure of RNF138 with posttranslational modifications may reveal insights into the overall mechanism of action. In summary, our investigation into RNF138 regulation not only sheds light on its intricate control mechanisms but raises a plethora of compelling questions, opening exciting avenues for further exploration into the precise regulation of RNF138 function.

### 5.4.3 Therapeutic Implications of RNF138

The therapeutic significance of RNF138 is underscored by its involvement in various cancer-related processes[125,127,128,135]. It is notably overexpressed in glioma cells, a particularly aggressive form of brain cancer, where it promotes proliferation, metastasis, and resistance to radiation therapy[135]. Furthermore, RNF138 has been implicated in conferring resistance to cisplatin in gastric cancer cells[128]. Changes in individual amino acids within proteins can sometimes cause minimal impact, but they often disrupt protein folding, activity, or stability[266–268]. While only a small portion of these variants has been studied experimentally (4 million missense variants, 2% clinically characterized as benign or pathogenic)[269], there is a vast amount of biological sequence data available that can be used for machine learning purposes[266–269]. AlphaMissense is a computational tool designed to assess the potential harmfulness of missense variants[269]. It uses a refined version of AlphaFold, which has been specifically adapted and optimized using data on missense variants found in human and primate populations[269]. By training on this variant population data and incorporating information about the frequency of these variants, AlphaMissense learns to predict whether a given missense variant is likely to cause harm or not[269,270]. For all 71 million possible single missense variants in the human genome, it provides a probability score indicating the likelihood of the variant being "likely benign" if the variant is probably harmless, "likely pathogenic" if it is likely to cause harm, or "uncertain" if the prediction is unclear[269,270]. AlphaMissense has predicted 511 possible missense variants to be likely pathogenic in RNF138. 20 likely pathogenic predictions fall on residues characterized in our studies **(Figure 5.3)**. Mutations occur at several key sites within RNF138: T27, a crucial phosphorylation site; R58, essential for interacting with the E2 and ubiquitin; M98, a residue located within the core of hydrophobic interactions between the RING domain and ZNF1; I191, key in the hydrophobic packing between ZNF2 and ZNF3; and finally, A236, which interacts with the ubiquitin hydrophobic patch within the UIM. Future characterization of RNF138 can use this tool as a starting point in characterizing these variants, to try to understand the underlying result of the mutations on RNF138 function. Understanding the effects of these mutations in patient tumors could help identify whether tumor cells are proficient or deficient in homologous recombination (HR), thus determining their sensitivity to treatment. This in turn can aid in future development of therapeutics, and personalized treatments for specific tumour profiles in patients.



**A**

| A.A. | Index |
|---|---|
| T27P | 0.886 |
| T27K | 0.822 |

**B**

| A.A. | Index |
|---|---|
| R58G | 0.9873 |
| R58S | 0.998 |
| R58H | 0.9263 |
| R58L | 0.9928 |
| R58P | 0.9976 |

**C**

| A.A. | Index |
|---|---|
| M98V | 0.9244 |
| M98R | 0.9897 |
| M98K | 0.9849 |
| M98T | 0.9974 |
| M98I | 0.9894 |

**D**

| A.A. | Index |
|---|---|
| I191N | 0.9873 |
| I191S | 0.998 |
| I191T | 0.9263 |
| I191M | 0.9928 |

**E**

| A.A. | Index |
|---|---|
| A236P | 0.9967 |
| A236T | 0.8706 |
| A236D | 0.9969 |
| A236G | 0.8821 |

**Figure 5.3. AlphaMissense Model of RNF138 Annotated with Predicted Pathogenic Mutations**

AlphaMissense model of RNF138 colored according to the likelihood of pathogenesis at every amino acid (https://alphamissense.hegelab.org/). Bening variants are scored 0.1-0.2 (blue). Ambiguous variants are typically scored 0.2-0.7 (white). Likely pathogenic variants are valued greater than 0.8 (red). These values are derived from various features extracted from the variant sequence, including its amino acid composition, predicted structural context, and evolutionary conservation. They are benchmarked against ClinVar, a database of known pathogenic and benign variants. 20 likely pathogenic missense variants are shown of 5 key amino acids characterized in our studies with an arrow pointing to their position in the model. **A.** T27. **B.** R58. **C.** M98. **D.** I191 **E.** A236.

# References

(1) Hakem, R. (2008) DNA-damage repair; the good, the bad, and the ugly. *EMBO J. 27*, 589–605.

(2) Hoeijmakers, J. H. J. (2009) DNA Damage, Aging, and Cancer. *N. Engl. J. Med. 361*, 1475–1485.

(3) Tiwari, V., and Wilson, D. M. (2019) DNA Damage and Associated DNA Repair Defects in Disease and Premature Aging. *Am. J. Hum. Genet. 105*, 237–257.

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

(5) De Bont, R., and van Larebeke, N. (2004) Endogenous DNA damage in humans: A review of quantitative data. *Mutagenesis 19*, 169–185.

(6) Predonzani, A. (2015) Spotlights on immunological effects of reactive nitrogen species: When inflammation says nitric oxide. *World J. Exp. Med. 5*, 64.

(7) Hang, B. (2010) Formation and Repair of Tobacco Carcinogen-Derived Bulky DNA Adducts. *J. Nucleic Acids* (Basu, A., Ed.) *2010*, 709521.

(8) Douki, T., Koschembahr, A. Von, and Cadet, J. (2017) Invited Review Insight in DNA Repair of UV-induced Pyrimidine Dimers by Chromatographic Methods †. *Photochem. Photobiol. 93*, 207–215.

(9) Li, G.-M. (2008) Mechanisms and functions of DNA mismatch repair. *Cell Res. 18*, 85–98.

(10) Ceccaldi, R., Sarangi, P., and D'Andrea, A. D. (2016) The Fanconi anaemia pathway: new players and new functions. *Nat. Rev. Mol. Cell Biol. 17*, 337–349.

(11) Makridakis, N. M., and Reichardt, J. K. V. (2012) Translesion DNA polymerases and cancer. *Front. Genet. 3*, 174.

(12) Pfeiffer, P., Goedecke, W., and Obe, G. (2000) Mechanisms of DNA double-strand break repair and their potential to induce chromosomal aberrations. *Mutagenesis 15*, 289–302.

(13) Chapman, J. R., Taylor, M. R. G., and Boulton, S. J. (2012) Review Playing the End Game : DNA Double-Strand Break Repair Pathway Choice. *Mol. Cell 47*, 497–510.

(14) Mckinnon, P. J. (2009) DNA Repair Deficiency and Neurological Disease Peter. *Nat. Rev. Neurosci. 10*, 100–112.

(15) Mahaney, B. L., Meek, K., and Lees-miller, S. P. (2009) Repair of ionizing radiation-induced DNA double strand breaks by non-homologous end-joining. *Biochem. J. 417*, 639–650.

(16) Gómez-herreros, F. (2019) DNA Double Strand Breaks and Chromosomal Translocations Induced by DNA Topoisomerase II. *Front. Mol. Biosci. 6*, 141.

(17) Staker, B. L., Hjerrild, K., Feese, M. D., Behnke, C. A., Burgin, A. B., and Stewart, L. (2002) The mechanism of topoisomerase I poisoning by a camptothecin analog. *Proc. Natl. Acad. Sci. 99*, 15387–15392.

(18) Srinivas, U. S., Tan, B. W. Q., and Vellayappan, B. A. (2019) Redox Biology ROS and the DNA damage response in cancer. *Redox Biol. 25*, 2213–2317.

(19) Symington, L. S., and Gautier, J. (2011) Double-strand break end resection and repair pathway choice. *Annu. Rev. Genet. 45*, 247–271.

(20) Dueva, R., and Iliakis, G. (2013) Alternative pathways of non-homologous end joining (NHEJ) in genomic instability and cancer. *Transl. Cancer Res. 2*, 163–177.

(21) Goodarzi, A. A., and Jeggo, P. A. (2013) The Repair and Signaling Responses to DNA Double-Strand Breaks. *Adv. Genet.* 1st ed. Elsevier Inc.

(22) Stinson, B. M., and Loparo, J. J. (2021) Repair of DNA Double-Strand Breaks by the Nonhomologous End Joining Pathway. *Annu. Rev. Biochem. 90*, 137–164.

(23) Blier, P. R., Griffith, A. J., Craft, J., and Hardin, J. A. (1993) Binding of Ku protein to DNA: Measurement of affinity for ends and demonstration of binding to nicks. *J. Biol. Chem. 268*, 7594–7601.

(24) Lee, K. J., Saha, J., Sun, J., Fattah, K. R., Wang, S. C., Jakob, B., Chi, L., Wang, S. Y., Taucher-Scholz, G., Davis, A. J., and Chen, D. J. (2015) Phosphorylation of Ku dictates DNA double-strand break (DSB) repair pathway choice in S phase. *Nucleic Acids Res. 44*, 1732–1745.

(25) Wu, D., Topper, L. M., and Wilson, T. E. (2008) Recruitment and Dissociation of Nonhomologous End Joining Proteins at a DNA Double-Strand Break in Saccharomyces cerevisiae. *Genetics 1249*, 1237–1249.

(26) Dna, R. I., Repair, B., Myler, L. R., Gallardo, I. F., Soniat, M. M., Kim, Y., Paull, T. T., Finkelstein, I. J., Myler, L. R., Gallardo, I. F., Soniat, M. M., Deshpande, R. A., and Gonzalez, X. B. (2017) Short Article Single-Molecule Imaging Reveals How Mre11- Short Article Single-Molecule Imaging Reveals How Mre11-Rad50-Nbs1 Initiates DNA Break Repair. *Mol. Cell 67*, 891–898.

(27) Chanut, P., Britton, S., Coates, J., Jackson, S. P., and Calsou, P. (2016) Coordinated nuclease activities counteract Ku at single-ended DNA double-strand breaks. *Nat. Commun. 7*, 12889.

(28) Lieber, M. R. (2008) The Mechanism of Human Nonhomologous DNA End Joining. *J. Biol. Chem. 283*, 1–5.

(29) Aceytuno, R. D., Piett, C. G., Havali-shahriari, Z., Edwards, R. A., Rey, M., Ye, R., Javed, F., Fang, S., Mani, R., Weinfeld, M., Hammel, M., Tainer, J. A., Schriemer, D. C., Lees-miller, S. P., and Glover, J. N. M. (2017) Structural and functional characterization of the PNKP – XRCC4 – LigIV DNA repair complex. *Nucleic Acids Res. 45*, 6238–6251.

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

(31) Ray, S., Breuer, G., Deveaux, M., Zelterman, D., Bindra, R., and Sweasy, J. B. (2018) DNA polymerase beta participates in DNA End-joining. *Nucleic Acids Res. 46*, 242–255.

(32) Schwertman, P., Bekker-Jensen, S., and Mailand, N. (2016) Regulation of DNA double-strand

break repair by ubiquitin and ubiquitin-like modifiers. *Nat. Rev. Mol. Cell Biol. 17*, 379–394.

(33) Polo, S. E., and Jackson, S. P. (2011) Dynamics of DNA damage response proteins at DNA breaks: A focus on protein modifications. *Genes Dev. 25*, 409–433.

(34) Bergink, S., and Jentsch, S. (2009) Principles of ubiquitin and SUMO modifications in DNA repair. *Nature 458*, 461–467.

(35) Yard, B. D., Reilly, N. M., Bedenbaugh, M. K., and Pittman, D. L. (2016) RNF138 interacts with RAD51D and is required for DNA interstrand crosslink repair and maintaining chromosome integrity. *DNA Repair (Amst). 42*, 82–93.

(36) Wertz, I. E., Rourke, K. M. O., Zhou, H., Eby, M., Aravind, L., Seshagiri, S., Wu, P., Wiesmann, C., and Dixit, V. M. (2004) ligase domains of A20 downregulate NF- k B signalling. *Nature 430*, 694–699.

(37) Ubersax, J. A., and Jr, J. E. F. (2007) Mechanisms of specificity in protein phosphorylation. *Nat. Rev. Mol. Cell Biol. 8*, 530–542.

(38) Blackford, A. N., and Jackson, S. P. (2017) ATM , ATR , and DNA-PK : The Trinity at the Heart of the DNA Damage Response. *Mol. Cell 66*, 801–817.

(39) Yue, X., Bai, C., Xie, D., Ma, T., and Zhou, P. (2020) DNA-PKcs : A Multi-Faceted Player in DNA Damage Response. *Front. Genet. 11*, 607428.

(40) Smith, G. C. M., and Jackson, S. P. (2010) The PIKK Family of Protein Kinases. *Handb. Glycomics* Second Edi. Elsevier Inc.

(41) Polo, S. E., Kaidi, A., Baskcomb, L., Galanty, Y., and Jackson, S. P. (2010) Regulation of DNA-damage responses and cell-cycle progression by the chromatin remodelling factor CHD4. *EMBO J. 29*, 3130–3139.

(42) Zou, L., and Elledge, S. J. (2003) Sensing DNA Damage Through ATRIP Recognition of RPA-ssDNA Complexes. *Science (80-. ). 300*, 1542–1548.

(43) Malumbres, M. (2014) Cyclin-dependent kinases. *Genome Biol. 15*, 122.

(44) Matthews, H. K., Bertoli, C., and Bruin, R. A. M. (2022) Cell cycle control in cancer. *Nat. Rev. Mol. Cell Biol. 23*, 74–88.

(45) Aylon, Y., Liefshitz, B., and Kupiec, M. (2004) The CDK regulates repair of double-strand breaks by homologous recombination during the cell cycle. *EMBO J. 23*, 4868–4875.

(46) Fernandez-capetillo, O., Lee, A., Nussenzweig, M., and Nussenzweig, A. (2004) H2AX : the histone guardian of the genome. *DNA Repair (Amst). 3*, 959–967.

(47) Lukas, J., Lukas, C., and Bartek, J. (2004) Mammalian cell cycle checkpoints : signalling pathways and their organization in space and time. *DNA Repair (Amst). 3*, 997–1007.

(48) Doil, C., Mailand, N., Bekker-jensen, S., Menard, P., Larsen, D. H., Pepperkok, R., Ellenberg, J., Panier, S., Durocher, D., Bartek, J., Lukas, J., and Lukas, C. (2009) RNF168 Binds and Amplifies Ubiquitin Conjugates on Damaged Chromosomes to Allow Accumulation of Repair Proteins. *Cell 136*, 435–446.

(49) Williams, R. S., Williams, J. S., and Tainer, J. A. (2007) Mre11–Rad50–Nbs1 is a keystone complex connecting DNA repair machinery, double-strand break signaling, and the chromatin template. *Biochem. Cell Biol. 85*, 509–520.

(50) Spycher, C., Miller, E. S., Townsend, K., Pavic, L., Morrice, N. A., Janscak, P., Stewart, G. S., and Stucki, M. (2008) Constitutive phosphorylation of MDC1 physically links the MRE11 – RAD50 – NBS1 complex to damaged chromatin. *J. Cell Biol. 181*, 227–240.

(51) Lee, M. S., Edwards, R. A., Thede, G. L., and Glover, J. N. M. (2005) Structure of the BRCT Repeat Domain of MDC1 and Its Specificity for the Free COOH- terminal End of the ▨ -H2AX Histone Tail . *J. Biol. Chem. 280*, 32053–32056.

(52) Stucki, M., Clapperton, J. A., Mohammad, D., Yaffe, M. B., Smerdon, S. J., and Jackson, S. P. (2005) MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks. *Cell 123*, 1213–1226.

(53) Melander, F., Bekker-jensen, S., Falck, J., Bartek, J., Mailand, N., and Lukas, J. (2008) Phosphorylation of SDT repeats in the MDC1 N terminus triggers retention of NBS1 at the DNA damage – modified chromatin. *J. Cell Biol. 181*, 213–226.

(54) Huen, M. S. Y., Grant, R., Manke, I., Minn, K., Yu, X., Yaffe, M. B., and Chen, J. (2007) RNF8 Transduces the DNA-Damage Signal via Histone Ubiquitylation and Checkpoint Protein Assembly. *Cell 131*, 901–914.

(55) Eddins, M. J., Carlile, C. M., Gomez, K. M., Pickart, C. M., and Wolberger, C. (2006) Mms2 – Ubc13 covalently bound to ubiquitin reveals the structural basis of linkage-specific polyubiquitin chain formation. *Nat. Struct. Mol. Biol. 13*, 915–920.

(56) Campbell, S. J., Edwards, R. A., Leung, C. C. Y., Neculai, D., Hodge, C. D., Dhe-paganon, S., and Glover, J. N. M. (2012) Molecular Insights into the Function of RING Finger ( RNF ) - containing Proteins hRNF8 and hRNF168 in Ubc13/Mms2-dependent Ubiquitylation . *J. Biol. Chem. 287*, 23900–23910.

(57) Kolas, N. K., Chapman, J. R., Nakada, S., Ylanko, J., Chahwan, R., Sweeney, F. D., Panier, S., Mendez, M., Wildenhain, J., Thomson, T. M., Pelletier, L., Jackson, S. P., and Durocher, D. (2007) Orchestration of the DNA-Damage Response by the RNF8 Ubiquitin Ligase. *Science (80-. ). 318*, 1637–1640.

(58) Wang, B., and Elledge, S. J. (2007) Ubc13 / Rnf8 ubiquitin ligases control foci formation of the Rap80 / Abraxas / Brca1 / Brcc36 complex in response to DNA damage. *Proc. Natl. Acad. Sci. 104*, 20759–20763.

(59) Mattiroli, F., Vissers, J. H. A., Dijk, W. J. Van, Ikpa, P., Citterio, E., Vermeulen, W., Marteijn, J. A., and Sixma, T. K. (2012) RNF168 Ubiquitinates K13-15 on H2A / H2AX to Drive DNA Damage Signaling. *Cell 150*, 1182–1195.

(60) Stewart, G. S., Panier, S., Townsend, K., Al-hakim, A. K., Kolas, N. K., Miller, E. S., Nakada, S., Ylanko, J., Olivarius, S., Mendez, M., Oldreive, C., Wildenhain, J., Tagliaferro, A., Pelletier, L., Taubenheim, N., Durandy, A., Byrd, P. J., Stankovic, T., Taylor, A. M. R., and Durocher, D. (2009) The RIDDLE Syndrome Protein Mediates a Ubiquitin-Dependent Signaling Cascade at Sites of DNA Damage. *Cell 136*, 420–434.

(61) Lys, H. A., Orthwein, A., Leung, C. C. Y., Huang, H., Noordermeer, S. M., Sicheri, F., Durocher, D., Landry, M., and Kitevski-leblanc, J. (2013) 53BP1 is a reader of the DNA-damage-. *Nature 499*, 50–56.

(62) Xie, A., Hartlerode, A., Stucki, M., Odate, S., Puget, N., Kwok, A., Nagaraju, G., Yan, C., Alt, F. W., Chen, J., and Jackson, S. P. (2007) Article Distinct Roles of Chromatin-Associated Proteins MDC1 and 53BP1 in Mammalian Double-Strand Break Repair. *Mol. Cell 28*, 1045–1057.

(63) Mailand, N., Bekker-jensen, S., Faustrup, H., Melander, F., Bartek, J., and Lukas, C. (2007) RNF8 Ubiquitylates Histones at DNA Double-Strand Breaks and Promotes Assembly of Repair Proteins. *Cell 131*, 887–900.

(64) Sato, Y., Yoshikawa, A., Mimura, H., Yamashita, M., Yamagata, A., and Fukai, S. (2009) Structural basis for specific recognition of Lys 63-linked polyubiquitin chains by tandem UIMs of RAP80. *EMBO J. 28*, 2461–2468.

(65) Pickart, C. M., and Eddins, M. J. (2004) Ubiquitin : structures , functions , mechanisms. *Biochim. Biophys. Acta 1695*, 55–72.

(66) Komander, D., and Rape, M. (2012) The Ubiquitin Code. *Annu. Rev. Biochem. 81*, 203–229.

(67) Al-hakim, A., Escribano-diaz, C., Landry, M., Donnell, L. O., Panier, S., Szilard, R. K., and Durocher, D. (2010) The ubiquitous role of ubiquitin in the DNA damage response. *DNA Repair (Amst). 9*, 1229–1240.

(68) George, A. J., Hoffiz, Y. C., Charles, A. J., Zhu, Y., and Mabb, A. M. (2018) A Comprehensive Atlas of E3 Ubiquitin Ligase Mutations in Neurological Disorders. *Front. Genet. 9*, 29.

(69) Zheng, N., and Shabek, N. (2017) Ubiquitin Ligases : Structure , Function , and Regulation. *Annu. Rev. Biochem. 86*, 129–157.

(70) Swatek, K. N., and Komander, D. (2016) Ubiquitin modifications. *Cell Res. 26*, 399–422.

(71) Dikic, I., and Schulman, B. A. (2023) An expanded lexicon for the ubiquitin code. *Nat. Rev. Mol. Cell Biol. 24*, 273–287.

(72) Chen, Z. J., and Sun, L. J. (2009) Review Nonproteolytic Functions of Ubiquitin in Cell Signaling. *Mol. Cell 33*, 275–286.

(73) Kulathu, Y., and Komander, D. (2012) Atypical ubiquitylation — the unexplored world of polyubiquitin beyond Lys48 and Lys63 linkages. *Nat. Rev. Mol. Cell Biol. 13*, 508–523.

(74) Walczak, H., Iwai, K., and Dikic, I. (2012) Generation and physiological roles of linear ubiquitin chains. *BMC Biol. 10*, 23.

(75) Rahighi, S., Ikeda, F., Kawasaki, M., Akutsu, M., Suzuki, N., Kato, R., Kensche, T., Uejima, T., Bloor, S., Komander, D., Randow, F., Wakatsuki, S., and Dikic, I. (2009) Specific Recognition of Linear Ubiquitin Chains by NEMO Is Important for NF- k B Activation. *Cell 136*, 1098–1109.

(76) Kim, W., Bennett, E. J., Huttlin, E. L., Guo, A., Li, J., Sowa, M. E., Rad, R., Rush, J., Comb, M. J., Wade, J., and Gygi, S. P. (2011) Systematic and quantitative assessment of the ubiquitin modified proteome. *Mol. Cell 44*, 325–340.

(77) Jin, J., Xie, X., Xiao, Y., Hu, H., Zou, Q., Cheng, X., and Sun, S. (2016) Epigenetic regulation of Il12 and Il23 gene expression and autoimmune inflammation by the deubiquitinase Trabid. *Nat. Immunol. 17*, 259–268.

(78) Wickliffe, K. E., Williamson, A., Meyer, H., Kelly, A., and Rape, M. (2011) K11-linked ubiquitin chains as novel regulators of cell division. *Trends Cell Biol. 21*, 656–663.

(79) Grice, G. L., Lobb, I. T., Michael, P., Gygi, S. P., Antrobus, R., Nathan, J. A., Grice, G. L., Lobb, I. T., Weekes, M. P., Gygi, S. P., Antrobus, R., and Nathan, J. A. (2015) The Proteasome Distinguishes between Heterotypic Report The Proteasome Distinguishes between Heterotypic and Homotypic Lysine-11-Linked Polyubiquitin Chains. *Cell Rep. 12*, 545–553.

(80) Gatti, M., Pinato, S., Aebersold, R., Penengo, L., Gatti, M., Pinato, S., Maiolica, A., Rocchio, F., Prato, M. G., and Aebersold, R. (2015) RNF168 Promotes Noncanonical K27 Ubiquitination to Signal DNA Damage. *Cell Rep. 10*, 226–238.

(81) Wauer, T., Swatek, K. N., Wagstaff, J. L., Gladkova, C., Pruneda, J. N., Michel, M. A., Gersch, M., Johnson, C. M., Freund, S. M. V, and Komander, D. (2015) Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis. *EMBO J. 34*, 307–325.

(82) Swaney, D. L., Rodríguez-mias, R. A., and Villén, J. (2015) Phosphorylation of ubiquitin at Ser 65 affects its polymerization , targets , and proteome-wide turnover. *EMBO Rep. 16*, 1131–1144.

(83) Jin, S. M., Lazarou, M., Wang, C., Kane, L. A., Narendra, D. P., and Youle, R. J. (2010) Mitochondrial membrane potential regulates PINK1 import and proteolytic destabilization by PARL. *J. Cell Biol. 191*, 933–942.

(84) Ohtake, F., Saeki, Y., Sakamoto, K., Ohtake, K., Nishikawa, H., Tsuchiya, H., Ohta, T., Tanaka, K., and Kanno, J. (2015) Ubiquitin acetylation inhibits polyubiquitin chain elongation. *EMBO Rep. 16*, 192–201.

(85) Pao, K., Wood, N. T., Knebel, A., Rafie, K., Stanley, M., Mabbitt, P. D., Sundaramoorthy, R., Hofmann, K., Aalten, D. M. F. Van, and Virdee, S. (2018) Activity-based E3 ligase profiling uncovers an E3 ligase with esterification activity. *Nature 556*, 381–485.

(86) Mabbitt, P. D., Loreto, A., Déry, M., Fletcher, A. J., Stanley, M., Pao, K., Wood, N. T., Coleman, M. P., and Virdee, S. (2020) Structural basis for RING-Cys-Relay E3 ligase activity and its role in axon integrity. *Nat. Chem. Biol. 16*, 1227–1236.

(87) Parp, R. D., Yang, C., Jividen, K., Spencer, A., Jeffery, E., Sherman, N. E., and Paschal, B. M. (2017) Ubiquitin Modification by the E3 Ligase / ADP- Article Ubiquitin Modification by the E3 Ligase / ADP-Ribosyltransferase Dtx3L/Parp9. *Mol. Cell 66*, 503–516.

(88) Mckenna, S., Moraes, T., Pastushok, L., Ptak, C., Xiao, W., Spyracopoulos, L., and Ellison, M. J. (2003) An NMR-based Model of the Ubiquitin-bound Human Ubiquitin Conjugation Complex Mms2.Ubc13. *J. Biol. Chem. 278*, 13151–13158.

(89) Hodge, C. D., Spyracopoulos, L., and Glove, J. N. M. (2016) Ubc13: the Lys63 ubiquitin chain building machine. *Oncotarget 7*, 64471–64504.

(90) Poulsen, S. L., Hansen, R. K., Wagner, S. A., van Cuijk, L., van Belle, G. J., Streicher, W.,

Wikström, M., Choudhary, C., Houtsmuller, A. B., Marteijn, J. A., Bekker-Jensen, S., and Mailand, N. (2013) RNF111/Arkadia is a SUMO-targeted ubiquitin ligase that facilitates the DNA damage response. *J. Cell Biol. 201*, 797–807.

(91) Brusky, J., Zhu, Y., and Xiao, W. (2000) UBC13, a DNA-damage-inducible gene, is a member of the error-free postreplication repair pathway in Saccharomyces cerevisiae. *Curr. Genet. 37*, 168–174.

(92) Motegi, A., Liaw, H.-J., Lee, K.-Y., Roest, H. P., Maas, A., Wu, X., Moinova, H., Markowitz, S. D., Ding, H., Hoeijmakers, J. H. J., and Myung, K. (2008) Polyubiquitination of proliferating cell nuclear antigen by HLTF and SHPRH prevents genomic instability from stalled replication forks. *Proc. Natl. Acad. Sci. 105*, 12411–12416.

(93) Yan, Z., Guo, R., Paramasivam, M., Shen, W., Ling, C., Fox, D., Wang, Y., Oostra, A. B., Kuehl, J., Lee, D.-Y., Takata, M., Hoatlin, M. E., Schindler, D., Joenje, H., de Winter, J. P., Li, L., Seidman, M. M., and Wang, W. (2012) A Ubiquitin-Binding Protein, FAAP20, Links RNF8-Mediated Ubiquitination to the Fanconi Anemia DNA Repair Network. *Mol. Cell 47*, 61–75.

(94) Rai, R., Li, J.-M., Zheng, H., Lok, G. T.-M., Deng, Y., Huen, M. S.-Y., Chen, J., Jin, J., and Chang, S. (2011) The E3 ubiquitin ligase Rnf8 stabilizes Tpp1 to promote telomere end protection. *Nat. Struct. Mol. Biol. 18*, 1400–1407.

(95) Wu, X., and Karin, M. (2015) Emerging roles of Lys63-linked polyubiquitylation in immune responses. *Immunol. Rev. 266*, 161–174.

(96) Madiraju, C., Novack, J. P., Reed, J. C., and Matsuzawa, S. (2022) K63 ubiquitination in immune signaling. *Trends Immunol. 43*, 148–162.

(97) Wertz, I. E., and Dixit, V. M. (2010) Signaling to NF- k B : Regulation by Ubiquitination. *Cold Spring Harb. Perspect. Biol. 2*, a003350.

(98) Moraes, T. F., Edwards, R. A., McKenna, S., Pastushok, L., Xiao, W., Glover, J. N. M., and Ellison, M. J. (2001) Crystal structure of the human ubiquitin conjugating enzyme complex, hMms2–hUbc13. *Nat. Struct. Biol. 8*, 669–673.

(99) Hodge, C. D., Ismail, I. H., Edwards, R. A., Hura, G. L., Xiao, A. T., Tainer, J. A., Hendzel, M. J., and Glover, J. N. M. (2016) RNF8 E3 Ubiquitin Ligase Stimulates Ubc13 E2 Conjugating Activity That Is Essential for DNA Double Strand Break Signaling and BRCA1 Tumor Suppressor Recruitment . *J. Biol. Chem. 291*, 9396–9410.

(100) Hodge, C. D., Edwards, R. A., Markin, C. J., Mcdonald, D., Pulvino, M., Huen, S. Y., Zhao, J., Spyracopoulos, L., Hendzel, M. J., and Glover, J. N. M. (2015) Covalent inhibition of Ubc13 affects ubiquitin signaling and reveals active site elements important for targeting. *ACS Chem. Biol. 10*, 1718–1728.

(101) Mckenna, S., Hu, J., Moraes, T., Xiao, W., Ellison, M. J., Spyracopoulos, L., Mckenna, S., Hu, J., Moraes, T., Xiao, W., Ellison, M. J., and Spyracopoulos, L. (2003) Energetics and Specificity of Interactions within Ub · Uev · Ubc13 Human Ubiquitin Conjugation Complexes Energetics and Specificity of Interactions within Ub , Uev , Ubc13 Human Ubiquitin. *Biochemistry 42*, 7922–7930.

(102) Mckenna, S., Spyracopoulos, L., Moraes, T., Pastushok, L., Ptak, C., Xiao, W., and Ellison, M. J. (2001) Noncovalent Interaction between Ubiquitin and the Human DNA Repair Protein Mms2 Is Required for Ubc13-mediated Polyubiquitination way but lack their characteristic active site cysteine. *J. Biol. Chem. 276*, 40120–40126.

(103) Rout, M. K., Hodge, C. D., Markin, C. J., Xu, X., Glover, J. N. M., Xiao, W., and Spyracopoulos, L. (2014) Stochastic Gate Dynamics Regulate the Catalytic Activity of Ubiquitination Enzymes. *J. Am. Chem. Soc. 136*, 17446–17458.

(104) Vandemark, A. P., Hofmann, R. M., Tsui, C., Pickart, C. M., and Wolberger, C. (2001) Molecular Insights into Polyubiquitin Chain Assembly : Crystal Structure of the Mms2 / Ubc13 Heterodimer. *Cell 105*, 711–720.

(105) Markin, C. J., Saltibus, L. F., Kean, M. J., McKay, R. T., Xiao, W., and Spyracopoulos, L. (2010) Catalytic Proficiency of Ubiquitin Conjugation Enzymes: Balancing pKa Suppression, Entropy, and Electrostatics. *J. Am. Chem. Soc. 132*, 17775–17786.

(106) Berndsen, C. E., Wiener, R., Yu, I. W., Ringel, A. E., and Wolberger, C. (2013) A conserved asparagine has a structural role in ubiquitin-conjugating enzymes. *Nat. Chem. Biol. 9*, 154–156.

(107) Soss, S. E., Yue, Y., Dhe-paganon, S., and Chazin, W. J. (2011) E2 Conjugating Enzyme Selectivity and Requirements for Function of the E3 Ubiquitin Ligase CHIP . *J. Biol. Chem. 286*, 21277–21286.

(108) Skaug, B., Jiang, X., and Chen, Z. J. (2009) The Role of Ubiquitin in NF- κ B Regulatory Pathways. *Annu. Rev. Biochem. 78*, 769–796.

(109) Plechanovová, A., Jaffray, E. G., Tatham, M. H., Naismith, J. H., and Hay, R. T. (2012) Structure of a RING E3 ligase and ubiquitin-loaded E2 primed for catalysis. *Nature 489*, 115–120.

(110) Pruneda, J. N., Littlefield, P. J., Soss, S. E., Nordquist, K. A., Chazin, W. J., Brzovic, P. S., and Klevit, R. E. (2012) Structure of an E3:E2~Ub Complex Reveals an Allosteric Mechanism Shared among RING/U-box Ligases. *Mol. Cell 47*, 933–942.

(111) Yamazaki, K., Gohda, J., Kanayama, A., Miyamoto, Y., Sakurai, H., Yamamoto, M., Akira, S., Hayashi, H., Su, B., and Inoue, J. (2009) Two Mechanistically and Temporally Distinct NF-κB Activation Pathways in IL-1 Signaling. *Sci. Signal. 2*, ra66–ra66.

(112) Wu, Z., Shen, S., Zhang, Z., Zhang, W., and Xiao, W. (2014) Ubiquitin-conjugating enzyme complex Uev1A-Ubc13 promotes breast cancer metastasis through nuclear factor-κB mediated matrix metalloproteinase-1 gene regulation. *Breast Cancer Res. 16*, R75.

(113) Wu, X., Zhang, W., Font-burgada, J., Palmer, T., Hamil, A. S., and Biswas, S. K. (2014) Ubiquitin-conjugating enzyme Ubc13 controls breast cancer metastasis through a TAK1-p38 MAP kinase cascade. *Proc. Natl. Acad. Sci. 111*, 13870–13875.

(114) Su, W.-P., Hsu, S.-H., Wu, C.-K., Chang, S.-B., Lin, Y.-J., Yang, W.-B., Hung, J.-J., Chiu, W.-T., Tzeng, S.-F., Tseng, Y.-L., Chang, J.-Y., Su, W.-C., and Liaw, H. (2014) Chronic treatment with cisplatin induces replication-dependent sister chromatid recombination to confer cisplatin-resistant phenotype in nasopharyngeal carcinoma. *Oncotarget 5*, 6323–6337.

(115) Sakai, E., Kato, H., Rotinsulu, H., Losung, F., Mangindaan, R. E. P., de Voogd, N. J.,

Yokosawa, H., and Tsukamoto, S. (2014) Variabines A and B: new β-carboline alkaloids from the marine sponge Luffariella variabilis. *J. Nat. Med. 68*, 215–219.

(116) Pulvino, M., Liang, Y., Oleksyn, D., DeRan, M., Van Pelt, E., Shapiro, J., Sanz, I., Chen, L., and Zhao, J. (2012) Inhibition of proliferation and survival of diffuse large B-cell lymphoma cells by a small-molecule inhibitor of the ubiquitin-conjugating enzyme Ubc13-Uev1A. *Blood 120*, 1668–1677.

(117) Strickson, S., Campbell, D. G., Emmerich, C. H., Knebel, A., Plater, L., Ritorto, M. S., Shpiro, N., and Cohen, P. (2013) The anti-inflammatory drug BAY 11-7082 suppresses the MyD88-dependent signalling network by targeting the ubiquitin system. *Biochem. J. 451*, 427–437.

(118) Cheng, J., Fan, Y.-H., Xu, X., Zhang, H., Dou, J., Tang, Y., Zhong, X., Rojas, Y., Yu, Y., Zhao, Y., Vasudevan, S. A., Zhang, H., Nuchtern, J. G., Kim, E. S., Chen, X., Lu, F., and Yang, J. (2014) A small-molecule inhibitor of UBE2N induces neuroblastoma cell death via activation of p53 and JNK pathways. *Cell Death Dis. 5*, e1079–e1079.

(119) N. GOMBODORJ, T. Ykobori, S Yoshiyama, R. Kawabata-Iwakawa, S. Rokudai, I. Horikoshi, M. Nishiyama, and T. N. (2017) Inhibition of Ubiquitin-conjugating Enzyme E2 May Activate the Degradation of Hypoxia-inducible Factors and , thus , Overcome Cellular Resistance to Radiation in Colorectal Cancer. *Anticancer Res. 37*, 2425–2436.

(120) Berndsen, C. E., and Wolberger, C. (2014) New insights into ubiquitin E3 ligase mechanism. *Nat. Struct. Mol. Biol. 21*, 301–307.

(121) Yamada, M., Ohnishi, J., Ohkawara, B., Iemura, S., Satoh, K., Hyodo-Miura, J., Kawachi, K., Natsume, T., and Shibuya, H. (2006) NARF, an Nemo-like Kinase (NLK)-associated Ring Finger Protein Regulates the Ubiquitylation and Degradation of T Cell Factor/Lymphoid Enhancer Factor (TCF/LEF) . *J. Biol. Chem. 281*, 20749–20760.

(122) Brooks, W. S., Banerjee, S., and Crawford, D. F. (2014) RNF138 / NARF is a Cell Cycle Regulated E3 Ligase that Poly- ubiquitinates G2E3. *JSM Cell Dev. Biol. 2*, 1005.

(123) Fu, S.-J., Jeng, C.-J., Ma, C.-H., Peng, Y.-J., Lee, C.-M., Fang, Y.-C., Lee, Y.-C., Tang, S.-C., Hu, M.-C., and Tang, C.-Y. (2017) Ubiquitin Ligase RNF138 Promotes Episodic Ataxia Type 2-Associated Aberrant Degradation of Human Ca&lt;sub&gt;v&lt;/sub&gt;2.1 (P/Q-Type) Calcium Channels. *J. Neurosci. 37*, 2485 LP – 2503.

(124) Yu, X., Li, W., Deng, Q., Liu, H., Wang, X., Hu, H., Cao, Y., Xu-Monette, Z. Y., Li, L., Zhang, M., Lu, Z., Young, K. H., and Li, Y. (2021) MYD88 L265P elicits mutation-specific ubiquitination to drive NF-κB activation and lymphomagenesis. *Blood 137*, 1615–1627.

(125) Lu, Y., Huang, R., Ying, J., Li, X., Jiao, T., Guo, L., Zhou, H., Wang, H., Tuersuntuoheti, A., Liu, J., Chen, Q., Wang, Y., Su, L., Guo, C., Xu, F., Wang, Z., Lu, Y., Li, K., Liang, J., Huang, Z., Chen, X., Yao, J., Hu, H., Cheng, X., Wan, Y., Chen, X., Zhang, N., Miao, S., Cai, J., Wang, L., Liu, C., Song, W., and Zhao, H. (2022) RING finger 138 deregulation distorts NF-κB signaling and facilities colitis switch to aggressive malignancy. *Signal Transduct. Target. Ther. 7*, 185.

(126) Xu, L., Lu, Y., Han, D., Yao, R., Wang, H., Zhong, S., Luo, Y., Han, R., Li, K., Fu, J., Zong, S., Miao, S., Song, W., and Wang, L. (2017) Rnf138 deficiency promotes apoptosis of

spermatogonia in juvenile male mice. *Cell Death Dis. 8*, e2795–e2795.

(127) Kim, W., Youn, H., Lee, S., Kim, E., Kim, D., Sub Lee, J., Lee, J.-M., and Youn, B. (2018) RNF138-mediated ubiquitination of rpS3 is required for resistance of glioblastoma cells to radiation-induced apoptosis. *Exp. Mol. Med. 50*, e434–e434.

(128) Lu, Y., Han, D., Liu, W., Huang, R., Ou, J., Chen, X., Zhang, X., Wang, X., Li, S., Wang, L., Liu, C., Miao, S., Wang, L., Ma, C., and Song, W. (2018) RNF138 confers cisplatin resistance in gastric cancer cells via activating Chk1 signaling pathway. *Cancer Biol. Ther. 19*, 1128–1138.

(129) Ismail, I. H., Gagné, J.-P., Genois, M.-M., Strickfaden, H., McDonald, D., Xu, Z., Poirier, G. G., Masson, J.-Y., and Hendzel, M. J. (2015) The RNF138 E3 ligase displaces Ku to promote DNA end resection and regulate DNA repair pathway choice. *Nat. Cell Biol. 17*, 1446–1457.

(130) Schmidt, C. K., Galanty, Y., Sczaniecka-Clift, M., Coates, J., Jhujh, S., Demir, M., Cornwell, M., Beli, P., and Jackson, S. P. (2015) Systematic E2 screening reveals a UBE2D–RNF138–CtIP axis promoting DNA repair. *Nat. Cell Biol. 17*, 1458–1470.

(131) Han, D., Liang, J., Lu, Y., Xu, L., Miao, S., Lu, L.-Y., Song, W., and Wang, L. (2016) Ubiquitylation of Rad51d Mediated by E3 Ligase Rnf138 Promotes the Homologous Recombination Repair Pathway. *PLoS One 11*, e0155476.

(132) Giannini, A. L., Gao, Y., and Bijlmakers, M.-J. (2008) T-cell regulator RNF125/TRAC-1 belongs to a novel family of ubiquitin ligases with zinc fingers and a ubiquitin-binding domain. *Biochem. J. 410*, 101–111.

(133) Iuchi, S. (2001) Three classes of C2H2 zinc finger proteins. *Cell. Mol. Life Sci. C. 58*, 625–635.

(134) Bijlmakers, M.-J., Teixeira, J. M. C., Boer, R., Mayzel, M., Puig-Sàrries, P., Karlsson, G., Coll, M., Pons, M., and Crosas, B. (2016) A C2HC zinc finger is essential for the RING-E2 interaction of the ubiquitin ligase RNF125. *Sci. Rep. 6*, 29232.

(135) Wu, H., Li, X., Feng, M., Yao, L., Deng, Z., Zao, G., Zhou, Y., Chen, S., and Du, Z. (2018) Downregulation of RNF138 inhibits cellular proliferation, migration, invasion and EMT in glioma cells via suppression of the Erk signaling pathway Corrigendum in /10.3892/or.2022.8337. *Oncol Rep 40*, 3285–3296.

(136) Hershko, A., and Ciechanover, A. (1998) The Ubiquitin System. *Annu. Rev. Biochem. 67*, 425–479.

(137) van Wijk, S. J. L., and Timmers, H. T. M. (2010) The family of ubiquitin-conjugating enzymes (E2s): deciding between life and death of proteins. *FASEB J. 24*, 981–993.

(138) Michelle, C., Vourc'h, P., Mignon, L., and Andres, C. R. (2009) What Was the Set of Ubiquitin and Ubiquitin-Like Conjugating Enzymes in the Eukaryote Common Ancestor? *J. Mol. Evol. 68*, 616–628.

(139) Jackson, S. P., and Durocher, D. (2013) Regulation of DNA Damage Responses by Ubiquitin and SUMO. *Mol. Cell 49*, 795–807.

(140) Hofmann, R. M., and Pickart, C. M. (1999) Noncanonical MMs2-Encoded Ubiquitin-

Conjugating Enzyme Functions in Assembly of Novel Polyubiquitin Chains for DNA Repair. *Cell 96*, 645–653.

(141) Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., Richardson, J. S., Richardson, D. C., Sammito, M. D., Sobolev, O. V., Stockwell, D. H., Terwilliger, T. C., Urzhumtsev, A. G., Videau, L. L., Williams, C. J., and Adams, P. D. (2019) Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. Sect. D Struct. Biol. 75*, 861–877.

(142) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 486–501.

(143) Rupesh, K. R., Smith, A., and Boehmer, P. E. (2014) Ligand induced stabilization of the melting temperature of the HSV-1 single-strand DNA binding protein using the thermal shift assay. *Biochem. Biophys. Res. Commun. 454*, 604–608.

(144) Lo, M.-C., Aulabaugh, A., Jin, G., Cowling, R., Bard, J., Malamas, M., and Ellestad, G. (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. *Anal. Biochem. 332*, 153–159.

(145) Moiani, D., Ronato, D. A., Brosey, C. A., Arvai, A. S., Syed, A., Masson, J.-Y., Petricci, E., and Tainer, J. A. (2018) Chapter Ten - Targeting Allostery with Avatars to Design Inhibitors Assessed by Cell Activity: Dissecting MRE11 Endo- and Exonuclease Activities, in *Mechanisms of DNA Recombination and Genome Rearrangements: Intersection between Homologous Recombination, DNA Replication and DNA Repair* (Spies, M., and Malkova, A. B. T.-M. in E., Eds.), pp 205–241. Academic Press.

(146) Rich, T., Allen, R. L., and Wyllie, A. H. (2000) Defying death after DNA damage. *Nature 407*, 777–783.

(147) Hartlerode, A. J., and Scully, R. (2009) Mechanisms of double-strand break repair in somatic mammalian cells. *Biochem. J. 423*, 157–168.

(148) Jasin, M., and Rothstein, R. (2013) Repair of Strand Breaks by Homologous Recombination. *Cold Spring Harb. Perspect. Biol. 5*, a012740.

(149) Curtin, N. J. (2012) DNA repair dysregulation from cancer driver to therapeutic target. *Nat. Rev. Cancer 12*, 801–817.

(150) van Wijk, S. J. L., de Vries, S. J., Kemmeren, P., Huang, A., Boelens, R., Bonvin, A. M. J. J., and Timmers, H. T. M. (2009) A comprehensive framework of E2–RING E3 interactions of the human ubiquitin–proteasome system. *Mol. Syst. Biol. 5*, 295.

(151) Metzger, M. B., Pruneda, J. N., Klevit, R. E., and Weissman, A. M. (2014) RING-type E3 ligases: Master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim. Biophys. Acta - Mol. Cell Res. 1843*, 47–60.

(152) Qian, H., Zhang, Y., Wu, B., Wu, S., You, S., Zhang, N., and Sun, Y. (2020) Structure and function of HECT E3 ubiquitin ligases and their role in oxidative stress *8*, 71–79.

(153) Middleton, A. J., Barzak, F. M., Fokkens, T. J., Nguyen, K., and Day, C. L. (2023) Zinc

finger 1 of the RING E3 ligase, RNF125, interacts with the E2 to enhance ubiquitylation. *Structure 31*, 1208-1219.e5.

(154) Evans, R., O'Neill, M., Pritzel, A., Antropova, N., Senior, A., Green, T., Žídek, A., Bates, R., Blackwell, S., Yim, J., Ronneberger, O., Bodenstein, S., Zielinski, M., Bridgland, A., Potapenko, A., Cowie, A., Tunyasuvunakool, K., Jain, R., Clancy, E., Kohli, P., Jumper, J., and Hassabis, D. (2022) Protein complex prediction with AlphaFold-Multimer. *bioRxiv* 2021.10.04.463034.

(155) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature 596*, 583–589.

(156) Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Žídek, A., Bridgland, A., Cowie, A., Meyer, C., Laydon, A., Velankar, S., Kleywegt, G. J., Bateman, A., Evans, R., Pritzel, A., Figurnov, M., Ronneberger, O., Bates, R., Kohl, S. A. A., Potapenko, A., Ballard, A. J., Romera-Paredes, B., Nikolov, S., Jain, R., Clancy, E., Reiman, D., Petersen, S., Senior, A. W., Kavukcuoglu, K., Birney, E., Kohli, P., Jumper, J., and Hassabis, D. (2021) Highly accurate protein structure prediction for the human proteome. *Nature 596*, 590–596.

(157) Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) ColabFold: making protein folding accessible to all. *Nat. Methods 19*, 679–682.

(158) Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Applying and improving AlphaFold at CASP14. *Proteins Struct. Funct. Bioinforma. 89*, 1711–1721.

(159) Xu, J., and Zhang, Y. (2010) How significant is a protein structure similarity with TM-score = 0.5? *Bioinformatics 26*, 889–895.

(160) Mariani, V., Biasini, M., Barbato, A., and Schwede, T. (2013) lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics 29*, 2722–2728.

(161) Middleton, A. J., Budhidarmo, R., Das, A., Zhu, J., Foglizzo, M., Mace, P. D., and Day, C. L. (2017) The activity of TRAF RING homo- and heterodimers is regulated by zinc finger 1. *Nat. Commun. 8*, 1788.

(162) Nayak, D., and Sivaraman, J. (2018) Structure of LNX1:Ubc13~Ubiquitin Complex Reveals the Role of Additional Motifs for the E3 Ligase Activity of LNX1. *J. Mol. Biol. 430*, 1173–1188.

(163) Nayak, D., and Sivaraman, J. (2015) Structural basis for the indispensable role of a unique zinc finger motif in LNX2 ubiquitination. *Oncotarget 6*, 34342–34357.

(164) Rennie, M. L., Chaugule, V. K., and Walden, H. (2020) Modes of allosteric regulation of the ubiquitination machinery. *Curr. Opin. Struct. Biol. 62*, 189–196.

(165) Das, R., Mariano, J., Tsai, Y. C., Kalathur, R. C., Kostova, Z., Li, J., Tarasov, S. G., McFeeters, R. L., Altieri, A. S., Ji, X., Byrd, R. A., and Weissman, A. M. (2009) Allosteric Activation of E2-RING Finger-Mediated Ubiquitylation by a Structurally Defined Specific E2-Binding Region of gp78. *Mol. Cell 34*, 674–685.

(166) Metzger, M. B., Liang, Y.-H., Das, R., Mariano, J., Li, S., Li, J., Kostova, Z., Byrd, R. A., Ji, X., and Weissman, A. M. (2013) A Structurally Unique E2-Binding Domain Activates Ubiquitination by the ERAD E2, Ubc7p, through Multiple Mechanisms. *Mol. Cell 50*, 516–527.

(167) Khanna, K. K., and Jackson, S. P. (2001) DNA double-strand breaks: signaling, repair and the cancer connection. *Nat. Genet. 27*, 247–254.

(168) Wright, W. D., Shah, S. S., and Heyer, W.-D. (2018) Homologous recombination and the repair of DNA double-strand breaks. *J. Biol. Chem. 293*, 10524–10535.

(169) Rothkamm, K., Krüger, I., Thompson, L. H., and Löbrich, M. (2003) Pathways of DNA Double-Strand Break Repair during the Mammalian Cell Cycle. *Mol. Cell. Biol. 23*, 5706–5715.

(170) Takata, M., Sasaki, M. S., Sonoda, E., Morrison, C., Hashimoto, M., Utsumi, H., Yamaguchi-Iwai, Y., Shinohara, A., and Takeda, S. (1998) 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. 17*, 5497–5508.

(171) Cejka, P., and Symington, L. S. (2021) DNA End Resection: Mechanism and Control. *Annu. Rev. Genet. 55*, 285–307.

(172) Anand, R., Ranjha, L., Cannavo, E., and Cejka, P. (2016) Phosphorylated CtIP Functions as a Co-factor of the MRE11-RAD50-NBS1 Endonuclease in DNA End Resection. *Mol. Cell 64*, 940–950.

(173) Deshpande, R. A., Lee, J.-H., Arora, S., and Paull, T. T. (2016) Nbs1 Converts the Human Mre11/Rad50 Nuclease Complex into an Endo/Exonuclease Machine Specific for Protein-DNA Adducts. *Mol. Cell 64*, 593–606.

(174) Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S. P. (2007) Human CtIP promotes DNA end resection. *Nature 450*, 509–514.

(175) Mimori, T., and Hardin, J. A. (1986) Mechanism of interaction between Ku protein and DNA. *J. Biol. Chem. 261*, 10375–10379.

(176) Pierce, A. J., Hu, P., Han, M., Ellis, N., and Jasin, M. (2001) Ku DNA end-binding protein modulates homologous repair of double-strand breaks in mammalian cells. *Genes Dev. 15*, 3237–3242.

(177) Fattah, F., Lee, E. H., Weisensel, N., Wang, Y., Lichter, N., and Hendrickson, E. A. (2010) Ku Regulates the Non-Homologous End Joining Pathway Choice of DNA Double-Strand Break Repair in Human Somatic Cells. *PLOS Genet. 6*, e1000855.

(178) Scully, R., Panday, A., Elango, R., and Willis, N. A. (2019) DNA double-strand break repair-

pathway choice in somatic mammalian cells. *Nat. Rev. Mol. Cell Biol. 20*, 698–714.

(179) Shao, Z., Davis, A. J., Fattah, K. R., So, S., Sun, J., Lee, K.-J., Harrison, L., Yang, J., and Chen, D. J. (2012) Persistently bound Ku at DNA ends attenuates DNA end resection and homologous recombination. *DNA Repair (Amst). 11*, 310–316.

(180) Hustedt, N., and Durocher, D. (2017) The control of DNA repair by the cell cycle. *Nat. Cell Biol. 19*, 1–9.

(181) Schwertman, P., Bekker-Jensen, S., and Mailand, N. (2016) Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat. Rev. Mol. Cell Biol. 17*, 379–394.

(182) Bekker-Jensen, S., and Mailand, N. (2015) RNF138 joins the HR team. *Nat. Cell Biol. 17*, 1375–1377.

(183) Sullivan, M. R., and Bernstein, K. A. (2018) RAD-ical New Insights into RAD51 Regulation. *Genes (Basel).*

(184) Békés, M., Okamoto, K., Crist, S. B., Jones, M. J., Chapman, J. R., Brasher, B. B., Melandri, F. D., Ueberheide, B. M., Lazzerini Denchi, E., and Huang, T. T. (2013) DUB-Resistant Ubiquitin to Survey Ubiquitination Switches in Mammalian Cells. *Cell Rep. 5*, 826–838.

(185) Ma, H., McLean, J. R., Chao, L. F.-I., Mana-Capelli, S., Paramasivam, M., Hagstrom, K. A., Gould, K. L., and McCollum, D. (2012) A Highly Efficient Multifunctional Tandem Affinity Purification Approach Applicable to Diverse Organisms . *Mol. Cell. Proteomics 11*, 501–511.

(186) Meierhofer, D., Wang, X., Huang, L., and Kaiser, P. (2008) Quantitative Analysis of global Ubiquitination in HeLa Cells by Mass Spectrometry. *J. Proteome Res. 7*, 4566–4576.

(187) Khurana, S., Kruhlak, M. J., Kim, J., Tran, A. D., Liu, J., Nyswaner, K., Shi, L., Jailwala, P., Sung, M.-H., Hakim, O., and Oberdoerffer, P. (2014) A Macrohistone Variant Links Dynamic Chromatin Compaction to BRCA1-Dependent Genome Maintenance. *Cell Rep. 8*, 1049–1062.

(188) Robert, X., and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res. 42*, 320–324.

(189) Whitman, B. T., Murray, C. R. A., Whitford, D. S., Paul, S. S., Fahlman, R. P., Glover, M. J. N., and Owttrim, G. W. (2022) Degron-mediated proteolysis of CrhR-like DEAD-box RNA helicases in cyanobacteria. *J. Biol. Chem. 298*, 101925.

(190) Vaughn, J.P., Davis, P.L., Jarboe, M.D., Huper, G., Craig Evans, A., Wiseman, R.W., Berchuck, A., Iglehart, J.D., Futreal, P.A., Marks, J. R. (1996) BRCA1 expression is induced before DNA synthesis in both normal and tumor-derived breast cells. *Cell Growth Differ. 7*, 711–715.

(191) Ruffner, H., and Verma, I. M. (1997) BRCA1 is a cell cycle-regulated nuclear phosphoprotein. *Proc. Natl. Acad. Sci. 94*, 7138–7143.

(192) Scully, R., Chen, J., Ochs, R. L., Keegan, K., Hoekstra, M., Feunteun, J., and Livingston, D. M. (1997) Dynamic Changes of BRCA1 Subnuclear Location and Phosphorylation State Are Initiated by DNA Damage. *Cell 90*, 425–435.

(193) Malumbres, M. (2014) Cyclin-dependent kinases. *Genome Biol. 15*, 122.

(194) Deans, A. J., Khanna, K. K., McNees, C. J., Mercurio, C., Heierhorst, J., and McArthur, G. A. (2006) Cyclin-Dependent Kinase 2 Functions in Normal DNA Repair and Is a Therapeutic Target in BRCA1-Deficient Cancers. *Cancer Res. 66*, 8219–8226.

(195) Ira, G., Pellicioli, A., Balijja, A., Wang, X., Fiorani, S., Carotenuto, W., Liberi, G., Bressan, D., Wan, L., Hollingsworth, N. M., Haber, J. E., and Foiani, M. (2004) DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature 431*, 1011–1017.

(196) Simoneau, A., Robellet, X., Ladouceur, A.-M., and D'Amours, D. (2014) Cdk1-dependent regulation of the Mre11 complex couples DNA repair pathways to cell cycle progression. *Cell Cycle 13*, 1078–1090.

(197) Falck, J., Forment, J. V, Coates, J., Mistrik, M., Lukas, J., Bartek, J., and Jackson, S. P. (2012) CDK targeting of NBS1 promotes DNA-end resection, replication restart and homologous recombination. *EMBO Rep. 13*, 561–568.

(198) Wohlbold, L., Merrick, K. A., De, S., Amat, R., Kim, J. H., Larochelle, S., Allen, J. J., Zhang, C., Shokat, K. M., Petrini, J. H. J., and Fisher, R. P. (2012) Chemical Genetics Reveals a Specific Requirement for Cdk2 Activity in the DNA Damage Response and Identifies Nbs1 as a Cdk2 Substrate in Human Cells. *PLOS Genet. 8*, e1002935.

(199) Huertas, P., Cortés-Ledesma, F., Sartori, A. A., Aguilera, A., and Jackson, S. P. (2008) CDK targets Sae2 to control DNA-end resection and homologous recombination. *Nature 455*, 689–692.

(200) Huertas, P., and Jackson, S. P. (2009) Human CtIP Mediates Cell Cycle Control of DNA End Resection and Double Strand Break Repair . *J. Biol. Chem. 284*, 9558–9565.

(201) Wang, H., Shi, L. Z., Wong, C. C. L., Han, X., Hwang, P. Y.-H., Truong, L. N., Zhu, Q., Shao, Z., Chen, D. J., Berns, M. W., Yates III, J. R., Chen, L., and Wu, X. (2013) The Interaction of CtIP and Nbs1 Connects CDK and ATM to Regulate HR–Mediated Double-Strand Break Repair. *PLOS Genet. 9*, e1003277.

(202) Chen, X., Niu, H., Chung, W.-H., Zhu, Z., Papusha, A., Shim, E. Y., Lee, S. E., Sung, P., and Ira, G. (2011) Cell cycle regulation of DNA double-strand break end resection by Cdk1-dependent Dna2 phosphorylation. *Nat. Struct. Mol. Biol. 18*, 1015–1019.

(203) Tomimatsu, N., Mukherjee, B., Catherine Hardebeck, M., Ilcheva, M., Vanessa Camacho, C., Louise Harris, J., Porteus, M., Llorente, B., Khanna, K. K., and Burma, S. (2014) Phosphorylation of EXO1 by CDKs 1 and 2 regulates DNA end resection and repair pathway choice. *Nat. Commun. 5*, 3561.

(204) Wang, C., Xu, H., Lin, S., Deng, W., Zhou, J., Zhang, Y., Shi, Y., Peng, D., and Xue, Y. (2020) GPS 5.0: An Update on the Prediction of Kinase-specific Phosphorylation Sites in Proteins. *Genomics. Proteomics Bioinformatics 18*, 72–80.

(205) Bach, S., Knockaert, M., Reinhardt, J., Lozach, O., Schmitt, S., Baratte, B., Koken, M., Coburn, S. P., Tang, L., Jiang, T., Liang, D., Galons, H., Dierick, J.-F., Pinna, L. A., Meggio, F., Totzke, F., Schächtele, C., Lerman, A. S., Carnero, A., Wan, Y., Gray, N., and Meijer, L. (2005)

Roscovitine Targets, Protein Kinases and Pyridoxal Kinase [boxs]. *J. Biol. Chem. 280*, 31208–31219.

(206) Byth, K. F., Thomas, A., Hughes, G., Forder, C., McGregor, A., Geh, C., Oakes, S., Green, C., Walker, M., Newcombe, N., Green, S., Growcott, J., Barker, A., and Wilkinson, R. W. (2009) AZD5438, a potent oral inhibitor of cyclin-dependent kinases 1, 2, and 9, leads to pharmacodynamic changes and potent antitumor effects in human tumor xenografts. *Mol. Cancer Ther. 8*, 1856–1866.

(207) Vassilev, L. T., Tovar, C., Chen, S., Knezevic, D., Zhao, X., Sun, H., Heimbrook, D. C., and Chen, L. (2006) Selective small-molecule inhibitor reveals critical mitotic functions of human CDK1. *Proc. Natl. Acad. Sci. 103*, 10660–10665.

(208) Himmels, S.-F., and Sartori, A. A. (2016) Controlling DNA-End Resection: An Emerging Task for Ubiquitin and SUMO. *Front. Genet. 7*.

(209) Locke, A. J., Hossain, L., McCrostie, G., Ronato, D. A., Fitieh, A., Rafique, T. A., Mashayekhi, F., Motamedi, M., Masson, J.-Y., and Ismail, I. H. (2021) SUMOylation mediates CtIP's functions in DNA end resection and replication fork protection. *Nucleic Acids Res. 49*, 928–953.

(210) Boeing, S., Williamson, L., Encheva, V., Gori, I., Saunders, R. E., Instrell, R., Aygün, O., Rodriguez-Martinez, M., Weems, J. C., Kelly, G. P., Conaway, J. W., Conaway, R. C., Stewart, A., Howell, M., Snijders, A. P., and Svejstrup, J. Q. (2016) Multiomic Analysis of the UV-Induced DNA Damage Response. *Cell Rep. 15*, 1597–1610.

(211) Elia, A. E. H., Boardman, A. P., Wang, D. C., Huttlin, E. L., Everley, R. A., Dephoure, N., Zhou, C., Koren, I., Gygi, S. P., and Elledge, S. J. (2015) Quantitative Proteomic Atlas of Ubiquitination and Acetylation in the DNA Damage Response. *Mol. Cell 59*, 867–881.

(212) Udeshi, N. D., Svinkina, T., Mertins, P., Kuhn, E., Mani, D. R., Qiao, J. W., and Carr, S. A. (2013) Refined Preparation and Use of Anti-diglycine Remnant (K-ε-GG) Antibody Enables Routine Quantification of 10,000s of Ubiquitination Sites in Single Proteomics Experiments . *Mol. Cell. Proteomics 12*, 825–831.

(213) Mertins, P., Qiao, J. W., Patel, J., Udeshi, N. D., Clauser, K. R., Mani, D. R., Burgess, M. W., Gillette, M. A., Jaffe, J. D., and Carr, S. A. (2013) Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nat. Methods 10*, 634–637.

(214) Grice, G. L., and Nathan, J. A. (2016) The recognition of ubiquitinated proteins by the proteasome. *Cell. Mol. Life Sci. 73*, 3497–3506.

(215) Pommier, Y. (2006) Topoisomerase I inhibitors: camptothecins and beyond. *Nat. Rev. Cancer 6*, 789–802.

(216) Raderschall, E., Golub, E. I., and Haaf, T. (1999) Nuclear foci of mammalian recombination proteins are located at single-stranded DNA regions formed after DNA damage. *Proc. Natl. Acad. Sci. 96*, 1921–1926.

(217) Mukherjee, B., Tomimatsu, N., and Burma, S. (2015) Immunofluorescence-Based Methods to Monitor DNA End Resection BT - Stress Responses: Methods and Protocols (Oslowski, C. M.,

Ed.), pp 67–75. Springer New York, New York, NY.

(218) Alani, E., Thresher, R., Griffith, J. D., and Kolodner, R. D. (1992) Characterization of DNA-binding and strand-exchange stimulation properties of y-RPA, a yeast single-strand-DNA-binding protein. *J. Mol. Biol. 227*, 54–71.

(219) Chen, H., Lisby, M., and Symington, L. S. (2013) RPA Coordinates DNA End Resection and Prevents Formation of DNA Hairpins. *Mol. Cell 50*, 589–600.

(220) Binz, S. K., Sheehan, A. M., and Wold, M. S. (2004) Replication Protein A phosphorylation and the cellular response to DNA damage. *DNA Repair (Amst). 3*, 1015–1024.

(221) Cruet-Hennequart, S., Glynn, M. T., Murillo, L. S., Coyne, S., and Carty, M. P. (2008) Enhanced DNA-PK-mediated RPA2 hyperphosphorylation in DNA polymerase η-deficient human cells treated with cisplatin and oxaliplatin. *DNA Repair (Amst). 7*, 582–596.

(222) Liu, W., Wang, Z., Liu, S., Zhang, X., Cao, X., and Jiang, M. (2023) RNF138 inhibits late inflammatory gene transcription through degradation of SMARCC1 of the SWI/SNF complex. *Cell Rep. 42*, 112097.

(223) Anand, R., Jasrotia, A., Bundschuh, D., Howard, S. M., Ranjha, L., Stucki, M., and Cejka, P. (2019) NBS1 promotes the endonuclease activity of the MRE11-RAD50 complex by sensing CtIP phosphorylation. *EMBO J. 38*, e101005.

(224) Mukherjee, S., Chakraborty, P., and Saha, P. (2016) Phosphorylation of Ku70 subunit by cell cycle kinases modulates the replication related function of Ku heterodimer. *Nucleic Acids Res. 44*, 7755–7765.

(225) Carballar, R., Martínez-Láinez, J. M., Samper, B., Bru, S., Bállega, E., Mirallas, O., Ricco, N., Clotet, J., and Jiménez, J. (2020) CDK-mediated Yku80 Phosphorylation Regulates the Balance Between Non-homologous End Joining (NHEJ) and Homologous Directed Recombination (HDR). *J. Mol. Biol. 432*, 166715.

(226) Postow, L., Ghenoiu, C., Woo, E. M., Krutchinsky, A. N., Chait, B. T., and Funabiki, H. (2008) Ku80 removal from DNA through double strand break–induced ubiquitylation . *J. Cell Biol. 182*, 467–479.

(227) Gatti, L., Hoe, K. L., Hayles, J., Righetti, S. C., Carenini, N., Bo, L. D., Kim, D. U., Park, H. O., and Perego, P. (2011) Ubiquitin-proteasome genes as targets for modulation of cisplatin sensitivity in fission yeast. *BMC Genomics 12*, 44.

(228) Vuorela, M., Pylkäs, K., and Winqvist, R. (2011) Mutation screening of the RNF8, UBC13 and MMS2 genes in Northern Finnish breast cancer families. *BMC Med. Genet. 12*, 98.

(229) Mather, B. D., Viswanathan, K., Miller, K. M., and Long, T. E. (2006) Michael addition reactions in macromolecular design for emerging technologies. *Prog. Polym. Sci. 31*, 487–531.

(230) Vass, M., Hruska, K., and Franek, M. (2008) Nitrofuran antibiotics: a review on the application, prohibition and residual analysis. *Vet. Med. (Praha). 53*, 469–500.

(231) Huang, A., Garraway, L. A., Ashworth, A., and Weber, B. (2020) Synthetic lethality as an engine for cancer drug target discovery. *Nat. Rev. Drug Discov. 19*, 23–38.

(232) Topatana, W., Juengpanich, S., Li, S., Cao, J., Hu, J., Lee, J., Suliyanto, K., Ma, D., Zhang, B., Chen, M., and Cai, X. (2020) Advances in synthetic lethality for cancer therapy: cellular mechanism and clinical translation. *J. Hematol. Oncol. 13*, 118.

(233) Lord, C. J., Tutt, A. N. J., and Ashworth, A. (2015) Synthetic Lethality and Cancer Therapy : Lessons Learned from the Development of PARP Inhibitors. *Annu. Rev. Med. 66*, 455–470.

(234) Liu, Q.-W., Yang, Z.-W., Tang, Q.-H., Wang, W.-E., Chu, D.-S., Ji, J.-F., Fan, Q.-Y., Jiang, H., Yang, Q.-X., Zhang, H., Liu, X.-Y., Xu, X.-S., Wang, X.-F., Liu, J.-B., Fu, D., Tao, K., and Yu, H. (2024) The power and the promise of synthetic lethality for clinical application in cancer treatment. *Biomed. Pharmacother. 172*, 116288.

(235) Masud, T., Soong, C., Xu, H., Biele, J., Bjornson, S., McKinney, S., and Aparicio, S. (2021) Ubiquitin-mediated DNA damage response is synthetic lethal with G-quadruplex stabilizer CX-5461. *Sci. Rep. 11*, 9812.

(236) Lopes, J., Piazza, A., Bermejo, R., Kriegsman, B., Colosio, A., Teulade-Fichou, M., Foiani, M., and Nicolas, A. (2011) G-quadruplex-induced instability during leading-strand replication. *EMBO J. 30*, 4033–4046.

(237) Morris, M. J., Negishi, Y., Pazsint, C., and Schonhoft, J. D. (2010) An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human VEGF IRES. *J. Am. Chem. Soc. 132*, 17831–17839.

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

(239) Jimeno, S., Camarillo, R., Mejı, F., Jesu, M., Soria-bretones, I., Prados-carvajal, R., and Huertas, P. (2018) The Helicase PIF1 Facilitates Resection over Sequences Prone to Forming G4 Structures Article The Helicase PIF1 Facilitates Resection over Sequences Prone to Forming G4 Structures. *Cell Rep. 24*, 3262–3273.

(240) Paeschke, K., Capra, J. A., and Zakian, V. A. (2011) DNA Replication through G-Quadruplex Motifs Is Promoted by the <em>Saccharomyces cerevisiae</em> Pif1 DNA Helicase. *Cell 145*, 678–691.

(241) Sarkies, P., Murat, P., Phillips, L. G., Patel, K. J., Balasubramanian, S., and Sale, J. E. (2012) FANCJ coordinates two pathways that maintain epigenetic stability at G-quadruplex DNA. *Nucleic Acids Res. 40*, 1485–1498.

(242) Eddy, S., Ketkar, A., Zafar, M. K., Maddukuri, L., Choi, J.-Y., and Eoff, R. L. (2014) Human Rev1 polymerase disrupts G-quadruplex DNA. *Nucleic Acids Res. 42*, 3272–3285.

(243) Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C. B., Proffitt, C., Trent, K., Whitten, J. P., Lim, J. K. C., Von Hoff, D., Anderes, K., and Rice, W. G. (2009) Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Res. 69*, 7653–7661.

(244) Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N. J., Santos, N. Dos, Silvester, J., Wei, V., Garcia, J., Kabeer, F., Lai, D., Soriano, P., Banáth, J., Chiu, D. S., Yap, D.,

Le, D. D., Ye, F. B., Zhang, A., Thu, K., Soong, J., Lin, S., Tsai, A. H. C., Osako, T., Algara, T., Saunders, D. N., Wong, J., Xian, J., Bally, M. B., Brenton, J. D., Brown, G. W., Shah, S. P., Cescon, D., Mak, T. W., Caldas, C., Stirling, P. C., Hieter, P., Balasubramanian, S., and Aparicio, S. (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat. Commun. 8*, 14432.

(245) Khot, A., Brajanovski, N., Cameron, D. P., Hein, N., Maclachlan, K. H., Sanij, E., Lim, J., Soong, J., Link, E., Blombery, P., Thompson, E. R., Fellowes, A., Sheppard, K. E., McArthur, G. A., Pearson, R. B., Hannan, R. D., Poortinga, G., and Harrison, S. J. (2019) First-in-Human RNA Polymerase I Transcription Inhibitor CX-5461 in Patients with Advanced Hematologic Cancers: Results of a Phase I Dose-Escalation Study. *Cancer Discov. 9*, 1036–1049.

(246) Negi, S. S., and Brown, P. (2015) Transient rRNA synthesis inhibition with CX-5461 is sufficient to elicit growth arrest and cell death in acute lymphoblastic leukemia cells. *Oncotarget 6*, 34846–34858.

(247) Lee, H. C., Wang, H., Baladandayuthapani, V., Lin, H., He, J., Jones, R. J., Kuiatse, I., Gu, D., Wang, Z., Ma, W., Lim, J., O'Brien, S., Keats, J., Yang, J., Davis, R. E., and Orlowski, R. Z. (2017) RNA Polymerase I Inhibition with CX-5461 as a Novel Therapeutic Strategy to Target MYC in Multiple Myeloma. *Br. J. Haematol. 177*, 80–94.

(248) Hein, N., Cameron, D. P., Hannan, K. M., Nguyen, N.-Y. N., Fong, C. Y., Sornkom, J., Wall, M., Pavy, M., Cullinane, C., Diesch, J., Devlin, J. R., George, A. J., Sanij, E., Quin, J., Poortinga, G., Verbrugge, I., Baker, A., Drygin, D., Harrison, S. J., Rozario, J. D., Powell, J. A., Pitson, S. M., Zuber, J., Johnstone, R. W., Dawson, M. A., Guthridge, M. A., Wei, A., McArthur, G. A., Pearson, R. B., and Hannan, R. D. (2017) Inhibition of Pol I transcription treats murine and human AML by targeting the leukemia-initiating cell population. *Blood 129*, 2882–2895.

(249) Hald, Ø. H., Olsen, L., Gallo-Oller, G., Elfman, L. H. M., Løkke, C., Kogner, P., Sveinbjörnsson, B., Flægstad, T., Johnsen, J. I., and Einvik, C. (2019) Inhibitors of ribosome biogenesis repress the growth of MYCN-amplified neuroblastoma. *Oncogene 38*, 2800–2813.

(250) Cornelison, R., Dobbin, Z. C., Katre, A. A., Jeong, D. H., Zhang, Y., Chen, D., Petrova, Y., Llaneza, D. C., Steg, A. D., Parsons, L., Schneider, D. A., and Landen, C. N. (2017) Targeting RNA-Polymerase I in Both Chemosensitive and Chemoresistant Populations in Epithelial Ovarian Cancer. *Clin. Cancer Res. 23*, 6529–6540.

(251) Rebello, R. J., Kusnadi, E., Cameron, D. P., Pearson, H. B., Lesmana, A., Devlin, J. R., Drygin, D., Clark, A. K., Porter, L., Pedersen, J., Sandhu, S., Risbridger, G. P., Pearson, R. B., Hannan, R. D., and Furic, L. (2016) The Dual Inhibition of RNA Pol I Transcription and PIM Kinase as a New Therapeutic Approach to Treat Advanced Prostate Cancer. *Clin. Cancer Res. 22*, 5539–5552.

(252) Hilton, J., Gelmon, K., Bedard, P. L., Tu, D., Xu, H., Tinker, A. V, Goodwin, R., Laurie, S. A., Jonker, D., Hansen, A. R., Veitch, Z. W., Renouf, D. J., Hagerman, L., Lui, H., Chen, B., Kellar, D., Li, I., Lee, S.-E., Kono, T., Cheng, B. Y. C., Yap, D., Lai, D., Beatty, S., Soong, J., Pritchard, K. I., Soria-Bretones, I., Chen, E., Feilotter, H., Rushton, M., Seymour, L., Aparicio, S., and Cescon, D. W. (2022) Results of the phase I CCTG IND.231 trial of CX-5461 in patients with advanced solid tumors enriched for DNA-repair deficiencies. *Nat. Commun. 13*, 3607.

(253) Zhang, B., Tang, C., Yao, Y., Chen, X., Zhou, C., Wei, Z., Xing, F., Chen, L., Cai, X., Zhang, Z., Sun, S., and Liu, Q. (2021) The tumor therapy landscape of synthetic lethality. *Nat. Commun. 12*, 1275.

(254) Krasnoselska, G. O., Dumoux, M., Gamage, N., Cheruvara, H., Birch, J., Quigley, A., and Owens, R. J. (2021) Transient Transfection and Expression of Eukaryotic Membrane Proteins in Expi293F Cells and Their Screening on a Small Scale: Application for Structural Studies BT - Structural Proteomics: High-Throughput Methods, in *Structural Proteomics* (Owens, R. J., Ed.), pp 105–128. Springer US, New York, NY.

(255) Greven, J. A., and Brett, T. J. (2022) Production of Eukaryotic Glycoproteins for Structural and Functional Studies Using Expi293F Cells. *Curr. Protoc. 2*, e512.

(256) Miller, S. L. H., Malotky, E., and O'Bryan, J. P. (2004) Analysis of the Role of Ubiquitin-interacting Motifs in Ubiquitin Binding and Ubiquitylation . *J. Biol. Chem. 279*, 33528–33537.

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

(258) Hura, G. L., Menon, A. L., Hammel, M., Rambo, R. P., Poole II, F. L., Tsutakawa, S. E., Jenney Jr, F. E., Classen, S., Frankel, K. A., Hopkins, R. C., Yang, S., Scott, J. W., Dillard, B. D., Adams, M. W. W., and Tainer, J. A. (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat. Methods 6*, 606–612.

(259) Hura, G. L., Budworth, H., Dyer, K. N., Rambo, R. P., Hammel, M., McMurray, C. T., and Tainer, J. A. (2013) Comprehensive macromolecular conformations mapped by quantitative SAXS analyses. *Nat. Methods 10*, 453–454.

(260) Brookes, E., Rocco, M., Vachette, P., and Trewhella, J. (2023) AlphaFold-predicted protein structures and small-angle X-ray scattering: insights from an extended examination of selected data in the Small-Angle Scattering Biological Data Bank. *J. Appl. Crystallogr. 56*, 910–926.

(261) Chinnam, N. B., Syed, A., Hura, G. L., Hammel, M., Tainer, J. A., and Tsutakawa, S. E. (2023) Chapter Twelve - Combining small angle X-ray scattering (SAXS) with protein structure predictions to characterize conformations in solution, in *Small Angle Scattering Part B: Methods for Structural Interpretation* (Tainer, J. A. B. T.-M. in E., Ed.), pp 351–376. Academic Press.

(262) Terwilliger, T. C., Afonine, P. V, Liebschner, D., Croll, T. I., McCoy, A. J., Oeffner, R. D., Williams, C. J., Poon, B. K., Richardson, J. S., Read, R. J., and Adams, P. D. (2023) Accelerating crystal structure determination with iterative AlphaFold prediction. *Acta Crystallogr. Sect. D 79*, 234–244.

(263) Terwilliger, T. C., Liebschner, D., Croll, T. I., Williams, C. J., McCoy, A. J., Poon, B. K., Afonine, P. V, Oeffner, R. D., Richardson, J. S., Read, R. J., and Adams, P. D. (2024) AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods 21*, 110–116.

(264) Hospenthal, M. K., Mevissen, T. E. T., and Komander, D. (2015) Deubiquitinase-based analysis of ubiquitin chain architecture using Ubiquitin Chain Restriction (UbiCRest). *Nat. Protoc. 10*, 349–361.

(265) Snyder, N. A., and Silva, G. M. (2021) Deubiquitinating enzymes (DUBs): Regulation, homeostasis, and oxidative stress response. *J. Biol. Chem. 297*.

(266) Lek, M., Karczewski, K. J., Minikel, E. V, Samocha, K. E., Banks, E., Fennell, T., O'Donnell-Luria, A. H., Ware, J. S., Hill, A. J., Cummings, B. B., Tukiainen, T., Birnbaum, D. P., Kosmicki, J. A., Duncan, L. E., Estrada, K., Zhao, F., Zou, J., Pierce-Hoffman, E., Berghout, J., Cooper, D. N., Deflaux, N., DePristo, M., Do, R., Flannick, J., Fromer, M., Gauthier, L., Goldstein, J., Gupta, N., Howrigan, D., Kiezun, A., Kurki, M. I., Moonshine, A. L., Natarajan, P., Orozco, L., Peloso, G. M., Poplin, R., Rivas, M. A., Ruano-Rubio, V., Rose, S. A., Ruderfer, D. M., Shakir, K., Stenson, P. D., Stevens, C., Thomas, B. P., Tiao, G., Tusie-Luna, M. T., Weisburd, B., Won, H.-H., Yu, D., Altshuler, D. M., Ardissino, D., Boehnke, M., Danesh, J., Donnelly, S., Elosua, R., Florez, J. C., Gabriel, S. B., Getz, G., Glatt, S. J., Hultman, C. M., Kathiresan, S., Laakso, M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Neale, B. M., Palotie, A., Purcell, S. M., Saleheen, D., Scharf, J. M., Sklar, P., Sullivan, P. F., Tuomilehto, J., Tsuang, M. T., Watkins, H. C., Wilson, J. G., Daly, M. J., MacArthur, D. G., and Consortium, E. A. (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature 536*, 285–291.

(267) Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., and Marchini, J. (2018) The UK Biobank resource with deep phenotyping and genomic data. *Nature 562*, 203–209.

(268) Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M. E., Aguilar Salinas, C. A., Ahmad, T., Albert, C. M., Ardissino, D., Atzmon, G., Barnard, J., Beaugerie, L., Benjamin, E. J., Boehnke, M., Bonnycastle, L. L., Bottinger, E. P., Bowden, D. W., Bown, M. J., Chambers, J. C., Chan, J. C., Chasman, D., Cho, J., Chung, M. K., Cohen, B., Correa, A., Dabelea, D., Daly, M. J., Darbar, D., Duggirala, R., Dupuis, J., Ellinor, P. T., Elosua, R., Erdmann, J., Esko, T., Färkkilä, M., Florez, J., Franke, A., Getz, G., Glaser, B., Glatt, S. J., Goldstein, D., Gonzalez, C., Groop, L., Haiman, C., Hanis, C., Harms, M., Hiltunen, M., Holi, M. M., Hultman, C. M., Kallela, M., Kaprio, J., Kathiresan, S., Kim, B.-J., Kim, Y. J., Kirov, G., Kooner, J., Koskinen, S., Krumholz, H. M., Kugathasan, S., Kwak, S. H., Laakso, M., Lehtimäki, T., Loos, R. J. F., Lubitz, S. A., Ma, R. C. W., MacArthur, D. G., Marrugat, J., Mattila, K. M., McCarroll, S., McCarthy, M. I., McGovern, D., McPherson, R., Meigs, J. B., Melander, O., Metspalu, A., Neale, B. M., Nilsson, P. M., O'Donovan, M. C., Ongur, D., Orozco, L., Owen, M. J., Palmer, C. N. A., Palotie, A., Park, K. S., Pato, C., Pulver, A. E., Rahman, N., Remes, A. M., Rioux, J. D., Ripatti, S., Roden, D. M., Saleheen, D., Salomaa, V., Samani, N. J., Scharf, J., Schunkert, H., Shoemaker, M. B., Sklar, P., Soininen, H., Sokol, H., Spector, T., Sullivan, P. F., Suvisaari, J., Tai, E. S., Teo, Y. Y., Tiinamaija, T., Tsuang, M., Turner, D., Tusie-Luna, T., Vartiainen, E., Vawter, M. P., Ware, J. S., Watkins, H., Weersma, R. K., Wessman, M., Wilson, J. G., Xavier, R. J., Neale, B. M., Daly, M. J., MacArthur, D. G., and Consortium, G. A. D. (2020)

The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature 581*, 434–443.

(269) Cheng, J., Novati, G., Pan, J., Bycroft, C., Žemgulytė, A., Applebaum, T., Pritzel, A., Wong, L. H., Zielinski, M., Sargeant, T., Schneider, R. G., Senior, A. W., Jumper, J., Hassabis, D., Kohli, P., and Avsec, Ž. (2024) Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science (80-. ). 381*, eadg7492.

(270) Minton, K. (2023) Predicting variant pathogenicity with AlphaMissense. *Nat. Rev. Genet. 24*, 804.

# Appendix

# Appendix A. Reversible and irreversible inhibitors of coronavirus Nsp15 endoribonuclease.

This chapter was published:

Jerry Chen[‡], Rabih Abou Farraj[‡], Daniel Limonta,[‡], Seyed Amir Tabatabaei Dakhili, Evan M. Kerek, Ashim Bhattacharya7, Filip M. Reformat, Ola M. Mabrouk, Benjamin Brigant, Tom A. Pfeifer, Mark T. McDermott, John R. Ussher, Tom C. Hobman, J. N. Mark Glover, and Basil P. Hubbard. Reversible and irreversible inhibitors of coronavirus Nsp15 endoribonuclease. *Journal of Biological Chemistry*.

[‡]These authors contributed equally.

JBC RESEARCH ARTICLE

# Reversible and irreversible inhibitors of coronavirus Nsp15 endoribonuclease

Received for publication, May 31, 2023, and in revised form, October 4, 2023 Published, Papers in Press, October 11, 2023,
https://doi.org/10.1016/j.jbc.2023.105341

**Jerry Chen**[1,‡], **Rabih Abou Farraj**[2,‡], **Daniel Limonta**[3,4,5,‡], **Seyed Amir Tabatabaei Dakhili**[6], **Evan M. Kerek**[1], **Ashim Bhattacharya**[7], **Filip M. Reformat**[1], **Ola M. Mabrouk**[8], **Benjamin Brigant**[1,9], **Tom A. Pfeifer**[10], **Mark T. McDermott**[8], **John R. Ussher**[6], **Tom C. Hobman**[3], **J. N. Mark Glover**[2], and **Basil P. Hubbard**[1,7,*]

*From the [1]Department of Pharmacology, [2]Department of Biochemistry, and [3]Department of Cell Biology, Li Ka Shing Institute of Virology, University of Alberta, Edmonton, Alberta, Canada; [4]Quantitative Biosciences Institute (QBI), University of California San Francisco, San Francisco, California, USA; [5]Gladstone Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, California, USA; [6]Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, Alberta, Canada; [7]Department of Pharmacology and Toxicology, University of Toronto, Toronto, Ontario, Canada; [8]Department of Chemistry, University of Alberta, Edmonton, Alberta, Canada; [9]Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology, Trondheim, Norway; [10]High Throughput Biology Facility, Life Sciences Institute, University of British Columbia, Vancouver, British Columbia, Canada*

Reviewed by members of the JBC Editorial Board. Edited by Craig Cameron

**The emergence of severe acute respiratory syndrome coronavirus 2, the causative agent of coronavirus disease 2019, has resulted in the largest pandemic in recent history. Current therapeutic strategies to mitigate this disease have focused on the development of vaccines and on drugs that inhibit the viral 3CL protease or RNA-dependent RNA polymerase enzymes. A less-explored and potentially complementary drug target is Nsp15, a uracil-specific RNA endonuclease that shields coronaviruses and other nidoviruses from mammalian innate immune defenses. Here, we perform a high-throughput screen of over 100,000 small molecules to identify Nsp15 inhibitors. We characterize the potency, mechanism, selectivity, and predicted binding mode of five lead compounds. We show that one of these, IPA-3, is an irreversible inhibitor that might act *via* covalent modification of Cys residues within Nsp15. Moreover, we demonstrate that three of these inhibitors (hexachlorophene, IPA-3, and CID5675221) block severe acute respiratory syndrome coronavirus 2 replication in cells at subtoxic doses. This study provides a pipeline for the identification of Nsp15 inhibitors and pinpoints lead compounds for further development against coronavirus disease 2019 and related coronavirus infections.**

Since its appearance in late 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has infected hundreds of millions of individuals globally and killed more than 18 million people (1, 2). Exposure to this virus can lead to respiratory problems, systemic inflammation, and multiorgan dysfunction, which collectively define coronavirus disease 2019 (COVID-19) (1). Coronaviruses are characterized by large single-stranded positive sense RNA genomes of roughly 30 kB and a "corona" structure in their envelope (3–5). SARS-CoV-2 is classified as a lineage B betacoronavirus alongside its most phylogenetically similar relative, SARS-CoV-1 (4). Translation of the SARS-CoV-2 viral genome is performed by host cell ribosomes and generates four structural proteins: an envelope protein involved in viral assembly and budding, a membrane protein involved in defining virion shape, a nucleocapsid protein that packages genomic RNA, and a spike protein involved in host attachment and entry (6, 7). SARS-CoV-2 also contains 16 nonstructural proteins (Nsps) including the 3C-Like main protease (3CLpro/MPro, Nsp5), the papain-like protease (PLPro, Nsp3), and the RNA-dependent RNA polymerase (RdRp) complex (Nsp7, Nsp8, and Nsp12) (5, 8, 9). Finally, nine open reading frames that encode additional accessory proteins have been annotated (5, 8, 9).

Current strategies to combat COVID-19 have focused on the development of prophylactic vaccines mostly directed against the SARS-CoV-2 spike protein (10) and small-molecule drugs targeting the 3CLpro (11), PLPro (12), and RdRp complex (13). Recently, the Food and Drug Administration approved the first two small-molecule drugs, nirmatrelvir/ritonavir and molnupiravir, which target the 3CLpro (14) and RdRp (15) complex, respectively, for therapeutic use. Despite these advances, transmission of SARS-CoV-2 continues in the population, and severe outcomes including death persist (16). This is in part because of the evolution of new viral variants that acquire mutations in genes that encode vaccine-induced antibody targets (16, 17). Furthermore, it has been suggested that current pharmaceutical treatments against COVID-19 could also be rendered ineffective in the future because of resistance (18). Thus, there is a need to broaden the repertoire of coronavirus antiviral treatments to additional targets in the SARS-CoV-2 proteome.

A promising but less explored SARS-CoV-2 drug target is Nsp15, a nidoviral endoribonuclease that selectively cleaves 3′ of uridylates (NendoU) to generate 2′,3′-cyclic phosphodiester and 5′-hydroxyl termini products (19, 20). Nsp15 cleaves both

---

‡ These authors contributed equally to this work.
* For correspondence: Basil P. Hubbard, basil.hubbard@utoronto.ca.

ssRNA and dsRNA but not DNA (20, 21). Structurally, it is comprised of a hexamer formed by the dimerization of two trimers (20). X-ray crystal (20) and cryo-EM structures (22) reveal that each monomeric unit has three domains: an N-terminal domain involved in multimerization, a middle domain, and the C-terminal catalytic core (20). The active site contains residues that are reminiscent of the catalytic triad found in RNase A, including two histidine residues (His235 and His250) that act as general acid–base catalysts and a lysine residue (Lys290) (19, 20). Two other important residues within the active site are Ser294 and Tyr343, which are thought to enforce uracil specificity (20). Unlike RNase A, Nsp15 requires coordinated manganese ions ($Mn^{2+}$) for catalysis (20). These ions are purported to maintain active site conformation and substrate binding (20).

The role of Nsp15 in coronavirus replication and pathogenesis is complex and multifaceted (5). Several studies have suggested an essential role for Nsp15 in viral replication, as mutation of Nsp15 blocks viral RNA synthesis in certain cell types (23) and attenuates disease phenotypes in mice (24). Experimental data suggest that Nsp15 also plays an important role in shielding viruses from innate host cell immunity (24). It has been proposed that Nsp15 facilitates evasion of dsRNA sensors that would normally activate a type I interferon (IFN) response by cleaving viral dsRNA outside the replication complexes (24). This is supported by studies with SARS-CoV-1 showing that Nsp15-defective viruses induce MDA5, PKR, and OAS/RNase L dsRNA sensing pathways (5, 24–26) and by studies with SARS-CoV-2 implicating Nsp15 in suppression of IFNβ production (27). Finally, recent work has shown that Nsp15 cleaves 5′ polyuridines from negative sense viral RNA (PUN RNA) and thereby decreases MDA5-mediated IFN response (28). While its multiple roles in viral biology are still being elucidated, it is apparent that Nsp15 plays a critical role in coronavirus pathogenesis (24) and is therefore an attractive candidate for drug design.

Only a few Nsp15 inhibitors have been experimentally validated to date (5). These include benzopurpurin B, which inhibits RNase A and numerous nidoviral Nsp15 homologs (19), betulonic acid derivatives that inhibit Nsp15 from HCoV-229E but not SARS-CoV-2 viruses (29), and tipiracil, which acts as a competitive inhibitor of Nsp15 (30). The β-amyloid antiaggregation molecule Exebryl-1 also robustly inhibits SARS-CoV-2 Nsp15 *in vitro* but displays weaker potency in cells ($EC_{50}$ of roughly 65 μM in Calu-3 cells) (5). While these and other studies (31, 32) have clearly demonstrated the potential to chemically inhibit Nsp15, small-molecule scaffolds that are potent, bioactive, and nontoxic remain elusive. Here, we perform a high-throughput screen (HTS) of over 100,000 diverse compounds to identify Nsp15 inhibitors that meet these criteria. We refine our initial list of hits by using a workflow that encompasses a series of biochemical and biophysical experiments to eliminate pan-assay interference (PAIN) compounds (33). We then characterize the potency, selectivity against other Nsp15 orthologs, kinetic mechanisms, predicted binding mode, and cellular activity/toxicity of five lead compounds. These include the antimicrobial drug hexachlorophene (34), the PAK1 kinase inhibitor IPA-3 (35), and three proprietary molecules (CID5220994, CID5266986, and CID5675221) from the ChemBridge DIVERSet library. We find that these inhibitors operate through reversible mechanisms, with the exception of IPA-3, which is irreversible and likely covalent. Moreover, we demonstrate that hexachlorophene, IPA-3, and CID5675221 inhibit SARS-CoV-2 replication in Vero CCL-81 cells at subtoxic doses. Overall, this work provides a comprehensive platform for the identification and validation of small-molecule Nsp15 inhibitors and identifies several potent, bioactive, and relatively nontoxic lead compounds with potential for further development into treatments against COVID-19 and related illnesses.

## Results

### *Measurement of SARS-CoV-2 Nsp15 enzyme kinetics*

To measure Nsp15 enzyme kinetics and determine the ideal parameters for a high-throughput compound screen, we adapted a previously described biochemical assay (21). This assay employs a monouridylated ssRNA substrate that is flanked by a 6-carboxyfluorescein (6-FAM) moiety at the 5′ end and a tetramethylrhodamine quencher at the 3′ end (21). Cleavage of the substrate by Nsp15 eliminates FRET between the fluorophore and quencher, resulting in a fluorescent signal (21). For our studies, we substituted the tetramethylrhodamine group with a 3′ black hole quencher (Fig. 1*A*), which has been demonstrated to yield more consistent results in biochemical assays (36). We purified recombinant SARS-CoV-2 Nsp15 from *Escherichia coli* and tested its ability to cleave several ssRNAs of different sequence length and composition alongside positive and negative controls (Fig. S1, *A* and *B*). We identified one sequence, RNA2 (5′-FAM-CAACUAAACGAAC-BHQ1-3′), which yielded low background and robust signal in the presence of Nsp15 (Fig. S1*B*). This substrate was used in subsequent experiments. Next, we verified that the signal being measured in the assay required Nsp15 activity using a catalytic histidine mutant (H250A) and an inactive monomeric truncated protein variant (Fig. 1*B*) (26, 37). In addition, we confirmed that the signal was enhanced by manganese ions (Fig. S1*C*) and abrogated by metal chelation with EDTA (Fig. S1*D*). These experiments validated the ability of our assay to reliably measure Nsp15 activity.

Enzyme and substrate concentrations and reaction time must be carefully chosen to ensure optimal hit identification when performing HTSs (38). To determine the $K_m$ of RNA2, we first measured fluorescence over a time window for reactions containing variable amounts of substrate at a fixed enzyme concentration (Fig. 1*C*). In addition, we serially diluted precleaved (fluorescent) RNA to generate a standard curve relating fluorescence signal to the amount of product produced (Fig. 1*D*). Next, we plotted the slopes (derivatives) of the lines in Figure 1*C versus* RNA substrate concentration and used the standard curve equation (Fig. 1*D*) to transform the *y*-axis to reaction rate (Fig. 1*E*). Fitting this plot to a Michaelis–Menten curve yielded a $K_m$ value of ~3 μM for RNA2 and a $V_{max}$ value

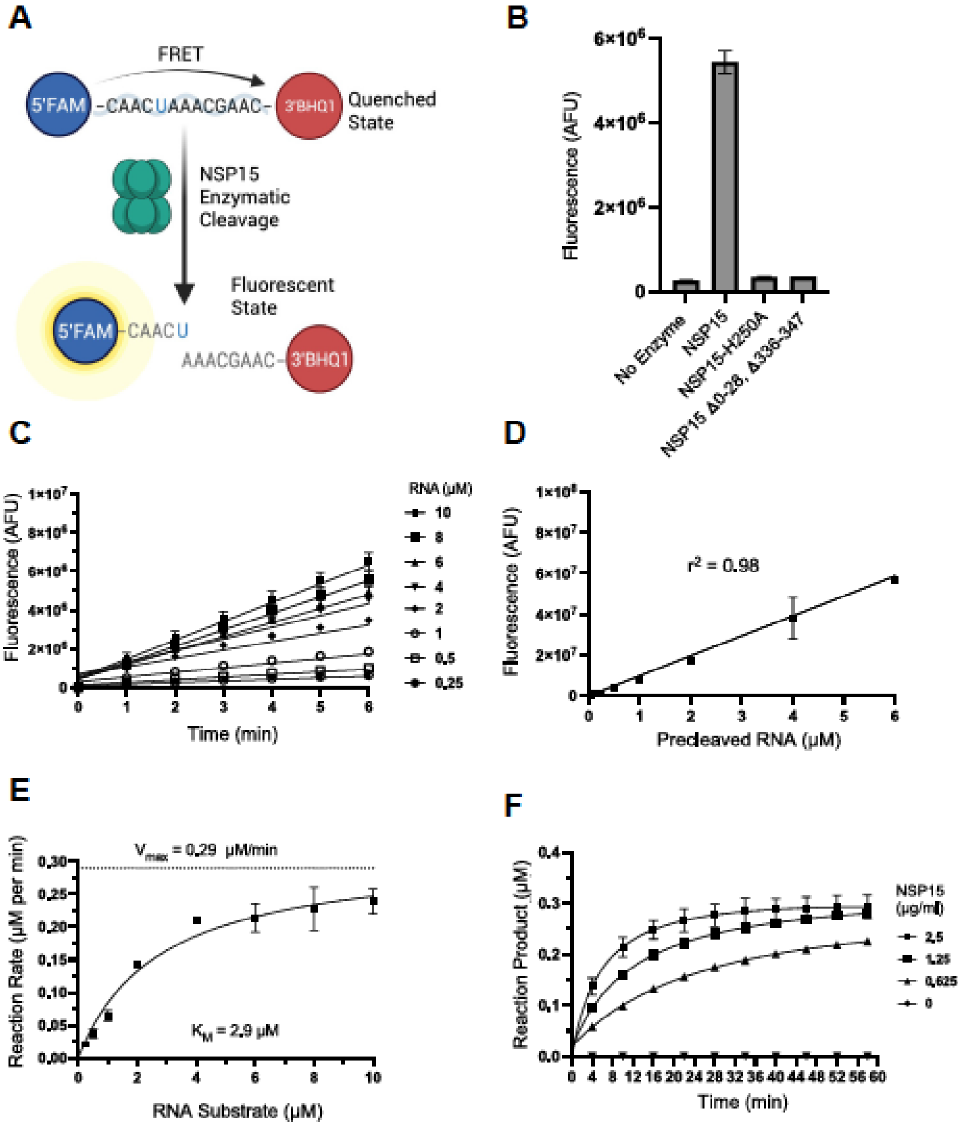


**Figure 1. Characterization of Nsp15 activity using an FRET-based assay.** *A*, schematic of Nsp15 activity assay. An RNA substrate flanked by 5′fluorescein (FAM) and 3′black hole quencher 1 (BHQ1) modifications is cleaved by Nsp15 in the presence of manganese. *B*, bar graph showing the activity of various Nsp15 mutant proteins using the FRET-based assay. Reactions were allowed to proceed for 40 min at 37 °C; mean ± SD is shown (n = 3). *C*, reaction progress over time at various concentrations of RNA substrate; mean ± SD shown (n = 3). Points were fit to linear equations. *D*, standard curve (linear fit) showing the relationship between RNA cleavage (using a positive control RNA lacking the quencher) and fluorescence; mean ± SD is shown (n = 3). *E*, plot showing the relationship between reaction rate and substrate concentration generated using data from *C*. $K_M$ and $V_{max}$ values were determined by fitting data to the Michaelis–Menten equation ($v = V_{max}[S]/(K_M + [S])$) using GraphPad Prism; mean ± SD is shown (n = 3). *F*, reaction progress curves using a fixed concentration of 0.5 µM RNA at different concentrations of Nsp15 enzyme; mean ± SD is shown (n = 3). All experiments were repeated three times with similar results. Nsp, nonstructural protein.

of ~0.3 µM/min, which are in agreement with recently published results (31). Finally, to optimize the enzyme concentration and time for our HTS, we generated reaction progress curves at a fixed substrate concentration (0.5 µM) at various enzyme concentrations. As shown in Figure 1*F*, we found that product cleavage progressed until saturation was reached between ~30 and 60 min, depending on the enzyme concentration used. These data were used to guide the selection of parameters for our fixed end-point high-throughput compound screen.

### *HTS for small-molecule inhibitors of SARS-CoV-2 Nsp15*

The identification of Nsp15 inhibitors was accomplished by carrying out a screen of over 108,000 small molecules, comprised of compounds sourced from Maybridge, Prestwick, Microsource Spectrum, LOPAC-1280, TimTec, and ChemBridge DIVERSet collections. Overlap between compound libraries was <0.1%. After acoustic dispensing of buffer, enzyme, and compound into wells of 384-well plates, an initial read was performed to measure background and account for autofluorescence. Next, substrate was added, and reactions were allowed to proceed for ~20 min before being stopped *via* the addition of 100 mM EDTA, at which time a second measurement was recorded. Finally, a 5′FAM RNA lacking a quencher was added to each reaction, and a third measurement was performed to account for potential signal quenching caused by each compound. Reactions with benzopurpurin B (inhibitor) and reactions without enzyme were used as positive controls on each plate (19), whereas reactions in the presence of 0.2% dimethyl sulfoxide (DMSO) were used as a negative control (no inhibition). The signal/noise ratio of accepted plates was >5 with an average Z′ score of 0.47 ± 0.11 (standard deviation). As this was a challenging assay with respect to read timing and reagent additions, plates with a Z′ score >0.3 were accepted as complete since the high signal/noise ratio enabled reasonably good distinction of actives. Plots summarizing the results of the screen are shown in Figure 2*A* and Fig. S2*A*. After filtering out compounds displaying high autofluorescence (read 1) or quenching (read 3–read 2), we identified 1280 reactions with fluorescence values >2 standard deviations from the mean. The corresponding

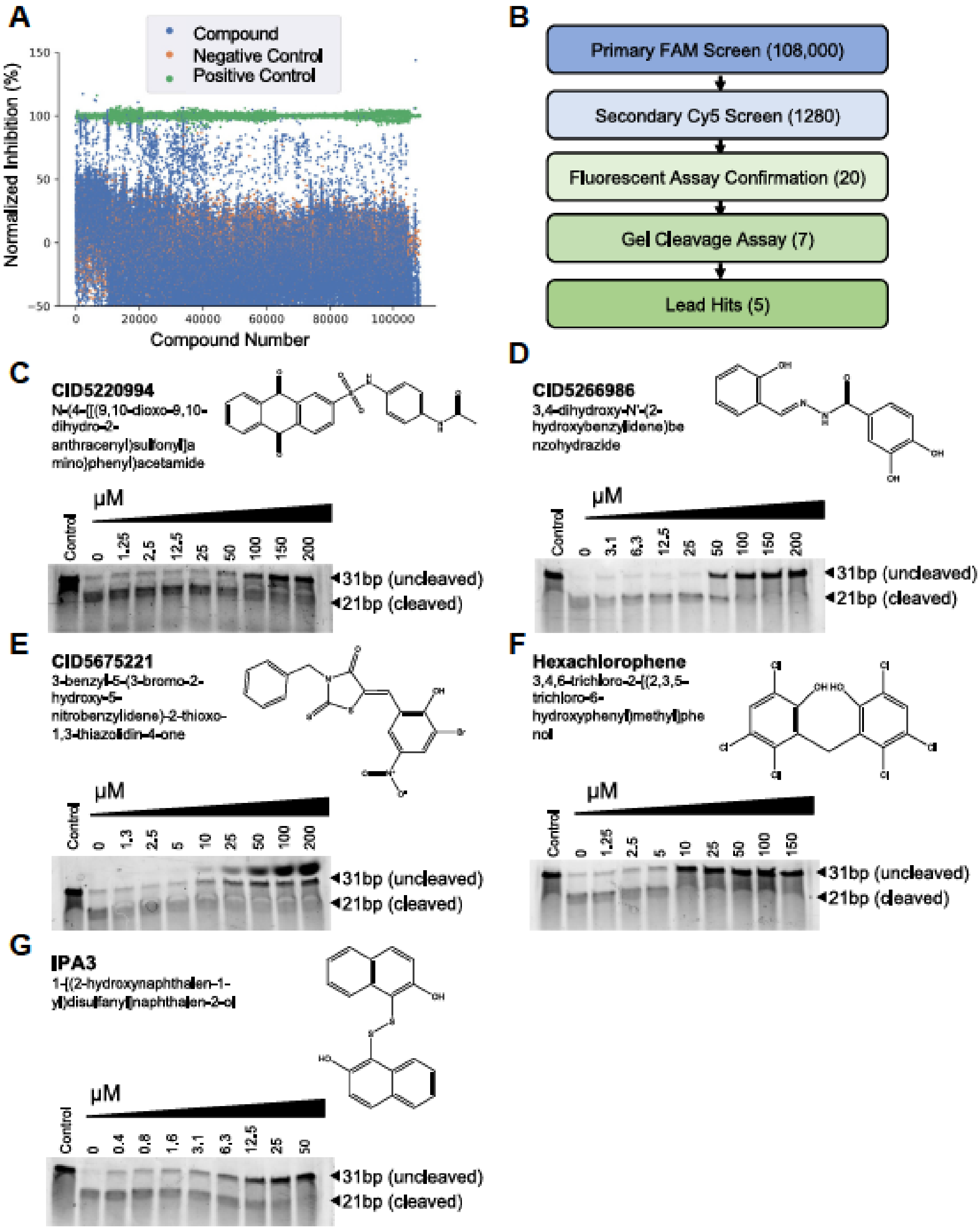


**Figure 2. Identification and validation of SARS-CoV-2 Nsp15 inhibitors.** *A*, plot summarizing the results of the primary ~108,000 compound high-throughput screen using the FAM-BHQ1 RNA substrate. DMSO was used as a negative control, and reactions without enzyme, or with enzyme in the presence of benzopurpurin B, were used as positive controls. *B*, schematic outlining the pipeline used to prioritize hits. Gels showing the results of Nsp15 native RNA cleavage assays in the presence of increasing concentrations of (*C*) CID5220994, (*D*) CID5266986, (*E*) CID5675221, (*F*) hexachlorophene, or (*G*) IPA-3. Nsp15 (7.5 ng/μl) and RNA substrate (25 ng/μl) were incubated with compounds and a 31-nt single "rU"-containing RNA substrate for 1 h at 37 °C. Bands representing the uncleaved substrate and the 21-nt cleaved product were visualized by performing denaturing gel electrophoresis followed by SYBR Gold staining. Representative gels are shown; experiments were repeated three times with similar results. BHQ1, black hole quencher 1; DMSO, dimethyl sulfoxide; FAM, carboxyfluorescein; Nsp, nonstructural protein; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

compounds from these reactions were selected for further inspection, as outlined in the workflow in Figure 2*B*.

We performed a secondary screen of our initial hits using a substrate in which 5′FAM was replaced with 5′Cy5 (a red-shifted dye) to eliminate compounds yielding artifactual inhibition because of interaction with the 5′FAM group. The results of this experiment are shown in Fig. S2, *B* and *C*. Reactions without enzyme and reactions with enzyme and DMSO only were used as positive and negative controls, respectively. The 20 most potent inhibitors identified in the Cy5 screen, taken from the pool of hits greater than three standard deviations away from the mean, were purchased or resynthesized and then individually tested for Nsp15 inhibition and fluorescence quenching using both the 5′FAM and 5′Cy5-conjugated substrates. The results of this analysis, shown in Fig. S3, identified eight compounds (CID5220994, CID5266986, CID5675221, CID5326429, hexachlorophene, IPA3, β-lapachone, Reactive Blue 2) that inhibited Nsp15 using either substrate, without appreciable quenching.

### Validation of HTS hits

One of our hits, β-lapachone, had previously been identified in an Nsp15 inhibitor screen (5) and was shown to induce nonspecific enzyme inhibition through production of reactive oxygen species. We confirmed this finding using an Amplex Red assay, which measures hydrogen peroxide formation caused by reducing agents such as DTT undergoing redox cycling in the presence of oxygen or certain compounds (Fig. S4) (39). Consequently, β-lapachone was excluded from further analysis.

Interactions with the substrate or artificial chemical moieties on the substrate can be a source of assay interference (40). To ensure that our compounds could inhibit Nsp15 activity on native substrates, we tested the remaining seven compounds in a PAGE-based RNA cleavage assay. Briefly, this assay employed a 31-nt poly(A) ssRNA substrate containing a single uridylate, which when cleaved by Nsp15 generates RNA fragments of 10-nt and 21-nt in length. We set up cleavage reactions with Nsp15 and native substrates in the absence or the presence of 25 μM or 50 μM of each compound and analyzed the reaction products on a SYBR Gold-stained gel. In our initial assay shown in Fig. S5, three of seven compounds displayed robust inhibitory activity, whereas the others displayed weaker effects. Inhibition of Nsp15 by CID5326429 was modest and sporadic, and Reactive Blue 2 produced a fluorescent artifact, resulting in these compounds being deprioritized. Titrations of the remaining five compounds, CID5220994, CID5266986, CID5675221, hexachlorophene, and IPA-3, using the native substrate assay revealed that all five displayed dose-dependent effects (Fig. 2, *C*–*G*).

PAIN compounds are nuisance molecules that commonly come up as hits in HTSs because of their ability to interfere with common biochemical assays or inactivate enzymes in a promiscuous manner (33). We performed a battery of tests to evaluate if any of our hits displayed these properties. First, we investigated if our lead hits could be forming colloidal aggregates that block Nsp15 activity by carrying out reactions in the presence of a nonionic detergent, to see if inhibition would be relieved (41). As shown in Fig. S6, there was an increase in Nsp15 activity with CID5266986, CID5675221, hexachlorophene, and IPA-3 and a higher increase in activity with CID5220994, when reactions were performed in the presence of either 0.01% Triton-X or CHAPS. However, we also observed a roughly twofold increase in the baseline activity of Nsp15 in the presence of detergents, which accounted for some of the increased enzyme activity observed with the inhibitors. This increase in enzyme activity in response to detergents has previously been documented and is speculated to occur because of reduced adhesion of active enzyme to plastic surfaces in wells (42). Second, we tested if any of our compounds inhibited Nsp15 by inducing protein aggregation and/or denaturation. For this purpose, a dynamic light scattering (DLS) assay that estimates Nsp15 protein particle diameter based on the rate at which scattered light fluctuates in the solution was performed, in the absence or the presence of each of our hits (43). The results of this analysis yielded a single sharp peak on the intensity size distribution plot for all samples, with polydispersity index values ranging from 0.06 to 0.09 and consistent particle diameters ranging from 12.2 to 13.0 nm. These data reflect a monodisperse and homogenous sample (44) and indicate that protein aggregation was not induced by the compounds at the tested concentrations (Fig. S7). Finally, to rule out general RNA–compound interactions, we performed fluorescence polarization assays with an RNA probe in the presence of each of the five compounds. As shown in Fig. S8, we did not detect any appreciable binding between the compounds and the RNA substrate. Collectively, these data imply that CID5220994, CID5266986, CID5675221, hexachlorophene, and IPA-3 inhibit Nsp15 activity *via* nontrivial mechanisms.

### Characterization of lead compounds

To characterize our validated hits, we performed detailed dose–response titrations to quantify the concentration at which each of them inhibited Nsp15 activity by 50% ($IC_{50}$). As shown in Figure 3, CID5220994 and CID5266986 had relatively high $IC_{50}$ values (~50 and 78 μM), whereas CID5675221 was a stronger inhibitor ($IC_{50}$ of ~20 μM), and hexachlorophene and IPA-3 were the most potent ($IC_{50s}$ of ~1 μM and ~7 μM, respectively). These values were generally consistent with the results of the semiquantitative native substrate assay (Figs. 2 and 3). The fact that CID567221 appears less potent in the gel-based assay than the fluorescent assay can be explained on the basis that it exhibits some quenching in the 5′ FAM channel (Fig. S3). Next, we performed titrations of RNA substrate in the presence of fixed concentrations of each inhibitor to gain insight into their kinetic mechanisms. Modeling these results using steady-state enzyme kinetics (38), we found that CID5266986, CID5675221, and hexachlorophene displayed mixed mechanisms of inhibition (increased $K_m$ and decreased $V_{max}$), whereas CID5220994 appeared to be competitive with substrate (increased $K_m$ only), and IPA-3 was noncompetitive

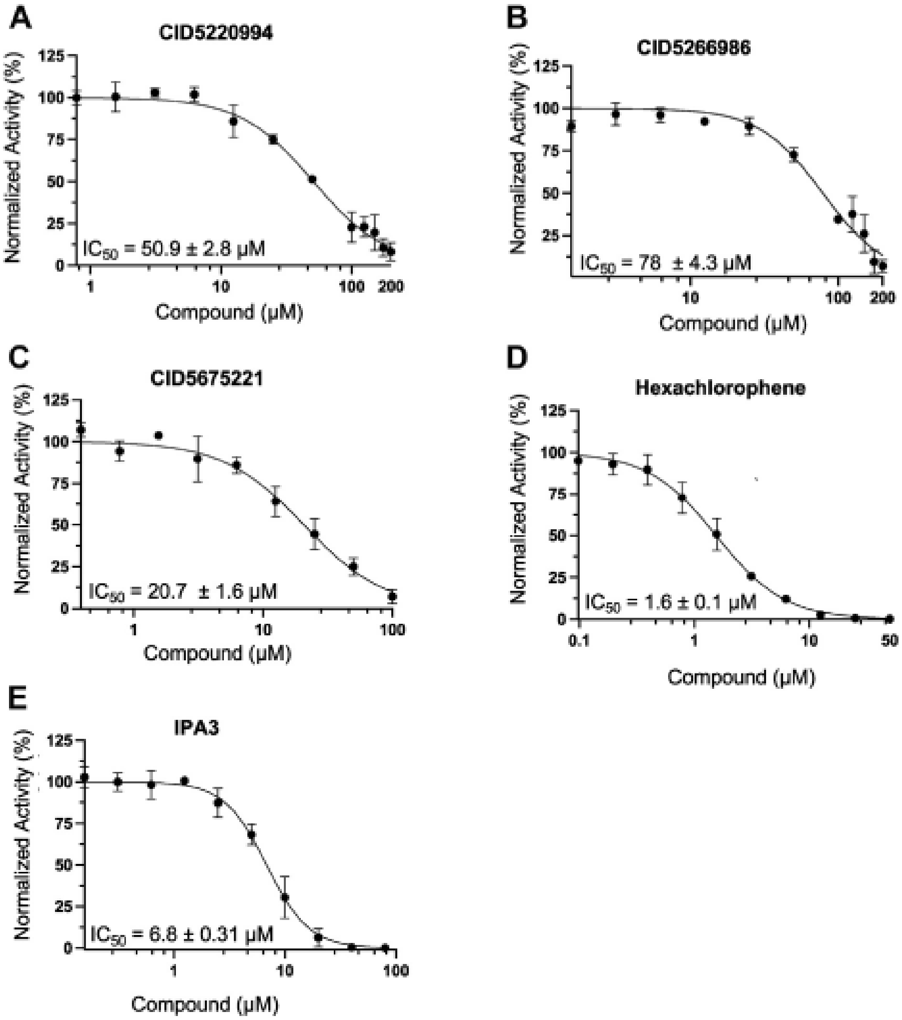


**Figure 3. IC$_{50}$ determination of lead compounds using the FRET-based Nsp15 activity assay.** Plots showing Nsp15 activity in the presence of increasing doses of (*A*) CID5220994, (*B*) CID5266986, (*C*) CID5675221, (*D*) hexachlorophene, or (*E*) IPA-3. Nsp15 (1 ng/µl) and RNA substrate (25 ng/µl) were incubated with compounds for ~12 min at 37 °C. Values were normalized to activity in the absence of inhibitor. IC$_{50}$ values were calculated using [inhibitor] *versus* normalized response variable slope curves generated by GraphPad Prism ($y = 100/(1 + (\mathrm{IC}_{50}/[\mathrm{I}])^{\mathrm{HillSlope}})$); mean ± SD of replicates is shown (n = 3). All experiments were repeated three times with similar results. Nsp, nonstructural protein.

(decreased $V_{\max}$ only, Fig. S9). Calculated $K_i$ values for the inhibitors ranged from ~1 to 73 µM (Fig. S9*F*). Finally, to establish if these inhibitors were reversible or irreversible, we obtained reaction progress curves following release of preincubated enzyme–inhibitor complexes by dilution (Fig. 4*A*). These data revealed that Nsp15 activity was restored once concentrations of CID5220994, CID5266986, CID5675221, and hexachlorophene were dramatically lowered, whereas IPA-3 inhibition was irreversible.

Previous studies have characterized IPA-3 as a selective, allosteric, and irreversible inhibitor of the Pak1 kinase (35, 45). IPA-3 contains a central disulfide bond (Fig. 2*G*), and while it does not form mixed disulfides with surface-exposed cysteines on Pak1, it covalently modifies cysteines within its regulatory domain and blocks binding to its upstream activator, Cdc42 (45). To test if the IPA-3 might be inhibiting Nsp15 *via* its reactive disulfide group, we first performed Nsp15 cleavage reactions in the presence of standard (1 mM) and high (10 mM) concentrations of DTT reducing agent. Previous work showed that DTT concentrations greater than 1 mM relieved Pak1 inhibition by IPA-3 (35). In line with these results, we found that excess DTT nearly completely reversed inhibition by IPA-3, implicating thiol groups in its mechanism of inhibition (Fig. S10).

We used TOF mass spectrometry to further characterize the interaction. We found that preincubation of Nsp15 with IPA-3 resulted in a ~333 Da peak shift *versus* the apoprotein (Figs. 4*B* and S11), corresponding roughly to the molecular weight of one IPA-3 molecule (350 Da). Next, we attempted to map the location of the modified residues using LC–MS/MS. There are a total of five Cys residues within Nsp15, and several of these are known to influence Nsp15 activity (46). While we were unable to detect any modified peptides, we did note the absence of a particular peptide species, corresponding to amino acids 291 to 308 in Nsp15, in the IPA-3-treated sample *versus* the control (Fig. 4*C*). Importantly, this peptide contained two cysteines, Cys291 and Cys293, that were recently shown to form adducts with β-mercaptoethanol (47). Given

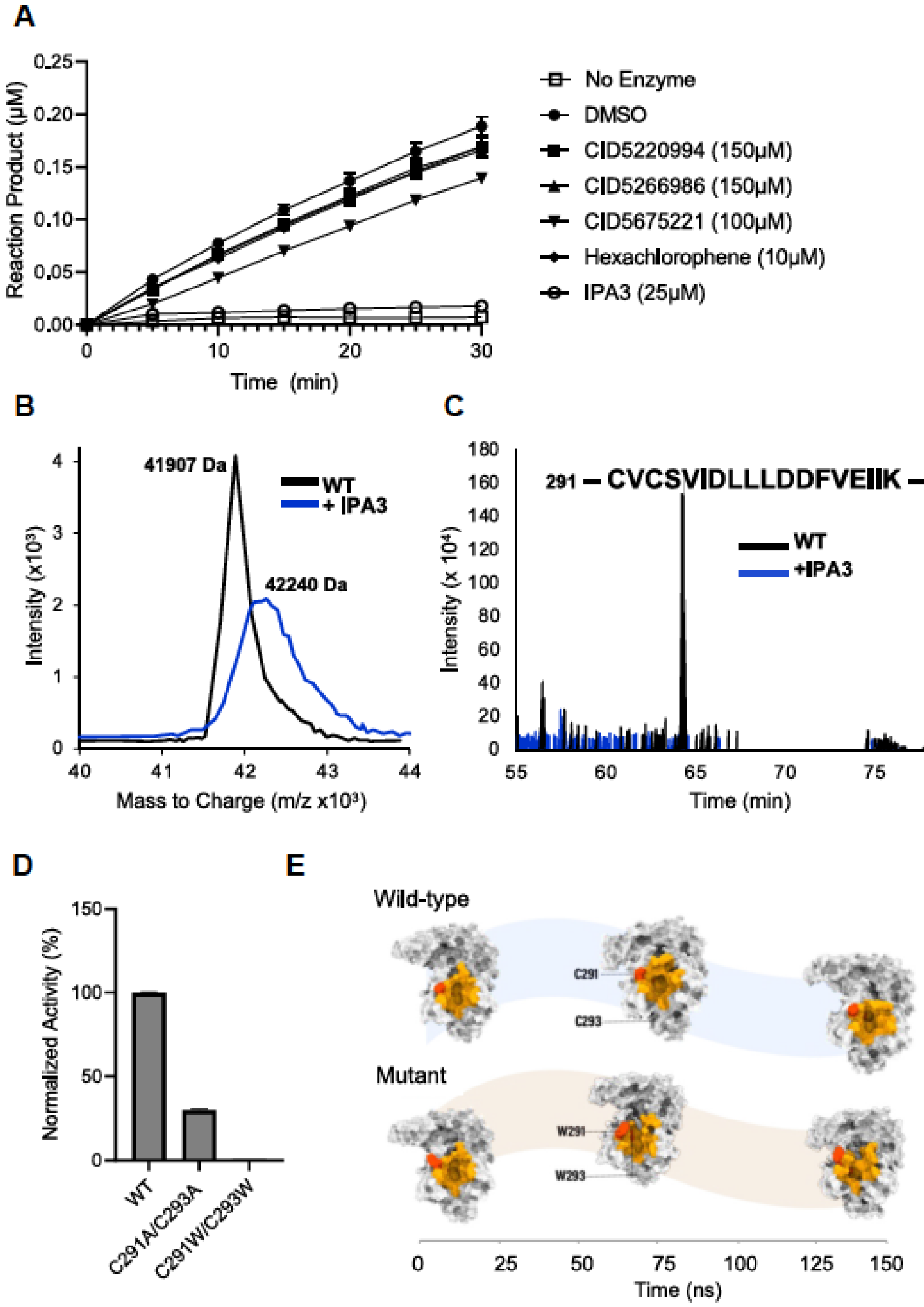


**Figure 4. Assessment of inhibitor reversibility and characterization of covalent modifications imparted by IPA-3.** *A*, plot showing Nsp15 activity over time following dilution of inhibitor. Enzyme (1 µg) was preincubated with the indicated concentrations of inhibitors in a cold room for 60 min and subsequently diluted 1:50 into reaction buffer containing 1 µM substrate, at which time fluorescence measurements were begun ($t = 0$). Measurements were taken every 5 min for a total of 30 min. Activity values were transformed to micromolar reaction product using a standard curve. Mean ± SD is shown (n = 3). Points were fit to linear equations. *B*, plot showing intensity *versus* mass to charge ratio distribution for untreated Nsp15 protein (*black*) or Nsp15 protein preincubated with IPA-3 (*blue*), as determined by MALDI-TOF mass spectrometry. The calculated mass is indicated at the top of each central peak. This experiment was repeated three times with similar results. *C*, plot showing intensity *versus* time, determined by LC–MS/MS, for peptides derived from untreated (*black*) or IPA-3-treated (*blue*) Nsp15 protein. The sequence of the peptide corresponding to amino acids 291 to 308 of Nsp15, which was absent in the treated sample, is indicated. *D*, comparison of the activity of wildtype Nsp15 with Nsp15-C291A/C293A and Nsp15-C291W/C293W mutant proteins using the FRET-based assay. Nsp15 (1 ng/µl) was incubated with 1 µM substrate, and reactions were allowed to proceed for 15 min. Activity was normalized to the wildtype protein. Mean ± SD is shown (n = 3). This experiment was repeated three times with similar results. *E*, diagram showing the results of an MD simulation comparing the active site of wildtype Nsp15 protein to that of a C291W/C293W-Nsp15 mutant protein; duration is indicated on the *x*-axis. The active site is colored in *orange*, and residues being substituted are shown in *red*. MD, molecular dynamics; Nsp, nonstructural protein.

this, we speculated that Cys291 and Cys293 might be undergoing split IPA-3 modification (half of the structure), which would also explain the corresponding peak shift. Moreover, as these residues are adjacent to Lys290 and Ser294, two key active site residues involved in catalysis (19, 20), we investigated what effects their modification might have on Nsp15 enzyme activity. We generated C291A/C293A and C291W/C293W mutant proteins and found that substitution of these Cys residues with Ala reduced enzyme activity, and that replacement with Trp residues, which sterically mimic the bulky diphenyl of the split IPA-3 modification, inactivated the enzyme (Fig. 4*D*). We analyzed the mutant proteins using size-exclusion chromatography. As shown in Fig. S12, we found that the wildtype protein and both mutants yielded single peaks at similar elution volumes (48). While global changes in structure and hexamerization cannot be fully ruled out as a possible explanation for the loss of activity of the mutants, molecular dynamics (MD) simulations comparing the structure of wildtype Nsp15 (20) to that of a computationally modeled C291W/C293W mutant protein revealed that the bulky residues likely cause a distortion of the active site. This could lead to impaired substrate binding or improper positioning within the catalytic core (Fig. 4*E*). Together, these data indicate that IPA-3 inhibits Nsp15 activity irreversibly through covalent modification of Cys residues, and that Cys291 and Cys293 are likely implicated in the mechanism of inhibition.

We attempted to cocrystallize the reversible inhibitors with Nsp15 to gain deeper insight into their mechanisms. While we were able to generate a novel structure of an Nsp15-H250A catalytic mutant protein (26) (Figs. 1*B* and S13), we were unable to obtain structures for the Nsp15–inhibitor complexes. Therefore, we used molecular modeling and MD simulation studies as an alternative strategy. In addition to the previously reported active site (20), we identified a deeper allosteric binding site on the surface of the enzyme that could mediate interactions with the noncompetitive inhibitors (Fig. 5*A*). Several residues within this pocket including Tyr279 have previously been implicated in chemical interactions with Exebryl-1 (5) and other Nsp15 inhibitors (32). Initially, we docked the four reversible inhibitors into both the active and allosteric sites in order to determine their differential affinity for each site. Subsequently, Nsp15–inhibitor complexes were subjected to a 150 ns MD simulation in order to investigate the binding dynamics and free energies of each inhibitor in both active and allosteric sites. In agreement with our experimental data, the MMPBSA binding affinity of the competitive inhibitor CID5220994 was higher in the active site compared with the allosteric site (Fig. 5*B*). Interestingly, the mixed inhibitors (CID5266986 and CID5675221) exhibited affinity for both binding sites, although the affinity for the allosteric site was significantly higher (Fig. 5*B*). Unfortunately, we were unable to model hexachlorophene binding because of its chlorine atoms interfering with the total simulation system charges. Binding modes of CID5675221 into the active site and allosteric site are shown in Figure 5, *C* and *D*, with the resulting active site distortion highlighted in Figure 5*E*. Depictions of the binding modes of CID5220994 and CID5266986 are presented in Fig. S14. Collectively, these data provide a theoretical basis for the kinetic mechanisms of the lead compounds and provide structural insight into how these molecules may be inhibiting Nsp15 activity.

### *Assay of ability of lead compounds to suppress SARS-CoV-2 replication in cells*

To translate our findings to cells, we first evaluated the toxicity of the five lead compounds using a luminescent ATP-based cell viability assay (CellTiter-Glo) in Vero CCL-81 cells. These are monkey kidney cells that have previously been used as a model for SARS-CoV-2 infection (49). CID5675221 had a very high $CC_{50}$ value (>240 μM), whereas CID5220994, CID5266986, hexachlorophene, and IPA-3 were moderately toxic to cells at higher doses, with $CC_{50}$ values ranging from ~15 to 42 μM (Fig. S15). Next, we used a plaque assay to test the ability of these compounds to inhibit production of infectious SARS-CoV-2 virions at subtoxic concentrations; remdesivir was used as a positive control. As shown in Figure 6*A*, CID5675221, hexachlorophene, IPA-3, and the positive control significantly reduced viral titers, whereas the effects of CID5220994 and CID5266986 were not statistically different from the control.

To establish the approximate selectivity indexes for the three bioactive compounds, we performed dose titrations and measured their inhibitory effects on viral replication and cell viability in parallel. Cellular $IC_{50}$ values were ~20 μM, ~1 μM, and 10 μM for CID5675221, hexachlorophene, and IPA-3, respectively (Fig. 6, *B–D*), mirroring our *in vitro* results (Fig. 3, *C–E*). Using these data in combination with the cytotoxicity measurements (Fig. 6, *B–D*), we determined the selectivity index of the small molecules to be CID5675221 ~15, hexachlorophene ~16, and IPA-3 ~5. Finally, as complementary assays for assessing virus replication and infection, respectively, we measured intracellular SARS-CoV-2 RNA using quantitative RT–PCR and assessed levels of the viral spike protein in cells using immunofluorescence in the absence or the presence of each compound. We found that all three compounds significantly reduced viral RNA in a dose-dependent manner (Fig. 6, *E–G*) and decreased the presence of viral spike proteins in Vero CCL-81 cells (Fig. S16). These findings establish several new Nsp15 inhibitors that actively inhibit SARS-CoV-2 replication in cells at nontoxic doses.

### *Selectivity of inhibitors against Nsp15 homologs and unrelated proteins*

Nsp15 from SARS-CoV-2 is highly similar to homologs from SARS-CoV-1 and Middle East respiratory syndrome (MERS) (Fig. S17), displaying amino acid identity of 88% and 51%, respectively (20). Based on this, we wondered if our lead inhibitors might also be effective in inhibiting these related enzymes. We tested the ability of the five lead compounds to inhibit the Nsp15 homologs from these viruses at two doses corresponding roughly to the SARS-CoV-2 $IC_{50}$ and $IC_{100}$ values. In agreement with our hypothesis, we found that the effects of the compounds on Nsp15 from SARS-CoV-1 were

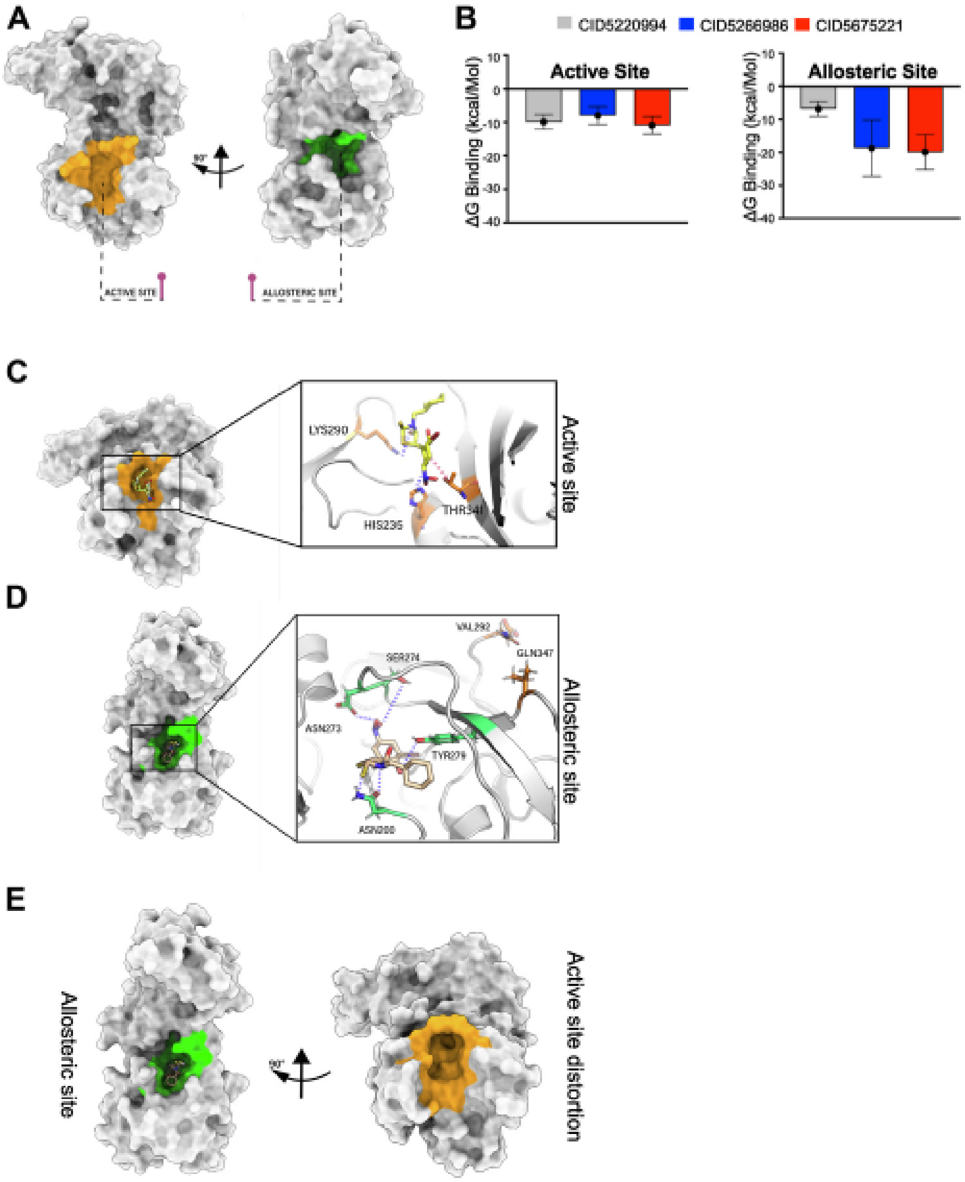


**Figure 5. Binding site analysis and molecular modeling of several lead Nsp15 inhibitors.** *A*, protein structures showing the Nsp15 catalytic (active) site colored in *orange* and the identified allosteric site colored in *green*. *B*, free energy of binding for select inhibitors in the active or allosteric pockets. Binding mode of the mixed inhibitor CID5675221 inside the (*C*) catalytic pocket or (*D*) allosteric pocket of Nsp15 after 150 ns of MD simulation. Hydrogen bonds and hydrophobic contacts are shown using *dashed blue* and *red lines*, respectively. *E*, diagram highlighting the active site distortion caused by binding of CID5675221 into the allosteric site. MD, molecular dynamics; Nsp, nonstructural protein.

virtually equivalent to those observed with the SARS-CoV-2 enzyme, with the exception of hexachlorophene, which was less potent (Fig. S18*A*). We also observed some degree of inhibition of MERS Nsp15 by all lead compounds, albeit with reduced potency. This was especially evident for CID5220994 and CID5266986, and to some extent IPA-3, which displayed only modest effects (Fig. S18*B*). We next tested if these compounds could inhibit the activity of RNAse A, a more distantly related RNA endonuclease that uses a similar catalytic mechanism. As shown in Fig. S18*C*, we did not observe any inhibition of this enzyme by the compounds. Finally, we tested the effects of the lead compounds on a completely unrelated enzyme, the NAD-dependent lysine deacetylase SIRT1 (40). As anticipated, none of the compounds inhibited SIRT1 enzymatic activity (Fig. S18*D*). In sum, these data show that CID5220994, CID5266986, CID5675221, hexachlorophene, and IPA-3 are able to selectively inhibit Nsp15 homologs from multiple coronaviruses without altering the activity of other RNA endonucleases and unrelated enzymes.

## Discussion

Outbreaks of SARS-CoV-2 around the world continue, and there remains a need to develop new therapeutics to ensure that treatments against evolved variants are available (1). From a screen of over 100,000 small molecules, we identified five promising lead compounds that inhibit Nsp15 activity in a native substrate assay *in vitro* (Fig. 2). Three of these, hexachlorophene, IPA-3, and CID5675221, were effective in blocking SARS-CoV-2 viral replication in cells (Fig. 6). The

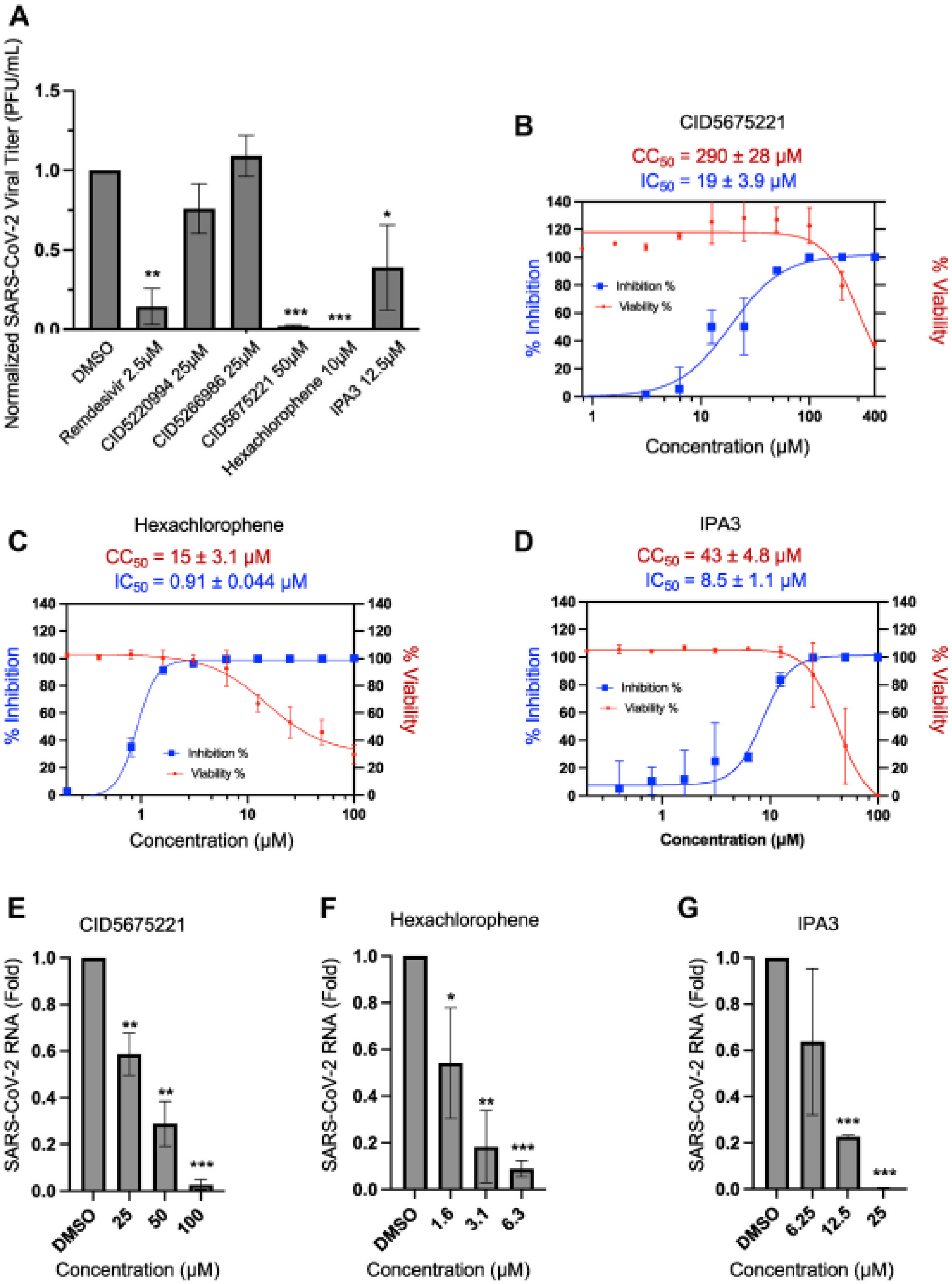


**Figure 6. Determination of anti-SARS-CoV-2 activity of Nsp15 inhibitors in cells.** *A*, bar graph showing SARS-CoV-2 replication, as determined by viral plaque assays, in the absence or the presence of Nsp15 inhibitors. Vero CCL-81 cells were infected with SARS-CoV-2 (MOI = 0.1) and treated with a single concentration of remdesivir (positive control), CID5220994, CID5266986, CID5675221, hexachlorophene, IPA-3, or vehicle (DMSO). Titers were determined by plaque assay of supernatants collected 24 h postinfection and normalized to the DMSO control. Mean ± SD is shown (n = 2). Plots showing the effect of increasing concentrations of (*B*) CID5675221, (*C*) hexachlorophene, or (*D*) IPA-3 on viral inhibition (*blue curve*), and cell viability (*red curve*). The % inhibition and % viability were calculated relative to a DMSO control sample. Viral titers were determined by plaque assay following a 24 h infection of Vero CCL-81 cells infected with SARS-CoV-2 (MOI = 0.1) in the absence or the presence of compounds. Cell viability was assessed at the same time point by measuring the intracellular ATP levels. $IC_{50}$ and $CC_{50}$ values are indicated above the curves and were calculated by fitting points to a four-parameter logistic curve ($y$ = bottom + ($[I]^{HillSlope}$) * (top − bottom)/($[I]^{HillSlope}$ + $IC_{50}^{HillSlope}$)) using GraphPad Prism. Mean ± SD of two biological replicates is shown. Effect of the indicated doses of (*E*) CID5675221, (*F*) hexachlorophene, or (*G*) IPA-3 on levels of SARS-CoV-2 viral RNA in Vero CCL-81 cells as determined by qRT–PCR from total cellular RNA following a 24 h infection with SARS-CoV-2 (MOI = 0.1). Mean ± SD (n = 2) is shown. * denotes $p < 0.05$, ** denotes $p < 0.01$, and *** denotes $p < 0.001$ significance using a one-tailed Student's *t* test *versus* the DMSO control sample. DMSO, dimethyl sulfoxide; MOI, multiplicity of infection; Nsp, nonstructural protein; qRT–PCR, quantitative RT–PCR; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

results of this study add to the growing body of evidence (5, 32) that suggests fully optimized Nsp15 inhibitors could 1 day be employed as alternative or complementary therapeutics to current state-of-the-art COVID-19 drugs.

Lipinski's rule of 5 is a classic predictor of how suitable a molecule may be as a drug (50). This theory states that a drug-like molecule should (1) have no more than five hydrogen bonds, (2) have no more than ten hydrogen bond acceptors, (3) have a molecular weight of <500 Da, and (4) have a lipophilicity (LogP) value of <5 (50). As shown in Fig. S19, our five lead compounds fulfilled all these criteria, with the exception of hexachlorophene, which had a larger LogP value of 7.25. However, our empirical data identified several properties of these first-generation inhibitors that would need to be revamped to realize their therapeutic potential. For example, both CID5220994 and CID5266986, the two weakest inhibitors, caused toxicity in cells at concentrations close to their *in vitro* $IC_{50}$ values (Figs. 3, 6 and S15). Interestingly, CID5266986 bears striking structural similarity to Exebryl-1, a compound currently in clinical trials for the treatment of Alzheimer's disease that was recently characterized as a mixed inhibitor of Nsp15 (5).

Hexachlorophene (Nabac) has a long history of use as a biocide in toothpaste, soaps, and topical treatments, although it has recently been removed from many nonprescription products because of concerns over neurotoxicity (51). Its antimicrobial activity results from its ability to inhibit the membrane-bound part of the electron transport chain, leading to inhibition of respiration and cell leakage (52). In addition, it inhibits a broad range of different enzymes, including phosphatases such as SHP2 (53), adenylyl cyclases (54), and succinate dehydrogenase (55). Our data show that hexachlorophene also potently inhibits the Nsp15 enzyme and blocks SARS-CoV-2 viral replication in cells with an $IC_{50}$ of ∼1 to 2 μM (Figs. 3 and 6). While we are the first to report inhibition of Nsp15 by this compound, several past studies have implicated hexachlorophene in related antiviral activity. For example, a previous publication showed that hexachlorophene abrogated SARS-CoV-2 replication in Vero cells (56). Moreover, a separate publication characterized hexachlorophene as a competitive inhibitor of the 3CL protease from SARS-CoV-1 (57). While its promiscuity, toxicity, and chemical properties preclude it from being used systemically, our findings suggest that hexachlorophene could be a promising additive to topical formulations aimed at reducing the spread of COVID-19 *via* hand and skin sanitization. For example, the prescription skin cleaner pHisoHex contains a 3% w/w emulsion of hexachlorophene, which corresponds to a concentration of 73.7 mM (58). While topical agents using this dose carry some risk because of absorption and toxicity (51), our results indicate that a dose ∼10,000 times lower (Fig. 6) would be sufficient to block SARS-CoV-2 replication.

SARS-CoV-2 Nsp15 contains five cysteine residues (Cys103, Cys117, Cys291, Cys293, and Cys334) that have been implicated in subunit oligomerization (59) and interactions with the RNA substrate (47). Our finding that IPA-3 inhibits Nsp15 activity through potential covalent modification of cysteines is supported by previous work demonstrating the same mode of action against the Pak1 kinase (60). While there is a ∼16 to 18 Da discrepancy between the sum of the masses of Nsp15 and IPA-3 (42,257 Da) and the experimentally determined mass for the main Nsp15-IPA-3 adduct (42,240 Da), this could be explained by the loss of one hydroxyl group on IPA-3, alone, or in combination with the loss of reduced cysteine hydrogens on the protein during covalent reaction. However, further investigation is needed to elucidate the exact chemistry of the reaction. Our assertion that Cys291 and Cys293 (Fig. 4) play a crucial role in this process is supported by a previous cryo-EM study that designated these residues as reactive (47). Moreover, modeling from this study indicated that Cys291 may participate in substrate interactions in the postcleavage state of the enzyme (47). Our independent docking and simulation studies show that modification of these residues by bulky aromatic groups, such as those found within IPA-3, causes a distortion in the active site that may impair substrate positioning and/or release (Fig. 4*E*). This finding was confirmed experimentally by our mutation studies showing that substitution of these residues with tryptophan results in a loss of enzyme activity (Fig. 4*D*). Finally, the fact that IPA-3 inhibits Nsp15 from SARS-CoV-1 and SARS-CoV-2 with equal potency, but shows reduced effect on the MERS homolog (Fig. S18), could be explained on the basis of the conservation of both Cys291/Cys293 amongst the SARS viruses but only Cys293 in MERS (Fig. S17).

IPA-3 has previously been characterized as an isoform-selective non-ATP competitive inhibitor of the p21-activated kinase Pak1, and it is currently being investigated as a potential anticancer drug (35, 60). While initial reports showed that IPA-3 did not react with surface-exposed cysteines on Pak1 (35), later reports demonstrated its mechanism of inhibition to be covalent modification of the protein regulatory domain (45). Covalent inhibitors bind to targets in two distinct steps: (1) equilibrium bond formation or reversible interaction and (2) covalent bond formation or irreversible interaction (61). Despite previously being classified as PAIN molecules (62), modern covalent inhibitors are emerging with advantages over traditional mechanistic inhibitors, including increased potency, the ability to target shallow binding sites, and defense against drug resistance (61). In fact, the recently approved COVID-19 drug paxlovid acts through reversible covalent inhibition of the 3CL protease (63). Although IPA-3 inhibits multiple enzymes and is not specific for Nsp15 (35), medicinal chemistry work could be performed to refine its binding selectivity. Moreover, the novel mechanism of covalent inhibition of Nsp15 involving Cys291/293 discovered in this study could be exploited to develop new rationally designed Nsp15 inhibitors.

CID5675221 is a novel compound based on a rhodanine scaffold, which is found in drugs such as the aldose reductase inhibitor epalrestat (64). This small molecule demonstrated efficacy against SARS-CoV-2 with an $IC_{50}$ of <20 μM and was the least toxic of any lead hit in our study (Figs. 6 and S15). Our kinetic experiments revealed that this compound had an apparent mixed mode of inhibition (Fig. S9). However, our structural studies suggest a primarily noncompetitive mechanism of inhibition for CID5675221, stemming from the formation of multiple hydrogen bonds with residues located in the N-terminal allosteric pocket, including Tyr279, Asn200, Ser274, and Asn273 (Fig. 5, *D* and *E*). We observed that binding of CID5675221 to the allosteric site distorts Gln347, an essential residue of the Nsp15 catalytic site (Fig. 5, *D* and *E*). More precisely, CID5675221 exerts structural strain on Gln347, shortening the distance between it and Val292 from 16.6 to 4.8 Å, causing closure of one side of the catalytic site (Fig. 5*E*). While we speculate that CID5675221-mediated Nsp15 inhibition is caused mostly by binding to the allosteric and not the active site, our observations during the MD simulation do account for mixed inhibitory effects (Fig. 5).

Based on this model, several refinements to the chemical structure of CID5675221 can be proposed. Foremost, we propose eliminating the nitro group from the molecule, as this moiety has classically been considered a toxicophore by medicinal chemists (65). Our *in silico* model indicates that bioisosteric replacement of this group with a carboxylic acid would not alter molecular stability or affinity of the inhibitor for the allosteric pocket. In addition, we hypothesize that introduction of a hydroxyl group to the ortho position of the benzene ring, resulting in a hydrogen bond gain, would impose additional strain on Tyr279 and Gln347. While these modifications are only based on preliminary structural studies, we speculate that CID5675221 is an attractive lead hit for further development using medicinal chemistry approaches.

Overall, this study identifies a set of five lead inhibitors for SARS-CoV-2 Nsp15, a promising coronavirus drug target (29), that operate through diverse reversible and irreversible (covalent) mechanisms. We propose that derivatization of these molecules through chemistry, supported by additional crystallographic studies, and/or rational design of new molecules that harness the mechanisms described in this work, could lead to second-generation inhibitors with improved potency and drug-like properties. As many of these scaffolds are active against Nsp15 variants from other coronaviruses (Fig. S18), optimized Nsp15 inhibitor therapeutics would be useful not only for the treatment of COVID-19 but also infections from SARS-CoV-1, MERS, and related coronaviruses that are known or yet to be discovered.

## Experimental procedures

### Constructs and cloning

The sequence encoding wildtype SARS-CoV-2 Nsp15 protein (Protein Data Bank [PDB] ID: 6VWW) (20) was codon optimized for bacterial expression and purchased as a custom gene synthesis plasmid from IDT (Table S1). This sequence was PCR amplified using primers found in Table S2 and inserted into a pET-based vector (derived from pC013; Addgene #90097) using the NEBuilder HiFi DNA assembly cloning kit (NEB), according to the manufacturer's instructions. The resulting construct was verified by Sanger sequencing using T7 primers (Table S2). The Δ0–28 Δ336–347 Nsp15 mutant construct was prepared in a similar fashion using its specific primers (Table S2). To prepare the Nsp15 H250A, C291A/C293A, and C291W/C293W mutants, Q5 site-directed mutagenesis (NEB) was performed on the wildtype Nsp15 construct using the primer sequences included in Table S2, as per the manufacturer's instructions. Sequences corresponding to SARS-CoV-1 Nsp15 and MERS Nsp15 were ordered as bacterial codon-optimized gBlocks from IDT (sequences in Table S1) and were cloned into the BamHI and HindIII restriction sites of a pET-28B(+) plasmid. All plasmids were validated by sequencing.

### Nsp15 protein purification

Wildtype and mutant Nsp15 proteins were purified as previously described (20), with several modifications. Briefly, BL21(DE3)pLYsS cells were transformed with plasmids encoding Nsp15. Starter cultures were grown overnight (~16 h) at 37 °C with 5 ml of LB broth in the presence of 50 μg/ml carbenicillin. About 1 ml of this culture was then used to inoculate 2 l of LB–carbenicillin, and this culture was grown as above until an absorbance of 0.6 at 600 nm was reached. Following this, IPTG was added to a final concentration of 1 mM, and the culture was incubated at 18 °C overnight. Cells were pelleted by centrifugation at 3500*g* for 15 min and resuspended in lysis buffer (20 mM Tris–Cl, pH 7.5, 250 mM NaCl, 5 mM imidazole, pH 8.0) supplemented with Roche Complete Ultra protease inhibitor (Sigma) and 0.1 M PMSF. The mixture was incubated on ice for 30 min before being sonicated (15 s pulse-on and 59 s pulse-off for a total of 15 min at 55% amplitude). Cellular debris was pelleted by centrifuging at 28,000*g* for 1 h. Next, the lysate was collected and subjected to filtering through a 0.45 μM polyvinylidene difluoride membrane. The filtered lysate was loaded onto a 1 ml HisTrap HP column (Sigma) and purified using an AKTA Start System (Cytiva). The column was washed with buffer (20 mM Tris–Cl, pH 8.0, 250 mM NaCl, 10 mM imidazole, pH 8.0) until UV baseline was reached and subsequently eluted in gradient fashion. The final elution buffer was comprised of 20 mM Tris–Cl, pH 8.0, 250 mM NaCl, 250 mM imidazole, pH 8.0. Pooled protein fractions were concentrated with a Pierce Protein concentrator 10K (Thermo). During concentration, the buffer was exchanged with 20 mM 30 Hepes–KOH, pH 7.5, 500 mM NaCl, 1 mM DTT, and 10% glycerol. Concentrated protein was aliquoted and stored at −80 °C until usage. Protein concentration was measured using the DTT-resistant Pierce 660 nM Protein BCA Assay kit (Thermo).

### Nsp15 FRET-based activity assay

We adapted a previously described FRET-based assay, which employs a uracil-containing RNA substrate that is flanked by fluorophore and quencher moieties (21). Sequences for the substrates are listed in Table S3. Reactions were set up in black 96-well flat-bottom polystyrene plates (Corning) in 60 μl volume of reaction buffer (25 mM Hepes, 50 mM NaCl, 5 mM $MnCl_2$, and 1 mM DTT) and contained 1 ng/μl Nsp15 and 1 μM substrate unless otherwise stated. Where applicable, compounds were dissolved in DMSO and subsequently added to the reaction (DMSO concentration was kept to less than

1%). Reactions were incubated at 37 °C for the indicated times and read at excitation/emission wavelengths of 490/520 nm for FAM or 645/670 nm for Cy5 using a SpectraMax i3x spectrophotometer (Molecular Devices).

### HTS

Screening was performed at the High Content Analysis Core facility at the University of Alberta and at the Biofactorial High Throughput Biology Facility at the University of British Columbia. The library of compounds screened was comprised of roughly 30,000 compounds from the Canadian Chemical Biology Network collection, 1280 from LOPAC, 3040 from the TimTec collection, 50,000 from the ChemBridge DIVERSet collection, and 24,000 ChemBridge compounds from the GlycoNet collection. Compound overlap between the collections was <0.1%. Reagents were distributed into 384-well black flat bottom plates (Greiner) using either a JANUS 384-well liquid handling system (PerkinElmer) or an Echo525 acoustic dispenser (Beckman Coulter) contained in a Labcyte Access Workstation. The reaction setup was similar to that described previously, in which the final concentration of Nsp15 was 1 ng/μl and that of RNA was 0.5 μM. The reaction volume was 20 μl. Sequences for the positive control, RNA2, and Cy5 RNA substrates are listed in Table S3. Compounds were dissolved in DMSO and screened at a final concentration of roughly 10 μM. DMSO without any inhibitors was used as a negative control, and either 100 μM of benzopurpurin B or a reaction without Nsp15 was used as a positive control. Readings were taken at excitation/emission wavelengths of 490/520 nm for FAM or 645/670 nm for Cy5. Read 1 measured autofluorescence in a reaction mixture containing buffer, Nsp15, and compound. Subsequently, RNA substrate was added, and following incubation at 37 °C in a humidified incubator for 20 min, the reaction was stopped by addition of 100 mM EDTA and read 2 was performed. After the addition of 1 μM of positive control FAM-RNA (Table S3) to the reaction, read 3 was taken to test for potential quenching. Percent inhibition values were calculated as follows: (read 2 − read 1)/average negative control. Compounds that were found to quench the positive control by >50% in read 3 were excluded from further analysis.

### Amplex Red assay

Redox reactivity of the lead compounds was assessed using the Amplex Red Hydrogen Peroxide/Peroxidase Assay Kit (Invitrogen), according to the manufacturer's instructions. Briefly, test compounds were diluted in assay buffer to a final concentration of 100 μM in the presence of the indicated concentrations of DTT in 96-well plates. Hydrogen peroxide at a concentration of 10 μM was used as a positive control. The reaction was started by addition of 0.2 U/ml horseradish peroxidase and 50 μM final concentration of Amplex Red reagent. The reaction was incubated in the dark for 15 min at room temperature. Plates were read using a SpectraMax i3x spectrophotometer at excitation and emission wavelengths of 560 nm and 590 nm, respectively.

### Nsp15 native RNA cleavage assay

The native RNA cleavage assays employed a 31 nucleotide (nt) ssRNA (IDT and Biosynthesis) with the sequence 5′-rArArArArArArArArArArArArArArArArArArArArArUrArArArArArArArArArA-3′, whose cleavage at the "rU" site results in 21 nt and 10 nt fragments. The reaction was set up in a 10 μl volume containing a final concentration of 7.5 ng/μl Nsp15 and 250 ng of RNA in assay buffer (25 mM Hepes, 50 mM NaCl, 5 mM $MnCl_2$, and 1 mM DTT) and allowed to run for the indicated time. Prior to gel loading, samples were prepared with 2× formamide-based RNA loading dye (NEB) and were boiled at 95 °C for 5 min. Samples were electrophoresed on a denaturing 15% Mini-Protean TBE–UREA polyacrylamide gel (Bio-Rad) at 200 V for ~40 min at room temperature. Gels were stained with SYBR Gold (Thermo Fisher) for 20 min and imaged at Cy3 fluorescent channels (520/605 nm excitation/emission) with an Amersham Imager 680 (Cytiva).

### DLS

DLS was performed with a solution containing 1 mg/ml of wildtype Nsp15 in storage buffer (50 mM Hepes [pH 8], 150 mM NaCl, and 1 mM Tris(2-carboxyethyl)phosphine along with a 1% DMSO), in the absence or the presence of the indicated compounds. Compounds were tested at a concentration of fivefold excess compared with protein, with a consistent 1% DMSO concentration. The solutions were spun down and subjected to analysis using a Zetasizer Ultra Red (Malvern Panalytical) instrument in reusable 50 μl cuvettes. Measurements were performed using a He–Ne laser emitting at 633 nm, at a scattering angle of 90°, at a temperature of 25 °C, and with a laser power of 10 mW. Samples were analyzed in three cycles, and the correlation function graphs were averaged. The particle's hydrodynamic diameter was derived from the correlation function using ZS Xplorer software (Malvern Panalytical, version 1.0).

### Fluorescence polarization assay

Increasing concentrations of Nsp15 protein or inhibitor were titrated against a constant concentration of RNA substrate. A labeled RNA substrate (20 nM) was used in a reaction volume of 20 μl and dispensed in OptiPlate 384 F black microplates (PerkinElmer) after a 10 min incubation. 6-FAM fluorescence was excited at 480 nm, and its emission was measured at 535 nm using an EnVision 2103 Multilabel Plate Reader (PerkinElmer, Inc). The change in polarization was plotted against the logarithmic concentration of the protein or inhibitor.

### MALDI-TOF mass spectroscopy

About 20 mg/ml of wildtype Nsp15 protein was preincubated overnight in the absence or the presence of a fivefold molar excess of each compound. Samples were diluted 10-fold (50% acetonitrile/water + 0.1% trifluoroacetic acid), and 1 μl of each was then mixed with 1 μl of sinapinic acid (10 mg/ml in 50% acetonitrile/water + 0.1% trifluoroacetic acid). The sample–matrix solutions were spotted onto a stainless-steel target plate and allowed to air dry. Mass spectra were acquired using an Autoflex Speed MALDI-TOF mass spectrometer (Bruker Daltonik) with a Smartbeam-II laser at a frequency of 2 KHz. Ions were analyzed in positive mode, and external calibration was performed using a standard protein mixture.

### *LC–MS/MS mapping of modified peptides*

Samples of recombinant Nsp15 protein (10 μg) were incubated in a buffer solution (25 mM Hepes/pH 7.5 with 50 mM NaCl, 5 mM $MnCl_2$, 0.1 mM DTT) in the absence or the presence of 67 μM IPA-3 for periods of either 4 h or 16 h, at 25 °C or 4 °C, respectively. Subsequently, samples were prepared for LC–MS/MS analysis. Samples were dissolved in 100 mM ammonium bicarbonate, reduced (DTT), alkylated (iodoacetamide), and subjected to digestion with trypsin (Promega sequencing grade) overnight at 37 °C. The pH of the samples was then adjusted to 3 to 4 with formic acid, and they were dried, dissolved in water + 0.2% formic acid, and desalted (Pierce C18 tips). The tryptic peptides were resolved and ionized by using nano flow HPLC (Easy-nLC 1000; Thermo Scientific) coupled to a Q Exactive orbitrap mass spectrometer (Thermo Scientific) with an EASY-Spray capillary HPLC column (catalog no.: ES902; Thermo Scientific). The mass spectrometer was operated in data-dependent acquisition mode, recording high-accuracy and high-resolution survey orbitrap spectra using external mass calibration, with a resolution of 35,000 and an *m/z* range of 300 to 1700. The 12 most intense multiply charged ions were sequentially fragmented by using high-emergy collisional dissociation, and spectra of their fragments were recorded in the orbitrap at a resolution of 17,500.

### *X-ray crystallography*

Crystals of the Nsp15 H250A mutant protein were grown by sitting drop vapor diffusion with a reservoir solution containing 0.2 M calcium acetate, 0.1 M imidazole, HCl (pH 8), and 10% (w/v) PEG 8000. Diffraction data were collected on a home source rotating anode, and indexed, integrated, and scaled using HKL2000 (HKL Research Inc). Molecular replacement was carried out using PHASER (search model PDB ID: 6VWW (https://www.rcsb.org/structure/6VWW), chain A). Data quality assessment was carried out using Xtriage (Phenix). Twinning was detected (Twin law: h, -h-k, and -l), and the twinning law was applied in further refinement steps. Refinement in PHENIX utilized automated non-crystallographic symmetry and reference model restraints throughout, with the latter being removed for the final refinement round.

### *Molecular modeling*

The crystal structure of Nsp15 was obtained from the PDB (PDB ID: 6WXC) (30), and Schrodinger's Mastro and EPIK (66) were used to prepare the protein structure by removing chain B, followed by a short minimization, addition of missing side chains, and equilibration of the protonated group to the biological pH (7.0). The SiteMap module of Schrodinger (67) was used to detect binding sites on the surface of Nsp15 protein. The 3D structures of the inhibitors were prepared for docking using the UCSF CHIMERA, V1.10.2 dock prep tool in the framework of the AMBER99SB force field (68). Using AutoDock Vina (69), all the compounds were docked into the two identified binding pockets with a grid box of 40 × 40 × 40 and a spacing of 0.375 Å. A total of 12 docking runs were performed with exhaustiveness of 40 for each inhibitor. Prior to doing the molecular dynamic simulation on the enzyme–inhibitor complexes, an MD simulation on the apoprotein was performed using the GROMACS 2021 package (70). TIP3P water models were utilized to solvate the protein with a 1 nm marginal cushion on each side. The box was then neutralized using NaCl, and the system was minimized using the AMBER99SB0ILDN force field. The system was heated to 300 K and equilibrated for 500 ps using the Berendsen Thermostat. Using the isothermal–isobaric ensemble at 1 bar with the Parrinello–Rahman barostat, an additional equilibration was also performed. A 150 ns production run was performed using the periodic boundary condition. The Lenard–Jones, the Coulomb (cutoff = 1.0 nm), and the particle mesh Ewald were used to calculate the Van Der Waals and electrostatic interactions. The enzyme–inhibitor complexes were analyzed using the same conditions. The AnteChamber Python Parser interfacE (71) was used for ligand parameterization. All the data were plotted using Schrodinger's PyMOL package and ChimeraX (72). The free energy of interaction between each inhibitor and Nsp15 was calculated using the gmx_MMPBSA tool (73). The C291W/C293W mutant protein structure was generated and prepared using Schrodinger's Maestro. MD simulation for this protein was performed as described previously, and snapshots were taken every 75 ns for structural comparison, assessment, and analysis.

### *Vero CCL-81 cell culture*

Cells were obtained from American Type Culture Collection and were previously authenticated and shown to be negative for mycoplasma at the time of purchase. Cells were maintained in high glucose Dulbecco's modified Eagle's medium (Thermo Fisher) supplemented with 10% fetal bovine serum, Canadian origin (Sigma), and 1× penicillin–streptomycin–glutamine (Thermo), and grown in a 37 °C humidified incubator with 5% $CO_2$.

### *Cell cytotoxicity assay*

Vero CCL-81 cells were seeded in 96-well plates (Greiner) at 10,000 cells per well overnight before compounds were added to wells at the indicated concentrations alongside a DMSO control. Twenty-four hours later, cell viability was assayed using the CellTiter-Glo Luminescent Cell Viability Assay (Promega), according to the manufacturer's instructions. Briefly, this assay relies on the quantification of ATP using a proprietary Ultra-Glo Luciferase that converts luciferin to luminescent oxyluciferin in the presence of ATP. Cells were incubated in 100 μl of complete media with 100 μl of reconstituted CellTiter-Glo

Reagent (buffer plus substrate). Samples were mixed by shaking for 10 min, and then luminescence was measured using a Spectramax i3x (Molecular Devices) device. Data from experimental wells were normalized to the appropriate DMSO control.

### *SARS-CoV-2 sourcing and propagation*

The Canadian clinical isolate of SARS-CoV-2 72B/CA/CALG (74) was propagated at the biosafety level 3 laboratory at the University of Alberta in Vero-E6 cells (American Type Culture Collection), grown in Dulbecco's modified Eagle's medium supplemented with 3% fetal bovine serum, 15 mM Hepes, 1× L-glutamine, and penicillin–streptomycin. All experiments using live virus were carried out at this facility, in accordance with approved protocols.

### *SARS-CoV-2 plaque assay*

SARS-CoV-2 supernatant samples were serially diluted (10-fold dilutions) in fresh media and used to infect monolayers of $1 \times 10^5$ Vero-CCL-81 cells in 24-well plates (Greiner) for 1 h. Subsequently, viral supernatants were removed, and cell monolayers were overlaid with a mixture of MEM (Thermo Fisher Scientific) and 0.75 to 1.5% carboxymethylcellulose (Sigma–Aldrich). Cells were maintained at 37 °C for 3 days to allow plaque development to occur. Before plaque counting, cells were fixed with 10% formaldehyde and stained with 1% crystal violet in 20% ethanol.

### *Assessment of SARS-CoV-2 RNA levels following drug treatment*

Vero CCL-81 cells were seeded in 96-well plates (Greiner) at $1 \times 10^4$ cells per well. The next day, cells were rinsed once with PBS, and SARS-CoV-2 at a multiplicity of infection of 0.1 was added to the wells using fresh media supplemented with 10% fetal bovine serum containing either DMSO or inhibitors. The cells were then incubated at 37 °C for 24 h. Next, following removal of the supernatant, total RNA from Vero CCL-81 cells was extracted using the NucleoSpin RNA kit (Macherey–Nagel), following the manufacturer's protocol. Total RNA was reverse transcribed using 0.5 to 1 μg of total RNA and ImProm-II Reverse Transcriptase (Promega), according to the manufacturer's protocol. Quantitative RT–PCR was performed with PerfeCTa SYBR Green SuperMix (Quanta BioSciences) using a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). The cycling conditions were 45 cycles of 94 °C for 30 s, 55 °C for 60 s, and 68 °C for 20 s. Gene expression (fold change) was calculated using the $2^{(-\Delta\Delta CT)}$ method using human β-actin messenger RNA as the housekeeping gene. The β-actin primers used for this analysis were 5′-TGGATCAGCAAGCAGGAGTATG-3′ and 5′-GCATTTGCGGTGGACGAT-3′. The SARS-CoV-2 spike primers were 5′-CAATGGTTTAACAGGCACAGG-3′ and 5′-CTCAAGTGTCTGTGGATCACG-3′. Remdesivir (Medkoo) at 2.5 μM was used as positive control in selected assays.

### *Assessment of SARS-CoV-2 protein levels via immunofluorescence following drug treatment*

Vero CCL-81 cells were grown overnight on coverslips at $1 \times 10^5$ cells per well in 12-well plates (Greiner). DMSO or inhibitors were added to the cells in fresh media combined with SARS-CoV-2 virus at a multiplicity of infection of 0.1. After 24 h of incubation, coverslips were collected for staining and imaging of viral antigens. The coverslips were fixed with 4% paraformaldehyde and permeabilized and blocked with a Triton X-100 (0.2%)–bovine serum albumin (3%) solution. Cells were incubated with a 1:250 dilution of mouse monoclonal anti-SARS-CoV/SARS-CoV-2 spike protein antibody (1A9; GeneTex) at room temperature for 1.5 h. Next, they were washed twice and then incubated with Alexa Fluor 647 secondary antimouse antibodies (Invitrogen), diluted at 1:1000, and 4′,6-diamidino-2-phenylindole (1 μg/ml) was added for 1 h at room temperature prior to being washed (two times). Antibodies were diluted in blocking buffer, and PBS containing 0.3% bovine serum albumin was used for the wash steps. Samples were visualized using an Olympus 1 × 81 spinning-disk confocal microscope. Images were analyzed using Volocity (PerkinElmer) or Gen5 (BioTek) software.

### *Data analysis and software*

GraphPad Prism (GraphPad Software, Inc) was used for general plotting and statistics. Chemical structures were generated using ChemDraw (ChemAxon), and chemical properties (Fig. S19) were predicted using Percepta software (ACD/Labs).

### *Quantification and statistical analysis*

Numbers of trial replicates and appropriate statistical measures and tests are indicated in the figure captions.

### Data availability

All data relating to this study are included in the article and supporting information files. Inquiries relating to this study and reagent requests should be directed to the corresponding author, Dr Basil P. Hubbard (basil.hubbard@utoronto.ca). All new plasmid constructs generated from this project will be made publicly available on Addgene ("http://www.addgene.org"). Structural data corresponding to Nsp15-H250A have been deposited in PDB (accession code: 8D34). Raw LC–MS/MS data files have been deposited in the MassIVE repository ("https://massive.ucsd.edu") and are freely available: MSV000092541.

---

*Supporting information*—This article contains supporting information (21, 75, 76).

*Acknowledgments*—We thank Ryan McLaughlin for the computational skills used to create several figures and recognize assistance from the UBC GREx Biological Resilience Initiative in support of Biofactorial, UBC's High Throughput Biology Facility. Additional support was provided by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.

sharcnet.ca) and the Digital Research Alliance of Canada ("https://alliancecan.ca/"). We also thank the developers of ChimeraX at the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.

*Author contributions*—J. C. and B. P. H. conceptualization; J. C. and B. P. H. methodology; J. C., R. A. F., D. L., S. A. T. D., E. M. K., A. B., F. M. R., O. M. M., B. B., and T. A. P. investigation; J. C., R. A. F., D. L., S. A. T. D., E. M. K., A. B., F. M. R., O. M. M., B. B., T. A. P., M. T. M., J. R. U., T. C. H., J. N. M. G., and B. P. H. writing–original draft; M. T. M., J. R. U., T. C. H., and J. N. M. G. supervision.

*Funding and additional information*—This work was supported by a COVID-19 and Emerging Pathogens Research Fund grant from the Li Ka Shing Institute of Virology (University of Alberta) and a COVID-19 Rapid Research Funding Operating Grant from the Canadian Institutes for Health Research (CIHR; with Francois Jean and Artem Cherkasov from the University of British Columbia as principal investigators) to B. P. H. Research in the T. C. H. laboratory was also funded in part by a CIHR COVID-19 Rapid Response Funding Operating Grant. S. A. T. D. is a CIHR postdoctoral fellow (grant no.: MFE-186352).

*Conflict of interest*—The authors declare that they have no conflicts of interest with the contents of this article.

*Abbreviations*—The abbreviations used are: COVID-19, coronavirus disease 2019; DLS, dynamic light scattering; DMSO, dimethyl sulfoxide; FAM, carboxyfluorescein; HTS, high-throughput screen; IFN, interferon; MD, molecular dynamics; MERS, Middle East respiratory syndrome; Nsp, nonstructural protein; PAIN, pan-assay interference; PDB, Protein Data Bank; SARS-CoV, severe acute respiratory syndrome coronavirus.

## References

1. Cardenas, G., Chávez-Canales, M., Espinosa, A. M., Jordán-Ríos, A., Malagon, D. A., Murillo, M. F. M., *et al.* (2022) Intranasal dexamethasone: a new clinical trial for the control of inflammation and neuroinflammation in COVID-19 patients. *Trials* **23**, 148
2. Zhang, W., Liu, L., Xiao, X., Zhou, H., Peng, Z., Wang, W., *et al.* (2023) Identification of common molecular signatures of SARS-CoV-2 infection and its influence on acute kidney injury and chronic kidney disease. *Front. Immunol.* **14**, 961642
3. Zhou, Z., Qiu, Y., and Ge, X. (2021) The taxonomy, host range and pathogenicity of coronaviruses and other viruses in the Nidovirales order. *Anim. Dis.* **1**, 5
4. Abdelrahman, Z., Li, M., and Wang, X. (2020) Comparative review of SARS-CoV-2, SARS-CoV, MERS-CoV, and influenza A respiratory viruses. *Front. Immunol.* **11**, 552909
5. Choi, R., Zhou, M., Shek, R., Wilson, J. W., Tillery, L., Craig, J. K., *et al.* (2021) High-throughput screening of the ReFRAME, pandemic box, and COVID box drug repurposing libraries against SARS-CoV-2 nsp15 endoribonuclease to identify small-molecule inhibitors of viral activity. *PLoS One* **16**, e0250019
6. Ke, Z., Oton, J., Qu, K., Cortese, M., Zila, V., McKeane, L., *et al.* (2020) Structures and distributions of SARS-CoV-2 spike proteins on intact virions. *Nature* **588**, 498–502
7. Mariano, G., Farthing, R. J., Lale-Farjat, S. L. M., and Bergeron, J. R. C. (2020) Structural characterization of SARS-CoV-2: where we are, and where we need to be. *Front. Mol. Biosci.* **7**, 605236
8. Kim, D., Lee, J. Y., Yang, J. S., Kim, J. W., Kim, V. N., and Chang, H. (2020) The architecture of SARS-CoV-2 transcriptome. *Cell* **181**, 914–921.e10
9. Decroly, E., Debarnot, C., Ferron, F., Bouvet, M., Coutard, B., Imbert, I., *et al.* (2011) Crystal structure and functional analysis of the SARS-coronavirus RNA cap 2'-O-methyltransferase nsp10/nsp16 complex. *PLoS Pathog.* **7**, e1002059
10. Kyriakidis, N. C., Lopez-Cortes, A., Gonzalez, E. V., Grimaldos, A. B., and Prado, E. O. (2021) SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *NPJ Vaccines* **6**, 28
11. Vuong, W., Khan, M. B., Fischer, C., Arutyunova, E., Lamer, T., Shields, J., *et al.* (2020) Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *Nat. Commun.* **11**, 4282
12. Lim, C. T., Tan, K. W., Wu, M., Ulferts, R., Armstrong, L. A., Ozono, E., *et al.* (2021) Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of Nsp3 papain-like protease. *Biochem. J.* **478**, 2517–2531
13. Kokic, G., Hillen, H. S., Tegunov, D., Dienemann, C., Seitz, F., Schmitzova, J., *et al.* (2021) Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.* **12**, 279
14. Hammond, J., Leister-Tebbe, H., Gardner, A., Abreu, P., Bao, W., Wisemandle, W., *et al.* (2022) Oral nirmatrelvir for high-risk, nonhospitalized adults with Covid-19. *N. Engl. J. Med.* **386**, 1397–1408
15. Jayk Bernal, A., Gomes da Silva, M. M., Musungaie, D. B., Kovalchuk, E., Gonzalez, A., Delos Reyes, V., *et al.* (2022) Molnupiravir for oral treatment of Covid-19 in nonhospitalized patients. *N. Engl. J. Med.* **386**, 509–520
16. Wang, P., Nair, M. S., Liu, L., Iketani, S., Luo, Y., Guo, Y., *et al.* (2021) Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135
17. Van Egeren, D., Novokhodko, A., Stoddard, M., Tran, U., Zetter, B., Rogers, M., *et al.* (2021) Risk of rapid evolutionary escape from biomedical interventions targeting SARS-CoV-2 spike protein. *PLoS One* **16**, e0250780
18. Kozlov, M. (2022) Why scientists are racing to develop more COVID antivirals. *Nature* **601**, 496
19. Ortiz-Alcantara, J., Bhardwaj, K., Palaninathan, S., Frieman, M., Baric, R., and Kao, C. (2010) Small molecule inhibitors of the SARS-CoV Nsp15 endoribonuclease. *Virus Adapt. Treat.* **2**, 125–133
20. Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., *et al.* (2020) Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci.* **29**, 1596–1605
21. Bhardwaj, K., Sun, J., Holzenburg, A., Guarino, L. A., and Kao, C. C. (2006) RNA recognition and cleavage by the SARS coronavirus endoribonuclease. *J. Mol. Biol.* **361**, 243–256
22. Pillon, M. C., Frazier, M. N., Dillard, L. B., Williams, J. G., Kocaman, S., Krahn, J. M., *et al.* (2021) Cryo-EM structures of the SARS-CoV-2 endoribonuclease Nsp15 reveal insight into nuclease specificity and dynamics. *Nat. Commun.* **12**, 636
23. Ivanov, K. A., Hertzig, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A. E., *et al.* (2004) Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 12694–12699
24. Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., *et al.* (2017) Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4251–E4260
25. Deng, X., and Baker, S. C. (2018) An "old" protein with a new story: coronavirus endoribonuclease is important for evading host antiviral defenses. *Virology* **517**, 157–163
26. Kindler, E., Gil-Cruz, C., Spanier, J., Li, Y., Wilhelm, J., Rabouw, H. H., *et al.* (2017) Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus replication. *PLoS Pathog.* **13**, e1006195
27. Shemesh, M., Aktepe, T. E., Deerain, J. M., McAuley, J. L., Audsley, M. D., David, C. T., *et al.* (2021) SARS-CoV-2 suppresses IFNbeta production mediated by NSP1, 5, 6, 15, ORF6 and ORF7b but does not suppress the effects of added interferon. *PLoS Pathog.* **17**, e1009800
28. Hackbart, M., Deng, X., and Baker, S. C. (2020) Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 8094–8103
29. Stevaert, A., Krasniqi, B., Van Loy, B., Nguyen, T., Thomas, J., Vandeput, J., *et al.* (2021) Betulonic acid derivatives interfering with human

coronavirus 229E replication via the nsp15 endoribonuclease. *J. Med. Chem.* **64**, 5632–5644

30. Kim, Y., Wower, J., Maltseva, N., Chang, C., Jedrzejczak, R., Wilamowski, M., *et al.* (2021) Tipiracil binds to uridine site and inhibits Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Commun. Biol.* **4**, 193
31. Canal, B., Fujisawa, R., McClure, A. W., Deegan, T. D., Wu, M., Ulferts, R., *et al.* (2021) Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of nsp15 endoribonuclease. *Biochem. J.* **478**, 2465–2479
32. Kumar, S., Gupta, Y., Zak, S. E., Upadhyay, C., Sharma, N., Herbert, A. S., *et al.* (2021) A novel compound active against SARS-CoV-2 targeting uridylate-specific endoribonuclease (NendoU/NSP15): in silico and in vitro investigations. *RSC Med. Chem.* **12**, 1757–1764
33. Baell, J. B., and Nissink, J. W. M. (2018) Seven year itch: pan-assay interference compounds (PAINS) in 2017-utility and limitations. *ACS Chem. Biol.* **13**, 36–44
34. Silvernale, J. N., Joswick, H. L., Corner, T. R., and Gerhardt, P. (1971) Antimicrobial actions of hexachlorophene: cytological manifestations. *J. Bacteriol.* **108**, 482–491
35. Deacon, S. W., Beeser, A., Fukui, J. A., Rennefahrt, U. E. E., Myers, C., Chernoff, J., *et al.* (2008) An isoform-selective, small-molecule inhibitor targets the autoregulatory mechanism of p21-activated kinase. *Chem. Biol.* **15**, 322–331
36. Yang, G. P., Erdman, D. D., Tondella, M. L., and Fields, B. S. (2009) Evaluation of tetramethylrhodamine and black hole quencher 1 labeled probes and five commercial amplification mixes in TaqMan real-time RT-PCR assays for respiratory pathogens. *J. Virol. Methods* **162**, 288–290
37. Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Buchmeier, M. J., Stevens, R. C., *et al.* (2007) Crystal structure of a monomeric form of severe acute respiratory syndrome coronavirus endonuclease nsp15 suggests a role for hexamerization as an allosteric switch. *J. Virol.* **81**, 6700–6708
38. Markossian, S., Grossman, A., Brimacombe, K., Arkin, M., Auld, D., Austin, C., *et al.* eds. (2004) *Assay Guidance Manual [Internet]*. Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda, MD
39. Johnston, P. A. (2011) Redox cycling compounds generate H2O2 in HTS buffers containing strong reducing reagents–real hits or promiscuous artifacts? *Curr. Opin. Chem. Biol.* **15**, 174–182
40. Hubbard, B. P., Gomes, A. P., Dai, H., Li, J., Case, A. W., Considine, T., *et al.* (2013) Evidence for a common mechanism of SIRT1 regulation by allosteric activators. *Science* **339**, 1216–1219
41. Feng, B. Y., and Shoichet, B. K. (2006) A detergent-based assay for the detection of promiscuous inhibitors. *Nat. Protoc.* **1**, 550–553
42. Ryan, A. J., Gray, N. M., Lowe, P. N., and Chung, C. W. (2003) Effect of detergent on "promiscuous" inhibitors. *J. Med. Chem.* **46**, 3448–3451
43. Lorber, B., Fischer, F., Bailly, M., Roy, H., and Kern, D. (2012) Protein analysis by dynamic light scattering: methods and techniques for students. *Biochem. Mol. Biol. Educ.* **40**, 372–382
44. Stetefeld, J., McKenna, S. A., and Patel, T. R. (2016) Dynamic light scattering: a practical guide and applications in biomedical sciences. *Biophys. Rev.* **8**, 409–427
45. Viaud, J., and Peterson, J. R. (2009) An allosteric kinase inhibitor binds the p21-activated kinase autoregulatory domain covalently. *Mol. Cancer Ther.* **8**, 2559–2565
46. Cao, J., Wu, C. C., and Lin, T. L. (2008) Turkey coronavirus non-structure protein NSP15–an endoribonuclease. *Intervirology* **51**, 342–351
47. Frazier, M. N., Dillard, L. B., Krahn, J. M., Perera, L., Williams, J. G., Wilson, I. M., *et al.* (2021) Characterization of SARS2 Nsp15 nuclease activity reveals it's mad about U. *Nucleic Acids Res.* **49**, 10136–10149
48. Saramago, M., Costa, V. G., Souza, C. S., Bárria, C., Domingues, S., Viegas, S. C., *et al.* (2022) The nsp15 nuclease as a good target to combat SARS-CoV-2: mechanism of action and its inactivation with FDA-approved drugs. *Microorganisms* **10**, 342
49. Ogando, N. S., Dalebout, T. J., Zevenhoven-Dobbe, J. C., Limpens, R. W. A. L., van der Meer, Y., Caly, L., *et al.* (2020) SARS-coronavirus-2 replication in Vero E6 cells: replication kinetics, rapid adaptation and cytopathology. *J. Gen. Virol.* **101**, 925–940
50. Benet, L. Z., Hosey, C. M., Ursu, O., and Oprea, T. I. (2016) BDDCS, the rule of 5 and drugability. *Adv. Drug Deliv. Rev.* **101**, 89–98
51. Shuman, R. M., Leech, R. W., and Alvord, E. C., Jr. (1975) Neurotoxicity of hexachlorophene in humans. II. A clinicopathological study of 46 premature infants. *Arch. Neurol.* **32**, 320–325
52. McDonnell, G., and Russell, A. D. (1999) Antiseptics and disinfectants: activity, action, and resistance. *Clin. Microbiol. Rev.* **12**, 147–179
53. Fu, N. J., Xi, R. Y., Shi, X. K., Li, R. Z., Zhang, Z. H., Li, L. Y., *et al.* (2022) Hexachlorophene, a selective SHP2 inhibitor, suppresses proliferation and metastasis of KRAS-mutant NSCLC cells by inhibiting RAS/MEK/ERK and PI3K/AKT signaling pathways. *Toxicol. Appl. Pharmacol.* **441**, 115988
54. Kleinboelting, S., Ramos-Espiritu, L., Buck, H., Colis, L., van den Heuvel, J., Glickman, J. F., *et al.* (2016) Bithionol potently inhibits human soluble adenylyl cyclase through binding to the allosteric activator site. *J. Biol. Chem.* **291**, 9776–9784
55. Lokanatha, V., Sailaja, P., and Rajendra, W. (1999) *In vitro* kinetics of the rat brain succinate dehydrogenase inhibition by hexachlorophene. *J. Biochem. Mol. Toxicol.* **13**, 303–306
56. Jeon, S., Ko, M., Lee, J., Choi, I., Byun, S. Y., Park, S., *et al.* (2020) Identification of antiviral drug candidates against SARS-CoV-2 from FDA-approved drugs. *Antimicrob. Agents Chemother.* **64**, e00819-20
57. Liu, Y. C., Huang, V., Chao, T. C., Hsiao, C. D., Lin, A., Chang, M. F., *et al.* (2005) Screening of drugs by FRET analysis identifies inhibitors of SARS-CoV 3CL protease. *Biochem. Biophys. Res. Commun.* **333**, 194–199
58. Itaba, N., Matsumi, Y., Okinaka, K., Ashla, A. A., Kono, Y., Osaki, M., *et al.* (2015) Human mesenchymal stem cell-engineered hepatic cell sheets accelerate liver regeneration in mice. *Sci. Rep.* **5**, 16169
59. Xu, X., Zhai, Y., Sun, F., Lou, Z., Su, D., Xu, Y., *et al.* (2006) New antiviral target revealed by the hexameric structure of mouse hepatitis virus nonstructural protein nsp15. *J. Virol.* **80**, 7909–7917
60. Verma, A., Artham, S., Alwhaibi, A., Adil, M. S., Cummings, B. S., and Somanath, P. R. (2020) PAK1 inhibitor IPA-3 mitigates metastatic prostate cancer-induced bone remodeling. *Biochem. Pharmacol.* **177**, 113943
61. Sutanto, F., Konstantinidou, M., and Domling, A. (2020) Covalent inhibitors: a rational approach to drug discovery. *RSC Med. Chem.* **11**, 876–884
62. Baell, J. B., and Holloway, G. A. (2010) New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53**, 2719–2740
63. Eng, H., Dantonio, A. L., Kadar, E. P., Obach, R. S., Di, L., Lin, J., *et al.* (2022) Disposition of PF-07321332 (nirmatrelvir), an orally bioavailable inhibitor of SARS-CoV-2 3CL protease, across animals and humans. *Drug Metab. Dispos.* **50**, 576–590
64. Iyer, S., Sam, F. S., DiPrimio, N., Preston, G., Verheijen, J., Murthy, K., *et al.* (2019) Repurposing the aldose reductase inhibitor and diabetic neuropathy drug epalrestat for the congenital disorder of glycosylation PMM2-CDG. *Dis. Model. Mech.* **12**, dmm040584
65. Nepali, K., Lee, H. Y., and Liou, J. P. (2019) Nitro-group-containing drugs. *J. Med. Chem.* **62**, 2851–2893
66. Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007) Epik: a software program for pK prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **21**, 681–691
67. Halgren, T. (2007) New method for fast and accurate binding-site identification and analysis. *Chem. Biol. Drug Des.* **69**, 146–148
68. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612
69. Trott, O., and Olson, A. J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461
70. Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: fast, flexible, and free. *J. Comput. Chem.* **26**, 1701–1718

71. Sousa da Silva, A. W., and Vranken, W. F. (2012) ACPYPE - AnteChamber PYthon Parser interfacE. *BMC Res. Notes* **5**, 367
72. Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., *et al.* (2021) UCSF ChimeraX: structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82
73. Miller, B. R., 3rd, McGee, T. D., Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012) MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* **8**, 3314–3321
74. Lin, Y. C., Malott, R. J., Ward, L., Kiplagat, L., Pabbaraju, K., Gill, K., *et al.* (2022) Detection and quantification of infectious severe acute respiratory coronavirus-2 in diverse clinical and environmental samples. *Sci. Rep.* **12**, 5418
75. Nedialkova, D. D., Ulferts, R., van den Born, E., Lauber, C., Gorbalenya, A. E., Ziebuhr, J., *et al.* (2009) Biochemical characterization of arterivirus nonstructural protein 11 reveals the nidovirus-wide conservation of a replicative endoribonuclease. *J Virol* **83**, 5671–5682
76. Guarino, L. A., Bhardwaj, K., Dong, W., Sun, J., Holzenburg, A., and Kao, C. (2005) Mutational analysis of the SARS virus Nsp15 endoribonuclease: identification of residues affecting hexamer formation. *J Mol Biol* **353**, 1106–1117

# Appendix B. The Antipsychotic Dopamine 2 Receptor Antagonist Diphenylbutylpiperidines Improve Glycemiain Experimental Obesity by Inhibiting Succinyl-CoA:3-Ketoacid CoA Transferase

This chapter was published:

Seyed Amirhossein Tabatabaei Dakhili, Amanda A. Greenwell, Kunyan Yang, Rabih Abou Farraj, Christina T. Saed, Keshav Gopal, Jordan S.F. Chan, Jadin J. Chahade, Farah Eaton, Crystal Lee, Carlos A. Velazquez-Martinez, Peter A. Crawford, J.N. Mark Glover, Rami Al Batran, and John R. Ussher. The Antipsychotic Dopamine 2 Receptor Antagonist Diphenylbutylpiperidines Improve Glycemiain Experimental Obesity by Inhibiting Succinyl-CoA:3-Ketoacid CoA Transferase. *Diabetes*.

**Diabetes** Volume 72, January 2023

# The Antipsychotic Dopamine 2 Receptor Antagonist Diphenylbutylpiperidines Improve Glycemia in Experimental Obesity by Inhibiting Succinyl-CoA:3-Ketoacid CoA Transferase

Seyed Amirhossein Tabatabaei Dakhili,[1,2,3] Amanda A. Greenwell,[1,2,3] Kunyan Yang,[1,2,3] Rabih Abou Farraj,[4] Christina T. Saed,[1,2,3] Keshav Gopal,[1,2,3] Jordan S.F. Chan,[1,2,3] Jadin J. Chahade,[1,2,3] Farah Eaton,[1,2,3] Crystal Lee,[1,2,3] Carlos A. Velázquez-Martínez,[1] Peter A. Crawford,[5] J.N. Mark Glover,[4] Rami Al Batran,[6] and John R. Ussher[1,2,3]

*Diabetes 2023;72:126–134 | https://doi.org/10.2337/db22-0221*

**Despite significant progress in understanding the pathogenesis of type 2 diabetes (T2D), the condition remains difficult to manage. Hence, new therapeutic options targeting unique mechanisms of action are required. We have previously observed that elevated skeletal muscle succinyl CoA:3-ketoacid CoA transferase (SCOT) activity, the rate-limiting enzyme of ketone oxidation, contributes to the hyperglycemia characterizing obesity and T2D. Moreover, we identified that the typical antipsychotic agent pimozide is a SCOT inhibitor that can alleviate obesity-induced hyperglycemia. We now extend those observations here, using computer-assisted in silico modeling and in vivo pharmacology studies that highlight SCOT as a noncanonical target shared among the diphenylbutylpiperidine (DPBP) drug class, which includes penfluridol and fluspirilene. All three DPBPs tested (pimozide, penfluridol, and fluspirilene) improved glycemia in obese mice. While the canonical target of the DPBPs is the dopamine 2 receptor, studies in obese mice demonstrated that acute or chronic treatment with a structurally unrelated antipsychotic dopamine 2 receptor antagonist, lurasidone, was devoid of glucose-lowering actions. We further observed that the DPBPs improved glycemia in a SCOT-dependent manner in skeletal muscle, suggesting that this older class of antipsychotic agents may have utility in being repurposed for the treatment of T2D.**

It is imperative that new treatments continue to be developed for type 2 diabetes (T2D), as the majority of individuals with T2D will eventually require multiple therapies to control their glycemia (1,2). Recent observations have demonstrated that high-fat diet (HFD)–induced obesity in mice increases the expression/activity of succinyl CoA:3-ketoacid CoA transferase (SCOT) in skeletal muscle, the rate-limiting enzyme of ketone oxidation (3). It was determined that a muscle-specific knockout (KO) of SCOT ($\text{SCOT}^{\text{MuscleKO}}$) in mice improves glycemia in response to experimental obesity, implying that this metabolic adaptation may contribute to the pathology of obesity-induced T2D. Intriguingly, it was demonstrated with in silico modeling that pimozide, an antipsychotic drug from the diphenylbutylpiperidine (DPBP) family, was a SCOT inhibitor. In vivo studies confirmed that pimozide inhibited SCOT, while recapitulating the glucose-lowering effect observed in obese $\text{SCOT}^{\text{MuscleKO}}$ mice. Moreover, pimozide failed to improve glycemia in obese $\text{SCOT}^{\text{MuscleKO}}$ mice, confirming that its ability to improve glycemia depended on SCOT inhibition and not its canonical antipsychotic actions where it primarily inhibits the dopamine 2 ($D_2$) receptor (4).

---

[1] Faculty of Pharmacy and Pharmaceutical Sciences, University of Alberta, Edmonton, Alberta, Canada
[2] Alberta Diabetes Institute, University of Alberta, Edmonton, Alberta, Canada
[3] Cardiovascular Research Centre, University of Alberta, Edmonton, Alberta, Canada
[4] Department of Biochemistry, University of Alberta, Edmonton, Alberta, Canada
[5] Division of Molecular Medicine, Department of Medicine, Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN
[6] Faculty of Pharmacy, University of Montreal, Montreal, Quebec, Canada

Corresponding author: John R. Ussher, jussher@ualberta.ca

Received 2 March 2022 and accepted 12 October 2022

A.A.G. and K.Y. contributed equally to this work.

This article contains supplementary material online at https://doi.org/10.2337/figshare.21318354.

Pimozide and other members of the DPBP family, were synthesized as $D_2$ receptor antagonists in the late 1960s to treat schizophrenia and other mental disorders (5–7). Pimozide is currently used to manage tics in Tourette syndrome, as typical antipsychotic agents have been predominantly supplanted by atypical antipsychotics (e.g., lurasidone) (8). In relation to previous findings demonstrating that pimozide improves glycemia in obesity through SCOT inhibition, our aim was to determine whether the ability of pimozide to inhibit SCOT is preserved among the DPBPs, which may suggest new therapeutic utility in repurposing this drug class for the potential treatment of T2D.

## RESEARCH DESIGN AND METHODS

An expanded version of our research design and methods can be found in the Supplementary Material.

### Animals

All animal procedures were approved by the institution's health sciences animal welfare committee, with animals receiving care according to guidelines from the Canadian Council on Animal Care. To induce experimental obesity, 8-week-old male C57BL/6J mice were fed either a low-fat diet (LFD) (D12450J; Research Diets) or an HFD (D12451; Research Diets) for 12 weeks. Male mice were randomly assigned to receive either vehicle control (corn oil) or pimozide, fluspirilene, or penfluridol (10 mg/kg once every 48 h) by oral gavage for 4 weeks. All animals were subjected to intraperitoneal (IP) glucose tolerance tests (IPGTTs) and IP insulin tolerance tests (IPITTs) at 2 and 3 weeks post-treatment, respectively. At study completion, animals were killed with euthanyl (12 mg), following which their tissues were extracted and snap frozen in liquid nitrogen for storage at −80°C.

### Molecular Modeling

The DPBP structures were prepared in the framework of the AMBER99SB force field using the University of California, San Francisco, Chimera Dock Prep tool (9). The docking was performed using AutoDock Vina (10), followed by two sets of molecular dynamics (MD) simulations, one for the apo-SCOT (Protein Data Bank ID: 3DLX) and the other for the SCOT-drug complex using GROMACS 2021 (11). The system was neutralized to the physiological salt concentration and the energy minimized using the AMBER99SB-ILDN force field (12). The system was equilibrated to 1 bar and heated to 300 K before performing the 40-ns production run for apo-SCOT and SCOT-drug complex using the periodic boundary. Root mean square deviation (RMSD), ligand positional RMSD, root mean square fluctuation (RMSF), and hydrogen bonding analyses were calculated using GROMACS. All figures were generated using ChimeraX, and Pymol and Drug Discovery Studio Visualizer were used for protein-drug interaction analysis (13).

### Assessment of Glucose Homeostasis

After an overnight or 6-h fast, IPGTTs and IPITTs were performed following administration of glucose (2 g/kg) or insulin (0.5 units/kg), respectively, with the Contour Next blood glucose monitoring system (Bayer) used to assess blood glucose levels sampled from tail vein whole blood.

### SCOT Activity

Frozen tissues were homogenized in PBS (pH 7.2) with protease inhibitors (Halt Protease Inhibitor Cocktail; Thermo Fisher Scientific). The resulting lysate was centrifuged at 20,000*g* for 20 min at 4°C, and the supernatant was collected for measuring the rate of acetoacetyl CoA formation as previously described (14).

### Statistical Analysis

All values are presented as mean ± SEM. Multiple groups were compared using a one- or two-way ANOVA, followed by Bonferroni post hoc analysis. Differences were considered significant when $P < 0.05$. GraphPad Prism 9 software was used for all data analyses.

### Data and Resource Availability

All data associated with this study are available from the corresponding author upon reasonable request.

## RESULTS

### DPBPs Analogously Bind and Inhibit SCOT

The binding modes of penfluridol and fluspirilene were compared with that of pimozide with MD, followed by an MD simulation to determine the stability and dynamics of the DPBP-enzyme complex. Analysis of the binding modes revealed that all DPBPs, analogous to pimozide, promptly aligned and established a hydrogen bond with two critical binding pocket residues, namely Lys368 and Gly322 (Fig. 1*A–D*).

RMSD analysis revealed that these compounds did not impose any significant distortion on the backbone of the SCOT (Fig. 1*E*), and the three DPBPs were sufficiently stable inside the oxyanion pocket of SCOT during MD simulation (Fig. 1*F*). The RMSF of the side chains analysis of the drug-enzyme complex and apo-SCOT structure confirmed the initial docking poses, indicating that the DPBPs under investigation linger in contact with the residues mentioned above throughout the MD simulation (Fig. 1*G*). Hydrogen bond interactions were also computed throughout the MD simulation, highlighting their importance in DPBP-SCOT interactions (Fig. 1*H*).

We next used a cell-free assay system to quantify acetoacetyl CoA formation by recombinant SCOT in the presence of its endogenous substrates (Fig. 1*I*). Consistent with our in silico model, all three DPBPs bind to SCOT and reduce SCOT's ability to synthesize acetoacetyl CoA (Fig. 1*J*).

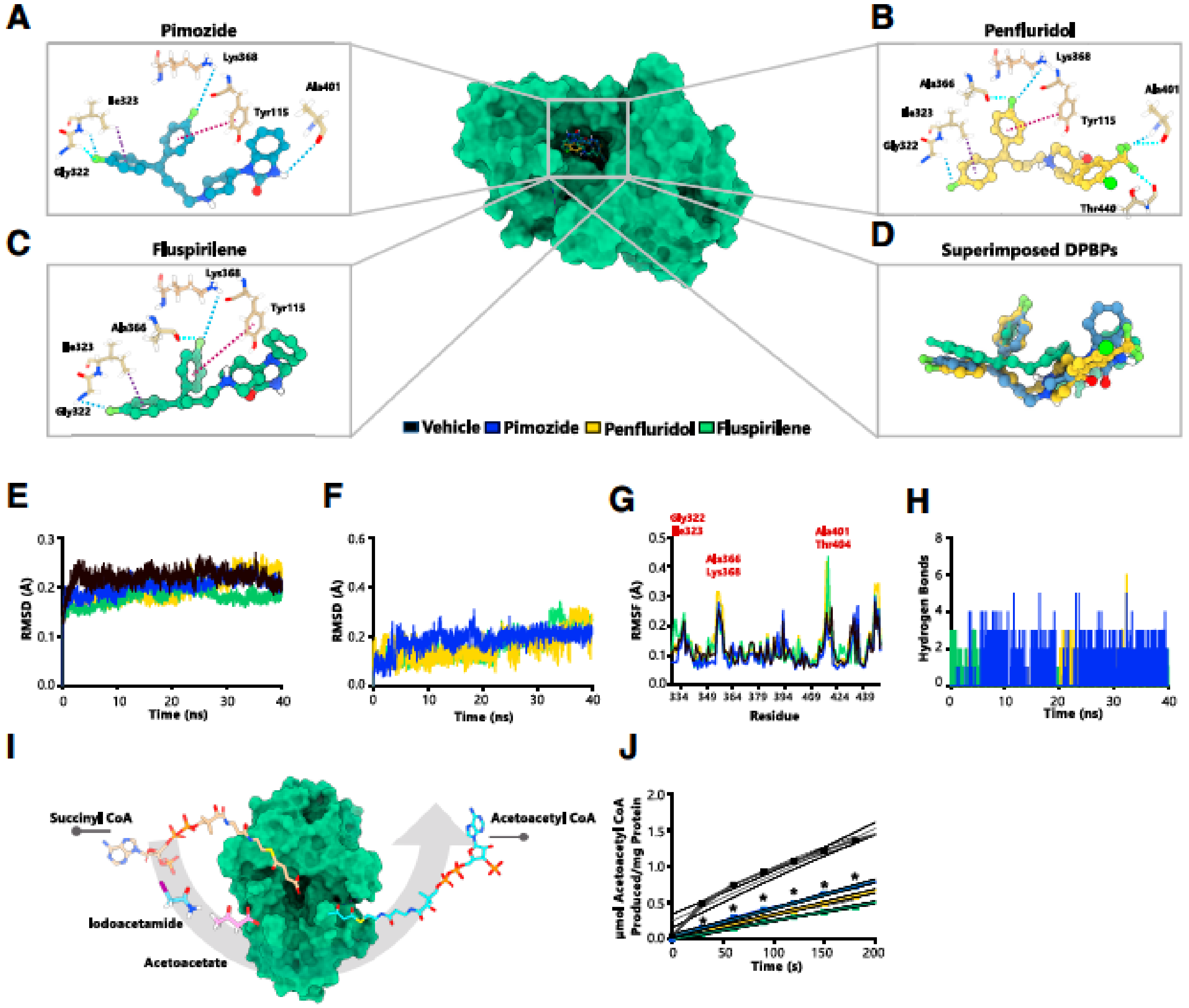


**Figure 1**—DPBPs bind SCOT and inhibit its catalytic activity. *A–D*: The predicted binding modes of DPBPs in the oxyanion pocket of SCOT. The X-ray crystal structure of human SCOT (Protein Data Bank ID: 3DLX) was acquired, and DPBPs were docked into its binding pocket. The docked structures' homologous binding modes and interacting residues of SCOT's catalytic pocket that contribute to pimozide (*A*), penfluridol (*B*), and fluspirilene (*C*) binding are illustrated. *D*: The overlaid structures of the DPBPs inside the oxyanion pocket of SCOT. Hydrogen bonds are represented with green, π–π interactions with red, halogen bonds with cyan, and π-alkyl interactions with a broken purple line. *E–H*: Analysis of the MD simulation of the SCOT-drug complexes. Backbone RMSD (*E*), ligand positional RMSD (*F*), side chain RMSF of apo-SCOT and SCOT-drug complex with the DPBPs (*G*), and the number of hydrogen bond contacts of DPBPs with SCOT's catalytic pocket during the length of the MD simulation (*H*) are shown. *I*: Schematic representation of the SCOT activity assay that measures formation of acetoacetyl CoA. *J*: SCOT activity (recombinant enzyme) in the presence of DMSO (vehicle control) and the three DPBPs (500 nmol/L). Data are mean ± SEM. Differences were determined using a two-way ANOVA, followed by Bonferroni post hoc analysis. *$P < 0.05$ vs. vehicle control.

### DPBP-Mediated SCOT Inhibition Reverses Obesity-Induced Hyperglycemia

Male mice were subjected to experimental obesity through HFD supplementation (obese mice), whereas their lean counterparts were provided an LFD (lean mice). Mice were treated with either vehicle control, pimozide, penfluridol, or fluspirilene (each at 10 mg/kg every 48 h through oral gavage) during the final 4 weeks (Fig. 2*A*). All three DPBPs elevated circulating β-hydroxybutyrate (β-OHB) levels in response to an overnight fast, though there was negligible impact on ad libitum/random-fed β-OHB levels (Fig. 2*B*). The improvement in glycemia following an IPGTT previously observed with pimozide (3) was mirrored by penfluridol and fluspirilene (Fig. 2*C*) while associated with lower circulating insulin levels (Fig. 2*D*). Conversely, treatment with all three DPBPs did not improve glycemia following an IPITT (Fig. 2*E* and *F*). Treatment with the DPBPs does not induce hypoglycemia, as glucose profiles were unaffected following an IPGTT in lean mice (Supplementary Fig. 1*A* and *B*).

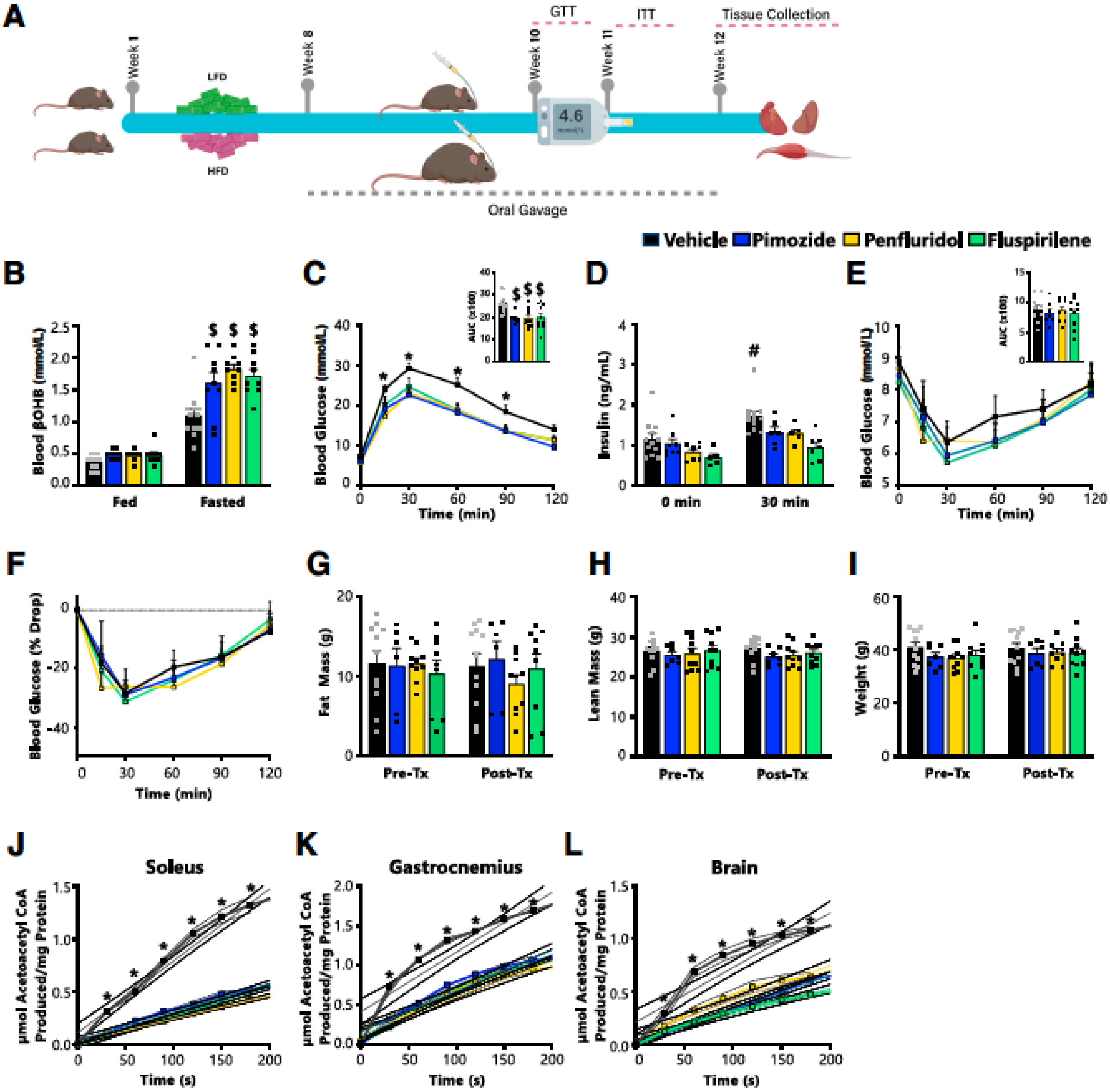


**Figure 2**—DPBPs improve glucose homeostasis in obese male mice. *A*: Schematic representation of the experimental design to determine whether DPBPs impact glycemia in obesity. Eight-week-old male C57BL/6J mice were fed either an HFD (obese mice) or LFD (lean mice) for 8 weeks before receiving pimozide, fluspirilene, and penfluridol (10 mg/kg every 48 h) orally for 4 weeks (week 12). IPGTT was performed at week 10, and IPITT was completed at week 11 before animal euthanasia and tissue collection at the end of the study (week 12). *B*: Circulating β-OHB levels in obese mice in the fed and fasted state (16 h) after 2 weeks of treatment with the DPBPs ($n = 7–10$). *C*: IPGTT and its associated areas under the curve (AUCs) in obese mice. Obese mice were fasted overnight (16 h) before IP administration of glucose (2 g/kg body weight). *D*: Plasma insulin levels before (0 min) and 30 min after IP glucose injection. *E* and *F*: IPITT (0.5 units/kg body weight) and the corresponding percentage drop in the blood glucose levels. *G–I*: Body composition analysis of the obese mice before and after 2 weeks of treatment with either vehicle control or the various DPBPs. Fat mass (*G*), lean mass (*H*), and body weight (*I*) were measured using quantitative nuclear magnetic resonance relaxometry. *J–L*: Measurement of SCOT activity in the isolated tissues of the obese mice treated with either vehicle control or the three DPBPs. The amount of acetoacetyl CoA produced by the endogenous SCOT in the soleus muscle (*J*), gastrocnemius muscle (*K*), and brain (*L*) of the DPBP-treated obese mice is shown. Data are mean ± SEM. Differences were determined using either a one- or two-way ANOVA, followed by Bonferroni post hoc analysis. $\$P < 0.05$ vs. vehicle control–treated counterparts; $\#P < 0.05$ vs. the respective 0-min counterpart; $*P < 0.05$ vs. all other groups. Tx, treatment.

DPBP administration in obese mice demonstrated no impact on fat mass, lean mass, or body weight (Fig. 2*G–I*), which was recapitulated in lean mice (Supplementary Fig. 1*C–E*). Furthermore, all three DPBPs inhibited SCOT activity in skeletal muscle and brain lysates collected at study completion (Fig. 2*J–L*).

### DPBPs Exert Glucose Lowering Through Suppressing Muscle and Brain SCOT Activity

Consistent with pimozide failing to improve glycemia in obese $SCOT^{MuscleKO}$ mice (3) (Supplementary Fig. 2*A–C*), all three DPBPs failed to improve glucose tolerance in obese $SCOT^{MuscleKO}$ mice, whereas they alleviated glucose intolerance in obese human α-skeletal actin-Cre ($HSA^{Cre}$) mice (Fig. 3*A–C*). This improvement in glycemia for all three DPBPs in obese $HSA^{Cre}$ mice was associated with lower circulating insulin levels (Fig. 3*D*), while all three DPBPs also increased circulating β-OHB levels following an overnight fast (Fig. 3*E*).

Ketones are an essential fuel source for the brain during prolonged periods of starvation (15), and brain energy metabolism can regulate whole-body glucose homeostasis (16,17). Thus, we generated brain-specific SCOT KO ($SCOT^{BrainKO}$) mice that lack SCOT activity within the brain (Supplementary Fig. 2*D–F*). Treatment of obese $nestin^{Cre}$ littermates with all three DPBPs once again improved glucose tolerance, whereas treatment with all three DPBPs failed to improve glucose tolerance in obese $SCOT^{BrainKO}$ mice (Fig. 3*F–H*). Although treatment with all three DPBPs was still associated with lower circulating insulin levels in the obese $nestin^{Cre}$ mice, DPBP-treated obese $SCOT^{BrainKO}$ mice consistently had higher circulating insulin levels after glucose administration than their $nestin^{Cre}$ mice counterparts (Fig. 3*I*). In addition, circulating β-OHB levels following an overnight fast were elevated in the obese $SCOT^{BrainKO}$ mice or in response to treatment with the DPBPs in the $nestin^{Cre}$ littermates (Fig. 3*J*).

### $D_2$ Receptor Inhibition Is Dispensable for DPBP-Mediated Glucose Lowering

To assess the potential contribution of the canonical actions of DPBPs as $D_2$ receptor antagonists toward our observations, we next treated obese mice with a single IP injection of a structurally unrelated $D_2$ receptor antagonist, lurasidone (10 mg/kg) (18,19) (Fig. 4*A*). When compared to treatment with pimozide, we observed that lurasidone had no impact on circulating β-OHB levels and was unable to improve glycemia during an IPGTT (Fig. 4*B* and *C*). In addition, lurasidone was devoid of actions to diminish SCOT activity (Fig. 4*D* and *E*). Similarly, treatment of obese mice with lurasidone (10 mg/kg) for 4 weeks yielded identical findings to acute treatment whereby no change in glucose tolerance, circulating β-OHB levels, or SCOT activity was observed (Fig. 4*F–I*).

## DISCUSSION

In this study, we demonstrated that the $D_2$ receptor antagonist DPBPs, originally developed to treat schizophrenia, also inhibit SCOT activity and induce glucose lowering in overnight-fasted obese mice. While this may not be representative of obese individuals with T2D where improvements in glycemia need to ideally be present throughout, we have observed that pimozide also induces glucose lowering with short-term fasting and in the ad libitum state (3). With the use of state-of-the-art in silico modeling techniques, we show that the DPBPs tightly bind to the catalytic pocket of SCOT to inhibit its enzymatic activity.

Our previous studies were limited because they did not rule out a possible role for the canonical antipsychotic actions of pimozide as a $D_2$ receptor antagonist in mediating glucose lowering. We posited that the glucose-lowering actions of pimozide and other DPBPs would be independent of these actions, since glycemia was improved in numerous studies following $D_2$ receptor activation (20–22). Furthermore, the partial $D_2$ receptor agonist bromocriptine exhibits antihyperglycemic effects in people with T2D (23). Nonetheless, to rule out these actions, we treated mice with a structurally unrelated antipsychotic agent that inhibits $D_2$ receptors, lurasidone. We specifically chose lurasidone as its distinct chemical structure should dictate that it does not inhibit SCOT activity (24). The inability of lurasidone to improve glucose tolerance is consistent with our data illustrating that the DPBP-mediated alleviation of obesity-induced hyperglycemia depends entirely on the ability of the DPBPs to inhibit SCOT.

We observed some discrepancies compared with our previous findings with pimozide (3), in particular, treatment with the DPBPs did not improve insulin tolerance in obese mice. This may be explained by a reduced fasting period (6 h) versus the overnight fast implemented in our previous study (3) and is consistent with our data following a 5-h fast, demonstrating that pimozide treatment does not improve skeletal muscle insulin sensitivity in obese mice during a hyperinsulinemic-euglycemic clamp. We posit that noninsulin-mediated glucose disposal accounts for DPBP-mediated glucose lowering. This may be attributed to increased glucose oxidation in muscle secondary to stimulation of pyruvate dehydrogenase (PDH) activity, the rate-limiting enzyme of glucose oxidation. Indeed, it has been suggested that increased mitochondrial oxidation of glucose is sufficient to improve glycemia independent of insulin signaling (25), and we did observe that DPBP treatment of obese mice decreased PDH phosphorylation (indicative of increased PDH activity) in gastrocnemius muscle but not in the brain (Supplementary Fig. 3*A–C*).

One of our surprising observations is that reductions in brain SCOT activity may contribute to the glucose-lowering actions of the DPBPs. Although the glucose clearance profiles of obese $SCOT^{BrainKO}$ mice were comparable to that of obese $nestin^{Cre}$ mice, the DPBPs failed to improve glycemia in $SCOT^{BrainKO}$ mice. As our previous findings revealed

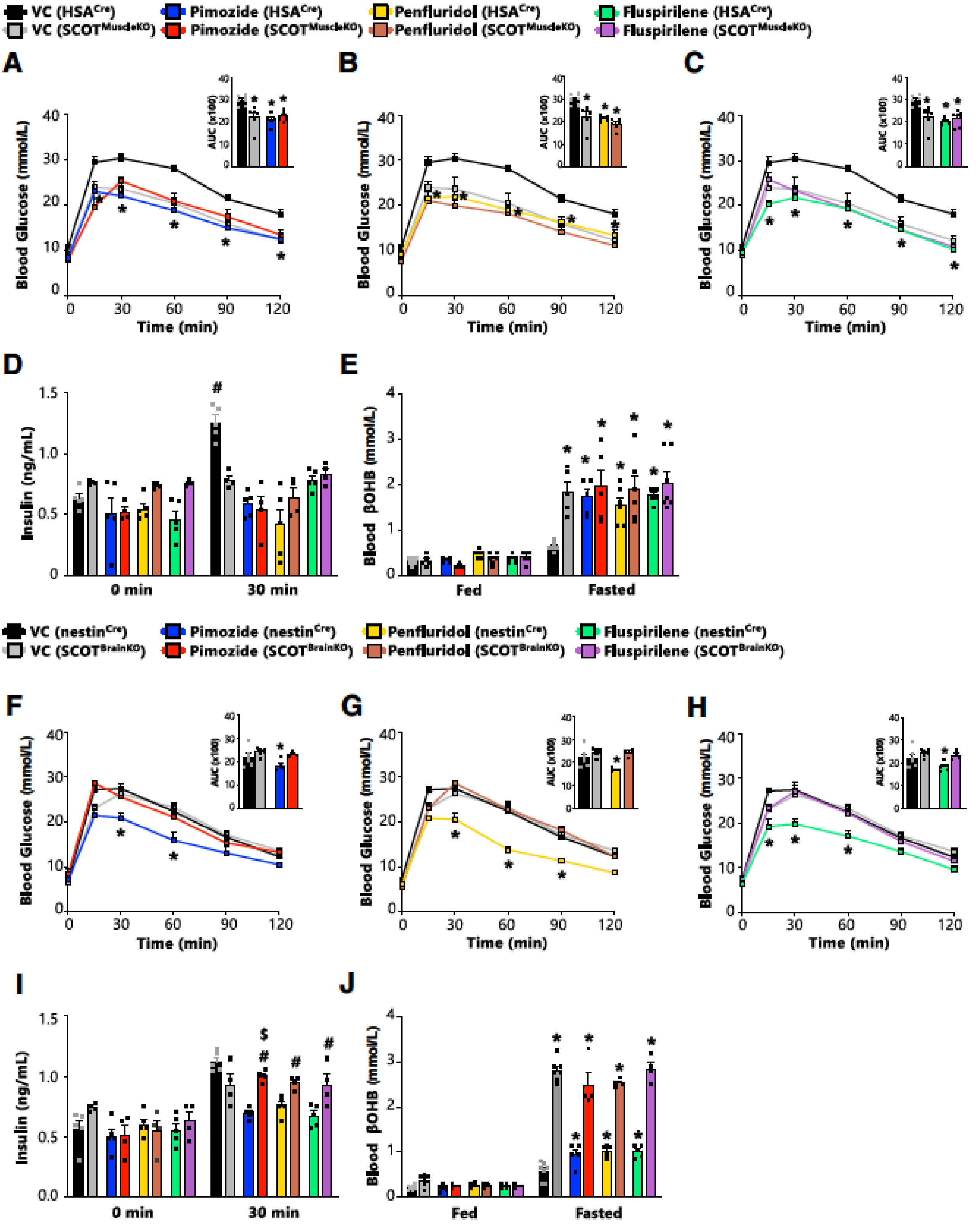


**Figure 3**—The glucose-lowering actions of DPBPs require inhibition of SCOT in both skeletal muscle and brain. *A–D*: IPGTTs in obese SCOT$^{MuscleKO}$ mice and their HSA$^{Cre}$ littermates treated with either vehicle control (VC) or pimozide (*A*), fluspirilene (*B*), or penfluridol (*C*) (10 mg/kg every 48 h) and their corresponding circulating insulin levels at 0 and 30 min (*D*) (*n* = 6–7). *E*: Circulating β-OHB levels in obese SCOT$^{MuscleKO}$ mice and their HSA$^{Cre}$ littermates treated with either VC or the three DPBPs measured before and after a 16-h fast (*n* = 6–7). *F–I*: IPGTTs in obese SCOT$^{BrainKO}$ mice and their nestin$^{Cre}$ littermates treated with either VC or pimozide (*F*), fluspirilene (*G*), or penfluridol (*H*) (10 mg/kg every 48 h) and their corresponding circulating insulin levels at 0 and 30 min (*I*) (*n* = 4–5). *J*: Circulating β-OHB levels in obese

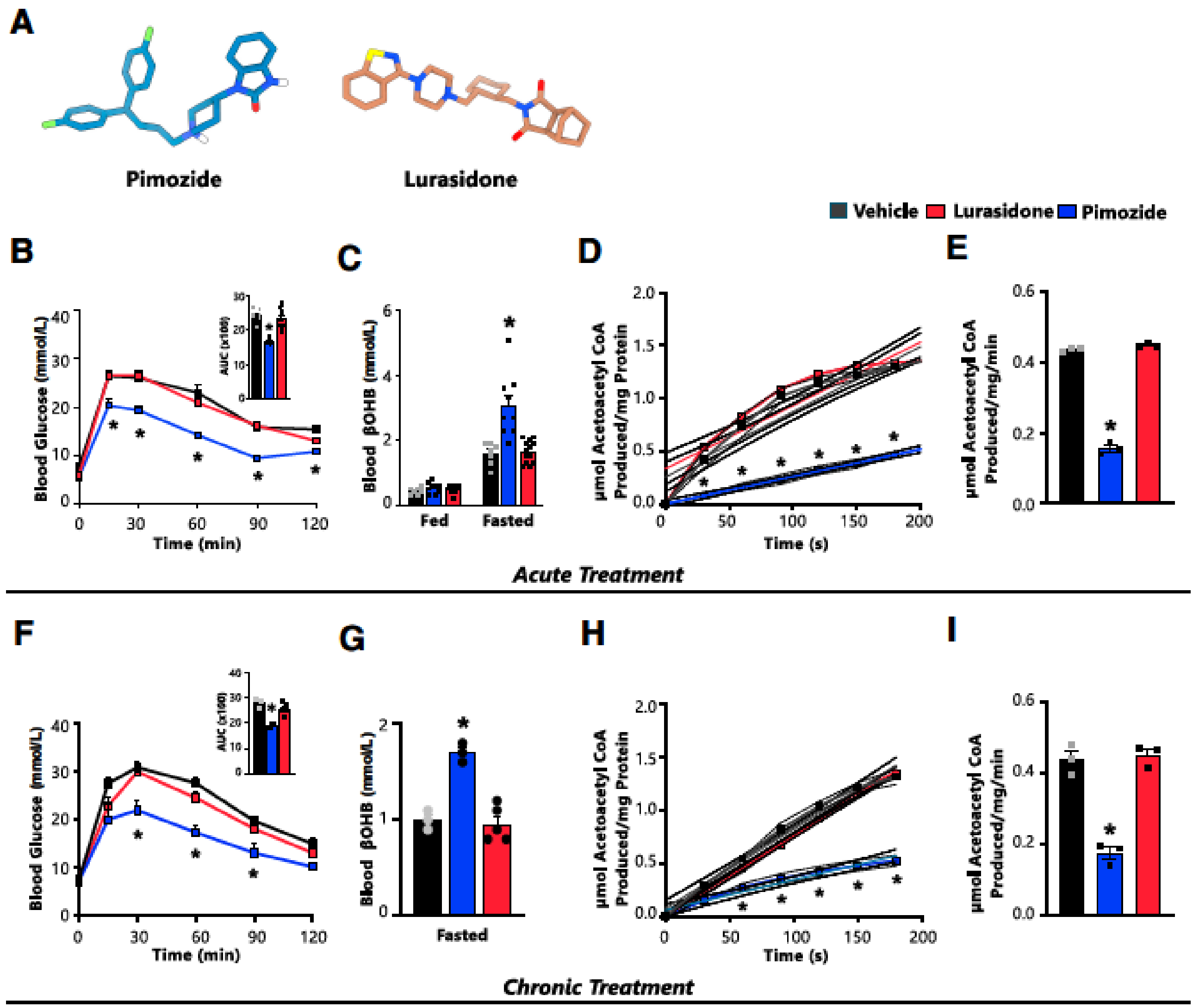


**Figure 4**—$D_2$ receptor antagonism does not improve glycemia in obese male mice. *A*: Structural comparison of pimozide and lurasidone. *B*: IPGTT and its corresponding area under the curve (AUC) in obese mice that received a single IP injection of either vehicle control (corn oil); a structurally dissimilar $D_2$ antagonist, lurasidone (10 mg/kg); or pimozide (10 mg/kg). *C*: Circulating β-OHB levels in each group ($n$ = 7–9). *D* and *E*: SCOT activity (recombinant enzyme) in the presence of DMSO (vehicle), pimozide (500 nmol/L), or lurasidone (500 nmol/L) and its corresponding activity rate ($n$ = 3). *F* and *G*: IPGTT and its corresponding AUC in obese mice treated for 2 weeks with either vehicle control (corn oil), lurasidone (10 mg/kg), or pimozide (10 mg/kg) and circulating β-OHB levels in each group ($n$ = 3–4). *H* and *I*: SCOT activity in gastrocnemius muscle and its corresponding activity rate ($n$ = 3). Data are mean ± SEM. Differences were determined using either a one- or two-way ANOVA, followed by Bonferroni post hoc analysis. *$P < 0.05$ vs. vehicle control–treated obese mice.

that pimozide inhibits SCOT in a noncompetitive manner (3), one potential explanation for the lack of effect in $\text{SCOT}^{\text{BrainKO}}$ mice may be due to the higher circulating ketones in these mice. Subsequent elevations in muscle acetoacetate would overcome the DPBP-mediated inhibition of SCOT, supporting the DPBPs' inability to promote glucose lowering. An alternative explanation could relate to the inhibition of brain SCOT activity playing a regulatory role in whole-body glucose homeostasis. This regulatory role appears to require simultaneous inhibition of brain and muscle SCOT activity/ketone oxidation, since the DPBPs remain capable of inhibiting brain SCOT activity in $\text{SCOT}^{\text{MuscleKO}}$ mice or inhibiting muscle SCOT activity in $\text{SCOT}^{\text{BrainKO}}$ mice. While a potential brain-muscle ketone

$\text{SCOT}^{\text{BrainKO}}$ mice and their $\text{nestin}^{\text{Cre}}$ littermates treated with either VC or the three DPBPs ($n$ = 4–5). Data are mean ± SEM. Differences were determined using either a one- or two-way ANOVA, followed by Bonferroni post hoc analysis. *$P < 0.05$ vs. VC-treated obese $\text{HSA}^{\text{Cre}}$ or obese $\text{nestin}^{\text{Cre}}$ mice; #$P < 0.05$ vs. the respective 0-min counterpart; \$$P < 0.05$ vs. the respective $\text{nestin}^{\text{Cre}}$ counterpart. AUC, area under the curve.

metabolism axis regulating glycemia appears intriguing, there are several concerns regarding nestin$^{\text{Cre}}$ mice and perturbed metabolic phenotypes that make such an interpretation ambiguous. Nestin$^{\text{Cre}}$ mice fed an HFD demonstrate improved insulin sensitivity with no improvement in glucose levels (26). In support of this, our comparisons of nestin$^{\text{Cre}}$ mice, floxed SCOT mice, and their Cre/floxed SCOT–negative wild-type littermates fed an HFD indicate that nestin$^{\text{Cre}}$ mice do exhibit improved insulin sensitivity with no change in fasting glucose levels, as well as a generalized growth defect (Supplementary Fig. 4*A–F*). As the DPBPs were still capable of improving glucose tolerance in nestin$^{\text{Cre}}$ mice but not in SCOT$^{\text{BrainKO}}$ mice, this suggests that brain SCOT activity and ketone oxidation regulate glycemia even if nestin$^{\text{Cre}}$ mice harbor these metabolic phenotypes. However, Cre expression driven by the nestin promoter is also leaky, with reported expression in muscle tissues (26). Although we observed normal SCOT expression in gastrocnemius muscles from SCOT$^{\text{BrainKO}}$ mice, it is possible that SCOT expression may have been decreased in other muscles, which could also account for the failure of the DPBPs to induce glucose lowering in these mice. As such, further interrogation of SCOT$^{\text{BrainKO}}$ mice is still necessary to determine whether brain ketone oxidation plays a critical role in the regulation of glycemia in obesity.

In summary, our study has demonstrated that not only pimozide but the DPBP drug class in general are also SCOT inhibitors. Through these actions and not their ability to inhibit D$_2$ receptors, they induce potent glucose lowering in obesity. While increases in circulating ketone levels secondary to SCOT inhibition may lead to ketoacidosis as a potential adverse effect, the extent of this increase appears comparable to that observed in individuals consuming ketogenic diets and is more reflective of a nutritional ketosis (15). As the DPBPs appear to be relatively safe in humans, where they have been previously approved for the treatment of schizophrenia/psychosis, this drug class may have utility in being repurposed for the treatment of T2D, though careful monitoring of circulating ketones will be necessary.

---

**Acknowledgments.** The in silico modeling studies required access to the Compute Canada and SHARCNET online servers.

**Funding.** This work was supported by a Canadian Institutes of Health Research project grant (to J.R.U.). J.R.U. is a Tier 2 Canada Research Chair (Pharmacotherapy of Energy Metabolism in Obesity).

**Duality of Interest.** No potential conflicts of interest relevant to this article were reported.

**Author Contributions.** S.A.T.D., A.A.G., K.Y., R.A.F., C.T.S., K.G., J.S.F.C., J.J.C., F.E., C.L., and R.A. performed the experiments. S.A.T.D., C.A.V.-M., P.A.C., J.N.M.G., R.A., and J.R.U. reviewed and edited the manuscript. S.A.T.D., R.A., and J.R.U. designed the experiments. S.A.T.D. and J.R.U. analyzed the data and wrote the manuscript. All authors read and approved the manuscript for submission. J.R.U. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

**Prior Presentation.** Parts of this study were presented in abstract form at the 82nd Scientific Sessions of the American Diabetes Association, New Orleans, LA, 3–7 June 2022.

## References

1. Riedel AA, Heien H, Wogen J, Plauschinat CA. Loss of glycemic control in patients with type 2 diabetes mellitus who were receiving initial metformin, sulfonylurea, or thiazolidinedione monotherapy. Pharmacotherapy 2007;27:1102–1110
2. Turner RC, Cull CA, Frighi V; UK Prospective Diabetes Study (UKPDS) Group. Glycemic control with diet, sulfonylurea, metformin, or insulin in patients with type 2 diabetes mellitus: progressive requirement for multiple therapies (UKPDS 49). JAMA 1999;281:2005–2012
3. Al Batran R, Gopal K, Capozzi ME, et al. Pimozide alleviates hyperglycemia in diet-induced obesity by inhibiting skeletal muscle ketone oxidation. Cell Metab 2020;31:909–919.e8
4. Tecott LH, Kwong LL, Uhr S, Peroutka SJ. Differential modulation of dopamine D2 receptors by chronic haloperidol, nitrendipine, and pimozide. Biol Psychiatry 1986;21:1114–1122
5. Carlsson A. Antipsychotic drugs, neurotransmitters, and schizophrenia. Am J Psychiatry 1978;135:165–173
6. Chouinard G, Annable L, Steinberg S. A controlled clinical trial of fluspirilene, a long-acting injectable neuroleptic, in schizophrenic patients with acute exacerbation. J Clin Psychopharmacol 1986;6:21–26
7. Chouinard G, Lehmann HE, Ban TA. Pimozide in the treatment of chronic schizophrenic patients. Curr Ther Res Clin Exp 1970;12:598–603
8. Meyer JM. Antipsychotic safety and efficacy concerns. J Clin Psychiatry 2007;68(Suppl. 14):20–26
9. Pettersen EF, Goddard TD, Huang CC, et al. UCSF Chimera–a visualization system for exploratory research and analysis. J Comput Chem 2004;25:1605–1612
10. Eberhardt J, Santos-Martins D, Tillack AF, Forli S. AutoDock Vina 1.2.0: new docking methods, expanded force field, and Python bindings. J Chem Inf Model 2021;61:3891–3898
11. Pronk S, Páll S, Schulz R, et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 2013;29:845–854
12. Lindorff-Larsen K, Piana S, Palmo K, et al. Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins 2010;78:1950–1958
13. Pettersen EF, Goddard TD, Huang CC, et al. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci 2021;30:70–82
14. Cotter DG, Schugar RC, Wentz AE, d'Avignon DA, Crawford PA. Successful adaptation to ketosis by mice with tissue-specific deficiency of ketone body oxidation. Am J Physiol Endocrinol Metab 2013;304:E363–E374
15. Puchalska P, Crawford PA. Multi-dimensional roles of ketone bodies in fuel metabolism, signaling, and therapeutics. Cell Metab 2017;25:262–284
16. Lam CK, Chari M, Lam TK. CNS regulation of glucose homeostasis. Physiology (Bethesda) 2009;24:159–170
17. Ruud J, Steculorum SM, Brüning JC. Neuronal control of peripheral insulin sensitivity and glucose metabolism. Nat Commun 2017;8:15259
18. Luoni A, Richetto J, Longo L, Riva MA. Chronic lurasidone treatment normalizes GABAergic marker alterations in the dorsal hippocampus of mice exposed to prenatal immune activation. Eur Neuropsychopharmacol 2017;27:170–179
19. Percelay S, Since M, Lagadu S, Freret T, Bouet V, Boulouard M. Antipsychotic lurasidone: behavioural and pharmacokinetic data in C57BL/6 mice. Pharmacol Biochem Behav 2020;194:172933

20. García-Tomadú I, Ornstein AM, Chamson-Reig A, et al. Disruption of the dopamine d2 receptor impairs insulin secretion and causes glucose intolerance. Endocrinology 2010;151:1441–1450

21. Stoelzel CR, Zhang Y, Cincotta AH. Circadian-timed dopamine agonist treatment reverses high-fat diet-induced diabetogenic shift in ventromedial hypothalamic glucose sensing. Endocrinol Diabetes Metab 2020;3:e00139

22. de Leeuw van Weenen JE, Parlevliet ET, Schröder-van der Elst JP, et al. Pharmacological modulation of dopamine receptor D2-mediated transmission alters the metabolic phenotype of diet induced obese and diet resistant C57Bl6 mice. Exp Diabetes Res 2011;2011:928523

23. Defronzo RA. Bromocriptine: a sympatholytic, d2-dopamine agonist for the treatment of type 2 diabetes [published correction appears in Diabetes Care 2011;34:1442]. Diabetes Care 2011;34:789–794

24. Ishibashi T, Horisawa T, Tokuda K, et al. Pharmacological profile of lurasidone, a novel antipsychotic agent with potent 5-hydroxytryptamine 7 (5-HT7) and 5-HT1A receptor activity. J Pharmacol Exp Ther 2010;334:171–181

25. Muoio DM. Metabolic inflexibility: when mitochondrial indecision leads to metabolic gridlock. Cell 2014;159:1253–1262

26. Harno E, Cottrell EC, White A. Metabolic pitfalls of CNS Cre-based technology. Cell Metab 2013;18:21–28

# Appendix C. Calnexin, More than Just a Molecular Chaperone

This chapter was published:

Tautvydas Paskevicius, Rabih Abou Farraj, Marek Michalak, and Luis B. Agellon. Calnexin, More Than Just a Molecular Chaperone. *Cells*.

cells

Review

# Calnexin, More Than Just a Molecular Chaperone

**Tautvydas Paskevicius [1], Rabih Abou Farraj [1], Marek Michalak [1,*] and Luis B. Agellon [2,*]**

[1] Department of Biochemistry, University of Alberta, Edmonton, AB T6G 2R3, Canada
[2] School of Human Nutrition, McGill University, Sainte Anne de Bellevue, QC H9X 3V9, Canada
* Correspondence: marek.michalak@ualberta.ca (M.M.); luis.agellon@mcgill.ca (L.B.A.)

**Abstract:** Calnexin is a type I integral endoplasmic reticulum (ER) membrane protein with an N-terminal domain that resides in the lumen of the ER and a C-terminal domain that extends into the cytosol. Calnexin is commonly referred to as a molecular chaperone involved in the folding and quality control of membrane-associated and secreted proteins, a function that is attributed to its ER-localized domain with a structure that bears a strong resemblance to another luminal ER chaperone and $Ca^{2+}$-binding protein known as calreticulin. Studies have discovered that the cytosolic C-terminal domain of calnexin undergoes distinct post-translational modifications and interacts with a variety of proteins. Here, we discuss recent findings and hypothesize that the post-translational modifications of the calnexin C-terminal domain and its interaction with specific cytosolic proteins play a role in coordinating ER functions with events taking place in the cytosol and other cellular compartments.

**Keywords:** calcium binding protein; cell signaling; endoplasmic reticulum; molecular chaperone; protein–protein interactions

check for updates

**Citation:** Paskevicius, T.; Farraj, R.A.; Michalak, M.; Agellon, L.B. Calnexin, More Than Just a Molecular Chaperone. *Cells* **2023**, *12*, 403. https://doi.org/10.3390/cells12030403

Academic Editor: Dmitry Lim

Received: 4 January 2023
Revised: 17 January 2023
Accepted: 17 January 2023
Published: 24 January 2023

## 1. Introduction

The endoplasmic reticulum (ER) is organized as a continuous membrane network of branching tubules and flattened sacs that envelop a single lumen. The ER performs a plethora of functions in cells, including lipid and steroid synthesis, $Ca^{2+}$ storage and signaling, protein synthesis and maturation involving protein folding and post-translational modification [1–3]. Human cells express approximately 10,000 different proteins at any given moment [4]. More than a third of all of these proteins are synthesized on ER membrane-bound ribosomes where proteins are either destined for residence in the ER, plasma membrane, Golgi apparatus, lysosomes or secreted from the cell [5]. Even though the native structure and conformation of a given protein is largely determined by its amino acid sequence [6], many newly synthesized proteins require assistance by molecular chaperones to reach their native fold at a biologically relevant time scale [7]. To help facilitate proper folding and quality control, the ER employs two major folding systems: the general pathway that is mediated by BiP (the ER homolog of the 70-kDa heat shock protein, Hsp70) together with protein disulfide isomerase PDIA1, and the N-linked glycoprotein pathway which is governed by lectin chaperones calnexin and calreticulin, commonly referred to as the calnexin/calreticulin cycle [8]. Both calreticulin and calnexin share structural similarity with respect to their lectin-like domains with calreticulin being an ER lumen-resident protein while calnexin is a type I integral ER membrane protein with a transmembrane helix and cytosol-exposed C-terminal domain. This review is focused on the role of calnexin not only as an ER chaperone involved in the protein quality control pathway but also on the emerging view on its importance in coordinating ER and cytosolic events via the unique interactions of calnexin with a variety of proteins at the ER–cytosol interface [9–14].

## 2. Discovery of Calnexin

In 1982 a novel ER-associated protein with an apparent mass of 90 kDa was detected using polyclonal antiserum raised against membrane fractions of rough ER [15]. Cell culture immunofluorescence analysis using these anti-rough ER antibodies revealed an extensive, reticular network of membrane occupying the entire cytoplasm and extending to the nuclear membrane [15]. However, it was not until a decade later when the 90 kDa phosphoprotein (referred to as pp90) was identified as being associated with ER signal sequence receptor complexes in canine pancreatic microsomes [16]. Two other research groups simultaneously reported the identification of p88 (reported as an 88 kDa protein) in transient association with partially assembled class I major histocompatibility molecules in murine lymphoma cell lines [17] and IP90 (reported as a 90 kDa intracellular protein) interacting with T-cell antigen and B-cell antigen receptor complexes [18]. Molecular cloning and characterization of the canine IP90 cDNA revealed that its encoded protein was a type I ER membrane protein [16]. Due to its $Ca^{2+}$-binding properties [16] and a high degree of amino acid sequence similarity with calreticulin [19], a major $Ca^{2+}$-binding ER resident protein, the pp90 protein was named calnexin [16,20–22]. Since then, calnexin has been observed to interact transiently with a wide array of nascent membrane or soluble N-linked glycoproteins.

Calnexin, together with calreticulin and ERp57 (also known as PDIA3) [23], forms the core components of a pathway that facilitates the folding and quality control of newly synthesized proteins with N-linked carbohydrate side chains [24,25]. This folding pathway has been widely studied and therefore well characterized and described in several excellent review articles [26–31]. Today it is well established that calnexin is ubiquitously expressed in all cells containing the ER membrane. It is highly conserved among different species (Figure 1), with its intraluminal domain responsible for chaperone function displaying the highest level of conservation and indicating the evolutionary importance of the chaperone domain of calnexin. Calnexin can be found distributed within different ER membrane subdomains, including a variety of ER membrane contact sites [32] such as perinuclear rough ER contacts with the ribosome–translocon complex [33,34], smooth ER, nuclear envelope and the mitochondria/ER contact sites (also referred to as the mitochondria-associated membrane [12,35]. The characteristic distribution of calnexin molecules within ER membranes is controlled by the post-translational modifications including palmitoylation and phosphorylation at its C-terminal domain, which are discussed later. Moreover, it was initially reported that calnexin is present on the cell surface of immature thymocytes in a complex with the CD3 antigen due to incomplete ER retention [36,37]. Other studies have detected small amounts of calnexin on the cell surface of various cell types [38]; the redistribution of calnexin between ER and plasma membranes was proposed to be controlled by the state of calnexin C-terminal domain phosphorylation and association with phosphofurin acidic cluster sorting protein 2 (PACS-2) [39]. Moreover, plasma membrane localization of calnexin has been detected in cancerous tumors such as oral squamous cell carcinoma and melanoma [40] while another study reported calnexin as being secreted in the serum of lung cancer patients, making calnexin a possible sero-diagnostic marker [41].

### *2.1. Calnexin-Deficient Animal Models*

Much of the early work on calnexin focused on biochemical and cellular aspects. Thus, animal models lacking calnexin were created to gain insight into its physiological importance. *D. melanogaster* has three genes encoding calnexin among which calnexin99A has the highest similarity with mammalian calnexin [44]. Mutations in calnexin99A affected the maturation and function of rhodopsin, which in turn led to age-dependent retinal degeneration [44]. Additionally, calnexin99A mutants displayed impaired $Ca^{2+}$ buffering, which contributed to the development of the retinal degeneration possibly due to $Ca^{2+}$ toxicity [44]. Inactivation of the calnexin gene in *C. elegans* resulted in developmental and reproductive defects that were temperature sensitive [45]. These mutant worms also exhibited growth impairment under calcium insufficiency [45]. Furthermore, RNAi-mediated silencing of the calnexin gene resulted in suppressed necrotic-like cell death [46]. In *D. rerio* (zebrafish), calnexin is required for the development of the mechanosensory system called the lateral line [47]. Upon the knockdown of calnexin, zebrafish exhibited reduced posterior lateral line cell proliferation and increased ER stress-dependent apoptosis [47].

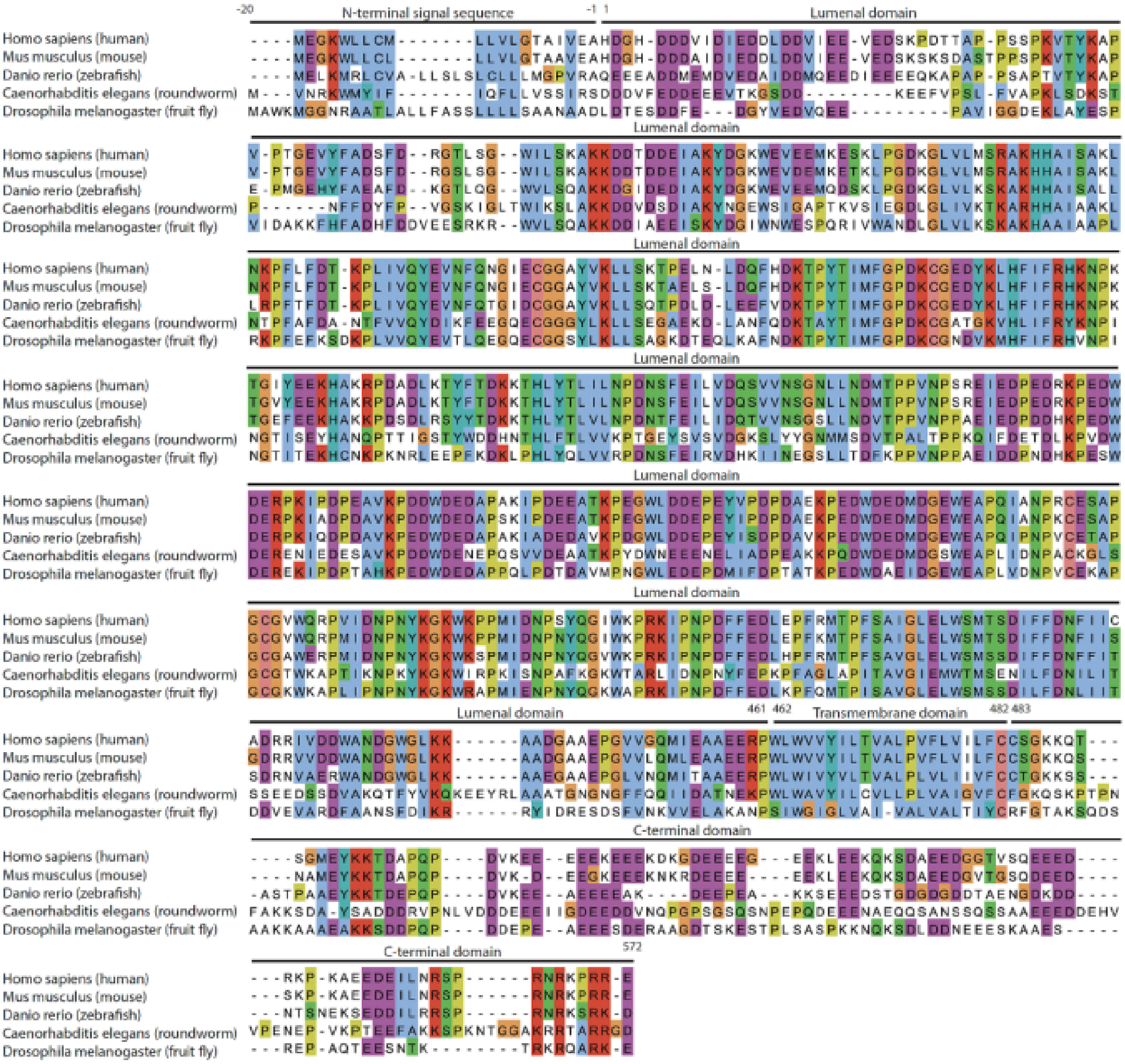


**Figure 1.** Amino acid sequence alignment of calnexin from different species. The default Clustal X color scheme was used for similar/identical amino acid residues [42]: blue for hydrophobic, red for positively charged, magenta for negatively charged, green for polar, cyan for aromatic, pink for cysteine, orange for glycine, yellow for proline and no color for amino acid residues that are not conserved. Dashes represent gaps in the amino acid sequence. The N-terminal signal sequence, intraluminal domain, transmembrane domain and C-terminal domain are indicated based on human calnexin protein topology. The numbering of amino acid residues referred to in the figure as well as the text corresponds to the mature human calnexin protein. Alignment was performed using Multiple Alignment using Fast Fourier Transform (MAFFT) high speed multiple sequence alignment program (https://toolkit.tuebingen.mpg.de/tools/mafft; accessed 31 August 2022). Jalview [43] was used to visualize amino acid alignment. Additional calnexin sequences from different species are shown in Supplementary Figure S1.

Two independent calnexin-deficient mouse strains have been generated [48,49]. Unlike calreticulin deficiency which is embryonic lethal due to impaired $Ca^{2+}$-dependent transcriptional regulation resulting in defective cardiac development [50], calnexin-deficient mice were born live but exhibited a high degree of neonatal lethality [48,49]. Calnexin-deficient mice were born with neurological disorders that included severe ataxia [48,49]; however, one functional calnexin allele is sufficient to prevent this defect [48]. The first study described the high early postnatal mortality of calnexin-deficient mice [49] and surviving mice exhibited ataxia due to the substantial loss of motor nerve fibers. The subsequent study by Kraus et al. [48] found however that surviving calnexin-deficient mice were fertile, had a normal life span, but were 30–50% smaller than their wild-type littermates. These mice developed peripheral neuropathy abnormalities manifesting as gait disturbance with instability, splaying of the hind limbs, tremors and a rolling walk but no reduction in the numbers of neuronal fibers was apparent [48]. Additionally, another mouse strain expressing a mutant form of calnexin-lacking amino acid residues 103–242 (encoded by exons 4–6), which deleted the regions involved in disulfide bond formation ($Cys^{141}$ and $Cys^{175}$) and carbohydrate binding ($Tyr^{145}$ and $Lys^{147}$), exhibited features that were identical to the calnexin-deficient mice, suggesting that the loss of chaperone function was responsible for the observed neurological defect [48,49]. Consistent with this idea, mice that only express the truncated version of calnexin lacking the C-terminal domain do not have apparent disturbances in motor function and display normal motor and sensory nerve conduction velocities [51]. Surprisingly, calnexin-deficient mice display no apparent aberrations in immune system development and function [48].

It is interesting to note that despite the ubiquitous presence of calnexin in all cells that possess an ER network, the loss of calnexin in the whole organism does not produce a common phenotype but rather manifests in a variety of phenotypes.

### *2.2. Calnexin as a Molecular Chaperone*

It is predicted that more than 30% of all eukaryotic proteins are glycoproteins with more than 90% of these containing N-linked sugars [52]. The folding and maturation of newly synthesized glycoproteins in the ER is assisted by calnexin and its soluble ER lumen-resident homolog, calreticulin. Whereas calnexin binds to glycans in protein domains that are close to membranes, calreticulin interacts with glycans that extend deeper into the ER lumen [31]. Immediately after the nascent polypeptide exits the translocon and enters the ER lumen, oligosaccharyltransferase transfers dolichol-pyrophosphate-bound branched core glycan to the sidechain nitrogen of the asparagine residue of the N-glycosylation consensus sequence motif (Asn-X-Ser/Thr, where X is any amino acid except for proline) [53]. The branched core oligosaccharide is comprised of three terminal glucoses, nine mannoses, and two N-acetyl-glucosamines ($Glc_3Man_9GlcNAc_2$). After the glycosylation, N-linked glycans are then processed by the subsequent action of endoplasmic reticulum (ER) glucosidases I (GI) and II (GII) removing the outer and middle glucose residues, respectively. As a result, the processing intermediate containing the single terminal glucose moiety is specifically recognized by calnexin and calreticulin [24,54,55]. Additionally, using their extended arm-like P-domains, calnexin and calreticulin recruit ERp57, cyclophilin B and ERp29 (also known as PDIA9) [56–60] to promote glycoprotein folding and maturation. Subsequently, the remaining innermost glucose moiety is removed by GII releasing the glycoprotein substrate from the calnexin–calreticulin complex. If the protein is correctly folded at this point, it is released from the ER and transferred to the Golgi apparatus to continue its journey along the secretory pathway. However, if the protein has not reached its native three dimensional conformation, it is re-glucosylated by the ER-folding sensor uridine diphosphate (UDP)-glucose:glycoprotein glucosyl transferase (UGGT) facilitating the re-association with calnexin and calreticulin for an additional round of the folding cycle [61]. Thus, the protein can re-enter the calnexin/calreticulin cycle multiple times until the native conformation is reached.

If numerous folding cycles fail to properly fold the protein, the misfolded protein is marked for degradation. Terminally misfolded glycoproteins and unassembled oligomers are retro-translocated to the cytosol and are degraded by the ubiquitin-proteasome system, a process known as ER-associated degradation (ERAD) [62]. This process is regulated by ER α-mannosidase I [63] and ER degradation-enhancing α-mannosidase-like protein (EDEM) [64]. Proteins undergoing multiple calnexin/calreticulin cycles are eventually subjected to mannose trimming by ER α-mannosidase I converting $Man_9GlcNAc_2$-glycans to $Man_8GlcNAc_2$ [65]. The slow kinetics of this enzyme essentially act as a timer for repeated glycoprotein folding cycles [66]. As a result, $Man_8GlcNAc_2$ becomes less efficient UGGT substrate [27] and instead is recognized by EDEM which acts as a signal triggering ERAD [67]. EDEM directly interacts with calnexin and accepts terminally misfolded glycoproteins upon mannose trimming [68,69]. This is followed by retro-translocation into the cytosol where the misfolded proteins are polyubiquitinated which are then degraded by the cytosolic 26S proteasome [70].

In the event of persistent ER stress and the extensive accumulation and aggregation of misfolded proteins, a selective form of autophagy named ER-phagy is used to ensure the timely removal of damaged ER [71]. During this process excessive or damaged portions of ER are fragmented and sequestered through ER-phagy receptors by double-membrane autophagic vesicles that eventually fuse with lysosomes for degradation [71]. It has been shown that calnexin makes a stable complex with the ER-phagy receptor FAM134B [72]. In collagen-producing cells calnexin acts as a co-receptor recognizing misfolded procollagen molecules in the ER lumen triggering FAM134B to recruit and bind the autophagosome membrane-associated protein LC3. In turn, this ER-phagy complex delivers a targeted portion of ER that contains both misfolded procollagen and calnexin to the lysosome for degradation [72]. Additionally, it has been shown that calnexin-FAM134B can facilitate the clearance of proteasome-resistant polymers of α1-antitrypsin Z in ER through a different vesicular transport pathway [73]. This pathway is known as the ER-to-lysosome-associated degradation (ERLAD), since ER-derived vesicles containing misfolded proteins are not encapsulated by autophagosomes but instead fuse with endosomes for degradation [73].

It was initially proposed that calnexin was an important factor in the development of the immune system, considering its importance in the folding and quality control of secreted and membrane-bound glycoproteins. It has been suggested that calnexin is involved in the folding and assembly of major histocompatibility complex (MHC) class I, although one study using a cultured cell model showed that these proteins can fold properly in the absence of calnexin [74–80]. Additionally, calnexin participates in MHC class II [81], T-cell antigen receptor (TCR) [21,82–85] and B-cell antigen receptor (BCR) [86–88] assembly and maturation. However, since mice with whole-body calnexin deficiency display normal immune function, it is apparent that calnexin is not essential for the development of the immune system in this species [48].

## 3. Structure of Calnexin

The gene for human calnexin (*CANX*) is located towards the distal end of the long arm of chromosome 5 (5q35.3 locus) spanning ~33 kbp and comprised of 15 exons [89] (Figure 2). It makes a mature transcript of 4915 bp that is translated into a 592 amino acid residue polypeptide.

Meanwhile, the mouse calnexin gene (*Canx*) is located on a reverse strand of the long arm of murine chromosome 11 and has a similar arrangement but with only 14 exons that are transcribed into a 4281 bp transcript which encodes a polypeptide of 591 amino acid residues. The human calnexin polypeptide is a 67 kDa type-I integral membrane protein but is often mistakenly referred to as a 90 kDa protein due to its high content of acidic residues which electrostatically repel SDS resulting in an insufficient electromotive incentive and a lowered electrophoretic mobility on SDS-PAGE. The calnexin polypeptide is composed of three topological domains (Figures 2 and 3): an N-terminal ER intraluminal domain, a transmembrane segment and a cytosol-facing C-terminal domain [16]. The ER luminal domain is responsible for chaperone function and thus is often referred to as the folding module [56,90]. The transmembrane segment anchors calnexin to the ER membrane and possibly also contributes to its chaperone function [91]. Finally, the 90 amino acid long C-terminal domain is oriented towards the cytosol and undergoes several distinct post-translational modifications [12,33,35,92–97].

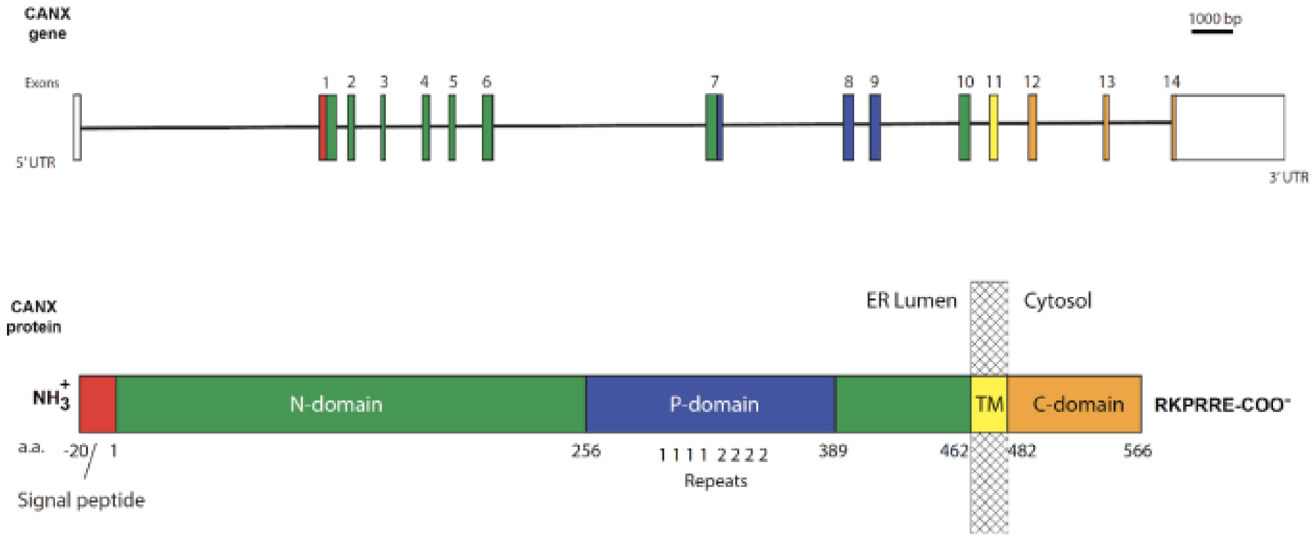


**Figure 2.** Schematic linear sequence of the calnexin gene and its encoded protein. The human calnexin gene is located on a forward strand of distal end of the long arm of chromosome 5 and is comprised of 15 exons (of which 14 are protein-coding exons) transcribed to a mature transcript of 4915 bp that is translated into a 592 amino acid residue polypeptide. Linear schematic representation of the calnexin protein and the corresponding exons encoding the specific protein domains of calnexin. The white boxes in the gene schematic diagram correspond to the untranslated regions of the first and last exons. The calnexin protein schematic diagram shows the signal peptide (red), the luminal domain (green for the N-domain and blue for the P-domain), the transmembrane domain (TM, yellow) and the cytosolic C-terminal domain (orange). The hatched box represents a portion of the ER membrane. Four repeats of Motif 1 and four repeats of Motif 2 are labelled as "11112222" and depict the proline-rich amino acid sequence repeats in the P-domain. The ER retention signal (RKPRRE) is shown as the most distal C-terminal amino acid sequence.

### 3.1. *The ER Lumen-Localized N-Terminal Domain*

The ER luminal domain of calnexin is responsible for its lectin-like chaperone function and is the site for interaction with cyclophilin B, ERp29 and ERp57 [56–60]. It also contains a 20 amino acid residue N-terminal signal sequence that is responsible for targeting calnexin into the ER. The crystal structure of the intraluminal portion of canine calnexin was solved by Schrag et al. [99] at 2.9 Å resolution and revealed the asymmetry featuring two distinct structural components comprised of a compact globular domain towards the N-terminus (referred to as the N-domain) and an elongated arm-like proline-rich domain (referred to as the P-domain) towards the C-terminus (Figure 4) [99]. The N-domain is composed of concave and convex antiparallel β sheets that have six and seven β strands, respectively, which together form a β-sandwich structure [99]. The calnexin luminal domain co-crystallized with α-D-glucose revealed the site for the carbohydrate binding within N-domain on the concave β sheet, where Tyr$^{144}$, Lys$^{146}$, Tyr$^{165}$, Glu$^{196}$ and Glu$^{405}$ (coordinates refer to the human calnexin amino acid sequence) form hydrogen bonds with glucose hydroxyl groups, while the Met$^{168}$ sidechain interacts with the glucose ring via van der Waals interactions [99].

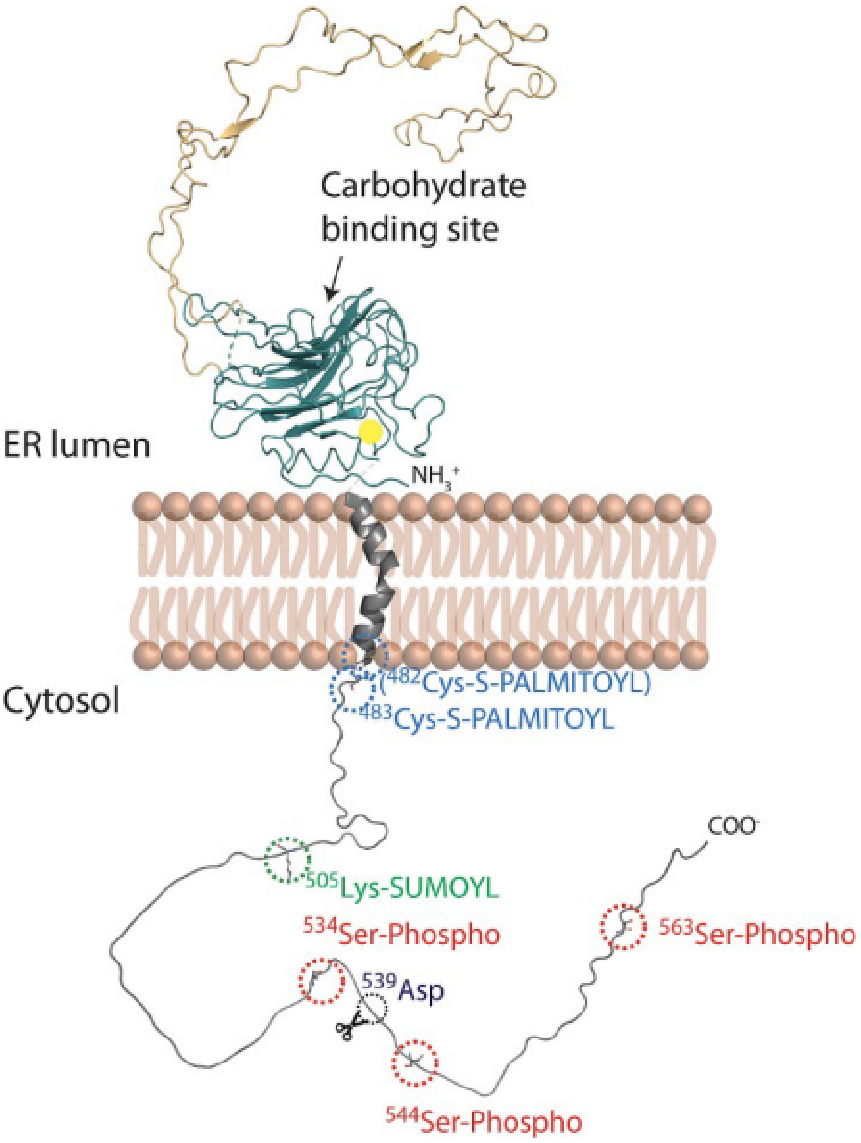


**Figure 3.** Schematic representation of full-length membrane-embedded calnexin. Calnexin (Protein Data Bank DOI: 10.2210/pdb1JHN/pdb) has three domains: luminal domain (comprised of N and P subdomains), transmembrane domain and C-terminal domain (no experimental structural information available). The yellow circle depicts a bound $Ca^{2+}$ ion. The numbering of amino acid residues in the C-terminal domain is relative to the mature N-terminus. The transmembrane domain was modelled based on molecular dynamics simulation performed by Lakkaraju et al. [34] and shows that $Pro^{494}$ introduces a kink in the helix located approximately at the midpoint of the domain. The C-terminal domain was modelled using AlphaFold [98]. The calnexin C-terminal domain undergoes distinct post-translational modifications including palmitoylation at $Cys^{482}$ and $Cys^{483}$ [35] (shown in blue); sumoylation at $Lys^{505}$ [96] (shown in green); and phosphorylation at $Ser^{534}$, $Ser^{544}$ and $Ser^{563}$ [97] (shown in red); $Asp^{539}$ proteolytic cleavage site (shown in black). Known and potential sites of post-translational modifications in the calnexin C-terminal domains of various species are shown in Supplementary Figure S1.

The P-domain [100] extends 140 Å away from the N- domain and forms a large hairpin loop (Figure 4). This loop consists of two types of motifs (Motif 1 and Motif 2) of proline-rich sequence repeats bearing the consensus sequence of I-DP(D/E)A-KPEDWD(D/E) and G-W-P-IN-P-Y, respectively. Each motif is repeated four times and arranged in a linear manner '11112222' where four repeats of motif 1 extend away from the N-globular domain and then fold back onto the strand with four repeats of motif 2. Every copy of motif 1 interacts with a copy of motif 2 in a head-to-tail fashion [99]. This hook-like arm is further stabilized via hydrophobic interactions of conserved isoleucine residues [26]. In addition, there are three regions of high amino acid sequence similarity, flanking the repeat motifs. Both N-globular and P-domain harbors one disulfide bond: $Cys^{140}$–$Cys^{174}$ and $Cys^{340}$–$Cys^{346}$, respectively, in the human calnexin sequence [99].

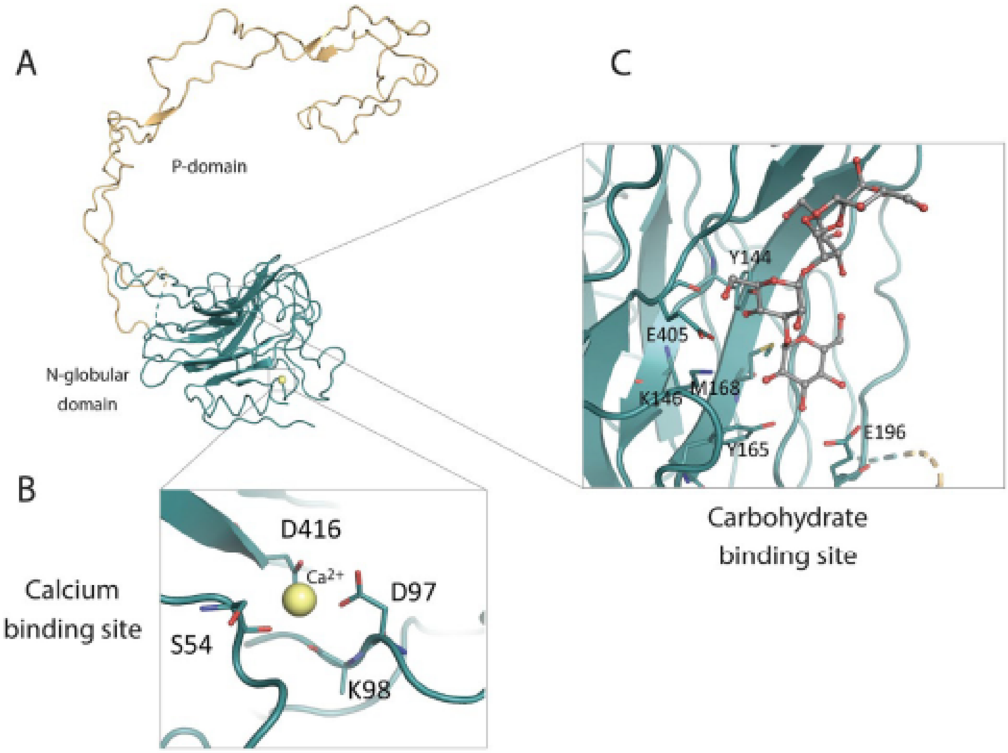


**Figure 4.** Crystal structure of the calnexin luminal domain and its characteristics. (**A**) Crystal structure of the calnexin intraluminal domain. The globular N-globular domain is shown in green while the P-domain is depicted in yellow. (**B**) Calnexin putative $Ca^{2+}$-binding site, showing a $Ca^{2+}$ ion (yellow circle) coordinated by $Asp^{416}$, $Asp^{97}$, $Ser^{54}$ and potentially $Lys^{98}$. (**C**) Calnexin carbohydrate binding site, showing the sidechains of $Tyr^{144}$, $Lys^{146}$, $Tyr^{165}$, $Glu^{196}$, $Glu^{405}$ and $Met^{168}$ involved in the binding of carbohydrate moieties.

Early studies indicated that calnexin harbous multiple low affinity $Ca^{2+}$ binding sites in both the N- and C-terminal regions [16,89]. However, the three-dimensional structure of the luminal domain revealed only a single putative $Ca^{2+}$ binding site coordinated by $Asp^{416}$, $Asp^{97}$ and $Ser^{54}$ [99]. This $Ca^{2+}$ binding site is highly conserved between calnexin and calreticulin, as the P-domain of calreticulin also binds a single $Ca^{2+}$ ion with high affinity [101] through $Asp^{328}$ (refers to human calreticulin; homologous to $Asp^{416}$ in calnexin), $Gln^{26}$, $Lys^{62}$, and $Lys^{64}$, with two water molecules [99]. Moreover, the binding of $Ca^{2+}$ to the ER luminal portion of calnexin plays a structural role by triggering $Ca^{2+}$-dependent conformational changes [102]. In addition, the calnexin luminal domain was shown to bind $Zn^{2+}$ and ATP; both regulate conformational changes [102], while ATP alone enhances the aggregation suppression abilities in vitro, even though no ATPase activity has been reported [103], and $Zn^{2+}$ facilitates the binding of ERp57 [104].

### 3.2. *The Transmembrane Domain*

Using a molecular dynamics simulation, a single transmembrane spanning domain comprised on an α-helix was predicted for calnexin. The $Pro^{494}$ (refers to the human calnexin sequence) at approximately the midpoint of the α-helix introduces a tilt of ~30° with respect to the surface of the membrane (Figure 3) [34]. Replacement of the $Pro^{494}$ with leucine to remove the kink in the transmembrane helix negatively affects the interaction between calnexin and the ribosome–translocon complex [34]. It was proposed that the anchoring of calnexin to the ER membrane facilitates its association with membrane-bound substrates, enhancing its chaperone function [91].

### 3.3. The Cytosolic C-Terminal Domain

Much of the work characterizing the function of calnexin has focused on the intraluminal and transmembrane domain and consequently little is known about the C-terminal domain. It is known that the C-terminal domain plays an important role in the retention of calnexin in the ER as this domain contains the RKPRRE motif, which acts as an ER retention sequence [105]. To date, no structural information about the calnexin C-terminal domain is available. This highly acidic 90 amino acid long segment (theoretical pI of 4.52 for human calnexin) faces the cytosol and is thought to be flexible and unstructured [16,26]. The acidic nature of this domain contributes to the unusual electrophoretic mobility of the calnexin in SDS-PAGE. Based on the number of amino acids encoded by the calnexin mRNA, the predicted molecular mass of calnexin is 67 kDa. However, the apparent molecular mass of calnexin on SDS-PAGE gels is dramatically increased to 90 kDa protein. The anionic character of the calnexin C-terminal domain imparts this domain with multiple low affinity, but high capacity $Ca^{2+}$-binding sites [87].

## 4. Post-Translational Modifications of the Calnexin C-Terminal Domain

Unlike the compact luminal domain of calnexin, the cytosol-exposed C-terminal domain appears to be unstructured which makes it easily accessible. Indeed, recent studies have found that the cytosol-exposed C-terminal domain undergoes post-translational modifications, which include palmitoylation [92–94], phosphorylation, sumoylation [12,33,35,95–97] and proteolytic cleavage [10] (Figure 3).

### 4.1. Palmitoylation

Calnexin is palmitoylated at both juxtamembranous cysteines $Cys^{502}$ and $Cys^{503}$ (or $Cys^{482}$ and $Cys^{483}$ counting from the mature N-terminus) by an ER palmitoyltransferase DHHC6 [34]. Over 90% of calnexin molecules are S-acylated at a steady state suggesting that the cell maintains constant pamitoylation-depalmitoylation cycles of calnexin [34]. Moreover, molecular dynamics simulations predicted that upon palmitoylation, the C-terminal domain adopts different conformations with respect to the transmembrane helix axis suggesting that palmitoylation might affect the capacity and/or selectivity of calnexin to interact with additional proteins outside the ER via its cytosolic C-terminal domain. Palmitoylation at $Cys^{503}$ was predicted to have a more prominent effect on the conformation of the C-terminal domain, suggesting possible functional/regulatory difference between $Cys^{502}$ and $Cys^{503}$ state of palmitoylation. Using both computational and experimental approaches, it has been shown that the calnexin half-life increases 9-fold upon palmitoylation [106]. As a functional consequence, the palmitoylation of both cysteines preferentially localizes calnexin to the perinuclear rough ER while also facilitating the association with the ribosome–translocon complex. This association was shown to be crucial for chaperone function, as calnexin can capture its substrates as they emerge through translocon [34]. Other studies have also shown that palmitoylated calnexin is localized to the mitochondria-ER contact sites [107,108], where it interacts and controls sarco-endoplasmic reticulum $Ca^{2+}$ transport ATPase 2b (SERCA2b) [13], an interaction that modulates ER-mitochondria $Ca^{2+}$ signaling [11,12,35]. It has been shown that upon short-term ER stress, the pool of palmitoylated calnexin is reduced, shifting non-palmitoylated calnexin localization to rough ER where it interacts with ERp57 to facilitate protein folding and quality control [12]. Further studies should help to clarify the impact of calnexin redistribution on the remodeling of cellular processes and the regulation of cellular function.

### 4.2. Phosphorylation

Calnexin harbors three phosphorylation sites in its C-terminal domain: Ser$^{534}$, Ser$^{544}$ and Ser$^{563}$ (in the human calnexin sequence). These sites are known to be phosphorylated by casein kinase CK2 (at Ser$^{534}$ and Ser$^{544}$) [97,109] and extracellular signal-regulated kinase-1 ERK-1 (at Ser$^{563}$) [33]. ERK-1 is activated by mitogen-activated protein kinase 1 MEK1 under conditions that promote protein misfolding [110]; therefore, the phosphorylation state of Ser$^{563}$ exemplifies calnexin function in ER quality control. Upon ER stress, detection of misfolded protein accumulation leads to ERK-1 activation and enhanced calnexin phosphorylation at Ser$^{563}$ which in turn leads to the specific recruitment of calnexin to ER-membrane bound ribosomes. This specific recruitment facilitates calnexin function as a chaperone to enhance glycoprotein quality control at the ER ribosome–translocon complex by prolonging its association with unfolded proteins [33,95]. In synergy with the phosphorylation of Ser$^{563}$, the phosphorylation of Ser$^{534}$ and Ser$^{544}$ by CK2 further promotes calnexin–ribosome interaction [33]. Moreover, the phosphorylation of these two residues disrupts calnexin interaction with phosphofurin acidic cluster sorting protein 2 (PACS-2), a key regulator of mitochondria/ER contact sites [111], thus shifting calnexin distribution from these sites to rough ER [39]. Another study also showed that calcineurin, a Ca$^{2+}$-dependent phosphatase, dephosphorylates Ser$^{563}$ [112] thereby controlling the phosphorylation status of Ser$^{563}$ which was shown to also modulate calnexin interaction with SERCA2b which in turn regulate intracellular Ca$^{2+}$ oscillations [13]. Taken together, phosphorylation of calnexin in its C-terminal domain illustrates another level of complexity in controlling calnexin distribution among different ER membrane subdomains and therefore influences its function.

### 4.3. SUMOylation

Sumoylation is a type of post-translational modification involving the covalent attachment of the small ubiquitin-related modifier (SUMO) protein (~10 kDa) to certain proteins [113]. Calnexin interacts with sumoylation E2 ligase UBC9 via its C-terminal domain and undergoes sumoylation at Lys$^{505}$ [96]. Sumoylated calnexin interacts with protein tyrosine phosphatase 1B PTP1B [96], linking the protein quality control pathway with insulin and leptin signaling [114].

### 4.4. Proteolytic Cleavage

It was previously reported that apoptotic stimuli caused calnexin to undergo proteolytic cleavage at its C-terminal domain by either caspase-3 or caspase-4 at a DXXD site resulting in the attenuation of apoptosis [9]. Another study showed that cells stimulated with epidermal growth factor (EGF) triggers caspase-8-dependent proteolytic cleavage of the calnexin C-terminal domain at Asp$^{519}$ yielding a 63 amino acid peptide [10] that translocates to the nucleus where it makes a stable complex with Protein Inhibitor of Activated STAT 3 (PIAS3) thus preventing PIAS3 from inhibiting Signal Transducer and Activator of Transcription 3 (STAT3). Palmitoylation of calnexin, which targets it to mitochondria/ER contact sites, is required for EGF-induced cleavage of calnexin whereas ER stress prevents its proteolytic cleavage [10].

Despite some variation in the calnexin C-terminal primary structure (Figure S1), the positions of Cys$^{502}$, Cys$^{503}$, Ser$^{563}$, Ser$^{534}$, Ser$^{544}$, Lys$^{505}$, which represent sites for post-translational modifications (Figure 3), are moderately conserved (ranging from 44 to 65%) suggesting similar modifications might also occur in different organisms.

## 5. The Calnexin C-Terminal Domain, a Cytosol-ER Regulatory Nexus

In addition to modifying enzymes, recent studies have found that the C-terminal domain of calnexin interacts with a variety of proteins [14,96,115–118] which, depending on the type of modifications introduced, have a substantial impact on cellular homeostasis and function. Some of the cellular processes affected by the post-translational modification of the C-terminal domain have already been described in the preceding section.

Endocytosis is an example of a cellular process that is altered via the calnexin C-terminal domain. It has been found that the binding of SGIP1 to the calnexin C-terminal domain inhibits clathrin-dependent endocytosis in neuronal cells, and the absence of calnexin in mice causes an increased endocytosis in the nervous system [15]. Cellular efflux of substrate cholesterol is similarly altered via the calnexin C-terminal domain. The binding of the human immunodeficiency virus (HIV) protein Nef to ABCA1, the main cellular cholesterol transporter, disrupts the interaction of calnexin with ABCA1, leading to its retention in the ER and eventual degradation, thereby inhibiting ABCA1-dependent cholesterol efflux [119]. The accumulation of HIV-infected cells accumulate cholesterol in HIV infected cells leads to the increased formation of plasma membrane lipid rafts which serve as sites of HIV entry, assembly, and budding [120]. Moreover, Nef promotes the interaction calnexin with the HIV glycoprotein protein gp160 enhancing HIV envelope protein maturation [119].

Further insight into the pathophysiological relevance of the calnexin C-terminal domain was provided by recent studies on a complex formed by the calnexin C-terminal domain and the cytosolic protein Fabp5 (also referred to as epidermal fatty acid binding protein). This unexpected complex was detected during a search for calnexin interaction partners using the yeast two-hybrid system [118]. While whole-body calnexin deficiency leads to myelinopathy in mice [48], loss of calnexin also causes resistance to induction of experimental autoimmune encephalomyelitis (EAE), a model of inflammatory central nervous system demyelination, coincident with the phenotype of whole-body Fabp5 deficiency [121,122]. In fact, deletion of the calnexin cytosolic C-terminal domain, the site for Fabp5 interaction, is sufficient to impart resistance to EAE induction [51]. This resistance due to the inhibition of circulating T-cell infiltration across endothelial cells of the blood–brain barrier when the formation of the complex between calnexin C-terminal domain and Fabp5 was prevented [51,121]. In contrast, experiments using a cell culture model of the blood–brain barrier demonstrated that promoting the formation of the calnexin/Fabp5 complex enabled T-cells to traverse the endothelial cells of the blood–brain barrier model [51]. It is possible that the stable interaction of Fabp5 with the calnexin C-terminal domain prevents calnexin from interacting with other regulatory proteins or from redistributing to subdomains of the ER membrane system.

The ability of the calnexin C-terminal domain to undergo distinct post-translational modifications and interact with regulatory proteins suggests that the calnexin C-terminal domain is a dynamically modifiable segment of a key ER-resident protein that acts as an interface for facilitating communication between cytosolic and ER processes.

## 6. Summary

Calnexin was first characterized as a molecular chaperone in the ER. The structure of calnexin resembles calreticulin, another ER-resident chaperone and $Ca^{2+}$-binding protein. Calnexin and calreticulin, along with additional accessory ER proteins that include ERp57, participate in the calnexin/calreticulin cycle responsible for the folding and maturation of glycosylated proteins synthesized in the ER some of which are destined for cellular export. Unlike calreticulin however, calnexin is anchored to the ER membrane via its transmembrane domain and its C-terminal domain that extends to the cytosol. Recent studies have discovered that the calnexin C-terminal domain is subject to post-translational modification featuring lipidation, sumoylation, phosphorylation and proteolytic cleavage. These modifications result in the redistribution of calnexin within ER membranes and are associated with the remodelling of cellular processes. Other cytosolic proteins have also been found to interact with the calnexin C-terminal domain and influence cellular function, indicating that these interactions are important in integrating cytosolic and ER events. These findings help pave the way towards the identification and characterization of new calnexin functions, in addition to its well-recognized role as a molecular chaperone.

**Supplementary Materials:** The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cells12030403/s1, Figure S1: Amino acid sequence alignment of calnexin from multiple species.

**Author Contributions:** Conceptualization, T.P., M.M., L.B.A.; writing—original draft preparation, T.P.; Computer modeling, T.P., R.A.F.; writing—review and editing, T.P., R.A.F., M.M., L.B.A.; funding acquisition, M.M., L.B.A. All authors have read and agreed to the published version of the manuscript.

**Funding:** Our research is supported by grants from the Canadian Institutes of Health Research and the Natural Sciences and Engineering Research Council of Canada in our laboratories (to L.B.A. and to M.M.), and a generous donation from the Kenneth and Sheelagh McCourt family and University Hospital Foundation to M.M.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new experimental data were created or analyzed in this study. Data sharing is not applicable to this article.

**Acknowledgments:** T.P. was a recipient of an endMS Personnel Award from the Multiple Sclerosis Society of Canada and by a Women and Children's Health Research Institute Studentship Award.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Fagone, P.; Jackowski, S. Membrane phospholipid synthesis and endoplasmic reticulum function. *J. Lipid Res.* **2009**, *50*, S311–S316. [CrossRef] [PubMed]
2. Braakman, I.; Hebert, D.N. Protein folding in the endoplasmic reticulum. *Cold Spring Harb. Perspect. Biol.* **2013**, *5*, a013201. [CrossRef]
3. Clapham, D.E. Calcium signaling. *Cell* **2007**, *131*, 1047–1058. [CrossRef] [PubMed]
4. Kulak, N.A.; Geyer, P.E.; Mann, M. Loss-less Nano-fractionator for High Sensitivity, High Coverage Proteomics. *Mol. Cell Proteom.* **2017**, *16*, 694–705. [CrossRef] [PubMed]
5. Anelli, T.; Sitia, R. Protein quality control in the early secretory pathway. *EMBO J.* **2008**, *27*, 315–327. [CrossRef] [PubMed]
6. Anfinsen, C.B. Principles that govern the folding of protein chains. *Science* **1973**, *181*, 223–230. [CrossRef]
7. Balchin, D.; Hayer-Hartl, M.; Hartl, F.U. In vivo aspects of protein folding and quality control. *Science* **2016**, *353*, aac4354. [CrossRef] [PubMed]
8. Ellgaard, L.; McCaul, N.; Chatsisvili, A.; Braakman, I. Co- and Post-Translational Protein Folding in the ER. *Traffic* **2016**, *17*, 615–638. [CrossRef]
9. Takizawa, T.; Tatematsu, C.; Watanabe, K.; Kato, K.; Nakanishi, Y. Cleavage of calnexin caused by apoptotic stimuli: Implica-tion for the regulation of apoptosis. *J. Biochem.* **2004**, *136*, 399–405. [CrossRef]
10. Lakkaraju, A.K.; van der Goot, F.G. Calnexin controls the STAT3-mediated transcriptional response to EGF. *Mol. Cell* **2013**, *51*, 386–396. [CrossRef]
11. Gutiérrez, T.; Qi, H.; Yap, M.C.; Tahbaz, N.; Milburn, L.A.; Lucchinetti, E.; Lou, P.H.; Zaugg, M.; LaPointe, P.G.; Mercier, P.; et al. The ER chaperone calnexin controls mitochondrial positioning and respiration. *Sci. Signal.* **2020**, *13*, eaax6660. [CrossRef] [PubMed]
12. Lynes, E.M.; Raturi, A.; Shenkman, M.; Ortiz Sandoval, C.; Yap, M.C.; Wu, J.; Janowicz, A.; Myhill, N.; Benson, M.D.; Camp-bell, R.E.; et al. Palmitoylation is the switch that assigns calnexin to quality control or ER $Ca^{2+}$ signaling. *J. Cell Sci.* **2013**, *126*, 3893–3903. [CrossRef]
13. Roderick, H.L.; Lechleiter, J.D.; Camacho, P. Cytosolic phosphorylation of calnexin controls intracellular $Ca^{2+}$ oscillations via an interaction with SERCA2b. *J. Cell Biol.* **2000**, *149*, 1235–1248. [CrossRef]
14. Li, H.D.; Liu, W.X.; Michalak, M. Enhanced clathrin-dependent endocytosis in the absence of calnexin. *PLoS ONE* **2011**, *6*, e21678. [CrossRef]
15. Louvard, D.; Reggio, H.; Warren, G. Antibodies to the Golgi complex and the rough endoplasmic reticulum. *J. Cell Biol.* **1982**, *92*, 92–107. [CrossRef] [PubMed]
16. Wada, I.; Rindress, D.; Cameron, P.H.; Ou, W.J.; Doherty, J.J., 2nd; Louvard, D.; Bell, A.W.; Dignard, D.; Thomas, D.Y.; Ber-geron, J.J. SSR alpha and associated calnexin are major calcium binding proteins of the endoplasmic reticulum membrane. *J. Biol. Chem.* **1991**, *266*, 19599–19610. [CrossRef] [PubMed]
17. Degen, E.; Williams, D.B. Participation of a novel 88-kD protein in the biogenesis of murine class I histocompatibility mole-cules. *J. Cell Biol.* **1991**, *112*, 1099–1115. [CrossRef] [PubMed]

18. Hochstenbach, F.; David, V.; Watkins, S.; Brenner, M.B. Endoplasmic reticulum resident protein of 90 kilodaltons associates with the T- and B-cell antigen receptors and major histocompatibility complex antigens during their assembly. *Proc. Natl. Acad. Sci. USA* **1992**, *89*, 4734–4738. [CrossRef]
19. Michalak, M.; Groenendyk, J.; Szabo, E.; Gold, L.I.; Opas, M. Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. *Biochem. J.* **2009**, *417*, 651–666. [CrossRef]
20. Ahluwalia, N.; Bergeron, J.J.; Wada, I.; Degen, E.; Williams, D.B. The p88 molecular chaperone is identical to the endoplasmic reticulum membrane protein, calnexin. *J. Biol. Chem.* **1992**, *267*, 10914–10918.
21. David, V.; Hochstenbach, F.; Rajagopalan, S.; Brenner, M.B. Interaction with newly synthesized and retained proteins in the endoplasmic reticulum suggests a chaperone function for human integral membrane protein IP90 (calnexin). *J. Biol. Chem.* **1993**, *268*, 9585–9592. [CrossRef]
22. Cala, S.E.; Ulbright, C.; Kelley, J.S.; Jones, L.R. Purification of a 90-kDa protein (Band VII) from cardiac sarcoplasmic reticu-lum. Identification as calnexin and localization of casein kinase II phosphorylation sites. *J. Biol. Chem.* **1993**, *268*, 2969–2975. [CrossRef] [PubMed]
23. Coe, H.; Michalak, M. ERp57, a multifunctional endoplasmic reticulum resident oxidoreductase. *Int. J. Biochem. Cell Biol.* **2010**, *42*, 796–799. [CrossRef] [PubMed]
24. Hammond, C.; Braakman, I.; Helenius, A. Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 913–917. [CrossRef] [PubMed]
25. Helenius, A. How N-linked oligosaccharides affect glycoprotein folding in the endoplasmic reticulum. *Mol. Biol. Cell* **1994**, *5*, 253–265. [CrossRef] [PubMed]
26. Kozlov, G.; Gehring, K. Calnexin cycle-structural features of the ER chaperone system. *FEBS J.* **2020**, *287*, 4322–4340. [CrossRef]
27. Parodi, A.J. Protein glucosylation and its role in protein folding. *Annu. Rev. Biochem.* **2000**, *69*, 69–93. [CrossRef] [PubMed]
28. Caramelo, J.J.; Parodi, A.J. Getting in and out from calnexin/calreticulin cycles. *J. Biol. Chem.* **2008**, *283*, 10221–10225. [CrossRef]
29. Helenius, A.; Aebi, M. Roles of N-linked glycans in the endoplasmic reticulum. *Annu. Rev. Biochem.* **2004**, *73*, 1019–1049. [CrossRef]
30. Lamriben, L.; Graham, J.B.; Adams, B.M.; Hebert, D.N. N-Glycan-based ER Molecular Chaperone and Protein Quality Con-trol System: The Calnexin Binding Cycle. *Traffic* **2016**, *17*, 308–326. [CrossRef]
31. Hebert, D.N.; Molinari, M. In and out of the ER: Protein folding, quality control, degradation, and related human diseases. *Physiol. Rev.* **2007**, *87*, 1377–1408. [CrossRef] [PubMed]
32. Agellon, L.B.; Michalak, M. The Endoplasmic Reticulum and the Cellular Reticular Network. *Adv. Exp. Med. Biol.* **2017**, *981*, 61–76. [CrossRef] [PubMed]
33. Chevet, E.; Wong, H.N.; Gerber, D.; Cochet, C.; Fazel, A.; Cameron, P.H.; Gushue, J.N.; Thomas, D.Y.; Bergeron, J.J. Phosphorylation by CK2 and MAPK enhances calnexin association with ribosomes. *EMBO J.* **1999**, *18*, 3655–3666. [CrossRef] [PubMed]
34. Lakkaraju, A.K.; Abrami, L.; Lemmin, T.; Blaskovic, S.; Kunz, B.; Kihara, A.; Dal Peraro, M.; van der Goot, F.G. Palmitoylated calnexin is a key component of the ribosome-translocon complex. *EMBO J.* **2012**, *31*, 1823–1835. [CrossRef]
35. Lynes, E.M.; Bui, M.; Yap, M.C.; Benson, M.D.; Schneider, B.; Ellgaard, L.; Berthiaume, L.G.; Simmen, T. Palmitoylated TMX and calnexin target to the mitochondria-associated membrane. *EMBO J.* **2012**, *31*, 457–470. [CrossRef]
36. Wiest, D.L.; Burgess, W.H.; McKean, D.; Kearse, K.P.; Singer, A. The molecular chaperone calnexin is expressed on the sur-face of immature thymocytes in association with clonotype-independent CD3 complexes. *EMBO J.* **1995**, *14*, 3425–3433. [CrossRef]
37. Wiest, D.L.; Bhandoola, A.; Punt, J.; Kreibich, G.; McKean, D.; Singer, A. Incomplete endoplasmic reticulum (ER) retention in immature thymocytes as revealed by surface expression of "ER-resident" molecular chaperones. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 1884–1889. [CrossRef]
38. Okazaki, Y.; Ohno, H.; Takase, K.; Ochiai, T.; Saito, T. Cell surface expression of calnexin, a molecular chaperone in the endo-plasmic reticulum. *J. Biol. Chem.* **2000**, *275*, 35751–35758. [CrossRef]
39. Myhill, N.; Lynes, E.M.; Nanji, J.A.; Blagoveshchenskaya, A.D.; Fei, H.; Carmine Simmen, K.; Cooper, T.J.; Thomas, G.; Sim-men, T. The subcellular distribution of calnexin is mediated by PACS-2. *Mol. Biol. Cell* **2008**, *19*, 2777–2788. [CrossRef]
40. Chen, Y.; Ma, D.; Wang, X.; Fang, J.; Liu, X.; Song, J.; Li, X.; Ren, X.; Li, Q.; Li, Q.; et al. Calnexin Impairs the Antitumor Im-munity of CD4+ and CD8+ T Cells. *Cancer Immunol. Res.* **2019**, *7*, 123–135. [CrossRef]
41. Kobayashi, M.; Nagashio, R.; Jiang, S.X.; Saito, K.; Tsuchiya, B.; Ryuge, S.; Katono, K.; Nakashima, H.; Fukuda, E.; Goshima, N.; et al. Calnexin is a novel sero-diagnostic marker for lung cancer. *Lung Cancer* **2015**, *90*, 342–345. [CrossRef] [PubMed]
42. Thompson, J.D.; Gibson, T.J.; Plewniak, F.; Jeanmougin, F.; Higgins, D.G. The CLUSTAL_X windows interface: Flexible strat-egies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **1997**, *25*, 4876–4882. [CrossRef] [PubMed]
43. Waterhouse, A.M.; Procter, J.B.; Martin, D.M.; Clamp, M.; Barton, G.J. Jalview Version 2—A multiple sequence alignment edi-tor and analysis workbench. *Bioinformatics* **2009**, *25*, 1189–1191. [CrossRef] [PubMed]
44. Rosenbaum, E.E.; Hardie, R.C.; Colley, N.J. Calnexin is essential for rhodopsin maturation, $Ca^{2+}$ regulation, and photorecep-tor cell survival. *Neuron* **2006**, *49*, 229–241. [CrossRef] [PubMed]
45. Lee, W.; Lee, T.H.; Park, B.J.; Chang, J.W.; Yu, J.R.; Koo, H.S.; Park, H.; Yoo, Y.J.; Ahnn, J. Caenorhabditis elegans calnexin is N-glycosylated and required for stress response. *Biochem. Biophys. Res. Commun.* **2005**, *338*, 1018–1030. [CrossRef]

46. Xu, K.; Tavernarakis, N.; Driscoll, M. Necrotic cell death in C. elegans requires the function of calreticulin and regulators of $Ca^{2+}$ release from the endoplasmic reticulum. *Neuron* **2001**, *31*, 957–971. [CrossRef]
47. Hung, I.C.; Cherng, B.W.; Hsu, W.M.; Lee, S.J. Calnexin is required for zebrafish posterior lateral line development. *Int. J. Dev. Biol.* **2013**, *57*, 427–438. [CrossRef]
48. Kraus, A.; Groenendyk, J.; Bedard, K.; Baldwin, T.A.; Krause, K.-H.; Dubois-Dauphin, M.; Dyck, J.; Rosenbaum, E.E.; Korn-gut, L.; Colley, N.J.; et al. Calnexin deficiency leads to dysmyelination. *J. Biol. Chem.* **2010**, *285*, 18928–18938. [CrossRef]
49. Denzel, A.; Molinari, M.; Trigueros, C.; Martin, J.E.; Velmurgan, S.; Brown, S.; Stamp, G.; Owen, M.J. Early postnatal death and motor disorders in mice congenitally deficient in calnexin expression. *Mol. Cell Biol.* **2002**, *22*, 7398–7404. [CrossRef]
50. Mesaeli, N.; Nakamura, K.; Zvaritch, E.; Dickie, P.; Dziak, E.; Krause, K.H.; Opas, M.; MacLennan, D.H.; Michalak, M. Calre-ticulin is essential for cardiac development. *J. Cell Biol.* **1999**, *144*, 857–868. [CrossRef]
51. Paskevicius, T.; Jung, J.; Pujol, M.; Eggleton, P.; Qin, W.; Robinson, A.; Gutowski, N.; Holley, J.; Smallwood, M.; Newcombe, J.; et al. The Fabp5/calnexin complex is a prerequisite for sensitization of mice to experimental autoimmune encephalomyeli-tis. *FASEB J.* **2020**, *34*, 16662–16675. [CrossRef]
52. Apweiler, R.; Hermjakob, H.; Sharon, N. On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim. Biophys. Acta* **1999**, *1473*, 4–8. [CrossRef] [PubMed]
53. Kornfeld, R.; Kornfeld, S. Assembly of asparagine-linked oligosaccharides. *Annu. Rev. Biochem.* **1985**, *54*, 631–664. [CrossRef] [PubMed]
54. Hebert, D.N.; Foellmer, B.; Helenius, A. Glucose trimming and reglucosylation determine glycoprotein association with cal-nexin in the endoplasmic reticulum. *Cell* **1995**, *81*, 425–433. [CrossRef] [PubMed]
55. Rodan, A.R.; Simons, J.F.; Trombetta, E.S.; Helenius, A. N-linked oligosaccharides are necessary and sufficient for association of glycosylated forms of bovine RNase with calnexin and calreticulin. *EMBO J.* **1996**, *15*, 6921–6930. [CrossRef] [PubMed]
56. Zapun, A.; Darby, N.J.; Tessier, D.C.; Michalak, M.; Bergeron, J.J.; Thomas, D.Y. Enhanced catalysis of ribonuclease B folding by the interaction of calnexin or calreticulin with ERp57. *J. Biol. Chem.* **1998**, *273*, 6009–6012. [CrossRef] [PubMed]
57. Frickel, E.M.; Riek, R.; Jelesarov, I.; Helenius, A.; Wuthrich, K.; Ellgaard, L. TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 1954–1959. [CrossRef] [PubMed]
58. Kozlov, G.; Bastos-Aristizabal, S.; Määttänen, P.; Rosenauer, A.; Zheng, F.; Killikelly, A.; Trempe, J.F.; Thomas, D.Y.; Gehring, K. Structural basis of cyclophilin B binding by the calnexin/calreticulin P-domain. *J. Biol. Chem.* **2010**, *285*, 35551–35557. [CrossRef]
59. Kozlov, G.; Muñoz-Escobar, J.; Castro, K.; Gehring, K. Mapping the ER Interactome: The P Domains of Calnexin and Calre-ticulin as Plurivalent Adapters for Foldases and Chaperones. *Structure* **2017**, *25*, 1415–1422.e1413. [CrossRef]
60. Sakono, M.; Seko, A.; Takeda, Y.; Ito, Y. PDI family protein ERp29 forms 1:1 complex with lectin chaperone calreticulin. *Biochem. Biophys. Res. Commun.* **2014**, *452*, 27–31. [CrossRef]
61. Tannous, A.; Patel, N.; Tamura, T.; Hebert, D.N. Reglucosylation by UDP-glucose:glycoprotein glucosyltransferase 1 delays glycoprotein secretion but not degradation. *Mol. Biol. Cell* **2015**, *26*, 390–405. [CrossRef] [PubMed]
62. Plemper, R.K.; Wolf, D.H. Retrograde protein translocation: ERADication of secretory proteins in health and disease. *Trends Biochem. Sci.* **1999**, *24*, 266–270. [CrossRef] [PubMed]
63. Cabral, C.M.; Choudhury, P.; Liu, Y.; Sifers, R.N. Processing by endoplasmic reticulum mannosidases partitions a secretion-impaired glycoprotein into distinct disposal pathways. *J. Biol. Chem.* **2000**, *275*, 25015–25022. [CrossRef] [PubMed]
64. Hosokawa, N.; Wada, I.; Hasegawa, K.; Yorihuzi, T.; Tremblay, L.O.; Herscovics, A.; Nagata, K. A novel ER alpha-mannosidase-like protein accelerates ER-associated degradation. *EMBO Rep.* **2001**, *2*, 415–422. [CrossRef]
65. Tremblay, L.O.; Herscovics, A. Cloning and expression of a specific human alpha 1,2-mannosidase that trims $Man_9GlcNAc_2$ to $Man_8GlcNAc_2$ isomer B during N-glycan biosynthesis. *Glycobiology* **1999**, *9*, 1073–1078. [CrossRef]
66. Meusser, B.; Hirsch, C.; Jarosch, E.; Sommer, T. ERAD: The long road to destruction. *Nat. Cell Biol.* **2005**, *7*, 766–772. [CrossRef]
67. Jakob, C.A.; Burda, P.; Roth, J.; Aebi, M. Degradation of misfolded endoplasmic reticulum glycoproteins in Saccharomyces cerevisiae is determined by a specific oligosaccharide structure. *J. Cell Biol.* **1998**, *142*, 1223–1233. [CrossRef]
68. Oda, Y.; Hosokawa, N.; Wada, I.; Nagata, K. EDEM as an acceptor of terminally misfolded glycoproteins released from cal-nexin. *Science* **2003**, *299*, 1394–1397. [CrossRef] [PubMed]
69. Molinari, M.; Calanca, V.; Galli, C.; Lucca, P.; Paganetti, P. Role of EDEM in the release of misfolded glycoproteins from the calnexin cycle. *Science* **2003**, *299*, 1397–1400. [CrossRef]
70. Christianson, J.C.; Ye, Y. Cleaning up in the endoplasmic reticulum: Ubiquitin in charge. *Nat. Struct. Mol. Biol.* **2014**, *21*, 325–335. [CrossRef] [PubMed]
71. Yang, M.; Luo, S.; Wang, X.; Li, C.; Yang, J.; Zhu, X.; Xiao, L.; Sun, L. ER-Phagy: A New Regulator of ER Homeostasis. *Front. Cell Dev. Biol.* **2021**, *9*, 684526. [CrossRef] [PubMed]
72. Forrester, A.; De Leonibus, C.; Grumati, P.; Fasana, E.; Piemontese, M.; Staiano, L.; Fregno, I.; Raimondi, A.; Marazza, A.; Bruno, G.; et al. A selective ER-phagy exerts procollagen quality control via a Calnexin-FAM134B complex. *EMBO J.* **2019**, *38*, e99847. [CrossRef]
73. Fregno, I.; Fasana, E.; Bergmann, T.J.; Raimondi, A.; Loi, M.; Soldà, T.; Galli, C.; D'Antuono, R.; Morone, D.; Danieli, A.; et al. ER-to-lysosome-associated degradation of proteasome-resistant ATZ polymers occurs via receptor-mediated vesicular transport. *EMBO J.* **2018**, *37*, e99259. [CrossRef] [PubMed]

74. Diedrich, G.; Bangia, N.; Pan, M.; Cresswell, P. A role for calnexin in the assembly of the MHC class I loading complex in the endoplasmic reticulum. *J. Immunol.* **2001**, *166*, 1703–1709. [CrossRef]
75. Jackson, M.R.; Cohen-Doyle, M.F.; Peterson, P.A.; Williams, D.B. Regulation of MHC class I transport by the molecular chap-erone, calnexin (p88, IP90). *Science* **1994**, *263*, 384–387. [CrossRef] [PubMed]
76. Suh, W.K.; Mitchell, E.K.; Yang, Y.; Peterson, P.A.; Waneck, G.L.; Williams, D.B. MHC class I molecules form ternary com-plexes with calnexin and TAP and undergo peptide-regulated interaction with TAP via their extracellular domains. *J. Exp. Med.* **1996**, *184*, 337–348. [CrossRef]
77. Zhang, Q.; Tector, M.; Salter, R.D. Calnexin recognizes carbohydrate and protein determinants of class I major histocompati-bility complex molecules. *J. Biol. Chem.* **1995**, *270*, 3944–3948. [CrossRef]
78. Vassilakos, A.; Cohen-Doyle, M.F.; Peterson, P.A.; Jackson, M.R.; Williams, D.B. The molecular chaperone calnexin facilitates folding and assembly of class I histocompatibility molecules. *EMBO J.* **1996**, *15*, 1495–1506. [CrossRef]
79. Prasad, S.A.; Yewdell, J.W.; Porgador, A.; Sadasivan, B.; Cresswell, P.; Bennink, J.R. Calnexin expression does not enhance the generation of MHC class I-peptide complexes. *Eur. J. Immunol.* **1998**, *28*, 907–913. [CrossRef]
80. Sadasivan, B.K.; Cariappa, A.; Waneck, G.L.; Cresswell, P. Assembly, peptide loading, and transport of MHC class I mole-cules in a calnexin-negative cell line. *Cold Spring Harb. Symp. Quant. Biol.* **1995**, *60*, 267–275. [CrossRef]
81. Anderson, K.S.; Cresswell, P. A role for calnexin (IP90) in the assembly of class II MHC molecules. *EMBO J.* **1994**, *13*, 675–682. [CrossRef] [PubMed]
82. Kearse, K.P.; Williams, D.B.; Singer, A. Persistence of glucose residues on core oligosaccharides prevents association of TCR alpha and TCR beta proteins with calnexin and results specifically in accelerated degradation of nascent TCR alpha proteins within the endoplasmic reticulum. *EMBO J.* **1994**, *13*, 3678–3686. [CrossRef]
83. Van Leeuwen, J.E.; Kearse, K.P. Calnexin associates exclusively with individual CD3 delta and T cell antigen receptor (TCR) alpha proteins containing incompletely trimmed glycans that are not assembled into multisubunit TCR complexes. *J. Biol. Chem.* **1996**, *271*, 9660–9665. [CrossRef] [PubMed]
84. Gardner, T.G.; Franklin, R.A.; Robinson, P.J.; Pederson, N.E.; Howe, C.; Kearse, K.P. T cell receptor assembly and expression in the absence of calnexin. *Arch. Biochem. Biophys.* **2000**, *378*, 182–189. [CrossRef]
85. Van Leeuwen, J.E.; Kearse, K.P. The related molecular chaperones calnexin and calreticulin differentially associate with nas-cent T cell antigen receptor proteins within the endoplasmic reticulum. *J. Biol. Chem.* **1996**, *271*, 25345–25349. [CrossRef] [PubMed]
86. Grupp, S.A.; Mitchell, R.N.; Schreiber, K.L.; McKean, D.J.; Abbas, A.K. Molecular mechanisms that control expression of the B lymphocyte antigen receptor complex. *J. Exp. Med.* **1995**, *181*, 161–168. [CrossRef] [PubMed]
87. Wu, Y.; Pun, C.; Hozumi, N. Roles of calnexin and Ig-alpha beta interactions with membrane Igs in the surface expression of the B cell antigen receptor of the IgM and IgD classes. *J. Immunol.* **1997**, *158*, 2762–2770. [CrossRef]
88. Foy, S.P.; Matsuuchi, L. Association of B lymphocyte antigen receptor polypeptides with multiple chaperone proteins. *Immunol. Lett.* **2001**, *78*, 149–160. [CrossRef]
89. Tjoelker, L.W.; Seyfried, C.E.; Eddy, R.L., Jr.; Byers, M.G.; Shows, T.B.; Calderon, J.; Schreiber, R.B.; Gray, P.W. Human, mouse, and rat calnexin cDNA cloning: Identification of potential calcium binding motifs and gene localization to human chromosome 5. *Biochemistry* **1994**, *33*, 3229–3236. [CrossRef]
90. Oliver, J.D.; Roderick, H.L.; Llewellyn, D.H.; High, S. ERp57 functions as a subunit of specific complexes formed with the ER lectins calreticulin and calnexin. *Mol. Biol. Cell* **1999**, *10*, 2573–2582. [CrossRef]
91. Ho, S.C.; Rajagopalan, S.; Chaudhuri, S.; Shieh, C.C.; Brenner, M.B.; Pillai, S. Membrane anchoring of calnexin facilitates its interaction with its targets. *Mol. Immunol.* **1999**, *36*, 1–12. [CrossRef] [PubMed]
92. Dowal, L.; Yang, W.; Freeman, M.R.; Steen, H.; Flaumenhaft, R. Proteomic analysis of palmitoylated platelet proteins. *Blood* **2011**, *118*, e62–e73. [CrossRef] [PubMed]
93. Kang, R.; Wan, J.; Arstikaitis, P.; Takahashi, H.; Huang, K.; Bailey, A.O.; Thompson, J.X.; Roth, A.F.; Drisdel, R.C.; Mastro, R.; et al. Neural palmitoyl-proteomics reveals dynamic synaptic palmitoylation. *Nature* **2008**, *456*, 904–909. [CrossRef] [PubMed]
94. Yount, J.S.; Moltedo, B.; Yang, Y.Y.; Charron, G.; Moran, T.M.; López, C.B.; Hang, H.C. Palmitoylome profiling reveals S-palmitoylation-dependent antiviral activity of IFITM3. *Nat. Chem. Biol.* **2010**, *6*, 610–614. [CrossRef]
95. Cameron, P.H.; Chevet, E.; Pluquet, O.; Thomas, D.Y.; Bergeron, J.J. Calnexin phosphorylation attenuates the release of par-tially misfolded alpha1-antitrypsin to the secretory pathway. *J. Biol. Chem.* **2009**, *284*, 34570–34579. [CrossRef] [PubMed]
96. Lee, D.; Kraus, A.; Prins, D.; Groenendyk, J.; Aubry, I.; Liu, W.-X.; Li, H.-D.; Julien, O.; Touret, N.; Sykes, B.D.; et al. UBC9-dependent association between calnexin and protein tyrosine phosphatase 1B (PTP1B) at the endoplasmic reticulum. *J. Biol. Chem.* **2015**, *290*, 5725–5738. [CrossRef]
97. Wong, H.N.; Ward, M.A.; Bell, A.W.; Chevet, E.; Bains, S.; Blackstock, W.P.; Solari, R.; Thomas, D.Y.; Bergeron, J.J. Conserved in vivo phosphorylation of calnexin at casein kinase II sites as well as a protein kinase C/proline-directed kinase site. *J. Biol. Chem.* **1998**, *273*, 17227–17235. [CrossRef] [PubMed]
98. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]
99. Schrag, J.D.; Bergeron, J.J.; Li, Y.; Borisova, S.; Hahn, M.; Thomas, D.Y.; Cygler, M. The Structure of calnexin, an ER chaper-one involved in quality control of protein folding. *Mol. Cell* **2001**, *8*, 633–644. [CrossRef]
100. Trombetta, E.S.; Helenius, A. Lectins as chaperones in glycoprotein folding. *Curr. Opin. Struct. Biol.* **1998**, *8*, 587–592. [CrossRef]

101. Baksh, S.; Michalak, M. Expression of calreticulin in Escherichia coli and identification of its $Ca^{2+}$ binding domains. *J. Biol. Chem.* **1991**, *266*, 21458–21465. [CrossRef] [PubMed]
102. Ou, W.J.; Bergeron, J.J.; Li, Y.; Kang, C.Y.; Thomas, D.Y. Conformational changes induced in the endoplasmic reticulum lu-minal domain of calnexin by Mg-ATP and $Ca^{2+}$. *J. Biol. Chem.* **1995**, *270*, 18051–18059. [CrossRef]
103. Ihara, Y.; Cohen-Doyle, M.F.; Saito, Y.; Williams, D.B. Calnexin discriminates between protein conformational states and functions as a molecular chaperone in vitro. *Mol. Cell* **1999**, *4*, 331–341. [CrossRef] [PubMed]
104. Leach, M.R.; Cohen-Doyle, M.F.; Thomas, D.Y.; Williams, D.B. Localization of the lectin, ERp57 binding, and polypeptide binding sites of calnexin and calreticulin. *J. Biol. Chem.* **2002**, *277*, 29686–29697. [CrossRef] [PubMed]
105. Rajagopalan, S.; Xu, Y.; Brenner, M.B. Retention of unassembled components of integral membrane proteins by calnexin. *Science* **1994**, *263*, 387–390. [CrossRef]
106. Dallavilla, T.; Abrami, L.; Sandoz, P.A.; Savoglidis, G.; Hatzimanikatis, V.; van der Goot, F.G. Model-Driven Understanding of Palmitoylation Dynamics: Regulated Acylation of the Endoplasmic Reticulum Chaperone Calnexin. *PLoS Comput. Biol.* **2016**, *12*, e1004774. [CrossRef]
107. Hayashi, T.; Su, T.P. Sigma-1 receptor chaperones at the ER-mitochondrion interface regulate $Ca^{2+}$ signaling and cell survival. *Cell* **2007**, *131*, 596–610. [CrossRef]
108. De Brito, O.M.; Scorrano, L. An intimate liaison: Spatial organization of the endoplasmic reticulum-mitochondria relation-ship. *EMBO J.* **2010**, *29*, 2715–2723. [CrossRef]
109. Ou, W.J.; Thomas, D.Y.; Bell, A.W.; Bergeron, J.J. Casein kinase II phosphorylation of signal sequence receptor alpha and the associated membrane chaperone calnexin. *J. Biol. Chem.* **1992**, *267*, 23789–23796. [CrossRef]
110. Nguyên, D.T.; Kebache, S.; Fazel, A.; Wong, H.N.; Jenna, S.; Emadali, A.; Lee, E.H.; Bergeron, J.J.; Kaufman, R.J.; Larose, L.; et al. Nck-dependent activation of extracellular signal-regulated kinase-1 and regulation of cell survival during endoplasmic reticulum stress. *Mol. Biol. Cell* **2004**, *15*, 4248–4260. [CrossRef]
111. Li, C.; Li, L.; Yang, M.; Zeng, L.; Sun, L. PACS-2: A key regulator of mitochondria-associated membranes (MAMs). *Pharmacol. Res.* **2020**, *160*, 105080. [CrossRef] [PubMed]
112. Bollo, M.; Paredes, R.M.; Holstein, D.; Zheleznova, N.; Camacho, P.; Lechleiter, J.D. Calcineurin interacts with PERK and dephosphorylates calnexin to relieve ER stress in mammals and frogs. *PLoS ONE* **2010**, *5*, e11925. [CrossRef]
113. Geiss-Friedlander, R.; Melchior, F. Concepts in sumoylation: A decade on. *Nat. Rev. Mol. Cell Biol.* **2007**, *8*, 947–956. [CrossRef] [PubMed]
114. Feldhammer, M.; Uetani, N.; Miranda-Saavedra, D.; Tremblay, M.L. PTP1B: A simple enzyme for a complex world. *Crit. Rev. Biochem. Mol. Biol.* **2013**, *48*, 430–445. [CrossRef]
115. Dudek, E.; Millott, R.; Liu, W.X.; Beauchamp, E.; Berthiaume, L.G.; Michalak, M. N-Myristoyltransferase 1 interacts with cal-nexin at the endoplasmic reticulum. *Biochem. Biophys. Res. Commun.* **2015**, *468*, 889–893. [CrossRef] [PubMed]
116. Hunegnaw, R.; Vassylyeva, M.; Dubrovsky, L.; Pushkarsky, T.; Sviridov, D.; Anashkina, A.A.; Üren, A.; Brichacek, B.; Vassylyev, D.G.; Adzhubei, A.A.; et al. Interaction Between HIV-1 Nef and Calnexin: From Modeling to Small Molecule In-hibitors Reversing HIV-Induced Lipid Accumulation. *Arterioscler. Thromb. Vasc. Biol.* **2016**, *36*, 1758–1771. [CrossRef]
117. Myrum, C.; Soulé, J.; Bittins, M.; Cavagnini, K.; Goff, K.; Ziemek, S.K.; Eriksen, M.S.; Patil, S.; Szum, A.; Nair, R.R.; et al. Arc Interacts with the Integral Endoplasmic Reticulum Protein, Calnexin. *Front. Cell Neurosci.* **2017**, *11*, 294. [CrossRef] [PubMed]
118. Jung, J.; Wang, J.; Groenendyk, J.; Lee, D.; Michalak, M.; Agellon, L.B. Fatty acid binding protein (Fabp) 5 interacts with the calnexin cytoplasmic domain at the endoplasmic reticulum. *Biochem. Biophys. Res. Commun.* **2017**, *493*, 202–206. [CrossRef]
119. Jennelle, L.; Hunegnaw, R.; Dubrovsky, L.; Pushkarsky, T.; Fitzgerald, M.L.; Sviridov, D.; Popratiloff, A.; Brichacek, B.; Buk-rinsky, M. HIV-1 protein Nef inhibits activity of ATP-binding cassette transporter A1 by targeting endoplasmic reticulum chaperone calnexin. *J. Biol. Chem.* **2014**, *289*, 28870–28884. [CrossRef]
120. Waheed, A.A.; Freed, E.O. Lipids and membrane microdomains in HIV-1 replication. *Virus Res.* **2009**, *143*, 162–176. [CrossRef]
121. Jung, J.; Eggleton, P.; Robinson, A.; Wang, J.; Gutowski, N.; Holley, J.; Newcombe, J.; Dudek, E.; Paul, A.M.; Zochodne, D.; et al. Calnexin is necessary for T cell transmigration into the central nervous system. *JCI Insight* **2018**, *3*, 98410. [CrossRef] [PubMed]
122. Rao, E.; Singh, P.; Li, Y.; Zhang, Y.; Chi, Y.I.; Suttles, J.; Li, B. Targeting epidermal fatty acid binding protein for treatment of experimental autoimmune encephalomyelitis. *BMC Immunol.* **2015**, *16*, 28. [CrossRef] [PubMed]